Diffusion Schrödinger Bridge

29/06/2023

大纲

- Motivation
- Methodology
- Results
- Main onclusions

Motivation

背景：

- 计算量
- 找到粒子在两个不同时间的分布下最可能的路径。

Methodology

> 🔰 Trivial

- 🔰 Schrödinger Bridges （模型）
- 🔰 Score-Based Diffusions （模型）
- 🔰 Iterative Proportional Fitting（算法）
- ✅ Diffusion Schrödinger Bridge （算法）

Methodology - Schrödinger Bridges

Schrödinger bridge 和 Static Schrödinger bridge

普通的Schrödinger Bridges

$$\pi^{\star} = \arg\min \left\{ \mathrm{KL}(\pi \mid p) : \pi \in \mathscr{P}_{N+1}, \pi_0 = p_{\text{data}}\,, \pi_N = p_{\text{prior}} \right\}$$

Methodology - Schrödinger Bridges

Static Schrödinger Bridge

$$\pi^{\mathrm{s},\star} = \arg\min\left\{\mathrm{KL}\left(\pi^{\mathrm{s}} \mid p_{0,N}\right) : \pi^{\mathrm{s}} \in \mathscr{P}_2, \pi_0^{\mathrm{s}} = p_{\mathrm{data}}\,, \pi_N^{\mathrm{s}} = p_{\mathrm{prior}}\,\right\}$$

π与π^{s}的主要差别在于：

- π 是 $N+1$ 维 （关注了从$0\cdots N$时刻）
- π^{s} 是 2 维 （只关注 $0, N$时刻）

Methodology - Schrödinger Bridges

Link with optimal transport

Original:

$$\pi^{\mathrm{s},\star} = \arg\min \left\{\mathrm{KL}\left(\pi^{\mathrm{s}} \mid p_{0,N}\right) : \pi^{\mathrm{s}} \in \mathscr{P}_2, \pi_0^{\mathrm{s}} = p_{\mathrm{data}}, \pi_N^{\mathrm{s}} = p_{\mathrm{prior}} \right\}$$

$$\pi^{\mathrm{s},\star} = \arg\min \left\{\mathrm{H}(\pi^{s}, p_{0,N}) - \mathrm{H}(\pi^{s}) : \pi^{\mathrm{s}} \in \mathscr{P}_2, \pi_0^{\mathrm{s}} = p_{\mathrm{data}}, \pi_N^{\mathrm{s}} = p_{\mathrm{prior}} \right\}$$

$$\pi^{\mathrm{s},\star} = \arg\min \left\{-\mathbb{E}_{\pi^{s}}[p_{0,N}] - \mathrm{H}(\pi^{s}) : \pi^{\mathrm{s}} \in \mathscr{P}_2, \pi_0^{\mathrm{s}} = p_{\mathrm{data}}, \pi_N^{\mathrm{s}} = p_{\mathrm{prior}} \right\}$$

$$(\text{How?})\pi^{\mathrm{s},\star} = \arg\min \left\{-\mathbb{E}_{\pi^{\mathrm{s}}}\left[\log p_{N|0}\left(X_N \mid X_0\right)\right] - \mathrm{H}\left(\pi^{\mathrm{s}}\right) : \pi^{\mathrm{s}} \in \mathscr{P}_2, \pi_0^{\mathrm{s}} = p_{\mathrm{data}}, \pi_N^{\mathrm{s}} = p_{\mathrm{prior}} \right\}$$

Methodology - Iterative Proportional Fitting (IPF)

这里的记号比较奇怪，和之前不太一样。上标代表的是当前的迭代次数，也就是说迭代n次代表n个“来回”：

$$
\begin{aligned}
\pi^{2n+1} &= \arg\min\left\{\mathrm{KL}\left(\pi \mid \pi^{2n}\right) : \pi \in \mathscr{P}_{N+1}, \pi_N = p_{\text{prior}}\right\} \\
\pi^{2n+2} &= \arg\min\left\{\mathrm{KL}\left(\pi \mid \pi^{2n+1}\right) : \pi \in \mathscr{P}_{N+1}, \pi_0 = p_{\text{data}}\right\}
\end{aligned}
$$

已知：

$$
p^0\left(x_{0:N}\right) = p\left(x_{0:N}\right)
$$

向后更新：

$$
q^n\left(x_{0:N}\right) = p_{\text{prior}}\left(x_N\right)\prod_{k=0}^{N-1} p_{k|k+1}^n\left(x_k \mid x_{k+1}\right), \text{with } p_{k|k+1}^n\left(x_k \mid x_{k+1}\right) = \frac{p_{k+1|k}^n\left(x_{k+1} \mid x_k\right) p_k^n\left(x_k\right)}{p_{k+1}^n\left(x_{k+1}\right)}
$$

向前更新：

$$
p^{n+1}\left(x_{0:N}\right) = p_{\text{data}}\left(x_0\right)\prod_{k=0}^{N-1} q_{k+1|k}^n\left(x_{k+1} \mid x_k\right), \text{with } q_{k+1|k}^n\left(x_{k+1} \mid x_k\right) = \frac{q_{k|k+1}^n\left(x_k \mid x_{k+1}\right) q_{k+1}^n\left(x_{k+1}\right)}{q_k^n\left(x_k\right)}
$$

(Before)Methodology - Diffusion Schrödinger Bridge

有:

$$\begin{aligned} p_{k|k+1}\left(x_k \mid x_{k+1}\right) &= \frac{p_k\left(x_k\right) p_{k+1|k}\left(x_{k+1} \mid x_k\right)}{p_{k+1}\left(x_{k+1}\right)} \\ &= p_{k+1|k}\left(x_{k+1} \mid x_k\right) \frac{p_k\left(x_k\right)}{p_{k+1}\left(x_{k+1}\right)} \\ &= p_{k+1|k}\left(x_{k+1} \mid x_k\right) \exp\left[\log p_k\left(x_k\right) - \log p_{k+1}\left(x_{k+1}\right)\right] \end{aligned} \tag{1}$$

和

$$\begin{aligned} p_{k+1|k}\left(x_{k+1} \mid x_k\right) &= \mathcal{N}\left(x_{k+1}; x_k + \gamma_{k+1} f\left(x_k\right), 2\gamma_{k+1}\mathbf{I}\right) \\ &= \frac{1}{\sqrt{4\pi\gamma_{k+1}}} \exp\left(-\frac{\left(x_{k+1} - x_k - \gamma_{k+1} f\left(x_k\right)\right)^2}{4\gamma_{k+1}}\right) \end{aligned} \tag{2}$$

(Before)Methodology - Diffusion Schrödinger Bridge

使用泰勒展开：我们对 $\log p_{k+1}$ 在 $x_k \approx x_{k+1}$ 处进行泰勒展开。泰勒展开是:

$$\log p_{k+1}\left(x_{k+1}\right) \approx \log p_{k+1}\left(x_k\right) + \frac{x_{k+1} - x_k}{1!} \nabla \log p_{k+1}\left(x_k\right)$$

代入泰勒展开: 将泰勒展开代入原始公式后并化简:

$$\begin{aligned}
p_{k|k+1}\left(x_k \mid x_{k+1}\right) &\approx p_{k+1|k}\left(x_{k+1} \mid x_k\right) \exp\left[\log p_k\left(x_k\right) - \log p_{k+1}\left(x_k\right) - \left(x_{k+1} - x_k\right) \nabla \log p_{k+1}\left(x_k\right)\right] \\
&= p_{k+1|k}\left(x_{k+1} \mid x_k\right) \exp\left[-2\left(x_{k+1} - x_k\right) \nabla \log p_{k+1}\left(x_k\right)\right] \\
&\vdots \quad \text{算不动了} \\
&= \mathcal{N}\left(x_k; x_{k+1} - \gamma_{k+1} f\left(x_{k+1}\right) + 2\gamma_{k+1} \nabla \log p_{k+1}\left(x_{k+1}\right), 2\gamma_{k+1} \mathbf{I}\right)
\end{aligned}$$

Methodology - Diffusion Schrödinger Bridge

对于第n次迭代，首先向前：

$$p_{k+1|k}^{n}\left(x_{k+1} \mid x_{k}\right)=\mathcal{N}\left(x_{k+1} ; x_{k}+\gamma_{k+1} f_{k}^{n}\left(x_{k}\right), 2 \gamma_{k+1} \mathbf{I}\right)$$
$$\text{with, } p^{0}=p \text{ and } f_{k}^{0}=f$$

然后，向后：

$$\begin{aligned} q_{k|k+1}^{n}\left(x_{k} \mid x_{k+1}\right) &= p_{k+1|k}^{n}\left(x_{k+1} \mid x_{k}\right) \exp \left[\log p_{k}^{n}\left(x_{k}\right)-\log p_{k+1}^{n}\left(x_{k+1}\right)\right] \\ &\approx \mathcal{N}\left(x_{k} ; x_{k+1}+\gamma_{k+1} b_{k+1}^{n}\left(x_{k+1}\right), 2 \gamma_{k+1} \mathbf{I}\right), \end{aligned}$$
$$\text{where } b_{k+1}^{n}\left(x_{k+1}\right)=-f_{k}^{n}\left(x_{k+1}\right)+2 \nabla \log p_{k+1}^{n}\left(x_{k+1}\right)$$

Methodology - Diffusion Schrödinger Bridge

现在我们要优化的目标就是:

$$q_{k|k+1}^{n}\left(x_{k} \mid x_{k+1}\right)=\mathcal{N}\left(x_{k}; B_{k+1}^{n}\left(x_{k+1}\right), 2\gamma_{k+1}\mathbf{I}\right), p_{k+1|k}^{n}\left(x_{k+1} \mid x_{k}\right)=\mathcal{N}\left(x_{k+1}; F_{k}^{n}\left(x_{k}\right), 2\gamma_{k+1}\mathbf{I}\right)$$

因此:

$$B_{k+1}^{n}=\arg\min_{\mathrm{B}\in\mathrm{L}^{2}(\mathbb{R}^{d},\mathbb{R}^{d})}\mathbb{E}_{p_{k,k+1}^{n}}\left[\left\|\mathrm{B}\left(X_{k+1}\right)-\left(X_{k+1}+F_{k}^{n}\left(X_{k}\right)-F_{k}^{n}\left(X_{k+1}\right)\right)\right\|^{2}\right]$$

$$F_{k}^{n+1}=\arg\min_{\mathrm{F}\in\mathrm{L}^{2}(\mathbb{R}^{d},\mathbb{R}^{d})}\mathbb{E}_{q_{k,k+1}^{n}}\left[\left\|\mathrm{F}\left(X_{k}\right)-\left(X_{k}+B_{k+1}^{n}\left(X_{k+1}\right)-B_{k+1}^{n}\left(X_{k}\right)\right)\right\|^{2}\right]$$

Methodology - Diffusion Schrödinger Bridge

每个iteration需要计算两次

Schrödinger Bridge (DSB) and is summarized in Algorithm
for an illustration.

Algorithm 1 Diffusion Schrödinger Bridge

for $n \in \{0, \dots, L\}$ **do**
  while not converged **do**
    Sample $\{X_k^j\}_{k,j=0}^{N,M}$, where $X_0^j \sim p_{\text{data}}$, and
    $X_{k+1}^j = F_{\alpha^n}(k, X_k^j) + \sqrt{2\gamma_{k+1}} Z_{k+1}^j$
    Compute $\hat{\ell}_n^b(\beta^n)$ approximating (12)
    $\beta^n \leftarrow$ Gradient Step$(\hat{\ell}_n^b(\beta^n))$
  end while
  while not converged **do**
    Sample $\{X_k^j\}_{k,j=0}^{N,M}$, where $X_N^j \sim p_{\text{prior}}$, and
    $X_{k-1}^j = B_{\beta^n}(k, X_k^j) + \sqrt{2\gamma_k} \tilde{Z}_k^j$
    Compute $\hat{\ell}_{n+1}^f(\alpha^{n+1})$ approximating (13)
    $\alpha^{n+1} \leftarrow$ Gradient Step$(\hat{\ell}_{n+1}^f(\alpha^{n+1}))$
  end while
end for
Output: (α^{L+1}, β^L)

3.4 Convergence of Iterative Proportional Fitting

Train

Loss function

$$\hat{\ell}^{b}_{n,I}(\beta) = M^{-1} \sum_{(k,y)\in I} \left\| B_{\beta}\left(k+1, X^{j}_{k+1}\right) - \left(X^{j}_{k+1} + F^{n}_{k}\left(X^{j}_{k+1}\right) - F^{n}_{k}\left(X^{j}_{k}\right)\right) \right\|^{2}$$

$$\hat{\ell}^{j}_{n+1,I}(\alpha) = M^{-1} \sum_{(k,y)\in I} \left\| F_{\alpha}\left(k, X^{j}_{k}\right) - \left(X^{j}_{k} + B^{n}_{k+1}\left(X^{j}_{k+1}\right) - B^{n}_{k+1}\left(X^{j}_{k}\right)\right) \right\|^{2}$$