Schrödinger Bridges

Charlie

Geometric AI Lab
KAIST

Weekly Meeting, 2023

Outline

Introduction

Introduction

This presentation will not present a single paper but multiple related papers. The goal of this presentation is to present a different perspective on Diffusion Models. through Optimal Transport with a focus on this paper:

- Likelihood training of SB using FBSDE Chen et al. (2021)

Background

Diffusion Model SDE (SGM)

Following Anderson (1982), we have: Given a Forward-time diffusion:

$$d\mathbf{X}_t = f(t, \mathbf{X}_t)dt + g(t)d\mathbf{W}_t, \quad \mathbf{X}_0 \sim p_{\text{data}} \tag{1}$$

the Reverse-time Diffusion has the form:

$$d\mathbf{X}_t = [f(t, \mathbf{X}_t)dt - g(t)^2 \nabla_x \log p_t(\mathbf{X}_t)]dt + g(t)d\mathbf{W}_t, \quad \mathbf{X}_T \sim p_{\text{prior}} \tag{2}$$

Weakness of SGM formulation

SGM formulation requires the drift term $f(t, \mathbf{X}_t)$ to be simple for p_{0t} to be closed form.
SGM requires a large number of time steps for prior to be Gaussian.

Weakness of SGM formulation

SGM formulation requires the drift term $f(t, \mathbf{X}_t)$ to be simple for p_{0t} to be closed form.
SGM requires a large number of time steps for prior to be Gaussian.
What if we require the final distribution to be Gaussian and the initial

distribution to be our data distribution and just find the path probability that matches that of SGM?

Schrödinger Bridge (SB) Problem

SB Problem is defined as an optimization problem with the following objective:

$$\min_{\mathbb{Q} \in \mathcal{P}(p_{\text{data}}, p_{\text{prior}})} D_{KL}(\mathbb{Q}||\mathbb{P}) \tag{3}$$

Here, $\mathcal{P}$ is the set of all possible path measures and $\mathbb{P}$ is the path measure arise from process in Eq. 1

Solving Schrödinger Bridge with Log-likelihood Objective

Schrödinger Bridge

Theorem 1 Let $\psi(t, \mathbf{x})$ and $\hat{\psi}(t, \mathbf{x})$ be the solution to the following PDE:

$$\frac{\partial \psi}{\partial t} = -f \cdot \nabla_x \psi - \frac{1}{2} Tr(g^2 \nabla_x^2 \psi) \tag{4}$$

$$\frac{\partial \hat{\psi}}{\partial t} = -\nabla_x \cdot (f \hat{\psi}) + \frac{1}{2} Tr(g^2 \nabla_x^2 \psi) \tag{5}$$

Where $\psi(0, \cdot)\hat{\psi}(0, \cdot) = p_{\text{data}}$ and $\psi(T, \cdot)\hat{\psi}(T, \cdot) = p_{\text{prior}}$ Then, the solution $\mathbb{Q}$ to Schrödinger Bridge Problem in Eq. **??** is also the path measure of these processes:

$$d\mathbf{X}_t = [f + g^2 \nabla_x \log \psi] dt + g d\mathbf{W}_t, \quad \mathbf{X}_0 \sim p_{\text{data}} \tag{6}$$

$$d\mathbf{X}_t = [f - g^2 \nabla_x \log \hat{\psi}] dt + g d\mathbf{W}_t, \quad \mathbf{X}_T \sim p_{\text{prior}} \tag{7}$$

Connection between Schrödinger Bridge and SDE

SDE Diffusion

$$dX_t = f dt + g dW_t, \quad X_0 \sim p_{\text{data}} \tag{8}$$
$$dX_t = [f - g^2 \nabla_x \log p_t(X_t)]dt + g dW_t, \quad X_T \sim p_{\text{prior}} \tag{9}$$

Schrödinger Bridge solution

$$dX_t = [f + g^2 \nabla_x \log \psi_t(X_t)]dt + g dW_t, \quad X_0 \sim p_{\text{data}} \tag{10}$$
$$dX_t = [f - g^2 \nabla_x \log \hat{\psi}_t(X_t)]dt + g dW_t, \quad X_T \sim p_{\text{prior}} \tag{11}$$

Connection between Schrödinger Bridge and SDE

SDE Diffusion

$$
\begin{aligned}
d\mathbf{X}_t &= f dt &&+ g d\mathbf{W}_t, \quad \mathbf{X}_0 \sim p_{\text{data}} && (12)\\
d\mathbf{X}_t &= [f - g^2 \nabla_x \log p_t(\mathbf{X}_t)] dt &&+ g d\mathbf{W}_t, \quad \mathbf{X}_T \sim p_{\text{prior}} && (13)
\end{aligned}
$$

Schrödinger Bridge solution

$$
\begin{aligned}
d\mathbf{X}_t &= [f + g^2 \nabla_x \log \psi_t(\mathbf{X}_t)] dt + g d\mathbf{W}_t, \quad \mathbf{X}_0 \sim p_{\text{data}} && (14)\\
d\mathbf{X}_t &= [f - g^2 \nabla_x \log \hat{\psi}_t(\mathbf{X}_t)] dt + g d\mathbf{W}_t, \quad \mathbf{X}_T \sim p_{\text{prior}} && (15)
\end{aligned}
$$

All of these equations belong to a class of SDE called *control-affine* SDE

$$\mathbf{X}_t = \mathbf{A}(t, \mathbf{X}_t) dt + \mathbf{B}(t, \mathbf{X}_t) \mathbf{u}(t, \mathbf{X}_t) dt + \mathbf{C}(t) d\mathbf{W}_t$$

with the same $(\mathbf{A}, \mathbf{B}, \mathbf{C}) = (f, \mathbf{I}, g)$ and different *control* variable $\mathbf{u}(t, \mathbf{X}_t)$

Solving Schrödinger Bridge

Theorem 2 (Forward-backward SDE (FBSDE) to SB optimality) Given the following SDEs, (let $\mathbf{u}(t) = \mathbf{Z}_t$)

$$\begin{aligned} d\mathbf{X}_t &= (f + g\mathbf{Z}_t)dt + g d\mathbf{W}_t \\ d\mathbf{Y}_t &= \frac{1}{2}\mathbf{Z}_t^T\mathbf{Z}_t dt + \mathbf{Z}_t^T d\mathbf{W}_t \\ d\hat{\mathbf{Y}}_t &= \left(\frac{1}{2}\hat{\mathbf{Z}}_t^T\hat{\mathbf{Z}}_t dt + \nabla_x \cdot \left(g\hat{\mathbf{Z}}_t - f\right) + \hat{\mathbf{Z}}_t^T\mathbf{Z}_t\right) dt + \hat{\mathbf{Z}}_t d\mathbf{W}_t \end{aligned} \tag{16}$$

Where $\mathbf{X}(0) = \mathbf{x}_0$ is given and $\mathbf{Y}_T + \hat{\mathbf{Y}}_T = \log p_{\text{prior}}(\mathbf{X}_T)$, then we have

$$\mathbf{Y}_t = \log \psi(t, \mathbf{X}_t) \quad \mathbf{Z}_t = g\nabla_x \log \psi(t, \mathbf{X}_t), \tag{17}$$

$$\hat{\mathbf{Y}}_t = \log \hat{\psi}(t, \mathbf{X}_t) \quad \hat{\mathbf{Z}}_t = g\nabla_x \log \hat{\psi}(t, \mathbf{X}_t), \tag{18}$$

And most importantly, $\mathbf{Y}_t + \hat{\mathbf{Y}}_t = \log p_t^{SB}(\mathbf{X}_t)$.

Solving Schrödinger Bridge

Theorem 3 We can compute the loglikelihood of FBSDE in Eq. 16

$$
\begin{aligned}
\mathcal{L}_{SB}(\mathbf{x}_0) = \log p_0^{SB}(\mathbf{x}_0) &= \mathbb{E}[\log p_{\text{prior}}(\mathbf{X}_T)] \\
&- \int_0^T \mathbb{E}\left[\frac{1}{2}||\mathbf{Z}_t||^2 + \frac{1}{2}||\hat{\mathbf{Z}}_t||^2 + \nabla_x \cdot (g\hat{\mathbf{Z}}_t - f) + \hat{\mathbf{Z}}_t^T \mathbf{Z}_t\right] dt
\end{aligned}
$$

In practice, we can train with simpler alternating objectives:

$$
\tilde{\mathcal{L}}_{SB}(\mathbf{x}_0; \phi) = -\int_0^T \mathbb{E}_{\mathbf{X}_t \sim Eq11}\left[\frac{1}{2}||\hat{\mathbf{Z}}_\phi(t, \mathbf{X}_t)||^2 + \frac{1}{2}\nabla_x \cdot \hat{\mathbf{Z}}_\phi(t, \mathbf{X}_t) + \mathbf{Z}_t^T \hat{\mathbf{Z}}_\phi(t, \mathbf{X}_t)\right] \tag{19}
$$

$$
\tilde{\mathcal{L}}_{SB}(\mathbf{x}_T; \theta) = -\int_0^T \mathbb{E}_{\mathbf{X}_t \sim Eq11}\left[\frac{1}{2}||\mathbf{Z}_\theta(t, \mathbf{X}_t)||^2 + \frac{1}{2}\nabla_x \cdot \mathbf{Z}_\theta(t, \mathbf{X}_t) + \hat{\mathbf{Z}}_t^T \mathbf{Z}_\theta(t, \mathbf{X}_t)\right] \tag{20}
$$

Training

Algorithm 1 Likelihood training of SB-FBSDE

Input: boundary distributions p_{data} and p_{prior}, parameterized policies $\mathbf{Z}(\cdot,\cdot;\theta)$ and $\widehat{\mathbf{Z}}(\cdot,\cdot;\phi)$
repeat
  if memory resource is affordable **then**
    run Algorithm 2.
  else
    run Algorithm 3.
  end if
until converges

Algorithm 2 Joint (diffusion flow-based) training

for $k = 1$ **to** K **do**
  Sample $\mathbf{X}_{t\in[0,T]}$ from (13a) where $\boldsymbol{x}_0 \sim p_{\text{data}}$ (computational graph retained).
  Compute $\mathcal{L}_{\text{SB}}(\boldsymbol{x}_0;\theta,\phi)$ with (16).
  Update (θ,ϕ) with $\nabla_{\theta,\phi}\mathcal{L}_{\text{SB}}(\boldsymbol{x}_0;\theta,\phi)$.
end for

Algorithm 3 Alternate (IPF-based) training

Input: Caching frequency M
for $k = 1$ **to** K **do**
  if $k\%M == 0$ **then**
    Sample $\mathbf{X}_{t\in[0,T]}$ from (13a) where $\boldsymbol{x}_0 \sim p_{\text{data}}$ (computational graph discarded).
  end if
  Compute $\widetilde{\mathcal{L}}_{\text{SB}}(\boldsymbol{x}_0;\phi)$ with (18).
  Update ϕ with gradient $\nabla_\phi\widetilde{\mathcal{L}}_{\text{SB}}(\boldsymbol{x}_0;\phi)$.
end for
for $k = 1$ **to** K **do**
  if $k\%M == 0$ **then**
    Sample $\mathbf{X}_{t\in[0,T]}$ from (7b) where $\boldsymbol{x}_T \sim p_{\text{prior}}$ (computational graph discarded).
  end if
  Compute $\mathcal{L}_{\text{SB}}(\boldsymbol{x}_T;\theta)$ with (19).
  Update θ with gradient $\nabla_\theta\widetilde{\mathcal{L}}_{\text{SB}}(\boldsymbol{x}_T;\theta)$.
end for

Sampling

Given a predefined variance scheduling at time step t and corrector step i:

$$\sigma_{t,i} = \frac{2r^2 g^2 ||\epsilon_i||^2}{||\mathbf{Z}(t, \mathbf{X}_{t,i}) + \hat{\mathbf{Z}}(t, \mathbf{X}_{t,i})||^2}$$

Algorithm 4 Generative Process of SB-FBSDE

Input: p_{prior}, policies $\mathbf{Z}(\cdot, \cdot; \theta)$ and $\widehat{\mathbf{Z}}(\cdot, \cdot; \phi)$
Sample $\mathbf{X}_T \sim p_{\text{prior}}$.
for $t = T$ **to** Δt **do**
  Sample $\epsilon \sim \mathcal{N}(\mathbf{0}, \boldsymbol{I})$.
  Predict $\mathbf{X}_{t,1} \leftarrow \mathbf{X}_t + g\, \widehat{\mathbf{Z}}_t \Delta t + \sqrt{g \Delta t}\, \epsilon$.
  for $i = 1$ **to** N **do**
    Sample $\epsilon_i \sim \mathcal{N}(\mathbf{0}, \boldsymbol{I})$.
    Compute $\nabla_{\boldsymbol{x}} \log p^{\text{SB}}_{t,i} \approx [\mathbf{Z}(t, \mathbf{X}_{t,i}) + \widehat{\mathbf{Z}}(t, \mathbf{X}_{t,i})]/g$.
    Compute $\sigma_{t,i}$ with (59).
    Correct $\mathbf{X}_{t,i+1} \leftarrow \mathbf{X}_{t,i} + \sigma_{t,i} \nabla_{\boldsymbol{x}} \log p^{\text{SB}}_{t,i} + \sqrt{2\sigma_{t,i}}\, \epsilon_i$.
  end for
  Propagate $\mathbf{X}_{t-\Delta t} \leftarrow \mathbf{X}_{t,N}$.
end for
return $\mathbf{X}_0$

SGM vs. FBSDE

When $\left(\mathbf{Z}_t, \hat{\mathbf{Z}}_t\right) := (0, g\mathbf{s}_t)$, FBSDE collapse to SGM and

$$\mathcal{L}_{SB} = \mathcal{L}_{SGM}$$

Probability flow of SB

we have the following ODE that characterize the optimal process of SB:

$$d\mathbf{X}_t = \left[f + g\mathbf{Z}(t, \mathbf{X}_t) - \frac{1}{2}g(\mathbf{Z}(t, \mathbf{X}_t) + \hat{\mathbf{Z}}(t, \mathbf{X}_t)\right]$$

Main Takeaways

Main Takeaways

We redefine the Diffusion Process in terms of a SB problem with the forward and backward equation:

$$
\begin{aligned}
d\mathbf{X}_t &= [f + g\mathbf{Z}_t]dt + g d\mathbf{W}_t, \quad \mathbf{X}_0 \sim p_{\text{data}} \\
d\mathbf{X}_t &= [f - g\hat{\mathbf{Z}}_t]dt + g d\mathbf{W}_t, \quad \mathbf{X}_T \sim p_{\text{prior}}
\end{aligned}
$$

Where $\mathbf{Z}_t$ and $\hat{\mathbf{Z}}_t$ can be learned with a maximum likelihood objective:

$$
\begin{aligned}
\mathcal{L}_{SB}(\mathbf{x}_0) = \log p_0^{SB}(\mathbf{x}_0) = \mathbb{E}[\log p_{\text{prior}}(\mathbf{X}_T)] \\
- \int_0^T \mathbb{E}\left[\frac{1}{2}||\mathbf{Z}_t||^2 + \frac{1}{2}||\hat{\mathbf{Z}}_t||^2 + \nabla_x \cdot (g\hat{\mathbf{Z}}_t - f) + \hat{\mathbf{Z}}_t^T \mathbf{Z}_t\right] dt
\end{aligned}
$$

Thank You!

References

Chen, T., Liu, G.-H., and Theodorou, E. A. (2021). Likelihood training of schrödinger bridge using forward-backward sdes theory. *arXiv preprint arXiv:2110.11291*.