On the Collapse Errors Induced by the Deterministic Sampler for Diffusion Models

HKU Musketeers Foundation
Institute of Data Science
香港大學同心基金數據科學研究院

Yi Zhang
The HKU Musketeers Foundation Institute of Data Science
The University of Hong Kong
yizhang101@connect.hku.hk

Difan Zou
Department of Computer Science
The University of Hong Kong
dzou@cs.hku.hk

NEURAL INFORMATION PROCESSING SYSTEMS

Introduction

We identify and investigate a critical issue in ODE-based sampling for diffusion models, referred to as the "collapse error", where samples tend to collapse locally.

- This phenomenon is observed across various datasets,
- Our analysis shows that this error occurs even in the early sampling process,
- To better understand the errors, we explore several factors that influence them.
- Furthermore, We apply a set of techniques to mitigate the collapse error.

We hope this paper will draw attention to the "collapse error" phenomenon and encourage further research to better understand and address this issue in diffusion models.

Background

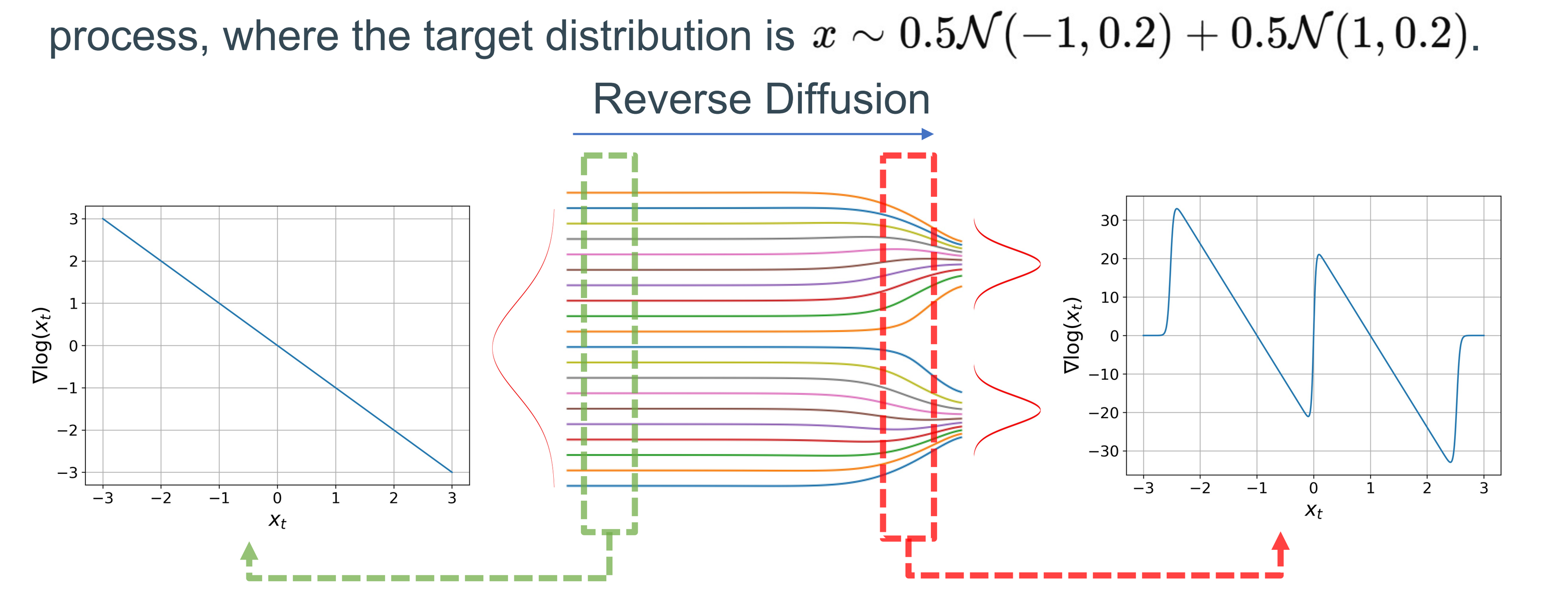
- A standard variance-preserving (VP) forward diffusion process follows:

$$\mathrm{d}\mathbf{x}_t = -\frac{1}{2}\beta(t)\mathbf{x}_t\,\mathrm{d}t + \sqrt{\beta(t)}\,\mathrm{d}\mathbf{w}$$

- The reverse diffusion process is to solve an ODE-based deterministic sampler:

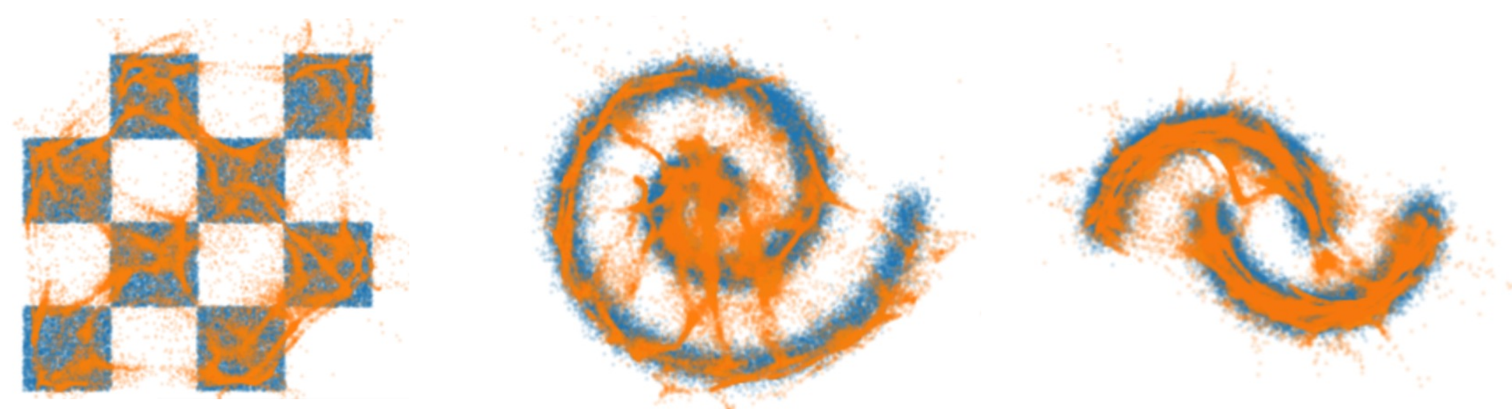
$$\mathrm{d}\mathbf{x}_t = \left[\frac{1}{2}\beta(1-t)\mathbf{x}_t + \frac{1}{2}\beta(1-t)\nabla_{\mathbf{x}_t}\log p_{1-t}(\mathbf{x}_t)\right]\mathrm{d}t$$

- The complexity of the score function varies across different t. To illustrate this, we visualize the score functions at the start and end of the reverse diffusion process, where the target distribution is $x \sim 0.5\mathcal{N}(-1, 0.2) + 0.5\mathcal{N}(1, 0.2)$.



$\nabla_{\mathbf{x}} \log p_t(\mathbf{x})$ at the start of reverse diffusion $\nabla_{\mathbf{x}} \log p_t(\mathbf{x})$ at the end of reverse diffusion

ODE Collapse Errors across Various Datasets



Chessboard Spiral Semi-circle

Ground Truth Samples
Samples by Deterministic Reverse Diffusion

We visualize the discrepancies between the ODE-generated data and the ground truth in the three synthetic datasets. We observe that ODE-generated samples are concentrated into fewer regions than expected.

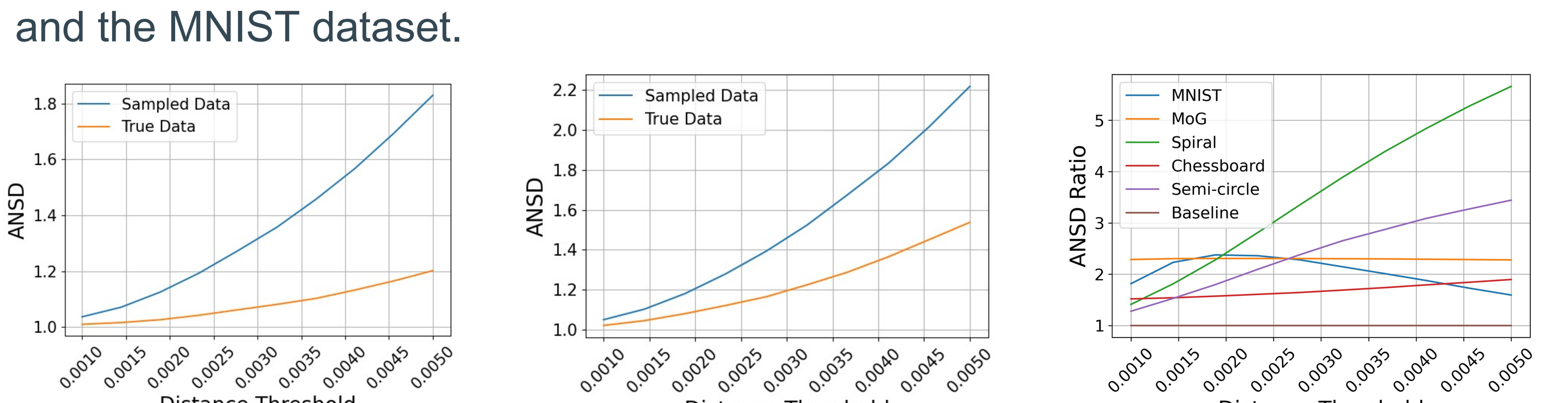
A Quantitative Criteria to Evaluate Collapse

We proposed average Number of Samples within a certain Distance (**ANSD**), i.e.,

$$\mathrm{ANSD}(\mathbf{X}, \epsilon) = \frac{1}{N}\sum_{i=1}^{N}\left(\sum_{j=1, j\neq i}^{N}\mathbb{I}(\|\mathbf{x}_i - \mathbf{x}_j\| \leq \epsilon)\right),$$

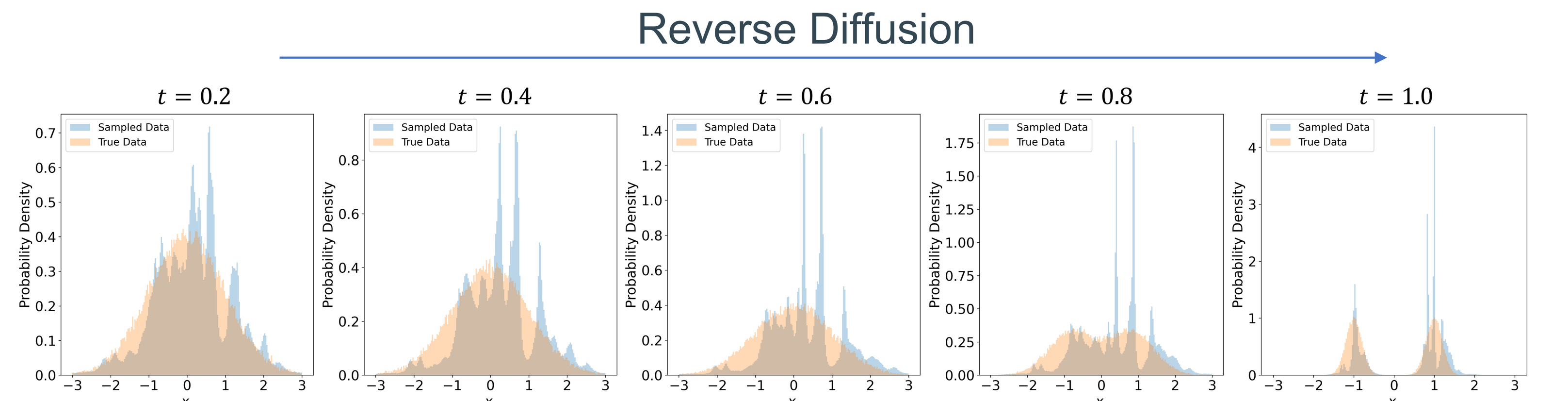
where X represents the sample set, $\|x_i - x_j\|$ is the Euclidean distance between samples x_i and x_j, N is the size of the neighborhood set, and $I(\cdot)$ is an indicator function that equals 1 if the condition inside is true and 0 otherwise.

A higher **ANSD** value for the generated data compared to the dataset indicates a more severe ODE collapse error, so we calculate the **ANSD Ratio** (ANSD of the generated data divided by the ANSD of the true dataset) to normalize this comparison. Below, We show the **ANSD** of chessboard and spiral-shape distribution, and **ANSD Ratio** of the above three synthetic datasets, a 1D MoG, and the MNIST dataset.



Chessboard Spiral ANSD Ratio across Datasets

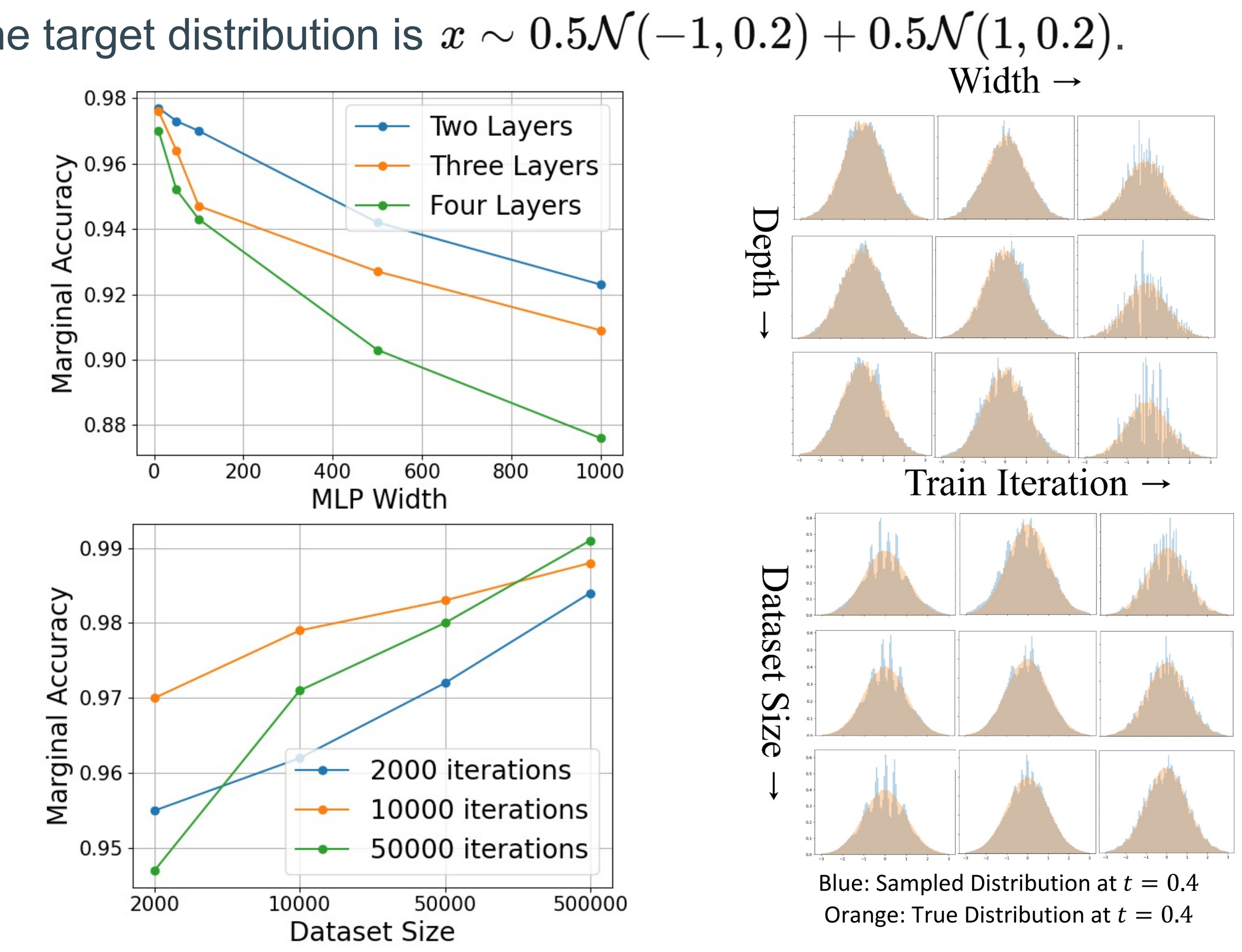
Collapse Errors Accumulate during Sampling



The collapse error is accumulated throughout the ODE sampling process, even from the early stages, and becomes more severe as sampling progresses. The reverse diffusion process illustrated here uses a 1D MoG as the target distribution, where $x \sim 0.5\mathcal{N}(-1, 0.2) + 0.5\mathcal{N}(1, 0.2)$.

Factors Influencing Collapse Errors

We evaluate the Marginal Accuracy of intermediate distributions generated by ODE-based sampling at $t = 0.4$ where the model is a MLP with tanh activation and the target distribution is $x \sim 0.5\mathcal{N}(-1, 0.2) + 0.5\mathcal{N}(1, 0.2)$.

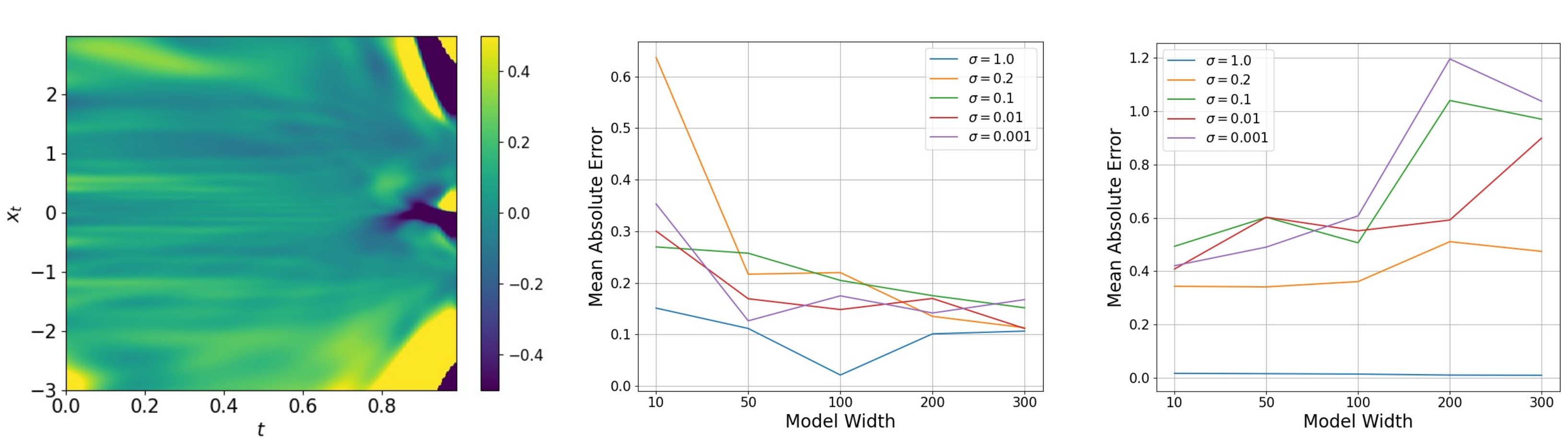


We find that the collapse error is influenced by the following factors:

- **Model Size**: Wider and deeper models are more prone to collapse errors.
- **Dataset Size**: Larger datasets generally help mitigate collapse errors.
- **Training Duration**: Longer training reduces collapse errors with sufficient data, but for small datasets, it may worsen the issue.

Analysis of Collapse Errors

We measure the following criteria in the setting of a two-layer MLP with tanh activation and a 1D MoG $x \sim 0.5\mathcal{N}(-1, \sigma) + 0.5\mathcal{N}(1, \sigma)$ as the target distributions.



Velocity Error Map MAE of Score Function at $t = 0.8$ MAE of Score Function at $t = 0.2$

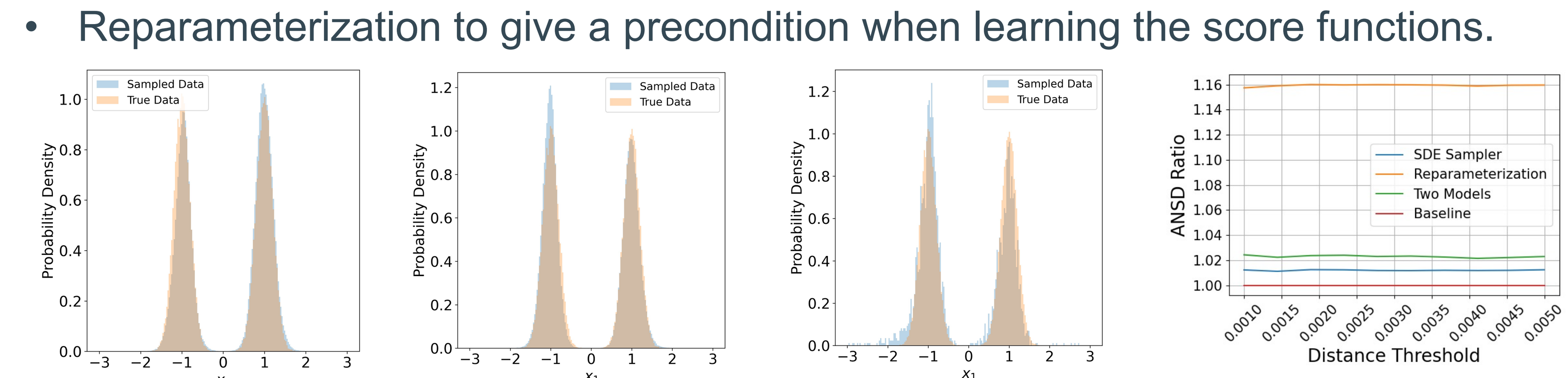
We find that the collapse error may be attributed to the following reasons:

- Velocity error propagates along t.
- Challenges in fitting score functions of different ts.

Techniques to Mitigate the Collapse Errors

With insights on the potential reasons of collapse error, we propose three techniques to mitigate it. We apply them on a 1D MoG as the target distribution.

- Using SDE-based sampler
- Separating the learning of score functions to two models
- Reparameterization to give a precondition when learning the score functions.



Using SDE Sampler Using Two Models Using Reparameterization ANSD Ratio across the three Methods

References

Ho, J., Jain, A., & Abbeel, P. (2020). Denoising diffusion probabilistic models. Advances in neural information processing systems, 33, 6840-6851.

Karras, T., Aittala, M., Aila, T., and Laine, S. (2022). Elucidating the design space of diffusion-based generative models. Advances in neural information processing systems, 35:26565–26577.

Song, Y., Sohl-Dickstein, J., Kingma, D. P., Kumar, A., Ermon, S., and Poole, B. (2020). Score-based generative modeling through stochastic differential equations. In International Conference on Learning Representations.