Introduction to Data Assimilation
Lecture 5 on Oct. 30, 2025

Quanling Deng*

Abstract

This lecture focuses on Itô stochastic integral, Itô's formula, Fokker-Planck equation, and Markov jump process.

1 Itô Stochastic Integral and Its Properties

1.1 Definition

Let $(W_t)_{t\geq 0}$ be a standard real Wiener process on a probability space $(\Omega, \mathcal{F}, \mathbb{P})$, and let $\{\mathcal{F}_t\}$ be its natural filtration. Let $\phi : [0,T] \times \Omega \to \mathbb{R}$ be an *adapted*, square-integrable process satisfying

$$\mathbb{E}\Big[\int_0^T \phi(t)^2\, dt\Big] < \infty.$$

The *Itô stochastic integral* of ϕ with respect to W_t over $[0,T]$ is defined as the $L^2(\Omega)$-limit of Riemann sums

$$\int_0^T \phi(t)\, dW_t := \lim_{n\to\infty} \sum_{k=0}^{n-1} \phi(t_k)\big(W_{t_{k+1}} - W_{t_k}\big), \tag{1.1}$$

where $0 = t_0 < t_1 < \cdots < t_n = T$ is a partition with mesh $\max_k(t_{k+1} - t_k) \to 0$.

Here $\phi(t_k)$ must be $\mathcal{F}_{t_k}$-measurable (non-anticipating), ensuring that the integrand depends only on the past information of the Wiener process.

1.2 Construction via Simple Processes

To make (1.1) precise, one first defines the integral for *simple processes*:

$$\phi(t,\omega) = \sum_{k=0}^{n-1} \phi_k(\omega)\, \mathbf{1}_{[t_k,t_{k+1})}(t),$$

where each ϕ_k is $\mathcal{F}_{t_k}$-measurable and square integrable. Then define

$$\int_0^T \phi(t)\, dW_t := \sum_{k=0}^{n-1} \phi_k\big(W_{t_{k+1}} - W_{t_k}\big). \tag{1.2}$$

*Yau Mathematical Science Center, Tsinghua University, Beijing, China 100048. E-mail address: qldeng@tsinghua.edu.cn

This random variable satisfies:

$$\mathbb{E}\Big[\int_0^T \phi(t)\,dW_t\Big] = 0, \qquad \mathbb{E}\Big[\Big|\int_0^T \phi(t)\,dW_t\Big|^2\Big] = \mathbb{E}\Big[\int_0^T \phi(t)^2\,dt\Big].$$

The second equality is known as the *Itô isometry* (proved below).

By completion, the mapping $\phi \mapsto \int_0^T \phi\,dW$ extends uniquely to all square-integrable adapted processes.

1.3 Itô Isometry (Main Property)

Theorem 1.1 (Itô isometry)**.** *For any $\phi \in L^2_{\mathrm{ad}}([0,T]\times\Omega)$,*

$$\mathbb{E}\Big[\Big(\int_0^T \phi(t)\,dW_t\Big)^2\Big] = \mathbb{E}\Big[\int_0^T \phi(t)^2\,dt\Big]. \tag{1.3}$$

Proof (step-by-step). For a simple process (1.2),

$$\int_0^T \phi(t)\,dW_t = \sum_k \phi_k\,\Delta W_k, \qquad \Delta W_k = W_{t_{k+1}} - W_{t_k}.$$

Then

$$\begin{aligned}
\mathbb{E}\Big[\Big(\sum_k \phi_k\,\Delta W_k\Big)^2\Big] &= \sum_{k,\ell} \mathbb{E}[\phi_k\phi_\ell\,\Delta W_k\,\Delta W_\ell] \\
&= \sum_k \mathbb{E}\big[\phi_k^2\,(\Delta W_k)^2\big] \qquad \text{(since ΔW_k are independent, mean zero)} \\
&= \sum_k \mathbb{E}\big[\phi_k^2\big]\,\mathbb{E}\big[(\Delta W_k)^2\big] \\
&= \sum_k \mathbb{E}\big[\phi_k^2\big]\,(t_{k+1}-t_k) = \mathbb{E}\Big[\int_0^T \phi(t)^2\,dt\Big].
\end{aligned}$$

Hence (1.3) holds for simple ϕ, and by density for all admissible ϕ.

1.4 Zero Mean Property

$$\mathbb{E}\Big[\int_0^T \phi(t)\,dW_t\Big] = 0. \tag{1.4}$$

Proof. For simple ϕ,

$$\mathbb{E}\Big[\sum_k \phi_k\,\Delta W_k\Big] = \sum_k \mathbb{E}[\phi_k\,\mathbb{E}[\Delta W_k|\mathcal{F}_{t_k}]] = 0,$$

because $\mathbb{E}[\Delta W_k|\mathcal{F}_{t_k}] = 0$. Extending by limit in L^2 proves the claim for all ϕ.

1.5 Cross-Covariance Property

If $\phi, \psi \in L^2_{\text{ad}}$, then

$$\mathbb{E}\Big[\Big(\int_0^T \phi(t)\, dW_t\Big)\Big(\int_0^T \psi(t)\, dW_t\Big)\Big] = \mathbb{E}\Big[\int_0^T \phi(t)\psi(t)\, dt\Big]. \tag{1.5}$$

Proof. Same as the Itô isometry, expanding cross terms:

$$\mathbb{E}\Bigg[\sum_{k,\ell} \phi_k\psi_\ell\, \Delta W_k\, \Delta W_\ell\Bigg] = \sum_k \mathbb{E}\big[\phi_k\psi_k\, (\Delta W_k)^2\big] = \sum_k \mathbb{E}[\phi_k\psi_k](t_{k+1} - t_k).$$

Taking the limit yields (1.5).

2 Fokker–Planck Equation

2.1 Itô's Formula: Derivation from SDE

Consider the stochastic differential equation (SDE)

$$dx_t = a_t\, dt + b_t\, dW_t, \tag{2.1}$$

where $a_t = a(x(t), t)$ and $b_t = b(x(t), t)$ are nonlinear functions of both x and t.

Remark 2.1. Herein, x_t is understood as a time-dependent random variable $x(t)$. This equation is sometimes rewritten as $\frac{dx_t}{dt} = a_t + b_t\dot{W}_t$, which is known as the physics formulation. Without the $b_t\dot{W}_t$, it is an ODE. $\dot{W}_t$, called the (Gaussian) white noise, can be understood as a stochastic process dW_t/dt, however, the Wiener process is a.s. nowhere differentiable.

Let $f(x)$ be a smooth deterministic function. By Itô's formula,

$$\begin{aligned} df(x_t) &= f'(x_t)\, dx_t + \tfrac{1}{2} f''(x_t)\, (dx_t)^2 \\ &= f'(x_t)\, (a_t\, dt + b_t\, dW_t) + \tfrac{1}{2} f''(x_t)(b_t)^2\, dt \\ &= \Big[f'(x_t)a_t + \tfrac{1}{2} f''(x_t) b_t^2\Big] dt + f'(x_t) b_t\, dW_t. \end{aligned} \tag{2.2}$$

In the stochastic setting, $(dW_t)^2 = dt$, so the second-order Taylor term is of the same order as dt and cannot be ignored.

Remark 2.2. The derivation uses the definition and removes the higher-order terms, keeping in mind that $(dW_t)^2 = dt$. If one consider a smooth function in the form of two variables, $f(x_t, t)$, then Itô's formula is

$$\begin{aligned} df(x_t, t) &= \frac{\partial f}{\partial t}\, dt + \frac{\partial f}{\partial x_t}\, dx_t + \tfrac{1}{2}\frac{\partial^2 f}{\partial x_t^2}(dx_t)^2 \\ &= \left(\frac{\partial f}{\partial t} + a_t\frac{\partial f}{\partial x_t} + \tfrac{1}{2} b_t^2 \frac{\partial^2 f}{\partial x_t^2}\right) dt + b_t \frac{\partial f}{\partial x_t}\, dW_t. \end{aligned} \tag{2.3}$$

2.2 From Itô's Formula to the Fokker–Planck Equation

Let $p(x,t)$ denote the probability density function (PDF) of $x(t)$. For any smooth function f, its expectation is

$$\langle f(x(t))\rangle = \int_{-\infty}^{\infty} f(x)\, p(x,t)\, dx. \tag{2.4}$$

Taking the time derivative gives

$$\frac{d}{dt}\langle f(x(t))\rangle = \int_{-\infty}^{\infty} f(x)\, \frac{\partial p(x,t)}{\partial t}\, dx. \tag{2.5}$$

Applying the expectation operator to (2.2) and using $\langle dW_t\rangle = 0$ gives

$$\frac{d}{dt}\langle f(x(t))\rangle = \left\langle a(x,t)\, f_x + \tfrac{1}{2}b(x,t)^2\, f_{xx}\right\rangle = \int_{-\infty}^{\infty}\left[a(x,t)f_x + \tfrac{1}{2}b(x,t)^2 f_{xx}\right]p(x,t)\, dx. \tag{2.6}$$

Integrating by parts and assuming $p(x,t) \to 0$ as $|x| \to \infty$, one obtains

$$\frac{d}{dt}\langle f(x(t))\rangle = \int_{-\infty}^{\infty} f(x)\left[-\frac{\partial}{\partial x}\big(a(x,t)p(x,t)\big) + \frac{1}{2}\frac{\partial^2}{\partial x^2}\big(b(x,t)^2 p(x,t)\big)\right]dx. \tag{2.7}$$

Comparing (2.5) and (2.7), the <u>Fokker–Planck equation</u> follows:

$$\boxed{\frac{\partial p(x,t)}{\partial t} = -\frac{\partial}{\partial x}\big(a(x,t)p(x,t)\big) + \frac{1}{2}\frac{\partial^2}{\partial x^2}\big(b(x,t)^2 p(x,t)\big).} \tag{2.8}$$

2.3 Interpretation

Equation (2.8) is a <u>linear</u> partial differential equation describing the time evolution of the PDF of the SDE solution.

- The <u>drift term</u>

$$-\frac{\partial}{\partial x}\big(a(x,t)p(x,t)\big)$$

represents deterministic advection of the probability density. - The <u>diffusion term</u>

$$\frac{1}{2}\frac{\partial^2}{\partial x^2}\big(b(x,t)^2 p(x,t)\big)$$

corresponds to the spreading of probability due to stochastic noise.

2.4 Two Special Cases

1. Deterministic case: If $dx = dt$ and $a_t = 1$, then (the following eqn is also known as the Liouville equation)

$$\frac{\partial p}{\partial t} = -\frac{\partial p}{\partial x}, \quad p(x,0) = \delta(x),$$

whose solution is

$$p(x,t) = \delta(x-t),$$

meaning the probability mass moves deterministically along $x = t$.

2. Pure diffusion (Brownian motion): If $dx = dW_t$ and $b_t = 1$, then

$$\frac{\partial p}{\partial t} = \frac{1}{2}\frac{\partial^2 p}{\partial x^2},$$

which is the **heat equation**, with Gaussian solution

$$p(x,t) = \frac{1}{\sqrt{2\pi t}}\, e^{-x^2/(2t)}.$$

2.5 Multi-Dimensional Extension

For the N-dimensional SDE

$$d\mathbf{x} = \boldsymbol{\mu}(\mathbf{x},t)\,dt + \boldsymbol{\sigma}(\mathbf{x},t)\,d\mathbf{W}_t,$$

where $\boldsymbol{\sigma}$ is an $N \times M$ matrix and $\mathbf{W}_t$ is an M-dimensional Wiener process, the Fokker–Planck equation becomes

$$\frac{\partial p(\mathbf{x},t)}{\partial t} = -\sum_{i=1}^{N} \frac{\partial}{\partial x_i}\big(\mu_i(\mathbf{x},t)p\big) + \frac{1}{2}\sum_{i,j=1}^{N} \frac{\partial^2}{\partial x_i \partial x_j}\big(D_{ij}(\mathbf{x},t)p(\mathbf{x},t)\big), \tag{2.9}$$

where

$$D_{ij}(\mathbf{x},t) = \sum_{k=1}^{M} \sigma_{ik}\sigma_{jk}$$

is the diffusion tensor.

Remark 2.3. Solving the Fokker–Planck equation (which is a PDE) directly becomes computationally prohibitive in high dimensions (beyond 3–4D) using standard PDE solvers. In weather and climate models, it can be millions or billions. Monte Carlo simulations or reduced-order models are typically used to approximate the evolution of the PDF.

3 Markov Jump Processes

3.1 Definition

A Markov jump process is a stochastic process whose state changes at random discrete times among a finite or countable set of states. Unlike SDEs with continuous-valued states, Markov jump processes are defined on discrete state spaces, making them ideal for modeling systems that switch between a finite number of regimes or modes.

Formally, if $\gamma(t)$ takes values in $\{s_1, s_2, \ldots, s_m\}$ and transitions with probabilities

$$\mathbb{P}\big[\gamma(t+\Delta t) = s_j \,|\, \gamma(t) = s_i\big] = q_{ij}\,\Delta t + o(\Delta t),$$

then $Q = [q_{ij}]$ is called the transition rate matrix or generator of the process.

3.2 A Motivating Example: Intermittent Time Series

Many physical systems exhibit intermittent instabilities, irregular switching between stable and unstable phases, such as the Rossby wave phenomena in atmospheric and oceanic flows.

Consider a two-dimensional system:

$$du = [(-\gamma + i\omega_u)u + f]\,dt + \sigma_u\,dW_u, \tag{3.1a}$$

$$\gamma \text{ follows a two-state Markov jump process.} \tag{3.1b}$$

Here, ω_u, f, and σ_u are constants, and γ randomly switches between two values (e.g.):

$$\gamma_+ = 2.27 \quad \text{(stable phase)}, \qquad \gamma_- = -0.04 \quad \text{(unstable phase)}.$$

When $\gamma = \gamma_+$, the system is damped, and u decays exponentially. When $\gamma = \gamma_-$, the damping becomes negative, and u grows exponentially — producing bursts of large amplitude.

The energy of u,

$$E(u) = (\Re[u])^2 + (\Im[u])^2,$$

grows rapidly in the unstable phase and decays in the stable phase. Random switching between the two regimes generates intermittent dynamics and non-Gaussian statistics.

Remark 3.1. This coupling of a linear SDE with a Markov jump process models systems that alternate randomly between stability and instability, producing intermittent behavior similar to turbulence bursts or climate transitions.

3.3 Statistical Interpretation

The Markov jump process introduces non-Gaussianity and fat tails in the probability distribution of the system variables. In the example (3.1), the real and imaginary parts of u display heavy-tailed PDFs due to random amplification during unstable intervals.

3.4 Practical Use

Markov jump processes are widely used for:

- Modeling regime-switching systems (e.g., weather regimes, market volatility states);
- Representing unresolved dynamics through stochastic parameterization;
- Capturing intermittent phenomena and non-Gaussian features in reduced models.