Diffusion Model for Solving Schrödinger Bridge Problem

School of Data science

The Chinese University of Hong Kong, Shenzhen

May 28, 2023

Outline

Part 1. Basic Knowledge of Optimal Transport

Static Optimal Transport

The static optimal transport (OT) map is defined as the coupling between π_0 and π_T minimizing the Wasserstein-2 metric, i.e.

$$\Pi_{0,T}^{\mathrm{OT}} = \operatorname{argmin}_{\Pi_{0,T}} \left\{ \mathbb{E}_{\Pi_{0,T}} \left[\|\mathbf{X}_0 - \mathbf{X}_T\|^2 \right] : \Pi_0 = \pi_0, \Pi_T = \pi_T \right\}.$$

Static Optimal Transport

The static optimal transport (OT) map is defined as the coupling between π_0 and π_T minimizing the Wasserstein-2 metric, i.e.

$$\Pi_{0,T}^{\mathrm{OT}} = \mathrm{argmin}_{\Pi_{0,T}} \left\{ \mathbb{E}_{\Pi_{0,T}} \left[\|\mathbf{X}_0 - \mathbf{X}_T\|^2 \right] : \Pi_0 = \pi_0, \Pi_T = \pi_T \right\}.$$

Under mild assumptions on π_0, π_T, the OT map can be obtained by solving a dynamic Optimal Transport problem.

Dynamic Optimal Transport

$$v_{\mathrm{OT}} = \operatorname{argmin}_v \left\{ \int_0^T \mathbb{E}_{\mathbb{P}_t} \left[\| v(t, \mathbf{X}_t) \|^2 \right] \mathrm{d}t, \right.$$
$$\mathrm{d}\mathbf{X}_t = v(t, \mathbf{X}_t) \, \mathrm{d}t, \mathbb{P}_0 = \pi_0, \mathbb{P}_T = \pi_T.$$

The OT map is then obtained by sampling $\mathbf{X}_0 \sim \pi_0$ followed by $\mathrm{d}\mathbf{X}_t = v_{\mathrm{OT}}(t, \mathbf{X}_t) \, \mathrm{d}t$. The joint samples at the initial and final times satisfy $(\mathbf{X}_0, \mathbf{X}_T) \sim \Pi_{0,T}^{\mathrm{OT}}$.

Part 2. Flow matching

Flow Matching

Conditional Flow Matching is proposed by [1,2,3] concurrently.

They learn an ODE/velocity field: Given a probability density path $p_t(x)$ ($p_1 = q$) and a corresponding vector field $u_t(x)$, which generates $p_t(x)$, we define the Flow Matching (FM) objective as

$$\mathcal{L}_{\mathrm{FM}}(\theta) = \mathbb{E}_{t,p_t(x)} \left\| v_t^{\theta}(x) - u_t(x) \right\|^2,$$

[1] Lipman, et. al. Flow Matching for Generative Modeling. ICLR 2023

[2] Tong et. al. Conditional Flow Matching: Simulation-Free Dynamic Optimal Transport. ArXiv, abs/2302.00482.

[3] Liu et. al. Flow Straight and Fast: Learning to Generate and Transfer Data with Rectified Flow. ICLR 2023.

Flow Matching

Conditional Flow Matching is proposed by [1,2,3] concurrently.
They learn an ODE/velocity field: Given a probability density path $p_t(x)$ ($p_1 = q$) and a corresponding vector field $u_t(x)$, which generates $p_t(x)$, we define the Flow Matching (FM) objective as

$$\mathcal{L}_{\text{FM}}(\theta) = \mathbb{E}_{t,p_t(x)} \left\| v_t^{\theta}(x) - u_t(x) \right\|^2,$$

Problem: $u_t(x)$ is intractable

[1] Lipman, et. al. Flow Matching for Generative Modeling. ICLR 2023

[2] Tong et. al. Conditional Flow Matching: Simulation-Free Dynamic Optimal Transport. ArXiv, abs/2302.00482.

[3] Liu et. al. Flow Straight and Fast: Learning to Generate and Transfer Data with Rectified Flow. ICLR 2023.

Flow Matching

Conditional Flow Matching is proposed by [1][2][3] concurrently.
They learn an ODE/velocity field: Given a probability density path $p_t(x)$ ($p_1 = q$) and a corresponding vector field $u_t(x)$, which generates $p_t(x)$, we define the Flow Matching (FM) objective as

$$\mathcal{L}_{\mathrm{FM}}(\theta) = \mathbb{E}_{t,p_t(x)} \left\| v_t^\theta(x) - u_t(x) \right\|^2,$$

Problem: $u_t(x)$ is intractable
Solution:

- construct a probability path through a mixture of simpler probability path
- marginalizing over the conditional vector flow

[1] Lipman, et. al. Flow Matching for Generative Modeling. ICLR 2023
[2] Tong et. al. Conditional Flow Matching: Simulation-Free Dynamic Optimal Transport. ArXiv, abs/2302.00482.
[3] Liu et. al. Flow Straight and Fast: Learning to Generate and Transfer Data with Rectified Flow. ICLR 2023.

Flow Matching

Construct a probability path through a mixture of simpler probability path:

Flow Matching

Construct a probability path through a mixture of simpler probability path: Given a data sample $x_1 \sim q$, construct $p_t(x \mid x_1)$, s.t., $p_0(x \mid x_1) = p_0(x)$ and $p_1(x \mid x_1)$ to be a distribution concentrated around $x = x_1$, e.g., $p_1(x \mid x_1) = \mathcal{N}\left(x \mid x_1, \sigma^2 I\right)$. Then,

$$p_t(x) = \int p_t(x \mid x_1) q(x_1) dx_1$$

$$p_1(x) = \int p_1(x \mid x_1) q(x_1) dx_1 \approx q(x)$$

Flow Matching

Marginalizing over the conditional vector flow:

Flow Matching

Marginalizing over the conditional vector flow:
Let $u_t(\cdot \mid x_1)$ be the conditional vector field that generates $p_t(\cdot \mid x_1)$

$$u_t(x) = \mathbb{E}_{q(x_1)}[\frac{u_t(x \mid x_1) p_t(x \mid x_1)}{p_t(x)}]$$

Flow Matching

Marginalizing over the conditional vector flow:
Let $u_t(\cdot \mid x_1)$ be the conditional vector field that generates $p_t(\cdot \mid x_1)$

$$u_t(x) = \mathbb{E}_{q(x_1)}[\frac{u_t(x \mid x_1)\, p_t(x \mid x_1)}{p_t(x)}]$$

Then, u_t is the vector field that generate p_t. (The proof is left to the end of this section)

Conditional Flow Matching

Thing can be much simpler:

$$\mathcal{L}_{\mathrm{CFM}}(\theta)=\mathbb{E}_{t,q(x_1),p_t(x|x_1)}\left\|v_t^\theta(x)-u_t\left(x\mid x_1\right)\right\|^2$$

with the guarantee

$$\nabla_\theta \mathcal{L}_{FM}(\theta)=\nabla_\theta \mathcal{L}_{CFM}(\theta)$$

All we need is to calculate the closed form of $u_t\left(x\mid x_1\right)$:

Conditional Flow Matching

Thing can be much simpler:

$$\mathcal{L}_{\mathrm{CFM}}(\theta) = \mathbb{E}_{t,q(x_1),p_t(x|x_1)} \left\| v_t^{\theta}(x) - u_t\left(x \mid x_1\right) \right\|^2$$

with the guarantee

$$\nabla_{\theta}\mathcal{L}_{FM}(\theta) = \nabla_{\theta}\mathcal{L}_{CFM}(\theta)$$

All we need is to calculate the closed form of $u_t\left(x \mid x_1\right)$:
Let

$$p_t\left(x \mid x_1\right) = \mathcal{N}\left(x \mid \mu_t\left(x_1\right), \sigma_t\left(x_1\right)^2 I\right),$$

then

$$u_t\left(x \mid x_1\right) = \frac{\sigma_t'\left(x_1\right)}{\sigma_t\left(x_1\right)}\left(x - \mu_t\left(x_1\right)\right) + \mu_t'\left(x_1\right).$$

Example[4]

We can define the $p_t\left(x \mid x_1\right)$ as different form:
Variance Exploding (VE) form:

$$p_t(x)=\mathcal{N}\left(x \mid x_1, \sigma_{1-t}^2 I\right),$$

then we get

$$u_t\left(x \mid x_1\right)=-\frac{\sigma_{1-t}^{\prime}}{\sigma_{1-t}}\left(x-x_1\right).$$

[4] Lipman, et. al. Flow Matching for Generative Modeling. ICLR 2023

Example[4]

We can define the $p_t(x \mid x_1)$ as different form:
Variance Exploding (VE) form:

$$p_t(x) = \mathcal{N}\left(x \mid x_1, \sigma_{1-t}^2 I\right),$$

then we get

$$u_t(x \mid x_1) = -\frac{\sigma_{1-t}'}{\sigma_{1-t}}(x - x_1).$$

Variance Preserving (VP) form

$$p_t(x \mid x_1) = \mathcal{N}\left(x \mid \alpha_{1-t} x_1, \left(1 - \alpha_{1-t}^2\right) I\right),$$

where $\alpha_t = e^{-\frac{1}{2}T(t)}, T(t) = \int_0^t \beta(s) ds$, then we get

$$u_t(x \mid x_1) = \frac{\alpha_{1-t}'}{1 - \alpha_{1-t}^2}(\alpha_{1-t} x - x_1)$$

[4]Lipman, et. al. Flow Matching for Generative Modeling. ICLR 2023

Example

Optimal Transport displacement map form:

$$\mu_t(x) = tx_1, \text{ and } \sigma_t(x) = 1 - (1 - \sigma_{\min})\, t,$$

then we get

$$u_t\left(x \mid x_1\right) = \frac{x_1 - (1 - \sigma_{\min})\, x}{1 - (1 - \sigma_{\min})\, t}$$

	CIFAR-10			ImageNet 32×32			ImageNet 64×64		
Model	NLL↓	FID↓	NFE↓	NLL↓	FID↓	NFE↓	NLL↓	FID↓	NFE↓
Ablations									
DDPM	3.12	7.48	274	3.54	7.19	262	3.32	17.16	264
Score Matching	3.16	19.94	242	3.56	6.21	178	3.40	19.55	441
ScoreFlow	3.09	20.78	428	3.55	15.05	195	3.36	24.18	601
Ours									
FM w/ Diffusion	3.10	8.06	183	3.54	6.65	193	3.33	16.28	187
FM w/ OT	**2.99**	**6.35**	**142**	**3.53**	**5.31**	**122**	**3.31**	**14.04**	**138**

Proof 1: u_t is the vector field that generate p_t

Tool 1: continuity equation

One method of testing if a vector field v_t generates a probability path p_t is the continuity equation. It is a Partial Differential Equation (PDE) providing a necessary and sufficient condition to ensuring that a vector field v_t generates p_t,

$$\frac{d}{dt} p_t(x) + \operatorname{div}\left(p_t(x) v_t(x)\right) = 0$$

where the divergence operator, div, is defined with respect to the spatial variable $x = \left(x^1, \ldots, x^d\right)$, i.e., $\operatorname{div} = \sum_{i=1}^{d} \frac{\partial}{\partial x^i}$.

Proof 1: u_t is the vector field that generate p_t

To verify this, we check that p_t and u_t satisfy the continuity equation (equation 25):

$$
\begin{aligned}
\frac{d}{dt} p_t(x) &= \int \left(\frac{d}{dt} p_t\left(x \mid x_1\right) \right) q\left(x_1\right) dx_1 \\
&= -\int \operatorname{div}\left(u_t\left(x \mid x_1\right) p_t\left(x \mid x_1\right)\right) q\left(x_1\right) dx_1 \\
&= -\operatorname{div}\left(\int u_t\left(x \mid x_1\right) p_t\left(x \mid x_1\right) q\left(x_1\right) dx_1 \right) \\
&= -\operatorname{div}\left(u_t(x) p_t(x)\right)
\end{aligned}
$$

Proof 2: L_{FM} and L_{CFM} are equivalent

objective: $\nabla_\theta \mathbb{E}_{t,p_t(x)} \left\| v_{\theta t(x) - u_t(x)} \right\|^2 = \nabla_\theta \mathbb{E}_{t,q(x_1),p_t(x|x_1)} \left\| v_{\theta t(x) - u_t(x|x_1)} \right\|^2$

$$\|v_t(x) - u_t(x)\|^2 = \|v_t(x)\|^2 - 2\langle v_t(x), u_t(x)\rangle + \|u_t(x)\|^2$$

$$\|v_t(x) - u_t(x \mid x_1)\|^2 = \|v_t(x)\|^2 - 2\langle v_t(x), u_t(x \mid x_1)\rangle + \|u_t(x \mid x_1)\|^2$$

Proof 2: L_{FM} and L_{CFM} are equivalent

objective: $\nabla_\theta \mathbb{E}_{t,p_t(x)} \left\| v_{\theta\, t(x) - u_t(x)} \right\|^2 = \nabla_\theta \mathbb{E}_{t,q(x_1),p_t(x|x_1)} \left\| v_{\theta\, t(x) - u_t(x|x_1)} \right\|^2$

$$\|v_t(x) - u_t(x)\|^2 = \|v_t(x)\|^2 - 2\langle v_t(x), u_t(x)\rangle + \|u_t(x)\|^2$$

$$\|v_t(x) - u_t(x \mid x_1)\|^2 = \|v_t(x)\|^2 - 2\langle v_t(x), u_t(x \mid x_1)\rangle + \|u_t(x \mid x_1)\|^2$$

$$\begin{aligned}\mathbb{E}_{p_t(x)} \|v_t(x)\|^2 &= \int \|v_t(x)\|^2 p_t(x) dx = \int \|v_t(x)\|^2 p_t(x \mid x_1) q(x_1) dx_1 dx \\ &= \mathbb{E}_{q(x_1),p_t(x|x_1)} \|v_t(x)\|^2,\end{aligned}$$

$$\begin{aligned}\mathbb{E}_{p_t(x)} \langle v_t(x), u_t(x)\rangle &= \int \left\langle v_t(x), \frac{\int u_t(x \mid x_1) p_t(x \mid x_1) q(x_1) dx_1}{p_t(x)} \right\rangle p_t(x) dx \\ &= \int \left\langle v_t(x), \int u_t(x \mid x_1) p_t(x \mid x_1) q(x_1) dx_1 \right\rangle dx \\ &= \int \langle v_t(x), u_t(x \mid x_1)\rangle p_t(x \mid x_1) q(x_1) dx_1 dx \\ &= \mathbb{E}_{q(x_1),p_t(x|x_1)} \langle v_t(x), u_t(x \mid x_1)\rangle\end{aligned}$$

Proof 3: The analytic form of conditional vector field

For notational simplicity let $w_t(x) = u_t(x \mid x_1)$. For the following ODE,

$$\frac{d}{dt}\psi_t(x) = w_t(\psi_t(x))$$

Let $x = \psi^{-1}(y)$ and get

$$\psi_t'\left(\psi^{-1}(y)\right) = w_t(y) \tag{1}$$

Inverting $\psi_t(x)$ provides

$$\psi_t^{-1}(y) = \frac{y - \mu_t(x_1)}{\sigma_t(x_1)}$$

Differentiating ψ_t with respect to t gives

$$\psi_t'(x) = \sigma_t'(x_1)\, x + \mu_t'(x_1).$$

Plugging these last two equations in equation (1) we get

$$w_t(y) = \frac{\sigma_t'(x_1)}{\sigma_t(x_1)}\left(y - \mu_t(x_1)\right) + \mu_t'(x_1)$$

Part 3. Basic Knowledge of Schrödinger Bridge

Dynamic Schrödinger Bridge

The dynamic Schrödinger Bridge (SB) problem consists in finding a path measure $\mathbb{P}^{\mathrm{SB}} \in \mathcal{P}(\mathcal{C}) := \mathcal{P}\left(\mathrm{C}\left([0, T], \mathbb{R}^d\right)\right)$ such that

$$\mathbb{P}^{\mathrm{SB}} = \operatorname{argmin}_{\mathbb{P}} \left\{\mathrm{KL}(\mathbb{P} \mid \mathbb{Q}) : \mathbb{P}_0 = \pi_0, \mathbb{P}_T = \pi_T\right\},$$

where $\mathbb{Q} \in \mathcal{P}(\mathcal{C})$ is a reference path measure, defined by the diffusion process $\mathrm{d}\mathbf{X}_t = f_t\left(\mathbf{X}_t\right) \mathrm{d}t + \sigma_t \, \mathrm{d}\mathbf{B}_t, \quad \mathbf{X}_0 \sim \mathbb{Q}_0$.

One form of the solution of Dynamic Schrödinger Bridge

The optimal solution admits the form: $\mathbb{P}^{\mathrm{SB}}=\int \mathbb{Q}_{\mid 0, T} \, \mathrm{d} \Pi_{0, T}^{\mathrm{SB}}$ where $\Pi_{0, T}^{\mathrm{SB}}=\operatorname{argmin}_{\Pi_{0, T}}\left\{\mathrm{KL}\left(\Pi_{0, T} \mid \mathbb{Q}_{0, T}\right): \Pi_0=\pi_0, \Pi_T=\pi_T\right\}$ is the solution of the static SB problem, and $\mathbb{Q}$ is a multivariate Brownian motion with standard deviation σ.

Static Schrödinger Bridge Link with Optimal Transport

The static SB problem can be solved by the entropy-regularized optimal transport

$$\Pi_{0,T}^{\mathrm{SB}} = \operatorname{argmin}_{\Pi_{0,T}} \Big\{ \mathbb{E}_{\Pi_{0,T}} \left[\|\mathbf{X}_0 - \mathbf{X}_T\|^2 - 2\sigma^2 T \mathrm{H}\left(\Pi_{0,T}\right) \right] \\ \Pi_0 = \pi_0, \Pi_T = \pi_T \Big\}$$

where $H(\Pi) = \int \ln \Pi(x, y) d\Pi(dx, dy)$ denotes the entropy, and $\Pi_{0,T}^{\mathrm{SB}}$ is an entropy-regularized OT plan.

Static Schrödinger Bridge Link with Optimal Transport

The static SB problem can also be solved by dynamic optimal transport

$$v_{\mathrm{SB}} = \operatorname{argmin}_v \left\{ \int_0^T \mathbb{E}_{\mathbb{P}_t} \left[\left\| v\left(t, \mathbf{X}_t\right) \right\|^2 \right] \mathrm{d}t \right.$$
$$\left. \mathrm{d}\mathbf{X}_t = v\left(t, \mathbf{X}_t\right) \mathrm{d}t + \sigma \mathrm{d}\mathbf{B}_t, \mathbb{P}_0 = \pi_0, \mathbb{P}_T = \pi_T \right\}.$$

Part 4. Different ways for solving Schrödinger Bridge Problem

Iterative Proportional Fitting

One classical solution is Sinkhorn algorithm:

$$\mathbb{P}^{2n+1} = \mathrm{argmin}_{\mathbb{P}} \left\{ \mathrm{KL}\left(\mathbb{P} \mid \mathbb{P}^{2n}\right) \mathbb{P}_T = \pi_T \right\},$$
$$\mathbb{P}^{2n+2} = \mathrm{argmin}_{\mathbb{P}} \left\{ \mathrm{KL}\left(\mathbb{P} \mid \mathbb{P}^{2n+1}\right) : \mathbb{P}_0 = \pi_0 \right\},$$

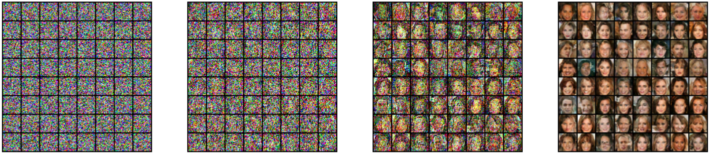
Bortoli et al [5] adopt IPF for solving Schrödinger Bridge by learning a diffusion model.

[5] Bortoli et al. Diffusion Schrödinger Bridge with Applications to Score-Based Generative Modeling. NIPS 2021.

One classical solution is Sinkhorn algorithm:

$$\mathbb{P}^{2n+1} = \operatorname{argmin}_{\mathbb{P}} \left\{ \mathrm{KL} \left(\mathbb{P} \mid \mathbb{P}^{2n} \right) \mathbb{P}_T = \pi_T \right\},$$
$$\mathbb{P}^{2n+2} = \operatorname{argmin}_{\mathbb{P}} \left\{ \mathrm{KL} \left(\mathbb{P} \mid \mathbb{P}^{2n+1} \right) : \mathbb{P}_0 = \pi_0 \right\},$$

Bortoli et al [5] adopt IPF for solving Schrödinger Bridge by learning a diffusion model.



$t = 0$ $t = 0.31$ $t = 0.60$ $t = 0.63$

Figure 4: Generative model for CelebA 32×32 after 10 DSB iterations with $N = 50$ ($T = 0.63$)

[5] Bortoli et al. Diffusion Schrödinger Bridge with Applications to Score-Based Generative Modeling. NIPS 2021.

Bridge

$\mathbb{P}^{\mathrm{SB}}=\int \mathbb{Q}_{|0,T}\,\mathrm{d}\Pi_{0,T}^{\mathrm{SB}}$, where $\mathbb{Q}_{|0,T}$ denotes the bridge that has the reference measure $\mathbb{Q}$ together with fixed initial and terminal point. Let

$$\mathbb{Q}: \mathrm{d}Z_t=b\left(Z_t, t\right)\mathrm{d}t+\sigma\left(Z_t, t\right)\mathrm{d}W_t,$$

By the result from h-transform, $\mathbb{Q}_{|0,T}$ obeys

$$\mathrm{d}Z_t=\left(b\left(Z_t, t\right)+\sigma^2\left(Z_t, t\right)\nabla_z \log q_{T|t}\left(x \mid Z_t\right)\right)\mathrm{d}t+\sigma\left(Z_t, t\right)\mathrm{d}W_t,$$

where $q_{T|t}(x \mid z)$ is the density function of the transition probability $\mathbb{Q}_{T|t}(Z_T=x \mid Z_t=z)$.

Bridge

If the reference measure is a Brownian motion, i.e., $\mathrm{d}Z_t = \sigma_t \, \mathrm{d}W_t$, we have $\mathbb{Q}_{T|t}(\cdot \mid z) = \mathcal{N}\left(z, \beta_T - \beta_t\right)$, where $\beta_t = \int_0^t \sigma_s^2 \, \mathrm{d}$, then $\mathbb{Q}_{|0,T}$ obeys[6]

$$\mathrm{d}Z_t = \sigma_t^2 \frac{x - Z_t}{\beta_T - \beta_t} \, \mathrm{d}t + \sigma_t \, \mathrm{d}W_t$$

Note that the drift grows to infinity in magnitude with a rate of $O\left(1/\left(\beta_T - \beta_t\right)\right)$ as $t \to T$, which ensures that $Z_t = x$ with probability one.

[6] Liu et. al. Learning Diffusion Model Bridges on Constrained Domains. ICLR 2023.

Bridge

For the general reference measure, it hard to get the analytic form of the corresponding bridge since the analytic form of the transition density is the solution of the backward Kolmogorov equation, which is hard to solve in high dimensional case.

Bortoli et. al.[7] propose to use score matching to learn the drift term of the bridge.

[7] Bortoli et. al. Simulating Diffusion Bridges with Score Matching. ArXiv, abs/2111.07243.

Forward-backward SDE[8]

Let $\Psi(t, \boldsymbol{x})$ and $\widehat{\Psi}(t, \boldsymbol{x})$ be the solutions to the following PDEs:

$$\begin{cases} \frac{\partial \Psi}{\partial t} = -\nabla_{\boldsymbol{x}} \Psi^{\top} f - \frac{1}{2} \operatorname{Tr}\left(g^2 \nabla_{\boldsymbol{x}}^2 \Psi\right) & \text{s.t. } \Psi(0, \cdot) \widehat{\Psi}(0, \cdot) = p_{\text{data}} \\ \frac{\partial \widehat{\Psi}}{\partial t} = -\nabla_{\boldsymbol{x}} \cdot (\widehat{\Psi} f) + \frac{1}{2} \operatorname{Tr}\left(g^2 \nabla_{\boldsymbol{x}}^2 \widehat{\Psi}\right) & \quad\;\; \Psi(T, \cdot) \widehat{\Psi}(T, \cdot) = p_{\text{prior}} \end{cases}$$

Then, the solution to the dynamic Schrödinger Bridge problem can be expressed by the following SDE:

$$\mathrm{d}\mathbf{X}_t = \left[f + g^2 \nabla_{\boldsymbol{x}} \log \Psi\left(t, \mathbf{X}_t\right)\right] \mathrm{d}t + g \, \mathrm{d}\mathbf{W}_t, \quad \mathbf{X}_0 \sim p_{\text{data}} ,$$

$$\mathrm{d}\mathbf{X}_t = \left[f - g^2 \nabla_{\boldsymbol{x}} \log \widehat{\Psi}\left(t, \mathbf{X}_t\right)\right] \mathrm{d}t + g \, \mathrm{d}\mathbf{W}_t, \quad \mathbf{X}_T \sim p_{\text{prior}} ,$$

where $\nabla_{\boldsymbol{x}} \log \Psi\left(t, \mathbf{X}_t\right)$ and $\nabla_{\boldsymbol{x}} \log \widehat{\Psi}\left(t, \mathbf{X}_t\right)$ are the optimal forward and backward drifts for SB. Also, the marginal density obeys $p_t^{\mathrm{SB}}\left(\mathbf{X}_t\right) = \Psi\left(t, \mathbf{X}_t\right) \widehat{\Psi}\left(t, \mathbf{X}_t\right)$.

[8] Chen et. al. Likelihood Training of Schrödinger Bridge using Forward-Backward SDEs Theory. ICLR 2022.