Flow-matching generative models

June 11, 2025

Part 1: Recall the previous lecture

Generative modeling

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:
   $Y_0 \sim p(\cdot, T) \to Y_T \sim p_{\text{target}}$.

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

Time-reversal

Define

$$q(x,t) = p(x, T-t), \quad \forall\, t \in [0,T]\,. \tag{1}$$

Theorem

q is the probability density of the following SDE

$$\begin{aligned} dY_t =& f^-(Y_t,t)\,dt + \sigma^-(t)dB_t\,, \quad t\in[0,T]\,,\\ Y_0 \sim& p(\cdot, T)\,, \end{aligned} \tag{2}$$

where

$$\begin{aligned} f^-(x,t) =& -f(x,T-t) + \sigma^2(T-t)\nabla \ln p(x, T-t)\\ \sigma^-(t) =& \sigma(T-t)\,. \end{aligned} \tag{3}$$

Loss objective

Proposition

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

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

where C_2, C_3 are constants independent of u.

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

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

Part 2: Dirac delta function

Dirac delta function

Definition

We define the Dirac delta function $\delta(x)$ such that

$$\int_{\mathbb{R}^d} \delta(x) f(x) dx = f(0) \tag{4}$$

for any bounded continuous function $f : \mathbb{R}^d \to \mathbb{R}$.

Dirac delta function

Definition

We define the Dirac delta function $\delta(x)$ such that

$$\int_{\mathbb{R}^d} \delta(x) f(x) dx = f(0) \tag{4}$$

for any bounded continuous function $f : \mathbb{R}^d \to \mathbb{R}$.

Lemma

We have the following two identities.

1. *For any $x_0 \in \mathbb{R}^d$, we have*

$$\int_{\mathbb{R}^d} \delta(x - x_0) f(x) dx = f(x_0)\,. \tag{5}$$

2. $\int_{\mathbb{R}^d} \delta(x) dx = 1$.

Dirac delta function

Remark

1. *$\delta(x - x_0)dx$ can be viewed as the probability distribution where the probability is one at $x = x_0$ and is zero elsewhere.*

Dirac delta function

Remark

1. *$\delta(x - x_0)dx$ can be viewed as the probability distribution where the probability is one at $x = x_0$ and is zero elsewhere.*
2. *Intuitively, delta function can be thought as the limit of Gaussian density*

$$\psi_\sigma(x) = (2\pi\sigma^2)^{-\frac{d}{2}} \mathrm{e}^{-\frac{|x|^2}{2\sigma^2}}, \quad x \in \mathbb{R}^d \tag{6}$$

as $\sigma \to 0+$. In fact, for a bounded smooth function $f : \mathbb{R}^d \to \mathbb{R}$,

$$\lim_{\sigma \to 0+} \int_{\mathbb{R}^d} f(x)\psi_\sigma(x)dx = f(0) = \int_{\mathbb{R}^d} \delta(x)f(x)dx\,. \tag{7}$$

Dirac delta function

Consider a C^1-smooth map $\xi : \mathbb{R}^d \to \mathbb{R}^k$, where $1 \leq k < d$. We will often use the integral

$$\int_{\mathbb{R}^d} f(x)\delta(z - \xi(x))dx\,, \tag{8}$$

where $z \in \mathbb{R}^k$.

Dirac delta function

Consider a C^1-smooth map $\xi : \mathbb{R}^d \to \mathbb{R}^k$, where $1 \leq k < d$. We will often use the integral

$$\int_{\mathbb{R}^d} f(x)\delta(z - \xi(x))dx\,, \tag{8}$$

where $z \in \mathbb{R}^k$.

Lemma

For two test functions $f : \mathbb{R}^d \to \mathbb{R}$ and $g : \mathbb{R}^k \to \mathbb{R}$, we have

$$\int_{\mathbb{R}^d} f(x)g(\xi(x))dx = \int_{\mathbb{R}^k} \Big(\int_{\mathbb{R}^d} f(x)\delta(z - \xi(x))dx\Big) g(z)dz\,. \tag{9}$$

Linear map ξ

1. Let $x = (y, z) \in \mathbb{R}^d$, where $y \in \mathbb{R}^{d-k}$ and $z \in \mathbb{R}^k$.
2. Linear map $\xi(x) = \xi(y, z) = z$, for $x \in \mathbb{R}^d$.

Linear map ξ

1. Let $x = (y, z) \in \mathbb{R}^d$, where $y \in \mathbb{R}^{d-k}$ and $z \in \mathbb{R}^k$.
2. Linear map $\xi(x) = \xi(y, z) = z$, for $x \in \mathbb{R}^d$.

Then,

$$\begin{aligned}
&\int_{\mathbb{R}^d} f(x')\delta(z - \xi(x'))dx' \\
&= \int_{\mathbb{R}^k} \int_{\mathbb{R}^{d-k}} f(y', z')\delta(z - z')dy'dz' \\
&= \int_{\mathbb{R}^{d-k}} \Big(\int_{\mathbb{R}^k} f(y', z')\delta(z - z')dz' \Big) dy' \\
&= \int_{\mathbb{R}^d} f(y', z)dy' \,.
\end{aligned}$$

Part 3: Conditional expectation

Marginal density

Assume that X is a random variable in $\mathbb{R}^d$ whose probability density is $p(x)$. Then, $Z = \xi(X)$ is a random variable taking values in $\mathbb{R}^k$.

Marginal density

Assume that X is a random variable in $\mathbb{R}^d$ whose probability density is $p(x)$. Then, $Z = \xi(X)$ is a random variable taking values in $\mathbb{R}^k$.

Lemma

The probability density of $Z = \xi(X)$ is given by

$$Q(z) = \int_{\mathbb{R}^d} p(x)\delta(z - \xi(x))dx\,, \quad z \in \mathbb{R}^k\,. \tag{10}$$

Conditional expectation

Definition

Let $z \in \mathbb{R}^k$ and $f : \mathbb{R}^d \to \mathbb{R}$, we define the expectation of $f(X)$ conditioned on the event that $\xi(X) = z$ as

$$\mathbb{E}(f(X)|\xi(X) = z) = \frac{\int_{\mathbb{R}^d} f(x)\delta(\xi(x) - z)p(x)dx}{Q(z)}, \tag{11}$$

where $Q(z)$ is defined in (10).

Conditional expectation, linear case

Consider the linear map $\xi(x) = z$, where $x = (y, z) \in \mathbb{R}^d$. We can derive

$$\begin{aligned} &\int_{\mathbb{R}^d} f(x)\delta(\xi(x) - z)p(x)dx = \int_{\mathbb{R}^{d-k}} f(y, z)p(y, z)dy \\ &Q(z) = \int_{\mathbb{R}^{d-k}} p(y, z)dy\,. \end{aligned} \tag{12}$$

Conditional expectation, linear case

Consider the linear map $\xi(x) = z$, where $x = (y, z) \in \mathbb{R}^d$. We can derive

$$\begin{aligned} &\int_{\mathbb{R}^d} f(x)\delta(\xi(x) - z)p(x)dx = \int_{\mathbb{R}^{d-k}} f(y, z)p(y, z)dy \\ &Q(z) = \int_{\mathbb{R}^{d-k}} p(y, z)dy\,. \end{aligned} \tag{12}$$

Therefore, in this case, the conditional expectation is

$$\mathbb{E}(f(X)|\xi(X) = z) = \frac{\int_{\mathbb{R}^{d-k}} f(y, z)p(y, z)dy}{\int_{\mathbb{R}^{d-k}} p(y, z)dy}\,. \tag{13}$$

Conditional expectation

Proposition (Law of total expectation)

For a test function $f : \mathbb{R}^d \to \mathbb{R}$, we have

$$\mathbb{E}(f(X)) = \mathbb{E}_{z\sim Q}\Big(\mathbb{E}\big(f(X)\big|\xi(X) = z\big)\Big)\,. \tag{14}$$

Conditional expectation

Proposition (Law of total expectation)

For a test function $f : \mathbb{R}^d \to \mathbb{R}$, we have

$$\mathbb{E}(f(X)) = \mathbb{E}_{z\sim Q}\Big(\mathbb{E}\big(f(X)\big|\xi(X) = z\big)\Big). \tag{14}$$

Proposition

Given a function $f : \mathbb{R}^d \to \mathbb{R}$, we define

$$\widetilde{f}(z) = \mathbb{E}\big(f(X)\big|\xi(X) = z\big), \quad z \in \mathbb{R}^k. \tag{15}$$

Then, for any $g : \mathbb{R}^k \to \mathbb{R}$, we have

$$\mathbb{E}\Big(\big|f(X) - \widetilde{f}(\xi(X))\big|^2\Big) \le \mathbb{E}\Big(|f(X) - g(\xi(X))|^2\Big). \tag{16}$$

Part 4: Flow-based generative models

Setup

1. Target density $p_1 = p_{\text{target}}$ on $\mathbb{R}^d$.
2. Prior density p_0 on $\mathbb{R}^d$, typically a Gaussian density, is chosen.
3. Define $p(\cdot, t)$ as the probability density of

$$X_t = (1 - t)X_0 + tX_1\,, \quad \text{where } X_0 \sim p_0 \text{ and } X_1 \sim p_1\,. \tag{17}$$

Then,

$$p(\cdot, 0) = p_0\,, \quad p(\cdot, 1) = p_1\,. \tag{18}$$

Setup

1. Target density $p_1 = p_{\text{target}}$ on $\mathbb{R}^d$.
2. Prior density p_0 on $\mathbb{R}^d$, typically a Gaussian density, is chosen.
3. Define $p(\cdot, t)$ as the probability density of

$$X_t = (1-t)X_0 + tX_1 \,, \quad \text{where } X_0 \sim p_0 \text{ and } X_1 \sim p_1 \,. \tag{17}$$

   Then,

$$p(\cdot, 0) = p_0 \,, \quad p(\cdot, 1) = p_1 \,. \tag{18}$$

Idea: learn an ODE

$$\frac{dY_t}{dt} = u(Y_t, t) \,, \quad t \in [0, 1] \tag{19}$$

such that, when $Y_0 \sim p_0$, then $Y_t \sim p(\cdot, t)$ for any $t \in [0, 1]$.

Continuity equation

Denoted by $q(\cdot, t)$ the probability density of

$$\frac{dY_t}{dt} = u(Y_t, t)\,, \quad t \in [0, 1] \tag{20}$$

Recall that q satisfies the continuity equation

$$\frac{\partial q}{\partial t} + \mathrm{div}(uq) = 0\,. \tag{21}$$

Equation of p

Let $f \in \mathbb{R}^d \times [0, 1] \to \mathbb{R}$ be C^1-smooth with compact support and $f(\cdot, 0) = f(\cdot, 1) = 0$.

Equation of p

Let $f \in \mathbb{R}^d \times [0,1] \to \mathbb{R}$ be C^1-smooth with compact support and $f(\cdot,0) = f(\cdot,1) = 0$.

$$\int_0^1 \int_{\mathbb{R}^d} f(x,t) \frac{\partial p(x,t)}{\partial t} dxdt = -\int_0^1 \int_{\mathbb{R}^d} \frac{\partial f(x,t)}{\partial t} p(x,t) dxdt$$

Equation of p

Let $f \in \mathbb{R}^d \times [0,1] \to \mathbb{R}$ be C^1-smooth with compact support and $f(\cdot, 0) = f(\cdot, 1) = 0$.

$$\int_0^1 \int_{\mathbb{R}^d} f(x,t) \frac{\partial p(x,t)}{\partial t} dxdt = -\int_0^1 \int_{\mathbb{R}^d} \frac{\partial f(x,t)}{\partial t} p(x,t) dxdt$$
$$= -\int_0^1 \mathbb{E}\Big(\frac{\partial f}{\partial t}(X_t, t)\Big) dt$$

Equation of p

Let $f \in \mathbb{R}^d \times [0,1] \to \mathbb{R}$ be C^1-smooth with compact support and $f(\cdot,0) = f(\cdot,1) = 0$.

$$\begin{aligned}
\int_0^1 \int_{\mathbb{R}^d} f(x,t)\frac{\partial p(x,t)}{\partial t} dxdt &= -\int_0^1 \int_{\mathbb{R}^d} \frac{\partial f(x,t)}{\partial t} p(x,t) dxdt \\
&= -\int_0^1 \mathbb{E}\Big(\frac{\partial f}{\partial t}(X_t,t)\Big) dt \\
&= -\int_0^1 \mathbb{E}\Big(\frac{df}{dt}(X_t,t) - \nabla f(X_t,t) \cdot \frac{dX_t}{dt}\Big) dt
\end{aligned}$$

Equation of p

Let $f \in \mathbb{R}^d \times [0,1] \to \mathbb{R}$ be C^1-smooth with compact support and $f(\cdot, 0) = f(\cdot, 1) = 0$.

$$\begin{aligned}
\int_0^1 \int_{\mathbb{R}^d} f(x,t) \frac{\partial p(x,t)}{\partial t} dxdt &= -\int_0^1 \int_{\mathbb{R}^d} \frac{\partial f(x,t)}{\partial t} p(x,t) dxdt \\
&= -\int_0^1 \mathbb{E}\Big(\frac{\partial f}{\partial t}(X_t, t)\Big) dt \\
&= -\int_0^1 \mathbb{E}\Big(\frac{df}{dt}(X_t, t) - \nabla f(X_t, t) \cdot \frac{dX_t}{dt}\Big) dt \\
&= \int_0^1 \mathbb{E}\Big(\nabla f(X_t, t) \cdot (X_1 - X_0)\Big) dt
\end{aligned}$$

Equation of p

Let $f \in \mathbb{R}^d \times [0,1] \to \mathbb{R}$ be C^1-smooth with compact support and $f(\cdot, 0) = f(\cdot, 1) = 0$.

$$
\begin{aligned}
\int_0^1 \int_{\mathbb{R}^d} f(x,t) \frac{\partial p(x,t)}{\partial t} dxdt &= -\int_0^1 \int_{\mathbb{R}^d} \frac{\partial f(x,t)}{\partial t} p(x,t) dxdt \\
&= -\int_0^1 \mathbb{E}\Big(\frac{\partial f}{\partial t}(X_t, t)\Big) dt \\
&= -\int_0^1 \mathbb{E}\Big(\frac{df}{dt}(X_t, t) - \nabla f(X_t, t) \cdot \frac{dX_t}{dt}\Big) dt \\
&= \int_0^1 \mathbb{E}\Big(\nabla f(X_t, t) \cdot (X_1 - X_0)\Big) dt \\
&= \int_0^1 \mathbb{E}_{x \sim p(x,t)} \Big[\mathbb{E}\Big(\nabla f(X_t, t) \cdot (X_1 - X_0) \Big| X_t = x\Big)\Big] dt
\end{aligned}
$$

Equation of p

Let $f \in \mathbb{R}^d \times [0,1] \to \mathbb{R}$ be C^1-smooth with compact support and $f(\cdot, 0) = f(\cdot, 1) = 0$.

$$
\begin{aligned}
\int_0^1 \int_{\mathbb{R}^d} f(x,t) \frac{\partial p(x,t)}{\partial t} dxdt &= -\int_0^1 \int_{\mathbb{R}^d} \frac{\partial f(x,t)}{\partial t} p(x,t) dxdt \\
&= -\int_0^1 \mathbb{E}\Big(\frac{\partial f}{\partial t}(X_t, t)\Big) dt \\
&= -\int_0^1 \mathbb{E}\Big(\frac{df}{dt}(X_t, t) - \nabla f(X_t, t) \cdot \frac{dX_t}{dt}\Big) dt \\
&= \int_0^1 \mathbb{E}\Big(\nabla f(X_t, t) \cdot (X_1 - X_0)\Big) dt \\
&= \int_0^1 \mathbb{E}_{x \sim p(x,t)} \Big[\mathbb{E}\Big(\nabla f(X_t, t) \cdot (X_1 - X_0) \Big| X_t = x\Big)\Big] dt \\
&= \int_0^1 \int_{\mathbb{R}^d} \nabla f(x,t) \cdot \mathbb{E}\big(X_1 - X_0 \big| X_t = x\big) p(x,t) dxdt
\end{aligned}
$$

Equation of p

Let $f \in \mathbb{R}^d \times [0,1] \to \mathbb{R}$ be C^1-smooth with compact support and $f(\cdot,0) = f(\cdot,1) = 0$.

$$\begin{aligned}
\int_0^1 \int_{\mathbb{R}^d} f(x,t)\frac{\partial p(x,t)}{\partial t} dxdt &= -\int_0^1 \int_{\mathbb{R}^d} \frac{\partial f(x,t)}{\partial t} p(x,t)dxdt \\
&= -\int_0^1 \mathbb{E}\Big(\frac{\partial f}{\partial t}(X_t,t)\Big)dt \\
&= -\int_0^1 \mathbb{E}\Big(\frac{df}{dt}(X_t,t) - \nabla f(X_t,t)\cdot\frac{dX_t}{dt}\Big)dt \\
&= \int_0^1 \mathbb{E}\Big(\nabla f(X_t,t)\cdot(X_1 - X_0)\Big)dt \\
&= \int_0^1 \mathbb{E}_{x\sim p(x,t)}\Big[\mathbb{E}\Big(\nabla f(X_t,t)\cdot(X_1 - X_0)\Big|X_t = x\Big)\Big]dt \\
&= \int_0^1 \int_{\mathbb{R}^d} \nabla f(x,t)\cdot\mathbb{E}\big(X_1 - X_0\big|X_t = x\big)p(x,t)dxdt \\
&= -\int_0^1 \int_{\mathbb{R}^d} f(x,t)\mathrm{div}\Big(\mathbb{E}\big(X_1 - X_0\big|X_t = x\big)p(x,t)\Big)dxdt\,,
\end{aligned}$$

Equation of p

To summarize, we have obtained

Proposition

The probability density $p(x,t)$ of X_t solves the equation

$$\frac{\partial p(x,t)}{\partial t} + \mathrm{div}\Big(\mathbb{E}(X_1 - X_0 | X_t = x) p(x,t)\Big) = 0$$

in a weak sense.

Equation of p

To summarize, we have obtained

Proposition

The probability density $p(x,t)$ of X_t solves the equation

$$\frac{\partial p(x,t)}{\partial t} + \mathrm{div}\Big(\mathbb{E}\big(X_1 - X_0 | X_t = x\big) p(x,t)\Big) = 0$$

in a weak sense.

Therefore, we can choose $u(x,t) = \mathbb{E}\big(X_1 - X_0 | X_t = x\big)$.

Loss objective

Idea: Define a loss objective, whose minimizer is $\mathbb{E}\big(X_1 - X_0 | X_t = x\big)$.

Loss objective

Idea: Define a loss objective, whose minimizer is $\mathbb{E}\big(X_1 - X_0 | X_t = x\big)$.

$$\int_0^1 \mathbb{E}_{x\sim p(x,t)}\Big(\big|u(x,t) - \mathbb{E}\big(X_1 - X_0\big|X_t = x\big)\big|^2\Big)dt$$

Loss objective

Idea: Define a loss objective, whose minimizer is $\mathbb{E}\big(X_1 - X_0 | X_t = x\big)$.

$$\int_0^1 \mathbb{E}_{x\sim p(x,t)}\Big(\big|u(x,t) - \mathbb{E}\big(X_1 - X_0\big|X_t = x\big)\big|^2\Big)dt$$
$$= \int_0^1 \mathbb{E}_{x\sim p(x,t)}\Big(|u(x,t)|^2 - 2u(x,t)\cdot\mathbb{E}\big(X_1 - X_0\big|X_t = x\big)\Big)dt + C$$

Loss objective

Idea: Define a loss objective, whose minimizer is $\mathbb{E}\big(X_1 - X_0 | X_t = x\big)$.

$$\int_0^1 \mathbb{E}_{x\sim p(x,t)}\Big(\big|u(x,t) - \mathbb{E}\big(X_1 - X_0\big|X_t = x\big)\big|^2\Big)dt$$

$$= \int_0^1 \mathbb{E}_{x\sim p(x,t)}\Big(|u(x,t)|^2 - 2u(x,t)\cdot\mathbb{E}\big(X_1 - X_0\big|X_t = x\big)\Big)dt + C$$

$$= \int_0^1 \mathbb{E}_{x\sim p(x,t)}\Big[\mathbb{E}\Big(|u(X_t,t)|^2 - 2u(X_t,t)\cdot(X_1 - X_0)\Big|X_t = x\Big)\Big]dt + C$$

Loss objective

Idea: Define a loss objective, whose minimizer is $\mathbb{E}\big(X_1 - X_0 | X_t = x\big)$.

$$\int_0^1 \mathbb{E}_{x\sim p(x,t)}\Big(\big|u(x,t) - \mathbb{E}\big(X_1 - X_0\big|X_t = x\big)\big|^2\Big)dt$$
$$= \int_0^1 \mathbb{E}_{x\sim p(x,t)}\Big(|u(x,t)|^2 - 2u(x,t)\cdot\mathbb{E}\big(X_1 - X_0\big|X_t = x\big)\Big)dt + C$$
$$= \int_0^1 \mathbb{E}_{x\sim p(x,t)}\Big[\mathbb{E}\Big(|u(X_t,t)|^2 - 2u(X_t,t)\cdot(X_1 - X_0)\Big|X_t = x\Big)\Big]dt + C$$
$$= \int_0^1 \mathbb{E}_{X_0\sim p_0, X_1\sim p_1}\Big(\big|u((1-t)X_0 + tX_1, t) - (X_1 - X_0)\big|^2\Big)dt + C$$

Loss objective

Idea: Define a loss objective, whose minimizer is $\mathbb{E}(X_1 - X_0 | X_t = x)$.

$$\int_0^1 \mathbb{E}_{x \sim p(x,t)} \left(\left| u(x,t) - \mathbb{E}(X_1 - X_0 \middle| X_t = x) \right|^2 \right) dt$$
$$= \int_0^1 \mathbb{E}_{x \sim p(x,t)} \left(|u(x,t)|^2 - 2u(x,t) \cdot \mathbb{E}(X_1 - X_0 \middle| X_t = x) \right) dt + C$$
$$= \int_0^1 \mathbb{E}_{x \sim p(x,t)} \left[\mathbb{E}\left(|u(X_t,t)|^2 - 2u(X_t,t) \cdot (X_1 - X_0) \middle| X_t = x \right) \right] dt + C$$
$$= \int_0^1 \mathbb{E}_{X_0 \sim p_0, X_1 \sim p_1} \left(\left| u((1-t)X_0 + tX_1, t) - (X_1 - X_0) \right|^2 \right) dt + C$$

Final loss:

$$\mathrm{Loss}(u) = \mathbb{E}_{t \sim U[0,1]} \mathbb{E}_{X_0 \sim p_0, X_1 \sim p_1} \left(\left| u((1-t)X_0 + tX_1, t) - (X_1 - X_0) \right|^2 \right). \quad (22)$$

Extension: General "interpolant"

Define

$$X_t = I_t(X_0, X_1)\,, \quad \text{where } X_0 \sim p_0, \quad X_1 \sim p_1\,, \tag{23}$$

with $I_0(x, y) = x$ and $I_1(x, y) = y$.

Extension: General "interpolant"

Define

$$X_t = I_t(X_0, X_1)\,, \quad \text{where } X_0 \sim p_0, \quad X_1 \sim p_1\,, \tag{23}$$

with $I_0(x, y) = x$ and $I_1(x, y) = y$.

Proposition

Density $p(x,t)$ of X_t in (23) solves

$$\frac{\partial p(x,t)}{\partial t} + \mathrm{div}\Big(\mathbb{E}\big(\partial_t I_t(X_0, X_1)\big|X_t = x\big)p(x,t)\Big) = 0\,.$$

Extension: General "interpolant"

Define

$$X_t = I_t(X_0, X_1)\,, \quad \text{where } X_0 \sim p_0, \quad X_1 \sim p_1\,, \tag{23}$$

with $I_0(x, y) = x$ and $I_1(x, y) = y$.

Proposition

Density $p(x, t)$ of X_t in (23) solves

$$\frac{\partial p(x,t)}{\partial t} + \mathrm{div}\Big(\mathbb{E}\big(\partial_t I_t(X_0, X_1)\big|X_t = x\big)p(x,t)\Big) = 0\,.$$

Loss function:

$$\mathrm{Loss}(u) = \mathbb{E}_{t\sim U[0,1]}\mathbb{E}_{X_0\sim p_0, X_1\sim p_1}\Big(\big|u\big(I_t(X_0, X_1), t\big) - \partial_t I_t(X_0, X_1)\big|^2\Big)\,.$$