Intrinsic Riemannian Cross-covariance for Manifold-valued Random Objects

Carlos J. Soto
Department of Mathematics and Statistics
University of Massachusetts Amherst
Amherst, MA 01002
carlossoto@umass.edu

Cheng Wang
Department of Mathematics and Statistics
University of Massachusetts Amherst
Amherst, MA 01002
chengwang@umass.edu

Yujing Huang
Department of Industrial and Systems Engineering
State University of New York at Buffalo
Buffalo, NY 14260
yujinghu@buffalo.edu

Xiaoyu Chen *
Department of Industrial and Systems Engineering
State University of New York at Buffalo
Buffalo, NY 14260
xchen325@buffalo.edu

Abstract

Covariance estimation yields a fundamental second-order statistic underlying representation learning, dimension reduction, and dependence modeling. While covariance has been well understood in Euclidean spaces, it is ill-defined for random objects residing on nonlinear Riemannian manifolds, which increasingly arise in modern machine learning applications involving shapes, symmetric positive definite (SPD) matrices, etc. This paper introduces an intrinsic Riemannian cross-covariance for manifold-valued random objects. Our approach defines covariance and correlation by transporting local variations to a common tangent space via parallel transport, yielding a second-order descriptor that is independent of arbitrary coordinate choices. We establish that the proposed covariance inherits desirable properties of its Euclidean counterparts and characterize its asymptotic behavior. Numerical studies on spheres and SPD manifolds, together with real-data experiments on heart valve shapes in Kendall's shape space, demonstrate the effectiveness of our estimators and verify the stated properties. Our results position the Riemannian covariance as a fundamental tool for second-order learning and analysis in non-Euclidean representation spaces.

1 Introduction

Understanding the dependence between two random objects is a natural and important task underlying representation learning. This dependence can typically be quantified via the cross-covariance matrix along with subsequent analysis including a scalar covariance and correlation. The foundations for the correlation analysis, however, are grounded in the existence of a vector space structure. For

*Corresponding Author.

40th Conference on Neural Information Processing Systems (NeurIPS 2026).

non-Euclidean manifold-valued data, such vector space structures are not well-defined, preventing the natural definition of cross-covariance and correlation.

A Riemannian manifold is a generalization of Euclidean space which need not be flat. Data that naturally lives on Riemannian manifolds is often encountered in statistical analysis. Extending linear, statistical methods to non-linear spaces is a ubiquitous endeavor; a basic example is the Fréchet mean, or Riemannian center of mass, which extends the notion of the Euclidean mean to metric spaces [Fréchet, 1948]. Data such as directional data [Mardia, 1975], meteorological data [Fu et al., 2025], protein data [Mardia et al., 2018], and discrete densities [Rao et al., 1945], can all be thought of as spherical data, a Riemannian manifold of constant, positive curvature. Other common examples include covariance matrices, for instance in brain imaging, as elements of the cone of symmetric positive definite matrices [Dai et al., 2019], point cloud shape data as elements of complex projective space [Kendall, 1984], and tree structured data [Lu and Miao, 2015]. Analyzing such data requires extensions of the standard Euclidean statistical tools.

Due to the prevalence of non-Euclidean data in machine learning analyses and the need to quantifiably analyze such data, generalizations of the notion of covariance have been proposed. Some such generalizations include, to Riemannian manifolds for a single random variable [Pennec, 2006], to Wasserstein space [Petersen and Müller, 2019a], to Riemannian functional data [Lin and Yao, 2019, Shao et al., 2022], and to Riemannian manifolds for more than one random variables sharing the same support [Abuqrais and Pigoli, 2026].

The contributions of this paper are as follows:

- we introduce the *intrinsic footpoint-invariant Riemannian cross-covariance*, a notion of covariance between two random variables that live on a Riemannian manifold, and subsequently define the *Riemannian correlation*;
- we elucidate the properties of the introduced estimators and establish their asymptotic behavior;
- we demonstrate the effectiveness of our estimators on simulated data on the sphere and the SPD manifold, as well as on experimental shape data via Kendall's shape space.

2 Background

In the following, we introduce the necessary background on differential geometry and Riemannian manifolds. More details can be found in Appendix B. For additional details, we refer to do Carmo [1992] and Lee [2018]. After introducing the foundational geometric tools, we give background on the state-of-the-art covariance estimation for Riemannian manifolds.

2.1 Riemannian Geometry

Let $\mathcal{M}$ be a d-dimensional smooth topological manifold. At every point $p \in \mathcal{M}$ there exists a neighborhood which is diffeomorphic to $\mathbb{R}^d$. Suppose $\alpha : [-\delta, \delta] \to \mathcal{M}$ is a smooth curve on $\mathcal{M}$ such that $\alpha(0) = p$ and denote by α' its derivative. The tangent space of $\mathcal{M}$ at p, denoted $T_p\mathcal{M}$, is the collection of all tangent vectors $\alpha'(0)$, that is $T_p\mathcal{M} = \{\alpha'(0) | \alpha(0) = p, \alpha : [-\delta, \delta] \to \mathcal{M}\}$. Each tangent space is a vector space with $\dim T_p\mathcal{M} = \dim \mathcal{M}$. We equip each tangent space with an inner product $g_p(\cdot, \cdot) = \langle \cdot, \cdot \rangle_{g(p)} : T_p\mathcal{M} \times T_p\mathcal{M} \to \mathbb{R}$ which varies smoothly on $\mathcal{M}$. A Riemannian manifold is a smooth manifold equipped with a Riemannian metric, $(\mathcal{M}, g)$.

With respect to the metric, we have a unique torsion-free affine connection ∇, the Levi–Civita connection. This connection thus allows us to define the notion of parallel vector fields. Given a smooth path $\alpha : [0, 1] \to \mathcal{M}$, the parallel transport of $v \in T_{\alpha(0)}\mathcal{M}$ is denoted as $\Gamma_{\alpha(0)}^{\alpha(1)} v$ and satisfies $\nabla_{\alpha'(t)} V(t) = 0$ with V being a smooth vector field with $V(\alpha(0)) = v$. Since parallel transport moves tangent vectors in a parallel fashion, it preserves inner products and vector norms; that is, parallel transport is an isometry, so for $v, w \in T_p\mathcal{M}$ we have $\langle v, w \rangle_{g(p)} = \langle \Gamma_p^q v, \Gamma_p^q w \rangle_{g(q)}$. For the remainder of this paper, we assume $\mathcal{M}$ to be a complete, finite-dimensional Riemannian manifold; concrete supports are restricted to geodesically convex sets via Assumption 3.1.

2.2 Related Literature - Covariance

The Fréchet mean [Fréchet, 1948] generalizes the notion of the arithmetic mean to metric spaces and is defined with respect to the Fréchet variance function. The Fréchet variance of a random variable X with probability measure λ at $x \in \mathcal{M}$ is defined as

$$\sigma_X^2(x) = \mathbb{E}(d^2(x, X)) = \int_{\mathcal{M}} d^2(x, y)\, \mathrm{d}\lambda(y).$$

The Fréchet mean is the point on the manifold that minimizes the variance, i.e., $\mathbb{E}(X) = \mu \coloneqq \operatorname{argmin}_{x\in\mathcal{M}} \sigma_X^2(x)$. While the Fréchet variance is defined at any point x, if no point is designated, then the variance is assumed to be at the mean, i.e., $\sigma_X^2 \coloneqq \sigma_X^2(\mu)$. The Fréchet variance may not have a unique minimizer on general manifolds, namely on positively curved manifolds [Karcher, 1977, Afsari, 2011]. To ensure the existence of a unique mean, one is required to bound the density in a geodesic ball centered at p with radius $r \leq \frac{1}{2}\min\{\operatorname{inj}\mathcal{M}, \pi/2\sqrt{\kappa}\}$, $B_r(p)$, [Karcher, 1977, Kendall, 1990], where $\operatorname{inj}\mathcal{M}$ denotes the injectivity radius and κ is an upper bound on the sectional curvature.

Pennec [2006] first extended the definition of covariance to Riemannian manifolds focusing on a single random variable. Suppose X is a random variable as above with a unique mean $\mathbb{E}(X) = \mu$ then the covariance of X is defined as

$$\Sigma_{X,X}^{Pennec} = \mathrm{Cov}_X(\mu) \coloneqq \mathbb{E}\left[(\log_\mu X)(\log_\mu X)^T\right] = \int_{\mathcal{M}} (\log_\mu x)(\log_\mu x)^T \mathrm{d}\lambda(x).$$

Similar to the Euclidean setting, we have that the trace of the covariance is the Fréchet variance, that is, $\operatorname{tr}(\Sigma_{X,X}^{Pennec}) = \sigma_X^2$. Given a dataset $D = \{x_1, \ldots, x_n\}$, the empirical estimate of the covariance is $\hat{\Sigma}_{X,X}^{Pennec} = \frac{1}{n}\sum_i \left[(\log_{\hat{\mu}} x_i)(\log_{\hat{\mu}} x_i)^T\right]$ where $\hat{\mu}$ is the empirical estimate of the Fréchet mean.

The previous definition, however, is limited to measuring the covariance of a single random variable. Extending the idea of covariance to more than one variable for complex spaces has been of recent interest. Lyons [2013] extended distance covariance from Euclidean space to general metric spaces and established that distance covariance characterizes independence when the underlying spaces are of strong negative type, a term they define. The corresponding distance correlation takes values in $[0, 1]$ and measures the strength of dependence. Building on this framework, Pan et al. [2017] introduced conditional local distance correlation to measure the association between two random objects at a given value. They developed kernel-based estimators and a bootstrap method for testing conditional independence and constructing confidence bands.

Abuqrais and Pigoli [2026] propose an extension of the covariance definition to two random variables on Riemannian manifolds. To this end, the authors require that both densities have the same support. That is, given random variables X and Y defined on $\mathcal{M}$, their respective measures are defined, almost surely, in the same geodesic ball $B_r(p)$. They define the covariance as follows.

Definition 2.1. *[Abuqrais and Pigoli [2026]] The covariance between random variables X and Y at point (footpoint) $p \in \mathcal{M}$ is*

$$\Sigma_{X,Y}^{A\&P}(p) = \mathbb{E}\left[(\log_p X)(\log_p Y)^T\right] - \mathbb{E}\left[\log_p X\right]\mathbb{E}\left[\log_p Y\right]^T.$$

The Riemannian covariance is defined as

$$\mathit{Rcov}(p)(X, Y) \coloneqq \operatorname{tr}(\Sigma_{X,Y}^{A\&P}(p)).$$

Note that if $Y = X$ and $p = \mu$ then we have that $\Sigma_{X,Y}^{A\&P}(p) = \Sigma_{X,X}^{Pennec}$ and so $\mathrm{Rcov}(\mu)(X, X) = \sigma_X^2$. Definition 2.1, by construction, requires both measures to exist in the same geodesic ball; this can be a restrictive limitation. Further, the covariance is dependent on the choice of footpoint p. Abuqrais and Pigoli [2026] give some guidance on how to choose p such as selecting the midpoint of the geodesic connecting the two sample Fréchet means. We aim to address both of these limitations. The empirical estimate in this case is $\hat{\Sigma}_{X,Y}^{A\&P}(p) = \frac{1}{n}\sum\left[(\log_p x_i)(\log_p y_i)^T\right] - \left(\frac{1}{n}\sum \log_p x_i\right)\left(\frac{1}{n}\sum \log_p y_i\right)^T$. The following proposition shows that this definition is not footpoint-invariant; the proof can be found in Appendix G.

Proposition 2.2. *On a Riemannian manifold with non-zero curvature, the Abuqrais-Pigoli covariance Rcov$(p)(X, Y)$ is not invariant with respect to the choice of footpoint p.*

In the space of probability distributions, Petersen and Müller [2019b] introduced a notion of covariance for univariate density functions endowed with the Wasserstein metric. Let $\mathcal{F}_1, \mathcal{F}_2$ denote two random densities with distribution functions F_1, F_2, and let $\bar{f}_i$ (with distribution function $\bar{F}_i$) be the Fréchet (Wasserstein) mean of $\mathcal{F}_i$. The optimal transport map from $\bar{f}_i$ to a realization f_i is $T_i = F_i^{-1} \circ \bar{F}_i$, and its tangent representation $T_i - \mathrm{id}$ lives in the tangent space at $\bar{f}_i$ (formally $L^2(\bar{f}_i)$). The squared Wasserstein distance is induced from this inner product, $d^2(\bar{f}_1, \bar{f}_2) = \int (T(u) - u)^2 \bar{f}_1(u)\, du$, where $T = \bar{F}_2^{-1} \circ \bar{F}_1$ is the optimal transport from $\bar{f}_1$ to $\bar{f}_2$.

Petersen and Müller [2019b] define the Wasserstein covariance as

$$\Sigma_{\mathcal{F}_1,\mathcal{F}_2} = \mathbb{E}(\langle \widetilde{T}_1, T_2 \rangle_{\bar{f}_2}) = \mathbb{E}\left(\int_{\mathbb{R}} (\widetilde{T}_1(u) - u)(T_2(u) - u) \bar{f}_2(u) du \right), \tag{1}$$

where $T_i = F_i^{-1} \circ \bar{F}_i$ is the optimal transport map and $\widetilde{T}_1 = T_1 \circ (\bar{F}_1^{-1} \circ \bar{F}_2) - (\bar{F}_1^{-1} \circ \bar{F}_2) + \mathrm{id}$ is the parallel transport of the optimal transport map to the tangent space of $\bar{f}_2$.

2.3 Related Literature - Correlation

Kim et al. [2014] introduced a way to define the correlation as an extension of canonical correlation analysis to manifold data. We summarize their method here, with more details in Appendix C. Traditional CCA finds linear subspaces at the means of two multivariate random variables and finds projections that maximize the correlation between them. Kim et al. [2014] extend this idea by finding linear subspaces of tangent spaces. Let π_S denote the projection operator onto a geodesic submanifold $S \subset \mathcal{M}$ defined as $\pi_S(x) = \mathrm{argmin}_{\tilde{x} \in S}\, d^2(x, \tilde{x})$.

Let the data be $\{x_i, y_i\}_{i=1}^n$, $\mathbf{a} = \{a_i\}$, $\mathbf{b} = \{b_i\}$, with $\bar{a}, \bar{b}$ the Euclidean means of $\mathbf{a}$ and $\mathbf{b}$, respectively, and μ, ν the Fréchet means of X and Y, respectively. Here W_x, W_y are the basis tangent vectors and π_{W_x} refers to the projection onto the space spanned by W_x. Following this idea, the generalized CCA is expressed as an optimization problem

$$\begin{aligned} \mathcal{R}_{X,Y}^{CCA} =& \mathrm{corr}(\pi_{W_x}(x), \pi_{W_y}(y)) \\ =& \max_{W_x, W_y, \mathbf{a}, \mathbf{b}} \frac{\sum_i (a_i - \bar{a})(b_i - \bar{b})}{\sqrt{\sum_i (a_i - \bar{a})^2} \sqrt{\sum_i (b_i - \bar{b})^2}} \end{aligned}$$

such that $a_i = \mathrm{argmin}_{a_i \in (-\epsilon, \epsilon)} \| \log_{\exp_\mu(a_i W_x)}(x_i) \|^2$, $b_i = \mathrm{argmin}_{b_i \in (-\epsilon, \epsilon)} \| \log_{\exp_\nu(b_i W_y)}(y_i) \|^2$, $\forall i \in \{1, \ldots, n\}$. This formulation finds the tangent vectors which determine the geodesic submanifolds such that the correlation of projection coefficients is maximized.

Additionally, Andrew et al. [2013] introduced Deep CCA, which jointly learns neural-network transformations of paired data views and maximizes canonical correlations between their output representations, while Buenfil and Lila [2025] developed asymmetric CCA for Riemannian manifold-valued functional data paired with high-dimensional Euclidean variables. Basile et al. [2025] took a different approach by proposing Intrinsic Dimension Correlation, which quantifies shared intrinsic degrees of freedom between paired data representations by comparing their marginal and joint intrinsic dimensions. The measure emphasizes shared dimensional structure rather than providing a general characterization of statistical dependence.

Abuqrais and Pigoli [2026] also propose a generalization for a Riemannian correlation defined in terms of their covariance. They define the correlation matrix as

$$\varrho_{X,Y}^{A\&P}(p) = \frac{\Sigma_{X,Y}^{A\&P}(p)}{\sqrt{\mathrm{tr}(\Sigma_{X,X}^{A\&P}(p))\ \mathrm{tr}(\Sigma_{Y,Y}^{A\&P}(p))}},$$

where $\Sigma_{X,Y}^{A\&P}(p)$ is as in Definition 2.1 and the correlation is $\mathcal{R}_{X,Y}^{A\&P}(p) = \mathrm{tr}(\varrho_{X,Y}^{A\&P}(p))$. The authors show that this satisfies the crucial property that $\mathcal{R}_{X,Y}^{A\&P}(p) \in [-1, 1]$ [Abuqrais and Pigoli, 2026, Proposition 3.5].

3 Intrinsic Footpoint-invariant Riemannian Cross-covariance

Proofs for Section 3.1 and Section 3.2 can be found in Appendix G and proofs for Section 3.3 can be found in Appendix H.

3.1 Riemannian Covariance

Suppose we have two random variables X, Y with unique, respective Fréchet means μ, ν.

Assumption 3.1. To ensure we have unique Fréchet means, each dataset $\{Z_1, Z_2, \ldots, Z_n\}$ is bounded in an open ball $B_r(p)$ with $r \leq \frac{1}{2}\min\{\operatorname{inj}\mathcal{M}, \pi/(2\sqrt{\kappa})\}$ [Karcher, 1977, Kendall, 1990], where $\operatorname{inj}\mathcal{M}$ denotes the injectivity radius and κ is an upper bound on the sectional curvature. If $\kappa \leq 0$, then r must simply be finite.

Further, our Definition 3.3 requires the use of parallel transport. This raises an additional issue of uniqueness. We elaborate on this point in Appendix B, but note that parallel transport is generally not unique unless it is restricted to geodesics. Thus, we have,

Assumption 3.2. To ensure a unique parallel transport, we require a unique geodesic between the Fréchet means. Thus, we assume $d(\mu, \nu) < \operatorname{inj}\mathcal{M}$.

Definition 3.3 (Riemannian covariance)**.** *Given random variables X and Y with respective Fréchet means μ and ν, let the Riemannian cross-covariance at footpoint p be defined as*

$$\Sigma_{X,Y}(p) = \mathbb{E}\left[(\Gamma_\mu^p \log_\mu X)(\Gamma_\nu^p \log_\nu Y)^T\right],$$

where the parallel transport is with respect to the geodesic α with $p \in \alpha$, $\alpha : [0,1] \to \mathcal{M}$, $\alpha(0) = \mu$, and $\alpha(1) = \nu$. The Riemannian covariance between X and Y at point p is thus

$$\operatorname{Cov}_{X,Y}(p) = \operatorname{tr}(\Sigma_{X,Y}(p)).$$

Remark 3.4. In Theorem 3.5 we show that our definition is invariant to the choice of p. When there is no canonical or reference choice for p, we define the covariance as

$$\Sigma_{X,Y} = \mathbb{E}\left[(\Gamma_\mu^\nu \log_\mu X)(\log_\nu Y)^T\right], \qquad \operatorname{Cov}_{X,Y} = \operatorname{tr}(\Sigma_{X,Y}).$$

In addition, our Riemannian covariance connects to the Wasserstein covariance (Eq. 1) by defining an operator-valued covariance in a common tangent space, whose trace yields a scalar dependence measure. In the one-dimensional Wasserstein setting, the covariance is defined directly as an L^2 inner product, which coincides with the trace of this operator under the flat geometry.

Theorem 3.5 (Footpoint Invariance)**.** *We have $\operatorname{Cov}_{X,Y}(p) = \operatorname{Cov}_{X,Y}(q)$ for all $p, q \in \alpha$ where α is the geodesic joining μ and ν.*

Our definition is similar to that proposed by Abuqrais and Pigoli [2026], however, we utilize a parallel transport to achieve a common tangent space rather than utilize the tangent space at some midpoint. The use of parallel transport is necessary as $\log_{\mu *}$ and $\log_{\nu *}$ push the densities forward onto different tangent spaces, so the outer (and inner) product is not well-defined. This issue is not present in the Euclidean setting, as the tangent space at every point is the space itself, $T_p\mathbb{R}^d \cong \mathbb{R}^d$. Further, note that the vectors $\Gamma_\mu^p \log_\mu X$ and $\Gamma_\nu^p \log_\nu Y$ have length equal to the distance between the data and their respective mean: $\|\Gamma_\mu^p \log_\mu X\|_p = \|\log_\mu X\|_\mu = d(\mu, X)$. So, our definition is a *centered* second moment. Definition 2.1 centers the second moment by subtracting the "mean" e.g., $\mathbb{E}\left[\log_p X\right]$. However, $\mathbb{E}\left[\log_p X\right] \neq \log_p \mathbb{E}\left[X\right]$, generally, i.e., the mean of the pushforward density onto the tangent space at an arbitrary point is not necessarily the $\log$ of the Fréchet mean of the density.

Proposition 3.6. *Let $\mathcal{H}$ denote the Hilbert space of Hilbert-Schmidt operators on $T_\nu M$, equipped with the inner product $\langle A, B\rangle_{\mathcal{H}} := \operatorname{tr}(AB^*)$. Then the Riemannian covariance between X and Y equals the Hilbert-Schmidt inner product between the transported intrinsic covariance operator $\widetilde{C}_{X,Y} = \mathbb{E}\big[(\Gamma_\mu^\nu \log_\mu X) \otimes \log_\nu Y\big] \in \mathcal{H}$ and the identity operator $\operatorname{id}_{T_\nu M}$.*

We characterize the behavior of our estimator in Proposition F.1. The proposed Definition 3.3 satisfies some standard results comparable to those of the Euclidean covariance such as generalization of the variance, symmetry, and equaling 0 for independent random variables.

3.2 Riemannian Correlation

Given the definition of Riemannian covariance, we introduce the notion of Riemannian correlation. This notion generalizes the Pearson correlation from Euclidean spaces to the manifold setting.

Definition 3.7 (Riemannian correlation)**.** *Let X and Y be random variables with respective means μ and ν satisfying Assumption 3.2. Let α be the geodesic with $\alpha(0) = \mu$ and $\alpha(1) = \nu$. For some $p \in \alpha$, the Riemannian cross-correlation at footpoint p is defined as*

$$\varrho_{X,Y}(p) = \frac{\Sigma_{X,Y}(p)}{\sqrt{\mathrm{Cov}_{X,X}(\mu)}\sqrt{\mathrm{Cov}_{Y,Y}(\nu)}},$$

and the corresponding Riemannian correlation coefficient is given by

$$\mathcal{R}_{X,Y}(p) = \mathrm{tr}(\varrho_{X,Y}(p)) = \frac{\mathrm{Cov}_{X,Y}(p)}{\sqrt{\mathrm{Cov}_{X,X}(\mu)}\sqrt{\mathrm{Cov}_{Y,Y}(\nu)}}.$$

Here, $\mathrm{Cov}_{X,X}(\mu)$ and $\mathrm{Cov}_{Y,Y}(\nu)$ are computed at μ and ν, respectively. However, benefiting from invariance under parallel transport, these quantities can be evaluated at any point. The properties of the Riemannian correlation, Proposition F.3, parallel those of the Riemannian covariance (i.e., Proposition F.1); we include these in Appendix F. A fundamental property of Riemannian correlation is its boundedness. Therefore, we have the following result, which is analogous to its Euclidean counterpart.

Proposition 3.8. *We have that $\mathcal{R}_{X,Y}(p) \in [-1, 1]$.*

3.3 Consistency and Asymptotic Normality

Given i.i.d. samples $\{(X_i, Y_i)\}_{i=1}^n$ and sample Fréchet means $\hat{\mu}, \hat{\nu}$, we first define the plug-in estimator $\widehat{\Sigma}_{X,Y}(p) \coloneqq \frac{1}{n}\sum_{i=1}^n \hat{U}_i \hat{V}_i^\top$, where $\hat{U}_i \coloneqq \Gamma_{\hat{\mu}}^p \log_{\hat{\mu}}(X_i)$ and $\hat{V}_i \coloneqq \Gamma_{\hat{\nu}}^p \log_{\hat{\nu}}(Y_i)$ at point p. We also define the scalar covariance $\mathrm{Cov}_{X,Y}(p) \coloneqq \mathrm{tr}(\Sigma_{X,Y}(p))$ and its estimator $\widehat{\mathrm{Cov}}_{X,Y}(p) \coloneqq \mathrm{tr}(\widehat{\Sigma}_{X,Y}(p))$.

We impose standard regularity conditions (A1)–(A6); the full statement is deferred to Assumption H.1 in Appendix H.1. In brief, (A1)–(A4) are the standard regularity conditions under which sample Fréchet means satisfy strong consistency and $\sqrt{n}$-asymptotics, while (A5)–(A6) ensure that the plug-in covariance admits a first-order expansion. Under these conditions, the following two lemmas will be used to establish the consistency and asymptotic properties of the estimators.

Lemma 3.9 (Strong consistency of sample Fréchet means (Bhattacharya and Patrangenaru, 2003, Thm. 2.3))**.** *Under (A1)–(A4), $\hat{\mu} \to \mu$ almost surely and $\hat{\nu} \to \nu$ almost surely.*

Lemma 3.10 ($\sqrt{n}$ expansion of sample Fréchet means (Bhattacharya and Patrangenaru, 2005, Thm. 2.1))**.** *Under (A1)–(A5), in $T_\mu\mathcal{M}$ and $T_\nu\mathcal{M}$ respectively,*

$$\sqrt{n}\,\log_\mu(\hat{\mu}) = H_X^{-1}\,\frac{1}{\sqrt{n}}\sum_{i=1}^n \xi_i + o_p(1), \qquad \sqrt{n}\,\log_\nu(\hat{\nu}) = H_Y^{-1}\,\frac{1}{\sqrt{n}}\sum_{i=1}^n \zeta_i + o_p(1), \tag{2}$$

where $\xi_i \in T_\mu\mathcal{M}$ and $\zeta_i \in T_\nu\mathcal{M}$ are i.i.d. mean-zero random vectors with finite second moments. Moreover, $\frac{1}{\sqrt{n}}\sum \xi_i \xrightarrow{d} \mathcal{N}(0, \Omega_X)$ and $\frac{1}{\sqrt{n}}\sum \zeta_i \xrightarrow{d} \mathcal{N}(0, \Omega_Y)$.

We then prove the strong consistency of the plug-in covariance estimator $\widehat{\Sigma}_{X,Y}(p)$ for a fixed footpoint $p \in \mathcal{U}$ and, hence, of its scalar trace $\widehat{\mathrm{Cov}}_{X,Y}(p)$. This almost-sure convergence justifies treating $\widehat{\Sigma}_{X,Y}(p)$ as a consistent plug-in quantity in the subsequent $\sqrt{n}$-asymptotic expansion.

Theorem 3.11 (Strong consistency)**.** *Under (A1)–(A4) and (A6), for any fixed $p \in \mathcal{U}$,*

$$\widehat{\Sigma}_{X,Y}(p) \xrightarrow{a.s.} \Sigma_{X,Y}(p).$$

Consequently, $\widehat{\mathrm{Cov}}_{X,Y}(p) \to \mathrm{Cov}_{X,Y}(p)$ almost surely.

Lemma 3.12 (Asymptotic linear representation of $\widehat{\Sigma}_{X,Y}(p)$)**.** *Assume (A1)–(A6) and fix $p \in \mathcal{U}$. Let* $\mathrm{vec}(\cdot)$ *stack columns of a $d \times d$ matrix into a d^2-vector and define*

$$\Delta_n := \mathrm{vec}(\widehat{\Sigma}_{X,Y}(p)) - \mathrm{vec}(\Sigma_{X,Y}(p)).$$

Then

$$\sqrt{n}\,\Delta_n = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \mathrm{IF}_i + o_p(1), \tag{3}$$

where

$$\mathrm{IF}_i := \Big(\mathrm{vec}(Z_i) - \mathbb{E}\mathrm{vec}(Z_i)\Big) + A_\mu H_X^{-1} \xi_i + A_\nu H_Y^{-1} \zeta_i. \tag{4}$$

Here, $Z_i := U_i V_i^\top$ with $U_i = \Gamma_\mu^p \log_\mu(X_i)$ and $V_i = \Gamma_\nu^p \log_\nu(Y_i)$, and A_μ, A_ν and ξ_i, ζ_i are as defined in (14)–(15) *and Lemma 3.10.*

Theorem 3.13 ($\sqrt{n}$-asymptotic normality)**.** *Under (A1)–(A6), for any fixed $p \in \mathcal{U}$,*

$$\sqrt{n}\Big(\mathrm{vec}(\widehat{\Sigma}_{X,Y}(p)) - \mathrm{vec}(\Sigma_{X,Y}(p))\Big) \xrightarrow{d} \mathcal{N}(0, \Xi),$$

where $\Xi = \mathrm{Var}(\mathrm{IF}_1) = \mathrm{Var}\left(\mathrm{vec}(Z_1) + A_\mu H_X^{-1} \xi_1 + A_\nu H_Y^{-1} \zeta_1\right)$.

Since $\widehat{\mathrm{Cov}}_{X,Y}(p) = \mathrm{tr}(\widehat{\Sigma}_{X,Y}(p))$ is a linear functional of $\widehat{\Sigma}_{X,Y}(p)$, the delta method is immediate. We hence have the corollary below stating the CLT for the scalar covariance.

Corollary 3.14 (CLT for scalar covariance)**.** *Under the conditions of Theorem 3.13,*

$$\sqrt{n}\Big(\widehat{\mathrm{Cov}}_{X,Y}(p) - \mathrm{Cov}_{X,Y}(p)\Big) \xrightarrow{d} \mathcal{N}\big(0,\ \sigma_{\mathrm{tr}}^2\big),$$

where $\sigma_{\mathrm{tr}}^2 = \mathrm{Var}\left(\langle \mathrm{vec}(I_d), \mathrm{IF}_1 \rangle\right)$ and I_d is the $d \times d$ identity matrix.

Remark 3.15 (Plug-in variance estimation and Wald inference). One can estimate Ξ by the sample covariance of estimated influence terms:

$$\widehat{\Xi} = \frac{1}{n} \sum_{i=1}^{n} \hat{\mathrm{IF}}_i \hat{\mathrm{IF}}_i^\top,$$

where $\hat{\mathrm{IF}}_i$ replaces $(\mu, \nu, H_X, H_Y, A_\mu, A_\nu, \xi_i, \zeta_i)$ by consistent estimators (e.g. empirical Hessians and empirical gradients for the Fréchet loss, and empirical averages for A_μ, A_ν). Then asymptotic Wald-type confidence regions follow:

$$\mathrm{vec}(\widehat{\Sigma}_{X,Y}(p)) \pm z_{\alpha/2}\, \frac{1}{\sqrt{n}} \sqrt{\mathrm{diag}(\widehat{\Xi})}.$$

For the trace, use $\hat{\sigma}_{\mathrm{tr}}^2 = \mathrm{vec}(I_d)^\top \widehat{\Xi}\, \mathrm{vec}(I_d)$.

Remark 3.16 (Correlation). The consistency and asymptotic normality properties of Riemannian cross-correlation $\hat{\varrho}_{X,Y}(p)$ and Riemannian correlation coefficient $\hat{\mathcal{R}}_{X,Y}(p)$ follow directly. This is since they are simply the compositions of continuous functions of $\widehat{\Sigma}_{X,Y}(p)$ and $\widehat{\mathrm{Cov}}_{X,Y}(p)$.

4 Examples

For the following simulations, we utilize code from ManOpt [Townsend et al.] and Geomstats [Miolane et al., 2020]. We include additional experiments in Appendix I.

4.1 Sphere

We first present some simulation studies on the unit sphere $\mathbb{S}^2$ to illustrate the behavior of Riemannian covariance and correlation, see Appendix B.1 for more details. The first dataset is sampled from a uniform distribution. A dependent dataset is then generated by rotating the first dataset, adding noise and transporting it to another point on the sphere, as detailed in Appendix I.1.

We compare our proposed Riemannian correlation with estimates from Abuqrais and Pigoli [2026] in different settings. Unless otherwise specified, we evaluate Riemannian correlation in the tangent

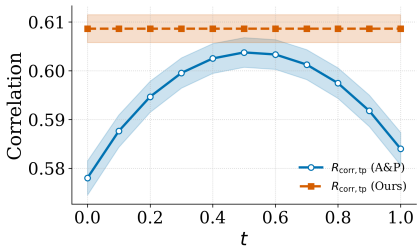

Figure 1: Geodesic Evaluation on Sphere. The shaded area indicates 95% confidence intervals (mean $\pm$ 2SE). The x-axis represents the location where the evaluation is taken along the geodesic connecting sample Fréchet means. Parameters: $n = 200$, $\tau = \pi/6$, $\eta = \pi/4$, $q = (0, -1, 0)$, $\sigma_\epsilon = 0.15$, MC $= 200$.

space of the Fréchet mean of the dependent dataset and evaluate $A\&P$ correlation in the tangent space of the midpoint of the geodesic connecting the two Fréchet means.

For Figure 1, we transport the dependent set to point $(0, -1, 0)$ and evaluate the correlation estimate along the geodesic connecting the Fréchet means of the two datasets. The shaded area denotes the 95% confidence interval based on two standard error. As shown in the plot, our estimate is numerically consistent along the geodesic, whereas the $A\&P$ correlation has an inverted U-shaped pattern with the largest estimate at 0.5 which represents the midpoint of the geodesic. This demonstrates that the $A\&P$ correlation is subject to the choice of footpoint, as discussed in Section 2.2, while our method is footpoint-invariant.

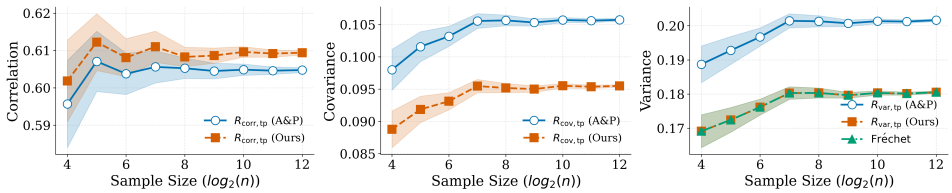

Figure 2: Comprehensive Comparison on Sphere with parallel-transported dataset. Correlation, covariance, and variance estimates versus sample size. Shaded regions indicate 95% confidence intervals. Parameters: $\tau = \pi/6$, $\eta = \pi/4$, $q = (0, -1, 0)$, $\sigma_\epsilon = 0.15$, MC $= 200$.

Figure 2 compares the correlation and covariance estimates from our method and those from $A\&P$ correlation. We transport the dependent set to the point $(0, -1, 0)$ and evaluate both methods at the midpoint, with different sample sizes. As before, the shaded region represents the confidence interval corresponding to two standard errors. The left plot shows the estimates of correlation, where both methods exhibit convergence as the sample size increases. Compared to our approach, $A\&P$ correlation yields a slightly smaller result. The two panels in the middle and right are the corresponding covariance and variance estimates used in computing the correlation. For the variance, we also include the estimate of Fréchet variance, computed using the Fréchet mean of the transported set. It can be seen that our method matches the Fréchet variance exactly, which is expected since parallel transport allows evaluation at the mean at all times. In contrast, the $A\&P$ method tends to produce a higher estimate, and could potentially lead to overestimation and bias.

4.2 Symmetric Positive Definite Matrices

We extend the simulation study to the manifold of Symmetric Positive Definite (SPD) matrices, $\mathcal{P}_d$, endowed with the Affine Invariant Riemannian Metric (AIRM), see Appendix B.2 for more details. This setting, characterized by non-positive (and non-constant) sectional curvature, provides a rigorous testbed for evaluating the proposed estimators. The data generation process, detailed in Appendix I.2, constructs correlated datasets by parallel transporting tangent vectors from a source distribution to a target location, thereby preserving the intrinsic geometric structure while varying the ambient configuration.

The fundamental property of footpoint invariance is first examined. Figure 9 illustrates the trajectory of correlation estimates as the evaluation point p traverses the geodesic connecting the sample means, for both the original dataset (non-transport) and the parallel-transported dataset (transport). We include this figure in Appendix I.2 as the results are similar to the spherical case in Figure 1. The proposed Riemannian correlation yields estimates with low variation (standard error $SE \approx 0.008$) across all 21 evaluation points, which is consistent with footpoint invariance (Theorem 3.5). In contrast, the A&P estimator exhibits substantial variation ($SE \approx 0.15$), revealing a systematic bias that depends on the evaluation point. This dependence arises because the logarithmic map $\log_p(\cdot)$ at varying footpoints distorts the metric structure differently in the presence of negative curvature. Our transport-based formulation mitigates this flaw by aligning tangent spaces via parallel transport, which preserves inner products under isometry. The comparison between the two panels demonstrates that our method produces nearly identical estimates for both original and parallel-transported datasets, which is consistent with the estimator capturing the intrinsic geometry invariant under isometry.

We next compare the estimators across sample sizes for both the original (non-transport) and parallel-transported (transport) datasets (Figures 10 and 3). For the non-transport dataset the two Fréchet means nearly coincide and the two methods agree. For the transport dataset, our variance matches the Fréchet variance (Proposition F.1, item 1), whereas the A&P variance, evaluated at the geodesic midpoint, underestimates it, with a ratio of 0.961 ± 0.001 that does not change with n. The A&P covariance is underestimated by nearly the same factor (ratio 0.960 ± 0.001), so the two effects cancel in the correlation and the A&P correlation stays close to ours. This agrees with the geometry of $\mathcal{P}_d$: under the AIRM the sectional curvature is non-positive, so $\log_p$ is non-expansive and a point cloud mapped to the tangent space at a point away from its own Fréchet mean is contracted. Our estimator avoids this by computing each log map at the variable's own Fréchet mean.

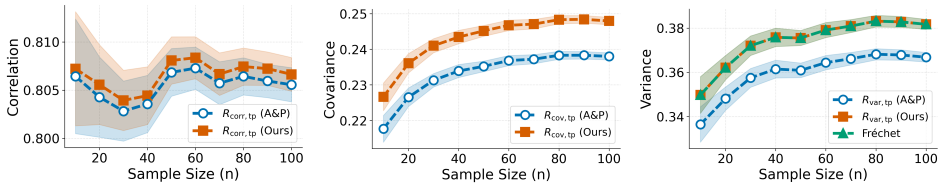

Figure 3: Comprehensive Comparison on the SPD Manifold with parallel-transported dataset. Correlation, covariance, and variance estimates versus sample size. Shaded regions indicate 95% confidence intervals. Parameters: $d = 3, \tau = 0.5, \sigma_\epsilon = 0.15$, MC $= 200$.

4.3 Kendall's Shape Space

Kendall's shape space [Kendall, 1984] is a geometric framework for shape analysis that treats objects as having the same shape regardless of their location, scale, and orientation. For more details, we refer to Appendix B.3. Kendall's shape space is identified with the quotient space $\mathcal{L}_{m,k}/\left(\mathbb{R}^m \rtimes (\mathbb{R}^+ \times SO(m))\right)$, with the translation group $\mathbb{R}^m$, scaling group $\mathbb{R}^+$, and rotation group $SO(m)$. Here $\rtimes$ denotes the semi-direct product of groups. In the special case of planar shapes $(m = 2)$, the shape space is a complex projective space $\mathbb{C}P^{k-2}$, which has a well-defined Riemannian geometry structure.

We evaluate our approach on the Sunnybrook Cardiac MRI dataset [Radau et al., 2009], which consists of two-dimensional cine magnetic resonance images (MRI) acquired in the short-axis view. The dataset includes four clinically distinct groups: Healthy controls (Normal), Hypertrophic cardiomyopathy (Hypertrophy), Heart failure without infarction (HF-Infarct), and Heart failure with infarction (HF+Infarct). Landmarks are extracted from the contour coordinates of each cardiac structure. We compute the correlation between the inner wall (endocardium) and outer wall (epicardium) of the left ventricular (LV) anatomy within each group. Our goal is to investigate whether the cardiac condition could affect the LV structure and symmetry.

Table 1 compares the correlation estimates from our proposed approach, $A\&P$ correlation, and CCA. As in the previous analysis, the $A\&P$ correlation is computed at the midpoint of the geodesic connecting the two sample Fréchet means. The proposed estimates are similar to those obtained using the $A\&P$ method, whereas CCA yields correlations equal to 1 across all four groups. This

Table 1: Correlation estimates for the four groups using our proposed method, the $A\&P$ correlation, and canonical correlation analysis (CCA).

	Shape Space		Preshape Space		
	R_{corr}(Ours)	$R_{corr}(A\&P)$	R_{corr}(Ours)	$R_{corr}(A\&P)$	CCA
Normal ($n = 9$)	0.637	0.637	0.253	0.253	1
Hypertrophy ($n = 12$)	0.393	0.394	-0.070	-0.069	1
HF-infarct ($n = 12$)	0.846	0.846	0.442	0.442	1
HF+infarct ($n = 12$)	0.659	0.659	0.405	0.405	1

reflects an inherent limitation of CCA: The sample canonical correlations are the cosines of the principal angles between the subspaces spanned by the two datasets. If the sum of their dimensions exceeds the dimension of the ambient space, these subspaces necessarily have a nontrivial intersection. Consequently, at least one canonical correlation equals one, indicating that there exist directions along which the two datasets are perfectly aligned.

Similar shapes in shape space are expected to exhibit a high degree of association. The estimated Riemannian correlation results suggest that the HF-infarct group has the highest correlation, while the hypertrophy group shows a relatively weak association. As illustrated in Figure 14, the inner and outer walls of the LV in the HF-infarct group are expected to remain geometrically symmetric. In contrast, MRI images of patients with hypertrophy typically display imbalanced wall thickness, leading to a disruption of symmetry. These observations are consistent with existing findings in the literature. For instance, Maron et al. [2014] reports that in hypertrophic cardiomyopathy, the inner wall becomes irregular and chaotic, whereas the outer wall remains relatively smooth, thereby breaking the geometric coupling structure. Moreover, Eshaghian et al. [2007] states that amyloid infiltration of the heart can lead to a uniformly thickened wall with a firm, rubbery consistency, resulting in heart failure without infarction. Overall, our proposed estimator provides reasonable results and presents a quantitative measure of structural coupling in shape geometry.

5 Discussion

We have introduced two novel definitions, the intrinsic footpoint-invariant Riemannian cross-covariance and the subsequent Riemannian correlation. These tools are fundamental for understanding the second-order information present in manifold-valued random objects. Our estimators enjoy many desirable properties (e.g., Proposition F.1), which makes them a robust and natural choice for manifold statistical analysis. Lastly, their asymptotic properties are vital for quantifying their behavior. The experimental results on the sphere, SPD matrices, and Kendall's shape space show the applicability of our method and validate its properties. Given its effectiveness, the method can further the area of geometric learning on Riemannian manifolds, including regression and principal component analysis.

The correlation matrix $\varrho_{X,Y}(p)$ introduced in Definition 3.7 differs from its conventional counterpart, as its diagonal entries do not represent correlations between the corresponding components in the chosen basis. A Riemannian analogue of the conventional correlation matrix could instead be defined as $D_X^{-1}\Sigma_{X,Y}D_Y^{-1}$, where $D_X = \text{diag}(\Sigma_{X,X})^{1/2}$ and $D_Y = \text{diag}(\Sigma_{Y,Y})^{1/2}$, thereby normalizing each covariance entry by the corresponding marginal standard deviations. However, the absolute value of the trace of this matrix is not generally bounded by 1. Our definition is therefore chosen to yield a coefficient that appropriately characterizes the association between the two objects.

We also note in regards to our central limit theorems that generalizations of Stein's method on Riemannian manifolds are not directly applicable nor straightforward. There are some recent works such as Qu and Vemuri [2022] and Le et al. [2024] which have made advancements towards this generalization, but a challenge with a generalized Stein's method is that the Gaussian distribution is defined over a Euclidean space not the manifold itself. Further, we explored a few manifolds but many others can be additionally considered. Lastly, our method can be extended to Lie groups, where the natural choice for a tangent space is the Lie algebra and one can consider the group logarithm instead.

Acknowledgements

This work was in part supported by the National Science Foundation [CAREER Award CMMI-2543434, EAGER Award CMMI-2430998] and the American Heart Association Collaborative Science [Award 23CSA1052735].

References

Meshal Abuqrais and Davide Pigoli. A riemannian covariance for manifold-valued data. *Journal of Mathematical Imaging and Vision*, 68(2):7, 2026.

Bijan Afsari. Riemannian l^{p} center of mass: existence, uniqueness, and convexity. *Proceedings of the American Mathematical Society*, 139(2):655–673, 2011.

Galen Andrew, Raman Arora, Jeff Bilmes, and Karen Livescu. Deep canonical correlation analysis. In *International conference on machine learning*, pages 1247–1255. PMLR, 2013.

Lorenzo Basile, Santiago Acevedo, Luca Bortolussi, Fabio Anselmi, and Alex Rodriguez. Intrinsic dimension correlation: uncovering nonlinear connections in multimodal representations. In *International Conference on Learning Representations*, volume 2025, pages 31989–32014, 2025.

Rabi Bhattacharya and Vic Patrangenaru. Large sample theory of intrinsic and extrinsic sample means on manifolds. *The Annals of Statistics*, 31(1):1–29, 2003. doi: 10.1214/aos/1046294456.

Rabi Bhattacharya and Vic Patrangenaru. Large sample theory of intrinsic and extrinsic sample means on manifolds—ii. *The Annals of Statistics*, 33(3):1225–1259, 2005. doi: 10.1214/009053605000000093.

James Buenfil and Eardi Lila. Asymmetric canonical correlation analysis of riemannian and high-dimensional data. *Electronic journal of statistics*, 19(2):6077, 2025.

Eduardo Castro, joycenv, Mustafa Cetin, Navin Cota, Rogers F Silva, and Vince. Mlsp 2014 schizophrenia classification challenge. `https://kaggle.com/competitions/mlsp-2014-mri`, 2014. Kaggle.

Hui Cheng and Krishna C Gupta. An historical note on finite rotations. *Journal of Applied Mechanics*, 56(1):139–145, 1989.

Mengyu Dai, Zhengwu Zhang, and Anuj Srivastava. Analyzing dynamical brain functional connectivity as trajectories on space of covariance matrices. *IEEE transactions on medical imaging*, 39 (3):611–620, 2019.

Manfredo Perdigao do Carmo. *Riemannian geometry*, volume 2. Springer, 1992.

Shervin Eshaghian, Sanjay Kaul, and Prediman K Shah. Cardiac amyloidosis: new insights into diagnosis and management. *Reviews in Cardiovascular Medicine*, 8(4):189–199, 2007.

P Thomas Fletcher, Conglin Lu, Stephen M Pizer, and Sarang Joshi. Principal geodesic analysis for the study of nonlinear statistics of shape. *IEEE transactions on medical imaging*, 23(8):995–1005, 2004.

Maurice Fréchet. Les éléments aléatoires de nature quelconque dans un espace distancié. In *Annales de l'institut Henri Poincaré*, volume 10, pages 215–310, 1948.

Tian Fu, Tianliang Yao, Haoyu Wang, and Bin Chen. Adaptive electromagnetic analysis via non-euclidean manifold learning for atmospheric precipitation understanding. *IEEE Geoscience and Remote Sensing Letters*, 2025.

Nicolas Guigui, Elodie Maignant, Alain Trouvé, and Xavier Pennec. Parallel transport on kendall shape spaces. In *International Conference on Geometric Science of Information*, pages 103–110. Springer, 2021.

Hermann Karcher. Riemannian center of mass and mollifier smoothing. *Communications on pure and applied mathematics*, 30(5):509–541, 1977.

David G Kendall. Shape manifolds, procrustean metrics, and complex projective spaces. *Bulletin of the London mathematical society*, 16(2):81–121, 1984.

Wilfrid S Kendall. Probability, convexity, and harmonic maps with small image i: uniqueness and fine existence. *Proceedings of the London Mathematical Society*, 3(2):371–406, 1990.

Hyunwoo J Kim, Nagesh Adluru, Barbara B Bendlin, Sterling C Johnson, Baba C Vemuri, and Vikas Singh. Canonical correlation analysis on riemannian manifolds and its applications. In *European conference on computer vision*, pages 251–267. Springer, 2014.

Huiling Le, Alexander Lewis, Karthik Bharath, and Christopher Fallaize. A diffusion approach to stein's method on riemannian manifolds. *Bernoulli*, 30(2):1079–1104, 2024.

John M Lee. *Introduction to Riemannian manifolds*, volume 2. Springer, 2018.

Zhenhua Lin and Fang Yao. Intrinsic Riemannian functional data analysis. *The Annals of Statistics*, 47(6):3533 – 3577, 2019. doi: 10.1214/18-AOS1787.

Na Lu and Hongyu Miao. Clustering tree-structured data on manifold. *IEEE transactions on pattern analysis and machine intelligence*, 38(10):1956–1968, 2015.

Russell Lyons. Distance covariance in metric spaces. *The Annals of Probability*, pages 3284–3305, 2013.

Kanti V Mardia, Jesper Illemann Foldager, and Jes Frellsen. Directional statistics in protein bioinformatics. In *Applied Directional Statistics*, pages 17–40. Chapman and Hall/CRC, 2018.

Kantilal Varichand Mardia. Statistics of directional data. *Journal of the Royal Statistical Society Series B: Statistical Methodology*, 37:349–371, 1975.

Barry J Maron, Steve R Ommen, Christopher Semsarian, Paolo Spirito, Iacopo Olivotto, and Martin S Maron. Hypertrophic cardiomyopathy: present and future, with translation into contemporary cardiovascular medicine. *Journal of the American College of Cardiology*, 64(1):83–99, 2014.

Nina Miolane, Nicolas Guigui, Alice Le Brigant, Johan Mathe, Benjamin Hou, Yann Thanwerdas, Stefan Heyder, Olivier Peltre, Niklas Koep, Hadi Zaatiti, et al. Geomstats: A python package for riemannian geometry in machine learning. *Journal of Machine Learning Research*, 21(223):1–9, 2020.

Wenliang Pan, Xueqin Wang, Canhong Wen, Martin Styner, and Hongtu Zhu. Conditional local distance correlation for manifold-valued data. In *International Conference on Information Processing in Medical Imaging*, pages 41–52. Springer, 2017.

Xavier Pennec. Intrinsic statistics on riemannian manifolds: Basic tools for geometric measurements. *Journal of Mathematical Imaging and Vision*, 25(1):127–154, 2006.

Alexander Petersen and Hans-Georg Müller. Fréchet regression for random objects with euclidean predictors. *The Annals of Statistics*, 47(2):691–719, 2019a.

Alexander Petersen and Hans-Georg Müller. Wasserstein covariance for multiple random densities. *Biometrika*, 106(2):339–351, 2019b.

Xiaoda Qu and Baba C Vemuri. A framework for improving the characterization scope of stein's method on riemannian manifolds. *arXiv preprint arXiv:2209.08424*, 2022.

Perry Radau, Yingli Lu, Kim Connelly, Gideon Paul, Alexander J Dick, and Graham A Wright. Evaluation framework for algorithms segmenting short axis cardiac mri. *The MIDAS Journal*, 2009.

C Radhakrishna Rao et al. Information and the accuracy attainable in the estimation of statistical parameters. *Bull. Calcutta Math. Soc*, 37(3):81–91, 1945.

Lingxuan Shao, Zhenhua Lin, and Fang Yao. Intrinsic riemannian functional data analysis for sparse longitudinal observations. *The Annals of Statistics*, 50(3):1696–1721, 2022.

James Townsend, Niklas Koep, and Sebastian Weichwald. Pymanopt: A python toolbox for optimization on manifolds using automatic differentiation. *Journal of Machine Learning Research*, (137): 1–5.

A Limitations

Throughout the paper we make Assumptions 3.1,3.2,H.1. Assumption 3.1 is a standard, necessary assumption for mean estimation on metric spaces. It is crucial to have unique means as we define our estimate on the geodesic between the means. For non-positively curved manifolds, this assumption poses no limitations. For positively curved manifolds, however, it poses a restriction. One can easily determine the radius of empirical data, however. This can be done, for instance by computing $d_{\mathcal{M}}(x_i, x_j)$ for all i, j and denoting that as the diameter. The Fréchet mean on positively curved manifolds is, generally, not robust. While we acknowledge this is a limitation, this is inherently a limitation of the Fréchet mean rather than directly of our proposed definitions.

Assumption 3.2 is necessary so that we have a unique geodesic between the means. This can be verified by determining the distance between the means and comparing this to the injectivity radius.

Assumption H.1 has some regularity conditions necessary for the asymptotic behavior. These are needed to avoid extreme cases such as having infinite curvature and infinite variance. Further they detail that we assume some continuity and differentiability which are, again, typical and non restrictive assumptions.

Moreover, our method is intrinsically defined on a manifold and therefore requires all datasets to lie on the same manifold. This assumption is essential because the method relies on establishing a relationship between the metrics of different datasets. Consequently, our approach is not applicable to datasets on different metric spaces or with different dimensions.

B Riemannian Geometry

Here we elaborate on more geometric tools which we require, and again we note that details can be found in sources such as Lee [2018] and do Carmo [1992].

The collection of all tangent spaces is referred to as the tangent bundle, $T\mathcal{M} = \{(p, v) | v \in T_p\mathcal{M}, p \in \mathcal{M}\}$. As we require the exponential map to be invertible, we need to consider on which domain it is as such. Riemannian manifolds are locally diffeomorphic to a Euclidean space, and the injectivity radius formally quantifies this locality. The injectivity radius at $p \in \mathcal{M}$, denoted as $\mathrm{inj}(p)$, is the largest (supremum) radius of an open ball in $T_p\mathcal{M}$ centered at the origin such that $\exp_p$ is defined and B_r and $\exp_p(B_r)$ are diffeomorphic. The injectivity radius of the manifold, $\mathrm{inj}\,\mathcal{M}$, is the smallest injectivity radius on the manifold, $\mathrm{inj}\,\mathcal{M} = \inf\{\mathrm{inj}(p) | p \in \mathcal{M}\}$.

A smooth vector field V is said to be parallel along a smooth path α if $\frac{D}{dt}V = 0$, with the covariant derivative defined with respect to the Riemannian connection. Given a smooth path $\alpha : [0, 1] \to \mathcal{M}$, $t \in [0, 1]$ with $u \in T_{\alpha(t)}\mathcal{M}$, one can show that there exists a unique parallel vector field V such that $V(t) = u$ [Lee, 2018, Theorem 4.32]. This however is still dependent on the chosen path. However, if we restrict to geodesics, and a unique geodesic exists, the parallel vector field is uniquely determined.

The Riemannian volume form $\mathrm{d}\omega$ is defined as $\mathrm{d}\omega = \sqrt{|g|}\,\mathrm{d}x^1 \wedge \cdots \wedge \mathrm{d}x^d$ where $|g|$ is the determinant of the matrix representation of the metric in local coordinates. The Riemannian volume form allows us to define densities with respect to this base measure.

For each manifold we require an exponential map, the logarithm map, and the parallel transport. These are standard tools, and we include them next for completeness.

B.1 Sphere

Let $\mathcal{M} = \mathbb{S}^d$ be the d-dimensional unit sphere in $\mathbb{R}^{d+1}$ with the metric induced from the ambient space. The exponential map is given by

$$\exp_p(v) = \cos(\|v\|)\, p + \sin(\|v\|)\, \frac{v}{\|v\|}.$$

The logarithm map is given by

$$\log_p(q) = \frac{\theta}{\sin(\theta)}\, (q - \cos(\theta)p),$$

with $q \notin \{p, -p\}$ and $\theta = \arccos\langle p, q\rangle$. Lastly, the parallel transport of a vector v along a geodesic from p to q is given by

$$\Gamma_p^q(v) = v - \frac{\langle \log_p q, v\rangle_p}{\theta^2}\,(\exp_p^{-1} q + \exp_q^{-1} p),$$

with θ as above.

B.2 SPDM

Let $\mathcal{M}$ be the space of SPD matrices with the affine invariant metric (the Rao-Fisher metric). The exponential map is given by

$$\exp_p(v) = p^{1/2}\,\mathrm{Exp}\Big(p^{-1/2} v p^{-1/2}\Big)\,p^{1/2},$$

and the log map is given by

$$\log_q(p) = q^{1/2}\,\mathrm{Log}\Big(q^{-1/2} p\, q^{-1/2}\Big)\,q^{1/2},$$

where Exp and Log are the matrix exponential and logarithm, respectively. The parallel transport of tangent vector $v \in T_p\mathcal{M}$ along a geodesic connecting p and q is given by

$$\Gamma_p^q v = (qp^{-1})^{1/2} v (p^{-1}q)^{1/2}.$$

B.3 Kendall's shape space

Define the ordered k-tuple landmark space

$$\mathcal{L}_{m,k} = \{X \in \mathbb{R}^{m\times k} |\, \dim(\mathrm{span}(X)) = m, k > m\}.$$

For an object $X \in \mathcal{L}_{m,k}$, we first remove the effect of translation and scaling

$$\hat{X} = \frac{X - \bar{x}\,\mathbf{1}_k^T}{\|X - \bar{x}\,\mathbf{1}_k^T\|_F},$$

where $\bar{x} \in \mathbb{R}^m$ is the mean across landmarks (i.e., the row-wise mean over the k columns) and $\|\cdot\|_F$ is the Frobenius norm. This defines the *preshape* space

$$\mathcal{S}_m^k = \Big\{X \in \mathcal{L}_{m,k} \,\Big|\, X\mathbf{1}_k = 0,\ \|X\|_F = 1\Big\},$$

which lies on a hypersphere $\mathcal{S}^{m(k-1)-1}$. The Kendall's shape space is further defined by taking quotient space of preshape space under the rotation group

$$\Sigma_m^k = \mathcal{S}_m^k / SO(m).$$

For each element $\tilde{X} \in \Sigma_m^k$, there corresponds an orbit in preshape space $\tilde{X} = [\hat{X}] = \{O\hat{X} | O \in SO(m)\}$. Therefore, the exponential map in Σ_m^k is defined (on a chosen representative $\hat{X}_1$ of $\tilde{X}_1$) by

$$\exp_{\tilde{X}_1}(v) = \cos(\|v\|_F)\hat{X}_1 + \sin(\|v\|_F)\frac{v}{\|v\|_F}.$$

The logarithm map is given by

$$\log_{\tilde{X}_1}(\tilde{X}_2) = \frac{\theta}{\sin(\theta)}(O^*\hat{X}_2 - \cos(\theta)\hat{X}_1),$$

where $\theta = \arccos(\langle \hat{X}_1, O^*\hat{X}_2\rangle_F)$ and $O^* = \mathrm{argmax}_{O\in SO(m)}\langle \hat{X}_1, O\hat{X}_2\rangle_F$. The spherical parallel transport cannot be applied on Kendall's shape space directly, but the transported vector can be approximated using a pole ladder algorithm [Guigui et al., 2021].

C Canonical Correlation

Here we include more details from Kim et al. [2014]. The geodesic submanifold Fletcher et al. [2004] $S \subset \mathcal{M}$ is defined as $S = \exp_p(\text{span}(\{b_i\}) \cap U)$ with $U \subset T_p\mathcal{M}$ and $b_i \in T_p\mathcal{M}$. The geodesic submanifold is the subspace of $\mathcal{M}$ generated by pushing the span of some vectors $\{b_i\}$ onto the manifold. Note that when $|\{b_i\}| = 1$ then S is simply a geodesic. Note that once an appropriate $\mathcal{R}^{CCA}_{X,Y}$ is obtained, one can determine the covariance by multiplying by the standard deviations of both sets of coefficients.

Notice that projection onto the geodesic submanifold can be expressed in terms of the inner product, thus using Euclidean metric, the Log-Euclidean framework of CCA is actually equivalent to the manifold-valued CCA:

$$\begin{aligned}\mathcal{R}^{CCA}_{X,Y} &= \max_{W_x,W_y} \frac{\sum_i \langle W_x, \log_{\mu_x}(x_i)\rangle \langle W_y, \log_{\mu_y}(y_i)\rangle}{\sqrt{\sum_i \langle W_x, \log_{\mu_x}(x_i)\rangle^2}\sqrt{\sum_i \langle W_y, \log_{\mu_y}(y_i)\rangle^2}} \\ &= \max_{W_x,W_y} \frac{W_x^T \operatorname{Cov}(X,Y)\, W_y}{\sqrt{W_x^T \operatorname{Cov}(X,X)\, W_x}\sqrt{W_y^T \operatorname{Cov}(Y,Y)\, W_y}},\end{aligned}$$

where $\operatorname{Cov}(X,Y) = \frac{1}{n}\sum_i [(\log_{\mu_x} x_i)(\log_{\mu_y} y_i)^T]$ and similar for $\operatorname{Cov}(X,X)$ and $\operatorname{Cov}(Y,Y)$. Unlike the general cross-covariance, which accounts for all possible linear combination between datasets, the CCA computes a normalized maximal cross-correlation, thus it tends to overestimate the correlation when datasets share the same coordinates.

D Euclidean Covariance and Correlation

Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space and let X and Y be two integrable random variables such that $X, Y \in L^1(\Omega, \mathcal{F}, \mathbb{P})$. The expectation of X is defined as

$$\mathbb{E}(X) = \int_\Omega X(\omega)\, \mathrm{d}\mathbb{P}(\omega),$$

and similarly for $\mathbb{E}(Y)$. The covariance of X and Y is given by

$$\operatorname{Cov}(X,Y) = \mathbb{E}((X - \mathbb{E}(X))(Y - \mathbb{E}(Y))).$$

Equivalently,

$$\operatorname{Cov}(X,Y) = \mathbb{E}(XY) - \mathbb{E}(X)\mathbb{E}(Y).$$

The covariance has the following properties:

1. Variance: $\operatorname{Cov}(X,X) = \operatorname{Var}(X)$;
2. Symmetry: $\operatorname{Cov}(X,Y) = \operatorname{Cov}(Y,X)$;
3. Scaling: $\operatorname{Cov}(aX,Y) = a\operatorname{Cov}(X,Y)$;
4. Uncorrelatedness: If X and Y are independent, then $\operatorname{Cov}(X,Y) = 0$;
5. Invariance to constant: $\operatorname{Cov}(X+c,Y) = \operatorname{Cov}(X,Y)$;
6. Variance of a constant: $\operatorname{Cov}(X,c) = 0$;
7. Additivity: $\operatorname{Cov}(X_1+X_2,Y) = \operatorname{Cov}(X_1,Y) + \operatorname{Cov}(X_2,Y)$. More generally,

$$\operatorname{Cov}\left(\sum_i a_i X_i, \sum_j b_j Y_j\right) = \sum_i \sum_j a_i b_j \operatorname{Cov}(X_i, Y_j).$$

With the variance given by

$$\operatorname{Var}(X) = \operatorname{Cov}(X,X),$$

the Pearson correlation is defined as

$$\rho(X,Y) = \frac{\operatorname{Cov}(X,Y)}{\sqrt{\operatorname{Var}(X)\operatorname{Var}(Y)}}.$$

Thus we have the following properties for correlation:

1. Bounded: $\rho(X, Y) \in [-1, 1]$;
2. Symmetry: $\rho(X, Y) = \rho(Y, X)$;
3. Scale invariance: $\rho(aX + b, cY + d) = \text{sign}(ac)\rho(X, Y)$;
4. Self-correlation: if X is nondegenerate, $\rho(X, X) = 1$.

E Covariance Under Basis Rotation

The parallel transport can be interpreted as a rotation of basis when the manifold is viewed as embedded in some ambient space, i.e. $\mathbb{R}^N$. Suppose that we want to move a vector w from the tangent space at $p \in \mathcal{M}$ to the tangent space at $q \in \mathcal{M}$ along the geodesic connecting p and q. The vector $w \in T_p\mathcal{M}$ can be written as

$$w = \sum_i a_i e_i^p,$$

where $\{e_i^p\}$ is a basis of $T_p\mathcal{M}$. Under parallel transport, the corresponding basis on $T_q\mathcal{M}$ is mapped by $e_i^q = De_i^p$, where $D \in SO(N)$, and the transported vector $w' \in T_q\mathcal{M}$ is

$$w' = \sum_i a_i e_i^q = Dw.$$

Therefore, we identify D with the parallel transport matrix, denoted $D \simeq \Gamma_\mu^\nu$.

Under this setting, for two random variables $X, Y \in \mathcal{M}$ with respective means μ, ν, we have random vectors $u = \log_\mu X \in T_\mu\mathcal{M}$ and $v = \log_\nu Y \in T_\nu\mathcal{M}$. Let $\gamma : [0, 1] \to \mathcal{M}$ be a geodesic with $\gamma(0) = \mu$, $\gamma(1) = \nu$, and for some $t \in [0, 1]$, the corresponding matrices for parallel transport are given by

$$A \simeq \Gamma_{\gamma(0)}^{\gamma(1)}, \quad B(t) \simeq \Gamma_{\gamma(0)}^{\gamma(t)}, \quad C(t) \simeq \Gamma_{\gamma(t)}^{\gamma(1)},$$

where $A, B, C \in SO(N)$, and hence we also have $A = C(t)B(t)$ and

$$A^T \simeq \Gamma_{\gamma(1)}^{\gamma(0)}, \quad B^T(t) \simeq \Gamma_{\gamma(t)}^{\gamma(0)}, \quad C^T(t) \simeq \Gamma_{\gamma(1)}^{\gamma(t)}.$$

The Riemannian cross-covariance at $p = \alpha(t)$ can be defined alternatively as

$$\begin{aligned}
\Sigma_{X,Y}(p) &= \mathbb{E}((\Gamma_\mu^p \log_\mu X)(\Gamma_\nu^p \log_\nu Y)^T) \\
&= \mathbb{E}((\Gamma_{\gamma(0)}^{\gamma(t)} u)(\Gamma_{\gamma(1)}^{\gamma(t)} v)^T) \\
&= \mathbb{E}((B(t)u)(C(t)^T v)^T) \\
&= \mathbb{E}(B(t)uv^T C(t)).
\end{aligned}$$

The covariance matrix itself depends on the choice of p; however, for the Riemannian covariance we have

$$\begin{aligned}
\text{Cov}_{X,Y}(p) &= \text{tr}(\Sigma_{X,Y}(p)) \\
&= \text{tr}\left(\mathbb{E}(B(t)uv^T C(t))\right) \\
&= \mathbb{E}\left(\text{tr}(C(t)B(t)uv^T)\right) \\
&= \mathbb{E}\left(\text{tr}(Auv^T)\right) \\
&= \text{tr}\left(\mathbb{E}\left[(\Gamma_\mu^\nu \log_\mu X)(\log_\nu Y)^T\right]\right) \\
&= \text{Cov}_{X,Y}(\nu),
\end{aligned}$$

similarly, we also have $\text{Cov}_{X,Y}(p) = \text{Cov}_{X,Y}(\mu)$. Since it holds for any $p = \alpha(t)$, it is enough to show the invariance of Riemannian covariance along the geodesic α.

F Riemannian Covariance Correlation Properties

Proofs for the following propositions can be found in Appendix G.

Proposition F.1. *The Riemannian Covariance as in Definition 3.3 has the following properties:*

1. $\mathrm{Cov}_{X,X}(\mu) = \sigma_X^2$ *(Fréchet variance special case);*
2. $\mathrm{Cov}_{X,Y}(p) = \mathrm{Cov}_{Y,X}(p)$ *(Symmetry);*
3. *if* $\mathrm{Cov}_{X,X}(\mu) = 0$*, then* X *is degenerate (Degeneracy);*
4. *if* X *and* Y *are independent, then* $\mathrm{Cov}_{X,Y}(p) = 0$ *(Uncorrelatedness);*
5. *for a fixed point* $q \in \mathcal{M}$*,* $\mathrm{Cov}_{X,q}(q) = 0$ *(Covariance of a constant);*
6. $\mathrm{Cov}_{(a,X),Y}(p) = a\,\mathrm{Cov}_{X,Y}(p)$ *where* (a, X) *is* $\exp_\mu(a \log_\mu(X))$ *and* $a \in \mathbb{R}$ *(Scaling).*

Remark F.2. As multiplication is not defined on Riemannian manifolds, item 6 of Proposition F.1 requires us to define what it means to scale a random variable. We propose a definition and mention the implications with regard to our assumptions in the proof. Further, items 6 and 2 in conjunction with 1 yield $\mathrm{Cov}_{(a,X),(a,X)}(\mu) = a^2\,\mathrm{Cov}_{X,X}(\mu) = a^2\sigma_X^2$.

Proposition F.3. *The Riemannian correlation coefficient as in Definition 3.7 has the following properties:*

1 $\mathcal{R}_{X,Y}(p) = \mathcal{R}_{X,Y}(q)$ *(Invariance);*

2 $\mathcal{R}_{X,Y}(p) = \mathcal{R}_{Y,X}(p)$ *(Symmetry);*

3 *If* X *and* Y *are independent, then* $\mathcal{R}_{X,Y}(p) = 0$ *(Uncorrelatedness);*

4 $\mathcal{R}_{(a,X),Y}(p) = \mathit{sign}(a)\mathcal{R}_{X,Y}(p)$ *(Scale invariance);*

5 *If X is nondegenerate,* $\mathcal{R}_{X,X}(p) = 1$ *(Self-correlation).*

G Additional Proofs

Proof of Proposition 2.2. Fix two distinct points $p, q \in \mathcal{M}$, and write $\xi = \log_p(q)$. Since $\mathcal{M}$ has nonzero curvature, the logarithmic map does not commute with parallel transport. In particular, for any x in a sufficiently small normal neighborhood of p, the logarithmic map admits the expansion

$$\log_q(x) = \Gamma_p^q\big(\log_p(x) - \xi\big) + C_p(\log_p(x), \xi) + O\big(\|\log_p(x)\|^3 + \|\xi\|^3\big)\,, \tag{5}$$

where Γ_p^q denotes parallel transport along the geodesic from p to q, the leading term $-\Gamma_p^q\xi$ is the translation in the basepoint (present already in the Euclidean limit), and $C_p(\cdot,\cdot)$ is a bilinear term involving the Riemann curvature tensor at p. The curvature term vanishes identically if and only if $\mathcal{M}$ is flat.

Applying (5) to X and Y and centering, the translation $-\Gamma_p^q\xi$ cancels in the centered second moment, so

$$\Sigma_{X,Y}^{A\&P}(q) = \Gamma_p^q\,\Sigma_{X,Y}^{A\&P}(p)\,(\Gamma_p^q)^\top + D_{XY},$$

where the correction term D_{XY} collects the curvature contributions

$$\begin{aligned} D_{XY} = \mathbb{E}\big[&\bar{C}_p(\log_p X, \xi)(\log_p Y - \mathbb{E}\log_p Y)^\top \\ &+ (\log_p X - \mathbb{E}\log_p X)\bar{C}_p(\log_p Y, \xi)^\top\big] + O(\cdot), \end{aligned}$$

with $\bar{C}_p$ involving the Riemann curvature tensor at p.

Taking traces and using that parallel transport is an isometry (so $\mathrm{tr}(\Gamma_p^q A(\Gamma_p^q)^\top) = \mathrm{tr}(A)$),

$$\mathrm{tr}\big(\Sigma_{X,Y}^{A\&P}(q)\big) = \mathrm{tr}\big(\Sigma_{X,Y}^{A\&P}(p)\big) + \mathrm{tr}(D_{XY}).$$

Since the curvature tensor is nonzero, one can choose distributions of X and Y for which $\mathrm{tr}(D_{XY}) \neq 0$. Consequently,

$$\mathrm{tr}\big(\Sigma_{X,Y}^{A\&P}(q)\big) \neq \mathrm{tr}\big(\Sigma_{X,Y}^{A\&P}(p)\big),$$

which proves that the Abuqrais-Pigoli covariance depends on the choice of footpoint. □

Proof of Theorem 3.5.

$$\begin{aligned}
\operatorname{tr}\left[\Sigma_{X,Y}(p)\right] &= \operatorname{tr}\left[\mathbb{E}\left[(\Gamma_\mu^p \log_\mu X)(\Gamma_\nu^p \log_\nu Y)^T\right]\right] \\
&= \mathbb{E}\left[\operatorname{tr}\left[(\Gamma_\mu^p \log_\mu X)(\Gamma_\nu^p \log_\nu Y)^T\right]\right] \\
&= \mathbb{E}\left[\operatorname{tr}\left[(\Gamma_p^q \Gamma_\mu^p \log_\mu X)(\Gamma_p^q \Gamma_\nu^p \log_\nu Y)^T\right]\right] \\
&= \mathbb{E}\left[\operatorname{tr}\left[(\Gamma_\mu^q \log_\mu X)(\Gamma_\nu^q \log_\nu Y)^T\right]\right] \\
&= \operatorname{tr}\left[\mathbb{E}\left[(\Gamma_\mu^q \log_\mu X)(\Gamma_\nu^q \log_\nu Y)^T\right]\right] \\
&= \operatorname{tr}\left[\Sigma_{X,Y}(q)\right].
\end{aligned}$$

Here the diagonal elements of $\left[(\Gamma_\mu^p \log_\mu X)(\Gamma_\nu^p \log_\nu Y)^T\right]$ are $\langle \Gamma_\mu^p \log_\mu X, e_i\rangle \cdot \langle \Gamma_\nu^p \log_\nu Y, e_i\rangle$ where $\{e_i\}$ is a local orthonormal basis in $T_p\mathcal{M}$. Noting that parallel transport is done with respect to a basis and thus applying the parallel transport we have $\langle \Gamma_p^q \Gamma_\mu^p \log_\mu X, \Gamma_p^q e_i\rangle \cdot \langle \Gamma_p^q \Gamma_\nu^p \log_\nu Y, \Gamma_p^q e_i\rangle$. Since p, q, μ, ν are all on the same geodesic we have that $\langle \Gamma_p^q \Gamma_\mu^p \log_\mu X, \Gamma_p^q e_i\rangle \cdot \langle \Gamma_p^q \Gamma_\nu^p \log_\nu Y, \Gamma_p^q e_i\rangle = \langle \Gamma_\mu^q \log_\mu X, \Gamma_p^q e_i\rangle \cdot \langle \Gamma_\nu^q \log_\nu Y, \Gamma_p^q e_i\rangle$. Parallel transport preserves inner products thus the trace (sum) is unaffected. Note that since p, q, μ are on the same geodesic $\Gamma_p^q \Gamma_\mu^p v = \Gamma_\mu^q v$, otherwise this does not generally hold.

□

Proof of Proposition 3.6. Set $U := \Gamma_\mu^\nu \log_\mu X$ and $V := \log_\nu Y$, both elements of $T_\nu M$. Since Γ_μ^ν is a linear isometry, $\|U\|_\nu = d(\mu, X)$ and $\|V\|_\nu = d(\nu, Y)$, so the second-moment hypotheses give $\mathbb{E}\|U\|_\nu^2, \mathbb{E}\|V\|_\nu^2 < \infty$. For $u, v \in T_\nu M$, define the rank-one operator $u \otimes v \in \mathcal{H}$ by $(u \otimes v)(w) = \langle v, w\rangle_\nu u$. A direct expansion in any orthonormal basis $\{e_i\}_{i=1}^d$ of $T_\nu M$ gives $\|u \otimes v\|_{\mathcal{H}} = \|u\|_\nu \|v\|_\nu$ together with the pairing identity $\langle u \otimes v, A\rangle_{\mathcal{H}} = \langle u, Av\rangle_\nu$ for every $A \in \mathcal{H}$, which specializes to $\langle u \otimes v, \mathrm{id}_{T_\nu M}\rangle_{\mathcal{H}} = \langle u, v\rangle_\nu$. Continuity of $(u, v) \mapsto u \otimes v$ makes $\omega \mapsto U(\omega) \otimes V(\omega)$ strongly measurable in $\mathcal{H}$, and by Cauchy–Schwarz

$$\mathbb{E}\|U \otimes V\|_{\mathcal{H}} = \mathbb{E}[\|U\|_\nu \|V\|_\nu] \le (\mathbb{E}\|U\|_\nu^2)^{1/2} (\mathbb{E}\|V\|_\nu^2)^{1/2} < \infty,$$

so $U \otimes V$ is Bochner integrable and $\widetilde{C}_{X,Y} = \mathbb{E}[U \otimes V] \in \mathcal{H}$ is well defined. Since $A \mapsto \langle Ae_j, e_i\rangle_\nu$ is a bounded linear functional on $\mathcal{H}$, it commutes with the Bochner expectation, and a direct calculation gives $\langle \widetilde{C}_{X,Y} e_j, e_i\rangle_\nu = \mathbb{E}[\langle V, e_j\rangle_\nu \langle U, e_i\rangle_\nu] = (\Sigma_{X,Y})_{ij}$, so $\Sigma_{X,Y}$ is precisely the matrix of $\widetilde{C}_{X,Y}$ in the basis $\{e_i\}$. The trace $\operatorname{tr}(\cdot) = \langle \cdot, \mathrm{id}_{T_\nu M}\rangle_{\mathcal{H}}$ is itself a bounded linear functional on $\mathcal{H}$, and applying it under the expectation together with the pairing identity above yields

$$\operatorname{tr}(\Sigma_{X,Y}) = \langle \widetilde{C}_{X,Y}, \mathrm{id}_{T_\nu M}\rangle_{\mathcal{H}} = \mathbb{E}\langle U \otimes V, \mathrm{id}_{T_\nu M}\rangle_{\mathcal{H}} = \mathbb{E}\langle U, V\rangle_\nu,$$

which is the claimed identity $\mathrm{Cov}_{X,Y} = \langle \widetilde{C}_{X,Y}, \mathrm{id}_{T_\nu M}\rangle_{\mathcal{H}}$.

□

Proof of Proposition F.1. ***Proof of Fréchet variance special case.***

As we only have one mean, the geodesic reduces to a single point μ. Note that Γ_μ^μ is the identity.

$$\begin{aligned}
\operatorname{tr}\left[\Sigma_{X,X}(\mu)\right] &= \operatorname{tr}\left[\mathbb{E}\left[(\Gamma_\mu^\mu \log_\mu X)(\Gamma_\mu^\mu \log_\mu X)^T\right]\right] \\
&= \operatorname{tr}\left[\mathbb{E}\left[(\log_\mu X)(\log_\mu X)^T\right]\right] \\
&= \mathbb{E}\left[\operatorname{tr}\left[(\log_\mu X)(\log_\mu X)^T\right]\right] \\
&= \mathbb{E}\left[\|\log_\mu X\|_\mu^2\right] \\
&= \mathbb{E}\left[d^2(\mu, X)\right] = \sigma_X^2(\mu).
\end{aligned}$$

The diagonal elements of $\left[(\log_\mu X)(\log_\mu X)^T\right]$ are $\langle \log_\mu X, e_i\rangle \cdot \langle \log_\mu X, e_i\rangle$ where $\{e_i\}$ is a local orthonormal basis in $T_\mu\mathcal{M}$. The trace is thus $\sum_i \langle \log_\mu X, e_i\rangle^2 = \|\log_\mu X\|_\mu^2$.

Proof of symmetry.

$$\begin{aligned}\mathrm{Cov}_{X,Y}(p) &= \mathrm{tr}(\Sigma_{X,Y}(p)) \\ &= \mathrm{tr}\left[\mathbb{E}\left[(\Gamma_\mu^p \log_\mu X)(\Gamma_\nu^p \log_\nu Y)^T\right]\right] \\ &= \mathbb{E}\left[\mathrm{tr}\left[(\Gamma_\mu^p \log_\mu X)(\Gamma_\nu^p \log_\nu Y)^T\right]\right] \\ &= \mathbb{E}\left[\mathrm{tr}\left[(\Gamma_\nu^p \log_\nu Y)(\Gamma_\mu^p \log_\mu X)^T\right]\right] \\ &= \mathrm{tr}\left[\mathbb{E}\left[(\Gamma_\nu^p \log_\nu Y)(\Gamma_\mu^p \log_\mu X)^T\right]\right] \\ &= \mathrm{Cov}_{Y,X}(p).\end{aligned}$$

Proof of degeneracy.

Suppose $\mathrm{Cov}_{X,X}(\mu) = 0$. By the Fréchet variance special case proved above, $\mathrm{Cov}_{X,X}(\mu) = \mathbb{E}\|\log_\mu X\|_\mu^2$. Vanishing of a non-negative expectation forces $\|\log_\mu X\|_\mu = 0$ almost surely, hence $X = \mu$ a.s. Thus X is degenerate.

Proof of uncorrelatedness.

Suppose X and Y are statistically independent random variables on a Riemannian manifold, i.e., $X \perp\!\!\!\perp Y$. Under Assumption 3.1, the Fréchet means μ, ν are unique interior minimizers, so the first-order optimality of the Fréchet variance functions yields

$$\mathbb{E}[\log_\mu X] = 0, \qquad \mathbb{E}[\log_\nu Y] = 0.$$

Independence of X, Y implies independence of $\log_\mu X$ and $\log_\nu Y$, hence

$$\begin{aligned}\Sigma_{X,Y}(p) &= \mathbb{E}\left[(\Gamma_\mu^p \log_\mu X)(\Gamma_\nu^p \log_\nu Y)^\top\right] \\ &= \left(\Gamma_\mu^p \,\mathbb{E}[\log_\mu X]\right)\left(\Gamma_\nu^p \,\mathbb{E}[\log_\nu Y]\right)^\top \\ &= 0,\end{aligned}$$

and consequently $\mathrm{Cov}_{X,Y}(p) = \mathrm{tr}(\Sigma_{X,Y}(p)) = 0$.

Proof that a constant does not vary.

This is clear as $\log_q q = 0$.

Proof of scaling.

First, we must define scaling. Given a constant $a \in \mathbb{R}$ define the scaling action of a on X as $(a, X) \coloneqq \exp_\mu(a \log_\mu(X))$. This scaling requires a footpoint which we choose to be μ here. The mean μ is the canonical choice as it is the only choice which does not affect the center of the density. The scaling $a \log_\mu(X)$ is well-defined as $T_\mu\mathcal{M}$ is a vector space. Note that μ remains the Fréchet mean of (a, X): by linearity, $\mathbb{E}[\log_\mu(a, X)] = a\,\mathbb{E}[\log_\mu X] = 0$ (since μ is the Fréchet mean of X), which is the first-order optimality condition; uniqueness follows from Assumption 3.1. Given Assumption 3.1, the support of X and (a, X) are both required to be within a ball with appropriate radius. Thus, $|a|$ has an upper bound dependent on the support of the original density. Given this setup we have,

$$\begin{aligned}\mathrm{Cov}_{(a,X),Y}(p) &= \mathrm{tr}(\Sigma_{(a,X),Y}(p)) \\ &= \mathrm{tr}\left[\mathbb{E}\left[(\Gamma_\mu^p\, a \log_\mu X)(\Gamma_\nu^p \log_\nu Y)^T\right]\right] \\ &= a\ \mathrm{tr}\left[\mathbb{E}\left[(\Gamma_\mu^p \log_\mu X)(\Gamma_\nu^p \log_\nu Y)^T\right]\right] \\ &= a\ \mathrm{Cov}_{X,Y}(p).\end{aligned}$$

□

Proof of Proposition 3.8. First, we have

$$\mathrm{Cov}_{X,Y}(p) = \mathrm{tr}(\mathbb{E}[(\Gamma_\mu^p \log_\mu X)(\Gamma_\nu^p \log_\nu Y)^T]).$$

Using Jensen's inequality and Cauchy-Schwarz inequality,

$$\begin{aligned}
|\operatorname{Cov}_{X,Y}(p)| &= |\operatorname{tr}(\mathbb{E}[(\Gamma_\mu^p \log_\mu X)(\Gamma_\nu^p \log_\nu Y)^T])| \\
&= |\mathbb{E}[\operatorname{tr}((\Gamma_\mu^p \log_\mu X)(\Gamma_\nu^p \log_\nu Y)^T)]| \\
&\le \mathbb{E}[|\operatorname{tr}((\Gamma_\mu^p \log_\mu X)(\Gamma_\nu^p \log_\nu Y)^T)|] \\
&\le \mathbb{E}(\|\Gamma_\mu^p \log_\mu X\|_p \|\Gamma_\nu^p \log_\nu Y\|_p) \\
&\le (\mathbb{E}\|\Gamma_\mu^p \log_\mu X\|_p^2)^{1/2} (\mathbb{E}\|\Gamma_\nu^p \log_\nu Y\|_p^2)^{1/2} \\
&= (\mathbb{E}\|\log_\mu X\|_\mu^2)^{1/2} (\mathbb{E}\|\log_\nu Y\|_\nu^2)^{1/2} \\
&= \operatorname{Cov}_{X,X}(\mu)^{1/2} \operatorname{Cov}_{Y,Y}(\nu)^{1/2}.
\end{aligned}$$

Therefore,

$$\mathcal{R}_{X,Y}(p) = \frac{\operatorname{Cov}_{X,Y}(p)}{\sqrt{\operatorname{Cov}_{X,X}(\mu)}\sqrt{\operatorname{Cov}_{Y,Y}(\nu)}} \in [-1, 1].$$

□

Theorem G.1. *When $\mathcal{M} = \mathbb{R}^d$, the Riemannian covariance reduces to the standard Euclidean covariance.*

Proof. This is clear noting that in Euclidean space $\log_p q = q - p$ and $\Gamma_p^q v = v$:

$$\begin{aligned}
\operatorname{Cov}_{X,Y}(p) &= \operatorname{tr}(\Sigma_{X,Y}(p)) \\
&= \operatorname{tr}\left[\mathbb{E}\left[(\Gamma_\mu^p \log_\mu X)(\Gamma_\nu^p \log_\nu Y)^T\right]\right] \\
&= \operatorname{tr}\left[\mathbb{E}\left[(X-\mu)(Y-\nu)^T\right]\right].
\end{aligned}$$

□

H Consistency and Asymptotic Normality Proofs

H.1 Regularity conditions

Assumption H.1 (Regularity conditions). We impose the following regularity conditions:

- (A1) There exists a geodesically convex open set $\mathcal{U} \subset \mathcal{M}$ containing μ, ν, p such that: (i) $\mathcal{U}$ is contained in a normal neighborhood of each point in $\mathcal{U}$, (ii) sectional curvature on $\mathcal{U}$ is bounded, and (iii) the injectivity radius on $\mathcal{U}$ is bounded away from 0.
- (A2) $\mathbb{E}\|\log_\mu(X)\|_\mu^2 < \infty$ and $\mathbb{E}\|\log_\nu(Y)\|_\nu^2 < \infty$.
- (A3) $\mathbb{P}(X \in \mathcal{U},\, Y \in \mathcal{U}) \ge 1 - \delta$ for some $\delta \ge 0$ (so the argument applies after truncation).
- (A4) μ and ν are unique Fréchet means and the population Fréchet variance functions are twice continuously differentiable on $\mathcal{U}$. Let

$$H_X := \nabla^2\Big(x \mapsto \mathbb{E}[d^2(X, x)]\Big)\Big|_{x=\mu},$$
$$H_Y := \nabla^2\Big(y \mapsto \mathbb{E}[d^2(Y, y)]\Big)\Big|_{y=\nu}.$$

  Assume H_X and H_Y are positive definite.
- (A5) $\mathbb{E}\|\log_\mu(X)\|_\mu^4 < \infty$ and $\mathbb{E}\|\log_\nu(Y)\|_\nu^4 < \infty$.
- (A6) Define transport maps into $T_p\mathcal{M}$:

$$\phi(x) := \Gamma_\mu^p \log_\mu(x), \qquad \psi(y) := \Gamma_\nu^p \log_\nu(y),$$

  and their plug-in counterparts

$$\hat{\phi}_{\hat{\mu}}(x) := \Gamma_{\hat{\mu}}^p \log_{\hat{\mu}}(x), \qquad \hat{\psi}_{\hat{\nu}}(y) := \Gamma_{\hat{\nu}}^p \log_{\hat{\nu}}(y).$$

  Assume ϕ, ψ are C^2 on $\mathcal{U}$ and the maps $m \mapsto \Gamma_m^p \log_m(z)$ are C^2 jointly in $(m, z) \in \mathcal{U} \times \mathcal{U}$ with uniformly bounded first derivatives on $\mathcal{U} \times \mathcal{U}$. (Under (A1) this holds for the Levi–Civita connection on compact subsets of $\mathcal{U}$.)

H.2 Proofs

Proof of Theorem 3.11. Define $Z_i \coloneqq U_i V_i^\top$ where $U_i \coloneqq \Gamma_\mu^p \log_\mu(X_i)$ and $V_i \coloneqq \Gamma_\nu^p \log_\nu(Y_i)$. Then $\Sigma_{X,Y}(p) = \mathbb{E}[Z_1]$ and the oracle sample average $\bar{Z}_n \coloneqq \frac{1}{n}\sum_{i=1}^n Z_i$ satisfies $\bar{Z}_n \to \mathbb{E}[Z_1]$ almost surely by the strong law of large numbers, provided $\mathbb{E}\|Z_1\|_F < \infty$. Given that $\|Z_1\|_F \le \|U_1\|\,\|V_1\|$ and parallel transport is an isometry, hence by Cauchy–Schwarz and (A2),

$$\begin{aligned}\mathbb{E}\|Z_1\|_F &\le \mathbb{E}\|\log_\mu(X)\|_\mu\,\|\log_\nu(Y)\|_\nu \\ &\le \sqrt{\mathbb{E}\|\log_\mu(X)\|_\mu^2}\,\sqrt{\mathbb{E}\|\log_\nu(Y)\|_\nu^2} < \infty.\end{aligned}$$

Therefore $\bar{Z}_n \xrightarrow{a.s.} \Sigma_{X,Y}(p)$.

It remains to show that $\|\widehat{\Sigma}_{X,Y}(p) - \bar{Z}_n\|_F \to 0$ almost surely. Write

$$\begin{aligned}\widehat{\Sigma}_{X,Y}(p) - \bar{Z}_n &= \frac{1}{n}\sum_{i=1}^n (\hat{U}_i\hat{V}_i^\top - U_iV_i^\top) \\ &= \frac{1}{n}\sum_{i=1}^n \Big[(\hat{U}_i - U_i)\hat{V}_i^\top + U_i(\hat{V}_i - V_i)^\top\Big].\end{aligned}$$

Thus by triangle inequality and submultiplicativity,

$$\begin{aligned}\|\widehat{\Sigma}_{X,Y}(p) - \bar{Z}_n\|_F \le &\underbrace{\frac{1}{n}\sum_{i=1}^n \|\hat{U}_i - U_i\|\,\|\hat{V}_i\|}_{T_{n,1}} \\ &+ \underbrace{\frac{1}{n}\sum_{i=1}^n \|U_i\|\,\|\hat{V}_i - V_i\|}_{T_{n,2}}.\end{aligned} \tag{6}$$

It suffices to control $T_{n,1}$ since the bound for $T_{n,2}$ follows by the same argument with (X,μ,U) replaced by (Y,ν,V). Under (A6), the map $(m,x) \mapsto \Gamma_m^p \log_m(x)$ is continuously differentiable on $\mathcal{U}\times\mathcal{U}$ with bounded derivative. Hence, there exists (random but a.s. finite) L such that for all large n (once $\hat{\mu} \in \mathcal{U}$),

$$\begin{aligned}\|\hat{U}_i - U_i\| &= \|\Gamma_{\hat{\mu}}^p \log_{\hat{\mu}}(X_i) - \Gamma_\mu^p \log_\mu(X_i)\| \\ &\le L\cdot d(\hat{\mu},\mu), \qquad \text{for all } i.\end{aligned}$$

Therefore,

$$T_{n,1} \le L\,d(\hat{\mu},\mu)\cdot\frac{1}{n}\sum_{i=1}^n \|\hat{V}_i\|.$$

By Lemma 3.9, $d(\hat{\mu},\mu) \to 0$ almost surely. Also $\|\hat{V}_i\| = \|\log_{\hat{\nu}}(Y_i)\|_{\hat{\nu}}$ since $\Gamma_{\hat{\nu}}^p$ is an isometry. Because $\hat{\nu} \to \nu$ a.s. and $(m,y) \mapsto \|\log_m(y)\|_m$ is continuous on $\mathcal{U}\times\mathcal{U}$, we have $\|\log_{\hat{\nu}}(Y_i)\|_{\hat{\nu}} \to \|\log_\nu(Y_i)\|_\nu$ pointwise in i along almost every sample path for large n. Moreover, by (A2) and uniform boundedness of derivatives on $\mathcal{U}$, one can dominate $\|\log_{\hat{\nu}}(Y_i)\|_{\hat{\nu}}$ by $C(1 + \|\log_\nu(Y_i)\|_\nu)$ for all large n and some constant C (a standard local Lipschitz bound). Hence, by the strong law of large numbers,

$$\frac{1}{n}\sum_{i=1}^n \|\hat{V}_i\| \xrightarrow{a.s.} \mathbb{E}\|V\| < \infty.$$

Therefore $T_{n,1} \xrightarrow{a.s.} 0$. A symmetric argument yields $T_{n,2} \xrightarrow{a.s.} 0$. Plugging into (6) gives $\|\widehat{\Sigma}_{X,Y}(p) - \bar{Z}_n\|_F \xrightarrow{a.s.} 0$. Combining with $\bar{Z}_n \xrightarrow{a.s.} \Sigma_{X,Y}(p)$ completes the proof. Finally, $\widehat{\mathrm{Cov}}_{X,Y}(p) = \mathrm{tr}(\widehat{\Sigma}_{X,Y}(p)) \xrightarrow{a.s.} \mathrm{tr}(\Sigma_{X,Y}(p))$ by continuity of trace. □

Proof of Lemma 3.12. Recall the transport maps ϕ,ψ and their plug-in versions from Assumption (A6). Define the population functional $T(\mu,\nu) \coloneqq \mathbb{E}[\phi(X)\psi(Y)^\top]$ and the empirical counterpart

$$T_n(m,m') \coloneqq \frac{1}{n}\sum_{i=1}^n \phi_m(X_i)\,\psi_{m'}(Y_i)^\top,$$

for $(m, m') \in \mathcal{U} \times \mathcal{U}$, so that $\Sigma_{X,Y}(p) = T(\mu, \nu)$ and $\widehat{\Sigma}_{X,Y}(p) = T_n(\hat{\mu}, \hat{\nu})$. Write $\widehat{\Sigma} \coloneqq \widehat{\Sigma}_{X,Y}(p)$ and $\Sigma \coloneqq \Sigma_{X,Y}(p)$. Then the decomposition

$$\begin{aligned}\sqrt{n}\big(\widehat{\Sigma} - \Sigma\big) = \sqrt{n}\Big(T_n(\mu, \nu) - T(\mu, \nu)\Big) \\ + \sqrt{n}\Big(T_n(\hat{\mu}, \hat{\nu}) - T_n(\mu, \nu)\Big),\end{aligned} \tag{7}$$

separates the empirical fluctuation term from the plug-in term.

For the empirical term, let $Z_i \coloneqq U_i V_i^\top$ with $\mathbb{E}Z_i = \Sigma_{X,Y}(p)$. Under (A5), $\mathbb{E}\|Z_i\|_F^2 < \infty$ and the multivariate CLT gives

$$\begin{aligned}\sqrt{n}\Big(\mathrm{vec}(T_n(\mu, \nu)) - \mathrm{vec}(T(\mu, \nu))\Big) = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \tilde{Z}_i \\ \xrightarrow{d} \mathcal{N}(0, \Xi_0),\end{aligned} \tag{8}$$

where $\tilde{Z}_i \coloneqq \mathrm{vec}(Z_i) - \mathbb{E}\mathrm{vec}(Z_i)$ and $\Xi_0 \coloneqq \mathrm{Var}(\mathrm{vec}(Z_1))$.

For the plug-in term, let $\delta_\mu \coloneqq \log_\mu(\hat{\mu}) \in T_\mu\mathcal{M}$ and $\delta_\nu \coloneqq \log_\nu(\hat{\nu}) \in T_\nu\mathcal{M}$. Under (A6), $(m, x) \mapsto \phi_m(x)$ and $(m', y) \mapsto \psi_{m'}(y)$ admit first-order Taylor expansions, uniformly over i on $\mathcal{U}$:

$$\phi_{\hat{\mu}}(X_i) = \phi_\mu(X_i) + D_\mu\phi_\mu(X_i)[\delta_\mu] + r_{i,\mu}, \tag{9}$$

$$\psi_{\hat{\nu}}(Y_i) = \psi_\nu(Y_i) + D_\nu\psi_\nu(Y_i)[\delta_\nu] + r_{i,\nu}, \tag{10}$$

with $\max_i \|r_{i,\mu}\| = O_p(\|\delta_\mu\|^2)$ and $\max_i \|r_{i,\nu}\| = O_p(\|\delta_\nu\|^2)$. Multiplying and keeping first-order terms yields

$$\hat{U}_i \hat{V}_i^\top = U_i V_i^\top + \Delta_{\mu,i} V_i^\top + U_i \Delta_{\nu,i}^\top + R_i, \tag{11}$$

where $\Delta_{\mu,i} \coloneqq D_\mu\phi_\mu(X_i)[\delta_\mu]$ and $\Delta_{\nu,i} \coloneqq D_\nu\psi_\nu(Y_i)[\delta_\nu]$. The remainder satisfies $\frac{1}{n}\sum_{i=1}^n \|R_i\|_F = O_p(\|\delta_\mu\|^2 + \|\delta_\nu\|^2 + \|\delta_\mu\|\,\|\delta_\nu\|)$, hence by Lemma 3.10,

$$\sqrt{n} \cdot \Big\|\frac{1}{n}\sum_{i=1}^{n} R_i\Big\|_F = o_p(1). \tag{12}$$

Using $\mathrm{vec}(ab^\top) = b \otimes a$ gives the linearization

$$\sqrt{n}\,\mathrm{vec}\big(\Delta T_n\big) = B_{\mu,n}\,\sqrt{n}\,\delta_\mu + B_{\nu,n}\,\sqrt{n}\,\delta_\nu + o_p(1), \tag{13}$$

where $\Delta T_n \coloneqq T_n(\hat{\mu}, \hat{\nu}) - T_n(\mu, \nu)$ and

$$B_{\mu,n} \coloneqq \frac{1}{n}\sum_{i=1}^{n} V_i \otimes \Big(D_\mu\phi_\mu(X_i)\Big),$$

$$B_{\nu,n} \coloneqq \frac{1}{n}\sum_{i=1}^{n} \Big(D_\nu\psi_\nu(Y_i)\Big) \otimes U_i.$$

By the law of large numbers and (A5)–(A6), $B_{\mu,n} \to_p A_\mu$ and $B_{\nu,n} \to_p A_\nu$, where

$$A_\mu \coloneqq \mathbb{E}\big[V \otimes D_\mu\phi_\mu(X)\big], \tag{14}$$

$$A_\nu \coloneqq \mathbb{E}\big[D_\nu\psi_\nu(Y) \otimes U\big]. \tag{15}$$

Thus

$$\sqrt{n}\,\mathrm{vec}\big(\Delta T_n\big) = A_\mu\,\sqrt{n}\,\delta_\mu + A_\nu\,\sqrt{n}\,\delta_\nu + o_p(1). \tag{16}$$

Substituting the $\sqrt{n}$-expansions of δ_μ, δ_ν from Lemma 3.10 yields

$$\begin{aligned}\sqrt{n}\,\mathrm{vec}\big(\Delta T_n\big) = A_\mu H_X^{-1}\left(\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\xi_i\right) \\ + A_\nu H_Y^{-1}\left(\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\zeta_i\right) + o_p(1).\end{aligned} \tag{17}$$

Finally, combining (7), the vectorized empirical term, and (17) gives the asymptotic linear representation (3) with influence term (4). □

Proof of Theorem 3.13. By Lemma 3.12, we have

$$\sqrt{n}\,\Delta_n = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \mathrm{IF}_i + o_p(1).$$

The summands IF_i are i.i.d. mean-zero. Under (A5)–(A6), they have finite second moments, hence by the multivariate CLT $\frac{1}{\sqrt{n}} \sum_{i=1}^{n} \mathrm{IF}_i \Rightarrow \mathcal{N}(0, \mathrm{Var}(\mathrm{IF}_1))$. Slutsky's theorem yields the stated limit with $\Xi = \mathrm{Var}(\mathrm{IF}_1)$. □

Proof of Corollary 3.14. Note that $\mathrm{tr}(M) = \langle I_d, M \rangle_F = \langle \mathrm{vec}(I_d), \mathrm{vec}(M) \rangle$. Apply the continuous mapping theorem to Theorem 3.13 with the linear map $v \mapsto \langle \mathrm{vec}(I_d), v \rangle$. □

I Simulation Results

The sphere experiments were conducted on a Macbook Air with an Apple M1 chip (8 cores: 4 performance and 4 efficiency), with a total wall-clock time of approximately 1 hours. The SPD experiments were run on a MacBook Pro with an Apple M4 Pro chip (14 cores: 10 performance and 4 efficiency) and 24 GB of unified memory, running macOS 26.3, with a total wall-clock time of approximately 3 hours. While the runtime for the real case study is negligible, the bootstrap method for Schizophrenia challenge dataset takes about 2 hours.

We utilize code from ManOpt [Townsend et al.] (`http://jmlr.org/papers/v17/16-177.html`) and Geomstats [Miolane et al., 2020] (`https://www.jmlr.org/papers/v21/19-027.html`).

I.1 Sphere

I.1.1 Data Generation and Experimental Setup

We start by generating a random data set $\{X_i\}_{i=1}^n$ that is uniformly distributed on a truncated cap of $\mathbb{S}^2$, which provides a simple way to ensure the data are bounded. Fixing the center at the north pole, points with coordinate $(x, y, z) \in \mathbb{R}^3$ are generated as

$$\begin{aligned} x &= \sqrt{1 - z^2}\cos(2\pi v), \\ y &= \sqrt{1 - z^2}\sin(2\pi v), \\ z &= 1 - w(1 - \cos\tau), \end{aligned}$$

where $v, w \sim U(0, 1)$ and τ controls the bound of the data. To guarantee the uniqueness of the Fréchet mean, we require $0 < \tau < \pi/2$. The dependent dataset $\{Y_i\}_{i=1}^n$ is controlled by defining it in terms of a rotation of the first set. The rotation matrix, given by

$$\begin{aligned} O(\eta) &= I + \sin(\eta)\, K + (1 - \cos(\eta))\, K^2, \\ K &= \begin{bmatrix} 0 & -u_z & u_y \\ u_z & 0 & -u_x \\ -u_y & u_x & 0 \end{bmatrix}, \end{aligned}$$

is a realization of Rodrigues' rotation formula [Cheng and Gupta, 1989]. Here η denotes the angle of counterclockwise rotation about an axis (u_x, u_y, u_z). Noise is then added to represent the random perturbation, after which the samples are parallel transported to the tangent space at another point q.

Therefore, the second sample set is generated as

$$Y_i = \exp_q \left[\Gamma_{\tilde{\nu}}^{q} \left(\frac{O(\eta) \log_{\hat{\mu}} X_i + \epsilon_i}{\| O(\eta) \log_{\hat{\mu}} X_i + \epsilon_i \|} \right) \right],$$

where $\epsilon_i \sim N_3(0, \sigma_\epsilon^2 I_3)$, $\hat{\mu}$ is the Fréchet mean of the set $\{X_i\}_{i=1}^n$ and $\tilde{\nu}$ is the Fréchet mean of the set $\{Y_i\}_{i=1}^n$ before the transportation. We can expect $q = \hat{\nu}$ to be the Fréchet mean of transported data, as there's no additional noise introduced.

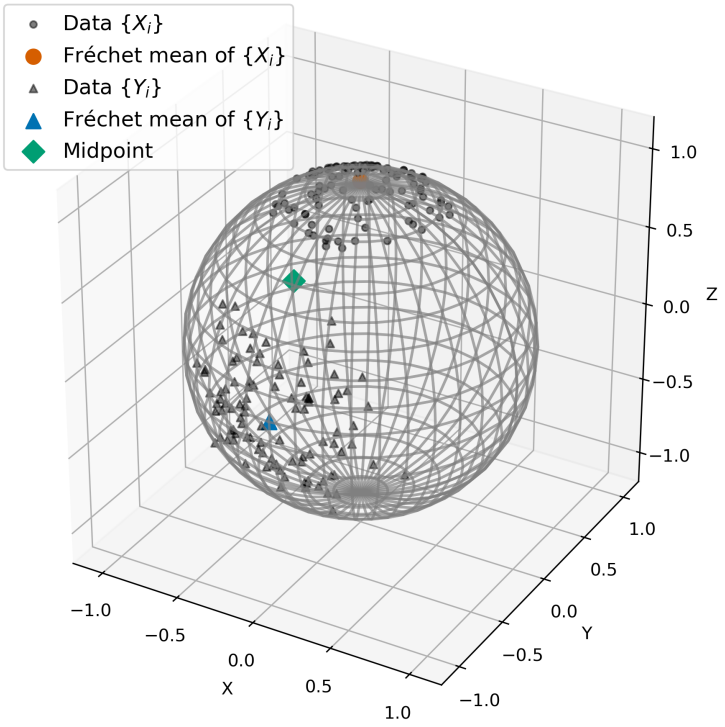

Figure 4: An illustration of a dataset $(n = 100)$ on $\mathbb{S}^2$ and its dependent dataset, together with the respective Fréchet mean and the midpoint of two Fréchet mean.

I.1.2 Simulation on Transport effects

In Figure 5 we evaluate the correlation estimates along a geodesic. The correlation is computed by transporting the data $\{Y_i\}$ to different points along a geodesic connecting $\tilde{\nu}$ and $(0, -1, 0)$. The shaded area represents the confidence interval of 2 standard errors. It can be seen that our estimates closely match the $A\&P$ correlation when the two datasets are close, but diverge as they are moved far apart. This indicates that the distance between datasets can affect the $A\&P$ estimate when it is evaluated on the tangent space of the midpoint.

In Figure 6, we compare our proposed Riemannian correlation with $A\&P$ correlation and CCA estimator. Solid lines correspond to the original set $\{Y_i\}_{i=1}^n$ without transport, while dotted lines show the results after transporting $\{Y_i\}_{i=1}^n$ to $(0, -1, 0)$. The CCA estimator gives the same estimation before and after transportation, but it consistently overestimates the correlation, as it removes the effect of the rotation matrix.

I.1.3 Sensitivity to Rotation and Independence

In Figure 7, we explore the effect of different rotation matrices on the samples, as shown in the left-hand plot. The samples $\{Y_i\}_{i=1}^n$ are generated by rotating $\{X_i\}_{i=1}^n$ by various angles and transported to the midpoint of the geodesic connecting $\tilde{\nu}$ and $(0, -1, 0)$. When no noise is added, we expect $\mathcal{R}_{X,Y}(\hat{\nu}) = 1$ for $\eta = 0$ and $\mathcal{R}_{X,Y}(\hat{\nu}) = 0$ for $\eta = \pi/2$, as the two datasets form orthogonal vectors in a tangent space. As before, CCA consistently ignores the rotation and overestimates the correlation. Our method gives estimates similar to those of the $A\&P$ approach, and it can be checked that the estimates are close to 0 at $\pi/2$.

Finally, we test the estimators on two independent datasets. The datasets are generated using a pseudo-random number generator (PRNG). Though they are not mathematically independent in a theoretical sense, the independence assumption is sufficient for simulation purposes. One dataset is transported along the geodesic, with results shown on the right of Figure 7. As can be seen, our method always has estimates around 0, indicating the reliability of the method. The $A\&P$ method

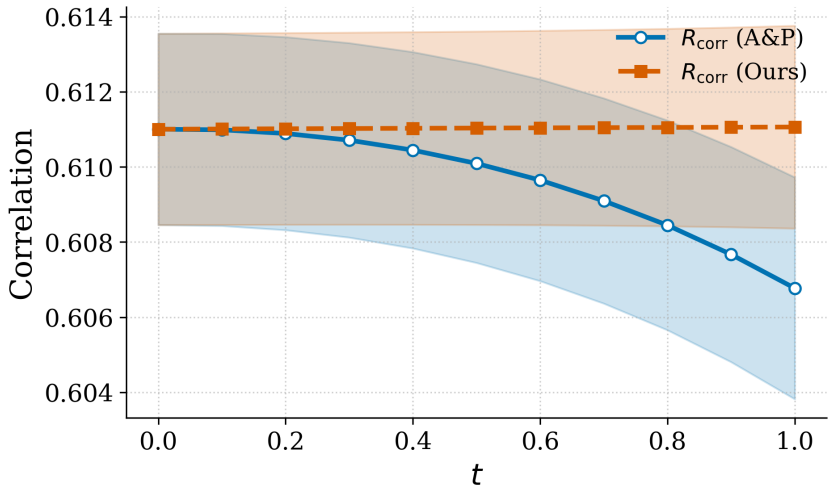

Figure 5: Comparison of Riemannian correlation (Ours), $A\&P$ correlation, and the CCA estimator. The set $\{Y_i\}$ is transported along the geodesic α with $\alpha(0) = \tilde{\nu}$ and $\alpha(1) = (0, -1, 0)$. The x-axis denotes the position of Fréchet mean of $\{Y_i\}$ along α. Parameters: $\tau = \pi/6$, $\eta = \pi/4$, $\sigma_\epsilon = 0.15$, MC $= 200$.

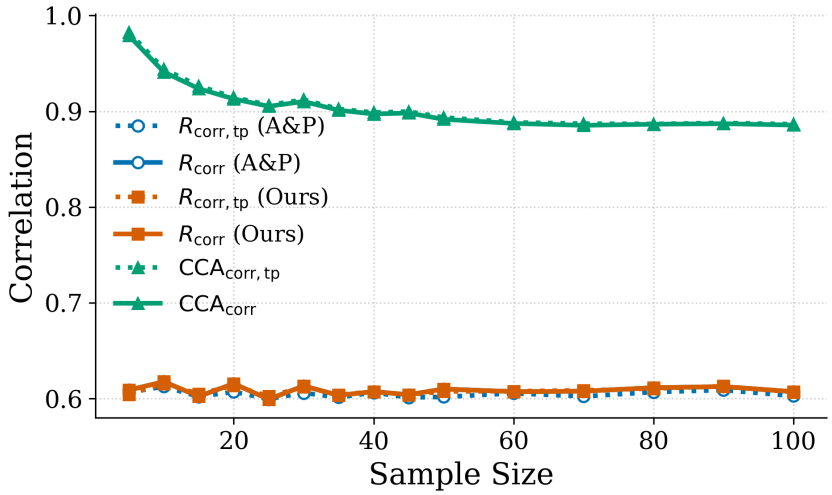

Figure 6: Comparison of Riemannian correlation, $A\&P$ correlation and CCA estimator. Solid lines correspond to samples sharing the same Fréchet mean, while dotted lines represent results from data $\{Y_i\}$ transported to $(0, -1, 0)$. Parameters: $\tau = \pi/6$, $\eta = \pi/4$, with noise level $\sigma_\epsilon = 0.15$.

also has a similar pattern. Since CCA is defined to maximize an optimization problem, it produces slightly higher correlation estimates.

I.1.4 Ground-truth Recovery Experiment

We here conduct a recovery simulation on $\mathbb{S}^2$. In the tangent space at the north pole $(0, 0, 1)$, we generate two sets of normally distributed samples with a known covariance matrix. The samples take the form $X = (x_1, x_2, 0)$ and $Y = (y_1, y_2, 0)$, where $(x_1, x_2, y_1, y_2) \sim \mathcal{N}_4(0, \Sigma)$ with

$$\Sigma = \begin{pmatrix} 0.1I_2 & 0.04J_2 \\ 0.04J_2 & 0.08I_2 \end{pmatrix}, \quad J_2 = \begin{pmatrix} 1 & 1 \\ 1 & 1 \end{pmatrix}.$$

The cross-covariance matrix of the two sets is the 3×3 matrix whose top-left block is $0.04J_2$ and all other entries are 0. The correlation, by construction, is $0.08/\sqrt{0.2 \cdot 0.16} \approx 0.447$. We generate

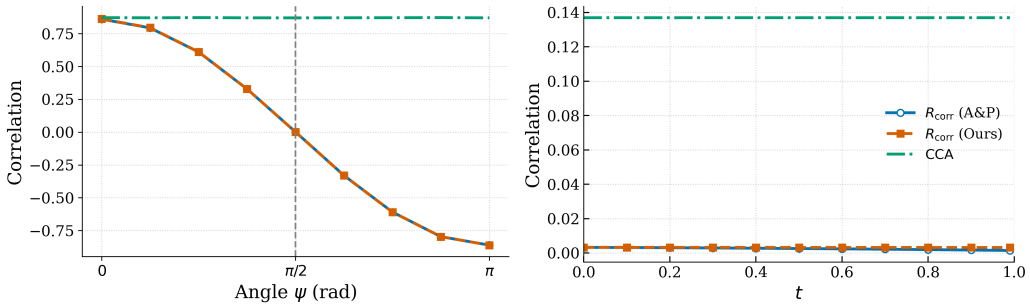

Figure 7: Comparison of Riemannian correlation (Ours), $A\&P$ correlation, and the CCA estimator. The parameter is set to $\tau = \pi/6$. **Left:** The set $\{Y_i\}$ is transported to $\alpha(1/2)$, with $\alpha(0) = \tilde{\nu}$ and $\alpha(1) = (0, -1, 0)$. The x-axis shows the rotation angle, with the vertical dashed line indicating $\pi/2$. The noise level is $\sigma_\epsilon = 0.15$. **Right:** Two independent sets with $\{Y_i\}$ transported along the geodesic connecting its Fréchet mean and $(0, -1, 0)$. The x-axis denotes the position of the Fréchet mean of $\{Y_i\}$ along α.

1000 pairs of data and repeat the experiment 100 times. Table 2 gives a numerical comparison of the methods addressed in the paper. The geodesic γ is defined such that $\gamma(0)$ is the midpoint of geodesic connecting Fréchet mean of $\{\exp_{(0,0,1)} X_i\}$ and $\{\exp_{(0,0,1)} Y_i\}$, and $\gamma(1) = (0, 1, 0)$.

Table 2: Covariance and correlation estimates for recovery experiment

	$\gamma(0)$	$\gamma(0.5)$	$\gamma(1)$
$R_{cov}(Ours)$	0.079	0.079	0.079
$R_{cov}(A\&P)$	0.079	0.087	0.081
$R_{corr}(Ours)$	0.446	0.446	0.446
$R_{corr}(A\&P)$	0.446	0.438	0.445
CCA	0.894	0.894	0.894

Specifically, the covariance matrix given by our method (at Fréchet mean of X set) is

$$\hat{\Sigma} = \begin{pmatrix} 0.0395 & 0.0396 & -5.96 \times 10^{-6} \\ 0.0398 & 0.0399 & -7.07 \times 10^{-6} \\ 4.715 \times 10^{-8} & -2.692 \times 10^{-6} & 9.971 \times 10^{-6} \end{pmatrix},$$

which is very close to the ground truth covariance matrix.

I.2 SPD Manifold

I.2.1 Data Generation and Experimental Setup

To rigorously assess the proposed estimator, we employ a generative model that decouples intrinsic geometric correlation from ambient curvature effects. The manifold of $d \times d$ Symmetric Positive Definite (SPD) matrices, $\mathcal{P}_d$, is endowed with the Affine Invariant Riemannian Metric (AIRM). We fix the dimensionality at $d = 3$ and the sample size at $n = 100$ for most experiments unless otherwise specified.

The primary dataset $\{X_i\}_{i=1}^n$ is generated by perturbing the identity matrix I_d. Tangent vectors $v_{X,i} \in T_{I_d}\mathcal{P}_d$ are sampled from a symmetric normal distribution with spread parameter $\tau = 0.5$ and subsequently mapped to the manifold via the Riemannian exponential map, denoted as $X_i = \text{Exp}_{I_d}(v_{X,i})$.

The dependent dataset $\{Y_i\}_{i=1}^n$ is constructed via a two-stage process to ensure that the correlation structure is intrinsic and transport-invariant. First, we generate tangent vectors $v_{Y,i} \in T_{I_d}\mathcal{P}_d$ that are locally correlated with $v_{X,i}$ according to the relation $v_{Y,i} = \rho \cdot v_{X,i} + \sqrt{1-\rho^2} \cdot \epsilon_i$. Here, $\rho \in [0, 1]$ controls the correlation strength, and ϵ_i is a random symmetric matrix representing noise, normalized such that $\|\epsilon_i\|_{I_d} = \|v_{X,i}\|_{I_d}$. This normalization ensures that the dispersion of Y matches

that of X. Subsequently, the entire vector cloud $\{v_{Y,i}\}$ is parallel transported from I_d to a target mean ν along the geodesic connecting them. Let $\Gamma_{I_d}^{\nu}$ denote the parallel transport operator; the final samples are given by $Y_i = \mathrm{Exp}_{\nu}\left(\Gamma_{I_d}^{\nu}(v_{Y,i})\right)$. By generating correlation locally and then transporting the distribution, we ensure that any observed decorrelation in the estimators is attributable to the estimation method itself rather than artifacts of the data generation process in curved space.

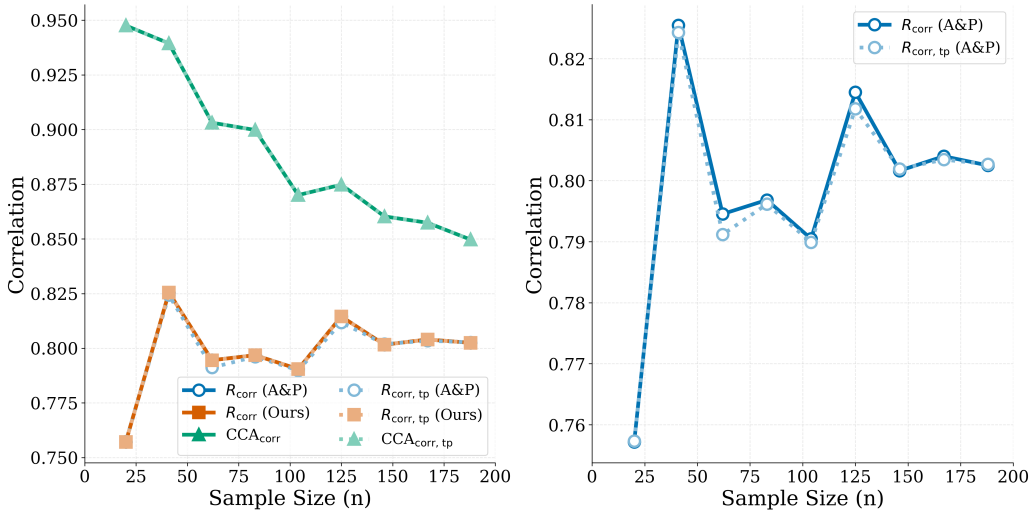

Figure 8: Convergence of Correlation Estimates on SPD Manifold. **Left:** Comparison of A&P, Ours, and CCA methods with increasing sample size. Solid lines: original dataset (non-transport); dashed lines: parallel-transported dataset. Our method produces nearly identical estimates for non-transport and transport scenarios, consistent with isometric invariance. Log-log regression suggests $\sqrt{n}$-convergence (slope -0.49 ± 0.03). **Right:** Detailed A&P vs Ours comparison. Shaded regions show 95% confidence intervals. Parameters: $d = 3, \tau = 0.5, \sigma_\epsilon = 0.15$, MC varies by n.

I.2.2 Asymptotic Consistency and Convergence

Asymptotic consistency and isometric invariance are assessed in Figure 8. Log-log regression analysis suggests $\sqrt{n}$-convergence for both methods: the convergence rate estimates are -0.49 ± 0.03 (Ours) and -0.51 ± 0.04 (A&P), which are consistent with the theoretical rate of -0.5 (95% confidence intervals include -0.5). The proposed method produces nearly identical estimates for the original and parallel-transported datasets across all sample sizes, which is consistent with the estimator capturing the intrinsic geometry invariant under isometry. This is a prerequisite for robust statistical inference on manifolds independent of the ambient coordinate system. The right panel provides a detailed comparison between A&P and Ours, highlighting the systematic differences in convergence behavior.

I.2.3 Geodesic Simulation and Transport Effects

Figure 11 examines the estimator's behavior under geometric transformations through two complementary experiments that test distinct theoretical properties. The left panel tests isometric invariance by physically transporting the dataset Y along a geodesic away from X and evaluating correlation at a fixed reference point. As expected from the definition of parallel transport (an isometry preserving inner products), the intrinsic correlation should remain invariant, and our estimator exhibits low variation (standard error $SE \approx 0.008$). The right panel tests footpoint invariance (Theorem 3.5) by fixing the data and evaluating correlation at different analytical points p along the geodesic connecting the sample means. Our estimator produces nearly constant estimates ($SE \approx 0.008$) across all evaluation points, while the A&P estimator exhibits substantial variation ($SE \approx 0.15$), depending on where it is computed. This dependence on an arbitrary reference tangent space is undesirable for a summary statistic, whereas our method renders results independent of this choice.

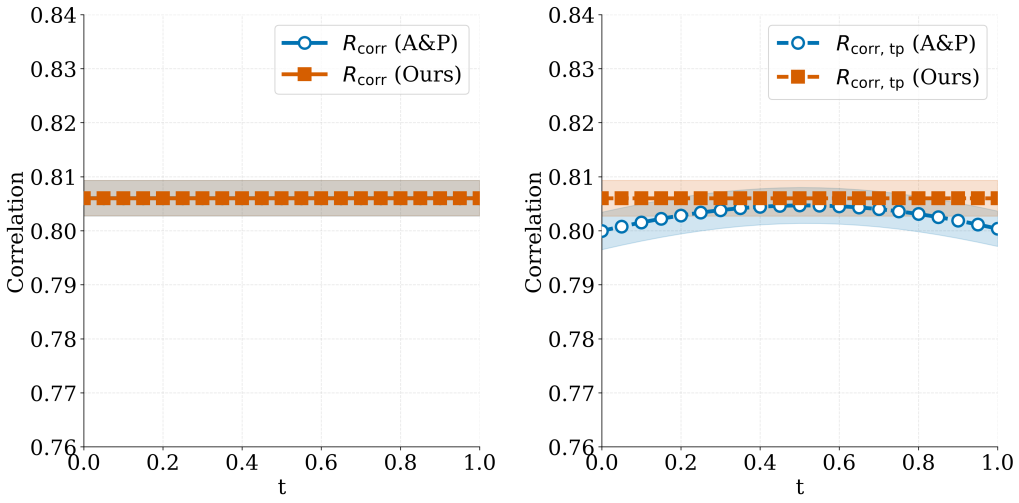

Figure 9: Geodesic Evaluation on SPD Manifold. **Left:** Original dataset (non-transport). **Right:** Parallel-transported dataset (transport). Our method produces nearly constant estimates ($SE \approx 0.008$) across all evaluation points, consistent with footpoint invariance (Theorem 3.5). The A&P method exhibits substantial variation ($SE \approx 0.15$) with an inverted U-shaped pattern. Shaded regions show 95% confidence intervals (mean $\pm$ 2 SE). Parameters: $n = 100, d = 3, \tau = 0.5, \sigma_\epsilon = 0.15$, MC $= 200$.

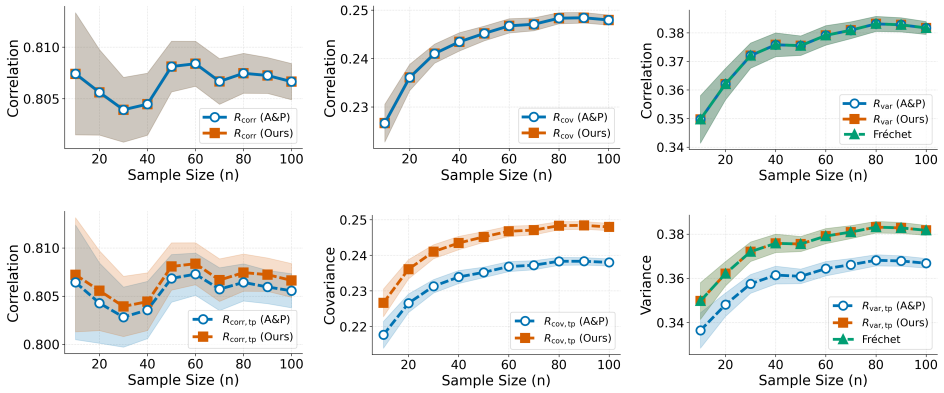

Figure 10: Comprehensive Comparison on SPD Manifold. **Top Row:** Original dataset (non-transport). **Bottom Row:** Parallel-transported dataset (transport). **Columns:** Correlation, covariance, and variance estimates versus sample size. Our method closely matches Fréchet variance, while A&P yields inflated variance (ratio $= 1.23 \pm 0.05$ at $n = 100$), associated with deflated correlation. Our method produces nearly identical estimates for non-transport/transport scenarios, consistent with isometric invariance. Shaded regions indicate 95% confidence intervals (mean $\pm$ 2 SE). Parameters: $n \in \{10, 20, \ldots, 100\}, d = 3, \tau = 0.5, \sigma_\epsilon = 0.15$, MC $= 200$ per n.

I.2.4 Sensitivity to Rotation and Independence

We analyze the sensitivity of the estimator to the underlying dependence structure using tangent space rotations. Figure 12 demonstrates that the correlation estimate decreases monotonically as the rotation angle η approaches $\pi/2$, reflecting the geometric decorrelation. The relationship follows the expected pattern $\mathcal{R}_{X,Y}(\eta) \approx \cos(\eta)$ in the noise-free limit, with our method closely tracking this theoretical curve (mean absolute error ≈ 0.04 across all angles). Furthermore, under the null hypothesis of independence, the estimator yields values close to zero (mean correlation $= 0.002 \pm 0.008$), which is consistent with the expected behavior for independent datasets. The stability across transport locations (right panel) further supports the estimator's invariance properties.

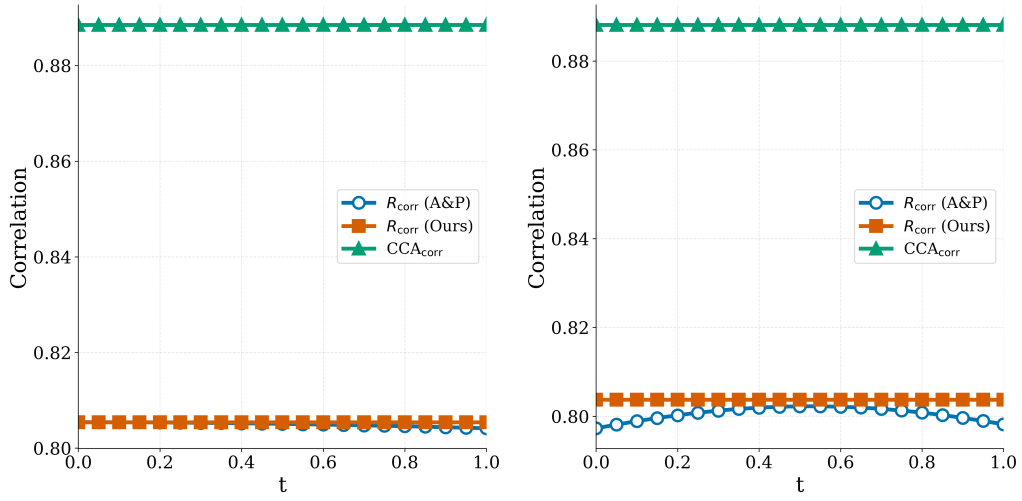

Figure 11: Geodesic Simulation and Transport Effects on SPD Manifold. **Left:** Isometric invariance test: correlation estimates remain stable ($SE \approx 0.008$) when dataset Y is physically transported along a geodesic. **Right:** Footpoint invariance test (Theorem 3.5): correlation estimates remain nearly constant ($SE \approx 0.008$) when evaluated at different points p along the geodesic. The A&P estimator varies substantially ($SE \approx 0.15$), depending on the evaluation point. Parameters: $n = 100, d = 3, \tau = 0.5, \sigma_\epsilon = 0.15$, MC $= 50$.

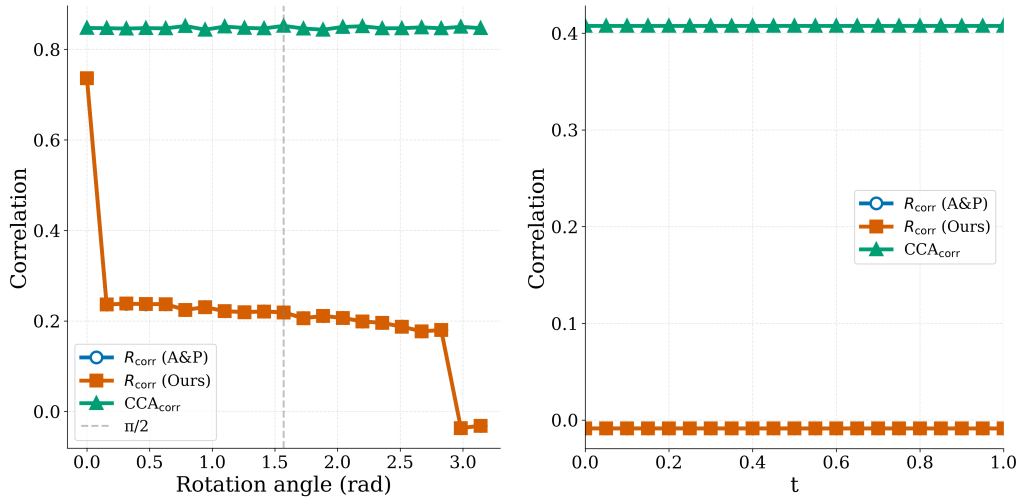

Figure 12: Angle Effect and Independence on SPD Manifold. **Left:** Correlation decreases monotonically with rotation angle η, following $\cos(\eta)$ pattern (mean absolute error ≈ 0.04). **Right:** Independent datasets yield correlation $= 0.002 \pm 0.008$, consistent with expected behavior for independent data. Parameters: $n = 100, d = 3, \tau = 0.5, \sigma_\epsilon = 0.15$, MC $= 60$.

I.2.5 Effect of Mean Separation

A critical challenge in non-positively curved spaces like $\mathcal{P}_d$ is that estimators relying on a single reference frame become increasingly biased as the geodesic distance $d(\mu_X, \mu_Y)$ between population means increases. This occurs because the logarithmic map $\log_p(\cdot)$ at a fixed footpoint distorts the metric structure differently as the data moves farther from the reference point. Figure 13 systematically examines how both estimators behave as the mean separation increases, while maintaining a fixed intrinsic correlation structure ($\rho = 0.8$). The Y dataset is generated with correlation to X at the source location, then transported as a unit to a target mean ν at varying geodesic distances from μ_X. This design ensures that any observed decorrelation is attributable to the estimation method rather than the data generation process.

The results reveal a fundamental difference: while the A&P estimator (evaluated at the geodesic midpoint) exhibits systematic bias that increases monotonically with separation distance, our proposed method remains stable and accurately recovers the intrinsic correlation (mean correlation $= 0.798 \pm 0.012$ across all separation distances, with no significant trend). At maximum separation ($d = 2.0$), the A&P method underestimates correlation by 0.23 ± 0.05 compared to the true value. The difference plot highlights the divergence, showing that the A&P method increasingly underestimates correlation as means separate, which is associated with variance inflation (variance ratio increases from 1.05 at $d = 0$ to 1.35 at $d = 2.0$) due to metric distortion at the midpoint evaluation point.

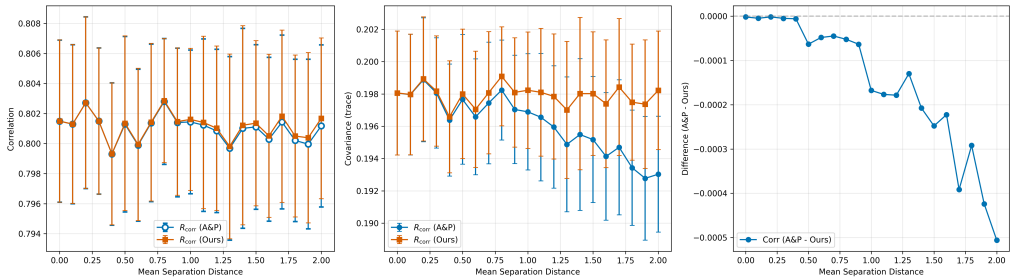

Figure 13: Effect of Mean Separation on Estimator Performance. **Left:** Correlation estimates versus geodesic distance $d(\mu_X, \mu_Y)$. A&P exhibits monotonic bias, while Ours remains stable (0.798 ± 0.012, with no significant trend). **Middle:** Covariance estimates. **Right:** Difference (A&P - Ours) highlighting systematic divergence. At $d = 2.0$, A&P underestimates by 0.23 ± 0.05. Intrinsic correlation fixed at $\rho = 0.8$. Parameters: $n = 100, d = 5, \tau_X = \tau_Y = 0.5$, MC $= 50$ per distance.

I.3 Real Case Studies

I.3.1 Sunnybrook Cardiac MRI dataset

Here we include an additional figure for our experimental results on Sunnybrook Cardiac MRI dataset (`https://www.cardiacatlas.org/sunnybrook-cardiac-data/`). Figure 14 displays examples of the inner and outer ventricular walls for the four groups. The landmarks of each shape are connected with a solid line to display the shape. In Table 1 we present the results for both shape and preshape space. In either case our estimates and those of $A\&P$ are similar to each other and different from those of the CCA (and its variants). We present results for both shape and preshape shape as rotational differences may be of interest to practitioners. That is, if all the data has been aligned in a medically meaningful way, then an analysis over shape space would remove that alignment dependence. As opposed to our simulations which show significant differences between our proposed correlation and $A\&P$ correlation, there is a small difference here. The reason behind this is that in all of these groups the respective group Fréchet means are very close. That is, the $A\&P$ correlation is most effected when the two group means are dissimilar to each other.

Table 3 reproduces Table 1 in shape space with two additional columns including results from ridge canonical correlation analysis (RidgeCCA) and partial least squares (PLS). RidgeCCA extends standard CCA by regularizing the covariance matrices, improving numerical stability when the number of features exceeds the sample size. PLS instead maximizes covariance subject to unit-length weight constraints, i.e., $\|W_x\|_2 = \|W_y\|_2 = 1$ in Appendix C. Both methods circumvent difficulties arising from singular covariance matrices. Although the RidgeCCA regularization parameter is

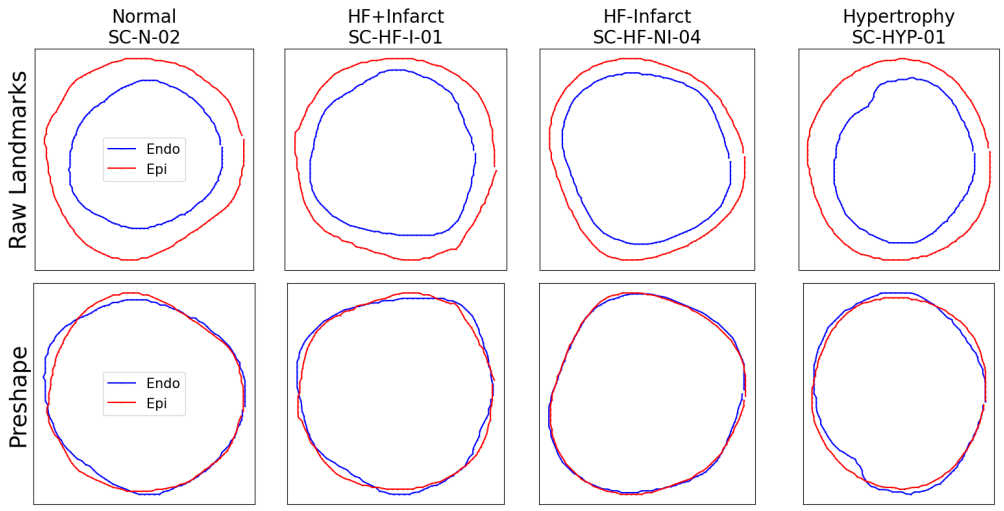

Figure 14: Illustrative examples of inner and outer ventricular walls from the four groups. **Top:** raw landmarks extracted from cardiac MRI. **Bottom:** preprocessed shapes in preshape space.

Table 3: Correlation estimates for the four groups.

	R_{corr}(Ours)	$R_{corr}(A\&P)$	PLS	RidgeCCA	CCA
Normal ($n = 9$)	0.637	0.637	0.910	0.9992	1
Hypertrophy ($n = 12$)	0.393	0.394	0.798	0.9950	1
HF-infarct ($n = 12$)	0.846	0.846	0.975	0.9969	1
HF+infarct ($n = 12$)	0.659	0.659	0.934	0.9992	1

typically selected by cross-validation, we set the penalty $\lambda = 1$ given the limited sample size and the lack of ground-truth information. As shown in the table, both methods have high estimates, while PLS exhibits a pattern similar to those of our proposed measure and the A&P correlation.

I.3.2 Schizophrenia Challenge dataset

We use the MLSP 2014 Schizophrenia Challenge dataset [Castro et al., 2014], available via `geomstats` [Miolane et al., 2020]: 86 subjects (46 healthy, 40 schizophrenic), each with a 28×28 SPD connectivity matrix. For a simple low-dimensional illustration, we use one pair of non-overlapping 4×4 principal sub-matrices as $X, Y \in \mathcal{P}_4$ and compute the Riemannian correlation within each group ($n = 46, 40$; 3–5$\times$ larger than Sunnybrook).

Result 1: Footpoint invariance (Fig. 15). Across 15 footpoints along the geodesic ($t \in [0, 2]$), our estimator is exactly constant, whereas A&P varies systematically. The absolute variation is small on this dataset (up to 0.003) because the group means are close, but the bias grows with mean separation (up to 0.23 in Figure 13). The key point is methodological: a correlation should not depend on an arbitrary analyst choice of footpoint.

Result 2: Group differentiation (Fig. 16). The healthy group yields $\hat{\mathcal{R}} = -0.096$ and the schizophrenic group $\hat{\mathcal{R}} = +0.106$, giving a sign reversal with $\Delta\hat{\mathcal{R}} = 0.20$. A two-sample permutation test confirms this group difference ($p = 0.032$, 1000 permutations). Bootstrap 95% CIs ($B = 1000$) are shown in Fig. 16.

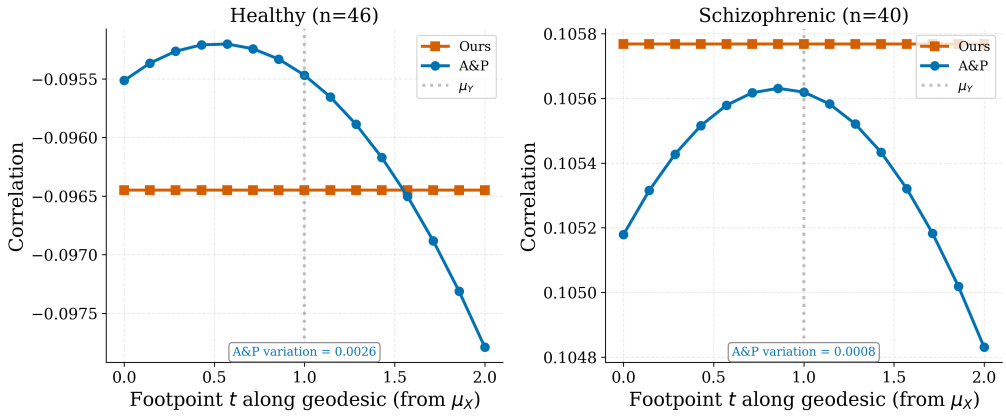

Figure 15: Footpoint invariance on connectome data. Our method (orange) is constant; A&P (blue) varies with the footpoint.

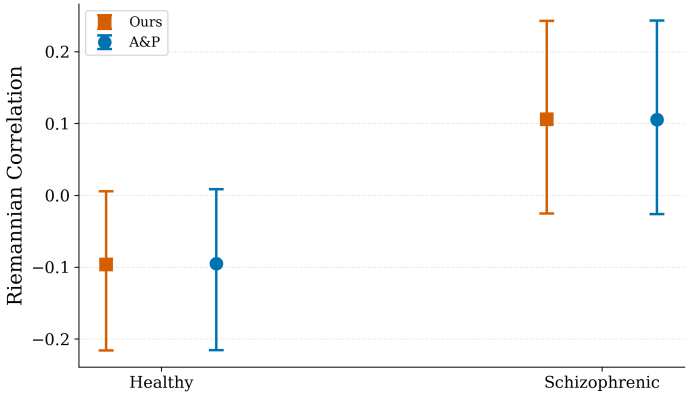

Figure 16: Bootstrap 95% CI for Riemannian correlation. Sign reversal between groups is significant ($p = 0.032$, two-sample permutation test).

NeurIPS Paper Checklist

1. **Claims**

   Question: Do the main claims made in the abstract and introduction accurately reflect the paper's contributions and scope?

   Answer: [Yes]

   Justification: The claimed contribution is introduced in Section 3. Our claims of invariance are supported theoretically in Section 3 and experimentally in Section 4.

   Guidelines:

   - The answer [N/A] means that the abstract and introduction do not include the claims made in the paper.
   - The abstract and/or introduction should clearly state the claims made, including the contributions made in the paper and important assumptions and limitations. A [No] or [N/A] answer to this question will not be perceived well by the reviewers.
   - The claims made should match theoretical and experimental results, and reflect how much the results can be expected to generalize to other settings.
   - It is fine to include aspirational goals as motivation as long as it is clear that these goals are not attained by the paper.

2. **Limitations**

   Question: Does the paper discuss the limitations of the work performed by the authors?

   Answer: [Yes]

   Justification: We include a Limitations section in Appendix A.

   Guidelines:

   - The answer [N/A] means that the paper has no limitation while the answer [No] means that the paper has limitations, but those are not discussed in the paper.
   - The authors are encouraged to create a separate "Limitations" section in their paper.
   - The paper should point out any strong assumptions and how robust the results are to violations of these assumptions (e.g., independence assumptions, noiseless settings, model well-specification, asymptotic approximations only holding locally). The authors should reflect on how these assumptions might be violated in practice and what the implications would be.
   - The authors should reflect on the scope of the claims made, e.g., if the approach was only tested on a few datasets or with a few runs. In general, empirical results often depend on implicit assumptions, which should be articulated.
   - The authors should reflect on the factors that influence the performance of the approach. For example, a facial recognition algorithm may perform poorly when image resolution is low or images are taken in low lighting. Or a speech-to-text system might not be used reliably to provide closed captions for online lectures because it fails to handle technical jargon.
   - The authors should discuss the computational efficiency of the proposed algorithms and how they scale with dataset size.
   - If applicable, the authors should discuss possible limitations of their approach to address problems of privacy and fairness.
   - While the authors might fear that complete honesty about limitations might be used by reviewers as grounds for rejection, a worse outcome might be that reviewers discover limitations that aren't acknowledged in the paper. The authors should use their best judgment and recognize that individual actions in favor of transparency play an important role in developing norms that preserve the integrity of the community. Reviewers will be specifically instructed to not penalize honesty concerning limitations.

3. **Theory assumptions and proofs**

   Question: For each theoretical result, does the paper provide the full set of assumptions and a complete (and correct) proof?

   Answer: [Yes]

Justification: We state our assumptions in 3.1, 3.2, and H.1. All proofs have been included. Due to space limitations proofs are presented in Sections G and H.

Guidelines:

- The answer [N/A] means that the paper does not include theoretical results.
- All the theorems, formulas, and proofs in the paper should be numbered and cross-referenced.
- All assumptions should be clearly stated or referenced in the statement of any theorems.
- The proofs can either appear in the main paper or the supplemental material, but if they appear in the supplemental material, the authors are encouraged to provide a short proof sketch to provide intuition.
- Inversely, any informal proof provided in the core of the paper should be complemented by formal proofs provided in appendix or supplemental material.
- Theorems and Lemmas that the proof relies upon should be properly referenced.

4. **Experimental result reproducibility**

   Question: Does the paper fully disclose all the information needed to reproduce the main experimental results of the paper to the extent that it affects the main claims and/or conclusions of the paper (regardless of whether the code and data are provided or not)?

   Answer: [Yes]

   Justification: In Section 3 we include the technicaly defintions, in Section 4 we outline the experiments, and in Appendix I we give details on synethic data generation and experimental setup.

   Guidelines:

   - The answer [N/A] means that the paper does not include experiments.
   - If the paper includes experiments, a [No] answer to this question will not be perceived well by the reviewers: Making the paper reproducible is important, regardless of whether the code and data are provided or not.
   - If the contribution is a dataset and/or model, the authors should describe the steps taken to make their results reproducible or verifiable.
   - Depending on the contribution, reproducibility can be accomplished in various ways. For example, if the contribution is a novel architecture, describing the architecture fully might suffice, or if the contribution is a specific model and empirical evaluation, it may be necessary to either make it possible for others to replicate the model with the same dataset, or provide access to the model. In general. releasing code and data is often one good way to accomplish this, but reproducibility can also be provided via detailed instructions for how to replicate the results, access to a hosted model (e.g., in the case of a large language model), releasing of a model checkpoint, or other means that are appropriate to the research performed.
   - While NeurIPS does not require releasing code, the conference does require all submissions to provide some reasonable avenue for reproducibility, which may depend on the nature of the contribution. For example
     (a) If the contribution is primarily a new algorithm, the paper should make it clear how to reproduce that algorithm.
     (b) If the contribution is primarily a new model architecture, the paper should describe the architecture clearly and fully.
     (c) If the contribution is a new model (e.g., a large language model), then there should either be a way to access this model for reproducing the results or a way to reproduce the model (e.g., with an open-source dataset or instructions for how to construct the dataset).
     (d) We recognize that reproducibility may be tricky in some cases, in which case authors are welcome to describe the particular way they provide for reproducibility. In the case of closed-source models, it may be that access to the model is limited in some way (e.g., to registered users), but it should be possible for other researchers to have some path to reproducing or verifying the results.

5. **Open access to data and code**

Question: Does the paper provide open access to the data and code, with sufficient instructions to faithfully reproduce the main experimental results, as described in supplemental material?

Answer: [No]

Justification: We mainly utilize code from open sources (ManOpt and Geomstats) and we point to them in Section 4 and Appendix I. In Appendix I.3 we include citations and URLs to the reference of the open access data. While we don't include our specific code, we include thorough instructions to replicate our results.

Guidelines:

- The answer [N/A] means that paper does not include experiments requiring code.
- Please see the NeurIPS code and data submission guidelines (`https://neurips.cc/public/guides/CodeSubmissionPolicy`) for more details.
- While we encourage the release of code and data, we understand that this might not be possible, so [No] is an acceptable answer. Papers cannot be rejected simply for not including code, unless this is central to the contribution (e.g., for a new open-source benchmark).
- The instructions should contain the exact command and environment needed to run to reproduce the results. See the NeurIPS code and data submission guidelines (`https://neurips.cc/public/guides/CodeSubmissionPolicy`) for more details.
- The authors should provide instructions on data access and preparation, including how to access the raw data, preprocessed data, intermediate data, and generated data, etc.
- The authors should provide scripts to reproduce all experimental results for the new proposed method and baselines. If only a subset of experiments are reproducible, they should state which ones are omitted from the script and why.
- At submission time, to preserve anonymity, the authors should release anonymized versions (if applicable).
- Providing as much information as possible in supplemental material (appended to the paper) is recommended, but including URLs to data and code is permitted.

6. **Experimental setting/details**

   Question: Does the paper specify all the training and test details (e.g., data splits, hyperparameters, how they were chosen, type of optimizer) necessary to understand the results?

   Answer: [Yes]

   Justification: All parameters are included in Section 4 and Appendix I.

   Guidelines:

   - The answer [N/A] means that the paper does not include experiments.
   - The experimental setting should be presented in the core of the paper to a level of detail that is necessary to appreciate the results and make sense of them.
   - The full details can be provided either with the code, in appendix, or as supplemental material.

7. **Experiment statistical significance**

   Question: Does the paper report error bars suitably and correctly defined or other appropriate information about the statistical significance of the experiments?

   Answer: [Yes]

   Justification: Error bars are included in almost all plots. In the few plots without error bars, the results are deterministic.

   Guidelines:

   - The answer [N/A] means that the paper does not include experiments.
   - The authors should answer [Yes] if the results are accompanied by error bars, confidence intervals, or statistical significance tests, at least for the experiments that support the main claims of the paper.

- The factors of variability that the error bars are capturing should be clearly stated (for example, train/test split, initialization, random drawing of some parameter, or overall run with given experimental conditions).
- The method for calculating the error bars should be explained (closed form formula, call to a library function, bootstrap, etc.)
- The assumptions made should be given (e.g., Normally distributed errors).
- It should be clear whether the error bar is the standard deviation or the standard error of the mean.
- It is OK to report 1-sigma error bars, but one should state it. The authors should preferably report a 2-sigma error bar than state that they have a 96% CI, if the hypothesis of Normality of errors is not verified.
- For asymmetric distributions, the authors should be careful not to show in tables or figures symmetric error bars that would yield results that are out of range (e.g., negative error rates).
- If error bars are reported in tables or plots, the authors should explain in the text how they were calculated and reference the corresponding figures or tables in the text.

8. **Experiments compute resources**

   Question: For each experiment, does the paper provide sufficient information on the computer resources (type of compute workers, memory, time of execution) needed to reproduce the experiments?

   Answer: [Yes]

   Justification: We include details on the computing resources in Appendix I.

   Guidelines:

   - The answer [N/A] means that the paper does not include experiments.
   - The paper should indicate the type of compute workers CPU or GPU, internal cluster, or cloud provider, including relevant memory and storage.
   - The paper should provide the amount of compute required for each of the individual experimental runs as well as estimate the total compute.
   - The paper should disclose whether the full research project required more compute than the experiments reported in the paper (e.g., preliminary or failed experiments that didn't make it into the paper).

9. **Code of ethics**

   Question: Does the research conducted in the paper conform, in every respect, with the NeurIPS Code of Ethics `https://neurips.cc/public/EthicsGuidelines`?

   Answer: [Yes]

   Justification: The authors have read and reviewed the NeurIPS Code of Ethics. No violations have been committed. The data which is used is open access and does not contain identifiable information.

   Guidelines:

   - The answer [N/A] means that the authors have not reviewed the NeurIPS Code of Ethics.
   - If the authors answer [No], they should explain the special circumstances that require a deviation from the Code of Ethics.
   - The authors should make sure to preserve anonymity (e.g., if there is a special consideration due to laws or regulations in their jurisdiction).

10. **Broader impacts**

    Question: Does the paper discuss both potential positive societal impacts and negative societal impacts of the work performed?

    Answer: [Yes]

    Justification: The authors do not foresee negative societal impacts. Section 5 mentions the possible positive societal impacts.

Guidelines:

- The answer [N/A] means that there is no societal impact of the work performed.
- If the authors answer [N/A] or [No], they should explain why their work has no societal impact or why the paper does not address societal impact.
- Examples of negative societal impacts include potential malicious or unintended uses (e.g., disinformation, generating fake profiles, surveillance), fairness considerations (e.g., deployment of technologies that could make decisions that unfairly impact specific groups), privacy considerations, and security considerations.
- The conference expects that many papers will be foundational research and not tied to particular applications, let alone deployments. However, if there is a direct path to any negative applications, the authors should point it out. For example, it is legitimate to point out that an improvement in the quality of generative models could be used to generate Deepfakes for disinformation. On the other hand, it is not needed to point out that a generic algorithm for optimizing neural networks could enable people to train models that generate Deepfakes faster.
- The authors should consider possible harms that could arise when the technology is being used as intended and functioning correctly, harms that could arise when the technology is being used as intended but gives incorrect results, and harms following from (intentional or unintentional) misuse of the technology.
- If there are negative societal impacts, the authors could also discuss possible mitigation strategies (e.g., gated release of models, providing defenses in addition to attacks, mechanisms for monitoring misuse, mechanisms to monitor how a system learns from feedback over time, improving the efficiency and accessibility of ML).

11. **Safeguards**

    Question: Does the paper describe safeguards that have been put in place for responsible release of data or models that have a high risk for misuse (e.g., pre-trained language models, image generators, or scraped datasets)?

    Answer: [N/A]

    Justification: The paper does not contain data or models which are of high risk for misuse.

    Guidelines:

    - The answer [N/A] means that the paper poses no such risks.
    - Released models that have a high risk for misuse or dual-use should be released with necessary safeguards to allow for controlled use of the model, for example by requiring that users adhere to usage guidelines or restrictions to access the model or implementing safety filters.
    - Datasets that have been scraped from the Internet could pose safety risks. The authors should describe how they avoided releasing unsafe images.
    - We recognize that providing effective safeguards is challenging, and many papers do not require this, but we encourage authors to take this into account and make a best faith effort.

12. **Licenses for existing assets**

    Question: Are the creators or original owners of assets (e.g., code, data, models), used in the paper, properly credited and are the license and terms of use explicitly mentioned and properly respected?

    Answer: [Yes]

    Justification: We include citations and URL of the code we used in Appendix I. In Appendix I.3 we include citations and URL for the real data.

    Guidelines:

    - The answer [N/A] means that the paper does not use existing assets.
    - The authors should cite the original paper that produced the code package or dataset.
    - The authors should state which version of the asset is used and, if possible, include a URL.
    - The name of the license (e.g., CC-BY 4.0) should be included for each asset.

- For scraped data from a particular source (e.g., website), the copyright and terms of service of that source should be provided.
- If assets are released, the license, copyright information, and terms of use in the package should be provided. For popular datasets, `paperswithcode.com/datasets` has curated licenses for some datasets. Their licensing guide can help determine the license of a dataset.
- For existing datasets that are re-packaged, both the original license and the license of the derived asset (if it has changed) should be provided.
- If this information is not available online, the authors are encouraged to reach out to the asset's creators.

13. **New assets**

    Question: Are new assets introduced in the paper well documented and is the documentation provided alongside the assets?

    Answer: [N/A]

    Justification: We include thorough detail of the main contribution in Section 3.

    Guidelines:

    - The answer [N/A] means that the paper does not release new assets.
    - Researchers should communicate the details of the dataset/code/model as part of their submissions via structured templates. This includes details about training, license, limitations, etc.
    - The paper should discuss whether and how consent was obtained from people whose asset is used.
    - At submission time, remember to anonymize your assets (if applicable). You can either create an anonymized URL or include an anonymized zip file.

14. **Crowdsourcing and research with human subjects**

    Question: For crowdsourcing experiments and research with human subjects, does the paper include the full text of instructions given to participants and screenshots, if applicable, as well as details about compensation (if any)?

    Answer: [N/A]

    Justification: The experiments did not have any human subjects. The data which was used is open source and we reference it in Section I.3.

    Guidelines:

    - The answer [N/A] means that the paper does not involve crowdsourcing nor research with human subjects.
    - Including this information in the supplemental material is fine, but if the main contribution of the paper involves human subjects, then as much detail as possible should be included in the main paper.
    - According to the NeurIPS Code of Ethics, workers involved in data collection, curation, or other labor should be paid at least the minimum wage in the country of the data collector.

15. **Institutional review board (IRB) approvals or equivalent for research with human subjects**

    Question: Does the paper describe potential risks incurred by study participants, whether such risks were disclosed to the subjects, and whether Institutional Review Board (IRB) approvals (or an equivalent approval/review based on the requirements of your country or institution) were obtained?

    Answer: [N/A]

    Justification: There were no human subjects, so this is not discussed.

    Guidelines:

    - The answer [N/A] means that the paper does not involve crowdsourcing nor research with human subjects.

- Depending on the country in which research is conducted, IRB approval (or equivalent) may be required for any human subjects research. If you obtained IRB approval, you should clearly state this in the paper.
- We recognize that the procedures for this may vary significantly between institutions and locations, and we expect authors to adhere to the NeurIPS Code of Ethics and the guidelines for their institution.
- For initial submissions, do not include any information that would break anonymity (if applicable), such as the institution conducting the review.

16. **Declaration of LLM usage**

    Question: Does the paper describe the usage of LLMs if it is an important, original, or non-standard component of the core methods in this research? Note that if the LLM is used only for writing, editing, or formatting purposes and does *not* impact the core methodology, scientific rigor, or originality of the research, declaration is not required.

    Answer: [N/A]

    Justification: LLMs were not used for any of the core methods in this research.

    Guidelines:

    - The answer [N/A] means that the core method development in this research does not involve LLMs as any important, original, or non-standard components.
    - Please refer to our LLM policy in the NeurIPS handbook for what should or should not be described.