Optimal Stopping in Latent Diffusion Models

Yu-Han Wu[1, 2], Quentin Berthet[2], Gérard Biau[1], Claire Boyer[3], Romuald Elie[2], Pierre Marion[4]

07/12/2025

[1]LPSM, Sorbonne Université

[2]Google DeepMind

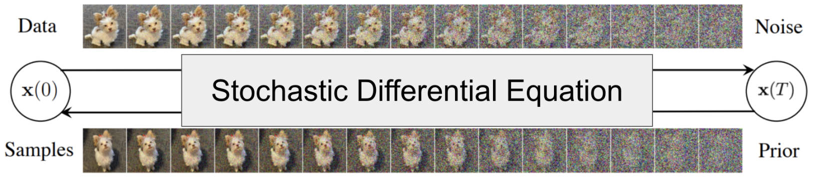
[3]LMO, Université Paris-Saclay

[4]Inria

Introduction

Diffusion Model

Goal: Learn the distribution p_{true} of images.



- Forward SDE:

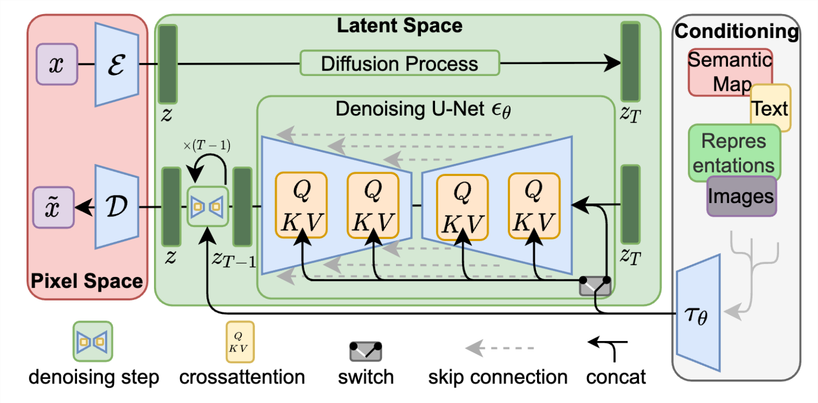
$$d\overrightarrow{X}_t = -\overrightarrow{X}_t dt + \sqrt{2}\, d\overrightarrow{B}_t, \quad \overrightarrow{X}_0 \sim p_{\text{true}},$$

- Backward SDE:

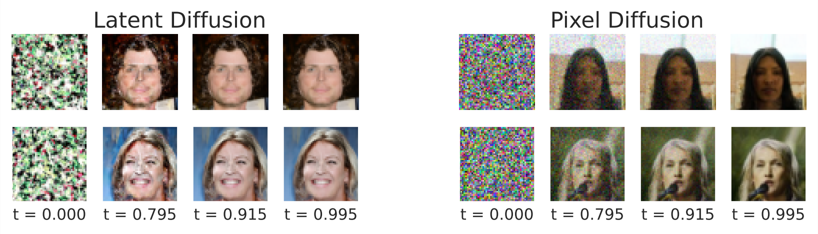
$$d\overleftarrow{X}_t = (\overleftarrow{X}_t + 2\underbrace{\nabla \log p_{T-t}(\overleftarrow{X}_t)}_{\substack{\text{score function}\\ \text{unknown}}})dt + \sqrt{2}\, d\overleftarrow{B}_t, \quad \overleftarrow{X}_0 \sim p_T,$$

where $\overrightarrow{X}_t \sim p_t$ and $\overleftarrow{X}_{T-t} \stackrel{\mathcal{D}}{=} \overrightarrow{X}_t$.

Latent Diffusion Model



Qualitative Results



FID on last few steps

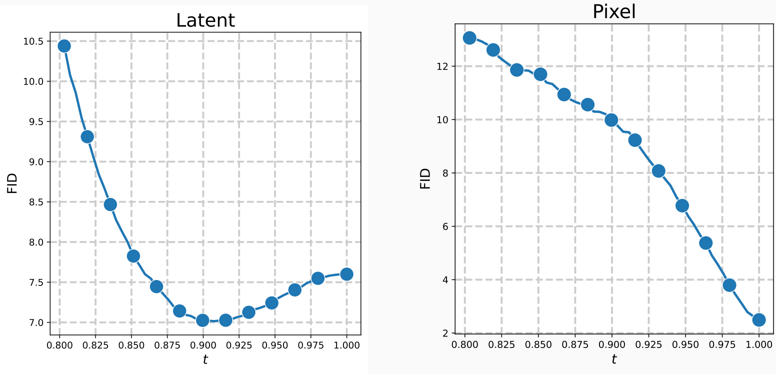


Figure 1: CelebA

FID on last few steps

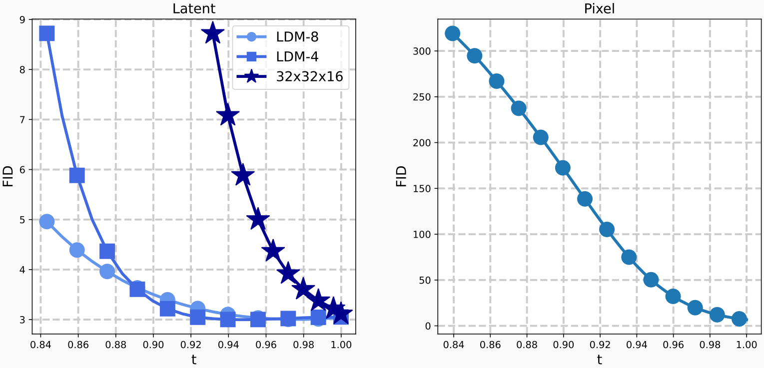


Figure 2: ImageNet

Same Observation in Prior Work

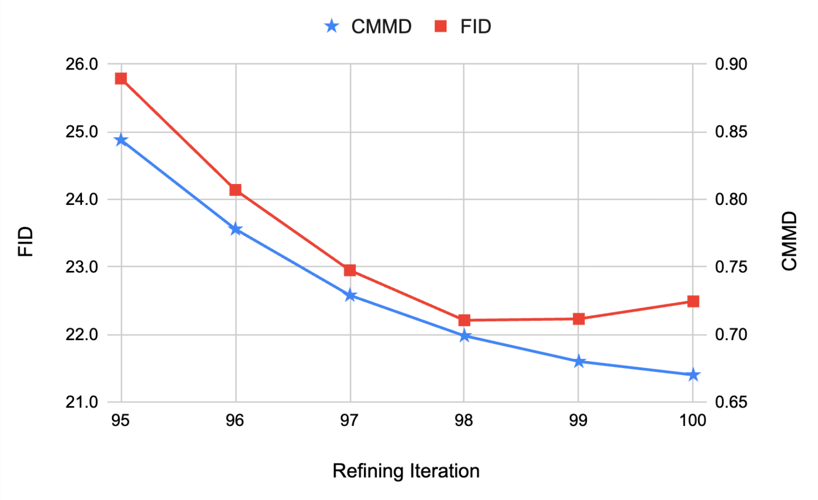


Figure 3: Jayasumana et al. (2024)

Problem Setup and Results

Problem Setup

- **Gaussian Data.** Data distribution is centered D-dimensional Guassian with independent components, i.e. $p_{\text{data}} = \mathcal{N}(0, \text{diag}(\sigma_1^2, \ldots, \sigma_D^2))$, where $\sigma_1 > \ldots > \sigma_D$.
- **Projected diffusion process.** We consider orthogonal projection matrices P_d (projecting on the first d coordinates) that map the diffusion process to a lower dimension

$$dP_d \overleftarrow{X_t} = (w_{T-t}^2 P_d \overleftarrow{X_t} + 2w_{T-t}^2 s_{P_d}(P_d \overleftarrow{X_t}, T-t))dt + \sqrt{2w_{T-t}^2} dP_d \overleftarrow{W_t}.$$

- **Training.** In this simpler setup with Gaussian distribution, learning the score boils down to covariance matrix estimation, and we assume that the estimation is equal to $\hat{\Sigma} = \text{diag}(\hat{\sigma}_1^2, \ldots, \hat{\sigma}_D^2)$.

Results

Non-Monotonicity

For $d \in \{1, \ldots, D\}$, the Fréchet distance $d_F(P_d^\top P_d \overleftarrow{X}_t, \overrightarrow{X}_0)$ is non-increasing with respect to t. On the other hand, $d_F(P_d^\top P_d \overleftarrow{\hat{X}}_t, \overrightarrow{X}_0)$ is non-increasing if and only if

$$\sum_{d'=1}^{d} (1 - \frac{\sigma_{d'}}{\hat{\sigma}_{d'}})(1 - \hat{\sigma}_{d'}^2) \geq 0.$$

Results

Optimal projection and optimal stopping time

Assume that $\Sigma = \text{diag}(\sigma^2, \dots, \sigma^2, 0, \dots, 0)$ with the last $D - d_0$ entries equal to 0. Let $\varepsilon \in (0, 1)$. Then, there exists $\hat{\delta}_{d_0} \in [0, T]$ such that with probability $1 - 2d_0 e^{-\frac{n}{8}}$,

$$d_F(P_{d_0}^\top P_{d_0} \overleftarrow{\hat{X}}_{T-\hat{\delta}_{d_0}}, \overrightarrow{X_0}) = \min_{\substack{t \in [0,T] \\ d' \in \{1,\dots,D\}}} d_F(P_{d'}^\top P_{d'} \overleftarrow{\hat{X}}_t, \overrightarrow{X_0}).$$

In other words, if the data lies in a linear subspace, there is an optimal stopping time and optimal projection that minimizes the data distribution and generated distribution.

Results (General)

General Gaussian data and estimation. We now consider $p_{\text{data}} = \mathcal{N}(0, \Sigma)$ to be a general centered Gaussian distribution and $\hat{\Sigma}$ an estimation of Σ.

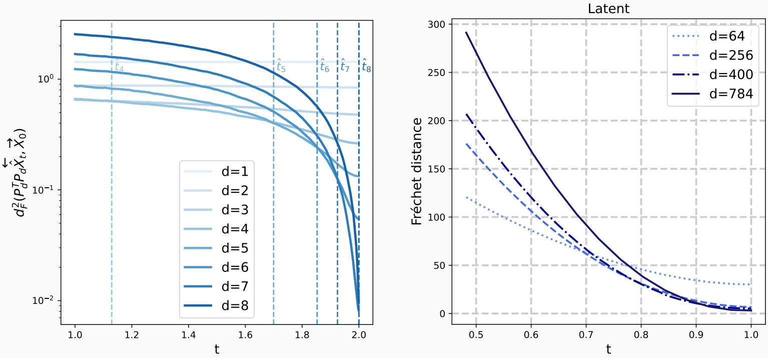
Optimal projection given t

For $d \in \{1, \ldots, D\}$ and any $t \in [\hat{T}_d(u), \hat{t}_{d+1}(u)]$, with probability $1 - 2e^{-u}$,

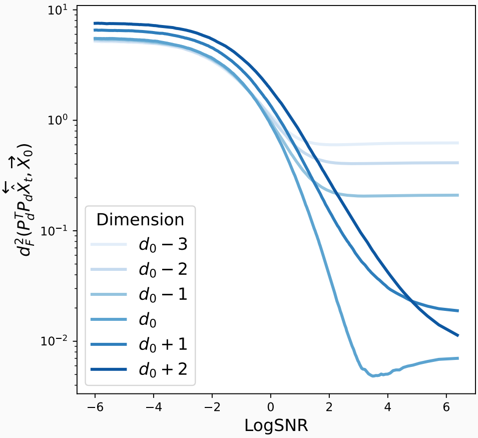
$$d \in \operatorname*{arg\,min}_{d' \in \{1,\ldots,D\}} d_F(\hat{O} P_{d'}^\top P_{d'} \hat{O}^\top \overleftarrow{\hat{X}}_t, \overrightarrow{X_0}).$$

Numerical Verification

Experiment: Optimal Projection at a given time



Experiment: Optimal stopping time



Conclusion

Conclusion

Summary:

- The last few steps of latent diffusion might be degrading the image quality.
- Our theoretical results indicate that there exists optimal stopping time and optimal projection for latent diffusions.

Thank you

Appendix

More metrics