Score-Based Diffusion Model

Spring 2025 DRP Presentation

Jiayi Zhang, Yingzhou Wang, Yuhan Wang, and Zikun Zhou
Mentor: Borong Zhang

University of Wisconsin–Madison

Department of Mathematics
COLLEGE OF LETTERS & SCIENCE
UNIVERSITY OF WISCONSIN–MADISON

Table of contents

Probability and SDE Formulation [1]

Definition (Probability Space)
A triple $(\Omega, \mathcal{A}, P)$ is called a *probability space* if:

- Ω is a set,
- $\mathcal{A}$ is a σ-algebra on Ω,
- P is a probability measure on $\mathcal{A}$.

Example: Coin Toss
Consider a fair coin toss:

$$\Omega = \{H, T\}, \quad \mathcal{A} = \big\{\emptyset, \{H\}, \{T\}, \{H, T\}\big\}, \quad P(\{H\}) = P(\{T\}) = \tfrac{1}{2}.$$

Here $(\Omega, \mathcal{A}, P)$ is a probability space modeling one toss of the coin.

Definition (Probability Space)
A triple $(\Omega, \mathcal{A}, P)$ is called a *probability space* if:

- Ω is a set,
- $\mathcal{A}$ is a σ-algebra on Ω,
- P is a probability measure on $\mathcal{A}$.

Example: Uniform on $\Omega = [0, 1]$
Suppose we choose a number uniformly at random from $[0, 1]$.

Questions:

- What is the probability space $(\Omega, \mathcal{A}, P)$?
- What is the probability $P(\mathbb{Q} \bigcap [0, 1])$?

Definition (Sigma-Algebra)
A σ-*algebra* on a set Ω is a collection $\mathcal{A}$ of subsets of Ω satisfying:

1. $\Omega \in \mathcal{A}$,
2. If $A \in \mathcal{A}$, then $A^c = \Omega \setminus A \in \mathcal{A}$,
3. If $A_1, A_2, \ldots \in \mathcal{A}$, then $\bigcup_{k=1}^{\infty} A_k \in \mathcal{A}$.

Definition (Probability Measure)
Let $\mathcal{A}$ be a σ-algebra of subsets of Ω. A function

$$P : \mathcal{A} \to [0, 1]$$

is called a *probability measure* if:

1. $P(\Omega) = 1$,
2. If $A_1, A_2, \ldots$ are pairwise disjoint sets in $\mathcal{A}$, then

$$P\left(\bigcup_{k=1}^{\infty} A_k\right) = \sum_{k=1}^{\infty} P(A_k).$$

Example: Uniform on $\Omega = [0,1]$

Suppose we choose a number uniformly at random from $[0,1]$.

Answer:

- The probability space is

$$(\Omega, \mathcal{A}, P) \; = \; \big([0,1], \mathcal{B}([0,1]), \text{ Lebesgue measure}\big),$$

  where $\mathcal{B}([0,1])$ is the Borel σ-algebra on $[0,1]$.
- To compute $P(\mathbb{Q} \cap [0,1])$: enumerate the rationals in $[0,1]$ as $\mathbb{Q} \cap [0,1] = \{q_1, q_2, \dots\}$.
  - By Axiom 3 of the σ-algebra, each singleton $\{q_k\} \in \mathcal{B}([0,1])$ and hence $\bigcup_{k=1}^{\infty} \{q_k\} = \mathbb{Q} \cap [0,1] \in \mathcal{A}$.
  - Each singleton has Lebesgue measure zero, so $P(\{q_k\}) = 0$.
  - By Axiom 2 (countable additivity) of the probability measure,

$$P(\mathbb{Q} \cap [0,1]) \; = \; P\Big(\bigcup_{k=1}^{\infty} \{q_k\}\Big) \; = \; \sum_{k=1}^{\infty} P(\{q_k\}) \; = \; \sum_{k=1}^{\infty} 0 \; = \; 0.$$

Definition (Random Variable)
Let $(\Omega, \mathcal{A}, P)$ be a probability space. A mapping

$$X : \Omega \to \mathbb{R}^n$$

is called an *n-dimensional random variable* if for each $B \in \mathcal{B}$, we have $X^{-1}(B) \in \mathcal{A}$. Equivalently, we say that X is $\mathcal{A}$-measurable.

Example: Coin Toss Random Variable

$$\Omega = \{H, T\}, \quad \mathcal{A} = \big\{\emptyset, \{H\}, \{T\}, \{H, T\}\big\}, \quad P(\{H\}) = P(\{T\}) = \tfrac{1}{2}.$$

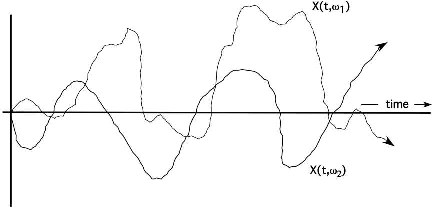
Define

$$X : \Omega \to \mathbb{R}, \quad X(H) = 1,\ X(T) = 0.$$

For any Borel set $B \subseteq \mathbb{R}$, the inverse image

$$X^{-1}(B) = \{\omega \in \Omega : X(\omega) \in B\}$$

is one of $\emptyset$, $\{H\}$, $\{T\}$, or $\{H, T\}$, all of which lie in $\mathcal{A}$.



Definition (Brownian Motion / Wiener Process)
A real-valued stochastic process $\{W(t,\cdot)\}_{t\geq 0}$ on a probability space $(\Omega, \mathcal{F}, P)$ is called a *Brownian motion* (or *Wiener process*) if:

1. $W(0,\cdot) = 0$ almost surely,
2. For $t \geq 0$, the random variable $W(t,\cdot)$ has distribution $\mathcal{N}(0,t)$,
3. For any times $0 \leq t_1 < t_2 < \cdots < t_n$, the increments are mutually independent:

$$W(t_1,\cdot), \quad W(t_2,\cdot) - W(t_1,\cdot), \quad \ldots, \quad W(t_n,\cdot) - W(t_{n-1},\cdot)$$

General SDE Formulation

Definition (SDE)

- Canonical form

$$dX_t \;=\; b(X_t)\,dt \;+\; \sigma(X_t)\,dW_t, \quad X_0 = x_0$$

- Drift term $b(x)$: deterministic trend (velocity field)
- Diffusion term $\sigma(x)$: amplitude of random fluctuations
- Key examples:
    - Brownian motion: $dX_t = dW_t$
    - Ornstein–Uhlenbeck: $dX_t = -X_t\,dt + \sqrt{2}\,dW_t$

Ito’s Formula and Fokker–Planck [2]

Motivation

$$\begin{cases} dX_t = b\left(X_t\right) dt + \sigma\left(X_t\right) dW_t, \\ X_0 = x_0. \end{cases}$$

A process X_t is said to *solve* the SDE if for all $t \geq 0$

$$X_t = x_0 + \int_0^t b\left(X_s\right) ds + \int_0^t \sigma\left(X_s\right) dW_s.$$

Theorem (Properties of the Itô Integral)
Let $a, b \in \mathbb{R}$ and let $\{G_s\}_{s\in[0,T]}$, $\{H_s\}_{s\in[0,T]}$ be adapted processes in $L^2\big(\Omega \times [0,T]\big)$. Then:

(i) $$\int_0^T \big(a\,G_s + b\,H_s\big)\,dW_s = a\int_0^T G_s\,dW_s + b\int_0^T H_s\,dW_s.$$

(ii) $$\mathbb{E}\left[\int_0^T G_s\,dW_s\right] = 0.$$

(iii) $$\mathbb{E}\left[\left(\int_0^T G_s\,dW_s\right)^2\right] = \mathbb{E}\left[\int_0^T G_s^2\,ds\right].$$

(iv) $$\mathbb{E}\left[\int_0^T G_s\,dW_s\ \int_0^T H_s\,dW_s\right] = \mathbb{E}\left[\int_0^T G_s\,H_s\,ds\right].$$

Itô's Formula

$$dX_t = b(X_t, t)\,dt + \sigma(X_t, t)\,dW_t.$$

For $f \in C^2(\mathbb{R})$, Itô's lemma starts with

$$df(X_t) = f'(X_t)\,dX_t + \tfrac{1}{2}\,f''(X_t)\,(dX_t)^2.$$

Now substitute dX_t and expand:

$$(dX_t)^2 = b^2\,dt^2 + 2\,b\,\sigma\,dt\,dW_t + \sigma^2\,(dW_t)^2.$$

By the rules of Itô calculus:

$$dt^2 \approx 0, \quad dt\,dW_t \approx 0, \quad (dW_t)^2 = dt,$$

Plugging back in and collecting terms gives the final Itô formula:

$$df(X_t) = \left[f'(X_t)\,b(X_t, t) + \tfrac{1}{2}\,f''(X_t)\,\sigma(X_t, t)^2\right]dt + f'(X_t)\,\sigma(X_t, t)\,dW_t.$$

Fokker–Planck: Expectation Identity

$$df(X_t) = \Big[f'(X_t)\, b(X_t,t) + \tfrac{1}{2} f''(X_t)\, \sigma^2(X_t,t)\Big]\, dt + f'(X_t)\, \sigma(X_t,t)\, dW_t.$$

Taking expectations and noting $\mathbb{E}[\int_0^T f'(X_s)\,\sigma(X_s,s)\,dW_s] = 0$, we obtain

$$\frac{d}{dt}\,\mathbb{E}[f(X_t)] = \mathbb{E}\Big[b(X_t,t)\, f'(X_t) + \tfrac{1}{2}\,\sigma^2(X_t,t)\, f''(X_t)\Big].$$

Fokker–Planck: From Expectation to PDE

Write

$$\mathbb{E}[f(X_t)] \;=\; \int_{\mathbb{R}} f(x)\, p(x,t)\, dx.$$

Then

$$\int_{\mathbb{R}} f(x)\, \partial_t p(x,t)\, dx = \int_{\mathbb{R}} \Big[b(x,t)\, f'(x) + \tfrac{1}{2}\, \sigma^2(x,t)\, f''(x) \Big]\, p(x,t)\, dx.$$

Integrate by parts (assuming decay at infinity):

$$\int_{\mathbb{R}} b\, f'\, p = -\int_{\mathbb{R}} f\, \partial_x \big[b\, p \big], \qquad \int_{\mathbb{R}} \tfrac{1}{2}\sigma^2\, f''\, p = \int_{\mathbb{R}} f\, \tfrac{1}{2}\, \partial_x^2 \big[\sigma^2\, p \big].$$

Hence,

$$\int_{\mathbb{R}} f(x)\, \partial_t p(x,t)\, dx = \int_{\mathbb{R}} f(x)\, \Big[-\partial_x \big[b(x,t)\, p(x,t) \big] + \tfrac{1}{2}\, \partial_x^2 \big[\sigma^2(x,t)\, p(x,t) \big] \Big]\, dx.$$

Since f was arbitrary,

$$\boxed{\partial_t p(x,t) = -\partial_x \big[b(x,t)\, p(x,t) \big] + \tfrac{1}{2}\, \partial_x^2 \big[\sigma^2(x,t)\, p(x,t) \big].}$$

Fokker–Planck: Scalar Examples

Pure drift: $dX_t = 1 \cdot dt$

$$\partial_t p(x,t) \;=\; -\partial_x\big(1 \cdot p(x,t)\big) \quad\Longrightarrow\quad p(x,t) \;=\; \delta\big(x-t\big).$$

Pure diffusion: $dX_t = 1 \cdot dW_t$

$$\partial_t p(x,t) = \tfrac{1}{2}\,\partial_x^2\big(1^2\,p(x,t)\big) \quad\Longrightarrow\quad p(x,t) = \frac{1}{\sqrt{2\pi\,t}}\,\exp\Big(-\tfrac{x^2}{2t}\Big).$$

- Drift case: mass travels at constant speed.
- Diffusion case: Gaussian distribution with mean 0 and variance t.

Diffusion Models [3]

Ornstein–Uhlenbeck (OU) Process

- **Stochastic differential equation:**

$$dX_t = -X_t\,dt + \sqrt{2}\,dW_t, \qquad X_0 \sim p_{\mathrm{data}}(x).$$

- Drift: $-x$ pulls the state toward the origin (mean-reverting).
- Diffusion: $\sqrt{2}\,dW_t$ injects isotropic Gaussian noise.

Fokker–Planck Equation for the OU Process

General 1-D formula

$$\partial_t p(x,t) \;=\; -\partial_x\big[a(x,t)\,p(x,t)\big] \;+\; \frac{1}{2}\,\partial_x^2\big[b(x,t)^2\,p(x,t)\big].$$

OU coefficients (time-homogeneous):

$$a(x) = -x, \qquad b(x) = \sqrt{2}.$$

Substitute into the general formula:

$$\partial_t p(x,t) \;=\; -\partial_x\big[(-x)\,p(x,t)\big] \;+\; \frac{1}{2}\,\partial_x^2\big[(\sqrt{2})^2\,p(x,t)\big] \;=\; \partial_x\big[x\,p(x,t)\big] \;+\; \partial_x^2 p(x,t).$$

Initial condition:

$$p(x,0) = p_{\mathrm{data}}(x).$$

Equilibrium: The unique stationary solution satisfying $\partial_t p = 0$ is the standard Gaussian $p_\infty(x) = \dfrac{1}{\sqrt{2\pi}} e^{-x^2/2}$.

Multi-dimensional SDE: $dX(t) = a\big(X(t), t\big)\, dt + B\, dW(t)$, $X(0) \sim p(\,\cdot\,, 0)$, where $B \in \mathbb{R}^{d\times d}$ and W is d-dim. Brownian motion.

General Fokker–Planck:

$$\partial_t p(x,t) \;=\; -\,\nabla_x \cdot \big[a(x,t)\, p(x,t)\big] \;+\; \tfrac{1}{2} \sum_{i,j=1}^{d} \partial_{x_i} \partial_{x_j} \big[\big(BB^\top\big)_{ij}\, p(x,t)\big].$$

Isotropic noise ($B = \sqrt{2}\, I_d$): $BB^\top = 2I_d \Rightarrow$

$$\partial_t p = -\,\nabla_x \cdot (a\, p) \;+\; \Delta_x p, \qquad \Delta_x := \sum_{k=1}^{d} \partial_{x_k}^2.$$

OU coefficients: $a(x) = -x$, $B = \sqrt{2} I_d \Longrightarrow$

$$\partial_t p(x,t) \;=\; \nabla_x \cdot \big[x\, p(x,t)\big] + \Delta_x p(x,t), \qquad p(x,0) = p_{\mathrm{data}}(x).$$

Equilibrium (all d): $p_\infty(x) = \dfrac{1}{(2\pi)^{d/2}} \exp\!\big(-\tfrac{1}{2}\|x\|^2\big)$.

Time-Reversal of the OU Process

Forward Fokker–Planck

$$\partial_t p = \nabla_x \cdot (xp) + \Delta_x p.$$

Substitute $t = T - \tau$ (τ increases backward):

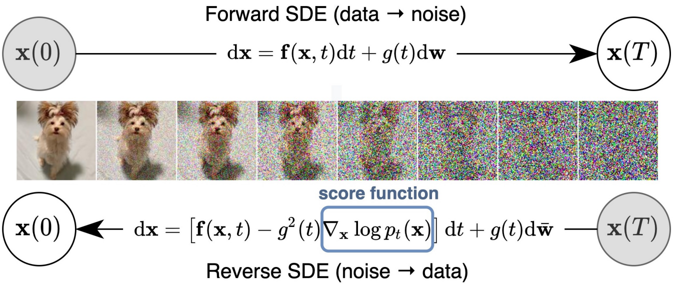
$$\partial_\tau p(x, T-\tau) = -\nabla_x \cdot \left[p\left(x + 2\nabla_x \log p\right)\right] + \Delta_x p.$$

Reverse SDE:

$$d\widetilde{X}_t = -\left(\widetilde{X}_t + 2\nabla_x \log p(\widetilde{X}_t, t)\right)dt + \sqrt{2}\, dW'_t.$$

Starting from $\widetilde{X}_T \sim \mathcal{N}(0, I)$ and integrating to $t = 0$ recovers p_{data}.

Application: Generative Modeling



- Forward SDE (data $\to$ noise):
  - Adds controlled noise via $d\mathbf{x} = f(\mathbf{x}, t)\, dt + g(t)\, d\mathbf{W}$.
  - Transforms complex data distributions into tractable priors.
- Reverse SDE (noise $\to$ data):
  - Removes noise by leveraging the *score* $\nabla_{\mathbf{x}} \log p_t(\mathbf{x})$.
  - Recovers original data samples by integrating backward in time.

References

[1] Lawrence Evans. An introduction to stochastic differential equation. *Book*, pages 1–139, 07 2006.

[2] Nan Chen, Stephen Wiggins, and Marios Andreou. Taming uncertainty in a complex world: The rise of uncertainty quantification—a tutorial for beginners. *Notices of the American Mathematical Society*, 72(3):250–260, March 2025.

[3] Borong Zhang, Martín Guerra, Qin Li, and Leonardo Zepeda-Núñez. Back-projection diffusion: Solving the wideband inverse scattering problem with diffusion models, 2025.

Thank you for your attention! Questions?