MaRS: A Fast Sampler for Mean Reverting Diffusion Based on ODE and SDE Solvers

Ao Li*, Wei Fang*, Hongbo Zhao*, Le Lu, Ge Yang†, Minfeng Xu†

University of Chinese Academy of Science

Computational Biology and Machine Intelligence, CASIA

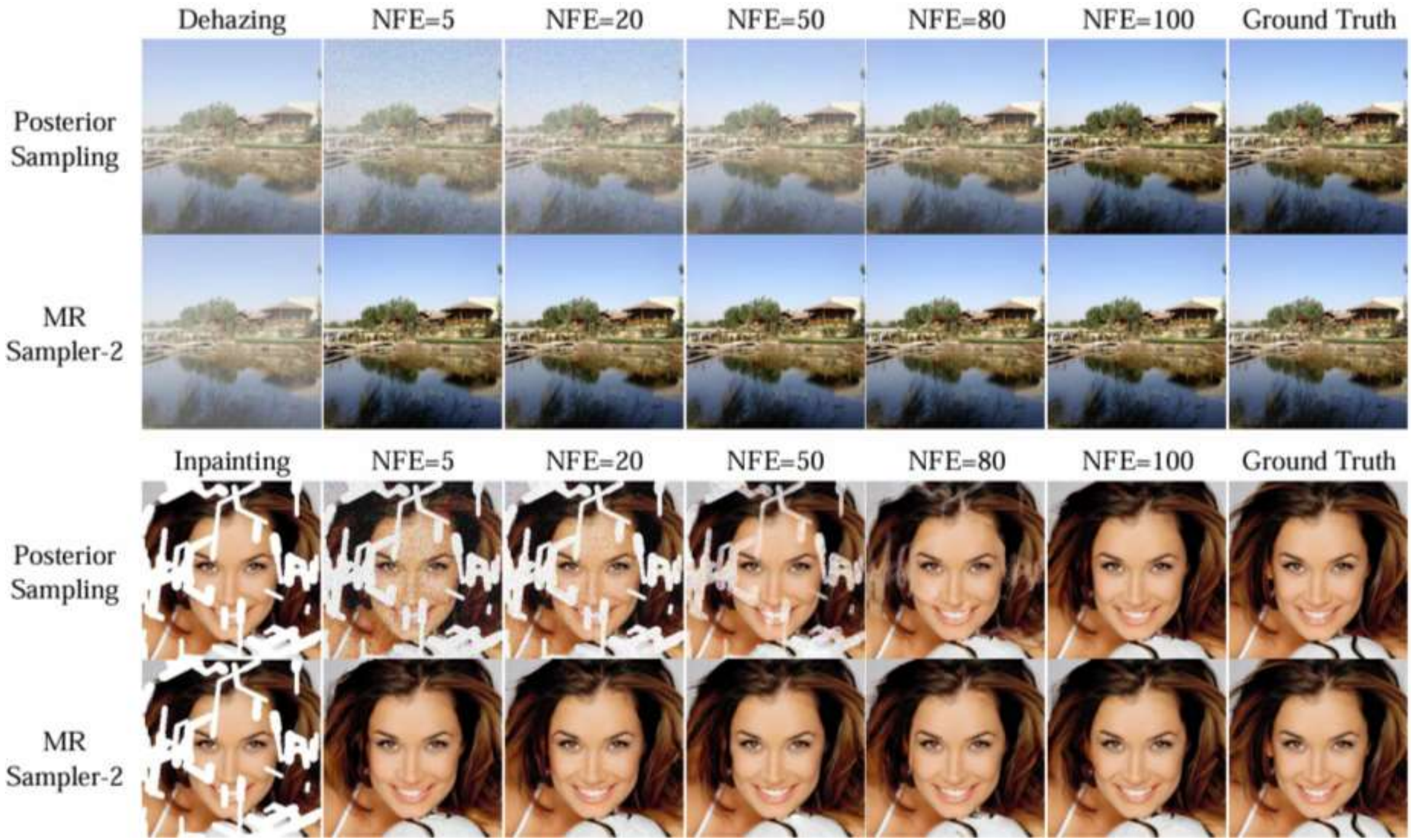
DAMO Academy, Alibaba

HuPan Lab

*These authors contributed equally to this work.
†Corresponding authors.

MaRS: A Fast Sampler for Mean Reverting Diffusion Based on ODE and SDE Solvers

Results Display: MR Sampler achieves stable performance with speedup factors ranging from 10 to 20.



MaRS: A Fast Sampler for Mean Reverting Diffusion Based on ODE and SDE Solvers

Background

- SDE of Diffusion Probabilistic Models

$$\mathrm{d}\boldsymbol{x} = f(t)\boldsymbol{x}\mathrm{d}t + g(t)\mathrm{d}\boldsymbol{w}$$

- Two commonly used DPMs

SDE	$f(t)$	$g(t0)$	$p(\boldsymbol{x}_t \mid \boldsymbol{x}_0)$
VP-SDE	$-\frac{1}{2}\beta(t)$	$\sqrt{\beta(t)}$	$\mathcal{N}(\boldsymbol{x}_t; \boldsymbol{x}_0 e^{-\frac{1}{2}\int_0^t \beta(\tau)\mathrm{d}\tau}, \boldsymbol{I} - \boldsymbol{I}e^{-\int_0^t \beta(\tau)\mathrm{d}\tau})$
VE-SDE	0	$\sqrt{\frac{\mathrm{d}[\sigma^2(t)]}{\mathrm{d}t}}$	$\mathcal{N}(\boldsymbol{x}_t; \boldsymbol{x}_0, [\sigma^2(t) - \sigma^2(0)]\,\boldsymbol{I})$

- Solving the probability flow ODE yields a fast sampler

$$\frac{\mathrm{d}\boldsymbol{x}}{\mathrm{d}t} = f(t)\boldsymbol{x} - \frac{1}{2}g^2(t)\nabla_{\boldsymbol{x}} \log p_t(\boldsymbol{x})$$

  - PNDM[1]

$$x_{t-\delta} = \frac{\sqrt{\bar{\alpha}_{t-\delta}}}{\sqrt{\bar{\alpha}_t}}x_t - \frac{(\bar{\alpha}_{t-\delta} - \bar{\alpha}_t)}{\sqrt{\bar{\alpha}_t}\left(\sqrt{(1-\bar{\alpha}_{t-\delta})\bar{\alpha}_t} + \sqrt{(1-\bar{\alpha}_t)\bar{\alpha}_{t-\delta}}\right)}\epsilon_t$$

  - DPM-Solver(++)[2]

$$\tilde{\boldsymbol{x}}_{t_i} = \frac{\sigma_{t_i}}{\sigma_{t_{i-1}}}\tilde{\boldsymbol{x}}_{t_{i-1}} + \sigma_{t_i}\sum_{n=0}^{k-1}\boldsymbol{x}_\theta^{(n)}(\hat{\boldsymbol{x}}_{\lambda_{t_{i-1}}}, \lambda_{t_{i-1}})\int_{\lambda_{t_{i-1}}}^{\lambda_{t_i}} e^{\lambda}\frac{(\lambda - \lambda_{t_{i-1}})^n}{n!}\mathrm{d}\lambda + \mathcal{O}(h_i^{k+1})$$

  - UniPC[3]

$$\tilde{\boldsymbol{x}}_{t_i}^c = \frac{\alpha_{t_i}}{\alpha_{t_{i-1}}}\tilde{\boldsymbol{x}}_{t_{i-1}}^c - \sigma_{t_i}(e^{h_i} - 1)\boldsymbol{\epsilon}_\theta(\tilde{\boldsymbol{x}}_{t_{i-1}}, t_{i-1}) - \sigma_{t_i}B(h_i)\sum_{m=1}^{p}\frac{a_m}{r_m}D_m$$

[1]. Luping Liu, Yi Ren, Zhijie Lin, and Zhou Zhao. Pseudo numerical methods for diffusion models on manifolds. *arXiv preprint arXiv*:2202.09778, 2022a.
[2]. Cheng Lu, Yuhao Zhou, Fan Bao, Jianfei Chen, Chongxuan Li, and Jun Zhu. Dpm-solver: A fast ode solver for diffusion probabilistic model sampling in around 10 steps. *Advances in Neural Information Processing Systems*, 35:5775–5787, 2022a.
[3]. Wenliang Zhao, Lujia Bai, Yongming Rao, Jie Zhou, and Jiwen Lu. Unipc: A unified predictor-corrector framework for fast sampling of diffusion models. *Advances in Neural Information Processing Systems*, 36, 2024.

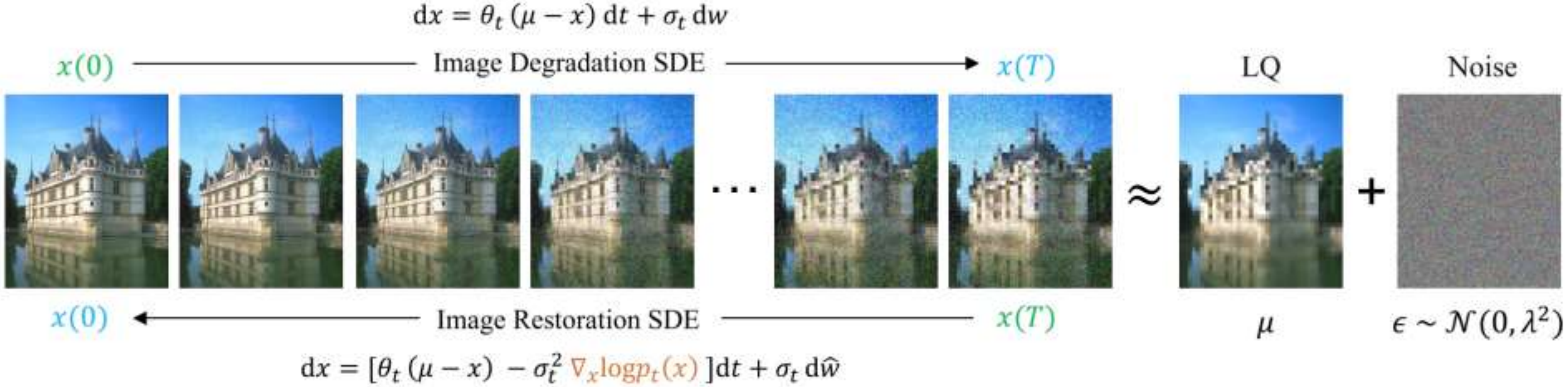
MaRS: A Fast Sampler for Mean Reverting Diffusion Based on ODE and SDE Solvers

Research Gap

- Mean-Reverting SDE

$$\mathrm{d}\boldsymbol{x} = f(t)(\boldsymbol{\mu} - \boldsymbol{x})\mathrm{d}t + g(t)\mathrm{d}\boldsymbol{w}$$

- Image restoration with mean-reverting diffusion*



- Samplers for DPM are not applicable
  - PF-ODE

$$\begin{cases} \frac{\mathrm{d}\boldsymbol{x}}{\mathrm{d}t} = f(t)\boldsymbol{x} - \frac{1}{2}g^2(t)\nabla_{\boldsymbol{x}} \log p_t(\boldsymbol{x}) \\ \frac{\mathrm{d}\boldsymbol{x}}{\mathrm{d}t} = f(t)(\boldsymbol{\mu} - \boldsymbol{x}) - \frac{1}{2}g^2(t)\nabla_{\boldsymbol{x}} \log p_t(\boldsymbol{x}) \end{cases}$$

  - Reverse-time SDE

$$\begin{cases} \mathrm{d}\boldsymbol{x} = \left[f(t)\boldsymbol{x} - g^2(t)\nabla_{\boldsymbol{x}} \log p_t(\boldsymbol{x})\right]\mathrm{d}t + g(t)\mathrm{d}\bar{\boldsymbol{w}} \\ \mathrm{d}\boldsymbol{x} = \left[f(t)(\boldsymbol{\mu} - \boldsymbol{x}) - g^2(t)\nabla_{\boldsymbol{x}} \log p_t(\boldsymbol{x})\right]\mathrm{d}t + g(t)\mathrm{d}\bar{\boldsymbol{w}} \end{cases}$$

*Ziwei Luo, Fredrik K Gustafsson, Zheng Zhao, Jens Sjölund, and Thomas B Schön. Image restoration with mean-reverting stochastic differential equations. *arXiv preprint arXiv*:2301.11699, 2023b.

Fast Samplers with Noise Prediction

- PF-ODE with noise prediction

$$\frac{\mathrm{d}\boldsymbol{x}}{\mathrm{d}t} = f(t)\left(\boldsymbol{\mu} - \boldsymbol{x}\right) + \frac{g^2(t)}{2\sigma_t}\boldsymbol{\epsilon}_\theta(\boldsymbol{x}_t, t)$$

- Solution

$$\boldsymbol{x}_t = \frac{\alpha_t}{\alpha_s}\boldsymbol{x}_s + \left(1 - \frac{\alpha_t}{\alpha_s}\right)\boldsymbol{\mu} - \alpha_t \int_{\lambda_s}^{\lambda_t} e^{-\lambda}\boldsymbol{\epsilon}_\theta(\boldsymbol{x}_\lambda, \lambda)\mathrm{d}\lambda$$

- Reverse-time SDE with noise prediction

$$\mathrm{d}\boldsymbol{x} = \left[f(t)\left(\boldsymbol{\mu} - \boldsymbol{x}\right) + \frac{g^2(t)}{\sigma_t}\boldsymbol{\epsilon}_\theta(\boldsymbol{x}_t, t)\right]\mathrm{d}t + g(t)\mathrm{d}\bar{\boldsymbol{w}}$$

- Solution

$$\boldsymbol{x}_t = \frac{\alpha_t}{\alpha_s}\boldsymbol{x}_s + \left(1 - \frac{\alpha_t}{\alpha_s}\right)\boldsymbol{\mu} - 2\alpha_t \int_{\lambda_s}^{\lambda_t} e^{-\lambda}\boldsymbol{\epsilon}_\theta(\boldsymbol{x}_\lambda, \lambda)\mathrm{d}\lambda + \sigma_t\sqrt{(e^{2(\lambda_t - \lambda_s)} - 1)}\boldsymbol{z}$$

- Estimate the integral part with *exponential integrators**

$$\int_{\lambda_s}^{\lambda_t} e^{-\lambda}\boldsymbol{\epsilon}_\theta(x_\lambda, \lambda)\mathrm{d}\lambda = e^{-\lambda_t}\sum_{n=0}^{k-1}\left[\boldsymbol{\epsilon}_\theta^{(n)}(\boldsymbol{x}_{\lambda_s}, \lambda_s)\left(e^h - \sum_{m=0}^{n}\frac{(h)^m}{m!}\right)\right] + \mathcal{O}(h^{k+1})$$

*Marlis Hochbruck and Alexander Ostermann. Exponential integrators. *Acta Numerica*, 19:209–286, 2010.

MaRS: A Fast Sampler for Mean Reverting Diffusion Based on ODE and SDE Solvers

Fast Samplers with Data Prediction

- PF-ODE with data prediction

$$\frac{\mathrm{d}\boldsymbol{x}}{\mathrm{d}t} = \left(\frac{g^2(t)}{2\sigma_t^2} - f(t)\right)\boldsymbol{x} + \left[f(t) - \frac{g^2(t)}{2\sigma_t^2}(1-\alpha_t)\right]\boldsymbol{\mu} - \frac{g^2(t)}{2\sigma_t^2}\alpha_t \boldsymbol{x}_\theta(\boldsymbol{x}_t, t)$$

- Solution

$$\boldsymbol{x}_t = \frac{\sigma_t}{\sigma_s}\boldsymbol{x}_s + \boldsymbol{\mu}\left(1 - \frac{\sigma_t}{\sigma_s} + \frac{\sigma_t}{\sigma_s}\alpha_s - \alpha_t\right) + \sigma_t \int_{\lambda_s}^{\lambda_t} e^{\lambda}\boldsymbol{x}_\theta(\boldsymbol{x}_\lambda, \lambda)\mathrm{d}\lambda$$

- Reverse-time SDE with noise prediction

$$\mathrm{d}\boldsymbol{x} = \left(\frac{g^2(t)}{\sigma_t^2} - f(t)\right)\boldsymbol{x} + \left[f(t) - \frac{g^2(t)}{\sigma_t^2}(1-\alpha_t)\right]\boldsymbol{\mu} - \frac{g^2(t)}{\sigma_t^2}\alpha_t \boldsymbol{x}_\theta(\boldsymbol{x}_t, t) + g(t)\mathrm{d}\bar{\boldsymbol{w}}$$

- Solution

$$\boldsymbol{x}_t = \frac{\sigma_t}{\sigma_s}e^{-(\lambda_t-\lambda_s)}\boldsymbol{x}_s + \boldsymbol{\mu}\left(1 - \frac{\alpha_t}{\alpha_s}e^{-2(\lambda_t-\lambda_s)} - \alpha_t + \alpha_t e^{-2(\lambda_t-\lambda_s)}\right)$$
$$+2\alpha_t \int_{\lambda_s}^{\lambda_t} e^{-2(\lambda_t-\lambda)}\boldsymbol{x}_\theta(\boldsymbol{x}_\lambda, \lambda)\mathrm{d}\lambda + \sigma_t\sqrt{1-e^{-2(\lambda_t-\lambda_s)}}\boldsymbol{z},$$

MaRS: A Fast Sampler for Mean Reverting Diffusion Based on ODE and SDE Solvers

Experiments

- Diffusion sampling process in NCSN[1]

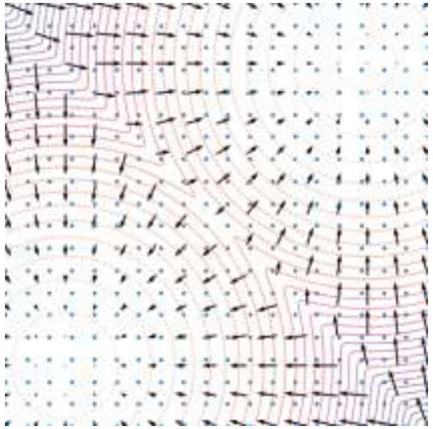
Images---Points in high-dimensional space

- Sampling Trajectory

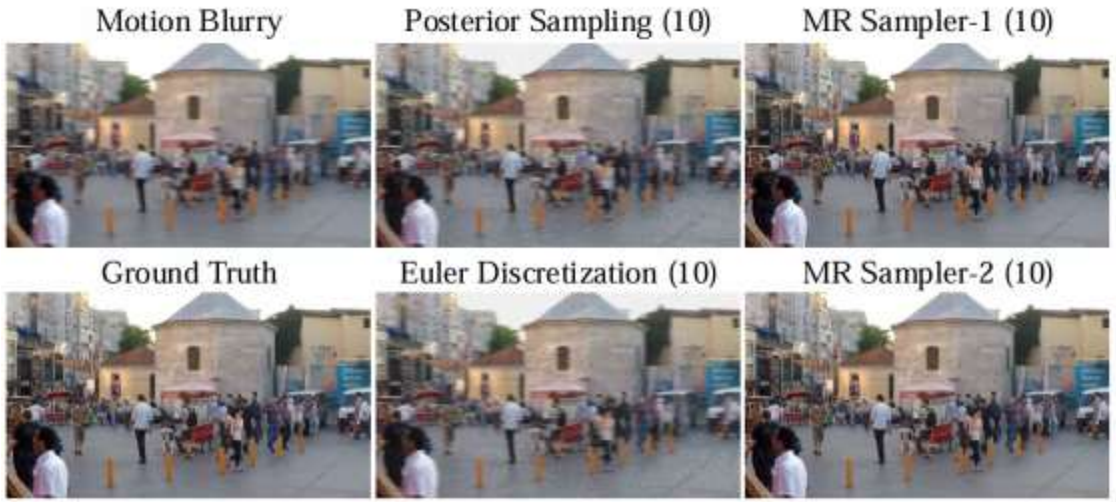
$\boldsymbol{x}_T$ (Gaussian distribution) → ⋯ → $\boldsymbol{x}_t$ → ⋯ → $\boldsymbol{x}_0$ (data distribution)

- Visualization

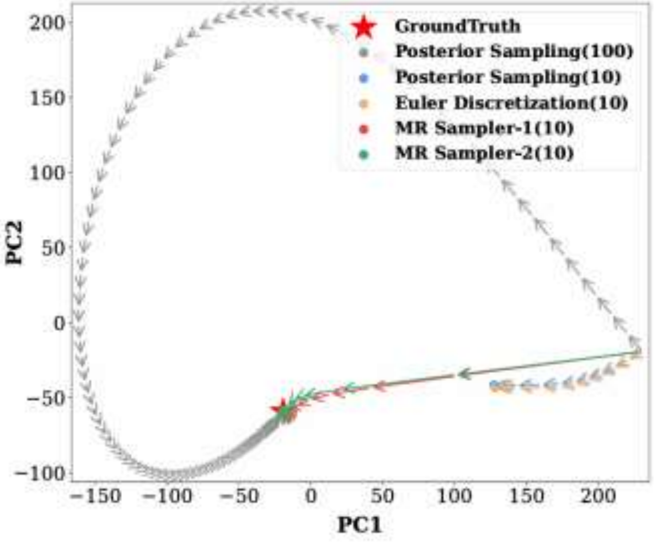
Use Principal Component Analysis (PCA) to obtain the projection of the sampling trajectory in 2D space.



Using Langevin dynamics to sample from a mixture of two Gaussians.[2]



(a) Sampling results.



(b) Trajectory.

[1]. Yang Song and Stefano Ermon. Generative modeling by estimating gradients of the data distribution. *Advances in neural information processing systems*, 32, 2019.
[2]. **Image Source**: Yang Song. Generative Modeling by Estimating Gradients of the Data Distribution. Online. [Accessed: January 9, 2025].

Experiments

- Numerical stability of two parameterizations

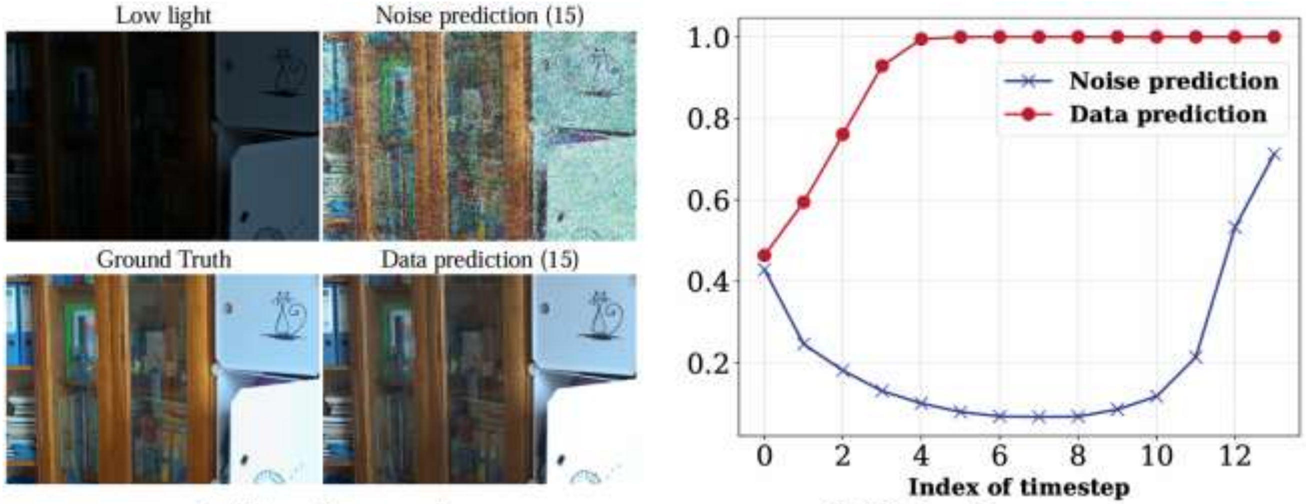
Sampler with noise prediction generate poor results in few NFEs.

- Condition of convergence for Taylor expansion

$$|\lambda - \lambda_s| < \boldsymbol{R}(s) \qquad \frac{1}{\boldsymbol{R}(t)} = \lim_{n\to\infty}\left|\frac{\boldsymbol{c}_{n+1}(t)}{\boldsymbol{c}_n(t)}\right|$$

- Ratio of convergence

$$\frac{\lambda_{t_i} - \lambda_{t_{i-1}}}{\lambda_{t_{i-1}} - \lambda_{t_{i-2}}} < \frac{f_\theta(\lambda_{t_{i-1}})}{f_\theta(\lambda_{t_{i-1}}) - f_\theta(\lambda_{t_{i-2}})}$$



(a) Sampling results. (b) Ratio of convergence.

Thanks for watching!

Paper on arxiv

Code available