DIFFUSION PROCESSES ON IMPLICIT MANIFOLDS

Anonymous authors
Paper under double-blind review

ABSTRACT

Simulating diffusion processes on a manifold typically requires access to tangent spaces and projection operators, which are usually unavailable when the manifold is known only through samples. We introduce Implicit Manifold-valued Diffusions (IMDs): a data-driven framework for constructing continuous stochastic dynamics on such implicit manifolds. Rather than reconstructing the geometry explicitly, IMDs estimate a manifold diffusion operator from the point cloud and use it to define the drift and diffusion coefficients of an ambient-valued SDE. The point cloud therefore serves only to estimate the diffusion operator. The induced process evolves continuously in the ambient space and becomes manifold-valued in the large-sample limit. We prove that strong-resolvent convergence of the estimated operator is sufficient for weak convergence of the induced stochastic process on path space. As a concrete realization, we construct a family of graph-based estimators whose induced IMDs range from density-biased Langevin dynamics to manifold Brownian motion. Experiments validate the predicted path space convergence and demonstrate geodesic recovery, density-debiased manifold exploration, and stochastic interpolation on image data.

1 INTRODUCTION

Stochastic differential equations (SDEs) model continuous-time dynamics under deterministic drift and random perturbations [57]. In machine learning, they form the common mathematical language describing many important methods, such as Langevin sampling [8; 55], Bayesian inference [76; 51], stochastic optimization [1; 10], and score-based generative modeling [64; 24]. While classical SDE methods are typically formulated in Euclidean space, the manifold hypothesis posits that high-dimensional data concentrate near a lower-dimensional manifold $\mathcal{M}^d \subset \mathbb{R}^D$ with $d \ll D$ [29; 48]. If the relevant state space is therefore $\mathcal{M}$ rather than the full ambient space, stochastic dynamics should respect its geometry, motivating manifold-valued extensions of standard Euclidean SDEs.

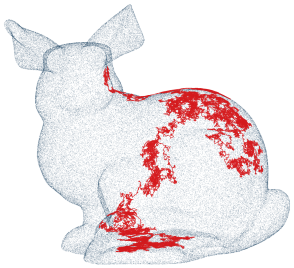


Figure 1: Our method allows the computation of a general family of continuous stochastic process (red) constrained to a compact Riemannian submanifold defined through a sampled point cloud.

Mathematically, the theory of SDEs on manifolds is a well-developed field [39; 2]. Associated numerical methods adapt Euclidean integration schemes to account for geometric primitives using tools from differential geometry [36], and form the basis for current SDE methods on data manifolds, such as sampling [17; 18; 8] or generative modeling [25; 78; 16].

In this work, we consider the *implicit manifold* $\mathcal{M}$, which is only available through a set of samples $X_N = \{x_i\}_{i=1}^N$ drawn from a data distribution q_{data} that is supported on $\mathcal{M}$ [44]. This creates a basic mismatch: manifold SDE methods require geometric primitives that are not available under the implicit manifold setting, raising the following question:

> *Can we construct continuous stochastic dynamics on $\mathcal{M}$ directly from samples, without first reconstructing its geometry or learning a global parametrization?*

We answer this question through an operator-based construction of diffusion processes. Indeed, it is well-known that diffusions may be characterized by a second order differential operator $\mathcal{L}$ called

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.

the *diffusion operator* or *infinitesimal generator* [2; 28; 39]. We briefly introduce this formalism in Section 3, and present a strictly $\mathcal{L}$-dependent construction of $\mathbb{R}^D$-valued diffusion processes that remain on $\mathcal{M}$ in Section 4.1. Section 4.2 then leverages the $\mathcal{L}$-dependent formulation, defining *Implicit Manifold-valued Diffusions* (IMDs) as the same type of process, only with a data-driven diffusion operator $\widehat{\mathcal{L}}$. Figure 1 shows a visual example of an IMD-simulated diffusion trajectory.

We further define a "suitable" estimator $\widehat{\mathcal{L}}$ in Section 5 by establishing a sufficient condition on $\widehat{\mathcal{L}}$, leading to convergence in the IMD's path measure. That is, the entire data-driven stochastic path $(X_t^N)_t$ converges to the true path $(X_t)_t$ generated by $\mathcal{L}$. This is a stronger guarantee than matching the marginal distribution which only tracks current position, as path-consistency also matches how the process moves between states. In Section 6 we propose a concrete realization of $\widehat{\mathcal{L}}$ based on graph Laplacians, for which we verify the above operator condition directly. Finally, Section 7 presents numerical experiments supporting our theoretical claims, thereby demonstrating that IMDs can infer geodesic distances, preserve both local and global information of manifold diffusion, and stochastically interpolate on the implicit manifold of real-world data.

To our knowledge, IMDs are the first framework to compute ambient-valued stochastic dynamics on $\mathcal{M}$ in settings where geometric priors are unavailable. We summarize our contributions below.

- We introduce IMDs as an operator-based, data-driven method to simulate diffusion processes on implicit manifolds.
- We provide sufficient conditions on the estimated operator to guarantee convergence of the process as a whole to the intended target process on $\mathcal{M}$.
- We instantiate IMDs using Diffusion Maps-type operators, provably satisfying the aforementioned sufficient conditions and leading to a family of Langevin IMDs on $\mathcal{M}$.
- A suite of numerical experiments on known manifolds validate the proposed theory. Further tests of IMDs on image data reveal that their understanding of geometry and exploration capabilities.

2 Related works

We briefly overview related literature, focusing on geometric SDEs, aspects of non-parametric manifold learning and recent data-driven approaches.

Manifold SDEs Simulating SDEs on known manifolds is well-understood, and is typically made possible through explicit use of the exponential map and retractions [21; 36; 61]. A similar strategy involves constructing a Markov chain through Geodesic Random Walks (GRWs), whose limit is a smooth diffusion on $\mathcal{M}$ [43; 66]. Both approaches have led to important geometry-based algorithms, such as Riemannian Langevin HMC [33], Langevin dynamics [58; 8] and MCMC [17; 18]. Recently, such techniques have been adapted to diffusion- and flow-based generative modeling settings [25; 53; 24; 16]. All the above still rely on geometric priors, and therefore do not apply to implicit $\mathcal{M}$.

Manifolds from Graphs A popular non-parametric method to estimate the implicit manifold's structure is via a proximity graph constructed from the samples. Pioneered by Belkin & Niyogi [5; 6; 7], such graph-based methods are able to access manifold geometry through continuum limits of Laplacian-type operators [37; 67; 31; 68; 32], which are tightly linked to manifold diffusion processes [2]. More recent work led by Jones [41; 42] leverages Diffusion Maps [22] to recover the carré-du-champ (CdC) from the data graph's diffusion operator, acting as a proxy for $\mathcal{M}$'s metric and therefore tangent spaces. In a similar spirit, a graph-based construction of tangent spaces was proposed in the context of optimization in [73]. IMDs exploit the convergence of such estimators to construct ambient drift and diffusion fields inducing a process confined to $\mathcal{M}$.

Data-driven Geometry Jones' CdC has recently been implemented as a tool for geometric flow matching and modeling of distributions [4; 3]. In the context of Riemannian optimization, Kharitenko et al. [44] concurrently showed that a pre-trained score model and its Jacobian asymptotically recover projection operators onto $\mathcal{M}$ and its tangent bundle by leveraging well-established connections between diffusion models and manifold data [60; 11; 15; 13; 49]. Our work will rather focus on constructing stochastic dynamics on $\mathcal{M}$ through estimated diffusion operators, with a formulation that leaves the door open to neural surrogates.

3 PRELIMINARIES: STOCHASTIC DYNAMICS ON EMBEDDED MANIFOLDS

We briefly recall the fundamentals of Euclidean diffusion processes in $\mathbb{R}^D$, as well as on embedded Riemannian manifolds. We refer readers to standard texts [28; 57; 36; 39] for more thorough treatments of the topic.

Notation We denote the space of probability distributions on a space $\mathcal{X}$ as $\mathcal{P}(\mathcal{X})$. Brownian motion on a manifold $\mathcal{M}$ is denoted $B_t^{\mathcal{M}}$. Estimators are always denoted with a $\widehat{\cdot}$ symbol, and continuous-time diffusion processes are denoted $(X_t)_t$, where $t \geq 0$ is the time variable.

Euclidean Diffusions We begin with a standard diffusion process $(X_t)_t$ in $\mathbb{R}^D$, described by the Itô stochastic differential equation (SDE)

$$dX_t = b(X_t)dt + \sigma(X_t)dB_t, \tag{1}$$

where $b : \mathbb{R}^D \to \mathbb{R}^D$ is the drift, $\sigma : \mathbb{R}^D \to \mathbb{R}^{D\times D}$ is the diffusion coefficient, and $(B_t)_t$ is standard $\mathbb{R}^D$-valued Brownian motion. A standard numerical approximation of Eq. (1) is the *Euler-Maruyama* (EM) integration scheme with timestep $h > 0$, given by

$$X_{k+1}^h = X_k^h + hb(X_k^h) + \sqrt{h}\sigma(X_k^h)\xi_k, \qquad \xi_k \sim \mathcal{N}(0, \mathbf{I}_D). \tag{2}$$

Embedded Manifold We now consider the case where the state space is a smooth d-dimensional Riemannian manifold $\mathcal{M}$ embedded in $\mathbb{R}^D$ through a given embedding $\iota : \mathcal{M} \to \mathbb{R}^D$. We identify $\mathcal{M}$ with its embedded image $\iota(\mathcal{M})$ for notational simplicity, and endow it with the Riemannian metric induced by the ambient Euclidean space. At each $x \in \mathcal{M}$, infinitesimal directions lie in the tangent space $T_x\mathcal{M}$, and a drift term on $\mathcal{M}$ is a now vector field $b(x) \in T_x\mathcal{M}$.

For a smooth function $f : \mathcal{M} \to \mathbb{R}$, the Riemannian gradient $\nabla_{\mathcal{M}} f(x) \in T_x\mathcal{M}$ is defined by the unique tangent vector obtained by projecting the gradient of an ambient extension $\tilde{f}$ of f onto $T_x\mathcal{M}$. We further denote the projection matrix onto the tangent space of $x \in \mathcal{M}$ as $P_{T_x\mathcal{M}}$. The corresponding Laplace-Beltrami operator (LBO) is

$$\Delta_{\mathcal{M}} f := \operatorname{div}_{\mathcal{M}}\big(\nabla_{\mathcal{M}} f\big), \tag{3}$$

and plays the role of the Euclidean Laplacian on $\mathcal{M}$ [27].

Geometric SDE Integrator When the geometry of $\mathcal{M}$ is explicitly available, the EM update (2) admits a natural geometric analogue: Geometric Euler-Maruyama (GEM). Given $X_k^h \in \mathcal{M}$, one first constructs a stochastic increment in the tangent space $T_{X_k^h}\mathcal{M}$ and then maps it back to the manifold through the exponential map $\exp_x : T_x\mathcal{M} \to \mathcal{M}$:

$$X_{k+1}^{\mathrm{GEM}} = \exp_{X_k^h}\left(h\, b(X_k^h) + \sqrt{h}\, \sigma(X_k^h)\xi_k^{\parallel}\right), \tag{4}$$

where $\xi_k^{\parallel} = P_{T_{X_k^h}}\xi_k$ is a Gaussian vector in $T_{X_k^h}\mathcal{M}$. GEM therefore requires explicit access to local tangent geometry and to the exponential map [80], which are generally unavailable under the implicit $\mathcal{M}$ setting discussed in the introduction.

4 IMPLICIT MANIFOLD-VALUED DIFFUSIONS

We now present our general framework to convert samples $x_1, \dots x_N \overset{\text{i.i.d.}}{\sim} q \in \mathcal{P}(\mathcal{M})$ to continuous-time stochastic dynamics on $\mathcal{M}$, dubbed *Implicit Manifold-valued Diffusions* (IMDs). We introduce the diffusion operator formalism along with notation and interpretation in Section 4.1. We then define IMDs as such a process whose the drift and diffusion terms are estimated from the data.

4.1 THE OPERATOR REPRESENTATION OF DIFFUSIONS

To shift to an operator-theoretic perspective, we consider a smooth observable of the SDE (1) $f{:}X_t \mapsto f(X_t)$. Its average change of rate along trajectories can be expressed as

$$\mathcal{L}f(X_t) = \lim_{t\to 0} \frac{1}{t}(\mathbb{E}\left[f(X_t)|X_0 = x\right] - f), \tag{5}$$

and corresponds to the backward Kolmogorov equation [2]. This construction is general and applies to both Euclidean and manifold diffusions [28]. Itô's lemma resolves this limit into a second-order differential operator identifying the explicit drift and diffusion terms, given by

$$\mathcal{L}f = \langle b, \nabla_{\mathcal{M}} f \rangle + \frac{1}{2}\Delta_{\mathcal{M}} f, \tag{6}$$

making use of the previously introduced Riemannian gradient and LBO (3). The *carré du champ* (CdC) associated to any such $\mathcal{L}$ captures the second-order (diffusive) terms as

$$\Gamma(f,g) := \frac{1}{2}\left[\Delta_{\mathcal{M}}(fg) - f\Delta_{\mathcal{M}} g - g\Delta_{\mathcal{M}} f\right]. \tag{7}$$

It is classical that when applied to embedding coordinates $\iota = (\iota^1, \ldots, \iota^D)$ of a Riemannian manifold $\mathcal{M}$, the LBO at $x \in \mathcal{M}$ gives the *mean curvature vector* $(\Delta_{\mathcal{M}}\iota)(x) = \vec{\nu}_x$ which is normal to $\mathcal{M}$ at x, while the CdC at $x \in \mathcal{M}$ is the tangent projection matrix $\Gamma(\iota,\iota)(x) = P_{T_x\mathcal{M}}$ [63; 3]. For a simple example, the process $dX_t = \frac{1}{2}\vec{\nu}_{X_t}dt + P_{T_{X_t}\mathcal{M}}B_t$ is Itô Brownian motion on $\mathcal{M}$ [2; 39], where $P_{T_{X_t}\mathcal{M}}B_t =: B_t^{\parallel}$ is a tangent diffusion step, corrected by the normal drift $\frac{1}{2}\vec{\nu}_{X_t}$. We therefore see that $\mathcal{L}$ from Eq. (6) with $b = 0$ is $\frac{1}{2}\Delta_{\mathcal{M}}$ and indeed generates manifold Brownian motion $B_t^{\mathcal{M}}$.

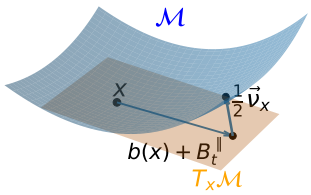


Figure 2: Intuitive decomposition of the operator-based manifold diffusion from Eq. (8). The LBO contribution corrects the tangent steps taken by b and $B_t^{\parallel}$.

Adding nonzero drift $b \neq 0$ yields a general diffusion process on $\mathcal{M}$ that is realizable by the $\mathbb{R}^D$-valued SDE

$$dX_t = \underbrace{b(X_t)dt + \frac{1}{2}\vec{\nu}_{X_t}dt}_{=(\mathcal{L}\iota)(X_t)dt} + \underbrace{dB_t^{\parallel}}_{=\Gamma(\iota,\iota)(X_t)dB_t}, \tag{8}$$

where $\mathcal{L}$ and Γ are as in Eqs. (6),(7), visually depicted in Figure 2. The following shows that our alternative, operator-based SDE (8) is equivalent to the GEM scheme (4) introduced in Section 3.

Proposition 1. *If the drift b is Lipschitz, the stochastic dynamics induced by the EM-discretization of Eq. (8) locally match those of Geometric Euler-Maruyama (4).*

More details on the construction of Eq. (8) are available in Appendix A.1, as well as a proof of Proposition 1 in Appendix A.2.

4.2 Data-driven diffusions

We now define an IMD $(X_t^N)_t$ by replacing the diffusion operator $\mathcal{L}$ of Section 4.1 with an operator $\widehat{\mathcal{L}}_N \in \mathbb{R}^{N\times N}$ estimated from the samples $\{x_i\}_{i=1}^N$, where $x_i \sim q \in \mathcal{P}(\mathcal{M})$. Let

$$\widehat{b}_N := \widehat{\mathcal{L}}_N \iota \in \mathbb{R}^D, \qquad \widehat{A}_N := \widehat{\Gamma}_N(\iota,\iota) \in \mathbb{R}^{D\times D}, \tag{9}$$

where $\widehat{\Gamma}_N$ denotes the CdC associated with $\widehat{\mathcal{L}}_N$ (7). We define the corresponding IMD as the $\mathbb{R}^D$-valued stochastic process

$$dX_t^N = \widehat{b}_N(X_t^N)dt + \widehat{A}_N^{1/2}(X_t^N)dB_t. \tag{10}$$

Equation (10) provides the central "lifting" mechanism of IMDs. The samples are used to estimate a diffusion operator, whose action on the embedding coordinates determines continuous drift and diffusion fields in the ambient space, generating a continuous stochastic trajectory in $\mathbb{R}^D$. Thus, the point cloud itself is not treated as the state space of the dynamics; rather, but as a discrete representation from which the continuum dynamics on the underlying manifold are inferred. Table 1 summarizes this distinction from classical geometric stochastic integrators and recent data-driven geometric methods.

A key consequence of this operator viewpoint is that IMDs are agnostic to how $\widehat{\mathcal{L}}_N$ is obtained. We will show in Section 5 that it is the consistency of $\widehat{\mathcal{L}}_N$ with respect to $\mathcal{L}$ that matters in order to ensure correctness of an induced IMD path.

Method	Geometry access			Nature
	No $T_x\mathcal{M}$	No $\exp_x$	Data-driven	
GEM, GRWs [36; 80; 66]	✗	✗	✗	Stochastic
DRGD / DLF [44], RMM [3]	✓	✓	✓	Deterministic
IMDs (ours)	✓	✓	✓	Stochastic

Table 1: IMDs combine data-driven methods with stochastic geometric analysis, complementing the current literature of geometric modelling [3] and optimization on implicit manifolds [44].

5 Operator consistency: a sufficient condition

Consistency of paths As mentioned in the introduction, the goal of this work is to produce correct stochastic processes on an implicit data manifold $\mathcal{M}$. We note that this is different from most current generative modeling or sampling methods, which aim to match the time-t marginal distribution $\mu_t := \text{Law}(X_t)$ of an intended stochastic process $(X_t)_t$ [38; 64; 18]. Rather, we rather aim for *dynamical consistency*, in that the produced $(X_t^N)_t$ matches the intended $(X_t)_t$ as a process on $[0, T]$. Such process-level matching can be measured through convergence of the probability measure on the *path space*: the space of continuous functions $F : [0, T] \to \mathbb{R}^D$, denoted $C([0, T], \mathbb{R}^D)$. That is, we consider the entire $\mathbb{R}^D$-valued stochastic process $(X_t^N)_t$ over a time interval $[0, T]$ as a random variable and its associated probability distribution $\mathbb{P}_N \in \mathcal{P}(C([0, T], \mathbb{R}^D))$ on the space of paths [9], and ask whether $\mathbb{P}_N$ converges in some sense to the intended process' path law $\mathbb{P}$.

Faithful path computations are highly relevant in tasks where one does not simply care about endpoints, but also intermediate or transition modes. For example, in molecular dynamics, reproducing the distribution of molecular conformations at a given time does not in general imply that the underlying dynamics need be correct. In applications such as protein folding, the *way* it folds is a quantity of high scientific interest [14; 79], but many current diffusion-based methods only focus on black-box generation of conformations [75; 77; 78]. Ensuring a correct stochastic path on $\mathcal{M}$ therefore ensures a notion of plausibility of each point along $(X_t^N)_t$. This enables computations of transition pathways, residence times, or temporal correlations, all of which are inaccessible through marginal matching. Path space convergence therefore provides a stronger notion of consistency than matching marginals, thereby being our desired notion of convergence. We now introduce a sufficient condition on the IMD's generator $\widehat{\mathcal{L}}_N$ to ensure weak convergence of path measures $\mathbb{P}_N \Rightarrow \mathbb{P}$.

Sufficient Condition To ensure path space convergence, we need a condition on the estimated operator $\widehat{\mathcal{L}}_N$ that is strong enough to control the process it generates, rather than only its pointwise action on individual functions. We use *strong-resolvent convergence* for this purpose, which informally means that an indexed sequence of operators $(A_r)_r$ tends to a limiting operator A along $r \to \infty$

$$(\lambda \mathbf{I} - A_r)^{-1} f \overset{r\to\infty}{\to} (\lambda \mathbf{I} - A)^{-1} f \tag{11}$$

for every suitable f. Under standard assumptions, Eq. (11) provides the operator-level route to convergence of the induced stochastic dynamics [46; 47].

Theorem 1. *Let $\mathcal{M}$ be a smooth, compact, connected, orientable and boundaryless d-dimensional Riemannian manifold, embedded in $\mathbb{R}^D$ via ι, and $T > 0$ be a time horizon. Furthermore, consider $\widehat{\mathcal{L}}_N$ a data-driven estimator of a diffusion operator $\mathcal{L}$ on $\mathcal{M}$, such that $\widehat{\mathcal{L}}_N$ converges in the strong-resolvent sense to $\mathcal{L}$ at $N \to \infty$. Then, the associated path space law $\mathbb{P}_N$ of the IMD constructed from $\widehat{\mathcal{L}}_N$ via Eq. (10) weakly converges to the true path law $\mathbb{P}$ on $C([0, T], \mathbb{R}^D)$.*

The above theorem is a consistency result. Informally, as the number of samples increases, the point cloud X_N resembles more and more $\mathcal{M}$, thereby inducing $\widehat{\mathcal{L}}_N \to \mathcal{L}$. We provide a proof in Appendix C, and empirical corroboration in Section 7.2. Therefore, any IMD $(X_t^N)_t$ whose data-driven generator $\widehat{\mathcal{L}}_N$ satisfies the condition in Theorem 1 automatically enjoys weak convergence of path measures to the intended diffusion $(X_t)_t$.

6 A FAMILY OF DIFFUSIONS ON IMPLICIT $\mathcal{M}$

We now instantiate the general IMD framework with a concrete family of data-driven diffusion operators $\widehat{\mathcal{L}}_N^{(\alpha)}$ that are closely related to the Diffusion Maps operators [22].[1] The target diffusion operator $\mathcal{L}^{(\alpha)}$ is parametrized by $\alpha \in [0,1]$, which adjusts the associated diffusion's invariant law between q^2 and the uniform density on $\mathcal{M}$, effectively capable of accentuating or attenuating density bias on $\mathcal{M}$. Such a construction finds particular use when interested in sampling from low-probability regions, or simply exploring the surface of $\mathcal{M}$ [26].

6.1 GRAPH-BASED CONSTRUCTION OF DIFFUSION OPERATORS

The estimator $\widehat{\mathcal{L}}_N^{(\alpha)}$ is based on random walks on graphs. We first build a proximity graph $\mathcal{G}_N := (X_N, E)$ whose node set is the point cloud X_N, and edges $(x_i, x_j) \in E$ are determined via a fixed-bandwidth cutoff kernel $\eta(x_i, x_j) = \mathbf{1}\{\|x_i - x_j\|_2 \leq \epsilon\}$. This specifies the entries of a symmetric unweighted adjacency matrix $W_{ij} = \eta(x_i, x_j)$ with per-node degree $m_i := \sum_{j=1}^N W_{ij}$.

For a chosen $\alpha \in [0,1]$, rescaling the effective weight $d_i^{(\alpha)} := \sum_j \frac{W_{ij}}{m_i^\alpha m_j^\alpha}$ induces an α-biased random walk graph Laplacian [20; 32],

$$(\widehat{\mathcal{L}}_N^{(\alpha)} u)(x_i) := \frac{c_\epsilon}{d_i^{(\alpha)}} \sum_{i \sim j} \frac{W_{ij}}{m_i^\alpha m_j^\alpha} \left(u(x_j) - u(x_i)\right), \tag{12}$$

where c_ϵ is a scaling constant that depends on the choice of the kernel, and $u : X_N \to \mathbb{R}$ is a function over the nodes. Further, we extend the action of $\widehat{\mathcal{L}}_N^{(\alpha)}$ from functions u defined on the nodes of $\mathcal{G}_N$, to the full embedding space $\mathbb{R}^D$ by computing a weighted k-nearest neighbours average $\frac{1}{k}\sum_{i=1}^k w_i u(x_i)$. We refer interested readers to Appendix B for further details on this construction.

Before we turn to showing that $\widehat{\mathcal{L}}_N^{(\alpha)}$ satisfies the conditions of Theorem 1, we emphasize that it is only one (among many) possible instantiations of the IMD drift and diffusion terms in Eq. (9). We use graph Laplacians because they are a well-studied class of manifold operator estimators with established convergence guarantees; see Section 2.

6.2 ASYMPTOTIC CONVERGENCE

We now show that for $\alpha \in [0,1]$, Eq. (12) satisfies the consistency conditions of Theorem 1, therefore implying asymptotic path-wise correctness of the induced IMD.

Proposition 2. *Let $\{x_i\}_{i=1}^N$ be a point cloud sampled from $\mathcal{M}$ with $x_i \overset{\text{i.i.d.}}{\sim} q \in \mathcal{P}(\mathcal{M})$. Then, the estimator $\widehat{\mathcal{L}}_N^{(\alpha)}$ constructed via Eq. (12) converges in the strong-resolvent sense (11) as $N \to \infty, \epsilon_N \to 0$ with rates specified in Assumption 1, to the manifold-valued operator*

$$\mathcal{L}_\alpha = (1-\alpha)\nabla_{\mathcal{M}} \log q \cdot \nabla_{\mathcal{M}} + \frac{1}{2}\Delta_{\mathcal{M}}. \tag{13}$$

The α parameter acts as a scale for the strength of density correction and directly influences the produced diffusion operator $\mathcal{L}^{(\alpha)}$. The corresponding induced IMDs are then interpretable as Langevin diffusion processes driven by a potential $(1-\alpha)V$ where $q \propto e^{-V}$, affecting the invariant law [59; 62]. See Table 2 for the interesting cases of $\alpha = 0, \frac{1}{2}, 1$.

This flexibility is especially interesting for manifold exploration, where the empirical sampling density may not coincide with the distribution one wishes to explore. In particular, choosing $\alpha = 1$ corresponds to simulating Brownian motion constrained to the manifold. This allows for uniform exploration of $\mathcal{M}$ despite only having access to non-uniformly distributed samples. We demonstrate this debiasing effect empirically in Section 7.3.

[1] For reasons detailed in Appendix B, we use a hard cutoff kernel rather than a typical Gaussian kernel, although a Gaussian kernel is surmised to also satisfy our required criteria for convergence [32].

Parameter	Limiting Operator	Invariant Law	Process Type
$\alpha = 0$	$-\langle \nabla_{\mathcal{M}} V, \nabla_{\mathcal{M}} \rangle + \frac{1}{2}\Delta_{\mathcal{M}}$	q^2	Langevin
$\alpha = \frac{1}{2}$	$-\frac{1}{2}\langle \nabla_{\mathcal{M}} V, \nabla_{\mathcal{M}} \rangle + \frac{1}{2}\Delta_{\mathcal{M}}$	q	Langevin
$\alpha = 1$	$\frac{1}{2}\Delta_{\mathcal{M}}$	uniform	Brownian motion

Table 2: Diffusion generators and associated invariant laws for $q \propto e^{-V}$.

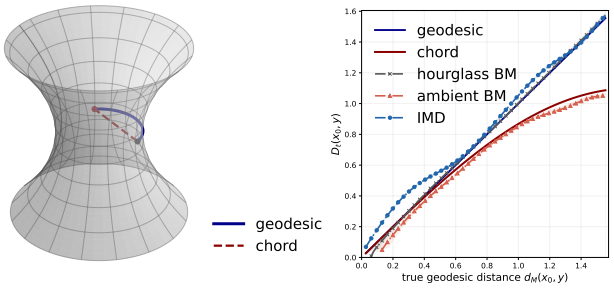
7 Numerical Experiments

In this section we present experimental results tested on numerical solutions of IMDs (88), validating all aspects of the developed theory. We begin by showing in Section 7.1 that IMDs can recover geodesic distances on an implicit $\mathcal{M}$. In Section 7.2, we support the path space convergence statement from Theorem 1 on a non-trivial geometry. In Section 7.3, we show that the constructed α-family of diffusions from Proposition 2 indeed enable flexible invariant laws, opening the doors to on-manifold exploration. Finally, in Section 7.4, we use a real-world image dataset to show that from geometry alone, IMDs are able to stochastically explore and interpolate data, while maintaining semantic meaning.

7.1 Geodesic estimation

We demonstrate that IMDs are able to estimate geodesic distances $d_{\mathcal{M}}(x, y)$ on the implicit $\mathcal{M}$ through Varadhan's formula (91), which relates the short-time transition law of $B_t^{\mathcal{M}}$ (heat kernel) to geometry [23; 54; 70; 34]. Results are shown in Figure 3, and details are deferred to Appendix F.2.

Figure 3: The empirical short-time transition law $\hat{p}_t(x, y)$ obtained through IMD Brownian motion ran on the waist of the displayed hourglass successfully recovers the intrinsic geodesic $d_{\mathcal{M}}(x, y)$. Ambient Gaussian noise, however, expectedly recovers the chord distance.

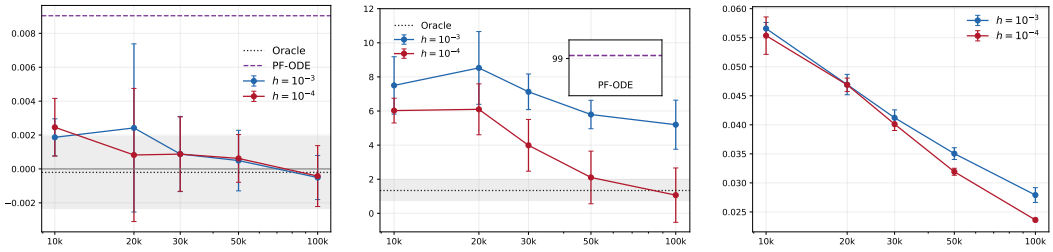


7.2 Path space convergence on non-trivial geometry

In this experiment we aim to evaluate the path convergence of IMDs on a point cloud sampled from a re-meshed Stanford Bunny $\mathcal{B}$ (Figure 1) [69]. We aim to show that both local and global aspects of path space are recovered by IMDs, and that off-manifold errors disappear in the $N \to \infty$ limit, supporting the claim of Theorem 1.

As a reference process, we consider approximate Brownian motion on $\mathcal{B}$ using an isotropic random walk, with steps traced by unfolding and reflection at boundary edges. This construction is motivated by geodesic random-walk diffusion limits [43; 72]. We measure the global discrepancy between path-level distributions via the maximum mean discrepancy (MMD) (93), and local fluctuations through quadratic variation (QV) (95). Finally, we report distance from its closest point to the mesh as a measure of geometric error. We refer the reader to Appendix F.3 for further details.

Results. We report all results in Figure 4, demonstrating that all metrics behave as expected as a function of the point cloud's size $N \in \{10k, 20k, 30k, 50k, 100k\}$ averaged over 5 seeds. We especially observe that the MMD is insensitive to discretization step size h, and that the QV matches oracle's within acceptable error range for sufficiently small h while the PF-ODE [50; 64] expectedly fails to retain any temporal correlations. Off-manifold errors also decay to zero as $N \to \infty$.



(a) Squared Path MMD (b) Quadratic Variation Error (%) (c) Off-manifold Distance

Figure 4: Comparison of results for IMDs, oracle $B_t^{\mathcal{B}}$, and a PF-ODE across sample sizes N.

7.3 Debiasing for manifold exploration

Recent advances in exploration of data manifolds [26] has motivated uniform sampling of $\mathcal{M}$, despite the accessible data being drawn from a non-uniform $q \in \mathcal{P}(\mathcal{M})$. The α=1 diffusion proposed in Section 6.1 is therefore perfectly suited to this task, as it not only has the uniform as invariant law (Table 2), but also *continuously and dynamically* explores $\mathcal{M}$. We exemplify on the torus $\mathbb{T}^2 \subset \mathbb{R}^3$.

Using a point cloud drawn from a bimodal distribution $q_{\text{vMF}} \in \mathcal{P}(\mathbb{T}^2)$ [52], we construct $\widehat{\mathcal{L}}_N^{(\alpha)}$ as per Eq. (12) as the generator of a torus-valued IMD. The invariant law is then estimated by sampling the IMD path from independent trajectories after a predetermined burn-in time. For each α, we report the empirical invariant law's energy distance (ED) (97) to $q_{\text{vMF}}^{2-2\alpha}$. Uniform exploration is further measured through a uniform mesh of the torus, on which we compare total variation (TV) distance between the empirical distribution of samples in each cell c_j and the expected equal-area (99), as well as occupancy (coverage) of each cell (98). We defer further details to Appendix F.4.

Results. Table 3 shows that the empirical invariant law progressively shifts to $q_{\text{vMF}}^{2-2\alpha}$ across all tested α. Both equal-area TV and coverage imply that the IMD associated to α=1 is indeed reaching low-probability regions, further supported by the clear spread in Figure 5. Notably, while its empirical invariant law remains close in ED to q_{vMF}, the α=1 case is nevertheless closest to uniform among all tested α, supporting the expected debiasing effect. We attribute the remaining finite-sample discrepancy to both limited samples and the implicit KDE estimation $q_{\text{vMF}}^{\alpha}(x_i) \approx d_i^{(\alpha)}$ in the graph Laplacian construction from Section 6.1. Nonetheless, the α=1 IMD asymptotically admits the uniform distribution as its invariant law, thereby providing max-entropy exploration of $\mathcal{M}$ [26].

α	**Energy Distance** ↓			**Equal-Area TV** ↓	**Coverage** ↑
	q_{vMF}^2	q_{vMF}	uniform		
$\alpha = 0$	$\mathbf{0.007 \pm 0.008}$	0.065 ± 0.021	0.289 ± 0.004	0.760 ± 0.002	0.558 ± 0.012
$\alpha = \frac{1}{2}$	0.046 ± 0.005	$\mathbf{0.007 \pm 0.003}$	0.164 ± 0.003	0.619 ± 0.003	0.807 ± 0.015
$\alpha = 1$	0.194 ± 0.008	0.045 ± 0.004	$\mathbf{0.065 \pm 0.005}$	$\mathbf{0.426 \pm 0.002}$	$\mathbf{0.965 \pm 0.006}$

Table 3: For each α, the associated IMD's invariant law approaches the intended $q_{\text{vMF}}^{2-2\alpha}$. We also observe that for α=1, the IMD reaches almost all cells c_j of the torus' mesh.

7.4 COIL-20: IMD bridge interpolation

We now turn to a more realistic setting and demonstrate that IMDs can both capture and navigate the implicit manifold structure of image data [12]. We use the COIL-20 dataset [56], which contains 32×32 grayscale images of 20 different objects photographed at 5° angular increments. For a chosen target image, we will be interested in stochastically interpolating two viewing angles using IMDs, and assessing the produced samples' validity.

We learn a shared viewpoint transformation from 19 objects and use it to augment sparsely observed views of the held-out target image. We then construct an IMD in raw pixel space and evaluate

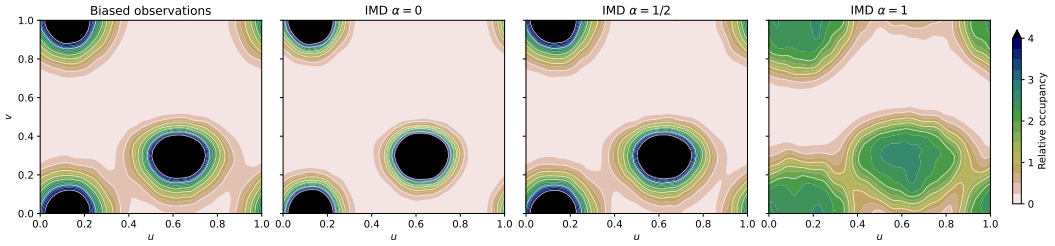


Figure 5: Contour map of invariant laws of IMDs from Table 3 along both axes of the torus. We observe the strong favoring of the modes for $\alpha = 0$, and increased exploratory behaviour for $\alpha = 1$.

whether its endpoint-conditioned trajectories recover withheld views more faithfully than deterministic linear interpolation. Reconstruction quality is measured using pixel-wise root mean squared error (RMSE) and the structural similarity index (SSIM), which compares local luminance, contrast, and image structure [74]. More details are deferred to Appendix F.5.

Results. Table 4 reports the mean and standard deviation of RMSE and SSIM scores averaged of 32 trajectories $\times$ 5 seeds. Figure 6 shows a single trajectory snippet between $0°$ and $45°$, demonstrating generalization-like behaviour by reconstructing the withheld angles $(5°, 15°, \dots)$ from the given geometric support.

Method	**RMSE** $\downarrow$	**SSIM** $\uparrow$
Linear	0.097	0.8113
IMD	$\mathbf{0.071 \pm 0.001}$	$\mathbf{0.879 \pm 0.004}$

Table 4: IMDs score higher than linear interpolations on similarity metrics.

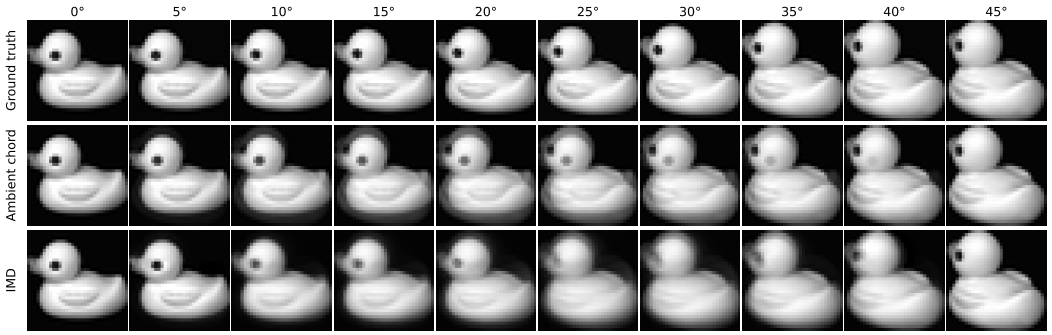


Figure 6: True (top) versus linear (middle) versus IMD interpolation (bottom) from $0°$ to $45°$.

8 Conclusion and Outlook

We proposed IMDs as an operator-based, data-driven framework to construct and simulate diffusion processes on implicit manifolds. IMDs replace explicit geometric reconstruction by operator estimation: an estimate of the diffusion operator determines, through its action on the embedding coordinates and its carré-du-champ, the drift and diffusion coefficients of an $\mathbb{R}^D$-valued SDE whose limiting state space is the data manifold. We showed that strong-resolvent convergence of the estimator is sufficient for consistency with the intended diffusion on path space, and instantiated the framework using graph-based operators, inducing a family of manifold Langevin IMDs. Numerical experiments support both the local and global dynamical predictions of the theory, and illustrated applications to geodesic estimation, manifold exploration, and stochastic interpolation.

Rather than targeting a particular task, we view IMDs as a general tool for constructing stochastic dynamics directly from sampled geometry. We hope this operator-to-SDE perspective can inspire new methods for geometrically faithful sampling, interpolation, exploration, data generation, and stochastic optimization in settings where the data's geometry is implicit. Beyond the presented graph-based instantiation, IMDs are made to accommodate different classes of estimators. Developing scalable neural surrogates to estimate $\mathcal{L}$ and Γ, along with non-asymptotic finite-data convergence rates, make for a natural and interesting direction of future work.

AI USE STATEMENT

In this work, the authors acknowledge the use of generative AI tools for setting up code, partial writing of the code, generation of synthetic datasets, aid in producing plots, and checking the writing/grammar of the article and consistency of the proposed proofs.

The authors have *not* used generative AI tools during the creative process of finding the research problem, ideation of IMDs as data-driven stochastic processes, or finding proof techniques. All AI-generated code was double-checked by the authors, which take full responsibility for the final content of this work.

ETHICS STATEMENT

The present work is a work of research on stochastic processes on geometries defined by data. Topics discussed include data generation. By no means do the authors intend nor condone this to be used maliciously, e.g. data retrieval or unlawful / unethical data generation of personal information.

REPRODUCIBILITY STATEMENT

In the spirit of open science, the authors intentionally made efforts to make this work as reproducible as possible by including extra derivations, intuition, and experimental details in the Appendix. Theoretical results are all clearly stated, along with their assumptions and references. Experimental results were fully described, and the code is provided in the Supplementary Material. The full repository will be publicly released upon acceptance of the manuscript.

REFERENCES

[1] P-A Absil, Robert Mahony, and Rodolphe Sepulchre. *Optimization algorithms on matrix manifolds*. Princeton University Press, 2008.

[2] Dominique Bakry, Ivan Gentil, and Michel Ledoux. *Analysis and geometry of Markov diffusion operators*, volume 348. Springer Science & Business Media, 2013.

[3] Jacob Bamberger, Adam Gosztolai, Pierre Vandergheynst, Michael M Bronstein, and Iolo Jones. Riemannian metric matching for Scalable Geometric Modelling of Distributions. In *ICML and ICLR 2026 Workshop on Geometry-grounded Representation Learning and Generative Modeling*, 2026.

[4] Jacob Bamberger, Iolo Jones, Dennis Duncan, Michael M. Bronstein, Pierre Vandergheynst, and Adam Gosztolai. Carré du champ flow matching: better quality-generalisation tradeoff in generative models. In *The Fourteenth International Conference on Learning Representations*, 2026.

[5] Mikhail Belkin and Partha Niyogi. Laplacian eigenmaps for dimensionality reduction and data representation. *Neural computation*, 15(6):1373–1396, 2003.

[6] Mikhail Belkin and Partha Niyogi. Semi-supervised learning on Riemannian manifolds. *Machine learning*, 56(1):209–239, 2004.

[7] Mikhail Belkin and Partha Niyogi. Towards a theoretical foundation for Laplacian-based manifold methods. *Journal of Computer and System Sciences*, 74(8):1289–1308, 2008.

[8] 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.

[9] Patrick Billingsley. *Convergence of probability measures*. John Wiley & Sons, 2013.

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

[11] Valentin De Bortoli. Convergence of denoising diffusion models under the manifold hypothesis. *Transactions on Machine Learning Research*, 2022. ISSN 2835-8856. Expert Certification.

[12] Bradley CA Brown, Anthony L Caterini, Brendan Leigh Ross, Jesse C Cresswell, and Gabriel Loaiza-Ganem. Verifying the union of manifolds hypothesis for image data. *arXiv preprint arXiv:2207.02862*, 2022.

[13] Sam Buchanan, Druv Pai, Yi Ma, and Valentin De Bortoli. On the edge of memorization in diffusion models. In *The Thirty-ninth Annual Conference on Neural Information Processing Systems*, 2026.

[14] Carlos Bustamante, Christian M Kaiser, Erik Lindahl, Robert Sosa, and Giovanni Volpe. How proteins fold. *Nature Reviews Molecular Cell Biology*, pp. 1–2, 2026.

[15] Saptarshi Chakraborty, Quentin Berthet, and Peter L Bartlett. Generalization Properties of Score-matching Diffusion Models for Intrinsically Low-dimensional Data. *arXiv preprint arXiv:2603.03700*, 2026.

[16] Ricky T. Q. Chen and Yaron Lipman. Flow matching on general geometries. In *The Twelfth International Conference on Learning Representations*, 2024.

[17] Xiang Cheng, Jingzhao Zhang, and Suvrit Sra. Efficient sampling on riemannian manifolds via langevin mcmc. In *Proceedings of the 36th International Conference on Neural Information Processing Systems*, NIPS '22, Red Hook, NY, USA, 2022. Curran Associates Inc. ISBN 9781713871088.

[18] Xiang Cheng, Jingzhao Zhang, and Suvrit Sra. Theory and algorithms for diffusion processes on Riemannian manifolds. *arXiv preprint arXiv:2204.13665*, 2022.

[19] Sinho Chewi. *Log-concave sampling*. Forthcoming, 2026. Available online at `https://chewisinho.github.io/`.

[20] Fan RK Chung. *Spectral graph theory*, volume 92. American Mathematical Soc., 1997.

[21] Budd C.J. and Piggott M.D. Geometric integration and its applications. *Handbook of Numerical Analysis*, pp. 35–139, 2003. ISSN 1570-8659. doi: 10.1016/s1570-8659(02)11002-7.

[22] Ronald R Coifman and Stéphane Lafon. Diffusion maps. *Applied and computational harmonic analysis*, 21(1):5–30, 2006.

[23] Keenan Crane, Clarisse Weischedel, and Max Wardetzky. Geodesics in heat: A new approach to computing distance based on heat flow. *ACM Transactions on Graphics (ToG)*, 32(5):1–11, 2013.

[24] Valentin De Bortoli, James Thornton, Jeremy Heng, and Arnaud Doucet. Diffusion schrödingerbridge with applications to score-based generative modeling. *Advances in neural information processing systems*, 34:17695–17709, 2021.

[25] 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.

[26] Riccardo De Santi, Marin Vlastelica, Ya-Ping Hsieh, Zebang Shen, Niao He, and Andreas Krause. Provable maximum entropy manifold exploration via diffusion models. *International Conference on Machine Learning (ICML)*, 2025.

[27] Manfredo Perdigao Do Carmo and J Flaherty Francis. *Riemannian geometry*, volume 2. Springer, 1992.

[28] Stewart N Ethier and Thomas G Kurtz. *Markov processes: characterization and convergence*. John Wiley & Sons, 2009.

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

[30] Masatoshi Fukushima, Yoichi Oshima, and Masayoshi Takeda. *Dirichlet forms and symmetric Markov processes*, volume 19. W. de Gruyter Berlin, 1994.

[31] Nicolás García Trillos and Dejan Slepčev. Continuum limit of total variation on point clouds. *Archive for rational mechanics and analysis*, 220(1):193–241, 2016.

[32] Nicolás García Trillos, Moritz Gerlach, Matthias Hein, and Dejan Slepčev. Error estimates for spectral convergence of the graph Laplacian on random geometric graphs toward the Laplace–Beltrami operator. *Foundations of Computational Mathematics*, 20(4):827–887, 2020.

[33] Mark Girolami and Ben Calderhead. Riemann manifold langevin and hamiltonian monte carlo methods. *Journal of the Royal Statistical Society Series B: Statistical Methodology*, 73(2): 123–214, 2011.

[34] Alexander Grigoryan. *Heat kernel and analysis on manifolds*, volume 47. American Mathematical Soc., 2009.

[35] Maria Letizia Guerra and Luciano Stefanini. A comparative simulation study for estimating diffusion coefficient. *Mathematics and computers in simulation*, 53(3):193–203, 2000.

[36] Ernst Hairer, Marlis Hochbruck, Arieh Iserles, and Christian Lubich. Geometric numerical integration. *Oberwolfach Reports*, 3(1):805–882, 2006.

[37] Matthias Hein, Jean-Yves Audibert, and Ulrike von Luxburg. Graph laplacians and their convergence on random neighborhood graphs. *Journal of Machine Learning Research*, 8(6), 2007.

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

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

[40] Jean Jacod and Philip Protter. *Discretization of processes*, volume 67. Springer Science & Business Media, 2011.

[41] Iolo Jones. Diffusion Geometry. *arXiv preprint arXiv:2405.10858*, 2024.

[42] Iolo Jones and David Lanners. Computing diffusion geometry. *arXiv preprint arXiv:2602.06006*, 2026.

[43] Erik Jørgensen. The central limit problem for geodesic random walks. *Zeitschrift für Wahrscheinlichkeitstheorie und verwandte Gebiete*, 32(1):1–64, 1975.

[44] Andrey Kharitenko, Zebang Shen, Riccardo de Santi, Niao He, and Florian Doerfler. Landing with the Score: Riemannian Optimization through Denoising, 2025.

[45] Takashi Kumagai and Karl-Theodor Sturm. Construction of diffusion processes on fractals, d-sets, and general metric measure spaces. *Journal of Mathematics of Kyoto University*, 45(2): 307–327, 2005.

[46] Kazuhiro Kuwae and Takashi Shioya. Convergence of spectral structures: A functional analytic theory and its applications to spectral geometry. *Communications in Analysis and Geometry*, 11(4):599–673, September 2003. ISSN 1019-8385.

[47] Peter D Lax. *Functional analysis*. John Wiley & Sons, 2014.

[48] Brigitte Le Roux and Henry Rouanet. *Geometric data analysis: from correspondence analysis to structured data analysis*. Springer, 2005.

[49] Xiang Li, Zebang Shen, Ya-Ping Hsieh, and Niao He. When Scores Learn Geometry: Rate Separations under the Manifold Hypothesis. In *The Fourteenth International Conference on Learning Representations*, 2026.

[50] Yaron Lipman, Ricky T. Q. Chen, Heli Ben-Hamu, Maximilian Nickel, and Matthew Le. Flow matching for generative modeling. In *The Eleventh International Conference on Learning Representations*, 2023.

[51] Stephan Mandt, Matthew D Hoffman, and David M Blei. Stochastic gradient descent as approximate bayesian inference. *Journal of Machine Learning Research*, 18(134):1–35, 2017.

[52] Kanti V Mardia and Peter E Jupp. *Directional statistics*. John Wiley & Sons, 2009.

[53] Emile Mathieu, Vincent Dutordoir, Michael Hutchinson, Valentin De Bortoli, Yee Whye Teh, and Richard Turner. Geometric neural diffusion processes. *Advances in Neural Information Processing Systems*, 36:53475–53507, 2023.

[54] Stanislav A Molchanov. Diffusion processes and riemannian geometry. *Russian Mathematical Surveys*, 30(1):1–63, 1975.

[55] Radford M Neal. MCMC using Hamiltonian dynamics. *Handbook of markov chain monte carlo*, pp. 47–95, 2011.

[56] Samer A. Nene, Shree K. Nayar, and Hiroshi Murase. Columbia object image library (coil-20). Technical Report CUCS-005-96, Department of Computer Science, Columbia University, February 1996.

[57] Bernt Øksendal. Stochastic differential equations. In *Stochastic differential equations: an introduction with applications*, pp. 38–50. Springer, 2003.

[58] Sam Patterson and Yee Whye Teh. Stochastic gradient Riemannian Langevin dynamics on the probability simplex. *Advances in neural information processing systems*, 26, 2013.

[59] Grigorios A Pavliotis. Stochastic processes and applications. *Texts in applied mathematics*, 60:41–43, 2014.

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

[61] 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.

[62] Hannes Risken. Fokker-planck equation. In *The Fokker-Planck equation: methods of solution and applications*, pp. 63–95. Springer, 1989.

[63] Felix Schulze, Fanny Bastien, and Hugo Béchet. F. Schulze - An introduction to weak mean curvature flow 1, June 2021. URL `https://hal.science/hal-03322467`.

[64] Yang Song, Jascha Sohl-Dickstein, Diederik P. Kingma, Abhishek Kumar, Stefano Ermon, and Ben Poole. Score-Based Generative Modeling through Stochastic Differential Equations, 2021.

[65] Karl-Theodor Sturm. Diffusion processes and heat kernels on metric spaces. *The Annals of Probability*, 26(1):1–55, 1998.

[66] Toshikazu Sunada. Geodesic flows and geodesic random walks. In *Geometry of geodesics and related topics*, volume 3, pp. 47–86. Mathematical Society of Japan, 1984.

[67] Nicolás Garcia Trillos and Dejan Slepčev. On the rate of convergence of empirical measures in ∞-transportation distance. *Canadian Journal of Mathematics*, 67(6):1358–1383, 2015.

[68] Nicolas Garcia Trillos and Dejan Slepčev. A variational approach to the consistency of spectral clustering. *Applied and Computational Harmonic Analysis*, 45(2):239–281, 2018.

[69] Greg Turk and Marc Levoy. Zippered polygon meshes from range images. In *Proceedings of the 21st annual conference on Computer graphics and interactive techniques*, pp. 311–318, 1994.

[70] Veeravalli S Varadarajan. On the convergence of sample probability distributions. *Sankhyā: The Indian Journal of Statistics (1933-1960)*, 19(1/2):23–26, 1958.

[71] Cédric Villani et al. *Optimal transport: old and new*, volume 338. Springer, 2009.

[72] Max-K. von Renesse. Intrinsic coupling on riemannian manifolds and polyhedra. *Electronic Journal of Probability*, 9(none):411, 01 2004. ISSN 1083-6489. doi: 10.1214/ejp.v9-205.

[73] Shiyu Wang, Mariam Avagyan, Yihan Shen, Arnaud Lamy, Tingran Wang, Szabolcs Márka, Zsuzsanna Márka, and John Wright. Fast, accurate manifold denoising by tunneling riemannian optimization. In *Proceedings of the 42nd International Conference on Machine Learning*, ICML'25. JMLR.org, 2025.

[74] Zhou Wang, Alan C Bovik, Hamid R Sheikh, and Eero P Simoncelli. Image quality assessment: from error visibility to structural similarity. *IEEE transactions on image processing*, 13 (4):600–612, 2004.

[75] Joseph L Watson, David Juergens, Nathaniel R Bennett, Brian L Trippe, Jason Yim, Helen E Eisenach, Woody Ahern, Andrew J Borst, Robert J Ragotte, Lukas F Milles, et al. De novo design of protein structure and function with rfdiffusion. *Nature*, 620(7976):1089–1100, 2023.

[76] Max Welling and Yee W Teh. Bayesian learning via stochastic gradient Langevin dynamics. In *Proceedings of the 28th international conference on machine learning (ICML-11)*, pp. 681–688, 2011.

[77] Jason Yim, Brian L Trippe, Valentin De Bortoli, Emile Mathieu, Arnaud Doucet, Regina Barzilay, and Tommi Jaakkola. SE (3) diffusion model with application to protein backbone generation. In *Proceedings of the 40th International Conference on Machine Learning*, pp. 40001–40039, 2023.

[78] Olga Zaghen, Floor Eijkelboom, Alison Pouplin, Cong Liu, Max Welling, Jan-Willem van de Meent, and Erik J Bekkers. Riemannian Variational Flow Matching for Material and Protein Design. In *The Fourteenth International Conference on Learning Representations*, 2026.

[79] Dirk Zahn. Exploring the mechanisms of reactions in solution from transition path sampling molecular dynamics simulations. *Journal of Chemical Theory and Computation*, 2(1):107–114, 2006.

[80] Zhiyuan Zhan and Masashi Sugiyama. Riemannian Langevin dynamics: Strong convergence of Geometric Euler-Maruyama scheme. arXiv preprint arXiv:2603.03626 [stat.ML], 2026.

A More on IMDs

A.1 Construction of the extrinsic representation

Here we explain in further detail the construction of $\mathbb{R}^D$-valued diffusion confined to $\mathcal{M}$ through the use of $\mathcal{L}$ and Γ from Section 4.1. We principally follow the construction from Chapter 1 of [39]. For the sake of completeness, we do *not* identify $\mathcal{M}^d$ with its embedded image $\iota(\mathcal{M}) \subset \mathbb{R}^D$, and keep everything separate. We will work entirely intrinsically, and formally recover the presented extrinsic formulation of Section 4 by passing all terms through the embedding ι.

Intrinsic to extrinsic diffusion operator The embedding map $\iota : \mathcal{M} \to \mathbb{R}^D$ may be understood through composition with the coordinates function $\pi^\alpha : \mathbb{R}^D \to \mathbb{R}$ given by $\pi^\alpha(y) = y^\alpha$ for $\alpha = 1, \dots, D$ as

$$\iota^\alpha := \pi^\alpha \circ \iota : \mathcal{M} \to \mathbb{R}. \tag{14}$$

This matches the type of assumed smooth test functions (or observables) $f : \mathcal{M} \to \mathbb{R}$, which we will denote $C^\infty(\mathcal{M}; \mathbb{R})$.

Intrinsically, the Riemannian gradient sends $f \in C^\infty(\mathcal{M}; \mathbb{R})$ to a vector field, which is a section of $\mathcal{M}$'s tangent bundle $T\mathcal{M}$, denoted $\Gamma(T\mathcal{M})$. That is, we have $\nabla_\mathcal{M} : C^\infty(\mathcal{M}; \mathbb{R}) \to \Gamma(T\mathcal{M})$. The Laplace-Beltrami operator has the type $\Delta\mathcal{M} : C^\infty(\mathcal{M}; \mathbb{R}) \to C^\infty(\mathcal{M}; \mathbb{R})$.

For a diffusion process on $\mathcal{M}$, the intrinsic SDE form would require extra introduction of Stratonovich stochastic integration, which we will avoid for the sake of remaining self-contained.[2] We therefore directly consider the diffusion operator characterization of a diffusion process with drift $b \in \Gamma(T\mathcal{M})$ and isotropic diffusion term (without loss of generality), given intrinsically by

$$\mathcal{L}f := \langle b, \nabla_\mathcal{M} f \rangle_\mathcal{M} + \frac{1}{2}\Delta_\mathcal{M} f. \tag{15}$$

Now let $x \in \mathcal{M}$ and consider the embedding itself ι^α as a vector of test functions, and plug it into Eq. (15) at x:

$$\mathcal{L}\iota(x) = (\langle b_x, \nabla_\mathcal{M} \iota^\alpha(x) \rangle_{g_x})_{\alpha=1}^D + \frac{1}{2}\Delta_\mathcal{M}\iota(x). \tag{16}$$

For each $\alpha = 1, \dots, D$, we look at the first term of Eq. (16):

$$\langle b_x, \nabla_\mathcal{M} \iota^\alpha(x) \rangle_{g_x} = d\iota_x^\alpha(b_x) \tag{17}$$

by definition of the Riemannian metric g at $x \in \mathcal{M}$, and we use $\iota^\alpha = \pi^\alpha \circ \iota$ to get

$$\langle b_x, \nabla_\mathcal{M} \iota^\alpha(x) \rangle_{g_x} = [d\iota_x(b_x)]^\alpha, \tag{18}$$

which is the α^{th} component of $d\iota_x(b_x) \in T_{\iota(x)}\iota(\mathcal{M})$, the vector field corresponding to b on the embedding of $\mathcal{M}$. Therefore, we may interpret applying $\mathcal{L}$ to the embedding ι on each ι^α, and stacking the result in a vector, leading to the extrinsic drift indeed being b fed through the embedding. Using the fact of $\Delta_\mathcal{M}\iota = \vec{\nu}$ where $\vec{\nu}$ is the mean curvature vector [63] therefore gives us the drift term of Eq. (8).

[2] Interested readers may refer themselves to Chapter 1 of [39] or Appendix B of [2].

The carré-du-champ We now discuss the derivation of Eq. (7). The CdC of a second-order differential operator $\mathcal{L}$ is generally given by the formula

$$\Gamma_{\mathcal{L}}(f,g) := \mathcal{L}(fg) - f\mathcal{L}(g) - g\mathcal{L}(f). \tag{19}$$

We now show that for any operator $\mathcal{L}$ of the type (15), its associated CdC only captures the diffusivity part, i.e. $\Gamma_{\mathcal{L}} = \Gamma_{\frac{1}{2}\Delta_{\mathcal{M}}}$, which we derive below. Let $f, g \in C^\infty(\mathcal{M};\mathbb{R})$, then we have

$$\begin{aligned}\Gamma_{\mathcal{L}}(f,g) &:= \mathcal{L}(fg) - f\mathcal{L}(g) - g\mathcal{L}(f)\\ &= \langle b, \nabla_{\mathcal{M}}(fg)\rangle_{\mathcal{M}} + \frac{1}{2}\Delta_{\mathcal{M}}(fg) - f\langle b, \nabla_{\mathcal{M}} g\rangle_{\mathcal{M}}\\ &\quad -\frac{1}{2}f\Delta_{\mathcal{M}} g - g\langle b, \nabla_{\mathcal{M}} f\rangle_{\mathcal{M}} - \frac{1}{2}g\Delta_{\mathcal{M}} f\\ &= \frac{1}{2}\left[\Delta_{\mathcal{M}}(fg) - f\Delta_{\mathcal{M}} g - g\Delta_{\mathcal{M}} f\right] = \Gamma_{\frac{1}{2}\Delta_{\mathcal{M}}}(f,g).\end{aligned} \tag{20}$$

Therefore, the CdC of an arbitrary diffusion operator only depends on the diffusivity term, which is given by the LBO (3) throughout our entire paper.

Extrinsic Riemannian operators Although the dynamics are defined intrinsically on the embedded manifold $\iota(\mathcal{M}) \subset \mathbb{R}^D$, it is often convenient to regard functions on $\iota(\mathcal{M})$ as restrictions of smooth ambient functions. Since $\iota(\mathcal{M})$ is a smooth embedded submanifold, any $f \in C^\infty(\iota(\mathcal{M}))$ admits a smooth extension $\widetilde{f}$ to a neighborhood of $\iota(\mathcal{M})$; locally, one may for instance take $\widetilde{f} = f \circ \varphi$, where φ is the smooth nearest-point projection onto $\mathcal{M}$ in a tubular neighborhood. This allows ambient derivatives such as $\nabla\widetilde{f}$ to be used while recovering the intrinsic quantities after restriction to $\iota(\mathcal{M})$, e.g. $\nabla_{\mathcal{M}} f = P\nabla\widetilde{f}$. The particular extension is immaterial for the intrinsic generator, which depends only on the tangential derivatives of f along $\iota(\mathcal{M})$, but it is essential for expressing the dynamics as an SDE in the ambient space $\mathbb{R}^D$.

IMD Algorithm For completeness, we summarize the IMD pipeline in the following algorithm.

Input: Samples $X_N = \{x_i\}_{i=1}^N$, step size h, steps K, initial point X_0
Output: IMD trajectory $\{X_k\}_{k=0}^K$
$\widehat{\mathcal{L}}_N \leftarrow$ estimate diffusion operator from X_N
$\widehat{b}_N \leftarrow \widehat{\mathcal{L}}_N \iota$
$\widehat{A}_N \leftarrow \widehat{\Gamma}_N(\iota,\iota)$ through Eq. (7) or data-driven
$\widehat{b}_N, \widehat{A}_N \leftarrow$ interpolate off-sample from X_N
for $k = 0, \ldots, K-1$ **do**
  $\xi_k \sim \mathcal{N}(0, \mathbf{I}_D)$
  $X_{k+1} \leftarrow X_k + h\,\widehat{b}_N(X_k) + \sqrt{h}\,\widehat{A}_N^{1/2}(X_k)\xi_k$
end
return $\{X_k\}_{k=0}^K$

Algorithm 1: Implicit Manifold-valued Diffusion

A.2 Proof of Proposition 1

Without loss of generality, we will consider the GEM scheme with no drift ($b = 0$) that is driven by isotropic Brownian motion ($\sigma = \mathbf{I}_D$). Following [80], we use a Taylor expansion of the exponential map around the geodesic $\sqrt{h} \mapsto \exp_{X_k}(\sqrt{h}\xi)$ to observe the local behaviour of the scheme:

$$\begin{aligned}X_{k+1}^{\text{GEM}} &= \exp_{X_k^{\text{GEM}}}\left(\sqrt{h}\xi_k^{\|}\right)\\ &= X_k^{\text{GEM}} + \sqrt{h}\xi_k^{\|} + \frac{h}{2}\Pi_{X_k^{\text{GEM}}}(\xi_k^{\|}, \xi_k^{\|}) + \mathcal{O}(\|\sqrt{h}\xi_k^{\|}\|^3),\end{aligned} \tag{21}$$

where Π_x is the second fundamental form of $\mathcal{M}$ at $x \in \mathcal{M}$. A simple Euler-Maruyama discretization of Eq. (8) gives us

$$X_{k+1} = X_k + \frac{1}{2}(\Delta_{\mathcal{M}}\iota)(X_k) + \Gamma(\iota)(X_k)\xi_k. \tag{22}$$

Therefore, conditionally on $X_k = X_k^{\text{GEM}} = x$, we have that

$$X_{k+1}^{\text{GEM}} = x + \sqrt{h}\xi_k^{\parallel} + \frac{h}{2}\Pi_x(\xi_k^{\parallel}, \xi_k^{\parallel}) \tag{23}$$

and

$$X_{k+1} = x + \frac{h}{2}\vec{\nu}_x + \sqrt{h}\tilde{\xi}_k^{\parallel}, \tag{24}$$

where we have used the fact that the Laplace-Beltrami operator applied to the embedding gives the mean curvature vector at x [63], and that $\Gamma(\iota)(x) = P_{T_x\mathcal{M}}$ the projection onto the tangent space at $x \in \mathcal{M}$. It remains to take expectations, as is done to describe the evolution of diffusions [2; 39], we obtain

$$\mathbb{E}[\|X_{k+1}^{\text{GEM}} - X_{k+1}\|_2^2 \mid X_k = x] = \mathcal{O}(h^2), \tag{25}$$

as we have that $\mathbb{E}[\Pi_x(\xi_k^{\parallel}, \xi_k^{\parallel})] = \vec{\nu}_x$ and both $\xi_k^{\parallel}, \tilde{\xi}_k^{\parallel}$ are zero mean Gaussian vectors.

□

B Dirichlet forms & diffusion operators

In this section we prove the strong resolvent convergence stated in Proposition 2. We first recall the Dirichlet-form viewpoint and the graph-to-manifold identification maps, then establish Mosco convergence of the α-corrected graph forms to their continuum counterpart. This approach is similar to construction of diffusion processes on metric spaces or fractals via Γ-convergence of a sequence of Dirichlet forms to the limiting target space [45; 65]. For notational simplicity, we rescale all operators by a factor of two, leading us to the convention that $\Delta_{\mathcal{M}}$ is the generator of Brownian motion.

B.1 Primer on Dirichlet forms

A *Dirichlet form* provides an energy-based description of a Laplace-type operator [30]. Given a reference measure ν on $\mathcal{M}$, a typical example is

$$\mathcal{E}(f, g) = \int_{\mathcal{M}} \langle \nabla_{\mathcal{M}} f, \nabla_{\mathcal{M}} g \rangle \, d\nu, \tag{26}$$

defined on an appropriate Sobolev space. Under standard regularity assumptions, such a symmetric nonnegative form determines a self-adjoint operator L through

$$\mathcal{E}(f, g) = \langle Lf, g \rangle_{L^2(\mathcal{M},\nu)}. \tag{27}$$

For the Riemannian volume measure this recovers the Laplace-Beltrami operator, while weighted measures give weighted Laplace-type operators [2].

Graph Dirichlet forms Following [32], let

$$w_{ij} := \frac{1}{N\epsilon_N^d}\eta\left(\frac{\|x_i - x_j\|_2}{\epsilon_N}\right), \tag{28}$$

where η is the kernel from Section 6 and

$$\sigma_\eta := \int_{\mathbb{R}^d} \eta(\|z\|_2) z_1^2 \, dz. \tag{29}$$

For $u : X_N \to \mathbb{R}$, define

$$(\delta u)(i, j) := \frac{u(x_i) - u(x_j)}{\epsilon_N}. \tag{30}$$

If each undirected edge is counted once, the corresponding graph Dirichlet form is

$$\mathcal{D}_N(u, v) := \frac{1}{N\sigma_\eta\epsilon_N^2} \sum_{i \sim j} w_{ij}(u_i - u_j)(v_i - v_j), \tag{31}$$

with $\mathcal{D}_N(u) := \mathcal{D}_N(u, u)$.

Manifold Dirichlet forms For a positive density $\rho : \mathcal{M} \to \mathbb{R}_{>0}$, let

$$d\nu_\rho(x) := \rho(x) d\,\mathrm{vol}_{\mathcal{M}}(x). \tag{32}$$

The corresponding continuum Dirichlet form is

$$\mathcal{D}_\rho(f, g) := \int_{\mathcal{M}} \langle \nabla_{\mathcal{M}} f, \nabla_{\mathcal{M}} g \rangle \rho \, d\,\mathrm{vol}_{\mathcal{M}}, \tag{33}$$

for $f, g \in H^1(\mathcal{M}, \nu_\rho)$. Its associated self-adjoint operator is

$$\mathcal{L}_\rho f = -\frac{1}{\rho} \mathrm{div}_{\mathcal{M}}(\rho \nabla_{\mathcal{M}} f)\,. \tag{34}$$

Continuum limit For intuition, if u is the restriction of a smooth $f : \mathcal{M} \to \mathbb{R}$, then Eq. equation 31 has the formal continuum limit

$$\mathcal{D}_N(u) \longrightarrow \frac{1}{\sigma_\eta \epsilon_N^{d+2}} \iint_{\mathcal{M}\times\mathcal{M}} \eta\left(\frac{\|x-y\|_2}{\epsilon_N}\right) (f(x) - f(y))^2 \, d\mu(x) d\mu(y) \tag{35}$$

$$\longrightarrow \int_{\mathcal{M}} q(x)^2 \|\nabla_{\mathcal{M}} f(x)\|^2 \, d\,\mathrm{vol}_{\mathcal{M}}(x), \tag{36}$$

where $d\mu = q\, d\,\mathrm{vol}_{\mathcal{M}}$. The second limit follows by writing $y = \exp_x(\epsilon_N z)$ and using $f(y) - f(x) = \epsilon_N \langle \nabla_{\mathcal{M}} f(x), z\rangle + O(\epsilon_N^2)$. This informal calculation explains the appearance of the sampling density in the limiting energy; the rigorous comparison uses the identification maps below.

B.2 Graph-to-manifold identification maps

The discrete and continuum forms live on different spaces. Following [32], let

$$\mu_N := \frac{1}{N} \sum_{i=1}^N \delta_{x_i}, \tag{37}$$

and let $\mathcal{T}_N : \mathcal{M} \to X_N$ satisfy $(\mathcal{T}_N)_\# \mu = \mu_N$. Define

$$e_N := \operatorname*{ess\,sup}_{x\in\mathcal{M}} d_{\mathcal{M}}(x, \mathcal{T}_N(x)), \qquad U_i := \mathcal{T}_N^{-1}(\{x_i\}). \tag{38}$$

We now introduce the required assumptions on the sampling density q, as well as required rates of decay of the kernel η in terms of the above quantities. The following is assumed throughout the rest of this section.

Assumption 1. *Throughout this section, we assume that the sampling density $q \in C^1(\mathcal{M})$ with $0 < q_{\min} \le q \le q_{\max} < \infty$, and that $\epsilon_N \to 0$ in a regime such that the ∞-transport distance satisfies*

$$\frac{e_N}{\epsilon_N} \longrightarrow 0. \tag{39}$$

The kernel η is also assumed to satisfy the standard admissibility conditions of [32]; in particular, the hard-cutoff kernel used in Sec (6.1) is admissible.

The discretization and piecewise-constant extension maps are

$$(P_N f)(x_i) := \frac{1}{\mu(U_i)} \int_{U_i} f(x)\, d\mu(x), \tag{40}$$

and

$$(P_N^* u)(x) := \sum_{i=1}^N u(x_i) \mathbf{1}_{U_i}(x) = u(\mathcal{T}_N(x)). \tag{41}$$

Since $P_N^* u$ is generally discontinuous, we use the smoothed interpolation

$$I_N u := \Lambda_{r_N} P_N^* u, \qquad r_N = \epsilon_N - 2e_N, \tag{42}$$

where Λ_r is the smoothing operator of [32]. Together with Lemma 1, these maps are asymptotically compatible with the weighted spaces $H_{N,\alpha}$ and H_α used below.

B.3 MOSCO CONVERGENCE

Because the discrete spaces vary with N, we use generalized Mosco convergence in the sense of Kuwae & Shioya [46]. In our setting, $\mathcal{D}_{N,\alpha} \xrightarrow{M} \mathcal{D}_\alpha$ means:

- if $I_N u_N \rightharpoonup f$ weakly in H_α, then
$$\mathcal{D}_\alpha(f) \le \liminf_{N\to\infty} \mathcal{D}_{N,\alpha}(u_N); \tag{43}$$
- for every $f \in H_\alpha$, there exists $u_N \in H_{N,\alpha}$ with $I_N u_N \to f$ strongly in H_α such that
$$\limsup_{N\to\infty} \mathcal{D}_{N,\alpha}(u_N) \le \mathcal{D}_\alpha(f). \tag{44}$$

For closed nonnegative forms, generalized Mosco convergence is equivalent to generalized strong resolvent convergence of the associated self-adjoint operators [46].

B.4 PROOF OF PROPOSITION 2

We now specialize the above construction to the α-corrected graph operator from Eq. (12). Define
$$\rho_\alpha(x) := q(x)^{2-2\alpha} \tag{45}$$
and
$$\mathcal{D}_\alpha(f,g) := \int_{\mathcal{M}} q(x)^{2-2\alpha} \langle \nabla_{\mathcal{M}} f(x), \nabla_{\mathcal{M}} g(x)\rangle d\,\mathrm{vol}_{\mathcal{M}}(x). \tag{46}$$
Thus
$$H_\alpha := L^2\big(\mathcal{M}, q^{2-2\alpha} d\,\mathrm{vol}_{\mathcal{M}}\big), \tag{47}$$
and the associated weighted LBO is
$$\mathcal{L}_\alpha f = -\frac{1}{q^{2-2\alpha}} \,\mathrm{div}_{\mathcal{M}}\big(q^{2-2\alpha} \nabla_{\mathcal{M}} f\big). \tag{48}$$
Expanding the divergence gives
$$\mathcal{L}_\alpha f = -\Delta_{\mathcal{M}} f - 2(1-\alpha)\langle \nabla_{\mathcal{M}} \log q, \nabla_{\mathcal{M}} f\rangle, \tag{49}$$
as claimed in Proposition 2.

Lemma 1 (Convergence of the corrected vertex measure)**.** *Under the assumptions of Proposition 2, let $d_i^{(\alpha)}$ be the corrected degree from Eq. equation 12 and define*
$$\bar{d}_{i,N}^{(\alpha)} := (N c_\eta \epsilon_N^d)^{2\alpha-1} d_i^{(\alpha)}, \qquad c_\eta := \int_{\mathbb{R}^d} \eta(\|z\|_2)\, dz. \tag{50}$$
On the sampling regime $\epsilon_N \to 0$ and $e_N/\epsilon_N \to 0$,
$$\max_{1\le i\le N} \left| \bar{d}_{i,N}^{(\alpha)} - q(x_i)^{1-2\alpha} \right| \longrightarrow 0. \tag{51}$$
Consequently,
$$\nu_{N,\alpha} := \frac{1}{N} \sum_{i=1}^N \bar{d}_{i,N}^{(\alpha)} \delta_{x_i} \rightharpoonup q^{2-2\alpha} d\,\mathrm{vol}_{\mathcal{M}} =: \nu_\alpha. \tag{52}$$
In particular, $H_{N,\alpha} := L^2(X_N, \nu_{N,\alpha})$ converges to the reference space H_α above.

Proof. See Appendix E. □

With the weighted vertex measure fixed, let $\mathcal{D}_{N,\alpha}$ denote the symmetric Dirichlet form associated with the random-walk operator $\widehat{\mathcal{L}}_{N,\alpha}$ from Eq. (12) on $H_{N,\alpha}$:
$$\mathcal{D}_{N,\alpha}(u,v) := \langle \widehat{\mathcal{L}}_{N,\alpha} u, v\rangle_{H_{N,\alpha}}. \tag{53}$$
Equivalently, if $i \sim j$ counts each undirected edge once,
$$\mathcal{D}_{N,\alpha}(u,v) = \frac{c_{\epsilon_N}}{N} (N c_\eta \epsilon_N^d)^{2\alpha-1} \sum_{i\sim j} \frac{W_{ij}}{m_i^\alpha m_j^\alpha} (u_i - u_j)(v_i - v_j). \tag{54}$$
Thus the row normalization in Eq. (12) is encoded by the weighted vertex measure $\nu_{N,\alpha}$.

Lemma 2 (Discrete-to-continuum energy comparison)**.** *Assume $q \in C^1(\mathcal{M})$ with $0 < q_{\min} \le q \le q_{\max} < \infty$, let $\epsilon_N \to 0$, and let $e_N = d_\infty(\mu_N, \mu)$ satisfy $e_N/\epsilon_N \to 0$. Then, for N sufficiently large and every $f \in H^1(\mathcal{M})$,*

$$\mathcal{D}_{N,\alpha}(P_N f) \le \left(1 + C'_{1,\alpha}\epsilon_N + C'_{2,\alpha}\frac{e_N}{\epsilon_N} + C'_{3,\alpha}\epsilon_N^2\right)\mathcal{D}_\alpha(f). \tag{55}$$

Likewise, for every $u : X_N \to \mathbb{R}$,

$$\mathcal{D}_\alpha(I_N u) \le \left(1 + C''_{1,\alpha}\epsilon_N + C''_{2,\alpha}\frac{e_N}{\epsilon_N} + C''_{3,\alpha}\epsilon_N^2\right)\mathcal{D}_{N,\alpha}(u). \tag{56}$$

The constants are independent of N.

Proof. See Appendix D. □

Set

$$\delta'_N := C'_{1,\alpha}\epsilon_N + C'_{2,\alpha}\frac{e_N}{\epsilon_N} + C'_{3,\alpha}\epsilon_N^2, \tag{57}$$

$$\delta''_N := C''_{1,\alpha}\epsilon_N + C''_{2,\alpha}\frac{e_N}{\epsilon_N} + C''_{3,\alpha}\epsilon_N^2. \tag{58}$$

By assumption, $\delta'_N, \delta''_N \to 0$.

Lower bound Let $u_N \in H_{N,\alpha}$ satisfy $I_N u_N \rightharpoonup f$ weakly in H_α. Since the closed convex quadratic form $\mathcal{D}_\alpha : H_\alpha \to [0,\infty]$ is weakly lower semicontinuous,

$$\begin{aligned}\mathcal{D}_\alpha(f) &\le \liminf_{N\to\infty} \mathcal{D}_\alpha(I_N u_N) \\ &\le \liminf_{N\to\infty} (1+\delta''_N)\mathcal{D}_{N,\alpha}(u_N),\end{aligned} \tag{59}$$

where the second inequality follows from Eq. (56). Since $\delta''_N \to 0$, this gives the Mosco lower bound.

Upper bound For $f \in H^1(\mathcal{M})$, take the recovery sequence $u_N = P_N f$. The identification estimates of [32], together with Lemma 1, give

$$I_N P_N f \longrightarrow f \qquad \text{strongly in } H_\alpha. \tag{60}$$

Using Eq. (55),

$$\limsup_{N\to\infty} \mathcal{D}_{N,\alpha}(P_N f) \le \limsup_{N\to\infty} (1+\delta'_N)\mathcal{D}_\alpha(f) = \mathcal{D}_\alpha(f). \tag{61}$$

This proves the recovery condition for every f with finite energy; when $\mathcal{D}_\alpha(f) = +\infty$, the Mosco upper bound is vacuous. Hence

$$\mathcal{D}_{N,\alpha} \xrightarrow{M} \mathcal{D}_\alpha. \tag{62}$$

Strong resolvent convergence By the Kuwae–Shioya equivalence between generalized Mosco convergence of closed nonnegative forms and generalized strong-resolvent convergence of their associated self-adjoint operators [46],

$$\widehat{\mathcal{L}}_{N,\alpha} \xrightarrow{\text{s.r.}} \mathcal{L}_\alpha, \tag{63}$$

in the sense of Eq. (11), which is exactly the claim of Proposition 2.

□

B.5 Consistency of the continuous IMD lift

Although $\widehat{\mathcal{L}}_N$ is a graph jump operator at finite N, Eq. (10) defines a continuous diffusion through its first two local moments (9). Let G_N denote the generator of this continuous IMD,

$$G_N \tilde{f} := \langle \widehat{b}_N, \nabla \tilde{f} \rangle + \frac{1}{2}\widehat{A}_N : \nabla^2 \tilde{f}, \tag{64}$$

for a smooth ambient extension $\tilde{f}$ of $f \in C^3(\mathcal{M})$.

Proposition 3. *Under Assumption 1, the graph generator $\widehat{\mathcal{L}}_N$ and the continuous IMD generator $\mathcal{G}_N$ are asymptotically equivalent under the identification maps of Appendix B.*

Proof. For neighboring samples, let $\delta_{ij} := \iota(x_j) - \iota(x_i)$. A second-order Taylor expansion gives

$$f(x_j) - f(x_i) = \nabla \tilde{f}(x_i)^\top \delta_{ij} + \frac{1}{2}\delta_{ij}^\top \nabla^2 \tilde{f}(x_i)\delta_{ij} + \mathcal{O}(\|\delta_{ij}\|^3). \tag{65}$$

Substituting this into the graph operator, the first two terms are exactly $\widehat{\mathcal{L}}_N \iota$ and $\widehat{\Gamma}_N(\iota, \iota)$, hence

$$\widehat{\mathcal{L}}_N f = G_N \tilde{f} + \mathcal{O}(\epsilon_N), \tag{66}$$

where the remainder vanishes as $\epsilon_N \to 0$ by locality of the graph and Assumption 1. Stability of the interpolation map preserves this convergence after identification with the continuum space. □

Thus, while the finite-N graph process is a jump process, the continuous IMD constructed from its drift and CdC has the same limiting generator. Consequently, the strong-resolvent limit established above also describes the limiting dynamics of Eq. (10).

C Proof of Theorem 1

Denote by $(X_t^N)_t$ the continuous IMD defined in Eq. (10), with generator G_N, and by $(X_t)_t$ the limiting manifold diffusion generated by $\mathcal{L}$. By Proposition 3, G_N is asymptotically equivalent to $\widehat{\mathcal{L}}_N$ under the identification maps. Thus, it suffices to study the semigroup convergence induced by $\widehat{\mathcal{L}}_N \to \mathcal{L}$. By standard consequences of strong resolvent convergence for self-adjoint operators, the associated Markov semigroups

$$T_t^{(N)} := e^{t\widehat{\mathcal{L}}_N}, \qquad T_t := e^{t\mathcal{L}} \tag{67}$$

satisfy

$$T_t^{(N)} f \to T_t f \qquad \text{strongly in } H \tag{68}$$

for every $f \in H$ and each fixed $t \geq 0$.

Let $0 \leq t_1 < \cdots < t_k \leq T$, and let $f_1, \ldots, f_k \in C_b(\mathcal{M})$.

By the Markov property,

$$\mathbb{E}_x\left[\prod_{j=1}^k f_j\left(X_{t_j}^N\right)\right] = T_{t_1}^{(N)}\Big(f_1\, T_{t_2-t_1}^{(N)}\big(f_2 \cdots T_{t_k-t_{k-1}}^{(N)} f_k\big)\Big)(x), \tag{69}$$

and similarly,

$$\mathbb{E}_x\left[\prod_{j=1}^k f_j\big(X_{t_j}\big)\right] = T_{t_1}\Big(f_1\, T_{t_2-t_1}\big(f_2 \cdots T_{t_k-t_{k-1}} f_k\big)\Big)(x). \tag{70}$$

Since $T_t^{(N)} \to T_t$ strongly for each fixed t, an iterative application of the semigroup convergence implies

$$\mathbb{E}_x\left[\prod_{j=1}^k f_j\left(X_{t_j}^N\right)\right] \to \mathbb{E}_x\left[\prod_{j=1}^k f_j\big(X_{t_j}\big)\right].$$

Hence the finite-dimensional distributions of $(X_t^N)_t$ converge to those of $(Z_t)_t$.

It remains to prove tightness of the laws of $(X_t^N)_t$ on $(C[0,T], \mathcal{M})$. Since $\mathcal{M}$ is compact, the coefficients of the limiting diffusion are smooth and therefore bounded. Under the standing assumptions on the graph approximation, the approximating processes inherit uniform moment bounds on increments. In particular, there exists a constant $C > 0$, independent of N, such that for all $0 \leq s \leq t \leq T$,

$$\sup_N \mathbb{E}_x\Big[\|X_t^N, X_s^N\|^4\Big] \leq C|t-s|^2. \tag{71}$$

By the Kolmogorov-Chentsov theorem [28], the family of laws of $(X_t^N)_t$ is tight in $C([0,T];\mathcal{M})$ for all N. Combining tightness with convergence of finite-dimensional distributions [9], we conclude that

$$X_t^N \Rightarrow X_t \qquad \text{in } C([0,T];\mathcal{M}), \tag{72}$$

proving intrinsic convergence in law on path space.

□

D Proof of Lemma 2

Set $s_\alpha := q^{1-\alpha}$. Since q is positive and Lipschitz on the compact manifold $\mathcal{M}$, so are $q^{-\alpha}$ and s_α.

The standard degree estimate for geometric graphs gives

$$\max_i |\hat{q}_i - q(x_i)| \le C\left(\epsilon_N + \frac{e_N}{\epsilon_N} + \epsilon_N^2\right). \tag{73}$$

Hence, since $q \ge q_{\min} > 0$, the map $t \mapsto t^{-\alpha}$ is Lipschitz on the relevant range and therefore

$$\max_i \left|\hat{q}_i^{-\alpha} - q(x_i)^{-\alpha}\right| \le C_\alpha\left(\epsilon_N + \frac{e_N}{\epsilon_N} + \epsilon_N^2\right). \tag{74}$$

Moreover, if $x \in U_i$, then $d_{\mathcal{M}}(x, x_i) \le e_N$, so Lipschitz continuity of $q^{-\alpha}$ yields

$$\hat{q}_i^{-\alpha} = q(x)^{-\alpha}\left[1 + O_\alpha\left(\epsilon_N + \frac{e_N}{\epsilon_N} + \epsilon_N^2\right)\right]. \tag{75}$$

Consequently the product of the two endpoint corrections satisfies, uniformly for $x \in U_i$, $y \in U_j$,

$$\hat{q}_i^{-\alpha}\hat{q}_j^{-\alpha} = q(x)^{-\alpha}q(y)^{-\alpha}\left[1 + O_\alpha(\delta_N)\right], \qquad \delta_N := \epsilon_N + \frac{e_N}{\epsilon_N} + \epsilon_N^2. \tag{76}$$

For the first inequality, Jensen's inequality on the transport cells gives

$$\mathcal{D}_{N,\alpha}(P_N f) \le (1 + C_\alpha\delta_N)\frac{1}{\sigma_\eta \epsilon_N^{d+2}} \iint_{\mathcal{M}^2} \eta_N(x,y)|f(x) - f(y)|^2 q(x)^{-\alpha}q(y)^{-\alpha}\, d\mu(x)d\mu(y), \tag{77}$$

where η_N differs from $\eta(d_{\mathcal{M}}(x,y)/\epsilon_N)$ only through the transport and extrinsic–intrinsic distance errors. As in Lemma 13 of [32], these contribute $O(e_N/\epsilon_N + \epsilon_N^2)$.

Since $d\mu = q\, d\,\mathrm{vol}_{\mathcal{M}}$, the remaining nonlocal energy is

$$\iint \eta\left(\frac{d_{\mathcal{M}}(x,y)}{\epsilon_N}\right)|f(x) - f(y)|^2 s_\alpha(x)s_\alpha(y) d\,\mathrm{vol}_{\mathcal{M}}(x) d\,\mathrm{vol}_{\mathcal{M}}(y). \tag{78}$$

Applying the usual nonlocal-to-local estimate with the density $s_\alpha = q^{1-\alpha}$ gives

$$\frac{1}{\sigma_\eta \epsilon_N^{d+2}} E_{\epsilon_N,\alpha}(f) \le (1 + C_\alpha\epsilon_N + C\epsilon_N^2)\mathcal{D}_\alpha(f), \tag{79}$$

which proves Eq. (55).

For the converse estimate, set $\widetilde{u} = P_N^* u$ and $I_N u = \Lambda_{\epsilon_N - 2e_N}\widetilde{u}$. The standard smoothing estimate, again applied with density s_α, yields

$$\mathcal{D}_\alpha(I_N u) \le (1 + C_\alpha\epsilon_N + C\epsilon_N^2)\frac{E_{\epsilon_N - 2e_N,\alpha}(\widetilde{u})}{\sigma_\eta(\epsilon_N - 2e_N)^{d+2}}. \tag{80}$$

Transporting this nonlocal energy back to the graph and using Eq. (76), using bounds on η from [32], gives

$$\frac{E_{\epsilon_N - 2e_N,\alpha}(\widetilde{u})}{\sigma_\eta(\epsilon_N - 2e_N)^{d+2}} \le \left(1 + C_\alpha\epsilon_N + C_\alpha\frac{e_N}{\epsilon_N} + C_\alpha\epsilon_N^2\right)\mathcal{D}_{N,\alpha}(u).$$

Combining the two estimates proves Eq. (56).

□

E PROOF OF LEMMA 1

Proof. Define the usual kernel density estimator at the sample points, $\hat{q}_{N,i} := \frac{m_i}{Nc_\eta \epsilon_N^d}$. Under the stated assumptions, the standard degree estimate gives

$$\max_i |\hat{q}_{N,i} - q(x_i)| \longrightarrow 0. \tag{81}$$

Since q is uniformly bounded away from zero, the map $t \mapsto t^{-\alpha}$ is Lipschitz on the relevant interval, and hence

$$\max_i \left| \hat{q}_{N,i}^{-\alpha} - q(x_i)^{-\alpha} \right| \longrightarrow 0. \tag{82}$$

Using $m_i = (Nc_\eta \epsilon_N^d)\hat{q}_{N,i}$, the rescaled corrected degree in Eq. (50) can be written as

$$\bar{d}_{i,N}^{(\alpha)} = \hat{q}_{N,i}^{-\alpha} \frac{1}{Nc_\eta \epsilon_N^d} \sum_{j=1}^N W_{ij} \hat{q}_{N,j}^{-\alpha}. \tag{83}$$

By Eq. (82), replacing $\hat{q}_{N,j}^{-\alpha}$ by $q(x_j)^{-\alpha}$ incurs a uniformly vanishing error. The remaining kernel average satisfies

$$\frac{1}{Nc_\eta \epsilon_N^d} \sum_{j=1}^N W_{ij} q(x_j)^{-\alpha} \longrightarrow q(x_i)^{1-\alpha} \tag{84}$$

uniformly in i. Indeed, its continuum counterpart is

$$\frac{1}{c_\eta \epsilon_N^d} \int_{\mathcal{M}} \eta\left(\frac{\|x_i - y\|}{\epsilon_N} \right) q(y)^{-\alpha} \, d\mu(y), \tag{85}$$

and since $d\mu(y) = q(y) d\,\mathrm{vol}_{\mathcal{M}}(y)$, the integrand carries the density $q(y)^{1-\alpha}$. Localization of the kernel around x_i therefore gives the limit $q(x_i)^{1-\alpha}$.

Combining this with $\hat{q}_{N,i}^{-\alpha} \to q(x_i)^{-\alpha}$ yields $\bar{d}_{i,N}^{(\alpha)} \to q(x_i)^{-\alpha} q(x_i)^{1-\alpha} = q(x_i)^{1-2\alpha}$, uniformly in i, proving Eq. (51).

It remains to identify the limiting vertex measure. For any $\varphi \in C(\mathcal{M})$,

$$\begin{aligned} \int_{\mathcal{M}} \varphi \, d\nu_{N,\alpha} &= \frac{1}{N} \sum_{i=1}^N \varphi(x_i) \bar{d}_{i,N}^{(\alpha)} \\ &= \frac{1}{N} \sum_{i=1}^N \varphi(x_i) q(x_i)^{1-2\alpha} + o(1). \end{aligned} \tag{86}$$

Since $\mu_N \rightharpoonup \mu$, we have

$$\frac{1}{N} \sum_{i=1}^N \varphi(x_i) q(x_i)^{1-2\alpha} \longrightarrow \int_{\mathcal{M}} \varphi(x) q(x)^{1-2\alpha} \, d\mu(x). \tag{87}$$

Finally, $q^{1-2\alpha} \, d\mu = q^{2-2\alpha} \, d\,\mathrm{vol}_{\mathcal{M}}$ implies $\nu_{N,\alpha} \rightharpoonup q^{2-2\alpha} d\,\mathrm{vol}_{\mathcal{M}}$, as claimed. □

F NUMERICAL IMPLEMENTATION DETAILS

This appendix contains all details surrounding Section 7.

F.1 DISCRETIZATION OF IMDS

We propose the following Euler-Maruyama (E-M) scheme for the numerical solution of Eq. (10)

$$X_{k+1}^N := X_k^N + \widehat{\mathcal{L}}_N(X_{kh}^N)h + \widehat{\Gamma}^{1/2}(X_k^N)\sqrt{h}\xi_{k+1}, \tag{88}$$

where $\xi_{k+1} \sim \mathcal{N}(0, \mathbf{I}_D)$ is an independently drawn Gaussian, and $h > 0$ is the step size such that $t \in [kh, (k+1)h)$ for some $k \in \mathbb{N}$. The following result demonstrates that Eq. (88) is a consistent discretization in Wasserstein-2 distance that preserves the intrinsic geometric structure in the limit of large-sample size and infinitesimal step size h.

Theorem 2. *Let X_t be the diffusion process satisfying Eq. (8), and X_k^N be the discretization of the data-driven IMD, given by Eq. (88) with step size $h > 0$. Assume that the coefficients $\widehat{\mathcal{L}}_N$ and $\widehat{\Gamma}_N$ satisfy regularity conditions of Assumption 2.*

Then, for any fixed $T > 0$, and any $\eta > 0$ and for all $t \in [kh, (k+1)h)$ with $hk \leq T$, we have

$$W_2\left(\mathit{Law}(X_t), \mathit{Law}(X_k^N)\right) \leq \eta, \tag{89}$$

as $N \to \infty, h \to 0, \epsilon \to 0$ under the assumed rates of Assumption 1 with probability $1 - o_N(1)$.

Proof. See Appendix G. □

F.2 Geodesic Estimation Details

Baseline The hourglass surface from Figure 3 is given by the following:

$$x(z, \phi) = \left((a + cz^2)\cos(\phi), (a + cz^2)\sin(\phi), z\right) \tag{90}$$

for $(a, c) = (0.55, 0.45)$.

Varadhan's formula ties the manifold's heat kernel $p_t(x, y)$ to geodesic distances $d_{\mathcal{M}}(x, y)$ as follows:

$$\lim_{t \to 0} t \log p_t(x, y) = -\frac{d_{\mathcal{M}}^2(x, y)}{4}. \tag{91}$$

Graph construction We set the bandwidth of the cutoff kernel η to be $\epsilon = (r_N/3)^2$, where r_N is the median 300^{th}-neighbour distance.

Heat kernel estimation We estimate the heat kernel through the empirical Gaussian KDE distribution $\hat{p}_t(x, y)$ produced by running $100,000$ IMDs constructed from $10,000$ points sampled from the hourglass (90). We average off-graph points via the 20 nearest-neighbours, with Gaussian weights $w_i(z) \propto \exp\left(\frac{\|z - x_i\|_2^2}{4\eta(z)}\right)$.

F.3 Path space convergence details

Baseline The reference bunny $\mathcal{B}$ has $14,944$ vertices, $29,808$ triangluar faces and 84 boundary edges. We construct the dataset by sampling $100,000$ points uniformly in mesh area by selecting faces proportionally to area. The varying sizes $N \in \{10k, 20k, 30k, 50k, 100k\}$ are nested prefixes of the original $100k$ samples.

The mesh Brownian motion is computed by drawing a uniform direction in the current face's tangent plane and travel geodesic length $2\sqrt{\Delta t}$. Adjacent triangles are unfolded into a common plane and the straight direction is continued on the next face before folding back. To avoid absorption, the 84 boundary edges are given reflecting boundary conditions. Our baseline is therefore a discretization with time step $h > 0$ of

$$dB_{k+1} = \exp_{B_k}\left(\sqrt{h}\xi_k^{\|}\right), \tag{92}$$

where ξ_k is an independently sampled Gaussian in the tangent plane of $B_k \in \mathcal{B}$.

Graph construction We chose a global radius of r_N to be the median 50^{th}-neighbour distance, dictating the cutoff of kernel η.

All computed IMDs only use these sampled points. Initial positions are uniformly sampled in the surface. We simulated for two different timesteps $h \in \{10^{-3}, 10^{-4}\}$, and average off-graph points by their 20-nearest neighbours with gaussian weighting exactly as in Section F.2. The computed drift and covariance terms are computed once per graph, and re-used for both numerical integration timesteps.

PF-ODE baseline We also compare IMDs with a probability-flow ODE (PF-ODE) [50; 64] on the mesh, aimed to match the marginals of the oracle Brownian motion (92). We simulate the heat operator on the mesh with velocity $\nabla \log \rho_t$ using the same face-unfolding tracer. This baseline remains on the mesh and is deterministic conditional on its initial position, providing control for the distinction between marginal transport and Brownian path roughness. It is therefore single-seed.

Metrics We use three complementary metrics in order to assess global path convergence, local fluctuation correctness, and off-manifold distance.

- The *path maximum mean discrepancy* (MMD) measures the discrepancy between the distributions of complete IMD and oracle (92) trajectories using a Gaussian kernel constructed from the discretized L^2 path distance

$$d^2_{\mathrm{path}}(X, Y) = \sum_{j=0}^{J} \|X_{t_j} - Y_{t_j}\|_2^2 \, \Delta t_{\mathrm{eval}}, \tag{93}$$

  so that this metric captures differences in the joint temporal law rather than only in the terminal distributions. In particular, we chose $\Delta t_{\mathrm{eval}} = 0.002$.
- Second, we report the relative error in ambient *quadratic variation* (QV)

$$\delta_{\mathrm{QV}} = \left| \frac{\mathbb{E}[\mathrm{QV}(X^{\mathrm{IMD}})]}{\mathbb{E}[\mathrm{QV}(X^{\mathrm{oracle}})]} - 1 \right|, \tag{94}$$

  where

$$\mathrm{QV}(X) = \sum_{j=0}^{J-1} \|X_{t_{j+1}} - X_{t_j}\|_2^2 \tag{95}$$

  measures local fluctuations of a process $(X_t)_t$. If QVs of two stochastic processes match, then they fluctuate the same at second order [57; 35]. This metric is complementary to the global MMD from Eq. (93).
- Finally, the *off-manifold error* is the mean Euclidean distance from each IMD state to its closest point on the reference triangular mesh, quantifying the geometric fidelity of the ambient simulation.

F.4 Torus Experiment Details

Baseline We fix one realization of $40,000$ points on a torus with radii $R = 2$ and $r = 0.7$. In (θ, ϕ)-coordinates, the chosen bimodal von Mises Fisher density is given as

$$q_{\mathrm{vMF}}(\theta, \phi) \propto \exp\left(3\cos(\theta - \frac{\pi}{4}) + 4\cos(\phi)\right) + \exp\left(3\cos(\theta - \frac{5\pi}{4}) + 4\cos(\phi - \frac{\pi}{2})\right). \tag{96}$$

Graph construction The proximity graph is built using r_N set to be the median 50^{th}-neighbour distance, and union it with a symmetric 5-nearest-neighbour scaffold to connect sparse regions in case it is too sparse, resulting in a single component. Drift and covariance are cached per choice of $\alpha \in \{0, \frac{1}{2}, 1\}$.

For each α and each seed, 100 paths of IMDs are started from uniformly sampled points. We integrate up to $T = 1500$ with timestep $h = 10^{-2}$, discard the initial 100 units, and retain ≈200 evenly spaced states per path.

Metrics We compare projected IMD samples $\{X_i\}_{i=1}^n$ with reference samples $\{Y_j\}_{j=1}^m$ using the empirical energy distance

$$\widehat{E}(X, Y) = \frac{2}{nm} \sum_{i=1}^{n} \sum_{j=1}^{m} \|X_i - Y_j\|_2 - \frac{1}{n^2} \sum_{i=1}^{n} \sum_{i'=1}^{n} \|X_i - X_{i'}\|_2 - \frac{1}{m^2} \sum_{j=1}^{m} \sum_{j'=1}^{m} \|Y_j - Y_{j'}\|_2. \tag{97}$$

IMD samples are projected onto the torus before evaluation, and distances are ambient Euclidean distances in $\mathbb{R}^3$. We use at most $2{,}000$ samples per distribution. To measure uniformity, we partition

the torus into $K = 32 \times 16{=}512$ equal-surface-area cells and denote by $\widehat{p}_j$ the empirical fraction of samples in cell j. We define coverage as

$$\text{Coverage}(\widehat{p}) = \frac{1}{K} \sum_{j=1}^{K} \mathbf{1}\{\widehat{p}_j > 0\}, \tag{98}$$

which measures the fraction of the manifold reached by the process. We additionally report the total-variation discrepancy from uniform surface measure,

$$\text{TV}(\widehat{p}, \text{Unif}) = \frac{1}{2} \sum_{j=1}^{K} \left| \widehat{p}_j - \frac{1}{K} \right|. \tag{99}$$

Coverage measures support exploration, whereas equal-area TV measures how uniformly probability mass is distributed across the explored manifold.

F.5 COIL-20 Dataset

Baseline The COIL-20 dataset contains $72{\times}20{=}1440$ images, which we convert to grayscale and resize to 32×32 pixels. We train a residual network to predict signed 5° viewpoints using all views from Objects 2-20, omitting Object 1 (the duck from Figure 6). The network comprises of five $3{\times}3$ convolutions with widths $2, 24, 24, 24, 12, 1$ respectively. We train for 40 epochs with batch size 128 and optimize using AdamW with learning rate 8×10^{-4}. We then use Object 1 views at angles $k \times 20^\circ$ to generate intermediate views. We use 18 ground-truth and 18 network-generated duck images to train IMDs.

Graph construction Given the observed images $\{x_i\}_{i=1}^{36}$, we construct the proximity graph using a Gaussian kernel with "bandwidth" ϵ set to the the median squared distance to the second nearest other support image. We chose the Gaussian kernel as it is believed to satisfy Proposition 2's assumptions [32] and guarantees a fully-connected graph in such a sparse regime.

Endpoint-conditioned IMDs For IMD computations, we retain the 16 largest eigenvalues (as done in [3] for image data), and interpolate off-graph points using the 4 nearest-neighbour graph images using the same gaussian weights as Section F.2. We condition the duck IMD on the real 45° endpoint image y by assigning Gaussian endpoint similarities to the graph nodes and propagating them through the graph generator A:

$$g_i = \exp\left(-\frac{\|x_i - y\|_2^2}{2\varepsilon}\right), \qquad h_\tau = e^{\tau A} g, \qquad \tau = T - t. \tag{100}$$

After smoothly interpolating h_τ into image space, we use the guided drift

$$b_{\text{guided}}(z, t) = b(z) + 8\,\Gamma(z)\nabla_z \log h_\tau(z) + \frac{\Pi(z)(y - z)}{\tau}, \tag{101}$$

where $\Gamma(z)$ is the interpolated rank-16 CDC covariance and $\Pi(z)$ averages local rank-2 tangent projectors obtained from the largest eigenvalues of the CDC. We additionally taper the noise near the endpoint, retract each proposal toward a local tangent plane, and apply a direct endpoint pull that strengthens near the final time. The last image is set exactly to y. Together, these steps define an approximate guided bridge, rather than an exact sample from the endpoint-conditioned IMD.

Ambient baseline We compare the IMD interpolation with the conditional mean of isotropic ambient Brownian motion between the same endpoints. Its conditional mean is the straight chord

$$\overline{x}_t^{\text{amb}} = \left(1 - \frac{t}{T}\right) x_{\text{start}} + \frac{t}{T} x_{\text{end}}. \tag{102}$$

The ambient baseline therefore interpolates directly through image space, whereas the IMD follow the implicit shape of the "manifold" of images.

G PROOF OF THEOREM 2

Proof outline. Recall that X_t is the true $\mathcal{L}$-diffusion generated by Eq. (8), X_t^N is the IMD from Eq. (10), and X_k^N is the Euler-Maruyama discretization of X_t^N. Our proof will introduce an intermediate time-continuous data-driven process X_{kh}^N and apply a triangle inequality on the Wasserstein-2 distance, bounding each appearing term individually using different classical tools.

G.1 PRELIMINARIES

We begin by discussing the regularity conditions we assume. Since we are in a computational setting, we only consider all stochastic processes $(X_t)_{t\geq 0}$ before their explosion time $e(X)$. A unique solution can be ensured to exist by assuming local Lipschitzness of the drift and diffusion terms [39], which is the setting we consider, formalized below.

Assumption 2 (Uniform second-moment control)**.** *We assume that the drift and diffusion terms $\widehat{b}_N$ and $\widehat{A}_N$ from Eq. (9) satisfy a local Lipschitz growth condition uniformly in N, i.e., there exists constants $C, K > 0$, independent of N, such that for all $x \in \mathbb{R}^D$,*

$$\|\widehat{b}_N(x) - \widehat{b}_N(y)\| + \| \widehat{A}_N^{1/2}(x) - \widehat{A}_N^{1/2}(y)\| \leq C(\|x-y\|), \tag{103}$$

$$\|\widehat{b}_N(x)\|^2 + \|\widehat{A}_N^{1/2}(x)\|^2 \leq K(1+\|x\|^2). \tag{104}$$

Consequently, for every finite time horizon $T > 0$, the IMD process from Eq. (10) has finite moments:

$$\sup_r \sup_{t\in[0,T]} \mathbb{E}\|(X_t^N)^r\|^2 < \infty, \qquad \sup_{t\in[0,T]} \mathbb{E}\|X_t\|^2 < \infty. \tag{105}$$

G.2 SPLITTING THE TERMS

We now consider the Euler-Maruyama discretization of the $\mathcal{L}$-diffusion from Eq. (8) with timestep $h > 0$, given by the difference equation

$$X_{k+1}^N := X_k^N + h\widehat{b}_N(X_k^N) + \sqrt{h}\widehat{A}_N^{1/2}(X_k^N)\xi_{\ell+1}, \tag{106}$$

where $\xi_{k+1} \sim \mathcal{N}(0, \mathbf{I}_D)$.

We recall the squared Wasserstein-2 distance [71],

$$W_2^2(\mu,\nu) := \inf_{\gamma\in\Gamma(\mu,\nu)} \mathbb{E}_{(x,y)}\left[\|x-y\|_2^2\right], \tag{107}$$

where $\Gamma(\mu,\nu)$ denotes the couplings of μ and ν. It is easy to see that for two random variables $X \sim \mu$, $Y \sim \nu$, we have that

$$W_2^2(\mu,\nu) \leq \mathbb{E}\left[\|X-Y\|_2^2\right] \tag{108}$$

by the optimality of the coupling γ in W_2.

We first introduce the intermediate continuous-time off-grid terms X_{kh}^N, and apply a triangle inequality on $W_2\left(\text{Law}(X_k^N), \text{Law}(X_t)\right)$, giving us

$$\begin{aligned} W_2\left(\text{Law}(X_k^N), \text{Law}(X_t)\right) \leq & \underbrace{W_2\left(\text{Law}(X_k^N), \text{Law}(X_{kh}^N)\right)}_{\text{E-M error}} + \underbrace{W_2\left(\text{Law}(X_{kh}^N), \text{Law}(X_t^N)\right)}_{\text{time discretization}} \\ & + \underbrace{W_2\left(\text{Law}(X_t^N), \text{Law}(X_t)\right)}_{\text{data approximation}}, \end{aligned} \tag{109}$$

which we will treat individually using different tools.

Data approximation We begin with the term labeled "data approximation", which may be asymptotically suppressed by Theorem 1. Indeed, the strong operator convergence ensures us that the data driven and off-grid processes will converge as $N \to \infty$.

Time discretization We now use Eq. (108) to upper bound the term by

$$W_2\left(\text{Law}(X_{kh}^N), \text{Law}(X_t^N)\right) \leq \left(\mathbb{E}\left[\|X_{kh}^N - X_t^N\|_2^2\right]\right)^{1/2}, \tag{110}$$

which is controlled under linear growth by *synchronous coupling*, i.e. both X_t^N and X_{kh}^N are driven by the same Brownian motion B_τ [19]. Without loss of generality, let $0 < s < t$ and consider the difference

$$X_s^N - X_t^N = \int_s^t \widehat{b}_N(X_\tau^N)d\tau + \int_s^t \widehat{A}_N^{1/2}(X_\tau^N)dB_\tau. \tag{111}$$

Plugging this into the second term of Eq. (109), we get

$$\begin{aligned}\left(\mathbb{E}\left[\|X_t^N - X_s^N\|_2^2\right]\right)^{1/2} &\leq \left(\mathbb{E}\left[\|\int_s^t \widehat{b}_N(X_\tau^N)d\tau\|_2^2\right]\right)^{1/2} + \left(\mathbb{E}\left[\|\int_s^t \widehat{A}_N^{1/2}(X_\tau^N)dB_\tau\|_2^2\right]\right)^{1/2} \\ &\overset{a)}{\leq} \left(\mathbb{E}\left[\sup_{\tau\in(s,t)}\|\widehat{b}_N(X_\tau^N)\|_2^2\right]^{1/2} + \mathbb{E}\left[\sup_{\tau\in(s,t)}\|\widehat{A}_N^{1/2}(X_\tau^N)\|_2^2\right]^{1/2}\right)\sqrt{|t-s|} \\ &\leq C_t\sqrt{|t-s|}, \end{aligned} \tag{112}$$

where we use the maximum principle in $a)$, and bound it with a constant C_t.

Thus, it follows by plugging $t \in (kh, (k+1)h)$ into Eqs. (110) and (112) that

$$W_2^2(\text{Law}(X_{kh}^N), \text{Law}(X_t^N)) \in \mathcal{O}(h). \tag{113}$$

Thus, by choosing an appropriate step size $h \to 0$ as $N \to \infty$, we obtain the desired convergence.

Euler-Maruyama error Finally, considering the regularity of Assumption 2, classical results of Euler-Maruyama schemes [19; 40] yield the bound

$$W_2\left(\text{Law}(X_k^N), \text{Law}(X_{kh}^N)\right) \leq \mathcal{O}(h^{1/2}) \tag{114}$$

where in $a)$ we use the bound from Eq. (108) on the two first terms. Notice that said terms purely pertain to discretization and "off-grid" errors of $\hat{X}$, whereas the last term is purely continuous, but compares the transportation cost from the $\mathcal{L}$-diffusion to the data-driven diffusion.

By taking $N \to \infty$, $\epsilon \to 0$ as in Lemma 2 and $h \to 0$, we combine all previous upper bounds and plug into Eq. (109) to obtain

$$W_2\left(\text{Law}(X_k^N), \text{Law}(X_t)\right) \in \mathcal{O}(h^{1/2}), \tag{115}$$

thus proving the claim.

□