Soft Mixture Denoising: Beyond the Expressive Bottleneck of Diffusion Models

Yangming Li, Boris van Breugel, Mihaela van der Schaar

Department of Applied Mathematics and Theoretical Physics

University of Cambridge

Outline

- Theory: Unbounded Approximation Errors
- Method: Soft Mixture Denoising
- Empirical Experiments on Image Generation

Theory

Proposition 3.1 (Non-Gaussian Inverse Probability). *For the diffusion process defined in Eq. (1), suppose that the real data follow a Gaussian mixture:* $q(\mathbf{x}_0) = \sum_{k=1}^{K} w_k \mathcal{N}(\mathbf{x}_0; \boldsymbol{\mu}_k, \boldsymbol{\Sigma}_k)$, *which consists of* K *Gaussian components with mixture weight* w_k, *mean vector* $\boldsymbol{\mu}_k$, *and covariance matrix* $\boldsymbol{\Sigma}_k$, *then the posterior forward probability* $q(\mathbf{x}_{t-1} \mid \mathbf{x}_t)$ *at every iteration* $t \in [1, T]$ *is another mixture of Gaussian distributions:*

$$q(\mathbf{x}_{t-1} \mid \mathbf{x}_t) = \sum_{k=1}^{K} w'_k \mathcal{N}(\mathbf{x}_{t-1}; \boldsymbol{\mu}'_k, \boldsymbol{\Sigma}'_k), \tag{6}$$

where $w'_k, \boldsymbol{\mu}'_k$ *depend on both variable* $\mathbf{x}_t$ *and* $\boldsymbol{\mu}_t$.

Remark 3.1. The Gaussian mixture in theory is a universal approximator of smooth probability densities (Dalal & Hall, 1983; Goodfellow et al., 2016). Therefore, this proposition implies that the posterior forward probability $q(\mathbf{x}_{t-1} \mid \mathbf{x}_t)$ can be arbitrarily complex.

Insights: the Gaussian denoisier is not expressive enough and the previous assumption of bounded errors is too strong

Theorem 3.1 (Uniformly Unbounded Denoising Error). *For the diffusion process defined in Eq. (1) and the Gaussian denoising process defined in Eq. (2), there exists a continuous data distribution* $q(\mathbf{x}_0)$ *(more specifically, Gaussian mixture) such that* $\mathcal{M}_t$ *is uniformly unbounded—given any real number* $N \in \mathbb{R}$, *the inequality* $\mathcal{M}_t > N$ *holds for every denoising iteration* $t \in [1, T]$.

Theorem 3.2 (Unbounded Approximation Error). *For the forward and backward processes respectively defined in Eq. (1) and Eq. (2), given any real number* $N \in \mathbb{R}$, *there exists a continuous data distribution* $q(\mathbf{x}_0)$ *(specifically, Gaussian mixture) such that* $\mathcal{E} > N$.

Insight: as expected, the local and global denoising errors are unluckily unbounded

Method

- A continuously relaxed Gaussian mixture (instead of simple Gaussian) for backward denoising

$$p_\theta(\mathbf{x}_{t-1} \mid \mathbf{x}_t) = \mathcal{N}(\mathbf{x}_{t-1}; \boldsymbol{\mu}_\theta(\mathbf{x}_t, t), \sigma_t \mathbf{I}), \longrightarrow p_\theta^{\mathrm{SMD}}(\cdot) = \int p_\theta^{\mathrm{SMD}}(\mathbf{x}_{t-1}, \mathbf{z}_t \mid \mathbf{x}_t) d\mathbf{z}_t = \int p_\theta^{\mathrm{SMD}}(\mathbf{z}_t \mid \mathbf{x}_t) p_\theta^{\mathrm{SMD}}(\mathbf{x}_{t-1} \mid \mathbf{x}_t, \mathbf{z}_t) d\mathbf{z}_t,$$

- Theoretical guarantee

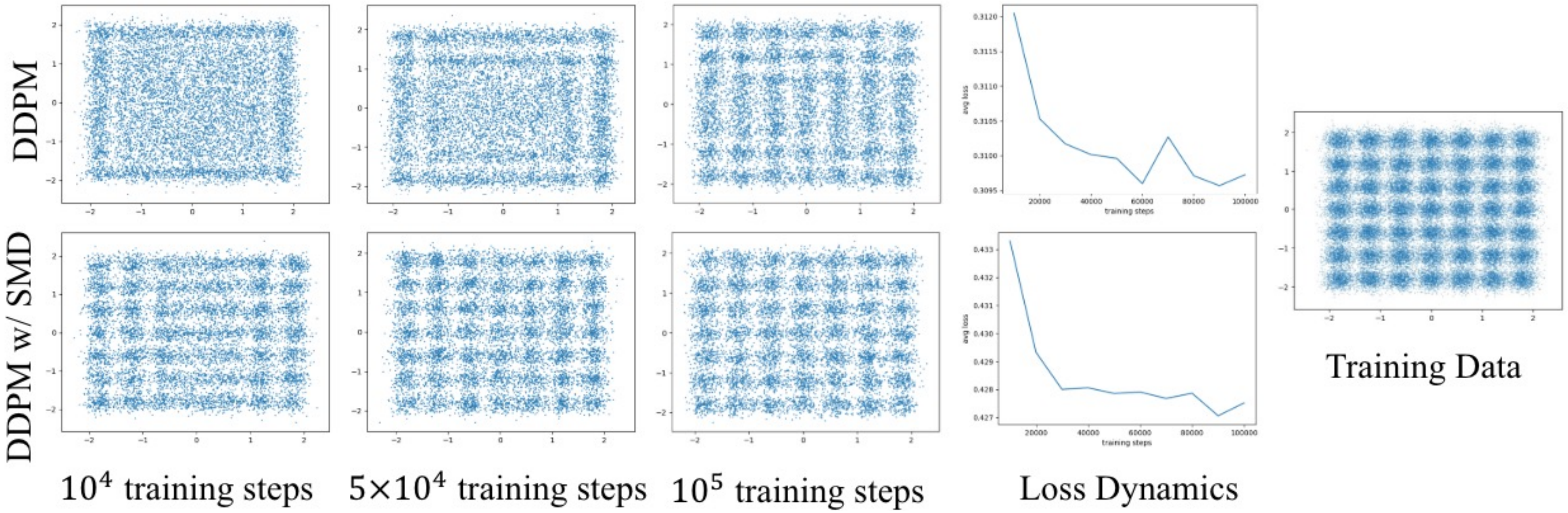
Theorem 4.1 (Expressive Soft Mixture Denoising). *For the diffusion process defined in Eq. (1), suppose soft mixture model $p_\theta^{\mathrm{SMD}}(\mathbf{x}_{t-1} \mid \mathbf{x}_t)$ is applied for backward denoising and data distribution $q(\mathbf{x}_0)$ is a Gaussian mixture, then both $\mathcal{M}_t = 0, \forall t \in [1, T]$ and $\mathcal{E} = 0$ hold.*

- Loss function for optimization

$$\mathcal{L}^{\mathrm{SMD}} = C + \sum_{t=1}^{T} \mathbb{E}_{\boldsymbol{\eta}, \boldsymbol{\epsilon}, \mathbf{x}_0} \Big[\Gamma_t \big\| \boldsymbol{\epsilon} - \boldsymbol{\epsilon}_{\theta \bigcup f_\phi(g_\xi(\cdot), t)} \big(\sqrt{\bar{\alpha}_t} \mathbf{x}_0 + \sqrt{1 - \bar{\alpha}_t} \boldsymbol{\epsilon}, t \big) \big\|^2 \Big],$$

Experiments – Part1

- Synthetic data



Experiments – Part2

- Few backward iterations

(a) Baseline: vanilla LDM; FID: 11.29.

(b) Our model: LDM w/ SMD; FID: 6.85.

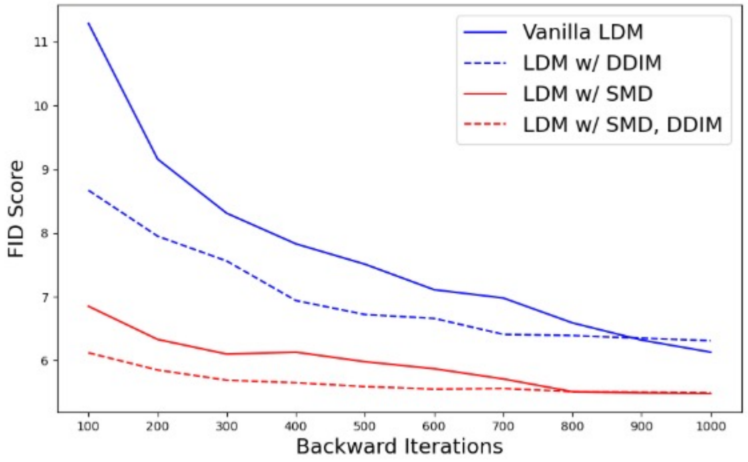


Figure 3: **SMD reduces the number of sampling steps.** Latent DDIM and DDPM for different iterations on CelebA-HQ (256×256).