Randomized midpoint method for sampling from diffusion models

Matthew Zhang[1] **Stephen Huan**[2] Jerry Huang[2] Nicholas Boffi[2] Sitan Chen[3] Sinho Chewi[4]

[1]UToronto [2]Carnegie Mellon University [3]Harvard [4]Yale

The problem: sampling diffusion models

Diffusion models are becoming popular for a variety of scientific problems, including generative modeling, sampling, tilting, and inverse problems. However, sampling these models usually involves solving an ODE or SDE, requiring many expensive score function evaluations for accurate results. Can we design efficient solvers to reduce this computational burden?

Setting the stage: a primer on diffusions

We turn data π_0 into Gaussian noise γ by an Ornstein–Uhlenbeck process

$$\mathrm{d}X_t^{\rightarrow} = -X_t^{\rightarrow}\,\mathrm{d}t + \sqrt{2}\,\mathrm{d}B_t^{\rightarrow}, \quad X_0^{\rightarrow} \sim \pi_0, \tag{OU}$$

where B_t is Brownian motion and let $\pi_t \coloneqq \mathrm{Law}(X_t^{\rightarrow})$; $\lim_{T\to\infty} \pi_T = \gamma$. The key insight of diffusions [1] is that this process admits a time-reversal

$$\mathrm{d}X_t^{\leftarrow} = \left\{+X_t^{\leftarrow} + 2\,\mathsf{s}_{T-s}(X_t^{\leftarrow})\right\}\mathrm{d}t + \sqrt{2}\,\mathrm{d}B_t^{\leftarrow}. \tag{rev-OU}$$

where $\mathsf{s}_t \coloneqq \nabla \pi_t$ is the *score*. In addition, there is a probability flow ODE

$$\mathrm{d}X_t^{\rightarrow} = \left\{-X_t^{\rightarrow} - \mathsf{s}_t(X_t^{\rightarrow})\right\}\mathrm{d}t, \quad X_0^{\rightarrow} \sim \pi_0, \tag{PF-ODE}$$
$$\mathrm{d}X_t^{\leftarrow} = \left\{+X_t^{\leftarrow} + \mathsf{s}_{T-t}(X_t^{\leftarrow})\right\}\mathrm{d}t, \quad X_0^{\leftarrow} \sim \pi_T. \tag{rev-ODE}$$

Writing $\widetilde{\mathsf{s}}_t \coloneqq \mathsf{s}_t - \nabla \log \gamma$ for the *relative score*, these equations read

$$\mathrm{d}X_t^{\leftarrow} = \left\{-X_t^{\leftarrow} + 2\,\widetilde{\mathsf{s}}_{T-s}(X_t^{\leftarrow})\right\}\mathrm{d}t + \sqrt{2}\,\mathrm{d}B_t^{\leftarrow}, \tag{rev-OU}$$
$$\mathrm{d}X_t^{\rightarrow} = -\widetilde{\mathsf{s}}_t(X_t^{\rightarrow})\,\mathrm{d}t, \quad X_0^{\rightarrow} \sim \pi_0, \tag{PF-ODE}$$
$$\mathrm{d}X_t^{\leftarrow} = +\widetilde{\mathsf{s}}_t(X_t^{\leftarrow})\,\mathrm{d}t, \quad X_0^{\leftarrow} \sim \pi_T. \tag{rev-ODE}$$

The upshot is that we can sample by solving the semi-linear SDE

$$\mathrm{d}X_t = (\lambda(t)X_t + f_t(X_t))\,\mathrm{d}t + g(t)\,\mathrm{d}B_t, \tag{SDE}$$

which generalizes to generic forward processes [2] and subsumes the ODE with $g(t) = 0$. For the ODE we recommend choosing $\lambda(t)$ to integrate the score and for the SDE we recommend integrating the relative score.

A panacea from theory

Shen and Lee (2019) originally proposed the *randomized midpoint method* for log-concave sampling. We port their algorithm to diffusion models and give a sharp analysis without the algorithmic tweaks of previous works.

The crux of the method

In integral form, the ODE $\mathrm{d}X_t = f_t(X_t)$ has solution on time $[0, h]$

$$X_h = X_0 + \int_0^h f_t(x_t)\,\mathrm{d}t.$$

Monte Carlo quadrature gives an unbiased estimator: $\tau \sim \mathrm{Unif}[0, h]$,

$$\hat{X}_h = \hat{X}_0 + h f_\tau(\hat{X}_\tau).$$

But we don't know $\hat{X}_\tau$! Bootstrap by approximating with (left) Euler,

$$\hat{X}_\tau^+ = \hat{X}_0 + \tau f_0(\hat{X}_0),$$
$$\hat{X}_h = \hat{X}_0 + h f_\tau(\hat{X}_\tau^+).$$

For the general algorithm, see Algorithm 1. SDEs are defined as stochastic integrals and we use integrating factor (with Itô's rule) for the linear term. The noise is simple since the coefficient does not depend on space X_t.

Algorithm 1: Generalized randomized midpoint kernel on $[t_0, t_1]$.

Input: current state $X_{t_0} \in \mathbb{R}^d$; step $h \coloneqq t_1 - t_0$; drift $f_t(\cdot)$; noise term $g(\cdot)$.
Input: scaling factor $\lambda(t)$; integrating factor $\omega(t) \coloneqq \exp(-\int_{t_0}^t \lambda)$; normalizing factor $\Omega(t) \coloneqq \int_{t_0}^t \omega$; inverse $\Omega^{-1}(\cdot)$; noise factor $\xi(t) \coloneqq \int_{t_0}^t (\omega g)^2$.

1. Draw the randomized midpoint. Sample $U \sim \mathrm{Unif}(0,1)$ and set

$$\tau = \Omega^{-1}((1-U)\,\Omega(t_0) + U\,\Omega(t_1)) \quad \text{i.e., with density } p(\tau) \propto 1_{[t_0,t_1]}(t)\,\omega(t).$$

2. Midpoint prediction for $X_{t_0+\tau}^+$. Draw $Z_1 \sim \mathcal{N}(0, I_d)$ and set the OU noise $\xi^+ \coloneqq [\sqrt{|\xi(t_0+\tau)|}/\omega(t_0+\tau)]\,Z_1$. Then

$$X_{t_0+\tau_k}^+ = \frac{\omega(t_0)}{\omega(t_0+\tau)}\,X_{t_0} + \frac{\Omega(t_0+\tau)}{\omega(t_0+\tau)}\,f_{t_0}(X_{t_0}) + \xi^+.$$

3. Full-step update for X_{t_1}. Draw $Z_2 \sim \mathcal{N}(0, I_d)$ independent of Z_1 and set

$$\xi = \frac{\omega(t_0+\tau)}{\omega(t_1)}\,\xi^+ + \frac{\sqrt{|\xi(t_1) - \xi(t_0+\tau)|}}{\omega(t_1)}\,Z_2.$$

Compute the score at the randomized time and update

$$X_{t_1} = \frac{\omega(t_0)}{\omega(t_1)}\,X_{t_0} + \frac{\Omega(t_1)}{\omega(t_1)}\,f_{t_0+\tau}(X_{t_0+\tau}^+) + \xi.$$

Quick notes on the theory

Assumptions standard in the literature: data has finite second moment, ε^2 accurate score, estimated score has (time-varying) Lipschitz constant. Result: $\mathcal{O}(\sqrt{d}/\epsilon)$ steps in W_2 + KL (not early stopping). Difficulties:

- Anticipating: midpoint "sees into the future", precluding Girsanov.
- Distinguish weak error ("bias") vs strong error ("variance").
- Typical Wasserstein analyses (Grönwall) blow up exponentially.

We use the framework of ✨shifted composition✨ to address these issues!

Denoisers, DDPMs, and linear interpolants

EDM [2] specialize (SDE) to the following parametrization in (c, σ, β),

$$\mathrm{d}X_t = \left[\frac{\dot{c}(t)}{c(t)}X_t - (c(t)^2\dot{\sigma}(t)\sigma(t) \pm \beta(t)\sigma(t)^2 c(t)^2)\mathsf{s}_t(X_t)\right]\mathrm{d}t + \sqrt{2\beta(t)}\sigma(t)c(t)\,\mathrm{d}B_t. \tag{EDM}$$

The score admits an intuitive *denoiser* representation by Tweedie's formula

$$\mathsf{s}_t(x) = \frac{D(x/c(t); \sigma(t)) - x/c(t)}{c(t)\,\sigma(t)^2},$$

where $D(\cdot; \sigma)$ predicts the data x given a noisy observation $x + \sigma z$, $z \sim \gamma$. Under the DDIM choice [4] of $\sigma(t) = t$, $c(t) = 1$, [2] observe a single step of Euler to $t = 0$ has the appealing property of returning the denoiser. In interpolants [5], the linear choice $I_t = (1-t)x_0 + tx_1$ also has this property and recovers the DDPM ancestral sampler $\mathbb{E}[I_{t+h} | I_t = X_t]$ [6].

Choosing the noise coefficient

By a standard calculation with Girsanov's theorem to minimize path KL,

$$\beta(t) = \dot{\sigma}(t)/\sigma(t), \tag{Girsanov-β}$$

recovering the choice of [4] where the forward process of (EDM) does not contain the score. We also reverse engineer the sampler of [2] for

$$\beta(t) = 1_{[S_{\mathsf{tmin}}, S_{\mathsf{tmax}}]}(t) S_{\mathsf{churn}}/T. \tag{EDM-β}$$

Sampling with the OU process

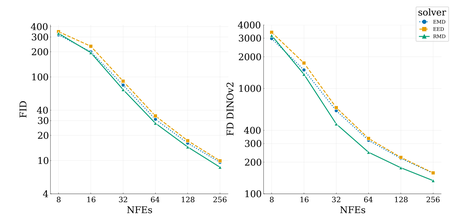


Figure: Sampling with the SDE on the OU process (rev-OU).

Sampling with the probability flow ODE

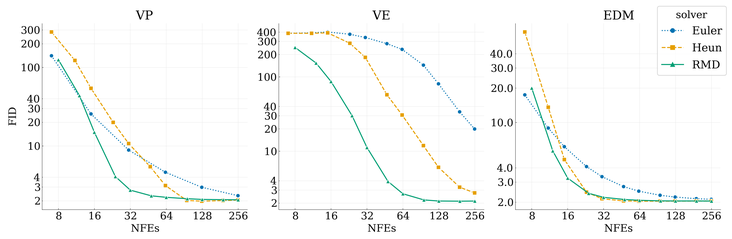


Figure: Sampling over a range of processes (EDM) with the ODE ($g(t) = 0$).

Sampling with the SDE and tuned noise

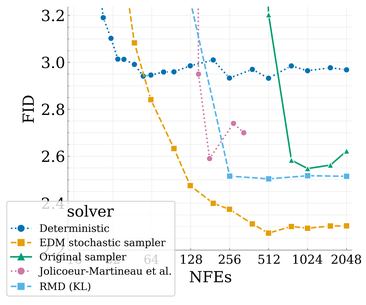


Figure: Sampling with the SDE and Girsanov-tuned noise coefficient (Girsanov-β).

Paper and code

https://arxiv.org/abs/2511.04844
https://github.com/stephen-huan/edm_rmd

References

[1] Y. Song, J. Sohl-Dickstein, D. P. Kingma, A. Kumar, S. Ermon, and B. Poole, "Score-based generative modeling through stochastic differential equations," in *International Conference on Learning Representations*, 2021.

[2] T. Karras, M. Aittala, T. Aila, and S. Laine, *Elucidating the Design Space of Diffusion-Based Generative Models*, Oct. 2022. DOI: 10.48550/arXiv.2206.00364. arXiv: 2206.00364 [cs, stat].

[3] R. Shen and Y. T. Lee, *The Randomized Midpoint Method for Log-Concave Sampling*, Sep. 2019. DOI: 10.48550/arXiv.1909.05503. arXiv: 1909.05503 [cs].

[4] J. Song, C. Meng, and S. Ermon, "Denoising diffusion implicit models," in *International Conference on Learning Representations*, 2021.

[5] M. S. Albergo, N. M. Boffi, and E. Vanden-Eijnden, *Stochastic Interpolants: A Unifying Framework for Flows and Diffusions*, Nov. 2023. DOI: 10.48550/arXiv.2303.08797. arXiv: 2303.08797 [cond-mat].

[6] J. Ho, A. Jain, and P. Abbeel, "Denoising diffusion probabilistic models," *Advances in Neural Information Processing Systems*, vol. 33, pp. 6840–6851, 2020.

DOE CSGF