

# Spherical Principal Curves

Jongmin Lee\*, Jang-Hyun Kim\*, Hee-Seok Oh

**Abstract**—This paper presents a new approach for dimension reduction of data observed on spherical surfaces. Several dimension reduction techniques have been developed in recent years for non-Euclidean data analysis. As a pioneer work, Hauberg [1] attempted to implement principal curves on Riemannian manifolds. However, this approach uses approximations to process data on Riemannian manifolds, resulting in distorted results. This study proposes a new approach to project data onto a continuous curve to construct principal curves on spherical surfaces. Our approach lies in the same line of Hastie and Stuetzle [2] that proposed principal curves for data on Euclidean space. We further investigate the stationarity of the proposed principal curves that satisfy the self-consistency on spherical surfaces. The results on the real data analysis and simulation examples show promising empirical characteristics of the proposed approach.

**Index Terms**—Dimension reduction, Feature extraction, Principal geodesic analysis, Principal curve, Spherical domain.

## 1 INTRODUCTION

A variety of dimension reduction techniques have been developed to represent and analyze data on Euclidean space. Recently, there have been growing interests in the analysis of non-Euclidean data with a variety of applications; directional data [3]–[5], shape data [6]–[9], and motion analysis [1], [9]. For example, Siddiqi and Pizer [10] and Cippitelli et al. [11] introduced a Cartesian product of sphere  $S^2$  and  $\mathbb{R}$  for medial representation and skeleton data, respectively. For these representations, the conventional dimension reduction methods on Euclidean space have been modified by considering geodesics on non-Euclidean space [7], [8], [12]–[15]. As a study closely related to our proposal, Hauberg [1] developed principal curves on Riemannian manifolds. However, Hauberg [1] uses an approximate method by projecting data onto a *finite* set of points, unlike the original principal curve [2] which projects data onto a *continuous* curve. This approximate projection causes a problem that may project different data points onto a single point mistakenly. This study proposes a new principal curve approach for spherical data by projecting the data onto a continuous curve without any approximations and improves the performance of dimension reduction. Our proposed approach is two-fold: One is an extrinsic approach that requires the setting of additional embedding space for a given manifold. The other is an intrinsic approach that does not need an embedding space. This intrinsic approach is difficult to calculate [16], but it is necessary to develop principal curves on generic manifolds. In this study, we investigate the stationarity of the principal curves on spherical surfaces from both approaches.

The main contributions of this study can be summarized as follows: (a) We propose both extrinsic and intrinsic ap-

proaches to form principal curves on  $d$ -sphere  $S^d$ ,  $d \geq 2$ . (b) We verify the stationarity of the proposed principal curves on  $S^d$ . (c) We show the usefulness of the proposed method through real data analysis and simulation study.

The detailed proofs of the theoretical properties of the proposed principal curves method are given in the Supplementary Material.

## 2 BACKGROUNDS

### 2.1 Principal Curves

The principal curve in [2] can be considered as a nonlinear generalization of PCA that finds an affine subspace maximizing the variance of the projections of data. A curve is a function from one-dimensional closed interval to a given space, and a curve  $f$  is called self-consistent or a *principal curve* of a random variable  $X$  if the curve satisfies

$$f(\lambda) = \mathbb{E}[X|\lambda_f(X) = \lambda], \quad (1)$$

where  $\lambda_f(x)$  is a projection index of a point  $x$  onto the curve  $f$ . It implies that  $f(\lambda)$  is the average of all data points projected onto  $f(\lambda)$  itself. One of the most important consequences of the self-consistency is that the principal curve is a critical point with respect to reconstruction error for small perturbations [2]. However, it is difficult to formulate a principal curve by solving the self-consistency equation of (1). Thus, Hastie and Stuetzle [2] represented a curve as the first order spline, connected by  $T$  points. Then, they iteratively updated the curve to achieve the self-consistency condition using the following two steps, *projection* and *expectation*: (a) In the projection step, the given data are projected onto the curve. (b) In the expectation step,  $T$  points of the curve are updated to satisfy the self-consistency.

### 2.2 Means on Manifolds

Manifold is a topological space that locally resembles a Euclidean space. *Riemannian manifold*  $M$  is a smooth manifold equipped with smoothly varying inner product on tangent space. A *geodesic* is the shortest curve between two

• Jongmin Lee and Hee-Seok Oh are with the Department of Statistics, Seoul National University, Seoul 08826, Korea.  
E-mail: heeseok@stats.snu.ac.kr  
• Jang-Hyun Kim is with the Department of Computer Science and Engineering, Seoul National University, Seoul 08826, Korea.

\* The first two authors contributed equally to this work.

points in  $M$  and its length is called *geodesic distance*, denoted by  $d_{Geo}(\cdot, \cdot)$ . The class of Riemannian manifold includes a variety of spaces, such as Euclidean space  $\mathbb{R}^d$ , sphere  $S^d$  [3]–[5],  $\mathbb{RP}^d$  and  $\mathbb{CP}^d$  (Kendall's shape space; [6]–[8]),  $PD(d)$  (space of  $d \times d$  symmetric positive definite matrices; [9], [17]), and product space of  $S^2$  (medial representations; [10], [12], [13]). For more details about Riemannian manifold, see [18].

The concept of the expected value of a distribution can be naturally extended to manifolds, called *Fréchet mean*. Given a probability distribution  $Q$  on  $M$  with a distance  $\rho(\cdot, \cdot)$ , the Fréchet mean  $m \in M$  is defined as

$$\arg \min_{m \in M} \int \rho^2(m, x) Q(dx).$$

The Fréchet mean with geodesic distance is termed *intrinsic mean* [19]. Meanwhile, by embedding a given manifold  $M$  into Euclidean space  $\mathbb{R}^d$ , the Fréchet mean can be calculated using Euclidean distance in  $\mathbb{R}^d$ , called *extrinsic mean*. With an embedding  $\xi : M \hookrightarrow \mathbb{R}^d$ , the extrinsic mean is defined as

$$\arg \min_{m \in M} \int \|\xi(m) - \xi(x)\|^2 Q(dx).$$

It is equivalent to the projection of the expectation in  $\mathbb{R}^d$  to  $M$  [19]. That is, given a projection mapping  $\pi : \mathbb{R}^d \rightarrow M$  defined as  $\pi(y) = \arg \min_{m \in M} \|\xi(m) - y\|$ , the extrinsic mean can be calculated as  $\pi(\int \xi(x) Q(dx))$ . The extrinsic mean is computationally efficient compared to the intrinsic mean [20], and for a distribution  $Q$  supported in a small region, the extrinsic mean is close with the intrinsic mean [21].

### 2.3 Principal Curves on Riemannian Manifolds

Hauberg [1] proposed principal curves on Riemannian manifolds by expressing a curve as a set of  $T$  points,  $f = \{C_1, \dots, C_T\}$ , joined by geodesics. The estimation algorithm of the curve follows that of [2] with an approximation. Specifically, the mean operation in the expectation step is performed by intrinsic mean, and the projection is conducted by finding the nearest point in  $f$  as

$$\text{proj}(x) = \arg \min_{C_i \in f} d_{Geo}(x, C_i),$$

which is not an *exact* projection onto the continuous curve.

## 3 PROPOSED PRINCIPAL CURVES

This section presents our new *exact* principal curves on  $d$ -sphere  $S^d$  for  $d \geq 2$  from both intrinsic and extrinsic perspectives. We further investigate the stationarity of the proposed principal curves.

### 3.1 Exact Projection Step on $S^d$

As mentioned in Section 2.3, the approach of Hauberg [1] does not perform the exact projections onto curves. On the other hand, the exact projections on  $S^d$  for  $d \geq 2$  are carried out in our method, which results in more elaborated principal curves. To this end, we parameterize the curve as a set of  $T$  points joined by geodesics as in [1]. Specifically, we first project the data point to each geodesic segment of the curve and then obtain the exact projection on the curve

by choosing the closest geodesic segment. Let  $\lambda_f(x)$  be the projection index of a point  $x$  to the curve  $f(\lambda)$  for  $\lambda \in [0, 1]$ ,

$$\lambda_f(x) = \arg \min_{\lambda} d_{Geo}(x, f(\lambda)). \quad (2)$$

The projection of  $x$  onto the curve can be obtained as  $f(\lambda_f(x))$ .

The following subsections describe a procedure for projecting a point onto a geodesic segment on  $S^d$ . Given  $A, B, C \in S^d \subset \mathbb{R}^{d+1}$ , we find the closest point to  $C$  on the geodesic segment joining  $A$  and  $B$ . When  $A = B$ , the process is obvious, and in the case of  $A = -B \in \mathbb{R}^{d+1}$ , there is no unique geodesic connecting  $A$  and  $B$ . Hence, we only consider the case that  $A$  and  $B$  are linearly independent, *i.e.*,  $(A \cdot B)^2 \neq 1$ , where  $\cdot$  denotes the dot product in  $\mathbb{R}^{d+1}$ . We first deal with the projection on  $S^2$  and then extend it into hyperspherical cases.

#### 3.1.1 Projection on $S^2$

Before describing the projection procedure on  $S^2$ , it is important to notice that  $(A \cdot B)^2 \neq 1$  is equivalent to  $A \times B \neq 0$ , where  $\times$  denotes the cross product in  $\mathbb{R}^3$ . In addition, if  $A \times B / \|A \times B\| = \pm C$ , then any points on geodesic through  $A$  and  $B$  have the same distance from  $C$ . From now on, we assume  $A \times B / \|A \times B\| \neq \pm C$ .

Figure 1 shows the projection procedure. We define the North Pole  $N$  concerning  $A$  and  $B$  as  $N = \frac{A \times B}{\|A \times B\|} \in S^2$  and a center  $Q$  of the great circle through  $N$  and  $C$  as  $Q = \frac{N \times C}{\|N \times C\|} \in S^2$  that is contained in the great circle through  $A$  and  $B$ . Then, the projection of  $C$  onto the great circle through  $A$  and  $B$ ,  $\text{proj}(C)$ , becomes an intersection point of two great circles, as shown in Figure 1a,

$$\text{proj}(C) = Q \times N = \frac{(A \times B) \times C}{\|(A \times B) \times C\|} \times \frac{(A \times B)}{\|A \times B\|} \in S^2.$$

Note that  $\text{proj}(C)$  is not always included in the geodesic segment  $\widehat{AB}$  joining  $A$  and  $B$  as Figure 1b. For this reason, we define an indicator  $I = -(A - \text{proj}(C)) \cdot (B - \text{proj}(C))$ , indicating whether  $\text{proj}(C)$  is inside  $\widehat{AB}$  or not, *i.e.*, orthogonally projected onto  $\widehat{AB}$  or not. Finally, the projection of  $C$  onto  $\widehat{AB}$ ,  $\text{proj}_{\widehat{AB}}(C)$ , is

$$\text{proj}_{\widehat{AB}}(C) = \begin{cases} \text{proj}(C), & \text{if } I \geq 0 \\ \arg \min_{E \in \{A, B\}} d_{Geo}(C, E), & \text{if } I < 0. \end{cases}$$

#### 3.1.2 Projection on Hypersphere

For  $A, B, C \in S^d \subset \mathbb{R}^{d+1}$ , if  $B \cdot C = C \cdot A = 0$ , then all points on  $\widehat{AB}$  have the same geodesic distance of  $\pi/2$  from  $C$ , which is verified in Appendix A of the Supplementary Material; hence, assume that  $A, B$ , and  $C$  do not satisfy  $B \cdot C = C \cdot A = 0$ . Let  $V$  be a two-dimensional vector space in  $\mathbb{R}^{d+1}$  spanned by  $A$  and  $B$ .

As shown in Figure 2, we aim to find the projection of  $C$  onto  $V \cap S^d$ ,  $\text{proj}(C)$ , by following two steps: (Step 1) Locate the projection of  $C$  onto  $V$ ,  $C'$ . (Step 2) Find the projection of  $C'$  onto  $V \cap S^d$ . Note that the resulting projection is equivalent to the projection of  $C$  onto  $V \cap S^d$ ,  $\text{proj}(C)$ . The rigorous justification of the above procedure is provided in Appendix A of the Supplementary Material.

(Step 1): We find the closest point  $C' \in V$  from  $C$ . Let  $C' = \mu A + \lambda B$  for  $\mu, \lambda \in \mathbb{R}$ . Then  $C'$  should satisfy the

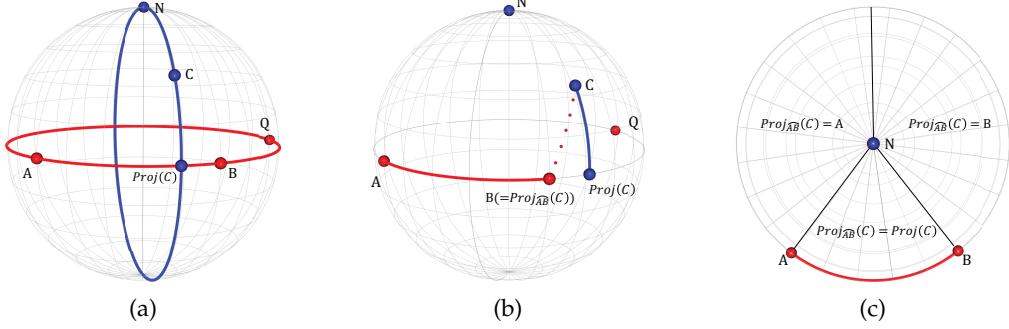


Figure 1: Illustration of the projection procedure on  $S^2$ . (a) The case that  $C$  is projected inside  $\widehat{AB}$ , i.e.,  $\text{proj}_{\widehat{AB}}(C) = \text{proj}(C)$  and  $I \geq 0$ . The projection of  $C$  is an intersection point of two great circles. (b) The case that  $C$  is projected onto  $B$  in a non-orthogonal way (red dotted line), i.e.,  $\text{proj}(C) \neq \text{proj}_{\widehat{AB}}(C) = B$  and  $I < 0$ . (c) An image of the sphere viewed from above the Northern Hemisphere in the projection of  $C$ .

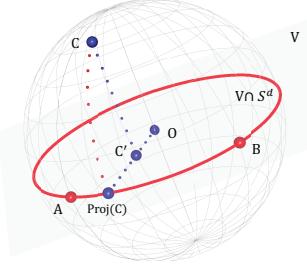


Figure 2: Illustration of the projection procedure on  $S^d$ .

orthogonal condition,  $(C - C') \cdot A = (C - C') \cdot B = 0$ . By plugging the equation  $C' = \mu A + \lambda B$  into the above condition and solving the systems of linear equations with respect to  $\mu$  and  $\lambda$ , it follows that

$$C' = \frac{C \cdot A - (A \cdot B)(B \cdot C)}{1 - (A \cdot B)^2} A + \frac{B \cdot C - (A \cdot B)(C \cdot A)}{1 - (A \cdot B)^2} B,$$

where the denominator is non-zero and  $C' \neq 0 \in \mathbb{R}^{d+1}$  because of the assumptions;  $(A \cdot B)^2 \neq 1$ , and  $A$ ,  $B$ , and  $C$  do not satisfy  $B \cdot C = C \cdot A = 0$ .

(Step 2): The projection of  $C'$  onto  $V \cap S^d$ ,  $\text{proj}(C')$ , is obtained by just normalizing  $C'$  so that it is in  $S^d$ . Therefore, we have

$$\begin{aligned} \text{proj}(C) &= \frac{C'}{\|C'\|} \\ &= \frac{(C \cdot A - (A \cdot B)(B \cdot C))A + (B \cdot C - (A \cdot B)(C \cdot A))B}{\|(C \cdot A - (A \cdot B)(B \cdot C))A + (B \cdot C - (A \cdot B)(C \cdot A))B\|}. \end{aligned}$$

Similarly, we define the indicator  $I = -(A - \text{proj}(C)) \cdot (B - \text{proj}(C))$  to find the projection of  $C$  onto  $\widehat{AB}$ ,  $\text{proj}_{\widehat{AB}}(C)$ . Due to the fact that  $A$ ,  $B$ , and  $\text{proj}(C)$  are in the one-dimensional unit circle  $V \cap S^d$ , we obtain  $I \neq 0$  unless  $\text{proj}(C) = A$  or  $B$ . Since  $I$  is continuous with respect to  $\text{proj}(C) \in V \cap S^d$ , it indicates that whether  $\text{proj}(C)$  is in  $\widehat{AB}$  or not. We finally obtain  $\text{proj}_{\widehat{AB}}(C)$  as

$$\text{proj}_{\widehat{AB}}(C) = \begin{cases} \text{proj}(C), & \text{if } I \geq 0 \\ \arg \min_{E \in \{A, B\}} d_{\text{Geo}}(C, E), & \text{if } I < 0. \end{cases}$$

Note that the distance between  $C$  and  $\widehat{AB}$  is the geodesic

distance from  $C$  to  $\text{proj}_{\widehat{AB}}(C)$ , which can be calculated as

$$d_{\text{Geo}}(C, \text{proj}_{\widehat{AB}}(C)) = \arccos(C \cdot \text{proj}_{\widehat{AB}}(C)). \quad (3)$$

### 3.2 Expectation Step on $S^d$

The expectation step follows the principal curve of Hauberg [1], i.e., updates the weighted average with smoothing that makes the curve closer to the self-consistency condition. Suppose that we have  $n$  data points  $D = \{x_i\}_{i=1}^n$  and the corresponding projection indices  $\{\lambda_i\}_{i=1}^n$ , where  $\lambda_i = \lambda_f(x_i)$  for  $i = 1, \dots, n$ . Let  $T$  denote the number of points of an initial curve. Then, the local weighted smoother iteratively updates the  $t^{\text{th}}$  point of the principal curve,  $C_t$ , with the weighted mean of data points. In this study, we use a quadratic kernel  $k(\lambda) = (1 - \lambda^2)^2 \cdot \delta_{|\lambda| \leq 1}$ , as [1], and the weight of each data point is given by  $w_{t,i} = k(|\lambda_f(C_t) - \lambda_i|/\sigma)$ , where  $\sigma = q \cdot (\text{length of } f)$ .

#### 3.2.1 Extrinsic Approach

The extrinsic mean on  $S^d$  can be calculated by considering the canonical embedding  $S^d \hookrightarrow \mathbb{R}^{d+1}$ . Specifically, for a curve  $f = \{C_1, \dots, C_T\}$  and each point  $C_t$ , the extrinsic mean is obtained by averaging the data points represented in Euclidean coordinates as

$$m_t(D, f) = \sum_{i=1}^n w_{t,i} x_i / \left\| \sum_{i=1}^n w_{t,i} x_i \right\|, \quad t = 1, \dots, T \quad (4)$$

where  $\|\cdot\|$  is the standard norm in  $\mathbb{R}^{d+1}$ . Then  $C_t$  is updated by  $m_t(D, f)$ . The extrinsic approach is advantageous in terms of the computational complexity compared to the intrinsic approach. Furthermore, the extrinsic way ensures the stationarity of the principal curves on hyperspheres  $S^d$  for  $d \geq 2$ , which is discussed in Section 3.4.

#### 3.2.2 Intrinsic Approach

From the intrinsic perspective, the weighted mean of data points can be obtained by the optimization

$$m_t(D, f) = \arg \min_x \sum_{i=1}^n w_{t,i} d_{\text{Geo}}^2(x, x_i), \quad t = 1, \dots, T, \quad (5)$$

and then each  $C_t$  is updated by  $m_t(D, f)$ . The intrinsic mean exists uniquely if the points are in an open hemisphere

of  $S^d$ , i.e.,  $\exists p \in S^d$  s.t.  $d_{Geo}(x_i, p) < \frac{\pi}{2}$  for  $1 \leq i \leq n$  [22]. Since the intrinsic mean cannot be obtained in a closed form, to solve Equation (5), algorithms based on tangent space approximation, such as [12], [22], can be used.

Before closing this section, as an alternative measure of the centrality of data, the geometric median can be considered to robustify the principal curves for a dataset that might contain outliers instead of the extrinsic or intrinsic mean. Median principal curves and their associated characteristics can be developed along with the same line of our procedure. Because of the limitation of space, this part is not discussed in the current paper.

### 3.3 Algorithm

#### 3.3.1 Initialization

For a better estimation of principal curves, we initialize a principal curve as a circle on  $d$ -sphere  $S^d$ . The detailed descriptions of the circle and its algorithm are provided in Appendix B of the Supplementary Material.

#### 3.3.2 Spherical Principal Curves

The proposed spherical principal curves on  $S^d$  can be obtained by Algorithm 1 below.

---

#### Algorithm 1 Spherical Principal Curves

---

```

Initialize curve  $f = \{C_1, \dots, C_T\}$ .
Parameterize the curve as  $f(\lambda)$  by unit speed.
Calculate  $\lambda_f(x_i)$  in Equation (2) for  $i = 1, \dots, n$ .
Calculate errors  $\delta(D, f) = \sum_{i=1}^n d_{Geo}^2(x_i, f(\lambda_f(x_i)))$ .
while ( $\Delta\delta(D, f) \geq \text{threshold}$ ) do
    (Expectation)  $C_t \leftarrow m_t(D, f)$  for  $t = 1, \dots, T$ .
    Reparameterize the curve by unit speed.
    (Projection) Calculate  $\lambda_f(x_i)$  for  $i = 1, \dots, n$ .
    Calculate  $\delta(D, f) = \sum_{i=1}^n d_{Geo}^2(x_i, f(\lambda_f(x_i)))$ .
end while

```

---

Note that  $d_{Geo}^2(x_i, f(\lambda_f(x_i)))$  is calculated by (3). As far as Euclidean space is concerned as embedding space, the extrinsic approach is advantageous for computational efficiency [20]. However, if the data points are not contained within local regions at the expectation step, the intrinsic method may have better performances than the extrinsic one. Furthermore, the intrinsic approach can be attractive because of its inherent metric.

### 3.4 Stationarity of Principal Curves

For a random vector  $X$  in  $\mathbb{R}^d$ ,  $d \in \mathbb{N}$ , the stationarity of the principal curve of  $X$  is given by [2] as

$$\frac{\partial \mathbb{E}_X[d^2(X, f + \epsilon g)]}{\partial \epsilon} \Big|_{\epsilon=0} = 0, \quad (6)$$

where  $f$  and  $g$  are smooth curves in  $\mathbb{R}^d$  satisfying  $\|g\| \leq 1$  and  $\|g'\| \leq 1$ , and  $d(X, f)$  denotes the distance from  $X$  to the curve  $f$ .

However, since spheres are not vector spaces such as  $\mathbb{R}^d$ , additions are not directly defined on spheres. Thus, it is necessary to redefine some concepts, such as addition

and perturbation, in order to extend the properties of the principal curves in Euclidean space to spheres. To this end, we conversely consider  $f + g$  instead of  $g$ . Specifically, let  $f$  and  $f + g$  be smooth curves on  $d$ -sphere parameterized with  $\lambda \in [0, 1]$ . Then, we define  $f + \epsilon g$  in a pointwise sense as follows.

**Definition 1.** For  $a, b \in S^d$  and  $\epsilon \in [0, 1]$ ,  $\text{div}(a, b, \epsilon)$  is a set of points on geodesics between  $a$  and  $b$  satisfying  $\forall c \in \text{div}(a, b, \epsilon)$ ,  $d_{Geo}(a, c) = \epsilon d_{Geo}(a, b)$  and  $c$  is on a geodesic between  $a$  and  $b$ .

Note that if  $d_{Geo}(a, b) < \pi$ , then the geodesic between  $a$  and  $b$  on  $S^d$  is unique. In this case,  $\text{div}(a, b, \epsilon)$  is a single point set and  $\text{div}(a, b, -\epsilon)$  can be defined as a reflection of  $\text{div}(a, b, \epsilon)$  with respect to  $a$ .

**Definition 2.** Let  $f$  and  $f + g$  be smooth curves on  $S^d$  parameterized with  $\lambda \in [0, 1]$  satisfying  $\|g\| < \pi$ , where  $\|g\| := \max_{\lambda \in [0, 1]} d_{Geo}(f(\lambda), (f+g)(\lambda))$ . Then, for  $\epsilon \in [-1, 1]$ ,  $f + \epsilon g$  is a curve on  $S^d$ , where  $(f + \epsilon g)(\lambda) = \text{div}(f(\lambda), (f+g)(\lambda), \epsilon)$ ,  $\forall \lambda \in [0, 1]$ .

Note that  $f + \epsilon g$  is a smooth curve on  $S^d$ . For a detailed proof, refer to Appendix C in the Supplementary Material. Let  $X$  be a random vector on  $S^d$  that has a probability density function. Then, we call  $f$  as an *extrinsic principal curve* of  $X$ , if  $f$  is self-consistent with  $X$  in the embedding space as

$$\pi(\mathbb{E}[\xi(X)|\lambda_f(X) = \lambda]) = f(\lambda) \text{ for a.e. } \lambda,$$

where  $\xi : S^d \hookrightarrow \mathbb{R}^{d+1}$  is the canonical embedding and  $\pi : \mathbb{R}^{d+1} \setminus \{0\} \rightarrow S^d$  by  $X \rightarrow \frac{X}{\|X\|}$  is the standard projection (retraction) from  $\mathbb{R}^{d+1}$  to  $S^d$ . In analogy to Equation (6), we provide the following theorem on spheres. Note that  $\cdot$  represents dot-product and  $d_{Geo}(X, f + \epsilon g)$  denotes the geodesic distance from  $X$  to the curve  $f + \epsilon g$ .

**Definition 3.**  $\|g'\| := \max_{\lambda \in [0, 1]} \|g'(\lambda)\|$ , where  $\|g'(\lambda)\| = \max_{\epsilon \in [0, 1]} \left\| \frac{\partial^2(f + \epsilon g)(\lambda)}{\partial \lambda \partial \epsilon} \right\|$ .

**Theorem 1.** Let  $f, f + g : [0, 1] \rightarrow S^d$ ,  $d \geq 2$  be smooth curves satisfying  $\|g\| < \pi$  and  $\|g'\| \leq 1$ . Let  $X$  be a random vector on  $S^2$  or a random vector on  $S^d$ ,  $d \geq 3$  with  $X \in C(\zeta)$ , where  $C(\zeta) := \{x \in S^d \mid |f''(\lambda_f(x)) \cdot x| > \zeta\}$  for a small  $\zeta > 0$ . Then  $f$  is an extrinsic principal curve of  $X$  if and only if

$$\frac{\partial \mathbb{E}_X[\cos(d_{Geo}(X, f + \epsilon g))] }{\partial \epsilon} \Big|_{\epsilon=0} = 0. \quad (7)$$

*Proof.* See Appendix C in the Supplementary Material.  $\square$

Note that since  $2 - 2 \cos x \approx x^2$  for small  $x$ , Equation (7) can be interpreted as an analogy of Equation (6).

We further consider the intrinsic perspective of the stationarity. We define a curve  $f$  as an *intrinsic principal curve* of  $X$  if the intrinsic mean of  $X$  conditioned on  $\lambda_f(X) = \lambda$  is equal to  $f(\lambda)$  for a.e.  $\lambda$ ,

$$\mathbb{E}_{int}[X|\lambda_f(X) = \lambda] = f(\lambda) \text{ for a.e. } \lambda,$$

where  $\mathbb{E}_{int}[\cdot]$  represents an intrinsic mean of a random variable on  $S^d$ .

Note that the intrinsic mean of a random variable  $Y$  on  $S^d$  is unique if  $d_{Geo}(Y, p) < \frac{\pi}{2}$  a.s. for  $\exists p \in S^d$ , i.e., the

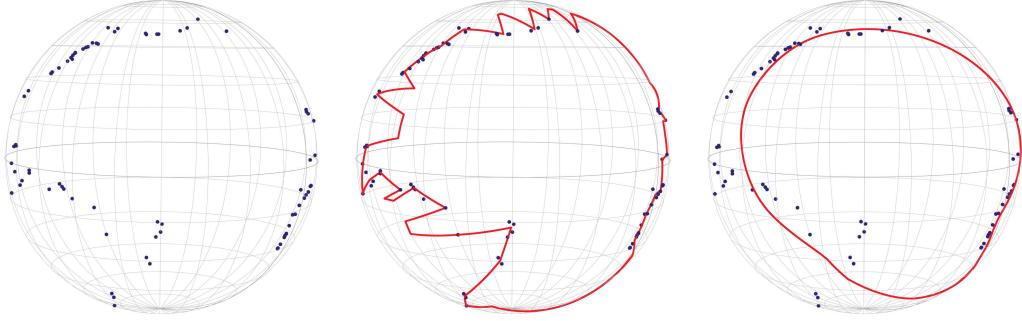


Figure 3: The distribution of earthquake data (left). The proposed extrinsic principal curves of  $T = 500$  with  $q = 0.01$  (middle) and  $q = 0.2$  (right). Blue points represent the observations and red lines are the fitted curves.

support of  $Y$  is in an open hemisphere [23]. We verify that the intrinsic principal curves on  $S^2$  satisfy the stationarity.

**Theorem 2.** Let  $f, f + g : [0, 1] \rightarrow S^2$  be smooth curves satisfying  $\|g\| < \pi$  and  $\|g'\| \leq 1$ . Let  $X$  a random vector on  $S^2$  with  $X \in B(\zeta)$ , where  $B(\zeta) := \{x \in S^2 \mid |f''(\lambda_f(x)) \cdot x| > \zeta\}$  for a small  $\zeta > 0$ . Then,  $f$  is an intrinsic principal curve of  $X$  if and only if

$$\frac{\partial \mathbb{E}_X[d_{Geo}^2(X, f + \epsilon g)]}{\partial \epsilon} \Big|_{\epsilon=0} = 0. \quad (8)$$

*Proof.* See Appendix C in the Supplementary Material.  $\square$

The constraints  $C(\zeta)$  and  $B(\zeta)$  in Theorems 1 and 2 are required to ensure the differentiation of the projection index  $\lambda_{f+\epsilon g}(X)$  with respect to  $\epsilon$ . Note that the constraints are almost negligible by setting  $\zeta$  infinitesimally small; see Lemmas 4 and 6 of Appendix C in the Supplementary Material for details.

We finally remark that the stationarity of the principal curves in Euclidean space provides a rationale for the principal curves of [2] that is a nonlinear generalization of the linear principal component. Following the same line, the above stationarity results provide a theoretical justification that the proposed approaches directly generalize the principal curves by [2] from Euclidean space to spheres. In the intrinsic approach, the case of  $S^d$  with  $d \geq 3$  remains a challenge.

## 4 NUMERICAL EXPERIMENTS

This section conducts numerical experiments with real data analysis and simulated examples to assess the practical performance of the proposed methods. The experiment can be reproduced at <https://github.com/Janghyun1230/Spherical-Principal-Curve>. Moreover, we provide R package, **spherepc** at <https://cran.r-project.org/package=spherepc>, which implements the spherical principal curves for a variety of datasets lying on  $S^2$ .

### 4.1 Real Data Analysis

#### 4.1.1 Earthquake Data on $S^2$

We consider earthquake data from the U.S. Geological Survey (<https://earthquake.usgs.gov/earthquakes/map/>) in Figure 3 that represent the distribution of significant earthquakes (8+ Mb magnitude) around the Pacific Ocean since

1900. As shown in the figure, 77 observations are distributed in the vicinity of the borders between the Pacific, Eurasian, and Nazca plates. Since the plates are gradually moving towards different directions, recognizing the unrevealed patterns of borders provides essential information about seismological events such as earthquakes and volcanoes [3], [24]. In the following experiment, we utilize the spherical principal curves to recover the plates' borders by extracting curvilinear features of the observations.

We have implemented the proposed principal curves connected by  $T = 500$ , with various values of hyperparameter  $q$  that is the bandwidth of kernel in the expectation step. Figure 3 shows the results with  $q = 0.01$  and  $0.2$ . We observe that a small  $q$  produces a wiggly and overfitted curve. It is noteworthy that the choice of  $q$  affects the quality of the fitted curve. Duchamp and Stuetzle [25] proved that principal curves are always the saddle point of the expectation of the squared distance from a particular random variable, pointing out that cross-validation is not reliable for the model selection of principal curves, *i.e.*, determination of  $q$ . Kégl et al. [26] defined principal curves that minimize reconstruction errors in the constraint of the curve length, but used a heuristic way to determine the corresponding hyperparameter, the length of the curves. In the current study, the value of  $q$  is selected by visual inspection through all our experiments. An objective way to select  $q$  is left for future research.

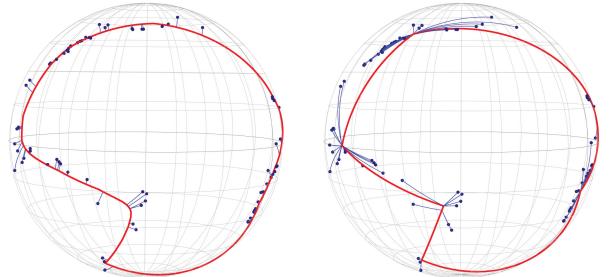


Figure 4: Projection results by the proposed extrinsic method (left) and Hauberg's method (right) with  $T = 77$  and  $q = 0.1$ .

As one can see, the proposed extrinsic curve represents a given data as a continuous curve, while the Hauberg method projects several local data at one point.

We further compare the proposed extrinsic principal curves with the method of Hauberg [1]. Figure 4 shows both results with  $q = 0.1$ , where the purple lines represent the fitted curves, and the blue lines represent the projections from the data to the curve. The proposed extrinsic principal curve continuously represents the given data on the curve, while the method of Hauberg projects several local points to a single location. The comparison is further summarized in Table 1. As a result, the number of distinct projections (# proj) by our method is much larger than that of Hauberg’s method. It implies that the proposed principal curve continuously represents the data, whereas the method of Hauberg tends to cluster the data. We also measure a reconstruction error (RE) defined as  $\sum_{i=1}^n d_{Geo}^2(x_i, \hat{f}(\lambda_{\hat{f}}(x_i)))$  with observations  $\{x_i\}_{i=1}^n$  and fitted values  $\{\hat{f}(\lambda_{\hat{f}}(x_i))\}_{i=1}^n$ . As listed in Table 1, our method outperforms Hauberg’s method in terms of the reconstruction error.

Table 1: The values of RE and # proj by the proposed methods and Hauberg’s method on the earthquake data

		Extrinsic	Intrinsic	Hauberg	
$T = 77$	$q = 0.2$	RE # proj	2.662 74/77	4.391 72/77	12.067 22/77
	$q = 0.1$	RE # proj	0.463 76/77	0.467 76/77	4.920 9/77
	$q = 0.05$	RE # proj	0.359 74/77	0.359 73/77	1.313 16/77
	$q = 0.01$	RE # proj	0.061 75/77	0.061 75/77	0.227 27/77
$T = 500$	$q = 0.2$	RE # proj	2.193 75/77	3.460 72/77	11.300 30/77
	$q = 0.1$	RE # proj	0.715 75/77	0.732 74/77	3.903 18/77
	$q = 0.05$	RE # proj	0.298 75/77	0.200 75/77	0.963 27/77
	$q = 0.01$	RE # proj	0.036 75/77	0.036 75/77	0.121 37/77

#### 4.1.2 Motion Capture Data on $S^2$

We now consider a benchmark data on  $S^2$ , motion capture data of a person walking in a circular pattern [1], [9], [27], [28]. The data represent the orientation of the person’s left *thigh bone* and naturally lie on  $S^2$ . There are 338 data points in the data set that are periodic.

Figure 5 shows both results with  $q = 0.03, 0.05$ , where the red and yellow lines represent the fitted curves, and the blue lines represent the projections from the data to the curves. The proposed extrinsic principal curve continuously represents the given data on the curve, while the method of Hauberg projects several local points to a single location. Furthermore, Table 2 lists the quantitative results of the proposed methods and the method of Hauberg [1]. As listed, the proposed methods outperform Hauberg’s method in terms of the reconstruction error and represent the data more precisely.

## 4.2 Simulation Study

### 4.2.1 Simulation on $S^2$

We consider two types of functions on the unit sphere with spherical coordinates  $(r = 1, \theta, \phi)$ , where  $\theta$  is the azimuthal

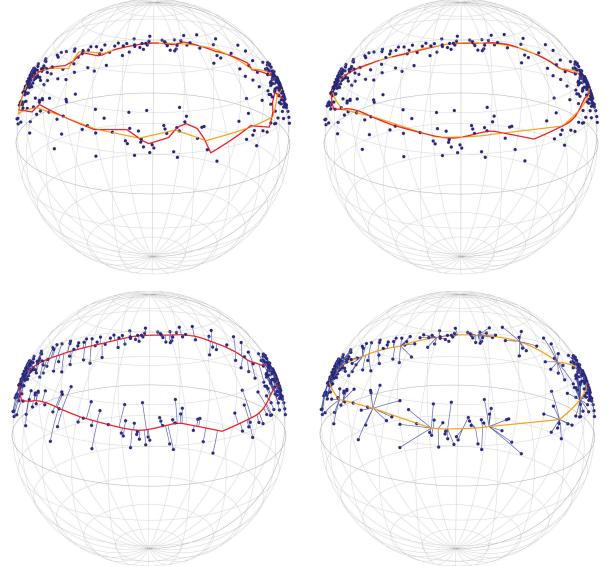


Figure 5: Results by the proposed extrinsic method (red) and Hauberg’s method (yellow) with  $T = 100$ . From top left to bottom right, results with  $q = 0.03$  and  $q = 0.05$ , projection results by the two methods (blue) with  $q = 0.05$ .

Table 2: The values of RE and # proj by the proposed methods and Hauberg’s method on the motion capture data

		Extrinsic	Intrinsic	Hauberg	
$T = 500$	$q = 0.05$	RE # proj	2.502 336/338	2.504 337/338	2.534 223/338
	$q = 0.03$	RE # proj	1.741 332/338	1.741 333/338	2.637 119/338
	$q = 0.01$	RE # proj	0.669 315/338	0.669 317/338	1.253 92/338

angle and  $\phi$  is the polar angle: (Circle) it is formed of  $(r = 1, \theta, \phi)$  with  $0 \leq \theta < 2\pi$  and  $\phi = \pi/4$ . (Wave) it is defined as  $(r = 1, \theta, \phi)$  with  $0 \leq \theta < 2\pi$  and  $\phi = \alpha \sin(\theta f) + \pi/2$ , where the frequency  $f = 4$  and the amplitude  $\alpha = 1/3$ .

For each type of functions, we generate  $n = 100$  data points by sampling  $\theta$  uniformly in  $[0, 2\pi]$  and adding Gaussian noises sampled from  $N(0, \sigma^2)$  to  $\phi$ . Figure 6 shows the results on the waveform data with  $T = 500$  and  $q = 0.05$ . Both extrinsic and intrinsic principal curves extract the true waveform effectively, while Hauberg’s approach yields a rather sharp curve. In the Supplementary Material, we provide additional visual results with various parameter settings.

We next quantify the performance of the proposed methods by measuring a reconstruction error between the fitted and true curves to measure the reconstruction ability of the methods. For the fitted curve  $\hat{f}$ , the reconstruction error is defined as  $\sum_{i=1}^n d_{Geo}^2(x_i, \hat{f}(\lambda_{\hat{f}}(\tilde{x}_i)))$ , where  $\{x_i\}_{i=1}^n$  denote the true values of the generating curves and  $\{\tilde{x}_i\}_{i=1}^n$  denote noisy sample values. We also count the number of distinct projection points to evaluate the continuity of resulting curves of the methods. Moreover, we compare the proposed spherical principal curves with Hauberg’s method over various settings  $T = 100, 500$ ,  $q = 0.05, 0.03, 0.01$ , and  $\sigma = 0.07, 0.1$ .

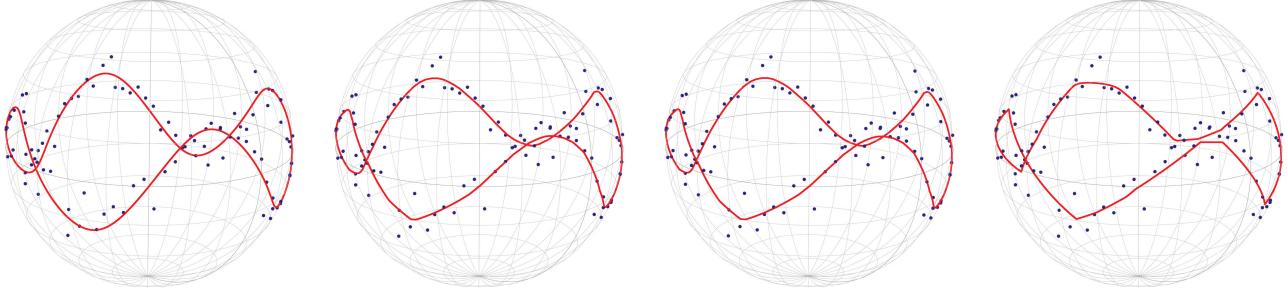


Figure 6: From left to right: True waveform and noisy data (blue dots), the extrinsic principal curve, the intrinsic principal curve, and the curve by Hauberg’s method with  $T = 100$ .

Table 3: Averages of reconstruction errors and their standard deviations in the parentheses by each method

	True form	Method	Noise level					
			$\sigma = 0.07$			$\sigma = 0.1$		
			$q = 0.05$	$q = 0.03$	$q = 0.01$	$q = 0.05$	$q = 0.03$	$q = 0.01$
$T = 100$	Circle	Proposed (intrinsic)	0.093 (0.026)	0.12 (0.027)	0.095 (0.013)	0.201 (0.048)	0.216 (0.046)	0.137 (0.025)
		Hauberg	0.117 (0.073)	0.408 (0.149)	0.298 (0.038)	0.370 (0.205)	0.74 (0.208)	0.494 (0.063)
	Wave	Proposed (intrinsic)	0.71 (0.114)	0.329 (0.097)	0.084 (0.023)	0.673 (0.150)	0.346 (0.113)	0.124 (0.038)
		Hauberg	2.444 (0.059)	2.158 (0.155)	0.568 (0.055)	2.544 (0.118)	2.103 (0.563)	0.796 (0.094)
$T = 500$	Circle	Proposed (intrinsic)	0.088 (0.026)	0.118 (0.023)	0.091 (0.018)	0.21 (0.050)	0.207 (0.043)	0.129 (0.018)
		Hauberg	0.089 (0.027)	0.205 (0.079)	0.269 (0.034)	0.233 (0.087)	0.453 (0.177)	0.397 (0.079)
	Wave	Proposed (intrinsic)	0.535 (0.065)	0.239 (0.056)	0.072 (0.020)	0.574 (0.094)	0.237 (0.082)	0.110 (0.031)
		Hauberg	2.006 (0.697)	1.831 (0.146)	0.529 (0.043)	1.906 (0.847)	1.756 (0.696)	0.688 (0.073)

Table 3 lists the average values of reconstruction errors and their standard deviations over 50 simulation sets. As listed, the proposed principal curves outperform the Hauberg’s method, recovering the true curves accurately. Table 4 provides the average values of distinct projection points and their standard deviations. The proposed method provides a very large number of distinct projection points compared to that of Hauberg’s method. Overall, as listed in Table 3 and 4, our methods perform better than Hauberg’s method, including the case that the number of points of the curves ( $T = 500$ ) is much larger than the number of data points ( $n = 100$ ). In addition, the results of the intrinsic and extrinsic principal curves are similar in terms of both reconstruction error and the number of distinct projection points, which appear with the fact that the intrinsic and extrinsic means are almost identical for localized data, as noted in [21]. The results of the extrinsic approach are almost identical to those of the intrinsic one, and hence are omitted.

#### 4.2.2 Simulation on Hypersphere

We conduct a simulation study on a hypersphere. To this end, we consider a waveform simulated data on  $S^4$  represented by four angular parameters  $\varphi_1, \varphi_2, \varphi_3 \in [0, \pi]$ , and  $\varphi_4 \in [0, 2\pi]$ . The explicit representation on  $S^d$ ,  $d \geq 3$  is given in Appendix B.3 of the Supplementary Material. Mimicking a waveform dataset on  $S^2$  in Section 4.2.1, we craft simulation sets ( $r = 1, \varphi_1, \varphi_2, \varphi_3, \varphi_4$ ) with  $\varphi_1 = \varphi_2 = \varphi_3 = \alpha \sin(\varphi_4 f) + \pi/2$  and  $0 \leq \varphi_4 < 2\pi$ , frequency  $f = 2$ , and amplitude  $\alpha = 1/2$ . Data points of  $n = 200$  are generated by sampling  $\varphi_4$  uniformly in  $[0, 2\pi]$  and adding the random noises sampled from  $N(0, \sigma^2)$  to  $\varphi_1$  with  $\sigma = 0.05$ . Table 5 lists the average values of reconstruction errors defined on Section 4.2.1 and their

standard deviations over 50 simulation sets for each method with  $T = 300$ . As listed, the proposed principal curves outperform Hauberg’s method, recovering the true curves more closely.

## 5 CONCLUDING REMARKS

In this paper, new principal curves are proposed for data on spheres. The extrinsic and intrinsic perspectives are considered, and the stationarity of the principal curves is investigated, supporting that the proposed methods are a direct generalization of the principal curves by [2] to spheres.

For the data on  $S^d$ , both extrinsic and intrinsic approaches yield similar performance. However, it is questionable whether the extrinsic approach of non-isotropic manifolds, like a torus, will still be valid. For some non-isotropic manifolds, the intrinsic approach may yield better performance because of its inherency. Finally, the principal curve algorithm proposed in this study is a top-down approach. It approximates the structure of data with an initial curve and then gradually improves the estimation. However, for complex structures divided into several pieces or containing intersections, the initialization can significantly affect the final estimate. To cope with this limitation, it is worth studying a bottom-up approach. This approach to spheres is left for future research.

## ACKNOWLEDGMENTS

This research was supported by the National Research Foundation of Korea (NRF) funded by the Korea government (2018R1D1A1B07042933; 2020R1A4A1018207).

Table 4: Averages of distinct projection points and their standard deviations in the parentheses

	True form	Method	Noise level					
			$\sigma = 0.07$		$\sigma = 0.1$			
			$q = 0.05$	$q = 0.03$	$q = 0.01$	$q = 0.05$	$q = 0.03$	$q = 0.01$
$T = 100$	Circle	proposed (intrinsic) Hauberg	99.02 (0.32) 87.70 (7.95)	98.92 (0.34) 56.68 (17.99)	98.84 (0.47) 64.70 (3.22)	99.08 (0.34) 69.80 (12.28)	98.72 (1.11) 47.04 (15.44)	98.20 (1.12) 60.42 (2.83)
	Wave	proposed (intrinsic) Hauberg	93.36 (4.47) 22.72 (2.77)	97.28 (2.13) 25.94 (2.65)	99.32 (0.51) 62.14 (2.49)	95.82 (3.77) 24.5 (3.63)	96.72 (2.22) 32.16 (16.72)	99.10 (0.65) 58.84 (3.04)
$T = 500$	Circle	proposed (intrinsic) Hauberg	99.08 (0.27) 97.8 (1.47)	99.02 (0.25) 89 (8.63)	98.76 (0.69) 79.28 (4.60)	99.1 (0.30) 93.64 (5.29)	99.04 (0.49) 78.72 (13.08)	99.30 (1.09) 73.86 (7.28)
	Wave	proposed (intrinsic) Hauberg	99.18 (0.39) 45.2 (24.8)	98.5 (1.27) 43.38 (3.72)	99.26 (0.56) 73.20 (3.42)	99.22 (0.42) 52.04 (26.81)	98.84 (1.20) 50.64 (19.99)	99.18 (0.66) 71.52 (4.38)

Table 5: A simulation result of waveform data on  $S^4$ 

Method	$q = 0.03$	$q = 0.005$	$q = 0.002$
Proposed (extrinsic)	0.211 (0.230)	0.179 (0.162)	0.199 (0.235)
Proposed (intrinsic)	0.729 (0.493)	0.267 (0.264)	0.150 (0.232)
Hauberg	1.990 (0.815)	0.481 (0.215)	0.357 (0.251)

## REFERENCES

- [1] S. Hauberg, "Principal curves on riemannian manifolds," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 38, no. 9, p. 1915, 2016.
- [2] T. Hastie and W. Stuetzle, "Principal curves," *Journal of the American Statistical Association*, vol. 84, no. 406, pp. 502–516, 1989.
- [3] K. V. Mardia and R. J. Gadsden, "A small circle of best fit for spherical data and areas of vulcanism," *Journal of the Royal Statistical Society: Series C*, vol. 26, pp. 238–245, 1977.
- [4] K. V. Mardia, *Statistics of Directional Data*. Academic press, 2014.
- [5] N. H. Gray, P. A. Geiser, and J. R. Geiser, "On the least-squares fit of small and great circles to spherically projected orientation data," *Journal of the International Association for Mathematical Geology*, vol. 12, no. 3, pp. 173–184, 1980.
- [6] D. G. Kendall, "Shape manifolds, procrustean metrics, and complex projective spaces," *Bulletin of the London Mathematical Society*, vol. 16, no. 2, pp. 81–121, 1984.
- [7] S. Huckemann and H. Ziezold, "Principal component analysis for riemannian manifolds, with an application to triangular shape spaces," *Advances in Applied Probability*, vol. 38, no. 2, pp. 299–319, 2006.
- [8] S. Huckemann, T. Hotz, and A. Munk, "Intrinsic shape analysis: Geodesic pca for riemannian manifolds modulo isometric lie group actions," *Statistica Sinica*, pp. 1–58, 2010.
- [9] A. Mallasto and A. Feragen, "Wrapped gaussian process regression on riemannian manifolds," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2018, pp. 5580–5588.
- [10] K. Siddiqi and S. Pizer, *Medial Representations: Mathematics, Algorithms and Applications*. Springer Science & Business Media, 2008, vol. 37.
- [11] E. Cippitelli, S. Gasparrini, E. Gambi, and S. Spinsante, "A human activity recognition system using skeleton data from rgbd sensors," *Computational Intelligence and Neuroscience*, vol. 2016, 2016.
- [12] P. T. Fletcher, C. Lu, S. M. Pizer, and S. Joshi, "Principal geodesic analysis for the study of nonlinear statistics of shape," *IEEE Transactions on Medical Imaging*, vol. 23, no. 8, pp. 995–1005, 2004.
- [13] S. Jung, M. Foskey, and J. Marron, "Principal arc analysis on direct product manifolds," *The Annals of Applied Statistics*, vol. 5, no. 1, pp. 578–603, 2011.
- [14] S. Jung, I. L. Dryden, and J. Marron, "Analysis of principal nested spheres," *Biometrika*, vol. 99, no. 3, pp. 551–568, 2012.
- [15] V. M. Panaretos, T. Pham, and Z. Yao, "Principal flows," *Journal of the American Statistical Association*, vol. 109, no. 505, pp. 424–436, 2014.
- [16] A. Srivastava and E. Klassen, "Monte carlo extrinsic estimators of manifold-valued parameters," *IEEE Transactions on Signal Processing*, vol. 50, no. 2, pp. 299–308, 2002.
- [17] P. T. Fletcher and S. Joshi, "Riemannian geometry for the statistical analysis of diffusion tensor data," *Signal Processing*, vol. 87, no. 2, pp. 250–262, 2007.
- [18] W. M. Boothby, *An Introduction to Differentiable Manifolds and Riemannian Geometry*. Academic Press, 1986.
- [19] R. Bhattacharya, V. Patrangenaru *et al.*, "Large sample theory of intrinsic and extrinsic sample means on manifolds," *The Annals of Statistics*, vol. 31, no. 1, pp. 1–29, 2003.
- [20] R. N. Bhattacharya, L. Ellingson, X. Liu, V. Patrangenaru, and M. Crane, "Extrinsic analysis on manifolds is computationally faster than intrinsic analysis with applications to quality control by machine vision," *Applied Stochastic Models in Business and Industry*, vol. 28, no. 3, pp. 222–235, 2012.
- [21] R. Bhattacharya and V. Patrangenaru, "Large sample theory of intrinsic and extrinsic sample means on manifolds: ii," *The Annals of statistics*, pp. 1225–1259, 2005.
- [22] S. R. Buss and J. P. Fillmore, "Spherical averages and applications to spherical splines and interpolation," *ACM Transactions on Graphics (TOG)*, vol. 20, no. 2, pp. 95–126, 2001.
- [23] X. Pennec, "Intrinsic statistics on riemannian manifolds: Basic tools for geometric measurements," *Journal of Mathematical Imaging and Vision*, vol. 25, no. 1, p. 127, 2006.
- [24] G. Biau and A. Fischer, "Parameter selection for principal curves," *IEEE Transactions on Information Theory*, vol. 58, no. 3, pp. 1924–1939, 2011.
- [25] T. Duchamp, W. Stuetzle *et al.*, "Extremal properties of principal curves in the plane," *The Annals of Statistics*, vol. 24, no. 4, pp. 1511–1520, 1996.
- [26] B. Kégl, A. Krzyzak, T. Linder, and K. Zeger, "Learning and design of principal curves," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 22, no. 3, pp. 281–297, 2000.
- [27] C. Ionescu, F. Li, and C. Sminchisescu, "Latent structured models for human pose estimation," in *2011 International Conference on Computer Vision*. IEEE, 2011, pp. 2220–2227.
- [28] C. Ionescu, D. Papava, V. Olaru, and C. Sminchisescu, "Human3.6m: Large scale datasets and predictive methods for 3d human sensing in natural environments," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 36, no. 7, pp. 1325–1339, 2013.

**Jongmin Lee** received a M.S. degree in Statistics from Seoul National University, Seoul, Korea, in 2019, where he is currently pursuing a Ph.D. degree. His current research interests include non-Euclidean data analysis.

**Jang-Hyun Kim** received a B.S. degree in Mathematics from Seoul National University, Seoul, Korea, in 2019, where he is currently pursuing a M.S. degree in Computer Science. His current research interests include signal/image processing and machine learning.

**Hee-Seok Oh** received a Ph.D. degree in Statistics from Texas A&M University, College Station, TX, USA, in 1999. Since 2004, he has been with the Department of Statistics, Seoul National University, Seoul, Korea, as a Professor. His current research interests include multiscale methods in statistics, non-Euclidean data analysis, and statistical dimension reduction.

# Spherical Principal Curves

## — Supplementary Material —

Jongmin Lee\*, Jang-Hyun Kim\*, Hee-Seok Oh

### APPENDIX A JUSTIFICATION OF THE PROJECTION STEPS ON $S^d$

Let  $A = (a_1, a_2, \dots, a_{d+1})$ ,  $B = (b_1, b_2, \dots, b_{d+1})$ ,  $C = (c_1, c_2, \dots, c_{d+1}) \in S^d \subset \mathbb{R}^{d+1}$  with  $(A \cdot B)^2 \neq 1$ . Any point  $P$  on  $\widehat{AB}$  is denoted by  $P = \mu A + \lambda B$  for  $\mu, \lambda \in \mathbb{R}_+$  with  $\mu^2 + \lambda^2 = 1$ . If  $B \cdot C = C \cdot A = 0$ , then we have

$$d_{Geo}(C, P) = \arccos(C \cdot (\mu A + \lambda B)) = \pi/2.$$

Hence, any points on  $\widehat{AB}$  have the same geodesic distance of  $\pi/2$  from  $C$ . We may assume that  $A$ ,  $B$ , and  $C$  do not satisfy  $B \cdot C = C \cdot A = 0$ .

The orthogonal complement of  $V$  in  $\mathbb{R}^{d+1}$ ,  $V^\perp$ , has a dimension of  $d-1$ , owing to the fact that  $\mathbb{R}^{d+1} = V \oplus V^\perp$  with  $\oplus$  denoting the direct sum. As a column vector notation, we choose an orthonormal basis for  $V$  as  $R_1, R_2 \in \mathbb{R}^{d+1}$  and an orthonormal basis for  $V^\perp$  as  $R_3, R_4, \dots, R_{d+1} \in \mathbb{R}^{d+1}$ . Define a  $(d+1) \times (d+1)$  matrix  $R = [R_1, R_2, R_3, \dots, R_d, R_{d+1}]^T$ . Clearly,  $R$  is a rotation (orthogonal) matrix, i.e.  $R \in O(n) = \{X \in M_{d+1,d+1}(\mathbb{R}) \mid X^T X = I\}$  and satisfies that  $RA = (\tilde{a}_1, \tilde{a}_2, 0, 0, \dots, 0)$  and  $RB = (\tilde{b}_1, \tilde{b}_2, 0, 0, \dots, 0)$ . Let  $\tilde{A} = RA$ ,  $\tilde{B} = RB$ , and  $\tilde{C} = RC = (\tilde{c}_1, \tilde{c}_2, \dots, \tilde{c}_d, \tilde{c}_{d+1})$ . Let  $\tilde{V}$  be a two-dimensional vector space spanned by  $\tilde{A}$  and  $\tilde{B}$ , as shown in the right panel of Figure 1. It follows that  $\tilde{V} = \{x = (x_1, x_2, x_3, \dots, x_{d+1}) \mid x_3 = x_4 = \dots = x_{d+1} = 0\}$ . We denote the projection of  $\tilde{C}$  onto  $\tilde{V}$  as  $\tilde{C}' = (\tilde{c}_1, \tilde{c}_2, 0, \dots, 0) \in \mathbb{R}^{d+1}$  with  $\tilde{c}_1^2 + \tilde{c}_2^2 \neq 0$ . For any  $\tilde{P} = (\tilde{p}_1, \tilde{p}_2, 0, \dots, 0) \in \tilde{V} \cap S^d$ , it follows that

$$d_{Geo}(\tilde{C}, \tilde{P}) = \arccos(\tilde{c}_1 \tilde{p}_1 + \tilde{c}_2 \tilde{p}_2) \geq \arccos(\sqrt{\tilde{c}_1^2 + \tilde{c}_2^2}), \quad (1)$$

where the last inequality holds due to the Cauchy-Schwarz inequality  $(\tilde{c}_1 \tilde{p}_1 + \tilde{c}_2 \tilde{p}_2)^2 \leq (\tilde{c}_1^2 + \tilde{c}_2^2)(\tilde{p}_1^2 + \tilde{p}_2^2) = (\tilde{c}_1^2 + \tilde{c}_2^2)$ . The equality of (1) holds when  $(\tilde{p}_1, \tilde{p}_2) = t(\tilde{c}_1, \tilde{c}_2)$  for some  $t \in \mathbb{R}_+$ . It means that the closest point  $\tilde{P}$  on  $\tilde{V} \cap S^d$  from  $\tilde{C}$  is found by normalizing  $\tilde{C}'$  so that it is in  $S^d$ . Since  $R$  is an orthogonal matrix, for any  $P \in V \cap S^d$  and  $\tilde{P} = RP \in \tilde{V} \cap S^d$ , it follows, as a column vector notation, that

$$\begin{aligned} d_{Geo}(\tilde{C}, \tilde{P}) &= \arccos(\tilde{C}^T \tilde{P}) = \arccos(C^T R^T R P) \\ &= \arccos(C^T P) \\ &= d_{Geo}(C, P). \end{aligned}$$

Accordingly,  $\text{proj}(C)$  is obtained by applying  $R^{-1}$  to  $\text{proj}(\tilde{C})$  that is the projection of  $\tilde{C}$  onto  $\tilde{V} \cap S^d$ . Since the rotation is

a rigid motion, it completes the proof.

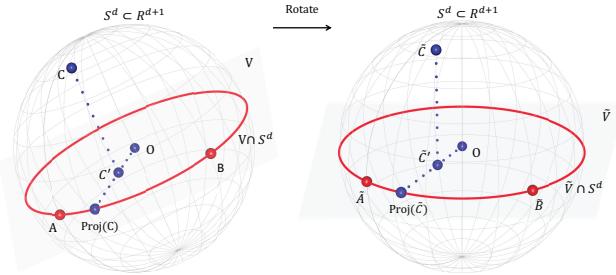


Figure 1: (Left) The projection process of  $C$  onto the one-dimensional great circle  $V \cap S^d$  (red) in a hypersphere  $S^d \subset \mathbb{R}^{d+1}$ : (i) find the projection of  $C$  onto  $V$ ,  $C'$  and (ii) obtain the projection of  $C'$  onto  $V \cap S^d$ ,  $\text{proj}(C)$ . (Right) The rotated configuration of the objects.

### APPENDIX B ENHANCEMENT OF PRINCIPAL CIRCLE FOR INITIALIZATION

Methods for fitting circles to data on  $S^2$  are actively used in many applications, especially in astronomy and geology, to recognize undisclosed patterns of data [1], [2]. This section improves the principal circle to use the initialization of the principal curves proposed in Section 3.

#### B.1 Principal Geodesic and Principal Circle

The principal curve algorithm of [3] uses the first principal component as the initial curve, which is easily calculated by singular value decomposition (SVD) of the data matrix in Euclidean space. Along with this line, the proposed principal curve algorithm in Section 4 requires an initial curve. The principal geodesic analysis (PGA) by [4] can be considered as a generalization of PCA that performs dimension reduction of data on the Cartesian product of simple manifolds, such as  $\mathbb{R}^3$ ,  $S^2$ , and  $\mathbb{R}_+$ . To this end, Fletcher et al. [4] projected each manifold component of the data into a tangent space at the intrinsic mean of each component. As a result of the tangent space approximation of each component, data are approximated by points in Euclidean space, so applying PCA allows dimension reduction to be performed through the inverse process of the tangent projection, i.e. exponential map that preserves a distance and angle at a

\* The first two authors contributed equally to this work.

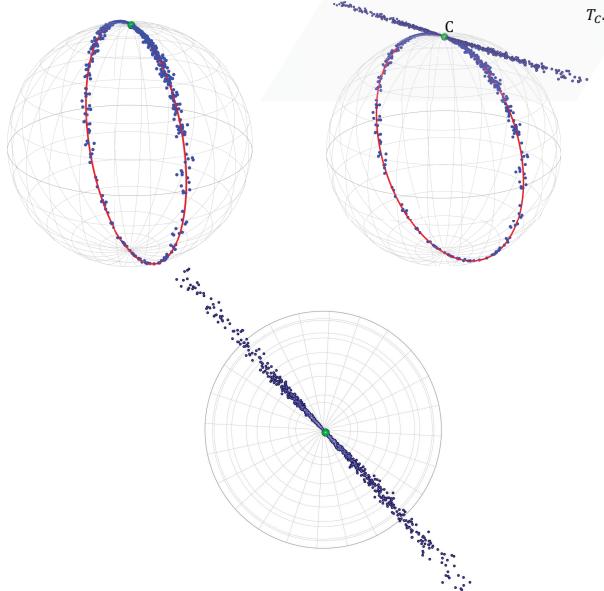
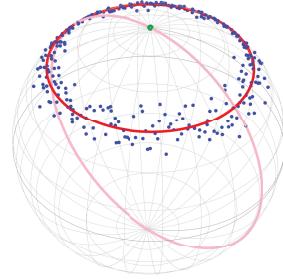
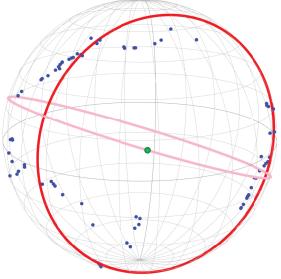


Figure 2: (Left) Spherical distribution of significant earthquakes (blue) with its intrinsic mean (green), the result (pink) by PGA, and the result (red) by our proposed principal circle. (Right) Circular simulated data (blue) with its intrinsic mean (green), the result (pink) by PGA, and the result (red) by our proposed principal circle.

base point. For spherical cases, they mainly perform tangent space projection using an inverse exponential map, called *log map*. The explicit forms of exponential and log maps of  $S^2$  are described in [4], [5], [6].

However, PGA always results in a great circle going through the intrinsic mean on the sphere, as shown in Figure 2, and the class of great circles on a sphere is sometimes limited to suitably fit a dataset on the sphere [5], [7]. For example, the left panel of Figure 2 shows earthquake data from the U.S. Geological Survey showing the location (blue dot) of significant earthquakes with Mb magnitude 8 or higher around the Pacific since 1900. The data will be analyzed in detail in Section 4. In Figure 2, while the result (pink) by PGA does not fit the data correctly, our principal circle (red) in Section B.2 of this Appendix improves the representation of the data. Further, in the right panel of Figure 2, our principal circle suitably fits the circular simulated data, whereas the result (pink) by PGA does not capture the variation of the data. The PGA’s failure stems from the fact that the above two data sets are far from their intrinsic means, as noted in [5], [6], [7].

In the literature, there is an attempt by [5] that generalizes the PGA to a circle on  $S^2$ . The circle on  $S^2$  that minimizes a reconstruction error is called *principal circle*, where the reconstruction error is defined as the total sum of squares of the geodesic distance between the curve and the data. Jung et al. [5] used a double iteration algorithm that uses the log map to project the data into the tangent space and then finds the principal circle. However, this approach has two problems. First, using the tangent approximation when minimizing the distance causes numerical errors. If the data points are located away from the mean, the numerical errors increase considerably because there is no local isometry between the sphere and its tangent plane according to the *Gauss’s Theorema Egregium* (see p. 363-370 of [8] or Ch 8 of [9] for details). For example, Figure 3 shows simulated data, where the underlying structure is a great circle, and the intrinsic mean is the North Pole  $(0, 0, 1)$ , where the data points are mostly concentrated around the North Pole. Second, due to the topological difference between the sphere and the plane, the existence of principal circles in the tangent plane is not guaranteed. From the compactness of the sphere, the least-squares circle always exists regardless

Figure 3: (Top left) Simulated data points (blue) with the intrinsic mean  $(0, 0, 1)$  (green), and the result by our proposed principal circle (red). (Top right) The projected points from the sphere onto the tangent plane at  $C = (0, 0, 1)$ . (Bottom) The projected points viewed from above the Northern Hemisphere.

of the data structure. It is an advantage of the intrinsic approaches. On the other hand, the least-squares circle does not exist if the data points projected onto the tangent space at their intrinsic mean are collinear, as shown in Figure 3. It coincides that several circle fitting procedures in a plane, such as [10] and [11], fail when the data points are collinear, as noted in [12]. Moreover in this case, the (tangent) plane cannot consider the periodicity of the data, as opposed to the sphere in Figure 3. Ignoring the periodic structure of data, as noted in [13], reduces the efficiency of a method.

This study proposes a new principal circle that does not rely on tangent projection for better initialization of the proposed principal curve presented in Section 3. We obtain the constraint-free optimization problem by expressing the center of the circle using the spherical coordinate system in Sections B.2 and B.3.

## B.2 Exact Principal Circle

For our principal circle, we consider an intrinsic optimization algorithm that does not use any approximations. Let  $d_{Geo}(x, y)$  be the geodesic distance between  $x, y \in S^2$ . For a given dataset  $D$  and a circle  $C$  on  $S^2$ , let  $\delta(D, C)$  be the sum of squares of distances between circle and data, defined as

$$\delta(D, C) = \sum_{x \in D} d_{Geo}(x, \text{proj}_C(x))^2,$$

where  $\text{proj}_C(x)$  denotes a projection of  $x$  on  $C$ . The goal is to find a circle  $C$  on  $S^2$  that minimizes  $\delta(D, C)$ . To solve this optimization problem, we represent a circle  $C$  by a center  $c$  of the circle and a radius  $r \in [0, \pi]$ , the geodesic distance between the center  $c$  and the circle  $C$ . This representation is not unique [5]. For example, let  $c' \in S^2$  be the antipodal

point of  $c$  that is diametrically opposite to  $c$  on  $S^2$ , then  $(c, r)$  and  $(c', \pi - r)$  represent the same circle  $C$ . Nevertheless, it is not crucial to the optimization problem because we simply find a representation of the least square circle. By using a spherical coordinate system, it is able to parameterize  $c$  as  $(\theta, \rho)$ , where  $\theta$  denotes the azimuthal angle and  $\rho$  is the polar angle. By symmetry of the circle,  $d_{Geo}(x, \text{proj}_C(x))$  can be easily calculated by

$$d_{Geo}(x, \text{proj}_C(x)) = d_{Geo}(x, c) - r.$$

Thus, we have

$$\delta(D, C) = \sum_{x \in D} (d_{Geo}(x, c) - r)^2. \quad (2)$$

With letting  $c = (\theta_c, \rho_c)$  and  $x = (\theta_x, \rho_x)$  in the spherical coordinate system, the geodesic distance  $d_{Geo}(x, c)$  is given by the spherical law of cosines with three points  $c, x$ , and the polar point (see Lemma 3 of Appendix C below for details)

$$d_{Geo}(x, c) = \arccos(\cos \rho_c \cos \rho_x + \sin \rho_c \sin \rho_x \cos(\theta_c - \theta_x)). \quad (3)$$

By putting (3) into (2), it follows that  $\delta(D, C)$  is represented as a three-parameter differentiable function  $\delta_D(\theta_c, \rho_c, r)$  in domain  $[0, 2\pi] \times [0, \pi] \times [0, \pi]$  as follows,

$$\delta_D(\theta_c, \rho_c, r) = \sum_{x \in D} (\arccos(\cos \rho_c \cos \rho_x + \sin \rho_c \sin \rho_x \cos(\theta_c - \theta_x)) - r)^2. \quad (4)$$

Since  $[0, 2\pi] \times [0, \pi] \times [0, \pi]$  is compact, the function  $\delta_D(\theta_c, \rho_c, r)$  holds a global minimum value. Thus, it can apply the gradient descent method to find the solution. Here is the algorithm to find a principal circle from the above description.

---

#### Algorithm 1 Exact Principal Circle by gradient descent

---

```

Initialize  $(\theta_c, \rho_c, r)$  as  $(\bar{\theta}, \bar{\rho}, \pi/2)$ 
while  $(\Delta\delta(D, C) \geq \text{threshold})$  do
     $(\theta_c, \rho_c, r) \leftarrow (\theta_c, \rho_c, r) - \beta \nabla \delta_D(\theta_c, \rho_c, r)$ 
end while

```

---

As in many nonlinear least-square algorithms, such as Gauss-Newton algorithm and Levenberg-Marquardt algorithm (see Ch 4 of [14] for details), the above Algorithm 1 may converge to a local minimum or a saddle point instead of the global minimum, since  $\delta_D(\theta_c, \rho_c, r)$  is non-convex. Thus, initial values should be selected carefully. If the data points in  $D$  are not too apart and localized, then it is reasonable to choose  $(\theta_x, \rho_x, \pi/2)$  for some  $x \in D$  as an initial. The spherical coordinates of the intrinsic mean of  $D$  with radius  $r = \pi/2$ , denoted by  $(\bar{\theta}, \bar{\rho}, \pi/2)$ , if necessary with varying  $r \in [0, \pi]$ , is also recommended as initial values. In the case of a non-localized data set, one can implement the algorithm with various initial settings as much as one wants, compare the consequences of  $\delta$ , and finally choose the circle with the lowest  $\delta$  as the principal circle. Note that, in existing methods for fitting circles to data on  $S^2$ , such as [2] and [5], there are no assurances that their algorithms finally achieve the circle minimizing (2). Although  $\delta$  is not convex globally, it is convex on a neighborhood of a global minimum point. Hence, it is reasonably expected that if an

initial value is suitably close to an optimum point, then Algorithm 1 converges to the optimum. A specification about the neighborhood for which  $\delta$  is convex, and rigorous proof for convergence of Algorithm 1 on that neighborhood remains a challenge. In the real data analysis and the simulated studies on Section 4, however, implementations of Algorithm 1 with several initial values result in almost the same principal circles and converge rapidly. Thus, there are no practical difficulties in our experiments. In addition,  $\beta$  is the step size of Algorithm 1, and it relies on data set  $D$ . The algorithm may diverge when  $\beta$  is large (e.g., greater than 0.01). In simulated examples and real data on Section 4, we use .001. Since too small  $\beta$  causes computational time to be high, an appropriate  $\beta$  should be selected properly throughout experiments from a relatively larger value of  $\beta$  to the lower one.

### B.3 Extension to Hyperspheres

In the case of high-dimensional spheres, to find a one-dimensional circle that attempts to represent a given data closely, we provide both extrinsic and intrinsic ways. The former is easy to implement and more computationally feasible because it uses an extrinsic approach and is not exactly found. The latter directly extends the exact principal circle in the previous section into higher-dimensional spheres using the framework of *principal nested spheres* [6]; however, it takes time to compute compared to the former approach.

#### B.3.1 Circle as an Initialization

In Section 4.2.2, we have used the following extrinsic method as an initial estimate of the spherical principal curves for waveform simulated data on  $S^4$ . Specifically, we consider  $S^d = \{y = (y_1, y_2, \dots, y_{d+1}) \in \mathbb{R}^{d+1} \mid \sum_{i=1}^{d+1} y_i^2 = 1\}$  for  $d \geq 2$ , as an embedded surface in the ambient space  $\mathbb{R}^{d+1}$ . That is,  $\{x_i\}_{i=1}^n \subset S^d \hookrightarrow \mathbb{R}^{d+1}$  are regarded as elements in  $\mathbb{R}^{d+1}$ , not taking into account a nonlinear dependence of the data; though, ensuring lower computational complexity. Note that any one-dimensional circle on  $S^d$  is an intersection of a two-dimensional plane and  $S^d$ . Hence, the strategy is to find the 2-plane  $P \subset \mathbb{R}^{d+1}$  that closely represents the data  $\{x_i\}_{i=1}^n$  with respect to the standard distance in  $\mathbb{R}^{d+1}$ , rather than geodesic distance in  $S^d$ . That is, the plane  $P$  is the two-dimensional vector subspace of  $\mathbb{R}^{d+1}$  spanned by first two principal components of the data, and then  $P \cap S^d$  is a one-dimensional circle to find. Although the extrinsic circle is capable of approximating the meaningful data, there may be some instances that need more precise initial estimate for the data.

#### B.3.2 Exact Principal Circle

For a better initial guess of the proposed principal curves, we provide an exact principal circle on  $S^d = \{y = (y_1, y_2, \dots, y_{d+1}) \in \mathbb{R}^{d+1} \mid \sum_{i=1}^{d+1} y_i^2 = 1\}$  for  $d \geq 2$ . The arguments in the previous section can be applied to higher-dimensional spheres  $S^d$  for  $d \geq 2$  if the geodesic distance of Equation (2) can be precisely calculated. To this end, let  $D = \{x_i\}_{i=1}^n$  be a dataset on  $S^d$ , and denote a  $(d-1)$ -dimensional subsphere on  $S^d$  as  $C$ . Using a spherical co-

ordinates for  $S^d$ ,  $x = (x_1, x_2, \dots, x_d, x_{d+1}) \in S^d \subset \mathbb{R}^{d+1}$  can be parametrized as

$$\begin{aligned} x_1 &= \cos(\varphi_1) \\ x_2 &= \sin(\varphi_1) \cos(\varphi_2), \\ x_3 &= \sin(\varphi_1) \sin(\varphi_2) \cos(\varphi_3) \\ &\vdots \\ x_d &= \sin(\varphi_1) \cdots \sin(\varphi_{d-1}) \cdot \cos(\varphi_d) \\ x_{d+1} &= \sin(\varphi_1) \cdots \sin(\varphi_{d-1}) \cdot \sin(\varphi_d), \end{aligned}$$

where  $\varphi_1, \varphi_2, \dots, \varphi_{d-1}, \varphi_d$  are angular coordinates with  $\varphi_d \in [0, 2\pi)$  and the others ranging over  $[0, \pi]$ . Note that  $d_{Geo}(x, c) = \arccos(x \cdot c)$ , where  $\cdot$  denotes the inner product in  $\mathbb{R}^{d+1}$ . Thus,

$$\begin{aligned} d_{Geo}(x, c) &= \arccos(\cos(\varphi_{1c}) \cos(\varphi_{1x}) + \\ &\sum_{k=1}^{d-2} [\prod_{i=1}^k \sin(\varphi_{ic}) \sin(\varphi_{ix})] \cdot \cos(\varphi_{(k+1)c}) \cos(\varphi_{(k+1)x}) \\ &+ [\prod_{i=1}^{d-1} \sin(\varphi_{dc}) \sin(\varphi_{dx})] \cdot \cos(\varphi_{dc} - \varphi_{dx})), \end{aligned} \quad (5)$$

where  $\{\varphi_{ic}\}_{i=1}^d$  and  $\{\varphi_{ix}\}_{i=1}^d$  are the corresponding angular coordinates of  $c$  and  $x$ , respectively. By putting (5) into (2), it follows that  $\delta(D, C)$  is represented as a  $(n+1)$ -parameter differentiable function  $\delta_D(\varphi_{1c}, \dots, \varphi_{dc}, r)$  in domain  $[0, \pi]^{d-1} \times [0, 2\pi] \times [0, \pi]$  as follows,

$$\begin{aligned} \delta_D(\varphi_{1c}, \dots, \varphi_{dc}, r) &= \sum_{x \in D} \left( \arccos(\cos(\varphi_{1c}) \cos(\varphi_{1x}) \right. \\ &+ \sum_{k=1}^{d-2} [\prod_{i=1}^k \sin(\varphi_{ic}) \sin(\varphi_{ix})] \cdot \cos(\varphi_{(k+1)c}) \cos(\varphi_{(k+1)x}) \\ &\left. + [\prod_{i=1}^{d-1} \sin(\varphi_{ic}) \sin(\varphi_{ix})] \cdot \cos(\varphi_{dc} - \varphi_{dx}) \right)^2 - r \right). \end{aligned} \quad (6)$$

Note that, in the case of  $d = 2$ , the above equation (6) becomes (4).  $\delta_D$  holds a global minimum value due to the compactness of the domain  $[0, \pi]^{d-1} \times [0, 2\pi] \times [0, \pi]$ . Therefore, an exact principal circle on  $S^d$  can be obtained by gradient descent, the same way in Algorithm 1, except that the number of parameters is  $d+1$ . Let  $(\bar{\varphi}_1, \bar{\varphi}_2, \dots, \bar{\varphi}_d)$  denote the spherical coordinates of the intrinsic mean of  $D$ . Here is the algorithm to find a principal circle on  $S^d$ .

---

**Algorithm 2** Exact Principal Nested Sphere on Hyper-sphere  $S^d$ 


---

```

Initialize  $(\varphi_{1c}, \varphi_{2c}, \dots, \varphi_{dc}, r)$  as  $(\bar{\varphi}_1, \bar{\varphi}_2, \dots, \bar{\varphi}_d, \pi/2)$ 
while  $(\Delta\delta(D, C) \geq \text{threshold})$  do
     $(\varphi_{1c}, \varphi_{2c}, \dots, \varphi_{dc}, r) \leftarrow (\varphi_{1c}, \varphi_{2c}, \dots, \varphi_{dc}, r)$ 
     $\quad -\beta \nabla \delta_D(\varphi_{1c}, \varphi_{2c}, \dots, \varphi_{dc}, r)$ 
end while

```

---

It is possible that Algorithm 2 converges to a local minimum or a saddle point of  $\delta_D$ , owing to its non-convexity. Therefore, an initial value should be carefully chosen, for instance, a data point in  $D$  and the intrinsic mean of  $D$ . The discussions about initial values and step size  $\beta$  are the same as those of Algorithm 1.

By applying the Algorithm 2 to a given data iteratively, we can obtain a one-dimensional sphere, *i.e.*, an exact principal circle on  $S^d$  that can be the initialization of the spherical principal curves. For more details about the procedure, see [6]. It is noteworthy that from the perspective of the principal nested spheres, our method can be applied to find nested spheres in an *exact* way.

Before closing this section, some useful non-convex optimization techniques, described in [15] and [16], have been recently developed to deal with the non-convex objective functions. The optimization techniques such as stochastic gradient descent, alternating minimization, and their variations in [15] can be applied to our problem. Specific applications of these approaches and related theoretical investigations remain as future work.

## APPENDIX C PROOF: STATIONARITY OF PRINCIPAL CURVES

Here we cover a smooth (infinitely differentiable) curve that does not cross on a sphere (*i.e.*,  $\lambda_1 \neq \lambda_2 \in [0, 1] \Rightarrow f(\lambda_1) \neq f(\lambda_2)$ ), including curves with end points and closed curves, which can be both parameterized over interval  $[0, 1]$  by a constant speed, *i.e.*  $f'(\lambda) = s > 0$  for any  $\lambda \in [0, 1]$ . In the latter case, a boundary condition is needed; any order partial derivatives of  $f$  at end points are the same, *i.e.*,  $f^{(k)}(0) = f^{(k)}(1)$  for all  $k \geq 0$ . For a random vector  $X$  on a sphere, we further assume that the curve  $f$  are not short enough to cover the support of  $X$  well, *i.e.*,  $\lambda_f(X) \neq 0, 1$  for *a.e.*  $X$ . For example, any closed curve satisfies the condition  $\lambda_f(X) \neq 0, 1$  for *a.e.*  $X$ , meaning that almost all  $X$  is orthogonally projected onto the curve  $f$ . Note that  $f$  is smooth on  $[0, 1]$ , *i.e.*  $f$  is smoothly extended on  $[0, 1]$ ; thus any order its derivatives are continuous on  $[0, 1]$ . Our main purpose is to prove the stationarity of extrinsic, intrinsic principal curves  $f : [0, 1] \rightarrow S^d$  for  $d \geq 2$  that satisfy the equations (8) and (9) in Theorems 1 and 2. We first consider the 2-sphere, and then extend  $d$ -spheres,  $d \geq 2$ .

When moving from Euclidean space to spherical surfaces, topological properties such as measurability and continuity are preserved, while the formula using specific distance should be modified. This modification could be obtained by embedding a spherical surface  $S^d$  into a  $(d+1)$ -dimensional Euclidean space. Specifically, we embed a spherical surface as a unit sphere centered at the origin, *i.e.*  $S^d \hookrightarrow \mathbb{R}^{d+1}$ , and investigate further derivations. When  $d = 2$ , for a smooth curve  $f : [0, 1] \rightarrow S^2$ , suppose that  $f$  is parameterized by a constant speed with respect to  $\lambda$  and is expressed as three-dimensional coordinates  $(f(\lambda)_1, f(\lambda)_2, f(\lambda)_3)$ . Then the following lemmas are held.

**Lemma 1.** A curve  $f$  satisfies  $f'(\lambda) \cdot f(\lambda) = 0$  and  $f''(\lambda) \cdot f'(\lambda) = 0$ ,  $\forall \lambda \in [0, 1]$ , where  $\cdot$  denotes inner product in  $\mathbb{R}^3$ .

*Proof.* It is directly obtained from  $f(\lambda) \cdot f(\lambda) = 1$  and  $f'(\lambda) \cdot f'(\lambda) = \text{constant}$ .  $\square$

**Lemma 2.** Suppose that  $f(\lambda)$  and  $x$  are expressed as three-dimensional vectors. Then, it follows that  $d_{Geo}(f(\lambda), x) = \arccos(f(\lambda) \cdot x)$ , where  $\arccos(f(\lambda) \cdot x)$  is the angle between  $f(\lambda)$  and  $x$ . Then,  $\frac{df}{d\lambda}(\lambda_f(0, 0, 1))_3 = 0$ . Thus, it follows that  $\frac{df}{d\lambda}(f(\lambda_f(0, 0, 1))) = a(-f(\lambda_f(0, 0, 1))_2, f(\lambda_f(0, 0, 1))_1, 0)$  for

some  $a \in \mathbb{R}$ . Note that  $\lambda_f(x)$  denotes the projection index of point  $x$  to the curve  $f$ .

*Proof.* For  $p = (0, 0, 1)$ , it follows that  $d_{Geo}(f(\lambda), p) = \arccos(f(\lambda)_3)$ . From the assumption that  $f(\lambda)$  is a smooth curve and the fact that  $d_{Geo}(f(\lambda), p)$  has the minimum at  $\lambda_f(0, 0, 1)$ , the remaining part of the lemma follows by differentiation with respect to  $\lambda$ .  $\square$

**Lemma 3.** (Spherical law of cosines) Let  $u, v, w$  be points on a sphere, and  $a, b$  and  $c$  denote  $d_{Geo}(w, u)$ ,  $d_{Geo}(w, v)$  and  $d_{Geo}(u, v)$ , respectively. If  $C$  is the angle between  $a$  and  $b$ , i.e., the angle of the corner opposite  $c$ , then,  $\cos c = \cos a \cos b + \sin a \sin b \cos C$ . Further, with three-dimensional vectors  $u, v, w$ , it follows that  $\sin a \sin b \cos C = (w \cdot u) \cdot (w \cdot v)$ , where  $\times$  denotes cross product in  $\mathbb{R}^3$ .

The following property can be obtained from Definition 2.

**Proposition 1.** Under the same conditions in Definition 2,  $f(\epsilon, \lambda) := (f + \epsilon g)(\lambda)$  is smooth on  $[-1, 1] \times [0, 1]$ . Hence, for each  $\epsilon \in [-1, 1]$ ,  $f + \epsilon g : [0, 1] \rightarrow S^d$  is a smooth curve on  $S^d$  and  $\lim_{\epsilon \rightarrow 0} (f + \epsilon g)(\lambda) = f(\lambda)$  for  $\lambda \in [0, 1]$ .

*Proof.* For simplicity, we denote  $f + g$  as  $h$ . Let  $R_{a,b}(\theta)$  be a rotation matrix that rotates points on  $S^d$  by  $\theta$  in the direction along the geodesic from  $a$  to  $b$  with  $a, b \in S^d \subset \mathbb{R}^{d+1}$  satisfying  $a \neq -b \in \mathbb{R}^{d+1}$  and  $\theta$  ranging over  $[0, \pi]$ . Then, it has a closed form; formally, as a column vector notation,  $R_{a,b}(\theta) = I_{d+1} + \sin(\theta)B + (\cos(\theta) - 1)(bb^T + cc^T)$ , where if  $a \neq b$ , then  $c = (a - b(b^T a)) / \|a - b(b^T a)\|$ , otherwise  $c = 0$ , and  $B = bc^T - cb^T$ . For more details, refer to the Section 8.1 in [6]. Hence, we obtain  $f(\epsilon, \lambda) := (f + \epsilon g)(\lambda) = R_{f(\lambda), h(\lambda)}(\theta)f(\lambda)$ , where  $\theta = \epsilon \arccos(f(\lambda) \cdot h(\lambda))$ . Thus,  $(f + \epsilon g)(\lambda) (= f(\epsilon, \lambda))$  is smooth on  $[-1, 1] \times [0, 1]$  since all functions  $f, h, R$ , and  $\theta$  are smooth. Therefore, for a fixed  $\epsilon \in [0, 1]$ , the smoothness of  $(f + \epsilon g)(\lambda)$  with respect to  $\lambda$  also follows. Moreover, the last equality is directly followed by the definition of  $f + \epsilon g$ .  $\square$

In Euclidean space, we have  $g(\lambda) = \frac{\partial}{\partial \epsilon} f_\epsilon(\lambda)$ , where  $f_\epsilon(\lambda) := f + \epsilon g$ . From this fact, the magnitude of perturbation  $\|h - f\|$  is defined as  $g(\epsilon_0, \lambda) := \frac{\partial}{\partial \epsilon}|_{\epsilon=\epsilon_0} f_\epsilon(\lambda)$ ,  $\|g(\lambda)\| := d_{Geo}(f(\lambda), g(\lambda)) = |g(\epsilon_0, \lambda)|$ ,  $\|g\| := \max_\lambda \|g(\lambda)\|$ , and finally  $\|h - f\| := \|g\| \neq \pi$ . The boundedness of  $\|g\|$  guarantees that  $\epsilon$ -internal division of the geodesic from  $f$  to  $h$  converges to  $f$  uniformly on  $\lambda \in [0, 1]$  as  $\epsilon$  goes to 0. Notice that from the compactness of the unit sphere,  $\|h - f\|$  is inherently equal or less than  $\pi$ ; thus, the assumption of  $\|h - f\| \neq \pi$  implies that  $\|h - f\| < \pi$ .

Moreover, the norm of derivative of perturbation  $\|(h - f)'\|$  is defined as  $g'(\epsilon, \lambda_0) := \frac{\partial}{\partial \lambda}|_{\lambda=\lambda_0} g(\epsilon, \lambda)$ ,  $\|g'(\lambda_0)\| := \max_\epsilon \|g'(\epsilon, \lambda_0)\|$ ,  $\|g'\| := \max_{\lambda_0} \|g'(\lambda_0)\|$ , and finally  $\|(h - f)'\| := \|g'\|$ .

Let  $x$  be a point on a sphere. By the continuity of  $f$  and the compactness of the domain set, it follows that  $\inf_{\lambda \in [0, 1]} d_{Geo}(x, f(\lambda))$  can be attained. Let  $d_{Geo}(x, f)$  denote the geodesic distance between  $x$  and  $f$ , i.e.,  $d_{Geo}(x, f) := \min_{\lambda \in [0, 1]} d_{Geo}(x, f(\lambda))$ . By the continuity of  $f$  again,  $\{\lambda \in [0, 1] \mid d_{Geo}(x, f(\lambda)) = d_{Geo}(x, f)\}$  is closed and therefore compact. Thus, the projection indices  $\lambda_f(x) = \inf\{\lambda \mid d_{Geo}(x, f(\lambda)) = d_{Geo}(x, f)\}$  and  $\lambda_{f+eg}$

are well defined. The latter holds due to the fact that  $f + \epsilon g$  is a continuous curve by Proposition 1. When  $\text{card}\{\lambda \mid d_{Geo}(x, f(\lambda)) = d_{Geo}(x, f)\} > 1$ , the point  $x$  is called an ambiguity point of  $f$ . The set of ambiguity points of the smooth curve has spherical measure 0; thus, the ambiguity points are negligible when calculating the expected value.

The next topological properties of the principal curves established in Euclidean space of [3] are still valid in spherical surfaces.

**Proposition 2.** (Measurability of index function) For a continuous curve  $f$  on  $S^d$ , the index function  $\lambda_f : S^d \rightarrow [0, 1]$  by  $x \mapsto \lambda_f(x)$  is measurable.

*Proof.* It follows that of Theorem 4.1 in [17]. It is enough to show that, for any constant  $c \in [0, 1]$ ,  $\{x \in S^d \mid \lambda_f(x) \geq c\}$  is a measurable set on  $S^d$ . If  $c = 0$ , then the set is  $S^d$ ; thus, we may assume that  $c \in (0, 1]$ . By the definition of the projection index  $\lambda_f(x) = \inf\{\lambda \mid d_{Geo}(x, f(\lambda)) = d_{Geo}(x, f)\}$ , the condition  $\lambda_f(x) \geq c$  is equivalent to the property that for any  $\lambda_1 \in [0, c]$ , there exists  $\lambda_2 \in [c, 1]$  such that  $d_{Geo}(x, f(\lambda_2)) < d_{Geo}(x, f(\lambda_1))$ . Technically, we aim to prove that

$\lambda_f(x) \geq c \stackrel{(1)}{\iff} \text{for any } \lambda_1 \in [0, c], \text{there exists } \lambda_2 \in [c, 1] \cap \mathbb{Q} \text{ such that } d_{Geo}(x, f(\lambda_2)) < d_{Geo}(x, f(\lambda_1)) \stackrel{(2)}{\iff} \text{for any } \lambda_1 \in [0, c] \cap \mathbb{Q}, \text{there exists } \lambda_2 \in [c, 1] \cap \mathbb{Q} \text{ such that } d_{Geo}(x, f(\lambda_2)) < d_{Geo}(x, f(\lambda_1))$ . Here  $\mathbb{Q}$  is the set of rational numbers. If (1) and (2) are verified, we obtain that

$$\begin{aligned} \{\lambda_f(x) \geq c\} &= \bigcap_{\lambda_1 \in [0, c] \cap \mathbb{Q}} \bigcup_{\lambda_2 \in [c, 1] \cap \mathbb{Q}} \{d_{Geo}(x, f(\lambda_2)) < d_{Geo}(x, f(\lambda_1))\}. \end{aligned}$$

Each set is measurable on  $S^d$  because for any  $\lambda_1$  and  $\lambda_2$ , the function  $x \mapsto d_{Geo}(x, f(\lambda_1)) - d_{Geo}(x, f(\lambda_2))$  is continuous. Accordingly,  $\{\lambda_f(x) \geq c\}$  is measurable due to the fact that countable unions and intersections of measurable sets are also measurable. It completes the proof.

*Proof of (1).* For any  $\lambda_1 \in [0, c]$ , there is  $\lambda_2 \in [c, 1]$  such that  $d_{Geo}(x, f(\lambda_2)) < d_{Geo}(x, f(\lambda_1))$ . Since  $\mathbb{Q}$  is dense in  $\mathbb{R}$  and  $f$  is continuous, there is  $\lambda'_2 \in [c, 1] \cap \mathbb{Q}$  such that  $d_{Geo}(x, f(\lambda'_2)) < d_{Geo}(x, f(\lambda_1))$ , which completes (1).

*Proof of (2).* We want to show that

$$\begin{aligned} D &:= \bigcap_{\lambda_1 \in [0, c]} \bigcup_{\lambda_2 \in [c, 1] \cap \mathbb{Q}} \{d_{Geo}(x, f(\lambda_2)) < d_{Geo}(x, f(\lambda_1))\} \\ &= \bigcap_{\lambda_1 \in [0, c] \cap \mathbb{Q}} \bigcup_{\lambda_2 \in [c, 1] \cap \mathbb{Q}} \{d_{Geo}(x, f(\lambda_2)) < d_{Geo}(x, f(\lambda_1))\} \\ &=: E. \end{aligned}$$

The inclusion  $D \subseteq E$  is clear. If  $x \in E$ , for any  $\lambda_1 \in [0, a] \cap \mathbb{Q}$ , there is  $\lambda_2 \in [a, 1] \cap \mathbb{Q}$  such that  $d_{Geo}(x, f(\lambda_2)) < d_{Geo}(x, f(\lambda_1))$ . For such  $\lambda_1$  and  $\lambda_2$ , owing to the continuity of  $f$ , there is  $\lambda_1 \in (\lambda_1, a) \cap \mathbb{Q}$  such that  $\forall \lambda \in [\lambda_1, \tilde{\lambda}_1] \cap \mathbb{Q} \Rightarrow d_{Geo}(x, f(\lambda_2)) < d_{Geo}(x, f(\lambda))$ . That is,

$$\begin{aligned} x \in &\bigcap_{\lambda \in [\lambda_1, \tilde{\lambda}_1] \cap \mathbb{Q}} \bigcup_{\lambda_3 \in [c, 1] \cap \mathbb{Q}} \{d_{Geo}(x, f(\lambda_3)) < d_{Geo}(x, f(\lambda))\} \\ &=: F_{\lambda_1, \tilde{\lambda}_1}. \end{aligned}$$

Note that  $\tilde{\lambda}_1$  is automatically chosen for each  $\lambda_1 \in [0, c] \cap \mathbb{Q}$ . Since the above derivation is satisfied for any  $\lambda_1 \in [0, c] \cap \mathbb{Q}$ , it

follows that

$$\begin{aligned} x &\in \bigcap_{\lambda_1 \in [0, c) \cap \mathbb{Q}, \lambda_1 < \tilde{\lambda}_1} F_{\lambda_1, \tilde{\lambda}_1} \\ &= \bigcap_{\lambda \in [0, c)} \bigcup_{\lambda_3 \in [c, 1] \cap \mathbb{Q}} \{d_{Geo}(x, f(\lambda_3)) < d_{Geo}(x, f(\lambda))\} \\ &= D. \end{aligned}$$

Hence, we obtain  $E \subseteq D$ ; thus,  $D = E$ .  $\square$

According to Proposition 2,  $\lambda_f(X)$  is a random variable with respect to  $X$  as long as  $X$  is a random vector on  $S^d$  for  $d \geq 2$ . Thus, a conditional expectation on  $\lambda_f(X)$  is feasible.

**Proposition 3.** (*Continuity of projection index under perturbation*) If  $x$  is not an ambiguity point for continuous curve  $f$ , then  $\lim_{\epsilon \rightarrow 0} \lambda_{f+\epsilon g}(x) = \lambda_f(x)$ .

*Proof.* The proof follows the line of the proof of Lemma 4.1 in [3]. It is enough to show that, for any small  $\eta > 0$  there exists  $\delta > 0$  such that  $|\epsilon| < \delta$  implies  $|\lambda_{f_\epsilon}(x) - \lambda_f(x)| < \eta$ . Define a set  $C := [0, 1] \cap (\lambda_f(x) - \eta, \lambda_f(x) + \eta)^c$  and  $d_C := \inf_{\lambda \in C} d_{Geo}(x, f(\lambda)) > d_{Geo}(x, f(\lambda_f(x)))$  where  $d_C$  is achieved by some  $\lambda \in C$  from the compactness of  $C$ , and the last inequality holds since  $x$  is not an ambiguity point of  $f$ . Choose  $\delta = \frac{1}{3} [d_C - d_{Geo}(x, f(\lambda_f(x)))] > 0$ . Then if  $|\epsilon| < \delta$ , it follows that

$$\begin{aligned} \inf_{\lambda \in C} d_{Geo}(x, f_\epsilon(\lambda)) - d_{Geo}(x, f_\epsilon(\lambda_f(x))) \\ \geq \inf_{\lambda \in C} d_{Geo}(x, f(\lambda)) - d_{Geo}(f(\lambda), f_\epsilon(\lambda)) \\ - d_{Geo}(x, f(\lambda_f(x))) - d_{Geo}(f(\lambda_f(x)), f_\epsilon(\lambda_f(x))) \\ \geq d_C - d_{Geo}(x, f(\lambda_f(x))) - \delta - \delta \\ = 3\delta - 2\delta > 0 \end{aligned}$$

By the definition of  $\lambda_{f_\epsilon}(x)$ , we obtain  $\lambda_{f_\epsilon}(x) \notin C$ ; thus,  $|\lambda_f(x) - \lambda_{f_\epsilon}(x)| < \eta$ . It completes the proof.  $\square$

In the proof of Proposition 3, it is possible to apply the triangle inequality on a sphere because the sphere is a metric space equipped with its geodesic distance. The following proposition is an important tool for verifying Theorem 2.

**Proposition 4.** (*Uniform continuity of projection index under perturbation*)  $\lim_{\epsilon \rightarrow 0} \lambda_{f+\epsilon g}(x) = \lambda_f(x)$  uniformly on the set of non-ambiguity points of  $f$ . That is, for every  $\eta > 0$ , there exists  $\delta > 0$  such that for any non-ambiguity points  $x$ , if  $|\epsilon| < \delta$ , then  $|\lambda_{f_\epsilon}(x) - \lambda_f(x)| < \eta$ .

For guaranteeing uniform continuity of projection index, it is required that  $|f''|$  is bounded. It is directly followed by smoothness of  $f$  and compactness of  $[0, 1]$ . A proof is similar to that of Proposition 3; thus, we omit the proof, which can be provided on request.

**Proposition 5.** *Spherical measure of the set of ambiguity points of smooth curve  $f$  is 0.*

Detailed steps for a proof of Proposition 5 are similar with those of [3]; thus, we omit the proof, which can be provided on request.

Meanwhile, to prove Theorem 2, it is essential to verify that  $\lambda_{f_\epsilon}$  is differentiable for  $\epsilon$  and its derivative is uniformly bounded. Thus, it is necessary to define a subset of  $S^2$  as  $B(\zeta) := \{x \in S^2 \mid |f''(\lambda_f(x)) \cdot x| > \zeta\}$  for  $\zeta \geq 0$ . Obviously,

as an inclusion of sets,  $\{B(\zeta)\}_{\zeta \geq 0}$  is decreasing as  $\zeta$  goes to 0. Moreover, the following lemma implies that, as  $\zeta$  goes to 0,  $B(\zeta)$  covers  $S^2$  almost everywhere.

**Lemma 4.** *The image of smooth function from  $[0, 1]$  to  $S^2$  has measure 0. Moreover,*

$$S^2 \setminus B(0) = \{x \in S^2 \mid |f''(\lambda_f(x)) \cdot x| = 0\}$$

is an union of images of two smooth functions from  $[0, 1]$  to  $S^2$ , which implies that  $S^2 \setminus B(0)$  are measure 0. Therefore, the measure of  $S^2 \setminus B(\zeta)$  goes to 0 as  $\zeta \rightarrow 0$ .

*Proof.* Suppose that  $I : [0, 1] \rightarrow S^2$  is smooth. The domain and range of  $I$  are the second countable (with usual topology) differentiable manifolds whose dimensions are 1 and 2, respectively. Since  $f$  is twice continuously differentiable function and the differential  $dI$  has rank 1 which is less than intrinsic dimension of  $S^2$ , by a generalization of Sard's Theorem, the image  $I([0, 1]) = \{I(x) \in S^2 \mid x \in [0, 1]\}$  has measure zero. Next, each point  $x \in S^2 \setminus B(0)$  satisfying  $\lambda_f(x) \neq 0, 1$  is characterized by two equations  $f'(\lambda) \cdot x = 0$  and  $f''(\lambda) \cdot x = 0$  for some  $\lambda \in [0, 1]$ . Therefore, we define functions  $I_1, I_2$  as follows: For all  $\lambda \in [0, 1]$ ,

$$\begin{aligned} I_1(\lambda) &= f'(\lambda) \times f''(\lambda) / \|f'(\lambda) \times f''(\lambda)\|, \\ I_2(\lambda) &= -f'(\lambda) \times f''(\lambda) / \|f'(\lambda) \times f''(\lambda)\|. \end{aligned}$$

It is well known that the curvature of a smooth curve lying on the unit sphere is more than 1. It implies that  $\kappa = \frac{|f'|}{s^2} \geq 1$ , where  $\kappa$  is the curvature of  $f$  and  $s := |f'(\lambda)| > 0$  for all  $\lambda \in [0, 1]$ , and hence  $f'' \neq 0$ . We have already known that  $f' \cdot f'' = 0$  by Lemma 1. Hence, it is obtained that  $f' \times f'' \neq 0$ . It implies that  $I_1$  and  $I_2$  are well defined and smooth. Therefore, we have  $S^2 \setminus B(0) = I_1([0, 1]) \cup I_2([0, 1])$ , which completes the proof.  $\square$

Lemma 4 means that the constraints of random vector  $X$  in Theorems 1 and 2 are almost negligible by setting  $\zeta$  infinitesimally small. Denote the set of ambiguity points of smooth curve  $f$  on a sphere as  $A$ , which is a measure zero set by Proposition 5.

**Lemma 5.** *Let  $A$  be the set of ambiguity points of smooth curve  $f$  on a sphere. Suppose that for any  $x \in S^2$ ,  $\lambda_f(x) \in (0, 1)$  and  $x \in A^c \cap B(\zeta)$  for some small  $\zeta > 0$ . Then  $\lambda(\epsilon) := \lambda_{f_\epsilon}(x)$  is a smooth function for  $\epsilon$  on an open interval containing 0. Moreover,  $\frac{\partial \lambda(\epsilon)}{\partial \epsilon}$  is uniformly bounded on  $A^c \cap B(\zeta)$ . That is, there are constants  $C > 0$  and  $\delta > 0$  such that if  $|\epsilon_0| < \delta$  and  $x \in A^c \cap B(\zeta)$ , then  $|\frac{\partial \lambda_{f_\epsilon}(x)}{\partial \epsilon}|_{\epsilon=\epsilon_0} < C$ .*

*Proof.* By the assumptions that  $x$  is a non-ambiguity point of  $f$  and  $\lambda_f(x) \neq 0, 1$  and Proposition 4, we obtain  $\lambda_{f+\epsilon g}(x) \neq 0, 1$  for sufficiently small values of  $|\epsilon|$ . Hence,  $\lambda(\epsilon)$  is characterized by orthogonality between  $f'_\epsilon(\lambda(\epsilon))$  and the geodesic through  $x$  and  $f_\epsilon(\lambda(\epsilon))$  on a small  $\epsilon$  near 0; that is,  $f'_\epsilon(\lambda) \cdot (x - f_\epsilon(\lambda)) = f'_\epsilon(\lambda) \cdot x = 0$  by the same argument in Lemma 1. Then, we define a map  $F : [-1, 1] \times [0, 1] \rightarrow \mathbb{R}$  as  $F(\epsilon, \lambda) = f'_\epsilon(\lambda) \cdot x$ .  $F$  is a smooth function by Proposition 1. It follows by the definition of  $B(\zeta)$  that

$$\frac{\partial}{\partial \lambda} F(\epsilon, \lambda) \Big|_{(0, \lambda_f)} = f''(\lambda_f) \cdot x \neq 0.$$

By implicit function theorem, for each  $x \in A^c \cap B(\zeta)$ ,  $\lambda(\epsilon) = \lambda_{f_\epsilon}(x)$  is a smooth function for  $\epsilon$  and  $F(\epsilon, \lambda(\epsilon)) = 0$

in an open interval containing zero. Next, in order to prove uniform boundedness of  $\frac{\partial \lambda(\epsilon)}{\partial \epsilon}$ , we should verify that  $f''_{\epsilon}(\lambda_{f_{\epsilon}}(x))$  uniformly converge to  $f''(\lambda_f(x))$  on  $A^c \cap B(\zeta)$  as  $\epsilon$  goes to 0. First of all, for each  $\lambda$ , we have

$$\begin{aligned} f_{\epsilon_0}(\lambda) &= f(\lambda) + \int_0^{\epsilon_0} g(\epsilon_0, \lambda) d\epsilon \\ \Rightarrow f''_{\epsilon_0}(\lambda) &= f''(\lambda) + \int_0^{\epsilon_0} g''(\epsilon, \lambda) d\epsilon \\ \Rightarrow \|f''_{\epsilon_0}(\lambda) - f''(\lambda)\| &\leq \int_0^{\epsilon_0} \|g''(\epsilon, \lambda)\| d\epsilon \leq \epsilon_0 M, \end{aligned} \quad (7)$$

for some  $M > 0$ . Note that the above derivatives are differentiation by  $\lambda$ . Also, the second equation holds since  $g(\epsilon, \cdot)$  is a twice continuously differentiable function for all  $\epsilon$ ; thus, it is able to change the order of derivative and the integration. The last inequality holds because  $g''(\epsilon, \lambda)( = \frac{\partial^2 g(\epsilon, \lambda)}{\partial \lambda^2})$  is continuous on  $[-1, 1] \times [0, 1]$ . Hence, it follows that

$$\begin{aligned} \|f''_{\epsilon}(\lambda_{f_{\epsilon}}(x)) - f''(\lambda_f(x))\| &\leq \|f''_{\epsilon}(\lambda_{f_{\epsilon}}(x)) - f''(\lambda_{f_{\epsilon}}(x))\| \\ &\quad + \|f''(\lambda_{f_{\epsilon}}(x)) - f''(\lambda_f(x))\| \\ &\rightarrow 0, \end{aligned} \quad (8)$$

as  $\epsilon \rightarrow 0$  uniformly on  $x \in A^c \cap B(\zeta)$ , because the first term uniformly converges to 0 by (7) and the last one also converges to 0 uniformly by Proposition 4 and the boundedness of  $f'''$ . We have that  $|x \cdot f''(\lambda_f(x))| > \zeta$  owing to  $x \in B(\zeta)$ , and from (8), there exists a constant  $\delta > 0$  such that  $|\epsilon| < \delta \Rightarrow |x \cdot f''(\lambda_{f_{\epsilon}}(x))| \geq \frac{\zeta}{2}$ . Since  $f_{\epsilon}(\lambda) = f(\epsilon, \lambda)$  has continuous second partial derivatives, it is able to change the order of partial derivatives by Schwarz's theorem, as

$$\begin{aligned} \frac{\partial}{\partial \epsilon} \Big|_{\epsilon=\epsilon_0} f'_{\epsilon}(\lambda(\epsilon_0)) &= \frac{\partial}{\partial \epsilon} \Big|_{\epsilon=\epsilon_0} \frac{\partial}{\partial \lambda} \Big|_{\lambda=\lambda(\epsilon_0)} f_{\epsilon}(\lambda) \\ &= \frac{\partial}{\partial \lambda} \Big|_{\lambda=\lambda(\epsilon_0)} \frac{\partial}{\partial \epsilon} \Big|_{\epsilon=\epsilon_0} f_{\epsilon}(\lambda) \\ &= g'(\epsilon_0, \lambda(\epsilon_0)), \end{aligned}$$

for all  $|\epsilon_0| < \delta$ . Therefore, if  $|\epsilon_0| < \delta$  by applying implicit function theorem to  $F$  again, then we obtain that  $\lambda(\epsilon)$  is differentiable at  $\epsilon = \epsilon_0$  and

$$\begin{aligned} |\lambda'(\epsilon_0)| &= \left| \frac{-\partial F(\epsilon, \lambda)/\partial \epsilon}{\partial F(\epsilon, \lambda)/\partial \lambda} \right|_{(\epsilon_0, \lambda(\epsilon_0))} = \frac{|x \cdot \frac{\partial}{\partial \epsilon}|_{\epsilon=\epsilon_0} f'_{\epsilon}(\lambda(\epsilon_0))|}{|x \cdot f'_{\epsilon_0}(\lambda(\epsilon_0))|} \\ &\leq \frac{\|g'\|}{\zeta/2} \leq \frac{2}{\zeta}, \end{aligned}$$

which completes the proof.  $\square$

We further consider principal curves on hypersphere  $S^d$  for  $d \geq 2$ . For a smooth curve  $f : [0, 1] \rightarrow S^d$ , suppose that  $f$  is parameterized by a constant speed with respect to  $\lambda$ . Lemma 1 and all of the Propositions are still valid on  $S^d$ . Moreover, Lemmas 4 and 5 can be extended onto  $S^d$  as follows.

**Lemma 6.** Define  $C(\zeta) = \{x \in S^d \mid |f''(\lambda_f(x)) \cdot x| > \zeta\}$  for  $d \geq 2$ . Then,

$$S^d \setminus C(0) = \{x \in S^d \mid |f''(\lambda_f(x)) \cdot x| = 0\}$$

has spherical ( $d$ -dimensional Hausdorff) measure zero. Hence, the

spherical measure of  $S^d \setminus C(\zeta)$  goes to 0 as  $\zeta \rightarrow 0$ .

**Lemma 7.** Let  $A$  be a set of ambiguity points of smooth curve  $f$  on  $S^d$  for  $d \geq 2$ . Suppose that  $x \in A^c \cap C(\zeta)$  for a small  $\zeta > 0$ , and  $\lambda_f(x) \in (0, 1)$ . Then  $\lambda(\epsilon) := \lambda_{f_{\epsilon}}(x)$  is a smooth function for  $\epsilon$  on an open interval containing zero.

Proofs of Lemmas 6 and 7 are similar to those of Lemmas 4 and 5, respectively. Thus, we omit the proofs.

## Proof of Theorem 1

*Proof.* First of all, we prove the theorem on  $S^2$ . If  $f = h$ , then nothing to prove. Thus, we assume that the curves  $f$  and  $f + g (= h)$  are not identical and further both are parameterized by  $\lambda \in [0, 1]$ . To prove the result, we need to show that the conditional expectation is zero after exchanging the order of the derivative and expectation.

First, for order exchange, it is necessary to show that the following random variable

$$\begin{aligned} Z_{\epsilon}(X) &= \frac{\cos(d_{Geo}(X, f + \epsilon g)) - \cos(d_{Geo}(X, f))}{\epsilon} \\ &= \frac{\cos(d_{Geo}(X, (f + \epsilon g)(\lambda_{f+\epsilon g}(X))))}{\epsilon} \\ &\quad - \frac{\cos(d_{Geo}(X, f(\lambda_f(X))))}{\epsilon} \end{aligned} \quad (9)$$

is uniformly bounded for any sufficiently small  $|\epsilon| > 0$ . Then we apply bounded convergence theorem. Since the projection index of  $X$  represents the closest point in the curve, it follows that

$$\begin{aligned} Z_{\epsilon}(X) &\leq \frac{\cos(d_{Geo}(X, (f + \epsilon g)(\lambda_{f+\epsilon g}(X))))}{\epsilon} \\ &\quad - \frac{\cos(d_{Geo}(X, f(\lambda_{f+\epsilon g}(X))))}{\epsilon}. \end{aligned} \quad (10)$$

For simplicity, let  $f_g(\lambda_{\epsilon}) := (f + g)(\lambda_{f+\epsilon g}(X))$ ,  $f_{\epsilon}(\lambda_{\epsilon}) := (f + \epsilon g)(\lambda_{f+\epsilon g}(X))$  and  $f(\lambda_{\epsilon}) := f(\lambda_{f+\epsilon g}(X))$ . By applying Lemma 3 to  $\cos(d_{Geo}(X, f_{\epsilon}(\lambda_{\epsilon})))$ , the inequality of (10) becomes

$$\begin{aligned} Z_{\epsilon}(X) &\leq \frac{\cos(d_{Geo}(X, f_{\epsilon}(\lambda_{\epsilon}))) - \cos(d_{Geo}(X, f(\lambda_{\epsilon})))}{\epsilon} \\ &= \frac{\cos(d_{Geo}(X, f(\lambda_{\epsilon}))) (\cos(d_{Geo}(f_{\epsilon}(\lambda_{\epsilon}), f(\lambda_{\epsilon}))) - 1)}{\epsilon} \\ &\quad + \frac{(f(\lambda_{\epsilon}) \times f_{\epsilon}(\lambda_{\epsilon})) \cdot (f(\lambda_{\epsilon}) \times X)}{\epsilon} \\ &= \frac{\cos(d_{Geo}(X, f(\lambda_{\epsilon}))) (\cos(\epsilon d_{Geo}(f_g(\lambda_{\epsilon}), f(\lambda_{\epsilon}))) - 1)}{\epsilon} \\ &\quad + \frac{A_{\epsilon}(f(\lambda_{\epsilon}) \times f_g(\lambda_{\epsilon})) \cdot (f(\lambda_{\epsilon}) \times X)}{\epsilon}, \end{aligned} \quad (11)$$

where

$$\begin{aligned} A_{\epsilon} &= |f(\lambda_{\epsilon}) \times f_{\epsilon}(\lambda_{\epsilon})| / |f(\lambda_{\epsilon}) \times f_g(\lambda_{\epsilon})| \\ &= \sin(\epsilon d_{Geo}(f(\lambda_{\epsilon}), f_g(\lambda_{\epsilon}))) / |f(\lambda_{\epsilon}) \times f_g(\lambda_{\epsilon})|. \end{aligned}$$

The last equality is done by Definition 1. To get the upper bound of  $Z_{\epsilon}(X)$ , we further use the following fact,  $|\frac{\sin \epsilon C}{\epsilon}| \leq |C|$  and  $|\frac{1-\cos \epsilon C}{\epsilon}| \leq \frac{|\epsilon| C^2}{2}$  for  $C \in \mathbb{R}$  and  $\epsilon \in \mathbb{R}$ . Then, we

have

$$\begin{aligned} Z_\epsilon(X) &\leq |\cos(d_{Geo}(X, f(\lambda_\epsilon)))| \frac{|\epsilon|B^2}{2} \\ &\quad + \frac{B}{|f(\lambda_\epsilon) \times f_g(\lambda_\epsilon)|} |(f(\lambda_\epsilon) \times f_g(\lambda_\epsilon)) \cdot (f(\lambda_\epsilon) \times X)|, \end{aligned}$$

where  $B = d_{Geo}(f(\lambda_\epsilon), f_g(\lambda_\epsilon)) \leq \|g\| < \pi$ . Note that any smallest geodesic distance on a unit sphere is smaller than  $\pi$ . In addition, we can assume that  $\epsilon$  is less than  $1/\pi$  because we are only interested in  $\epsilon$  near 0. Thus, we obtain the upper bound of  $Z_\epsilon(X)$  in (11)

$$Z_\epsilon(X) \leq \frac{\pi}{2} + \pi = \frac{3\pi}{2}.$$

A lower bound of  $Z_\epsilon(X)$  can be similarly obtained. Let  $f_g(\lambda) := (f+g)(\lambda_f(X))$ ,  $f_\epsilon(\lambda) := (f+\epsilon g)(\lambda_f(X))$  and  $f(\lambda) := f(\lambda_f(X))$ . By following the same path, we have

$$\begin{aligned} Z_\epsilon(X) &\geq \frac{\cos(d_{Geo}(X, (f+\epsilon g)(\lambda_f(X))))}{\epsilon} \\ &\quad - \frac{\cos(d_{Geo}(X, f(\lambda_f(X))))}{\epsilon} \\ &= \frac{\cos(d_{Geo}(X, f(\lambda))) (\cos(d_{Geo}(f_\epsilon(\lambda), f(\lambda))) - 1)}{\epsilon} \\ &\quad + \frac{(f(\lambda) \times f_\epsilon(\lambda)) \cdot (f(\lambda) \times X)}{\epsilon} \\ &= \frac{\cos(d_{Geo}(X, f(\lambda))) (\cos(ed_{Geo}(f_g(\lambda), f(\lambda))) - 1)}{\epsilon} \\ &\quad + \frac{B_\epsilon(f(\lambda) \times f_g(\lambda)) \cdot (f(\lambda) \times X)}{\epsilon}, \end{aligned} \quad (12)$$

where

$$\begin{aligned} B_\epsilon &= |f(\lambda) \times f_\epsilon(\lambda)| / |f(\lambda) \times f_g(\lambda)| \\ &= \sin(\epsilon d_{Geo}(f(\lambda), f_g(\lambda))) / |f(\lambda) \times f_g(\lambda)|. \end{aligned}$$

By the same way, it can be shown that  $Z_\epsilon(X) \geq -\frac{3\pi}{2}$ . Hence, we show that

$$|Z_\epsilon(X)| \leq \frac{3\pi}{2},$$

which is bounded for any  $0 \neq |\epsilon| \leq 1/\pi$ . Then, by the bounded convergence theorem, it follows that

$$\begin{aligned} \frac{\partial \mathbb{E}_X \cos(d_{Geo}(X, f+\epsilon g))}{\partial \epsilon} \Big|_{\epsilon=0} \\ = \mathbb{E}_X \frac{\partial \cos(d_{Geo}(X, f+\epsilon g))}{\partial \epsilon} \Big|_{\epsilon=0}. \end{aligned}$$

Thus, the proof is completed provided that the following equation holds

$$\mathbb{E} \left[ \frac{\partial \cos(d_{Geo}(X, f+\epsilon g))}{\partial \epsilon} \Big|_{\epsilon=0} \middle| \lambda_f(X) = \lambda \right] = 0, \text{ for a.e. } \lambda.$$

By the definition of derivative,

$$\frac{\partial \cos(d_{Geo}(X, f+\epsilon g))}{\partial \epsilon} \Big|_{\epsilon=0} = \lim_{\epsilon \rightarrow 0} Z_\epsilon(X),$$

and as shown,  $Z_\epsilon(X)$  is bounded. Since  $f$  and  $f+g$  are continuous, by Proposition 3, if  $X$  is not an ambiguity point of  $f$  and  $f+g$ , then

$$\lim_{\epsilon \rightarrow 0} f_g(\lambda_\epsilon) = f_g(\lambda), \quad \lim_{\epsilon \rightarrow 0} f(\lambda_\epsilon) = f(\lambda).$$

Next, to show the limit of  $Z_\epsilon$ , we use the fact that

$\lim_{\epsilon \rightarrow 0} \frac{\sin \epsilon C}{\epsilon} = C$  and  $\lim_{\epsilon \rightarrow 0} \frac{1-\cos \epsilon C}{\epsilon} = 0$  for  $C \in \mathbb{R}$  and  $\epsilon \in \mathbb{R}$ . When  $f_g(\lambda) \neq f(\lambda)$ , it follows that

$$\begin{aligned} \lim_{\epsilon \rightarrow 0} \text{RHS of (11)} \\ &= d_{Geo}(f(\lambda), f_g(\lambda)) \frac{f(\lambda) \times f_g(\lambda)}{|f(\lambda) \times f_g(\lambda)|} \cdot (f(\lambda) \times X) \\ &= \mu(\lambda) \cdot (f(\lambda) \times X), \end{aligned}$$

where  $\mu(\lambda) = d_{Geo}(f(\lambda), f_g(\lambda))(f(\lambda) \times f_g(\lambda)) / |f(\lambda) \times f_g(\lambda)|$  if  $f(\lambda) \neq (f+g)(\lambda)$  and  $\mu(\lambda) = 0$  otherwise. Similarly, we obtain that

$$\lim_{\epsilon \rightarrow 0} \text{RHS of (12)} = \mu(\lambda) \cdot (f(\lambda) \times X).$$

In summary, if  $X$  is not an ambiguity point of  $f$  and  $f+g$ , and  $f(\lambda_f(X)) \neq (f+g)(\lambda_f(X))$ , then we have

$$\frac{\partial \cos(d_{Geo}(X, f+\epsilon g))}{\partial \epsilon} \Big|_{\epsilon=0} = \mu(\lambda_f(X)) \cdot (f(\lambda_f(X)) \times X). \quad (13)$$

In the case of  $f(\lambda_f(X)) = (f+g)(\lambda_f(X))$ , the equation (13) also hold because its left and right hand side are 0. From Proposition 4, the limit of (13) is established for a.e.  $X$ . Note that, since  $X$  is a random vector and  $\lambda_f(X)$  is measurable with respect to  $X$  according to the Proposition 2,  $\lambda_f(X)$  is also a random variable depending on  $X$ . It implies that conditional expectation on  $\lambda_f(X)$  is feasible. Hence, the following equality holds

$$\begin{aligned} \mathbb{E}_X \left[ \frac{\partial \cos(d_{Geo}(X, f+\epsilon g))}{\partial \epsilon} \Big|_{\epsilon=0} \right] \\ = \mathbb{E}_X [\mu(\lambda_f(X)) \cdot (f(\lambda_f(X)) \times X)]. \end{aligned} \quad (14)$$

Finally, if  $f$  is an extrinsic principal curve, then

$$\mathbb{E}[X | \lambda_f(X) = \lambda] = cf(\lambda)$$

for  $\exists c \in \mathbb{R}$ . Hence, it follows that

$$\begin{aligned} \mathbb{E}[\mu(\lambda_f(X)) \cdot (f(\lambda_f(X)) \times X) | \lambda_f(X) = \lambda] \\ = \mathbb{E}[\mu(\lambda) \cdot (f(\lambda) \times X) | \lambda_f(X) = \lambda] \\ = \mu(\lambda) \cdot (f(\lambda) \times cf(\lambda)) = 0. \end{aligned}$$

Hence, we have

$$\begin{aligned} \text{LHS of (14)} &= \mathbb{E}_X [\mu(\lambda) \cdot (f(\lambda) \times X)] \\ &= \mathbb{E}_\lambda [\mathbb{E}[\mu(\lambda) \cdot (f(\lambda) \times X) | \lambda_f(X) = \lambda]] = 0. \end{aligned}$$

To prove the converse, we assume that

$$\begin{aligned} \mathbb{E}_\lambda [\mathbb{E}(\mu(\lambda) \cdot (f(\lambda) \times X) | \lambda_f(X) = \lambda)] \\ = \mathbb{E}_\lambda [\mu(\lambda) \cdot \mathbb{E}[f(\lambda) \times X | \lambda_f(X) = \lambda]] = 0, \end{aligned}$$

for all smooth  $f+g$  satisfying  $\|g\| \neq \pi$  and  $\|g'\| \leq 1$ . Since  $f+g$  is only concerned with  $\mu(\lambda)$ , it follows that

$$\begin{aligned} \mathbb{E}[f(\lambda) \times X | \lambda_f(X) = \lambda] &= f(\lambda) \times \mathbb{E}[X | \lambda_f(X) = \lambda] \\ &= 0, \text{ for a.e. } \lambda. \end{aligned}$$

Therefore, we have

$$\mathbb{E}[X | \lambda_f(X) = \lambda] = cf(\lambda)$$

for  $\exists c \geq 0$ , which completes the proof.

Next, we consider the hypersphere case  $S^d$  for  $d \geq 3$ .

For given smooth curves  $f$  and  $h (= f + g)$  parametrized by  $\lambda \in [0, 1]$ , if  $f = h$ , the result is obvious. Thus, we assume that  $f$  and  $f + g (= h)$  are not identical. Suppose that  $X \in A^c \cap B(\zeta)$  for a small  $\zeta > 0$  and  $\lambda_f(X) \in (0, 1)$  for a.e.  $X$ , where  $A$  denotes the set of ambiguity points of  $f$ . As the proof of the case of  $S^2$ , we use the bounded convergence theorem to change the order of derivative and expectation. Since  $d_{Geo}(x, y) = \arccos(x \cdot y)$ ,  $\forall x, y \in S^d \subset \mathbb{R}^{d+1}$ , we have

$$\begin{aligned} Z_\epsilon(X) &:= \frac{\cos(d_{Geo}(X, f + \epsilon g)) - \cos(d_{Geo}(X, f))}{\epsilon} \\ &= \frac{\cos(d_{Geo}(X, (f + \epsilon g)(\lambda_{f+\epsilon g}(X))))}{\epsilon} \\ &\quad - \frac{\cos(d_{Geo}(X, f(\lambda_f(X))))}{\epsilon} \\ &\leq \frac{\cos(d_{Geo}(X, (f + \epsilon g)(\lambda_{f+\epsilon g}(X))))}{\epsilon} \\ &\quad - \frac{\cos(d_{Geo}(X, f(\lambda_{f+\epsilon g}(X))))}{\epsilon} \\ &= \frac{X \cdot (f + \epsilon g)(\lambda_{f+\epsilon g}(X)) - X \cdot f(\lambda_{f+\epsilon g}(X))}{\epsilon} \\ &= X \cdot \frac{(f + \epsilon g)(\lambda_{f+\epsilon g}(X)) - f(\lambda_{f+\epsilon g}(X))}{\epsilon}, \end{aligned} \quad (15)$$

where  $\cdot$  denotes the standard inner product in  $\mathbb{R}^{d+1}$ . Hence, we obtain the upper bound of  $Z_\epsilon(X)$ ,

$$\begin{aligned} Z_\epsilon(X) &\leq \|X\| \frac{\|(f + \epsilon g)(\lambda_{f+\epsilon g}(X)) - f(\lambda_{f+\epsilon g}(X))\|}{\epsilon} \\ &\leq \frac{d_{Geo}((f + \epsilon g)(\lambda_{f+\epsilon g}(X)), f(\lambda_{f+\epsilon g}(X)))}{\epsilon} \\ &\leq \|g(\lambda_{f+\epsilon g}(X))\| \leq \|g\| \\ &\leq \pi, \end{aligned}$$

where  $\|\cdot\|$  denotes the standard norm in  $\mathbb{R}^{d+1}$ . Similarly, it follows that

$$\begin{aligned} Z_\epsilon(X) &\geq \frac{\cos(d_{Geo}(X, (f + \epsilon g)(\lambda_f(X))))}{\epsilon} \\ &\quad - \frac{\cos(d_{Geo}(X, f(\lambda_f(X))))}{\epsilon} \\ &= \frac{X \cdot (f + \epsilon g)(\lambda_f(X)) - X \cdot f(\lambda_f(X))}{\epsilon} \\ &= X \cdot \frac{(f + \epsilon g)(\lambda_f(X)) - f(\lambda_f(X))}{\epsilon} \\ &\geq -\|X\| \frac{\|(f + \epsilon g)(\lambda_f(X)) - f(\lambda_f(X))\|}{\epsilon} \\ &\geq -\frac{d_{Geo}((f + \epsilon g)(\lambda_f(X)), f(\lambda_f(X)))}{\epsilon} \\ &\geq -\|g(\lambda_f(X))\| \geq -\|g\| \\ &\geq -\pi. \end{aligned}$$

It means that  $Z_\epsilon(X)$  is uniformly bounded for  $0 \neq |\epsilon| \leq 1$ . Next, to find the limit of  $Z_\epsilon(X)$ , we have

$$\begin{aligned} Z_\epsilon(X) &= \frac{\cos(d_{Geo}(X, (f + \epsilon g)(\lambda_{f+\epsilon g}(X))))}{\epsilon} \\ &\quad - \frac{\cos(d_{Geo}(X, f(\lambda_f(X))))}{\epsilon} \\ &= \frac{X \cdot (f + \epsilon g)(\lambda_{f+\epsilon g}(X)) - X \cdot f(\lambda_f(X))}{\epsilon} \\ &= X \cdot \frac{(f + \epsilon g)(\lambda_{f+\epsilon g}(X)) - f(\lambda_f(X))}{\epsilon}. \end{aligned}$$

According to the Proposition 3,

$$\begin{aligned} \lim_{\epsilon \rightarrow 0} Z_\epsilon(X) &= X \cdot \lim_{\epsilon \rightarrow 0} \frac{(f + \epsilon g)(\lambda_{f+\epsilon g}(X)) - f(\lambda_f(X))}{\epsilon} \\ &=: X \cdot \phi(\lambda_f(X)). \end{aligned}$$

For each  $X \in A^c \cap B(\zeta)$ , define a curve  $C : I \rightarrow S^d$  by  $\epsilon \mapsto C(\epsilon) = (f + \epsilon g)(\lambda_{f+\epsilon g}(X)) \in S^d \subset \mathbb{R}^{d+1}$ , where  $I$  is an open interval containing zero and  $C(0) = f(\lambda_f(X))$ . For convenience, let  $(f + \epsilon g)(\lambda) = f(\epsilon, \lambda)$ ,  $\lambda_f(X) = \lambda(0)$  and  $\lambda_{f+\epsilon g}(X) = \lambda(\epsilon)$ . According to Lemma 7,  $\lambda(\epsilon)$  is a smooth function on an interval  $I$  containing zero. As  $f(\cdot, \cdot)$  is smooth on  $[-1, 1] \times [0, 1]$  by the Proposition 1 and  $\lambda(\epsilon)$  is smooth on  $\epsilon \in I$ ,  $C(\epsilon) = f(\epsilon, \lambda(\epsilon))$  is also smooth on  $\epsilon \in I$ . Thus,  $\phi(\lambda)$  is well defined. Hence, by the definition of tangent space via tangent curves, it follows that

$$\phi(\lambda) = \lim_{\epsilon \rightarrow 0} \frac{C(\epsilon) - C(0)}{\epsilon} = C'(0) \in T_{f(\lambda)} S^d,$$

where  $T_{f(\lambda)} S^d$  is the tangent space of  $S^d$  at  $f(\lambda)$ . Note that, by the symmetry of spheres, any tangent vector in  $T_{f(\lambda)} S^d$  is orthogonal to the vector  $f(\lambda)$ , i.e.,  $\phi(\lambda) \cdot f(\lambda) = 0$ . Finally, if  $f$  is an extrinsic principal curve, then

$$\mathbb{E}[X | \lambda_f(X) = \lambda] = c f(\lambda)$$

for  $\exists c \in \mathbb{R}$ . Hence, it follows, by the bounded convergence theorem, that

$$\begin{aligned} &\frac{\partial \mathbb{E}_X [\cos(d_{Geo}(X, f + \epsilon g))] }{\partial \epsilon} \Big|_{\epsilon=0} \\ &= \lim_{\epsilon \rightarrow 0} \frac{\mathbb{E}_X [\cos(d_{Geo}(X, f + \epsilon g))] - \mathbb{E}_X [\cos(d_{Geo}(X, f))] }{\epsilon} \\ &= \mathbb{E}_X \left[ \lim_{\epsilon \rightarrow 0} \frac{\cos(d_{Geo}(X, f + \epsilon g)) - \cos(d_{Geo}(X, f))}{\epsilon} \right] \\ &= \mathbb{E}_\lambda \left[ \mathbb{E} \left[ \lim_{\epsilon \rightarrow 0} Z_\epsilon(X) \mid \lambda_f(X) = \lambda \right] \right] \\ &= \mathbb{E}_\lambda \left[ \mathbb{E}[\phi(\lambda) \cdot X \mid \lambda_f(X) = \lambda] \right] \\ &= \mathbb{E}_\lambda [\phi(\lambda) \cdot \mathbb{E}[X \mid \lambda_f(X) = \lambda]] \\ &= \mathbb{E}_\lambda [\phi(\lambda) \cdot c f(\lambda)] \\ &= 0. \end{aligned}$$

To prove the converse, we assume that  $f$  satisfies

$$\begin{aligned} 0 &= \frac{\partial \mathbb{E}_X [\cos(d_{Geo}(X, f + \epsilon g))] }{\partial \epsilon} \Big|_{\epsilon=0} \\ &= \lim_{\epsilon \rightarrow 0} \frac{\mathbb{E}_X [\cos(d_{Geo}(X, f + \epsilon g))] - \mathbb{E}_X [\cos(d_{Geo}(X, f))] }{\epsilon} \\ &= \mathbb{E}_X \left[ \lim_{\epsilon \rightarrow 0} \frac{\cos(d_{Geo}(X, f + \epsilon g)) - \cos(d_{Geo}(X, f))}{\epsilon} \right] \\ &= \mathbb{E}_\lambda \left[ \mathbb{E} \left[ \lim_{\epsilon \rightarrow 0} Z_\epsilon(X) \mid \lambda_f(X) = \lambda \right] \right] \\ &= \mathbb{E}_\lambda \left[ \mathbb{E}[\phi(\lambda) \cdot X \mid \lambda_f(X) = \lambda] \right] \\ &= \mathbb{E}_\lambda [\phi(\lambda) \cdot \mathbb{E}[X \mid \lambda_f(X) = \lambda]], \end{aligned}$$

for any smooth curve  $h : [0, 1] \rightarrow S^d$ . Since  $h$  is arbitrary,  $\phi$  can become any vector in  $T_{f(\lambda)} S^d$ . In addition,  $h$  is only concerned with  $\phi$ . We thus obtain, for a.e.  $\lambda$ , the following condition:

$$\phi \cdot \mathbb{E}[X \mid \lambda_f(X) = \lambda] = 0 \text{ for any } \phi \in T_{f(\lambda)} S^d.$$

It means that  $\mathbb{E}[X \mid \lambda_f(X) = \lambda]$  is orthogonal to  $T_{f(\lambda)} S^d$ .

Therefore, it follows that

$$\mathbb{E}[X \mid \lambda_f(X) = \lambda] = cf(\lambda)$$

for  $\exists c \geq 0$ , which completes the proof.  $\square$

## Proof of Theorem 2

*Proof.* In the case of  $f = h$ , the result is obvious. We thus assume that  $f$  and  $f + g (= h)$  are not identical. Further, suppose that  $X \in A^c \cap B(\zeta)$  for a small  $\zeta > 0$  and  $\lambda_f(X) \in (0, 1)$  for a.e.  $X$ . As the proof of Theorem 1, we use the bounded convergence theorem to change the order of derivative and expectation. For this purpose, we define

$$\begin{aligned} Z_\epsilon(X) &= \frac{d_{Geo}^2(X, f + \epsilon g) - d_{Geo}^2(X, f)}{\epsilon} \\ &= \frac{d_{Geo}^2(X, f_\epsilon(\lambda_{f_\epsilon})) - d_{Geo}^2(X, f(\lambda_f))}{\epsilon}, \end{aligned}$$

where  $f_\epsilon := f + \epsilon g$  for  $|\epsilon| \leq 1$ . Let  $\theta(\lambda, X)$  be the angle between segments of geodesics from  $f(\lambda)$  to  $X$  and from  $f(\lambda)$  to  $(f + g)(\lambda)$ . Then, from Lemma 3, it follows that

$$\begin{aligned} F(\epsilon) &:= \cos(d_{Geo}(X, f_\epsilon(\lambda_{f_\epsilon}))) \\ &= \cos(d_{Geo}(X, f(\lambda_{f_\epsilon}))) \cdot \cos(\epsilon \|g(\lambda_{f_\epsilon})\|) \\ &\quad + \sin(d_{Geo}(X, f(\lambda_{f_\epsilon}))) \cdot \sin(\epsilon \|g(\lambda_{f_\epsilon})\|) \cdot \cos(\theta(\lambda_{f_\epsilon}, X)), \end{aligned}$$

where  $\|g(\lambda)\| = d_{Geo}(f(\lambda), (f + g)(\lambda)) < \pi$ .

Firstly, we verify that  $Z_\epsilon(X)$  is uniformly bounded for a small  $|\epsilon| > 0$ . By Lemma 5, there are constants  $C > 0$  and  $\eta > 0$  such that if  $0 < |\epsilon_0| < \eta$ , then  $\lambda(\epsilon)$  is differentiable at  $\epsilon = \epsilon_0$  and  $|\frac{\partial \lambda(\epsilon)}{\partial \epsilon}|_{\epsilon=\epsilon_0} < C$ , where  $\lambda(\epsilon) = \lambda_{f_\epsilon}(X)$ . For convenience, let  $\lambda_{f_\epsilon}(X) = \lambda_\epsilon$  and  $\lambda_f(X) = \lambda_0$ . If  $0 < |\epsilon_0| < \eta$ , then by the triangle inequality on sphere and mean value theorem, it follows that

$$\begin{aligned} |Z_{\epsilon_0}(X)| &= \left| \frac{d_{Geo}(X, f_{\epsilon_0}(\lambda_{f_{\epsilon_0}})) - d_{Geo}(X, f(\lambda_f))}{\epsilon_0} \right| \\ &\quad \cdot \left( d_{Geo}(X, f_{\epsilon_0}(\lambda_{f_{\epsilon_0}})) + d_{Geo}(X, f(\lambda_f)) \right) \\ &\leq 2\pi \cdot \frac{d_{Geo}(f(\lambda_0), f_{\epsilon_0}(\lambda_{\epsilon_0}))}{\epsilon_0} \\ &\leq 2\pi \cdot \left[ \frac{d_{Geo}(f(\lambda_0), f(\lambda_{\epsilon_0}))}{\epsilon} \right. \\ &\quad \left. + \frac{d_{Geo}(f(\lambda_{\epsilon_0}), f_{\epsilon_0}(\lambda_{\epsilon_0}))}{\epsilon_0} \right] \\ &< 2\pi \cdot \left( s \cdot \frac{|\lambda_0 - \lambda_{\epsilon_0}|}{\epsilon_0} + \|g(\lambda_{\epsilon_0})\| \right) \\ &\leq 2\pi \cdot (s \cdot C + \pi), \end{aligned}$$

where  $s = |f'(\lambda)|$  for all  $\lambda$ . Therefore,  $Z_\epsilon(X)$  is uniformly bounded on  $X \in A^c \cap B(\zeta)$  for  $0 < |\epsilon| < \eta$ .

Secondly, we aim to find the limit of  $Z_\epsilon(X)$ . For this purpose, we define a map  $u : (-1, 1] \rightarrow (1, \infty)$  by  $u(x) = \arccos(x) \cdot \frac{1}{\sqrt{1-x^2}}$  if  $x \in (-1, 1)$ , and  $u(1) = 1$ . By simple calculations,  $u$  is a monotone decreasing continuous function on  $(-1, 1]$ . Note that  $F(\epsilon)$  is differentiable for  $|\epsilon| < \eta$ . By the

mean value theorem to find the limit of  $Z_\epsilon(X)$ , we have

$$\begin{aligned} Z_{\epsilon_0}(X) &= \frac{d_{Geo}^2(X, f_{\epsilon_0}(\lambda_{f_{\epsilon_0}})) - d_{Geo}^2(X, f(\lambda_f))}{\epsilon_0} \\ &= \frac{\arccos^2(F(\epsilon_0)) - \arccos^2(F(0))}{\epsilon_0} \\ &= -2 \arccos(F(\epsilon_1)) \cdot \frac{1}{\sqrt{1-F^2(\epsilon_1)}} \cdot \frac{dF(\epsilon)}{d\epsilon} \Big|_{\epsilon=\epsilon_1} \end{aligned} \quad (16)$$

for  $0 < |\epsilon_1| < |\epsilon_0| < \eta$ . When  $F(\epsilon_1) = 1$ , the last equality is considered as a limit that is well-defined, because  $\lim_{x \rightarrow 1} u(x) = 1$  and  $u(x)$  is smoothly extended on an open interval containing 1 such that  $u(x)$  is differentiable at  $x = 1$ . By applying chain rule to the derivative of  $F$ , we obtain

$$\begin{aligned} \lim_{\epsilon_0 \rightarrow 0} \frac{\partial F(\epsilon)}{\partial \epsilon} \Big|_{\epsilon=\epsilon_0} &= \lim_{\epsilon_0 \rightarrow 0} \left[ \sin(d_{Geo}(X, f(\lambda_{f_{\epsilon_0}}))) \right. \\ &\quad \cdot \cos(\theta(\lambda_{f_{\epsilon_0}}, X)) \cdot \left( \|g(\lambda_{f_{\epsilon_0}})\| + \epsilon_0 \cdot \frac{\partial \|g(\lambda_{f_\epsilon})\|}{\partial \epsilon} \Big|_{\epsilon=\epsilon_0} \right) \\ &\quad \left. - \lim_{\epsilon_0 \rightarrow 0} \left[ \sin(d_{Geo}(X, f(\lambda_{f_{\epsilon_0}}))) \cdot \frac{\partial d_{Geo}(X, f(\lambda_{f_\epsilon}))}{\partial \epsilon} \Big|_{\epsilon=\epsilon_0} \right] \right]. \end{aligned}$$

In addition,

$$\frac{\partial \|g(\lambda_{f_\epsilon})\|}{\partial \epsilon} \Big|_{\epsilon=\epsilon_0} = \frac{\partial \|g(\lambda)\|}{\partial \lambda} \Big|_{\lambda=\lambda_{f_{\epsilon_0}}} \frac{\partial \lambda(\epsilon)}{\partial \epsilon} \Big|_{\epsilon=\epsilon_0},$$

which exists and does not diverge as  $\epsilon_0$  goes to 0, since  $\|g(\lambda)\| = d_{Geo}(f(\lambda), (f + g)(\lambda))$  is continuously differentiable for  $\lambda$  and  $\frac{\partial \lambda(\epsilon)}{\partial \epsilon} \Big|_{\epsilon=0}$  is bounded by Lemma 5. Moreover,

$$\begin{aligned} \lim_{\epsilon_0 \rightarrow 0} \frac{\partial d_{Geo}(X, f(\lambda_{f_\epsilon}))}{\partial \epsilon} \Big|_{\epsilon=\epsilon_0} &= \lim_{\epsilon_0 \rightarrow 0} \frac{\partial d_{Geo}(X, f(\lambda))}{\partial \lambda} \Big|_{\lambda=\lambda_{f_{\epsilon_0}}} \cdot \frac{\partial \lambda(\epsilon)}{\partial \epsilon} \Big|_{\epsilon=\epsilon_0} \\ &= \frac{\partial d_{Geo}(X, f(\lambda))}{\partial \lambda} \Big|_{\lambda=\lambda_f} \frac{\partial \lambda(\epsilon)}{\partial \epsilon} \Big|_{\epsilon=0} = 0, \end{aligned}$$

where  $\lambda(\epsilon) = \lambda_{f_\epsilon}$ . The last equality is done by the definition of  $\lambda_f$ . Therefore, we have

$$\begin{aligned} \lim_{\epsilon \rightarrow 0} \frac{\partial F(\epsilon)}{\partial \epsilon} &= \|g(\lambda_f)\| \cdot \cos(\theta(\lambda_f, X)) \\ &\quad \cdot \sin(d_{Geo}(X, f(\lambda_f))). \end{aligned} \quad (17)$$

Thirdly, it follows from (16) and (17) that

$$\begin{aligned} \lim_{\epsilon_0 \rightarrow 0} Z_{\epsilon_0}(X) &= \lim_{\epsilon_1 \rightarrow 0} \left[ -2 \arccos(F(\epsilon_1)) \cdot \frac{1}{\sqrt{1-F^2(\epsilon_1)}} \right. \\ &\quad \left. \cdot \frac{dF(\epsilon)}{d\epsilon} \Big|_{\epsilon=\epsilon_1} \right] \\ &= -2u \left( \cos(d_{Geo}(X, f(\lambda_f))) \right) \cdot \|g(\lambda_f)\| \\ &\quad \cdot \cos(\theta(\lambda_f, X)) \cdot \sin(d_{Geo}(X, f(\lambda_f))) \\ &= -2d_{Geo}(X, f(\lambda_f)) \cdot \frac{1}{\sin(d_{Geo}(X, f(\lambda_f)))} \\ &\quad \cdot \|g(\lambda_f)\| \cdot \cos(\theta(\lambda_f, X)) \cdot \sin(d_{Geo}(X, f(\lambda_f))) \\ &= -2d_{Geo}(X, f(\lambda_f)) \cdot \|g(\lambda_f)\| \cdot \cos(\theta(\lambda_f, X)), \end{aligned} \quad (18)$$

In the case of  $d_{Geo}(X, f(\lambda_f)) = 0$ , the same result follows since both (18) and (19) are zero. Thus, by Proposition 5, the equation (19) is established for a.e.  $X \in B(\zeta)$ . Next,

we notice that, for a smooth curve  $f$ , it can be shown that  $M_\lambda := \{x \in S^2 \mid \lambda_f(x) = \lambda\}$  is a subset of the great circle perpendicular to  $f$  at  $f(\lambda)$  by Lemma 2. Let  $S_\lambda$  be the great circle perpendicular to  $f$  at  $f(\lambda)$ . That is,  $M_\lambda \subset S_\lambda \cong S^1$ . Moreover, a connected proper subset of  $S_\lambda$  is isometric to a line with the same length in  $\mathbb{R}$ , which makes the intrinsic mean on  $M_\lambda$  feasible. Note that if the length is less than  $\pi/2$ , the intrinsic mean is unique. Thus,  $f$  is an intrinsic principal curve of  $X$ , by the definition of  $\theta(\lambda_f, X)$  and  $\cos(\pi - \theta) = -\cos(\theta)$ , if and only if

$$\mathbb{E}[d_{Geo}(X, f(\lambda_f)) \cdot \cos(\theta(\lambda_f, X)) \mid \lambda_f(X) = \lambda] = 0, \quad \text{for a.e. } \lambda.$$

Finally, it follows, from (19) and by the bounded convergence theorem, that

$$\begin{aligned} & \frac{\partial \mathbb{E}_X[d_{Geo}^2(X, f + \epsilon g)]}{\partial \epsilon} \Big|_{\epsilon=0} \\ &= \lim_{\epsilon \rightarrow 0} \left[ \frac{\mathbb{E}_X[d_{Geo}^2(X, f + \epsilon g)] - \mathbb{E}_X[d_{Geo}^2(X, f)]}{\epsilon} \right] \\ &= \mathbb{E}_X \left[ \lim_{\epsilon \rightarrow 0} \frac{d_{Geo}^2(X, f + \epsilon g) - d_{Geo}^2(X, f)}{\epsilon} \right] \\ &= \mathbb{E}_\lambda \left[ \mathbb{E} \left[ \lim_{\epsilon \rightarrow 0} Z_\epsilon(X) \mid \lambda_f(X) = \lambda \right] \right] \\ &= -2\mathbb{E}_\lambda \left[ \mathbb{E}[d_{Geo}(X, f(\lambda_f(X))) \cdot \|g(\lambda_f(X))\| \cdot \cos(\theta(\lambda_f, X)) \mid \lambda_f(X) = \lambda] \right] \\ &= -2\mathbb{E}_\lambda \left[ \|g(\lambda)\| \cdot \mathbb{E}[d_{Geo}(X, f(\lambda_f(X))) \cdot \cos(\theta(\lambda_f, X)) \mid \lambda_f(X) = \lambda] \right] \\ &= 0. \end{aligned}$$

Conversely, we assume that

$$\mathbb{E}_\lambda \left[ \|g(\lambda)\| \cdot \mathbb{E}[d_{Geo}(X, f(\lambda_f)) \cdot \cos(\theta(\lambda_f, X)) \mid \lambda_f(X) = \lambda] \right] = 0,$$

for all  $f + g (= h)$  such that  $\|g\| \neq \pi$  and  $\|g'\| \leq 1$ . It follows that

$$\mathbb{E}[d_{Geo}(X, f(\lambda_f)) \cdot \cos(\theta(\lambda_f, X)) \mid \lambda_f(X) = \lambda] = 0, \quad \text{for a.e. } \lambda,$$

which is equivalent to that  $f$  is an intrinsic principal curve of  $X$ .  $\square$

## APPENDIX D INFLUENCE OF $T$ AND $q$ .

Here we discuss the influence of the hyperparameters  $T$  and  $q$ . To this end, we consider the waveform simulated data used in Section 4.2.1. Figure 4 visualizes the fitted curves by the proposed extrinsic method for various  $q$ 's in the range of  $[0.01, 0.1]$  at intervals of 0.01 with a fixed  $T = 500$ . As shown in the top panels of Figure 4, the resulting curve with  $q = 0.01$  is wiggly, and the curve with  $q = 0.1$  is almost flat. In general, the curves tend to overfit data when the  $q$  value is small, whereas the curves tend to underfit data when the  $q$  value is large. On the other hand, the bottom panels of Figure 4 show the fitted curves by the same method for a fixed  $q = 0.06$  and varying  $T$  in  $\{10, 20, 50, 100, 200, 500\}$ . The curve of the bottom left panel implemented by a small  $T$  value, such as  $T = 10$ , does not represent the data well. For appropriate  $T$  values, the spherical principal curves of

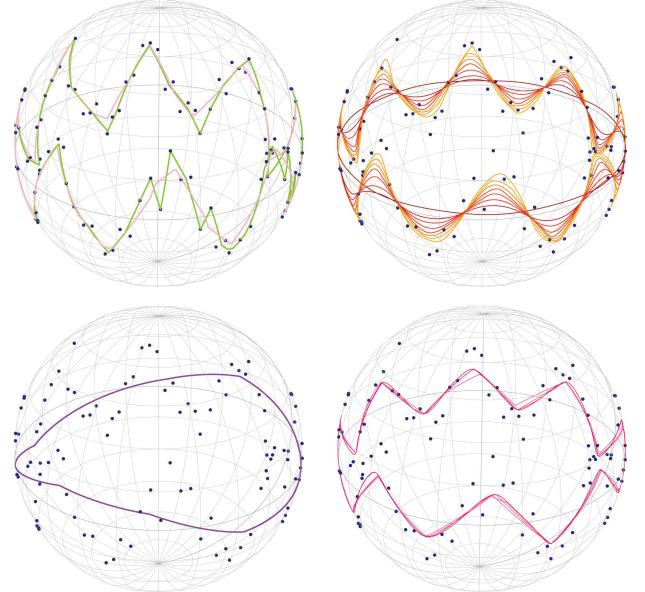


Figure 4: Noisy waveform simulated data (blue). (Top left) Extrinsic principal curves with  $q = 0.01$  (green) and  $0.02$  (pink) for fixed  $T = 500$ . (Top right) Influence of varying  $q$  over  $[0.03, 0.1]$  with a step size 0.01 (from yellow to brown) for fixed  $T = 500$ . (Bottom left) Extrinsic principal curves (purple) with  $T = 10$  and  $q = 0.06$ . (Bottom right) Influence of varying  $T$  in  $\{20, 50, 100, 200, 500\}$  (from violet to red) for a fixed  $q = 0.06$ .

the right panel successfully recover the underlying structure of the data.

## REFERENCES

- [1] K. V. Mardia and R. J. Gadsden, "A small circle of best fit for spherical data and areas of vulcanism," *Journal of the Royal Statistical Society: Series C*, vol. 26, pp. 238–245, 1977.
- [2] N. H. Gray, P. A. Geiser, and J. R. Geiser, "On the least-squares fit of small and great circles to spherically projected orientation data," *Journal of the International Association for Mathematical Geology*, vol. 12, no. 3, pp. 173–184, 1980.
- [3] T. Hastie and W. Stuetzle, "Principal curves," *Journal of the American Statistical Association*, vol. 84, no. 406, pp. 502–516, 1989.
- [4] P. T. Fletcher, C. Lu, S. M. Pizer, and S. Joshi, "Principal geodesic analysis for the study of nonlinear statistics of shape," *IEEE Transactions on Medical Imaging*, vol. 23, no. 8, pp. 995–1005, 2004.
- [5] S. Jung, M. Foskey, and J. Marron, "Principal arc analysis on direct product manifolds," *The Annals of Applied Statistics*, vol. 5, no. 1, pp. 578–603, 2011.
- [6] S. Jung, I. L. Dryden, and J. Marron, "Analysis of principal nested spheres," *Biometrika*, vol. 99, no. 3, pp. 551–568, 2012.
- [7] S. Hauberg, "Principal curves on riemannian manifolds," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 38, no. 9, p. 1915, 2016.
- [8] W. M. Boothby, *An Introduction to Differentiable Manifolds and Riemannian Geometry*. Academic Press, 1986.
- [9] L. W. Tu, *Differential Geometry: Connections, Curvature, and Characteristic Classes*. Springer, 2017, vol. 275.
- [10] I. D. Coope, "Circle fitting by linear and nonlinear least squares," *Journal of Optimization Theory and Applications*, vol. 76, no. 2, pp. 381–388, 1993.
- [11] I. Kåsa, "A circle fitting procedure and its error analysis," *IEEE Transactions on Instrumentation and Measurement*, no. 1, pp. 8–14, 1976.
- [12] D. Umbach and K. N. Jones, "A few methods for fitting circles to data," *IEEE Transactions on Instrumentation and Measurement*, vol. 52, no. 6, pp. 1881–1885, 2003.

- [13] B. Eltzner, S. Huckemann, K. V. Mardia *et al.*, "Torus principal component analysis with applications to rna structure," *The Annals of Applied Statistics*, vol. 12, no. 2, pp. 1332–1359, 2018.
- [14] L. Scales, *Introduction to Non-linear Optimization*. Macmillan International Higher Education, 1985.
- [15] P. Jain and P. Kar, "Non-convex optimization for machine learning," *Foundations and Trends® in Machine Learning*, vol. 10, pp. 142–336, 2017.
- [16] Y. Zhang, Q. Qu, and J. Wright, "From symmetry to geometry: Tractable nonconvex problems," *arXiv preprint arXiv:2007.06753*, 2020.
- [17] T. Hastie, "Principal curves and surfaces," Ph.D. dissertation, Stanford University, 1984.