Tight bounds for Schrödinger potential estimation in unpaired image-to-image translation problems

N. Puchkin[1], D. Suchkov[1,2], A. Naumov[1], and D. Belomestny[1,3]

[1]HSE University [2]Skoltech [3]Duisburg-Essen University

HSE Study Center, Voronovo, October 28, 2025

Diffusion-based generative models

$X_0 \sim p_0$ $\quad$ $dX_t = f(X_t, t)\,dt + g(X_t, t)dt$ $\quad$ $X_T \sim p_T$

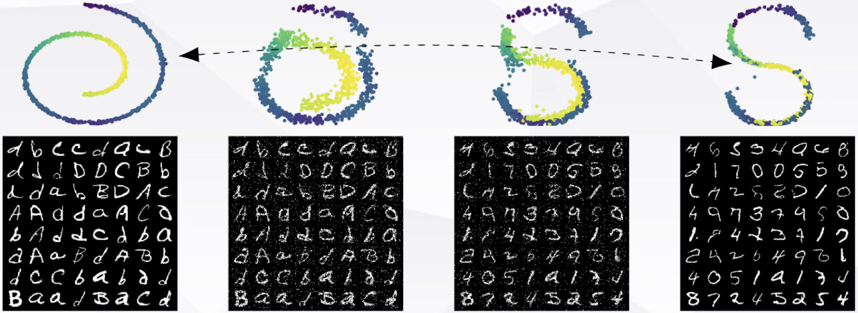


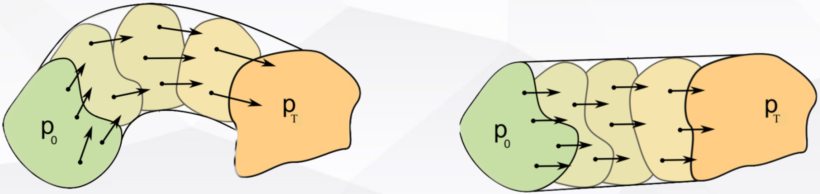
Figure: diffusion Schrödinger bridge. Image courtesy: V. de Bortoli et al. "Diffusion Schrödinger Bridge with Applications to Score-Based Generative Modeling", NeurIPS, 2021

Stochastic optimal control approach to generative modelling

$$X_0 \sim \mathrm{p}_0 \qquad dX_t = f(X_t, t)\,dt + g(X_t, t)dW_t \qquad X_T \sim \mathrm{p}_T$$

Observation 1: there exist multiple ways to transfer p_0 to p_T

Observation 2: simpler trajectories are easier to simulate



Base process:

$$X_0^0 \sim \mathrm{p}_0, \quad \mathrm{d}X_t^0 = b(m - X_t^0)\,\mathrm{d}t + \Sigma^{1/2}\,\mathrm{d}W_t, \quad 0 < t < T$$

Controlled process with Schrödinger log-potential φ:

$$X_t^\varphi \sim \mathrm{p}_0, \quad \mathrm{d}X_t^\varphi = \big(bm - bX_t^\varphi + \Sigma\,\nabla \log \mathcal{T}_{T-t}[e^\varphi](X_t^\varphi)\big)\,\mathrm{d}t + \Sigma^{1/2}\,\mathrm{d}W_t, \quad 0 < t < T$$

where $\mathcal{T}_t$ is the Ornstein-Uhlenbeck operator:

$$\mathcal{T}_t[g](x) = \mathbb{E}_{\xi \sim \mathcal{N}(m_t(x), \Sigma_t)}\, g(\xi), \quad m_t(x) = (1 - e^{-bt})m + e^{-bt}x, \quad \Sigma_t = \frac{1 - e^{-2bt}}{2b}\,\Sigma$$

Controlled process with Schrödinger log-potential φ:

$$X_t^\varphi \sim \mathrm{p}_0, \quad \mathrm{d}X_t^\varphi = \big(bm - bX_t^\varphi + \Sigma\,\nabla \log \mathcal{T}_{T-t}[e^\varphi](X_t^\varphi)\big)\,\mathrm{d}t + \Sigma^{1/2}\,\mathrm{d}W_t, \quad 0 < t < T$$

Remark 1: the joint density of X_0^φ and X_T^φ has a form

$$\pi^\varphi(x, y) = \frac{\mathrm{p}_0(x)\,\mathrm{q}_T(y \,|\, x)\,e^{\varphi(y)}}{\mathcal{T}_T[e^\varphi](x)}, \quad \mathrm{q}_T(y \,|\, x) \text{ – density of } \mathcal{N}(m_t(x), \Sigma_t)$$

The marginal density of X_T^φ is given by

$$\mathrm{p}_T^\varphi(y) = \int \pi^\varphi(x, y)\,\mathrm{d}x = \int \frac{\mathrm{p}_0(x)\,\mathrm{q}_T(y \,|\, x)\,e^{\varphi(y)}}{\mathcal{T}_T[e^\varphi](x)}\,\mathrm{d}x$$

Controlled process with Schrödinger log-potential φ:

$$X_t^\varphi \sim \mathrm{p}_0, \quad \mathrm{d}X_t^\varphi = \left(bm - bX_t^\varphi + \Sigma \nabla \log \mathcal{T}_{T-t}[e^\varphi](X_t^\varphi)\right)\mathrm{d}t + \Sigma^{1/2}\,\mathrm{d}W_t, \quad 0 < t < T$$

Remark 2: for any φ the function $u^\varphi(x,t) = \Sigma \nabla \log \mathcal{T}_{T-t}[e^\varphi](x)$ solves the stochastic optimal control problem [Dai Pra, 1991, Theorem 3.2]:

$$\begin{cases} \int_0^T \mathbb{E}\left\|\Sigma^{-1/2} u(X_t, t)\right\|^2 \longrightarrow \min_u \\ X_0 \sim \mathrm{p}_0, \quad X_T \sim \mathrm{p}_T^\varphi \\ \mathrm{d}X_t = \left(bm - bX_t + u(X_t, t)\right)\mathrm{d}t + \Sigma^{1/2}\,\mathrm{d}W_t, \quad 0 < t < T \end{cases}$$

$u^\varphi(x,t) = \Sigma \nabla \log \mathcal{T}_{T-t}[e^\varphi](x)$ **transforms** p_0 **into** p_T^φ **in an optimal way**

Statistical setup

The initial and target densities p_0 and p_T are unknown, but we have i.i.d. samples

$$Z_1, \ldots, Z_n \sim \mathrm{p}_0 \quad \text{and} \quad Y_1, \ldots, Y_n \sim \mathrm{p}_T$$

Schrödinger log-potential for transferring p_0 into p_T:

$$\pi^*(x, y) = \frac{\mathrm{p}_0(x)\, \mathrm{q}_T(y \,|\, x)\, e^{\varphi^*(y)}}{\mathcal{T}_T[e^{\varphi^*}](x)}, \qquad \mathrm{p}_T(y) = \int \frac{\mathrm{p}_0(x)\, \mathrm{q}_T(y \,|\, x)\, e^{\varphi^*(y)}}{\mathcal{T}_T[e^{\varphi^*}](x)}\, \mathrm{d}x$$

Remark 1. The sample sizes are assumed equal to ease the presentation

Remark 2. The joint distribution of (Z_i, Y_i) may differ from π^*

Statistical setup

The initial and target densities p_0 and p_T are unknown, but we have i.i.d. samples

$$Z_1, \dots, Z_n \sim p_0 \quad \text{and} \quad Y_1, \dots, Y_n \sim p_T$$

Proximity measure:

$$\mathrm{KL}(\pi^*, \pi^\varphi) = \int \log \frac{\pi^*(z,y)}{\pi^\varphi(z,y)}\, \pi^*(z,y)\, \mathrm{d}z\mathrm{d}y$$

Observation:

$$\mathrm{KL}(\pi^*, \pi^\varphi) = \mathcal{L}(\varphi) - \mathcal{L}(\varphi^*), \quad \text{where} \quad \mathcal{L}(\varphi) = \mathbb{E}_{Z\sim p_0} \log \mathcal{T}_T[e^\varphi](Z) - \mathbb{E}_{Y\sim p_T} \varphi(Y)$$

$\mathcal{L}(\varphi)$ includes only marginal distributions!

Statistical setup

The initial and target densities p_0 and p_T are unknown, but we have i.i.d. samples

$$Z_1, \dots, Z_n \sim \mathrm{p}_0 \quad \text{and} \quad Y_1, \dots, Y_n \sim \mathrm{p}_T$$

Estimation:

$$\widehat{\pi} = \pi^{\widehat{\varphi}}, \quad \widehat{\varphi} \in \operatorname*{Argmin}_{\varphi \in \Phi} \left\{ \frac{1}{n} \sum_{j=1}^{n} \log \mathcal{T}_T[e^{\varphi}](Z_j) - \frac{1}{n} \sum_{i=1}^{n} \varphi(Y_i) \right\},$$

where Φ is a fixed reference class of functionals

Remark. A similar empirical functional (with $\mathcal{T}_T$ replaced by a different integral operator) was considered in [Korotin et al., 2024]

Illustration

Figure: examples of unpaired image-to-image translation with the suggested approach. Top: adult to child. Middle and bottom: male to female. In all the cases $\Sigma = \varepsilon I_d$ and $\exp\left\{\widehat{\varphi}(x) - b\varepsilon^{-1}\|x\|^2/(1 - e^{-bT})\right\}$ is a Gaussian mixture

Assumptions

1. The density p_0 has a bounded support: $\|\Sigma^{-1/2}(x - m)\| \leqslant R$ for all $x \in \operatorname{supp}(\mathsf{p}_0)$
2. The density p_T has sub-Gaussian tails $\mathbb{E}_{Y \sim \rho_T} e^{u^\top Y} \leqslant e^{\mathrm{v}^2 \|u\|^2/2}$ for any $u \in \mathbb{R}^d$
3. Every $\varphi \in \Phi \cup \{\varphi^*\}$ satisfies $-L\|\Sigma^{-1/2}(x - m)\|^2 - M \leqslant \varphi(x) \leqslant M$ for all $x \in \mathbb{R}^d$
4. Normalization condition: for any $\varphi \in \Phi \cup \{\varphi^*\}$ it holds that $\mathcal{T}_\infty \varphi = 0$
5. The class Φ is parametric: $\Phi = \{\varphi_\theta : \theta \in \Theta\}$, where $\Theta \subseteq [-1, 1]^D$
6. Lipschitz parametrization: there exists $\Lambda > 0$ such that

$$|\varphi_\theta(x) - \varphi_{\theta'}(x)| \leqslant \Lambda \left(1 + \|x\|^2\right) \|\theta - \theta'\|_\infty \quad \text{for all } \theta, \theta' \in \Theta \text{ and all } x \in \mathbb{R}^d$$

Theorem

Grant the aforementioned assumptions. For any $n \in \mathbb{N}$ and $\delta \in (0, 1)$, introduce

$$\Upsilon(n, \delta) = D(M + d) \left(1 \vee \frac{L}{b}\right) (M + \log((L \vee \Lambda)nd) + \log(1/\delta)) \frac{\log n}{n} \quad \text{and} \quad \Delta = \inf_{\varphi \in \Phi} \mathrm{KL}(\pi^*, \pi^\varphi).$$

Then there exists

$$T_0 \lesssim \frac{1}{b} \log\log \frac{1}{\Upsilon(n, 1)} \vee \frac{1}{b} \log\left(d + LR^2 + \frac{Ld^2}{b} + M\right) \lesssim \log\log n$$

such that for any $\delta \in (0, 1)$ and any $T \geqslant T_0$ with probability at least $(1 - \delta)$ it holds that

$$\mathrm{KL}(\pi^*, \widehat{\pi}) - \Delta \lesssim \sqrt{\Upsilon(n, \delta)\Delta \left(1 \vee \log \frac{(1 \vee L/b)(M + d)}{\Delta}\right)} + \Upsilon(n, \delta) \left(1 \vee \log \frac{(1 \vee L/b)(M + d)}{\Delta}\right).$$

More details are in our paper:

N. Puchkin, D. Suchkov, A. Naumov, D. Belomestny,
“Tight bounds for Schrödinger potential estimation in unpaired image-to-image translation problems”

arxiv.org/pdf/2508.07392.pdf

Thank you for attention!

References

[Dai Pra, 1991] Dai Pra, P. (1991).
A stochastic control approach to reciprocal diffusion processes.
Applied Mathematics and Optimization, 23(1):313–329.

[Korotin et al., 2024] Korotin, A., Gushchin, N., and Burnaev, E. (2024).
Light Schrödinger bridge.
In *The Twelfth International Conference on Learning Representations*.