PNAS RESEARCH ARTICLE | COMPUTER SCIENCES

On the universal hyperbolic structure in generative diffusion dynamics

Defang Chen[a,1], Zhenyu Zhou[b], Zeyu Zheng[c], Can Wang[b], and Siwei Lyu[a,1]

This manuscript was compiled on April 17, 2026

Diffusion-based generative modeling has emerged as a prominent computational framework for simulating high-dimensional, structured data. Recent studies have revealed striking geometric regularities in deterministic sampling trajectories, yet a theoretical explanation remains elusive. Here, we introduce an exactly solvable Gaussian flow to investigate generative diffusion dynamics. Our analysis establishes that sampling trajectories exhibit a universal hyperbolic structure and are asymptotically congruent in high dimensions. The predictions of this tractable model closely match the empirical behavior of deep neural network-based generative models.

generative diffusion | sampling dynamics | universality | Gaussian flow

Rooted in nonequilibrium statistical physics (1, 2), diffusion-based generative modeling provides a differentiable, data-driven framework for representing high-dimensional and complex data distributions (3, 4). This paradigm has achieved unprecedented success across diverse applications, including text-controlled image synthesis (5), drug discovery (6), and weather forecasting (7).

It employs a diffusion process to progressively perturb a sample $\mathbf{x}_0 \in \mathbb{R}^d$ from an unknown data distribution into an unstructured noise. The forward trajectory $\{\mathbf{x}_t\}_{t\in[0,T]}$ is typically governed by a stochastic differential equation $\mathrm{d}\mathbf{x}_t = \sqrt{2t}\,\mathrm{d}\mathbf{w}_t$, where $\mathbf{w}_t$ denotes the standard Wiener process (3, 8). Remarkably, the noise can be transformed back into the data by reversing the stochastic diffusion process, or through the equivalent *deterministic probability flow* $\mathrm{d}\mathbf{x}_t = -t\nabla_{\mathbf{x}_t}\log p_t(\mathbf{x}_t)\,\mathrm{d}t$. In practice, the score function $\nabla_{\mathbf{x}_t}\log p_t(\mathbf{x}_t)$ is estimated using a deep neural network trained on a large-scale dataset (3).

Recent empirical studies (8, 9) have revealed a striking phenomenon: deterministic sampling trajectories in generative diffusion models exhibit highly regular geometric patterns—all share an almost identical "boomerang" shape in low-dimensional subspaces, regardless of the model architecture, applied conditions, or generated content. Given the model variety and the heterogeneity of the data on which they are trained, the persistence of such regularities is surprising. Establishing a theoretical explanation for this phenomenon is essential for deepening our understanding of generative dynamics and for developing more efficient accelerated sampling algorithms from first principles.

In this work, we aim to achieve this goal using an exactly solvable model that yields closed-form solutions for deterministic sampling trajectories. Our analysis is built on approximating the centered data distribution with a single Gaussian distribution $q_0 \coloneqq \mathcal{N}(\mathbf{0}, \mathbf{\Sigma})$, where the covariance matrix $\mathbf{\Sigma} \in \mathbb{R}^{d\times d}$ is estimated empirically from the given dataset. The covariance matrix admits the spectral decomposition $\mathbf{\Sigma} = \mathbf{U}\mathbf{\Lambda}\mathbf{U}^\top$, where the diagonal matrix $\mathbf{\Lambda} = \mathrm{diag}\,(\lambda_1, \cdots, \lambda_d)$ contains the non-negative eigenvalues sorted in descending order, and the orthogonal matrix $\mathbf{U} = [\mathbf{u}_1, \cdots, \mathbf{u}_d] \in \mathbb{R}^{d\times d}$ contains the corresponding eigenvectors. Under this formulation, the random process admits the marginal distribution $q_t \coloneqq \mathcal{N}(\mathbf{0}, \mathbf{\Sigma} + t^2\mathbf{I}_d)$, and the sampling trajectory starting from the terminal state $\mathbf{x}_T \sim q_T = \mathcal{N}(\mathbf{0}, \mathbf{\Sigma} + T^2\mathbf{I}_d)$ can be analytically derived by solving the *deterministic Gaussian flow* $\mathrm{d}\mathbf{x}_t = t(\mathbf{\Sigma} + t^2\mathbf{I}_d)^{-1}\,\mathbf{x}_t\,\mathrm{d}t$. In particular, the solution in the spectral representation $(\mathbf{y}_t = \mathbf{U}^\top\mathbf{x}_t)$ is

$$\mathbf{y}_t = \left(\mathbf{\Lambda} + t^2\mathbf{I}_d\right)^{1/2}\left(\mathbf{\Lambda} + T^2\mathbf{I}_d\right)^{-1/2}\mathbf{y}_T. \qquad [1]$$

We first prove that the deterministic Gaussian flow induces hyperbolic structures in the sampling trajectories when projected onto any two-dimensional spectral subspace. Furthermore, these sampling trajectories are asymptotically

Author affiliations: [a]Department of Computer Science and Engineering, State University of New York at Buffalo, Buffalo, NY 14260; [b]College of Computer Science, Zhejiang University, Hangzhou, 310027, China; [c]Department of Industrial Engineering and Operations Research, University of California, Berkeley, CA 94720

Author contributions: D.C. and S.L. designed research; D.C. and Z.Zhou performed research; D.C. and Z.Zheng contributed new reagents/analytic tools; D.C., Z.Zhou, Z.Zheng, W.C. and S.L. wrote the paper.

The authors declare no competing interest.

[1]To whom correspondence should be addressed. E-mail: defangch@buffalo.edu, siweilyu@buffalo.edu

Draft

congruent in high dimensions, that is, they differ from each other only by rigid motions. Finally, we demonstrate that our theoretical model aligns accurately with empirical observations. Our conclusion also applies to other types of forward diffusion processes.

Hyperbolic structures. We begin by simplifying Eq. 1 as $\mathbf{y}_t = (\mathbf{\Lambda} + t^2\mathbf{I}_d)^{1/2}\bar{\mathbf{y}}_T$, with $\bar{\mathbf{y}}_T = (\mathbf{\Lambda} + T^2\mathbf{I}_d)^{-1/2}\mathbf{y}_T \sim \mathcal{N}(\mathbf{0}, \mathbf{I_d})$, to decouple the deterministic and stochastic components.* We observe that the i-th coordinate, $y_t^{(i)}$, of the sampling trajectory tracks the restoration of data features at a deterministic rate $z_t^{(i)} := y_t^{(i)}/\bar{y}_T^{(i)} = \sqrt{\lambda_i + t^2}$ (for $\bar{y}_T^{(i)} \neq 0$) across different realizations, while the initial stochasticity $\bar{\mathbf{y}}_T$ acts exclusively as a scaling factor for these coordinates. Furthermore, projecting the trajectory onto any two-dimensional coordinate plane spanned by $\mathbf{u}_i$ and $\mathbf{u}_j$ with non-zero initial conditions ($\bar{y}_T^{(i)}\bar{y}_T^{(j)} \neq 0$) yields a hyperbolic segment

$$\left(z_t^{(i)}\right)^2 - \left(z_t^{(j)}\right)^2 = \lambda_i - \lambda_j, \qquad [2]$$

with $t \in [0, T]$. Geometrically, each relation (Eq. 2) 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. In the asymptotic limit ($t \to \infty$), the trajectory's curvature vanishes, as captured by the Taylor expansion $\mathbf{y}_t \approx (t\mathbf{I}_d + \mathbf{\Lambda}/(2t))\bar{\mathbf{y}}_T$.

Moreover, two-dimensional linear representations can effectively capture the entire sampling dynamics with a relative residual error of $\mathcal{O}(T^{-2})$, since the evolution of all spatial coordinates is highly synchronized (see *SI Appendix*). Our analysis of the deterministic generative process complements the spectral analysis performed on the forward diffusion process in ref. (10), which yields Lissajous curves.

Asymptotic congruence. Sampling trajectories starting from distinct terminal states are geometrically congruent (equivalent under orthogonal transformations) if their corresponding Gramian kernels are equal. The Gramian kernel of the continuous-time trajectory is defined as the inner product, $G(t,s) = \mathbf{y}_t^\top \mathbf{y}_s$, and its expectation over the terminal distribution is given by $\mathbb{E}[G(t,s)] = \sum_{i=1}^d (\sqrt{\lambda_i + t^2}\sqrt{\lambda_i + s^2})$. We then compute the *relative fluctuation* of $G(t,s)$ to quantify deviations from this deterministic limit caused by different realizations:

$$\frac{\sqrt{\mathrm{Var}\left(G(t,s)\right)}}{\mathbb{E}\left[G(t,s)\right]} = \frac{\sqrt{2\sum_{i=1}^d (\lambda_i + t^2)(\lambda_i + s^2)}}{\sum_{i=1}^d \sqrt{(\lambda_i + t^2)(\lambda_i + s^2)}}. \qquad [3]$$

This relative fluctuation depends on the data eigenspectrum and the noise level. 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 manifold. Here, the effective data dimension r_{eff} is defined as $(\sum_{i=1}^d \lambda_i)^2/(\sum_{i=1}^d \lambda_i^2)$. In high dimensions ($d \to \infty$), these statistical fluctuations are microscopically small compared to the macroscopic deterministic behavior of the Gaussian flow. Although the fluctuation increases near the data manifold, rendering the trajectory geometry slightly more flexible and varied, the overall structural congruence remains robust.

Remark 1. *The eigenspectrum of natural images typically exhibits a power-law decay (11), $\lambda_i \propto i^{-\alpha}$. At the critical boundary ($\alpha = 1$), the quantity $r_{\text{eff}}^{-1/2}$ scales as $\mathcal{O}(1/\ln d)$. In the steep-decay regime ($\alpha > 1$), it converges to a finite constant expressed in terms of 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 $\alpha \to 1^+$. In the slow-decay regime ($0 \leq \alpha < 1$), it converges to zero linearly or sublinearly.*

Detailed derivations are provided in *SI Appendix*, where we further characterize the overall structural discrepancy using the *relative trajectory fluctuation*. It also converges to zero at the rate of $\mathcal{O}(d^{-1/2})$ in high dimensions.

Results

We evaluate the effectiveness of the Gaussian flow model by comparing its predicted hyperbolic structures with the results of state-of-the-art diffusion models. Due to space limitations, we employ a recent Diffusion Transformer (DiT) model (12) as an illustrative example, which was trained on ImageNet at 512×512 resolution (13). By aligning the original implementation with our defined forward diffusion process via Itô's lemma (8), we obtain a maximum noise level $T \approx 152$, with fluctuations (Eq. 3) below 1.2% across the majority (96.1%) of the sampling trajectory in this setting.

We first empirically estimate the covariance of the centered data to obtain the Gaussian approximation q_0, which captures the second-order statistics of the image distribution p_0. The distributional discrepancy during the stochastic diffusion process, measured by the Kullback-Leibler (KL) divergence $\mathcal{D}_{\text{KL}}(p_t \| q_t)$, decreases monotonically with respect to t according to *de Bruijn's identity* (14). Consequently, the terminal distributions p_T and q_T become indistinguishable at a sufficiently large terminal time T (Fig. 1(a)).

Next, we randomly select an initial noise realization to deterministically generate both a Gaussian trajectory and a real trajectory using the Gaussian flow ($\mathrm{d}\mathbf{x}_t = t(\mathbf{\Sigma} + t^2\mathbf{I}_d)^{-1}\,\mathbf{x}_t\,\mathrm{d}t$) and the deep neural network-based probability flow ($\mathrm{d}\mathbf{x}_t = -t\nabla_{\mathbf{x}_t}\log p_t(\mathbf{x}_t)\,\mathrm{d}t$), respectively. These two trajectories gradually diverge, as visualized in Fig. 1(b). Importantly, we demonstrate that such trajectory differences arise primarily from distinct orientations rather than structural discrepancies, and they can be largely eliminated via an orthogonal transformation. Notably, the transformation matrix that minimizes the trajectory discrepancy is shared across all time steps and can be efficiently derived via singular value decomposition (see *SI Appendix*).

Furthermore, the Gaussian trajectory projected onto any two-dimensional spectral subspace spanned by the data eigenvectors $\mathbf{u}_i$ and $\mathbf{u}_j$ exhibits a hyperbolic structure (Eq. 2), the curvature of which is determined by the eigenvalue difference $(\lambda_i - \lambda_j)$. We therefore provide a quantitative comparison between the theoretical result and the real trajectory generated by the DiT model (12) at each time step by projecting the real trajectory (after the orthogonal transformation) onto the same

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

DRAFT

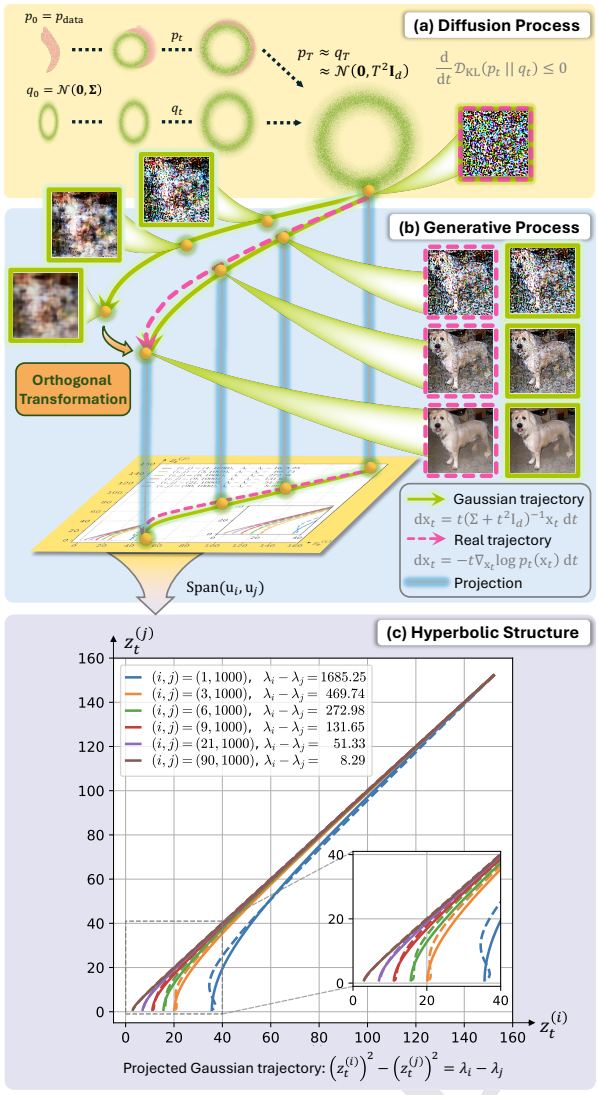


Fig. 1. (a) The Gaussian approximation q_0 shares the same covariance as the centered data distribution p_0, and their marginal distributions, p_t and q_t, are defined by convolution with the density of $\mathcal{N}(\mathbf{0}, t^2\mathbf{I}_d)$. Duing the stochastic diffusion process, the distributional discrepancy decreases monotonically. (b) In the deterministic generative process, the Gaussian trajectory (solid green curve, left) gradually deviates from the real trajectory (dashed pink curve) in terms of visual quality, yet they become closely aligned following an orthogonal transformation (solid green curve, right). (c) The original trajectories in $\mathbb{R}^d$ are projected onto two-dimensional subspaces spanned by the data eigenvectors $\mathbf{u}_i$ and $\mathbf{u}_j$. The universal hyperbolic structures effectively characterize the behavior of the real trajectory.

spectral subspace. Fig. 1(c) depicts several representative hyperbolic curves derived from the analytic Gaussian trajectory (solid lines) as a reference. These curves become asymptotically linear as $t \to \infty$ and degenerate into globally linear segments as the eigenvalue difference approaches zero.

As shown in Fig. 1(c), the Gaussian trajectory provides a remarkably effective approximation of the real trajectory (dashed lines). Similar results are observed when trajectories originate from distinct initial realizations, where a translation may be employed as a pre-processing step. Therefore, in high dimensions, all sampling trajectories are approximately rigid transformations of a single, universal "master trajectory" defined by Eq. 1. The above analysis accounts for the trajectory alignment and curvature observed in empirical studies of generative diffusion models (8, 9). These geometric properties arise intrinsically from the diffusion dynamics and the second-order statistics of the data distribution.

Discussion

Elucidating the geometric structure of generative diffusion dynamics is fundamental to explaining empirical observations and developing principled techniques for model training and inference. In this work, we advance this understanding by leveraging the Gaussian flow—the exact *minimum mean square error* (MMSE) linear approximation of the original probability flow—to rigorously characterize the universal hyperbolic structure of sampling trajectories in high dimensions, accurately recovering the regularities identified in previous studies (8, 9). While existing analyses typically focus on the evolution of probability distributions (15) or the local differential equations governing sampling dynamics (3, 8), we examine the geometry of full sampling trajectories to provide a global perspective on the generative process. Our findings open new avenues for the design of diffusion-based generative models, including the identification of optimal sampling configurations and acceleration of model training through prior knowledge of the underlying trajectory structure.

Data, Materials, and Software Availability. We utilized the open-source repository of DiT (https://github.com/facebookresearch/dit) to prepare the ImageNet dataset (https://www.kaggle.com/competitions/imagenet-object-localization-challenge/data) and employed its pre-trained model weights (https://dl.fbaipublicfiles.com/DiT/models/DiT-XL-2-512x512.pt). All code and scripts used in this work will be available in a public repository upon publication.

ACKNOWLEDGMENTS. This work is supported by the University at Buffalo Office of Vice President for Research and Economic Development.

1. C Jarzynski, Equilibrium free-energy differences from nonequilibrium measurements: A master-equation approach. *Phys. Rev. E* **56**, 5018 (1997).
2. J Sohl-Dickstein, E Weiss, N Maheswaranathan, S Ganguli, Deep unsupervised learning using nonequilibrium thermodynamics in *International Conference on Machine Learning*. (PMLR), Vol. 37, pp. 2256–2265 (2015).
3. Y Song, et al., Score-based generative modeling through stochastic differential equations in *International Conference on Learning Representations*. (2021).
4. D Ghio, Y Dandi, F Krzakala, L Zdeborová, Sampling with flows, diffusion, and autoregressive neural networks from a spin-glass perspective. *Proc. Natl. Acad. Sci.* **121**, e2311810121 (2024).
5. P Esser, et al., Scaling rectified flow transformers for high-resolution image synthesis in *International Conference on Machine Learning*. (PMLR), Vol. 235, pp. 12606–12633 (2024).
6. JL Watson, et al., De novo design of protein structure and function with rfdiffusion. *Nature* **620**, 1089–1100 (2023).
7. I Price, et al., Probabilistic weather forecasting with machine learning. *Nature* **637**, 84–90 (2025).
8. D Chen, Z Zhou, C Wang, C Shen, S Lyu, On the trajectory regularity of ode-based diffusion sampling in *International Conference on Machine Learning*. (PMLR), Vol. 235, pp. 7905–7934 (2024).
9. D Chen, Z Zhou, C Wang, S Lyu, Geometric regularity in deterministic sampling dynamics of diffusion-based generative models. *J. Stat. Mech. Theory Exp.* **2025**, 124002 (2025).
10. J Antognini, J Sohl-Dickstein, Pca of high dimensional random walks with comparison to neural network training in *Advances in Neural Information Processing Systems*. Vol. 31, pp. 10328–10337 (2018).
11. DL Ruderman, W Bialek, Statistics of natural images: Scaling in the woods. *Phys. Rev. Lett.* **73**, 814–817 (1994).
12. W Peebles, S Xie, Scalable diffusion models with transformers in *Proceedings of the IEEE/CVF International Conference on Computer Vision*. pp. 4195–4205 (2023).
13. O Russakovsky, et al., Imagenet large scale visual recognition challenge. *Int. J. Comput. Vis.* **115**, 211–252 (2015).
14. AJ Stam, Some inequalities satisfied by the quantities of information of fisher and shannon. *Inf. Control.* **2**, 101–112 (1959).
15. S Chewi, J Niles-Weed, P Rigollet, *Statistical optimal transport*. (Springer), (2025).