Non-asymptotic convergence guarantees for probability flow ODEs under weak log-concavity

Gitte Kremling, Francesco Iafrate, Mahsa Taheri, Johannes Lederer

Department of Mathematics, University of Hamburg

Universität Hamburg
DER FORSCHUNG | DER LEHRE | DER BILDUNG

Motivation

- Score-based generative models produce very good results in practice.
- Are there theoretical guarantees for their good performance?
- Drawbacks of existing works:
  - Difference measured in TV or KL distance, as in [1] (less interpretable and stable in high dimensions)
  - Strong assumptions on data distribution, as in [2] (e.g. excl. Gaussian mixtures)
  - Only apply to specific forward SDEs, as in [3] (e.g. OU process)

Aim of our project:
Establish general error bounds in $\mathcal{W}_2$-distance relying on weaker assumptions on the data distribution

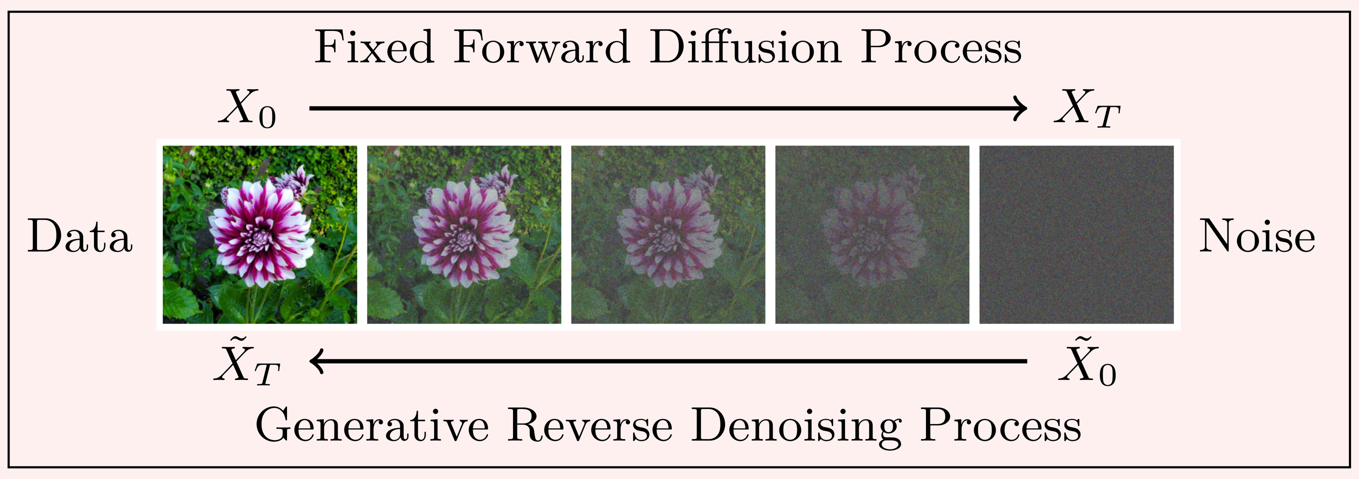
Probability Flow ODE

Forward SDE $(X_t \sim p_t)$

$$\mathrm{d}X_t = -f(t)X_t\,\mathrm{d}t + g(t)\,\mathrm{d}B_t, \quad t \in [0,T]$$
$$X_0 \sim p_0$$

Reverse ODE $(\tilde{X}_t \sim p_{T-t})$

$$\frac{\mathrm{d}\tilde{X}_t}{\mathrm{d}t} = f(T-t)\tilde{X}_t + \frac{1}{2}g^2(T-t)\nabla \log p_{T-t}(\tilde{X}_t)$$
$$\tilde{X}_0 \sim p_T$$



Implementation of Reverse ODE based on:

Approximation 1: Initialization $\widehat{p}_T \approx p_T$

$$\frac{\mathrm{d}Y_t}{\mathrm{d}t} = f(T-t)Y_t + \frac{1}{2}g^2(T-t)\nabla \log p_{T-t}(Y_t)$$
$$Y_0 \sim \widehat{p}_T$$

Approximation 2: Discretization $\overset{h}{\longleftrightarrow}$ $0 \quad t_1 \quad \cdots \quad t_{K-1} \quad T$

$$\frac{\mathrm{d}\widehat{Y}_t}{\mathrm{d}t} = f(T-t)\widehat{Y}_t + \frac{1}{2}g^2(T-t)\nabla \log p_{T-t_{k-1}}(\widehat{Y}_{t_{k-1}})$$
$$\widehat{Y}_0 \sim \widehat{p}_T$$

Approximation 3: Score matching $s_\theta(x,t) \approx p_t(x)$

$$\frac{\mathrm{d}\widehat{Z}_t}{\mathrm{d}t} = f(T-t)\widehat{Z}_t + \frac{1}{2}g^2(T-t)s_\theta(\widehat{Z}_{t_{k-1}}, T-t_{k-1})$$
$$\widehat{Z}_0 \sim \widehat{p}_T$$

$\rightarrow$ New sample $\widehat{Z}_T$ approximately follows distribution p_0

Main Result

Non-asymptotic error bound for the distance between the approximated sample distribution and the true data distribution (under the assumptions listed below):

$$\mathcal{W}_2\Big(\mathcal{L}\Big(\widehat{Z}_T\Big), p_0\Big) \leq E_0(f,g,T) + E_1(f,g,K,h) + E_2(f,g,K,h,\mathcal{E})$$

Key properties of the individual error components:

	$E_0(f,g,T)$	$E_1(f,g,K,h)$	$E_3(f,g,K,h,\mathcal{E})$
Error source	Initialization	Discretization	Score matching
Vanishes with	$T \to \infty$	$h \to 0$	$\mathcal{E} \to 0$
OU process*	$\mathcal{O}\left(e^{-T}\sqrt{d}\right)$	$\mathcal{O}\left(e^{Th}Th\left(\sqrt{d}+T\right)\right)$	$\mathcal{O}(e^{Th}T\mathcal{E})$
Error $\leq \varepsilon$ if*	$T \geq \mathcal{O}\left(\log\left(\frac{\sqrt{d}}{\varepsilon}\right)\right)$	$h \leq \mathcal{O}\left(\frac{\varepsilon}{\sqrt{d}\log\left(\frac{\sqrt{d}}{\varepsilon}\right)}\right)$	$\mathcal{E} \leq \mathcal{O}\left(\frac{\varepsilon}{\log\left(\frac{\sqrt{d}}{\varepsilon}\right)}\right)$

*Specific choice: $f(t) \equiv 1$ and $g(t) \equiv \sqrt{2}$

Main finding: Same asymptotics as under the stronger assumption in [2]!

Assumptions

1. **(Regularity of the data distribution)**
   - $p_0 \in C^2(\mathbb{R}^d)$ and positive everywhere
   - p_0 is (α_0, M_0)-weakly log-concave, i.e.
     $\langle \nabla \log p_0(x) - \nabla \log p_0(y), x-y\rangle \leq -\alpha_0\|x-y\|^2 + 2\sqrt{M_0}\tanh\left(\frac{\sqrt{M_0}}{2}\|x-y\|\right)\|x-y\|$
   - $\log p_0$ is L_0-smooth, i.e.
     $\|\nabla \log p_0(x) - \nabla \log p_0(y)\| \leq L_0\|x-y\|$
2. **(Lipschitz-continuity in time of the score function)**
   $$\sup_{k,t\in[t_{k-1},t_k]} \left\|\nabla \log p_{T-t}(x) - \nabla \log p_{T-t_{k-1}}(x)\right\| \leq L_1 h(1+\|x\|)$$
3. **(Boundedness of the score matching error)**
   $$\sup_k \left\|\nabla \log p_{T-t_{k-1}}(\widehat{Z}_{t_{k-1}}) - s_\theta(\widehat{Z}_{t_{k-1}}, T-t_{k-1})\right\|_{L_2} \leq \mathcal{E}$$

More Examples

Other choices of the drift f and the diffusion g result in the following heuristics for the choice of hyperparameters:

f	g	T	h	$\mathcal{E}$
0	ae^{bt}	$\mathcal{O}\left(\log\left(\frac{\sqrt{d}}{\varepsilon}\right)\right)$	$\mathcal{O}\left(\frac{\varepsilon^3}{d^{\frac{3}{2}}}\right)$	$\mathcal{O}\left(\frac{\varepsilon^2}{\sqrt{d}}\right)$
0	$(b+at)^c$	$\mathcal{O}\left(\left(\frac{d}{\varepsilon^2}\right)^{\frac{1}{2c+1}}\right)$	$\mathcal{O}\left(\frac{\varepsilon^3}{d^{\frac{3}{2}}}\right)$	$\mathcal{O}\left(\frac{\varepsilon^2}{\sqrt{d}}\right)$
$\frac{b}{2}$	$\sqrt{b}$	$\mathcal{O}\left(\log\left(\frac{\sqrt{d}}{\varepsilon}\right)\right)$	$\mathcal{O}\left(\frac{\varepsilon}{\sqrt{d}\log\left(\frac{\sqrt{d}}{\varepsilon}\right)}\right)$	$\mathcal{O}\left(\frac{\varepsilon}{\log\left(\frac{\sqrt{d}}{\varepsilon}\right)}\right)$
$\frac{b+at}{2}$	$\sqrt{b+at}$	$\mathcal{O}\left(\left(\log\left(\frac{\sqrt{d}}{\varepsilon}\right)\right)^{\frac{1}{2}}\right)$	$\mathcal{O}\left(\frac{\varepsilon}{\sqrt{d}\log\left(\frac{\sqrt{d}}{\varepsilon}\right)}\right)$	$\mathcal{O}\left(\frac{\varepsilon}{\log\left(\frac{\sqrt{d}}{\varepsilon}\right)}\right)$
$\frac{(b+at)^\rho}{2}$	$(b+at)^{\frac{\rho}{2}}$	$\mathcal{O}\left(\left(\log\left(\frac{\sqrt{d}}{\varepsilon}\right)\right)^{\frac{1}{\rho+1}}\right)$	$\mathcal{O}\left(\frac{\varepsilon}{\sqrt{d}\log\left(\frac{\sqrt{d}}{\varepsilon}\right)}\right)$	$\mathcal{O}\left(\frac{\varepsilon}{\log\left(\frac{\sqrt{d}}{\varepsilon}\right)}\right)$

References

[1] A. Wibisono and K. Yang. Convergence in kl divergence of the inexact langevin algorithm with application to score-based generative models. *arXiv:2211.01512*, 2022.

[2] X. Gao and L. Zhu. Convergence analysis for general probability flow odes of diffusion models in wasserstein distances. *arXiv preprint arXiv:2401.17958*, 2024.

[3] M. Gentiloni-Silveri and A. Ocello. Beyond log-concavity and score regularity: Improved convergence bounds for score-based generative models in w2-distance. *arXiv preprint arXiv:2501.02298*, 2025.