LECTURE 3

1. Approximation of Brownian motion

By the definition via increments, the Brownian motion at time t_{n+1} can be approximated by

$$\sum_{i=0}^{n} \Delta W_i = \sum_{i=0}^{n} \sqrt{\Delta \mathsf{t}_i}\xi_i, \text{ where } \Delta W_i = W(\mathsf{t}_{i+1}) - W(\mathsf{t}_i), \text{ and } \Delta \mathsf{t}_i = \mathsf{t}_{i+1} - \mathsf{t}_i, \tag{1.1}$$

where ξ_i's are i.i.d. standard Gaussian random variables.

One popular approximation of Brownian motion in continuous time is piecewise linear approximation (also known as polygonal approximation,

$$W^{(n)}(t) = W(\mathsf{t}_i) + (W(\mathsf{t}_{i+1}) - W(\mathsf{t}_i))\frac{t - \mathsf{t}_i}{\mathsf{t}_{i+1} - \mathsf{t}_i}, \quad t \in [\mathsf{t}_i, \mathsf{t}_{i+1}). \tag{1.2}$$

Another way to approximate Brownian motion is by a truncated *orthogonal expansion*:

$$W^{(\mathsf{n})}(t) = \sum_{i=1}^{\mathsf{n}} \xi_i \int_0^t m_i(s)\, ds, \quad \xi_j =: \int_0^T m_j(t)\, dW, \quad t \in [0, T], \tag{1.3}$$

where $\{m_i(t)\}$ is a CONS in $L^2([0,T])$ and ξ_j are mutually independent standard Gaussian random variables.

We can use the cosine basis $\{m_l(s)\}_{l\geq 1}$ given by

$$m_1(s) = \frac{1}{\sqrt{T}}, \quad m_l(s) = \sqrt{\frac{2}{T}}\cos(\frac{\pi(l-1)s}{T}), \quad l \geq 2, \quad 0 \leq s \leq T, \tag{1.4}$$

or a piecewise version of spectral expansion (1.3) by taking a partition of $[0,T]$, e.g. $0 = \mathsf{t}_0 < \mathsf{t}_1 < \cdots \mathsf{t}_{\mathsf{K}-1} < \mathsf{t}_{\mathsf{K}} = T$. We then have

$$W^{(\mathsf{n},\mathsf{K})}(t) = \sum_{k=1}^{\mathsf{K}} \sum_{l=1}^{\mathsf{n}} M_{k,l}(t)\xi_{k,l}, \quad \xi_{k,l} = \int_{I_k} m_{k,l}(s)\, dW(s), \tag{1.5}$$

where $M_{k,l}(t) = \int_{\mathsf{t}_{k-1}}^{\mathsf{t}_k \wedge t} m_{k,l}(s)\, ds$ ($\mathsf{t}_k \wedge t = \min(\mathsf{t}_k, t)$), $\{m_{k,l}\}_{l=1}^{\infty}$ is a CONS in $L^2(I_k)$ and $I_k = [\mathsf{t}_{k-1}, \mathsf{t}_k)$. The random variables $\xi_{k,l}$ are i.i.d. standard Gaussian random variables. Sometimes (1.5) is written as

$$W^{(\mathsf{n},\mathsf{K})}(t) = \sum_{k=1}^{\mathsf{K}} \sum_{l=1}^{\mathsf{n}} \int_0^t \mathbf{1}_{I_k}(s) m_{k,l}(s)\, ds \xi_{k,l}, \tag{1.6}$$

where $\mathbf{1}.$ is the indicator function. On I_k, $W^{(\mathsf{n},\mathsf{K})}(t) = W^{(\mathsf{n},\mathsf{K})}(t_{k-1}) + \sum_{l=1}^{\mathsf{n}} M_{k,l}(t)\xi_{k,l}$ if $m_{k,1}$ is a constant for all k.

Here different choices of CONS lead to different representations. The orthonormal piecewise constant basis over time interval $I_k = [\mathbf{t}_{k-1}, \mathbf{t}_k)$, with $h_k = (\mathbf{t}_k - \mathbf{t}_{k-1})/\sqrt{\mathsf{n}}$, is

$$m_{k,l}(t) = \frac{1}{\sqrt{h}}\chi_{[\mathbf{t}_{k-1}+(l-1)h_k, \mathbf{t}_{k-1}+lh_k)}, \quad l = 1, 2, \cdots, \mathsf{n}. \tag{1.7}$$

Date: October 4, 2019.

When $\mathsf{n} = 1$, this basis gives the classical piecewise linear interpolation (1.2). The orthonormal Fourier basis gives Wiener's representation

$$m_{k,1} = \frac{1}{\sqrt{\mathbf{t}_k - \mathbf{t}_{k-1}}}, \quad m_{k,2l} = \sqrt{\frac{2}{\mathbf{t}_k - \mathbf{t}_{k-1}}} \sin\big(2l\pi \frac{t - \mathbf{t}_{k-1}}{\mathbf{t}_k - \mathbf{t}_{k-1}}\big),$$

$$m_{k,2l+1}(t) = \sqrt{\frac{2}{\mathbf{t}_k - \mathbf{t}_{k-1}}} \cos\big(2l\pi \frac{t - \mathbf{t}_{k-1}}{\mathbf{t}_k - \mathbf{t}_{k-1}}\big), \; t \in [\mathbf{t}_{k-1}, \mathbf{t}_k). \tag{1.8}$$

Note that taking only $m_{k,1}$ ($\mathsf{n} = 1$) in (1.8) again leads to the piecewise linear interpolation (1.2). Besides, we can also use the Haar wavelet basis, which gives the Levy-Ciesielsky representation [Karatzas and Shreve, 1991].

Remark 1.1. *Once we have a formal representation (approximation) of Brownian motion, we then can readily obtain a formal representation (approximation) of white noise, and thus we skip the formulas for white noise.*

Lemma 1.2. *Consider a uniform partition of $[0, T]$, i.e., $\mathsf{t}_k = k\Delta$, $\mathsf{K}\Delta = T$. For $t \in [0, T]$, there exists a constant $C > 0$ such that*

$$\mathbb{E}[\big(W(t) - W^{(\mathsf{n},\mathsf{K})}(t)\big)^2] \le C\frac{\Delta}{n},$$

and for sufficient small $\epsilon > 0$

$$\left|W(t) - W^{(\mathsf{n},\mathsf{K})}(t)\right| \le \mathcal{O}\left(\left(\frac{\Delta}{n}\right)^{1/2-\epsilon}\right)^{1}. \tag{1.9}$$

For $t = \mathsf{t}_k$, we have

$$W(\mathsf{t}_k) - W^{(\mathsf{n},\mathsf{K})}(\mathsf{t}_k) = 0, \tag{1.10}$$

if the CONS $\{m_{k,l}\}_{l=1}^{\infty}$ contains $\frac{1}{\sqrt{\mathsf{t}_k - \mathsf{t}_{k-1}}}$ as its elements, i.e., $\int_{\mathsf{t}_{k-1}}^{\mathsf{t}_k} m_{k,l}(s) \frac{1}{\sqrt{\mathsf{t}_k - \mathsf{t}_{k-1}}}\, ds = \delta_{l,1}$.

Proof. By the spectral approximation of $W(t)$ (1.5) and the fact that $\xi_{k,l,i}$ are i.i.d, we have

$$\begin{aligned}
\mathbb{E}[(W(t) - W^{(\mathsf{n},\mathsf{K})}(t))^2] &= \sum_{k=1}^{\mathsf{K}} \sum_{l=\mathsf{n}+1}^{\infty} \left(\int_{\mathsf{t}_{k-1}}^{t\wedge \mathsf{t}_k} m_{k,l}(s)\, ds\right)^2 \\
&= \sum_{l=\mathsf{n}+1}^{\infty} \left(\int_{\mathsf{t}_{k-1}}^{\mathsf{t}_k} \chi_{[0,t]}(s) m_{k,l}(s)\, ds\right)^2, \quad (t \in I_k) \\
&\le C\frac{\Delta}{n},
\end{aligned}$$

where $\mathsf{t}_k \wedge t = \min(\mathsf{t}_k, t)$ and we have applied the standard estimate for L^2-projection using piecewise orthonormal basis $m_{k,l}(s)$, and have used the fact that the indicator function $\chi_{[0,t]}(s)$ belongs to the Sobolev space $H^{1/2}((0,T))$ for any $T > t$. The $H^{\theta}((0,T))$ $(0 < \theta < 1)$ is endowed with the norm $\|v\| + \big(\int_0^T \int_0^T \frac{|v(t)-v(s)|^2}{|t-s|^{2\theta+1}}\, ds\, dt\big)^{1/2}$.

Once we have the L^2-estimate, we can apply the Borel-Cantelli lemma to obtain the almost sure (a.s.) convergence (1.9).

[1]The big "$\mathcal{O}$" implies that the error is bounded by a positive constant times the term in the parenthesis.

If $t = \mathsf{t_k}$, we have $\int_{\mathsf{t}_{k-1}}^{\mathsf{t}_k} m_{k,l}(s)\,ds = 0$ for any $l \geq 2$ if $m_{k,1} = \frac{1}{\sqrt{\mathsf{t}_k - \mathsf{t}_{k-1}}}$, and thus (1.10) holds. □

Though any CONS in $L^2([0,T])$ can be used in the spectral approximation (1.3), we use a CONS containing a constant in the basis. Consequently, we have the following relation

$$\int_{t_n}^{t_{n+1}} dW^{(\mathsf{n})}(t) = \Delta W_n, \quad \Delta W_n = W(t_{n+1}) - W(t_n). \tag{1.11}$$

2. Stochastic Integration

As Brownian motion $W(t)$ is not a process of bounded variation, the integral $\int_0^t f(t)\,dW(t)$ cannot be interpreted using Riemann-Stieltjes integration or Lebesgue-Stieltjes integration, even for very smooth stochastic process f. However, it can be understood as Ito integral or Stratonovich integral. For an adapted process $f(t)$ (with respect to the natural filtration of Brownian motion), the Ito integral is defined as, see e.g. [Øksendal, 2003], for all partitions of the interval $[0,T]$,

$$\lim_{|\Pi_n|\to 0} \mathbb{E}[\Big(\int_0^T f(t)\cdot\, dW - \sum_{i=1}^n f(t_{i-1})\Delta W_i\Big)^2] = 0,$$

where $\Pi_n = \{0 = t_0 < t_1 < t_2 < \cdots < t_n = T\}$ is a partition of the interval $[0,T]$ and $|\Pi_n| = \max_{0\leq i\leq n-1} |t_{i+1} - t_i|$. The finite sum in this definition is defined at the *left-hand points* in each subinterval of the partition. For Stratonovich calculus, the finite sum is defined at the *midpoints* in each subinterval of the partition. i.e.,

$$\int_0^T f(t)\circ\, dW = \lim_{|\Pi_n|\to 0} \sum_{i=1}^n f(\frac{t_{i-1}+t_i}{2})\Delta W_i.$$

Again, the limit is understood in the mean-square sense, see e.g. [Øksendal, 2003].

Example 2.1. *It can be readily checked that*

$$\int_0^T W(t)\,dW(t) = \frac{W^2(T) - T}{2}, \quad \int_0^T W(t)\circ dW(t) = \frac{W^2(T)}{2}. \tag{2.1}$$

Let us show that the first formula holds. By simple calculation and the properties of increments of Brownian motion, for $0 = t_0 < t_1 < t_2 < \cdots < t_n = T$, we have

$$\begin{aligned}
&\mathbb{E}[\Big(\frac{W^2(T)-T}{2} - \sum_{i=0}^{n-1} W(t_i)\Delta W_i\Big)^2] \\
=\ &\mathbb{E}[\Big(\frac{W^2(T)-T}{2} - \sum_{i=0}^{n-1} \frac{W(t_i)+W(t_{i+1})}{2}\Delta W_i + \frac{1}{2}\sum_{i=1}^n (\Delta W_i)^2\Big)^2] \\
=\ &\mathbb{E}[\Big(\frac{W^2(T)-T}{2} - \sum_{i=0}^{n-1} \frac{W^2(t_{i+1})-W^2(t_i)}{2} + \frac{1}{2}\sum_{i=1}^n (\Delta W_i)^2\Big)^2] \\
=\ &\mathbb{E}[\Big(\frac{-T}{2} + \frac{1}{2}\sum_{i=1}^n (\Delta W_i)^2\Big)^2] \to 0, \quad n\to\infty.
\end{aligned}$$

The second formula can be derived similarly.

Exercise 2.2. *Show that Brownian motion on $[0,T]$ has a bounded quadratic variation which is equal to T.*

Theorem 2.3 (Conversion of a Stratonovich integral to an Ito integral)**.** *A Stratonovich integral can be computed via the Ito integral:*

$$\int_0^T f(t, W(t)) \circ dW(t) = \int_0^T f(t, W(t))\, dW(t) + \frac{1}{2}\int_0^T \partial_x f(t, W(t))\, dt.$$

Here $f(t, W(t))$ is a scalar function and $\partial_x f$ is the derivative with respect to the second argument of f. When $f \in \mathbb{R}^{m\times n}$ is a matrix function, then

$$[\int_0^T f(t, W(t)) \circ dW(t)]_i = [\int_0^T f(t, W(t))\, dW(t)]_i + \frac{1}{2}\int_0^T \sum_{j=1}^n \partial_{x_j} f_{i,j}(t, W(t))\, dt, \quad i = 1, 2, \cdots, m.$$

Here v_i means the i-th component of a vector v.

The proof can be done using the definition of two integrals and mean value theorem. We leave the proof to interested readers.

Remark 2.4. *When we want to define the integral $\int_0^t f(t)\, dW(t)$ via a spectral representation of Brownian motion instead of using increments of Brownian motion, we have to use the so-called Ito-Wick product (Wick product) and Stratonovich product.*

Definition 2.5 (Quadratic covariation)**.** *The quadratic covariation of two processes X and Y is*

$$[X, Y]_t = \lim_{|\Pi_n|\to 0} \sum_{k=1}^n \left(X(t_k) - X(t_{k-1})\right)\left(Y(t_k) - Y(t_{k-1})\right).$$

Here $\Pi_n = \{0 = t_0 < t_1 < \cdots < t_{n-1} < t_n = t\}$ is an arbitrary partition of the interval $[0, t]$.

When $X = Y$, the quadratic covariation becomes the quadratic variation:

$$[X]_t = [X, X]_t = \lim_{|\Pi_n|\to 0} \sum_{k=1}^n (X(t_k) - X(t_{k-1}))^2.$$

The quadratic covariation can be computed by the polarization identity:

$$[X, Y]_t = \tfrac{1}{4}([X + Y]_t - [X - Y]_t).$$

With the definition of quadratic covariation, we have

$$\int_0^T f(t, W(t) \circ dW(t) = \int_0^T f(t, W(t))\, dW(t) + \frac{1}{2}\int_0^T \partial_x f(t, W(t))\, d[W, W]_t.$$

More generally, we have the following conversion rule

$$\int_0^t Y(s) \circ dX(s) = \int_0^t Y(s)\, dX(s) + \frac{1}{2}[X, Y]_t. \tag{2.2}$$

For Ito integral, we have the following properties. Define

$$\mathbb{L}^2_{\text{ad}}(\Omega; L^2([a, b])) = \left\{ f_t(\omega) | f_t(\omega) \text{ is } \mathcal{F}_t\text{-measurable and } \mathbb{E}[\int_a^b f_s^2\, ds] < \infty \right\}.$$

Theorem 2.6. *For $f, g \in \mathbb{L}^2_{ad}(\Omega; L^2([0, T]))$, we have*

- *(linear property)* $\int_0^t (af(s) + bg(s))\, dW(s) = a\int_0^t f(s)\, dW(s) + b\int_0^t g(s)\, dW(s), \quad a, b \in \mathbb{R}$,

- *(Ito isometry)* $\mathbb{E}[\big(\int_0^t f(s)\,dW(s)\big)^2] = \int_0^t \mathbb{E}[f^2(s)]\,ds$,
- *(Generalized Ito isometry)* $\mathbb{E}[\int_0^t f(s)\,dW(s)\int_0^t g(s)\,dW(s)] = \int_0^t \mathbb{E}[f(s)g(s)]\,dt$,
- $M_t = \int_0^t f(s)\,dW(s)$ *is a continuous martingale. Moreover, the quadratic variation of* M_t *is* $|M|_{2,\mathrm{TV}} = \int_0^t f^2(s)\,ds$, $0 \le t \le T$ *and*

$$\mathbb{E}[\sup_{0\le t\le T}\big(\int_0^t f(s)\,dW(s)\big)^2] \le 4\int_0^t \mathbb{E}[f^2(s)]\,ds.$$

Example 2.7 (Multiple Ito integral)**.** *Assume that* $W(t)$ *is a standard Brownian Motion, show that*

$$\frac{1}{n!}\int_0^t\int_0^{t_n}\cdots\int_0^{t_2} dW(t_1)\cdots dW(t_n) = t^{n/2}H_n(\frac{W(t)}{\sqrt{t}}).$$

Here H_n *is the n-th Hermite polynomial:*

$$H_n(x) = (-1)^n e^{x^2/2}\frac{d^n}{dx^n}e^{-x^2/2}. \tag{2.3}$$

When $n = 0$, $H_n(x) = 1$ and we use the convention that when $n < 1$ the integral is defined as 1. When $n = 1$, $\int_0^t dW(t_1) = W(t) = t^{1/2}H_1(\frac{W(t)}{\sqrt{t}})$ as $H_1(x) = x$. Then it can be shown by induction that the integrand in the left hand side is in $\mathbb{L}^2_{\mathrm{ad}}(\Omega; L^2([0,t]))$, and thus the multiple integral is indeed an Ito integral and is equal to the right hand side.

2.1. **Applications of stochastic integrals.** *Application I.* The **fractional Brownian motion** $B_H(t)$, $t \ge 0$, can be introduced. It is a centered Gaussian process with the following covariance function

$$\mathbb{E}[B_H(t)B_H(s)] = \tfrac{1}{2}(|t|^{2H} + |s|^{2H} - |t-s|^{2H}), \quad 0 < H < 1.$$

The constant H is called the Hurst index or Hurst parameter. The fractional Brownian motion can be represented by

$$B_H(t) = B_H(0) + \frac{1}{\Gamma(H+1/2)}\left\{\int_{-\infty}^0 \left[(t-s)^{H-1/2} - (-s)^{H-1/2}\right] dW(s) + \int_0^t (t-s)^{H-1/2}\,dW(s)\right\}.$$

Application II. **Simulation of Gaussian processes with given stationary kernels**. Recall the power spectrum

$$f(\omega) = \frac{1}{2\pi}\int_{\mathbb{R}} e^{-\mathrm{i}\,\omega t}K(t)\,dt. \tag{2.4}$$

When $K(t) = K^\top(t)$ (e.g., when the kernel is isotropic), the spectral density becomes

$$f(\omega) = \frac{1}{2\pi}\int_0^\infty \cos(\omega t)[K(t) + K^\top(t)]\,dt. \tag{2.5}$$

In this case, $\mathrm{Im}\, f = 0$, we can set

$$\sqrt{2}z(\,d\omega) = (\mathrm{Re}\, f)^{1/2}\,dW(\omega) + i(\mathrm{Re}\, f)^{1/2}\,dB(\omega),$$

where W and B are independent standard Brownian motions and

$$Y_t = \sqrt{2}\int_{\mathbb{R}^+}\cos(\omega t)\,(\mathrm{Re}\, f)^{1/2}\,dW(d\omega) - \sqrt{2}\int_{\mathbb{R}^+}\sin(\omega t)\,(\mathrm{Re}\, f)^{1/2}\,dB(d\omega). \tag{2.6}$$

Then we can obtain a discretization using the trapezoidal rule

$$\tilde{Y}_t = \sqrt{2}\sum_{j=0}^{N-1} (\mathrm{Re}\, f)^{1/2}(\omega_{j+\frac{1}{2}})(\cos(\omega_{j+\frac{1}{2}}t)\,\Delta W_j + \sin(\omega_{j+\frac{1}{2}}t)\,\Delta B_j),$$

where $\Delta W_j = W(\omega_{j-1}) - W(\omega_j)$ and $\Delta B_j = B(\omega_{j-1}) - B(\omega_j)$ and ω_j's form a partition of $[0, L]$ with a large enough L while $\omega_{j+\frac{1}{2}} = \frac{\omega_j+\omega_{j+1}}{2}$. Replacing the increments of Brownian motions ΔB_j and ΔW_j with Gaussian random variables, we obtain

$$\boxed{\tilde{Y}_t = \sqrt{2}\sum_{j=0}^{N-1} (\mathrm{Re}\, f)^{1/2}(\omega_{j+\frac{1}{2}})(\cos(\omega_{j+\frac{1}{2}}t)\,\xi_j + \sin(\omega_{j+\frac{1}{2}}t)\,\eta_j)\sqrt{\Delta t_j}.} \tag{2.7}$$

Here ξ_j's and η_j's are independent standard normal random vectors.

Remark 2.8. *Here $(\cdot)^{1/2}$ refers to the Cholesky decomposition of $(\cdot)$ instead of the square root of a matrix.*

Remark 2.9. *When the Gaussian process is smooth, one can also apply the spectral element approximation of Brownian motions.*

Example 2.10. *Suppose that $X(t) \sim \mathcal{N}(0, \exp(-\frac{|x-y|^2}{l^2}))$. Then*

$$X(t) = \sqrt{2}(\pi)^{\frac{1}{4}}\int_{\mathbb{R}^+} (l)^{\frac{1}{2}}\cos(\omega t)\,\exp(-\frac{l^2\omega^2}{8})\,dW(\omega) - \sqrt{2}(\pi)^{\frac{1}{4}}\int_{\mathbb{R}^+}\sin(\omega t)\,(l)^{\frac{1}{2}}\exp(-\frac{l^2\omega^2}{8})\,dB(\omega). \tag{2.8}$$

Sometimes it is useful to apply the change of variable $\lambda = l\omega$ and write $X(t)$ as

$$X(t) = \sqrt{2}(\pi)^{\frac{1}{4}}\int_{\mathbb{R}^+}\cos(\frac{\lambda}{l}t)\,\exp(-\frac{\lambda^2}{8})\,dW(\lambda) - \sqrt{2}(\pi)^{\frac{1}{4}}\int_{\mathbb{R}^+}\sin(\frac{\lambda}{l}t)\,\exp(-\frac{\lambda^2}{8})\,dB(\lambda). \tag{2.9}$$

Application III. **Estimate the error of linear interpolation of Gaussian process.** Applying a linear interpolation Π_1 on the interval $[t_i, t_{i+1}]$, then

$$\begin{aligned}
\mathbb{E}[(X(t) - \Pi_1 X(t))^2] &= 2(\pi)^{\frac{1}{2}}\int_{\mathbb{R}^+}((I-\Pi_1)\cos(\frac{\lambda}{l}t))^2\,\exp(-\frac{\lambda^2}{4})\,d\lambda \\
&\quad + 2(\pi)^{\frac{1}{2}}\int_{\mathbb{R}^+}((I-\Pi_1)\sin(\frac{\lambda}{l}t))^2\,\exp(-\frac{\lambda^2}{4})\,d\lambda \\
&\le (\pi)^{\frac{1}{2}}(t_{i+1}-t_i)^4\int_{\mathbb{R}^+}(\frac{\lambda}{l})^4\,\exp(-\frac{\lambda^2}{4})\,d\lambda \\
&= 24\pi\frac{(t_{i+1}-t_i)^4}{l^4}.
\end{aligned}$$

where we realled the error estimate of the linear interpolation on $[a, b]$ (by Taylor's expansion)

$$|(I-\Pi_1)g(t)| = \frac{1}{2}(b-a)^2\left|g''(\xi)\right|, \tag{2.10}$$

where ξ lies in between a and b. Then by the Borel-Cantelli lemma, we have

$$|X(t) - \Pi_1 X(t)| \le C\frac{(t_{i+1}-t_i)^{2-\epsilon}}{l^2}, \quad \epsilon > 0, \tag{2.11}$$

where C is an absolute value of a Gaussian random variable with a finite variance.

3. Stochastic chain rules

One motivation of the Ito formula is to evaluate Ito integral with a complicated integrand.

Theorem 3.1 (Ito formula in the simplest form). *If f and its first two derivatives are continuous on $\mathbb{R}$, then it holds with probability one (almost surely, a.s.) that*

$$f(W(t)) = f(W(t_0)) + \int_{t_0}^{t} f'(W(s))\,dW(s) + \frac{1}{2}\int_{t_0}^{t} f''(W(s))\,ds.$$

From the theorem, we can compute the Ito integral $\int_{t_0}^{t} f'(W(s))\,dW(s)$ by

$$\int_{t_0}^{t} f'(W(s))\,dW(s) = f(W(t)) - f(W(t_0)) - \frac{1}{2}\int_{t_0}^{t} f''(W(s))\,ds.$$

Definition 3.2 (Ito process). *An Ito process is a stochastic process of the form*

$$X_t = X(t_0) + \int_{t_0}^{t} a(s)\,ds + \int_{t_0}^{t} \sigma(s)\,dW(s),$$

where $X(t_0)$ is $\mathcal{F}_{t_0}$-measurable, a_s and σ_s are adapted w.r.t. $\mathcal{F}_s$ and

$$\int_{t_0}^{t} |a(s)|\,ds < \infty, \quad \int_{t_0}^{t} \|\sigma(s)\|^2\,ds < \infty \ a.s..$$

The filtration $\{\mathcal{F}_s,\, t_0 \le s \le t\}$ is defined such that

- for any s, B_s, $a(s)$ and $\sigma(s)$ are $\mathcal{F}_s$-measurable;
- for any $t_1 \le t_2$, $B_{t_2} - B_{t_1}$ is independent of $\mathcal{F}_{t_1}$.

Suppose that $X(t)$ exists a.s. such that

$$X(t) = X(t_0) + \int_{t_0}^{t} a(s, X(s))\,ds + \sum_{r=1}^{m}\int_{t_0}^{t} \sigma_r(s, X(s))\,dW_r(s).$$

Here $X(t)$, $X(t_0), a$, $\sigma_r \in \mathbb{R}^d$ and $\sigma \in \mathbb{R}^{d\times m}$. Also, $W_r(s)$'s are mutually independent Brownian motions, $a(s)$ and $\sigma(s)$ are adapted w.r.t. $\mathcal{F}_s$, and

$$\int_{t_0}^{t} |a(s, X(s))|\,ds < \infty, \quad \sum_{r=1}^{m}\int_{t_0}^{t} |\sigma_r(s, X(s))|^2\,ds < \infty \text{ a.s..}$$

The filtration $\{\mathcal{F}_s,\, t_0 \le s \le t\}$ is defined such that for any s, $W_r(s)$ is $\mathcal{F}_s$-measurable and for any $t_1 \le t_2$, $W_r(t_2) - W_r(t_1)$ is independent of $\mathcal{F}_{t_1}$. Ito formula for a $C^1([0,T]; C^2(\mathbb{R}^d))$ function $f(t,\cdot)$ is

$$f(t, X(t)) = f(t_0, X(t_0)) + \sum_{r=1}^{m}\int_{t_0}^{t} \Lambda_r f(s, X(s))\,dW_r(s) + \int_{t_0}^{t} \mathcal{L} f(s, X(s))\,ds,$$

where

$$\begin{aligned}
\Lambda_r &= \sigma_r^\top \nabla = \sum_{i=1}^{d} \sigma_{i,r}\frac{\partial}{\partial x_i}, \quad \nabla = (\frac{\partial}{\partial x_1}, \frac{\partial}{\partial x_2}, \cdots, \frac{\partial}{\partial x_d}),\\
\mathcal{L} &= \frac{\partial}{\partial t} + a^\top \nabla + \frac{1}{2}\sum_{r=1}^{m}\sum_{i,j=1}^{d} \sigma_{i,r}\sigma_{j,r}\frac{\partial^2}{\partial x_i \partial x_j}.
\end{aligned}$$

Remark 3.3. *For the multi-dimensional Ito formula, we can use the following table to memorize the formula when $W_j(t)$ are mutually independent Brownian motions.*

$\times$	$dW_j(t)$	dt
$dW_i(t)$	$\mathbf{1}_{\{i=j\}}dt$	0
dt	0	0

Theorem 3.4 (Integration by parts formula)**.** *Let $X(t)$ and Y_t be Ito processes defined in Definition 3.2. Then the following integration by parts formula holds*

$$X(t)Y(t) = X(t_0)Y(t_0) + \int_{t_0}^{t} X(s)\,dY(s) + \int_{t_0}^{t} Y(s)\,dX(s) + \int_{t_0}^{t} dX(s)\,dY(s).$$

Here $dX(s)\,dY(s)$ can be formally computed using the table in Remark 3.3.

This integration by parts formula is a corollary of multi-dimensional Ito formula for Ito processes.

Consider two Ito processes, X and Y: $dX = a^X(t)\,dt + \sigma^X(t)\,dW(t)$ and $dY = a^Y(t)\,dt + \sigma^Y(t)\,dW(t)$. Then we have from Remark 3.3 that

$$dXdY = (a^X(s)\,dt + \sigma^X(t)\,dW(t)(a^Y(s)\,dt + \sigma^Y(t)\,dW(t))) = \sigma^X(t)\sigma^Y(t)\,dt.$$

Appendix A. Naive stochastic integration is impossible

Let (Π_m) be a refining sequence of partition of $[0,T]$. Let $\phi : [0,T] \to R$ be a right-continuous function. For $n \in \mathbb{N}$, define an operator $S_\phi^{(m)} : C([0,T] \to R)$ by

$$S_\phi^{(m)}(\mu) = \sum_{t_k, t_{k-1} \in \Pi_m} \mu(t_{k-1})(\phi(t_k) - \phi(t_{k-1}).$$

If $S_\phi^{(m)}(\mu)$ converges to a finite limit as $m \to \infty$, for all $\mu \in C([0,T])$, then ϕ must be of bounded variation.

Proof. Pick any continuous function h which take values $\operatorname{sign}(\phi(t_k) - \phi(t_{k-1})$ at t_{k-1} and $\|h\| = 1$. By the uniform convergence, we see that

$$S_\phi^{(m)}(h) = \sum_{t_k, t_{k-1} \in \Pi_m} |\phi(t_k) - \phi(t_{k-1})| < \infty.$$

By the assumption and Uniform Boundedness Principle (Banach-Steinhaus theorem), $S_\phi^{(m)}$ is a bounded operator. Note that

$$\sum_{t_k, t_{k-1} \in \Pi_m} |\phi(t_k) - \phi(t_{k-1})| = S_\phi^{(m)}(h) \le \left\| S_\phi^{(m)} \right\| \|h\| < \infty.$$ □

Appendix B. Regular stochastic process

A linear regular stationary process $X(t)$ admits a regular representation

$$X(t) = \int_{-\infty}^{t} C(t-s)\,dW(s)$$

A necessary and sufficient condition for regularity of a (one-dimensional) stationary process is the existence of a spectral density such that

$$\int_{\mathbb{R}} \frac{\ln f(\lambda)}{1+\lambda^2}\,d\lambda > -\infty. \tag{B.1}$$

Appendix C. Bramble-Hilbert Lemma

Lemma C.1 (Bramble-Hilbert Lemma)**.** *For a function u that has m derivatives on interval (a, b).*

$$\inf_{v \in P_{m-1}} \left\| u^{(k)} - v^{(k)} \right\|_{L^p(a,b)} \leq C\left(m\right)\left(b-a\right)^{m-k} \left\| u^{(m)} \right\|_{L^p(a,b)},$$

where P_{m-1} is the space of all polynomials of order at most $m-1$.

References

[Karatzas and Shreve, 1991] Karatzas, I. and Shreve, S. E. (1991). *Brownian motion and stochastic calculus*, volume 113 of *Graduate Texts in Mathematics*. Springer-Verlag, New York, second edition.

[Øksendal, 2003] Øksendal, B. (2003). *Stochastic differential equations*. Universitext. Springer-Verlag, Berlin, sixth edition. An introduction with applications.