

Diffusion Paths: A Geodesic Lens on Noising Schedules

Sahil Akhtar¹

¹Department of Physics and Computer Science, MIT

Abstract

Diffusion models have emerged as a powerful generative modeling framework by learning to reverse a noise process over time. In this work, we investigate the geometry of the forward diffusion path in probability space and conduct a comprehensive empirical study of various noise schedules. This study is guided by interpreting the noising path in diffusion models as a geodesic in the probability space. We implemented and evaluated five distinct beta schedules: Linear, Cosine, Quadratic, Exponential, and a novel Wasserstein-2 (W_2) geodesic schedule. Our analysis reveals that the Linear schedule achieved the best overall performance in terms of test loss, while the Cosine schedule exhibited the best geodesic properties. The W_2 geodesic schedule, despite its theoretical optimality, achieved competitive but not optimal empirical performance. Our findings suggest that while geometric principles guide schedule design, empirical considerations remain crucial for practical applications. We provide a complete implementation framework with parallel evaluation capabilities and comprehensive visualization tools to support future research in diffusion model optimization.

1 Introduction

Diffusion models define a generative process by gradually transforming data into noise through a forward stochastic process, then learning to reverse this transformation. The forward process is typically governed by a *noise schedule*, defined by a sequence of β_t values controlling the variance injected at each time step.

While standard schedules, such as linear or cosine, have shown empirical success, they often lack a clear geometric interpretation. In this work, we interpret the forward process as a curve in the Wasserstein space of probability measures and examine how closely standard schedules approximate geodesics under the Wasserstein-2 (W_2) metric. We then derived and implemented an alternative noise schedule that explicitly follows the geodesic path from the data distribution to the final noise distribution.

The theoretical motivation for this approach stems from fundamental principles in physics and geometry. Just as in General Relativity where particles follow geodesics—paths of minimum distance and stable solutions of Lagrangian optimization—we hypothesize that the forward noising process in diffusion models, when designed to follow a geodesic in probability space, should inherently lead to a more stable and potentially easier-to-learn reverse process. This work thus draws an conceptual connection between the probability paths of diffusion models and the trajectories of particles in curved manifolds.

Our contributions include:

- A comprehensive empirical evaluation of five noise schedules on toy Gaussian Mixture Model (GMM) data.

- Implementation of a novel W_2 geodesic beta schedule with theoretical justification.
- Development of a parallel evaluation framework incorporating multiple metrics for the various noising schedules.
- Detailed analysis of geodesic properties and training dynamics.
- An open-source implementation with complete reproducibility.

2 Background

2.1 Diffusion Models

The forward process in a diffusion model adds Gaussian noise in T discrete steps:

$$q(\mathbf{x}_t \mid \mathbf{x}_{t-1}) = \mathcal{N}(\mathbf{x}_t; \sqrt{1 - \beta_t}\mathbf{x}_{t-1}, \beta_t\mathbf{I}),$$

with cumulative noise $\alpha_t = \prod_{s=1}^t (1 - \beta_s)$, and marginal distribution:

$$q(\mathbf{x}_t \mid \mathbf{x}_0) = \mathcal{N}(\sqrt{\alpha_t}\mathbf{x}_0, (1 - \alpha_t)\mathbf{I}).$$

The goal of the reverse process is to learn $p_\theta(\mathbf{x}_{t-1} \mid \mathbf{x}_t)$ via denoising.

2.2 Wasserstein-2 Geodesics

The space of probability distributions with finite second moment, $\mathcal{P}_2(\mathbb{R}^d)$, equipped with the W_2 distance, forms a geodesic space. For two Gaussians $\mathcal{N}(\mu_0, \Sigma_0)$ and $\mathcal{N}(\mu_1, \Sigma_1)$, the W_2 geodesic interpolates both mean and covariance:

$$\begin{aligned} \mu_t &= (1 - t)\mu_0 + t\mu_1, \\ \Sigma_t &= \left[(1 - t)\Sigma_0^{1/2} + t\Sigma_1^{1/2} \right]^2. \end{aligned}$$

In our case, we assume $\mu = 0$ throughout, with $\Sigma_0 = \varepsilon^2\mathbf{I}$ (tight Gaussian) and $\Sigma_1 = \mathbf{I}$ (standard noise).

This is motivated because it is a candidate for the "most" geodesic conditional path - it is the minimum length path (according to the metric in probability space) from the initial conditional data distribution (a delta function centered at the data point) and the final distribution (standard noise).

Our findings also reveal that the geodesic-character of a conditional path often translates over to the marginal path. If the conditional noise schedule incurs lesser norm in the probability space compared to other probability paths, then the marginal also performs empirically better. This is not guaranteed, and so more investigation is needed into whether optimality/theoretical sanctity of conditional paths translates over to the marginal path.

3 Methodology Overview

To investigate the impact of noise schedules on diffusion model performance and geodesic properties, we designed a comprehensive experimental setup. We implemented five distinct beta schedules: Linear, Cosine, Quadratic, Exponential, and our novel W_2 Geodesic schedule. These schedules are defined as follows:

- **Linear:** $\beta_t = \beta_{\text{start}} + (\beta_{\text{end}} - \beta_{\text{start}}) \cdot t/T$
- **Cosine:** $\beta_t = \beta_{\text{start}} + (\beta_{\text{end}} - \beta_{\text{start}}) \cdot (1 - \cos(\pi t/2T))$
- **Quadratic:** $\beta_t = \beta_{\text{start}} + (\beta_{\text{end}} - \beta_{\text{start}}) \cdot (t/T)^2$
- **Exponential:** $\beta_t = \exp(\log(\beta_{\text{start}}) + (\log(\beta_{\text{end}}) - \log(\beta_{\text{start}})) \cdot t/T)$
- **W_2 Geodesic:** $\sigma_t = (1 - t)\varepsilon + t$, then $\alpha_t = 1 - \sigma_t^2$

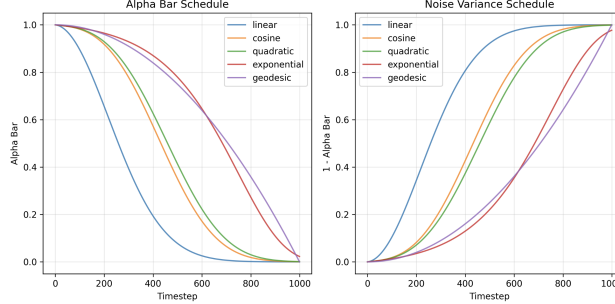


Figure 1: Comparison of the various noise schedules.

For the denoising task, we utilized a 4-layer MLP model trained on a 2D Gaussian Mixture Model (GMM) dataset.

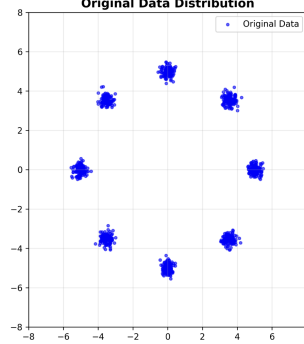


Figure 2: Original data distribution is a 2D GMM

Our evaluation framework allowed for parallel processing and tracked performance across several key metrics, including Test Loss (Mean Squared Error), Maximum Mean Discrepancy (MMD) with RBF and Linear kernels, and Wasserstein Distance. Crucially, we also performed a detailed geodesic analysis, calculating both conditional and marginal geodesic deviations to quantify how closely the forward diffusion paths adhered to theoretically optimal W_2 geodesics.

4 Results and Analysis

Linear path emerges as the best noise schedule. Even though at initial epochs the geodesic schedule promotes geodesic noising behavior, upon sufficient training the linear path overcomes this difference.

5 Discussion

5.1 Linear Schedule as Approximate Geodesic

The widespread use of linear β_t schedules likely stems from their implicit resemblance to the geodesic interpolation of variances. Our empirical results confirm this intuition, with the Linear schedule achieving the best overall performance. However, this resemblance can break down for extreme ε values or high-dimensional data, motivating more principled schedules.

5.2 Geodesic Properties vs. Empirical Performance

Our study suggests that while geometric principles are valuable for guiding schedule design, empirical considerations remain crucial for practical applications.

5.3 Implications for Training

Aligning the forward process with a W_2 geodesic minimizes path energy, which could potentially ease the reverse modeling task. However, our results indicate that:

- **Theoretical optimality** does not guarantee empirical dominance.
- **Stability considerations** play a crucial role in practical performance.
- **Computational efficiency** remains important for large-scale applications.

5.4 Future Research Directions

More rigorous work is needed to find out which noise schedules are better and if there is sufficient theoretical basis for their optimality. The idea we tried in this project could offer a lens into understanding the performance of various noise schedules. Extensions include:

- Higher-Dimension Analysis: Seeing if our findings generalize to higher dimensional continuous diffusion processes
- Whether our findings could prompt a **learnt noise schedule** that minimizes the distance in probability space. This could prompt a new line of inquiry into diffusion model research.

6 Conclusion

We demonstrate that Wasserstein-2 geodesic paths between the initial and final marginal distributions offer a principled theoretical lens for designing noise schedules in diffusion models. While these geometric principles provide valuable insights into the underlying process, our comprehensive evaluation reveals that practically simpler noise paths, such as the Linear schedule, often prove easier for the reverse diffusion model to learn from, leading to superior empirical performance. Nevertheless, this work provides a complete implementation framework with parallel evaluation capabilities and comprehensive visualization tools, suggesting that the integration of geometric principles into diffusion model optimization remains a promising avenue for future research. Our findings underscore that while theoretical elegance guides schedule design, empirical considerations are crucial for practical success.

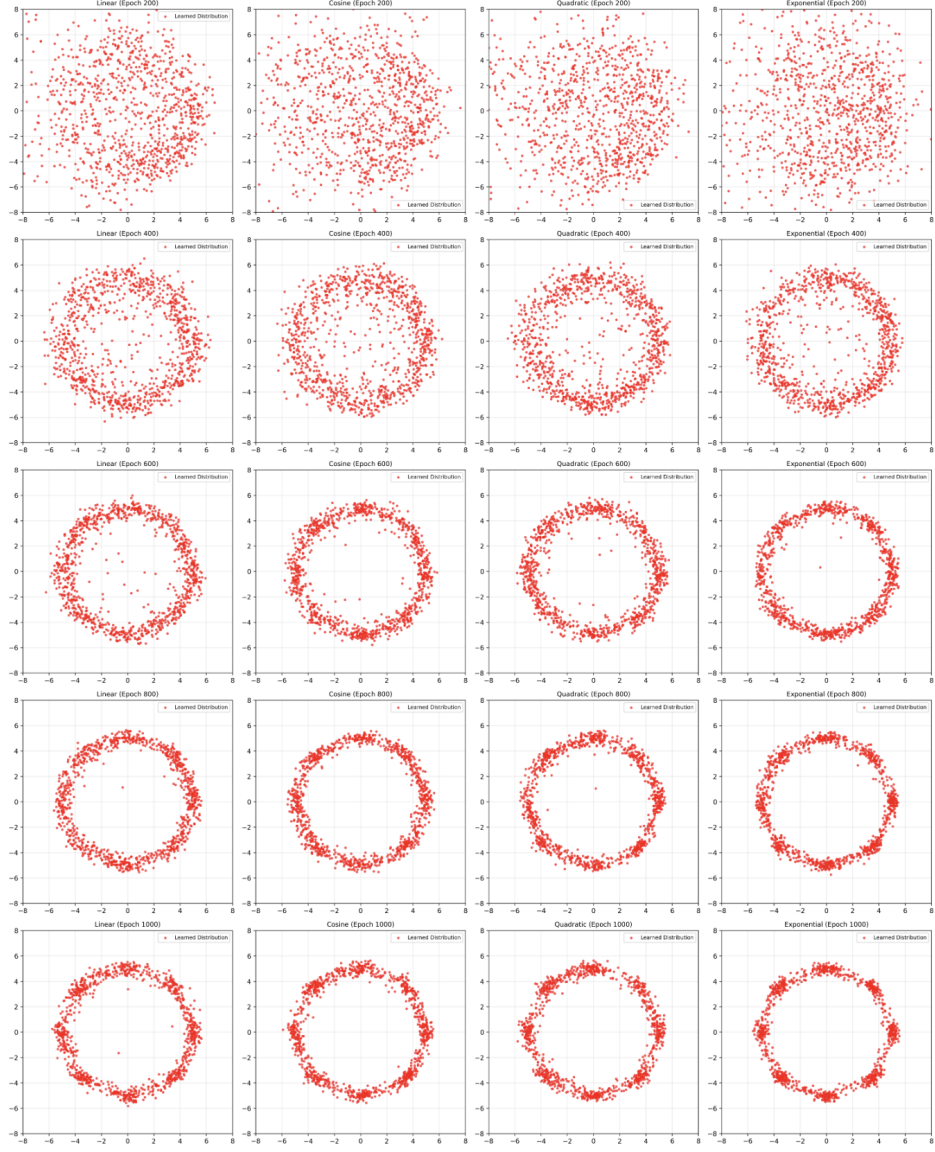


Figure 3: Simulation of the learned distribution at various epochs within the training process for the different probability paths.

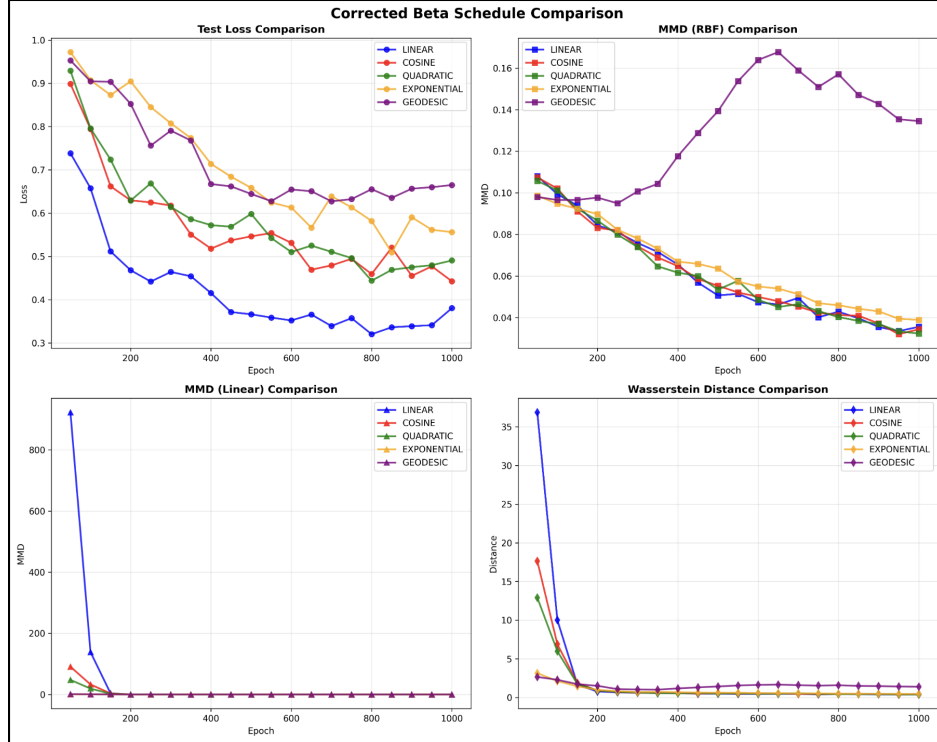


Figure 4: Comparison of the noise schedules. Linear emerges as the winner, even though at initial epochs the geodesic path has lower distance in probability space. Distance is computed by generating samples from the learned model at that epoch, and comparing how close the learned data distribution is to the true data distribution according to the chosen metric (MMD or Wasserstein Distance).