

STAT 6900: Stochastic Processes in Modern  
Machine Learning

Lecture 3: Stochastic Differential Equations and  
Diffusion Models

**Farzad Sabzikar**

Office: 2218 Snedecor Hall      [sabzikar@iastate.edu](mailto:sabzikar@iastate.edu)

February 8, 2026

## Session Objective

- Understand stochastic differential equations as models for diffusion processes.
- Study existence and uniqueness of solutions and their qualitative behavior.
- Analyze explicit examples, with emphasis on the Ornstein–Uhlenbeck process.
- Introduce numerical approximation schemes that motivate diffusion limits of stochastic optimization algorithms.

# Stochastic Differential Equations

## Definition 3.1

Let  $\{B_t; t \geq 0\}$  be a Brownian motion defined on a probability space  $(\Omega, \mathcal{F}, \mathbb{P})$ . Let  $\{\mathcal{F}_t; t \geq 0\}$  be a filtration such that  $B_t$  is  $\mathcal{F}_t$ -adapted and, for any  $0 \leq s < t$ , the increment  $B_t - B_s$  is independent of  $\mathcal{F}_s$ . We consider stochastic differential equations of the form

$$dX_t = b(t, X_t) dt + \sigma(t, X_t) dB_t, \quad (1)$$

with initial condition  $X_0$ , where  $X_0$  is a random variable independent of the Brownian motion  $\{B_t\}$ . The functions  $b(t, x)$  and  $\sigma(t, x)$  are called the *drift* and *diffusion* coefficients, respectively.

If the diffusion coefficient vanishes, that is,  $\sigma(t, x) \equiv 0$ , then the stochastic differential equation (1) reduces to the ordinary differential equation  $\frac{dX_t}{dt} = b(t, X_t)$ . In the linear case  $b(t, x) = b(t)x$ , the solution is given explicitly by  $X_t = X_0 \exp\left(\int_0^t b(s) ds\right)$ .

# Stochastic Interpretation

The stochastic differential equation (1) admits the following heuristic interpretation. For a small time increment  $\Delta t > 0$ , the increment

$$\Delta X_t = X_{t+\Delta t} - X_t$$

can be approximately decomposed as

$$\Delta X_t \approx b(t, X_t) \Delta t + \sigma(t, X_t) \Delta B_t,$$

where  $\Delta B_t = B_{t+\Delta t} - B_t$  is interpreted as a random impulse. Since  $\Delta B_t \sim \mathcal{N}(0, \Delta t)$ , the approximate distribution of  $\Delta X_t$  is normal with mean  $b(t, X_t)\Delta t$  and variance  $\sigma(t, X_t)^2\Delta t$ .

# Integral Form of an SDE

A formal meaning of equation (1) is obtained by rewriting it in integral form using stochastic integrals:

$$X_t = X_0 + \int_0^t b(s, X_s) ds + \int_0^t \sigma(s, X_s) dB_s. \quad (2)$$

That is, the solution  $\{X_t; t \geq 0\}$  is an Itô process. The solutions of stochastic differential equations are called *diffusion processes*.

# Existence and Uniqueness of Solutions

## Theorem 3.1

Fix a time interval  $[0, T]$ . Suppose that the coefficients of equation (1) satisfy the following Lipschitz and linear growth conditions: there exist constants  $D_1, D_2, C_1, C_2 > 0$  such that, for all  $x, y \in \mathbb{R}$  and all  $t \in [0, T]$ ,

$$|b(t, x) - b(t, y)| \leq D_1|x - y|, \quad (3)$$

$$|\sigma(t, x) - \sigma(t, y)| \leq D_2|x - y|, \quad (4)$$

$$|b(t, x)| \leq C_1(1 + |x|), \quad (5)$$

$$|\sigma(t, x)| \leq C_2(1 + |x|). \quad (6)$$

Assume that  $X_0$  is a random variable independent of the Brownian motion  $\{B_t; 0 \leq t \leq T\}$  and satisfies  $\mathbb{E}(X_0^2) < \infty$ . Then there exists a unique continuous and  $\{\mathcal{F}_t\}$ -adapted stochastic process  $\{X_t; t \in [0, T]\}$  such that

$$\mathbb{E}\left(\int_0^T |X_s|^2 ds\right) < \infty,$$

and which satisfies equation (2).

# Remarks on Existence and Uniqueness

## Remark 3.1

- (1) The existence and uniqueness result remains valid in higher dimensions. More precisely, if  $\{B_t\}$  is an  $m$ -dimensional Brownian motion,  $\{X_t\}$  is an  $n$ -dimensional process, and the coefficients satisfy

$$b : [0, T] \times \mathbb{R}^n \rightarrow \mathbb{R}^n, \quad \sigma : [0, T] \times \mathbb{R}^n \rightarrow \mathbb{R}^{n \times m},$$

under analogous Lipschitz and linear growth conditions, then the same conclusion holds.

- (2) The linear growth conditions (5)–(6) ensure that the solution does not explode on the time interval  $[0, T]$ .

## Remarks on Existence and Uniqueness (continued)

### Example 3.1

Consider the deterministic differential equation

$$\frac{dX_t}{dt} = X_t^2, \quad X_0 = 1, \quad 0 \leq t \leq 1.$$

The unique solution is

$$X_t = \frac{1}{1-t}, \quad 0 \leq t < 1,$$

which diverges at time  $t = 1$ .

## Remarks on Existence and Uniqueness (continued)

### Remark 3.2

- (3) The Lipschitz conditions (3)–(4) guarantee uniqueness of the solution.

### Example 3.2

Consider the deterministic differential equation

$$\frac{dX_t}{dt} = 3X_t^{2/3}, \quad X_0 = 0.$$

This equation admits infinitely many solutions. Indeed, for any  $a > 0$ , the function

$$X_t = \begin{cases} 0, & t \leq a, \\ (t - a)^3, & t > a, \end{cases}$$

is a solution. In this case, the coefficient  $b(x) = 3x^{2/3}$  does not satisfy the Lipschitz condition, since its derivative is unbounded near  $x = 0$ .

## Remarks on Existence and Uniqueness (continued)

### Remark 3.3

- (4) If the coefficients  $b(t, x)$  and  $\sigma(t, x)$  are differentiable with respect to  $x$ , then the Lipschitz conditions are equivalent to the boundedness of the partial derivatives  $\partial b/\partial x$  and  $\partial \sigma/\partial x$ , with bounds given by  $D_1$  and  $D_2$ , respectively.

# Explicit Solutions of Stochastic Differential Equations

## Example 3.3 Linear equations

The *geometric Brownian motion*

$$X_t = X_0 \exp\left(\left(\mu - \frac{\sigma^2}{2}\right)t + \sigma B_t\right)$$

solves the linear stochastic differential equation

$$dX_t = \mu X_t dt + \sigma X_t dB_t.$$

## Linear Stochastic Differential Equations (continued)

More generally, consider the homogeneous linear stochastic differential equation

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

where  $b(t)$  and  $\sigma(t)$  are continuous functions. Its solution is given explicitly by

$$X_t = X_0 \exp\left(\int_0^t \left(b(s) - \frac{1}{2}\sigma^2(s)\right) ds + \int_0^t \sigma(s) dB_s\right).$$

# Explicit Solutions of Stochastic Differential Equations

## Example 3.4 Ornstein–Uhlenbeck process

Consider the stochastic differential equation

$$dX_t = a(m - X_t) dt + \sigma dB_t, \quad X_0 = x,$$

where  $a, \sigma > 0$  and  $m \in \mathbb{R}$ . This is a nonhomogeneous linear stochastic differential equation. To solve it, we make use of the method of variation of constants. Consider first the associated homogeneous equation

$$dx_t = -ax_t dt, \quad x_0 = x,$$

whose solution is

$$x_t = xe^{-at}.$$

We make the change of variables

$$X_t = Y_t e^{-at}, \quad \text{that is,} \quad Y_t = X_t e^{at}.$$

# Ornstein–Uhlenbeck Process (continued)

## Example 3.4

The process  $\{Y_t\}$  satisfies

$$dY_t = aX_t e^{at} dt + e^{at} dX_t = ame^{at} dt + \sigma e^{at} dB_t.$$

Thus,

$$Y_t = x + m(e^{at} - 1) + \sigma \int_0^t e^{as} dB_s.$$

This implies

$$X_t = m + (x - m)e^{-at} + \sigma e^{-at} \int_0^t e^{as} dB_s.$$

# Ornstein–Uhlenbeck Process (continued)

## Example 3.4

The stochastic process  $\{X_t\}$  is Gaussian. Its mean and covariance function are given by

$$\mathbb{E}(X_t) = m + (x - m)e^{-at}.$$

Moreover,

$$\begin{aligned}\text{Cov}(X_t, X_s) &= \sigma^2 e^{-a(t+s)} \mathbb{E} \left[ \left( \int_0^t e^{ar} dB_r \right) \left( \int_0^s e^{ar} dB_r \right) \right] \\ &= \sigma^2 e^{-a(t+s)} \int_0^{t \wedge s} e^{2ar} dr \\ &= \frac{\sigma^2}{2a} \left( e^{-a|t-s|} - e^{-a(t+s)} \right).\end{aligned}$$

# Ornstein–Uhlenbeck Process (continued)

## Example 3.4

The law of  $X_t$  is the normal distribution

$$X_t \sim \mathcal{N}\left(m + (x - m)e^{-at}, \frac{\sigma^2}{2a}(1 - e^{-2at})\right).$$

As  $t \rightarrow \infty$ , this distribution converges to the normal law

$$\nu = \mathcal{N}\left(m, \frac{\sigma^2}{2a}\right).$$

This distribution is called the *invariant* or *stationary* distribution. If the initial condition  $X_0$  has distribution  $\nu$ , then for each  $t > 0$  the law of  $X_t$  is also  $\nu$ . Indeed, in this case,

$$X_t = m + (X_0 - m)e^{-at} + \sigma e^{-at} \int_0^t e^{as} dB_s.$$

# Ornstein–Uhlenbeck Process (continued)

## Example 3.4

If the initial condition  $X_0$  has the invariant distribution

$$\nu = \mathcal{N}\left(m, \frac{\sigma^2}{2a}\right),$$

then  $\mathbb{E}(X_0) = m$  and  $\text{Var}(X_0) = \frac{\sigma^2}{2a}$ . Therefore,

$$\mathbb{E}(X_t) = m + (\mathbb{E}(X_0) - m)e^{-at} = m.$$

Moreover,

$$\begin{aligned}\text{Var}(X_t) &= e^{-2at} \text{Var}(X_0) + \sigma^2 e^{-2at} \mathbb{E}\left[\left(\int_0^t e^{as} dB_s\right)^2\right] \\ &= \frac{\sigma^2}{2a}.\end{aligned}$$

# Ornstein–Uhlenbeck Process: Why It Matters

## Remark 3.4

The Ornstein–Uhlenbeck (OU) process

$$dX_t = a(m - X_t) dt + \sigma dB_t$$

will reappear throughout the rest of this course.

- Near local minima, diffusion and SME approximations of SGD lead to OU-type dynamics.
- Linearized continuous-time models in reinforcement learning yield OU processes.
- OU processes provide the local stability model for neural SDEs near equilibria.

Thus, the OU process is not only an explicit example, but also a key local model for learning dynamics studied later in the course.

# Numerical Approximation of SDEs: Euler–Maruyama

Most stochastic differential equations cannot be solved explicitly. We therefore seek numerical schemes that approximate sample paths of their solutions.

Consider the stochastic differential equation

$$dX_t = b(X_t) dt + \sigma(X_t) dB_t, \quad X_0 = x,$$

where  $b$  and  $\sigma$  satisfy Lipschitz and linear growth conditions.  
Fix a time horizon  $[0, T]$  and a partition

$$t_i = \frac{iT}{n}, \quad i = 0, 1, \dots, n,$$

with step size  $\Delta t = \frac{T}{n}$ .

## Euler–Maruyama Scheme

The Euler–Maruyama scheme is defined recursively by

$$X^{(n)}(t_i) = X^{(n)}(t_{i-1}) + b\left(X^{(n)}(t_{i-1})\right)\Delta t + \sigma\left(X^{(n)}(t_{i-1})\right)\Delta B_i,$$

for  $i = 1, \dots, n$ , where

$$\Delta B_i = B_{t_i} - B_{t_{i-1}}.$$

The initial condition is

$$X^{(n)}(0) = x.$$

Inside each interval  $(t_{i-1}, t_i)$ , the process  $X^{(n)}$  is defined by linear interpolation.

# Simulation Interpretation

Since Brownian increments satisfy

$$\Delta B_i \sim \mathcal{N}(0, \Delta t),$$

a numerical simulation of the Euler–Maruyama scheme is obtained by generating independent standard normal random variables  $\xi_i \sim \mathcal{N}(0, 1)$  and setting

$$\Delta B_i = \sqrt{\Delta t} \xi_i.$$

Thus, the update rule becomes

$$X^{(n)}(t_i) = X^{(n)}(t_{i-1}) + b\left(X^{(n)}(t_{i-1})\right) \Delta t + \sigma\left(X^{(n)}(t_{i-1})\right) \sqrt{\Delta t} \xi_i.$$

This form makes the connection with stochastic gradient methods transparent.

## Error and Convergence of Euler–Maruyama

Let  $X_t$  denote the exact solution of the SDE and  $X_t^{(n)}$  its Euler–Maruyama approximation. Define the strong error at time  $T$  by

$$e_n = \left( \mathbb{E} \left[ \left| X_T - X_T^{(n)} \right|^2 \right] \right)^{1/2}.$$

Under standard Lipschitz and linear growth assumptions, there exist constants  $c > 0$  and  $n_0$  such that

$$e_n \leq c (\Delta t)^{1/2}, \quad \text{for all } n \geq n_0.$$

Thus, Euler–Maruyama converges with strong order 1/2.

# Euler–Maruyama and Stochastic Gradient Descent

The Euler–Maruyama scheme has the same structure as stochastic gradient descent:

$$X_{k+1} = X_k + \eta b(X_k) + \sqrt{\eta} \sigma(X_k) \xi_k.$$

Here:

- the step size  $\eta$  corresponds to  $\Delta t$ ,
- the noise term  $\xi_k$  models stochastic gradient noise,
- the drift term  $b$  represents the deterministic gradient flow.

This observation provides the basis for diffusion and SME approximations of SGD studied later in the course.

# Essential Concept 1: Markov Property

## Definition 3.2 Markov property

A stochastic process  $\{X_t; t \geq 0\}$  is a *Markov process* if, for every  $0 \leq s < t$ , the conditional law of the future given the past depends only on the present state  $X_s$ . Equivalently, for any bounded Borel function  $f$ ,

$$\mathbb{E}[f(X_t) \mid \sigma(X_r : r \leq s)] = \mathbb{E}[f(X_t) \mid X_s].$$

## Essential Concept 2: Diffusion Processes

### Definition 3.3 Diffusion process

A stochastic process  $\{X_t; t \geq 0\}$  is called a *diffusion process* if it is a solution of a stochastic differential equation of the form

$$dX_t = b(X_t) dt + \sigma(X_t) dB_t, \quad X_0 = x,$$

where  $b$  is the *drift* coefficient,  $\sigma$  is the *diffusion* coefficient, and  $\{B_t\}$  is a Brownian motion.

### Remark 3.5

Diffusion processes are (under standard conditions) continuous and possess the Markov property.

## Essential Concept 3: Invariant Distribution

### Definition 3.4 Invariant (stationary) distribution

Let  $\{X_t\}$  be a Markov process. A probability distribution  $\nu$  on  $\mathbb{R}^n$  is called *invariant* (or *stationary*) for  $\{X_t\}$  if

$$X_0 \sim \nu \implies X_t \sim \nu \quad \text{for all } t > 0.$$

### Remark 3.6

Invariant distributions describe long-time behavior and equilibrium fluctuations (e.g., for the Ornstein–Uhlenbeck process and for diffusion/SME limits of SGD).

# Summary

- Introduced stochastic differential equations and diffusion processes.
- Discussed existence and uniqueness under Lipschitz and linear growth conditions.
- Studied explicit solutions, highlighting the Ornstein–Uhlenbeck process and its stationary behavior.
- Presented the Euler–Maruyama scheme as a numerical approximation of SDEs.
- Established conceptual foundations for diffusion and SME approximations of SGD.