A Primer on Stochastic Differential Equations

Pramook Khungurn

January 15, 2023

This note gives basic information on stochastic differential equations. The materials come primarity two books: [Evans, 2013], [Mörters and Peres, 2012], [Särkkä, 2012], and [Mikosch, 1998].

What got me interested in the subject was an attempt to understand recent works on deep generativ models, score-based models, in particular. I read a blog post by Yang Song [Song, 2021], and I found that this body of work involves the Langevin equation:

$$\mathrm{d}\mathbf{x} = \frac{1}{2}\nabla \log \pi(\mathbf{x})\,\mathrm{d}t + \mathrm{d}\mathbf{W}.$$

And I have to admit that I have no idea what this equation is about. This note is an attempt to understand the subject to the level that allows me to carry out further reading into the subject.

1 Introduction

- We study ordinary differential equations to be able to solve the initial value problem: find a function $\mathbf{x} : [0, \infty) \to \mathbb{R}^d$ that satisfies the quations

$$\begin{aligned}\frac{\mathrm{d}\mathbf{x}(t)}{\mathrm{d}t} &= \mathbf{b}(\mathbf{x}(t), t),\\ \mathbf{x}(0) &= \mathbf{x}_0,\end{aligned}$$

  where $\mathbf{b} : \mathbb{R}^d \times \mathbb{R} \to \mathbb{R}^d$ is a smooth, time-varying vector field, and $\mathbf{x}_0 \in \mathbb{R}^d$ is a point in $\mathbb{R}^d$.

- Because the vector field $\mathbf{b}$ is smooth, the trajectory of $\mathbf{x}$ would be smooth.

- In many applications such as molecular simulation and modeling of stock prices, however, the trajectories we want to model are not at all smooth: they are influenced by random noise. It is thus common to change the differential equation to

$$\frac{\mathrm{d}\mathbf{x}(t)}{\mathrm{d}t} = \mathbf{b}(\mathbf{x}(t), t) + B(\mathbf{x}(t), t)\boldsymbol{\xi}(t)$$

  where $B : \mathbb{R}^d \times \mathbb{R} \to \mathbb{R}^{n\times m}$ is a matrix-valued function, and $\boldsymbol{\xi}(t) : [0, \infty) \to \mathbb{R}^m$ is an m-dimensional "white noise" function.

- We will go into details about what a white noise is later, but it suffices to say that it corresponds to noise that is i.i.d. in time.

- It would turn out that our white noise is the time derivative of the **standard Brownian motion**:

$$\frac{\mathrm{d}\mathbf{W}(t)}{\mathrm{d}t} = \boldsymbol{\xi}(t).$$

  (I think we use the letter $\mathbf{W}$ because the standard Brownian motion has another name: the **Wiener process**.) We will go into more details on what a Brownian motion is later.

- Hence, we can rewrite the differential equation as

$$\mathrm{d}\mathbf{x}(t) = \mathbf{b}(\mathbf{x}(t), t)\,\mathrm{d}t + B(\mathbf{x}(t), t)\,\mathrm{d}\mathbf{W}(t) \tag{1}$$

  or simply

$$\mathrm{d}\mathbf{x} = \mathbf{b}(\mathbf{x}, t)\,\mathrm{d}t + B(\mathbf{x}, t)\,\mathrm{d}\mathbf{W},$$

  and this is a **stochastic differential equation** (SDE).

- The standard Brownian motion and the solution to the SDE above are, of course, functions. However, they are not deterministic, but random. Random functions are called **stochastic processes** in literature. We will of course go deeper into what they are later.

- Examples of stochastic differential equations include the **Langevin equation**:

$$\mathrm{d}\mathbf{x} = \frac{1}{2}\nabla \log \pi(\mathbf{x})\,\mathrm{d}t + \mathrm{d}\mathbf{W}.$$

  Here, in the context of probabilistic modeling, $\pi : \mathbb{R}^d \to [0,1]$ is a probability density function of the $\mathbf{x}$'s.

- Another example is from financial modeling. A stock price is often modeled as a **geometric Brownian motion**, which is governed by the following equation:

$$\mathrm{d}S(t) = \mu S(t)\,\mathrm{d}t + \sigma S(t)\,\mathrm{d}W(t).$$

  Here, $S : [0,\infty) \to \mathbb{R}$ is the scalar stock price, $\mu \in \mathbb{R}$ is called the **percentage drift**, $\sigma \in \mathbb{R}^+$ is called the **percentage volatility**, and $W : [0,\infty) \to \mathbb{R}$ is the 1D standard Brownian motion. This model is, in turn, used in the famous Black–Scholes formula.[1]

- To solve an SDE, we integrate both sides of Equation 1 to obtain

$$\mathbf{x}(t) = \mathbf{x}_0 + \int_0^t \mathbf{b}(\mathbf{x}(t), t)\,\mathrm{d}t + \int_0^t B(\mathbf{x}(t), t)\,\mathrm{d}\mathbf{W}(t).$$

- The non-obvious part is how to integrate with respect to $\mathrm{d}\mathbf{W}(t)$. This integral is neither the Riemann or Lebesgue integral, but a new type of integral called the **Itô integral**. It is a main object of study of this note.

- Lastly, we will also study how we solve SDEs numerically and will at least cover the Euler–Maruyama method.

2 Stochastic Processes

- In the rest of this note, we will be working with a probability space $(\Omega, \mathcal{F}, P)$ where Ω is the sample space, $\mathcal{F}$ is a σ-algebra on Ω, and P is the probability measure on $(\Omega, \mathcal{F})$.

- **Definition 1.** *A* **stochastic process** *is a collection* $\{\mathbf{X}(t) \in \mathbb{R}^d : t \geq 0\}$ *of random variables. For each point* $\omega \in \Omega$*, the mapping* $t \mapsto \mathbf{X}(t, \omega)$ *is called the* **sample path***.*

- From the above definition, we see that there are three ways to view a stochastic process.
  - When viewed as collection of random variables, we see it as a collection of (potentially correlated) random values on the real line.

[1] `https://en.wikipedia.org/wiki/Black%E2%80%93Scholes_model`

  - When viewed from the lens of the sample path, it becomes a random function.
  - We can also view $\mathbf{X}$ as a function that maps an ordered pair $(t, \omega) \in [0, \infty) \times \Omega$ to a point in $\mathbb{R}^d$.

- A stochastic process can be characterized by many of its aspects. One of these aspects is called its *finite-dimensional distribution.*

  Definition 2. *The* **finite-dimensional distributions (fidis)** *of the stochastic process* $\{\mathbf{X}(t) \in \mathbb{R}^d : t \geq 0\}$ *are the distributions of the finite dimensional vectors* $(\mathbf{X}(t_1), \mathbf{X}(t_2), \ldots, \mathbf{X}(t_n))$ *for all* $t_1, t_2, \ldots, t_n \geq 0$ *and every* $n \geq 1$.

- **Definition 3.** *A stochastic process* $\{\mathbf{X}(t) \in \mathbb{R}^d : t \geq 0\}$ *is called a* **Gaussian process** *if, for any* $0 \leq t_1 < t_2 < \ldots < t_k$, *the vector* $(\mathbf{X}(t_1), \mathbf{X}(t_2), \ldots, \mathbf{X}(t_k))$ *has a (multi-variate) Gaussian distribution. Equivalently, it is a Gaussian process if every linear combination* $\sum_{i=1}^{k} a_i \mathbf{X}(t_i)$ *is either identically zero or has a (multi-variate) Gausssian distribution.*

- Another way of characterizing a stochastic process is through its *dependence structure*, which is mainly concerns with how $\mathbf{X}(t)$ is dependent (or independent) of $\mathbf{X}(s)$ for $s \neq t$.

- **Definition 4.** *A stochastic process* $\{\mathbf{X}(t) \in \mathbb{R}^d : t \geq 0\}$ *is* **strictly stationary** *if the fidis are invariant under shifts of the time t:*

  $$\big(\mathbf{X}(t_1), \mathbf{X}(t_2), \ldots, \mathbf{X}(t_n)\big) \text{ has the same distribution as } \big(\mathbf{X}(t_1 + h), \mathbf{X}(t_2 + h), \ldots, \mathbf{X}(t_n + h)\big)$$

  for all possible choices of times $t_1, t_2, \ldots, t_n \geq 0$, $n \geq 1$, *and* h *such that* $t_1 + h, t_2 + h, \ldots, t_n + h \geq 0$.

- **Definition 5.** *Let* $X(t)$ *be a real-valued stochastic process with* $E[X^2(t)] < \infty$ *for all* $t \geq 0$. *The* **autocorrelation function of** X *is the function*

  $$r(t, s) = E[X(t)X(s)]$$

  defined for $t, s \geq 0$.

- **Definition 6.** *We call a stochastic process* $\{X(t) : t \geq 0\}$ **stationary in the wide sence** *the following two properties are satisfied the following properties.*

  1. *Its autocorrelation function depends only on the difference between the times. In other words,* $r(t, s) = c(t - s)$ *for some function* $c : \mathbb{R} \to \mathbb{R}$ *and for all* $t, s \geq 0$.
  2. *Its expectation is constant. That is,* $E[X(t)] = E[X(s)]$ *for all* $t, s \geq 0$

- We can also impose stationary properties on the increments of a process.

  Definition 7. *We say that* $\{\mathbf{X}(t) : t \geq 0\}$ *have* **stationary increments** *if*

  $$\mathbf{X}(t) - \mathbf{X}(s) \text{ has the same distribution as } \mathbf{X}(t + h) - \mathbf{X}(s + h)$$

  for all $t, s \geq 0$ *and* h *such that* $t + h, s + h \geq 0$.

  Definition 8. *We say that* $\{\mathbf{X}(t) : t \geq 0\}$ *have* **independent increments** *if*

  $$\mathbf{X}(t_2) - \mathbf{X}(t_1), \mathbf{X}(t_3) - \mathbf{X}(t_2), \ldots, \mathbf{X}(t_n) - \mathbf{X}(t_{n-1})$$

  are independent random variables for all choices of $0 \leq t_1 < t_2 < \cdots < t_n$ *and* $n \geq 1$.

3 Brownian Motion

3.1 Definition and Basic Properties

- **Definition 9.** *A stochastic process $\{W(t) \in \mathbb{R} : t \geq 0\}$ is called a* **Brownian motion** *starting at $x_0 \in \mathbb{R}$ if the following properties hold.*

  (1) $W(0) = x_0$.

  (2) The process has independent increments.

  (3) For all $t > s \geq 0$, the increment $W(t) - W(s)$ is normally distributed with expectation 0 and variance $t - s$. In other words, $W(t) - W(s) \sim \mathcal{N}(0, t-s)$.

  (4) The sample path $t \mapsto W(t, \omega)$ is continuous almost surely (i.e. with probability 1).

  When $x_0 = 0$, we call it a **standard Brownian motion**.

- **Theorem 10 (Weiner 1923).** *The standard Brownian motion exists.*

  Proof sketch. We present a construction by Lévy and Ciesielski. We first construct the standard Brownian motion on the interval $[0, 1]$. Then, the Brownian motion can be extended to $[0, \infty)$ by "tiling."

  We start by a family of $\{h_k(\cdot)\}_{k=0}^{\infty}$ of **Haar functions**, where each h_k has signature $[0, 1] \to \mathbb{R}$. The functions are defined as follows.

$$
\begin{aligned}
h_0(t) &= 1, \\
h_1(t) &= \begin{cases} 1, & t \in [0, 1/2] \\ -1, & t \in (1/2, 1] \end{cases}.
\end{aligned}
$$

  For $2^n \leq k < 2^{n+1}$,

$$
h_k(t) = \begin{cases} 1, & t \in [\frac{k-2^n}{2^n}, \frac{k+1/2-2^n}{2^n}] \\ -1, & t \in [\frac{k+1/2-2^n}{2^n}, \frac{k+1-2^n}{2^n}] \\ 0, & \text{otherwise} \end{cases}
$$

  We have that $\{h_k(\cdot)\}_{k=0}^{\infty}$ is an orthonormal basis of the set $L^2(0, 1)$ of functions $f : \mathbb{R} \to \mathbb{R}$ such that $\int_0^1 |f(x)|^2 \, \mathrm{d}x$ is finite.

  From the Harr functions, define the **Schauder functions** as

$$
s_k(t) = \int_0^t h_k(u) \, \mathrm{d}u
$$

  for $t \in [0, 1]$. The graph of s_k is a tent of height $2^{-n/2-1}$ on the interval $[\frac{k-2^n}{2^n}, \frac{k+1-2^n}{2^n}]$.

  Let $\{A_k\}_{k=0}^{\infty}$ be a sequence of independent $\mathcal{N}(0, 1)$ random variables. We can define

$$
W(t, \omega) = \sum_{k=0}^{\infty} A_k(\omega) s_k(t),
$$

  and it can be shown that this function has all the properties of the Brownian motion. □

- The existence of the standard Brownian motion implies the existence of all other Brownian motions. Now that we know that they exist, let's examine some of their properties.

- For any $t > 0$, we have that $W(t) = x_0 + W(t) - W(0)$. Hence, $W(t)$ is distributed according to $\mathcal{N}(x_0, t)$ for any $t > 0$. (For $t = 0$, we may say that $W(0)$ is a Gaussian distribution with mean x_0 and variance 0.) As a result, a Brownian motion is a Gaussian process.
- Because $W(t) \sim \mathcal{N}(x_0, t)$, we have that $E[X(t)] = x_0$ for all $t \geq 0$. So, a Brownian motion has constant mean, and the mean of the standard Brownian motion is always 0.
- Property (3) in Definition 9 implies that a Brownian motion has stationary increments.
- **Lemma 11.** *If $W(t)$ is the standard Brownian motion, we have that $E[W(t)] = 0$, $E[W^2(t)] = t$, and*

  $$E[W(t)W(s)] = t \wedge s = \min(t, s)$$

  for all $t, s \geq 0$.

  Proof. Note that, because $W(\cdot)$ is the standard Brownian motion, we have that $W(t) \sim \mathcal{N}(0, t)$. So, obviously, $E[W(t)] = 0$. Moreover, we have that

  $$E[W^2(t)] = E[W^2(t)] - 0 = E[W^2(t)] - E[(W(t))^2] = \mathrm{Var}(W(t)) = t.$$

  Now, Assume $t \geq s \geq 0$. We have that.

  $$\begin{aligned} E[W(t)W(s)] &= E[(W(s) + W(t) - W(s))W(s)] \\ &= E[W^2(s)] + E[(W(t) - W(s))W(s)] \\ &= s + E[(W(t) - W(s))(W(s) - W(0))] \\ &= s + E[W(t) - W(s)]E[W(s) - W(0)] \\ &= s + (E[W(t)] - E[W(s)])(E[W(s)] - E[W(0)]) \\ &= s = \min(t, s) = t \wedge s \end{aligned}$$

  as required. □

- So, while a Brownian motion has constant mean, it is not stationary in the wide sense because the autocorrelation function $r(t, s) = E[W(t)W(s)]$ is not a function of t_s.
- It can be shown that, if $X(t)$ is a Gaussian process such that $E[X(t)X(s)] = t \wedge s$ and $\mathrm{Var}(X(0)) = 0$, then it is a Brownian motion.
- We can extend the standard Brownian motion in $\mathbb{R}$ to one in $\mathbb{R}^d$.

  Definition 12. *A stochastic process $\{\mathbf{W}(t) \in \mathbb{R}^d : t \geq 0\}$ where $\mathbf{W}(t) = (W_1(t), W_2(t), \ldots, W_n(t))$ is an n-**dimensional Brownian motion** if it satisfies the following conditions.*

  - *For each $k = 1, 2, \ldots, n$, we have that $W_k(t)$ is a one-dimensional Brownian motion.*
  - *The σ-algebras $\mathcal{W}_k = \sigma\left(\bigcup_{t \geq 0} \sigma(W_k(t))\right)$, for $k = 1, 2, \ldots, n$, are independent of one another.*

- For an n-dimensional Brownian motion, we have that

  $$\begin{aligned} E[W_k(t)W_l(t)] &= (t \wedge s)\delta_{kl} \\ E[(W_k(t) - W_k(s))(W_l(t) - W_l(s))] &= (t - s)\delta_{kl} \end{aligned}$$

  where

  $$\delta_{kl} = \begin{cases} 1, & k = l \\ 0, & k \neq l \end{cases}$$

  is the Kronecker delta function.

3.2 Self Similarity and Non-Differentiability

- **Definition 13.** *We say a stochastic process* $\{\mathbf{X}(t) : t \geq 0\}$ *is* H*-self-similar for some* $H > 0$ *if its fidis satisfy the condition*

  $$(c^H \mathbf{X}(t_1), c^H \mathbf{X}(t_2), \ldots, c^H \mathbf{X}(t_n)) \text{ has the same distribution as } (\mathbf{X}(ct_1), \mathbf{X}(ct_2), \ldots, \mathbf{X}(ct_n))$$

  for every $c > 0$ *and any choice of* $t_1, t_2, \ldots, t_n \geq 0$ *and* $n \geq 1$.

- Self-similarity means that the properly scaled patterns of a sample path in any small or large time interval have a simlar shape.

- **Theorem 14.** *A Brownian motion is* 0.5*-self-similar.*

- **Theorem 15.** *Let* $\{X(t) : t \geq 0\}$ *be a stochastic process. If* $X(t)$ *is* H*-self-similar for some* $H \in (0, 1)$ *and has stationary increments, then, for any fixed* t_0,

  $$\limsup_{t \to t_0^+} \frac{|X(t) - X(t_0)|}{t - t_0} = \infty$$

  almost surely. In other words, sample paths of H*-self-similar stochastic processes are nowhere differentiable almost surely.*

- This means that a path of a Brownian motion is not differentiable everywhere almost surely.

3.3 White Noise

- We said earlier that it turned out that the derivative of the Brownian motion is the "white noise." In one dimension, this means that

  $$\mathrm{d}W(t)/\mathrm{d}t = \xi(t)$$

  where $\xi(t)$ is the one-dimensional white noise.

- However, we learned that the sample path $t \mapsto W(t, \omega)$ is not differentiable at any $t \geq 0$ with probability 1. So, the derivative does not really exist. Let us suspend our disbelief and derive some of $\xi(t)$'s properties though.

- The first is that $E[\xi(t)] = 0$ for all t. This is because

  $$\xi(t) = \lim_{h \to 0} \frac{W(t+h) - W(t)}{h}.$$

  We know that $W(t+h) - W(t) \sim \mathcal{N}(0, h)$, and so $E\Big[\frac{W(t+h)-W(h)}{h}\Big] = 0$ for all h. As a result, $E[\xi(t)]$ should be 0 as well.

- We can also show that

  $$E[\xi(t)\xi(s)] = \delta(t - s)$$

  where δ is the Dirac delta function.

  The "proof" is as follows. Fix $h > 0$ and $t > 0$. Define

  $$\begin{aligned}
  \phi_h(s) &= E\bigg[\bigg(\frac{W(t+h) - W(t)}{h}\bigg)\bigg(\frac{W(s+h) - W(s)}{h}\bigg)\bigg] \\
  &= \frac{1}{h^2}\Big(E[W(t+h)W(s+h)] - E[W(t+h)W(s)] - E[W(t)W(s+h)] + E[W(t)W(s)]\Big) \\
  &= \frac{1}{h^2}\Big((t+h) \wedge (s+h) - (t+h) \wedge s - t \wedge (s+h) + t \wedge s\Big).
  \end{aligned}$$

  There are 4 cases.

1. If $t + h < s$, then $\phi_s(s) = (t + h - t + h - t + t)/h^2 = 0$.
2. If $t \leq s < t + h$, then $\phi_s(s) = (t + h - s - t + t)/h^2 = (h + t - s)/h^2$.
3. If $t - h \leq s < t$, then $\phi_s(s) = (s + h - s - t + s)/h^2 = (h - t + s)/h^2$.
4. If $s < t - h$, then $\phi_s(s) = (s + h - s - s + h + s)/h^2 = 0$.

As a result, $\phi_h(s)$ is a tent function of height $1/h$ over the interval $[t - h, t + h]$. It follows that $\int \phi_h(s)\ \mathrm{d}s = 1$, and $\lim_{h\to 0} \phi_h(s) = 0$ when $s \neq t$. As a result, $\lim_{h\to 0} \phi_h(s) = \delta(t - s)$ where δ is the Direct delta function.

- Because $\xi(t)$ has constant mean and its autocorrelation function depends only on $t - s$, we have that $\xi(t)$ is stationary in the wide sense.

- **Definition 16.** *Let $X(t)$ be a stochastic process that is stationary in the wide sense with autocorrelation function $c(\cdot)$. The process's* **spectral density** *is the Fourier transform of the autocorrelation function:*

$$f(\lambda) = \frac{1}{2\pi} \int_{-\infty}^{\infty} e^{-i\lambda t} c(t)\ \mathrm{d}t$$

for any $\lambda \in \mathbb{R}$.

- $\xi(t)$'s spectral density is given by

$$f(\lambda) = \frac{1}{2\pi} \int_{-\infty}^{\infty} e^{-i\lambda t} \delta(t)\ \mathrm{d}t = \frac{1}{2\pi}$$

for all λ. This is why it is called "white" noise.

3.4 Unbounded Variation

- **Definition 17.** *A* **partition** *$\mathcal{P}$ of the interval $[a, b]$ is a set of real numbers where*

$$\{a = t_0 < t_1 < t_2 < \cdots < t_k = b\}.$$

The **mesh size** *of $\mathcal{P}$ is given by*

$$|\mathcal{P}| = \max_{0<j\leq k} |t_k - t_{k-1}|.$$

- **Definition 18.** *Consider a sequence of partitions $\{\mathcal{P}^{(n)}\}_{n=1}^{\infty}$ where*

$$\mathcal{P}^{(n)} = \{a = t_1^{(n)} < t_2^{(n)} < \cdots < t_{k(n)}^{(n)} = b\}.$$

We call the sequence **nested** *if $\mathcal{P}^{(n)}$ is a proper subset of $\mathcal{P}^{(n+1)}$ for all n. In other words, at least one more point is added to each subsequent partition.*

- **Definition 19.** *Let $f : \mathbb{R} \to \mathbb{R}$ be a real-valued function. The* **variation** *of f on the $[a, b]$ is given by*

$$V_{[a,b]}^{(1)}(f) = \lim_{\substack{n\to\infty \\ |\mathcal{P}^{(n)}|\to 0}} \sum_{j=1}^{k(n)} |f(t_j^{(n)}) - f(t_{j-1}^{(n)})|.$$

The limit is taken over any nested sequence of partitions $\{\mathcal{P}^{(n)}\}_{n=1}^{\infty}$ such that $|\mathcal{P}^{(n)}| \to 0$ as $n \to \infty$. On the other hand, the **quadratic variation** *of f on $[a, b]$ is*

$$V_{[a,b]}^{(2)}(f) = \lim_{\substack{n\to\infty \\ |\mathcal{P}^{(n)}|\to 0}} \sum_{j=1}^{k(n)} \Big(f(t_j^{(n)}) - f(t_{j-1}^{(n)})\Big)^2.$$

- **Theorem 20.** *Let $W(t)$ be a Brownian motion. We have that*

$$V^{(2)}_{[a,b]}(W) = b - a$$
$$V^{(1)}_{[a,b]}(W) = \infty$$

for any $0 \leq a < b$. In other words, a Brownian motion has finite quadratic variation but infinite variation.

- The fact that the paths of a Brownian motion are not differentiable and have infinite variation makes it difficult to apply Riemann integration to them. As a result, we need another way to define integrals that involve a Brownian motion.

3.5 Markov Properties

- Consider a stochastic process $\{\mathbf{X}(t) : t \geq 0\}$. Informally, we say that the process has **Markov property** if, when we want to predict the future $\{\mathbf{X}(t) : t \geq s\}$ for some $s \geq 0$ using information from the past $\{X(t) : 0 \leq t \leq s\}$, then the only useful information is the value of $X(s)$.

- A process is called a **(time-homogeneous) Markov process** if starts afresh at any fixed time s. In other words, the time-shifted process $\{\mathbf{X}(s+t) : t \geq 0\}$ has the same distribution as the processed starting at $\mathbf{X}(s)$ at time 0.

- **Theorem 21 (Markov property).** *A Brownian motion is a Markov process. More precisely, let $\{\mathbf{W}(t) : t \geq 0\}$ be a Brownian motion starting at $\mathbf{x}_0 \in \mathbb{R}^d$, and let $s > 0$. Then, the process $\{\mathbf{W}(t+s) - \mathbf{W}(s) : t \geq 0\}$ is again a Brownian motion starting at the origin, and it is independent of the process $\{\mathbf{W}(t) : 0 \leq t \leq s\}$.*

- **Definition 22.** *A* **filtration** *on a probability space $(\Omega, \mathcal{F}, P)$ is a family $\{\mathcal{F}(t) : t \geq 0\}$ of σ-algebras such that $\mathcal{F}(s) \subseteq \mathcal{F}(t)$ for all $s < t$. In other words, it is a stream of increasingly more information.*

  A probabilty space together with a filtration is called a **the filtered probability space**.

  A stochastic process $\{X(t) : t \geq 0\}$ defined on a filtered probability space with filtration $\{\mathcal{F}(t) : t \geq 0\}$ is said to be **adapted to the filtration** *if $X(t)$ is $\mathcal{F}(t)$-measurable for any $t \geq 0$.*

- The Markov property means that a Brownian motion is started anew at each deterministic time instance. However, this is also true for a class of random times called "stopping time."

- Intuitively, a stopping time T is a random variable such that we can deduce whether $T \leq t$ by only observing the path of the stochastic process up to time t. In other words, the set $\{T \leq t\}$ (which is an abbreviation for $\{\omega \in \Omega : T(\omega) \leq t\}$) is an event in $\mathcal{F}(t)$.

  Definition 23. *A random variable T with values in $[0, \infty)$ defined on a probability space with filtration $\{\mathcal{F}(t) : t \geq 0\}$ is called a* **stopping time** *with respect to the filtration if $\{T \leq t\} \in \mathcal{F}(t)$ for all $t \geq 0$.*

- **Theorem 24 (Strong Markov property).** *For every almost surely finite stopping time T, the process $\{\mathbf{W}(T+t) - \mathbf{W}(T) : t \geq 0\}$ is a Brownian motion starting at $\mathbf{0}$ independent of $\mathcal{F}^+(T)$.*

3.6 Martingale Properites

- A stochastic process $\mathbf{X}$ is always adapted to the **natural filtration generated by X**

$$\mathcal{F}(t) = \sigma\bigg(\bigcup_{0 \leq s \leq t} \sigma(\mathbf{X}(s)) \bigg).$$

- **Definition 25.** *A stochastic process* $\{\mathbf{X}(t) : t \geq 0\}$ *is called a* **(continuous-time) martingale with respect to the filtration** $\{\mathcal{F}(t) : t \geq 0\}$ *if the following conditions are satisfied.*

  (1) $E\big[\|\mathbf{X}(t)\|_1\big] < \infty$ *for all* $t \geq 0$. *(Here,* $\|\cdot\|_1$ *denote the* 1*-norm.)*

  (2) $\mathbf{X}$ *is adapted to* $\{\mathcal{F}(t) : t \geq 0\}$.

  (3) $E[\mathbf{X}(t)|\mathcal{F}(s)] = \mathbf{X}(s)$ *for all* $0 \leq s < t$.

- **Theorem 26.** *A martingale's expectation is constant.*

  Proof. Let $t > 0$.

  $$E[\mathbf{X}(0)] = E[E[\mathbf{X}(t)|\mathcal{F}(0)]] = E[\mathbf{X}(t)].$$

  We are done. □

- **Theorem 27.** *A Brownian motion is a martingale with respect to its natural filtration.*

  Proof. Let $0 \leq s < t$. We have that

  $$\begin{aligned} E[\mathbf{W}(t)|\mathcal{F}(s)] &= E[\mathbf{W}(t) - \mathbf{W}(s) + \mathbf{W}(s)|\mathcal{F}(s)] \\ &= E[\mathbf{W}(t) - \mathbf{W}(s)|\mathcal{F}(s)] + E[\mathbf{W}(s)|\mathcal{F}(s)]. \end{aligned}$$

  According to the Markov property (Theorem 21), $\mathbf{W}(t) - \mathbf{W}(s)$ is independent of $\mathcal{F}(s)$, we have that

  $$E[\mathbf{W}(t) - \mathbf{W}(s)|\mathcal{F}(s)] = E[\mathbf{W}(t) - \mathbf{W}(s)] = E[\mathbf{W}(t)] - E[\mathbf{W}(s)] = 0.$$

  Now, $E[\mathbf{W}(s)|\mathcal{F}(s)] = E[\mathbf{W}(s)]$. As a result, $E[\mathbf{W}(t)|\mathcal{F}(s)] = \mathbf{W}(s)$. □

4 Stochastic Integrals

- Recall that our end goal is to solve the initial value problem

  $$\begin{aligned} \mathrm{d}\mathbf{X}(t) &= \mathbf{b}(\mathbf{X}(t), t)\,\mathrm{d}t + B(\mathbf{X}(t), t)\,\mathrm{d}\mathbf{W}(t) \\ \mathbf{X}(0) &= \mathbf{x}_0 \end{aligned}$$

  where $\mathbf{X}(t)$ is a stochastic process, and $\mathbf{W}(t)$ is the standard Brownian motion in $\mathbb{R}^d$.

- We said in the introduction that the solution would be

  $$\mathbf{X}(t) = \mathbf{x}_0 + \int_0^t \mathbf{b}(\mathbf{X}(t), t)\,\mathrm{d}t + \int_0^t B(\mathbf{X}(t), t)\,\mathrm{d}\mathbf{W}(t)$$

  As a result, we need to define the integral of the form

  $$\int_0^t \mathbf{G}(t)\,\mathrm{d}\mathbf{W}(t)$$

  where $\mathbf{G}$ is a stochastic process.

- Note that, in real analysis, there is a way to define integrals of the form

  $$\int_a^b f(x)\,\mathrm{d}\alpha(x)$$

  where f and α are both functions. This is the **Riemann–Stieltjes integral**, which is defined as follows.

  - We start with a sequence of nested partitions $\{\mathcal{P}^{(n)}\}_{n=1}^{\infty}$ of $[a,b]$.
  - Given a partition $\mathcal{P}^{(n)}$, we define the Riemann–Stieltjes sums:

$$\mathcal{S}(f,\alpha,\mathcal{P}^{(n)}) = \sum_{j=1}^{k(n)} f(\tau_j^{(n)})[\alpha(t_j^{(n)}) - \alpha(t_{j-1}^{(n)})]$$

    where $\tau_j^{(n)} \in [t_{j-1}^{(n)}, t_j^{(n)}]$.
  - The Riemann–Stieltjes integral is the defined as

$$\int_a^b f(x)\,\mathrm{d}\alpha(x) = \lim_{\substack{n\to\infty \\ |\mathcal{P}^{(n)}|\to 0}} \mathcal{S}(f,\alpha,\mathcal{P}^{(n)})$$

    provided that the limit exists.

- If α is differentiable on $[a,b]$, we have that

$$\int_a^b f(x)\,\mathrm{d}\alpha(x) = \int_a^b f(x)\alpha'(x)\,\mathrm{d}x.$$

  Hence, the integral $\int \mathbf{G}(t)\,\mathrm{d}\mathbf{W}(t)$ that we want to compute would correspond to $\int \mathbf{G}(t)\boldsymbol{\xi}(t)\,\mathrm{d}t$, and the form $\mathbf{G}(t)\boldsymbol{\xi}(t)$ is the one we started our modeling with.

- It is tempting to use the Riemann–Stieltjes integral to define the stochastic integral. However, the existence of the Riemann–Stieltjes integrals rests on the premise that $\mathcal{S}(f,\alpha,\mathcal{P}^{(n)})$ does not change based on the choice of $\tau_j^{(n)}$ as we take the limit. This is true for deterministic functions, but is not true for the Brownian motion.

- Before we go ahead to demonstrate that the value of $\mathcal{S}(f,\alpha,\mathcal{P}^{(n)})$ depends on the particular choise of $\tau_j^{(n)}$ we make, let us remind ourselves of the notion of convergence or random variables that we will use.

  Definition 28. *Let $(\Omega,\mathcal{F},P)$ be a probability space. Let $X:\Omega\to\mathbb{R}$ be a real-value random variable. The* **$\mathcal{L}^p$-norm of X** *is given by*

$$\|X\|_p = \Big(E[|X|^p]\Big)^{1/p} = \bigg(\int_\Omega |X|^p\,\mathrm{d}P\bigg)^{1/p}.$$

  Definition 29. *Let $\mathcal{L}^p(\Omega,\mathcal{F},P)$ denote the set of random variables X such that $\|X\|_p<\infty$.*

  Definition 30. *Let $\{X_n\}_{n=1}^{\infty}$ be a sequence of random variables in $\mathcal{L}^p(\Omega,\mathcal{F},P)$. Let X be another random variable in $\mathcal{L}^p(\Omega,\mathcal{F},P)$. We say that $\{X_n\}_{n=1}^{\infty}$* **converges to X in $\mathcal{L}^p$** *if*

$$\lim_{n\to\infty} \|X_n - X\|_p = 0.$$

  We write $X_n \xrightarrow{\mathcal{L}^p} X$ to denote this fact.

- With the notion of convergence defined, one can show that, for any $0\le\lambda\le 1$, if we pick

$$\tau_j^{(n)} = (1-\lambda)t_{j-1}^{(n)} + \lambda t_j^{(n)},$$

  then

$$\mathcal{S}(W,W,\mathcal{P}^{(n)}) \xrightarrow{\mathcal{L}^2} \frac{(W(T))^2}{2} + \Big(\lambda - \frac{1}{2}\Big)T$$

  where $\{\mathcal{P}^{(n)}\}_{n=1}^{\infty}$ is a sequence of nested partition of $[0,T]$ such that $|\mathcal{P}^{(n)}|\to 0$, and W is the standard Brownian motion. See the proof in the appendix (Proposition 64).

- Itô's definition of stochastic integral uses $\lambda = 0$. So,

$$\int_0^T W \, \mathrm{d}W = \frac{W(T)^2}{2} - \frac{T}{2}.$$

This shows that stochastic integrals are different from deterministic integrals. Who could have expected the $-T/2$ term to show up?

4.1 One-Dimensional Itô Integral

- The Itô integral is defined using a combination of three techniques.
  (a) Integration with respect to a function as done in the Riemann–Stieltjes integral.
  (b) Approximation by simple functions as used in the Lebesgue integral.
  (c) Fixing $\tau_j^{(n)} = t_j^{(n)}$ (i.e., $\lambda = 0$) as discussed above when defining the integral of simple functions.

  Using (a) and (b) together gets you the "Lebesgue–Stieltjes integral," adding (c) yields the Itô integral.

- First, however, we need to discuss the class of stochastic processes G upon which the integral $\int G \, \mathrm{d}W$ can be defined.

- **Definition 31.** *A stochastic process* $\mathbf{X} : [0, \infty) \times \Omega \to \mathbb{R}^d$ *defined on a probability space* $(\Omega, \mathcal{F}, P)$ *is called* **progressively measurable** *with respect to the filtration* $\{\mathcal{F}(t) : t \geq 0\}$ *if, for each* $t \geq 0$*, the mapping* $\mathbf{X} : [0, t] \times \Omega \to \mathbb{R}^d$ *is measurable with respect to the* σ*-algebra* $\mathcal{B}([0, t]) \times \mathcal{F}(t)$.

- Note that being progressively measurable implies being adapted.

- **Lemma 32.** *Any process which is adapted and is either right or left continuous is progressively measurable.*

- We then proceed with the definition of simple functions in the context of stochastic processes.

  Definition 33. *A* **step process on** $[0, T]$ **adapted to filteration** $\{\mathcal{F}_t : t \geq 0\}$ *is a real-valued stochastic process* $H : [0, T] \times \Omega \to \mathbb{R}$ *of the form*

  $$H(t, \omega) = \sum_{i=1}^{k} A_i(\omega) \mathbb{1}_{[t_i, t_{i+1})}(t)$$

  where

  – $0 = t_1 \leq t_2 \leq \cdots \leq t_{k+1} = T$,
  – A_i *is a random variable that is* $\mathcal{F}(t_i)$*-measurable, and*
  – $\mathbb{1}_{[t_{i-1}, t_{i+1})}(t)$ *is the indicator function of the interval* $[t_i, t_{i+1})$*. That is,*

  $$\mathbb{1}_{[t_i, t_{i+1})}(t) = \begin{cases} 1, & t_i \leq t < t_{i+1} \\ 0, & \textit{otherwise} \end{cases}.$$

  Note that the path of a step process takes only finitely many values.

- Because a step process is adapted to $\{\mathcal{F}(t) : t \geq 0\}$ and is left continous, it is progressively measurable.

- **Definition 34.** *Let* H *be a real-valued step process on* $[0, T]$ *adapted to the natural filtration* $\{\mathcal{F}(t) : t \geq 0\}$ *of the standard Brownian motion* W*. Define the* **Itô integral of** H **with respect to** W *to be*

  $$\int_0^T H(t) \, \mathrm{d}W(t) = \sum_{i=1}^{k} A_i(W(t_{i+1}) - W(t_i)).$$

- For a step process H on $[0,T]$, its Itô integral $\int_0^T H\,\mathrm{d}W$ is a random variable that maps a member $\omega \in \Omega$ to a real number.

- **Definition 35.** *Let $\{H(t) : t \geq 0\}$ be a real-valued stochastic process. The* $\mathcal{L}^p$ **norm of** H **on** $[0,T]$ *is given by*

$$\|H\|_p = \left(E\bigg[\int_0^T |H(t)|^p\,\mathrm{d}t\bigg]\right)^{1/p}$$

where the integration on the RHS is the Lebesgue integral on the real line performed on a sample path of H with respect to the Borel measure on $\mathbb{R}$.

- **Definition 36.** *We denote by $\mathbb{L}^p(0,T)$ the set of all real-valued, progressively measurable stochastic process H such that $\|H\|_p < \infty$.*

- **Definition 37.** *For all constants, $a, b \in \mathbb{R}$ and for all step processes $G, H \in \mathbb{L}^2(0,T)$, the following identities hold.*

$$\int_0^T aG + bH\,\mathrm{d}W = a\int_0^T G\,\mathrm{d}W + b\int_0^T H\,\mathrm{d}W.$$

$$E\bigg[\int_0^T G\,\mathrm{d}W\bigg] = 0.$$

$$E\bigg[\bigg(\int_0^T G\,\mathrm{d}W\bigg)^2\bigg] = E\bigg[\int_0^T G^2\,\mathrm{d}t\bigg]$$

- The third identity above is very important. It gives a sense that $(\mathrm{d}W)^2 = \mathrm{d}t$, which agrees with the fact that $E[W^2(t)] = t$. This will show up again when we discuss the Itô's chain rule.

- More over, it implies two more facts.

Proposition 38. *If H be a step process in $\mathbb{L}^2(0,T)$, then $\int_0^T H\,\mathrm{d}W \in \mathcal{L}^2(\Omega, \mathcal{F}, P)$.*

Proof. We have that

$$E\bigg[\bigg(\int_0^T H\mathrm{d}W\bigg)^2\bigg] = E\bigg[\int H^2\,\mathrm{d}t\bigg] = \|H\|_2^2 < \infty.$$

This implies that

$$\bigg\|\int_0^T H\mathrm{d}W\bigg\|_2 = \left(E\bigg[\bigg(\int_0^T H\mathrm{d}W\bigg)^2\bigg]\right)^{1/2} < \infty$$

as required. □

Proposition 39. *If G and H are step processes in $\mathbb{L}^2(0,T)$, then*

$$\bigg\|\int G\,\mathrm{d}W - \int H\,\mathrm{d}H\bigg\|_2 = \|G - H\|_2.$$

Proof. We have that

$$\begin{aligned}
\bigg\|\int G\,\mathrm{d}W - \int H\,\mathrm{d}H\bigg\|_2 &= E\bigg[\bigg(\int_0^T G\,\mathrm{d}W - \int_0^T H\,\mathrm{d}W\bigg)^2\bigg] \\
&= E\bigg[\bigg(\int_0^T (G-H)\,\mathrm{d}W\bigg)^2\bigg] \\
&= E\bigg[\int_0^T (G-H)^2\,\mathrm{d}t\bigg] \\
&= \|G - H\|_2^2
\end{aligned}$$

as required. $\square$

- The definition of the Lebegue integral relies on the fact that every measurable function has a sequence of simple functions that converges to it. The construcion involes a series of increasing and non-negative step functions, which converges as a result of the monotone convergence theorem. The construction of the Itô integral, however, typically uses a different notion of convergence: convergence in the $\mathcal{L}^2$ norm.

- **Lemma 40.** *If $G \in \mathbb{L}^2(0,T)$, there exists a sequence of bounded step processes $\{H^{(n)}\}_{n=1}^{\infty}$ in $\mathbb{L}^2(0,T)$ such that $\lim_{n\to 0} \|G - H^{(n)}\|_2 = 0$.*

- **Definition 41.** *Let $G \in \mathbb{L}^2(0,T)$. Let $\{H^{(n)}\}_{n=1}^{\infty}$ be a sequence of bounded step processes in $\mathbb{L}^2(0,T)$ such that $\lim_{n\to 0} \|G - H^{(n)}\|_2 = 0$. We have that the Itô integrals $\int_0^T H^{(n)}\,\mathrm{d}W$, as random variables, would converge in $\mathcal{L}^2$ to a random variable. The* **Itô integral of** G **on** $[0,T]$ *is defined to that limit. In other words,*

$$\int_0^T G\,\mathrm{d}W = \lim_{n\to 0} \int_0^T H^{(n)}\,\mathrm{d}W.$$

- Note that the sequence $\left\{ \int_0^T H^{(n)}\,\mathrm{d}W \right\}_{n=1}^{\infty}$ is Cauchy sequence of random variables in $L^2(\Omega,\mathcal{F},P)$. Because $L^2(\Omega,\mathcal{F},P)$ is complete, $\int_0^T H^{(n)}\,\mathrm{d}W$ would converge to a random variable in $\mathcal{L}^2(\Omega,\mathcal{F},P)$, and we define this random variable to be $\int_0^T G\,\mathrm{d}W$. To repeat, the convergence here is the convergence in $\mathcal{L}^2$ of random variables. In other words, we have that $\int_0^T G\,\mathrm{d}W$ and $\int_0^T H^{(n)}\,\mathrm{d}W$ are random variables, and

$$\int_0^T H^{(n)}\,\mathrm{d}W \xrightarrow{\mathcal{L}^2} \int_0^T G\,\mathrm{d}W$$

or

$$\lim_{n\to\infty} E\left[\left(\int H^{(n)}\,\mathrm{d}W - \int G\,\mathrm{d}W \right)^2 \right] = 0.$$

- We have defined the Itô integrals for functions in $\mathbb{L}^2(0,T)$ whose every member G satisfies

$$E\left[\int_0^T G^2\,\mathrm{d}t \right] < \infty.$$

It is possible to extend the definition to $\mathcal{M}^2(0,T)$, which is the class of progressively measurable real-value stochastic processes such that

$$\int_0^T G^2\,\mathrm{d}t < \infty \text{ almost surely.}$$

Notice that $\mathbb{L}^2(0,T) \subseteq \mathcal{M}^2(0,T)$ because the probabily that $\int G^2\,\mathrm{d}t = \infty$ must be 0 if the expectation is finite.

- **Theorem 42.** *For all constants $a, b \in \mathbb{R}$ and for all $G, H \in \mathbb{L}^2(0,T)$, we have that*

$$\int_0^T (aG + bH)\,\mathrm{d}W = a\int_0^T G\,\mathrm{d}W + b\int_0^T H\,\mathrm{d}W,$$

$$E\left[\int_0^T G\,\mathrm{d}W \right] = 0,$$

$$E\left[\left(\int_0^T G\,\mathrm{d}W \right)^2 \right] = E\left[\int_0^T G^2\,\mathrm{d}t \right],$$

$$E\left[\left(\int_0^T G\,\mathrm{d}W \right) \left(\int_0^T H\,\mathrm{d}W \right) \right] = E\left[\int_0^T GH\,\mathrm{d}t \right].$$

4.2 Itô's Chain and Product Rules

- **Definition 43.** *Suppose that* $\{X(t) : t \geq 0\}$ *is a real-valued stochastic process satisfying*

$$X(r) = X(s) + \int_s^r F\,\mathrm{d}t + \int_s^r G\,\mathrm{d}W$$

for some $F \in \mathbb{L}^1(0,T)$, $G \in \mathbb{L}^2(0,T)$, *and all times* $0 \leq s \leq t \leq T$. *We say that* $X(t)$ *has a* **stochastic differential**

$$\mathrm{d}X = F\,\mathrm{d}t + G\,\mathrm{d}W$$

for $0 \leq t \leq T$.

- **Theorem 44 (Itô's chain rule).** *Suppose that* $X(t)$ *has a stochastic differential* $\mathrm{d}X = F\,\mathrm{d}t + G\,\mathrm{d}W$. *Let* $u : \mathbb{R} \times [0,T] \to \mathbb{R}$ *be a constinuous function that* $u(x,t)$ *has continuous partial derivatives*

$$u_t = \frac{\partial u}{\partial t},\ u_x = \frac{\partial u}{\partial x},\ \text{and } \mu_{xx} = \frac{\partial^2 u}{\partial x^2}.$$

Then, $Y(t) = u(X(t),t)$ *has a stochastic differential*

$$\begin{aligned}\mathrm{d}Y = \mathrm{d}u(X,t) &= u_t\,\mathrm{d}t + u_x\,\mathrm{d}X + \frac{1}{2}u_{xx}G^2\,\mathrm{d}t \\ &= \left(u_t + u_x F + \frac{1}{2}u_{xx}G^2\right)\mathrm{d}t + u_x G\,\mathrm{d}W.\end{aligned}$$

- Note that, for the conventional chain rule, we would have that

$$\mathrm{d}Y = u_x\,\mathrm{d}X + u_t\,\mathrm{d}t = u_x(F\,\mathrm{d}t + G\,\mathrm{d}W) + u_t\,\mathrm{d}t = (u_x F + u_t)\,\mathrm{d}t + G\,\mathrm{d}W.$$

The Itô's chain rule add the term $0.5u_{xx}G^2\,\mathrm{d}t$. To see why this is should be the case, let's expand $\mathrm{d}Y$ with Taylor expansion.

$$\mathrm{d}Y = u_x\,\mathrm{d}X + u_t\,\mathrm{d}t + \frac{1}{2!}\big(u_{xx}\,(\mathrm{d}X)^2 + u_{xt}\,\mathrm{d}X\mathrm{d}t + u_{tt}\,(\mathrm{d}t)^2\big) + \frac{1}{3!}O\big((\mathrm{d}X + \mathrm{d}t)^3\big) + \cdots$$

When working with differentials, we consider $(\mathrm{d}t)^2 = 0$ and $\mathrm{d}W\mathrm{d}t = 0$. However, as we learned earlier that $(\mathrm{d}W)^2$ should be equal to $\mathrm{d}t$ in a sense. So,

$$\begin{aligned}&u_{xx}\,(\mathrm{d}X)^2 + u_{xt}\,\mathrm{d}X\mathrm{d}t + u_{tt}\,(\mathrm{d}t)^2 \\ &= u_{xx}(F\,\mathrm{d}t + G\,\mathrm{d}W)^2 + u_{xt}(F\,\mathrm{d}t + G\,\mathrm{d}W)\mathrm{d}t + u_{tt}\,(\mathrm{d}t)^2 \\ &= u_{xx}(F^2\,(\mathrm{d}t)^2 + FG\,\mathrm{d}W\mathrm{d}t + G^2\,(\mathrm{d}W)^2) + u_{xt}(F\,(\mathrm{d}t)^2 + G\,\mathrm{d}W\mathrm{d}t) + u_{tt}\,(\mathrm{d}t)^2 \\ &= u_{xx}G^2\,(\mathrm{d}W)^2 = u_x x G^2\,\mathrm{d}t.\end{aligned}$$

For higher order terms, we will always have $\mathrm{d}W\mathrm{d}t$ or $(\mathrm{d}t)^2$, so all those terms will be 0. As a result,

$$\mathrm{d}Y = u_x\,\mathrm{d}X + u_t\,\mathrm{d}t + \frac{1}{2}u_x x\,\mathrm{d}t = \left(u_t + u_x F + \frac{1}{2}u_{xx}G^2\right)\mathrm{d}t + u_x G\,\mathrm{d}W.$$

- **Example 45.** *Let* $X = W$, *and* $u(x) = x^m$. *We have that* $u_t(W) = 0$, $u_x(W) = mW^{m-1}$, *and* $u_{xx}(W) = m(m-1)W^{m-2}$. *Moreover,* $F = 0$ *and* $G = 1$, *so the Itô's chain rule yields*

$$\begin{aligned}\mathrm{d}(W^m) = \mathrm{d}u(W) &= \left(u_t(W) + u_x(W)F + \frac{1}{2}u_{xx}(W)G^2\right)\mathrm{d}t + u_x(W)G\,\mathrm{d}W \\ &= \frac{1}{2}m(m-1)W^{m-2}\,\mathrm{d}t + mW^{m-1}\,\mathrm{d}W.\end{aligned}$$

- **Theorem 46 (Itô's product rule).** *Suppose*

$$dX_1 = F_1\,\mathrm{d}t + G_1\,\mathrm{d}W$$
$$dX_2 = F_2\,\mathrm{d}t + G_2\,\mathrm{d}W$$

for $0 \le t \le dT$*,* $F_i \in \mathbb{L}^1(0,T)$*, and* $G_i \in \mathbb{L}^2(0,T)$*. Then,*

$$\mathrm{d}(X_1X_2) = X_2\,\mathrm{d}X_1 + X_1\,\mathrm{d}X_2 + G_1G_2\,\mathrm{d}t.$$

The expression $G_1G_2\,\mathrm{d}t$ *is called the* **Itô correction term***.*

Again, we may "derive" the product rule through Taylor expansion and requiring that $(\mathrm{d}W)^2 = \mathrm{d}t$.

- The integrated version of the above expression gives the **Itô integration-by-parts formula.**

$$\int_s^r X_2\,\mathrm{d}X_1 = X_1(r)X_2(r) - X_1(s)X_2(s) - \int_r^s X_1\,\mathrm{d}X_2 - \int_r^s G_1G_2\,\mathrm{d}t.$$

- The Itô's chain rule can be generalized into one that involves multi-variable u.

Theorem 47. *Suppose that*

$$\mathrm{d}X_i = F_i\,\mathrm{d}t + G_i\,\mathrm{d}W$$

with $F_i \in \mathbb{L}^1(0,T)$*,* $G_i \in \mathbb{L}^2(0,T)$ *for* $i = 1,\dots,n$*. If* $u(x_1,x_2,\dots,x_n,t) : \mathbb{R}^d \times [0,T] \to \mathbb{R}$ *is a continuous function with continuous partial derivatives* u_t*,* u_{x_i}*, and* $u_{x_ix_j}$ *for* $i,j = 1,2,\dots,n$*, then*

$$\mathrm{d}u = u_t\,\mathrm{d}t + \sum_{i=1}^n u_{x_i}\,\mathrm{d}X_i + \frac{1}{2}\sum_{i,j} u_{x_ix_j}G_iG_j\,\mathrm{d}t.$$

4.3 Itô Integral in Higher Dimensions

- We use $\mathbf{W}(t) = (W_1(t), W_2(t), \dots, W_n(t))$ to denote n-dimensional Brownian motion. Here, each component is an independent one-dimensional Brownian motion.

- **Lemma 48.** *If* W_1 *and* W_2 *are independent one-dimensional Brownian motions, then*

$$\mathrm{d}(W_1W_2) = W_1\,\mathrm{d}W_2 + W_2\,\mathrm{d}W_1.$$

- **Lemma 49 (Itô's product rule with several Brownian motions).** *Suppose that*

$$dX_1 = F_1\,\mathrm{d}t + \sum_{k=1}^m G_{1k}\,\mathrm{d}W_k$$
$$dX_2 = F_2\,\mathrm{d}t + \sum_{k=1}^m G_{2k}\,\mathrm{d}W_k$$

where the W_k*'s are independent Brownian motions. Then,*

$$\mathrm{d}(X_1X_2) = X_1\,\mathrm{d}X_2 + X_2\,\mathrm{d}X_1 + \sum_{k=1}^m G_{1k}G_{2k}\,\mathrm{d}t.$$

- We let $\mathbb{L}^2_n(0,T)$ to denote the set of vector-valued functions $\mathbf{F} = (F_1, F_2, \ldots, F_n)$ where each F_i is a member of $\mathbb{L}^2(0,T)$. Also, let $\mathbb{L}^2_{n\times m}(0,T)$ denote the set of matrix value function

$$G = [G_{ij}] = \begin{bmatrix} G_{11} & G_{12} & \cdots & G_{1m} \\ G_{21} & G_{22} & \cdots & G_{2m} \\ \vdots & \vdots & \ddots & \vdots \\ G_{n1} & G_{n2} & \cdots & G_{nm} \end{bmatrix}$$

where each G_{ij} is a member of $\mathbb{L}^2(0,T)$.

- In this way, if $G \in \mathbb{L}^2_{n\times m}(0,T)$ and $\mathcal{W}$ is the m-dimensional Brownian motion, then

$$\int_0^T G \,\mathrm{d}\mathbf{W}$$

is an $\mathbb{R}^d$-value random variable whose i-th component is

$$\sum_{j=1}^m \int_0^T G_{ij} \,\mathrm{d}W_j.$$

- **Lemma 50.** *If $G \in \mathbb{L}^2_{n\times m}(0,T)$, then*

$$E\left[\int_0^T G \,\mathrm{d}\mathbf{W}\right] = \mathbf{0},$$

and

$$E\left[\left\|\int_0^T G(t) \,\mathrm{d}\mathbf{W}(t)\right\|^2\right] = E\left[\int_0^T \|G(t)\|^2 \,\mathrm{d}t\right]$$

where $\|G(t)\|$ is the Frobenius norm of the matrix:

$$\|G(t)\| = \left(\sum_{i,j} (G_{ij}(t))^2\right)^{1/2}.$$

- Let $\mathbf{X}(t) = (X_1(t), \ldots, X_n(t))$. When we write

$$\mathrm{d}\mathbf{X} = \mathbf{F}\,\mathrm{d}t + G\,\mathrm{d}\mathbf{W},$$

we mean that $\mathbf{X}$ is the stochastic process such that

$$\mathbf{X}(r) = \mathbf{X}(s) + \int_r^s \mathbf{F}\,\mathrm{d}t + \int_s^r G\,\mathrm{d}\mathbf{W},$$

which means

$$X_i(r) = X_i(s) + \int_s^r F_i(t)\,\mathrm{d}t + \sum_{j=1}^m \int_r^s G_{ij}\,\mathrm{d}W_j$$

or, when written with differentials,

$$\mathrm{d}X_i = F_i\,\mathrm{d}t + \sum_{j=1}^m G_{ij}\,\mathrm{d}W_j$$

for all $i = 1, 2, \ldots, n$.

- **Lemma 51 (Itô's chain rule in n-dimension).** *Suppose that*

$$\mathrm{d}\mathbf{X} = \mathbf{F}\,\mathrm{d}t + \mathbf{G}\,\mathrm{d}\mathbf{W}.$$

Let $u : \mathbb{R}^d \times [0,T]$ be a continuous function with continuous partial derivatives u_t, u_{x_i}, and $u_{x_i x_j}$ for $i, j = 1, 2, \ldots, n$. Then,

$$\mathrm{d}(u(\mathbf{X}(t),t)) = u_t\ \mathrm{d}t + \sum_{i=1}^{n} u_{x_i} \mathrm{d}X_i + \frac{1}{2}\left(\sum_{i=1}^{n}\sum_{j=1}^{n} u_{x_i x_j} \sum_{l=1}^{m} G_{il} G_{jl} \right) \mathrm{d}t.$$

- Note that the Itô chain rule corresponds to evaluating the expression

$$d(u(\mathbf{X},t)) = u_t\,\mathrm{d}t + \sum_{i=1}^{n} u_{x_i}\,\mathrm{d}X_i + \frac{1}{2}\sum_{i=1}^{n}\sum_{j=1}^{n} u_{x_i x_j}\,\mathrm{d}X_i \mathrm{d}X_j.$$

We then expand each $\mathrm{d}X_i$ into $F_i\,\mathrm{d}t + \sum_{k=1}^{m} G_{ij}\,\mathrm{d}W_j$, multiply everything out, and then eliminate terms according to the following rules:

$$\begin{aligned} (\mathrm{d}t)^2 &= 0, \\ \mathrm{d}W_i\,\mathrm{d}t &= 0, \\ \mathrm{d}W_i\,\mathrm{d}W_j &= \delta_{ij}\mathrm{d}t. \end{aligned}$$

5 Stochastic Differential Equations

- **Definition 52.** *Let $\mathbf{W}$ be the m-dimensinal Brownian motion and $\mathbf{X}_0$ be a random variable that is independent of $\mathbf{W}$. We say that an $\mathbb{R}^d$-valued stochastic process $\{\mathbf{X} : 0 \le t \le T\}$ is a solution of the differential equation*

$$\begin{aligned} \mathrm{d}\mathbf{X} &= \mathbf{b}(\mathbf{X},t)\,\mathrm{d}t + B(\mathbf{X},t)\,\mathrm{d}\mathbf{W} \\ \mathbf{X}(0) &= \mathbf{X}_0 \end{aligned}$$

for $0 \le t \le T$ if the following conditions are satisfied.

1. *$\mathbf{X}(t)$ is progressively measurable with respect the filtration $\{\mathcal{F}(t) : 0 \le t \le T\}$ where $\mathcal{F}(t)$ denotes $\sigma(\mathbf{X}_0, \mathbf{W}(s) : 0 \le s \le t)$, which is the σ-algebra generated by $\mathbf{X}_0$ and $\mathbf{W}(\cdot)$ up to time t.*
2. *$\mathbf{b}(\mathbf{X}(t),t) \in \mathbb{L}^1_n(0,T)$.*
3. *$B(\mathbf{X}(t),t) \in \mathbb{L}^2_{n\times m}(0,T)$.*
4. *For all time $0 \le t \le T$, we have that*

$$\mathbf{X}(t) = \mathbf{X}_0 + \int_0^t \mathbf{b}(\mathbf{X}(s),s)\,\mathrm{d}s + \int_0^t B(\mathbf{X}(s),s)\,\mathrm{d}\mathbf{W}(s)$$

almost surely.

5.1 Examples

- **Example 53.** *Suppose f, g are continuous functions t (not random variables). Consider the initial value problem*

$$\begin{aligned} \mathrm{d}X &= fX\,\mathrm{d}t + gX\,\mathrm{d}W, \\ X(0) &= 1. \end{aligned}$$

Then, the solution is

$$X(t) = \exp\left(\int_0^t f(s)\,\mathrm{d}s - \frac{1}{2}\int_0^t g^2(s)\,\mathrm{d}s + \int_0^t g(s)\,\mathrm{d}W(s)\right).$$

To check this, we take the time derivative. Take

$$Y(t) = \int_0^t f(s)\,\mathrm{d}s - \frac{1}{2}\int_0^t g^2(s)\,\mathrm{d}s + \int_0^t g(s)\,\mathrm{d}W(s),$$

and so

$$\mathrm{d}Y = f\,\mathrm{d}t - \frac{1}{2}g^2\,\mathrm{d}t + g\,\mathrm{d}W.$$

Using the Itô's chain rule with $u(Y) = e^Y = X$*, we have that* $u_t = 0$*,* $u_Y = e^Y = X$*,* $u_{YY} = e^Y = X$*, and*

$$\begin{aligned}
\mathrm{d}X = \mathrm{d}u(Y) &= u_t + u_Y\,\mathrm{d}Y + \frac{1}{2}u_{YY}(\mathrm{d}Y)^2 \\
&= X\left(f\,\mathrm{d}t - \frac{1}{2}g^2\,\mathrm{d}t + g\,\mathrm{d}W\right) + \frac{1}{2}X\left(f\,\mathrm{d}t - \frac{1}{2}g^2\,\mathrm{d}t + g\,\mathrm{d}W\right)^2 \\
&= X\left(f\,\mathrm{d}t - \frac{1}{2}g^2\,\mathrm{d}t + g\,\mathrm{d}W\right) + \frac{1}{2}Xg^2\,\mathrm{d}t \\
&= fX\,\mathrm{d}t + gX\,\mathrm{d}W.
\end{aligned}$$

- **Example 54 (Stock prices).** *Let* $S(t)$ *denote the price of a stock at time* t*. Recall from the introduction that we often model it with*

$$\mathrm{d}S = \mu S\,\mathrm{d}t + \sigma S\,\mathrm{d}W.$$

where μ *and* σ *are constants. Taking* $S(0) = 1$*, we have that the solution is given by:*

$$\begin{aligned}
S(t) &= \exp\left(\int_0^t \mu\,\mathrm{d}s - \frac{1}{2}\int_0^t \sigma^2\,\mathrm{d}s + \int_0^t \sigma\,\mathrm{d}W\right) \\
&= \exp\left(\mu t - \frac{\sigma^2}{2}t + \sigma W(t)\right) \\
&= \exp\left(\sigma W(t) + \left(\mu - \frac{\sigma^2}{2}\right)t\right).
\end{aligned}$$

Moreover, if $S(0) = s_0$*, one can easily check that the solution is*

$$S(t) = s_0 \exp\left(\sigma W(t) + \left(\mu - \frac{\sigma^2}{2}\right)t\right).$$

Let us compute $E[S(t)]$*. We have that*

$$S(t) = s_0 + \int_0^t \mu S\,\mathrm{d}s + \int_0^t \sigma S\,\mathrm{d}W.$$

So,

$$\begin{aligned}
E[S(t)] &= s_0 + E\left[\int_0^t \mu S\,\mathrm{d}s\right] + E\left[\int_0^t \sigma S\,\mathrm{d}W\right] \\
&= s_0 + \mu\int_0^T E[S(s)]\,\mathrm{d}s + E\left[\int_0^t \sigma S\,\mathrm{d}W\right].
\end{aligned}$$

By Theorem 42,

$$E\left[\int_0^t \sigma S \,\mathrm{d}W\right] = 0.$$

As a result,

$$E[S(t)] = s_0 + \mu \int_0^t E[S(s)]\,\mathrm{d}s.$$

Differentiating both sides with respect to t, we have that

$$\frac{\mathrm{d}E[S(t)]}{\mathrm{d}t} = \mu E[S(t)],$$

which implies that $E[S(t)] = s_0 e^{\mu t}$, which is the solution of the ODE $\mathrm{d}S = \mu S\,\mathrm{d}t$.

- **Example 55 (Ornstein–Uhlenbeck process).** *The Ornstein–Uhlenbeck equation*

  $$\frac{\mathrm{d}^2 Y}{\mathrm{d}t^2} = -b\frac{\mathrm{d}Y}{\mathrm{d}t} + \sigma\xi$$

  describes the motion of a particle under two forces: the damping force $-b\,\mathrm{d}Y/\mathrm{d}t$ and the random perturbation $\sigma\xi$ where ξ is the white noise.

  Let $X = \mathrm{d}Y/\mathrm{d}t$. We have that

  $$\begin{aligned}\frac{\mathrm{d}X}{\mathrm{d}t} &= -bX + \sigma\xi\\ \frac{\mathrm{d}X}{\mathrm{d}t} + bX &= \sigma\xi.\end{aligned}$$

  To get a solution, we solve it like a normal ODE. First, multiplying both sides by e^{bt}, we have

  $$\begin{aligned}e^{bt}\frac{X}{\mathrm{d}t} + be^{bt}X &= \sigma e^{bt}\xi\\ \frac{\mathrm{d}}{\mathrm{d}t}(e^{bt}X) &= \sigma e^{bt}\xi\\ e^{bt}X(t) &= X(0) + \sigma\int_0^t e^{bs}\xi(s)\,\mathrm{d}s\\ X(t) &= e^{-bt}X(0) + \sigma\int_0^t e^{-b(t-s)}\,\mathrm{d}W(s).\end{aligned}$$

  Because $X = \mathrm{d}Y/\mathrm{d}t$, we have that

  $$Y(t) = Y(0) + \int_0^t X(s)\,\mathrm{d}s,$$

  and we can expand this out to get an expression for $Y(t)$.

5.2 Properties of Solution

- The first thing we have to worry about stochastic different equations is whether a solution exist or not. We know that $\mathbf{X}$ must satisfy

  $$\mathbf{X}(t) = \mathbf{X}_0 + \int_0^t \mathbf{b}(\mathbf{X}(s), s)\,\mathrm{d}s + \int_0^t B(\mathbf{X}(s), s)\,\mathrm{d}\mathbf{W}(s).$$

  One of the problem with the above expression is that we don't know what $\mathbf{X}$ is, so we don't know whether $\int_0^t \mathbf{b}(\mathbf{X}(s), s)\,\mathrm{d}s$ and $\int_0^t B(\mathbf{X}(s), s)\,\mathrm{d}\mathbf{W}(s)$ are even well defined. The following theorem gives us a sufficient condition for the solution to exist.

- **Theorem 56.** *Let $\mathbf{b} : \mathbb{R}^d \times [0,T] \to \mathbb{R}^d$ and $B : \mathbb{R}^d \times [0,T] \to \mathbb{R}^{m \times n}$ be uniformly Lipschitz continuous, meaning that there exists a constant L such that*

$$\begin{aligned}
\|\mathbf{b}(\mathbf{x},t) - \mathbf{b}(\mathbf{y},t)\| &\leq L\|\mathbf{x} - \mathbf{y}\| \\
\|B(\mathbf{x},t) - B(\mathbf{y},t)\| &\leq L\|\mathbf{x} - \mathbf{y}\| \\
\|\mathbf{b}(\mathbf{x},t)\| &\leq L(1 + \|\mathbf{x}\|) \\
\|B(\mathbf{x},t)\| &\leq L(1 + \|\mathbf{x}\|)
\end{aligned}$$

for all $0 \leq t \leq T$, $\mathbf{x}, \mathbf{y} \in \mathbb{R}^d$. Let $\mathbf{X}_0$ be a random variable such that $E[\|\mathbf{X}_0\|^2] < \infty$ that is independent of the m-dimensional Brownian motion $\mathbf{W}(\cdot)$. Then, there exists a unique solution $\mathbf{X} \in \mathbb{L}^2(0,T)$ of the stochastic differential equation

$$\begin{aligned}
\mathrm{d}\mathbf{X} &= \mathbf{b}(\mathbf{X},t)\,\mathrm{d}t + B(\mathbf{X},t)\,\mathrm{d}\mathbf{W} \\
\mathbf{X}(0) &= \mathbf{X}_0
\end{aligned}$$

for $0 \leq t \leq T$. By "unique," we mean that, if $\widetilde{\mathbf{X}}$ is another solution of the SDE, $\mathbf{X}(t) = \widetilde{\mathbf{X}}(t)$ for all $0 \leq t \leq T$ almost surely.

- The solution can be found by using an algorithm called **Picard's iteration**, which is also used in the proof for existence and uniqueness of solution to deteriministic ODE.

  1. Start with $\mathbf{X}^{(0)} \leftarrow \mathbf{X}_0$.
  2. Having computed $\mathbf{X}^{(n)}$ in the previous iteration, compute the next esimate by

  $$\mathbf{X}^{(n+1)} \leftarrow \mathbf{X}_0 + \int_0^t \mathbf{b}(\mathbf{X}^{(n)}(s), s)\,\mathrm{d}s + \int_0^t B(\mathbf{X}^{(n)}(s), s)\,\mathrm{d}\mathbf{W}.$$

  The proof of the above theorem involves showing that the sequence $\{\mathbf{X}^{(n)}\}_{n=0}^{\infty}$ coverges in the L^2 norm to a function in $\mathbb{L}^2(0,T)$, which clearly solves the SDE.

- **Theorem 57.** *The probability density $p(\mathbf{X},t)$ of the solution of the SDE in Theorem 56 solves the partial differential equation*

$$\frac{\partial p(\mathbf{X},t)}{\partial t} = -\sum_{i=1}^{n} \frac{\partial}{\partial x_i}[b_i(\mathbf{X},t)p(\mathbf{X},t)] + \frac{1}{2}\sum_{i,j} \frac{\partial^2}{\partial X_i \partial X_j}\Big([B(\mathbf{X},t)(B(\mathbf{X},t))^T]_{ij}\, p(\mathbf{X},t)\Big).$$

  The equation above is called the **Fokker–Planck equation** *or the* **Kolmogorov forward equation**.

- **Theorem 58.** *Let $\mathbf{X}(t)$ be the solution of the SDE in Theorem 56. Let*

$$\begin{aligned}
\mathbf{m}(t) &= E[\mathbf{X}(t)], \\
C(t) &= E[(\mathbf{X}(t) - \mathbf{m}(t))(\mathbf{X}(t) - \mathbf{m}(t))^T].
\end{aligned}$$

  be the mean and the covariance matrix of the solution as a function of time. Then, these two functions are solutions to following differential equations

$$\begin{aligned}
\frac{\mathrm{d}\mathbf{m}}{\mathrm{d}t} &= E[\mathbf{b}(\mathbf{X},t)], \\
\frac{\mathrm{d}C}{\mathrm{d}t} &= E[\mathbf{b}(\mathbf{X},t)(\mathbf{X} - \mathbf{m})^T] + E[(\mathbf{X} - \mathbf{m})(\mathbf{b}(\mathbf{X},t))^T] + E[B(\mathbf{X},t)B(\mathbf{X},t)^T].
\end{aligned}$$

6 Numerical Solution to SDE

- Recall that we wish to solve

$$\begin{aligned} \mathrm{d}\mathbf{X} &= \mathbf{b}(\mathbf{X}, t)\,\mathrm{d}t + B(\mathbf{X}, t)\,\mathrm{d}\mathbf{W} \\ \mathbf{X}(0) &= \mathbf{X}_0 \end{aligned}$$

  where $\mathbf{X}_0$ is a random variable that is independent from the Brownian motion $\mathbf{W}$.

- The simplest numerical integration scheme is the **Euler–Maruyama method**, which is just the Euler method applied in a very straightforward way to SDE.

  The algorithm goes as follows.

  1. Divide the interval $[0, T]$ into K subintevals of equal width. Let us say that the width of each subinterval is Δt.
  2. Sample $\widehat{\mathbf{X}}[0] \sim p(\mathbf{X}_0)$.
  3. For $k = 1, 2, \ldots, K$, do the following.
     (a) Sample $\Delta\mathbf{W}[k] \sim \mathcal{N}(0, \Delta t\, I)$.
     (b) Compute

$$\widehat{\mathbf{X}}[k] \leftarrow \widehat{\mathbf{X}}[k-1] + \mathbf{b}(\widehat{\mathbf{X}}[k-1], (k-1)\Delta t)\Delta t + B(\widehat{\mathbf{X}}[k-1], (k-1)\Delta t)\Delta\mathbf{W}[k].$$

  We that the sequence $\widehat{\mathbf{X}}[0], \widehat{\mathbf{X}}[1], \ldots \widehat{\mathbf{X}}[K]$ should approximate $\mathbf{X}(0)$, $\mathbf{X}(\Delta t)$, $\ldots$, $\mathbf{X}(K\Delta t)$, respectively.

- More sophisticated methods include the Milstein method and the stochastic Runge–Kutta method, but we are not discussing them here in this note.

7 Girsanov's Theorem

- We have been working with a fixed probability measure P. This is used to define our standard Brownian motion and everything that follows from it. Naturally, if we change P to another probability measure Q, the standard Brownian motion would no longer be a Brownian motion because, say, $W(t)$ might not be distributed according to $\mathcal{N}(0, t)$ any more.

- Girsanov's theorem identifies a class of probability measures Q that are "equivalent" to P, and the standard Brownian motions with respect to those probability measures.

- **Definition 59.** *Let $(\Omega, \mathcal{F})$ be measurable spaces. Let P and Q be two probability measures on $\mathcal{F}$. We say that Q is* **absolutely continuous with respect to** *P if $P(E) = 0 \implies Q(E) = 0$ for all event $E \in \mathcal{F}$. We write $Q \ll P$ to denote this fact.*

- **Theorem 60 (Radon–Nikodym).** *Let P and Q be probability measures such that $Q \ll P$. Then, there exists a positive $\mathcal{F}$-measurable function f such that*

$$Q(E) = \int_E f\,\mathrm{d}P$$

  for all event $E \in \mathcal{F}$. The function f is uniquely determined P-almost everywhere. It is called the **Radon–Nikodym derivative of Q with respect to P** *and denoted by $\mathrm{d}Q/\mathrm{d}P$.*

- **Definition 61.** *Let P and Q be two probability measures defined on measurable space $(\Omega, \mathcal{F})$. We say that P and Q are* **equivalent** *if $P \ll Q$ and $Q \ll P$. (In other words, P and Q are equivalent if $P(E) = 0 \iff Q(E) = 0$ for all $E \in \mathcal{F}$)*

- **Theorem 62 (Girsanov's).** *Let $(\Omega, \mathcal{F}, P)$ be a probability space. Let W be the standard Brownian motion defined on $(\Omega, \mathcal{F}, P)$, and let $\{\mathcal{F}(t) : t \geq 0\}$ be its natural filtration. Let $q \neq 0$ be a real constant. Then, the stochastic process*
$$M(t) = \exp\left(- qW(t) - \frac{1}{2}q^2 t\right)$$
is a martingale with respect to $\{\mathcal{F}(t) : t \geq 0\}$. For any $T > 0$, the function
$$Q(E) = \int_E M(T, \omega)\, \mathrm{d}P(\omega)$$
defines a probability measure Q on $\mathcal{F}$ which is equivalent to P. Under Q, the stochastic process $\{\widetilde{W}(t) : 0 \leq t \leq T\}$, where
$$\widetilde{W}(t) = W(t) + qt,$$
is the standard Brownian motion with respect to Q. It is adapted to $\{\mathcal{F}(t) : t \geq 0\}$.

- **Example 63.** *Solve the SDE*
$$\mathrm{d}X = cX\, \mathrm{d}t + \sigma X\, \mathrm{d}W.$$
where $c, \sigma \in \mathbb{R}$, and $\sigma > 0$.

Proof. Let
$$\widetilde{W}(t) = W(t) + \frac{c}{\sigma}t.$$
By Girsanov's theorem, $\widetilde{W}$ is the standard Brownian motion under the probability measure
$$Q(E) = \int_E \exp\left(- \frac{c}{\sigma}W(T) - \frac{c^2}{2\sigma^2}t\right) \mathrm{d}P$$
for any $T > 0$. Now,
$$\begin{aligned} \mathrm{d}\widetilde{W} &= \mathrm{d}W + \frac{c}{\sigma}\mathrm{d}t \\ \mathrm{d}W &= \mathrm{d}\widetilde{W} - \frac{c}{\sigma}\mathrm{d}t. \end{aligned}$$
Thus,
$$\mathrm{d}X = cX\, \mathrm{d}t + \sigma X \left(\mathrm{d}\widetilde{W} - \frac{c}{\sigma}\mathrm{d}t\right) = \sigma X\, \mathrm{d}\widetilde{W}.$$
Assume that $X = u(t, \widetilde{W})$. We have that
$$\mathrm{d}X = \sigma u(t, \widetilde{W})\, \mathrm{d}\widetilde{W},$$
and, by Itô's chain rule,
$$\mathrm{d}X = \left(u_1(t, \widetilde{W}) + \frac{1}{2}u_{22}(t, \widetilde{W})\right) \mathrm{d}t + u_2(t, \widetilde{W})\, \mathrm{d}\widetilde{W}.$$
As a result,
$$\begin{aligned} u_1(t, \widetilde{W}) + \frac{1}{2}u_{22}(t, \widetilde{W}) &= 0, \\ u_2(t, \widetilde{W}) &= \sigma u(t, \widetilde{W}). \end{aligned}$$

By the second equation, we have that $u(t,w) = C\exp(f(t) + \sigma w)$ for some function f. The first equation then becomes

$$\begin{aligned} f'(t)u(t,\widetilde{W}) + \frac{1}{2}\sigma^2 u(t,\widetilde{W}) = 0 \\ f'(t) = -\frac{1}{2}\sigma^2 \\ f(t) = -\frac{1}{2}\sigma^2 t. \end{aligned}$$

Hence,

$$X = C\exp(-0.5\sigma^2 t + \sigma\widetilde{W}) = C\exp(-0.5\sigma^2 t + \sigma(W + ct/\sigma)) = C\exp((c - 0.5\sigma^2)t + \sigma W).$$

The constant C is simply $X(0)$, the initial condition. □

A Proofs

- **Proposition 64.** *Let W be the standard Brownian motion. Let $\{\mathcal{P}^{(n)}\}_{n=1}^{\infty}$ be a sequence of nested partition of $[0,T]$ such that $\mathcal{P}^{(n)} \to 0$. Let $\lambda \in (0,1)$. Pick*

$$\tau_j^{(n)} = (1-\lambda)t_{j-1}^{(n)} + \lambda t_j^{(n)}.$$

Then,

$$\mathcal{S}(W, W, \mathcal{P}^{(n)}) \xrightarrow{\mathcal{L}^2} \frac{(W(T))^2}{2} + \left(\lambda - \frac{1}{2}\right)T.$$

Proof. We have that

$$\begin{aligned} \mathcal{S}(W, W, \mathcal{P}^{(n)}) &= \sum_{j=1}^{n} W(\tau_j^{(n)})\Big(W(t_j^{(n)}) - W(t_{j-1}^{(n)})\Big) \\ &= \sum_{j=1}^{n} W(\tau_j^{(n)})\Big(W(t_j^{(n)}) - W(\tau_j^{(n)}) + W(\tau_j^{(n)}) - W(t_{j-1}^{(n)})\Big) \\ &= \sum_{j=1}^{n} \bigg(W(\tau_j^{(n)})\Big(W(t_j^{(n)}) - W(\tau_j^{(n)})\Big) + W(\tau_j^{(n)})\Big(W(\tau_j^{(n)}) - W(t_{j-1}^{(n)})\Big)\bigg). \end{aligned}$$

Consider the $W(\tau_j^{(n)})\Big(W(t_j^{(n)}) - W(\tau_j^{(n)})\Big)$ term. We have that

$$\begin{aligned} &W(\tau_j^{(n)})\Big(W(t_j^{(n)}) - W(\tau_j^{(n)})\Big) \\ &= W(\tau_j^{(n)})W(t_j^{(n)}) - W(\tau_j^{(n)})^2 \\ &= \frac{1}{2}W(t_j^{(n)})^2 - \frac{1}{2}W(t_j^{(n)})^2 + W(\tau_j^{(n)})W(t_j^{(n)}) - \frac{1}{2}W(\tau_j^{(n)})^2 - \frac{1}{2}W(\tau_j^{(n)})^2 \\ &= \frac{1}{2}W(t_j^{(n)})^2 - \frac{1}{2}\Big(W(t_j^{(n)}) - W(\tau_j^{(n)})\Big)^2 - \frac{1}{2}W(\tau_j^{(n)})^2. \end{aligned}$$

For the $W(\tau_j^{(n)})\Big(W(\tau_j^{(n)}) - W(t_{j-1}^{(n)})\Big)$ term, we have that

$$\begin{aligned}
&W(\tau_j^{(n)})\Big(W(\tau_j^{(n)}) - W(t_{j-1}^{(n)})\Big) \\
&= W(\tau_j^{(n)})^2 - W(\tau_j^{(n)})W(t_{j-1}^{(n)}) \\
&= \frac{1}{2}W(\tau_j^{(n)})^2 + \frac{1}{2}W(\tau_j^{(n)})^2 - W(\tau_j^{(n)})W(t_{j-1}^{(n)}) + \frac{1}{2}W(t_{j-1}^{(n)})^2 - \frac{1}{2}W(t_{j-1}^{(n)})^2 \\
&= \frac{1}{2}W(\tau_j^{(n)})^2 + \frac{1}{2}\Big(W(\tau_j^{(n)}) - W(t_{j-1}^{(n)})\Big)^2 - \frac{1}{2}W(t_{j-1}^{(n)})^2.
\end{aligned}$$

As a result,

$$\begin{aligned}
&\mathcal{S}(W, W, \mathcal{P}^{(n)}) \\
&= \sum_{j=1}^{n}\left(\frac{1}{2}W(t_j^{(n)})^2 - \frac{1}{2}\Big(W(t_j^{(n)}) - W(\tau_j^{(n)})\Big)^2 + \frac{1}{2}\Big(W(\tau_j^{(n)}) - W(t_{j-1}^{(n)})\Big)^2 - \frac{1}{2}W(t_{j-1}^{(n)})^2\right) \\
&= \frac{1}{2}W(t_n^{(n)})^2 - \frac{1}{2}W(t_0^{(n)})^2 - \frac{1}{2}\sum_{j=1}^{n}\left(\Big(W(t_j^{(n)}) - W(\tau_j^{(n)})\Big)^2 - \Big(W(\tau_j^{(n)}) - W(t_{j-1}^{(n)})\Big)^2\right) \\
&= \frac{1}{2}W(t)^2 - \frac{1}{2}\sum_{j=1}^{n}\left(\Big(W(t_j^{(n)}) - W(\tau_j^{(n)})\Big)^2 - \Big(W(\tau_j^{(n)}) - W(t_{j-1}^{(n)})\Big)^2\right).
\end{aligned}$$

Let

$$Q^{(n)}(\lambda) := \sum_{j=1}^{n}\left(\Big(W(t_j^{(n)}) - W(\tau_j^{(n)})\Big)^2 - \Big(W(\tau_j^{(n)}) - W(t_{j-1}^{(n)})\Big)^2\right).$$

We have that

$$\begin{aligned}
E[Q^{(n)}(\lambda)] &= \sum_{j=1}^{n}\left(E\Big[\Big(W(t_j^{(n)}) - W(\tau_j^{(n)})\Big)^2\Big] - E\Big[\Big(W(\tau_j^{(n)}) - W(t_{j-1}^{(n)})\Big)^2\Big]\right) \\
&= \sum_{j=1}^{n}\Big((t_j^{(n)} - \tau_j^{(n)}) - (\tau_j^{(n)} - t_{j-1}^{(n)})\Big) = \sum_{j=1}^{n}\Big(t_j^{(n)} + t_{j-1}^{(n)} - 2\tau_j^{(n)}\Big) \\
&= \sum_{j=1}^{n}\Big(t_j^{(n)} + t_{j-1}^{(n)} - 2\big((1-\lambda)t_{j-1}^{(n)} + \lambda t_j^{(n)}\big)\Big) = \sum_{j=1}^{n}(1-2\lambda)(t_j^{(n)} - t_{j-1}^{(n)}) \\
&= (1-2\lambda)(t_n^{(n)} - t_0^{(n)}) = (1-2\lambda)T.
\end{aligned}$$

Also,

$$
\begin{aligned}
\mathrm{Var}(Q^{(n)}(\lambda)) &= \sum_{j=1}^{n} \mathrm{Var}\left(\left(W(t_j^{(n)}) - W(\tau_j^{(n)})\right)^2\right) + \sum_{j=1}^{n} \mathrm{Var}\left(\left(W(\tau_j^{(n)}) - W(t_{j-1}^{(n)})\right)^2\right) \\
&= \sum_{j=1}^{n} E\left[\left(W(t_j^{(n)}) - W(\tau_j^{(n)})\right)^4\right] - \sum_{j=1}^{n} E\left[\left(W(t_j^{(n)}) - W(\tau_j^{(n)})\right)^2\right]^2 \\
&\quad + \sum_{j=1}^{n} E\left[\left(W(\tau_j^{(n)}) - W(t_{j-1}^{(n)})\right)^4\right] - \sum_{j=1}^{n} E\left[\left(W(\tau_j^{(n)}) - W(t_{j-1}^{(n)})\right)^2\right]^2 \\
&= \sum_{j=1}^{n} 3(t_j^{(n)} - \tau_j^{(n)})^2 - \sum_{j=1}^{n} (t_j^{(n)} - \tau_j^{(n)})^2 + \sum_{j=1}^{n} 3(\tau_j^{(n)} - t_{j-1}^{(n)})^2 - \sum_{j=1}^{n} (\tau_j^{(n)} - t_{j-1}^{(n)})^2 \\
&= 2\sum_{j=1}^{n} (t_j^{(n)} - \tau_j^{(n)})^2 + 2\sum_{j=1}^{n} (\tau_j^{(n)} - t_{j-1}^{(n)})^2 \\
&\le 2|\mathcal{P}^{(n)}| \sum_{j=1}^{n} (t_j^{(n)} - \tau_j^{(n)}) + 2|\mathcal{P}^{(n)}| \sum_{j=1}^{n} (\tau_j^{(n)} - t_{j-1}^{(n)}) \\
&= 2|\mathcal{P}^{(n)}|T.
\end{aligned}
$$

So, as $n \to \infty$ and $|\mathcal{P}^{(n)}| \to 0$, we have that $\mathrm{Var}(Q^{(n)}(\lambda)) \to 0$. This means that

$$
\begin{aligned}
\lim_{n\to\infty} E\left[\left(Q^{(n)}(\lambda) - \left(\lambda - \frac{1}{2}\right)T\right)^2\right] &= 0 \\
\lim_{n\to\infty} E\left[\left(\frac{W(T)^2}{2} + Q^{(n)}(\lambda) - \frac{W(T)^2}{2} - \left(\lambda - \frac{1}{2}\right)T\right)^2\right] &= 0 \\
\lim_{n\to\infty} E\left[\left(\mathcal{S}(W, W, \mathcal{P}^{(n)}) - \frac{W(T)^2}{2} - \left(\lambda - \frac{1}{2}\right)T\right)^2\right] &= 0
\end{aligned}
$$

In other words,

$$
\mathcal{S}(W, W, \mathcal{P}^{(n)}) \xrightarrow{\mathcal{L}^2} \frac{W(T)^2}{2} + \left(\lambda - \frac{1}{2}\right)T
$$

as required. □

References

[Evans, 2013] Evans, L. C. (2013). *An Introduction to Stochastic Differential Equations.* American Mathematical Society, New York.

[Mikosch, 1998] Mikosch, T. (1998). *Elementary Stochastic Calculus with Finance in View.* World Scientific, Singapore.

[Mörters and Peres, 2012] Mörters, P. and Peres, Y. (2012). Brownian motion. `https://people.bath.ac.uk/maspm/book.pdf`. Accessed: 2022-01-21.

[Särkkä, 2012] Särkkä, S. (2012). Applied stochastic differential equations. `https://users.aalto.fi/~ssarkka/course_s2012/pdf/sde_course_booklet_2012.pdf`. Accessed: 2022-01-21.

[Song, 2021] Song, Y. (2021). Generative modeling by estimating gradients of the data distribution. `https://yang-song.github.io/blog/2021/score/index.html`. Accessed: 2022-01-21.