Score-based diffusion models

May 28, 2025

Part 1: Introduction

Sampling tasks

In many applications, we want to sample from a target distribution p_{target}.

Sampling tasks

In many applications, we want to sample from a target distribution p_{target}.

1. simple, e.g. uniform distribution, Gaussian distributions
   $\Longrightarrow$ direct sampling (independent samples)

Sampling tasks

In many applications, we want to sample from a target distribution p_{target}.

1. simple, e.g. uniform distribution, Gaussian distributions
   $\Longrightarrow$ direct sampling (independent samples)

2. known up to a constant, e.g. Boltzmann distribution: $\frac{1}{Z}\mathrm{e}^{-\beta V(x)}$
   $\Longrightarrow$ sampling SDEs, Markov chain Monte Carlo (MCMC)

Sampling tasks

In many applications, we want to sample from a target distribution p_{target}.

1. simple, e.g. uniform distribution, Gaussian distributions
   $\Longrightarrow$ direct sampling (independent samples)
2. known up to a constant, e.g. Boltzmann distribution: $\frac{1}{Z}\mathrm{e}^{-\beta V(x)}$
   $\Longrightarrow$ sampling SDEs, Markov chain Monte Carlo (MCMC)
3. unknown, but data is available, e.g. images
   $\Longrightarrow$ generative modeling

Generative modelling

Different approaches:

1. Variational AutoEncoders (VAEs)
2. Generative Adversarial Networks (GANs)
3. Normalizing Flows (NFs)
4. Diffusion generative models
5. Flow-based generative models

Score-based diffusion models

Time interval: $[0, T]$. p_{prior} is a simple distribution.

1. forward process X_t:
   $X_0 \sim p_{\text{target}} \to X_T \sim p(\cdot, T)$.
2. backward process Y_t:
   $Y_0 \sim p(\cdot, T) \to Y_T \sim p_{\text{target}}$.

Score-based diffusion models

Time interval: $[0, T]$. p_{prior} is a simple distribution.

1. forward process X_t:
   $$X_0 \sim p_{\text{target}} \to X_T \sim p(\cdot, T).$$
2. backward process Y_t:
   $$Y_0 \sim p(\cdot, T) \to Y_T \sim p_{\text{target}}.$$

Steps:

1. choose a forward process X_t
   linear SDE $\Longrightarrow p(\cdot, T)$ is (approximately) Gaussian p_{prior}.

Score-based diffusion models

Time interval: $[0, T]$. p_{prior} is a simple distribution.

1. forward process X_t:
   $X_0 \sim p_{\text{target}} \to X_T \sim p(\cdot, T)$.
2. backward process Y_t:
   $Y_0 \sim p(\cdot, T) \to Y_T \sim p_{\text{target}}$.

Steps:

1. choose a forward process X_t
   linear SDE $\Longrightarrow p(\cdot, T)$ is (approximately) Gaussian p_{prior}.
2. learn SDE of the backward process Y_t

Score-based diffusion models

Time interval: $[0, T]$. p_{prior} is a simple distribution.

1. forward process X_t:
   $$X_0 \sim p_{\text{target}} \to X_T \sim p(\cdot, T).$$
2. backward process Y_t:
   $$Y_0 \sim p(\cdot, T) \to Y_T \sim p_{\text{target}}.$$

Steps:

1. choose a forward process X_t
   linear SDE $\Longrightarrow p(\cdot, T)$ is (approximately) Gaussian p_{prior}.
2. learn SDE of the backward process Y_t
3. generate new data: sample $Y_0 \sim p_{\text{prior}}$ and simulate the backward process to get Y_T

Part 2: Forward process

Forward process

SDE in $\mathbb{R}^d$

$$\begin{aligned} dX_t &= f(X_t, t)\, dt + \sigma(t) dB_t\,, \quad t \in [0, T]\,, \\ X_0 &\sim p_0 = p_{\text{target}}\,. \end{aligned} \tag{1}$$

Forward process

Fokker-Planck equation

$$\frac{\partial p}{\partial t} = \mathcal{L}_t^\top p\,, \quad t \in [0, T]\,, \\ p(x,0) = p_0(x)\,. \tag{2}$$

$\mathcal{L}_t$ is the generator of (1) at time t:

$$(\mathcal{L}_t g)(x) = f(x,t) \cdot \nabla g(x) + \frac{\sigma^2(t)}{2} \Delta g(x)\,, \quad g : \mathbb{R}^d \to \mathbb{R}\,, \tag{3}$$

Forward process

Fokker-Planck equation

$$\begin{aligned} \frac{\partial p}{\partial t} &= \mathcal{L}_t^\top p\,, \quad t \in [0, T]\,, \\ p(x,0) &= p_0(x)\,. \end{aligned} \tag{2}$$

$\mathcal{L}_t$ is the generator of (1) at time t:

$$(\mathcal{L}_t g)(x) = f(x,t)\cdot\nabla g(x) + \frac{\sigma^2(t)}{2}\Delta g(x)\,, \quad g : \mathbb{R}^d \to \mathbb{R}\,, \tag{3}$$

and $\mathcal{L}_t^\top$ is the adjoint operator of $\mathcal{L}_t$:

$$(\mathcal{L}_t^\top g)(x) = -\mathrm{div}(f(x,t)g(x)) + \frac{\sigma^2(t)}{2}\Delta g(x)\,, \quad g : \mathbb{R}^d \to \mathbb{R}\,. \tag{4}$$

Example: Linear SDEs

$$dX_t = -\alpha(t) X_t dt + \sqrt{\beta(t)} dB_t \, ,$$

Example: Linear SDEs

$$dX_t = -\alpha(t) X_t dt + \sqrt{\beta(t)} dB_t \,,$$

1.

$$X_t = \mathrm{e}^{-\int_0^t \alpha(s)ds} X_0 + \int_0^t \mathrm{e}^{-\int_s^t \alpha(r)dr} \sqrt{\beta(s)} dB_s \,.$$

Example: Linear SDEs

$$dX_t = -\alpha(t) X_t dt + \sqrt{\beta(t)} dB_t \,,$$

1. $$X_t = \mathrm{e}^{-\int_0^t \alpha(s)ds} X_0 + \int_0^t \mathrm{e}^{-\int_s^t \alpha(r)dr} \sqrt{\beta(s)} dB_s \,.$$

2. For fixed $X_0 = x_0$, X_t is Gaussian:

$$X_t \sim \mathcal{N}\left(\mathrm{e}^{-\int_0^t \alpha(s)ds} x_0, \eta^2(t) \mathbf{1}_d\right), \quad \text{where } \eta^2(t) = \int_0^t \mathrm{e}^{-2\int_s^t \alpha(r)dr} \beta(s) ds \,.$$

Example: Linear SDEs

$$dX_t = -\alpha(t) X_t dt + \sqrt{\beta(t)} dB_t \,,$$

1. $$X_t = \mathrm{e}^{-\int_0^t \alpha(s)ds} X_0 + \int_0^t \mathrm{e}^{-\int_s^t \alpha(r)dr} \sqrt{\beta(s)} dB_s \,.$$

2. For fixed $X_0 = x_0$, X_t is Gaussian:

$$X_t \sim \mathcal{N}\left(\mathrm{e}^{-\int_0^t \alpha(s)ds} x_0, \eta^2(t) \mathbf{1}_d\right), \quad \text{where } \eta^2(t) = \int_0^t \mathrm{e}^{-2\int_s^t \alpha(r)dr} \beta(s) ds \,.$$

$$\implies \quad p(x, t | x_0) = \left(2\pi \eta(t)^2\right)^{-\frac{d}{2}} \mathrm{e}^{-\frac{1}{2\eta(t)^2} |x - \mathrm{e}^{-\int_0^t \alpha(s)ds} x_0|^2}, \quad x \in \mathbb{R}^d \,.$$

Linear SDEs

Fokker-Planck equation:

$$\begin{aligned} \frac{\partial p}{\partial t} &= \mathcal{L}_t^\top p\,, \quad t \in [0, T]\,, \\ p(x, 0) &= p_0(x)\,. \end{aligned} \tag{5}$$

Fokker-Planck equation:

$$\begin{aligned}\frac{\partial p}{\partial t} &= \mathcal{L}_t^\top p\,, \quad t \in [0, T]\,, \\ p(x, 0) &= p_0(x)\,.\end{aligned} \tag{5}$$

1. When $p_0(x) = \delta(x - x_0)$,

$$p(x,t) = p(x,t|x_0) = \big(2\pi\eta^2(t)\big)^{-\frac{d}{2}} \mathrm{e}^{-\frac{1}{2\eta^2(t)}|x - \mathrm{e}^{-\int_0^t \alpha(s)ds} x_0|^2}\,.$$

Fokker-Planck equation:

$$\begin{aligned} \frac{\partial p}{\partial t} &= \mathcal{L}_t^\top p\,, \quad t \in [0, T]\,, \\ p(x, 0) &= p_0(x)\,. \end{aligned} \tag{5}$$

1. When $p_0(x) = \delta(x - x_0)$,

$$p(x, t) = p(x, t|x_0) = \left(2\pi\eta^2(t)\right)^{-\frac{d}{2}} \mathrm{e}^{-\frac{1}{2\eta^2(t)}|x - \mathrm{e}^{-\int_0^t \alpha(s)ds} x_0|^2}\,.$$

2. For general p_0,

$$p(x, t) = \int_{\mathbb{R}^d} p(x, t|x_0) p_0(x_0) dx_0\,.$$

Part 3: Backward process

Time-reversal

Define

$$q(x,t) = p(x, T-t), \quad \forall\, t \in [0,T]\,. \tag{6}$$

Time-reversal

Define

$$q(x,t) = p(x, T-t), \quad \forall\, t \in [0, T]\,. \tag{6}$$

Theorem

q is the probability density of the following SDE

$$\begin{aligned} dY_t =& f^-(Y_t, t)\, dt + \sigma^-(t) dB_t\,, \quad t \in [0, T]\,, \\ Y_0 \sim& p(\cdot, T)\,, \end{aligned} \tag{7}$$

where

$$\begin{aligned} f^-(x,t) =& -f(x, T-t) + \sigma^2(T-t)\nabla \ln p(x, T-t) \\ \sigma^-(t) =& \sigma(T-t)\,. \end{aligned} \tag{8}$$

Time-reversal

Define

$$q(x,t) = p(x, T-t), \quad \forall\, t \in [0,T]\,. \tag{6}$$

Theorem

q is the probability density of the following SDE

$$\begin{aligned} dY_t =& f^-(Y_t, t)\,dt + \sigma^-(t)dB_t\,, \quad t \in [0,T]\,, \\ Y_0 \sim& p(\cdot, T)\,, \end{aligned} \tag{7}$$

where

$$\begin{aligned} f^-(x,t) =& -f(x, T-t) + \sigma^2(T-t)\nabla \ln p(x, T-t) \\ \sigma^-(t) =& \sigma(T-t)\,. \end{aligned} \tag{8}$$

Proof.

p solves Fokker-Planck equation associated to X_t. Using this fact to show that q solves the Fokker-Planck equation associated to Y_t. □

Backward process

From the previous theorem:

Backward process

From the previous theorem:

1. (backward) process Y_t: $p(\cdot, T) \implies p_0 = p_{\text{target}}$.

Backward process

From the previous theorem:

1. (backward) process Y_t: $\quad p(\cdot, T) \implies p_0 = p_{\text{target}}$.
2. SDE is known.

$$\begin{aligned} dY_t =& \Big(-f(Y_t, T-t) + \sigma^2(T-t)\nabla \ln p(Y_t, T-t) \Big)\, dt + \sigma(T-t) dB_t\,, \\ Y_0 \sim& p(\cdot, T)\,. \end{aligned} \tag{9}$$

Backward process

From the previous theorem:

1. (backward) process Y_t: $\quad p(\cdot, T) \implies p_0 = p_{\text{target}}$.
2. SDE is known.

$$
\begin{aligned}
dY_t =& \Big(-f(Y_t, T-t) + \sigma^2(T-t)\nabla \ln p(Y_t, T-t) \Big)\, dt + \sigma(T-t) dB_t\,, \\
Y_0 \sim& p(\cdot, T)\,.
\end{aligned} \tag{9}
$$

3. Coefficients depend on $\nabla \ln p$ (score function).

Part 4: Learning the score function

Overall method

Given: dataset sampled from some (unknown) p_{target}.

Overall method

Given: dataset sampled from some (unknown) p_{target}.
Goal: train the model on dataset to generate new data.

Overall method

Given: dataset sampled from some (unknown) p_{target}.
Goal: train the model on dataset to generate new data.

Recall the steps:

1. choose a forward process X_t
   linear process $\Longrightarrow p(\cdot, T)$ is (approximately) Gaussian p_{prior}.

Overall method

Given: dataset sampled from some (unknown) p_{target}.
Goal: train the model on dataset to generate new data.

Recall the steps:

1. choose a forward process X_t
   linear process $\Longrightarrow p(\cdot, T)$ is (approximately) Gaussian p_{prior}.
2. learn SDE of backward process $Y_t \iff$ learn the score $\nabla \ln p$.

Overall method

Given: dataset sampled from some (unknown) p_{target}.
Goal: train the model on dataset to generate new data.

Recall the steps:

1. choose a forward process X_t
   linear process $\implies p(\cdot, T)$ is (approximately) Gaussian p_{prior}.

2. learn SDE of backward process $Y_t \iff$ learn the score $\nabla \ln p$.

3. sample $Y_0 \sim p_{\text{prior}}$ and simulate Y_t to get Y_T
   $\implies \quad Y_T \sim q(\cdot, T) = p(\cdot, 0) = p_{\text{target}}$

Starting point

Starting point

Clearly, the score $\nabla \ln p$ solves the minimization problem:

$$
\begin{aligned}
&\min_{u:\mathbb{R}^d\times[0,T]\to\mathbb{R}^d} \mathbb{E}_{t\sim U([0,T])}\mathbb{E}_{x\sim p(\cdot,t)}\left[\frac{1}{2}\left|u(x,t)-\nabla \ln p(x,t)\right|^2 w(t)\right] \\
=&\min_{u:\mathbb{R}^d\times[0,T]\to\mathbb{R}^d} \left[\frac{1}{T}\int_0^T \left(\int_{\mathbb{R}^d} \frac{1}{2}\left|u(x,t)-\nabla \ln p(x,t)\right|^2 p(x,t)dx\right) w(t)dt\right],
\end{aligned} \tag{10}
$$

where $w(t):[0,T]\to\mathbb{R}^+$ is a weight function.

Starting point

Clearly, the score $\nabla \ln p$ solves the minimization problem:

$$\begin{aligned} &\min_{u:\mathbb{R}^d\times[0,T]\to\mathbb{R}^d} \mathbb{E}_{t\sim U([0,T])}\mathbb{E}_{x\sim p(\cdot,t)}\left[\frac{1}{2}\left|u(x,t)-\nabla \ln p(x,t)\right|^2 w(t)\right] \\ =&\min_{u:\mathbb{R}^d\times[0,T]\to\mathbb{R}^d}\left[\frac{1}{T}\int_0^T\left(\int_{\mathbb{R}^d}\frac{1}{2}\left|u(x,t)-\nabla \ln p(x,t)\right|^2 p(x,t)dx\right)w(t)dt\right], \end{aligned} \tag{10}$$

where $w(t):[0,T]\to\mathbb{R}^+$ is a weight function.

However, (10) is **not useful** because the density p is unknown.

Loss objective

Recall that

$$p(x,t) = \int_{\mathbb{R}^d} p(x,t|x_0)p_0(x_0)dx_0\,,$$

Loss objective

Recall that

$$p(x,t) = \int_{\mathbb{R}^d} p(x,t|x_0)p_0(x_0)dx_0\,,$$

and, for linear SDEs, we have

$$p(x,t|x_0) = \left(2\pi\eta^2(t)\right)^{-\frac{d}{2}} \mathrm{e}^{-\frac{1}{2\eta^2(t)}|x-\mathrm{e}^{-\int_0^t \alpha(s)ds}x_0|^2}\,.$$

Loss objective

Recall that

$$p(x,t) = \int_{\mathbb{R}^d} p(x,t|x_0)p_0(x_0)dx_0\,,$$

and, for linear SDEs, we have

$$p(x,t|x_0) = \left(2\pi\eta^2(t)\right)^{-\frac{d}{2}} \mathrm{e}^{-\frac{1}{2\eta^2(t)}|x - \mathrm{e}^{-\int_0^t \alpha(s)ds} x_0|^2}\,.$$

Idea: replace $p(x,t)$ in the loss by $p(x,t|x_0)$.

Loss objective

$$\int_0^T \left[\int_{\mathbb{R}^d} \frac{1}{2} |u(x,t) - \nabla \ln p(x,t)|^2 p(x,t) dx \right] w(t) dt$$

Loss objective

$$\int_0^T \Big[\int_{\mathbb{R}^d} \frac{1}{2} |u(x,t) - \nabla \ln p(x,t)|^2 p(x,t) dx \Big] w(t) dt$$
$$= \int_0^T \Big[\int_{\mathbb{R}^d} \frac{1}{2} \Big| u(x,t) - \frac{\nabla p(x,t)}{p(x,t)} \Big|^2 p(x,t) dx \Big] w(t) dt$$

Loss objective

$$
\begin{aligned}
&\int_0^T \Big[\int_{\mathbb{R}^d} \frac{1}{2} |u(x,t) - \nabla \ln p(x,t)|^2 p(x,t) dx \Big] w(t) dt \\
= &\int_0^T \Big[\int_{\mathbb{R}^d} \frac{1}{2} \Big| u(x,t) - \frac{\nabla p(x,t)}{p(x,t)} \Big|^2 p(x,t) dx \Big] w(t) dt \\
= &\int_0^T \Big[\int_{\mathbb{R}^d} \Big(\frac{1}{2} |u(x,t)|^2 p(x,t) - u(x,t) \cdot \nabla p(x,t) + \frac{1}{2} \frac{|\nabla p(x,t)|^2}{p(x,t)} \Big) dx \Big] w(t)\, dt
\end{aligned}
$$

Loss objective

$$
\begin{aligned}
&\int_0^T \Big[\int_{\mathbb{R}^d} \frac{1}{2}|u(x,t) - \nabla \ln p(x,t)|^2 p(x,t)dx \Big] w(t)dt \\
&= \int_0^T \Big[\int_{\mathbb{R}^d} \frac{1}{2}\Big|u(x,t) - \frac{\nabla p(x,t)}{p(x,t)}\Big|^2 p(x,t)dx\Big]\, w(t)dt \\
&= \int_0^T \Big[\int_{\mathbb{R}^d} \Big(\frac{1}{2}|u(x,t)|^2 p(x,t) - u(x,t)\cdot\nabla p(x,t) + \frac{1}{2}\frac{|\nabla p(x,t)|^2}{p(x,t)}\Big)dx\Big] w(t)\, dt \\
&= \int_0^T \Big[\int_{\mathbb{R}^d} \Big(\frac{1}{2}|u(x,t)|^2 p(x,t) - u(x,t)\cdot\nabla p(x,t)\Big)dx\Big]\, w(t)dt + C_1\,,
\end{aligned}
$$

where C_1 is a constant independent of u.

Loss objective

Using $p(x,t) = \int_{\mathbb{R}^d} p(x,t|x_0)p_0(x_0)dx_0$, we derive

$$\int_0^T \Big[\int_{\mathbb{R}^d} \Big(\frac{1}{2}|u(x,t)|^2 p(x,t) - u(x,t)\cdot\nabla p(x,t) \Big) dx \Big] w(t)dt$$

Loss objective

Using $p(x,t) = \int_{\mathbb{R}^d} p(x,t|x_0)p_0(x_0)dx_0$, we derive

$$\int_0^T \Big[\int_{\mathbb{R}^d} \Big(\frac{1}{2}|u(x,t)|^2 p(x,t) - u(x,t)\cdot\nabla p(x,t)\Big)dx\Big]\, w(t)dt$$

$$= \int_0^T \Big[\int_{\mathbb{R}^d} \Big(\frac{1}{2}|u(x,t)|^2 \int_{\mathbb{R}^d} p(x,t|x_0)p_0(x_0)dx_0$$

$$- u(x,t)\cdot\int_{\mathbb{R}^d} \nabla p(x,t|x_0)p_0(x_0)dx_0\Big)dx\Big]\, w(t)dt$$

Loss objective

Using $p(x,t) = \int_{\mathbb{R}^d} p(x,t|x_0)p_0(x_0)dx_0$, we derive

$$\int_0^T \Big[\int_{\mathbb{R}^d} \Big(\frac{1}{2}|u(x,t)|^2 p(x,t) - u(x,t)\cdot \nabla p(x,t) \Big) dx \Big] w(t)dt$$

$$= \int_0^T \Big[\int_{\mathbb{R}^d} \Big(\frac{1}{2}|u(x,t)|^2 \int_{\mathbb{R}^d} p(x,t|x_0)p_0(x_0)dx_0$$
$$- u(x,t)\cdot \int_{\mathbb{R}^d} \nabla p(x,t|x_0)p_0(x_0)dx_0 \Big) dx \Big] w(t)dt$$

$$= \int_0^T \Big[\int_{\mathbb{R}^d}\int_{\mathbb{R}^d} \Big(\frac{1}{2}|u(x,t)|^2 p(x,t|x_0) - u(x,t)\cdot \nabla p(x,t|x_0) \Big) p_0(x_0)dx_0 dx \Big] w(t)dt$$

Loss objective

Using $p(x,t) = \int_{\mathbb{R}^d} p(x,t|x_0)p_0(x_0)dx_0$, we derive

$$\int_0^T \Big[\int_{\mathbb{R}^d} \Big(\frac{1}{2}|u(x,t)|^2 p(x,t) - u(x,t)\cdot\nabla p(x,t)\Big)dx\Big]\, w(t)dt$$

$$= \int_0^T \Big[\int_{\mathbb{R}^d} \Big(\frac{1}{2}|u(x,t)|^2 \int_{\mathbb{R}^d} p(x,t|x_0)p_0(x_0)dx_0 - u(x,t)\cdot\int_{\mathbb{R}^d} \nabla p(x,t|x_0)p_0(x_0)dx_0\Big)dx\Big]\, w(t)dt$$

$$= \int_0^T \Big[\int_{\mathbb{R}^d}\int_{\mathbb{R}^d} \Big(\frac{1}{2}|u(x,t)|^2 p(x,t|x_0) - u(x,t)\cdot\nabla p(x,t|x_0)\Big)p_0(x_0)dx_0 dx\Big]\, w(t)dt$$

$$= \int_0^T \Big[\int_{\mathbb{R}^d}\int_{\mathbb{R}^d} \Big(\frac{1}{2}|u(x,t)|^2 - u(x,t)\cdot\nabla \ln p(x,t|x_0)\Big)p(x,t|x_0)p_0(x_0)dx_0 dx\Big]\, w(t)dt$$

Loss objective

Using $p(x,t) = \int_{\mathbb{R}^d} p(x,t|x_0)p_0(x_0)dx_0$, we derive

$$\int_0^T \Big[\int_{\mathbb{R}^d} \Big(\frac{1}{2}|u(x,t)|^2 p(x,t) - u(x,t)\cdot\nabla p(x,t)\Big)dx\Big]\, w(t)dt$$

$$= \int_0^T \Big[\int_{\mathbb{R}^d} \Big(\frac{1}{2}|u(x,t)|^2 \int_{\mathbb{R}^d} p(x,t|x_0)p_0(x_0)dx_0$$

$$- u(x,t)\cdot\int_{\mathbb{R}^d} \nabla p(x,t|x_0)p_0(x_0)dx_0\Big)dx\Big]\, w(t)dt$$

$$= \int_0^T \Big[\int_{\mathbb{R}^d}\int_{\mathbb{R}^d} \Big(\frac{1}{2}|u(x,t)|^2 p(x,t|x_0) - u(x,t)\cdot\nabla p(x,t|x_0)\Big)p_0(x_0)dx_0dx\Big]\, w(t)dt$$

$$= \int_0^T \Big[\int_{\mathbb{R}^d}\int_{\mathbb{R}^d} \Big(\frac{1}{2}|u(x,t)|^2 - u(x,t)\cdot\nabla \ln p(x,t|x_0)\Big)p(x,t|x_0)p_0(x_0)dx_0dx\Big]\, w(t)dt$$

$$= \int_0^T \Big[\int_{\mathbb{R}^d}\int_{\mathbb{R}^d} \frac{1}{2}\Big|u(x,t) - \nabla \ln p(x,t|x_0)\Big|^2 p(x,t|x_0)p_0(x_0)dx_0dx\Big]\, w(t)dt + C_2$$

Loss objective

Using $p(x,t) = \int_{\mathbb{R}^d} p(x,t|x_0)p_0(x_0)dx_0$, we derive

$$\int_0^T \Big[\int_{\mathbb{R}^d} \Big(\frac{1}{2}|u(x,t)|^2 p(x,t) - u(x,t)\cdot\nabla p(x,t)\Big)dx\Big]\, w(t)dt$$

$$= \int_0^T \Big[\int_{\mathbb{R}^d} \Big(\frac{1}{2}|u(x,t)|^2 \int_{\mathbb{R}^d} p(x,t|x_0)p_0(x_0)dx_0 \\ - u(x,t)\cdot \int_{\mathbb{R}^d} \nabla p(x,t|x_0)p_0(x_0)dx_0\Big)dx\Big]\, w(t)dt$$

$$= \int_0^T \Big[\int_{\mathbb{R}^d}\int_{\mathbb{R}^d} \Big(\frac{1}{2}|u(x,t)|^2 p(x,t|x_0) - u(x,t)\cdot\nabla p(x,t|x_0)\Big)p_0(x_0)dx_0 dx\Big]\, w(t)dt$$

$$= \int_0^T \Big[\int_{\mathbb{R}^d}\int_{\mathbb{R}^d} \Big(\frac{1}{2}|u(x,t)|^2 - u(x,t)\cdot\nabla \ln p(x,t|x_0)\Big)p(x,t|x_0)p_0(x_0)dx_0 dx\Big]\, w(t)dt$$

$$= \int_0^T \Big[\int_{\mathbb{R}^d}\int_{\mathbb{R}^d} \frac{1}{2}\Big|u(x,t) - \nabla \ln p(x,t|x_0)\Big|^2 p(x,t|x_0)p_0(x_0)dx_0 dx\Big]\, w(t)dt + C_2$$

$$= T\mathbb{E}_{t\sim U([0,T])}\mathbb{E}_{x_0\sim p_0}\mathbb{E}_{x\sim p(x,t|x_0)}\Big[\frac{1}{2}\Big|u(x,t) - \nabla \ln p(x,t|x_0)\Big|^2 w(t)\Big] + C_2$$

Loss objective

Using $p(x,t) = \int_{\mathbb{R}^d} p(x,t|x_0)p_0(x_0)dx_0$, we derive

$$
\begin{aligned}
&\int_0^T \Big[\int_{\mathbb{R}^d} \Big(\frac{1}{2}|u(x,t)|^2 p(x,t) - u(x,t)\cdot\nabla p(x,t)\Big)dx\Big]\, w(t)dt \\
=&\int_0^T \Big[\int_{\mathbb{R}^d} \Big(\frac{1}{2}|u(x,t)|^2 \int_{\mathbb{R}^d} p(x,t|x_0)p_0(x_0)dx_0 \\
&\qquad - u(x,t)\cdot \int_{\mathbb{R}^d} \nabla p(x,t|x_0)p_0(x_0)dx_0\Big)dx\Big]\, w(t)dt \\
=&\int_0^T \Big[\int_{\mathbb{R}^d}\int_{\mathbb{R}^d} \Big(\frac{1}{2}|u(x,t)|^2 p(x,t|x_0) - u(x,t)\cdot\nabla p(x,t|x_0)\Big)p_0(x_0)dx_0dx\Big]\, w(t)dt \\
=&\int_0^T \Big[\int_{\mathbb{R}^d}\int_{\mathbb{R}^d} \Big(\frac{1}{2}|u(x,t)|^2 - u(x,t)\cdot\nabla \ln p(x,t|x_0)\Big)p(x,t|x_0)p_0(x_0)dx_0dx\Big]\, w(t)dt \\
=&\int_0^T \Big[\int_{\mathbb{R}^d}\int_{\mathbb{R}^d} \frac{1}{2}\Big|u(x,t) - \nabla \ln p(x,t|x_0)\Big|^2 p(x,t|x_0)p_0(x_0)dx_0dx\Big]\, w(t)dt + C_2 \\
=&T\mathbb{E}_{t\sim U([0,T])}\mathbb{E}_{x_0\sim p_0}\mathbb{E}_{x\sim p(x,t|x_0)}\Big[\frac{1}{2}\Big|u(x,t) - \nabla \ln p(x,t|x_0)\Big|^2 w(t)\Big] + C_2 \\
=&T\mathbb{E}_{t\sim U([0,T])}\mathbb{E}_{x_0\sim p_0}\mathbb{E}_{x\sim p(x,t|x_0)}\Big[\Big(\frac{1}{2}|u(x,t)|^2 - u(x,t)\cdot\nabla \ln p(x,t|x_0)\Big)w(t)\Big] + C_3,
\end{aligned}
$$

where C_2, C_3 are constants independent of u.

Loss objective

To summarize, we have obtained the following result.

Loss objective

To summarize, we have obtained the following result.

Proposition

For $u : \mathbb{R}^d \times [0, T] \to \mathbb{R}^d$, we have

$$\begin{aligned}
&\mathbb{E}_{t\sim U([0,T])}\mathbb{E}_{x\sim p(\cdot,t)}\left[\frac{1}{2}\left|u(x,t) - \nabla \ln p(x,t)\right|^2 w(t)\right] \\
=&\mathbb{E}_{t\sim U([0,T])}\mathbb{E}_{x_0\sim p_0}\mathbb{E}_{x\sim p(\cdot,t|x_0)}\left[\frac{1}{2}\left|u(x,t) - \nabla \ln p(x,t|x_0)\right|^2 w(t)\right] + C_2 \\
=&\mathbb{E}_{t\sim U([0,T])}\mathbb{E}_{x_0\sim p_0}\mathbb{E}_{x\sim p(\cdot,t|x_0)}\left[\left(\frac{1}{2}|u(x,t)|^2 - u(x,t)\cdot\nabla \ln p(x,t|x_0)\right) w(t)\right] + C_3,
\end{aligned} \tag{11}$$

where C_2, C_3 are constants independent of u.

Practical issue

Loss function:

$$\mathrm{Loss}(u) = \mathbb{E}_{t\sim U([0,T])}\mathbb{E}_{x_0\sim p_0}\mathbb{E}_{x\sim p(\cdot,t|x_0)}\left[\left(\frac{1}{2}|u(x,t)|^2 - u(x,t)\cdot\nabla \ln p(x,t|x_0)\right)w(t)\right].$$

Practical issue

Loss function:

$$\mathrm{Loss}(u) = \mathbb{E}_{t\sim U([0,T])}\mathbb{E}_{x_0\sim p_0}\mathbb{E}_{x\sim p(\cdot,t|x_0)}\left[\left(\frac{1}{2}|u(x,t)|^2 - u(x,t)\cdot\nabla \ln p(x,t|x_0)\right)w(t)\right].$$

Numerical issue: $\nabla \ln p(x,t|x_0)$ diverges as $t \to 0$.

Practical issue

Loss function:

$$\mathrm{Loss}(u) = \mathbb{E}_{t\sim U([0,T])}\mathbb{E}_{x_0\sim p_0}\mathbb{E}_{x\sim p(\cdot,t|x_0)}\left[\left(\frac{1}{2}|u(x,t)|^2 - u(x,t)\cdot\nabla \ln p(x,t|x_0)\right)w(t)\right].$$

Numerical issue: $\nabla \ln p(x,t|x_0)$ diverges as $t \to 0$.

1. Draw $t \sim U([\epsilon, T])$ with $\epsilon > 0$.
2. use perturbed data $\Longleftrightarrow$ approximate $p(x,t|x_0)$
3. choose $w(t)$ such that $w(t) \to 0$ as $t \to 0$.

VESDE

1. SDE $dX_t = \sqrt{\beta(t)} dB_t$, where $0 < \eta_{\min} \le \eta_{\max}$ and

$$\beta(t) = \frac{2\eta_{\min}^2}{T} \left(\frac{\eta_{\max}}{\eta_{\min}} \right)^{2t/T} \ln \left(\frac{\eta_{\max}}{\eta_{\min}} \right), \quad t \in [0, T]. \tag{12}$$

VESDE

1. SDE $dX_t = \sqrt{\beta(t)}dB_t$, where $0 < \eta_{\min} \le \eta_{\max}$ and

$$\beta(t) = \frac{2\eta_{\min}^2}{T}\left(\frac{\eta_{\max}}{\eta_{\min}}\right)^{2t/T} \ln\left(\frac{\eta_{\max}}{\eta_{\min}}\right), \quad t \in [0, T]. \tag{12}$$

2. $X_t = X_0 + \int_0^t \sqrt{\beta(s)}dB_s \quad \Longrightarrow \quad X_t \sim \mathcal{N}\Big(x_0, \eta^2(t)\mathbf{1}_d\Big)$, where

$$\eta(t) = \left(\int_0^t \beta(s)ds\right)^{\frac{1}{2}} = \eta_{\min}\sqrt{\left(\frac{\eta_{\max}}{\eta_{\min}}\right)^{2t/T} - 1}.$$

VESDE

1. SDE $dX_t = \sqrt{\beta(t)}dB_t$, where $0 < \eta_{\min} \le \eta_{\max}$ and

$$\beta(t) = \frac{2\eta_{\min}^2}{T}\left(\frac{\eta_{\max}}{\eta_{\min}}\right)^{2t/T} \ln\left(\frac{\eta_{\max}}{\eta_{\min}}\right), \quad t \in [0, T]. \tag{12}$$

2. $X_t = X_0 + \int_0^t \sqrt{\beta(s)}dB_s \quad \Longrightarrow \quad X_t \sim \mathcal{N}\Big(x_0, \eta^2(t)\mathbf{1}_d\Big)$, where

$$\eta(t) = \left(\int_0^t \beta(s)ds\right)^{\frac{1}{2}} = \eta_{\min}\sqrt{\left(\frac{\eta_{\max}}{\eta_{\min}}\right)^{2t/T} - 1}.$$

3. When $X_0 \sim \mathcal{N}(x_0, \eta_{\min}^2\mathbf{1}_d)$,

$$X_t = (x_0 + \eta_{\min} z) + \int_0^t \sqrt{\beta(s)}dB_s \sim \mathcal{N}(x_0, \tilde{\eta}^2(t)\mathbf{1}_d)$$

where $z \sim \mathcal{N}(0, \mathrm{I}_d)$ and

VESDE

1. SDE $dX_t = \sqrt{\beta(t)}dB_t$, where $0 < \eta_{\min} \le \eta_{\max}$ and

$$\beta(t) = \frac{2\eta_{\min}^2}{T}\left(\frac{\eta_{\max}}{\eta_{\min}}\right)^{2t/T}\ln\left(\frac{\eta_{\max}}{\eta_{\min}}\right), \quad t \in [0, T]\,. \tag{12}$$

2. $X_t = X_0 + \int_0^t \sqrt{\beta(s)}dB_s \quad \Longrightarrow \quad X_t \sim \mathcal{N}\Big(x_0, \eta^2(t)\mathbf{1}_d\Big)$, where

$$\eta(t) = \left(\int_0^t \beta(s)ds\right)^{\frac{1}{2}} = \eta_{\min}\sqrt{\left(\frac{\eta_{\max}}{\eta_{\min}}\right)^{2t/T} - 1}\,.$$

3. When $X_0 \sim \mathcal{N}(x_0, \eta_{\min}^2\mathbf{1}_d)$,

$$X_t = (x_0 + \eta_{\min} z) + \int_0^t \sqrt{\beta(s)}dB_s \sim \mathcal{N}(x_0, \tilde{\eta}^2(t)\mathbf{1}_d)$$

where $z \sim \mathcal{N}(0, \mathrm{I}_d)$ and

$$\tilde{\eta}(t) = \sqrt{\eta^2(t) + \eta_{\min}^2} = \eta_{\min}\left(\frac{\eta_{\max}}{\eta_{\min}}\right)^{t/T}.$$

VESDE

1. Forward process:

$$X_t = (x_0 + \eta_{\min} z) + \int_0^t \sqrt{\beta(s)} dB_s \sim \mathcal{N}(x_0, \tilde{\eta}^2(t)\mathbf{1}_d)$$

whose density is $\tilde{p}(x, t|x_0) = \left(2\pi\tilde{\eta}^2(t)\right)^{-\frac{d}{2}} \mathrm{e}^{-\frac{1}{2}|x-x_0|^2/\tilde{\eta}^2(t)}$, where $\tilde{\eta}(t) = \eta_{\min}\left(\frac{\eta_{\max}}{\eta_{\min}}\right)^{t/T}$.

VESDE

1. Forward process:

$$X_t = (x_0 + \eta_{\min} z) + \int_0^t \sqrt{\beta(s)} dB_s \sim \mathcal{N}(x_0, \tilde{\eta}^2(t)\mathbf{1}_d)$$

whose density is $\tilde{p}(x, t|x_0) = \left(2\pi\tilde{\eta}^2(t)\right)^{-\frac{d}{2}} \mathrm{e}^{-\frac{1}{2}|x-x_0|^2/\tilde{\eta}^2(t)}$, where $\tilde{\eta}(t) = \eta_{\min}\left(\frac{\eta_{\max}}{\eta_{\min}}\right)^{t/T}$.

2. $x \sim \tilde{p}(\cdot, t|x_0) \Longrightarrow x = x_0 + \tilde{\eta}(t) z$, where $z \sim \mathcal{N}(0, \mathrm{I}_d)$.

VESDE

1. Forward process:

$$X_t = (x_0 + \eta_{\min} z) + \int_0^t \sqrt{\beta(s)} dB_s \sim \mathcal{N}(x_0, \tilde{\eta}^2(t)\mathbf{1}_d)$$

whose density is $\tilde{p}(x,t|x_0) = \left(2\pi\tilde{\eta}^2(t)\right)^{-\frac{d}{2}} \mathrm{e}^{-\frac{1}{2}|x-x_0|^2/\tilde{\eta}^2(t)}$, where $\tilde{\eta}(t) = \eta_{\min}\left(\frac{\eta_{\max}}{\eta_{\min}}\right)^{t/T}$.

2. $x \sim \tilde{p}(\cdot, t|x_0) \implies x = x_0 + \tilde{\eta}(t)z$, where $z \sim \mathcal{N}(0, \mathrm{I}_d)$.
3. Choose $w(t) = \tilde{\eta}^2(t)$, we obtain

$$\mathrm{Loss}(u) = \mathbb{E}_{t\sim U([0,T])}\mathbb{E}_{x_0\sim p_0}\mathbb{E}_{x\sim\tilde{p}(\cdot,t|x_0)}\left[\left(\frac{1}{2}|u(x,t)|^2 - u(x,t)\cdot\nabla \ln \tilde{p}(x,t|x_0)\right)w(t)\right]$$

VESDE

1. Forward process:

$$X_t = (x_0 + \eta_{\min} z) + \int_0^t \sqrt{\beta(s)} dB_s \sim \mathcal{N}(x_0, \tilde{\eta}^2(t)\mathbf{1}_d)$$

whose density is $\tilde{p}(x, t|x_0) = \left(2\pi\tilde{\eta}^2(t)\right)^{-\frac{d}{2}} \mathrm{e}^{-\frac{1}{2}|x-x_0|^2/\tilde{\eta}^2(t)}$, where $\tilde{\eta}(t) = \eta_{\min}\left(\frac{\eta_{\max}}{\eta_{\min}}\right)^{t/T}$.

2. $x \sim \tilde{p}(\cdot, t|x_0) \Longrightarrow x = x_0 + \tilde{\eta}(t)z$, where $z \sim \mathcal{N}(0, \mathrm{I}_d)$.
3. Choose $w(t) = \tilde{\eta}^2(t)$, we obtain

$$\begin{aligned}\mathrm{Loss}(u) =& \mathbb{E}_{t\sim U([0,T])}\mathbb{E}_{x_0\sim p_0}\mathbb{E}_{x\sim\tilde{p}(\cdot,t|x_0)}\left[\left(\frac{1}{2}|u(x,t)|^2 - u(x,t)\cdot\nabla\ln\tilde{p}(x,t|x_0)\right)w(t)\right]\\ =& \mathbb{E}_{t\sim U([0,T])}\mathbb{E}_{x_0\sim p_0}\mathbb{E}_{x\sim\tilde{p}(\cdot,t|x_0)}\left[\left(\frac{1}{2}|u(x,t)|^2 + u(x,t)\cdot\frac{x-x_0}{\tilde{\eta}(t)^2}\right)\tilde{\eta}^2(t)\right]\end{aligned}$$

VESDE

1. Forward process:

$$X_t = (x_0 + \eta_{\min} z) + \int_0^t \sqrt{\beta(s)} dB_s \sim \mathcal{N}(x_0, \tilde{\eta}^2(t)\mathbf{1}_d)$$

whose density is $\tilde{p}(x, t|x_0) = \left(2\pi\tilde{\eta}^2(t)\right)^{-\frac{d}{2}} \mathrm{e}^{-\frac{1}{2}|x-x_0|^2/\tilde{\eta}^2(t)}$, where $\tilde{\eta}(t) = \eta_{\min}\left(\frac{\eta_{\max}}{\eta_{\min}}\right)^{t/T}$.

2. $x \sim \tilde{p}(\cdot, t|x_0) \Longrightarrow x = x_0 + \tilde{\eta}(t)z$, where $z \sim \mathcal{N}(0, \mathrm{I}_d)$.
3. Choose $w(t) = \tilde{\eta}^2(t)$, we obtain

$$\begin{aligned}
\mathrm{Loss}(u) =& \mathbb{E}_{t\sim U([0,T])}\mathbb{E}_{x_0\sim p_0}\mathbb{E}_{x\sim\tilde{p}(\cdot,t|x_0)}\left[\left(\frac{1}{2}|u(x,t)|^2 - u(x,t)\cdot\nabla \ln \tilde{p}(x,t|x_0)\right)w(t)\right] \\
=& \mathbb{E}_{t\sim U([0,T])}\mathbb{E}_{x_0\sim p_0}\mathbb{E}_{x\sim\tilde{p}(\cdot,t|x_0)}\left[\left(\frac{1}{2}|u(x,t)|^2 + u(x,t)\cdot\frac{x-x_0}{\tilde{\eta}(t)^2}\right)\tilde{\eta}^2(t)\right] \\
=& \mathbb{E}_{t\sim U([0,T])}\mathbb{E}_{x_0\sim p_0}\mathbb{E}_{z\sim\mathcal{N}(0,\mathrm{I}_d)}\left[\left(\frac{1}{2}|u(x,t)|^2 + \frac{u(x,t)\cdot z}{\tilde{\eta}(t)}\right)\tilde{\eta}^2(t)\right]
\end{aligned}$$