Lecture 13 SDE and Ito's formula *

Tiejun Li

1 White noise

In physics literature, the physicists usually use the stochastic differential equations (SDEs) like

$$\dot{X}_t = b(X_t, t) + \sigma(X_t, t)\dot{W}_t, \quad X|_{t=0} = X_0, \tag{1.1}$$

where $\dot{W}_t$ is called the *temporal Gaussian white noise*, which is the formal derivative of the Brownian motion W_t with respect to time. Its formal definition is that it is a Gaussian process with mean and covariance functions as

$$m(t) = \mathbb{E}(\dot{W}_t) = 0, \quad K(s,t) = \mathbb{E}(\dot{W}_s\dot{W}_t) = \delta(t-s).$$

It can be formally understood as

$$m(t) = \frac{d}{dt}\mathbb{E}(W_t) = 0, \quad K(s,t) = \frac{\partial^2}{\partial s \partial t}\mathbb{E}(W_sW_t) = \frac{\partial^2}{\partial s \partial t}(s \wedge t) = \delta(t-s).$$

The name *white noise* comes from its power spectral density (PSD) $S(\omega)$ defined as the Fourier transform of its autocorrelation function $R(t) = \mathbb{E}(\dot{W}_0\dot{W}_t) = \delta(t)$, thus $S(\omega) = \widehat{(\delta(t))} = 1$ which corresponds to a flat constant at all frequencies ω. We call it white as an analogy to the frequency spectrum of white light. If the frequency spectrum of the noise is not flat, it is called colored noise. From practical point view, the white noise is not physical since it requires infinite energy

$$E = \int_{-\infty}^{\infty} S(\omega)d\omega = \infty.$$

From the regularity theory of the Brownian motion, the function $\dot{W}$ is meaningless since W_t has less than half order smoothness. In fact, it is not a traditional function but a distribution [1]. However, the rigorous mathematical foundation of the white noise calculus can be also established [2]. But we will only introduce the Itô's classical way to establish the well-posedness of the stochastic differential equations.

*School of Mathematical Sciences, Peking University, Beijing 100871, P.R. China

Mathematically, the SDEs (1.1) are often denoted as

$$dX_t = b(X_t,t)dt + \sigma(X_t,t)dW_t, \tag{1.2}$$

to avoid the ambiguity of the white noise, where W_t is the standard Wiener process. X_t may be viewed as a process induced by W_t. If there is no term $\sigma(X_t,t)dW_t$, it is a deterministic ODEs. The effect of $b(X_t,t)$ is to drive the mean position of the system, while the effect of $\sigma(X_t,t)dW_t$ is to diffuse around the mean position which we will see later. To make sense of (1.2), one natural way is to define X_t through its integral form

$$X_t = X_0 + \int_0^t b(X_s,s)ds + \int_0^t \sigma(X_s,s)dW_s. \tag{1.3}$$

We will show the first mathematical issue is how to define the integral $\int_0^t \sigma(X_s,s)dW_s$ involving Brownian motion.

2 Itô integral

First suppose X_t is continuous with respect to time t. For a fixed sample ω, we borrow the idea for defining the Riemann-Stieljes integral to make the definition

$$\int_0^t \sigma(X_s,s)dW_s = \lim_{|\Delta|\to 0} \sum_j \sigma(X_j,t_j^*)\Big(W_{t_{j+1}} - W_{t_j}\Big),$$

where Δ is a subdivision of $[0,t]$, X_j is the function value $X_{t_j^*}$ and t_j^* is chosen from the interval $[t_j,t_{j+1}]$. One critical issue about the above definition is that it depends on the choice of t_j^* when we are handling W_t, which has unbounded variation in any interval almost surely.

To have a sense on this, consider the Riemann-Stieltjes integral to $\int_a^b f(t)dg(t)$, where f and g are all assumed continuous. So

$$\int_a^b f(t)dg(t) \approx \sum_j f_j\Big(g(t_{j+1}) - g(t_j)\Big). \tag{2.1}$$

If one takes another value for f_j in $[t_j,t_{j+1}]$ under the same subdivision, then

$$\int_a^b f(t)dg(t) \approx \sum_j \tilde{f}_j\Big(g(t_{j+1}) - g(t_j)\Big).$$

If $g(t)$ has bounded total variation, we subtract the right hand side of the above two definitions and obtain

$$\begin{aligned}\Big|\sum_j (f_j - \tilde{f}_j)\Big(g(t_{j+1}) - g(t_j)\Big)\Big| &\le \max_j |f_j - \tilde{f}_j| \sum_j \Big|g(t_{j+1}) - g(t_j)\Big| \\ &\le \max_j |f_j - \tilde{f}_j| V(g;[a,b]) \to 0\end{aligned}$$

as $|\Delta| \to 0$ by the uniform continuity of f on $[a, b]$. Thus we get a well-defined definition which is independent of the choice of reference point in the approximation. If $g(t) = W_t(\omega)$, let us see what will happen in this case.

Example 2.1. *Different choices for the stochastic integral $\int_0^T W_t dW_t$.*

Choice 1: Leftmost endpoint integral.

$$\int_0^T W_t dW_t \approx \sum_j W_{t_j}(W_{t_{j+1}} - W_{t_j}) := I_N^L.$$

Choice 2: Rightmost endpoint integral.

$$\int_0^T W_t dW_t \approx \sum_j W_{t_{j+1}}(W_{t_{j+1}} - W_{t_j}) := I_N^R.$$

Choice 3: Midpoint integral.

$$\int_0^T W_t dW_t \approx \sum_j W_{t_{j+\frac{1}{2}}}(W_{t_{j+1}} - W_{t_j}) := I_N^M.$$

Without looking into the exact pathwise result for the three choices, we have the following identities from the statistical average sense.

$$\begin{aligned}
\mathbb{E}(I_N^L) &= \sum_j \mathbb{E}W_{t_j}\mathbb{E}(W_{t_{j+1}} - W_{t_j}) = 0, \\
\mathbb{E}(I_N^R) &= \sum_j \Big[\mathbb{E}(W_{t_{j+1}} - W_{t_j})^2 + \mathbb{E}W_{t_j}\mathbb{E}(W_{t_{j+1}} - W_{t_j})\Big] = \sum_j \Delta t_j = T, \\
\mathbb{E}(I_N^M) &= \mathbb{E}\Big[\sum_j W_{t_{j+\frac{1}{2}}}(W_{t_{j+1}} - W_{t_{j+\frac{1}{2}}}) + \sum_j W_{t_{j+\frac{1}{2}}}(W_{t_{j+\frac{1}{2}}} - W_{t_j})\Big] \\
&= \sum_j \mathbb{E}(W_{t_{j+\frac{1}{2}}} - W_{t_j})^2 = \sum_j (t_{j+\frac{1}{2}} - t_j) = \frac{T}{2}.
\end{aligned}$$

The reason is that the Brownian motion has unbounded variations for any finite interval. The example above also shows that we should take special attention to stochastic integrals.

One important remark on the definition of stochastic integrals like (2.1) is that it can not be defined for arbitrary continuous functions f, otherwise the function g must have bounded variations on compacts [9]. To overcome this issue, one rescue is to restrict the integrands to be a special class of functions, the adapted processes. That is the key point of the well-known Itô integral to be introduced below.

The first stochastic integral which is studied rigorously in the history is Itô's leftmost endpoint integral [4], which is named Itô integral from then on. It turns out that the different choices of the reference point correspond to different consistent definitions of stochastic integrals under suitable conditions, but they can be connected by some simple transformation rules (See [9], Theorem 30 in Chapter 5). To understand Itô's definition for stochastic integral, we take the filtration generated by standard Wiener process as $\mathcal{F}_t^W$ (we also assume all of the sets of measure zero has been contained in $\mathcal{F}_t^W$). The construction of Itô integral takes the leftmost endpoint approximation

$$\int_0^T f(t,\omega)dW_t \approx \sum_j f_{t_j}(W_{t_{j+1}} - W_{t_j}).$$

Mathematically, to understand Itô integral, we need the concept *simple function* which takes the form

$$f(t,\omega) = \sum_{j=1}^n e_j(\omega)\chi_{[t_j,t_{j+1})}(t), \tag{2.2}$$

where $e_j(\omega)$ is $\mathcal{F}_{t_j}^W$-measurable and $\chi_{[t_j,t_{j+1})}(t)$ is the indicator function on $[t_j, t_{j+1})$. It is natural to define

$$\int_0^T f(t,\omega)dW_t = \sum_j e_j(\omega)(W_{t_{j+1}} - W_{t_j}) \tag{2.3}$$

for this choice of simple functions.

Lemma 2.2. *For any $S \le T$, the stochastic integral for the simple functions satisfies*

$$(1) \qquad \mathbb{E}\left(\int_S^T f(t,\omega)dW_t\right) = 0, \tag{2.4}$$

$$(2) \qquad \textit{(Itô isometry)} \quad \mathbb{E}\Big(\int_S^T f(t,\omega)dW_t\Big)^2 = \mathbb{E}\left(\int_S^T f^2(t,\omega)dt\right). \tag{2.5}$$

Proof. The first property is straightforward by the independence between $\Delta W_j := W_{t_{j+1}} - W_{t_j}$ and $e_j(\omega)$ and $\Delta W_j \sim N(0, t_{j+1} - t_j)$. For the second property we have

$$\begin{aligned}
\mathbb{E}\left(\int_S^T f(t,\omega)dW_t\right)^2 &= \mathbb{E}\Big(\sum_j e_j \Delta W_j\Big)^2 = \mathbb{E}\Big(\sum_{j,k} e_j e_k \Delta W_j \Delta W_k\Big) \\
&= \mathbb{E}\Big(\sum_j e_j^2 \Delta W_j^2 + 2\sum_{j<k} e_j e_k \Delta W_j \Delta W_k\Big) \\
&= \sum_j \mathbb{E}e_j^2 \cdot \mathbb{E}\Delta W_j^2 + \sum_{j<k} \mathbb{E}(f_j f_k \Delta W_j) \cdot \mathbb{E}(\Delta W_k) \\
&= \sum_j \mathbb{E}e_j^2 \Delta t_j = \mathbb{E}\left(\int_S^T f^2(t,\omega)dt\right).
\end{aligned}$$

where the last third identity holds because of the independence between ΔW_k and $e_j e_k \Delta W_j$ for $j < k$.

Now for $f(t,\omega)$ which belongs to the class of functions $\mathcal{V}[S,T]$ defined as

(i) f is $\mathcal{B}([0,\infty)) \times \mathcal{F}$-measurable as a function from (t,ω) to $\mathbb{R}$,

(ii) $f(t,\omega)$ is $\mathcal{F}_t^W$-adapted,

(iii) $f \in L_P^2 L_t^2$, that is $\mathbb{E}\left(\int_S^T f^2(t,\omega)dt\right) < \infty$,

we have the approximation property through simple functions $\phi_n(t,\omega)$

$$\mathbb{E}\left(\int_S^T (f(t,\omega) - \phi_n(t,\omega))^2 dt\right) \to 0, \tag{2.6}$$

i.e. $\phi_n \to f$ in $L_P^2 L_t^2$ (c.f. [5,8]). With this setup, we can define the Itô integral as

$$\int_S^T f(t,\omega)dW_t = \lim_{n\to\infty} \int_S^T \phi_n(t,\omega)dW_t \qquad \text{in} \;\; L_P^2. \tag{2.7}$$

From (2.5), $\int_S^T \phi_n(t,\omega)dW_t$ is in L_P^2 for any simple function $\phi_n(t,\omega)$. Furthermore we have

$$\mathbb{E}\Big(\int_S^T \phi_n dW_t - \int_S^T \phi_m dW_t\Big)^2 = \mathbb{E}\left(\int_S^T (\phi_n - \phi_m)^2 dt\right). \tag{2.8}$$

From (2.6), the approximation sequence $\{\phi_n\}$ is a Cauchy sequence in $L_P^2(\Omega; L_t^2[S,T])$. This implies $\{\int_S^T \phi_n dW_t\}$ is also a Cauchy sequence in L_P^2. From the completeness of $L_P^2(\Omega)$, it has a unique limit and we define it as

$$\int_S^T f(t,\omega)dW_t$$

in the definition (2.7). The independence on the choice of the approximating sequence $\{\phi_n\}$ is left as an exercise.

As a natural extension of Lemma 2.2, we have

Theorem 2.3. *For $f \in \mathcal{V}[S,T]$, the Itô integral satisfies*

$$(1) \qquad \mathbb{E}\left(\int_S^T f(t,\omega)dW_t\right) = 0, \tag{2.9}$$

$$(2) \qquad \textit{(Itô isometry)} \quad \mathbb{E}\Big(\int_S^T f(t,\omega)dW_t\Big)^2 = \mathbb{E}\left(\int_S^T f^2(t,\omega)dt\right). \tag{2.10}$$

Proof. Based on Lemma 2.2, we have

$$\begin{aligned}\left|\mathbb{E}\left(\int_S^T f(t,\omega)dW_t\right)\right| &= \left|\mathbb{E}\left(\int_S^T f(t,\omega)dW_t - \int_S^T \phi_n(t,\omega)dW_t\right)\right| \\ &\le \mathbb{E}\left(\int_S^T f(t,\omega)dW_t - \int_S^T \phi_n(t,\omega)dW_t\right)^2 \to 0\end{aligned}$$

by Hölder's inequality and the definition (2.7).

It is a standard result that if $X_n \to X$ in a Hilbert space H, then $|X_n| \to |X|$ and thus $|X_n|^2 \to |X|^2$, where $|\cdot|$ is the correspoding norm in Hilbert space H. So we have

$$\mathbb{E}\left(\int_S^T \phi_n(t,\omega)dW_t\right)^2 \to \mathbb{E}\left(\int_S^T f(t,\omega)dW_t\right)^2 \qquad \text{in } L_P^2(\Omega)$$

and

$$\mathbb{E}\left(\int_S^T \phi_n^2(t,\omega)dt\right) \to \mathbb{E}\left(\int_S^T f^2(t,\omega)dt\right) \qquad \text{in } L_P^2(\Omega; L_t^2[S,T])$$

From the Itô isometry for simple functions, we obtain (2.10) immediately.

The following properties can be proved for the Itô integral easily.

Proposition 2.4. *For $f, g \in \mathcal{V}[S,T]$ and $U \in [S,T]$, we have*

(i) $\int_S^T f dW_t = \int_S^U f dW_t + \int_U^T f dW_t$ *a.s..*

(ii) $\int_S^T (f + cg) dW_t = \int_S^T f dW_t + c\int_S^T g dW_t$ *(c is a constant) a.s..*

(iii) $\int_S^T f dW_t$ *is $\mathcal{F}_t^W$-measurable.*

Furthermore, we have the regularity of the path of the process defined via Itô integral, whose proof may be referred to [5,8,10].

Lemma 2.5. *For $f \in \mathcal{V}[0,T]$, $X_t := \int_0^t f(s,\omega)dW_s$ has continuous trajectories in the almost sure sense.*

We remark that the class of functions $\mathcal{V}[0,T]$ to make sense of the Itô integral and keep the above properties can be weakened by replacing the conditions (ii) and (iii) in $\mathcal{V}[0,T]$ as

(ii)' f is $\mathcal{F}_t$-adapted, where $\{\mathcal{F}_t\}$ is a filtration such that W_t is a $\mathcal{F}_t$-martingale.

(iii)' $\int_0^T f^2(s,\omega)ds < \infty$ almost surely.

The readers may be referred to [5,8,10] for more details. With this weaker setup, one can define the multi-dimensional Ito integral

$$\int_0^T \boldsymbol{\sigma}(t,\omega)\cdot d\boldsymbol{W}_t,$$

where $\boldsymbol{W}_t$ is an m-dimensional Wiener process, and $\boldsymbol{\sigma} \in \mathbb{R}^{n\times m}$ is $\mathcal{F}_t^{\boldsymbol{W}}$-adapted. To compute their expectation, We have the similar property as the Ito isometry

$$\mathbb{E}\left(\int_S^T \sigma(t,\omega)dW_t^j\right) = 0, \qquad \mathbb{E}\Big(\int_S^T \sigma(t,\omega)dW_t^j\Big)^2 = \mathbb{E}\left(\int_S^T \sigma^2(t,\omega)dt\right), \quad \forall j.$$

and especially the cross product expectation

$$\mathbb{E}\left(\int_S^T \sigma_1(t,\omega)dW_t^i \cdot \int_S^T \sigma_2(t,\omega)dW_t^j\right) = 0, \qquad \forall i \neq j,$$
$$\mathbb{E}\Big(\int_S^T \sigma_1(t,\omega)dW_t^j \int_S^T \sigma_2(t,\omega)dW_t^j\Big) = \mathbb{E}\left(\int_S^T \sigma_1(t,\omega)\sigma_2(t,\omega)dt\right), \quad \forall j.$$

Example 2.6. *With Itô integral we have*

$$\int_0^t W_s dW_s = \frac{W_t^2}{2} - \frac{t}{2}. \tag{2.11}$$

Proof. From the definition of Itô integral

$$\begin{aligned}\int_0^t W_s dW_s &\approx \sum_j W_{t_j}(W_{t_{j+1}} - W_{t_j}) = \sum_j \frac{2W_{t_j}W_{t_{j+1}} - 2W_{t_j}^2}{2} \\ &= \sum_j \frac{W_{t_{j+1}}^2 - W_{t_j}^2}{2} - \sum_j \frac{W_{t_{j+1}}^2 - 2W_{t_{j+1}}W_{t_j} + W_{t_j}^2}{2} \\ &= \frac{W_t^2}{2} - \frac{1}{2}\sum_j (W_{t_{j+1}} - W_{t_j})^2 \to \frac{W_t^2}{2} - \frac{t}{2},\end{aligned}$$

where the last limit is due to the fact $\langle W, W\rangle_t = t$ in Proposition **??**.

3 Itô's formula

Let's take the differential form of the identity (2.11), then we have

$$dW_t^2 = 2W_t dW_t + dt.$$

Note that it is different from the traditional Newton-Leibnitz calculus which suggests $dW_t^2 = 2W_t dW_t$ with chain rule. This exactly manifests the specialty of Itô calculus to be introduced in this section. To further understand the previous specific example, we consider a more general situation.

Proposition 3.1. *For any bounded and continuous function $f(t,\omega)$ in t,*

$$\sum_j f(t_j^*,\omega)(W_{t_{j+1}} - W_{t_j})^2 \to \int_0^t f(s,\omega)ds, \quad \textit{for any} \;\; t_j^* \in [t_j, t_{j+1}]$$

in probability when the subdivision size goes to zero.

Proof. Straightforward calculation shows

$$\begin{aligned}\mathbb{E}\left(\sum_j f(t_j)\Delta W_{t_j}^2 - \sum_j f(t_j)\Delta t_j\right)^2 &= \mathbb{E}\left(\sum_{j,k} f(t_j)f(t_k)(\Delta W_{t_j}^2 - \Delta t_j)(\Delta W_{t_k}^2 - \Delta t_k)\right)\\ &= \mathbb{E}\left(\sum_j f^2(t_j)\cdot \mathbb{E}\left((\Delta W_{t_j}^2 - \Delta t_j)^2|\mathcal{F}_{t_j}\right)\right)\\ &= 2\sum_j \mathbb{E}f^2(t_j)\Delta t_j^2 \to 0.\end{aligned}$$

At the same time, we have

$$|\sum_j (f(t_j^*) - f(t_j))\Delta W_{t_j}^2| \leq \sup_j |f(t_j^*) - f(t_j)| \cdot \sum_j \Delta W_{t_j}^2.$$

The first term on the right hand side goes to zero almost surely because of the uniform continuity of f on $[0,t]$, and the second term converges to the quadratic variation of W_t in probability. Combining the results above leads to the desired conclusion.

It is exactly this reason that we simply denoted it as

$$dW_t^2 = dt$$

for calculations. The Itô's formula to be introduced below gives this a rigorous foundation.

Now let us consider the Itô process defined as

$$X_t = X_0 + \int_0^t b(s,\omega)ds + \int_0^t \sigma(s,\omega)dW_s,$$

which is usually denoted as

$$dX_t = b(t,\omega)dt + \sigma(t,\omega)dW_t, \quad X_t|_{t=0} = X_0 \tag{3.1}$$

for functions

$$\sigma \in \mathcal{W}[0,T], \quad b \text{ is } \mathcal{F}_t\text{-adapted and } \int_0^T |b(t,\omega)|dt < \infty \text{ a.s.}$$

We have the following important result, whose rigorous proof can be referred to [3,5].

Theorem 3.2 (1D Itô's formula)**.** *If X_t is an Itô process as in Equation (3.1), $Y_t = f(X_t)$ where f is a twice differentiable function. Then Y_t is also an Itô process and*

$$dY_t = f'(X_t)dX_t + \frac{1}{2}f''(X_t)(dX_t)^2,$$

where the rule of simplification is $dt^2 = 0$, $dtdW_t = dW_tdt = 0$ and $(dW_t)^2 = dt$, i.e.

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

Thus finally

$$dY_t = \left(b(t,\omega)f'(X_t) + \frac{1}{2}\sigma^2(t,\omega)f''(X_t)\right)dt + \sigma(t,\omega)f'(X_t)dW_t.$$

Sketch of Proof. We will only consider the situation that f, f' and f'' are bounded and continuous here. At first, if b and σ are simple functions, we have

$$Y_t - Y_0 = \sum_j (f(X_{t_{j+1}}) - f(X_{t_j})) = \sum_j \left(f'(X_{t_j})\Delta X_{t_j} + \frac{1}{2} f''(X_{t_j})\Delta X_{t_j}^2 + R_j \right),$$

where $\Delta X_{t_j} = X_{t_{j+1}} - X_{t_j}$ and $R_j = o(|\Delta X_{t_j}|^2)$. Without loss of generality we assume the discontinuity of the step functions are embedded in the current subdivision grid points. We obtain

$$\begin{aligned}\sum_j f'(X_{t_j})\Delta X_{t_j} &= \sum_j f'(X_{t_j}) b(t_j)\Delta t_j + \sum_j f'(X_{t_j})\sigma(t_j)\Delta W_{t_j} \\ &\to \int_0^t b(s) f'(X_s) ds + \int_0^t \sigma(s) f'(X_s) dW_s\end{aligned}$$

and

$$\sum_j f''(X_{t_j})\Delta X_{t_j}^2 = \sum_j f''(X_{t_j}) \left(b^2(t_j)\Delta t_j^2 + 2b(t_j)\sigma(t_j)\Delta t_j \Delta W_{t_j} + \sigma^2(t_j)\Delta W_{t_j}^2 \right).$$

We have

$$|\sum_j f''(X_{t_j}) b^2(t_j)\Delta t_j^2| \le K \sum_j \Delta t_j^2 \le KT \sup_j \Delta t_j \to 0,$$

$$|\sum_j f''(X_{t_j}) b(t_j)\sigma(t_j)\Delta t_j \Delta W_{t_j}| \le K \sum_j |\Delta t_j \Delta W_{t_j}| \le KT \sup_j |\Delta W_{t_j}| \to 0$$

as the subdivision size goes to zero, where K is the bound of b, σ and f''. From Proposition 3.1, we get

$$\sum_j f''(X_{t_j})\sigma^2(t_j)\Delta W_{t_j}^2 \to \int_0^t \sigma^2(s) f''(X_s) ds \qquad \text{in} \quad L_P^2.$$

The general situation can be done by taking approximation through simple functions.

The above result can be generalized to multidimensional case as

Theorem 3.3 (Multidimensional Ito formula)**.** *If* $d\boldsymbol{X}_t = \boldsymbol{b}(t,\omega)dt + \boldsymbol{\sigma}(t,\omega)\cdot d\boldsymbol{W}_t$, *where* $\boldsymbol{X}_t \in \mathbb{R}^n$, $\boldsymbol{\sigma} \in \mathbb{R}^{n\times m}$, $\boldsymbol{W} \in \mathbb{R}^m$. *Define* $Y_t = f(\boldsymbol{X}_t)$, *where* f *is a twice differentiable function. Then*

$$dY_t = \nabla f(\boldsymbol{X}_t)\cdot d\boldsymbol{X}_t + \frac{1}{2}(d\boldsymbol{X}_t)^T \cdot \nabla^2 f(\boldsymbol{X}_t)\cdot (d\boldsymbol{X}_t),$$

where the rule of simplification is $dt^2 = 0$, $dtdW_t^i = dW_t^i dt = dW_t^i dW_t^j = 0$ $(i \ne j)$, $(dW_t^i)^2 = dt$. *That is*

$$\begin{aligned}(d\boldsymbol{X}_t)^T \cdot \nabla^2 f(\boldsymbol{X}_t)\cdot (d\boldsymbol{X}_t) &= \sum_{l,k,i,j} dW_t^l \sigma_{il}\partial_{ij}^2 f \sigma_{jk} dW_t^k \\ &= \sum_{k,i,j} \sigma_{ik}\sigma_{jk}\partial_{ij}^2 f dt = \boldsymbol{\sigma}\boldsymbol{\sigma}^T : \nabla^2 f dt,\end{aligned}$$

where $\boldsymbol{A} : \boldsymbol{B} = \sum_{ij} a_{ij} b_{ji}$ *is the twice contraction for second order tensors. Finally*

$$dY_t = (\boldsymbol{b} \cdot \nabla f + \frac{1}{2}\boldsymbol{\sigma}\boldsymbol{\sigma}^T : \nabla^2 f)dt + \nabla f \cdot \boldsymbol{\sigma} \cdot d\boldsymbol{W}_t.$$

Example 3.4. *Integration by part*

$$\int_0^t s dW_s = tW_t - \int_0^t W_s ds. \tag{3.2}$$

Proof. Define $f(x, y) = xy$, $X_t = t$, $Y_t = W_t$, then from multidimensional Itô's formula

$$df(X_t, Y_t) = X_t dY_t + Y_t dX_t + dX_t dY_t.$$

With the rule $dt dW_t = 0$, we obtain $d(tW_t) = t dW_t + W_t dt$ and the result follows.

Example 3.5. *Iterated Itô integrals*

$$\int_0^t dW_{t_1} \int_0^{t_1} dW_{t_2} \dots \int_0^{t_{n-1}} dW_{t_n} = \frac{1}{n!} t^{\frac{n}{2}} h_n \left(\frac{W_t}{\sqrt{t}} \right), \tag{3.3}$$

where $h_n(x)$ *is the* n*-th order Hermite polynomial*

$$h_n(x) = (-1)^n e^{\frac{1}{2}x^2} \frac{d^n}{dx^n} \left(e^{-\frac{1}{2}x^2} \right).$$

Proof. It is easy to verify that

$$\int_0^t W_s dW_s = \frac{t}{2!} h_2 \left(\frac{W_t}{\sqrt{t}} \right),$$

where $h_2(x) = x^2 - 1$ is the second order Hermite polynomial. In the same fashion, we have

$$\int_0^t \Big(\int_0^s W_u dW_u \Big) dW_s = \frac{1}{2} \int_0^t (W_s^2 - s) dW_s.$$

Using Itô's formula, we have

$$\int_0^t W_s^2 dW_s = \frac{1}{3} W_t^3 - \int_0^t W_s ds.$$

Hence, using (3.2) we obtain

$$\int_0^t \Big(\int_0^s W_u dW_u) dW_s = \frac{1}{6} W_t^3 - \frac{1}{2} t W_t = \frac{1}{3!} t^{\frac{3}{2}} h_3 \left(\frac{W_t}{\sqrt{t}} \right),$$

where $h_3(x) = x^3 - 3x$ is the third order Hermite polynomial. The general case is left as an exercise.

4 SDE

4.1 Wellposed-ness

With the help of Itô's integral, we can establish the classical well-posedness result for the stochastic differential equations

$$d\boldsymbol{X}_t = \boldsymbol{b}(\boldsymbol{X}_t, t)dt + \boldsymbol{\sigma}(\boldsymbol{X}_t, t) \cdot d\boldsymbol{W}_t, \tag{4.1}$$

through Picard-type iterations.

Theorem 4.1. *Let $\boldsymbol{X} \in \mathbb{R}^n, \boldsymbol{W} \in \mathbb{R}^m$. Suppose the coefficients $\boldsymbol{b} \in \mathbb{R}^n, \boldsymbol{\sigma} \in \mathbb{R}^{n\times m}$ satisfy global Lipschitz and linear growth conditions as*

$$|\boldsymbol{b}(\boldsymbol{x}, t) - \boldsymbol{b}(\boldsymbol{y}, t)| + |\boldsymbol{\sigma}(\boldsymbol{x}, t) - \boldsymbol{\sigma}(\boldsymbol{y}, t)| \le K|\boldsymbol{x} - \boldsymbol{y}|, \tag{4.2}$$

$$|\boldsymbol{b}(\boldsymbol{x}, t)|^2 + |\boldsymbol{\sigma}(\boldsymbol{x}, t)|^2 \le K(1 + |\boldsymbol{x}|^2) \tag{4.3}$$

for any $\boldsymbol{x}, \boldsymbol{y} \in \mathbb{R}^n, t \in [0, T]$, where K is a positive constant and $|\cdot|$ means the Frobenius norm, that is

$$|\boldsymbol{b}|^2 := \sum_i b_i^2, \quad |\boldsymbol{\sigma}|^2 := \sum_{i,j} \sigma_{ij}^2.$$

Assume the initial value $\boldsymbol{X}_0 = \xi$ is a random variable which is independent of $\mathcal{F}_\infty^{\boldsymbol{W}}$ and satisfies $\mathbb{E}|\xi|^2 < \infty$. Then (4.1) has a unique t-continuous solution $\boldsymbol{X}_t \in \mathcal{V}[0, T]$.

The proof can be referred to [5].

4.2 Diffusion process

The SDEs driven by Wiener processes is the typical Markov process which is also called the *diffusion processes* in stochastic analysis. Mathematically, the diffusion process is defined for a Markov process $\{\boldsymbol{X}_t\}$ with continuous trajectory and its transition density $p(\boldsymbol{x}, t|\boldsymbol{y}, s)$ $(t \ge s)$ satisfies the following conditions for any $\epsilon > 0$:

$$\lim_{t\to s} \frac{1}{t-s} \int_{|\boldsymbol{x}-\boldsymbol{y}|<\epsilon} (\boldsymbol{x} - \boldsymbol{y}) p(\boldsymbol{x}, t|\boldsymbol{y}, s) d\boldsymbol{x} = \boldsymbol{b}(\boldsymbol{y}, s) + O(\epsilon), \tag{4.4}$$

$$\lim_{t\to s} \frac{1}{t-s} \int_{|\boldsymbol{x}-\boldsymbol{y}|<\epsilon} (\boldsymbol{x} - \boldsymbol{y})(\boldsymbol{x} - \boldsymbol{y})^T p(\boldsymbol{x}, t|\boldsymbol{y}, s) d\boldsymbol{x} = \boldsymbol{a}(\boldsymbol{y}, s) + O(\epsilon), \tag{4.5}$$

where $\boldsymbol{b}(\boldsymbol{y}, s)$ is called the drift of the considered diffusion process and $\boldsymbol{a}(\boldsymbol{y}, s)$ is called the diffusion matrix at time s and position $\boldsymbol{y}$. The conditions (4.4) and (4.5) can also be represented as

$$\lim_{t\to s} \frac{1}{t-s} \mathbb{E}^{\boldsymbol{y},s}(\boldsymbol{X}_t - \boldsymbol{y}) = \boldsymbol{b}(\boldsymbol{y}, s), \tag{4.6}$$

$$\lim_{t\to s} \frac{1}{t-s} \mathbb{E}^{\boldsymbol{y},s}(\boldsymbol{X}_t - \boldsymbol{y})(\boldsymbol{X}_t - \boldsymbol{y}) = \boldsymbol{a}(\boldsymbol{y}, s). \tag{4.7}$$

It is easy to find that the diffusion matrix $\boldsymbol{a} = \boldsymbol{\sigma}\boldsymbol{\sigma}^T$ in (4.1).

4.3 Simple SDEs

Example 4.2 (Ornstein-Uhlenbeck process)**.**

$$dX_t = -\gamma X_t dt + \sigma dW_t. \tag{4.8}$$

The Ornstein-Uhlenbeck process (OU process) has fundamental importance in statistical physics since it serves as the simplest model for many complex diffusion dynamics.

Solution. The equation above is equivalent to

$$dX_t + \gamma X_t dt = \sigma dW_t. \tag{4.9}$$

By applying Ito's formula to $e^{\gamma t} X_t$, we get

$$d(e^{\gamma t} X_t) = \gamma e^{\gamma t} X_t dt + e^{\gamma t} dX_t.$$

Integrating from 0 to t we have

$$e^{\gamma t} X_t - X_0 = \int_0^t (\gamma e^{\gamma s} X_s ds + e^{\gamma s} dX_s).$$

Timing $e^{\gamma t}$ to both sides of (4.9) and taking integration, we get

$$e^{\gamma t} X_t - X_0 = \int_0^t \sigma e^{\gamma s} dW_s.$$

Thus the solution

$$X_t = e^{-\gamma t} X_0 + \sigma \int_0^t e^{-\gamma(t-s)} dW_s.$$

If we define $Q_t := \int_0^t e^{-\gamma(t-s)} dW_s$, then it is not difficult to know that Q_t is a Gaussion process with

$$\mathbb{E}Q_t = 0, \qquad \mathbb{E}Q_t^2 = \int_0^t \mathbb{E}e^{-2\gamma(t-s)} ds = \frac{1}{2\gamma}(1 - e^{-2\gamma t}).$$

From this result we can observe that X_t is also a Gaussian process if X_0 is Gaussian, and the limit behavior of X_t is

$$X_t \xrightarrow{d} N\left(0, \frac{\sigma^2}{2\gamma}\right), \qquad (t \to +\infty).$$

This equation is called the SDE with additive noise since the coefficient of dW_t term is just a constant.

Example 4.3 (Geometric Brownian motion)**.**

$$dN_t = rN_t dt + \alpha N_t dW_t, \quad r, \alpha > 0. \tag{4.10}$$

This model has strong background in mathematical finance, in which N_t represents the asset price, r is the interest rate and α is called the volatility.

Solution. Divide N_t to both sides we have $dN_t/N_t = rdt + \alpha dW_t$. In deterministic calculus $1/N_t dN_t = d(\log N_t)$, so we apply Ito's formula to $\log N_t$, then

$$\begin{aligned} d(\log N_t) &= \frac{1}{N_t} dN_t - \frac{1}{2N_t^2}(dN_t)^2 \\ &= \frac{1}{N_t} dN_t - \frac{1}{2N_t^2}\alpha^2 N_t^2 dt \\ &= \frac{1}{N_t} dN_t - \frac{\alpha^2}{2} dt. \end{aligned}$$

Substitute the equation of dN_t we get

$$d(\log N_t) = (r - \frac{\alpha^2}{2})dt + \alpha dW_t.$$

Integrate from 0 to t to both sides

$$\begin{aligned} \log N_t - \log N_0 &= (r - \frac{\alpha^2}{2})t + \alpha W_t, \\ N_t &= N_0 \exp\left\{(r - \frac{\alpha^2}{2})t + \alpha W_t\right\}. \end{aligned}$$

This equation is called the SDE with multiplicative noise since the coefficient of dW_t term depends on N_t.

4.4 Brownian motion: revisited

Example 4.4 (Langevin equation)**.** *Mathematically a mesoscopic particle obeys the following well-known Langevin equation by Newton's Second Law*

$$\begin{cases} d\boldsymbol{X}_t &= \boldsymbol{V}_t dt, \\ m d\boldsymbol{V}_t &= \Big(-\gamma \boldsymbol{V}_t - \nabla V(\boldsymbol{X}_t) \Big) dt + \sqrt{2\sigma} d\boldsymbol{W}_t, \end{cases}$$

where γ is frictional coefficient, $V(\boldsymbol{X})$ is external potential, $\boldsymbol{W}_t$ is standard Wiener process, and σ is the strength of fluctuating force.

This example is used to show that the strength of fluctuating force must be related to the frictional coefficient in a physical setup. In principle the fluctuating force must be independent of external potential. In the case that the external force is zero, we have

$$m d\boldsymbol{V}_t = -\gamma \boldsymbol{V}_t dt + \sqrt{2\sigma} d\boldsymbol{W}_t.$$

This is exactly an Ornstein-Uhlenbeck process for $\boldsymbol{V}_t$. In the limit $t \to \infty$, we have

$$\langle \frac{1}{2} m \boldsymbol{V}^2 \rangle = \frac{3\sigma}{2\gamma}.$$

From equilibrium thermodynamics, the average kinetic energy must obey the rule

$$\langle \frac{1}{2} m \boldsymbol{V}^2 \rangle = \frac{3k_B T}{2}.$$

Thus we obtain the well-known *fluctuation-dissipation relation*:

$$\sigma = k_B T \gamma.$$

It can be proved that in this case the diffusion coefficient

$$D := \lim_{t \to \infty} \frac{\langle (\boldsymbol{X}_t - \boldsymbol{X}_0)^2 \rangle}{6t} = \frac{k_B T}{\gamma} \tag{4.11}$$

which is called *Einstein's relation.*

For more general forms of fluctuation-dissipation relation, the readers may be referred to [7].

Example 4.5 (Brownian dynamics)**.** *In the high γ case, the velocity $\boldsymbol{V}_t$ will always stay at an equilibrium Gaussian distribution, which means formally we can take $d\boldsymbol{V}_t = 0$. Then the Langevin equation is approximated by*

$$d\boldsymbol{X}_t = -\frac{1}{\gamma} \nabla V(\boldsymbol{X}_t) dt + \sqrt{\frac{2k_B T}{\gamma}} d\boldsymbol{W}_t,$$

which is called Brownian dynamics or Smoluchowski approximation. A mathematically rigorous derivation of Brownian dynamics from Langevin equations may be referred to [6] and the references therein.

5 Stratonovich integral

Another very important definition of the stochastic integral is the so-called Stratonovich (or Fisk-Stratonovich) integral which is defined as the limit of the following approximation

$$\int_0^T f(t,\omega) \circ dW_t \approx \sum_j \frac{f(t_j) + f(t_{j+1})}{2} (W_{t_{j+1}} - W_{t_j}).$$

Note that we use the special notation $\circ$ for stochastic integral to distinguish the Ito and Stratonovich integrals. As one can follow the similar way as in the definition for the Ito integral, we can also establish a consistent stochastic calculus based on the Stratonovich integral. It turns out that If X_t satisfies the SDE

$$dX_t = b(X_t, t)dt + \sigma(X_t, t) \circ dW_t \tag{5.1}$$

in the Stratonovich sense, then X_t satisfies the modified Ito SDE

$$dX_t = \Big(b(X_t, t) + \frac{1}{2}\partial_x \sigma \sigma(X_t, t)\Big)dt + \sigma(X_t, t)dW_t. \tag{5.2}$$

To understand this, we assume the solution X_t of the Stratonovich SDE satisfies

$$dX_t = \alpha(X_t, t)dt + \beta(X_t, t)dW_t. \tag{5.3}$$

Then by the definition of the Stratonovich integral

$$\int_0^t \sigma(X_s, s) \circ dW_s \approx \sum_j \frac{1}{2}(\sigma(X_{t_j}, t_j) + \sigma(X_{t_{j+1}}, t_{j+1}))(W_{t_{j+1}} - W_{t_j}).$$

From (5.3) we have

$$X_{t_{j+1}} = X_{t_j} + \alpha(X_{t_j}, t_j)\Delta t_j + \beta(X_{t_j}, t_j)\Delta W_{t_j} + h.o.t.,$$

and thus

$$\begin{aligned}\sum_j \sigma(X_{t_{j+1}}, t_{j+1})\Delta W_{t_j} &= \sum_j \Big(\sigma(X_{t_j}, t_j)\Delta W_{t_j} + \partial_t \sigma(X_{t_j}, t_j)\Delta t_j \Delta W_{t_j} \\ &\quad + \partial_x \sigma \alpha(X_{t_j}, t_j)\Delta t_j \Delta W_{t_j} + \partial_x \sigma \beta(X_{t_j}, t_j)\Delta W_{t_j}^2 + h.o.t.\Big) \\ &\to \int_0^t \sigma(X_s, s)dW_s + \int_0^t \partial_x \sigma \beta(X_s, s)ds\end{aligned}$$

from the fact $dW_t^2 = dt$. Summarizing the above results we obtain that X_t satisfies

$$dX_t = \Big(b(X_t, t) + \frac{1}{2}\partial_x \sigma \beta(X_t, t)\Big)dt + \sigma(X_t, t)dW_t. \tag{5.4}$$

To make (5.3) and (5.3) consistent, we take

$$\beta(x, t) = \sigma(x, t), \quad \alpha(x, t) = b(x, t) + \frac{1}{2}\partial_x \sigma \sigma(x, t).$$

In the high dimensions, one can derive similarly

$$d\boldsymbol{X}_t = \Big(\boldsymbol{b}(\boldsymbol{X}_t, t) + \frac{1}{2}\nabla_x \boldsymbol{\sigma} : \boldsymbol{\sigma}(\boldsymbol{X}_t, t)\Big)dt + \boldsymbol{\sigma}(\boldsymbol{X}_t, t) \cdot d\boldsymbol{W}_t \tag{5.5}$$

where $(\nabla_x \boldsymbol{\sigma} : \boldsymbol{\sigma})_i := \sum_{jk} \partial_k \sigma_{ij} \sigma_{kj}$ in the index notation if $\boldsymbol{X}$ satisfies

$$d\boldsymbol{X}_t = \boldsymbol{b}(\boldsymbol{X}_t, t)dt + \boldsymbol{\sigma}(\boldsymbol{X}_t, t) \circ d\boldsymbol{W}_t. \tag{5.6}$$

With this connection, we can check that the Stratonovich integral satisfies the Newton-Leibnitz chain rule

$$df(X_t) = f'(X_t) \circ dX_t = f'(X_t)b(X_t, t)dt + f'(X_t)\sigma(X_t, t) \circ dW_t$$

and its corresponding multi-dimensional form is

$$df(\boldsymbol{X}_t) = \nabla f(\boldsymbol{X}_t) \circ d\boldsymbol{X}_t = \nabla f(\boldsymbol{X}_t) \cdot \boldsymbol{b}(\boldsymbol{X}_t, t)dt + \nabla f(\boldsymbol{X}_t) \cdot \boldsymbol{\sigma}(\boldsymbol{X}_t, t) \circ d\boldsymbol{W}_t.$$

We finally remark here that the Ito isometry and mean zero property no longer hold for the Stratonovich integral, which can be easily observed from (5.2).

One reason that the Stratonovich interpretation is important is due to the following Wong-Zakai type theorem. The motivation is to intuitively understand the SDE (1.2) in the pathwise sense, i.e. for each fixed realization ω of W_t, we want to solve X_t by treating $W.(\omega)$ like a deterministic forcing term. But the issue is that the ordinary differential equation can not be solved in the classical case because of the rough property of the path of the Brownian motion. Since the C^1 functions on $[0, T]$ are dense in $C[0, T]$, so if we regularize the Brownian motion path from the following way

$$W^m \to W \text{ in } L^\infty[0, T] \text{ norm} \quad \text{as } m \to \infty,$$

where $W^m \in C^1[0, T]$, the differential equation

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

can be solved in the classical sense. We denote the solution as X_t^m. Then it can be proved that

$$X^m \to X \text{ in } L^\infty[0, T] \text{ norm}, \quad m \to \infty, \ a.s.$$

and the limit X_t is precisely the Stratonovich solution of the SDE (see [11, 12] for more details).

Now a rationale for why Stratonovich interpretation is useful in physics may be as follows. In realistic situations, the noise term $\dot{W}$ in (1.1) is usually not "white" but a smoothed colored noise since the idealistic white noise must be supplied with infinite energy from external environment. This smoothed colored noise exactly corresponds to some regularization of the white noise, which falls into the regime in the Wong-Zakai type smoothing argument.

Homeworks

1. Prove that with midpoint approximation

$$\int_0^t W_s dW_s \approx \sum_j W_{t_{j+\frac{1}{2}}}(W_{t_{j+1}} - W_{t_j}) \to \frac{W_t^2}{2}$$

and the rightmost approximation

$$\int_0^t W_s dW_s \approx \sum_j W_{t_{j+1}}(W_{t_{j+1}} - W_{t_j}) \to \frac{W_t^2}{2} + \frac{t}{2}$$

in $L_P^2(\Omega)$ as $|\Delta| \to 0$.

2. Prove the relation (3.3) through the following steps:

   (a) Prove that the Hermite polynomials satisfy

   $$\sum_{n=0}^{\infty} \frac{u^n}{n!} h_n(x) = \exp\left(ux - \frac{u^2}{2}\right)$$

   and

   $$\sum_{n=0}^{\infty} \frac{u^n}{n!} H_n(x,a) = \exp\left(ux - \frac{au^2}{2}\right),$$

   where $H_n(x,a) = a^{n/2} h_n(x/\sqrt{a})$ $(a > 0)$ and $H_n(x,0) = x^n$.

   (b) Prove that

   $$\left(\frac{1}{2}\frac{\partial^2}{\partial x^2} + \frac{\partial}{\partial a}\right) H_n(x,a) = 0 \quad \text{and} \quad \frac{\partial}{\partial x} H_n(x,a) = n H_{n-1}(x,a).$$

   (c) Prove the relation (3.3) through Itô's formula.

3. Solving the SDE

   (a) $dX_t = -X_t/(1+t)dt + 1/(1+t)dW_t$ with initial $X_0 = 0$.

   (b) $dX_t = -X_t dt + e^{-t} dW_t$ with initial X_0.

4. For the multidimensional OU process

   $$d\boldsymbol{X}_t = \boldsymbol{A}\boldsymbol{X}_t dt + \boldsymbol{\sigma} \cdot d\boldsymbol{W}_t,$$

   derive the relations that the stationary mean and covariance matrix should satisfy.

5. Prove that if one takes the right-most endpoint integral (backward stochastic integral) like

   $$\int_0^T f(t,\omega) * dW_t \approx \sum_j f(t_{j+1})(W_{t_{j+1}} - W_{t_j}).$$

   Then the SDE defined as

   $$dX_t = b(X_t,t)dt + \sigma(X_t,t) * dW_t \tag{5.7}$$

   can be related to the Ito SDE as

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

References

[1] I.M. Gelfand and G.E. Shilov. *Generalized functions*, volume Vol. 1-5. Academic Press, New York and London, 1964-1966.

[2] T. Hida and Si Si. *Lectures on White noise functionals.* World Scientific Publishing Co., Singapore, 2008.

[3] N. Ikeda and S. Watanabe. *Stochastic differential equations and diffusion processes.* North-Holland Publishing Company, Amsterdam, Oxford and New York, 1981.

[4] K. Ito. Stochastic integral. *Proc. Imp. Acad. Tokyo*, 20:519–524, 1944.

[5] I. Karatzas and S.E. Shreve. *Brownian motion and stochastic calculus.* Springer-Verlag, Berlin, Heidelberg and New York, 1991.

[6] P.R. Kramer and A.J. Majda. Stochastic mode reduction for particle-based simulation methods for complex microfluid systems. *SIAM J. Appl. Math.*, 64:401–422, 2003.

[7] R Kubo. The fluctuation-dissipation theorem. *Rep. Prog. Phys.*, 29:255, 1966.

[8] B. Oksendal. *Stochastic differential equations: An introduction with applications.* Springer-Verlag, New York, Berlin, Heidelberg and Tokyo, 4th edition edition, 1998.

[9] P.E. Protter. *Stochastic integral and differential equations.* Springer-Verlag, Berlin, Heidelberg and New York, Second edition edition, 2004.

[10] D. Revuz and M. Yor. *Continuous martingales and Brownian motion.* Springer-Verlag, Berlin and Heidelberg, 3rd edition edition, 2005.

[11] H.J. Sussman. On the gap between deterministic and stochastic ordinary differential equations. *Ann. Prob.*, 60:19–41, 1978.

[12] E. Wong and M. Zakai. On the convergence of ordinary integrals to stochastic integrasl. *Ann. Math. Stat.*, 36:1560–1564, 1965.