Stochastic Calculus with Application to non-Linear Filtering Notes

Francesco N. Chotuck

Contents

Include section with review exercises as the problem sheets

Notation

Let $a, b \in \mathbb{R}$.

- $\triangleq$ means "is defined to be".
- $a \wedge b \triangleq \min(a, b)$.
- $a \vee b \triangleq \max(a, b)$.
- $a^+ \triangleq \max(a, 0)$.
- $a^- \triangleq -\min(a, 0) = \max(-a, 0)$.
- $\mathbb{R}_+ = [0, \infty)$.

1 Introduction

The price of a stock is not a smooth function of time, and standard calculus tools cannot be used to effectively model it. A commonly used technique is to model the price S as a *geometric Brownian motion*, given by the *stochastic differential equation (SDE)*

$$dS(t) = \alpha S(t)\, dt + \sigma S(t)\, dW(t),$$

where α and σ are parameters, and W is a Brownian motion. If $\sigma = 0$, this is simply the ordinary differential equation

$$dS(t) = \alpha S(t)\, dt \quad \text{or} \quad \partial_t S = \alpha S(t).$$

This is the price assuming it grows at a rate α. The σdW term models *noisy fluctuations* and the first goal of this course is to understand what this means. The mathematical tools required for this are Brownian motion, and Itô integrals, which we will develop and study.

These notes are based on [KS91].

2 Martingales in Continuous Time

Remark 2.1. We always work on a probability space $(\Omega, \mathcal{F}, P)$ where

- Ω is an abstract space,
- $\mathcal{F}$ is a σ-algebra (σ-field), and
- $\mathbb{P}$ is a probability measure.

2.1 Continuous time processes

Definition 2.2. A **stochastic process in continuous time**, is a family of random variables $(X_t)_{t\geq 0}$ parametrised by $t \geq 0$. For every $\omega \in \Omega$, the function $t \mapsto X_t(\omega)$ is called the **sample path**.

> **Note 2.3.** A stochastic process is a mathematical model for the occurence, at each moment after the initial time, of a random phenomenon.

Definition 2.4. A stochastic process $(X_t)_{t\geq 0}$ is called **measurable** if for every Borel set $A \in \mathcal{B}(\mathbb{R})$ the set $\{(t, w) \in \mathbb{R}_+ \times \Omega : X_t(\omega) \in A\}$ belongs to the product σ-field $\mathbb{R}_+ \otimes \mathcal{F}$.

> **Remark 2.5.** This definition is required for technical reasons.

As time elapses, the available information to the observer of a stochastic process increases. In other words, the dependence on time suggests a flow of information to the observer. The following definition formalises this.

Definition 2.6. A family of σ-fields $\mathcal{F}_{t\geq 0}$ on a given probability space $(\Omega, \mathcal{F}, \mathbb{P})$, is called a **filtration** if $\mathcal{F}_s \subseteq \mathcal{F}_t \subset \mathcal{F}$ for all $0 \leq s \leq t$.

Definition 2.7. A stochastic process $(X_t)_{t\geq 0}$ is called **adapted** if for $t \geq 0$, X_t is $\mathcal{F}_t$-measurable.

Example 2.8

The simplest choice of a filtration is that generated by the process itself, which we call the **natural filtration** of a process X, defined by

$$\mathcal{F}_t^X \triangleq \sigma(X_s : 0 \leq s \leq t).$$

That is, the smallest σ-field with respect to which X_s is measurable for every $s \in [0, t]$.

Trivially, a process is adapted with respect to its natural filtration.

Definition 2.9. A probability space on which a filtration is defined is called a **filtered probability space**.

Definition 2.10. A σ-field $\mathcal{F}$ is said to be **complete** if, whenever $A \in \mathcal{F}$ with $\mathbb{P}(A) = 0$ and $B \subset A$, it follows that $B \in \mathcal{F}$.

Definition 2.11. Let $\mathcal{N} = \{A \in \mathcal{F} : \mathbb{P}(A) = 0\}$, we shall say that a filtration $\{\mathcal{F}_t\}$ satisfies the usual conditions if $\mathcal{N} \subset \mathcal{F}_0$ and $\mathcal{F}_t = \bigcap_{\varepsilon>0} \mathcal{F}_{t+\varepsilon}$ for all $t \geq 0$.

> **Remark 2.12.** Most of the statements in this course will be "almost sure" i.e. they hold with probability 1. To avoid technical difficulties, the above two definitions are technical requirements that ensure desirable properties. Henceforth, we shall assume that most filtrations are complete and satisfy the usual conditions.

Definition 2.13. A stochastic process is said to be **right/left continuous** if the paths are right/left continuous almost surely. That is, the $\mathbb{R}$-valued function $t \mapsto X_t(\omega)$ is right/left continuous almost surely.

Remark 2.14. In this course we will be working with continuous processes.

Notation. The abbreviation RCLL or càdlàg is often used for processes which are almost surely right continuous with left limits i.e. it is right continuous and the left exists. LCRL for the opposite.

Exam Questions 2.15 (Exercise [KS91, 1.7 Exercise])

Let X be a real value process whose every path is RCLL. If a denotes the event that X is continuous on $[0, t)$ for $t > 0$ show that $A \in \mathcal{F}_t^X$.

Solution.

To do, see solution of problem sheets

Definition 2.16. Let X_t be a process. If for any $0 \leq s < t$ the distribution of the random variable $X_t - X_s$ depends on s, t only through the difference $t - s$ we say that the process has **stationary increments**.

Note 2.17. In other words, amount by which it changes over a fixed period of time depends only on the length of that time period, not on when the period starts.

Definition 2.18. Consider the same setting as the previous definition. If for any $0 \leq t_1 < t_2 < \cdots < t_n$ the random variables $X_{t_2} - X_{t_1}, \ldots, X_{t_n} - X_{t_{n-1}}$ are independent, we say that the process has **independent increments**.

2.2 Stopping times

Definition 2.19. A **random time** is a random varibale $T : (\Omega, \mathcal{F}) \to [0, \infty]$ which is $\mathcal{F}$-measurable.

Definition 2.20. A random time T is called a **stopping time** if for every $t > 0$ we have that $\{\omega : T(\omega) \leq t\} \in \mathcal{F}_t$.

Exam Questions 2.21 (Exercise)

Let X be a real valued stochastic process with continuous paths and $c > 0$ be a positive constant. Set $T = \inf \{t \geq 0 : X_t \geq c\}$. Then T is a stopping time.

TO do

Definition 2.22. Suppose that T is a stopping time with respect to the filtration $\{\mathcal{F}_t\}$. We define the σ-field

$$\mathcal{F}_T := \{A \in \mathcal{F} : A \cap \{T \leq t\} \in \mathcal{F}_t, \forall t \geq 0\}.$$

Note 2.23. This σ-field can be thought of all the information that can be gathered before the occurence of a stopping time T.

Exam Questions 2.24

Show that $\mathcal{F}_T$ is indeed a σ-field and that T is $\mathcal{F}_T$-measurable.

To do

Proposition 2.25

Let $\{X_t\}_{t\geq 0}$ be a continuous stochastic process adapted to a filtration $\{\mathcal{F}_t\}$. Let T be a stopping time with respect to the same filtration. Define the random variable:

$$X_T(\omega) := X_{T(\omega)}(\omega) \quad \text{on the set } \{T < \infty\}.$$

Then the following hold:

1. X_T is measurable with respect to the σ-field $\mathcal{F}_T$.
2. The stopped process $\{X_t^T\}_{t\geq 0}$ defined by

$$X_t^T := X_{T\wedge t}, \quad t \geq 0,$$

   - Continuous: X_t^T is continuous as a function of t.
   - Adapted: X_t^T is adapted to the filtration $\{\mathcal{F}_t\}$.

2.3 Martingales in continuous time

Definition 2.26. The process M_t as above is said to be a **martingale** (**sub-martingale, supermartingale**)

- M_t is adapted to $\mathcal{F}_t$.
- M_t is integrable for all $t \in \mathbb{R}_+$.
- For all s, t with $s < t$ we have $\mathbb{E}(M_t \mid \mathcal{F}_s) = M_s$, (respectively $\mathbb{E}(M_t \mid \mathcal{F}_s) \geq M_s, \mathbb{E}(M_t \mid \mathcal{F}_s) \leq M_s$ almost surely).

Note 2.27. We can think of each type of martingale as the following:

- Martingales: a fair game with no advantage or disadvantage, where the expected future value equals the current value.
- Supermartingale: a game where you are more likely to lose than to win; the expected value can only decrease or stay the same over time.
- Submartingale: a game where youa are more likely to win than to lose; the expected value can only increase or stay the same over time.

Definition 2.28. We write $\mathcal{M}^2$ for the space of martingales such that $\mathbb{E}(M_t^2) < \infty$ for all t and call them **square integrable** martingales. We also denote by $\mathcal{M}_c^2$ the subspace of $\mathcal{M}^2$ whose elements have continuous paths.

Remark 2.29. In these notes, and everywhere else, if no explicit reference is made to the filtration used, a process is checked to be a martingale always with respect ot its natural filtration.

Exam Questions 2.30

Let X_t for $t \in \mathbb{R}_+$ be a real valued martingale and $g : \mathbb{R} \to \mathbb{R}$ be a convex function such that $\mathbb{E}\left(|g(X_t)|\right) < \infty$ for all t. Show that the process $g(X_t)$ for $t \geq 0$ is a submartingale.

To do

.

Definition 2.31. Let N_t for $t \geq 0$ be an adapted process to a filtration $\{\mathcal{F}_t\}$ with RCLL paths such that $N_0 = 0$. Assume that for any $s \in [0, t) N_t - N_s$ is independent of $\mathcal{F}_s$ and is distributed as a Poisson random variable with parameter $\lambda(t-s)$. We call N_t a **Poisson process**.

Definition 2.32. We define the **compensated Poisson process** as $M_t \triangleq N_t - \lambda t$.

Exam Questions 2.33

Show that the compensated Poisson process is an $[\mathcal{F}_t]$-martingale.

To do

Proposition 2.34 (Doob's maximal inequality)

If $(M_t, \mathcal{F}_t)$ for $t \geq 0$ is a continuous martingale we have

$$\mathbb{E}\left(\sup_{0 \leq s \leq T} |M_s|^p\right) \leq \left(\frac{p}{p-1}\right)^p \mathbb{E}\left(|M_T|^p\right)$$

for $p > 1$.

Exam Questions 2.35 (Exercise)

Let N_t be a Poisson process with parameter λ. Given a complex number $z \in \mathbb{C}$, consider the process

$$X_t = \exp\left(izN_t - \lambda t\left(e^{iz} - 1\right)\right), \quad t \geq 0$$

where i denotes the imaginary unit. Show that $\mathrm{Re}(X_t)$, $\mathrm{Im}(X_t)$ are both martingales.

To do

Exam Questions 2.36 (Exercise)

Show that a submartingale of constant expectation is a martingale.

To do

2.4 The Doob-Meyer decomposition

Note 2.37. This section is devoted to the decomposition of certain submartingales as the summation of a martingale and an increasing process. In other words, we can write a submartingale as

$$X_n = \text{martingale} + \text{increasing process}.$$

Proposition 2.38. Let $(X_n, \mathcal{F}_n)$ for $n \in \mathbb{N}$ be a discrete time submartingale. We then define the discrete time process

$$\begin{aligned} A_0 &= 0 \\ A_{n+1} &= A_n + \mathbb{E}(X_{n+1} \mid \mathcal{F}_n) - X_n \end{aligned}$$

for $n \geq 1$. The process A_n is an increasing process in the sense that $A_n \leq A_{n+1}$ almost surely for all n. Moreover, the process $M_n := X_n - A_n$ is a martingale.

Note 2.39. In other words, our submartingale can be written as $X_n = \text{martingale} + \text{increasing process}$.

Proposition 2.40. The process A_n as above, is $\mathcal{F}_{n-1}$-measurable. Such processes are called **previsible**.

Exam Questions 2.41 (Exercise)

Prove that if one requires that the increasing process is also previsible in the above decomposition, then the decomposition is unique.

To do

Definition 2.42. A class $\mathcal{C}$ of random variables is **uniformly integrable** if given any $\varepsilon > 0$ there exists a $K > 0$ such that $\mathbb{E}\left(|X| \, \mathbf{1}_{|X|>K}\right) < \varepsilon$ for all $X \in \mathcal{C}$, where

$$\mathbf{1}_{|X|\geq K} = \begin{cases} 1 & \text{if } |X| \geq K, \\ 0 & \text{if } |X| < K. \end{cases}$$

is the indicator function.

Note 2.43. Uniform integrability controls the "tail behaviour" of the random variables, preventing them from having excessively large values that could disrupt averages or limits. Even if large values are possible, they are so rare that they do not significantly impact the overall expectation.

Theorem 2.44 (Doob-Meyer decomposition)

Let $(X_t, \mathcal{F}_t)$ for $t \geq 0$ be a submartingale with continuous paths. Assume that $\mathcal{F}_t$ satisfies the usual conditions and define the family of stopping times

$$\mathcal{J}_a := \{T \text{ a stopping time and } \mathbb{P}(T \leq a) = 1\} .$$

Assume that the family of random variables $\{X_T\}_{T\in\mathcal{J}_a}$ is uniformly integrable for any $a > 0$. Then X_t can be written as

$$X_t = M_t + A_t$$

where M_t is a continuous martingale and A_t is an increasing process. The latter is also continuous and with this additional property the decomposition is unique.

Exam Questions 2.45

Show that for an element $M \in \mathcal{M}_c^2$ the family $\{M_T^2\}_{T\in\mathcal{J}_a}$ is uniformly integrable for any $a > 0$.

to do

Note 2.46. In other words, for an element $M \in \mathcal{M}_c^2$ the process M_t^2 admits the Doob-Meyer decomposition.

Definition 2.47. For $M_t \in \mathcal{M}_c^2$, we define the **quadratic variation** of M_t to be the process $[M]_t \triangleq A_t$, where A_t is the natural increasing process in the Doob-Meyer decomposition of M_t^2. In other words, $[M]_t$ is that unique (up to indistinguishability) adapted, increasing process, for which $M_t^2 - [M]_t$ is a martingale.

Exam Questions 2.48 (Exercise)

Let M_t be compensated Poisson process of parameter λ. Show that $[M]_t = \lambda t$.

To do

Proposition 2.49

For any two elements $X, Y \in \mathcal{M}_c^2$ the processes

$$(X + Y)^2 - [X + Y] \quad \text{and} \quad (X - Y)^2 - [X - Y]$$

are martingales, and therefore so is their difference

$$4XY - ([X + Y] - [X - Y]) .$$

Definition 2.50. For any two martingales $X, Y \in \mathcal{M}_c^2$, we define their **cross-variation process** by

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

for $t \geq 0$.

Proposition 2.51. The process $XY - [X, Y]$ is a martingale.

Definition 2.52. Two elements $X, Y \in \mathcal{M}_c^2$ are called (**strongly**) **orthogonal** if

$$[X, Y]_t = 0$$

almost surely for any $t \geq 0$.

Remark 2.53. In view of the identities

$$\begin{aligned}\mathbb{E}\left((X_t - X_s)(Y_t - Y_s) \mid \mathcal{F}_s\right) &= \mathbb{E}\left(X_tY_t - X_sY_s \mid \mathcal{F}_s\right) \\ &= \mathbb{E}\left([X, Y]_t - [X, Y]_s \mid \mathcal{F}_s\right) \text{ almost surely}\end{aligned}$$

valid for any $0 < s < t$, the orthogonality of $X, Y \in \mathcal{M}_c^2$ is equivalent to the statements

- XY is a martingale, or
- the increments of X and Y over $[s, t]$ are orthogonal, conditional on $\mathcal{F}_s$.

Exam Questions 2.54

Show that $[\cdot, \cdot]$ is a bilinear form on $\mathcal{M}_c^2$.

To do

Definition 2.55. Let $(M_t, \mathcal{F}_t)$ be an adapted process with continuous paths. Suppose that we can fine a non-decreasing sequence of stopping times that converges to $+\infty$ almost surely, i.e. $\mathbb{P}(T_n \to +\infty) = 1$, such that for every n the process $M_t^{T_n}$ is a continuous martingale. Then we call M_t a **local martingale**. The space of these process is denoted by $\mathcal{M}_{c,loc}$.

Note 2.56. A local martingale behaves like a martingale up to a stopping time, meaning it may not be a true martingale globally, but locally (within finite time intervals), it satisfies the martingale property.

Definition 2.57. Let X_t be a $\{\mathcal{F}_t\}$-adapted process with continuous paths. If X_t can be decomposed as

$$X_t = M_t + V_t$$

where M_t is an $\{\mathcal{F}_t\}$ martingale with continuous paths and the paths of V_t are of finite variataion then we call X_t a **semimartingale**.

3 Brownian motion

3.1 Definition and properties

Example 3.1

A one-dimensional Brownian motion may be thought of as an infinitesimal random walk. That is, a process which at every infinitesimal time interval has equal (and infinitesimal) probabilities of moving up or down. In fact, a formal construction of Brownian motion is based on this approach. Hence, a first look at the discrete analogue of Brownian motion, the discrete-time random walk, makes sense.

Let $\{\xi_i\}_i$ be a countable family of independent binomially distributed random variables with

$$P(\xi_i = +1) = P(\xi_i = -1) = \frac{1}{2}.$$

Then $S_n = \sum_{i=1}^{n} \xi_i$ is a simple symmetric random walk.

Exam Questions 3.2

Show tgat S_n as defined above is a discrete time martingale and that it has stationary and independent increments (with the equivalent discrete time definitions).

To do

Remark 3.3. With the above exercise, we can show that for any $m, n \in \mathbb{N}$ we have that

$$\text{Cov}(S_m, S_n) = n \wedge m.$$

Definition 3.4. Consider a real-valued stochastic process W_t for $t \geq 0$ on a probability space $(\Omega, \mathcal{F}, \mathbb{P})$. It is a **standard Brownian motion** if the following are satisfied

1. $W_t(\omega)$ is continuous almost surely and $W_0 = 0$.
2. W_t has independent increments. That is, if $0 \leq t_1 < t_2 < \cdots < t_n$, the n random variables
$$W_{t_2} - W_{t_1}, W_{t_3} - W_{t_s}, \ldots, W_{t_n} - W_{t_{n-1}}$$
are independent random variables.
3. For any $s < t$ the random variable $W_t - W_s$ is normally distributed wiht mean 0 and variance $t - s$. In shorthand, $W_t - W_s \sim \mathcal{N}(0, t - s)$.

Remark 3.5. The adjective “standard” above comes from the fact that $W_0 = 0$.

Proposition 3.6. If W_t is a standard Brownian motion, $x + W_t$ is a Brownian motion starting from x for $x \in \mathbb{R}$.

Definition 3.7. Given d independent Brownian motions $B_t^1, \ldots, B_t^d$ we set

$$B_t = (B_t^1, \ldots, B_t^d)$$

and call it a d**-dimensional Brownian motion.**

Exam Questions 3.8 (Exercise)

Let $Z \sim \mathcal{N}(0,1)$ be a random variable. Is $X_t = Z\sqrt{t}$ a Brownian motion?

To do

Definition 3.9. We define the **finite dimensional distribution** of W_t as

$$\mathbb{P}(W_{t_1} \in A_1, \ldots, W_{t_n} \in A_n) = \int_{A_1} \cdots \int_{A_n} p(t_1, 0, x_1) p(t_2 - t_1, x_1, x_2) \cdots \\ \cdots p(t_n - t_{n-1}, x_{n-1}, x_n) \, dx_1 \cdots dx_n$$

where

$$p(t, x, y) = \frac{1}{\sqrt{2\pi t}} \exp\left(-\frac{(x-y)^2}{2t}\right)$$

is the **transition density** (or the *heat kernel*).

Note 3.10. A note on notation. Consider the set $A = (-1, 1)$. The notation

$$\mathbb{P}(W_t \in A) = \mathbb{P}(-1 \leq W_t \leq 1)$$

simply expresses the probability that W_t takes a value within the interval $(-1, 1)$.

More generally, for a random variable X_n with probability density function $p(x)$, the probability of X_n belonging to a set A is given by

$$\mathbb{P}(X_n \in A) = \int_A p(x) \, dx.$$

Lemma 3.11

For any $s, t > 0$

1. $\mathbb{E}(W_t \mid \mathcal{F}_s^W) = W_s$ almost surely.
2. $C(W_t, W_s) = s \wedge t$

Exam Questions 3.12 (Exercise)

Use the "independent increments" property of W_t to prove the above lemma. Show that $[W]_t = t$, almost surely.

To do

Exam Questions 3.13 (Exercise)

Let $z \in \mathbb{R}^d$ and B_t be a standard d-dimensional Brownian motion. Show that

$$Z_t = \exp\left(iz \cdot B_t + \frac{t}{2}\|z\|^2\right)$$

is a complex-valued martingale.

To do

Definition 3.14. An $\mathbb{R}$-valued process X_t, for $t \geq 0$ is called **Gaussian** if, for any integer $k \geq 1$ and any $t_1 < \cdots < t_k$, the vector random variable

$$(X_{t_1}, \ldots, X_{t_k})$$

has a multivariate normal distribution.

Remark 3.15. The finite-dimensional distributions of a Gaussian process are determined by its mean vector $m(t) = \mathbb{E}(X_t)$ and covariance function

$$\rho(s,t) = \mathbb{E}\left[(X_s - m(s))(X_t - m(t))^t\right].$$

Definition 3.16. If $m(t) = 0$, for all t, we say that X_t is a **zero-mean Gaussian process**.

Example 3.17

A one-dimensional Brownian motion is a zero-mean Gaussian process with covariance function

$$\rho(s,t) = s \wedge t.$$

Reciprocally, any zero-mean Gaussian process X_t for $t \geq 0$ with continuous paths and covariance function $\rho(s,t) = s \wedge t$ is a Brownian motion.

Definition 3.18. An adapted d-dimensional process $\{X_t, \mathcal{F}_t\}_{t\geq 0}$ on some probability space is said to be a **Markov process**, or to satisfy the **Markov property**, if for any $\Gamma \in \mathcal{B}(\mathbb{R}^d)$ and any $s, t \geq 0$,

$$\mathbb{P}(X_{t+s} \in \Gamma \mid \mathcal{F}_s) = \mathbb{P}(X_{t+s} \in \Gamma \mid X_s),$$

almost surely.

Note 3.19. A Markov process is a stochastic process where the future depends only on the present, not on the past.

Theorem 3.20 (Chapman-Kolmogorov equation)
Let $\{X_t, \mathcal{F}_t\}_{t\geq 0}$ be a Markov process with transition density function given by

$$\mathbb{P}(X_t \in A \mid X_s = x) = \int_A p(s,t,x,y)\,dy.$$

Then, for any $s < u < t$,

$$p(s,t,x,y) = \int_{\mathbb{R}} p(s,u,x,z)p(u,t,z,y)\,dz.$$

Remark 3.21. If X_t has a density for every t, then the Markov property can be expressed using the equation above.

Exam Questions 3.22 (Exercise)
Verify by direct computation that the Chapman-Kolmogorov equation holds for the Brownian motion transition density.

to do

3.2 Processes with finite p variation

Note 3.23. One important fact about Brownian motion sample paths is that *Brownian motion has paths of infinite variation almost surely.*

Definition 3.24. Let $f : [0,T] \to \mathbb{R}$ be a continuous real-valued function. A partition of the interval $[0,T]$, denoted by π, is a collection of points in $[0,T]$:

$$\pi = \{0 = t_0 < t_1 < \cdots < t_n = T\}.$$

Definition 3.25. The **mesh** of the partition (defined as above), denoted $|\pi|$, is defined as

$$|\pi| = \max_{1\leq i\leq n} |t_i - t_{i-1}|\,.$$

Definition 3.26. For $p > 0$, the **p-variation** of f over π (as above) is defined as

$$V_\pi^{(p)}(f) \triangleq \sum_{i=1}^{n} |f(t_i) - f(t_{i-1})|^p\,.$$

The total p-variation of f is then

$$V^{(p)}(f) \triangleq \lim_{|\pi|\to 0} V_\pi^{(p)}(f).$$

For $p = 1$, this is simply called the **variation** of f. A function is said to have **finite variation** if $V^{(1)}(f) < \infty$.

Note 3.27. The p-variation measures how much a function oscillates over time. Finite variation means the total variation remains bounded as the partition gets finer.

Definition 3.28. A sequence of random variables X_n **converges in probability** to X if, for any $\varepsilon > 0$ we have
$$\lim_{n\to\infty} \mathbb{P}(|X_n - X| > \varepsilon) = 0.$$

Lemma 3.29. Let $\{W_t, \mathcal{F}_t\}$ be a Brownian motion. Consider a family of partitions $\pi^n := \{0 = t_0 < t_1 < \cdots < t_n = t\}$, where $|\pi^n| \to 0$ as $n \to \infty$. Define the corresponding sum
$$S_n(t) = \sum_{i=0}^{n-1} (W_{t_{i+1}} - W_{t_i})^2.$$
Then,
$$S_n(t) \to t, \quad n \to \infty, \quad \text{in probability.}$$

Note 3.30. This means that quadratic variation of Brownian motion is finite.

Sketch of Proof. To show $S_n(t) \to t$ in L^2, observe that
$$S_n(t) - t = \sum_{i=0}^{n-1} \left((W_{t_{i+1}} - W_{t_i})^2 - (t_{i+1} - t_i) \right).$$
Denoting each term as X_i, these are independent with mean 0 but not identically distributed. Then,
$$\mathbb{E}\left[(S_n(t) - t)^2\right] = \sum_{i=1}^{n-1} (t_{i+1} - t_i) \mathbb{E}\left[\left(\frac{(W_{t_{i+1}} - W_{t_i})^2}{t_{i+1} - t_i} - 1 \right)^2 \right].$$
Since $\frac{(W_{t_{i+1}} - W_{t_i})^2}{t_{i+1} - t_i} - 1 \sim Z^2 - 1$ with $Z \sim \mathcal{N}(0,1)$, we obtain
$$\mathbb{E}\left[(S_n(t) - t)^2\right] \leq \mathbb{E}\left[|Z^2 - 1|^2\right] t \max_{1 \leq i \leq n} |t_{i+1} - t_i| \xrightarrow{n\to\infty} 0.$$
This proves convergence in L^2, which implies convergence in probability. To prove the almost sure converges requires the Borel-Cantelli lemma. $\square$

Lemma 3.31. Let $X_t, \mathcal{F}_t$ be a continuous process such that for each fixed $t > 0$ and $p > 0$,
$$\lim_{\|\pi\|\to 0} V_\pi^{(p)}(X_t) = L_t, \quad \text{in probability,}$$
where L_t is an almost surely, finite random variable. Then,

- If $q > p$, then $\lim_{\|\pi\|\to 0} V_\pi^{(q)} = 0$ in probability.
- If $0 < q < p$, then $\lim_{\|\pi\|\to 0} V_\pi^{(q)} = \infty$ in probability, on the event $\{L_t > 0\}$.

Note 3.32. If a process has finite quadratic variation but infinite first variation, it means it oscillates too much to be of finite variation. This means Brownian motion paths are *too rough* to define a standard Riemann integral.

Corollary 3.33. Brownian sample paths have infinite first variation almost surely

Proof. Combine the previous lemmas. □

3.3 Exercises

Exercise. (Exponential martingale). Let W_t be a one-dimensional $\mathcal{F}_t$ Brownian motion. Show that the process

$$\exp\left(\sigma W_t - \frac{\sigma^2}{2}t\right)$$

is a martingale. This is known as the **exponential martingale**.

To do

Exercise. (Using the Markov Property). Let W_t be a one-dimensional $\mathcal{F}_t$ Brownian motion, and let $A, B \in \mathcal{B}(\mathbb{R})$. Use the Markov property and the transition density of W_t to compute as explicitly as possible the probability

$$P(W_s \in A, W_t \in B).$$

Hint: Observe that

$$P(W_s \in A, W_t \in B) = \mathbb{E}(\mathbf{1}_A(W_s)1_B(W_t)).$$

Then, use conditioning and the Markov property.

To do

Exercise.(Linear Combination of Brownian Motions). Let $W_t, \widetilde{W}_t$ be two independent Brownian motions. Define

$$B_t = \rho W_t + \sqrt{1-\rho^2}\widetilde{W}_t, \quad -1 < \rho < 1.$$

Is B_t a Brownian motion?

To do

Exercise. (Geometric Brownian Motion and Digital Options). In the *Black-Scholes model*, the stock price S_t is assumed to follow a **Geometric Brownian Motion**:

$$S_t = S_0 \exp\left((r - \frac{\sigma^2}{2})t + \sigma B_t\right),$$

where:

- S_0 is the initial stock price,
- r is the interest rate,
- σ is the volatility,
- B_t is a standard Brownian motion.

Compute the probability

$$\mathbb{P}(S_T > K)$$

for $T, K > 0$, i.e., compute the price of a **digital option** (modulo discounting).

To do

4 Stochastic calculus

4.1 Motivation

Assume that the price of a stock S_t evolves as a martingale, but trading occurs only at discrete times $0 = t_0 < t_1 < \cdots < t_n$. We trade on this stock following a self-financing trading strategy $\{\phi_j\}_{j=0}^n$, where at time t_j we hold ϕ_j units of the stock. The gains or losses from trading up to time t_k are given by:

$$V_k = \sum_{j=1}^{k-1} \phi_j (S_{t_{j+1}} - S_{t_j}), \quad k = 1, \ldots, n-1.$$

Exam Questions 4.1 (Exercise)

Show that $\{V_k\}_{k=0}^n$ is a martingale.

To do

However, we would like to extend this procedure to continuous time, where sums are replaced by integrals and finite differences by differentials. That is, we take the limit of the above as $\max_j |t_{j+1} - t_j| \to 0$. If S_t had paths of finite variation, we could define the integral $\int \phi_t dS_t$ using Riemann-Stieltjes integration [1]

4.2 The stochastic integral w.r.t. Brownian motion

Definition 4.2. A stochastic process $(X_t)_{t\geq 0}$ is called $\mathcal{F}_t$**-previsible** or $\mathcal{F}_t$**-predictable** if the mapping

$$(t, \omega) \in \mathbb{R}_+ \times \Omega \mapsto X_t(\omega)$$

is measurable with respect to the σ-field $\mathcal{S}$, where $\mathcal{S}$ is the σ-field generated by all left-continuous, $\mathcal{F}_t$-adapted processes.

Note 4.3. A process is predictable if its value at time t can be determined using past information.

Remark 4.4. If a process is adapted and left-continuous, then it is also predictable. In particular, all adapted continuous processes are predictable.

Since the classical Riemann-Stieltjes integral cannot handle Brownian motion as an integrator, we develop Itô's stochastic integral.

[1] The Riemann-Stieltjes integral of a function f with respect to another function g over $[a, b]$ is defined as:

$$\int_a^b f(x)\, dg(x) = \lim_{\|\pi\|\to 0} \sum_i f(x_i^*)(g(x_{i+1}) - g(x_i))$$

where the sum is taken over a partition π of $[a, b]$ and x_i^* is a sample point in each subinterval. It generalises the Riemann integral, where $g(x) = x$.

4.4 Definition (Simple Processes)

Definition 4.5. Let $\{t_i\}_{i\geq 1}$ be a strictly increasing sequence of non-negative numbers such that $\lim_n t_n = +\infty$. A **simple process** takes the form

$$X_t(\omega) = \sum_{i=1}^{\infty} a_i(\omega)\mathbf{1}_{(t_i,t_{i+1}]}(t),$$

where a_i is a square-integrable, $\mathcal{F}_i$-measurable random variable.

Note 4.6. Simple processes take constant values over intervals.

Definition 4.7. If X_s is a simple random variable, then so is $X_s 1_{[0,t]}(s)$. The **stochastic integral for simple processes** is defined as:

$$\int_0^t X_s \, dW_s \triangleq \sum_{i=1}^{\infty} a_i(\omega)(W_{s_{i+1}\wedge t} - W_{s_i \wedge t}).$$

Lemma 4.8

Let X_t be a simple process. Then:

1. The process $\int_0^t X_s \, dW_s$ is a continuous martingale with respect to $\mathcal{F}_t$.
2. $\mathbb{E}\left[\left(\int_0^t X_s dW_s\right)^2\right] = \mathbb{E}\left[\int_0^t X_s^2 ds\right]$.
3. $\mathbb{E}\left[\sup_{t\leq T}\left(\int_0^t X_s dW_s\right)^2\right] \leq 4\mathbb{E}\left[\int_0^T X_s^2 ds\right]$.

Remark 4.9. We have that:

- The stochastic integral is a martingale.
- The second result is known as Itô isometry.
- The third result is a version of *Doob's inequality.*

Exam Questions 4.10 (Exercise)

Prove the first assertion of the lemma.

To do

Definition 4.11 (Class of integrable processes)**.** Define $\mathcal{H}_T$ as the set of processes $X : \mathbb{R}_+ \times \Omega \to \mathbb{R}$ such that X_t is $\mathcal{F}_t$-previsible and satisfies:

$$\mathbb{E}\left[\int_0^T X_t^2 dt\right] < \infty.$$

The space $\mathcal{H}$ consists of all processes in $\mathcal{H}_T$ for any T.

Definition 4.12. For any $X \in \mathcal{H}_T$, there exists a sequence of simple processes X^n such that:

$$\mathbb{E}\left[\int_0^T |X_t - X_t^n| \, dt\right] \xrightarrow{n\to\infty} 0.$$

We define the **stochastic integral**:

$$\int_0^T X_t \, dW_t \triangleq \lim_{n\to\infty} \int_0^T X_t^n \, dW_t.$$

Theorem 4.13

Suppose $\{W_t, \mathcal{F}_t\}$ is a Brownian motion and X_t is a previsible process in $\mathcal{H}_T$. Then there exists a linear mapping $J : \mathcal{H} \to \mathcal{M}_c^2$ such that:

1. If X_t is simple,
$$J(X)_t = \int_0^t X_s dW_s.$$
2. For $t \leq T$,
$$\mathbb{E}\left[J(X)_t^2\right] = \mathbb{E}\left[\int_0^t X_s^2 ds\right].$$
3. We have
$$\mathbb{E}\left[\sup_{t\leq T} J(X)_t^2\right] \leq 4\mathbb{E}\left[\int_0^T X_s^2 ds\right].$$

Note 4.14. The above theorem confirms the construction of the stochastic integral.

Definition 4.15. For $X \in \mathcal{H}_T$, we write:

$$J(X)_T = \int_0^T X_t \, dW_t$$

and call this the **stochastic integral of X_t against W_t**

4.3 Extensions

4.3.1 Integration against martingales

Assume that $M_t \in \mathcal{M}_c^2$, where M_t is a martingale with respect to some filtration $\mathcal{F}_t$. Working in analogy with the Brownian motion case, we define the integral for a simple integrand X_t as:

$$X_t = \sum_{i=1}^n a_i(\omega)\mathbf{1}_{I_i}(t)$$

and set

$$\int_0^t X_s dM_s \triangleq \sum_{i=1}^n a_i(\omega)(M_{s_{i+1}\wedge t} - M_{s_i\wedge t}).$$

Passing to the limit, we define

$$J^M(X)_t = \int_0^t X_s dM_s.$$

For any process $X \in \mathcal{H}_T^M$, where $\mathcal{H}_T^M$ is the space of all predictable processes such that

$$\mathbb{E}\left[\int_0^T X_s^2 d[M]_s\right] < \infty,$$

the integral $J^M(X)$ is extended analogously to the theorem above.

Theorem 4.16 (Stochastic Integral Against a Martingale)
Suppose that M_t is a martingale with respect to a filtration $\mathcal{F}_t$, and X_t is a predictable process in $\mathcal{H}_T^M$. There exists a linear mapping $J^M : \mathcal{H}_T^M \to \mathcal{M}_c^2$ such that:

1. If X_t is simple and $t \leq T$,

$$J^M(X)_t = \int_0^t X_s dM_s.$$

2. For $t \leq T$,

$$\mathbb{E}\left[J^M(X)_t^2\right] = \mathbb{E}\left[\int_0^t X_s^2 d[M]_s\right],$$

where $[M]$ is the quadratic variation process associated with M_t.

3. We have

$$\mathbb{E}\left[\sup_{t\leq T} J^M(X)_t^2\right] \leq 4\mathbb{E}\left[\int_0^T X_s^2 d[M]_s\right].$$

Definition 4.17. For $X \in \mathcal{H}_T^M$, we define

$$J^M(X)_T = \int_0^T X_t dM_t$$

and call this the **stochastic integral of X_t against M_t**.

4.3.2 Integration Against Semimartingales

Extending integration beyond martingales, we consider semimartingales as integrators. Since we have defined integrals against martingales, the Riemann-Stieltjes integral handles integrals against finite-variation processes.

Definition 4.18. If $X_t = M_t + A_t$ is a semimartingale, where M_t is a martingale and A_t has finite variation, then the integral $\int Y_t dX_t$ decomposes as:

$$\int_0^t Y_s dX_s \triangleq \int_0^t Y_s dM_s + \int_0^t Y_s dA_s.$$

Here:

- The first term is a stochastic integral.
- The second term is a Riemann-Stieltjes integral.

Remark 4.19. With the use of stopping times, one can localise the construction of stochastic integrals and consider integrators from the more general class of local martingales.

4.4 Itô's formula

Note 4.20. Itô's formula is the stochastic analogue of the chain rule.

Theorem 4.21 (Itô's formula)
Let $f : \mathbb{R}_+ \times \mathbb{R}^d \to \mathbb{R}$ be a function with partial derivatives $\frac{\partial}{\partial x} f$, $\frac{\partial}{\partial t} f$, and $\frac{\partial^2}{\partial x^2} f$ that are continuous, and assume $\frac{\partial}{\partial x} f(t, M_t) \in \mathcal{H}$. Given a d-dimensional martingale $M \in \mathcal{M}_c^2$, Itô's formula states:

$$f(t, M_t) = f(0, M_0) + \int_0^t \sum_{i=1}^d \frac{\partial}{\partial x_i} f(s, M_s) dM_s^i + \int_0^t \frac{\partial}{\partial t} f(s, M_s) ds + \frac{1}{2} \int_0^t \sum_{i,j=1}^d \frac{\partial^2}{\partial x_i \partial x_j} f(s, M_s) d[M^i, M^j]_s.$$

Remark 4.22. The formula is often written in differential form:

$$df(t, M_t) = \sum_{i=1}^d \frac{\partial}{\partial x_i} f(t, M_t) dM_t^i + \frac{\partial}{\partial t} f(t, M_t) dt + \frac{1}{2} \sum_{i,j=1}^d \frac{\partial^2}{\partial x_i \partial x_j} f(t, M_t) d[M^i, M^j]_t.$$

Here, $X_t dM_t$ is a formal notation representing $dZ_t = X_t dM_t$ where:

$$Z_t = Z_0 + \int_0^t X_s dM_s.$$

Note 4.23. This is the general form of Itô's formula. In the one-dimensional case, the quadratic variation satisfies $d[W]_t = dt$, eliminating the need for summation symbols. Thus, we can simplify it to:

$$df(M_t) = f'(M_t)\, dM_t + \frac{1}{2} f''(M_t)\, dt.$$

Example 4.24

We use Itô's formula to compute $\mathbb{E}(\cos(W_t))$, where W_t is a standard Brownian motion. Applying Itô's formula to $\cos(W_t)$, where W_t is standard Brownian motion, we use the general form

$$df(W_t) = f'(W_t)dW_t + \frac{1}{2}f''(W_t)dt.$$

For $f(x) = \cos(x)$, we compute its derivatives:

$$f'(x) = -\sin(x), \quad f''(x) = -\cos(x).$$

Substituting into Itô's formula gives

$$d\cos(W_t) = -\sin(W_t)dW_t - \frac{1}{2}\cos(W_t)dt.$$

Integrating both sides from 0 to t,

$$\cos(W_t) = \cos(W_0) + \int_0^t -\sin(W_s)dW_s - \frac{1}{2}\int_0^t \cos(W_s)ds.$$

Since $W_0 = 0$, we have $\cos(W_0) = 1$, so the equation simplifies to

$$\cos(W_t) = 1 + \int_0^t -\sin(W_s)dW_s - \frac{1}{2}\int_0^t \cos(W_s)ds.$$

Taking expectations on both sides,

$$\mathbb{E}[\cos(W_t)] = \mathbb{E}\left[1 + \int_0^t -\sin(W_s)dW_s - \frac{1}{2}\int_0^t \cos(W_s)ds\right].$$

By the martingale property of stochastic integrals, the expectation of $\int_0^t -\sin(W_s)dW_s$ is zero, giving

$$\mathbb{E}[\cos(W_t)] = 1 - \frac{1}{2}\int_0^t \mathbb{E}[\cos(W_s)]ds.$$

This is now an integral equation of the form

$$y(t) = 1 - \frac{1}{2}\int_0^t y(s)ds.$$

Differentiating both sides with respect to t,

$$\frac{d}{dt}y(t) = -\frac{1}{2}y(t).$$

This is a first-order linear differential equation with general solution

$$y(t) = Ce^{-t/2}.$$

Using the initial condition $y(0) = 1$, we find $C = 1$, so

$$\mathbb{E}[\cos(W_t)] = e^{-t/2}.$$

Example 4.25

We now show without using properties of the normal distribution, that $\mathbb{E}(W_t^3) = 0$. Now applying Itô's formula to $f(x) = x^3$, we differentiate:

$$f'(x) = 3x^2, \quad f''(x) = 6x.$$

Using Itô's formula,

$$dW_t^3 = 3W_t^2 dW_t + \frac{1}{2} \cdot 6W_t dt = 3W_t^2 dW_t + 3W_t dt.$$

Integrating from 0 to t,

$$W_t^3 = W_0^3 + \int_0^t 3W_s^2 dW_s + \int_0^t 3W_s ds.$$

Since $W_0 = 0$, this simplifies to

$$W_t^3 = \int_0^t 3W_s^2 dW_s + 3\int_0^t W_s ds.$$

Taking expectations,

$$\mathbb{E}[W_t^3] = \mathbb{E}\left[\int_0^t 3W_s^2 dW_s\right] + 3\mathbb{E}\left[\int_0^t W_s ds\right].$$

Again, by the martingale property,

$$\mathbb{E}\left[\int_0^t 3W_s^2 dW_s\right] = 0.$$

Thus,

$$\mathbb{E}[W_t^3] = 3\int_0^t \mathbb{E}[W_s] ds.$$

Since $\mathbb{E}[W_s - W_0] = \mathbb{E}[E_s] = 0$ for all s (by the defining properties of Brownian motion), we conclude

$$\mathbb{E}[W_t^3] = 0.$$

Exam Questions 4.26 (Exercise)
Assume S_t follows geometric Brownian motion:

$$S_t = S_0 \exp\left(\left(\nu - \frac{1}{2}\sigma^2\right)t + \sigma W_t\right).$$

Using Itô's formula, show that:

$$\mathbb{E}[S_t] = S_0 e^{\nu t}.$$

To do

Theorem 4.27 (Itô's Formula for Semimartingales)
For semimartingales $X_t^1, \dots, X_t^d$ and $f : \mathbb{R}_+ \times \mathbb{R}^d \to \mathbb{R}$ with appropriate smoothness, we have:

$$\begin{aligned} f(t, X_t) = f(0, X_0) + \int_0^t \frac{\partial}{\partial t} f(s, X_s) ds + \int_0^t \sum_{i=1}^d \frac{\partial}{\partial x_i} f(s, X_s) dX_s^i \\ + \frac{1}{2} \sum_{i,j=1}^d \int_0^t \frac{\partial^2}{\partial x_i \partial x_j} f(s, X_s) d[X^i, X^j]_s. \end{aligned}$$

4.5 Stochastic Integration by Parts

Theorem 4.28 (Stochastic integration by parts formula for semimartingale)
For semimartingales X_t^1, X_t^2:

$$X_t^1 X_t^2 = X_0^1 X_0^2 + \int_0^t X_s^1 dX_s^2 + \int_0^t X_s^2 dX_s^1 + [X^1, X^2]_t.$$

Example 4.29
For a geometric Brownian motion satisfying

$$dS_t = \mu S_t dt + \sigma S_t dW_t,$$

we compute $\mathbb{E}[S_t^n]$ using Itô's formula for $f(S_t) = S_t^n$. The first and second derivatives are

$$\frac{\partial}{\partial x} f(x) = nx^{n-1}, \quad \frac{\partial^2}{\partial x^2} f(x) = n(n-1)x^{n-2}.$$

Applying Itô's formula,

$$dS_t^n = nS_t^{n-1} dS_t + \frac{1}{2} n(n-1) S_t^{n-2} d[S]_t.$$

Since $d[S]_t = \sigma^2 S_t^2 dt$, substituting dS_t from the original SDE,

$$dS_t^n = nS_t^{n-1}(\mu S_t dt + \sigma S_t dW_t) + \frac{1}{2} n(n-1) S_t^{n-2} \sigma^2 S_t^2 dt.$$

Expanding and simplifying,

$$dS_t^n = S_t^n \left(n\mu + \frac{1}{2} n(n-1)\sigma^2 \right) dt + n\sigma S_t^n dW_t.$$

Taking expectations, the stochastic integral vanishes due to the martingale property, leaving

$$\frac{d}{dt}\mathbb{E}[S_t^n] = \left(n\mu + \frac{1}{2} n(n-1)\sigma^2 \right) \mathbb{E}[S_t^n].$$

This is a first-order linear ODE with solution

$$\mathbb{E}[S_t^n] = S_0^n \exp\left[\left(n\mu + \frac{1}{2} n(n-1)\sigma^2 \right) t \right].$$

The additional term $\frac{1}{2} n(n-1)\sigma^2$ arises from the quadratic variation of S_t, accounting for the effects of stochastic fluctuations beyond the deterministic drift.

How to compute quadratic variation in the example above

4.6 Martingale Representations

Theorem 4.30 (Lévy Characterization of Brownian Motion)
Let M_t be a continuous $\mathcal{F}_t$-adapted d-dimensional martingale such that $M_0 = 0$ and $[M]_t = tI$ for $t \geq 0$. Then M_t is a $\mathcal{F}_t$-adapted d-dimensional Brownian motion.

Exam Questions 4.31 (Exercise)
Let W_t be d-dimensional Brownian motion and A a $d \times d$ orthogonal matrix. Show that $B_t = AW_t$ is also d-dimensional Brownian motion.
To do

Theorem 4.32 (Martingale representation)
Let W_t be a Brownian motion and M_t a square-integrable martingale with respect to $\mathcal{F}_t$. Then there exist adapted processes $\phi_t^1, \ldots, \phi_t^d$ such that:

$$\mathbb{E}\left[\int_0^t (\phi_s^i)^2 ds\right] < \infty, \quad \forall i,$$

and

$$M_t = M_0 + \sum_{i=1}^{d} \int_0^t \phi_s^i dW_s^i, \quad a.s.$$

These processes are essentially unique, and the result extends to local martingales.

Note 4.33. In short, this theorem states that any square-integrable martingale M_t can be expressed as a stochastic integral with respect to Brownian motion. This means that Brownian motion 'generates' all martingales, in the sense that every martingale can be written as an integral of some adapted process ϕ_t against Brownian motion.

5 Stochastic differential equations

5.1 Definitions

Definition 5.1. On a finite time horizon $T > 0$, we consider two functions:

$$b : [0, T] \times \mathbb{R}^d \to \mathbb{R}^d, \quad \sigma : [0, T] \times \mathbb{R}^d \to \mathbb{R}^{d \times d}$$

which are assumed to be sufficiently well-behaved. Additionally, we consider a d-dimensional Brownian motion W_t, and define the SDE:

$$dZ_t = b(t, Z_t)dt + \sigma(t, Z_t)dW_t.$$

An equation of this form with deterministic functions b and σ is called a **stochastic differential equation** for Z_t. The function b is called the **drift**, and σ is referred to as the **diffusion** or **volatility matrix**.

Remark 5.2. A solution to an SDE is always adapted to the filtration generated by the driving Brownian motion.

Proposition 5.3. Componentwise, the SDE above can be rewritten as:

$$dZ_t^i = b^i(t, Z_t)dt + \sum_{j=1}^{d} \sigma_{ij}(t, Z_t)dW_t^j, \quad i = 1, 2, \ldots, d.$$

Here, $\sigma_{ij}(t,x)$ represents the (i,j)-th entry of the matrix-valued function $\sigma(t,x)$.

Example 5.4

Consider the stochastic differential equation (SDE):

$$dX_t = X_t^3 dt + X_t^2 dW_t, \quad X_0 = 1.$$

We seek a solution of the form $X_t = f(t, W_t)$. Applying Itô's formula, we compute:

$$dX_t = \left(f_t(t, W_t) + \frac{1}{2} f_{xx}(t, W_t) \right) dt + f_x(t, W_t) dW_t.$$

Substituting into the given SDE, we equate terms:

$$\left(f_t(t, W_t) + \frac{1}{2} f_{xx}(t, W_t) \right) dt + f_x(t, W_t) dW_t = f(t, W_t)^3 dt + f(t, W_t)^2 dW_t.$$

Matching the coefficients of dW_t gives:

$$f_x(t, W_t) = f(t, W_t)^2.$$

Similarly, matching the coefficients of dt leads to:

$$f_t(t, W_t) + \frac{1}{2} f_{xx}(t, W_t) = f(t, W_t)^3.$$

Solving the first equation as a separable differential equation:

$$\frac{df}{dx} = f^2 \quad \Rightarrow \quad \int \frac{1}{f^2} df = \int dx,$$

which integrates to:

$$-\frac{1}{f} = x + C.$$

Rearranging gives:

$$f(t, x) = \frac{1}{-x + C}.$$

Using the initial condition $X_0 = 1$, we solve for C. Setting $f(0,0) = 1$, we obtain $C = 1$, leading to the final solution:

$$X_t = \frac{1}{1 - W_t}.$$

This solution exhibits a finite-time explosion, as Brownian motion W_t will almost surely hit the value 1 in finite time, causing the denominator to become zero.

5.2 Existence and uniqueness results

Note 5.5. A natural question that arises when studying stochastic differential equations (SDEs) is what kind of restrictions on the drift term $b(t,x)$ and diffusion term $\sigma(t,x)$ ensure that solutions exist and are unique.

Definition 5.6. A function $f : \mathbb{R}^d \to \mathbb{R}$ is called **Lipschitz continuous** if there exists a constant $K > 0$ such that for all $x, y \in \mathbb{R}^d$,

$$|f(x) - f(y)| \leq K\,|x - y|\,,$$

where $|x|$ is the Euclidean norm in $\mathbb{R}^d$.

Theorem 5.7 (Existence and Uniqueness Theorem)
Suppose the functions $b(t,x)$ and $\sigma(t,x)$ satisfy the following conditions for some positive constant K:

$$|b(t,x) - b(t,y)| + \|\sigma(t,x) - \sigma(t,y)\| \leq K|x - y|,$$

$$|b(t,x)|^2 + \|\sigma(t,x)\|^2 \leq K(1 + |x|^2).$$

for all $t \geq 0$ and $x, y \in \mathbb{R}^d$, where

$$\|\sigma\|^2 = \sum_{i,j=1}^{d} \sigma_{ij}^2.$$

Then there exists a unique process X_t adapted to the filtration of W_t that solves the SDE:

$$dX_t = b(t, X_t)dt + \sigma(t, X_t)dW_t.$$

with initial condition $X_0 = x$.

Moreover, the solution is square-integrable, meaning that for some constant $C > 0$ depending on K and T,

$$\mathbb{E}[|X_t|^2] \leq C(1 + |x|^2)e^{Ct}, \quad 0 \leq t \leq T.$$

Remark 5.8. The conclusion of the theorem remains valid if we substitue the initial condition $X_0 = x$ by $X_0 = \xi$ where ξ is an independent from W random variables with finite second moment. In this case the estimate for the second moment of the solution becomes

$$\mathbb{E}\left(|X_t|^2\right) \leq C\left(1 + \mathbb{E}\left(|\xi|^2\right)\right)e^{Ct}$$

for $T \in [0, T]$.

Example 5.9 (The Ornstein-Uhlenbeck Process)
One of the most well-known SDEs is:

$$dX_t = -aX_t dt + \sigma dW_t, \quad a, \sigma > 0.$$

The solution is obtained using integration:

$$X_t = X_0 e^{-at} + \sigma \int_0^t e^{-a(t-s)} dW_s, \quad 0 \leq t.$$

This is known as the Ornstein-Uhlenbeck (O-U) process. The mean is given by:

$$m(t) = \mathbb{E}[X_0]e^{-at} = m(0)e^{-at}.$$

For the variance:

$$V(t) = \mathbb{E}[X_t^2] - \mathbb{E}[X_t]^2 = \mathbb{E}[X_0^2]e^{-2at} + \sigma^2 \mathbb{E}\left[\int_0^t e^{-2a(t-s)} ds\right].$$

Using the orthogonality of X_0 and W_t, this simplifies to:

$$V(t) = \frac{\sigma^2}{2a} + \left(V(0) - \frac{\sigma^2}{2a}\right) e^{-2at}.$$

If the initial condition X_0 follows a normal distribution with mean zero and variance $\sigma^2/(2a)$, the process remains stationary with covariance function:

$$C(X_s, X_t) = \frac{\sigma^2}{2a} e^{-a|t-s|}.$$

Exam Questions 5.10 (Exercise)
Show that the covariance of an O-U process is given by

$$C(X_s, X_t) = \left[V(0) + \frac{\sigma^2}{2a}\left(e^{2at\wedge s} - 1\right)\right] e^{-a(t+s)}.$$

TO do

Remark 5.11. If the initial random variable $X_0 \sim \mathcal{N}\left(0, \frac{\sigma^2}{2a}\right)$, then the O-U process X is a stationary zero-mean Gaussian process with covariance function $C(X_s, X_t) = \left(\frac{\sigma^2}{2a}\right) e^{-a|t-s|}$.

Exam Questions 5.12 (Exercise)
Suppose that there exists a real-valued function $u(t,y)$ such that

$$\frac{\partial u}{\partial t}(t,y) = b(t,u(t,y)), \quad \frac{\partial u}{\partial y}(t,y) = \sigma(t,u(t,y))$$

where $b(t,x)$ is Lipschitz continuous and $\sigma(t,x) \in C^{1,2}([0,\infty) \times \mathbb{R})$, meaning it is once continuously differentiable in t and twice in x. Show that the process

$$X_t = u(t, W_t)$$

solves the stochastic differential equation

$$dX_t = \left(b(t,X_t) + \frac{1}{2}\sigma(t,X_t)\frac{\partial \sigma}{\partial x}(t,X_t) \right) dt + \sigma(t,X_t)dW_t.$$

To do

Exam Questions 5.13 (Exercise)
Solve the SDE:

$$dX_t = \left(\sqrt{1+X_t^2} + \frac{1}{2}X_t \right) dt + \sqrt{1+X_t^2} dW_t,$$

where $X_0 = x$. Using the ansatz:

$$X_t = \sinh(C + t + W_t),$$

where $\sinh(x) = \frac{e^x - e^{-x}}{2}$,

to do

5.3 Girsanov's theorem

Theorem 5.14 (Girsanov's Theorem)
Suppose that $\{W_t\}_{t\geq 0}$ is a $\mathbb{P}$-Brownian motion with a natural filtration $\{\mathcal{F}_t\}_{t\geq 0}$ and that $\{\theta_t\}_{t\geq 0}$ is an $\{\mathcal{F}_t\}_{t\geq 0}$-adapted process such that

$$\mathbb{E}\left[\exp\left(\frac{1}{2}\int_0^T \theta_t^2 dt\right)\right] < \infty.$$

Define

$$L_t \triangleq \exp\left(-\int_0^t \theta_s dW_s - \frac{1}{2}\int_0^t \theta_s^2 ds\right).$$

Let $\mathbb{Q}$ be the probability measure defined by

$$\mathbb{Q}[A] \triangleq \int_A L_t d\mathbb{P}, \quad A \in \mathcal{F}_t.$$

Then under the probability measure $\mathbb{Q}$, the process $\{\overline{W}_t\}_{0\leq t\leq T}$ defined by

$$\overline{W}_t = W_t + \int_0^t \theta_s ds$$

is a Brownian motion. The process L_t is often called the **Girsanov kernel**.

Note 5.15. This result tells us how to transform a probability measure so that a Brownian motion with drift becomes a standard Brownian motion under the new measure. The key idea is that by adjusting the drift of the Brownian motion via the process θ_t, we define a new probability measure $\mathbb{Q}$ under which the drift disappears. This is particularly useful in stochastic control and finance, where it allows changing the measure to simplify computations (e.g., risk-neutral pricing).

Remark 5.16. Some remarks.

1. When
$$\mathbb{E}\left[\exp\left(\frac{1}{2}\int_0^T \theta_t^2 dt\right)\right] < \infty$$
the process L_t is a martingale, called the *exponential martingale*. This is known as the **Novikov condition**. Since L_t is trivially positive and $\mathbb{E}[L_t] = 1$ for all t, $\mathbb{Q}$ is a probability measure.

2. Two probability measures $\mathbb{P}_1$, $\mathbb{P}_2$ are said to be equivalent if $\mathbb{P}_1(A) = 0$ if and only if $\mathbb{P}_2(A) = 0$. In this sense, $\mathbb{P}$ and $\mathbb{Q}$ are equivalent.

3. When passing from one measure to another, Girsanov's theorem gives, for any bounded $\mathcal{F}_t$-measurable h_t:
$$\mathbb{E}^{\mathbb{Q}}[h_t] = \mathbb{E}^{\mathbb{P}}[h_t L_t]$$
and more generally,
$$\mathbb{E}^{\mathbb{Q}}[h_t|\mathcal{F}_s] = \mathbb{E}^{\mathbb{P}}\left[h_t\frac{L_t}{L_s} \mid \mathcal{F}_s\right] = \frac{1}{L_s}\mathbb{E}^{\mathbb{P}}[h_t L_t|\mathcal{F}_s],$$

as L is adapted.

Proposition 5.17

Let W_t be a standard Brownian motion and let θ_t be an adapted process satisfying appropriate measurability conditions. The **Novikov condition** states that if:

$$\mathbb{E}\left[\exp\left(\frac{1}{2}\int_0^T \theta_s^2 ds\right)\right] < \infty,$$

then the stochastic exponential:

$$Z_t = \exp\left(-\int_0^t \theta_s dW_s - \frac{1}{2}\int_0^t \theta_s^2 ds\right)$$

is a martingale for $t \in [0, T]$.

Example 5.18

Consider the process

$$B_t = \sigma W_t + \mu t.$$

This is a Brownian motion with a drift, where σ, μ are fixed constants and W_t is a $\mathbb{P}$-Brownian motion. Set $\theta = \mu/\sigma$ and define

$$L_t = \exp\left(-\theta W_t - \frac{1}{2}\theta^2 t\right), \quad \mathbb{Q}[A] = \mathbb{P}[\mathbf{1}_A L_t].$$

Then

$$W_t^L = W_t + \frac{\mu}{\sigma}t$$

is a $\mathbb{Q}$-Brownian motion, and $B_t = \sigma W_t^L$ is a scaled Brownian motion under $\mathbb{Q}$. Of course, L_t is also a $\mathbb{Q}$-martingale.

Notice that under $\mathbb{P}$:

$$\mathbb{E}^{\mathbb{P}}[B_t^2] = \mathbb{E}^{\mathbb{P}}[\sigma^2 W_t^2 + 2\sigma\mu t W_t + \mu^2 t^2] = \sigma^2 t + \mu^2 t^2.$$

Whereas under $\mathbb{Q}$:

$$\mathbb{E}^{\mathbb{Q}}[B_t^2] = \mathbb{E}^{\mathbb{Q}}[\sigma^2 W_t^2] = \sigma^2 t.$$

6 Conditional Expectation

Definition 6.1. Let ξ be an integrable $\mathcal{F}$-measurable random variable and let $\mathcal{G}$ be a sub σ-algebra of $\mathcal{F}$, i.e. $\mathcal{G} \subseteq \mathcal{F}$. The **conditional expectation** of ξ given $\mathcal{G}$ is defined as the (unique) random variable $\mathbb{E}[\xi \mid \mathcal{G}]$ satisfying:

1. It is integrable: $\mathbb{E}(|\mathbb{E}[\xi \mid \mathcal{G}]|) < \infty$.
2. $\mathbb{E}[\xi \mid \mathcal{G}]$ is $\mathcal{G}$-measurable.

3. For any $A \in \mathcal{G}$,
$$\int_A \mathbb{E}[\xi \mid \mathcal{G}]dP = \int_A \xi dP.$$

Note 6.2. Heuristically, we can think of $\mathbb{E}[\xi \mid \mathcal{G}]$ as the best approximation of ξ given the information contained in $\mathcal{G}$. If $\xi \in L^2$, then $\mathbb{E}[\xi \mid \mathcal{G}]$ is the best least-squares predictor of ξ among all $\mathcal{G}$-measurable random variables Y, meaning it minimizes $\mathbb{E}[(\xi - Y)^2]$.

6.1 Properties of Conditional Expectation

Theorem 6.3. We list the properties of the conditional expectation.

1. **Linearity.** For any real numbers a, b and any two random variables ξ, ζ,
$$\mathbb{E}[a\xi + b\zeta \mid \mathcal{G}] = a\mathbb{E}[\xi \mid \mathcal{G}] + b\mathbb{E}[\zeta \mid \mathcal{G}].$$
2. $\mathbb{E}[\mathbb{E}[\xi \mid \mathcal{G}]] = \mathbb{E}[\xi]$.
3. **Taking out what is known.** If ξ is $\mathcal{G}$-measurable, then
$$\mathbb{E}[\xi \mid \mathcal{G}] = \xi \quad \text{a.s.}$$
Intuition: If ξ is already determined by $\mathcal{G}$, then conditioning on $\mathcal{G}$ does nothing.
4. If ξ is independent of $\mathcal{G}$, then
$$\mathbb{E}[\xi \mid \mathcal{G}] = \mathbb{E}[\xi].$$
5. **Tower property.** If $\mathcal{H} \subseteq \mathcal{G}$, then
$$\mathbb{E}[\mathbb{E}[\xi \mid \mathcal{G}] \mid \mathcal{H}] = \mathbb{E}[\xi \mid \mathcal{H}].$$
Intuition: Conditioning stepwise still results in the same expectation when conditioning on a smaller σ-algebra.
6. **Positivity.** If $\xi \geq 0$ almost surely, then
$$\mathbb{E}[\xi \mid \mathcal{G}] \geq 0 \quad \text{a.s.}$$
Intuition: Expectation respects order; conditioning does not introduce negativity.
7. **Conditional Jensen's inequality.** If $h : \mathbb{R} \to \mathbb{R}$ is convex, then
$$h(\mathbb{E}[\xi \mid \mathcal{G}]) \leq \mathbb{E}[h(\xi) \mid \mathcal{G}], \quad \text{a.s.}$$
8. **Conditional Monotone Convergence.** If $0 \leq \xi_n \uparrow \xi$, then
$$\mathbb{E}[\xi_n \mid \mathcal{G}] \uparrow \mathbb{E}[\xi \mid \mathcal{G}].$$
Intuition: Like the usual monotone convergence theorem, this holds even when conditioning.
9. **Conditional Dominated Convergence.** If $\mid \xi_n \mid \leq Y$ for all n, $\mathbb{E}[Y] < \infty$, and $\xi_n \to \xi$, then
$$\mathbb{E}[\xi_n \mid \mathcal{G}] \to \mathbb{E}[\xi \mid \mathcal{G}] \quad \text{a.s.}$$
Intuition: Just like the standard Dominated Convergence Theorem, but in the conditional setting.

7 The filtering framework

7.1 The signal process

In many real-world applications, we are interested in tracking the evolution of a system over time. A standard approach is to describe the system using a dynamical equation. If the system follows deterministic laws, its evolution can be described by an ordinary differential equation (ODE):

$$\frac{dX_t}{dt} = f(X_t).$$

This equation describes how X_t changes over time, where:

- $f(X_t)$ is the drift (system's natural tendency to evolve in a particular direction).
- Given an initial condition X_0, the ODE solution fully determines X_t for all t.

However, in many practical scenarios, real-world systems experience random fluctuations due to uncertainty, environmental noise, or measurement errors. To account for this randomness, we modify the ODE by adding a stochastic perturbation:

$$\frac{dX_t}{dt} = f(X_t) + \text{“noise”}.$$

Introducing Stochasticity: From ODEs to SDEs

A more refined way to introduce randomness is to scale the noise based on the state X_t using a function $\sigma(X_t)$, leading to:

$$\frac{dX_t}{dt} = f(X_t) + \sigma(X_t) \cdot \text{”noise”}.$$

Here, $\sigma(X_t)$ represents the magnitude of randomness (e.g., if $\sigma(X_t)$ is large, the system is highly volatile). A natural mathematical model for noise is Brownian motion V_t, whose derivative is often informally written as “white noise”:

$$\frac{dX_t}{dt} = f(X_t) + \sigma(X_t)\frac{dV_t}{dt}.$$

However, since Brownian motion is nowhere differentiable, the term $\frac{dV_t}{dt}$ does not make rigorous sense. Instead, we define the equation in an integral form, leading to a stochastic differential equation (SDE):

$$X_t = X_0 + \int_0^t f(X_s)ds + \int_0^t \sigma(X_s)dV_s.$$

Definition 7.1. A function $f : \mathbb{R} \to \mathbb{R}$ (or more generally $f : \mathbb{R}^n \to \mathbb{R}^m$) is said to be **Lipschitz continuous** if there exists a constant $L > 0$ such that for all $x, y \in \mathbb{R}$:

$$|f(x) - f(y)| \le L\,|x - y|\,.$$

Definition 7.2. Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space with filtration $\mathcal{F}_t$. Let $X = \{X_t : t \ge 0\}$ be an $\mathcal{F}_t$-adapted process that solves the SDE

$$X_t = X_0 + \int_0^t f(X_s)ds + \int_0^t \sigma(X_s)dV_s.$$

where

- X_0 is the **initial value of the signal**;
- $V = \{V_t : t \geq 0\}$ is a Brownian motion independent of X_0;
- $f, g : \mathbb{R} \to \mathbb{R}$ are bounded Lipschitz continuous functions;
- f is the **drift coefficient** ;
- σ is the **diffusion** (or **volatility**) coefficient.

Then we call X the **signal process**.

7.2 The observation process

In real-world applications, we often do not observe the true state X_t of a system directly. Instead, we obtain 'noisy measurements' of X_t through an observation process Y_t, which is affected by sensor inaccuracies and random disturbances.

Definition 7.3. Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space with filtration $\mathcal{F}_t$. Let $Y = \{Y_t : t \geq 0\}$ be an $\mathcal{F}_t$-adapted process that satisfies the evolution equation

$$Y_t = \int_0^t \mathfrak{h}(X_s)\, ds + W_t,$$

where $\mathfrak{h} : \mathbb{R} \to \mathbb{R}$ is a bounded and measurable "sensor" function and W is a Brownian motion independent of V. We call Y the **observation process**.

Why Does This Model Make Sense?

In an ideal system, we might observe $\mathfrak{h}(X_t)$ directly. However, in practice, sensors are imperfect and introduce noise into the measurement:

$$\mathfrak{h}(X_t) + \text{“measurement noise”}.$$

A natural mathematical model for noise is white noise, which is formally written as the derivative of Brownian motion:

$$\frac{dW_t}{dt}.$$

Since Brownian motion is nowhere differentiable, this expression is only formal. Instead, we define the observation process as an integral equation, which avoids differentiation issues.

Integrating the noisy observation equation over time:

$$Y_t - Y_0 = \int_0^t \mathfrak{h}(X_s) ds + W_t.$$

For simplicity, (and from now on) we assume $Y_0 = 0$, meaning that there are no observations at the initial time. Thus, the final equation simplifies to:

$$Y_t = \int_0^t \mathfrak{h}(X_s) ds + W_t.$$

7.3 The filtering problem

Definition 7.4. The filtering problem consists of computing the *conditional distribution* of the signal X_t, given the observation process Y_t up to time t. The goal is to determine the **posterior distribution**:

$$\Pi_t(\varphi) = \mathbb{E}[\varphi(X_t) \mid \mathcal{Y}_t].$$

where $\mathcal{Y}_t = \sigma(Y_s, s \in [0,t])$ represents the **observation filtration**.

Note 7.5. Imagine you are tracking the position of a car that is moving randomly but with some drift (like wind affecting its motion). You cannot see the car directly, but you receive noisy sensor readings Y_t. The **filtering problem** asks:

Given all observations up to time t, what is the best estimate of the car's true position X_t

Mathematically, we express this as the *posterior distribution*:

$$\Pi_t(\varphi) = \mathbb{E}[\varphi(X_t) \mid \mathcal{Y}_t].$$

This means we are taking the expectation of a function $\varphi(X_t)$, given the information we have so far (i.e., all past sensor readings).

Remark 7.6. The probability measure Π_t depends on the entire past observation path $Y_t(\omega)$:

$$\Pi_t = \Gamma_t(t \to Y_t(\omega))$$

which means that the randomness of Y_t induces randomness in Π_t.

Definition 7.7. Two related problems in filtering are:

- **Smoothing**: Estimating X_t given future observations $\mathcal{Y}_{t'}$ where $t' > t$.
- **Prediction**: Estimating X_t given past observations $\mathcal{Y}_{t'}$ where $t' < t$.

7.3.1 Prior and Posterior Distributions

Definition 7.8. The distribution (or the law) of X_t is called the **prior distribution**, denoted by p_t, is given by:

$$p_t(\varphi) = \mathbb{E}[\varphi(X_t)]$$

where $\phi : \mathbb{R} \to \mathbb{R}$.

Example 7.9

Examples:

- If $\varphi(x) = x$, then $p_t(\varphi) = \mathbb{E}[X_t]$ (mean of X_t).
- If $\varphi(x) = x^2$, then $p_t(\varphi) = \mathbb{E}[X_t^2]$ (second moment of X_t).
- If $\varphi(x) = \mathbf{1}_A(x)$, then $p_t(\varphi) = \mathbb{P}(X_t \in A)$ (probability of X_t in set A).

The posterior Π_t is absolutely continuous with respect to the prior p_t, meaning there exists a density $d_t(x)$ such that:

$$\Pi_t(A) = \int_A d_t(x) p_t(dx).$$

Definition 7.10. We say $A : \mathcal{C}_b^2 \to \mathcal{C}_b^2$ is the **infinitesimal generator** of X, where X is Markov process if it has the property that

$$p_0(A\varphi) = \lim_{t\to 0} \frac{p_t(\varphi) - p_0(\varphi)}{t}.$$

Remark 7.11. $\mathcal{C}_b^2$ is the space of all continuous twice differentiable functions, with bounded first and second derivative.

Note 7.12. A describes the rate of change of expected function values over infinitesimally small time intervals

Proposition 7.13

For any $\varphi \in C_b^2(\mathbb{R})$, the prior satisfies:

$$p_t(\varphi) = p_0(\varphi) + \int_0^t p_s(A\varphi) ds,$$

where the *infinitesimal generator* A is defined as:

$$A\varphi = f\varphi' + \frac{1}{2}\sigma^2\varphi''.$$

Remark 7.14. If the initial condition is a Dirac delta measure $p_0 = \delta_x$, then the generator satisfies:

$$A\varphi(x) = \lim_{t\to 0} \frac{p_t(\varphi) - p_0(\varphi)}{t}.$$

Remark 7.15. The domain $\mathcal{D}(A)$ of the infinitesimal generator consists of all functions φ for which:

$$A\varphi = f\varphi' + \frac{1}{2}\sigma^2\varphi''$$

is well-defined and remains continuous and bounded.

$$C_b^2(\mathbb{R}) \subset \mathcal{D}(A).$$

Remark 7.16. The generator satisfies the Kolmogorov forward equation:

$$\frac{d}{dt} p_t(\varphi) = p_t(A\varphi).$$

Proof. By Itô's formula, we expand $\varphi(X_t)$:

$$\varphi(X_t) = \varphi(X_0) + \int_0^t \varphi'(X_s) f(X_s) ds + \int_0^t \varphi'(X_s)\sigma(X_s) dV_s + \frac{1}{2}\int_0^t \varphi''(X_s)\sigma^2(X_s) ds.$$

Since the stochastic integral is a martingale with zero expectation:

$$\mathbb{E}\left[\int_0^t \varphi'(X_s)\sigma(X_s) dV_s\right] = 0,$$

we take expectations to obtain:

$$\mathbb{E}[\varphi(X_t)] = \mathbb{E}[\varphi(X_0)] + \mathbb{E}\left[\int_0^t A\varphi(X_s) ds\right].$$

Since $A\varphi(X_s)$ is a function of X_s, we take expectations inside the integral:

$$\mathbb{E}\left[\int_0^t A\varphi(X_s) ds\right] = \int_0^t \mathbb{E}[A\varphi(X_s)] ds = \int_0^t p_s(A\varphi) ds.$$

Thus, we conclude:

$$p_t(\varphi) = p_0(\varphi) + \int_0^t p_s(A\varphi) ds.$$

□

Exam Questions 7.17

Prove the integral representation:

$$p_t(\varphi) = p_0(\varphi) + \int_0^t p_s(A\varphi) ds.$$

to do

7.4 The strong version of the prior distribution equation

We consider the prior distribution of the process X_t, which describes the evolution of the probability law of X_t before any observations are made.

Exam Questions 7.18 (Exercise)

Assuming that $X_0 = x$, prove that the prior distribution satisfies:

$$\lim_{t\to 0} \frac{p_t(\varphi) - p_0(\varphi)}{t} = (A\varphi)(x).$$

To do

Now, assuming that the probability measure p_t has a density with respect to the Lebesgue measure, we denote this density as:

$$x \to \widetilde{p}_t(x).$$

Thus, the expectation of $\varphi(X_t)$ can be rewritten as:

$$p_t(\varphi) = \int \varphi(x)\widetilde{p}_t(x)dx.$$

Applying the evolution equation,

$$p_t(\varphi) = p_0(\varphi) + \int_0^t p_s(A\varphi)ds.$$

Substituting the density representation,

$$p_t(\varphi) = \int_{\mathbb{R}} \varphi(x)\widetilde{p}_0(x)dx + \int_0^t \int_{\mathbb{R}} A\varphi(x)\widetilde{p}_s(x)dxds.$$

Functional Interpretation in $L^2(\mathbb{R})$

We now introduce the function space $L^2(\mathbb{R})$, which consists of all square-integrable functions:

$$L^2(\mathbb{R}) = \left\{ \varphi : \mathbb{R} \to \mathbb{R} : \int \varphi^2(x)dx < \infty \right\}.$$

This space is equipped with the inner product:

$$\langle \varphi, \psi \rangle = \int_{\mathbb{R}} \varphi(x)\psi(x)dx.$$

The integral term in our equation can now be rewritten using this inner product:

$$\int_{\mathbb{R}} A\varphi(x)\widetilde{p}_s(x)dx = \langle A\varphi, \widetilde{p}_s \rangle.$$

Adjoint Operator A^* and its Role

To further analyse the prior distribution, we introduce the **adjoint operator** A^*, defined as:

$$A^* : C_b^2(\mathbb{R}) \to C_b(\mathbb{R}),$$

and given by:

$$A^*\varphi = -(f\varphi)' + \frac{1}{2}(\sigma^2\varphi)''.$$

Expanding this,

$$A^*\varphi = -f\varphi' - f'\varphi + \frac{1}{2}(2\sigma\sigma'\varphi + \sigma^2\varphi')'.$$

Rearranging,

$$A^*\varphi = (-f' + (\sigma')^2 + \sigma\sigma'')\varphi + (-f + 2\sigma\sigma')\varphi' + \frac{1}{2}\sigma^2\varphi''.$$

Note 7.19. A^* describes the backward evolution of probability densities.

Applying this operator, we obtain the strong form of the prior density evolution equation:

$$\int A\varphi \widetilde{p}_s dx = \int \varphi A^* \widetilde{p}_s dx.$$

This implies that the density $\widetilde{p}_t(x)$ satisfies the PDE:

$$\frac{d}{dt}\widetilde{p}_t(x) = A^* \widetilde{p}_t(x).$$

Exam Questions 7.20 (Exercise)

Using integration by parts, rewrite the previous equation as:

$$\langle A\varphi, \widetilde{p}_s \rangle = \langle \varphi, A^* \widetilde{p}_s \rangle.$$

To do

Proposition 7.21

In other words, A^* is the adjoint operator of A.

In other words, A^* is the adjoint operator of A. We can deduce that

$$\begin{aligned}\int \varphi(x)\widetilde{p}_t(x)dx &= \int \varphi(x)\widetilde{p}_0(x)dx + \int_0^t \int \varphi(x) A^* \widetilde{p}_s(x) dx ds \\ &= \int \varphi(x) \left[\widetilde{p}_0(x) + \int_0^t A^* \widetilde{p}_s(x) ds\right] dx.\end{aligned}$$

Denote by

$$g_t(x) = \widetilde{p}_t(x) - \widetilde{p}_0(x) - \int_0^t A^* \widetilde{p}_s(x) ds$$

so that

$$\int \varphi(x) g_t(x) dx = 0.$$

This means that g_t is orthogonal to any $\varphi \in C_c^2(\mathbb{R})$ and therefore in $L^2(\mathbb{R})$ (since $C_c^2(\mathbb{R})$ is dense in $L^2(\mathbb{R})$).

Exam Questions 7.22 (Exercise)

Prove that $C_c^2(\mathbb{R})$ is dense in $L^2(\mathbb{R})$ in the $\|\cdot\|_2$ norm.

Solution. If $g_t \in L^2(\mathbb{R})$ and is orthogonal to every function in $C_c^2(\mathbb{R})$, then $g_t = 0$. If g_t is not in $L^2(\mathbb{R})$, a more technical argument is required, but it follows that $g_t = 0$ almost everywhere.

Exam Questions 7.23 (Exercise)
Do the following.

1. Prove that $\text{var}(X_t) = \min_{a\in\mathbb{R}} \mathbb{E}[(X_t - a)^2]$.
2. Let $\hat{a} \in \mathbb{R}$ be such that $\text{var}(X_t) = \mathbb{E}[(X_t - \hat{a})^2]$. Show that $\hat{a} = \mathbb{E}[X_t]$.

To do

Exam Questions 7.24 (Exercise)
Assume that $X = \{X_t, t \geq 0\}$ follows the Ornstein-Uhlenbeck process:

$$X_t = X_0 - \int_0^t aX_s ds + \sigma W_t.$$

1. Show that if X satisfies the Ornstein-Uhlenbeck process, then

$$X_t = e^{-at}X_0 + \sigma e^{-at} \int_0^t e^{as} dV_s.$$

2. Show that the posterior density satisfies

$$p_t = N(x_t, \alpha_t),$$

where

$$x_t = x_0 e^{-at}, \quad \alpha_t = \alpha_0 e^{-2at} + \frac{\sigma^2}{2a}(1 - e^{-2at}).$$

To do

Exam Questions 7.25 (Exercise)
Let $f : [0,t] \to \mathbb{R}$ be a square-integrable function such that $\int_0^t f^2(s)ds < \infty$. Define

$$\xi = \int_0^t f(s)dW_s.$$

1. Use Novikov's condition to show that
$$s \to Z_s = \exp\left(\lambda \int_0^s f(r)dW_r - \frac{\lambda^2}{2}\int_0^s f(r)^2 dr\right)$$
is a martingale.
2. Deduce that
$$\mathbb{E}[\exp(\lambda\xi)] = e^{\frac{\lambda^2}{2}\int_0^t f^2(s)ds}.$$
3. Conclude that
$$\xi \sim \mathcal{N}\left(0, \int_0^t f^2(s)ds\right).$$

To do

If X satisfies the O-U process then the prior distribuiton satisfies

$$\frac{d}{dt}p_t(\varphi) = p_t(A\varphi)$$

where

$$A\varphi = -\alpha x\varphi' + \frac{1}{2}\sigma^2\varphi''$$

and

$$\begin{aligned} A^*\varphi &= \frac{\mathrm{d}}{\mathrm{d}x}\left(-ax\varphi(x)\right)_{\frac{1}{2}}(\sigma^2\varphi)'' \\ &= a\varphi_a x\varphi' + \frac{1}{2}\sigma^2\varphi'' \end{aligned}$$

The **strong version** of the above is

$$\frac{\mathrm{d}}{\mathrm{d}t}\widetilde{p}_t(x) = a\widetilde{p}_t(x) + ax\widetilde{p}_t(x)' + \frac{1}{2}\sigma^2\widetilde{p}_t''(x)$$

The solution to this is

$$\widetilde{p}_t = \frac{1}{\sqrt{2\pi\alpha_t}}\exp\left(-\frac{(x-x_t)^2}{2\alpha_t}\right)$$

with initial condition

$$\widetilde{p}_0 = \frac{1}{\sqrt{2\pi\alpha_0}}\exp\left(-\frac{(x-x_0)^2}{2\alpha_0}\right)$$

Exam Questions 7.26 (Exercise)
What can you say about the solution of the PDE

$$\frac{dp_t(x)}{dt} = -(axp_t(x))' + \frac{1}{2}(\sigma^2 x^2 p_t(x))''?$$

To do

7.5 The importance of the prior distribution

Choosing the right model for the signal (that is, the prior distribution) is crucial for successfully solving the filtering problem. The prior distribution must satisfy specific criteria to ensure accurate estimation and avoid misleading results. The two key properties are:

- The prior must be sufficiently general (uninformative) to allow all possible trajectories of the signal and to handle outliers (cases where measurement noise is large). Mathematically, the evolution of the observation process can be approximated as:

  $$Y_{t+1} - Y_t \approx \mathfrak{h}(X_t)\delta t + \Delta W_t$$

  where ΔW_t represents the noise component, which is typically small but can become large with small probability.

- A more informative prior results in a better posterior approximation.

Challenges in choosing a good Prior

Selecting an appropriate prior is a challenging problem in mathematics, statistics, and engineering. Even if a good model is found, one still needs to estimate its parameters. In engineering, the term *training period* refers to the numerical simulations used to estimate these parameters.

If the observations are sufficiently accurate (small measurement noise), then the filtering problem has a stable solution. Over time, the system gradually **forgets** its initial solution. This implies that even if one starts with a bad prior, in the long run, it does not significantly impact the final outcome. However, the convergence rate depends on the quality of the prior.

7.6 From the Prior to the Posterior Distribution

If the signal process X follows an Ornstein-Uhlenbeck process, its prior distribution is given by:

$$p_t = N(x_t, \alpha_t)$$

where

$$x_t = x_0 e^{-at}, \quad \text{and} \quad \alpha_t = \frac{\sigma^2}{2a}(1 - e^{-2at}).$$

Now, suppose that we observe X through the observation process:

$$dY_t = hX_t dt + dW_t.$$

Then, given the observations, the posterior distribution of X_t is also normally distributed:

$$\Pi_t = N(\widetilde{x}_t, \widetilde{\alpha}_t),$$

where the mean and variance evolve as:

$$\begin{aligned}
\widetilde{x}_t &= 2\widetilde{\alpha}_t \left[\frac{x_0 \overline{a}}{\sigma^2 \sinh(t\overline{a})} + h \int_0^t \frac{\sinh(s\overline{a})}{\sinh(t\overline{a})} dY_s \right], \\
\widetilde{\alpha}_t &= \frac{\sigma^2}{a + \overline{a}\coth(t\overline{a})}, \\
\overline{a} &= \sqrt{a^2 + h^2\sigma^2}.
\end{aligned}$$

Exam Questions 7.27

Answer the following.

1. What happens to Π_t when $h \to 0$ and $\mathfrak{h} \to \infty$?
2. Discuss the asymptotic behaviour of Π_t as $t \to \infty$.
3. What happens when $\sigma \to 0$?

To do

The posterior measure Π_t is governed by a stochastic partial differential equation (PDE). Specifically, it depends on the observation process through the integral:

$$\int_0^t \sinh(s\hat{a}) dY_s.$$

Since Y_s is a random process, Π_t itself is a random measure. This means that Π_t is measurable with respect to the observed data Y_t, making it the conditional probability of X_t given Y_t.

Kushner-Stratonovich Equation

The posterior Π_t satisfies a nonlinear stochastic PDE, known as the Kushner-Stratonovich equation:

$$d\Pi_t(\varphi) = \Pi_t(A\varphi)dt + (\Pi_t(h\varphi) - \Pi_t(h)\Pi_t(\varphi))\, dY_t.$$

Note 7.28. Interpretation.

- The first term $\Pi_t(A\varphi)dt$ represents the evolution due to the system dynamics.
- The second term acts as a correction factor based on incoming observations. The term $\Pi_t(h\varphi) - \Pi_t(h)\Pi_t(\varphi)$ ensures that the update accounts for the difference between the actual observation and the expected observation.

Duncan-Mortensen-Zakai Equation

An alternative way to describe the filtering process is via an unnormalised measure ξ_t, which satisfies the linear stochastic PDE:

$$d\xi_t(\varphi) = \xi_t(A\varphi)dt + \xi_t(h\varphi)dY_t.$$

In teh following we will assume that $Y_0 = 0$ (there are no observation available at time $t = 0$). That measn that

$$p_0 = \Pi_0.$$

The observation at tiem $t = 0$ with filtration $\mathcal{F}_0 = (\varnothing, \Omega)$

$$\Pi_0(\varphi) = \mathbb{E}\left[\varphi(X_0) \mid \mathcal{Y}_0\right] = \mathbb{E}\left[\varphi(X_0)\right] = p_0(\varphi).$$

It follows that the D-M-Z equation in its integral form is

$$\xi_t(\varphi) = \Pi_0(\varphi) + \int_0^t \xi_s(A\varphi)ds + \int_0^t \xi_s(h\varphi)dY_s.$$

7.7 The change of probability method

In filtering theory, a common approach to studying the filtering equations is by using a *change of probability measure*. The goal is to transform the original probability measure $\mathbb{P}$ into a new measure $\widetilde{\mathbb{P}}$ under which the observation process Y_t behaves like a Brownian motion. This allows us to simplify the filtering equations.

To perform a change of measure, we introduce the likelihood process Z_t, which is given by:

$$Z_t = \exp\left(-\int_0^t \mathfrak{h}(X_s)dW_s - \frac{1}{2}\int_0^t \mathfrak{h}(X_s)^2 ds\right).$$

To ensure that Z_t is a martingale (which is necessary for it to define a valid change of measure), we check Novikov's condition:

$$\mathbb{E}\left[\exp\left(\frac{1}{2}\int_0^t \mathfrak{h}(X_s)^2 ds\right)\right] < \infty.$$

Using an upper bound,

$$\mathbb{E}\left[\exp\left(\frac{1}{2}\int_0^t \mathfrak{h}(X_s)^2 ds\right)\right] \leq \exp\left(\frac{t\|\mathfrak{h}\|_\infty^2}{2}\right) < \infty,$$

we conclude that Z_t is a martingale.

Using Z_t, we define a new probability measure $\widetilde{\mathbb{P}}$ via:

$$\left.\frac{d\widetilde{\mathbb{P}}}{d\mathbb{P}}\right|_{\mathcal{F}_t} = Z_t.$$

Remark 7.29. We can do this change of measure via the Radon-Nikodym theorem.

Under this new measure, for any event $A \in \mathcal{F}_t$, we have:

$$\widetilde{\mathbb{P}}(A) = \int_A Z_t(\omega)\mathbb{P}(d\omega) = \mathbb{E}[\mathbf{1}_A Z_t].$$

This change of measure effectively removes the influence of the observation process from the dynamics, simplifying computations.

Exam Questions 7.30 (Exercise)

Prove that Z_t satisfies the stochastic differential equation:

$$Z_t = 1 - \int_0^t Z_s \mathfrak{h}(X_s) dW_s.$$

The stochastic integral above is well-defined because

$$\mathbb{E}\left[\int_0^t (Z_s \mathfrak{h}(X_s))^2 \, ds\right] < \infty.$$

To do

Exam Questions 7.31 (Exercise)

Do the following.

1. Prove that $Z_t^2 \leq e^{\|h\|^2 t} \widetilde{Z}_t$ where $\widetilde{Z}_t = \exp\left(-\int_0^t 2\mathfrak{h}(X_s)dW_s - \frac{1}{2}\int_0^t (2\mathfrak{h}(X_s))^2 ds\right)$.
2. Show that $\widetilde{Z}_t$ is a martingale.
3. Use (1) and (2) to deduce the expectation condition of the previous exercise.

To do

We have previously defined

$$Z_t = \exp\left(-\int_0^t \mathfrak{h}(X_s)dW_s - \frac{1}{2}\int_0^t \mathfrak{h}(X_s)^2 ds\right).$$

Since Z_t is a martingale under $\mathbb{P}$, we can define a new probability measure $\widetilde{\mathbb{P}}$ via:

$$\left.\frac{d\widetilde{\mathbb{P}}}{d\mathbb{P}}\right|_{\mathcal{F}_t} = Z_t.$$

To reverse this transformation, we define $\widetilde{Z} = \left\{\widetilde{Z}_t : t \geq 0\right\}$ where

$$\begin{aligned}\widetilde{Z}_t &= \frac{1}{Z_t} \\ &= \exp\left(\int_0^t \mathfrak{h}(X_s)dW_s + \frac{1}{2}\int_0^t \mathfrak{h}(X_s)^2 ds\right) \\ &= \exp\left(\int_0^t \mathfrak{h}(X_s)dY_s - \frac{1}{2}\int_0^t \mathfrak{h}(X_s)^2 ds\right).\end{aligned}$$

The process $\widetilde{Z}_t$ is a martingale under $\widetilde{\mathbb{P}}$

$$\frac{d\widetilde{\mathbb{P}}}{d\mathbb{P}} = Z_t$$

which implies that

$$\frac{d\mathbb{P}}{d\widetilde{\mathbb{P}}} = \frac{1}{Z_t} = \widetilde{Z}_t.$$

Since $\widetilde{Z}_t$ is a martingale under $\widetilde{\mathbb{P}}$, we can relate expectations under the two probability measures:

$$\mathbb{E}[\xi] = \widetilde{\mathbb{E}}[\xi Z_t], \quad \forall \mathcal{F}_t\text{-measurable } \xi.$$

Rewriting using $\widetilde{Z}_t$:

$$\widetilde{\mathbb{E}}[\xi \widetilde{Z}_t] = \mathbb{E}[\xi \widetilde{Z}_t Z_t] = \mathbb{E}[\xi].$$

Lemma 7.32
The observation process $Y = \{Y_t : t \geq 0\}$ is Brownian motion under $\widetilde{\mathbb{P}}$.

Proof. This follows immediately from Girsanov's theorem, which states that under $\widetilde{\mathbb{P}}$, the new process:

$$Y_t = \widetilde{W}_t := W_t + \underbrace{\int_0^t \mathfrak{h}(X_s)ds}_{\text{drift}}$$

is a standard Brownian motion under $\widetilde{\mathbb{P}}$. □

7.8 Kallianpur-Striebel's formula

A key result in filtering theory is the Kallianpur-Striebel formula, which expresses the posterior distribution Π_t in terms of the unnormalised measure ρ_t. This formula provides a way to compute the conditional probability distribution of the hidden state X_t given observations Y_t.

Definition 7.33. Define the measure-valued process:

$$\rho_t(A) = \widetilde{\mathbb{E}}[\mathbf{1}_A(X_t)\widetilde{Z}_t \mid \mathcal{Y}_t], \quad \forall A \in \mathcal{B}(\mathbb{R}).$$

Exam Questions 7.34 (Exercise)
Prove that ρ_t is a measure (but not a probability measure):

1. $\rho_t(A) \geq 0$ for all $A \in \mathcal{B}(\mathbb{R})$.
2. $\rho_t\left(\bigcup_{i=1}^{\infty} A_i\right) = \sum_{i=1}^{\infty} \rho_t(A_i)$ if $A_i \cap A_j = \varnothing$.
3. $\rho_t(\mathbb{R}) \neq 1$, $\mathbb{P}$-almost surely, ($\widetilde{\mathbb{P}}$-almost surely).

TO do

Lemma 7.35
For any Borel measurable function $\varphi : \mathbb{R} \to \mathbb{R}$:

$$\rho_t(\varphi) = \int_{\mathbb{R}} \varphi(x)\rho_t(dx) = \widetilde{\mathbb{E}}[\varphi(X_t)\widetilde{Z}_t \mid \mathcal{Y}_t].$$

Theorem 7.36 (Kallianpur-Striebel Formula)
For any Borel set $A \in \mathcal{B}(\mathbb{R})$,

$$\Pi_t(A) = \frac{\rho_t(A)}{\rho_t(\mathbb{R})}, \quad \mathbb{P}\text{-almost surely, } (\widetilde{\mathbb{P}}\text{-almost surely}),$$

which is equivalent to the expectation form:

$$\Pi_t(\varphi) = \frac{\rho_t(\varphi)}{\rho_t(\mathbb{R})}.$$

Note 7.37. The Kallianpur-Striebel formula allows us to compute the posterior expectation $\Pi_t(\varphi)$ by normalising the unnormalised measure ρ_t.

Proof. We aim to prove the following relation:

$$\Pi_t(\varphi)\rho_t(1) = \rho_t(\varphi).$$

By definition of the measure ρ_t, we have:

$$\rho_t(1) = \int_{\mathbb{R}} 1 \cdot \rho_t(dx) = \rho_t(\mathbb{R}).$$

The conditional expectation representation of $\Pi_t(\varphi)$ is:

$$\Pi_t(\varphi) = \mathbb{E}[\varphi(X_t) \mid \mathcal{Y}_t].$$

Using the definition of conditional expectation:

$$\mathbb{E}[\mathbf{1}_A \Pi_t(\varphi)] = \mathbb{E}[\mathbf{1}_A \varphi(X_t)].$$

for any $A \in \mathcal{Y}_t$. Using the change of measure and the likelihood ratio $\widetilde{Z}_t$, we rewrite the expectation:

$$\widetilde{\mathbb{E}}[\mathbf{1}_A \Pi_t(\varphi)\widetilde{Z}_t] = \widetilde{\mathbb{E}}[\mathbf{1}_A \varphi(X_t)\widetilde{Z}_t].$$

Since conditional expectation preserves linearity:

$$\widetilde{\mathbb{E}}[\mathbb{E}[\mathbf{1}_A \Pi_t(\varphi)\widetilde{Z}_t \mid \mathcal{Y}_t]] = \widetilde{\mathbb{E}}[\mathbb{E}[\mathbf{1}_A \varphi(X_t)\widetilde{Z}_t \mid \mathcal{Y}_t]].$$

Applying the definition of $\rho_t(\varphi)$:

$$\widetilde{\mathbb{E}}[\mathbf{1}_A \Pi_t(\varphi)\widetilde{Z}_t] = \widetilde{\mathbb{E}}[\mathbf{1}_A \varphi(X_t)\widetilde{Z}_t].$$

By recognising that:

$$\rho_t(1) = \widetilde{\mathbb{E}}[1(X_t)\widetilde{Z}_t \mid \mathcal{Y}_t] = \widetilde{\mathbb{E}}[\widetilde{Z}_t \mid \mathcal{Y}_t],$$

we obtain:

$$\widetilde{\mathbb{E}}[\mathbf{1}_A \Pi_t(\varphi)\rho_t(1)] = \widetilde{\mathbb{E}}[\mathbf{1}_A \rho_t(\varphi)].$$

Since $\Pi_t(\varphi)\rho_t(1) - \rho_t(\varphi)$ is $\mathcal{Y}_t$-measurable, it follows that:

$$\Pi_t(\varphi)\rho_t(1) - \rho_t(\varphi) = 0.$$

Rearranging the equation:

$$\Pi_t(\varphi) = \frac{\rho_t(\varphi)}{\rho_t(1)},$$

as required. □

Exam Questions 7.38 (Exercise)

Let ξ be a $\mathcal{Y}_t$-measurable random variable such that:

$$\widetilde{\mathbb{E}}[\mathbf{1}_A \xi] = 0, \quad \forall A \in \mathcal{Y}_t.$$

Prove that:

1. $\widetilde{\mathbb{P}}(\{\omega \in \Omega \mid \xi(\omega) > 0\}) = 0$.
2. $\widetilde{\mathbb{P}}(\{\omega \in \Omega \mid \xi(\omega) < 0\}) = 0$.
3. Deduce that $\xi = 0$, $\widetilde{\mathbb{P}}$-almost surely.

To do

In its explicit integral form, the formula becomes:

$$\Pi_t(\varphi) = \frac{\rho_t(\varphi)}{\rho_t(1)} = \frac{\widetilde{\mathbb{E}}\left[\varphi(X_t)\exp\left(\int_0^t \mathfrak{h}(X_s)dY_s - \frac{1}{2}\int_0^t \mathfrak{h}(X_s)^2 ds\right) \mid \mathcal{Y}_t\right]}{\widetilde{\mathbb{E}}\left[\exp\left(\int_0^t \mathfrak{h}(X_s)dY_s - \frac{1}{2}\int_0^t \mathfrak{h}(X_s)^2 ds\right) \mid \mathcal{Y}_t\right]}.$$

Exam Questions 7.39 (Exercise)
Do the following.

1. Prove that V and the observation process Y are mutually independent under $\widetilde{\mathbb{P}}$.
2. Prove that the signal process X and the observation process Y are mutually independent under $\widetilde{\mathbb{P}}$.

to do

Remark 7.40. At first, it seems counter-intuitive that X and Y are independent under $\widetilde{\mathbb{P}}$. The key reason is that the change of measure absorbs all the information about X into the density function $\widetilde{Z}_t$.

7.9 Numerical approximation of Π_t

Computing the posterior expectation $\Pi_t(\varphi)$ exactly is often infeasible. Instead, we approximate it using discretisation methods.

Proposition 7.41
A useful property of conditional expectations states that if ξ and η are mutually independent random variables, then:

$$\mathbb{E}[\varphi(\xi, \eta)|\sigma(\eta)](\omega) = \mathbb{E}[\varphi(\xi, \eta(\omega))], \quad \mathbb{P} - \text{almost surely.}$$

In other words, to compute $\mathbb{E}[\varphi(\xi, \eta) \mid \sigma(\eta)]$, we fix $\eta = \eta(\omega)$ and compute $\mathbb{E}[\varphi(\xi, \eta(\omega))]$. That is equivalent to integrating the function

$$x \mapsto \varphi(x, \eta(\omega))$$

with respect to the law of ξ:

$$\mathbb{E}[\varphi(\xi, \eta) \mid \sigma(\eta)](\omega) = \int_{\mathbb{R}} \varphi(x, \eta(\omega))\mathbb{P}_\xi(dx).$$

Application to Filtering: computing $\rho_t(\varphi)$

Since X and Y are independent, we apply this property to compute:

$$\rho_t(\varphi) = \widetilde{\mathbb{E}}[\varphi(X_t)\widetilde{Z}_t \mid \mathcal{Y}_t].$$

However, the challenge arises when dealing with integrals of the form:

$$\int_0^t q(X_s)dY_s(\omega).$$

Here, X_s remains random, while Y_s is fixed to a given observation path. Since the integration path should depend on X_s, we approximate it using an *Itô-sum*.

Definition 7.42. We approximate the stochastic integral as:

$$\int_0^t h(X_s)dY_s \approx \sum_{i=0}^{n-1} \mathfrak{h}(X_{t_i})(Y_{t_{i+1}} - Y_{t_i}),$$

corresponding to the partition π

$$0 = t_0 < t_1 < \cdots < t_n = t.$$

We call the LHS an **Itô-sum**.

> **Remark 7.43.** Why is this valid? This sum converges to the integral as the partition becomes finer.

As the partition π becomes finer, we define the following discrete approximation for $\rho_t(\varphi)$:

$$\rho_t^\pi(\varphi) = \widetilde{\mathbb{E}}\left[\varphi(X_t)\exp\left(\sum_{i=0}^{n-1} \mathfrak{h}(X_{t_i})(Y_{t_{i+1}} - Y_{t_i}) - \frac{1}{2}\sum_{i=0}^{n-1} \mathfrak{h}(X_{t_i})^2(t_{i+1} - t_i)\right) \mid \mathcal{Y}_t\right].$$

Since the sum inside the exponent approximates the integral $\int_0^t h(X_s)dY_s$, we expect that:

$$\rho_t(\varphi) \approx \rho_t^\pi(\varphi).$$

One can prove that the approximation error satisfies:

$$\widetilde{\mathbb{E}}\left[(\rho_t(\varphi) - \rho_t^\pi(\varphi))^2\right] \leq Ch^2\|\varphi\|_\infty^2,$$

where

$$h = \max_{i=0,\dots,n-1}(t_{i+1} - t_i)$$

is the largest step size in the partition.

> **Note 7.44.** The error bound shows that the mean square error in our approximation scales quadratically with h. This means that finer partitions (smaller h) lead to more accurate results.

Exam Questions 7.45 (Exercise)

Using the error bound, show that:

$$\mathbb{E}\left[\left|\Pi_t(\varphi) - \frac{\rho_t^\pi(\varphi)}{\rho_t^\pi(1)}\right|\right] \leq Ch\|\varphi\|_\infty.$$

TO do

Approximation of X_t

To make the approximation fully computable, we approximate the signal X_t using the Euler approximation method. We define the discrete-time process X_t^π corresponding to the partition π given by $0 = t_0 < t_1 < \cdots < t_n = t$ as:

$$\begin{aligned} X_{t_0}^\pi &= X_0 \\ X_{t_{i+1}}^\pi &= X_{t_i}^\pi + f(X_{t_i}^\pi)(t_{i+1} - t_i) + \sigma(X_{t_i}^\pi)(V_{t_{i+1}} - V_{t_i}). \end{aligned}$$

Using X_t^π, we define the fully discretised approximation:

$$\rho_t^{\pi,\pi}(\varphi) = \widetilde{\mathbb{E}}\left[\varphi(X_t^\pi)\exp\left(\sum_{i=0}^{n-1}\mathfrak{h}(X_{t_i}^\pi)(Y_{t_{i+1}} - Y_{t_i}) - \frac{1}{2}\sum_{i=0}^{n-1}\mathfrak{h}(X_{t_i}^\pi)^2(t_{i+1} - t_i)\right)\middle|\mathcal{Y}_t\right].$$

One can prove:

$$\widetilde{\mathbb{E}}\left[(\rho_t^\pi(\varphi) - \rho_t^{\pi,\pi}(\varphi))^2\right] \le Ch^2\|\varphi\|_\infty^2.$$

Thus, the Euler approximation error is also quadratic in $\mathfrak{h}$, confirming that finer partitions lead to better approximations.

Exam Questions 7.46 (Exercise)

Prove that

$$\mathbb{E}\left[\left|\Pi_t - \frac{\rho_t^{\pi,\pi}(\varphi)}{\rho_t^{\pi,\pi}(1)}\right|\right] \le Ch\,\|\varphi\|_\infty.$$

to do

Monte Carlo approximation

To estimate expectations numerically, we use Monte Carlo sampling. We generate N independent Euler-discretised trajectories $X_t^{\pi,k}$, where $k \in \{1, \dots, N\}$ indexes the sample:

$$X_{t_{i+1}}^{\pi,k} = X_{t_i}^{\pi,k} + f(X_{t_i}^{\pi,k})(t_{i+1} - t_i) + \sigma(X_{t_i}^{\pi,k})(V_{t_{i+1}}^k - V_{t_i}^k),$$

where $\xi_i^k = (V_{t_{i+1}}^k - V_{t_i}^k) \sim \mathcal{N}(0, t_{i+1} - t_i)$ are independent random variables.

Then we approximate $\rho_t^{\pi,\pi}$ as:

$$\rho_t^{\pi,\pi,N}(\varphi) = \frac{1}{N}\sum_{k=1}^{N}\varphi(X_t^{\pi,k})a^k,$$

where:

$$a^k = \exp\left(\sum_{i=0}^{n-1}h(X_{t_i}^{\pi,k})(Y_{t_{i+1}} - Y_{t_i}) - \frac{1}{2}\sum_{i=0}^{n-1}h(X_{t_i}^{\pi,k})^2(t_{i+1} - t_i)\right).$$

Exam Questions 7.47 (Exercise)
Since Monte Carlo introduces sampling error, we define:

$$\widetilde{\mathbb{E}}\left[(\rho_t^{\pi,\pi}(\varphi) - \rho_t^{\pi,\pi,N}(\varphi))^2\right] = \frac{c_t(\varphi)}{N}.$$

Prove this equality.

To do

Interpretation

In other words, the mean squared error of the approximation decays inversely proportional (as $\frac{1}{N}$) to the number of trajectories (samples) that we take from X, meaning that increasing the number of samples reduces variance. Therefore, more samples lead to better approximations, but computational cost increases.

This leads to the final bound:

$$\mathbb{E}\left[|\Pi_t(\varphi) - \Pi_t^N(\varphi)|\right] \le c_t\left[h + \frac{1}{\sqrt{N}}\right],$$

where

$$\Pi_t^N(\varphi) = \frac{\rho_t^{\pi,\pi,N}}{\rho_t^{\pi,\pi,N}(1)}$$

Challenges of Monte Carlo methods

The constant $c_t(\varphi)$ is the variance of the random variable

$$\varphi(X_t^{\pi,k})\exp\left(\sum_{i=0}^{n-1}\mathfrak{h}(X_{t_i}^{\pi,k})(Y_{t_{i+1}} - Y_{t_i}) - \frac{1}{2}\sum_{i=0}^{n-1}\mathfrak{h}(X_{t_i}^{\pi,k})^2(t_{i+1} - t_i)\right).$$

The variance can explode (exponentially with time). Thus, for large t, we need more samples to maintain accuracy. This makes standard Monte Carlo methods inefficient for long-time filtering.

Summary

The Kallianpur-Striebel formula provides a general framework for approximating Π_t:

1. the approximation of the functional we wish to integrate;
2. the approximation of the signal (the approximation of the prior)
3. the Monte Carlo approximation of the resulting functional.

References

[KS91] I. Karatzas and S. Shreve. *Brownian Motion and Stochastic Calculus*. Graduate Texts in Mathematics (113) (Book 113). Springer New York, 1991.