GEOMETRY-AWARE ORNSTEIN–UHLENBECK DIFFUSION VIA STRUCTURED DRIFT DESIGN

Anonymous authors
Paper under double-blind review

ABSTRACT

Diffusion-based generative models have achieved remarkable success in high-dimensional data generation; however, they fundamentally rely on isotropic diffusion processes that destroy meaningful geometric structures in the forward process. For complex, multimodal, and highly correlated distributions—such as biologically constrained genetic data—isotropic noise merges distinct modes and distorts intrinsic dependencies. This forces the reverse process to recover structure from heavily degraded signals, leading to slow convergence, mode averaging, and biologically implausible samples. To address this, we introduce the Geometry-aware Ornstein–Uhlenbeck (GOU) process, a structured drift design that embeds data geometry into forward and backward dynamics. By employing a variance-aware anisotropic drift, GOU contracts low-variance directions rapidly while preserving high-variance directions longer, maintaining key multimodal structures as stable channels over time. We provide rigorous theoretical analysis of the proposed dynamics, including the evolution of moments in the forward process and bounds on the initialization error of the backward process when using a Gaussian mixture model with explicitly derived means and covariance matrix. Furthermore, we theoretically analyze the optimal finite stopping time that prevents mode merging during the forward process. Experiments on synthetic and real-world biological data demonstrate that GOU achieves superior structural preservation and sample quality across both statistical and biological metrics.

1 INTRODUCTION

Diffusion models have achieved remarkable success in learning complex high-dimensional data distributions (Song and Ermon, 2019; Ho et al., 2020; Song et al., 2021; Taheri and Lederer, 2025). However, standard formulations rely on isotropic Gaussian perturbations to transform data into pure noise and a corresponding reverse process to reconstruct samples from the noise distribution. Both processes treat all directions equally and ignore the underlying geometric and correlation structure present in many scientific datasets. A crucial question araises here: What if we consider the geometry of data along the diffusion process?

Isotropic diffusion treats all directions uniformly, causing distinct modes in multimodal distributions to progressively blur together in the forward process. The reverse process must then recover separated components from severely degraded signals, which is challenging both theoretically and empirically. This can lead to inefficient sample generation, with convergence rates that depend on properties of the target distribution (e.g., log-concavity and smoothness). Existing convergence analyses show that diffusion samplers might mix slowly in non-log-concave landscapes, particularly in regions of low density between modes (Block et al., 2020; Tzen and Raginsky, 2019; Eberle, 2016; Lee et al., 2023; Gao et al., 2025; Kremling et al., 2025).

Moreover, accurately recovering the correlation structure of data is crucial in applications such as genetics and biology, where covariance patterns—particularly linkage disequilibrium (LD)—capture functional, biological, and evolutionary constraints. Even small deviations from these dependency structures can result in samples that are biologically implausible or statistically invalid. If the forward diffusion process substantially disrupts these correlations, the reverse process may be unable to recover them accurately, potentially leading to unrealistic generated genotypes (Lampis et al., 2026; Sadia et al., 2025; Si et al., 2026). While such sensitivity may be less pronounced in comparatively

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.

simple domains such as natural images, it becomes crucial in scientific applications where fine-scale dependency structures carry essential information. Recent works in genomic and single-cell generative modeling have highlighted that accurately preserving these correlations is necessary for generating realistic and scientifically meaningful samples (Kenneweg et al., 2025; Li et al., 2025).

These limitations motivate diffusion processes that explicitly incorporate the geometry of the data into the diffusion dynamics. Diffusion on manifolds replaces Euclidean noise with geometry-aware operators (De Bortoli et al., 2022; Huang et al., 2022), but typically requires explicit knowledge of the underlying manifold and introduces substantial analytical and computational complexity. Preconditioned stochastic dynamics adapt sampling procedures using local geometry (Girolami and Calderhead, 2011; Ma et al., 2015), yet these approaches mainly target the backward process. Transport-based approaches, including neural ODEs and flow matching (Chen et al., 2018; Lipman et al., 2023), learn flexible mappings between distributions, but do not provide a tractable or interpretable mechanism to encode geometry directly into the diffusion process.

Taken together, existing approaches largely treat diffusion under fixed isotropic dynamics or attempt to incorporate geometry via complex or external mechanisms—such as diffusion in manifold, conditional sampling, or structured noise. However, these strategies are often intractable or suboptimal to integrate into a unified generative framework. As a result, using the forward diffusion process to directly encode data-dependent structure remains largely unexplored. Building on this perspective, we propose a structured diffusion framework that introduces anisotropic feature-dependent drifts, maintaining mode separation and preserving correlations among features for significantly long time steps.

Two closely related works explore structured prior distributions and local manifold structures in the latent space to guide generation in diffusion models and flow matching (Jia et al., 2024; Luo et al., 2026). Jia et al. (2024) employ a Gaussian Mixture Model (GMM) as a structured prior for diffusion models, applying diffusion dynamics locally over data clusters. Extending this concept to latent space and flow matching, Luo et al. (2026) utilize vision foundation models as encoders to reveal union-of-manifold structures. They employ a multimodal GMM source distribution alongside mode coupling, implying flow matching locally in the latent space. While these works share our core insight—that multi-modal priors provide a superior generative foundation, our work takes a complementary perspective by directly modifying the drift function during the diffusion process. Rather than relying on explicit mode coupling and pre-trained foundation encoders to partition the space, our structured drift design continuously guides the diffusion trajectory along local linear subspaces, preserving the data structure and multi-modality of the data for a longer time without auxiliary architectural constraints.

Contributions The primary contributions of this work are summarized as follows:

1. We introduce *Geometry-aware Ornstein–Uhlenbeck (GOU)* process, a novel structure-preserving diffusion framework explicitly designed to mitigate the rapid decay of feature correlations and delay mode merging during the forward process through a structured, anisotropic drift.
2. We provide a rigorous *theoretical analysis* mapping the exact analytical evolution of moments (means and covariances) under the proposed GOU dynamics (Theorem 2 and Theorem 3).
3. We show that under the GOU process, local modes remain well-separated at finite time T (Theorem 4). Leveraging this property, we introduce an effective backward initialization scheme that approximates the terminal distribution using a Gaussian mixture model. This reduces the initialization error so that it scales with individual local cluster variances rather than the global dataset variance, avoiding the initialization mismatch common in standard isotropic diffusion.
4. We demonstrate empirically that GOU significantly improves *mode separation, latent interpretability, and sample fidelity* over standard diffusion baselines across both synthetic benchmarks and real-world biologically structured datasets.

Paper outline We review the related literature in Section 2. In Section 3, we introduce the proposed *geometry-aware Ornstein–Uhlenbeck (GOU)* process. In Section 4, we present a theoreti-

cal analysis of the evolution of the mean and covariance under the GOU process, highlighting its structure-preserving properties. In Section 5, we study the reverse process of the GOU framework and analyze the associated initialization error. Finally, in Section 6, we empirically evaluate our approach on synthetic and biological datasets. Additional theoretical and empirical results, complete proofs, and further discussion are provided in the Appendix.

2 Related works

A growing body of research provides theoretical guarantees for diffusion models and their sampling procedures. Existing analyses establish convergence under various assumptions on the data distribution, score function, and discretization schemes (Block et al., 2020; De Bortoli, 2022; Wibisono and Yang, 2022; Chen et al., 2023c;a; Lee et al., 2023; Silveri and Ocello, 2025; Liang et al., 2024; Gao and Zhu, 2024; Gao et al., 2025; Kremling et al., 2025; Taheri and Lederer, 2025; Lee et al., 2022; Chen et al., 2023b; Conforti et al., 2025; Beyler and Bach, 2025). These results cover convergence in total variation, Kullback–Leibler divergence, and Wasserstein distances. Despite these advances, most analyses in Wasserstein distance rely on strong global assumptions on the target density, such as log-concavity, as well as regularity conditions on the score function, including Lipschitz continuity and smoothness. More recent works aim to relax these assumptions by introducing weaker conditions, such as weak log-concavity or dissipativity (Silveri and Ocello, 2025; Bruno and Sabanis, 2025; Brigati and Pedrotti, 2024; Ishige, 2024; Kremling et al., 2025). While these approaches broaden the theoretical scope, they still fall short of covering realistic data settings, such as non-sub-Gaussian or heavy-tailed distributions, and often yield conservative (i.e., loose) convergence rates. Moreover, these works focus primarily on the reverse-time dynamics, leaving the forward process largely unexplored as a design component.

In a different direction, several works aim to incorporate geometric structure into generative modeling. Diffusion on manifolds replaces Euclidean noise with geometry-aware operators (De Bortoli et al., 2022; Huang et al., 2022), but typically requires explicit knowledge of the underlying manifold and introduces substantial analytical and computational complexity. However, these approaches fail to preserve the multimodality of the distribution and the structural relationships between features on the manifold. Preconditioned stochastic dynamics adapt sampling procedures using local geometry (Girolami and Calderhead, 2011; Ma et al., 2015), yet these approaches are designed to target just the backward process. Transport-based approaches, including neural ODEs and flow matching (Chen et al., 2018; Lipman et al., 2023), learn flexible mappings between distributions, but do not provide a tractable or interpretable mechanism to encode geometry directly into the diffusion process.

A recent line of research (empirically) investigates non-isotropic diffusion processes employing structured diffusion coefficients, including learning matrix-valued diffusion coefficients (Liu et al., 2026), edge-aware perturbations (Vandersanden et al., 2024), blue-noise masks (Huang et al., 2024), covariance-informed diffusion (Huk and Damoulas, 2026), and subspace- or frequency-restricted dynamics (Jing et al., 2022), all of which introduce anisotropic noise to better match the data geometry. For instance, Liu et al. (2026) proposed a framework for anisotropic diffusion models by replacing the traditional scalar noise schedule with a matrix-valued path. This monotone PSD matrix dynamically allocates noise across distinct directions and subspaces, and is optimized through a trajectory-level objective that jointly trains the noise schedule and the score network. In contrast, our approach formulates anisotropy through a generalized Ornstein–Uhlenbeck dynamics that explicitly encodes geometric structure in the drift function rather than in the noise, providing a unified perspective on how geometry governs diffusion. Despite these differences in modeling choices and application domains, these works collectively underscore the importance of incorporating data geometry into diffusion-based modeling, motivating our present study, which aims to integrate geometric structure directly into the forward dynamics in a principled and analytically tractable manner.

Notations For a vector $X \in \mathbb{R}^d$, we denote by $\|X\|$ its Euclidean norm. For any two probability measures $\mu, \nu \in \mathcal{P}_2(\mathbb{R}^d)$, the space of measures on $\mathbb{R}^d$ with finite second moment, the 2-Wasserstein distance, based on the Euclidean norm, is defined as

$$\mathcal{W}_2(\mu,\nu) := \left(\inf_{X\sim\mu, Y\sim\nu} \mathbb{E}\|X-Y\|^2 \right)^{\frac{1}{2}},$$

where the infimum is taken over all possible couplings of μ and ν.

3 GEOMETRY-AWARE DIFFUSION MODELS

Standard denoising diffusion models use a forward Stochastic Differential Equation (SDE) that gradually turns data into noise, together with a reverse-time SDE or ODE (Ordinary Differential Equation) that reconstructs the data using a learned score function (Song et al., 2021). While this framework performs well for generic data such as natural images, it ignores the underlying geometry of the data, including multimodality and the intrinsic dependency structure that characterizes many scientific datasets. Examples include correlations among genes, single nucleotide polymorphisms (SNPs), and other physical or biological measurements, where preserving these structures is essential for generating realistic samples. In particular, the forward process is typically isotropic, meaning that it treats all directions in the data space equally. For example variance-preserving (VP) SDE is of the form

$$\mathrm{d}X_t = -\frac{1}{2}\beta(t)X_t\,\mathrm{d}t + \sqrt{\beta(t)}\,\mathrm{d}W_t, \tag{1}$$

where $X_t \in \mathbb{R}^d$, W_t is a d-dimensional Brownian motion, and $\beta(t) > 0$ is a time dependent drift function.

In this work, we propose anisotropic stochastic processes, in which data are diffused respecting their underlying geometric structure. Our primary contribution is a forward diffusion process that mitigates mode mixing in multimodal distributions and preserves the intrinsic geometric structure of the data over extended diffusion times.

3.1 GEOMETRY-AWARE ORNSTEIN-UHLENBECK

We design a drift matrix $M \in \mathbb{R}^{d\times d}$ for the Ornstein-Uhlenbeck (OU) process that contracts directions proportionally to the variance of data, allowing for controlled diffusion. Let $\Sigma \coloneqq \mathrm{Cov}_{p_0}(X)$ denote the covariance matrix of the target distribution p_0, with eigen-decomposition

$$\Sigma = U\Lambda U^\top, \qquad \Lambda = \mathrm{diag}(\lambda_1, \ldots, \lambda_d),$$

where U contains the principal directions with $UU^T = \mathbb{I}_d$ and $\{\lambda_i\}_{i=1}^d \geq 0$ are the corresponding variances along the principal directions.

We propose variance-aware contraction rate along each principal direction

$$c_i := -b\left(\gamma + (1-\gamma)\left(1 - \frac{\lambda_i}{\lambda_{\max}}\right)\right), \qquad b, \gamma \geq 0, \tag{2}$$

where $\lambda_{\max}$ is the largest eigenvalue and b and γ are hyperparameters to be chosen. This implies $c_i < 0$ for all $i \in \{1, \ldots, d\}$. Equation equation 2 implies directions of high variance to contract slowly, whereas low-variance directions contract quickly:

$$\lambda_i \text{ large } \implies |c_i| \text{ small (slow contraction)}, \qquad \lambda_i \text{ small } \implies |c_i| \text{ large (fast contraction)}.$$

We then employ the drift matrix

$$M := U\,\mathrm{diag}(c_1, \ldots, c_d)\,U^\top, \tag{3}$$

which is fixed and shared across the whole diffusion process. Employing the designed drift matrix M, we define the geometry-aware Ornstein-Uhlenbeck as

$$\mathrm{d}X_t = M(X_t - \mu)\,\mathrm{d}t + \sqrt{\beta}\,\mathrm{d}W_t, \tag{4}$$

where $\mu \in \mathbb{R}^d$ may be chosen fixed as a global mean or another reference center and $\beta > 0$ is a constant diffusion coefficient. In fact, the diffusion adapts the contraction rates to the estimated covariance geometry. Directions with small empirical variance are contracted more strongly, reducing the influence of poorly represented directions while preserving slower dynamics along dominant principal components.

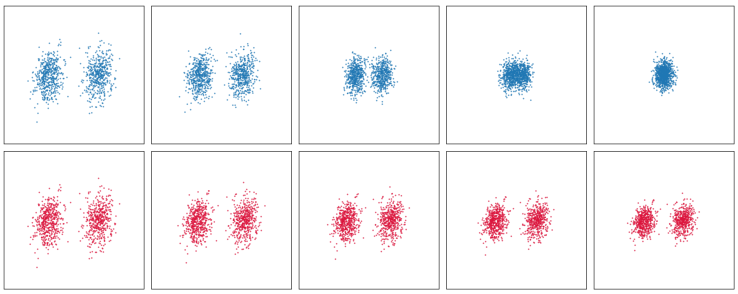


Figure 1: Comparison of the isotropic diffusion (first row) with GOU process (second row) in the forward dynamics (for $t \in \{0, 0.2, 0.5, 0.7, 1\}$ from left to right). The GOU process maintains separation between modes along the X-axis, while isotropic diffusion quickly mixes samples across modes.

3.2 Mode separation under GOU process

In this section we study how local modes are preserved under the GOU forward process introduced in equation 4. For simplicity, we consider a bimodal data distribution, although the results extend naturally to general multimodal settings.

Theorem 1 (Mode separation under GOU)**.** *Let $\mu_1(0), \mu_2(0) \in \mathbb{R}^d$ denote the initial means of two modes with X_0 has finite first and second moments. Let $\mu_k(t) = \mathbb{E}[X_t^{(k)}]$ be the corresponding mean trajectories under equation 4, initialized at $\mu_k(0)$ for $k \in \{1, 2\}$. Define the initial separation of two modes*

$$\Delta_0 \coloneqq \mu_1(0) - \mu_2(0)$$

and let define $U^\top \Delta_0 =: \alpha$, with components α_i for $i \in \{1, \dots, d\}$, which are coordinates in the rotated system. Then, for all $t \geq 0$,

$$\Delta(t) \coloneqq \mu_1(t) - \mu_2(t) = e^{Mt}\Delta_0 = U \operatorname{diag}(e^{c_1 t}, \dots, e^{c_d t})\, \alpha = \sum_{i=1}^{d} \alpha_i e^{c_i t} u_i \,,$$

where u_i denotes the i-th column of U. Moreover,

$$\|\Delta(t)\|^2 = \sum_{i=1}^{d} \alpha_i^2 e^{2c_i t} \,.$$

In particular, each principal-direction component evolves independently with the rate $e^{c_i t}$.

Theorem 1 establishes that the separation between modes in high-variance directions decays significantly more slowly than in low-variance directions. We defer the detailed proof to Section B.4. We next state an immediate consequence of Theorem 1.

Corollary 1 (Persistence of mode separation under GOU)**.** *Let Δ_0 denote the initial mean difference between two modes, and write $\alpha = U^\top \Delta_0$. If there exists at least one index i such that $\alpha_i \neq 0$ and $|c_i|$ is small compared to the other $|c_j|$, then the corresponding component of the separation $\Delta(t)$ decays more slowly and thus persists for a longer time.*

For a visual illustration, we refer to Figure 1, which shows that the GOU forward process preserves mode separation along the X-axis, whereas isotropic diffusion rapidly mixes samples from the two modes in a mixture of two Gaussians.

4 Forward process under GOU

In this section, we analyze the evolution of the mean and covariance matrix under the GOU forward process and subsequently characterize its stationary distribution.

4.1 MEAN AND COVARIANCE EVOLUTION UNDER GOU

We now turn to study how the mean and covariance matrix diffuse under the GOU forward process. We study the mean evolution in Theorem 2 and the covariance evolution in Theorem 3.

Theorem 2 (Mean evolution under GOU)**.** *Let $\mu_t := \mathbb{E}[X_t]$ and assume X_0 has finite first and second moments. Then*

$$\mu_t = \mu + e^{Mt}(\mu_0 - \mu)\,. \tag{5}$$

Equivalently, writing the spectral decomposition of M, we obtain

$$\mu_t = \mu + U\,\mathrm{diag}(e^{c_1 t}, \ldots, e^{c_d t})\,U^\top(\mu_0 - \mu)\,. \tag{6}$$

The explicit solution equation 5 shows that the mean of the forward process of GOU evolves deterministically under the linear flow induced by the drift matrix M, independently of the noise level β. In contrast to isotropic diffusion, where the mean decays uniformly, the structured drift induces direction-dependent contraction toward the reference center μ. For M diagonalizable (as we defined), each principal component of the mean evolves independently at a rate determined by the corresponding eigenvalue of M, allowing directions with high variance to persist for longer time steps (see Figure 1).

Theorem 3 (Covariance evolution under GOU)**.** *Let $\Sigma_t := \mathrm{Cov}(X_t)$. Then Σ_t satisfies the Lyapunov differential equation*

$$\frac{\mathrm{d}\Sigma_t}{\mathrm{d}t} = M\Sigma_t + \Sigma_t M^\top + \beta\mathbb{I}_d\,, \qquad \Sigma_0 = \mathrm{Cov}(X_0)\,. \tag{7}$$

Moreover, the unique solution is given by

$$\Sigma_t = e^{Mt}\Sigma_0 e^{M^\top t} + \beta\int_0^t e^{M(t-s)}e^{M^\top(t-s)}\,\mathrm{d}s\,. \tag{8}$$

Theorem 3 describes how the uncertainty of the diffusion process evolves over time. The covariance matrix Σ_t changes through two mechanisms. First, the existing covariance Σ_0 is transported by the linear dynamics induced by the drift matrix M, which reshapes the correlations in the data according to the geometric structure encoded in M. Second, stochastic noise continuously injects additional variance into the system at rate β. The resulting covariance therefore consists of two parts: a transformed version of the initial covariance (first term on the right-hand side of equation 8) and an accumulated noise contribution (second term on the right-hand side of equation 8). Although the injected noise is isotropic, the drift matrix M steers its propagation so that the evolving covariance remains anisotropic and aligned with the principal directions of the data. Note that the drift matrix M appears in the diffusion part of the covariance evolution because the same linear dynamics that contract the mean also act on past noise increments. At each infinitesimal step, isotropic noise $\sqrt{\beta}\,\mathrm{d}W_t$ is injected equally in all directions. However, once injected, each noise increment is subsequently transported by the drift propagator $e^{M(t-s)}$ from its injection time s to the present time t. Directions with slow contraction (small $|c_i|$) retain past noise for longer, leading to larger accumulated variance, while fast-contracting directions (large $|c_i|$) quickly suppress past noise. Thus, although the instantaneous noise injection is isotropic, the total accumulated covariance becomes anisotropic precisely because M controls the memory of past fluctuations. The proof of the Theorem 3 is deferred to the Appendix B.3.

4.2 STATIONARY DISTRIBUTION

A direct consequence of Theorem 2 yields the mean of the stationary distribution in the limit as $t \to \infty$.

Corollary 2 (Stationary mean of the GOU process)**.** *Under the same assumptions as in Theorem 2, the mean u_t converges as $t \to \infty$ to u.*

Employing Theorem 3, we can directly compute the stationary covariance in the limit $t \to \infty$.

Lemma 1 (Stationary covariance of the GOU process)**.** *Under the same assumptions as in Theorem 3, the covariance matrix Σ_t converges as $t \to \infty$ to a unique stationary covariance Σ_∞, given by the unique solution of the Lyapunov equation*

$$M\Sigma_\infty + \Sigma_\infty M^\top + \beta\mathbb{I}_d = 0\,. \tag{9}$$

Moreover, as M is symmetric, this simplifies to

$$\Sigma_\infty = -\frac{\beta}{2} M^{-1} . \tag{10}$$

The proof of Lemma 1 is deferred to the Appendix B.5.

Remark 1 (Interpretation of the stationary covariance). Lemma 1 shows that the stationary covariance is fully determined by the drift matrix M and the noise level β.

- *Relation to contraction rates:* From the definition of c_i in Eq. equation 2, directions with weaker contraction (i.e., smaller $|c_i|$) correspond to larger stationary variance, while strongly contracting directions lead to smaller stationary variance.
- *Covariance preserving:* In a low-dimensional setting $d \ll n$, set $c_i = -1/\lambda_i$ and $\beta = 2$, yields $\Sigma_\infty = \Sigma_0$, showing that the stationary covariance recovers the data covariance exactly.

Overall, the stationary distribution is anisotropic and encodes the geometry induced by the drift matrix M, in contrast to standard isotropic diffusion models.

5 Backward process under GOU

We now move to study the reverse-time dynamics of the GOU process in equation 4.

5.1 Reverse-time SDE

Consider the forward GOU process defined in equation 4, and let $p_t(x)$ denote the marginal density of X_t at time t. From the general time-reversal theory for diffusion processes (Anderson, 1982), the reverse-time process $\{\bar{X}_t\}_{0 \le t \le T}$, defined such that $\bar{X}_t = X_{T-t}$ in distribution, satisfies the stochastic differential equation

$$\mathrm{d}\bar{X}_t = \big[-M(\bar{X}_t - \mu) + \beta \nabla_x \log p_{T-t}(\bar{X}_t) \big] \mathrm{d}t + \sqrt{\beta}\, \mathrm{d}\bar{W}_t , \tag{11}$$

where $\bar{W}_t$ is a d-dimensional Brownian motion. The reverse dynamics consist of two components. The first term, $-M(\bar{X}_t - \mu)$, corresponds to the deterministic drift of the forward GOU process run in the opposite direction. The second term, involving the score function $\nabla_x \log p_{T-t}(\cdot)$, compensates for the stochastic diffusion introduced in the forward process and guides the trajectories toward regions of high probability under the evolving distribution. Together, these dynamics transform samples from a prior distribution back into structured data while preserving the geometry induced by the matrix M. We then follow the framework of score-based generative modeling (Song and Ermon, 2019), training a neural network $s_\theta(x, t)$ that approximates scores $s_\theta(x, t) \approx \nabla_x \log p_t(x)$ to employ in equation 11.

5.2 Initialization of the backward process

Unlike standard isotropic diffusion, which converges to a stationary distribution given by a standard Gaussian, the GOU forward process does not admit such a simple limiting distribution. Consequently, the backward generative process cannot be initialized from standard Gaussian noise. In GOU, the correct initialization for the reverse-time SDE is given by the distribution of the forward process at the terminal time T using the expressions for the mean μ_T and covariance Σ_T derived in Theorem 2 and Theorem 3. An effective approximation is to initialize the backward process from a mixture distribution constructed across a set of structural anchors (e.g., cluster centers or representative data samples). Specifically, we define the practical initialization distribution $\hat{p}_T$ by assigning a localized Gaussian profile to each anchor:

$$\hat{p}_T(x) = \mathbb{E}_{\hat{X}_0 \sim p_{\text{anchor}}} \Big[\mathcal{N}\big(x; \mu + e^{MT}(\hat{X}_0 - \mu), \Sigma_T\big) \Big] , \tag{12}$$

where p_{anchor} is the empirical distribution over the chosen structural anchors $\{\hat{X}_0^{(k)}\}_{k=1}^K$. To sample from $\hat{p}_T$ at the start of generation, one simply selects an anchor $\hat{X}_0$ (e.g., a cluster center where

centers are obtained via a clustering algorithm) at random and draws a sample from its corresponding shifted Gaussian component. Recall that Σ_T only needs to be computed once via equation 8.

This initialization scheme ensures that sample generation originates directly from a multimodal base distribution tailored to the target data geometry. Indeed, because we consider a finite time T to preserve mode separation (see Theorem 4), the GOU forward process does not reach a stationary distribution. Instead, we approximate the terminal distribution at time T using a Gaussian mixture model (GMM), leveraging the property that GMMs are strong universal approximators for such terminal distributions (Nguyen et al., 2020). Our empirical observations indicate that this approach performs remarkably well in practice. Below, we study the initialization error induced employing the initialization scheme in equation 12.

Lemma 2 (Wasserstein bound on the initialization error)**.** *Let p_T be the true distribution of the forward GOU process at terminal time T initialized from the data distribution $X_0 \sim p_0$. Let $\hat{p}_T$ be the practical initialization distribution of the backward process defined in equation 12. Let $\pi(X_0, \hat{X}_0)$ be the joint distribution (coupling) matching data points to their corresponding structural anchors. Then, we have*

$$\mathcal{W}_2(p_T, \hat{p}_T) \leq e^{c_{\max} T} \sqrt{\mathbb{E}_{(X_0, \hat{X}_0)\sim\pi}\left[\|X_0 - \hat{X}_0\|^2\right]}\,, \tag{13}$$

where $c_{\max} \coloneqq \max_i c_i$ is the maximum (least negative) eigenvalue of the contraction drift matrix M.

To see the practical importance of Lemma 2, consider a multimodal target distribution partitioned into distinct clusters. Because $\hat{p}_T$ explicitly models these modes via the structural anchors, the mathematical definition of the $\mathcal{W}_2$-distance admits the following admissible coupling, $\pi(X_0, \hat{X}_0)$ that pairs each true data point X_0 with its respective cluster center $\hat{X}_0 = \mu^{(k)}$. Under this assignment, the joint expectation $\mathbb{E}_{(X_0, \hat{X}_0)\sim\pi}\left[\|X_0 - \hat{X}_0\|^2\right]$ collapses into a weighted sum of local cluster variances. Consequently, the initialization error scales down proportionally to the local variance of each individual cluster, rather than the global variance of the entire dataset. This stands in stark contrast to classical isotropic diffusion models, where the initialization error is fundamentally scales with the second moment $\mathbb{E}\left[\|X_0\|^2\right]$ (Gao and Zhu, 2024, Lemma 16).

6 Empirical studies

We evaluate the proposed GOU process on both synthetic and real-world datasets. As motivated from the outset, our application domain focuses on biologically structured data, where isotropic drifts fail to adequately capture the underlying structure of the data. We present the results for the Paul15 dataset here; additional datasets are provided in Appendix C. We compare the proposed GOU process against an isotropic-drift baseline. For all experiments, both methods are trained using identical neural score network architectures and hyperparameters to ensure a fair comparison. Specifically, we use a multilayer perceptron (MLP) with four hidden layers of width 512, SiLU activations, and a learned 64-dimensional time embedding concatenated to the input. We employed $b = 1, \gamma = 0.4$, and $K = 30$ anchors across all experiments for GOU, though we observe that performance remains robust when varying $K \in \{10, 20, 40\}$. All experiments were run locally on a MacBook M2 with 16 GB of RAM.

To evaluate empirical performance, we focus on seven key statistical and biological metrics covering distinct aspects of distributional fidelity, diversity, and domain structure. Classifier Two-Sample AUC (C2ST) measures overall high-dimensional indistinguishability, with values near 0.5 indicating synthetic samples are indistinguishable from real data. Fréchet Distance (FD) measures feature-representation distance by comparing joint multivariate mean and covariance structures, while Sliced Wasserstein-2 quantifies global distribution alignment across random one-dimensional projections. To evaluate coverage, Manifold Recall measures the proportion of the real data manifold captured, ensuring the generator avoids mode collapse. Biologically, Co-expression Spearman assesses the preservation of feature-feature correlation networks and regulatory dependencies. At the population level, Cell-state Proportion JSD (Jensen-Shannon Divergence) quantifies whether synthetic generation distorts relative cell-type abundances. Finally, Marker-effect Spearman (or LD-decay Spearman for genetic variant data) verifies that fine-grained biological signatures—such as cluster-specific expression programs or physical linkage disequilibrium decay—are structurally preserved.

Single-cell RNA-seq Paul15 dataset We evaluate our method on the Paul15 dataset (Paul et al., 2015), which consists of single-cell RNA sequencing profiles capturing mouse hematopoietic stem cell differentiation across a continuous developmental trajectory. The dataset contains expression measurements for $2{,}700$ cells and over $5{,}000$ genes, and exhibits a well-characterized branching structure, where cells progressively differentiate into distinct blood lineages, making it a standard benchmark for trajectory inference and generative modeling of dynamic biological processes. For simplicity, we restrict our analysis to the first 200 genes. Overall, the empirical evaluation demonstrates that the structured model consistently outperforms the isotropic baseline across all primary statistical and biological fidelity metrics. Comprehensive statistical and biological results are detailed in Table 1, while Figure 3 provides a visual comparison of the low-dimensional representations and correlation heatmaps of the generated data. On statistical fidelity, the structured approach achieves superior high-dimensional indistinguishability (C2ST AUC $= 0.520$ vs. 0.622), lower geometric distance (Gaussian $W_2 = 0.106$ vs. 0.484), and higher manifold recall (0.528 vs. 0.345), indicating better multivariate alignment and reduced mode collapse. Biologically, it better preserves critical gene co-expression dependencies (0.779 vs. 0.454), maintains accurate cell-state proportions (JSD $= 0.006$ vs. 0.236), and faithfully recovers cell-type identity programs (Marker-effect Spearman $= 0.805$ vs. 0.573). Together, these paired differences confirm that incorporating structured priors significantly enhances sample realism while preserving fine-grained biological organization.

Table 1: Primary statistical and biological fidelity metrics comparing structured and isotropic methods across 10 paired units over paul15 dataset

Category	Metric	Ideal	Structured [95% CI]	Isotropic [95% CI]
Statistical	Classifier two-sample AUC	Near 0.5	0.520 [0.512, 0.530]	0.622 [0.615, 0.629]
	Gaussian W2 (FD)	Lower	0.106 [0.103, 0.109]	0.484 [0.481, 0.488]
	Sliced Wasserstein-2	Lower	0.093 [0.088, 0.099]	0.437 [0.433, 0.442]
	Manifold recall	Higher	0.528 [0.521, 0.535]	0.345 [0.338, 0.354]
Biological	Co-expression Spearman	Higher	0.779 [0.776, 0.782]	0.454 [0.449, 0.459]
	Cell-state proportion JSD	Lower	0.006 [0.005, 0.008]	0.236 [0.228, 0.243]
	Marker-effect Spearman	Higher	0.805 [0.793, 0.816]	0.573 [0.565, 0.582]

CONCLUSION

In this work, we introduced Geometry-aware Ornstein-Uhlenbeck diffusion models, which incorporate the intrinsic geometric structure of data into the forward diffusion process. By designing a variance-dependent drift matrix, the GOU process contracts low-variance directions quickly while preserving high-variance directions, effectively preventing mode merging in multimodal data. We derived explicit solutions for the evolution of both the mean and covariance, showing how the process transports correlations and injects noise in a geometry-aware manner. The corresponding reverse-time dynamics provide a principled generative mechanism that can reconstruct structured data while respecting the underlying geometry. Overall, GOU offers a simple yet powerful framework to integrate data-dependent structure into diffusion models, bridging the gap between classical isotropic diffusion and geometry-preserving generative modeling.

Limitation and future work. While the GOU process effectively incorporates geometric structure into diffusion dynamics, relying on a static global drift matrix M based on overall data covariance may not fully capture complex, non-linear, or localized structures across different regions of the data manifold. To address this, a promising direction is employing adaptive or state-dependent local drift matrices, which can also guarantee stationary multimodal distributions as time approaches infinity. Furthermore, while existing convergence analyses for reverse-time diffusion processes often rely on global (or weak) log-concavity assumptions (Block et al., 2020; Lee et al., 2023; Gao et al., 2025; Kremling et al., 2025), such conditions are rarely satisfied by complex real-world data. We conjecture that the geometry-aware structure of the GOU process allows for substantially weaker assumptions—such as local log-concavity or monotonicity—to guarantee global contraction in Wasserstein distance. Establishing such non-asymptotic guarantees for multimodal and heavy-tailed distributions remains an important direction for future work.

REFERENCES

B. Anderson. Reverse-time diffusion equation models. *Stochastic Processes and their Applications*, 12(3):313–326, 1982.

E. Beyler and F. Bach. Convergence of deterministic and stochastic diffusion-model samplers: A simple analysis in wasserstein distance. *arXiv:2508.03210*, 2025.

A. Block, Y. Mroueh, and A. Rakhlin. Generative modeling with denoising auto-encoders and langevin sampling. *arXiv:2002.00107*, 2020.

G. Brigati and F. Pedrotti. Heat flow, log-concavity, and lipschitz transport maps. *arXiv:2404.15205*, 2024.

S. Bruno and S. Sabanis. Wasserstein convergence of score-based generative models under semi-convexity and discontinuous gradients. *TMLR*, 2025.

H. Chen, H. Lee, and J. Lu. Improved analysis of score-based generative modeling: User-friendly bounds under minimal smoothness assumptions. In *Proc. ICML*, pages 4735–4763, 2023a.

R. Chen, Y. Rubanova, J. Bettencourt, and D. Duvenaud. Neural ordinary differential equations. In *Proc. NIPS*, volume 31, 2018.

S. Chen, S. Chewi, H. Lee, Y. Li, J. Lu, and A. Salim. The probability flow ode is provably fast. In *Proc. NIPS*, volume 36, 2023b.

S. Chen, S. Chewi, J. Li, Y. Li, A. Salim, and A. Zhang. Sampling is as easy as learning the score: theory for diffusion models with minimal data assumptions. In *Proc. ICLR*, 2023c.

G. Conforti, A. Durmus, and M. Gentiloni-Silveri. Kl convergence guarantees for score diffusion models under minimal data assumptions. *SIAM J. Math. Data Sci.*, 7(1):86–109, 2025.

V. De Bortoli. Convergence of denoising diffusion models under the manifold hypothesis. *arXiv:2208.05314*, 2022.

V. De Bortoli, E. Mathieu, M. Hutchinson, J. Thornton, Y. Teh, and A. Doucet. Riemannian score-based generative modelling. In *Proc. NIPS*, volume 35, pages 2406–2422, 2022.

A. Eberle. Reflection coupling and wasserstein contractivity without convexity. *Comptes Rendus Mathematique*, 354(2):131–135, 2016.

X. Gao and L. Zhu. Convergence analysis for general probability flow odes of diffusion models in wasserstein distances. *arXiv:2401.17958*, 2024.

X. Gao, H. Nguyen, and L. Zhu. Wasserstein convergence guarantees for a general class of score-based generative models. *J. Mach. Learn. Res.*, 26(43):1–54, 2025.

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

J. Ho, A. Jain, and P. Abbeel. Denoising diffusion probabilistic models. In *Proc. NIPS*, volume 33, pages 6840–6851, 2020.

C. Huang, M. Aghajohari, J. Bose, P. Panangaden, and A. Courville. Riemannian diffusion models. In *Proc. NIPS*, volume 35, pages 2750–2761, 2022.

X. Huang, C. Salaun, C. Vasconcelos, C. Theobalt, C. Oztireli, and G. Singh. Blue noise for diffusion models. In *ACM SIGGRAPH 2024 conference papers*, pages 1–11, 2024.

D. Huk and T. Damoulas. Diffusion and flow-based copulas: Forgetting and remembering dependencies. In *Proc. ICLR*, 2026.

K. Ishige. Eventual concavity properties of the heat flow. *Mathematische Annalen*, 390(4):5883–5922, 2024.

N. Jia, T. Zhu, H. Liu, and Z. Zheng. Structured diffusion models with mixture of gaussians as prior distribution. *arXiv:2410.19149*, 2024.

B. Jing, G. Corso, R. Berlinghieri, and T. Jaakkola. Subspace diffusion generative models. In *European conference on computer vision*, pages 274–289. Springer, 2022.

P. Kenneweg, R. Dandinasivara, X. Luo, B. Hammer, and A. Schönhuth. Generating synthetic genotypes using diffusion models. *Bioinformatics*, 41(Supplement_1):i484–i492, 2025.

G. Kremling, F. Iafrate, M. Taheri, and J. Lederer. Non-asymptotic error bounds for probability flow odes under weak log-concavity. *arXiv:2510.17608*, 2025.

A. Lampis, M. Massi, N. Pirastu, F. Ieva, M. Matteucci, and E. Di Angelantonio. Snpgen: Phenotype-supervised genotype representation and synthetic data generation via latent diffusion. *arXiv:2603.10873*, 2026.

H. Lee, J. Lu, and Y. Tan. Convergence for score-based generative modeling with polynomial complexity. In *Proc. NIPS*, volume 35, pages 22870–22882, 2022.

H. Lee, J. Lu, and Y. Tan. Convergence of score-based generative modeling for general data distributions. *ALT*, pages 946–985, 2023.

Y. Li, S. Liu, S. Zhan, Z. Fang, and T. Ji. Poster: Genomic data generation via correlation-guided and privacy-preserving diffusion. In *Proceedings of the Network and Distributed System Security Symposium (NDSS)*, 2025. Poster.

Y. Liang, P. Ju, Y. Liang, and N. Shroff. Non-asymptotic convergence of discrete-time diffusion models: New approach and improved rate. In *Proc. ICLR*, 2024.

Y. Lipman, R. Chen, H. Ben-Hamu, M. Nickel, and M. Le. Flow matching for generative modeling. *arXiv:2210.02747*, 2023.

P. Liu, Z. Li, and X. Cheng. Variational trajectory optimization of anisotropic diffusion schedules. *arXiv:2602.19512*, 2026.

G. Luo, Frank Cole, S. Zhang, Y. Wan, Y. Lu, and J. Sun. Flow matching for multimodal distributions. In *CVPR*, pages 23260–23271, 2026.

Y. Ma, T. Chen, and E. Fox. A complete recipe for stochastic gradient mcmc. In *Proc. NIPS*, volume 28, 2015.

T. Nguyen, H. Nguyen, F. Chamroukhi, and G. McLachlan. Approximation by finite mixtures of continuous density functions that vanish at infinity. *Cogent Mathematics & Statistics*, 7(1): 1750861, 2020.

F. Paul, Y. Arkin, et al. Transcriptional heterogeneity and lineage commitment in myeloid progenitors. *Nature Biotechnology*, 2015.

R. T. Sadia, M. A. Ahamed, and Qian Cheng. Causalgendiff: Generative causal diffusion bridges scrna-seq and spatial transcriptomics. *Journal of Biomedical Informatics*, 2025.

Q. Si, P. Wang, Y. Wu, Y. Jiao, X. Liu, X. Guo, Y. Qi, and Y. Cheng. Flag: Foundation model representation with latent diffusion alignment via graph for spatial gene expression prediction. *arXiv:2605.18055*, 2026.

M. Silveri and A. Ocello. Beyond log-concavity and score regularity: Improved convergence bounds for score-based generative models in w2-distance. In *Proc. ICML*, 2025.

Y. Song and S. Ermon. Generative modeling by estimating gradients of the data distribution. In *Proc. NIPS*, volume 32, 2019.

Y. Song, J. Sohl-Dickstein, D. Kingma, A. Kumar, S. Ermon, and B. Poole. Score-based generative modeling through stochastic differential equations. In *Proc. ICLR*, 2021.

M. Taheri and J. Lederer. Regularization can make diffusion models more efficient. *arXiv:2502.09151*, 2025.

B. Tzen and M. Raginsky. Nonconvex learning via stochastic gradient langevin dynamics: A nonasymptotic analysis. In *Proc. COLT*, pages 313–338, 2019.

J. Vandersanden, S. Holl, X. Huang, and G. Singh. Edge-preserving noise for diffusion models. In *Computer Graphics Forum*, page e70383. Wiley Online Library, 2024.

A. Wibisono and K. Yang. Convergence in kl divergence of the inexact langevin algorithm with application to score-based generative models. *arXiv:2211.01512*, 2022.

G. Zheng, J. Terry, P. Belgrader, P. Ryvkin, Z. Bent, R. Wilson, S. Ziraldo, T. Wheeler, G. McDermott, J. Zhu, et al. Massively parallel digital transcriptional profiling of single cells. *Nature communications*, 8(1):14049, 2017.

A Discussion: advantages of structured drift over isotropic diffusion

The GOU framework, through its structured forward and backward process, offers several significant advantages over conventional isotropic diffusion models. These benefits stem directly from the geometry-aware drift matrix M, which encodes the covariance structure of the data.

1. **Mode preservation.** By incorporating the data geometry through the drift matrix M, the diffusion process rapidly contracts low-variance directions toward Gaussian noise while preserving separation along high-variance directions (Corollary 1). As a result, distinct modes remain distinguishable for longer time during the forward process (Theorem 4), allowing the reverse process to recover them more faithfully. Moreover, the score matching objective effectively reduces to a low-dimensional problem, since the scores in directions that quickly become Gaussian are already well-known. In contrast, isotropic diffusion contracts all directions uniformly, causing premature mode mixing and potentially leading to mode collapse.
2. **Preservation of correlation structure.** The anisotropic drift enables feature-specific evolution rates, allowing the process to retain critical statistical dependencies encoded in the data covariance. Theorem 3 demonstrates that the covariance Σ_t evolves through a combination of transported initial correlations and geometry-aware noise injection, thereby preserving key structural properties of the data (Remark 1). In contrast, isotropic diffusion treats all dimensions identically, failing to distinguish informative correlations from noise and disrupting the underlying data structure.
3. **Initialization error is governed by local variance.** While standard diffusion error scales with the overall spread of the dataset, GOU isolates the initial mismatch to localized variances around structural anchors (Lemma 2). This structural alignment dramatically reduces initial error and ensures rapid contractive convergence.

These properties establish the GOU framework as a principled extension of diffusion-based generative modeling to structured data. By respecting the intrinsic geometry of the data distribution—rather than treating all directions uniformly—the method achieves better mode coverage, preserves statistical dependencies, improves initialization error, and as a conjuncture converges under milder theoretical conditions. These advantages position GOU as particularly well-suited for scientific applications where correlations (e.g., between genes, SNPs, or physical measurements) carry essential information, and where mode collapse would compromise downstream analysis.

While our GOU process shares a similar motivation with Riemannian diffusion (De Bortoli et al., 2022; Huang et al., 2022)—namely, to incorporate geometric structure into the diffusion dynamics—the two approaches differ fundamentally in how this geometry is modeled. Riemannian diffusion introduces a position-dependent metric tensor, leading to locally adaptive and nonlinear diffusion that reflects the curvature of an underlying manifold. In contrast, our approach employs a fixed, data-driven linear operator derived from the covariance structure of the target distribution.

As a result, the induced geometry is global rather than local, and the dynamics remain analytically tractable. Despite this simplification, the GOU process captures key aspects of anisotropic behavior by aligning the diffusion with principal directions of the data, thereby preserving important structural features such as mode separation. In this sense, our method can be viewed as a linear and computationally efficient approximation to more general geometry-aware diffusion processes.

B Further technical results and proofs

We first present some further theoretical results, followed by the proofs of the principal theorems.

Lemma 3 (Matrix exponential of a diagonalizable matrix)**.** *Let $M \in \mathbb{R}^{d\times d}$ be a symmetric matrix with eigendecomposition*

$$M = U\,\mathrm{diag}(c_1,\dots,c_d)\,U^\top,$$

where U is orthogonal and $c_1,\dots,c_d \in \mathbb{R}$. Then the matrix exponential satisfies

$$e^{Mt} = U\,\mathrm{diag}(e^{c_1 t},\dots,e^{c_d t})\,U^\top.$$

Theorem 4 (Fisher separability under GOU Process)**.** *Let $R_i(t)$ denote the Fisher Linear Discriminant ratio measuring mode separability along the principal component direction u_i:*

$$R_i(t) := \frac{|\langle \Delta(t), u_i\rangle|}{\sqrt{\langle \Sigma_t u_i, u_i\rangle}}\,. \tag{14}$$

Under the GOU forward process defined in equation 4, for large t and within high-variance directions (where $|c_i|$ is small), the squared Fisher ratio scales inversely with time ($\alpha_i = u_i^T\Delta_0$):

$$R_i^2(t) \approx \frac{\alpha_i^2}{\beta t}\,.$$

Consequently, the local modes remain distinctly separated along the u_i direction ($R_i^2(t) > 2$) for diffusion times satisfying:

$$t < \frac{\alpha_i^2}{2\beta}\,.$$

Theorem 4 establishes that the GOU process successfully preserves local mode boundaries by leveraging the data geometry. While standard isotropic diffusion blurs cluster structures uniformly across all dimensions, our anisotropic framework creates protective "channels" along high-variance directions where the rate of mode mixing is significantly slowed down. Specifically, the theorem provides an explicit upper bound demonstrating that local modes remain statistically distinct for extended diffusion durations proportional to their initial separation (Δ_0) and inversely proportional to the injected noise (β). Intuitively, the condition $R_i^2(t) > 2$ implies that the squared distance between the two mode centers remains more than twice as large as the variance along that direction. This ensures that the physical separation between the modes easily dominates the blurring effect of the diffusion process, preventing them from merging.

B.1 Proof of Theorem 1

Proof. We study the evolution of the mean of the solution X_t of the SDE defined in equation 4 in 4 steps:

Step 1 (ODE for the mean): Consider the SDE

$$\mathrm{d}X_t = M(X_t-\mu)\,\mathrm{d}t + \sqrt{\beta}\,\mathrm{d}W_t, \qquad X_0 \text{ has finite first moment.}$$

Integrating both sides from 0 to t yields

$$\int_0^t \mathrm{d}X_s = \int_0^t M(X_s-\mu)\,\mathrm{d}s + \sqrt{\beta}\int_0^t \mathrm{d}W_s\,.$$

Using the identities (recall $W_0 = 0$)

$$\int_0^t \mathrm{d}X_s = X_t - X_0\,, \qquad \int_0^t \mathrm{d}W_s = W_t\,,$$

we obtain the integral representation

$$X_t - X_0 = \int_0^t M(X_s - \mu)\,\mathrm{d}s + \sqrt{\beta}\,W_t\,.$$

Taking expectations from both sides, using linearity of expectation, together with $\mathbb{E}[W_t] = 0$, we obtain

$$\mathbb{E}[X_t] - \mathbb{E}[X_0] = \int_0^t M(\mathbb{E}[X_s] - \mu)\,\mathrm{d}s\,.$$

The interchange of expectation and integration is justified by Fubini's theorem since

$$\int_0^t \mathbb{E}\big[|X_s|\big]\,\mathrm{d}s < \infty\,.$$

Differentiating both sides with respect to t then gives the ODE for the mean:

$$\frac{\mathrm{d}}{\mathrm{d}t}\mathbb{E}[X_t] = M(\mathbb{E}[X_t] - \mu)\,, \qquad \mathbb{E}[X_0] = \mu_0\,.$$

Step 2 (Difference of mean of two solutions): Consider two solutions of the SDE with initial means $\mu_1(0)$ and $\mu_2(0)$, denoted by $X_t^{(1)}$ and $X_t^{(2)}$. Let the mean difference be $\Delta(t) := \mathbb{E}[X_t^{(1)}] - \mathbb{E}[X_t^{(2)}]$. Each mean satisfies

$$\frac{\mathrm{d}}{\mathrm{d}t}\mathbb{E}[X_t^{(1)}] = M(\mathbb{E}[X_t^{(1)}] - \mu)\,, \qquad \frac{\mathrm{d}}{\mathrm{d}t}\mathbb{E}[X_t^{(2)}] = M(\mathbb{E}[X_t^{(2)}] - \mu)\,.$$

Subtracting gives a linear ODE for the mean difference:

$$\frac{\mathrm{d}}{\mathrm{d}t}\Delta(t) = M\Delta(t), \qquad \Delta(0) = \mu_1(0) - \mu_2(0) = \Delta_0\,.$$

Step 3 (Solution of the linear ODE): The solution of this linear ODE with constant coefficients is

$$\Delta(t) = e^{Mt}\Delta_0\,.$$

Step 4 (Componentwise representation): Diagonalize $M = UCU^\top$, where $C = \mathrm{diag}(c_1, \ldots, c_d)$ and U is an orthogonal matrix of eigenvectors. Using Lemma 3 in the Appendix,

$$\Delta(t) = Ue^{Ct}U^\top\Delta_0$$

and the squared norm can be written as

$$\|\Delta(t)\|^2 = \|Ue^{Ct}U^\top\Delta_0\|^2 = \sum_{i=1}^{d}\alpha_i^2 e^{2c_i t}\,.$$

This explicitly shows how each component of the initial mean difference evolves along the eigendirections of M, concluding the proof. □

B.2 Proof of Theorem 2

Proof. Taking expectations on both sides of the SDE equation 4 yields

$$\frac{\mathrm{d}}{\mathrm{d}t}\mathbb{E}[X_t] = M(\mathbb{E}[X_t] - \mu)\,,$$

where we used $\mathbb{E}[\mathrm{d}W_t] = 0$ (see the proof of Theorem 1 for further details).

For $\mu_t = \mathbb{E}[X_t]$, it satisfies the linear ODE

$$\frac{\mathrm{d}\mu_t}{\mathrm{d}t} = M(\mu_t - \mu)\,, \qquad \mu_0 = \mathbb{E}[X_0]\,.$$

Define $\tilde{\mu}_t := \mu_t - \mu$. Then $\tilde{\mu}_t$ satisfies

$$\frac{\mathrm{d}\tilde{\mu}_t}{\mathrm{d}t} = M\tilde{\mu}_t, \qquad \tilde{\mu}_0 = \mu_0 - \mu\,.$$

The unique solution is given by

$$\tilde{\mu}_t = e^{Mt}(\mu_0 - \mu)\,.$$

Therefore,

$$\mu_t = \mu + e^{Mt}(\mu_0 - \mu)\,,$$

which proves equation 5.

Employing Lemma 3 gives equation 6, as desired.

□

B.3 Proof of Theorem 3

Proof. We do the proof in four steps:

Step 1: Centering the process.

Let $\mu_t = \mathbb{E}[X_t]$ and define the centered process

$$\bar{X}_t := X_t - \mu_t\,.$$

By construction, $\mathbb{E}[\bar{X}_t] = 0$.

Using the SDE for X_t and the dynamics of μ_t (similar approach as in the proof of Theorem 2), we compute

$$\mathrm{d}\bar{X}_t = \mathrm{d}X_t - \mathrm{d}\mu_t\,.$$

Substituting,

$$\begin{aligned}\mathrm{d}\bar{X}_t &= M(X_t - \mu)\,\mathrm{d}t + \sqrt{\beta}\,\mathrm{d}W_t - M(\mu_t - \mu)\,\mathrm{d}t \\ &= M\big[(X_t - \mu) - (\mu_t - \mu)\big]\,\mathrm{d}t + \sqrt{\beta}\,\mathrm{d}W_t \\ &= M\bar{X}_t\,\mathrm{d}t + \sqrt{\beta}\,\mathrm{d}W_t\,.\end{aligned}$$

Step 2: Applying Itô's formula.

We now study the matrix-valued process $\bar{X}_t\bar{X}_t^\top$. By Itô's product rule,

$$\mathrm{d}\big(\bar{X}_t\bar{X}_t^\top\big) = \mathrm{d}\bar{X}_t\,\bar{X}_t^\top + \bar{X}_t\,\mathrm{d}\bar{X}_t^\top + \mathrm{d}\bar{X}_t\,\mathrm{d}\bar{X}_t^\top\,.$$

We compute each term separately.

Using $\mathrm{d}\bar{X}_t = M\bar{X}_t\,\mathrm{d}t + \sqrt{\beta}\,\mathrm{d}W_t$, we obtain

$$\begin{aligned}\mathrm{d}\bar{X}_t\,\bar{X}_t^\top &= (M\bar{X}_t\,\mathrm{d}t + \sqrt{\beta}\,\mathrm{d}W_t)\,\bar{X}_t^\top \\ &= M\bar{X}_t\bar{X}_t^\top\,\mathrm{d}t + \sqrt{\beta}\,\mathrm{d}W_t\,\bar{X}_t^\top\,,\end{aligned}$$

and similarly,

$$\begin{aligned}\bar{X}_t\,\mathrm{d}\bar{X}_t^\top &= \bar{X}_t(\bar{X}_t^\top M^\top \mathrm{d}t + \sqrt{\beta}\,\mathrm{d}W_t^\top) \\ &= \bar{X}_t\bar{X}_t^\top M^\top \mathrm{d}t + \sqrt{\beta}\,\bar{X}_t\mathrm{d}W_t^\top\,.\end{aligned}$$

We now compute the quadratic variation term $\mathrm{d}\bar{X}_t\,\mathrm{d}\bar{X}_t^\top$ explicitly.

First, expand the product:

$$\begin{aligned}\mathrm{d}\bar{X}_t\,\mathrm{d}\bar{X}_t^\top &= (M\bar{X}_t\,\mathrm{d}t + \sqrt{\beta}\,\mathrm{d}W_t)(M\bar{X}_t\,\mathrm{d}t + \sqrt{\beta}\,\mathrm{d}W_t)^\top \\ &= (M\bar{X}_t\,\mathrm{d}t)(M\bar{X}_t\,\mathrm{d}t)^\top \\ &\quad + (M\bar{X}_t\,\mathrm{d}t)(\sqrt{\beta}\,\mathrm{d}W_t)^\top \\ &\quad + (\sqrt{\beta}\,\mathrm{d}W_t)(M\bar{X}_t\,\mathrm{d}t)^\top \\ &\quad + (\sqrt{\beta}\,\mathrm{d}W_t)(\sqrt{\beta}\,\mathrm{d}W_t)^\top\,.\end{aligned}$$

We now analyze each term separately using standard Itô rules.

(i) Drift–drift term.

$$(M\bar{X}_t\,\mathrm{d}t)(M\bar{X}_t\,\mathrm{d}t)^\top = M\bar{X}_t\bar{X}_t^\top M^\top\,\mathrm{d}t^2\,.$$

Since $\mathrm{d}t^2 = 0$ in Itô calculus, this term vanishes.

(ii) Drift–noise term.

$$(M\bar{X}_t\,\mathrm{d}t)(\sqrt{\beta}\,\mathrm{d}W_t)^\top = \sqrt{\beta}\,M\bar{X}_t\,\mathrm{d}t\,\mathrm{d}W_t^\top\,.$$

Since $\mathrm{d}t\,\mathrm{d}W_t = 0$, this term vanishes.

(iii) Noise–drift term.

$$(\sqrt{\beta}\,\mathrm{d}W_t)(M\bar{X}_t\,\mathrm{d}t)^\top = \sqrt{\beta}\,\mathrm{d}W_t\,\mathrm{d}t\,\bar{X}_t^\top M^\top\,.$$

Since $\mathrm{d}W_t\,\mathrm{d}t = 0$, this term also vanishes.

(iv) Noise–noise term.

$$(\sqrt{\beta}\,\mathrm{d}W_t)(\sqrt{\beta}\,\mathrm{d}W_t)^\top = \beta\,\mathrm{d}W_t\mathrm{d}W_t^\top\,.$$

Using the quadratic variation identity of Brownian motion,

$$\mathrm{d}W_t\mathrm{d}W_t^\top = \mathbb{I}_d\,\mathrm{d}t\,,$$

we obtain

$$(\sqrt{\beta}\,\mathrm{d}W_t)(\sqrt{\beta}\,\mathrm{d}W_t)^\top = \beta\mathbb{I}_d\,\mathrm{d}t\,.$$

Combining all four terms yields

$$\mathrm{d}\bar{X}_t\,\mathrm{d}\bar{X}_t^\top = \beta\mathbb{I}_d\,\mathrm{d}t\,.$$

Summing all contributions,

$$\begin{aligned}\mathrm{d}\big(\bar{X}_t\bar{X}_t^\top\big) = {} & M\bar{X}_t\bar{X}_t^\top\mathrm{d}t + \bar{X}_t\bar{X}_t^\top M^\top\mathrm{d}t \\ & + \sqrt{\beta}\,\mathrm{d}W_t\bar{X}_t^\top + \sqrt{\beta}\,\bar{X}_t\mathrm{d}W_t^\top + \beta\mathbb{I}_d\mathrm{d}t\,.\end{aligned}$$

Step 3: Taking expectations.

Define the covariance matrix

$$\Sigma_t := \mathbb{E}[\bar{X}_t\bar{X}_t^\top]\,.$$

Taking expectations and using $\mathbb{E}[\mathrm{d}W_t] = 0$, the stochastic integral terms vanish:

$$\mathbb{E}[\mathrm{d}W_t\bar{X}_t^\top] = 0, \quad \mathbb{E}[\bar{X}_t\mathrm{d}W_t^\top] = 0\,.$$

Hence,

$$\mathbb{E}\Big[\mathrm{d}\big(\bar{X}_t\bar{X}_t^\top\big)\Big] = M\Sigma_t\mathrm{d}t + \Sigma_t M^\top\mathrm{d}t + \beta\mathbb{I}_d\mathrm{d}t\,.$$

Since $\bar{X}_t$ has finite second moments, we can interchange expectation and differentiation, yielding

$$\frac{\mathrm{d}}{\mathrm{d}t}\Sigma_t = M\Sigma_t + \Sigma_t M^\top + \beta\mathbb{I}_d, \qquad \Sigma_0 = \mathrm{Cov}(X_0)\,.$$

Step 4: Solving the matrix differential equation.

Define

$$A_t := e^{-Mt}\Sigma_t e^{-M^\top t}\,.$$

We differentiate using the product rule:

$$\begin{aligned}\frac{\mathrm{d}}{\mathrm{d}t}A_t &= (-Me^{-Mt})\Sigma_t e^{-M^\top t} + e^{-Mt}\dot{\Sigma}_t e^{-M^\top t} + e^{-Mt}\Sigma_t(-M^\top e^{-M^\top t}) \\ &= e^{-Mt}\left(\dot{\Sigma}_t - M\Sigma_t - \Sigma_t M^\top\right)e^{-M^\top t}\,.\end{aligned}$$

Substituting the differential equation for Σ_t,

$$\dot{\Sigma}_t - M\Sigma_t - \Sigma_t M^\top = \beta \mathbb{I}_d \,,$$

we obtain

$$\frac{\mathrm{d}}{\mathrm{d}t} A_t = \beta e^{-Mt} e^{-M^\top t} \,.$$

Integrating from 0 to t,

$$A_t = A_0 + \beta \int_0^t e^{-Ms} e^{-M^\top s} \mathrm{d}s \,.$$

Since $A_0 = \Sigma_0$, this gives

$$A_t = \Sigma_0 + \beta \int_0^t e^{-Ms} e^{-M^\top s} \mathrm{d}s \,.$$

Multiplying from the left by e^{Mt} and from the right by $e^{M^\top t}$ yields

$$\Sigma_t = e^{Mt} \Sigma_0 e^{M^\top t} + \beta \int_0^t e^{M(t-s)} e^{M^\top (t-s)} \mathrm{d}s \,,$$

which is the desired result.

□

B.4 Proof of Lemma 3

Proof. By definition of the matrix exponential,

$$e^{Mt} := \sum_{k=0}^{\infty} \frac{(Mt)^k}{k!} = \sum_{k=0}^{\infty} \frac{t^k M^k}{k!}.$$

We first show that

$$M^k = U \,\mathrm{diag}(c_1^k, \ldots, c_d^k)\, U^\top.$$

We proceed by induction. For $k = 1$, the statement is trivial. Assume it holds for some $k \geq 1$. Then

$$M^{k+1} = M^k M = \left(U \Lambda^k U^\top\right) \left(U \Lambda U^\top\right) = U \Lambda^k (U^\top U) \Lambda U^\top = U \Lambda^{k+1} U^\top,$$

where $\Lambda = \mathrm{diag}(c_1, \ldots, c_d)$ and we used $U^\top U = I$. Hence the claim holds for all $k \geq 0$.

Substituting into the series expansion yields

$$e^{Mt} = \sum_{k=0}^{\infty} \frac{t^k}{k!} U \Lambda^k U^\top = U \left(\sum_{k=0}^{\infty} \frac{t^k \Lambda^k}{k!} \right) U^\top.$$

Since Λ is diagonal, powers act entrywise:

$$\sum_{k=0}^{\infty} \frac{t^k \Lambda^k}{k!} = \mathrm{diag}\left(\sum_{k=0}^{\infty} \frac{(c_1 t)^k}{k!}, \ldots, \sum_{k=0}^{\infty} \frac{(c_d t)^k}{k!} \right).$$

Each scalar series equals the exponential function, i.e.,

$$\sum_{k=0}^{\infty} \frac{(c_i t)^k}{k!} = e^{c_i t}.$$

Therefore,

$$e^{Mt} = U \,\mathrm{diag}(e^{c_1 t}, \ldots, e^{c_d t}) U^\top.$$

□

B.5 Proof of Lemma 1

Proof. From Theorem 3, we have

$$\Sigma_t = e^{Mt}\Sigma_0 e^{M^\top t} + \beta \int_0^t e^{M(t-s)} e^{M^\top (t-s)} \, \mathrm{d}s \, .$$

We analyze the limit $t \to \infty$.

Step 1: Decay of the initial condition.

Since M is negative definite, all eigenvalues have strictly negative real parts, hence $\lim_{t\to\infty} e^{Mt}$ goes to zero matrix, which implies

$$\lim_{t\to\infty} e^{Mt}\Sigma_0 e^{M^\top t} = 0_{d\times d} \, .$$

Step 2: Limit of the noise term.

Define

$$\Sigma_\infty := \beta \int_0^\infty e^{Mu} e^{M^\top u} \, \mathrm{d}u \, .$$

The integral is well-defined because e^{Mu} decays exponentially fast as $u \to \infty$.

Step 3: Lyapunov equation.

Let

$$J := \int_0^\infty e^{Mu} e^{M^\top u} \, \mathrm{d}u \, .$$

We compute

$$\frac{\mathrm{d}}{\mathrm{d}u}\big(e^{Mu} e^{M^\top u}\big) = M e^{Mu} e^{M^\top u} + e^{Mu} e^{M^\top u} M^\top \, .$$

Integrating over $[0, \infty)$ yields

$$\big[e^{Mu} e^{M^\top u}\big]_0^\infty = MJ + JM^\top \, .$$

Since $e^{Mu} e^{M^\top u} \to 0$ as $u \to \infty$ and equals $\mathbb{I}_d$ at $u = 0$, we obtain

$$MJ + JM^\top = -\mathbb{I}_d \, . \tag{15}$$

This is a continuous-time Lyapunov equation. Because M is negative definite, the solution J is unique.

Step 4: Identification of the stationary covariance.

We have

$$\Sigma_\infty = \beta J \, ,$$

and multiplying equation 15 by β gives

$$M\Sigma_\infty + \Sigma_\infty M^\top + \beta \mathbb{I}_d = 0 \, .$$

Step 5: Symmetric case.

For M symmetric, then $M = M^\top$ and the Lyapunov equation becomes

$$M\Sigma_\infty + \Sigma_\infty M + \beta \mathbb{I}_d = 0 \, .$$

Since M and Σ_∞ commute in this case (they share eigenvectors), we get

$$2M\Sigma_\infty + \beta \mathbb{I}_d = 0 \, ,$$

hence

$$\Sigma_\infty = -\frac{\beta}{2} M^{-1} \, .$$

This completes the proof. □

B.6 Proof of Lemma 2

Proof. By definition, $\mathcal{W}_2^2(p_T, \hat{p}_T)$ is defined as the infimum of the expected squared Euclidean distance over all valid joint distributions whose marginals match p_T and $\hat{p}_T$. We upper-bound this distance by constructing a specific synchronous coupling between the true forward trajectory X_T and the generative initialization trajectory $\hat{X}_T$, where the pair of initial states $(X_0, \hat{X}_0)$ is drawn from the anchor assignment distribution π.

Recalling the SDE definition in equation 4, the exact solution for the true terminal state X_T given an initial data point X_0 is:

$$X_T = \mu + e^{MT}(X_0 - \mu) + \sqrt{\beta}\int_0^T e^{M(T-s)}\,\mathrm{d}W_s\,. \tag{16}$$

We initialize the approximate trajectory using the corresponding anchor $\hat{X}_0$ subjected to the identical realization of the Brownian motion W_s:

$$\hat{X}_T = \mu + e^{MT}(\hat{X}_0 - \mu) + \sqrt{\beta}\int_0^T e^{M(T-s)}\,\mathrm{d}W_s\,. \tag{17}$$

Since the marginal distributions of X_T and $\hat{X}_T$ under this coupling are p_T and $\hat{p}_T$ respectively, we have:

$$\mathcal{W}_2^2(p_T, \hat{p}_T) \le \mathbb{E}_{(X_0,\hat{X}_0)\sim\pi}\big[\|X_T - \hat{X}_T\|^2\big]\,.$$

Subtracting equation 17 from equation 16, the stochastic integral terms cancel out identically, yielding a purely deterministic structural error expression for the coupled states:

$$X_T - \hat{X}_T = e^{MT}(X_0 - \hat{X}_0)\,.$$

Taking the expectation over the joint distribution $\pi(X_0, \hat{X}_0)$ of the squared Euclidean norm on both sides, we obtain:

$$\mathbb{E}_{(X_0,\hat{X}_0)\sim\pi}\big[\|X_T - \hat{X}_T\|^2\big] = \mathbb{E}_{(X_0,\hat{X}_0)\sim\pi}\big[\|e^{MT}(X_0 - \hat{X}_0)\|^2\big]\,.$$

To explicitly extract the final bound involving $c_{\max}$, we evaluate the matrix exponential using the spectral decomposition of the symmetric drift matrix $M = U\,\mathrm{diag}(c_1, \ldots, c_d)U^\top$, where U is an orthogonal matrix preserving the Euclidean norm:

$$\begin{aligned}\mathbb{E}_{(X_0,\hat{X}_0)\sim\pi}\big[\|e^{MT}(X_0 - \hat{X}_0)\|^2\big] &= \mathbb{E}_{(X_0,\hat{X}_0)\sim\pi}\big[\|U\,\mathrm{diag}\left(e^{c_1 T}, \ldots, e^{c_d T}\right)U^\top(X_0 - \hat{X}_0)\|^2\big] \\ &= \sum_{i=1}^{d} e^{2c_i T}\mathbb{E}_{(X_0,\hat{X}_0)\sim\pi}\big[|(U^\top X_0)_i - (U^\top \hat{X}_0)_i|^2\big]\,.\end{aligned}$$

Since $c_i < 0$ for all directions, meaning the exponential decay is strictly bounded above by the maximum (least negative) eigenvalue $c_{\max} = \max_i c_i$. Factoring this term out of the summation yields:

$$\begin{aligned}\sum_{i=1}^{d} e^{2c_i T}\mathbb{E}_{(X_0,\hat{X}_0)\sim\pi}\big[|(U^\top X_0)_i - (U^\top \hat{X}_0)_i|^2\big] &\le e^{2c_{\max}T}\sum_{i=1}^{d}\mathbb{E}_{(X_0,\hat{X}_0)\sim\pi}\big[|(U^\top X_0)_i - (U^\top \hat{X}_0)_i|^2\big] \\ &= e^{2c_{\max}T}\mathbb{E}_{(X_0,\hat{X}_0)\sim\pi}\big[\|U^\top(X_0 - \hat{X}_0)\|^2\big] \\ &= e^{2c_{\max}T}\mathbb{E}_{(X_0,\hat{X}_0)\sim\pi}\big[\|X_0 - \hat{X}_0\|^2\big]\,.\end{aligned}$$

Taking the square root on both sides directly produces the final contractive metric bound:

$$\mathcal{W}_2(p_T, \hat{p}_T) \le e^{c_{\max}T}\sqrt{\mathbb{E}_{(X_0,\hat{X}_0)\sim\pi}\big[\|X_0 - \hat{X}_0\|^2\big]}\,.$$

This completes the proof. □

B.7 Proof of Theorem 4

Proof. A standard metric for mode separability along a directional vector v (such as a principal component u_i) is the Fisher Linear Discriminant ratio:

$$R_i(t) = \frac{|\langle \Delta(t), u_i \rangle|}{\sqrt{\langle \Sigma_t u_i, u_i \rangle}},$$

where the numerator computes the distance of the difference vector when projected onto the direction u_i, and the denominator extracts the scalar variance along that same axis.

Form Theorem 1, the projection of the separation along u_i is

$$|\langle \Delta(t), u_i \rangle| = |\alpha_i| e^{c_i t},$$

where $\alpha_i = u_i^T \Delta_0$.

Also from Theorem 3 and for large t

$$\langle \Sigma_t u_i, u_i \rangle \approx \beta \int_0^t e^{2c_i(t-s)} \mathrm{d}s = \frac{\beta}{2|c_i|}(1 - e^{2c_i t}).$$

Collecting the pieces, we obtain

$$R_i^2(t) = \frac{2|c_i|\alpha_i^2}{\beta(e^{2|c_i|t} - 1)}.$$

In high-variance directions (small $|c_i|$), we employ the fact that $e^x \approx 1 + x$ to obtain

$$R_i^2(t) \approx \frac{2|c_i|\alpha_i^2}{2\beta|c_i|t} = \frac{\alpha_i^2}{\beta t}.$$

This implies that modes keep separated as long as $\alpha_i^2/\beta t > 2$ that holds for $t < \alpha_i^2/\beta^2$. □

C Complementary empirical studies

To further validate our approach, we present extensive empirical evaluations on various simulated and real-world datasets.

Performance across structural metrics and out-of-batch experiments on Paul15 To evaluate generative fidelity and domain generalization, we benchmarked our structured drift model against isotropic diffusion across structural biological metrics and a zero-shot out-of-batch experiment. On full dataset evaluations, structured drift significantly enhanced network structure preservation, reducing Gene Regulatory Network (GRN) Graph Laplacian error by over 90% (0.0637 vs. 0.6877) and improving gene co-expression Spearman correlation from 0.8048 to a near-perfect 0.9903. When evaluated on a held-out experimental batch to test out-of-distribution generalization, structured drift reduced Maximum Mean Discrepancy (MMD) by over 75% (0.0136 vs. 0.0549) and maintained robust gene co-expression rank correlation (0.7557 vs. 0.4308). These findings demonstrate that explicitly incorporating structural priors into the drift operator successfully preserves global regulatory topology while preventing the generation of non-biological batch artifacts.

Mixture of three Gaussians To evaluate the performance of the GOU process against isotropic drift, we conduct a 2D synthetic experiment on a three-component Gaussian mixture model ($n = 2000$). For each approach, a neural score network is trained to approximate the score function and generate synthetic samples via reverse sampling. As shown in Figure 2, GOU consistently yields samples closer to the target distribution. Quantitatively, GOU achieves superior statistical fidelity over isotropic drift, substantially reducing both Gaussian $\mathcal{W}_2$ (0.032 vs. 0.489) and Sliced Wasserstein-2 distance (0.114 vs. 0.671). Furthermore, GOU achieves higher manifold recall (0.996 vs. 0.986) and stays closer to the optimal value for classifier two-sample AUC (0.514 vs. 0.663).

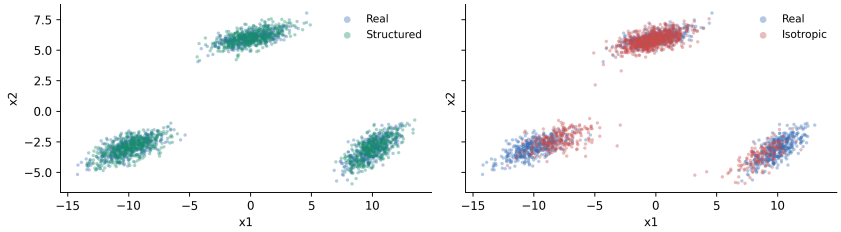


Figure 2: Qualitative comparison of synthetic samples generated via the GOU process and isotropic drift against the ground-truth 2D three-component Gaussian mixture dataset.

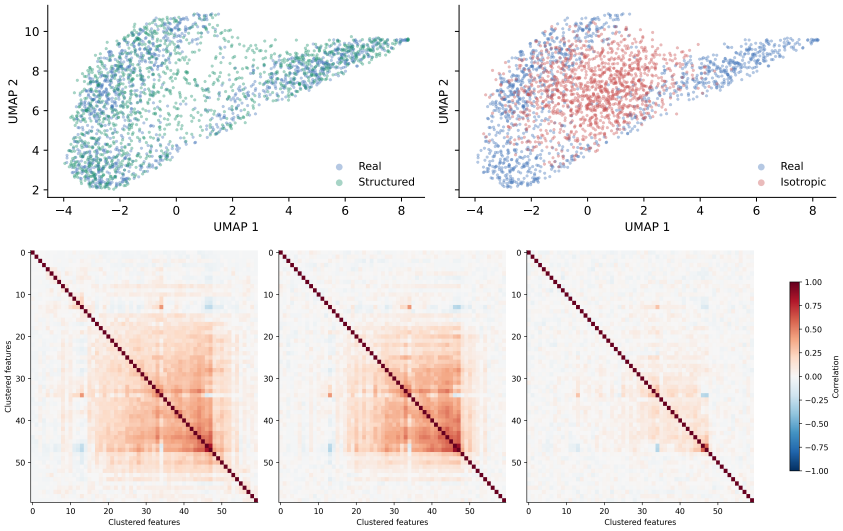


Figure 3: Synthetic sample fidelity and biological co-expression structure. (First row) Joint low-dimensional (UMAP) representation comparing generated samples from structured (left) and isotropic models (right) against real data. (Second row) Clustered feature correlation heatmaps across real (left), structured (middle), and isotropic (right) models for top 60 feature. The GOU accurately recovers real data distribution and feature-feature dependence and co-expression networks, whereas the isotropic baseline exhibits weak coverage and significant correlation attenuation.

Synthetic LD-like simulation To evaluate the models on data with dependency patterns, we generated a synthetic genetic dataset consisting of $n = 2000$ individuals and $d = 200$ single-nucleotide polymorphisms (SNPs). The SNPs were simulated to mimic the local dependency structure observed in real genomic data, where nearby variants exhibit stronger correlations due to linkage disequilibrium (LD). Specifically, the SNPs were partitioned into blocks of fixed size 50, and within each block correlated latent Gaussian variables were generated using an exponentially decaying covariance structure (decay rate 0.5). These latent variables were then discretized to obtain genotype values in $\{0, 1, 2\}$, corresponding to the number of minor alleles carried by an individual. This procedure produces a dataset with realistic marginal distributions and local correlation patterns similar to those observed in genomic studies. We then evaluate the performance of GOU vs isotropic diffusion models. Results are provided in Table 2 and demonstrate that structured modeling generally outperforms or matches isotropic baselines across both statistical fidelity and biological structure metrics. Specifically, structured methods show key gains in manifold recall and optimal transport distances (Gaussian and Sliced Wasserstein-2) while maintaining tighter confidence intervals.

Table 2: Primary statistical and biological fidelity metrics comparing structured and isotropic methods across 10 paired units over LD-like simulated dataset

Category	Metric	Ideal	Structured [95% CI]	Isotropic [95% CI]
Statistical	Classifier two-sample AUC	Near 0.5	0.536 [0.530, 0.543]	0.532 [0.527, 0.538]
	Gaussian W2 (FD)	Lower	0.220 [0.219, 0.222]	0.280 [0.279, 0.281]
	Sliced Wasserstein-2	Lower	0.191 [0.189, 0.193]	0.264 [0.262, 0.265]
	Manifold recall	Higher	0.333 [0.326, 0.342]	0.158 [0.153, 0.163]
Biological	Co-expression Spearman	Higher	0.092 [0.089, 0.096]	0.084 [0.080, 0.088]
	Cell-state proportion JSD	Lower	0.003 [0.002, 0.004]	0.004 [0.003, 0.005]
	Marker-effect Spearman	Higher	0.663 [0.648, 0.678]	0.671 [0.658, 0.684]

Single-cell PBMC transcriptomic dataset We evaluate our method on the PBMC3K dataset (Zheng et al., 2017), which contains approximately 2,700 single-cell RNA sequencing profiles across more than 13,000 genes. For simplicity, we restrict our analysis to the first 200 genes. The data exhibits clear heterogeneity across multiple cell types (including T cells, B cells, NK cells, and monocytes) and a moderate gene–gene correlation structure, making it a standard benchmark for evaluating generative models on high-dimensional and structured data. Table 3 summarizes our experimental results. Overall, the empirical evaluation demonstrates that the structured model consistently outperforms the isotropic baseline across key statistical and biological fidelity metrics; see Table 3. On statistical fidelity, the structured approach achieves better high-dimensional indistinguishability (Classifier two-sample AUC $= 0.534$ vs. 0.513), lower geometric distance (Gaussian $W_2 = 0.235$ vs. 0.317; Sliced $W_2 = 0.231$ vs. 0.322), and higher manifold coverage (Recall $= 0.514$ vs. 0.377), demonstrating superior alignment with the true distribution. Biologically, it more accurately preserves gene correlations (Co-expression Spearman $= 0.250$ vs. 0.168) and retains fine-grained population structures (Cell-state proportion JSD $= 0.022$ vs. 0.046), while maintaining competitive marker-effect recovery (0.611 vs. 0.617). Together, these results confirm that structured priors effectively capture key multivariate properties and global biological organization.

Table 3: Primary statistical and biological fidelity metrics comparing structured and isotropic methods across 10 paired units over PBMC dataset

Category	Metric	Ideal	Structured [95% CI]	Isotropic [95% CI]
Statistical	Classifier two-sample AUC	Near 0.5	0.534 [0.522, 0.546]	0.513 [0.508, 0.520]
	Gaussian W2 (FD)	Lower	0.235 [0.230, 0.240]	0.317 [0.309, 0.325]
	Sliced Wasserstein-2	Lower	0.231 [0.223, 0.240]	0.322 [0.312, 0.331]
	Manifold recall	Higher	0.514 [0.502, 0.526]	0.377 [0.370, 0.385]
Biological	Co-expression Spearman	Higher	0.250 [0.245, 0.255]	0.168 [0.166, 0.170]
	Cell-state proportion JSD	Lower	0.022 [0.018, 0.027]	0.046 [0.038, 0.054]
	Marker-effect Spearman	Higher	0.611 [0.578, 0.644]	0.617 [0.590, 0.639]