Optimality of score based generative models through score regularity

4 December 2025

CREST-ENSAE

joint work with Eddie Aamari and Clément Levrard

Generative modeling

Setup:
We have access to a data $X^{(1)}, ..., X^{(n)} \sim p \in \mathcal{P}(\mathbb{R}^d)$ iid.

Goal:
Find an algorithm $\mathcal{A}_{X^{(1)},...,X^{(n)}} : \mathbb{R}^d \to \mathbb{R}^d$ such for an independent seed $Z \sim \mathcal{N}(0, \mathrm{Id})$, the new point $\mathcal{A}_{X^{(1)},....,X^{(n)}}(Z) \in \mathbb{R}^d$ "resemble" the dataset.

We aim at finding $\mathcal{A}_{X^{(1)},....,X^{(n)}}$ such that

$$(\mathcal{A}_{X^{(1)},....,X^{(n)}})_{\#}\gamma_d = p,$$

for γ_d the d-dimensional Gaussian measure.

Generative modeling

Setup:
We have access to a data $X^{(1)}, ..., X^{(n)} \sim p \in \mathcal{P}(\mathbb{R}^d)$ iid.

Goal:
Find an algorithm $\mathcal{A}_{X^{(1)},...,X^{(n)}} : \mathbb{R}^d \to \mathbb{R}^d$ such for an independent seed $Z \sim \mathcal{N}(0, \mathrm{Id})$, the new point $\mathcal{A}_{X^{(1)},....,X^{(n)}}(Z) \in \mathbb{R}^d$ "resemble" the dataset.

We aim at finding $\mathcal{A}_{X^{(1)},....,X^{(n)}}$ such that

$$(\mathcal{A}_{X^{(1)},....,X^{(n)}})_{\#}\gamma_d = p,$$

for γ_d the d-dimensional Gaussian measure.

Generative modeling

Setup:
We have access to a data $X^{(1)}, ..., X^{(n)} \sim p \in \mathcal{P}(\mathbb{R}^d)$ iid.

Goal:
Find an algorithm $\mathcal{A}_{X^{(1)},...,X^{(n)}} : \mathbb{R}^d \to \mathbb{R}^d$ such for an independent seed $Z \sim \mathcal{N}(0, \mathrm{Id})$, the new point $\mathcal{A}_{X^{(1)},....,X^{(n)}}(Z) \in \mathbb{R}^d$ "resemble" the dataset.

We aim at finding $\mathcal{A}_{X^{(1)},....,X^{(n)}}$ such that

$$(\mathcal{A}_{X^{(1)},....,X^{(n)}})_{\#}\gamma_d = p,$$

for γ_d the d-dimensional Gaussian measure.

Sampling with Score-Based Generative models

Forward Process (OU):

$$d\vec{X}_t = -\vec{X}_t dt + \sqrt{2} dB_t, \quad \vec{X}_0 \sim p$$

$$\mathrm{law}(\vec{X}_t) = \mathrm{law}\Big(e^{-t}\vec{X}_0 + \sqrt{1-e^{-2t}}\,Y\Big), \quad \vec{X}_0 \sim p \text{ and } Y \sim \mathcal{N}(0, \mathrm{Id}) \text{ ind.}$$

Fix $T > 0, \quad b_t \geq 0$ and write $p_t := \mathrm{law}(\vec{X}_t), \quad s(t,x) := \nabla \log p_t(x)$

Backward Process:

$$dX_t = \Big(X_t + (1+b_t^2)\cdot s(T-t, X_t)\Big)dt + \sqrt{2} b_t dB_t, \quad X_0 \sim p_T$$

$$\mathrm{law}(X_{T-t}) = \mathrm{law}(\vec{X}_t)$$

$$\mathrm{law}(X_T) = p$$

Sampling with Score-Based Generative models

Forward Process (OU):

$$d\overrightarrow{X}_t = -\overrightarrow{X}_t dt + \sqrt{2} dB_t, \quad \overrightarrow{X}_0 \sim p$$

$$\mathrm{law}\big(\overrightarrow{X}_t\big) = \mathrm{law}\Big(e^{-t}\overrightarrow{X}_0 + \sqrt{1-e^{-2t}}\,Y\Big), \quad \overrightarrow{X}_0 \sim p \text{ and } Y \sim \mathcal{N}(0, \mathrm{Id}) \text{ ind.}$$

Fix $T > 0$, $b_t \geq 0$ and write $p_t := \mathrm{law}(\overrightarrow{X}_t)$, $s(t,x) := \nabla \log p_t(x)$

Backward Process:

$$dX_t = \Big(X_t + (1+b_t^2) \cdot s(T-t, X_t)\Big)dt + \sqrt{2} b_t dB_t, \quad X_0 \sim p_T$$

$$\mathrm{law}(X_{T-t}) = \mathrm{law}(\overrightarrow{X}_t)$$

$$\mathrm{law}(X_T) = p$$

Sampling with Score-Based Generative models

Forward Process (OU):

$$d\vec{X}_t = -\vec{X}_t dt + \sqrt{2} dB_t, \quad \vec{X}_0 \sim p$$

$$\mathrm{law}(\vec{X}_t) = \mathrm{law}\Big(e^{-t}\vec{X}_0 + \sqrt{1-e^{-2t}}\,Y\Big), \quad \vec{X}_0 \sim p \text{ and } Y \sim \mathcal{N}(0, \mathrm{Id}) \text{ ind.}$$

$$\text{Fix } T > 0, \quad b_t \geq 0 \qquad \text{and write} \qquad p_t := \mathrm{law}(\vec{X}_t), \quad s(t,x) := \nabla \log p_t(x)$$

Backward Process:

$$dX_t = \Big(X_t + (1 + b_t^2) \cdot s(T-t, X_t)\Big) dt + \sqrt{2} b_t dB_t, \quad X_0 \sim p_T$$

$$\mathrm{law}(X_{T-t}) = \mathrm{law}(\vec{X}_t)$$

$$\mathrm{law}(X_T) = p$$

Sampling with Score-Based Generative models

Forward Process (OU):

$$d\overrightarrow{X}_t = -\overrightarrow{X}_t dt + \sqrt{2} dB_t, \quad \overrightarrow{X}_0 \sim p$$

$$\mathrm{law}\big(\overrightarrow{X}_t\big) = \mathrm{law}\Big(e^{-t}\overrightarrow{X}_0 + \sqrt{1-e^{-2t}}\,Y\Big), \quad \overrightarrow{X}_0 \sim p \text{ and } Y \sim \mathcal{N}(0, \mathrm{Id}) \text{ ind.}$$

$$\text{Fix } T > 0, \quad b_t \geq 0 \qquad \text{and write} \qquad p_t := \mathrm{law}(\overrightarrow{X}_t), \quad s(t,x) := \nabla \log p_t(x)$$

Backward Process:

$$dX_t = \Big(X_t + (1+b_t^2)\cdot s(T-t, X_t)\Big)dt + \sqrt{2} b_t dB_t, \quad X_0 \sim p_T$$

$$\mathrm{law}(X_{T-t}) = \mathrm{law}(\overrightarrow{X}_t)$$

$$\mathrm{law}(X_T) = p$$

Sampling with Score-Based Generative models

Forward Process (OU):

$$d\vec{X}_t = -\vec{X}_t dt + \sqrt{2} dB_t, \quad \vec{X}_0 \sim p$$

$$\mathrm{law}(\vec{X}_t) = \mathrm{law}\Big(e^{-t}\vec{X}_0 + \sqrt{1-e^{-2t}}Y\Big), \quad \vec{X}_0 \sim p \text{ and } Y \sim \mathcal{N}(0, \mathrm{Id}) \text{ ind.}$$

$$\text{Fix } T > 0, \quad b_t \geq 0 \qquad \text{and write} \qquad p_t := \mathrm{law}(\vec{X}_t), \quad s(t,x) := \nabla \log p_t(x)$$

Backward Process:

$$dX_t = \Big(X_t + (1 + b_t^2) \cdot s(T-t, X_t)\Big) dt + \sqrt{2} b_t dB_t, \quad X_0 \sim p_T$$

$$\mathrm{law}(X_{T-t}) = \mathrm{law}(\vec{X}_t)$$

$$\mathrm{law}(X_T) = p$$

Sampling with Score-Based Generative models

Forward Process (OU):

$$d\vec{X}_t = -\vec{X}_t dt + \sqrt{2} dB_t, \quad \vec{X}_0 \sim p$$

$$\mathrm{law}(\vec{X}_t) = \mathrm{law}\Big(e^{-t}\vec{X}_0 + \sqrt{1-e^{-2t}}\,Y\Big), \quad \vec{X}_0 \sim p \text{ and } Y \sim \mathcal{N}(0, \mathrm{Id}) \text{ ind.}$$

$$\text{Fix } T > 0, \quad b_t \geq 0 \qquad \text{and write} \qquad p_t := \mathrm{law}(\vec{X}_t), \quad s(t,x) := \nabla \log p_t(x)$$

Backward Process:

$$dX_t = \Big(X_t + (1+b_t^2)\cdot s(T-t, X_t)\Big)dt + \sqrt{2} b_t dB_t, \quad X_0 \sim p_T$$

$$\mathrm{law}(X_{T-t}) = \mathrm{law}(\vec{X}_t)$$

$$\mathrm{law}(X_T) = p$$

Forward Process (OU):

$$d\overrightarrow{X}_t = -\overrightarrow{X}_t dt + \sqrt{2} dB_t, \quad \overrightarrow{X}_0 \sim p$$

$$\text{Fix } T > 0, \quad b_t \geq 0 \qquad \text{and write} \qquad p_t := \text{law}(\overrightarrow{X}_t), \quad s(t, x) := \nabla \log p_t(x)$$

Backward Process:

$$dX_t = \Big(X_t + (1 + b_t^2) \cdot s(T - t, X_t) \Big) dt + \sqrt{2} b_t dB_t, \quad X_0 \sim p_T$$

$$\text{law}(X_{T-t}) = \text{law}(\overrightarrow{X}_t)$$

In practice:

$$d\hat{X}_t = \Big(\hat{X}_t + (1 + b_t^2) \cdot \hat{s}(T - t, \hat{X}_t) \Big) dt + \sqrt{2} b_t dB_t, \quad \hat{X}_0 \sim \mathcal{N}(0, \text{Id})$$

with $\hat{s}$ an estimator of s.

Score: definition & how to compute it

Score function:

$$s_t(x) = \nabla_x \log p_t(x),$$

with p_t the law of

$$e^{-t}\overrightarrow{X}_0 + \sqrt{1-e^{-2t}}\,Y,$$

$\overrightarrow{X}_0 \sim p$ and $Y \sim \mathcal{N}(0, \mathrm{Id})$ independent.

$$s_t \in \underset{\phi_t \in L^2(\mathbb{R}^d, \mathbb{R}^d)}{\arg\min}\ \mathbb{E}_{\substack{\overrightarrow{X}_0 \sim p \\ Y \sim \mathcal{N}(0,\mathrm{Id})}} \left[\| \phi_t(e^{-t}\overrightarrow{X}_0 + \sqrt{1-e^{-2t}}\,Y) + \frac{1}{\sqrt{1-e^{-2t}}} Y \|^2 \right].$$

How to approximate it in practice (Denoising score matching):

Choose a neural network class $\mathcal{S}$ and compute

$$\hat{s}_t \in \underset{\phi_t \in \mathcal{S}}{\arg\min}\ \frac{1}{n}\sum_{i=1}^{n} \| \phi_t(e^{-t}X^{(i)} + \sqrt{1-e^{-2t}}\,Y^{(i)}) + \frac{1}{\sqrt{1-e^{-2t}}} Y^{(i)} \|^2,$$

with $X^{(i)} \sim p$ and $Y^{(i)} \sim \mathcal{N}(0, \mathrm{Id})$.

Score: definition & how to compute it

Score function:

$$s_t(x) = \nabla_x \log p_t(x),$$

with p_t the law of

$$e^{-t}\overrightarrow{X}_0 + \sqrt{1-e^{-2t}}\,Y,$$

$\overrightarrow{X}_0 \sim p$ and $Y \sim \mathcal{N}(0, \mathrm{Id})$ independent.

$$s_t \in \underset{\phi_t \in L^2(\mathbb{R}^d, \mathbb{R}^d)}{\arg\min} \; \mathbb{E}_{\substack{\overrightarrow{X}_0 \sim p \\ Y \sim \mathcal{N}(0,\mathrm{Id})}} \left[\left\| \phi_t(e^{-t}\overrightarrow{X}_0 + \sqrt{1-e^{-2t}}\,Y) + \frac{1}{\sqrt{1-e^{-2t}}} Y \right\|^2 \right].$$

How to approximate it in practice (Denoising score matching):

Choose a neural network class $\mathcal{S}$ and compute

$$\hat{s}_t \in \underset{\phi_t \in \mathcal{S}}{\arg\min} \; \frac{1}{n} \sum_{i=1}^{n} \left\| \phi_t(e^{-t}X^{(i)} + \sqrt{1-e^{-2t}}\,Y^{(i)}) + \frac{1}{\sqrt{1-e^{-2t}}} Y^{(i)} \right\|^2,$$

with $X^{(i)} \sim p$ and $Y^{(i)} \sim \mathcal{N}(0, \mathrm{Id})$.

Sampling with Score-Based Generative models

$$\text{Fix } T \approx \log(n), b_t \geq 0$$

Backward Process:

$$dX_t = \Big(X_t + (1+b_t^2)\cdot s(T-t, X_t)\Big)dt + \sqrt{2}b_t dB_t, \quad X_0 \sim p_T$$

In practice:

$$d\hat{X}_t = \Big(\hat{X}_t + (1+b_t^2)\cdot \hat{s}(T-t, \hat{X}_t)\Big)dt + \sqrt{2}b_t dB_t, \quad \hat{X}_0 \sim \mathcal{N}(0, \mathrm{Id})$$

with

$$\hat{s}_t \in \underset{\phi_t \in \mathcal{S}}{\arg\min}\ \frac{1}{n}\sum_{i=1}^{n} \| \phi_t(e^{-t}X^{(i)} + \sqrt{1-e^{-2t}}Y^{(i)}) + \frac{1}{\sqrt{1-e^{-2t}}} Y^{(i)} \|^2, \quad X^{(i)} \sim p, Y^{(i)} \sim \mathcal{N}(0, \mathrm{Id}).$$

Our estimator of p is

$$\hat{p} = \mathrm{law}(\hat{X}_T).$$

Sampling with Score-Based Generative models

$$\text{Fix } T \approx \log(n), b_t \geq 0$$

Backward Process:

$$dX_t = \Big(X_t + (1 + b_t^2) \cdot s(T - t, X_t)\Big)dt + \sqrt{2} b_t dB_t, \quad X_0 \sim p_T$$

In practice:

$$d\hat{X}_t = \Big(\hat{X}_t + (1 + b_t^2) \cdot \hat{s}(T - t, \hat{X}_t)\Big)dt + \sqrt{2} b_t dB_t, \quad \hat{X}_0 \sim \mathcal{N}(0, \mathrm{Id})$$

with

$$\hat{s}_t \in \underset{\phi_t \in \mathcal{S}}{\arg\min} \ \frac{1}{n} \sum_{i=1}^{n} \| \phi_t(e^{-t} X^{(i)} + \sqrt{1 - e^{-2t}} Y^{(i)}) + \frac{1}{\sqrt{1 - e^{-2t}}} Y^{(i)} \|^2, \quad X^{(i)} \sim p, Y^{(i)} \sim \mathcal{N}(0, \mathrm{Id}).$$

Our estimator of p is

$$\hat{p} = \mathrm{law}(\hat{X}_T).$$

Minimax Rates for density estimation

Question:

$$\mathbb{E}_{X^{(1)},\ldots,X^{(n)}\sim p}\left[W_1\big(p,\hat{p}\big)\right] \leq ?$$

Setup: Nonparametric density estimation

- Observe $X^{(1)},\ldots,X^{(n)} \sim p$ iid on $\mathbb{R}^d$
- Assume there exist $\beta, K > 0$

$$p \in \mathcal{H}_K^\beta = \Big\{ f \in C^{\lfloor\beta\rfloor}([0,1]^d,\mathbb{R}) \;\Big|\; \sum_{i=1}^{\lfloor\beta\rfloor} \|\nabla^i f\|_\infty + \|\nabla^{\lfloor\beta\rfloor} f\|_{\beta-\lfloor\beta\rfloor} \leq K \Big\}.$$

Minimax Rate:

$$\inf_{p_n} \sup_{p\in\mathcal{H}_K^\beta} \mathbb{E}_{X^{(1)},\ldots,X^{(n)}}\left[W_1\big(p,p_n\big)\right] \underset{\mathrm{polylog}(n)}{\approx} n^{-\frac{\beta+1}{2\beta+d}},$$

with $\hat{p}_n$ an estimator having only access to the data $X^{(1)},\ldots,X^{(n)}$.

Minimax Rates for density estimation

Question:

$$\mathbb{E}_{X^{(1)},\dots,X^{(n)}\sim p}\left[W_1\big(p,\hat{p}\big)\right] \leq ?$$

Setup: Nonparametric density estimation

- Observe $X^{(1)},\dots,X^{(n)} \sim p$ iid on $\mathbb{R}^d$
- Assume there exist $\beta, K > 0$

$$p \in \mathcal{H}_K^\beta = \Big\{ f \in C^{\lfloor\beta\rfloor}([0,1]^d,\mathbb{R}) \;\Big|\; \sum_{i=1}^{\lfloor\beta\rfloor} \|\nabla^i f\|_\infty + \|\nabla^{\lfloor\beta\rfloor} f\|_{\beta-\lfloor\beta\rfloor} \leq K \Big\}.$$

Minimax Rate:

$$\inf_{p_n} \sup_{p\in\mathcal{H}_K^\beta} \mathbb{E}_{X^{(1)},\dots,X^{(n)}}\left[W_1\big(p,p_n\big)\right] \underset{\mathrm{polylog}(n)}{\approx} n^{-\frac{\beta+1}{2\beta+d}},$$

with $\hat{p}_n$ an estimator having only access to the data $X^{(1)},\dots,X^{(n)}$.

Minimax Rates for density estimation

Question:

$$\mathbb{E}_{X^{(1)},\ldots,X^{(n)}\sim p}\left[W_1\big(p,\hat{p}\big)\right] \leq \,?$$

Setup: Nonparametric density estimation

- Observe $X^{(1)},\ldots,X^{(n)} \sim p$ iid on $\mathbb{R}^d$
- Assume there exist $\beta, K > 0$

$$p \in \mathcal{H}_K^\beta = \Big\{ f \in C^{\lfloor\beta\rfloor}([0,1]^d,\mathbb{R}) \;\Big|\; \sum_{i=1}^{\lfloor\beta\rfloor} \|\nabla^i f\|_\infty + \|\nabla^{\lfloor\beta\rfloor} f\|_{\beta-\lfloor\beta\rfloor} \leq K \Big\}.$$

Minimax Rate:

$$\inf_{p_n}\sup_{p\in\mathcal{H}_K^\beta} \mathbb{E}_{X^{(1)},\ldots,X^{(n)}}\left[W_1\big(p,p_n\big)\right] \underset{\mathrm{polylog}(n)}{\approx} n^{-\frac{\beta+1}{2\beta+d}},$$

with $\hat{p}_n$ an estimator having only access to the data $X^{(1)},\ldots,X^{(n)}$.

Prior Works on Minimax Rates for SGMs

Several works already establish minimax rates:

- Oko et al. (2023): diffusion model for densities in $\mathcal{H}^\beta([0,1]^d)$ bounded below in W_1 and TV via NN scores.
- Zhang et al. (2024): diffusion model for densities in $\mathcal{H}^\beta(\mathbb{R}^d)$ subgaussian with $\beta \leq 2$ in TV, via kernel scores.
- Cai and Li (2025): ODE model for densities in $\mathcal{H}^\beta(\mathbb{R}^d)$ subgaussian with $\beta \leq 2$ in TV, via kernel scores.

Limits:

- Each proof is tailored to a specific model class and sampling method.
- Long and technically involved proofs.

Objectives

- **Unification.** Provide a unified proof that handles both stochastic (SDE) and deterministic (ODE) sampling schemes.
- **Flexible model.** The framework encompasses both compactly supported and unbounded distributions.
- **Short proof.** No need to assume regularity of the score: it follows from the regularity of p. This yields a significantly more concise argument, especially in the approximation theory.

Objectives

- **Unification.** Provide a unified proof that handles both stochastic (SDE) and deterministic (ODE) sampling schemes.
- **Flexible model.** The framework encompasses both compactly supported and unbounded distributions.
- **Short proof.** No need to assume regularity of the score: it follows from the regularity of p. This yields a significantly more concise argument, especially in the approximation theory.

Objectives

- **Unification.** Provide a unified proof that handles both stochastic (SDE) and deterministic (ODE) sampling schemes.
- **Flexible model.** The framework encompasses both compactly supported and unbounded distributions.
- **Short proof.** No need to assume regularity of the score: it follows from the regularity of p. This yields a significantly more concise argument, especially in the approximation theory.

Key step I: Stability of SDEs

$$\text{True backward process:} \quad dX_t = \Big(X_t + (1+b_t^2)\cdot s(T-t, X_t)\Big)dt + \sqrt{2}b_t dB_t, \quad X_0 \sim p_T$$

$$\text{In practice:} \quad d\hat{X}_t = \Big(\hat{X}_t + (1+b_t^2)\cdot \hat{s}(T-t, \hat{X}_t)\Big)dt + \sqrt{2}b_t dB_t, \quad \hat{X}_0 \sim \mathcal{N}(0, \mathrm{Id}).$$

Write

$$p_t = \mathrm{law}(X_{T-t}), \quad p = \mathrm{law}(X_T), \quad \hat{p} = \mathrm{law}(\hat{X}_t).$$

Stability bound:

$$\begin{aligned} \mathrm{W}_1(p, \hat{p}) &\le d(p_T, \gamma_d) + \int_0^T \int_{\mathbb{R}^d} e^{\int_0^t \lambda_{\max}(\nabla \hat{s}(u,\cdot)+\mathrm{Id})\, du} \|\hat{s}(t,x) - s(t,x)\|\, dp_t(x)dt \\ &\le \epsilon + C \int_0^T \int_{\mathbb{R}^d} \|\hat{s}(t,x) - s(t,x)\|\, dp_t(x)dt \end{aligned}$$

Key step I: Stability of SDEs

True backward process: $dX_t = \Big(X_t + (1+b_t^2)\cdot s(T-t, X_t)\Big)dt + \sqrt{2}b_t dB_t, \quad X_0 \sim p_T$

In practice: $d\hat{X}_t = \Big(\hat{X}_t + (1+b_t^2)\cdot \hat{s}(T-t, \hat{X}_t)\Big)dt + \sqrt{2}b_t dB_t, \quad \hat{X}_0 \sim \mathcal{N}(0, \mathrm{Id}).$

Write

$$p_t = \mathrm{law}(X_{T-t}), \quad p = \mathrm{law}(X_T), \quad \hat{p} = \mathrm{law}(\hat{X}_t).$$

Stability bound:

$$\mathrm{W}_1(p, \hat{p}) \le d(p_T, \gamma_d) + \int_0^T \int_{\mathbb{R}^d} e^{\int_0^t \lambda_{\max}(\nabla \hat{s}(u,\cdot)+\mathrm{Id})\, du} \|\hat{s}(t,x) - s(t,x)\|\, dp_t(x)dt$$

$$\le \epsilon + C \int_0^T \int_{\mathbb{R}^d} \|\hat{s}(t,x) - s(t,x)\|\, dp_t(x)dt$$

Key step I: Stability of SDEs

True backward process: $dX_t = \Big(X_t + (1+b_t^2)\cdot s(T-t, X_t)\Big)dt + \sqrt{2}b_t dB_t, \quad X_0 \sim p_T$

In practice: $d\hat{X}_t = \Big(\hat{X}_t + (1+b_t^2)\cdot \hat{s}(T-t, \hat{X}_t)\Big)dt + \sqrt{2}b_t dB_t, \quad \hat{X}_0 \sim \mathcal{N}(0, \mathrm{Id}).$

Write

$$p_t = \mathrm{law}(X_{T-t}), \quad p = \mathrm{law}(X_T), \quad \hat{p} = \mathrm{law}(\hat{X}_t).$$

Stability bound:

$$\begin{aligned}
\mathrm{W}_1(p,\hat{p}) &\leq d(p_T, \gamma_d) + \int_0^T \int_{\mathbb{R}^d} e^{\int_0^t \lambda_{\max}(\nabla \hat{s}(u,\cdot)+\mathrm{Id})\,du} \|\hat{s}(t,x) - s(t,x)\|\, dp_t(x)dt \\
&\leq \epsilon + C\int_0^T \int_{\mathbb{R}^d} \|\hat{s}(t,x) - s(t,x)\|\, dp_t(x)dt
\end{aligned}$$

Key step II: Score approximation

$$\hat{s} \in \arg\min_{\phi \in \mathcal{S}} \ \frac{1}{n} \sum_{i=1}^{n} \int_0^1 \| \, \phi_t(e^{-t} X^{(i)} + \sqrt{1-e^{-2t}} Y^{(i)}) + \frac{1}{\sqrt{1-e^{-2t}}} Y^{(i)} \, \|^2 dt, .$$

with $X^{(i)} \sim p$ and $Y^{(i)} \sim \mathcal{N}(0, \mathrm{Id})$. Aims to approximate

$$s = \arg\min_{\phi \in L^2} \iiint \|\phi(t, e^{-t}x + \sqrt{1-e^{-2t}} y) + \frac{y}{\sqrt{1-e^{-2t}}}\|^2 dp(x) d\gamma_d(y) dt$$

$$= \arg\min_{\phi \in L^2} \iint \|\phi(t,x) - s(t,x)\|^2 dp_t(x) dt$$

Bias–variance trade-off:

$$\mathbb{E}\left[\iint \|\hat{s}(t,x) - s(t,x)\|^2 dp_t(x) dt\right] \lesssim \inf_{\phi \in \mathcal{S}} \iint \|\phi(t,x) - s(t,x)\|^2 dp_t(x) dt + \frac{1}{n} \log \mathcal{N}(\mathcal{S}, \|\cdot\|_\infty, n^{-1})$$

Key step II: Score approximation

$$\hat{s} \in \operatorname*{arg\,min}_{\phi \in \mathcal{S}} \ \frac{1}{n} \sum_{i=1}^{n} \int_0^1 \| \, \phi_t(e^{-t} X^{(i)} + \sqrt{1-e^{-2t}} Y^{(i)}) + \frac{1}{\sqrt{1-e^{-2t}}} Y^{(i)} \, \|^2 dt, .$$

with $X^{(i)} \sim p$ and $Y^{(i)} \sim \mathcal{N}(0, \mathrm{Id})$. Aims to approximate

$$s = \operatorname*{arg\,min}_{\phi \in L^2} \iiint \|\phi(t, e^{-t}x + \sqrt{1-e^{-2t}} y) + \frac{y}{\sqrt{1-e^{-2t}}}\|^2 dp(x) d\gamma_d(y) dt$$

$$= \operatorname*{arg\,min}_{\phi \in L^2} \iint \|\phi(t,x) - s(t,x)\|^2 dp_t(x) dt$$

Bias–variance trade-off:

$$\mathbb{E}\left[\iint \|\hat{s}(t,x) - s(t,x)\|^2 dp_t(x) dt\right] \lesssim \inf_{\phi \in \mathcal{S}} \iint \|\phi(t,x) - s(t,x)\|^2 dp_t(x) dt + \frac{1}{n} \log \mathcal{N}(\mathcal{S}, \| \cdot \|_\infty, n^{-1})$$

Key step II: Score approximation

$$\hat{s} \in \operatorname*{arg\,min}_{\phi \in \mathcal{S}} \ \frac{1}{n} \sum_{i=1}^{n} \int_0^1 \| \phi_t(e^{-t} X^{(i)} + \sqrt{1-e^{-2t}} Y^{(i)}) + \frac{1}{\sqrt{1-e^{-2t}}} Y^{(i)} \|^2 dt, .$$

with $X^{(i)} \sim p$ and $Y^{(i)} \sim \mathcal{N}(0, \mathrm{Id})$. Aims to approximate

$$\begin{aligned} s &= \operatorname*{arg\,min}_{\phi \in L^2} \iiint \|\phi(t, e^{-t}x + \sqrt{1-e^{-2t}} y) + \frac{y}{\sqrt{1-e^{-2t}}}\|^2 dp(x) d\gamma_d(y) dt \\ &= \operatorname*{arg\,min}_{\phi \in L^2} \iint \|\phi(t,x) - s(t,x)\|^2 dp_t(x) dt \end{aligned}$$

Bias–variance trade-off:

$$\mathbb{E}\left[\iint \|\hat{s}(t,x) - s(t,x)\|^2 dp_t(x) dt\right] \lesssim \inf_{\phi \in \mathcal{S}} \iint \|\phi(t,x) - s(t,x)\|^2 dp_t(x) dt + \frac{1}{n} \log \mathcal{N}(\mathcal{S}, \|\cdot\|_\infty, n^{-1})$$

Key step II: Score approximation

$$\hat{s} \in \arg\min_{\phi \in \mathcal{S}} \; \frac{1}{n} \sum_{i=1}^{n} \int_0^1 \| \, \phi_t(e^{-t} X^{(i)} + \sqrt{1 - e^{-2t}} Y^{(i)}) + \frac{1}{\sqrt{1 - e^{-2t}}} Y^{(i)} \, \|^2 dt, .$$

with $X^{(i)} \sim p$ and $Y^{(i)} \sim \mathcal{N}(0, \mathrm{Id})$. Aims to approximate

$$\begin{aligned} s &= \arg\min_{\phi \in L^2} \iiint \|\phi(t, e^{-t} x + \sqrt{1 - e^{-2t}} y) + \frac{y}{\sqrt{1 - e^{-2t}}}\|^2 dp(x) d\gamma_d(y) dt \\ &= \arg\min_{\phi \in L^2} \iint \|\phi(t, x) - s(t, x)\|^2 dp_t(x) dt \end{aligned}$$

Bias–variance trade-off:

$$\begin{aligned} \mathbb{E}\left[\iint \|\hat{s}(t, x) - s(t, x)\|^2 dp_t(x) dt\right] &\lesssim \inf_{\phi \in \mathcal{S}} \iint \|\phi(t, x) - s(t, x)\|^2 dp_t(x) dt \\ &\quad + \frac{1}{n} \log \mathcal{N}(\mathcal{S}, \| \cdot \|_\infty, n^{-1}) \end{aligned}$$

Works using Lipschitz-type assumptions on the score

Diffusion (SDE)

- Lee et al. (2023)
- Chen et al. (2022, 2023a)
- Wibisono et al. (2024)

Probability flow (ODE)

- Chen et al. (2023b)
- Benton et al. (2023)
- Huang et al. (2024)

Assumptions on the score $s_t(x) = \nabla \log p_t(x)$

$$\textbf{Global Lipschitz:} \quad \sup_{x \in \mathbb{R}^d} \|\nabla s_t(x)\|_{op} \leq f(t)$$

$$\textbf{One-sided Lipschitz:} \quad \sup_{x \in \mathbb{R}^d} \lambda_{\max}(\nabla s_t(x)) \leq f(t),$$

Lipschitz regularity of the score

Assumptions 1: The probability density p is of the form $p(x) = \exp(-u(x) + a(x))$ with:

1. $u \in C^2$ on the interior of the support of p and there exists $\alpha > 0$ such that $\nabla^2 u \succeq \alpha \mathrm{Id}$,
2. there exists $\beta > 0$ and $K > 0$ such that $a \in \mathcal{H}_K^\beta$.

Lipschitz regularity of the score

Assumptions 1: The probability density p is of the form $p(x) = \exp(-u(x) + a(x))$ with:

1. $u \in C^2$ on the interior of the support of p and there exists $\alpha > 0$ such that $\nabla^2 u \succeq \alpha \mathrm{Id}$,
2. there exists $\beta > 0$ and $K > 0$ such that $a \in \mathcal{H}_K^\beta$.

Lipschitz regularity of the score

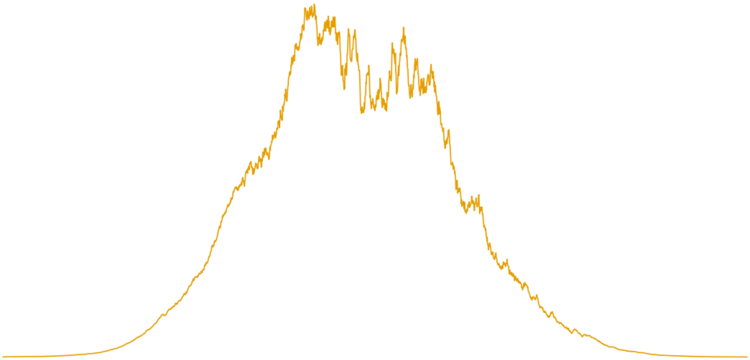
Assumptions 1: The probability density p is of the form $p(x) = \exp(-u(x) + a(x))$ with:

1. $u \in C^2$ on the interior of the support of p and there exists $\alpha > 0$ such that $\nabla^2 u \succeq \alpha \mathrm{Id}$,
2. there exists $\beta > 0$ and $K > 0$ such that $a \in \mathcal{H}_K^\beta$.

Lipschitz regularity of the score

Assumptions 1: The probability density p is of the form $p(x) = \exp(-u(x) + a(x))$ with:

1. $u \in C^2$ on the interior of the support of p and there exists $\alpha > 0$ such that $\nabla^2 u \succeq \alpha \mathrm{Id}$,
2. there exists $\beta > 0$ and $K > 0$ such that $a \in \mathcal{H}_K^\beta$.



Lipschitz regularity of the score

Assumptions 1: The probability density p is of the form $p(x) = \exp(-u(x) + a(x))$ with:

1. $u \in C^2$ on the interior of the support of p and there exists $\alpha > 0$ such that $\nabla^2 u \succeq \alpha \mathrm{Id}$,
2. there exists $\beta > 0$ and $K > 0$ such that $a \in \mathcal{H}_K^\beta$.

Theorem: *Let $p : \mathbb{R}^d \to \mathbb{R}$ a probability density satisfying Assumption 1. Then, the largest eigenvalue of the Jacobian of the score function satisfies*

$$\sup_{x \in \mathbb{R}^d} \lambda_{\max}\big(\nabla s(t,x) + Id\big) \leq Ce^{-2t}(1 + t^{\frac{\beta \wedge 1}{2} - 1}).$$

In particular, we have

$$\int_0^\infty \sup_{x \in \mathbb{R}^d} \lambda_{\max}\big(\nabla s(t,x) + \mathrm{Id}\big) dt < \infty.$$

Lipschitz regularity of the score

Assumptions 1: The probability density p is of the form $p(x) = \exp(-u(x) + a(x))$ with:

1. $u \in C^2$ on the interior of the support of p and there exists $\alpha > 0$ such that $\nabla^2 u \succeq \alpha \mathrm{Id}$,
2. there exists $\beta > 0$ and $K > 0$ such that $a \in \mathcal{H}_K^\beta$.

Theorem: *Let $p : \mathbb{R}^d \to \mathbb{R}$ a probability density satisfying Assumption 1. Then, the largest eigenvalue of the Jacobian of the score function satisfies*

$$\sup_{x \in \mathbb{R}^d} \lambda_{\max}\big(\nabla s(t,x) + Id\big) \leq Ce^{-2t}(1 + t^{\frac{\beta \wedge 1}{2} - 1}).$$

In particular, we have

$$\int_0^\infty \sup_{x \in \mathbb{R}^d} \lambda_{\max}\big(\nabla s(t,x) + \mathrm{Id}\big) dt < \infty.$$

Higher order regularity of the score

Assumptions 2: The probability density p is of the form $p(x) = \exp(-u(x) + a(x))$ with:

1. $u \in C^2$ on the interior of the support of p and there exists $\alpha > 0$ such that $\nabla^2 u \succeq \alpha \mathrm{Id}$,
2. there exists $\beta > 0$ and $K > 0$ such that $a \in \mathcal{H}^{\beta}_{K}$.
3. The potential u satisfies a mild growth assumption.

Theorem: *Let $p : \mathbb{R}^d \to \mathbb{R}$ a probability density satisfying Assumption 2. Then, for $t, \epsilon \in (0,1)$ there exists a convex set A_t^ϵ such that the law of the Ornstein–Uhlenbeck process at time t satisfies $p_t(A_t^\epsilon) \geq 1 - \epsilon$ and for all $\gamma \geq 0$ we have*

$$\|s(t,\cdot)\|_{\mathcal{H}^{\gamma}(A_t^\epsilon)} \underset{\mathrm{polylog}(\epsilon^{-1})}{\lesssim} t^{-\frac{1}{2}\left((1+\gamma-\beta)\vee 0\right)}.$$

Higher order regularity of the score

Assumptions 2: The probability density p is of the form $p(x) = \exp(-u(x) + a(x))$ with:

1. $u \in C^2$ on the interior of the support of p and there exists $\alpha > 0$ such that $\nabla^2 u \succeq \alpha \mathrm{Id}$,
2. there exists $\beta > 0$ and $K > 0$ such that $a \in \mathcal{H}^\beta_K$.
3. The potential u satisfies a mild growth assumption.

Theorem: *Let $p : \mathbb{R}^d \to \mathbb{R}$ a probability density satisfying Assumption 2. Then, for $t, \epsilon \in (0,1)$ there exists a convex set A_t^ϵ such that the law of the Ornstein–Uhlenbeck process at time t satisfies $p_t(A_t^\epsilon) \geq 1 - \epsilon$ and for all $\gamma \geq 0$ we have*

$$\|s(t,\cdot)\|_{\mathcal{H}^\gamma(A_t^\epsilon)} \underset{\mathrm{polylog}(\epsilon^{-1})}{\lesssim} t^{-\frac{1}{2}\left((1+\gamma-\beta)\vee 0\right)}.$$

Score approximation with neural networks

Assumptions 2: The probability density p is of the form $p(x) = \exp(-u(x) + a(x))$ with:

1. $u \in C^2$ on the interior of the support of p and there exists $\alpha > 0$ such that $\nabla^2 u \succeq \alpha \mathrm{Id}$,
2. there exists $\beta > 0$ and $K > 0$ such that $a \in \mathcal{H}_K^\beta$.
3. The potential u satisfies a mild growth assumption.

Corollary: *Let $p : \mathbb{R}^d \to \mathbb{R}$ a probability density satisfying Assumption 2. Then, there exists a neural network class $\mathcal{S}$ such that*

$$\frac{1}{n} \log \mathcal{N}(\mathcal{S}, \|\cdot\|_\infty, n^{-1}) \underset{\mathrm{polylog}(n)}{\lesssim} n^{-\frac{2(\beta+1)}{2\beta+d}}$$

and

$$\inf_{\phi \in \mathcal{S}} \iint \|\phi(t,x) - s(t,x)\|^2 dp_t(x) dt \underset{\mathrm{polylog}(n)}{\lesssim} n^{-\frac{2(\beta+1)}{2\beta+d}}.$$

Minimax rates through score regularity

Step I: stability of SDEs

$$\mathrm{W}_1(p,\hat{p}) \le d(p_T,\gamma_d) + \int_0^T \int_{\mathbb{R}^d} e^{\int_0^t \lambda_{\max}(\nabla \hat{s}(u,\cdot)+\mathrm{Id})\,du} \|\hat{s}(t,x)-s(t,x)\|\,dp_t(x)dt$$

$$\le \epsilon + C\int_0^T \int_{\mathbb{R}^d} \|\hat{s}(t,x)-s(t,x)\|\,dp_t(x)dt$$

Step II: score approximation

$$\mathbb{E}\Big[\iint \|\hat{s}(t,x)-s(t,x)\|^2 dp_t(x)dt\Big] \lesssim \inf_{\phi\in\mathcal{S}} \iint \|\phi(t,x)-s(t,x)\|^2 dp_t(x)dt + \frac{1}{n}\log\mathcal{N}(\mathcal{S},\|\cdot\|_\infty,n^{-1})$$

$$\underset{\mathrm{polylog}(n)}{\lesssim} n^{-\frac{2(\beta+1)}{2\beta+d}}.$$

Minimax rates through score regularity

Step I: stability of SDEs

$$\mathrm{W}_1(p,\hat{p}) \leq d(p_T,\gamma_d) + \int_0^T \int_{\mathbb{R}^d} e^{\int_0^t \lambda_{\max}(\nabla \hat{s}(u,\cdot)+\mathrm{Id})\,du} \|\hat{s}(t,x)-s(t,x)\|\,dp_t(x)dt$$

$$\leq \epsilon + C \int_0^T \int_{\mathbb{R}^d} \|\hat{s}(t,x)-s(t,x)\|\,dp_t(x)dt$$

Step II: score approximation

$$\mathbb{E}\Big[\iint \|\hat{s}(t,x)-s(t,x)\|^2 dp_t(x)dt\Big] \lesssim \inf_{\phi\in\mathcal{S}} \iint \|\phi(t,x)-s(t,x)\|^2 dp_t(x)dt$$

$$+ \frac{1}{n}\log\mathcal{N}(\mathcal{S},\|\cdot\|_\infty,n^{-1})$$

$$\underset{\mathrm{polylog}(n)}{\lesssim} n^{-\frac{2(\beta+1)}{2\beta+d}}.$$

Minimax rates through score regularity

Step I: stability of SDEs

$$\mathrm{W}_1(p, \hat{p}) \leq d(p_T, \gamma_d) + \int_0^T \int_{\mathbb{R}^d} e^{\int_0^t \lambda_{\max}(\nabla \hat{s}(u,\cdot)+\mathrm{Id})\, du} \|\hat{s}(t,x) - s(t,x)\| \, dp_t(x)dt$$
$$\leq \epsilon + C \int_0^T \int_{\mathbb{R}^d} \|\hat{s}(t,x) - s(t,x)\| \, dp_t(x)dt$$

Step II: score approximation

$$\mathbb{E}\Big[\iint \|\hat{s}(t,x) - s(t,x)\|^2 dp_t(x)dt \Big] \lesssim \inf_{\phi \in \mathcal{S}} \iint \|\phi(t,x) - s(t,x)\|^2 dp_t(x)dt$$
$$+ \frac{1}{n} \log \mathcal{N}(\mathcal{S}, \|\cdot\|_\infty, n^{-1})$$
$$\underset{\mathrm{polylog}(n)}{\lesssim} \quad n^{-\frac{2(\beta+1)}{2\beta+d}}.$$

Minimax rates through score regularity

Step I: stability of SDEs

$$\begin{aligned}\mathrm{W}_1(p,\hat{p}) &\le d(p_T,\gamma_d) + \int_0^T \int_{\mathbb{R}^d} e^{\int_0^t \lambda_{\max}(\nabla \hat{s}(u,\cdot)+\mathrm{Id})\,du} \|\hat{s}(t,x) - s(t,x)\|\, dp_t(x)dt \\ &\le \epsilon + C \int_0^T \int_{\mathbb{R}^d} \|\hat{s}(t,x) - s(t,x)\|\, dp_t(x)dt\end{aligned}$$

Step II: score approximation

$$\begin{aligned}\mathbb{E}\Big[\iint \|\hat{s}(t,x) - s(t,x)\|^2 dp_t(x)dt\Big] &\lesssim \inf_{\phi\in\mathcal{S}} \iint \|\phi(t,x) - s(t,x)\|^2 dp_t(x)dt \\ &\quad + \frac{1}{n}\log \mathcal{N}(\mathcal{S}, \|\cdot\|_\infty, n^{-1}) \\ &\underset{\mathrm{polylog}(n)}{\lesssim} n^{-\frac{2(\beta+1)}{2\beta+d}}.\end{aligned}$$

Minimax rates through score regularity

Step I: stability of SDEs

$$\begin{aligned}\mathrm{W}_1(p,\hat{p}) &\leq d(p_T,\gamma_d) + \int_0^T \int_{\mathbb{R}^d} e^{\int_0^t \lambda_{\max}(\nabla \hat{s}(u,\cdot)+\mathrm{Id})\,du} \|\hat{s}(t,x)-s(t,x)\|\, dp_t(x)dt \\ &\leq \epsilon + C\int_0^T \int_{\mathbb{R}^d} \|\hat{s}(t,x)-s(t,x)\|\, dp_t(x)dt\end{aligned}$$

Step II: score approximation

$$\begin{aligned}\mathbb{E}\Big[\iint \|\hat{s}(t,x)-s(t,x)\|^2 dp_t(x)dt\Big] &\lesssim \inf_{\phi\in\mathcal{S}} \iint \|\phi(t,x)-s(t,x)\|^2 dp_t(x)dt \\ &\qquad + \frac{1}{n}\log\mathcal{N}(\mathcal{S},\|\cdot\|_\infty, n^{-1}) \\ &\underset{\mathrm{polylog}(n)}{\lesssim} n^{-\frac{2(\beta+1)}{2\beta+d}}.\end{aligned}$$

Minimax convergence rates for SGMs

Assumptions 2: The probability density p is of the form $p(x) = \exp(-u(x) + a(x))$ with:

1. $u \in C^2$ on the interior of the support of p and there exists $\alpha > 0$ such that $\nabla^2 u \succeq \alpha \mathrm{Id}$,
2. there exists $\beta > 0$ and $K > 0$ such that $a \in \mathcal{H}_K^\beta$.
3. The potential u satisfies a mild growth assumption.

Theorem: *Let $p : \mathbb{R}^d \to \mathbb{R}$ a probability density satisfying Assumption 2. Let $\hat{s}$ be the estimated score obtained by score matching over a well-chosen neural network class $\mathcal{S}$. Then for well-chosen $0 < \underline{T} < \overline{T} < \infty$, the distribution $\hat{p}$ of the early-stopped backward diffusion $\hat{X}_{\overline{T}-\underline{T}}$ satisfies*

$$\mathbb{E}_{X^{(1)},\dots,X^{(n)}\sim p}\left[\mathrm{W}_1(p,\hat{p})\right] \underset{\mathrm{polylog}(n)}{\lesssim} n^{-\frac{\beta+1}{2\beta+d}}.$$

Flow matching

The goal is to sample from $p \in \mathcal{P}(\mathbb{R}^d)$ by approximating the trajectories of a well chosen ODE:

$$dX_t = v_t(X_t)dt, \quad X_0 \sim \mathcal{N}(0, \mathrm{Id}),$$

satisfying $\mathrm{law}(X_1) = p$. Choose a coupling $\pi \in \mathcal{P}(\mathbb{R}^d \times \mathbb{R}^d)$ with $\pi_1 = p$ and $\pi_0 = \gamma_d$ and take the law of X_t to be equal to

$$p_t(x_t) = \int p_t(x_t|x_0, x_1)d\pi(x_0, x_1), \quad \text{with } p_t(x_t|x_0, x_1) \text{ the pdf of } \mathcal{N}(\mu_t(x_0, x_1), \sigma_t^2).$$

Property:

$$v_t(x_t) = \mathbb{E}\left[v_t(x_t|X_0, X_1)\middle|X_t = x_t\right],$$

with $v_t(x_t|x_0, x_1) = \frac{\sigma_t^{'}}{\sigma_t}(x_t - \mu_t(x_0, x_1)) + \mu_t^{'}(x_0, x_1)$.

Flow matching

The goal is to sample from $p \in \mathcal{P}(\mathbb{R}^d)$ by approximating the trajectories of a well chosen ODE:

$$dX_t = v_t(X_t)dt, \quad X_0 \sim \mathcal{N}(0, \mathrm{Id}),$$

satisfying $\mathrm{law}(X_1) = p$. Choose a coupling $\pi \in \mathcal{P}(\mathbb{R}^d \times \mathbb{R}^d)$ with $\pi_1 = p$ and $\pi_0 = \gamma_d$ and take the law of X_t to be equal to

$$p_t(x_t) = \int p_t(x_t|x_0, x_1)d\pi(x_0, x_1), \quad \text{with } p_t(x_t|x_0, x_1) \text{ the pdf of } \mathcal{N}(\mu_t(x_0, x_1), \sigma_t^2).$$

Property:

$$v_t(x_t) = \mathbb{E}\left[v_t(x_t|X_0, X_1)\middle|X_t = x_t\right],$$

with $v_t(x_t|x_0, x_1) = \frac{\sigma_t'}{\sigma_t}(x_t - \mu_t(x_0, x_1)) + \mu_t'(x_0, x_1)$.

Flow matching

The goal is to sample from $p \in \mathcal{P}(\mathbb{R}^d)$ by approximating the trajectories of a well chosen ODE:

$$dX_t = v_t(X_t)dt, \quad X_0 \sim \mathcal{N}(0, \mathrm{Id}),$$

satisfying law$(X_1) = p$. Choose a coupling $\pi \in \mathcal{P}(\mathbb{R}^d \times \mathbb{R}^d)$ with $\pi_1 = p$ and $\pi_0 = \gamma_d$ and take the law of X_t to be equal to

$$p_t(x_t) = \int p_t(x_t|x_0, x_1)d\pi(x_0, x_1), \quad \text{with } p_t(x_t|x_0, x_1) \text{ the pdf of } \mathcal{N}(\mu_t(x_0, x_1), \sigma_t^2).$$

Property:

$$v_t(x_t) = \mathbb{E}\left[v_t(x_t|X_0, X_1)\middle|X_t = x_t\right],$$

with $v_t(x_t|x_0, x_1) = \frac{\sigma_t^{'}}{\sigma_t}(x_t - \mu_t(x_0, x_1)) + \mu_t^{'}(x_0, x_1)$.

Diffusion Flow matching

For $\overrightarrow{X}_t$ the forward process in diffusion models

$$\mathrm{law}(\overrightarrow{X}_t) = \mathrm{law}\Big(e^{-t}\overrightarrow{X}_0 + \sqrt{1-e^{-2t}}\,Y\Big), \quad \overrightarrow{X}_0 \sim p \text{ and } Y \sim \mathcal{N}(0, \mathrm{Id}) \text{ ind.}$$

Rescaling, we can define X_t satisfying

$$\mathrm{law}(X_t) = \mathrm{law}\Big(t\overrightarrow{X}_0 + \sqrt{1-t^2}\,Y\Big), \quad \overrightarrow{X}_0 \sim p \text{ and } Y \sim \mathcal{N}(0, \mathrm{Id}) \text{ ind}$$

by taking

$$p_t(x_t) = \int p_t(x_t|x_0, x_1)d\pi(x_0, x_1), \quad \text{with } p_t(x_t|x_0, x_1) \text{ the pdf of } \mathcal{N}(\mu_t(x_0, x_1), \sigma_t^2).$$

for $\pi = p \otimes \gamma_d$ and $\mu_t((x_0, x_1)) = tx_1$, $\sigma_t = \sqrt{1-t^2}$. Then,

$$\begin{aligned} v_t(x_t) &= \mathbb{E}\left[v_t(x_t|X_0, X_1)\Big|X_t = x_t\right] \\ &= \frac{1}{t}\big(s(\log(t^{-1}), x_t) + x_t\big), \end{aligned}$$

with s the score function.

Diffusion Flow matching

For $\overrightarrow{X}_t$ the forward process in diffusion models

$$\mathrm{law}(\overrightarrow{X}_t) = \mathrm{law}\Big(e^{-t}\overrightarrow{X}_0 + \sqrt{1-e^{-2t}}Y\Big), \quad \overrightarrow{X}_0 \sim p \text{ and } Y \sim \mathcal{N}(0, \mathrm{Id}) \text{ ind.}$$

Rescaling, we can define X_t satisfying

$$\mathrm{law}(X_t) = \mathrm{law}\Big(t\overrightarrow{X}_0 + \sqrt{1-t^2}Y\Big), \quad \overrightarrow{X}_0 \sim p \text{ and } Y \sim \mathcal{N}(0, \mathrm{Id}) \text{ ind}$$

by taking

$$p_t(x_t) = \int p_t(x_t|x_0, x_1) d\pi(x_0, x_1), \quad \text{with } p_t(x_t|x_0, x_1) \text{ the pdf of } \mathcal{N}(\mu_t(x_0, x_1), \sigma_t^2).$$

for $\pi = p \otimes \gamma_d$ and $\mu_t((x_0, x_1)) = tx_1$, $\sigma_t = \sqrt{1-t^2}$. Then,

$$\begin{aligned} v_t(x_t) &= \mathbb{E}\left[v_t(x_t|X_0, X_1)\Big|X_t = x_t\right] \\ &= \frac{1}{t}\big(s(\log(t^{-1}), x_t) + x_t\big), \end{aligned}$$

with s the score function.

Diffusion Flow matching

For $\overrightarrow{X}_t$ the forward process in diffusion models

$$\mathrm{law}\big(\overrightarrow{X}_t\big) = \mathrm{law}\Big(e^{-t}\overrightarrow{X}_0 + \sqrt{1-e^{-2t}}\,Y\Big), \quad \overrightarrow{X}_0 \sim p \text{ and } Y \sim \mathcal{N}(0, \mathrm{Id}) \text{ ind.}$$

Rescaling, we can define X_t satisfying

$$\mathrm{law}\big(X_t\big) = \mathrm{law}\Big(t\overrightarrow{X}_0 + \sqrt{1-t^2}\,Y\Big), \quad \overrightarrow{X}_0 \sim p \text{ and } Y \sim \mathcal{N}(0, \mathrm{Id}) \text{ ind}$$

by taking

$$p_t(x_t) = \int p_t(x_t|x_0, x_1)d\pi(x_0, x_1), \quad \text{ with } p_t(x_t|x_0, x_1) \text{ the pdf of } \mathcal{N}(\mu_t(x_0, x_1), \sigma_t^2).$$

for $\pi = p \otimes \gamma_d$ and $\mu_t((x_0, x_1)) = tx_1$, $\sigma_t = \sqrt{1-t^2}$. Then,

$$\begin{aligned} v_t(x_t) &= \mathbb{E}\left[v_t(x_t|X_0, X_1)\middle| X_t = x_t\right] \\ &= \frac{1}{t}\big(s(\log(t^{-1}), x_t) + x_t\big), \end{aligned}$$

with s the score function.

Unified minimax rates

Conjecture: Diffusion models, Flow matching and Shrodinger bridges can be shown to be minimax under the same framework:

1. There exists a flow of measures $(p_t)_{t\in[0,1]}$ such that $p_1 = p$ and $p_t = \text{law}(X_t)$ with

$$dX_t = v_t(X_t)dt + b_t dB_t, \quad X_0 \sim p_0.$$

2. There exists a conditional vector field $v_t(\cdot|)$ such that

$$v_t(x_t) = \mathbb{E}[v_t(x_t|X_0, X_1)\big|X_t = x_t]$$

3. The drift is learnt the following way

$$\hat{v} \in \arg\min_{\phi\in\mathcal{S}} \frac{1}{n}\sum_{i=1}^{n}\int_0^1 \mathbb{E}_{X_t\sim p_t(\cdot|X_1^{(i)},X_0^{(i)})}\big[\|\phi_t(X_t) - v_t(X_t|X_0^{(i)}, X_1^{(i)})\|^2\big]dt,$$

with $\mathcal{S}$ a nn class, $(X_0^{(i)}, X_1^{(i)}) \sim \pi \subset \Pi(p, \gamma_d)$ iid.

Unified minimax rates

Conjecture: Diffusion models, Flow matching and Shrodinger bridges can be shown to be minimax under the same framework:

1. There exists a flow of measures $(p_t)_{t\in[0,1]}$ such that $p_1 = p$ and $p_t = \text{law}(X_t)$ with
$$dX_t = v_t(X_t)dt + b_t dB_t, \quad X_0 \sim p_0.$$
2. There exists a conditional vector field $v_t(\cdot|)$ such that
$$v_t(x_t) = \mathbb{E}[v_t(x_t|X_0, X_1)\big|X_t = x_t]$$
3. The drift is learnt the following way
$$\hat{v} \in \arg\min_{\phi\in\mathcal{S}} \frac{1}{n}\sum_{i=1}^{n}\int_0^1 \mathbb{E}_{X_t\sim p_t(\cdot|X_1^{(i)},X_0^{(i)})}\left[\|\phi_t(X_t) - v_t(X_t|X_0^{(i)}, X_1^{(i)})\|^2\right]dt,$$
with $\mathcal{S}$ a nn class, $(X_0^{(i)}, X_1^{(i)}) \sim \pi \subset \Pi(p, \gamma_d)$ iid.

Unified minimax rates

Conjecture: Diffusion models, Flow matching and Shrodinger bridges can be shown to be minimax under the same framework:

1. There exists a flow of measures $(p_t)_{t\in[0,1]}$ such that $p_1 = p$ and $p_t = \text{law}(X_t)$ with
$$dX_t = v_t(X_t)dt + b_t dB_t, \quad X_0 \sim p_0.$$

2. There exists a conditional vector field $v_t(\cdot|)$ such that
$$v_t(x_t) = \mathbb{E}[v_t(x_t|X_0, X_1)\big|X_t = x_t]$$

3. The drift is learnt the following way
$$\hat{v} \in \arg\min_{\phi\in\mathcal{S}} \frac{1}{n}\sum_{i=1}^{n}\int_0^1 \mathbb{E}_{X_t\sim p_t(\cdot|X_1^{(i)},X_0^{(i)})}\big[\|\phi_t(X_t) - v_t(X_t|X_0^{(i)}, X_1^{(i)})\|^2\big]dt,$$
with $\mathcal{S}$ a nn class, $(X_0^{(i)}, X_1^{(i)}) \sim \pi \subset \Pi(p, \gamma_d)$ iid.

Unified minimax rates

Conjecture: Diffusion models, Flow matching and Shrodinger bridges can be shown to be minimax under the same framework:

1. There exists a flow of measures $(p_t)_{t\in[0,1]}$ such that $p_1 = p$ and $p_t = \text{law}(X_t)$ with
$$dX_t = v_t(X_t)dt + b_t dB_t, \quad X_0 \sim p_0.$$
2. There exists a conditional vector field $v_t(\cdot|)$ such that
$$v_t(x_t) = \mathbb{E}[v_t(x_t|X_0, X_1)\big|X_t = x_t]$$
3. The drift is learnt the following way
$$\hat{v} \in \underset{\phi\in\mathcal{S}}{\arg\min} \frac{1}{n}\sum_{i=1}^{n}\int_0^1 \mathbb{E}_{X_t\sim p_t(\cdot|X_1^{(i)},X_0^{(i)})}\big[\|\phi_t(X_t) - v_t(X_t|X_0^{(i)}, X_1^{(i)})\|^2\big]\,dt,$$
with $\mathcal{S}$ a nn class, $(X_0^{(i)}, X_1^{(i)}) \sim \pi \subset \Pi(p, \gamma_d)$ iid.

Manifold setting

Key insight: Minimax rates are achieved as soon as the score inherits the regularity of p.

- High-order Hölder regularity on high-probability sets of p_t suffices.
- Suggests that the score likely belongs to Sobolev-type spaces with respect to p_t.

Conjecture for manifold-data: Let p be supported on a $\mathcal{C}^{\beta+1}$ compact submanifold with density in $\mathcal{H}^{\beta}$. Then for all $t > 0$, the score satisfies:

$$s(t, \cdot) \in W^{\beta,2}_{t^{-1/2}}(\mathbb{R}^d, p_t).$$

Manifold setting

Key insight: Minimax rates are achieved as soon as the score inherits the regularity of p.

- High-order Hölder regularity on high-probability sets of p_t suffices.
- Suggests that the score likely belongs to Sobolev-type spaces with respect to p_t.

Conjecture for manifold-data: Let p be supported on a $\mathcal{C}^{\beta+1}$ compact submanifold with density in $\mathcal{H}^{\beta}$. Then for all $t > 0$, the score satisfies:

$$s(t, \cdot) \in W^{\beta,2}_{t^{-1/2}}(\mathbb{R}^d, p_t).$$

References

Joe Benton, George Deligiannidis, and Arnaud Doucet. Error bounds for flow matching methods. *arXiv preprint arXiv:2305.16860*, 2023.

Changxiao Cai and Gen Li. Minimax optimality of the probability flow ode for diffusion models. *arXiv preprint arXiv:2503.09583*, 2025.

Hongrui Chen, Holden Lee, and Jianfeng Lu. Improved analysis of score-based generative modeling: User-friendly bounds under minimal smoothness assumptions. In *Proceedings of the 40th International Conference on Machine Learning*, volume 202, pages 4735–4763. PMLR, 23–29 Jul 2023a.

Sitan Chen, Sinho Chewi, Jerry Li, Yuanzhi Li, Adil Salim, and Anru R Zhang. Sampling is as easy as learning the score: theory for diffusion models with minimal data assumptions. *arXiv preprint arXiv:2209.11215*, 2022.

Sitan Chen, Sinho Chewi, Holden Lee, Yuanzhi Li, Jianfeng Lu, and Adil Salim. The probability flow ode is provably fast. *Advances in Neural Information Processing Systems*, 36: 68552–68575, 2023b.

Daniel Zhengyu Huang, Jiaoyang Huang, and Zhengjiang Lin. Convergence analysis of probability flow ode for score-based generative models. *arXiv preprint arXiv:2404.09730*, 2024.

Holden Lee, Jianfeng Lu, and Yixin Tan. Convergence of score-based generative modeling for general data distributions. In *International Conference on Algorithmic Learning Theory*, pages 946–985. PMLR, 2023.

Kazusato Oko, Shunta Akiyama, and Taiji Suzuki. Diffusion models are minimax optimal distribution estimators. In *International Conference on Machine Learning*, pages 26517–26582. PMLR, 2023.

Andre Wibisono, Yihong Wu, and Kaylee Yingxi Yang. Optimal score estimation via empirical bayes smoothing. In *The Thirty Seventh Annual Conference on Learning Theory*, pages 4958–4991. PMLR, 2024.

Kaihong Zhang, Heqi Yin, Feng Liang, and Jingbo Liu. Minimax optimality of score-based diffusion models: Beyond the density lower bound assumptions. In *Proceedings of the 41st International Conference on Machine Learning*, volume 235, pages 60134–60178. PMLR, 21–27 Jul 2024.