Journal of Multivariate Analysis 212 (2026) 105528

Contents lists available at ScienceDirect

Journal of Multivariate Analysis

journal homepage: www.elsevier.com/locate/jmva

Recent advances in principal component analysis for directional data☆

Anahita Nodehi [a,*], Meisam Moghimbeygi [b], Christophe Ley [c,*]

[a] Population Health Sciences, Bristol Medical School, University of Bristol, Bristol, United Kingdom
[b] Department of Mathematics, Faculty of Mathematics and Computer Science, Kharazmi University, Tehran, Iran
[c] Department of Mathematics, Faculty of Science, Technology and Medicine, University of Luxembourg, Esch-sur-Alzette, Luxembourg

ARTICLE INFO

AMS 2020 subject classifications:
62H11
62H25

Keywords:
Non-Euclidean spaces
Principal component analysis
Riemannian manifolds
Spheres
Torus

ABSTRACT

The high dimensionality of the input data can pose multiple problems when implementing statistical techniques. The presence of many dimensions in the data can lead to challenges in visualizing the data, higher computational demands, and a higher probability of over-fitting or under-fitting in modeling. Furthermore, the curse of dimensionality contributes to these issues by stating that the necessary number of observations for accurate modeling increases exponentially as the number of dimensions increases. Dimension reduction tools help overcome this challenge. Principal Component Analysis (PCA) is the most widely used technique, intensively studied in classical linear spaces. However, in applied sciences such as biology, bioinformatics, astronomy and geology, there are many instances in which the data's support are non-Euclidean spaces. In fact, the available data often include elements of Riemannian manifolds such as the unit circle, torus, sphere, and their extensions. Therefore, the terms "manifold-valued" or "directional" data are used in the literature for these situations. When dealing with directional data, the linear nature of PCA might pose a challenge to achieve accurate data reduction. This paper therefore reviews and investigates the methodological aspects of PCA on directional data and their practical applications.

1. Introduction

Nowadays statisticians and data analysts face a significant challenge in analyzing large and complex datasets, typically known as "big data". As a result, significant research efforts have focused on high-dimensional challenges in the past several decades. Richard E. Bellman proposed the term "curse of dimensionality" to describe the challenges encountered in dynamic programming, as mentioned in his work [1]. It refers to the difficulties of organizing and analyzing data in high-dimensional spaces. In this sense, feature selection and extraction are employed to address problems resulting from the "curse of dimensionality" [2,3]. Feature selection aims to select some variables from the entire set of variables based on their relevance. Such feature selection algorithms can be categorized into four groups: filter methods (such as ANOVA, Pearson correlation, and variance thresholding), wrapper methods (including forward, backward, and stepwise selection), embedding methods (such as Lasso, Ridge, and Decision Tree), and hybrid methods (as mentioned by [4]). Feature extraction on the other hand involves reducing the number of dimensions by creating a smaller collection of linear or non-linear transformations of the input variables. The primary distinction between feature selection and extraction is rooted in the fact that the former retains a subset of the original features, while the latter creates new ones [2].

☆ This article is part of a Special issue entitled: 'JMVA_Dimension Reduction in Multivariate Analysis' published in Journal of Multivariate Analysis.

* Corresponding authors.
E-mail addresses: anahita.nodehi@bristol.ac.uk (A. Nodehi), christophe.ley@uni.lu (C. Ley).

https://doi.org/10.1016/j.jmva.2025.105528
Received 17 June 2024; Received in revised form 7 August 2025; Accepted 7 November 2025

Available online 15 November 2025

Principal Component Analysis (PCA), also known as proper orthogonal decomposition (POD), is a famous method in feature extraction to overcome difficulties caused by the curse of dimensionality and has attracted much attention in the literature. Pearson [5] was the first to introduce it and Hotelling [6] developed it independently. PCA may be described, according to Hotelling [6], as the *orthogonal* projection of data onto a lower dimension's linear space, called the principal subspace, so that the variance of the projected data is maximized. On the other hand, according to Pearson [5], PCA is described as the *linear* projection that minimizes the average projection cost, defined as the mean squared distance between the data points and their projections. It also creates a linear transformation of variables such that newly combined variables are almost uncorrelated [5]. In some situations, concerns are related to the use of PCA. For example, when there are outliers in the data, PCA is strongly impacted, which led to the development of robust PCA in computer vision for images that can be corrupted by noisy pixels. And with non-Euclidean data sets, PCA's linearity may be an obstacle to data reduction and compression [3].

The statistical field of directional statistics was explicitly developed for use with observations that are directions supported on non-Euclidean spaces [7]. A direction, such as the wind direction, observed in the plane $\mathbb{R}^2$, can be represented by an angle θ, often in the intervals $[0, 2\pi)$ or $[-\pi, \pi)$ and measured in a specified direction from a specified origin, or by the unit vector $x = (\cos\theta, \sin\theta)^\top$ for which $\|x\| = \sqrt{x^\top x} = 1$. These types of data are referred to as *circular data* because the natural support for such directions is the circumference of the unit circle $\mathbb{S}^1$. The term circular data is also used to distinguish them from data with the real line $\mathbb{R}$ (or some subset of it) as their support, which henceforth we will term as linear data. Observations made in directions in $\mathbb{R}^3$ can be represented by pairs of angles that have as natural support the unit sphere, $\mathbb{S}^2$ (known as *spherical data*), or the unit torus, $\mathbb{T}^2 = \mathbb{S}^1 \times \mathbb{S}^1$ (referred to as *toroidal data*). These forms of data are also known as "non-Euclidean data" or "manifold-valued data" because the natural supports for the data under consideration are Riemannian manifolds like the unit circle, torus, sphere, and their extensions. Shape analysis [8–10] is a crucial related field with spherical data where a pre-shape corresponding to a configuration of $k \geq 0$ landmarks in $\mathbb{R}^d$ can be regarded as a point on $\mathbb{S}^{d(k-1)-1}$, $d \in \mathbb{N}$.

In the literature, there are many books that cover numerous aspects of directional statistics, including [7,11–19]. In addition, some articles on the topic review direction statistics. To name two essential ones, Jupp and Mardia [20] conducted a comprehensive review that covered the vast majority of works on directional statistics between 1975 and 1988, and recently Pewsey and García-Portugués [21] provided a comprehensive overview of the overall developments in directional statistics over the last two decades. Despite all these efforts, so far no paper has covered dimension reduction techniques, particularly manifold-valued PCA, in depth. Of course, this is also due to the recency of many of the essential results in that domain. For example, a concise section (8.1) in Pewsey and García-Portugués [21] discusses PCA. In shape analysis, Huckemann and Eltzner [22,23] reviewed statistical methods generalizing PCA to non-Euclidean spaces. Nevertheless, Huckemann and Eltzner [22,23] does not cover other forms of manifold-valued data such as toroidal data. David and Jacobs [24] and Kitao [25] reviewed the methodological aspects of PCA and related methods and their applications to investigate protein dynamics. Given the absence of a detailed treatment of the topic, and in view of the numerous recent developments regarding PCA for directional data, we believe that a review of the latest progress in PCA-related research for directional data is essential.

In this paper, we will provide an overview of PCA, related methods, and their applications in manifold-valued data sets. Furthermore, some basic concepts and essential findings obtained during the early years of this field are revisited for the benefit of non-experts, as well as a review of the latest progress in PCA-related research. The remainder of this paper is organized as follows. Section 2 describes the classical Euclidean PCA (Section 2.1) and provides guidance for its extension to general manifolds. This is achieved by reviewing essential concepts of non-Euclidean spaces and manifolds (Sections 2.2–2.3). Then, Section 3 presents developments for geodesic-based and tangent-based PCA. Sections 4 and 5 review various extensions of PCA for spherical and toroidal data, respectively. This is followed by Section 6, which contains PCA extensions for particular spaces, such as shape and symmetric spaces. In addition, Sections 7 and 8 provide a brief overview of available R packages and conclude with final remarks, respectively. Further results and a comparative example of PCA-related methods are included in the Supplementary Materials.

2. Towards PCA on general manifolds

Before exploring manifold-valued PCA, it is important to review the classical Euclidean PCA and some fundamental concepts of manifolds, including their intrinsic and extrinsic means.

2.1. Principal component analysis (PCA)

PCA is a widely used technique in lossy data compression, feature extraction, and data visualization [26]. It can be described as the problem of fitting a low-dimensional affine subspace to a set of data points in a high-dimensional space [5,6].

Consider a random d-vector $\mathbf{X} = (X_1, \ldots, X_d)^\top$. PCA seeks to substitute the collection of d (unordered and correlated) input variables, $X_1, X_2, \ldots, X_d$, with a smaller set of $p \leq d$ (ordered and uncorrelated) linear projections $Y_1, \ldots, Y_p$ of the input variables under the form

$$\boldsymbol{Y}_j = \boldsymbol{b}_j^\top \mathbf{X} = b_{j1}X_1 + \cdots + b_{jd}X_d, \ j \in \{1, \ldots, p\}, \tag{2.1}$$

in such a way that we minimize the loss of information (defined as the *total variation* of the original input variables). The linear projections (2.1) are known as the first m principal components of $\boldsymbol{X}$. The jth coefficient vector, $\boldsymbol{b}_j = (b_{1j}, \ldots, b_{dj})^\top$, is chosen so that:

- The first p linear projections $\boldsymbol{Y}_j$ (for $j \in \{1, \ldots, p\}$) of $\mathbf{X}$ are ranked in order of importance through their variances $\mathrm{var}(Y_j)$, which are listed in decreasing order of magnitude: $\mathrm{var}(Y_1) \geq \mathrm{var}(Y_2) \geq \cdots \geq \mathrm{var}(Y_p)$.
- $\boldsymbol{Y}_j$ is uncorrelated with $\boldsymbol{Y}_k$ for $k \neq j$.

The following theorem shows that the principal components of $\boldsymbol{X}$ can be computed from the eigenvectors of its covariance matrix $\boldsymbol{\Sigma}_X$, a $d \times d$ symmetric real-valued matrix.

Theorem 2.1 (*Principal Components of a Random Variable [27]*)**.** *Assume that* $\mathrm{rank}\,(\boldsymbol{\Sigma}_X) \geq p$*. Then, the first* p *principal components of a zero-mean multivariate random vector* $\boldsymbol{X}$*, denoted by* $\boldsymbol{Y}_j$ *for* $j \in \{1, \ldots, p\}$*, are given by* $\boldsymbol{Y}_j = \boldsymbol{b}_j^\top \boldsymbol{X}$ *where* b_j *are* p *orthonormal eigenvectors of* $\boldsymbol{\Sigma}_X$ *associated with its* p *largest eigenvalues* λ_j*. Moreover,* $\lambda_j = \mathrm{var}(\boldsymbol{Y}_j)$ *for* $j \in \{1, \ldots, p\}$.

2.2. Manifold background

As indicated by Karcher [28], a non-Euclidean space can be distinguished by several unique bases. Consequently, the distance between the points is a function of the bases. Therefore, PCA based on these distances, while ignoring the topological characteristics of the non-Euclidean space, can be invoked. Directional data live in non-Euclidean spaces, and therefore we use nonlinear statistics [29] to analyze this kind of data. In other words, the data under consideration are elements of a *manifold* and the term manifold-valued statistics is also used in the literature, particularly by Karcher [28]. To explore the notions of measure of centrality such as mean and median and/or of variability, we review some relevant concepts of manifolds based on Karcher [28], Boothby [30], Fletcher et al. [31], Pennec [29] and Bishop and Crittenden [32].

Riemannian manifold. To define a Riemannian manifold, it is essential that we take into account a set of fundamental definitions. These foundational concepts provide the framework for our subsequent formal delineation of a Riemannian manifold.

Definition 2.1 (*Manifold*)**.** A manifold M of dimension $d \in \mathbb{N}$, or d-manifold, is a topological space with the following properties:

i. M is Hausdorff (i.e., if any two distinct points in M can be separated),
ii. M is locally Euclidean of dimension d, and
iii. M has a countable basis of open sets such that for each point $x \in M$ there is a neighborhood U of x that is homeomorphic to $\mathbb{R}^d$.

In a manifold, every point $x \in M$ has a neighborhood U with a diffeomorphism $\phi : U \longrightarrow \mathbb{R}^d$. The pair (U, ϕ) is called a coordinate chart, and a collection of charts whose domains cover M is called an atlas.

Definition 2.2 (*Smooth Manifold*)**.** A smooth manifold is a topological space M together with a maximal atlas. The maximal atlas is also called a differentiable structure on M.

Lines and circles are one-dimensional manifolds. Manifolds in two dimensions are also known as surfaces. The plane, the sphere, and the torus are examples. The concept of a manifold is fundamental to numerous aspects of geometry and non-Euclidean spaces because it allows the characterization of complex structures in terms of the well-understood topological properties of simpler spaces. According to Karcher [28], for each manifold $M \subset \mathbb{R}^d$, it is possible to associate a linear subspace of $\mathbb{R}^d$ to each point $x \in M$. Such space is called the *tangent space* at x, and is written as $T_x M$.

Definition 2.3 (*Tangent Space and Metric*)**.** The tangent space $T_x M$ is the set of tangent vectors of all curves passing through the point $x \in M$. Intuitively, it is a linear subspace that best approximates the manifold M in a neighborhood of the point x. It is equipped with an inner product

$$g_x : T_x M \times T_x M \longrightarrow \mathbb{R}$$

along with a norm metric $\|\cdot\| : T_x M \to \mathbb{R}$ defined by $\|v\| := \sqrt{g_x(v, v)}$ for any $v \in T_x M$. In the literature, the metric is sometimes referred to by brackets $\langle \cdot, \cdot \rangle$ instead of $g_x(\cdot, \cdot)$.

Definition 2.4 (*Riemannian Manifold*)**.** A Riemannian manifold M is a smooth manifold M equipped with a metric g_x, which is a smoothly varying inner product on the tangent space $T_x M$ of M at each point $x \in M$. The family g_x of inner products is called a Riemannian metric.

It is a well-established fact that any Riemannian manifold M can be embedded isometrically into some Euclidean space with sufficiently high dimension. The dimensional requirements vary based on the nature of the embedding; for continuously differentiable embeddings (C^1) [33] and analytic or smooth embeddings of a particular class (C^k, $k \geq 3$) [34], different upper bounds are considered. At times, a suitable non-isometric embedding $M \hookrightarrow \mathbb{R}^d$ may be preferred. It is important to note that in the case of an isometric embedding, the Riemannian metric g_x on $T_x M$ corresponds to the standard Euclidean inner product.

Definition 2.5 (*Riemannian Distance*)**.** The Riemannian distance between two points, $a, b \in M$, denoted $\mathrm{d}(a, b)$, is defined as the minimum length over all possible smooth curves between a and b.

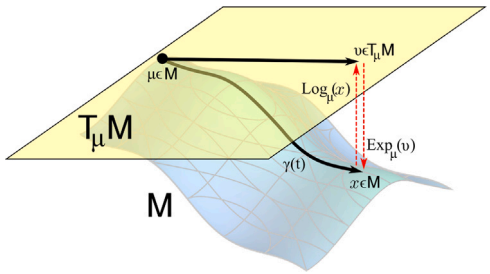

Fig. 1. Illustration of a geodesic $\gamma(t)$ constructed between two points $\mu, x \in M$, along with the associated tangent space vector $v \in T_\mu M$. The exponential map Exp_μ and the logarithmic map Log_μ are also shown.

The length of the manifold's shortest curve between a pair of points, known as a *geodesic*, is defined as follows.

Definition 2.6 (*Geodesics and its Curve*)**.**

(i) A differentiable or smooth curve $\gamma : [a, b] \to M$ assigns to each $t \in (a, b),\ a, b \in \mathbb{N}$, a tangent vector $\boldsymbol{\gamma}'(t) \in T_{\boldsymbol{\gamma}(t)}M$. Its length, $L(\boldsymbol{\gamma})$, is computed by integrating the norm of the tangent vectors along the curve (i.e., arc length):

$$L(\boldsymbol{\gamma}) := \int_a^b \|\boldsymbol{\gamma}'(t)\| dt.$$

(ii) Given a couple of points on a Riemannian manifold $(p, x) \in M$ and the set of all curves $\boldsymbol{\gamma} : [a, b] \to M$ such that $\gamma(a) = p$ and $\gamma(b) = x$, the *geodesic curve* $\bar{\gamma}$ is the curve with the shortest total length $L(\bar{\gamma})$ as

$$L(\bar{\gamma}) := \inf\{L(\gamma) \mid \boldsymbol{\gamma} : [a, b] \to M, \text{with } \gamma(a) = p, \gamma(b) = x\}.$$

In other words, $\bar{\gamma}$ locally minimizes the length between two points.

(iii) A Riemannian manifold M is said to be complete if every geodesic segment $\boldsymbol{\gamma} : [a, b] \to M$ can be extended to a geodesic from all of $\mathbb{R}$ to M.

Theorem 2.2 (*Hopf–Rinow Theorem [35]*)**.** *If M is a complete, connected Riemannian manifold, then the distance metric $\mathrm{d}(\cdot, \cdot)$ induced on M is complete. Furthermore, between any two points on M, there exists a geodesic.*

This result shows that in the complete manifold M, each geodesic extends indefinitely in both directions (even though it wraps around itself). This implies that between any two points, there exists a length-minimizing geodesic. For instance, the Euclidean space $\mathbb{R}^d$ and the sphere $\mathbb{S}^2$ are complete, and thus by the Hopf–Rinow Theorem 2.2, there is a geodesic (for antipodal points on $\mathbb{S}^2$ the geodesic is not unique in this case).

The tangent space of a Riemannian manifold M and geodesics γ have direct relationships through the *exponential map* and the *logarithmic map*. For a tangent vector $v \in T_x M$, constructed at point $x \in M$, there exists a unique geodesic $\gamma_v : [0, 1] \to M$ such that $\gamma_v(0) = x$ and with $\gamma_v'(0) = v$ as initial velocity. The vector $v \in T_\mu M$ is mapped to the endpoint of the geodesic $\gamma_v(1) \in M$ through the exponential map $\mathrm{Exp}_x : T_x M \to M$ such that $\mathrm{Exp}_x(v) = \gamma_v(1)$. The inverse of the exponential map is the *logarithmic map* $\mathrm{Log}_x : M \to T_x M$ such that $\mathrm{Log}_x(\gamma_v(1)) = v$, which maps the point $\gamma_v(1)$ in the neighborhood of $x \in M$ to a vector in the tangent space $v \in T_x M$. These functions provide a mapping from the manifold's surface to the tangent space of a given point. Fig. 1 gives a schematic representation of these maps.

2.3. Means on manifolds

While various measures of a central point exist in the literature, including the Fermat–Weber point [36], the projection median [37], and the Tukey center point [38], which are all variants of the geometric median, this paper focuses specifically on the mean as the central point.

The PCA approaches on manifolds can be classified as *extrinsic* (based on the tangent space) or *intrinsic* (geodesic-based). In an extrinsic approach, the manifold is typically approximated by a linearized space, and the majority of the analysis is performed there before the results are projected back. In the intrinsic approach, no approximation procedures are used, but instead mathematical tools are applied directly to the Riemannian manifold [29]. Taking terminology from stepwise multiple linear regression, alternative approaches to classify manifold PCA is *backward* versus *forward* stepwise PCA [39,40]. Typically, in the *forward* approach, PCA and approximating subspaces are constructed from lower to higher dimensions, while PCA is constructed in reverse order from the largest dimension to the lowest dimension in the *backward* approach [41]. In Euclidean data sets, forward and backward PCA are equivalent. For further details, the interested reader is referred to Marron et al. [39], Damon and Marron [40], Marron and Alonso [42], and Pennec [43].

The distances on a manifold play a key role in defining some measures of centrality and dispersion in directional statistics. Thus, after Definition 2.5 of a distance between two points, extending statistical methods to manifolds is straightforward.

Definition 2.7 (*Intrinsic Mean, Fréchet [44], Karcher [28]*)**.** Let $x_1, x_2, \ldots, x_n \in M$ be a data set on a manifold M. The *intrinsic mean* μ is defined as

$$\mu := \arg\min_{a \in M} \sum_{i=1}^{n} \mathsf{d}(a, x_i)^2, \tag{2.2}$$

where $\mathsf{d}(\cdot,\cdot)$ denotes the Riemannian distance on M.

In a study conducted by Fréchet [44], the *intrinsic mean* is defined for a general metric space. Accordingly, it is sometimes called the Fréchet mean. The properties of the intrinsic mean on a Riemannian manifold have been studied by Karcher [28]. Kendall [45] shows that the intrinsic mean exists and is unique if the data are well-localized. For our case, having the intrinsic mean μ, $T_\mu M$ will be the tangent space to the intrinsic mean. This construction is particularly useful when the statistical analysis cannot be invoked directly on the manifold. In that case, the tangent space in the vicinity of the intrinsic mean is a suitable place to carry out those required analyses and project the results back. However, there are numerous benefits to employing mathematical tools directly in the Riemannian manifold rather than employing any approximation procedure [29].

The intrinsic mean is computed by solving the minimization problem in (2.2). Let us assume that our data $x_1, \ldots, x_n$ lie in a strongly convex neighborhood (i.e., a neighborhood U such that any two points in U are connected by a unique geodesic contained completely within U) to guarantee a unique solution. We must minimize the sum-of-squared distance function

$$f(x) = \frac{1}{2n} \sum_{i=1}^{n} \mathsf{d}(x, x_i)^2, x \in M.$$

Pennec [46] proposed a gradient descent algorithm to minimize f (for more details about gradient algorithms, see [47,48]). The gradient descent algorithm takes successive steps in the negative gradient direction. Given a current estimate $\mu_j, j \geq 1$, for the intrinsic mean, the following equation can be used to update the mean by taking a step in the negative gradient direction as

$$\mu_{j+1} := \mathrm{Exp}_{\mu_j}\left(\frac{k}{n} \sum_{i=1}^{n} \mathrm{Log}_{\mu_j}(x_i)\right),$$

where $k \in \mathbb{N}$ is the step size. Since the gradient descent algorithm only converges locally and the data is assumed to be well-localized, Fletcher et al. [31] recommended considering initial estimate of μ_0 as one of the data points, say x_1, and $k = 1$ for spherical data. It is important to note that if M is a vector space, then the gradient descent algorithm with $k = 1$ simplifies to linear averaging, resulting in convergence in just one step. Additionally, when $M = \mathbb{R}^+$, which represents the Lie group of positive real numbers under multiplication, the algorithm with $k = 1$ corresponds to the geometric average, also leading to convergence in a single step.

The gradient-based approach for calculating means on a manifold, such as that described by Pennec [46], generally demonstrates linear convergence under certain conditions. However, Cazals et al. [49] provide further insights into convergence behavior specifically on the unit circle $\mathbb{S}^1$ when computing the Fréchet mean. By partitioning $\mathbb{S}^1$ into arc intervals and performing structured calculations within these intervals, they achieve efficient convergence even for large sample sizes. Additionally, Cazals et al. [49] address the challenge of local minima by analyzing the mean's behavior on $\mathbb{S}^1$ through interval decomposition, ensuring that each interval contains at most one local minimum. Using a combination of algebraic and transcendental expressions to assess derivatives, their approach reliably identifies intervals where local minima may occur. In summary, there is the following algorithm for computing the intrinsic mean of manifold data:

Algorithm 1 Intrinsic Mean

Input: $x_1, \ldots, x_n \in M$
Output: $\mu \in M$, the intrinsic mean
$\mu_0 = x_1$, the initial value;
Do
$\Delta_\mu = \frac{k}{n} \sum_{i=1}^{n} \mathrm{Log}_{\mu_j}(x_i)$
$\mu_{j+1} = \mathrm{Exp}_{\mu_j}(\Delta\mu)$
while $||\Delta_\mu|| > \epsilon$. [where $||\cdot||$ denotes the Euclidean norm]

Another way of defining the distance on M is to embed it in a Euclidean space and use the Euclidean distance between points. This notion is called the *extrinsic distance* on M. Similarly to the above, we can define the extrinsic mean, which depends on the ambient space and the choice of embedding.

Definition 2.8 (*Extrinsic Mean*)**.** Given an embedding $\Phi : M \to \mathbb{R}^d$, the *extrinsic mean* μ_Φ of a collection of points $x_1, x_2, \ldots, x_n \in M$ is defined as

$$\mu_\Phi := \arg\min_{a \in M} \sum_{i=1}^{n} \|\Phi(a) - \Phi(x_i)\|^2, \tag{2.3}$$

where $\|\cdot\|$ denotes the Euclidean norm.

Given the embedding of M, we can also calculate the arithmetic (Euclidean) mean of the embedded points and then project this mean onto the manifold M. This projected mean is equivalent to the above definition of the extrinsic mean [50]. On the issues relating to the existence and uniqueness of μ_Φ, see Bhattacharya and Patrangenaru [51].

3. Geodesic-based and tangent-based PCA

In the past few decades, two broad categories of PCA on manifolds have been developed: *tangent-based PCA* and *geodesic-based PCA*. The former seeks to exploit manifolds' locally Euclidean property (see Fletcher et al. [31]), whereas the latter removes the constraint of geodesics passing through the mean (see Huckemann and Ziezold [52]). Geodesic-based and tangent-based PCA have been applied to many problems, including statistical shape analysis [10,31,53–55], diffeomorphic image registration, Vaillant et al. [56], tensor field classification in medical image analysis [57,58], wrist rotation in Robotics [59] and neuroimaging data [60].

3.1. Principal geodesic analysis (PGA)

Fletcher et al. [31] introduced Principal geodesic analysis (PGA), which is a generalization of PCA to manifolds. More precisely, they described the variability of the manifold data in a way analogous to PCA. For this, they used the Riemannian property of the non-Euclidean space to derive the eigenvectors and eigenvalues on the manifold through the exponential and logarithmic maps. In PGA, the *forward* approach is taken to PCA using geodesics in the appropriate manifolds. In other words, the PGA employs an appropriate mapping to connect the manifold and tangent space. It utilizes the tangent space of the manifold at the geodesic mean as the linear space. This results in finding the best-fitting geodesic among those passing through the geodesic mean.

PGA aims to represent variability in data on a manifold, similarly to how PCA represents variability in Euclidean space. This requires addressing the following three key concepts in the manifold setting:

- **Variance:** Fréchet [44] stated that sample variance is the expected value of the squared Riemannian deviation from the mean. Given a set of points $x_1, x_2, \dots, x_n$ in a complete connected manifold M, the sample variance of the data is defined as

$$\sigma^2 := \frac{1}{n}\sum_{i=1}^{n} \mathsf{d}(\mu, x_i)^2,$$

  where μ is the intrinsic mean of the x_i.
- **Geodesic sub-manifolds:** A sub-manifold H of M is said to be geodesic at $x \in H$ if all geodesics of H passing through x are also geodesics of M [31].
- **Projection:** The projection of a point $x \in M$ onto a geodesic submanifold H of M is defined as the point on H that is nearest to x in Riemannian distance. Thus, we define the projection operator $\pi_H : M \to H$ as

$$\pi_H(x) := \operatorname*{argmin}_{y \in H} \mathsf{d}(x, y)^2.$$

Due to the computational cost of non-linear least-squares optimization for the geodesic subspace, and the absence of a closed-form solution for the projection operator π_H on general manifolds, we employ a linear approximation within the tangent space of M for projecting onto a geodesic submanifold. This approximation, implemented in Algorithm 2, provides a practical approach to PGA. The algorithm's core principle, minimizing residual norms or unexplained variance, aligns with the objective of manifold dimensionality reduction.

Algorithm 2 Principal Geodesic Analysis

Input: $x_1, \dots, x_n \in M$
Output: Principal directions, $v_k \in T_\mu M$
$\lambda_k \in \mathbb{R}$
μ = intrinsic mean of x_i (Algorithm 1)
$u_i = \mathrm{Log}_\mu(x_i)$
$\mathbf{S} = \frac{1}{n}\sum_{i=1}^{n} u_i u_i^\top$
$\{v_k, \lambda_k\}$ = eigenvectors/eigenvalues of $\mathbf{S}$

Sommer et al. [61] developed a method for calculating the exact PGA in a wide range of manifolds. Specifically, they proposed an algorithm for numerically solving optimization problems over geodesic spaces by integrating Jacobi fields and second-order geodesic derivatives. They investigated linearized and exact algorithms for computing PGA by applying both algorithms to synthetic data.

3.2. Geodesic principal component analysis (GPCA)

The PGA method involves analyzing geodesic subspaces, and it has been noted that the mean value is always found within these subspaces, even if it falls outside the expected range of values. Huckemann and Ziezold [52] and Huckemann et al. [62] have proposed Geodesic principal component analysis (GPCA) whereby the first-order component of the analysis is initiated with the geodesic that best fits the data, without necessarily adhering to the mean value. This alternative approach can help to improve the accuracy of the PGA procedure when working with data that lie beyond the expected distribution range. In that respect, they introduced a new notion of the center point, the PC's mean, which is an intersection of the first two principal geodesics. The main improvement of GPCA over PGA is that the best fitting geodesic is drawn from the set of all geodesics instead of being constrained to go through the geodesic mean [62]. Thus, it falls within the framework of a *backward* approach to PCA.

To expand on the existing framework that relaxes the assumption that second and higher-order components should cross at a single point, Sommer [63] has suggested an alternative method for defining these components. Specifically, Sommer [63] has proposed a parallel transport of the second direction along the first principal geodesic, which is used to establish the second coordinates. This process is then iteratively repeated to define higher-order coordinates by horizontally developing along the previous modes. The method was named Horizontal component analysis (HCA). By adopting this, it is possible to achieve a more flexible method for defining principal components, which can improve the accuracy of subsequent analyses.

3.3. Probabilistic principal geodesic analysis (PPGA)

PGA is an extension of PCA that is designed to work with non-linear manifolds [31]. The method is founded on geometric principles and utilizes least-squares estimation techniques to determine subspaces that minimize the sum-of-squared geodesic distances to the available data. Although the original PCA formulation inspired the development of PGA, the current component analysis techniques for manifolds do not possess a probabilistic interpretation, similar to PCA.

Tipping and Bishop [64] introduced a latent variable model for PCA called probabilistic PCA (PPCA). Roweis also proposed a similar formulation [65]. Inspired by probabilistic PCA, a latent variable model for PGA was presented in a paper by Zhang and Fletcher [66] called Probabilistic PGA (PPGA).

The main idea of probabilistic PCA is to model a d-dimensional Euclidean random vector $\boldsymbol{y}$ as

$$\boldsymbol{y} = \boldsymbol{\mu} + \boldsymbol{B}\boldsymbol{x} + \boldsymbol{\epsilon},$$

where $\mu \in \mathbb{R}^d$ is the mean of $\boldsymbol{y}$, $\boldsymbol{x}$ is a p-dimensional latent random vector ($p < d$) which follows a normal distribution with mean zero and identity covariance matrix $\boldsymbol{I}_p$, the factor matrix $\boldsymbol{B}$ is a $d \times p$ matrix that relates $\boldsymbol{x}$ and $\boldsymbol{y}$, and the error is represented by $\boldsymbol{\epsilon}$, which follows a normal distribution with mean zero and covariance matrix $\sigma^2 \boldsymbol{I}_d$. To simplify the factor matrix, it is common to consider $\boldsymbol{B} = \boldsymbol{W}\boldsymbol{\Lambda}$, where the columns of $\boldsymbol{W}$ are mutually orthogonal and $\boldsymbol{\Lambda}$ is a diagonal matrix of scale factors. Tipping and Bishop [64] demonstrated how the principal axes of a set of observed data vectors may be determined through maximum likelihood estimation of parameters in a latent variable model closely related to factor analysis. They introduced two methods to estimate the parameters: matrix decomposition and the EM algorithm.

To apply the model to random variables on Riemannian manifolds, Zhang and Fletcher [66] used a noise model based on the normal distribution for such manifolds [29]. Consider a random vector $\boldsymbol{y}$ that takes values on a Riemannian manifold M and that is defined by a probability density function expressed as

$$\Pr(\boldsymbol{y} \mid \boldsymbol{\mu}, \tau) := \frac{1}{C(\boldsymbol{\mu}, \tau)} \exp\left(-\frac{\tau}{2}\mathrm{d}(\boldsymbol{\mu}, \boldsymbol{y})^2\right),$$

$$C(\boldsymbol{\mu}, \tau) = \int_M \exp\left(-\frac{\tau}{2}\mathrm{d}(\boldsymbol{\mu}, \boldsymbol{y})^2\right) dy,$$

where $C(\boldsymbol{\mu}, \tau)$ is the normalizing constant depending on $\boldsymbol{\mu} \in M$ which is a location parameter on the manifold, and $\tau \in \mathbb{R}^+$ is a dispersion parameter that is similar to the precision of a Gaussian distribution. We term this distribution a *Riemannian normal distribution* and use the notation $\boldsymbol{y} \sim N_M(\boldsymbol{\mu}, \tau^{-1})$ to denote it. Given that there is no explicit formula for the normalizing constant, this estimation is only applicable to symmetric spaces such as Euclidean space, spheres, Kendall shape spaces, Grassman or Stiefel manifolds, and so on. Hence, the PPGA model for a random variable y on a smooth Riemannian manifold M is based on

$$\boldsymbol{y} \mid \boldsymbol{x} \sim N_M(\mathrm{Exp}(\boldsymbol{\mu}, \boldsymbol{z}), \tau^{-1}), \ \boldsymbol{z} = \boldsymbol{B}\boldsymbol{x},$$

where $\boldsymbol{x}$ represents a set of latent random variables in $\mathbb{R}^p$, μ is a base point on M and τ is a scale parameter for the noise. In this model, both $\boldsymbol{B}$ (a linear combination of $\boldsymbol{W}$ and $\boldsymbol{\Lambda}$) and the latent variables $\boldsymbol{x}$ form a new tangent vector $\boldsymbol{z} \in T_\mu M$. Subsequently, the exponential map is used to move the base point μ along the tangent vector $\boldsymbol{z} \in T_\mu M$, resulting in the generation of the location parameter for a Riemannian normal distribution. It is worth noting that in Euclidean space, the exponential map corresponds to simple addition, i.e., $\mathrm{Exp}(\boldsymbol{\mu}, \boldsymbol{z}) = \boldsymbol{\mu} + \boldsymbol{z}$ which coincides with the PPCA model [66].

Zhang and Fletcher [66] proposed a Monte Carlo Expectation Maximization (MCEM) algorithm to estimate the parameters in the maximum likelihood function. The expectation in this algorithm cannot be computed in closed form and is instead approximated by Hamiltonian Monte Carlo (HMC) sampling of the latent variables. The MCEM algorithm is used to compute the subspace spanned by d principal components in PPGA. This algorithm involves iterative steps and the time required for each step depends on the complexity of certain mathematical concepts such as the exponential map, log map, and Jacobi field. The computational cost of the gradient algorithm increases with the data size, dimensionality, and the number of samples drawn. Nonetheless, MCEM can work simultaneously for each data point, thus making it a parallelizable algorithm.

Suppose, we have observed data y_i, $i \in \{1, \ldots, n\}$, on a Riemannian manifold M, where each data point y_i is associated with a corresponding p-dimensional latent variable $x_i \in \mathbb{R}^p$ and $\boldsymbol{z}_i = \mathbf{W}\boldsymbol{\Lambda}x_i$ where $\boldsymbol{z}_i \in T_\mu M$. To sample from the posterior distribution $\Pr(x \mid y; \boldsymbol{\eta})$, where the parameters are denoted as $\boldsymbol{\eta} = (\boldsymbol{\mu}, \boldsymbol{W}, \boldsymbol{\Lambda}, \tau)$, using the HMC method, we compute the log of the posterior distribution

$$\ln \prod_{i=1}^{n} \Pr(x_i \mid y_i; \boldsymbol{\eta}) \propto -n \ln C(\boldsymbol{\mu}, \tau) - \sum_{i=1}^{n} \frac{\tau}{2}\mathrm{d}(\mathrm{Exp}(\boldsymbol{\mu}, \boldsymbol{z}), y_i)^2 - \frac{\|x_i\|^2}{2} \tag{3.1}$$

and use this in the MCEM algorithm to estimate the parameters η. The procedure contains two main steps:

step 1: For every x_i in the data, generate a set of q samples from the posterior distribution (3.1) using the HMC. In iteration $(k+1)$, with the estimated parameters of iteration k, denoted as η^k, let x_{ij} represent the jth sample generated for x_i. The Monte Carlo approximation is calculated as follows:

$$Q(\boldsymbol{\eta} \mid \eta^k) = \mathbb{E}_{x_i|y_i,\eta^k}\Big[\sum_{i=1}^{n} \ln \Pr(x_i \mid y_i;\eta^k)\Big] \approx \frac{1}{q}\sum_{j=1}^{q}\sum_{i=1}^{n} \ln \Pr(x_{ij} \mid y_i;\eta^k). \tag{3.2}$$

step 2: In this stage, calculate the maximization step to update the parameters $\boldsymbol{\eta} = (\boldsymbol{\mu}, \boldsymbol{W}, \boldsymbol{\Lambda}, \tau)$ by maximizing the HMC approximation of the Q function as given in (3.2). This procedure involves a gradient ascent algorithm [47,48] since closed-form solutions are not available for all the parameters.

The details of this algorithm are summarized in Algorithm 3.

Algorithm 3 MCEM for PPGA

Input: Data set $Y \in M$, reduced dimension p
Initialize $\boldsymbol{\eta} = (\boldsymbol{\mu}, \boldsymbol{W}, \boldsymbol{\Lambda}, \tau)$
repeat
Sample X and do Hamiltonian Monte Carlo (HMC) method
Update $\boldsymbol{\eta} = (\boldsymbol{\mu}, \boldsymbol{W}, \boldsymbol{\Lambda}, \tau)$ using the gradient ascent algorithm
until convergence.

Zhang and Fletcher [67] proposed a Bayesian inference procedure for the estimation of model parameters and the simultaneous detection of the effective dimensionality of the latent space. Also, Zhang et al. [68] introduced a PPGA mixture model in subpopulations, where the parameters of the principal subspaces are automatically estimated by employing an EM algorithm, and Sommer [69] proposed a solution for a probabilistic model focusing on PGA based on the normal distribution in the manifold M, derived from the marginal distributions of a diffusion process in M with constant infinitesimal covariance.

3.4. Principal flow

Principal flow, introduced by Panaretos et al. [70], extends the concept of principal curves to Riemannian manifolds. Principal flows are flexible curves that pass through the mean of the data points and can capture non-geodesic patterns of variation both locally and globally, making them a powerful tool for analyzing data on manifolds. The Principal Flows move along a path of maximal data variation subject to smoothness constraints. Specifically, Panaretos et al. [70] studied the problem of constructing a smooth curve on a manifold passing through a given data center (such as the intrinsic mean) whose tangent velocity at each point aligns with the eigenvector of the corresponding tangent local covariance matrix, subject to smoothness constraints. In essence, the curve is defined as a solution to a Euler–Lagrange problem on the manifold.

Suppose that M is a complete Riemannian manifold of dimension p embedded in $\mathbb{R}^d$, $p < d$, and assume that the embedding is explicitly known, in that $M := \{x \in \mathbb{R}^d : F(x) = 0\}$ for a given differentiable function $F : \mathbb{R}^d \to \mathbb{R}^p$. The corresponding bundle of tangent spaces $T_X M$ of $x \in M$ will be denoted by

$$T_X M = \{y \in \mathbb{R}^d : \boldsymbol{\nabla F}(y) = 0\}$$

with associated metric tensor $\langle \cdot, \cdot \rangle$ and induced Riemannian metric $\mathrm{d}(\cdot, \cdot)$. Here, $\boldsymbol{\nabla F}$ is the derivative matrix $p \times d$ of F evaluated at $x \in M$, assumed to be of full rank everywhere in M. Throughout, we assume that a configuration $\{x_1, \ldots, x_n\} \in M$ is localized enough that there exists a connected open set $B \subset M$ covering this configuration and such that the logarithmic map $(\mathrm{Log}_{\bar{x}}(x_i))$ is well defined for all $\bar{x} \in B$ ($\bar{x}$ could be a Fréchet mean [44]).

Definition 3.1 (*Tangent PCA [70]*)**.** Let $x_1, \ldots, x_n \in M$ and $B \subset M$ be as above. The tangent principal components $\{e_1(\bar{x}), \ldots, e_r(\bar{x})\}$ at $\bar{x} \in B$ are the basis for $T_{\bar{x}}M$, given by the first p eigenvectors (corresponding eigenvalues are $\{\lambda_1(\bar{x}), \ldots, \lambda_p(\bar{x})\}$) of the $d \times d$ tangent covariance matrix defined by

$$\boldsymbol{\Sigma}(\bar{x}) := \frac{1}{n}\sum_{i=1}^{n} \mathrm{Log}_{\bar{x}}(x_i) \otimes \mathrm{Log}_{\bar{x}}(x_i),$$

where the tensor product here is defined by $\boldsymbol{y} \otimes \boldsymbol{y} := \boldsymbol{y}\boldsymbol{y}^\top$.

It is also possible to define local versions of the tangent covariance matrix $\boldsymbol{\Sigma}$. To do so, a scale parameter $h > 0$ and a smooth non-increasing univariate kernel K within the range of $[0, \infty)$ must be selected.

Definition 3.2 (*Local Tangent Covariance [70]*)**.** Given $h > 0$ and

$$\kappa_h(x, \bar{x}) := K(h^{-1} \| \mathrm{Log}_{\bar{x}}(x - \bar{x})\|),$$

a scale h local tangent covariance matrix $\boldsymbol{\Sigma}_h$ at $\bar{x} \in B$ is defined as

$$\boldsymbol{\Sigma}_h(\bar{x}) := \frac{1}{\sum_{i=1}^{n} \kappa_h(x_i, \bar{x})}\sum_{i=1}^{n} \mathrm{Log}_{\bar{x}}(x_i) \otimes \mathrm{Log}_{\bar{x}}(x_i)\kappa_h(x_i, \bar{x}). \tag{3.3}$$

Initially, the focus is on the first principal flow, and for simplicity, the first eigenvector and eigenvalue of a matrix $\boldsymbol{\Sigma}_h(x)$ are denoted $e(x)$ and $\lambda(x)$, respectively. These define an orientation field $\{V(x) : x \in B\}$ comprised of orientations of "maximal variation", that is,

$$V(x) = \{-\lambda(x)e(x), \lambda(x)e(x) : x \in B\}.$$

The fact that the orientation field $V(x)$ can be converted into a vector field, at least locally, will be crucial in what follows. We can choose the eigenvectors $\{e(x) : x \in N\}$ to be "pointing in the same direction" within an open neighborhood $N \subset B$.

Proposition 1 (*Panaretos et al. [70]*). *Let $\{x_1, \ldots, x_n\} \subset M$ and $B \subset M$ be as above, and assume that the tangent covariance matrix $\boldsymbol{\Sigma}_h(x)$ has distinct first and second eigenvalues for all $x \in B$. Given any $\bar{x} \in B$, the tangent vector $\lambda(\bar{x})e(\bar{x})$ can be extended to a tangent vector field $\boldsymbol{W} = \{W(x) : x \in N(\bar{x})\}$ defined over an open neighborhood $N(\bar{x})$ of $\bar{x}$ and such that*

- *$W(x)$ is independent of the local coordinates of $T_x M$ and can be expressed in coordinates of the ambient space $\mathbb{R}^d$.*
- *$\Sigma_h(x)W(x) = \lambda(x)W(x)$ (i.e., $W(x) \in V(x)$).*
- *$W : N(\bar{x}) \to \mathbb{R}^d$ is a differentiable mapping.*

With Proposition 1, we can now understand how to integrate over the orientation field defined by the eigenvectors and describe the class of functions that are candidates for a principal flow. Given $\bar{x}$ as a starting point on M, one can define the flow class for $\bar{x} \in M$ as

$$\Gamma(\bar{x}, v) = \Big\{ \gamma : [0, p] \to M : \gamma \in C^2(M), \quad p \leqslant 1, \ \gamma(s) \neq \gamma(s') \ \text{ for } \ s \neq s', \\ \gamma(0) = \bar{x}, \quad \dot{\gamma}(0) = v, \quad \ell(\gamma[0, t]) = t \text{ for all } 0 \leqslant t \leqslant p \leqslant 1 \Big\}$$

where $\ell(\gamma[0, t])$ represents the length of the parametric curve γ from $\gamma(0)$ to $\gamma(t)$, for all $0 \leqslant t \leqslant 1$. Here, v is a unit tangent vector at the point $\bar{x}$ in the tangent space $T_{\bar{x}}M$. Specifically, $\Gamma(\bar{x}, v)$ is the set of all twice differentiable, non-intersecting, and unit-speed curves on the manifold M. These curves start at $\bar{x}$ with velocity vector v at $\bar{x}$, and their length can be up to 1.

The principal flow is defined as a smooth curve γ on the manifold and needs to satisfy a variational problem. The curve γ passes through a given point $\bar{x}$ (e.g., a Fréchet mean) and has a derivative vector field $\dot{\gamma}$ that is maximally compatible with the vector field $\{V(x) : x \in B\}$ of orientations of maximal variation.

Definition 3.3 (*Principal Flow at Scale h, [70]*). Suppose $\bar{x} \in B$ and $\{\boldsymbol{\Sigma}_h(x) : x \in B\}$ have distinct first and second eigenvalues. Let W be the extension of $\lambda(\bar{x})e(\bar{x})$ as mentioned in Proposition 1. A principal flow of the set $\{x_1, \ldots, x_n\}$ through $\bar{x}$ at scale h is a union of two curves γ_1 and γ_2 that meet the conditions

$$\gamma_1 := \arg \sup_{\gamma \in \Gamma(\bar{x}, e(\bar{x}))} \int_0^{\ell(\gamma)} \langle \dot{\gamma}(t), W(\gamma(t)) \rangle \, dt$$

$$\gamma_2 := \arg \inf_{\gamma \in \Gamma(\bar{x}, -e(\bar{x}))} \int_0^{\ell(\gamma)} \langle \dot{\gamma}(t), W(\gamma(t)) \rangle \, dt.$$

With Proposition 1, the curve γ_1 starts at $\bar{x}$ and follows the direction of the vector field that extends $\lambda(\bar{x})e(\bar{x})$, while the curve γ_2 starts at $\bar{x}$ and follows the opposite direction $-\lambda(\bar{x})e(\bar{x})$. The degree of rigidity or flexibility of the flow is determined by the scale parameter h. For large values of h, the flow will be insensitive to local variations in the sample, instead focusing on global variation properties. Conversely, for small values of h, the flow will be more sensitive to local variations, capturing finer-scale variation properties.

Panaretos et al. [70] showed that the problem of finding a principal flow can be reformulated as an Euler–Lagrange problem on the manifold, provided conditions for the existence and uniqueness of a solution, and presented an algorithm for its practical application. Recently, Liu et al. [71] combined the level set method with the Principal Flow algorithm to obtain a fully implicit formulation, so that the curves obtained in the manifold fit the data set well.

Fixed Boundary Flow is an extension of Principal Flows that aims to find a flow with fixed starting and ending points for multivariate data sets lying on an embedded non-linear Riemannian manifold [72]. Unlike principal flows which start from the center of the data cloud, both points in Fixed Boundary Flow are given in advance using the intrinsic metric on the manifolds. From a geometric perspective, the Fixed Boundary Flow is defined as an optimal curve that moves within the data cloud, maximizing the inner product of the vector field, calculated locally, and the tangent vector of the flow at any point on the curve. To achieve this, an Euler–Lagrange problem is formulated and the solution is reduced to that of a Differential Algebraic Equation. In Yao et al. [72], conditions for the existence and uniqueness of a solution are given, and an algorithm for its practical implementation is also presented.

Instead of handing curves with an explicit parameterization in principal flows, Liu et al. [71] combine the level set method with the Principal Flow algorithm to obtain a fully implicit formulation, so that the obtained surface on the manifold fits the data set well.

4. PCA for spherical data

Spherical data typically arise in astronomy (many astronomical observations are points on the celestial sphere; see, e.g., [15,19,73]), earth sciences [19,70,74–76], shape studies [10,77–80], text mining [18,81], archeology [82,83], biology [79,84,85], medicine (the brain scan for instance [86,87] and cardiology [88,89]) and bioinformatics (Fotouhi and Golalizadeh [55], Hamelryck et al. [90], Dryden and Mardia [10], Ley and Verdebout [18]).

In geometry, the $(p-1)$-dimensional unit sphere (or a hyper-sphere in $\mathbb{R}^p$, where $p \geq 3$) is denoted as $\mathbb{S}^{p-1} := \{x \in \mathbb{R}^p : \|x\| = 1\}$. The computation of the spherical mean for a set of data points $x_1, \ldots, x_n$ is defined as $\frac{\bar{x}}{\|\bar{x}\|}$, where $\bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i$. The geodesic distance d on a unit sphere is defined as:

$$\mathsf{d}(x, y) = \arccos(x^\top y)$$

where $x, y \in \mathbb{S}^{p-1}$ and $\mathsf{d}(x, y)$ is the geodesic between x and y. Note that the shortest path between two points is unique, except for $x^\top y = -1$.

4.1. Principal arc analysis (PAA)

Jung et al. [41] introduced Principal Arc Analysis (PAA) as an extension of PCA on a special class of manifold namely on the *direct product manifold* $M = M_1 \oplus \cdots \oplus M_m$, where each M_i, $i \in \{1, \ldots, m\}$, is a simple manifold such as the unit circle $\mathbb{S}^1$, unit sphere $\mathbb{S}^2$, $\mathbb{R}_+$ and $\mathbb{R}^p$. For instance, when multiple 3D directions are considered simultaneously, the sample space is $\mathbb{S}^2 \oplus \cdots \oplus \mathbb{S}^2$, which is a direct finite product manifold. Another illustration is the medial representation of shapes known as m-reps [91], which is well-known in image analysis. Some necessary background on direct product manifolds can be found in [30].

Note that the curvature of direct product manifolds is mainly due to the spherical part, which motivated Jung et al. [41] to define the principal circles and the principal circle mean when the manifold is $\mathbb{S}^2$. They suggested transforming the data points in $\mathbb{S}^2$ into the linear space by a special mapping utilizing the principal circles for more complex direct product manifolds. Regarding other components of the manifolds, logarithmic mapping can be used to map the data to a linear space as done in Fletcher et al. [31]. Once directional data are mapped onto the linear space, the classical PCA can be applied to find principal components in the transformed linear space. In their real data application, Jung et al. [41] use *principal arcs* to model the prostate in medical images.

4.2. Principal nested spheres (PNS)

Jung et al. [92] suggested a decomposition technique called Principal Nested Spheres (PNS) that can be seen as an extension of the PAA approach to spheres. PNS was developed in response to concerns that geodesic based methods might not give an effective decomposition of the space when non-geodesic variation is major. Instead, PNS can be viewed as a *backwards* generalization of PCA, a method that sequentially reduces the dimension of the fitted sphere through iterative steps. PNS yields a sequence of submanifolds $A_1, A_2, \ldots, A_{d-2}$ of $\mathbb{S}^{d-1}$ such that

$$A_1 \subset A_2 \subset \ldots \subset A_{d-2} \subset \mathbb{S}^{d-1},$$

where a subsphere A_{d-2} of $\mathbb{S}^{d-1}$ is defined by an axis v and a distance $r \in (0, \pi/2]$ as

$$A_{d-2}(v, r) = \{x \in \mathbb{S}^{d-1} : \mathsf{d}(x, v) = r\}.$$

The subsphere A_{d-2} can be mapped onto the unit hypersphere $\mathbb{S}^{d-2}$ through an invertible mapping f_1 (see Jung et al. [92]); therefore, it can be identified with $\mathbb{S}^{d-2}$ and treated as a unit sphere. The next subsphere A_{d-3} can then be built from $\mathbb{S}^{d-2}$ (given a new axis $v^* \in \mathbb{S}^{d-2}$ and a new distance $r^* \in (0, \pi/2]$) and identified with a submanifold of $\mathbb{S}^{d-1}$ via $\mathsf{f}_1^{-1}(A_{d-3}) \subset A_{d-2}$. Continuing this process, the $(d-k)$-dimensional nested sphere of $\mathbb{S}^{d-1}$ is defined as

$$A_{d-k} = \begin{cases} \mathsf{f}_1^{-1} \circ \ldots \circ \mathsf{f}_{k-2}^{-1}(A_{d-k}) & \text{if } k = 3, \ldots, d-1 \\ A_{d-2} & \text{if } k = 2, \end{cases}$$

where the f_j's are the consecutive invertible mappings linking subspheres A_{d-j-1} with unit spheres $\mathbb{S}^{d-j-1}$.

From a practical point of view, suppose that we have a sample $x_1, \ldots, x_n$ with values on $\mathbb{S}^{d-1}$. The residual of a point $x \in \mathbb{S}^{d-1}$ for the subsphere $A_{d-2}(v_1, r_1)$ is defined as the signed length of the geodesic that connects x to $A_{d-2}(v_1, r_1)$, given by $\mathsf{d}(x, v_1) - r_1$. The best fitting least squares subsphere is obtained after determining

$$(\hat{v}_1, \hat{r}_1) = \underset{(v,r)\in(\mathbb{S}^{d-1}\times(0,\pi/2])}{\operatorname{argmin}} \sum_{i=1}^{n} (\mathsf{d}(x_i, v) - r)^2. \tag{4.1}$$

The selection of the objective function in Eq. (4.1) is not unique, and using different types of objective functions may enhance the robustness of the estimator. The empirical nested sphere $\hat{A}_{d-2}$, which has a dimension of $(d-2)$, is represented by the symbol $A_{d-2}(\hat{v}_1, \hat{r}_1)$. Next, the data are projected onto this nested sphere along the minimal geodesic connecting each observation to $\hat{A}_{d-2}$. The same process as described in Eq. (4.1) is then applied to these projected data to obtain the second best fitting sphere $\hat{A}_{d-3}$ and hence $\hat{A}_{d-3} \subset \hat{A}_{d-2} \subset \mathbb{S}^{d-1}$.

As an extension of PNS for addressing m-rep data, Pizer et al. [93] proposed the method of *composite principal nested spheres*. Their approach consists of sphere decomposition in the data using PNS and then combining the scores into a Euclidean vector that is then analyzed by PCA. For further details, the interested reader is referred to Huckemann and Eltzner [22,23].

4.3. Barycentric subspace analysis of PCA

The geodesic based methods such as PGA and GPCA consist of a reference point (e.g., intrinsic mean) and k tangent vectors at that specific point. This assigns a special role to the reference point, unlike in higher-dimensional descriptors. Additionally, while these spaces are entirely geodesic at the reference point, they generally are not elsewhere. This asymmetry might not be optimal for multimodal distributions lacking a single 'pole'.

In order to have a symmetric and 'multi-pole' description of subspaces, Pennec [43] introduced a novel and general set of subspaces in manifolds, known as *barycentric subspaces*. These subspaces are implicitly defined as a locus of weighted means of $k+1$ reference points with positive or negative weights summing up to one. Barycentric subspaces provide a natural nesting property, enabling the construction of nested subspaces that may be used for creating forward or backward nested subspaces to approximate data points. Depending on the definition of the mean, the Fréchet, Karcher, or exponential barycentric subspaces (FBS/KBS/EBS) are used.

Consider a Riemannian manifold M with a Riemannian metric on each tangent space T_xM and a logarithmic map $\mathrm{Log}_x : M \to T_xM$. In addition, suppose $x_0, \dots, x_k \in M$ be $k+1$ distinct reference points and $\pi_0, \dots, \pi_k \in \mathbb{R}$ be $k+1$ weights such that $\sum_i \pi_i \neq 0$, and define $\mathscr{P}_k^* = \{(\pi_0, \dots, \pi_k)^T \in \mathbb{R}^{k+1} \text{ such that } \sum_i \pi_i \neq 0\}$. The (normalized) weighted variance at point x with weight $\pi \in \mathscr{P}_k^*$ is

$$\sigma^2(x, \pi) = \frac{1}{2} \sum_{i=0}^{k} \bar{\pi}_i \mathrm{d}^2(x, x_i) = \frac{1}{2} \sum_{i=0}^{k} \frac{\pi_i}{\sum_{j=0}^{k} \pi_j} \mathrm{d}^2(x, x_i)$$

where we call $\bar{\pi}_i = \dfrac{\pi_i}{\sum_{j=0}^{k} \pi_j}$ normalized weights. The *Fréchet barycentric subspace* (FBS) is the locus of weighted Fréchet means of these points, i.e., the set of absolute global minima of the weighted variance:

$$FBS(x_0, \dots, x_k) = \underset{x \in M}{\operatorname{argmin}}\, \sigma^2(x, \pi), \ \ \pi \in \mathscr{P}_k^*.$$

The Karcher barycentric subspace (KBS) is defined similarly with local minima instead of global ones. The exponential barycentric subspace (EBS) generated by the affinely independent reference points $(x_0, \dots, x_k) \in M$ is defined as

$$EBS(x_0, \dots, x_k) = \{x \in M \mid \exists \pi \in \mathbb{R}^{k+1} : \sum_{i=0}^{k} \pi_i \mathrm{Log}_x(x_i) = 0\}.$$

According to Pennec [43], the definition is only valid for points x that lie outside the reference points' cut locus.

Barycentric subspaces can be characterized similarly to the Euclidean case, either through the singular value decomposition of a specific matrix or via the diagonalization of the covariance and Gram matrices [43]. Pennec [43] illustrated the use of barycentric subspaces in cardiac imaging with the estimation, analysis and reconstruction of cardiac motion from sequences of images.

4.4. Spherical coordinate-based kernel PCA

Guo and Ling [94] introduced an extension of PCA based on kernel PCA [95,96] using spherical coordinates. In this method, the kernel function refers to a nonlinear transform from the Cartesian coordinate system to the spherical coordinate system. To be more precise, the vectors originally represented in the Cartesian coordinate system are converted into vectors defined in the spherical coordinate system. Subsequently, the rotational angles or vector radii are reset to their corresponding mean values. Finally, the obtained vectors in the spherical coordinate system are transformed back into the Cartesian coordinate system. As the degrees of freedom of the processed vectors decrease in the spherical coordinate system, the dimension of the manifold of the processed vectors in the Cartesian coordinate system also reduces. The computational cost of the spherical coordinate-based kernel PCA is significantly lower compared to the PCA and kernel PCA methods. This makes it important for applications such as real-time video processing, where computational complexity is crucial. As a result, spherical coordinate-based kernel PCA outperforms both PCA and kernel PCA in dealing with Euclidean data such as video process applications.

4.5. Spherical Rotation Component Analysis (SRCA)

Luo et al. [97] presented Spherical Rotation Component Analysis (SRCA), a method that employs spheres to represent high-dimensional data in low-dimensional spaces while preserving the geometric structure and cyclic nature of biological processes. This rotation-based method utilizes a geometric-induced loss function to minimize the point-to-sphere distance between the original and target spaces. This method is an essential part of the process of revealing the structure of and performing additional biological research on large biological datasets, such as single-cell RNA sequencing data.

5. PCA for toroidal data

Toroidal data are an extension of circular data on a torus and play a critical role in various scientific fields, such as astronomy [98], bioinformatics [19,21,98–105], meteorology [106], environmental science [107] and medicine [105,108]. In Supplementary Material (Section SM-1), we provide an example comparing different PCA-related methods.

In geometry, a torus is a surface of revolution that is produced by rotating a circle in a three-dimensional space about a coplanar axis. Let the radius from the center of the hole to the center of the torus tube be R, and let the radius of the tube be r. The parametric equations for a torus are

$$x = (R + r\cos\psi)\cos\phi$$
$$y = (R + r\cos\psi)\sin\phi$$
$$z = r\sin\psi$$

where the angles $(\psi, \phi) \in [0, 2\pi]$. Depending on the relative sizes of R and r, three types of torus, known as standard tori, are feasible: $R > r$ corresponds to a *ring torus*, $R = r$ corresponds to a *horn torus* which is tangent to itself at the point $(0, 0, 0)^T$, and $R < r$ corresponds to a *self-intersecting spindle torus*. When $R = 0$, the torus degenerates into the sphere. In topology, a ring torus is homeomorphic to the Cartesian product of two circles, that is, $\mathbb{T}^2 = \mathbb{S}^1 \times \mathbb{S}^1$.

The $d-$torus (hyper-torus) is the $d-$fold product group $\mathbb{T}^d = \underbrace{\mathbb{S}^1 \times \cdots \times \mathbb{S}^1}_{d \text{ times}}$.

5.1. Dihedral angles principal component analysis (dPCA)

Mu et al. [99] and Altis et al. [100] proposed dihedral angles principal component analysis (dPCA) as an extension of PCA for dihedral angles in protein structures. Dihedral angles seem highly advantageous when considering biomolecules because other internal coordinates like bond lengths and bond angles usually do not change large amplitude. PCA utilizing Cartesian coordinates does not produce the correct rugged free-energy landscapes [99] as a result of the mixing of internal and overall motion. It is worth noting that a study of free-energy landscapes provides an understanding of how proteins fold and function [109–111].

To avoid problems arising from the circularity of these variables, trigonometric functions are applied to dihedral angles. Subsequently, Mu et al. [99] used this method to describe the energy landscape of small peptides in the protein structure. The basic idea of the dPCA is to perform PCA on sin- and cos-transformed dihedral angles

$$q_{2j-1} = \cos\theta_j, \quad q_{2j} = \sin\theta_j, \quad \theta_j \in [0, 2\pi] \quad \text{and} \quad j \in \{1, \ldots, n\}$$

where n is the total number of peptide backbone and side-chain dihedral angles used in the analysis. Hence, the covariance matrix of the dPCA uses $2n$ variables q_j. In Altis et al. [100], it was shown that the dPCA provides a one-to-one representation of the original angle distribution. Furthermore, its principal components can be readily characterized by the corresponding conformational changes of the peptide.

To systematically construct a low-dimensional free-energy landscape of RNA from a classical molecular dynamics simulation, Riccardi et al. [112] investigated the performance of various other variants of PCA, including a PCA on the Cartesian coordinates of all atoms (cPCA) and an angular PCA (aPCA) based on applying PCA to toroidal data centered on their circular means. It was shown that the conformational heterogeneity of RNA hairpins could only be resolved by dPCA, not by the PCA using Cartesian coordinates [112,113]. Furthermore, cPCA completely failed to reproduce the qualitative features of the folding free-energy landscape. On the other hand, the free-energy landscape obtained from the aPCA appears quite similar to the results for the dPCA. This is because, in this case, the dihedral angles of the RNA hairpins stay close to their circular mean, for which the aPCA is a correct representation.

The dPCA has various drawbacks. To name but a few, it requires double the number of variables, neglects the identity $\cos^2\theta + \sin^2\theta = 1$, $\theta \in [0, 2\pi]$, and disregards the topological space of angles. These limitations motivate researchers to find other extensions of the dPCA. Moreover, when determining the minimal dimensionality of free-energy landscapes, Maisuradze et al. [114] showed that trajectories with large fluctuation amplitudes require multidimensional landscapes to accurately capture folding dynamics. This is due to the complex hierarchical structure often observed in the first few PCs, with each component capturing only a small portion of the fluctuations. Consequently, for trajectories with substantial fluctuation amplitudes, not all peaks in these hierarchical PCs represent distinct conformational states. Therefore, a careful examination of the structures in each local minimum is essential.

Complex dPCA. Altis et al. [100] proposed Complex dPCA as an alternative extension to the sin/cos transformation in dPCA which maps n angles on $2n$ real numbers. They suggest transforming the angles θ_j into complex numbers using Euler's formula

$$z_j = e^{i\theta_j} \in \mathbb{C}, \quad j \in \{1, \ldots, n\}, \quad \text{where i denotes the complex number.}$$

The covariance matrix corresponding to the complex variables z_j is defined as

$$\boldsymbol{C}_{kj} = \langle (z_k - \langle z_k \rangle)(z_j^* - \langle z_j^* \rangle) \rangle,$$

where $k, j \in \{1, \ldots, n\}$, z^* is the complex conjugate of z, and $\langle \cdot \rangle$ denotes the average over all sampled conformations. The matrix $\mathbf{C} = (C_{kj})$ is a Hermitian matrix with n real-valued eigenvalues λ_j and n complex eigenvectors v_j, where the eigenvectors are unique

up to a phase $\theta_0 \in [0, 2\pi]$. The complex principal components are $\boldsymbol{Y} = v^\top \boldsymbol{z}$. Despite its similarity to the sin/cos dPCA, the Complex dPCA might be advantageous because the representation of n angular variables yields directly n eigenvalues and n eigenvectors (i.e., there is no duplication of variables as in the sin/cos dPCA) and it may enable their direct interpretation in terms of simple physical variables [100].

PCA on torus (dPCA+). The dPCA+ is an extension of dPCA for protein data [115]. The main idea is that the (periodicity-induced) projection error can be minimized by transforming the data so that the maximal gap of the sampling is shifted to the periodic boundary. After that, the covariance matrix and its eigen-decomposition can be computed in a standard manner. This method preserves the topological feature of the torus by defining the data points' correct neighborhoods. In this regard, the main assumption underlying dPCA+ is that the data indeed show a significant gap in their distribution. In general, there are many ways to maximize the gap between data points by selecting the optimal cut. For instance, one can (1) increase the squared distance of the cut from the nearest points, (2) minimize the number of data points within a corridor surrounding the cut, (3) utilize Molecular Dynamics simulations, which generate trajectories illustrating atomic motions by presenting atomic coordinates at specific time intervals, provide the ability to examine changes over time trajectories, and select a cut that minimizes the number of crossings, as outlined by [115]. For the description of maximal gaps in the dihedral distributions of proteins, [115] have chosen a procedure based on the second approach. To highlight some shortcomings of dPCA+, it should be noted that a significant gap in data distribution is a limitation, in general, and this method is relatively comparable to aPCA [112] in its practical application.

GeoPCA. Inspired by GPCA (see Section 3.2), Sargsyan et al. [116] proposed GeoPCA as a developed tool for multivariate analysis of dihedral angles based on principal component geodesics. GeoPCA uses a noninjective mapping from $\mathbb{T}^d$ to $\mathbb{S}^d$ that equates toroidal angles in $[-\pi, \pi)^d$ to hyperspherical coordinate angles, even though the latter are defined in $[0, \pi]^{d-1} \times [-\pi, \pi)$, and then applies PGA. Furthermore, Sargsyan et al. [116] applied the principal component geodesic approach in the clustering of RNA conformations. Later, Sargsyan et al. [117] developed a Python package called `Clustangles` to analyze and cluster conformations of molecular dynamics trajectories using angular data.

Mixture modeling of dihedral angles. Understanding the distributions of dihedral angles in proteins has been a central focus in protein structural and biophysical research [118–120]. A significant advancement in this area was proposed by Mardia et al. [121], who introduced a model based on mixtures of bivariate von Mises distributions. This model provides a way to characterize the secondary structures of proteins through their conformational angles, often visualized in a Ramachandran plot [122].

Expanding on previous work, Amarasinghe et al. [123] applied a Bayesian model selection criterion (that is Minimum Message Length (MML) [124]) to model mixtures of von Mises distributions. MML is a Bayesian approach for hypothesis/model selection. By applying it, Amarasinghe et al. [123] demonstrated that mixture models based on MML are less prone to common issues of under- or over-fitting. This approach thus represents a robust alternative to model the distribution of protein dihedral angles. Ameijeiras-Alonso and Ley [125] introduced sine-skewed toroidal distributions to model amino acid dihedral angles via mixtures. Furthermore, Jung et al. [126] developed a clustering method for toroidal data using conformal prediction sets. This method applies a mixture of model-based prediction sets to identify clusters and assign cluster memberships, outperforming simple modifications of standard clustering algorithms for toroidal data. More recently, Xu and Wang [127] proposed a semiparametric mixture model for estimating multidimensional toroidal data, showing that their mixture-based density estimator generally outperforms the kernel density estimator.

5.2. PCA for the wrapped normal torus model

Kent and Mardia [128] suggested a trigonometric moment characterization of the covariance matrix in a wrapped normal model on $\mathbb{T}^d$, applying PCA on it. Suppose that θ is a vector of angles following a wrapped normal torus distribution, that is, $\theta_j = X_j \bmod 2\pi$, $j \in \{1, \dots, d\}$, where $X \sim N(\mathbf{0}, \boldsymbol{\Sigma})$ with $\boldsymbol{\Sigma}$ a $d \times d$ symmetric positive definite real-valued matrix. Then it can be shown that

$$\mathbb{V}\mathrm{ar}(\cos\theta) = \boldsymbol{D}\boldsymbol{A}\boldsymbol{D} - \boldsymbol{c}\boldsymbol{c}^\top, \ \mathbb{V}\mathrm{ar}(\sin\theta) = \boldsymbol{D}\boldsymbol{B}\boldsymbol{D}, \ \mathbb{C}\mathrm{ov}(\cos\theta, \sin\theta) = \mathbf{0}, \tag{5.1}$$

where the elements of the vector $\boldsymbol{c}$ and matrices $\boldsymbol{A}$, $\boldsymbol{B}$ are given by

$$a_{jk} = c_j c_k \cosh(\Sigma_{jk}), \quad b_{jk} = c_j c_k \sinh(\Sigma_{jk}) \ \text{ with } \ c_j = \exp\left\{-\frac{1}{2}\Sigma_{jj}\right\} \ \ \forall\, j \text{ and } k \in \{1, \dots, d\}$$

and $\boldsymbol{D} = \mathrm{diag}(c)$ is a $d \times d$ matrix with vector c along its main diagonal. Thus, $\boldsymbol{\Sigma}$ can be recovered from the trigonometric moments through the equation

$$\boldsymbol{\Sigma} = \sinh^{-1}(\boldsymbol{D}^{-1}\mathbb{V}\mathrm{ar}(\sin\theta)\boldsymbol{D}^{-1}), \tag{5.2}$$

where $\sinh^{-1}(u) = \log(u + \sqrt{u^2 + 1})$ for $u \in \mathbb{R}$. This means that the inverse sinh function is applied to each element of the matrix. These results suggest a method to estimate $\boldsymbol{\Sigma}$ from a $n \times d$ matrix of toroidal data:

i. Calculate the first-order trigonometric moments of the sample for the d angles, and rotate each angle so that the resultant vector points towards the positive horizontal axis, leading to an estimate of $\boldsymbol{D}$.
ii. Calculate the sample second trigonometric moments to get an estimate corresponding to $\mathbb{V}\mathrm{ar}(\sin\theta)$ as (5.1).
iii. Now, use (5.2) to produce an estimate of $\boldsymbol{\Sigma}$.

Similarly, we can estimate $\boldsymbol{\Sigma}$ from $\mathbb{V}\mathrm{ar}(\cos\theta)$ using $\boldsymbol{\Sigma} = \cosh^{-1}(\boldsymbol{D}^{-1}\mathbb{V}\mathrm{ar}(\cos\theta)\boldsymbol{D}^{-1}) + \mathbf{1}\mathbf{1}^\top$, where $\mathbf{1}$ is a $d-$dimensional vector of ones. By pooling the cosine and sine information, one can obtain a pooled estimate of $\boldsymbol{\Sigma}$, as suggested by [128].

5.3. Dihedral angles principal geodesic analysis (dPGA)

Nodehi et al. [102] introduced dihedral angles principal geodesic analysis (dPGA) as an extension of PGA designed specifically for dihedral angles within protein structures, employing a *forward* approach. The primary objective of this method is to effectively capture the variability of a dataset situated on a torus, with a specific emphasis on dihedral angles. To achieve this, the method involves the derivation of logarithmic and exponential maps on the torus. A torus, defined topologically as the product of two circles, requires the approximation of current maps using the logarithmic and exponential maps of a circle.

Expanding on this foundation, Nodehi et al. [102] advocated for the seamless integration of dPGA into the statistical toolkit for molecular dynamics simulations, which generate intricate high-dimensional molecular structure data. The efficacy of dPGA is assessed by applying it to a real data set, and a comparative analysis with existing methods is performed.

Both the PCA for the wrapped normal torus model and dPGA yield principal directions that almost surely wrap around infinitely. Addressing this issue, Kent and Mardia [128] discussed desirable properties for Principal Component Curves (PCC) in $[-\pi, \pi)^2$.

5.4. Torus principal component analysis (TPCA)

Eltzner et al. [103] proposed TPCA as an extension of PCA to toroidal data. This method deforms the torus into a sphere and then uses PNS [92]. Moreover, TPCA involves the small sphere fitting step, providing a test to avoid overfitting. However, deforming the torus into a sphere creates singularities. To address this problem, they introduce a data-adaptive preclustering technique. In detail, the authors proposed two data-driven orderings of variables, *SI ordering* and *SO ordering*, corresponding to sorting the variables in terms of decreasing and increasing circular spread, respectively. However, Zoubouloglou et al. [98] showed that SI ordering and SO ordering yield significantly different outcomes that may affect subsequent analyses (as depicted in Figure SM-1 in the Supplementary Material for a detailed discussion, see Section 5.2 in [98]). Furthermore, they evaluated the performance of TPCA with two RNA structures. Recently, Mardia et al. [105] have provided some applications of TPCA on clustering for multivariate analysis. Although there is a principled approach in TPCA, the transformation in TPCA is not invariant under the permutations of variables. Furthermore, as described in Remark 2.3 of Eltzner et al. [103], TPCA is only applicable when there are structural data gaps in all angles except for at most two. This limitation arises due to the formation of a singularity set, which forms a subtorus of dimension $d-2$. Moreover, deformation to $\mathbb{S}^{d-1}$, which is designed to cut the torus at a point with minimal distortion, may induce artifacts in particular data sets. These deformation artifacts can create inaccurate cluster structures in datasets containing no cluster structures (See Figure SM-2 in the Supplementary Material for a graphical overview, or Section 5.1 of [98] for a more comprehensive understanding).

5.5. Torus probabilistic PCA (TPPCA)

Nodehi et al. [104] introduced Torus Probabilistic PCA (TPPCA) as an extension of PCA designed for toroidal data which are represented as a latent variable model modulo 2π, assuming Gaussianity for the latent variables. An EM algorithm is developed to estimate the parameters within the model and introduced a Likelihood Ratio Test (LRT) to determine the optimal number of components.

Let $\boldsymbol{X} \in \mathbb{R}$ be a real random variable and define a circular random variable as $\boldsymbol{\Theta} = \boldsymbol{X} \bmod 2\pi \in [0, 2\pi)$ or $\boldsymbol{X} = \boldsymbol{\Theta} + 2\pi k$ for $k \in \mathbb{Z}$. The TPPCA is a latent variable model that seeks to relate a d-dimensional observation vector $\boldsymbol{\Theta} \in \mathbb{T}^d$ to a corresponding p-dimensional vector of latent (or unobserved) variables $\boldsymbol{Z} \in \mathbb{R}^p$ $(p < d)$, that is,

$$\begin{aligned} \boldsymbol{\Theta} &= \boldsymbol{X} \bmod 2\pi \\ \boldsymbol{X} &= \boldsymbol{\mu} + \boldsymbol{W}\boldsymbol{Z} + \boldsymbol{\epsilon}, \end{aligned} \tag{5.3}$$

where $\boldsymbol{W}$ is a $(d \times p)$ matrix that relates the two sets of variables, $\boldsymbol{\Theta} = (\Theta_1, \ldots, \Theta_d)^\top \in \mathbb{T}^d$, $\boldsymbol{X} = (X_1, \ldots, X_d)^\top \in \mathbb{R}^d$, $\boldsymbol{Z} \sim N_p(0, I_p)$ is a p-dimensional Gaussian latent vector with zero mean and covariance $\boldsymbol{I}_p$, and $\boldsymbol{\epsilon} \sim N_d(\boldsymbol{0}, \sigma^2 \boldsymbol{I}_d)$ is a d-dimensional zero-mean Gaussian-distributed noise vector with covariance $\sigma^2 \boldsymbol{I}_d$. Assume that $\mathbb{C}\mathrm{ov}(\boldsymbol{Z}, \boldsymbol{\epsilon}) = \boldsymbol{0}$ and there exists a random vector $\boldsymbol{K}$ such that $\boldsymbol{\Theta} = \boldsymbol{X} - 2\pi\boldsymbol{K} \in \mathbb{T}^d$. For a given point $x \in \mathbb{R}^d$, define a parameter vector $\boldsymbol{k} \in \mathbb{Z}^d$ so that $\boldsymbol{\theta} = x - 2\pi\boldsymbol{k} \in \mathbb{T}^d$, then

$$f(\theta, x, z, k) = f(\theta, k|x, z) f(x|z) f(z) = f(\theta, k|x) f(x|z) f(z)$$

where

$$f(\theta, k|x) = \mathbb{1}(\theta, k, x) = \begin{cases} 1 & \text{if } \theta \in \mathbb{T}^d \text{ and } k = \dfrac{x - \theta}{2\pi} \\ 0 & \text{otherwise.} \end{cases}$$

The conditional distribution $f(x|z)$ is modeled as $N_d(\boldsymbol{\mu} + \boldsymbol{W}\boldsymbol{Z}, \sigma^2 \boldsymbol{I}_d)$. This leads to the joint distribution

$$\begin{aligned} &f(\theta, x, z, k) \propto \mathbb{1}(\theta, x, k)(\sigma^2)^{-\frac{d}{2}} \times \\ &\exp\left\{ -\frac{(\theta + 2\pi k - \boldsymbol{W}z - \mu)^\top (\theta + 2\pi k - \boldsymbol{W}z - \mu)}{2\sigma^2} - \frac{z^\top z}{2} \right\}. \end{aligned}$$

Let $\boldsymbol{\vartheta} = (\theta_1, \ldots, \theta_n)$ be a sample of size n and $\boldsymbol{K} = (k_1, \ldots, k_n)$ a set of missing values from the random vector K. The logarithm of the likelihood function, denoted as $\ell(\boldsymbol{\mu}, \boldsymbol{W}, \sigma^2, \boldsymbol{K})$, is given by

$$\ell(\boldsymbol{\mu}, \boldsymbol{W}, \sigma^2, \boldsymbol{K}) = \sum_{j=1}^{n} \ln f(\theta_j, x_j, z_j, k_j)$$
$$\propto \sum_{j=1}^{n} \left[-\frac{d}{2} \ln \sigma^2 - \frac{(x_j - \boldsymbol{W} z_j - \mu)^\top (x_j - \boldsymbol{W} z_j - \mu)}{2\sigma^2} - \frac{z_j^\top z_j}{2} \right] \times \mathbb{1}(\theta_j, k_j, x_j).$$

TPPCA now involves calculating the expectation of the log-likelihood with respect to latent variables based on the data and initial values (for more in-depth information, we refer the reader to [104]). The overall procedure can be summarized as follows.

Step 0: Initial values for all the parameters are given by a procedure described in Section 4.3 in Nodehi et al. [104].
Step 1: In this step, $\boldsymbol{\mu}$ is updated, the missing values of $\boldsymbol{K}$ are imputed using the Classification Expectation-Maximization algorithm (CEM algorithm), as described by Nodehi et al. [129].
Step 2: After obtaining updated values for $\boldsymbol{\mu}$ and $\boldsymbol{K}$, the next step is to update the estimates of $\boldsymbol{W}$ and σ^2. This is done by maximizing the log-likelihood function with respect to these parameters.
Step 3: Repeat **Step 1** and **Step 2** until convergence. Convergence happens when there are no significant changes in the parameters or the log-likelihood function.

5.6. Scaled torus principal component analysis (ST-PCA)

Zouboulouglou et al. [98] recently proposed scaled torus principal component analysis (ST-PCA) to multivariate toroidal data. The ST-PCA seeks a data-driven map from a torus to a sphere with the same size and radius as the torus. This method was inspired by the work of Eltzner et al. [103], who demonstrated that spherical embeddings are often more suitable than Euclidean ones for data on $\mathbb{T}^d$. The analyses are primarily conducted on spheres, and upon identifying the optimal data fit, the results can be inverted back to the torus. This optimal fit is achieved through Multidimensional Scaling, which minimizes the discrepancies between pairwise geodesic distances on both the torus and the corresponding sphere.

As highlighted in Zouboulouglou et al. [98], the foundation of ST-PCA is based on TPCA [103], with several key reasons supporting the use of spherical embeddings.

1. The spherical space $\mathbb{S}^d$ allows for the embedding of arbitrary distance matrices under specific conditions, unlike a Euclidean space. Schoenberg [130] provided the fundamental conditions necessary for embedding arbitrary distances isometrically into a sphere, a result further formalized in Theorem 3.22 by Pekalska and Duin [131].
2. For all $d \in \mathbb{N}$, both $\mathbb{S}^d$ and $\mathbb{T}^d$ are compact Riemannian manifolds, which makes them different from $\mathbb{R}^d$ with the Euclidean distance.
3. The topology of $\mathbb{S}^d$ also allows for an isomorphic subspace to $\mathbb{S}^1$, helping to preserve the periodicity inherent in $\mathbb{T}^d$. This contrasts with $\mathbb{R}^d$, where periodic structures are not supported.
4. As discussed in Cox and Cox [132], convex hulls in $\mathbb{S}^d$ and $\mathbb{T}^d$ present challenges compared to $\mathbb{R}^d$. This is because on Euclidean space, convex hulls have a natural notion of a boundary.
5. From a statistical perspective, the use of PNS within $\mathbb{S}^d$ provides better clustering compared to applying PCA in $\mathbb{R}^d$. For example, when there are three clusters, $\mathbb{S}^d$ can effectively represent all clusters through a one-dimensional periodic mode of variation (specifically, $\mathbb{S}^1$), which is challenging in $\mathbb{R}^d$ unless the clusters have colinear centers. Embedding data in $\mathbb{S}^d$ and using PNS thus provides a more flexible and accurate representation of clusters than embedding in $\mathbb{R}^d$ and using PCA.

As previously introduced, let $\mathbb{S}^d := \{a \in \mathbb{R}^{d+1} : \|a\| = 1\}$, where $\|\cdot\|$ is the Euclidean norm, and $\mathbb{T}^d := (\mathbb{S}^1)^d := \underbrace{\mathbb{S}^1 \times \cdots \times \mathbb{S}^1}_{d \text{ times}}$ or equivalently $\mathbb{T}^d = [-\pi, \pi)^d$. The geodesic distance on $\mathbb{S}^d$ (denoted as $\delta_{\mathbb{S}^d}$) is given by

$$\delta_{\mathbb{S}^d}(a, b) = \arccos(a^\top b), \quad a, b \in \mathbb{S}^d.$$

For angular coordinates on $\mathbb{S}^1$, $\delta_{\mathbb{S}^1}(\phi, \psi) = \min\{|\phi - \psi|, 2\pi - |\phi - \psi|\}$, $\phi, \psi \in [-\pi, \pi)$. The distance on $\mathbb{T}^d$ is defined as

$$\delta_{\mathbb{T}^d}(\boldsymbol{\Phi}, \boldsymbol{\Psi}) := \left(\sum_{i=1}^{d} \delta_{\mathbb{S}^1}(\phi_i, \psi_i)^2 \right)^{\frac{1}{2}}, \quad \boldsymbol{\Phi}, \boldsymbol{\Psi} \in \mathbb{T}^d = [-\pi, \pi)^d.$$

Given a sample $x_1, \ldots, x_n \in \mathbb{T}^d$, ST-PCA runs in three main steps as follows:

1. Transform the original sample on $\mathbb{T}^d$ to a similar configuration of points on a $d-$dimensional sphere by using Spherical MultiDimensional Scaling (SMDS) [132,133]. SMDS is an optimization procedure designed to minimize the squared differences between the sample pairwise geodesic distances on $\mathbb{T}^d$ and the corresponding pairwise geodesic distances of the transformed sample on a $d-$sphere. In other words, a configuration $\hat{y}_1, \ldots, \hat{y}_n \in \mathbb{S}^d$ is obtained by solving the SMDS problem

$$(\hat{y}_1, \ldots, \hat{y}_n) = \arg \min_{(y_1, \ldots, y_n) \in (\mathbb{S}^d)^n} \frac{1}{n(n-1)} \sum_{i \neq j} (\delta_{\mathbb{T}^d}(x_i, x_j) - \delta_{\mathbb{S}^d}(y_i, y_j))^2.$$

2. Obtain S^j, $j = d-1, \dots, 1$, the nested sequence of subspaces that are isomorphic to $\mathbb{S}^j$ and that best fit $\hat{y}_1, \dots, \hat{y}_n$ according to PNS.
3. Optionally, "invert" S^1 through a reverse SMDS problem to obtain a principal curve on $\mathbb{T}^d$.

One key limitation of ST-PCA arises from its dependence on SMDS. The SMDS optimization process, crucial for finding the "best fit" sphere onto which toroidal data projects, becomes increasingly computationally intensive as the dataset size grows [98]. This limitation poses challenges in applying ST-PCA to large-scale toroidal datasets, necessitating careful consideration of computational resources and potential optimization strategies.

5.7. Toroidal PCA via density ridges

Toroidal Ridge PCA (TR-PCA) is a recent extension of PCA for bivariate toroidal data that utilizes density ridges as a flexible analog to the first principal component [134]. García-Portugués and Prieto-Tirado [134] have shown that two reference toroidal distributions, the Bivariate Sine von Mises (BSvM) and the Bivariate Wrapped Cauchy (BWC), provide stable connected components that pass across the distributions' modes by applying this method to bivariate data. This method derives PCA-like scores by obtaining signed distances along the ridge, signed projections onto the ridge, and Fréchet means within the ridge. Facilitated by Fourier approximations of the ridge curve, these operations are executed with a fast analytical approach. In their data application, García-Portugués and Prieto-Tirado [134] demonstrated the usefulness of TR-PCA in the study of ocean currents at the coast of Santa Barbara.

Density ridges are higher-dimensional extensions of the mode concept, offering valuable insights into the essential features of a density distribution. A mode, defined as a local maximum of the density function f, is identified by a null gradient and negative Hessian eigenvalues-excluding degenerate cases. The eigen-decomposition of the Hessian matrix of a density function f, evaluated at a point $t \in \mathbb{R}^d$, is expressed as $\mathrm{H}f(t) = \mathbf{U}(\mathbf{t})\Lambda(t)(\mathbf{U}(\mathbf{t}))^\top$, where $\mathbf{U}(\mathbf{t}) = (u_1(t), \dots, u_d(t))$ is a matrix with columns representing the eigenvectors of $\mathrm{H}f(t)$, and $\boldsymbol{\Lambda}(t) = \mathrm{diag}(\lambda_1(t), \dots, \lambda_d(t))$, arranged in descending order, comprises their corresponding eigenvalues. Denoting $\mathbf{U_{d-1}}(\mathbf{t}) = (u_2(t), \dots, u_d(t))$ and defining the projected gradient on $(u_2(t), \dots, u_d(t))$ as $\nabla^{(d-1)}f(t) = \mathbf{U_{d-1}}(\mathbf{t})(\mathbf{U_{d-1}}(\mathbf{t}))^\top \times \nabla f(t)$, where $\nabla f(t)$ denotes the column vector gradient, the density ridge is defined by [135] as follows.

Definition 5.1 (*Density Ridge [135]*)**.** The density ridge of f is the set

$$\mathscr{R}(f) = \left\{ t \in \mathbb{R}^d \,:\, \left\| \nabla^{(d-1)} f(t) \right\| = 0,\ \lambda_2(t), \dots, \lambda_d(t) < 0 \right\}.$$

Clearly, two cases imply that $t \in \mathbb{R}^d$ satisfies $\nabla^{(d-1)}f(t) = 0$. The first is $\nabla f(t) = 0$. In this case, if $\lambda_2(t), \dots, \lambda_d(t) < 0$, t is either a local maximum or a saddle point. The second is that $\nabla f(t)$ is perpendicular to $\mathbf{U_{d-1}}(\mathbf{t})(\mathbf{U_{d-1}}(\mathbf{t}))^\top$, meaning that the gradient is parallel to $u_1(t)$. In this case, the directions of maximum ascent (gradient) and "minimum signed curvature" (u_1) coincide. Two important properties of density ridges are "ridge invariance to translations and rotations" and "ridges for elliptically symmetric densities", as established in García-Portugués and Prieto-Tirado [134]. These properties have several important consequences. Firstly, they simplify the computation of density ridges by focusing on those centered at a specific origin, providing insights into utilizing the symmetries of f to reduce computational costs in evaluating $\mathscr{R}(f)$. Secondly, there is a close connection between $\mathscr{R}(f)$ and the first principal component of PCA for elliptically symmetric densities, as the subspace generated by the latter direction is included in it. Thirdly, like the first principal component, the density ridge includes the center in elliptically symmetric distributions. The TR-PCA is defined in *three* main steps as represented in Algorithm 4.

Algorithm 4 Toroidal PCA via density ridges

Input: $x_1, \dots, x_n \in \mathbb{T}^2$
Step 1: Modeling
(i) Fit the BSvM and/or BWC models with maximum likelihood estimation. If both models are fit, select the one with the smallest Bayesian Information Criterion (BIC).
(ii) Inspect edge cases using LRTs (Sections 3.1 and 4.1 in García-Portugués and Prieto-Tirado [134]) at 5% significance level.
(iii) Retrieve $\hat{f}$, $\hat{\mu}$ (location parameter), and $\hat{f}$ (index of the lowest concentration).
Step 2: Ridge computation
Determine a grid of $\mathcal{R}(\hat{f})$ with the implicit equation approach (using Fourier approximations of the ridge curve).
Step 3: Scores and the Proportion of variance Explained (PVE) computation as described in Sections 5.2 and 5.3 in García-Portugués and Prieto-Tirado [134].

The limitation of TR-PCA to bivariate cases poses a substantial challenge. The complexity arises when attempting to extend this method to multivariate toroidal models. In the multivariate von Mises distribution, obtaining an equivalent of the implicit equation is more challenging, so one may need to rely on the (computationally expensive) Euler algorithm [134]. Also, currently, there does not exist a multivariate extension of the BWC.

6. Application to shape spaces and symmetric spaces

This section explores extensions of PCA and unique techniques for shape and symmetric spaces, emphasizing how these techniques improve the efficiency of PCA in obtaining information from spaces with unique geometric characteristics.

6.1. Statistical shape analysis

Shape, as defined by [136], refers to the geometry of objects that remains invariant under scaling, rotation, and translation. In this framework, a shape is a point on a high-dimensional, non-linear manifold called a shape space, which can be characterized as a hypersphere. To approximate the manifold, a linearized tangent space is commonly used, derived from the tangent plane at a specific point on the shape space. Due to the complexity of working directly in the tangent space of the shape space [137], the tangent space to the pre-shape sphere is considered, identifying the "horizontal part" that is invariant to rotation.

PCA on shape space can be categorized into various types based on different tangent coordinates [10]. The pioneering work of Goodall and Mardia [138], Cootes et al. [139], and Kent [140] marked the initial applications of PCA to shape tangent space. Goodall and Mardia [138] and Cootes et al. [139] employed Procrustes residuals, while Kent [140] proposed using PCA of partial Procrustes coordinates. Larsen and Hilger [141] extended PCA on the tangent space at the full Procrustes mean to describe noise variance across landmarks. For symmetric shape analysis, Kolamunnage and Kent [142] proposed a PCA extension in Procrustes tangent coordinates. Additionally, PCA on quotient spaces derived from an isometric Lie group action on a Riemannian manifold was proposed by Huckemann et al. [62]. Abboud et al. [143] studied outlier detection and robust tangent PCA, while Severn et al. [144] further computed PCA in tangent space, projecting it back onto the shape. For foundational concepts in shape statistics, the readers are encouraged to consult Dryden and Mardia [10] for a thorough introduction.

There are lots of applications in diverse areas where it is important analyze shape data using PCA, such as 3D model face shape variation [145–147], human sex classification from laser-scanned human heads [148], distinguishing between healthy and diseased cardiac ventricles [149], the female facial attractiveness [150], exploring the variability of DNA molecules in shape space [55], morphometric variation of fish scales [151], femur Shape variability in female patients with Hip dysplasia [152], shape clustering in the human body [153], classification of fishes using the Procrustes analysis, clustering the peptides in the dynamic molecule [78].

6.2. Lie group action and symmetric space analysis

An essential tool in image analysis is diffusion tensor magnetic resonance imaging (DT-MRI). DT-MRI generates a 3D diffusion tensor, a 3×3 symmetric and positive-definite matrix, for each voxel in an imaging volume. Notably, the space of diffusion tensors does not constitute a vector space, making standard linear statistical procedures inapplicable. In other words, standard PCA does not preserve the positive definiteness of diffusion tensors. Addressing this challenge, Fletcher and Joshi [154] demonstrated that diffusion tensors exist on a Riemannian symmetric space, and developed methods for producing statistics like mean and modes of variation within this space.

A *Lie group* is a mathematical structure that combines the properties of an algebraic group G with those of a differentiable manifold. In a Lie group, the operations of multiplication and inversion are smooth mappings. Fletcher and his colleagues also introduced PGA to compute the modes of variation of data in Lie groups [155].

Subsequently, Huckemann et al. [62] introduced PCA on quotient spaces arising from an isometric Lie group action on a Riemannian manifold. As the PGA relies on reference points like intrinsic mean and does not account for the curvature of the manifold, Curry et al. [156] introduced a specialized variant known as principal symmetric space analysis, based on nested sequences of totally geodesic submanifolds of symmetric spaces. These submanifolds provide excellent approximating spaces, being the flattest or most straightforward possible lower-dimensional data representations.

6.3. Principal curve

A *principal curve* is a smooth one-dimensional curve that passes through the "middle" of a d-dimensional data set, providing a summary of a nonlinear data set. It is a set of points that effectively represents the mean of data densities. Various definitions of principal curves have been proposed in the literature. Hastie and Stuetzle [157] introduced one of the earliest definitions, based on 'self-consistency,' where the curve coincides at each position with the expected value of the data projected to that position. Tibshirani [158] proposed a probabilistic approach, defining the principal curve as one that minimizes a penalized log-likelihood measure using Gaussian mixtures and generalized EM algorithms. Kégl et al. [159] and Verbeek et al. [160] provided alternative definitions, describing principal curves as continuous curves of a specified length that minimize the expected squared distance between the curve and randomly chosen points from a given distribution.

Since their introduction by Hastie and Stuetzle [157], principal curves have been developed and applied across various fields. To name but a few, for example, they have been used to identify the outlines of ice floes in satellite images [161], construct a principal curve clustering algorithm for spatial point patterns [162], and analyze the nonlinear quality of life index for 171 countries by finding a principal curve through a 4D dataset and using the projections for ranking [163]. In addition, the principal curves have shown close connections to self-organizing maps (SOM) [164,165], captured patterns of variation in seismological data [70], and defined functional data clustering [166].

7. R software

In the past, the use of directional statistics was limited due to the lack of software that could implement its methodology. However, this issue has been partially resolved with the emergence of the R statistical computing environment and its related packages. Currently, several packages are available that support directional statistics. Pewsey and García-Portugués [21],Pewsey [167] provided an overview of many of these packages. The `circular` and `directional` packages in R are essential tools that offer a range of functions for handling, summarizing, visualizing, simulating, modeling, and analyzing circular data. These packages provide the foundational functionality required to work with such data.

The available software for handling spherical and toroidal data is limited. Specifically, the Principal Nested Spheres method can be implemented using the `pns` function in the `shapes` package [79]. Additionally, implementing tangent-based Shape PCA in the `shapes` library is straightforward. The necessary calculations are carried out using the `procGPA` command, and the results are displayed using `shapepca`. Furthermore, Sargsyan et al. [117] released the Python library `Clustangles` for performing GeoPCA. Toroidal ridge PCA is implemented in the R package `ridgetorus` [134] and principal curves are available in the R package `princurve` [168]. Moreover, PGA can be applied using the `PGA` function within the `spherepc` package [169]. Currently, we are not aware of any R package that supports all methodologies, highlighting a potential gap for further research.

8. Conclusions

This paper thoroughly explores various PCA developments for directional data, as summarized in Supplementary Table SM-2. Furthermore, Supplementary Table SM-3 outlines the advantages and disadvantages of PCA-related techniques for directional data. Together, these tables provide a comprehensive reference that details the various PCA methods discussed throughout the paper. This resource is meant to be a valuable guide for both experienced researchers and newcomers, facilitating the exploration of PCA applications within directional data analysis.

We wish to stress that, while most directional data sets lie on the respective manifolds by their nature (e.g., angles on circles), sometimes the manifold assumption model is not met in practice, and data can resemble a stratified space rather than a manifold.

Our review contributes to the literature on directional statistics, shedding light on the practical uses of PCA and providing a foundation for further exploration and application of PCA within the intricate field of directional statistics. We aim to inspire future research work and the development of novel approaches designed to address the unique challenges presented by directional data. In doing so, this review contributes to the ongoing dialog within the research community on the analysis of directional data sets.

Finally, we acknowledge the potential value of a comprehensive software package that unifies the methods discussed in this survey. Currently, this represents a significant gap in the literature, as no integrated framework exists to make these approaches easily accessible for researchers. Developing such a resource, while beyond the scope of this manuscript, would be a valuable contribution to the field. We consider this gap a promising area for future work and a potential focus for further development.

CRediT authorship contribution statement

Anahita Nodehi: Conceptualization, Visualization Writing – original draft, Formal analysis, Project administration. **Meisam Moghimbeygi:** Conceptualization, Writing – review & editing , Formal analysis. **Christophe Ley:** Conceptualization, Writing – review & editing, Formal analysis, Project administration, Supervision.

Acknowledgments

The authors thank Guendalina Palmirotta for her very useful comments on a draft version of this article. They are also very grateful to the Editor, Associate Editor and two anonymous reviewers for their suggestions that helped improve the present paper.

Appendix A. Supplementary data

Supplementary material related to this article can be found online at https://doi.org/10.1016/j.jmva.2025.105528.

References

[1] R. Bellman, Adaptive Control Processes, Princeton University Press, New Jersey, 1961.
[2] C.M. Bishop, Pattern Recognition and Machine Learning, vol. 4, (4) Springer, 2006.
[3] A.J. Izenman, Modern Multivariate Statistical Techniques: Regression, Classification, and Manifold Learning, Springer Science & Business Media, 2009.
[4] R. Zebari, A. Abdulazeez, D. Zeebaree, D. Zebari, J. Saeed, A comprehensive review of dimensionality reduction techniques for feature selection and feature extraction, J. Appl. Sci. Technol. Trends 1 (2) (2020) 56–70.
[5] K. Pearson, On lines and planes of closest fit to systems of points in space, Lond. Edinb. Dublin Philos. Mag. J. Sci. 2 (1901) 559–572.
[6] H. Hotelling, Analysis of a complex of statistical variables into principal components, J. Educ. Psychol. 24 (1933) 417–441.
[7] K.V. Mardia, Statistics of Directional data, Academic Press, London, 1972.
[8] D.G. Kendall, D. Barden, T.K. Carne, H. Le, Shape and Shape Theory, John Wiley & Sons, 1999.
[9] D.G. Kendall, D. Barden, T.K. Carne, H. Le, Shape and Shape Theory, John Wiley & Sons, 2009.
[10] I.L. Dryden, K.V. Mardia, Statistical Shape Analysis: with Applications in R, vol. 995, John Wiley & Sons, 2016.
[11] E. Batschelet, Circular Statistics in Biology, Academic Press, New York, 1981.
[12] G.S. Watson, Statistics on Spheres, Wiley, New York, 1983.

[13] N.I. Fisher, T. Lewis, B.J.J. Embleton, Statistical Analysis Of Spherical Data, Cambridge University Press, Cambridge, 1987.
[14] N.I. Fisher, Statistical Analysis Of Circular Data, Cambridge University Press, Cambridge, 1993.
[15] K.V. Mardia, P. Jupp, Directional Statistics, Wiley, Chichester, 2000.
[16] S.R. Jammalamadaka, A. SenGupta, Topics In Circular Statistics, Multivariate Analysis, World Scientific, Singapore, 2001.
[17] A. Pewsey, M. Neuhäuser, G.D. Ruxton, Circular Statistics In R, Oxford University Press, England, 2013.
[18] C. Ley, T. Verdebout, Modern Directional Statistics, Chapman and Hall/CRC Press, Boca Raton, 2017.
[19] C. Ley, T. Verdebout (Eds.), Applied Directional Statistics: Modern Methods and Case Studies, Chapman and Hall/CRC Press, Boca Raton, 2018.
[20] P.E. Jupp, K.V. Mardia, A unified view of the theory of directional statistics, 1975–1988, Int. Stat. Review/Revue Int. de Stat. 57 (3) (1989) 261–294.
[21] A. Pewsey, E. García-Portugués, Recent advances in directional statistics, Test 30 (1) (2021) 1–58.
[22] S. Huckemann, B. Eltzner, Statistical methods generalizing principal component analysis to non-euclidean spaces, in: Handbook of Variational Methods for Nonlinear Geometric Data, Springer, 2020, pp. 317–338.
[23] S. Huckemann, B. Eltzner, Data analysis on nonstandard spaces, Wiley Interdiscip. Rev.: Comput. Stat. (2020) e1526.
[24] C.C. David, D.J. Jacobs, Principal component analysis: a method for determining the essential dynamics of proteins, in: Protein Dynamics, Springer, 2014, pp. 193–226.
[25] A. Kitao, Principal component analysis and related methods for investigating the dynamics of biological macromolecules, J 5 (2) (2022) 298–317.
[26] I. Jolliffe, Principal Component Analysis, Springer, New York, 2002.
[27] R. Vidal, Y. Ma, S.S. Sastry, Generalized Principal Component Analysis, Springer, 2016.
[28] H. Karcher, Riemannian center of mass and mollifier smoothing, Commun. Pure Appl. Math 30 (5) (1977) 509–541.
[29] X. Pennec, Intrinsic statistics on Riemannian manifolds: Basic tools for geometric measurements, J. Math. Imaging Vis. 25 (2006) 127–154.
[30] W.M. Boothby, An Introduction to Differentiable Manifolds and Riemannian Geometry, Academic Press, New York, 1986.
[31] P.T. Fletcher, C.P. Lu, S. Pizar, S.C. Joshi, Principal geodesic analysis for the study of nonlinear statistics of shape, IEEE Trans. Med. Imaging 23 (2004) 995–1005.
[32] R.L. Bishop, R.J. Crittenden, Geometry of Manifolds, Academic Press, 2011.
[33] J. Nash, C^1 isometric imbeddings, Ann. Math. 60 (3) (1954) 383–396.
[34] J. Nash, The imbedding problem for Riemannian manifolds, Ann. Math. 63 (1) (1956) 20–63.
[35] H. Hopf, W. Rinow, Ueber den Begriff der vollständigen differentialgeometrischen Fläche, Comment. Math. Helv. 3 (1) (1931) 209–225.
[36] V. Boltyanski, H. Martini, V. Soltan, Geometric Methods and Optimization Problems, vol. 4, Springer Science & Business Media, 1998.
[37] R. Basu, B.B. Bhattacharya, T. Talukdar, The projection median of a set of points in d, Discrete Comput. Geom. 47 (2) (2012) 329–346.
[38] J.W. Tukey, Mathematics and the picturing of data, in: Proceedings of the International Congress of Mathematicians, 2, Vancouver, 1975, pp. 523–531.
[39] J.S. Marron, S. Jung, I.L. Dryden, Speculation on the generality of the backward stepwise view of PCA, in: Proceedings of the International Conference on Multimedia Information Retrieval, 2010, pp. 227–230.
[40] J. Damon, J.S. Marron, Backwards principal component analysis and principal nested relations, J. Math. Imaging Vision 50 (1) (2014) 107–114.
[41] S. Jung, X. Liu, J.S. Marron, S. Pizer, Generalized PCA via the backward stepwise approach in image analysis, Brain, Body Mach. 83 (2010) 111–123.
[42] J.S. Marron, A.M. Alonso, Overview of object oriented data analysis, Biom. J. 56 (5) (2014) 732–753.
[43] X. Pennec, Barycentric subspace analysis on manifolds, Ann. Statist. 46 (6A) (2018) 2711–2746.
[44] M. Fréchet, Les éléments aléatoires de nature quelconque dans un espace distancié, Ann. de L'Institut Henri PoincarÉ 10 (1948) 215–310.
[45] W.S. Kendall, Probability, convexity, and harmonic maps with small image I: uniqueness and fine existence, Proc. Lond. Math. Soc. 3 (2) (1990) 371–406.
[46] X. Pennec, Probabilities and statistics on Riemannian manifolds: Basic tools for geometric measurements., in: NSIP, 3, 1999, pp. 194–198.
[47] A. Cauchy, Méthode générale pour la résolution des systèmes d'équations simultanées, Comptes Rendus Hebd. Séances Acad. Sci. 25 (1847) (1847) 536–538.
[48] S. Bubeck, Theory of convex optimization for machine learning, Found. Trends Mach. Learn. 8 (3–4) (2015) 231–358.
[49] F. Cazals, B. Delmas, T. O'Donnell, Fréchet mean and P-mean on the unit circle: Decidability, algorithm, and applications to clustering on the flat torus, in: SEA 2021-19th Symposium on Experimental Algorithms, 2021.
[50] A. Srivastava, E. Klassen, Monte Carlo extrinsic estimators of manifold-valued parameters, IEEE Trans. Signal Process. 50 (2) (2002) 299–308.
[51] R. Bhattacharya, V. Patrangenaru, Nonparametic estimation of location and dispersion on Riemannian manifolds, J. Statist. Plann. Inference 108 (1–2) (2002) 23–35.
[52] S. Huckemann, H. Ziezold, Principal component analysis for Riemannian manifolds, with application to triangular shape spaces, Adv. in Appl. Probab. 38 (2006) 299–319.
[53] S. Huckemann, T. Hotz, Principal component geodesics for planar shape spaces, J. Multivariate Anal. 100 (4) (2009) 699–714.
[54] J.J. Faraway, C.A. Trotman, Shape change along geodesics with application to cleft lip surgery, J. R. Stat. Soc. Ser. C. Appl. Stat. 60 (5) (2011) 743–755.
[55] H. Fotouhi, M. Golalizadeh, Exploring the variability of DNA molecules via principal geodesic analysis on the shape space, J. Appl. Stat. 39 (10) (2012) 2199–2207.
[56] M. Vaillant, M.I. Miller, L. Younes, A. Trouvé, Statistics on diffeomorphisms via tangent space representations, NeuroImage 23 (2004) S161–S169.
[57] Y. Xie, B.C. Vemuri, J. Ho, Statistical analysis of tensor fields, in: Medical Image Computing and Computer-Assisted Intervention–MICCAI 2010: 13th International Conference, Beijing, China, September 20-24, 2010, Proceedings, Part I 13, Springer, 2010, pp. 682–689.
[58] H. Salehian, R. Chakraborty, E. Ofori, D. Vaillancourt, B.C. Vemuri, An efficient recursive estimator of the fréchet mean on a hypersphere with applications to medical image analysis, Math. Found. Comput. Anat. 3 (2015) 143–154.
[59] D. Lazar, L. Lin, Scale and curvature effects in principal geodesic analysis, J. Multivariate Anal. 153 (2017) 64–82.
[60] H. Zhu, T. Li, B. Zhao, Statistical learning methods for neuroimaging data analysis with applications, Annu. Rev. Biomed. Data Sci. 6 (2023) 73–104.
[61] S. Sommer, F. Lauze, M. Nielsen, Optimization over geodesics for exact principal geodesic analysis, Adv. Comput. Math. 40 (2014) 283–313.
[62] S. Huckemann, T. Hotz, A. Munk, Intrinsic shape analysis: Geodesic PCA for Riemannian manifolds modulo isometric Lie group actions, Statist. Sinica 20 (2010) 1–100.
[63] S. Sommer, Horizontal dimensionality reduction and iterated frame bundle development, in: Geometric Science of Information: First International Conference, GSI 2013, Paris, France, August 28-30, 2013. Proceedings, Springer, 2013, pp. 76–83.
[64] M.E. Tipping, C.M. Bishop, Probabilistic principal component analysis, 1997, Technical Report.
[65] S. Roweis, EM algorithms for PCA and SPCA, Adv. Neural Inf. Process. Syst. 10 (1998) 626–632.
[66] M. Zhang, P.T. Fletcher, Probabilistic principal geodesic analysis, 2013, Nevada, United States.
[67] M. Zhang, P.T. Fletcher, Bayesian principal geodesic analysis for estimating intrinsic diffeomorphic image variability, Med. Image Anal. 25 (2015) 37–44.
[68] Y. Zhang, J. Xing, M. Zhang, Mixture probabilistic principal geodesic analysis, in: Multimodal Brain Image Analysis and Mathematical Foundations of Computational Anatomy: 4th International Workshop, MBIA 2019, and 7th International Workshop, MFCA 2019, Springer, 2019, pp. 196–208.
[69] S. Sommer, An infinitesimal probabilistic model for principal component analysis of manifold valued data, Sankhya A 81 (1) (2019) 37–62.
[70] V.M. Panaretos, T. Pham, Z. Yao, Principal flows, J. Amer. Statist. Assoc. 109 (2014) 424–436.
[71] H. Liu, Z. Yao, S. Leung, T.F. Chan, A level set based variational principal flow method for nonparametric dimension reduction on Riemannian manifolds, SIAM J. Sci. Comput. 39 (4) (2017) A1616–A1646.
[72] Z. Yao, Y. Xia, Z. Fan, Fixed boundary flows, 2019, arXiv preprint arXiv:1904.11332.

[73] N.I. Fisher, T. Lewis, B.J.J. Embleton, Statistical Analysis of Spherical Data, Cambridge University Press, 1993.
[74] R.A. Fisher, Dispersion on a sphere, Proc. R. Soc. Lond. Ser. A. Math. Phys. Sci. 217 (1130) (1953) 295–305.
[75] T. Chang, Spherical regression and the statistics of tectonic plate reconstructions, Int. Stat. Review/Revue Int. de Stat. (1993) 299–316.
[76] A. SenGupta, B.C. Arnold, Directional statistics for innovative applications: A bicentennial tribute to florence nightingale, Springer, 2022.
[77] C.R. Anderson, Object recognition using statistical shape analysis (Ph.D. thesis), University of Leeds (Department of Statistics), 1996.
[78] I.L. Dryden, K.R. Kim, C.A. Laughton, H. Le, Principal nested shape space analysis of molecular dynamics data, Ann. Appl. Stat. 13 (4) (2019) 2213–2234.
[79] I.L. Dryden, 'shapes': Statistical Shape Analysis, R package version 1.2.6, 2021.
[80] M. Moghimbeygi, A. Nodehi, Multinomial principal component logistic regression on shape data, J. Classification 39 (2022) 578–599.
[81] A. Banerjee, I.S. Dhillon, J. Ghosh, S. Sra, G. Ridgeway, Clustering on the unit hypersphere using von mises-Fisher distributions., J. Mach. Learn. Res. 6 (9) (2005).
[82] D.G. Kendall, W.S. Kendall, Alignments in two-dimensional random sets of points, Adv. in Appl. Probab. 12 (2) (1980) 380–424.
[83] C.G. Small, Techniques of shape analysis on sets of points, Int. Stat. Rev. (1988) 243–257.
[84] K.V. Mardia, I.L. Dryden, Shape distributions for landmark data, Adv. in Appl. Probab. 21 (4) (1989) 742–755.
[85] P. O'Higgins, I.L. Dryden, Sexual dimorphism in hominoids: further studies of craniofacial shape differences in pan, gorilla and pongo, J. Hum. Evol. 24 (3) (1993) 183–205.
[86] F.L. Bookstein, Biometrics, biomathematics and the morphometric synthesis, Bull. Math. Biol. 58 (1996) 313–365.
[87] C.J. Brignell, I.L. Dryden, S.A. Gattone, B. Park, S. L, W.J. Browne, S. Flynn, Surface shape analysis with an application to brain surface asymmetry in schizophrenia, Biostatistics 11 (4) (2010) 609–630.
[88] T.D. Downs, J. Liebman, Statistical methods for vectorcardiographic directions, IEEE Trans. Biomed. Eng. (1) (1969) 87–94.
[89] A.L. Gould, A regression technique for angular variates, Biometrics (1969) 683–700.
[90] T. Hamelryck, K.V. Mardia, J. Ferkinghoff-Borg, Bayesian methods in structural bioinformatics, Springer, 2012.
[91] K. Siddiqi, S. Pizer, Medial Representations: Mathematics, Algorithms and Applications, vol. 37, Springer Science & Business Media, 2008.
[92] S. Jung, I.L. Dryden, J.S. Marron, Analysis of principal nested spheres, Biometrika 99 (2012) 551–568.
[93] S.M. Pizer, S. Jung, D. Goswami, J. Vicory, X. Zhao, R. Chaudhuri, J.N. Damon, S. Huckemann, J.S. Marron, Nested sphere statistics of skeletal models, in: Innovations for Shape Analysis, Springer, 2013, pp. 93–115.
[94] Y. Guo, B.W.K. Ling, Spherical coordinate-based kernel principal component analysis, Signal, Image Video Process. 15 (2021) 511–518.
[95] B. Schölkopf, A. Smola, K.R. Müller, Kernel principal component analysis, in: International Conference on Artificial Neural Networks, 1997, pp. 583–588.
[96] B. Schölkopf, A. Smola, K.R. Müller, Nonlinear component analysis as a kernel eigenvalue problem, Neural Comput. 10 (5) (1998) 1299–1319.
[97] H. Luo, J.E. Purvis, D. Li, Spherical rotation dimension reduction with geometric loss functions, 2022, arXiv preprint arXiv:2204.10975.
[98] P. Zoubouloglou, E. García-Portugués, J.S. Marron, Scaled torus principal component analysis, J. Comput. Graph. Statist. 32 (2023) 1024–1035.
[99] Y. Mu, P.H. Nguyen, G. Stock, Energy landscape of a small peptide revealed by dihedral angles principal component analysis, Proteins 58 (2005) 45.
[100] A. Altis, P. Nguyen, R. Hegger, G. Stock, Dihedral angles principal component analysis of molecular dynamics simulations, J. Chem. Phys. 26 (2007) 244111.1–244111.
[101] K.V. Mardia, G. Hughes, C.C. Taylor, H. Singh, A multivariate von mises distribution with applications to bioinformatics, Canad. J. Statist. 36 (1) (2008) 99–109.
[102] A. Nodehi, M. Golalizadeh, A. Heydari, Dihedral angles principal geodesic analysis using nonlinear statistics, J. Appl. Stat. 42 (2015) 1962–1972.
[103] B. Eltzner, S. Huckemann, K.V. Mardia, Torus principal component analysis with applications to RNA structure, Ann. Appl. Stat. 12 (2) (2018) 1332–1359.
[104] A. Nodehi, M. Golalizadeh, M. Maadooliat, C. Agostinelli, Torus probabilistic principal component analysis, J. Classification (2025).
[105] K.V. Mardia, H. Wiechers, B. Eltzner, S. Huckemann, Principal component analysis and clustering on manifolds, J. Multivariate Anal. 188 (2022) 104862.
[106] J.J. Fernández-Durán, M.M. Gregorio-Domínguez, 'Circnntsr': an R package for the statistical analysis of circular, multivariate circular, and spherical data using nonnegative trigonometric sums, J. Stat. Softw. 70 (2016) 1–19.
[107] F. Lagona, M. Mingione, Nonhomogeneous hidden semi-Markov models for toroidal data, 2023, arXiv preprint arXiv:2312.14719.
[108] A.C. Walls, Y.J. Park, M.A. Tortorici, A. Wall, A.T. McGuire, D. Veesler, Structure, function, and antigenicity of the SARS-CoV-2 spike glycoprotein, Cell 181 (2) (2020) 281–292.
[109] H. Frauenfelder, S.G. Sligar, P.G. Wolynes, The energy landscapes and motions of proteins, Science 254 (5038) (1991) 1598–1603.
[110] D.J. Wales, Energy landscapes, in: Atomic Clusters and Nanoparticles. Agregats Atomiques Et Nanoparticules: Les Houches Session LXXIII 2–28 July 2000, Springer, 2002, pp. 437–507.
[111] S.A. Adcock, J.A. McCammon, Molecular dynamics: survey of methods for simulating the activity of proteins, Chem. Rev. 106 (5) (2006) 1589–1615.
[112] L. Riccardi, P.H. Nguyen, G. Stock, Free-energy landscape of RNA hairpins constructed via dihedral angle principal component analysis, J. Phys. Chem. B 113 (52) (2009) 16660–16668.
[113] F. Sittel, A. Jain, G. Stock, Principal component analysis of molecular dynamics: On the use of cartesian vs. internal coordinates, J. Chem. Phys. 141 (1) (2014) 07B605_1.
[114] G.G. Maisuradze, A. Liwo, H.A. Scheraga, Relation between free energy landscapes of proteins and dynamics, J. Chem. Theory Comput. 6 (2) (2010) 583–595.
[115] F. Sittel, T. Filk, G. Stock, Principal component analysis on a torus: Theory and application to protein dynamics, J. Chem. Phys. 147 (2017) 244101.1–244101.12.
[116] K. Sargsyan, J. Wright, C. Lim, GeoPCA: a new tool for multivariate analysis of dihedral angles based on principal component geodesics, Nucleic Acids Res. 40 (2012) 25.
[117] K. Sargsyan, Y.H. Hua, C. Lim, Clustangles: An open library for clustering angular data, J. Chem. Inf. Model. 55 (8) (2015) 1517–1520.
[118] J. Janin, S. Wodak, M. Levitt, B. Maigret, Conformation of amino acid side-chains in proteins, J. Mol. Biol. 125 (3) (1978) 357–386.
[119] M.V. Shapovalov, J. Dunbrack, Roland L., Statistical and conformational analysis of the electron density of protein side chains, Proteins: Struct. Funct. Bioinform. 66 (2) (2007) 279–303.
[120] M.V. Shapovalov, J. Dunbrack, L. Roland, A smoothed backbone-dependent rotamer library for proteins derived from adaptive kernel density estimates and regressions, Structure 19 (6) (2011) 844–858.
[121] K.V. Mardia, C.C. Taylor, G.K. Subramaniam, Protein bioinformatics and mixtures of bivariate von mises distributions for angular data, Biometrics 63 (2) (2007) 505–512.
[122] G.N. Ramachandran, C. Ramakrishnan, V. Sasisekharan, Stereochemistry of polypeptide chain configurations, J. Mol. Biol. 7 (1963) 95–99.
[123] P.R. Amarasinghe, L. Allison, P.J. Stuckey, M. Garcia de la Banda, A.M. Lesk, A.S. Konagurthu, Getting '$\phi\psi\chi$al'with proteins: minimum message length inference of joint distributions of backbone and sidechain dihedral angles, Bioinformatics 39 (Supplement_1) (2023) i357–i367.
[124] C.S. Wallace, Statistical and Inductive Inference by Minimum Message Length, Springer Science & Business Media, 2005.
[125] J. Ameijeiras-Alonso, C. Ley, Sine-skewed toroidal distributions and their application in protein bioinformatics, Biostatistics 23 (3) (2022) 685–704.
[126] S. Jung, K. Park, B. Kim, Clustering on the torus by conformal prediction, Ann. Appl. Stat. 15 (4) (2021) 1583–1603.
[127] D. Xu, Y. Wang, Density estimation for toroidal data using semiparametric mixtures, Stat. Comput. 33 (6) (2023) 140.
[128] J.T. Kent, K.V. Mardia, Principal component analysis for the wrapped normal torus model, in: Proceedings of the Leeds Annual Statistical Research (LASR) Workshop, 2009.

[129] A. Nodehi, M. Golalizadeh, M. Maadooliat, C. Agostinelli, Estimation of parameters in multivariate wrapped models for data on a p-torus, Comput. Statist. 36 (1) (2021) 193–215.
[130] I.J. Schoenberg, On certain metric spaces arising from Euclidean spaces by a change of metric and their imbedding in Hilbert space, Ann. Math. 38 (4) (1937) 787–793.
[131] E. Pekalska, R.P.W. Duin, The dissimilarity representation for pattern recognition: foundations and applications, vol. 64, World scientific, 2005.
[132] T.F. Cox, M.A.A. Cox, Multidimensional scaling on a sphere, Comm. Statist. Theory Methods 20 (9) (1991) 2943–2953.
[133] M.A.A. Cox, T.F. Cox, Multidimensional scaling, in: Handbook of Data Visualization, Springer, 2008, pp. 315–347.
[134] E. García-Portugués, A. Prieto-Tirado, Toroidal PCA via density ridges, Stat. Comput. 33 (107) (2023).
[135] C.R. Genovese, M. Perone-Pacifico, I. Verdinelli, L. Wasserman, Nonparametric ridge estimation, Ann. Statist. 42 (4) (2014) 1511–1545.
[136] D.G. Kendall, The diffusion of shape, Adv. in Appl. Probab. 9 (3) (1977) 428–430.
[137] H. Le, On geodesics in euclidean shape spaces, J. Lond. Math. Soc. 2 (2) (1991) 360–372.
[138] C.R. Goodall, K.V. Mardia, A geometrical derivation of the shape density, Adv. in Appl. Probab. 23 (3) (1991) 496–514.
[139] T.F. Cootes, C.J. Taylor, D.H. Cooper, J. Graham, Training models of shape from sets of examples, in: BMVC92, Springer, 1992, pp. 9–18.
[140] J.T. Kent, The complex bingham distribution and shape analysis, J. R. Stat. Soc. Ser. B Stat. Methodol. 56 (2) (1994) 285–299.
[141] R. Larsen, K.B. Hilger, Statistical shape analysis using non-euclidean metrics, Med. Image Anal. 7 (4) (2003) 417–423.
[142] R. Kolamunnage, J.T. Kent, Principal component analysis for shape variation about an underlying symmetric shape, Stoch. Geom. Biological Struct. Images 72 (2003) 137–139.
[143] M. Abboud, A. Benzinou, K. Nasreddine, A robust tangent PCA via shape restoration for shape variability analysis, Pattern Anal. Appl. 23 (2) (2020) 653–671.
[144] K.E. Severn, I.L. Dryden, S.P. Preston, Manifold valued data analysis of samples of networks, with applications in corpus linguistics, Ann. Appl. Stat. 16 (1) (2022) 368–390.
[145] V. Blanz, T. Vetter, A morphable model for the synthesis of 3D faces, in: Proceedings of the 26th Annual Conference on Computer Graphics and Interactive Techniques, 1999, pp. 187–194.
[146] A. Patel, W.A.P. Smith, 3D morphable face models revisited, in: 2009 IEEE Conference on Computer Vision and Pattern Recognition, IEEE, 2009, pp. 1327–1334.
[147] F. Yang, J. Wang, E. Shechtman, L. Bourdev, D. Metaxas, Expression flow for 3D-aware face component transfer, in: ACM SIGGRAPH 2011 Papers, 2011, pp. 1–10.
[148] A.J. O'toole, T. Vetter, N.F. Troje, H.H. Bülthoff, Sex classification is better with three-dimensional head structure than with image intensity information, Perception 26 (1) (1997) 75–84.
[149] R. Pilgram, K.D. Fritscher, R.H. Zwick, M.F. Schocke, T. Trieb, O. Pachinger, R. Schubert, Shape analysis of healthy and diseased cardiac ventricles using PCA, in: International Congress Series, 1281, Elsevier, 2005, pp. 357–362.
[150] D.R. Valenzano, A. Mennucci, G. Tartarelli, A. Cellerino, Shape analysis of female facial attractiveness, Vis. Res. 46 (8–9) (2006) 1282–1291.
[151] A.L. Ibáñez, L.A. Jawad, Morphometric variation of fish scales among some species of rattail fish from New Zealand waters, J. Mar. Biol. Assoc. UK 98 (8) (2018) 1991–1998.
[152] B.M.M. Gaffney, T.J. Hillen, J.J. Nepple, J.C. Clohisy, M.D. Harris, Statistical shape modeling of femur shape variability in female patients with hip dysplasia, J. Orthop. Res. 37 (3) (2019) 665–673.
[153] Y.M. Xing, Z.J. Wang, J.P. Wang, Y. Kan, N. Zhang, X.M. Shi, Human body shape clustering for apparel industry using PCA-based probabilistic neural network, in: Advances in Computer Communication and Computational Sciences, Springer, 2019, pp. 343–354.
[154] P.T. Fletcher, S. Joshi, Principal geodesic analysis on symmetric spaces: Statistics of diffusion tensors, in: Computer Vision and Mathematical Methods in Medical and Biomedical Image Analysis, Springer, 2004, pp. 87–98.
[155] P.T. Fletcher, C. Lu, S. Joshi, Statistics of shape via principal geodesic analysis on Lie groups, in: 2003 IEEE Computer Society Conference on Computer Vision and Pattern Recognition, 2003. Proceedings., 1, IEEE, 2003, p. I.
[156] C. Curry, S. Marsland, R. McLachlan, Principal symmetric space analysis, J. Comput. Dyn. 6 (2) (2019) 251.
[157] T. Hastie, W. Stuetzle, Principal curves, J. Amer. Statist. Assoc. 84 (1989) 502–516.
[158] R. Tibshirani, Principal curves revisited, Stat. Comput. 2 (1992) 183–190.
[159] B. Kégl, A. Krzyzak, T. Linder, K. Zeger, Learning and design of principal curves, IEEE Trans. Pattern Anal. Mach. Intell. 22 (3) (2000) 281–297.
[160] J.J. Verbeek, N. Vlassis, B. Kröse, A k-segments algorithm for finding principal curves, Pattern Recognit. Lett. 23 (8) (2002) 1009–1017.
[161] J.D. Banfield, A.E. Raftery, Ice floe identification in satellite images using mathematical morphology and clustering about principal curves, J. Amer. Statist. Assoc. 87 (417) (1992) 7–16.
[162] D.C. Stanford, A.E. Raftery, Finding curvilinear features in spatial point patterns: principal curve clustering with noise, IEEE Trans. Pattern Anal. Mach. Intell. 22 (6) (2000) 601–609.
[163] A.N. Gorban, A. Zinovyev, Principal manifolds and graphs in practice: from molecular biology to dynamical systems, Int. J. Neural Syst. 20 (03) (2010) 219–232.
[164] F. Mulier, V. Cherkassky, Self-organization as an iterative kernel smoothing process, Neural Comput. 7 (6) (1995) 1165–1177.
[165] H. Yin, Learning nonlinear principal manifolds by self-organising maps, in: Principal Manifolds for Data Visualization and Dimension Reduction, Springer, 2008, pp. 68–95.
[166] R. Wu, B. Wang, A. Xu, Functional data clustering using principal curve methods, Comm. Statist. Theory Methods 51 (20) (2022) 7264–7283.
[167] A. Pewsey, Applied directional statistics with r: an overview, Appl. Dir. Stat. (2018) 293–306.
[168] T. Hastie, A. Weingessel, K. Hornik, H. Bengtsson, R. Cannoodt, M.R. Cannoodt, Package 'princurve', 2022.
[169] J. Lee, J.H. Kim, H.S. Oh, 'Spherepc': An R package for dimension reduction on a sphere, R J. 14 (1) (2022).