On the universal structure in deterministic generative diffusion

Defang Chen[1] Zhenyu Zhou[2] Can Wang[2] Zeyu Zheng[1] Siwei Lyu[3]
[1]UC Berkeley [2]Zhejiang University [3]University at Buffalo

{defchern,zyzheng}@berkeley.edu zhyzhou@zju.edu.cn

Abstract

Diffusion-based generative modeling has emerged as a general computational framework for simulating high-dimensional, structured data across various domains. Understanding its generative mechanisms is paramount for advancing model training and inference in a principled way. Here, we demonstrate that deterministic sampling trajectories share a universal geometric structure invariant to data modalities and diffusion model implementations. This structure can be largely explained by an exactly solvable model based on the optimal linear denoiser.

1 Introduction

A recurring theme in scientific inquiry is that systems of immense complexity often exhibit unexpectedly simple regularities. For example, Kepler's three laws first reduced intricate planetary trajectories to simple geometric form, and Newtonian mechanics then unified celestial and terrestrial dynamics under the same laws of motion. Identifying such universal principles, which transcend the details of individual instances, deepens our understanding and often yields predictive or practical power. Today, this paradigm extends beyond the natural world to artificial neural systems, as exemplified by *double descent* [Belkin et al., 2019] and *scaling laws* [Hestness et al., 2017, Bahri et al., 2024].

Modern generative diffusion models, rooted in nonequilibrium thermodynamics [Jarzynski, 1997, Sohl-Dickstein et al., 2015], are among the most successful AI systems to date [Song et al., 2021]. Their power lies in overcoming a long-standing obstacle in generative modeling, namely the difficulty of representing and sampling from distributions that are at once high-dimensional and richly structured. Instead of modeling the data distribution directly, diffusion models amortize computation over iterative denoising steps that begin from an uninformative Gaussian prior. The denoising direction that guides the data-generating process is learned from the predefined diffusion process that perturbs data into noise. This framework now extends from realistic multimedia synthesis to pressing scientific challenges.

Considering the tremendous complexity of diffusion models and the heterogeneity of the data on which they are trained, it is surprising to us that their *deterministic sampling trajectories follow a common geometric pattern*. Our previous studies identified such a strong regularity in the image domain, covering both unconditional and conditional generation in data and latent spaces [Chen et al., 2024, 2025]. Whether this pattern reflects the statistics of natural images or is a more general feature of generative diffusion has remained unclear. Here, we confirm its universality through

An early version of this manuscript (April 2026) is available here; the present version was reorganized with the help of Claude Opus 5.5. Code is available here.

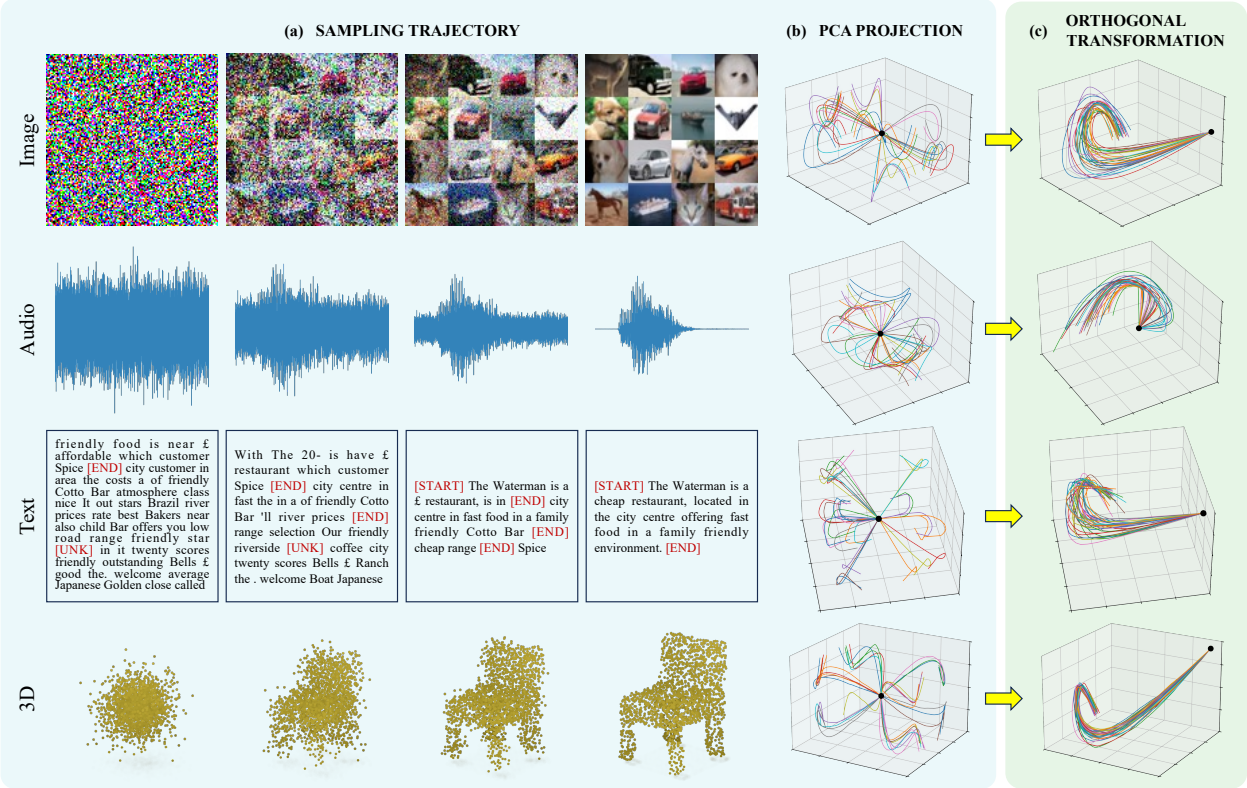


Figure 1: Deterministic sampling trajectories share a universal structure across data modalities and diffusion model implementations. (a) Intermediate states along sampling trajectories, from random noise (left) to generated data (right). (b) Each trajectory projected onto its three leading principal components, with all starting points aligned (black point). (c) The trajectories aligned through an orthogonal transformation.

experiments spanning a diverse collection of data modalities and diffusion model implementations (Figure 1). We further show that a simple hyperbolic structure accounts for the dominant trajectory behaviors observed in practice (Figure 2). This geometric structure naturally arises in an exactly solvable model built on the linear minimum mean-square-error (LMMSE) denoiser, which coincides at every noise level with the Bayes-optimal denoiser for a Gaussian with the same mean and covariance as the data. We call this model the Gaussian flow and discuss its asymptotic trajectory congruence in high dimensions.

2 Results

We analyze sampling trajectories across fundamentally different data modalities, namely images [Karras et al., 2022], audio [Kong et al., 2021], text [Li et al., 2022], and 3D point clouds [Zhou et al., 2021], each with a representative implementation. Details are provided in the Supplementary material. Figure 1(a) visualizes several intermediate states of the deterministic sampling trajectories in data spaces with thousands of dimensions, where various data structures emerge gradually from random noise. We then project each randomly initialized trajectory onto its three leading principal components (Figure 1(b)). The first component alone captures more than 99% of the trajectory variance for all four models, i.e., each trajectory is nearly a straight line from noise toward data.

However, a straight line cannot describe how the data structure emerges along the trajectory; this information lies in the small deviation from the line, which traces a characteristic curve in the second and third components. We further align each trajectory with an arbitrarily chosen reference trajectory through an orthogonal transformation. As shown in Figure 1(c), the aligned trajectories share approximately the same geometric shape, which persists across heterogeneous data modalities and model implementations. Such persistence makes it unlikely that these regularities originate in the statistics of any particular dataset or in the design of any particular model implementation. They point instead to a general mechanism intrinsic to how generative diffusion transports distributions. We next take a step toward understanding this universal structure by investigating an exactly solvable model that yields closed-form solutions for deterministic sampling trajectories.

2.1 Gaussian reference frame

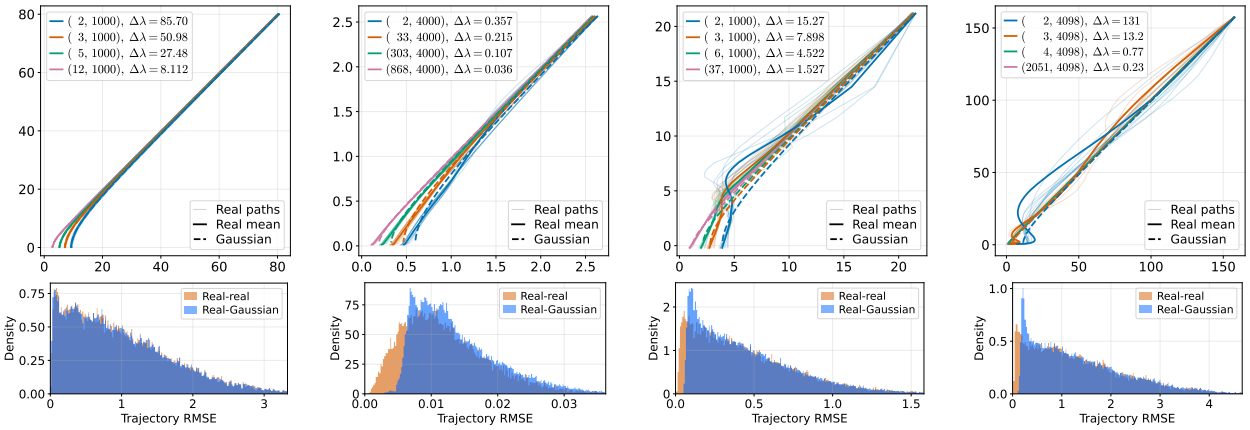
Without loss of generality, we consider the simplest forward diffusion process $\mathrm{d}\mathbf{x}_t = \sqrt{2t}\,\mathrm{d}\mathbf{w}_t$. Here, $\mathbf{x}_0 \sim p_{\text{data}}$ is a data sample from a centered data distribution in $\mathbb{R}^d$, and $\mathbf{w}_t$ is a standard Wiener process. The sample is corrupted by isotropic Gaussian noise of standard deviation t, i.e., $\mathbf{x}_t \mid \mathbf{x}_0 \sim \mathcal{N}(\mathbf{x}_0, t^2\mathbf{I}_d)$. At a sufficiently large time T, the noise dominates the data structure, and $\mathbf{x}_T$ approximately follows $\mathcal{N}(\mathbf{0}, T^2\mathbf{I}_d)$. The (real) sampling trajectory is generated by solving the deterministic probability flow $\mathrm{d}\mathbf{x}_t = (1/t)(\mathbf{x}_t - \mathbb{E}(\mathbf{x}_0 \mid \mathbf{x}_t))\,\mathrm{d}t$ backward, where the conditional expectation is estimated using a deep neural network (the denoiser) trained on a large-scale dataset. If we instead estimate $\mathbb{E}(\mathbf{x}_0 \mid \mathbf{x}_t)$ with the LMMSE denoiser, or equivalently model the data as Gaussian, we have a deterministic Gaussian flow $\mathrm{d}\bar{\mathbf{x}}_t = t(\mathbf{\Sigma} + t^2\mathbf{I}_d)^{-1}\,\bar{\mathbf{x}}_t\,\mathrm{d}t$. Here, $\mathbf{\Sigma}$ denotes the data covariance matrix and admits the spectral decomposition $\mathbf{\Sigma} = \mathbf{U}\mathbf{\Lambda}\mathbf{U}^\top$, where the diagonal matrix $\mathbf{\Lambda} = \mathrm{diag}(\lambda_1, \ldots, \lambda_d)$ contains the non-negative eigenvalues sorted in descending order, and the orthogonal matrix $\mathbf{U} = [\mathbf{u}_1, \ldots, \mathbf{u}_d] \in \mathbb{R}^{d\times d}$ contains the corresponding eigenvectors.

One interesting property of such an exactly solvable model is that its induced Gaussian sampling trajectory has an invariant hyperbolic structure irrespective of the initial noise. In particular, the solution in the spectral representation ($\bar{\mathbf{y}}_t = \mathbf{U}^\top\bar{\mathbf{x}}_t$) evolves independently across dimensions and decouples the deterministic and stochastic components: the i-th coordinate $\bar{y}_{t,i} = \alpha_{t,i}\,\tilde{y}_{T,i}$ tracks the restoration of data features with a deterministic amplitude $\alpha_{t,i} \coloneqq \sqrt{\lambda_i + t^2}$ across different realizations, and the initial stochasticity $\tilde{y}_{T,i}$, which approximately follows $\mathcal{N}(0,1)$, acts as a scaling factor for each coordinate. Furthermore, projecting the trajectory onto the plane spanned by $\mathbf{u}_i$ and $\mathbf{u}_j$ with non-zero initial conditions ($\tilde{y}_{T,i}\,\tilde{y}_{T,j} \neq 0$) yields a hyperbolic segment

$$(\bar{y}_{t,i}/\tilde{y}_{T,i})^2 - (\bar{y}_{t,j}/\tilde{y}_{T,j})^2 = \alpha_{t,i}^2 - \alpha_{t,j}^2 = \lambda_i - \lambda_j, \tag{1}$$

with $t \in [0, T]$. Geometrically, each relation (Eq. 1) defines a hyperbolic cylinder in $\mathbb{R}^d$, and the intersection of these $(d-1)$ hyperbolic cylinders (if non-degenerate) confines the generative dynamics to a one-dimensional trajectory $\bar{\mathbf{y}}_t = (\mathbf{\Lambda}+t^2\mathbf{I}_d)^{1/2}\tilde{\mathbf{y}}_T$. The trajectory's curvature vanishes as $t \to \infty$, and its leading principal component already captures all but a fraction $\mathcal{O}(T^{-3})$ of its variance (see the Supplementary material). This invariant structure makes the Gaussian flow suitable as a yardstick to measure the trajectory variation and the nonlinear deviation.

In this Gaussian reference frame, each real trajectory closely follows these hyperbolas when the eigenvalue gap is moderate (Figure 2, top row). This comparison, like the "After" columns of the table, examines the trajectory shape rather than its evolution along the original directions ("Before" columns). After the transformation, each real trajectory deviates on average less from its Gaussian counterpart than from the other three references, and about as much from another Gaussian trajectory as from another real one (Figure 2, bottom row).



(a) Image. (b) Audio. (c) Text. (d) 3D.

	Gaussian chord		Brownian bridge		Isotropic Gaussian flow		Gaussian flow	
	Before	After	Before	After	Before	After	Before	After
Image [Karras et al., 2022]	1	0.6157	1.4136	0.7535	1.5703	0.9960	0.5600	**0.1148**
Audio [Kong et al., 2021]	1	0.3736	1.4147	0.7479	0.5031	0.3039	0.5731	**0.1581**
Text [Li et al., 2022]	1	0.6241	1.4148	0.7610	0.7669	0.2941	0.7122	**0.1780**
3D [Zhou et al., 2021]	1	0.9047	1.4138	1.0307	0.7361	0.4189	0.5873	**0.3419**

Figure 2: The real trajectories, after an orthogonal transformation, closely follow the Gaussian hyperbolas for most eigenmode pairs (top row; legend: (i, j) and the eigenvalue gap $\Delta\lambda = \lambda_i - \lambda_j$), except at the largest $\Delta\lambda$. The bottom row shows the distributions of the shape deviation, measured by the root-mean-square error (RMSE), between two independently initialized real trajectories and between a real trajectory and the Gaussian counterpart of another such trajectory. The table reports the deviation of each reference from the real trajectory before and after the transformation, normalized by the Gaussian chord deviation before the transformation. The Gaussian chord is the straight line from the initial noise to the endpoint of the Gaussian flow, the Brownian bridge is pinned to the same two endpoints, and the isotropic Gaussian flow replaces $\mathbf{\Sigma}$ with the scaled identity $(\operatorname{tr}\mathbf{\Sigma}/d)\,\mathbf{I}_d$ (see the Supplementary material).

Furthermore, Gaussian sampling trajectories starting from distinct initial states are approximately congruent (equivalent under orthogonal transformations) in high dimensions. The *relative fluctuation* of the Gram kernel $G(t, s) = \bar{\mathbf{y}}_t^\top \bar{\mathbf{y}}_s$ quantifies how far a single realization deviates from its expectation over the initial noise:

$$\frac{\sqrt{\operatorname{Var}\left(G(t,s)\right)}}{\mathbb{E}\left[G(t,s)\right]} = \frac{\sqrt{2\sum_{i=1}^{d}(\lambda_i + t^2)(\lambda_i + s^2)}}{\sum_{i=1}^{d}\sqrt{(\lambda_i + t^2)(\lambda_i + s^2)}}. \tag{2}$$

The fluctuation depends on the data eigenspectrum and the noise levels. Along the diagonal $t = s$, it gradually transitions from $\mathcal{O}(d^{-1/2})$ to $\mathcal{O}(r_{\text{eff}}^{-1/2})$ as the reverse process travels from pure noise to the data end. Here, $r_{\text{eff}} := (\operatorname{tr}\mathbf{\Lambda})^2/\operatorname{tr}(\mathbf{\Lambda}^2)$ is the effective data dimension. For a power-law eigenspectrum $\lambda_i \propto i^{-\alpha}$, as is typical of natural images, $r_{\text{eff}}^{-1/2}$ vanishes as $d \to \infty$ when $\alpha \leq 1$ and otherwise tends to a positive constant (see the Supplementary material). Although the fluctuation increases near the data end, rendering the trajectory geometry slightly more varied, the overall structural congruence remains robust.

The mathematical formulation of Gaussian flows has previously been adopted to connect diffusion models with optimal transport [Khrulkov et al., 2023], to derive Wasserstein errors of numerical samplers [Pierret and Galerne, 2025], and to illustrate the score decomposition of low-dimensional data [Chen et al., 2023], but none of them considers the universal geometric shape or its congruence studied in this paper.

3 Conclusion

Elucidating the geometric structure of generative diffusion dynamics is fundamental to explaining empirical observations and developing principled techniques for model training and inference. We find that deterministic sampling trajectories share approximately the same geometric shape across data modalities and model implementations, beyond the image domain of our previous studies [Chen et al., 2024, 2025]. Using the Gaussian flow built on the LMMSE denoiser, we rigorously characterize the hyperbolic structure of its sampling trajectories in high dimensions and show that this structure approximately recovers the observed shape. Our findings open new avenues for the design of diffusion-based generative models, including the identification of optimal sampling configurations and the acceleration of model training through prior knowledge of the trajectory structure.

4 Author contributions

D.C.: Conceptualization, Methodology, Formal analysis, Writing – original draft; Z. Zhou: Software, Investigation, Visualization, Writing – review & editing; C.W., Z. Zheng, S.L.: Resources, Supervision, Writing – review & editing.

References

Yasaman Bahri, Ethan Dyer, Jared Kaplan, Jaehoon Lee, and Utkarsh Sharma. Explaining neural scaling laws. *Proceedings of the National Academy of Sciences*, 121(27):e2311878121, 2024.

Mikhail Belkin, Daniel Hsu, Siyuan Ma, and Soumik Mandal. Reconciling modern machine-learning practice and the classical bias–variance trade-off. *Proceedings of the National Academy of Sciences*, 116(32):15849–15854, 2019.

Defang Chen, Zhenyu Zhou, Can Wang, Chunhua Shen, and Siwei Lyu. On the trajectory regularity of ODE-based diffusion sampling. In *International Conference on Machine Learning*, volume 235, pages 7905–7934. PMLR, 2024.

Defang Chen, Zhenyu Zhou, Can Wang, and Siwei Lyu. Geometric regularity in deterministic sampling dynamics of diffusion-based generative models. *Journal of Statistical Mechanics: Theory and Experiment*, 2025(12):124002, 2025.

Minshuo Chen, Kaixuan Huang, Tuo Zhao, and Mengdi Wang. Score approximation, estimation and distribution recovery of diffusion models on low-dimensional data. In *International Conference on Machine Learning*, volume 202, pages 4672–4712. PMLR, 2023.

Joel Hestness, Sharan Narang, Newsha Ardalani, Gregory Diamos, Heewoo Jun, Hassan Kianinejad, Md Mostofa Ali Patwary, Yang Yang, and Yanqi Zhou. Deep learning scaling is predictable, empirically. *arXiv preprint arXiv:1712.00409*, 2017.

Christopher Jarzynski. Equilibrium free-energy differences from nonequilibrium measurements: A master-equation approach. *Physical Review E*, 56(5):5018, 1997.

Tero Karras, Miika Aittala, Timo Aila, and Samuli Laine. Elucidating the design space of diffusion-based generative models. In *Advances in Neural Information Processing Systems*, volume 35, pages 26565–26577, 2022.

Valentin Khrulkov, Gleb Ryzhakov, Andrei Chertkov, and Ivan Oseledets. Understanding DDPM latent codes through optimal transport. In *International Conference on Learning Representations*, 2023.

Zhifeng Kong, Wei Ping, Jiaji Huang, Kexin Zhao, and Bryan Catanzaro. DiffWave: A versatile diffusion model for audio synthesis. In *International Conference on Learning Representations*, 2021.

Xiang Li, John Thickstun, Ishaan Gulrajani, Percy S Liang, and Tatsunori B Hashimoto. Diffusion-LM improves controllable text generation. In *Advances in Neural Information Processing Systems*, volume 35, pages 4328–4343, 2022.

Emile Pierret and Bruno Galerne. Diffusion models for Gaussian distributions: Exact solutions and Wasserstein errors. In *International Conference on Machine Learning*, volume 267, pages 49355–49381, 2025.

Jascha Sohl-Dickstein, Eric Weiss, Niru Maheswaranathan, and Surya Ganguli. Deep unsupervised learning using nonequilibrium thermodynamics. In *International Conference on Machine Learning*, volume 37, pages 2256–2265. PMLR, 2015.

Yang Song, Jascha Sohl-Dickstein, Diederik P. Kingma, Abhishek Kumar, Stefano Ermon, and Ben Poole. Score-based generative modeling through stochastic differential equations. In *International Conference on Learning Representations*, 2021.

Linqi Zhou, Yilun Du, and Jiajun Wu. 3D shape generation and completion through point-voxel diffusion. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 5826–5835, 2021.

A Supplementary material

A.1 Generative diffusion dynamics

We consider a diffusion process $\{\mathbf{x}_t\}_{t\in[0,T]}$ in $\mathbb{R}^d$, governed by the linear stochastic differential equation $\mathrm{d}\mathbf{x}_t = f(t)\mathbf{x}_t\mathrm{d}t + g(t)\,\mathrm{d}\mathbf{w}_t$, where $\mathbf{w}_t$ is the Wiener process, and $f(t)$ and $g(t)$ are the drift and diffusion coefficients, respectively. The initial condition $\mathbf{x}_0$ is drawn from a data distribution p_0, which can be, for example, a distribution of images, audio, or text. The marginal density p_t evolves according to the Fokker–Planck equation $\partial p_t(\mathbf{x}_t)/\partial t = -\nabla\cdot[f(t)\mathbf{x}_t p_t(\mathbf{x}_t) - (g^2(t)/2)\nabla_{\mathbf{x}_t} p_t(\mathbf{x}_t)]$. As $t \to T$, the data distribution is smoothly transformed into an approximately Gaussian distribution for suitably chosen $f(t)$ and $g(t)$. This stochastic diffusion process can be reversed with the (Stein) score function $\nabla_{\mathbf{x}_t}\log p_t(\mathbf{x}_t)$, i.e., the gradient of the log marginal density with respect to $\mathbf{x}_t$. The property that makes fast data generation possible is that the same marginal evolution can be reproduced by a *deterministic* process characterized by the probability flow ordinary differential equation,

$$\mathrm{d}\mathbf{x}_t = \left[f(t)\mathbf{x}_t - \frac{1}{2}g^2(t)\nabla_{\mathbf{x}_t}\log p_t(\mathbf{x}_t)\right]\mathrm{d}t. \tag{3}$$

Integrating Eq. 3 backward in time from initial Gaussian noise yields the synthesized datum and the deterministic sampling trajectory, whose geometry is the central object of this paper.

This formulation underlies most diffusion- and flow-based models, which use various designs of the drift and diffusion coefficients, as well as iterative samplers based on numerical solvers that discretize Eq. 3 into reliable finite-step sampling trajectories. In practice, the score function is estimated using a deep neural network trained on a large-scale dataset with denoising score matching or equivalent objectives [Song et al., 2021]. More details are provided in our previous study [Chen et al., 2025].

A.2 Gaussian flows

We collect the notation of the deterministic Gaussian flow used in the main text. The data have mean $\boldsymbol{\mu} = \mathbf{0}$ and covariance matrix $\boldsymbol{\Sigma} \in \mathbb{R}^{d\times d}$, which is positive semi-definite with the spectral decomposition $\boldsymbol{\Sigma} = \mathbf{U}\boldsymbol{\Lambda}\mathbf{U}^\top$, where $\boldsymbol{\Lambda} = \mathrm{diag}(\lambda_1,\ldots,\lambda_d)$ with $\lambda_1 \geq \cdots \geq \lambda_d \geq 0$, and $\mathbf{U} = [\mathbf{u}_1,\ldots,\mathbf{u}_d] \in \mathbb{R}^{d\times d}$ is orthogonal. Starting from an initial state $\bar{\mathbf{x}}_T$, the deterministic Gaussian flow $\mathrm{d}\bar{\mathbf{x}}_t = t(\boldsymbol{\Sigma} + t^2\mathbf{I}_d)^{-1}\bar{\mathbf{x}}_t\,\mathrm{d}t$ has the closed-form solution

$$\bar{\mathbf{x}}_t = \left(\boldsymbol{\Sigma} + t^2\mathbf{I}_d\right)^{1/2}\left(\boldsymbol{\Sigma} + T^2\mathbf{I}_d\right)^{-1/2}\bar{\mathbf{x}}_T, \quad t\in[0,T]. \tag{4}$$

For data with a nonzero mean $\boldsymbol{\mu}$, as in our experiments, Eq. 4 applies to $\bar{\mathbf{x}}_t - \boldsymbol{\mu}$ and $\bar{\mathbf{x}}_T - \boldsymbol{\mu}$. In the eigenbasis, $\bar{\mathbf{y}}_t = \mathbf{U}^\top\bar{\mathbf{x}}_t$ evolves coordinate-wise as $\bar{y}_{t,i} = \alpha_{t,i}\,\bar{y}_{T,i}$, where $\alpha_{t,i} := \sqrt{\lambda_i + t^2}$ and

$\tilde{y}_{T,i} \coloneqq \bar{y}_{T,i}/\alpha_{T,i}$. When $\bar{\mathbf{x}}_T \sim \mathcal{N}(\mathbf{0}, \mathbf{\Sigma} + T^2\mathbf{I}_d)$[1], we have $\tilde{\mathbf{y}}_T \sim \mathcal{N}(\mathbf{0}, \mathbf{I}_d)$. The trajectory deviates only slightly from a straight line, since all coordinates share the common linear growth, as captured by the Taylor expansion $\bar{\mathbf{y}}_t \approx (t\mathbf{I}_d + \mathbf{\Lambda}/(2t))\tilde{\mathbf{y}}_T$ for $t \to \infty$.

A.2.1 Straight-line decomposition

We split the deterministic amplitude $\alpha_{t,i}$ into a common linear part and a bounded correction, $\alpha_{t,i} = t + \rho_{t,i}$, with

$$\rho_{t,i} \coloneqq \sqrt{\lambda_i + t^2} - t = \frac{\lambda_i}{\sqrt{\lambda_i + t^2} + t}. \tag{5}$$

The correction $\rho_{t,i}$ has three properties: (i) it decreases in t from its maximum $\rho_{0,i} = \sqrt{\lambda_i}$ at the data end, since $\partial\rho_{t,i}/\partial t = t/\sqrt{\lambda_i + t^2} - 1 \leq 0$, and thus $\rho_{t,i} \in [0, \sqrt{\lambda_i}]$; (ii) $\rho_{t,i} \leq \lambda_i/(2t)$ and $\rho_{t,i} = \lambda_i/(2t) + \mathcal{O}(t^{-3})$ for $t \gg \sqrt{\lambda_i}$; (iii) it increases in λ_i, since $\partial\rho_{t,i}/\partial\lambda_i = 1/(2\sqrt{\lambda_i + t^2}) > 0$, and vanishes when $\lambda_i = 0$.

The same split holds for the matrix square root in Eq. 4, i.e., $\left(\mathbf{\Sigma} + t^2\mathbf{I}_d\right)^{1/2} = t\,\mathbf{I}_d + \mathbf{R}_t$, where $\mathbf{R}_t \coloneqq \mathbf{U}\,\mathrm{diag}\left(\rho_{t,1}, \ldots, \rho_{t,d}\right)\mathbf{U}^\top$ is positive semi-definite and shares the eigenvectors of $\mathbf{\Sigma}$. Substituting this into Eq. 4 gives

$$\bar{\mathbf{x}}_t = (t\,\mathbf{I}_d + \mathbf{R}_t)\left(\mathbf{\Sigma} + T^2\mathbf{I}_d\right)^{-1/2}\bar{\mathbf{x}}_T = t\,\mathbf{a} + \mathbf{r}_t, \tag{6}$$

where

$$\mathbf{a} \coloneqq \left(\mathbf{\Sigma} + T^2\mathbf{I}_d\right)^{-1/2}\bar{\mathbf{x}}_T = \mathbf{U}\tilde{\mathbf{y}}_T, \quad \mathbf{r}_t \coloneqq \mathbf{R}_t\mathbf{a}. \tag{7}$$

The straight part $t\mathbf{a}$ thus depends only on the initial noise and vanishes at $t = 0$, i.e., the line passes through the data mean (the origin), so the generated structure is carried entirely by the deviation $\mathbf{r}_t$, to which directions with $\lambda_i = 0$ do not contribute (property (iii)). Away from the data end, this deviation is small relative to the straight part: by property (iii), $\mathbf{R}_t$ has the largest eigenvalue $\rho_{t,1}$, so $\|\mathbf{r}_t\|_2 \leq \rho_{t,1}\|\mathbf{a}\|_2$, and property (ii) gives

$$\frac{\|\mathbf{r}_t\|_2}{\|t\,\mathbf{a}\|_2} \leq \frac{\rho_{t,1}}{t} \leq \frac{\lambda_1}{2t^2}, \tag{8}$$

which is below one for $t > \sqrt{\lambda_1/2}$.

Temporal statistics. Let ν be a probability measure on $[0, T]$ that weights the time points of the trajectory, e.g., the uniform distribution for the continuous-time trajectory or the empirical distribution of the time steps of a numerical sampler. We write $\mathbb{E}_t$, Var_t, and Cov_t for the mean, variance, and covariance over $t \sim \nu$. The principal components of the trajectory are the eigenvectors of its temporal covariance

$$\mathbf{\Sigma}_{\text{path}} \coloneqq \mathrm{Cov}_t(\bar{\mathbf{x}}_t) = \mathbb{E}_t\left[(\bar{\mathbf{x}}_t - \mathbb{E}_t\bar{\mathbf{x}}_t)(\bar{\mathbf{x}}_t - \mathbb{E}_t\bar{\mathbf{x}}_t)^\top\right], \tag{9}$$

with eigenvalues $\mu_1 \geq \cdots \geq \mu_d \geq 0$. We measure the fraction of the trajectory variance not captured by its k leading principal components by $\varepsilon_k \coloneqq \sum_{j>k}\mu_j / \sum_{j=1}^{d}\mu_j$.

[1] In practice, we initialize the reverse process from the isotropic noise distribution $\mathcal{N}(\mathbf{0}, T^2\mathbf{I}_d)$, which accurately approximates p_T provided T is sufficiently large to dominate the data structure, and this determinism-stochasticity dichotomy still holds.

Proposition 1. *For any positive semi-definite $\mathbf{\Sigma}$, any initial state with $\tilde{\mathbf{y}}_T \neq \mathbf{0}$, and any probability measure ν on $[0,T]$ with $\mathrm{Var}_t(t) > 0$, define*

$$\delta := \frac{\mathbb{E}_t\|\mathbf{r}_t\|_2^2}{\mathrm{Var}_t(t)\,\|\mathbf{a}\|_2^2} = \frac{\sum_{i=1}^d \tilde{y}_{T,i}^2\,\mathbb{E}_t\left[\rho_{t,i}^2\right]}{\mathrm{Var}_t(t)\sum_{i=1}^d \tilde{y}_{T,i}^2}. \tag{10}$$

(a) If $\delta < 1$, then $\varepsilon_2 \le \varepsilon_1 \le \delta/(1-\sqrt{\delta})^2$. (b) If ν is the uniform distribution on $[0,T]$, then

$$\delta \le 8\,\frac{\sum_{i=1}^d \tilde{y}_{T,i}^2\left(\lambda_i/T^2\right)^{3/2}}{\sum_{i=1}^d \tilde{y}_{T,i}^2} \le 8\left(\frac{\lambda_1}{T^2}\right)^{3/2}. \tag{11}$$

Hence, $\varepsilon_2 \le \varepsilon_1 = \mathcal{O}(T^{-3})$ as $T \to \infty$.

Proof. *(a)* Let $\hat{\mathbf{a}} = \mathbf{a}/\|\mathbf{a}\|_2$ and $\mathbf{P} = \mathbf{I}_d - \hat{\mathbf{a}}\hat{\mathbf{a}}^\top$. By the variational characterization of the leading eigenvalue, $\mu_1 \ge \hat{\mathbf{a}}^\top \mathbf{\Sigma}_{\mathrm{path}}\hat{\mathbf{a}}$, and thus

$$\begin{aligned}\sum_{j=2}^d \mu_j &= \mathrm{tr}(\mathbf{\Sigma}_{\mathrm{path}}) - \mu_1 \le \mathrm{tr}\left(\mathbf{P}\mathbf{\Sigma}_{\mathrm{path}}\mathbf{P}\right)\\ &= \mathbb{E}_t\left\|\mathbf{P}(\mathbf{r}_t - \mathbb{E}_t\mathbf{r}_t)\right\|_2^2 \le \mathbb{E}_t\|\mathbf{r}_t\|_2^2,\end{aligned} \tag{12}$$

where the equality uses $\mathbf{P}\bar{\mathbf{x}}_t = \mathbf{P}\mathbf{r}_t$ (Eq. 6), and the last step uses that $\mathbf{P}$ is an orthogonal projection and that $\mathbb{E}_t\|\mathbf{r}_t - \mathbb{E}_t\mathbf{r}_t\|_2^2 = \mathbb{E}_t\|\mathbf{r}_t\|_2^2 - \|\mathbb{E}_t\mathbf{r}_t\|_2^2$. For the total variance, the triangle inequality in $L^2(\nu)$ gives

$$\begin{aligned}\sqrt{\mathrm{tr}(\mathbf{\Sigma}_{\mathrm{path}})} &= \left(\mathbb{E}_t\left\|(t - \mathbb{E}_t t)\,\mathbf{a} + (\mathbf{r}_t - \mathbb{E}_t\mathbf{r}_t)\right\|_2^2\right)^{1/2}\\ &\ge \sqrt{\mathrm{Var}_t(t)}\,\|\mathbf{a}\|_2 - \sqrt{\mathbb{E}_t\|\mathbf{r}_t\|_2^2}\\ &= \left(1 - \sqrt{\delta}\right)\sqrt{\mathrm{Var}_t(t)}\,\|\mathbf{a}\|_2,\end{aligned} \tag{13}$$

which is positive when $\delta < 1$. Dividing the first bound by the square of the second yields $\varepsilon_1 \le \delta/(1-\sqrt{\delta})^2$, and $\varepsilon_2 \le \varepsilon_1$ holds since $\mu_2 \ge 0$.

(b) For the uniform distribution, $\mathrm{Var}_t(t) = T^2/12$. The function $F(t) = \lambda_i t + \frac{2}{3}t^3 - \frac{2}{3}(\lambda_i + t^2)^{3/2}$ satisfies $F'(t) = \rho_{t,i}^2$, $F(0) = -\frac{2}{3}\lambda_i^{3/2}$, and $F(t) \to 0$ as $t \to \infty$. Therefore,

$$\mathbb{E}_t\left[\rho_{t,i}^2\right] = \frac{1}{T}\int_0^T \rho_{t,i}^2\,\mathrm{d}t \le \frac{1}{T}\int_0^\infty \rho_{t,i}^2\,\mathrm{d}t = \frac{2\lambda_i^{3/2}}{3T}. \tag{14}$$

Substituting this bound and $\mathrm{Var}_t(t) = T^2/12$ into δ gives the first inequality, and $\lambda_i \le \lambda_1$ gives the second. Since $\delta = \mathcal{O}(T^{-3})$, part (a) yields $\varepsilon_1 = \mathcal{O}(T^{-3})$. □

Proposition 1 holds for every realization of the initial noise and requires no low-rank assumption on $\mathbf{\Sigma}$. Part (a) also applies to a principal component analysis that weights every saved state equally, in which case ν is the empirical distribution of their time steps, so that δ, and hence the bound on ε_1, depends on the schedule. A schedule that places most of its steps near the data end, such as that of EDM [Karras et al., 2022], gives more weight to small t, where $\rho_{t,i}$ is largest (property (i)), and increases the numerator of δ. Weighting each state by its step size instead makes ν nearly uniform, as in part (b).

A.3 Implementation details

A.3.1 Models and common protocol

We consider four data modalities and use a representative unconditional diffusion model for each: images generated by EDM [Karras et al., 2022] on CIFAR-10[2], raw audio waveforms generated by DiffWave [Kong et al., 2021] on SC09[3], text generated by Diffusion-LM [Li et al., 2022] on E2E[4], and 3D point clouds generated by PVD [Zhou et al., 2021] on the Chair category of ShapeNet[5]. We use the official implementations and their publicly released checkpoints[6,7,8,9]. For each model, we generate 200 sampling trajectories using a deterministic 200-step Euler sampler with its default noise schedule, and transform them into variance-exploding (VE) coordinates, i.e., the coordinates of the drift-free diffusion. The terminal noise levels are $T = 80$, 2.56, 21.2, and 157 for EDM, DiffWave, Diffusion-LM, and PVD, respectively. The effective data dimension $r_{\text{eff}} \coloneqq (\operatorname{tr} \mathbf{\Lambda})^2 / \operatorname{tr}(\mathbf{\Lambda}^2)$ is 9.2 for CIFAR-10, 1136.5 for SC09, 48.1 for E2E, and 83.8 for ShapeNet Chair.

- CIFAR-10 contains 50,000 RGB training images of resolution 32×32 in 10 classes, i.e., the data belong to $\mathbb{R}^{3\times32\times32}$. Following the EDM[10] preprocessing protocol, we linearly rescale pixel values to $[-1, 1]$ and estimate the mean and a rank-1,024 approximation of the covariance from the entire training set.

- SC09 contains the 38,908 utterances of the digits from "zero" to "nine", of which 31,158 utterances by 2,032 speakers form the training split [Kong et al., 2021]. Following the released DiffWave–SaShiMi[11,12] loader, we truncate or right-pad each utterance to 16,000 samples, i.e., the data belong to $\mathbb{R}^{1\times16000}$. We estimate the mean and a rank-4,096 approximation of the covariance from all 38,908 utterances, since the released loader also trains on all of them.

- E2E contains more than 50,000 crowdsourced restaurant descriptions, each paired with a meaning representation of 3 to 8 attributes, such as food type and price range. Following the Diffusion-LM[13] preprocessing protocol, we use only the 42,061 descriptions in the training split, add `START` and `END` tokens to each description, and truncate or right-pad it to 64 tokens. Each token ID is mapped through the checkpoint's word-embedding matrix to a 16-dimensional embedding, i.e., the data belong to $\mathbb{R}^{64\times16}$. We estimate the mean and the full covariance, of rank 1,024, and include the isotropic timestep-zero embedding noise used during training. During sampling, each predicted clean token embedding is mapped to its nearest vocabulary embedding before it enters the next update.

- ShapeNetCore, a subset of the ShapeNet repository of 3D CAD models, contains about 51,300 models in 55 categories. We use its Chair point clouds, each with 15,000 points sampled uniformly from the mesh surface, and the official split with 4,612 training shapes. Following

[2] https://www.cs.toronto.edu/~kriz/cifar.html
[3] http://download.tensorflow.org/data/speech_commands_v0.02.tar.gz
[4] https://github.com/tuetschek/e2e-dataset
[5] https://github.com/stevenygd/PointFlow
[6] https://nvlabs-fi-cdn.nvidia.com/edm/pretrained/edm-cifar10-32x32-uncond-vp.pkl
[7] https://github.com/albertfgu/diffwave-sashimi/blob/checkpoints/exp/unet_d128_n6_pool_2_expand2_ff2_T200_betaT0.02_uncond/checkpoint/800000.pkl
[8] https://drive.google.com/drive/folders/110CA22rwu_3EcllPYGhql0TnYeOBY77d (`ema_0.9999_200000.pt`)
[9] https://drive.google.com/drive/folders/1bwJ9_S8HW7oq1zNPW2lDLNUrKcW-_mRQ (`chair_1799.pth`)
[10] https://github.com/nvlabs/edm
[11] https://github.com/lmnt-com/diffwave
[12] https://github.com/albertfgu/diffwave-sashimi
[13] https://github.com/XiangLi1999/Diffusion-LM

the PVD[14] preprocessing protocol, we subtract the per-axis mean and divide by a single standard deviation, both computed over the training split, and keep the first 10,000 points of each shape for training. Each model input contains 2,048 points sampled uniformly with replacement from these 10,000 points, i.e., the data belong to $\mathbb{R}^{2048\times 3}$. We average over this random subsampling in closed form when estimating the mean and the covariance, which has full rank 6,144. When each point cloud is flattened coordinate by coordinate, the covariance is $\mathbf{W} \otimes \mathbf{I}_{2048} + \mathbf{B} \otimes \mathbf{1}\mathbf{1}^\top$, where $\otimes$ is the Kronecker product, $\mathbf{1} \in \mathbb{R}^{2048}$ is the all-ones vector, $\mathbf{W} \in \mathbb{R}^{3\times 3}$ is the mean within-shape covariance, and $\mathbf{B} \in \mathbb{R}^{3\times 3}$ is the covariance of the shape centroids. It has three common-point modes, in which all points move together, with the eigenvalues of $\mathbf{W} + 2048\,\mathbf{B}$, and three residual eigenspaces of dimension 2,047, in which the points move relative to their centroid, with the eigenvalues of $\mathbf{W}$. The predicted clean point cloud is not clipped.

For CIFAR-10 and SC09, the covariance is estimated by randomized principal component analysis with the stated rank. Directions outside the retained rank are given zero data variance, so the Gaussian flow contracts the initial noise along them toward the data mean in proportion to t. The isotropic Gaussian flow instead uses the trace of the full sample covariance.

A.3.2 Trajectory visualization

Sampling trajectories are hard to compare and describe precisely in the original space. Beyond the subspace projection of our previous studies [Chen et al., 2024, 2025], we introduce the Gaussian reference frame as a new visualization technique.

- **Subspace projection**. For each modality, we randomly select 32 of the 200 trajectories. Each real trajectory is translated to a common starting point and projected onto its own three leading temporal principal components, obtained by principal component analysis (PCA). Since independently estimated PCA bases have arbitrary orientations, we choose one real trajectory as the reference and align the others to it using the standard orthogonal Procrustes transformation fitted over the complete trajectory. Fitting this transformation in the three-dimensional PCA space or in the original space gives similar results.

- **Two-eigenmode projection**. Since the standardized Gaussian trajectory has an invariant hyperbolic structure irrespective of the initial noise, we align each real trajectory with its Gaussian counterpart, which starts from the same noise, through one ambient-space orthogonal Procrustes transformation fitted over the complete trajectory, with the common initial noise subtracted before the fit and added back afterwards. We then project the aligned trajectory onto two eigenvectors $\mathbf{u}_i$ and $\mathbf{u}_j$ of the data covariance and divide each coordinate by its initial stochasticity $\tilde{y}_{T,i}$ or $\tilde{y}_{T,j}$, so that the Gaussian flow (dashed) traces the hyperbola of Eq. 1, from near (T, T) at the initial noise to $(\sqrt{\lambda_i}, \sqrt{\lambda_j})$ at the data end. The residual of this alignment depends only on the Gram kernels of the two start-aligned trajectories, so a trajectory with the same Gram kernel but unrelated directions would follow the hyperbolas exactly. The projection therefore visualizes, mode by mode, the shape mismatch that the "After" columns of Figure 2 quantify, whereas the "Before" columns compare the unaligned trajectories. The Gaussian trajectory, however, spans fewer directions than the real one: at most six for PVD, whose covariance has six distinct eigenvalues, and effectively few for the other models, whose eigenmodes with similar eigenvalues have nearly identical time profiles. The alignment thus leaves the rest of the real trajectory free to rotate without changing the

[14] https://github.com/alexzhou907/PVD

residual, so the curves show one of several equally optimal views; the "After" columns are the same for all of them. We select four eigenmode pairs. The reference mode j is the 1,000th eigenvector for CIFAR-10 and E2E and the 4,000th for SC09, and the mode i is chosen such that $\Delta\lambda$ is approximately 1, 0.5–0.6, 0.3, and 0.1 times the gap of the second eigenmode; the leading eigenmode is excluded because its gap is far larger than all others. For PVD, we compare the second and third common-point modes and the first two residual eigenspaces with the third residual eigenspace serving as the reference mode j. Solid lines show the pointwise mean over the real trajectories whose initial stochasticity satisfies $|\tilde{y}_{T,k}| \geq 0.01$ for all plotted modes (187 to 193 of 200), and thin lines show the ten real trajectories whose deviation from this mean is closest to the median.

A.3.3 Trajectory deviations

A dominant first principal component and trajectory congruence alone cannot validate a theoretical model, since simple baselines share both properties in high dimensions. We therefore compare real trajectories with their Gaussian counterparts and with three simpler references, all starting from the same initial noise. Let $\Delta\mathbf{x}_t = \mathbf{x}_t - \mathbf{x}_T$ and $\Delta\bar{\mathbf{x}}_t = \bar{\mathbf{x}}_t - \bar{\mathbf{x}}_T$ denote the start-aligned real and Gaussian paths. Define progress as $u_t = (T-t)/(T-t_{\min})$, where $t_{\min}$ is the noise level of the final saved state. The Gaussian chord is

$$\mathbf{c}_t = u_t \Delta\bar{\mathbf{x}}_{t_{\min}}, \tag{15}$$

so it uses the same start, endpoint, and noise-level progress as the Gaussian counterpart while discarding its curved shape. We also include a Brownian bridge around this Gaussian chord as the null shape model, to control for potential artifacts of principal-component-based dimensionality reduction,

$$\mathbf{b}_t = \mathbf{c}_t + \sigma_B(\mathbf{w}_{u_t} - u_t\mathbf{w}_1), \tag{16}$$

where $\mathbf{w}_u$ is a standard Wiener process in $\mathbb{R}^d$, so the bridge is pinned to the same two endpoints. Its amplitude is chosen independently for each real trajectory according to

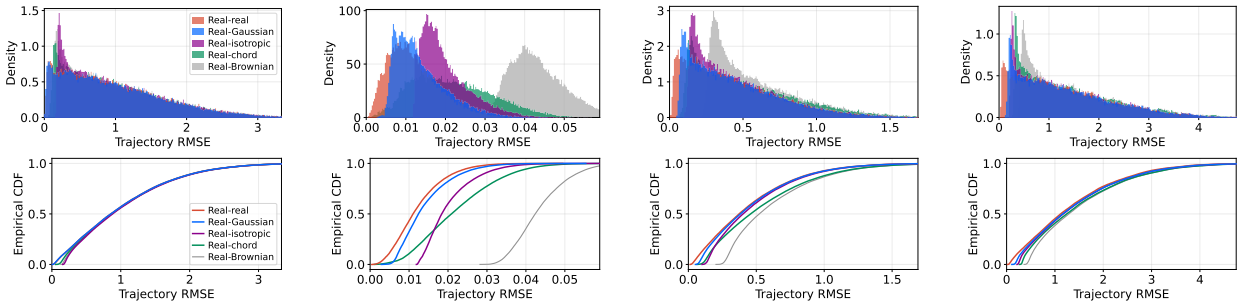
$$\sigma_B^2 = \frac{\mathbb{E}_t\|\Delta\mathbf{x}_t - \mathbf{c}_t\|_2^2}{d\,\mathbb{E}_t\left[u_t(1-u_t)\right]}, \tag{17}$$

where $\mathbb{E}_t$ is the uniform average over the saved states, i.e., ν is their empirical distribution. Thus, the bridge matches the real trajectory's mean squared deviation from the Gaussian chord but retains a random shape. Since the random shape $\mathbf{w}_{u_t} - u_t\mathbf{w}_1$ is independent of the real trajectory, which sets only its amplitude σ_B, the bridge's mean squared deviation from the real trajectory before alignment is about twice that of the chord, giving the ratio of about $\sqrt{2}$ in Figure 2. The isotropic Gaussian flow replaces $\mathbf{\Sigma}$ with the scaled identity $(\operatorname{tr}\mathbf{\Sigma}/d)\,\mathbf{I}_d$, which keeps the total data variance and the noise-level dependence of the Gaussian flow but none of its covariance structure; it moves along a straight line from the initial noise toward the data mean.

In both comparisons below, every trajectory is translated so that its initial state lies at the origin, as for $\Delta\mathbf{x}_t$ above, and its L saved states form the rows of a trajectory matrix in $\mathbb{R}^{L\times d}$.

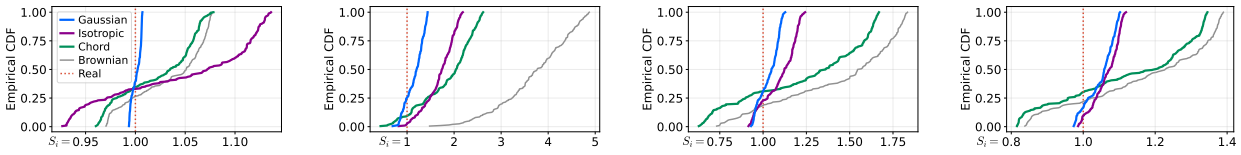
- **Identical-start trajectory deviation**. We evaluate the Gaussian counterpart, the isotropic Gaussian flow, the Gaussian chord, and the Brownian bridge before and after alignment. For any two trajectory matrices $\mathbf{A}, \mathbf{B} \in \mathbb{R}^{L\times d}$, we define the trajectory root-mean-square error (RMSE) over the saved states as

$$d(\mathbf{A},\mathbf{B}) = \min_{\mathbf{Q}^\top\mathbf{Q}=\mathbf{I}} \left[\frac{1}{Ld}\,\|\mathbf{A}-\mathbf{B}\mathbf{Q}\|_{\mathrm{F}}^2\right]^{1/2}, \tag{18}$$



(a) Image. (b) Audio. (c) Text. (d) 3D.

Figure 3: Distributions of the trajectory RMSE for independently initialized paths. The top row shows the error densities, and the bottom row shows the corresponding empirical cumulative distribution functions. In addition to the real–real and real–Gaussian comparisons in the bottom row of Figure 2, we include the isotropic Gaussian flow, Gaussian chord, and Brownian bridge.



(a) Image. (b) Audio. (c) Text. (d) 3D.

Figure 4: For a reference family $\mathcal{M}$, we define $S_i^{(\mathcal{M})} = \text{mean}_{j\neq i}\, d(\mathbf{X}_i, \mathbf{M}_j)/\,\text{mean}_{j\neq i}\, d(\mathbf{X}_i, \mathbf{X}_j)$, where $d(\cdot,\cdot)$ is the optimally aligned trajectory RMSE and $\mathbf{M}_j$ uses the same initialization as real trajectory $\mathbf{X}_j$. The curves show the distribution of this score over the 200 real trajectories. The dotted line marks $S_i = 1$, matching the average real–real error; lower values mean closer agreement.

where the minimum over $\mathbf{Q}$ is the alignment, a single ambient-space orthogonal Procrustes transformation fitted over the complete trajectory, and $\mathbf{Q} = \mathbf{I}_d$ before alignment. We normalize each deviation by that of the Gaussian chord before alignment, so the chord before alignment has ratio one by construction, and a ratio below one indicates an improvement over this endpoint-matched linear reference. The table in Figure 2 reports the mean of these ratios over the 200 trajectories of each model.

- **Independent-start trajectory deviation**. In addition, we compare independently initialized paths. This setting is the more realistic one, since real trajectories from different initial noise are themselves congruent only up to an orthogonal transformation (Figure 1(c) and our previous study [Chen et al., 2024]). Let $\mathbf{X}_i$ denote the trajectory matrix of real trajectory i, and let $\mathbf{G}_j$ denote that of the Gaussian trajectory starting from the same noise as real trajectory j. The bottom row of Figure 2 reports the distributions of $d(\mathbf{X}_i, \mathbf{X}_j)$ and $d(\mathbf{X}_i, \mathbf{G}_j)$ over all 200×199 ordered pairs with $i \neq j$, and Figure 3 adds the other references. In Figure 4, the scores of the Gaussian counterpart are the most concentrated and lie closest to one, matched only by the isotropic Gaussian flow in (d).

Together, the identical-start and independent-start comparisons support the hyperbolic structure of the Gaussian flow as a reasonable approximation of real trajectories.

As discussed in the main text, at $t = s$ the relative fluctuation of Gaussian sampling trajectories

(Eq. 2) equals $\sqrt{2/r_{\text{eff}}}$ at the data end and decays quickly as t grows, since adding t^2 to every eigenvalue flattens the spectrum. Once t^2 exceeds the root-mean-square eigenvalue $\sqrt{\text{tr}(\mathbf{\Lambda}^2)/d}$, it reaches the $\mathcal{O}(d^{-1/2})$ level, i.e., $\sqrt{2/d} \leq \sqrt{\text{Var}(G(t,t))}/\mathbb{E}[G(t,t)] \leq 2/\sqrt{d}$: the lower bound follows from the Cauchy–Schwarz inequality, and the upper bound from $\text{tr}(\mathbf{\Lambda}^2) \leq d\,t^4$. At the data end, we note the following for power-law eigenspectra.

Remark 1. The eigenspectrum of natural images typically exhibits a power-law decay, $\lambda_i \propto i^{-\alpha}$. In the slow-decay regime ($0 \leq \alpha < 1$), the quantity $r_{\text{eff}}^{-1/2}$ converges to zero as $\mathcal{O}(d^{-1/2})$ for $\alpha < 1/2$, as $\mathcal{O}(\sqrt{\ln d/d})$ at $\alpha = 1/2$, and as $\mathcal{O}(d^{-(1-\alpha)})$ for $1/2 < \alpha < 1$. At the critical boundary ($\alpha = 1$), it converges to zero as $\mathcal{O}(1/\ln d)$. In the steep-decay regime ($\alpha > 1$), it converges to a finite constant given by the Riemann zeta function $\zeta(\cdot)$: $\lim_{d\to\infty} r_{\text{eff}}^{-1/2} = \sqrt{\zeta(2\alpha)}/\zeta(\alpha)$, which monotonically approaches zero as $\alpha \to 1^+$.