BOOTSTRAP INFERENCE FOR RIEMANNIAN OPTIMIZATION PROBLEMS

Anonymous authors
Paper under double-blind review

ABSTRACT

We develop an online multiplier bootstrap for manifold-valued parameters estimated by averaging the iterates of Riemannian stochastic approximation. The method runs weighted iterations alongside the original algorithm, with each path using the same observations while updating its own state and intrinsic average. The weighted averages provide uncertainty assessments without a plug-in asymptotic covariance estimator. Under local stability and regularity conditions, we establish bootstrap consistency and show that a fixed number of paths suffices for asymptotically valid Student intervals and intrinsic confidence regions. The theory accommodates samplewise nonsmooth scores and identifies a critical oracle-bias regime in which the estimator's limiting distribution shifts but the centered bootstrap distribution does not. Applications include spherical means, positive-definite matrix means and medians, and principal-subspace estimation. Simulations show coverage close to nominal in regular settings, and an analysis of North Atlantic storm locations illustrates descriptive online inference for a spherical center.

1 INTRODUCTION

Many stochastic optimization problems have parameters that are not naturally represented by unconstrained vectors. Principal subspaces lie on Grassmann manifolds, covariance and diffusion tensors lie on the cone of positive-definite matrices, and directional parameters lie on spheres. Retraction-based stochastic approximation respects these constraints without sacrificing sequential computation (Absil et al., 2008; Bonnabel, 2013); iterative intrinsic averaging is the corresponding Polyak-Ruppert construction and can recover root-n statistical accuracy under local stability and smoothness (Ruppert, 1988; Polyak and Juditsky, 1992; Tripuraneni et al., 2018).

The inferential obstacle is estimating the asymptotic covariance. For a regular root-n averaged estimator, the limiting covariance has the form

$$V = H^{-1}Q(H^{-1})^{\dagger},$$

where H is the derivative of the population vector field, Q is the covariance operator of the noise term at the target, and $A^{\dagger}$ denotes the adjoint of a linear operator A. A plug-in estimator must estimate both objects in compatible coordinates, which is difficult when the stochastic field is samplewise nonsmooth, when H contains nuisance structure, or when a streaming procedure cannot retain matrix summaries. Geometry adds a second obstacle: fields and derivatives at different iterates belong to different tangent spaces.

We construct an online bootstrap by running randomly weighted copies of the original algorithm alongside it. Each copy has independent weights and its own estimate and intrinsic average, but shares the observations and auxiliary random inputs with the original algorithm; variation among the weighted averages provides the uncertainty assessment. Since two manifold estimates cannot generally be subtracted as vectors, we compare them in the tangent space at the observed average and prove that this computable comparison reproduces the estimator's sampling fluctuation to first order.

We make four contributions.

(i) We establish valid online bootstrap inference for averaged Riemannian stochastic approximation without estimating an asymptotic covariance; under local stability and regularity conditions, the theory accommodates samplewise nonsmooth scores and a nonsymmetric population derivative.

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.

(ii) We justify comparing weighted averages at the observed estimate rather than at the unknown target. This resolves the need to compare different tangent spaces and makes the bootstrap statistic computable.

(iii) We show that a fixed number of weighted copies suffices for asymptotically valid Student intervals and intrinsic confidence regions. As the number of copies grows, the same construction yields consistent conditional quantiles and, under additional integrability, covariance estimates.

(iv) We specialize the theory to spherical means, positive-definite matrix means and medians, and principal subspaces. For approximate updates, we identify when accumulated bias is negligible and when it shifts the estimator's limit without being reproduced by the centered bootstrap.

For Euclidean smooth averaged SGD with exponential weights, our iteration and fixed-B scalar interval reduce exactly to the Cheap Online Bootstrap (COnB) of Lam and Wang (2026); our results extend this construction to averaged estimates on manifolds of fixed dimension through a local theory.

The numerical evidence follows this local scope: regular designs on three geometries have coverage near the nominal level, the localized skew spherical experiment gives coverage closer to nominal than a matched offline trajectory bootstrap at $n = 100$, and the IBTrACS analysis shows the same paths providing descriptive online uncertainty for a spherical location summary.

1.1 Related work

Full-path multipliers avoid explicit sandwich estimation: Fang et al. (2018) evolve perturbed stochastic-gradient paths beside the original iteration, evaluating each score at its own perturbed state, and Lam and Wang (2026) combine this construction with a fixed number of paths and a Student pivot, so that COnB is valid for smooth strongly convex averaged SGD even when B is fixed. Other Euclidean approaches estimate covariance online (Zhu et al., 2023), use batch means for nonsmooth or nonmonotone stochastic approximation (Jiang et al., 2025), treat local nonconvex inference (Zhong et al., 2023), or establish quantitative approximations for linear stochastic approximation and smooth SGD (Samsonov et al., 2024; Sheshukova et al., 2026). In the derivative-free setting, Jin et al. (2021) derive a biased central limit theorem for Euclidean zeroth-order Polyak-Ruppert averages and estimate its covariance online. Our approximation-bias analysis instead concerns manifold full-path multipliers and the cumulative predictable bias relative to the unsmoothed target. Offline bootstrap theory for manifold-valued means is classical (Bhattacharya and Patrangenaru, 2005), but it does not resolve comparison across tangent spaces for a continuously updated stochastic-approximation path.

Streaming principal-component inference is the closest geometric comparison: existing results give high-dimensional approximations and online multiplier procedures for the rank-one Oja error (Lunde et al., 2021), entrywise inference with subsampling alternatives (Kumar et al., 2025), and Gaussian and multiplier approximations for the final general-rank Oja iterate (Xu and Li, 2026). These results exploit PCA structure and target final iterates in high-dimensional regimes, whereas we study averaged iterations on a general manifold of fixed dimension, with Grassmann PCA as one specialization.

Riemannian stochastic approximation and iterative averaging provide the geometric foundation (Bonnabel, 2013; Tripuraneni et al., 2018); the additional task here is to compare complete weighted averages at the estimated base point with an influence representation in a fixed tangent space, which requires separate control of state stability, full-average localization, and geometric remainders. For geometric medians, Cheng et al. (2024) already develop online bootstrap inference using averaged stochastic gradients. Our median application treats the Riemannian normalized-distance score under explicit population-field and localization conditions.

The rest of the paper is organized as follows. Section 2 defines the online paths and intervals, Section 3 establishes their guarantees, and Section 4 evaluates finite-sample coverage; the appendices develop primitive conditions, large-replication inference, and geometric applications, alongside proofs, simulation details, and the North Atlantic storm analysis.

2 METHODOLOGY

2.1 RIEMANNIAN STOCHASTIC APPROXIMATION AND INTRINSIC AVERAGING

Let $(\mathcal{M}, g)$ be a connected d-dimensional Riemannian manifold, and let $x_\star \in \mathcal{M}$ be a locally unique zero of a population vector field. Write $d(x, y)$ for geodesic distance (d alone is the manifold dimension), $T_x\mathcal{M}$ for the tangent space at x with Riemannian norm $\|\cdot\|_x$, R_x for a retraction with locally defined inverse R_x^{-1}, and $\mathcal{P}_{x\to y}$ for parallel transport along the unique minimizing geodesic when x, y lie in a common normal neighborhood.

Let $\Psi_k(x; D_k) \in T_x\mathcal{M}$ be the stochastic oracle, the stochastic vector-field estimate used to update x; its input $D_k = (Z_k, S_k)$ consists of the observation Z_k and an auxiliary random input S_k generated at step k, such as a sampled direction, minibatch, or solver seed. The original iteration is

$$x_k = R_{x_{k-1}}\{-\eta_k \Psi_k(x_{k-1}; D_k)\}, \qquad \eta_k = \eta_0 k^{-\alpha}, \quad \eta_0 > 0, \quad \alpha \in (1/2, 1). \tag{1}$$

Its iterative intrinsic average is

$$\bar{x}_1 = x_1, \qquad \bar{x}_k = R_{\bar{x}_{k-1}}\left\{\frac{1}{k} R^{-1}_{\bar{x}_{k-1}}(x_k)\right\}, \qquad k \geq 2. \tag{2}$$

In Euclidean space, (2) computes the running arithmetic mean $\bar{x}_k = k^{-1}\sum_{j=1}^k x_j$.

2.2 FULL-PATH MULTIPLIER TRAJECTORIES

Let $(W_k)_{k\geq 1}$ be i.i.d., independent of the complete data and oracle-randomness stream, with

$$\mathbb{E}W_k = 1, \qquad \mathbb{E}(W_k - 1)^2 = 1. \tag{3}$$

A multiplier path starts from $x_0^\star = x_0$ and evolves according to

$$x_k^\star = R_{x_{k-1}^\star}\{-\eta_k W_k \Psi_k(x_{k-1}^\star; D_k)\}, \tag{4}$$

$$\bar{x}_1^\star = x_1^\star, \qquad \bar{x}_k^\star = R_{\bar{x}_{k-1}^\star}\left\{\frac{1}{k} R^{-1}_{\bar{x}_{k-1}^\star}(x_k^\star)\right\}. \tag{5}$$

The original and multiplier paths share D_k, but each path evaluates the oracle at its own state $x_{k-1}^\star$; this distinguishes the full-path procedure from a linearized resampling scheme that only reweights stored gradients.

2.3 INTERVALS AND IMPLEMENTATION

For a differentiable scalar functional ϕ and fixed $B \geq 1$, run B multiplier paths with independent weights. With confidence error probability $\tau \in (0, 1)$, define

$$s^2_{\phi,n,B} = \frac{1}{B}\sum_{b=1}^B \{\phi(\bar{x}_n^{\star(b)}) - \phi(\bar{x}_n)\}^2, \tag{6}$$

$$I_{\phi,n,B}(1-\tau) = \left[\phi(\bar{x}_n) - t_{B,1-\tau/2} s_{\phi,n,B},\ \phi(\bar{x}_n) + t_{B,1-\tau/2} s_{\phi,n,B}\right]. \tag{7}$$

Here $t_{B,1-\tau/2}$ is the Student quantile with B degrees of freedom; the differences are centered at the original estimate. If $s_{\phi,n,B} = 0$, report $\mathbb{R}$. For intrinsic regions, use the second moment of the vectors $\sqrt{n}R^{-1}_{\bar{x}_n}(\bar{x}_n^{\star(b)})$ and the Hotelling-type critical value $dBF_{d,B-d+1;1-\tau}/(B-d+1)$, with $B \geq d$, where $F_{\nu_1,\nu_2;q}$ is the q-quantile of the F distribution. The completion is defined below; the singularity fallback and the intrinsic and large-B regions are specified in Section C.1.1. When $B = B_n \to \infty$, empirical conditional quantiles yield the basic interval (35).

All paths share each observation and oracle seed; any safeguard is applied after the complete oracle is multiplied, and each path updates its own full average with the required inverse maps defined along its history. Paths for a chosen horizon are initialized before the stream begins. Algorithm 1 gives the complete procedure with reporting fallbacks; for an $O(d)$ state representation, the extra storage is $O(Bd)$ and each observation costs B additional oracle evaluations and retractions.

Algorithm 1 Online full-path multiplier inference

Require: Initial state x_0; steps (η_k); retraction and local inverse;
oracle and any state safeguard; B paths; multiplier law; reporting rule.

Initialize $x_0^{\star(b)} \leftarrow x_0$ for $b = 1, \dots, B$.
for $k = 1, \dots, n$ **do**
  Read Z_k, generate the oracle seed S_k, and set $D_k = (Z_k, S_k)$.
  Evaluate $\Psi_k(x_{k-1}; D_k)$ and update x_k by (1), including any specified safeguard.
  If $k = 1$, set $\bar{x}_1 \leftarrow x_1$; otherwise update $\bar{x}_k$ by (2).
  for $b = 1, \dots, B$ **do**
    Draw $W_k^{(b)}$ independently of the original stream and other weights.
    Evaluate $\Psi_k(x_{k-1}^{\star(b)}; D_k)$ at this path's own state.
    Update $x_k^{\star(b)}$ by (4), multiplying the entire oracle before applying any specified safeguard.
    If $k = 1$, set $\bar{x}_1^{\star(b)} \leftarrow x_1^{\star(b)}$; otherwise update $\bar{x}_k^{\star(b)}$ by (5).
At reporting time, form the observable completed vectors $\widetilde{T}_{n,b}^{\star}$ in (21) if needed.
if fixed-B calibration is requested **then**
  Report (7) or (33), with their stated fallbacks.
else
  Report (35) or (36); use (38) only
  under the covariance conditions and with its rank/singularity fallback.

3 Theoretical Results

Work in $T_{x_\star}\mathcal{M}$ with its norms and inner products, and write $x(a) = R_{x_\star}(a)$, $u_k = R_{x_\star}^{-1}(x_k)$ and $u_k^\star = R_{x_\star}^{-1}(x_k^\star)$. Let $\mathcal{F}_k = \sigma(D_1, \dots, D_k)$ and $\mathcal{G}_k = \mathcal{F}_k \vee \sigma(W_1, \dots, W_k)$. Let $\mathcal{F}_\infty = \sigma(\cup_{k\geq 1}\mathcal{F}_k)$. Matrix symbols are upright, including coordinate covariance matrices; H, Q and V denote basis-free operators on $T_{x_\star}\mathcal{M}$. For $a, b \in T_{x_\star}\mathcal{M}$, $(a \otimes b)v = \langle b, v\rangle\, a$. The subscript $\star$ denotes the target, the superscript $\star$ a multiplier path, and a star on probability or law denotes conditioning on $\mathcal{F}_\infty$. The metric d_{BL} is the bounded-Lipschitz distance, and $\rightsquigarrow$ denotes convergence in distribution.

3.1 Conditions and main guarantees

The assumptions control the innovation law, accumulated iteration errors, and geometric comparison; the predictable covariance and conditional Lindeberg conditions below are standard martingale central limit conditions (Hall and Heyde, 1980, Chapter 3), and the theory appendix supplies sufficient conditions, state bounds and geometric examples.

Assumption 3.1 (Innovation and multiplier conditions)**.** There are $T_{x_\star}\mathcal{M}$-valued martingale differences $(\xi_k, \mathcal{F}_k)$ and a deterministic positive-semidefinite operator Q such that

$$\frac{1}{n}\sum_{k=1}^{n} \mathbb{E}(\xi_k \otimes \xi_k \mid \mathcal{F}_{k-1}) \xrightarrow{p} Q, \tag{8}$$

$$\frac{1}{n}\sum_{k=1}^{n} \mathbb{E}\left[\|\xi_k\|^2\, \mathbf{1}\{\|\xi_k\| > \varepsilon\sqrt{n}\} \mid \mathcal{F}_{k-1}\right] \xrightarrow{p} 0 \quad \text{for every } \varepsilon > 0. \tag{9}$$

The multipliers are i.i.d., independent of $\mathcal{F}_\infty$, and satisfy

$$\mathbb{E}W = 1, \qquad \mathbb{E}(W-1)^2 = 1. \tag{10}$$

Assumption 3.2 (Averaged local iterations)**.** There is a nonsingular deterministic linear operator $H : T_{x_\star}\mathcal{M} \to T_{x_\star}\mathcal{M}$ and adapted arrays $b_k, m_k, e_k, m_k^\star, e_k^\star$ such that, on the local chart,

$$u_k = u_{k-1} - \eta_k\{Hu_{k-1} + \xi_k + b_k + m_k + e_k\}, \tag{11}$$

$$u_k^\star = u_{k-1}^\star - \eta_k\{Hu_{k-1}^\star + W_k\xi_k + W_k b_k + m_k^\star + e_k^\star\}. \tag{12}$$

Here b_k is $\mathcal{F}_{k-1}$-measurable, m_k is an $(\mathcal{F}_k)$ martingale difference, and $m_k^\star$ is a $(\mathcal{G}_k)$ martingale difference and may include the centered multiplier fluctuation $(W_k - 1)Hu_{k-1}^\star$. The accumulated

conditions are

$$\frac{1}{\sqrt{n}}\sum_{k=1}^{n} b_k \xrightarrow{p} 0, \qquad \frac{1}{n}\sum_{k=1}^{n} \|b_k\|^2 \xrightarrow{p} 0, \tag{13}$$

$$\frac{1}{n}\sum_{k=1}^{n} \mathbb{E}(\|m_k\|^2 \mid \mathcal{F}_{k-1}) \xrightarrow{p} 0, \qquad \frac{1}{n}\sum_{k=1}^{n} \mathbb{E}(\|m_k^\star\|^2 \mid \mathcal{G}_{k-1}) \xrightarrow{p} 0, \tag{14}$$

$$\frac{1}{\sqrt{n}}\sum_{k=1}^{n} e_k \xrightarrow{p} 0, \qquad \frac{1}{\sqrt{n}}\sum_{k=1}^{n} e_k^\star \xrightarrow{p} 0 \quad \text{under the joint law.} \tag{15}$$

Finally,

$$\sup_{k\geq 1} \eta_k^{-1}\mathbb{E}\,\|u_k\|^2 < \infty, \qquad \sup_{k\geq 1} \eta_k^{-1}\mathbb{E}\mathbb{E}^\star\,\|u_k^\star\|^2 < \infty. \tag{16}$$

Here $\mathbb{E}^\star$ denotes expectation over the multiplier stream conditional on $\mathcal{F}_\infty$.

Assumption 3.3 (Intrinsic averaging and comparison at the estimated base point)**.** For a, b in a sufficiently small ball in $T_{x_\star}\mathcal{M}$ and $0 \leq t \leq 1$, define

$$A_t(a,b) = R_{x_\star}^{-1}\left[R_{x(a)}\{tR_{x(a)}^{-1}(x(b))\}\right]. \tag{17}$$

There is a constant C such that

$$\|A_t(a,b) - \{(1-t)a + tb\}\| \leq Ct(\|a\|^2 + \|b\|^2). \tag{18}$$

Moreover, with

$$\Gamma(a,b) = \mathcal{P}_{x(a)\to x_\star} R_{x(a)}^{-1}\{x(b)\}, \tag{19}$$

we have

$$\|\Gamma(a,b) - (b-a)\| \leq C(\|a\| + \|b\|)\,\|b-a\|. \tag{20}$$

The original and weighted states and iterative averages have finite coordinates in this chart under localization and enter a common comparison ball $B_{T_{x_\star}\mathcal{M}}(0,r)$ after an almost surely finite path-specific index. The radius r is sufficiently small for the intrinsic averaging bound to yield the $O(k^{-\alpha})$ average rate. It is permissible for the safeguarded update to differ from its local coordinate formula at finitely many path-specific indices, provided the exact coordinate discrepancies at those indices are included in the arrays e_k and $e_k^\star$ in (11)–(12). The reported averages are the full iterative averages in (2) and (5); no common all-path restart is used.

These conditions retain familiar requirements of averaged stochastic approximation. Fang et al. (2018, Assumptions A1–A6) likewise use $\alpha \in (1/2, 1)$, population smoothness and a local L^2 score modulus, without samplewise differentiability. The state rate in (16) and local retraction requirements parallel Assumptions 1–2 of Tripuraneni et al. (2018). Unlike the samplewise smooth, strongly convex setting of Lam and Wang (2026, Assumption 1), our sufficient conditions allow nonsmooth scores and a nonsymmetric population derivative. The additional geometric requirement is joint control of the original and weighted states and full averages in a common chart; this must be verified for each application.

3.2 Full-path validity

One sufficient route separates population smoothness from samplewise regularity. A locally $C^{1,\rho}$ population field has remainder $O(\|u\|^{1+\rho})$, while the centered stochastic field needs only a vanishing local L^2 modulus. Given the state bounds, $\alpha(1+\rho) > 1$ makes the accumulated population remainder negligible; for $\rho = 1$, $\alpha > 1/2$ suffices. Second-order update compatibility controls geometric errors. Full-average localization also needs verification: an invariant contracting averaging basin suffices, whereas tangent-step clipping alone does not ensure a common chart. The theory appendix develops this route.

Choose a normal neighborhood $U_\star$ of $x_\star$ on which transport from $\bar{x}_n$ to $x_\star$ is unique. Define the completed comparison vectors by

$$\begin{aligned} \widetilde{T}_n^\star &= \begin{cases} \sqrt{n}R_{\bar{x}_n}^{-1}(\bar{x}_n^\star), & \text{if the inverse retraction is available,} \\ 0_{\bar{x}_n}, & \text{otherwise,} \end{cases} \\ T_n^\star &= \begin{cases} \mathcal{P}_{\bar{x}_n\to x_\star}\widetilde{T}_n^\star, & \bar{x}_n \in U_\star, \\ 0_{x_\star}, & \text{otherwise.} \end{cases} \end{aligned} \tag{21}$$

The transport to $x_\star$ is used only in the analysis; reported intervals and regions use the observable differences at $\bar{x}_n$, and the assumptions make the completion events negligible. The joint expansions remove the state-dependent fluctuations, and the random-base comparison then justifies centering at the observed average.

Theorem 3.4 (Joint asymptotic linearity)**.** *Suppose Assumptions 3.1 to 3.3 hold with $\eta_k = \eta_0 k^{-\alpha}$ and $\alpha \in (1/2, 1)$. Then*

$$\sqrt{n}\, R_{x_\star}^{-1}(\bar{x}_n) = -H^{-1} \frac{1}{\sqrt{n}} \sum_{k=1}^{n} \xi_k + o_p(1), \tag{22}$$

$$\sqrt{n}\, R_{x_\star}^{-1}(\bar{x}_n^\star) = -H^{-1} \frac{1}{\sqrt{n}} \sum_{k=1}^{n} W_k \xi_k + o_{\mathbb{P}^\star}(1) \quad \textit{in probability}. \tag{23}$$

Theorem 3.5 (Conditional full-path multiplier validity)**.** *Under the assumptions of Theorem 3.4, let $T_n^\star$ be as in* (21) *and $V = H^{-1} Q (H^{-1})^\dagger$. Then*

$$T_n^\star = -H^{-1} \frac{1}{\sqrt{n}} \sum_{k=1}^{n} (W_k - 1) \xi_k + o_{\mathbb{P}^\star}(1) \quad \textit{in probability}, \tag{24}$$

and

$$d_{\mathrm{BL}}\{\mathcal{L}^\star(T_n^\star), N(0, V)\} \xrightarrow{p} 0. \tag{25}$$

Moreover,

$$d_{\mathrm{BL}}\left[\mathcal{L}^\star(T_n^\star), \mathcal{L}\{\sqrt{n} R_{x_\star}^{-1}(\bar{x}_n)\}\right] \xrightarrow{p} 0. \tag{26}$$

Summation by parts reduces each local iteration to its target innovations and negligible accumulated errors. The intrinsic averaging bound makes the iterative manifold average first-order equivalent to its coordinate average, so both averaged coordinates are $O_p(n^{-1/2})$ under the joint law. Substituting them into (20) makes the inverse-retraction error $O_p(n^{-1})$, negligible after multiplication by $\sqrt{n}$. Subtracting the linear representations leaves the centered multiplier sum in (24). Geometry changes this comparison step, but not the first-order covariance.

3.3 Fixed-path inference

Put $Z_n = \sqrt{n} R_{x_\star}^{-1}(\bar{x}_n)$ and construct $T_{n,b}^\star$ from conditionally independent multiplier paths. Their joint independence in the limit accounts for the uncertainty from using finitely many paths.

Theorem 3.6 (Fixed-replication joint limit)**.** *Under Theorem 3.5, for every fixed integer $B \geq 1$,*

$$(Z_n, T_{n,1}^\star, \ldots, T_{n,B}^\star) \rightsquigarrow (Z_0, Z_1, \ldots, Z_B), \tag{27}$$

where $Z_0, \ldots, Z_B$ are independent centered Gaussian tangent vectors, each with covariance V.

Corollary 3.7 (Student and Hotelling-type regions)**.** *Suppose Theorem 3.6 holds.*

(i) Let $\phi : \mathcal{M} \to \mathbb{R}$ be differentiable at $x_\star$, with differential $D\phi_{x_\star} : T_{x_\star}\mathcal{M} \to \mathbb{R}$, suppose $\sigma_\phi^2 = D\phi_{x_\star} V (D\phi_{x_\star})^\dagger > 0$. For every fixed $B \geq 1$, the interval (7) *satisfies*

$$\mathbb{P}\{\phi(x_\star) \in I_{\phi,n,B}(1-\tau)\} \longrightarrow 1 - \tau. \tag{28}$$

(ii) If V is positive definite and $B \geq d$, the intrinsic region (33) *satisfies*

$$\mathbb{P}\{x_\star \in \mathcal{C}_{n,B}^{\mathrm{fix}}(1-\tau)\} \longrightarrow 1 - \tau. \tag{29}$$

The scalar estimation error is asymptotically independent of B Gaussian bootstrap differences with the same variance, so their root mean square gives a Student-t_B denominator; the degrees of freedom are B, not $B - 1$, because centering uses the original estimate rather than the empirical bootstrap mean. Fixed-B calibration thus accounts for replication uncertainty without estimating the entire bootstrap distribution.

The t_B penalty can make $B = 1$ intervals wide. With growing replication, conditional quantiles remain valid, while covariance estimation additionally requires conditional uniform integrability; the theory appendix gives both guarantees.

3.4 APPROXIMATION BIAS AND ITS FAILURE BOUNDARY

Approximate gradients, random-direction secants, and smoothed nonsmooth scores can introduce a predictable bias b_k; the multiplier paths reproduce the random influence term but not automatically a deterministic root-n shift.

Proposition 3.8 (Polynomial approximation-bias regimes and subcritical diagnostic)**.** *For constants $h_0 > 0$, $r > 0$, $\beta > 0$, a nonzero $B_0 \in T_{x_\star}\mathcal{M}$, and a remainder $\widetilde{b}_k$, suppose*

$$b_k = h_0^r k^{-r\beta} B_0 + \widetilde{b}_k, \qquad \frac{1}{\sqrt{n}} \sum_{k=1}^{n} \widetilde{b}_k \xrightarrow{p} 0, \qquad \frac{1}{n} \sum_{k=1}^{n} \left\| \widetilde{b}_k \right\|^2 \xrightarrow{p} 0. \tag{30}$$

Put $\gamma = r\beta$. For $\gamma \geq 1/2$, suppose in addition that Assumptions 3.1 and 3.3 hold and that all parts of the abstract iteration other than the first condition in (13)*, including the state bounds* (16)*, hold with the displayed b_k.*

(i) If $\gamma > 1/2$, the approximation bias is negligible at the root-n scale, and the conclusions of Theorems 3.4 and 3.5 hold.

(ii) If $\gamma = 1/2$, with $V = H^{-1}Q(H^{-1})^\dagger$,

$$\sqrt{n} R_{x_\star}^{-1}(\bar{x}_n) \rightsquigarrow N\{-2h_0^r H^{-1} B_0, V\}. \tag{31}$$

The centered multiplier law still satisfies (25)*, so it does not reproduce the shifted sampling law about $x_\star$. A scalar functional can remain unshifted if its derivative annihilates $H^{-1}B_0$.*

(iii) If $0 < \gamma < 1/2$ and a bias-aware stability analysis separately justifies the coordinate Abel rearrangement, its deterministic leading term is

$$-\frac{h_0^r}{1-\gamma} n^{1/2-\gamma} H^{-1} B_0 \{1 + o(1)\}, \tag{32}$$

whereas the centered weighted bias fluctuation is $o_{\mathbb{P}^\star}(1)$. This coordinate-level calculation alone does not establish the state, intrinsic-average, or random-base remainders for the implemented manifold statistic, and no bootstrap-validity conclusion is asserted in this subcritical regime.

At the boundary, undersmoothing, an analytically justified bias correction, or a multiscale extrapolation that removes the leading B_0 term is required for centered inference about $x_\star$. The subcritical formula is retained only as a diagnostic for the larger deterministic coordinate scale.

For smooth Euclidean averaged SGD with exponential weights, the procedure reduces exactly to COnB, see C.1. On a manifold, the same inference requires localization of the weighted states and full averages and a comparison at the observed base point; primitive conditions and geometric comparisons give sufficient routes for these steps. The geometric applications specialize the result to spherical means, positive-definite means and medians, and principal subspaces, see C.3

4 NUMERICAL EXPERIMENTS

We evaluate the finite-sample coverage of the proposed online multiplier intervals (Proposed; fixed-path Student or 200-path quantile versions) for scalar functionals: first with few bootstrap paths in spherical examples, then across three manifold geometries and with a samplewise nonsmooth score, and finally under different multiplier distributions. All intervals are nominally 95%, with empirical coverage from 500 independent data streams per configuration. Unless otherwise stated, the iterations use step sizes $\eta_k = \eta_0 k^{-0.7}$ and exponential multiplier weights; design specifications, tuning procedures, and comparator definitions are in the simulation appendix.

4.1 COVERAGE WITH FEW BOOTSTRAP PATHS

To assess fixed-path coverage in small samples, we use two laws on $\mathbb{S}^2$: a symmetric truncated tangent-normal law with $n \in \{250, 500\}$ and an asymmetric four-point law with $n \in \{100, 250\}$. Both have intrinsic mean at the north pole $\mathbf{e}_3$ and scalar estimand $\phi(x) = x_1$. Independent pilots

selected $\eta_0 = 0.8$ and 1.0, respectively, by base-estimator mean squared geodesic error before the 500-stream evaluations.

Table 1 compares Proposed Student intervals, Proposed (quantile), a tangent sandwich interval and a matched offline basic bootstrap with 200 trajectory reruns. Methods share observations, initialization, tuning and the base estimate; all paths average from the first update without burn-in. The offline method replays the same iteration on each resampled ordered dataset.

Table 1: Nominal 95% coverage in localized spherical designs. Proposed Student and quantile intervals use (7) and (35); the quantile and offline methods each use 200 replicates.

	Symmetric		Skew	
Method	$n = 250$	$n = 500$	$n = 100$	$n = 250$
Proposed ($B = 1$)	0.940	0.938	0.954	0.950
Proposed ($B = 3$)	0.956	0.954	0.942	0.938
Proposed ($B = 5$)	0.960	0.950	0.938	0.942
Proposed ($B = 10$)	0.952	0.946	0.936	0.932
Proposed (quantile)	0.968	0.944	0.932	0.924
Sandwich	0.942	0.930	0.910	0.910
Offline	0.904	0.926	0.840	0.878

With only three to ten multiplier paths, the proposed method achieves coverage near nominal and improves on offline replay across all four settings. The gains are most pronounced under the skew law at $n = 100$, highlighting the effectiveness of few-path inference even in a curved, asymmetric problem with a small sample. Coverage is also broadly comparable to that of sandwich and 200-path quantile intervals. The improved calibration relative to offline replay comes with wider intervals, although their lengths decrease substantially as B increases; full length comparisons are in the simulation appendix.

A wider-cap comparison uses $n = 1000$ observations from a von Mises–Fisher law on $\mathbb{S}^2$, with scalar estimand $\phi(x) = x_1$. Table 2 shows coverage near nominal for Proposed Student and sandwich intervals, versus 0.900 for offline replay.

Table 2: Nominal 95% coverage in the wider-cap spherical design ($q = 2$, $n = 1000$). Proposed (quantile) and offline basic intervals each use 200 replicates.

Proposed (Student)				Proposed (quantile)		Offline
$B = 1$	$B = 3$	$B = 5$	$B = 10$	$B = 200$	Sandwich	$B = 200$
0.948	0.958	0.950	0.970	0.964	0.948	0.900

4.2 Coverage across geometries and nonsmooth scores

We next consider spherical means, affine-invariant positive-definite matrix means, and principal subspaces on Grassmann manifolds. Their scalar functionals are, respectively, the coordinate x_1, the largest log eigenvalue at a simple-eigenvalue matrix target, and the projector entry $(\mathrm{U}\mathrm{U}^{\mathsf{T}})_{r,r+1}$, where U is an orthonormal basis of the principal subspace. Dimensions and distributional parameters are specified in the simulation appendix.

With $B = 10$, the Proposed intervals achieve coverage near nominal for spherical means, affine-invariant positive-definite means and Grassmann principal subspaces at $n \in \{5000, 20000\}$, without estimating population derivatives (Figure 1).

To assess a samplewise nonsmooth score, we compare the positive-definite Fréchet mean and geometric median. In the specified Gaussian model with sign-symmetric contamination, affine-invariant reflection symmetry and uniqueness preserve their common target. For 3×3 positive-definite matrices at $n = 5000$, Table 3 shows similar coverage near nominal at contamination probabilities 0, 0.05 and 0.15.

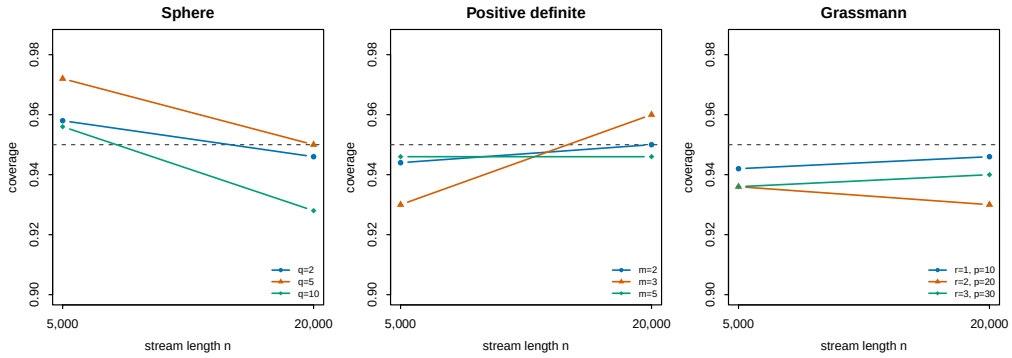

Figure 1: Proposed ($B = 10$) coverage in regular designs; the dashed line marks 95%. Spherical laws differ between horizons, so connecting segments are visual guides rather than same-law sample-size trends.

Table 3: Nominal 95% coverage of Proposed ($B = 10$) for the largest log eigenvalue of the common positive-definite mean and median (3×3 matrices, $n = 5000$).

Contamination probability	Fréchet mean	Geometric median
0	0.930	0.936
0.05	0.952	0.948
0.15	0.952	0.956

4.3 Sensitivity to the multiplier distribution

A separate cap design at $n = 5000$ compares weights on identical streams, with concentration 6, radius 1.20, start 0.45 radians from the north pole, $\eta_0 = 0.6$ and estimand $\phi(x) = x_1$ as in the wider-cap comparison. Exponential, Poisson and signed two-point weights give Proposed ($B = 10$) coverage of 0.946, 0.954 and 0.964. The signed law assigns probabilities 0.2 and 0.8 to -1 and 1.5, giving mean and variance one. All three laws therefore preserve the target innovation covariance in the first-order multiplier representation. Conditionally on the original stream, the multiplier sum $n^{-1/2}\sum_{k=1}^{n}(W_k - 1)\xi_k$ has mean zero and covariance $n^{-1}\sum_{k=1}^{n}\xi_k \otimes \xi_k$. Only the unit weight variance enters this covariance, whose convergence in probability to Q follows from Assumption 3.1. They need not produce identical finite-sample paths: Poisson weights can omit an update, while a negative weight reverses its direction. The comparable coverage illustrates similar behavior under these distinct perturbations, without showing superiority of signed weights. Their use still requires the weighted stability and localization conditions of the theorem. Implementation details are in the simulation appendix; the data appendix gives descriptive inference for 738 North Atlantic storm locations.

5 Concluding Remarks

Full-path multipliers provide online inference for Riemannian Polyak-Ruppert averages without plug-in derivative or covariance estimation. The proposed construction supports valid fixed-path Student intervals and intrinsic regions with few multiplier paths, while growing replication yields conditional quantiles. The random-base comparison resolves the difficulty of comparing tangent spaces and allows the same weighted trajectories to be used directly for manifold-valued inference. The framework also accommodates samplewise nonsmooth scores and identifies when accumulated approximation bias remains negligible or shifts the limiting law.

The guarantees require local stability and control of accumulated approximation bias in fixed dimension. Dependence-aware weights and bias-corrected oracles are natural extensions beyond the present martingale and negligible-bias conditions.

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 dedicated reproduction script for each figure and table, allowing the reported results to be reproduced individually. The `README` provides instructions for running the code.

REFERENCES

P.-A. Absil and J. Malick. Projection-like retractions on matrix manifolds. *SIAM Journal on Optimization*, 22(1):135–158, 2012.

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

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

M. Bačák. *Convex Analysis and Optimization in Hadamard Spaces*. De Gruyter, 2014.

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

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

S. Bonnabel. Stochastic gradient descent on Riemannian manifolds. *IEEE Transactions on Automatic Control*, 58(9):2217–2229, 2013.

G. Cheng, Q. Xiong, and R. Lin. Online bootstrap inference for the geometric median. *Computational Statistics and Data Analysis*, 197:107992, 2024. doi:10.1016/j.csda.2024.107992.

Y. Fang, J. Xu, and L. Yang. Online bootstrap confidence intervals for the stochastic gradient descent estimator. *Journal of Machine Learning Research*, 19(78):1–21, 2018.

P. T. Fletcher, S. Venkatasubramanian, and S. Joshi. The geometric median on Riemannian manifolds with application to robust atlas estimation. *NeuroImage*, 45(1 Suppl.):S143–S152, 2009.

J. Gahtan, K. R. Knapp, C. J. Schreck, H. J. Diamond, J. P. Kossin, and M. C. Kruk. International Best Track Archive for Climate Stewardship (IBTrACS) Project, version 4r01. NOAA National Centers for Environmental Information, doi:10.25921/82ty-9e16, 2024; accessed 21 August 2026.

P. Hall and C. C. Heyde. *Martingale Limit Theory and Its Application*. Academic Press, 1980.

T. Hotz, H. Le, and A. T. A. Wood. Central limit theorem for intrinsic Fréchet means in smooth compact Riemannian manifolds. *Probability Theory and Related Fields*, 189:1219–1246, 2024.

L. Jiang, A. Roy, K. Balasubramanian, D. Davis, D. Drusvyatskiy, and S. Na. Online covariance estimation in nonsmooth stochastic approximation. In *Proceedings of the 38th Conference on Learning Theory*, volume 291 of *Proceedings of Machine Learning Research*, pages 3079–3123, 2025.

Y. Jin, T. Xiao, and K. Balasubramanian. Statistical inference for Polyak-Ruppert averaged zeroth-order stochastic gradient algorithm. arXiv:2102.05198v3, 2021.

K. R. Knapp, M. C. Kruk, D. H. Levinson, H. J. Diamond, and C. J. Neumann. The International Best Track Archive for Climate Stewardship (IBTrACS): Unifying tropical cyclone data. *Bulletin of the American Meteorological Society*, 91(3):363–376, 2010.

S. Kumar, S. Pandey, and P. Sarkar. Beyond sin-squared error: Linear time entrywise uncertainty quantification for streaming PCA. In *Proceedings of the 41st Conference on Uncertainty in Artificial Intelligence*, volume 286 of *Proceedings of Machine Learning Research*, pages 2396–2430, 2025.

H. Lam and Z. Wang. Cheap bootstrap for fast uncertainty quantification of stochastic gradient descent. *Journal of Machine Learning Research*, 27(93):1–42, 2026.

C. J. Li and M. I. Jordan. Nonconvex stochastic scaled-gradient descent and generalized eigenvector problems. In *Proceedings of the Thirty-Ninth Conference on Uncertainty in Artificial Intelligence*, volume 216 of *Proceedings of Machine Learning Research*, pages 1230–1240, 2023.

R. Lunde, P. Sarkar, and R. Ward. Bootstrapping the error of Oja's algorithm. In *Advances in Neural Information Processing Systems 34*, pages 6240–6252, 2021.

X. Pennec, P. Fillard, and N. Ayache. A Riemannian framework for tensor computing. *International Journal of Computer Vision*, 66:41–66, 2006.

B. T. Polyak and A. B. Juditsky. Acceleration of stochastic approximation by averaging. *SIAM Journal on Control and Optimization*, 30(4):838–855, 1992.

D. Ruppert. Efficient estimations from a slowly convergent Robbins–Monro process. Technical Report 781, School of Operations Research and Industrial Engineering, Cornell University, 1988.

S. Samsonov, E. Moulines, Q.-M. Shao, Z.-S. Zhang, and A. Naumov. Gaussian approximation and multiplier bootstrap for Polyak-Ruppert averaged linear stochastic approximation with applications to TD learning. In *Advances in Neural Information Processing Systems 37*, 2024.

M. Sheshukova, S. Samsonov, D. Belomestny, E. Moulines, Q.-M. Shao, Z.-S. Zhang, and A. Naumov. Gaussian approximation and multiplier bootstrap for stochastic gradient descent. In *Proceedings of the 29th International Conference on Artificial Intelligence and Statistics*, volume 300 of *Proceedings of Machine Learning Research*, pages 1378–1386, 2026.

N. Tripuraneni, N. Flammarion, F. Bach, and M. I. Jordan. Averaging stochastic gradient descent on Riemannian manifolds. In *Proceedings of the 31st Conference on Learning Theory*, volume 75 of *Proceedings of Machine Learning Research*, pages 650–687, 2018.

C. Wang, L. Wang, X. Wang, D. Wang, and L. Wu. North–south variations of tropical storm genesis locations in the Western Hemisphere. *Geophysical Research Letters*, 43(21):11367–11374, 2016. doi:10.1002/2016GL071440.

H. Xu and H. Li. Inference and uncertainty quantification for streaming r-PCA. arXiv:2608.18374v1, 2026.

L. Yang. Riemannian median and its estimation. *LMS Journal of Computation and Mathematics*, 13:461–479, 2010.

Y. Zhong, T. Kuffner, and S. N. Lahiri. Online bootstrap inference with nonconvex stochastic gradient descent estimator. arXiv:2306.02205, 2023.

W. Zhu, X. Chen, and W. B. Wu. Online covariance matrix estimation in stochastic gradient descent. *Journal of the American Statistical Association*, 118(541):393–404, 2023.

A Simulation Details and Additional Results

This appendix specifies the experiments in the main text and reports additional comparisons for spherical means. Coverage uses all 500 streams in each configuration; estimation error is the median geodesic distance of the localized spherical estimator from its population target.

A.1 Protocol and implementation

All iterations use $\alpha = 0.7$. The step constants are $\eta_0 = 0.6$ on the original sphere designs, 0.8 and 1.0 on the symmetric and skew localized sphere designs, 0.3 for positive-definite means, 0.2 for positive-definite medians, and 0.1 on Grassmann manifolds. The proposed fixed-path intervals are evaluated with $B \in \{1, 3, 5, 10\}$ replicas, using exponential weights and the Student calibration in (7). We additionally compare Proposed (quantile) with $B = 200$, a tangent plug-in sandwich interval, and an offline trajectory bootstrap with 200 resampled trajectories. Methods compared within a configuration receive the same original stream and base fit; the two 200-replicate procedures use equal replication budgets. The offline comparator uses basic quantiles, not the fixed-replication Student calibration of COnB (Lam and Wang, 2026). Full regions are formed only when B is at least the tangent dimension.

The sphere designs use cap-supported von Mises–Fisher laws and rotational mixtures in dimensions $q \in \{2, 5, 10\}$ and the scalar functional $\phi(x) = \mathbf{e}_1^{\mathsf{T}} x = x_1$, whose target is zero at the north pole. Positive-definite observations have the form $\mathrm{Y} = \mathrm{A}^{1/2} \exp(\mathrm{E}) \mathrm{A}^{1/2}$ for $m \in \{2, 3, 5\}$, with sign-symmetric Gaussian or contaminated E. The target A has a simple largest eigenvalue, so its largest-log-eigenvalue functional satisfies (64). Grassmann designs use $(r, p) \in \{(1, 10), (2, 20), (3, 30)\}$ and Gaussian observations with eigengap 0.6. Their scalar target is the smooth projector entry $(\mathrm{UU}^{\mathsf{T}})_{r,r+1}$. The cross-geometry comparison uses $n \in \{5000, 20000\}$.

The iterations average all post-update states from $k = 1$ without burn-in. In the two localized spherical families, independent pilots selected a common step constant for every method by minimizing pooled mean squared geodesic error of the base average, not by optimizing coverage. The symmetric pilot used 40 streams per (n, η_0) cell and $\eta_0 \in \{0.4, 0.6, 0.8\}$; the skew pilot used 50 and $\eta_0 \in \{0.4, 0.6, 0.8, 1.0\}$. The selected constants were fixed before the 500-stream evaluations, whose seeds are disjoint from the pilots. All reported coverage denominators include all 500 streams. Coverage differences are interpreted together with their paired Monte Carlo variability; small numerical differences do not establish an ordering of methods.

A.2 Design details

Sphere designs. Write $\mathbf{e}_{q+1}$ for the north pole. In a cap design, $T = \langle Y, \mathbf{e}_{q+1} \rangle$ has density proportional to

$$\exp(\kappa t)(1 - t^2)^{(q-2)/2}, \qquad \cos(r_{\mathrm{cap}}) \le t \le 1,$$

and, conditional on T, the direction in $\mathbf{e}_{q+1}^{\perp}$ is uniform. A rotational mixture draws two components with equal probability after rotating them by ± 0.25 radians in the first–last coordinate plane; the component cap radius is $r_{\mathrm{cap}} - 0.25$. The primary $(n, \text{design}, \kappa, r_{\mathrm{cap}})$ settings are $(5000, \text{cap}, 8, 1.45)$ and $(20000, \text{mixture}, 6, 1.10)$ for $q = 2$; $(5000, \text{cap}, 8, 1.20)$ and $(20000, \text{mixture}, 6, 1.10)$ for $q = 5$; and $(5000, \text{mixture}, 8, 1.20)$ and $(20000, \text{cap}, 6, 1.10)$ for $q = 10$. The representative $q = 2$, $n = 1000$ cell in Table 2 is a cap design with $(\kappa, r_{\mathrm{cap}}) = (4, 1.20)$; the weight comparison uses $(q, n, \kappa, r_{\mathrm{cap}}) = (2, 5000, 6, 1.20)$. Reported cells start from $\mathrm{Exp}_{\mathbf{e}_{q+1}}(0.45\mathbf{e}_1)$, use the exponential map and logarithm, and rescale steps above 0.75.

Localized small-sample sphere designs. Both laws take $Y = \mathrm{Exp}_{\mathbf{e}_3}(v)$ for a two-dimensional tangent vector v. In the symmetric design, $v \sim N(0, 0.3^2 \mathrm{I}_2)$ conditional on $\|v\| \le 0.70$. In the skew design, its support is

$$(0.55, 0.05), \quad (-0.35, 0.45), \quad (-0.45, -0.25), \quad (-17/60, -0.55),$$

with probabilities $0.40, 0.25, 0.20, 0.15$, respectively. The support radii are 0.70 and approximately 0.619 radians. Both tangent means are zero; together with strong convexity on the supporting cap, this identifies the same intrinsic target without imposing symmetry on the second law.

The symmetric configurations use $n = 250, 500$, start at $\mathrm{Exp}_{\mathbf{e}_3}(0.30\mathbf{e}_1)$, take $\eta_0 = 0.8$, and rescale proposed tangent steps above 0.60. The skew configurations use $n = 100, 250$, start at $\mathrm{Exp}_{\mathbf{e}_3}(0.25\mathbf{e}_1)$, take $\eta_0 = 1.0$, and rescale steps above 0.55. In both designs, the observations lie in a cap of radius below $\pi/4$. Applying Corollary C.11 additionally requires the original and weighted states to remain in such a common cap, converge to the target, and satisfy the stated rates; tangent-step rescaling alone does not certify these path conditions. A single 200-path multiplier run supplies the nested fixed-B and quantile intervals, and each configuration also uses 200 offline trajectory resamples.

The spherical score is $\Psi(x; Y) = -\operatorname{Log}_x(Y)$. Each path moves toward the incoming observation along a geodesic, with the multiplier changing its step before tangent-step rescaling. Its iterative average also uses the exponential map and logarithm, rather than normalizing an arithmetic average of the observations.

Interval lengths. Table 4 reports median lengths of the 95% intervals in the localized and wider-cap comparisons. Length is the upper endpoint minus the lower endpoint, in units of $\phi(x) = x_1$, without truncation to its range $[-1, 1]$. Each median uses all 500 streams. The long $B = 1$ intervals reflect the Student critical value with one degree of freedom; using three to ten paths substantially reduces this cost.

Table 4: Median 95% interval lengths for the spherical coverage comparisons. Proposed (quantile) and offline methods each use 200 replicates; all methods share the base estimator within a configuration.

	Symmetric		Skew		Wider cap
Method	$n = 250$	$n = 500$	$n = 100$	$n = 250$	$n = 1000$
Proposed ($B = 1$)	0.3280	0.2316	0.8337	0.4994	0.2814
Proposed ($B = 3$)	0.1075	0.0720	0.2782	0.1775	0.0963
Proposed ($B = 5$)	0.0927	0.0634	0.2384	0.1515	0.0818
Proposed ($B = 10$)	0.0841	0.0573	0.2217	0.1370	0.0737
Proposed (quantile)	0.0764	0.0526	0.1996	0.1256	0.0662
Sandwich	0.0684	0.0484	0.1832	0.1162	0.0609
Offline	0.0740	0.0516	0.1966	0.1232	0.0652

Representative spherical comparators. At $\bar{x}_n$, let the columns of E be an orthonormal basis of $T_{\bar{x}_n}\mathbb{S}^q$, set $z_i = \mathrm{E}^{\mathsf{T}} \operatorname{Log}_{\bar{x}_n}(Y_i)$ and $r_i = \|z_i\|$, and put

$$\widehat{\mathrm{H}} = \frac{1}{n}\sum_{i=1}^{n}\left[\nu_i\nu_i^{\mathsf{T}} + \frac{r_i}{\tan r_i}\{\mathrm{I} - \nu_i\nu_i^{\mathsf{T}}\}\right], \qquad \widehat{\mathrm{Q}} = \frac{1}{n}\sum_{i=1}^{n}(z_i - \bar{z})(z_i - \bar{z})^{\mathsf{T}},$$

where $\bar{z} = n^{-1}\sum_{i=1}^{n} z_i$, $\nu_i = z_i/r_i$ when $r_i > 0$ and the corresponding Hessian summand is I when $r_i = 0$. The plug-in sandwich is $\widehat{\mathrm{V}} = \widehat{\mathrm{H}}^{-1}\widehat{\mathrm{Q}}(\widehat{\mathrm{H}}^{-1})^{\mathsf{T}}$, and its interval for $\phi(x) = x_1$ uses the standard error $\{(\mathrm{E}^{\mathsf{T}}\mathbf{e}_1)^{\mathsf{T}}\widehat{\mathrm{V}}(\mathrm{E}^{\mathsf{T}}\mathbf{e}_1)/n\}^{1/2}$ and the standard-normal critical value. For each of the 200 offline trajectory replicates, n indices are drawn independently with replacement, the base iteration and its full iterative average are replayed on the resulting ordered sample with the same initialization, step schedule, retraction, and step-rescaling rule, and the first coordinate is recorded. The reported offline interval is the basic interval formed from the type-8 empirical 0.025 and 0.975 quantiles of those replicate-minus-original differences.

Positive-definite designs. For $m \in \{2, 3, 5\}$, let

$$\mathrm{A} = \mathrm{diag}\{\exp(\ell_1), \ldots, \exp(\ell_m)\}, \qquad \ell_j = -0.3 + \frac{0.6(j-1)}{m-1}, \qquad j = 1, \ldots, m,$$

and set $\mathrm{Y} = \mathrm{A}^{1/2}\exp(\mathrm{E})\mathrm{A}^{1/2}$. The independent diagonal entries of the symmetric Gaussian matrix E have variance 0.25^2, and its independent upper-triangular entries have variance $0.25^2/2$. In the robustness cells, with probability $\varepsilon \in \{0, 0.05, 0.15\}$ we add

$$S\,1.2\ \mathrm{diag}(c_1, \ldots, c_m), \qquad \mathbb{P}(S = 1) = \mathbb{P}(S = -1) = 1/2,$$

to E, where $c_j = -1 + \frac{2(j-1)}{m-1}, \qquad j = 1, \ldots, m$. Primary mean cells use $n \in \{5000, 20000\}$ for each m; the mean–median comparison uses $m = 3$, $n = 5000$. All paths start at I_m, use the affine-invariant exponential and logarithm, and rescale steps above 0.8.

Grassmann designs. For $(r, p) \in \{(1, 10), (2, 20), (3, 30)\}$, the target is the span of the first r coordinate vectors. Gaussian observations have independent coordinates with covariance eigenvalues. Specifically, for $r = 2$, set $\lambda_1 = 2.15$; for $r = 3$, set $(\lambda_1, \lambda_2) = (2.30, 2.00)$. In all cases, $\lambda_r = 1 + \delta, \qquad \lambda_{r+1:p} = 1$. The regular cells use $\delta = 0.6$ and $n \in \{5000, 20000\}$. Initialization rotates the last active target vector toward the first inactive coordinate by 0.10 radians. Updates use the QR retraction and graph inverse in Lemma C.22, with step rescaling above 0.8.

B North Atlantic Tropical-Cyclone Locations

B.1 Data, estimand, and preprocessing

The International Best Track Archive for Climate Stewardship (IBTrACS) reconciles tropical-cyclone best-track records from multiple agencies (Knapp et al., 2010). We use the publicly accessible official NOAA/NCEI North Atlantic subset, version 4r01, downloaded on 21 August 2026 (Gahtan et al., 2024).

The spatial location at which a storm first reaches a tropical-storm threshold is itself a scientifically studied summary: for example, Wang et al. (2016) use such locations to study meridional variability across the North Atlantic and eastern North Pacific. Our selection below is deliberately more conservative in its labeling. It is defined by the first qualifying IBTrACS archive category and includes subtropical records, so we call it a first recorded tropical/subtropical location rather than physical genesis.

We first restrict to complete seasons 1980–2025, North Atlantic main-track records with finite coordinates, and archive times at which the combined `NATURE` field classifies a storm as tropical or subtropical (`TS` or `SS`). Within each storm identifier we then retain the earliest qualifying archive time, and finally order the storm-level observations chronologically. The retained point is therefore the first *recorded* tropical or subtropical location in the archive, not a physical genesis location. The resulting dataset contains 738 storms, of which 89.3% first enter the retained set as tropical.

Locations are represented as unit vectors on $\mathbb{S}^2$, and the estimand is the intrinsic center of the pooled storm-level distribution. Every retained point is less than $\pi/2$ from the reported center. This describes the observed spatial configuration, but does not certify common-chart membership of the original and weighted trajectories or their full averages. We report descriptive intervals under a working-independence model for the chronological archive; robustness to serial dependence or distributional change is not asserted.

B.2 Online analysis and uncertainty

We process storms chronologically from the north pole using the spherical exponential map, $\eta_k = 0.6k^{-0.7}$, and exponential weights. Proposed tangent steps are rescaled only when their norm exceeds 0.75. The first ten weighted paths give fixed-$B = 10$ intervals and all 200 paths give large-B basic intervals. Longitude is unwrapped locally around the estimate, so longitude and latitude are smooth scalar coordinates on the observed cap. The reported intervals are marginal for each coordinate, not a simultaneous two-coordinate confidence region.

Table 5: North Atlantic first recorded tropical/subtropical locations, 1980–2025. Coordinates and interval endpoints are in degrees; west longitudes are negative. All displayed intervals are descriptive working-independence multiplier intervals.

Quantity	Estimate	95% fixed $B = 10$	95% large $B = 200$
Longitude	-59.88	$[-61.77, -57.98]$	$[-61.57, -58.37]$
Latitude	21.92	$[21.06, 22.78]$	$[21.11, 22.49]$

The chronological online center is $59.88°$W, $21.92°$N. The fixed-$B = 10$ interval places its longitude between $61.77°$W and $57.98°$W and its latitude between $21.06°$N and $22.78°$N. The large-$B = 200$ intervals are similar, locating the pooled center in the tropical western Atlantic with closely agreeing uncertainty summaries.

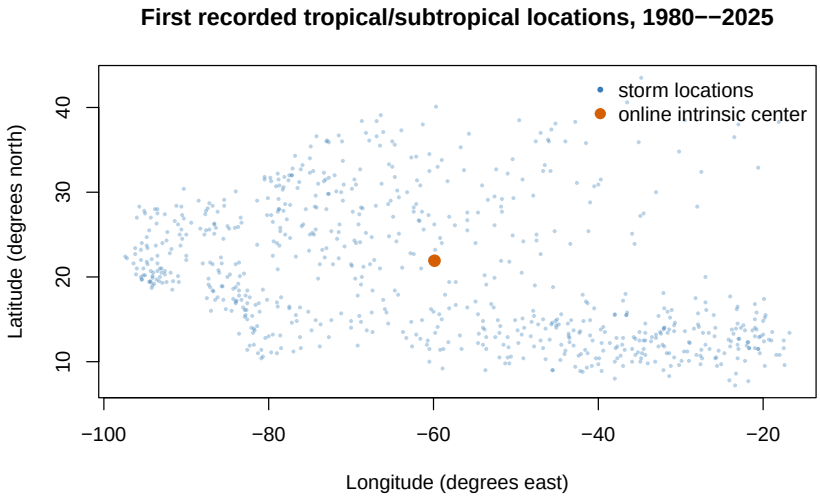

Figure 2: Retained storm-level locations and the chronological online intrinsic center. The plotted points are first archive records satisfying the tropical/subtropical filter, not physical genesis estimates.

C Theory Supplement

C.1 Calibration and implementation details

The theorem permits signed weights. Examples satisfying (3) are $W \sim \mathrm{Exp}(1)$, $W \sim \mathrm{Poisson}(1)$, $W = 1+\epsilon$ with $\epsilon \in \{-1, 1\}$ equiprobably, and $W = 1+G$ with $G \sim N(0,1)$. Nonnegative weights may reduce finite-sample path instability, but nonnegativity is not needed for first-order validity once the weighted state bound is verified.

C.1.1 Two calibration regimes

Fix $0 < \tau < 1$ as the confidence error probability. For tangent vectors, $(a \otimes b)v = \langle b, v\rangle\, a$ denotes the rank-one operator. The completed vectors are defined in (21). On $\{\bar{x}_n \in U_\star\}$, which depends only on the original stream, the same isometry preserves inner products for every replicate. The fixed-space representation is used only for analysis; the algorithm does not use $x_\star$.

Fixed-replication calibration. Let $B \geq 1$ be fixed and let $\bar{x}_n^{\star(1)}, \ldots, \bar{x}_n^{\star(B)}$ be conditionally independent multiplier averages. For a scalar functional $\phi : \mathcal{M} \to \mathbb{R}$, use (6)–(7), where $t_{B,1-\tau/2}$ is the Student-t quantile with B degrees of freedom. The denominator is centered at the observed original estimate, so the degrees of freedom are B, not $B-1$. If $s_{\phi,n,B} = 0$, define $I_{\phi,n,B}(1-\tau) = \mathbb{R}$; under the positive limiting-variance condition in Corollary 3.7, this fallback event has probability tending to zero.

For a full intrinsic region, let $\widetilde{T}^\star_{n,b}$ be the completed statistic in (21) for path b and put

$$\widehat{S}_{n,B} = \frac{1}{B}\sum_{b=1}^{B} \widetilde{T}^\star_{n,b} \otimes \widetilde{T}^\star_{n,b}.$$

When $B \geq d$ and $\widehat{S}_{n,B}$ is invertible, restrict attention to the retractive neighborhood of $\bar{x}_n$ on which $R^{-1}_{\bar{x}_n}$ is single valued and define

$$\mathcal{C}^{\mathrm{fix}}_{n,B}(1-\tau) = \left\{ x :\ n \left\langle R^{-1}_{\bar{x}_n}(x), \widehat{S}^{-1}_{n,B} R^{-1}_{\bar{x}_n}(x) \right\rangle_{\bar{x}_n} \leq c_{d,B,\tau} \right\}, \tag{33}$$

with

$$c_{d,B,\tau} = \frac{dB}{B-d+1} F_{d,B-d+1;1-\tau}, \tag{34}$$

where $F_{a,b;\gamma}$ denotes the γ quantile of the F distribution with (a, b) degrees of freedom. For $d = 1$, the critical-value identity $F_{1,B;1-\tau} = t^2_{B,1-\tau/2}$ makes (33) coincide with (7) for the Euclidean identity functional, or for the one-dimensional local coordinate used to define the intrinsic region. For a nonlinear scalar functional, the two procedures instead have the same first-order pivotal calibration through the delta method. If $\widehat{S}_{n,B}$ is singular, define $\mathcal{C}^{\mathrm{fix}}_{n,B}(1-\tau) = \mathcal{M}$. Positive-definite V and $B \geq d$ make this fallback event asymptotically negligible. The borderline choice $B = d$ is valid but has one denominator degree of freedom and is consequently not a practical default.

Large-replication calibration. When $B = B_n \to \infty$, empirical bootstrap quantiles approximate the full conditional law. Let $q^\star_\gamma$ be the empirical γ quantile of

$$\sqrt{n}\{\phi(\bar{x}_n^{\star(b)}) - \phi(\bar{x}_n)\}, \qquad b = 1, \ldots, B_n.$$

The basic interval is

$$\left[\phi(\bar{x}_n) - \frac{q^\star_{1-\tau/2}}{\sqrt{n}},\quad \phi(\bar{x}_n) - \frac{q^\star_{\tau/2}}{\sqrt{n}} \right]. \tag{35}$$

If $c^\star_{1-\tau}$ is the empirical $(1-\tau)$ quantile of $\left\| \widetilde{T}^\star_{n,b} \right\|_{\bar{x}_n}$, the norm region is

$$\mathcal{C}_{n,1-\tau} = \left\{ R_{\bar{x}_n}(v/\sqrt{n}) : v \in T_{\bar{x}_n}\mathcal{M},\ \|v\|_{\bar{x}_n} \leq c^\star_{1-\tau} \right\}. \tag{36}$$

The set in (36) includes only vectors for which $v/\sqrt{n}$ lies in the domain of $R_{\bar{x}_n}$. A large-B bootstrap covariance estimator is

$$\widehat{V}^{\star}_{n,B} = \frac{1}{B}\sum_{b=1}^{B}(\widetilde{T}^{\star}_{n,b} - \overline{T}^{\star}_{n,B}) \otimes (\widetilde{T}^{\star}_{n,b} - \overline{T}^{\star}_{n,B}), \qquad \overline{T}^{\star}_{n,B} = \frac{1}{B}\sum_{b=1}^{B}\widetilde{T}^{\star}_{n,b}. \tag{37}$$

If the estimator in (37) is invertible, restrict the corresponding Wald region to a retractive neighborhood $U_{\bar{x}_n}$ on which the inverse is single-valued:

$$\left\{ x \in U_{\bar{x}_n} : n \left\langle R^{-1}_{\bar{x}_n}(x), (\widehat{V}^{\star}_{n,B})^{-1} R^{-1}_{\bar{x}_n}(x) \right\rangle_{\bar{x}_n} \le \chi^2_{d,1-\tau} \right\}. \tag{38}$$

Because the B centered vectors in (37) have rank at most $B-1$, invertibility requires $B \ge d+1$. If this rank condition fails or $\widehat{V}^{\star}_{n,B}$ is singular, define the Wald region to be $\mathcal{M}$. Under the positive-definite limit in Proposition C.4 and $B_n \to \infty$, this fallback event has probability tending to zero. Fixed-B calibration and large-B conditional-law approximation are different uses of the same paths. The former gives asymptotic coverage with very few paths but uses Gaussian pivotality; the latter estimates the conditional shape of the bootstrap law and requires Monte Carlo growth.

C.1.2 Implementation details

All inverse maps used by the iterative averages must be defined along the implemented paths; the reporting completion does not define a missing earlier update. When a safeguard is used, it is part of each state update. For a state representation using $O(d)$ numbers, the additional path storage is $O(Bd)$ and every observation requires B extra oracle evaluations and retractions. A Grassmann QR representation instead stores $p \times r$ matrices, and forming a full reporting covariance requires $O(d^2)$ additional storage. The data point and base seed can be discarded after all paths have been updated. An offline replay using the same ordered data, seeds, weights, localization rule, and arithmetic produces the same paths; the method does not rely on resampling stored observations. For a growing replication count B_n, all paths for a chosen horizon are initialized before processing its stream; adding a path after discarding earlier data does not reproduce a full-history path.

C.1.3 Fixed-coordinate notation and probability conventions

Fix a relatively compact retractive neighborhood U of $x_\star$ and write

$$x(a) = R_{x_\star}(a), \qquad u_k = R^{-1}_{x_\star}(x_k), \qquad u^{\star}_k = R^{-1}_{x_\star}(x^{\star}_k).$$

All unlabelled norms and inner products are those of $T_{x_\star}\mathcal{M}$. Let

$$\mathcal{F}_k = \sigma(D_1, \dots, D_k), \qquad \mathcal{G}_k = \mathcal{F}_k \vee \sigma(W_1, \dots, W_k).$$

Write $\mathcal{F}_\infty = \sigma(\bigcup_{k\ge 1}\mathcal{F}_k)$; the starred probability and expectation operators are defined in the main text.

For random variables $Y^{\star}_n$ depending on both streams, $Y^{\star}_n = o_{\mathbb{P}^\star}(1)$ in probability means that, for every $\varepsilon > 0$,

$$\mathbb{P}\{\mathbb{P}^\star(\|Y^{\star}_n\| > \varepsilon) > \varepsilon\} \longrightarrow 0.$$

For probability measures on the finite-dimensional space $T_{x_\star}\mathcal{M}$, d_{BL} denotes the bounded-Lipschitz distance, and $N(0, V)$ denotes the centered Gaussian law with covariance V. Matrix subscripts F indicate the Frobenius inner product or norm; vec stacks matrix columns and I is the identity of the relevant space. For a linear operator A, $A^\dagger$ denotes its adjoint. Upright symbols denote matrices; tangent vectors and basis-free linear operators retain their intrinsic notation.

C.2 Additional discussions

Moment conditions Equations (8)–(9) are the covariance and conditional Lindeberg assumptions of the martingale central limit theorem (Hall and Heyde, 1980). A uniform $2+\delta$ moment is sufficient for (9) but not necessary, while (8) must be verified separately. Likewise, (10) is enough for the conditional multiplier central limit theorem. Higher multiplier moments or nonnegativity are stability conveniences rather than requirements for weak convergence.

Approximation bias. The full oracle is multiplied by W_k, so a predictable approximation bias appears as $W_k b_k$. The decomposition

$$\frac{1}{\sqrt{n}}\sum_{k=1}^{n} W_k b_k = \frac{1}{\sqrt{n}}\sum_{k=1}^{n} b_k + \frac{1}{\sqrt{n}}\sum_{k=1}^{n}(W_k - 1)b_k$$

explains the two conditions in (13). The vector-sum condition allows directional cancellation; the conditional variance of the second term is $n^{-1}\sum \|b_k\|^2$.

Localization. A safeguard, when used, is part of the implemented state iteration. Localization does not restart or replace its reported full iterative averages. Eventual inactivity means that the ordinary update is applied at the current safeguarded state; it does not identify this path with a separately run unsafeguarded path having a different history. The assumption above concerns both the states and their full iterative averages: state convergence alone does not imply that the average enters the comparison chart. Lemma C.9 gives a sufficient route for the latter conclusion when the averaging iteration has an invariant contracting basin. For each path, let τ denote the last index at which its safeguard is active or its local proposal formula is unavailable; τ need not be a stopping time. When $\tau < \infty$ almost surely, the finitely many exact coordinate discrepancies have an almost surely finite sum and therefore contribute $o_p(1)$ after the $n^{-1/2}$ normalization in (15). The same statement holds under the joint law for a weighted path and hence conditionally in probability. Because every weighted trajectory uses its own path-specific finite-exception index and its full iterative average is retained, the reported trajectories remain conditionally independent and identically distributed. No maximum of entry times over B_n paths is required for empirical inference from the completed statistic in (21): for each n, the multiplier paths form an i.i.d. row from its conditional law. In contrast, an implementation that aborts a whole row whenever any inverse is unavailable would require separate rate control and is not used here. Lemma C.19 records the corresponding invariance of the intrinsic average.

Unweighted expansion. Under smooth RSGD conditions, the unweighted representation is a Polyak-Ruppert benchmark (Tripuraneni et al., 2018). It is retained because the weighted result must be proved on the same probability space and under the abstract averaged conditions above. The full-path extension is nontrivial: the weighted oracle is evaluated at a weighted state, yet only the target innovation is multiplied in the first-order representation.

Geometric centering. Equation (24) is not obtained by formally subtracting manifold-valued points. The proof first expands both averages in $T_{x_\star}\mathcal{M}$ and then shows that the implemented vector $R_{\bar{x}_n}^{-1}(\bar{x}_n^\star)$ is first-order equivalent to their coordinate difference. This is the geometric step absent in Euclidean COnB.

Proposition C.1 (Exact Euclidean reduction to COnB)**.** *Let $\mathcal{M} = \mathbb{R}^d$, $R_x(v) = x + v$, use no state safeguard and the arithmetic Polyak-Ruppert average, let $\Psi_k(x; D_k) = \nabla h(x, Z_k)$ with i.i.d. data, and take independent* $\mathrm{Exp}(1)$ *multipliers. Up to an immaterial indexing convention, Algorithm 1 and, for the coordinate functional $\phi_i(x) = x_i$, the interval* (7) *are exactly Algorithm 2 and the ASGD confidence interval of Lam and Wang (2026). Consequently, Theorem 3.6 and Corollary 3.7 recover their fixed-B online coordinatewise result whenever both sets of assumptions hold.*

Theoretical comparison. The overlap in Proposition C.1 is exact at the algorithmic level, while the assumption regimes are nonnested. Lam and Wang (2026) assume a twice continuously differentiable sample loss, global samplewise strong convexity and bounded Hessian, a sample Hessian that is Lipschitz on a bounded target neighborhood uniformly over observations, a fourth moment for the target gradient, and Glivenko–Cantelli conditions; they allow $\alpha = 1$, establish an offline cheap bootstrap and last-iterate results, derive quantitative normal-approximation bounds, and study a sparse high-dimensional extension. The present theorem is local and fixed-dimensional for averaged iterations with $\alpha < 1$, and extends the construction to nonsymmetric derivatives, martingale innovations, samplewise nonsmooth fields controlled in L^2, finite-second-moment multipliers, and manifold-valued parameters.

C.2.1 LARGE-REPLICATION INFERENCE

When the goal is to estimate the conditional distribution itself, the number of paths must grow. Weak convergence is enough for continuous quantiles; estimating a covariance asks for the additional tail control stated after the next corollary.

Corollary C.2 (Smooth functionals and intrinsic norm regions)**.** *Let $\phi : \mathcal{M} \to \mathbb{R}^p$ be differentiable at $x_\star$. Under Theorem 3.5, conditionally in probability,*

$$\sqrt{n}\{\phi(\bar{x}_n^\star) - \phi(\bar{x}_n)\} \rightsquigarrow N_p\{0, D\phi_{x_\star} V D\phi_{x_\star}^\dagger\}, \tag{39}$$

and its conditional law consistently estimates the sampling law of $\sqrt{n}\{\phi(\bar{x}_n) - \phi(x_\star)\}$. If the limiting scalar variance is positive and $B = B_n \to \infty$, (35) *has asymptotic coverage $1-\tau$. If the distribution of $\|Z\|$, $Z \sim N(0, V)$, is continuous at its $(1-\tau)$ quantile, then* (36) *has asymptotic coverage $1-\tau$.*

Definition C.3 (Conditional uniform integrability in probability)**.** A sequence $(Y_n^\star)$ is conditionally uniformly integrable in probability if, for every $\varepsilon > 0$,

$$\lim_{K\to\infty} \limsup_{n\to\infty} \mathbb{P}\left\{\mathbb{E}^\star\left[|Y_n^\star| \mathbf{1}\{|Y_n^\star| > K\}\right] > \varepsilon\right\} = 0. \tag{40}$$

Proposition C.4 (Bootstrap covariance and Wald regions)**.** *Suppose Theorem 3.5 holds, $B = B_n \to \infty$, and $\|T_n^\star\|^2$ is conditionally uniformly integrable in probability. Then*

$$\mathcal{P}_{\bar{x}_n \to x_\star} \widehat{V}_{n,B_n}^\star \mathcal{P}_{x_\star \to \bar{x}_n} \xrightarrow{p} V. \tag{41}$$

If V is positive definite, the Wald region (38) *has asymptotic coverage $1-\tau$.*

Additional moment conditions for covariance estimation. Conditional uniform integrability of $\|T_n^\star\|^2$ is needed only to pass from weak convergence to covariance convergence; it is not implied by (10) and the weak bootstrap theorem. A transparent sufficient condition is that, for some $\delta > 0$, $\mathbb{E}^\star \|T_n^\star\|^{2+\delta} = O_p(1)$. Conditional Markov inequality then gives (40). This bound is an additional sufficient condition, not a consequence of the second-moment state bounds in Proposition C.8. A fourth moment is not required by the general covariance proposition, although a particular quadratic oracle, such as streaming PCA, may itself require a fourth data moment to have a finite innovation variance.

C.2.2 PRIMITIVE CONDITIONS FOR RIEMANNIAN STOCHASTIC FIELDS

The abstract iteration is useful only if its remainders can be verified from an oracle and a manifold update. We now give one such route. It replaces samplewise derivatives by a centered local L^2 modulus, isolates population smoothness in a deterministic remainder, and uses a separate stability argument for the state rate.

For u in a ball $B_{T_{x_\star}\mathcal{M}}(0, r)$, define the transported oracle

$$\psi_k(u) = \mathcal{P}_{x(u)\to x_\star} \Psi_k\{x(u); D_k\} \in T_{x_\star}\mathcal{M}.$$

Assumption C.5 (Primitive oracle expansion)**.** For $u \in B_{T_{x_\star}\mathcal{M}}(0, r)$,

$$\psi_k(u) = Hu + \xi_k + b_k + \chi_k(u) + r_{\mathrm{pop}}(u), \tag{42}$$

where H is nonsingular, (ξ_k) satisfies Assumption 3.1, b_k is $\mathcal{F}_{k-1}$-measurable and satisfies (13), and the following conditions hold.

(i) $\chi_k(0) = 0$. For every $\mathcal{G}_{k-1}$-measurable U_{k-1} taking values in $B_{T_{x_\star}\mathcal{M}}(0, r)$,

$$\mathbb{E}\{\chi_k(U_{k-1}) \mid \mathcal{G}_{k-1}\} = 0, \tag{43}$$

$$\mathbb{E}\{\|\chi_k(U_{k-1})\|^2 \mid \mathcal{G}_{k-1}\} \le \omega(\|U_{k-1}\|), \tag{44}$$

where ω is bounded on $[0, r]$ and $\omega(t) \to 0$ as $t \downarrow 0$. The analogous statements hold for $\mathcal{F}_{k-1}$-measurable arguments.

(ii) For some $\rho \in (0, 1]$,

$$\|r_{\mathrm{pop}}(u)\| \le C\|u\|^{1+\rho}. \tag{45}$$

(iii) The local oracle has a uniformly bounded conditional second moment:

$$\sup_{k\geq 1}\sup_{u\in B_{T_{x_\star}\mathcal{M}}(0,r)}\mathbb{E}\{\|\psi_k(u)\|^2 \mid \mathcal{F}_{k-1}\}\leq C \quad \text{almost surely.} \tag{46}$$

The current oracle input D_k is conditionally fresh given the past, so (43)–(46) remain valid after enlargement by past multipliers.

The conditional bounds are understood through common measurable conditional-kernel versions. At $u = 0$, (42) and (46) give $\mathbb{E}(\|\xi_k\|^2 \mid \mathcal{F}_{k-1}) + \|b_k\|^2 \leq C$, since $\psi_k(0) = \xi_k + b_k$ and ξ_k is centered. Independence of the weight sequence preserves this bound after enlargement to $\mathcal{G}_{k-1}$. Together with (44) and the bounded deterministic terms on the local ball, it follows that $\mathbb{E}\{\|\psi_k(U_{k-1})\|^2 \mid \mathcal{G}_{k-1}\} \leq C'$ for every predictable local argument U_{k-1}. Thus the second-moment bound applies at the actual local states without an additional fourth-moment assumption.

For i.i.d. data, let $\psi(u; D_k)$ be a measurable random field and $G(u) = \mathbb{E}\psi(u; D_k)$. Then

$$\xi_k = \psi(0; D_k) - G(0), \qquad \chi_k(u) = \psi(u; D_k) - \psi(0; D_k) - \{G(u) - G(0)\}$$

provides the centered decomposition. The condition on χ_k is then an ordinary local L^2 modulus and does not require a samplewise derivative.

Assumption C.6 (Second-order update compatibility)**.** For $u, z \in T_{x_\star}\mathcal{M}$ sufficiently small, define

$$\Phi(u, z) = R_{x_\star}^{-1}\left[R_{x(u)}\{\mathcal{P}_{x_\star\to x(u)}z\}\right]. \tag{47}$$

There is a constant C such that

$$\|\Phi(u, z) - u - z\| \leq C\{\|u\|^2\|z\| + \|z\|^2\}. \tag{48}$$

The exponential map and parallel transport satisfy Assumptions 3.3 and C.6 by normal-coordinate and Jacobi-field expansions. A general first-order retraction does not automatically give all three inequalities. For non-exponential retractions, the chosen retraction, inverse, and transport must be checked together; projection-like second-order retractions on matrix manifolds provide an important class (Absil and Malick, 2012).

Proposition C.7 (Primitive verification of the averaged iteration)**.** *Suppose Assumptions 3.3, C.5 and C.6 and the state bounds* (16) *hold, with eventual localization in a common ball where all three assumptions apply. If a safeguard is used, assume it is eventually inactive for each fixed path; Proposition C.8 supplies this conclusion. If*

$$\alpha(1 + \rho) > 1, \tag{49}$$

then the actual updates (1) *and* (4) *satisfy Assumption 3.2, with their finitely many exceptional coordinate discrepancies absorbed into e_k and $e_k^\star$. Consequently, Theorems 3.4 to 3.6 hold for the reported full iterative averages.*

The smoothness threshold. The condition (49) controls

$$\frac{1}{\sqrt{n}}\sum_{k=1}^{n}\mathbb{E}\,\|u_{k-1}\|^{1+\rho} \lesssim \frac{1}{\sqrt{n}}\sum_{k=1}^{n}\eta_k^{(1+\rho)/2}.$$

For a locally C^2 population field, $\rho = 1$, so no restriction is added beyond $\alpha > 1/2$. For $\rho < 1$, (49) is sufficient rather than necessary; a direct accumulated remainder bound can replace it.

Proposition C.8 (A second-moment route to the $O(\eta_k)$ state bounds)**.** *Let $K \subset \mathrm{int}(K^+)$ be compact neighborhoods of $x_\star$, with K geodesically convex and squared distance twice continuously differentiable on K^+. Initialize $x_0 \in K$ and all weighted paths at the same point. Suppose the update uses, when necessary, a measurable safeguard $S(x, v) \in K$ for an attempted tangent step $v \in T_x\mathcal{M}$. There is an $r_0 > 0$ such that, uniformly for $x \in K$, $R_x(-v)$ is defined and lies in K^+ whenever $\|v\|_x \leq r_0$, and*

$$d\{S(x, v), x_\star\} \leq d\{R_x(-v), x_\star\} \quad \textit{on short proposals.}$$

For arbitrary v, the safeguard may include tangent-step clipping before applying the retraction, but membership of its output in K is a separate requirement: clipping alone does not imply this

invariance. There are an open neighborhood U_0 of $x_\star$, with $\overline{U}_0 \subset K$, and a constant $\delta_0 > 0$ such that, uniformly for $x \in U_0$,

$$S(x, v) = R_x(-v) \qquad \textit{whenever } \|v\|_x \le \delta_0. \tag{50}$$

Let G be a deterministic mean field. The following conditional bounds use common measurable versions and hold at every predictable K-valued state under the original and enlarged filtrations. For some $\mu > 0$, suppose uniformly for $x \in K$ that

$$\langle G(x), \mathrm{Log}_x(x_\star)\rangle_x \le -\mu d(x, x_\star)^2, \tag{51}$$

$$\|\mathbb{E}\{\Psi_k(x; D_k) \mid \mathcal{F}_{k-1}\} - G(x)\|_x \le a_k, \qquad a_k^2 \le C\eta_k, \tag{52}$$

$$\mathbb{E}\{\|\Psi_k(x; D_k)\|_x^2 \mid \mathcal{F}_{k-1}\} \le C. \tag{53}$$

Assume the current oracle is conditionally fresh given the past, and the multipliers satisfy (10)*. For sufficiently small η_0, the localized paths satisfy*

$$\mathbb{E}d(x_k, x_\star)^2 \le C\eta_k, \qquad \mathbb{E}\mathbb{E}^\star d(x_k^\star, x_\star)^2 \le C\eta_k. \tag{54}$$

Moreover, the safeguard is applied only finitely often almost surely for each fixed path.

The preceding proposition controls the states. The next deterministic result supplies the missing implication for the unchanged full average whenever the chosen geometry admits an invariant averaging basin.

Lemma C.9 (Averaging-basin localization of the full average)**.** *Let $\alpha \in (0, 1)$ and let $K \subset \mathcal{M}$ contain $x_\star$. Suppose that, for every $x, y \in K$ and $t \in [0, 1]$, the inverse retraction is defined, the averaging map $\mathsf{A}_t(x, y) = R_x\{tR_x^{-1}(y)\}$ belongs to K, and, for constants $\kappa \in (\alpha, 1]$ and $C < \infty$,*

$$d\{\mathsf{A}_t(x, y), x_\star\}^2 \le (1 - \kappa t)d(x, x_\star)^2 + Ct\, d(y, x_\star)^2. \tag{55}$$

If a state path (x_k) lies in K, converges to $x_\star$, and its full iterative average satisfies $\bar{x}_1 = x_1$ and $\bar{x}_k = \mathsf{A}_{1/k}(\bar{x}_{k-1}, x_k)$, then $\bar{x}_k$ remains in K and converges to $x_\star$. If additionally $\mathbb{E}d(x_k, x_\star)^2 \le C_0 k^{-\alpha}$, then $\mathbb{E}d(\bar{x}_k, x_\star)^2 \le C_1 k^{-\alpha}$. The same conclusions hold under the joint original–multiplier law after replacing $\mathbb{E}$ by $\mathbb{E}\mathbb{E}^\star$.

This lemma is a sufficient primitive route to the average-localization clause of Assumption 3.3; the abstract theorem continues to allow that clause to be verified directly.

A concrete safeguard and averaging basin. On a Hadamard manifold, choose a specified compact geodesically convex set K containing $x_\star$ in its interior and a fixed $c \in K$. For a short proposal use $S(x, v) = \Pi_K\{R_x(-v)\}$, where Π_K is metric projection; for a long or unavailable proposal set $S(x, v) = c$. Projection does not increase distance to any point of K, so the output is invariant in K and the distance inequality in Proposition C.8 holds. Compactness supplies a uniform short-step neighborhood, and sufficiently small steps from a neighborhood of $x_\star$ already lie in K, giving (50). The rule evaluates neither $x_\star$ nor an entry time; $x_\star \in \mathrm{int}\, K$ is a model restriction. After verifying the drift and moment bounds on K, Proposition C.8 yields convergence and eventual inactivity. For exponential averaging, geodesic convexity of squared distance gives (55) with $\kappa = C = 1$. This completes the state and full-average verification without assuming either conclusion.

Why stability is stronger than nonsingularity. Strong monotonicity is not part of the abstract bootstrap theorem. It is one route to the state rate. It follows locally from a positive-definite Riemannian Hessian for geodesically convex objectives and from an eigengap for principal-subspace estimation. A model-specific Lyapunov function or contraction argument may replace it. The bound $a_k^2 \lesssim \eta_k$ is used only for state stability; inference still requires the accumulated-bias condition (13).

C.2.3 How the assumptions divide the work

The abstract theorem takes the state rate and complete-path localization as inputs. A primitive route first verifies the drift and moments for an actual invariant safeguard, then obtains state convergence from Proposition C.8 and full-average localization from Lemma C.9; only then does Proposition C.7 supply the averaged iteration. The following distinctions explain which assumptions serve inference

and which provide sufficient verification devices. Nonsingularity of H is structural: without a locally invertible derivative, the usual linear influence representation need not hold. Symmetry and positive definiteness are not required by the abstract Abel argument. They enter only through a chosen stability proof, as in strongly geodesically convex objectives or eigengap problems. This distinction is standard in stochastic approximation, where the averaging limit and the Lyapunov analysis need not use identical hypotheses (Polyak and Juditsky, 1992; Tripuraneni et al., 2018).

The step range $\alpha \in (1/2, 1)$ is conventional for averaged stochastic approximation. The lower bound makes $\sum \eta_k^2 < \infty$ and controls geometric one-step errors; the upper bound removes the Abel boundary and is used in the $O(\eta_k)$ comparison induction. The endpoint $\alpha = 1$ is possible in Euclidean and Riemannian models, but the admissible step constant then depends on the spectrum of H and requires a separate argument. Lam and Wang (2026), for example, include $\alpha = 1$ under their global smooth strongly convex assumptions. More general regularly varying steps can be used whenever the summation-by-parts bounds in Lemma C.17 remain valid.

The martingale covariance and Lindeberg conditions are close to the standard minimal assumptions for a finite-dimensional martingale central limit theorem (Hall and Heyde, 1980). A $2 + \delta$ score moment is a transparent sufficient condition for the Lindeberg condition, not a required high-order assumption. The same applies to the multipliers: mean one preserves the population drift, variance one calibrates Q, and a finite second moment is enough for the conditional Lindeberg argument. The fourth moment in the Grassmann application is different: there the stochastic gradient is quadratic in the data, so a fourth data moment supplies a sufficient square-integrable envelope for the score and its local increments.

The primitive $C^{1,\rho}$ condition is imposed on the population field, not on every sample loss. It can be replaced by the direct accumulated remainder in (15). The local L^2 modulus similarly permits nonsmooth sample paths; the geometric median application below is an example. Relative to the quantitative Euclidean analysis of Lam and Wang (2026), these conditions replace global samplewise Hessian control by local population and L^2 requirements, while adding explicit state, localization, and geometric hypotheses. The assumption sets are therefore non-nested, and the present conclusion is only first-order and fixed-dimensional.

The geometric inequalities (18), (20), and (48) are local second-order coordinate identities. They are automatic for exponential-map normal coordinates with parallel transport and can be checked for common second-order matrix retractions (Absil et al., 2008; Absil and Malick, 2012). They are not probabilistic smoothness assumptions. Localization is a substantive path condition, separate from these deterministic identities; Proposition C.8 shows eventual safeguard inactivity separately for each path, and Lemma C.19 covers the unchanged full averages without an online or common restart time.

Finally, the bias conditions are substantive. Pointwise convergence $b_k \to 0$ is insufficient because the Polyak-Ruppert expansion accumulates b_k before dividing by $\sqrt{n}$. Proposition 3.8 proves the fixed boundary statement under the abstract state conditions and records the subcritical Abel term only as a coordinate-level diagnostic pending a bias-aware manifold remainder analysis. Conditional uniform integrability is needed only for covariance estimation, not for fixed-B pivotal intervals or large-B quantile inference.

Bias accumulates before averaging. The multiplier perturbation of the bias has conditional variance $n^{-1} \sum \|b_k\|^2$, which vanishes for every polynomial rate $k^{-\gamma}$ with $\gamma > 0$. It therefore cannot mimic a fixed or diverging deterministic shift in the sampling law.

Let $h_k = h_0 k^{-\beta}$, with $h_0, r, \beta > 0$ as in Proposition 3.8. For a state-dependent approximation bias $h_k^r B(u)$, controlling only $b_k = h_k^r B(0)$ leaves an additional drift $h_k^r\{B(u_{k-1}) - B(0)\}$. If B is locally Lipschitz and the state rate holds, its root-n accumulated expectation is bounded by $Cn^{-1/2} \sum_{k \le n} h_k^r \eta_k^{1/2}$, which vanishes when $\alpha + 2r\beta > 1$. This is a sufficient check on the remaining drift, not a replacement for bias-aware state stability in a subcritical regime.

C.3 Non-Euclidean Applications

The examples below use i.i.d. observations and exponential-map notation for clarity. A non-exponential implementation may be substituted only after its retraction, local inverse, and transport jointly satisfy the three geometric comparisons. Lemma C.22 verifies the normalized spherical

projection and Grassmann QR/graph-inverse triples used below. In each case the displayed limiting law is also the conditional limit of the full-path multiplier difference, and both the fixed-B and large-B procedures apply under their respective moment and nondegeneracy conditions.

C.3.1 Intrinsic means on a sphere

Let $\mathcal{M} = \mathbb{S}^q$ with its canonical metric and $F(x) = \frac{1}{2}\mathbb{E}d(x, Y)^2$. Assume that, for some $x_\star \in \mathbb{S}^q$ and $r_0 < \pi/2$,

$$\mathbb{P}\{d(x_\star, Y) \le r_0\} = 1, \qquad \mathbb{E}\,\mathrm{Log}_{x_\star}(Y) = 0. \tag{56}$$

Let $v = \mathrm{Log}_{x_\star}(Y)$, $r = \|v\|$, and $e = v/r$ when $r > 0$. With the entire Hessian integrand defined to be I at $r = 0$, define

$$H_{\mathbb{S}} = \mathbb{E}\left[e \otimes e + r\cot(r)\{I - e \otimes e\}\right], \tag{57}$$

$$Q_{\mathbb{S}} = \mathbb{E}(v \otimes v). \tag{58}$$

The support condition excludes the cut locus and makes every eigenvalue in the integrand of (57) positive. It is a transparent sufficient condition for a regular local mean, not the most general condition for a spherical Fréchet-mean central limit theorem; cut-locus configurations can require additional terms (Hotz et al., 2024).

Corollary C.10 (Spherical intrinsic mean)**.** *Suppose* (56) *holds, the original and weighted states and full iterative averages satisfy the localization clause of Assumption 3.3 in a common strongly convex neighborhood of $x_\star$, and their state rates satisfy* (16)*. Use the stochastic gradient $\Psi(x; Y) = -\mathrm{Log}_x(Y)$. Then*

$$\sqrt{n}\,\mathrm{Log}_{x_\star}(\bar{x}_n) \rightsquigarrow N\left(0, H_{\mathbb{S}}^{-1} Q_{\mathbb{S}} H_{\mathbb{S}}^{-1}\right), \tag{59}$$

and the multiplier law is conditionally consistent. The fixed-B scalar intervals in Corollary 3.7 are valid for differentiable spherical functionals with nonzero limiting variance; the full intrinsic region additionally requires the limiting covariance in (59) *to be positive definite.*

The localization condition in Corollary C.10 has a direct sufficient verification for the two spherical averaging maps used most often in practice.

Corollary C.11 (A cap-preserving route for spherical full averages)**.** *Let $K_\rho = \{x \in \mathbb{S}^q : d(x, x_\star) \le \rho\}$ with $\rho < \pi/4$. Suppose the original state and every fixed multiplier state are initialized in K_ρ, are kept in K_ρ by a measurable cap-preserving localization rule, converge almost surely to $x_\star$, and satisfy the $O(k^{-\alpha})$ second-moment state bounds under the original and joint laws. If the full iterative averages use either the exponential map and logarithm or the normalized projection retraction and its local inverse, then every full average remains in K_ρ, converges almost surely to $x_\star$, and has the $O(k^{-\alpha})$ second-moment rate. Consequently, each base or fixed multiplier average enters the comparison chart after its own almost surely finite index, without a restart or a maximum entry time across multiplier paths.*

The cap-membership clause concerns the output of the localization rule. Rescaling a tangent proposal controls its norm but does not by itself certify that the retracted state remains in K_ρ.

The radius $r_0 < \pi/2$ is stronger than uniqueness alone, but it gives cut-locus avoidance, a smooth bounded score, and positive local curvature in one condition; related convexity radii are standard in the theory of Riemannian centers of mass (Afsari, 2011). If the support approaches the antipode, neither the displayed Hessian nor a standard Gaussian limit should be used without a separate cut-locus analysis.

C.3.2 Hadamard Fréchet means and positive-definite matrices

Let $\mathcal{M}$ be a finite-dimensional Hadamard manifold and let

$$F(x) = \frac{1}{2}\mathbb{E}d(x, Y)^2, \qquad x_\star = \arg\min_x F(x).$$

Assume that Y is supported in a compact set, and localize the iterations to a compact geodesically convex neighborhood containing $x_\star$. Hadamard geometry removes the cut locus and makes squared distance geodesically strongly convex (Bačák, 2014). Put

$$H_F = \mathrm{Hess}\, F(x_\star), \qquad Q_F = \mathbb{E}\{\mathrm{Log}_{x_\star}(Y) \otimes \mathrm{Log}_{x_\star}(Y)\}.$$

Corollary C.12 (Hadamard Fréchet mean)**.** *Under the preceding support and compact-localization conditions, suppose the original and weighted states and full iterative averages satisfy the localization clause of Assumption 3.3. Use $\Psi(x;Y) = -\operatorname{Log}_x(Y)$ and suppose the $O(\eta_k)$ state rates hold. Then*

$$\sqrt{n}\,\operatorname{Log}_{x_\star}(\bar{x}_n) \rightsquigarrow N(0, H_F^{-1} Q_F H_F^{-1}), \tag{60}$$

and the full-path multiplier law is conditionally consistent.

A principal instance is the $d = m(m+1)/2$ dimensional manifold $\mathcal{P}_m$ of $m \times m$ symmetric positive-definite matrices with the affine-invariant metric $\langle \mathrm{U}, \mathrm{V} \rangle_{\mathrm{X}} = \operatorname{tr}(\mathrm{X}^{-1}\mathrm{U}\mathrm{X}^{-1}\mathrm{V})$. Its exponential and logarithm maps are

$$\operatorname{Exp}_{\mathrm{X}}(\mathrm{U}) = \mathrm{X}^{1/2} \exp(\mathrm{X}^{-1/2}\mathrm{U}\mathrm{X}^{-1/2})\mathrm{X}^{1/2}, \tag{61}$$

$$\operatorname{Log}_{\mathrm{X}}(\mathrm{Y}) = \mathrm{X}^{1/2} \log(\mathrm{X}^{-1/2}\mathrm{Y}\mathrm{X}^{-1/2})\mathrm{X}^{1/2}. \tag{62}$$

This geometry is complete, simply connected, and nonpositively curved (Pennec et al., 2006). If the eigenvalues of Y lie almost surely in a fixed interval $[a, b] \subset (0, \infty)$, all local score and smoothness conditions hold on an eigenvalue-bounded neighborhood. For $\phi(\mathrm{X}) = \log\det \mathrm{X}$,

$$D\phi_{\mathrm{X}_\star}[\mathrm{U}] = \operatorname{tr}(\mathrm{X}_\star^{-1}\mathrm{U}), \tag{63}$$

so Corollaries C.2 and 3.7 give direct intervals for the log generalized volume without an explicit estimator of H_F, provided $D\phi_{\mathrm{X}_\star} V (D\phi_{\mathrm{X}_\star})^\dagger > 0$. For $\phi_{\max}(\mathrm{X}) = \log \lambda_{\max}(\mathrm{X})$, the same scalar theory applies only when the largest eigenvalue of $\mathrm{X}_\star$ is simple. If $v_{\max}$ is its normalized leading eigenvector and the limiting variance is nonzero, then

$$D\phi_{\max,\mathrm{X}_\star}[\mathrm{U}] = \frac{v_{\max}^{\mathsf{T}} \mathrm{U} v_{\max}}{\lambda_{\max}(\mathrm{X}_\star)}. \tag{64}$$

A positive target eigengap is therefore an explicit regularity condition for the largest-log-eigenvalue intervals below; no such inference is claimed at a repeated top eigenvalue.

Corollary C.13 (Affine-invariant log-normal moment route)**.** *Let* E *be a symmetric random matrix whose law is non-atomic, with* $\mathrm{E} \overset{d}{=} -\mathrm{E}$ *and* $\mathbb{E}\,\|\mathrm{E}\|_F^{2+\delta} < \infty$ *for some* $\delta > 0$*, and set* $\mathrm{Y} = \mathrm{A}^{1/2} \exp(\mathrm{E}) \mathrm{A}^{1/2}, \quad \mathrm{A} \in \mathcal{P}_m$*. If the original and weighted states and full iterative averages are localized in a common compact neighborhood of* A *in the sense of Assumption 3.3, satisfy* (16)*, and the population mean field satisfies* (45) *for some* $\rho \in (0, 1]$ *with* $\alpha(1+\rho) > 1$ *and nonsingular derivative, then* A *is the unique affine-invariant Fréchet mean and the conclusions of Corollary C.12 hold. For the normalized median score, assume separately that its original and weighted states and full iterative averages satisfy the same localization requirements, that its states satisfy* (16)*, that* A *is the unique median, and that its population field is locally $C^{1,\rho}$ with nonsingular derivative and* $\alpha(1+\rho) > 1$*. Then Corollary C.14 applies to the median iteration.*

The bounded-eigenvalue condition packages compact localization, finite score moments, and uniform derivative bounds, but is not necessary. Corollary C.13 gives an unbounded score-moment route: on a compact state set, the affine-invariant score and its local derivative admit an envelope of order $1 + \|\mathrm{E}\|_F$, so the stated moment controls the score second moment and local L^2 modulus. The population remainder (45) is still required separately. The unweighted Gaussian limit for intrinsic means is classical (Bhattacharya and Patrangenaru, 2003; 2005); the result here supplies the online full-path conditional law and fixed-replication calibration.

C.3.3 Riemannian geometric medians

Let $\mathcal{M}$ be a finite-dimensional Hadamard manifold, assume $\mathbb{E}d(x_0, Y) < \infty$ for some $x_0 \in \mathcal{M}$, and consider the population objective $x \mapsto \mathbb{E}d(x, Y)$. Suppose $x_\star$ is its unique minimizer, $\mathbb{P}(Y = x_\star) = 0$, and the population field

$$G_{\mathrm{med}}(x) = -\mathbb{E}\left\{ \frac{\operatorname{Log}_x(Y)}{d(x, Y)} \right\}$$

exists and is $C^{1,\rho}$ in a neighborhood of $x_\star$, with nonsingular derivative $H_{\mathrm{med}} = DG_{\mathrm{med}}(x_\star)$. Define

$$Q_{\mathrm{med}} = \mathbb{E}\left[\frac{\operatorname{Log}_{x_\star}(Y)}{d(x_\star, Y)} \otimes \frac{\operatorname{Log}_{x_\star}(Y)}{d(x_\star, Y)} \right].$$

The stochastic score has norm one away from $Y = x$ and is samplewise nonsmooth at the observation.

Corollary C.14 (Riemannian geometric median)**.** *Use the score*

$$\Psi(x;Y) = -\frac{\mathrm{Log}_x(Y)}{d(x,Y)}, \qquad Y \neq x,$$

with any measurable convention of uniformly bounded norm at $Y = x$. Suppose the original and weighted states and full iterative averages satisfy the localization clause of Assumption 3.3, the states satisfy the local rates in (16)*, $\alpha(1+\rho) > 1$, and the law of Y is non-atomic on their common localized neighborhood. Thus, for every fixed path and time, the exceptional events $Y_k = x_{k-1}$ and $Y_k = x^\star_{k-1}$ have probability zero. Then*

$$\sqrt{n}\ \mathrm{Log}_{x_\star}(\bar{x}_n) \rightsquigarrow N\left(0, H_{\mathrm{med}}^{-1} Q_{\mathrm{med}} H_{\mathrm{med}}^{-\dagger}\right), \tag{65}$$

and the conditional multiplier approximation is valid. Fixed-B scalar intervals and full intrinsic regions follow under, respectively, the nonzero-variance and positive-definiteness conditions of Corollary 3.7.

Existence, uniqueness, and algorithms for Riemannian medians are studied by Fletcher et al. (2009) and Yang (2010). The substantive condition here is differentiability of the *population* field, not of the sample distance. A concrete sufficient model, including the population remainder and state stability, is given in Proposition C.15. Non-atomicity excludes equality at either predictable state, and dominated convergence yields the local L^2 modulus. The example therefore shows how first-order full-path inference can accommodate a samplewise nonsmooth score.

Proposition C.15 (A verifiable nonsmooth median model)**.** *On $\mathcal{P}_m$ with the affine-invariant metric, let $m \geq 2$, $d = m(m+1)/2$, and fix $\mathrm{A} \in \mathcal{P}_m$. Suppose Y has density $p_{\mathrm{A}}(y) = c_R h\{d(\mathrm{A}, y)^2\}$ with respect to Riemannian volume, where c_R normalizes the density, $R > 0$, and*

$$h(s) = \begin{cases} \exp\{-1/(R^2 - s)\}, & s < R^2, \\ 0, & s \geq R^2. \end{cases}$$

Then A is its unique geometric median, the population median field is locally $C^{1,1}$, and its derivative and innovation covariance at A are positive definite. Choose a specified compact geodesically convex K with $\mathrm{A} \in \mathrm{int}\, K$, initialize in K, and use the projection safeguard described above with exponential updates and averaging. For sufficiently small η_0, $\alpha \in (1/2, 1)$, and independent multipliers satisfying (10)*, the state-rate, full-average localization, and primitive conditions hold. Hence Corollary C.14 and its fixed-B conclusions apply, including with signed multipliers.*

The proof in Section C.4.5 verifies the conditions from a compact smooth density. This example establishes attainability of the nonsmooth theory; the separate unbounded log-normal model in Corollary C.13 retains its explicit population-regularity and localization hypotheses.

C.3.4 Streaming principal-subspace estimation

Let $Y \in \mathbb{R}^p$ be i.i.d., centered, with covariance

$$\Sigma = [\mathrm{U}_\star, \mathrm{U}_\perp] \begin{pmatrix} \Lambda_1 & 0 \\ 0 & \Lambda_2 \end{pmatrix} [\mathrm{U}_\star, \mathrm{U}_\perp]^{\mathsf{T}},$$

where $\mathrm{U}_\star$ spans the leading r-dimensional eigenspace and $\lambda_r > \lambda_{r+1}$. On $\mathrm{Gr}(r,p)$, consider

$$F([\mathrm{U}]) = -\frac{1}{2}\mathbb{E}\left\|\mathrm{U}^{\mathsf{T}} Y\right\|^2, \tag{66}$$

with sample Riemannian gradient $\Psi([\mathrm{U}]; Y) = -(\mathrm{I} - \mathrm{U}\mathrm{U}^{\mathsf{T}}) Y Y^{\mathsf{T}} \mathrm{U}$. A horizontal tangent vector at $[\mathrm{U}_\star]$ is $\mathrm{U}_\perp \mathrm{K}$, with $\mathrm{K} \in \mathbb{R}^{(p-r)\times r}$. Define

$$\mathcal{H}(\mathrm{K}) = \mathrm{K}\Lambda_1 - \Lambda_2 \mathrm{K}, \tag{67}$$

$$\mathrm{Q}_{\mathrm{PCA}} = \mathrm{Cov}\{\mathrm{vec}(Y_2 Y_1^{\mathsf{T}})\}, \qquad Y_1 = \mathrm{U}_\star^{\mathsf{T}} Y, \quad Y_2 = \mathrm{U}_\perp^{\mathsf{T}} Y. \tag{68}$$

The eigengap makes $\mathcal{H}$ positive definite. Let $\mathrm{H}_{\mathrm{vec}}$ denote its $r(p-r) \times r(p-r)$ matrix in the vectorized coordinates $\mathrm{vec}(\mathrm{K})$.

Corollary C.16 (Grassmann principal-subspace bootstrap)**.** *Assume* $\mathbb{E}\,\|Y\|^4 < \infty$*. Use the exponential map and parallel transport, or a named retraction–inverse–transport triple verified to satisfy* (18)*,* (20)*, and* (48)*. Suppose the original and weighted states and full iterative averages satisfy the localization clause of Assumption 3.3 in the attraction basin of* $[\mathrm{U}_\star]$*, and the states satisfy* (16)*. Then*

$$\sqrt{n}\ \mathrm{vec}\{\mathrm{U}_\perp^{\mathsf{T}} R_{[\mathrm{U}_\star]}^{-1}(\bar{\mathrm{U}}_n)\} \rightsquigarrow N\left(0, \mathrm{H}_{\mathrm{vec}}^{-1}\mathrm{Q}_{\mathrm{PCA}}\mathrm{H}_{\mathrm{vec}}^{-\mathsf{T}}\right), \tag{69}$$

and the multiplier law is conditionally consistent.

If Y *is Gaussian in the eigenbasis, tangent coordinate* (a, j) *has asymptotic variance*

$$\frac{\lambda_{r+a}\lambda_j}{(\lambda_j - \lambda_{r+a})^2}, \qquad a = 1, \dots, p-r, \quad j = 1, \dots, r. \tag{70}$$

For a symmetric matrix A *and the projector functional* $\phi_{\mathrm{A}}([\mathrm{U}]) = \mathrm{tr}(\mathrm{AUU}^{\mathsf{T}})$*, the differential at* $[\mathrm{U}_\star]$ *is*

$$D\phi_{\mathrm{A}}[\mathrm{U}_\perp \mathrm{K}] = 2\left\langle \mathrm{U}_\perp^{\mathsf{T}}\mathrm{AU}_\star, \mathrm{K}\right\rangle_F. \tag{71}$$

Thus Corollary 3.7 gives a fixed-replication interval for a specified projector contrast whenever $D\phi_{\mathrm{A}} V (D\phi_{\mathrm{A}})^\dagger > 0$*. A contrast with zero differential, such as* $\mathrm{A} = \mathrm{I}$*, is not covered by the nondegenerate scalar pivot.*

The fourth data moment supplies a uniform square-integrable envelope for the quadratic score and its local increments; it is sufficient, rather than equivalent to a target-coordinate innovation moment. The eigengap is the local identifiability condition, and (70) shows its inverse-square effect on uncertainty. The scalar delta method applies to smooth projector contrasts; principal-angle maxima are nonregular at coincident subspaces and require a different limit theory.

Existing Riemannian averaging and generalized-eigenvector results provide the unweighted local benchmark (Tripuraneni et al., 2018; Li and Jordan, 2023). Rank-one online Oja bootstrap and entrywise inference are studied by Lunde et al. (2021) and Kumar et al. (2025). The general-rank online multiplier method of Xu and Li (2026) concerns the high-dimensional final iterate with a run-constant step and bounded nonnegative weights. The result above follows from a fixed-dimensional decreasing-step Polyak-Ruppert theorem, permits signed finite-second-moment multipliers for weak validity, and supplies fixed-B pivots for smooth contrasts.

C.4 Proofs

The following subsections collect every proof and analytic derivation. Result statements needed within a proof are included at their point of use.

C.4.1 Geometric and Averaging Lemmas

This appendix proves the deterministic and probabilistic lemmas used in the main argument. Throughout, constants may change from line to line.

Lemma C.17 (Abel boundary)**.** *Let* $\eta_k = \eta_0 k^{-\alpha}$ *with* $\alpha \in (0, 1)$ *and suppose* $\mathbb{E}\,\|u_k\|^2 \le C\eta_k$*. Then*

$$\frac{1}{\sqrt{n}} \sum_{k=1}^{n} \frac{u_{k-1} - u_k}{\eta_k} = o_{L^1}(1). \tag{72}$$

The same conclusion holds under the joint original–multiplier law if $\mathbb{E}\mathbb{E}^\star\,\|u_k^\star\|^2 \le C\eta_k$*.*

Proof. Summation by parts gives

$$\sum_{k=1}^{n} \frac{u_{k-1} - u_k}{\eta_k} = \frac{u_0}{\eta_1} + \sum_{k=1}^{n-1} u_k \left(\frac{1}{\eta_{k+1}} - \frac{1}{\eta_k}\right) - \frac{u_n}{\eta_n}.$$

By Cauchy–Schwarz, $\mathbb{E}\,\|u_k\| \le C\eta_k^{1/2} = Ck^{-\alpha/2}$. Since $\eta_{k+1}^{-1} - \eta_k^{-1} \le Ck^{\alpha-1}$,

$$\frac{1}{\sqrt{n}} \sum_{k=1}^{n-1} \mathbb{E}\,\|u_k\| \left(\frac{1}{\eta_{k+1}} - \frac{1}{\eta_k}\right) \le \frac{C}{\sqrt{n}} \sum_{k=1}^{n-1} k^{\alpha/2-1} \le Cn^{(\alpha-1)/2} \to 0.$$

The terminal term satisfies

$$\frac{\mathbb{E}\,\|u_n\|}{\sqrt{n}\eta_n} \le Cn^{-1/2}n^{\alpha}n^{-\alpha/2} = Cn^{(\alpha-1)/2} \to 0,$$

and the initial term is $O(n^{-1/2})$. This proves (72). Replacing $\mathbb{E}$ by $\mathbb{E}\mathbb{E}^\star$ gives the weighted statement. □

Lemma C.18 (Local expansion of the iterative intrinsic average)**.** *Suppose* (18) *holds on* $B_{T_{x_\star}\mathcal{M}}(0,r)$. *Let* $v_k = R_{x_\star}^{-1}(\bar{x}_k)$ *and assume* $u_k, v_k \in B_{T_{x_\star}\mathcal{M}}(0,r)$. *Then*

$$v_k = (1-k^{-1})v_{k-1} + k^{-1}u_k + \Delta_k, \qquad \|\Delta_k\| \le \frac{C}{k}(\|v_{k-1}\|^2 + \|u_k\|^2). \tag{73}$$

If $\mathbb{E}\,\|u_k\|^2 \le Ck^{-\alpha}$ *for* $\alpha \in (0,1)$ *and* r *is sufficiently small, then*

$$\mathbb{E}\,\|v_k\|^2 \le Ck^{-\alpha}. \tag{74}$$

Consequently, if $\alpha > 1/2$,

$$\sqrt{n}\left(v_n - \frac{1}{n}\sum_{k=1}^{n} u_k\right) = o_{L^1}(1).$$

All conclusions hold under the joint law for a weighted trajectory.

Proof. The definition (2) gives $v_k = A_{1/k}(v_{k-1}, u_k)$, and (18) yields (73). Put $w_k = (1-k^{-1})v_{k-1} + k^{-1}u_k$. Convexity of the squared norm gives

$$\|w_k\|^2 \le (1-k^{-1})\,\|v_{k-1}\|^2 + k^{-1}\,\|u_k\|^2\,.$$

For $u_k, v_{k-1} \in B(0,r)$, (73) implies $\|\Delta_k\| \le Crk^{-1}(\|v_{k-1}\| + \|u_k\|)$. Expanding $\|w_k + \Delta_k\|^2$ and increasing C gives

$$\|v_k\|^2 \le \left(1 - \frac{1-Cr}{k}\right)\|v_{k-1}\|^2 + \frac{1+Cr}{k}\,\|u_k\|^2\,.$$

Choose r so that $\lambda = 1 - Cr > \alpha$. With $A_k = \mathbb{E}\,\|v_k\|^2$, $A_k \le (1-\lambda/k)A_{k-1} + Ck^{-1-\alpha}$. A standard comparison, obtained by multiplying by k^λ and summing, gives $A_k \le Ck^{-\alpha}$, proving (74).

Multiplying (73) by k and telescoping yields

$$v_n - \frac{1}{n}\sum_{k=1}^{n} u_k = \frac{1}{n}\sum_{k=2}^{n} k\Delta_k.$$

Hence

$$\begin{aligned}\mathbb{E}\left\|\sqrt{n}(v_n - n^{-1}\textstyle\sum_{k=1}^{n} u_k)\right\| &\le \frac{C}{\sqrt{n}}\sum_{k=2}^{n}\{\mathbb{E}\,\|v_{k-1}\|^2 + \mathbb{E}\,\|u_k\|^2\} \\ &\le \frac{C}{\sqrt{n}}\sum_{k=1}^{n} k^{-\alpha} = O(n^{1/2-\alpha}) \to 0.\end{aligned}$$

The proof is pathwise up to expectations and therefore applies under the joint law. □

Lemma C.19 (Finite-prefix invariance of the intrinsic average)**.** *Let* $u_k = R_{x_\star}^{-1}(x_k)$ *and* $v_k = R_{x_\star}^{-1}(\bar{x}_k)$ *be the coordinates of the actual state and full iterative average. Suppose the state bound in* (16) *holds and there is an almost surely finite, path-specific index* τ *such that* $u_k, v_{k-1} \in B_{T_{x_\star}\mathcal{M}}(0,r)$ *and* (73) *holds for every* $k > \tau$. *Assume* r *is sufficiently small that the squared-norm comparison in the proof of Lemma C.18 holds with some* $\lambda \in (\alpha, 1)$. *Then, for* $\alpha > 1/2$,

$$\sqrt{n}\left(v_n - \frac{1}{n}\sum_{k=1}^{n} u_k\right) = o_p(1). \tag{75}$$

The same conclusion is $o_{\mathbb{P}^\star}(1)$ *in probability for a weighted trajectory under the joint state bound. The result uses the reported full average and does not restart it at* τ.

Proof. For $n > \tau$, multiplying (73) by k and telescoping only over the indices after τ gives the exact identity

$$v_n - \frac{1}{n}\sum_{k=1}^{n} u_k = \frac{\tau v_\tau - \sum_{k=1}^{\tau} u_k}{n} + \frac{1}{n}\sum_{k=\tau+1}^{n} k\Delta_k.$$

The first numerator is an almost surely finite random vector. Take $\tau \geq 1$ without loss of generality. For $M \geq 1$ and $L < \infty$, put

$$E_{M,L} = \{\tau \leq M,\ \max_{1\leq j\leq M} \|v_j\| \leq L\}.$$

On this event, for every $k > M$, the pathwise squared-norm comparison gives

$$\|v_k\|^2 \mathbf{1}(E_{M,L}) \leq (1 - \lambda/k)\|v_{k-1}\|^2 \mathbf{1}(E_{M,L}) + (C/k)\|u_k\|^2 \mathbf{1}(E_{M,L}).$$

Taking unconditional expectations, using the state bound, and iterating from M yields

$$\mathbb{E}\{\|v_k\|^2 \mathbf{1}(E_{M,L})\} \leq C_{M,L}k^{-\lambda} + Ck^{-\alpha}.$$

Only a deterministic inequality has been multiplied by the indicator, so no martingale centering conditional on $E_{M,L}$ is required. The remainder bound now gives

$$\mathbb{E}\left\{\mathbf{1}(E_{M,L})\frac{1}{\sqrt{n}}\sum_{k=M+1}^{n} k\|\Delta_k\|\right\} \leq Cn^{1/2-\alpha} + C_{M,L}n^{1/2-\lambda} = o(1).$$

On $E_{M,L}$, the prefix through M has a finite numerator and is negligible after division by $\sqrt{n}$. Let $n \to \infty$, then $L \to \infty$ and $M \to \infty$. Since τ and all prefix coordinates are finite almost surely, the excluded probability tends to zero. This proves (75). The same proof under the joint law gives the weighted result. For a jointly negligible remainder $R_n^\star$ and $\varepsilon, \delta > 0$,

$$\mathbb{P}\{\mathbb{P}^\star(\|R_n^\star\| > \varepsilon) > \delta\} \leq \delta^{-1}\mathbb{P}(\|R_n^\star\| > \varepsilon) \to 0,$$

which proves the conditional conclusion. □

Lemma C.20 (Change of base point)**.** *Under Assumption 3.3, let $a_n, b_n \in T_{x_\star}\mathcal{M}$ satisfy $a_n = O_p(n^{-1/2})$ and $b_n = O_{\mathbb{P}^\star}(n^{-1/2})$ in probability. Put $x_n = R_{x_\star}(a_n)$ and $y_n = R_{x_\star}(b_n)$. Then*

$$\sqrt{n}\,\mathcal{P}_{x_n\to x_\star} R_{x_n}^{-1}(y_n) = \sqrt{n}(b_n - a_n) + o_{\mathbb{P}^\star}(1) \quad \textit{in probability}. \tag{76}$$

In particular,

$$\sqrt{n}\,\mathcal{P}_{x_n\to x_\star} R_{x_n}^{-1}(x_\star) = -\sqrt{n}a_n + o_p(1). \tag{77}$$

Proof. By (20),

$$\Gamma(a_n, b_n) = b_n - a_n + R_n, \qquad \|R_n\| \leq C(\|a_n\| + \|b_n\|)\|b_n - a_n\|.$$

The assumed orders imply $\sqrt{n}\|R_n\| = O_{\mathbb{P}^\star}(n^{-1/2})$, proving (76). Taking $b_n = 0$ gives (77). □

Lemma C.21 (One-step coordinate remainder)**.** *Under Assumption C.6, let $q \in \mathbb{R}$, $u \in B_{T_{x_\star}\mathcal{M}}(0, r)$, and let $\psi \in T_{x_\star}\mathcal{M}$ be such that $z = -\eta q\psi$ lies in the coordinate domain. Then*

$$\Phi(u, -\eta q\psi) = u - \eta q\psi + \epsilon(u, q, \psi, \eta),$$

where

$$\|\epsilon(u, q, \psi, \eta)\| \leq C\{\eta|q|\,\|u\|^2\,\|\psi\| + \eta^2 q^2\,\|\psi\|^2\}. \tag{78}$$

Proof. Substitute $z = -\eta q\psi$ into (48). The resulting upper bound is exactly (78). □

Lemma C.22 (Two non-exponential implementation triples)**.** *Use Riemannian parallel transport in the comparison maps.*

(i) On $\mathbb{S}^q$, let

$$R_x^{\mathrm{proj}}(v) = \frac{x+v}{\|x+v\|}, \qquad (R_x^{\mathrm{proj}})^{-1}(y) = \frac{y}{\langle x, y\rangle} - x, \quad \langle x, y\rangle > 0.$$

(ii) On $\mathrm{Gr}(r,p)$*, for a horizontal matrix* Δ *with* $\mathrm{U}^\mathsf{T}\Delta = 0$*, let*

$$R^{\mathrm{QR}}_{[\mathrm{U}]}(\Delta) = [\mathrm{qf}(\mathrm{U}+\Delta)], \qquad (R^{\mathrm{QR}}_{[\mathrm{U}]})^{-1}([\mathrm{V}]) = (\mathrm{I}-\mathrm{U}\mathrm{U}^\mathsf{T})\mathrm{V}(\mathrm{U}^\mathsf{T}\mathrm{V})^{-1},$$

whenever $\mathrm{U}^\mathsf{T}\mathrm{V}$ *is invertible. Here* $\mathrm{qf}(\mathrm{Z})$ *is the orthonormal* Q *factor in a thin QR factorization of a full-column-rank matrix* Z*; the Grassmann point is independent of the column-sign convention.*

On sufficiently small compact coordinate balls, each displayed retraction, inverse, and parallel-transport triple satisfies (18)*,* (20)*, and* (48)*.*

Proof. For the sphere, $\|x+v\| = (1+\|v\|^2)^{1/2}$ because $v \perp x$, so $R^{\mathrm{proj}}_x(v) = x + v - \frac{1}{2}\|v\|^2 x + O(\|v\|^3)$. This agrees with $\mathrm{Exp}_x(v)$ through second order, and direct substitution verifies the displayed local inverse. For the Grassmannian, $[\mathrm{qf}(\mathrm{U}+\Delta)]$ equals the polar-retraction subspace $[(\mathrm{U}+\Delta)(\mathrm{I}+\Delta^\mathsf{T}\Delta)^{-1/2}]$ exactly. The latter agrees with the Grassmann exponential through second order (Absil et al., 2008; Absil and Malick, 2012). If $[\mathrm{V}] = R^{\mathrm{QR}}_{[\mathrm{U}]}(\Delta)$, then $\mathrm{V} = (\mathrm{U}+\Delta)\mathrm{C}$ for an invertible C, and the displayed graph inverse returns Δ.

It remains to translate second-order agreement into the three stated comparisons. In normal coordinates at the target, parallel transport has identity differential and no term linear in the base coordinate. For either smooth second-order triple, the maps in (17), (19), and (47) therefore have, respectively, first-order parts $(1-t)a+tb$, $b-a$, and $u+z$. More explicitly, the Jacobi-field expansion of the normal-coordinate logarithm, composed with radial parallel transport, gives $D_z\Phi(u,0) = I + O(\|u\|^2)$ uniformly on a smaller coordinate ball; second-order agreement of the displayed retractions and inverses preserves this derivative estimate. Taylor's formula in the step coordinate then yields

$$\Phi(u,z) - u - z = \{D_z\Phi(u,0) - I\}z + \int_0^1 (1-s)D^2_{zz}\Phi(u,sz)[z,z]\,ds,$$

and the second derivative is uniformly bounded. The other Taylor remainders vanish when $t=0$ and when $a=b$. Uniform bounds on the same compact ball consequently give

$$Ct(\|a\|^2 + \|b\|^2), \qquad C(\|a\| + \|b\|)\,\|b-a\|\,, \qquad C\{\|u\|^2\,\|z\| + \|z\|^2\},$$

which are exactly (18), (20), and (48). □

C.4.2 Proof of the Abstract Bootstrap Theorem

Lemma C.23 (Negligible martingale remainders)**.** *Let* $(z_k, \mathcal{H}_k)$ *be a finite-dimensional martingale-difference sequence. If*

$$\frac{1}{n}\sum_{k=1}^n \mathbb{E}(\|z_k\|^2 \mid \mathcal{H}_{k-1}) \xrightarrow{p} 0,$$

then $n^{-1/2}\sum_{k=1}^n z_k \to_p 0$*. The same statement holds under the joint original–multiplier law.*

Proof. Choose an orthonormal basis $e_1, \dots, e_d$ of the state space. For each n, define the martingale $M^{(n)}_{m,j} = n^{-1/2}\sum_{k=1}^m \langle z_k, e_j\rangle$ for $0 \le m \le n$. Its predictable quadratic variation is bounded by

$$\langle M^{(n)}_{\cdot,j}\rangle_n \le \frac{1}{n}\sum_{k=1}^n \mathbb{E}(\|z_k\|^2 \mid \mathcal{H}_{k-1}).$$

For $a, \gamma > 0$, Lenglart's inequality gives

$$\mathbb{P}(|M^{(n)}_{n,j}| > a) \le \frac{\gamma}{a^2} + \mathbb{P}\{\langle M^{(n)}_{\cdot,j}\rangle_n > \gamma\}.$$

First let $n \to \infty$, then $\gamma \downarrow 0$. Hence each coordinate converges to zero in probability. Finite dimensionality gives the vector conclusion. The same proof applies to the joint filtration. □

Lemma C.24 (Abel expansion for the base path)**.** *Under Assumption 3.2,*

$$\frac{1}{\sqrt{n}}\sum_{k=1}^n u_{k-1} = -H^{-1}\frac{1}{\sqrt{n}}\sum_{k=1}^n \xi_k + o_p(1). \tag{79}$$

Proof. Rearrange (11):

$$Hu_{k-1} = \frac{u_{k-1} - u_k}{\eta_k} - \xi_k - b_k - m_k - e_k.$$

Sum from $k = 1$ to n and divide by $\sqrt{n}$. The first term on the right is $o_p(1)$ by Lemma C.17 and (16). The bias and drift terms are $o_p(1)$ by (13) and (15). The normalized martingale remainder is $o_p(1)$ by Lemma C.23 and (14). Multiplication by the fixed inverse H^{-1} proves (79). □

Lemma C.25 (Abel expansion for the weighted path)**.** *Under Assumption 3.2,*

$$\frac{1}{\sqrt{n}} \sum_{k=1}^{n} u^\star_{k-1} = -H^{-1} \frac{1}{\sqrt{n}} \sum_{k=1}^{n} W_k \xi_k + o_{\mathbb{P}^\star}(1) \tag{80}$$

in probability.

Proof. Rearranging (12) gives

$$Hu^\star_{k-1} = \frac{u^\star_{k-1} - u^\star_k}{\eta_k} - W_k \xi_k - W_k b_k - m^\star_k - e^\star_k. \tag{81}$$

The normalized Abel boundary is negligible under the joint law by Lemma C.17 and (16). The martingale and drift remainders are negligible under the joint law by Lemma C.23, (14), and (15).

For the bias, write

$$\frac{1}{\sqrt{n}} \sum_{k=1}^{n} W_k b_k = \frac{1}{\sqrt{n}} \sum_{k=1}^{n} b_k + \frac{1}{\sqrt{n}} \sum_{k=1}^{n} (W_k - 1) b_k.$$

The first term is $o_p(1)$. Conditionally on the complete original stream, the second term has mean zero and second moment

$$\mathbb{E}^\star \left\| \frac{1}{\sqrt{n}} \sum_{k=1}^{n} (W_k - 1) b_k \right\|^2 = \frac{1}{n} \sum_{k=1}^{n} \|b_k\|^2 \xrightarrow{p} 0. \tag{82}$$

Conditional Chebyshev therefore makes it $o_{\mathbb{P}^\star}(1)$ in probability.

All remaining remainders converge to zero under the joint law. If $R^\star_n \to 0$ in joint probability, then for every $\varepsilon > 0$,

$$\mathbb{P}\{\mathbb{P}^\star(\|R^\star_n\| > \varepsilon) > \varepsilon\} \le \varepsilon^{-1} \mathbb{P}(\|R^\star_n\| > \varepsilon) \to 0.$$

Thus joint convergence implies conditional convergence in probability. Multiplying (81) by H^{-1} proves (80). □

Proof of Theorem 3.4. Apply Lemmas C.24 and C.25. By Lemma C.18 when the local formula holds throughout, or by Lemma C.19 under the finite-exception localization formulation, together with (16),

$$\sqrt{n} \left\{ R^{-1}_{x_\star}(\bar{x}_n) - \frac{1}{n} \sum_{k=1}^{n} u_k \right\} = o_p(1),$$

$$\sqrt{n} \left\{ R^{-1}_{x_\star}(\bar{x}^\star_n) - \frac{1}{n} \sum_{k=1}^{n} u^\star_k \right\} = o_{\mathbb{P}^\star}(1) \quad \text{in probability.}$$

Changing the summation index from u_{k-1} to u_k contributes only $(u_n - u_0)/\sqrt{n} = o_p(1)$, and similarly under the joint law for the weighted path. Combining these relations with (79) and (80) proves (22)–(23). □

C.4.3 Martingale and Conditional Multiplier Central Limit Theorems

Lemma C.26 (Optional covariance and realized Lindeberg condition)**.** *Under Assumption 3.1,*

$$\frac{1}{n}\sum_{k=1}^{n}\xi_k\otimes\xi_k \xrightarrow{p} Q, \tag{83}$$

$$\frac{1}{n}\sum_{k=1}^{n}\|\xi_k\|^2\,\mathbf{1}\{\|\xi_k\|>\varepsilon\sqrt{n}\}\xrightarrow{p}0 \quad \textit{for every } \varepsilon>0. \tag{84}$$

Proof. Set $X_{n,k}=n^{-1/2}\xi_k$. For fixed $\varepsilon>0$, the nonnegative adapted process

$$A_{n,j}=\sum_{k=1}^{j}\|X_{n,k}\|^2\,\mathbf{1}\{\|X_{n,k}\|>\varepsilon\}$$

has predictable compensator

$$\widetilde{A}_{n,j}=\sum_{k=1}^{j}\mathbb{E}[\|X_{n,k}\|^2\,\mathbf{1}\{\|X_{n,k}\|>\varepsilon\}\mid\mathcal{F}_{k-1}].$$

By (9), $\widetilde{A}_{n,n}\to_p 0$. Lenglart's domination inequality for nonnegative submartingales gives, for $a,b>0$, $\mathbb{P}(A_{n,n}>a)\le b/a+\mathbb{P}(\widetilde{A}_{n,n}>b)$. Let $n\to\infty$ and then $b\downarrow 0$ to obtain (84).

It remains to compare optional and predictable covariances. Define

$$\mathrm{Y}_{n,k}^{(\varepsilon)}:=X_{n,k}\otimes X_{n,k}\mathbf{1}\{\|X_{n,k}\|\le\varepsilon\}-\mathbb{E}[X_{n,k}\otimes X_{n,k}\mathbf{1}\{\|X_{n,k}\|\le\varepsilon\}\mid\mathcal{F}_{k-1}].$$

These are matrix-valued martingale differences. On the event

$$\sum_{k=1}^{n}\mathbb{E}(\|X_{n,k}\|^2\mid\mathcal{F}_{k-1})\le M,$$

the predictable sum of their squared Frobenius norms is at most $4\varepsilon^2M$. Coordinatewise Lenglart therefore implies $\sum_{k=1}^{n}\mathrm{Y}_{n,k}^{(\varepsilon)}=O_p(\varepsilon)$ after first restricting to the preceding event. The event has probability tending to one for sufficiently large M by (8). The omitted large-increment parts are $o_p(1)$ by (9), (84), and another application of Lenglart. Letting $\varepsilon\downarrow 0$ yields

$$\sum_{k=1}^{n}X_{n,k}\otimes X_{n,k}-\sum_{k=1}^{n}\mathbb{E}(X_{n,k}\otimes X_{n,k}\mid\mathcal{F}_{k-1})\xrightarrow{p}0.$$

Combining this with (8) proves (83). □

Lemma C.27 (Base martingale central limit theorem)**.** *Under Assumption 3.1,*

$$\frac{1}{\sqrt{n}}\sum_{k=1}^{n}\xi_k\rightsquigarrow N(0,Q). \tag{85}$$

Proof. For every fixed $a\in T_{x_\star}\mathcal{M}$, the scalar martingale array $n^{-1/2}\langle a,\xi_k\rangle$ has predictable variance converging to $\langle a,Qa\rangle$ by (8). Its conditional Lindeberg condition follows from (9) because $|\langle a,\xi_k\rangle|\le\|a\|\,\|\xi_k\|$. The scalar martingale central limit theorem (Hall and Heyde, 1980, Chapter 3) gives convergence of every linear functional. Cramér–Wold and finite dimensionality prove (85). □

Lemma C.28 (Conditional multiplier central limit theorem with second moments)**.** *Under Assumption 3.1,*

$$d_{\mathrm{BL}}\left[\mathcal{L}^\star\left\{\frac{1}{\sqrt{n}}\sum_{k=1}^{n}(W_k-1)\xi_k\right\},N(0,Q)\right]\xrightarrow{p}0. \tag{86}$$

Proof. Condition on the realized sequence $\xi_1, \ldots, \xi_n$. The summands are independent, centered, and their conditional covariance is $n^{-1}\sum_{k=1}^{n} \xi_k \otimes \xi_k$, which converges to Q in probability by Lemma C.26.

It remains to verify the conditional Lindeberg condition. Put $U = |W - 1|$. For $M > 0$ and $\varepsilon > 0$,

$$\begin{aligned}
&\frac{1}{n}\sum_{k=1}^{n} \mathbb{E}^\star\left[U_k^2 \left\|\xi_k\right\|^2 \mathbf{1}\{U_k \left\|\xi_k\right\| > \varepsilon\sqrt{n}\}\right] \\
&\quad \le \mathbb{E}\{U^2 \mathbf{1}(U > M)\}\frac{1}{n}\sum_{k=1}^{n} \left\|\xi_k\right\|^2 \\
&\qquad + \mathbb{E}(U^2)\frac{1}{n}\sum_{k=1}^{n} \left\|\xi_k\right\|^2 \mathbf{1}\{\left\|\xi_k\right\| > \varepsilon\sqrt{n}/M\}.
\end{aligned}$$

The empirical second moment is $O_p(1)$ by (83). Since $\mathbb{E}U^2 < \infty$, the first coefficient tends to zero as $M \to \infty$. For fixed M, the second line tends to zero in probability by (84). Hence the conditional Lindeberg condition holds in probability.

The conditional Lindeberg–Feller theorem now gives convergence of every conditional linear functional along every subsequence on which the covariance and Lindeberg conditions hold almost surely. Every subsequence has a further subsequence with this property, so the convergence holds in probability. Finite dimensionality upgrades Cramér–Wold convergence to bounded-Lipschitz convergence, proving (86). □

Proof of Theorem 3.5. The change-of-base lemma requires conditional root-n tightness of the weighted average. From (23),

$$\sqrt{n} R_{x_\star}^{-1}(\bar{x}_n^\star) = -H^{-1}\left\{\frac{1}{\sqrt{n}}\sum_{k=1}^{n} \xi_k + \frac{1}{\sqrt{n}}\sum_{k=1}^{n}(W_k - 1)\xi_k\right\} + o_{\mathbb{P}^\star}(1).$$

The first sum is $O_p(1)$ by Lemma C.27. Conditionally on the original stream, the second sum has second moment $n^{-1}\sum_k \left\|\xi_k\right\|^2 = O_p(1)$ by Lemma C.26; conditional Markov inequality therefore makes it $O_{\mathbb{P}^\star}(1)$ in probability. Hence $R_{x_\star}^{-1}(\bar{x}_n^\star) = O_{\mathbb{P}^\star}(n^{-1/2})$, as required.

Apply Theorem 3.4 and Lemma C.20 with $a_n = R_{x_\star}^{-1}(\bar{x}_n)$ and $b_n = R_{x_\star}^{-1}(\bar{x}_n^\star)$. On the event that the local inverse is available, subtracting (22) from (23) and using Lemma C.20 gives (24). The conditional probability of the complementary event tends to zero in probability by the root-n tightness just established and the local inverse domain, so the measurable completion in (21) preserves the expansion. Lemma C.28 and conditional Slutsky yield (25). By Lemma C.27 and (22), the sampling law of $\sqrt{n} R_{x_\star}^{-1}(\bar{x}_n)$ converges to $N(0, V)$. The triangle inequality for d_{BL} proves (26). □

Proof of Theorem 3.6. Let $f_0, \ldots, f_B$ be bounded continuous functions on $T_{x_\star}\mathcal{M}$. Conditional on the original stream, the B multiplier paths are independent and identically distributed. Hence

$$\begin{aligned}
&\mathbb{E}\left[f_0(Z_n)\prod_{b=1}^{B} f_b(T_{n,b}^\star)\right] \\
&\qquad = \mathbb{E}\left[f_0(Z_n)\prod_{b=1}^{B} \mathbb{E}^\star f_b(T_n^\star)\right].
\end{aligned}$$

By (25), for each b, $\mathbb{E}^\star f_b(T_n^\star) \to_p \mathbb{E} f_b(Z_b)$, where $Z_b \sim N(0, V)$. Boundedness makes this convergence hold in L^1, and therefore the product of the conditional expectations converges in L^1 to $\prod_{b=1}^{B} \mathbb{E} f_b(Z_b)$. By Lemma C.27 and (22), $Z_n \rightsquigarrow Z_0 \sim N(0, V)$. Consequently,

$$\mathbb{E}\left[f_0(Z_n)\prod_{b=1}^{B} f_b(T_{n,b}^\star)\right] \longrightarrow \prod_{b=0}^{B} \mathbb{E} f_b(Z_b).$$

Finite products of bounded continuous functions form a convergence-determining class on the finite product space. This proves (27). □

Proof of Corollary 3.7. For the scalar statement, differentiability and Lemma C.20 give jointly for $b = 1, \dots, B$,

$$\sqrt{n}\{\phi(\bar{x}_n) - \phi(x_\star)\} = D\phi_{x_\star} Z_n + o_p(1),$$
$$\sqrt{n}\{\phi(\bar{x}_n^{\star(b)}) - \phi(\bar{x}_n)\} = D\phi_{x_\star} T_{n,b}^\star + o_p(1).$$

By Theorem 3.6, after division by $\sigma_\phi > 0$, the first term and the B bootstrap differences converge to independent standard normal variables $N_0, N_1, \dots, N_B$. Therefore

$$\frac{\phi(\bar{x}_n) - \phi(x_\star)}{s_{\phi,n,B}} \rightsquigarrow \frac{N_0}{\{B^{-1}\sum_{b=1}^B N_b^2\}^{1/2}} \sim t_B.$$

The limiting denominator is positive almost surely, so $\mathbb{P}(s_{\phi,n,B} = 0) \to 0$. The Student quantile is continuous, and the fallback convention is asymptotically irrelevant, which proves (28).

For the intrinsic region, transport $\widehat{S}_{n,B}$ from $T_{\bar{x}_n}\mathcal{M}$ to $T_{x_\star}\mathcal{M}$ and identify the latter with $\mathbb{R}^d$ through an orthonormal basis. In the following Gaussian calculation, S_B and V denote the corresponding covariance matrices. By Theorem 3.6 and the continuous mapping theorem,

$$\mathcal{P}_{\bar{x}_n \to x_\star} \widehat{S}_{n,B} \mathcal{P}_{x_\star \to \bar{x}_n} \rightsquigarrow \mathrm{S}_B := \frac{1}{B}\sum_{b=1}^B Z_b \otimes Z_b.$$

Since V is positive definite and $B \geq d$, S_B is invertible almost surely. Also, Lemma C.20 gives

$$\sqrt{n}\, \mathcal{P}_{\bar{x}_n \to x_\star} R_{\bar{x}_n}^{-1}(x_\star) = -Z_n + o_p(1).$$

Thus the quadratic form determining membership of $x_\star$ in (33) converges to $Z_0^\mathsf{T} \mathrm{S}_B^{-1} Z_0$. Since $B\mathrm{S}_B = \sum_{b=1}^B Z_b Z_b^\mathsf{T}$, $B\mathrm{S}_B \sim \mathsf{W}_d(\mathrm{V}, B)$, the Wishart law with scale V and B degrees of freedom, and is independent of $Z_0 \sim N_d(0, \mathrm{V})$. The standard normal–Wishart identity gives

$$\frac{B - d + 1}{dB} Z_0^\mathsf{T} \mathrm{S}_B^{-1} Z_0 \sim F_{d, B-d+1}.$$

The limiting inverse-retraction vector lies in the regular local domain and S_B is nonsingular almost surely; hence the finite-sample fallback event has probability tending to zero. The distribution is continuous, so (34) gives (29). □

Proof of Proposition C.1. In Euclidean space, $R_x(v) = x + v$, inverse retraction is subtraction, parallel transport is the identity, and (2) is the arithmetic mean. Equations (1) and (4) therefore become

$$x_k = x_{k-1} - \eta_k \nabla h(x_{k-1}, Z_k), \qquad x_k^{\star(b)} = x_{k-1}^{\star(b)} - \eta_k W_k^{(b)} \nabla h(x_{k-1}^{\star(b)}, Z_k),$$

which are the original and perturbed ASGD paths in COnB. For $\phi_i(x) = x_i$, Equation (6) is the uncentered squared coordinate displacement from the observed averaged path, and (7) uses the t_B critical value. These are exactly the COnB ASGD coordinate interval after matching the update index. The final statement follows from Theorem 3.6 and Corollary 3.7. □

Proof of Corollary C.2. Differentiability at $x_\star$ gives, uniformly for $h \to 0$ in the local chart,

$$\phi\{R_{x_\star}(h)\} - \phi(x_\star) = D\phi_{x_\star} h + o(\|h\|).$$

The two averages are $O_p(n^{-1/2})$ and $O_{\mathbb{P}^\star}(n^{-1/2})$ in fixed coordinates by Theorem 3.4. Subtracting their two expansions therefore yields

$$\sqrt{n}\{\phi(\bar{x}_n^\star) - \phi(\bar{x}_n)\} = D\phi_{x_\star} T_n^\star + o_{\mathbb{P}^\star}(1),$$

where Lemma C.20 aligns the tangent difference. Conditional Slutsky proves (39).

For scalar basic intervals, conditional weak convergence to a continuous nondegenerate Gaussian distribution implies convergence of the relevant conditional quantiles. The empirical bootstrap quantiles with $B_n \to \infty$ converge to the conditional quantiles by the conditional Glivenko–Cantelli theorem. The intrinsic-region statement follows in the same way from continuity of the norm quantile and the isometry of parallel transport. □

Proof of Proposition C.4. On the original-stream event $\{\bar{x}_n \in U_\star\}$, whose probability tends to one, $T_n^\star = \mathcal{P}_{\bar{x}_n \to x_\star} \widetilde{T}_n^\star$ and parallel transport is an isometry. Thus conditional weak convergence and conditional uniform integrability of $\|T_n^\star\|^2$ imply, by conditional truncation,

$$\mathbb{E}^\star T_n^\star \xrightarrow{p} 0, \tag{87}$$

$$\mathbb{E}^\star (T_n^\star \otimes T_n^\star) \xrightarrow{p} V. \tag{88}$$

For completeness, take a bounded continuous radial truncation $h_K(t)$ equal to t on $\|t\| \le K$. Conditional weak convergence gives convergence of the first two moments of $h_K(T_n^\star)$. The conditional uniform-integrability condition makes the truncation error vanish in probability as $K \to \infty$, which proves (87)–(88).

Given the original stream, the B_n bootstrap replicates are i.i.d. from the conditional law. For fixed K, the sample first and second moments of the truncated variables differ from their conditional expectations by $o_p(1)$ because their conditional variances are bounded by constants times K^2/B_n and K^4/B_n. The average tail contribution is controlled by conditional Markov inequality and uniform integrability. Letting first $n \to \infty$ and then $K \to \infty$ shows that the transported empirical first and second moments converge to (87)–(88). Their centered difference is therefore $V + o_p(1)$, proving (41).

If V is positive definite, inversion is continuous in a neighborhood of V. By (77), the transported tangent vector from $\bar{x}_n$ to $x_\star$ has asymptotic law $N(0, V)$ after multiplication by $\sqrt{n}$. Slutsky's theorem yields a χ^2_d limit for the Wald quadratic form and hence the coverage claim. □

C.4.4 Verification of the Primitive Conditions and State Bounds

Proof of Proposition C.7. Choose a radius $a > 0$ within the oracle and comparison domains whose ball contains the eventual localization ball. For every k, define the predictable representatives $\widehat{u}_{k-1} = u_{k-1}\mathbf{1}\{\|u_{k-1}\| < a\}$ and $\widehat{u}^\star_{k-1} = u^\star_{k-1}\mathbf{1}\{\|u^\star_{k-1}\| < a\}$. Define, at every index,

$$m_k = \chi_k(\widehat{u}_{k-1}), \qquad m_k^\star = W_k \chi_k(\widehat{u}^\star_{k-1}) + (W_k - 1) H \widehat{u}^\star_{k-1}.$$

These are martingale differences by (43) and independence of the current weight from the original stream and past weights. The oracle need not be independent of the past. Define $e_k, e_k^\star$ by the exact residual identities

$$e_k = \eta_k^{-1}(u_{k-1} - u_k) - H u_{k-1} - \xi_k - b_k - m_k,$$
$$e_k^\star = \eta_k^{-1}(u^\star_{k-1} - u^\star_k) - H u^\star_{k-1} - W_k \xi_k - W_k b_k - m_k^\star.$$

All arrays are now defined globally, and (11)–(12) hold identically. No martingale array is truncated at a last-exception time or by a current-proposal event.

The conditional moment bound at predictable local arguments gives, for a sufficiently small fixed $r_1 > 0$,

$$\mathbb{P}\{\eta_k \|\psi_k(\widehat{u}_{k-1})\| > r_1 \mid \mathcal{F}_{k-1}\} + \mathbb{P}\{\eta_k |W_k| \left\|\psi_k(\widehat{u}^\star_{k-1})\right\| > r_1 \mid \mathcal{G}_{k-1}\} \le C\eta_k^2.$$

Since $\sum_k \eta_k^2 < \infty$, only finitely many such large proposals occur almost surely. By the assumed eventual chart entry and safeguard inactivity, each path has only finitely many indices at which its representative differs from its actual state or the local update formula fails. These are its exceptional indices, separately for each multiplier path.

At regular indices, Lemma C.21 gives

$$u_k = u_{k-1} - \eta_k \psi_k(u_{k-1}) + \epsilon_k. \tag{89}$$

Define ϵ_k to be this coordinate error at regular indices and zero at exceptional ones. Then, at all indices,

$$\|\epsilon_k\| \le C\{\eta_k \|\widehat{u}_{k-1}\|^2 \|\psi_k(\widehat{u}_{k-1})\| + \eta_k^2 \|\psi_k(\widehat{u}_{k-1})\|^2\}.$$

Define $\epsilon_k^\star$ in the same way for the weighted local update. Its bound is

$$\|\epsilon_k^\star\| \le C\{\eta_k |W_k| \left\|\widehat{u}^\star_{k-1}\right\|^2 \left\|\psi_k(\widehat{u}^\star_{k-1})\right\| + \eta_k^2 W_k^2 \left\|\psi_k(\widehat{u}^\star_{k-1})\right\|^2\}. \tag{90}$$

Substituting (42) into (89) shows that e_k agrees with $r_{\mathrm{pop}}(\widehat{u}_{k-1}) - \eta_k^{-1}\epsilon_k$ at every regular index. Likewise, $e_k^\star$ agrees there with $W_k r_{\mathrm{pop}}(\widehat{u}_{k-1}^\star) - \eta_k^{-1}\epsilon_k^\star$. Their differences at the exceptional indices have an almost surely finite sum, because all actual coordinates and residuals are finite. Dividing that sum by $\sqrt{n}$ gives $o_p(1)$ under the original or joint law. This use of finite exceptions does not require them to form stopping times.

For the martingale variance, (44) gives

$$\mathbb{E}(\|m_k\|^2 \mid \mathcal{F}_{k-1}) \le \omega(\|\widehat{u}_{k-1}\|), \qquad \mathbb{E}(\|m_k^\star\|^2 \mid \mathcal{G}_{k-1}) \le C\{\omega(\|\widehat{u}_{k-1}^\star\|) + \|\widehat{u}_{k-1}^\star\|^2\}.$$

The representatives converge to zero in probability by (16) and satisfy $\|\widehat{u}\| \le \|u\|$. Boundedness of ω and $\omega(t) \to 0$ imply that the expectations of these bounds tend to zero. Cesàro averaging and Markov's inequality establish (14), under the joint law for the weighted path.

For the population drift, Jensen's inequality and the state bound imply

$$\frac{1}{\sqrt{n}}\sum_{k=1}^{n}\mathbb{E}\,\|r_{\mathrm{pop}}(\widehat{u}_{k-1})\| \le \frac{C}{\sqrt{n}}\left\{1 + \sum_{k=2}^{n}\eta_{k-1}^{(1+\rho)/2}\right\} = o(1) \tag{91}$$

by (49). The weighted counterpart follows using $\mathbb{E}|W| < \infty$ and the joint state bound. Conditional Cauchy–Schwarz at the predictable representatives gives $\mathbb{E}\left\|\eta_k^{-1}\epsilon_k\right\| \le C\eta_k$ for $k \ge 2$, and (90) gives $\mathbb{E}\mathbb{E}^\star\left\|\eta_k^{-1}\epsilon_k^\star\right\| \le C\eta_k$. Here only $\mathbb{E}W^2 = 2$ and the conditional oracle second moment are used. These bounds hold without conditioning on a non-stopping regularity event, since the errors are bounded by nonnegative envelopes at all indices. Their root-n sums vanish because $\alpha > 1/2$; together with (91) and the finite exceptional residuals, this proves (15).

The bias conditions are assumed, and (82) controls their multiplier fluctuation. Thus Assumption 3.2 holds. All constructions are separate for each weighted path and retain conditional i.i.d. sampling. The complete averages keep their exceptional prefixes, whose first-order negligibility follows from Lemma C.19. □

Proof of Proposition C.8. Write $\mathcal{V}(x) = d(x, x_\star)^2$. Because K is compact and $\mathcal{V}$ is C^2 on K^+, there are constants $r_0, C > 0$ such that, for $x \in K$ and every tangent vector v with $\|v\|_x \le r_0$,

$$\mathcal{V}\{R_x(-v)\} \le \mathcal{V}(x) + 2\left\langle v, \mathrm{Log}_x(x_\star)\right\rangle_x + C\,\|v\|_x^2\,. \tag{92}$$

Decrease the safeguard radius r_0, if necessary, so that every short retraction proposal remains in K^+. Let $q = 1$ for the base path and $q = W_k$ for a weighted path, and abbreviate $\Psi = \Psi_k(x; D_k)$, $\eta = \eta_k$. Define the short-proposal event $A = \{\eta|q|\,\|\Psi\|_x \le r_0\}$. On A, apply (92) with $v = \eta q \Psi$. On A^c, the safeguarded next state lies in K, so its squared distance is bounded by the squared diameter D_K^2.

We first control the discarded tail in the drift. With $X = |q|\,\|\Psi\|_x$,

$$\mathbb{E}\{X\mathbf{1}(\eta X > r_0) \mid \text{past}\} \le \frac{\eta}{r_0}\mathbb{E}(X^2 \mid \text{past}), \tag{93}$$

and

$$\mathbb{P}(\eta X > r_0 \mid \text{past}) \le \frac{\eta^2}{r_0^2}\mathbb{E}(X^2 \mid \text{past}). \tag{94}$$

For the base path, (53) bounds the conditional second moment of X. For a weighted path, independence and (10) give the same bound multiplied by $\mathbb{E}W^2 = 2$.

Taking conditional expectations in the short-event Taylor inequality, restoring the omitted drift by (93), and controlling the long event by (94) gives, for either path,

$$\mathbb{E}\{\mathcal{V}(x_k) \mid \text{past}\} \le \mathcal{V}(x) + 2\eta\left\langle \bar{G}_k(x), \mathrm{Log}_x(x_\star)\right\rangle_x + C\eta^2, \tag{95}$$

where $\bar{G}_k(x) = \mathbb{E}\{\Psi_k(x; D_k) \mid \mathcal{F}_{k-1}\}$; for the weighted path we used $\mathbb{E}(W_k\Psi_k \mid \text{past}) = \bar{G}_k$. By (51)–(52),

$$\begin{aligned}\left\langle \bar{G}_k(x), \mathrm{Log}_x(x_\star)\right\rangle_x &\le -\mu\mathcal{V}(x) + a_k\sqrt{\mathcal{V}(x)} \\ &\le -\frac{\mu}{2}\mathcal{V}(x) + \frac{a_k^2}{2\mu}.\end{aligned}$$

Since $a_k^2 \le C\eta_k$, (95) becomes

$$\mathbb{E}\{\mathcal{V}(x_k) \mid \text{past}\} \le (1 - c\eta_k)\mathcal{V}(x_{k-1}) + C\eta_k^2 \tag{96}$$

for some $c > 0$. The safeguard does not increase squared distance on short proposals and is already accounted for by the bounded long-event contribution.

Taking expectations, let $A_k = \mathbb{E}\mathcal{V}(x_k)$ for the base path or $A_k = \mathbb{E}\mathbb{E}^\star\mathcal{V}(x_k^\star)$ for the weighted path. Then $A_k \le (1-c\eta_k)A_{k-1} + C\eta_k^2$. Because $\eta_{k-1} - \eta_k = o(\eta_k^2)$ when $\alpha < 1$, a comparison induction proves $A_k \le C_0\eta_k$ for a sufficiently large C_0. Indeed, if $A_{k-1} \le C_0\eta_{k-1}$, then

$$A_k - C_0\eta_k \le C_0(\eta_{k-1} - \eta_k) - cC_0\eta_k\eta_{k-1} + C\eta_k^2,$$

and the negative middle term dominates the first and third terms for large k after C_0 is chosen large enough. This proves (54).

For almost sure inactivity, (96), $\sum \eta_k = \infty$, and $\sum \eta_k^2 < \infty$ allow the Robbins–Siegmund theorem to be applied. It gives convergence of $\mathcal{V}(x_k)$ and $\sum \eta_k \mathcal{V}(x_{k-1}) < \infty$ almost surely, hence $\mathcal{V}(x_k) \to 0$. State convergence puts the path in U_0 eventually. Apply the same tail estimate as (94) at the smaller uniform identity threshold δ_0 from (50):

$$\sum_{k=1}^{\infty} \mathbb{P}(\eta_k|q_k|\,\|\Psi_k\| > \delta_0 \mid \text{past}) < \infty \quad \text{almost surely.}$$

Conditional Borel–Cantelli implies that only finitely many proposals exceed δ_0. Therefore, after a finite random index, the state lies in U_0 and every proposal meets (50); the safeguarded update then equals the original retraction update forever. □

Proof of Lemma C.9. Invariance and induction give $\bar{x}_k \in K$. Put $a_k = d(\bar{x}_k, x_\star)^2$ and $b_k = d(x_k, x_\star)^2$. Then (55) gives

$$a_k \le \left(1 - \frac{\kappa}{k}\right) a_{k-1} + \frac{C}{k} b_k. \tag{97}$$

For fixed $m < k$, let

$$p_{j,k} = \prod_{\ell=j+1}^{k} \left(1 - \frac{\kappa}{\ell}\right).$$

Iteration of (97) yields

$$a_k \le p_{m,k} a_m + C \sum_{j=m+1}^{k} \frac{p_{j,k}}{j} b_j.$$

Here $p_{m,k} \to 0$ and

$$\sum_{j=m+1}^{k} \frac{p_{j,k}}{j} = \frac{1 - p_{m,k}}{\kappa}.$$

Given $\varepsilon > 0$, choose m so that $b_j \le \varepsilon$ for $j \ge m$. It follows that $\limsup_k a_k \le C\varepsilon/\kappa$; letting $\varepsilon \downarrow 0$ proves $\bar{x}_k \to x_\star$.

For the moment bound, take expectations in (97). Since $p_{j,k} \le C_2(j/k)^\kappa$,

$$\mathbb{E}a_k \le p_{1,k}\mathbb{E}a_1 + C_3 \sum_{j=2}^{k} p_{j,k} j^{-1-\alpha} \le C_4 k^{-\kappa} \left(1 + \sum_{j=2}^{k} j^{\kappa-\alpha-1}\right) = O(k^{-\alpha}),$$

where the last equality uses $\kappa > \alpha$. The argument is deterministic before expectations are taken, so replacing $\mathbb{E}$ by $\mathbb{E}\mathbb{E}^\star$ proves the joint-law statement. □

C.4.5 Proofs for the Non-Euclidean Applications

Proof of Corollary C.10. For $f_y(x) = \frac{1}{2}d(x,y)^2$ away from the cut locus, $\operatorname{grad} f_y(x) = -\operatorname{Log}_x(y)$. Hence the target innovation in fixed coordinates is $\xi = -v$, where $v = \operatorname{Log}_{x_\star}(Y)$, and it is centered by (56). Its covariance is $Q_{\mathbb{S}}$ in (58).

The Hessian of f_y at $x_\star$ on the unit sphere is

$$\operatorname{Hess} f_y(x_\star) = e \otimes e + r\cot(r)\{I - e \otimes e\},$$

with the continuous convention at $r = 0$. Taking expectations gives $H_{\mathbb{S}}$. Since $r \le r_0 < \pi/2$, both eigenvalues 1 and $r\cot r$ are strictly positive, so $H_{\mathbb{S}}$ is positive definite.

On a common strongly convex neighborhood containing the support and localized paths, the log map and its derivatives are bounded. The stochastic score is therefore square integrable, locally Lipschitz in L^2, and its population field has a quadratic remainder. The exponential map and parallel transport satisfy the geometric assumptions. The state-rate hypothesis and Proposition C.7 yield the abstract iteration with $\rho = 1$. Applying Theorem 3.4 and using $-H_{\mathbb{S}}^{-1}\xi = H_{\mathbb{S}}^{-1}v$ gives (59). Theorem 3.5 and Corollaries C.2 and 3.7 give the inferential statements. □

Proof of Corollary C.11. The cap K_ρ is strongly geodesically convex, and squared distance from $x_\star$ is geodesically convex on it. Exponential averaging follows the minimizing geodesic segment and therefore satisfies (55) with $\kappa = C = 1$.

For normalized projection, take $x, y \in K_\rho$, write $\theta = d(x,y) \le 2\rho < \pi/2$, and express $y = \cos(\theta)x + \sin(\theta)u$ with $u \perp x$ and $\|u\| = 1$. For

$$R_x^{\mathrm{proj}}(v) = \frac{x+v}{\|x+v\|}, \qquad (R_x^{\mathrm{proj}})^{-1}(y) = \frac{y}{\langle x, y\rangle} - x,$$

direct substitution gives

$$R_x^{\mathrm{proj}}\{t(R_x^{\mathrm{proj}})^{-1}(y)\} = \operatorname{Exp}_x\{s(t,\theta)\operatorname{Log}_x(y)\}, \qquad s(t,\theta) = \frac{\arctan\{t\tan(\theta)\}}{\theta},$$

with the continuous convention at $\theta = 0$. Concavity of $t \mapsto \arctan\{t\tan(\theta)\}$ gives $s(t,\theta) \ge t$, while monotonicity of $\tan(\theta)/\theta$ gives

$$s(t,\theta) \le \frac{\tan(2\rho)}{2\rho}t =: C_\rho t.$$

Moreover, $0 \le s(t,\theta) \le 1$, so the update lies on the same minimizing geodesic segment and remains in K_ρ. Geodesic convexity now yields

$$\begin{aligned} d\{\mathsf{A}_t(x,y), x_\star\}^2 &\le (1-s)d(x,x_\star)^2 + sd(y,x_\star)^2 \\ &\le (1-t)d(x,x_\star)^2 + C_\rho t d(y,x_\star)^2. \end{aligned}$$

Thus Lemma C.9 applies with $\kappa = 1 > \alpha$ to the base path and, under the joint law, separately to every fixed multiplier path. Convergence then gives the path-specific entry times into the comparison chart. □

Proof of Corollary C.12. A Hadamard manifold has no cut locus, so $f_y(x) = \frac{1}{2}d(x,y)^2$ is smooth and $\operatorname{grad} f_y(x) = -\operatorname{Log}_x(y)$ globally. At the minimizer, $\mathbb{E}\operatorname{Log}_{x_\star}(Y) = 0$. Thus the target innovation is $\xi = -\operatorname{Log}_{x_\star}(Y)$ with covariance Q_F, and the derivative of the population field is $H_F = \operatorname{Hess} F(x_\star)$.

Compact support and compact localization make the score, its local derivatives, and the relevant curvature terms bounded. They imply the local L^2 modulus, a quadratic population remainder, and the oracle second-moment envelope. Strong geodesic convexity of squared distance supplies local monotonicity, while the state rate is assumed in the corollary. The primitive theorem therefore applies with $\rho = 1$. Since H_F is self-adjoint, Theorem 3.4 gives (60), and Theorem 3.5 and Corollaries C.2 and 3.7 give conditional and fixed-replication inference.

For the positive-definite specialization, formulas (61)–(62) are the affine-invariant exponential and logarithm. An eigenvalue-bounded set is compact in finite dimension and stays a positive distance from the boundary, so the preceding local bounds apply. Finally, the usual matrix differential gives $D\log\det(\mathrm{X})[\mathrm{U}] = \mathrm{tr}(\mathrm{X}^{-1}\mathrm{U})$, proving (63). □

Proof of Corollary C.13. Affine congruence by $\mathrm{A}^{-1/2}$ is an isometry, and (62) gives

$$\mathrm{A}^{-1/2}\,\mathrm{Log}_{\mathrm{A}}(\mathrm{Y})\mathrm{A}^{-1/2} = \mathrm{E}, \qquad d(\mathrm{A},\mathrm{Y}) = \|\mathrm{E}\|_F\,.$$

Sign symmetry therefore centers the mean score at A. Finite second distance moment gives existence, and strict geodesic convexity on the Hadamard manifold gives uniqueness. On any compact set of states, Hessian comparison for the affine-invariant metric bounds the transported score and its local derivative by $C_K(1 + \|\mathrm{E}\|_F)$. The assumed $2 + \delta$ moment consequently supplies the score second moment, local L^2 modulus, and oracle envelope directly, without compact support for Y. The quantitative population remainder and its step-size threshold are separate hypotheses of the corollary; with them, the stated localization, and the state rates, Proposition C.7 applies.

For the Gaussian contamination model in the numerical experiments, the map $s_{\mathrm{A}}(\mathrm{Y}) = \mathrm{A}\mathrm{Y}^{-1}\mathrm{A}$ is an affine-invariant isometry and sends E to $-\mathrm{E}$. The independent symmetric contamination preserves this invariance and finite distance moments. The nondegenerate Gaussian component gives a positive density on $\mathcal{P}_m$, so the law is not supported on a geodesic. Convexity of distance on this Hadamard manifold is then strict after expectation along every nonconstant geodesic, giving a unique median. Isometry invariance fixes both unique minimizers under s_{A}, whose only positive-definite fixed point is A. This establishes the common target for the simulated laws; the path and regularity conditions for inference remain those stated in the corollary.

For the median, the normalized score at A is $-\mathrm{E}/\|\mathrm{E}\|_F$ under the same isometry, with any bounded convention at $\mathrm{E} = 0$. Its expectation is zero by sign symmetry. The supremum-envelope argument in the proof of Corollary C.14 supplies the local L^2 modulus. Together with the median-specific localization and state-rate assumptions and its stated population regularity, this verifies the hypotheses of Proposition C.7. The mean and median conclusions then follow from Theorem 3.5 and Corollaries C.2 and 3.7. □

Proof of Corollary C.14. At $x_\star$, the target innovation is

$$\xi = \Psi(x_\star; Y) = -\frac{\mathrm{Log}_{x_\star}(Y)}{d(x_\star, Y)}.$$

The no-atom condition makes this well defined almost surely, and the first-order condition for the unique differentiable population median gives $\mathbb{E}\xi = 0$. The innovation has norm one and covariance Q_{med}. Hence the innovation covariance and Lindeberg conditions hold without a moment assumption beyond measurability.

It remains to verify the local L^2 modulus despite samplewise nonsmoothness. Bound the equality convention for $\Psi(x; y)$ by $M \geq 1$, enlarging the bound if necessary. On a separable compact local ball, define the measurable supremum envelope

$$G_t(y) = \sup_{d(x,x_\star)\leq t} \|\mathcal{P}_{x\to x_\star}\Psi(x; y) - \Psi(x_\star; y)\|^2\,;$$

the supremum may equivalently be taken over a countable dense subset, with equality points included separately. For every fixed $y \neq x_\star$, eventually $t < d(y, x_\star)/2$; smoothness away from the diagonal and the shrinking compact ball then give $G_t(y) \to 0$. Uniformly, $G_t(y) \leq (M + 1)^2$. Since $\mathbb{P}(Y = x_\star) = 0$, dominated convergence yields $\mathbb{E}G_t(Y) \to 0$, and hence

$$\sup_{d(x,x_\star)\leq t} \mathbb{E}\,\|\mathcal{P}_{x\to x_\star}\Psi(x; Y) - \Psi(x_\star; Y)\|^2 \leq \mathbb{E}G_t(Y) \longrightarrow 0.$$

Centering the increment changes the bound by at most a constant factor, so (44) holds for both original and weighted predictable arguments. Non-atomicity on the localized neighborhood excludes their equality events. The population $C^{1,\rho}$ assumption gives (45), and bounded score norm gives the oracle second-moment condition. The stated inequality $\alpha(1 + \rho) > 1$, together with the state-rate and geometry assumptions, now allows Proposition C.7 to be applied. Theorem 3.4 yields (65), and the bootstrap conclusions follow from Theorem 3.5 and Corollaries C.2 and 3.7. □

Proof of Proposition C.15. The density is smooth, bounded, compactly supported, and positive near A. The affine-invariant symmetry $s_{\mathrm{A}}(\mathrm{y}) = \mathrm{A}\mathrm{y}^{-1}\mathrm{A}$ is an isometry preserving this density and Riemannian volume, and reverses $\mathrm{Log}_{\mathrm{A}}(\mathrm{y})$. Thus the population normalized score vanishes at A.

Write $r_y(x) = d(x, y)$. On compact state and observation sets, smooth normal coordinates near the diagonal and smoothness away from it give

$$\left\|\nabla^2 r_y(x)\right\| \le C(1 + r_y(x)^{-1}), \qquad \left\|\nabla_x \nabla^2 r_y(x)\right\| \le C(1 + r_y(x)^{-2}).$$

These follow by differentiating the local distance expression: its first, second, and third derivatives have orders 1, r^{-1}, and r^{-2}, respectively; bounded metric derivatives on a compact enlargement control the remaining terms. Uniform volume comparison on that enlargement bounds the integral of r^{-2} over a ball of radius ε by $C \int_0^\varepsilon r^{d-3}\, dr$. Since $d \ge 3$ and the density is bounded, this tends to zero uniformly in the state. Differentiation of the expected score is justified by first excluding such a ball and then taking its radius to zero. Applying the fundamental theorem of calculus to the Hessian along state geodesics and Fubini's theorem gives a locally Lipschitz population derivative. Observation points lying on a fixed state geodesic form a volume-null set and do not affect that argument.

Hadamard Hessian comparison gives, for $y \ne x$,

$$\operatorname{Hess} r_y(x)[v, v] \ge \frac{\|v\|_x^2 - \langle v, e_{x,y}\rangle_x^2}{d(x, y)}, \qquad e_{x,y} = \operatorname{Log}_x(y)/d(x, y).$$

The density is positive on an open set, so the expectation is strictly positive for every nonzero v. Continuity and compactness of the unit tangent bundle over K give a uniform positive lower bound there. The expected distance is strictly geodesically convex; its stationary point A is the unique minimizer. Integrating its Hessian along the geodesic from A to any $x \in K$ yields (51). The same open support makes the target normalized-score covariance positive definite.

The sample score is bounded by one, has zero approximation bias, and has a local L^2 modulus by the dominated-convergence argument in the proof of Corollary C.14. The population remainder has $\rho = 1$. The projection safeguard satisfies every requirement of Proposition C.8, which therefore proves the state rate and almost sure convergence. Exponential averaging satisfies Lemma C.9 with $\kappa = C = 1$. On a Hadamard manifold all logarithms exist, compactness gives finite coordinates, and both states and full averages eventually enter every comparison neighborhood of A. Exponential-map geometry and Proposition C.7 now establish the claimed bootstrap conclusions. □

Proof of Corollary C.16. Work in the standard horizontal tangent representation at $\mathrm{U}_\star$. A tangent perturbation has the form $\mathrm{U}_\perp \mathrm{K}$. The population Riemannian gradient of (66) is $G(\mathrm{U}) = -(\mathrm{I} - \mathrm{U}\mathrm{U}^\mathsf{T})\Sigma\mathrm{U}$. Differentiating at $\mathrm{U}_\star$ in direction $\mathrm{U}_\perp \mathrm{K}$ and using $\Sigma\mathrm{U}_\star = \mathrm{U}_\star \Lambda_1$ and $\Sigma\mathrm{U}_\perp = \mathrm{U}_\perp \Lambda_2$ gives

$$DG(\mathrm{U}_\star)[\mathrm{U}_\perp \mathrm{K}] = \mathrm{U}_\perp(\mathrm{K}\Lambda_1 - \Lambda_2 \mathrm{K}) = \mathrm{U}_\perp \mathcal{H}(\mathrm{K}).$$

Every tangent eigenvalue is $\lambda_j - \lambda_{r+a} > 0$, so $\mathcal{H}$ is positive definite.

At the target,

$$\Psi(\mathrm{U}_\star; Y) = -(\mathrm{I} - \mathrm{U}_\star \mathrm{U}_\star^\mathsf{T}) Y Y^\mathsf{T} \mathrm{U}_\star = -\mathrm{U}_\perp Y_2 Y_1^\mathsf{T}.$$

The vectorized coordinate innovation is therefore $\xi = -\operatorname{vec}(Y_2 Y_1^\mathsf{T})$. It is centered because $\mathrm{U}_\perp^\mathsf{T} \Sigma \mathrm{U}_\star = 0$, and its covariance is $\mathrm{Q}_{\mathrm{PCA}}$ in (68). The score obeys $\|\Psi(\mathrm{U}; Y)\| \le \|Y\|^2$, and its local Lipschitz envelope is bounded by $C\|Y\|^2$. Thus $\mathbb{E}\|Y\|^4 < \infty$ gives both the innovation second moment and the local L^2 modulus. The population field is smooth on the compact Grassmann manifold and has a quadratic remainder. The state-rate and second-order retraction assumptions complete the hypotheses of Proposition C.7. Equation (69) follows from Theorem 3.4; the bootstrap conclusions follow from Theorem 3.5 and Corollaries C.2 and 3.7.

If Y is Gaussian in the eigenbasis, Y_1 and Y_2 are independent with diagonal covariance matrices Λ_1 and Λ_2. Hence

$$\operatorname{Cov}(Y_{2,a} Y_{1,j}, Y_{2,b} Y_{1,\ell}) = \lambda_{r+a} \lambda_j \mathbf{1}(a = b) \mathbf{1}(j = \ell).$$

The inverse of (67) divides coordinate (a, j) by $\lambda_j - \lambda_{r+a}$, proving (70). For $P(\mathrm{U}) = \mathrm{U}\mathrm{U}^\mathsf{T}$, $dP[\mathrm{U}_\perp \mathrm{K}] = \mathrm{U}_\perp \mathrm{K} \mathrm{U}_\star^\mathsf{T} + \mathrm{U}_\star \mathrm{K}^\mathsf{T} \mathrm{U}_\perp^\mathsf{T}$. Taking the trace against symmetric A gives (71). □

C.4.6 Proof of the Approximation-Bias Proposition

Proof of Proposition 3.8. Whenever the coordinate Abel rearrangement is valid, its nonnegligible-bias term is

$$-H^{-1}\frac{1}{\sqrt{n}}\sum_{k=1}^{n} b_k. \tag{98}$$

For $0 < \gamma < 1$,

$$\sum_{k=1}^{n} k^{-\gamma} = \frac{n^{1-\gamma}}{1-\gamma}\{1 + o(1)\},$$

while $n^{-1/2}\sum_{k=1}^{n} k^{-1/2} \to 2$. If $\gamma = 1$, the sum is $O(\log n)$, and if $\gamma > 1$, it is bounded. Therefore the leading part of (98) tends to zero when $\gamma > 1/2$, tends to $-2h_0^r H^{-1} B_0$ when $\gamma = 1/2$, and equals (32) when $0 < \gamma < 1/2$. The remainder $\widetilde{b}_k$ is negligible by assumption. In the first two regimes the other abstract state and remainder conditions were assumed, so this Abel term combines with Lemmas C.18, C.20 and C.27 as in the main theorem; at the boundary this gives (31). In the subcritical regime the calculation is only conditional on a separate bias-aware justification of the Abel identity and does not invoke those two geometric lemmas.

The bootstrap difference contains the centered bias term

$$-H^{-1}\frac{1}{\sqrt{n}}\sum_{k=1}^{n}(W_k - 1)b_k. \tag{99}$$

Its conditional second moment is bounded by a constant times $n^{-1}\sum_{k=1}^{n}\|b_k\|^2$. For the leading term in (30), this average is $O(n^{-2\gamma})$ if $0 < \gamma < 1/2$, $O((\log n)/n)$ if $\gamma = 1/2$, and $O(n^{-1})$ if $\gamma > 1/2$. The squared-remainder condition and Cauchy–Schwarz control the remaining terms. Hence (99) is $o_{\mathbb{P}^\star}(1)$ for every $\gamma > 0$. For $\gamma \geq 1/2$, the remaining terms in the random-base difference are those in the proof of Theorem 3.5; Lemma C.28 therefore gives (25), including the centered conditional law at the boundary. This proves the fixed boundary statement under the abstract conditions and the stated coordinate-level diagnostic below the boundary; it does not turn the latter into a manifold-bootstrap theorem. □