Lecture 3: The Itô Integral and Stochastic Differential Equations

Dr. Abhishek Chaudhary
Numerical Analysis Group, Department of Mathematics,
University of Tübingen

Today's Agenda

1. Motivation: Why move from discrete to continuous models?
2. Introduction to probability spaces and stochastic processes.
3. Wiener process and its properties.
4. Itô integral: construction and properties.
5. Stochastic differential equations (SDEs) and the Itô formula.
6. Worked examples and exercises.

1. The Itô integral and stochastic differential equations

The model considered in 1.5 is clearly too simple: only two discrete times, only two possible prices of $S(T)$.
Goal: Construct a more realistic model for the dynamics of $S(t)$.
Ansatz:

$$\underbrace{\frac{dS}{dt} = f(t, S)}_{\text{ordinary differential equation}} \qquad + \underbrace{\text{random noise}}_{?}$$

2.1 Some definitions from probability theory

Definition 2.1.1 (Probability space) The triple $(\Omega, \mathcal{F}, \mathbb{P})$ is called a probability space, if the following holds:

1. $\Omega \neq \emptyset$ is a set, and $\mathcal{F}$ is a σ-algebra (or σ-field) on Ω, i.e. a family of subsets of Ω with the following properties:

   - $\emptyset \in \mathcal{F}$
   - If $F \in \mathcal{F}$, then $\Omega \setminus F \in \mathcal{F}$
   - If $F_i \in \mathcal{F}$ for all $i \in \mathbb{N}$, then $\bigcup_{i=1}^{\infty} F_i \in \mathcal{F}$

   The pair $(\Omega, \mathcal{F})$ is called a measurable space.

2. $\mathbb{P} : \mathcal{F} \to [0, 1]$ is a probability measure, i.e.

   - $\mathbb{P}(\emptyset) = 0$ and $\mathbb{P}(\Omega) = 1$
   - If $F_i \in \mathcal{F}$ for all $i \in \mathbb{N}$ are pairwise disjoint (i.e. $F_i \cap F_j = \emptyset$ for $i \neq j$), then

$$\mathbb{P}\Big(\bigcup_{i=1}^{\infty} F_i\Big) = \sum_{i=1}^{\infty} \mathbb{P}(F_i).$$

A probability space is complete if $\mathcal{F}$ contains all subsets G of Ω with $\mathbb{P}$-outer measure zero, i.e. with

$$\mathbb{P}^*(G) := \inf\{\mathbb{P}(F) : F \in \mathcal{F} \text{ and } G \subset F\} = 0.$$

Any probability space can be completed. Hence, we can assume that every probability space in this lecture is complete.

Definition 2.1.2 (Borel σ-algebra) If $\mathcal{U}$ is a family of subsets of Ω, then the σ-algebra generated by $\mathcal{U}$ is

$$\mathcal{F}_{\mathcal{U}} = \bigcap \{\mathcal{F} : \mathcal{F} \text{ is a } \sigma\text{-algebra of } \Omega \text{ and } \mathcal{U} \subset \mathcal{F}\}.$$

If $\mathcal{U}$ is the collection of all open subsets of a topological space Ω (e.g. $\Omega = \mathbb{R}^d$), then $\mathcal{B} = \mathcal{F}_{\mathcal{U}}$ is called the Borel σ-algebra on Ω. The elements $B \in \mathcal{B}$ are called Borel sets.

For the rest of this section $(\Omega, \mathcal{F}, \mathbb{P})$ is a probability space.

Definition 2.1.3 (Measurable functions, random variables)

- A function $X : \Omega \to \mathbb{R}^d$ is called $\mathcal{F}$-measurable if

$$X^{-1}(B) := \{\omega \in \Omega : X(\omega) \in B\} \in \mathcal{F}$$

  for all Borel sets $B \in \mathcal{B}$. If $(\Omega, \mathcal{F}, \mathbb{P})$ is a probability space, then every $\mathcal{F}$-measurable function is called a random variable.

- Random variables $X_1, \ldots, X_n$ are called independent if

$$\mathbb{P}\Big(\bigcap_{i=1}^{n} X_i^{-1}(A_i)\Big) = \prod_{i=1}^{n} \mathbb{P}(X_i^{-1}(A_i))$$

  for all $A_1, \ldots, A_n \in \mathcal{B}$.

- If $X : \Omega \to \mathbb{R}^d$ is any function, then the σ-algebra generated by X is the smallest σ-algebra on Ω containing all the sets $X^{-1}(B)$ for all $B \in \mathcal{B}$. Notation: $\mathcal{F}^X = \sigma\{X\}$. $\mathcal{F}^X$ is the smallest σ-algebra where X is measurable.

Definition 2.1.4 (Stochastic process) Let T be an ordered set (e.g. $T = [0, \infty)$, $T = \mathbb{N}$). A stochastic process is a family $X = \{X_t : t \in T\}$ of random variables

$$X_t : \Omega \to \mathbb{R}^d.$$

Below, we will often simply write X_t instead of $\{X_t : t \in T\}$. Equivalent notations: $X(t, \omega)$, $X(t)$, $X_t(\omega)$, $X_{t_1}, \ldots, X_{t_n}$. For a fixed $\omega \in \Omega$, the function $t \mapsto X_t(\omega)$ is called a realization (or path or trajectory) of X.

The path of a stochastic process is associated to some $\omega \in \Omega$. As time evolves, more information about ω becomes available.

Example. Toss a coin three times. Possible results:

ω_1	ω_2	ω_3	ω_4	ω_5	ω_6	ω_7	ω_8
HHH	HHT	HTH	HTT	THH	THT	TTH	TTT

(H = heads, T = tails).

- Before the first toss, we only know that $\omega \in \Omega$.
- After the first toss, we know if the final result will belong to $\{HHH, HHT, HTH, HTT\}$ or to $\{THH, THT, TTH, TTT\}$. These sets are "resolved by the information". Hence, we know in which of the sets $\{\omega_1, \omega_2, \omega_3, \omega_4\}$, $\{\omega_5, \omega_6, \omega_7, \omega_8\}$ ω is.
- After the second toss, the sets $\{HHH, HHT\}$, $\{HTH, HTT\}$, $\{THH, THT\}$, $\{TTH, TTT\}$ are resolved, and we know in which of the sets $\{\omega_1, \omega_2\}$, $\{\omega_3, \omega_4\}$, $\{\omega_5, \omega_6\}$, $\{\omega_7, \omega_8\}$ ω is.

Definition 2.1.5 (Filtration)

- A filtration is a family $\{\mathcal{F}_t : t \geq 0\}$ of sub-σ-algebras of $\mathcal{F}$ such that $\mathcal{F}_s \subset \mathcal{F}_t$ for all $t \geq s \geq 0$.
- If $\{X_t : t \geq 0\}$ is a family of random variables and X_t is $\mathcal{F}_t$-measurable, then $\{X_t : t \geq 0\}$ is adapted to (or nonanticipating with respect to) $\{\mathcal{F}_t : t \geq 0\}$. Interpretation: At time t we know for each set $S \in \mathcal{F}_t$ if $\omega \in S$ or not.
- $\{\mathcal{F}_t : t \geq 0\}$ is called the *natural filtration* of a stochastic process X_t if $\mathcal{F}_t$ is the smallest σ-algebra which contains $\mathcal{F}_s^X$ for all $s \in [0, t]$, i.e. $\mathcal{F}_t = \sigma\{X_s, s \in [0, t]\}$. *This is the smallest filtration to which X_t is adapted.*

Definition 2.1.6 (Normal distribution) *A random variable* $X : \Omega \to \mathbb{R}^d$ *with* $d \in \mathbb{N}$ *is normal if it has a multivariate normal (Gaussian) distribution with mean* $\mu \in \mathbb{R}^d$ *and a symmetric, positive definite covariance matrix* $\Sigma \in \mathbb{R}^{d\times d}$*, i.e. if*

$$\mathbb{P}(X \in B) = \int_B \frac{1}{\sqrt{(2\pi)^d \det(\Sigma)}} \exp\left(-\frac{1}{2}(x-\mu)^T \Sigma^{-1}(x-\mu)\right) dx$$

for all Borel sets $B \subset \mathbb{R}^d$*. Notation:* $X \sim \mathcal{N}(\mu, \Sigma)$

Remarks:

1. If $X \sim \mathcal{N}(\mu, \Sigma)$, then $\mathbb{E}(X) = \mu$ and $\Sigma = (\sigma_{ij})$ with $\sigma_{ij} = \mathbb{E}[(X_i - \mu_i)(X_j - \mu_j)]$.
2. Standard normal distribution $\Leftrightarrow \mu = 0$, $\Sigma = I$. (I identity matrix)
3. If $X \sim \mathcal{N}(\mu, \Sigma)$ and $Y = v + TX$ for some $v \in \mathbb{R}^d$ and a regular matrix $T \in \mathbb{R}^{d\times d}$, then
$$Y \sim \mathcal{N}(v + T\mu, T\Sigma T^T).$$

2.2 The Wiener process

A very important stochastic process is the Wiener process. This process will serve as the "source of randomness" in our model of the financial market.
Robert Brown 1827, Louis Bachelier 1900, Albert Einstein 1905, Norbert Wiener 1923

Definition 2.2.1 (Wiener process, Brownian motion)
A continuous-time stochastic process $\{W_t : t \in [0,T)\}$ *is called a standard Brownian motion or standard Wiener process if it has the following properties:*

1. $W_0 = 0$ *(with probability one)*
2. *Independent increments: For all* $0 \le t_1 < t_2 < \cdots < t_n < T$ *the random variables*
$$W_{t_2} - W_{t_1}, W_{t_3} - W_{t_2}, \ldots, W_{t_n} - W_{t_{n-1}}$$
are independent.
3. $W_t - W_s \sim \mathcal{N}(0, t-s)$ *for any* $0 \le s < t < T$.
4. *There is a* $\hat{\Omega} \subset \Omega$ *with* $\mathbb{P}(\hat{\Omega}) = 1$ *such that* $t \mapsto W_t(\omega)$ *is continuous for all* $\omega \in \hat{\Omega}$.

If $W_t^{(1)}, \ldots, W_t^{(d)}$ *are independent one-dimensional Wiener processes, then* $W_t = \left(W_t^{(1)}, \ldots, W_t^{(d)}\right)$ *is called a d-dimensional Wiener process, and*

$$W_t - W_s \sim \mathcal{N}(0, (t-s)I).$$

Numerical simulation of a Wiener process

Choose step-size $h > 0$, put $t_n = n \cdot h$ and $\tilde{W}_0 = (0, \dots, 0) \in \mathbb{R}^d$.

For $n = 0, 1, 2, 3, \dots$

- Generate random vector $Z_n \sim \mathcal{N}(0, I)$
- $\tilde{W}_{n+1} = \tilde{W}_n + \sqrt{h} Z_n$

For $h \to 0$ the interpolation of $\tilde{W}_n$, $n \in \mathbb{N}$ approximates a path of Brownian motion.

How smooth is a path of a Wiener process? Consider only $d = 1$.

Hölder continuity and non-differentiability

A function $f : (a, b) \to \mathbb{R}$ is **Hölder continuous of order** α for some $\alpha \in [0, 1]$ if there is a constant C such that

$$|f(t) - f(s)| \leq C|t - s|^\alpha \qquad \text{for all } s, t \in (a, b).$$

If $\alpha = 1$, then f is Lipschitz continuous.
If $\alpha > 0$, then f is uniformly continuous.
If $\alpha = 0$, then f is bounded.

A path of Brownian motion on a bounded interval is Hölder continuous for any $\alpha \in (0, \frac{1}{2})$ with probability one.
For $\alpha \geq \frac{1}{2}$, however, the path is not Hölder continuous with probability one.
In particular, a path of Brownian motion is nowhere differentiable with probability one.

Proofs: [Ste01], chapter 5

Unbounded total variation

Let $[a, b]$ be an interval and let

$$P_N = (t_n)_{n=0}^N, \qquad a = t_0 < t_1 < \cdots < t_N = b$$

be a partition of $[a, b]$ with $|P_N| = \max_n |t_n - t_{n-1}|$.
Example: equidistant partition, $h = (b - a)/N$, $t_n = a + n \cdot h$. Then, the **total variation** of a function $f : (a, b) \to \mathbb{R}$ is

$$TV_{a,b}(f) = \lim_{\substack{N \to \infty \\ |P_N| \to 0}} \sum_{n=1}^N |f(t_n) - f(t_{n-1})|.$$

If f is differentiable and f' is integrable, then (exercise)

$$TV_{a,b}(f) = \int_a^b |f'(t)| \, dt$$

Conversely: If a function f has bounded total variation, then its derivative exists for almost all $x \in [a, b]$.
Consequence: A path of the Wiener process has unbounded total variation with probability one.

Quadratic variation

The quadratic variation of a function $f : (a, b) \to \mathbb{R}$ is

$$QV_{a,b}(f) = \lim_{\substack{N\to\infty \\ |P_N|\to 0}} \sum_{n=1}^{N} \left(f(t_n) - f(t_{n-1})\right)^2 .$$

If f is continuously differentiable, then (exercise)

$$QV_{a,b}(f) = 0.$$

For a path $t \mapsto W_t(\omega)$ with $t \in [0, T]$, however, it can be shown (exercise) that

$$\lim_{\substack{N\to\infty \\ |P_N|\to 0}} \left\| \sum_{n=1}^{N} \Big(W_{t_n}(\omega) - W_{t_{n-1}}(\omega)\Big)^2 - T \right\|_{L^2(\mathbb{P})} = 0,$$

where

$$\|X\|_{L^2(\mathbb{P})} = \sqrt{\mathbb{E}(X^2)} = \left(\int_{\omega\in\Omega} X^2(\omega)\; d\mathbb{P}(\omega)\right)^{1/2}.$$

By choosing a suitable subsequence, it can be concluded that $QV_{0,t}(t \mapsto W_t(\omega)) = t$ with probability one.

The standard Brownian filtration

The natural filtration of Brownian motion on $[0, T]$ is given by

$$\{\mathcal{F}_t : t \in [0, T]\}, \qquad \mathcal{F}_t = \sigma\{W_s,\ s \in [0, t]\}$$

(cf. Definition 2.1.5). For technical reasons, however, it is more advantageous to use the **augmented filtration**:

- For fixed t let $\mathcal{Z} = \{S \in \sigma\{W_s,\, s \in [0, T]\} : \mathbb{P}(S) = 0\}$.
- Let $\widehat{\mathcal{Z}} = \{\hat{S} \subset \Omega : \hat{S} \subset S \text{ for some } S \in \mathcal{Z}\}$.
- Extend $\mathbb{P}$ by defining $\mathbb{P}(\hat{S}) = 0$ for all $\hat{S} \in \widehat{\mathcal{Z}}$.
- Re-define $\mathcal{F}_t$ as the smallest σ-algebra which contains $\sigma\{W_s,\, s \in [0, t]\}$ and $\widehat{\mathcal{Z}}$.

This filtration is called the **standard Brownian filtration**.

- Consequence: $\mathcal{F}_t$ fulfills the “usual conditions”, i.e.

– If $\hat{S} \in \widehat{\mathcal{Z}}$, then $\hat{S} \in \mathcal{F}_0$.

– Right continuity: $\mathcal{F}_t = \bigcap_{s>t} \mathcal{F}_s$

2.3 The Itô integral

Itô Kiyoshi 1944

Motivation

Goal: Define stochastic differential equations. Naïve Ansatz:

$$\frac{dX}{dt} = \underbrace{f(t,X)}_{\text{ordinary differential equation}} + \underbrace{g(t,X)Z(t)}_{\text{random noise}}, \qquad Z(t) = ?$$

Apply explicit Euler method: Choose $t \geq 0$ and step-size $N \in \mathbb{N}$, let $h = t/N$, $t_n = n \cdot h$ and define approximations $X_n \approx X(t_n)$ by

$$X_{n+1} = X_n + hf(t_n, X_n) + hg(t_n, X_n)Z(t_n) \qquad (n = 0, 1, 2, \ldots).$$

In the special case $f(t,X) = 0$ and $g(t,X) = 1$, we want that $X_n = W(t_n)$ is the Wiener process, i.e. we postulate that

$$W(t_{n+1}) \stackrel{!}{=} W(t_n) + hZ(t_n).$$

This yields

$$X_{n+1} = X_n + hf(t_n, X_n) + g(t_n, X_n)\left(W(t_{n+1}) - W(t_n)\right)$$

and after N steps

$$X_N = X_0 + h\sum_{n=0}^{N-1} f(t_n, X_n) + \sum_{n=0}^{N-1} g(t_n, X_n)\left(W(t_{n+1}) - W(t_n)\right). \tag{2.2}$$

Keep t fixed, let $N \to \infty$, $h = t/N \to 0$. Then, (2.2) should somehow converge to

$$X(t) = X(0) + \int_0^t f(s, X(s))\, ds + \int_0^t g(s, X(s))\, dW(s). \tag{2.3}$$

Problem: We cannot define ($\star$) as a pathwise Riemann-Stieltjes integral! When $N \to \infty$, the sum

$$\sum_{n=0}^{N-1} g(t_n, X_n(\omega))\left(W(t_{n+1}, \omega) - W(t_n, \omega)\right)$$

diverges with probability one, because a path of the Wiener process has unbounded total variation with probability one.

New goal: Define the integral

$$\mathcal{I}_t[u](\omega) = \int_0^t u(s,\omega)\, dW_s(\omega)$$

in a "reasonable" way for the following class of functions.

Definition 2.3.1 *Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space, and let $\{\mathcal{F}_t : t \in [0,T]\}$ be the standard Brownian filtration. Then, we define $\mathcal{H}^2 = \mathcal{H}^2[0,T]$ to be the class of functions*

$$u = u(t,\omega), \qquad u : [0,T] \times \Omega \longrightarrow \mathbb{R}$$

with the following properties:

- $(t,\omega) \mapsto u(t,\omega)$ is $(\mathcal{B} \times \mathcal{F})$-measurable.
- u is adapted to $\{\mathcal{F}_t : t \in [0,T]\}$, i.e. $u(t,\cdot)$ is $\mathcal{F}_t$-measurable.
- $\mathbb{E}\left(\int_0^T u^2(t,\omega)\, dt\right) < \infty$

Step 1: Itô integral for elementary functions

Definition 2.3.2 (Elementary functions) *A function $\phi \in \mathcal{H}^2$ is called elementary if it is a stochastic step function of the form*

$$\begin{aligned}\phi(t,\omega) &= a_0(\omega)\, 1_{[0,t_1)}(t) + \sum_{n=0}^{N-1} a_n(\omega)\, 1_{[t_n,t_{n+1})}(t) \\ &= a_0(\omega)\, 1_{[0,t_1)}(t) + \sum_{n=1}^{N-1} a_n(\omega)\, 1_{[t_n,t_{n+1})}(t)\end{aligned}$$

where a_n is $\mathcal{F}_{t_n}$-measurable with $\mathbb{E}(a_n^2) < \infty$. Here and below,

$$1_{[c,d]}(t) = \begin{cases} 1 & \text{if } t \in [c,d] \\ 0 & \text{else} \end{cases} \tag{2.4}$$

is the indicator function of an interval $[c,d]$.

For $0 \leq c < d \leq T$, the only reasonable way to define the Itô integral of an indicator function $1_{[c,d]}$ is

$$\mathcal{I}_T[1_{[c,d]}](\omega) = \int_0^T 1_{[c,d]}(s)\, dW(s,\omega) = \int_c^d dW(s,\omega) = W(d,\omega) - W(c,\omega).$$

Hence, by linearity, we define the Itô integral of an elementary function by

$$\mathcal{I}_T[\phi](\omega) = \sum_{n=0}^{N-1} a_n(\omega)\left(W(t_{n+1},\omega) - W(t_n,\omega)\right).$$

Lemma 2.3.3 (Itô isometry for elementary functions) *For all elementary functions we have*

$$\mathbb{E}\left(\mathcal{I}_T[\phi]^2\right) = \mathbb{E}\left(\int_0^T \phi^2(t,\omega)\,dt\right)$$

or equivalently

$$\|\mathcal{I}_T[\phi]\|_{L^2(\mathbb{P})} = \|\phi\|_{L^2(dt\times d\mathbb{P})}$$

with

$$\|\phi\|_{L^2(dt\times d\mathbb{P})} = \left(\int_\Omega \int_0^T \phi^2(t,\omega)\,dt\,d\mathbb{P}\right)^{\frac{1}{2}} = \left(\mathbb{E}\left(\int_0^T \phi^2(t,\omega)\,dt\right)\right)^{\frac{1}{2}}.$$

Proof. Since

$$\phi^2(t,\omega) = a_0^2(\omega)1_{[0,t_1)}(t) + \sum_{n=0}^{N-1} a_n^2(\omega)1_{[t_n,t_{n+1})}(t)$$

we obtain

$$\mathbb{E}\left(\int_0^T \phi^2(t,\omega)\,dt\right) = \sum_{n=0}^{N-1} \mathbb{E}\left(a_n^2\right)(t_{n+1}-t_n) \tag{2.5}$$

for the right-hand side. If we let $\Delta W_n = W(t_{n+1}) - W(t_n)$, then

$$\mathcal{I}_T[\phi]^2 = \left(\sum_{n=0}^{N-1} a_n \Delta W_n\right)^2 = \sum_{n=0}^{N-1}\sum_{m=0}^{N-1} a_n a_m \Delta W_n \Delta W_m. \tag{2.6}$$

By definition, the Wiener process has independent increments with $\mathbb{E}(\Delta W_n) = 0$ and $\mathbb{E}(\Delta W_n^2) = \mathbb{V}(\Delta W_n) = t_{n+1} - t_n$. Since a_n is independent of ΔW_n, it follows that

$$\mathbb{E}(a_n a_m \Delta W_n \Delta W_m) = \begin{cases} 0 & \text{if } n \neq m \\ \mathbb{E}(a_n^2)(t_{n+1}-t_n) & \text{if } n = m \end{cases}$$

and taking the expectation of (2.6) gives

$$\mathbb{E}\left(\mathcal{I}_T[\phi]^2\right) = \sum_{n=0}^{N-1} \mathbb{E}(a_n^2)(t_{n+1}-t_n). \tag{2.7}$$

Comparing (2.5) and (2.7) yields the assertion. ■

Step 2: Itô integral on $\mathcal{H}^2$

Lemma 2.3.4 *For any $u \in \mathcal{H}^2$ there is a sequence $(\phi_k)_{k\in\mathbb{N}}$ of elementary functions $\phi_k \in \mathcal{H}^2$ such that*

$$\lim_{k\to\infty} \|u - \phi_k\|_{L^2(dt\times d\mathbb{P})} = 0$$

Proof: Section 6.6 in [Ste01].

Let $u \in \mathcal{H}^2$ and let $(\phi_k)_{k\in\mathbb{N}}$ be elementary functions such that

$$u = \lim_{k\to\infty} \phi_k \qquad \text{in } L^2(dt \times d\mathbb{P})$$

as in Lemma 2.3.4. The linearity of $\mathcal{I}_T[\cdot]$ and Lemma 2.3.3 yield

$$\|\mathcal{I}_T[\phi_j] - \mathcal{I}_T[\phi_k]\|_{L^2(\mathbb{P})} = \|\mathcal{I}_T[\phi_j - \phi_k]\|_{L^2(\mathbb{P})} = \|\phi_j - \phi_k\|_{L^2(dt\times d\mathbb{P})} \longrightarrow 0$$

for $j, k \to \infty$. Hence, $(\mathcal{I}_T[\phi_k])_k$ is a Cauchy sequence in the Hilbert space $L^2(d\mathbb{P})$. Thus, $(\mathcal{I}_T[\phi_k])_k$ converges in $L^2(d\mathbb{P})$, and we can define

$$\mathcal{I}_T[u] = \lim_{k\to\infty} \mathcal{I}_T[\phi_k].$$

The choice of the sequence does not matter: If $(\psi_k)_{k\in\mathbb{N}}$ are elementary functions with $u = \lim_{k\to\infty} \psi_k$ in $L^2(dt \times d\mathbb{P})$, then by Lemma 2.3.3 we obtain for $k \to \infty$

$$\begin{aligned}\|\mathcal{I}_T[\phi_k] - \mathcal{I}_T[\psi_k]\|_{L^2(\mathbb{P})} &= \|\mathcal{I}_T[\phi_k - \psi_k]\|_{L^2(\mathbb{P})} \\ &= \|\phi_k - \psi_k\|_{L^2(dt\times d\mathbb{P})} \\ &\le \|\phi_k - u\|_{L^2(dt\times d\mathbb{P})} + \|u - \psi_k\|_{L^2(dt\times d\mathbb{P})} \longrightarrow 0.\end{aligned}$$

Theorem 2.3.5 (Itô isometry) *For all $u \in \mathcal{H}^2$ we have*

$$\|\mathcal{I}_T[u]\|_{L^2(\mathbb{P})} = \|u\|_{L^2(dt\times d\mathbb{P})}.$$

Proof: Exercise.

Step 3: The Itô integral as a process

So far we have defined the Itô integral $\mathcal{I}_T[u](\omega)$ over the interval $[0, T]$ for fixed T. For applications in mathematical finance, however, we want to consider $\{\mathcal{I}_t[u](\omega) : t \in [0, T]\}$ as a stochastic process. Therefore, we let

$$m_t(s, \omega) = \begin{cases} 1 & \text{if } s \in [0, t] \\ 0 & \text{else.} \end{cases}$$

If $u \in \mathcal{H}^2$, then $m_t u \in \mathcal{H}^2$. Can we define $\mathcal{I}_t[u](\omega)$ by $\mathcal{I}_T[m_t u](\omega)$? **No! Problem:** The integral $\mathcal{I}_T[m_t u](\omega)$ is only defined in $L^2(d\mathbb{P})$. Hence, the value $\mathcal{I}_T[m_t u](\omega)$ is arbitrary on sets of $\mathbb{P}$-measure zero. This is the case for every $t \in [0, T]$, and since the set $[0, T]$ is uncountable, the union

$$\bigcup_{t\in[0,T]} \{Z_t \in \mathcal{F}_t : \mathbb{P}(Z_t) = 0\}$$

(i.e. the set where the process is not well-defined) could be “very large”!

Let $(\phi_k)_{k\in\mathbb{N}}$ be elementary functions with $\lim_{k\to\infty}\phi_k = u$. Define a continuous process by

$$X^{(k)}(t,\omega) = \mathcal{I}_T[m_t\phi_k](\omega).$$

It can be shown that there is a sub-sequence $(X^{(k_j)})_j$ such that

$$\max_{t\in[0,T]} |X^{(k_j)}(t,\omega) - X^{(k_{j'})}(t,\omega)| \longrightarrow 0 \qquad \text{for } i,j \longrightarrow \infty$$

with probability one. Hence, $(X^{(k_j)})_j$ converges uniformly on $[0,T]$ to a continuous process X with probability one. The assumption

$$\lim_{k\to\infty} m_t\phi_k = m_t u \qquad \text{wrt. } \|\cdot\|_{L^2(dt\times d\mathbb{P})}$$

implies

$$\lim_{k\to\infty} \mathcal{I}_T[m_t\phi_k] = \mathcal{I}_T[m_t u] \qquad \text{wrt. } \|\cdot\|_{L^2(d\mathbb{P})},$$

and since it also can be shown that

$$\lim_{k\to\infty} X^{(k_j)}(t,\omega) = X(t,\omega) \qquad \text{wrt. } \|\cdot\|_{L^2(d\mathbb{P})},$$

it follows that for each $t\in[0,T]$

$$X(t,\omega) = \mathcal{I}_T[m_t u](\omega)$$

with probability one. Details: Theorem 6.2 in [Ste01].

Step 4: The Itô integral on $\mathcal{L}^2_{\text{loc}}$

So far we have defined the Itô integral for functions $u\in\mathcal{H}^2([0,T])$; cf. Definition 2.3.1. Such functions must satisfy

$$\mathbb{E}\left(\int_0^T u^2(t,\omega)\,dt\right) < \infty, \tag{2.8}$$

and this condition is sometimes too restrictive. With some more work, the Itô integral can be extended to all functions

$$u = u(t,\omega), \qquad u : [0,T]\times\Omega \longrightarrow \mathbb{R}$$

with the following properties:

- $(t,\omega)\mapsto u(t,\omega)$ is $(\mathcal{B}\times\mathcal{F})$-measurable.
- u is adapted to $\{\mathcal{F}_t : t\in[0,T]\}$.
- $\mathbb{P}\left(\int_0^T u^2(t,\omega)\,dt < \infty\right) = 1$

This class is called $\mathcal{L}^2_{\text{loc}}[0,T]$. The first two conditions are the same as for $\mathcal{H}^2$, but the third condition is weaker than (2.8). If $y : \mathbb{R} \longrightarrow \mathbb{R}$ is continuous, then $u(t,\omega) = y(W(t,\omega)) \in \mathcal{L}^2_{\text{loc}}[0,T]$, because $\omega \mapsto y(W(t,\omega))$ is continuous with probability one and hence bounded on $[0,T]$.

Details: Chapter 7 in [Ste01].

Notation

The process X constructed above is called the **Itô integral** of $u \in \mathcal{L}^2_{\text{loc}}[0,T]$ and is denoted by

$$X(t,\omega) = \int_0^t u(s,\omega)\, dW(s,\omega).$$

The Itô integral over an arbitrary interval $[a,b] \subset [0,T]$ is defined by

$$\int_a^b u(s,\omega)\, dW(s,\omega) = \int_0^b u(s,\omega)\, dW(s,\omega) - \int_0^a u(s,\omega)\, dW(s,\omega).$$

Alternative notations:

$$\int_a^b u(s,\omega)\, dW(s,\omega) = \int_a^b u(s,\omega)\, dW_s(\omega) = \int_a^b u_s(\omega)\, dW_s(\omega) = \int_a^b u_s\, dW_s$$

Properties of the Itô integral

Lemma 2.3.6 *The Itô integral on $[a,b] \subset [0,T]$ has the following properties:*

1. **Linearity:** For all $c \in \mathbb{R}$ and $u, v \in \mathcal{L}^2_{\text{loc}}$, we have

$$\int_a^b (cu(s,\omega) + v(s,\omega))\, dW_s(\omega) = c\int_a^b u(s,\omega)\, dW_s(\omega) + \int_a^b v(s,\omega)\, dW_s(\omega)$$

with probability one.

2.

$$\mathbb{E}\left(\int_a^b u(s,\omega)\, dW_s(\omega)\right) = 0$$

3. $\int_a^t u(s,\omega)\, dW_s(\omega)$ is $\mathcal{F}_t$-measurable for $t \geq a$.

4. **Itô isometry on $[a,b]$:** For all $u \in \mathcal{L}^2_{\text{loc}}$ we have

$$\mathbb{E}\left(\left(\int_a^b u(s,\omega)\, dW_s(\omega)\right)^2\right) = \mathbb{E}\left(\int_a^b u^2(s,\omega)\, ds\right)$$

Proof. Show these properties for elementary functions and pass to the limit. ∎

Important example.

The Itô integral of $u(s,\omega) = W(s,\omega)$ on $[0,t]$ yields (exercise!)

$$\int_0^t W(s,\omega)\,dW(s,\omega) = \frac{1}{2}W^2(t,\omega) - \frac{1}{2}t. \tag{2.9}$$

The term $-\frac{1}{2}t$ is surprising, because for the corresponding Riemann-Stieltjes integral of a continuously differentiable (deterministic) function $v : [0,t] \longrightarrow \mathbb{R}$ with $v(0) = 0$, we obtain

$$\int_0^t v(s)\,dv(s) = \int_0^t v(s)v'(s)\,ds = \frac{1}{2}\int_0^t \frac{d}{ds}(v^2(s))\,ds = \frac{1}{2}v^2(t). \tag{2.10}$$

The reason for the "strange" behaviour of the Itô integral will be revealed in the next subsection.

2.4 Stochastic differential equations and the Itô formula

Definition 2.4.1 (SDE) *A* ***stochastic differential equation (SDE)*** *is an equation of the form*

$$X(t) = X(0) + \int_0^t f(s,X(s))\,ds + \int_0^t g(s,X(s))\,dW(s). \tag{2.11}$$

The functions $f : \mathbb{R}\times\mathbb{R} \longrightarrow \mathbb{R}$ and $g : \mathbb{R}\times\mathbb{R} \longrightarrow \mathbb{R}$ are called drift and diffusion coefficients, respectively. These functions are typically given while $X(t) = X(t,\omega)$ is unknown. The solution $X(t)$ is called an ***Itô process****.*

This equation is actually not a **differential** equation, but an **integral** equation Often people write

$$dX_t = f(t,X_t)dt + g(t,X_t)dW_t$$

as a shorthand notation for (2.11). Some people even "divide by dt" in order to make the equation look like a differential equation, but this is more than audacious since "dW_t/dt" does not make sense.

Two special cases:

- If $g(t,X(t)) \equiv 0$, then (2.11) is reduced to

$$X(t) = X(0) + \int_0^t f(s,X(s))\,ds.$$

  If $X(t)$ is differentiable, this is equivalent to the initial value problem

$$\frac{dX(t)}{dt} = f(t,X(t)), \qquad X(0) = X_0.$$

- For $f(t, X(t)) \equiv 0$, $g(t, X(t)) \equiv 1$ and $X(0) = 0$, (2.11) turns into

$$X(t) = \underbrace{X(0)}_{=0} + \underbrace{\int_0^t f(s, X(s))\, ds}_{=0} + \int_0^t \underbrace{g(s, X(s))}_{=1}\, dW(s) = W(t) - W(0) = W(t).$$

Computing Riemann integrals via the basic definition is usually very tedious. The fundamental theorem of calculus provides an alternative which is more convenient in most cases. For Itô integrals, the situation is similar: The approximation via elementary functions is rarely used to compute the integral. What is the counterpart of the fundamental theorem of calculus for the Itô integral?

Theorem 2.4.2 (Itô formula) *Let X_t be the solution of the SDE*

$$dX_t = f(t, X_t)dt + g(t, X_t)dW_t$$

and let $F(t, x)$ be a function with continuous partial derivatives $\frac{\partial F}{\partial t}$, $\frac{\partial F}{\partial x}$, and $\frac{\partial^2 F}{\partial x^2}$. Then, we have for $Y_t := F(t, X_t)$ that

$$\begin{aligned} dY_t &= \frac{\partial F}{\partial t}dt + \frac{\partial F}{\partial x}dX_t + \frac{1}{2}\frac{\partial^2 F}{\partial x^2}g^2 dt \\ &= \left(\frac{\partial F}{\partial t} + \frac{\partial F}{\partial x}f + \frac{1}{2}\frac{\partial^2 F}{\partial x^2}g^2\right) dt + \frac{\partial F}{\partial x} g dW_t. \end{aligned} \tag{2.12}$$

with $f = f(t, X_t)$, $g = g(t, X_t)$, $\frac{\partial F}{\partial x} = \frac{\partial F(t, X_t)}{\partial x}$, and so on. **Notation.** From now on, the partial derivatives of some function $u(t, x)$ will be denoted by

$$\partial_t u := \frac{\partial u}{\partial t}, \qquad \partial_x u := \frac{\partial u}{\partial x}, \qquad \partial_x^2 u := \frac{\partial^2 u}{\partial x^2}$$

and so on. Evaluations of the derivatives of F are to be understood in the sense of, e.g.,

$$\partial_x F(s, X_s) := \partial_x F(t, x)|_{(t,x)=(s,X_s)}$$

and so on.

Remarks:

1. The Itô formula can be considered as a stochastic chain rule, but the term $\frac{1}{2}\partial_x^2 F \cdot g^2 dt$ is surprising since such a term does not appear in the chain rule for deterministic functions: If X_t and $F(t, x)$ are smooth deterministic functions, then the derivative of $t \mapsto F(t, X_t)$ is

$$\partial_t F(t, X_t) + \partial_x F(t, X_t) \cdot \frac{dX_t}{dt}, \qquad \text{i.e. } dF = \partial_t F dt + \partial_x F dX_t.$$

2. Let $f(t, X_t) = 0$, $g(t, X_t) = 1$, $X_t = W_t$ and suppose that $F(t, x) = F(x)$. Then,

the Itô formula yields for $Y_t := F(W_t)$ that

$$dY_t = F'(W_t)dW_t + \frac{1}{2}F''(W_t)dt$$

which is the shorthand notation for

$$F(W_t) = F(W_0) + \int_0^t F'(W_s)dW_s + \frac{1}{2}\int_0^t F''(W_s)ds.$$

This can be seen as a counterpart of the fundamental theorem of calculus. Again, the last term is surprising, because for a suitable deterministic function $v(t) = v_t$ we obtain

$$F(v_t) = F(v_0) + \int_0^t F'(v_s)dv_s.$$

Sketch of the proof of Theorem 2.4.2.

- Equation (2.12) is the shorthand notation for

$$\begin{aligned} Y_t = Y_0 + \int_0^t \Big(\partial_t F(s, X_s) + \partial_x F(s, X_s) \cdot f(s, X_s) + \frac{1}{2}\partial_x^2 F(s, X_s) \cdot g^2(s, X_s)\Big)\, ds \\ + \int_0^t \partial_x F(s, X_s) \cdot g(s, X_s)dW_s \end{aligned}$$

Assume that F is twice continuously differentiable with bounded partial derivatives. (Otherwise F can be approximated by such functions with uniform convergence on compact subsets of $[0, \infty) \times \mathbb{R}$.) Assume that $(t, \omega) \mapsto f(t, X_t(\omega))$ and $(t, \omega) \mapsto g(t, X_t(\omega))$ are elementary functions. (Otherwise approximate by elementary functions.) Hence, there is a partition $0 = t_0 < t_1 < \ldots < t_N = t$ such that

$$f(t, X_t(\omega)) = f(0, X_0(\omega))1_{[0,t_1)}(t) + \sum_{n=1}^{N-1} f(t_n, X_{t_n}(\omega))1_{[t_n,t_{n+1})}(t)$$

and the same equation with f replaced by g.

Notation: For the rest of the proof, we define

$$f^{(n)} := f(t_n, X_{t_n}), \qquad F^{(n)} := F(t_n, X_{t_n}),$$

$$g^{(n)} := g(t_n, X_{t_n}), \qquad \partial_t F^{(n)} := \partial_t F(t_n, X_{t_n})$$

and so on, and

$$\Delta t_n = t_{n+1} - t_n, \qquad \Delta X_n = X_{t_{n+1}} - X_{t_n}, \qquad \Delta W_n = W_{t_{n+1}} - W_{t_n}.$$

Since f and g are elementary functions, we have

$$X_{t_n} = X_0 + \int_0^{t_n} f(s, X_s)\, ds + \int_0^{t_n} g(s, X_s)\, dW_s$$

$$X_{t_n} = X_0 + \sum_{k=0}^{n-1} f(t_k, X_{t_k})\Delta t_k + \sum_{k=0}^{n-1} g(t_k, X_{t_k})\Delta W_k.$$

and hence

$$\Delta X_n = X_{t_{n+1}} - X_{t_n} = f^{(n)}\Delta t_n + g^{(n)}\Delta W_n.$$

Telescoping sum:

$$Y_t = Y_{t_N} = Y_0 + \sum_{n=0}^{N-1}(Y_{t_{n+1}} - Y_{t_n}) = Y_0 + \sum_{n=0}^{N-1}(F^{(n+1)} - F^{(n)})$$

Apply Taylor's theorem:

$$\begin{aligned} F^{(n+1)} - F^{(n)} &= \partial_t F^{(n)} \cdot \Delta t_n + \partial_x F^{(n)} \cdot \Delta X_n + \frac{1}{2}\partial_x^2 F^{(n)} \cdot (\Delta X_n)^2 + \partial_t\partial_x F^{(n)} \cdot \Delta t_n \Delta X_n \\ &\quad + \frac{1}{2}\partial_t^2 F^{(n)} \cdot (\Delta t_n)^2 + R_n(\Delta t_n, \Delta X_n) \end{aligned}$$

with a remainder term R_n. Insert this into the telescoping sum.

- Consider the limit $N \to \infty$, $\Delta t_n \to 0$ with respect to $\|\cdot\|_{L^2(d\mathbb{P})}$. For the first two terms, this yields

$$\lim_{N\to\infty} \sum_{n=0}^{N-1} \partial_t F^{(n)} \cdot \Delta t_n = \lim_{N\to\infty} \sum_{n=0}^{N-1} \partial_t F(t_n, X_{t_n}) \cdot \Delta t_n = \int_0^t \partial_t F(s, X_s)\, ds$$

and

$$\begin{aligned} \lim_{N\to\infty} \sum_{n=0}^{N-1} \partial_x F^{(n)} \cdot \Delta X_n &= \lim_{N\to\infty} \sum_{n=0}^{N-1} \partial_x F^{(n)} \cdot f^{(n)}\Delta t_n + \lim_{N\to\infty} \sum_{n=0}^{N-1} \partial_x F^{(n)} \cdot g^{(n)}\Delta W_n \\ &= \int_0^t \partial_x F(s, X_s) \cdot f(s, X_s)\, ds + \int_0^t \partial_x F(s, X_s) \cdot g(s, X_s)\, dW_s. \end{aligned}$$

- Next, we investigate the "$\partial_x^2 F^{(n)}$ term". Since

$$(\Delta X_n)^2 = \left(f^{(n)}\Delta t_n + g^{(n)}\Delta W_n\right)^2$$

we have

$$\begin{aligned} \frac{1}{2}\sum_{n=0}^{N-1} \partial_x^2 F^{(n)} \cdot (\Delta X_n)^2 &= \frac{1}{2}\sum_{n=0}^{N-1} \partial_x^2 F^{(n)} \cdot (f^{(n)})^2(\Delta t_n)^2 && (2.13) \\ &\quad + \sum_{n=0}^{N-1} \partial_x^2 F^{(n)} \cdot f^{(n)} g^{(n)} \Delta t_n \Delta W_n && (2.14) \\ &\quad + \frac{1}{2}\sum_{n=0}^{N-1} \partial_x^2 F^{(n)} \cdot (g^{(n)})^2(\Delta W_n)^2. && (2.15) \end{aligned}$$

For the right-hand side of (2.13), we obtain

$$\left\| \sum_{n=0}^{N-1} \partial_x^2 F^{(n)} \cdot (f^{(n)})^2 (\Delta t_n)^2 \right\|_{L^2(d\mathbb{P})}^2 = \mathbb{E}\left[\left(\sum_{n=0}^{N-1} \partial_x^2 F^{(n)} \cdot (f^{(n)})^2 (\Delta t_n)^2 \right)^2 \right] \longrightarrow 0.$$

With the abbreviation $\alpha^{(n)} := \partial_x^2 F^{(n)} \cdot f^{(n)} g^{(n)}$ we obtain for the right-hand side of (2.14) that

$$\left\| \sum_{n=0}^{N-1} \alpha^{(n)} \Delta t_n \Delta W_n \right\|_{L^2(d\mathbb{P})}^2 = \mathbb{E}\left[\left(\sum_{n=0}^{N-1} \alpha^{(n)} \Delta t_n \Delta W_n \right)^2 \right] = \sum_{n=0}^{N-1} \sum_{m=0}^{N-1} \mathbb{E}\left(\alpha^{(n)} \alpha^{(m)} \Delta t_n \Delta t_m \Delta W_n \Delta W_m \right).$$

Since

$$\mathbb{E}\left(\alpha^{(n)} \alpha^{(m)} \Delta W_n \Delta W_m \right) = \mathbb{E}\left(\alpha^{(n)} \alpha^{(m)} \right) \mathbb{E}\left(\Delta W_n \Delta W_m \right) = 0$$

for $n < m$ and similar for $m < n$, only the terms with $n = m$ have to be considered, which yields

$$\left\| \sum_{n=0}^{N-1} \alpha^{(n)} \Delta t_n \Delta W_n \right\|_{L^2(d\mathbb{P})}^2 = \sum_{n=0}^{N-1} \mathbb{E}\left((\alpha^{(n)})^2 \right) (\Delta t_n)^2 \underbrace{\mathbb{E}\left[(\Delta W_n)^2 \right]}_{=\Delta t_n} \longrightarrow 0.$$

The third term (2.15), however, has a non-zero limit: We show that

$$\lim_{N\to\infty} \frac{1}{2} \sum_{n=0}^{N-1} \partial_x^2 F^{(n)} \cdot (g^{(n)})^2 \, (\Delta W_n)^2 = \frac{1}{2} \int_0^t \partial_x^2 F(s, X_s) \cdot \left(g(s, X_s) \right)^2 ds$$

which yields the strange additional term in the Itô formula. With the abbreviation $\beta^{(n)} = \frac{1}{2} \partial_x^2 F^{(n)} \cdot (g^{(n)})^2$ we have

$$\left\| \sum_{n=0}^{N-1} \beta^{(n)} \left((\Delta W_n)^2 - \Delta t_n \right) \right\|_{L^2(d\mathbb{P})}^2 = \mathbb{E}\left[\left(\sum_{n=0}^{N-1} \beta^{(n)} \left((\Delta W_n)^2 - \Delta t_n \right) \right)^2 \right]$$

$$= \mathbb{E}\left[\sum_{n=0}^{N-1} \sum_{m=0}^{N-1} \beta^{(n)} \beta^{(m)} \left((\Delta W_n)^2 - \Delta t_n \right) \left((\Delta W_m)^2 - \Delta t_m \right) \right].$$

For $n < m$ we have

$$\begin{aligned} &\mathbb{E}\left[\beta^{(n)} \beta^{(m)} \left((\Delta W_n)^2 - \Delta t_n \right) \left((\Delta W_m)^2 - \Delta t_m \right) \right] \\ &= \mathbb{E}\left[\beta^{(n)} \beta^{(m)} \left((\Delta W_n)^2 - \Delta t_n \right) \right] \underbrace{\mathbb{E}\left[\left((\Delta W_m)^2 - \Delta t_m \right) \right]}_{=0} = 0 \end{aligned}$$

and vice versa for $n > m$. Hence, only the terms with $n = m$ have to be considered, and we obtain

$$\left\| \sum_{n=0}^{N-1} \beta^{(n)} \left((\Delta W_n)^2 - \Delta t_n \right) \right\|_{L^2(d\mathbb{P})}^2 = \mathbb{E}\left[\sum_{n=0}^{N-1} \left(\beta^{(n)} \right)^2 \left((\Delta W_n)^2 - \Delta t_n \right)^2 \right]$$

$$= \sum_{n=0}^{N-1} \mathbb{E}\left[\left(\beta^{(n)}\right)^2\right] \mathbb{E}\left[\left((\Delta W_n)^2 - \Delta t_n\right)^2\right]$$

according to Exercise 5.

- With essentially the same arguments, it can be shown that

$$\lim_{N\to\infty} \frac{1}{2} \sum_{n=0}^{N-1} \partial_t^2 F^{(n)} \cdot (\Delta t_n)^2 = 0$$

$$\lim_{N\to\infty} \sum_{n=0}^{N-1} \partial_t \partial_x F^{(n)} \cdot \Delta t_n \Delta X_n = 0$$

and that the remainder term from the Taylor expansion can be neglected when the limit is taken. ■