Characterization of Schrödinger Bridges

Grigoriy Ksenofontov, Mikhail Persiianov

Moscow, 2026

Overview

Recap: Dynamic Schrödinger Bridge Problem

Characterization of Schrödinger Bridges

Light Schrödinger Bridges

Recap: Dynamic Schrödinger Bridge Problem

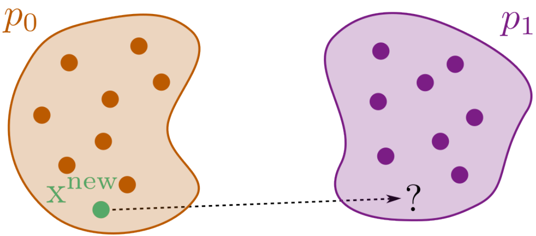
Unsupervised Domain Translation: Problem Setup

Unsupervised setting.

We observe two datasets

$$\{x^{(i)}\}_{i=1}^{M} \sim p_0, \qquad \{y^{(i)}\}_{i=1}^{M} \sim p_1,$$

with **no paired correspondences**.



Goal: learn a mapping

$$T : \mathcal{X} \to \mathcal{Y}$$

such that

$$T_\sharp p_0 \approx p_1.$$

Key difficulty: the pushforward constraint does not uniquely determine T.

Why Static Transport Models Are Not Enough

In practice, the discussed optimal transport corresponds to learning a **one-step generator**:

$$y = T(x, z), \qquad z \sim p(z).$$

Complex distributions are difficult to generate in a single step.

Thus, modern generative models are inherently **dynamic**, i.e.:

$$x \longrightarrow x_{t_1} \longrightarrow \cdots \longrightarrow x_{t_N} \longrightarrow y$$

Examples of state-of-the-art approaches:

- Diffusion models (DDPM, DDIM, NCSN, etc.);
- Flow matching / continuous normalizing flows (FM, Rectified Flows, etc.).

QUESTION: Can we bring dynamics into Optimal Transport?

From Random Variables to Random Paths

Static OT: random variable x on state space $\mathbb{R}^d$

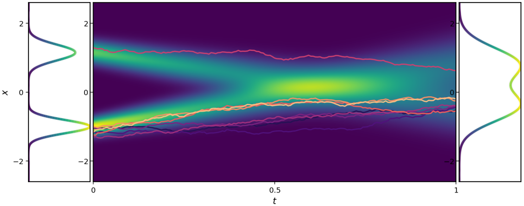
$$\Downarrow$$

Dynamic OT: random path[1] $x(t) \stackrel{\text{def}}{=} x_t$ on path space $C([0,1];\mathbb{R}^d)$

Evaluating the path at times $t \in [0,1]$ yields random variables $\{x_t\}_{t\in[0,1]}$, which together form a stochastic process, or formally:

$$(C([0,T];\mathbb{R}^d), \mathcal{B}(C([0,T];\mathbb{R}^d)), \mathbb{P}),$$

where $\mathbb{P}$ denotes path distribution.



The target probabilistic object is now the path distribution.

[1]We abuse notation by using x_t both for the random path and for its realization, denoted as trajectories.

Path Distributions Induced by SDEs

Given the scaled Wiener process, we define a new path distribution $\mathbb{P}$ by combining deterministic dynamics with Wiener noise:

$$T : dx_t = \underbrace{v(x_t, t)\, dt}_{\substack{\text{drift,} \\ \text{defines deterministic motion}}} + \underbrace{\sqrt{\epsilon}\, dW_t,}_{\substack{\text{scaled Wiener noise,} \\ \text{defines stochastic motion}}} \qquad x_0 \sim p_0,$$

Over a small time step Δt, the SDE can be interpreted as

$$x_{t+\Delta t} \approx x_t + v(x_t, t)\Delta t + \sqrt{\epsilon \Delta t}\, z_t, \qquad z_t \sim \mathcal{N}(0, \mathbb{I}).$$

The trajectories generated by this SDE induce a path distribution $\mathbb{P}$ on $C([0,1]; \mathbb{R}^d)$.

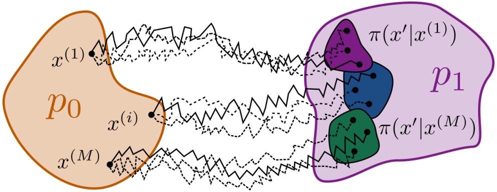
One useful notation is the bridge distribution $\mathbb{P}_{|x,y} \stackrel{\text{def}}{=} \mathbb{P}(\cdot | x_0 = x,\ x_1 = y)$, that is, the path distribution of the SDE conditioned on fixed endpoints.

Dynamic Schrödinger Bridge Formulation

Suppose we are given marginal distributions p_0 and p_1 and a reference scaled Wiener process $\mathbb{W}^\epsilon$. Then, the dynamic Schrödinger Bridge (SB) problem[2] is defined as

$$\mathbb{P}^* = \underset{\mathbb{P}\in\Pi(p_0,p_1)}{\arg\min}\ \mathrm{KL}\left(\mathbb{P}\|\mathbb{W}^\epsilon\right), \qquad \text{(DynSB)}$$

where $\Pi(p_0, p_1)$ is the set of all path distributions with initial and terminal marginals p_0 and p_1, respectively.



Intuitively, the problem seeks the path distribution $\mathbb{P}$ that is closest to the Wiener path distribution $\mathbb{W}^\epsilon$ in KL divergence, while matching the prescribed marginals p_0 and p_1 at times $t = 0$ and $t = 1$, respectively.

[2] Erwin Schrödinger (1931). **Über die umkehrung der naturgesetze.** Verlag der Akademie der Wissenschaften in Kommission bei Walter De Gruyter u ...

Four Equivalent Views of Entropic Transport

Entropic Optimal Transport $\min_{\pi \in \Pi(p_0, p_1)} \mathbb{E}_{x,y\sim\pi}[c(x,y)] - \epsilon \mathbb{E}_{x\sim p_0} \mathrm{H}(\pi(\cdot \vert x))$ (EOT)	**Static Schrödinger Bridge** $\min_{\pi \in \Pi(p_0, p_1)} \mathrm{KL}(\pi \Vert \pi^{\mathrm{ref}})$ (StatSB)
Dynamic Schrödinger Bridge $\min_{\mathbb{P} \in \Pi(p_0, p_1)} \mathrm{KL}(\mathbb{P} \Vert \mathbb{W}^{\epsilon})$ (DynSB)	**Stochastic Optimal Control** $\min_{v} \frac{1}{2\epsilon} \mathbb{E}_{\mathbb{P}} \left[\int_0^1 \Vert v(x_t, t) \Vert^2 dt \right]$ (SOC) s.t. $dx_t = v(x_t, t)\, dt + \sqrt{\epsilon}\, d\widetilde{W}_t$

From drift control to a path-space view of Schrödinger Bridge

From the stochastic optimal control formulation, the dynamic Schrödinger Bridge solves

$$\min_{v} \frac{1}{2\epsilon} \mathbb{E}_{\mathbb{P}} \left[\int_0^1 \|v(x_t, t)\|^2 dt \right]$$

subject to

$$dx_t = v(x_t, t)\, dt + \sqrt{\epsilon}\, d\widetilde{W}_t, \qquad x_0 \sim p_0, \qquad x_1 \sim p_1.$$

We have already seen that the SB solution keeps the same Wiener noise and changes only the drift. However, the search space over path distributions $\mathbb{P}$ is still completely arbitrary, and therefore does not automatically enforce these properties.

QUESTION: Can we represent the SB directly on path space as a special transform of the reference process, so that the stochastic dynamics are already built in, and only the endpoint marginals remain to be matched?

Characterization of Schrödinger Bridges

Doob's h-transform of path distributions

Let $\mathbb{P}^{\mathrm{ref}}$ be a Markov reference path distribution (e.g. the Wiener path distribution $\mathbb{W}^{\epsilon}$), and let $h_t(x) \in (0, \infty)$ be a positive space-time harmonic function, such that

$$h_t(x) \stackrel{\text{def}}{=} \mathbb{E}_{\mathbb{P}^{\mathrm{ref}}}[h_1(x_1)|x_t = x]\,.$$

The Doob's h-transform defines a new path distribution $\mathbb{P}^h$ by reweighting $\mathbb{P}^{\mathrm{ref}}$ with h_t:

$$\frac{d\mathbb{P}^h}{d\mathbb{P}^{\mathrm{ref}}} = \frac{h_1(x_1)}{h_0(x_0)} = \frac{h_1(x_1)}{\underbrace{\mathbb{E}_{\mathbb{P}^{\mathrm{ref}}}[h_1(x_1)|x_0 = x]}_{\text{normalization}}}. \tag{1}$$

For any measurable set B of future path segments $(x_s)_{t \le s \le 1}$, we have

$$\mathbb{P}^h((x_s)_{t\le s\le 1} \in B|(x_s)_{0\le s\le t}) = \mathbb{E}_{\mathbb{P}^h}\left[1_{\{(x_s)_{t\le s\le 1}\in B\}}|(x_s)_{0\le s\le t}\right] =$$

$$= \frac{\mathbb{E}_{\mathbb{P}^{\mathrm{ref}}}\left[1_{\{(x_s)_{t\le s\le 1}\in B\}}\, h_1(x_1)|(x_s)_{0\le s\le t}\right]}{\mathbb{E}_{\mathbb{P}^{\mathrm{ref}}}[h_1(x_1)|(x_s)_{0\le s\le t}]} = \mathbb{E}_{\mathbb{P}^h}\left[1_{\{(x_s)_{t\le s\le 1}\in B\}}|x_t\right].$$

> If the reference path distribution $\mathbb{P}^{\mathrm{ref}}$ is Markov, then its Doob-h transform $\mathbb{P}^h$ remains Markov.

Doob-h transform as a drift correction

Consider a reference path distribution $\mathbb{P}^{\mathrm{ref}}$ induced by the SDE

$$dx_t = b(x_t, t)\, dt + \sigma(x_t, t)\, dW_t, \qquad a(x, t) \stackrel{\text{def}}{=} \sigma(x, t)\sigma(x, t)^\top.$$

If the path distribution is transformed by a positive space-time harmonic function $h_t(x)$, then the Doob-h transformed diffusion has the same diffusion coefficient and modified drift

$$dx_t = \Big(b(x_t, t) + a(x_t, t)\nabla \log h_t(x_t)\Big)\, dt + \sigma(x_t, t)\, d\widetilde{W}_t.$$

In the Wiener reference ($\mathbb{P}^{\mathrm{ref}} = \mathbb{W}^\epsilon$) case we obtain $a(x, t) = \epsilon\mathbb{I}$ leading to

$$dx_t = \sqrt{\epsilon}\, dW_t, \qquad \text{under } \mathbb{W}^\epsilon,$$

this becomes

$$dx_t = \epsilon\nabla \log h_t(x_t)\, dt + \sqrt{\epsilon}\, d\widetilde{W}_t. \qquad \text{under } \mathbb{P}^h$$

Doob-h transform keeps the noise unchanged and only **adds a drift correction**.

Doob-h transform: one-step transition reweighting

Let the reference path distribution $\mathbb{P}^{\mathrm{ref}}$ be induced by

$$dx_t = b(x_t, t)\,dt + \sigma(x_t, t)\,dW_t, \qquad a(x, t) \stackrel{\text{def}}{=} \sigma(x, t)\sigma(x, t)^\top.$$

To identify the one-step transition distribution under $\mathbb{P}^h$, consider any measurable set $B \subset \mathbb{R}^d$. Then

$$\begin{aligned}
\mathbb{P}^h(x_{t+\Delta t} \in B | x_t) = \mathbb{E}_{\mathbb{P}^h}\left[1_{\{x_{t+\Delta t} \in B\}} | x_t\right] &= \frac{\mathbb{E}_{\mathbb{P}^{\mathrm{ref}}}\left[1_{\{x_{t+\Delta t} \in B\}}\, h_1(x_1) | x_t\right]}{\mathbb{E}_{\mathbb{P}^{\mathrm{ref}}}\left[h_1(x_1) | x_t\right]} \\
&= \frac{\mathbb{E}_{\mathbb{P}^{\mathrm{ref}}}\left[1_{\{x_{t+\Delta t} \in B\}}\, h_{t+\Delta t}(x_{t+\Delta t}) | x_t\right]}{h_t(x_t)} \\
&= \frac{1}{h_t(x_t)} \int_B q^{\mathrm{ref}}(x_{t+\Delta t} | x_t)\, h_{t+\Delta t}(x_{t+\Delta t})\, dx_{t+\Delta t}.
\end{aligned}$$

Therefore, the transition density under $\mathbb{P}^h$ is

$$q^h(x_{t+\Delta t} | x_t) = q^{\mathrm{ref}}(x_{t+\Delta t} | x_t) \frac{h_{t+\Delta t}(x_{t+\Delta t})}{h_t(x_t)}.$$

What is still missing?

At this point, the connection between dynamic Schrödinger Bridge and Doob-h transform is already suggestive:

- both modify a reference path distribution by reweighting paths,
- both preserve the stochastic part of the dynamics,
- both appear through a drift correction.

However, the derivation so far shows how Doob-h transforms modify a reference process, but it does not yet identify the optimal transform corresponding to SB.

To solve the problem, we recall the equivalence of the static and dynamic SB problems.

Step 1: static SB as a constrained variational problem

Recall the static Schrödinger Bridge problem:

$$\pi^* = \underset{\pi^{\mathbb{P}} \in \Pi(p_0, p_1)}{\arg\min} \ \mathrm{KL}\left(\pi^{\mathbb{P}} \| \pi^{\mathbb{W}^\epsilon}\right).$$

Let's introduce Lagrange multipliers $\varphi(x_0)$ and $\psi(x_1)$ to relax constraints, and consider the Lagrangian

KL

$$\mathcal{L}(\pi^{\mathbb{P}}, \varphi, \psi) = \int \pi^{\mathbb{P}}(x_0, x_1) \log \frac{\pi^{\mathbb{P}}(x_0, x_1)}{\pi^{\mathbb{W}^\epsilon}(x_0, x_1)} dx_0 dx_1$$
$$+ \int \varphi(x_0) \left(\pi_0^{\mathbb{P}}(x_0) - p_0(x_1)\right) dx_0 + \int \psi(x_1) \left(\pi_1^{\mathbb{P}}(x_1) - p_1(x_1)\right) dx_1.$$

Using $\int \varphi(x_0)\, \pi_0^{\mathbb{P}}(x_0) dx_0 = \int \varphi(x_0)\, \pi^{\mathbb{P}}(x_0, x_1) dx_0 dx_1$, and similarly for ψ, this becomes

$$\mathcal{L}(\pi^{\mathbb{P}}, \varphi, \psi) = \int \left[\log \frac{\pi^{\mathbb{P}}(x_0, x_1)}{\pi^{\mathbb{W}^\epsilon}(x_0, x_1)} + \varphi(x_0) + \psi(x_1)\right] \pi^{\mathbb{P}}(x_0, x_1) dx_0 dx_1 -$$
$$- \int \varphi(x_0)\, d p_0(x_0) - \int \psi(x_1)\, d p_1(x_1).$$

Step 1: optimality condition and factorization

To find the minimizer, take the first-order optimality condition, i.e., $\frac{\delta \mathcal{L}}{\delta \pi^{\mathbb{P}}}(x_0, x_1) = 0$. Now, since

$$\frac{\delta}{\delta \pi^{\mathbb{P}}} \left[\int \log \pi^{\mathbb{P}}(x_0, x_1) \frac{\pi^{\mathbb{P}}(x_0, x_1)}{\pi^{\mathbb{W}^\epsilon}(x_0, x_1)} \, dx_0 \, dx_1 \right] = \log \frac{\pi^{\mathbb{P}}(x_0, x_1)}{\pi^{\mathbb{W}^\epsilon}(x_0, x_1)} + 1,$$

one obtains

$$\log \frac{\pi^*(x_0, x_1)}{\pi^{\mathbb{W}^\epsilon}(x_0, x_1)} + 1 + \varphi(x_0) + \psi(x_1) = 0.$$

Rearranging and absorbing constants into the potentials gives[3]

$$\pi^*(x_0, x_1) = f(x_0)\, \pi^{\mathbb{W}^\epsilon}(x_0, x_1)\, g(x_1). \tag{2}$$

where

$$f(x_0) \stackrel{\text{def}}{=} e^{-\varphi(x_0)}, \qquad g(x_1) \stackrel{\text{def}}{=} e^{-1-\psi(x_1)}.$$

Thus, the optimal static SB coupling is obtained by reweighting the reference coupling only through endpoint potentials.

[3] Since static SB is equivalent to EOT, f and g coincide with the exponentiated EOT dual potentials.

Step 2: dynamic SB is the endpoint Doob-h transform

Recall the KL decomposition:

$$\mathrm{KL}\left(\mathbb{P}\|\mathbb{W}^{\epsilon}\right) = \mathrm{KL}\left(\pi^{\mathbb{P}}\|\pi^{\mathbb{W}^{\epsilon}}\right) + \int \mathrm{KL}\left(\mathbb{P}_{|x_0,x_1}\|\mathbb{W}^{\epsilon}_{|x_0,x_1}\right)\, d\pi^{\mathbb{P}}(x_0,x_1).$$

As discussed before, for a fixed endpoint coupling $\pi^{\mathbb{P}}$, the second KL term is minimized by $\mathbb{P}_{|x_0,x_1} = \mathbb{W}^{\epsilon}_{|x_0,x_1}$. Hence, for the optimal dynamic SB solution $\mathbb{P}^*$, the interior of the path does not contribute to the Radon-Nikodym derivative - only the endpoint reweighting remains.

Since the endpoint coupling is exactly the optimal static SB solution, it follows that the full path-space Radon-Nikodym derivative (SB characterization) is[4]

$$\frac{d\mathbb{P}^*}{d\mathbb{W}^{\epsilon}}(x_{[0,1]}) = \frac{\pi^{\mathbb{P}^*}(x_0,x_1)}{\pi^{\mathbb{W}^{\epsilon}}(x_0,x_1)} = f(x_0)\, g(x_1). \tag{3}$$

Hence, the dynamic Schrödinger Bridge is exactly the endpoint Doob-h transform of the reference Wiener process.

[4] Absolute continuity $\mathbb{P} \ll \mathbb{W}^{\epsilon}$ is implicit whenever the path-space KL divergence is finite.

Dynamic SB as an SDE with potential-induced drift

Notably, the SB characterization is two-sided: it involves two generally different endpoint factors, f and g. In contrast, the Doob-h transform is one-sided, with a potential propagated from the terminal time $t = 1$.

To obtain the Doob-h form, define $h_t(x) \overset{\text{def}}{=} \mathbb{E}_{\mathbb{W}^\epsilon}[g(x_1)|x_t = x]$. Then the factor $f(x_0)$ is absorbed into the prescribed initial distribution.[5] Indeed,

$$p_0(x_0) = \mathbb{P}_0^*(x_0) = f(x_0)\, h_0(x_0)\, p_0^{\mathbb{W}^\epsilon}(x_0).$$

Hence, once the process is started with $x_0 \sim p_0$, the reweighting is the Doob-h transform:

$$\frac{d\mathbb{P}^*_{|x_0}}{d\mathbb{W}^\epsilon_{|x_0}}(x_{(0,1]}) = \frac{\pi^{\mathbb{P}^*}(x_1|x_0)}{\pi^{\mathbb{W}^\epsilon}(x_1|x_0)} = \frac{g(x_1)}{\mathbb{E}_{\mathbb{W}^\epsilon}[g(x_1)|x_0]}, \quad \text{s.t. } dx_t = \epsilon \nabla \log h_t(x_t)\, dt + \sqrt{\epsilon}\, d\widetilde{W}_t, \quad x_0 \sim p_0$$

Thus, after fixing the initial distribution p_0, the SB characterization becomes the Doob-h transform given by the potential g, which yields the SB SDE with potential-induced drift.

[5] Analogously, the factor f generates the backward-time characterization.

Light Schrödinger Bridges

Direct optimization via the optimal-form parameterization

To build a practical method, we would like to avoid solving the constrained SB problem:

$$\arg\min_{\pi \in \Pi(p_0, p_1)} \mathrm{KL}\left(\pi(x_0, x_1) \| \pi^{\mathbb{W}^\epsilon}(x_0, x_1)\right).$$

Under the SB characterization we have obtained, we restrict π to the family

$$\pi(x_0, x_1) = p_0(x_0)\,\pi(x_1|x_0) = p_0(x_0)\frac{p^{\mathbb{W}^\epsilon}(x_1|x_0)g(x_1)}{c(x_0)}, \qquad c(x_0) \stackrel{\text{def}}{=} \mathbb{E}_{\mathbb{W}^\epsilon}[g(x_1)|x_0].$$

Then minimizing $\mathrm{KL}(\pi^* \| \pi)$ over this family is equivalent to

$$\min_g \left(\mathbb{E}_{p_0}[\log c(x_0)] - \mathbb{E}_{p_1}[\log g(x_1)]\right), \qquad c(x_0) \stackrel{\text{def}}{=} \mathbb{E}_{\mathbb{W}^\epsilon}[g(x_1)|x_0].$$

By restricting π to the SB form, the marginal constraints are absorbed into the parameterization, and the optimization becomes unconstrained over g.

Deriving the learning objective

Under the given characterization of $\pi(x_0, x_1)$ we have

$$
\begin{aligned}
\mathrm{KL}\left(\pi^* \| \pi\right) &= \int \pi^*(x_0, x_1) \log \frac{\pi^*(x_0, x_1)}{\pi(x_0, x_1)}\, dx_0\, dx_1 \\
&= C - \int \pi^*(x_0, x_1) \log \pi(x_1 | x_0)\, dx_0\, dx_1 \\
&= C - \int \pi^*(x_0, x_1) \log \frac{p^{\mathbb{W}^\epsilon}(x_1 | x_0)\, g(x_1)}{c(x_0)}\, dx_0\, dx_1 \\
&= \widetilde{C} + \int \pi^*(x_0, x_1) \log c(x_0)\, dx_0\, dx_1 - \int \pi^*(x_0, x_1) \log g(x_1)\, dx_0\, dx_1 \\
&= \widetilde{C} + \mathbb{E}_{x_0 \sim p_0}[\log c(x_0)] - \mathbb{E}_{x_1 \sim p_1}[\log g(x_1)].
\end{aligned}
$$

Therefore, up to an additive constant independent of π, minimizing $\mathrm{KL}\left(\pi^* \| \pi\right)$ is equivalent to minimizing

$$
\mathcal{L}(g) \stackrel{\text{def}}{=} \mathbb{E}_{x_0 \sim p_0}[\log c(x_0)] - \mathbb{E}_{x_1 \sim p_1}[\log g(x_1)].
$$

Gaussian parameterization

For an arbitrary potential g, $c(x_0)$ may be difficult to compute. To address this issue, LightSB[6] proposes to parameterize the potential g_θ by a Gaussian mixture model (GMM).

Specifically, since for $x_0 = 0$ the factor $g_\theta(x_1)$ acts as an unnormalized density, we set

$$g_\theta(x_1) \stackrel{\text{def}}{=} \sum_{k=1}^{K} \alpha_k \mathcal{N}(x_1 | r_k, \epsilon S_k), \text{where } \theta \stackrel{\text{def}}{=} \{\alpha_k, r_k, S_k\}_{k=1}^{K}, \ \alpha_k \geq 0, \ r_k \in \mathbb{R}^D, \ 0 \prec S_k \in \mathbb{R}^{D \times D}.$$

For this parameterization, the conditional distribution admits the explicit form

$$\pi_\theta(x_1|x_0) = \frac{1}{c_\theta(x_0)} \sum_{k=1}^{K} \widetilde{\alpha}_k(x_0) \mathcal{N}(x_1 | r_k(x_0), \epsilon S_k), \qquad r_k(x_0) \stackrel{\text{def}}{=} r_k + S_k x_0,$$

$$\widetilde{\alpha}_k(x_0) \stackrel{\text{def}}{=} \alpha_k \exp\left(\frac{x_0^\top S_k x_0 + 2 r_k^\top x_0}{2\epsilon}\right), \qquad c_\theta(x_0) \stackrel{\text{def}}{=} \sum_{k=1}^{K} \widetilde{\alpha}_k(x_0).$$

[6] Alexander Korotin, Nikita Gushchin, and Evgeny Burnaev (2024). **"Light Schrödinger Bridge".** In: *The Twelfth International Conference on Learning Representations*. URL: https://openreview.net/forum?id=WhZoCLRWYJ.

Training

Under GMM parametrization, the

$$\min_{\theta} \mathcal{L}(\theta) = \min_{\theta} \left(\mathbb{E}_{x_0 \sim p_0}[\log c_\theta(x_0)] - \mathbb{E}_{x_1 \sim p_1}[\log g_\theta(x_1)] \right).$$

The empirical optimization functional.

Since the marginals p_0 and p_1 are accessible only through samples

$$\{x_0^{(}1), \ldots, x_0^{(}M)\} \sim p_0, \qquad \{x_1^{(}1), \ldots, x_1^{(}M)\} \sim p_1,$$

we optimize the empirical counterpart of $\mathcal{L}(\theta)$:

$$\widehat{\mathcal{L}}(\theta) \stackrel{\text{def}}{=} \frac{1}{M}\sum_{i=1}^{M} \log c_\theta(x_0^{(}i)) - \frac{1}{M}\sum_{i=1}^{M} \log g_\theta(x_1^{(}i)) \approx \mathcal{L}(\theta).$$

We then apply minibatch gradient descent with respect to the parameters θ.

Static and Dynamic inference

The conditional distribution $\pi_\theta(x_1|x_0)$ is a GMM:

$$\pi_\theta(x_1|x_0) = \sum_{k=1}^{K} \omega_k(x_0)\mathcal{N}(x_1|r_k(x_0), \epsilon S_k),$$

$$\omega_k(x_0) \stackrel{\text{def}}{=} \frac{\widetilde{\alpha}_k(x_0)}{c_\theta(x_0)}.$$

Hence, ancestral sampling is straightforward:

$$k \sim \mathrm{Cat}\big(\omega_1(x_0), \dots, \omega_K(x_0)\big),$$

$$x_1|x_0, k \sim \mathcal{N}(r_k(x_0), \epsilon S_k).$$

Therefore, sampling (x_0, x_1) is very fast.

To sample a full trajectory, it is enough to sample from the Wiener bridge $\mathbb{W}^{\epsilon}_{|x_0,x_1}$.

As discussed before, the Doob-h transform also yields an SDE induced by g_θ:

$$dx_t = b_\theta(x_t, t)\, dt + \sqrt{\epsilon}\, dW_t, \quad x_0 \sim p_0,$$

$$b_\theta(x, t) \stackrel{\text{def}}{=} \epsilon \nabla_x \log\Bigg(\mathcal{N}(x|0, \epsilon(1-t)\mathbb{I}_d)$$

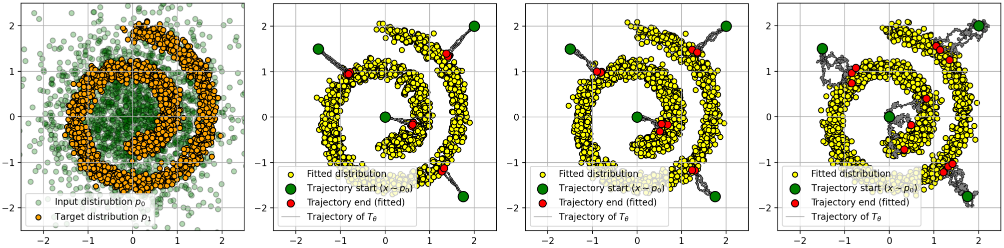
$$\sum_{k=1}^{K}\Big\{\alpha_k \mathcal{N}(r_k|0, \epsilon S_k)\mathcal{N}(h_k(x,t)|0, A_k^t)\Big\}\Bigg),$$

$$\text{where } A_k^t \stackrel{\text{def}}{=} \frac{t}{\epsilon(1-t)}\mathbb{I}_D + \frac{S_k^{-1}}{\epsilon},$$

$$h_k(x,t) \stackrel{\text{def}}{=} \frac{x}{\epsilon(1-t)} + \frac{S_k^{-1} r_k}{\epsilon}.$$

To sample, SDE solver must be used.

Gaussian to Swiss Roll Translation



(a) $x_0 \sim p_0$, $x_1 \sim p_1$. **(b)** $\epsilon = 2 \cdot 10^{-3}$. **(c)** $\epsilon = 0.01$. **(d)** $\epsilon = 0.1$.

Male to Female Translation on CelebA

Dim=512

(a) Male → Female. **(b)** Female → Male. **(c)** Adult → Child. **(d)** Child → Adult.

Conclusions & Takeaways

- The Schrödinger Bridge solution can be characterized on path space as an **endpoint Doob-h transform** of the Wiener process.
- This characterization preserves the **stochastic part** of the dynamics and changes only the **drift** through a potential-induced correction.
- After restricting to the SB form, the constrained problem over couplings becomes an **unconstrained optimization** over the endpoint potential g.
- LightSB parameterizes this potential by a **Gaussian mixture model**, which makes the conditional plan explicit and enables efficient learning from samples.
- The learned model supports both **static inference** via fast sampling of (x_0, x_1) and Wiener bridges, and **dynamic inference** via the corresponding SB SDE.

Main message: Dynamic Schrödinger Bridge admits an explicit path-space characterization through Doob-h transforms, and this leads to practical sample-based methods such as LightSB.