Universität Hamburg
DER FORSCHUNG | DER LEHRE | DER BILDUNG

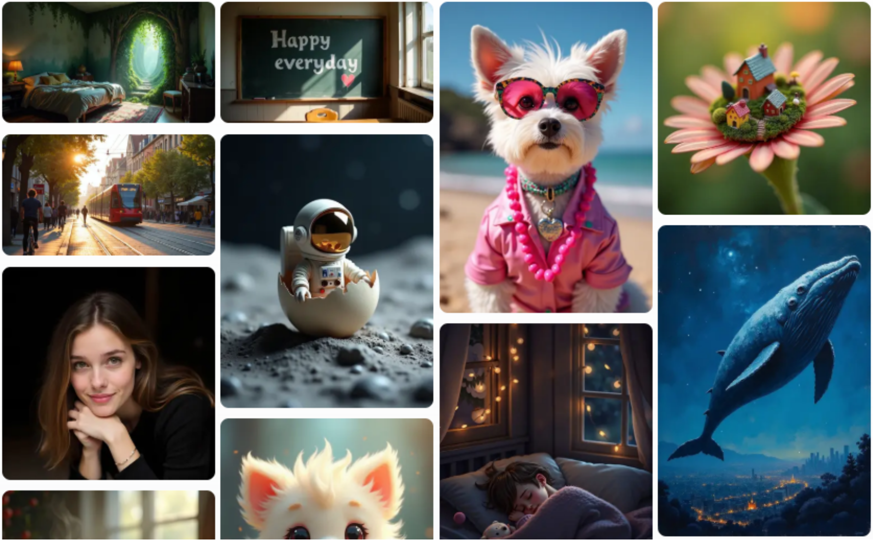
Non-asymptotic error bounds for probability flow ODEs under weak log-concavity

Joint work with F. Iafrate, M. Taheri and J. Lederer

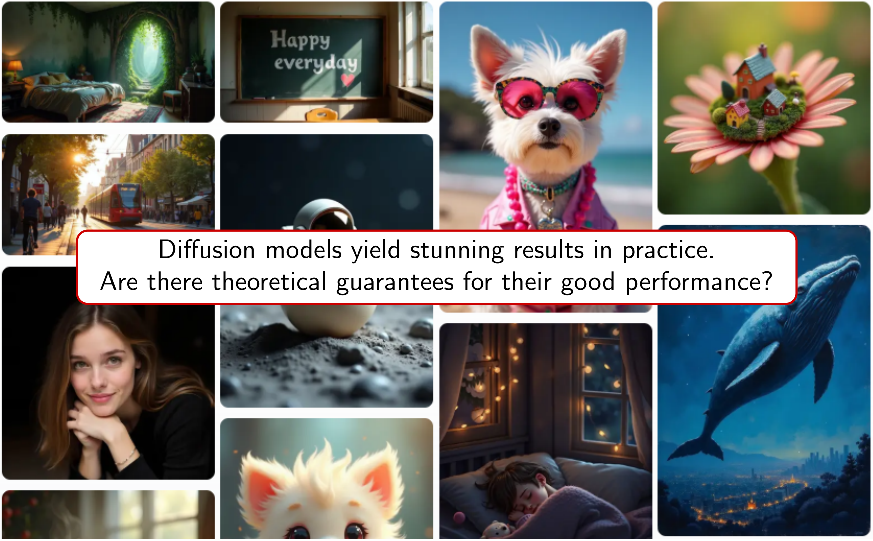
Gitte Kremling

December 15, 2025

Motivation



Motivation



Diffusion models yield stunning results in practice.
Are there theoretical guarantees for their good performance?

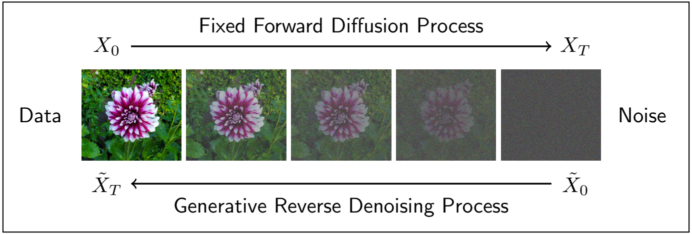
Source: stablediffusionweb.com

Diffusion Models — General Idea

Goal: Generate new samples from a probability density function p_0 given a data set of i.i.d. observations.

Training: Systematically transform the given data samples into noise and learn how to reverse this process.

Generation of new samples: Start with noise and use the reverse process to transform it to a data sample.



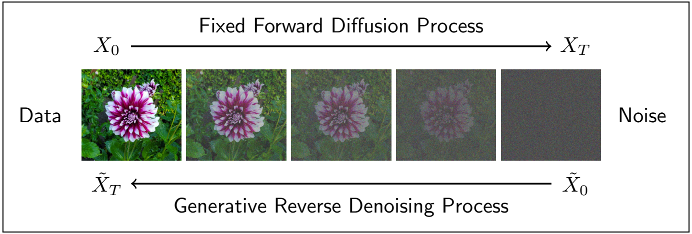
Probability Flow ODE

Forward SDE $\big(X_t \sim p_t\big)$

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

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

$$\frac{\mathrm{d}\tilde{X}_t}{\mathrm{d}t} = f(T-t)\tilde{X}_t + \frac{1}{2}g^2(T-t)\nabla \log p_{T-t}(\tilde{X}_t), \quad t \in [0,T]$$
$$\tilde{X}_0 \sim p_T$$



Probability Flow ODE

Forward SDE $\big(X_t \sim p_t\big)$

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

Remark: Choice of f and g

$$X_t = e^{-\int_0^t f(s)\,\mathrm{d}s}X_0 + \underbrace{\int_0^t e^{-\int_s^t f(v)\,\mathrm{d}v} g(s)\,\mathrm{d}B_s}_{\sim\mathcal{N}\left(0,\int_0^t e^{-2\int_s^t f(v)\,\mathrm{d}v} g^2(s)\,\mathrm{d}s\cdot I_d\right)=:\hat{p}_t} \quad \sim p_t$$

Variance-**E**xploding **SDE**: $\quad f(t) = 0, \quad g(t) = \sqrt{\frac{\mathrm{d}[\sigma^2(t)]}{\mathrm{d}t}}$

$$\Rightarrow \hat{p}_t = \mathcal{N}(0,\sigma^2(t))$$

Variance-**P**reserving **SDE**: $\quad f(t) = \frac{1}{2}\beta(t), \quad g(t) = \sqrt{\beta(t)}$

$$\Rightarrow \hat{p}_t = \mathcal{N}(0, 1 - e^{\int_0^t \beta(s)\,\mathrm{d}s}), \quad p_t \xrightarrow[t\to\infty]{} \mathcal{N}(0,1)$$

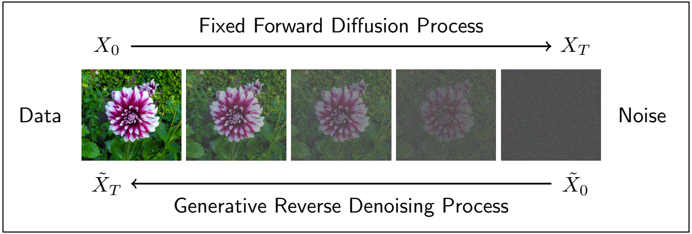
Probability Flow ODE

Forward SDE $\big(X_t \sim p_t\big)$

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

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

$$
\begin{aligned}
\frac{\mathrm{d}\tilde{X}_t}{\mathrm{d}t} &= f(T-t)\tilde{X}_t + \frac{1}{2}g^2(T-t)\nabla \log p_{T-t}(\tilde{X}_t), \quad t \in [0,T] \\
\tilde{X}_0 &\sim p_T
\end{aligned}
$$



Implementation of Probability Flow ODE

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

$$\frac{\mathrm{d}Y_t}{\mathrm{d}t} = f(T-t)Y_t + \frac{1}{2}g^2(T-t)\nabla \log p_{T-t}(Y_t), \quad t \in [0,T]$$
$$Y_0 \sim \widehat{p}_T$$

Recall:

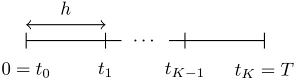
$$X_t = e^{-\int_0^t f(s)\,\mathrm{d}s} X_0 + \underbrace{\int_0^t e^{-\int_s^t f(v)\,\mathrm{d}v} g(s)\,\mathrm{d}B_s}_{\sim \mathcal{N}\left(0, \int_0^t e^{-2\int_s^t f(v)\,\mathrm{d}v} g^2(s)\,\mathrm{d}s \cdot I_d\right) =: \hat{p}_t} \sim p_t$$

Implementation of Probability Flow ODE

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

$$\frac{\mathrm{d}Y_t}{\mathrm{d}t} = f(T-t)Y_t + \frac{1}{2}g^2(T-t)\nabla \log p_{T-t}(Y_t), \quad t \in [0,T]$$
$$Y_0 \sim \widehat{p}_T$$

Approximation 2: Discretization



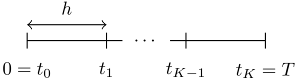
$$\frac{\mathrm{d}\widehat{Y}_t}{\mathrm{d}t} = f(T-t)\widehat{Y}_t + \frac{1}{2}g^2(T-t)\nabla \log p_{T-t_{k-1}}(\widehat{Y}_{t_{k-1}}), \quad t \in [t_{k-1}, t_k]$$
$$\widehat{Y}_0 \sim \widehat{p}_T$$

Implementation of Probability Flow ODE

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

$$\frac{\mathrm{d}Y_t}{\mathrm{d}t} = f(T-t)Y_t + \frac{1}{2}g^2(T-t)\nabla \log p_{T-t}(Y_t), \quad t \in [0,T]$$
$$Y_0 \sim \widehat{p}_T$$

Approximation 2: Discretization



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

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

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

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

Establishing Theoretical Guarantees

Evaluate performance of diffusion models by comparing $\mathcal{L}(\widehat{Z}_T)$ and p_0.

Drawbacks of existing works:

- Difference measured in TV or KL distance, as in Wibisono and Yang (2022) $\rightarrow$ less interpretable and stable in high dimensions
- Strong assumptions on the data distribution, as in Gao and Zhu (2024) $\rightarrow$ very restrictive, e.g. excluding Gaussian mixtures
- Only considering specific forward SDEs, as in Gentiloni-Silveri and Ocello (2025) $\rightarrow$ e.g. OU process $f(t) = 1$ and $g(t) = \sqrt{2}$

Establishing Theoretical Guarantees

Evaluate performance of diffusion models by comparing $\mathcal{L}(\widehat{Z}_T)$ and p_0.

Drawbacks of existing works:

- Difference measured in TV or KL distance, as in Wibisono and Yang (2022) $\rightarrow$ less interpretable and stable in high dimensions

- Strong assumptions on the data distribution, as in Gao and Zhu (2024) $\rightarrow$ very restrictive, e.g. excluding Gaussian mixtures

- Only considering specific forward SDEs, as in Gentiloni-Silveri and Ocello (2025) $\rightarrow$ e.g. OU process $f(t) = 1$ and $g(t) = \sqrt{2}$

Aim of our project:

Establish error bounds in $\mathcal{W}_2$-distance for general f and g relying on weaker assumptions on the data distribution

Assumptions in Gao and Zhu (2024)

(1) Regularity of the data distribution

- $p_0 \in C^2(\mathbb{R}^d)$ and positive everywhere
- p_0 is α_0-strongly log-concave, i.e.
  $\langle \nabla \log p_0(x) - \nabla \log p_0(y), x - y \rangle \leq -\alpha_0 \|x - y\|^2$
- $\nabla \log p_0$ is L_0-Lipschitz continuous, 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

$$\sup_k \left\| \nabla \log p_{T-t_{k-1}}(\widehat{Z}_{t_{k-1}}) - s_\theta(\widehat{Z}_{t_{k-1}}, T - t_{k-1}) \right\|_{L_2} \leq \mathcal{E}$$

Assumptions in our work

(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.
  $\langle \nabla \log p_0(x) - \nabla \log p_0(y), x - y \rangle \leq -\alpha_0 \|x - y\|^2 + f_{M_0}(\|x - y\|)\|x - y\|$
- $\nabla \log p_0$ is L_0-Lipschitz continuous, 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

$$\sup_k \left\| \nabla \log p_{T-t_{k-1}}(\widehat{Z}_{t_{k-1}}) - s_\theta(\widehat{Z}_{t_{k-1}}, T - t_{k-1}) \right\|_{L_2} \leq \mathcal{E}$$

Assumptions in our work

(1) **Regularity of the data distribution**

Strongly log-concave on large scales, while allowing small local non-concave fluctuations

- $p_0 \in C^2(\mathbb{R}^d)$ and positive everywhere
- p_0 is (α_0, M_0)-weakly log-concave, i.e.
  $\langle \nabla \log p_0(x) - \nabla \log p_0(y), x - y \rangle \leq -\alpha_0 \|x - y\|^2 + f_{M_0}(\|x - y\|)\|x - y\|$
- $\nabla \log p_0$ is L_0-Lipschitz continuous, i.e.
  $\|\nabla \log p_0(x) - \nabla \log p_0(y)\| \leq L_0 \|x - y\|$

Includes multi-modal distributions such as Gaussian mixture models

(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**

$$\sup_k \left\| \nabla \log p_{T-t_{k-1}}(\widehat{Z}_{t_{k-1}}) - s_\theta(\widehat{Z}_{t_{k-1}}, T - t_{k-1}) \right\|_{L_2} \leq \mathcal{E}$$

Weak log-concavity

Definition: p_0 is said to be (α_0, M_0)-weakly log-concave if

$$\langle \nabla \log p_0(x) - \nabla \log p_0(y), x - y \rangle \leq -\alpha_0 \|x - y\|^2 + f_{M_0}(\|x - y\|)\|x - y\|$$

where $f_M(r) = 2\sqrt{M} \tanh\left(\frac{1}{2}\sqrt{M} r\right)$.

- Implies that $\langle \nabla \log p_0(x) - \nabla \log p_0(y), x - y \rangle \leq -(\alpha_0 - M_0)\|x - y\|^2$
$$\Leftrightarrow \nabla^2 \log p_0(x) \preceq -(\alpha_0 - M_0) I_d$$
  and hence strong log-concavity of p_0 if $\alpha_0 - M_0 > 0$.

Weak log-concavity

Definition: p_0 is said to be (α_0, M_0)-weakly log-concave if

$$\langle \nabla \log p_0(x) - \nabla \log p_0(y), x - y \rangle \leq -\alpha_0 \|x - y\|^2 + f_{M_0}(\|x - y\|)\|x - y\|$$

where $f_M(r) = 2\sqrt{M} \tanh\left(\frac{1}{2}\sqrt{M} r\right)$.

- Implies that $\langle \nabla \log p_0(x) - \nabla \log p_0(y), x - y \rangle \leq -(\alpha_0 - M_0)\|x - y\|^2$
$$\Leftrightarrow \nabla^2 \log p_0(x) \preceq -(\alpha_0 - M_0) I_d$$
  and hence strong log-concavity of p_0 if $\alpha_0 - M_0 > 0$.
- $f_M(r)$ can be replaced by a different function satisfying specific requirements. (Conforti, 2024)
- Every GMM with full rank covariance matrices is weakly log-concave. (Gentiloni-Silveri and Ocello, 2025)
- Weak log-concavity requires sub-gaussian tails.

Main Result

Non-asymptotic error bound for the distance between the approximated sample distribution and the true data distribution under Assumptions (1)-(3):

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

Key properties of the individual error components:

	$E_0(f,g,T)$	$E_1(f,g,K,h)$	$E_2(f,g,K,h,\mathcal{E})$
Error source	Initialization	Discretization	Score matching
Vanishes with	$T \to \infty$	$h \to 0$	$\mathcal{E} \to 0$
OU process*	$\mathcal{O}\Big(e^{-T}\sqrt{d}\Big)$	$\mathcal{O}\Big(e^{Th}Th\Big(\sqrt{d}+T\Big)\Big)$	$\mathcal{O}(e^{Th}T\mathcal{E})$
Error $\leq \varepsilon$ if*	$T \geq \mathcal{O}\Big(\log\Big(\frac{\sqrt{d}}{\varepsilon}\Big)\Big)$	$h \leq \mathcal{O}\left(\frac{\varepsilon}{\sqrt{d}\log\left(\frac{\sqrt{d}}{\varepsilon}\right)}\right)$	$\mathcal{E} \leq \mathcal{O}\left(\frac{\varepsilon}{\log\left(\frac{\sqrt{d}}{\varepsilon}\right)}\right)$

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

Main Result

Non-asymptotic error bound for the distance between the approximated sample distribution and the true data distribution under Assumptions (1)-(3):

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

Key properties of the individual error components:

	$E_0(f,g,T)$	$E_1(f,g,K,h)$	$E_2(f,g,K,h,\mathcal{E})$
Error source	Initialization	Discretization	Score matching
Vanishes with	$T \to \infty$	$h \to 0$	$\mathcal{E} \to 0$
OU process*	$\mathcal{O}\Big(e^{-T}\sqrt{d}\Big)$	$\mathcal{O}\Big(e^{Th}Th\Big(\sqrt{d}+T\Big)\Big)$	$\mathcal{O}(e^{Th}T\mathcal{E})$
Error $\leq \varepsilon$ if*	$T \geq \mathcal{O}\Big(\log\Big(\frac{\sqrt{d}}{\varepsilon}\Big)\Big)$	$h \leq \mathcal{O}\left(\frac{\varepsilon}{\sqrt{d}\log\left(\frac{\sqrt{d}}{\varepsilon}\right)}\right)$	$\mathcal{E} \leq \mathcal{O}\left(\frac{\varepsilon}{\log\left(\frac{\sqrt{d}}{\varepsilon}\right)}\right)$

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

Main finding: Same asymptotics as under the strong log-concavity assumption analyzed in Gao and Zhu (2024)!

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	$(b+at)^c$	$\mathcal{O}\left(\left(\frac{d}{\varepsilon^2}\right)^{\frac{1}{2c+1}}\right)$	$\mathcal{O}\left(\frac{\varepsilon^3}{d^{\frac{3}{2}}}\right)$	$\mathcal{O}\left(\frac{\varepsilon^2}{\sqrt{d}}\right)$
$\frac{b}{2}$	$\sqrt{b}$	$\mathcal{O}\left(\log\left(\frac{\sqrt{d}}{\varepsilon}\right)\right)$	$\mathcal{O}\left(\frac{\varepsilon}{\sqrt{d}\log\left(\frac{\sqrt{d}}{\varepsilon}\right)}\right)$	$\mathcal{O}\left(\frac{\varepsilon}{\log\left(\frac{\sqrt{d}}{\varepsilon}\right)}\right)$
$\frac{b+at}{2}$	$\sqrt{b+at}$	$\mathcal{O}\left(\left(\log\left(\frac{\sqrt{d}}{\varepsilon}\right)\right)^{\frac{1}{2}}\right)$	$\mathcal{O}\left(\frac{\varepsilon}{\sqrt{d}\log\left(\frac{\sqrt{d}}{\varepsilon}\right)}\right)$	$\mathcal{O}\left(\frac{\varepsilon}{\log\left(\frac{\sqrt{d}}{\varepsilon}\right)}\right)$
$\frac{(b+at)^\rho}{2}$	$(b+at)^{\frac{\rho}{2}}$	$\mathcal{O}\left(\left(\log\left(\frac{\sqrt{d}}{\varepsilon}\right)\right)^{\frac{1}{\rho+1}}\right)$	$\mathcal{O}\left(\frac{\varepsilon}{\sqrt{d}\log\left(\frac{\sqrt{d}}{\varepsilon}\right)}\right)$	$\mathcal{O}\left(\frac{\varepsilon}{\log\left(\frac{\sqrt{d}}{\varepsilon}\right)}\right)$

Main finding: Same asymptotics as under the strong log-concavity assumption analyzed in Gao and Zhu (2024)!

Proof Sketch

Recall: $p_0 \sim \tilde{X}_T \overset{\text{Initialization}}{\approx} Y_T \overset{\text{Discretization}}{\approx} \widehat{Y}_T \overset{\text{Score matching}}{\approx} \widehat{Z}_T$

- $\mathcal{W}_2\Big(p_0, \mathcal{L}(\widehat{Z}_T)\Big) \leq \mathcal{W}_2\Big(p_0, \mathcal{L}(Y_T)\Big) + \mathcal{W}_2\Big(\mathcal{L}(Y_T), \mathcal{L}(\widehat{Z}_T)\Big)$

Proof Sketch

Recall: $p_0 \sim \tilde{X}_T \overset{\text{Initialization}}{\approx} Y_T \overset{\text{Discretization}}{\approx} \widehat{Y}_T \overset{\text{Score matching}}{\approx} \widehat{Z}_T$

- $\mathcal{W}_2\Big(p_0, \mathcal{L}(\widehat{Z}_T)\Big) \leq \mathcal{W}_2\Big(p_0, \mathcal{L}(Y_T)\Big) + \mathcal{W}_2\Big(\mathcal{L}(Y_T), \mathcal{L}(\widehat{Z}_T)\Big)$
- $\mathcal{W}_2\Big(p_0, \mathcal{L}(Y_T)\Big) = \mathcal{W}_2\Big(\mathcal{L}(\tilde{X}_T), \mathcal{L}(Y_T)\Big) \leq E_0(f, g, T)$

Proof Sketch

Recall: $p_0 \sim \tilde{X}_T \overset{\text{Initialization}}{\approx} Y_T \overset{\text{Discretization}}{\approx} \widehat{Y}_T \overset{\text{Score matching}}{\approx} \widehat{Z}_T$

- $\mathcal{W}_2\Big(p_0, \mathcal{L}(\widehat{Z}_T)\Big) \leq \mathcal{W}_2\Big(p_0, \mathcal{L}(Y_T)\Big) + \mathcal{W}_2\Big(\mathcal{L}(Y_T), \mathcal{L}(\widehat{Z}_T)\Big)$
- $\mathcal{W}_2\Big(p_0, \mathcal{L}(Y_T)\Big) = \mathcal{W}_2\Big(\mathcal{L}(\tilde{X}_T), \mathcal{L}(Y_T)\Big) \leq E_0(f, g, T)$
- Pick a coupling between Y_t and $\widehat{Z}_t$ such that $Y_0 = \widehat{Z}_0$ a.s., then

$$\begin{aligned}\mathcal{W}_2\Big(\mathcal{L}(Y_T), \mathcal{L}(\widehat{Z}_T)\Big) &\leq \Big\| Y_{t_K} - \widehat{Z}_{t_K} \Big\|_{L_2} \leq \alpha_{K,h} \Big\| Y_{t_{K-1}} - \widehat{Z}_{t_{K-1}} \Big\|_{L_2} + \beta_{K,h} \\ &\leq \sum_{k=1}^{K} \left(\prod_{j=k+1}^{K} \alpha_{k,h} \right) \beta_{k,h} \\ &= E_1(f, g, K, h) + E_2(f, g, K, h, \mathcal{E})\end{aligned}$$

Proof Sketch

Recall: $p_0 \sim \tilde{X}_T \overset{\text{Initialization}}{\approx} Y_T \overset{\text{Discretization}}{\approx} \widehat{Y}_T \overset{\text{Score matching}}{\approx} \widehat{Z}_T$

- $\mathcal{W}_2\Big(p_0, \mathcal{L}(\widehat{Z}_T)\Big) \leq \mathcal{W}_2\Big(p_0, \mathcal{L}(Y_T)\Big) + \mathcal{W}_2\Big(\mathcal{L}(Y_T), \mathcal{L}(\widehat{Z}_T)\Big)$
- $\mathcal{W}_2\Big(p_0, \mathcal{L}(Y_T)\Big) = \mathcal{W}_2\Big(\mathcal{L}(\tilde{X}_T), \mathcal{L}(Y_T)\Big) \leq E_0(f, g, T)$
- Pick a coupling between Y_t and $\widehat{Z}_t$ such that $Y_0 = \widehat{Z}_0$ a.s., then

$$\begin{aligned}
\mathcal{W}_2\Big(\mathcal{L}(Y_T), \mathcal{L}(\widehat{Z}_T)\Big) &\leq \Big\| Y_{t_K} - \widehat{Z}_{t_K} \Big\|_{L_2} \leq \alpha_{K,h} \Big\| Y_{t_{K-1}} - \widehat{Z}_{t_{K-1}} \Big\|_{L_2} + \beta_{K,h} \\
&\leq \sum_{k=1}^{K} \left(\prod_{j=k+1}^{K} \alpha_{k,h} \right) \beta_{k,h} \\
&= E_1(f, g, K, h) + E_2(f, g, K, h, \mathcal{E})
\end{aligned}$$

Highlighted inequalities make use of the fact:
$\log p_0$ weakly concave and Lipschitz $\Rightarrow$ $\log p_t$ weakly concave and Lipschitz

Regime Shifting Property

Known result: The forward SDE preserves strong log-concavity , i.e.

$$\log p_0 \text{ is strongly concave} \quad \Rightarrow \quad \log p_t \text{ is strongly concave}$$

New result: The forward SDE induces strong log-concavity, i.e.
for some (precisely defined) τ, it holds that

$$\log p_0 \text{ is weakly concave} \quad \Rightarrow \quad \begin{cases} \log p_t \text{ is weakly concave for } t \in [0, \tau) \\ \log p_t \text{ is strongly concave for } t \in [\tau, \infty) \end{cases}$$

This property is crucial to show that the asymptotics remain the same under the weakly log-concave assumption.

Conclusion

Link to paper:

- Extended convergence theory for score-based generative models to more realistic data distributions
- Relaxed strong to weak log-concavity assumption, allowing for multi-modal distributions such as GMMs
- Remarkably, the asymptotics remain the same under the weaker assumption
- Result applies to general drift and diffusion functions f and g
- Error bound can be translated to concrete heuristics for the choice of time horizon T, step size h and score-matching error $\mathcal{E}$

Conclusion

Link to paper:

- Extended convergence theory for score-based generative models to more realistic data distributions
- Relaxed strong to weak log-concavity assumption, allowing for multi-modal distributions such as GMMs
- Remarkably, the asymptotics remain the same under the weaker assumption
- Result applies to general drift and diffusion functions f and g
- Error bound can be translated to concrete heuristics for the choice of time horizon T, step size h and score-matching error $\mathcal{E}$

Outlook: Reverse SDE, vector-valued f and g, even weaker assumptions, conditional sampling, lower intrinsic dimension, . . .

Conclusion

Link to paper:

- Extended convergence theory for score-based generative models to more realistic data distributions
- Relaxed strong to weak log-concavity assumption, allowing for multi-modal distributions such as GMMs
- Remarkably, the asymptotics remain the same under the weaker assumption
- Result applies to general drift and diffusion functions f and g
- Error bound can be translated to concrete heuristics for the choice of time horizon T, step size h and score-matching error $\mathcal{E}$

Outlook: Reverse SDE, vector-valued f and g, even weaker assumptions, conditional sampling, lower intrinsic dimension, ...

What do you think? What is unclear? What could be investigated more?

References

- Conforti, G. (2024). "Weak semiconvexity estimates for Schrödinger potentials and logarithmic Sobolev inequality for Schrödinger bridges". In: *Probability Theory and Related Fields* 189.3, pp. 1045–1071.
- Gao, X. and Zhu, L. (2024). "Convergence analysis for general probability flow odes of diffusion models in wasserstein distances". In: *arXiv preprint arXiv:2401.17958*.
- Gao, X., Nguyen, H. M., and Zhu, L. (2025). "Wasserstein convergence guarantees for a general class of score-based generative models". In: *Journal of machine learning research* 26.43, pp. 1–54.
- Gentiloni-Silveri, M. and Ocello, A. (2025). "Beyond log-concavity and score regularity: Improved convergence bounds for score-based generative models in w2-distance". In: *arXiv preprint arXiv:2501.02298*.
- Wibisono, A. and Yang, K. (2022). "Convergence in KL divergence of the inexact Langevin algorithm with application to score-based generative models". In: *arXiv:2211.01512*.