

# Deep Generative Models: Diffusion Models – SDE perspective

Fall Semester 2025

René Vidal

Director of the Center for Innovation in Data Engineering and Science (IDEAS)

Rachleff University Professor, University of Pennsylvania

Amazon Scholar & Chief Scientist at NORCE



# Taxonomy of Generative Models

What we've learned:

- MMs, HMMs, LDSs, RNNs, SSMs

## Deep Generative Models

**Autoregressive models**  
(e.g., PixelCNN)

**Flow-based models**  
(e.g., RealNVP)

**Latent variable models**

**Energy-based models**

**Implicit models**  
(e.g., GANs)

**Prescribed models**  
(e.g., VAEs)

What we've learned:

- PPCA
- VAE
- Latent Diffusion

What we study now:

- SDE perspective of Diffusion process

# Stochastic Differential Equation

- SDE is a differential equation in which one or more terms is a stochastic process.

$$dx_t = \mu(x_t, t)dt + \sigma(x_t, t)dw_t$$

- $\mu(x_t, t)$  is called as the drift coefficient,
- $\sigma(x_t, t)$  is called the diffusion coefficient,
- $w_t$  is called the Weiner process and has the following properties:
  1. Initial zero condition:  $w_0 = 0$  a.s. (acronym for “almost surely” – any event that satisfies with probability 1)
  2. Independent increments: For every  $t, u \geq 0$ , the increments  $w_{t+u} - w_t$  are independent of  $w_{<t}$ .
  3. Gaussian increments: For every  $t, u \geq 0$ , the increments  $w_{t+u} - w_t \sim \mathcal{N}(0, uI)$ .
  4. Continuity paths:  $w_t$  is continuous in  $t$  a.s.
- How to obtain  $x_t$  given  $x_0$ ?

$$x_t - x_0 = \int_0^t \mu(x_t, t)dt + \int_0^t \sigma(x_t, t)dw_t$$

Lebesgue integral

$\mu$ , and  $\sigma$  must be Lipschitz continuous in  $(x, t)$  for the uniqueness and existence of solution.

Itô integral

# Fokker-Plank Equation

$$dx_t = \mu(x_t, t)dt + \sigma(x_t, t)dw_t$$

- Every SDE has an associated ODE in probability space; which is characterized by Fokker-Plank equation.
- Lemma: The probability ODE is given by:
$$\frac{dp(x_t, t)}{dt} = -\nabla_x \cdot \left( \mu(x_t, t) \cdot p(x_t, t) - \frac{1}{2} \nabla_x (\sigma(x_t, t)^2 \cdot p(x_t, t)) \right)$$
  - $\nabla_x \cdot g(x_t, t) = \sum_i \partial_{x_i} (g(x_t, t))_i$  is the divergence operation.
- Conservation law: The net change of probability at point  $(x_t, t)$  is the same as the flow due to drift and diffusion coefficient.
- In the forward process the time progresses from  $t: 0 \rightarrow 1$ , and initial distribution  $p_0$ . The probability ODE can be obtained.
- To obtain reverse process we need to modify the SDE with initial distribution  $p_1$  such that probability ODE matches with forward.

# Itô's lemma: Scalar case

$$dx_t = \mu(x_t, t)dt + \sigma(x_t, t)dw_t$$

- **Itô** is pronounced as ***ee-toh***.
- Lemma: Suppose  $y_t = u(x_t, t)$  is another process defined on  $x_t$ , then we have

$$dy_t = \partial_t u(x_t, t)dt + \partial_x u(x_t, t)dx_t + \frac{1}{2} \partial_x^2 u(x_t, t)\sigma(x_t, t)^2 dt.$$

- Intuition: Let us compute the finite difference  $u(x_{t+\Delta t}, t + \Delta t) - u(x_t, t)$  using Taylor's series:

$$\begin{aligned} &u(x_{t+\Delta t}, t + \Delta t) - u(x_t, t) \\ &= \partial_t u(x_t, t)\Delta t + \partial_x u(x_t, t)(x_{t+\Delta t} - x_t) + \frac{1}{2} \partial_x^2 u(x_t, t)(x_{t+\Delta t} - x_t)^2 + o(\Delta t^{1.5}). \end{aligned}$$

- Now as  $\Delta t \rightarrow 0$ , we have  $(x_{t+\Delta t} - x_t)^2 = \sigma(x_t, t)^2 (w_{t+\Delta t} - w_t)^2 + o(\Delta t^{1.5})$ .
- By Gaussian increment property we have  $w_{t+\Delta t} - w_t \sim \mathcal{N}(0, \Delta t)$ . Therefore,  $w_{t+\Delta t} - w_t$  scales proportionally with  $\Delta t$ , due to which the 3<sup>rd</sup> term survives.
- The actual proof requires more rigorous arguments refer to [Stochastic Calculus For Finance, Shreve](#).

# Proof for Fokker-Plank Equations

$$dx_t = \mu(x_t, t)dt + \sigma(x_t, t)dw_t$$

- From Ito's Lemma we have that for any  $u$  we have:

$$du(x_t, t) = \partial_t u(x_t, t)dt + \partial_x u(x_t, t)dx_t + \frac{1}{2} \partial_x^2 u(x_t, t)\sigma(x_t, t)^2 dt.$$

- Using the forward SDE we have

$$du(x_t, t) = \left[ \partial_t u(x_t, t) + \mu(x_t, t)\partial_x u(x_t, t) + \frac{1}{2} \sigma(x_t, t)^2 \partial_x^2 u(x_t, t) \right] dt + \partial_x u(x_t, t) \sigma(x_t, t)dw_t.$$

- Now apply expectation on both the sides:

$$d\mathbb{E}_p[u(x_t, t)] = \mathbb{E}_p \left[ \partial_t u(x_t, t) + \mu(x_t, t)\partial_x u(x_t, t) + \frac{1}{2} \sigma(x_t, t)^2 \partial_x^2 u(x_t, t) \right] dt.$$

- By exchanging the derivative with the integral, we have:

$$\mathbb{E}_p[\partial_t u(x_t, t)] + \mathbb{E}_{\partial_t p}[u(x_t, t)] = \mathbb{E}_p[\partial_t u(x_t, t)] + \mathbb{E}_p \left[ \partial_x u(x_t, t)\mu(x_t, t) + \frac{1}{2} \partial_x^2 u(x_t, t)\sigma(x_t, t)^2 \right].$$

- Integration by parts gives us:  $\mathbb{E}_{\partial_t p}[u(x_t, t)] = \mathbb{E}_{\partial_x(\mu \cdot p)}[-u(x_t, t)] + \frac{1}{2} \mathbb{E}_{\partial_x(\sigma^2 \cdot p)}[u(x_t, t)].$
- Since these arguments hold true for any  $u$ ; then the increments must match.

# Reversal of the diffusion process

$$dx_t = \mu(x_t, t)dt + \sigma(x_t, t)dw_t$$

- [Brian Aderson](#) showed this in 1982 which later become a building block in [score-matching diffusion models](#) introduced in 2020.
- Theorem: The time-reversal of the SDE  $dx_t = \mu(x_t, t)dt + \sigma(x_t, t)dw_t$  with  $x_0 \sim p_0$  is given by  $dx'_t = \left( \frac{\nabla_x \left( \sigma(x'_{1-t}, 1-t)^2 \cdot p(x'_{1-t}, 1-t) \right)}{p(x'_t, t)} - \mu(x'_{1-t}, 1-t) \right) dt + \sigma(x'_{1-t}, 1-t)d\bar{w}_t$  with  $x'_0 \sim p_1$ , and  $(w_t, \bar{w}_t)$  are Wiener process.
- Proof: Let us compute the Fokker-Plank equations for reverse SDE.

$$\begin{aligned} \frac{dp(x'_{1-t}, 1-t)}{dt} &= -\nabla_x \cdot \left( \left( -\mu(x'_{1-t}, 1-t) + \frac{\nabla_x \left( \sigma(x'_{1-t}, 1-t)^2 \cdot p(x'_{1-t}, 1-t) \right)}{p(x'_{1-t}, 1-t)} \right) \cdot p(x'_{1-t}, 1-t) - \frac{1}{2} \nabla_x^\top \left( \sigma(x'_{1-t}, 1-t)^2 \cdot p(x'_{1-t}, 1-t) \right) \right), \\ &= -(-\nabla_x \left( \mu(x'_{1-t}, 1-t) \cdot p(x'_{1-t}, 1-t) \right) + \nabla_x \cdot (\nabla_x \sigma(x'_{1-t}, 1-t)^2 \cdot p(x'_{1-t}, 1-t))) = \frac{dp(x_{1-t}, 1-t)}{d(1-t)}. \end{aligned}$$

- The probability ODEs match in both the direction match.
- Due to uniqueness of the solution at each  $t$  the solutions to ODE /  $p(x_t, t)$  match in both direction.

# Special case: SDE $\longrightarrow$ DDPM

- Set  $\mu(x, t) = -\frac{1}{2}\sqrt{1 - N\alpha_t} \cdot x_t$ , and  $\sigma(x, t) = N\sqrt{1 - \alpha_t}$ .
- We have the forward process:  $dx_t = -\frac{1}{2}\sqrt{1 - N\alpha_t} \cdot x_t dt + N\sqrt{1 - \alpha_t} \cdot dw_t$ .
- Now we discretize the time with  $N$  scales:  $t = \frac{i}{N}$ , and  $\Delta t = \frac{1}{N}$ .
- We can write  $dx_t \approx x_{i+1} - x_i$ ,  $-\frac{1}{2}\sqrt{1 - N\alpha_t} \cdot x_t dt = -\frac{1}{2}\sqrt{1 - N\alpha_i} \cdot x_i \Delta t$ , and  $N\sqrt{1 - \alpha_t}dw_t \sim \mathcal{N}(\epsilon_i; 0, \sqrt{1 - \alpha_t}I)$ .
- We obtain  $x_{i+1} = \left(1 - \frac{1}{2N}\sqrt{1 - N\alpha_i}\right)x_i + \sqrt{1 - \alpha_t} \cdot \epsilon_i$ .
- Let  $N \rightarrow \infty$ , then we have  $\left(1 - \frac{1}{2N}\sqrt{1 - N\alpha_i}\right) \approx \sqrt{1 - \frac{1}{N} + \alpha_i} \approx \sqrt{\alpha_i}$ .
- Finally, we have  $x_{i+1} \approx \sqrt{\alpha_i}x_i + \sqrt{1 - \alpha_t}\epsilon_i$ . This is the same as DDPM.



# Special case: DDIM $\longrightarrow$ SDE

- Recall the forward process in diffusion models:

$$x_t = \sqrt{\alpha_t} x_{t-1} + \sqrt{1 - \alpha_t} \epsilon_t, \quad \epsilon_t \sim \mathcal{N}(\epsilon_t; 0, I)$$

This is in the discrete time.

- $D\alpha(t) = N \cdot (1 - \alpha_t), \Delta t = \frac{1}{N}$ , where  $N$  is the number of noise scales.
- In the limit  $N \rightarrow \infty$  we have,
  - $x_{t+\Delta t} = \sqrt{1 - \alpha(t)\Delta t} \cdot x_t + \sqrt{\Delta t \cdot \alpha(t)} \cdot \epsilon_t,$
  - $x_{t+\Delta t} \approx \left(1 - \frac{1}{2}\right) x_t + \sqrt{\alpha(t)} \cdot \sqrt{\Delta t} \cdot \epsilon_t,$
  - $x_{t+\Delta t} - x_t \approx -\frac{\alpha(t)}{2} x_t \cdot \Delta t + \sqrt{\alpha(t)} \cdot \sqrt{\Delta t} \cdot \epsilon_t,$
  - $dx_t \approx -\frac{\alpha(t)}{2} x_t dt + \sqrt{\alpha(t)} \sqrt{dt} \cdot \epsilon_t.$
- $\sqrt{dt} \cdot \epsilon_t$  is infinitesimal Normal random variable or Brownian / Weiner process.

# Special case: SDE $\longrightarrow$ Score-based models

- Suppose we have zero drift and sample independent noise variance:

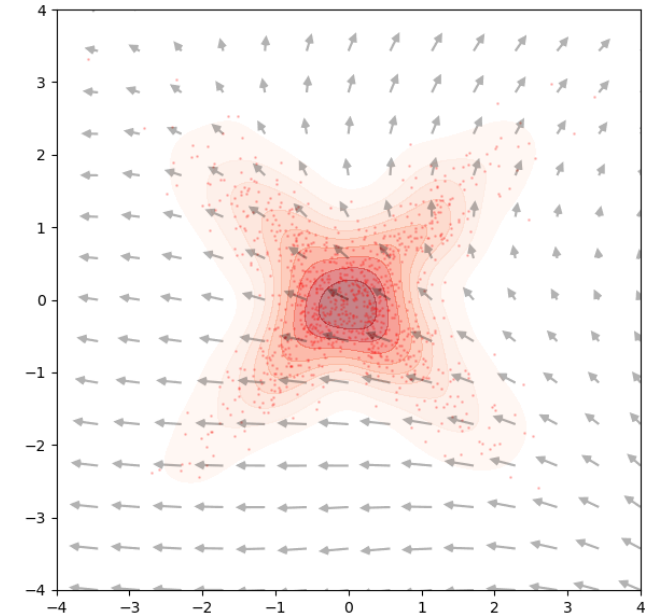
Forward:  $dx_t = \sigma(t)dw_t; x_0 \sim p_0,$

Reverse:  $dx'_t = \sigma(1-t)^2 \underbrace{\nabla_x \ln(p(x'_{1-t}, 1-t))}_{\text{Stein's score function}} dt + \sigma(1-t)d\bar{w}_t; x'_0 \sim p_1.$

Stein's score function

- In the reverse process, we are forcing the samples to drift to high-likelihood regions just like in score-based diffusion model.
- After discretizing the reverse process, we obtain:

$$x'_{t+\Delta t} = x'_t + \sigma(1-t)^2 \Delta t \nabla_x \ln(p(x'_{1-t}, 1-t)) + \sigma(1-t) \sqrt{\Delta t} \cdot \epsilon_t$$



Vector field of score function