A Primer on Stochastic Differential Equations

Pramook Khungurn

January 24, 2022

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

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 $\mathbf{x} : [0,\infty) \to \mathbb{R}^n$ 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 $\mathbf{b} : \mathbb{R}^n \times \mathbb{R} \to \mathbb{R}^n$ is a smooth, time-varying vector field, and $\mathbf{x}_0 \in \mathbb{R}^n$ is a point in $\mathbb{R}^n$.

- 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 $B : \mathbb{R}^n \times \mathbb{R} \to \mathbb{R}^{n\times m}$ 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 Brownion 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, $\pi : \mathbb{R}^n \to [0, 1]$ 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 Brownian Motion and White Noise

2.1 Definition and Basic Properties

- In the rest of this note, we will be working with a probability space (Ω, Σ, P) where Ω is the sample space, Σ is a σ-algebra on Ω, and P is the probability measure on (Ω, Σ).

- **Definition 1.** *A* **stochastic process** *is a collection* $\{\mathbf{X}(t) \in \mathbb{R}^n : t \geq 0\}$ *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 two ways to view a stochastic process.

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

- When viewed as collection of random variables, we see it as a collection of (potentially correlated) random values on the real line.
- When viewed from the lens of the sample path, it becomes a random function.

- **Definition 2.** *A stochastic process* $\{\mathbf{X}(t) \in \mathbb{R}^n : t \geq 0\}$ *is called a* **Gaussian process** *if, for any* $0 \leq t_1 < t_2 < \ldots < t_k$*, the vector* $(X(t_1), X(t_2), \ldots, X(t_k))$ *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.*

- **Definition 3.** *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.*
  - $W(0) = x_0$.
  - *The process has independent increments. That is, for all times* $0 \leq t_1 < t_2 < \cdots < t_n$*, the increments* $W(t_2) - W(t_1)$*,* $W(t_3) - W(t_2)$*, ...,* $W(t_n) - W(t_{n-1})$ *are independent random variables.*
  - *For all* $t \geq 0$ *and* $h > 0$*, the increment* $W(t + h) - W(t)$ *is normally distributed with expectation* 0 *and variance* h*. In other words,* $W(t + h) - W(h) \sim \mathcal{N}(0, h)$.
  - *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**.

- Note that because, 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 always a Gaussian process.

- **Theorem 4 (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. $\square$

- **Lemma 5.** *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. $\square$

- 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 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, it also turns out that the sample path $t \mapsto W(t,\omega)$ is not differentiable at any $t \geq 0$. So, the derivative does not really exist.

- Neverless, we can show that, in a sense,

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

  where δ is the Dirac delta function.

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

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

  There are 4 cases.

1. If $t+h<s$, then $\phi_s(s) = (t+h-t+h-t+t)/h^2 = 0$.
2. If $t \le s < t+h$, then $\phi_s(s) = (t+h-s-t+t)/h^2 = (h+t-s)/h^2$.
3. If $t-h \le s < t$, 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 $s \ne t$. As a result, $\lim_{h\to 0}\phi_h(s) = \delta(t-s)$ where δ is the Direct delta function.

- **Definition 6.** *Let $X(t)$ be a real-valued stochastic process with $E[X^2(t)] < \infty$ for all $t \ge 0$. The* **autocorrelation function of** X *is the function*

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

defined for $t, s \ge 0$.

- **Definition 7.** *We call a stochastic process $X(t)$* **stationary in the wide sence** *if*
  - *$r(t-s) = c(t-s)$ for some function $c : \mathbb{R} \to \mathbb{R}$, and*
  - *$E[X(t)] = E[X(s)]$*

  for all $t, s \ge 0$.

- **Definition 8.** *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 $\lambda \in \mathbb{R}$.

- Note that the white noise $\xi(t)$ is, in a sense, stationary in the wide sense (i.e., its mean should be zero), and its autocorrelation is given by $c(t-s) = \delta(t-s)$. Its 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.

- We can extend the standard Brownian motion in $\mathbb{R}$ to one in $\mathbb{R}^n$.

  Definition 9. *A stochastic process $\{\mathbf{W}(t) \in \mathbb{R}^n : t \ge 0\}$ where $\mathbf{W}(t) = (W_1(t), W_2(t), \dots, W_n(t))$ is an n***-dimensional Brownian motion** *if it satisfies the following conditions.*
  - *For each $k = 1, 2, \dots, n$, we have that $W_k(t)$ is a one-dimensional Brownian motion.*
  - *The σ-algebras $\mathcal{W}_k = \sigma\Big(\bigcup_{t\ge 0}\sigma(W_k(t))\Big)$, for $k = 1, 2, \dots, n$, 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

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

is the Kronecker delta function.

2.2 Properties of Sample Paths

- **Definition 10.** *A function $f : [0,\infty) \to \mathbb{R}$ is said to be* **locally α-Hölder continuous at** $x \geq 0$ *if there exists $\varepsilon > 0$ and $c > 0$ such that*

$$|f(x) - f(y)| \leq c|x-y|^\alpha$$

for all $y \geq 0$ such that $|y - x| < \varepsilon$. We refer to $\alpha > 0$ as the **Hölder exponent** *and to $c > 0$ as the* **Hölder constant**.

- **Theorem 11.** *Let $W(t)$ be a Brownian motion in 1D.*
  - *If $\alpha < 1/2$, then the sample path $t \mapsto W(t,\omega)$ is everywhere locally α-Hölder continuous almost surely.*
  - *If $\alpha > 1/2$, however, it is not locally α-Hölder anywhere almost surely.*

- **Theorem 12.** *For all $0 < a < b < \infty$, the sample path $t \mapsto W(t,\omega)$ is not monotone on the interval $[a,b]$ almost surely.*

- **Definition 13.** *For a function $f : \mathbb{R} \to \mathbb{R}$, the* **upper** *and* **lower right derivatives** *are defined as*

$$D^* f(t) = \limsup_{h\to 0} \frac{f(t+h) - f(t)}{h},$$
$$D_* f(t) = \liminf_{h\to 0} \frac{f(t+h) - f(t)}{h}.$$

- Note that the derivative of f at t exists if and only if $D^* f(t) = D_* f(t)$.

- **Theorem 14 (Paley, Wiener, and Zygmund 1933).** *The sample path $t \mapsto W(t,\omega)$ is nowhere differentiable almost surely, Furthermore, for all $t \geq 0$, either $D^* W(t) = \infty$ or $D_* W(t) = -\infty$, or both almost surely.*

- **Definition 15.** *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 16.** *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 17.** *A function $f : \mathbb{R} \to \mathbb{R}$ is the* **variation** *of f on the $[a,b]$ is given by*

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

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*

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

- **Theorem 18.** *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.

- Note that the fact that $V^{(1)}_{[a,b]}(W)$ follows from the fact that $V^{(2)}_{[a,b]}(W)$ is finite. To see this, note that $W(t)$ is locally α-Hölder. So, for n large enough, we would have that, for any $0 < \alpha < 1/2$.

$$\begin{aligned}
|W(t_j^{(n)}) - W(t_{j-1}^{(n)})| &\leq c|t_j^{(n)} - t_{j-1}^{(n)}|^\alpha \leq c|\mathcal{P}^{(n)}|^\alpha \\
\frac{1}{|W(t_j^{(n)}) - W(t_{j-1}^{(n)})|} &\geq \frac{1}{c|\mathcal{P}^{(n)}|^\alpha} \\
\frac{(W(t_j^{(n)}) - W(t_{j-1}^{(n)}))^2}{|W(t_j^{(n)}) - W(t_{j-1}^{(n)})|} &\geq \frac{1}{c|\mathcal{P}^{(n)}|^\alpha}(W(t_j^{(n)}) - W(t_{j-1}^{(n)}))^2 \\
|W(t_j^{(n)}) - W(t_{j-1}^{(n)})| &\geq \frac{1}{c|\mathcal{P}^{(n)}|^\alpha}(W(t_j^{(n)}) - W(t_{j-1}^{(n)}))^2 \\
\sum_{j=1}^{k(n)} |W(t_j^{(n)}) - W(t_{j-1}^{(n)})| &\geq \frac{1}{c|\mathcal{P}^{(n)}|^\alpha} \sum_{j=1}^{k(n)} (W(t_j^{(n)}) - W(t_{j-1}^{(n)}))^2.
\end{aligned}$$

  Taking the limit on both sides, we have that

$$V^{(1)}_{[a,b]}(W) \geq V^{(2)}_{[a,b]}(W)\left(\lim_{|\mathcal{P}^{(n)}| \to 0} \frac{1}{c|\mathcal{P}^{(n)}|^\alpha} \right).$$

  So, if $V^{(2)}_{[a,b]}(W)$ is a positive, then $V^{(1)}_{[a,b]}(W)$ would have to be infinite.

2.3 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 19 (Markov property).** *A Brownian motion is a Markov process. More precisely, let $\{\mathbf{W}(t) : t \geq 0\}$ be a Brownian motion starting at $\mathbf{x}_0 \in \mathbb{R}^n$, 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 20.** *A* **filtration** *on a probability space (Ω, Σ, P) is a family $\{\Sigma(t) : t \geq 0\}$ of σ-algebras such that $\Sigma(s) \subseteq \Sigma(t)$ for all $s < t$.*

  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 $\{\Sigma(t) : t \geq 0\}$ is said to be **adapted to the filtration** *if $X(t)$ is $\Sigma(t)$-measurable for any $t \geq 0$.*

- Let $\mathbf{W}(t)$ be a Brownian motion defined on a probabilty space (Ω, Σ, P). We can define a filtration

$$\Sigma^0(t) = \sigma(W(s) : 0 \leq s \leq t) = \sigma\bigg(\bigcup_{0 \leq s \leq t} \sigma(\mathbf{W}(s))\bigg)$$

  to be the σ-algebra generated by the random variables $\mathbf{W}(s)$ for $0 \leq s \leq t$. We have that the Brownian motion is adapted to the filtration. The Markov property says that $\{\mathbf{W}(t+s) - \mathbf{W}(s) : t \geq 0\}$ is independent of $\Sigma^0(s)$.

- The Markov property can be slightly improved. Define

$$\Sigma^+(s) = \bigcap_{t > s} \Sigma^0(t).$$

  Intuitively, it contains all information common to the future of the Brownian motion from time s. We have that the family $\{\Sigma^+(t) : t \geq 0\}$ is a filtration. Moreover, $\Sigma^0(s) \subseteq \Sigma^+(s)$ because $\Sigma^+(s)$ has an infinitesimally more information about the future.

- **Theorem 21 (Slightly stronger Markov property).** *The process* $\{\mathbf{W}(t+s) - \mathbf{W}(s) : t \geq 0\}$ *is independent of the* σ*-algebra* $\Sigma^+(s)$.

- 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 $\Sigma(t)$.

  Definition 22. *A random variable* T *with values in* $[0, \infty)$ *defined on a probability space with filtration* $\{\Sigma(t) : t \geq 0\}$ *is called a* **stopping time** *with respect to the filtration if* $\{T \leq t\} \in \Sigma(t)$ *for all* $t \geq 0$.

- **Theorem 23 (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* $\Sigma^+(T)$.

3 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 $\mathbb{R}^n$.

- 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 choise of $\tau_j^{(n)}$ as we take the limit. This is true for deterministic functions, but is not true for the Brownian motion.

- In fact, one can show that, for any $0 \le \lambda \le 1$, one can show that, if we pick

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

then

$$\lim_{\substack{n\to\infty \\ |\mathcal{P}^{(n)}|\to 0}} \mathcal{S}(W,W,\mathcal{P}^{(n)}) = \frac{(W(T))^2}{2} + \left(\lambda - \frac{1}{2}\right)T$$

when $\{\mathcal{P}^{(n)}\}_{n=1}^{\infty}$ partitions $[0,T]$.

- 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?

3.1 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.

  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 24.** *A stochastic process $\{X(t,\omega) : t \geq 0, \omega \in \Omega\}$ defined on a probability space (Ω, Σ, P) that is adapted to a filtration $\{\Sigma(t) : t \geq 0\}$ is called* **progressively measurable** *if, for each $t \geq 0$, the mapping $X : [0,t] \times \Omega \to \mathbb{R}$ is measurable with respect to the σ-algebra $\mathcal{B}([0,,1]) \times \Sigma(t)$.*

- **Lemma 25.** *Any process which is adapted and is either right or left continuous is progressively measurable.*

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

  Definition 26. *A* **step process on** $[0,T]$ *is a real-valued stochastic process $\{H(t,\omega) : t \geq 0, \omega \in \Omega\}$ of the form*

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

  where $0 = t_1 \leq t_2 \leq \cdots \leq t_{k+1} = T$, A_i is a $\Sigma(t)$-measurable random variable (which only depends on ω), and $\chi_{[t_{i-1}, t_{i+1})}(t)$ is the charactistic function of the interval $[t_i, t_{i+1})$. That is,

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

  Note that a step process only takes finitely many values.

- **Definition 27.** *Let $\{H(t,\omega) : t \geq 0, \omega \in \Omega\}$ be a real-valued progressively measurable step process on $[0,T]$ defined on a probability space (Ω, Σ, P) with filtration $\{\Sigma(t) : t \geq 0\}$ to which the standard Brownian motion $\{W(t) : t \geq 0\}$ is adapted. Define the* **Itô integral of H with respect to W** *to be*

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

- 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 L^2 norm.

- **Definition 28.** *Let $\{H(t) : t \geq 0\}$ be a real-valued stochastic process. The* **L^p norm of H on** $[0,T]$ *is given by*

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

  where the integration on the RHS is the Lebesgue integral on the real line performed on a sample path of H. We denote by $\mathcal{L}^p(0,T)$ the space of all real-valued, progressively measurable stochastic process H such that $\|H\|_p < \infty$.

- **Lemma 29.** *If $G \in \mathcal{L}^2(0,T)$, there exists a sequence of bounded step processes $\{H^{(n)}\}_{n=1}^{\infty}$ in $\mathcal{L}^2(0,T)$ such that $\lim_{n\to 0} \|G - H^{(n)}\|_2 = 0$.*

- **Definition 30.** *Let $G \in \mathcal{L}^2(0,T)$. The* **Itô integral of** G **on** $[0,T]$ *is defined to be*

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

where $\{H^{(n)}\}_{n=1}^{\infty}$ is a sequence of step processes in $\mathcal{L}^2(0,T)$ such that $\lim_{n\to 0} \|G - H^{(n)}\|_2 = 0$. The integral defined in this way does not depend on the particular sequence of step processes used to approximate G.

- We have defined the Itô integrals for functions in $\mathcal{L}^2(0,T)$ 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 $\mathcal{L}^2(0,T) \subseteq \mathcal{M}^2(0,T)$ because, if the expectation is finite, then the probabily that $\int G^2 \,\mathrm{d}t = \infty$ must be 0.

- **Theorem 31.** *For all constants $a, b \in \mathbb{R}$ and for all $G, H \in \mathcal{L}^2(0,T)$, we have that*

$$\begin{aligned}
\int_0^T (aG + bH) \,\mathrm{d}W &= a \int_0^T G \,\mathrm{d}W + b \int_0^T H \,\mathrm{d}W, \\
E\left[\int_0^T G \,\mathrm{d}W\right] &= 0, \\
E\left[\left(\int_0^T G \,\mathrm{d}W\right)^2\right] &= E\left[\int_0^T G^2 \,\mathrm{d}t\right], \\
E\left[\left(\int_0^T G \,\mathrm{d}W\right)\left(\int_0^T H \,\mathrm{d}W\right)\right] &= E\left[\int_0^T GH \,\mathrm{d}t\right].
\end{aligned}$$

3.2 Itô's Chain and Product Rules

- **Definition 32.** *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 $F \in \mathcal{L}^1(0,T)$, $G \in \mathcal{L}^2(0,T)$, 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 33 (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

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

- **Example 34.** *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 35 (Itô's product rule).** *Suppose*

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

for $0 \leq t \leq dT$, $F_i \in \mathcal{L}^1(0 =, T)$, and $G_i \in \mathcal{L}^2(0, T)$. 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**.

- The integrated versino 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 36. *Suppose that*

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

with $F_i \in \mathcal{L}^1(0,T)$, $G_i \in \mathcal{L}^2(0,T)$ for $i = 1, \dots, n$. If $u(x_1, x_2, \dots, x_n, t) : \mathbb{R}^n \times [0,T] \to \mathbb{R}$ 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

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

3.3 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 37.** *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 38 (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,

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

- We let $\mathcal{L}_n^2(0,T)$ 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 $\mathcal{L}^2(0,T)$. Also, let $\mathcal{L}_{n\times m}^2(0,T)$ 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 $\mathcal{L}^2(0,T)$.

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

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

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

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

- **Lemma 39.** *If $G \in \mathcal{L}_{n\times m}^2(0,T)$, 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:

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

- Moreover, 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 40 (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 $u : \mathbb{R}^n \times [0, T]$ 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:

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

4 Stochastic Differential Equations

- **Definition 41.** *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 $\mathbb{R}^n$-valued stochastic process $\{\mathbf{X} : 0 \le t \le T\}$ is a solution of the differential equation*

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

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

1. *$\mathbf{X}(t)$ is progressively measurable with respect the filtration $\{\Sigma(t) : 0 \le t \le T\}$ where $\Sigma(t)$ 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. $\mathbf{b}(\mathbf{X}(t),t) \in \mathcal{L}^1_n(0,T)$.

3. $B(\mathbf{X}(t),t) \in \mathcal{L}^2_{n\times m}(0,T)$.

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.

4.1 Examples

- **Example 42.** *Suppose* f,g *are continuous functions* t *(not random variables). Consider the initial value problem*

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

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 43 (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 31,

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

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 44 (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 $Y(t)$.

4.2 Properties of Solution

- **Theorem 45.** *Let $\mathbf{b} : \mathbb{R}^n \times [0,T] \to \mathbb{R}^n$ and $B : \mathbb{R}^n \times [0,T] \to \mathbb{R}^{m\times n}$ be uniformly Lipschitz continuous, meaning that there exists a constant L such that*

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

for all $0 \le t \le T$, $\mathbf{x},\mathbf{y} \in \mathbb{R}^n$. 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 $\mathbf{X} \in \mathcal{L}^2(0,T)$ 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 $0 \le t \le T$. By "unique," we mean that, if $\widetilde{\mathbf{X}}$ is another solution of the SDE, $\mathbf{X}(t) = \widetilde{\mathbf{X}}(t)$ for all $0 \le t \le T$ 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 $\mathcal{L}^2(0,T)$, which clearly solves the SDE.

- **Theorem 46.** *The probability density $p(\mathbf{X},t)$ of the solution of the SDE in Theorem 45 solves the partial differential equation*

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

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

- **Theorem 47.** *Let $\mathbf{X}(t)$ be the solution of the SDE in Theorem 45. 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

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

5 Numerical Solution to SDE

- Recall that we wish to solve

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

  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.

References

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

[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.