LECTURE 3

1. Approximation of Brownian motion

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

$$\sum_{i=0}^{n} \Delta W_i = \sum_{i=0}^{n} \sqrt{\Delta t_i} \xi_i, \text{ where } \Delta W_i = W(t_{i+1}) - W(t_i), \text{ and } \Delta t_i = t_{i+1} - t_i,$$

where $\xi_i$'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,

$$W^{(n)}(t) = W(t_i) + (W(t_{i+1}) - W(t_i)) \frac{t - t_i}{t_{i+1} - t_i}, \quad t \in [t_i, t_{i+1}).$$

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

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

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

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

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

or a piecewise version of spectral expansion (1.3) by taking a partition of $[0, T]$, e.g. $0 = t_0 < t_1 < \cdots < t_{K-1} < t_K = T$. We then have

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

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

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

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

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

$$m_{k,l}(t) = \frac{1}{\sqrt{h_k}} \chi_{[t_{k-1}+(l-1)h_k, t_{k-1}+lh_k]}, \quad l = 1, 2, \cdots, n.$$
When $n = 1$, this basis gives the classical piecewise linear interpolation (1.2). The orthonormal Fourier basis gives Wiener’s representation

$$m_{k,1} = \frac{1}{\sqrt{t_k - t_{k-1}}}, \quad m_{k,2l} = \sqrt{\frac{2}{t_k - t_{k-1}}} \sin \left(2l\pi \frac{t - t_{k-1}}{t_k - t_{k-1}}\right),$$

$$m_{k,2l+1}(t) = \sqrt{\frac{2}{t_k - t_{k-1}}} \cos \left(2l\pi \frac{t - t_{k-1}}{t_k - t_{k-1}}\right), \quad t \in [t_{k-1}, t_k).$$

(1.8)

Note that taking only $m_{k,1}$ ($n = 1$) 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., $t_k = k\Delta, \quad K\Delta = T$. For $t \in [0, T]$, there exists a constant $C > 0$ such that

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

and for sufficient small $\epsilon > 0$

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

For $t = t_k$, we have

$$W(t_k) - W^{(n,K)}(t_k) = 0, \quad (1.10)$$

if the CONS $\{m_{k,l}\}_{l=1}^{\infty}$ contains $\frac{1}{\sqrt{t_k - t_{k-1}}}$ as its elements, i.e.,

$$\int_{t_{k-1}}^{t_k} \frac{m_{k,l}(s)}{\sqrt{t_k - t_{k-1}}} ds = \delta_{l,1}.$$  

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

$$\mathbb{E}[|W(t) - W^{(n,K)}(t)|^2] = \sum_{k=1}^{K} \sum_{l=n+1}^{\infty} \left(\int_{t_{k-1}}^{t_k} m_{k,l}(s) ds\right)^2$$

$$= \sum_{l=n+1}^{\infty} \left(\int_{t_{k-1}}^{t_k} \chi_{[0,t]}(s)m_{k,l}(s) ds\right)^2, \quad (t \in I_k)$$

$$\leq C \frac{\Delta}{n},$$

where $t_k \wedge t = \min(t_k, t)$ 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 $\|v\| + \left(\int_0^T \int_{[0,t]} \frac{|w(t) - w(s)|^2}{|s-t|^{1+\theta}} ds dt\right)^{1/2}$.  

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

---

1The big “$\mathcal{O}$” implies that the error is bounded by a positive constant times the term in the parenthesis.
If \( t = t_k \), we have \( \int_{t_{k-1}}^{t_k} m_{k,l}(s) \, ds = 0 \) for any \( l \geq 2 \) if \( m_{k,1} = \frac{1}{\sqrt{t_k - t_{k-1}}} \), 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

\[
\int_{t_n}^{t_{n+1}} dW^{(n)}(t) = \Delta W_n, \quad \Delta W_n = W(t_{n+1}) - W(t_n).
\]

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 \([0, T]\),

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

where \( \Pi_n = \{0 = t_0 < t_1 < t_2 < \cdots < t_n = T\} \) is a partition of the interval \([0, T]\) and \(|\Pi_n| = \max_{0 \leq i < n-1} |t_{i+1} - t_i|\). 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.,

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

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

**Example 2.1.** It can be readily checked that

\[
(2.1) \quad \int_0^T W(t) \, dW(t) = \frac{W^2(T) - T}{2}, \quad \int_0^T W(t) \, dW(t) = \frac{W^2(T)}{2}.
\]

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

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

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

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

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

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

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:

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

The quadratic covariation can be computed by the polarization identity:

$$[X, Y]_t = \frac{1}{2} ([X + Y]_t - [X - Y]_t).$$  

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

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

For Ito integral, we have the following properties. Define

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

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

- (linear property) \( \int_0^t (af(s)+bg(s)) dW(s) = a \int_0^t f(s) dW(s)+b \int_0^t g(s) dW(s) \), \( a, b \in \mathbb{R} \),
• (Ito isometry) \( \mathbb{E}[\left( \int_0^t f(s) \, dW(s) \right)^2] = \int_0^t \mathbb{E}[f^2(s)] \, ds \),

• (Generalized Ito isometry) \( \mathbb{E}\left[ \int_0^t f(s) \, dW(s) \int_0^t g(s) \, dW(s) \right] = \int_0^t \mathbb{E}[f(s)g(s)] \, dt \),

• \( M_t = \int_0^t f(s) \, dW(s) \) is a continuous martingale. Moreover, the quadratic variation of \( M_t \) is \( |M|_{2,\text{TV}} = \int_0^t f^2(s) \, ds, 0 \leq t \leq T \) and

\[
\mathbb{E}\left[ \sup_{0 \leq t \leq T} \left( \int_0^t f(s) \, dW(s) \right)^2 \right] \leq 4 \int_0^t \mathbb{E}[f^2(s)] \, ds.
\]

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

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

Here \( H_n \) is the \( n \)-th Hermite polynomial:

\[
H_n(x) = (-1)^n e^{x^2/2} \frac{d^n}{dx^n} e^{-x^2/2}.
\]

When \( n = 0 \), \( H_n(x) = 1 \) and we use the convention that when \( n < 1 \) the integral is defined as 1. When \( n = 1 \), \( \int_0^t dW(t_1) = W(t) = t^{1/2} H_1\left( \frac{W(t)}{\sqrt{t}} \right) \) as \( H_1(x) = x \). Then it can be shown by induction that the integrand in the left hand side is in \( L^2_{\text{ad}}(\Omega; L^2([0, t])) \), 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 \( B_H(t), t \geq 0 \), can be introduced. It is a centered Gaussian process with the following covariance function

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

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

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

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

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

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

\[
f(\omega) = \frac{1}{2\pi} \int_{0}^{\infty} \cos(\omega t) [K(t) + K^T(t)] \, dt.
\]

In this case, \( \text{Im} f = 0 \), we can set

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

where \( W \) and \( B \) are independent standard Brownian motions and

\[
Y_t = \sqrt{2} \int_{\mathbb{R}^+} \cos(\omega t) (\text{Re} f)^{1/2} \, dW(\omega) - \sqrt{2} \int_{\mathbb{R}^+} \sin(\omega t) (\text{Re} f)^{1/2} \, dB(\omega).
\]
Then we can obtain a discretization using the trapezoidal rule

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

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

\[ \tilde{Y}_t = \sqrt{2} \sum_{j=0}^{N-1} (\text{Re} f)^{1/2}(\omega_{j+1/2}) (\cos(\omega_{j+1/2} t) \xi_j + \sin(\omega_{j+1/2} t) \eta_j) \sqrt{\Delta t}_j. \]

Here \( \xi_j \)'s and \( \eta_j \)'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 \( X(t) \sim N(0, \exp(-\frac{|x-y|^2}{t^2})) \). Then

\[ X(t) = \sqrt{2} (\pi)^{\frac{1}{4}} \int_{\mathbb{R}^+} (\lambda)^{\frac{1}{2}} \cos(\lambda t) \exp(-\lambda^2 t^2) d\lambda X(\omega) - \sqrt{2} (\pi)^{\frac{1}{4}} \int_{\mathbb{R}^+} \sin(\lambda t) \exp(-\lambda^2 t^2) d\lambda. \]

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

\[ X(t) = \sqrt{2} (\pi)^{\frac{1}{4}} \int_{\mathbb{R}^+} \cos(\frac{\lambda}{t} t) \exp(-\frac{\lambda^2}{8} t) d\lambda - \sqrt{2} (\pi)^{\frac{1}{4}} \int_{\mathbb{R}^+} \sin(\frac{\lambda}{t} t) \exp(-\frac{\lambda^2}{8} t) d\lambda. \]

**Application III.** Estimate the error of linear interpolation of Gaussian process. Applying a linear interpolation \( \Pi_1 \) on the interval \([t_i, t_{i+1}]\), then

\[ \mathbb{E}[(X(t) - \Pi_1 X(t))^2] = 2(\pi)^{\frac{1}{4}} \int_{\mathbb{R}^+} \left[ (I - \Pi_1)(\frac{\lambda}{t} t) \right]^2 \exp(-\frac{\lambda^2}{4} t) d\lambda \]

\[ + 2(\pi)^{\frac{1}{4}} \int_{\mathbb{R}^+} \left[ (I - \Pi_1)(\frac{\lambda}{t} t) \right]^2 \exp(-\frac{\lambda^2}{4} t) d\lambda \]

\[ \leq (\pi)^{\frac{1}{4}} (t_{i+1} - t_i)^4 \int_{\mathbb{R}^+} \left( \frac{\lambda}{t} t \right)^4 \exp(-\frac{\lambda^2}{4} t) d\lambda \]

\[ = 24 \pi \frac{(t_{i+1} - t_i)^4}{t^4}. \]

where we recalled the error estimate of the linear interpolation on \([a, b]\) (by Taylor’s expansion)

\[ (2.10) \quad |(I - \Pi_1)g(t)| = \frac{1}{2} (b-a)^2 |g''(\xi)|, \]

where \( \xi \) lies in between \( a \) and \( b \). Then by the Borel-Cantelli lemma, we have

\[ (2.11) \quad |X(t) - \Pi_1 X(t)| \leq C \frac{(t_{i+1} - t_i)^{2-\epsilon}}{t^2}, \quad \epsilon > 0, \]

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 $\sigma_s$ are adapted w.r.t. $\mathcal{F}_s$ and

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

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

- for any $s$, $B_s$, $a(s)$ and $\sigma(s)$ are $\mathcal{F}_s$-measurable;
- for any $t_1 \leq t_2$, $B_{t_2} - B_{t_1}$ 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 $\{\mathcal{F}_s, t_0 \leq s \leq t\}$ is defined such that for any $s$, $W_r(s)$ is $\mathcal{F}_s$-measurable and for any $t_1 \leq t_2$, $W_r(t_2) - W_r(t_1)$ 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

$$\Lambda_r = \sigma_r^T \nabla = \sum_{i=1}^{d} \sigma_{i,r} \frac{\partial}{\partial x_i}, \quad \nabla = \left( \frac{\partial}{\partial x_1}, \frac{\partial}{\partial x_2}, \ldots, \frac{\partial}{\partial x_d} \right),$$

$$\mathcal{L} = \frac{\partial}{\partial t} + a^T \nabla + \frac{1}{2} \sum_{r=1}^{m} \sum_{i,j=1}^{d} \sigma_{i,r} \sigma_{j,r} \frac{\partial^2}{\partial x_i \partial x_j}.$$
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.

<table>
<thead>
<tr>
<th>$\times$</th>
<th>$dW_i(t)$</th>
<th>$dt$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$dW_i(t)$</td>
<td>$\mathbf{1}_{i=j} dt$</td>
<td>0</td>
</tr>
<tr>
<td>$dt$</td>
<td>0</td>
<td>$0$</td>
</tr>
</tbody>
</table>

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

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

APPENDIX A. NAIVE STOCHASTIC INTEGRATION IS IMPOSSIBLE

Let $(\Pi_m)$ 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 $S^{(m)}_{\phi} : C([0, T] \to R)$ by

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

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

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

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

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

$$\sum_{t_k, t_{k-1} \in \Pi_m} |\phi(t_k) - \phi(t_{k-1})| = S^{(m)}_{\phi}(h) \leq \left\|S^{(m)}_{\phi}\right\| \|h\| < \infty. \quad \Box$$

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

$$(B.1) \quad \int_{\mathbb{R}} \frac{\ln f(\lambda)}{1 + \lambda^2} d\lambda > -\infty.$$
Appendix C. Bramble-Hilbert Lemma

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

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

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

References
