Diffusion Models Chapter 1: Reverse-Time SDE

Generative AI and Foundation Models

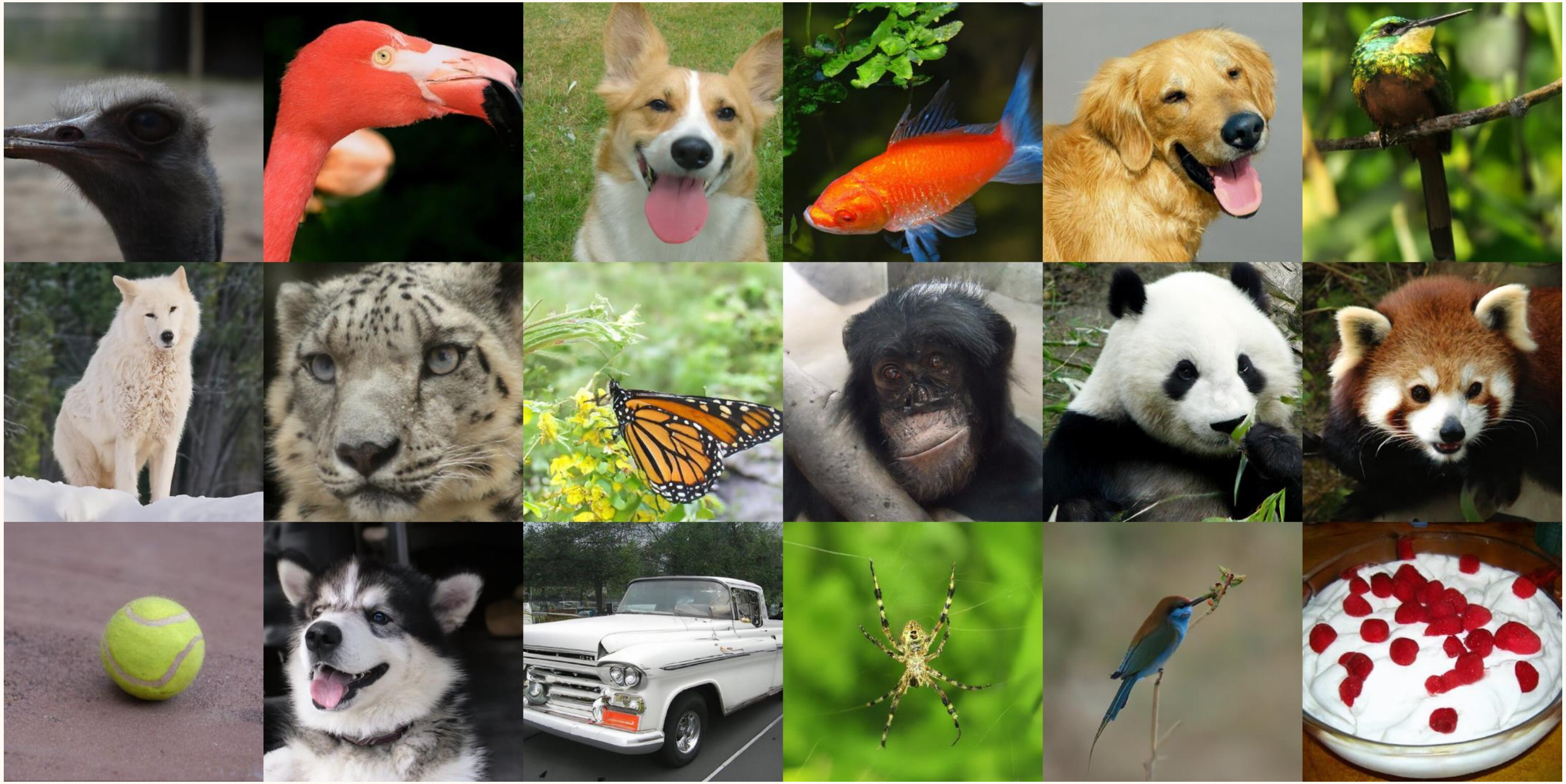
Spring 2024

Department of Mathematical Sciences

Ernest K. Ryu

Seoul National University

Diffusion models are SOTA



P. Dhariwal and A. Nichol, Diffusion models beat GANs on image synthesis, *NeurIPS*, 2021.

Ordinary differential equation

Consider the ordinary differential equation (ODE)

$$\frac{dX}{dt}(t) = f(X(t), t)$$

which we also express as

$$dX_t = f(X_t, t)dt$$

where $X(t), f(X(t), t) \in \mathbb{R}^d$. Then, $\{X(t)\}_t$ is a deterministic curve.

We can think of the ODE as the limit

$$X_{k+1} = X_k + \Delta t f(X_k, k\Delta t), \quad k = 0, 1, \ldots$$

under $\Delta t \to 0$, where $t = k\Delta t$. Precisely, $\{X_{\lfloor t/\Delta t \rfloor}\}_t \to \{X_t\}_t$ uniformly on compact intervals.

Solution for ODE

$\{X(t)\}_{t=0}^{T}$ solves ODE if it satisfies the

- differential form of the ODE

$$\frac{dX}{dt}(t) = f(X(t), t)$$

- or the integral form of the ODE

$$X(t) = X_0 + \int_0^t f(X(s), s) ds$$

- Example:

$$\frac{dX}{dt} = -\beta X$$

$$X(t) = X(0) e^{-\beta t}$$

Stochastic differential equation

Consider the stochastic differential equation (SDE)

$$dX_t = f(X_t, t)dt + g(t)dW_t$$

where $X_t(t), f(X_t, t) \in \mathbb{R}^d$, $g(t) \in \mathbb{R}^{d\times d}$, and W_t is a d-dimensional Brownian motion or Wiener process. $\{X_t\}_t$ is a random process. (We can allow g to also depend on X_t, but this makes the equations more complicated.)

We can think of the SDE as the limit

$$X_{k+1} = X_k + \Delta t f(X_k, k\Delta t) + g(k\Delta t)\sqrt{\Delta t}Z_k, \qquad k = 0, 1, \ldots$$

under $\Delta t \to 0$, where $t = k\ \Delta t$ and $Z_0, Z_1, \ldots \sim \mathcal{N}(0, I)$. Precisely, $\{X_{\lfloor t/\Delta t \rfloor}\}_t \overset{\mathcal{D}}{\to} \{X_t\}_t$ on compact intervals.

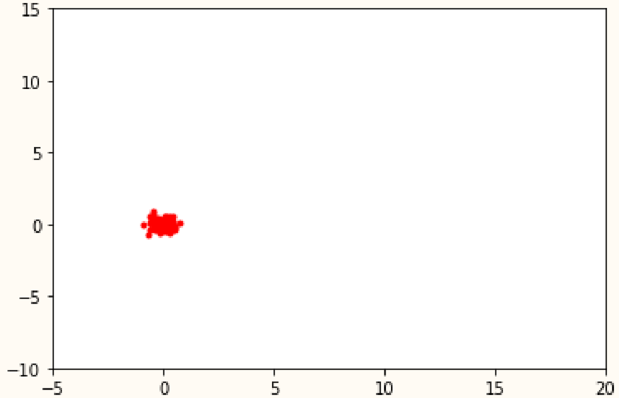
B. Øksendal, *Stochastic Differential Equations: An Introduction with Applications*, 6th edition, Springer, 2003.

Example: Forward

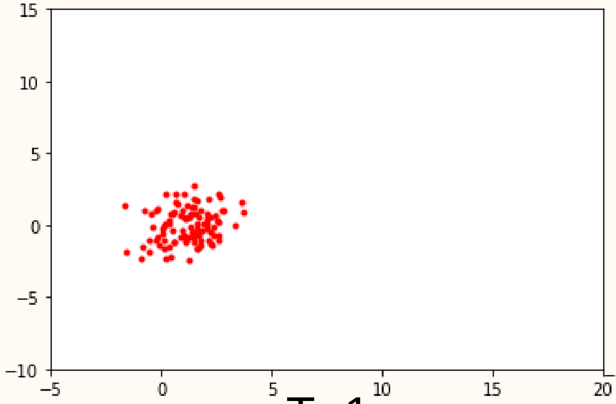
$$P_0 \sim \mathcal{N}\left(\begin{bmatrix}0\\0\end{bmatrix}, 0.1\begin{bmatrix}1 & 0\\0 & 1\end{bmatrix}\right)$$

$$f(x,t) = \begin{bmatrix}1\\0\end{bmatrix}, g(x,t) = \begin{bmatrix}1 & 0\\0 & 1\end{bmatrix}$$

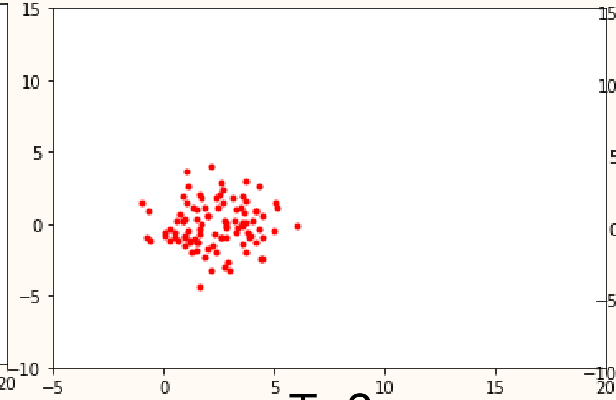
$$P_t \sim \mathcal{N}\left(\begin{bmatrix}t\\0\end{bmatrix}, (0.1+t)\begin{bmatrix}1 & 0\\0 & 1\end{bmatrix}\right)$$



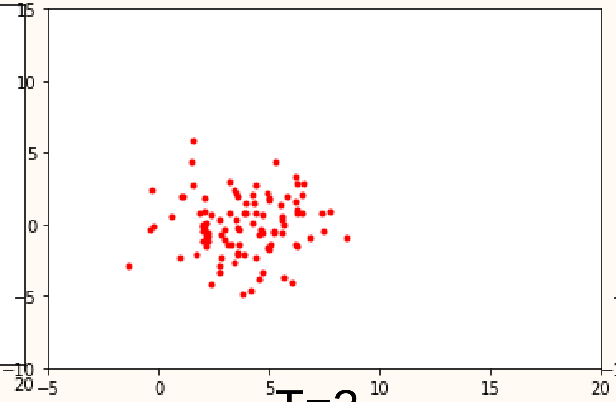
T=0



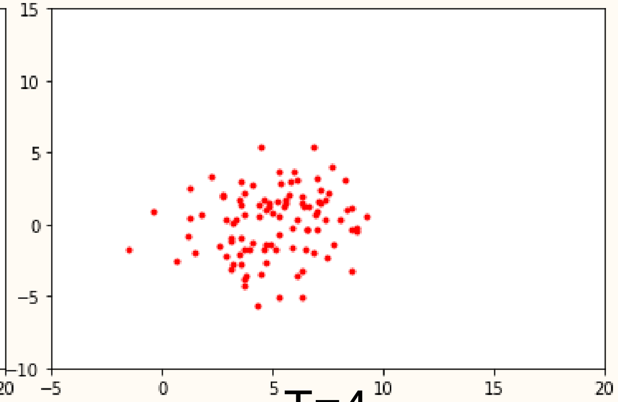
T=1



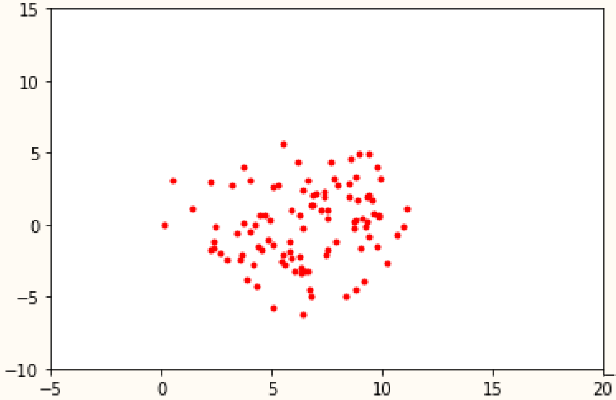
T=2



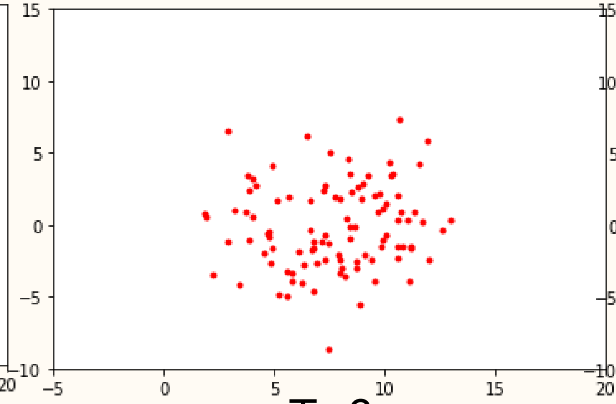
T=3



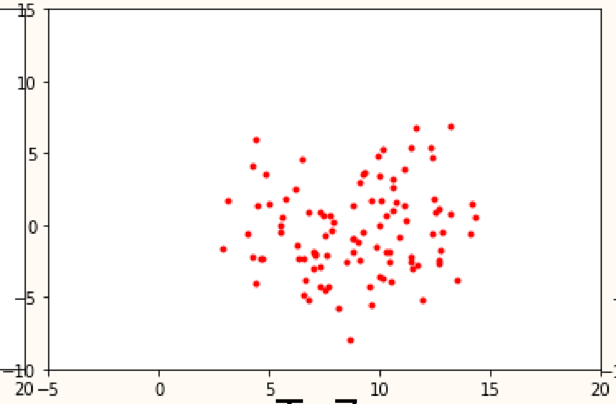
T=4



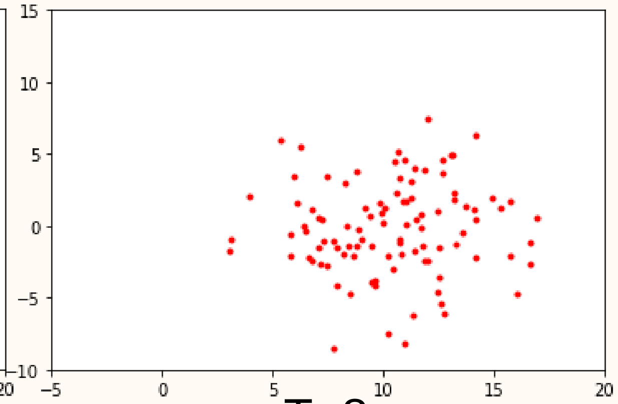
T=5



T=6



T=7



T=8

Solution for SDE

$\{X_t\}_{t=0}^{T}$ is a solution path for SDE if $\{X_t\}_{t=0}^{T}$ is nice# with probability distribution defined by

$$X_t = X_0 + \int_0^t f(X_s, s)ds + \int_0^t g(X_s, s)dW_s$$

where the Itô stochastic integral is defined as

$$\int_0^t g(X_s, s)dW_s = \lim_{\Delta t \to 0} \sum_{k=0}^{\lfloor t/\Delta t \rfloor} g(X_{k\Delta t}, \varepsilon k)\sqrt{\Delta t} Z_k$$

where $Z_1, Z_2, \ldots \sim \mathcal{N}(0, I)$ are IID.

#right-continuous with left limits (càdlàg)

Solution for SDE

For a given fixed path $\{X_t\}_{t=0}^{T}$, we cannot determine whether it was generated as an instance of the SDE. (Given a fixed sequence 00110011, can you determine whether it was generated as 8 independent Bernoulli random variables?)

Rather, we can talk about whether a distribution of paths solve the SDE. A “solution” of an SDE is a probability distribution of $\{X_t\}_{t=0}^{T}$ (the joint distribution over all X_t for $t \in [0, T]$).

For diffusion probabilistic models, we will consider a weaker notion: The marginal probability distributions $\{p_t\}_{t=0}^{T}$ such that $X_t \sim p_t$ for all $t \in [0, T]$.

Our question of interest is: How does p_t evolve as a function of time t?

Fokker–Planck equation 1D

The time evolution of p_t under the SDE $dX_t = f(X_t, t)dt + g(t)dW_t$ is governed by the Fokker–Planck (FP) equation.

For $d = 1$, the FP equation is

$$\partial_t p_t = -\partial_x(f p_t) + \frac{g^2}{2}\partial_x^2(p_t)$$

More precisely, this means

$$\partial_t p_t(x) = -\partial_x(f(x,t)p_t(x)) + \frac{g^2(t)}{2}\partial_x^2(p_t(x))$$

for all $t > 0$ and $x \in \mathbb{R}$. This is a partial differential equation (PDE).

Integration by parts

Let $\varphi : \mathbb{R} \to \mathbb{R}$ and $f : \mathbb{R} \to \mathbb{R}$. Assume φ and f are sufficiently smooth and decay sufficiently quickly as $|x| \to \infty$. Then

$$\int_{\mathbb{R}} \varphi(x) f'(x) \, dx = -\int_{\mathbb{R}} \varphi'(x) f(x) \, dx$$

Let $\varphi : \mathbb{R}^d \to \mathbb{R}^d$ and $f : \mathbb{R}^d \to \mathbb{R}$. Assume φ and f are sufficiently smooth and decay sufficiently quickly as $\|x\| \to \infty$. Then

$$\int_{\mathbb{R}^d} \varphi(x) \cdot \nabla f(x) \, dx = -\int_{\mathbb{R}^d} (\nabla \cdot \varphi(x)) f(x) \, dx$$

(The usual integration by parts has boundary terms, but they vanish under the decay assumption.)

https://en.wikipedia.org/wiki/Integration_by_parts#Higher_dimensions

Derivation of FP equation

Let $d = 1$. Let $\{p_t\}_{t=0}^{T}$ be a family of pdfs such that $X_t \sim p_t$ for $0 \leq t \leq T$. For any $\varphi \in \mathcal{C}_c^\infty(\mathbb{R})$ (set of smooth compactly supported functions on $\mathbb{R}$), we have

$$\begin{aligned}
\partial_t \mathbb{E}_{X\sim p_t}[\varphi(X)] &\approx \frac{1}{\varepsilon}\left(\mathbb{E}_{X\sim p_{t+\varepsilon}}[\varphi(X)] - \mathbb{E}_{X\sim p_t}[\varphi(X)]\right) \\
&\approx \frac{1}{\varepsilon}\mathbb{E}_{\substack{X\sim p_t \\ Z\sim\mathcal{N}(0,I)}}[\varphi(X+\varepsilon f+\sqrt{\varepsilon} gZ) - \varphi(X)] \\
&\approx \frac{1}{\varepsilon}\mathbb{E}_{\substack{X\sim p_t \\ Z\sim\mathcal{N}(0,I)}}[\varphi(X) + \varepsilon\varphi'(X)f(X,t) + \sqrt{\varepsilon}\varphi'(X)g(t)Z + \frac{1}{2}\varphi''(X)g^2(t)\varepsilon Z^2 + \mathcal{O}(\varepsilon^{3/2}) - \varphi(X)] \\
&\approx \mathbb{E}_{X\sim p_t}[\varphi'(X)f(X,t) + \frac{1}{2}\varphi''(X)g^2(t)]
\end{aligned}$$

Therefore,

$$\partial_t \int \varphi(x)p_t(x)dx = \int \varphi'(x)f(x,t)p_t(x)dx + \frac{1}{2}\int \varphi''(x)g^2(t)p_t(x)dx$$

$$\int \varphi(x)\partial_t p_t(x)dx = \int \varphi(x)(-\partial_x(fp_t))dx + \frac{1}{2}\int \varphi(x)g^2\partial_x^2(p_t)dx$$

$$\partial_t p_t = -\partial_x fp_t + \frac{g^2}{2}\partial_x^2(p_t) \quad \blacksquare$$

Fokker–Planck equation (multi-dim)

The multi-dimensional Fokker–Planck equation is

$$
\begin{aligned}
\partial_t p_t(x) &= -\sum_{i=1}^{d} \frac{\partial}{\partial x_i}(f_i(x,t)p_t(x)) + \frac{1}{2}\sum_{i=1}^{d}\sum_{j=1}^{d} \frac{\partial^2}{\partial x_i \partial x_j} p_t(x) \sum_{k=1}^{d} g_{ik}(t) g_{jk}(t) \\
&= -\sum_{i=1}^{d} \frac{\partial}{\partial x_i}(f_i(x,t)p_t(x)) + \frac{1}{2}\sum_{i=1}^{d}\sum_{j=1}^{d} \frac{\partial^2}{\partial x_i \partial x_j} p_t(x) g_{i,:}(t) g_{j,:}^{\intercal}(t) \\
&= -\nabla_x \cdot (f p_t) + \frac{1}{2}\mathrm{Tr}(g g^{\intercal} \nabla_x^2 p_t) \\
&= -\nabla_x \cdot (f p_t) + \frac{1}{2}\mathrm{Tr}(g^{\intercal} \nabla_x^2 p_t g)
\end{aligned}
$$

Example SDE: Ornstein–Uhlenbeck process

Example:

$$dX_t = -\beta X_t dt + \sigma dW_t$$

$$X_t \mid X_0 \sim \mathcal{N}\left(e^{-\beta t} X_0, \frac{\sigma^2}{2\beta}(1 - e^{-2\beta t})\right)$$

If $X_0 \sim \mathcal{N}(0, \sigma^2/\beta)$

$$X_t \sim \mathcal{N}\left(0, \frac{\sigma^2}{2\beta}\right)$$

$$p_t(X_t) = \frac{1}{\sqrt{\pi\sigma^2/\beta}} \exp\left[-\frac{\beta}{\sigma^2}(X_t)^2\right]$$

With direct calculations, we can verify that p_t satisfies the FP equation.

$$\begin{aligned} 0 = \partial_t p_t(x) &= -\partial_x(f p_t) + \frac{g^2}{2}\partial_x^2(p_t) \\ &= \partial_x(\beta x p_t(x)) + \frac{\sigma^2}{2}\partial_x^2(p_t(x)) \\ &= 0 \end{aligned}$$

Corruption via Ornstein–Uhlenbeck

The Ornstein–Uhlenbeck process

$$dX_t = -\beta X_t dt + \sigma dW_t$$

with $\beta \geq 0$ and $\sigma > 0$ adds noise to the a datapoint X_0. As $T \to \infty$, all information is lost.

Since $X_t \mid X_0 \sim \mathcal{N}\left(e^{-\beta t} X_0, \frac{\sigma^2}{2\beta}(1 - e^{-2\beta t}) I\right)$, we have X_T is approximately distributed as $\mathcal{N}\left(0, \frac{\sigma^2}{2\beta} I\right)$ if $\beta > 0$ and $T \approx \infty$.

Q) Samping $X_T \sim \mathcal{N}\left(0, \frac{\sigma^2}{2\beta} I\right)$ is easy. Can we reverse the SDE to sample X_0?

Forward-time ODE

To simulate
$$\frac{dX}{dt}(t) = f(X(t), t), \qquad X(0) \text{ given}$$

for $0 < t$, set $X_0 = X(0)$ and compute
$$X_{k+1} = X_k + \Delta t f(X_k, k\Delta t), \quad k = 0, 1, \ldots$$

for sufficiently small Δt and set $t = k\Delta t$.

Reverse-time ODE

To simulate

$$\frac{dX}{dt}(t) = f(X(t), t), \qquad X(T) \text{ given}$$

for $0 < t < T$, set $K = \lfloor T/\Delta t \rfloor$ and $X_K = X(T)$ and compute

$$X_{k-1} = X_k - \Delta t f(X_k, k\Delta t), \quad k = K, K-1, \ldots, 2, 1$$

for sufficiently small Δt and set $t = k\Delta t$.

Reversing time for ODEs is easy.
(Mapping from $X(0)$ to $X(T)$ is, after all, a one-to-one map.)

Forward-time SDE

To simulate

$$dX_t = f(X_t)dt + g(t)dW_t, \qquad X_0 \sim p_0$$

for $0 < t$, sample $X_0 \sim p_0$ and compute

$$X_{k+1} = X_k + \Delta t f(X_k, k\Delta t) + g(k\Delta t)\sqrt{\Delta t} Z_k, \qquad k = 0, 1, \ldots$$

for sufficiently small Δt and set $t = k\Delta t$, where $Z_1, Z_2, \ldots \sim \mathcal{N}(0, I)$.

Reverse-time SDE

To simulate

$$dX_t = f(X_t, t)dt + g(t)dW_t, \qquad X_T \sim p_T$$

for $0 < t < T$, set $X_{\lfloor T/\Delta t \rfloor} = X_T$, and compute

$$X_{k-1} = X_k - \Delta t f(X_k, k\Delta t) - g(k\Delta t)\sqrt{\Delta t} Z_k, \qquad k = K, K-1, \ldots, 2, 1$$

This does not work!

Rewinding time in SDEs takes more care

Reverse-time SDE

Example:

See code

Anderson's reverse-time SDE theorem

Instead, given the forward-time SDE

$$dX_t = f(X_t, t)dt + g(t)dW_t, \qquad X_0 \sim p_0$$

the corresponding *reverse-time SDE* is

$$d\overline{X}_t = (f(\overline{X}_t, t) - g^2(t)\nabla_x \log p_t(\overline{X}_t))dt + g(t)d\overline{W}_t, \qquad \overline{X}_T \sim p_T$$

where $\overline{W}_t$ is the reverse-time Brownian motion and p_T is the pdf of $\overline{X}_T$ defined by the forward-time SDE.

Alternatively, define $\{Y_t\}_{t=0}^T$ via

$$dY_t = -(f(Y_t, T-t) - g^2(T-t)\nabla_x \log p_{T-t}(Y_t))dt + g(T-t)dW_t, \qquad Y_0 \sim p_T$$

(Note that $dW_t \overset{\mathcal{D}}{=} -dW_t$.) If we set $\overline{X}_{T-t} = Y_t$, then $X_t \overset{\mathcal{D}}{=} \overline{X}_t = Y_{T-t}$.

B. D. O. Anderson, Reverse-time diffusion equation models, *Stochastic Processes and their Applications*, 1982.

Reverse-time SDE

To simulate the reverse-time SDE,

$$d\overline{X}_t = (f(\overline{X}_t, t) - g^2(t)\nabla_x \log p_t(\overline{X}_t))dt + g(t)d\overline{W}_t, \qquad \overline{X}_T \sim p_T$$

for $0 < t < T$, sample $\bar{X}_T \sim p_T$, set $K = \lfloor T/\Delta t \rfloor$ and $\bar{X}_K = \bar{X}_T$, and compute

$$\overline{X}_{k-1} = \overline{X}_k - \Delta t\left(f(\overline{X}_k, k\Delta t) - g^2(k\Delta t)\nabla_x \log p_{k\Delta t}(\overline{X}_k)\right) + g(k\Delta t)\sqrt{\Delta t} Z_k, \qquad k = K, K-1, \ldots, 2, 1$$

where $Z_1, \ldots, Z_K \sim \mathcal{N}(0, I)$. More concisely,

$$\overline{X}_{k-1} = \overline{X}_k - \Delta t(f - g^2\nabla_x \log p_{k\Delta t}(\overline{X}_k)) + g\sqrt{\Delta t} Z_k, \qquad k = K, K-1, \ldots, 2, 1$$

B. D. O. Anderson, Reverse-time diffusion equation models, *Stochastic Processes and their Applications*, 1982.

Example: Reverse

See code

Marginal vs. joint distributions

Note that Anderson's theorem is claiming [$X_t \overset{\mathcal{D}}{=} \overline{X}_t$ for all $0 \leq t \leq T$], which is a weaker statement than $\{X_t\}_{t=0}^T \overset{\mathcal{D}}{=} \{\overline{X}_t\}_{t=0}^T$.

The latter $\{X_t\}_{t=0}^T \overset{\mathcal{D}}{=} \{\overline{X}_t\}_{t=0}^T$ asserts that the two processes have equal (joint) distributions, while the former [$X_t \overset{\mathcal{D}}{=} \overline{X}_t$ for all $0 \leq t \leq T$] asserts that the marginal distributions are equal for all t.

Diffusion probabilistic models are concerned with the marginal distributions.

Anderson's theorem proof

Let $d = 1$. Let $\{p_t\}_{t=0}^T$ be marginal densities of the forward SDE

$$dX_t = f(X_t)dt + g(t)dW_t, \qquad X_0 \sim p_0$$

Remember that $\{p_t\}_{t=0}^T$ satisfies the FP equation

$$\begin{aligned}\partial_t p_t &= -\partial_x(f(x,t)p_t(x)) + \partial_x(g^2(t)\partial_x p_t(x)) - \frac{g^2(t)}{2}\partial_x^2(p_t(x)) \\ &= -\partial_x(f(x,t)p_t(x)) + \frac{g^2(t)}{2}\partial_x^2(p_t(x))\end{aligned}$$

Anderson's theorem proof

Let $\{q_t\}_{t=0}^T$ be marginal densities of

$$dY_t = -(f(Y_t, T-t) - g^2(T-t)\partial_{Y_t} \log p_{T-t}(Y_t))dt + g(T-t)dW_t, \qquad Y_0 \sim p_T$$

Then $\{q_t\}_{t=0}^T$ satisfies the FP equation

$$\partial_t q_t(y) = \partial_y(f(y, T-t) - g^2(T-t)\partial_y \log p_{T-t}(y))q_t(y) + \frac{g^2(T-t)}{2}\partial_y^2(q_t(y))$$

Let $\{\overline{p}_t\}_{t=0}^T$ be marginal densities of the reverse-time SDE

$$d\overline{X}_t = (f(\overline{X}_t, t) - g^2(t)\partial_x \log p_t(\overline{X}_t))dt + g(t)d\overline{W}_t, \qquad \overline{X}_T \sim p_T$$

Since $\overline{p}_t = q_{T-t}$, the densities $\{\overline{p}_t\}_{t=0}^T$ satisfies

$$\partial_t \overline{p}_t = -\partial_x \left((f(x,t) - g^2(t)\partial_x \log p_t(x))\overline{p}_t(x)\right) - \frac{g^2(t)}{2}\partial_x^2(\overline{p}_t(x))$$

Anderson's theorem proof

If we plug in $\{\bar{p}_t\}_{t=0}^T = \{p_t\}_{t=0}^T$ into

$$\partial_t \overline{p}_t = -\partial_x \big((f(x,t) - g^2(t)\partial_x \log p_t(x))\overline{p}_t(x) \big) - \frac{g^2(t)}{2}\partial_x^2(\overline{p}_t(x))$$

we get the FP equation for $\{p_t\}_{t=0}^T$

$$\partial_t p_t = -\partial_x \big((f(x,t)p_t(x)) + g^2(t)\partial_x p_t(x) \big) - \frac{g^2(t)}{2}\partial_x^2(p_t(x))$$

In other words, we have verified that $\{\bar{p}_t\}_{t=0}^T = \{p_t\}_{t=0}^T$ solves the FP equation for $\{\bar{p}_t\}_{t=0}^T$, which proves $\{\bar{p}_t\}_{t=0}^T = \{p_t\}_{t=0}^T$ provided that the solution to the PDE is unique. We omit the uniqueness argument. ∎

Sample generation via SDE

Let $X_0 \sim p_0$, where p_0 corresponds to the MNIST or ImageNet dataset.

$$dX_t = f dt + g dW_t, \qquad X_0 \sim p_0$$

Then the forward-time SDE produces $X_T \sim p_T$.

If we can sample $\bar{X}_T \sim p_T$ and run the reverse-time SDE

$$d\overline{X}_t = (f - g^2 \nabla \log p_t(\overline{X}_t))dt + g d\overline{W}_t, \qquad \overline{X}_T \sim p_T$$

tis would be a generative model producing images $X_0 \sim p_0$.

Sample generation via SDE

Consider the Ornstein–Uhlenbeck forward-time SDE

$$dX_t = -\beta X_t dt + \sigma dW_t, \qquad X_0 \sim p_0$$

Remember that

$$X_t \mid X_0 \sim \mathcal{N}\left(e^{-\beta t} X_0, \sigma_t^2 I\right), \qquad \sigma_t^2 = \frac{\sigma^2}{2\beta}(1 - e^{-2\beta t})$$

If T is sufficiently large, $p_T \approx \mathcal{N}(0, \sigma_T^2 I)$.

Consider the reverse-time counterpart

$$d\overline{X}_t = (-\beta \overline{X}_t - \sigma^2 \nabla \log p_t(\overline{X}_t))dt + \sigma d\overline{W}_t, \qquad \overline{X}_T \sim \mathcal{N}(0, \sigma_T^2 I)$$

(It would be better to sample $\overline{X}_T \sim p_T$ exactly, but we do not know p_T because we do not know $p_0 = p_{\text{data}}$.)

Sample generation via SDE

Set $K = \lfloor T/\Delta t \rfloor$ and sample $\bar{X}_K \sim \mathcal{N}(0, \sigma_T^2 I)$. Using a standard discretization (Euler–Maruyama), we get

$$
\begin{aligned}
&\overline{X}_K \sim \mathcal{N}(0, \sigma_T^2 I)\\
&\text{for } k = K, K-1, \ldots, 2, 1\\
&\qquad Z_k \sim \mathcal{N}(0, I)\\
&\qquad \overline{X}_{k-1} = \overline{X}_k - \Delta t(-\beta \overline{X}_k - \sigma^2 \nabla \log p_{k\Delta t}(\overline{X}_k)) + \sigma\sqrt{\Delta t} Z_k\\
&\text{end}
\end{aligned}
$$

Output is $\bar{X}_0$ approximately distributed as p_0.

Interestingly, there is randomness in the generation process.

To clarify, this is not yet implementable since we do not have access to $\nabla \log p_t$.

Reverse-time ODE

Let $\{p_t\}_{t=0}^T$ be the marginal density functions of the forward-time SDE

$$dX_t = f dt + g dW_t, \qquad X_0 \sim p_0$$

and reverse-time SDE

$$d\overline{X}_t = (f(\overline{X}_t, t) - g^2(t)\nabla \log p_t(\overline{X}_t))dt + g(t)d\overline{W}_t, \qquad \overline{X}_T \sim p_T$$

Then, $\{p_t\}_{t=0}^T$ is also the marginal density functions of the following reverse-time ODE

$$d\overline{X}_t = \left(f(\overline{X}_t, t) - \frac{g^2(t)}{2}\nabla \log p_t(\overline{X}_t)\right) dt, \qquad \overline{X}_T \sim p_T$$

This ODE defines a flow model, a one-to-one mapping between $\overline{X}_T$ and $\overline{X}_0$.

Proof) Same reasoning as Anderson's theorem with the Fokker–Planck equation. ■

Sample generation via ODE

Consider the particular forward-time SDE

$$dX_t = -\beta X_t dt + \sigma dW_t, \qquad X_0 \sim p_0$$

If T is sufficiently large, $p_T \approx \mathcal{N}(0, \sigma_T^2 I)$. Consider the reverse-time ODE

$$d\overline{X}_t = \left(\frac{\sigma^2}{2}\nabla \log p_t(\overline{X}_t) - \beta \overline{X}_t\right) dt, \qquad \overline{X}_T \sim \mathcal{N}(0, \sigma_T^2 I)$$

Y. Song, J. Sohl-Dickstein, D. P. Kingma, A. Kumar, S. Ermon, and B. Poole, Score-based generative modeling through stochastic differential equations, *ICLR*, 2021.

Sample generation via ODE

Set $K = \lfloor T/\Delta t \rfloor$ and sample $\overline{X}_K \sim \mathcal{N}(0, \sigma_T^2 I)$. Using a standard discretization (Euler), we get

$$\overline{X}_K \sim \mathcal{N}(0, \sigma_T^2 I)$$

for $k = K, K-1, \ldots, 2, 1$

$$\overline{X}_{k-1} = \overline{X}_k - \Delta t \left(-\beta \overline{X}_k - \frac{\sigma^2}{2} \nabla \log p_{k\Delta t}(\overline{X}_k) \right)$$

end

Output is $\bar{X}_0$ approximately distributed as p_0.

Only source of randomness is in the initial generation of $\overline{X}_K$.

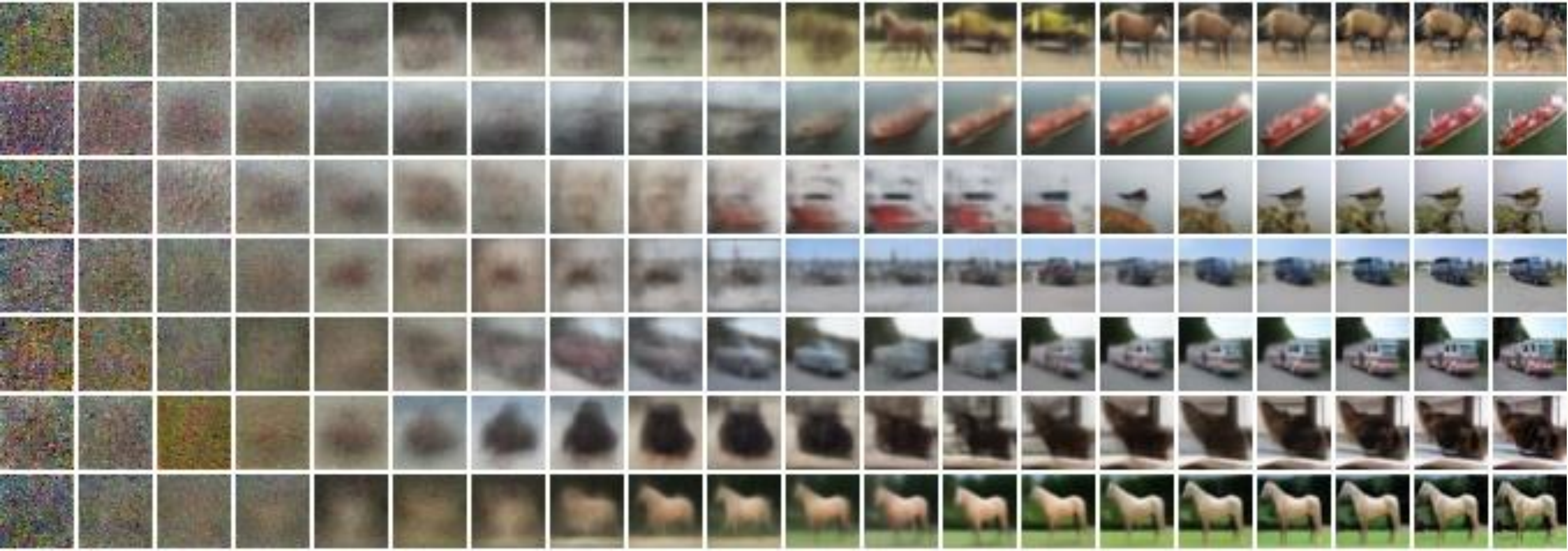
To clarify, this is not yet implementable since we do not have access to $\nabla \log p_t$.

Y. Song, J. Sohl-Dickstein, D. P. Kingma, A. Kumar, S. Ermon, and B. Poole, Score-based generative modeling through stochastic differential equations, *ICLR*, 2021.

Sample generation via (discretized) SDE

J. Ho, A. Jain, and P. Abbeel, Denoising diffusion probabilistic models, *NeurIPS*, 2020.

Sample generation via (discretized) SDE



J. Ho, A. Jain, and P. Abbeel, Denoising diffusion probabilistic models, *NeurIPS*, 2020.