Diffusion Schrödinger Bridge with Applications to Score-Based Generative Modeling

Mamadou DIOUF, Abdoul Wahide MAMA

Université Gustave Eiffel

04-07-2025

Introduction

- A generative model is a machine learning model designed to create new data that is similar to its training data.
- A diffusion model is kind of generating modeling technique that takes inspiration from physics(non-equilibrium statistical physics and stochastic differential equations to be more exact).
- The main idea is to convert a well-known and simple base distribution (like gaussian) to the target(data) distribution iteratively, with small step sizes, via a Markov chain.

Introduction

Schrödinger Brigde

- Generative modeling as a Schrödinger bridge problem is a famous entropy regularized Optimal Transport (OT) problem introduced by Schrödinger (1932). Given a reference diffusion with finite time horizon T, a data distribution and a prior distribution, solving the SB amounts to finding the closest diffusion to the reference (in terms of Kullback–Leibler divergence on path spaces) which admits the data distribution as marginal at time $t = 0$ and the prior at time $t = T$.
- The reverse-time diffusion solving this SB problem provides a new Score-Based Generative Modeling (SGM) algorithm which enables approximate sample generation from the data distribution using shorter time intervals compared to the original SGM methods.

Discrete time diffusion and score based

- Consider a data distribution with positive density p_{data}, a positive prior density p_{prior} on $\mathbb{R}^d$
- And a Markov chain with initial density $p_0 = p_{data}$ on $\mathbb{R}^d$ evolving according to transition densities $p_{k+1|k}$.
- Hence for any $x_{0:N} = \{x_k\}_{k=0}^N \in \mathcal{X} = (\mathbb{R}^d)^{N+1}$, the joint density may be expressed as

$$p(x_{0:N}) = p_0(x_0) \prod_{k=0}^{N-1} p_{k+1|k}(x_{k+1}|x_k)$$

This joint density also admits the backward decomposition

$$p(x_{0:N}) = p_N(x_N) \prod_{k=0}^{N-1} p_{k|k+1}(x_k|x_{k+1})$$

where

$$p_{k|k+1}(x_k|x_{k+1}) = \frac{p_k(x_k)\, p_{k+1|k}(x_{k+1}|x_k)}{p_{k+1}(x_{k+1})}$$

Discrete time diffusion and score based

- For the purpose of generative modeling, we will choose transition densities such that $p_N(x_N) = p(x_{0:N})dx_{0:N-1} \approx p_{prior}(x_N)$ for large N.
- One can (approximately) sample from p_{data} by first sampling from p_{prior} followed by $X_k \sim p_{k|k+1}(\cdot|X_{k+1})$ for $k \in [\![0, N-1]\!]$
- The reverse-time transitions may be approximated if we consider a forward transition density of the form

$$p_{k+1|k}(x_{k+1}|x_k) = \mathcal{N}(x_{k+1}; x_k + \gamma_{k+1} f(x_k), 2\gamma_{k+1}\mathbf{I})$$

  with drift $f : \mathbb{R}^d \to \mathbb{R}^d$ and stepsize $\gamma_{k+1} > 0$.
- We can approximate $p_{k|k+1}(x_k|x_{k+1}) \approx \mathcal{N}(x_k; x_{k+1} - \gamma_{k+1} f(x_{k+1}) + 2\gamma_{k+1}\nabla log p_{k+1}(x_{k+1}), 2\gamma_{k+1}\mathbf{I})$

Discrete time diffusion and score based

- With $\nabla logp_{k+1}(x_{k+1}) = \mathbb{E}_{p_{0|k+1}}[\nabla_{x_{k+1}} logp_{k+1|0}(x_{k+1}|X_0)]$
- We can therefore formulate score estimation as a regression problem and use a flexible class of functions, $e.g.$ neural networks, to parametrize an approximation $s_{\theta^*}(k, x_k) \approx \nabla logp_k(x_k)$ such that

$$\theta^* = \arg\min_{\theta} \sum_{k=1}^{N} \mathbb{E}_{p_{0,k}}[||s_\theta(k, X_k) - \nabla_{x_k} log_{p_{k|0}}(X_k|X_0)||^2],$$

,

- If $p_{k|0}$ is not available, we use θ^* s.t.

$$\theta^* = \arg\min_{\theta} \sum_{k=1}^{N} \mathbb{E}_{p_{0,k}}[||s_\theta(k, X_k) - \nabla_{x_k} log_{p_{k-1|k}}(X_k|X_{k-1})||^2],$$

Discrete time diffusion and score based

- In summary, Scored-Based Generative Modeling (SGM) involves first estimating the score function s_{θ^*} from noisy data, and then sampling X_0 using $X_N \sim p_{prior}$ and the approximation

$$X_k = X_{k+1} - \gamma_{k+1} f(X_{k+1}) + 2\gamma_{k+1} \nabla log p_{k+1}(X_{k+1}) + \sqrt{2\gamma_{k+1}} Z_{k+1}$$

  $Z_{k+1} \sim \mathcal{N}(0, \mathbf{I})$ iid.
- The random variable X_0 is approximately $p_0 = p_{data}$ distributed if $p_N(x_N) \approx p_{prior}(x_N)$.
- In what follows, we let $\{Y_k\}_{k=0}^N = \{X_{N-k}\}_{k=0}^N$ and remark that $\{Y_k\}_{k=0}^N$ satisfies a forward recursion.

Continuous time diffusion and score based

- The Markov chain with kernel corresponds to an Euler–Maruyama discretization of $(\mathbf{X}_t)_{t\in[0,T]}$ solving the following SDE

$$d\mathbf{X}_t = f(\mathbf{X}_t)dt + \sqrt{2}d\mathbf{B}_t, \quad \mathbf{X}_0 \sim p_0 = p_{\text{data}}$$

  where $(\mathbf{B}_t)_{t\in[0,T]}$ is a Brownian motion and $f : \mathbb{R}^d \to \mathbb{R}^d$ is regular enough so that (strong) solutions exist.
- The reverse-time process $(\mathbf{Y}_t)_{t\in[0,T]} = (\mathbf{X}_{T-t})_{t\in[0,T]}$ satisfies

$$d\mathbf{Y}_t = \{-f(\mathbf{Y}_t) + 2\nabla \log p_{T-t}(\mathbf{Y}_t)\}dt + \sqrt{2}d\mathbf{B}_t$$

  with initialization $\mathbf{Y}_0 \sim p_T$, where p_t denotes the marginal density of $\mathbf{X}_t$.

- The reverse-time Markov chain $\{\mathbf{Y}_k\}_{k=0}^N$ is an Euler–Maruyama discretization , where the score functions $\nabla \log p_t(x)$ are approximated by $s_{\theta^*}(t, x)$.

Diffusion Schrödinger Bridge and Generative Modeling

- **The SB problem** is a classical problem appearing in applied mathematics, optimal control and probability. In the discrete-time setting, it takes the following (dynamic) form. Consider as reference density $p(x_{0:N})$, describing the process adding noise to the data. We aim to find $\pi^\star \in \mathcal{P}_{N+1}$ such that

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

- Assuming $\pi^\star$ is available, a generative model can be obtained by sampling $X_N \sim p_{\text{prior}}$, followed by the reverse-time dynamics $X_k \sim \pi^\star_{k|k+1}(\cdot|X_{k+1})$ for $k \in \{N-1, \ldots, 0\}$.

- **Static Schrödinger bridge problem.** First, we recall that the dynamic formulation admits a static analogue. The following decomposition holds for any $\pi \in \mathcal{P}_{N+1}$,

$$\mathrm{KL}(\pi \| p) = \mathrm{KL}(\pi_0, \pi_N \| p_{0,N}) + \mathbb{E}_{\pi_{0,N}}[\mathrm{KL}(\pi_{|0,N} \| p_{|0,N})]$$

where for any $\mu \in \mathcal{P}_{N+1}$ we have $\mu = \mu_{0,N}\mu_{|0,N}$, with $\mu_{|0,N}$ the conditional distribution of $X_{1:N-1}$ given X_0, X_N.

- Hence we have $\pi^\star(x_{0:N}) = \pi^{s,\star}(x_0, x_N) p_{|0,N}(x_{1:N-1}|x_0, x_N)$ where $\pi^{s,\star} \in \mathcal{P}_2$ with marginals $\pi_0^\star$ and $\pi_N^\star$ is the solution of the static SB problem

$$\pi^{s,\star} = \arg\min\{\mathrm{KL}(\pi^s \| p_{0,N}) : \pi^s \in \mathcal{P}_2,\ \pi_0^s = p_{\mathrm{data}},\ \pi_N^s = p_{\mathrm{prior}}\}.$$

- **Link with optimal transport.** Under mild assumptions, the static SB problem can be seen as an entropy-regularized optimal transport problem which is equivalent to

$$\pi^{s,\star} = \arg\min \Big\{ -\mathbb{E}_{\pi^s}[\log p_{N|0}(X_N|X_0)] - \mathcal{H}(\pi^s) :$$

$$\pi^s \in \mathcal{P}_2,\ \pi_0^s = p_{\text{data}},\ \pi_N^s = p_{\text{prior}} \Big\}$$

- If $p_{k+1|k}(x_{k+1}|x_k) = \mathcal{N}(x_{k+1}; x_k, \sigma_k^2)$, then $p_{N|0}(x_N|x_0) = \mathcal{N}(x_N; x_0, \sigma^2)$ with $\sigma^2 = \sum_{k=1}^{N} \sigma_k^2$ which induces a quadratic cost and

$$\pi^{s,\star} = \arg\min \Big\{ \mathbb{E}_{\pi^s}[\|X_0 - X_N\|^2] - 2\sigma^2\mathcal{H}(\pi^s) :$$

$$\pi^s \in \mathcal{P}_2,\ \pi_0^s = p_{\text{data}},\ \pi_N^s = p_{\text{prior}} \Big\}.$$

Iterative Proportional Fitting and Time Reversal

- The SB problem does not admit a closed-form solution. However, it can be solved using Iterative Proportional Fitting (IPF) which is defined by the following recursion for $n \in \mathbb{N}$ with initialization $\pi^0 = p$ given in:

$$\pi^{2n+1} = \arg\min \left\{ \mathrm{KL}(\pi \| \pi^{2n}) : \pi \in \mathcal{P}_{N+1},\ \pi_N = p_{\text{prior}} \right\}$$

$$\pi^{2n+2} = \arg\min \left\{ \mathrm{KL}(\pi \| \pi^{2n+1}) : \pi \in \mathcal{P}_{N+1},\ \pi_0 = p_{\text{data}} \right\}.$$

- This sequence is well-defined if there exists $\bar{\pi} \in \mathcal{P}_{N+1}$ such that $\bar{\pi}_0 = p_{\text{data}}, \bar{\pi}_N = p_{\text{prior}}$ and $\mathrm{KL}(\bar{\pi} \| p) < +\infty$. A standard representation of π^n is obtained by updating the joint density p using potential functions.

- However, this representation of the IPF iterates is difficult to approximate as it requires approximating the potentials. Our methodology builds upon an alternative representation that is better suited to numerical approximations for generative modeling where one has access to samples of p_{data} and p_{prior}.
- Assume that $\mathrm{KL}(p_{\mathrm{data}} \otimes p_{\mathrm{prior}} \| p_{0:N}) < +\infty$. Then for any $n \in \mathbb{N}$, π^{2n} and π^{2n+1} admit positive densities w.r.t. the Lebesgue measure denoted as p^n resp. q^n and for any $x_{0:N} \in \mathcal{X}$, we have

$$p^0(x_{0:N}) = p(x_{0:N}) \text{ and } q^n(x_{0:N}) = p_{\mathrm{prior}}(x_N) \prod_{k=0}^{N-1} p^n_{k+1|k}(x_{k+1}|x_k)$$

$$p^{n+1}(x_{0:N}) = p_{\mathrm{data}}(x_0) \prod_{k=0}^{N-1} q^n_{k+1|k}(x_{k+1}|x_k).$$

Iterative Proportional Fitting and Time Reversal

- In practice we have access to $p^n_{k+1|k}$ and $q^n_{k|k+1}$. Hence, to compute $p^n_{k+1|k}$ and $q^n_{k|k+1}$ we use

$$p^n_{k+1|k}(x_{k+1}|x_k) = \frac{p^n_{k+1|k}(x_{k+1}|x_k)p^n_k(x_k)}{p^n_k(x_k)}$$

$$q^n_{k|k+1}(x_k|x_{k+1}) = \frac{q^n_{k|k+1}(x_k|x_{k+1})q^n_{k+1}(x_{k+1})}{q^n_{k+1}(x_{k+1})}.$$

- We may interpret these formulas as follows. At iteration $2n$, we have $\pi^{2n} = p^n$ with $p^0 = p$. This forward process initialized with $p^n_0 = p_{\text{data}}$ defines reverse-time transitions $p^n_{k|k+1}$ which can combined with an initialization point at step N defines the reverse-time process $\pi^{2n+1} = q^n$. The forward transitions $q^n_{k+1|k}$ associated to q^n are then used to obtain $\pi^{2n+2} = p^{n+1}$. IPF then iterates this procedure.

Iterative Mean-Matching Proportional Fitting

- Approximate the IPF recursion: . If at step $n \in \mathbb{N}$ we have $p^n_{k+1|k}(x_{k+1}|x_k) = \mathcal{N}(x_{k+1}; x_k + \gamma^n_k f^n_k(x_k), 2\gamma^n_k I)$ where $p^0 = p$ and $f^n_0 = f$, then we can approximate the reverse-time transition by

$$q^n_{k|k+1}(x_k|x_{k+1}) \approx \mathcal{N}(x_k; x_{k+1} + \gamma^n_k b^{n+1}_k(x_{k+1}), 2\gamma^n_k I)$$

  with $b^{n+1}_k(x_{k+1}) = -f^n_k(x_{k+1}) + 2\nabla \log p^n_{k+1}(x_{k+1})$.
- We can also approximate the forward transitions by

$$p^{n+1}_{k+1|k}(x_{k+1}|x_k) \approx \mathcal{N}(x_{k+1}; x_k + \gamma^n_k f^{n+1}_k(x_k), 2\gamma^n_k I)$$

  with $f^{n+1}_k(x_k) = -b^{n+1}_k(x_k) + 2\nabla \log q^n_k(x_k)$.

Iterative Mean-Matching Proportional Fitting

- Hence we have $f_k^{n+1}(x_k) = f_k^n(x_k) - 2\nabla \log p_{k+1}^n(x_k) + 2\nabla \log q_k^n(x_k)$. It follows that one could estimate f_k^{n+1}, b_k^{n+1} by using score-matching to approximate $\{\nabla \log p_k^n(x)\}_{n=0}^N, \{\nabla \log q_k^n(x)\}_{n=0}^N$. This approach is prohibitively costly in terms of memory and compute. We follow an alternative approach which avoids these difficulties.

Diffusion Schrödinger Bridge

Proposition

Assume that for any $n \in \mathbb{N}$ and $k \in \{0, \dots, N-1\}$,

$$q^n_{k|k+1}(x_k|x_{k+1}) = \mathcal{N}(x_k; B^n_{k+1}(x_{k+1}), 2\gamma^n_k I),$$

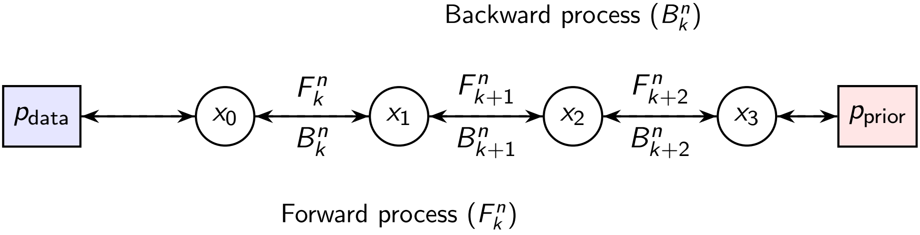
$$p^{n+1}_{k+1|k}(x_{k+1}|x_k) = \mathcal{N}(x_{k+1}; F^{n+1}_k(x_k), 2\gamma^n_k I),$$

with $B^n_{k+1}(x) = x + \gamma^n_k b^{n+1}_k(x)$, $F^n_k(x) = x + \gamma^n_k f^n_k(x)$ for any $x \in \mathbb{R}^d$.
Then we have for any $n \in \mathbb{N}$ and $k \in \{0, \dots, N-1\}$:

$$B^n_{k+1} = \arg \min_{B \in L^2(\mathbb{R}^d, \mathbb{R}^d)} \mathbb{E}_{p^n_{k+1|k}} \left[\| B(X_{k+1}) - (X_{k+1} + F^n_k(X_k) - F^n_k(X_{k+1})) \|^2 \right]$$

$$F^{n+1}_k = \arg \min_{F \in L^2(\mathbb{R}^d, \mathbb{R}^d)} \mathbb{E}_{q^n_{k|k+1}} \left[\| F(X_k) - (X_k + B^n_{k+1}(X_{k+1}) - B^n_{k+1}(X_k)) \|^2 \right]$$

Diffusion Schrödinger Bridge Map



Diffusion Schrödinger Bridge Algorithm

- In practice, we use neural networks $B_{\beta^n}(k, x) \approx B_k^n(x)$ and $F_{\alpha^n}(k, x) \approx F_k^n(x)$. Note that the networks could also be learned jointly. In this case, at equilibrium, we would obtain a bridge between p_{data} and p_{prior} but not necessarily the Schrödinger bridge.
- Network parameters α^n, β^n are learnt through gradient descent to minimize empirical versions of the sum over k of the loss functions computed using M samples and denoted as $\hat{\ell}^b_{n,k}(\beta)$ and $\hat{\ell}^f_{n,k}(\alpha)$.
- The resulting algorithm approximating $L \in \mathbb{N}$ IPF iterations is called *Diffusion Schrödinger Bridge (DSB)* with $Z_k^j, \bar{Z}_k^j \overset{\text{i.i.d.}}{\sim} \mathcal{N}(0, I)$.

Diffusion Schrödinger Bridge Algorithm

for $n \in \{0, \ldots, L\}$ **do**
    while not converged do **do**
        Sample $\{X_k^j\}_{n,k,j=0}^{N,M}$ where $X_0^j \sim p_{\text{data}}$, and

$$X_{k+1}^j = F_{\alpha^n}(k, X_k^j) + \sqrt{2\gamma_k^n} Z_{k+1}^j$$

        Compute $\hat{\ell}_{n,k}^b(\beta^n)$ approximating (**??**)
        $\beta^n \leftarrow$ Gradient Step$(\hat{\ell}_{n,k}^b(\beta^n))$
    end while
    while not converged do **do**
        Sample $\{X_k^j\}_{n,k,j=0}^{N,M}$ where $X_N^j \sim p_{\text{prior}}$, and

$$X_{k-1}^j = B_{\beta^n}(k, X_k^j) + \sqrt{2\gamma_k^n} \bar{Z}_k^j$$

        Compute $\hat{\ell}_{n,k}^f(\alpha^n)$
        $\alpha^{n+1} \leftarrow$ Gradient Step$(\hat{\ell}_{n,k}^f(\alpha^n))$
    end while
end for
Output: (α^{L+1}, β^L)

Convergence of Iterative Proportional Fitting

In this non-compact setting we require only the following mild assumption for the convergence of IPF:

$p_N, p_{\text{prior}} > 0$, $|\mathbb{H}(p_{\text{prior}})| < +\infty$,
$\int_{\mathbb{R}^d} |\log p_{N|0}(x_N|x_0)| p_{\text{data}}(x_0) p_{\text{prior}}(x_N) dx_0 dx_N < +\infty$.

Proposition: Assume this conditions, we have $(\pi^n)_{n\in\mathbb{N}}$ is well-defined and for any $n \geq 1$ we have

$$\mathrm{KL}(\pi^{n+1}\|\pi^n) \leq \mathrm{KL}(\pi^n\|\pi^{n-1}), \quad \mathrm{KL}(\pi^n\|\pi^{n+1}) \leq \mathrm{KL}(\pi^{n-1}\|\pi^n).$$

In addition, $(\|\pi^{n+1} - \pi^n\|_{TV})_{n\in\mathbb{N}}$ *and* $(J(\pi^n, \pi^{n+1}))_{n\in\mathbb{N}}$ *are non-increasing. Finally, we have*

$$\lim_{n\to\infty} \{\mathrm{KL}(\pi_0^n\|p_{\text{data}}) + \mathrm{KL}(\pi_N^n\|p_{\text{prior}})\} = 0.$$

Convergence of Iterative Proportional Fitting

Then there exists a solution $\pi^\star \in \mathcal{P}_{N+1}$ to the SB problem. Assume that $\lim_{n\to+\infty} \|\pi^n - \pi^\infty\|_{\mathrm{TV}} = 0$ with $\pi^\infty \in \mathcal{P}_{N+1}$. Let $h = p_{0,N}/(p_0 \otimes p_N)$ and assume that $h \in C((\mathbb{R}^d)^2, (0,+\infty))$ and that there exist $\Phi_0, \Phi_N \in C(\mathbb{R}^d, (0,+\infty))$ such that

$$\int_{\mathbb{R}^d\times\mathbb{R}^d} \left(|\log h(x_0,x_N)| + |\log \Phi_0(x_0)| + |\log \Phi_N(x_N)|\right) p_{\mathrm{data}}(x_0) p_{\mathrm{prior}}(x_N) dx_0 dx_N < +\infty$$

with $h(x_0,x_N) \le \Phi_0(x_0)\Phi_N(x_N)$. *If p is absolutely continuous w.r.t. π^∞ then $\pi^\infty = \pi^\star$.*

Continuous-time IPF

- Given a reference measure $\mathbb{P} \in \mathcal{P}(\mathcal{C})$, the continuous formulation of the SB involves solving the following problem

$$\Pi^\star = \arg\min\{\mathrm{KL}(\Pi \| \mathbb{P}) : \Pi \in \mathcal{P}(\mathcal{C}),\, \Pi_0 = p_{\mathrm{data}},\, \Pi_T = p_{\mathrm{prior}}\}, \quad T = \sum_{k=0}^{N-1} \gamma_{k+1}.$$

- Similarly to the discretization, we define the IPF $(\Pi^n)_{n \in \mathbb{N}}$ with $\Pi^0 = \mathbb{P}$ associated with forward EDS and for any $n \in \mathbb{N}$

$$\Pi^{2n+1} = \arg\min\{\mathrm{KL}(\Pi \| \Pi^{2n}) : \Pi \in \mathcal{P}(\mathcal{C}),\, \Pi_T = p_{\mathrm{prior}}\},$$
$$\Pi^{2n+2} = \arg\min\{\mathrm{KL}(\Pi \| \Pi^{2n+1}) : \Pi \in \mathcal{P}(\mathcal{C}),\, \Pi_0 = p_{\mathrm{data}}\}.$$

- For any $n \in \mathbb{N}$, $\Pi^n = \pi^{S,n}_{[0,T]}$, with $(\pi^{S,n})_{n \in \mathbb{N}}$ the IPF for the static SB problem. Then the DSB can be seen as a discretization of the continuous IPF.

Continuous-time IPF

Proposition

Assume mild condition and that there exist $\mathbb{M} \in \mathcal{P}(\mathcal{C})$, $U \in C^1(\mathbb{R}^d, \mathbb{R})$, $C \geq 0$ such that for any $n \in \mathbb{N}$, $x \in \mathbb{R}^d$, $\mathrm{KL}(\Pi^n \| \mathbb{M}) < +\infty$, $\langle x, \nabla U(x) \rangle \geq -C(1 + \|x\|^2)$ and $\mathbb{M}$ is associated with

$$dX_t = -\nabla U(X_t)dt + \sqrt{2}dB_t$$

with X_0 distributed according to the invariant distribution of (14). Then, for any $n \in \mathbb{N}$ we have:

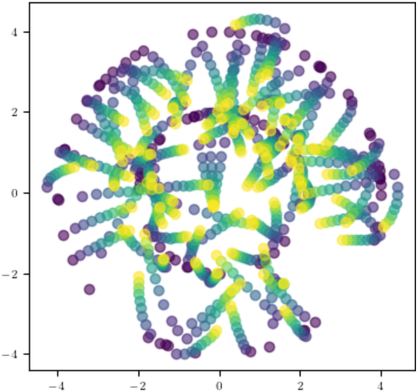
Continuous-time IPF

- $(\Pi^{2n+1})^R$ is associated with $dY_t^{2n+1} = b_t^n(Y_t^{2n+1})dt + \sqrt{2}dB_t$ with $Y_0^{2n+1} \sim p_{\text{prior}}$
- Π^{2n+2} is associated with $dX_t^{2n+2} = f_t^n(X_t^{2n+2})dt + \sqrt{2}dB_t$ with $X_0^{2n+2} \sim p_{\text{data}}$

where for any $n \in \mathbb{N}$, $t \in [0, T]$ and $x \in \mathbb{R}^d$, $b_t^n(x) = -f_{T-t}^n(x) + 2\nabla \log p_t^n(x)$, $f_t^n(x) = f(x)$, and p_t^n, q_t^n the densities of Π^n and $(\Pi^n)^R$.

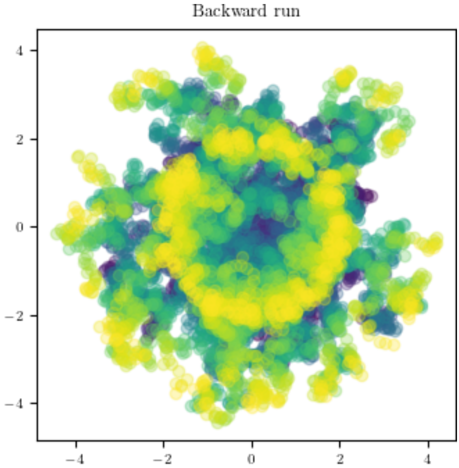
Applications

Visualization a few trajectories from the Ornstein-Uhlenbeck process $\mathrm{d}\mathbf{X}_t = -(1/2)\mathbf{X}_t\mathrm{d}t + \mathrm{d}\mathbf{B}_t$. Similarly, we can also plot the mean and standard deviation of the forward trajectories which should converge to 0, respectively 1.



Applications

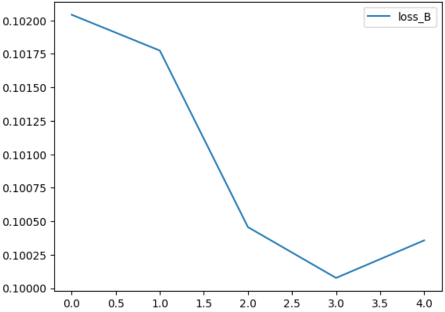
We sample approximately from the backward SDE
$\mathrm{d}\mathbf{Y}_t = \{(1/2)\mathbf{Y}_t + \nabla \log p_{T-t}(\mathbf{Y}_t)\}\mathrm{d}t + \mathrm{d}\mathbf{B}_t, \qquad \mathbf{Y}_0 \sim \pi_T,$



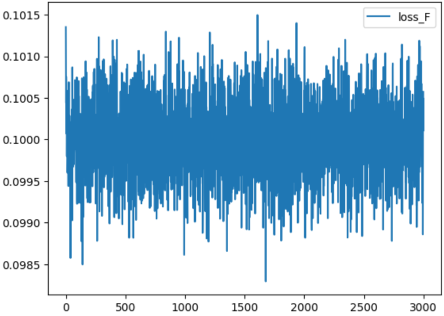
Applications

- We choose 100 images of MNIST, 10 for each number.
- We use DSB to train the forward and backward process with number of DSB iterations 5 ,time steps 10, noise level 0.05 and 1 epoch.
- We plot the loss of the diffusion models

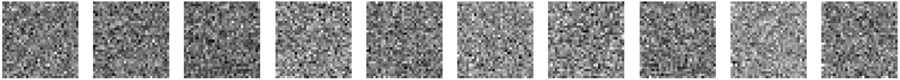
Loss backward model



Loss forward model



Noising images



Conclusion

Thanks for your attention!