DIFFERENTIALLY PRIVATE INFERENCE FOR GEODESIC REGRESSION ON RIEMANNIAN MANIFOLDS

Anonymous authors
Paper under double-blind review

ABSTRACT

In this paper, we study inference for geodesic regression with Euclidean predictors and a manifold-valued response under Gaussian differential privacy (GDP). We develop a one-shot GDP inference procedure via sample splitting. The released outputs include a private parameter estimator, a private Wald region for the footpoint and shooting-vector parameters, and pointwise private confidence regions for the fitted geodesic. We derive sensitivity bounds and a GDP guarantee, establish consistency and asymptotic normality, and construct a consistent private covariance estimator. We also illustrate our methodology on several common manifolds and explain how the main theorems specialize in these settings, including Euclidean spaces, spheres, Hadamard manifolds, affine-invariant or log-Euclidean SPD manifolds, and Grassmann manifolds. Extensive numerical experiments provide positive support for the proposed GDP inference framework. We further apply the methodology to the tropical-cyclone track data.

1 INTRODUCTION

Manifold-valued data now appear routinely in statistics and machine learning whenever the object of interest is a direction, shape, covariance matrix, diffusion tensor, rotation, or subspace. Representative examples include directional observations on spheres, diffusion tensors and covariance matrices on symmetric positive-definite (SPD) manifolds, and subspace descriptors on Grassmann manifolds (Pennec et al., 2006; Arsigny et al., 2006; Kim et al., 2014). In such problems the response space lacks global linear structure, so Euclidean regression can distort distances, ignore curvature, and depend on arbitrary coordinates. This has motivated a substantial literature on geometry-aware regression for manifold-valued responses, including population shape regression from random-design data (Davis et al., 2010), intrinsic regression models (Shi et al., 2009), local polynomial regression for SPD responses (Yuan et al., 2012), geodesic regression (Niethammer et al., 2011; Fletcher, 2013; Kim et al., 2014), intrinsic polynomial models (Hinkle et al., 2014), regression on Riemannian symmetric spaces (Cornea et al., 2017), extrinsic local regression (Lin et al., 2017), and Fréchet regression (Petersen & Müller, 2019). Among these approaches, geodesic regression is the closest intrinsic analogue of Euclidean linear regression: it models the conditional center of the response by a geodesic determined by a footpoint together with one or more shooting vectors.

The motivating applications often involve sensitive data. Medical-imaging data, subject-specific covariance objects, trajectory data, and related geometric summaries are typically derived from records that should not be disclosed at the individual level. Differential privacy (DP) provides a rigorous framework for releasing statistical information while limiting the information revealed about any single observation (Dwork et al., 2006; Dwork & Roth, 2014). In Euclidean regression, the literature has progressed from private empirical-risk minimization and objective perturbation (Chaudhuri et al., 2011) and private ordinary least squares (Sheffet, 2019) to inferential procedures based on bootstrap corrections, noisy optimization, and analytic covariance adjustments (Ferrando et al., 2022; Avella-Medina et al., 2023; Lin et al., 2025; Sopa et al., 2026). Recent work has also developed intrinsic privacy mechanisms on Riemannian manifolds (Reimherr et al., 2021), shape- and structure-preserving releases (Soto et al., 2022), Riemannian Gaussian mechanisms (Jiang et al., 2023), and private parameter estimation for geodesic regression (Kulkarni & Soto, 2025).

What is still missing is private inference for geodesic regression. The regression parameter is naturally a footpoint together with a collection of tangent directions, living on a smooth parameter bundle

Reviewers: please read the Reviewer Guidelines (iclr.cc/Conferences/2027/ReviewerGuidelines) and the AI Policy for Reviewers (iclr.cc/Conferences/2027/AIPolicyForReviewers). If you used AI to expand, edit, or polish your review, please provide the input text to the LLM. Better still, consider skipping the LLM and submitting your original text: we, and the authors, are much more interested in your unedited thoughts than in what an LLM has to say. AI-assisted or not, you are putting your name and reputation behind your review: LLM-generated falsehoods, hallucinations or misrepresentations are subject to disciplinary action, which may include desk-rejecting all papers you have authored.

over the response manifold. Consequently, the score and Hessian are defined in local coordinates on the parameter space, while the loss retains the Riemannian distance between the observed and fitted responses. Moving the footpoint changes both the fitted regression map and the tangent spaces in which the shooting vectors lie (Niethammer et al., 2011; Fletcher, 2013; Hinkle et al., 2014; Kim et al., 2014). This joint parameterization differs from fixing a tangent or ambient representation of the responses before fitting a regression (Pigoli et al., 2016; Lin et al., 2017; Xiong et al., 2022). The squared-distance loss also inherits geometric singularities near cut loci. Curvature and the working neighborhood affect derivative bounds, conditioning, and the amount of clipping, even though the sensitivity bound for a clipped average depends only on the chosen threshold and sample size.

There is also an inferential difficulty that does not arise in private point estimation. In scientific applications, one needs uncertainty quantification for both the finite-dimensional regression parameter and the fitted manifold-valued curve. Under privacy, the uncertainty has two sources: sampling fluctuation and the injected Gaussian noise. We combine one-step asymptotics with privacy noise accounting by releasing a small set of derivative summaries once. Fitted-response estimates, covariance estimates, Wald regions, and pointwise fitted-response regions are then obtained by post-processing. This construction makes the privacy cost independent of the number of fitted-point queries.

The main contributions are as follows.

- We introduce a split-sample, one-shot GDP algorithm for inference in geodesic regression. Conditional on a public or private pilot, the method privatizes only clipped averages of the score, Hessian, and score second moment in a local chart on the parameter bundle.
- We derive explicit sensitivity bounds and a GDP guarantee, including privacy accounting for a private pilot and the parallel composition induced by sample splitting.
- We establish consistency, an asymptotic linear representation, a central limit theorem, and a consistent private covariance estimator with an analytic correction for the injected score noise.
- We construct parameter Wald regions and pointwise fitted-response regions, and examine pilot sensitivity, active clipping, and finite-sample performance on common manifolds.

The closest existing work is Kulkarni & Soto (2025), which develops an intrinsic K-norm-gradient mechanism for private geodesic regression estimation. We develop an inference procedure that takes an independent preliminary estimate as input and provides private covariance estimates and confidence regions for the parameter and fitted curve. This also complements Euclidean DP inference (Ferrando et al., 2022; Avella-Medina et al., 2023; Lin et al., 2025) by accounting for the geometry of both the parameter bundle and the fitted-response map.

The rest of the paper is organized as follows. Section 2 introduces the proposed DP geodesic-regression inference framework. Section 3 develops sensitivity, privacy, and asymptotic theory. Section 4 presents the numerical experiments. Section 5 gives a real data application. Section 6 concludes. Discussions, additional numerical results and all proofs are given in the Appendix.

2 METHODOLOGY

2.1 GEODESIC REGRESSION AND PARAMETER-BUNDLE COORDINATES

Let $(\mathcal{M}, \rho)$ be a d-dimensional connected Riemannian manifold, and let $(\boldsymbol{X}, Y) \in \mathbb{R}^{d_x} \times \mathcal{M}$ denote a Euclidean predictor and a manifold-valued response. We use $\boldsymbol{x} = 0$ as the reference predictor and write the geodesic regression model

$$\gamma_\theta(\boldsymbol{x}) = \mathrm{Exp}_p\left(\sum_{j=1}^{d_x} x_j v_j\right), \qquad \boldsymbol{\theta} = (p, \boldsymbol{v}), \quad \boldsymbol{v} = (v_1, \ldots, v_{d_x}), \quad v_j \in \mathcal{T}_p\mathcal{M},$$

where $p \in \mathcal{M}$ is the footpoint and the v_j are shooting vectors. The parameter bundle is $\mathcal{TM} = \{(p, \boldsymbol{v}) : p \in \mathcal{M},\ v_j \in \mathcal{T}_p\mathcal{M}\}$, of dimension $d_\theta = d(1 + d_x)$. For $d_x = 1$, it is the tangent bundle. For $d_x > 1$, it is the direct sum of d_x tangent bundles over a common footpoint. For i.i.d.

observations $W_i = (\boldsymbol{X}_i, Y_i)$, the empirical and population least-squares risks are

$$L_n(\boldsymbol{\theta}) = \frac{1}{2n}\sum_{i=1}^{n} \rho^2\{Y_i, \gamma_\theta(\boldsymbol{X}_i)\}, \qquad L(\boldsymbol{\theta}) = \frac{1}{2}\mathbb{E}\big[\rho^2\{Y, \gamma_\theta(\boldsymbol{X})\}\big].$$

Let $\boldsymbol{\theta}_0 = (p_0, \boldsymbol{v}_0)$ denote the unique minimizer of L. Since the parameter $\boldsymbol{\theta}_0$ lies on $\mathcal{TM}$ rather than in a vector space, we work in a publicly fixed smooth chart $\phi : \mathcal{U} \to \phi(\mathcal{U}) \subset \mathbb{R}^{d_\theta}$ on a neighborhood of $\boldsymbol{\theta}_0$. Write $\boldsymbol{\eta} = \phi(\boldsymbol{\theta})$ and $\boldsymbol{\eta}_0 = \phi(\boldsymbol{\theta}_0)$. The loss in parameter coordinates is

$$\ell(w, \boldsymbol{\eta}) = \frac{1}{2}\rho^2\{y, \gamma_{\phi^{-1}(\eta)}(\boldsymbol{x})\}, \qquad w = (\boldsymbol{x}, y). \tag{1}$$

The score function, Hessian matrix, and score second moment in this chart are $\boldsymbol{\psi}(w, \boldsymbol{\eta}) = \nabla_\eta \ell(w, \boldsymbol{\eta})$, $\mathbf{H}(w, \boldsymbol{\eta}) = \nabla_\eta^2 \ell(w, \boldsymbol{\eta})$ and $\mathbf{M}(w, \boldsymbol{\eta}) = \boldsymbol{\psi}(w, \boldsymbol{\eta})\boldsymbol{\psi}(w, \boldsymbol{\eta})^\top$, respectively. We further write $\mathbf{A}_0 := \mathbb{E}\{\mathbf{H}(W, \boldsymbol{\eta}_0)\}$, $\mathbf{J}_0 := \mathrm{Var}\{\boldsymbol{\psi}(W, \boldsymbol{\eta}_0)\}$. Under the differentiability conditions below, $\mathbb{E}\{\boldsymbol{\psi}(W, \boldsymbol{\eta}_0)\} = 0$ at the population minimizer, so $\mathbf{J}_0 = \mathbb{E}\{M(W, \boldsymbol{\eta}_0)\}$. Appendix A.1 gives explicit formulas for the loss, score, and Hessian on $\mathbb{S}^1$ and log-Euclidean SPD manifolds.

2.2 A ONE-SHOT PRIVATE ONE-STEP ESTIMATOR

Definition 1 (Gaussian differential privacy)**.** *A randomized mechanism $\mathcal{A}$ is μ-GDP if for all adjacent datasets $\mathcal{D} \simeq \mathcal{D}'$ the trade-off function of testing $\mathcal{A}(\mathcal{D})$ against $\mathcal{A}(\mathcal{D}')$ is bounded below by the trade-off function of testing $N(0,1)$ against $N(\mu, 1)$ (Dong et al., 2022).*

The algorithm releases private uncertainty information by privatizing only three finite-dimensional averages in the chosen chart. We use a disjoint data split $\mathcal{D}_n = \mathcal{D}_1 \cup \mathcal{D}_2$ with $|\mathcal{D}_1| = n_1$ and $|\mathcal{D}_2| = n_2$, where the split is independent of the data values. The first part is used to construct a pilot, which is a preliminary estimator for the one-step update, and the second part is used for the private score, Hessian, and second-moment releases. Let $\mathcal{F}_1 = \sigma(\mathcal{D}_1, \widetilde{\boldsymbol{\eta}}_n)$ denote the first-stage information, including the realized pilot. The inference sample and the fresh release noises are independent of $\mathcal{F}_1$.

The pilot may be obtained from a private geodesic-regression estimator on $\mathcal{D}_1$ or from public data or a reference geodesic independent of $\mathcal{D}_n$. We treat pilot construction as a separate estimation step and focus on the subsequent private inference. The asymptotic results require the accuracy condition in Assumption 6. The privacy guarantee in Theorem 1 holds for every pilot value in K, which is a public working set defined below in Assumption 3, with the pilot's privacy budget included in the accounting. A public pilot has $\mu_0 = 0$ and permits $\mathcal{D}_1 = \varnothing$ and $\mathcal{D}_2 = \mathcal{D}_n$. Our experiments use a KNG-based pilot following Kulkarni & Soto (2025), with implementation details in Appendix B and Appendix C.1. Appendix C.3 examines the finite-sample effects of public warm-start misspecification on estimation error and coverage.

For a vector $\boldsymbol{z} \in \mathbb{R}^k$ and a matrix $\mathbf{A}$, define the clipping operators

$$\mathrm{clip}_c(\boldsymbol{z}) = \boldsymbol{z}\min\{1, c/\|\boldsymbol{z}\|_2\}, \qquad \mathrm{clip}_c(\mathbf{A}) = \mathbf{A}\min\{1, c/\|\mathbf{A}\|_F\},$$

with the convention $\mathrm{clip}_c(0) = 0$. Clipping bounds each contribution and hence the sensitivity of the sample average. The thresholds can be fixed using public information or chosen by privately estimating the 95% quantiles of the corresponding norms on $\mathcal{D}_2$ after obtaining the pilot. The cost should be included in the privacy budget. At a fixed total budget, this leaves less budget for the releases. Finally, define $\mathrm{sym}(\mathbf{B}) = \big(\mathbf{B} + \mathbf{B}^\top\big)/2$. Let $\mathbb{S}_+^{d_\theta}$ denote the cone of $d_\theta \times d_\theta$ symmetric positive semidefinite matrices, and let $\Pi_{\mathbb{S}_+^{d_\theta}}$ be the Frobenius-norm projection.

Algorithm 1 summarizes our proposed inference procedure. Let $K \subset \phi(\mathcal{U})$ be the publicly fixed nonempty closed convex set in Assumption 3, and let Π_K denote Euclidean projection onto K. Before evaluating derivatives, replace the pilot by $\widetilde{\boldsymbol{\eta}}_n \leftarrow \Pi_K(\widetilde{\boldsymbol{\eta}}_n)$. We use the same symbol for this projected pilot. Define the raw and final updates by

$$\widehat{\boldsymbol{\eta}}_{n,\mathrm{raw}}^{dp} = \widetilde{\boldsymbol{\eta}}_n - (\widehat{\mathbf{H}}_n^{dp})^{-1}\boldsymbol{\psi}_n^{dp}, \qquad \widehat{\boldsymbol{\eta}}_n^{dp} = \Pi_K(\widehat{\boldsymbol{\eta}}_{n,\mathrm{raw}}^{dp}),$$

where the private score $\boldsymbol{\psi}_n^{dp}$ and the private Hessian $\widehat{\mathbf{H}}_n^{dp}$ are defined in Algorithm 1. This update can be viewed as a one-step Newton correction initialized at the pilot estimator $\widetilde{\boldsymbol{\eta}}_n$. The algorithm is

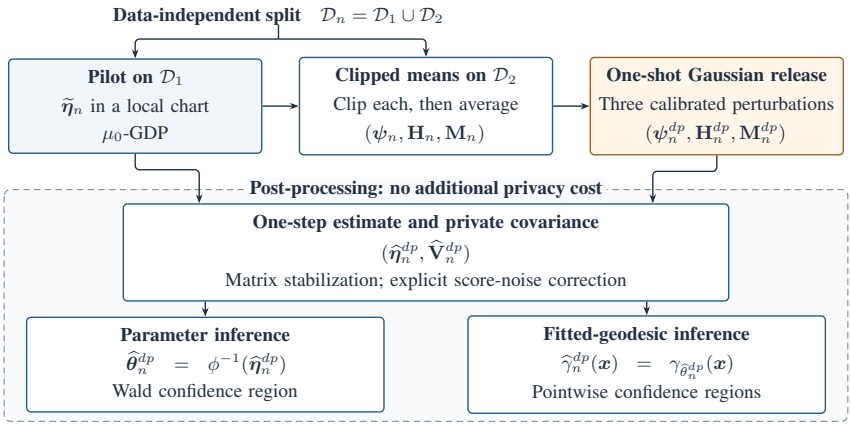


Figure 1: The workflow of Algorithm 1.

deliberately one-shot: rather than privatizing the fitted values $\widehat{\gamma}_n^{dp}(\boldsymbol{x}) = \gamma_{\widehat{\theta}_n^{dp}}(\boldsymbol{x})$ one query point at a time, it releases a single finite-dimensional private object, namely the parameter estimate together with its covariance estimate. The entire fitted curve, as well as any finite or infinite collection of queries along the curve, can then be obtained by post-processing.

Algorithm 1 Split-sample GDP inference for geodesic regression

Require: Data $\mathcal{D}_n = \{W_i\}_{i=1}^n$, chart ϕ on $\mathcal{TM}$, clipping levels (c_ψ, c_H, c_M), ridge λ_n, privacy budgets $(\mu_0, \mu_\psi, \mu_H, \mu_M)$, publicly fixed nonempty closed convex set $K \subset \phi(\mathcal{U})$

1: Split $\mathcal{D}_n$ into two disjoint parts $\mathcal{D}_1$ and $\mathcal{D}_2$
2: Obtain a pilot $\widetilde{\boldsymbol{\eta}}_n$ either
   from a μ_0-GDP mechanism on $\mathcal{D}_1$,
   or from public information independent of $\mathcal{D}_n$ if such information is available.
3: Set $\widetilde{\boldsymbol{\eta}}_n \leftarrow \Pi_K(\widetilde{\boldsymbol{\eta}}_n)$.
4: **for** $i \in \mathcal{D}_2$ **do**
5: Compute $\boldsymbol{\psi}_i = \boldsymbol{\psi}(W_i, \widetilde{\boldsymbol{\eta}}_n)$, $\mathbf{H}_i = \mathbf{H}(W_i, \widetilde{\boldsymbol{\eta}}_n)$, and $\mathbf{M}_i = \boldsymbol{\psi}_i \boldsymbol{\psi}_i^\top$
6: Clip $\widetilde{\boldsymbol{\psi}}_i = \mathrm{clip}_{c_\psi}(\boldsymbol{\psi}_i)$, $\widetilde{\mathbf{H}}_i = \mathrm{clip}_{c_H}(\mathbf{H}_i)$, and $\widetilde{\mathbf{M}}_i = \mathrm{clip}_{c_M}(\mathbf{M}_i)$
7: **end for**
8: Empirical means $\boldsymbol{\psi}_n = n_2^{-1} \sum_{i \in \mathcal{D}_2} \widetilde{\boldsymbol{\psi}}_i$, $\mathbf{H}_n = n_2^{-1} \sum_{i \in \mathcal{D}_2} \widetilde{\mathbf{H}}_i$, $\mathbf{M}_n = n_2^{-1} \sum_{i \in \mathcal{D}_2} \widetilde{\mathbf{M}}_i$.
9: Release private versions $\boldsymbol{\psi}_n^{dp} = \boldsymbol{\psi}_n + \boldsymbol{Z}_{\psi,n}$, $\mathbf{H}_n^{dp} = \mathbf{H}_n + \mathbf{Z}_{H,n}$, $\mathbf{M}_n^{dp} = \mathbf{M}_n + \mathbf{Z}_{M,n}$, where

$$\boldsymbol{Z}_{\psi,n} \sim N\left(0, \sigma_{\psi,n}^2 \mathbf{I}_{d_\theta}\right), \quad \mathrm{vec}(\mathbf{Z}_{H,n}) \sim N\left(0, \sigma_{H,n}^2 \mathbf{I}_{d_\theta^2}\right), \quad \mathrm{vec}\left(\mathbf{Z}_{M,n}\right) \sim N\left(0, \sigma_{M,n}^2 \mathbf{I}_{d_\theta^2}\right),$$

   with $\sigma_{\psi,n} = 2c_\psi/(\mu_\psi n_2)$, $\sigma_{H,n} = 2c_H/(\mu_H n_2)$, and $\sigma_{M,n} = 2c_M/(\mu_M n_2)$.
   The three Gaussian perturbations are mutually independent and independent of the data and the pilot randomness.
10: Set $\widehat{\mathbf{H}}_n^{dp} = \Pi_{\mathbb{S}_+^{d_\theta}}\big(\mathrm{sym}(\mathbf{H}_n^{dp})\big) + \lambda_n \mathbf{I}_{d_\theta}$, $\widehat{\mathbf{M}}_n^{dp} = \Pi_{\mathbb{S}_+^{d_\theta}}\big(\mathrm{sym}(\mathbf{M}_n^{dp})\big)$
11: Compute the raw update $\widehat{\boldsymbol{\eta}}_{n,\mathrm{raw}}^{dp} = \widetilde{\boldsymbol{\eta}}_n - (\widehat{\mathbf{H}}_n^{dp})^{-1} \boldsymbol{\psi}_n^{dp}$, and set $\widehat{\boldsymbol{\eta}}_n^{dp} = \Pi_K(\widehat{\boldsymbol{\eta}}_{n,\mathrm{raw}}^{dp})$.
12: Compute the private covariance estimator $\widehat{\mathbf{V}}_n^{dp} = \left(\widehat{\mathbf{H}}_n^{dp}\right)^{-1} \left(\frac{\widehat{\mathbf{M}}_n^{dp}}{n_2} + \sigma_{\psi,n}^2 \mathbf{I}_{d_\theta}\right) \left(\widehat{\mathbf{H}}_n^{dp}\right)^{-1}$
13: Output $\widehat{\boldsymbol{\theta}}_n^{dp} = \phi^{-1}(\widehat{\boldsymbol{\eta}}_n^{dp})$, the covariance $\widehat{\mathbf{V}}_n^{dp}$, and the fitted geodesic $\widehat{\gamma}_n^{dp}(\boldsymbol{x}) = \gamma_{\widehat{\theta}_n^{dp}}(\boldsymbol{x})$, $\boldsymbol{x} \in \mathbb{R}^{d_x}$.

2.3 Assumptions

We separate assumptions on the regression model and local geometry from those on the pilot and regularization. Throughout the main theory, d, d_x, positive inference budgets, and clipping levels are fixed, and $n_2 \to \infty$.

Assumption 1 (bounded predictor support)**.** *There exists $B_X < \infty$ such that $\|\boldsymbol{X}\|_2 \le B_X$ a.s.*

Assumption 2 (local geodesic neighborhood)**.** *There exists a compact geodesically convex set $K_{\mathcal{M}} \subset \mathcal{M}$ such that: (a) $Y \in K_{\mathcal{M}}$ almost surely; (b) for every $\boldsymbol{\theta}$ in the parameter neighborhood under consideration and every $\|\boldsymbol{x}\|_2 \le B_X$, the point $\gamma_\theta(\boldsymbol{x})$ belongs to $K_{\mathcal{M}}$; (c) on $K_{\mathcal{M}} \times K_{\mathcal{M}}$, the squared distance function $(y_1, y_2) \mapsto \rho^2(y_1, y_2)$ is C^3.*

Assumption 3 (bundle chart and derivative bounds)**.** *There exists an open neighborhood $\mathcal{U} \subset \mathcal{TM}$ of $\boldsymbol{\theta}_0$ and a smooth chart $\phi : \mathcal{U} \to \phi(\mathcal{U}) \subset \mathbb{R}^{d_\theta}$ with $d_\theta = d(d_x + 1)$ such that $\phi(\mathcal{U})$ contains a publicly fixed closed convex set K with $\boldsymbol{\eta}_0 \in \mathrm{int}(K)$ and the projected pilot satisfies $\widetilde{\boldsymbol{\eta}}_n \in K$ for every realization. Moreover, on K the loss $\ell(w, \boldsymbol{\eta})$ is three times continuously differentiable in $\boldsymbol{\eta}$ and there exist finite constants G_ψ, G_H, L_H such that for all w in the support of W and all $\boldsymbol{\eta}, \boldsymbol{\eta}' \in K$, $\|\boldsymbol{\psi}(w, \boldsymbol{\eta})\|_2 \le G_\psi$, $\|\mathbf{H}(w, \boldsymbol{\eta})\|_F \le G_H$, $\|\mathbf{H}(w, \boldsymbol{\eta}) - \mathbf{H}(w, \boldsymbol{\eta}')\|_F \le L_H \|\boldsymbol{\eta} - \boldsymbol{\eta}'\|_2$.*

Assumption 4 (identifiability)**.** *The population risk $L(\boldsymbol{\theta})$ has a unique minimizer $\boldsymbol{\theta}_0 \in \mathcal{U}$, $\mathbb{E}\{\boldsymbol{\psi}(W, \boldsymbol{\eta}_0)\} = 0$, and the information matrix $\mathbf{A}_0 = \mathbb{E}\{\mathbf{H}(W, \boldsymbol{\eta}_0)\}$ is positive definite.*

Assumption 5 (inactive clipping)**.** *The clipping levels in Algorithm 1 satisfy $c_\psi \ge G_\psi$, $c_H \ge G_H$, $c_M \ge G_\psi^2$. Hence clipping is theoretically inactive on the support.*

Assumption 6 (pilot)**.** *The pilot $\widetilde{\boldsymbol{\eta}}_n$ is independent of $\mathcal{D}_2$. For consistency we require $\widetilde{\boldsymbol{\eta}}_n - \boldsymbol{\eta}_0 = o_p(1)$. For the asymptotic linear representation and central limit theorem we require the stronger rate $\|\widetilde{\boldsymbol{\eta}}_n - \boldsymbol{\eta}_0\|_2 = o_p(n_2^{-1/4})$.*

Assumption 7 (ridge scaling)**.** *$\lambda_n > 0$, $\sqrt{n_2}\lambda_n \to 0$ and $\log(1/\lambda_n) = o(n_2)$ as $n \to \infty$.*

Assumptions 1–3 give a smooth local loss with bounded derivatives. Similar local coordinates are also used for asymptotic inference in intrinsic regression (Cornea et al., 2017). Hadamard manifolds have no cut locus. On spheres and Grassmannians, the cut-locus condition holds when the responses and fitted points lie in a compact subset of a common convex normal neighborhood. Algorithm 1 projects the pilot onto K before evaluating derivatives and the raw update onto K before applying ϕ^{-1}. These projections preserve GDP, ensure valid chart evaluations, and do not increase pilot error. Since $\widehat{\boldsymbol{\eta}}_{n,\mathrm{raw}}^{dp} \xrightarrow{p} \boldsymbol{\eta}_0 \in \mathrm{int}(K)$, the final projection is asymptotically inactive. Assumption 4 is the usual local identifiability condition for M-estimation. Assumption 5 removes clipping bias from the main asymptotic results. Bounded scores and derivatives are also standard in private ERM and M-estimation literature (Chaudhuri et al., 2011; Avella-Medina et al., 2023). Appendix F analyzes active clipping and gives conditions for valid inference. Appendices C.2 and C.5 examine threshold sensitivity and its interaction with curvature, showing that our method performs robustly under moderate clipping. Assumption 6 gives the initial accuracy required by the one-step correction. The $o_p(n_2^{-1/4})$ rate makes the quadratic pilot remainder negligible at the $n_2^{-1/2}$ scale. In particular, a pilot with $O_p(n_1^{-1/2})$ error suffices when $n_1/n_2 \to r \in (0, \infty)$. An independent public estimator of the same $\boldsymbol{\eta}_0$ with $O_p(m^{-1/2})$ error also satisfies it when $m/\sqrt{n_2} \to \infty$, where m is the public sample size. A fixed deterministic warm start meets this rate only if it equals $\boldsymbol{\eta}_0$. Nevertheless, the simulations in Appendix C.3 suggest that the method can tolerate moderate misspecification of such warm starts in practice. Assumption 7 stabilizes inversion of the noisy Hessian while making the ridge negligible at the $n_2^{-1/2}$ scale.

3 Theoretical Results

In this section, we control the private releases in the Euclidean chart. Theorem 1 gives the sensitivity and GDP guarantee, Theorems 2–3 establish consistency and the one-step linear expansion, and Theorem 4 establishes consistency of the private covariance estimator. Then, we map these chart-based results back to the geometry. Corollary 1 gives a Wald region on $\mathcal{TM}$, and Theorem 5 propagates the parameter covariance through the fitted-geodesic map to obtain pointwise confidence regions on $\mathcal{M}$. Figure 2 illustrates the relationships among the notations, mappings, and theorems presented in this section.

3.1 Sensitivity and privacy

Theorem 1 (Sensitivity and GDP guarantee)**.** *Conditional on any pilot value in K, the deterministic clipping step in Algorithm 1 implies that the global ℓ_2/Frobenius sensitivities of the three averages*

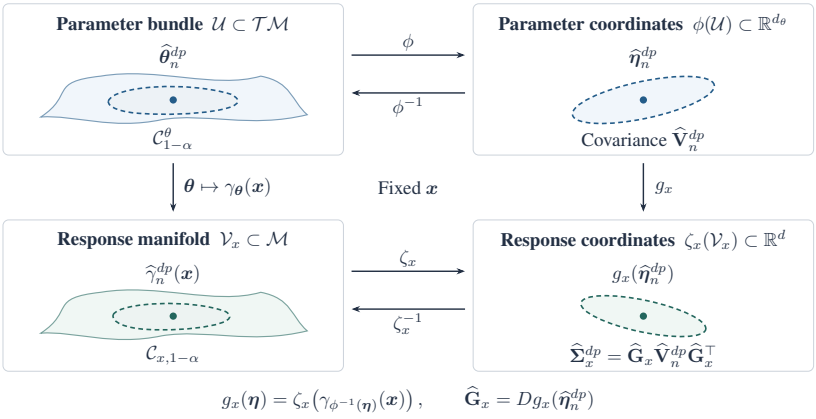


Figure 2: Illustrations of Theorems 3–5.

released from $\mathcal{D}_2$ satisfy $\Delta_{\psi,n} \le 2c_\psi/n_2$, $\Delta_{H,n} \le 2c_H/n_2$, $\Delta_{M,n} \le 2c_M/n_2$. Consequently, the private score, Hessian, and second-moment releases on $\mathcal{D}_2$ are μ_ψ-, μ_H-, and μ_M-GDP, respectively, and their joint release is $\mu_{\mathcal{D}_2} = \sqrt{\mu_\psi^2 + \mu_H^2 + \mu_M^2}$-GDP. If the pilot mechanism on $\mathcal{D}_1$ is μ_0-GDP, or if the pilot is public and independent of $\mathcal{D}_n$ with $\mu_0 = 0$, and if the split $\mathcal{D}_n = \mathcal{D}_1 \cup \mathcal{D}_2$ is data-independent, then the full mechanism in Algorithm 1 is $\mu_{\mathrm{tot}} = \max\Big\{\mu_0, \sqrt{\mu_\psi^2 + \mu_H^2 + \mu_M^2}\Big\}$-GDP.

The privacy guarantee in Theorem 1 is independent of the number of fitted-point queries. With fixed clipping thresholds, geometry affects the derivative distributions and statistical accuracy, while the sensitivity bounds remain unchanged.

3.2 Consistency and asymptotic linearity

Theorem 2 (Consistency)**.** *Under Assumptions 1–7, if $\widetilde{\boldsymbol{\eta}}_n - \boldsymbol{\eta}_0 = o_p(1)$, then $\widehat{\boldsymbol{\eta}}_n^{dp} \xrightarrow{\mathbb{P}} \boldsymbol{\eta}_0$, $\widehat{\boldsymbol{\theta}}_n^{dp} = \phi^{-1}(\widehat{\boldsymbol{\eta}}_n^{dp}) \xrightarrow{\mathbb{P}} \boldsymbol{\theta}_0$.*

Theorem 3 (Asymptotic linear representation and CLT)**.** *Suppose Assumptions 1–7 hold, and in Assumption 6, $\|\widetilde{\boldsymbol{\eta}}_n - \boldsymbol{\eta}_0\|_2 = o_p(n_2^{-1/4})$. Then*

$$\widehat{\boldsymbol{\eta}}_n^{dp} - \boldsymbol{\eta}_0 = -\mathbf{A}_0^{-1}\left\{ n_2^{-1} \sum_{i \in \mathcal{D}_2} \boldsymbol{\psi}(W_i, \boldsymbol{\eta}_0) + \boldsymbol{Z}_{\psi,n} \right\} + \boldsymbol{r}_n, \quad \|\boldsymbol{r}_n\|_2 = o_p\left(n_2^{-1/2}\right), \tag{2}$$

where $\boldsymbol{Z}_{\psi,n} \sim N(0, \sigma_{\psi,n}^2 \mathbf{I}_{d_\theta})$ is the score-noise term in Algorithm 1. Consequently, $\sqrt{n_2}\,(\widehat{\boldsymbol{\eta}}_n^{dp} - \boldsymbol{\eta}_0) \Rightarrow N\big(0,\ \mathbf{A}_0^{-1}\mathbf{J}_0\mathbf{A}_0^{-1}\big)$. Recall that $\mathcal{F}_1 = \sigma(\mathcal{D}_1, \widetilde{\boldsymbol{\eta}}_n)$ denote the first-stage information, including the realized pilot. The inference sample and the fresh release noises are independent of $\mathcal{F}_1$.

$$\mathrm{Cov}\big(\widehat{\boldsymbol{\eta}}_n^{dp} \mid \mathcal{F}_1\big) = \mathbf{A}_0^{-1}\left(\frac{\mathbf{J}_0}{n_2} + \sigma_{\psi,n}^2 \mathbf{I}_{d_\theta}\right)\mathbf{A}_0^{-1} + o_p(n_2^{-1}), \tag{3}$$

where the first term is the sampling contribution and the second is the finite-sample privacy correction.

Theorem 3 has the same qualitative message as several recent Euclidean DP-inference papers (Avella-Medina et al., 2023; Lin et al., 2025; Sopa et al., 2026). The score noise is first-order negligible under the stated assumptions, while its known contribution can still be retained in the covariance approximation. Here that happens because the released score average has sensitivity $O\left(n_2^{-1}\right)$, so the Gaussian score noise is of order $O_p\left(n_2^{-1}\right)$, smaller than the sampling fluctuation $O_p\left(n_2^{-1/2}\right)$. It relies on working in a local chart on $\mathcal{TM}$, on the bounded-derivative assumptions, and on the availability of a sufficiently accurate pilot.

3.3 PRIVATE COVARIANCE ESTIMATION AND CONFIDENCE REGIONS

Recall the definition of the private covariance estimator from Algorithm 1: $\widehat{\mathbf{V}}_n^{dp} = \left(\widehat{\mathbf{H}}_n^{dp}\right)^{-1}\left(\widehat{\mathbf{M}}_n^{dp}/n_2 + \sigma_{\psi,n}^2\mathbf{I}_{d_\theta}\right)\left(\widehat{\mathbf{H}}_n^{dp}\right)^{-1}$.

Theorem 4 (Consistency of the private covariance estimator)**.** *Under the assumptions of Theorem 3,* $\widehat{\mathbf{V}}_n^{dp} - \mathbf{A}_0^{-1}\left(\mathbf{J}_0/n_2 + \sigma_{\psi,n}^2\mathbf{I}_{d_\theta}\right)\mathbf{A}_0^{-1} = o_p(n_2^{-1})$. *In particular,* $n_2\,\widehat{\mathbf{V}}_n^{dp} \xrightarrow{\mathbb{P}} \mathbf{A}_0^{-1}\mathbf{J}_0\mathbf{A}_0^{-1}$.

Corollary 1 (Private Wald region on the parameter bundle)**.** *Under the assumptions of Theorem 4, suppose additionally that* $\mathbf{J}_0$ *is positive definite. Then* $(\widehat{\boldsymbol{\eta}}_n^{dp} - \boldsymbol{\eta}_0)^\top(\widehat{\mathbf{V}}_n^{dp})^{-1}(\widehat{\boldsymbol{\eta}}_n^{dp} - \boldsymbol{\eta}_0) \Rightarrow \chi^2_{d_\theta}$. *Hence an asymptotic* $(1-\alpha)$ *private confidence region for* $\boldsymbol{\theta}_0$ *is*

$$\mathcal{C}^\theta_{1-\alpha} = \left\{\boldsymbol{\theta} \in \mathcal{U} : \left(\phi(\boldsymbol{\theta}) - \widehat{\boldsymbol{\eta}}_n^{dp}\right)^\top(\widehat{\mathbf{V}}_n^{dp})^{-1}\left(\phi(\boldsymbol{\theta}) - \widehat{\boldsymbol{\eta}}_n^{dp}\right) \le \chi^2_{d_\theta,1-\alpha}\right\}.$$

This corollary yields an ellipsoid in terms of the $\boldsymbol{\eta}$-coordinates in $\mathbb{R}^{d_\theta}$ and maps its intersection with $\phi(U)$ back via ϕ^{-1} to the parameter space $\mathcal{TM}$ of θ, obtaining a confidence region. Because ϕ is one-to-one, $\boldsymbol{\theta}_0$ belongs to this region exactly when $\boldsymbol{\eta}_0$ belongs to the coordinate ellipsoid. The inverse map can change the shape and volume without changing this coverage event. The resulting set need not be an intrinsic ellipsoid. However, a different chart can change the clipping and Gaussian perturbations, so finite-sample estimates and regions can depend on the chosen coordinates.

3.4 POINTWISE PRIVATE INFERENCE FOR THE FITTED RESPONSE

Fix $\|\boldsymbol{x}\|_2 \le B_X$ and let $\mathcal{V}_x \subset \mathcal{M}$ be a neighborhood of $\gamma_{\theta_0}(\boldsymbol{x})$ with local coordinate mapping $\zeta_x : \mathcal{V}_x \to \mathbb{R}^d$. Define the fitted-response map $g_x(\boldsymbol{\eta}) = \zeta_x\left(\gamma_{\phi^{-1}(\eta)}(\boldsymbol{x})\right)$, where the domain is restricted to a small enough neighborhood of $\boldsymbol{\eta}_0$ so that the fitted point belongs to $\mathcal{V}_x$. Let $\mathbf{G}_x = Dg_x(\boldsymbol{\eta}_0) \in \mathbb{R}^{d\times d_\theta}, \widehat{\mathbf{G}}_x = Dg_x(\widehat{\boldsymbol{\eta}}_n^{dp})$.

Theorem 5 (Delta-method CLT for fitted response points)**.** *Under the assumptions of Theorem 3,*

$$\sqrt{n_2}\left(g_x(\widehat{\boldsymbol{\eta}}_n^{dp}) - g_x(\boldsymbol{\eta}_0)\right) \Rightarrow N\left(0,\ \mathbf{G}_x\mathbf{A}_0^{-1}\mathbf{J}_0\mathbf{A}_0^{-1}\mathbf{G}_x^\top\right).$$

For $\widehat{\boldsymbol{\Sigma}}_x^{dp} = \widehat{\mathbf{G}}_x\,\widehat{\mathbf{V}}_n^{dp}\,\widehat{\mathbf{G}}_x^\top$, *one has* $\widehat{\boldsymbol{\Sigma}}_x^{dp} - \mathbf{G}_x\mathbf{A}_0^{-1}\left(\mathbf{J}_0/n_2 + \sigma_{\psi,n}^2\mathbf{I}_{d_\theta}\right)\mathbf{A}_0^{-1}\mathbf{G}_x^\top = o_p\left(n_2^{-1}\right)$. *If* $\mathbf{G}_x\mathbf{A}_0^{-1}\mathbf{J}_0\mathbf{A}_0^{-1}\mathbf{G}_x^\top$ *is positive definite, then an asymptotic* $(1-\alpha)$ *private confidence region for the response point* $\gamma_{\theta_0}(\boldsymbol{x})$ *is*

$$\mathcal{C}_{x,1-\alpha} = \left\{y \in \mathcal{V}_x : \left(\zeta_x(y) - g_x(\widehat{\boldsymbol{\eta}}_n^{dp})\right)^\top(\widehat{\boldsymbol{\Sigma}}_x^{dp})^{-1}\left(\zeta_x(y) - g_x(\widehat{\boldsymbol{\eta}}_n^{dp})\right) \le \chi^2_{d,1-\alpha}\right\}.$$

The fitted response inherits uncertainty from the parameter estimator through the Jacobian $\mathbf{G}_x$. On constant-curvature manifolds, this Jacobian can be represented by Jacobi fields along the fitted geodesic. Appendix A.2 describes the specialization to common manifolds. Corollary 1 and Theorem 5 provide analytic private inference for geodesic regression. There is no need to rely exclusively on simulation-based inferential devices such as a DP bootstrap. The term $\sigma_{\psi,n}^2\mathbf{I}_{d_\theta}$ in the covariance is the explicit penalty for privatizing the mean score. The parameter covariance can be factorized once at a cost of $O(d_\theta^3)$. Each pointwise region then requires $O(dd_\theta^2 + d^3)$ arithmetic operations, plus evaluation of the fitted point and its Jacobian. At fixed dimensions and numerical accuracy, computation grows linearly with the number of queries, with no additional privacy cost.

4 SIMULATIONS

We evaluate the performance of Algorithm 1 in two settings: geodesic regression on $\mathbb{S}^2$ and log-Euclidean geodesic regression on $\mathrm{SPD}(2)$. We consider $n \in \{100, 200, 400\}$ and $\mu \in \{0.5, 1, 2\}$, with 500 replications per configuration. We report mean geodesic curve error on $\mathcal{G}_5 = \{-1, -0.5, 0, 0.5, 1\}$, empirical coverage of the nominal 95% Wald region for the tangent-bundle parameter, and average pointwise coverage of the nominal 95% curve regions on $\mathcal{G}_5$. Our method (Ours) uses a 40/60 split and the joint KNG pilot in Appendix B, with GDP budget μ for

each stage and overall guarantee μ-GDP. The GDP budget for the inference stage is allocated proportionally to $(0.45, 0.78, 0.45)$ across the score, Hessian, and score-second-moment releases, so the Hessian release is mildly favored. We compare Ours with four methods. Full LS is the full-sample nonprivate estimator obtained by minimizing the geodesic least-squares loss, without sample splitting. No-noise uses the same private pilot, sample split, and one-step update, but computes inference from raw summaries without Gaussian perturbations, providing a benchmark for the cost of privatizing the inference stage. Seq-E and Seq-M use all observations for both the private pilot and inference, allowing us to examine the loss of data-splitting. Seq-E assigns equal pilot-inference stage budgets $(\mu/\sqrt{2}, \mu/\sqrt{2})$. Seq-M uses $(0.8\mu, 0.6\mu)$ so that its larger inference sample offsets its smaller inference budget, matching the release noise of Ours. All private inference stages use the same clipping thresholds and relative budget allocation. Data generation and implementation details are given in Appendix C.1. Appendix C.2 reports dense pointwise diagnostics, sensitivity and stress tests, and a privacy-accounting comparison. Appendices C.3–C.5 examine public warm starts, multiple predictors, and the interaction between curvature and clipping thresholds, respectively. All simulations are run in R on a 2022 MacBook Air with an Apple M2 chip and 16 GB unified memory, without GPU, cluster, or distributed computation.

Table 1 reports the results on $\mathbb{S}^2$ and log-Euclidean $\mathrm{SPD}(2)$, with the $\mathrm{SPD}(2)$ results shown in parentheses. In both settings, the curve error of Ours decreases as either n or μ increases, and the gaps from Full LS and No-noise generally narrow. With larger samples and a moderate privacy budget, our method provides reliable private inference for both the regression parameter and individual curve points, with coverage close to the nominal 95% level in both settings. The small-sample, strong-privacy regime remains difficult because the private Hessian and covariance estimates are noisier, leading to possible undercoverage. Additional robustness studies show that the performance is stable under moderate changes in the sample split ratio, clipping scale, privacy budget allocation, and error distribution, with noticeable deterioration mainly occurring under small n, small μ, contaminated noise, or a budget-allocation scheme that does not favor the Hessian.

5 Real data application

We analyze two tropical-cyclone tracks from NOAA/NCEI IBTrACS (Gahtan et al., 2024; Knapp et al., 2010): Pancho–Helinda (South Indian Ocean, 1997) and Tina (Eastern Pacific, 1992). Storm-center locations are mapped from latitude and longitude to $\mathbb{S}^2$ as $Y_i = \big(\cos(\mathrm{lat}_i)\cos(\mathrm{lon}_i),\ \cos(\mathrm{lat}_i)\sin(\mathrm{lon}_i),\ \sin(\mathrm{lat}_i)\big) \in \mathbb{S}^2$, with elapsed time since genesis as the predictor. We use a 40/60 split, the joint KNG pilot in Appendix B, and GDP budget 2 for each stage, with the inference allocation in Section 4. Time is rescaled during optimization, and shooting-vector estimates are reported in radians per hour. We compute the score and Hessian by centered finite differences and use the covariance in Algorithm 1. We treat the observations as iid for this illustration, focusing on how privacy noise affects the fitted tracks and their confidence regions. Appendix D reports the parameter estimates and confidence intervals.

Figure 3 shows that the private geodesics closely follow the No-noise fits, with wider confidence regions consistent with the covariance correction in Theorem 4. Pancho–Helinda moves mainly westward and slightly southward, whereas Tina moves westward and northward with a larger fitted angular speed. All fitted points and pointwise regions are obtained by post-processing one private parameter-and-covariance release, so additional queries incur no privacy cost.

6 Concluding remarks

We proposed a one-shot GDP framework for inference in geodesic regression with manifold-valued responses. The method works in a local chart on the parameter bundle, privatizes clipped score/Hessian/second-moment averages, and then obtains parameter and fitted-curve uncertainty by post-processing. The theory shows that the Gaussian privacy noise is first-order negligible under fixed dimension and fixed clipping levels, while the known score-noise contribution is retained explicitly within a first-order covariance approximation.

Several limitations remain. The analysis is local: it requires a chart neighborhood that avoids cut-locus singularities and supports uniform derivative bounds. The cleanest theory also assumes inactive clipping and a sufficiently accurate private or public pilot estimator. These assumptions are stan-

Table 1: Simulation results over 500 replications on $\mathbb{S}^2$, with log-Euclidean $\mathrm{SPD}(2)$ in parentheses. Nominal coverage is 0.95.

n	μ	Ours	Full LS	No-noise	Seq-E	Seq-M
(a) Mean curve error						
100	0.5	0.624 (0.724)	0.039 (0.049)	0.149 (0.171)	0.614 (0.744)	0.673 (0.802)
	1.0	0.425 (0.519)	0.040 (0.050)	0.101 (0.110)	0.388 (0.482)	0.447 (0.567)
	2.0	0.241 (0.287)	0.038 (0.050)	0.079 (0.077)	0.197 (0.246)	0.233 (0.293)
200	0.5	0.419 (0.523)	0.027 (0.034)	0.096 (0.102)	0.376 (0.491)	0.425 (0.572)
	1.0	0.238 (0.281)	0.027 (0.034)	0.080 (0.064)	0.198 (0.244)	0.226 (0.284)
	2.0	0.132 (0.148)	0.028 (0.035)	0.055 (0.049)	0.108 (0.125)	0.119 (0.145)
400	0.5	0.241 (0.293)	0.020 (0.024)	0.077 (0.058)	0.211 (0.246)	0.232 (0.297)
	1.0	0.132 (0.145)	0.019 (0.024)	0.044 (0.039)	0.117 (0.124)	0.127 (0.143)
	2.0	0.067 (0.074)	0.020 (0.024)	0.029 (0.033)	0.056 (0.062)	0.062 (0.071)
(b) Parameter coverage						
100	0.5	0.918 (0.928)	0.924 (0.886)	0.760 (0.830)	0.922 (0.886)	0.930 (0.904)
	1.0	0.936 (0.908)	0.902 (0.920)	0.778 (0.812)	0.934 (0.886)	0.924 (0.862)
	2.0	0.956 (0.940)	0.918 (0.890)	0.892 (0.904)	0.954 (0.926)	0.952 (0.934)
200	0.5	0.942 (0.894)	0.944 (0.954)	0.780 (0.938)	0.942 (0.868)	0.934 (0.856)
	1.0	0.920 (0.914)	0.928 (0.924)	0.810 (0.926)	0.932 (0.934)	0.942 (0.938)
	2.0	0.920 (0.948)	0.938 (0.934)	0.904 (0.964)	0.934 (0.944)	0.940 (0.940)
400	0.5	0.904 (0.892)	0.942 (0.952)	0.706 (0.918)	0.908 (0.908)	0.926 (0.892)
	1.0	0.918 (0.932)	0.960 (0.932)	0.872 (0.966)	0.910 (0.910)	0.928 (0.926)
	2.0	0.958 (0.936)	0.922 (0.938)	0.952 (0.968)	0.944 (0.950)	0.964 (0.958)
(c) Average pointwise coverage						
100	0.5	0.904 (0.912)	0.936 (0.918)	0.843 (0.898)	0.916 (0.910)	0.914 (0.916)
	1.0	0.902 (0.919)	0.919 (0.924)	0.850 (0.877)	0.913 (0.909)	0.907 (0.913)
	2.0	0.938 (0.934)	0.942 (0.923)	0.926 (0.938)	0.964 (0.957)	0.965 (0.964)
200	0.5	0.958 (0.956)	0.949 (0.950)	0.759 (0.787)	0.966 (0.947)	0.970 (0.963)
	1.0	0.942 (0.949)	0.938 (0.938)	0.850 (0.947)	0.948 (0.942)	0.950 (0.944)
	2.0	0.940 (0.953)	0.938 (0.938)	0.930 (0.970)	0.941 (0.948)	0.950 (0.956)
400	0.5	0.934 (0.930)	0.943 (0.954)	0.846 (0.929)	0.935 (0.933)	0.948 (0.919)
	1.0	0.930 (0.929)	0.951 (0.937)	0.903 (0.967)	0.922 (0.925)	0.935 (0.939)
	2.0	0.957 (0.959)	0.938 (0.945)	0.959 (0.971)	0.950 (0.959)	0.958 (0.964)

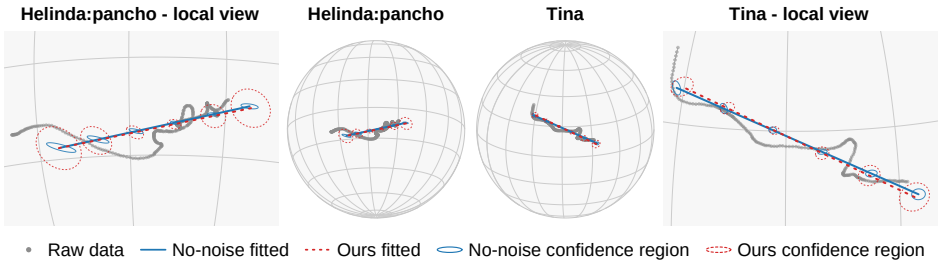


Figure 3: Sphere-valued geodesic regression for two tropical cyclone tracks. Gray points are raw storm-center locations. The blue solid curve is the No-noise fitted geodesic, and the red dashed curve is the DP fitted geodesic. Ellipses show representative pointwise 95% confidence regions on $\mathbb{S}^2$.

dard in one-step inference, but practical implementations should incorporate clipping diagnostics and robustness checks. Extensions worthy of further study include adaptive private clipping, simultaneous confidence bands for the entire geodesic, high-dimensional tangent-bundle parameters, and user-level privacy for longitudinal manifold-valued data. These problems are both interesting and challenging, and we leave them for future research.

AI USE STATEMENT

Generative AI tools were used to revise wording and improve the clarity and readability of the manuscript; to identify potentially relevant literature, references, and search directions, and to assist with research execution, including drafting, debugging, and refining code. For mathematical proofs, AI was used primarily to review and check the authors arguments for potential errors, inconsistencies, or gaps. AI-assisted manuscript revisions, literature-search suggestions, feedback on research ideas and proofs, and generated or modified code were reviewed and verified by the authors before incorporation into the final work. The authors independently assessed the accuracy and appropriateness of these outputs and retain full responsibility for the methodology, proofs, implementation, analysis, and conclusions presented in the paper.

REPRODUCIBILITY STATEMENT

The code and an accompanying `README` file are submitted with the manuscript. We provide a reproduction script for the main results, allowing the reported results to be reproduced individually. The `README` provides instructions for running the code.

REFERENCES

P.-A. Absil, R. Mahony, and R. Sepulchre. *Optimization Algorithms on Matrix Manifolds*. Princeton University Press, Princeton, NJ, 2008.

Vincent Arsigny, Pierre Fillard, Xavier Pennec, and Nicholas Ayache. Log-Euclidean metrics for fast and simple calculus on diffusion tensors. *Magnetic Resonance in Medicine*, 56(2):411–421, 2006. doi: 10.1002/mrm.20965.

Marco Avella-Medina, Casey Bradshaw, and Po-Ling Loh. Differentially private inference via noisy optimization. *The Annals of Statistics*, 51(5):2067–2092, 2023. doi: 10.1214/23-AOS2321.

Rajendra Bhatia. *Positive Definite Matrices*. Princeton University Press, Princeton, NJ, 2007.

Kamalika Chaudhuri, Claire Monteleoni, and Anand D. Sarwate. Differentially private empirical risk minimization. *Journal of Machine Learning Research*, 12:1069–1109, 2011.

Emil Cornea, Hongtu Zhu, Peter Kim, and Joseph G. Ibrahim. Regression models on Riemannian symmetric spaces. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 79(2):463–482, 2017. doi: 10.1111/rssb.12169.

Brad C. Davis, P. Thomas Fletcher, Elizabeth Bullitt, and Sarang C. Joshi. Population shape regression from random design data. *International Journal of Computer Vision*, 90(2):255–266, 2010. doi: 10.1007/s11263-010-0367-1.

Jinshuo Dong, Aaron Roth, and Weijie J. Su. Gaussian differential privacy. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 84(1):3–37, 2022. doi: 10.1111/rssb.12454.

Cynthia Dwork and Aaron Roth. The algorithmic foundations of differential privacy. *Foundations and Trends in Theoretical Computer Science*, 9(3–4):211–407, 2014. doi: 10.1561/0400000042.

Cynthia Dwork, Frank McSherry, Kobbi Nissim, and Adam Smith. Calibrating noise to sensitivity in private data analysis. In *Theory of Cryptography*, volume 3876 of *Lecture Notes in Computer Science*, pp. 265–284. Springer, 2006. doi: 10.1007/11681878_14.

Cecilia Ferrando, Shufan Wang, and Daniel Sheldon. Parametric bootstrap for differentially private confidence intervals. In *Proceedings of the Twenty Fifth International Conference on Artificial Intelligence and Statistics*, volume 151 of *Proceedings of Machine Learning Research*, pp. 1598–1618. PMLR, 2022.

P. Thomas Fletcher. Geodesic regression and the theory of least squares on Riemannian manifolds. *International Journal of Computer Vision*, 105(2):171–185, 2013. doi: 10.1007/s11263-012-0591-y.

Jennifer Gahtan, Kenneth R. Knapp, Carl J. Schreck, Howard J. Diamond, James P. Kossin, and Michael C. Kruk. International Best Track Archive for Climate Stewardship (IBTrACS) Project, Version 4r01. `https://doi.org/10.25921/82ty-9e16`, 2024.

Jacob Hinkle, P. Thomas Fletcher, and Sarang Joshi. Intrinsic polynomials for regression on Riemannian manifolds. *Journal of Mathematical Imaging and Vision*, 50(1–2):32–52, 2014. doi: 10.1007/s10851-013-0489-5.

Yangdi Jiang, Xiaotian Chang, Yi Liu, Lei Ding, Linglong Kong, and Bei Jiang. Gaussian differential privacy on Riemannian manifolds. In *Advances in Neural Information Processing Systems*, volume 36, pp. 14665–14684, 2023.

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*, pp. 2705–2712, 2014. doi: 10.1109/CVPR.2014.352.

Kenneth R. Knapp, Michael C. Kruk, David H. Levinson, Howard J. Diamond, and Charles J. Neumann. The International Best Track Archive for Climate Stewardship (IBTrACS): Unifying Tropical Cyclone Best Track Data. *Bulletin of the American Meteorological Society*, 91(3):363–376, 2010. doi: 10.1175/2009BAMS2755.1.

Aditya Kulkarni and Carlos Soto. Differentially private geodesic and linear regression. *arXiv preprint arXiv:2504.11304v1*, 2025. URL https://arxiv.org/abs/2504.11304v1. Version 1, 15 April 2025.

Lizhen Lin, Brian St. Thomas, Hongtu Zhu, and David B. Dunson. Extrinsic local regression on manifold-valued data. *Journal of the American Statistical Association*, 112(519):1261–1273, 2017. doi: 10.1080/01621459.2016.1208615.

Shurong Lin, Aleksandra Slavković, and Deekshith Reddy Bhoomireddy. Differentially private linear regression and synthetic data generation with statistical guarantees. *arXiv preprint arXiv:2510.16974*, 2025.

Marc Niethammer, Yang Huang, and François-Xavier Vialard. Geodesic regression for image time-series. In *Medical Image Computing and Computer-Assisted Intervention – MICCAI 2011*, volume 6892 of *Lecture Notes in Computer Science*, pp. 655–662. Springer, 2011. doi: 10.1007/978-3-642-23629-7_80.

Xavier Pennec, Pierre Fillard, and Nicholas Ayache. A Riemannian framework for tensor computing. *International Journal of Computer Vision*, 66(1):41–66, 2006. doi: 10.1007/s11263-005-3222-z.

Alexander Petersen and Hans-Georg Müller. Fréchet regression for random objects with Euclidean predictors. *The Annals of Statistics*, 47(2):691–719, 2019. doi: 10.1214/17-AOS1624.

Davide Pigoli, Alessandra Menafoglio, and Piercesare Secchi. Kriging prediction for manifold-valued random fields. *Journal of Multivariate Analysis*, 145:117–131, 2016. doi: 10.1016/j.jmva.2015.12.006.

Matthew Reimherr, Karthik Bharath, and Carlos Soto. Differential privacy over Riemannian manifolds. In *Advances in Neural Information Processing Systems*, volume 34, pp. 12292–12303, 2021.

Or Sheffet. Differentially private ordinary least squares. *Journal of Privacy and Confidentiality*, 9(1), 2019. doi: 10.29012/jpc.654.

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 2009*, volume 5762 of *Lecture Notes in Computer Science*, pp. 192–199. Springer, 2009. doi: 10.1007/978-3-642-04271-3_24.

Getoar Sopa, Marco Avella Medina, and Cynthia Rush. Differentially private inference for longitudinal linear regression. *arXiv preprint arXiv:2601.10626*, 2026.

Carlos Soto, Karthik Bharath, Matthew Reimherr, and Aleksandra Slavković. Shape and structure preserving differential privacy. In *Advances in Neural Information Processing Systems*, volume 35, pp. 24693–24705, 2022.

Di Xiong, Shihui Ying, and Hongtu Zhu. Intrinsic partial linear models for manifold-valued data. *Information Processing & Management*, 59(4):102954, 2022. doi: 10.1016/j.ipm.2022.102954.

Ying Yuan, Hongtu Zhu, Weili Lin, and J. S. Marron. Local polynomial regression for symmetric positive definite matrices. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 74(4):697–719, 2012. doi: 10.1111/j.1467-9868.2011.01022.x.

A ILLUSTRATIVE EXAMPLES

A.1 DETAILED EXAMPLES FOR THE ABSTRACT LOSS FUNCTION, SCORE, AND HESSIAN

Scalar geodesic regression on $\mathbb{S}^1$. Let $p(\alpha) = (\cos\alpha, \sin\alpha)$ and let $e_\alpha = (-\sin\alpha, \cos\alpha)$ be the unit tangent at $p(\alpha)$. For a scalar predictor x, write the shooting vector as $v = \beta e_\alpha$ and use the chart $\boldsymbol{\eta} = (\alpha, \beta)^\top$. Then

$$\gamma_\eta(x) = \mathrm{Exp}_{p(\alpha)}(x\beta e_\alpha) = (\cos(\alpha + x\beta), \sin(\alpha + x\beta)).$$

If $Y = (\cos U, \sin U)$, define the signed local residual

$$r(w, \boldsymbol{\eta}) = \mathrm{wrap}_\pi(U - \alpha - x\beta),$$

where wrap_π denotes the representative in $(-\pi, \pi]$. On a local chart where the residual does not cross the cut point, the squared geodesic loss is

$$\ell(w, \boldsymbol{\eta}) = \frac{1}{2} r(w, \boldsymbol{\eta})^2.$$

With $\mathbf{d}(x) = (1, x)^\top$, the derivative objects are

$$\boldsymbol{\psi}(w, \boldsymbol{\eta}) = -r(w, \boldsymbol{\eta})\mathbf{d}(x), \qquad \mathbf{H}(w, \boldsymbol{\eta}) = \mathbf{d}(x)\mathbf{d}(x)^\top, \qquad \mathbf{M}(w, \boldsymbol{\eta}) = r(w, \boldsymbol{\eta})^2 \mathbf{d}(x)\mathbf{d}(x)^\top.$$

Thus, in this chart, the score is a signed arc residual times the design vector, the Hessian is the local curvature of the squared arc loss, and the score second moment is the outer product of the individual score. At the population minimizer, $\mathbf{A}_0 = \mathbb{E}\{\mathbf{d}(X)\mathbf{d}(X)^\top\}$ and $\mathbf{J}_0 = \mathrm{Var}\{r(W, \boldsymbol{\eta}_0)\mathbf{d}(X)\}$.

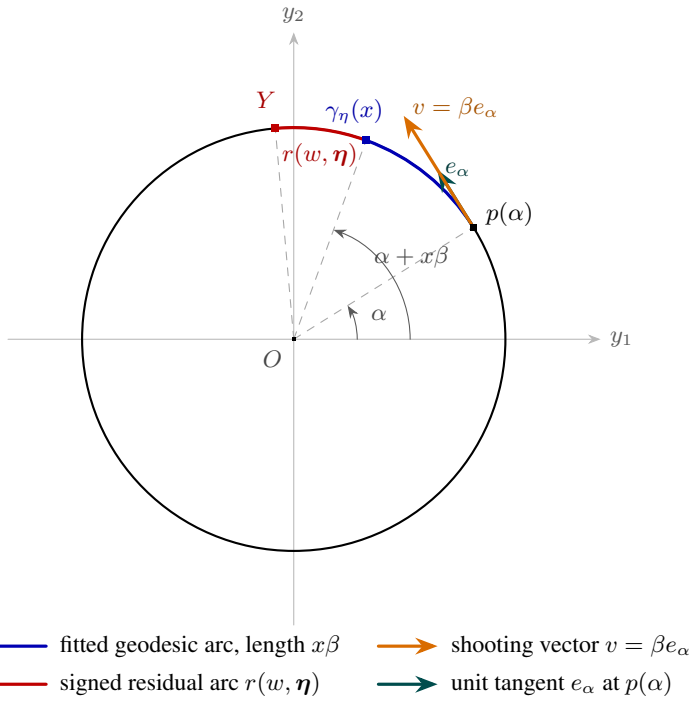


Figure A.1: Scalar geodesic regression on $\mathbb{S}^1$. The local chart $\boldsymbol{\eta} = (\alpha, \beta)^\top$ determines the footpoint $p(\alpha)$ and shooting vector $v = \beta e_\alpha$. For predictor value x, the fitted point is $\gamma_\eta(x) = \mathrm{Exp}_{p(\alpha)}(x\beta e_\alpha)$, and the signed residual $r(w, \boldsymbol{\eta}) = \mathrm{wrap}_\pi(U - \alpha - x\beta)$ is the signed arc length from the fitted point to $Y = (\cos U, \sin U)$.

Log-Euclidean $\mathrm{SPD}(q)$. Let $\mathcal{M} = \mathrm{SPD}(q)$ with the log-Euclidean metric. Put $s = q(q+1)/2$ and write $\boldsymbol{z} = \mathrm{svec}(\log Y) \in \mathbb{R}^s$. In log coordinates, the footpoint is represented by $\boldsymbol{a} \in \mathbb{R}^s$, and the

shooting block is represented by $\mathbf{B} = [\boldsymbol{b}_1, \ldots, \boldsymbol{b}_{d_x}] \in \mathbb{R}^{s \times d_x}$. Thus, $\mathrm{svec}\{\log \gamma_{(\boldsymbol{a},B)}(\boldsymbol{x})\} = \boldsymbol{a} + \mathbf{B}\boldsymbol{x}$. Let $\boldsymbol{\eta} = (\boldsymbol{a}^\top, \mathrm{vec}(\mathbf{B})^\top)^\top$, $\mathbf{Q}_{\boldsymbol{x}} = \big(\mathbf{I}_s,\ \boldsymbol{x}^\top \otimes \mathbf{I}_s\big)$. Then

$$\ell(w, \boldsymbol{\eta}) = \frac{1}{2}\|\boldsymbol{z} - \mathbf{Q}_{\boldsymbol{x}}\boldsymbol{\eta}\|_2^2 = \frac{1}{2}\|\boldsymbol{z} - \boldsymbol{a} - \mathbf{B}\boldsymbol{x}\|_2^2.$$

Therefore,

$$\boldsymbol{\psi}(w, \boldsymbol{\eta}) = -\mathbf{Q}_{\boldsymbol{x}}^\top(\boldsymbol{z} - \mathbf{Q}_{\boldsymbol{x}}\boldsymbol{\eta}), \qquad \mathbf{H}(w, \boldsymbol{\eta}) = \mathbf{Q}_{\boldsymbol{x}}^\top\mathbf{Q}_{\boldsymbol{x}}, \qquad \mathbf{M}(w, \boldsymbol{\eta}) = \boldsymbol{\psi}(w, \boldsymbol{\eta})\boldsymbol{\psi}(w, \boldsymbol{\eta})^\top.$$

Consequently, $\mathbf{A}_0 = \mathbb{E}\{\mathbf{Q}_{\boldsymbol{X}}^\top\mathbf{Q}_{\boldsymbol{X}}\}$ and $\mathbf{J}_0 = \mathrm{Var}\{\boldsymbol{\psi}(W, \boldsymbol{\eta}_0)\}$. After the logarithmic coordinate transformation, the model is exactly fixed-dimensional multivariate Euclidean linear regression.

Remark on curvature. The two examples considered in this appendix are intentionally geometrically simple. $\mathbb{S}^1$ has vanishing intrinsic Riemannian curvature and the log–Euclidean SPD manifold is globally Euclidean after the logarithm map, so the Hessian expressions reduce to design-matrix-type formulas.

A.2 Common manifolds examples

Euclidean space $\mathbb{R}^d$. Geodesic regression becomes classical linear regression:

$$\gamma_{(p,V)}(\boldsymbol{x}) = p + V\boldsymbol{x}, \qquad p \in \mathbb{R}^d, \quad V \in \mathbb{R}^{d \times d_x}.$$

There is no cut locus and the global chart is the identity. Assumption 2 reduces to boundedness of the responses and fitted values on the working set, and Assumption 3 follows from bounded predictors and compactness of the parameter set. The theory is exactly fixed-dimensional private multivariate linear-regression inference. If $D_x = (I_d, \boldsymbol{x}^\top \otimes I_d)$, then Theorem 5 uses $\mathbf{G}_x = D_x$ and the pointwise confidence region is the usual Euclidean ellipsoid for $p + V\boldsymbol{x}$.

The sphere $\mathbb{S}^d$. The geodesic is a great-circle arc

$$\gamma_{(p,v)}(\boldsymbol{x}) = \cos(\big\|\boldsymbol{x}^\top\boldsymbol{v}\big\|)\, p + \sin(\big\|\boldsymbol{x}^\top\boldsymbol{v}\big\|)\, \frac{\boldsymbol{x}^\top\boldsymbol{v}}{\|\boldsymbol{x}^\top\boldsymbol{v}\|}$$

with the usual continuous extension at $\boldsymbol{v} = 0$, where $u(\boldsymbol{x}) = \boldsymbol{x}^\top\boldsymbol{v} = \sum_{j=1}^{d_x} x_j v_j \in \mathcal{T}_p\mathbb{S}^d$. The main local restriction is that the fitted arc and the responses remain inside an open hemisphere or, more generally, inside a compact subset bounded away from the cut locus. In that regime the Jacobians $D_\theta\gamma_{\boldsymbol{\theta}}(t)$ are classical Jacobi fields with explicit sine–cosine formulas, so Assumption 3 can be verified directly. Theorem 5 then yields pointwise confidence ellipsoids for great-circle predictions.

Hadamard manifolds. If $\mathcal{M}$ is Hadamard (complete, simply connected, nonpositively curved), there is no cut locus and the exponential map is globally defined. The local geometric restrictions are therefore milder. Uniform derivative bounds and local identifiability are still needed for the inferential results.

Affine-invariant $\mathrm{SPD}(q)$. Under the affine-invariant Riemannian metric (Pennec et al., 2006; Bhatia, 2007), let $\boldsymbol{v} = (V_1, \ldots, V_{d_x})$, where $V_j \in \mathcal{T}_P\mathrm{SPD}(q)$, and put $V(\boldsymbol{x}) = \sum_{j=1}^{d_x} x_j V_j$. Then the geodesics are

$$\gamma_{(P,\boldsymbol{v})}(\boldsymbol{x}) = P^{1/2}\exp\big(P^{-1/2}V(\boldsymbol{x})P^{-1/2}\big)P^{1/2}.$$

The manifold is Hadamard, so the cut-locus issue disappears. To verify Assumption 3, one usually restricts attention to a compact spectral band, which bounds the differentials of the matrix exponential and logarithm. The fitted-curve Jacobian is again a matrix-valued Jacobi operator.

Log-Euclidean $\mathrm{SPD}(q)$. Under the log-Euclidean metric (Arsigny et al., 2006), write the shooting directions in log coordinates as symmetric matrices $\boldsymbol{v} = (B_1, \ldots, B_{d_x})$ and put $B(\boldsymbol{x}) = \sum_{j=1}^{d_x} x_j B_j$. Then

$$\gamma_{(P,\boldsymbol{v})}(\boldsymbol{x}) = \exp\{\log P + B(\boldsymbol{x})\}.$$

After vectorizing the log matrix in an orthonormal symmetric basis, this is exactly Euclidean multivariate linear regression with response $\mathrm{svec}(\log Y)$ and design $(\mathbf{I}_s, \boldsymbol{x}^\top \otimes \mathbf{I}_s)$. Thus, all of our theorems reduce to the Euclidean case after the logarithmic coordinate change.

Grassmann manifolds. On $\mathrm{Gr}(r, m)$, geodesics and Jacobi fields admit explicit matrix sine–cosine representations in terms of principal angles (Absil et al., 2008). The main technical requirement is that the fitted curve stay in a chart bounded away from the cut locus, equivalently away from principal angles equal to $\pi/2$. Within such a chart, the derivative bounds in Assumption 3 are again finite.

B Additional comments on pilot constructions

The main text deliberately treats the pilot as modular. We record here two useful options.

A private pilot We give a joint K-norm-gradient (KNG) construction in the working coordinates, using the gradient-norm mechanism underlying Kulkarni & Soto (2025). The pilot is a single draw for the full parameter, including the footpoint and all shooting vectors. Take K to be compact, with $\boldsymbol{\eta}_0 \in \mathrm{int}(K)$, and use uniform Lebesgue measure on K as the public reference distribution. Set

$$c_{0,n} = n_1^{1/4}, \qquad \epsilon_0 = \log\left\{\frac{\Phi(\mu_0/2)}{1-\Phi(\mu_0/2)}\right\},$$

where $\mu_0 > 0$ is the pilot GDP budget and Φ is the standard normal c.d.f. Draw $\widetilde{\boldsymbol{\eta}}_n$ from

$$f(\boldsymbol{\eta} \mid \mathcal{D}_1) \propto \exp\left[-\frac{\epsilon_0 n_1}{4c_{0,n}}\left\|\frac{1}{n_1}\sum_{i\in\mathcal{D}_1}\mathrm{clip}_{c_{0,n}}\{\boldsymbol{\psi}(W_i,\boldsymbol{\eta})\}\right\|_2\right], \qquad \boldsymbol{\eta}\in K. \tag{4}$$

The threshold $c_{0,n}$ is used only for the pilot and is separate from the three inference thresholds. It grows with n_1 to avoid a fixed clipping bias in the pilot score. All settings and the reference measure are fixed independently of the protected observations. At points where a score is undefined, we assign the zero vector before clipping.

Replacing one record changes the norm in (4) by at most $2c_{0,n}/n_1$. Therefore, the unnormalized densities differ by at most a factor $e^{\epsilon_0/2}$, and the normalizing constants contribute the same factor. The pilot is pure ϵ_0-DP. The trade-off formulation of privacy (Dong et al., 2022) implies that a pure ϵ-DP mechanism is $\kappa(\epsilon)$-GDP, where

$$\kappa(\epsilon) = 2\Phi^{-1}\left(\frac{e^\epsilon}{1+e^\epsilon}\right).$$

Our choice gives $\kappa(\epsilon_0) = \mu_0$. This conversion uses the pure-DP trade-off guarantee and does not fix a value of δ. Equation (4) specifies an exact sampling distribution. One exact implementation repeatedly proposes $\boldsymbol{\eta}$ uniformly on K and accepts it with probability equal to the exponential in that equation. The first accepted proposal is the pilot. Proposal counts and rejected values are not released. The acceptance probability is positive on compact K, so this procedure terminates almost surely. This reference sampler requires no initialization or burn-in, although its cost can be large as the dimension increases. A fixed-length MCMC approximation is not interchangeable with this exact mechanism for privacy accounting. For fixed dimension and fixed $\mu_0 > 0$, Assumptions 1–4 and a unique population score zero in K imply $\|\widetilde{\boldsymbol{\eta}}_n - \boldsymbol{\eta}_0\|_2 = O_p(n_1^{-1/2})$. Thus, Assumption 6 holds when $n_1/n_2 \to r \in (0,\infty)$.

A public warm start. When an atlas, baseline template, or scientifically natural reference geodesic is available, one can take a deterministic warm start and set $\mathcal{D}_1 = \varnothing$. Under this specialization the total privacy budget is simply $\sqrt{\mu_\psi^2 + \mu_H^2 + \mu_M^2}$ on the full sample. We also provide in Appendix C the simulation results for the public warm start pilot under different degrees of perturbation. The results indicate that our method is relatively robust with respect to the choice of pilot or the perturbation added.

C Comprehensive simulation studies

This appendix gives the simulation details and supplementary empirical evidence beyond the main text. The goal of the simulations is not to benchmark a new optimizer for geodesic regression, but to evaluate the private inference layer. Given an independent pilot, Algorithm 1 privatizes only three finite-dimensional averages and uses the resulting noise-aware covariance for Wald inference.

C.1 Data generation mechanisms and implementation details

For the sphere experiment, $\mathcal{M} = \mathbb{S}^2$ and $\boldsymbol{\eta}_0 = (0.6, 0.3, 0.4, -0.2)^\top$. The first two coordinates are the longitude and latitude of the footpoint, and the last two are the shooting-vector coefficients in the associated orthonormal tangent basis. We generate independent predictors $X_i \sim \mathrm{Unif}[-1, 1]$ and responses

$$Y_i = \mathrm{Exp}_{\gamma_{\boldsymbol{\eta}_0}(X_i)}(\Xi_i), \qquad \Xi_i \sim N(0, 0.2^2 \mathbf{I}_2),$$

where the Gaussian noise is expressed in an orthonormal basis of the tangent plane at the true curve point. For the log-Euclidean experiment, $\mathcal{M} = \mathrm{SPD}(2)$. We represent symmetric matrices by

$$\mathrm{svec}\begin{pmatrix} a & b \\ b & c \end{pmatrix} = (a, c, \sqrt{2}b)^\top.$$

The true intercept and slope in log coordinates are $\boldsymbol{p}_0 = (0.2, -0.1, 0.12)^\top$ and $\boldsymbol{v}_0 = (0.3, -0.2, 0.08)^\top$, respectively. We generate $X_i \sim \mathrm{Unif}[-1, 1]$ and

$$\mathrm{svec}(\log Y_i) = \boldsymbol{p}_0 + X_i \boldsymbol{v}_0 + \varepsilon_i, \qquad \varepsilon_i \sim N(0, 0.2^2 \mathbf{I}_3).$$

Under the log-Euclidean metric, this is multivariate Euclidean linear regression in log coordinates. Thus, it provides a matrix-manifold sanity check for the general theory.

We use the following public parameter domains:

$$\begin{aligned} &K = [0, 1.2] \times [-0.3, 0.9] \times [-0.5, 0.5]^{2d_x} && \text{on } \mathbb{S}^2, \\ &K = [-0.5, 0.5]^{3(d_x+1)} && \text{in log-Euclidean coordinates.} \end{aligned}$$

Here $d_x = 1$ for the main experiments. For multiple predictors, the shooting-vector coordinates are concatenated in the same order as the predictors.

The main experiments use $n \in \{100, 200, 400\}$ and $\mu \in \{0.5, 1, 2\}$, with 500 replications per setting. In each replication, we randomly split the observations, independently of their values, into $\mathcal{D}_1$ and $\mathcal{D}_2$, with $n_1 = \lfloor 0.4n \rfloor$ and $n_2 = n - n_1$. We draw the pilot from (4), using $c_{0,n} = n_1^{1/4}$, $\mu_0 = \mu$, and the rejection sampler described in Appendix B. The footpoint and all shooting-vector coordinates are sampled jointly. We compute the pilot score by automatic differentiation on $\mathbb{S}^2$ and analytically in log-Euclidean coordinates. The inference stage uses $\mathcal{D}_2$ and a GDP budget μ. Its score, Hessian, and second-moment budgets are proportional to $(0.45, 0.78, 0.45)$, normalized to have squared sum μ^2. Since the two stages use disjoint observations, Ours satisfies μ-GDP. For the inference score and Hessian on $\mathbb{S}^2$, we use centered finite differences with steps 10^{-5} and 10^{-4}, respectively. Both derivatives are computed analytically in the log-Euclidean experiment. The inference clipping levels are $(c_\psi, c_H, c_M) = (2.50, 3.00, 2.32)$ on $\mathbb{S}^2$ and $(2.50, 3.50, 3.87)$ on $\mathrm{SPD}(2)$. The ridge is $6/n_{\mathrm{infer}}$ in both settings. For Ours and No-noise, $6/n_{\mathrm{infer}} = n_2$. For Full LS, Seq-E and Seq-M, $6/n_{\mathrm{infer}} = n$.

Full LS minimizes the geodesic least-squares loss over K using the full sample. No-noise uses the same split and realized pilot as Ours, but computes the inference averages without clipping or Gaussian noise and uses the usual sandwich covariance. Seq-E and Seq-M use the full sample for both pilot estimation and inference. Their pilots follow (4), with n_1 replaced by n throughout, including $c_{0,n} = n^{1/4}$. The pilot and inference GDP budgets are $(\mu/\sqrt{2}, \mu/\sqrt{2})$ for Seq-E and $(0.8\mu, 0.6\mu)$ for Seq-M. Within each inference stage, we use the same relative budget allocation and clipping levels as Ours. GDP composition gives an overall budget μ for both full-sample variants. The split-sample asymptotic results concern Ours.

We report mean curve error on $\mathcal{G}_5 = \{-1, -0.5, 0, 0.5, 1\}$, parameter-region coverage, and average pointwise coverage on $\mathcal{G}_5$. All confidence regions have nominal coverage 95%. The Gaussian experiments assess finite-sample behavior beyond the bounded-support assumptions of the main theory. Appendix F discusses inference with active clipping.

To examine the robustness of the algorithm, we conduct the following additional analyses. Unless otherwise stated, supplementary experiments use 100 replications at $(n, \mu) \in \{(100, 0.5), (200, 1), (400, 2)\}$.

Dense pointwise diagnostics. To examine how coverage varies along the predictor axis rather than only at the five summary points, we evaluate pointwise coverage on the dense grid $\mathcal{G}_{50} = \{-1 = t_1 < t_2 < \cdots < t_{50} = 1\}$ of 50 equally spaced points. This experiment tests whether the pointwise regions remain well calibrated across the full fitted curve.

Sensitivity to the $\mathcal{D}_1/\mathcal{D}_2$ split. We vary the pilot fraction over $|\mathcal{D}_1|/n \in \{0.20, 0.30, 0.40, 0.50\}$, corresponding to 20/80, 30/70, 40/60, and 50/50 splits, while keeping all other settings fixed. This experiment probes the trade-off between pilot accuracy and inference-stage sample size.

Sensitivity to clipping thresholds. Starting from the baseline clipping triplet (c_ψ, c_H, c_M), we multiply all three clipping thresholds by $0.85, 1.00, 1.5, 2.0$. This experiment assesses the robustness of estimation and coverage to moderate under- or over-calibration of clipping.

Sensitivity to the inference-stage budget allocation. We compare five relative allocations of the inference-stage budget across the score, Hessian, and score-second-moment releases:

$$(0.45, 0.78, 0.45),\ (0.58, 0.58, 0.58),\ (0.90, 0.35, 0.35),\ (0.35, 0.90, 0.35),\ (0.35, 0.35, 0.90),$$

each normalized to have unit ℓ_2 norm before multiplying by μ. These correspond, respectively, to the main experiment allocation, a balanced allocation, and score-heavy, Hessian-heavy, and moment-heavy allocations. The goal here is to identify which private release is most influential for estimation and uncertainty.

Geometry and noise stress tests. Finally, we compare four stress-test scenarios on each manifold: main experiment setting, hard geometry, contaminated noise, and heavy-tailed noise. In the hard-geometry scenario, we increased both the predictor range and the signal strength by taking $X_i \sim \mathrm{Unif}[-1.4, 1.4]$. In the contaminated-noise scenario, each noise draw was replaced with probability 0.1 by a Gaussian draw with four times the baseline standard deviation. In the heavy-tailed scenario, the noise was sampled from a t_5 distribution rescaled to match the baseline variance. These experiments probe robustness to stronger curvature, larger extrapolation ranges, and departures from the nominal Gaussian model.

C.2 Results

Figure A.2 evaluates pointwise coverage on a dense grid of 50 equally spaced values in $[-1, 1]$ for the representative configuration $(n, \mu) = (400, 2)$. Both manifolds show stable coverage across the predictor range, with no persistent undercoverage at the endpoints or the center. This supports the use of a single parameter covariance and the fitted-curve Jacobian $\mathbf{G}_x$ to propagate uncertainty along the whole curve. Figure A.3 compares empirical coverage with the average size of the corresponding confidence regions. For a parameter covariance matrix $\widehat{\mathbf{V}}$, we report the diameter-style quantity $L_\theta(\widehat{\mathbf{V}}) = 2\left\{\chi^2_{d_\theta,1-\alpha}\, d_\theta^{-1}\,\mathrm{tr}(\widehat{\mathbf{V}})\right\}^{1/2}$. For a pointwise fitted-curve region at predictor value $\boldsymbol{x}$, with covariance $\widehat{\boldsymbol{\Sigma}}_x = \widehat{\mathbf{G}}_x \widehat{\mathbf{V}} \widehat{\mathbf{G}}_x^\top$, we use $L_x(\widehat{\boldsymbol{\Sigma}}_x) = 2\left\{\chi^2_{d,1-\alpha}\, d^{-1}\,\mathrm{tr}(\widehat{\boldsymbol{\Sigma}}_x)\right\}^{1/2}$, and average this quantity over $\mathcal{G}_5$. The results show that the Ours achieves the nominal coverage rate with longer confidence regions than the No-noise method. This is reasonable, as the score-noise term is explicitly included in $\widehat{\mathbf{V}}_n^{dp}$. Thus, the finite-sample effect of privacy is primarily an uncertainty inflation rather than a loss of accuracy.

Figures A.4 and A.5 show that the proposed DP inference procedure is broadly robust to moderate changes in the pilot–inference split, clipping scale, budget allocation, and noise specification. Across both manifolds, the main empirical patterns remain stable once the sample size or privacy budget is not too small. In the sample split experiment, very small pilot fractions can destabilize the pilot, whereas very large pilot fractions reduce the effective inference sample size. The 40/60 split used in the main text is a stable compromise. Performance is also fairly insensitive to moderate clipping perturbations, indicating that the method is not finely tuned to a single threshold choice. Among the budget-allocation schemes, under-budgeting the Hessian is the most detrimental, which is consistent with its dual role in the one-step update and the sandwich covariance estimator. Therefore, we recommend allocating the privacy budget in a way that mildly favors Hessian. The stress

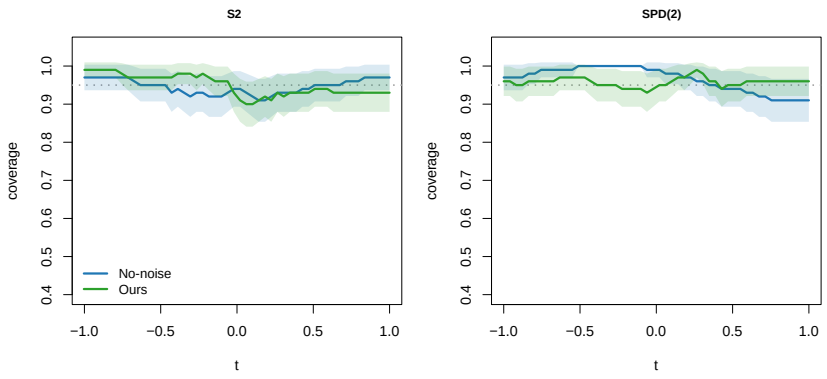


Figure A.2: Dense pointwise coverage of nominal 95% fitted-curve regions on a 50-point grid for $(n, \mu) = (400, 2)$. The left panel shows the $\mathbb{S}^2$ experiment and the right panel shows the log-Euclidean $\mathrm{SPD}(2)$ experiment. The horizontal reference line marks the nominal 95% level. Shaded bands indicate Monte Carlo standard-error bands.

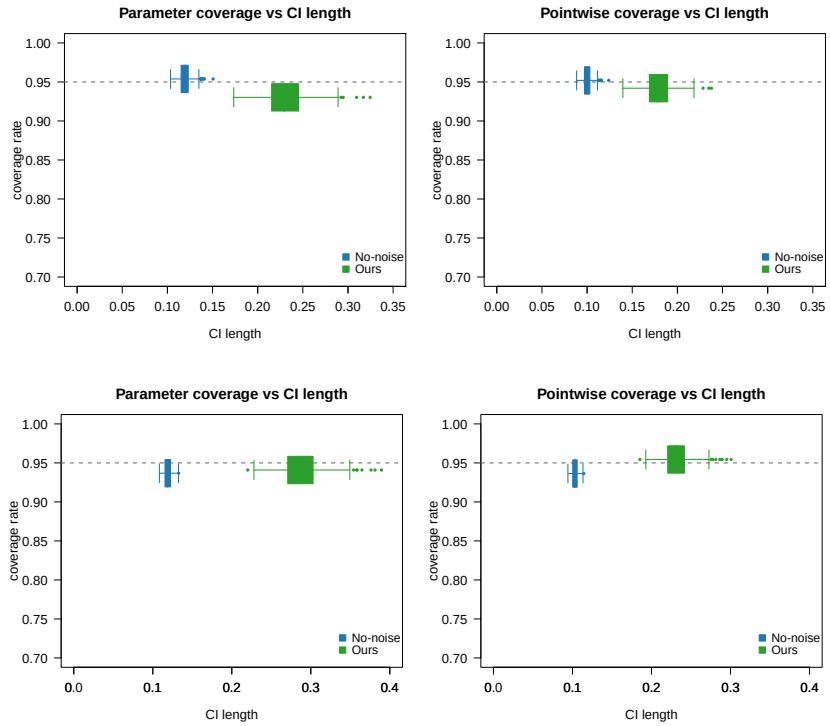


Figure A.3: Coverage–size comparison at $(n, \mu) = (400, 2)$. The top row corresponds to the $\mathbb{S}^2$ experiment and the bottom row to the log-Euclidean $\mathrm{SPD}(2)$ experiment. The horizontal reference line marks the nominal 95% coverage level. The horizontal axis summarizes the average length of the corresponding confidence regions.

tests can make the pilot and the finite-difference Hessian become less stable, while the error and coverage rate can return to nominal levels as n or μ increases. Overall, these experiments support the practical robustness of our method while suggesting that a Hessian-favored allocation may improve performance.

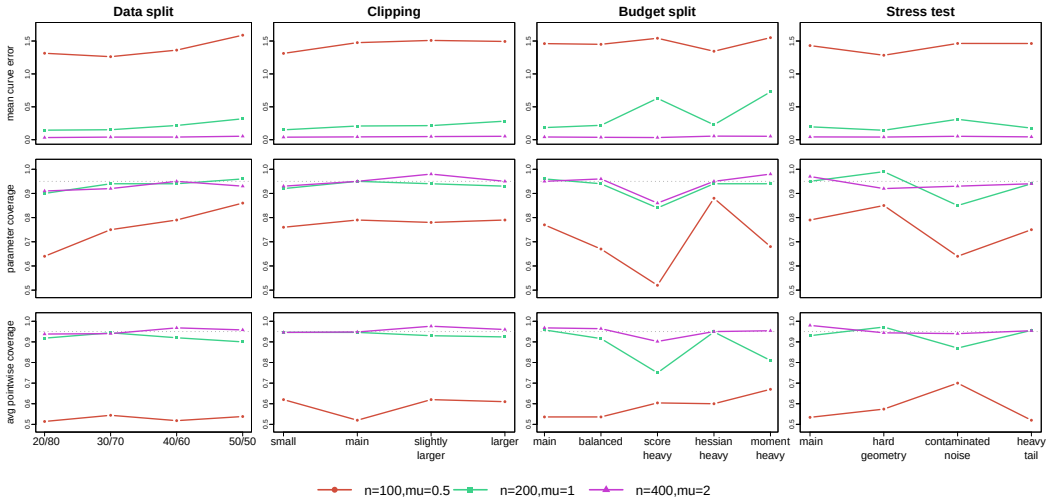


Figure A.4: Sensitivity analyses for the $\mathbb{S}^2$ simulation. Columns correspond to the pilot–inference split, clipping scale, inference-stage budget allocation, and stress-test scenario. Rows report mean curve error, parameter coverage, and average pointwise coverage.

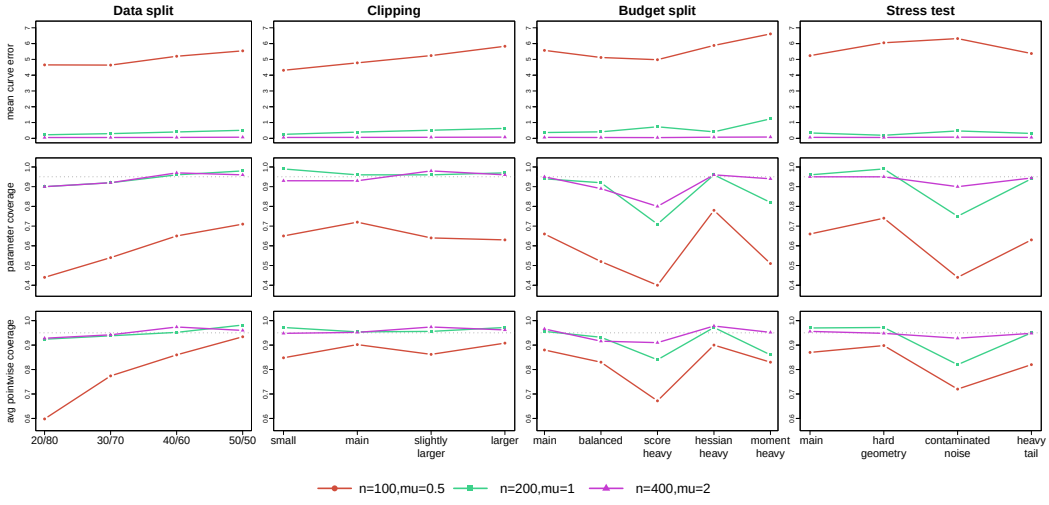


Figure A.5: Sensitivity analyses for the log-Euclidean $\mathrm{SPD}(2)$ simulation. Columns correspond to the pilot–inference split, clipping scale, inference-stage budget allocation, and stress-test scenario. Rows report mean curve error, parameter coverage, and average pointwise coverage.

Finally, Figure A.6 illustrates the privacy-accounting advantage of releasing the tangent-bundle parameter once. If instead one privatized N fitted points separately and split a fixed GDP budget equally across the N releases, GDP composition would assign budget $\mu/\sqrt{N}$ to each pointwise release, so the Gaussian noise standard deviation for each query would grow as $\sqrt{N}$. Algorithm 1 avoids this deterioration. It releases a finite-dimensional parameter and covariance once, and all

subsequent fitted values and confidence regions are post-processing. The privacy cost is therefore independent of the number of queried time points.

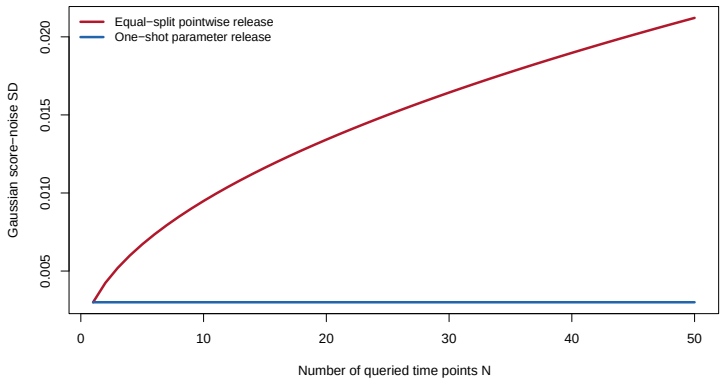


Figure A.6: Deterministic GDP accounting comparison between equal-budget pointwise releases and the one-shot parameter release used by Algorithm 1.

C.3 Additional simulation based on public warm start pilot

In this subsection, we complement the private-pilot experiments in Section C.1 by considering a deterministic public warm start, and assess finite-sample robustness to pilot misspecification. The data generation mechanisms on $\mathbb{S}^2$ and log-Euclidean SPD(2) are exactly the same as in Section C.1. For each manifold, the public pilot takes the deterministic form

$$\widetilde{\boldsymbol{\eta}} = \boldsymbol{\eta}_0 + a\boldsymbol{\delta},$$

where $\boldsymbol{\eta}_0$ is the true parameter, $\boldsymbol{\delta}$ is a fixed public offset, and $a \geq 0$ controls the degree of pilot misspecification. The main public-warm-start experiment uses $a = 1$. For the sphere experiment, we take $\boldsymbol{\delta}_{\mathbb{S}^2} = (0.15, -0.10, 0.10, -0.05)^\top$, and for the log-Euclidean $\mathrm{SPD}(2)$ experiment, we take $\boldsymbol{\delta}_{\mathrm{SPD}} = (0.05, -0.025, 0.03, 0.075, -0.05, 0.02)^\top$. The offsets are fixed public perturbations of moderate magnitude. Their errors do not vanish with n, so they violate the convergence requirements in Assumption 6. The pilot sensitivity experiment varies the offset scale over $a \in \{0.5, 1.0, 1.5, 2.0\}$. This construction yields increasingly misspecified public pilots.

We set $\mathcal{D}_1 = \varnothing$ and $\mathcal{D}_2 = \mathcal{D}_n$, so all methods use the full sample for inference. The public pilot consumes no privacy budget, allowing the full budget μ to be allocated to inference. Full MLE is unaffected by the pilot. Ours, Seq-E, and Seq-M coincide. The allocation of the inference budget across the three releases is unchanged. The clipping constants are fixed at $(c_\psi, c_H, c_M) = (1.00, 4.00, 1.20)$ on $\mathbb{S}^2$ and $(1.00, 5.20, 1.50)$ on $\mathrm{SPD}(2)$. The main public-warm-start experiment uses the full grid $n \in \{100, 200, 400\}$ and $\mu \in \{0.5, 1, 2\}$ with 500 Monte Carlo replications. The pilot-quality sensitivity study is reported at the representative configurations $(n, \mu) \in \{(100, 0.5), (200, 1), (400, 2)\}$ with 100 Monte Carlo replications.

Table A.1 presents results consistent with Table 1. As n and μ increase, the error of the DP method decreases to a level comparable to that of the Full MLE method, and the coverage rates attain the nominal level. Figure A.7 examines finite-sample sensitivity to fixed public offsets, indicating that the algorithm is robust to mild pilot misspecification.

C.4 Inference with multiple predictors

We examine whether the private Wald regions remain useful as the number of predictors increases. With multiple predictors, the regression parameter consists of a footpoint and one tangent shooting vector for each predictor. With more predictors, more perturbed quantities enter estimation and inference. We assess the resulting effect through prediction error, empirical coverage rate, and region size.

Table A.1: Main public warm-start results with $a = 1$ on $\mathbb{S}^2$, with log-Euclidean $\mathrm{SPD}(2)$ in parentheses. Nominal coverage is 0.95.

n	μ	Ours	Full LS	No-noise
(a) Mean curve error				
100	0.5	0.882 (1.097)	0.040 (0.049)	0.040 (0.045)
	1.0	0.227 (0.280)	0.039 (0.049)	0.039 (0.045)
	2.0	0.084 (0.093)	0.038 (0.049)	0.039 (0.045)
200	0.5	0.219 (0.308)	0.028 (0.035)	0.028 (0.032)
	1.0	0.076 (0.088)	0.027 (0.034)	0.028 (0.032)
	2.0	0.044 (0.050)	0.027 (0.034)	0.028 (0.032)
400	0.5	0.078 (0.084)	0.020 (0.024)	0.020 (0.024)
	1.0	0.040 (0.044)	0.019 (0.025)	0.020 (0.024)
	2.0	0.026 (0.030)	0.019 (0.024)	0.020 (0.024)
(b) Parameter coverage				
100	0.5	0.932 (0.952)	0.916 (0.902)	0.962 (0.926)
	1.0	0.962 (0.948)	0.910 (0.898)	0.962 (0.920)
	2.0	0.940 (0.936)	0.932 (0.902)	0.976 (0.922)
200	0.5	0.936 (0.942)	0.948 (0.930)	0.972 (0.942)
	1.0	0.948 (0.932)	0.944 (0.930)	0.982 (0.938)
	2.0	0.948 (0.944)	0.954 (0.938)	0.972 (0.946)
400	0.5	0.920 (0.940)	0.930 (0.946)	0.960 (0.930)
	1.0	0.928 (0.926)	0.948 (0.952)	0.968 (0.928)
	2.0	0.934 (0.922)	0.954 (0.938)	0.984 (0.922)
(c) Average pointwise coverage				
100	0.5	0.628 (0.958)	0.925 (0.920)	0.967 (0.935)
	1.0	0.933 (0.965)	0.932 (0.920)	0.967 (0.935)
	2.0	0.957 (0.957)	0.941 (0.923)	0.971 (0.938)
200	0.5	0.918 (0.942)	0.941 (0.934)	0.968 (0.945)
	1.0	0.941 (0.938)	0.952 (0.938)	0.980 (0.946)
	2.0	0.948 (0.938)	0.944 (0.940)	0.971 (0.952)
400	0.5	0.936 (0.948)	0.943 (0.947)	0.966 (0.937)
	1.0	0.945 (0.937)	0.952 (0.946)	0.966 (0.940)
	2.0	0.952 (0.934)	0.946 (0.943)	0.962 (0.935)

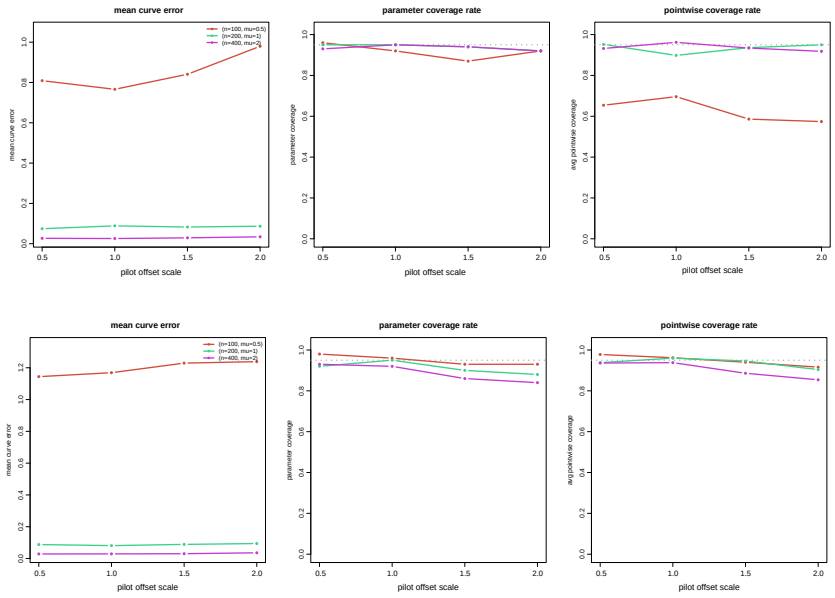


Figure A.7: Sensitivity to the pilot offset scale. $\mathbb{S}^2$ is shown on the top. $\mathrm{SPD}(2)$ is shown on the bottom.

We consider $\mathbb{S}^2$ and log-Euclidean $\mathrm{SPD}(2)$ with $d_x \in \{5, 10\}$, $n \in \{400, 800\}$, and $\mu \in \{1, 2\}$, using 500 replications per configuration. The corresponding parameter dimensions are 12 and 22 on $\mathbb{S}^2$, and 18 and 33 on $\mathrm{SPD}(2)$. We use the joint pilot, 40/60 split, and stage budgets in Appendix C.1, with all shooting vectors included in the same pilot draw. Clipping thresholds are fixed using public calibration, and the Full LS serves as the nonprivate comparator. We also examine three inference-budget allocations and four ridge values. Tables A.2 and A.3 report results with inference budgets $(\mu_\psi, \mu_H, \mu_M)/\mu \approx (0.45, 0.78, 0.45)$ and $\lambda_n = 0.10$. All regions have nominal coverage 95%. Region size is measured by the mean Wald semi-axis length in the working coordinates, averaged over axes and replications. Pointwise lengths are also averaged over the evaluation points.

Table A.2: Multiple-predictor results on $\mathbb{S}^2$ over 500 replications. Each entry is Ours / Full LS.

			Mean curve	Coverage		Mean semi-axis	
d_x	n	μ	error	Parameter	Pointwise	Parameter	Pointwise
5	400	1	0.160/0.034	0.942/0.938	0.958/0.945	0.334/0.071	0.322/0.066
5	400	2	0.076/0.034	0.918/0.884	0.956/0.944	0.168/0.072	0.156/0.066
5	800	1	0.072/0.024	0.914/0.926	0.946/0.939	0.152/0.051	0.140/0.046
5	800	2	0.045/0.024	0.858/0.950	0.920/0.947	0.089/0.051	0.082/0.047
10	400	1	0.466/0.046	0.968/0.892	0.967/0.945	0.825/0.093	0.957/0.090
10	400	2	0.157/0.046	0.956/0.896	0.971/0.941	0.329/0.093	0.345/0.090
10	800	1	0.148/0.033	0.958/0.920	0.968/0.943	0.295/0.066	0.303/0.063
10	800	2	0.071/0.033	0.898/0.930	0.965/0.946	0.150/0.066	0.148/0.064

Table A.3: Multiple-predictor results on log-Euclidean $\mathrm{SPD}(2)$ over 500 replications. Each entry is Ours / Full LS.

			Mean curve	Coverage		Mean semi-axis	
d_x	n	μ	error	Parameter	Pointwise	Parameter	Pointwise
5	400	1	0.228/0.045	0.906/0.914	0.956/0.938	0.409/0.086	0.454/0.078
5	400	2	0.102/0.045	0.912/0.908	0.969/0.938	0.188/0.086	0.185/0.078
5	800	1	0.097/0.032	0.922/0.924	0.951/0.944	0.174/0.061	0.163/0.055
5	800	2	0.054/0.032	0.846/0.934	0.942/0.945	0.096/0.061	0.092/0.055
10	400	1	0.519/0.061	0.914/0.832	0.968/0.936	0.922/0.113	1.054/0.106
10	400	2	0.243/0.061	0.960/0.844	0.976/0.944	0.442/0.113	0.469/0.106
10	800	1	0.229/0.043	0.946/0.916	0.968/0.950	0.368/0.080	0.408/0.075
10	800	2	0.074/0.043	0.898/0.902	0.964/0.947	0.166/0.080	0.165/0.075

The two manifolds show the same main pattern. At fixed n and μ, increasing the number of predictors increases the prediction error and region sizes, which can be mitigated by increasing n. Pointwise coverage rate is generally close to or above the nominal level. Thus, more predictors mainly reduce the precision of pointwise inference rather than consistently lowering its coverage. Joint parameter coverage is less stable: at $(n, \mu) = (800, 2)$, it ranges from 0.846 to 0.898 across the two manifolds and predictor dimensions. The additional budget-allocation and ridge comparisons suggest that the main bottleneck in moderate to high dimensions is inverting the noisy Hessian. Ridge values of 0.05 or 0.10 improve numerical stability relative to the smaller values considered. We recommend a slightly larger ridge and reporting the released Hessian's condition number and Wald semi-axis lengths as stability diagnostics.

C.5 Curvature and clipping thresholds

We examine how curvature affects the choice of clipping thresholds and the resulting estimation and coverage. Curvature enters indirectly through the derivatives of the geodesic loss, changing the distributions of $\|\psi\|_2$, $\|\mathbf{H}\|_F$ and $\|\mathbf{M}\|_F$ and hence the amount of clipping at a fixed threshold. We study spheres of radius $R \in \{2, 1, 0.5\}$, with curvature $K = R^{-2} \in \{0.25, 1, 4\}$, using $n = 400$, $\mu = 2$, and 500 replications per configuration. We keep the predictor distribution and all signal and noise fixed in geodesic-distance units across curvature levels. An independent public reference sample of 5000 observations at $K = 1$ is used to calibrate the score and Hessian thresholds at the

empirical 0.95-quantiles of $\|\psi_i\|_2$ and $\|\mathbf{H}_i\|_F$, respectively. We jointly multiply these reference thresholds by $a \in \{0.5, 0.75, 1, 1.25, 1.5, 2\}$ and use the resulting thresholds at all three curvatures. Increasing a reduces clipping but increases the standard deviations of the corresponding Gaussian perturbations in proportion to a. In addition to private curve error and coverage, we report Δ_{clip}, the mean geodesic curve error of the clipped estimator minus that of the unclipped estimator, with inference-stage Gaussian perturbations set to zero in both. This measures the change in estimation error due to clipping before noise is added. Positive values indicate increased error.

Table A.4: Curvature and clipping results over 500 replications.

K	a	Δ_{clip}	Mean curve error	Parameter coverage	Average pointwise coverage
0.25	0.50	0.015	0.035	0.930	0.920
	0.75	0.002	0.023	0.970	0.969
	1.00	0.000	0.023	0.962	0.960
	1.25	0.000	0.026	0.954	0.951
	1.50	0.000	0.028	0.936	0.948
	2.00	0.000	0.034	0.920	0.948
1	0.50	0.012	0.033	0.916	0.929
	0.75	0.001	0.024	0.956	0.961
	1.00	0.000	0.024	0.948	0.958
	1.25	0.000	0.027	0.932	0.953
	1.50	0.000	0.029	0.930	0.948
	2.00	0.000	0.035	0.928	0.942
4	0.50	0.006	0.033	0.924	0.919
	0.75	-0.001	0.027	0.942	0.935
	1.00	0.000	0.029	0.942	0.940
	1.25	0.000	0.032	0.928	0.934
	1.50	0.000	0.034	0.914	0.931
	2.00	0.000	0.040	0.908	0.940

Table A.4 shows a similar tradeoff at all three curvatures. At $a = 0.5$, clipping increases curve error, and both parameter and pointwise coverage fall below nominal. The scales $a = 0.75$–1 give the smallest private errors and coverage close to the nominal level. The favorable threshold range does not change much with curvature. We further compare the common public thresholds with thresholds recalibrated separately at each K. The common/recalibrated curve errors are $0.023/0.024$, $0.024/0.024$, and $0.029/0.029$ for $K = 0.25, 1, 4$, respectively, with little change in coverage. Thus, separate calibration by curvature provides little benefit in these settings.

D Real Data Analysis Details

This appendix contains the summaries for the storm-track application. The response is represented on $\mathbb{S}^2$, and the fitted curves and pointwise uncertainty summaries are post-processing of a single private parameter-and-covariance release.

For the IBTrACS storm-track analysis in Section 5, after sorting by elapsed time and retaining complete records in the required location and unit-vector fields, Pancho–Helinda has $n = 350$ observations and Tina has $n = 195$ observations. Let h_i denote elapsed hours from storm genesis. The fitted model uses $t_i = h_i/h_{\max}$, with $h_{\max} = 774$ hours for Pancho–Helinda and $h_{\max} = 582$ hours for Tina. The time origin and observation window, including $h_{\max}$, are fixed public design constants. The reported shooting-vector coordinates are converted to radians per hour by dividing the scaled-time coefficients by $h_{\max}$. We use the 40/60 split and joint KNG pilot in Appendix B, with $c_{0,n} = n_1^{1/4}$ and $\mu_0 = 2$. The inference stage uses a total GDP budget $\mu = 2$, with normalized allocation proportional to $(0.45, 0.78, 0.45)$ across the score, Hessian, and score-second-moment releases. The fixed clipping constants are $(c_\psi, c_H, c_M) = (1.32, 4.00, 2.32)$. Centered finite differences use steps 10^{-5} for the score and 10^{-4} for the Hessian. The ridge added to the one-step Hessian is 10^{-3}. The private and No-noise analyses use the same data split, pilot, and iid approximation. This common setup allows us to assess how privatizing the inference summaries changes the fitted tracks and confidence regions. We parameterize the footpoint as

$$p(\lambda, \varphi) = \big(\cos\varphi\cos\lambda,\ \cos\varphi\sin\lambda,\ \sin\varphi\big),$$

where λ is longitude and φ is latitude. At $p(\lambda, \varphi)$ we use the orthonormal tangent basis

$$e_\lambda = (-\sin\lambda, \cos\lambda, 0), \qquad e_\varphi = (-\sin\varphi\cos\lambda, -\sin\varphi\sin\lambda, \cos\varphi),$$

and write the shooting vector as $v = ae_\lambda + be_\varphi$. Table A.5 reports the estimates and Wald intervals for (λ, φ, a, b). The Ours and No-noise point estimates are close, whereas our intervals are wider because they include the explicit privacy-noise contribution. For Pancho–Helinda, our footpoint is approximately $(105.4°, -9.6°)$ and the fitted coefficients a and b are negative, indicating a westward and slightly southward motion. For Tina, $a < 0$ and $b > 0$, indicating a westward and northward motion. The fitted angular speed is about $0.0526°$ per hour for Pancho–Helinda under our fit and $0.0773°$ per hour for Tina, so Tina is faster in angular units.

Table A.5: Parameter estimates and 95% Wald confidence intervals for the two IBTrACS storm tracks.

Storm	Parameter	Method	Estimate (95% CI)
Pancho–Helinda	λ	No-noise	105.108 ($[103.602,\ 106.614]$)
		Ours	105.430 ($[101.968,\ 108.892]$)
	φ	No-noise	-9.291 ($[-9.671,\ -8.911]$)
		Ours	-9.636 ($[-12.714,\ -6.559]$)
	a	No-noise	-8.62 ($[-9.47,\ -7.78]$)
		Ours	-8.82 ($[-10.39,\ -7.25]$)
	b	No-noise	-2.54 ($[-2.74,\ -2.34]$)
		Ours	-2.53 ($[-3.90,\ -1.15]$)
Tina	λ	No-noise	-96.633 ($[-97.637,\ -95.629]$)
		Ours	-97.386 ($[-99.604,\ -95.169]$)
	φ	No-noise	7.997 ($[7.246,\ 8.748]$)
		Ours	7.607 ($[5.675,\ 9.539]$)
	a	No-noise	-12.17 ($[-12.49,\ -11.85]$)
		Ours	-11.67 ($[-12.70,\ -10.65]$)
	b	No-noise	6.66 ($[6.16,\ 7.15]$)
		Ours	6.76 ($[5.90,\ 7.62]$)

E TECHNICAL PROOFS

In this section we give full proofs of all stated results. The pilot is projected onto the publicly fixed closed convex set K before derivative evaluation, so $\widetilde{\boldsymbol{\eta}}_n \in K$ for every realization. Since $\boldsymbol{\eta}_0 \in K$, this projection cannot increase the pilot error and therefore preserves the rates in Assumption 6. Exact one-step-update identities below apply to $\widehat{\boldsymbol{\eta}}^{dp}_{n,\mathrm{raw}}$. Then, we transfer the conclusions to $\widehat{\boldsymbol{\eta}}^{dp}_{n} = \Pi_K(\widehat{\boldsymbol{\eta}}^{dp}_{n,\mathrm{raw}})$. All norms on vectors are Euclidean and all norms on matrices are Frobenius unless otherwise stated.

E.1 AUXILIARY LEMMAS

Lemma 1. *The clipping map $\boldsymbol{x} \mapsto \mathrm{clip}_c(\boldsymbol{x})$ is 1-Lipschitz in Euclidean norm, the matrix clipping map $\mathbf{A} \mapsto \mathrm{clip}_c(\mathbf{A})$ is 1-Lipschitz in Frobenius norm, and the projection map $\mathbf{B} \mapsto \Pi_{\mathbb{S}^{d_\theta}_+}(\mathrm{sym}(\mathbf{B}))$ is 1-Lipschitz in Frobenius norm.*

Proof. The vector and matrix clipping maps are radial projections onto closed Euclidean/Frobenius balls, and orthogonal projection onto a closed convex set in a Hilbert space is nonexpansive. The set of symmetric positive-semidefinite matrices is a closed convex cone in the Hilbert space of symmetric matrices equipped with the Frobenius inner product. Therefore the projection map is nonexpansive. Since symmetrization itself is linear with operator norm at most one in Frobenius norm, the composition $\mathbf{B} \mapsto \Pi_{\mathbb{S}^{d_\theta}_+}(\mathrm{sym}(\mathbf{B}))$ is also nonexpansive. □

Lemma 2. *With the noise scales in Algorithm 1,*

$$\|\boldsymbol{Z}_{\psi,n}\|_2 = O_p(n_2^{-1}), \qquad \|\mathbf{Z}_{H,n}\|_F = O_p(n_2^{-1}), \qquad \|\mathbf{Z}_{M,n}\|_F = O_p(n_2^{-1}).$$

Proof. Each noise term is centered Gaussian in a fixed finite dimension, with entrywise standard deviation proportional to n_2^{-1}. In a fixed dimension, the Euclidean or Frobenius norm of such a Gaussian vector/matrix is of the same order as its entrywise standard deviation. □

For the proofs below, distinguish the clipped empirical means in Algorithm 1 from the raw empirical functions on $\mathcal{D}_2$. For $\boldsymbol{\eta} \in K$, define

$$\boldsymbol{S}^\circ_n(\boldsymbol{\eta}) = \frac{1}{n_2}\sum_{i\in\mathcal{D}_2} \boldsymbol{\psi}(W_i,\boldsymbol{\eta}), \quad \mathbf{H}^\circ_n(\boldsymbol{\eta}) = \frac{1}{n_2}\sum_{i\in\mathcal{D}_2} \mathbf{H}(W_i,\boldsymbol{\eta}), \quad \mathbf{M}^\circ_n(\boldsymbol{\eta}) = \frac{1}{n_2}\sum_{i\in\mathcal{D}_2} \mathbf{M}(W_i,\boldsymbol{\eta}).$$

Lemma 3. *For any $\boldsymbol{\eta}, \boldsymbol{\eta}' \in K$,*

$$\boldsymbol{S}^\circ_n(\boldsymbol{\eta}) - \boldsymbol{S}^\circ_n(\boldsymbol{\eta}') = \left\{\int_0^1 \mathbf{H}^\circ_n\big(\boldsymbol{\eta}' + u(\boldsymbol{\eta}-\boldsymbol{\eta}')\big)\, du\right\}(\boldsymbol{\eta}-\boldsymbol{\eta}').$$

In particular, if $\boldsymbol{\eta}' = \boldsymbol{\eta}_0$ and Assumption 3 holds, then

$$\boldsymbol{S}^\circ_n(\boldsymbol{\eta}) - \boldsymbol{S}^\circ_n(\boldsymbol{\eta}_0) = \mathbf{H}^\circ_n(\boldsymbol{\eta})(\boldsymbol{\eta}-\boldsymbol{\eta}_0) + \boldsymbol{R}_n(\boldsymbol{\eta}), \qquad \|\boldsymbol{R}_n(\boldsymbol{\eta})\|_2 \le L_H \|\boldsymbol{\eta}-\boldsymbol{\eta}_0\|_2^2 .$$

Proof. For each fixed observation W_i, the map $\boldsymbol{\eta} \mapsto \boldsymbol{\psi}(W_i,\boldsymbol{\eta})$ is continuously differentiable on K with derivative $\mathbf{H}(W_i,\boldsymbol{\eta})$. Applying the mean-value formula to each observation and averaging gives the first claim. For the second claim, write

$$\boldsymbol{S}^\circ_n(\boldsymbol{\eta}) - \boldsymbol{S}^\circ_n(\boldsymbol{\eta}_0) = \mathbf{H}^\circ_n(\boldsymbol{\eta})(\boldsymbol{\eta}-\boldsymbol{\eta}_0) + \left[\int_0^1 \mathbf{H}^\circ_n\big(\boldsymbol{\eta}_0 + u(\boldsymbol{\eta}-\boldsymbol{\eta}_0)\big)\, du - \mathbf{H}^\circ_n(\boldsymbol{\eta})\right](\boldsymbol{\eta}-\boldsymbol{\eta}_0).$$

The bracketed term has norm at most

$$\int_0^1 \big\|\mathbf{H}^\circ_n\big(\boldsymbol{\eta}_0 + u(\boldsymbol{\eta}-\boldsymbol{\eta}_0)\big) - \mathbf{H}^\circ_n(\boldsymbol{\eta})\big\|_F \, du \le \int_0^1 L_H(1-u)\,\|\boldsymbol{\eta}-\boldsymbol{\eta}_0\|_2 \, du \le L_H \|\boldsymbol{\eta}-\boldsymbol{\eta}_0\|_2 ,$$

which proves the bound. □

Lemma 4. *Under Assumptions 3 and 6,*

$$\mathbf{H}^\circ_n(\widetilde{\boldsymbol{\eta}}_n) \xrightarrow{\mathbb{P}} \mathbf{A}_0, \qquad \mathbf{M}^\circ_n(\widetilde{\boldsymbol{\eta}}_n) \xrightarrow{\mathbb{P}} \mathbf{J}_0.$$

Proof. We start with the Hessian. By the triangle inequality,

$$\|\mathbf{H}_n^\circ(\widetilde{\boldsymbol{\eta}}_n) - \mathbf{A}_0\|_F \le \|\mathbf{H}_n^\circ(\widetilde{\boldsymbol{\eta}}_n) - \mathbf{H}_n^\circ(\boldsymbol{\eta}_0)\|_F + \|\mathbf{H}_n^\circ(\boldsymbol{\eta}_0) - \mathbf{A}_0\|_F .$$

The second term converges to zero in probability by the ordinary law of large numbers, because $\mathbf{H}(W, \boldsymbol{\eta}_0)$ is bounded by Assumption 3. For the first term,

$$\|\mathbf{H}_n^\circ(\widetilde{\boldsymbol{\eta}}_n) - \mathbf{H}_n^\circ(\boldsymbol{\eta}_0)\|_F \le \frac{1}{n_2} \sum_{i \in \mathcal{D}_2} \|\mathbf{H}(W_i, \widetilde{\boldsymbol{\eta}}_n) - \mathbf{H}(W_i, \boldsymbol{\eta}_0)\|_F \le L_H \|\widetilde{\boldsymbol{\eta}}_n - \boldsymbol{\eta}_0\|_2 ,$$

which is $o_p(1)$ by Assumption 6. Therefore $\mathbf{H}_n^\circ(\widetilde{\boldsymbol{\eta}}_n) \xrightarrow{\mathbb{P}} \mathbf{A}_0$.

For the second-moment matrix,

$$\|\mathbf{M}_n^\circ(\widetilde{\boldsymbol{\eta}}_n) - \mathbf{M}_n^\circ(\boldsymbol{\eta}_0)\|_F \le \frac{1}{n_2} \sum_{i \in \mathcal{D}_2} \big\|\boldsymbol{\psi}(W_i, \widetilde{\boldsymbol{\eta}}_n)\boldsymbol{\psi}(W_i, \widetilde{\boldsymbol{\eta}}_n)^\top - \boldsymbol{\psi}(W_i, \boldsymbol{\eta}_0)\boldsymbol{\psi}(W_i, \boldsymbol{\eta}_0)^\top\big\|_F .$$

Using the identity $ab^\top - cd^\top = (a-c)b^\top + c(b-d)^\top$ with $a = b = \boldsymbol{\psi}(W_i, \widetilde{\boldsymbol{\eta}}_n)$ and $c = d = \boldsymbol{\psi}(W_i, \boldsymbol{\eta}_0)$, together with Assumption 3, gives

$$\|\mathbf{M}_n^\circ(\widetilde{\boldsymbol{\eta}}_n) - \mathbf{M}_n^\circ(\boldsymbol{\eta}_0)\|_F \le 2G_\psi G_H \|\widetilde{\boldsymbol{\eta}}_n - \boldsymbol{\eta}_0\|_2 = o_p(1).$$

Since $\mathbf{M}(W, \boldsymbol{\eta}_0)$ is bounded and $\mathbb{E}\{\boldsymbol{\psi}(W, \boldsymbol{\eta}_0)\} = 0$, the law of large numbers implies

$$\mathbf{M}_n^\circ(\boldsymbol{\eta}_0) \xrightarrow{\mathbb{P}} \mathbb{E}\{\mathbf{M}(W, \boldsymbol{\eta}_0)\} = \mathbf{J}_0.$$

Combining the last two displays proves the lemma. □

E.2 Proof of Theorem 1

Proof. We first prove the sensitivity bounds on $\mathcal{D}_2$. Consider the score release. After clipping, each clipped score contribution $\widetilde{\boldsymbol{\psi}}_i$ has Euclidean norm at most c_ψ. If two datasets in $\mathcal{D}_2$ are adjacent, then their score averages differ in exactly one summand. Hence

$$\Delta_{\psi,n} = \sup_{\mathcal{D}_2 \simeq \mathcal{D}_2'} \left\| \frac{1}{n_2} \sum_{i \in \mathcal{D}_2} \widetilde{\boldsymbol{\psi}}_i - \frac{1}{n_2} \sum_{i \in \mathcal{D}_2'} \widetilde{\boldsymbol{\psi}}_i' \right\|_2 \le \frac{1}{n_2} \sup_{\|u\|_2, \|u'\|_2 \le c_\psi} \|u - u'\|_2 \le \frac{2c_\psi}{n_2}.$$

The same argument, with the Frobenius norm in place of the Euclidean norm, yields

$$\Delta_{H,n} \le \frac{2c_H}{n_2}, \qquad \Delta_{M,n} \le \frac{2c_M}{n_2}.$$

Now calibrate the Gaussian mechanisms to these sensitivities. By the definition of the Gaussian mechanism under GDP (Dong et al., 2022), the releases

$$\boldsymbol{\psi}_n^{dp} = \boldsymbol{\psi}_n + \boldsymbol{Z}_{\psi,n}, \qquad \mathbf{H}_n^{dp} = \mathbf{H}_n + \mathbf{Z}_{H,n}, \qquad \mathbf{M}_n^{dp} = \mathbf{M}_n + \mathbf{Z}_{M,n},$$

with standard deviations

$$\sigma_{\psi,n} = 2c_\psi/(\mu_\psi n_2), \qquad \sigma_{H,n} = 2c_H/(\mu_H n_2), \qquad \sigma_{M,n} = 2c_M/(\mu_M n_2),$$

are respectively μ_ψ-, μ_H-, and μ_M-GDP. Because they are released from the same subset $\mathcal{D}_2$, GDP composition gives

$$\mu_{\mathcal{D}_2} = \sqrt{\mu_\psi^2 + \mu_H^2 + \mu_M^2}.$$

Finally, if the changed record belongs to $\mathcal{D}_2$, then the pilot stage has the same distribution under the two adjacent full datasets. Conditional on any pilot value, the inference-stage release is $\mu_{\mathcal{D}_2}$-GDP, and averaging over the common pilot randomness preserves the same GDP guarantee. If the changed record belongs to $\mathcal{D}_1$, then the final output is a randomized post-processing of the μ_0-GDP pilot output, using the fixed $\mathcal{D}_2$ data and fresh Gaussian noises as auxiliary randomness, which is a μ_0-GDP case. Therefore, the total privacy budget is

$$\mu_{\mathrm{tot}} = \max\Big\{\mu_0, \sqrt{\mu_\psi^2 + \mu_H^2 + \mu_M^2}\Big\}.$$

All remaining operations in Algorithm 1, including projection of the pilot and final coordinate update onto K, symmetrization, projection to the positive-semidefinite cone, matrix inversion after adding ridge, the map ϕ^{-1}, and evaluation of the fitted geodesic, are deterministic post-processing and do not affect privacy. This proves the theorem. □

E.3 Proof of Theorem 2

Proof. Write

$$\widehat{\mathbf{H}}_n^{dp} = \Pi_{\mathbb{S}_+^{d_\theta}}(\mathrm{sym}(\mathbf{H}_n^{dp})) + \lambda_n \mathbf{I}_{d_\theta}.$$

We first show that $\widehat{\mathbf{H}}_n^{dp} \xrightarrow{\mathbb{P}} \mathbf{A}_0$ and that its inverse is stochastically bounded. By Lemma 1,

$$\begin{aligned}\left\|\Pi_{\mathbb{S}_+^{d_\theta}}(\mathrm{sym}(\mathbf{H}_n^{dp})) - \mathbf{A}_0\right\|_F &= \left\|\Pi_{\mathbb{S}_+^{d_\theta}}(\mathrm{sym}(\mathbf{H}_n^{dp})) - \Pi_{\mathbb{S}_+^{d_\theta}}(\mathbf{A}_0)\right\|_F \\ &\le \left\|\mathrm{sym}(\mathbf{H}_n^{dp}) - \mathbf{A}_0\right\|_F \le \left\|\mathbf{H}_n^{dp} - \mathbf{A}_0\right\|_F.\end{aligned}$$

Because clipping is inactive by Assumption 5, we have

$$\left\|\widehat{\mathbf{H}}_n^{dp} - \mathbf{A}_0\right\|_F \le \|\mathbf{Z}_{H,n}\|_F + \|\mathbf{H}_n^\circ(\widetilde{\boldsymbol{\eta}}_n) - \mathbf{A}_0\|_F + \lambda_n\sqrt{d_\theta}.$$

The first term is $o_p(1)$ by Lemma 2, the second term is $o_p(1)$ by Lemma 4, and the third term is $o(1)$ by Assumption 7. Therefore,

$$\widehat{\mathbf{H}}_n^{dp} \xrightarrow{\mathbb{P}} \mathbf{A}_0.$$

Since $\mathbf{A}_0$ is positive definite by Assumption 4, its smallest eigenvalue is strictly positive. Thus, standard eigenvalue perturbation implies that with probability tending to one, $\widehat{\mathbf{H}}_n^{dp}$ is invertible and

$$\left\|(\widehat{\mathbf{H}}_n^{dp})^{-1}\right\|_{\mathrm{op}} = O_p(1).$$

Next expand the empirical score at the pilot. By Lemma 3,

$$\boldsymbol{S}_n^\circ(\widetilde{\boldsymbol{\eta}}_n) = \boldsymbol{S}_n^\circ(\boldsymbol{\eta}_0) + \mathbf{A}_n(\widetilde{\boldsymbol{\eta}}_n)(\widetilde{\boldsymbol{\eta}}_n - \boldsymbol{\eta}_0),$$

where

$$\mathbf{A}_n(\widetilde{\boldsymbol{\eta}}_n) = \int_0^1 \mathbf{H}_n^\circ\big(\boldsymbol{\eta}_0 + u(\widetilde{\boldsymbol{\eta}}_n - \boldsymbol{\eta}_0)\big)\, du.$$

Using the definition of the raw one-step update,

$$\widehat{\boldsymbol{\eta}}_{n,\mathrm{raw}}^{dp} - \boldsymbol{\eta}_0 = \widetilde{\boldsymbol{\eta}}_n - \boldsymbol{\eta}_0 - (\widehat{\mathbf{H}}_n^{dp})^{-1}\Big(\boldsymbol{S}_n^\circ(\widetilde{\boldsymbol{\eta}}_n) + \boldsymbol{Z}_{\psi,n}\Big).$$

Substituting the previous display gives

$$\widehat{\boldsymbol{\eta}}_{n,\mathrm{raw}}^{dp} - \boldsymbol{\eta}_0 = \widetilde{\boldsymbol{\eta}}_n - \boldsymbol{\eta}_0 - (\widehat{\mathbf{H}}_n^{dp})^{-1}\Big(\boldsymbol{S}_n^\circ(\boldsymbol{\eta}_0) + \mathbf{A}_n(\widetilde{\boldsymbol{\eta}}_n)(\widetilde{\boldsymbol{\eta}}_n - \boldsymbol{\eta}_0) + \boldsymbol{Z}_{\psi,n}\Big).$$

Add and subtract $(\widehat{\mathbf{H}}_n^{dp})^{-1}\widehat{\mathbf{H}}_n^{dp}(\widetilde{\boldsymbol{\eta}}_n - \boldsymbol{\eta}_0)$:

$$\widehat{\boldsymbol{\eta}}_{n,\mathrm{raw}}^{dp} - \boldsymbol{\eta}_0 = -(\widehat{\mathbf{H}}_n^{dp})^{-1}\Big(\boldsymbol{S}_n^\circ(\boldsymbol{\eta}_0) + \boldsymbol{Z}_{\psi,n}\Big) + (\widehat{\mathbf{H}}_n^{dp})^{-1}\Big(\widehat{\mathbf{H}}_n^{dp} - \mathbf{A}_n(\widetilde{\boldsymbol{\eta}}_n)\Big)(\widetilde{\boldsymbol{\eta}}_n - \boldsymbol{\eta}_0).$$

We now show that both terms are $o_p(1)$. The first term is $o_p(1)$ because $\left\|(\widehat{\mathbf{H}}_n^{dp})^{-1}\right\|_{\mathrm{op}} = O_p(1)$, because $\boldsymbol{S}_n^\circ(\boldsymbol{\eta}_0) \xrightarrow{\mathbb{P}} 0$ by the law of large numbers and $\mathbb{E}\{\boldsymbol{\psi}(W, \boldsymbol{\eta}_0)\} = 0$, and because $\boldsymbol{Z}_{\psi,n} = o_p(1)$ by Lemma 2.

For the second term, note that

$$\widehat{\mathbf{H}}_n^{dp} - \mathbf{A}_n(\widetilde{\boldsymbol{\eta}}_n) = \Big(\widehat{\mathbf{H}}_n^{dp} - \mathbf{H}_n^\circ(\widetilde{\boldsymbol{\eta}}_n)\Big) + \Big(\mathbf{H}_n^\circ(\widetilde{\boldsymbol{\eta}}_n) - \mathbf{A}_n(\widetilde{\boldsymbol{\eta}}_n)\Big).$$

The first bracket is $o_p(1)$ by the argument at the start of the proof. For the second bracket, Assumption 3 and the definition of $\mathbf{A}_n(\widetilde{\boldsymbol{\eta}}_n)$ imply

$$\|\mathbf{H}_n^\circ(\widetilde{\boldsymbol{\eta}}_n) - \mathbf{A}_n(\widetilde{\boldsymbol{\eta}}_n)\|_F \le \int_0^1 \|\mathbf{H}_n^\circ(\widetilde{\boldsymbol{\eta}}_n) - \mathbf{H}_n^\circ(\boldsymbol{\eta}_0 + u(\widetilde{\boldsymbol{\eta}}_n - \boldsymbol{\eta}_0))\|_F\, du \le L_H \|\widetilde{\boldsymbol{\eta}}_n - \boldsymbol{\eta}_0\|_2.$$

Therefore

$$\left\|(\widehat{\mathbf{H}}_n^{dp})^{-1}\big(\widehat{\mathbf{H}}_n^{dp} - \mathbf{A}_n(\widetilde{\boldsymbol{\eta}}_n)\big)(\widetilde{\boldsymbol{\eta}}_n - \boldsymbol{\eta}_0)\right\|_2 \le O_p(1)\cdot o_p(1)\cdot o_p(1) = o_p(1).$$

Hence $\widehat{\boldsymbol{\eta}}_{n,\mathrm{raw}}^{dp} - \boldsymbol{\eta}_0 = o_p(1)$. Nonexpansiveness of Π_K and $\Pi_K(\boldsymbol{\eta}_0) = \boldsymbol{\eta}_0$ imply $\|\widehat{\boldsymbol{\eta}}_n^{dp} - \boldsymbol{\eta}_0\|_2 \le \|\widehat{\boldsymbol{\eta}}_{n,\mathrm{raw}}^{dp} - \boldsymbol{\eta}_0\|_2 = o_p(1)$. Moreover, because $\boldsymbol{\eta}_0 \in \mathrm{int}(K)$, the final projection is inactive with probability tending to one. Since ϕ^{-1} is continuous on K, the continuous mapping theorem implies

$$\widehat{\boldsymbol{\theta}}_n^{dp} = \phi^{-1}(\widehat{\boldsymbol{\eta}}_n^{dp}) \xrightarrow{\mathbb{P}} \phi^{-1}(\boldsymbol{\eta}_0) = \boldsymbol{\theta}_0.$$

□

E.4 Proof of Theorem 3

Proof. We start from the raw-update decomposition obtained in the proof of Theorem 2:

$$\widehat{\boldsymbol{\eta}}^{dp}_{n,\mathrm{raw}} - \boldsymbol{\eta}_0 = -(\widehat{\mathbf{H}}^{dp}_n)^{-1}\Big(\boldsymbol{S}^{\circ}_n(\boldsymbol{\eta}_0) + \boldsymbol{Z}_{\psi,n}\Big) + (\widehat{\mathbf{H}}^{dp}_n)^{-1}\Big(\widehat{\mathbf{H}}^{dp}_n - \mathbf{A}_n(\widetilde{\boldsymbol{\eta}}_n)\Big)(\widetilde{\boldsymbol{\eta}}_n - \boldsymbol{\eta}_0). \quad (5)$$

We show that the second term is $o_p(n_2^{-1/2})$ and that $(\widehat{\mathbf{H}}^{dp}_n)^{-1}$ may be replaced by $\mathbf{A}_0^{-1}$ up to the same order.

Step 1: the remainder term is $o_p(n_2^{-1/2})$. First,

$\|\mathbf{H}^{\circ}_n(\widetilde{\boldsymbol{\eta}}_n) - \mathbf{A}_0\|_F \le \|\mathbf{H}^{\circ}_n(\boldsymbol{\eta}_0) - \mathbf{A}_0\|_F + L_H\|\widetilde{\boldsymbol{\eta}}_n - \boldsymbol{\eta}_0\|_2 = O_p(n_2^{-1/2}) + o_p(n_2^{-1/4}) = o_p(n_2^{-1/4})$. Since $\mathbf{A}_0 \succ 0$, $\Pi_{\mathbb{S}^{d_\theta}_+}(\mathbf{A}_0) = \mathbf{A}_0$. By nonexpansiveness of symmetrization followed by PSD projection,

$$\begin{aligned}
\|\widehat{\mathbf{H}}^{dp}_n - \mathbf{A}_0\|_F &\le \Big\|\Pi_{\mathbb{S}^{d_\theta}_+}\{\mathrm{sym}(\mathbf{H}^{dp}_n)\} - \mathbf{A}_0\Big\|_F + \lambda_n\sqrt{d_\theta} \\
&\le \|\mathbf{H}^{dp}_n - \mathbf{A}_0\|_F + \lambda_n\sqrt{d_\theta} \\
&\le \|\mathbf{H}^{\circ}_n(\widetilde{\eta}_n) - \mathbf{A}_0\|_F + \|\mathbf{Z}_{H,n}\|_F + \lambda_n\sqrt{d_\theta} \\
&= o_p(n_2^{-1/4}),
\end{aligned}$$

where we use Lemma 2 and $\sqrt{n_2}\lambda_n \to 0$. Next,

$$\begin{aligned}
\|\mathbf{A}_n(\widetilde{\eta}_n) - \mathbf{A}_0\|_F &\le \int_0^1 \|\mathbf{H}^{\circ}_n\{\boldsymbol{\eta}_0 + u(\widetilde{\boldsymbol{\eta}}_n - \boldsymbol{\eta}_0)\} - \mathbf{H}^{\circ}_n(\boldsymbol{\eta}_0)\|_F\, du + \|\mathbf{H}^{\circ}_n(\boldsymbol{\eta}_0) - \mathbf{A}_0\|_F \\
&\le L_H\|\widetilde{\boldsymbol{\eta}}_n - \boldsymbol{\eta}_0\|_2 + O_p(n_2^{-1/2}) = o_p(n_2^{-1/4}).
\end{aligned}$$

Therefore,

$$\|\widehat{\mathbf{H}}^{dp}_n - \mathbf{A}_n(\widetilde{\boldsymbol{\eta}}_n)\|_F = o_p(n_2^{-1/4}).$$

Multiplying by $\|\widetilde{\boldsymbol{\eta}}_n - \boldsymbol{\eta}_0\|_2 = o_p(n_2^{-1/4})$ yields

$$\Big\|\big(\widehat{\mathbf{H}}^{dp}_n - \mathbf{A}_n(\widetilde{\boldsymbol{\eta}}_n)\big)(\widetilde{\boldsymbol{\eta}}_n - \boldsymbol{\eta}_0)\Big\|_2 = o_p(n_2^{-1/2}).$$

Since $\Big\|(\widehat{\mathbf{H}}^{dp}_n)^{-1}\Big\|_{\mathrm{op}} = O_p(1)$ from the proof of Theorem 2, the second term in (5) is indeed $o_p(n_2^{-1/2})$.

Step 2: replace $(\widehat{\mathbf{H}}^{dp}_n)^{-1}$ by $\mathbf{A}_0^{-1}$. From the proof of Theorem 2,

$$\Big\|\widehat{\mathbf{H}}^{dp}_n - \mathbf{A}_0\Big\|_F = o_p(1).$$

Matrix inversion is continuous at positive-definite matrices, so

$$(\widehat{\mathbf{H}}^{dp}_n)^{-1} \xrightarrow{\mathbb{P}} \mathbf{A}_0^{-1}.$$

Moreover,

$$\|\boldsymbol{S}^{\circ}_n(\boldsymbol{\eta}_0)\|_2 = O_p(n_2^{-1/2})$$

by the multivariate central limit theorem applied to the bounded mean-zero vectors $\boldsymbol{\psi}(W_i, \boldsymbol{\eta}_0)$, and

$$\|\boldsymbol{Z}_{\psi,n}\|_2 = O_p(n_2^{-1})$$

by Lemma 2. Therefore

$$\Big((\widehat{\mathbf{H}}^{dp}_n)^{-1} - \mathbf{A}_0^{-1}\Big)\Big(\boldsymbol{S}^{\circ}_n(\boldsymbol{\eta}_0) + \boldsymbol{Z}_{\psi,n}\Big) = o_p(1)\cdot O_p(n_2^{-1/2}) = o_p(n_2^{-1/2}).$$

Combining Step 1 with Step 2 in (5) yields the raw-update representation

$$\widehat{\boldsymbol{\eta}}^{dp}_{n,\mathrm{raw}} - \boldsymbol{\eta}_0 = -\mathbf{A}_0^{-1}\Big(\boldsymbol{S}^{\circ}_n(\boldsymbol{\eta}_0) + \boldsymbol{Z}_{\psi,n}\Big) + \boldsymbol{r}_n, \qquad \|\boldsymbol{r}_n\|_2 = o_p(n_2^{-1/2}).$$

The consistency proof shows that $\widehat{\boldsymbol{\eta}}^{dp}_n = \widehat{\boldsymbol{\eta}}^{dp}_{n,\mathrm{raw}}$ with probability tending to one. Thus the projected update has the same asymptotic linear representation, proving (2). The symbol $\boldsymbol{r}_n$ below continues to denote the raw-update remainder; Step 4 separately controls the projected remainder in conditional L^2.

Step 3: central limit theorem. Apply the corresponding representation for the projected update and multiply by $\sqrt{n_2}$:

$$\sqrt{n_2}(\widehat{\boldsymbol{\eta}}_n^{dp} - \boldsymbol{\eta}_0) = -\mathbf{A}_0^{-1}\left\{\frac{1}{\sqrt{n_2}}\sum_{i\in\mathcal{D}_2}\boldsymbol{\psi}(W_i,\boldsymbol{\eta}_0)\right\} - \mathbf{A}_0^{-1}\sqrt{n_2}\boldsymbol{Z}_{\psi,n} + o_p(1).$$

The first term converges in distribution to

$$N(0,\ \mathbf{A}_0^{-1}\mathbf{J}_0\mathbf{A}_0^{-1})$$

by the multivariate central limit theorem. The second term converges to zero in probability because $\sqrt{n_2}\boldsymbol{Z}_{\psi,n} = o_p(1)$ by Lemma 2. Slutsky's theorem therefore proves the stated CLT.

Step 4: first-order covariance decomposition. Throughout this step, we condition on $\mathcal{F}_1$, so that the realized pilot is fixed. From the raw-update linear decomposition,

$$\boldsymbol{r}_n = -\Big\{(\widehat{\mathbf{H}}_n^{dp})^{-1} - \mathbf{A}_0^{-1}\Big\}\Big(\boldsymbol{S}_n^{\circ}(\boldsymbol{\eta}_0) + \boldsymbol{Z}_{\psi,n}\Big) + (\widehat{\mathbf{H}}_n^{dp})^{-1}\Big(\widehat{\mathbf{H}}_n^{dp} - \mathbf{A}_n(\widetilde{\boldsymbol{\eta}}_n)\Big)(\widetilde{\boldsymbol{\eta}}_n - \boldsymbol{\eta}_0).$$

Let

$$\mathcal{E}_n = \Big\{\|\widehat{\mathbf{H}}_n^{dp} - \mathbf{A}_0\|_F \le \lambda_{\min}(\mathbf{A}_0)/2\Big\}.$$

By Assumptions 3–7, Lemma 2, and standard bounded-summand concentration for the fixed-dimensional empirical Hessian, $\mathbb{P}(\mathcal{E}_n^c \mid \mathcal{F}_1) \le C\exp(-cn_2)$ with probability tending to one. On $\mathcal{E}_n$, $\|(\widehat{\mathbf{H}}_n^{dp})^{-1}\|_{\mathrm{op}} \le C$ and $\|(\widehat{\mathbf{H}}_n^{dp})^{-1} - \mathbf{A}_0^{-1}\|_{\mathrm{op}} \le C\|\widehat{\mathbf{H}}_n^{dp} - \mathbf{A}_0\|_F$ for some constant $C > 0$. Moreover, conditional on $\mathcal{F}_1$,

$$\mathbb{E}\big[\|\boldsymbol{S}_n^{\circ}(\boldsymbol{\eta}_0) + \boldsymbol{Z}_{\psi,n}\|_2^4 \mid \mathcal{F}_1\big] = O(n_2^{-2}),$$

and

$$\mathbb{E}\Big[\|\widehat{\mathbf{H}}_n^{dp} - \mathbf{A}_0\|_F^4 \mid \mathcal{F}_1\Big] = O_p\big(n_2^{-2} + \|\widetilde{\boldsymbol{\eta}}_n - \boldsymbol{\eta}_0\|_2^4 + \lambda_n^4\big),$$
$$\mathbb{E}\Big[\|\widehat{\mathbf{H}}_n^{dp} - \mathbf{A}_n(\widetilde{\boldsymbol{\eta}}_n)\|_F^2 \mid \mathcal{F}_1\Big] = O_p\big(n_2^{-1} + \|\widetilde{\boldsymbol{\eta}}_n - \boldsymbol{\eta}_0\|_2^2 + \lambda_n^2\big).$$

Therefore, by Cauchy's inequality,

$$\begin{aligned}\mathbb{E}\big[\|\boldsymbol{r}_n\|_2^2\mathbf{1}_{\mathcal{E}_n} \mid \mathcal{F}_1\big] &\le C\left\{\mathbb{E}\Big[\|\widehat{\mathbf{H}}_n^{dp} - \mathbf{A}_0\|_F^4 \mid \mathcal{F}_1\Big]\right\}^{1/2}\left\{\mathbb{E}\big[\|\boldsymbol{S}_n^{\circ}(\boldsymbol{\eta}_0) + \boldsymbol{Z}_{\psi,n}\|_2^4 \mid \mathcal{F}_1\big]\right\}^{1/2}\\ &\quad + C\|\widetilde{\boldsymbol{\eta}}_n - \boldsymbol{\eta}_0\|_2^2\mathbb{E}\Big[\|\widehat{\mathbf{H}}_n^{dp} - \mathbf{A}_n(\widetilde{\boldsymbol{\eta}}_n)\|_F^2 \mid \mathcal{F}_1\Big]\\ &= o_p(n_2^{-1}).\end{aligned}$$

On $\mathcal{E}_n^c$, the ridge gives $\|(\widehat{\mathbf{H}}_n^{dp})^{-1}\|_{\mathrm{op}} \le \lambda_n^{-1}$. Hence, by Cauchy's inequality and the exponential bound for $\mathcal{E}_n^c$,

$$\mathbb{E}\big[\|\boldsymbol{r}_n\|_2^2\mathbf{1}_{\mathcal{E}_n^c} \mid \mathcal{F}_1\big] \le C\lambda_n^{-2}\left(n_2^{-1} + \|\widetilde{\boldsymbol{\eta}}_n - \boldsymbol{\eta}_0\|_2^2\right)\exp(-cn_2/2) = o_p(n_2^{-1}),$$

where the last equality uses $\log(1/\lambda_n) = o(n_2)$ and $\|\widetilde{\boldsymbol{\eta}}_n - \boldsymbol{\eta}_0\|_2 = o_p(n_2^{-1/4})$. Thus

$$\mathbb{E}\big[\|\boldsymbol{r}_n\|_2^2 \mid \mathcal{F}_1\big] = o_p(n_2^{-1}).$$

It follows that

$$\|\mathrm{Cov}(\boldsymbol{r}_n \mid \mathcal{F}_1)\|_F \le \mathbb{E}\big[\|\boldsymbol{r}_n\|_2^2 \mid \mathcal{F}_1\big] = o_p(n_2^{-1}),$$

and

$$\big\|\mathrm{Cov}\big(-\mathbf{A}_0^{-1}\{\boldsymbol{S}_n^{\circ}(\boldsymbol{\eta}_0) + \boldsymbol{Z}_{\psi,n}\}, \boldsymbol{r}_n \mid \mathcal{F}_1\big)\big\|_F = o_p(n_2^{-1}).$$

Finally, conditional on $\mathcal{F}_1$, $\boldsymbol{S}_n^{\circ}(\boldsymbol{\eta}_0)$ depends only on $\mathcal{D}_2$ and is independent of $\boldsymbol{Z}_{\psi,n}$. Therefore,

$$\mathrm{Cov}\{\boldsymbol{S}_n^{\circ}(\boldsymbol{\eta}_0) + \boldsymbol{Z}_{\psi,n} \mid \mathcal{F}_1\} = \frac{\mathbf{J}_0}{n_2} + \sigma_{\psi,n}^2\mathbf{I}_{d_\theta}.$$

Combining the last displays gives

$$\mathrm{Cov}(\widehat{\boldsymbol{\eta}}_{n,\mathrm{raw}}^{dp} \mid \mathcal{F}_1) = \mathbf{A}_0^{-1}\left(\frac{\mathbf{J}_0}{n_2} + \sigma_{\psi,n}^2\mathbf{I}_{d_\theta}\right)\mathbf{A}_0^{-1} + o_p(n_2^{-1}),$$

For the raw update, write $\widehat{\boldsymbol{\eta}}^{dp}_{n,\mathrm{raw}} - \boldsymbol{\eta}_0 = \boldsymbol{L}_n + \boldsymbol{r}_n$, where $\boldsymbol{L}_n = -\mathbf{A}_0^{-1}\{\boldsymbol{S}^\circ_n(\boldsymbol{\eta}_0) + \boldsymbol{Z}_{\psi,n}\}$. Choose $a > 0$ such that the closed ball centered at $\boldsymbol{\eta}_0$ with radius a is contained in K, and define $\boldsymbol{r}_n^K = \widehat{\boldsymbol{\eta}}^{dp}_n - \boldsymbol{\eta}_0 - \boldsymbol{L}_n$. Nonexpansiveness of Π_K gives

$$\|\boldsymbol{r}_n^K\|_2 \le \|\boldsymbol{r}_n\|_2 + \|\boldsymbol{L}_n\|_2 \mathbf{1}\{\|\boldsymbol{L}_n\|_2 > a\}.$$

Consequently,

$$\mathbb{E}(\|\boldsymbol{r}_n^K\|_2^2 \mid \mathcal{F}_1) \le 2\mathbb{E}(\|\boldsymbol{r}_n\|_2^2 \mid \mathcal{F}_1) + 2a^{-2}\mathbb{E}(\|\boldsymbol{L}_n\|_2^4 \mid \mathcal{F}_1) = o_p(n_2^{-1}).$$

Thus the conditional covariance expansion also holds for the projected update.

□

E.5 Proof of Theorem 4 and Corollary 1

Proof of Theorem 4. We already know from the proof of Theorem 2 that

$$\widehat{\mathbf{H}}^{dp}_n \xrightarrow{\mathbb{P}} \mathbf{A}_0.$$

By Lemma 4,

$$\mathbf{M}^\circ_n(\widetilde{\boldsymbol{\eta}}_n) \xrightarrow{\mathbb{P}} \mathbf{J}_0.$$

Since clipping is inactive and $\mathbf{Z}_{M,n} = o_p(1)$ by Lemma 2,

$$\mathbf{M}^{dp}_n = \mathbf{M}^\circ_n(\widetilde{\boldsymbol{\eta}}_n) + \mathbf{Z}_{M,n} \xrightarrow{\mathbb{P}} \mathbf{J}_0.$$

Applying Lemma 1 to the positive-semidefinite projection then gives

$$\widehat{\mathbf{M}}^{dp}_n \xrightarrow{\mathbb{P}} \mathbf{J}_0.$$

Therefore

$$\frac{\widehat{\mathbf{M}}^{dp}_n}{n_2} + \sigma^2_{\psi,n}\mathbf{I}_{d_\theta} - \left(\frac{\mathbf{J}_0}{n_2} + \sigma^2_{\psi,n}\mathbf{I}_{d_\theta}\right) = o_p(n_2^{-1}).$$

Because matrix inversion is continuous at $\mathbf{A}_0$ and $\widehat{\mathbf{H}}^{dp}_n \xrightarrow{\mathbb{P}} \mathbf{A}_0$,

$$(\widehat{\mathbf{H}}^{dp}_n)^{-1} \xrightarrow{\mathbb{P}} \mathbf{A}_0^{-1}.$$

Combining the last two displays and using the continuous mapping theorem proves

$$\begin{aligned}\widehat{\mathbf{V}}^{dp}_n - \mathbf{A}_0^{-1}\left(\frac{\mathbf{J}_0}{n_2} + \sigma^2_{\psi,n}\mathbf{I}_{d_\theta}\right)\mathbf{A}_0^{-1} &= \{(\widehat{\mathbf{H}}^{dp}_n)^{-1} - \mathbf{A}_0^{-1}\}\left(\frac{\widehat{\mathbf{M}}^{dp}_n}{n_2} + \sigma^2_{\psi,n}\mathbf{I}_{d_\theta}\right)(\widehat{\mathbf{H}}^{dp}_n)^{-1} \\ &\quad + \mathbf{A}_0^{-1}\left(\frac{\widehat{\mathbf{M}}^{dp}_n}{n_2} - \frac{\mathbf{J}_0}{n_2}\right)(\widehat{\mathbf{H}}^{dp}_n)^{-1} \\ &\quad + \mathbf{A}_0^{-1}\left(\frac{\mathbf{J}_0}{n_2} + \sigma^2_{\psi,n}\mathbf{I}_{d_\theta}\right)\{(\widehat{\mathbf{H}}^{dp}_n)^{-1} - \mathbf{A}_0^{-1}\}.\end{aligned}$$

The three terms on the right-hand side are respectively

$$o_p(1)\,O_p(n_2^{-1})\,O_p(1), \qquad O(1)\,o_p(n_2^{-1})\,O_p(1), \qquad O(1)\,O(n_2^{-1})\,o_p(1)$$

since $\sigma^2_{\psi,n} = O(n_2^{-2})$ for fixed $\mu_\psi > 0$. Thus,

$$\widehat{\mathbf{V}}^{dp}_n - \mathbf{A}_0^{-1}\left(\frac{\mathbf{J}_0}{n_2} + \sigma^2_{\psi,n}\mathbf{I}_{d_\theta}\right)\mathbf{A}_0^{-1} = o_p(n_2^{-1}).$$

Finally, since $n_2\sigma^2_{\psi,n} = 4c_\psi^2/(\mu_\psi^2 n_2) \to 0$, multiplying by n_2 yields

$$n_2\widehat{\mathbf{V}}^{dp}_n \xrightarrow{\mathbb{P}} \mathbf{A}_0^{-1}\mathbf{J}_0\mathbf{A}_0^{-1}.$$

□

Proof of Corollary 1. Write

$$\boldsymbol{\Sigma}_n := \mathbf{A}_0^{-1}\left(\frac{\mathbf{J}_0}{n_2}\right)\mathbf{A}_0^{-1}.$$

Theorem 3 states that

$$\boldsymbol{\Sigma}_n^{-1/2}(\widehat{\boldsymbol{\eta}}_n^{dp} - \boldsymbol{\eta}_0) \Rightarrow N(0, \mathbf{I}_{d_\theta}).$$

Equivalently,

$$(\widehat{\boldsymbol{\eta}}_n^{dp} - \boldsymbol{\eta}_0)^\top \boldsymbol{\Sigma}_n^{-1}(\widehat{\boldsymbol{\eta}}_n^{dp} - \boldsymbol{\eta}_0) \Rightarrow \chi^2_{d_\theta}.$$

Theorem 4 implies $\widehat{\mathbf{V}}_n^{dp} - \boldsymbol{\Sigma}_n = o_p(n_2^{-1})$ because the extra term $\sigma^2_{\psi,n}\mathbf{I}_{d_\theta}$ is of order $o(n_2^{-1})$. Hence

$$\widehat{\mathbf{V}}_n^{dp}\boldsymbol{\Sigma}_n^{-1} \xrightarrow{\mathbb{P}} \mathbf{I}_{d_\theta} \qquad \text{and} \qquad (\widehat{\mathbf{V}}_n^{dp})^{-1} - \boldsymbol{\Sigma}_n^{-1} = o_p(n_2).$$

A standard Slutsky argument then gives

$$(\widehat{\boldsymbol{\eta}}_n^{dp} - \boldsymbol{\eta}_0)^\top (\widehat{\mathbf{V}}_n^{dp})^{-1}(\widehat{\boldsymbol{\eta}}_n^{dp} - \boldsymbol{\eta}_0) \Rightarrow \chi^2_{d_\theta}.$$

The confidence region statement is the usual inversion of the Wald test. □

E.6 Proof of Theorem 5

Proof. Fix $\|\boldsymbol{x}\|_2 \le B_X$. By Assumptions 2 and 3, the map

$$g_x(\boldsymbol{\eta}) = \zeta_x\big(\gamma_{\phi^{-1}(\boldsymbol{\eta})}(\boldsymbol{x})\big)$$

is continuously differentiable in a neighborhood of $\boldsymbol{\eta}_0$. Therefore, a first-order Taylor expansion gives

$$g_x(\widehat{\boldsymbol{\eta}}_n^{dp}) - g_x(\boldsymbol{\eta}_0) = \mathbf{G}_x(\widehat{\boldsymbol{\eta}}_n^{dp} - \boldsymbol{\eta}_0) + \boldsymbol{r}_{x,n},$$

where

$$\frac{\|\boldsymbol{r}_{x,n}\|_2}{\left\|\widehat{\boldsymbol{\eta}}_n^{dp} - \boldsymbol{\eta}_0\right\|_2} \xrightarrow{\mathbb{P}} 0.$$

Since $\widehat{\boldsymbol{\eta}}_n^{dp} - \boldsymbol{\eta}_0 = O_p(n_2^{-1/2})$ by Theorem 3, it follows that $\boldsymbol{r}_{x,n} = o_p(n_2^{-1/2})$.

Multiply by $\sqrt{n_2}$:

$$\sqrt{n_2}\big(g_x(\widehat{\boldsymbol{\eta}}_n^{dp}) - g_x(\boldsymbol{\eta}_0)\big) = \mathbf{G}_x\sqrt{n_2}(\widehat{\boldsymbol{\eta}}_n^{dp} - \boldsymbol{\eta}_0) + o_p(1).$$

Theorem 3 implies

$$\sqrt{n_2}(\widehat{\boldsymbol{\eta}}_n^{dp} - \boldsymbol{\eta}_0) \Rightarrow N(0, \mathbf{A}_0^{-1}\mathbf{J}_0\mathbf{A}_0^{-1}),$$

so the multivariate delta method yields

$$\sqrt{n_2}\big(g_x(\widehat{\boldsymbol{\eta}}_n^{dp}) - g_x(\boldsymbol{\eta}_0)\big) \Rightarrow N\big(0,\ \mathbf{G}_x\mathbf{A}_0^{-1}\mathbf{J}_0\mathbf{A}_0^{-1}\mathbf{G}_x^\top\big).$$

For the covariance estimator, define

$$\widehat{\mathbf{G}}_x = Dg_x(\widehat{\boldsymbol{\eta}}_n^{dp}).$$

Because g_x is continuously differentiable and $\widehat{\boldsymbol{\eta}}_n^{dp} \xrightarrow{\mathbb{P}} \boldsymbol{\eta}_0$ by Theorem 2, $\widehat{\mathbf{G}}_x \xrightarrow{\mathbb{P}} \mathbf{G}_x$. Theorem 4 gives

$$\widehat{\mathbf{V}}_n^{dp} - \mathbf{A}_0^{-1}\left(\frac{\mathbf{J}_0}{n_2} + \sigma^2_{\psi,n}\mathbf{I}_{d_\theta}\right)\mathbf{A}_0^{-1} = o_p(n_2^{-1}).$$

Hence,

$$\begin{aligned}
&\widehat{\boldsymbol{\Sigma}}_x^{dp} - \mathbf{G}_x\mathbf{A}_0^{-1}\left(\frac{\mathbf{J}_0}{n_2} + \sigma^2_{\psi,n}\mathbf{I}_{d_\theta}\right)\mathbf{A}_0^{-1}\mathbf{G}_x^\top \\
=&\widehat{\mathbf{G}}_x\widehat{\mathbf{V}}_n^{dp}\widehat{\mathbf{G}}_x^\top - \mathbf{G}_x\mathbf{A}_0^{-1}\left(\frac{\mathbf{J}_0}{n_2} + \sigma^2_{\psi,n}\mathbf{I}_{d_\theta}\right)\mathbf{A}_0^{-1}\mathbf{G}_x^\top \\
=&\widehat{\mathbf{G}}_x\left(\widehat{\mathbf{V}}_n^{dp} - \mathbf{A}_0^{-1}\left(\frac{\mathbf{J}_0}{n_2} + \sigma^2_{\psi,n}\mathbf{I}_{d_\theta}\right)\mathbf{A}_0^{-1}\right)\widehat{\mathbf{G}}_x^\top \\
&+\left(\widehat{\mathbf{G}}_x - \mathbf{G}_x\right)\mathbf{A}_0^{-1}\left(\frac{\mathbf{J}_0}{n_2} + \sigma^2_{\psi,n}\mathbf{I}_{d_\theta}\right)\mathbf{A}_0^{-1}\widehat{\mathbf{G}}_x^\top \\
&+\mathbf{G}_x\mathbf{A}_0^{-1}\left(\frac{\mathbf{J}_0}{n_2} + \sigma^2_{\psi,n}\mathbf{I}_{d_\theta}\right)\mathbf{A}_0^{-1}\left(\widehat{\mathbf{G}}_x - \mathbf{G}_x\right)^\top.
\end{aligned}$$

The three terms on the right-hand side are respectively

$$O_p(1)\, o_p(n_2^{-1})\, O_p(1), \qquad o_p(1)\, O(n_2^{-1})\, O_p(1), \qquad O(1)\, O(n_2^{-1})\, o_p(1).$$

Thus,

$$\widehat{\mathbf{\Sigma}}_x^{dp} - \mathbf{G}_x \mathbf{A}_0^{-1} \left(\frac{\mathbf{J}_0}{n_2} + \sigma_{\psi,n}^2 \mathbf{I}_{d_\theta} \right) \mathbf{A}_0^{-1} \mathbf{G}_x^\top = o_p(n_2^{-1}).$$

The confidence-region statement follows by the same Wald inversion argument as in Corollary 1, now applied in the local chart of the response manifold at predictor value $\boldsymbol{x}$. □

F INFERENCE WITH ACTIVE CLIPPING

We examine the effect of active clipping on Algorithm 1. The dimension and positive privacy budgets remain fixed. The first part retains Assumptions 1–4, 6, and 7, without imposing Assumption 5. Expectations evaluated at $\widetilde{\boldsymbol{\eta}}_n$ are taken over an independent observation W, holding the pilot fixed. The pilot belongs to K for every realization because it is projected before derivative evaluation.

Fixed thresholds. Suppose the thresholds are fixed, $\widetilde{\boldsymbol{\eta}}_n \xrightarrow{\mathbb{P}} \boldsymbol{\eta}_0$, and $\mathbb{E}\big[\mathrm{clip}_{c_H}\{\mathbf{H}(W, \boldsymbol{\eta}_0)\}\big] \succ 0$. Then

$$\widehat{\boldsymbol{\eta}}_{n,\mathrm{raw}}^{dp} \xrightarrow{\mathbb{P}} \boldsymbol{\eta}_0 - \big[\mathbb{E}\{\mathrm{clip}_{c_H}(\mathbf{H}(W, \boldsymbol{\eta}_0))\}\big]^{-1} \mathbb{E}\{\mathrm{clip}_{c_\psi}(\boldsymbol{\psi}(W, \boldsymbol{\eta}_0))\}, \tag{6}$$

and

$$n_2 \widehat{\mathbf{V}}_n^{dp} \xrightarrow{\mathbb{P}} \big[\mathbb{E}\{\mathrm{clip}_{c_H}(\mathbf{H}(W, \boldsymbol{\eta}_0))\}\big]^{-1} \mathbb{E}\{\mathrm{clip}_{c_M}(\mathbf{M}(W, \boldsymbol{\eta}_0))\} \times \big[\mathbb{E}\{\mathrm{clip}_{c_H}(\mathbf{H}(W, \boldsymbol{\eta}_0))\}\big]^{-1}. \tag{7}$$

Let $\boldsymbol{\eta}_\infty$ denote the right-hand side of (6). This limit concerns the raw coordinate update. Continuity of projection gives $\widehat{\boldsymbol{\eta}}_n^{dp} \xrightarrow{\mathbb{P}} \Pi_K(\boldsymbol{\eta}_\infty)$. To obtain the two displayed limits, pilot consistency and the law of large numbers imply that each clipped average converges to its expectation at $\boldsymbol{\eta}_0$. The Gaussian release noise is $O_p(n_2^{-1})$. The Hessian limit is positive definite, so the projection leaves $\mathrm{sym}(\mathbf{H}_n^{dp})$ unchanged with probability tending to one. The second-moment limit is positive semidefinite, and its projection has the same limit by Lemma 1. Substituting these limits into Algorithm 1, with $\lambda_n \to 0$ and $n_2 \sigma_{\psi,n}^2 \to 0$, gives (6)–(7).

For inference on $\boldsymbol{\eta}_0$, we add back the expected score and Hessian contributions removed by clipping. The score contribution has the explicit form

$$\mathbb{E}\big[\boldsymbol{\psi}(W, \boldsymbol{\eta}) - \mathrm{clip}_{c_\psi}\{\boldsymbol{\psi}(W, \boldsymbol{\eta})\}\big] = \mathbb{E}\left[\left(1 - \frac{c_\psi}{\|\boldsymbol{\psi}(W, \boldsymbol{\eta})\|_2} \right)_+ \boldsymbol{\psi}(W, \boldsymbol{\eta}) \right], \tag{8}$$

where the integrand is zero when the score is zero. The same identity holds for the Hessian, with threshold c_H and the Frobenius norm. Define

$$\widehat{\mathbf{H}}_n^{bc} = \Pi_{\mathbb{S}_+^{d_\theta}} \Big(\mathrm{sym}\Big[\mathbf{H}_n^{dp} + \mathbb{E}\big\{ \mathbf{H}(W, \widetilde{\boldsymbol{\eta}}_n) - \mathrm{clip}_{c_H}(\mathbf{H}(W, \widetilde{\boldsymbol{\eta}}_n)) \big\} \Big] \Big) + \lambda_n \mathbf{I}_{d_\theta}, \tag{9}$$

and

$$\begin{aligned} \widehat{\boldsymbol{\eta}}_{n,\mathrm{raw}}^{bc} &= \widetilde{\boldsymbol{\eta}}_n - (\widehat{\mathbf{H}}_n^{bc})^{-1} \Big[\boldsymbol{\psi}_n^{dp} + \mathbb{E}\big\{ \boldsymbol{\psi}(W, \widetilde{\boldsymbol{\eta}}_n) - \mathrm{clip}_{c_\psi}(\boldsymbol{\psi}(W, \widetilde{\boldsymbol{\eta}}_n)) \big\} \Big], \\ \widehat{\boldsymbol{\eta}}_n^{bc} &= \Pi_K(\widehat{\boldsymbol{\eta}}_{n,\mathrm{raw}}^{bc}). \end{aligned} \tag{10}$$

For this corrected procedure, suppose $\|\widetilde{\boldsymbol{\eta}}_n - \boldsymbol{\eta}_0\|_2 = o_p(n_2^{-1/4})$. Then

$$\widehat{\boldsymbol{\eta}}_n^{bc} - \boldsymbol{\eta}_0 = -\mathbf{A}_0^{-1} \left[\frac{1}{n_2} \sum_{i \in \mathcal{D}_2} \Big\{ \mathrm{clip}_{c_\psi}(\boldsymbol{\psi}(W_i, \boldsymbol{\eta}_0)) - \mathbb{E}[\mathrm{clip}_{c_\psi}\{\boldsymbol{\psi}(W, \boldsymbol{\eta}_0)\}] \Big\} + \boldsymbol{Z}_{\psi,n} \right] + o_p(n_2^{-1/2}). \tag{11}$$

Consequently,

$$\sqrt{n_2}(\widehat{\boldsymbol{\eta}}_n^{bc} - \boldsymbol{\eta}_0) \Rightarrow N\Big(0, \mathbf{A}_0^{-1} \mathrm{Var}[\mathrm{clip}_{c_\psi}\{\boldsymbol{\psi}(W, \boldsymbol{\eta}_0)\}] \mathbf{A}_0^{-1}\Big). \tag{12}$$

For covariance estimation, take $c_M = c_\psi^2$. The identity $\left\|\boldsymbol{\psi}\boldsymbol{\psi}^\top\right\|_F = \|\boldsymbol{\psi}\|_2^2$ then gives $\mathrm{clip}_{c_M}(\boldsymbol{\psi}\boldsymbol{\psi}^\top) = \mathrm{clip}_{c_\psi}(\boldsymbol{\psi})\mathrm{clip}_{c_\psi}(\boldsymbol{\psi})^\top$. Thus, the covariance estimator based on the three releases is

$$\widehat{\mathbf{V}}_n^{bc} = (\widehat{\mathbf{H}}_n^{bc})^{-1}\left[\frac{1}{n_2}\Pi_{\mathbb{S}_+^{d_\theta}}\left\{\mathrm{sym}(\mathbf{M}_n^{dp}) - \boldsymbol{\psi}_n^{dp}(\boldsymbol{\psi}_n^{dp})^\top\right\} + \sigma_{\psi,n}^2\mathbf{I}_{d_\theta}\right](\widehat{\mathbf{H}}_n^{bc})^{-1}. \tag{13}$$

Then,

$$n_2\widehat{\mathbf{V}}_n^{bc} \xrightarrow{\mathbb{P}} \mathbf{A}_0^{-1}\mathrm{Var}[\mathrm{clip}_{c_\psi}\{\boldsymbol{\psi}(W,\boldsymbol{\eta}_0)\}]\mathbf{A}_0^{-1}.$$

If the clipped-score covariance is positive definite, it follows that

$$(\widehat{\boldsymbol{\eta}}_n^{bc} - \boldsymbol{\eta}_0)^\top(\widehat{\mathbf{V}}_n^{bc})^{-1}(\widehat{\boldsymbol{\eta}}_n^{bc} - \boldsymbol{\eta}_0) \Rightarrow \chi_{d_\theta}^2.$$

Proof. The corrected score in (10) equals

$$\mathbb{E}\{\boldsymbol{\psi}(W,\widetilde{\boldsymbol{\eta}}_n)\} + \boldsymbol{Z}_{\psi,n} + \frac{1}{n_2}\sum_{i\in\mathcal{D}_2}\left[\mathrm{clip}_{c_\psi}\{\boldsymbol{\psi}(W_i,\widetilde{\boldsymbol{\eta}}_n)\} - \mathbb{E}\{\mathrm{clip}_{c_\psi}(\boldsymbol{\psi}(W,\widetilde{\boldsymbol{\eta}}_n))\}\right].$$

By Lemma 1 and Assumption 3, $\left\|\mathrm{clip}_{c_\psi}\{\boldsymbol{\psi}(W,\widetilde{\boldsymbol{\eta}}_n)\} - \mathrm{clip}_{c_\psi}\{\boldsymbol{\psi}(W,\boldsymbol{\eta}_0)\}\right\|_2 \le G_H\|\widetilde{\boldsymbol{\eta}}_n - \boldsymbol{\eta}_0\|_2$. Therefore, conditional on $\mathcal{F}_1$, the mean squared norm of the centered empirical increment is at most $G_H^2\|\widetilde{\boldsymbol{\eta}}_n - \boldsymbol{\eta}_0\|_2^2/n_2$. Thus, replacing the pilot by $\boldsymbol{\eta}_0$ in the centered sum above costs $o_p(n_2^{-1/2})$. Taylor expansion of the population score gives

$$\mathbb{E}\{\boldsymbol{\psi}(W,\widetilde{\boldsymbol{\eta}}_n)\} = \mathbf{A}_0(\widetilde{\boldsymbol{\eta}}_n - \boldsymbol{\eta}_0) + O_p(\|\widetilde{\boldsymbol{\eta}}_n - \boldsymbol{\eta}_0\|_2^2).$$

The final term is $o_p(n_2^{-1/2})$.

The corrected Hessian before projection has mean $\mathbb{E}\{\mathbf{H}(W,\widetilde{\boldsymbol{\eta}}_n)\}$ and centered sampling error $O_p(n_2^{-1/2})$. Lemma 1 and the ridge condition imply

$$\left\|\widehat{\mathbf{H}}_n^{bc} - \mathbf{A}_0\right\|_F = O_p\big(n_2^{-1/2} + \|\widetilde{\boldsymbol{\eta}}_n - \boldsymbol{\eta}_0\|_2\big) + O(\lambda_n) = o_p(n_2^{-1/4}).$$

Substitution into the raw update in (10) cancels the leading pilot term. The remaining pilot contribution is

$$(\widehat{\mathbf{H}}_n^{bc})^{-1}(\widehat{\mathbf{H}}_n^{bc} - \mathbf{A}_0)(\widetilde{\boldsymbol{\eta}}_n - \boldsymbol{\eta}_0) = o_p(n_2^{-1/2}).$$

Replacing the inverse by $\mathbf{A}_0^{-1}$ in the centered score term gives (11) first for $\widehat{\boldsymbol{\eta}}_{n,\mathrm{raw}}^{bc}$. This expansion implies raw-update consistency, so $\boldsymbol{\eta}_0 \in \mathrm{int}(K)$ makes projection inactive with probability tending to one. The same expansion therefore holds for $\widehat{\boldsymbol{\eta}}_n^{bc}$ as stated. The central limit theorem gives (12). The released score and second moment converge to the mean and uncentered second moment of the clipped score, respectively. Subtracting the outer product of the released score and applying Lemma 1 therefore consistently estimates its covariance.This proves covariance consistency and the Wald limit. □

The same conclusions hold if the score and Hessian correction expectations are replaced by estimates whose errors at the pilot are $o_p(n_2^{-1/2})$ in Euclidean norm and $o_p(n_2^{-1/4})$ in Frobenius norm, respectively. An empirical tail correction based on $\mathcal{D}_2$ generally has error of order $n_2^{-1/2}$, whose variance and covariance with the clipped score must be included. Any correction from protected data also requires additional privacy accounting.

Growing thresholds. The original estimator can also retain valid inference if publicly specified thresholds increase with n_2. For this part, retain the chart and localization conditions in Assumption 3 and Assumptions 4, 6, and 7. Suppose the loss is twice continuously differentiable on K and, in place of the deterministic derivative bounds,

$$\begin{gathered}\mathbb{E}\left[\sup_{\boldsymbol{\eta}\in K}\|\boldsymbol{\psi}(W,\boldsymbol{\eta})\|_2^q\right] < \infty \quad \text{for some } q > 2,\\ \mathbb{E}\left[\sup_{\boldsymbol{\eta}\in K}\|\mathbf{H}(W,\boldsymbol{\eta})\|_F^2\right] < \infty,\\ \mathbb{E}\left[\sup_{\boldsymbol{\eta},\boldsymbol{\eta}'\in K,\boldsymbol{\eta}\neq\boldsymbol{\eta}'}\frac{\|\mathbf{H}(W,\boldsymbol{\eta}) - \mathbf{H}(W,\boldsymbol{\eta}')\|_F}{\|\boldsymbol{\eta} - \boldsymbol{\eta}'\|_2}\right] < \infty.\end{gathered} \tag{14}$$

Assume the $o_p(n_2^{-1/4})$ pilot rate and let $c_\psi \asymp n_2^{1/q}$, $c_H \asymp n_2^{1/2}$, $c_M = c_\psi^2$. Thus, the thresholds depend on the sample size. Using the raw empirical functions defined before Lemma 3, we have

$$\begin{aligned} \|\boldsymbol{\psi}_n - \boldsymbol{S}_n^\circ(\widetilde{\boldsymbol{\eta}}_n)\|_2 &= o_p(n_2^{-1/2}), \\ \|\mathbf{H}_n - \mathbf{H}_n^\circ(\widetilde{\boldsymbol{\eta}}_n)\|_F &= O_p(n_2^{-1/2}), \\ \|\mathbf{M}_n - \mathbf{M}_n^\circ(\widetilde{\boldsymbol{\eta}}_n)\|_F &= o_p(1). \end{aligned} \tag{15}$$

For example, the expectation of the uniform score-clipping error is bounded by

$$\mathbb{E}\left[\sup_{\boldsymbol{\eta}\in K}\left\|\boldsymbol{\psi}(W,\boldsymbol{\eta}) - \mathrm{clip}_{c_\psi}\{\boldsymbol{\psi}(W,\boldsymbol{\eta})\}\right\|_2\right] \le \frac{\mathbb{E}[\sup_{\boldsymbol{\eta}\in K}\|\boldsymbol{\psi}(W,\boldsymbol{\eta})\|_2^q]}{c_\psi^{q-1}} = O(n_2^{-1+1/q}).$$

This is $o(n_2^{-1/2})$. The corresponding Hessian and second-moment bounds are

$$\frac{\mathbb{E}[\sup_{\boldsymbol{\eta}\in K}\|\mathbf{H}(W,\boldsymbol{\eta})\|_F^2]}{c_H} = O(n_2^{-1/2}), \qquad \frac{\mathbb{E}[\sup_{\boldsymbol{\eta}\in K}\|\boldsymbol{\psi}(W,\boldsymbol{\eta})\|_2^q]}{c_\psi^{q-2}} = O(n_2^{-1+2/q}).$$

Markov's inequality gives (15). The Gaussian noise satisfies

$$\|\boldsymbol{Z}_{\psi,n}\|_2 = O_p(n_2^{-1+1/q}), \quad \|\mathbf{Z}_{H,n}\|_F = O_p(n_2^{-1/2}), \quad \|\mathbf{Z}_{M,n}\|_F = O_p(n_2^{-1+2/q}).$$

The last condition in (14) makes the empirical Hessian Lipschitz constant $O_p(1)$. The Taylor argument in Lemma 3 therefore gives

$$\boldsymbol{S}_n^\circ(\widetilde{\boldsymbol{\eta}}_n) = \boldsymbol{S}_n^\circ(\boldsymbol{\eta}_0) + \mathbf{H}_n^\circ(\widetilde{\boldsymbol{\eta}}_n)(\widetilde{\boldsymbol{\eta}}_n - \boldsymbol{\eta}_0) + o_p(n_2^{-1/2}).$$

Moreover, the Hessian moment condition and the pilot rate imply $\mathbf{H}_n^\circ(\widetilde{\boldsymbol{\eta}}_n) - \mathbf{A}_0 = o_p(n_2^{-1/4})$. By (15), the Hessian noise bound, and $\mathbf{A}_0 \succ 0$,

$$\left\|\widehat{\mathbf{H}}_n^{dp} - \mathbf{H}_n^\circ(\widetilde{\boldsymbol{\eta}}_n)\right\|_F = O_p(n_2^{-1/2}).$$

Substitution into the raw one-step update yields the expansion below first for $\widehat{\boldsymbol{\eta}}_{n,\mathrm{raw}}^{dp}$. It implies raw-update consistency, and hence the projection is inactive with probability tending to one. Thus the projected update also satisfies

$$\widehat{\boldsymbol{\eta}}_n^{dp} - \boldsymbol{\eta}_0 = -\mathbf{A}_0^{-1}\left\{\frac{1}{n_2}\sum_{i\in\mathcal{D}_2}\boldsymbol{\psi}(W_i,\boldsymbol{\eta}_0) + \boldsymbol{Z}_{\psi,n}\right\} + o_p(n_2^{-1/2}). \tag{16}$$

The score noise is $o_p(n_2^{-1/2})$ because $q > 2$. The central limit theorem therefore gives

$$\sqrt{n_2}(\widehat{\boldsymbol{\eta}}_n^{dp} - \boldsymbol{\eta}_0) \Rightarrow N(0, \mathbf{A}_0^{-1}\mathbf{J}_0\mathbf{A}_0^{-1}).$$

Finally, the score and Hessian moment conditions imply

$$\mathbb{E}\left[\sup_{\boldsymbol{\eta}\in K}\|\boldsymbol{\psi}(W,\boldsymbol{\eta})\|_2 \sup_{\boldsymbol{\eta}\in K}\|\mathbf{H}(W,\boldsymbol{\eta})\|_F\right] < \infty$$

by Cauchy–Schwarz. Together with pilot consistency, this gives $\mathbf{M}_n^\circ(\widetilde{\boldsymbol{\eta}}_n) \xrightarrow{\mathbb{P}} \mathbf{J}_0$ by the argument of Lemma 4. The clipping and release errors vanish, so

$$n_2\widehat{\mathbf{V}}_n^{dp} \xrightarrow{\mathbb{P}} \mathbf{A}_0^{-1}\mathbf{J}_0\mathbf{A}_0^{-1}.$$

The original Wald limit follows if $\mathbf{J}_0 \succ 0$.