

# Statistical inference on the Hilbert sphere with application to random densities\*

Xiongtao Dai

*Department of Statistics, Iowa State University, Ames, Iowa 50011 USA*  
*e-mail: [xdai@iastate.edu](mailto:xdai@iastate.edu)*

**Abstract:** The infinite-dimensional Hilbert sphere  $\mathcal{S}^\infty$  has been widely employed to model density functions and shapes, extending the finite-dimensional counterpart. We consider the Fréchet mean as an intrinsic summary of the central tendency of data lying on  $\mathcal{S}^\infty$ . For sound statistical inference, we derive properties of the Fréchet mean on  $\mathcal{S}^\infty$  by establishing its existence and uniqueness as well as a root- $n$  central limit theorem (CLT) for the sample version, overcoming obstructions from infinite-dimensionality and lack of compactness on  $\mathcal{S}^\infty$ . Intrinsic CLTs for the estimated tangent vectors and covariance operator are also obtained. Asymptotic and bootstrap hypothesis tests for the Fréchet mean based on projection and norm are then proposed and are shown to be consistent. The proposed two-sample tests are applied to make inference for daily taxi demand patterns over Manhattan, modeled as densities, of which the square root densities are analyzed on the Hilbert sphere. Numerical properties of the proposed hypothesis tests which utilize the spherical geometry are studied in the real data application and simulations, where we demonstrate that the tests based on the intrinsic geometry compare favorably to those based on an extrinsic or flat geometry.

**MSC2020 subject classifications:** Primary 62G20; secondary 62G05, 62G99.

**Keywords and phrases:** Intrinsic mean, functional data analysis, Hilbert geometry, Riemannian manifold, large sample property.

Received October 2020.

## Contents

|   |  |     |
|---|--|-----|
| 1 | Introduction . . . . .   | 701 |
| 2 | Hilbert sphere geometry . . . . .                                | 704 |
| 3 | Properties of the Fréchet mean on $\mathcal{S}^\infty$ . . . . . | 705 |
|   | 3.1 Intrinsic mean . . . . .                                     | 705 |
|   | 3.2 Tangent vectors and the covariance operator . . . . .        | 707 |
| 4 | Hypothesis tests for the intrinsic mean . . . . .                | 709 |
|   | 4.1 General setup . . . . .                                      | 709 |
|   | 4.2 Asymptotic tests . . . . .                                   | 711 |

---

arXiv: [2101.00527](https://arxiv.org/abs/2101.00527)

\*The research is supported in part by NSF grant DMS-2113713.

|     |  |     |
|-----|--|-----|
| 4.3 | Bootstrap tests . . . . .  | 711 |
| 5   | Simulation studies . . . . .                                     | 713 |
| 6   | Data application: taxi demands . . . . .                         | 714 |
| A   | Fréchet derivatives . . . . .                                    | 717 |
| B   | Proof of <a href="#">Theorem 3.1</a> . . . . .                   | 719 |
| C   | Proofs of other results in <a href="#">Section 3.1</a> . . . . . | 721 |
| D   | Proofs of the results in <a href="#">Section 3.2</a> . . . . .   | 723 |
| E   | Differentiating the squared geodesic distance . . . . .          | 725 |
| F   | Proofs for Corollaries in <a href="#">Section 4</a> . . . . .    | 729 |
| G   | Paired-sample test . . . . .                                     | 729 |
| H   | Additional simulation results for the one-sample test . . . . .  | 731 |
|     | Acknowledgments . . . . .  | 731 |
|     | References . . . . .   | 731 |

## 1. Introduction

We aim to develop statistical theory and inferential methods for analyzing a sample of random elements taking values on the Hilbert sphere  $\mathcal{S}^\infty = \{f \in \mathbb{H} \mid \|f\|_{\mathbb{H}} = 1\}$ , where  $\mathbb{H}$  is a separable infinite-dimensional Hilbert space with inner product  $\langle \cdot, \cdot \rangle_{\mathbb{H}}$  and norm  $\|\cdot\|_{\mathbb{H}}$ . The spherical Hilbert geometry gives rise to invariance properties and efficient calculation of geometric quantities, for instance the geodesics, which promotes much of its prevailing applications to model different data objects:

- *Probability distributions* supported on a compact interval  $I$  is represented by density functions in  $\text{Dens}(I) = \{y : I \rightarrow \mathbb{R} \mid \int_I y(s) ds = 1, y(s) > 0, s \in I\}$ . To analyze the densities in a reparametrization-invariant fashion, Srivastava, Jermyn and Joshi (2007) proposed a nonparametric Fisher-Rao metric with a reparametrization-invariant property, generalizing the well-known finite-dimensional version (Rao, 1945). The Fisher-Rao metric is conveniently expressed and calculated through representing the densities by the corresponding square root densities in  $\mathcal{X} = \{x : I \rightarrow \mathbb{R} \mid x(s) = \sqrt{y(s)}, s \in I, y \in \text{Dens}(I)\}$ , where  $\mathcal{X}$  is the positive orthant of the Hilbert sphere  $\mathcal{S}^\infty$  modeled in  $\mathbb{H} = L^2(I)$ . The square root density framework has been widely applied in modeling the orientation distribution functions in high angular resolution diffusion images (Cheng et al., 2009; Du et al., 2014) and time-warping functions (Tucker, Wu and Srivastava, 2013; Yu, Lu and Marron, 2017).
- *Plane and space curves* such as shape contours and motion trajectories are often compared in a group action-invariant fashion for pattern recognition. To compare the shape of curves with the translation and scaling effects removed, an observed smooth parametrized curve in  $\mathbb{R}^d$ ,  $d = 2$  or  $3$  is centered and scaled to have unit length, obtaining a centered-and-scaled curve in  $\mathcal{F} = \{f : [0, 1] \rightarrow \mathbb{R}^d \mid f(0) = 0, \int_0^1 |f'(s)| ds = 1\}$ . Each  $f \in \mathcal{F}$  is represented by its square root velocity function (SRVF) (Joshi et al., 2007)  $\beta_f : [0, 1] \rightarrow \mathbb{R}^d$ ,  $\beta_f(s) = |f'(s)|^{-1/2} f'(s)$ ,  $s \in [0, 1]$ . Now, the SRVF

$\beta_f$  lies on the Hilbert sphere  $\mathcal{S}^\infty$  in the Hilbert space  $\mathbb{H} = L^2([0, 1], \mathbb{R}^d) = \{h : [0, 1] \rightarrow \mathbb{R}^d \mid \int |h(s)|^2 ds < \infty\}$ , equipped with the inner product  $\langle h_1, h_2 \rangle_{\mathbb{H}} = \int_0^1 h_1(s)^T h_2(s) ds$  for  $h_1, h_2 \in \mathbb{H}$ . The spherical geometry on  $\mathcal{S}^\infty$  for the SRVFs induces a special case of the elastic metric (Younes, 1998) on the space of centered-and-scaled curves  $\mathcal{F}$ , so that the distance between curves are given by the square root of the minimal energy to transform between them. This geometry has demonstrated attractive and interpretable performance in practical curve matching tasks (Su et al., 2014; Bauer, Eslitzbichler and Grasmair, 2017; Xie et al., 2017; Strait, Chkrebti and Kurtek, 2019).

Applications of Hilbert sphere in computer vision, medical imaging, and biology necessitates hypothesis tests backed by solid theory, but there are so far no available asymptotic results to support hypothesis tests under an intrinsic geometry on  $\mathcal{S}^\infty$ ; see for example Wu and Srivastava (2014); Henning and Srivastava (2016) who applied U-statistics to perform two-sample comparisons. For random densities,  $F$ -tests have been considered by Petersen, Liu and Divani (2019) under a Wasserstein geometry and Dubey and Müller (2019) under a more general object-oriented framework, but the latter utilizes an entropy number condition in the proof which has not been verified on  $\mathcal{S}^\infty$ . This provides motivation for our work, which aims to derive asymptotic distributional results on  $\mathcal{S}^\infty$ , develop consistent hypothesis tests, and showcase applications of the proposed hypothesis tests to random densities.

Let  $(\mathcal{M}, \rho)$  be a metric space where  $\rho : \mathcal{M} \times \mathcal{M} \rightarrow \mathbb{R}$  is the distance metric. Consider an  $\mathcal{M}$ -valued random element  $X : \Omega \rightarrow \mathcal{M}$  measurable between a complete probability space  $(\Omega, \mathcal{A}, P)$  and  $\sigma$ -algebra  $\mathcal{B}(\mathcal{M})$  induced on  $\mathcal{M}$  by  $\rho$ . The *Fréchet mean* (Fréchet, 1948) is a commonly used location descriptor for non-Euclidean data objects and is termed the *intrinsic mean* if  $\mathcal{M}$  is a Riemannian manifold. If there exists a unique minimizer of the Fréchet functional

$$M(\cdot) = \mathbb{E}\rho^2(X, \cdot), \quad (1)$$

then the minimizer  $\mu = \arg \min_{p \in \mathcal{M}} M(p)$  is called the Fréchet mean of  $X$ . Similarly, for independent realizations  $X_1, \dots, X_n$  of  $X$ , the sample Fréchet functional is

$$M_n(\cdot) = \frac{1}{n} \sum_{i=1}^n \rho^2(X_i, \cdot), \quad (2)$$

and the sample Fréchet mean is  $\hat{\mu} = \arg \min_{p \in \mathcal{M}} M_n(p)$  given existence and uniqueness. This work focuses on  $\mathcal{M} = \mathcal{S}^\infty$  with  $\rho$  being the geodesic distance.

When the Riemannian manifold  $\mathcal{M}$  is finite-dimensional, theory and methods for the intrinsic Fréchet mean have been well investigated. Extensive study has characterized the existence and uniqueness (Karcher, 1977; Ziezold, 1977; Le, 2001; Afsari, 2011), consistency (Bhattacharya and Patrangenaru, 2003), and central limit theorems (CLTs) (Bhattacharya and Patrangenaru, 2005; Bhattacharya and Lin, 2017) for the Fréchet mean on a general finite-dimensional

$\mathcal{M}$ . A slower-than- $n^{1/2}$  CLT has been studied on the circle (Hotz and Huckemann, 2015) and, more recently, high-dimensional spheres and general Riemannian manifolds (Eltzner and Huckemann, 2019). Various statistical methods for manifold-valued data have also been considered, for example, confidence intervals (Bhattacharya and Patrangenaru, 2005), hypothesis tests (Huckemann, 2012), regression (Zhu et al., 2009), and principal component analysis (Fletcher et al., 2004).

Also well-studied are random objects lying in an infinite-dimensional Hilbert space, which are termed functional data (Wang, Chiou and Müller, 2016). Thanks to the flat geometry and vector space structure, the definition of the mean element is straightforward, and the asymptotic theory can be obtained through extending the multivariate case (Hsing and Eubank, 2015). The Hilbert mean extends to a class of Hilbert manifolds through the extrinsic (Ellingson, Patrangenaru and Ruymgaart, 2013) or the transformation method (Petersen and Müller, 2016) by mapping the original data into a Hilbert space and then perform analysis there. However, these approaches depend on the embedding or transformation and does not preserve the intrinsic geometry on  $\mathcal{S}^\infty$ , so the theory derived for these methods cannot be applied to obtain an intrinsic analysis for objects lying on  $\mathcal{S}^\infty$ .

Little is known about the theory for the Fréchet mean on curved infinite-dimensional geometries. In particular, on the Hilbert sphere  $\mathcal{S}^\infty$ , the existence and uniqueness of the Fréchet mean has not been established, and no distributional results are available for the inference of the intrinsic mean. Difficulties in deriving these results on  $\mathcal{S}^\infty$  include the lack of compactness and the positive curvature, which prevents proof techniques developed for a finite-dimensional manifold (Bhattacharya and Patrangenaru, 2005; Afsari, 2011) or Hilbert space (Hsing and Eubank, 2015) to be applicable. General results for the convergence rate of the sample Fréchet mean in abstract settings has been established by Gouic et al. (2019); Ahidar-Coutrix, Le Gouic and Paris (2019); Schötz (2019). The latter two works applied empirical process theory under entropy conditions, which are challenging to verify for the infinite-dimensional positively curved manifold  $\mathcal{S}^\infty$ ; no distributional results was established were provided there, and the uniqueness and existence of the Fréchet mean were assumed.

Our contributions include establishing the uniqueness and existence of the Fréchet mean on  $\mathcal{S}^\infty$  and its large sample properties, and applying these results to derive valid hypothesis tests for data objects. We state in [Section 3](#) theoretical properties of the intrinsic data analysis on  $\mathcal{S}^\infty$ . The existence and uniqueness of the intrinsic mean are shown in [Theorem 3.1](#), and large sample properties of the sample intrinsic mean  $\hat{\mu}$  including a strong LLN and a CLT are stated in [Proposition 3.1](#) and [Theorem 3.2](#), respectively. Theoretical hurdles in verifying tightness and Lipschitz continuity are overcome by utilizing weak compactness, Hilbert differential geometry (Lang, 1999), and a careful analysis of the spherical geometry. The asymptotic covariance of the limiting Gaussian element in the CLT of  $\hat{\mu}$  is shown to be consistently estimated by an empirical version in [Proposition 3.2](#). To quantify the deviation of data observations around the intrinsic mean, we formulate intrinsic CLTs for the tangent vectors

in [Corollary 3.2](#) and the covariance operator in [Theorem 3.3](#) based on parallel transport. Unlike the case in a flat Hilbert space, additional terms manifest in the asymptotic covariance of the sample covariance operator due to the curvature of  $\mathcal{S}^\infty$ . Results on the covariance will be useful to derive theoretical properties of principal component analysis (e.g. Tucker, Wu and Srivastava, [2013](#)) on  $\mathcal{S}^\infty$ , though the latter exposition is out of scope of this work.

Asymptotic and bootstrap hypothesis tests for the population means are proposed in [Section 4](#) as a result of the CLTs, with consistent properties of the test derived in [Corollary 4.1–Corollary 4.5](#). Two test statistics are constructed using the norm and the projections of the intrinsic sample means expressed on a chart, respectively, analogous to those developed for functional data (Horváth, Kokoszka and Reeder, [2013](#); Aue, Rice and Sönmez, [2018](#)). Simulation studies in [Section 5](#) demonstrate that the hypothesis tests based on the intrinsic mean have smaller bias than that based on the extrinsic mean when data are asymmetrically distributed around the mean. The asymptotics is shown to kick in with a moderate sample size of 50 despite the infinite-dimensionality. A real data study of daily taxi demands in Manhattan is presented in [Section 6](#), where the taxi demands are modeled as densities for which the square roots are analyzed on  $\mathcal{S}^\infty$ . The spherical geometry demonstrates advantage over alternative geometries in detecting changes in the demand patterns. The proofs of the main results are deferred to the Appendix.

## 2. Hilbert sphere geometry

Let  $\mathbb{H}$  denote a separable Hilbert space with inner product  $\langle \cdot, \cdot \rangle_{\mathbb{H}}$  and induced norm  $\|\cdot\|_{\mathbb{H}}$ . The infinite-dimensional Hilbert sphere  $\mathcal{S}^\infty$  is the unit sphere in  $\mathbb{H}$  with well-known geometry and explicit expressions for its geometric quantities. In particular, the *tangent space* of  $\mathcal{S}^\infty$  at  $p \in \mathcal{S}^\infty$  is the subspace  $T_p\mathcal{S}^\infty = \{v \in \mathbb{H} \mid \langle v, p \rangle_{\mathbb{H}} = 0\}$  of  $\mathbb{H}$  with codimension one. The *metric tensor*  $\langle u, v \rangle_p$  for tangent vectors  $u, v \in T_p\mathcal{S}^\infty$  at  $p \in \mathcal{S}^\infty$  is induced from and equal to the  $\mathbb{H}$ -inner product  $\langle u, v \rangle_{\mathbb{H}}$ . The *geodesic distance*  $\rho : \mathcal{S}^\infty \times \mathcal{S}^\infty \rightarrow \mathbb{R}$  is given by  $\rho(f, g) = \arccos(\langle f, g \rangle_{\mathbb{H}})$ . The Riemannian *exponential map* at  $p \in \mathcal{S}^\infty$  is  $\exp_p : T_p\mathcal{S}^\infty \rightarrow \mathcal{S}^\infty$ ,  $\exp_p v = \cos(\|v\|_p)p + \sin(\|v\|_p)\|v\|_p^{-1}v$ , a map from a tangent space to the sphere that preserves the distance to the origin, i.e.  $\|v\|_p = \rho(p, \exp_p v)$ . The inverse exponential map, or the *logarithm map*, at  $p \in \mathcal{S}^\infty$  is defined as  $\log_p : \mathcal{S}^\infty \setminus \{-p\} \rightarrow T_p\mathcal{S}^\infty$ ,  $\log_p x = \arccos(\langle p, x \rangle_p)\|u\|_p^{-1}u$ , where  $-p$  is the antipodal point of  $p$  and  $u = x - \langle p, x \rangle_p p$ . A chart  $\tau : U \subset \mathcal{S}^\infty \rightarrow \mathbb{G}$  is a homeomorphism that maps  $U$  onto open subset  $\tau(U)$  of a separable Hilbert space  $\mathbb{G}$  for coordinates. We require  $\tau$  to be smooth, so that  $\tau^{-1} : \tau(U) \subset \mathbb{G} \rightarrow \mathbb{H}$  is a smooth diffeomorphism. For example,  $\tau_p(\cdot) = \log_p(\cdot)$  is a smooth chart defined on  $U_p = \mathcal{S}^\infty \setminus \{-p\}$ , for  $p \in \mathcal{S}^\infty$ . We refer to Lang ([1999](#)) for general background in Hilbert differential geometry.

For concreteness, we consider in our numerical illustrations the Hilbert space  $\mathbb{H} = L^2(I) = \{f : I \rightarrow \mathbb{R} \mid \int_I f(s)^2 ds < \infty\}$  of (equivalent classes of) square-integrable functions on a compact Euclidean subset  $I$ , while we note that the

definition of a Hilbert sphere is fully intrinsic through isometry (Lang, 1999).

### 3. Properties of the Fréchet mean on $\mathcal{S}^\infty$

#### 3.1. Intrinsic mean

Unlike the case in a Euclidean space, the Fréchet mean for data lying on a manifold may not exist and may not be unique. A neighborhood condition (A1) is needed to guarantee the existence and uniqueness of the intrinsic means  $\mu$  defined in (1).

(A1) The support  $\mathcal{U} := \text{supp}(X) \subset \mathcal{S}^\infty$  of  $X$  satisfies  $\sup_{x,y \in \mathcal{U}} \rho(x,y) \leq \pi/2$ .

A closely related neighborhood condition, originally posed for a general finite-dimensional Riemannian manifold but stated here for a  $d$ -dimensional sphere  $\mathcal{S}^d$  for comparison, is (A2) The support  $\mathcal{U}$  of  $X$  lies in the geodesic ball  $B(q,r)$  centered at some  $q \in \mathcal{S}^d$  with radius  $r < r^*$ . With  $r^* = \pi/4$ , Karcher (1977) showed the existence and local uniqueness of the Fréchet mean within  $B(q,r)$ , while Le (2001) concluded global uniqueness over  $\mathcal{S}^d$ . Stronger uniqueness results assuming  $r^* = \pi/2$  were established by Kendall (1990) for the local uniqueness within the geodesic ball and by Afsari (2011) for the global uniqueness. The arguments of Afsari (2011) rely on compactness for showing the existence and uniqueness, and the uniqueness proof utilizes the Poincaré–Hopf theorem for compact manifolds, which does not apply on the infinite-dimensional Hilbert sphere  $\mathcal{S}^\infty$  due to a lack of compactness. We interpret the additional concentration required by (A1) as compared to (A2) with  $r^* = \pi/2$  as a compensation for a lack of compactness on  $\mathcal{S}^\infty$ . The next theorem establishes the uniqueness and existence of the Fréchet mean on  $\mathcal{S}^\infty$ .

**Theorem 3.1.** *If (A1) holds, then there exists a unique intrinsic mean  $\mu$  of  $X$  on  $\mathcal{S}^\infty$ . Furthermore,  $\mathbb{P}(\rho(\mu, X) < \pi/2) = 1$ .*

The existence proof is based on the weak compactness of the unit ball in  $\mathbb{H}$  by the Banach–Alaoglu theorem (Rudin, 1973) and an analysis of  $\rho^2(x, \cdot)$ , the uniqueness result makes use of the convexity of  $\rho$ , and the proximity of  $\mu$  and  $X$  is obtained through a reflection argument used in Afsari (2011). With the well-definedness of the population and sample intrinsic means  $\mu$  and  $\hat{\mu}$  guaranteed by Theorem 3.1, we derive the consistency for  $\hat{\mu}$  based on  $M$ -estimation and empirical process (see, e.g., van der Vaart and Wellner, 1996).

**Proposition 3.1.** *If (A1) holds, then*

$$\rho(\hat{\mu}, \mu) = o(1) \quad a.s.$$

Some notations are needed to state the CLT for  $\hat{\mu}$ . Let  $\mathcal{B}(E_1, E_2)$  be the Banach space of bounded linear operators between Banach spaces  $E_1$  and  $E_2$ . For any Hilbert space  $(\mathbb{F}, \langle \cdot, \cdot \rangle_{\mathbb{F}})$ , by the Riesz representation theorem, an element  $f \in \mathbb{F}$  is in one-to-one correspondence with bounded linear functional  $f^* :=$

$\langle f, \cdot \rangle_{\mathbb{F}}$  in the dual Hilbert space  $\mathbb{F}^* = \mathcal{B}(\mathbb{F}, \mathbb{R})$  equipped with inner product  $\langle f^*, g^* \rangle_{\mathbb{F}^*} = \langle f, g \rangle_{\mathbb{F}}$ . Let  $\mathcal{A}^*$  denote the adjoint of a linear operator  $\mathcal{A}$  between Hilbert spaces. The tensor product  $\otimes : \mathbb{F} \times \mathbb{F} \rightarrow \mathcal{B}(\mathbb{F}, \mathbb{F})$  of two elements  $f, g \in \mathbb{F}$  is defined according to  $(f \otimes g)h = \langle f, h \rangle_{\mathbb{F}} g$  for any  $h \in \mathbb{F}$ . Given a smooth chart  $\tau : U \subset \mathcal{S}^\infty \rightarrow \mathbb{G}$  defined on a neighborhood  $U$  of  $\mu$ , we re-express  $\mu$ ,  $\hat{\mu}$ , and  $\rho$  as  $\mu_\tau = \tau(\mu)$ ,  $\hat{\mu}_\tau = \tau(\hat{\mu})$ , and  $\rho_\tau : \mathcal{S}^\infty \times \mathbb{G} \rightarrow \mathbb{R}$ ,  $\rho_\tau(x, e) = \rho(x, \tau^{-1}(e))$ . Let  $D_2$  denote the partial (Fréchet) derivative of a multivariate function w.r.t. the second argument, where the definition of Fréchet derivatives is reviewed in [Appendix A](#). Set  $\psi_\tau : \mathcal{S}^\infty \times \mathbb{G} \rightarrow \mathcal{B}(\mathbb{G}, \mathbb{R})$ ,  $\psi_\tau(x, e) = D_2 \rho_\tau^2(x, e)$ ;  $\mathcal{F}_\tau = \mathbb{E}[\psi_\tau(X, \mu_\tau) \otimes \psi_\tau(X, \mu_\tau)] \in \mathcal{B}(\mathbb{G}^*, \mathbb{G}^*)$ ; and  $\Lambda_\tau = \mathbb{E}[D_2^2 \rho_\tau^2(X, \mu_\tau)] \in \mathcal{B}(\mathbb{G}, \mathcal{B}(\mathbb{G}, \mathbb{R}))$ ; the explicit forms of these quantities are obtained in [Appendix E](#).

**Theorem 3.2.** *Let  $(U, \tau)$  be a chart of  $\mathcal{S}^\infty$  in a neighborhood of  $\mu$ . Under (A1),*

$$\sqrt{n}(\hat{\mu}_\tau - \mu_\tau) \xrightarrow{L} Z$$

as  $n \rightarrow \infty$ , where  $Z$  is a Gaussian random element in  $\mathbb{G}$  with mean zero and covariance operator  $\mathcal{T} = (\Lambda_\tau^{-1}) \mathcal{F}_\tau (\Lambda_\tau^{-1})^* \in \mathcal{B}(\mathbb{G}, \mathbb{G})$ , satisfying

$$\langle h_1, \mathcal{T} h_2 \rangle_{\mathbb{G}} = \text{cov}(\langle \Lambda_\tau^{-1} \psi_\tau(X, \mu_\tau), h_1 \rangle_{\mathbb{G}}, \langle \Lambda_\tau^{-1} \psi_\tau(X, \mu_\tau), h_2 \rangle_{\mathbb{G}})$$

for  $h_1, h_2 \in \mathbb{G}$ . The operator  $\Lambda_\tau$  is continuously invertible.

[Theorem 3.2](#) follows from a linearization argument applied to the chart representation  $M_{n, \tau} : \mathbb{G} \rightarrow \mathbb{R}$ ,  $M_{n, \tau}(e) = M_n(\tau^{-1}(e))$  of the Fréchet functional  $M_n(\cdot)$  in a neighborhood of  $\mu_\tau$ , where the uniform convergence of the residual is obtained due to the simple dependency of the geodesic distance  $\rho$  on the inner product of its arguments. The key difficulty is to verify the Lipschitz continuity of the criterion function under the infinite-dimensional setup, which is handled by a careful analysis of the Hilbert sphere geometry. The CLT is intrinsic in the sense that if  $\eta : V \rightarrow \mathbb{G}$  is another chart defined on a neighborhood  $V$  of  $\mu$ , then  $\sqrt{n}[\eta(\hat{\mu}) - \eta(\mu)]$  converges in law to  $D(\eta \circ \tau^{-1})(\mu_\tau)Z$  where  $Z$  is the limiting Gaussian element under  $\tau$ .

The asymptotic distribution in [Theorem 3.2](#) must be estimated in inferential tasks such as hypothesis testing which will be detailed in [Section 4](#). The next result states that the asymptotic distribution of the limiting Gaussian element can be estimated consistently. Define  $\hat{\mathcal{T}} = (\hat{\Lambda}_\tau^{-1}) \hat{\mathcal{F}}_\tau (\hat{\Lambda}_\tau^{-1})^*$ ,  $\hat{\mathcal{F}}_\tau = n^{-1} \sum_{i=1}^n \psi_\tau(X_i, \hat{\mu}_\tau) \otimes \psi_\tau(X_i, \hat{\mu}_\tau)$ , and  $\hat{\Lambda}_\tau = n^{-1} \sum_{i=1}^n D_2^2 \rho_\tau^2(X_i, \hat{\mu}_\tau)$ . For an operator  $\mathcal{C} : E_1 \rightarrow E_2$  between Banach spaces  $(E_1, \|\cdot\|_{E_1})$  and  $(E_2, \|\cdot\|_{E_2})$ , we let  $\|\mathcal{C}\| = \sup_{\|h\|_{E_1}=1} \|\mathcal{C}h\|_{E_2}$  denote the operator norm; if  $E_1 = E_2 = \mathbb{F}$  for a Hilbert space  $(\mathbb{F}, \langle \cdot, \cdot \rangle_{\mathbb{F}})$ , let  $\|\mathcal{C}\|_1 = \sum_{j=1}^\infty \langle e_j, (\mathcal{C}^* \mathcal{C})^{1/2} e_j \rangle_{\mathbb{F}}$  denote the trace norm, where  $\{e_j\}_{j=1}^\infty$  is an arbitrary complete orthonormal basis of  $\mathbb{F}$ .

**Proposition 3.2.** *Under the conditions of [Theorem 3.2](#), as  $n \rightarrow \infty$ ,*

$$\left\| \hat{\mathcal{T}} - \mathcal{T} \right\|_1 = o(1) \quad \text{a.s.}$$

Let  $Z_n$  and  $Z$  be zero-mean Gaussian random elements in  $\mathbb{G}$  with covariance  $\hat{\mathcal{T}}$  and  $\mathcal{T}$ , respectively. Then, as  $n \rightarrow \infty$ ,

$$Z_n \xrightarrow{L} Z.$$

### 3.2. Tangent vectors and the covariance operator

On a nonlinear manifold, the deviation of an observation around the intrinsic mean is commonly characterized by the corresponding tangent vector constructed from the logarithm map. The tangent vector of  $X$  and its empirical version are respectively defined as

$$V = \log_{\mu} X, \quad \hat{V} = \log_{\hat{\mu}} X.$$

Similarly, let  $V_i = \log_{\mu} X_i$  and  $\hat{V}_i = \log_{\hat{\mu}} X_i$  denote the corresponding quantities for an observation  $X_i$ ,  $i = 1, \dots, n$ .

**Proposition 3.3.** *If  $\mu$  is the intrinsic mean of an  $\mathcal{S}^{\infty}$ -valued random variable  $X$  with  $\mathbb{P}(\rho(\mu, X) = \pi) = 0$ , then  $\mathbb{E}V = 0$ .*

Proposition 3.3 states that the logarithm map centers the observations around the intrinsic mean on  $\mathcal{S}^{\infty}$ , a property that has been established on finite-dimensional Riemannian manifolds (Karcher, 1977; Bhattacharya and Patrangenaru, 2003).

The tangent space  $T_{\mu}\mathcal{S}^{\infty}$  at  $\mu$  is a Hilbert space with inner product given by the metric tensor. We obtain first a CLT of  $\hat{\mu}$  mapped to the tangent space  $T_{\mu}\mathcal{S}^{\infty}$  as a corollary of Theorem 3.2 by considering  $\tau = \log_{\mu}(\cdot)$ , in which case  $\psi_{\tau}(X, \mu_{\tau}) = -2\langle V, \cdot \rangle_{\mu}$  as will be shown in the proof of Proposition 3.3. Here  $\Lambda_{\tau}$  becomes  $\Lambda = \mathbb{E}[D_2^2 \rho^2(X, \mu)] \in \mathcal{B}(T_{\mu}\mathcal{S}^{\infty}, \mathcal{B}(T_{\mu}\mathcal{S}^{\infty}, \mathbb{R}))$ , identified with the  $T_{\mu}\mathcal{S}^{\infty}$ -valued linear map  $\Lambda_1 \in \mathcal{B}(T_{\mu}\mathcal{S}^{\infty}, T_{\mu}\mathcal{S}^{\infty})$ , so that  $\Lambda v = \langle \Lambda_1 v, \cdot \rangle_{\mu}$  for  $v \in T_{\mu}\mathcal{S}^{\infty}$ .

**Corollary 3.1.** *Under the conditions of Theorem 3.2,  $\sqrt{n} \log_{\mu} \hat{\mu}$  converges in distribution to a Gaussian random element on  $T_{\mu}\mathcal{S}^{\infty}$  with mean zero and covariance  $4\Lambda_1^{-1} \mathbb{E}[V \otimes V](\Lambda_1^{-1})^*$ .*

The variation of  $X$  around the intrinsic mean is summarized by the covariance operator, which is a linear characterization of the intrinsic variation that has been applied to generalize the principal component analysis to Riemannian manifolds (Fletcher et al., 2004; Lazar and Lin, 2017). The population and empirical covariance operators are defined as  $\mathcal{G} : T_{\mu}\mathcal{S}^{\infty} \rightarrow T_{\mu}\mathcal{S}^{\infty}$ ,  $\mathcal{G} = \mathbb{E}[V \otimes V]$  and  $\hat{\mathcal{G}} : T_{\hat{\mu}}\mathcal{S}^{\infty} \rightarrow T_{\hat{\mu}}\mathcal{S}^{\infty}$ ,  $\hat{\mathcal{G}} = n^{-1} \sum_{i=1}^n \hat{V}_i \otimes \hat{V}_i$ , respectively, so

$$\mathcal{G}h = \mathbb{E}[\langle V, h \rangle_{\mu} V], \quad \hat{\mathcal{G}}g = \frac{1}{n} \sum_{i=1}^n \langle \hat{V}_i, g \rangle_{\hat{\mu}} \hat{V}_i$$

for  $h \in T_{\mu}\mathcal{S}^{\infty}$  and  $g \in T_{\hat{\mu}}\mathcal{S}^{\infty}$ .

The next result compares the estimated and true tangent vectors and will be needed to assess the estimation of the covariance operator. Parallel transport under the Levi-Civita connection on  $\mathcal{S}^\infty$  (Lang, 1999) is applied to compare tangent vectors on different tangent spaces in an intrinsic manner (Lin and Yao, 2019). Let  $P_x^y : T_x\mathcal{S}^\infty \rightarrow T_y\mathcal{S}^\infty$  denote the parallel transport of tangent vectors on  $T_x\mathcal{S}^\infty$  along the geodesic leaving from  $x$  to  $y$  on  $\mathcal{S}^\infty$ , given that  $x$  and  $y$  are not antipodal and thus the geodesic is unique. Write  $\Lambda_x = D_2^2\rho^2(x, \mu) \in \mathcal{B}(T_\mu\mathcal{S}^\infty, \mathcal{B}(T_\mu\mathcal{S}^\infty, \mathbb{R}))$  and we identify it with  $\Lambda_{x1} \in \mathcal{B}(T_\mu\mathcal{S}^\infty, T_\mu\mathcal{S}^\infty)$  such that  $\Lambda_x v = \langle \Lambda_{x1} v, \cdot \rangle_\mu$  for any  $v \in T_\mu\mathcal{S}^\infty$ . Recall that  $\Lambda_1$  is defined before Corollary 3.1.

**Corollary 3.2.** *If (A1) holds, then  $v = \log_\mu x$  and  $\hat{v} = \log_{\hat{\mu}} x$  are well defined for any  $x \in \mathcal{S}^\infty$  with  $\rho(x, \mu) < \pi$ , the latter almost surely as  $n \rightarrow \infty$ . Moreover, as  $n \rightarrow \infty$ ,*

$$\sqrt{n}(P_{\hat{\mu}}^\mu \hat{v} - v) \xrightarrow{L} Z_v,$$

where  $Z_v$  is a Gaussian random element in  $T_\mu\mathcal{S}^\infty$  with mean 0 and covariance  $\Lambda_{x1}\Lambda_1^{-1}\mathbb{E}[V \otimes V]\Lambda_1^{-1}\Lambda_{x1}$ .

To derive a central limit theorem for the covariance, operators  $\mathcal{G}$  and  $\hat{\mathcal{G}}$  are analyzed as Hilbert–Schmidt operators on the tangent spaces and are compared through the parallel transport. For  $x \in \mathcal{S}^\infty$  lying in a small enough neighborhood of  $\mu$ , let  $\mathcal{P}_x^\mu : \mathcal{B}(T_x\mathcal{S}^\infty, T_x\mathcal{S}^\infty) \rightarrow \mathcal{B}(T_\mu\mathcal{S}^\infty, T_\mu\mathcal{S}^\infty)$  be the parallel transport of operators from  $x$  to  $\mu$  such that for any operator  $\mathcal{A} \in \mathcal{B}(T_x\mathcal{S}^\infty, T_x\mathcal{S}^\infty)$ ,  $\mathcal{P}_x^\mu(\mathcal{A})v = P_x^\mu \mathcal{A} P_x^\mu v$  for  $v \in T_\mu\mathcal{S}^\infty$ . For a Hilbert space  $\mathbb{F}$ , let  $\mathcal{B}_{\text{HS}}(\mathbb{F}, \mathbb{F})$  denote the Hilbert space of Hilbert–Schmidt operators on  $\mathbb{F}$  equipped with the inner product

$$\langle \mathcal{F}_1, \mathcal{F}_2 \rangle_{\text{HS}} = \sum_{j=1}^{\infty} \langle \mathcal{F}_1 e_j, \mathcal{F}_2 e_j \rangle_{\mathbb{F}},$$

where  $\mathcal{F}_1, \mathcal{F}_2$  are operators on  $\mathbb{F}$  with finite Hilbert–Schmidt norm induced by this inner product, and  $\{e_j\}_{j=1}^{\infty}$  is a complete orthonormal basis of  $\mathbb{F}$ . For clarity, let  $\otimes_{\text{HS}}$  denote the tensor product of elements in  $\mathcal{B}_{\text{HS}}(T_\mu\mathcal{S}^\infty, T_\mu\mathcal{S}^\infty)$ . Define  $\mathcal{H} : T_\mu\mathcal{S}^\infty \rightarrow \mathcal{B}_{\text{HS}}(T_\mu\mathcal{S}^\infty, T_\mu\mathcal{S}^\infty)$ ,  $\mathcal{H}(\cdot) = \mathbb{E}[\Lambda_{X1}(\cdot) \otimes V]$  where  $\Lambda_{X1}$  is  $\Lambda_{x1}$  at a random  $x = X$ .

**Theorem 3.3.** *If (A1) holds, then as  $n \rightarrow \infty$ ,*

$$\sqrt{n}(\mathcal{P}_{\hat{\mu}}^\mu \hat{\mathcal{G}} - \mathcal{G}) \xrightarrow{L} Z_{\mathcal{G}}$$

in  $\mathcal{B}_{\text{HS}}(T_\mu\mathcal{S}^\infty, T_\mu\mathcal{S}^\infty)$ , where  $Z_{\mathcal{G}}$  is a Gaussian random element with mean 0 and covariance  $\mathbb{E}[\mathcal{G}_0 \otimes_{\text{HS}} \mathcal{G}_0]$ , where

$$\mathcal{G}_0 = V \otimes V - \mathcal{G} + \mathcal{H}\Lambda_1^{-1}V + (\mathcal{H}\Lambda_1^{-1}V)^*.$$

Theorem 3.3 extends the central limit theorem for the covariance operator of a Hilbert random element to a random element on the Hilbert sphere. On

a Hilbert sphere, additional terms involving  $\mathcal{H}$  appear in the random Hilbert–Schmidt operator  $\mathcal{G}_0$  that generates the asymptotic covariance due to the positive curvature of  $\mathcal{S}^\infty$ , as compared to the corresponding result (e.g. Theorem 8.1.2 in Hsing and Eubank, 2015) in a Hilbert space where the parallel transport is simply the identity and  $\mathcal{G}_0 = V \otimes V - \mathcal{G}$ . Theorem 3.3 can be applied to derive asymptotic theory for the principal component analysis (Tucker, Wu and Srivastava, 2013; Dai and Müller, 2018; Lin and Yao, 2019) based on  $\hat{\mathcal{G}}$ , an exposition beyond the scope of this work.

#### 4. Hypothesis tests for the intrinsic mean

##### 4.1. General setup

We obtain here one- and two-sample hypothesis tests of the intrinsic mean as a result of the CLT in Theorem 3.2; for completeness, a paired-sample test is formulated in Appendix G. To handle the manifold constraint, these tests make use of a chart  $(U, \tau)$  so that test statistics based on the intrinsic mean can be constructed analogously to those in a Hilbert space (Berkes et al., 2009; Aue, Rice and Sönmez, 2018).

##### One-sample test

Given a sample  $X_1, \dots, X_n$  on  $\mathcal{S}^\infty$  with unknown mean  $\mu$ , consider the one-sample hypothesis

$$H_0 : \mu = \mu_0, \quad H_1 : \mu \neq \mu_0,$$

for a pre-specified  $\mu_0 \in \mathcal{S}^\infty$ . We propose a norm-based and a projection-based test statistic, defined, respectively, as

$$T_1 = n \|\tau(\hat{\mu}) - \tau(\mu_0)\|_{\mathbb{G}}^2, \\ S_1 = n \sum_{k=1}^K \frac{\langle \tau(\hat{\mu}) - \tau(\mu_0), \hat{\phi}_k \rangle_{\mathbb{G}}^2}{\hat{\lambda}_k},$$

where the projection-based test utilizes  $K < \infty$  projections. Here  $S_1$  is the empirical estimate of  $\tilde{S}_1 = n \sum_{k=1}^K \lambda_k^{-1} \langle \tau(\hat{\mu}) - \tau(\mu_0), \phi_k \rangle_{\mathbb{G}}^2$ , and the  $(\lambda_k, \phi_k)$  and  $(\hat{\lambda}_k, \hat{\phi}_k)$  are, respectively, the eigenvalue–eigenfunction pairs of the true and estimated covariance operator  $\mathcal{T}$  or  $\hat{\mathcal{T}}$  of the limiting element  $Z$  in Theorem 3.2, respectively satisfying  $\mathcal{T}\phi_k = \lambda_k\phi_k$  and  $\hat{\mathcal{T}}\hat{\phi}_k = \hat{\lambda}_k\hat{\phi}_k$ . Note that for infinite-dimensional data a Hotelling’s  $T^2$ -like test statistic cannot be obtained by normalizing the mean with the inverse of its covariance operator, since the resulting test statistic diverges (Hájek, 1962). A natural choice of chart is  $\tau(\cdot) = \log_{\mu_0}(\cdot)$ , under which the norm-based test statistic reduces to  $T_1 = n\rho^2(\hat{\mu}, \mu_0)$ .

By Slutsky’s theorem and Theorem 3.2, the limiting distribution for  $T_1$  is  $\|Z\|_{\mathbb{G}}^2$  and that for  $S_1$  is  $\chi_K^2$ . Summarizing,

**Corollary 4.1.** *If the conditions of Theorem 3.2 hold, then as  $n \rightarrow \infty$ ,*

$$\begin{aligned} T_1 &\xrightarrow{L} \sum_{k=1}^{\infty} \lambda_k W_k, \\ S_1 &\xrightarrow{L} \chi_K^2, \end{aligned}$$

where  $W_k$  are i.i.d.  $\chi_1^2$  random variables.

### Two-sample test

The two-sample setting is that we have i.i.d. observations  $X_1^{[g]}, \dots, X_{n_g}^{[g]}$  on  $\mathcal{S}^\infty$  for  $g = 1, 2$ , where  $n_g$  is the sample size in Population  $g$  and  $n = n_1 + n_2$  is the total sample size. Let  $\mu_g$  be the unknown intrinsic mean in Population  $g$ , which is estimated by the empirical mean  $\hat{\mu}_g$ . The two-sample hypothesis is

$$H_0 : \mu_1 = \mu_2, \quad H_1 : \mu_1 \neq \mu_2.$$

Under chart  $\tau$ , let  $\mathcal{T}_g = (\Lambda_{g\tau}^{-1})\mathcal{F}_{g\tau}(\Lambda_{g\tau}^{-1})^*$  be the asymptotic covariance of  $\tau(\hat{\mu}_g)$  given by Theorem 3.2, where  $\mathcal{F}_{g\tau} = \mathbb{E}[\psi_\tau(X_i^{[g]}, \tau(\mu_g)) \otimes \psi_\tau(X_i^{[g]}, \tau(\mu_g))]$ ,  $\Lambda_{g\tau} = \mathbb{E}[D_2^2 \rho_\tau^2(X_i^{[g]}, \tau(\hat{\mu}_g))]$ ,  $g = 1, 2$ . Let  $\mathcal{T}_{\text{pool}} = (n/n_1)\mathcal{T}_1 + (n/n_2)\mathcal{T}_2$  be the pooled covariance operator, which has eigenvalue–eigenfunction pairs  $(\lambda_{k,\text{pool}}, \phi_{k,\text{pool}})$  satisfying  $\mathcal{T}_{\text{pool}}\phi_{k,\text{pool}} = \lambda_{k,\text{pool}}\phi_{k,\text{pool}}$ , for  $k = 1, 2, \dots$ . The following corollary of Theorem 3.2 serves as a basis for the two-sample tests.

**Corollary 4.2.** *Suppose that  $n_1, n_2 \rightarrow \infty$  and  $n_1/n_2 \rightarrow q$  for some  $q \in (0, 1)$ . If  $H_0$  is true and the conditions of Theorem 3.2 hold for both populations, then as  $n \rightarrow \infty$ ,*

$$\sqrt{n}[\tau(\hat{\mu}_1) - \tau(\hat{\mu}_2)] \xrightarrow{L} Z_2,$$

where  $Z_2$  is a Gaussian random element in  $\mathbb{G}$  with mean 0 and covariance  $\mathcal{T}_{\text{pool}}$ .

The norm-based and projection-based two-sample test statistics are

$$\begin{aligned} T_2 &= n \|\tau(\hat{\mu}_1) - \tau(\hat{\mu}_2)\|_{\mathbb{G}}^2, \\ S_2 &= n \sum_{k=1}^K \frac{\langle \tau(\hat{\mu}_1) - \tau(\hat{\mu}_2), \hat{\phi}_{k,\text{pool}} \rangle_{\mathbb{G}}^2}{\hat{\lambda}_{k,\text{pool}}}. \end{aligned}$$

Here  $(\hat{\lambda}_{k,\text{pool}}, \hat{\phi}_{k,\text{pool}})$ ,  $k = 1, 2, \dots$  are the eigenvalue–eigenfunction pairs of the estimated pooled covariance  $\hat{\mathcal{T}}_{\text{pool}} = (n/n_1)\hat{\mathcal{T}}_1 + (n/n_2)\hat{\mathcal{T}}_2$ , satisfying  $\hat{\mathcal{T}}_{\text{pool}}\hat{\phi}_{k,\text{pool}} = \hat{\lambda}_{k,\text{pool}}\hat{\phi}_{k,\text{pool}}$  for  $k = 1, 2, \dots$ ;  $\hat{\mathcal{T}}_g = (\hat{\Lambda}_{g\tau}^{-1})\hat{\mathcal{F}}_{g\tau}(\hat{\Lambda}_{g\tau}^{-1})^*$ ,  $\hat{\mathcal{F}}_{g\tau} = n_g^{-1} \sum_{i=1}^{n_g} \psi_\tau(X_i^{[g]}, \tau(\hat{\mu}_g)) \otimes \psi_\tau(X_i^{[g]}, \tau(\hat{\mu}_g))$ , and  $\hat{\Lambda}_{g\tau} = n_g^{-1} \sum_{i=1}^{n_g} D_2^2 \rho_\tau^2(X_i^{[g]}, \tau(\hat{\mu}_g))$ ,  $g = 1, 2$ .

**Corollary 4.3.** *Let  $W_1, W_2, \dots$  be i.i.d.  $\chi_1^2$  random variables. If the conditions of Corollary 4.2 hold, then as  $n \rightarrow \infty$ ,*

$$\begin{aligned} T_2 &\xrightarrow{L} \sum_{k=1}^{\infty} \lambda_{k,pool} W_k, \\ S_2 &\xrightarrow{L} \chi_K^2. \end{aligned}$$

#### 4.2. Asymptotic tests

In performing asymptotic tests, the asymptotic null distributions of the norm-based test statistics involve unknown eigenvalues that need to be estimated. A natural question is whether the asymptotic  $p$ -values can be estimated consistently in light of the infinite sum in the limiting distribution. The answer is affirmative: In the one-sample scenario, the asymptotic null distribution of  $T_1$  is  $\|Z\|_{\mathbb{G}}^2$ , consistently estimated by the distribution of  $\hat{T}_1 := \|Z_n\|_{\mathbb{G}}^2 = \sum_{k=1}^{\infty} \hat{\lambda}_k W_k$ . By Proposition 3.2 and the continuous mapping theorem, the limiting distributions of  $T_1$  and  $\hat{T}_1$  are the same, so the asymptotic cumulative distribution function and the quantiles/ $p$ -values of  $T_1$  can be estimated uniformly consistently by those of  $\hat{T}_1$  in the large sample limit. Although the asymptotic distribution is non-pivotal, Theorem 3.2 leads to an explicit form of the asymptotic distribution of the test statistics, enabling efficient implementation. In practice, quantiles of  $\hat{T}_1$  are obtained by Monte Carlo.

For the projection-based tests, the tuning parameter  $K$  needs to be chosen in order to maximally capture the difference in the means. While an optimal choice may depend on the sample size and the stochastic structure of the data, we find in our numerical studies that following the Fraction of Variance Explained (FVE) criterion leads to reasonable test performance; namely, set  $K = K^*$  where

$$K^* = \min\{K \geq 1 \mid \frac{\sum_{j=1}^K \hat{\lambda}_j}{\sum_{k=1}^{\infty} \hat{\lambda}_k} \geq r\}, \quad (3)$$

with a given threshold  $r \in (0, 1)$ . The projection-based tests will be powerful as long as the mean difference (on the chart) is not orthogonal to the subspace spanned by the first  $K$  eigenfunctions. Our experience, in agreement with Berkes et al. (2009); Horváth, Kokoszka and Reeder (2013), suggests that in practice the mean difference is usually well captured by the first few projections. Even though the norm-based tests always capture the mean difference, the projection-based tests can often be more powerful in our numerical studies since they utilize a pivotal test statistic and focus on the leading components with higher signal-to-noise ratio. The asymptotic tests in the two- and paired-sample scenarios follow analogous development.

#### 4.3. Bootstrap tests

Alternative bootstrap tests are proposed to improve finite-sample performance. The bootstrap tests avoid the computation of the covariance components of the

norm-based test statistics and enjoy a second order accuracy when using the pivotal projection-based test statistics (Hall, 1992).

In the one-sample scenario, let  $X_1^*, \dots, X_n^*$  be a nonparametric bootstrap sample drawn from  $X_1, \dots, X_n$  with replacement and  $\hat{\mu}^*$  be the bootstrap sample intrinsic mean. The bootstrap distributions are derived from

$$T_1^* = n \|\tau(\hat{\mu}^*) - \tau(\hat{\mu})\|_{\mathbb{G}}^2,$$

$$S_1^* = n \sum_{k=1}^K \frac{\langle \tau(\hat{\mu}^*) - \tau(\hat{\mu}), \hat{\phi}_k^* \rangle_{\mathbb{G}}^2}{\hat{\lambda}_k^*},$$

where the  $(\hat{\lambda}_k^*, \hat{\phi}_k^*)$  are the eigenpairs of  $\hat{\mathcal{T}}_{\tau}^*$  constructed analogously to the sample covariance  $\hat{\mathcal{T}}_{\tau}$  but with the bootstrap sample. The validity of the bootstrap test is guaranteed by the following corollary of Theorem 3.2 and the bootstrap theorem for infinite-dimensional  $Z$ -estimators by Wellner and Zhan (1996).

**Corollary 4.4.** *Under the conditions of Theorem 3.2, the bootstrap intrinsic mean  $\hat{\mu}^*$  is consistent for  $\mu$ . Furthermore, as  $n \rightarrow \infty$ ,*

$$\sqrt{n}[\tau(\hat{\mu}^*) - \tau(\hat{\mu})] \xrightarrow{L} Z^*,$$

where  $Z^*$  is a Gaussian random element sharing the same distribution with  $Z$  in Theorem 3.2. Hence, the asymptotic distributions are the same for  $T_1^*$  and  $T_1$ , as well as for  $S_1^*$  and  $S_1$ .

For the two-sample test, let  $X_1^{[g]*}, \dots, X_{n_g}^{[g]*}$  be a nonparametric bootstrap sample of size  $n_g$  from Population  $g = 1, 2$ . Let  $\hat{\mu}_g^*$  be the intrinsic mean and  $\hat{\mathcal{T}}_g^*$  the sample covariance operator of the bootstrap sample in Population  $g$ ,  $\hat{\mathcal{T}}_{\text{pool}}^* = (n/n_1)\hat{\mathcal{T}}_1^* + (n/n_2)\hat{\mathcal{T}}_2^*$  the bootstrap pooled covariance, and  $(\hat{\lambda}_{k,\text{pool}}^*, \hat{\phi}_{k,\text{pool}}^*)$ ,  $k = 1, 2, \dots$  the eigenpairs of  $\hat{\mathcal{T}}_{\text{pool}}^*$ . The bootstrap versions of for  $T_2$  and  $S_2$  are, respectively,

$$T_2^* = n \|\tau(\hat{\mu}_1^*) - \tau(\hat{\mu}_1) - [\tau(\hat{\mu}_2^*) - \tau(\hat{\mu}_2)]\|_{\mathbb{G}}^2,$$

$$S_2^* = n \sum_{k=1}^K \frac{\langle [\tau(\hat{\mu}_1^*) - \tau(\hat{\mu}_1)] - [\tau(\hat{\mu}_2^*) - \tau(\hat{\mu}_2)], \hat{\phi}_{k,\text{pool}}^* \rangle_{\mathbb{G}}^2}{\hat{\lambda}_{k,\text{pool}}^*},$$

where the bootstrap statistics are constructed as such to increase power.

**Corollary 4.5.** *Under the conditions of Corollary 4.2, the bootstrap intrinsic mean  $\hat{\mu}_g^*$  is consistent for  $\mu_g$ ,  $g = 1, 2$ . Moreover, as  $n \rightarrow \infty$ ,*

$$\sqrt{n} \{[\tau(\hat{\mu}_1^*) - \tau(\hat{\mu}_1)] - [\tau(\hat{\mu}_2^*) - \tau(\hat{\mu}_2)]\} \xrightarrow{L} Z_2^*,$$

where  $Z_2^*$  is a Gaussian random element sharing the same distribution as  $Z_2$  in Corollary 4.2. Hence, the asymptotic distributions are the same for  $T_2^*$  and  $T_2$ , as well as for  $S_2^*$  and  $S_2$ .

## 5. Simulation studies

Simulation studies were performed on  $\mathcal{S}^\infty \subset \mathbb{H} = L^2([0, 1])$  to demonstrate the numerical properties of the proposed hypothesis tests. We focus on the two-sample case here, while analogous results for the one-sample scenario are included in [Appendix H](#). In the two-sample scenario, we set the intrinsic population mean  $\mu_1$  to the square root of the Beta(2, 1) density function and  $\mu_2 = \exp_{\mu_1} \delta v$ , where  $\delta \in [-0.4, 0.4]$  is the effect size,  $v = K_\mu^{-1/2} \sum_{k=1}^{K_\mu} \phi_{1k}$ ,  $K_\mu \in \{1, 3, 5\}$  is the number of mean components, and the  $\phi_{1k}$  are orthonormal functions to be described shortly. Independent observations in Population  $g = 1, 2$  were generated according to  $X_i^{[g]}(s) = \exp_{\mu_g} \{(-1)^{g-1} \sum_{k=1}^{K_X} \xi_{gik} \phi_{gk}(s)\}$ ,  $s \in [0, 1]$  with  $K_X = 50$  components. The  $k$ th scores  $\xi_{gik}$  were i.i.d. real-valued random variables with mean 0 and variance  $\theta_k = 3^{-k}$ , generated from either the normal distribution or the centered exponential distribution, i.e.  $\xi_{gik} = \eta_{gik} - \mathbb{E}\eta_{gik}$  and  $\eta_{gik}$  follows  $\text{Exponential}(\theta_k)$  in the latter case, for  $i = 1, \dots, n_g$ ,  $g = 1, 2$ . The orthonormal basis functions  $\phi_{gk}$  were defined as  $\phi_{gk} = R_{\mu_g}(\psi_{k+1})$ ,  $k = 1, \dots, K_X$ , where  $\{\psi_j\}_{j=1}^\infty = \{\psi_1(s) = 1, \psi_{2k}(s) = 2^{1/2} \sin(2k\pi s), \psi_{2k-1}(s) = 2^{1/2} \cos(2(k-1)\pi s), s \in [0, 1], \text{ for } k \in \mathbb{N}\}$  is the trigonometric basis on  $[0, 1]$ , and  $R_q : \mathbb{H} \rightarrow \mathbb{H}$  is the rotation operator from  $\psi_1$  to  $q \neq -\psi_1$  along the shortest geodesic, defined by

$$R_q(p) = p + \sin(\rho_q)(\langle u, p \rangle_{\mathbb{H}} q - \langle q, p \rangle_{\mathbb{H}} u) + (\cos(\rho_q) - 1)(\langle q, p \rangle_{\mathbb{H}} q + \langle u, p \rangle_{\mathbb{H}} u),$$

where  $\rho_q = \rho(\psi_1, q)$  and  $u = (\psi_1 - \langle \psi_1, q \rangle_{\mathbb{H}} q) / (1 - \langle \psi_1, q \rangle_{\mathbb{H}}^2)^{1/2}$ . The mean squared geodesic distance from the observations to the intrinsic mean was  $\sum_{k=1}^{K_X} \theta_k = 0.5$ , and the cumulative FVE by the first  $J = 1, \dots, 5$  components, defined as  $\text{FVE}(J) = \sum_{j=1}^J \theta_j / \sum_{k=1}^{K_X} \theta_k$ , were 66.7%, 88.9%, 96.3%, 98.8%, and 99.6%, respectively.

The simulation settings consisted of all combinations of

- sample size  $n_g \in \{10, 25, 50\}$ ;
- number of components  $K_\mu \in \{1, 3, 5\}$  in the mean difference; and
- either symmetric or asymmetric data generated around the mean, corresponding to the normally (norm) or exponentially (exp) distributed  $\xi_{gik}$ , respectively.

We compared the proposed asymptotic and bootstrap tests which are intrinsic to  $\mathcal{S}^\infty$ , as well as a norm-based bootstrap test of the extrinsic means (Ellingson, Patrangenaru and Ruyngaert, 2013) in the ambient space  $\mathbb{H}$  projected back onto  $\mathcal{S}^\infty$ . The number of components  $K$  for our projection-based tests were chosen according to the FVE criterion (3) with threshold  $r = 0.8, 0.95, \text{ or } 0.99$ , which correspond roughly to  $K = 2, 3, \text{ and } 4$  projections in our settings, respectively.

With 1000 Monte Carlo iterations each with 499 bootstrap samples, the empirical power curves for the two-sample tests over effect size  $\delta \in [-0.4, 0.4]$  are displayed in [Figure 1](#), noting that  $H_0$  holds if and only if  $\delta = 0$ . As a visual aid, dark and light paired colors were used to denote the proposed asymptotic

and bootstrap tests, respectively. Bootstrap tests were overall more conservative and had better control of the type I error rate (size) than the asymptotic tests, which is most apparent for  $n_g = 10$  or 25. When  $n_g = 10$ , the asymptotic projection-based tests were over-liberal, while the corresponding bootstrap tests properly controlled the size to be around the nominal level  $\alpha = 0.05$ . This can be attributed to the second-order correctness for the bootstrap tests when the test statistic is pivotal (Hall, 1992), of which the effect is most prominent in small samples. The asymptotic and bootstrap tests had almost identical performance under  $n_g = 50$  (3rd and 6th columns, Figure 1), showing that the asymptotics comes into force under this moderate sample size even if the data lie on an infinite-dimensional curved manifold  $\mathcal{S}^\infty$ .

When  $n_g = 10$ , the norm-based tests had higher power than the projection-based tests while controlling the size near the nominal level. In this small-sample scenario, the norm-based tests avoid estimating the projection directions and gain stability as compared to the projection-based tests. With  $n_g = 25$  or 50, the projection-based tests were more powerful than the norm-based test when larger FVE thresholds were chosen to capture mean differences in multiple projections. Specifically, when  $K_\mu = 3$  and 5, the projection-based tests with FVE = 0.95 and 0.99 were the most powerful, respectively. This is due to the fact that the projection-based tests focus on the mean differences only in the directions with high signal-to-noise ratios. When  $K_\mu = 1$ , the norm-based tests and the projection-based tests with FVE = 0.80 had similar performance and were both among the best performers.

The extrinsic and intrinsic norm-based tests performed similarly in the symmetric scenarios (1st–3rd columns, Figure 1) but rather differently in the asymmetric scenarios (4th–6th columns, Figure 1). All tests suffered from finite-sample biases to different extents in the asymmetric scenarios, which is reflected by the lack of power when  $\delta$  was slightly below 0. The bias for the proposed intrinsic methods reduced as  $n_g$  increased in the asymmetric scenarios, reaching near-unbiasedness when  $n_g = 50$ , but the bias for the extrinsic tests remained even when  $n_g = 50$ . This underlines the importance of respecting the intrinsic geometry when data is asymmetrically generated on the manifold.

## 6. Data application: taxi demands

Understanding the demand in public transportation will provide key insights for more reliable and economic infrastructures. Modeling taxi demands has been of increasing interest (Chu and Chen, 2019; Dubey and Müller, 2020) as the popularity of app-based for-hire vehicle services such as Uber and Lyft increases and data become available. We analyzed the demand patterns of taxi and other for-hire vehicles in New York City, which were extracted from a total of 1.1 billion trips records of for-hire vehicles including the yellow and green cabs, Uber, Lyft, etc. The trip record data were made public following Freedom of Information Law (FOIL) requests, and our analysis was built upon a database compiled by Todd Schneider available on <https://github.com/toddwschneider/nyc-taxi-data>.

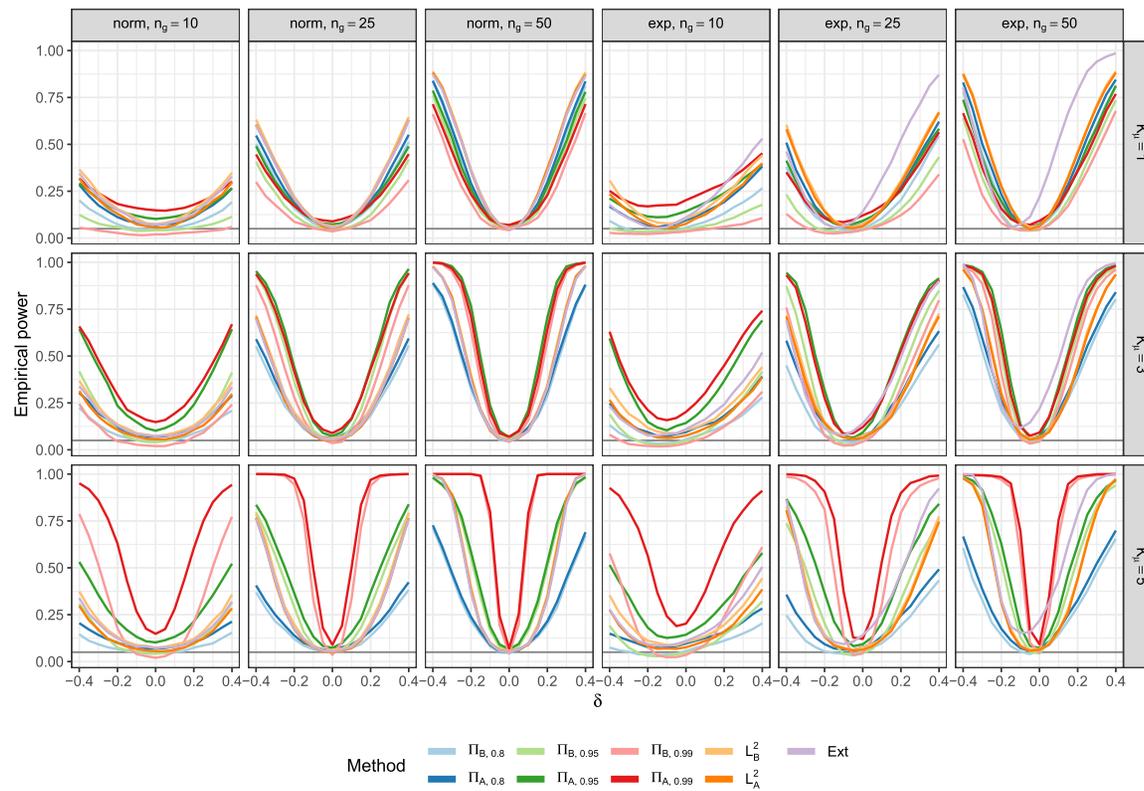


FIG 1. Empirical power curves for the two-sample tests. Columns correspond to different generating distributions for the principal components  $\xi_{gik}$  and sample sizes  $n_g$ , and rows correspond to different numbers of components  $K_\mu$  in the mean difference. The horizontal gray lines indicate the nominal level  $\alpha = 0.05$ .  $\Pi_{A,r}$  and  $\Pi_{B,r}$ , projection-based tests with FVE threshold  $r$ ;  $L_A^2$  and  $L_B^2$ , norm-based tests; *Ext*, the extrinsic bootstrap test. Subscripts *A* and *B* stand for the proposed tests in the asymptotic and bootstrap versions, respectively.

For the interest of monitoring evolving demand patterns, our goal is to compare the daily taxi demands in the year of 2016 and 2017. The demand pattern in the  $i$ th day of year  $g \in \{2016, 2017\}$  is modeled as the spatial density function  $Y_{gi}(s)$ ,  $s \in \mathcal{Z}$  of the locations where passengers were picked up, where  $\mathcal{Z}$  denotes the collection of taxi zones in Manhattan. To compare the daily demand patterns in 2016 and 2017 we performed two-sample tests with the daily pick-up densities  $Y_{gi}$  as the observational units. To apply the proposed two-sample tests on  $\mathcal{S}^\infty$ , density  $Y_{gi}$  is first transformed into square root density  $X_{gi}$ , obtaining  $X_{gi}(s) = \sqrt{Y_{gi}(s)}$ ,  $s \in \mathcal{S}^\infty$ . Since  $\|X_{gi}\|_{\mathbb{H}}^2 = \int_{\mathcal{Z}} X_{gi}^2(s) ds = \int_{\mathcal{Z}} Y_{gi}(s) ds = 1$ , the  $X_{gi}$  are objects lying on the unit Hilbert sphere  $\mathcal{S}^\infty$ . The proposed tests are then applied on these square root density objects. The average demands in 2016 and 2017 as measured by the intrinsic means of the square root densities  $X_{gi}$  within each year appear overall similar, as displayed in the left panels of Figure 2, but there is a substantial decrease in demand concentrated near the Upper East Side of Manhattan as shown in the right panel of Figure 2. The different is likely due to the opening of three Second Avenue subway stations on January 1, 2017.

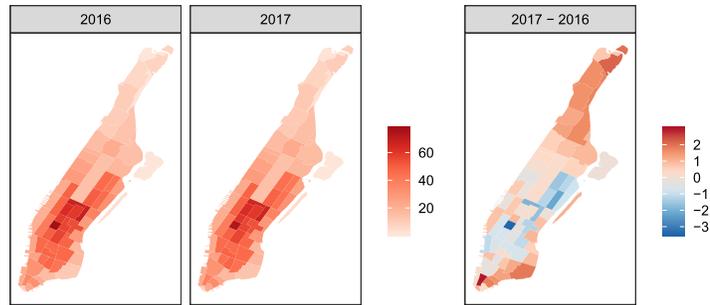


FIG 2. The sample intrinsic means and their (Euclidean) difference.

We next compared the power of two-sample hypothesis tests under the Hilbert sphere geometry and the flat geometry. In the former case the random objects were the square root densities  $X_{gi}$ , and in the latter the observations were the original densities  $Y_{gi}$ . Two scenarios were considered, namely the Equal Mean scenario where the random objects for both populations were sampled without replacement from year 2017, representing  $H_0$ , and the Unequal Mean scenario where the observations for the two populations were sampled from 2016 and 2017, respectively, representing  $H_1$ . The number of random samples  $n_g$  from each population varied among 10, 15, 20 and 25.

The proposed intrinsic tests and the extrinsic test based on the spherical geometry were applied on the square root densities  $X_{gi}$ , and an alternative norm-based bootstrap test was performed on the original densities  $Y_{gi}$ . The number of projections for the projection-based tests were selected according to the FVE criterion with threshold  $r = 0.8, 0.9$ , and  $0.95$ , which corresponds to  $K = 3, 5$ , and  $9$  components, respectively. The empirical sizes and powers are

reported in Table 1 for the nominal level  $\alpha = 0.05$ , calculated from 2000 Monte Carlo iterations and 999 bootstrap samples.

TABLE 1

Proportions of rejected  $H_0$  for the taxi demands at the nominal level  $\alpha = 0.05$  with  $n_g \in \{10, 15, 20, 25\}$  observations per population.  $\Pi_{A,r}$  and  $\Pi_{B,r}$ , projection-based tests with FVE threshold  $r$ ;  $L_A^2$  and  $L_B^2$ , norm-based tests; Ext, the extrinsic bootstrap test on  $\mathcal{S}^\infty$ ; Dens, the bootstrap test using the original densities. Subscripts A and B stand for the proposed asymptotic and bootstrap tests, respectively.

| Equal mean ( $H_0$ )   |         |               |               |                |         |               |               |                |       |       |  |
|------------------------|---------|---------------|---------------|----------------|---------|---------------|---------------|----------------|-------|-------|--|
| $n_g$                  | $L_A^2$ | $\Pi_{A,0.8}$ | $\Pi_{A,0.9}$ | $\Pi_{A,0.95}$ | $L_B^2$ | $\Pi_{B,0.8}$ | $\Pi_{B,0.9}$ | $\Pi_{B,0.95}$ | Ext   | Dens  |  |
| 10                     | 0.073   | 0.046         | 0.052         | 0.063          | 0.088   | 0.043         | 0.042         | 0.028          | 0.088 | 0.09  |  |
| 15                     | 0.062   | 0.043         | 0.04          | 0.05           | 0.078   | 0.044         | 0.038         | 0.026          | 0.076 | 0.076 |  |
| 20                     | 0.057   | 0.039         | 0.046         | 0.047          | 0.064   | 0.044         | 0.044         | 0.031          | 0.064 | 0.064 |  |
| 25                     | 0.044   | 0.035         | 0.041         | 0.04           | 0.048   | 0.038         | 0.034         | 0.026          | 0.05  | 0.052 |  |
| Unequal mean ( $H_1$ ) |         |               |               |                |         |               |               |                |       |       |  |
| $n_g$                  | $L_A^2$ | $\Pi_{A,0.8}$ | $\Pi_{A,0.9}$ | $\Pi_{A,0.95}$ | $L_B^2$ | $\Pi_{B,0.8}$ | $\Pi_{B,0.9}$ | $\Pi_{B,0.95}$ | Ext   | Dens  |  |
| 10                     | 0.203   | 0.215         | 0.596         | 0.846          | 0.259   | 0.215         | 0.572         | 0.776          | 0.259 | 0.124 |  |
| 15                     | 0.336   | 0.352         | 0.868         | 0.989          | 0.388   | 0.36          | 0.863         | 0.976          | 0.394 | 0.134 |  |
| 20                     | 0.532   | 0.484         | 0.975         | 0.999          | 0.591   | 0.5           | 0.972         | 0.999          | 0.59  | 0.148 |  |
| 25                     | 0.766   | 0.607         | 0.993         | 1              | 0.805   | 0.618         | 0.993         | 1              | 0.794 | 0.173 |  |

Under the Equal Mean scenario ( $H_0$ ), the proportion of rejection for all methods were below 0.1 and approached the nominal level  $\alpha = 0.05$  as  $n_g$  increased, indicating that the tests have approximately the correct size. The proposed norm-based bootstrap test  $L_B^2$  was slightly more liberal than its asymptotic version  $L_A^2$ , while the projection-based bootstrap test  $\Pi_{B,0.95}$  was slightly more conservative than its asymptotic version  $\Pi_{A,0.95}$ . In the Unequal Mean scenario ( $H_1$ ), all tests based on the square root densities (our proposed intrinsic tests and the extrinsic test) outperformed the bootstrap test based on the original density (last column, Table 1). This highlights that the spherical geometry on  $\mathcal{S}^\infty$  is more appropriate for detecting small changes in the demand distributions than a flat geometry for the original densities. The proposed projection-based tests with FVE threshold  $r = 0.9$  and  $0.95$  had the highest power for all sample sizes, outperforming the intrinsic and extrinsic norm-based bootstrap tests.

### Appendix A: Fréchet derivatives

The Fréchet derivative reviewed here follows Chapter I in Lang (1999). In this section, let  $E, E_j, F$  be Banach spaces with norms  $\|\cdot\|_E, \|\cdot\|_{E_j}$ , and  $\|\cdot\|_F$ , respectively, for  $j = 1, \dots, p$ . Denote  $\mathcal{B}(E, F)$  as the space of continuous linear maps from  $E$  into  $F$ , which is a Banach space equipped with the operator norm  $\|g\| = \sup_{\|e\|_E=1} \|g(e)\|_F$  for  $g \in \mathcal{B}(E, F)$ . Also let  $\mathcal{B}(E_1, \dots, E_p; F)$  denote the Banach space of multilinear maps equipped with the operator norm

$$\|h\| = \sup_{\|e_1\|_{E_1}=\dots=\|e_p\|_{E_p}=1} \|h(e_1, \dots, e_p)\|_F,$$

and write for short  $\mathcal{B}(E^p, F) = \mathcal{B}(E, \dots, E; F)$ . Repeated linear operator

$$g_{\text{rep}} \in \mathcal{B}(\underbrace{E, \mathcal{B}(E, \dots, \mathcal{B}(E, F) \dots)}_{p \text{ times}})$$

is isometrically identified with a multilinear map  $g_{\text{mult}} \in \mathcal{B}(E^p, F)$ , as

$$g_{\text{mult}}(e_1, \dots, e_p) = g_{\text{rep}}(e_1) \dots (e_p).$$

This identification gives rise to a Banach space isomorphism  $\mathcal{B}(E, \mathcal{B}(E, \dots, \mathcal{B}(E, F) \dots)) \simeq \mathcal{B}(E^p, F)$  (Proposition 2.4, p7, Lang, 1999); we use the same notation to denote both maps.

Let  $f : U \subset E \rightarrow F$  be a continuous map.

**Definition A.1.** Function  $f : U \subset E \rightarrow F$  is said to be (*Fréchet*) *differentiable* at a point  $x_0 \in U$  if there exists a continuous linear map  $l$  of  $E$  into  $F$  such that for  $y \in E$ ,

$$f(x_0 + y) = f(x_0) + l(y) + \epsilon(y),$$

where  $\|\epsilon(y)\|_F \rightarrow 0$  as  $\|y\|_E \rightarrow 0$ . The linear map  $l$  is called the (*Fréchet*) *derivative* of  $f$  at  $x_0$ , denoted as  $Df(x_0)$ . If  $f$  is differentiable at every point in  $U$ , then the derivative  $Df$  is a map

$$Df : U \rightarrow \mathcal{B}(E, F).$$

**Definition A.2.** Map  $f : U \subset E \rightarrow F$  is said to be *directional differentiable* at a point  $x_0 \in U$  if there exists a function  $l : E \rightarrow F$  such that

$$l(y) = \lim_{t \rightarrow 0} \frac{f(x_0 + ty) - f(x_0)}{t}$$

exists for all  $y \in E$ . The linear map  $l$  is called the *directional derivative* of  $f$  at  $x_0$ .

If a map  $f$  is Fréchet differentiable, then it is also directional differentiable and the two derivatives match. In what follows, “differentiability” refers to Fréchet differentiability unless otherwise noted, and the directional differentiation is used for calculation. Higher order derivatives and partial derivatives are defined in a recursive manner. Since the derivative  $Df(x_0)$  is in  $\mathcal{B}(E, F)$ , a Banach space, the *pth order derivative*  $D^p f$  is defined as  $D(D^{p-1}f)$ , a map of  $U$  into  $\mathcal{B}(E, \mathcal{B}(E, \dots, \mathcal{B}(E, F) \dots)) \simeq \mathcal{B}(E^p, F)$ . A map is said to be smooth if the derivatives of all orders exist. For a bivariate map  $h : E_1 \times E_2 \rightarrow F$ , the *partial derivative* with respect to the first argument at  $(x_0, y_0) \in U \times V \subset E_1 \times E_2$  is denoted as  $D_1 h(x_0, y_0)$ , where

$$D_1 h : U \times V \rightarrow \mathcal{B}(E_1, F), \quad D_1 h(x_0, y_0) = Df_{y_0}(x_0),$$

for  $f_{y_0}(x) = h(x, y_0)$ . The partial derivative  $D_2 h$  w.r.t. the second argument is similarly defined.

**Proposition A.1** (Proposition 3.1, page 9 in Lang (1999)). *If  $f : U \rightarrow V$  is differentiable at  $x_0$ , and  $g : V \rightarrow W$  is differentiable at  $f(x_0)$ , then  $g \circ f$  is differentiable at  $x_0$ , and*

$$D(g \circ f)(x_0) = Dg \circ f(x_0)(Df(x_0)).$$

For  $f : U \rightarrow \mathcal{B}(E, F)$  and  $g : U \rightarrow \mathcal{B}(F, G)$  defined on an open set  $U \subset E$ , denote  $f \cdot g$  as the function  $u \mapsto f(u) \circ g(u)$ . The chain rule can be compactly written as

$$D(g \circ f) = Dg \circ f \cdot Df.$$

## Appendix B: Proof of Theorem 3.1

*Proof.* We first show that for any  $q \in \mathcal{S}^\infty$  with  $\mathbb{P}(\rho(q, X) > \pi/2) > 0$ , there exists  $q' \in \mathcal{S}^\infty$  such that  $M(q') < M(q)$  and  $\mathbb{P}(\rho(q', X) \leq \pi/2) = 1$ . Denote  $B(b, r)$  as the open geodesic ball  $\{x \in \mathcal{S}^\infty \mid \rho(b, x) < r\}$  and  $\bar{B}(b, r)$  the closed geodesic ball  $\{x \in \mathcal{S}^\infty \mid \rho(b, x) \leq r\}$  centered at  $b \in \mathcal{S}^\infty$  with radius  $r > 0$ . Let  $q$  be a point on the sphere satisfying  $\mathbb{P}(\rho(q, X) > \pi/2) > 0$ . By (A1), there exists  $o$  in the support of  $X$  such that  $\rho(o, q) > \pi/2$  and  $\mathbb{P}(X \in B(o, \pi/2)) > 0$ . We follow the reflection argument used in the proof of Theorem 2.1 in Afsari (2011) to show that  $q$  cannot be a Fréchet mean. Due to the Gauss lemma (Theorem 5.7 in Lang, 1999), there exists a minimal geodesic from  $o$  to  $q$  which intersects perpendicularly the boundary of  $B(o, \pi/2)$  at a unique point  $q_c$ . Let  $q'$  be the reflection of  $q$  along the minimal geodesic into the boundary of  $B(o, \pi/2)$ . Let  $s$  be a point in  $\bar{B}(o, \pi/2)$  and  $\alpha$  the angle made by geodesic  $q_c o$  and  $q_c s$ . Any three points on  $\mathcal{S}^\infty$  and the geodesics between them lie on a 2-dimensional unit sphere, so the spherical law of cosines applies, which states that for any geodesic triangle with side lengths  $a, b, c$  and angle  $C$  for the corner opposite  $c$ , one has

$$\cos c = \cos a \cos b + \sin a \sin b \cos C. \quad (4)$$

By (4),  $\alpha \leq \pi/2$  for all  $s \in \bar{B}(o, \pi/2)$ , and thus two applications of (4) on geodesic triangles  $\Delta q q_c s$  and  $\Delta q' q_c s$  yield

$$\rho(q', s) \leq \rho(q, s),$$

where the inequality is strict if and only if  $s$  lies in the interior of  $\bar{B}(o, \pi/2)$ . Thus  $M(q') < M(q)$  by construction. Hence, if an intrinsic mean  $\mu$  exists, it must satisfy  $\mathbb{P}(\rho(\mu, X) \leq \pi/2) = 1$ .

We next show that if  $\mu$  is a minimizer of  $M(\cdot)$  on  $\mathcal{S}^\infty$ , then  $\mathbb{P}(\rho(\mu, X) = \pi/2) = 0$ . Suppose not, then there exists  $p$  in the support of  $X$  with  $\rho(\mu, p) = \pi/2$ . By (A1),  $p$  can be taken such that  $\mathbb{P}(\rho(p, X) < \pi/2) > 0$ , and  $\mathbb{P}(\rho(p, X) \leq \pi/2) = 1$ . Let  $t \mapsto \gamma(t)$  be a unit-speed minimal geodesic leaving from  $\mu$  at  $t = 0$  towards  $p$ . For any  $s \in \bar{B}(\mu, \pi/2)$ ,

$$\left. \frac{d}{dt} \rho^2(\gamma(t), s) \right|_{t=0} = -2\rho(\mu, s) \cos \alpha_s,$$

where  $\alpha_s$  is the angle made by minimal geodesics  $\mu p$  and  $\mu s$ . Arbitrarily define  $\cos \alpha_s = 0$  if  $s = \mu$ , noting that the former display is immaterial of this definition. Since  $\mu$  minimizes  $M(\cdot)$ , by the dominated convergence theorem (DCT) we have

$$0 = \left. \frac{d}{dt} M(\gamma(t)) \right|_{t=0} = -2\mathbb{E}[\rho(\mu, X) \cos \alpha_X]. \tag{5}$$

However, the construction of  $p$  and the spherical law of cosines imply  $\mathbb{P}(\cos \alpha_X \geq 0) = 1$  and  $\mathbb{P}(\cos \alpha_X > 0) > 0$ . Since  $\rho(\mu, X)$  is positive when  $\cos \alpha_X > 0$ , this leads to a contradiction to (5). Therefore, a necessary condition for  $\mu$  to be a minimizer of  $M(\cdot)$  is that

$$\mathbb{P}(\rho(\mu, X) < \pi/2) = 1. \tag{6}$$

We now prove the existence of a minimizer of  $M(\cdot)$  on  $\mathcal{S}^\infty$ . Let  $\{p_n\}_{n=1}^\infty$  be a sequence of points on  $\mathcal{S}^\infty$  such that  $\lim_{n \rightarrow \infty} M(p_n) = \inf_{p \in \mathcal{S}^\infty} M(p)$ . By the reflection argument, we can WLOG require  $\mathbb{P}(\rho(p_n, X) \leq \pi/2) = 1$ . Expand the domain of  $M(\cdot) = \mathbb{E} \arccos^2 \langle X, \cdot \rangle_{\mathbb{H}}$  to the closed unit ball  $B_{\mathbb{H}}(1) = \{p \in \mathbb{H} \mid \|p\|_{\mathbb{H}} \leq 1\}$  in  $\mathbb{H}$ . The Banach–Alaoglu theorem (e.g., Theorem 3.15 in Rudin, 1973) states that the closed unit ball in the Hilbert space is weakly compact, so any sequence  $\{a_n\}_{n=1}^\infty \subset \mathbb{H}$  with  $\|a_n\|_{\mathbb{H}} \leq 1$  has a subsequence  $a_{n_k}$  which converges to its weak limit  $a \in \mathbb{H}$  with  $\|a\|_{\mathbb{H}} \leq 1$ , satisfying  $\lim_{k \rightarrow \infty} \langle a_{n_k}, f \rangle_{\mathbb{H}} = \langle a, f \rangle_{\mathbb{H}}$  for all  $f \in \mathbb{H}$ . Let  $p_{n_k}$  be a subsequence of  $p_n$  converging weakly to  $p_\infty \in \mathbb{H}$ . By the weak continuity of  $M(\cdot)$  and the DCT,  $M(p_\infty) = \lim_{k \rightarrow \infty} M(p_{n_k}) = \inf_{p \in \mathcal{S}^\infty} M(p)$ , so for  $p_\infty$  to be a Fréchet mean it remains to be shown that  $p_\infty$  lies on  $\mathcal{S}^\infty$ . The Portmanteau theorem implies that

$$\mathbb{P}(\langle p_\infty, X \rangle_{\mathbb{H}} \geq 0) = \mathbb{P}(\lim_{k \rightarrow \infty} \langle p_{n_k}, X \rangle_{\mathbb{H}} \geq 0) \geq \limsup_{k \rightarrow \infty} \mathbb{P}(\rho(p_{n_k}, X) \leq \pi/2) = 1. \tag{7}$$

There exists  $o_1 \in \mathcal{S}^\infty$  such that  $\mathbb{P}(\rho(o_1, X) < 1/2) > 0$  and  $\mathbb{P}(\rho(o_1, X) \leq \pi/2) = 1$ , so  $M(o_1) < M(0)$ . This implies that  $p_\infty \neq 0$ , and

$$\mathbb{P}(\langle p_\infty, X \rangle_{\mathbb{H}} > 0) > 0. \tag{8}$$

Take  $p^* = p_\infty / \|p_\infty\|_{\mathbb{H}} \in \mathcal{S}^\infty$ . Due to the monotonicity of  $\arccos(\cdot)$ , equation (7), and (8), if  $\|p_\infty\|_{\mathbb{H}} < 1$  then one must have  $M(p^*) < M(p_\infty)$ , contradicting to the minimizing property of  $p_\infty$ . Hence,  $p_\infty$  must have unit norm and thus is the strong limit of  $p_{n_k}$ , and  $\mu = p_\infty$  is a minimizer of  $M(\cdot)$  over  $\mathcal{S}^\infty$ .

Finally, we establish uniqueness. Suppose  $\mu_1 \neq \mu_2$  are both minimizers of  $M(\cdot)$  on  $\mathcal{S}^\infty$ . Let  $\tilde{\mu} = \lambda\mu_1 + (1 - \lambda)\mu_2$  for some  $\lambda \in (0, 1)$ , which has norm  $0 < \|\tilde{\mu}\|_{\mathbb{H}} < 1$  by (6). By linearity and (6),  $\langle \tilde{\mu}, X \rangle_{\mathbb{H}}$  is positive with probability one, and thus

$$\begin{aligned} M(\tilde{\mu} / \|\tilde{\mu}\|_{\mathbb{H}}) &< M(\tilde{\mu}) = \mathbb{E} \arccos^2 \langle \lambda\mu_1 + (1 - \lambda)\mu_2, X \rangle_{\mathbb{H}} \\ &\leq \lambda \mathbb{E} \arccos^2 \langle \mu_1, X \rangle_{\mathbb{H}} + (1 - \lambda) \mathbb{E} \arccos^2 \langle \mu_2, X \rangle_{\mathbb{H}} \\ &= \inf_{p \in \mathcal{S}^\infty} M(p) \end{aligned}$$

by the strict monotonicity and convexity of  $\arccos^2(\cdot)$ , leading to a contradiction. We conclude that  $\mu = p_\infty$  is the unique minimizer of  $M(\cdot)$  over  $\mathcal{S}^\infty$ .  $\square$

**Appendix C: Proofs of other results in Section 3.1**

*Proof of Proposition 3.1.* Obtain

$$\begin{aligned} & \sup_{p \in \mathcal{S}^\infty} |M_n(p) - M(p)| \\ &= \sup_{p \in \mathcal{S}^\infty} \left| \frac{1}{n} \sum_{i=1}^n \arccos^2 \langle X_i, p \rangle_{\mathbb{H}} - \mathbb{E} \arccos^2 \langle X, p \rangle_{\mathbb{H}} \right| \\ &\leq \sup_{\|p\|_{\mathbb{H}} \leq 1} \left| \frac{1}{n} \sum_{i=1}^n \arccos^2 \langle X_i, p \rangle_{\mathbb{H}} - \mathbb{E} \arccos^2 \langle X, p \rangle_{\mathbb{H}} \right| = o(1) \quad a.s., \end{aligned}$$

where the last equality is due to the preservation of Glivenko–Cantelli class (van der Vaart and Wellner, 2000) and that  $\sup_{\|p\|_{\mathbb{H}} \leq 1} |n^{-1} \sum_{i=1}^n \langle X_i, p \rangle_{\mathbb{H}} - \mathbb{E} \langle X_1, p \rangle_{\mathbb{H}}| = \|\bar{X} - \mathbb{E}X_1\|_{\mathbb{H}} = o(1)$  a.s. by the strong LLN in  $\mathbb{H}$  (Theorem 2.4, Bosq, 2000). For  $\epsilon > 0$ , let  $S_\epsilon = \{p \in \mathcal{S}^\infty \mid \rho(p, \mu) \geq \epsilon\}$ . The uniqueness of the minimizer  $\mu$  of  $M(\cdot)$  and the weak limit argument in Theorem 3.1 imply

$$\inf_{p \in S_\epsilon} M(p) - M(\mu) > 0$$

for any  $\epsilon > 0$ , showing the well-separatedness of the minimizer  $\mu$  of  $M(\cdot)$ . The consistency of the sample intrinsic mean  $\hat{\mu}$  for  $\mu$  then follows from applying standard arguments for  $M$ -estimators (Section 3.2 in van der Vaart and Wellner, 1996).  $\square$

*Proof of Theorem 3.2.* Under (A1),  $\rho_\tau^2(X, \cdot)$  is smooth on a neighborhood of  $\mu_\tau$  with probability one. Write

$$\begin{aligned} \psi_\tau : \mathcal{S}^\infty \times \mathbb{G} &\rightarrow \mathcal{B}(\mathbb{G}, \mathbb{R}), & \psi_\tau(x, e) &= D_2 \rho_\tau^2(x, e), \\ \Psi_\tau : \mathbb{G} &\rightarrow \mathcal{B}(\mathbb{G}, \mathbb{R}), & \Psi_\tau(e) &= \mathbb{E} \psi_\tau(X, e), \\ \Psi_{n,\tau} : \mathbb{G} &\rightarrow \mathcal{B}(\mathbb{G}, \mathbb{R}), & \Psi_{n,\tau}(e) &= \frac{1}{n} \sum_{i=1}^n \psi_\tau(X_i, e). \end{aligned}$$

Since  $\mu$  and  $\hat{\mu}$  are minimizers of  $M(\cdot)$  and  $M_n(\cdot)$ , respectively, we have  $\Psi_\tau(\mu_\tau) = \Psi_{n,\tau}(\hat{\mu}_\tau) = 0$ , recalling that  $\mu_\tau = \tau(\mu)$  and  $\hat{\mu}_\tau = \tau(\hat{\mu})$ . By Taylor’s theorem,

$$\begin{aligned} 0 &= \sqrt{n} \Psi_{n,\tau}(\hat{\mu}_\tau) \\ &= \sqrt{n} \Psi_{n,\tau}(\mu_\tau) + D\Psi_{n,\tau}(\mu_\tau)[\sqrt{n}(\hat{\mu}_\tau - \mu_\tau)] + R_n, \end{aligned} \tag{9}$$

where  $R_n$  is the remainder that satisfies

$$\|R_n\| \leq \|\sqrt{n}(\hat{\mu}_\tau - \mu_\tau)\|_{\mathbb{G}} \sup_u \|D\Psi_{n,\tau}(u) - D\Psi_{n,\tau}(\mu_\tau)\|. \tag{10}$$

Here  $D$  denotes the Fréchet derivative, the supremum is taken over all  $u$  lying between  $\hat{\mu}_\tau$  and  $\mu_\tau$ , the norm of  $R_n$  is the operator norm in  $\mathcal{B}(\mathbb{G}, \mathbb{R})$  (i.e. the dual norm), and the last norm is the operator norm in  $\mathcal{B}(\mathbb{G}, \mathcal{B}(\mathbb{G}, \mathbb{R}))$ . By (A1),

Proposition 3.1, and Lemma E.3 in Appendix E, the supremum term on the RHS of (10) is  $o_p(1)$  as  $n \rightarrow \infty$ . Thus

$$\|R_n\| = o_p(\|\sqrt{n}(\hat{\mu}_\tau - \mu_\tau)\|_{\mathbb{G}}). \tag{11}$$

By (9), (11), and the LLN applied to  $D\Psi_{n,\tau}(\mu_\tau)$ ,

$$(\Lambda_\tau + o_p(1))\sqrt{n}(\hat{\mu}_\tau - \mu_\tau) = -\sqrt{n}\Psi_{n,\tau}(\mu_\tau).$$

Apply this and the continuity of  $\Lambda_\tau^{-1}$  given by Lemma E.2 to find that

$$\sqrt{n}(\hat{\mu}_\tau - \mu_\tau) = -\Lambda_\tau^{-1}\sqrt{n}\Psi_{n,\tau}(\mu_\tau) + o_p(1). \tag{12}$$

By the central limit theorem (Theorem 2.7, Bosq, 2000),

$$\sqrt{n}\Psi_{n,\tau}(\mu_\tau) = \frac{1}{\sqrt{n}} \sum_{i=1}^n \psi_\tau(X_i, \mu_\tau) \xrightarrow{L} Z_0, \tag{13}$$

where  $Z_0$  is a zero-mean Gaussian random element in  $\mathcal{B}(\mathbb{G}, \mathbb{R})$  with covariance operator  $\mathcal{F}_\tau = \mathbb{E}[\psi_\tau(X, \mu_\tau) \otimes \psi_\tau(X, \mu_\tau)]$ . The CLT and the form of the covariance operator of  $\sqrt{n}(\hat{\mu}_\tau - \mu_\tau)$  follow from (12), (13), and the continuous mapping theorem.  $\square$

*Proof of Proposition 3.2.* Recall that  $\Lambda_\tau$  and  $\mathcal{F}_\tau$  are defined before Theorem 3.2, and  $\hat{\Lambda}_\tau$  and  $\hat{\mathcal{F}}_\tau$  are defined before Proposition 3.2. Let  $\mathcal{A} = (\Lambda_\tau^{-1})^* \Lambda_\tau^{-1}$  and  $\hat{\mathcal{A}} = (\hat{\Lambda}_\tau^{-1})^* \hat{\Lambda}_\tau^{-1}$ , and denote  $\text{tr}(\mathcal{C}) = \sum_{j=1}^n \langle \mathcal{C}e_j, e_j \rangle_{\mathbb{G}}$  as the trace of an operator  $\mathcal{C} \in \mathcal{B}(\mathbb{G}, \mathbb{G})$  in an arbitrary complete orthonormal basis  $\{e_j\}_{j=1}^\infty$  of  $\mathbb{G}$ . We will make use of the well-known facts that  $\text{tr}(\cdot)$  is linear,  $\text{tr}(\mathcal{C}\mathcal{D}) = \text{tr}(\mathcal{D}\mathcal{C})$  for operators  $\mathcal{C}$  and  $\mathcal{D}$  defined on the same Hilbert space, and that the trace norm  $\|\mathcal{C}\|_1$  equals the sum of the singular values of  $\mathcal{C}$  if the sum is finite. Then

$$\begin{aligned} \|\|\hat{\mathcal{F}}\|_1 - \|\mathcal{F}\|_1\| &= |\text{tr}(\hat{\mathcal{F}} - \mathcal{F})| = |\text{tr}(\hat{\mathcal{A}}\hat{\mathcal{F}}_\tau - \mathcal{A}\mathcal{F}_\tau)| \\ &= |\text{tr}((\hat{\mathcal{A}} - \mathcal{A})\mathcal{F}_\tau) + \text{tr}(\hat{\mathcal{A}}(\hat{\mathcal{F}}_\tau - \mathcal{F}_\tau))| \\ &\leq \|(\hat{\mathcal{A}} - \mathcal{A})\mathcal{F}_\tau\|_1 + \|\hat{\mathcal{A}}(\hat{\mathcal{F}}_\tau - \mathcal{F}_\tau)\|_1 \\ &\leq \|\hat{\mathcal{A}} - \mathcal{A}\| \|\mathcal{F}_\tau\|_1 + \|\hat{\mathcal{A}}\| \|\hat{\mathcal{F}}_\tau - \mathcal{F}_\tau\|_1, \end{aligned} \tag{14}$$

where the first and second inequalities are due to Theorem 3.1 and Theorem 1.6 in Simon (2005), respectively. By the strong LLN (Theorem 2.4, Bosq, 2000) and continuous mapping theorem,

$$\|\|\hat{\Lambda}_\tau^{-1} - \Lambda_\tau^{-1}\| = o(1) \quad a.s., \tag{15}$$

$$\|\|\hat{\mathcal{A}} - \mathcal{A}\| = o(1) \quad a.s., \tag{16}$$

$$\|\|\hat{\mathcal{A}}\| = O(1) \quad a.s., \tag{17}$$

where (15) is due to that the inversion operation is continuous at continuously invertible operator  $\Lambda_\tau$  (Proposition 3.9, p11, Lang, 1999). Since covariance operators are in the trace class (Section 4.5 in Hsing and Eubank, 2015),

$$\|\mathcal{F}_\tau\|_1 < \infty. \tag{18}$$

By the triangle inequality

$$\begin{aligned} \|\hat{\mathcal{F}}_\tau - \mathcal{F}_\tau\|_1 &\leq \left\| \frac{1}{n} \sum_{i=1}^n [\psi_\tau(X_i, \hat{\mu}_\tau) \otimes \psi_\tau(X_i, \hat{\mu}_\tau) - \psi_\tau(X_i, \mu_\tau) \otimes \psi_\tau(X_i, \mu_\tau)] \right\|_1 \\ &\quad + \left\| \frac{1}{n} \sum_{i=1}^n \psi_\tau(X_i, \mu_\tau) \otimes \psi_\tau(X_i, \mu_\tau) - \mathbb{E}[\psi_\tau(X, \mu_\tau) \otimes \psi_\tau(X, \mu_\tau)] \right\|_1. \end{aligned}$$

The second term in the upper bound is  $o(1)$  a.s. by the LLN in the Banach space of trace class operators in the trace norm (Theorem 2.7, Simon, 2005). The summand operator in the first term of the upper bound has a rank up to 2, so its trace norm is no larger than 2 times its operator norm. By the triangle inequality, Taylor’s theorem, Proposition 3.1, and Lemma E.3, the first term in the upper bound is also  $o(1)$  a.s. Thus

$$\|\hat{\mathcal{F}}_\tau - \mathcal{F}_\tau\|_1 = o(1) \quad a.s. \tag{19}$$

Combining (14) and (15)–(19) we have  $|\|\hat{\mathcal{S}}\|_1 - \|\mathcal{S}\|_1| = o(1)$  a.s. The conditions of Theorem 2.19 in Simon (2005) are thus verified, obtaining the first result. Theorem 2.7.21 in Bogachev (2018) then leads to a proof of the second statement.  $\square$

**Appendix D: Proofs of the results in Section 3.2**

*Proof of Proposition 3.3.* By the assumption and the spherical geometry, the tangent vector  $\log_\mu(X)$  is well defined with probability one. For any  $v \in T_\mu \mathcal{S}^\infty$ , let  $\gamma : (-\epsilon, \epsilon) \rightarrow \mathcal{S}^\infty$  be a geodesic starting at  $\gamma(0) = \mu$  with initial velocity  $\gamma'(t) = v$ . By the first variation formula (XI, §1, Lang, 1999), for any  $s \in \mathcal{S}^\infty$  with  $\rho(\mu, s) < \pi$ ,

$$\left. \frac{d}{dt} \rho^2(\gamma(t), s) \right|_{t=0} = -2\langle v, \log_\mu s \rangle_\mu. \tag{20}$$

Since  $\mu$  is a stationary point of  $M(\cdot) = \mathbb{E}\rho^2(\cdot, X)$ , by the DCT,

$$0 = \frac{d}{dt} M(\gamma(t)) = \left. \frac{d}{dt} \mathbb{E}\rho^2(\gamma(t), X) \right|_{t=0} = -2E\langle v, \log_\mu X \rangle_\mu.$$

Thus  $\mathbb{E}\log_\mu X = 0$ .  $\square$

*Proof of Corollary 3.2.* On a sufficiently small open neighborhood  $U$  of  $\mu$  such that  $U \subset \mathcal{S}^\infty \setminus \{-x\}$ , define  $g = (p \mapsto \log_p x)$  and  $f : U \rightarrow T_\mu \mathcal{S}^\infty$ ,  $f(p) = P_p^\mu g(p)$ , for  $p \in U$ . Viewing  $P_p^\mu$  as an operator in the Banach space  $\mathcal{B}(T_p \mathcal{S}^\infty, T_\mu \mathcal{S}^\infty)$ ,  $p \mapsto P_p^\mu$  is smooth due to the smooth dependency of the ordinary differential equations defining the parallel transport on the initial conditions. Combine this and the smoothness of  $\rho^2$  when the arguments are not antipodals we find that  $f$  is also smooth. Taylor’s theorem under a normal coordinate around  $\mu$  implies that

$$f(\tilde{\mu}) - f(\mu) = Df(\mu) \log_\mu \tilde{\mu} + O(\rho(\tilde{\mu}, \mu)^2) \tag{21}$$

as  $\rho(\tilde{\mu}, \mu) \rightarrow 0$ , where  $Df(\mu) : T_\mu \mathcal{S}^\infty \rightarrow T_\mu \mathcal{S}^\infty$  is the differential of  $f$  at  $\mu$  calculated along radial directions. For a tangent vector  $v \in T_\mu \mathcal{S}^\infty$ , let  $\alpha : [-\epsilon, \epsilon] \rightarrow \mathcal{S}^\infty$  be a unit-speed geodesic lying in  $U$  with  $\alpha(0) = \mu$  and  $\alpha'(0) = v$ , and denote  $f_\alpha = f \circ \alpha$  and  $g_\alpha = g \circ \alpha$ . By Proposition 5.1, p263 in Lang (1999),

$$f_\alpha(t) - f_\alpha(0) = P_{\alpha(t)}^{\alpha(0)} g_\alpha(t) - g_\alpha(0) = tD_{\alpha'} g_\alpha(0) + O(t^2), \tag{22}$$

where  $D_{\alpha'}$  is the derivative along the curve  $\alpha$ . Let  $\text{gr} F$  denote the gradient vector field of a function  $F : \mathcal{S}^\infty \rightarrow \mathbb{R}$  defined by  $\langle \text{gr} F(x), v \rangle_x = DF(x)v$  for  $v \in T_x \mathcal{S}^\infty$ , and analogously let  $\text{gr}_2$  be the gradient w.r.t. the second argument of a function. Match up the first derivative terms in (21) and (22) to obtain

$$Df(\mu)v = D_{\alpha'} g_\alpha(0).$$

For  $u \in T_\mu \mathcal{S}^\infty$ ,

$$\langle D_{\alpha'} g_\alpha(0), u \rangle_\mu = -\frac{1}{2} \langle D_2 \text{gr}_2 \rho^2(x, \mu)v, u \rangle_\mu = -\frac{1}{2} D_2^2 \rho^2(x, \mu)(v, u),$$

where the first and second equations are due to (20) and Theorem 1.1, p344 in Lang (1999), respectively. The previous two displays lead to  $Df(\mu) = \Lambda_{x1}$ , a self-adjoint operator. Expansion (21) applied to  $\tilde{\mu} = \hat{\mu}$  is then written as

$$\sqrt{n}(P_{\hat{\mu}}^\mu \hat{v} - v) = \sqrt{n} \Lambda_{x1} \log_\mu \hat{\mu} + o_p(1) = n^{-1/2} \Lambda_{x1} \Lambda_1^{-1} \sum_{i=1}^n V_i + o_p(1), \tag{23}$$

where the first and the second equations further apply Proposition 3.1 and the expansion (12) in the proof of Theorem 3.2, respectively. Finally, invoke Slutsky’s theorem and the Hilbert space CLT to conclude.  $\square$

The last  $o_p$  term in the proof of Corollary 3.2 is seen to be uniform in  $x$  when  $\rho(x, \mu) \leq c$  for some constant  $c < \pi$ .

*Proof of Theorem 3.3.* Denote  $\tilde{V}_i = P_{\hat{\mu}}^\mu \hat{V}_i \in T_\mu \mathcal{S}^\infty$ ,  $i = 1, \dots, n$ . Since the parallel transport is a linear isometry,  $\mathcal{P}_\mu^\mu \hat{\mathcal{G}} = n^{-1} \sum_{i=1}^n \mathcal{P}_\mu^\mu (\hat{V}_i \otimes \hat{V}_i) = n^{-1} \sum_{i=1}^n \tilde{V}_i \otimes \tilde{V}_i$ . Then

$$\sqrt{n}(\mathcal{P}_\mu^\mu \hat{\mathcal{G}} - \mathcal{G}) = \frac{1}{\sqrt{n}} \sum_{i=1}^n (V_i \otimes V_i - \mathcal{G})$$

$$\begin{aligned}
& + \frac{1}{\sqrt{n}} \sum_{i=1}^n (\tilde{V}_i - V_i) \otimes V_i + \frac{1}{\sqrt{n}} \sum_{i=1}^n V_i \otimes (\tilde{V}_i - V_i) \\
& + \frac{1}{\sqrt{n}} \sum_{i=1}^n (\tilde{V}_i - V_i) \otimes (\tilde{V}_i - V_i). \tag{24}
\end{aligned}$$

Let  $\Lambda_{X_i 1}$  denote the operator  $\Lambda_{x1}$  evaluated at  $x = X_i$  and  $\mathcal{H}_i(\cdot) = \Lambda_{X_i 1}(\cdot) \otimes V_i$  be a bounded linear operator in  $\mathcal{B}(T_\mu \mathcal{S}^\infty, \mathcal{B}_{\text{HS}}(T_\mu \mathcal{S}^\infty, T_\mu \mathcal{S}^\infty))$ ,  $i = 1, 2, \dots$ . By the one-term Taylor expansion (23),  $\sqrt{n}(\tilde{V}_i - V_i) = \mathcal{H}_i(\Lambda_1^{-1} n^{-1/2} \sum_{j=1}^n V_j) + o_p(1)$ , where the  $o_p(1)$  term is uniform over  $i = 1, \dots, n$ . Since  $\mathbb{E}\mathcal{H}_i = \mathcal{H}$  where  $\mathcal{H}$  is defined before Theorem 3.3, the second term in (24) equals

$$\left( \frac{1}{n} \sum_{i=1}^n \mathcal{H}_i \right) \left( \frac{1}{\sqrt{n}} \sum_{j=1}^n \Lambda_1^{-1} V_j \right) + o_p(1) = \mathcal{H} \left( \frac{1}{\sqrt{n}} \sum_{j=1}^n \Lambda_1^{-1} V_j \right) + o_p(1)$$

upon applying the LLN. The same argument implies that the third and the last terms in (24) are  $[\mathcal{H}(n^{-1/2} \sum_{j=1}^n \Lambda_1^{-1} V_j)]^* + o_p(1)$  and  $o_p(1)$ , respectively. Thus, (24) becomes

$$\sqrt{n} \mathcal{P}_\mu^\mu(\hat{\mathcal{G}} - \mathcal{G}) = \frac{1}{\sqrt{n}} \sum_{i=1}^n [V_i \otimes V_i - \mathcal{G} + \mathcal{H} \Lambda_1^{-1} V_i + (\mathcal{H} \Lambda_1^{-1} V_i)^*] + o_p(1),$$

where the term inside the bracket is a random element in the Hilbert space  $\mathcal{B}_{\text{HS}}(T_\mu \mathcal{S}^\infty, T_\mu \mathcal{S}^\infty)$  with mean zero by Proposition 3.3 and finite second moment by the boundedness of  $\mathcal{S}^\infty$ . The proof concludes with an application of the CLT and Slutsky's theorem.  $\square$

## Appendix E: Differentiating the squared geodesic distance

For a smooth function  $f : \mathcal{S}^\infty \rightarrow \mathbb{R}$  defined on the manifold, let  $Df$  denote its differential, which maps  $p \in \mathcal{S}^\infty$  to a linear functional in  $\mathcal{B}(T_p \mathcal{S}^\infty, \mathbb{R})$  according to  $Df(p)v = d/dt f(\exp_p tv)|_{t=0}$  for  $v \in T_p \mathcal{S}^\infty$ . The Hessian  $D^2f$  of  $f$  maps  $p \in \mathcal{S}^\infty$  to a linear operator in  $\mathcal{B}(T_p \mathcal{S}^\infty, \mathcal{B}(T_p \mathcal{S}^\infty, \mathbb{R}))$  identified with a bilinear operator in  $\mathcal{B}(T_p^2 \mathcal{S}^\infty, \mathbb{R})$ . As a bilinear operator,  $D^2f(p)$  is symmetric and is specified by the ‘‘diagonal’’ values  $D^2f(p)(v, v) = d^2/dt^2 f(\exp_p tv)|_{t=0}$  (p344, Lang, 1999),  $v \in T_p \mathcal{S}^\infty$ . The differential and Hessian of  $\rho^2$  w.r.t. the second parameter are, for  $p, q \in \mathcal{S}^\infty \subset \mathbb{H}$  and  $h \in T_q \mathcal{S}^\infty$ ,

$$\begin{aligned}
D_2 \rho^2(p, q)h &= \left. \frac{d}{dt} \rho^2(p, \exp_q th) \right|_{t=0} = -2 \frac{\arccos(\langle p, q \rangle_{\mathbb{H}}) \langle h, p \rangle_{\mathbb{H}}}{\sqrt{1 - \langle p, q \rangle_{\mathbb{H}}^2}}, \\
D_2^2 \rho^2(p, q)(h, h) &= \left. \frac{d^2}{dt^2} \rho^2(p, \exp_q th) \right|_{t=0} \\
&= \frac{-2 \arccos(\langle p, q \rangle_{\mathbb{H}}) \|h\|_{\mathbb{H}}^2 \langle p, q \rangle_{\mathbb{H}} (-1 + \langle p, q \rangle_{\mathbb{H}}^2)}{(1 - \langle p, q \rangle_{\mathbb{H}}^2)^{3/2}}
\end{aligned}$$

$$\begin{aligned}
 & \frac{2\langle h, p \rangle_{\mathbb{H}}^2 \left( \arccos(\langle p, q \rangle_{\mathbb{H}}) \langle p, q \rangle_{\mathbb{H}} - \sqrt{1 - \langle p, q \rangle_{\mathbb{H}}^2} \right)}{(1 - \langle p, q \rangle_{\mathbb{H}}^2)^{3/2}} \\
 & =: F(p, q)(h, h), \tag{25}
 \end{aligned}$$

where  $F$  is maps  $(p, q) \in \mathcal{S}^\infty \times \mathcal{S}^\infty$ ,  $\rho(p, q) < \pi$  to a symmetric bilinear operator in  $\mathcal{B}(T_q^2 \mathcal{S}^\infty, \mathbb{R})$ , and the quotients are defined through the limit if  $\langle p, q \rangle_{\mathbb{H}} = 1$ .

Let  $\tau : U \subset \mathcal{S}^\infty \rightarrow \mathbb{G}$  be a smooth chart with  $q \in U$ . For a bivariate function  $g$ , denote  $g \circ_2 f$  as  $g(\cdot_1, f(\cdot_2))$ , so  $\rho_\tau = \rho \circ_2 \tau^{-1}$ . Applying the chain rule twice, for  $q_\tau = \tau(q)$  and  $u, v \in T_q \mathcal{S}^\infty$ ,

$$D_2^2 \rho_\tau^2(p, q_\tau) = D_2^2 \rho^2(p, q) \cdot D\tau^{-1}(q_\tau) \cdot D\tau^{-1}(q_\tau) + D_2 \rho^2(p, q) \cdot D^2 \tau^{-1}(q_\tau), \tag{26}$$

$$\begin{aligned}
 D_2^2 \rho_\tau^2(p, q_\tau)(u, v) &= D_2^2 \rho^2(p, q)(D\tau^{-1}(q_\tau)u, D\tau^{-1}(q_\tau)v) \\
 &+ D_2 \rho^2(p, q) D^2 \tau^{-1}(q_\tau)(u, v). \tag{27}
 \end{aligned}$$

**Lemma E.1.** *For all  $p, q \in \mathcal{S}^\infty$  with  $\rho(p, q) \leq \pi/2$ , there exists  $c \geq 0$  depending on  $\rho(p, q)$  and  $C > 0$  not involving  $(p, q)$  such that  $C \|h\|_{\mathbb{H}}^2 \geq F(p, q)(h, h) \geq c \|h\|_{\mathbb{H}}^2$ . If  $\rho(p, q) < \pi/2$  then  $c$  can be chosen to be strictly positive.*

*Proof.* We analyze  $F(p, q)$  defined in (25). Since  $F(p, q)$  is bilinear, WLOG suppose  $\|h\|_{\mathbb{H}} = 1$ . Let  $a = \langle p, q \rangle_{\mathbb{H}}$  and  $b = \langle h, p \rangle_{\mathbb{H}}$ , which are subject to  $0 \leq a \leq 1$  and  $-1 \leq b \leq 1$  under the assumptions. Rewrite

$$F(p, q)(h, h) = \frac{2a \arccos(a)}{(1 - a^2)^{1/2}} + \frac{2b^2 (-a \arccos(a) + \sqrt{1 - a^2})}{(1 - a^2)^{3/2}} =: G(a, b).$$

Function  $G(a, b)$  is quadratic in  $b$  and the quadratic coefficient is nonnegative, so

$$G(a, b) \geq G(a, 0) = \frac{2a \arccos(a)}{(1 - a^2)^{1/2}},$$

where the RHS is strictly increasing in  $a$ . Thus, constant  $c$  can be taken as  $G(\cos(\rho(p, q)), 0)$ , which is nonnegative when  $\rho(p, q) \leq \pi/2$  and positive when  $\rho(p, q) < \pi/2$ . Also,

$$G(a, b) \leq G(a, 1) \leq G(1, 1) = 8/3,$$

since  $G(a, 1)$  is increasing in  $a$  for  $0 \leq a \leq 1$ , so  $C$  can be set as  $8/3$ . □

**Lemma E.2.** *Under the conditions of Theorem 3.2,  $\Lambda_\tau$  is continuously invertible.*

*Proof.* First view  $\Lambda_\tau$  as a bilinear operator in  $\mathcal{B}(\mathbb{G}^2, \mathbb{R})$ . Plug in  $X$  for  $p$  and  $\mu_\tau$  for  $q_\tau$  and take expected value on both sides of (26),

$$\Lambda_\tau = \mathbb{E} D_2^2 \rho^2(X, \mu) \cdot D\tau^{-1}(\mu_\tau) \cdot D\tau^{-1}(\mu_\tau) + \mathbb{E} D_2 \rho^2(X, \mu) \cdot D^2 \tau^{-1}(\mu_\tau). \tag{28}$$

The second term in the last equality vanishes since

$$\mathbb{E}D_2\rho^2(X, \mu) = DM(\mu) = 0, \tag{29}$$

by the minimizing property of  $\mu$  and the DCT. By [Theorem 3.1](#) and [Lemma E.1](#),

$$\mathbb{E}D_2^2\rho^2(X, \mu)(h, h) \geq c_0 \|h\|_\mu^2 \tag{30}$$

for some constant  $c_0 > 0$ . By the smoothness of  $\tau^{-1}$ , there is a constant  $c_1 > 0$  such that for any  $v \in \mathbb{G}$

$$\|D\tau^{-1}(\mu_\tau)v\|_\mu \geq c_1 \|v\|_\mathbb{G}. \tag{31}$$

Combine (28)–(31) to obtain

$$\inf_{\|v\|_\mathbb{G}=1} \Lambda_\tau(v, v) > 0.$$

Now work on the repeated linear operator  $\Lambda_\tau$  in  $\mathcal{B}(\mathbb{G}, \mathcal{B}(\mathbb{G}, \mathbb{R}))$ . Identifying  $\mathcal{B}(\mathbb{G}, \mathbb{R})$  with  $\mathbb{G}$ , our conclusion drops out by applying [Theorem 3.5.4](#) and [Theorem 3.5.1](#) in [Hsing and Eubank \(2015\)](#).  $\square$

**Lemma E.3.** *Under the conditions of [Theorem 3.2](#),*

$$\sup_u \sup_{\rho_\tau(x, \mu_\tau) \leq \pi/2} \|D_2^2\rho_\tau^2(x, u) - D_2^2\rho_\tau^2(x, \mu_\tau)\| = o(1) \quad a.s.$$

as  $n \rightarrow \infty$ , where the first supremum is taken over all points  $u \in \mathbb{G}$  lying between  $\hat{\mu}_\tau$  and  $\mu_\tau$ .

*Proof.* By the chain rule as in (26) and (27),

$$\begin{aligned} & D_2^2\rho_\tau^2(x, u) - D_2^2\rho_\tau^2(x, \mu_\tau) \\ &= [D_2^2\rho^2(x, \tau^{-1}(u)) \cdot D\tau^{-1}(u) \cdot D\tau^{-1}(u) - D_2^2\rho^2(x, \mu) \cdot D^{-1}\tau(\mu_\tau) \cdot D\tau^{-1}(\mu_\tau)] \\ &\quad + [D_2\rho^2(x, \tau^{-1}(u)) \cdot D^2\tau^{-1}(u) - D_2\rho^2(x, \mu_\tau) \cdot D^2\tau^{-1}(\mu_\tau)] \\ &= [D_2^2\rho^2(x, \tau^{-1}(u)) \cdot D\tau^{-1}(u) \cdot D\tau^{-1}(u) - D_2^2\rho^2(x, \mu) \cdot D\tau^{-1}(u) \cdot D\tau^{-1}(u)] \\ &\quad + [D_2^2\rho^2(x, \mu) \cdot D\tau^{-1}(u) \cdot D\tau^{-1}(u) - D_2^2\rho^2(x, \mu) \cdot D\tau^{-1}(u) \cdot D\tau^{-1}(\mu_\tau)] \\ &\quad + [D_2^2\rho^2(x, \mu) \cdot D\tau^{-1}(u) \cdot D\tau^{-1}(\mu_\tau) - D_2^2\rho^2(x, \mu) \cdot D\tau^{-1}(\mu_\tau) \cdot D\tau^{-1}(\mu_\tau)] \\ &\quad + [D_2\rho^2(x, \tau^{-1}(u)) \cdot D^2\tau^{-1}(u) - D_2\rho^2(x, \mu_\tau) \cdot D^2\tau^{-1}(u)] \\ &\quad + [D_2\rho^2(x, \mu_\tau) \cdot D^2\tau^{-1}(u) - D_2\rho^2(x, \mu_\tau) \cdot D^2\tau^{-1}(\mu_\tau)]. \end{aligned}$$

By triangle inequality and the (bi-)linearity of the Fréchet derivatives,

$$\begin{aligned} & \sup_u \sup_{\rho_\tau(x, \mu_\tau) \leq \pi/2} \|D_2^2\rho_\tau^2(x, u) - D_2^2\rho_\tau^2(x, \mu_\tau)\| \\ & \leq \sup_u \sup_{\rho_\tau(x, \mu_\tau) \leq \pi/2} \|D_2^2\rho^2(x, \tau^{-1}(u)) - D_2^2\rho^2(x, \mu)\| \sup_u \|D\tau^{-1}(u)\|^2 \end{aligned}$$

$$\begin{aligned}
 &+ \sup_{\rho_\tau(x, \mu_\tau) \leq \pi/2} \|D_2^2 \rho^2(x, \mu)\| \sup_u \|D\tau^{-1}(u)\| \sup_u \|D\tau^{-1}(u) - D\tau^{-1}(\mu_\tau)\| \\
 &+ \sup_{\rho_\tau(x, \mu_\tau) \leq \pi/2} \|D_2^2 \rho^2(x, \mu)\| \sup_u \|D\tau^{-1}(u) - D\tau^{-1}(\mu_\tau)\| \|D\tau^{-1}(\mu_\tau)\| \\
 &+ \sup_u \sup_{\rho_\tau(x, \mu_\tau) \leq \pi/2} \|D_2 \rho^2(x, \tau^{-1}(u)) - D_2 \rho^2(x, \mu)\| \sup_u \|D^2 \tau^{-1}(u)\| \\
 &+ \sup_{\rho_\tau(x, \mu_\tau) \leq \pi/2} \|D_2 \rho^2(x, \mu)\| \sup_u \|D^2 \tau^{-1}(u) - D^2 \tau^{-1}(\mu_\tau)\|, \tag{32}
 \end{aligned}$$

where the norms appearing here are appropriate operator norms. By Lemma E.1,  $\sup_{\rho_\tau(x, \mu_\tau) \leq \pi/2} \|D_2 \rho^2(x, \mu)\|$  is finite, and by the triangle inequality  $\sup_u \|D\tau^{-1}(u)\| \leq \|D\tau^{-1}(\mu_\tau)\| + \sup_u \|D\tau^{-1}(u) - D\tau^{-1}(\mu_\tau)\|$ . Thus, to obtain the convergence of (32), it suffices to show that

$$\sup_{\rho(x, \mu) \leq \pi/2} \|D_2^2 \rho^2(x, \tau^{-1}(u)) - D_2^2 \rho^2(x, \mu)\| = o_p(1), \tag{33}$$

$$\sup_{\rho(x, \mu) \leq \pi/2} \|D_2 \rho^2(x, \tau^{-1}(u)) - D_2 \rho^2(x, \mu)\| = o_p(1), \tag{34}$$

$$\sup_u \|D\tau^{-1}(u) - D\tau^{-1}(\mu_\tau)\| = o_p(1), \tag{35}$$

$$\sup_u \|D^2 \tau^{-1}(u) - D^2 \tau^{-1}(\mu_\tau)\| = o_p(1). \tag{36}$$

Equation (35)–(36) are consequences of the smoothness of  $\tau^{-1}$  and Proposition 3.1. Since (33) and (34) can be derived in a similar fashion, we show the proof for the former only. Viewing  $D_2^2 \rho^2$  as a mapping from  $\mathcal{S}^\infty \times \mathcal{S}^\infty$  to  $\mathcal{B}(\mathbb{H}^2, \mathbb{R})$ , for  $h \in \mathbb{H}$ ,  $\rho(p, q_j) \leq \pi/2$ ,  $j = 1, 2$ , and  $q_1$  and  $q_2$  sufficiently close,

$$|D_2^2 \rho^2(p, q_1)(h, h) - D_2^2 \rho^2(p, q_2)(h, h)| \leq \sup \|D(D_2^2 \rho^2(p, \cdot)(h, h))(u)\| \rho(q_1, q_2) \tag{37}$$

by Taylor’s theorem, where the supremum is taken over all  $u$  lying on the geodesic connecting  $q_1$  and  $q_2$ . The third derivative of  $\rho^2$  is, for  $v \in \mathbb{H}$ ,

$$\begin{aligned}
 D [D_2^2 \rho^2(p, \cdot)(h, h)](q)v &= \frac{d}{dt} [D_2^2 \rho^2(p, \exp_q tv)(h, h)] \Big|_{t=0} \\
 &= -2(1 - \langle p, q \rangle_{\mathbb{H}}^2)^{-5/2} \langle v, p \rangle_{\mathbb{H}} \\
 &\quad \left\{ \langle h, p \rangle_{\mathbb{H}}^2 \left( \arccos(\langle p, q \rangle_{\mathbb{H}}) + 2 \arccos(\langle p, q \rangle_{\mathbb{H}}) \langle p, q \rangle_{\mathbb{H}}^2 - 3 \langle p, q \rangle_{\mathbb{H}} \sqrt{1 - \langle p, q \rangle_{\mathbb{H}}^2} \right) \right. \\
 &\quad \left. + \|h\|_{\mathbb{H}}^2 (-1 + \langle p, q \rangle_{\mathbb{H}}^2) \left( \arccos(\langle p, q \rangle_{\mathbb{H}}) - \langle p, q \rangle_{\mathbb{H}} \sqrt{1 - \langle p, q \rangle_{\mathbb{H}}^2} \right) \right\}.
 \end{aligned}$$

The formula above depends on  $p, q, h$ , and  $v$  only through  $(\langle p, q \rangle_{\mathbb{H}}, \langle v, p \rangle_{\mathbb{H}}, \|h\|_{\mathbb{H}}^2, \langle h, p \rangle_{\mathbb{H}}) \in \mathbb{R}^4$ . By compactness and continuity,

$$\sup_{\substack{\|v\|_{\mathbb{H}} = \|h\|_{\mathbb{H}} = 1 \\ \rho(p, q) \leq \pi/2}} |D [D_2^2 \rho^2(p, \cdot)(h, h)](q)v| \leq C_0 \tag{38}$$

for some finite constant  $C_0 > 0$ . Proposition 3.1, (37), and (38) imply (33).  $\square$

**Appendix F: Proofs for Corollaries in Section 4**

*Proof of Corollary 4.2.* Under  $H_0 : \mu_1 = \mu_2$ ,

$$\begin{aligned} \sqrt{n}[\tau(\hat{\mu}_1) - \tau(\hat{\mu}_2)] &= \sqrt{\frac{n}{n_1}}\sqrt{n_1}[\tau(\hat{\mu}_1) - \tau(\mu_1)] - \sqrt{\frac{n}{n_2}}\sqrt{n_2}[\tau(\hat{\mu}_2) - \tau(\mu_2)] \\ &= \sqrt{\frac{n}{n_1}}Z_1 - \sqrt{\frac{n}{n_2}}Z_2, \end{aligned}$$

where  $Z_g$  is the limiting zero-mean Gaussian random element for  $n_g^{1/2}[\tau(\hat{\mu}_g) - \tau(\mu_g)]$ ,  $g = 1, 2$ . The claim follows from the continuous mapping theorem (Theorem 1.3.6, van der Vaart and Wellner, 1996).  $\square$

*Proof of Corollary 4.1 and Corollary 4.3.* We show the proof for the one-sample case, while the proof for the two-sample case is analogously derived based on Corollary 4.2. Let  $Z$  be a zero-mean Gaussian element with covariance  $\mathcal{I}$ . Applying basis expansion,  $Z = \sum_{k=1}^{\infty} \xi_k \phi_k$  with probability 1 (Theorem 7.2.7, Hsing and Eubank, 2015), where the  $\xi_k = \langle Z, \phi_k \rangle_{\mathbb{G}}$  are independent Gaussian random variables with variance  $\lambda_k$ . By Parseval’s identity,  $\|Z\|_{\mathbb{G}}^2 = \sum_{k=1}^{\infty} \xi_k^2 \stackrel{L}{=} \sum_{k=1}^{\infty} \lambda_k W_k$ , where  $\stackrel{L}{=}$  denotes equal in distribution. By Theorem 3.2 and the continuous mapping theorem, we have  $T_1 \xrightarrow{L} \sum_{k=1}^{\infty} \lambda_k W_k$ .

By the perturbation theory (Section 5.1 in Hsing and Eubank, 2015), Theorem 3.2, and the consistency of  $\hat{\mathcal{I}}$  for  $\mathcal{I}$  given in Proposition 3.2, we have  $S_1 - \tilde{S}_1 = o_p(1)$ . Thus the continuous mapping theorem implies that  $\tilde{S} \xrightarrow{L} \chi_K^2$  and that  $S_1$  converges to the same limit.  $\square$

*Proof of Corollary 4.4.* The consistency of  $\hat{\mu}^*$  to  $\mu$  is shown analogously to the proof of Proposition 3.1. The sample intrinsic mean  $\hat{\mu}_{\tau}$  on chart  $\tau$  is an  $Z$ -estimator since  $\Psi_{n,\tau}(\hat{\mu}_{\tau}) = 0$ , targeting  $\mu_{\tau}$  which satisfies  $\Psi_{\tau}(\mu_{\tau}) = 0$ . Theorem 3.1 of Wellner and Zhan (1996) is then applied to derive the bootstrap limiting distribution after verifying its conditions: Condition A.1–A.5 in Wellner and Zhan (1996) are guaranteed by the smoothness of  $\Psi_{n,\tau}$  and  $\Psi_{\tau}$  and (A1); the bootstrap weights conditions B.1–B.5 are satisfied by the nonparametric bootstrap weights.  $\square$

**Appendix G: Paired-sample test**

The setup is that we make correlated and paired observations  $(X_1^{[1]}, X_1^{[2]}), \dots, (X_n^{[1]}, X_n^{[2]})$ , where the two populations have unknown intrinsic means  $\mu_1$  and  $\mu_2$ , respectively. The paired-sample hypothesis is

$$H_0 : \mu_1 = \mu_2, \quad H_1 : \mu_1 \neq \mu_2.$$

Let  $\hat{\mu}_g$  be the sample intrinsic mean in Population  $g = 1, 2$ . We need the following CLT for paired samples.

**Corollary G.1.** *Suppose the conditions of Theorem 3.2 are satisfied by both populations. Under  $H_0$ ,*

$$\sqrt{n}[\tau(\hat{\mu}_1) - \tau(\hat{\mu}_2)] \xrightarrow{L} Z_3,$$

where  $Z_3$  is a zero-mean Gaussian random elements with covariance operator

$$\begin{aligned} \mathcal{T}_{paired} &= (\Lambda_{1\tau}^{-1})\mathcal{F}_{11}(\Lambda_{1\tau}^{-1})^* + (\Lambda_{2\tau}^{-1})\mathcal{F}_{22}(\Lambda_{2\tau}^{-1})^* - (\Lambda_{1\tau}^{-1})\mathcal{F}_{12}(\Lambda_{2\tau}^{-1})^* \\ &\quad - (\Lambda_{2\tau}^{-1})\mathcal{F}_{21}(\Lambda_{1\tau}^{-1})^*, \end{aligned}$$

where  $\Lambda_{g\tau} = \mathbb{E}D_2^2\rho_\tau^2(X_\tau^{[g]}, \mu_\tau)$ ,  $\mathcal{F}_{gl} = \mathbb{E}[\psi_\tau(X^{[g]}, \mu_\tau) \otimes \psi_\tau(X^{[l]}, \mu_\tau)]$ ,  $\mu_\tau = \tau(\mu_1) = \tau(\mu_2)$ , for  $g, l = 1, 2$ .

*Proof of Corollary G.1.* Under  $H_0 : \mu_1 = \mu_2$ ,

$$\sqrt{n}[\tau(\hat{\mu}_1) - \tau(\hat{\mu}_2)] = \sqrt{n}[\tau(\hat{\mu}_1) - \tau(\mu_1)] - \sqrt{n}[\tau(\hat{\mu}_2) - \tau(\mu_2)],$$

and by the proof of Theorem 3.2,

$$\sqrt{n}[\tau(\hat{\mu}_g) - \tau(\mu_g)] = -\Lambda_{g\tau}^{-1} \frac{1}{\sqrt{n}} \sum_{i=1}^n \psi_\tau(X_i^{[g]}, \tau(\mu_g)) + o_p(1)$$

for  $g = 1, 2$ . By the CLT and the continuous mapping theorem,  $\sqrt{n}[\tau(\hat{\mu}_1) - \tau(\hat{\mu}_2)]$  converges to a zero-mean Gaussian random element  $Z_3$  in  $\mathbb{G}$ . Its covariance operator is  $\mathbb{E}[D \otimes D]$ , where

$$D = \Lambda_{1\tau}^{-1}\psi_\tau(X_1^{[1]}, \mu_\tau) - \Lambda_{2\tau}^{-1}\psi_\tau(X_1^{[2]}, \mu_\tau),$$

which gives the covariance  $\mathcal{T}_{paired}$  as stated. □

We estimate  $\mathcal{T}_{paired}$  by the empirical version  $\hat{\mathcal{T}}_{paired}$ , and let  $(\hat{\lambda}_{k,paired}, \hat{\phi}_{k,paired})$  denote the  $k$ th eigenpair of  $\hat{\mathcal{T}}_{paired}$ . The paired-sample test statistics are

$$\begin{aligned} T_3 &= n \|\tau(\hat{\mu}_1) - \tau(\hat{\mu}_2)\|_{\mathbb{G}}^2, \\ S_3 &= n \sum_{k=1}^K \frac{\langle \tau(\hat{\mu}_1) - \tau(\hat{\mu}_2), \hat{\phi}_{k,paired} \rangle_{\mathbb{G}}^2}{\hat{\lambda}_{k,paired}}. \end{aligned}$$

By the continuous mapping theorem, the asymptotic distributions of  $T_3$  and  $S_3$  are  $\sum_{k=1}^\infty \lambda_{k,paired} W_k$  and  $\chi_K^2$ , respectively, where the  $W_k$  are i.i.d.  $\chi_1^2$  random variables. For bootstrapping the paired-sample test statistic, we resample  $n$  paired observations with replacement and obtain bootstrap samples  $(X_1^{[1]}, X_1^{[2]})^*, \dots, (X_n^{[1]}, X_n^{[2]})^*$  and their respective intrinsic means  $\hat{\mu}_1^*$  and  $\hat{\mu}_2^*$ . Also let  $\hat{\mathcal{T}}_{paired}^*$  be the bootstrap pooled covariance, and the  $(\hat{\lambda}_k^*, \hat{\phi}_k^*)$  be the bootstrap eigenpairs. The bootstrap test statistics are

$$\begin{aligned} T_3^* &= n \|\tau(\hat{\mu}_1^*) - \tau(\hat{\mu}_2^*)\|_{\mathbb{G}}^2, \\ S_3^* &= n \sum_{k=1}^K \frac{\langle \tau(\hat{\mu}_1^*) - \tau(\hat{\mu}_2^*), \hat{\phi}_k^* \rangle_{\mathbb{G}}^2}{\hat{\lambda}_k^*}. \end{aligned}$$

**Corollary G.2.** Under the conditions of [Corollary G.1](#), the bootstrap intrinsic mean  $\hat{\mu}_g^*$  is a consistent estimator of  $\mu_g$ ,  $g = 1, 2$ . Furthermore,

$$\sqrt{n} \{[\tau(\hat{\mu}_1^*) - \tau(\hat{\mu}_1)] - [\tau(\hat{\mu}_2^*) - \tau(\hat{\mu}_2)]\} \xrightarrow{L} Z_3^*,$$

where  $Z_3^*$  is a Gaussian random element sharing the same distribution with  $Z_3$ . Thus the asymptotic distributions for  $T_3^*$  and  $T_3$  are the same, as well as for  $S_3^*$  and  $S_3$ .

## Appendix H: Additional simulation results for the one-sample test

For the one-sample test, the true intrinsic population mean was  $\mu = \exp_{\mu_0} \delta v$  for  $\delta \in [-0.4, 0.4]$ , where  $\mu_0$  was the square root density of  $\text{Beta}(2, 1)$ ,  $v = K_\mu^{-1/2} \sum_{k=1}^{K_\mu} \phi_k$ , and  $\phi_k = R_\mu(\psi_{k+1})$ ,  $k = 1, \dots, \infty$ . All other setups were identical to the two-sample test described in the main text except that the sample size  $n$  ranged among  $\{20, 50, 100\}$ ,

The empirical power curves for the one-sample tests are displayed in [Figure 3](#). All three methods had similar performance in the symmetric case (first three columns, [Figure 3](#)) and all controlled the size approximately at the nominal level  $\alpha = 0.05$ . In contrast, when the observations were distributed asymmetrically around the mean (last three columns, [Figure 3](#)), the extrinsic method was biased for all  $n$  and had much larger sizes than the nominal level when  $\delta = 0$  and smallest empirical powers at  $\delta = 0.05$  or  $0.1$ . The biases for the extrinsic bootstrap test became more apparent and the size increased as  $n$  increased. The proposed intrinsic asymptotic and bootstrap tests performed overall similarly well. Both intrinsic methods had smaller bias as  $n$  increased, which became almost negligible when  $n = 50$  or  $100$ .

## Acknowledgments

The author is grateful to the reviewers for their constructive comments. He would also like to thank Hans-Georg Müller for bringing up this topic.

## References

- AFSARI, B. (2011). Riemannian  $L^p$  Center of Mass: Existence, Uniqueness, and Convexity. *Proceedings of the American Mathematical Society* **139** 655–673. [MR2736346](#)
- AHIDAR-COUTRIX, A., LE GOUIC, T. and PARIS, Q. (2019). Convergence Rates for Empirical Barycenters in Metric Spaces: Curvature, Convexity and Extendable Geodesics. *Probab. Theory Relat. Fields*. [MR4095017](#)
- AUE, A., RICE, G. and SÖNMEZ, O. (2018). Detecting and Dating Structural Breaks in Functional Data without Dimension Reduction. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* **80** 509–529. [MR3798876](#)

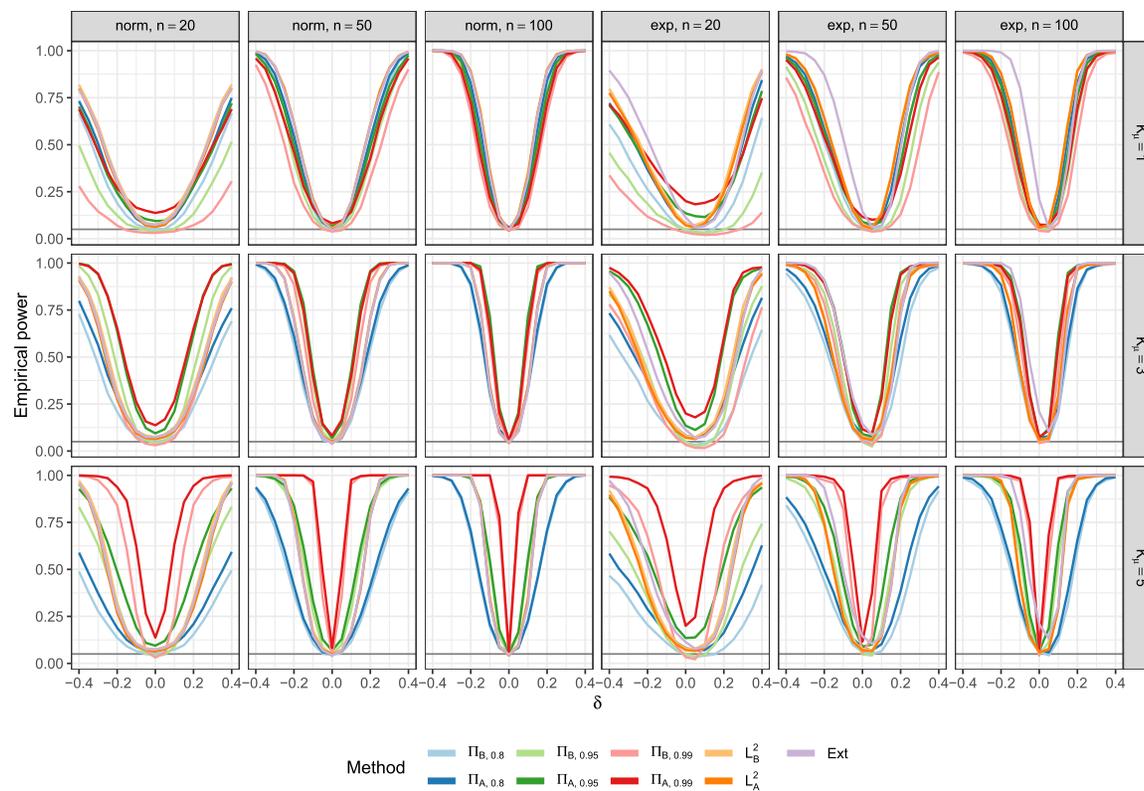


FIG 3. Empirical power curves for the one-sample tests. Columns correspond to different generating distributions for the principal components  $\xi_{gik}$  and sample sizes  $n$ , and rows correspond to different numbers of components  $K_\mu$  in the mean difference. The horizontal gray lines indicate the nominal level  $\alpha = 0.05$ .  $\Pi_{A,r}$  and  $\Pi_{B,r}$ , projection-based tests with FVE threshold  $r$ ;  $L_A^2$  and  $L_B^2$ , norm-based tests; *Ext*, the extrinsic bootstrap test. Subscripts *A* and *B* stand for the proposed tests in the asymptotic and bootstrap versions, respectively.

- BAUER, M., ESLITZBICHLER, M. and GRASMAIR, M. (2017). Landmark-Guided Elastic Shape Analysis of Human Character Motions. *Inverse Problems & Imaging* **11** 601–621. [MR3668343](#)
- BERKES, I., GABRYS, R., HORVÁTH, L. and KOKOSZKA, P. (2009). Detecting Changes in the Mean of Functional Observations. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* **71** 927–946. [MR2750251](#)
- BHATTACHARYA, R. and LIN, L. (2017). Omnibus CLTs for Fréchet Means and Nonparametric Inference on Non-Euclidean Spaces. *Proceedings of the American Mathematical Society* **145** 413–428. [MR3565392](#)
- BHATTACHARYA, R. and PATRANGENARU, V. (2003). Large Sample Theory of Intrinsic and Extrinsic Sample Means on Manifolds - I. *Annals of Statistics* **31** 1–29. [MR1962498](#)
- BHATTACHARYA, R. and PATRANGENARU, V. (2005). Large Sample Theory of Intrinsic and Extrinsic Sample Means on Manifolds - II. *Annals of statistics* **33** 1225–1259. [MR2195634](#)
- BOGACHEV, V. I. (2018). *Weak Convergence of Measures*. American Mathematical Society, Providence. [MR3837546](#)
- BOSQ, D. (2000). *Linear Processes in Function Spaces: Theory and Applications*. Springer-Verlag, New York. [MR1783138](#)
- CHENG, J., GHOSH, A., JIANG, T. and DERICHE, R. (2009). A Riemannian Framework for Orientation Distribution Function Computing. In *Medical Image Computing and Computer-Assisted Intervention – MICCAI 2009*, (G.-Z. Yang, D. Hawkes, D. Rueckert, A. Noble and C. Taylor, eds.) **5761** 911–918. Springer Berlin Heidelberg, Berlin, Heidelberg.
- CHU, L. and CHEN, H. (2019). Asymptotic Distribution-Free Change-Point Detection for Multivariate and Non-Euclidean Data. *Ann. Statist.* **47** 382–414. [MR3910545](#)
- DAI, X. and MÜLLER, H.-G. (2018). Principal Component Analysis for Functional Data on Riemannian Manifolds and Spheres. *Annals of Statistics* **46** 3334–3361. [MR3852654](#)
- DU, J., GOH, A., KUSHNAREV, S. and QIU, A. (2014). Geodesic Regression on Orientation Distribution Functions with Its Application to an Aging Study. *NeuroImage* **87** 416–426.
- DUBEY, P. and MÜLLER, H.-G. (2019). Fréchet Analysis of Variance for Random Objects. *Biometrika* **106** 803–821. [MR4031200](#)
- DUBEY, P. and MÜLLER, H.-G. (2020). Functional Models for Time-Varying Random Objects. *J. R. Stat. Soc. B* **82** 275–327. [MR4084166](#)
- ELLINGSON, L., PATRANGENARU, V. and RUYMGAART, F. (2013). Nonparametric Estimation of Means on Hilbert Manifolds and Extrinsic Analysis of Mean Shapes of Contours. *Journal of Multivariate Analysis* **122** 317–333. [MR3189326](#)
- ELTZNER, B. and HUCKEMANN, S. F. (2019). A Smearly Central Limit Theorem for Manifolds with Application to High-Dimensional Spheres. *Ann. Statist.* **47** 3360–3381. [MR4025745](#)
- FLETCHER, P. T., LU, C., PIZER, S. M. and JOSHI, S. (2004). Principal Geodesic Analysis for the Study of Nonlinear Statistics of Shape. *IEEE Trans-*

- actions on Medical Imaging* **23** 995–1005.
- FRÉCHET, M. (1948). Les Éléments Aléatoires de Nature Quelconque Dans Un Espace Distancié. *Annales de l'Institut Henri Poincaré* **10** 215–310. [MR0027464](#)
- GOUIC, T. L., PARIS, Q., RIGOLLET, P. and STROMME, A. J. (2019). Fast Convergence of Empirical Barycenters in Alexandrov Spaces and the Wasserstein Space. *arXiv*.
- HÁJEK, J. (1962). On Linear Statistical Problems in Stochastic Processes. *Czechoslovak Mathematical Journal* **12** 404–444. [MR0152090](#)
- HALL, P. (1992). *The Bootstrap and Edgeworth Expansion*. Springer, New York. [MR1145237](#)
- HENNING, W. and SRIVASTAVA, A. (2016). A Two-Sample Test for Statistical Comparisons of Shape Populations. In *2016 IEEE Winter Conference on Applications of Computer Vision (WACV)* 1–9. IEEE.
- HORVÁTH, L., KOKOSZKA, P. and REEDER, R. (2013). Estimation of the Mean of Functional Time Series and a Two-Sample Problem. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* **75** 103–122. [MR3008273](#)
- HOTZ, T. and HUCKEMANN, S. (2015). Intrinsic Means on the Circle: Uniqueness, Locus and Asymptotics. *Ann Inst Stat Math* **67** 177–193. [MR3297863](#)
- HSING, T. and EUBANK, R. (2015). *Theoretical Foundations of Functional Data Analysis, with an Introduction to Linear Operators*. Wiley, Hoboken. [MR3379106](#)
- HUCKEMANN, S. F. (2012). On the Meaning of Mean Shape: Manifold Stability, Locus and the Two Sample Test. *Ann Inst Stat Math* **64** 1227–1259. [MR2981621](#)
- JOSHI, S. H., KLASSEN, E., SRIVASTAVA, A. and JERMYN, I. (2007). A Novel Representation for Riemannian Analysis of Elastic Curves in  $R^n$ . In *2007 IEEE Conference on Computer Vision and Pattern Recognition* 1–7. IEEE.
- KARCHER, H. (1977). Riemannian Center of Mass and Mollifier Smoothing. *Communications on Pure and Applied Mathematics* **30** 509–541. [MR0442975](#)
- KENDALL, W. S. (1990). Probability, Convexity, and Harmonic Maps with Small Image I: Uniqueness and Fine Existence. *Proceedings of the London Mathematical Society* **s3-61** 371–406. [MR1063050](#)
- LANG, S. (1999). *Fundamentals of Differential Geometry*. Springer, New York. [MR1666820](#)
- LAZAR, D. and LIN, L. (2017). Scale and Curvature Effects in Principal Geodesic Analysis. *Journal of Multivariate Analysis* **153** 64–82. [MR3578839](#)
- LE, H. (2001). Locating Fréchet Means with Application to Shape Spaces. *Advances in Applied Probability* **33** 324–338. [MR1842295](#)
- LIN, Z. and YAO, F. (2019). Intrinsic Riemannian Functional Data Analysis. *The Annals of Statistics* **47** 3533–3577. [MR4025751](#)
- PETERSEN, A., LIU, X. and DIVANI, A. A. (2019). Wasserstein  $F$ -Tests and Confidence Bands for the Fréchet Regression of Density Response Curves. *arXiv*. [MR4206692](#)
- PETERSEN, A. and MÜLLER, H.-G. (2016). Functional Data Analysis for Den-

- sity Functions by Transformation to a Hilbert Space. *The Annals of Statistics* **44** 183–218. [MR3449766](#)
- RAO, C. R. (1945). Information and the Accuracy Attainable in the Estimation of Statistical Parameters. *Bulletin of Calcutta Mathematical Society* **37** 81–91. [MR0015748](#)
- RUDIN, W. (1973). *Functional Analysis*. McGraw-Hill, New York. [MR0365062](#)
- SCHÖTZ, C. (2019). Convergence Rates for the Generalized Fréchet Mean via the Quadruple Inequality. *Electron. J. Statist.* **13** 4280–4345. [MR4023955](#)
- SIMON, B. (2005). *Trace Ideals and Their Applications*. American Mathematical Society, Providence. [MR2154153](#)
- SRIVASTAVA, A., JERMYN, I. and JOSHI, S. (2007). Riemannian Analysis of Probability Density Functions with Applications in Vision. In *2007 IEEE Conference on Computer Vision and Pattern Recognition* 1–8.
- STRAIT, J., CHKREBTHI, O. and KURTEK, S. (2019). Automatic Detection and Uncertainty Quantification of Landmarks on Elastic Curves. *Journal of the American Statistical Association* 1–23. [MR4011754](#)
- SU, J., KURTEK, S., KLASSEN, E. and SRIVASTAVA, A. (2014). Statistical Analysis of Trajectories on Riemannian Manifolds: Bird Migration, Hurricane Tracking and Video Surveillance. *Annals of Applied Statistics* **8** 530–552. [MR3192001](#)
- TUCKER, J. D., WU, W. and SRIVASTAVA, A. (2013). Generative Models for Functional Data Using Phase and Amplitude Separation. *Computational Statistics & Data Analysis* **61** 50–66. [MR3063000](#)
- VAN DER VAART, A. and WELLNER, J. (1996). *Weak Convergence and Empirical Processes: With Applications to Statistics*. Springer, New York. [MR1385671](#)
- VAN DER VAART, A. and WELLNER, J. A. (2000). Preservation Theorems for Glivenko–Cantelli and Uniform Glivenko–Cantelli Classes. In *High Dimensional Probability II* 115–133. Springer, New York. [MR1857319](#)
- WANG, J.-L., CHIOU, J.-M. and MÜLLER, H.-G. (2016). Functional Data Analysis. *Annual Review of Statistics and its Application* **3** 257–295.
- WELLNER, J. A. and ZHAN, Y. (1996). Bootstrapping Z-Estimators Technical Report.
- WU, W. and SRIVASTAVA, A. (2014). Analysis of Spike Train Data: Alignment and Comparisons Using the Extended Fisher-Rao Metric. *Electronic Journal of Statistics* **8** 1776–1785. [MR3273594](#)
- XIE, W., KURTEK, S., BHARATH, K. and SUN, Y. (2017). A Geometric Approach to Visualization of Variability in Functional Data. *Journal of the American Statistical Association* **112** 979–993. [MR3735354](#)
- YOUNES, L. (1998). Computable Elastic Distances between Shapes. *SIAM Journal on Applied Mathematics* **58** 565–586. [MR1617630](#)
- YU, Q., LU, X. and MARRON, J. (2017). Principal Nested Spheres for Time-Warped Functional Data Analysis. *Journal of Computational and Graphical Statistics* **26** 144–151. [MR3610415](#)
- ZHU, H., CHEN, Y., IBRAHIM, J. G., LI, Y., HALL, C. and LIN, W. (2009).

Intrinsic Regression Models for Positive-Definite Matrices with Applications to Diffusion Tensor Imaging. *Journal of the American Statistical Association* **104** 1203–1212. [MR2750245](#)

ZIEZOLD, H. (1977). On Expected Figures and a Strong Law of Large Numbers for Random Elements in Quasi-Metric Spaces. In *Transactions of the Seventh Prague Conference on Information Theory, Statistical Decision Functions, Random Processes and of the 1974 European Meeting of Statisticians* 591–602. [MR0501230](#)