BEYOND MEAN CURVATURE: PATHWISE LANGEVIN SAMPLING ON LEARNED DATA MANIFOLDS

Anonymous authors
Paper under double-blind review

ABSTRACT

Underdetermined inverse problems admit multiple plausible solutions, yet learned priors entangle data-manifold geometry with training-density preferences, hindering controlled exploration. We propose *geometric-prior exploration*: the pretrained diffusion model defines the manifold of plausible solutions, while task potential and inverse temperature control density relative to manifold volume. To sample this target when the manifold is known only through a pretrained diffusion model, we introduce $\mathrm{II}L$, a manifold Langevin discretization combining soft landing with a *pathwise curvature correction* based on the second fundamental form while avoiding explicit computation of the mean-curvature vector. We construct the required directional-curvature estimator directly from diffusion scores without forming a tangent basis or the full curvature tensor. Under inexact geometric access, we establish stationary weak-bias guarantees. Our pathwise correction improves the deviation of the stationary distribution from the manifold from $O(\sqrt{\Delta t})$ under the standard mean-curvature correction to $O(\Delta t)$, where Δt is the discretization step size. Experiments on synthetic manifolds, image inpainting, human-pose editing, and image inverse problems validate the predicted numerical behavior, demonstrate temperature-controlled exploration of task-compatible solutions, and improve reconstruction refinement at comparable runtime.

1 INTRODUCTION

An inverse problem seeks to recover an unknown $x^\star$ from noisy observations y

$$y = \mathcal{A}(x^\star) + \eta, \tag{1}$$

where the potentially nonlinear measurement operator $\mathcal{A}$ is assumed known and η denotes a Gaussian noise. We focus on the underdetermined setting, in which the observations alone do not identify a unique candidate. Prior information is therefore typically introduced to distinguish plausible candidates from other observation-compatible ones.

A natural source of prior information about plausibility is the geometry of the data distribution: under the manifold hypothesis, plausible samples lie on a low-dimensional manifold $\mathcal{M} \subset \mathbb{R}^d$ (Fefferman et al., 2016). A standard way to encode prior information from data is through a *learned prior* ν_{data}, represented by a pretrained generative model (Song et al., 2021; Mardani et al., 2024).

Define the task potential $U(x) := \frac{1}{2}\|y - \mathcal{A}(x)\|^2$ and introduce an inverse temperature $\beta > 0$. Let $\mathrm{vol}_{\mathcal{M}}$ denote the Riemannian volume measure on $\mathcal{M}$ induced by the ambient Euclidean metric. Assuming that ν_{data} is absolutely continuous with respect to $\mathrm{vol}_{\mathcal{M}}$, define its density by $f_{\text{data}}(x) := \nu_{\text{data}}(\mathrm{d}x)/\,\mathrm{vol}_{\mathcal{M}}(\mathrm{d}x)$. For the inverse problem (1), the target posterior with a learned-prior ν_{data} is

$$q(\mathrm{d}x \mid y) \propto \exp(-\beta U(x))\,\nu_{\text{data}}(\mathrm{d}x) = \exp(-\beta U(x))\,f_{\text{data}}(x)\,\mathrm{vol}_{\mathcal{M}}(\mathrm{d}x). \tag{2}$$

Diffusion Posterior Sampling (DPS) (Chung et al., 2023) is motivated as an approximate sampler for this posterior; more broadly, this learned-prior perspective underlies many diffusion-based inverse-problem solvers (Kawar et al., 2022; Wang et al., 2022; Zhang et al., 2025).

Even after incorporating such prior information, multiple candidates may remain both plausible and compatible with the same observations. We refer to variation among these remaining candidates as *solution diversity*. This residual non-uniqueness need not be a defect: in many applications, access

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.

to alternative valid solutions is itself valuable. For example, in image inpainting, users may compare several realistic completions consistent with the observed pixels (Guillemot & Le Meur, 2013). In pose editing, an artist may inspect multiple natural-looking body configurations satisfying the same hand-position constraints before selecting one to refine (Ci et al., 2023; Agrawal et al., 2023). Our goal is therefore not to eliminate non-uniqueness altogether, but to control the exploration of these plausible, task-compatible alternatives.

Posterior-sampling methods such as DPS can already generate multiple plausible solutions, but the frequency with which they produce these solutions is influenced by their relative frequencies in the training data. These learned frequencies need not reflect desirable preferences for a particular test-time task. Under the learned-prior posterior in equation 2, two regions that are equally compatible with the observation can receive very different probability density simply because one was more frequent under the training distribution. Adjusting the temperature changes the strength of task conditioning, but does not remove this learned-density weighting. We instead seek an exploration distribution in which the data model determines the set and geometry of plausible solutions, while their relative density per unit manifold volume is determined *only* by the task potential. This is useful, for example, when two natural-looking pose families satisfy the same editing constraints but have substantially different frequencies under the learned prior.

Geometric-prior based exploration. We consider an alternative target distribution that retains the learned data manifold $\mathcal{M}$ but removes the learned density f_{data} appearing in equation 2. Specifically, we adopt the manifold volume measure $\text{vol}_{\mathcal{M}}$ as a geometric prior and define

$$\mu(\mathrm{d}x) \propto \exp\big(-\beta U(x)\big)\ \text{vol}_{\mathcal{M}}(\mathrm{d}x), \qquad \beta > 0. \quad (3)$$

Consequently, the corresponding sampler no longer automatically favors solutions that were more common in the training data, but treats solutions with the same observation compatibility on more equal footing. Figure 1 illustrates the resulting distinction: changing the training density changes the learned-prior posterior through f_{data}, while leaving the ideal geometric-prior target unchanged when the underlying manifold and task potential are held fixed. From a technical perspective, this formulation separates the roles of geometry, task, and temperature: $\mathcal{M}$ specifies where exploration takes place, U weights alternatives relative to manifold volume, and β modulates the concentration of that weighting. Moreover, the target (3) approaches manifold volume as $\beta \to 0$ and increasingly concentrates near the global minimizers of U as $\beta \to \infty$.

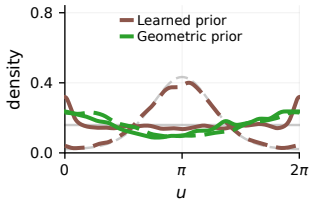
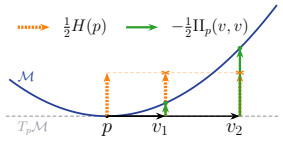


Figure 1: Sampled posterior densities under learned and geometric priors for two different training distributions (gray; matching line styles).

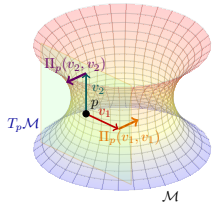
Having specified the desired Gibbs target in equation 3, the remaining challenge is how to sample from it. A natural continuous-time sampler for this Gibbs distribution is manifold Langevin dynamics (Hsu, 2002; Zhang, 2020; Bharath et al., 2025). In ambient Itô form, it reads

$$\begin{aligned} dX_t = &\big[-P(X_t)\nabla U(X_t) + \beta^{-1} H(X_t)\big]\, dt \\ &+ \sqrt{2\beta^{-1}}\, P(X_t)\, dW_t, \qquad X_0 \in \mathcal{M}, \end{aligned} \quad (4)$$

where $P(x)$ is the orthogonal projector onto $T_x\mathcal{M}$, W_t is Brownian motion in $\mathbb{R}^d$, and $H(x) = -\sum_{i=1}^{\dim \mathcal{M}} \text{II}_x(e_i, e_i)$ is the mean-curvature vector, with II_x the second fundamental form and $\{e_i\}$ an orthonormal tangent basis.



(a) Magnitude-dependent correction: on a 1-D parabola.



(b) Direction-dependent correction on a catenoid ($H \equiv 0$).

Figure 2: Curvature correction via the mean-curvature vector H and the second fundamental form II.

Discrete-time sampling dynamics over data manifold. In practice, the geometry of a data manifold is *unknown*, creating two challenges. First, we need geometric access to the manifold. Prior work provides score-derived approximations to the nearest projection and tangent projection (Kharitenko et al., 2025); building on these operators, we construct score-based estimators of the mean-curvature vector H and the directional second fundamental form II.

Second, these learned geometric operators are necessarily inexact, so the continuous-time dynamics in equation 4 must be discretized under inexact geometry. To understand the resulting difficulty, we first recall that, with exact geometric access, a standard Euler–Maruyama discretization is

$$X_{k+1} = X_k + \Delta t\big(-P(p_k)\nabla U(p_k) + \beta^{-1}H(p_k)\big) + \sqrt{2\beta^{-1}\Delta t}\, v_k, \tag{5}$$

where Δt is the step size, $p_k = \pi(X_k)$ with nearest projection $\pi = \pi_{\mathcal{M}}$, and $v_k = P(p_k)\xi_k$, with $\xi_k \sim \mathcal{N}(0, I_d)$, is tangent-projected Gaussian noise.

Due to the curvature of the manifold $\mathcal{M}$, trajectories of (5) generally drift away from $\mathcal{M}$. When exact projection is available, a standard remedy is to project back onto the manifold after every stochastic Euler step (Zhang, 2020; Sharma & Zhang, 2021; Bharath et al., 2025). With only learned geometric access, however, the exact projection must be replaced by an approximation $\hat{\pi}$. The resulting projection error is incurred at every step and can therefore accumulate over the iterations. In fact, we show that *decreasing the step size can counterintuitively increase the stationary sampling bias*. Example 1 makes this mechanism explicit by exhibiting an inverse-step-size amplification of the projection error. This motivates avoiding repeated hard projection. Instead, we adopt the soft-landing approach of Jeon et al. (2026), which pulls the iterate toward $\mathcal{M}$ through $\ell\Delta t(p_k - X_k)$.

The mean-curvature term H in equation 5 also deserves closer attention. In continuous time, H gives the correct Itô drift. For a finite tangent displacement v, however, the leading-order deviation of $p + v$ from $\mathcal{M}$ is governed by the direction-specific quantity $\mathrm{II}_p(v, v)$; see equation 7 and equation 11. By contrast, H averages curvature over all tangent directions and therefore need not capture the geometric effect of an individual stochastic update. Figure 2 illustrates both sources of mismatch: magnitude dependence on a one-dimensional parabola and direction dependence on a catenoid, where directional curvatures are nonzero even though $H \equiv 0$. Example 3 gives an explicit finite-step instance on the circle, showing that the mean-curvature correction does not in general reproduce the pathwise projection correction. This motivates a pathwise finite-step correction based directly on the second fundamental form II.

Based on the above considerations, we propose the IIL-scheme, combining pathwise curvature correction with soft landing for robust discretization for manifold Langevin under learned geometry:

$$X_{k+1} = X_k \underbrace{-\Delta t\, P(p_k)\nabla U(p_k) + \sqrt{2\beta^{-1}\Delta t}\, v_k}_{\text{tangent Euler step}} \underbrace{- \beta^{-1}\Delta t\, \mathrm{II}_{p_k}(v_k, v_k)}_{\text{pathwise curvature correction}} \underbrace{+ \ell\Delta t\,(p_k - X_k)}_{\text{landing term}}. \tag{6}$$

Note that the update of IIL avoids the computation of the mean curvature H, which is a computational advantage over the standard Euler scheme (5).

Contributions. We develop a practical discretization of geometric-prior exploration on unknown data manifolds and establish quantitative guarantees on its stationary accuracy:

- **Geometric-prior exploration.** We formulate inverse problems through a geometric prior that separates the data-manifold geometry, task-dependent density, and temperature-controlled concentration. This separation provides explicit control over the target distribution being explored.
- **A discretization tailored to learned geometry.** We introduce the IIL scheme for manifold Langevin discretization, combining soft landing with a directional curvature correction adapted to the realized stochastic increment. We construct the required directional-curvature estimator from a pre-trained diffusion score and bound its error, without forming a tangent basis, the full second fundamental form, or an explicit mean-curvature trace.
- **Stationary error control under inexact geometry.** Building on the generator formulation of Stein's method (Mattingly et al., 2010; Bharath et al., 2025), we develop a framework that decomposes stationary bias into off-manifold deviation and on-manifold sampling error. Under explicit regularity and geometric-accuracy assumptions, we show that IIL, which uses pathwise curvature correction based on the second fundamental form, improves deviation of the stationary distribution from $O(\sqrt{\Delta t})$ for mean-curvature correction to $O(\Delta t)$.

Experiments evaluate both the numerical design of IIL and its usefulness for controlled exploration. Synthetic manifolds test the predicted projection-error and pathwise-curvature mechanisms. MNIST inpainting examines diverse completions initialized from a selected reference, while source-initialized Human3.6M pose editing evaluates temperature-controlled exploration in unconstrained

joints. At comparable runtime, IIL provides greater pose diversity than DPS (Chung et al., 2023) at matched editing error over most of the evaluated range. Beyond exploration, terminal IIL refinement within Decoupled Annealing Posterior Sampling (DAPS) (Zhang et al., 2025) improves reconstruction on all eight FFHQ-256 tasks over native DAPS hosts at comparable measured runtime. Together, these studies demonstrate the usefulness of learned-manifold dynamics for exploring task-compatible alternatives and refining existing reconstructions.

1.1 Related Work

Learned and Geometric Priors. Many inverse solvers use pretrained models as distributional priors or learned regularizers. DPS/DAPS pursue posterior sampling (Chung et al., 2023; Zhang et al., 2025), while DDRM/DDNM exploit linear measurement structure (Kawar et al., 2022; Wang et al., 2022). PnP-ULA uses denoiser-based Bayesian sampling (Laumont et al., 2022), RED-diff and DPoser/DPoser-X use variational inference (Mardani et al., 2024; Lu et al., 2023; 2025), and GFPose supports conditional pose generation (Ci et al., 2023). We instead extract manifold geometry from the pretrained model and adopt manifold volume as a *geometric prior*, with target density determined by the task potential and inverse temperature. A closely related work is Li et al. (2026), who study tempered-score Langevin for recovering manifold-volume targets in the small-noise limit.

Manifold Langevin discretizations. Classical manifold Langevin discretizations are generally formulated under exact geometric access. Intrinsic schemes map tangent stochastic increments through exponential maps or retractions (Cheng et al., 2022; Schwarz et al., 2025; Bharath et al., 2025; Zhan & Sugiyama, 2026). Extrinsic schemes discretize the embedded SDE using either projection, sometimes combined with mean-curvature drift (Zhang, 2020; Yu et al., 2023), or normal relaxation (Zhang et al., 2022; Jeon et al., 2026). Although some consider numerical approximations (Armstrong & King, 2022; Sharma & Zhang, 2021) or data-driven geometric estimates (Kawasaki-Borruat et al., 2026), these works still build on the above idealized discretizations.

Geometry from diffusion scores. A growing line of work investigates data-manifold geometry through diffusion models and related score-based constructions, including manifold support, intrinsic dimension, and local geometric structure (Pidstrigach, 2022; Permenter & Yuan, 2023; Stanczuk et al., 2024; Kamkari et al., 2024; Tang & Yang, 2024; Diepeveen et al., 2024; Ventura et al., 2025; Li et al., 2026; Azeglio & Di Bernardo, 2025). Complementary approaches estimate curvature from point clouds using diffusion-generated samples (Wang et al., 2025) or diffusion-map operators (Jones, 2026). Closest to our setting, Kharitenko et al. (2025) use score-derived denoisers and their Jacobians to approximate nearest and tangent projections, enabling deterministic manifold optimization. We extend this score-based geometric access to a *directional curvature oracle*.

2 Preliminaries and Notation

Let $\mathcal{M} \subset \mathbb{R}^d$ be a compact, smooth, embedded manifold without boundary of intrinsic dimension r, with tangent space $T_p\mathcal{M}$ and distance $\mathsf{d}(x) := \operatorname{dist}(x, \mathcal{M})$. Define the tubular neighborhoods $\mathcal{T}_\rho := \{x : \mathsf{d}(x) < \rho\}$ and $\mathcal{K}_\rho := \{x : \mathsf{d}(x) \le \rho\}$. Under Assumption 5, there exists $\tau > 0$ such that the nearest projection $\pi : \mathcal{T}_\tau \to \mathcal{M}$ is uniquely defined (Lee, 2013). For $p \in \mathcal{M}$, the tangent and normal projectors are $P(p) = D\pi(p)$ and $N(p) = I_d - P(p)$. For $u, w \in T_p\mathcal{M}$, define

$$\mathrm{II}_p(u, w) := -N(p)DP(p)[u]w, \qquad H(p) := -\sum\nolimits_{j=1}^{r} \mathrm{II}_p(e_j, e_j), \tag{7}$$

where $\{e_j\}_{j=1}^r$ is any orthonormal basis of $T_p\mathcal{M}$. Thus $\mathrm{II}_p(v, v)$ gives directional normal curvature, while H averages over tangent directions. A core identity is that, for $v \sim \mathcal{N}(0, I_{T_p\mathcal{M}})$,

$$H(p) = -\sum\nolimits_{j=1}^{r} \mathrm{II}_p(e_j, e_j) = -\sum\nolimits_{i,j=1}^{r} \mathbb{E} v_i v_j \mathrm{II}_p(e_i, e_j) = -\mathbb{E}\mathrm{II}_p(v, v).$$

We assume access only to approximations $\widehat{\pi}$, $\widehat{P}$, $\widehat{\mathrm{II}}$, and $\widehat{H}$ satisfying the following bounds.

Assumption 1 (Inexact geometric access)**.** For a compact tube $\mathcal{K}_{\tau_{\mathrm{or}}} \Subset \mathcal{T}_\tau$ it holds uniformly for $x \in \mathcal{K}_{\tau_{\mathrm{or}}}$ that

$$\begin{aligned} &\|\widehat{\pi}(x) - \pi(x)\| \le \varepsilon_\pi, && \|\widehat{P}(\widehat{p}) - P(p)\|_{\mathrm{op}} \le \varepsilon_P, \\ &\|\widehat{H}(\widehat{p}) - H(p)\| \le \varepsilon_H, && \left(\mathbb{E}_{\xi \sim \mathcal{N}(0, I_d)} \|\widehat{\mathrm{II}}_{\widehat{p}}(\widehat{v}, \widehat{v}) - \mathrm{II}_p(v, v)\|^2\right)^{1/2} \le \varepsilon_{\mathrm{II}}, \end{aligned} \tag{8}$$

for some tolerances $\varepsilon_\pi, \varepsilon_P, \varepsilon_{\mathrm{II}}, \varepsilon_H > 0$, where $p = \pi(x)$, $\widehat{p} = \widehat{\pi}(x)$, $v = P(p)\xi$, and $\widehat{v} = \widehat{P}(\widehat{p})\xi$.

2.1 Hard projection and projection-bias amplification

A direct Euler–Maruyama step in equation 5 generally leaves $\mathcal{M}$. The π-scheme[1] (Zhang, 2020; Sharma & Zhang, 2021; Bharath et al., 2025),

$$X_{k+1}^{\pi} = \pi\Big(X_k - \Delta t\, P(p_k)\nabla U(p_k) + \sqrt{2\beta^{-1}\Delta t}\, v_k\Big), \tag{9}$$

projects every update back to the manifold. With a learned projection, however, its persistent bias is injected at every step and can be amplified as Δt decreases as the following example shows.

Example 1 (Projection-bias amplification)**.** *Let $\mathcal{M} = \mathbb{R}\times\{0\}$, $U(x) = x_1^2/2$, and $\beta = 1$, so that the target is $\mu = \mathcal{N}(0,1)\otimes\delta_0$. Suppose only the final projection is perturbed as $\widehat{\pi}(x) = (x_1 + \epsilon_\parallel, \epsilon_\perp)$,. For $0 < \Delta t < 1$, the π-scheme has stationary law and Wasserstein W_2-distance*

$$\mu_{\Delta t}^{\pi} = \mathcal{N}\left(\frac{\epsilon_\parallel}{\Delta t}, \frac{2}{2-\Delta t}\right)\otimes\delta_{\epsilon_\perp}, \quad W_2(\mu,\mu_{\Delta t}^{\pi})^2 = \frac{\epsilon_\parallel^2}{\Delta t^2} + \left(\sqrt{\frac{2}{2-\Delta t}} - 1\right)^2 + \epsilon_\perp^2,$$

for fixed $\epsilon_\parallel \neq 0$, the stationary error grows without bound as $\Delta t \to 0$.

This motivates replacing unscaled hard projection by a step-size-scaled correction.

2.2 Mean-curvature landing and finite-step curvature error

Mean-curvature landing (HL) (Jeon et al., 2026) avoids the final hard projection and instead adds a soft landing term $\ell(p_k - X_k)$:

$$X_{k+1}^{HL} = X_k - \Delta t\, P(p_k)\nabla U(p_k) + \sqrt{2\beta^{-1}\Delta t}\, v_k + \beta^{-1}\Delta t\, H(p_k) + \ell\Delta t\,(p_k - X_k), \tag{10}$$

where $\ell > 0$ is the landing strength.

Example 2 (Projection bias versus landing, continued)**.** *In the setting of Example 1, using the same biased projection in the landing term gives*

$$\mu_{\Delta t}^{HL} = \mathcal{N}\left(\ell\epsilon_\parallel, \frac{2}{2-\Delta t}\right)\otimes\delta_{\epsilon_\perp}, \quad W_2(\mu,\mu_{\Delta t}^{HL})^2 = \ell^2\epsilon_\parallel^2 + \left(\sqrt{\frac{2}{2-\Delta t}} - 1\right)^2 + \epsilon_\perp^2,$$

so fixed ℓ avoids inverse-step-size amplification, although geometric bias remains.

However, HL introduces a different finite-step issue even with exact geometry. Let $v = P(p)\xi \in T_p\mathcal{M}$ with $p \in \mathcal{M}$ and $p + v \in \mathcal{T}_\tau$, the local projection expansion is (Moser, 2005, Section 3.1)

$$\pi(p+v) = p + v - \tfrac{1}{2}\mathrm{II}_p(v,v) + \mathcal{O}(\|v\|^3), \quad \mathbb{E}_{\xi\sim\mathcal{N}(0,I_d)}[-\mathrm{II}_p(P(p)\xi, P(p)\xi)] = H(p)\,. \tag{11}$$

Thus the leading correction depends on the pathwise realized tangent increment, whereas HL uses only its directional *averaged* $H(p)$ irrespective of the particular v chosen during a realization.

Example 3 (Mean versus realized curvature on the circle)**.** *Let $\mathcal{M} = S^1$, $U \equiv 0$, and $\beta = \ell = 1$. At $p \in S^1$, write $v = \xi e$ with unit tangent $e = p^\perp$ and $\xi \sim \mathcal{N}(0,1)$. Then $H(p) = -p$ and*

$$H(p) = -p, \qquad -\mathrm{II}_p(v,v) = \xi^2 H(p).$$

The pathwise correction retains the sampled ξ^2 and removes the centered order-Δt fluctuation. For stationary laws $\mu_{\Delta t}^{HL}$ and $\mu_{\Delta t}^{\mathrm{IIL}}$,

$$\mathbb{E}_{X\sim\mu_{\Delta t}^{HL}}\big|\|X\| - 1\big| = \sqrt{\Delta t/\pi} + o(\sqrt{\Delta t}), \qquad \mathbb{E}_{X\sim\mu_{\Delta t}^{\mathrm{IIL}}}\big|\|X\| - 1\big| = \tfrac{3}{4}\Delta t + O(\Delta t^2)\,.$$

Thus, the realized-curvature correction keeps the stationary mean distance to the manifold at order $O(\Delta t)$, whereas mean-curvature landing yields an $O(\sqrt{\Delta t})$ deviation.

Beyond this magnitude effect curvature can also vary across equal-norm tangent directions in higher intrinsic dimensions, e.g. the catenoid in Figure 2b has $H \equiv 0$ despite generally nonzero $\mathrm{II}_p(v,v)$. More importantly, however, is that the curvature effect in example 3 is not confined to synthetic manifolds: in Appendix K.8 (see Table 7) we observe substantial differences between directional and averaged curvature estimates on *real data*, further motivating the pathwise correction considered in the next section.

[1]Note that in this exact setting we do not need any curvature correction terms due to directly projecting to the manifold.

3 PATHWISE CURVATURE-CORRECTED LANDING

The limitations identified in Sections 2.1 and 2.2 motivate our IIL discretization, which combines soft landing with pathwise curvature correction. We next analyze its stationary accuracy under inexact geometry and construct the required geometric operators from a pretrained diffusion score.

3.1 THE IIL-SCHEME

The two failure modes identified in Sections 2.1 and 2.2 suggest two corresponding design choices. First, to avoid repeatedly injecting projection error, we replace hard projection by the landing drift $\ell(p_k - X_k)$. Second, to match the curvature correction to the realized stochastic tangent update, we replace the mean-curvature correction $H(p_k)$ by the realized directional curvature $\mathrm{II}_{p_k}(v_k, v_k)$. Together, these modifications yield the IIL-scheme in equation 6.

On an implicit data manifold, however, the exact geometric operators π, P, and II are unavailable. We therefore use approximations $\widehat{\pi}$, $\widehat{P}$, and $\widehat{\mathrm{II}}$, giving

$$X_{k+1} = X_k + \Delta t \left[-\widehat{P}_k \nabla U(\widehat{p}_k) - \beta^{-1} \widehat{\mathrm{II}}_{\widehat{p}_k}(\widehat{v}_k, \widehat{v}_k) + \ell(\widehat{p}_k - X_k) \right] + \sqrt{2\beta^{-1}\Delta t}\, \widehat{v}_k \,, \tag{12}$$

where $\xi_k \sim \mathcal{N}(0, I_d)$, $\widehat{p}_k = \widehat{\pi}(X_k)$, $\widehat{P}_k = \widehat{P}(\widehat{p}_k)$, and $\widehat{v}_k = \widehat{P}_k \xi_k$. The required approximations $\widehat{\pi} \approx \pi$, $\widehat{P} \approx P$, and $\widehat{\mathrm{II}} \approx \mathrm{II}$ are specified below.

3.2 STATIONARY ERROR UNDER INEXACT GEOMETRY

We next analyze the stationary bias under inexact geometric access. For the stationary analysis, we use localized versions of the schemes detailed in Appendices B–C; when the localization safeguards are inactive, the IIL transition reduces exactly to equation 12, with analogous reductions for the π- and HL-schemes. Let $\mu_{\Delta t}$ denote an invariant law of the π-, HL-, or IIL-scheme, and let μ denote the invariant law of the continuous-time manifold Langevin dynamics equation 4. We measure their discrepancy using the integral probability metric

$$\mathrm{dist}_{\mathcal{F}}(\mu_{\Delta t}, \mu) := \sup_{g \in \mathcal{F}} |\mu_{\Delta t}(g) - \mu(g|_{\mathcal{M}})| \,, \qquad \mathcal{F} := \left\{ g : \|g\|_{C^{2,\alpha}(\mathcal{K}_{\tau_\star})} \le 1 \right\} \,, \tag{13}$$

where we abbreviate $\nu(g) := \int g \, d\nu$ and $C^{2,\alpha}(\mathcal{K}_{\tau_\star})$ denotes the set of C^2-functions with α-Hölder continuous second derivative on $\mathcal{K}_{\tau_\star} \subseteq \mathcal{K}_{\tau_{\mathrm{or}}}$. Integral probability metrics such as equation 13 are a common way to measure stationary bias of discretizations Mattingly et al. (2010). In Appendix E we give further details on $\mathrm{dist}_{\mathcal{F}}$ and show that $\mathrm{dist}_{\mathcal{F}}$ is, up to a power-scaling, equivalent to the well-known Wasserstein W_1-distance.

Our first major step consists in showing that the above distance can be split into the average distance to the manifold and the error in the infinitesimal generators of the corresponding process.

Proposition 2 (Projection-based bias decomposition, informal)**.** *Let $\mu_{\Delta t} \in \{\mu^{\pi}, \mu^{HL}, \mu^{\mathrm{II}L}\}$ be the invariant measure for one of three discretizations (9), (10) and (12) and μ stationary for (4). Then*

$$\mathrm{dist}_{\mathcal{F}}(\mu_{\Delta t}, \mu) \le \mathbb{E}_{X \sim \mu_{\Delta t}} \mathcal{R}_{\perp}(X) + \sup_{g \in \mathcal{F}} \mathbb{E}_{X \sim \mu_{\Delta t}} \mathcal{R}_{\mathrm{gen}}(X; g). \tag{14}$$

where

$$\mathcal{R}_{\perp}(X) = \mathrm{dist}(X, \mathcal{M}) \,, \quad \mathcal{R}_{\mathrm{gen}}(X; g) = |(\mathcal{L} u_g)(\pi(X)) - (\mathcal{L}_{\Delta t}(u_g \circ \pi))(X)| \,, \tag{15}$$

are the distance to manifold and the distance between the infinitesimal generator $\mathcal{L}$ of equation 4 and the discrete generator $\mathcal{L}_{\Delta t}$ of the corresponding discretization and u_g denotes the solution to the Poisson equation $\mathcal{L} u_g = g|_{\mathcal{M}} - \mu(g|_{\mathcal{M}})$ with $\mu(u_g) = 0$; see Appendices C–H for precise definitions and a formal statement.

Thus to achieve a small error in $\mathrm{dist}_{\mathcal{F}}$ a discretization has to enforce a small deviation in both, the manifold distance and dynamics on the manifold itself.

We now state informally our main result concerning the asymptotic dependence on the discretization step size Δt and the geometric errors in Assumption 1; see Appendix H for the precise statement.

Theorem 3 (Stationary accuracy of discretizations, informal)**.** *Under Assumption 1 and further regularity assumptions we have the following asymptotic dependence of equation 15 as* $\Delta t, \varepsilon_\pi, \varepsilon_P, \varepsilon_H, \varepsilon_{\mathrm{II}} \to 0$ *uniformly over* $g \in \mathcal{F}$ *given by*

Scheme	$\mathbb{E}\mathcal{R}_\perp$	$\mathbb{E}\mathcal{R}_{\mathrm{gen}}$	$\mathrm{dist}_{\mathcal{F}}(\mu_{\Delta t}, \mu)$
π	$\mathcal{O}(\varepsilon_\pi)$	$\mathcal{O}\big(\varepsilon_P + \Delta t + \varepsilon_\pi + \frac{\varepsilon_\pi}{\Delta t}\big)$	$\mathcal{O}\big(\varepsilon_P + \varepsilon_\pi + \Delta t + \frac{\varepsilon_\pi}{\Delta t}\big)$
HL	$\mathcal{O}\Big(\frac{\eta_{H,\ell}+\Delta t^{1/2}}{\ell^{1/2}}\Big)$	$\mathcal{O}(\eta_{H,\ell} + \Delta t + \mathbb{E}\mathcal{R}_\perp)$	$\mathcal{O}\Big(\eta_{H,\ell} + \Delta t + \frac{\eta_{H,\ell}+\Delta t^{1/2}}{\ell^{1/2}}\Big)$
IIL ***(ours)***	$\mathcal{O}\Big(\frac{\eta_{\mathrm{II},\ell}+\Delta t}{\ell^{1/2}}\Big)$	$\mathcal{O}(\eta_{\mathrm{II},\ell} + \Delta t + \mathbb{E}\mathcal{R}_\perp)$	$\mathcal{O}\big((1+\ell^{-1/2})(\eta_{\mathrm{II},\ell} + \Delta t)\big)$

with $\eta_{\mathrm{II},\ell} := \varepsilon_P + \varepsilon_{\mathrm{II}} + (1+\ell)\varepsilon_\pi$ *and* $\eta_{H,\ell} := \varepsilon_P + \varepsilon_H + (1+\ell)\varepsilon_\pi$.

Thus the generator bound of the π-scheme contains a possible $\varepsilon_\pi/\Delta t$, whereas soft landing avoids this error amplification. Compared to HL-scheme, our pathwise curvature correction improves the discretization contribution to the normal-deviation bound from $\mathcal{O}(\Delta t^{1/2})$ to $\mathcal{O}(\Delta t)$. With exact geometry, both π and IIL have first-order weak-bias bounds.

Remark 4. The upper bounds in Theorem 3 are not conservative: Example 1 demonstrates inverse-step-size projection-bias amplification in the π-scheme. On S^1 with $U = 0$ and $\beta = \ell = 1$, the exact-geometry HL- and IIL-schemes attain stationary mean normal deviations $\mathbb{E}\mathcal{R}_\perp$ of orders $\Theta(\sqrt{\Delta t})$ and $\Theta(\Delta t)$, respectively, as shown in Example 3 and proved in Appendix I.1.

3.3 Directional curvature from score-derived geometry

To implement the IIL-scheme on an implicit data manifold, we require estimators of the projection π, the tangent projector P, and the directional second fundamental form $\mathrm{II}(v, v)$. Following prior work in inverse problems with diffusion priors Song et al. (2021) we assume that this manifold $\mathcal{M}$ is given by the support $\operatorname{supp} q = \mathcal{M}$ of a distribution q, which is represented by a diffusion model score $s_\sigma(x) = \nabla \log q_\sigma(x)$, where $q_\sigma(x) = \mathcal{N}(0, \sigma^2 I) * q$. Our novel contribution in this section is to show how we can obtain the quantity $\widehat{\mathrm{II}}$ from the score s_σ *directly*.

For this purpose we first note that in Kharitenko et al. (2025) it has been shown that for any given tolerances $\varepsilon_\pi, \varepsilon_P > 0$ and $\mathcal{K}_{\tau_{\mathrm{or}}} \Subset \mathcal{T}_\tau$, there exists some $\sigma > 0$ such that the corresponding Tweedie denoiser $\pi_\sigma(x) = x + \sigma^2 s_\sigma(x)$ satisfies

$$\|\pi_\sigma(x) - \pi(x)\| \le \varepsilon_\pi\,, \quad \|D\pi_\sigma(\pi_\sigma(x)) - D\pi(\pi(x))\| \le \varepsilon_P\,, \ \text{ for all } \ x \in \mathcal{K}_{\tau_{\mathrm{or}}}\,. \tag{16}$$

We therefore assume that we have access to such a denoiser π_σ and set

$$\hat{\pi}(p) = \pi_\sigma(p)\,, \quad \hat{P}(p) = D\pi_\sigma(p) = D\widehat{\pi}(p)$$

for some fixed, small $\sigma > 0$ and propose the finite difference estimator

$$\widehat{\mathrm{II}}_{\widehat{p}}(\widehat{z}, \widehat{z}) := -\|\widehat{z}\|^2 (I - \widehat{P}(\hat{p}))\frac{\widehat{P}(\widehat{p} + \delta\widehat{q})\widehat{q} - \widehat{P}(\widehat{p} - \delta\widehat{q})\widehat{q}}{2\delta}\,, \tag{17}$$

where $\delta > 0$, $\widehat{p} = \widehat{\pi}(p)$ and $\widehat{q} = \widehat{z}/\|\widehat{z}\|$ and $\widehat{\mathrm{II}}_{\widehat{p}}(0,0) := 0$. Note that each II$L$-update equation 12 evaluates equation 17 only along $\widehat{z} = D\widehat{\pi}(\widehat{p})\xi$. The correction therefore requires only Jacobian-vector products, without constructing a tangent basis or the full second fundamental form. Under the corresponding uniform accuracy conditions of equation 16, Theorem 39 in Appendix J gives

$$\|\widehat{H}(\widehat{p}) - H(p)\| \le \varepsilon_H\,, \quad \mathbb{E}_{\xi\sim\mathcal{N}(0,I_d)}\|\widehat{\mathrm{II}}_{\widehat{p}}(\widehat{P}(\widehat{p})\xi, \widehat{P}(\widehat{p})\xi) - \mathrm{II}_p(P(p)\xi, P(p)\xi)\|^2 \le \varepsilon_{\mathrm{II}}^2\,. \tag{18}$$

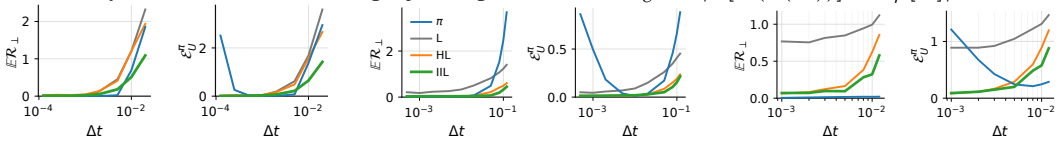
The explicit δ- and oracle-error dependence is given there.

4 Numerical Experiments

Our experiments evaluate both the numerical mechanisms underlying the stationary analysis and the practical behavior of geometric-prior exploration. We evaluate the IIL-scheme with its geometric estimators against the π- and HL-schemes, and the L-scheme (landing without curvature correction), and task-specific baselines such as diffusion posterior sampling. For implicit data manifolds, we use the score-derived estimators from Section 3.3.

4.1 CONTROLLED EXPERIMENTS ON SYNTHETIC MANIFOLDS

Task and evaluation. We first verify the stationary-error mechanisms developed in Section 3.2 on synthetic manifolds: a catenoid and a deformed torus, both 2-dimensional manifolds in $\mathbb{R}^3$, and an 8-dimensional bumpy sphere in $\mathbb{R}^{10}$. We estimate geometry from diffusion model, and compare the π-, L, HL-, and IIL-schemes across step sizes Δt. Motivated by the decomposition in Proposition 2, we directly measure $\mathbb{E}\mathcal{R}_\perp$ and the projected potential bias $\mathcal{E}_U^\pi := |\widehat{\mathbb{E}}[U(\pi(X))] - \mathbb{E}_\mu[U]|$.



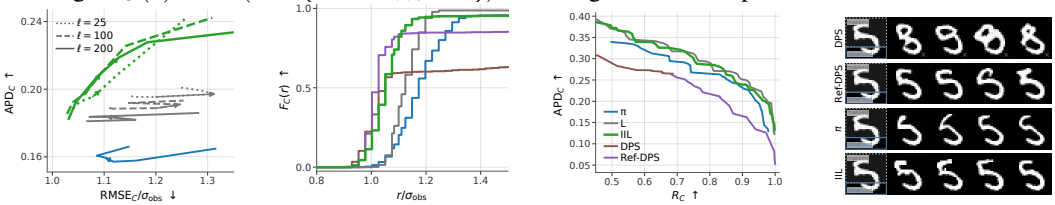
(a) Catenoid. (b) Deformed torus. (c) Bumpy sphere.

Figure 3: Controlled synthetic experiments under score-derived inexact geometry.

Effect of inexact geometry. Figure 3 shows an amplification of $\mathcal{E}_U^\pi$ in the π-scheme in the small-step regime, which does not emerge in landing-based schemes. Among the curvature-corrected schemes, IIL stays closer to the manifold than L and HL, with lower projected potential error.

4.2 IMAGE INPAINTING ON MNIST

Task and evaluation. We study MNIST inpainting (LeCun et al., 2010), using an unconditional DDPM (Ho et al., 2020) to approximate the data-manifold geometry. Given a selected completion x_{ref}, we initialize IIL, L, and π at x_{ref} and explore observation-consistent alternatives; class consistency is used only for evaluation, not as part of the target. We compare with DPS (Chung et al., 2023) and reference-penalized Ref-DPS. Let C denote finite outputs for which two separately-trained classifiers both predict the selected digit class. We report selected-class retention $R_C = \Pr(C)$, conditional observation error RMSE_C, missing-region diversity APD_C, and joint coverage $F_C(r) = \Pr\left(C \cap \{\text{RMSE}_{\text{obs}} \le r\}\right)$ over all generated samples.



(a) Discretization. (b) Coverage. (c) Retention-diversity. (d) Completions.

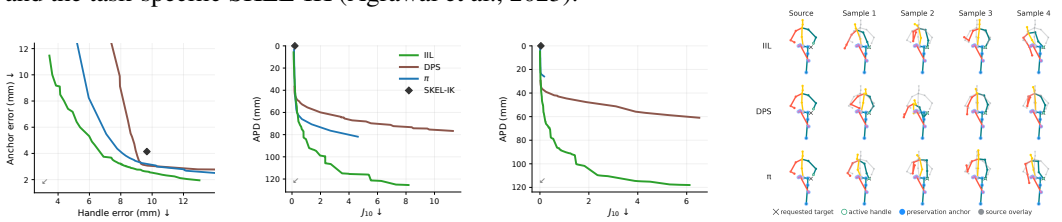
Figure 4: **MNIST inpainting. (a)** $\beta = 0.5$; arrows: increasing Δt. **(b)** $F_C(r)$ at a fixed configuration. **(c)** Frontiers across parameter sweeps. **(d)** One observation; crosses mark out-of-class outputs.

Discretization and conditional exploration. Across step sizes and landing strengths, IIL achieves a more favorable fidelity–diversity trade-off than the π- and L-schemes: at comparable conditional observation error, it attains higher missing-region diversity (Figure 4(a)). Equal-runtime L with smaller Δt shows only partial recovery (Appendix K.2). This advantage also appears in joint semantic-and-measurement coverage, with IIL attaining greater $F_C(r)$ than L and π through the tight-residual transition region (Figure 4(b)). Across the broader parameter sweeps, IIL further provides greater diversity at comparable selected-class retention than DPS and Ref-DPS (Figure 4(c)). Panel (d) illustrates diverse completions given a selected digit class.

4.3 HUMAN POSE EDITING ON HUMAN3.6M

Task and evaluation. We use 3D pose editing to test the control of the constraint–diversity trade-off through the inverse temperature parameter β. Given a source pose x^{src}, handle joints $\mathcal{H}$ with targets $\{y_j\}$, anchor joints $\mathcal{A}$, and a length scale τ, let $E_{\text{h}} = \frac{1}{3|\mathcal{H}|}\sum_{j\in\mathcal{H}} \|x_j - y_j\|_2^2/\tau^2$ and $E_{\text{a}} = \frac{1}{3|\mathcal{A}|}\sum_{j\in\mathcal{A}} \|x_j - x_j^{\text{src}}\|_2^2/\tau^2$. We sample from $U_r = (E_{\text{h}} + rE_{\text{a}})/(1+r)$ and report accuracy as $J_{10} = (E_{\text{h}} + E_{\text{a}})/2$ and diversity as average pairwise distance (APD) over unconstrained joints.

We use Human3.6M (Ionescu et al., 2013) with a pretrained GFPose model (Ci et al., 2023) on two benchmarks: *Pair*, a controlled stress test with competing handle and anchor objectives, and *Real-edit*, constructed from real motion. We examine the handle–anchor error trade-off and the accuracy–diversity trade-off. We compare against DPS (Chung et al., 2023) at comparable runtime and the task-specific SKEL-IK (Agrawal et al., 2023).



(a) Pair. (b) Pair. (c) Real-edit. (d) Solutions to one edit.

Figure 5: **Pose editing on H36M.** Pareto frontiers and diverse solutions to one sparse edit.

Temperature-controlled constrained exploration. Across parameter sweeps over β, IIL improves the handle–anchor trade-off over DPS on *Pair* (a) and achieves jointly lower handle and anchor errors on *Real-edit*. On both benchmarks, it also provides greater unconstrained-joint diversity than DPS at matched J_{10} over most of the evaluated range, at comparable runtime (b,c). Panel (d) shows multiple solutions to one sparse edit.

4.4 Image inverse problems on FFHQ-256

Task and evaluation. We apply the learned-manifold dynamics to eight inverse problems on FFHQ-256 (Chung et al., 2023; Karras et al., 2019), using DAPS (Zhang et al., 2025) as a controlled host solver: for a fixed DAPS trajectory, the native and IIL variants share the same upstream reconstruction and branch only in the terminal conditional-Langevin stage, where DAPS+IIL replaces the native updates with manifold Langevin. We report PSNR, SSIM, and LPIPS against native DAPS at comparable runtime, with compute-matched DPS best-of-K by minimum measurement residual; all methods share the same pretrained diffusion model.

Table 1: FFHQ-256 inverse problems at comparable runtime.

				Linear			*Nonlinear*		
	Method	Random	Box	SR×4	Gaussian	Motion	Nonlinear	Phase	HDR
PSNR↑	DPS-BoK	29.31	24.24	24.73	26.07	27.36	26.64	18.76	26.30
	Native DAPS	30.70	24.39	29.26	29.51	31.58	28.42	27.34	27.79
	DAPS+IIL	**32.41**	**24.79**	**30.18**	**30.41**	**32.83**	**28.55**	**27.73**	**27.85**
SSIM↑	DPS-BoK	.839	**.826**	.705	.731	.774	.768	.543	.835
	Native DAPS	.802	.749	.782	.786	.838	.794	.741	.844
	DAPS+IIL	**.891**	.774	**.850**	**.851**	**.898**	**.816**	**.787**	**.876**
LPIPS↓	DPS-BoK	.171	**.154**	.257	.204	.181	.180	.397	.146
	Native DAPS	.155	.175	.192	.175	.134	.181	.227	.156
	DAPS+IIL	**.101**	.162	**.160**	**.143**	**.098**	**.168**	**.221**	**.136**

Reconstruction quality at comparable runtime. Table 1 shows that DAPS+IIL improves PSNR, SSIM, and LPIPS over native DAPS on all eight tasks at comparable runtime, by up to 1.7 dB PSNR, and by 11% and 35% in relative SSIM and LPIPS, attaining the highest PSNR on all eight and the best SSIM and LPIPS on seven of eight among all compared methods. Under compute controls sweeping host size and suffix length (Appendix K.7), allocating part of the compute budget to IIL gives higher reconstruction quality than spending it entirely on native DAPS.

5 Conclusion

We introduced geometric-prior exploration on unknown data manifolds and proposed IIL, a manifold Langevin discretization that combines pathwise curvature correction with soft landing. This design avoids repeated-projection bias amplification and explicit mean-curvature computation, while achieving first-order stationary weak bias in the step size up to geometric approximation errors. Experiments validate these mechanisms, demonstrating controlled geometric-prior exploration on MNIST and Human3.6M and improved DAPS refinement on FFHQ-256 at comparable runtime.

AI USE STATEMENT

In this work, we used generative AI tools to formulate mathematical claims, provide critical ingredients for proving mathematical claims, assist in the writing of proofs, propose or refine hypotheses, design or provide feedback on research methodology or experiments, implement methods, support qualitative and thematic data analysis, and interpret results.

We have not used generative AI tools to generate synthetic datasets, develop theoretical models or conceptual frameworks, assist with translation, or clean and reformat datasets. There are no remaining required-disclosure tasks applicable to this work.

Additionally, we used generative AI tools to create or modify scientific figures or images through AI-assisted plotting code, create or edit software code, draft parts of the research paper, summarize or analyze existing literature, discover research topics or identify gaps, brainstorm, source or search for information, edit the research paper to improve readability, identify relevant literature, suggest a structure for the research paper, and propose a title or keywords for the research paper.

We have reviewed all AI-assisted work. AI-assisted mathematical claims, proof arguments, methodological suggestions, experimental suggestions, and interpretations of results were reviewed and verified by the authors. AI-assisted code, including plotting code used to generate scientific figures, was manually reviewed and tested for correctness, and the underlying data and resulting visualizations were checked by the authors. AI-assisted literature searches and summaries were verified against the original publications. AI-assisted research ideas were checked against the relevant literature, and AI-assisted manuscript text was reviewed and revised by the authors for technical accuracy, clarity, and consistency.

We take responsibility for the final content of this work, including text, claims, code, figures, and other artifacts produced with the aid of generative AI.

ETHICS STATEMENT

This work does not involve new data collection or direct interaction with human participants. Our experiments use existing benchmark datasets and pretrained models, including MNIST, Human3.6M, and FFHQ-256, together with procedurally defined synthetic-manifold experiments. Because some of these benchmarks represent people, the experiments may inherit biases and limitations present in the underlying datasets and pretrained models. The proposed methodology is intended for research on controlled exploration, inverse problems, and editing, and can produce multiple plausible reconstructions or edits consistent with partial observations or constraints. Such outputs should not be interpreted as uniquely determined ground truth, and downstream applications should consider potential risks related to misuse, privacy, and representational bias. We do not release a new human-subject dataset as part of this work. Use of existing datasets and pretrained models remains subject to their respective licenses, terms, and intended research settings.

REPRODUCIBILITY STATEMENT

We have taken several steps to support the reproducibility of this work. The main paper specifies the proposed IIL discretization, the score-derived geometric operators, and the theoretical assumptions and results. Detailed formulations, regularity and localization assumptions, intermediate results, and complete proofs are provided in Appendices B–J. Experimental protocols and implementation details are provided in Appendix K. In particular, Appendix K.1 describes the synthetic-manifold experiments; Appendices K.2–K.5 provide the MNIST experimental and robustness details; Appendix K.6 describes the Human3.6M pose-editing experiments; Appendix K.7 provides the FFHQ-256 inverse-problem settings and compute controls; and Appendix K.8 describes the directional-curvature diagnostics. These appendices document the relevant dataset and task construction, pretrained models, parameter sweeps, random seeds where applicable, evaluation procedures, validation protocols, and runtime-measurement settings used to obtain the reported results.

REFERENCES

Dhruv Agrawal, Martin Guay, Jakob Buhmann, Dominik Borer, and Robert W. Sumner. Pose and skeleton-aware neural ik for pose and motion editing. In *SIGGRAPH Asia 2023 Conference Papers*, pp. 1–10, 2023.

John Armstrong and Tim King. Curved schemes for stochastic differential equations on, or near, manifolds. *Proceedings of the Royal Society A: Mathematical, Physical and Engineering Sciences*, 478(2262), 2022.

Simone Azeglio and Arianna Di Bernardo. What's inside your diffusion model? a score-based riemannian metric to explore the data manifold. *arXiv preprint arXiv:2505.11128*, 2025.

Jöran Bergh and Jörgen Löfström. *Interpolation spaces: an introduction*. Springer Science & Business Media, 2012.

Karthik Bharath, Alexander Lewis, Akash Sharma, and Michael V Tretyakov. Sampling and estimation on manifolds using the langevin diffusion. *Journal of Machine Learning Research*, 26(71): 1–50, 2025.

Jean-Daniel Boissonnat, André Lieutier, and Mathijs Wintraecken. The reach, metric distortion, geodesic convexity and the variation of tangent spaces: J.-d. boissonnat et al. *Journal of applied and computational topology*, 3(1):29–58, 2019.

Honglin Chen, Karran Pandey, Rundi Wu, Matheus Gadelha, Yannick Hold-Geoffroy, Ayush Tewari, Niloy J Mitra, Changxi Zheng, and Paul Guerrero. Vips: Video-informed pose spaces for auto-rigged meshes. *arXiv preprint arXiv:2604.17623*, 2026.

Xiang Cheng, Jingzhao Zhang, and Suvrit Sra. Efficient sampling on riemannian manifolds via langevin mcmc. *Advances in Neural Information Processing Systems*, 35:5995–6006, 2022.

Hyungjin Chung, Byeongsu Sim, Dohoon Ryu, and Jong Chul Ye. Improving diffusion models for inverse problems using manifold constraints. *Advances in neural information processing systems*, 35:25683–25696, 2022.

Hyungjin Chung, Jeongsol Kim, Michael Thompson Mccann, Marc Louis Klasky, and Jong Chul Ye. Diffusion posterior sampling for general noisy inverse problems. In *The Eleventh International Conference on Learning Representations*, 2023. URL `https://openreview.net/forum?id=OnD9zGAGT0k`.

Hai Ci, Mingdong Wu, Wentao Zhu, Xiaoxuan Ma, Hao Dong, Fangwei Zhong, and Yizhou Wang. Gfpose: Learning 3d human pose prior with gradient fields. In *2023 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 4800–4810. IEEE, 2023.

Valentin De Bortoli, Emile Mathieu, Michael Hutchinson, James Thornton, Yee Whye Teh, and Arnaud Doucet. Riemannian score-based generative modelling. *Advances in neural information processing systems*, 35:2406–2422, 2022.

Prafulla Dhariwal and Alexander Nichol. Diffusion models beat gans on image synthesis. *Advances in neural information processing systems*, 34:8780–8794, 2021.

Willem Diepeveen, Georgios Batzolis, Zakhar Shumaylov, and Carola-Bibiane Schönlieb. Score-based pullback riemannian geometry: Extracting the data manifold geometry using anisotropic flows. *arXiv preprint arXiv:2410.01950*, 2024.

Charles Fefferman, Sanjoy Mitter, and Hariharan Narayanan. Testing the manifold hypothesis. *Journal of the American Mathematical Society*, 29(4):983–1049, 2016.

Christine Guillemot and Olivier Le Meur. Image inpainting: Overview and recent advances. *IEEE signal processing magazine*, 31(1):127–144, 2013.

Jonathan Ho, Ajay Jain, and Pieter Abbeel. Denoising diffusion probabilistic models. *Advances in neural information processing systems*, 33:6840–6851, 2020.

Elton P Hsu. *Stochastic analysis on manifolds*. Number 38. American Mathematical Soc., 2002.

Catalin Ionescu, Dragos Papava, Vlad Olaru, and Cristian Sminchisescu. Human3. 6m: Large scale datasets and predictive methods for 3d human sensing in natural environments. *IEEE transactions on pattern analysis and machine intelligence*, 36(7):1325–1339, 2013.

Kijung Jeon, Michael Muehlebach, and Molei Tao. Fast non-log-concave sampling under nonconvex equality and inequality constraints with landing. *Advances in Neural Information Processing Systems*, 38:158640–158701, 2026.

Jaehyeong Jo and Sung Ju Hwang. Generative modeling on manifolds through mixture of riemannian diffusion processes. *arXiv preprint arXiv:2310.07216*, 2023.

Iolo Jones. Manifold diffusion geometry: Curvature, tangent spaces, and dimension. *SIAM Journal on Mathematics of Data Science*, 8(2):257–277, 2026.

Hamidreza Kamkari, Brendan L Ross, Rasa Hosseinzadeh, Jesse C Cresswell, and Gabriel Loaiza-Ganem. A geometric view of data complexity: Efficient local intrinsic dimension estimation with diffusion models. *Advances in Neural Information Processing Systems*, 37:38307–38354, 2024.

Tero Karras, Samuli Laine, and Timo Aila. A style-based generator architecture for generative adversarial networks. In *2019 IEEE/CVF conference on computer vision and pattern recognition (CVPR)*, pp. 4396–4405. IEEE, 2019.

Bahjat Kawar, Michael Elad, Stefano Ermon, and Jiaming Song. Denoising diffusion restoration models. *Advances in neural information processing systems*, 35:23593–23606, 2022.

Victor Kawasaki-Borruat, Clara Grotehans, Pierre Vandergheynst, and Adam Gosztolai. Diffusion processes on implicit manifolds. *arXiv preprint arXiv:2604.07213*, 2026.

Andrey Kharitenko, Zebang Shen, Riccardo de Santi, Niao He, and Florian Doerfler. Landing with the score: Riemannian optimization through denoising. *arXiv preprint arXiv:2509.23357*, 2025.

Seunghoi Kim, Henry FJ Tregidgo, Matteo Figini, Chen Jin, Sarang Joshi, and Daniel C Alexander. Tackling hallucination from conditional models for medical image reconstruction with dynamicdps. In *International Conference on Medical Image Computing and Computer-Assisted Intervention*, pp. 593–603. Springer, 2025.

Rémi Laumont, Valentin De Bortoli, Andrés Almansa, Julie Delon, Alain Durmus, and Marcelo Pereyra. Bayesian imaging using plug & play priors: when langevin meets tweedie. *SIAM Journal on Imaging Sciences*, 15(2):701–737, 2022.

Adrien Laurent and Gilles Vilmart. Order conditions for sampling the invariant measure of ergodic stochastic differential equations on manifolds. *Foundations of Computational Mathematics*, 22 (3):649–695, 2022.

Yann LeCun, Corinna Cortes, and CJ Burges. Mnist handwritten digit database. *ATT Labs [Online]. Available: http://yann.lecun.com/exdb/mnist*, 2, 2010.

John M Lee. Smooth manifolds. In *Introduction to smooth manifolds*, pp. 1–31. Springer, 2013.

Tony Lelievre, Mathias Rousset, and Gabriel Stoltz. Langevin dynamics with constraints and computation of free energy differences. *Mathematics of computation*, 81(280):2071–2125, 2012.

Mufan Li and Murat A Erdogdu. Riemannian langevin algorithm for solving semidefinite programs. *Bernoulli*, 29(4):3093–3113, 2023.

Xiang Li, Zebang Shen, Ya-Ping Hsieh, and Niao He. When scores learn geometry: Rate separations under the manifold hypothesis. In *International Conference on Learning Representations*, volume 2026, pp. 40627–40660, 2026.

Junzhe Lu, Jing Lin, Hongkun Dou, Ailing Zeng, Yue Deng, Yulun Zhang, and Haoqian Wang. Dposer: Diffusion model as robust 3d human pose prior. *arXiv preprint arXiv:2312.05541*, 2023.

Junzhe Lu, Jing Lin, Hongkun Dou, Ailing Zeng, Yue Deng, Xian Liu, Zhongang Cai, Lei Yang, Yulun Zhang, Haoqian Wang, et al. Dposer-x: Diffusion model as robust 3d whole-body human pose prior. In *2025 IEEE/CVF International Conference on Computer Vision (ICCV)*, pp. 9988–9997. IEEE, 2025.

Alessandra Lunardi et al. *Analytic semigroups and optimal regularity in parabolic problems*, volume 16. Birkhäuser Basel, 1995.

Dawid Malarz, Artur Kasymov, Maciej Zięba, Jacek Tabor, and Przemysław Spurek. Classifier-free guidance with adaptive scaling. In *ECAI 2025: 28th European Conference on Artificial Intelligence 25-30 October 2025, Bologna, Italy-Including 14th Conference on Prestigious Applications of Intelligent Systems (PAIS 2025) Proceedings*, pp. 435–442. SAGE Publications 1 Oliver's Yard, 55 City Road, London, EC1Y 1SP, 2025.

Morteza Mardani, Jiaming Song, Jan Kautz, and Arash Vahdat. A variational perspective on solving inverse problems with diffusion models. In *International Conference on Learning Representations*, volume 2024, pp. 28027–28053, 2024.

Jonathan C Mattingly, Andrew M Stuart, and Michael V Tretyakov. Convergence of numerical time-averaging and stationary measures via poisson equations. *SIAM Journal on Numerical Analysis*, 48(2):552–577, 2010.

Roger Moser. *Partial regularity for harmonic maps and related problems*. World Scientific, 2005.

Du Nguyen and Stefan Sommer. Second-order differential operators, stochastic differential equations and brownian motions on embedded manifolds. *arXiv preprint arXiv:2406.02879*, 2024.

Frank Permenter and Chenyang Yuan. Interpreting and improving diffusion models from an optimization perspective. *arXiv preprint arXiv:2306.04848*, 2023.

Jakiw Pidstrigach. Score-based generative models detect manifolds. *Advances in Neural Information Processing Systems*, 35:35852–35865, 2022.

Marc J Piggott and Victor Solo. Geometric euler–maruyama schemes for stochastic differential equations in so (n) and se (n). *SIAM Journal on Numerical Analysis*, 54(4):2490–2516, 2016.

Simon Schwarz, Michael Herrmann, Anja Sturm, and Max Wardetzky. Efficient random walks on riemannian manifolds. *Foundations of Computational Mathematics*, 25(1):145–161, 2025.

Upanshu Sharma and Wei Zhang. Nonreversible sampling schemes on submanifolds. *SIAM Journal on Numerical Analysis*, 59(6):2989–3031, 2021.

Yang Song, Liyue Shen, Lei Xing, and Stefano Ermon. Solving inverse problems in medical imaging with score-based generative models. *arXiv preprint arXiv:2111.08005*, 2021.

Jan Pawel Stanczuk, Georgios Batzolis, Teo Deveney, and Carola-Bibiane Schönlieb. Diffusion models encode the intrinsic dimension of data manifolds. In *Forty-first International Conference on Machine Learning*, 2024.

Rong Tang and Yun Yang. Adaptivity of diffusion models to manifold structures. In *International conference on artificial intelligence and statistics*, pp. 1648–1656. PMLR, 2024.

Darshan Thaker, Abhishek Goyal, and René Vidal. Frequency-guided posterior sampling for diffusion-based image restoration. In *2025 IEEE/CVF International Conference on Computer Vision (ICCV)*, pp. 12873–12882. IEEE, 2025.

Enrico Ventura, Beatrice Achilli, Gianluigi Silvestri, Carlo Lucibello, and Luca Ambrogioni. Manifolds, random matrices and spectral gaps: The geometric phases of generative diffusion. In *International Conference on Learning Representations*, volume 2025, pp. 88317–88338, 2025.

Jason Wang, Bobak Kiani, and Melanie Weber. Curvature estimation on data manifolds via diffusion-augmented sampling. In *NeurIPS 2025 Workshop on Symmetry and Geometry in Neural Representations*, 2025.

Yinhuai Wang, Jiwen Yu, and Jian Zhang. Zero-shot image restoration using denoising diffusion null-space model. *arXiv preprint arXiv:2212.00490*, 2022.

Tianmin Yu, Shixin Zheng, Jianfeng Lu, Govind Menon, and Xiangxiong Zhang. Riemannian langevin monte carlo schemes for sampling psd matrices with fixed rank. *arXiv preprint arXiv:2309.04072*, 2023.

Honglin Yuan, Xiaoyi Gu, Rongjie Lai, and Zaiwen Wen. Global optimization with orthogonality constraints via stochastic diffusion on manifold. *Journal of Scientific Computing*, 80(2):1139–1170, 2019.

Zhiyuan Zhan and Masashi Sugiyama. Riemannian langevin dynamics: Strong convergence of geometric euler-maruyama scheme. *arXiv preprint arXiv:2603.03626*, 2026.

Bingliang Zhang, Wenda Chu, Julius Berner, Chenlin Meng, Anima Anandkumar, and Yang Song. Improving diffusion inverse problem solving with decoupled noise annealing. In *2025 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 20895–20905. IEEE, 2025.

Ruqi Zhang, Qiang Liu, and Xin Tong. Sampling in constrained domains with orthogonal-space variational gradient descent. *Advances in Neural Information Processing Systems*, 35:37108–37120, 2022.

Wei Zhang. Ergodic sdes on submanifolds and related numerical sampling schemes. *ESAIM: Mathematical Modelling & Numerical Analysis (ESAIM: M2AN)*, 54(2):391, 2020.

A Related work

A.1 Geometric Priors and Task-Conditioned Sampling

Learned-prior inference for images. Many inverse solvers use pretrained models as distributional priors or learned regularizers, incorporating information about the data distribution beyond measurement consistency. In Bayesian formulations, the observation likelihood and learned prior determine the posterior, as in equation 2. Diffusion Posterior Sampling (DPS) (Chung et al., 2023) seeks approximate posterior samples by combining reverse-diffusion updates with approximate likelihood guidance. Decoupled Annealing Posterior Sampling (DAPS) (Zhang et al., 2025) also seeks posterior samples, using decoupled noise annealing with denoising, measurement-conditioned Langevin updates, and re-noising. For linear restoration, DDRM (Kawar et al., 2022) exploits the measurement operator's singular-value structure, while DDNM (Wang et al., 2022) separates measurement-determined and null-space components. These methods use pretrained diffusion priors but differ in their inference procedures and approximation guarantees.

Learned-prior inference is not restricted to reverse-diffusion sampling. PnP-ULA (Laumont et al., 2022) combines an explicit likelihood with denoiser-derived approximations to the score of a smoothed prior in an ambient-space Langevin sampler. RED-diff (Mardani et al., 2024) adopts a variational posterior formulation and optimizes an objective combining measurement fidelity with diffusion-based regularization. Within reverse-diffusion methods, manifold-constrained gradients (MCG) (Chung et al., 2022) use local manifold information to guide measurement updates. Thus, the distinction from our approach is not that existing solvers ignore geometry. Moreover, projection onto a measurement-consistency set is different from the nearest-point projection onto the data manifold used in our discretizations.

Pose priors and editing models. Learned pose models similarly support several forms of task-conditioned inference. GFPose (Ci et al., 2023) learns score-based pose distributions and studies conditional inference, imputation, and their combination. DPoser and DPoser-X (Lu et al., 2023; 2025) formulate pose-centric tasks as inverse problems and combine task-specific measurement objectives with diffusion priors through variational inference. Complementary editing models incorporate additional structure directly into their architectures or learned pose spaces. SKEL-IK (Agrawal et al., 2023) conditions inverse kinematics on a base pose and sparse editing constraints, while ViPS (Chen et al., 2026) distills video-model priors into editable pose spaces. These approaches provide relevant mechanisms for task-conditioned generation and editing, but should not all be interpreted as samplers from the same learned-prior posterior.

Geometric-prior sampling. Our distinction concerns the reference weighting, rather than whether task and prior contributions can be specified separately. We use the pretrained model to estimate data-manifold geometry and adopt normalized manifold volume, induced by the ambient Euclidean metric, as the geometric prior. Under our manifold-support formulation, suppose that q admits a density r_q with respect to $\mathrm{vol}_{\mathcal{M}}$. For the Gaussian observation model, with $U(x) = \frac{1}{2}\|y - \mathcal{A}(x)\|^2$ and $\beta = \sigma^{-2}$, the learned-prior posterior and geometric-prior target satisfy

$$q(dx \mid y) \propto e^{-\beta U(x)} r_q(x)\ \mathrm{vol}_{\mathcal{M}}(dx),$$
$$\mu(dx) \propto e^{-\beta U(x)}\ \mathrm{vol}_{\mathcal{M}}(dx).$$

The geometric-prior target omits the explicit learned data-density factor r_q. Equal-cost solutions therefore have equal target density per unit manifold volume, although semantic families occupying different volumes need not have equal probabilities. This changes the reference prior and, in general, the desired exploration distribution, rather than merely providing another approximation to the original posterior. The geometric operators remain learned from data; omitting r_q from the ideal target does not make their estimation independent of the training distribution.

Related work by Li et al. (2026) analyzes tempered-score Langevin dynamics whose stationary laws converge to uniform or likelihood-weighted manifold volume as the diffusion smoothing noise vanishes, under their stated assumptions. Under their uniform score-discrepancy assumptions, they also establish weaker accuracy conditions for manifold concentration than for full distributional recovery. These results support a geometric use of pretrained scores, but do not by themselves guarantee the accuracy of the geometric derivatives needed by our estimator. Their manifold-volume sampling analysis concerns continuous-time stationary laws and explicitly leaves numerical discretization error unquantified.

Relation to our work. We build on this geometric-prior perspective and develop its numerical realization for controlled exploration. The task potential and inverse temperature determine density relative to manifold volume, while IIL implements the dynamics through approximate projection, tangent, and directional-curvature operators. Our analysis quantifies how finite time steps and geometric approximation errors perturb the prescribed target, establishing stationary smooth weak-bias bounds for localized IIL under explicit assumptions and at fixed stable landing strength. This complements the choice of geometric prior with a discretization designed for learned geometry. We also use IIL as terminal refinement within DAPS; this evaluates its practical compatibility with an existing solver, rather than establishing that the complete hybrid pipeline samples the standalone geometric-prior target.

A.2 Manifold Langevin and Its Discretization

Geometric formulations. Manifold Brownian motion and Langevin dynamics admit intrinsic and extrinsic constructions (Hsu, 2002), with suitable Langevin dynamics preserving Gibbs measures relative to Riemannian volume (Lelievre et al., 2012; Zhang, 2020; Bharath et al., 2025). Related stochastic dynamics also support constrained optimization (Yuan et al., 2019; Li & Erdogdu, 2023) and manifold-based generative modeling (De Bortoli et al., 2022; Jo & Hwang, 2023). Classical discretizations are generally defined in terms of exact geometric maps or operators, although their practical evaluation may involve numerical approximations.

Intrinsic discretizations. Intrinsic schemes map tangent stochastic increments onto the manifold through exponential maps or retractions (Cheng et al., 2022; Schwarz et al., 2025; Bharath et al., 2025). Related geometric Euler–Maruyama constructions incorporate manifold structure into the stochastic update (Piggott & Solo, 2016; Nguyen & Sommer, 2024; Zhan & Sugiyama, 2026). Exponential maps transport increments along geodesics, while retractions provide alternatives when evaluating the exponential map is expensive or unavailable in closed form. Their use requires suitable consistency conditions to approximate the intended diffusion. For Langevin sampling on compact Riemannian manifolds, Bharath et al. (2025) establish first-order error bounds for ensemble- and time-averaging estimators and for stationary measures against smooth test functions. Their analysis also accommodates retractions satisfying additional consistency conditions.

Extrinsic discretizations and normal relaxation. Extrinsic schemes work in ambient coordinates and account for the manifold through constrained updates or projection (Lelievre et al., 2012; Zhang,

2020; Sharma & Zhang, 2021), sometimes together with explicit geometric drift corrections (Yu et al., 2023). These approaches use an available representation of the constraints or the associated geometric operators. Higher-order accuracy for sampling invariant measures is also possible: Laurent & Vilmart (2022) derive order conditions for constrained stochastic Runge–Kutta methods and construct a second-order sampler.

Normal relaxation provides an alternative to projecting each update back onto the manifold (Zhang et al., 2022; Jeon et al., 2026). Such methods combine motion along the constrained domain with a component that reduces constraint violation. In particular, Jeon et al. (2026) develop Overdamped Langevin with Landing (OLLA) for equality and inequality constraints and establish continuous-time convergence under their assumptions. Their OLLA-H implementation uses Euler–Maruyama discretization with Hutchinson trace estimation for the mean-curvature correction. These constructions avoid repeated nonlinear projection solves but still use the prescribed constraint functions and their differential information.

Numerical approximations and data-driven geometry. The numerical realization of geometric operations has also been analyzed. Armstrong & King (2022) construct curved stochastic updates using first- and second-order geometric information and allow approximate curved steps, including implementations based on numerical ODE solvers or Taylor truncation. Their analysis relates approximation accuracy to the temporal and stochastic increments and establishes finite-horizon strong and weak convergence and manifold-deviation estimates. Sharma & Zhang (2021) study numerical approximation of an ODE-defined constraint projection and quantify its contribution to averaging error.

For implicit manifolds represented by point clouds, Kawasaki-Borruat et al. (2026) construct diffusions from graph-based generator estimates, establish convergence to the underlying manifold diffusion, and analyze the associated ambient Euler–Maruyama discretization. These developments extend geometric numerical methods through approximate maps or data-driven coefficients, with guarantees tied to the specified geometric representations and approximation conditions. Replacing these operations with biased learned estimates therefore requires a corresponding error analysis. Our focus is discretization design and stationary accuracy under approximate projection, tangent, and curvature operators.

A.3 Geometry from Diffusion Scores

Manifold support and intrinsic dimension. A growing literature investigates the geometric information captured by diffusion models and related score-based constructions. Pidstrigach (2022) establish conditions under which score-based generative models produce samples on the underlying data manifold, even when the generated distribution does not reproduce its relative probability weights. Li et al. (2026) further distinguish manifold concentration from distributional recovery through small-noise score-accuracy conditions. From a statistical perspective, Tang & Yang (2024) derive distribution-estimation rates adapted to intrinsic dimension without requiring explicit manifold estimation. Other methods extract dimensional information directly: Stanczuk et al. (2024) relate small-noise scores to normal directions and use this relationship to estimate intrinsic dimension, while Kamkari et al. (2024) construct a local intrinsic-dimension estimator from the Fokker–Planck equation associated with a diffusion model. These results concern support properties, statistical adaptation, or dimensional estimation, rather than a common construction of all geometric operators needed for manifold computation.

Local structure and score-informed metrics. Score derivatives and learned distributions also provide information about local geometric structure. Ventura et al. (2025) analyze score-Jacobian spectra, relating spectral gaps to manifold dimensionality and different phases of the generative process. Diepeveen et al. (2024) develop score-based pullback Riemannian geometry and instantiate their framework using anisotropic normalizing flows. Azeglio & Di Bernardo (2025) construct a score-informed ambient Riemannian metric and use its geodesics for interpolation and extrapolation. These approaches characterize local structure or construct geometries for navigating learned distributions; their geometric objects need not coincide with the Euclidean nearest-point projection and its derivatives for a fixed embedded data manifold.

Complementary point-cloud curvature estimation. Curvature can also be estimated through sampled geometric representations. Wang et al. (2025) use diffusion-generated samples to increase point-cloud density for downstream curvature estimation. Jones (2026) construct differential operators from diffusion-maps approximations of the Laplace–Beltrami operator and estimate the second fundamental form through Hessian information. In the latter construction, diffusion maps describe a point-cloud approximation of manifold diffusion, not the score of a pretrained generative diffusion model. These methods therefore provide alternative access to curvature through sampled data and estimated differential operators.

Score-derived projection and tangent operators. A more direct connection between denoising and geometric computation is developed by Permenter & Yuan (2023), who interpret denoising as approximate projection and diffusion sampling as approximate gradient descent on a distance function. Closest to our setting, Kharitenko et al. (2025) show that the ideal Gaussian denoiser and its Jacobian approximate the nearest-point projection and its derivative uniformly on suitable tubular neighborhoods as the smoothing noise vanishes, under their manifold and density assumptions. At manifold points, the projection derivative is the orthogonal tangent projector. They use these operators for deterministic optimization on implicit manifolds, including a denoising landing flow and a Riemannian gradient-descent construction. Their guarantees for learned implementations additionally require accuracy of the network-derived denoiser and its Jacobian, distinguishing properties of the ideal score from approximation conditions on a trained model.

Directional curvature access. Our construction extends this score-based geometric access to the directional second fundamental form needed by each IIL update. Rather than reconstructing curvature from a point cloud, we query the pretrained denoiser and use centered finite differences of its Jacobian-vector products to estimate $\mathrm{II}_p(v, v)$ along the sampled tangent increment. This avoids assembling a point-cloud differential operator, a tangent basis, or the full second fundamental form. Under uniform local projection and projection-Jacobian accuracy assumptions, including at the shifted finite-difference queries, we bound the directional-curvature error in terms of the operator errors and the finite-difference scale. The resulting oracle-error bound then enters our stationary smooth weak-bias analysis for localized IIL.

B Details on IIL-scheme

We now present in detail the proposed finite-step discretization scheme for manifold Langevin with inverse temperature $\beta > 0$ for inexact-geometry setting. For an iterate X_k, let $\widehat{p}_k = \widehat{\pi}(X_k)$, $\widehat{P}_k = \widehat{P}(\widehat{p}_k)$, $\widehat{v}_k = \widehat{P}_k \xi_k$, and $\xi_k \sim \mathcal{N}(0, I_d)$. Define the learned residual $\widehat{\mathsf{d}}_k := \|X_k - \widehat{p}_k\|$ and a smooth cutoff $\vartheta_k := \vartheta(\widehat{\mathsf{d}}_k) \in [0, 1]$, where $\vartheta = 1$ on $[0, \rho_0]$ and $\vartheta = 0$ on $[\rho_1, \infty)$ for $0 < \rho_0 < \rho_1$. For $\gamma \in (0, \frac{1}{2})$, let $\chi_k := \mathbf{1}\{\|\xi_k\| \le r_{\Delta t}\}$ with $r_{\Delta t} := \sqrt{d} + \Delta t^{-\gamma}$. Our II$L$-scheme is

$$X^{\mathrm{II}L}_{k+1} = X_k + \ell \Delta t\,(\widehat{p}_k - X_k) + \chi_k \Big[-\Delta t\, \vartheta_k^2 \widehat{P}_k \nabla U(\widehat{p}_k) + \sqrt{2\beta^{-1}\Delta t}\, \vartheta_k \widehat{v}_k - \beta^{-1} \Delta t\, \vartheta_k^2 \widehat{\mathrm{II}}_{\widehat{p}_k}(\widehat{v}_k, \widehat{v}_k) \Big]. \tag{19}$$

The landing term stabilizes deviations from the learned manifold, while the directional curvature term corrects the finite-step normal displacement associated with the *same* Gaussian direction that generates the stochastic increment. The factors ϑ_k^2 match the covariance scaling induced by the stochastic amplitude ϑ_k. When $\widehat{\mathsf{d}}_k \le \rho_0$ and $\chi_k = 1$, the localization is inactive and equation 19 reduces exactly to the nominal IIL update,

$$X^{\mathrm{II}L,\mathrm{nom}}_{k+1} = X_k \underbrace{-\Delta t\, \widehat{P}_k \nabla U(\widehat{p}_k) + \sqrt{2\beta^{-1}\Delta t}\, \widehat{v}_k}_{\text{tangent Euler step}} - \underbrace{\beta^{-1} \Delta t\, \widehat{\mathrm{II}}_{\widehat{p}_k}(\widehat{v}_k, \widehat{v}_k)}_{\text{pathwise curvature correction}} + \underbrace{\ell \Delta t\,(\widehat{p}_k - X_k)}_{\text{landing term}}. \tag{20}$$

The localization is needed only for the infinite-horizon analysis under local geometric-oracle assumptions; its forward-invariance and invariant-law properties are established in Appendix C. Analogous localized versions of the reference schemes are deferred in Appendix C.

C LOCALIZED THEORY: PRELIMINARIES

The main text states only the scheme-level oracle errors needed to expose the leading rates. We collect here the regularity and localization assumptions used to make the infinite-horizon argument rigorous. Throughout the appendix, the Hölder exponent $\alpha \in (0,1)$ is fixed as in Section 3.2.

Assumption 5 (Geometric and analytic regularity)**.** The manifold $\mathcal{M} \subset \mathbb{R}^d$ is compact, connected, boundaryless, and C^5. There exists $\tau > 0$ such that the nearest-point projection $\pi : \mathcal{T}_\tau \to \mathcal{M}$ is single-valued and C^4. The potential U has a C^4 extension to $\mathcal{T}_\tau$. All derivatives of π and U used below are uniformly bounded on compact sub-tubes of $\mathcal{T}_\tau$.

For $p \in \mathcal{M}$, the sign convention used throughout is

$$\mathrm{II}_p(u,w) := -N(p)DP(p)[u]w, \qquad H(p) := -\sum_{j=1}^{r} \mathrm{II}_p(e_j,e_j), \tag{21}$$

where $N(p) = I - P(p)$ and $\{e_j\}_{j=1}^r$ is any orthonormal basis of $T_p\mathcal{M}$.

Proposition 6 (Pathwise curvature correction)**.** *Under the geometric regularity assumptions of Appendix C, there exists $\rho > 0$ such that, uniformly for $p \in \mathcal{M}$, $a \in [0,1]$, and $u_a := \sqrt{2\beta^{-1}\Delta t}\, aP(p)\xi$ with $\|u_a\| \le \rho$, we have*

$$p + u_a - \pi(p+u_a) = \beta^{-1}\Delta t\, a^2 \mathrm{II}_p(P(p)\xi, P(p)\xi) + \mathcal{O}\Big((\beta^{-1}\Delta t)^{3/2} a^3 \|\xi\|^3\Big). \tag{22}$$

Moreover, for $\xi \sim \mathcal{N}(0, I_d)$,

$$\mathbb{E}_\xi[-\mathrm{II}_p(P(p)\xi, P(p)\xi)] = H(p). \tag{23}$$

For equation 19, taking $a = \vartheta_k$ shows that scaling the stochastic increment by ϑ_k scales its leading curvature displacement by ϑ_k^2, exactly matching the proposed correction. Proposition 6 identifies two distinct geometric roles. Under exact geometry, the pathwise curvature correction term in equation 20 cancels the leading normal residual of the *realized* tangent proposal, so that the corrected additive update matches the local retraction through quadratic order. In contrast, the classical mean-curvature drift retains only the Gaussian mean of this directional correction.

Assumption 7 (Oracle regularity and localization parameters)**.** Choose fixed constants

$$0 < 2\varepsilon_{\pi,0} < \rho_0 < \rho_1, \qquad \rho_1 + \varepsilon_{\pi,0} < \tau_\star < \tau_{\mathrm{or}} < \tau, \tag{24}$$

and restrict $0 \le \varepsilon_\pi \le \varepsilon_{\pi,0}$. Let $\vartheta \in C^\infty([0,\infty);[0,1])$ satisfy $\vartheta(s) = 1$ for $s \le \rho_0$ and $\vartheta(s) = 0$ for $s \ge \rho_1$. Fix $0 < \gamma < \frac{1}{2}$ and define

$$R_{h,d} := \sqrt{d} + h^{-\gamma}, \qquad G_h := \{\|\xi\| \le R_{h,d}\}. \tag{25}$$

Fix a finite landing cap $\ell_+ > 0$, assume $0 \le \ell \le \ell_+$, and take h sufficiently small that $\ell_+ h \le 1$.

In addition to the approximation bounds in Assumption 1, assume that the learned maps are continuous on the compact query sets and that, for fixed finite constants $C_P, C_H, C_{\mathrm{II}}$,

$$\|\widehat{P}\|_{\mathrm{op}} \le C_P, \qquad \|\widehat{H}\| \le C_H, \qquad \|\widehat{\mathrm{II}}_{\widehat{p}}(w,w)\| \le C_{\mathrm{II}}\|w\|^2.$$

For the diagonal curvature map, joint continuity in the base point and w is assumed on the compact sets on which it is queried. These conditions are satisfied by the finite-difference construction in Appendix J along the admissible parameter regime stated there.

Throughout the appendix set $h := \Delta t$ and $\sigma := \sqrt{2\beta^{-1}}$. For a current state x, write

$$\widehat{p} := \widehat{\pi}(x), \qquad \widehat{P} := \widehat{P}(\widehat{p}), \qquad \widehat{v} := \widehat{P}\xi,$$

and define

$$\widehat{\mathsf{d}}(x) := \|x - \widehat{\pi}(x)\|, \qquad \vartheta_x := \vartheta(\widehat{\mathsf{d}}(x)), \qquad \chi_h(\xi) := \mathbf{1}_{G_h}. \tag{26}$$

For completeness, the localized transition maps used in the stationary analysis are

$$\Phi_h^{\mathrm{IIL,loc}}(x,\xi) = x + \ell h(\widehat{p} - x) + \chi_h(\xi)\Big[-h\vartheta_x^2 \widehat{P}\nabla U(\widehat{p}) + \sigma\sqrt{h}\,\vartheta_x \widehat{v} - \beta^{-1} h \vartheta_x^2 \widehat{\mathrm{II}}_{\widehat{p}}(\widehat{v},\widehat{v})\Big], \tag{27}$$

$$\Phi_h^{HL,\mathrm{loc}}(x,\xi) = x + \ell h(\widehat{p} - x) + \chi_h(\xi)\Big[-h\vartheta_x^2 \widehat{P}\nabla U(\widehat{p}) + \sigma\sqrt{h}\,\vartheta_x \widehat{v} + \beta^{-1} h \vartheta_x^2 \widehat{H}(\widehat{p})\Big]. \tag{28}$$

The first map is exactly the safeguarded IIL update in equation 19, after identifying $h = \Delta t$. For the inexact hard-projection comparison, define

$$Y_h^{\pi,\mathrm{loc}}(x,\xi) := x + \chi_h(\xi)\left[-h\vartheta_x^2 \widehat{P}\nabla U(\widehat{p}) + \sigma\sqrt{h}\,\vartheta_x \widehat{v}\right], \tag{29}$$

and

$$\Phi_h^{\pi,\mathrm{loc}}(x,\xi) := \widehat{\pi}\Big(Y_h^{\pi,\mathrm{loc}}(x,\xi)\Big). \tag{30}$$

For $s \in \{\mathrm{II}L, HL, \pi\}$, let $K_h^{s,\mathrm{loc}}$ be the Markov kernel induced by $\Phi_h^{s,\mathrm{loc}}$. More explicitly, for every Borel set $A \subseteq \mathcal{K}_{\tau_\star}$,

$$K_h^{s,\mathrm{loc}}(x,A) := \mathbb{P}_\xi\Big(\Phi_h^{s,\mathrm{loc}}(x,\xi) \in A\Big).$$

Its action on a bounded measurable function f is

$$(K_h^{s,\mathrm{loc}} f)(x) := \int_{\mathcal{K}_{\tau_\star}} f(y)\, K_h^{s,\mathrm{loc}}(x,dy) = \mathbb{E}_\xi\Big[f\Big(\Phi_h^{s,\mathrm{loc}}(x,\xi)\Big)\Big].$$

We define the corresponding discrete generator by

$$\mathcal{L}_h^{s,\mathrm{loc}} := \frac{K_h^{s,\mathrm{loc}} - I}{h}. \tag{31}$$

For the finite-horizon comparison only, define the corresponding nominal learned-geometry transitions by removing ϑ_x and χ_h from the active terms; for IIL this is equation 20, while for HL and π it is the direct learned-oracle analogue of the reference updates in equation 10 and equation 9.

Throughout this appendix, Assumptions 5– 7 are in force. Constants denoted by C may change from line to line. They may depend on the fixed manifold, potential, inverse temperature, localization radii, ambient dimension, and fixed landing cap ℓ_+, but not on h or on the oracle errors within their prescribed caps.

For $x \in \mathcal{T}_\tau$, write

$$p = \pi(x), \qquad r(x) = x - \pi(x), \qquad q(x) = \|r(x)\|^2 = \mathrm{dist}(x,\mathcal{M})^2. \tag{32}$$

C.1 Residual certificate and truncated Gaussian moments

Lemma 8 (Learned-residual certificate)**.** *For every $x \in \mathcal{K}_{\tau_{\mathrm{or}}}$,*

$$|\widehat{\mathsf{d}}(x) - \mathrm{dist}(x,\mathcal{M})| \le \varepsilon_\pi. \tag{33}$$

Consequently,

$$\vartheta_x > 0 \quad \Longrightarrow \quad \mathrm{dist}(x,\mathcal{M}) < \rho_1 + \varepsilon_{\pi,0} < \tau_\star. \tag{34}$$

Moreover, with $\delta_0 = \rho_0 - \varepsilon_{\pi,0} > 0$,

$$1 - \vartheta_x^2 \le \mathbf{1}_{\{\widehat{\mathsf{d}}(x) > \rho_0\}} \le \mathbf{1}_{\{\mathrm{dist}(x,\mathcal{M}) > \delta_0\}}. \tag{35}$$

Proof. The reverse triangle inequality and the projection-oracle bound give

$$\begin{aligned} |\widehat{\mathsf{d}}(x) - \mathrm{dist}(x,\mathcal{M})| &= \big|\|x - \widehat{\pi}(x)\| - \|x - \pi(x)\|\big| \\ &\le \|\widehat{\pi}(x) - \pi(x)\| \le \varepsilon_\pi. \end{aligned}$$

If $\vartheta_x > 0$, then $\widehat{\mathsf{d}}(x) < \rho_1$, which proves equation 34. If $1 - \vartheta_x^2 > 0$, then $\widehat{\mathsf{d}}(x) > \rho_0$; applying the first inequality and $\varepsilon_\pi \le \varepsilon_{\pi,0}$ yields $\mathrm{dist}(x,\mathcal{M}) > \delta_0$. □

For the dimension-aware threshold in equation 25, let

$$p_h := \mathbb{P}(G_h), \qquad a_h := \mathbb{E}[\xi_1^2 \mathbf{1}_{G_h}], \qquad \xi \sim \mathcal{N}(0, I_d), \tag{36}$$

where ξ_1 denotes any coordinate of the isotropic Gaussian vector ξ; by rotational symmetry the definition is independent of the coordinate choice.

Lemma 9 (Weighted Gaussian tails and symmetric moments)**.** *For every integer $k, m \geq 0$, there exists $C_{k,m,d} < \infty$ such that, for all sufficiently small h,*

$$\mathbb{E}\big[(1 + \|\xi\|^k)\mathbf{1}_{G_h^c}\big] \leq C_{k,m,d} h^m. \tag{37}$$

The event G_h is radial and symmetric, so

$$\mathbb{E}[\xi \mathbf{1}_{G_h}] = 0, \qquad \mathbb{E}[\xi^{\otimes 3} \mathbf{1}_{G_h}] = 0, \tag{38}$$

and rotational invariance gives

$$\mathbb{E}[\xi\xi^\top \mathbf{1}_{G_h}] = a_h I_d, \qquad |1 - p_h| + |1 - a_h| \leq C_{m,d} h^m. \tag{39}$$

In particular, for any matrix A independent of ξ, conditionally on A,

$$\mathbb{E}[(A\xi)(A\xi)^\top \mathbf{1}_{G_h} \mid A] = a_h A A^\top. \tag{40}$$

Proof. Recall that

$$G_h = \left\{ \|\xi\| \leq \sqrt{d} + h^{-\gamma} \right\}, \qquad \xi \sim \mathcal{N}(0, I_d).$$

We first establish the weighted tail estimate. The map $z \mapsto \|z\|$ is 1-Lipschitz, since

$$\big| \|z\| - \|z'\| \big| \leq \|z - z'\|.$$

Moreover, by Jensen's inequality,

$$\mathbb{E}\|\xi\| \leq \big(\mathbb{E}\|\xi\|^2\big)^{1/2} = \sqrt{d}.$$

Hence Gaussian concentration for 1-Lipschitz functions gives

$$\begin{aligned} \mathbb{P}(G_h^c) &= \mathbb{P}\Big(\|\xi\| > \sqrt{d} + h^{-\gamma} \Big) \\ &\leq \mathbb{P}\big(\|\xi\| > \mathbb{E}\|\xi\| + h^{-\gamma} \big) \\ &\leq \exp\left(-\frac{1}{2} h^{-2\gamma} \right). \end{aligned}$$

For any integer $k \geq 0$, Cauchy–Schwarz yields

$$\begin{aligned} \mathbb{E}\big[(1 + \|\xi\|^k)\mathbf{1}_{G_h^c}\big] &\leq \big(\mathbb{E}(1 + \|\xi\|^k)^2\big)^{1/2} \left(\mathbb{E}\mathbf{1}_{G_h^c}^2\right)^{1/2} \\ &= \big(\mathbb{E}(1 + \|\xi\|^k)^2\big)^{1/2} \mathbb{P}(G_h^c)^{1/2}. \end{aligned}$$

Since $\mathbf{1}_{G_h^c}^2 = \mathbf{1}_{G_h^c}$ and all Gaussian polynomial moments are finite,

$$C_{k,d} := \big(\mathbb{E}(1 + \|\xi\|^k)^2\big)^{1/2} < \infty.$$

Therefore

$$\mathbb{E}\big[(1 + \|\xi\|^k)\mathbf{1}_{G_h^c}\big] \leq C_{k,d} \exp\left(-\frac{1}{4} h^{-2\gamma} \right).$$

For every fixed integer $m \geq 0$,

$$\exp\left(-\frac{1}{4} h^{-2\gamma} \right) \leq C_{m,\gamma} h^m$$

for all sufficiently small h. Indeed, setting $s = h^{-2\gamma}$ gives

$$h^m = s^{-m/(2\gamma)},$$

and

$$\sup_{s \geq 1} s^{m/(2\gamma)} e^{-s/4} < \infty.$$

Thus

$$\mathbb{E}\big[(1 + \|\xi\|^k)\mathbf{1}_{G_h^c}\big] \leq C_{k,m,d} h^m,$$

which proves equation 37.

We next prove the symmetry identities. Since both the standard Gaussian law and the event G_h are invariant under $\xi \mapsto -\xi$,

$$\mathbb{E}[\xi \mathbf{1}_{G_h}] = \mathbb{E}[-\xi \mathbf{1}_{G_h}] = -\mathbb{E}[\xi \mathbf{1}_{G_h}],$$

and hence

$$\mathbb{E}[\xi \mathbf{1}_{G_h}] = 0.$$

The same argument applies to the third tensor power because $(-\xi)^{\otimes 3} = -\xi^{\otimes 3}$, giving

$$\mathbb{E}[\xi^{\otimes 3} \mathbf{1}_{G_h}] = 0.$$

Now consider

$$M_h := \mathbb{E}[\xi \xi^\top \mathbf{1}_{G_h}].$$

For $i \neq j$, reflection of the i-th coordinate leaves both the Gaussian law and G_h unchanged but changes the sign of $\xi_i \xi_j$. Therefore

$$\mathbb{E}[\xi_i \xi_j \mathbf{1}_{G_h}] = 0, \qquad i \neq j.$$

Furthermore, permutation of any two coordinates leaves both the Gaussian law and G_h unchanged, so all diagonal entries are equal:

$$\mathbb{E}[\xi_i^2 \mathbf{1}_{G_h}] = \mathbb{E}[\xi_1^2 \mathbf{1}_{G_h}] = a_h.$$

Consequently,

$$\mathbb{E}[\xi \xi^\top \mathbf{1}_{G_h}] = a_h I_d.$$

Finally,

$$1 - p_h = \mathbb{P}(G_h^c),$$

while, since $\mathbb{E}\xi_1^2 = 1$,

$$\begin{aligned} 1 - a_h &= \mathbb{E}[\xi_1^2 \mathbf{1}_{G_h^c}] \\ &\leq \mathbb{E}[\|\xi\|^2 \mathbf{1}_{G_h^c}]. \end{aligned}$$

Applying equation 37 with $k = 0$ and $k = 2$ therefore gives, for every fixed m,

$$|1 - p_h| + |1 - a_h| \leq C_{m,d} h^m.$$

For any deterministic matrix A,

$$\begin{aligned} \mathbb{E}[(A\xi)(A\xi)^\top \mathbf{1}_{G_h}] &= A\, \mathbb{E}[\xi \xi^\top \mathbf{1}_{G_h}]\, A^\top \\ &= a_h A A^\top, \end{aligned}$$

which proves the final claim. □

C.2 Forward invariance, Feller property, and invariant laws

Lemma 10 (Deterministic forward invariance)**.** *There exists $h_0 > 0$, uniform over all oracle errors satisfying Assumption 1 and the bounds of Assumption 7, such that for every $0 < h \leq h_0$, every $0 \leq \ell \leq \ell_+$, every $x \in \mathcal{K}_{\tau_\star}$, and every $\xi \in \mathbb{R}^d$,*

$$\Phi_h^{s,\mathrm{loc}}(x, \xi) \in \mathcal{K}_{\tau_\star}, \qquad s \in \{\mathrm{III}, \mathrm{HL}, \pi\}.$$

Moreover, every argument at which the scheme-level oracle bounds of Assumption 1 are invoked lies in $\mathcal{K}_{\tau_{\mathrm{or}}}$. The one-step line segments used below for the additive schemes, and the preprojection segments for the π-scheme, remain in a fixed compact sub-tube of $\mathcal{T}_\tau$.

Proof. We first prove the claim for the localized III transition. Fix $x \in \mathcal{K}_{\tau_\star}$ and $\xi \in \mathbb{R}^d$, and define the one-step output

$$x^+ := \Phi_h^{\mathrm{III},\mathrm{loc}}(x, \xi).$$

Set

$$p := \pi(x), \qquad \widehat{p} := \widehat{\pi}(x), \qquad \widehat{v} := \widehat{P}(\widehat{p})\xi.$$

Since $\mathcal{K}_{\tau_\star} \subset \mathcal{K}_{\tau_{\mathrm{or}}}$, the projection-oracle bound applies at x, and hence

$$\|\widehat{p} - p\| \leq \varepsilon_\pi \leq \varepsilon_{\pi,0}.$$

In particular,

$$\mathrm{dist}(\widehat{p}, \mathcal{M}) \leq \varepsilon_{\pi,0} < \tau_{\mathrm{or}} < \tau,$$

so the learned geometric quantities appearing in the update are evaluated inside their prescribed query region. Shrink h_0, if necessary, so that $\ell_+ h \leq 1$ for every $0 < h \leq h_0$.

Case 1: fallback branch. Suppose that either $\chi_h(\xi) = 0$ or $\vartheta_x = 0$. Then all active drift, noise, and curvature terms vanish, and

$$x^+ = x + \ell h(\widehat{p} - x) = (1 - \ell h)x + \ell h\widehat{p}.$$

Since $0 \le \ell h \le 1$,

$$\begin{aligned}\operatorname{dist}(x^+, \mathcal{M}) &\le \|x^+ - p\| \\ &= \|(1 - \ell h)(x - p) + \ell h(\widehat{p} - p)\| \\ &\le (1 - \ell h)\operatorname{dist}(x, \mathcal{M}) + \ell h\|\widehat{p} - p\| \\ &\le (1 - \ell h)\tau_\star + \ell h\varepsilon_{\pi,0} \le \tau_\star.\end{aligned}$$

Thus $x^+ \in \mathcal{K}_{\tau_\star}$.

The same argument applies to every point

$$x_t := x + t(x^+ - x) = (1 - t\ell h)x + t\ell h\widehat{p}, \qquad 0 \le t \le 1,$$

and gives

$$\operatorname{dist}(x_t, \mathcal{M}) \le \tau_\star.$$

Hence the entire update segment remains in $\mathcal{K}_{\tau_\star} \Subset \mathcal{T}_\tau$.

Case 2: active $\mathrm{II}L$ branch. Suppose that $\chi_h(\xi) = 1$ and $\vartheta_x > 0$. By Lemma 8,

$$\operatorname{dist}(x, \mathcal{M}) < \rho_1 + \varepsilon_{\pi,0}.$$

Define the fixed positive margin

$$\delta_\star := \tau_\star - (\rho_1 + \varepsilon_{\pi,0}) > 0.$$

Thus

$$\operatorname{dist}(x, \mathcal{M}) < \tau_\star - \delta_\star.$$

On this branch, the one-step increment is

$$\begin{aligned}x^+ - x = \ell h(\widehat{p} - x) &- h\vartheta_x^2\widehat{P}\nabla U(\widehat{p}) \\ &+ \sigma\sqrt{h}\,\vartheta_x\widehat{v} - \beta^{-1}h\vartheta_x^2\widehat{\mathrm{II}}_{\widehat{p}}(\widehat{v}, \widehat{v}).\end{aligned}$$

We bound these four terms separately.

Since $\vartheta_x > 0$, $\widehat{\mathsf{d}}(x) = \|x - \widehat{p}\| < \rho_1$, so

$$\|\ell h(\widehat{p} - x)\| \le \ell_+\rho_1 h.$$

Uniform boundedness of $\widehat{P}$ and ∇U on the fixed query region gives

$$h\vartheta_x^2\|\widehat{P}\nabla U(\widehat{p})\| \le Ch.$$

Moreover,

$$\|\widehat{v}\| \le \|\widehat{P}\|_{\mathrm{op}}\|\xi\| \le C_P\|\xi\|,$$

and therefore

$$\sigma\sqrt{h}\,\vartheta_x\|\widehat{v}\| \le C\sqrt{h}\,\|\xi\|.$$

Finally, the quadratic-growth assumption on the learned second fundamental form gives

$$\begin{aligned}\beta^{-1}h\vartheta_x^2\left\|\widehat{\mathrm{II}}_{\widehat{p}}(\widehat{v}, \widehat{v})\right\| &\le Ch\|\widehat{v}\|^2 \\ &\le Ch\|\xi\|^2.\end{aligned}$$

Combining the four bounds,

$$\|x^+ - x\| \le C\left(h + \sqrt{h}\,\|\xi\| + h\|\xi\|^2\right). \tag{41}$$

Because $\chi_h(\xi) = 1$, we have $\|\xi\| \le R_{h,d} = \sqrt{d} + h^{-\gamma}$. Hence

$$\begin{aligned}\|x^+ - x\| &\le C\left(h + \sqrt{h}\,R_{h,d} + hR_{h,d}^2\right) \\ &\le C\left(h + \sqrt{dh} + h^{1/2-\gamma} + dh + 2\sqrt{d}\,h^{1-\gamma} + h^{1-2\gamma}\right).\end{aligned}$$

Since $0 < \gamma < 1/2$, every term on the right tends to zero as $h \downarrow 0$. We may therefore choose $h_0 > 0$, depending only on the fixed bounds and localization radii, such that

$$\|x^+ - x\| \le \frac{\delta_\star}{2} \qquad \text{for all } 0 < h \le h_0.$$

Using the 1-Lipschitz property of the distance function,

$$\begin{aligned}\operatorname{dist}(x^+, \mathcal{M}) &\le \operatorname{dist}(x, \mathcal{M}) + \|x^+ - x\| \\ &< \tau_\star - \delta_\star + \frac{\delta_\star}{2} = \tau_\star - \frac{\delta_\star}{2} < \tau_\star.\end{aligned}$$

Thus $x^+ \in \mathcal{K}_{\tau_\star}$.

Furthermore, for every $t \in [0, 1]$,

$$\begin{aligned}\operatorname{dist}\big(x + t(x^+ - x), \mathcal{M}\big) &\le \operatorname{dist}(x, \mathcal{M}) + t\|x^+ - x\| \\ &< \tau_\star - \frac{\delta_\star}{2}.\end{aligned}$$

Hence the whole active update segment lies in the fixed compact sub-tube $\mathcal{K}_{\tau_\star - \delta_\star/2} \Subset \mathcal{T}_\tau$.

For HL, the fallback branch is identical. On the active branch, the landing, potential-drift, and Gaussian terms satisfy the same bounds, while

$$\beta^{-1} h \vartheta_x^2 \|\widehat{H}(\widehat{p})\| \le Ch$$

by uniform boundedness of $\widehat{H}$. Hence its active increment is bounded by $C(h + \sqrt{h}\, R_{h,d})$, and the same margin argument gives forward invariance and segment containment.

For π, let $Y = Y_h^{\pi,\mathrm{loc}}(x, \xi)$ denote the preprojection proposal. If $\chi_h(\xi) = 0$ or $\vartheta_x = 0$, then $Y = x$. Otherwise, $\vartheta_x > 0$ gives the same fixed margin from the boundary and

$$\|Y - x\| \le C(h + \sqrt{h}\, R_{h,d}),$$

so $Y \in \mathcal{K}_{\tau_\star}$ for sufficiently small h. Since $\mathcal{K}_{\tau_\star} \subset \mathcal{K}_{\tau_{\mathrm{or}}}$, the projection-oracle bound applies at Y, and therefore

$$\operatorname{dist}\big(\widehat{\pi}(Y), \mathcal{M}\big) \le \|\widehat{\pi}(Y) - \pi(Y)\| \le \varepsilon_\pi < \tau_\star.$$

Thus $\Phi_h^{\pi,\mathrm{loc}}(x, \xi) \in \mathcal{K}_{\tau_\star}$. This proves the claimed deterministic forward invariance. □

Remark (query domains for finite-difference curvature). Lemma 10 concerns the scheme-level oracle queries appearing in the localized Markov transitions. When $\widehat{\mathrm{II}}$ or $\widehat{H}$ is instantiated by the finite-difference construction of Appendix J, evaluating one such curvature oracle additionally requires first-order queries of $\widehat{J}$ at shifted points

$$\widehat{p} \pm \delta \widehat{q}, \qquad \|\widehat{q}\| = 1.$$

These auxiliary query points need not belong to $\mathcal{K}_{\tau_{\mathrm{or}}}$, and their containment is not part of the forward-invariance argument above. Instead, Assumption 37 introduces the separate compact query region $\mathcal{K}_{\tau_{\mathrm{fd}}}$. Indeed, under that assumption,

$$\|\widehat{p} - p\| \le \eta_\pi, \qquad p \in \mathcal{M},$$

and therefore, for $0 < \delta \le \delta_{\mathrm{fd}}$,

$$\begin{aligned}\operatorname{dist}\big(\widehat{p} \pm \delta \widehat{q}, \mathcal{M}\big) &\le \|\widehat{p} - p\| + \delta \\ &\le \eta_\pi + \delta_{\mathrm{fd}} < \tau_{\mathrm{fd}}.\end{aligned}$$

Hence all shifted Jacobian evaluations used by the finite-difference curvature construction lie in $\mathcal{K}_{\tau_{\mathrm{fd}}}$, where the first-order oracle error bound of Assumption 37 is valid.

The following well-posedness result is used only to ensure that the localized numerical kernels possess stationary laws: continuity of the transitions yields the Feller property, while forward invariance of the compact tube then gives existence of an invariant measure.

A probability measure ρ on $\mathcal{K}_{\tau_\star}$ is invariant for a Markov kernel K if

$$\rho K = \rho,$$

where

$$(\rho K)(A) := \int_{\mathcal{K}_{\tau_\star}} K(x, A)\, \rho(dx).$$

Equivalently,

$$\rho(Kf) = \rho(f)$$

for every bounded measurable f. We reserve μ for the target Gibbs probability measure on $\mathcal{M}$ and use $\rho_h^{s,\mathrm{loc}}$ for invariant probability measures of the localized numerical kernels.

Proposition 11 (Feller kernels and invariant laws)**.** *For every $0 < h \le h_0$, each localized kernel*

$$K_h^{s,\mathrm{loc}}, \qquad s \in \{\mathrm{IIL}, \mathrm{HL}, \pi\},$$

is a Feller Markov kernel on the compact metric space $\mathcal{K}_{\tau_\star}$. Consequently, each kernel admits at least one invariant probability measure

$$\rho_h^{s,\mathrm{loc}} \in \mathcal{P}(\mathcal{K}_{\tau_\star}).$$

Proof. By Lemma 10,

$$\Phi_h^{s,\mathrm{loc}}(x, \xi) \in \mathcal{K}_{\tau_\star}$$

for every $x \in \mathcal{K}_{\tau_\star}$ and every $\xi \in \mathbb{R}^d$. Hence $K_h^{s,\mathrm{loc}}$ defines a Markov kernel on $\mathcal{K}_{\tau_\star}$.

We prove the Feller property. Fix $\xi \in \mathbb{R}^d$. Although $\chi_h(\xi)$ is discontinuous as a function of ξ, for fixed ξ it is simply a scalar independent of x. By the continuity assumptions on $\widehat{\pi}$, $\widehat{P}$, $\widehat{H}$, the diagonal map $(\widehat{p}, w) \mapsto \widehat{\Pi}_{\widehat{p}}(w, w)$, and the cutoff ϑ, the maps

$$x \longmapsto \Phi_h^{s,\mathrm{loc}}(x, \xi)$$

are continuous on $\mathcal{K}_{\tau_\star}$ for $s \in \{\mathrm{IIL}, \mathrm{HL}, \pi\}$. For fixed x, these transition maps are also Borel measurable in ξ, so the corresponding expectations are well defined.

Let $x_n \to x$ in $\mathcal{K}_{\tau_\star}$ and $f \in C(\mathcal{K}_{\tau_\star})$. Then, for every fixed ξ,

$$f\Big(\Phi_h^{s,\mathrm{loc}}(x_n, \xi)\Big) \longrightarrow f\Big(\Phi_h^{s,\mathrm{loc}}(x, \xi)\Big).$$

Since $\mathcal{K}_{\tau_\star}$ is compact, f is bounded, and Lemma 10 gives

$$\Big|f\Big(\Phi_h^{s,\mathrm{loc}}(x_n, \xi)\Big)\Big| \le \|f\|_\infty.$$

Dominated convergence therefore yields

$$\begin{aligned} K_h^{s,\mathrm{loc}} f(x_n) &= \mathbb{E}_\xi\left[f\Big(\Phi_h^{s,\mathrm{loc}}(x_n, \xi)\Big)\right] \\ &\longrightarrow \mathbb{E}_\xi\left[f\Big(\Phi_h^{s,\mathrm{loc}}(x, \xi)\Big)\right] = K_h^{s,\mathrm{loc}} f(x). \end{aligned}$$

Thus

$$K_h^{s,\mathrm{loc}} : C(\mathcal{K}_{\tau_\star}) \longrightarrow C(\mathcal{K}_{\tau_\star}),$$

so the kernel is Feller.

Finally, $\mathcal{K}_{\tau_\star}$ is compact because $\mathcal{M}$ is compact and $\mathcal{K}_{\tau_\star}$ is its closed $\tau_\star$-neighborhood in $\mathbb{R}^d$. The Krylov–Bogolyubov theorem therefore guarantees at least one invariant probability measure $\rho_h^{s,\mathrm{loc}} \in \mathcal{P}(\mathcal{K}_{\tau_\star})$. □

D PROJECTION AND CURVATURE IDENTITIES

For $p \in \mathcal{M}$, recall $P(p) = D\pi(p)$ and $N(p) = I - P(p)$.

Lemma 12 (Local projection expansion)**.** *There exist $\rho, C > 0$ such that, for every $p \in \mathcal{M}$ and $u \in T_p\mathcal{M}$ with $\|u\| \le \rho$,*

$$\pi(p+u) = p + u - \frac{1}{2}\mathrm{II}_p(u,u) + R_3(p,u), \qquad \|R_3(p,u)\| \le C\|u\|^3. \tag{42}$$

Proof. Taylor expansion gives $\pi(p+u) = p + D\pi(p)u + \frac{1}{2}D^2\pi(p)[u,u] + R_3$. Since $D\pi(p) = P(p)$ and u is tangent, $D\pi(p)u = u$. Differentiating $P^2 = P$ in the tangent direction u shows that $DP(p)[u]u$ is normal. By equation 21, $D^2\pi(p)[u,u] = DP(p)[u]u = -\mathrm{II}_p(u,u)$. Uniformity of the remainder follows from compactness and the C^3 bound on π. □

Lemma 13 (Mean curvature as the averaged pathwise correction)**.** *If $v = P(p)\xi$ with $\xi \sim \mathcal{N}(0, I_d)$, then*

$$\mathbb{E}[-\mathrm{II}_p(v,v)] = H(p). \tag{43}$$

For the truncated event G_h,

$$\mathbb{E}[-\mathrm{II}_p(v,v)\mathbf{1}_{G_h}] = a_h H(p). \tag{44}$$

Proof. In an orthonormal tangent basis, write $v = \sum_i \eta_i e_i$. The full Gaussian identity follows from $\mathbb{E}[\eta_i\eta_j] = \delta_{ij}$ and equation 21. On G_h, Lemma 9 gives $\mathbb{E}[vv^\top \mathbf{1}_{G_h}] = a_h P(p)$, and the same trace calculation yields the second identity. □

Proof of Proposition 6. Apply Lemma 12 with $u = u_a = \sqrt{2\beta^{-1}h}\, aP(p)\xi$. Since $\frac{1}{2}\mathrm{II}_p(u_a,u_a) = \beta^{-1}h\, a^2\mathrm{II}_p(P(p)\xi, P(p)\xi)$ and the cubic remainder is bounded by $C(\beta^{-1}h)^{3/2}a^3\|\xi\|^3$, rearranging equation 42 gives equation 22 in Appendix C. The identity equation 23 is exactly Lemma 13 without truncation. □

For a normal vector $n \in N_p\mathcal{M}$, let $S_p[n]$ denote the shape operator defined by

$$\langle S_p[n]u, w\rangle = \langle \mathrm{II}_p(u,w), n\rangle. \tag{45}$$

Lemma 14 (Hessian of the nearest projection)**.** *For $z_1, z_2 \in \mathbb{R}^d$,*

$$D^2\pi(p)[z_1,z_2] = -\mathrm{II}_p(Pz_1, Pz_2) - S_p[Nz_1](Pz_2) - S_p[Nz_2](Pz_1). \tag{46}$$

Proof. For tangent arguments, differentiate $D\pi = P$ and use equation 21. For two normal arguments, π is constant along each sufficiently short normal fiber, so the second derivative vanishes. For a mixed pair, differentiate $\pi(\gamma(t) + sn(t)) = \gamma(t)$ along a tangent curve γ and a parallel normal field $n(t)$; the Weingarten identity gives $D^2\pi(p)[n,u] = -S_p[n]u$. Bilinearity completes the proof. □

Lemma 15 (Residual expansion in a tube)**.** *Let $x = p + r \in \mathcal{T}_\tau$, $p = \pi(x)$, and suppose the line segment $x + tz$, $0 \le t \le 1$, lies in a fixed compact subtube. Then*

$$r(x+z) = r(x) + Dr(x)z + \frac{1}{2}D^2r(x)[z,z] + \frac{1}{6}D^3r(x)[z,z,z] + \mathcal{R}_4(x,z), \tag{47}$$

where $\|\mathcal{R}_4(x,z)\| \le C\|z\|^4$. Along a normal fiber,

$$\begin{aligned} Dr(x)n &= n \quad (n \in N_p\mathcal{M}), \quad \|Dr(x)P(p)\|_{\mathrm{op}} \le C\|r\|, \\ \|D^2r(x)[P(p)u, P(p)u] - \mathrm{II}_p(u,u)\| &\le C\|r\|\|u\|^2. \end{aligned} \tag{48}$$

Proof. The Taylor formula follows from $r = I - \pi \in C^4$. Since π is constant on the affine normal fiber $p + N_p\mathcal{M}$ inside the tube, $D\pi(x)n = 0$ for every $n \in N_p\mathcal{M}$; hence $Dr(x)n = n$. At p, $Dr(p)P(p) = 0$ and, by Lemma 14, $D^2r(p)[u,u] = \mathrm{II}_p(u,u)$ for tangent u. Lipschitz continuity of Dr and D^2r between p and $x = p + r$ proves the remaining estimates. □

E TARGET GENERATOR AND POISSON REGULARITY

Lemma 16 (Ambient representation of the target generator)**.** *Let $\bar{\varphi}$ be a smooth extension of $\varphi : \mathcal{M} \to \mathbb{R}$. Then*

$$(\mathcal{L}\varphi)(p) = \nabla\bar{\varphi}(p)^\top \big[-P(p)\nabla U(p) + \beta^{-1}H(p)\big] + \beta^{-1}\operatorname{tr}\big(P(p)\nabla^2\bar{\varphi}(p)\big). \tag{49}$$

For the normal extension $\widetilde{\varphi} = \varphi \circ \pi$, $\nabla\widetilde{\varphi}(p)$ is tangent and therefore

$$(\mathcal{L}\varphi)(p) = -D\widetilde{\varphi}(p)P(p)\nabla U(p) + \beta^{-1}\operatorname{tr}\big(P(p)D^2\widetilde{\varphi}(p)\big). \tag{50}$$

Proof. Choose a local orthonormal tangent frame $\{e_i\}_{i=1}^r$ satisfying $\nabla^{\mathcal{M}}_{e_i} e_i(p) = 0$. Then

$$\Delta_{\mathcal{M}}\varphi(p) = \sum_i e_i^\top \nabla^2\bar{\varphi}(p)e_i + \nabla\bar{\varphi}(p)^\top \sum_i D_{e_i}e_i(p).$$

The first term is $\operatorname{tr}(P\nabla^2\bar{\varphi})$ and the normal trace of the ambient accelerations is H under our sign convention. Substitution into equation 59 proves equation 49. For a normal extension, the gradient is tangent, so its inner product with H vanishes. □

Hölder-space conventions. Fix a Hölder exponent $\alpha \in (0,1)$. For an open set $\Omega \subset \mathbb{R}^d$, define

$$[f]_{C^{0,\alpha}(\Omega)} := \sup_{\substack{x,y\in\Omega \\ x\neq y}} \frac{|f(x)-f(y)|}{\|x-y\|^\alpha},$$

and, for an integer $k \geq 0$,

$$\|f\|_{C^{k,\alpha}(\Omega)} := \sum_{|\nu|\leq k} \|D^\nu f\|_{L^\infty(\Omega)} + \sum_{|\nu|=k} [D^\nu f]_{C^{0,\alpha}(\Omega)}.$$

Here ν denotes a multi-index.

On the compact manifold $\mathcal{M}$, the norm $\|\cdot\|_{C^{k,\alpha}(\mathcal{M})}$ is defined using any fixed finite smooth atlas and subordinate partition of unity. Since $\mathcal{M}$ is compact, different such choices give equivalent norms.

Fix also an open neighborhood $\mathcal{O}$ satisfying

$$\mathcal{K}_{\tau_\star} \Subset \mathcal{O} \Subset \mathcal{T}_\tau.$$

Since $\mathcal{K}_{\tau_\star}$ is closed, we interpret $C^{2,\alpha}(\mathcal{K}_{\tau_\star})$ as the restriction space

$$C^{2,\alpha}(\mathcal{K}_{\tau_\star}) := \left\{\bar{g}|_{\mathcal{K}_{\tau_\star}} : \bar{g} \in C^{2,\alpha}(\mathcal{O})\right\},$$

equipped with the quotient norm

$$\|g\|_{C^{2,\alpha}(\mathcal{K}_{\tau_\star})} := \inf_{\substack{\bar{g}\in C^{2,\alpha}(\mathcal{O}) \\ \bar{g}|_{\mathcal{K}_{\tau_\star}} = g}} \|\bar{g}\|_{C^{2,\alpha}(\mathcal{O})}. \tag{51}$$

For a function f defined on a neighborhood of $\mathcal{K}_{\tau_\star}$, we use

$$\|f\|_{C^4(\mathcal{K}_{\tau_\star})} := \sum_{j=0}^4 \sup_{x\in\mathcal{K}_{\tau_\star}} \|D^j f(x)\|.$$

Lemma 17 (Poisson equation and Schauder estimate)**.** *For every $g \in C^{2,\alpha}(\mathcal{K}_{\tau_\star})$, let*

$$g_{\mathcal{M}} := g|_{\mathcal{M}}.$$

Then the centered Poisson equation

$$\mathcal{L}u_g = g_{\mathcal{M}} - \mu(g_{\mathcal{M}}), \qquad \mu(u_g) = 0, \tag{52}$$

has a unique solution $u_g \in C^{4,\alpha}(\mathcal{M})$ satisfying

$$\|u_g\|_{C^{4,\alpha}(\mathcal{M})} \leq C_{\mathrm{P}}\|g_{\mathcal{M}}\|_{C^{2,\alpha}(\mathcal{M})}. \tag{53}$$

Moreover,

$$\|g_{\mathcal{M}}\|_{C^{2,\alpha}(\mathcal{M})} \leq C\|g\|_{C^{2,\alpha}(\mathcal{K}_{\tau_\star})}. \tag{54}$$

Consequently, the normal extension

$$\widetilde{u}_g := u_g \circ \pi$$

satisfies

$$\|\widetilde{u}_g\|_{C^4(\mathcal{K}_{\tau_\star})} \leq C\|g\|_{C^{2,\alpha}(\mathcal{K}_{\tau_\star})}. \tag{55}$$

Proof. The generator can be written in weighted divergence form as

$$\mathcal{L}f = \beta^{-1} e^{\beta U} \operatorname{div}_{\mathcal{M}} \left(e^{-\beta U} \nabla_{\mathcal{M}} f \right).$$

Hence $\mathcal{L}$ is a smooth uniformly elliptic second-order operator on the compact boundaryless manifold $\mathcal{M}$ and is symmetric in $L^2(\mu)$, where

$$d\mu = Z^{-1} e^{-\beta U} d \operatorname{vol}_{\mathcal{M}}.$$

Since $\mathcal{M}$ is connected, the kernel of $\mathcal{L}$ consists of the constant functions.

Set

$$f_g := g_{\mathcal{M}} - \mu(g_{\mathcal{M}}).$$

By construction, $\mu(f_g) = 0$. Standard elliptic Fredholm theory on compact manifolds therefore gives a solution of

$$\mathcal{L}u_g = f_g$$

unique up to an additive constant; the normalization $\mu(u_g) = 0$ makes the solution unique.

Since $f_g \in C^{2,\alpha}(\mathcal{M})$, global Schauder estimates give

$$u_g \in C^{4,\alpha}(\mathcal{M}), \qquad \|u_g\|_{C^{4,\alpha}(\mathcal{M})} \le C_{\mathrm{P}} \|f_g\|_{C^{2,\alpha}(\mathcal{M})} \le C \|g_{\mathcal{M}}\|_{C^{2,\alpha}(\mathcal{M})}.$$

This proves equation 53.

Next, let $\bar{g} \in C^{2,\alpha}(\mathcal{O})$ be any extension of g. Continuity of restriction to the fixed smooth compact submanifold $\mathcal{M}$ gives

$$\|g_{\mathcal{M}}\|_{C^{2,\alpha}(\mathcal{M})} \le C \|\bar{g}\|_{C^{2,\alpha}(\mathcal{O})}.$$

Taking the infimum over all such extensions yields

$$\|g_{\mathcal{M}}\|_{C^{2,\alpha}(\mathcal{M})} \le C \|g\|_{C^{2,\alpha}(\mathcal{K}_{\tau_\star})}.$$

Finally, π is C^4 on a neighborhood of $\mathcal{K}_{\tau_\star}$. Thus, for $\widetilde{u}_g = u_g \circ \pi$, the chain rule gives

$$\|\widetilde{u}_g\|_{C^4(\mathcal{K}_{\tau_\star})} \le C \|u_g\|_{C^4(\mathcal{M})} \le C \|g\|_{C^{2,\alpha}(\mathcal{K}_{\tau_\star})}.$$

□

Finally we show that for the compact tube $\mathcal{K}_{\tau_\star}$ the dual norm $\|\cdot\|_{C^{2,\alpha}(\mathcal{K})^*}$ that defines the integral probability metric in Section 3.2 via $\operatorname{dist}_{\mathcal{F}}(\nu,\mu) = \|\nu - \mu\|_{C^{2,\alpha}(\mathcal{K})^*}$ is equivalent to the classical W_1-distance up to a power scaling.

Theorem 18. *Let $\mathcal{K} = \mathcal{K}_{\tau_\star} \subseteq \mathcal{T}_\tau$ be a compact subtube of the tubular neighborhood with $\tau_\star < \tau$ and ν, μ be two probability measures with $\operatorname{supp}\nu \subseteq \mathcal{K}$ and $\operatorname{supp}\mu \subseteq \mathcal{M}$. Then there exist some constants $c, C > 0$ depending on d, α, $\tau_\star$ and $R := \operatorname{diam}(\mathcal{K})$ such that*

$$c W_1(\nu,\mu) \le \operatorname{dist}_{\mathcal{F}}(\nu,\mu)^{1/(2+\alpha)} \le C W_1(\nu,\mu)^{1/(2+\alpha)}. \tag{56}$$

Proof. First let us show that first inequality. For brevity write $\sigma := \nu - \mu$ and identify σ with the (signed) functional $\sigma(f) = \int f d\sigma$ for functions f on $\mathbb{R}^d$. When σ is considered as a functional on $C_b(\mathbb{R}^d)$, we have $\|\sigma\|_{C_b(\mathbb{R}^d)^*} \le \|\sigma\|_{\mathrm{TV}} \le 2$ and when σ is seen as a functional on $C_b^{2,\alpha}(\mathbb{R}^d)$, then $\|\sigma\|_{C_b^{2,\alpha}(\mathbb{R}^d)^*} \le \operatorname{dist}_{\mathcal{F}}(\nu,\mu)$. By the classical interpolation theorem (Bergh & Löfström, 2012, Theorem 3.1.2), σ defines a continuous functional on the interpolation space $(C_b(\mathbb{R}^d), C_b^{2,\alpha}(\mathbb{R}^d))_{\theta,\infty}$ for $\theta = 1/(2+\alpha)$ with operator norm $\|\sigma\|_{(C_b(\mathbb{R}^d), C_b^{2,\alpha}(\mathbb{R}^d))^*_{\theta,\infty}} \le \|\sigma\|^{1-\theta}_{C_b(\mathbb{R}^d)^*} \|\sigma\|^{\theta}_{C_b^{2,\alpha}(\mathbb{R}^d)^*} \le 2 \operatorname{dist}_{\mathcal{F}}(\nu,\mu)^{1/(2+\alpha)}$. Further, since C_b^s for $s \in \{0, 2+\alpha\}$ agree with the Zygmund classes $\mathcal{C}_b^s$ (Lunardi et al., 1995, Theorem 1.2.17), the result (Lunardi et al., 1995, Corollary 1.2.18) implies that $(C_b(\mathbb{R}^d), C_b^{2,\alpha}(\mathbb{R}^d))_{\theta,\infty} \cong \mathcal{C}_b^1$ with equivalent norms. This implies also that $\|\cdot\|_{(C_b(\mathbb{R}^d), C_b^{2,\alpha}(\mathbb{R}^d))^*_{\theta,\infty}} \ge c' \|\cdot\|_{(\mathcal{C}_b^1)^*}$ where the constant c' can depend on α, d and R. Next, since $\|f\|_{\mathcal{C}_b^1} \le \|f\|_\infty + \operatorname{Lip}(f)$ (Lunardi et al., 1995, §0.2, proof of Proposition 0.2.2), we have for the W_1-distance that

$$\begin{aligned} W_1(\mu,\nu) &= \sup_{\operatorname{Lip}(f)\le 1} |\sigma(f)| = \sup_{f(x_0)=0\,,\,\operatorname{Lip}(f)\le 1} |\sigma(f)| \le \sup_{\|f\|_{\mathcal{C}_b^1} \le (R+1)} |\sigma(f)| \\ &= (1+R)\|\sigma\|_{(\mathcal{C}_b^1)^*} \le \frac{R+1}{c'} \|\sigma\|_{(C_b(\mathbb{R}^d), C_b^{2,\alpha}(\mathbb{R}^d))^*_{\theta,\infty}} \le \frac{2(R+1)}{c'} \operatorname{dist}_{\mathcal{F}}(\nu,\mu)^{\frac{1}{2+\alpha}}, \end{aligned}$$

where $x_0 \in \mathcal{K}$ is arbitrary and we have used the fact that from $f(x_0) = 0$ and $\mathrm{Lip}(f) \leq 1$ it follows that $\|f\|_\infty \leq R$. Setting $c = \frac{c'}{2(R+1)}$ yields the first inequality. To show the second inequality we first establish that $|f(x)-f(y)| \leq L\|f\|_{C^{2,\alpha}(\mathcal{K})}\|x-y\|$ for any $x \in \mathcal{K}$ and $y \in \mathcal{M}$ and $f \in C^{2,\alpha}(\mathcal{K})$. To see the latter inequality, note that if $x \in \mathcal{K}$ and $y \in \mathcal{M}$ are such that $\|x-y\| < \tau_\star$, then $\|\pi(x) - y\| < 2\tau$ and so $d_{\mathcal{M}}(\pi(x), y) \leq \pi\|x-y\|$ by (Boissonnat et al., 2019, Theorem 1, Lemma 3). Using the mean value theorem along the straight line $x \to_{\mathbb{R}^d} \pi(x)$ and then along the manifold $\pi(x) \to_{\mathcal{M}} y$ we obtain $|f(x) - f(y)| \leq \|\nabla f\|_\infty(\|\pi(x) - x\| + d_{\mathcal{M}}(\pi(x), y)) \leq (1+\pi)\|f\|_{C^{2,\alpha}(\mathcal{K})}\|x-y\|$. If on the other hand $\|x-y\| > \tau_\star$, then trivially $|f(x) - f(y)| \leq 2\|f\|_\infty \leq \frac{2}{\tau_\star}\|f\|_{C^{2,\alpha}(\mathcal{K})}\|x-y\|$. Thus, in all, the claimed inequality holds with $L = \max\{\frac{2}{\tau_\star}, (1+\pi)\}$. Let γ be now any coupling between ν and μ. We obtain

$$|\nu(f) - \mu(f)| = |\int (f(x) - f(y))d\gamma(x,y)| \leq L\|f\|_{C^{2,\alpha}(\mathcal{K})} \int |x-y| d\gamma(x,y)\,.$$

Taking the infimum over γ on the right and supremum over $\|f\|_{C^{2,\alpha}(\mathcal{K})} \leq 1$ on the left yields then the second inequality with $C = L^{1/(2+\alpha)}$. □

F Proof of the projection-based bias decomposition

In this section the goal is to prove the following, formal version of Proposition 2.

Proposition 19 (Projection-based bias decomposition)**.** *Let $\rho_{\Delta t}$ be invariant for a localized kernel $K_{\Delta t}$, and $\mathcal{L}_{\Delta t} = (K_{\Delta t} - I)/\Delta t$. For $g \in \mathcal{F}$, set $g_{\mathcal{M}} = g|_{\mathcal{M}}$ and let u_g solve $\mathcal{L}u_g = g_{\mathcal{M}} - \mu(g_{\mathcal{M}})$, $\mu(u_g) = 0$. With $\widetilde{u}_g = u_g \circ \pi$, stationarity gives*

$$\rho_{\Delta t}(g) - \mu(g_{\mathcal{M}}) = \rho_{\Delta t}(g - g\circ\pi) + \rho_{\Delta t}\big((\mathcal{L}u_g)\circ\pi - \mathcal{L}_{\Delta t}\widetilde{u}_g\big). \tag{57}$$

Define $\mathcal{R}_\perp(X) = \mathrm{dist}(X, \mathcal{M})$ and $\mathcal{R}_{\mathrm{gen}}(X; g) = |(\mathcal{L}u_g)(\pi(X)) - (\mathcal{L}_{\Delta t}\widetilde{u}_g)(X)|$. Then

$$\mathrm{dist}_{\mathcal{F}}(\rho_{\Delta t}, \mu) \leq \mathbb{E}_{\rho_{\Delta t}}\mathcal{R}_\perp + \sup_{g\in\mathcal{F}} \mathbb{E}_{\rho_{\Delta t}}\mathcal{R}_{\mathrm{gen}}(X; g). \tag{58}$$

Here the generator of equation 4 is defined as

$$(\mathcal{L}\varphi)(p) = -\langle \nabla_{\mathcal{M}} U(p), \nabla_{\mathcal{M}}\varphi(p)\rangle + \beta^{-1}\Delta_{\mathcal{M}}\varphi(p), \qquad p \in \mathcal{M}, \tag{59}$$

with invariant Gibbs law μ on $\mathcal{M}$.

Fix an open neighborhood $\mathcal{O}$ satisfying $\mathcal{K}_{\tau_\star} \Subset \mathcal{O} \Subset \mathcal{T}_\tau$. We use the restriction-space norm equation 51 introduced above.

Lemma 20 (Stationarity of the discrete generator)**.** *If K_h has invariant law ρ_h and $\mathcal{L}_h = (K_h - I)/h$, then*

$$\rho_h(\mathcal{L}_h\psi) = 0 \tag{60}$$

for every integrable ψ for which the displayed quantities are finite.

Proof. Invariance gives $\rho_h(K_h\psi) = \rho_h(\psi)$; subtract and divide by h. □

Proof of Proposition 19. Write

$$\rho_h(g) - \mu(g_{\mathcal{M}}) = \rho_h(g - g\circ\pi) + \rho_h(g\circ\pi) - \mu(g_{\mathcal{M}}).$$

Since $g\circ\pi = g_{\mathcal{M}}\circ\pi$ and

$$\mathcal{L}u_g = g_{\mathcal{M}} - \mu(g_{\mathcal{M}}),$$

the second term satisfies

$$\rho_h(g\circ\pi) - \mu(g_{\mathcal{M}}) = \rho_h((\mathcal{L}u_g)\circ\pi)\,.$$

Moreover, stationarity of ρ_h and Lemma 20 give

$$\rho_h(\mathcal{L}_h\widetilde{u}_g) = 0.$$

Therefore

$$\rho_h(g) - \mu(g_{\mathcal{M}}) = \rho_h(g - g \circ \pi) + \rho_h((\mathcal{L}u_g) \circ \pi - \mathcal{L}_h \widetilde{u}_g),$$

which proves equation 57.

It remains to bound the two terms. Let $\bar{g} \in C^{2,\alpha}(\mathcal{O})$ be any extension of g. For $x \in \mathcal{K}_{\tau_\star}$, the normal segment joining x to $\pi(x)$ remains in $\mathcal{K}_{\tau_\star} \subset \mathcal{O}$. The mean-value theorem therefore gives

$$\begin{aligned} |g(x) - g(\pi(x))| &= |\bar{g}(x) - \bar{g}(\pi(x))| \\ &\le \|\nabla \bar{g}\|_{L^\infty(\mathcal{O})} \operatorname{dist}(x, \mathcal{M}) \\ &\le \|\bar{g}\|_{C^{2,\alpha}(\mathcal{O})} \operatorname{dist}(x, \mathcal{M}). \end{aligned}$$

Since this holds for every admissible extension, taking the infimum over all such $\bar{g}$ yields

$$|g(x) - g(\pi(x))| \le \|g\|_{C^{2,\alpha}(\mathcal{K}_{\tau_\star})} \operatorname{dist}(x, \mathcal{M}).$$

For $g \in \mathcal{F}$, the norm on the right is at most one, and hence

$$|\rho_h(g - g \circ \pi)| \le \mathbb{E}_{\rho_h} \mathcal{R}_\perp.$$

The second term is bounded directly by

$$|\rho_h((\mathcal{L}u_g) \circ \pi - \mathcal{L}_h \widetilde{u}_g)| \le \mathbb{E}_{\rho_h} \mathcal{R}_{\mathrm{gen}}(X; g).$$

Combining the two estimates and taking the supremum over $g \in \mathcal{F}$ proves equation 58. □

G One-Step Normal and Generator Estimates

G.1 Localized increments and elementary moments

Fix $x \in \mathcal{K}_{\tau_\star}$, set $p = \pi(x)$, $r = x - p$, $\theta = \vartheta_x$, $\chi = \chi_h(\xi)$, $\sigma = \sqrt{2\beta^{-1}}$, and

$$a(p) := -P(p)\nabla U(p), \qquad v = P(p)\xi. \tag{61}$$

The exact comparison increments use the same learned cutoff scalar θ as the inexact increments:

$$\Delta_{\mathrm{II}} := -\ell h r + \chi\Big[\sqrt{h}\,\sigma\theta v + h\theta^2 a(p) - \beta^{-1} h\theta^2 \mathrm{II}_p(v, v)\Big], \tag{62}$$

$$\Delta_H := -\ell h r + \chi\Big[\sqrt{h}\,\sigma\theta v + h\theta^2 a(p) + \beta^{-1} h\theta^2 H(p)\Big]. \tag{63}$$

The hatted increments are obtained by replacing exact operators by their inexact counterparts and replacing $-r$ in the landing term by $\widehat{p} - x$.

For later weak-generator comparisons, define the exact-operator comparison transitions

$$\Phi_h^{\mathrm{II}L,\mathrm{ex}}(x, \xi) := x + \Delta_{\mathrm{II}}, \qquad \Phi_h^{HL,\mathrm{ex}}(x, \xi) := x + \Delta_H,$$

and let $K_h^{\mathrm{II}L,\mathrm{ex}}$ and $K_h^{HL,\mathrm{ex}}$ denote their induced Markov kernels.

Here "ex" refers to replacing the learned geometric operators by their exact counterparts while retaining the same localization scalars $\theta = \vartheta_x$ and $\chi = \chi_h(\xi)$ used in the learned comparison.

Lemma 21 (Localized increment moments)**.** *Fix $x \in \mathcal{K}_{\tau_\star}$, and let all expectations below be with respect to $\xi \sim \mathcal{N}(0, I_d)$. Uniformly in $x \in \mathcal{K}_{\tau_\star}$ and $s \in \{\mathrm{II}, \mathrm{H}\}$,*

$$\begin{aligned} \mathbb{E}\|\Delta_s\|^2 \le Ch, \quad \mathbb{E}\|\Delta_s\|^4 \le Ch^2, \\ \big\|\mathbb{E}[\Delta_s^{\otimes 3}]\big\| \le Ch^2, \quad \mathbb{E}\|\Delta_s\|^8 \le Ch^4. \end{aligned} \tag{64}$$

The eighth-moment estimate is used only to control the second moment of the fourth-order Taylor remainder in Lemma 23. The same estimates hold for the corresponding hatted increments. In addition,

$$\mathbb{E}[\chi v] = 0, \quad \mathbb{E}[\chi v v^\top] = a_h P(p), \quad \mathbb{E}[\chi v^{\otimes 3}] = 0. \tag{65}$$

Proof. Write

$$\Delta_s = L + S + B_s,$$

where

$$L = -\ell h r, \qquad S = \chi\sigma\sqrt{h}\,\theta v,$$

and B_s is the corresponding order-h drift–curvature term. Set

$$C_s := L + B_s.$$

Since $x \in \mathcal{K}_{\tau_\star}$, the residual r is uniformly bounded. Moreover, $\theta \in [0,1]$, $P(p)$ is an orthogonal projector, and all deterministic drift and curvature coefficients are uniformly bounded on the relevant compact sets. Therefore

$$\|S\| \le C\sqrt{h}\,\|\xi\|.$$

For the IIL increment, bilinearity and uniform boundedness of the second fundamental form give

$$\|B_{\mathrm{II}}\| \le Ch(1 + \|\xi\|^2),$$

whereas for the mean-curvature increment

$$\|B_{\mathrm{H}}\| \le Ch.$$

Consequently, for every integer $1 \le q \le 8$,

$$\mathbb{E}\|S\|^q \le C_q h^{q/2}, \qquad \mathbb{E}\|C_s\|^q \le C_q h^q, \tag{66}$$

using finiteness of all Gaussian moments.

For $q \in \{2, 4, 8\}$, the elementary inequality

$$\|a + b\|^q \le C_q(\|a\|^q + \|b\|^q)$$

and $\Delta_s = S + C_s$ give

$$\begin{aligned}\mathbb{E}\|\Delta_s\|^q &\le C_q\left(\mathbb{E}\|S\|^q + \mathbb{E}\|C_s\|^q\right)\\ &\le C_q\left(h^{q/2} + h^q\right) \le C_q h^{q/2},\end{aligned}$$

for $0 < h \le 1$. This proves the second-, fourth-, and eighth-moment bounds in equation 64.

We next bound the third tensor moment. By multilinearity,

$$\begin{aligned}\mathbb{E}[\Delta_s^{\otimes 3}] = \mathbb{E}[S^{\otimes 3}] &+ \mathrm{Sym}\,\mathbb{E}[S^{\otimes 2} \otimes C_s]\\ &+ \mathrm{Sym}\,\mathbb{E}[S \otimes C_s^{\otimes 2}] + \mathbb{E}[C_s^{\otimes 3}],\end{aligned}$$

where Sym denotes the sum over the corresponding tensor permutations.

Since

$$S = \chi\sigma\sqrt{h}\,\theta P(p)\xi$$

and the truncated Gaussian law is invariant under $\xi \mapsto -\xi$, its third tensor moment vanishes exactly:

$$\mathbb{E}[S^{\otimes 3}] = 0.$$

Using Hölder's inequality and equation 66,

$$\begin{aligned}\left\|\mathbb{E}[S^{\otimes 2} \otimes C_s]\right\| &\le \left(\mathbb{E}\|S\|^4\right)^{1/2}\left(\mathbb{E}\|C_s\|^2\right)^{1/2} \le Ch^2,\\ \left\|\mathbb{E}[S \otimes C_s^{\otimes 2}]\right\| &\le \left(\mathbb{E}\|S\|^2\right)^{1/2}\left(\mathbb{E}\|C_s\|^4\right)^{1/2} \le Ch^{5/2},\\ \left\|\mathbb{E}[C_s^{\otimes 3}]\right\| &\le \mathbb{E}\|C_s\|^3 \le Ch^3.\end{aligned}$$

Hence, for $0 < h \le 1$,

$$\left\|\mathbb{E}[\Delta_s^{\otimes 3}]\right\| \le Ch^2.$$

The same argument applies to the hatted increments. Indeed, conditional on the fixed state x, the learned noise

$$\widehat{v} = \widehat{P}(\widehat{p})\xi$$

is still a deterministic linear image of the centrally symmetric truncated Gaussian. Assumption 7 gives

$$\|\widehat{P}\|_{\mathrm{op}} \le C_P, \qquad \left\|\widehat{\mathrm{II}}_{\widehat{p}}(w,w)\right\| \le C_{\mathrm{II}}\|w\|^2,$$

so the same moment bounds hold uniformly.

Finally, since $v = P(p)\xi$, Lemma 9 gives

$$\begin{aligned}
\mathbb{E}[\chi v] &= P(p)\,\mathbb{E}[\xi \mathbf{1}_{G_h}] = 0,\\
\mathbb{E}[\chi v v^\top] &= P(p)\,\mathbb{E}[\xi\xi^\top \mathbf{1}_{G_h}]\,P(p)^\top\\
&= a_h P(p) P(p)^\top = a_h P(p),\\
\mathbb{E}[\chi v^{\otimes 3}] &= P(p)^{\otimes 3}\mathbb{E}[\xi^{\otimes 3}\mathbf{1}_{G_h}] = 0,
\end{aligned}$$

where we used $P(p) = P(p)^\top = P(p)^2$. This proves equation 65. □

G.2 A GENERIC STATIONARY NORMAL-MOMENT ARGUMENT

Lemma 22 (From one-step contraction to stationary distance)**.** *Suppose a kernel K_h with invariant law ρ_h satisfies*

$$K_h q(x) \le (1 - 2\alpha h) q(x) + B^2 h + D h^{1+2\nu}, \qquad \alpha > 0, \quad \nu > 0. \tag{67}$$

Then

$$\rho_h(q) \le \frac{B^2 + D h^{2\nu}}{2\alpha}, \tag{68}$$

and

$$\rho_h(\sqrt{q}) \le C\alpha^{-1/2}(B + h^\nu). \tag{69}$$

Proof. Integrating equation 67 under ρ_h and using invariance cancels the two occurrences of $\rho_h(q)$. Divide by $2\alpha h$ to obtain the second-moment bound. Jensen's inequality gives the first-moment bound. □

G.3 NORMAL RECURSION FOR LOCALIZED IIL

The next two lemmas isolate the two parts of the argument: the exact geometric cancellation, and the perturbation caused by the inexact oracle. This makes all orders used in the Lyapunov estimate explicit.

Lemma 23 (Exact residual bookkeeping)**.** *Fix $x = p + r \in \mathcal{K}_{\tau_\star}$ and abbreviate $\theta = \vartheta_x$, $\chi = \chi_h(\xi)$, $S = \chi\sigma\sqrt{h}\,\theta v$, and $L = -\ell h r$. Define*

$$\begin{aligned}
B_{\mathrm{II}} &:= \chi h \theta^2 \big(a(p) - \beta^{-1}\mathrm{II}_p(v,v)\big),\\
B_H &:= \chi h \theta^2 \big(a(p) + \beta^{-1} H(p)\big), \qquad \Delta_s := L + S + B_s.
\end{aligned} \tag{70}$$

For $s \in \{\mathrm{II}, \mathrm{H}\}$ write

$$r(x + \Delta_s) = (1 - \ell h) r + \zeta_s. \tag{71}$$

There is a constant C, uniform in x and h, such that

$$\begin{aligned}
\|\mathbb{E}[\zeta_{\mathrm{II}} \mid x]\| &\le C(\sqrt{q(x)}\,h + h^2),\\
\mathbb{E}[\|\zeta_{\mathrm{II}}\|^2 \mid x] &\le C q(x) h + C h^3,
\end{aligned} \tag{72}$$

and

$$\begin{aligned}
\|\mathbb{E}[\zeta_H \mid x]\| &\le C(\sqrt{q(x)}\,h + h^2),\\
\mathbb{E}[\|\zeta_H\|^2 \mid x] &\le C q(x) h + C h^2.
\end{aligned} \tag{73}$$

Proof. Let $C_s = L + B_s$. Expanding $r(x + S + C_s)$ to third order with fourth-order remainder gives the exact decomposition

$$\begin{aligned}
\zeta_s &= A_1 + A_{2,s} + A_{3,s},\\
A_1 &:= Dr(x)S,\\
A_{2,s} &:= Dr(x)B_s + \frac{1}{2}D^2r(x)[S,S],\\
A_{3,s} &:= D^2r(x)[S,C_s] + \frac{1}{2}D^2r(x)[C_s,C_s]\\
&\quad + \frac{1}{6}D^3r(x)[S,S,S] + \frac{1}{2}D^3r(x)[S,S,C_s]\\
&\quad + \frac{1}{2}D^3r(x)[S,C_s,C_s] + \frac{1}{6}D^3r(x)[C_s,C_s,C_s] + \mathcal{R}_4(x,\Delta_s).
\end{aligned} \tag{74}$$

The landing term has disappeared from the leading residual because $Dr(x)L = -\ell hr$ by equation 48.

We bound the three groups separately. First, radial symmetry gives $\mathbb{E}[A_1 \mid x] = 0$, and

$$\mathbb{E}[\|A_1\|^2 \mid x] \le Ch\|Dr(x)P(p)\|_{\mathrm{op}}^2\mathbb{E}\|\xi\|^2 \le Cq(x)h. \tag{75}$$

For IIL, using $Dr(x)n = n$ for $n \in N_p\mathcal{M}$ and $\sigma^2/2 = \beta^{-1}$,

$$\begin{aligned}
A_{2,\mathrm{II}} &= \chi h\theta^2 Dr(x)a(p)\\
&\quad + \chi\beta^{-1}h\theta^2\big(D^2r(x)[v,v] - \mathrm{II}_p(v,v)\big).
\end{aligned} \tag{76}$$

The drift is tangent and the residual Hessian estimate in equation 48 therefore imply

$$\|A_{2,\mathrm{II}}\| \le C\chi h\sqrt{q(x)}\,(1 + \|\xi\|^2). \tag{77}$$

This display is the pathwise cancellation: the two order-h normal terms have cancelled before expectation. Consequently,

$$\|\mathbb{E}A_{2,\mathrm{II}}\| \le C\sqrt{q}\,h, \qquad \mathbb{E}\|A_{2,\mathrm{II}}\|^2 \le Cqh^2. \tag{78}$$

For HL, the corresponding identity is

$$\begin{aligned}
A_{2,H} &= \chi h\theta^2 Dr(x)a(p)\\
&\quad + \chi\beta^{-1}h\theta^2\big(H(p) + \mathrm{II}_p(v,v)\big) + E_H,
\end{aligned} \tag{79}$$

where $\|E_H\| \le C\chi h\sqrt{q}\,\|\xi\|^2$. By Lemma 13,

$$\mathbb{E}\big[\chi(H + \mathrm{II}(v,v))\big] = (p_h - a_h)H, \tag{80}$$

which is smaller than every algebraic power of h, whereas its second moment is uniformly bounded. Hence

$$\|\mathbb{E}A_{2,H}\| \le C(\sqrt{q}\,h + h^2), \qquad \mathbb{E}\|A_{2,H}\|^2 \le Cqh^2 + Ch^2. \tag{81}$$

It remains to control $A_{3,s}$. Uniform boundedness of the derivatives of r, Lemma 21, and

$$\mathbb{E}\|S\|^{2j} \le C_jh^j, \qquad \mathbb{E}\|C_s\|^j \le C_jh^j, \qquad j = 1,2,3,4, \tag{82}$$

show that every term in $A_{3,s}$ has second moment at most Ch^3. For completeness, the only terms of size $h^{3/2}$ are $D^2r[S,C_s]$ and $D^3r[S,S,S]/6$. In the former, the component containing L or the deterministic drift has one Gaussian factor, while the component containing the IIL quadratic correction has three; every such conditional expectation is zero by the radial symmetry of G_h. The pure cubic term has zero mean for the same reason. All terms with nonzero mean contain either two factors of S and one factor of C_s, or at least two factors of C_s, and are therefore $O(h^2)$. Finally,

$$\mathbb{E}\|\mathcal{R}_4(x,\Delta_s)\| \le C\mathbb{E}\|\Delta_s\|^4 \le Ch^2, \qquad \mathbb{E}\|\mathcal{R}_4(x,\Delta_s)\|^2 \le C\mathbb{E}\|\Delta_s\|^8 \le Ch^4. \tag{83}$$

Thus

$$\|\mathbb{E}A_{3,s}\| \le Ch^2, \qquad \mathbb{E}\|A_{3,s}\|^2 \le Ch^3. \tag{84}$$

Combining equation 75, equation 78–equation 84, and $\|a+b+c\|^2 \le 3(\|a\|^2+\|b\|^2+\|c\|^2)$ proves equation 72 and equation 73. □

Recall the aggregate oracle errors $\eta_{\mathrm{II},\ell}$ and $\eta_{H,\ell}$ defined in Theorem 3. For $s \in \{\mathrm{II}, \mathrm{H}\}$, we use the shorthand $\eta_{s,\ell} = \eta_{\mathrm{II},\ell}$ when $s = \mathrm{II}$ and $\eta_{s,\ell} = \eta_{H,\ell}$ when $s = \mathrm{H}$.

Lemma 24 (Propagation of oracle perturbations through the residual map)**.** *Couple the exact and inexact localized increments with the same ξ and the same scalar $\theta = \vartheta_x$. For $s \in \{\mathrm{II}, \mathrm{H}\}$, let $\delta\Delta_s = \widehat{\Delta}_s - \Delta_s$ and*

$$\delta r_s^+ := r(x + \widehat{\Delta}_s) - r(x + \Delta_s). \tag{85}$$

Then, for sufficiently small oracle errors,

$$\begin{aligned} \|\mathbb{E}[\delta\Delta_s \mid x]\| &\le C\eta_{s,\ell} h, \\ \mathbb{E}[\|\delta\Delta_s\|^2 \mid x] &\le C\varepsilon_P^2 h + C\eta_{s,\ell}^2 h^2, \\ \mathbb{E}[\|\delta\Delta_s\|^4 \mid x] &\le C\varepsilon_P^4 h^2 + Ch^4, \end{aligned} \tag{86}$$

and

$$\begin{aligned} \|\mathbb{E}[\delta r_s^+ \mid x]\| &\le C\eta_{s,\ell} h, \\ \mathbb{E}[\|\delta r_s^+\|^2 \mid x] &\le C\varepsilon_P^2 h + C\eta_{s,\ell}^2 h^2 + Ch^4. \end{aligned} \tag{87}$$

Proof. For IIL, the increment difference is exactly

$$\begin{aligned} \delta\Delta_{\mathrm{II}} = \ell h(\widehat{p} - p) &- \chi h\theta^2[\widehat{P}\nabla U(\widehat{p}) - P\nabla U(p)] \\ &+ \chi\sigma\sqrt{h}\,\theta(\widehat{P} - P)\xi \\ &- \chi\beta^{-1}h\theta^2[\widehat{\mathrm{II}}_{\widehat{p}}(\widehat{v}, \widehat{v}) - \mathrm{II}_p(v, v)]. \end{aligned} \tag{88}$$

For HL, replace the last line by $\chi\beta^{-1}h\theta^2(\widehat{H} - H)$. The order-$\sqrt{h}$ term is centered because $\mathbb{E}[\chi\xi] = 0$. The oracle assumptions and the quadratic-growth bound on the curvature estimator give equation 86; the fourth-moment bound uses Gaussian eighth moments and the fact that all non-projector perturbations carry a factor h.

Put $y = x + \Delta_s$ and $d = \delta\Delta_s$. The exact integral Taylor formula is

$$\delta r_s^+ = Dr(y)d + \int_0^1 (1 - t)D^2 r(y + td)[d, d]dt. \tag{89}$$

The segment $y + td$ remains in a fixed compact subtube: both increments are uniformly small on the active branch, while the outer tube has the fixed margin $\tau - \tau_\star$ on the fallback branch. Expand

$$Dr(y) = Dr(x) + D^2 r(x)[\Delta_s, \cdot] + R_D, \qquad \|R_D\|_{\mathrm{op}} \le C\|\Delta_s\|^2. \tag{90}$$

The term $Dr(x)\mathbb{E}d$ is $O(\eta_{s,\ell}h)$. In $\mathbb{E}D^2 r(x)[\Delta_s, d]$, the only product containing two order-$\sqrt{h}$ factors is the covariance of $\chi\sigma\sqrt{h}\,\theta P\xi$ with $\chi\sigma\sqrt{h}\,\theta(\widehat{P} - P)\xi$; its norm is $O(\varepsilon_P h)$. Every other product contains at least one order-h factor and is bounded by Cauchy–Schwarz using Lemma 21 and equation 86. The $R_D d$ term is $O(\varepsilon_P h^{3/2} + \eta_{s,\ell}h^2)$, and the integral remainder is bounded in mean by $C\mathbb{E}\|d\|^2$. Since $\varepsilon_P \le \eta_{s,\ell}$ and the errors are bounded by a fixed small constant, this proves the first line of equation 87. Squaring equation 89, using boundedness of Dr and $D^2 r$, and applying the second and fourth moment bounds for d proves the second line. □

Lemma 25 (One-step normal recursion for localized IIL)**.** *There exist constants $c_{\mathrm{II}}, C, h_0, \varepsilon_0 > 0$ such that, if $\ell > c_{\mathrm{II}}$, $0 < h \le h_0$, and the oracle errors are at most ε_0, then with $\alpha_{\mathrm{II}} = \ell - c_{\mathrm{II}} > 0$,*

$$\mathbb{E}[q(X_{k+1}) \mid X_k = x] \le (1 - 2\alpha_{\mathrm{II}}h)q(x) + C\eta_{\mathrm{II},\ell}^2 h + Ch^3 \tag{91}$$

uniformly for $x \in \mathcal{K}_{\tau_\star}$.

Proof. For the exact comparison transition, Lemma 23 and equation 71 give

$$\begin{aligned} \mathbb{E}q(x + \Delta_{\mathrm{II}}) &= (1 - \ell h)^2 q(x) + 2(1 - \ell h)\langle r, \mathbb{E}\zeta_{\mathrm{II}}\rangle + \mathbb{E}\|\zeta_{\mathrm{II}}\|^2 \\ &\le [1 - (2\ell - C_0)h]q(x) + Ch^3, \end{aligned} \tag{92}$$

where $\sqrt{q}\,h^2 \le \kappa q h + C_\kappa h^3$ and $\ell^2 h^2 q \le Chq$ were used for $h \le h_0$. For the inexact transition, $r(x+\widehat{\Delta}_{\mathrm{II}}) = r(x+\Delta_{\mathrm{II}}) + \delta r_{\mathrm{II}}^+$. The perturbation lemma and Young's inequality imply, for any fixed $\kappa > 0$,

$$2\big|\mathbb{E}\langle r(x+\Delta_{\mathrm{II}}), \delta r_{\mathrm{II}}^+\rangle\big| \le \kappa h q(x) + C_\kappa \eta_{\mathrm{II},\ell}^2 h + Ch^3, \tag{93}$$

and $\mathbb{E}\|\delta r_{\mathrm{II}}^+\|^2 \le C\eta_{\mathrm{II},\ell}^2 h + Ch^3$. Indeed, split $r(x+\Delta_{\mathrm{II}}) = (1-\ell h)r + \zeta_{\mathrm{II}}$; the first cross term uses the mean bound in equation 87, while the second uses conditional Cauchy–Schwarz together with equation 72. Choose κ smaller than the positive margin $2\ell - C_0$ and choose the fixed geometric constant c_{II} large enough that all qh terms are absorbed whenever $\ell > c_{\mathrm{II}}$. This proves equation 91. □

Corollary 26 (Stationary normal concentration for localized IIL)**.** *Every invariant probability measure $\rho_h^{\mathrm{IIL,loc}}$ of $K_h^{\mathrm{IIL,loc}}$ satisfies*

$$\rho_h^{\mathrm{IIL,loc}}(q) \le C\alpha_{\mathrm{II}}^{-1}\big(\eta_{\mathrm{II},\ell}^2 + h^2\big), \tag{94}$$

and

$$\mathbb{E}_{\rho_h^{\mathrm{IIL,loc}}} \operatorname{dist}(X, \mathcal{M}) \le C\alpha_{\mathrm{II}}^{-1/2}\big(\eta_{\mathrm{II},\ell} + h\big). \tag{95}$$

Moreover,

$$\mathbb{E}_{\rho_h^{\mathrm{IIL,loc}}}[1 - \vartheta_X^2] \le C\delta_0^{-2}\alpha_{\mathrm{II}}^{-1}\big(\eta_{\mathrm{II},\ell}^2 + h^2\big). \tag{96}$$

Proof. Integrating the one-step recursion equation 91 against the invariant law $\rho_h^{\mathrm{IIL,loc}}$ and using stationarity gives

$$\rho_h^{\mathrm{IIL,loc}}(q) \le (1 - 2\alpha_{\mathrm{II}} h)\rho_h^{\mathrm{IIL,loc}}(q) + C\eta_{\mathrm{II},\ell}^2 h + Ch^3.$$

Hence

$$\rho_h^{\mathrm{IIL,loc}}(q) \le C\alpha_{\mathrm{II}}^{-1}(\eta_{\mathrm{II},\ell}^2 + h^2),$$

which proves equation 94.

Since $q(X) = \operatorname{dist}(X, \mathcal{M})^2$, Jensen's inequality yields

$$\begin{aligned}\mathbb{E}\operatorname{dist}(X, \mathcal{M}) &= \mathbb{E}\sqrt{q(X)} \le \sqrt{\mathbb{E}q(X)} \\ &\le C\alpha_{\mathrm{II}}^{-1/2}(\eta_{\mathrm{II},\ell} + h),\end{aligned}$$

proving equation 95.

Finally, by equation 35,

$$1 - \vartheta_X^2 \le \mathbf{1}_{\{q(X) > \delta_0^2\}}.$$

Therefore Markov's inequality and equation 94 give

$$\begin{aligned}\mathbb{E}[1 - \vartheta_X^2] &\le \mathbb{P}(q(X) > \delta_0^2) \\ &\le \delta_0^{-2}\mathbb{E}q(X) \\ &\le C\delta_0^{-2}\alpha_{\mathrm{II}}^{-1}(\eta_{\mathrm{II},\ell}^2 + h^2).\end{aligned}$$

□

G.4 Normal recursion for localized HL

Lemma 27 (One-step normal recursion for localized HL)**.** *There exist constants $c_H, C, h_0, \varepsilon_0 > 0$ such that, if $\ell > c_H$, $0 < h \le h_0$, and the oracle errors are at most ε_0, then with $\alpha_H = \ell - c_H > 0$,*

$$\mathbb{E}[q(X_{k+1}) \mid X_k = x] \le (1 - 2\alpha_H h)q(x) + C\eta_{H,\ell}^2 h + Ch^2 \tag{97}$$

uniformly for $x \in \mathcal{K}_{\tau_\star}$.

Proof. For the exact comparison transition, Lemma 23 gives $r(x+\Delta_H) = (1-\ell h)r + \zeta_H$ with the moment bounds equation 73. Therefore

$$\begin{aligned}\mathbb{E}q(x+\Delta_H) &= (1-\ell h)^2 q(x) + 2(1-\ell h)\langle r, \mathbb{E}\zeta_H\rangle + \mathbb{E}\|\zeta_H\|^2 \\ &\le [1-(2\ell - C_0)h]q(x) + Ch^2. \end{aligned} \tag{98}$$

The order-h^2 term in the general upper bound arises because, at $r=0$, the leading residual is

$$\chi\beta^{-1}h\theta^2\big(H(p) + \mathrm{II}_p(v,v)\big), \tag{99}$$

which is centered up to the super-algebraic difference $p_h - a_h$ but is not canceled pathwise. Its second moment is generally of order h^2 (and may vanish in special geometries).

For the inexact transition, apply Lemma 24 with $s = \mathrm{H}$ and write $r(x+\widehat{\Delta}_H) = r(x+\Delta_H) + \delta r_H^+$. The same split as in equation 93 yields, for every fixed $\kappa > 0$,

$$\begin{aligned} 2\big|\mathbb{E}\langle r(x+\Delta_H), \delta r_H^+\rangle\big| &\le \kappa h q(x) + C_\kappa \eta_{H,\ell}^2 h + Ch^2, \\ \mathbb{E}\|\delta r_H^+\|^2 &\le C\eta_{H,\ell}^2 h + Ch^2. \end{aligned} \tag{100}$$

Choose κ below the positive margin in equation 98 and choose c_H large enough to absorb all fixed geometric qh terms. This proves equation 97. □

Corollary 28 (Stationary normal concentration for localized HL)**.** *Every invariant probability measure $\rho_h^{HL,\mathrm{loc}}$ of $K_h^{HL,\mathrm{loc}}$ satisfies*

$$\begin{aligned} \rho_h^{HL,\mathrm{loc}}(q) &\le C\alpha_H^{-1}(\eta_{H,\ell}^2 + h), \\ \mathbb{E}\,\mathrm{dist}(X,\mathcal{M}) &\le C\alpha_H^{-1/2}(\eta_{H,\ell} + h^{1/2}), \\ \mathbb{E}[1-\vartheta_X^2] &\le C\delta_0^{-2}\alpha_H^{-1}(\eta_{H,\ell}^2 + h). \end{aligned} \tag{101}$$

Proof. Apply Lemma 22 to equation 97, then use equation 35 exactly as in Corollary 26. □

G.5 Off-manifold control for localized hard projection

Lemma 29 (Distance and cutoff inactivity after inexact projection)**.** *For the transition equation 30,*

$$\mathrm{dist}(X_{k+1}, \mathcal{M}) \le \varepsilon_\pi. \tag{102}$$

Moreover,

$$\widehat{\mathsf{d}}(X_{k+1}) \le 2\varepsilon_\pi \le 2\varepsilon_{\pi,0} < \rho_0, \tag{103}$$

so $\vartheta_{X_k} = 1$ for every $k \ge 1$. Every invariant law $\rho_h^{\pi,\mathrm{loc}}$ therefore satisfies

$$\mathbb{E}_{\rho_h^{\pi,\mathrm{loc}}}\mathrm{dist}(X,\mathcal{M}) \le \varepsilon_\pi, \qquad \vartheta_X = 1 \quad \rho_h^{\pi,\mathrm{loc}}\text{-a.s.} \tag{104}$$

Proof. Let $Y = Y_h^{\pi,\mathrm{loc}}(x,\xi)$. Forward invariance ensures $Y \in \mathcal{K}_{\tau_\star} \subset \mathcal{K}_{\tau_{\mathrm{or}}}$, so

$$\mathrm{dist}(\widehat{\pi}(Y), \mathcal{M}) \le \|\widehat{\pi}(Y) - \pi(Y)\| \le \varepsilon_\pi.$$

For $z = \widehat{\pi}(Y)$,

$$\widehat{\mathsf{d}}(z) \le \|z - \pi(z)\| + \|\pi(z) - \widehat{\pi}(z)\| \le 2\varepsilon_\pi.$$

The invariant-law statements follow because every transition output has these properties. □

G.6 Projected generator consistency

The projected-generator estimate is common to the two landing schemes: for a normal extension $\widetilde{\varphi} = \varphi \circ \pi$, purely normal first-order corrections do not contribute. The distinction between IIL and HL therefore enters primarily through their stationary normal concentration, proved above. We give the full generator argument for IIL and then record the corresponding bounds for HL and hard projection.

Lemma 30 (Projected-generator consistency for localized IIL). *Let $\varphi \in C^4(\mathcal{M})$ and $\widetilde{\varphi} = \varphi \circ \pi$. For every integer $m \geq 2$, there exist constants $C, C_m > 0$, independent of h, the oracle errors, and φ, such that uniformly for $x \in \mathcal{K}_{\tau_\star}$,*

$$\begin{aligned}\Big|(\mathcal{L}_h^{\mathrm{IIL},\mathrm{loc}}\widetilde{\varphi})(x) - \vartheta_x^2(\mathcal{L}\varphi)(\pi(x))\Big| \\ \leq C\|\varphi\|_{C^4(\mathcal{M})}\left[\eta_{\mathrm{II},\ell} + h + \mathrm{dist}(x,\mathcal{M})\right] + C_m\|\varphi\|_{C^4(\mathcal{M})}h^m.\end{aligned} \tag{105}$$

Consequently,

$$\begin{aligned}\Big|(\mathcal{L}_h^{\mathrm{IIL},\mathrm{loc}}\widetilde{\varphi})(x) - (\mathcal{L}\varphi)(\pi(x))\Big| \\ \leq C\|\varphi\|_{C^4(\mathcal{M})}\left[\eta_{\mathrm{II},\ell} + h + \mathrm{dist}(x,\mathcal{M})\right] \\ + \|\mathcal{L}\varphi\|_\infty(1-\vartheta_x^2) + C_m\|\varphi\|_{C^4(\mathcal{M})}h^m.\end{aligned} \tag{106}$$

Proof. Fix $x \in \mathcal{K}_{\tau_\star}$ and let all expectations below be with respect to $\xi \sim \mathcal{N}(0, I_d)$. Set

$$f := \widetilde{\varphi}, \qquad p := \pi(x), \qquad r := x - p, \qquad \theta := \vartheta_x.$$

Step 1: exact-geometry comparison increment. For the exact IIL increment Δ_{II} in equation 62, radial symmetry and Lemma 13 give

$$\mathbb{E}_\xi\Delta_{\mathrm{II}} = -\ell h r + p_h h\theta^2 a(p) + \beta^{-1}a_h h\theta^2 H(p). \tag{107}$$

Moreover,

$$\mathbb{E}_\xi[\Delta_{\mathrm{II}}\Delta_{\mathrm{II}}^\top] = 2\beta^{-1}a_h h\theta^2 P(p) + R_2, \qquad \|R_2\| \leq Ch^2. \tag{108}$$

Indeed, the displayed term is the covariance of the tangent Gaussian increment. Every cross term containing exactly one stochastic factor has an odd Gaussian integrand and therefore vanishes; all remaining terms contain at least two order-h factors.

Lemma 21 also gives

$$\left\|\mathbb{E}_\xi[\Delta_{\mathrm{II}}^{\otimes 3}]\right\| \leq Ch^2, \qquad \mathbb{E}_\xi\|\Delta_{\mathrm{II}}\|^4 \leq Ch^2. \tag{109}$$

Taylor expansion of f to third order therefore yields

$$\begin{aligned}K_h^{\mathrm{IIL},\mathrm{ex}}f(x) - f(x) = Df(x)\mathbb{E}_\xi\Delta_{\mathrm{II}} \\ + \frac{1}{2}D^2f(x) : \mathbb{E}_\xi[\Delta_{\mathrm{II}}\Delta_{\mathrm{II}}^\top] \\ + \frac{1}{6}D^3f(x) : \mathbb{E}_\xi[\Delta_{\mathrm{II}}^{\otimes 3}] + R_{4,f},\end{aligned} \tag{110}$$

where

$$|R_{4,f}| \leq C\|f\|_{C^4(\mathcal{K}_{\tau_\star})}\mathbb{E}_\xi\|\Delta_{\mathrm{II}}\|^4 \leq C\|\varphi\|_{C^4(\mathcal{M})}h^2.$$

Because $x = p + r$ lies on the normal fiber through p, $r \in N_p\mathcal{M}$. Moreover, $f = \varphi \circ \pi$ is constant along this normal fiber, so

$$Df(x)n = 0, \qquad n \in N_p\mathcal{M}.$$

Since both r and $H(p)$ are normal,

$$Df(x)r = 0, \qquad Df(x)H(p) = 0.$$

Thus the landing contribution and the mean of the IIL curvature correction do not enter the projected generator.

Using $a(p) = -P(p)\nabla U(p)$, equation 107– equation 108 give, after division by h,

$$\begin{aligned}\frac{K_h^{\mathrm{IIL},\mathrm{ex}}f(x) - f(x)}{h} = p_h\theta^2 Df(x)a(p) \\ + \beta^{-1}a_h\theta^2\,\mathrm{tr}\big(P(p)D^2f(x)\big) \\ + O(\|\varphi\|_{C^4}h)\,.\end{aligned} \tag{111}$$

Since f is C^4 on the fixed tube,

$$\|Df(x) - Df(p)\| + \|D^2 f(x) - D^2 f(p)\| \le C\|\varphi\|_{C^4(\mathcal{M})}\|x - p\|.$$

Furthermore, Lemma 9 gives, for every $m \ge 2$,

$$|1 - p_h| + |1 - a_h| \le C_m h^m.$$

Hence, using equation 50,

$$\begin{aligned}\left|\frac{K_h^{\mathrm{IIL,ex}} f(x) - f(x)}{h} - \theta^2(\mathcal{L}\varphi)(p)\right| &\le C\|\varphi\|_{C^4(\mathcal{M})}\big[h + \mathrm{dist}(x, \mathcal{M})\big] \\ &\quad + C_m\|\varphi\|_{C^4(\mathcal{M})} h^m.\end{aligned} \tag{112}$$

Step 2: perturbation by the learned geometry. Let

$$\delta\Delta_{\mathrm{II}} := \widehat{\Delta}_{\mathrm{II}} - \Delta_{\mathrm{II}}.$$

The oracle bounds of Assumption 1, together with the centered projector-noise perturbation, imply

$$\begin{aligned}\Big\|\mathbb{E}_\xi[\widehat{\Delta}_{\mathrm{II}} - \Delta_{\mathrm{II}}]\Big\| &\le C\eta_{\mathrm{II},\ell} h, \\ \Big\|\mathbb{E}_\xi[\widehat{\Delta}_{\mathrm{II}}\widehat{\Delta}_{\mathrm{II}}^\top - \Delta_{\mathrm{II}}\Delta_{\mathrm{II}}^\top]\Big\| &\le C\eta_{\mathrm{II},\ell} h + C\eta_{\mathrm{II},\ell}^2 h^2, \\ \Big\|\mathbb{E}_\xi[\widehat{\Delta}_{\mathrm{II}}^{\otimes 3} - \Delta_{\mathrm{II}}^{\otimes 3}]\Big\| &\le C\eta_{\mathrm{II},\ell} h^{3/2} + Ch^2, \\ \mathbb{E}_\xi\|\widehat{\Delta}_{\mathrm{II}}\|^4 + \mathbb{E}_\xi\|\Delta_{\mathrm{II}}\|^4 &\le Ch^2.\end{aligned} \tag{113}$$

For the leading covariance perturbation, one uses

$$\mathbb{E}_\xi\Big[\chi(\theta\widehat{P}\xi)(\theta\widehat{P}\xi)^\top\Big] = a_h\theta^2\widehat{P}\widehat{P}^\top.$$

Writing $\widehat{P} = P + E_P$ with $\|E_P\|_{\mathrm{op}} \le \varepsilon_P$ gives

$$\widehat{P}\widehat{P}^\top - P = PE_P^\top + E_P P + E_P E_P^\top,$$

and therefore

$$\|\widehat{P}\widehat{P}^\top - P\|_{\mathrm{op}} \le 2\varepsilon_P + \varepsilon_P^2. \tag{114}$$

Apply the Taylor expansion to $\widehat{\Delta}_{\mathrm{II}}$ and subtract the exact expansion. After division by h, the first two moment perturbations contribute $C\eta_{\mathrm{II},\ell}$, the third contributes at most $C\eta_{\mathrm{II},\ell}\sqrt{h} + Ch$, and the fourth-order remainders contribute Ch. Combining these estimates with equation 112 proves equation 105.

Finally,

$$\begin{aligned}&(\mathcal{L}_h^{\mathrm{IIL,loc}}\widetilde{\varphi})(x) - (\mathcal{L}\varphi)(p) \\ &= \Big[(\mathcal{L}_h^{\mathrm{IIL,loc}}\widetilde{\varphi})(x) - \theta^2(\mathcal{L}\varphi)(p)\Big] - (1 - \theta^2)(\mathcal{L}\varphi)(p).\end{aligned}$$

Taking absolute values gives equation 106. □

Corollary 31 (Projected-generator consistency for localized HL)**.** *Under the assumptions of Lemma 30, uniformly for $x \in \mathcal{K}_{\tau_\star}$,*

$$\begin{aligned}&\Big|(\mathcal{L}_h^{HL,\mathrm{loc}}\widetilde{\varphi})(x) - \vartheta_x^2(\mathcal{L}\varphi)(\pi(x))\Big| \\ &\quad \le C\|\varphi\|_{C^4(\mathcal{M})}\left[\eta_{H,\ell} + h + \mathrm{dist}(x, \mathcal{M})\right] + C_m\|\varphi\|_{C^4(\mathcal{M})} h^m.\end{aligned} \tag{115}$$

Consequently,

$$\begin{aligned}&\Big|(\mathcal{L}_h^{HL,\mathrm{loc}}\widetilde{\varphi})(x) - (\mathcal{L}\varphi)(\pi(x))\Big| \\ &\quad \le C\|\varphi\|_{C^4(\mathcal{M})}\left[\eta_{H,\ell} + h + \mathrm{dist}(x, \mathcal{M})\right] \\ &\qquad + \|\mathcal{L}\varphi\|_\infty(1 - \vartheta_x^2) + C_m\|\varphi\|_{C^4(\mathcal{M})} h^m.\end{aligned} \tag{116}$$

Proof. The proof is the same Taylor argument as for IIL. For the exact HL increment,

$$\mathbb{E}_\xi \Delta_H = -\ell h r + p_h h\theta^2 a(p) + \beta^{-1} p_h h\theta^2 H(p),$$

while its leading covariance is again

$$2\beta^{-1} a_h h\theta^2 P(p) + O(h^2).$$

Because $\widetilde{\varphi} = \varphi \circ \pi$ is constant along the normal fiber through p,

$$D\widetilde{\varphi}(x) r = D\widetilde{\varphi}(x) H(p) = 0.$$

Thus the different explicit curvature drift does not affect the leading projected generator. Repeating the learned-oracle perturbation argument with ε_H in place of $\varepsilon_{\mathrm{II}}$ gives $\eta_{H,\ell}$ and proves equation 115. The full-target bound follows by adding and subtracting $\vartheta_x^2(\mathcal{L}\varphi)(\pi(x))$. □

Lemma 32 (Projected-generator consistency for localized hard projection)**.** *Let $\varphi \in C^4(\mathcal{M})$ and $\widetilde{\varphi} = \varphi \circ \pi$. For every integer $m \geq 2$, uniformly for $x \in \mathcal{K}_{\tau_\star}$,*

$$\begin{aligned} &\left| (\mathcal{L}_h^{\pi,\mathrm{loc}} \widetilde{\varphi})(x) - \vartheta_x^2(\mathcal{L}\varphi)(\pi(x)) \right| \\ &\qquad \leq C\|\varphi\|_{C^4(\mathcal{M})} \left[\varepsilon_P + \varepsilon_\pi + h + \mathrm{dist}(x, \mathcal{M}) + \frac{\varepsilon_\pi}{h} \right] + C_m \|\varphi\|_{C^4(\mathcal{M})} h^m. \end{aligned} \tag{117}$$

Proof. Fix x, put $p = \pi(x)$, $\theta = \vartheta_x$, and $f = \widetilde{\varphi}$. Define the exact preprojection proposal

$$Y^0 := x + \chi \left[h\theta^2 a(p) + \sigma\sqrt{h}\, \theta v \right].$$

The same deterministic increment estimate used in Lemma 10 shows that Y^0 remains in a fixed compact sub-tube of $\mathcal{T}_\tau$ for sufficiently small h.

Since $f = \varphi \circ \pi$,

$$f(\pi(Y^0)) = f(Y^0).$$

Hence the exact projected step is analyzed by the same weak Taylor expansion as above, now without landing or an explicit curvature term:

$$\left| \frac{\mathbb{E}_\xi[f(\pi(Y^0)) - f(x)]}{h} - \theta^2(\mathcal{L}\varphi)(p) \right| \leq C\|\varphi\|_{C^4} \left[h + \mathrm{dist}(x, \mathcal{M}) \right] + C_m \|\varphi\|_{C^4} h^m.$$

Let $\widehat{Y}$ denote the learned preprojection proposal. The drift perturbation contributes $O(h(\varepsilon_P + \varepsilon_\pi))$, while the covariance perturbation contributes $O(h\varepsilon_P)$. After division by h, these give

$$O(\varepsilon_P + \varepsilon_\pi).$$

The final learned projection introduces a different error mechanism. Because

$$f(\widehat{Y}) = f(\pi(\widehat{Y})),$$

we have

$$\begin{aligned} |f(\widehat{\pi}(\widehat{Y})) - f(\widehat{Y})| &= |f(\widehat{\pi}(\widehat{Y})) - f(\pi(\widehat{Y}))| \\ &\leq C\|\varphi\|_{C^1(\mathcal{M})} \|\widehat{\pi}(\widehat{Y}) - \pi(\widehat{Y})\| \\ &\leq C\|\varphi\|_{C^1(\mathcal{M})} \varepsilon_\pi. \end{aligned}$$

Since this $O(\varepsilon_\pi)$ error is incurred at each discrete step, division by h in the discrete generator produces

$$\frac{\varepsilon_\pi}{h}.$$

Combining the estimates proves equation 117. □

H Proof of the Main IIL Stationary Theorem

In this section we prove the formal version of Theorem 3 for the IIL-discretization. The corresponding results for the π- and HL-schemes are deferred to the next section.

Theorem 33 (Stationary accuracy of IIL)**.** *Under Assumptions 1, 5, and 7, there exist $\Delta t_0, \varepsilon_0, c_{\mathrm{II}}, C > 0$ such that the following holds. Let $c_{\mathrm{II}} < \ell \le \ell_+$ and $\alpha_{\mathrm{II}} := \ell - c_{\mathrm{II}} > 0$. For $0 < \Delta t \le \Delta t_0$ and $\max\{\varepsilon_\pi, \varepsilon_P, \varepsilon_H, \varepsilon_{\mathrm{II}}\} \le \varepsilon_0$, every invariant probability measure $\rho_{\Delta t}^{\mathrm{II}L,\mathrm{loc}}$ of the localized* IIL *kernel satisfies, uniformly over $g \in \mathcal{F}$,*

$$\begin{aligned} \mathbb{E}_{\rho_{\Delta t}^{\mathrm{II}L,\mathrm{loc}}} \mathcal{R}_\perp &\le C\alpha_{\mathrm{II}}^{-1/2} \left(\eta_{\mathrm{II},\ell} + \Delta t\right), \\ \mathbb{E}_{\rho_{\Delta t}^{\mathrm{II}L,\mathrm{loc}}} \mathcal{R}_{\mathrm{gen}}(X; g) &\le C\left(1 + \alpha_{\mathrm{II}}^{-1/2}\right)\left(\eta_{\mathrm{II},\ell} + \Delta t\right) + C\delta_0^{-2}\alpha_{\mathrm{II}}^{-1}\left(\eta_{\mathrm{II},\ell}^2 + \Delta t^2\right). \end{aligned} \tag{118}$$

equation 57 then transfers normal and generator control to weak bias,

$$\mathrm{dist}_{\mathcal{F}}(\rho_{\Delta t}^{\mathrm{II}L,\mathrm{loc}}, \mu) \le C\left(1 + \alpha_{\mathrm{II}}^{-1/2}\right)\left(\eta_{\mathrm{II},\ell} + \Delta t\right) + C\delta_0^{-2}\alpha_{\mathrm{II}}^{-1}\left(\eta_{\mathrm{II},\ell}^2 + \Delta t^2\right). \tag{119}$$

Here $\delta_0 = \rho_0 - \varepsilon_{\pi,0} > 0$ is the fixed localization margin defined in Lemma 8. For fixed stable ℓ and fixed δ_0, the quadratic term is lower order, giving $\mathrm{dist}_{\mathcal{F}}\left(\rho_{\Delta t}^{\mathrm{II}L}, \mu\right) = \mathcal{O}(\eta_{\mathrm{II},\ell} + \Delta t)$.

Proof of Theorem 33. Set $h = \Delta t$, and let $\rho_h^{\mathrm{II}L,\mathrm{loc}}$ be any invariant probability measure of the localized IIL kernel. Existence follows from Proposition 11.

Off-manifold error. Corollary 26 gives

$$\mathbb{E}_{\rho_h^{\mathrm{II}L,\mathrm{loc}}} \mathrm{dist}(X, \mathcal{M}) \le C\alpha_{\mathrm{II}}^{-1/2}(\eta_{\mathrm{II},\ell} + h),$$

which proves the bound on $\mathcal{R}_\perp$. The same corollary gives

$$\mathbb{E}_{\rho_h^{\mathrm{II}L,\mathrm{loc}}}[1 - \vartheta_X^2] \le C\delta_0^{-2}\alpha_{\mathrm{II}}^{-1}(\eta_{\mathrm{II},\ell}^2 + h^2).$$

Projected-generator error. Fix $g \in \mathcal{F}$, let u_g solve the Poisson equation, and set $\widetilde{u}_g = u_g \circ \pi$. Applying equation 106 with $\varphi = u_g$ yields

$$\begin{aligned} \mathbb{E}_{\rho_h^{\mathrm{II}L,\mathrm{loc}}} \mathcal{R}_{\mathrm{gen}}(X; g) \le C\|u_g\|_{C^4(\mathcal{M})} \Big[\eta_{\mathrm{II},\ell} + h + \mathbb{E}_{\rho_h^{\mathrm{II}L,\mathrm{loc}}} \mathrm{dist}(X, \mathcal{M})\Big] \\ + \|\mathcal{L}u_g\|_\infty \mathbb{E}_{\rho_h^{\mathrm{II}L,\mathrm{loc}}}[1 - \vartheta_X^2] + C_m h^m. \end{aligned}$$

By Lemma 17,

$$\|u_g\|_{C^4(\mathcal{M})} + \|\mathcal{L}u_g\|_\infty \le C\|g\|_{C^{2,\alpha}(\mathcal{K}_{\tau_\star})}.$$

Since $g \in \mathcal{F}$, this quantity is uniformly bounded. Substituting the stationary normal and cutoff-mass estimates gives the claimed bound on $\mathcal{R}_{\mathrm{gen}}$.

Weak stationary bias. Finally, Proposition 19 gives

$$\mathrm{dist}_{\mathcal{F}}(\rho_h^{\mathrm{II}L,\mathrm{loc}}, \mu) \le \mathbb{E}_{\rho_h^{\mathrm{II}L,\mathrm{loc}}} \mathcal{R}_\perp + \sup_{g \in \mathcal{F}} \mathbb{E}_{\rho_h^{\mathrm{II}L,\mathrm{loc}}} \mathcal{R}_{\mathrm{gen}}(X; g).$$

Combining the preceding two estimates proves equation 119. For fixed $\alpha_{\mathrm{II}} > 0$ and fixed $\delta_0 > 0$, the quadratic cutoff term is lower order, yielding

$$\mathrm{dist}_{\mathcal{F}}(\rho_h^{\mathrm{II}L,\mathrm{loc}}, \mu) = O(\eta_{\mathrm{II},\ell} + h).$$

□

I Complete Localized Baseline Bounds

Theorem 34 (Stationary bounds for localized HL and π)**.** *Under Assumptions 5–7 and Assumption 1, there exist $h_0, \varepsilon_0, c_H, C > 0$ such that the following holds for $0 < h \le h_0$ and oracle errors at most ε_0. For HL, assume $c_H < \ell \le \ell_+$ and set $\alpha_H := \ell - c_H > 0$. Every invariant law $\rho_h^{HL,\mathrm{loc}}$ satisfies, uniformly over $g \in \mathcal{F}$,*

$$\begin{aligned} \mathbb{E}\mathcal{R}_\perp &\le C\alpha_H^{-1/2}(\eta_{H,\ell} + h^{1/2}), \\ \mathbb{E}\mathcal{R}_{\mathrm{gen}}(X;g) &\le C\Big[\eta_{H,\ell} + h + \alpha_H^{-1/2}(\eta_{H,\ell} + h^{1/2}) + \delta_0^{-2}\alpha_H^{-1}(\eta_{H,\ell}^2 + h)\Big], \\ \mathrm{dist}_{\mathcal{F}}(\rho_h^{HL,\mathrm{loc}}, \mu) &\le C\Big[\eta_{H,\ell} + h + \alpha_H^{-1/2}(\eta_{H,\ell} + h^{1/2}) + \delta_0^{-2}\alpha_H^{-1}(\eta_{H,\ell}^2 + h)\Big]. \end{aligned} \tag{120}$$

For the localized hard-projection scheme, no positive landing condition is needed, and every invariant law $\rho_h^{\pi,\mathrm{loc}}$. satisfies

$$\begin{aligned} \mathbb{E}\mathcal{R}_\perp &\le \varepsilon_\pi, \\ \mathbb{E}\mathcal{R}_{\mathrm{gen}}(X;g) &\le C\left(\varepsilon_P + \varepsilon_\pi + h + \frac{\varepsilon_\pi}{h}\right), \\ \mathrm{dist}_{\mathcal{F}}(\rho_h^{\pi,\mathrm{loc}}., \mu) &\le C\left(\varepsilon_P + \varepsilon_\pi + h + \frac{\varepsilon_\pi}{h}\right). \end{aligned} \tag{121}$$

Proof. For HL, Proposition 11 supplies an invariant law and Corollary 28 gives the first line of equation 120 and the cutoff-mass estimate. Let $g \in \mathcal{F}$ and use the Poisson solution from Lemma 17. Integrating equation 116 with $s = HL$ and using the uniform Schauder estimate gives

$$\begin{aligned} \mathbb{E}\mathcal{R}_{\mathrm{gen}}(X;g) \le C\Big[&\eta_{H,\ell} + h + \alpha_H^{-1/2}(\eta_{H,\ell} + h^{1/2}) \\ &+ \delta_0^{-2}\alpha_H^{-1}(\eta_{H,\ell}^2 + h)\Big], \end{aligned}$$

which is the second line. Proposition 19 then gives the third line.

For π, Proposition 11 again gives an invariant law. Lemma 29 shows that it is supported where $\vartheta_X = 1$ and $\mathrm{dist}(X, \mathcal{M}) \le \varepsilon_\pi$. Integrating equation 117, using the uniform Schauder estimate, and then applying Proposition 19 proves equation 121. □

Corollary 35 (Exact-geometry specialization)**.** *Set $\varepsilon_\pi = \varepsilon_P = \varepsilon_H = \varepsilon_{\mathrm{II}} = 0$. For fixed landing strengths satisfying $\ell > \max\{c_{\mathrm{II}}, c_H\}$,*

$$\begin{aligned} \mathrm{dist}_{\mathcal{F}}(\rho_h^{\mathrm{II}L}, \mu) &= \mathcal{O}(h), \\ \mathrm{dist}_{\mathcal{F}}(\rho_h^{HL,\mathrm{loc}}, \mu) &= \mathcal{O}(h^{1/2}), \\ \mathrm{dist}_{\mathcal{F}}(\rho_h^{\pi}, \mu) &= \mathcal{O}(h). \end{aligned} \tag{122}$$

The π conclusion does not require a landing-strength condition.

Proof. Apply Theorem 33 and Theorem 34 with all oracle errors set to zero. □

I.1 Stationary normal-deviation separation on the circle

The preceding results provide general upper bounds for localized schemes. We now analyze the nominal exact-geometry schemes directly on S^1, showing that the different stationary normal-deviation orders can actually occur.

Proposition 36. *On $\mathcal{M} = S^1$, take $U \equiv 0$, $\beta = \ell = 1$, and $h := \Delta t \in (0, 1)$. The nominal exact-geometry HL and $\mathrm{II}L$ schemes, without cutoffs, admit invariant laws. For any such laws, their stationary radii satisfy, as $h \downarrow 0$,*

$$\mathbb{E}|R_h^{\mathrm{HL}} - 1| = \sqrt{h/\pi} + o(\sqrt{h}), \qquad \mathbb{E}|R_h^{\mathrm{II}L} - 1| = \tfrac{3}{4}h + O(h^2). \tag{123}$$

Thus their stationary mean manifold distances have orders $\Theta(\sqrt{h})$ and $\Theta(h)$, respectively.

Proof. Write $X = Rp$, where $p = X/\|X\|$, and let $t(p)$ be a unit tangent. Under our convention, $H(p) = -p$ and $\mathrm{II}_p(v, v) = Z^2 p$ for $v = Zt(p)$, $Z \sim \mathcal{N}(0, 1)$. Hence $-\mathrm{II}_p(v, v) = Z^2 H(p)$: even in one dimension, Gaussian magnitude distinguishes the corrections. Setting $a = 1 - h$, the updates give

$$(R_{k+1}^{\mathrm{HL}})^2 = a^2 (R_k^{\mathrm{HL}})^2 + 2hZ_k^2, \tag{124}$$

$$(R_{k+1}^{\mathrm{IIL}})^2 = \left[aR_k^{\mathrm{IIL}} + h(1 - Z_k^2)\right]^2 + 2hZ_k^2, \tag{125}$$

with independent standard Gaussian Z_k. Positive radii remain positive, so the projections are well defined.

Stationarity. Both radius maps F satisfy $|F(r, z) - F(s, z)| \le a|r - s|$ by the reverse triangle inequality. Since $F(1, Z)$ has all moments, successive backward-iteration differences have L^q norms bounded by $C_q a^n$. Their summability gives a stationary radial law with finite moments; synchronous contraction gives uniqueness. An independent uniform angle lifts this law to an invariant planar law, and every invariant planar law has this same radial marginal.

HL. The stationary squared radius is

$$S_h = (R_h^{\mathrm{HL}})^2 \overset{d}{=} 2h \sum_{j \ge 0} a^{2j} Z_j^2, \qquad \mathbb{E}S_h = \frac{2}{2 - h}, \qquad \mathrm{Var}(S_h) = \frac{8h^2}{1 - a^4} = 2h + O(h^2).$$

The centered summands have Lyapunov third-moment ratio $O(\sqrt{h})$, so $(S_h - 1)/\sqrt{h} \Rightarrow \mathcal{N}(0, 2)$. Consequently, $(\sqrt{S_h} - 1)/\sqrt{h} \Rightarrow \mathcal{N}(0, 1/2)$. Moreover, $\mathbb{E}(\sqrt{S_h} - 1)^2/h \le \mathbb{E}(S_h - 1)^2/h = O(1)$. Uniform integrability therefore gives $\mathbb{E}|R_h^{\mathrm{HL}} - 1|/\sqrt{h} \to 1/\sqrt{\pi}$.

IIL. For $e \ge 0$, its radius map satisfies

$$F(1 + e, z)^2 = 1 + 2ae + (ae - hz^2)^2 \ge 1, \qquad 0 \le F(1 + e, z) - 1 \le ae + \tfrac{1}{2}h^2 z^4.$$

Thus $[1, \infty)$ is invariant, and the stationary excess $e_h = R_h^{\mathrm{IIL}} - 1$ is nonnegative. Taking L^2 norms in the second inequality at stationarity gives $\|e_h\|_2 \le (h/2)\sqrt{\mathbb{E}Z^8} = O(h)$. Taking expectations in equation 125 yields

$$\mathbb{E}(R_h^{\mathrm{IIL}})^2 = \frac{2h + 2h^2}{1 - a^2}, \qquad \mathbb{E}e_h = \frac{1}{2}\left(\frac{3h}{2 - h} - \mathbb{E}e_h^2\right) = \tfrac{3}{4}h + O(h^2).$$

Since $e_h \ge 0$, this is the stationary mean distance to S^1. □

The example establishes attained normal-deviation orders for the nominal schemes, not a sharp separation in smooth weak bias.

J Finite-difference estimators for II and H

Assumption 37 (Uniform score-operator accuracy for curvature queries)**.** Fix a query radius $0 < \tau_{\mathrm{fd}} < \tau$ and a maximal finite-difference scale $\delta_{\mathrm{fd}} > 0$. Suppose that $\widehat{\pi}$ and $\widehat{J}$ are continuous on the compact sets below and that there exist $\eta_\pi, \eta_J \ge 0$ such that

$$\begin{aligned} \sup_{x \in \mathcal{K}_{\tau_{\mathrm{or}}}} \|\widehat{\pi}(x) - \pi(x)\| &\le \eta_\pi, \\ \sup_{y \in \mathcal{K}_{\tau_{\mathrm{fd}}}} \|\widehat{J}(y) - D\pi(y)\|_{\mathrm{op}} &\le \eta_J. \end{aligned} \tag{126}$$

Assume additionally that

$$\eta_\pi + \delta_{\mathrm{fd}} < \tau_{\mathrm{fd}}. \tag{127}$$

Then every shifted point $\widehat{p} \pm \delta q$ appearing in equation 152, with $0 < \delta \le \delta_{\mathrm{fd}}$ and $\|q\| = 1$, belongs to $\mathcal{K}_{\tau_{\mathrm{fd}}}$.

Proposition 38 (Score-derived directional curvature accuracy)**.** *Suppose $\widehat{\pi}_\sigma$ and $\widehat{J}_\sigma$ approximate π and $D\pi$ uniformly with errors η_π and η_J, respectively. Assume that all shifted evaluations $\widehat{p} \pm \delta\widehat{q}$ in*

equation 17 remain in the region where the Jacobian error bound holds, and that $0 < \delta \le \delta_0$ *and* $\eta_J + C_0\eta_\pi \le c_0$*, for fixed geometric constants* $\delta_0, C_0, c_0 > 0$*. Then*

$$\varepsilon_{\mathrm{II}} \le C_{\mathrm{II}} \left(\delta^2 + \frac{\eta_J}{\delta} + \eta_J + \eta_\pi \right).$$

Here C_{II} *depends only on fixed geometric and dimensional quantities. If* $\delta \asymp \eta_J^{1/3}$ *also satisfies* $\delta \le \delta_0$*, then* $\varepsilon_{\mathrm{II}} = \mathcal{O}(\eta_J^{2/3} + \eta_\pi)$*. The precise query regions, constants, and the construction of a continuous mean-curvature oracle for the HL reference scheme are given in Appendix J.*

Theorem 39 (Finite-difference curvature approximation)**.** *Suppose Assumptions 5 and 37 hold. There exist constants* $c, C > 0$ *such that, whenever* $0 < \delta \le \delta_{\mathrm{fd}}$ *and* $\eta_J + C\eta_\pi \le c$*, the following statements hold.*

For exact zeroth- and first-order operators,

$$\begin{aligned} \|\mathrm{II}_p^{(\delta)}(v,v) - \mathrm{II}_p(v,v)\| &\le C\delta^2\|v\|^2, \\ \|H^{(\delta)}(p) - H(p)\| &\le Cr\delta^2. \end{aligned} \tag{128}$$

For the learned estimator in equation 152, coupled with $v = P(p)\xi$ *through the same Gaussian vector as* $\widehat{v} = \widehat{J}(\widehat{p})\xi$*,*

$$\begin{aligned} \left(\mathbb{E}_\xi \left[\left\| \widehat{\mathrm{II}}_{\delta,\widehat{p}}(\widehat{v},\widehat{v}) - \mathrm{II}_p(v,v) \right\|^2 \right] \right)^{1/2} &\le C_d \left(\delta^2 + \frac{\eta_J}{\delta} + \eta_J + \eta_\pi \right), \\ \|\widehat{H}^{(\delta)}(\widehat{p}) - H(p)\| &\le Cr \left(\delta^2 + \frac{\eta_J}{\delta} + \eta_J + \eta_\pi \right). \end{aligned} \tag{129}$$

Moreover,

$$\begin{aligned} \varepsilon_\pi &\le \eta_\pi, \\ \varepsilon_P &\le C(\eta_J + \eta_\pi), \\ \varepsilon_{\mathrm{II}} &\le C_d \left(\delta^2 + \frac{\eta_J}{\delta} + \eta_J + \eta_\pi \right), \\ \varepsilon_H &\le Cr \left(\delta^2 + \frac{\eta_J}{\delta} + \eta_J + \eta_\pi \right). \end{aligned} \tag{130}$$

For $0 < \eta_J \le c\delta_{\mathrm{fd}}^3$*, choosing* $\delta \asymp \eta_J^{1/3}$ *(and, more generally, truncating this choice at* δ_{fd}*) gives*

$$\varepsilon_{\mathrm{II}} = \mathcal{O}(\eta_J^{2/3} + \eta_\pi), \qquad \varepsilon_H = \mathcal{O}\Big(r(\eta_J^{2/3} + \eta_\pi) \Big). \tag{131}$$

Along this optimized choice, the quadratic-growth constants of the diagonal II *estimator and the fixed-ambient-frame* H *estimator are bounded uniformly for* η_J *in a fixed sufficiently small interval. When* $\eta_J = 0$*, the optimized statement is understood by taking* $\delta \downarrow 0$*.*

This section contains the complete consistency and error analysis for the finite-difference estimators. A key notational point is that

$$J(y) := D\pi(y), \qquad y \in \mathcal{T}_\tau, \tag{132}$$

is the projection Jacobian on the whole tube. Only at a manifold point $p \in \mathcal{M}$ does $J(p) = P(p)$ become the orthogonal tangent projector. The finite difference evaluates J at off-manifold points, so this distinction is essential.

Fix the scale δ_{fd} from Assumption 37; all shifted query points below remain in the compact tube $\mathcal{K}_{\tau_{\mathrm{fd}}}$. For a matrix-valued map A, a point z, and a unit vector q, define

$$\mathcal{D}_\delta A(z;q) := \frac{A(z+\delta q)q - A(z-\delta q)q}{2\delta}, \qquad 0 < \delta \le \delta_{\mathrm{fd}}. \tag{133}$$

J.1 Exact zeroth- and first-order operators

For $p \in \mathcal{M}$ and $v \in T_p\mathcal{M}$, define, when $v \neq 0$,

$$\begin{aligned} q_v &:= \frac{v}{\|v\|}, \\ \mathrm{II}_p^{(\delta)}(v,v) &:= -\|v\|^2 N(p) \mathcal{D}_\delta J(p; q_v), \end{aligned} \tag{134}$$

with $\mathrm{II}_p^{(\delta)}(0,0) = 0$.

Lemma 40 (Second fundamental form from the projection Jacobian)**.** *For $p \in \mathcal{M}$ and $u, v \in T_p\mathcal{M}$,*

$$\mathrm{II}_p(u, v) = -N(p)DJ(p)[u]v. \tag{135}$$

Proof. The identity is precisely the convention in equation 21. For completeness, consider the tangent field $V(y) = P(y)v$ along $\mathcal{M}$. Its ambient directional derivative at p in direction u is $DP(p)[u]v = DJ(p)[u]v$. The normal component is the classical vector-valued second fundamental form up to the sign convention adopted in the paper, giving equation 135. □

Lemma 41 (Second-order centered-difference error)**.** *Let q be a unit vector and suppose $\{p + tq : |t| \le \delta\} \subset \mathcal{T}_\tau$. Then*

$$\|\mathcal{D}_\delta J(p; q) - DJ(p)[q]q\| \le \frac{\delta^2}{6} \sup_{|t|\le\delta} \|D^3 J(p + tq)\|_{\mathrm{op}}. \tag{136}$$

Proof. Set $g(t) = J(p + tq)q$. Since $J = D\pi$ and $\pi \in C^4$, the function g is C^3. Taylor's theorem with integral remainder gives

$$\begin{aligned} g(\delta) &= g(0) + \delta g'(0) + \frac{\delta^2}{2} g''(0) + R_+, \\ g(-\delta) &= g(0) - \delta g'(0) + \frac{\delta^2}{2} g''(0) + R_-, \end{aligned} \tag{137}$$

where $\|R_+\|, \|R_-\| \le \delta^3 \sup_{|t|\le\delta} \|g^{(3)}(t)\|/6$. Subtracting and dividing by 2δ yields

$$\left\| \frac{g(\delta) - g(-\delta)}{2\delta} - g'(0) \right\| \le \frac{\delta^2}{6} \sup_{|t|\le\delta} \|g^{(3)}(t)\|. \tag{138}$$

Moreover,

$$g'(0) = DJ(p)[q]q, \qquad g^{(3)}(t) = D^3 J(p + tq)[q, q, q]q. \tag{139}$$

Because $\|q\| = 1$, the last display implies $\|g^{(3)}(t)\| \le \|D^3 J(p + tq)\|_{\mathrm{op}}$, proving the claim. Notice that $D^3 J = D^4\pi$; this is why fourth-order regularity of π is required. □

Proposition 42 (Exact finite-difference errors for II and H)**.** *Let $M_3 := \sup_{y\in\mathcal{K}_{\tau_{\mathrm{fd}}}} \|D^3 J(y)\|_{\mathrm{op}}$. Then*

$$\left\| \mathrm{II}_p^{(\delta)}(v, v) - \mathrm{II}_p(v, v) \right\| \le \frac{M_3}{6}\delta^2\|v\|^2. \tag{140}$$

If $\{e_j\}_{j=1}^r$ is an orthonormal basis of $T_p\mathcal{M}$ and

$$H^{(\delta)}(p) := -\sum_{j=1}^{r} \mathrm{II}_p^{(\delta)}(e_j, e_j), \tag{141}$$

then

$$\|H^{(\delta)}(p) - H(p)\| \le \frac{rM_3}{6}\delta^2. \tag{142}$$

Proof. For $v \neq 0$, apply Lemma 41 with $q = q_v$. By Lemma 40, bilinearity gives

$$\mathrm{II}_p(v, v) = -\|v\|^2 N(p)DJ(p)[q_v]q_v. \tag{143}$$

Multiplying equation 136 by $\|v\|^2$ and using $\|N(p)\|_{\mathrm{op}} = 1$ proves equation 140. The case $v = 0$ is immediate. Finally, use the sign convention $H(p) = -\sum_{j=1}^r \mathrm{II}_p(e_j, e_j)$ and sum the directional bounds with $\|e_j\| = 1$ to obtain equation 142. □

Remark (basis dependence at finite difference scale). At a fixed nonzero δ, the map $v \mapsto \mathrm{II}_p^{(\delta)}(v, v)$ is two-homogeneous but need not be the diagonal of an exactly bilinear form. Consequently, $H^{(\delta)}(p)$ in equation 141 may depend on the chosen tangent orthonormal basis. The error bound equation 142 is uniform over that choice. The learned mean-curvature oracle below is therefore defined by a continuous fixed-ambient-frame sum rather than by an arbitrary learned eigenbasis.

Remark (unnormalized finite difference). If one instead evaluates $J(p \pm \delta v)v$, then the same proof gives the less uniform bound $\|\mathrm{II}_p^{(\delta)}(v,v) - \mathrm{II}_p(v,v)\| \le C\delta^2\|v\|^4$. The normalized construction in equation 134 is preferable for Gaussian vectors because the query radius is fixed and the truncation error grows only quadratically in $\|v\|$.

J.2 Uniform stability of the exact centered derivative

The learned-operator proof uses the fact that changing the common base point of an exact centered difference does not create a $1/\delta$ factor.

Lemma 43 (Integral representation and uniform Lipschitz bound)**.** *For every unit vector q,*

$$\mathcal{D}_\delta J(z;q) = \frac{1}{2}\int_{-1}^{1} DJ(z+s\delta q)[q]q\,\mathrm{d}s. \tag{144}$$

Consequently, for $0 < \delta \le \delta_{\mathrm{fd}}$, there is a constant C independent of δ such that

$$\begin{aligned} \|\mathcal{D}_\delta J(z;q)\| &\le C, \\ \|\mathcal{D}_\delta J(z;q) - \mathcal{D}_\delta J(z';q')\| &\le C\left(\|z-z'\| + \|q-q'\|\right) \end{aligned} \tag{145}$$

for all admissible z, z' and unit q, q'.

Proof. The fundamental theorem of calculus gives

$$J(z+\delta q)q - J(z-\delta q)q = \int_{-\delta}^{\delta} DJ(z+tq)[q]q\,\mathrm{d}t, \tag{146}$$

which becomes equation 144 after the change of variables $t = s\delta$. Uniform boundedness of DJ proves the first inequality. For the second, subtract the two integral representations. Lipschitz continuity of DJ and bilinearity in its two vector arguments give, uniformly in $s \in [-1,1]$,

$$\begin{aligned} &\|DJ(z+s\delta q)[q]q - DJ(z'+s\delta q')[q']q'\| \\ &\quad\le C\left(\|z-z'\| + \delta\|q-q'\| + \|q-q'\|\right). \end{aligned} \tag{147}$$

Since $\delta \le \delta_{\mathrm{fd}}$, integration proves the claim. □

For $w \neq 0$, define the exact finite-difference map on the tube by

$$\mathcal{Q}_\delta(z,w) := -\|w\|^2\big(I - J(z)\big)\mathcal{D}_\delta J\left(z; \frac{w}{\|w\|}\right), \tag{148}$$

and set $\mathcal{Q}_\delta(z,0) = 0$. At a manifold point, $\mathcal{Q}_\delta(p,v) = \mathrm{II}_p^{(\delta)}(v,v)$.

Lemma 44 (Stability in the base point and direction)**.** *There exists $C > 0$, independent of $\delta \in (0, \delta_{\mathrm{fd}}]$, such that*

$$\begin{aligned} \|\mathcal{Q}_\delta(z,w) - \mathcal{Q}_\delta(z',w')\| \le{}& C\big(\|w\|^2 + \|w'\|^2\big)\|z-z'\| \\ &+ C\big(\|w\| + \|w'\|\big)\|w-w'\|. \end{aligned} \tag{149}$$

Proof. Write $F_z(q) := \mathcal{D}_\delta J(z;q)$ for $\|q\| = 1$. By Lemma 43, F_z is uniformly bounded and is uniformly Lipschitz in both z and q. Define $G_z(0) = 0$ and $G_z(w) = \|w\|^2 F_z(w/\|w\|)$ for $w \neq 0$. We first control the direction variable. If one of w, w' is zero, then

$$\|G_z(w) - G_z(w')\| \le C(\|w\| + \|w'\|)\|w - w'\|.$$

Otherwise, assume without loss of generality that $a := \|w\| \le b := \|w'\|$, and put $q = w/a$, $q' = w'/b$. Then

$$\begin{aligned} \|G_z(w) - G_z(w')\| &\le |a^2 - b^2|\,\|F_z(q')\| + a^2\|F_z(q) - F_z(q')\| \\ &\le C(a+b)\|w-w'\| + Ca^2\|q-q'\|. \end{aligned}$$

The normalization inequality $a\|q-q'\| \le 2\|w-w'\|$ yields the required direction bound. The base-point bound follows directly from the Lipschitz dependence of F_z and $J(z)$ on z:

$$\|\mathcal{Q}_\delta(z,w) - \mathcal{Q}_\delta(z',w)\| \le C\|w\|^2\|z-z'\|.$$

Splitting the full difference into a base-point part and a direction part proves equation 149. No differentiability of the angular map at $w = 0$ is needed. □

J.3 Learned zeroth- and first-order operators

Let $\widehat{\pi}$ and $\widehat{J}$ denote learned approximations of π and $J = D\pi$. Under Assumption 37,

$$\begin{aligned} \sup_{x \in \mathcal{K}_{\tau_{\mathrm{or}}}} \|\widehat{\pi}(x) - \pi(x)\| &\le \eta_\pi, \\ \sup_{y \in \mathcal{K}_{\tau_{\mathrm{fd}}}} \|\widehat{J}(y) - J(y)\|_{\mathrm{op}} &\le \eta_J, \end{aligned} \tag{150}$$

and $\eta_\pi + \delta_{\mathrm{fd}} < \tau_{\mathrm{fd}}$. Hence all learned finite-difference query points lie in the region where the first-order error bound is valid. For a current x, set

$$\begin{aligned} p &= \pi(x), & \widehat{p} &= \widehat{\pi}(x), \\ P &= J(p), & \widehat{P}_0 &= \widehat{J}(\widehat{p}), \\ N &= I - P, & \widehat{N}_0 &= I - \widehat{P}_0, \\ v &= P\xi, & \widehat{v} &= \widehat{P}_0 \xi. \end{aligned} \tag{151}$$

For $\widehat{v} \neq 0$, define

$$\widehat{\mathrm{II}}_{\delta,\widehat{p}}(\widehat{v}, \widehat{v}) := -\|\widehat{v}\|^2 \widehat{N}_0 \mathcal{D}_\delta \widehat{J}\left(\widehat{p}; \frac{\widehat{v}}{\|\widehat{v}\|}\right), \tag{152}$$

with value zero at $\widehat{v} = 0$.

Lemma 45 (Base tangent and direction perturbations)**.** *Let* $M_1 := \sup_{y \in \mathcal{K}_{\tau_{\mathrm{fd}}}} \|DJ(y)\|_{\mathrm{op}}$ *and set* $\eta_0 := \eta_J + M_1 \eta_\pi$. *Then*

$$\begin{aligned} \|\widehat{P}_0 - P\|_{\mathrm{op}} = \|\widehat{N}_0 - N\|_{\mathrm{op}} &\le \eta_0, \\ \|\widehat{v} - v\| &\le \eta_0 \|\xi\|, \\ \|v\| + \|\widehat{v}\| &\le C\|\xi\|, \end{aligned} \tag{153}$$

provided $\eta_0 \le 1$.

Proof. Insert and subtract $J(\widehat{p})$:

$$\begin{aligned} \|\widehat{J}(\widehat{p}) - J(p)\|_{\mathrm{op}} &\le \|\widehat{J}(\widehat{p}) - J(\widehat{p})\|_{\mathrm{op}} + \|J(\widehat{p}) - J(p)\|_{\mathrm{op}} \\ &\le \eta_J + M_1 \eta_\pi. \end{aligned} \tag{154}$$

The remaining claims follow by applying the two operators to ξ and using uniform boundedness of J and $\widehat{J}$. □

Lemma 46 (Pointwise learned-J perturbation of the finite difference)**.** *For every admissible* z *and unit vector* q,

$$\|\mathcal{D}_\delta \widehat{J}(z; q) - \mathcal{D}_\delta J(z; q)\| \le \frac{\eta_J}{\delta}. \tag{155}$$

Consequently, for $w \neq 0$,

$$\begin{aligned} &\left\| -\|w\|^2 \big(I - \widehat{J}(z)\big) \mathcal{D}_\delta \widehat{J}\left(z; \frac{w}{\|w\|}\right) - \mathcal{Q}_\delta(z, w) \right\| \\ &\qquad \le C\left(\frac{\eta_J}{\delta} + \eta_J\right) \|w\|^2, \end{aligned} \tag{156}$$

when $\eta_J \le 1$.

Proof. Each of the two evaluations of $\widehat{J}$ differs from the corresponding evaluation of J by at most η_J in operator norm. Thus

$$\begin{aligned} &\|\mathcal{D}_\delta \widehat{J}(z; q) - \mathcal{D}_\delta J(z; q)\| \\ &\qquad \le \frac{\|\widehat{J}(z + \delta q) - J(z + \delta q)\|_{\mathrm{op}} + \|\widehat{J}(z - \delta q) - J(z - \delta q)\|_{\mathrm{op}}}{2\delta} \le \frac{\eta_J}{\delta}. \end{aligned} \tag{157}$$

Also $\|\widehat{J}(z) - J(z)\|_{\mathrm{op}} \le \eta_J$, while Lemma 43 bounds $\mathcal{D}_\delta J(z; q)$ uniformly. Splitting the difference into the normal-factor perturbation and the centered-derivative perturbation proves equation 156. The product η_J^2/δ is absorbed because $\eta_J \le 1$. □

Proposition 47 (Learned finite-difference error for the pathwise II term)**.** *Under equation 150, for* $0 < \delta \le \delta_{\mathrm{fd}}$ *and sufficiently small* η_0*,*

$$\begin{aligned}&\left\|\widehat{\mathrm{II}}_{\delta,\widehat{p}}(\widehat{v},\widehat{v}) - \mathrm{II}_p(v,v)\right\| \\ &\qquad \le C\left(\delta^2 + \frac{\eta_J}{\delta} + \eta_J + \eta_\pi\right)\|\xi\|^2.\end{aligned} \tag{158}$$

Hence

$$\begin{aligned}&\left(\mathbb{E}_\xi\left[\left\|\widehat{\mathrm{II}}_{\delta,\widehat{p}}(\widehat{v},\widehat{v}) - \mathrm{II}_p(v,v)\right\|^2\right]\right)^{1/2} \\ &\qquad \le C_d\left(\delta^2 + \frac{\eta_J}{\delta} + \eta_J + \eta_\pi\right),\end{aligned} \tag{159}$$

where C_d *contains the Gaussian fourth moment.*

Proof. Add and subtract the exact finite-difference maps at $(\widehat{p},\widehat{v})$ and (p,v):

$$\begin{aligned}&\widehat{\mathrm{II}}_{\delta,\widehat{p}}(\widehat{v},\widehat{v}) - \mathrm{II}_p(v,v) \\ &\quad = \left[\widehat{\mathrm{II}}_{\delta,\widehat{p}}(\widehat{v},\widehat{v}) - \mathcal{Q}_\delta(\widehat{p},\widehat{v})\right] \\ &\qquad + \left[\mathcal{Q}_\delta(\widehat{p},\widehat{v}) - \mathcal{Q}_\delta(p,v)\right] \\ &\qquad + \left[\mathrm{II}_p^{(\delta)}(v,v) - \mathrm{II}_p(v,v)\right].\end{aligned} \tag{160}$$

Lemma 46 and Lemma 45 bound the first bracket by

$$C\left(\frac{\eta_J}{\delta} + \eta_J\right)\|\xi\|^2. \tag{161}$$

For the second bracket, Lemma 44, $\|\widehat{p} - p\| \le \eta_\pi$, and equation 153 give

$$\begin{aligned}\|\mathcal{Q}_\delta(\widehat{p},\widehat{v}) - \mathcal{Q}_\delta(p,v)\| &\le C(\eta_\pi + \eta_0)\|\xi\|^2 \\ &\le C(\eta_J + \eta_\pi)\|\xi\|^2.\end{aligned} \tag{162}$$

The final bracket is bounded by $C\delta^2\|v\|^2 \le C\delta^2\|\xi\|^2$ using Proposition 42. This proves the pathwise estimate. Taking the L^2 norm and using $(\mathbb{E}\|\xi\|^4)^{1/2} < \infty$ proves equation 159. □

Corollary 48 (Curvature-oracle rate and uniform growth)**.** *The estimator equation 152 satisfies the diagonal curvature part of Assumption 1 with*

$$\varepsilon_{\mathrm{II}} \le C_d\left(\delta^2 + \frac{\eta_J}{\delta} + \eta_J + \eta_\pi\right). \tag{163}$$

It also satisfies

$$\left\|\widehat{\mathrm{II}}_{\delta,\widehat{p}}(w,w)\right\| \le C\left(1 + \frac{\eta_J}{\delta}\right)\|w\|^2. \tag{164}$$

Hence, if $0 < \eta_J \le c\delta_{\mathrm{fd}}^3$ *and* $\delta \asymp \eta_J^{1/3}$*, both the oracle error* $\varepsilon_{\mathrm{II}} = O(\eta_J^{2/3} + \eta_\pi)$ *and the quadratic-growth constant are uniform for* η_J *in a fixed sufficiently small interval.*

Proof. The error bound is Proposition 47. For the growth estimate, use uniform boundedness of $I - \widehat{J}(\widehat{p})$ and

$$\|\mathcal{D}_\delta\widehat{J}(\widehat{p};q)\| \le \|\mathcal{D}_\delta J(\widehat{p};q)\| + \frac{\eta_J}{\delta} \le C + \frac{\eta_J}{\delta}$$

for every unit q. Under $\delta \asymp \eta_J^{1/3}$, $\eta_J/\delta = O(\eta_J^{2/3})$, which proves the uniform-growth statement. □

J.4 A CONTINUOUS FIXED-AMBIENT-FRAME LEARNED ESTIMATOR FOR MEAN CURVATURE

The raw learned Jacobian need not be an exact rank-r orthogonal projector. Define

$$A_0 := \mathrm{sym}(\widehat{P}_0) = \frac{1}{2}(\widehat{P}_0 + \widehat{P}_0^\top), \tag{165}$$

and let $\widehat{\Pi}$ be the spectral projector onto the r largest eigenvalues of A_0.

Lemma 49 (Stability and continuity of the learned tangent projector)**.** *If $\eta_0 < 1/4$, the top-r spectral cluster of A_0 is separated from the remaining spectrum and*

$$\|\widehat{\Pi} - P\|_{\mathrm{op}} \le C\eta_0. \tag{166}$$

Moreover, $\widehat{\Pi}$ depends continuously on the state whenever $\widehat{J}$ and $\widehat{\pi}$ do.

Proof. Since P is an orthogonal rank-r projector and $\|A_0 - P\|_{\mathrm{op}} \le \eta_0$, Weyl's inequality places the leading r eigenvalues of A_0 in $[1-\eta_0, 1+\eta_0]$ and the remaining ones in $[-\eta_0, \eta_0]$. The gap is at least $1 - 2\eta_0$, and the Davis–Kahan projector estimate gives

$$\|\widehat{\Pi} - P\|_{\mathrm{op}} \le \frac{2\eta_0}{1-2\eta_0} \le 4\eta_0.$$

The Riesz spectral-projector formula on a contour separating the two clusters then gives continuity under the uniform gap. □

Fix once and for all an ambient orthonormal basis $\{f_a\}_{a=1}^d$ of $\mathbb{R}^d$, set $w_a = Pf_a$ and $\widehat{w}_a = \widehat{\Pi} f_a$, and define the fixed-ambient-frame estimator

$$\widehat{H}^{(\delta)}(\widehat{p}) := -\sum_{a=1}^{d} \widehat{\mathrm{II}}_{\delta,\widehat{p}}(\widehat{w}_a, \widehat{w}_a). \tag{167}$$

At finite δ, the learned sum may depend on the chosen fixed ambient orthonormal basis because the diagonal finite-difference map is not exactly quadratic. The basis is fixed globally, so the construction remains continuous; the error estimate below is uniform over the choice of fixed ambient basis. The exact bilinearity of II implies the ambient trace identity

$$H(p) = -\sum_{a=1}^{d} \mathrm{II}_p(w_a, w_a). \tag{168}$$

Indeed, expanding Pf_a in any orthonormal tangent basis and using Parseval's identity reduces the right-hand side to $-\sum_{j=1}^{r} \mathrm{II}_p(e_j, e_j)$.

Proposition 50 (Fixed-ambient-frame finite-difference error for mean curvature)**.** *Under the preceding assumptions,*

$$\|\widehat{H}^{(\delta)}(\widehat{p}) - H(p)\| \le Cr\left(\delta^2 + \frac{\eta_J}{\delta} + \eta_J + \eta_\pi\right). \tag{169}$$

The map $x \mapsto \widehat{H}^{(\delta)}(\widehat{\pi}(x))$ is continuous for every fixed $\delta > 0$, and

$$\|\widehat{H}^{(\delta)}(\widehat{p})\| \le Cr\left(1 + \frac{\eta_J}{\delta}\right). \tag{170}$$

Consequently, the mean-curvature part of Assumption 1 holds with

$$\varepsilon_H \le Cr\left(\delta^2 + \frac{\eta_J}{\delta} + \eta_J + \eta_\pi\right). \tag{171}$$

Along $\delta \asymp \eta_J^{1/3}$, the error is $O(r(\eta_J^{2/3} + \eta_\pi))$ and the growth constant is uniform for small η_J.

Proof. Add and subtract $\mathcal{Q}_\delta(\widehat{p}, \widehat{w}_a)$ and $\mathcal{Q}_\delta(p, w_a)$ in each summand of equation 167. The learned-J perturbation contributes

$$C\left(\frac{\eta_J}{\delta} + \eta_J\right)\sum_{a=1}^{d} \|\widehat{w}_a\|^2 = Cr\left(\frac{\eta_J}{\delta} + \eta_J\right),$$

because $\widehat{\Pi}$ is an orthogonal rank-r projector. The base-point part of Lemma 44 contributes $Cr\eta_\pi$. For the direction part, Cauchy–Schwarz gives

$$\sum_{a=1}^{d}(\|w_a\| + \|\widehat{w}_a\|)\|\widehat{w}_a - w_a\| \\ \leq (\sqrt{\operatorname{tr} P} + \sqrt{\operatorname{tr} \widehat{\Pi}})\|\widehat{\Pi} - P\|_{\mathrm{F}} \leq Cr\eta_0,$$

where $\|\widehat{\Pi} - P\|_{\mathrm{F}} \leq \sqrt{2r}\|\widehat{\Pi} - P\|_{\mathrm{op}}$. Finally, the exact truncation errors sum to

$$C\delta^2 \sum_{a=1}^{d} \|Pf_a\|^2 = Cr\delta^2.$$

Together with equation 168, these estimates prove equation 169.

For continuity, the spectral projector is continuous by Lemma 49. The diagonal map in equation 152 is continuous at $w = 0$ because its norm is bounded by $C(1+\eta_J/\delta)\|w\|^2$, and it is plainly continuous away from zero. Thus the finite sum in equation 167 is continuous. The same quadratic bound, summed over a and using $\sum_a \|\widehat{\Pi} f_a\|^2 = r$, proves equation 170. □

J.5 Instantiation of the scheme-level oracle assumption

Combining Lemma 45, Corollary 48, and Proposition 50 gives

$$\begin{aligned} \varepsilon_\pi &\leq \eta_\pi, \\ \varepsilon_P &\leq C(\eta_J + \eta_\pi), \\ \varepsilon_{\mathrm{II}} &\leq C_d \left(\delta^2 + \frac{\eta_J}{\delta} + \eta_J + \eta_\pi\right), \\ \varepsilon_H &\leq Cr \left(\delta^2 + \frac{\eta_J}{\delta} + \eta_J + \eta_\pi\right). \end{aligned} \tag{172}$$

Thus uniform approximation of the zeroth-order projection and first-order projection Jacobian supplies the diagonal curvature errors, continuity, and uniform growth conditions required by Assumption 1 along the admissible or optimized finite-difference choices. Only the pointwise learned-J error is amplified by $1/\delta$. The base-point error is shared by both sides of the centered difference and enters linearly through the uniform stability of the exact derivative established in Lemma 43.

Proof of Theorem 39. The exact second-fundamental-form and mean-curvature estimates in equation 128 are precisely Proposition 42. The learned pathwise estimate for II follows from Proposition 47 and Corollary 48. The learned mean-curvature estimate follows from Proposition 50; the fixed-ambient-frame construction is well-defined and continuous by Lemma 49 under the smallness condition $\eta_J + C\eta_\pi \leq c$.

The projection estimate $\varepsilon_\pi \leq \eta_\pi$ is part of Assumption 37. Lemma 45 gives $\varepsilon_P \leq C(\eta_J + \eta_\pi)$; the base-point and projector perturbations also make the drift mismatch $O(\eta_J + \eta_\pi)$. Combining these estimates gives equation 130. Finally, when $\eta_J > 0$ is small enough that the selected radius is below δ_{fd}, balancing the centered-difference truncation term δ^2 with the learned first-order operator term η_J/δ yields $\delta \asymp \eta_J^{1/3}$ and equation 131. When $\eta_J = 0$, the same rate is understood by taking $\delta \downarrow 0$. □

Proof of Proposition 38. Identify the main-text directional estimator in equation 17 with equation 152. The query-containment assumption allows us to invoke Assumption 37. Corollary 48 then gives

$$\varepsilon_{\mathrm{II}} \leq C_{\mathrm{II}} \left(\delta^2 + \frac{\eta_J}{\delta} + \eta_J + \eta_\pi\right).$$

Balancing δ^2 and η_J/δ gives $\delta \asymp \eta_J^{1/3}$, provided this choice remains within the admissible query radius, and hence

$$\varepsilon_{\mathrm{II}} = O(\eta_J^{2/3} + \eta_\pi).$$

□

K Additional details and results on experiments

K.1 Controlled synthetic manifolds

We use three manifolds with known embeddings to study how errors in learned geometry affect discretizations of manifold Langevin dynamics. Exact geometric operators are available from the known parametrizations, while learned variants approximate them from frozen diffusion score models. We compare the π-, L-, HL-, and $\mathrm{II}L$-schemes. Table 2 summarizes the score-training and simulation settings.

K.1.1 Manifolds and target distributions

Catenoid. For $u \in [0, 2\pi)$ and $t \in \mathbb{R}$, we use the scaled catenoid

$$F_{\mathrm{cat}}(u,t) = a \begin{bmatrix} \cosh(t)\cos u \\ \cosh(t)\sin u \\ t \end{bmatrix}, \qquad a = 2. \tag{173}$$

The catenoid is minimal: its mean-curvature vector satisfies $H \equiv 0$, while its second fundamental form is nonzero. It therefore isolates the difference between the mean-curvature correction in HL and the realized directional-curvature correction in $\mathrm{II}L$. We use

$$U_{\mathrm{cat}}(x) = \tfrac{3}{2}(x_3/a)^2 - 2\cos\big(3\,\mathrm{atan2}(x_2, x_1)\big), \qquad \mu_{\mathrm{cat}}(\mathrm{d}x) \propto e^{-2U_{\mathrm{cat}}(x)}\,\mathrm{d}\,\mathrm{vol}_{\mathcal{M}}(x).$$

In intrinsic coordinates, the corresponding density is proportional to $e^{4\cos(3u)}\cosh^2(t)e^{-3t^2}$, which allows the reference expectation to be computed accurately.

Deformed torus. For $(u, v) \in [0, 2\pi)^2$, we use the two-dimensional surface

$$F_{\mathrm{tor}}(u,v) = \big(1 + 0.05\sin(4u)\cos(2v)\big) \begin{bmatrix} (8 + 2\cos v)\cos u \\ (8 + 2\cos v)\sin u \\ 2\sin v \end{bmatrix} \in \mathbb{R}^3. \tag{174}$$

Training points and initial states are obtained from independent uniform (u, v), so they are uniform in parameter space rather than surface area. The potential is

$$U_{\mathrm{tor}}(x) = \tfrac{3}{2}x_3^2 - 2\cos\big(3\,\mathrm{atan2}(x_2, x_1)\big), \qquad \mu_{\mathrm{tor}}(\mathrm{d}x) \propto e^{-2U_{\mathrm{tor}}(x)}\,\mathrm{d}\,\mathrm{vol}_{\mathcal{M}}(x).$$

Bumpy sphere. The third example is an eight-dimensional deformation of $\mathbb{S}^8 \subset \mathbb{R}^9$ embedded in $\mathbb{R}^{10}$:

$$r(u) = 2\left[1 + 0.03\tanh\left(3^{-1/2}\sum_{j=1}^{3} w_j \sin(1.5a_j^\top u + \phi_j)\right)\right], \tag{175}$$

$$F_{\mathrm{bump}}(u) = \big[r(u)u^\top,\ 0.01\sin(b^\top u + \psi)\big]^\top. \tag{176}$$

The directions a_j, b, weights w_j, and phases ϕ_j, ψ are fixed random coefficients shared by all experiments. We use

$$U_{\mathrm{bump}}(x) = \tfrac{3}{2}x_3^2 - 2\{\cos(6x_1) + \cos(6x_2)\}, \qquad \mu_{\mathrm{bump}}(\mathrm{d}x) \propto e^{-2U_{\mathrm{bump}}(x)}\,\mathrm{d}\,\mathrm{vol}_{\mathcal{M}}(x).$$

For the torus and bumpy sphere, reference expectations are computed by importance weighting latent samples with the induced volume element and the corresponding Gibbs density.

K.1.2 Exact and score-derived geometry

For all three manifolds, the exact projection π, tangent projector P, second fundamental form II, and mean-curvature vector H are computed from the known geometry. For learned geometry, the frozen score model approximates the nearest projection by

$$\widehat{\pi}_\sigma(x) = x + \sigma^2 s_\theta(x, \sigma),$$

from which we construct the corresponding tangent and curvature operators as described in Section 3.3 and Appendix J. All three experiments use $\sigma_{\mathrm{eval}} = 0.005$. The score models are residual

Table 2: Synthetic experiment settings.

Setting	Catenoid	Deformed torus	Bumpy sphere
Intrinsic / ambient dimension	$2/3$	$2/3$	$8/10$
Training-set size	5,000	6,000	8,000
MLP width / residual blocks	$256/4$	$256/4$	$512/4$
Training noise range	$[10^{-3}, 0.5]$	$[10^{-4}, 1]$	$[10^{-3}, 0.5]$
Simulation seeds	0–4	0–9	0–4
Chains per seed	100	100	50
Landing coefficient ℓ	16	16	15
Horizon / burn-in	$200/50$	$1000/200$	$400/100$
Evaluation noise σ	0.005	0.005	0.005
Curvature FD step	0.1	0.2	0.1
Step size Δt values	8 (1.25×10^{-4}–0.02)	10 (5×10^{-4}–0.12)	7 (10^{-3}–0.012)

MLPs trained by variance-exploding denoising score matching, with the configurations in Table 2. For the catenoid, we use the finite-data score model trained on $5{,}000$ manifold samples.

The retained step-size grids are

$$\begin{aligned} \text{Catenoid:} &\quad \{0.000125, 0.00025, 0.0005, 0.001, 0.002, 0.005, 0.01, 0.02\}, \\ \text{Deformed torus:} &\quad \{0.0005, 0.001, 0.002, 0.005, 0.01, 0.02, 0.05, 0.08, 0.1, 0.12\}, \\ \text{Bumpy sphere:} &\quad \{0.001, 0.002, 0.003, 0.005, 0.008, 0.01, 0.012\}. \end{aligned}$$

Within each seed, all schemes use matched initializations and matched Gaussian noise.

K.1.3 Evaluation and additional results

Motivated by Proposition 19, we measure the normal deviation

$$\mathbb{E}\mathcal{R}_\perp = \widehat{\mathbb{E}}\,\mathrm{dist}(X, \mathcal{M}).$$

The generator discrepancy $\mathcal{R}_{\mathrm{gen}}$ is not estimated directly. Instead, using U as a fixed test function, we report

$$\mathcal{E}_U^\pi = \left|\widehat{\mathbb{E}}[U(\pi(X))] - \mathbb{E}_\mu[U]\right|, \qquad \mathcal{E}_U^{\widehat{\pi}} = \left|\widehat{\mathbb{E}}[U(\widehat{\pi}(X))] - \mathbb{E}_\mu[U]\right|. \tag{177}$$

Here $\mathcal{E}_U^\pi$ uses the known manifold only for evaluation, whereas $\mathcal{E}_U^{\widehat{\pi}}$ evaluates the same statistic through the learned projection and is therefore not a ground-truth stationary error. The absolute potential error is computed within each seed before averaging across seeds.

Figures 6, 7, and 8 show the exact- and learned-geometry sweeps on the three manifolds. Across all three settings, the π-scheme exhibits increasing projected potential error in the small-step regime under learned geometry, consistent with amplification of projection error as Δt decreases, which does not emerge under exact geometry. The landing-based L, HL, and $\mathrm{II}L$ schemes do not exhibit the same characteristic amplification, and $\mathrm{II}L$ generally provides stronger control of normal deviation over the finite step size Δt ranges. The three manifolds give complementary tests of the same mechanism: the catenoid isolates directional curvature through $H \equiv 0$ but $\mathrm{II} \neq 0$, the deformed torus tests a nonuniform curved surface in $\mathbb{R}^3$, and the bumpy sphere extends the comparison to eight intrinsic dimensions. Together, these results support the finite-step error mechanisms identified in Section 3.2.

K.2 Additional MNIST results

Task and purpose. We study noisy MNIST inpainting. Given a partial observation y and one feasible completion x_{ref}, we seek different completions that remain consistent with the observation and preserve the digit class of x_{ref}. The experiment serves two purposes: first, to compare $\mathrm{II}L$ with alternative discretizations under learned geometry; second, to compare its class-preserving conditional exploration with diffusion posterior-sampling baselines. We use 24 fixed inpainting problems with observation noise $\sigma_{\mathrm{obs}} = 0.05$ and generate 32 samples for each method and configuration.

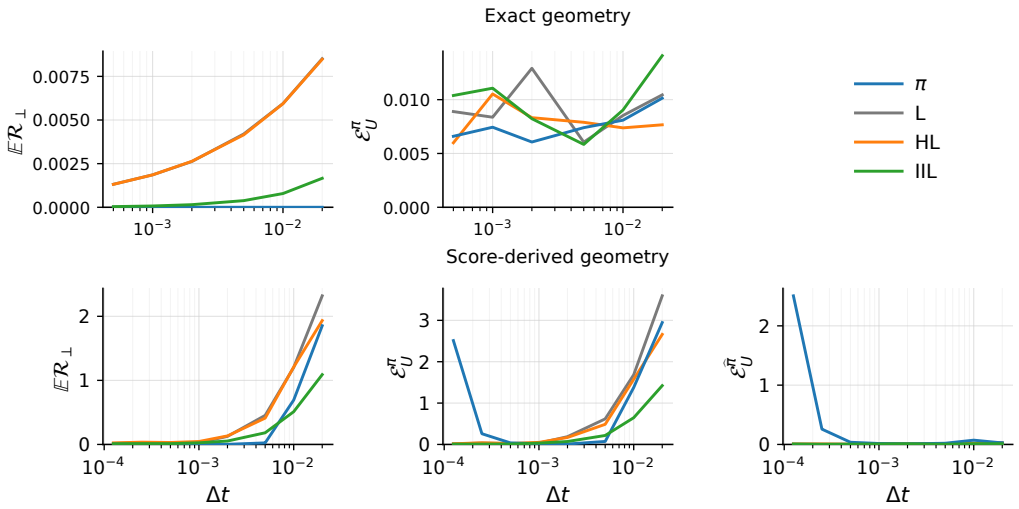


Figure 6: Catenoid sweep under exact and score-derived geometry

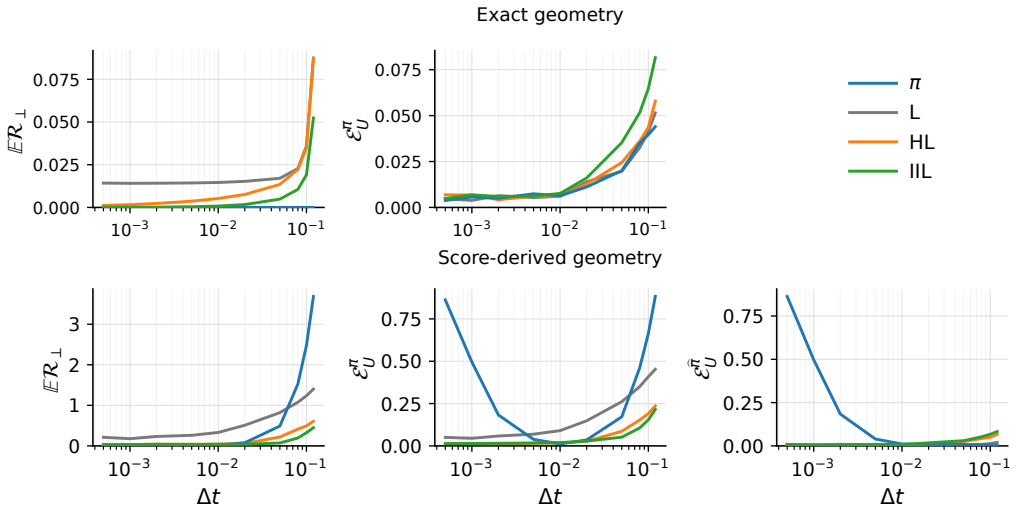


Figure 7: Deformed-torus sweep under exact and score-derived geometry

Compared methods. For the discretization comparison, IIL, Landing-only L, and the projection scheme π all start from x_{ref} and use the same score-derived geometry. For the conditional-exploration comparison, we additionally use standard from-noise DPS (Chung et al., 2023). Standard DPS is conditioned on the partial observation but has no information about which feasible completion we wish to explore around. We therefore also include Ref-DPS, our reference-guided adaptation that gives DPS access to x_{ref}.

Reference-guided DPS. Writing the standard DPS guidance update as

$$x_{i-1}^{\text{DPS}} = x'_{i-1} - \zeta_i \nabla_{x_i} \|y - \mathcal{A}(\hat{x}_0)\|_2^2 , \tag{178}$$

Ref-DPS keeps the original measurement-guidance term and adds only a reference penalty on the missing region:

$$x_{i-1}^{\text{Ref-DPS}} = x_{i-1}^{\text{DPS}} - \zeta_i \lambda_{\text{ref}} \nabla_{x_i} \left[\frac{\|M_{\text{miss}}(\hat{x}_0 - x_{\text{ref}})\|_2^2}{|M_{\text{miss}}|} \right] , \tag{179}$$

where M_{miss} selects the unobserved pixels. Ref-DPS is therefore a reference-aware adaptation introduced for this comparison, rather than an official DPS variant.

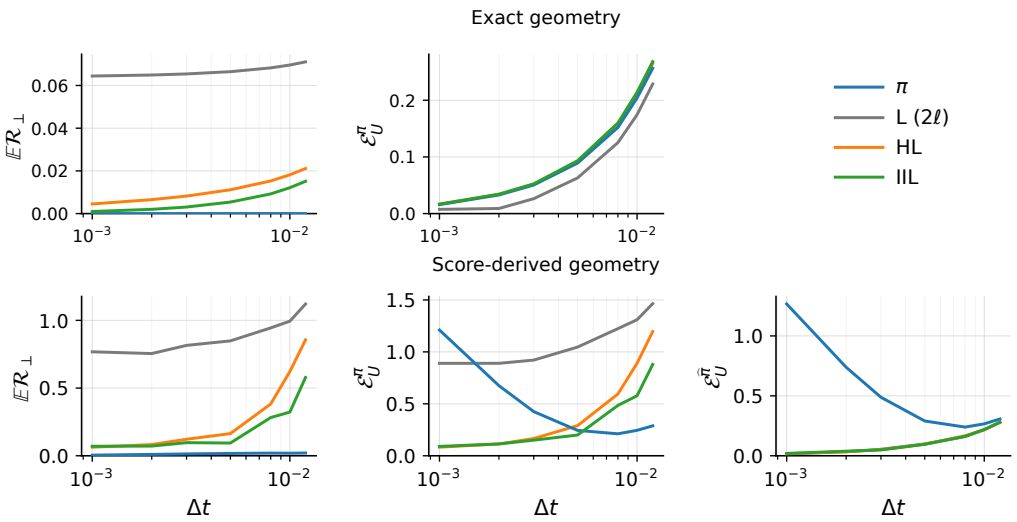


Figure 8: Bumpy-sphere sweep under exact and score-derived geometry.

Evaluation. Our primary semantic criterion, used in the main text, is class consistency. A finite generated image satisfies C when both separately trained frozen MNIST classifiers predict the selected digit class, and we denote its retention rate by $R_C = \Pr(C)$. Observation fidelity is measured against the noisy observation y on the observed pixels; RMSE_C denotes this error averaged over samples satisfying C. Diversity is measured only on the missing region: APD_C is the mean pairwise missing-region distance among accepted completions. We additionally report

$$F_C(r) = \Pr(C(X) \text{ and } \mathrm{RMSE}_{\mathrm{obs}}(X, y) \le r)\,, \tag{180}$$

which measures the fraction of all generated samples that simultaneously preserve the selected class and satisfy an observation-error tolerance r. As a stricter sensitivity check, we use $S \subset C$, which additionally requires both classifiers to assign calibrated probability at least 0.9 to the selected class; the corresponding quantities are denoted S-rate, RMSE_S, APD_S, and $F_S(r)$.

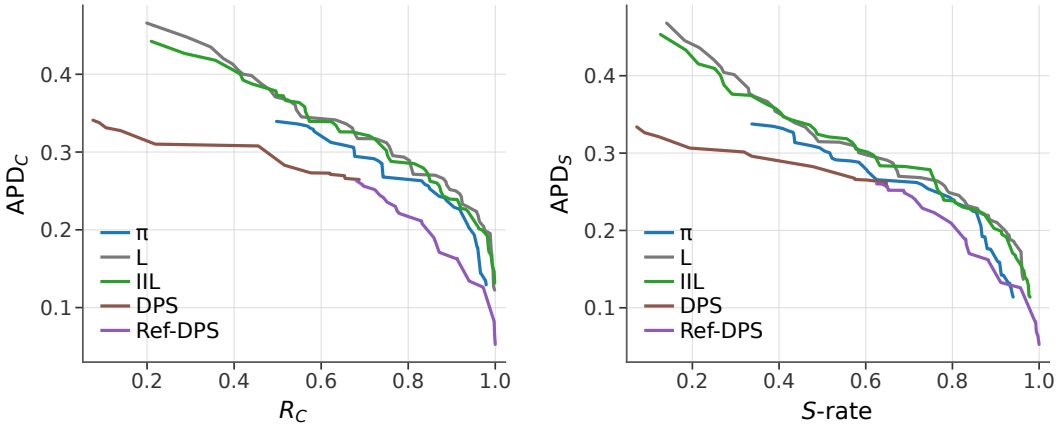
K.3 Retention–diversity trade-offs

Full-range and stricter-gate results. The main paper shows the C-based retention–diversity comparison over a focused display range. Figure 9(a) shows the complete R_C–APD_C frontiers over the evaluated parameter sweeps. DPS extends into a low-retention, high-diversity regime, whereas IIL maintains substantial diversity at higher class retention. Ref-DPS moves DPS toward higher retention by guiding samples toward x_{ref}, with a corresponding reduction in diversity. Figure 9(b) repeats the comparison under the stricter criterion S. The generated samples and parameter sweeps are unchanged; only the semantic acceptance rule differs.

K.4 Discretization robustness

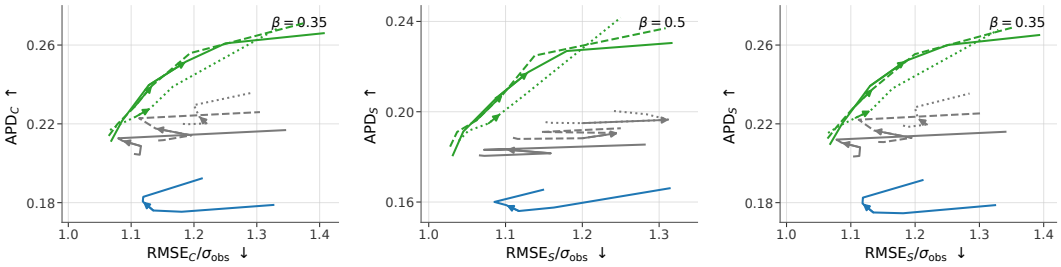
Step-size robustness. The main paper reports the discretization comparison under the primary criterion C at $\beta = 0.5$. Figure 10(a) repeats the same comparison at $\beta = 0.35$, while panels (b,c) use the stricter criterion S at $\beta = 0.5$ and 0.35, respectively. Thus the three panels test whether the observed fidelity–diversity behavior depends on either the inverse temperature or the semantic acceptance rule.

Landing-strength robustness. We next test how the fidelity–diversity behavior changes with the soft-landing strength. We vary $\ell \in \{12.5, 25, 40, 63, 100, 160, 200, 400\}$ at three fixed step sizes $\Delta t \in \{0.001, 0.0016, 0.002\}$. Figure 11 shows the resulting trajectories in the same fidelity–diversity plane as the step-size study. Each trajectory fixes the method and Δt, while the points follow increasing ℓ; arrows indicate this direction. We report both semantic criteria C and S and both inverse temperatures $\beta \in \{0.5, 0.35\}$.



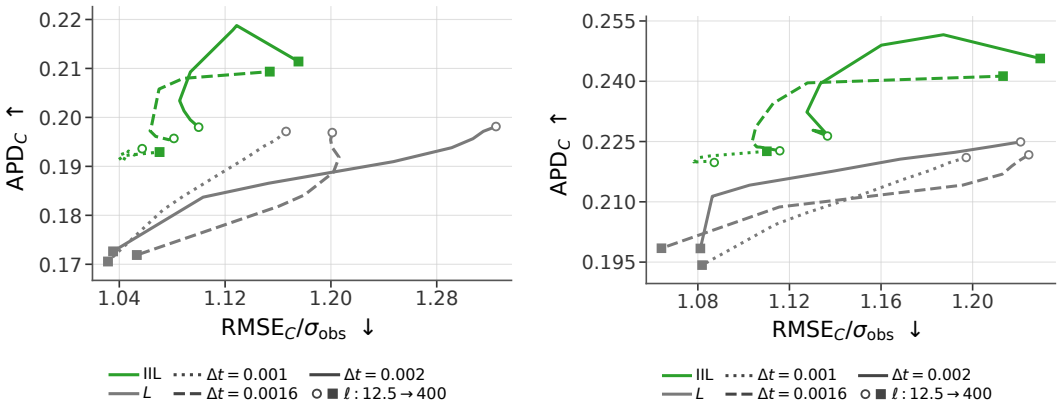
(a) Broad class criterion C. (b) Stricter semantic criterion S.

Figure 9: **Full-range retention–diversity trade-offs. (a)** Broad class-consistency criterion C used in the main text. **(b)** Stricter criterion S, which additionally requires both classifiers to assign calibrated probability at least 0.9 to the selected class.

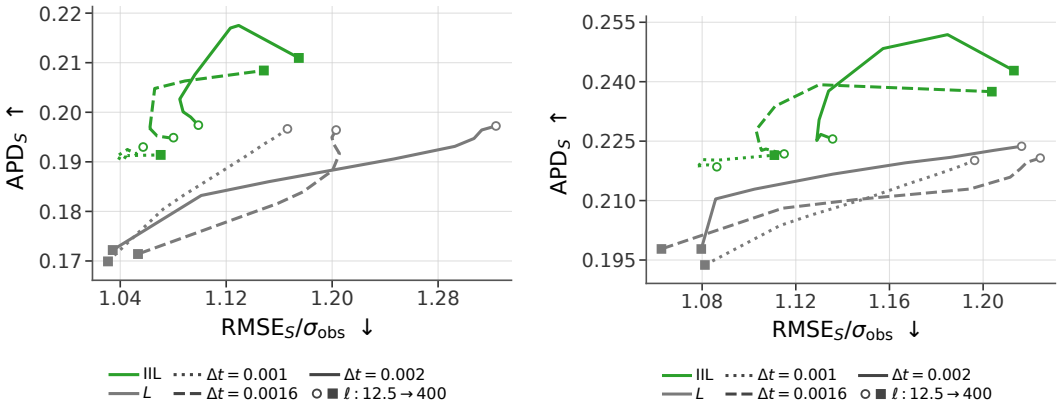


(a) C, $\beta = 0.35$. (b) S, $\beta = 0.5$. (c) S, $\beta = 0.35$.

Figure 10: **Step-size robustness.** The main text reports criterion C at $\beta = 0.5$. **(a)** changes the inverse temperature to $\beta = 0.35$; **(b,c)** use the stricter semantic criterion S at $\beta = 0.5$ and 0.35. Arrows follow increasing step size Δt.

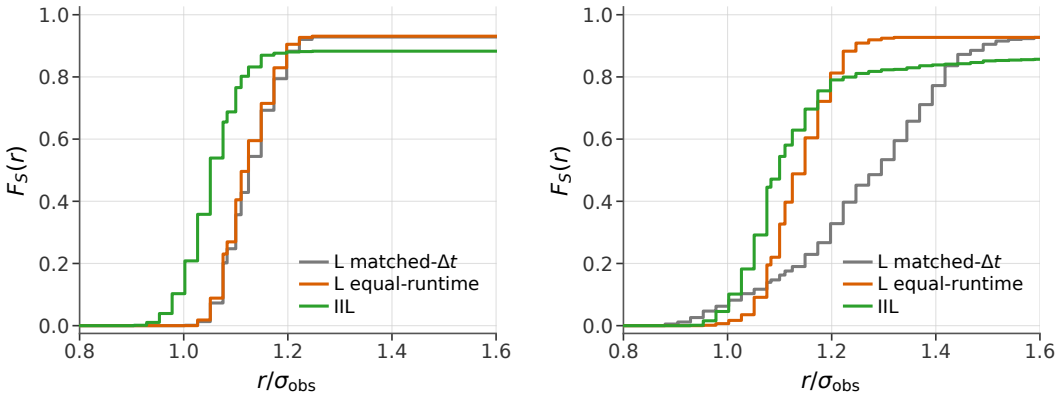


(a) C, $\beta = 0.5$. (b) C, $\beta = 0.35$.



(c) S, $\beta = 0.5$. (d) S, $\beta = 0.35$.

Figure 11: **Landing-strength robustness.** Each trajectory fixes the method and step size, with arrows indicating increasing landing strength ℓ. Green and gray denote IIL and L, respectively; dotted, dashed, and solid curves denote $\Delta t = 0.001$, 0.0016, and 0.002. **(a,b)** use the primary class-consistency criterion C; **(c,d)** use the stricter criterion S.



(a) $\Delta t = 0.001$. (b) $\Delta t = 0.002$.

Figure 12: **Equal-runtime Landing controls.** For each IIL setting, the lower per-step cost of L is reinvested in a finer discretization with approximately matched runtime. The curves report the stricter-gate coverage $F_S(r)$.

K.5 Equal-runtime control

Finer Landing steps. An L update is cheaper than an IIL update, so a natural alternative is to spend the saved computation on more, smaller L steps. We test this at approximately matched sampler runtime using the stricter S-based evaluation. At $\Delta t = 0.001$, finer-step L recovers only 7.2% of the matched-step RMSE_S gap. At $\Delta t = 0.002$, it recovers 81.4% of the gap, but its APD_S remains lower than that of IIL. Thus compute reallocation can substantially improve Landing in some regimes, but does not reproduce the same fidelity–diversity behavior.

Scope. The MNIST experiments use fixed development problems and evaluated parameter sweeps rather than an untouched held-out benchmark with equal hyperparameter-search budgets across methods. They therefore characterize empirical discretization and conditional-exploration behavior under learned geometry, rather than correctness relative to an unavailable true conditional manifold distribution.

K.6 Human-pose editing

Task and metrics. We consider underdetermined editing of 3D Human3.6M poses represented by 17 Cartesian joints. Given a source pose x^{src}, a set of handle joints $\mathcal{H}$ with target positions $\{y_j\}_{j\in\mathcal{H}}$, and anchor joints $\mathcal{A}$ that should remain close to the source, we define

$$E_{\mathrm{h}}(x) = \frac{1}{3|\mathcal{H}|\tau^2}\sum_{j\in\mathcal{H}}\|x_j - y_j\|_2^2, \qquad E_{\mathrm{a}}(x) = \frac{1}{3|\mathcal{A}|\tau^2}\sum_{j\in\mathcal{A}}\|x_j - x_j^{\mathrm{src}}\|_2^2,$$

with $\tau = 10\,\mathrm{mm}$. Sampling uses

$$U_r(x) = \frac{E_{\mathrm{h}}(x) + rE_{\mathrm{a}}(x)}{1+r},$$

where r controls the handle–anchor trade-off, while evaluation uses the r-independent quantity

$$J_{10}(x) = \frac{E_{\mathrm{h}}(x) + E_{\mathrm{a}}(x)}{2}.$$

We additionally report handle and anchor errors in millimeters. Diversity is measured by average pairwise distance (APD) over the joints that are neither handles nor anchors, so that diversity reflects variation only in the unconstrained part of the pose.

Benchmarks. We evaluate on two complementary suites.

Pair is a controlled wrist-edit stress test: each of its 48 instances moves a single left or right wrist to a synthetic target 10% beyond the source arm reach while anchoring the torso and shoulders. Because

Table 3: **Robustness to DPS guidance design.** All variants are swept over the same common validation grid. Max APD is reported in mm.

DPS variant	Pair APD	Real-edit APD	Score NFE	Fixed-compute
Original	**75.979**	**60.942**	2001	yes
FGPS-style schedule	68.932	45.579	2001	yes
Wolfe-style adaptation	74.828	55.192	4179	no
Beta-scheduled	74.514	57.269	2001	no

the constraint template is fixed across instances, Pair isolates the accuracy–diversity behavior of the sampler under a consistent, mildly conflicting edit.

Real-edit tests broader editing behavior using four constraint types—left wrist, right wrist, right ankle, and two hands—with targets taken from future frames of real Human3.6M motion. Thus Pair provides a controlled comparison, while Real-edit evaluates transfer across different handles and naturally occurring target configurations.

Learned pose geometry. We use the released pretrained Human3.6M GFPose model (Ci et al., 2023) without retraining. Projection, tangent, and directional-curvature operators are constructed from the same score model at a fixed diffusion noise level throughout each Langevin trajectory. For the stochastic phase we use the GFPose noise parameter $\sigma_{\mathrm{geom}} = 0.2$. The corresponding score-derived projection defines the tangent operator through its Jacobian-vector product, and the directional second fundamental form used by IIL is evaluated along the same tangent Gaussian direction used for the Brownian increment.

Main fixed-budget comparison. All methods generate 32 hypotheses per editing problem. For IIL, all chains start from the source pose and we use $N = 225$ stochastic steps. We sweep

$$\Delta t \in \{5 \times 10^{-5},\, 7.5 \times 10^{-5},\, 10^{-4}\}, \qquad \beta \in \{0.25, 0.5, 1, 2, 4, 8, 16, 32, 64, 128\},$$

and landing fractions

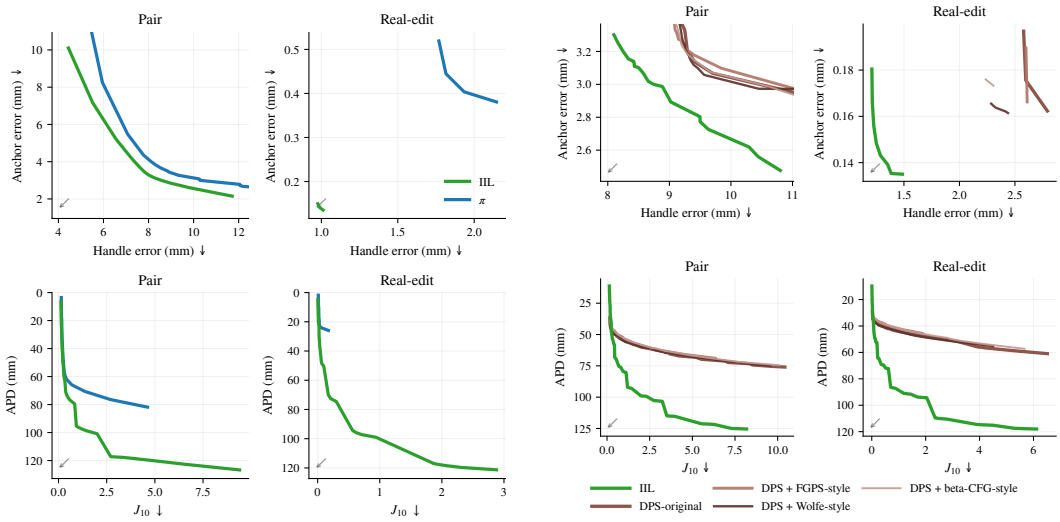
$$\rho \in \{0.075, 0.10, 0.15\}, \qquad \rho = \ell \Delta t.$$

The task-potential multiplier is fixed to $\gamma = 20$. After the stochastic phase, we apply 50 deterministic zero-temperature IIL updates using a lower geometry noise level ($\sigma_{\mathrm{geom}} \approx 0.006$).

DPS uses its native cold Gaussian initialization and 1000 reverse-diffusion steps, with its guidance strength tuned over the validation sweep. The IIL step count was chosen from measured runtime so that the resulting configurations require approximately 0.94–0.98 times the DPS wall-clock cost. Both methods finally receive the same target-independent fixed-skeleton correction.

Comparison with the π-scheme. We additionally isolate the effect of the manifold discretization by comparing IIL with the hard-projection π-scheme. Here both methods use 32 source-initialized chains, $N = 1000$, $\Delta t = 5 \times 10^{-5}$, the same GFPose geometry, and matched Gaussian directions. IIL uses landing with $\rho = 0.10$ and the directional curvature correction, whereas the π-scheme instead applies the learned projection after every stochastic update. Each method uses its frozen deterministic terminal refinement before the same final fixed-skeleton correction. This experiment is intended as a matched stochastic-phase discretization comparison rather than an equal-wall-clock comparison.

Robustness to DPS guidance design. We additionally evaluate three literature-motivated guidance strength variants of DPS while keeping the GFPose prior, editing tasks, objective, initialization, postprocessing, and evaluation protocol fixed. Specifically, we consider an FGPS-style early-to-late cosine schedule (Thaker et al., 2025), a Wolfe-style adaptive guidance rule (Kim et al., 2025), and a beta-CFG-style $\mathrm{Beta}(2, 2)$ schedule (Malarz et al., 2025). The alternative guidance-strength rules substantially reshape the tight-accuracy and interior DPS frontier, confirming that the canonical rule is not the only competitive DPS parametrization. However, none increases the high-diversity ceiling: original DPS still attains the largest APD among all DPS variants on both Pair and Real-edit benchmarks.



(a) IIL versus the π-scheme. (b) Robustness to DPS guidance design.

Figure 13: **Human-pose editing on H36M: discretization and baseline robustness. (a)** Pareto frontiers of IIL and the hard-projection π-scheme on Pair and Real-edit, comparing handle–anchor error (top) and accuracy–diversity trade-offs (bottom). **(b)** Comparison with original DPS and alternative literature-motivated guidance schedules. The alternative guidance rules change substantial portions of the DPS frontier, but do not extend its high-diversity regime. For visualization, diversity is plotted as $-\mathrm{APD}$ so that lower is better on both axes.

K.7 Additional details for FFHQ image inverse problems

Tasks and measurement model. We evaluate the eight FFHQ-256 inverse problems from Section 4.4 on the canonical 100 test images `49000.png`–`49099.png`. Measurements are generated once as

$$y = \mathcal{A}(x) + \sigma_{\mathrm{obs}}\varepsilon, \qquad \varepsilon \sim \mathcal{N}(0, I), \qquad \sigma_{\mathrm{obs}} = 0.05,$$

and are shared across methods. In DAPS, the Gaussian measurement likelihood used by the conditional Langevin sampler is parameterized by a measurement scale β_y,

$$p(y \mid x) \propto \exp\left(-\frac{\|\mathcal{A}(x) - y\|^2}{2\beta_y^2}\right).$$

Following Zhang et al. (2025), we treat β_y as an algorithmic hyperparameter rather than fixing it to the true observation-noise level, and use $\beta_y = 0.01$ for all DAPS and DAPS+IIL experiments. Thus $\sigma_{\mathrm{obs}} = 0.05$ specifies the noise used to generate the observations, whereas $\beta_y = 0.01$ specifies the measurement-conditioning strength used inside the sampler.

DAPS host configurations. DAPS (Zhang et al., 2025) proceeds over N_A annealing levels. At each level, it solves a probability-flow ODE using N_{ode} score-network evaluations to obtain a clean-image estimate, followed by $N_{\mathrm{inner}} = 100$ measurement-conditioned Langevin updates. Following the DAPS notation, we use DAPS-K to denote a host with approximately

$$K = N_A N_{\mathrm{ode}}$$

score-network evaluations. The label therefore describes score NFE rather than total wall-clock cost; all compute comparisons below use measured runtime.

Our standard hosts coincide with the DAPS scales used for the corresponding task families: DAPS-1K for the five linear problems and DAPS-4K for the three nonlinear problems. The reduced- and higher-compute hosts used in our controls are listed in Table 4. DAPS-5K and DAPS-7K are controlled higher-compute extensions used only as stronger native-DAPS controls.

Table 4: **DAPS host configurations used in the FFHQ experiments.**

Host	N_A	N_{ode}	Score NFE	N_{inner}	Role
DAPS-200	50	4	200	100	reduced host
DAPS-400	100	4	400	100	reduced host
DAPS-1K	200	5	1000	100	standard linear host
DAPS-2K	250	8	2000	100	intermediate/control host
DAPS-4K	400	10	4000	100	standard nonlinear host
DAPS-5K	500	10	5000	100	stronger control
DAPS-7K	700	10	7000	100	stronger control

Attaching terminal $\text{II}L$ to DAPS. Our modification leaves the complete upstream DAPS trajectory unchanged. At the final annealing level, native DAPS performs a 100-step conditional Langevin block. We write DAPS+$\text{II}L_M$ for the configuration that retains the first $100 - M$ native terminal updates and replaces only the final M updates by $\text{II}L$ dynamics. Thus $M = 0$ is native DAPS, while $M = 100$ is full terminal replacement. For partial replacement, the endpoint of the native prefix is handed directly to $\text{II}L$. No native DAPS update is applied after the $\text{II}L$ suffix.

The manifold operators are constructed from the same pretrained unconditional FFHQ-256 model used by DAPS (Zhang et al., 2025), which adopts the ADM architecture (Dhariwal & Nichol, 2021) and the public `ffhq_10m.pt` checkpoint released by Chung et al. (2023). We use tangent-projected drift and noise, curvature correction, and soft landing, with geometry refreshed at every $\text{II}L$ update. The landing strength and the finite-difference stepsize are selected on a validation cohort disjoint from the reported test images and are then frozen.

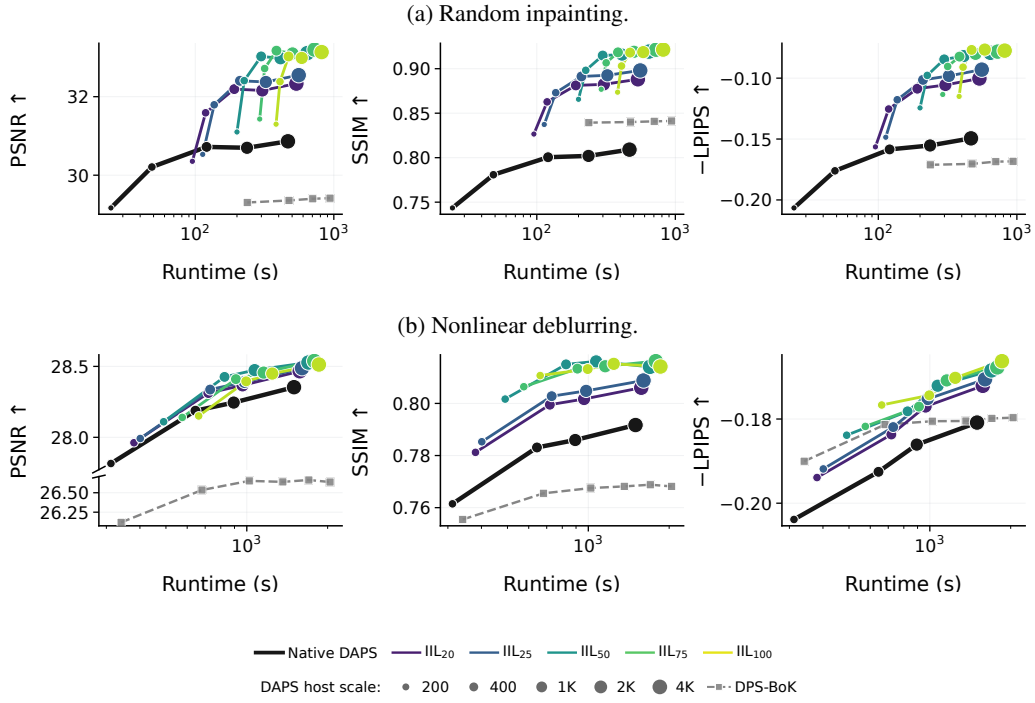


Figure 14: **Runtime–quality scaling on FFHQ-256.** Colored curves show DAPS+$\text{II}L_M$, where the final M native terminal updates are replaced by $\text{II}L$; marker size denotes the DAPS host scale. Black curves show native DAPS and gray dashed curves show DPS-BoK. The nonlinear-deblurring PSNR panel uses a broken vertical axis only to omit an empty interval.

Runtime protocol and compute controls. Score NFE alone does not capture the additional projection, tangent, and curvature operations required by $\text{II}L$, so we compare compute using measured sampler wall-clock time. Timing is measured on an H100 NVL using a frozen 20-image batch after model loading and warm-up, with metric evaluation and serialization excluded.

Table 5: **Reduced-host compute reallocation on FFHQ-256.** A reduced DAPS host with terminal IIL is compared with the standard DAPS host. Runtime is sampler wall-clock time in seconds. Bold denotes the better quality value within each task and metric.

Task	Method	Runtime (s)	PSNR ↑	SSIM ↑	LPIPS ↓
Linear tasks					
Random	DAPS-1K	121.26	30.720	.8005	.1585
	DAPS-400+IIL_{21}	123.01	**31.634**	**.8650**	**.1237**
Box	DAPS-1K	121.57	**24.644**	.7459	.1760
	DAPS-400+IIL_{21}	123.23	24.479	**.7512**	**.1749**
SR $\times 4$	DAPS-1K	132.76	29.378	.7827	.1921
	DAPS-400+IIL_{21}	129.06	**29.969**	**.8352**	**.1754**
Gaussian	DAPS-1K	471.58	29.611	.7864	.1795
	DAPS-200+IIL_{100}	467.21	**30.379**	**.8520**	**.1475**
Motion	DAPS-1K	475.01	31.653	.8366	.1373
	DAPS-200+IIL_{100}	472.39	**32.968**	**.9018**	**.0994**
Nonlinear tasks					
Nonlinear	DAPS-4K	1499.09	28.353	.7917	.1808
	DAPS-2K+IIL_{100}	1243.74	**28.450**	**.8152**	**.1702**
Phase	DAPS-4K	532.16	25.171	.6663	.2867
	DAPS-2K+IIL_{75}	539.97	**26.281**	**.7492**	**.2511**
HDR	DAPS-4K	466.30	27.320	.8412	.1594
	DAPS-2K+IIL_{50}	411.65	**27.592**	**.8610**	**.1471**

We use two complementary compute controls. The *reduced-host* comparison decreases upstream DAPS computation and spends the saved budget on terminal IIL, whereas the *stronger-DAPS* comparison allocates a comparable budget to a larger native DAPS host. These controls test whether the improvement is due to the terminal geometric dynamics rather than additional computation alone.

To examine this trade-off more systematically, we additionally perform a dense runtime–quality sweep on one representative linear task (random inpainting) and one nonlinear task (nonlinear deblurring). We vary both the DAPS host scale and $M \in \{0, 20, 25, 50, 75, 100\}$, the number of final native Langevin updates replaced by IIL; $M = 0$ is native DAPS. Figure 14 reports the resulting runtime–quality curves, together with DPS-BoK scaling based on increasing numbers of independent complete DPS trajectories. Tables 5 and 6 then report the two compute controls across all eight tasks.

K.8 Directional curvature on learned data geometries.

A natural question is whether the direction-specific correction $\widehat{\mathrm{II}}(v, v)$ is necessary on learned data geometries, or whether the estimated curvature is approximately the same across tangent directions. We select 30 samples from each of H36M, MNIST, and FFHQ. For each sample x, we evaluate the learned geometry at $p = \widehat{\pi}(x)$ and, keeping p fixed, probe $m = 32$ mutually orthogonal tangent directions $\{u_i\}_{i=1}^m$. We evaluate the normalized vector-valued curvature

$$C_i = \frac{\widehat{\mathrm{II}}_p(u_i, u_i)}{\|u_i\|_2^2}. \tag{181}$$

We separately measure how its magnitude and direction change with the tangent direction.

For magnitude variation, we use

$$CV_{\mathrm{norm}} = \frac{\mathrm{std}_i \, \|C_i\|_2}{\mathrm{mean}_i \, \|C_i\|_2 + \epsilon}. \tag{182}$$

For visualization, we additionally define the relative directional curvature magnitude

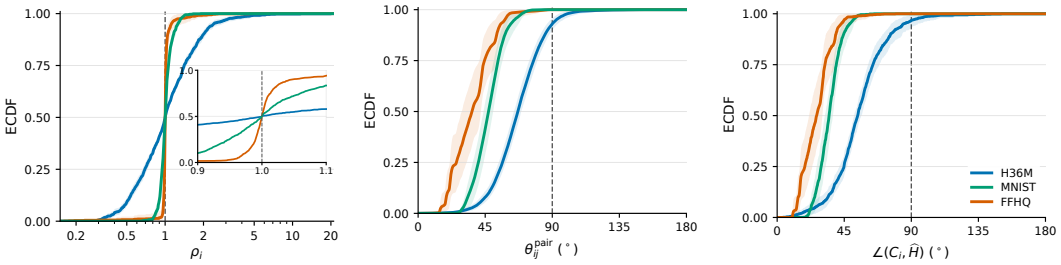
$$\rho_i = \frac{\|C_i\|_2}{\mathrm{median}_j \, \|C_j\|_2}. \tag{183}$$

Table 6: **Stronger-DAPS compute control on FFHQ-256.** DAPS+IIL is compared with a higher-compute native DAPS host at similar measured wall-clock cost. Nonlinear-task rows retain their frozen multi-seed stress-test protocol. Bold denotes the better quality within each task and metric.

Task	Method	Runtime (s)	PSNR ↑	SSIM ↑	LPIPS ↓
Linear tasks					
Random	DAPS-2K	236.83	30.698	.8020	.1553
	DAPS-1K+IIL_{25}	210.35	**32.405**	**.8913**	**.1014**
Box	DAPS-2K	236.64	24.385	.7487	.1749
	DAPS-1K+IIL_{25}	210.50	**24.787**	**.7736**	**.1619**
SR $\times 4$	DAPS-2K	251.18	29.264	.7815	.1915
	DAPS-1K+IIL_{25}	222.03	**30.181**	**.8495**	**.1599**
Gaussian	DAPS-2K	677.66	29.509	.7855	.1747
	DAPS-1K+IIL_{25}	560.87	**30.412**	**.8509**	**.1431**
Motion	DAPS-2K	677.47	31.584	.8377	.1337
	DAPS-1K+IIL_{25}	560.74	**32.826**	**.8977**	**.0978**
Nonlinear tasks					
Nonlinear	DAPS-5K	1897.31	28.420	.7941	.1812
	DAPS-4K+IIL_{100}	1867.27	**28.551**	**.8163**	**.1681**
Phase	DAPS-7K	4478.76	27.340	.7413	.2267
	DAPS-4K+IIL_{100}	4347.70	**27.730**	**.7874**	**.2207**
HDR	DAPS-7K	3903.32	27.793	.8439	.1563
	DAPS-4K+IIL_{100}	3993.11	**27.850**	**.8756**	**.1364**

For directional variation, we directly compare curvature vectors obtained from different tangent directions through

$$\theta_{ij} = \arccos \frac{\langle C_i, C_j \rangle}{\|C_i\|_2 \|C_j\|_2 + \epsilon}, \qquad i < j. \tag{184}$$



(a) Relative magnitude ρ_i (b) Pairwise angle θ_{ij} (c) Angle to $\widehat{H}$

Figure 15: Directional variation of the learned second fundamental form around held-out data samples. **(a)** Relative directional curvature magnitude ρ_i. Concentration near one indicates little variation in curvature magnitude across tangent directions. **(b)** Pairwise angles between curvature vectors obtained from different tangent directions at the same sample. **(c)** Angular deviation of each directional curvature vector from the estimated direction-averaged curvature $\widehat{H}$. Each dataset contains 30 held-out samples and 32 tangent directions per sample.

The median within-sample pairwise angles are $68.1°$ on H36M, $47.4°$ on MNIST, and $36.4°$ on FFHQ, while the corresponding magnitude variations are $CV_{\text{norm}} = 0.542$, 0.104, and 0.018, respectively. Thus, on H36M both curvature magnitude and direction vary substantially. MNIST also exhibits substantial directional variation with relatively modest magnitude variation. On FFHQ, the curvature-vector magnitudes are nearly unchanged across directions, while curvature vectors from different tangent directions are still separated by a median angle of $36.4°$, indicating that the variation is predominantly in normal-space direction.

We also compare each C_i with an estimate of the direction-averaged curvature at the same sample,

$$\widehat{H} = \frac{1}{K} \sum_{k=1}^{K} \widehat{\mathrm{II}}_p(\widehat{P} z_k, \widehat{P} z_k), \qquad z_k \sim \mathcal{N}(0, I), \quad K = 128. \tag{185}$$

Table 7: Directional curvature around selected data samples. Entries are medians across 30 samples. A_{pair} is the within-sample median pairwise angle between direction-specific curvature vectors, CV_{norm} measures variation in their magnitudes, and A_H is the median angular deviation from the estimated direction-averaged curvature $\widehat{H}$.

Dataset	A_{pair}	CV_{norm}	A_H
H36M	68.1°	0.542	54.4°
MNIST	47.4°	0.104	34.7°
FFHQ	36.4°	0.018	26.5°

The median angular deviations from $\widehat{H}$ are 54.4°, 34.7°, and 26.5° for H36M, MNIST, and FFHQ, respectively. Hence individual direction-specific curvature vectors can also differ substantially from the direction-averaged curvature. On a fixed subset of samples, increasing the trace estimate from $K = 128$ to $K = 256$ changes its direction by only 3.84°, 2.67°, and 1.42°, respectively, much less than the observed directional deviations.

These results show that, within the learned geometries used by IIL, curvature is not well described as direction independent. Changing the tangent direction can substantially change the full vector-valued second fundamental form, either in magnitude, normal-space direction, or both. This provides empirical support for retaining the direction-specific correction $\widehat{\mathrm{II}}(v, v)$ rather than replacing it by a single direction-averaged curvature vector. These diagnostics characterize our learned approximation of the geometry, not the exact curvature of the unknown true data manifolds.