Generalization error bound for denoising score matching under relaxed manifold assumption

Konstantin Yakovlev, Nikita Puchkin
HSE University

At a Glance

- We suggest a statistical model with relaxed manifold assumption;
- We show that the denoising score matching estimate enjoys a rate of convergence $\mathcal{O}(n^{-2\beta/(2\beta+d)})$;
- We also carefully track how the hidden constant behind $\mathcal{O}(\cdot)$ depends on the ambient dimension D and stopping time t_0;

Preliminaries: Denoising Score Matching

Denoising Diffusion Probalistic Models are based on the Ornstein-Uhlenbeck

OU Forward: $\mathrm{d}X_t = -X_t\,\mathrm{d}t + \sqrt{2}\,\mathrm{d}W_t, \quad 0 \leqslant t \leqslant T, \quad X_0 \sim \mathsf{p}_0^*,$

OU Reverse: $\mathrm{d}Z_t = \big(Z_t + 2\nabla \log \mathsf{p}^*_{T-t}(Z_t)\big)\,\mathrm{d}t + \sqrt{2}\,\mathrm{d}B_t, \quad Z_0 \sim \mathsf{p}_T^*$

The score function

$$s^*(y,t) = \nabla \log \mathsf{p}_t^*(y), \quad (y,t) \in \mathbb{R}^D \times [0,T],$$

and should be estimated from i.i.d. samples $Y_1, \dots, Y_n \sim \mathsf{p}_0^*$.
Score matching approach aims to minimize

$$\int_{t_0}^{T} \mathbb{E}_{X_t} \left\| s(X_t,t) - s^*(X_t,t) \right\|^2 \mathrm{d}t,$$

Since it is intractable, introduce a new loss function

$$\ell(s, X_0) = \int_{t_0}^{T} \mathbb{E}\left[\left\| s(X_t,t) + \frac{X_t - m_t X_0}{\sigma_t^2} \right\|^2 \,\Big|\, X_0 \right] \mathrm{d}t.$$

Given i.i.d. samples $Y_1, \dots, Y_n \sim \mathsf{p}_0^*$, the denoising score matching estimate is defined as an empirical risk minimizer

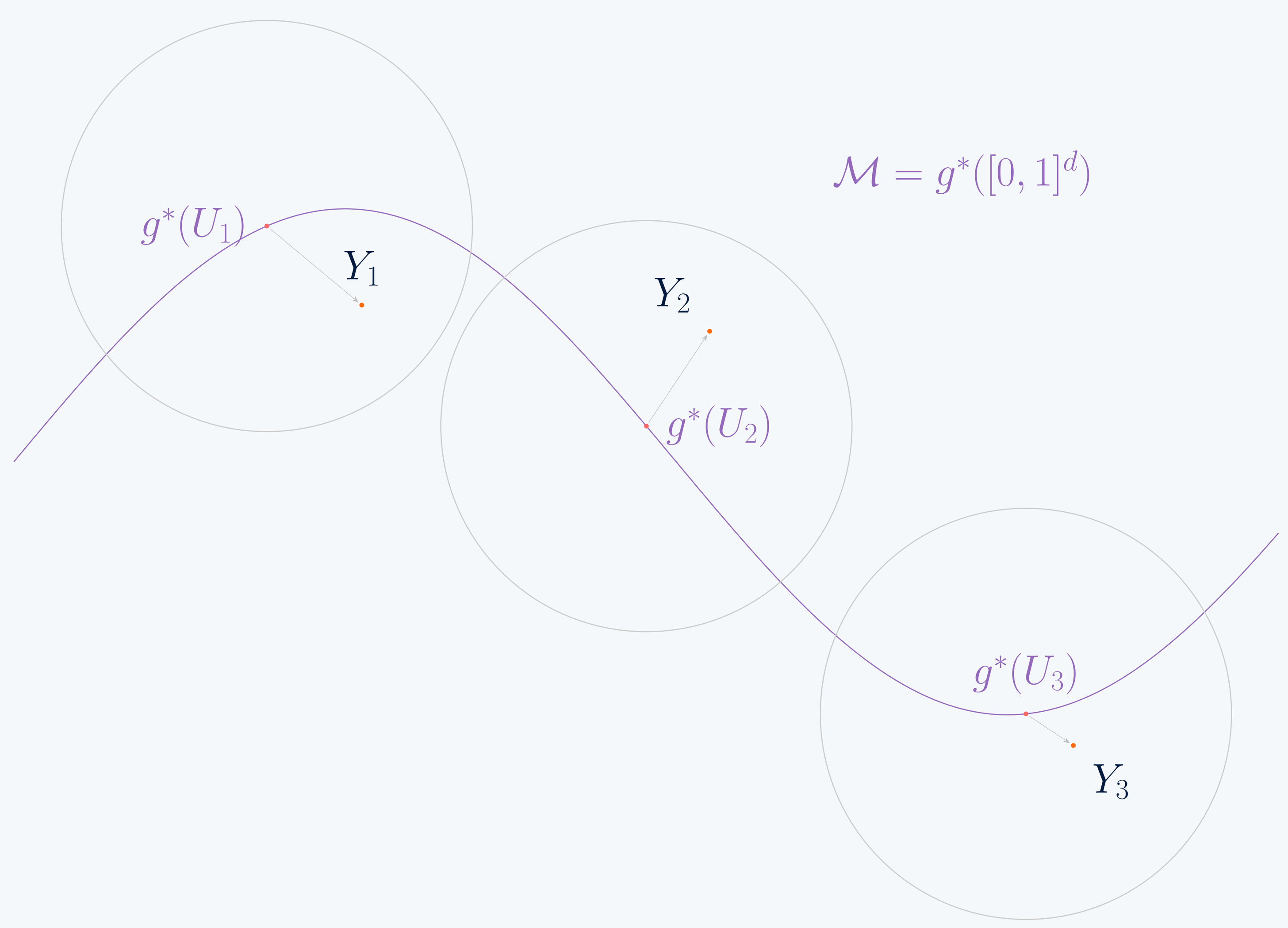
$$\widehat{s} \in \operatorname*{argmin}_{s \in \mathcal{S}} \left\{ \frac{1}{n} \sum_{i=1}^{n} \ell(s, Y_i) \right\}.$$

Relaxed Manifold Assumption

Given a *generator* from the Hölder ball $g^* \in \mathcal{H}^\beta([0,1]^d, \mathbb{R}^D, H)$, $\|g^*\|_{L^\infty([0,1]^d)} \leqslant 1$, and $\sigma_{\text{data}} \in [0,1)$, the observed samples $Y_1, \dots, Y_n$ are i.i.d. copies of a random element $X_0 \in \mathbb{R}^D$ generated from the model

$$X_0 = g^*(U) + \sigma_{\text{data}}\,\xi,$$

where $U \sim \mathrm{Un}([0,1]^d)$ and $\xi \sim \mathcal{N}(0, I_D)$ are independent.



The Class of Score Estimators

The class of neural score estimators $\mathcal{S}(L,W,S,B)$ is defined as

$$\mathcal{S}(L,W,S,B) = \left\{ s(y,t) := -\frac{y}{m_t^2\sigma^2 + \sigma_t^2} + \frac{m_t \mathrm{clip}_2(f(y,t))}{m_t^2\sigma^2 + \sigma_t^2} \,:\, f \in \mathsf{NN}(L,W,S,B),\ \sigma \in [0,1) \right\}.$$

Score Function Approximation

Theorem. Assume that $g^* \in \mathcal{H}^\beta([0,1]^d, \mathbb{R}^D, H)$ and let $s^*(y,t) = \nabla \log \mathsf{p}_t^*(y)$ be the corresponding score function. Fix an arbitrary $\varepsilon \in (0,1)$ such that

$$D\varepsilon\sqrt{\log(1/\varepsilon)} \leqslant \widetilde{\sigma}_{t_0}^2, \quad H\varepsilon\sqrt{D} \leqslant \widetilde{\sigma}_{t_0}, \quad \text{and} \quad \frac{H d^{\lfloor\beta\rfloor}\varepsilon^\beta\sqrt{D}}{\lfloor\beta\rfloor!} \leqslant 1 \wedge \widetilde{\sigma}_{t_0}.$$

Then there exists a score function $s \in \mathcal{S}(L,W,S,B)$ with $L \lesssim D^2 + \log^4(1/\varepsilon)$, $\log B \lesssim D + \log^2(1/\varepsilon)$, and

$$\|W\|_\infty \lesssim D^2\varepsilon^{-d}\left(\frac{1}{t_0 + \sigma_{\text{data}}^2} \vee 1\right)\left(D + \log^2(1/\varepsilon)\right)^3,$$

$$S \lesssim D^2\varepsilon^{-d}\left(\frac{1}{t_0 + \sigma_{\text{data}}^2} \vee 1\right)\left(D + \log^2(1/\varepsilon)\right)^3 + D\varepsilon^{-d}\left(D + \log^2\frac{1}{\varepsilon}\right)^{2\binom{d+\lfloor\beta\rfloor}{d}+5}$$

such that

$$\int_{t_0}^{T} \mathbb{E}_{X_t}\left\| s^*(X_t,t) - s(X_t,t) \right\|^2 \mathrm{d}t \lesssim \frac{D\varepsilon^{2\beta}}{\widetilde{\sigma}_{t_0}^2}.$$

Score Function Estimation

Theorem. Assume that the conditions of score approximation theorem hold. Let also $T \geqslant 1$ and let the sample size be sufficiently large, that is, it fulfils

$$\sigma_{t_0}^2 n \geqslant \left(\frac{D\sqrt{\log(n\sigma_{t_0}^2)}}{\sigma_{t_0}^2\sqrt{2\beta+d}}\right)^{2\beta+d} \vee \left(\frac{H d^{\lfloor\beta\rfloor}\sqrt{D}}{\lfloor\beta\rfloor!\,\sigma_{t_0}}\right)^{2+d/\beta} \vee \left(\frac{H\sqrt{D}}{\sigma_{t_0}}\right)^{2\beta+d}.$$

Then, for any $\delta \in (0,1)$, with probability at least $(1-\delta)$ the excess risk of an empirical risk minimizer over the class $\mathcal{S}(L,W,S,B)$ with

$$L \lesssim D^2 \log^4 n, \quad \|W\|_\infty \lesssim t_0^{-1} D^5 (n\sigma_{t_0}^2)^{\frac{d}{2\beta+d}} \log^6 n,$$
$$S \lesssim t_0^{-1} D^{6+2\binom{d+\lfloor\beta\rfloor}{d}} (n\sigma_{t_0}^2)^{\frac{d}{2\beta+d}} (\log n)^{10+4\binom{d+\lfloor\beta\rfloor}{d}}, \quad \log B \lesssim D\log^2 n$$

satisfies the inequality

$$\int_{t_0}^{T} \mathbb{E}_{X_t}\|\widehat{s}(X_t,t) - s^*(X_t,t)\|^2\,\mathrm{d}t \lesssim \frac{T^2 D^{12+2\binom{d+\lfloor\beta\rfloor}{d}}}{\sigma_{t_0}^2}(n\sigma_{t_0}^2)^{-\frac{2\beta}{2\beta+d}} L(t_0,n)\log(4/\delta),$$

where

$$L(t_0,n) = (\log n)^{20+4\binom{d+\lfloor\beta\rfloor}{d}} \log T \log D \log^3(1/t_0).$$

Distribution Estimation

Theorem. Assume that $\sigma_{\text{data}} > 0$ and suppose that the conditions of the score estimation theorem hold for $t_0 = \sigma_{\text{data}}^{(2\beta+d)/(3\beta+d)} n^{-\beta/(3\beta+d)}$ and $T \asymp \log n$. Then, for any $\delta \in (0,1)$, with probability at least $1-\delta$, the following holds:

$$\mathrm{TV}(\widehat{Z}_{T-t_0}, X_0) \lesssim D^{6+\binom{d+\lfloor\beta\rfloor}{d}} \sigma_{\text{data}}^{-\frac{1}{3\beta+d}} n^{-\frac{\beta}{6\beta+2d}} (\log n)^{25+4\binom{d+\lfloor\beta\rfloor}{d}} \log D \log^{1/2}(4/\delta).$$

References

- Yakovlev, Konstantin, and Puchkin, Nikita. "Generalization error bound for denoising score matching under relaxed manifold assumption." Conference on Learning Theory. PMLR, 2025.
- Oko, Kazusato, Shunta Akiyama, and Taiji Suzuki. "Diffusion models are minimax optimal distribution estimators." International Conference on Machine Learning. PMLR, 2023.
- Benton, Joe, et al. "Nearly d-Linear Convergence Bounds for Diffusion Models via Stochastic Localization." The Twelfth International Conference on Learning Representations, 2024.