Riemannian Ordinary Least Squares

Xiaoyu Chen[1*], **Yujing (Zipan) Huang**[1], **Yingyan Zeng**[2]
[1]University at Buffalo [2]University of Cincinnati

Abstract

Many scientific responses take values in Riemannian manifolds, with predictors that are scalar- and/or manifold-valued. A regression analysis in such settings must provide accurate prediction with hypothesis testing for individual coefficients. Yet no current manifold-regression framework accommodates both predictor types and enables per-coefficient hypothesis tests. We propose Riemannian Ordinary Least Squares (ROLS) as a tangent space conditional-mean model, which is the ordinary least squares (OLS) analogue for Riemannian manifolds. ROLS is well-defined under an explicit injectivity condition that keeps the logarithm maps single-valued, with cut-locus conditions for branch-sensitive cases outside it. Within the local geometric setting it admits a closed-form estimator, an $O(n^{-1})$ excess prediction-error bound with explicit curvature bias, and asymptotic t- and Wald tests for individual coefficients. With synthetic simulations and three real-data case studies, ROLS is shown to be competitive with the state-of-the-art baselines. In addition, the per-coefficient tests agree with effects reported in literature. We envision ROLS to be a standard tool for manifold-valued regression with coefficient-level inference to support scientific discovery, which can be interpreted in the same way as Euclidean OLS. The source code is available at `https://github.com/XiaoyuChenPub/ROLS`.

1 Introduction

In scientific applications, regression analysis is used not only to model predictor-response relationships, but also to determine which covariates have statistically significant effects. For Euclidean data, ordinary least squares (OLS) is the standard tool, because the estimator is closed form and each coefficient can be tested. Analogous questions arise when responses lie on Riemannian manifolds, including diffusion tensors in $\mathrm{Sym}^+(3)$ for brain microstructure [Pennec et al., 2006, Brooks et al., 2019], protein backbone dihedral angles on the flat torus $\mathbb{T}^2$, and wind directions on S^2 in geophysics. In such applications, the covariates are typically heterogeneous, combining manifold-valued and scalar predictors, yet scientific interpretation still requires the same tractable estimation and coefficient-level inference that OLS provides in Euclidean regression.

Existing manifold-regression frameworks each address only a strict subset of these estimation and inference requirements. Geodesic regression [Fletcher, 2013, Hinkle et al., 2012], multivariate general linear models (MGLM) on Riemannian manifolds [Kim et al., 2014], and symmetric-space regression [Shi et al., 2009, Cornea et al., 2017] provide inference for scalar covariates [Fletcher, 2013, Kim et al., 2014, Cornea et al., 2017], but none of them accommodates manifold-valued predictors. Symmetric-space regression further restricts the response to a Riemannian symmetric space, excluding spaces such as the Wasserstein space. Fréchet regression [Petersen and Müller, 2019a, Lin and Müller, 2021] and its extensions [Bhattacharjee and Müller, 2023, Lin et al., 2023, Han and Song, 2023, Zhang et al., 2024] accommodate general metric-space responses. However, they yield neither closed-form estimators nor individual coefficient tests, aside from specialized F-tests for Bures-Wasserstein

*Corresponding author: `xchen325@buffalo.edu`.

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

Table 1: Comparison of regression methods for manifold-valued responses.

Method	Manifold predictors	Scalar predictors	Individual coeff. tests	Response
Geodesic reg. [Fletcher, 2013]†	×	✓	✓	General manifolds
Polynomial reg. [Hinkle et al., 2012]	×	✓	×	General manifolds
MGLM [Kim et al., 2014]	×	✓	✓	General manifolds
Cornea et al. [2017]	×	✓	✓	Sym. sp.
Fréchet reg. [Petersen and Müller, 2019a]	×	✓	×	Metric sp.
Single-idx Fr. [Bhattacharjee and Müller, 2023]	×	✓	×	Metric sp.
Additive mfld. [Lin et al., 2023]	×	✓	×	Lie grp.
NL Fréchet [Bhattacharjee et al., 2025]	✓	✓	×	Metric sp.
DIDO reg. [Chen et al., 2024]	✓	×	×	$\mathcal{P}_2(\mathbb{R})$
Riemannian OLS (this paper)	✓	✓	✓	General manifolds

† The tangent-space formulation of geodesic regression (base point fixed at the Fréchet mean) is a special case of this paper's method with $p{=}0,\ q{=}1$; this paper additionally supplies an asymptotic Wald test for the geodesic-regression coefficient.

responses [Xu and Li, 2025]. Intrinsic Gaussian-process regression [Mallasto and Feragen, 2018] similarly focuses on prediction without providing coefficient-level inference. Conversely, methods that accept manifold-valued or non-Euclidean predictors [Petersen and Müller, 2019b, Zhou and Müller, 2022, Capitaine et al., 2024] are either domain-specific or nonparametric. In particular, distribution-in-distribution-out (DIDO) regression [Chen et al., 2024] is restricted to $\mathcal{P}_2(\mathbb{R})$ without scalar covariates, while nonlinear Fréchet (NL-Fréchet) regression [Bhattacharjee et al., 2025] is purely predictive. Consequently, no existing framework simultaneously provides (i) mixed manifold-valued and scalar covariates, (ii) individual coefficient tests, including tests for manifold-predictor coefficients, and (iii) explicit injectivity conditions ensuring that the tangent-space linearization is well defined (Table 1 summarizes (i) and (ii)).

To address these gaps, we propose Riemannian Ordinary Least Squares (ROLS), a regression framework based on tangent-space linearization at the Fréchet mean. ROLS lifts each manifold-valued variable to the tangent space at its Fréchet mean via the Riemannian logarithm map, parallel transports the lifted manifold predictors to the response tangent space, and fits a linear conditional mean model for the lifted response. Because the logarithm map and geodesic parallel transport are only locally unique, we formalize explicit support and injectivity conditions in Section 3.2 to separate the well-posed local setting from the branch-sensitive case near the cut locus. Under these local conditions, the resulting block normal equations admit a closed-form solution, and the linear tangent-space structure yields asymptotic t-tests for manifold predictors and Wald tests for scalar predictors (Sections 3 and 4). ROLS thus plays the same role for manifold-valued responses that classical OLS plays in Euclidean regression, serving as both an interpretable inferential tool and a natural baseline.

Specifically, this paper makes four contributions:

- We develop a unified, closed-form regression estimator for manifold-valued responses with mixed manifold and scalar predictors, which recovers tangent-space geodesic regression, DIDO regression, and Euclidean Fréchet regression as special cases (Appendix F).
- We clarify when tangent-space linearization is geometrically valid by characterizing the local support and injectivity conditions required for estimation and inference, and by quantifying how cut-locus branching compromises identifiability outside this neighborhood (Theorem 3.4).
- We enable coefficient-level hypothesis testing via asymptotic t- and Wald tests backed by a consistent pairs bootstrap that accounts for estimated Fréchet means and transports (Corollary 4.6). We also establish an $O(n^{-1})$ excess prediction-error bound that explicitly isolates the curvature-induced bias of tangent-space linearization (Theorem 4.2).
- Across five simulated manifolds and three real-data applications on $\mathrm{Sym}^+(3)$, $\mathbb{T}^2$, and S^2, we show that ROLS achieves superior predictive accuracy over state-of-the-art baselines (Tables 5 and 2) while uncovering domain-consistent covariate effects that purely predictive methods cannot test.

2 Notation and Preliminaries

Throughout the paper, $i = 1, \dots, n$ indexes observations, $j = 1, \dots, p$ indexes manifold-valued covariates, and $k = 1, \dots, q$ indexes scalar covariates. The response is $Y \in \mathcal{M}$, the manifold covariates are $X_1, \dots, X_p \in \mathcal{M}$, and the scalar covariates are $z_1, \dots, z_q \in \mathbb{R}$. The response and all manifold-valued covariates are assumed to lie in the same Riemannian manifold $\mathcal{M}$, since the model uses parallel transport between their Fréchet means. We write $\bar{Y}$ and $\bar{X}_j$ for population Fréchet means, $\hat{\bar{Y}}$ and $\hat{\bar{X}}_j$ for their empirical analogues, $\bar{z}_k = \mathbb{E}z_k$, and $\tilde{z}_k = z_k - \bar{z}_k$. All logarithm, exponential, and transport maps are written with their base points as explicit subscripts. Norms and inner products are taken in the tangent space indicated by context, usually $T_{\bar{Y}}\mathcal{M}$ after linearization. The ROLS parameter is $\boldsymbol{\beta} = (\alpha_1, \dots, \alpha_p, v_1, \dots, v_q)$, with scalar coefficients $\alpha_j \in \mathbb{R}$ for manifold covariates and tangent-vector coefficients $v_k \in T_{\bar{Y}}\mathcal{M}$ for scalar covariates.

Let $(\mathcal{M}, g)$ be a complete, connected Riemannian manifold of dimension d with geodesic distance $d_{\mathcal{M}}(p, q)$. The tangent space $T_p\mathcal{M}$ at each point p is a d-dimensional inner product space under the Riemannian metric g_p. The exponential map $\mathrm{Exp}_p \colon T_p\mathcal{M} \to \mathcal{M}$ sends a tangent vector v to the endpoint of the unit-time geodesic starting at p with initial velocity v, and its local inverse is the logarithm map Log_p. Let inj_p denote the injectivity radius at p and $\mathrm{inj}(\mathcal{M}) = \inf_{p \in \mathcal{M}} \mathrm{inj}_p$ the global injectivity radius. On the geodesic ball $B(p, \mathrm{inj}_p)$, Log_p is single-valued and smooth. The cut locus $\mathrm{Cut}(p)$ is the set where geodesics from p cease to minimize length or cease to be unique; equivalently, it marks the boundary beyond which Log_p may become multi-valued.

Parallel transport relates tangent vectors at different base points. For a curve γ from p to q, we write $\Gamma^{\gamma}_{p \to q} \colon T_p\mathcal{M} \to T_q\mathcal{M}$ for parallel transport along γ via the Levi-Civita connection; this map is a linear isometry. When $d_{\mathcal{M}}(p, q) < \min(\mathrm{inj}_p, \mathrm{inj}_q)$ and the minimizing geodesic is unique, we use that geodesic and abbreviate $\Gamma^{\gamma}_{p \to q}$ to $\Gamma_{p \to q}$. If a unique minimizing geodesic is not available, ROLS requires a fixed transport convention specified before fitting.

For a random variable X on $\mathcal{M}$ with finite second moment, the Fréchet mean is any minimizer of $p \mapsto \mathbb{E}[d^2_{\mathcal{M}}(X, p)]$ [Bhattacharya and Patrangenaru, 2003, 2005]. Under the Afsari condition (the support of X lies in a geodesic ball of radius $\rho < \frac{1}{2}\min(\mathrm{inj}_p, \pi/\sqrt{\kappa_+})$ around some center p, where κ_+ is an upper bound on the sectional curvature), the Fréchet mean exists uniquely [Afsari, 2011] and satisfies a central limit theorem (CLT) [Bhattacharya and Lin, 2017]; we denote it $\bar{X}$.

Curvature controls the distortion between tangent-space and manifold geometry. Let $\kappa_- \le \mathrm{Sec}(\mathcal{M}) \le \kappa_+$ denote sectional curvature bounds. The Hessian and Rauch comparison theorems [Cordero-Erausquin et al., 2001] introduce two distortion factors (explicit formulas in Appendix A): $\eta(\kappa_+, r) \in (0, 1]$, which contracts distances on positively curved manifolds, and $\tilde{\eta}(\kappa_-, r) \ge 1$, which expands distances on negatively curved ones. On non-negatively curved manifolds (S^2, $\mathrm{SO}(3)$), neither factor inflates ($\eta \le 1$, $\tilde{\eta} = 1$); on $\mathrm{Sym}^+(k)$ ($\kappa_- < 0$), $\tilde{\eta} > 1$.

3 Riemannian Ordinary Least Squares

We formulate the ROLS model and derive its closed-form estimator, then characterize the cut-locus condition under which the estimator is identifiable, together with the branch-sensitive failure mode.

3.1 ROLS Model and Estimation

ROLS operates in the tangent space at the response Fréchet mean $T_{\bar{Y}}\mathcal{M}$, see the construction in Figure 1. For each observation i, every manifold-valued covariate is first mapped to the tangent space at its own mean via the logarithm map, then parallel-transported to $T_{\bar{Y}}\mathcal{M}$, giving $u_{ij} \coloneqq \Gamma_{\bar{X}_j \to \bar{Y}}\big(\mathrm{Log}_{\bar{X}_j}(X_{ij})\big) \in T_{\bar{Y}}\mathcal{M}$ for $j = 1, \dots, p$. The response is lifted to the tangent space as $w_i \coloneqq \mathrm{Log}_{\bar{Y}}(Y_i) \in T_{\bar{Y}}\mathcal{M}$.

Definition 3.1 (Riemannian Ordinary Least Squares)**.** *ROLS specifies a conditional mean model after lifting all variables to the common tangent space $T_{\bar{Y}}\mathcal{M}$:*

$$\mathrm{Log}_{\bar{Y}}(Y) = \sum_{j=1}^{p} \alpha_j \Gamma_{\bar{X}_j \to \bar{Y}}\big(\mathrm{Log}_{\bar{X}_j}(X_j)\big) + \sum_{k=1}^{q} (z_k - \bar{z}_k)\, v_k + \varepsilon, \tag{1}$$

where $\varepsilon \in T_{\bar{Y}}\mathcal{M}$ is the residual at the population Fréchet least-squares (FLS) minimizer.

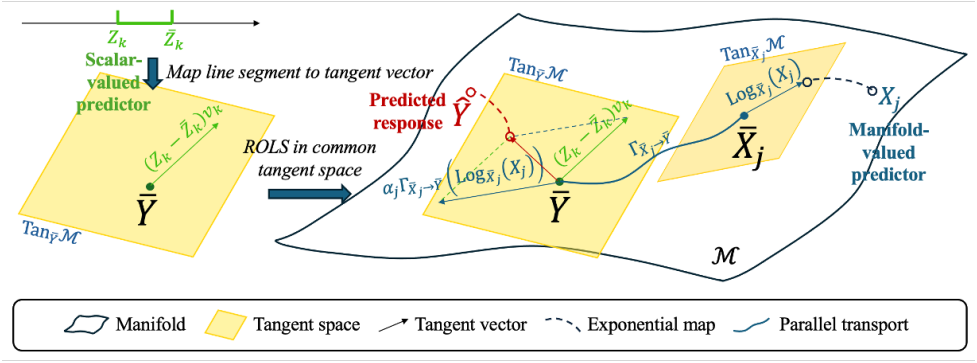


Figure 1: Geometric overview of ROLS. Each manifold predictor X_j is log-mapped to $T_{\bar{X}_j}\mathcal{M}$ and parallel-transported to the common tangent space $T_{\bar{Y}}\mathcal{M}$, where it is weighted by the scalar coefficient α_j. Each scalar predictor z_k contributes a centered tangent direction $(z_k - \bar{z}_k)\, v_k$. The predicted response is $\hat{Y} = \mathrm{Exp}_{\bar{Y}}\big(\sum_j \alpha_j u_j + \sum_k (z_k - \bar{z}_k) v_k\big)$, where u_j is the transported predictor.

Remark 3.2 (Special cases)**.** *Several regression models are recovered by restricting the spaces or the covariate classes. In Euclidean space, ROLS becomes a centered vector-valued linear model with scalar block coefficients for vector covariates, and it reduces to ordinary centered multiple linear regression when $d = 1$. With no manifold predictors and one scalar predictor, it gives the tangent-space form of geodesic regression. Appendix F gives special cases.*

To estimate the parameters, define the FLS objective as the tangent-space quadratic

$$L(\boldsymbol{\alpha}, \boldsymbol{v}) := \frac{1}{2}\mathbb{E}\Big[\|w - m_{\boldsymbol{\beta}}(X, z)\|^2_{g_{\bar{Y}}}\Big], \tag{2}$$

where $m_{\boldsymbol{\beta}}(X, z) = \sum_j \alpha_j u_j + \sum_k \tilde{z}_k v_k$ is the conditional mean prediction in $T_{\bar{Y}}\mathcal{M}$. On flat manifolds L coincides with the manifold Fréchet risk $\frac{1}{2}\mathbb{E}[d^2_{\mathcal{M}}(Y, \mathrm{Exp}_{\bar{Y}}(m_{\boldsymbol{\beta}}(X, z)))]$; on curved manifolds the two are related by a multiplicative factor at most $\tilde{\eta}(\kappa_-, r_0)^2$, where r_0 denotes the data spread $\sup_i d_{\mathcal{M}}(Y_i, \bar{Y})$ (Remark 4.4). Since L is quadratic in $(\boldsymbol{\alpha}, \boldsymbol{v})$, the optimal coefficients satisfy a block linear system analogous to the Euclidean normal equations. Define the design operator for observation i as the $(p+q)$-tuple $\boldsymbol{Z}_i = (u_{i1}, \ldots, u_{ip}, \tilde{z}_{i1}\,\mathrm{id}, \ldots, \tilde{z}_{iq}\,\mathrm{id})$, and stack the parameter vector $\boldsymbol{\beta} = (\alpha_1, \ldots, \alpha_p, v_1, \ldots, v_q)^\top$.

Theorem 3.3 (Generalized Normal Equations on the mixed design space)**.** *Fix an orthonormal frame of $T_{\bar{Y}}\mathcal{M}$ and identify it with $\mathbb{R}^d$. On the mixed Hilbert design space $\mathcal{H} := \mathbb{R}^p \oplus (T_{\bar{Y}}\mathcal{M})^q \cong \mathbb{R}^{p+qd}$, the population FLS minimizer $\boldsymbol{\beta}^* = (\alpha_1^*, \ldots, \alpha_p^*, v_1^*, \ldots, v_q^*) \in \mathcal{H}$ satisfies the block normal equations*

$$\Sigma_{\boldsymbol{ZZ}}\, \boldsymbol{\beta}^* = C_{\boldsymbol{ZY}}, \tag{3}$$

where $\Sigma_{\boldsymbol{ZZ}} \in \mathbb{R}^{(p+qd)\times(p+qd)}$ is the self-adjoint block Gram matrix

$$\Sigma_{\boldsymbol{ZZ}} = \begin{pmatrix} \big[\mathbb{E}\, g_{\bar{Y}}(u_j, u_\ell)\big]_{j,\ell=1}^{p} & \big[\mathbb{E}\, \tilde{z}_k\, u_j^\top\big]_{j,k} \\ \big[\mathbb{E}\, \tilde{z}_k\, u_j\big]_{k,j} & \big[\mathbb{E}\, \tilde{z}_k \tilde{z}_\ell\, I_d\big]_{k,\ell=1}^{q} \end{pmatrix}, \tag{4}$$

and $C_{\boldsymbol{ZY}} \in \mathbb{R}^{p+qd}$ has components $\mathbb{E}\, g_{\bar{Y}}(u_j, w) \in \mathbb{R}$ and $\mathbb{E}\, \tilde{z}_k\, w \in \mathbb{R}^d$. When $\Sigma_{\boldsymbol{ZZ}} \succ 0$, the unique minimizer is $\boldsymbol{\beta}^* = \Sigma_{\boldsymbol{ZZ}}^{-1} C_{\boldsymbol{ZY}}$.

The block structure in Theorem 3.3 is the direct analogue of Euclidean multiple regression. The upper-left block records inner products between transported manifold predictors, the lower-right block is the usual scalar-covariate covariance matrix repeated across tangent coordinates, and the off-diagonal blocks are the cross-covariances between scalar and manifold covariates. Solving (3) therefore estimates both sets of parameters jointly within a single linear system, adjusting each effect for the collinearity between manifold and scalar predictors. In this parameterization, each scalar covariate is associated with a directional coefficient vector $v_k \in T_{\bar{Y}}\mathcal{M}$, whereas each manifold predictor is weighted by a scalar coefficient $\alpha_j \in \mathbb{R}$ multiplying its transported tangent displacement. The proof follows by reducing the FLS objective to a quadratic in the Hilbert space $(T_{\bar{Y}}\mathcal{M}, g_{\bar{Y}})$ and differentiating, see Appendix B.

In practice, $\bar{Y}$ and $\bar{X}_j$ are approximated by empirical Fréchet means, and expectations by sample averages. The point estimator is therefore closed-form. We compute $\hat{\bar{Y}}$ and $\hat{\bar{X}}_j$, transport predictors $\hat{u}_{ij} = \Gamma^{\gamma_j}_{\hat{\bar{X}}_j \to \hat{\bar{Y}}}\big(\mathrm{Log}_{\hat{\bar{X}}_j}(X_{ij})\big)$, center $\tilde{z}_{ik}$, log-map $w_i = \mathrm{Log}_{\hat{\bar{Y}}}(Y_i)$, and solve the block normal equations for $\hat{\boldsymbol{\beta}} = (\hat{\alpha}, \hat{v})$. The maps are evaluated only at the $p+1$ estimated means, so a fit needs $O(np)$ logarithm and transport evaluations (timings are reported in Appendix H.8).

3.2 Geometric Identifiability

The basic operations of ROLS are meaningful only when the logarithm maps at the predictor and response Fréchet means are single-valued on the observed supports. In Euclidean space this condition is automatic. However, on a curved manifold, the logarithm map can become multi-valued at the cut locus. So the tangent design may depend on the chosen log branch, and Fréchet means near the cut locus can obey non-standard (smeary) limit theorems [Eltzner and Huckemann, 2019]. The next theorem isolates the geometric condition under which the estimator and statistical theory operate, then records the branch-sensitive failure mode outside it. See proof in Appendix C.

Theorem 3.4 (Cut-locus identifiability)**.** *Assume* $\Sigma_{\boldsymbol{ZZ}} \succ 0$.

(a) *If every manifold predictor and the response are supported in strongly convex normal balls around their Fréchet means,*

$$\mathrm{supp}(X_j) \subset B(\bar{X}_j, r_j), \quad r_j < \mathrm{inj}_{\bar{X}_j}, \qquad \mathrm{supp}(Y) \subset B(\bar{Y}, r_Y), \quad r_Y < \mathrm{inj}_{\bar{Y}}, \tag{5}$$

then all log maps used by ROLS are single-valued and smooth, and the population FLS objective has a unique minimizer.

(b) *If a predictor X_j admits two measurable log branches ℓ, ℓ' on a positive-probability set S, write $\boldsymbol{Z}^{(\ell)}$ for the design matrix built using branch ℓ in slot j and define $\Sigma^{(\ell)} = \mathbb{E}[\boldsymbol{Z}^{(\ell)}(\boldsymbol{Z}^{(\ell)})^\top]$, $C^{(\ell)} = \mathbb{E}[\boldsymbol{Z}^{(\ell)} W]$ with $W = \mathrm{Log}_{\bar{Y}}(Y)$ (analogously for ℓ'), $\Delta\Sigma = \Sigma^{(\ell)} - \Sigma^{(\ell')}$, $\Delta C = C^{(\ell)} - C^{(\ell')}$, and $\Delta Z_{ij} = \Gamma_{\bar{X}_j \to \bar{Y}}\{\ell(X_{ij}) - \ell'(X_{ij})\}\mathbf{1}\{X_{ij} \in S\}$. Let $P_j^\perp$ denote the projection in $L^2(T_{\bar{Y}}\mathcal{M})$ orthogonal to the remaining design (Appendix C). The branch-induced coefficient gap obeys the exact identity $\boldsymbol{\beta}^*_{(\ell)} - \boldsymbol{\beta}^*_{(\ell')} = (\Sigma^{(\ell')})^{-1}\big(\Delta C - \Delta\Sigma\, \boldsymbol{\beta}^*_{(\ell)}\big)$.*

Remark 3.5 (Predictively non-trivial gaps)**.** *When $P_j^\perp \Delta Z_j = 0$ in L^2, the branch perturbation is absorbed by the remaining covariates, so $\boldsymbol{\beta}^*_{(\ell)}$ and $\boldsymbol{\beta}^*_{(\ell')}$ parameterize the same population predictor even when they differ. A gap that changes the fitted predictor requires both $P_j^\perp \Delta Z_j \neq 0$ in L^2 and $\Delta C - \Delta\Sigma\, \boldsymbol{\beta}^*_{(\ell)} \neq 0$ in $\mathbb{R}^{p+qd}$ (Appendix C).*

4 Statistical Properties

Under the conditions of Subsection 3.2, the closed-form ROLS estimator reduces to ordinary least squares in the product tangent space $\mathbb{R}^p \times (T_{\bar{Y}}\mathcal{M})^q$, with three manifold-specific modifications. Curvature distorts the relation between tangent residuals and geodesic errors, tangent linearization can introduce approximation bias, and the tangent spaces themselves are anchored at empirical Fréchet means. The theory below follows the estimator's order of operations. First we show that the fitted conditional mean attains the expected parametric prediction behavior. We then pass from prediction to coefficient error, and finally to the asymptotic distribution needed for individual tests.

Assumption 4.1. *(i) The Fréchet means $\bar{Y}$ and each $\bar{X}_j$ exist uniquely, the variables are supported within the corresponding Afsari ball, and they have finite second moments there. (ii) The injectivity condition* (5) *holds almost surely, and the predictions $\mathrm{Exp}_{\bar{Y}}(m_{\hat{\boldsymbol{\beta}}}(X, z))$ lie within radius r_0 of $\bar{Y}$. (iii) The block Gram matrix $\Sigma_{\boldsymbol{ZZ}}$ has smallest eigenvalue $\lambda_* > 0$, excluding degenerate mixed-design configurations. (iv) The population residual $\varepsilon = w - \sum_j \alpha_j^* u_j - \sum_k \tilde{z}_k v_k^*$ at the FLS minimizer $\boldsymbol{\beta}^*$ has finite second moment, decomposed as $\mathbb{E}[\|\varepsilon\|^2] = \sigma^2 + b^2$, where σ^2 is the irreducible noise variance and $b^2 \geq 0$ is the squared tangent-linear misspecification bias. Under correct specification $\mathbb{E}[\varepsilon \mid X_1, \ldots, X_p, z_1, \ldots, z_q] = 0$ (invoked in Theorem 4.5), $b = 0$ and $\sigma^2 = \mathbb{E}[\|\varepsilon\|^2]$. Population orthogonality $\mathbb{E}[\boldsymbol{Z}\,\varepsilon] = 0$ holds automatically at $\boldsymbol{\beta}^*$. The observations are independent and identically distributed.*

4.1 Consistency and Convergence Rate

The first question is whether the closed-form normal equations in Theorem 3.3 lead to a statistically stable predictor. In Euclidean OLS, given a full-rank design, the leading prediction-error term is the familiar variance contribution $\sigma^2(p+qd)/n$. On a manifold the tangent residual and the geodesic residual differ by a Rauch factor, and a nonlinear relation in normal coordinates adds a deterministic bias. Theorem 4.2 states the Euclidean rate, the curvature distortion, and the bias term explicitly.

Theorem 4.2 (Consistency and convergence rate)**.** *Under Assumption 4.1 and finite fourth moments, let $\boldsymbol{\beta}^* = (\alpha_1^*, \dots, \alpha_p^*, v_1^*, \dots, v_q^*)$ be the population FLS minimizer and write $b^2 := \mathbb{E}[\|w - \sum_j \alpha_j^* u_j - \sum_k \tilde{z}_k v_k^*\|^2] - \sigma^2$ for the squared tangent-linear misspecification bias. For a new observation (X, z, Y), independent of the sample, the empirical FLS estimator satisfies*

$$\mathbb{E}\big[d_{\mathcal{M}}^2(\hat{Y}, Y)\big] \le \tilde{\eta}(\kappa_-, r_0)^2 \left\{ \frac{C\sigma^2(p+qd)}{n\lambda_*} + \sigma^2 + b^2 \right\},$$

where C depends on fourth moments and $\tilde{\eta}$ is the Rauch distortion factor. If $b = 0$, then $\hat{\boldsymbol{\beta}}_n \xrightarrow{P} \boldsymbol{\beta}^*$ and the estimation-driven excess prediction term is $O((p+qd)/n)$ in expectation.

Theorem 4.2 decomposes the prediction error into estimation variance, curvature distortion, and linearization bias. The leading term $C\sigma^2(p+qd)/(n\lambda_*)$ matches the Euclidean OLS variance in the mixed product tangent space and depends on the ratio $n/(p+qd)$. The term b^2 is the approximation error from higher-order curvature effects. Ambient curvature scales the squared-error bound by $\tilde{\eta}^2$. The root mean squared error (RMSE) bound has the factor $\tilde{\eta}$ from (8), and the parametric rate is unchanged. In particular, $\tilde{\eta} = 1$ on flat and non-negatively curved manifolds ($\mathbb{T}^2$, S^2, $\mathrm{SO}(3)$) and remains finite on non-positively curved manifolds ($\mathrm{Sym}^+(3)$, $\mathbb{H}^3$) under bounded support.

The prediction bound controls the fitted conditional mean. For inference we also need the coefficients themselves to concentrate in the product tangent space.

Corollary 4.3 (Parameter estimation rate)**.** *Under the same conditions, $\|\hat{\boldsymbol{\beta}}_n - \boldsymbol{\beta}^*\|^2 = O_P((p+qd)/n)$ in the product tangent space, with constants depending on λ_* and the fourth moments M_4.*

Remark 4.4 (Curvature distortion and bias)**.** *When the true relationship is genuinely geodesic on the manifold rather than linear in $T_{\bar{Y}}\mathcal{M}$, normal-coordinate expansions give a leading curvature correction of order Kr_0^3 in tangent radius [do Carmo, 1992, Pennec et al., 2006]. The constant $K = \max(|\kappa_-|, \kappa_+)$ bounds the absolute sectional curvature. Squaring gives $b^2 = O(K^2 r_0^6)$. The Afsari support condition therefore keeps the bias controlled by restricting the data to a normal neighborhood where higher-order curvature terms remain small. On flat manifolds, including $\mathcal{P}_2(\mathbb{R})$, $\tilde{\eta} = 1$ and the curvature contribution to b is zero. ROLS then reduces to the Euclidean FLS estimator.*

Corollary 4.3 transfers the prediction-error bound to the parameter level in the product tangent space. The coefficient tests in the next subsection use this scale.

4.2 Asymptotic Normality and Wald Tests

The final step is coefficient-level inference. In Euclidean OLS this follows from a CLT for the score. In ROLS, the score is still the ordinary tangent-space score, but it is evaluated after estimating the Fréchet means that define the log maps and transports. Those means fluctuate at order $n^{-1/2}$, so their influence functions appear in the same limit as the score. The next theorem records the resulting asymptotic normality and the tests it supports.

Theorem 4.5 (Asymptotic normality and statistical inference)**.** *Assume Assumption 4.1, finite fourth moments, correct specification ($b = 0$), and C^2 dependence of the log and transport maps on the Fréchet-mean base points in a geodesic neighborhood. Assume also that the Fréchet functions of Y and X_j are twice continuously differentiable near their means with nonsingular Hessians, so that the empirical Fréchet means admit asymptotically linear representations. Writing $O_i = (Y_i, X_{i1}, \dots, X_{ip}, z_{i1}, \dots, z_{iq})$ for the i-th observation tuple,*

$$\sqrt{n}(\hat{\boldsymbol{\beta}}_n - \boldsymbol{\beta}^*) = n^{-1/2} \sum_{i=1}^{n} \mathrm{IF}(O_i) + o_P(1) \xrightarrow{d} \mathcal{N}(0, \Sigma_{\boldsymbol{\beta}}), \tag{6}$$

where IF *is the influence function* (45)*,* $\Sigma_{\boldsymbol{\beta}} = \mathbb{E}[\mathrm{IF}(O)\mathrm{IF}(O)^\top]$*, and the difference* $\hat{v}_k - v_k^*$ *is taken after parallel-transporting the frame from* $\bar{Y}$ *to* $\hat{\bar{Y}}$*. Consequently,* t*-tests for manifold coefficients* $H_0 : \alpha_j = 0$ *and Wald* χ^2_d *tests for scalar-predictor tangent coefficients* $H_0 : v_k = 0$ *are valid for any consistent estimator of* $\Sigma_{\boldsymbol{\beta}}$*, provided the corresponding scalar variance* $[\Sigma_{\boldsymbol{\beta}}]_{jj}$ *or tangent-vector block* $\Sigma^{kk}_{\boldsymbol{\beta}}$ *is nonsingular. The proof is in Appendix E.*

Corollary 4.6 (Pairs bootstrap validity)**.** *Under the assumptions of Theorem 4.5, let* $\hat{\boldsymbol{\beta}}^*_n$ *denote the ROLS estimator computed on a pairs bootstrap resample, including recomputation of the empirical Fréchet means, log maps, transports, and block normal equations. Conditional on the data,* $\sqrt{n}(\hat{\boldsymbol{\beta}}^*_n - \hat{\boldsymbol{\beta}}_n) \xrightarrow{d_P} \mathcal{N}(0, \Sigma_{\boldsymbol{\beta}})$*, and the second-moment matrix of this difference, truncated at a level that grows sufficiently slowly, consistently estimates* $\Sigma_{\boldsymbol{\beta}}$*. Thus the resulting standard errors and studentized coefficient tests are asymptotically valid.*

Appendix E.1 proves Corollary 4.6 by treating the Fréchet means and $\boldsymbol{\beta}$ as one Z-estimator. The same truncation produces a covariance estimator for the full $\Sigma_{\boldsymbol{\beta}}$. Plugging any such covariance into the tests of Theorem 4.5 gives $T_j = \hat{\alpha}_j / \hat{\mathrm{se}}(\hat{\alpha}_j)$ for $H_0 : \alpha_j = 0$ and $W_k = n\hat{v}_k^\top (\hat{\Sigma}^{kk}_{\boldsymbol{\beta}})^{-1} \hat{v}_k$ for $H_0 : v_k = 0$. For the standard errors reported in the simulations and Section 6, we use the model-based estimator of the tangent-space component of $\Sigma_{\boldsymbol{\beta}}$. With tangent design $\boldsymbol{Z}_i = [u_{i1}, \ldots, u_{ip}, \tilde{z}_{i1} I_d, \ldots, \tilde{z}_{iq} I_d]^\top \in \mathbb{R}^{(p+qd)\times d}$, Gram matrix $\hat{A} = \sum_i \boldsymbol{Z}_i \boldsymbol{Z}_i^\top$, and residual $r_i = w_i - \boldsymbol{Z}_i^\top \hat{\boldsymbol{\beta}}$, this estimator is $n\hat{\sigma}^2 \hat{A}^{-1}$, where $\hat{\sigma}^2 = \sum_i \|r_i\|^2 / \nu$ and $\nu = nd - p - qd - d$ removes d degrees of freedom for the estimated response Fréchet mean. Thus T_j is compared with t_ν, while W_k is compared with χ^2_d. The χ^2_d approximation over-rejects when n/d is small, because $\hat{A}$ then estimates a Gram matrix of dimension $p + qd$ (Appendix H.7). A pairs-bootstrap calibration of these same statistics recomputes the model-based covariance on each resample, and parallel-transports $\hat{v}_k^*$ to $T_{\hat{\bar{Y}}}\mathcal{M}$ before forming $\hat{v}_k^* - \hat{v}_k$ (pairs variant of Algorithm 1).

Misspecification-robust standard errors. The covariance $n\hat{\sigma}^2 \hat{A}^{-1}$ pools the tangent residuals into one variance. When $b > 0$, $\boldsymbol{\beta}^*$ is the best tangent-linear approximation at the Fréchet mean and the residuals vary with the design, so this matrix does not estimate their contribution to $\Sigma_{\boldsymbol{\beta}}$. Under the same tangent equation $w_i = \boldsymbol{Z}_i^\top \boldsymbol{\beta} + \varepsilon_i$, the sensitivity analysis replaces it by the sandwich $n\hat{A}^{-1}\hat{\Omega}\hat{A}^{-1}$, where $\hat{\Omega} = \sum_i \boldsymbol{Z}_i r_i r_i^\top \boldsymbol{Z}_i^\top$. The heteroskedasticity-consistent estimator HC3 [MacKinnon and White, 1985] forms $\hat{\Omega}$ from the leverage-adjusted residuals $(I_d - H_{ii})^{-1} r_i$, with $H_{ii} = \boldsymbol{Z}_i^\top \hat{A}^{-1} \boldsymbol{Z}_i$. Either matrix is substituted into the same t_ν and χ^2_d statistics. Appendix H.5 checks their level under quadratic alternatives, and Appendix Table 13 reports the HC3 p-values for the case studies.

5 Simulation Studies

We evaluate ROLS on five manifolds, $\mathbb{S}^2$, $\mathrm{SO}(3)$, $\mathrm{Sym}^+(3)$, the flat torus $\mathbb{T}^2$, and the Poincaré ball $\mathbb{H}^3$, which include positive, zero, and negative curvature and trivial vs. non-trivial topology. Throughout, we use the canonical metrics, namely the round metric on $\mathbb{S}^2$, the flat metric on $\mathbb{T}^2$, the bi-invariant metric on $\mathrm{SO}(3)$, the affine-invariant metric on $\mathrm{Sym}^+(3)$, and the Poincaré metric on $\mathbb{H}^3$. Explicit formulas are summarized in Appendix F. Six experiments test, in turn, the cut-locus condition of Theorem 3.4, the parameter-error rate of Corollary 4.3, the calibration of the asymptotic tests of Theorem 4.5, their power, predictive accuracy under varying noise and sample size, and robustness to tangent-space misspecification. Data follow Model (1) with α taken from a fixed alternating-sign pool with leading entries $(0.5, -0.3, 0.2, -0.15, \ldots)$, $\|v_k\| = 0.2$ in random tangent directions at $\bar{Y}$, $z_{ik} \overset{\text{iid}}{\sim} \mathcal{N}(0, 1)$, and isotropic tangent noise of scale τ. The default configuration is n=200, p=2 manifold-valued and q=3 scalar predictors, τ=0.15. Prediction error is reported as Fréchet RMSE. See Appendix H for full data-generating process, including the selected baselines and hyperparameters.

Cut-locus condition on $\mathbb{S}^2$. We sweep the data spread r_0 over $[0.1, 3.0]$ (50 trials per spread, $n = 300$). Test RMSE is stable at ≈ 0.11 for $r_0 \leq 0.8$ and rises sharply once r_0 approaches the Karcher convexity radius $\pi/2$, reaching 1.10 at r_0 near the cut locus π (Fig. 2(a); detailed sweep in Appendix Fig. 5). As r_0 approaches $\pi/2$, both the empirical squared-RMSE ratio and $\eta(\kappa_+, r_0)$ from (7) fall toward zero, reflecting the tangent-linearization bias b^2 of Theorem 4.2.

Parameter convergence. We regress $\log \|\hat{\beta}-\beta^*\|$ on $\log n$ over $n \in \{50, 100, 200, 500, 1000, 2000\}$ (100 trials each). The slopes for $\hat{\alpha}$ and $\hat{v}$ on all five manifolds lie in $[-0.54, -0.49]$ with $R^2 \geq 0.98$, consistent with the $-1/2$ rate of Corollary 4.3 (Fig. 2(b); per-manifold slopes in Appendix Table 6). On $\mathbb{H}^3$ the fit uses the Lorentz model with a converged Fréchet mean (Appendix H.1).

Calibration of the asymptotic tests. Across $n \in \{50, 100, 200, 500\}$ and both null hypotheses, all 40 Type I rejection rates lie in $[0.026, 0.068]$ at the nominal 0.05 level (500 trials per cell; Fig. 2(c) summarizes power and Type I error, and Appendix Fig. 8 gives the full sample-size grid). The pairs-bootstrap test stays at or below 0.063 in all 20 settings of Appendix Table 7.

Power. The t-test for $H_0 : \alpha = 0$ saturates at $\geq 96\%$ rejection by $|\alpha| = 0.06$ on $\mathbb{T}^2$, SO(3), and $\mathrm{Sym}^+(3)$, by $|\alpha| = 0.10$ on $\mathbb{S}^2$, and requires $|\alpha| \approx 0.25$ on $\mathbb{H}^3$. The gap on $\mathbb{H}^3$ comes from the Poincaré conformal factor at the response base point, which inflates the effective tangent-space noise by roughly $2\times$ relative to the predictor variation. The Wald test for $H_0 : v = 0$ saturates at $\geq 99\%$ by $\|v\| = 0.06$ on all five manifolds because the model is linear in v and the coefficient is unaffected by curvature.

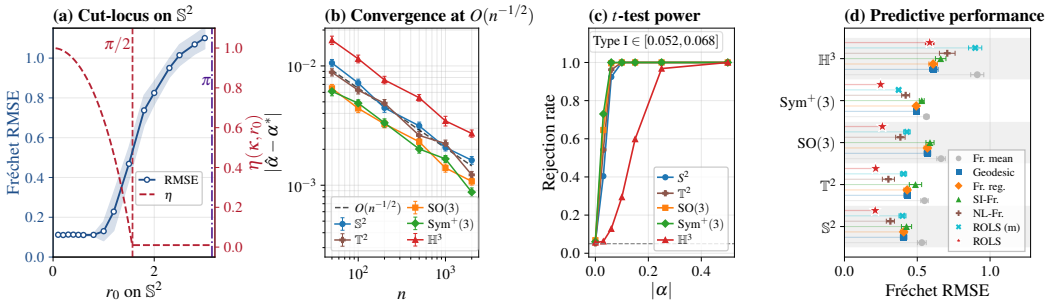


Figure 2: Simulation summary. (a) Cut-locus on $\mathbb{S}^2$: test RMSE vs. data spread r_0; the dashed curve is $\eta(\kappa_+, r_0)$ from (7). (b) Parameter convergence: $|\hat{\alpha} - \alpha^*|$ vs. n on five manifolds ($\mathbb{H}^3$ in the Lorentz model); dashed line is the $O(n^{-1/2})$ rate of Corollary 4.3. (c) t-test power for $H_0 : \alpha = 0$; dashed line is the nominal $\alpha = 0.05$; inset shows the Type I range at $|\alpha| = 0$, $n = 200$. (d) RMSE at the default cell (n, p, q, τ)=$(200, 2, 3, 0.15)$ for six baselines on five manifolds; ROLS shown as a red star. Full sweeps and per-cell numbers in Appendix H.

Predictive accuracy. We benchmark against six baselines (Fréchet mean, geodesic, Fréchet regression, single-index Fréchet (SI-Fr.), NL-Fr., and manifold-only ROLS (m)) across configurations varying τ, q, n, and p on each manifold, with 100 trials each (baselines and full grid in Appendix H). ROLS has the lowest mean RMSE in 44 of the 45 configurations, with a median reduction of 28% over the strongest baseline, which is NL-Fréchet in 27 configurations (Fig. 2(d); sweeps in Appendix Fig. 4, full table in Appendix Table 5). The exception is $\mathbb{H}^3$ at $\tau = 0.30$, where Fréchet regression is 0.9% lower. At $q = 0$, ROLS coincides with ROLS (m). The tangent linear-ridge ablation produces RMSE comparable to ROLS (median 1.2% higher, range -1.5% to 6.9% across the 45 cells), so the scalar α_j coefficients show no loss relative to a full $d \times d$ map.

Robustness under tangent-space misspecification. We add a quadratic distortion $\gamma \|u_i\|^2 \hat{n}$ to the $\mathbb{S}^2$ data-generating process. ROLS reaches RMSE 0.070 at $\gamma = 0$ (a 68% reduction over NL-Fréchet's 0.218) and retains a 31% reduction over NL-Fréchet at $\gamma = 0.6$ (0.300 vs. 0.438). In this design the omitted term contributes a squared bias of order γ^2, which is the term b^2 in Theorem 4.2. Under this misspecification the model-based test of $\alpha = 0$ over-rejects, up to 0.36 at $\gamma = 0.6$. The sandwich test stays at or below 0.084, and a targeted lack-of-fit diagnostic detects the departure (Appendix H.5). Base-point sensitivity, tangent dimension up to $d = 55$, and computational cost are reported in Appendices H.6–H.8.

6 Real-Data Case Studies

The real-data studies test whether the simulation findings extend to scientific prediction and inference tasks with mixed covariates and manifold-valued responses. We evaluate ROLS on three public

datasets on three manifolds. Diffusion tensor imaging of the corpus callosum (DTI-CC, $\mathrm{Sym}^+(3)$) data are extracted from the Autism Brain Imaging Data Exchange (ABIDE-II) [Di Martino et al., 2014], protein backbone dihedral angles ($\mathbb{T}^2$) from the Protein Data Bank (PDB) [Berman et al., 2000], and 850 hPa winds (S^2) from the National Centers for Environmental Prediction/National Center for Atmospheric Research (NCEP/NCAR) Reanalysis 1 [Kalnay et al., 1996].

We compare the same method classes used in simulation and report Fréchet RMSE under cross-validation (CV). The splitting scheme respects the sampling structure of each dataset, with subject-blocked folds for diffusion tensors, repeated folds for protein residues, and blocked temporal folds for wind. Dataset construction and preprocessing are reported in Appendix I. Table 2 summarizes the prediction results and Table 3 the primary coefficient-level inference.

Table 2: Real-data case studies: Fréchet RMSE (mean ± std across CV folds). **Bold**: lowest mean RMSE.

		Scalar-only baselines				*Manifold-aware methods*		
Dataset	**Manifold**	**Fr. mean**	**Geodesic**	**Fr. reg**	**SI-Fr.**	**NL-Fr.**	**ROLS (m)**	**ROLS**
DTI-CC	$\mathrm{Sym}^+(3)$	$1.138_{\pm.011}$	$1.129_{\pm.012}$	$1.130_{\pm.012}$	$1.128_{\pm.012}$	$1.134_{\pm.012}$	$\mathbf{.400}_{\pm.017}$	$\mathbf{.400}_{\pm.017}$
Protein backbone	$\mathbb{T}^2$	$1.779_{\pm.009}$	$1.215_{\pm.014}$	$1.357_{\pm.014}$	$1.220_{\pm.014}$	$1.494_{\pm.014}$	$1.395_{\pm.016}$	$\mathbf{1.177}_{\pm.016}$
Wind	S^2	$.546_{\pm.114}$	$.447_{\pm.040}$	$.462_{\pm.071}$	$.456_{\pm.041}$	$.414_{\pm.078}$	$.098_{\pm.009}$	$\mathbf{.095}_{\pm.005}$

ROLS improves on the strongest scalar-only baseline and the manifold-aware NL-Fréchet baseline across all three datasets (Table 2) and provides individual coefficient tests that the nonparametric baselines do not. Per-dataset numbers are in Appendix I.

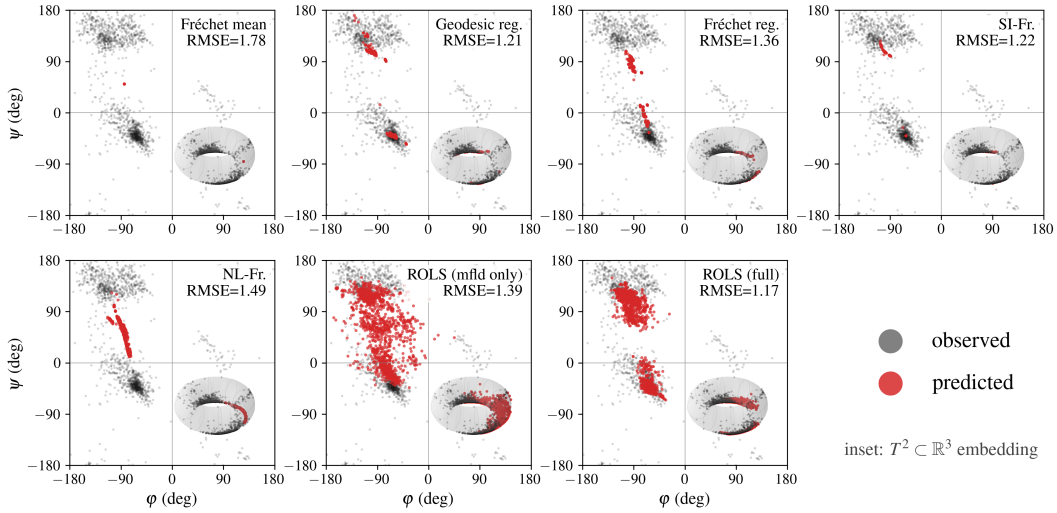


Figure 3: Ramachandran predictions for seven estimators on PDB. Each panel overlays predicted (red) and observed (grey) (φ, ψ) on the flat torus $[-180°, 180°]^2$, with $T^2 \subset \mathbb{R}^3$ in the inset.

Figure 3 shows how the estimators differ on the protein-backbone torus, where the Ramachandran plot is the canonical check. Observed (φ, ψ) pairs form two geometrically distant basins, the α-helix near $(-60°, -45°)$ and the β-sheet near $(-120°, +120°)$, so a useful predictor has to place each residue in the correct basin. Scalar-only baselines collapse to a point or a narrow cluster. NL-Fréchet places predictions across the $\varphi < 0$ band because the kernel averages residues from both basins, and manifold-only ROLS spreads predictions over the $\varphi < 0$ half-plane without separating the two basins. Full ROLS is the only estimator that recovers separate α- and β-clusters in both basins. This is consistent with complementary information in the neighboring dihedrals and the scalar descriptors.

Beyond prediction, the coefficient tests in Table 3 agree with effects reported in the literature. In DTI-CC, the along-tract tensor neighbors and the genu-versus-body contrast are significant, in agreement with the anterior-posterior microstructural gradient of the corpus callosum [Hofer and Frahm, 2006]. Autism spectrum disorder (ASD) diagnosis, age, sex, and full-scale intelligence quotient (FIQ) are

not significant once subject-level clustering is accounted for, consistent with later quality-controlled reanalyses that revised early autism-DTI claims downward [Koldewyn et al., 2014]. On the protein backbone, (φ, ψ) at offsets $-2, -1, +1$ dominate, recovering the local-context sequence-structure rule used in fragment-based folding [Simons et al., 1997]. The significance of the first Kidera principal component (PC1) agrees with the hydrophobicity-driven secondary-structure preference of Chou and Fasman [1974], as quantified by the leading factor of Kidera et al. [1985]. The designed null, Kidera PC3, remains non-significant. In the wind study, the three neighboring grid points dominate, consistent with synoptic transport, and sea-level pressure is not significant. Where the tangent-linear model does not fit exactly, α_j and v_k are the coefficients of the best-fitting tangent-linear approximation at the Fréchet mean. The lack-of-fit diagnostic in Appendix H.5 attributes about 1.6% of the residual variation on the protein data and 10% on DTI-CC to omitted quadratic terms. Every manifold-predictor coefficient has $p < 10^{-5}$ under both the model-based and the heteroskedasticity-robust Wald tests, and Benjamini–Hochberg adjustment changes no conclusion (Appendix I).

Table 3: Coefficient-level tests for three case studies.

DTI-CC			**Protein backbone**			**Wind**		
Predictor	Type	p	Predictor	Type	p	Predictor	Type	p
Tensor neighbor 1	Manifold	0.002	Offset -2	Manifold	$<10^{-5}$	Grid point 1	Manifold	$<10^{-5}$
Tensor neighbor 2	Manifold	0.002	Offset -1	Manifold	$<10^{-5}$	Grid point 2	Manifold	$<10^{-5}$
CC subregion	Scalar	0.002	Offset $+1$	Manifold	$<10^{-5}$	Grid point 3	Manifold	$<10^{-5}$
Age	Scalar	0.301	Kidera PC1	Scalar	$<10^{-5}$	Pressure	Scalar	0.734
Sex	Scalar	0.291	Kidera PC2	Scalar	$<10^{-5}$	Temperature	Scalar	$<10^{-5}$
ASD diagnosis	Scalar	0.549	Kidera PC3	Scalar	0.345	Season	Scalar	0.008
FIQ	Scalar	0.447	α-helix	Scalar	$<10^{-5}$			
			Solvent-accessibility	Scalar	2.8×10^{-4}			

p: subject-blocked permutation (DTI-CC), Wald test on the full-data fit (protein), Benjamini–Hochberg-adjusted Wald test (wind).

7 Conclusions

ROLS is a closed-form tangent-space analogue of Euclidean OLS for manifold-valued responses with mixed manifold and scalar predictors. Under explicit geometric regularity conditions, ROLS admits parametric-rate estimation and individual coefficient tests, and the cut-locus analysis identifies when the tangent-space design is well defined. Across simulations and three case studies, ROLS shows superior predictive performance and enables coefficient-level inference.

This study opens many future directions. First, the scalar coefficients can be extended to a matrix- or operator-valued linear map, so that a transported tangent vector can be rotated for directional estimation. Second, ROLS restricts manifold predictors to lie on the response manifold. Allowing predictors from different manifolds would make ROLS applicable in more complex case studies. For example, Choi and Park [2025] use a Hilbert–Schmidt operator to map predictors from other manifolds to the response manifold. Third, the exponential, logarithm, and parallel-transport maps can be replaced by a second-order retraction and a vector transport, which applies to manifolds where those maps are unavailable, including tensor responses of fixed multilinear rank [Kressner et al., 2014]. Fourth, strongly separated modes, or modes that cross the cut locus, call for cluster-specific centers or a mixture of local ROLS models. Fifth, the present theory fixes the dimension, and sparse designs in which d grows with n remain open. Finally, each new batch now requires a full refit. Recursive updates are available for Fréchet means [Ho et al., 2013, Cheng et al., 2016], and an online version of ROLS would need a separate analysis.

Acknowledgments and Disclosure of Funding

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

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

Miguel A. Arcones and Evarist Giné. On the bootstrap of M-estimators and other statistical functionals. In Raoul LePage and Lynne Billard, editors, *Exploring the Limits of Bootstrap*, pages 13–47. Wiley, 1992.

Marcel Berger. *A Panoramic View of Riemannian Geometry*. Springer, Berlin, 2003.

Christoph Bergmeir and José M. Benítez. On the use of cross-validation for time series predictor evaluation. *Information Sciences*, 191:192–213, 2012.

Helen M. Berman, John Westbrook, Zukang Feng, Gary Gilliland, T. N. Bhat, Helge Weissig, Ilya N. Shindyalov, and Philip E. Bourne. The Protein Data Bank. *Nucleic Acids Research*, 28(1):235–242, 2000.

Satarupa Bhattacharjee and Hans-Georg Müller. Single index Fréchet regression. *Annals of Statistics*, 51(4):1770–1798, 2023.

Satarupa Bhattacharjee, Bing Li, and Lingzhou Xue. Nonlinear global Fréchet regression for random objects via weak conditional expectation. *Annals of Statistics*, 53(1):117–143, 2025.

Rabi Bhattacharya and Lizhen Lin. Omnibus CLTs for Fréchet means and nonparametric inference on non-Euclidean spaces. *Proceedings of the American Mathematical Society*, 145(1):413–428, 2017.

Rabi Bhattacharya and Victor Patrangenaru. Large sample theory of intrinsic and extrinsic sample means on manifolds I. *Annals of Statistics*, 31(1):1–29, 2003.

Rabi Bhattacharya and Victor Patrangenaru. Large sample theory of intrinsic and extrinsic sample means on manifolds II. *Annals of Statistics*, 33(3):1225–1259, 2005.

Daniel Brooks, Olivier Schwander, Frédéric Barbaresco, Jean-Yves Schneider, and Matthieu Cord. Riemannian batch normalization for SPD neural networks. In *Advances in Neural Information Processing Systems*, volume 32, 2019.

Louis Capitaine, Jérémie Bigot, Rodolphe Thiébaut, and Robin Genuer. Fréchet random forests for metric space valued regression with non Euclidean predictors. *Journal of Machine Learning Research*, 25(355):1–41, 2024.

Jeff Cheeger and David G. Ebin. *Comparison Theorems in Riemannian Geometry*, volume 9 of *North-Holland Mathematical Library*. North-Holland, Amsterdam, 1975.

Xiaoyu Chen, Yujing Huang, and Xinwei Deng. Distribution-in-distribution-out regression. *arXiv preprint arXiv:2405.11626*, 2024.

Guang Cheng, Jeffrey Ho, Hesamoddin Salehian, and Baba C. Vemuri. Recursive computation of the Fréchet mean on non-positively curved Riemannian manifolds with applications. In *Riemannian Computing in Computer Vision*. Springer, 2016.

Victor Chernozhukov, Denis Chetverikov, Mert Demirer, Esther Duflo, Christian Hansen, Whitney Newey, and James Robins. Double/debiased machine learning for treatment and structural parameters. *The Econometrics Journal*, 21(1):C1–C68, 2018.

Changwon Choi and Byeong U. Park. High-dimensional Hilbert–Schmidt linear regression with Hilbert manifold variables. *The Annals of Statistics*, 53(6):2673–2701, 2025. doi: 10.1214/25-AOS2559.

Peter Y. Chou and Gerald D. Fasman. Prediction of protein conformation. *Biochemistry*, 13(2): 222–245, 1974.

Dario Cordero-Erausquin, Robert J. McCann, and Michael Schmuckenschläger. A Riemannian interpolation inequality à la Borell, Brascamp and Lieb. *Inventiones Mathematicae*, 146(2): 219–257, 2001.

Emil Cornea, Hongtu Zhu, Pauline Kim, and Joseph G. Ibrahim. Regression models on Riemannian symmetric spaces. *Journal of the Royal Statistical Society: Series B*, 79(2):463–482, 2017.

Adriana Di Martino, Chao-Gan Yan, Qingyang Li, Erin Denio, Francisco X. Castellanos, Kaat Alaerts, Jeffrey S. Anderson, Michal Assaf, Susan Y. Bookheimer, Mirella Dapretto, et al. The autism brain imaging data exchange: towards a large-scale evaluation of the intrinsic brain architecture in autism. *Molecular Psychiatry*, 19(6):659–667, 2014.

Manfredo Perdigão do Carmo. *Riemannian Geometry*. Mathematics: Theory & Applications. Birkhäuser, Boston, 1992. Translated by Francis Flaherty.

Benjamin Eltzner and Stephan F. Huckemann. A smeary central limit theorem for manifolds with application to high-dimensional spheres. *The Annals of Statistics*, 47(6):3360–3381, 2019.

P. Thomas Fletcher. Geodesic regression and the theory of least squares on Riemannian manifolds. *International Journal of Computer Vision*, 105(2):171–185, 2013.

David A. Freedman. Bootstrapping regression models. *The Annals of Statistics*, 9(6):1218–1228, 1981.

Octavian-Eugen Ganea, Gary Bécigneul, and Thomas Hofmann. Hyperbolic neural networks. In *Advances in Neural Information Processing Systems*, volume 31, 2018.

Kyunghee Han and Dogyoon Song. Errors-in-variables Fréchet regression with low-rank covariate approximation. In *Advances in Neural Information Processing Systems*, volume 36, 2023.

Jacob Hinkle, Prasanna Muralidharan, P. Thomas Fletcher, and Sarang Joshi. Polynomial regression on Riemannian manifolds. In *European Conference on Computer Vision*, pages 1–14. Springer, 2012.

Jeffrey Ho, Guang Cheng, Hesamoddin Salehian, and Baba C. Vemuri. Recursive Karcher expectation estimators and geometric law of large numbers. In *Proceedings of the Sixteenth International Conference on Artificial Intelligence and Statistics (AISTATS)*, volume 31 of *PMLR*, pages 325–332, 2013.

Sabine Hofer and Jens Frahm. Topography of the human corpus callosum revisited—comprehensive fiber tractography using diffusion tensor magnetic resonance imaging. *NeuroImage*, 32(3):989–994, 2006.

E. Kalnay, M. Kanamitsu, R. Kistler, W. Collins, D. Deaven, L. Gandin, M. Iredell, S. Saha, G. White, J. Woollen, Y. Zhu, A. Leetmaa, R. Reynolds, M. Chelliah, W. Ebisuzaki, W. Higgins, J. Janowiak, K. C. Mo, C. Ropelewski, J. Wang, R. Jenne, and D. Joseph. The NCEP/NCAR 40-year reanalysis project. *Bulletin of the American Meteorological Society*, 77(3):437–471, 1996.

Akinori Kidera, Yasuo Konishi, Mizuho Oka, Tatsuo Ooi, and Harold A. Scheraga. Statistical analysis of the physical properties of the 20 naturally occurring amino acids. *Journal of Protein Chemistry*, 4(1):23–55, 1985.

Hyunwoo J. Kim, Nagesh Adluru, Maxwell D. Collins, Moo K. Chung, Barbara B. Bendlin, Sterling C. Johnson, Richard J. Davidson, and Vikas Singh. Multivariate general linear models (MGLM) on Riemannian manifolds with applications to statistical analysis of diffusion weighted images. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 2705–2712. IEEE Computer Society, 2014.

Kami Koldewyn, Anastasia Yendiki, Sarah Weigelt, Hyowon Gweon, Joshua Julian, Hilary Richardson, Caitlin Malloy, Rebecca Saxe, Bruce Fischl, and Nancy Kanwisher. Differences in the right inferior longitudinal fasciculus but no general disruption of white matter tracts in children with autism spectrum disorder. *Proceedings of the National Academy of Sciences*, 111(5):1981–1986, 2014.

Daniel Kressner, Michael Steinlechner, and Bart Vandereycken. Low-rank tensor completion by Riemannian optimization. *BIT Numerical Mathematics*, 54(2):447–468, 2014.

Zhenhua Lin and Hans-Georg Müller. Total variation regularized Fréchet regression for metric-space valued data. *Annals of Statistics*, 49(6):3510–3533, 2021.

Zhenhua Lin, Hans-Georg Müller, and Byeong U. Park. Additive models for symmetric positive-definite matrices and Lie groups. *Biometrika*, 110(2):361–379, 2023.

James G. MacKinnon and Halbert White. Some heteroskedasticity-consistent covariance matrix estimators with improved finite sample properties. *Journal of Econometrics*, 29(3):305–325, 1985.

Anton Mallasto and Aasa Feragen. Wrapped Gaussian process regression on Riemannian manifolds. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 5580–5588, 2018.

Enno Mammen. Bootstrap and wild bootstrap for high dimensional linear models. *The Annals of Statistics*, 21(1):255–285, 1993.

Xavier Pennec, Pierre Fillard, and Nicholas Ayache. A Riemannian framework for tensor computing. *International Journal of Computer Vision*, 66(1):41–66, 2006.

Alexander Petersen and Hans-Georg Müller. Fréchet regression for random objects with Euclidean predictors. *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.

Xiaoyan Shi, Martin Styner, Jeffrey Lieberman, Joseph G. Ibrahim, Weili Lin, and Hongtu Zhu. Intrinsic regression models for manifold-valued data. In *Medical Image Computing and Computer-Assisted Intervention (MICCAI)*, volume 5762 of *Lecture Notes in Computer Science*, pages 192–199. Springer, 2009.

Kim T. Simons, Charles Kooperberg, Enoch Huang, and David Baker. Assembly of protein tertiary structures from fragments with similar local sequences using simulated annealing and Bayesian scoring functions. *Journal of Molecular Biology*, 268(1):209–225, 1997.

Aad W. van der Vaart. *Asymptotic Statistics*, volume 3 of *Cambridge Series in Statistical and Probabilistic Mathematics*. Cambridge University Press, 1998.

Aad W. van der Vaart and Jon A. Wellner. *Weak Convergence and Empirical Processes: With Applications to Statistics*. Springer Series in Statistics. Springer, New York, 1996.

Jon A. Wellner and Yihui Zhan. Bootstrapping Z-estimators. Technical Report 308, Department of Statistics, University of Washington, 1996.

Haoshu Xu and Hongzhe Li. Wasserstein F-tests for Fréchet regression on Bures–Wasserstein manifolds. *Journal of Machine Learning Research*, 2025.

Qi Zhang, Bing Li, and Lingzhou Xue. Dimension reduction for Fréchet regression. *Journal of the American Statistical Association*, 119(548), 2024.

Yaqing Zhou and Hans-Georg Müller. Network regression with graph Laplacians. *Journal of Machine Learning Research*, 23(320):1–41, 2022.

This appendix collects the supporting material in the following order. Appendix A records the Hessian and Rauch comparison factors used throughout the proofs. Appendices B–E contain the complete proofs of Theorems 3.3–4.5. Appendix E.1 proves the pairs bootstrap by embedding the Fréchet means and $\boldsymbol{\beta}$ in one Z-estimator. Appendix F derives the special cases summarized in Remark 3.2 and records the manifold-specific operations used in the experiments. Appendix G gives the fixed-base residual bootstrap as a within-tangent-space diagnostic. Appendix H provides additional simulation results, and Appendix I provides the full real-data case-study details. We use the same notation as the main text throughout. Standard references for Riemannian geometry are do Carmo [1992], Cheeger and Ebin [1975], and Berger [2003].

A Curvature Distortion Factors

The Hessian comparison theorem for $\frac{1}{2}d^2_{\mathcal{M}}(p,\cdot)$ [Cordero-Erausquin et al., 2001] yields the contraction factor

$$\eta(\kappa_+, r) = \begin{cases} \sqrt{\kappa_+}\, r \,/\, \tan(\sqrt{\kappa_+}\, r) & \kappa_+ > 0, \\ 1 & \kappa_+ \le 0, \end{cases} \tag{7}$$

defined for $r < \pi/(2\sqrt{\kappa_+})$ when $\kappa_+ > 0$ and for all $r \ge 0$ otherwise. This equals the minimum eigenvalue of $\mathrm{Hess}(\frac{1}{2}d^2_{\mathcal{M}}(p,\cdot))$ at distance r, with $\eta \in (0,1]$. The Rauch comparison theorem gives the expansion factor

$$\tilde{\eta}(\kappa_-, r) = \begin{cases} \sinh(\sqrt{|\kappa_-|}\, r) \,/\, (\sqrt{|\kappa_-|}\, r) & \kappa_- < 0, \\ 1 & \kappa_- \ge 0, \end{cases} \tag{8}$$

with $\tilde{\eta} \ge 1$.

B Proof of the Generalized Normal Equations (Theorem 3.3)

We restate the theorem for convenience.

Theorem 3.3. *Fix an orthonormal frame of $T_{\bar{Y}}\mathcal{M}$ and identify it with $\mathbb{R}^d$. On the mixed Hilbert design space $\mathcal{H} := \mathbb{R}^p \oplus (T_{\bar{Y}}\mathcal{M})^q \cong \mathbb{R}^{p+qd}$, the population FLS minimizer $\boldsymbol{\beta}^* = (\alpha_1^*, \dots, \alpha_p^*, v_1^*, \dots, v_q^*) \in \mathcal{H}$ satisfies the block normal equations $\Sigma_{\boldsymbol{ZZ}}\boldsymbol{\beta}^* = C_{\boldsymbol{ZY}}$, where $\Sigma_{\boldsymbol{ZZ}} \in \mathbb{R}^{(p+qd)\times(p+qd)}$ is the self-adjoint block Gram matrix with blocks $\big[\mathbb{E}\, g_{\bar{Y}}(u_j, u_\ell)\big]_{j,\ell}$, $\big[\mathbb{E}\, \tilde{z}_k\, u_j^\top\big]_{j,k}$, $\big[\mathbb{E}\, \tilde{z}_k\, u_j\big]_{k,j}$, $\big[\mathbb{E}\, \tilde{z}_k \tilde{z}_\ell\, I_d\big]_{k,\ell}$, and $C_{\boldsymbol{ZY}}$ has entries $\mathbb{E}\, g_{\bar{Y}}(u_j, w)$ and $\mathbb{E}\, \tilde{z}_k\, w$. When $\Sigma_{\boldsymbol{ZZ}} \succ 0$, the unique minimizer is $\boldsymbol{\beta}^* = \Sigma_{\boldsymbol{ZZ}}^{-1} C_{\boldsymbol{ZY}}$.*

Proof. Recall that $w = \mathrm{Log}_{\bar{Y}}(Y) \in T_{\bar{Y}}\mathcal{M}$ is the lifted response and $\hat{w} = \sum_{j=1}^p \alpha_j u_j + \sum_{k=1}^q \tilde{z}_k v_k$ is the model prediction in $T_{\bar{Y}}\mathcal{M}$. The manifold Fréchet risk is $\frac{1}{2}\mathbb{E}[d^2_{\mathcal{M}}(Y, \mathrm{Exp}_{\bar{Y}}(\hat{w}))]$. Under the injectivity condition (5), the logarithm map $\mathrm{Log}_{\bar{Y}}$ is a diffeomorphism on the geodesic ball $B(\bar{Y}, \mathrm{inj}_{\bar{Y}})$, and every observation Y_i and every prediction $\mathrm{Exp}_{\bar{Y}}(\hat{w}_i)$ lie in this ball. Because the ROLS model (1) postulates a tangent-space linear relationship, the natural working objective is the tangent-space quadratic

$$L(\boldsymbol{\alpha}, \boldsymbol{v}) := \frac{1}{2}\mathbb{E}\big[\|w - \hat{w}\|^2_{g_{\bar{Y}}}\big]. \tag{9}$$

On a flat manifold this equals $\frac{1}{2}\mathbb{E}[d^2_{\mathcal{M}}(Y, \mathrm{Exp}_{\bar{Y}}(\hat{w}))]$ exactly. On a curved manifold the two objectives differ by a factor controlled by the Rauch comparison theorem (see Theorem 4.2 and its proof in Appendix D). The manifold prediction error is bounded above by $\tilde{\eta}(\kappa_-, r_0)^2$ times $2L$. Since L is a quadratic functional in $(\boldsymbol{\alpha}, \boldsymbol{v})$, the normal equations derived below are exact for L.

Since $T_{\bar{Y}}\mathcal{M}$ equipped with $g_{\bar{Y}}$ is a finite-dimensional real Hilbert space, the objective L is a quadratic functional in the parameters $(\boldsymbol{\alpha}, \boldsymbol{v})$. We compute its derivatives by differentiating under the expectation sign, which is justified by the square-integrability assumption $\mathbb{E}[\|w\|^2] < \infty$.

We begin by differentiating L with respect to the scalar coefficients. For any fixed $j^* \in \{1, \dots, p\}$,

$$\begin{aligned}\frac{\partial L}{\partial \alpha_{j^*}} &= \frac{\partial}{\partial \alpha_{j^*}} \frac{1}{2} \mathbb{E}\big[\|w - \hat{w}\|^2_{g_{\bar{Y}}}\big] \\ &= \mathbb{E}\big[g_{\bar{Y}}\big(w - \hat{w},\ -\tfrac{\partial \hat{w}}{\partial \alpha_{j^*}}\big)\big] \\ &= -\mathbb{E}\big[g_{\bar{Y}}\big(w - \hat{w},\ u_{j^*}\big)\big], \end{aligned} \tag{10}$$

since $\frac{\partial \hat{w}}{\partial \alpha_{j^*}} = u_{j^*}$. Setting this to zero and expanding $\hat{w}$:

$$\begin{aligned} 0 &= \mathbb{E}\big[g_{\bar{Y}}\big(w - \hat{w},\ u_{j^*}\big)\big] \\ &= \mathbb{E}\big[g_{\bar{Y}}(w, u_{j^*})\big] - \mathbb{E}\left[g_{\bar{Y}}\left(\sum_{j=1}^{p} \alpha_j u_j + \sum_{k=1}^{q} \tilde{z}_k v_k,\ u_{j^*} \right) \right]. \end{aligned} \tag{11}$$

Using bi-linearity of $g_{\bar{Y}}(\cdot, \cdot)$ and linearity of expectation:

$$\mathbb{E}\big[g_{\bar{Y}}(w, u_{j^*})\big] = \sum_{j=1}^{p} \alpha_j \mathbb{E}\big[g_{\bar{Y}}(u_j, u_{j^*})\big] + \sum_{k=1}^{q} \mathbb{E}\big[\tilde{z}_k\, g_{\bar{Y}}(v_k, u_{j^*})\big]. \tag{12}$$

The coefficient $v_k \in T_{\bar{Y}}\mathcal{M}$ does not depend on the observation, so $\mathbb{E}[\tilde{z}_k\, g_{\bar{Y}}(v_k, u_{j^*})] = g_{\bar{Y}}(v_k, \mathbb{E}[\tilde{z}_k\, u_{j^*}]) = g_{\bar{Y}}(v_k, \operatorname{cov}(z_k, X_{j^*}))$.

Identifying $\operatorname{cov}(X_j, X_{j^*}) = \mathbb{E}[g_{\bar{Y}}(u_j, u_{j^*})]$, $\operatorname{cov}(X_{j^*}, Y) = \mathbb{E}[g_{\bar{Y}}(u_{j^*}, w)]$, and $\operatorname{cov}(z_k, X_{j^*}) = \mathbb{E}[\tilde{z}_k\, u_{j^*}] \in T_{\bar{Y}}\mathcal{M}$, equation (12) reads as the j^*-th row of the block normal equation

$$\sum_{j=1}^{p} \operatorname{cov}(X_j, X_{j^*})\, \alpha_j + \sum_{k=1}^{q} g_{\bar{Y}}\big(v_k,\ \operatorname{cov}(z_k, X_{j^*})\big) = \operatorname{cov}(X_{j^*}, Y). \tag{13}$$

We next compute the functional derivative with respect to the tangent-vector parameters. For any fixed $k^* \in \{1, \dots, q\}$ and any direction $h \in T_{\bar{Y}}\mathcal{M}$, the directional (Gateaux) derivative of L at $\boldsymbol{v}$ in direction $e_{k^*} \otimes h$ (i.e., perturbing v_{k^*} by h) is

$$\frac{d}{d\varepsilon}\bigg|_{\varepsilon=0} L(\boldsymbol{\alpha},\ v_{k^*} + \varepsilon h,\ \{v_k\}_{k \neq k^*}) = -\mathbb{E}\big[g_{\bar{Y}}\big(w - \hat{w},\ \tilde{z}_{k^*}\, h\big)\big] = -\mathbb{E}\big[\tilde{z}_{k^*}\, g_{\bar{Y}}(w - \hat{w},\ h)\big]. \tag{14}$$

Setting this to zero for all $h \in T_{\bar{Y}}\mathcal{M}$, by the non-degeneracy of $g_{\bar{Y}}$,

$$\begin{aligned} 0 &= \mathbb{E}\big[\tilde{z}_{k^*}(w - \hat{w})\big] \\ &= \mathbb{E}[\tilde{z}_{k^*} w] - \sum_{j=1}^{p} \alpha_j \mathbb{E}[\tilde{z}_{k^*}\, u_j] - \sum_{k=1}^{q} \mathbb{E}[\tilde{z}_{k^*}\, \tilde{z}_k]\, v_k, \end{aligned} \tag{15}$$

where in the last line we used bi-linearity of $g_{\bar{Y}}$ and the fact that v_k is a deterministic parameter, so $\mathbb{E}[\tilde{z}_{k^*} \tilde{z}_k\, v_k] = \mathbb{E}[\tilde{z}_{k^*} \tilde{z}_k]\, v_k$. Identifying $\operatorname{cov}(z_{k^*}, Y) = \mathbb{E}[\tilde{z}_{k^*}\, w] \in T_{\bar{Y}}\mathcal{M}$, $\operatorname{cov}(z_{k^*}, X_j) = \mathbb{E}[\tilde{z}_{k^*}\, u_j] \in T_{\bar{Y}}\mathcal{M}$, and $\operatorname{cov}(z_{k^*}, z_k) = \mathbb{E}[\tilde{z}_{k^*} \tilde{z}_k] \in \mathbb{R}$, equation (15) is the k^*-th row of the block normal equation

$$\sum_{j=1}^{p} \alpha_j \operatorname{cov}(z_{k^*}, X_j) + \sum_{k=1}^{q} \operatorname{cov}(z_{k^*}, z_k)\, v_k = \operatorname{cov}(z_{k^*}, Y). \tag{16}$$

Collecting equations (13) for $j^* = 1, \dots, p$ and (16) for $k^* = 1, \dots, q$ into matrix form gives $\Sigma_{\boldsymbol{ZZ}}\, \boldsymbol{\beta} = C_{\boldsymbol{ZY}}$, with the block Gram matrix

$$\Sigma_{\boldsymbol{ZZ}} = \begin{pmatrix} \big[\operatorname{cov}(X_j, X_\ell)\big]_{j,\ell=1}^{p} & \big[\operatorname{cov}(z_k, X_j)^\top\big]_{j,k} \\ \big[\operatorname{cov}(z_k, X_j)\big]_{k,j} & \big[\operatorname{cov}(z_k, z_\ell)\, I_d\big]_{k,\ell=1}^{q} \end{pmatrix} \tag{17}$$

and right-hand side $C_{\boldsymbol{ZY}}$ with components $\operatorname{cov}(X_j, Y)$ $(j = 1, \dots, p)$ and $\operatorname{cov}(z_k, Y)$ $(k = 1, \dots, q)$. The scalar block is repeated on each of the d tangent coordinates. The matrix has order $p + qd$ as in Theorem 3.3.

It remains to verify positive definiteness. The matrix $\Sigma_{\boldsymbol{ZZ}}$ is a block Gram matrix of second moments of random elements in $T_{\bar{Y}}\mathcal{M} \cong \mathbb{R}^d$ and $\mathbb{R}$. Hence $\Sigma_{\boldsymbol{ZZ}}$ is symmetric positive semi-definite by construction. Under the positive-definiteness hypothesis ($\Sigma_{\boldsymbol{ZZ}} \succ 0$), the block normal equation has the unique solution $\boldsymbol{\beta}^* = \Sigma_{\boldsymbol{ZZ}}^{-1} C_{\boldsymbol{Z}Y}$.

Finally, since $L(\boldsymbol{\alpha}, \boldsymbol{v})$ is a quadratic (convex) functional in the finite-dimensional parameter $\boldsymbol{\beta} = (\boldsymbol{\alpha}, \boldsymbol{v}) \in \mathbb{R}^p \times (T_{\bar{Y}}\mathcal{M})^q \cong \mathbb{R}^{p+qd}$, the critical point found above is a global minimum. When $\Sigma_{\boldsymbol{ZZ}} \succ 0$, the Hessian of L (viewed as a $(p+qd) \times (p+qd)$ matrix acting on the parameter space $\mathbb{R}^{p+qd}$) equals $\Sigma_{\boldsymbol{ZZ}}$, which is positive definite, confirming the unique minimizer. □

C Proof of Cut-Locus Identifiability (Theorem 3.4)

We restate the theorem for convenience.

Theorem 3.4. *Assume $\Sigma_{\boldsymbol{ZZ}} \succ 0$.* (a) Safe case. *If for each j the support of X_j sits inside a strongly convex normal ball $B(\bar{X}_j, r_j)$ with $r_j < \mathrm{inj}_{\bar{X}_j}$ (preferably $r_j < \mathrm{conv}_{\bar{X}_j}$) and $\mathrm{supp}(Y) \subset B(\bar{Y}, r_Y)$ with $r_Y < \mathrm{inj}_{\bar{Y}}$, where inj_p and conv_p denote the pointwise injectivity and convexity radii at p, then each $\mathrm{Log}_{\bar{X}_j}$ is single-valued and smooth on $\mathrm{supp}(X_j)$, L is a strictly convex quadratic on $\mathbb{R}^{p+qd}$, and the population minimizer $\boldsymbol{\beta}^*$ is unique.* (b) Conditional branch sensitivity. *Suppose on a measurable set S with $\mathbb{P}(S) > 0$ there exist two admissible measurable log-branches $\ell, \ell' \colon S \to T_{\bar{X}_j}\mathcal{M}$ with $\mathrm{Exp}_{\bar{X}_j}(\ell(x)) = \mathrm{Exp}_{\bar{X}_j}(\ell'(x)) = x$ and $\|\ell(x)\| = \|\ell'(x)\| = d_{\mathcal{M}}(\bar{X}_j, x)$, and set $\Delta Z_{ij} = \Gamma_{\bar{X}_j \to \bar{Y}}\{\ell(X_{ij}) - \ell'(X_{ij})\}\mathbf{1}\{X_{ij} \in S\}$. Let $P_j^{\perp}$ denote the orthogonal projection in $L^2(T_{\bar{Y}}\mathcal{M})$ onto the complement of the closed linear span of the remaining design $\{u_l\}_{l \neq j} \cup \{\tilde{z}_k e_a\}_{k,a}$. Then $\boldsymbol{\beta}^*_{(\ell)} - \boldsymbol{\beta}^*_{(\ell')} = (\Sigma^{(\ell')})^{-1}\big(\Delta C - \Delta\Sigma\, \boldsymbol{\beta}^*_{(\ell)}\big)$.*

Proof. **Part (a): Safe case via the inverse function theorem.**

Let $p \in \mathcal{M}$ and let inj_p denote the pointwise injectivity radius at p. By a standard result in Riemannian geometry [do Carmo, 1992, Chapter 3, Proposition 3.6], $\mathrm{Exp}_p \colon B(0, \mathrm{inj}_p) \subset T_p\mathcal{M} \to B(p, \mathrm{inj}_p) \subset \mathcal{M}$ is a diffeomorphism onto the open geodesic ball of radius inj_p. Equivalently, the differential $d\mathrm{Exp}_p(v)$ is nonsingular for all $\|v\| < \mathrm{inj}_p$ and no two points in $B(0, \mathrm{inj}_p)$ are sent to the same image, so by the inverse function theorem the logarithm $\mathrm{Log}_p \colon B(p, \mathrm{inj}_p) \to T_p\mathcal{M}$ is a smooth diffeomorphism. When additionally $r < \mathrm{conv}_p$ (the convexity radius), the ball $B(p, r)$ is strongly convex, and any two of its points are joined by a unique minimizing geodesic lying entirely in $B(p, r)$. The radii here are the pointwise injectivity and convexity radii at the base point.

Under the safe-case hypothesis, $\mathrm{supp}(X_j) \subset B(\bar{X}_j, r_j)$ with $r_j < \mathrm{inj}_{\bar{X}_j}$, and $\mathrm{supp}(Y) \subset B(\bar{Y}, r_Y)$ with $r_Y < \mathrm{inj}_{\bar{Y}}$. Hence $\mathrm{Log}_{\bar{X}_j}(X_{ij})$ and $\mathrm{Log}_{\bar{Y}}(Y_i)$ are each uniquely defined and vary smoothly with the observations.

Moreover, parallel transport $\Gamma_{\bar{X}_j \to \bar{Y}} \colon T_{\bar{X}_j}\mathcal{M} \to T_{\bar{Y}}\mathcal{M}$ along the (unique) minimal geodesic is a linear isometry [do Carmo, 1992, Chapter 2], hence injective. Therefore $X_{ij} \mapsto u_{ij} = \Gamma_{\bar{X}_j \to \bar{Y}}\big(\mathrm{Log}_{\bar{X}_j}(X_{ij})\big)$ is the composition of a smooth diffeomorphism with a linear isometry, a smooth injection on $\mathrm{supp}(X_j)$.

We now confirm that $\Sigma_{\boldsymbol{ZZ}} \succ 0$ yields an ordinary strictly convex quadratic objective. If $\Sigma_{\boldsymbol{ZZ}}\boldsymbol{c} = 0$ for some nonzero $\boldsymbol{c} = (a_1, \ldots, a_p, b_1, \ldots, b_q) \in \mathbb{R}^{p+qd}$, then

$$0 = \boldsymbol{c}^{\top}\Sigma_{\boldsymbol{ZZ}}\boldsymbol{c} = \mathbb{E}\left[\left\|\sum_{j=1}^{p} a_j u_j + \sum_{k=1}^{q} \tilde{z}_k\, b_k\right\|_{g_{\bar{Y}}}^2\right], \tag{18}$$

which forces an a.s. linear dependence, contradicting $\Sigma_{\boldsymbol{ZZ}} \succ 0$. Hence $L(\boldsymbol{\beta}) = \frac{1}{2}\boldsymbol{\beta}^{\top}\Sigma_{\boldsymbol{ZZ}}\boldsymbol{\beta} - C_{\boldsymbol{Z}Y}^{\top}\boldsymbol{\beta} + \mathrm{const}$ is a strictly convex quadratic and the population minimizer $\boldsymbol{\beta}^* = \Sigma_{\boldsymbol{ZZ}}^{-1} C_{\boldsymbol{Z}Y}$ is unique.

Part (b): Conditional branch sensitivity.

Assume the hypotheses of Part (b): $\mathrm{Log}_{\bar{X}_j}$ admits two distinct measurable branches $\ell, \ell' \colon S \to T_{\bar{X}_j}\mathcal{M}$ on a set S with $\mathbb{P}(S) > 0$, each returning minimal geodesics, with ΔZ_j as defined in Theorem 3.4. Off S the two branches agree, so the induced tangent-space design differs in the j-th block only. Write $\boldsymbol{Z}^{(\ell)}$ and $\boldsymbol{Z}^{(\ell')}$ for the two design matrices in $\mathbb{R}^{(p+qd)\times d}$ obtained by using branch ℓ or ℓ' in the j-th slot, and $W = \mathrm{Log}_{\bar{Y}}(Y)$ for the response tangent vector. The induced Gram and cross-covariance matrices are $\Sigma^{(\ell)} = \mathbb{E}[\boldsymbol{Z}^{(\ell)}(\boldsymbol{Z}^{(\ell)})^\top]$ and $C^{(\ell)} = \mathbb{E}[\boldsymbol{Z}^{(\ell)} W]$, with analogous definitions for ℓ', and population minimizers $\boldsymbol{\beta}^*_{(\ell)} = (\Sigma^{(\ell)})^{-1}C^{(\ell)}$ and $\boldsymbol{\beta}^*_{(\ell')} = (\Sigma^{(\ell')})^{-1}C^{(\ell')}$. Let $\Delta\Sigma = \Sigma^{(\ell)} - \Sigma^{(\ell')}$ and $\Delta C = C^{(\ell)} - C^{(\ell')}$, both supported in the j-th row/column through ΔZ_j.

Subtracting the two normal equations $\Sigma^{(\ell)}\boldsymbol{\beta}^*_{(\ell)} = C^{(\ell)}$ and $\Sigma^{(\ell')}\boldsymbol{\beta}^*_{(\ell')} = C^{(\ell')}$ yields the exact identity

$$\boldsymbol{\beta}^*_{(\ell)} - \boldsymbol{\beta}^*_{(\ell')} = (\Sigma^{(\ell')})^{-1}\big(\Delta C - \Delta\Sigma\,\boldsymbol{\beta}^*_{(\ell)}\big). \tag{19}$$

Abbreviating $\Sigma = \Sigma^{(\ell')}$ and $\boldsymbol{\beta}^* = \boldsymbol{\beta}^*_{(\ell)}$, the j-th coordinate of (19) is

$$\alpha^*_{j,(\ell)} - \alpha^*_{j,(\ell')} = \boldsymbol{e}_j^\top\,\Sigma^{-1}\big(\Delta C - \Delta\Sigma\,\boldsymbol{\beta}^*\big), \tag{20}$$

where $\boldsymbol{e}_j$ selects the j-th manifold-predictor slot of the concatenated design.

The non-degeneracy condition (ND), $P_j^\perp \Delta Z_j \neq 0$ in L^2, ensures that the branch perturbation ΔZ_j is not absorbed by the remaining design at the L^2 level, so $\Delta\Sigma$ contributes a non-trivial perturbation to the Gram matrix. Conversely, when (ND) fails so that ΔZ_j lies in the L^2-span of the remaining design, the perturbation block of $\Delta\Sigma$ is rank-deficient and the two branch conventions become observationally equivalent.

Applying Cauchy–Schwarz in $\mathbb{R}^{p+qd}$ directly to (20) yields the parameter-space upper bound

$$\big|\alpha^*_{j,(\ell)} - \alpha^*_{j,(\ell')}\big| \;\le\; \big\|(\Sigma^{(\ell')})^{-1}\boldsymbol{e}_j\big\|_2\,\big\|\Delta C - \Delta\Sigma\,\boldsymbol{\beta}^*_{(\ell)}\big\|_2, \tag{21}$$

where $\|(\Sigma^{(\ell')})^{-1}\boldsymbol{e}_j\|_2$ is the j-th column norm of $(\Sigma^{(\ell')})^{-1}$ and quantifies the sensitivity of the j-th coefficient to perturbations of the right-hand side. The vector $\Delta C - \Delta\Sigma\,\boldsymbol{\beta}^*_{(\ell)}$ is zero when the response is uncorrelated with the branch perturbation in the Σ-weighted sense. The two branches then share the same population minimizer even though (ND) holds. A gap that changes the predictor requires both (ND) and $\Delta C - \Delta\Sigma\,\boldsymbol{\beta}^*_{(\ell)} \neq 0$. □

Remark C.1 (On the role of non-degeneracy)**.** *Without condition* (ND) *the branch perturbation ΔZ_j lies in the L^2-span of the remaining design, $P_j^\perp \Delta Z_j$ vanishes, and the two branch conventions are observationally equivalent at the population level, with the branch swap on S absorbable into a redefinition of the remaining coefficients. The two branches then induce the same population predictor, and the j-th coefficient is no longer separately identified.*

Remark C.2 (Relation to global injectivity)**.** *Two log branches at $\bar{X}_j$ appear at distance $\mathrm{inj}_{\bar{X}_j}$. This radius can exceed the global infimum $\mathrm{inj}(\mathcal{M}) = \inf_p \mathrm{inj}_p$, so a point at distance $\mathrm{inj}(\mathcal{M})$ from $\bar{X}_j$ can still lie inside the injectivity ball of $\bar{X}_j$. Part (a) therefore uses the pointwise radius.*

D Proof of Consistency and Convergence Rate (Theorem 4.2)

We restate the theorem for convenience.

Theorem 4.2. *Under Assumption 4.1 (Afsari condition, injectivity condition, $\Sigma_{\boldsymbol{ZZ}} \succ 0$ with smallest eigenvalue $\lambda_* > 0$, and finite residual second moment) and finite fourth moments, let $\boldsymbol{\beta}^*$ denote the population FLS minimizer and $b^2 := \mathbb{E}[\big\|w - \sum_j \alpha_j^* u_j - \sum_k \tilde{z}_k v_k^*\big\|^2] - \sigma^2$ the squared model misspecification bias. Then the empirical FLS estimator $\hat{\boldsymbol{\beta}}_n$ satisfies*

$$\mathbb{E}\big[d^2_{\mathcal{M}}(\hat{Y}, Y)\big] \le \tilde{\eta}(\kappa_-, r_0)^2 \left(\frac{C\,\sigma^2(p+qd)}{n\lambda_*} + \sigma^2 + b^2\right), \tag{22}$$

where $r_0 = \sup_i d_{\mathcal{M}}(Y_i, \bar{Y})$, $\kappa_- \le \mathrm{Sec}(\mathcal{M})$ *is the sectional curvature lower bound,* $C > 0$ *depends on the fourth moments of the covariates, and*

$$\tilde{\eta}(\kappa_-, r) = \begin{cases} \dfrac{\sinh(\sqrt{|\kappa_-|}\, r)}{\sqrt{|\kappa_-|}\, r} & \kappa_- < 0, \\ 1 & \kappa_- \ge 0. \end{cases} \tag{23}$$

When the model is correctly specified ($b = 0$*),* $\hat{\boldsymbol{\beta}}_n \xrightarrow{P} \boldsymbol{\beta}^*$ *and* $\mathbb{E}[d^2_{\mathcal{M}}(\hat{Y}, Y)] \le \tilde{\eta}^2\, \sigma^2 + O(n^{-1})$.

Proof. Under the injectivity condition, the response Y_i lies in $B(\bar{Y}, \mathrm{inj}_{\bar{Y}})$ and $\hat{Y} = \mathrm{Exp}_{\bar{Y}}(\hat{w})$ lies in the same ball by Assumption 4.1(ii).

We require two-sided comparison geometry. Let $\kappa_- \le \mathrm{Sec}(\mathcal{M}) \le \kappa_+$ denote two-sided sectional curvature bounds. For points $p, q, q' \in \mathcal{M}$ with $r = \max\{d_{\mathcal{M}}(p, q), d_{\mathcal{M}}(p, q')\}$ satisfying $r < \min(\mathrm{inj}_p, \pi/(2\sqrt{\kappa_+}))$ (with the convention $\pi/(2\sqrt{\kappa_+}) = \infty$ when $\kappa_+ \le 0$), we use two comparison bounds.

For the **lower bound**, the Hessian comparison theorem for $\frac{1}{2} d^2_{\mathcal{M}}(p, \cdot)$ (see Cordero-Erausquin et al. [2001]; do Carmo [1992, Chapter 8]) gives

$$d^2_{\mathcal{M}}(q, q') \ge \eta(\kappa_+, r)\, \left\| \mathrm{Log}_p(q) - \mathrm{Log}_p(q') \right\|^2_{g_p}, \tag{24}$$

where $\eta(\kappa_+, r) \in (0, 1]$ is defined in (7). The factor $\eta = \sqrt{\kappa_+}\, r / \tan(\sqrt{\kappa_+}\, r)$ is the minimum eigenvalue of $\mathrm{Hess}(\frac{1}{2} d^2_{\mathcal{M}}(p, \cdot))$ at distance r from p; positive curvature contracts the exponential map, so the tangent-space norm overestimates the geodesic distance by this factor. For the **upper bound**, the Rauch comparison theorem (Jacobi field divergence under negative curvature; see Cheeger and Ebin [1975, Theorem 1.28]) gives

$$d^2_{\mathcal{M}}(q, q') \le \tilde{\eta}(\kappa_-, r)^2\, \left\| \mathrm{Log}_p(q) - \mathrm{Log}_p(q') \right\|^2_{g_p}, \tag{25}$$

where $\tilde{\eta}(\kappa_-, r) \ge 1$ is defined in (8). If $\kappa_+ \le 0$, then $\eta = 1$ and $d_{\mathcal{M}} \ge \left\| \mathrm{Log}_p(q) - \mathrm{Log}_p(q') \right\|$. If $\kappa_- \ge 0$, as on S^2 and $\mathrm{SO}(3)$, then $\tilde{\eta} = 1$ and (25) reads $d^2_{\mathcal{M}} \le \left\| \mathrm{Log}_p(q) - \mathrm{Log}_p(q') \right\|^2$. If $\kappa_- < 0$, as on $\mathrm{Sym}^+(k)$ and $\mathbb{H}^d$, then $\tilde{\eta} > 1$ and (25) exceeds the tangent norm.

Let (X, z, Y) be a new observation, independent of the sample. Assumption 4.1(ii) bounds both $d_{\mathcal{M}}(Y, \bar{Y})$ and $d_{\mathcal{M}}(\hat{Y}, \bar{Y})$ by r_0. Write $\mathrm{Log}_{\bar{Y}}(\hat{Y}) = \hat{w} + \delta$, where $\hat{w} = \boldsymbol{Z}(\theta^*)^\top \hat{\boldsymbol{\beta}}_n$ is the tangent prediction at the population base points. Applying (25) with $p = \bar{Y}$, $q = Y$, and $q' = \hat{Y}$:

$$\mathbb{E}[d^2_{\mathcal{M}}(\hat{Y}, Y)] \le \tilde{\eta}(\kappa_-, r_0)^2\, \mathbb{E}\big[\|w - \hat{w} - \delta\|^2_{g_{\bar{Y}}}\big]. \tag{26}$$

Remark D.1. *The lower bound* (24) *uses* κ_+ *and bounds* $d^2_{\mathcal{M}}$ *from below; the upper bound* (25) *uses* κ_- *and bounds* $d^2_{\mathcal{M}}$ *from above. These are different directions of the comparison and require different curvature bounds. On non-negatively curved manifolds (*$\kappa_- \ge 0$*),* $\tilde{\eta} = 1$*, so the tangent-space residual bounds the geodesic residual from above. On non-positively curved manifolds with* $\kappa_- < 0$*,* $\tilde{\eta} > 1$*, and the geodesic residual can exceed the tangent-space residual.*

We now work entirely in the Hilbert space $(T_{\bar{Y}}\mathcal{M}, g_{\bar{Y}})$. Define $\hat{w}^*_i = \sum_j \alpha^*_j u_{ij} + \sum_k \tilde{z}_{ik} v^*_k$ as the population best linear predictor (the projection of w onto the linear span of the covariates in $T_{\bar{Y}}\mathcal{M}$), and define $\varepsilon_i = w_i - \hat{w}^*_i$ as the residual. By the definition of $\boldsymbol{\beta}^*$ as the population FLS minimizer, $\hat{w}^*$ is the orthogonal projection of w onto the linear model space, so $\mathbb{E}[g_{\bar{Y}}(\varepsilon, f)] = 0$ for every function f in the span of the covariates.

Under misspecification, $\mathbb{E}[\varepsilon \mid X, Z]$ may be nonzero. The excess second moment is the bias $b^2 = \mathbb{E}[\|\varepsilon\|^2] - \sigma^2$, where σ^2 is the irreducible noise variance in Assumption 4.1(iv). Under correct specification, $b = 0$ and ε is that noise term.

The empirical FLS estimator satisfies $\hat{\boldsymbol{\beta}}_n = \hat{\Sigma}^{-1}_{\boldsymbol{ZZ}} \hat{C}_{\boldsymbol{Z}Y}$. The tangent-space prediction error decomposes as

$$\begin{aligned} \mathbb{E}\big[\|w - \hat{w}\|^2_{g_{\bar{Y}}}\big] &= \mathbb{E}\big[\|(\hat{w}^* - \hat{w}) + \varepsilon\|^2\big] \\ &= \underbrace{\mathbb{E}\big[\|\hat{w}^* - \hat{w}\|^2\big]}_{\text{estimation variance}} + \underbrace{\mathbb{E}[\|\varepsilon\|^2]}_{\sigma^2 + b^2}, \end{aligned} \tag{27}$$

Conditional on the sample, the cross term equals $(\boldsymbol{\beta}^* - \hat{\boldsymbol{\beta}}_n)^\top \mathbb{E}[\boldsymbol{Z}\varepsilon]$. It is zero because the new observation is independent of $\hat{\boldsymbol{\beta}}_n$, and the population normal equations give $\mathbb{E}[\boldsymbol{Z}\varepsilon] = 0$. This holds under correct specification and under misspecification.

The fitted prediction uses the empirical base points $\hat{\theta}$. The difference $\delta = \mathrm{Log}_{\hat{Y}}(\hat{Y}) - \hat{w}$ in (26) is a smooth function of $\hat{\theta} - \theta^*$. The mean-square bound for the Fréchet means, proved below, gives $\mathbb{E}\|\delta\|^2 = O(1/n)$. In the expansion of $\|w - \hat{w} - \delta\|^2$, the Cauchy–Schwarz inequality bounds the cross term in $\hat{w}^* - \hat{w}$ by $O(1/n)$. The cross term in ε has expectation $\mathbb{E}\langle \mathbb{E}(\varepsilon \mid X, z), \mathbb{E}(\delta \mid X, z)\rangle$. This expectation is zero under correct specification. Under misspecification it is $O(1/n)$. The expansion $\hat{\theta} - \theta^* = n^{-1}\sum_i \psi_i + r_n$ has $\mathbb{E}\psi_i = 0$ and $\|r_n\| \le C\|\hat{\theta} - \theta^*\|^2$, so $\|\mathbb{E}(\hat{\theta} - \theta^*)\| \le C\,\mathbb{E}\|\hat{\theta} - \theta^*\|^2$. These bounds give $\mathbb{E}\|w - \hat{w} - \delta\|^2 = \mathbb{E}\|w - \hat{w}\|^2 + O(1/n)$.

The variance term measures how well $\hat{\Sigma}_{\boldsymbol{ZZ}}$ and $\hat{C}_{\boldsymbol{ZY}}$ approximate their population counterparts. By the law of large numbers, $\hat{\Sigma}_{\boldsymbol{ZZ}} \xrightarrow{\text{a.s.}} \Sigma_{\boldsymbol{ZZ}}$ and $\hat{C}_{\boldsymbol{ZY}} \xrightarrow{\text{a.s.}} C_{\boldsymbol{ZY}}$.

For the convergence rate, consider the error $\hat{\boldsymbol{\beta}}_n - \boldsymbol{\beta}^* = \hat{\Sigma}_{\boldsymbol{ZZ}}^{-1}\hat{C}_{\boldsymbol{ZY}} - \Sigma_{\boldsymbol{ZZ}}^{-1}C_{\boldsymbol{ZY}}$. Writing $\hat{\Sigma} = \Sigma + \Delta$ where $\Delta = \hat{\Sigma} - \Sigma$, on the event $E_n = \{\|\Delta\|_{\mathrm{op}} < \lambda_*/2\}$, which has probability tending to one by Chebyshev's inequality and (29), the Neumann series $\hat{\Sigma}^{-1} = (\Sigma + \Delta)^{-1} = (\mathrm{id} + \Sigma^{-1}\Delta)^{-1}\Sigma^{-1}$ converges and $\left\|\hat{\Sigma}^{-1}\right\|_{\mathrm{op}} \le 2/\lambda_*$, giving

$$\begin{aligned}\left\|\hat{\boldsymbol{\beta}}_n - \boldsymbol{\beta}^*\right\| &\le \left\|\hat{\Sigma}^{-1}\right\|_{\mathrm{op}} \left\|\hat{C} - C\right\| + \left\|\hat{\Sigma}^{-1} - \Sigma^{-1}\right\|_{\mathrm{op}} \|C\| \\ &\le \frac{2}{\lambda_*}\left\|\hat{C} - C\right\| + \frac{2}{\lambda_*^2}\|\Delta\|_{\mathrm{op}}\,\|C\|\,, \end{aligned} \tag{28}$$

on the same event.

The empirical Gram matrix $\hat{\Sigma}$ and cross-covariance $\hat{C}$ are constructed using design vectors $\hat{Z}_i$ that depend on the empirical Fréchet means $\hat{\bar{Y}}, \hat{\bar{X}}_j$, which are computed from all n observations. We decompose $\hat{\Sigma} - \Sigma = (\hat{\Sigma}_{\mathrm{orcl}} - \Sigma) + (\hat{\Sigma} - \hat{\Sigma}_{\mathrm{orcl}})$, where $\hat{\Sigma}_{\mathrm{orcl}} = n^{-1}\sum_i Z_i(\theta^*)Z_i(\theta^*)^\top$ uses the population Fréchet means $\theta^* = (\bar{Y}, \bar{X}_j)$. The oracle term has independent and identically distributed summands. The Frobenius norm and the second moments of the entries give $\mathbb{E}\left\|\hat{\Sigma}_{\mathrm{orcl}} - \Sigma\right\|_{\mathrm{op}}^2 \le (p + qd)^2 M_4/n$. For fixed p, q, and d, $(p + qd)M_4$ is constant and is included in C_1 below. The base-point correction $\hat{\Sigma} - \hat{\Sigma}_{\mathrm{orcl}}$ is bounded by a Taylor expansion in θ around θ^* and by the mean-square Fréchet-mean rate $\mathbb{E}\|\hat{\theta} - \theta^*\|^2 = O(1/n)$. Under the Afsari condition the Fréchet functions are strongly convex on the support balls. On these compact supports the empirical Fréchet functions are averages of bounded functions, and the bounded-differences inequality gives this mean-square rate. The design is Lipschitz in the base points, because the supports lie strictly inside the injectivity balls. Both contributions have mean square $O(1/n)$, so

$$\mathbb{E}\big[\left\|\hat{\Sigma} - \Sigma\right\|_{\mathrm{op}}^2\big] \le \frac{C_1\,(p + qd)}{n}, \qquad \mathbb{E}\big[\left\|\hat{C} - C\right\|^2\big] \le \frac{C_1'\,(p + qd)\,M_4}{n}, \tag{29}$$

where $M_4 = \max(\mathbb{E}[\|u_j\|^4], \mathbb{E}[\tilde{z}_k^4], \mathbb{E}[\|w\|^4])$ is the largest fourth moment of the covariates and the response, and C_1, C_1' are constants depending on the curvature of the manifold and the smoothness of the log and transport maps in the base points. Therefore, on E_n,

$$\mathbb{E}\big[\left\|\hat{\boldsymbol{\beta}}_n - \boldsymbol{\beta}^*\right\|^2 \mathbf{1}_{E_n}\big] \le \frac{C_2\,M_4\,(p + qd)}{n\lambda_*^2} \tag{30}$$

for a constant $C_2 > 0$. By Chebyshev's inequality, $P(E_n^c) \le 4\,\mathbb{E}\,\|\Delta\|_{\mathrm{op}}^2/\lambda_*^2 \le 4C_1(p + qd)/(n\lambda_*^2)$. This is the bound in Corollary 4.3. Equation (30) and $P(E_n^c) \to 0$ give $\|\hat{\boldsymbol{\beta}}_n - \boldsymbol{\beta}^*\|^2 = O_P((p + qd)/n)$.

On E_n, the variance term in (27) satisfies $\mathbb{E}[\|\hat{w}^* - \hat{w}\|^2 \mathbf{1}_{E_n}] \le \|\Sigma_{\boldsymbol{ZZ}}\|_{\mathrm{op}}\,\mathbb{E}[\left\|\hat{\boldsymbol{\beta}}_n - \boldsymbol{\beta}^*\right\|^2 \mathbf{1}_{E_n}]$, since (X, z, Y) is independent of the sample. Combining this bound with (30) gives

$$\mathbb{E}\big[\|\hat{w}^* - \hat{w}\|^2 \mathbf{1}_{E_n}\big] \le \frac{C_3\,M_4\,(p + qd)}{n\lambda_*}, \tag{31}$$

where C_3 absorbs $\|\Sigma_{\boldsymbol{ZZ}}\|_{\mathrm{op}}/\lambda_*$. Under the Afsari condition, M_4 is bounded by a constant depending on the relevant injectivity radius and the covariate distribution.

By Assumption 4.1(ii), $d_{\mathcal{M}}(\hat{Y}, Y) \le 2r_0$, so $\mathbb{E}[d^2_{\mathcal{M}}(\hat{Y}, Y)] \le \mathbb{E}[d^2_{\mathcal{M}}(\hat{Y}, Y)\mathbf{1}_{E_n}] + 4r_0^2\, P(E_n^c)$. The second term, the $O(1/n)$ correction for δ, and the variance term are each $O((p+qd)/n)$ and are included in C_3. Substituting into (26) and using (27) on E_n gives

$$\begin{aligned}\mathbb{E}\big[d^2_{\mathcal{M}}(\hat{Y}, Y)\big] &\le \tilde{\eta}(\kappa_-, r_0)^2\, \mathbb{E}\big[\|w - \hat{w} - \delta\|^2_{g_{\bar{Y}}}\, \mathbf{1}_{E_n}\big] + 4r_0^2\, P(E_n^c) \\ &\le \tilde{\eta}(\kappa_-, r_0)^2 \left(\frac{C_3\, M_4\, (p+qd)}{n\lambda_*} + \sigma^2 + b^2\right). \end{aligned} \tag{32}$$

Setting $C \coloneqq C_3 M_4/\sigma^2$ (finite under the Afsari condition and the finite fourth-moment hypothesis) collects the fourth-to-second moment ratio and recovers the $\tilde{\eta}^2(C\sigma^2(p+qd)/(n\lambda_*) + \sigma^2 + b^2)$ form stated in Theorem 4.2.

As $n \to \infty$ the right-hand side tends to $\tilde{\eta}(\kappa_-, r_0)^2(\sigma^2 + b^2)$, so under correct specification ($b = 0$) the excess risk is bounded by $\tilde{\eta}^2\sigma^2 + O(n^{-1})$, and (30) with $P(E_n^c) \to 0$ gives $\hat{\boldsymbol{\beta}}_n \xrightarrow{P} \boldsymbol{\beta}^*$. The variance contribution therefore decays at the parametric n^{-1} rate, with a purely geometric factor $\tilde{\eta}^2 \ge 1$ in front. When $\kappa_- < 0$, as on $\mathbb{H}^d$, the factor is strictly larger than one. When $\kappa_- \ge 0$, as on S^2 and $\mathrm{SO}(3)$, $\tilde{\eta} = 1$ and the excess-risk rate is the Euclidean $O(n^{-1})$. □

E Proof of Asymptotic Normality (Theorem 4.5)

The main claim of Theorem 4.5 is the asymptotic-linearity expansion $\sqrt{n}(\hat{\boldsymbol{\beta}}_n - \boldsymbol{\beta}^*) = n^{-1/2}\sum_i \mathrm{IF}(O_i) + o_P(1)$ for an influence function IF built from the oracle tangent-space score and the Fréchet-mean influence functions. The limiting covariance is the sandwich $\Sigma_{\boldsymbol{\beta}} = \mathbb{E}[\mathrm{IF}(O)\mathrm{IF}(O)^\top]$. The tangent-space summands may be heteroskedastic. Under condition (D2) in the proof, the sandwich reduces to $\Sigma_{\mathrm{tan}} + \Sigma_{\mathrm{curv}}$.

Proof. Recall $d = \dim \mathcal{M}$ from Section 2. Fix an orthonormal basis of $T_{\bar{Y}}\mathcal{M}$ and identify the product tangent space $\mathcal{H} = \mathbb{R}^p \times (T_{\bar{Y}}\mathcal{M})^q$ with $\mathbb{R}^{p+qd}$.

We first establish a central limit theorem for the oracle FLS estimator $\tilde{\boldsymbol{\beta}}_n$ computed with the true Fréchet means $\bar{Y}$ and $\bar{X}_j$. By the decomposition $w_i = \boldsymbol{Z}_i^\top \boldsymbol{\beta}^* + \varepsilon_i$, the oracle estimation error is

$$\tilde{\boldsymbol{\beta}}_n - \boldsymbol{\beta}^* = \hat{\Sigma}^{-1}_{\boldsymbol{ZZ}}\, \frac{1}{n}\sum_{i=1}^n \boldsymbol{Z}_i \varepsilon_i. \tag{33}$$

Define $\xi_i \coloneqq \boldsymbol{Z}_i\varepsilon_i \in \mathbb{R}^{p+qd}$. The summands $\xi_1, \dots, \xi_n$ are independent and identically distributed with $\mathbb{E}[\xi_i] = 0$ (by the population normal equations $\mathbb{E}[\boldsymbol{Z}\varepsilon] = 0$) and covariance

$$\Omega \coloneqq \mathbb{E}\big[\xi_i \xi_i^\top\big] = \mathbb{E}\big[\boldsymbol{Z}_i\, \varepsilon_i \varepsilon_i^\top\, \boldsymbol{Z}_i^\top\big], \tag{34}$$

which allows heteroskedastic and design-dependent tangent-space noise. The finite fourth-moment hypothesis guarantees $\mathbb{E}[\|\xi_i\|^4] < \infty$ and hence the Lindeberg condition. By the multivariate central limit theorem,

$$\frac{1}{\sqrt{n}}\sum_{i=1}^n \xi_i \;\xrightarrow{d}\; \mathcal{N}(0, \Omega). \tag{35}$$

By the strong law of large numbers, $\hat{\Sigma}_{\boldsymbol{ZZ}} \xrightarrow{\text{a.s.}} \Sigma_{\boldsymbol{ZZ}}$, and $\Sigma_{\boldsymbol{ZZ}}$ is invertible by Assumption 4.1(iii). Slutsky's lemma applied to (33) and (35) yields

$$\sqrt{n}\big(\tilde{\boldsymbol{\beta}}_n - \boldsymbol{\beta}^*\big) \;\xrightarrow{d}\; \mathcal{N}\big(0,\; \Sigma^{-1}_{\boldsymbol{ZZ}}\Omega\Sigma^{-1}_{\boldsymbol{ZZ}}\big). \tag{36}$$

To bridge from this oracle result to the plug-in estimator, we record a Neyman-orthogonality property of the FLS score with respect to the base-point nuisance parameters. In practice, the true Fréchet means are replaced by their empirical counterparts $\hat{\theta}_n = (\hat{\bar{Y}}, \hat{\bar{X}}_1, \dots, \hat{\bar{X}}_p)$. Write $\theta^* = (\bar{Y}, \bar{X}_1, \dots, \bar{X}_p)$. For a base-point configuration θ in a geodesic neighborhood of θ^*, let $\boldsymbol{Z}_i(\theta)$,

$w_i(\theta)$, $\Sigma(\theta)$, $C(\theta)$ denote the θ-dependent design vector, response lift, Gram matrix, and cross-covariance. The population minimizer at θ is $\boldsymbol{\beta}^*(\theta) = \Sigma(\theta)^{-1}C(\theta)$, and the population residual is $\varepsilon_i(\theta) = w_i(\theta) - \boldsymbol{Z}_i(\theta)^\top \boldsymbol{\beta}^*(\theta)$.

Since $\boldsymbol{\beta}^*(\theta)$ minimizes the FLS objective at base points θ, the population normal equations give

$$\mathbb{E}\big[\boldsymbol{Z}_i(\theta)\,\varepsilon_i(\theta)\big] = 0 \quad \text{for all } \theta \text{ near } \theta^*. \tag{37}$$

Under the Afsari support condition, the logarithmic map and parallel transport are smooth in the base point (see Cheeger and Ebin [1975, Chapter 1]), so the left-hand side of (37) is a differentiable function of θ. Since (37) holds identically in θ, differentiation yields

$$\partial_\theta\, \mathbb{E}\big[\boldsymbol{Z}_i(\theta)\,\varepsilon_i(\theta)\big]\Big|_{\theta=\theta^*} = 0. \tag{38}$$

Equation (38) is the Neyman-orthogonality condition of the FLS score with respect to the nuisance parameter θ [Chernozhukov et al., 2018].

With (38) in hand, we can now show that the plug-in and oracle scores are asymptotically equivalent. Define the empirical score $S_n(\theta) = n^{-1}\sum_{i=1}^n \boldsymbol{Z}_i(\theta)\,\varepsilon_i(\theta)$. The plug-in estimator satisfies

$$\hat{\boldsymbol{\beta}}_n - \boldsymbol{\beta}^*(\hat\theta_n) = \hat\Sigma(\hat\theta_n)^{-1}\, S_n(\hat\theta_n). \tag{39}$$

Working in normal coordinates around θ^*, a first-order Taylor expansion of $\sqrt{n}\,S_n(\hat\theta_n)$ gives

$$\sqrt{n}\,S_n(\hat\theta_n) = \sqrt{n}\,S_n(\theta^*) + \big[\partial_\theta S_n(\theta^*)\big]\,\sqrt{n}(\hat\theta_n - \theta^*) + R_n. \tag{40}$$

The smoothness of the geometric maps under the Afsari condition together with the finite fourth-moment hypothesis yields a uniform bound on $\partial_\theta^2 S_n$ in a neighborhood of θ^*. Combined with the Fréchet mean estimation rate $\|\hat\theta_n - \theta^*\| = O_P(n^{-1/2})$ [Bhattacharya and Patrangenaru, 2005], this gives $R_n = O_P(\sqrt{n}\,\|\hat\theta_n - \theta^*\|^2) = O_P(n^{-1/2}) = o_P(1)$. For the first-order term, the law of large numbers gives

$$\partial_\theta S_n(\theta^*) \;\xrightarrow{P}\; \partial_\theta\, \mathbb{E}\big[\boldsymbol{Z}(\theta)\varepsilon(\theta)\big]\big|_{\theta^*} = 0$$

by the Neyman condition (38), where the interchange of differentiation and expectation is justified by the dominated convergence theorem under the bounded support guaranteed by the Afsari condition. Since $\sqrt{n}(\hat\theta_n - \theta^*) = O_P(1)$, the first-order term is $o_P(1)\cdot O_P(1) = o_P(1)$. Substituting into (40) yields the score equivalence

$$\sqrt{n}\,S_n(\hat\theta_n) = \sqrt{n}\,S_n(\theta^*) + o_P(1). \tag{41}$$

Since $\hat\Sigma(\hat\theta_n) \xrightarrow{P} \Sigma_{\boldsymbol{ZZ}}$ by the law of large numbers and the continuity of the geometric operations in θ, combining (39) and (41) gives

$$\sqrt{n}\big(\hat{\boldsymbol{\beta}}_n - \boldsymbol{\beta}^*(\hat\theta_n)\big) = \sqrt{n}\big(\tilde{\boldsymbol{\beta}}_n - \boldsymbol{\beta}^*\big) + o_P(1). \tag{42}$$

It remains to incorporate the base-point sensitivity $\boldsymbol{\beta}^*(\hat\theta_n) - \boldsymbol{\beta}^*$ into the limit distribution. Write $J_\theta = \partial_\theta \boldsymbol{\beta}^*(\theta^*)$. Since $\boldsymbol{\beta}^*(\theta) = \Sigma(\theta)^{-1}C(\theta)$, the chain rule yields

$$J_\theta = \Sigma_{\boldsymbol{ZZ}}^{-1}\big[\partial_\theta C(\theta^*) - \big(\partial_\theta \Sigma(\theta^*)\big)\,\boldsymbol{\beta}^*\big].$$

The Fréchet mean centering property $\mathbb{E}[\boldsymbol{Z}(\theta^*)] = 0$ and $\mathbb{E}[w(\theta^*)] = 0$ cancels the leading terms in $\partial_\theta\Sigma(\theta^*)$ and $\partial_\theta C(\theta^*)$. The remainders are curvature-driven. Let $K = \max(|\kappa_-|, \kappa_+)$. This constant bounds the absolute sectional curvature. Expanding $\mathrm{Log}_{\theta+\delta}$ and the associated parallel transport in normal coordinates around θ^* to second order [Pennec et al., 2006] and averaging against the distribution of $(\boldsymbol{u}, \tilde{\boldsymbol{z}})$, each surviving contribution is the product of a component of the Riemann tensor with a second-order or third-order moment of the covariates, so

$$J_\theta \;=\; \mathcal{R}\big(\mu_2,\, \mu_3^{(\tilde z)},\, \mu_3^{(\boldsymbol{u})}\big) \;+\; O\big(K^2 r_0^4 \lambda_*^{-1}\big), \tag{43}$$

where μ_2 denotes the cross moments of each u_j with the remaining design columns. The terms $\mu_3^{(\tilde z)}$ and $\mu_3^{(\boldsymbol{u})}$ denote the third cumulants of the scalar covariates and of the transported manifold-predictor log maps, and $\mathcal{R}$ is the linear operator contracting them against the Riemann curvature at θ^*, with operator norm $\|\mathcal{R}\| = O(K)$. The second-order terms are the holonomy of $\Gamma_{\bar X_j \to \bar Y}$ as the base point

varies, and they vanish when $\bar{X}_j = \bar{Y}$ for every j. Consequently $J_\theta = 0$ whenever the Riemann tensor vanishes, equivalently when $\kappa_+ = \kappa_- = 0$. On $\mathrm{Sym}^+(k)$ one has $\kappa_+ = 0$ and $\kappa_- < 0$. The Riemann tensor can then be nonzero, and so can J_θ. As a concrete illustration, on $\mathbb{S}^2$ with $p = 0$, $q = 1$, and scalar $\tilde{z}$ with third cumulant $\mu_3 := \mathbb{E}[\tilde{z}^3]$, (43) reduces at leading order in r_0 to $\partial_{\bar{Y}} \boldsymbol{\beta}^*(\theta^*) = \big(\mu_3/(3\mathrm{Var}(\tilde{z}))\big)\big(\|v\|^2 I - vv^\top\big)$, which is non-zero whenever $\mu_3 \neq 0$ and $v \neq 0$.

By the nonsingular-Hessian condition in Theorem 4.5, the joint empirical Fréchet mean admits the asymptotically linear representation

$$\sqrt{n}(\hat{\theta}_n - \theta^*) = n^{-1/2} \sum_{i=1}^{n} \psi_\theta(O_i) + o_P(1), \qquad \mathbb{E}[\psi_\theta(O)] = 0,$$

with covariance $\Sigma_\theta = \mathbb{E}[\psi_\theta(O)\psi_\theta(O)^\top]$ [Bhattacharya and Patrangenaru, 2005, Bhattacharya and Lin, 2017]. Taylor expansion therefore gives

$$\sqrt{n}\big(\boldsymbol{\beta}^*(\hat{\theta}_n) - \boldsymbol{\beta}^*\big) \;=\; n^{-1/2} \sum_{i=1}^{n} J_\theta \psi_\theta(O_i) + o_P(1). \tag{44}$$

Combining (42) and (44), the full plug-in estimator has influence function

$$\mathrm{IF}(O) = \Sigma_{\boldsymbol{ZZ}}^{-1} \boldsymbol{Z}\varepsilon + J_\theta \psi_\theta(O). \tag{45}$$

The multivariate CLT applied to this joint summand gives the sandwich limit

$$\sqrt{n}\big(\hat{\boldsymbol{\beta}}_n - \boldsymbol{\beta}^*\big) \;=\; n^{-1/2} \sum_{i=1}^{n} \mathrm{IF}(O_i) + o_P(1) \;\xrightarrow{d}\; \mathcal{N}(0, \Sigma_{\boldsymbol{\beta}}), \tag{46}$$

where

$$\Sigma_{\boldsymbol{\beta}} = \mathbb{E}\Big[\big\{\Sigma_{\boldsymbol{ZZ}}^{-1}\boldsymbol{Z}\varepsilon + J_\theta\psi_\theta(O)\big\}\big\{\Sigma_{\boldsymbol{ZZ}}^{-1}\boldsymbol{Z}\varepsilon + J_\theta\psi_\theta(O)\big\}^\top\Big]. \tag{47}$$

Equivalently,

$$\Sigma_{\boldsymbol{\beta}} = \Sigma_{\boldsymbol{ZZ}}^{-1}\Omega\Sigma_{\boldsymbol{ZZ}}^{-1} + J_\theta\Sigma_\theta J_\theta^\top + \Sigma_{\boldsymbol{ZZ}}^{-1}\mathbb{E}[\boldsymbol{Z}\varepsilon\,\psi_\theta(O)^\top]J_\theta^\top + J_\theta\mathbb{E}[\psi_\theta(O)\varepsilon^\top\boldsymbol{Z}^\top]\Sigma_{\boldsymbol{ZZ}}^{-1}.$$

The final two terms are the cross-covariances between the oracle tangent score and the Fréchet-mean influence function. They vanish under (D2), which requires the oracle tangent score to be uncorrelated with the Fréchet-mean influence functions. Under the correct specification in Theorem 4.5, (D2) holds when the conditional covariance of ε given (X, z) does not depend on (X, z). The identity $\mathbb{E}[\boldsymbol{Z}\varepsilon\, w^\top] = \mathbb{E}[\boldsymbol{Z}\,\mathbb{E}(\varepsilon\varepsilon^\top \mid X, z)]$ and the centering $\mathbb{E}[\boldsymbol{Z}] = 0$ then give $\mathbb{E}[\boldsymbol{Z}\varepsilon\, w^\top] = 0$. The covariance is then $\Sigma_{\tan} + \Sigma_{\mathrm{curv}}$, with $\Sigma_{\tan} = \Sigma_{\boldsymbol{ZZ}}^{-1}\Omega\Sigma_{\boldsymbol{ZZ}}^{-1}$ and $\Sigma_{\mathrm{curv}} = J_\theta\Sigma_\theta J_\theta^\top$. Appendix G uses two further conditions. Under (D1) the tangent noise is isotropic, and $\Sigma_{\tan}$ is proportional to $\Sigma_{\boldsymbol{ZZ}}^{-1}$. Under (D3) the Fréchet-mean covariance is block diagonal, and Σ_{curv} is a sum over the $p + 1$ means.

The stated t- and Wald limits now follow by projection. For a tested manifold coefficient, the marginal distribution of $\sqrt{n}(\hat{\alpha}_j - \alpha_j^*)$ has variance $[\Sigma_{\boldsymbol{\beta}}]_{jj}$. If $[\Sigma_{\boldsymbol{\beta}}]_{jj} > 0$ and $[\hat{\Sigma}_{\boldsymbol{\beta}}]_{jj} \to_P [\Sigma_{\boldsymbol{\beta}}]_{jj}$, Slutsky's lemma gives

$$T_j = \frac{\hat{\alpha}_j - \alpha_j^*}{\{[\hat{\Sigma}_{\boldsymbol{\beta}}]_{jj}/n\}^{1/2}} \xrightarrow{d} N(0, 1).$$

For a scalar-predictor tangent coefficient, the sub-vector $\hat{v}_k$ has asymptotic covariance $\Sigma_{\boldsymbol{\beta}}^{kk}/n$. If $\Sigma_{\boldsymbol{\beta}}^{kk}$ is nonsingular and $\hat{\Sigma}_{\boldsymbol{\beta}}^{kk} \to_P \Sigma_{\boldsymbol{\beta}}^{kk}$, the continuous mapping theorem gives

$$W_k = n(\hat{v}_k - v_k^*)^\top(\hat{\Sigma}_{\boldsymbol{\beta}}^{kk})^{-1}(\hat{v}_k - v_k^*) \xrightarrow{d} \chi_d^2.$$

Corollary 4.6 is the pairs-bootstrap counterpart of (6). It follows from the $\boldsymbol{\beta}$-block of Proposition E.1 below. □

E.1 Validity of the Pairs Bootstrap

Proposition E.1 is the Z-estimator statement behind Corollary 4.6. Conditions (ii), (iii), and (v) are the Hessian, differentiability, and fourth-moment hypotheses of Theorem 4.5, and condition (iv) is Assumption 4.1(iii). Condition (i) is Assumption 4.1(i)–(ii), namely that the Fréchet means are unique and the supports lie in the balls (5), whose closures are compact and strictly inside the injectivity radius. No further moment or design condition is imposed.

Proposition E.1 (Pairs bootstrap)**.** *Assume that (i) Y and each X_j have unique Fréchet means, with supports in compact sets strictly inside the injectivity radius of the respective mean; (ii) the Fréchet functions $F_Y(m) = \frac{1}{2}\mathbb{E}d^2_{\mathcal{M}}(Y,m)$ and $F_j(m) = \frac{1}{2}\mathbb{E}d^2_{\mathcal{M}}(X_j,m)$ are twice continuously differentiable near their minimizers with nonsingular Hessians H_Y and H_j; (iii) the logarithm and parallel-transport maps are jointly C^1 in the base points near the means and in their arguments; (iv) $\Sigma_{\boldsymbol{ZZ}}$ is nonsingular; and (v) the finite fourth moments of Theorem 4.5 hold, in particular $\mathbb{E}\|z\|^4 < \infty$. Let $\hat{\boldsymbol{\beta}}^*_n$ be the ROLS estimator on a pairs bootstrap resample, with the Fréchet means, logarithm maps, transports, and normal equations recomputed. Express $\hat{v}^*_k$ and $\hat{v}_k$ in one fixed orthonormal frame of $T_{\bar{Y}}\mathcal{M}$ by parallel transport. On a resample whose normal equations are singular, which has conditional probability tending to zero, define $\hat{\boldsymbol{\beta}}^*_n$ arbitrarily. Then, conditionally on the data and in probability, $\sqrt{n}(\hat{\boldsymbol{\beta}}^*_n - \hat{\boldsymbol{\beta}}_n) \rightsquigarrow \mathcal{N}(0, \Sigma_{\boldsymbol{\beta}})$ with $\Sigma_{\boldsymbol{\beta}} = \mathrm{Var}(\mathrm{IF}_{\boldsymbol{\beta}}(O))$, where $\mathrm{IF}_{\boldsymbol{\beta}}$ is given in* (48)*. The second-moment matrix of $\sqrt{n}(\hat{\boldsymbol{\beta}}^*_n - \hat{\boldsymbol{\beta}}_n)$, truncated at a level $\tau_n \to \infty$ that grows sufficiently slowly, converges in probability to $\Sigma_{\boldsymbol{\beta}}$.*

Proof. Write P for expectation under the data law and P_n for the empirical measure. For a pairs bootstrap resample, P^*_n denotes its empirical measure.

We first define the charts to differentiate a map that takes values in a moving tangent space. Fix orthonormal frames $E_Y\colon \mathbb{R}^d \to T_{\bar{Y}}\mathcal{M}$ and $E_j\colon \mathbb{R}^d \to T_{\bar{X}_j}\mathcal{M}$ at the population means, and write $m_Y(a) = \mathrm{Exp}_{\bar{Y}}(E_Y a)$ and $m_j(a) = \mathrm{Exp}_{\bar{X}_j}(E_j a)$ for $a \in \mathbb{R}^d$ near 0. A tangent vector at $m_Y(a)$ is expressed in $\mathbb{R}^d$ by $\iota = E_{\bar{Y}}^{-1} \circ \Gamma_{m_Y(a)\to\bar{Y}}$, and a tangent vector at $m_j(a)$ by $\iota_j = E_j^{-1} \circ \Gamma_{m_j(a)\to\bar{X}_j}$. By (i) and (iii) these identifications are C^1 for base points near the means, so $\vartheta = (a_Y, a_1, \ldots, a_p, \mu_z, \boldsymbol{\beta})$ may be treated as a Euclidean parameter, with true value $\vartheta^* = (0, \ldots, 0, \bar{z}, \boldsymbol{\beta}^*)$.

In these coordinates we stack the Fréchet-mean equations and the normal equations into one estimating function. For base points and centering given by ϑ, let w_ϑ, $u_{j,\vartheta}$, and $\tilde{z}_\vartheta = z - \mu_z$ be the lifted response, the transported predictors, and the centered covariates in the coordinates ι, and set $r_\vartheta = w_\vartheta - \sum_j \alpha_j u_{j,\vartheta} - \sum_k \tilde{z}_{\vartheta,k} v_k$. Define $\psi(O;\vartheta) = (\psi_Y, \psi_1, \ldots, \psi_p, \psi_z, \psi_{\boldsymbol{\beta}})$ by

$$\psi_Y = \iota \,\mathrm{Log}_{m_Y(a_Y)}(Y), \quad \psi_j = \iota_j \,\mathrm{Log}_{m_j(a_j)}(X_j), \quad \psi_z = z - \mu_z,$$
$$\psi_{\boldsymbol{\beta}} = \big(\langle u_{j,\vartheta},\, r_\vartheta\rangle\big)_{j\le p} \oplus \big(\tilde{z}_{\vartheta,k}\, r_\vartheta\big)_{k\le q}.$$

At ϑ^*, $P\psi_Y = -\iota\,\nabla F_Y(\bar{Y}) = 0$, and likewise $P\psi_j = P\psi_z = 0$. The block $P\psi_{\boldsymbol{\beta}}(\cdot;\vartheta^*) = 0$ is the population normal equation of Theorem 3.3, so $P\psi(\cdot;\vartheta^*) = 0$. Let Θ_0 be the neighborhood defined in the next paragraph. On the event that the sample means lie in Θ_0, the equations $\sum_i \mathrm{Log}_{\hat{\bar{Y}}}(Y_i) = 0$ and $\sum_i \mathrm{Log}_{\hat{\bar{X}}_j}(X_{ij}) = 0$ are interior critical-point equations by the choice of Θ_0, and on the event that the Gram matrix $\hat{A}$ of Section 4.2 is nonsingular, $P_n\psi_{\boldsymbol{\beta}} = 0$ is the ROLS normal equation, so $\hat{\vartheta}_n$ solves $P_n\psi(\cdot;\hat{\vartheta}_n) = 0$. On the corresponding event for the resample, $\hat{\vartheta}^*_n$ solves the same equation with P_n replaced by P^*_n. These events have probability tending to one, as shown below.

The estimating functions form a Donsker class on a neighborhood of ϑ^*. Let Θ_0 be a compact convex neighborhood of ϑ^*, small enough that the supports in (i) remain strictly inside the injectivity radius at every base point in Θ_0. Convexity keeps the segment between any two parameters inside Θ_0. On this neighborhood, (i) and (iii) make w_ϑ, $u_{j,\vartheta}$, and their ϑ-derivatives bounded, so ψ_Y and ψ_j are bounded and $\psi_z = O(1 + \|z\|)$. Since $\boldsymbol{\beta}$ ranges over a compact set, $r_\vartheta = O(1 + \|z\|)$, and therefore $\langle u_{j,\vartheta},\, r_\vartheta\rangle = O(1 + \|z\|)$ and $\tilde{z}_{\vartheta,k} r_\vartheta = O(1 + \|z\|^2)$. The same orders hold for the ϑ-derivatives. The largest is $\partial_{v_{k'}}(\tilde{z}_{\vartheta,k} r_\vartheta) = -\tilde{z}_{\vartheta,k}\tilde{z}_{\vartheta,k'} I_d$. Hence $\|\psi(O;\vartheta_1) - \psi(O;\vartheta_2)\| \le L(O)\|\vartheta_1 - \vartheta_2\|$ and $\sup_{\Theta_0}\|\psi(O;\vartheta)\| \le L(O)$, with $L(O) \le C(1 + \|z\|^2)$. Condition (v) gives $\mathbb{E}L(O)^2 < \infty$. A class that is Lipschitz in a finite-dimensional parameter, with a square-integrable Lipschitz function, is P-Donsker [van der Vaart, 1998, Example 19.7].

Differentiating $P\psi$ at ϑ^* gives the Jacobian $-M$. In the block order $(a_Y, a_1, \ldots, a_p, \mu_z, \boldsymbol{\beta})$,

$$M = \begin{pmatrix} H_Y & & & & & \\ & H_1 & & & & \\ & & \ddots & & & \\ & & & I_q & & \\ J_Y & J_1 & \cdots & J_z & \Sigma_{\boldsymbol{ZZ}} \end{pmatrix}.$$

The diagonal blocks of M are $H_Y, H_1, \ldots, H_p$, I_q, and $\Sigma_{\boldsymbol{ZZ}}$, because at ϑ^* the derivatives of $P\psi_Y$, $P\psi_j$, $P\psi_z$, and $P\psi_{\boldsymbol{\beta}}$ in their own parameters are $-H_Y$, $-H_j$, $-I_q$, and $-\Sigma_{\boldsymbol{ZZ}}$, and the derivative of $P\psi$ is $-M$. Every block strictly above the diagonal is zero, because ψ_Y, ψ_j, and ψ_z do not depend on $\boldsymbol{\beta}$ or on one another's parameters. The blocks $J_\bullet = -\partial_\bullet P\psi_{\boldsymbol{\beta}}(\cdot; \vartheta^*)$ exist by (iii). In particular $J_z = 0$, because $\mathbb{E}w = \mathbb{E}u_j = 0$ at the Fréchet means and $\mathbb{E}\tilde{z} = 0$. Thus M is block lower triangular, and it is nonsingular by (ii) and (iv).

This nonsingular Jacobian gives the asymptotic linearity of the sample estimator. After shrinking Θ_0 if necessary, ϑ^* is the only root of $P\psi(\cdot; \vartheta) = 0$ in Θ_0. At a root, a_Y is a critical point of F_Y near $\bar{Y}$, which is isolated by (ii), and likewise each a_j, then $\mu_z = \bar{z}$, and then $\boldsymbol{\beta} = \boldsymbol{\beta}^*$ by (iv). The sample Fréchet means converge to the population means [Bhattacharya and Patrangenaru, 2003], since the means are unique and the supports are compact. The entries of $n^{-1}\hat{A}$ and of $n^{-1}\sum_i \boldsymbol{Z}_i w_i$, as functions of the base points and of μ_z, are Lipschitz with envelope $C(1 + \|z\|^2)$, so they converge uniformly over Θ_0 by the Glivenko–Cantelli theorem. At the sample means, $n^{-1}\hat{A}$ therefore converges to $\Sigma_{\boldsymbol{ZZ}}$, which is nonsingular by (iv), and $\hat{\mu}_z$ and $\hat{\boldsymbol{\beta}}_n$ converge to $\bar{z}$ and $\boldsymbol{\beta}^*$. Hence the sample normal equations hold and $\hat{\vartheta}_n$ lies in Θ_0 with probability tending to one, and $\hat{\vartheta}_n \to \vartheta^*$ in probability. The Z-estimator theorem [van der Vaart, 1998, Theorem 5.21] then yields

$$\sqrt{n}(\hat{\vartheta}_n - \vartheta^*) = n^{-1/2} \sum_{i=1}^{n} M^{-1}\psi(O_i; \vartheta^*) + o_P(1).$$

The $\boldsymbol{\beta}$-block of $M^{-1}\psi(\cdot; \vartheta^*)$ is

$$\mathrm{IF}_{\boldsymbol{\beta}} = \Sigma_{\boldsymbol{ZZ}}^{-1}\Big(\psi_{\boldsymbol{\beta}} - \sum_{m \in \{Y, 1, \ldots, p\}} J_m H_m^{-1} \psi_m\Big). \tag{48}$$

At ϑ^*, $\psi_{\boldsymbol{\beta}} = \boldsymbol{Z}\varepsilon$. The term $H_m^{-1}\psi_m$ is the influence function of the corresponding Fréchet mean, and $-\Sigma_{\boldsymbol{ZZ}}^{-1} J_m$ is the corresponding block of J_θ. Therefore (48) is the same influence function as (45). Thus $\Sigma_{\boldsymbol{\beta}} = \mathrm{Var}(\mathrm{IF}_{\boldsymbol{\beta}})$ under the conditions of the proposition. Condition (D2) of Appendix E is the further assumption under which this variance splits as $\Sigma_{\boldsymbol{\beta}} = \Sigma_{\tan} + \Sigma_{\mathrm{curv}}$. The leading term of (48) then contributes $\Sigma_{\tan}$, and the sum over Fréchet means contributes Σ_{curv}.

The pairs bootstrap obeys the same expansion conditionally on the data. Because $\{\psi(\cdot; \vartheta) : \vartheta \in \Theta_0\}$ is Donsker, the bootstrap empirical process $\sqrt{n}(P_n^* - P_n)$ converges conditionally in probability to the same tight Gaussian limit as $\sqrt{n}(P_n - P)$ [van der Vaart and Wellner, 1996, Theorem 3.6.1]. A distribution supported in the ball $B(\bar{Y}, r_Y)$ has every Fréchet minimizer in the closed ball $B(\bar{Y}, 2r_Y)$, because outside that ball its Fréchet function exceeds $\frac{1}{2}r_Y^2$, which is at least its value at $\bar{Y}$, and this closed ball is compact because $\mathcal{M}$ is complete. On this ball, $d_{\mathcal{M}}^2(y, \cdot)$ is bounded and Lipschitz uniformly over y in the support, so the resampled Fréchet functions converge there uniformly, conditionally in probability, and uniqueness of the mean gives conditional consistency of the resampled mean of Y. The same argument applies to each X_j. The entries of $n^{-1}\hat{A}$ and of $n^{-1}\sum_i \boldsymbol{Z}_i w_i$ form Donsker classes by the same Lipschitz bound, so Theorem 3.6.1 of van der Vaart and Wellner [1996] gives their bootstrap convergence, uniformly over Θ_0 and conditionally in probability. Hence $\hat{\mu}_z^*$ and $\hat{\boldsymbol{\beta}}_n^*$ are conditionally consistent, the resampled normal equations are nonsingular and $\hat{\vartheta}_n^*$ lies in Θ_0 with conditional probability tending to one in probability, and $\hat{\vartheta}_n^* \to \vartheta^*$ conditionally in probability. The derivative $-M$ is nonsingular, and the Lipschitz bound implies the L^2-continuity $P\|\psi(\cdot; \vartheta) - \psi(\cdot; \vartheta^*)\|^2 \to 0$ as $\vartheta \to \vartheta^*$. The bootstrap Z-estimator theorem [Wellner and Zhan, 1996, Arcones and Giné, 1992] therefore gives, conditionally in probability,

$$\sqrt{n}(\hat{\vartheta}_n^* - \hat{\vartheta}_n) \rightsquigarrow \mathcal{N}\big(0,\, M^{-1}\mathrm{Var}\{\psi(O; \vartheta^*)\}M^{-\top}\big).$$

The $\boldsymbol{\beta}$-block of this covariance is $\Sigma_{\boldsymbol{\beta}}$. For each fixed τ, conditional weak convergence implies that the second-moment matrix of $\sqrt{n}(\hat{\boldsymbol{\beta}}_n^* - \hat{\boldsymbol{\beta}}_n)$, truncated at squared norm τ, converges in probability

to the same truncation of the Gaussian covariance, because the limit puts no mass on the sphere of radius $\sqrt{\tau}$. A diagonal argument extends this to some sequence $\tau_n \to \infty$. The second moment without truncation can fail to converge: a resample of identical observations has positive conditional probability, the normal equations are then singular, and $\hat{\boldsymbol{\beta}}_n^*$ was defined arbitrarily on that event. Section 4.2 instead compares $\hat{v}_k^*$ with $\hat{v}_k$ after transporting $\hat{v}_k^*$ to $T_{\hat{Y}}\mathcal{M}$. The two differences differ by the holonomy around a geodesic triangle with sides $O_P(n^{-1/2})$. That holonomy is $O_P(n^{-1})$, so the $\sqrt{n}$-scaled statistics differ by $O_P(n^{-1/2})$ and have the same limit law in any orthonormal frame at $\hat{Y}$ that converges to a frame at $\bar{Y}$, and the Wald statistic does not depend on the frame. □

F Special Cases and Manifold-Specific Derivations

This appendix gives the formulas behind the special cases summarized in Remark 3.2, namely Euclidean linear regression, geodesic regression, parametric Fréchet regression, and DIDO regression. It then records the manifold-specific operations used in the experiments.

F.1 Euclidean Linear Regression ($\mathcal{M} = \mathbb{R}^d$)

When $\mathcal{M} = \mathbb{R}^d$ with the standard inner product, all geometric operations collapse to linear algebra: $\mathrm{Log}_p(q) = q - p$, $\mathrm{Exp}_p(v) = p + v$, $\Gamma_{p\to q} = \mathrm{id}$, and the Fréchet mean coincides with the ordinary expectation. The model (1) becomes

$$Y_i - \bar{Y} = \sum_{j=1}^{p} \alpha_j (X_{ij} - \bar{X}_j) + \sum_{k=1}^{q} \tilde{z}_{ik}\, v_k + \varepsilon_i,$$

which is a centered Euclidean vector-valued linear model with scalar block coefficients for the vector covariates. When $d = 1$, this is ordinary centered multiple linear regression. The block normal equation (3) reduces to the ordinary normal equations $\hat{\Sigma}_{\boldsymbol{ZZ}}\hat{\beta} = \hat{C}_{\boldsymbol{ZY}}$ with the usual sample covariance matrix. In the correctly specified homoskedastic Gaussian case at fixed d, the Wald tests of Theorem 4.5 are the exact F-test for vector coefficients and the exact t-test for scalar coefficients. Outside that case the reference distributions are asymptotic. When n/d is small, the χ^2_d reference over-rejects relative to the F reference (Appendix H.7).

F.2 Geodesic Regression ($p = 0$, Single Scalar Predictor)

When $p = 0$ and $q = 1$, the model has no manifold predictors and one scalar covariate z, yielding

$$\mathrm{Log}_{\bar{Y}}(Y_i) = (z_i - \bar{z})\, v + \varepsilon_i.$$

The normal equation gives the closed-form solution $\hat{v} = \mathrm{cov}(z, Y)/\mathrm{var}(z)$, where $\mathrm{cov}(z, Y) = \frac{1}{n}\sum_i \tilde{z}_i\, w_i \in T_{\bar{Y}}\mathcal{M}$. The predicted geodesic is $\hat{Y}(z) = \mathrm{Exp}_{\bar{Y}}\big((z - \bar{z})\,\hat{v}\big)$, which is a single geodesic ray from $\bar{Y}$ in the direction $\hat{v}$, precisely the tangent-space formulation of geodesic regression [Fletcher, 2013].

More generally, in Euclidean space for $p = 0$ and $q \geq 2$, ROLS reduces to the parametric version of Fréchet regression [Petersen and Müller, 2019a] in which the weight function $w_i(z)$ is linear in z. The local version of Fréchet regression uses kernel weights $w_i(z) = K_h(z - z_i)/\sum_\ell K_h(z - z_\ell)$, making it nonparametric. On a curved or periodic manifold, Fréchet regression with linear weights minimizes a weighted sum of squared manifold distances and so differs from ROLS with $p = 0$.

F.3 DIDO Regression ($\mathcal{M} = \mathcal{P}_2(\mathbb{R})$)

Distribution-in-distribution-out (DIDO) regression [Chen et al., 2024] places both predictors and responses in the 2-Wasserstein space $\mathcal{P}_2(\mathbb{R})$. Via the quantile embedding $Q\colon \mathcal{P}_2(\mathbb{R}) \to L^2([0,1])$, the Wasserstein distance becomes the L^2 norm on quantile functions, and the tangent space at $\bar{Y}$ is identified with a closed subspace of $L^2([0,1])$.

On $\mathcal{P}_2(\mathbb{R})$ the sectional curvature vanishes (the space is flat under the quantile embedding) and the optimal-transport map between any two quantile functions is the identity rearrangement. Therefore

the parallel transport along any geodesic connecting $\bar{X}_j$ to $\bar{Y}$ reduces to the identity map in the quantile representation. Substituting into (1), the ROLS model becomes

$$Q_{Y_i} - Q_{\bar{Y}} = \sum_{j=1}^{p} \alpha_j \big(Q_{X_{ij}} - Q_{\bar{X}_j}\big) + \sum_{k=1}^{q} \tilde{z}_{ik}\, v_k + \varepsilon_i,$$

which is the linear model in the quantile Hilbert space studied by [Chen et al., 2024], broadened here to admit scalar covariates alongside distributional predictors.

Remark F.1 (Discretization)**.** *The quantile embedding is flat: curvature vanishes, the cut locus is empty, and parallel transport is the identity, so the model is linear regression in $L^2([0,1])$. In the computations, quantile functions are discretized on a grid of m points. The fit is then ordinary linear regression in $\mathbb{R}^m$, and Theorems 4.2–4.5 apply to that discretization.*

F.4 Manifold-Specific Derivations

We record the explicit forms of the geometric operations for each manifold used in the experiments and for $\mathcal{P}_2(\mathbb{R})$.

Sphere $S^d \subset \mathbb{R}^{d+1}$. $\mathrm{Exp}_p(v) = \cos(\|v\|)\, p + \sin(\|v\|)\, v/\|v\|$; $\mathrm{Log}_p(q) = \theta\,(q - \langle p, q\rangle p)/\|q - \langle p, q\rangle p\|$ where $\theta = \arccos(\langle p, q\rangle)$. Parallel transport along the geodesic from p to q: $\Gamma_{p\to q}(v) = v - \frac{\langle \mathrm{Log}_p(q), v\rangle}{\|\mathrm{Log}_p(q)\|^2}\big(\mathrm{Log}_p(q) + \mathrm{Log}_q(p)\big)$. Sectional curvature $\kappa = 1$ and $\mathrm{inj}(S^d) = \pi$.

$\mathrm{SO}(3)$**.** Identified with unit quaternions modulo sign. $\mathrm{Log}_R(S) = [\log(R^\top S)]^\vee \in \mathbb{R}^3$, where log is the matrix logarithm and $[\cdot]^\vee$ the inverse hat map. $\mathrm{Exp}_R(v) = R\exp([v]^\wedge)$. Parallel transport (Levi-Civita connection of the bi-invariant metric, in body-frame coordinates): $\Gamma_{R\to S}(v) = \exp([-\xi/2]^\wedge)\, v$ with $\xi = \mathrm{Log}_R(S)$. Both the left-Cartan and bi-invariant Levi-Civita transports are linear isometries that preserve tangent norms, so the geometric guarantees of Theorem 3.4 and the Rauch-comparison consistency bound (Theorem 4.2) apply under either convention; numerical coefficient values depend on the chosen convention. Bi-invariant metric with $\kappa = 1/4$ and $\mathrm{inj}(\mathrm{SO}(3)) = \pi$.

Flat torus $\mathbb{T}^d = \mathbb{R}^d/\Lambda$ (lattice Λ). $d_{\mathcal{M}}(p, q) = \min_{n\in\Lambda} \|p - q + n\|$; $\mathrm{Log}_p(q) = q - p + n^*$ with $n^* = \arg\min_{n\in\Lambda} \|q - p + n\|$; $\mathrm{Exp}_p(v) = (p + v) \bmod \Lambda$. The connection is flat, so parallel transport is the identity in coordinates, $\Gamma_{p\to q}(v) = v$, and the holonomy of every loop is trivial. Sectional curvature $\kappa = 0$ and $\mathrm{inj}(\mathbb{T}^d) = \frac{1}{2}\,\mathrm{sys}(\mathbb{T}^d)$. Trivial holonomy means parallel transport on $\mathbb{T}^d$ does not depend on the path.

$\mathrm{Sym}^+(k)$ **(affine-invariant metric).** $d_{\mathcal{M}}(P, Q) = \|\log(P^{-1/2}QP^{-1/2})\|_F$. $\mathrm{Log}_P(Q) = P^{1/2}\log(P^{-1/2}QP^{-1/2})P^{1/2}$; $\mathrm{Exp}_P(V) = P^{1/2}\exp(P^{-1/2}VP^{-1/2})P^{1/2}$. Parallel transport: $\Gamma_{P\to Q}(V) = E_0\, V\, E_0^\top$, where $E_0 = (QP^{-1})^{1/2}$ [Pennec et al., 2006]. Non-positive curvature ($\kappa_+ = 0$), $\mathrm{inj}(\mathrm{Sym}^+(k)) = \infty$ (Hadamard manifold).

Poincaré ball $\mathbb{H}^d$ (curvature $\kappa = -1$). $d_{\mathcal{M}}(x, y) = \mathrm{arcosh}\big(1 + \frac{2\|x-y\|^2}{(1-\|x\|^2)(1-\|y\|^2)}\big)$. The conformal factor $\lambda_x = 2/(1 - \|x\|^2)$ rescales the Euclidean inner product: $g_x(u, v) = \lambda_x^2\langle u, v\rangle$. Exponential and logarithm maps follow the Möbius gyrovector formalism [Ganea et al., 2018]. Parallel transport uses the gyroisometry: $\Gamma_{x\to y}(v) = \frac{\lambda_x}{\lambda_y}\,\mathrm{gyr}[y, -x]\, v$. Sectional curvature $\kappa = -1$ and $\mathrm{inj}(\mathbb{H}^d) = \infty$ (Hadamard manifold). The experiments generate data in the ball and fit in the isometric Lorentz model (Appendix H.1).

Wasserstein space $\mathcal{P}_2(\mathbb{R})$. Probability measures on $\mathbb{R}$ with finite second moment under the W_2 metric. Via the quantile-function map $\mu \mapsto F_\mu^{-1} : [0, 1] \to \mathbb{R}$, $\mathcal{P}_2(\mathbb{R})$ is isometric to the convex cone of non-decreasing functions in $L^2([0, 1])$ and inherits its flat L^2 geometry [Petersen and Müller, 2019b]. $d_{\mathcal{M}}(\mu, \nu) = \|F_\mu^{-1} - F_\nu^{-1}\|_{L^2([0,1])}$; $\mathrm{Log}_\mu(\nu) = F_\nu^{-1} - F_\mu^{-1}$ (in the quantile representation); $\mathrm{Exp}_\mu(\xi) = (F_\mu^{-1} + \xi)_{\#}\mathrm{Unif}[0, 1]$ when $F_\mu^{-1} + \xi$ is non-decreasing. Parallel transport is the identity in the quantile representation: $\Gamma_{\mu\to\nu}(\xi) = \xi$, since the embedded cone is flat in $L^2([0, 1])$. Sectional curvature $\kappa = 0$ and $\mathrm{inj}(\mathcal{P}_2(\mathbb{R})) = \infty$.

G Fixed-Base Residual Bootstrap (Oracle-Variance Diagnostic)

The nonparametric pairs bootstrap after Theorem 4.5 estimates the full covariance $\Sigma_{\boldsymbol{\beta}}$. Each replicate resamples whole observations and recomputes the Fréchet means, transports, and ROLS coefficients, so the covariance includes the cross term between the tangent-space score and the Fréchet-mean influence functions. This appendix records a simpler pivotal bootstrap that conditions on the empirical Fréchet mean $\hat{\bar{Y}}_n$ and the transported tangent design, resampling only the tangent-space residuals. By construction this fixed-base scheme targets the oracle tangent-space component $\Sigma_{\tan}$ of the asymptotic variance.

Algorithm 1 Pivotal Residual Bootstrap for ROLS

Require: Original data $\{(Y_i, X_{i1}, \ldots, X_{ip}, z_{i1}, \ldots, z_{iq})\}_{i=1}^n$, ROLS fit $(\hat{\boldsymbol{\beta}}, \hat{w}_i, \hat{\varepsilon}_i)$ at the empirical Fréchet mean $\hat{\bar{Y}}_n$, tangent design $(u_{ij}, \tilde{z}_{ik})$, number of bootstrap replicates B

1: **for** $b = 1, \ldots, B$ **do**

2: Draw $\tilde{\varepsilon}_i^{*(b)} \overset{\text{iid}}{\sim} \text{Uniform}\{\hat{\varepsilon}_1, \ldots, \hat{\varepsilon}_n\}$ and explicitly center: $\varepsilon_i^{*(b)} = \tilde{\varepsilon}_i^{*(b)} - n^{-1}\sum_{i'=1}^n \tilde{\varepsilon}_{i'}^{*(b)}$

3: Construct bootstrap responses in $T_{\hat{\bar{Y}}_n}\mathcal{M}$: $w_i^{*(b)} = \hat{w}_i + \varepsilon_i^{*(b)}$

4: Refit the tangent-space normal equations with $w_i^{*(b)}$ as responses and the same transported predictors u_{ij} and centered covariates $\tilde{z}_{ik}$ to obtain $\hat{\boldsymbol{\beta}}^{*(b)} = (\hat{\alpha}_j^{*(b)}, \hat{v}_k^{*(b)})$ and residual scale $\hat{\sigma}^{*(b)}$

5: Compute the bootstrap pivotal statistic:

$$T_j^{*(b)} = \frac{\hat{\alpha}_j^{*(b)} - \hat{\alpha}_j}{\widehat{\text{se}}(\hat{\alpha}_j^{*(b)})} \quad \text{or} \quad W_k^{*(b)} = n\,(\hat{v}_k^{*(b)} - \hat{v}_k)^\top (\hat{\Sigma}_{\boldsymbol{\beta}}^{kk,*(b)})^{-1} (\hat{v}_k^{*(b)} - \hat{v}_k)$$

6: **end for**

7: Compute the bootstrap p-value: for the t-test, $p_j^{\text{boot}} = \big(1 + \sum_{b=1}^B \mathbf{1}\{|T_j^{*(b)}| \geq |T_j|\}\big)/(B+1)$, where $T_j = \hat{\alpha}_j/\widehat{\text{se}}(\hat{\alpha}_j)$ is the original test statistic. Analogously for the Wald test using $W_k^{*(b)} \geq W_k$. Both T_j and $T_j^{*(b)}$ (and W_k, $W_k^{*(b)}$) use the model-based covariance of their own fit.

8: *Pairs variant.* Replace lines 2–4 by resampling whole observations with replacement and refitting the Fréchet means, transports, and coefficients; parallel-transport $\hat{v}_k^{*(b)}$ from $T_{\hat{\bar{Y}}^{*(b)}}\mathcal{M}$ to $T_{\hat{\bar{Y}}_n}\mathcal{M}$ before forming $\hat{v}_k^{*(b)} - \hat{v}_k$.

Under (D2) of Appendix E, the oracle tangent score is uncorrelated with the Fréchet-mean influence functions, and the variance splits as $\Sigma_{\boldsymbol{\beta}} = \Sigma_{\tan} + \Sigma_{\text{curv}}$. Under (D1) the tangent noise is isotropic, and $\Sigma_{\tan}$ is proportional to $\Sigma_{\boldsymbol{ZZ}}^{-1}$. Under (D3) the Fréchet-mean covariance is block diagonal, and Σ_{curv} is a sum over the $p+1$ means. On a flat manifold $\Sigma_{\text{curv}} = 0$. Otherwise Σ_{curv} is controlled by the curvature and the Afsari radius.

Suppose we wish to test $H_0 \colon \alpha_j = 0$ (or $H_0 \colon v_k = 0$). We have the ROLS fit on the original data: estimated parameters $\hat{\boldsymbol{\beta}} = (\hat{\alpha}_1, \ldots, \hat{\alpha}_p, \hat{v}_1, \ldots, \hat{v}_q)$, fitted values $\hat{w}_i = \sum_j \hat{\alpha}_j u_{ij} + \sum_k \tilde{z}_{ik}\hat{v}_k$, and residuals $\hat{\varepsilon}_i = w_i - \hat{w}_i \in T_{\hat{\bar{Y}}}\mathcal{M}$.

The pivotal construction recomputes both the numerator and denominator of $T_j^{*(b)}$ on the bootstrap sample, so the statistic is pivotal and its distribution does not depend on σ^2.

Conditional on $\hat{\bar{Y}}_n$ and the transported tangent design, and under isotropic homoskedastic noise, the conditional law of $\sqrt{n}(\hat{\boldsymbol{\beta}}^{*(b)} - \hat{\boldsymbol{\beta}})$ converges to $\mathcal{N}(0, \Sigma_{\tan})$, and the conditional distribution of $T_j^{*(b)}$ converges in probability to $N(0,1)$. The argument keeps the Fréchet-mean terms $J_m H_m^{-1}\psi_m$ in (48) fixed. Proposition E.1 treats the pairs bootstrap, which recomputes those terms.

Condition throughout on the empirical Fréchet mean $\hat{\bar{Y}}_n$, which is held fixed across all bootstrap replicates in Algorithm 1. Given $\hat{\bar{Y}}_n$, the tangent responses $w_i \in T_{\hat{\bar{Y}}_n}\mathcal{M}$, transported predictors

u_{ij}, and centered covariates $\tilde{z}_{ik}$ all live in the fixed Euclidean space $T_{\hat{\bar{Y}}_n}\mathcal{M} \cong \mathbb{R}^d$, and the ROLS normal equations reduce to an ordinary least squares system for $\hat{\boldsymbol{\beta}} = (\hat{\alpha}_1, \ldots, \hat{\alpha}_p, \hat{v}_1^\top, \ldots, \hat{v}_q^\top)^\top$ with design matrix $\mathbf{X} \in \mathbb{R}^{nd \times (p+qd)}$ whose rows encode $(u_{ij}, \tilde{z}_{ik} e_\ell)$. The resampled responses $w_i^{*(b)} = \hat{w}_i + \varepsilon_i^{*(b)}$ are generated by Efron's residual bootstrap applied to this Euclidean system with fixed design $\mathbf{X}$. Algorithm 1 recenters the resampled residuals. The coefficient $\hat{\boldsymbol{\beta}}^{*(b)}$ is unchanged, because $\sum_i \boldsymbol{Z}_i = 0$ at the empirical Fréchet means with centered covariates. Freedman [1981, Theorem 2.2, p. 1223] then applies directly: under the regression-model assumptions (1.1)–(1.4) of that paper (fixed design, finite error variance, and $n^{-1}\mathbf{X}^\top\mathbf{X} \to V$ positive definite, verified here by the full-rank and moment conditions underlying Theorem 4.5), along almost all sample sequences,

(a) the conditional law of $\sqrt{n}(\hat{\boldsymbol{\beta}}^{*(b)} - \hat{\boldsymbol{\beta}})$ given $\{(Y_i, X_{i1}, \ldots, z_{iq})\}_{i=1}^n$ converges weakly to $N(0, \sigma^2 V^{-1})$;

(b) the bootstrap variance estimator satisfies $\hat{\sigma}^{*(b)} \to \sigma$ in conditional probability; and

(c) consequently, the conditional law of the pivotal quantity $(\mathbf{X}^\top\mathbf{X})^{1/2}(\hat{\boldsymbol{\beta}}^{*(b)} - \hat{\boldsymbol{\beta}})/\hat{\sigma}^{*(b)}$ converges to the standard normal distribution on $\mathbb{R}^{p+qd}$.

Applying (c) to the j-th coordinate yields $\sup_t |P^*(T_j^{*(b)} \le t \mid \text{data}) - \Phi(t)| \xrightarrow{P} 0$, where Φ is the standard normal distribution function. The same continuous-mapping argument gives the Wald pivot $W_k^{*(b)}$. Mammen [1993] treats the corresponding Euclidean bootstrap when the dimension grows with n.

The Fréchet mean is held fixed, so this conditional variance is $\Sigma_{\tan}$. The term $\Sigma_{\text{curv}} = J_\theta \Sigma_\theta J_\theta^\top$ comes from the $O_P(n^{-1/2})$ fluctuation of $\hat{\bar{Y}}_n$ [Bhattacharya and Lin, 2017] and appears when the mean is recomputed, as in the pairs bootstrap. On a flat manifold, Appendix E gives $\Sigma_{\text{curv}} = 0$, and the residual bootstrap is then consistent for $\Sigma_{\boldsymbol{\beta}}$.

Each bootstrap replicate requires one ROLS fit, which is $O(n(p+qd)^2)$ (dominated by assembling the Gram matrix). The Riemannian operations (log, exp, parallel transport) are shared with the original fit. The total cost is therefore B times that of one refit.

H Additional Simulation Results

This appendix provides the full numerical simulation results and details for the theorem-verification experiments summarized in Section 5.

H.1 Data-Generating Process and Baseline Implementations

All synthetic experiments use Model (1) with manifold predictors $X_{i1}, \ldots, X_{ip}$ sampled near per-manifold centers $\bar{X}_j$, scalar covariates $z_{ik} \overset{\text{iid}}{\sim} \mathcal{N}(0,1)$, response base point $\bar{Y}$, and additive tangent noise of scale τ.

Coefficients. The manifold-predictor coefficients are taken from the fixed pool $(0.5, -0.3, 0.2, -0.15, 0.1, -0.08, 0.06, -0.04)$, retaining the first p entries. For each scalar predictor k, the tangent-vector coefficient $v_k \in T_{\bar{Y}}\mathcal{M}$ is drawn as an isotropic Gaussian direction projected to $T_{\bar{Y}}\mathcal{M}$ and rescaled to $\|v_k\| = 0.2$; on $\text{Sym}^+(3)$, the random vector is first symmetrized and then rescaled in the Frobenius norm.

Per-manifold base points and predictor spreads. Table 4 lists $\bar{Y}$, the predictor centers $\bar{X}_1, \ldots, \bar{X}_4$ used cyclically when $p > 4$, and the geodesic spread r_X that controls how far the X_{ij} deviate from $\bar{X}_j$. Sampling at $\bar{X}_j$ uses a centered, scaled tangent draw followed by $\text{Exp}_{\bar{X}_j}$, with a 0.95-norm projection on $\mathbb{H}^3$ to keep samples in the open ball. In every experiment, the $\mathbb{H}^3$ samples are mapped isometrically to the Lorentz model and fitted there with closed-form logarithm, exponential, and parallel-transport maps. The Fréchet mean is computed by the damped fixed-point iteration $\mu \leftarrow \text{Exp}_\mu\big(n^{-1}\sum_i \text{Log}_\mu(Y_i)\big)$ until the gradient reaches machine precision, and the error of $\hat{v}$ is measured in orthonormal frames after parallel transport.

Table 4: Synthetic-experiment fixtures. Predictor centers $\bar{X}_1, \ldots, \bar{X}_4$ are reused cyclically when $p > 4$. r_X is the predictor sampling spread (in radians for $\mathbb{S}^2$, $\mathrm{SO}(3)$; in coordinate units for $\mathbb{T}^2$, $\mathbb{H}^3$; and in the $\mathrm{Sym}^+(3)$ tangent norm at $\bar{X}_j$).

Manifold	$\bar{Y}$	r_X	Predictor centers $\bar{X}_1, \ldots, \bar{X}_4$
$\mathbb{S}^2$	$(0, 0, 1)$	0.6	$e_1,\ e_2,\ \frac{1}{\sqrt{2}}(e_1+e_2),\ \frac{1}{\sqrt{2}}(-e_1+e_2)$
$\mathbb{T}^2$	$(5.0, 1.2)$	0.45	$(4.8, 1.0),\ (5.4, 1.6),\ (4.4, 1.8),\ (5.8, 0.8)$
$\mathrm{SO}(3)$	I_3	0.5	$\exp(\hat{\omega}_j)$ for $\omega_j \in \{(0.5, 0.3, 0.2),\ (-0.4, 0.6, -0.1),\ (0.2, -0.5, 0.4),\ (-0.3, -0.2, 0.7)\}$
$\mathrm{Sym}^+(3)$	$1.5\, I_3$	0.3	$\mathrm{Exp}_{c_j I_3}(\frac{1}{2}(A_j + A_j^\top))$ for $c_j \in \{1.2, 0.8, 1.0, 1.4\},\ A_j \sim 0.1\,\mathcal{N}(0, I_{3\times 3})$
$\mathbb{H}^3$	$(0.2, 0, 0)$	0.15	$0.3\, d_j$ for $d_j \in \{(0.3, 0.1, 0),\ (-0.1, 0.3, 0),\ (0, -0.1, 0.3),\ (0.2, 0.2, -0.1)\}$

Tangent noise. The additive tangent noise ε_i is built from independent $\mathcal{N}(0, \tau^2)$ entries. On $\mathbb{S}^2$ and $\mathbb{T}^2$ it has intrinsic dimension 2 and is expressed in an orthonormal frame at $\bar{Y}$. On $\mathrm{SO}(3)$ and $\mathbb{H}^3$ it has intrinsic dimension 3 and is expressed in the standard frame, with $\mathrm{SO}(3)$ noise realized as a skew matrix and $\mathbb{H}^3$ noise applied in ambient Poincaré coordinates so that the Riemannian magnitude scales with the conformal factor $\lambda_{\bar{Y}} = 2/(1 - \|\bar{Y}\|^2) \approx 2.083$. On $\mathrm{Sym}^+(3)$ a 3×3 Gaussian matrix A with iid $\mathcal{N}(0, \tau^2)$ entries is symmetrized as $\frac{1}{2}(A + A^\top)$, giving a 6-dimensional symmetric tangent draw. In every case the linear predictor $m_i := m_{\boldsymbol{\beta}}(X_i, z_i)$ is formed in $T_{\bar{Y}}\mathcal{M}$ and the response is $Y_i = \mathrm{Exp}_{\bar{Y}}(m_i + \varepsilon_i)$.

Configurations and replications. The default configuration is $(n, p, q, \tau) = (200, 2, 3, 0.15)$; the prediction-accuracy table sweeps $\tau \in \{0.05, 0.15, 0.30\}$, $q \in \{0, 3, 6\}$, $n \in \{100, 200, 500\}$, and $p \in \{1, 2, 4\}$, giving 9 cells per manifold (45 in total) with 100 Monte Carlo replications per cell and $n = 100$ held-out test observations.

Baseline implementations. All baselines are evaluated on the same train/test splits as ROLS. The six baselines used in Fig. 2(d) and Table 5 are as follows. Fréchet mean predicts the empirical Karcher mean of $\{Y_i\}$. Geodesic regression fits a multi-scalar tangent-linear model in $T_{\bar{Y}}\mathcal{M}$, the p=0, $q \geq 1$ special case of ROLS (the original single-scalar geodesic regression of Fletcher [2013] extends to multiple scalars by stacking tangent-vector coefficients v_k). Fréchet regression [Petersen and Müller, 2019a] predicts the weighted Fréchet mean $\arg\min_y \sum_i w_i(z)\, d^2(Y_i, y)$ with the global linear weights $w_i(z) = n^{-1}\{1 + (z_i - \bar{z})^\top \hat{S}_z^{-1}(z - \bar{z})\}$, where $\hat{S}_z$ is the sample covariance of the scalar covariates and the weights may be negative. Single-index Fréchet (SI-Fr.) [Bhattacharjee and Müller, 2023] searches 500 random unit directions θ, selects a bandwidth for each by 5-fold cross-validation, and keeps the direction with the smallest in-sample loss. Its prediction is the local linear weighted Fréchet mean on the index $z^\top\theta$ with a Gaussian kernel. We compute the loss on the whole sample, which Bhattacharjee and Müller [2023] allow in place of their binned loss, and we evaluate the loss and the cross-validation error with tangent residuals at $\bar{Y}$ to keep the search feasible. NL-Fréchet (NL-Fr.) [Bhattacharjee et al., 2025] predicts the weighted Fréchet mean with weights $w(x) = n^{-1}\mathbf{1} + G(G + \epsilon I)^{-1}Q(K + \epsilon I)^{-1}Q\, k(x)$, where K is the kernel matrix, $k(x)$ is the vector of kernel values at x, $Q = I - n^{-1}\mathbf{1}\mathbf{1}^\top$, and $G = QKQ$. The kernel is the Gaussian radial basis function $\exp(-\gamma\|f - f'\|^2)$ on the features f (manifold-predictor logarithms at each $\hat{X}_j$, stacked with z), with $\gamma = (2\sigma^2)^{-1}$ and σ^2 the mean squared pairwise distance. We use these features because Bhattacharjee et al. [2025] require a positive definite kernel, and the Gaussian kernel of the geodesic distance has negative eigenvalues on $\mathbb{S}^2$ and $\mathrm{SO}(3)$ in our simulations. The constant ϵ minimizes their generalized cross-validation criterion over $\{10^{-6}, \ldots, 10^2\}$, which extends their grid above 10^{-1} because the criterion often has its minimum there. All three weighted Fréchet means are computed by gradient descent, started from the best, by objective value, of the weighted tangent average and 16 sampled responses. ROLS (m) is the manifold-only ablation that drops the scalar block. Separately, an ROLS ablation called tangent linear ridge replaces each scalar coefficient α_j with a full $d \times d$ matrix A_j in $T_{\bar{Y}}\mathcal{M}$ and adds a ridge penalty of strength 10^{-3}, sharing ROLS's tangent design otherwise; its RMSE is reported inline in Section 5 rather than included in the figures or table.

H.2 Prediction Accuracy Across Manifolds

Table 5 reports Fréchet RMSE on $\mathbb{S}^2$, $\mathrm{SO}(3)$, $\mathrm{Sym}^+(3)$, $\mathbb{H}^3$, and the flat torus $\mathbb{T}^2$ under varying noise level τ, scalar-covariate count q, sample size n, and manifold-predictor count p, with 100 Monte Carlo replications per cell. The default configuration is n=200, p=2, q=3, τ=0.15. ROLS is compared against the same six baselines as in Fig. 2(d) (Fréchet mean, geodesic regression, Fréchet regression, single-index Fréchet (SI-Fr.), nonlinear Fréchet (NL-Fr.), and the manifold-only ablation ROLS (m)). Figure 4 shows the same grid, with panel (a) the default cell from Fig. 2(d) and panels (b)–(e) the sweeps of τ, q, n, and p on $\mathbb{S}^2$.

Table 5: Full simulation grid: Fréchet RMSE (mean$_{\pm\text{std}}$, 100 trials) under varying τ, q, n, and p on $\mathbb{S}^2$, $\mathrm{SO}(3)$, $\mathrm{Sym}^+(3)$, $\mathbb{H}^3$, and the flat torus $\mathbb{T}^2$ (rightmost block). Default: n=200, p=2, q=3, τ=0.15. SI-Fr. = single-index Fréchet regression; NL-Fr. = nonlinear Fréchet regression. **Bold**: best per column within each three-column group.

	$\mathbb{S}^2$			$\mathrm{SO}(3)$			$\mathrm{Sym}^+(3)$			$\mathbb{H}^3$			$\mathbb{T}^2$		
τ	.05	.15	.30	.05	.15	.30	.05	.15	.30	.05	.15	.30	.05	.15	.30
Fr. mean	$.49_{\pm.03}$	$.53_{\pm.03}$	$.64_{\pm.04}$	$.62_{\pm.03}$	$.66_{\pm.03}$	$.80_{\pm.03}$	$.51_{\pm.02}$	$.56_{\pm.02}$	$.71_{\pm.02}$	$.76_{\pm.04}$	$.91_{\pm.04}$	$1.31_{\pm.06}$	$.51_{\pm.03}$	$.55_{\pm.03}$	$.66_{\pm.03}$
Geodesic	$.35_{\pm.02}$	$.41_{\pm.02}$	$.55_{\pm.03}$	$.51_{\pm.02}$	$.57_{\pm.02}$	$.73_{\pm.03}$	$.43_{\pm.01}$	$.49_{\pm.01}$	$.65_{\pm.02}$	$.27_{\pm.01}$	$.61_{\pm.03}$	$1.17_{\pm.06}$	$.38_{\pm.02}$	$.43_{\pm.02}$	$.57_{\pm.03}$
Fr. reg	$.35_{\pm.02}$	$.41_{\pm.02}$	$.55_{\pm.03}$	$.51_{\pm.02}$	$.57_{\pm.02}$	$.73_{\pm.03}$	$.43_{\pm.01}$	$.49_{\pm.01}$	$.65_{\pm.02}$	$.27_{\pm.01}$	$.61_{\pm.03}$	$\mathbf{1.15}_{\pm.05}$	$.38_{\pm.02}$	$.43_{\pm.02}$	$.57_{\pm.03}$
SI-Fr.	$.37_{\pm.03}$	$.43_{\pm.03}$	$.57_{\pm.04}$	$.53_{\pm.03}$	$.59_{\pm.03}$	$.75_{\pm.05}$	$.48_{\pm.01}$	$.53_{\pm.02}$	$.69_{\pm.04}$	$.38_{\pm.02}$	$.66_{\pm.04}$	$1.18_{\pm.06}$	$.44_{\pm.04}$	$.49_{\pm.04}$	$.63_{\pm.06}$
NL-Fr.	$.23_{\pm.03}$	$.32_{\pm.03}$	$.50_{\pm.04}$	$.26_{\pm.05}$	$.38_{\pm.03}$	$.62_{\pm.04}$	$.29_{\pm.07}$	$.42_{\pm.03}$	$.63_{\pm.03}$	$.38_{\pm.06}$	$.71_{\pm.05}$	$1.21_{\pm.05}$	$.19_{\pm.05}$	$.30_{\pm.04}$	$.51_{\pm.05}$
ROLS (m)	$.35_{\pm.02}$	$.40_{\pm.02}$	$.54_{\pm.03}$	$.35_{\pm.02}$	$.43_{\pm.02}$	$.62_{\pm.03}$	$.29_{\pm.01}$	$.37_{\pm.01}$	$.57_{\pm.02}$	$.74_{\pm.04}$	$.90_{\pm.05}$	$1.30_{\pm.06}$	$.35_{\pm.02}$	$.41_{\pm.02}$	$.55_{\pm.03}$
ROLS	$\mathbf{.07}_{\pm.00}$	$\mathbf{.21}_{\pm.01}$	$\mathbf{.42}_{\pm.02}$	$\mathbf{.09}_{\pm.00}$	$\mathbf{.26}_{\pm.01}$	$\mathbf{.52}_{\pm.02}$	$\mathbf{.08}_{\pm.00}$	$\mathbf{.25}_{\pm.01}$	$\mathbf{.50}_{\pm.01}$	$\mathbf{.20}_{\pm.01}$	$\mathbf{.59}_{\pm.03}$	$1.16_{\pm.05}$	$\mathbf{.07}_{\pm.00}$	$\mathbf{.21}_{\pm.01}$	$\mathbf{.43}_{\pm.02}$
q	0	3	6	0	3	6	0	3	6	0	3	6	0	3	6
Fr. mean	$.41_{\pm.02}$	$.53_{\pm.03}$	$.64_{\pm.03}$	$.57_{\pm.03}$	$.66_{\pm.03}$	$.75_{\pm.03}$	$.49_{\pm.01}$	$.56_{\pm.02}$	$.63_{\pm.02}$	$.57_{\pm.02}$	$.91_{\pm.04}$	$1.17_{\pm.05}$	$.43_{\pm.02}$	$.55_{\pm.03}$	$.65_{\pm.04}$
Geodesic	$.41_{\pm.02}$	$.41_{\pm.02}$	$.41_{\pm.02}$	$.57_{\pm.03}$	$.57_{\pm.02}$	$.57_{\pm.03}$	$.49_{\pm.01}$	$.49_{\pm.01}$	$.50_{\pm.01}$	$.57_{\pm.02}$	$.61_{\pm.03}$	$.66_{\pm.03}$	$.43_{\pm.02}$	$.43_{\pm.02}$	$.43_{\pm.02}$
Fr. reg	$.41_{\pm.02}$	$.41_{\pm.02}$	$.41_{\pm.02}$	$.57_{\pm.03}$	$.57_{\pm.02}$	$.57_{\pm.03}$	$.49_{\pm.01}$	$.49_{\pm.01}$	$.50_{\pm.01}$	$.57_{\pm.02}$	$.61_{\pm.03}$	$.66_{\pm.03}$	$.43_{\pm.02}$	$.43_{\pm.02}$	$.43_{\pm.02}$
SI-Fr.	$.41_{\pm.02}$	$.43_{\pm.03}$	$.50_{\pm.03}$	$.57_{\pm.03}$	$.59_{\pm.03}$	$.65_{\pm.03}$	$.49_{\pm.01}$	$.53_{\pm.02}$	$.58_{\pm.02}$	$.57_{\pm.02}$	$.66_{\pm.04}$	$.91_{\pm.05}$	$.43_{\pm.02}$	$.49_{\pm.04}$	$.53_{\pm.03}$
NL-Fr.	$.27_{\pm.03}$	$.32_{\pm.03}$	$.38_{\pm.04}$	$.33_{\pm.03}$	$.38_{\pm.03}$	$.45_{\pm.05}$	$.32_{\pm.02}$	$.42_{\pm.03}$	$.47_{\pm.03}$	$.57_{\pm.02}$	$.71_{\pm.05}$	$.84_{\pm.11}$	$.26_{\pm.03}$	$.30_{\pm.04}$	$.36_{\pm.05}$
ROLS (m)	$\mathbf{.21}_{\pm.01}$	$.40_{\pm.02}$	$.53_{\pm.03}$	$\mathbf{.26}_{\pm.01}$	$.43_{\pm.02}$	$.55_{\pm.03}$	$\mathbf{.25}_{\pm.01}$	$.37_{\pm.01}$	$.47_{\pm.02}$	$\mathbf{.55}_{\pm.02}$	$.90_{\pm.05}$	$1.16_{\pm.05}$	$\mathbf{.21}_{\pm.01}$	$.41_{\pm.02}$	$.53_{\pm.03}$
ROLS	$.21_{\pm.01}$	$\mathbf{.21}_{\pm.01}$	$\mathbf{.21}_{\pm.01}$	$.26_{\pm.01}$	$\mathbf{.26}_{\pm.01}$	$\mathbf{.26}_{\pm.01}$	$.25_{\pm.01}$	$\mathbf{.25}_{\pm.01}$	$\mathbf{.25}_{\pm.01}$	$.55_{\pm.02}$	$\mathbf{.59}_{\pm.03}$	$\mathbf{.63}_{\pm.03}$	$.21_{\pm.01}$	$\mathbf{.21}_{\pm.01}$	$\mathbf{.22}_{\pm.01}$
n	100	200	500	100	200	500	100	200	500	100	200	500	100	200	500
Fr. mean	$.53_{\pm.03}$	$.53_{\pm.03}$	$.53_{\pm.03}$	$.67_{\pm.03}$	$.66_{\pm.03}$	$.66_{\pm.03}$	$.56_{\pm.02}$	$.56_{\pm.02}$	$.56_{\pm.02}$	$.91_{\pm.04}$	$.91_{\pm.04}$	$.91_{\pm.04}$	$.55_{\pm.03}$	$.55_{\pm.03}$	$.55_{\pm.03}$
Geodesic	$.41_{\pm.02}$	$.41_{\pm.02}$	$.40_{\pm.02}$	$.58_{\pm.03}$	$.57_{\pm.02}$	$.57_{\pm.02}$	$.50_{\pm.01}$	$.49_{\pm.01}$	$.49_{\pm.01}$	$.62_{\pm.03}$	$.61_{\pm.03}$	$.61_{\pm.03}$	$.44_{\pm.02}$	$.43_{\pm.02}$	$.43_{\pm.02}$
Fr. reg	$.41_{\pm.02}$	$.41_{\pm.02}$	$.40_{\pm.02}$	$.58_{\pm.03}$	$.57_{\pm.02}$	$.57_{\pm.02}$	$.50_{\pm.01}$	$.49_{\pm.01}$	$.49_{\pm.01}$	$.62_{\pm.03}$	$.61_{\pm.03}$	$.61_{\pm.03}$	$.44_{\pm.02}$	$.43_{\pm.02}$	$.43_{\pm.02}$
SI-Fr.	$.47_{\pm.08}$	$.43_{\pm.03}$	$.42_{\pm.03}$	$.61_{\pm.05}$	$.59_{\pm.03}$	$.58_{\pm.02}$	$.54_{\pm.02}$	$.53_{\pm.02}$	$.53_{\pm.01}$	$.68_{\pm.06}$	$.66_{\pm.04}$	$.66_{\pm.03}$	$.54_{\pm.09}$	$.49_{\pm.04}$	$.47_{\pm.02}$
NL-Fr.	$.35_{\pm.04}$	$.32_{\pm.03}$	$.29_{\pm.03}$	$.44_{\pm.05}$	$.38_{\pm.03}$	$.34_{\pm.03}$	$.45_{\pm.04}$	$.42_{\pm.03}$	$.35_{\pm.02}$	$.69_{\pm.05}$	$.71_{\pm.05}$	$.66_{\pm.03}$	$.34_{\pm.05}$	$.30_{\pm.04}$	$.27_{\pm.03}$
ROLS (m)	$.40_{\pm.02}$	$.40_{\pm.02}$	$.40_{\pm.02}$	$.43_{\pm.02}$	$.43_{\pm.02}$	$.43_{\pm.02}$	$.37_{\pm.01}$	$.37_{\pm.01}$	$.37_{\pm.01}$	$.90_{\pm.04}$	$.90_{\pm.05}$	$.90_{\pm.04}$	$.41_{\pm.02}$	$.41_{\pm.02}$	$.40_{\pm.02}$
ROLS	$\mathbf{.21}_{\pm.01}$	$\mathbf{.21}_{\pm.01}$	$\mathbf{.21}_{\pm.01}$	$\mathbf{.26}_{\pm.01}$	$\mathbf{.26}_{\pm.01}$	$\mathbf{.26}_{\pm.01}$	$\mathbf{.25}_{\pm.01}$	$\mathbf{.25}_{\pm.01}$	$\mathbf{.25}_{\pm.01}$	$\mathbf{.59}_{\pm.03}$	$\mathbf{.59}_{\pm.03}$	$\mathbf{.58}_{\pm.03}$	$\mathbf{.22}_{\pm.01}$	$\mathbf{.21}_{\pm.01}$	$\mathbf{.21}_{\pm.01}$
p	1	2	4	1	2	4	1	2	4	1	2	4	1	2	4
Fr. mean	$.50_{\pm.03}$	$.53_{\pm.03}$	$.56_{\pm.03}$	$.61_{\pm.03}$	$.66_{\pm.03}$	$.70_{\pm.03}$	$.48_{\pm.02}$	$.56_{\pm.02}$	$.59_{\pm.02}$	$.91_{\pm.04}$	$.91_{\pm.04}$	$.92_{\pm.05}$	$.51_{\pm.02}$	$.55_{\pm.03}$	$.57_{\pm.03}$
Geodesic	$.36_{\pm.02}$	$.41_{\pm.02}$	$.44_{\pm.02}$	$.51_{\pm.02}$	$.57_{\pm.02}$	$.61_{\pm.02}$	$.39_{\pm.01}$	$.49_{\pm.01}$	$.52_{\pm.02}$	$.60_{\pm.03}$	$.61_{\pm.03}$	$.62_{\pm.03}$	$.38_{\pm.02}$	$.43_{\pm.02}$	$.46_{\pm.02}$
Fr. reg	$.36_{\pm.02}$	$.41_{\pm.02}$	$.44_{\pm.02}$	$.51_{\pm.02}$	$.57_{\pm.02}$	$.61_{\pm.02}$	$.39_{\pm.01}$	$.49_{\pm.01}$	$.52_{\pm.02}$	$.60_{\pm.03}$	$.61_{\pm.03}$	$.61_{\pm.03}$	$.38_{\pm.02}$	$.43_{\pm.02}$	$.46_{\pm.02}$
SI-Fr.	$.39_{\pm.04}$	$.43_{\pm.03}$	$.45_{\pm.02}$	$.53_{\pm.02}$	$.59_{\pm.03}$	$.63_{\pm.03}$	$.44_{\pm.02}$	$.53_{\pm.02}$	$.56_{\pm.02}$	$.65_{\pm.03}$	$.66_{\pm.04}$	$.67_{\pm.04}$	$.44_{\pm.04}$	$.49_{\pm.04}$	$.52_{\pm.04}$
NL-Fr.	$.29_{\pm.03}$	$.32_{\pm.03}$	$.34_{\pm.03}$	$.35_{\pm.03}$	$.38_{\pm.03}$	$.43_{\pm.03}$	$.36_{\pm.03}$	$.42_{\pm.03}$	$.45_{\pm.03}$	$.69_{\pm.05}$	$.71_{\pm.05}$	$.70_{\pm.05}$	$.28_{\pm.03}$	$.30_{\pm.04}$	$.33_{\pm.04}$
ROLS (m)	$.40_{\pm.02}$	$.40_{\pm.02}$	$.40_{\pm.02}$	$.43_{\pm.02}$	$.43_{\pm.02}$	$.43_{\pm.02}$	$.37_{\pm.01}$	$.37_{\pm.01}$	$.37_{\pm.01}$	$.90_{\pm.04}$	$.90_{\pm.05}$	$.90_{\pm.05}$	$.40_{\pm.02}$	$.41_{\pm.02}$	$.41_{\pm.02}$
ROLS	$\mathbf{.21}_{\pm.01}$	$\mathbf{.21}_{\pm.01}$	$\mathbf{.21}_{\pm.01}$	$\mathbf{.26}_{\pm.01}$	$\mathbf{.26}_{\pm.01}$	$\mathbf{.26}_{\pm.01}$	$\mathbf{.25}_{\pm.01}$	$\mathbf{.25}_{\pm.01}$	$\mathbf{.25}_{\pm.01}$	$\mathbf{.58}_{\pm.03}$	$\mathbf{.59}_{\pm.03}$	$\mathbf{.59}_{\pm.03}$	$\mathbf{.21}_{\pm.01}$	$\mathbf{.21}_{\pm.01}$	$\mathbf{.21}_{\pm.01}$

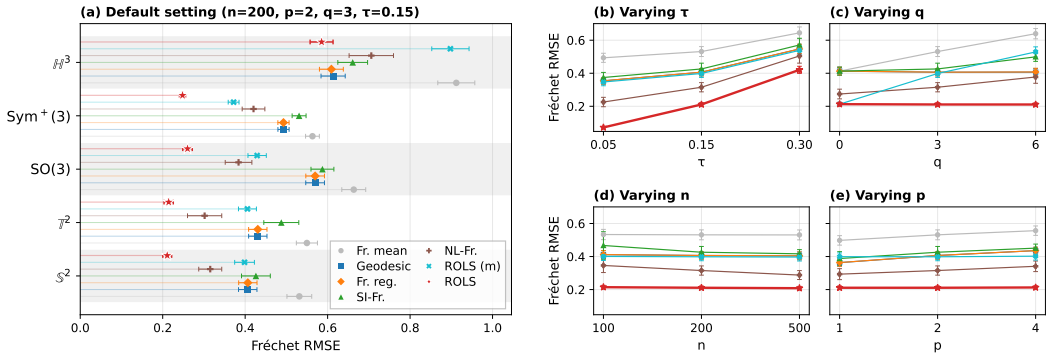


Figure 4: Predictive accuracy across configurations. (a) Default-cell Fréchet RMSE across the five response manifolds (same data as main-text Fig. 2(d), shown here for reference). (b)–(e) Sensitivity on $\mathbb{S}^2$ to noise level τ, scalar count q, sample size n, and manifold-predictor count p, sharing a unified y-axis. Full numerical entries are in Table 5.

H.3 Cut-Locus Sweep and Convergence Rate

On S^2 ($\kappa = 1$, $\mathrm{inj}(S^2) = \pi$), we probe the injectivity condition (5) by varying the data spread $r_0 \in [0.1, 3.0]$. Figure 5(a) shows that RMSE is stable at approximately 0.11 for $r_0 \leq 0.8$ and rises as r_0 approaches the Karcher convexity radius $\pi/2$ (where Afsari support fails and the squared distance ceases to be convex), well before the cut locus at π. This behavior illustrates Theorem 3.4.

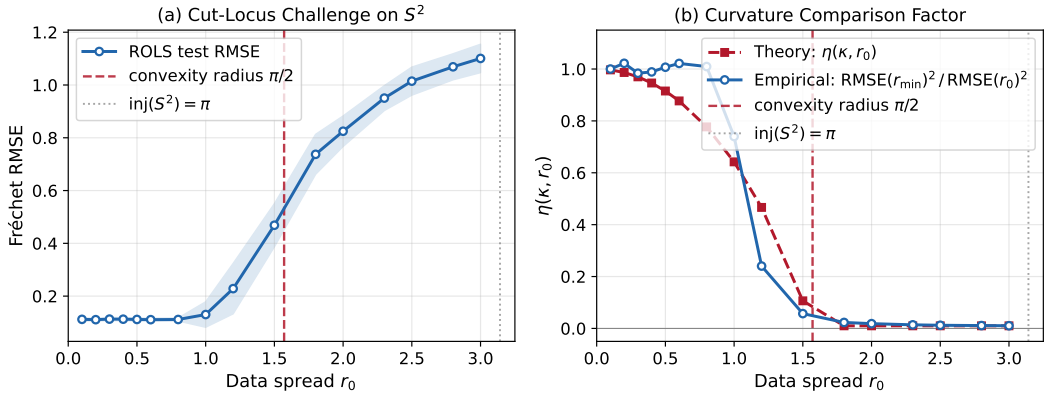


Figure 5: Cut-locus challenge on S^2 (Theorems 3.4 & 4.2). (a) ROLS test RMSE versus data spread r_0; the red dashed line marks the Karcher convexity radius $\pi/2$ and the gray dotted line marks $\mathrm{inj}(S^2) = \pi$. (b) Hessian comparison factor $\eta(1, r_0)$ from (7) (red) and empirical ratio $\mathrm{RMSE}(r_{\min})^2/\mathrm{RMSE}(r_0)^2$ (blue), with the same two reference lines, illustrating the degradation of the tangent-space approximation as data spread beyond the origin.

Panel (b) plots the Hessian comparison factor $\eta(\kappa_+, r_0)$ from (7) alongside the empirical ratio $\mathrm{RMSE}(r_{\min})^2/\mathrm{RMSE}(r_0)^2$. On S^2, $\kappa_- = \kappa_+ = 1$, so the factor $\tilde{\eta}$ in Theorem 4.2 equals one and does not grow with r_0. The rise in RMSE tracks the misspecification bias b^2 as the tangent-linear approximation picks up higher-order curvature terms. The empirical ratio stays near one for $r_0 \leq 0.8$ while η declines, falls below η by $r_0 = 1.2$, and both approach zero near $\pi/2$. The lower bound is $d^2_{\mathcal{M}} \geq \eta \, \|w - \hat{w}\|^2$. This bound controls how faithfully the tangent-space residual $\|w - \hat{w}\|$ represents the geodesic residual $d_{\mathcal{M}}(\hat{Y}, Y)$, and this relationship deteriorates as data spread further from the Fréchet mean.

Figure 6 plots parameter estimation errors versus n on a log–log scale for all five manifolds used elsewhere in the simulations, with training sample size n up to 2,000 (100 replications at each n). Table 6 reports per-manifold log–log slopes and goodness-of-fit (R^2).

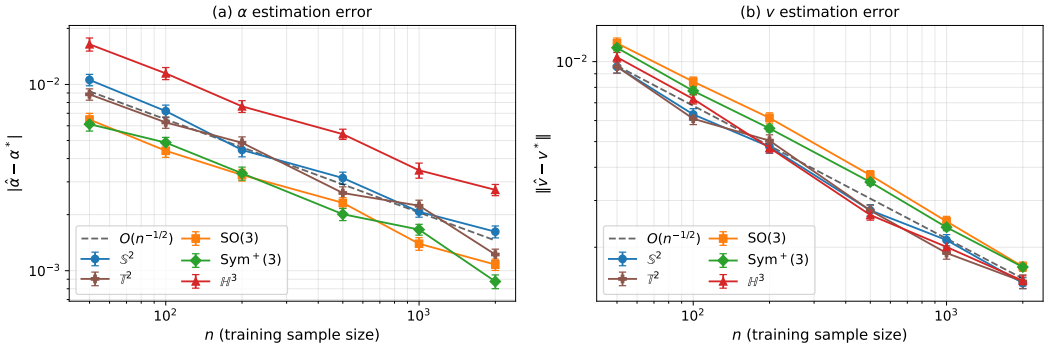


Figure 6: Convergence of ROLS parameter estimates as n grows (up to 2,000): (a) error $|\hat{\alpha} - \alpha^*|$; (b) error $\|\hat{v} - v^*\|$. Curves: $\mathbb{S}^2$, $\mathbb{T}^2$, SO(3), Sym$^+$(3), $\mathbb{H}^3$ (fitted in the exact Lorentz model, Appendix H.1). Dashed line: reference $O(n^{-1/2})$ slope in log–log space.

Table 6: Log–log regression slopes for parameter estimation errors vs. n on $\mathbb{S}^2$, the flat torus $\mathbb{T}^2$, SO(3), Sym$^+$(3), and $\mathbb{H}^3$ (100 trials at each n; $n \in \{50, 100, 200, 500, 1000, 2000\}$; $\mathbb{H}^3$ is fitted in the exact Lorentz model with a converged Fréchet mean, Appendix H.1). Theory predicts slope -0.50 for $\hat{\alpha}$ and $\hat{v}$ at this rate.

	$\mathbb{S}^2$	$\mathbb{T}^2$	SO(3)	Sym$^+$(3)	$\mathbb{H}^3$
$\hat{\alpha}$ slope	-0.51	-0.52	-0.49	-0.51	-0.49
$\hat{\alpha}$ R^2	0.99	0.98	0.99	0.98	1.00
$\hat{v}$ slope	-0.50	-0.51	-0.53	-0.51	-0.54
$\hat{v}$ R^2	1.00	0.99	1.00	1.00	0.99

On all five manifolds, both $\hat{\alpha}$ and $\hat{v}$ log–log slopes are close to -0.50 with $R^2 \geq 0.98$, in line with the $O(n^{-1/2})$ rate in Corollary 4.3.

H.4 Type I Error Calibration and Power

We evaluate the inference procedure of Theorem 4.5 via Monte Carlo on five manifolds (S^2, $\mathbb{T}^2$, SO(3), Sym$^+$(3), $\mathbb{H}^3$) with $n = 200$, $\tau = 0.15$, and 500 replications. Effect size is varied from 0 (null) to 0.5.

Figure 7 shows the power curves. Table 7 reports Type I error calibration.

Table 7: Type I error of the v-Wald test at nominal level 0.05 (300 trials, B=199 bootstrap replicates, $\tau = 0.5$). Asym.: asymptotic Wald test (Theorem 4.5) with the model-based covariance. Pairs-t: nonparametric pairs bootstrap with full Fréchet-mean refit on each replicate, with the observed and replicate statistics both studentized by the model-based covariance and aligned by parallel transport. Pairs-emp: the same bootstrap with the observed statistic studentized by the empirical bootstrap covariance. Resid.: fixed-base residual bootstrap of Appendix G, conditioning on the empirical Fréchet mean and resampling tangent-space residuals, targeting the oracle $\Sigma_{\tan}$ component.

	$n = 50$				$n = 100$				$n = 200$				$n = 500$			
Manifold	Asym.	Pairs-t	Pairs-emp	Resid.	Asym.	Pairs-t	Pairs-emp	Resid.	Asym.	Pairs-t	Pairs-emp	Resid.	Asym.	Pairs-t	Pairs-emp	Resid.
S^2	.063	.063	.097	.057	.047	.043	.057	.043	.053	.043	.063	.040	.060	.053	.063	.053
$\mathbb{T}^2$	.037	.027	.063	.023	.053	.053	.073	.043	.047	.057	.070	.050	.057	.043	.043	.047
$\mathbb{H}^3$	.053	.053	.097	.040	.057	.033	.060	.033	.027	.033	.053	.030	.060	.050	.067	.043
SO(3)	.063	.060	.123	.037	.070	.060	.087	.060	.063	.053	.067	.053	.057	.040	.053	.037
Sym$^+$(3)	.053	.040	.163	.033	.053	.043	.113	.040	.047	.037	.083	.040	.037	.033	.060	.033

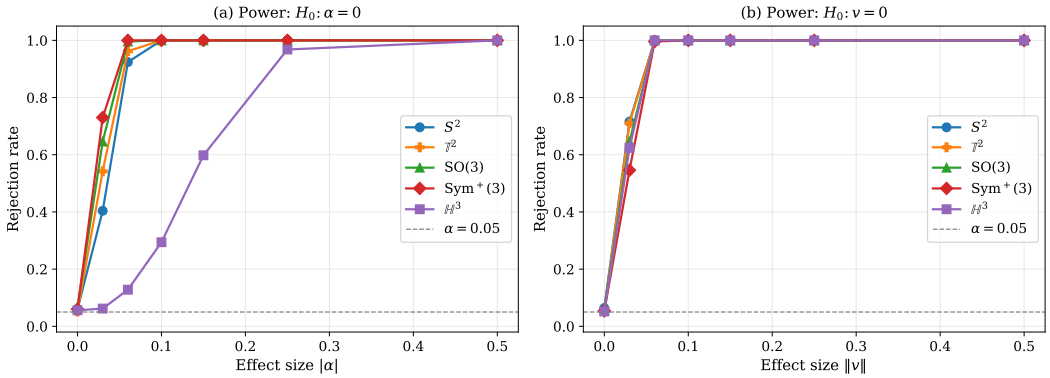


Figure 7: Power curves of ROLS hypothesis tests ($n = 200$, $\tau = 0.15$, 500 replications) on five manifolds. (a) t-test for $H_0 \colon \alpha = 0$. $\mathbb{H}^3$ shows reduced power because the Poincaré ball conformal factor inflates the effective Riemannian noise relative to the predictor variation. (b) Wald test for $H_0 \colon v = 0$, which converges quickly across all manifolds since the model is linear in v in the tangent space. Dashed line at nominal $\alpha = 0.05$.

In panel (a) of Figure 7, the t-test for $H_0 \colon \alpha = 0$ reaches near-saturation by $|\alpha| = 0.06$ on $\mathrm{SO}(3)$, $\mathrm{Sym}^+(3)$, and $\mathbb{T}^2$ ($\geq 96\%$ rejection), with S^2 at 92% and $\mathbb{H}^3$ trailing at 13%. $\mathbb{H}^3$ requires $|\alpha| \approx 0.25$ to attain 95% rejection, because the Poincaré conformal factor at the response base point inflates the effective Riemannian noise by roughly $2\times$ while simultaneously reducing the Riemannian variation of the predictor relative to its ambient spread; the resulting effective signal-to-noise ratio is about an order of magnitude smaller than on S^2. In panel (b), the Wald test for $H_0 \colon v = 0$ saturates by $\|v\| = 0.06$ on all five manifolds, since the model is linear in v in tangent coordinates and this coefficient is unaffected by the curvature-induced distortions of the manifold predictor. In Table 7, Pairs-t stays at or below 0.063, against the nominal 0.05, in all 20 settings. Pairs-emp studentizes the observed statistic with the empirical bootstrap covariance and each replicate with the model-based covariance. The resulting errors are largest on $\mathrm{Sym}^+(3)$, where $d = 6$, for $n \leq 200$. On S^2, $\mathbb{T}^2$, and $\mathrm{Sym}^+(3)$ the pairs-bootstrap standard error is 1.02, 1.00, and 0.97 times the fixed-base standard error ($n = 200$, $B = 200$, 60 trials). A companion check at $\tau = 0.15$, the noise level of Figure 7, uses 500 trials per cell. All 40 rejection rates of the asymptotic tests lie in $[0.026, 0.068]$, on all five manifolds, including $\mathbb{T}^2$. The residual bootstrap of Appendix G estimates $\Sigma_{\tan}$. On the flat torus $\Sigma_{\mathrm{curv}} = 0$, so that estimate is the full variance, in agreement with the $\mathbb{T}^2$ row of Table 7. Figure 8 summarizes the $\tau = 0.15$ companion check, plotting Type I rejection rate against n for each manifold and each null hypothesis.

H.5 Robustness Under Model Misspecification

To probe ROLS's sensitivity to the tangent-space linearity assumption, we generate data on S^2 via the nonlinear tangent-space model

$$\mathrm{Log}_{\bar{Y}}(Y_i) = \alpha\, u_i + z_i\, v + \gamma\, \|u_i\|^2\, \hat{n} + \varepsilon_i,$$

where $\hat{n}$ is a fixed tangent direction and $\gamma \in \{0, 0.1, 0.2, 0.3, 0.4, 0.5, 0.6\}$ controls the degree of model misspecification ($\gamma = 0$ is correctly specified). We compare ROLS, NL-Fréchet (kernel regression), and geodesic regression over 30 trials with $n = 300$. Figure 9 shows that ROLS has lower RMSE than both baselines at every value of γ, and its margin over the better baseline narrows as γ grows beyond 0.2.

In this design the omitted term contributes a squared bias of order γ^2, the misspecification term b^2 in Theorem 4.2.

Inference under misspecification. We rerun the design of Figure 9 with $\alpha = 0$, a second scalar covariate, $n = 300$, and 500 trials. Under $b > 0$, $\hat{\alpha}$ stays centered at its projection value 0, but its sampling standard deviation is about twice the model standard error, because the omitted quadratic

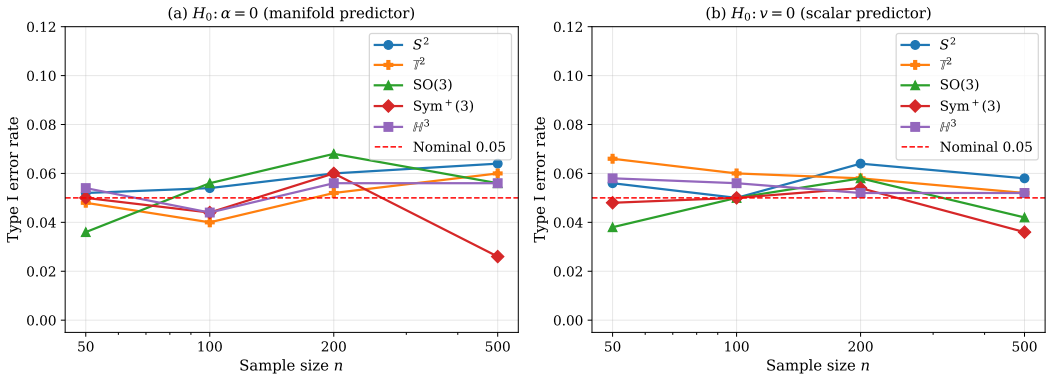


Figure 8: Type I error of the asymptotic ROLS tests at $\tau = 0.15$ across 500 trials per cell on five manifolds. (a) $H_0 \colon \alpha = 0$ (manifold predictor, t-test). (b) $H_0 \colon \nu = 0$ (scalar predictor, Wald test). Of the 40 entries, 39 lie within the binomial 95% envelope of nominal 0.05 (red dashed). The corresponding bootstrap-versus-asymptotic comparison at $\tau = 0.5$ is in Table 7.

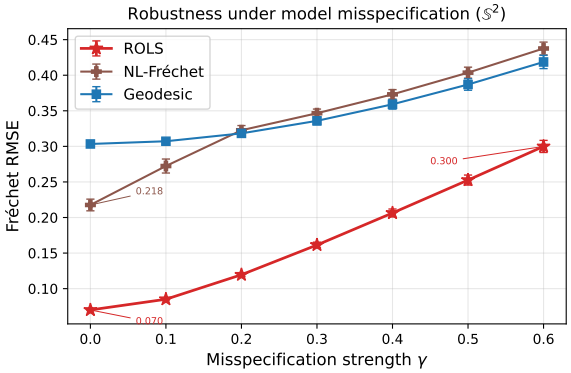


Figure 9: ROLS RMSE under increasing model misspecification (γ) on S^2. At $\gamma = 0$, ROLS has RMSE 0.070 and NL-Fréchet has RMSE 0.218. At $\gamma = 0.6$ the two RMSEs are 0.300 and 0.438.

term adds design-correlated variance. Table 8 reports the Type I error of the model-based Wald test and of the same test with a sandwich standard error, together with the rejection rate of the targeted lack-of-fit diagnostic. The diagnostic appends $\|u\|^2$, z_1^2, and z_2^2 to the tangent regression as additional scalar covariates and tests their coefficients jointly.

Table 8: Rejection rates at nominal level 0.05 under the misspecification of Figure 9 ($H_0 \colon \alpha = 0$, $n = 300$, 500 trials).

γ	0	0.1	0.2	0.3	0.4	0.5	0.6
Model Wald	.046	.092	.208	.266	.290	.330	.360
Robust α	.050	.038	.074	.076	.054	.062	.084
Lack-of-fit	.050	1.0	1.0	1.0	1.0	1.0	1.0

The leverage-corrected heteroskedasticity-consistent (HC3) version [MacKinnon and White, 1985] of the sandwich covariance of Section 4.2 replaces r_i in $\hat{\Omega}$ by $(I_d - H_{ii})^{-1} r_i$ with $H_{ii} = \boldsymbol{Z}_i^\top \hat{A}^{-1} \boldsymbol{Z}_i$. Table 9 applies this estimator to all blocks. The omitted term is centered, standardized, and placed either in u or in z_2 ($n = 200$, 1500 trials). The block targeted by the omitted term shows the dominant distortion, reaching 0.32 for α and 0.37 for v_2, while the untargeted block stays much closer to nominal. HC3 brings both targeted blocks back to 0.05–0.07.

Table 9: Type I error at nominal level 0.05 of the model-based and HC3 Wald tests when the omitted term involves u or z_2 ($n = 200$, 1500 trials).

Omitted term	Block	Test	$\gamma = 0$	0.1	0.2	0.3
in u	α	model	.043	.131	.252	.324
in u	α	HC3	.043	.063	.055	.069
in u	v_2	model	.052	.055	.065	.077
in z_2	v_2	model	.052	.168	.319	.369
in z_2	v_2	HC3	.051	.059	.066	.063
in z_2	α	model	.043	.043	.050	.053

The lack-of-fit diagnostic is targeted. Its basis matches the alternative of Figure 9, and its power against alternatives outside the basis is lower at the same standardized strength (Table 10).

Table 10: Rejection rate of the targeted lack-of-fit diagnostic at nominal level 0.05.

Alternative	In basis	$\gamma = 0$	0.1	0.2	0.3
Quadratic in u (Figure 9)	yes	.045	1.000	.999	.997
Cubic in z_1	no	.045	.228	.501	.675
$u \times z_1$ interaction	no	.045	.189	.353	.475

H.6 Sensitivity to the Base Point

On $\mathrm{Sym}^+(3)$ with the affine-invariant metric (200 trials), we refit the response and the predictor at four kinds of base point, namely the Fréchet mean, the Fréchet mean displaced by geodesic distance $\delta \cdot (\text{spread})$ in a random direction, the projected ambient mean, and a random observation. Table 11 reports the coefficient error $|\hat{\alpha} - \alpha^*|$ and the held-out Fréchet RMSE. The coefficient error stays near the Fréchet-mean value at $\delta = 0.1$ and rises once the base point moves by a quarter of the spread. In the expansion of Appendix E, the population coefficient $\boldsymbol{\beta}^*(\theta)$ depends on the base point through J_θ, so a displacement of that size changes the target of the estimator.

Table 11: Coefficient error and held-out Fréchet RMSE on $\mathrm{Sym}^+(3)$ under different base points (200 trials).

Base	Fréchet	$\delta = 0.1$	$\delta = 0.25$	$\delta = 0.5$	$\delta = 1.0$	Ambient	Random
Coeff. error	0.0153	0.0158	0.0373	0.1116	0.285	0.0217	0.256
RMSE	0.495	0.501	0.533	0.624	0.865	0.500	0.858

H.7 Higher Tangent Dimension

We grow the tangent dimension on $\mathrm{Sym}^+(k)$ up to $d = 55$ with one manifold predictor and one scalar covariate. At $d = 21$ the log–log slope of the parameter error is -0.495. The leading variance term of Theorem 4.2 is $C\sigma^2(p + qd)/(n\lambda_*)$, so the rate depends on $n/(p + qd)$. The geometric factor is $\tilde{\eta}^2$, and the rate is the same. Table 12 reports the Type I error of the v-Wald test. Calibration follows n/d, because the $(p + qd)$-dimensional Gram matrix is poorly estimated when n/d is small, as the χ^2 entry of the $d = 6$ row shows at the smallest sample size in that row, $n = 11$. The F reference brings that entry to .062 and changes the others by at most .006.

H.8 Computational Cost

ROLS evaluates the logarithm and transport maps only at the $p + 1$ estimated means, which gives $O(np)$ map evaluations, whereas kernel and Fréchet regression need $O(n^2)$ pairwise evaluations. A full refit on $\mathrm{Sym}^+(3)$ with $p = q = 1$ takes $0.002\,\mathrm{s}$ at $n = 100$ and $0.14\,\mathrm{s}$ at $n = 10^4$ (median over repeated single-threaded runs on one machine; timings depend on the implementation). A batch update requires a full refit, and there is no exact $O(1)$ streaming update, because a change in any of the $p + 1$ means changes the logarithm maps and transports of all earlier observations. The Fréchet

Table 12: Type I error of the v-Wald test on $\mathrm{Sym}^+(k)$ at nominal level 0.05 (500 trials, $n = \max\{\mathrm{round}(md),\ d+5\}$ for $m \in \{1.5, 3, 6, 12\}$), with the χ^2_d and F references.

d (χ^2 / F)	$n \approx 1.5d$	$n \approx 3d$	$n \approx 6d$	$n \approx 12d$
6	.086 / .062	.062 / .056	.064 / .062	.062 / .060
55	.060 / .058	.044 / .044	.040 / .040	.040 / .040

means themselves can be updated recursively with a Karcher estimator [Ho et al., 2013, Cheng et al., 2016], but an approximate online ROLS would need a separate analysis.

I Real-Data Case Study Details

This appendix provides complete descriptions of all three real-data case studies in Section 6.

For every dataset the ROLS model specializes the canonical form

$$\mathrm{Log}_{\bar{Y}}(Y_i) = \sum_{j=1}^{p} \alpha_j \, \Gamma_{\bar{X}_j \to \bar{Y}}\big(\mathrm{Log}_{\bar{X}_j}(X_{ij})\big) + V(z_i - \bar{z}) + \varepsilon_i, \tag{49}$$

where $\bar{Y}, \bar{X}_j$ are empirical Fréchet means, $\bar{z} = \mathbb{E}z$, and $V \in \mathbb{R}^{d \times q}$ collects the scalar-covariate tangent vectors $v_k \in T_{\bar{Y}}\mathcal{M}$ as columns. All reported RMSEs use the Fréchet distance on the respective manifold as the evaluation metric. Nominal categorical covariates are one-hot encoded with the first level dropped as the reference category, and periodic covariates are $(\sin, \cos)$-encoded.

Table 13 reports HC3 Wald p-values at the full-data fit. These tests treat observations as independent. For DTI-CC, the p-values that respect within-subject clustering are the subject-blocked permutation p-values in Table 3.

Table 13: HC3 Wald p-values at the full-data fit for the three case studies, treating observations as independent. Predictors are as in Table 3.

DTI-CC		**Protein backbone**		**Wind**	
Predictor	HC3	Predictor	HC3	Predictor	HC3
Tensor neigh. 1	$<10^{-5}$	Offset -2	$<10^{-5}$	Grid pt. 1	$<10^{-5}$
Tensor neigh. 2	$<10^{-5}$	Offset -1	$<10^{-5}$	Grid pt. 2	$<10^{-5}$
CC subregion	$<10^{-5}$	Offset $+1$	$<10^{-5}$	Grid pt. 3	$<10^{-5}$
Age	$<10^{-5}$	Kidera PC1	$<10^{-5}$	Pressure	0.700
Sex	3.1×10^{-5}	Kidera PC2	$<10^{-5}$	Temperature	$<10^{-5}$
ASD diagnosis	1.3×10^{-3}	Kidera PC3	0.118	Season	0.007
FIQ	4.6×10^{-2}	α-helix	$<10^{-5}$		
		Solvent-accessibility	1.8×10^{-3}		

I.1 Diffusion Tensor Field on the Corpus Callosum in $\mathrm{Sym}^+(3)$

Scientific question. Diffusion tensors summarize local water-diffusion anisotropy in white matter and are naturally represented as 3×3 symmetric positive-definite (SPD) matrices under the affine-invariant metric [Pennec et al., 2006]. The corpus callosum is the largest commissural tract, and its tensor field varies systematically along the genu–body axis. We ask whether a local tensor can be predicted from neighboring along-tract tensors after adjusting for subject demographics and anatomical subregion.

Data and preprocessing. We use ABIDE-II diffusion magnetic resonance imaging (MRI) data [Di Martino et al., 2014] from three sites with compatible single-shell acquisitions. Preprocessing comprises motion and eddy-current correction, tensor fitting, atlas-based corpus-callosum extraction, site harmonization in the matrix-log embedding, and removal of subjects with excessive head motion. The final cohort contains 141 subjects and $n = 112{,}694$ corpus-callosum voxels, with approximately 800 voxels sampled per subject.

Variables and model. The response $Y_i \in \mathrm{Sym}^+(3)$ is the diffusion tensor at a corpus-callosum voxel. The two manifold predictors are along-tract neighbor tensors X_{i1} and X_{i2} sampled at ± 2 mm by deterministic tractography. Scalar covariates z_i comprise age, sex, ASD diagnosis, FIQ, and a within-subject indicator for genu versus body. The fitted ROLS model is the canonical specification (49) with $p = 2$ manifold predictors and $q = 5$ scalar covariates.

Riemannian structure and CV design. On $\mathrm{Sym}^+(3)$ with the affine-invariant metric, Log, Exp, and parallel transport admit closed forms via the matrix exponential. Empirical Fréchet means lie in the interior of the cone, so injectivity at $\bar{Y}, \bar{X}_1, \bar{X}_2$ holds without restriction and the hypotheses of Theorem 3.4 are satisfied. Cross-validation is subject-blocked so that voxels from the same subject do not co-occur in training and test folds, which prevents within-subject leakage. Because voxel observations are clustered within subjects, we report both asymptotic Wald tests and block-restricted permutation p-values with $B = 500$ and subject-level exchangeability blocks; the scheme shuffles between-subject covariates at the subject level and within-subject covariates within subject blocks.

Findings. ROLS reduces Fréchet RMSE from 1.128 for the strongest scalar-only baseline to $.400$ when along-tract tensor predictors are included (Table 2). Both manifold predictors attain the permutation floor $p = 0.002$, and the genu-vs-body indicator is also significant at $p = 0.002$ (Table 3). The analytic Wald test treats the $112{,}694$ voxels as independent and gives p-values $< 10^{-5}$, 1.9×10^{-5}, 5.1×10^{-3}, and 1.6×10^{-3} for age, sex, ASD diagnosis, and FIQ. HC3 uses the same independence assumption and gives $< 10^{-5}$, 3.1×10^{-5}, 1.3×10^{-3}, and 4.6×10^{-2}. The subject-blocked permutation p-value for FIQ is 0.447. Benjamini–Hochberg adjustment of the permutation p-values is 0.005 for the two tensor neighbors and the genu-vs-body indicator, and at least 0.42 for the four demographic covariates. The lack-of-fit diagnostic of Appendix H.5 attributes about 10% of the residual variation to omitted quadratic terms.

I.2 Protein Backbone Dihedral Angles on $\mathbb{T}^2$

Scientific question. Protein backbone conformation depends on the residue-level dihedral angles (φ, ψ) on the flat torus $\mathbb{T}^2 = S^1 \times S^1$. The torus is flat, so the tangent-space fit has no curvature contribution, and the remaining geometric feature is the periodicity. We ask whether local backbone geometry is driven primarily by neighboring dihedral angles, residue chemistry, or basin-level secondary-structure indicators.

Data and preprocessing. We extract $n = 6{,}141$ residue windows from 30 high-resolution chains in the Protein Data Bank [Berman et al., 2000]. For each central residue with a complete local window we record its dihedral pair, neighboring pairs at offsets $-2, -1, +1$, the Kidera physicochemical descriptors of the central residue, an α-helix-basin indicator derived from secondary-structure assignment, and a solvent-accessibility proxy based on neighboring C_α distances.

Variables and model. The response $Y_i = (\varphi_i, \psi_i) \in \mathbb{T}^2$ is the central-residue dihedral pair. The three manifold predictors are neighboring dihedral pairs at offsets $-2, -1, +1$; the five scalar predictors are Kidera PC1–PC3, the α-helix indicator, and the solvent-accessibility proxy. The fitted ROLS model is the canonical specification (49) with $p = 3$ manifold predictors and $q = 5$ scalar covariates.

Riemannian structure and CV design. On $\mathbb{T}^2$ the logarithm and exponential maps are componentwise shortest-arc operations on S^1 and parallel transport is the identity, so (49) reduces to vector addition in $\mathbb{R}^2$-tangent representation. The injectivity radius is π in each S^1 component; the empirical Fréchet means are well-defined, and the shortest-arc logarithm is single-valued off the cut locus. Prediction is evaluated by three-fold cross-validation repeated three times with Fréchet RMSE in radians, and the fitted model is compared with the Fréchet mean, geodesic regression, scalar Fréchet regression, single-index Fréchet regression, nonlinear Fréchet regression, and a manifold-only ROLS ablation.

Findings. The Ramachandran conditional distribution has distinct α-helix and β-sheet basins. Full ROLS attains Fréchet RMSE 1.177, geodesic regression on the scalar predictors attains 1.215, and manifold-only ROLS attains 1.395 (Table 2). The neighboring angles and the residue-level

descriptors each reduce the error relative to the fit that uses the other set alone. All three neighboring dihedral predictors, Kidera PC1–PC2, the α-helix indicator, and the solvent-accessibility proxy are significant in Table 3. Kidera PC3, kept as a null control, has Wald $p = 0.345$ and HC3 $p = 0.118$ at the full-data fit. Benjamini–Hochberg adjustment leaves Kidera PC3 at 0.345 and the solvent-accessibility proxy at 3.2×10^{-4}. The lack-of-fit diagnostic attributes about 1.6% of the residual variation to omitted quadratic terms. On the Ramachandran plot (Figure 3) the full ROLS estimator places predictions in both the α-helix and β-sheet basins. Scalar-only methods concentrate in a point or narrow clusters, nonlinear Fréchet smooths across the two dominant basins, and manifold-only ROLS spreads predictions over the $\varphi < 0$ half-plane without separating the two basins.

I.3 Atmospheric Winds on S^2

Scientific question. Mid-latitude weather over central Europe is organized by synoptic systems that translate eastward across the continent, so winds at nearby grid points should share a common spatial signal once seasonal and thermodynamic covariates are removed. We ask whether the wind at a central European grid point is predictable from the winds at three neighboring grid points after adjusting for sea-level pressure, temperature, and day-of-year.

Data and preprocessing. We use $850\,\text{hPa}$ winds and surface meteorological covariates from the NCEP/NCAR Reanalysis 1 [Kalnay et al., 1996] for the year 2023, sampled at four grid points ($50°$N, $10°$E for the response; $52.5°$N, $5°$E; $47.5°$N, $5°$E; and $50°$N, $15°$E for the predictors) at daily resolution. Each daily horizontal wind (u, v) is mapped to the point on S^2 with azimuth $\operatorname{atan2}(v, u)$ and polar angle $(\pi/2)\, s/s_{\max}$, where s is the wind speed and $s_{\max}$ its annual maximum at that grid point. Sea-level pressure and near-surface temperature are extracted from the same reanalysis product at the response grid point.

Variables and model. The response $Y_i \in S^2$ is the daily wind at the response grid point. The three manifold predictors are the winds at the three neighboring grid points, each in S^2. Day-of-year is encoded by two orthogonal features $(\sin 2\pi\,\text{doy}/365, \cos 2\pi\,\text{doy}/365)$, yielding four scalar predictors (pressure, temperature, and the two seasonal features). A combined "season" p-value pools the two seasonal-feature coefficient tests via the joint Wald test on these two coefficients. The fitted ROLS model is the canonical specification (49) with $p = 3$ and $q = 4$.

Riemannian structure and CV design. S^2 is equipped with the round metric inherited from $\mathbb{R}^3$. The exponential map uses the Rodrigues formula, the logarithm map is its inverse, and parallel transport is rotation about the normal to the geodesic plane. The injectivity radius is π and all daily wind points lie in the same hemisphere, so the injectivity condition (5) of Theorem 3.4 is satisfied at every empirical Fréchet mean. The 365 daily observations are strongly autocorrelated because days within the same week share the same synoptic weather system, so random day-level folds would leak synoptic information across train and test. We therefore use blocked-temporal CV in the sense of Bergmeir and Benítez [2012], with five contiguous day-blocks, no random shuffling, and a single pass.

Findings. The sample has $n = 365$ days and manifold dimension $d = 2$. The four grid points share synoptic systems, and including the neighboring winds produces the RMSE reduction in Table 2. All three predictor winds are significant in Table 3. Temperature and the combined seasonal factor remain significant after Benjamini–Hochberg adjustment and under HC3. Sea-level pressure does not. The directional signal comes from the neighboring winds, and the seasonal and thermodynamic covariates account for smaller adjustments.

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 abstract and introduction state the cut-locus well-definedness condition (Theorem 3.4), a closed-form estimator (Theorem 3.3), convergence rates (Theorem 4.2), and asymptotic inference (Theorem 4.5). All are proved in the appendix and validated empirically in Sections 5–6.

   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: Sections 3–7 discuss limitations including the scalar coefficient restriction ($\alpha_j \in \mathbb{R}$ rather than a full linear map), the injectivity and tangent-linearity assumptions, the need for a fixed transport convention, and the requirement for closed-form geometric primitives, together with the same-manifold restriction, retraction-based approximations, multimodal responses, fixed dimension, and batch updates.

   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: All theorems are stated with explicit assumptions (Assumption 4.1 and conditions in each theorem statement). Complete proofs are given in the appendix.

   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: The model and data-generating process are specified in Eq. (1) (Model) and described in detail in Appendix H.1 with all parameter values. Section 3.1 gives the complete estimation procedure. Appendix H provides full experimental details for all simulations and Appendix I for all case studies, including data sources, preprocessing, and hyperparameters.

   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: [Yes]

   Justification: The core computer programs implementing ROLS are provided at `https://github.com/XiaoyuChenPub/ROLS`. The paper provides algorithmic details and experimental settings in the main text and appendix, and all datasets used are publicly available and cited.

   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: Section 5 specifies default parameters (n, p, q, τ, true coefficients). Appendix I details data splits (subject-blocked, repeated, and blocked-temporal CV as appropriate) and all preprocessing steps.

   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: Simulation results report mean $\pm$ standard deviation across Monte Carlo trials, with 100 trials for the prediction-accuracy and convergence-rate experiments (Tables 5, 6), 500 trials for Type I error and power, 300 trials for the bootstrap calibration (Table 7), and 30 trials for the misspecification probe. Case study RMSEs in Table 2 report mean $\pm$ standard deviation across cross-validation folds. Error bars in Figure 2 are ± 1 Monte Carlo standard deviation across trials.

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: [No]

   Justification: The paper does not report specific compute resources. All experiments are lightweight (closed-form linear algebra on small-to-moderate datasets) and run on a standard laptop CPU in minutes. No GPU is required. Single-threaded timings of one ROLS fit are reported in Appendix H.8.

   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 research is theoretical and methodological with no human subjects, no personally identifiable data, and no foreseeable negative societal impact.

   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: [N/A]

    Justification: This is foundational statistical methodology for regression on Riemannian manifolds. It is not tied to a specific application that poses foreseeable societal risks. The case studies use publicly available scientific datasets.

    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 release pre-trained models, generative models, or scraped datasets. The method is a statistical estimator with no dual-use risk.

    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: All datasets are cited with original references or public repositories (ABIDE-II diffusion MRI, Protein Data Bank, NCEP/NCAR Reanalysis 1). Data sources and preprocessing are provided in Appendix I.

    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: [Yes]

    Justification: The ROLS code is released at `https://github.com/XiaoyuChenPub/ROLS` under the CC BY-NC 4.0 license, with installation instructions and a usage example.

    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 paper does not involve crowdsourcing or research with human subjects.

    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: The paper does not involve human subjects research.

    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 are not used as a component of the core methodology.

    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.