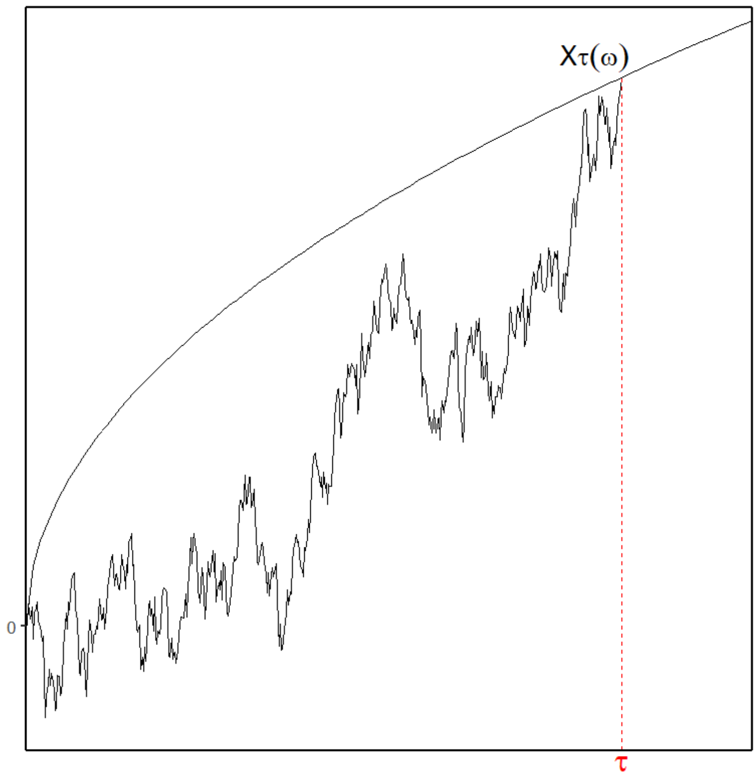
Partial Differential Equations with Applications in Finance

Topias Tolonen-Weckström[1]

April 1, 2025



[1]This material is originally intended for a course of the same name on Spring Term 2024 at the Department of Mathematics, Uppsala University, and it is updated for the same course on Spring Term 2025. Suggestions and comments are continuously accepted at topias.tolonen@math.uu.se.

Contents

Chapter 0

Introduction

Cover Picture

The cover picture shows when a simulated path of an Itô diffusion

$$\mathrm{d}X_t = \frac{1}{\sqrt{2}}\,\mathrm{d}W_t,$$

where W_t is the standard one-dimensional Brownian motion, first hits the curve $\sqrt{t}$, inducing a stopping time

$$\tau := \inf\left\{t \geq 0 : X_t(\omega) \notin \left(-\infty, \sqrt{t}\right)\right\}.$$

Foreword for 2nd Version, 2025

A sure sign of spring is when this course starts. Generally, I'm not too happy with my initial version from 2024. It was created hastily from the scratch and includes both sloppy writing and mathematics – not that there exists a text without some sloppiness in it, but too much is too much. That being said, the material is updated on-the-go from March to May in 2025. Apart from general polishing work, one should expect major rewriting especially in the areas Itô Integral and Stochastic control.

Foreword

This course material is prepared for the course of the same name in the Department of Mathematics, Uppsala University, Sweden, during the Spring term of 2024. The course is aimed for Masters' students (second-cycle) in mathematics, probability theory, stochastic processes, mathematical finance, and related subjects. I'm doing my best to suit the material for the target audience and balance between exhaustive presentation and pedagogical approach.

The material is largely based on the lecture notes by the previous lecturer, Yuqiong Wang, of the course Partial Differential Equations with Applications in Finance in Uppsala University: Wang 2023. In addition, the following sources are used: Kohn 2014, Björk 2020, and Karatzas 1991, as well as the lecture notes Klimek 2019 of Maciej Klimek, also a previous lecturer of the course at Uppsala University. In addition to these references, the material loans some probabilistic machinery from standard references (picked mostly to suit the preferences of the author) in probability theory and related areas: Gut 2013 for probability theory, Øksendal 2003 for stochastic differential equations, Evans 2010 for partial differential equations, Williams 1991 for martingales, Bass 2013 for stochastic processes and martingales, and Peskir 2006 for optimal stopping problems.

As an avid reader realizes, materials like these are often written hastily in between preparing lectures, researching, and doing other tasks that have stealthily found their way into a routines of a researcher in the beginning of their career. That is – suggestions and comments are continuously accepted at topias.tolonen@math.uu.se.

Introduction

As the name of the material suggests, our business here is to combine partial differential equations (PDE) with mathematical finance. In order to do that, we need to combine deterministic PDE's with random processes, which is a non-trivial feat to achieve. How could a deterministic equation be represented by a random process, something that is by definition non-deterministic?

The answer, of course, lies in expectations and probabilities, which are associated with random variables and random processes, but which are often deterministically defined or calculated. Consider, for example, a stochastic process

$$\mathrm{d}X_s = \mu \,\mathrm{d}s + \sigma \,\mathrm{d}W_s$$

for some constants μ, σ and a standard Brownian motion W_t. This process fluctuates randomly, i.e. we have no way of saying where exactly the process is moving at some given point, say, X_{100}. For example, one can examine the cover picture, which deciphers a special case of this process. It is in no way guaranteed that a randomly realized Brownian motion hits the curve $\sqrt{t}$ in any reasonable time, or in a way that makes a picturesque cover picture!

However, consider, instead of random realizations of a process, we examine its expectation

$$u(x,t) = \mathbb{E}_x \left[\Phi(X_T)\right]$$

with some initial value $X_t = x$ for some $t \leq s < T$, and some (reasonable) terminal value function Φ. Then, it can be shown that the expectation u solves a certain deterministic partial differential equation with a terminal condition:

$$\begin{cases} u_t + \mu u_x + \frac{1}{2}\sigma^2 u_{xx} &= 0 \\ \qquad\qquad u(x,T) &= \Phi(x). \end{cases}$$

This is rather fascinating, as the result[1] combines a stochastic process, which moves randomly by the law of a normal distribution, with a deterministic second order partial differential equation! More interestingly, the result can be used as an easy way to solve for a terminal value problem. In fact, it turns out that there exists many similar results which are covered in this material: those with a background in mathematical finance can be interested in option pricing, and on the other hand, students with a background in real-world matters might be interested in the Heat equation. The reader should learn a few key concepts regarding both stochastic differential equations and partial differential equations, their links, and finally, some applications in Finance: the Black-Scholes Formula for call options, Barrier Options, Local Volatility, Merton's allocation problem, and finally, finding an optimal exercise time to an American option.

The material ends up with solving financially applied examples in the field of Optimal Stopping. The field is rich in recent research, and most often a good stopping problem has applied foundations. Quite unintuitively, it sometimes appears that the best problems which are mathematically involved are inspired applications considering some real-world problems. At the end of this material, we finish with a conclusion that applied problems inspire optimal stopping, which induce the so-called free-boundary problems:

$$\text{Applications} \longleftrightarrow \text{Optimal Stopping} \longleftrightarrow \text{Free-Boundary Problems.}$$

At the summit of the materials's taxonomy, the above diagram summarizes the supposed learning outcomes of this material quite well.

Preview

The material is arranged as follows. To make skimming through the material easier, the material is equipped with an index at the very end. All the definitions and results (theorems, proposals, etc.) share the same running number counter per each chapter. The examples follow their own running counter.

In Chapter 1, the reader will be reminded – or taught in a fast manner – the basics of probability theory and stochastic processes, up until Itô integrals and the Itô formula. The chapter is written under the assumptions that the reader knows basics of probability theory, preferably from a measure theoretic approach.

In Chapter 2, we will discuss the connection between stochastic and partial differential equations. The chapter introduces Feynman-Kac theorem as an application of the Itô Formula, dives deeper into the properties of Brownian motions, and uses these both to guide the reader with different Boundary-Value problems.

In Chapter 3, we will discuss the Heat equation, namely solving it with the help of SDE's. The chapter covers many basic examples and applications of Heat equations – and closely related Cauchy initial value problems – and has a section devoted to financial applications, particularly to pricing of barrier options.

[1] Later known as the Feynman-Kac formula.

In Chapter 4, we introduce the notion of Markov processes and related Kolmogorov equations: the Chapman-Kolmogorov equation, Kolmogorov Backward equation, and the Fokker-Planck equation. Moreover, we highlight some important applications to the Fokker-Planck equation, namely regarding local volatility and the so-called Dupire's formula.

In Chapter 5, we motivate and introduce stochastic control problems, where a certain value function is optimized, often maximized, through selecting a suitable control variable. Moreover, we introduce the corresponding Hamilton-Jacobi-Bellman equations and discuss the verification of such problems. We also highlight some key applications, such as the Merton's problem.

In Chapter 6, we discuss a special case of a stochastic control problem called the optimal stopping problem, where the control is a stopping time. We finish with the important application of solving the price of an American option with our newfound machinery.

Each chapter ends with a set of suitable exercises for the reader. Some exercises are required for the user to hands-on practice the tools given – for example, in the case of the Itô formula – and other exercises are meant as more involved problems for advanced students and to pique the interest of the reader. These problems largely follow a collection of exercises by Yuqiong Wang from the 2023 course of Partial Differential Equations with Applications in Finance, Uppsala University.

Finally

This material can also be viewed through the lenses of the democratic aspect of higher education. Doing so, the main aim of the material would be to deepen the readers' enjoyment in probability theory, and to be inspired by the applications of it. I'm doing my best to write by this ideal and the expectations built by it, and I hope that at least some readers find solace in this material for whatever reason, despite how small.

Good luck!

Chapter 1

Crash Course in Probability and Itô Calculus

1.1 Elementary Probability

We begin with some core definitions in probability theory. Traditionally one starts with the notion of a probability space.

Definition 1.1. *(Probability Space and Probability Measure).* A triplet $(\Omega, \mathcal{F}, \mathbb{P})$ is called a probability space, where Ω denotes the sample space of all possible outcomes, $\mathcal{F}$ is a σ-algebra, the collection of events that are subsets of Ω, and $\mathbb{P}$ is the probability measure, that is, $\mathbb{P}$ satisfies the so-called Kolmogorov axioms[1]:

- $\mathbb{P}(E) \geq 0$ for all $E \in \mathcal{F}$,
- $\mathbb{P}(\Omega) = 1$, and
- $\mathbb{P}$ satisfies a so-called σ-additivity, i.e. it's countably additive under disjoint collections of sets A_i:
$$\mathbb{P}\left(\bigcup_i A_i\right) = \sum_i \mathbb{P}(A_i).$$

Recall that $\mathcal{F}$ is a $\sigma-$algebra if $\Omega \in \mathcal{F}$, $\mathcal{A} \in \mathcal{F}$ implies $\mathcal{A}^c \in \mathcal{F}$[2], and $\mathcal{F}$ is closed under countable unions of sets. The notion of σ-algebra is required in order to note for the collections of events, but for the scope of this course, it mostly only stays as a motor in our core definitions. However, let's quickly review some of the basics.

[1]This material doesn't discuss much measure theory and for us it suffices to reduce most of measure theory to this notion of probability measure. The reader can access for example Gut 2013 for more exhaustive notion of measure theory in probability.

[2]Here with the notation A^c we denote the complementary event of A.

Shortly on Measure Theory

Remark 1.2. Recall that we call a pair $(\Omega, \mathcal{F})$ a measurable space. If we then equip the space with some measure μ (which is then a function $\Omega \to \mathcal{F}$ and satisfies similar axioms as above), then the triplet $(\Omega, \mathcal{F}, \mu)$ becomes a measurable space. Thus, if $\mu \equiv \mathbb{P}$, we call the measurable space a probability space.

Moreover, the following definition will follow us throughout the whole material.

Definition 1.3. *(Measurable Function).* Let $(\Omega, \mathcal{F})$ be a measurable space. Then, a function $f : \Omega \to \mathbb{R}$ is called $\mathcal{F}$-measurable if its preimage is in $\mathcal{F}$, i.e. when

$$f^{-1}(A) = \{x \in \Omega : f(x) \in A\} \in \mathcal{F}$$

for all $A \in \mathbb{R}$.

Now that we have defined measurable functions, we present a short theorem regarding their convergence for both later use and for life in general.

Theorem 1.4. *(Dominated Convergence Theorem).* Let $(\Omega, \mathcal{F}, \mu)$ be a measure space, and let $f_1, f_2, \ldots$ be a sequence of functions in the measure space. Assume that

$$\lim_{n\to\infty} f_n(x) = f(x)$$

and that $|f_n(x)| \leq g(x)$ for some g such that $\int_\Omega |g| \, \mathrm{d}\mu < \infty$, for all n, and for all $x \in \Omega$. Then f is integrable and

$$\lim_{n\to\infty} \int_\Omega f_n \, \mathrm{d}\mu = \int_\Omega f \, \mathrm{d}\mu.$$

Proof. Omitted. A clear and well presented proof lies in Salamon 2020 Theorem 1.45. □

Back to Probability Theory

Since Kolmogorov, the concept of a probability space has been central in probability theory. Indeed, note that

- Ω, the sample space, denotes the space of all possible outcomes,
- $\mathcal{F}$ serves as a collection of those events lying in the sample space, and
- $\mathbb{P}$ assigns a probability, a value in $[0, 1]$, to each event in $\mathcal{F}$.

It should be noted that often to determine the likelihood of individual events, for discrete events (when the sample space is *countable*) a suitable choice for $\mathcal{F}$ is the power set 2^Ω of the sample space.

In addition, the following useful properties of the probability measure follow from its axioms[3]:

[3] The reader can prove these for a simple exercise. These properties hold, in fact, only if the set A_i is measurable – for the purpose of this course, let's assume all sets discussed to be measurable.

- $\mathbb{P}(\emptyset) = 0$,
- $\mathbb{P}(A^c) = 1 - \mathbb{P}(A)$, and
- $A_1 \subseteq A_2 \Rightarrow \mathbb{P}(A_1) \leq \mathbb{P}(A_2)$.

The concepts here can appear abstract if the reader isn't properly accustomed to the ways of probability theory. Two simple examples might be of help:

Example 1.1. *(Tossing two coins).* When modelling a coin toss of two coins, one can begin with assuming if the coin is fair, i.e. if $\mathbb{P}$("Individual coin lands on heads") $=$ $\mathbb{P}$("Individual coin lands on tails") $= 0.5$. Abbreviating the event descriptions as H's and T's, we can then define the sample space as $\Omega = \{TT, HT, TH, HH\}$.

However, the set of the events $\mathcal{F}$ is then not fixed. Often one wants to explore the smallest σ-algebra combined with some sets of interest. For example, one can be interested in the probability of having at least one heads. Then, a natural choice would be

$$\mathcal{F} = \{\emptyset, \{TT\}, \{HH, HT, TH\}, \Omega\}\,.$$

Alternatively, it's possible to let the set of events be the *power set* of the sample space, with a total of $2^4 = 16$ elements:

$$\mathcal{F} = \{\{\emptyset\}, \{TT\}, \{HT\}, \{TH\}, \{HH\}, \{TT, HT\}, \ldots, \{TT, HT, TH, HH\}\}$$

Either way, the probability of getting at least one heads in two coin tosses would be 0.75.

Example 1.2. *(Infinitely many coin tosses).* Now each single outcome is an infinite binary string: $\Omega = \{H, T\}^\infty$. If we are interested in a probability of the first coin landing as heads, we can choose

$$\mathcal{F} = \{\{\emptyset\}, \{HH, \ldots\}, \{HT, \ldots\}, \ldots\}\,.$$

Further analysis of this example can be done with e.g. so-called Borel-Cantelli Lemma (for this, check e.g. Øksendal 2003 or Gut 2013).

Sometimes, it's necessary for us to know if two events influence each other, or if our knowledge of the first event happening doesn't affect our judgement regarding the probability of some other event. For these types of discussions, one needs a concept of independence.

Definition 1.5. *(Independence of Events).* Two events A and B are independent if and only if $\mathbb{P}(A \cap B) = \mathbb{P}(A)\mathbb{P}(B)$. For more than two events, we say that events $\{A_k\}_{k=1}^n$ are independent if and only if

$$\mathbb{P}\left(\bigcap_i A_{i_k}\right) = \prod \mathbb{P}\left(A_{i_k}\right),$$

where intersections and products are taken of all possible combinations of $\{1, \ldots, n\}$.

In addition, independence of events raises a reason for us to discuss conditional probability.

Definition 1.6. *(Conditional Probability).* Given two events A and B with $P(B) > 0$, the conditional probability of A given B is then

$$\mathbb{P}(A|B) = \frac{\mathbb{P}(A \cap B)}{\mathbb{P}(B)}.$$

It should be noted that $\mathbb{P}(A \cap B) = \mathbb{P}(B \cap A) = \mathbb{P}(A)\mathbb{P}(B|A)$, and if the events A and B are independent, then $\mathbb{P}(A|B) = P(A)$. In addition, it turns out that the conditional probability itself $\mathbb{P}(A|B)$ satisfies the Kolmogorov axioms, and is thus a stand-alone probability measure, as well!

As one might recall from previous studies in the craft of probability, calculating probabilities with only events becomes cumbersome when one requires to calculate more complex situations. Random variables – which are not random nor variables – are often brought to rescue.

Definition 1.7. *(Random Variable).* Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space. random variable is a $\mathcal{F}$-measurable function $X : \Omega \to \mathbb{R}$ which maps an element from the sample space to some real-valued number so that the preimage of X is in $\mathcal{F}$, i.e.

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

for every *Borel-set*[4] A.

Despite their technical definition – and to reiterate, despite not being random or variable, and rather a deterministic function for each $\omega \in \Omega$ – random variables are inherently useful. Often we use a shorthand notation $\{\omega \in \Omega : X(\omega) \in A\} = \{X \in A\}$ to save us few moments of our scarce time upon this planet.

Remark 1.8. A random variable X induces a probability measure:

$$\mathbb{P}_X(A) = \mathbb{P}(X^{-1}(A)) = \mathbb{P}(\{\omega \in \Omega : X(\omega) \in A\}) =: \mathbb{P}(X \in A),$$

where we used a shorthand notation to hide all the ω's, $A \in \mathbb{R}$, $X^{-1}(A) \in \mathcal{F}$, $\mathbb{P}_X$ refers to the distribution of X.

The previous remark gives us motivation to also define distributions functions for random variables.

Definition 1.9. *(Distribution Function).* Let X be a random variable and $\mathbb{P}_X(A) = \mathbb{P}(X^{-1}(A))$ its distribution for some set A. Then its distribution function is $F_X(x) = \mathbb{P}_X([-\infty, x]) = \mathbb{P}(X \leq x)$.

If we only consider one random variable, we write $F_X =: F$.

[4]Borel-sets are sets that can be formed from open sets with countable unions, intersections, and complements. We can denote, for example, that $\mathcal{A} \in \mathcal{B}([0, \infty))$ denotes a collection collection of Borel-sets on the half-line.

An important example of distributions is the so-called Gaussian or normal distribution.

Definition 1.10. *(Normal Distribution).* We say that a random variable X follows a normal distribution with parameters μ and σ, denoted $X \sim N(\mu, \sigma^2)$, if its distribution function is of the form

$$F_X(x) = \int_{\mathbb{R}} \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2}.$$

If $\mu = 0$ and $\sigma = 1$, the distribution is called a standard normal distribution.

The concept of normal distributions generalizes neatly to a multi-dimensional case. There are a few ways to present the following definition, but we picked this one for its notational simplicity.

Definition 1.11. *(Multivariate Normal Distribution).* We say that a random vector $(X_1, \ldots, X_n)$ has a multivariate normal distribution if, for some constant vector $(a_1, \ldots, a_n)$, every linear combination

$$\sum_{i=1}^{n} a_i X_i$$

has a normal distribution.

It is also meaningful to define when two random variables are equivalent. We review two different ways to define this phenomenon.

Definition 1.12. *(Almost Sure, and in Distribution, Equivalent Random Variables).* Let X and Y be random variables. We say that X and Y are almost surely equivalent, $X =^{a.s.} Y$, if

$$\mathbb{P}(\{\omega \in \Omega : X(\omega) \neq Y(\omega)\}) = 0.$$

Similarly, we say that X and Y are equivalent in distribution, $X =^{d} Y$, if for all sets A

$$\mathbb{P}(X \in A) = \mathbb{P}(Y \in A).$$

Note that almost sure equivalency implies equivalency in distribution. For later use, it is also sensible to define few types of convergence for random variables.

Definition 1.13. *(Almost Sure and in Distribution Convergence).* Let $X_1, X_2, \ldots$ be random variables. We say that X_n converges almost surely (a.s.) to the limit random variable X if and only if

$$\mathbb{P}\left(\{\omega : X_n(\omega) \to X(\omega) \text{ as } n \to \infty\}\right) = 1.$$

We then denote $X_n \longrightarrow^{a.s.} X$.

Moreover, let $F_1, F_2, \ldots$ be the distribution functions of the respective random variables $X_1, X_2, \ldots$. We say that X_n converges in distribution (d) to the random variable X with distribution function F if

$$F_n(x) \longrightarrow F(x),$$

as $n \to \infty$ for every $x \in \mathbb{R}$ at which F is continuous. We then denote $X_n \longrightarrow^{d} X$.

A useful result, which may be of use, is tied to these concepts of convergence.

Proposition 1.14. *(Fatous's Lemma).* Let X and $X_1, X_2, \ldots$ be random variables, and assume that $X_n \to X$ almost surely or in distribution as $n \to \infty$. Then

$$\mathbb{E}[|X|] \leq \liminf_{n\to\infty} \mathbb{E}[|X_n|].$$

Proof. Omitted. See, for example Gut 2013. □

As mentioned in Chapter 0, expectation of random variables plays a crucial role in dancing between the deterministic and stochastic worlds. After all, expectation is a property of a random variable and its distribution, a single statistic which can be calculated deterministically based on random variables distribution regardless of how ω turn out to realize – in fact, it is an operator (on a function X) defined to be the Lebesgue-Stieltjes Integral.

Definition 1.15. *(Expectation of a Random Variable).* Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space and X a random variable. Then, the expectation of X is[5]

$$\mathbb{E}[X] = \int_\Omega X(\omega)\,\mathrm{d}\mathbb{P}(\omega) = \int_\mathbb{R} x\,\mathrm{d}\mathbb{P}_X(x).$$

Moreover, if $f : \mathbb{R} \to \mathbb{R}$ is a Borel-measurable and integrable[6], then

$$\mathbb{E}[f(X)] = \int_\Omega f(X(\omega))\,\mathrm{d}\mathbb{P}(\omega) = \int_\mathbb{R} f(x)\,\mathrm{d}\mathbb{P}_X(x).$$

Expectation is important tool for us, and to utilize it fully, we require the notion of conditional expectation.

Definition 1.16. *(Conditional Expectation).* Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space, X a random variable, and A some Borel-measurable event. Then, the conditional expectation of X given A is

$$\mathbb{E}[X|A] = \frac{\mathbb{E}[X\mathbb{1}_A]}{\mathbb{P}(A)},$$

where $\mathbb{1}_A$ denotes the indicator function with respect to A.

From this definition raises an interesting observation. If $\sigma(A_1, \ldots, A_n) = \mathcal{G} \in \mathcal{F}$ is the smallest σ-algebra containing sets $A_1, \ldots, A_n$, we have

$$\mathbb{E}[X|\mathcal{G}] = \sum_{k=1}^{n} \mathbb{E}[X|A_k]\mathbb{1}_{A_k}.$$

Moreover, if $\mathcal{G} \in \mathcal{F}$, the conditional expectation is a uniquely determined function from Ω to $\mathbb{R}$ and has the following properties:

[5]This definition is indeed well-defined. If we were to be punctual, we would need to start the definition for simple random variables (weighted sums of indicator functions), generalize to non-negative random variables, and use their sums to construct the general expectation for X

[6]For this, it suffices to say that the integral of $|f|$ exists and is not infinite.

- $\mathbb{E}[X|\mathcal{G}]$ is $\mathcal{G}$-measurable,
- $\int_A \mathbb{E}[X|\mathcal{G}]\, \mathrm{d}\mathbb{P} = \int_A x \, \mathrm{d}\mathbb{P}$ for all $A \in \mathcal{G}$,
- $\mathbb{E}\left[\mathbb{E}[X|\mathcal{G}]\right] = \mathbb{E}[X]$ (Law of iterated expectation),
- If X is $\mathcal{G}$-measurable, then $\mathbb{E}[X|\mathcal{G}] = X$, and
- If $X \perp\!\!\!\perp \mathcal{G}$, then $\mathbb{E}[X|\mathcal{G}] = \mathbb{E}[X]$.

Besides expectation, another important operator for random variables is the variance.

Definition 1.17. *(Variance of a Random Variable).* Let X be a random variable. Then, the variance of X is

$$\mathrm{Var}[X] = \mathbb{E}\left[(\mathbb{E}[X] - X)^2\right] = \mathbb{E}[X^2] - \mathbb{E}[X]^2.$$

Variance has many convenient properties. For example for any $a \in \mathbb{R}$, we have $\mathrm{Var}[aX] = a^2\mathrm{Var}[X]$ and $\mathrm{Var}[X] \geq 0$.

Now that we have introduced random variables, distributions, and expectations, we present – for later use – perhaps the single most useful theorem in classical statistics, albeit in a very simple setting.

Theorem 1.18. *(Central Limit Theorem).* Let $X_1, X_2, \dots$ be independent and identically distributed[7] random variables with $\mathbb{E}[X_i] = \mu < \infty$ and $\mathrm{Var}[X_i] = \sigma^2 < \infty$, and set $S_n = \sum_{i=1}^{n}$ for some $n \geq 1$. Then

$$\frac{S_n - n\mu}{\sigma\sqrt{n}} \longrightarrow^d \mathrm{N}(0, 1)$$

as $n \to \infty$.

Proof. Omitted, but surprisingly simple with characteristic functions, which are not covered in this material. See Gut 2013 Theorem 1.1 in chapter 7 for more details. □

We have now hurriedly covered some of the basics of elementary probability. Now imagine: what if we took a collection on x random variables $X_1, \dots, X_n$, and formed them into a single vector?

[7] The notion of independent and identically distributed random variables is often abbreviated as *i.i.d.*

1.2 Stochastic Processes

Stochastic processes form the backbone of this course. We are interested in defining them, how they evolve, and what kind of properties they have. Without further ado, let's define them for further analysis.

Definition 1.19. *(Stochastic Process).* Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space. A stochastic process is a collection of random variables $\{X_t\}$ in the probability space indexed by $t \in \mathcal{T}$[8] so that for each t, we have a random variable

$$\omega \mapsto X_t(\omega), \quad \omega \in \Omega,$$

and for each ω, we have a trajectory[9]

$$t \mapsto X_t(\omega), \quad t \in \mathcal{T}.$$

So, in other words, X can also seen as a map $X_t : \mathcal{T} \times \Omega \to \mathbb{R}$. Due to the dual nature of trajectories as realizations of stochastic processes and random variables as the underlying mechanisms, there is a plethora of different notations for stochastic processes: $X_t, X(t), X_t(\omega), X(t, \omega)$, and so on. In this material, similarly to the individual random variables, often omit the ω from the notation, and decide to use time t in the subscript: X_t.

One important example of stochastic processes is a Brownian motion. Brownian motion was originally discovered by a Scottish botanist Robert Brown (!) as an irregular motion of pollen grains in liquid, although the mathematical formulation came a bit later.

Definition 1.20. *(Brownian Motion).* A stochastic process W_t is called the (standard) Brownian Motion[10] if

(i) $W_0 = 0$ almost surely[11],

(ii) W_t has continuous trajectories,

(iii) W_t has independent increments, i.e. if $t_1 \leq \cdots \leq t_n$, then for all $i, k \in 1, \ldots, n$ such that $t_{i-1} > t_k$,
$$W_{t_i} - W_{t_{i-1}} \perp\!\!\!\perp W_{t_k} - W_{t_{k-1}},$$

(iv) the increments of W_t are Gaussian, and if $s < t$, then

$$W_t - W_s \sim N(0, t - s).$$

[8] In this material, we consider mostly $\mathcal{T} = [0, \infty)$.

[9] Also known as a path.

[10] Often also denoted as the Wiener Process.

[11] The reader is advised to be vigilant, as in the material, we have a bad habit of sometimes describing some stochastic process as a Brownian motion starting at $a \in \mathbb{R}^n$ for some $a \neq 0$. This is a shorthand notation for us to describe a process $\mathrm{d}X_t = \mathrm{d}W_t$ with some initial value $X_0 = a$, i.e. that in integral form the process is described as $X_t = a + W_t$.

Moreover, we denote the n-dimensional Brownian motion as $W_t = (W_t^1, \dots, W_t^n)$, where $W^1, \dots, W^n$ are independent.

Remark 1.21. The existence and well-definition of the Brownian motion, for example, with the respect of continuous trajectories, is not self-evident nor trivial. See, for example, Chapter 2 in Øksendal 2003 for more discussion.

The following example might help with understanding the properties of Brownian motion. The purpose here is to highlight that despite being a stochastic process defined in multiple steps, its properties are easy to calculate.

Proposition 1.22. Let W_t be a Brownian motion. Then

(i) $\mathbb{E}[W_t^2] = t$,

(ii) $\mathrm{Cov}[W_t, W_s] = t \wedge s$, and

(iii) $\mathbb{E}\left[(W_t - W_s)^2\right] = t - s$ for $s < t$.

Proof. The proofs are fairly straight-forward and follow mostly from the property of independent and normal increments.

(i) $\mathbb{E}[W_t^2] = \mathrm{Var}[W_t] = \mathrm{Var}[W_t - W_0] = t$.

(ii) For the covariance, we consider arbitrarily the case $s < t$ and note that $W_t W_s = W_s^2 + W_s(W_t - W_s)$. Now by the first part and independent increments

$$\begin{aligned}\mathrm{Cov}[W_t, W_s] &= \mathbb{E}[W_t W_s] - \mathbb{E}[W_t]\mathbb{E}[W_s] \\ &= \mathbb{E}\left[W_s^2 + W_s(W_t - W_s)\right] \\ &= \mathbb{E}\left[W_s^2\right] = s.\end{aligned}$$

Since $s < t$ was chosen arbitrarily, combining with the case $t > s$ we get $\mathrm{Cov}[W_t, W_s] = t \wedge s$.

(iii) By using the same trick as in the second part, we get

$$\mathbb{E}\left[(W_t - W_s)^2\right] = \mathbb{E}\left[(W_t)^2 - 2W_tW_s + (W_s)^2\right] = t - 2s + s.$$

□

However, one needs to be careful:

Remark 1.23. Brownian motion W_t is nowehere differentiable.

Often stochastic processes are dependent on their historical trajectories. This raises an interesting problem: we never actually know which ω represents the current reality, as we can only observe the process up until this time point. This notion is explored with filtrations.

Definition 1.24. *(Filtration and Filtered Probability Space).* A filtration $\{\mathcal{F}_t\}_{t\geq 0}$ is a family of increasing sub-σ-algebras[12] of $\mathcal{F}$, i.e. $\mathcal{F}_s \subset \mathcal{F}_t$ for $s < t$. Moreover, a probability space $(\Omega, \mathcal{F}, \mathbb{P})$ can be equipped with a filtration: then we end up with a filtered probability space $(\Omega, \mathcal{F}, \{\mathcal{F}_t\}_{t\geq 0}, \mathbb{P})$.

A short example to illustrate this:

Example 1.3. *(Tossing 2 coins, revisited).* Let $\Omega = \{TT, HT, TH, HH\}$ and let X_i be the result of the ith toss, $X_i \in \{0, 1\}$. Then assume that the first toss lands on a head. Now we denote $X_1\{HH\} = X_1\{HT\} = 1$, and $X_1\{TT\} = X_1\{TH\} = 0$, and moreover,

$$\mathcal{F}_1 = \sigma(X_1) = \{\{HH, HT\}, \{TT, TH\}, \emptyset, \Omega\} \subset \mathcal{F}.$$

We can denote the filtration generated by a process X, up to the time t, by $\mathcal{F}_t^X$. This, simply put, includes all information of the process up until time t. Note also that since we require filtration to be increasing, later times always include information of previous time points. Filtration has a lot of useful properties. For example, if by observing X_t, we can determine if some event $A \in \mathcal{F}$ has occurred or not. In this case, we denote $A \in \mathcal{F}_t^X$. These notions are made useful by the following (important!) definitions:

Definition 1.25. *(Adapted and Measurable Processes).* Let Z be a random variable. If we can determine the value of Z by F_t^X, then we say that Z is $\mathcal{F}_t^X$-measurable and $Z \in \mathcal{F}_t^X$.

Moreover, if X and Y are stochastic processes and $Y_t \in \mathcal{F}_t^X$ for all $t \geq 0$, then we say that Y is F^X-adapted, i.e. $Y \in \mathcal{F}^X$.

It's easy to illustrate elementary examples on the topic.

Example 1.4. *(Adapted processes).* It is easy to see that

$$Y_t = \sup_{0\leq s\leq t} X_s \in \mathcal{F}_t^X,$$

whereas

$$Y_t = \sup_{0\leq s\leq t+1} X_s \notin \mathcal{F}_t^X.$$

These concepts flow smoothly to our next hammer in the toolkit of stochastic processes.

Definition 1.26. *(Stopping time).* Let $\{\mathcal{F}_t\}_{t\geq 0}$ be a filtration and $(\Omega, \mathcal{F}, \{\mathcal{F}_t\}_{t\geq 0}, \mathbb{P})$ be a filtered probability space. A function function $\tau : \Omega \to [0, \infty)$ is called a stopping time with respect to $\{\mathcal{F}_t\}_{t\geq 0}$ if

$$\{\omega \in \Omega : \tau(\omega) \leq t\} \in \mathcal{F}_t, \qquad \text{for all } t.$$

Moreover, the following claims are equivalent[13]:

[12] A subset of $\mathcal{F}$ is called a sub-σ-algebra of $\mathcal{F}$ if it also is a $\sigma-$algebra.

[13] In mathematics professional lingo jargon, one often writes TFAE.

- τ is a $\mathcal{F}_t$-stopping time.
- $\mathbb{1}_{[0,\tau]}$ is $\mathcal{F}_t$-adapted.

Remark 1.27. τ is a random variable.

Proposition 1.28. *(Properties of Stopping times).* Let τ_1 and τ_2 be $\mathcal{F}_t$-stopping times. Then $\tau_1 \wedge \tau_2$, $\tau_1 \vee \tau_2$, and $\tau_1 + \tau_2$ are also $\mathcal{F}_t$-stopping times. However, $\tau_1 - \tau_2$ is not.

Proof. Left as exercises (see Section 1.6 and exercise 102). $\square$

A short example might be in place to illustrate the utility brought by stopping times.

Example 1.5. Every deterministic $t \in [0, \infty)$ is a stopping time.

Example 1.6. *(Hitting and exit times).* Let X_t be a stochastic process in $\mathbb{R}^n$, and let $X_0 \in D \subset \mathbb{R}^n$. We define the first exit time of X_t from D as

$$\tau_D := \inf\{t \geq 0 | X_t \notin D\}.$$

Example 1.7. *(A non-example.)* Let X_t be a stochastic process in $\mathbb{R}$ and $C \in \mathbb{R}$. Then,

$$\tau_C(\omega) := \sup\{t \geq 0 | X_t(\omega) = C\}$$

is not a stopping time.

1.3 Itô Integrals

Now that we have a solid foundation of probability theory, we have the tools to move over to stochastic calculus, namely, (heurestically) constructing the Itô integral

$$\int_a^b X_t \, \mathrm{d}W_t.$$

One could reasonably ask that why are we not satisfied with the Lebesgue-Stieltjes integral? Quite simply, it is due to the fact that the Brownian motion is nowhere differentiable, i.e. it cannot be differentiated with respect to time. This phenomenon is closely related to the following two concepts: total variation and quadratic variation.

Definition 1.29. *(Total Variation and Quadratic Variation).* Let f be a real-valued and continuous function. Then the Total Variation of f is defined on an interval $[a, b] \subset \mathbb{R}$ as

$$\mathrm{TV}(f(t), a, b) = \lim_{\Delta t_k \to 0} \sum_{t_k \leq t} |f(t_{k+1}) - f(t_k)|,$$

where $a = t_1 < t_2 < \ldots < t_n = b$ and $\Delta t_k = t_{k+1} - t_k$. Moreover, the Quadratic Variation of f is then similarly

$$\langle f, f \rangle(t) = \lim_{\Delta t_k \to 0} \sum_{t_k \leq t} |f(t_{k+1}) - f(t_k)|^2.$$

In addition, in the case of $a = 0$ and $b = t$, we often omit the arguments in the notation of total variation.

Remark 1.30. Note that both total and quadratic variations can thus be defined for any continuous stochastic process $X_t(\omega)$. In addition, we say that if $\mathrm{TV}(f(t)) = \infty$, then f has infinite variation.

The natural question that follows is that does the Brownian motion have infinite variation?

Lemma 1.31. Let W_t be the Brownian Motion. Then $\mathrm{TV}(W_t, a, b) = \infty$ almost surely for all $a < b$. Moreover, $\langle W, W\rangle(b) = b - a$ almost surely for any partition of the interval $[a, b]$.

That is, the Brownian motion has infinite variation but *finite* total variation. In fact, by letting $a = 0$ and $b = t$, the lemma above implies $\langle W, W\rangle(t) = t$. Informally, this leads to an observation

$$\mathrm{d}W_t\,\mathrm{d}W_t = \mathrm{d}t,$$

to which we return later. The calculations above and the non-zero quadratic variation would cause troubles in traditional integration. That is, the need for the Itô integral exists, but what does it mean?

Following the suite of Øksendal 2003, which contains a more thorough introduction the for Itô integral, we assume that we are interested in modeling "noise" around some signal or information: let $b(t, x)$ $\sigma(t, x)$ be some functions, and then a reasonable model for a "noisy" process would be

$$\frac{\mathrm{d}X_t}{\mathrm{d}t} = b(t, X_t) + \sigma(t, X_t)V_t,$$

where V_t represents random fluctuations in the process. Discreticizing this equation with respect to a partition $0 = t_0 < t_1 < \ldots < t_n = t$ of interval $[0, t]$ yields incremental equation

$$X_k - X_{k-1} = b(t, X_k)\Delta t_k + \sigma(t, X_k)V_{t_k}\Delta t_k.$$

If we wish that V_t represents "noise" in a reasonable manner, we would like that it satisfied the conditions (i)–(iii) in the definition of the Brownian motion. It turns out that the only process that can be represented as $V_{t_k}\Delta t_k$ while satisfying the conditions (i)–(iii) indeed is the Brownian motion W_t. Using this observation, we can use the above incremental equation to solve for X_k:

$$X_k = X_0 + \sum_{j=0}^{k-1} b(t_j, X_j)\Delta t_j + \sum_{j=0}^{k-1} \sigma(t_j, X_j)\Delta W_j.$$

If we took the limit $\Delta t_j \to 0$, would the right hand side of the above equation survive? If yes, using the standard integral notation would lead us to

$$X_t = X_0 + \int_0^t b(s, X_s)\,\mathrm{d}s + \text{"}\int_0^t \sigma(s, X_s)\,\mathrm{d}W_s\text{"},$$

where, slightly changing notation to highlight the ω-dependence and considering a certain function f, the term

$$\text{"} \int_0^t f(s,\omega)\, \mathrm{d}W_s(\omega) \text{"},$$

given that it exists, would be our desired Itô Integral. The remainder of the integral equation above, for example its solvability, is treated in Section 1.4.

That is, we will now present a ladder of lemmas leading up to the definition of an Itô Integral for the following class of functions:

Definition 1.32. (*Class $\mathcal{V}$ of functions.*) Let $\mathcal{V} = \mathcal{V}(S,T)$ be the class of functions

$$f(t,\omega) : [0,\infty) \times \Omega \to \mathbb{R}$$

which has the following properties[14] for some $S < T$:

(i) $(t,\omega) \to f(t,\omega)$ is $\mathcal{B}([0,\infty)) \times \mathcal{F}$-measurable,

(ii) $f(t,\omega)$ is $\mathcal{F}_t$-adapted, and

(iii) $\mathbb{E}[\int_a^b f(t,\omega)^2\, \mathrm{d}t] < \infty$.

For most of these lecture notes we assume our functions to belong to the class $\mathcal{V}(0,T)$.

Traditional method in constructing many integrals was hinted in the the footnote in Definition 1.15: take piecewise constant simple functions and generalize your findings. Here we consider elementary functions of the form

$$\phi(t,\omega) = \sum_j e_j(\omega) 1_{[t_j,t_{j+1}]}(t)$$

with $\phi \in \mathcal{V}$ and e_j being constants (in time, they still depend on ω). Again, following Øksendal 2003, we take the following steps, leading up to Itô integral for $f \in \mathcal{V}(0,T)$:

(Step 1): Approximate bounded and continuous $g \in \mathcal{V}$ with elementary functions.

(Step 2): Then, approximate bounded $h \in \mathcal{V}$ with g.

(Step 3): Finally, approximate $f \in \mathcal{V}$ with a h.

Roughly, this seems like a good plan. We will follow this suite to defining the Itô integral. Full proof of the existence of the integral would be outside the scope of this material, so we construct a sequence of lemmas without proof to outline the proper construction of the Itô integral. However, we first need to start with definitions. For a more exhaustive presentation, see e.g. Øksendal 2003. Note, that all the following convergences happen in L^2.

[14] In other words, $f(t,\omega) \in \mathrm{L}^2[S,T]$

Lemma 1.33. (Step 1): For $g \in \mathcal{V}$ bounded and continuous, there exists a sequence $\phi_n \in \mathcal{V}$ of elementary functions for which

$$\mathbb{E}\left[\int_0^T (g-\phi_n)^2 \,\mathrm{d}t\right] \to 0 \quad \text{as } n \to \infty.$$

Lemma 1.34. (Step 2): For $h \in \mathcal{V}$ bounded, there exists a sequence $g_n \in \mathcal{V}$ of bounded and continuous functions for which

$$\mathbb{E}\left[\int_0^T (h-g_n)^2 \,\mathrm{d}t\right] \to 0 \quad \text{as } n \to \infty.$$

Lemma 1.35. (Step 3): For $f \in \mathcal{V}$, there exists a sequence $h_n \in \mathcal{V}$ of bounded functions for which

$$\mathbb{E}\left[\int_0^T (f-h_n)^2 \,\mathrm{d}t\right] \to 0 \quad \text{as } n \to \infty.$$

Together, the three lemmas motivate and secure the existence of the following definition:

Definition 1.36. *(Itô Integral).* Let $f \in \mathcal{V}(0,T)$ and $\phi_n(t,\omega)$ a sequence of elementary functions. Then, we define the Itô integral of f as

$$\int_0^T f(t,\omega)\,\mathrm{d}W_t(\omega) = \lim_{n\to\infty}\int_0^T \phi_n(t,\omega)\,\mathrm{d}W_t(\omega)$$

in L^2 with

$$\mathbb{E}\left[\int_0^T \left(f(t,\omega-\phi_n(t,\omega)\right)^2 \,\mathrm{d}t\right] \to 0 \quad \text{as } n \to \infty.$$

The definition basically states, that the value of the integral should be reasonable close to

$$\lim_{\Delta t_k\to 0}\sum_{t_k\le t} X_{t_j}\left(W_{t_{k+1}}-W_{t_k}\right).$$

With a sufficiently tight partition, this definition is fairly close "standard" definitions of integrals: we multiply the value of the integrand, X, with the length of a narrow interval $W(t_{k+1})-W(t_k)$, and sum over these multiplications.

Remark 1.37. The Itô integral is a random variable.

Moreover, the Itô integral has many extremely useful properties.

Proposition 1.38. *(Itô Isometry).* Let $f(t,\omega)$ satisfy the conditions in definition 1.36. Then

$$\mathbb{E}\Big[\Big(\int_0^T f(t,\omega)\,\mathrm{d}W_t\Big)^2\Big] = \mathbb{E}\Big[\int_0^T f(t,\omega)^2\,\mathrm{d}t\Big].$$

Proof. Outline of the proof: We note, that for simple functions we have $\mathbb{E}[e_i e_k \Delta W_i \Delta W_k] = \mathbb{E}[e_k^2](t_{k+1} - t_k)$ if $i = j$, and 0 otherwise due to the independent increments of the Brownian motion. Thus, if f is a simple function between a and b with increments e_k, then by calculating the square with the definition of the Itô sum, we get

$$\mathbb{E}\Big[\Big(\int_a^b f\,\mathrm{d}W_t\Big)^2\Big] = \sum_k \mathbb{E}[e_k^2](t_{k+1} - t_k) = \mathbb{E}\Big[\int_a^b f^2\,\mathrm{d}W_t\Big].$$

Then, the result follows by approximating more general function X with a sequence of simple functions. □

A similiar, but perhaps a bit more counterintuitive result, is the following:

Proposition 1.39. Let f satisfy the conditions in 1.36. Then

$$\mathbb{E}\Big[\int_0^T f\,\mathrm{d}W_t\Big] = 0.$$

Proof. The result follows from similar calculations as the Itô Isometry. Note, indeed, that for simple functions f and $s < t$ we have that

$$\mathbb{E}\Big[\int_0^T f_s\,\mathrm{d}W_s\Big] = \sum_k \mathbb{E}[f_{t_k}](W_{t_{k+1}} - W_{t_k}) = 0,$$

since $W_{t_{k+1}} - W_{t_k}$ is independent of f_{t_k} by the definition of W (independent increments) as f is $\mathcal{F}_t$-adapted. The result follows by approximating more general functions with a sequence of simple functions. □

It should be noted that despite our usage of the class $\mathcal{V}(0, T)$, the integral can be equally defined for over any $a < b$. Moreover, the interval we integrate over can be separated as with ordinary integrals, and the integral as an operator preserves affinity. The following example is crucial in understanding some details in the particular construction.

Example 1.8. In the definition 1.36, it is crucial that we consider forward increments. Indeed, consider

$$\int_a^b W_t\,\mathrm{d}W_t.$$

Now the integrand W_t is not simple, and we approximate the integral as

$$\int_a^b W_t\,\mathrm{d}W_t = \lim_{\Delta t_k \to 0} \sum_i W_{t^*}(W_{t_{i+1}} - W_{t_i})$$

for $t^* \in \{t_{i+1}, t_i\}$. Consider the two following cases:

(i) Let $t^* = t_i$. Now W_{t^*} and $(W_{t_{i+1}} - W_{t_i})$ in the above sum are independent due to independent increments, and thus $\mathbb{E}\left[\int_a^b W_t\,\mathrm{d}W_t\right] = 0$.

(ii) Let $t^* = t_{i+1}$. Now W_{t^*} and $(W_{t_{i+1}} - W_{t_i})$ are not independent, and by multiplying them we get

$$\int_a^b W_t \,\mathrm{d}W_t = \lim_{\Delta t_k \to 0} \sum_i \Big((W_{t_{i+1}} - W_{t_i})^2 + W_{t_i}(W_{t_{i+1}} - W_{t_i}) \Big) ,$$

and moreover by similar calculations as in proposition 1.22 and noting that the partitions Δt_i indeed partitite the interval $b - a$,

$$\mathbb{E}\left[\int_a^b W_t \,\mathrm{d}W_t \right] = \sum_i \Delta t_i = b - a.$$

That is, with slightly altering the point on which the integral's value is evaluated, the integral breaks. In fact, it is possible to define an Itô integral using alternative evaluation points, for example, as the so-called Stratonovich Integral[15].

The following concept is closely related to the expectations of Brownian motions and Itô integrals.

Definition 1.40. *(Martingale).* Let M_t be an integrable stochastic process on $(\Omega, \mathcal{F}, \{\mathcal{F}_t\}_{t\geq 0}, \mathbb{P})$. We say that M_t is a martingale with respect to $\mathcal{F}_t$ if the following conditions hold:

(i) $M_t \in \mathcal{F}_t$ for each t, and

(ii) $\mathbb{E}\left[M_t | \mathcal{F}_s\right] = M_s$ for $s \leq t$.

In the above definition, the second condition is called the martingale condition and its the defining feature in martingale processes. To reiterate, if a stochastic process is a martingale, it means that on expectation its value will stay same over the upcoming time period. In gambling and other similar applications, it corresponds to the notion of a ”fair game”. The following example highlights a usual technique in showing if a process is a martingale.

Example 1.9. A Brownian motion is a martingale. Indeed, by the properties of conditional expectation we have for $s < t$ that

$$\mathbb{E}\left[W_t | \mathcal{F}_s\right] = \mathbb{E}\left[(W_t - W_s) + W_s | \mathcal{F}_s\right] = \mathbb{E}\left[W_t - W_s | \mathcal{F}_s\right] + W_s = W_s.$$

For more similar exercises, one should look begin with problem 105. The following could also be listed as a practical example, but it is fairly groundbreaking for the contents of this material so we will present it separately.

Proposition 1.41. Let X be a stochastic process and W the Brownian motion and $s < t$. Then, the Itô integral $\int_0^t X_s \,\mathrm{d}W_s$ is a martingale.

[15] For this, see, again, Øksendal 2003.

Proof. Follows with similar calculations as in example 1.9 after noting that

$$\mathbb{E}\left[\int_0^t X_s \,\mathrm{d}W_s - \int_0^s X_h \,\mathrm{d}W_h \middle| \mathcal{F}_s\right] = 0.$$

□

For the next section, recall the equation we used to motivate the need for the Itô integral

$$X_t = X_0 + \int_0^t b(s, X_s)\,\mathrm{d}s + \int_0^t \sigma(s, X_s)\,\mathrm{d}W_s.$$

Next, we discuss its solvability.

1.4 Stochastic Differential Equations

Recall what we often mean by a (deterministic) differential equation with respect to time t: for some function $k(t)$, we are interested in solving for

$$\mathrm{d}Y(t) = k(t)\,\mathrm{d}t.$$

In the stochastic world, we build upon a similar idea but add some noise in to the equation. In other words, we incorporate "randomness" into the differential equation by modelling noise with Brownian motions precisely as we did when introducing the Itô integral. The following definition will be a core building block for the rest of this material.

Definition 1.42. *(Itô Process).* Let W_t be a d-dimensional Brownian motion, μ_t a $n \times 1$ drift coefficient vector, and σ_t a $n \times d$ diffusion coefficient matrix, with individual elements of these matrices satisfying $\mu \in \mathcal{V}$ and $\sigma \in \mathcal{V}$[16]. Then, an d-dimensional Itô process X_t driven by W_t has a stochastic differential equation (SDE) of the following form:

$$\mathrm{d}X_t = \mu_t\,\mathrm{d}t + \sigma_t\,\mathrm{d}W_t,$$

or equivalently,

$$X_t - X_0 = \int_0^t \mu_s\,\mathrm{d}s + \int_0^t \sigma_s\,\mathrm{d}W_s.$$

First, note that X_t is a $\mathcal{F}_t^W$-adapted stochastic process. We often use the following definition for Itô processes. Note that in this sense, the concept of Itô diffusion generalizes Brownian motion.

Definition 1.43. *(Itô Diffusion).* We call X_t an Itô diffusion if $\mu = \mu(t, X_t)$ and $\sigma = \sigma(t, X_t)$ are deterministic functions of t and X_t. Moreover, X_t is a time-homogeneous Itô diffusion if $\mu = \mu(X_t)$ and $\sigma = \sigma(X_t)$ depend on time only through X_t.

[16] Particularly so that their integrals are finite in expectation – however, this is not very precise.

Time-homogeneity has some nice properties from the perspectives of martingale processes and stopping times. However, while we defined Itô diffusions and processes, did we forget something crucial? Do these definitions even make any sense? Recall that they are represented as some stochastic differential equations – for such equations, solvability often equals some type of existence and uniqueness result.

Theorem 1.44. *(Existence and Uniqueness of an Itô process).* Let $\mu : [0,T] \times \mathbb{R}^n \to \mathbb{R}^n$ and $\mu : [0,T] \times \mathbb{R}^n \to \mathbb{R}^{n \times d}$ be measurable functions satisfying the $C-$Lipschitz

$$|\mu(t,x) - \mu(t,y)| + |\sigma(t,x) - \sigma(t,y)| \leq C|x-y|$$

and the linear growth

$$|\mu(t,x)| + |\sigma(t,x)| \leq D\,(1 + |x|)$$

conditions for some constants C and D, $C, D < \infty$. Then the SDE

$$\begin{cases} \mathrm{d}X_t = & \mu(t,X_t)\,\mathrm{d}t + \sigma(t,X_t)\,\mathrm{d}W_t \\ \hat{X}_t = & x_0, \end{cases}$$

has a unique solution X_t satisfying the conditions in the definition 1.36.

Proof. Omitted by being too technical and lengthy, see Theorem 5.2.1 in Øksendal 2003. □

Do note, that this means that if either of the Lipschitz or linear growth conditions fail, there might not be a solution! The condition, roughly speaking, ensures that the solution does not explode, i.e. that $|X_t(\omega)| \nrightarrow \infty$ even for the wildest realizations of ω. For such an example, consider the following from the deterministic world.

Example 1.10. *(Non-example, where Lipschitz is violated).* Consider the following two ordinary differential equations (ODEs):

(i) The following ODE has no global solution:

$$\begin{cases} \mathrm{d}X_t = & X_t^2\,\mathrm{d}t, \quad t > 1 \\ X_t = & \frac{1}{1-t} \quad t \in [0,1), \end{cases}$$

(ii) The following ODE has multiple solutions, namely $X_t = (t-a)^3 \vee 0$ for any $a > 0$ and 0 for non-positive a, i.e. it is ill-defined:

$$\begin{cases} \mathrm{d}X_t = & 3X_t^{\frac{2}{3}}\,\mathrm{d}t \\ X_0 = & x_0. \end{cases}$$

It should be noted, that as for many differential equations, solution concepts for SDEs can be either strong or weak. The uniqueness we obtain in the above result is of strong type, and in cases where a weak uniqueness is achieved, we refer that the possible multiple solutions are identical in their distribution.

1.5 Itô's Formula

This material, along with many others, is devoted to study the dynamics of (stochastic) differential equations. Again, an example from the deterministic world: consider the fundamental theorem of calculus. Recall that if F is a function defined for all $x \in (a,b)$ for which

$$F(x) = \int_a^x f(t)\,\mathrm{d}t,$$

then F is uniformly continuous and differentiable on the interval with $\frac{\mathrm{d}F(t)}{\mathrm{d}t} = f(x)$. Does this work with Itô diffusion and Brownian motion? Of course not, as Brownian motion is nowhere differentiable! However, Itô has granted us a very powerful tool to further understand the dynamics of Brownian motion despite the difficulties in differentiating it.

Proposition 1.45. *(Itô's Formula for Brownian Motion).* Let $f \in C^2(\mathbb{R})$[17] and W_t the standard Brownian motion. Then $f(x)$ has a stochastic differential given by

$$\mathrm{d}f(x) = \frac{\partial f(x)}{\partial x}\,\mathrm{d}W_t + \frac{1}{2}\frac{\partial^2 f(x)}{\partial x^2}\,\mathrm{d}t.$$

Proof. Sketch of the proof[18]. First, assume that f and its derivatives are bounded. Then, take a suitable partition from 0 to t, and consider

$$f(W_t) - f(0) = \sum_{t_k \leq t} \left(f(W_{t_k}) - f(W_{t_{k-1}})\right).$$

By Taylor's formula we get

$$f(W_{t_k}) - f(W_{t_{k-1}}) = \frac{\partial f(W_{t_k})}{\partial W_{t_k}}\left(f(W_{t_k}) - f(W_{t_{k-1}})\right) + \frac{1}{2}\frac{\partial^2 f(W_{t_k})}{\partial W_{t_k}^2}\left(f(W_{t_k}) - f(W_{t_{k-1}})\right)^2$$

by noting that the higher order terms vanish and the "approximation" is exact. Then

$$f(W_t) - f(0) = \sum_{t_k \leq t} \frac{\partial f(W_{t_k})}{\partial W_{t_k}}\left(f(W_{t_k}) - f(W_{t_{k-1}})\right) + \frac{1}{2}\frac{\partial^2 f(W_{t_k})}{\partial W_{t_k}^2}\left(f(W_{t_k}) - f(W_{t_{k-1}})\right)^2,$$

and by the construction of the Itô integral we obtain

$$\sum_{t_k \leq t} \frac{\partial f(W_{t_k})}{\partial W_{t_k}}\left(f(W_{t_k}) - f(W_{t_{k-1}})\right) \to \int_0^t \frac{\partial f(W_t)}{\partial W_t}\,\mathrm{d}W_t.$$

What is left for us to show is that

$$\sum_{t_k \leq t} \frac{\partial^2 f(W_{t_k})}{\partial W_{t_k}^2}\left[\left(W_{t_k} - W_{t_{k-1}}\right)^2 - \left(t_k - t_{k-1}\right)\right] \to 0$$

[17]This denotes to a class of functions which have continuous second derivatives.

[18]For more information on a proper proof, although not a proper full proof, check Remark 4.5.1. in Björk 2020.

in expectation[19]. This is done by noting that $\sum_{t_k<t} \Delta t_k = t$, and so we can examine the second moment by independent increments and the properties of variance:

$$\begin{aligned}
\mathbb{E}\left[\left(\sum_{t_k \leq t} (\Delta W_{t_k})^2 - t\right)^2\right] &= \operatorname{Var}\left[\sum_{t_k \leq t} (\Delta W_{t_k})^2 - t\right] \\
&= \sum_{t_k \leq t} \operatorname{Var}\left[(\Delta W_{t_k})^2\right] \\
&= \sum_{t_k \leq t} \mathbb{E}\left[(\Delta W_{t_k})^4\right] - \mathbb{E}\left[(\Delta W_{t_k})^2\right]^2 \\
&= \sum_{t_k \leq t} 3(\Delta t_k)^2 - (\Delta t_k)^2 \underset{\Delta t_k \to 0}{\longrightarrow} 0,
\end{aligned}$$

where the last equality comes from the properties of the fourth moment of the normal distribution. □

The Itô formula also handily generalizes for functions which are not time-homogeneous.

Proposition 1.46. *(Itô's Formula for Itô diffusion).* Let X_t be an Itô diffusion, $f(t,x) \in C^{1,2}([0,\infty) \times \mathbb{R})$, and $Y_t = f(t,x)$. Then

$$\mathrm{d}Y_t = \frac{\partial f(t,x)}{\partial x}\,\mathrm{d}W_t + \left(\frac{\partial f(t,x)}{\partial t} + \frac{1}{2}\frac{\partial^2 f(t,x)}{\partial x^2}\right)\mathrm{d}t.$$

Proof. Follows a similar suit as the proof in proposition 1.45. □

Remark 1.47. There are some notations that are a bit hasty in the above calculations. Indeed, note that

$$\int_0^t (\,\mathrm{d}W_s)^2 = \lim_{\delta t_k \to 0} \sum_{t_k \leq t} (\Delta W_{t_{k-1}}) = t = \int_0^t \mathrm{d}s.$$

This would give us a reason to argue that $(\,\mathrm{d}W_t)^2 = \mathrm{d}t$. With a similar set of heuristics, we use the following notation:

(i) $(\,\mathrm{d}t)^2 = 0$,

(ii) $\mathrm{d}t\,\mathrm{d}W_t = 0$, and

(iii) $(\,\mathrm{d}W_t)^2 = \mathrm{d}t$

Indeed, consider some $p > 1$ and partition P with N intervals. Then

$$\int_0^t (\,\mathrm{d}s)^p = \lim_{\Delta t_k \to 0} \sum_{t_k \leq t} (\Delta t_k)^p = \lim_{\Delta t_k \to 0} N \frac{t}{N}^p = 0.$$

[19] I.e. in L^2 sense.

As usual, an example may be in place.

Example 1.11. Use Itô's formula to calculate

$$\int_0^t W_s \,\mathrm{d}W_s.$$

We consider $f(t,x) = \frac{1}{2}x^2$, which is clearly continuously differentiable, and let $Y_t = f(t, W_t)$. Note that now $Y_0 = 0$. Now by Itô's formula:

$$\mathrm{d}Y_t = W_t \,\mathrm{d}W_t + \frac{1}{2}\,\mathrm{d}t,$$

which is equivalent to

$$\frac{1}{2}W_t^2 = Y_0 + \int_0^t W_s \,\mathrm{d}W_s + \int_0^t \frac{1}{2}\,\mathrm{d}t.$$

Rearranging the equation and calculating the integral over a constant, we find that

$$\int_0^t W_s \,\mathrm{d}W_s = \frac{1}{2}\left(W_s^2 - t\right).$$

Let us generalize the Itô's formula to Itô processes.

Proposition 1.48. *(Itô's formula for Itö Process).* Let X_t be an Itô process given by

$$\mathrm{d}X_t = \mu(t)\,\mathrm{d}t + \sigma(t)\,\mathrm{d}W_t{}^{20},$$

and let $f(t,x) \in C^{1,2}([0,\infty) \times \mathbb{R})$, and $Y_t = f(t, X_t)$. Then

$$\mathrm{d}Y_t \left(f_t + f_x\mu(t) + \frac{1}{2}f_{xx}\sigma(t)^2\right)\mathrm{d}t + f_x\sigma_t\,\mathrm{d}W_t.$$

Proof. Rough idea of the argument (for more exhaustive proof, see e.g. partial proof in under proposition 1.45 and a referred discussion in Björk 2020. However, the idea is to use our earlier iteration of the Itô formula, omitting arguments:

$$\begin{aligned}\mathrm{d}Y_t =& f_t\,\mathrm{d}t + f_x\,\mathrm{d}X_t + \frac{1}{2}f_{xx}(\mathrm{d}X_t)^2 \\ =& f_t\,\mathrm{d}t + f_x\,\mathrm{d}\left(\mu(t)\,\mathrm{d}t + \sigma(t)\,\mathrm{d}W_t\right) + \frac{1}{2}f_{xx}\left(\mu(t)\,\mathrm{d}t + \sigma(t)\,\mathrm{d}W_t\right)^2,\end{aligned}$$

from which the result follows by expanding the squared binomial and rearranging terms. □

[20] Note that here we slightly alter the notation for $\mu_t = \mu(t)$ and $\sigma_t = \sigma(t)$ to avoid confusion with the subscript f_t, which for function f denote the partial derivative with respect to t. Similarly, f_x and f_{xx} respectively denote the first and second partial derivatives with respect to x.

This is a very useful result, and many of our subsequent applications follow from this result. In addition, Itô processes itself are useful from the perspective of mathematical finance.

Example 1.12. *(Some Useful Financial Models).*

(i) Bacheler (1900): $\mathrm{d}X_t = \mu \, \mathrm{d}t + \sigma \, \mathrm{d}W_t$.

(ii) Geometric Brownian Motion, Black-Scholes-Merton (1973): $\mathrm{d}X_t = \mu X_t \, \mathrm{d}t + \sigma X_t \, \mathrm{d}W_t$.

(iii) Ornstein-Uhlenbeck, Vaseicek model: for some positive $\theta > 0$: $\mathrm{d}X_t = -\theta X_t \, \mathrm{d}t + \sigma \, \mathrm{d}W_t$.

One of these is of particular interest to us. Let us solve it!

Example 1.13. *(Solving the Geometric Brownian Motion).* Let $Y_t = \log(X_t)$ with some $X_0 \neq 0$. Now by Itô's formula:

$$\begin{aligned}
\mathrm{d}Y_t =& \frac{1}{X_t} \, \mathrm{d}X_t - \frac{1}{2} \frac{1}{X_t^2} (\, \mathrm{d}X_t)^2 \\
=& \mu \, \mathrm{d}t + \sigma \, \mathrm{d}W_t + \left(-\frac{1}{2} \sigma^2 \, \mathrm{d}t \right) \\
=& \left(\mu - \frac{1}{2} \sigma \right) \, \mathrm{d}t + \sigma \, \mathrm{d}W_t.
\end{aligned}$$

This is equivalent with

$$\begin{aligned}
Y_t =& Y_0 + \int_0^t \left(\mu - \frac{1}{2} \sigma^2 \right) \, \mathrm{d}s + \int_0^t \sigma \, \mathrm{d}W_s \\
=& \log(X_0) + \left(\mu - \frac{1}{2} \sigma^2 \right) t + \sigma W_t.
\end{aligned}$$

Finally,

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

In addition, Itô's formula is very useful when calculating expectations of transformations of the Brownian motion.

Example 1.14. Let us calculate $\mathbb{E}\left[e^{\alpha W_t}\right]$ for some $\alpha \in \mathbb{R}$. From elementary probability courses, we are familiar with the fact that if $X \text{lognormal}(\mu, \sigma^2)$, then $\mathbb{E}[X] = e^{\mu + \frac{1}{2}}$. Note that $\alpha W_t \sim N(0, \alpha^2 t)$, and thus

$$\mathbb{E}\left[e^{\alpha W_t}\right] = e^{\frac{1}{2}\alpha^2 t}.$$

In a normal world, this would suffice as a result. However, this is not a course where we like to remember statistical properties of distributions (unless we are forced to). To calculate the expectation with the help of Itô's formula, let $Y_t = e^{\alpha W_t}$. Now, by Itô's formula, and by noting that $Y_0 = 1$,we have

$$\begin{aligned}\mathrm{d}Y_t =&\alpha e^{\alpha W_t}\,\mathrm{d}W_t + \frac{1}{2}\alpha^2 e^{\alpha W_t}\,\mathrm{d}t\\ =&\alpha Y_t\,\mathrm{d}W_t + \frac{1}{2}\alpha^2 Y_t\,\mathrm{d}t.\end{aligned}$$

The corresponding integral form is

$$Y_t = Y_0 + \frac{1}{2}\alpha^2\int_0^t Y_s\,\mathrm{d}s + \alpha\int_0^t Y_s\,\mathrm{d}W_s.$$

Taking expectations and denoting $M_t := \mathbb{E}\left[Y_t\right]$, we notice that the above reduces to an ODE

$$\mathrm{d}M_t = \frac{1}{2}\alpha^2 M_t$$

with an initial condition $M_0 = 1$. This is clearly solved by

$$M_t = M_0 e^{\frac{1}{2}\alpha^2 t} = e^{\frac{1}{2}\alpha^2 t},$$

which answers our original question.

Finally, after using examples to understand the lower-dimension Itô's formula, we are ready for the $n-$dimensional formula. However, we begin with a simple remark about notations.

Remark 1.49. *(Matrix Notations and Froberius Product).* We denote $A = [a_{ij}]$ and $B = [b_{ij}]$ to be $m \times n$ matrices, and define their Froberius product as

$$\langle A, B\rangle = \sum_{i,j} a_{ij}b_{ij} = \mathrm{Tr}(AB^T),$$

where $\mathrm{Tr}(\cdot)$ denotes a trace of a matrix.

Now, we can define the $n-$dimensional Itô Formula only to notice how it's strikingly similar to our single-dimension cases with the notation presented above.

Proposition 1.50. *(n-Dimensional Itô Formula).* Let $f(t,x) \in C^{1,2}$ and denote

$$f_x = \left(\frac{\partial f}{\partial x_1}, \ldots, \frac{\partial f}{\partial x_n}\right)^T, \quad \text{and} \quad f_{xx} = \left[\frac{\partial^2 f}{\partial x_i x_j}\right].$$

In addition, let X_t be an $n-$dimensional Itô process as in the definition 1.42, and let $Y_t = f(t, X_t)$. Then, omitting arguments,

$$\begin{aligned}\mathrm{d}Y_t =&f_t + \sum_{i=1}^n \frac{\partial f}{\partial x_i}\,\mathrm{d}X_t^i + \frac{1}{2}\sum_{i=1}^n\sum_{j=1}^n \frac{\partial^2 f}{\partial x_i x_j}\,\mathrm{d}X_t^i\,\mathrm{d}X_t^j\\ =&\left(f_t + \langle \mu, f_x\rangle + \frac{1}{2}\langle \sigma\sigma^T, f_{xx}\rangle\right)\mathrm{d}t + \langle f_x, \sigma\,\mathrm{d}W_t\rangle.\end{aligned}$$

Proof. Omitted. □

Remark 1.51. Note that here the each separate process X_t^i include a separate Brownian motion W_t^i, which are assumed to be independent. When the Brownian motions are independent, we get

(i) $\mathrm{d}W_t^i \,\mathrm{d}W_t^j = 0$ for $i \neq j$,

(ii) $(\mathrm{d}W_t^i)^2 = \mathrm{d}t$, and

(iii) $(\mathrm{d}W_t^j)^2 = \mathrm{d}t$.

When dealing with physical phenomena, one example is particularly useful.

Example 1.15. Consider the Euclidian norm $||\cdot||$ and let W_t be a $n-$dimensional Brownian motion starting from some $x_0 \in \mathbb{R}^n$. Moreover, let $f(x) = ||x||^2$. Now

$$f(W_t) = (W_t^1)^2 + \cdots + (W_t^n)^2,$$

and

$$f_x(W_t) = 2W_t, \quad f_{xx} = 2\mathrm{I}_n,$$

where I_n is the $n-$dimensional identity matrix. Now

$$\begin{aligned}\mathrm{d}f(W_t) =& \left(0 + 0 + \frac{1}{2}\langle \mathrm{I}_n, 2\mathrm{I}_n\rangle\right) \mathrm{d}t + \langle 2W_t, \mathrm{d}W_t\rangle \\ =& n\,\mathrm{d}t + \langle 2W_t, \mathrm{d}W_t\rangle,\end{aligned}$$

or equivalently in the integral form,

$$||W_t||^2 = ||x_0||^2 + nt + 2\sum_{i=1}^{n} \int_0^t W_s^i \,\mathrm{d}W_s^i.$$

In the previous iteration of Itô formula we assumed that the individual Brownian motions were uncorrelated[21]. What if they were correlated? We present the final iteration of the Itô formula regarding this notion.

Proposition 1.52. *(Itô formula for correlated diffusions.)* Consider a similar set-up as in 1.50, but with the notion that the individual components of the Brownian motions are not independent. More specifically, assume that for $s < t$ we have

$$\mathrm{Corr}\left(W_t^i - W_s^i, W_t^j - W_s^j\right) = \frac{1}{t-s}\mathbb{E}\left[W_t^i - W_s^i, W_t^j - W_s^j\right] = \rho_{ij},$$

where $\rho = [\rho_{ij}]$ is the so-called correlation matrix. Then

$$\mathrm{d}Y_t = \left(f_t + \langle \mu, f_x\rangle + \frac{1}{2}\langle \sigma\rho\sigma^T, f_{xx}\rangle\right) \mathrm{d}t + \langle f_x, \sigma \,\mathrm{d}W_t\rangle.$$

[21] Recall, that in the case of the increments of Brownian motions, uncorrelatedness also implies independence.

Proof. Omitted. See discussion in, for example, Björk 2020 section 4.8. □

Remark 1.53. The reader is advised to cautious and vigilant behaviour, as now the cross differentials for different Brownian motions are not zero. Instead, under the assumptions of 1.52 we have

$$\mathrm{d}W_t^i \,\mathrm{d}W_t^j = \rho_{ij}\,\mathrm{d}t.$$

We have now presented four different iterations for the Itô formula, all of which are different generalizations of each other. Working with Itô formula is a staple skill for any aspiring person dealing with stochastic processes, and the reader is advised in honing their practical skills with exercises in 1.6. Before moving on to discuss the relationship between SDE's and ODE's, we finish the chapter with an important application of the Itô formulas.

Theorem 1.54. *(Martingale Representation Theorem).* Let M_t be an $\mathcal{F}_t^W$-martingale with a deterministic initial point M_0 such that $M_t \in L^2$. Then there exists a unique $\mathcal{F}_T^W$-adapted process $\{H_t\}_{t\leq T} \in L^2$ such that

$$M_t = M_0 + \int_0^t H_s \,\mathrm{d}W_s \text{a.s.}$$

for all $t \leq T$.

Proof. Follows from the Itô formula used to an arbitrary martingale. □

This theorem is quite remarkable. We knew previously that any Itô integral is a martingale, but the result of the theorem gives us the inverse statement: every martingale can be represented as an Itô integral (hence the name of the theorem). The theorem has many practical applications in, e.g., finance, where one could research a relationship between a contract value, hedging strategy, and a stock price modelled as some transformation of W_t. Moreover, there is another example which ties this chapter to the next one, so we present it as a bridge between the two.

Example 1.16. *(Martingale repr. theorem and the Heat Equation).* Consider a simple case where $M_t = f(t, W_t)$. Now, by Itô's formula,

$$M_t + M_0 + \int_+^t f_s + \frac{1}{2} f_{xx} \,\mathrm{d}s + \int_0^t f_x \,\mathrm{d}W_s.$$

Based on the martingale representation theorem, we have to have that

$$f_s + \frac{1}{2} f_{xx} = 0,$$

which is precisely the so-called Heat Equation.

The example was meant to be a very short teaser to the heat equation, which will be more closely examined in chapter 3.

1.6 Exercises for Chapter 1

In the exercises, a sign (*) before the problem number denotes a slightly more involved exercise.

Probability

101. Throw a 6-sided die, write down the σ-algebra generated by 1 and 4. (Hint: if you can answer 'did I get a 1' and 'did I get a 4', you can answer all the yes/no questions in this σ-algebra)

102. Let τ_1, τ_2 be $\mathcal{F}_t$-stopping times. Argue that

    (a) $\tau_1 \wedge \tau_2$ is a stopping time

    (b) $\tau_1 \vee \tau_2$is a stopping time

    (c) $\tau_1 + \tau_2$ is a stopping time

    (d) $\tau_1 - \tau_2$ is not a stopping time (Hint: give a counter-example or argue with filtrations)

103. **(Moments of BM)**. By the properties of Brownian motion, we know that $W_t \sim \mathbb{N}(0, t)$ This implies that (by the characteristic function of a Gaussian random variable)
$$\mathbb{E}[e^{iuW_t}] = e^{-\frac{1}{2}u^2 t}.$$
Apply the power series expansion of the exponential function on both sides, compare the terms with the same power of u and show that
$$\mathbb{E}[W_t^4] = 3t^2.$$
More generally, with tiny bit more effort, one can show that for $k \in \mathbb{N}$,
$$\mathbb{E}[W_t^{2k}] = \frac{(2k)!}{2^k k!} t^k.$$

104. **(Moments of BM revisited)**. Let $W_t \in \mathbb{R}$, $W_0 = 0$ be a Brownian motion. Define
$$\beta_k(t) = \mathbb{E}[W_t^k], \quad k = 0, 1, \ldots, t \geq 0.$$

    (a) Use Itô's formula to show that
$$\beta_k(t) = \frac{1}{2}k(k-1)\int_0^t \beta_{k-2}(s)ds, \quad k \geq 2.$$

(b) Show that

$$\mathbb{E}[W_t^{2k+1}] = 0$$

and

$$\mathbb{E}[W_t^{2k}] = \frac{(2k)!t^k}{2^k k!}$$

for $k = 1, 2, \ldots$.

Martingales and Itô Integral

105. **(Martingales)**. Show that the following stochastic processes are martingales:

   (a) $M_t = W_t^2 - t$, (Hint: Write $W_t = W_t - W_s + W_s$, for $s < t$)

   (b) $M_t = \mathbb{E}[X|\mathcal{F}_t]$[22],

   (c) $M_t = t^2 W_t - 2\int_0^t sW_s ds$,

   (d) $M_t = W_1(t)W_2(t)$, where (W_1, W_2) is a 2-dimensional standard Brownian motion.

106. ***(Construction of the Itô integral)**. Let W_t be a standard Brownian motion. Find a sequence of piecewise constant process $\phi_n(t,\omega)$ such that

$$\mathbb{E}[\int_0^t (\phi_n(s) - W_s)^2 ds] \to 0.$$

Compute $\int_0^t \phi_n(s)dW_s$ and show that it converges (in what sense?) to $\frac{1}{2}(W_t^2 - t)$ if we consider finer and finer partitions. Deduce that

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

107. **(Continuation of 106)**. What does the result of 106 tell you? How does it relate to problem 105 (a)?

108. **(Stratonovich integral)**. When using $W_{t_{j^*}}$ to approximate W_t on an interval $[t_j,\ t_{j+1}]$, if $t_{j^*} = t_j$ it yields an Itô integral. Now take $t_{j^*} = \frac{1}{2}(t_j + t_{j+1})$, define the Stratonovich integral by

$$\int_0^T X(t,\omega) \circ dW_t(\omega) = \lim_{\Delta t_j \to 0} X(t_{j^*},\omega)\Delta W_j$$

Use the example $\int_0^T W_t \circ dW_t$ to see that the Stratonovich integral follows the usual chain rule. Compare to exercise 202, does Itô follow the normal chain rule?

[22] This is known as the "Doob's martingale". It can be thought of as the evolving sequence of best approximations to the random variable X based on accumulated information.

109. Argue that the Itô integral $\int_0^t s dW_s$ is a normal random variable. Show that it follows the distribution $N(0, \frac{1}{3}t^2)$.

Remark. In fact, integrate any deterministic function $f(t)$ with respect to the Brownian motion yields a Gaussian process

$$I_t(f) = \int_0^t f(s) dW_s \sim N(0, \int_0^t f^2(s) ds).$$

Itô's Formula and some SDEs

Note that the problems 110–115 are meant to get the reader familiar with a practical usage of the Itô's formula.

In the following exercises, we refer to a Geometric Brownian Motion (GBM) with a the process

$$dX_t = \mu X_t dt + \sigma X_t dW_t, X_0 = x.$$

In this set of exercises, it might be useful to use the idendity

$$\mathrm{d}(tW_t) = W_t \, \mathrm{d}t + t \, \mathrm{d}W_t.$$

110. Use Itô's formula to show that the expectation of a GBM satisfies

$$\mathbb{E}[X_t] = xe^{\mu t}.$$

111. Use Itô's formula to calculate $\int_0^t W_s^2 \, \mathrm{d}W_s$.

112. i) Show that $\int_0^t W_s \, \mathrm{d}s \sim N(0, \frac{1}{3}t^3)$.
   ii) Use Itô's formula to calculate $\int_0^t s \, \mathrm{d}W_s$. (Hint: use $f(t, x) = tx$).

113. Let α_t be some L^2 stochastic process, and define the 1-dimensional stochastic process

$$Z_t = \exp(\int_0^t \alpha_s dW_s - \frac{1}{2} \int_0^t \alpha_s^2 ds).$$

Use Itô's formula to show that $dZ_t = \alpha_t Z_t dW_t$, and conclude that it is a martingale (given that $\alpha_t Z_t \in L^2$). (Remark: This exercise provides a way to construct martingales.)

114. Solve the 1-dimensional SDE $dX_t = X_t dt + dW_t$ with $X_0 = x$ (Hint: multiply both sides with e^{-t}).

115. Solve the 1-dimensional SDE $dY_t = rdt + \alpha Y dW_t$, where $Y_0 = 0, r, \alpha \in \mathbb{R}$. (Hint: multiply both sides with $F_t = \exp(-\alpha W_t + \frac{1}{2}\alpha^2 t))$.

116. **(Brownian bridge)**. Let X_t be the solution to the following SDE with $X_0 = 0$:

$$dX_t = -\frac{1}{1-t} X_t dt + dW_t, \quad 0 \le t < 1.$$

(a) Use Itô's formula to show that $X_t = (1-t) \int_0^t \frac{dW_s}{1-s}$ solves the above SDE. (Hint: consider the process $Y_t = \int_0^t \frac{dW_s}{1-s}$.)

(b) Argue that for $t \in [0,1)$, X_t is a Gaussian random variable with mean 0 and variance $t(1-t)$. Moreover, show that X_t converges to 0 in L^2 as $t \to 1$:

$$\lim_{t \to 1} \mathbb{E}[X_t^2] = 0.$$

(Remark: the process X_t is pinned on both ends, hence a "bridge")

117. **(SDE on a circle)**. Consider the following pair of equations:

$$dX_t = -\frac{1}{2} X_t dt - Y_t dW_t,$$
$$dY_t = -\frac{1}{2} Y_t dt + X_t dW_t.$$

Let $(X_0, Y_0) = (x, y)$ such that $x^2 + y^2 = 1$. Show that $X_t^2 + Y_t^2 = 1$ for all t, i.e. that the process (X, Y) moves within the unit circle.

118. **(Hyperbolic SDE)**. Similarly, consider the following pair of equations

$$dX_t = \frac{1}{2} X_t dt + Y_t dW_t,$$
$$dY_t = \frac{1}{2} Y_t dt + X_t dW_t.$$

Show that $X_t^2 - Y_t^2$ is constant for all t.

119. Let X_t, Y_t be two one-dimensional Itô processes. Prove the following Itô product rule:

$$dX_t Y_t = X_t dY_t + Y_t dX_t + dX_t dY_t.$$

120. ***(Complex BM)**. Given a two-dimensional Brownian motion $(W^{(1)}, W^{(2)})$, define the complex Brownian motion

$$W_t^c := W_t^{(1)} + i W_t^{(2)}.$$

Let $f : \mathbb{C} \to \mathbb{C}$ be a function of the form $f(z) = f_{\mathbb{R}}(z) + i f_{\mathbb{C}}(z)$, for any $z \in \mathbb{C}$, and $f_{\mathbb{R}}, f_{\mathbb{C}} : \mathbb{C} \to \mathbb{R}$. If f is is analytic, i.e. satisfies the Cauchy-Riemann equations:

$$\frac{\partial f_{\mathbb{R}}}{\partial x} = \frac{\partial f_{\mathbb{C}}}{\partial y}, \quad \frac{\partial f_{\mathbb{R}}}{\partial y} = -\frac{\partial f_{\mathbb{C}}}{\partial x},$$

where $z = x + iy$. Show that the identity $df(W_t^c) = f'(W_t^c) dW_t^c$ holds, where f' denotes the complex derivative of f.

Chapter 2

Feynman-Kac Theorem with Applications

We begin by exploring the connection of Partial Differential Equations and Stochastic Differential Equations. Recall from the previous chapter that the solution to

$$\mathrm{d}X_t = \mu(X_t)\,\mathrm{d}t + \sigma(X_t)\,\mathrm{d}W_t$$

is called an Itô diffusion[1]. Originally, the Itô diffusion was used to model a diffusion of a pollen particle in water, and the random fluctuations were explained as the noise driven by a Brownian motion. The same phenomenom, however, could be explained deterministically: the pollen moves in a seemingly random way, but the fluctuations deterministically follow some laws of physics and known dynamics. To explore this notion further, we need to continue building our machinery.

2.1 Infinitesimal Generator

Definition 2.1. *(Infinitesimal Generator).* Let $X_t \in \mathbb{R}^n$ be a time-homogeneous Itô diffusion, and let $f \in C_o^2(\mathbb{R}^n), f : \mathbb{R}^n \to \mathbb{R}$[2]. Then the infinitesimal generator $\mathcal{L}$ of X_t is defined as

$$\left(\mathcal{L}f\right)(x) = \lim_{t\to 0} \frac{1}{t}\left(\mathbb{E}_x\left[f(X_t)\right] - f(x)\right),$$

whenever the limit exists.

Few remarks ought to be in place. First, in the definition we denote an expectation conditionaled with some initial value $\mathbb{E}_x := \mathbb{E}[\cdot|X_0 = x]$. Secondly, the reader noticed that we restrict our definition to the class of time-homogeneous Itô diffusions – this is purely out of laziness. Note that we can always make a non-time-homogeneous diffusion time-homogeneous with a suitable drift coefficient. One remark deserves its own pedestal.

[1] Recall definition 1.43.

[2] Here the notation $C_o^2(\mathbb{R}^n)$ means a collection of twice-differentiable functions with a compact support.

Remark 2.2. When $f(x) = \mathbb{1}_{\{X_t \in D\}}$ for some $D \subset \mathbb{R}^n$, the infinitesimal generator describes the change of a probability distribution.

However, the definition 2.1 is quite abstract. It tells that we are interested in a limit of differences between a function and its expectation under very short time intervals, but it is difficult to calculate when needed. The following theorem is of help.

Theorem 2.3. *(Infinitesimal Generator for Itô diffusion).* Let X_t be a $1-$dimensional Itô diffusion and let $f \in C_o^2(\mathbb{R}^n)$. Then $(\mathcal{L}f)(x)$ exists for all $x \in \mathbb{R}$, and

$$(\mathcal{L}f)(x) = \mu(x)\frac{\partial f}{\partial x} + \frac{1}{2}\sigma^2(x)\frac{\partial^2 f}{\partial x^2}.$$

Moreover, if X_t is an $n-$dimensional Itô diffusion, then

$$(\mathcal{L}f)(x) = \sum_i \mu_i(x)\frac{\partial f}{\partial x_i} + \frac{1}{2}\sum_{i,j}(\sigma\sigma^T)_{ij}\frac{\partial^2 f}{\partial x_i \partial x_j}.$$

Proof. It suffices to show that for a $1-$dimensional Brownian motion[3],

$$\mathcal{L}f = \frac{1}{2}\frac{\partial^2 f}{\partial x^2}.$$

By applying Itô's formula to $f(W_{t+s})$ with $W_t = x$ and using a shorthand notation $f_x = \frac{\partial f}{\partial x}$, we get

$$\begin{aligned}
(\mathcal{L}f)(x) &= \lim_{s\to 0}\frac{1}{s}\left(\mathbb{E}_x\left[f(W_{t+s})\right] - f(x)\right) \\
&= \lim_{s\to 0}\frac{1}{s}\left(\mathbb{E}_x\left[f(x) + \int_t^{t+s} f_x(W_h)\,\mathrm{d}W_h + \int_t^{t+s}\frac{1}{2}f_{xx}(W_h)\,\mathrm{d}h\right] - f(x)\right) \\
&= \frac{1}{2}\frac{\mathrm{d}}{\mathrm{d}s}\left(\int_t^{t+s}\mathbb{E}\left[f_{xx}(W_h)\right]\,\mathrm{d}h\right) \\
&= \frac{1}{2}f_{xx}(W_t) = \frac{1}{2}f_{xx}(x).
\end{aligned}$$

□

Remark 2.4. If we consider $f = f(t,x) \in C_o^{1,2}$, the definition still holds as the infinitesimal generator is only applied with respect to x, not time t.

Does the formula for the infinitesimal generator for Itô diffusion make the concept clearer? Let's take a look at the Itô's formula now:

$$\begin{aligned}
\mathrm{d}f(t,X_t) &= f_t\,\mathrm{d}t + \mu f_x\,\mathrm{d}t + \frac{1}{2}\sigma^2 f_{xx}\,\mathrm{d}t + \sigma f_x\,\mathrm{d}W_t \\
&= \left(f_t + (\mathcal{L}f)(t,X_t)\right)\mathrm{d}t + \sigma f_{xx}\,\mathrm{d}W_t.
\end{aligned}$$

[3] Note that the following condition means, in other words, that the generator of a Brownian motion is $\frac{1}{2}\Delta$, the so-called Laplacian operator.

That is, the generator describes the dynamics of the process over an infinitesimal time period (hence the name)!

The following quick example might be useful for the reader.

Example 2.1. *(Generator of a Geometric Brownian motion).* Let r and σ be real-valued constants. If X_t solves

$$\mathrm{d}X_t = rX_t\,\mathrm{d}t + \sigma X_t\,\mathrm{d}W_t,$$

then

$$(\mathcal{L}f)(x) = rxf_x + \frac{1}{2}\sigma^2 x^2 f_{xx}.$$

2.2 Stochastic Representation for a PDE: the Feynman-Kac Formula

In this section, we introduce the so-called Feynman-Kac formula as an application of the Itô formula.

Consider a PDE of the following form:

$$\begin{cases} u_t + \mathcal{L}u = 0 \\ u(T,x) = \Phi(x), \end{cases}$$

where $(\mathcal{L}u)(t,x) = \mu u_x - \frac{1}{2}\sigma^2 u_{xx}$, Φ is a known function, and we assume that this PDE has a solution in $C^{1,2}$. What happens if we apply Itô's formula to $u(t, X_t)$ from t to T? Here t stands for the time of our initial observation and $T > t$ is a so-called terminal time, i.e. when we are to stop observing the process. So, by Itô:

$$\underbrace{u(T,X_T)}_{\Phi(X_T)} = \underbrace{u(t,X_t)}_{\text{Deterministic}} + \int_t^T \underbrace{(u_s + \mathcal{L}u)}_{0 \text{ by the PDE}}\,\mathrm{d}s + \underbrace{\int_t^T \sigma u_x\,\mathrm{d}W_s}_{0 \text{ in } \mathbb{E}[\cdot]}.$$

Moreover, by taking expectations[4] we have

$$\mathbb{E}_{t,x}[\Phi(X_T)] = \mathbb{E}_{t,x}[u(t,X_t)] + 0,$$

which gives,

$$u(t,x) = \mathbb{E}_{t,x}[\Phi(X_T)].$$

To reiterate, we have now connected a solution to a PDE with an expectation to a random variable. This observation is formalized and generalized by the following theorem:

[4] With the notation $\mathbb{E}_{t,x}[\cdot]$ we emphasize the conditional expectation given the initial value x at some initial observation time t, i.e. $\mathbb{E}[\cdot|X_t = x]$.

Theorem 2.5. *(Feynman-Kac Formula).* Consider a stochastic process X_t which solves

$$\mathrm{d}X_s = \mu(s, X_s)\,\mathrm{d}s + \sigma(s, X_s)\,\mathrm{d}W_s$$

with deterministic $X_t = x$ and $0 < t < s \leq T$. Moreover, let $D \in \mathbb{R}^n$ be a connected[5] open domain. Consider the following deterministic functions

$$\begin{aligned}
&r : \mathbb{R}_+ \times \mathbb{R}^n \to \mathbb{R}_+ \quad \text{(discount rate function)},\\
&\Psi : \mathbb{R}_+ \times \mathbb{R}^n \to \mathbb{R} \quad \text{(running payoff or dividend function)},\\
&\Phi : \mathbb{R}^n \to \mathbb{R} \quad \text{(terminal payoff function)}.
\end{aligned}$$

Then the unique solution to the PDE[6]

$$\begin{cases} u_t + \mathcal{L}u - ru + \Psi = 0 & \text{in } [0, T) \times D \\ u(T, x) = \Phi(x) & \text{in } ([0, T) \times \partial D) \cup (\{T\} \times D)\,, \end{cases}$$

where $u \in C^{1,2}$ is given by

$$u(t, x) = \mathbb{E}_{t,x}\left[\exp\left\{-\int_t^T r(s, X_s)\,\mathrm{d}s\right\}\Phi(X_T) + \int_t^T \exp\left\{-\int_t^s r(s', X_{s'})\,\mathrm{d}s'\right\}\Psi(s, X_s)\,\mathrm{d}s\right].$$

Proof. The simplified sketch of the proof is above. □

Remark 2.6. The statement of the theorem simplifies remarkably when $D = \mathbb{R}^n$ and r is a known constant:

$$u(t, x) = \mathbb{E}_{t,x}\left[e^{-r(T-t)}\Phi(X_T) + \int_t^T e^{-r(s-t)}\Psi(s, X_s)\,\mathrm{d}s\right].$$

To make flesh out the remark, let $\Psi = r = 0$. Then, the contents of the statement are clearly

$$u(t, x) = \mathbb{E}_{t,x}\Big[\Phi(X_T)\Big],$$

as motivated earlier.

Not surprisingly, the theorem is a very useful application of the Itô's formula, and particularly it has many applications in solving terminal and initial value problems.

[5] So-called path-connectedness suffices for us: a domain is connected if we can shrink any closed curve within a domain to a point which remains inside the domain – or that any two points within a set can be connected with a continuous curve that lies within the domain.

[6] Here the notation ∂D refers to the boundary of the set D.

Example 2.2. Solve the following PDE:

$$\begin{cases} u_t + \frac{1}{2}u_{xx} = 0 \\ u(T,x) = x^2. \end{cases}$$

Let $\mathrm{d}X_s = \mathrm{d}W_s$ for $t < s$, but note that $X_t = 0$ while $W_t = 0$. Then, by Feynman-Kac, we have

$$\begin{aligned} u(t,x) &= \mathbb{E}_{t,x}\left[X_T^2\right] \\ &= \mathbb{E}\left[W_T^2 | X_t = x\right] \\ &= \mathbb{E}\left[(x + X_{T-t})^2\right] \\ &= x^2 + \mathbb{E}\left[W_{T-t}^2\right] + 2x\mathbb{E}\left[W_{T_t}\right] \\ &= x^2 + T - t. \end{aligned}$$

If the reader plugs in the solution $x^2 + T - t$ to the original PDE, they notice that the PDE is indeed solved by our solution.

One should also notice that the Feynman-Kac indeed is defined only for a deterministic T. What follows now is a short discussion on if it holds for a stopping time τ, and how it could be generalized. In order to introduce the coveted Dynkin's theorem, we require the following helpful result. Despite being sidelined in this material, Doob's Optional Sampling is very relevant in any stopping time related business.

Proposition 2.7. *(Doob's Optional Sampling).* Let X_t be an integrable martingale and τ a bounded F_t-stopping time. Then

$$\mathbb{E}\left[X_\tau\right] = \mathbb{E}\left[X_{\tau\wedge t}\right] = \mathbb{E}\left[X_0\right].$$

Proof. Omitted. See Gut 2013 section 7 and more specifically, corollary 7.2. □

Now, back to Dynkin's. The following result eerily resembles the Itô formula but for stopping times – note how the structure of initial value summed with a time integral fits our earlier observation that an Itô integral vanishes in expectation.

Theorem 2.8. *(Dynkin's Formula).* Let X_t be an Itô diffusion with $X_0 = x$ and $f \in C_o^2(\mathbb{R}^n)$. Moreover, let τ be a $\mathcal{F}_t$-stopping time such that $\mathbb{E}[\tau] < \infty$. Then

$$\mathbb{E}_x\left[f(X_\tau)\right] = f(x) = \mathbb{E}_x\left[\int_0^\tau \mathcal{L}f(X_s)\,\mathrm{d}s\right].$$

Proof. The proof relies on two facts. First, for any integer n, we recall that $\tau \wedge n$ is a stopping time. Secondly, Doob's optional sampling theorem.

By an application of Itô's formula, it suffices to show that

$$\mathbb{E}_x\left[\sum_{i,k}\int_0^{\tau}\sigma_{ik}f_{x_i}\,\mathrm{d}W_k\right]=0.$$

Noting that $f(X_s)$ is $\mathcal{F}_s$-measurable, we get

$$\mathbb{E}_x\left[\int_0^{k\wedge\tau}f(X_s)\,\mathrm{d}W_s\right]=\mathbb{E}_x\left[\int_0^{k}f(X_s)\mathbb{1}_{s\leq\tau}\,\mathrm{d}W_s\right]=0$$

for some integer k. Moreover, since any bounded function f is bounded such that $|f|\leq M$ for some $M<\infty$, we have by Itô isometry

$$\mathbb{E}_x\left[\left(\int_0^{\tau}f(X_s)\,\mathrm{d}W_s-\int_0^{k\wedge\tau}f(X_s)\,\mathrm{d}W_s\right)^2\right]=\mathbb{E}_x\left[\int_{k\wedge\tau}^{\tau}f(X_s)^2\,\mathrm{d}s\right]\leq M^2\mathbb{E}_x\left[\tau-\tau\wedge k\right].$$

Now, when $k\to\infty$, we see that

$$\int_0^{k\wedge\tau}f(X_s)\,\mathrm{d}W_s\to^{L^2}\int_0^{\tau}f(X_s)\,\mathrm{d}W_s,$$

and the above inequality implies that

$$\mathbb{E}_x\left[\int_0^{\tau}f(X_s)\,\mathrm{d}W_s\right]=0.$$

The result then follows by noting that for all indeces i, f_{x_i} is a bounded function for $f\in C_o^2$. $\square$

Dynkin's formula has a lot of uses. Consider the following example.

Example 2.3. *(Hitting times).* Let $W_t\in\mathbb{R}$ be a Brownian motion. Let τ be a first time for W_t to exit an interval $(-a,a)$, i.e.

$$\tau=\inf\left\{t\leq 0:W_t\notin(-a,a)\right\}.$$

What is $E[\tau]$?

First, we assume that $E[\tau]<\infty$[7]. Then, to utilize the Dynkin's formula, we let $f(x)=x^2$. Now, by Dynkin's, we have

$$\mathbb{E}[f(W_\tau)]=f(W_0)+\mathbb{E}\left[\int_0^{\tau}\frac{1}{2}f_{xx}(W_s)\,\mathrm{d}s\right]=\mathbb{E}\left[\int_0^{\tau}\mathrm{d}s\right]=\mathbb{E}[\tau].$$

Noting that at time τ, the possible values for W_τ are precisely $-a$ or a and thus $\mathbb{E}[f(W_\tau)]=\mathbb{E}[(W_\tau)^2]=a^2$. From this we find that

$$\mathbb{E}[\tau]=a^2.$$

[7] This will be proved soon in the material.

The example makes sense knowing that $\mathrm{Var}[W_t] = t$. The reader is also advised to tackle the similar problem 206 for practice.

In the above example, we assumed that the stopping time is finite. This assumption is also required in the formula, and things stop making sense fairly fast if the assumption is violated: see the following example.

Example 2.4. *(Dynkin's fails for infinite τ).* Let

$$\tau := \inf\{t \geq 0 : W_t = a\}$$

for some $a > 0$. It turns out that $\mathbb{E}[\tau] = \infty$. Let $f(x) = x$, and now by "Dynkin's" and knowing that $\mathbb{E}[W_\tau] = a$, we get

$$\mathbb{E}[W_\tau] = W_0 + \mathbb{E}\left[\int_0^\tau 0 \, \mathrm{d}t\right] = 0 \neq a = \mathbb{E}[W_\tau],$$

which is a contradiction!

We finish off by finally generalizing Feynman-Kac to the realm of stopping times.

Corollary 2.9. *(Feynman-Kac for a Stopping Time).* Let $D \subset \mathbb{R}^n$ and $\tau := \inf\{t \geq 0 : W_t \notin D\}$, such that $\mathbb{E}_x[\tau_D] < \infty$ for all initial values $x \in D$. Moreover, let $T > 0$ be some terminal time and $\tau = \tau_D \wedge T$, and functions r, Ψ, Φ, X_s and the PDE as in Theorem 2.5. Then

$$u(t,x) = \mathbb{E}_{t,x}\left[\exp\left\{-\int_t^\tau r(s, X_s)\,\mathrm{d}s\right\}\Phi(X_\tau) + \int_t^\tau \exp\left\{-\int_t^s r(s', X_{s'})\,\mathrm{d}s'\right\}\Psi(s, X_s)\,\mathrm{d}s\right].$$

Proof. Follows from the Feynman-Kac and Dynkin's formulas. □

So far, we have discussed the connection between PDEs and SDEs for some finite terminal times T. Next, we will explore so-called boundary-value problems with Brownian motions.

2.3 Brownian Motion revisited

Consider following discrete-time example.

Example 2.5. *(Discrete Random Walk).* Let

$$S_n := \sum_{i=1}^n X_i + x$$

be a discrete time stochastic process called a Random walk for some $S_0 = x$ and for i.i.d. random variables X_i with

$$\mathbb{P}(X_i = 1) = \mathbb{P}(X_i = -1) = \frac{1}{2}.$$

Moreover, let $\tau := \inf \{n \geq 1 : S_n \in \{0, N\}\}$ for some $N \in \mathbb{Z}$, and consider $M_n := S_{n \wedge \tau}$. We claim that M_n is a martingale.

Indeed, it is easy to see that by mimicking some of the martingale-proof strategies presented in Chapter1[8]:

$$\begin{aligned}\mathbb{E}\left[M_{n+1}|\mathcal{F}_n\right] &= \mathbb{E}\left[M_n + \mathbb{1}_{\tau>n}\left(M_{n+1} - M_n\right)|\mathcal{F}_n\right] \\ &= M_n + \mathbb{E}\left[S_{n+1} - S_n|\mathcal{F}_n\right] = M_n + E\left[X_{n+1}\right] = 0.\end{aligned}$$

Now, by Doob's Optional Sampling 2.7, we have that

$$x = M_0 = \mathbb{E}[M_n] = \mathbb{E}[S_{\tau \wedge n}].$$

Letting $n \to \infty$, we get

$$x = \mathbb{E}[S_\tau] = 0 \cdot \mathbb{P}(S_\tau = 0) + N \cdot \mathbb{P}(S_\tau = N) \Leftrightarrow \mathbb{P}(S_\tau = N) = \frac{x}{N}.$$

The above example could have been made with another martingale, for example with

$$\tilde{M}_n := S_{n \wedge \tau}^2 - n \wedge \tau.$$

Why are we discussing discrete time stochastic processes? Random walk is an interesting concept, and it is a staple in any research related discrete or combinatiorial probabilities[9]. However, from the perspective of this material, it is natural to consider if there is any connection between a Random walk and the Brownian motion? It is not surprising that there indeed is such a connection!

Proposition 2.10. *(Donsker).* Let $X_1, X_2, \ldots$ be a sequence of i.i.d. random variables with $\mathbb{E}[X_i] = 0$ and $\mathrm{Var}[X_i] = 1$, and let $S_n = \sum_{i=1}^n X_i$ be a random walk. Define

$$W^{(n)}(t) := \frac{S_{\lfloor nt \rfloor}}{\sqrt{n}} \quad \text{for } t \in [0, 1].$$

Then

$$W^{(n)}(t) \longrightarrow^d W_t$$

as $n \to \infty$.

Proof. Omitted. See discussion in Gut 2013 section 7.9. □

Very intuitively, the above result is sometimes also called the functional Central Limit Theorem.

Remark 2.11. Consider the above proposition.

[8] And by noting that we have so far defined martingales only for continuous-time stochastic processes, but the idea is the same: to show that $\mathbb{E}\left[M_{n+1}|\mathcal{F}_n\right] = \mathbb{E}\left[M_n\right]$ for some discrete time structure.

[9] For example, random walk is interestingly connected to the concept of random trees!

(i) If $t = 1$, the result reduces to the classical Central Limit theorem 1.18. Indeed, recall that for a standard Brownian motion, $W_1 \sim \mathrm{N}(0, 1)$.

(ii) A direct consequence of the result is that Brownian motion is a limit of a random walk.

(iii) The presented random walk is not the only process which has Brownian motion as its limit!

If Brownian motion can be a limit for several stochastic processes, how can we identify a process as a Brownian motion? In the world of elementary random variables, the answer would be easy as a random variable is uniquely determined by its distribution function. Luckily, a similar strategy works for Brownian motion. Following definition is a good beginning.

Definition 2.12. *(Gaussian Process).* Let X_t be a stochastic process. We say that it is Gaussian if for any choice of time periods $0 < t_1 < \cdots < t_n$ the random vector $(X_{t_1}, \ldots, X_{t_n})$ is a multivariate normal distribution (see definition 1.11).

The following result follows straightforwardly.

Proposition 2.13. The following claims are equivalent:

(i) X_t is a standard Brownian motion, and

(ii) X_t is a Gaussian process, which starts at 0, with continuous trajectories, and for all $s, t \in \mathbb{R}_+$,
$$\mathrm{Cov}\,(X_s, X_t) = s \wedge t.$$

The second item we actually proved already within proposition 1.22, and it is often called the covariance condition.

The proposition has many uses, and it can be used to solve many Brownian motion related problems and properties, for example that Brownian motion scales in a "zoom"-like[10] way, that is,
$$\tilde{W}_t = \frac{1}{c} W_{c^2 t}$$
is a standard Brownian motion, or that Brownian motion has a property of memorylessness:
$$\tilde{W}_t = W_{t+T} - W_T$$
is also a standard Brownian motion. These properties, coincidentally, are left as exercises 209 and 210.

[10] Zoom as in camera lense zooming in on something, not Zoom as the platform to hold digital lectures.

2.4 Boundary-Value, Dirichlet, and Poisson Problems

Now that we have more Brownian motion-related tools, let us get back to examining boundary -value problems, hitting times, and similar problems. Here we consider many results with an example-driven approach, and highlight the results, for example regarding exit times, from multiple standpoints. In particular, certain exit time problems are solved both with Dynkin's formula and as a so-called Poisson process.

Let us first return to example 2.3, and namely to the question of whether $\mathbb{E}[\tau] < \infty$ or not.

Proposition 2.14. *(Finiteness of* $\mathbb{E}[\tau]$*).* Let W_t be an $n-$dimensional Brownian motion which starts from some fixed $x \in \mathbb{R}^n$ with $|x| < R < \infty$. Let

$$D_R = \{x \in \mathbb{R}^n : ||x|| < R\},$$

and $\tau_R = \inf\{t \geq 0 : X_t \notin D_r\}$. Then $\mathbb{E}_x[\tau_R] < \infty$.

Proof. Let $f(x) = ||x||^2 = \sum_{i=1}^n x_i^2$ and let $\tau = \tau_R \wedge k$ for some arbitrary k. Note that surely $\mathbb{E}[\tau] < \infty$, and thus by Dynkin we have

$$\mathbb{E}_x [f(W_t)] = ||x||^2 + \mathbb{E}_x \left[\int_0^\tau n \, \mathrm{d}t\right] = ||x||^2 + n\mathbb{E}_x [\tau].$$

By noting that $\tau \leq \tau_r$, we get

$$\mathbb{E}_x [f(W_t)] \leq \mathbb{E}_x [f(W_{\tau_R})] = \mathbb{E}_x \left[||W_{\tau_R}||^2\right] = R^2.$$

Combining these, we get

$$||x||^2 + n\mathbb{E}_x[\tau] \leq R^2,$$

and finally

$$\mathbb{E}_x[\tau] = \mathbb{E}_x [\tau_R \wedge k] \leq \frac{R^2 - ||x||^2}{n}.$$

Now by Dominated Convergence (see 1.4) and Fubini's Theorem,

$$\lim_{k\to\infty} \mathbb{E}_x [\tau_R \wedge k] = \mathbb{E}_x \left[\lim_{k\to\infty} \tau_R \wedge k\right] = \mathbb{E}_x [\tau_R] < \infty.$$

□

That is, under the assumptions of the previous proposition, we can apply Dynkin's theorem directly with τ_R for any bounded domain R. Let's try our newfound power out!

Example 2.6. *(Exit time from an interval).* Let $W_t \in \mathbb{R}$ and $\tau = \inf\{t \geq 0 : W_t \notin (-a, b)\}$ for some $a, b > 0$. Let us find $\mathbb{E}[\tau]$.

Let $f(x) = x$. Now by Dynkin or by OST, we have

$$0 = \mathbb{E}[W_0] = \mathbb{E}[W_\tau] = -a\mathbb{P}(W_\tau = -a) + b\mathbb{P}(W_\tau = b) = -a\mathbb{P}(W_\tau = -a) + b\left(1 - \mathbb{P}(W_\tau = b)\right),$$

from which we get $\mathbb{P}(W_\tau = -a) = \frac{b}{a+b}$. This isn't sufficient, we need another iteration with $g(x) = x^2$:

$$0 = \mathbb{E}[W_\tau^2] = (-a)^2\mathbb{P}(W_\tau = -a) + b^2\mathbb{P}(W_\tau = b) = \frac{a^2 b}{a+b} + \frac{b^2 a}{a+b} = ab.$$

This indeed implies that $\mathbb{E}[\tau] = \mathbb{E}[W_\tau^2] = ab < \infty$.

Let us move now back to PDE's. Earlier, we considered Itô diffusions inside a domain D and denoted stopping times

$$\tau_D := \inf\{t \geq 0 : X_t \notin D\} \quad \text{and} \quad \tau := \tau_D \wedge T$$

for some $T > 0$. For repetition, by letting $r = 0$, Feynman-Kac for stopping times 2.9 states now that the function

$$u(t,x) = \mathbb{E}_{t,x}\left[\Phi(X_\tau) + \int_t^\tau \Psi(X_s)\right]$$

solves

$$\begin{cases} u_t + \mathcal{L}u + \Psi = 0 & \text{in } [0,T) \times D \\ u(T,x) = \Phi(x) & \text{in } \left([0,T) \times \partial D\right) \cup \left(\{T\} \times D\right). \end{cases}$$

If we now let $T \to \infty$ and define

$$\tau := \lim_{T \to \infty} \tau_D \wedge T = \tau_D,$$

the problem becomes time-homogeneous by the properties of Itô diffusion, and $u_t = 0$! The following result follows, which is the time-homogeneous form of the Feynman-Kac but known under a different name.

Proposition 2.15. *(Dirichlet-Poisson Problem).* Assume $D \subset \mathbb{R}^n$, $X_0 = x \in D$, and $\mathbb{E}[\tau] < \infty$. Then

$$u(t,x) = \mathbb{E}_x\left[\Phi(X_\tau) + \int_t^\tau \Psi(X_s)\right]$$

solves uniquely

$$\begin{cases} \mathcal{L}u + \Psi = 0 & \text{in } D \\ u(x) = \Phi(x) & \text{on } \partial D. \end{cases}$$

Proof. Omitted, very similar to the regular Feynman-Kac. □

This result provides a solid foundation for applications in Boundary-Value problems.

Corollary 2.16. *(Poisson Problem)* Let $u(x) = \mathbb{E}_x[\tau]$, where $\tau := \inf\{t \geq 0 : X_t \notin D\}$. Let $\Psi = 1$ and $\Phi = 0$. Then, clearly, $u(x) = \mathbb{E}_x[0 + \int_0^\tau 1 \, \mathrm{d}t]$ solves for

$$\begin{cases} \mathcal{L}u + 1 = 0 & \text{in } D \\ u = 0 & \text{on } \partial D. \end{cases}$$

The Poisson problem formulation can be used neatly to revisit some of the previous problems, consider for example 2.6.

Example 2.7. *(Exit time from an interval revisited).* Let $u(x) = \mathbb{E}_x[\tau_{(-a,b}]$ for some $a, b > 0$. Then

$$\begin{cases} \frac{1}{2}\Delta u + 1 = 0 & \text{in } (-a, b) \\ u = 0 & \text{on } \partial(-a, b) = \{-a, b\}. \end{cases}$$

We note that $u = 0$ on the boundary, an indeed, such condition would be satisfied when the function would be of the form $(x + a)(x - b)$. Based on this, let us make a guess $u = C(x + a)(x - b)$ for some constant C. By plugging this partial solution on the boundary to the PDE inside the interval, we gain

$$\frac{1}{2}\Delta u + 1 = \frac{1}{2}\Delta C(x + a)(x - b) + 1 = C + 1 = 0,$$

and thus $C = -1$, and the particular solution is

$$u(x) = (x + a)(b - x).$$

Plugging in $x = 0$, which is equivalent tosetting the starting point of the process as 0, leads to $u(0) = \mathbb{E}_0[\tau_{(-a,b)}] = ab$, which is precisely the same result as in example 2.6!

Of course, these examples rely only on "guesses" or ansatzes, and for to properly argumentate the result, we would need to verify the result. More on that later.

For another useful example, let us revisit the proposition 2.14 regarding finiteness of a more general stopping time.

Example 2.8. *(Finiteness of* $\mathbb{E}[\tau]$ *revisited).* To reiterate, we are looking at a $n-$dimensional Brownian motion W_t which starts from some fixed $x \in \mathbb{R}^n$ with $||x|| < R < \infty$. In addition,

$$D_R = \{x \in \mathbb{R}^n : ||x|| < R\},$$

and $\tau_R = \inf\{t \geq 0 : X_t \notin D_r\}$. As in the previous example, let

$$u(x) = \mathbb{E}_x[\tau_R],$$

which solves

$$\begin{cases} \frac{1}{2}\Delta u + 1 = 0 & \text{in } D_R \\ u = 0 & \text{on } \partial D_R. \end{cases}$$

Very similar to the previous example, we notice that the boundary condition must satisfy $||x||^2 - R^2 = 0$ on ∂D_R. We are thus looking for u of the form $u(x) = C(||x||^2 - R^2)$ for some constant C. It is easy to verify that $\Delta||x||^2 = 2n$, and so by plugging in the guess into our original PDE inside the domain we get

$$\frac{1}{2}\Delta u + 1 = \frac{C}{2}2n + 1 = 0,$$

from which $C = -\frac{1}{n}$ follows. Finally, from this we get

$$u(x) = \frac{R^2 - ||x||^2}{n},$$

which is equivalent to our solution in the proposition 2.14.

Similarly to the Poisson problem, the following result is immediate from the time-homogeneous Feynman-Kac, after noting that

$$E_x\left[\mathbb{1}_{X_\tau \in H}\right] = E_x\left[\mathbb{1}_{X_\tau \in H} + \int_0^\tau 0\,\mathrm{d}t\right].$$

Corollary 2.17. *(Dirichlet Problem).* Let $H \in \partial D$ for some domain D, $\Psi = 0$,

$$\Phi = \begin{cases} 1 & \text{if } x \in H, \\ 0 & \text{if } x \in \partial D \setminus H, \end{cases}$$

and let

$$u(x) = \mathbb{P}_x\left(X_\tau \in H\right) = E_x\left[\mathbb{1}_{X_\tau \in H}\right].$$

Then u solves

$$\begin{cases} \mathcal{L}u = 0 & \text{in } D, \\ u = 0 & \text{in } \partial D \setminus H, \\ u = 1 & \text{in } H. \end{cases}$$

That is, in the Dirichlet problem there are two separate boundary conditions. Let us visit a concrete example in the guise of old friends 2.6 and2.7.

Example 2.9. Consider the set-up of a problem 2.6, but instead let

$$u(x) = \mathbb{P}_x(X_\tau = -a)$$

be the probability of X_t hitting $-a$ before b. Then u solves

$$\begin{cases} \frac{1}{2}\Delta u = 0 & \text{in } (-a, b), \\ u(-a) = 1, \\ u(b) = 0. \end{cases}$$

Similarly to the earlier examples, we now see that as $u = 0$ on the boundary, it should have a component $b - x$. Building on this with respect to the boundary condition at $-a$, we notice that if

$$u(x) = \frac{b - x}{a + b},$$

both of the boundary conditions are satisfied. In fact, this is already the particular solution as the in-domain PDE is homogeneous. Thus, if the initial point is 0, we arrive at

$$\mathbb{P}(X_\tau = -a) = \frac{b}{a + b}.$$

We close this chapter by examining an important example, which serves as a bridge between the next topic, heat equations, and our current boundary-value problems, even though the next notions make next to no sense if one has not studied harmonic analysis. Consider the Laplace equation

$$\Delta u = 0.$$

Formulating this as a Dirichlet problem, we gain

$$\begin{cases} \Delta u = 0 & \text{in } D, \\ u = f & \text{on } \partial D \end{cases}$$

for some domain D and function f. We call the class of functions satisfying the Laplace equation the class of harmonic functions, if the function is twice continuously differentiable.

Now, the Laplace equation inside the domain extends f from ∂D – this is called the harmonic extension of f. Moreover, for readers with knowledge in Harmonic analysis and PDE's it's clear that if $u \in C^2(D)$, then the Laplace function is equivalent with f having the so-called mean-value property. Moreover, Maximum principle of harmonic functions ensures that u is uniquely determined by f.

Now, let us move to exercises and afterwards to the Heat equation.

2.5 Exercises for Chapter 2

In the exercises, a sign (*) before the problem number denotes a slightly more involved exercise.

Feynman-Kac and Dynkin's Formula

201. Find the infinitesimal generator for the following Itô processes:

    (a) $\mathrm{d}X_t = \beta X_t \,\mathrm{d}t + \sigma \,\mathrm{d}W_t$,

    (b) $\mathrm{d}X_t = \mu X_t \,\mathrm{d}t + \sigma X_t \,\mathrm{d}W_t$,

(c) the n-dimensional Brownian motion,

(d) the $(n+1)$-dimensional stochastic process (t, W_t), where $W_t \in \mathbb{R}^n$.

202. Find an Itô process whose generator is the following:

(a) $\mathcal{L}f(x) = f_x(x) + f_{xx}(x)$,

(b) $\mathcal{L}f(x) = rf_x(x) + \frac{1}{2}\alpha x^2 f_{xx}(x)$.

203. Prove the 1-dimensional general Feynman-Kac theorem when $D = \mathbb{R}$, and $r > 0$ is a constant. *(Hint: apply Itô's formula on $Y_s = e^{-r(s-t)}u(s, X_s)$).*

204. Use Feynman-Kac to solve the PDE with terminal condition:

$$\begin{cases} u_t + \frac{1}{2}u_{xx} = 0, \\ u(x,T) = x^4. \end{cases}$$

205. Use Feynman-Kac to solve the PDE with initial condition:

$$\begin{cases} u_t - \mu x u_x - \frac{1}{2}\sigma^2 u_{xx} = 0, \\ u(0,x) = \Phi(x). \end{cases}$$

(Hint: use $v(t,x) = u(T-t,x)$).

206. Let $\tau = \inf\{t > 0 : W_t \notin (-a,b)\}$, where $a, b > 0$. Use Dynkin's formula to determine $\mathbb{E}[\tau]$.

207. Let $J = \int_0^1 t \, \mathrm{d}dW_t$. Use Dynkin's formula to find the moment generating function $m\,(u) = \mathbb{E}[e^{uJ}]$, and show that $J \sim N(0, \frac{1}{3})$. *(Hint: let $f(x) = e^{ux}$).*

208. **(The Ornstein-Uhlenbeck Process)**. For a given standard Brownian motion W on some probability space $(\Omega, \mathcal{F}, \mathbb{P})$, consider the Ornstein-Uhlenbeck process which solves the following SDE:

$$\mathrm{d}X_t = \mu X_t \,\mathrm{d}t + \sigma \,\mathrm{d}W_t, \quad X_0 = x, \tag{2.1}$$

where $\mu, \sigma \in \mathbb{R}$.

(a) Show that equation (2.1) admits a unique strong solution.

(b) Use Itô's formula to find the solution to the above equation *(Hint: consider the process $e^{-\mu t}X_t$).*

(c) Fix $T \geq 0$, find $\mathbb{E}[X_T]$ and $\mathrm{Var}[X_T]$.

(d) We wish now to compute the characteristic function ϕ of X_T:

$$\phi_{X_T}(\xi) := \mathbb{E}[e^{i\xi X_T}|X_0 = x], \text{ for } \xi \in \mathbb{R}.$$

Fix $\xi \in \mathbb{R}$, use the Feynman-Kac theorem to show that the function $u : [0,T] \times \mathbb{R} \to \mathbb{R}$ defined by $u(t,x) := \mathbb{E}[e^{i\xi X_T}|X_t = x]$ satisfy the following PDE:

$$u_t + \mu x u_x + \frac{1}{2}\sigma^2 u_{xx} = 0, \text{ for all } (t,x) \in [0,T) \times \mathbb{R}.$$

Determine the terminal condition $u(T,x)$.

Exit Times, more on Brownian motion

209. Show that the process $\tilde{W}_t = W_{t+T} - W_T$ is a standard Brownian motion.

210. **(Brownian scaling)**. Let W_t be a 1-dimensional standard Brownian motion and let $c > 0$ be a constant. Show that

$$\tilde{W}_t := \frac{1}{c} W_{c^2 t}$$

is also a standard Brownian motion.

211. Let $X_t = \mu dt + \sigma dW_t$ be a be a process with a drift, driven by a Brownian motion, where $\mu, \sigma \neq 0$, Let $X_0 = x \in (a,b)$. Use Dynkin's formula to show that the expected exit time of X_t from the interval (a,b) is finite.

212. Let X_t be a geometric Brownian motion with

$$dX_t = \mu X_t dt + \sigma X_t dW_t, \quad X_0 = x \in (a,b)$$

Convince yourself that the expected exit time of X_t from the interval (a,b) is finite. *(Hint: Distinguish the case where $\mu - \frac{1}{2}\sigma^2 = 0$ and $\mu - \frac{1}{2}\sigma^2 \neq 0$).*

213. Characterise all harmonic functions $u : \mathbb{R}^n \setminus \{0\} \to \mathbb{R}^+$ of the form $u(x) = h(||x||)$.

214. Suppose that $r > 0, a_j > 0, b_j, c_j \in \mathbb{R}$ for $j = 1, \ldots, n$. Assume that $\sum_{j=1}^n c_j < r$. Consider the set

$$D = \{(x_1, \ldots, x_n) \in \mathbb{R}^n : \sum_{j=1}^n (a_j x_j^2 + b_j x_j + c_j) < r\}.$$

Let W_t be an n-dimensional Brownian motion starting at $(0, \ldots, 0) \in \mathbb{R}^n$. Determine the expected exit time of W_t from D.

215. Let $W_t \in \mathbb{R}^2$ with $W_0 = x$ where $r < ||x|| < R$, i.e. x is in the two-dimensional annulus

$$A(r, R) := \left\{ x \in \mathbb{R}^d : r \leq ||x|| \leq R \right\}$$

for $d = 2$. Determine the probability that W_t hits r before R (or, the probability that W_t exits the annulus from r). *(Hint: to find explicit solutions, it may be useful to consider only harmonic functions u, for which $\Delta u = 0$ and assume the spherical symmetry of the solution, i.e. that for some solution u there exists a function $\psi : [r, R] \to \mathbb{R}$ so that $u(x) = \psi\left(||x||^2\right)$).*

Chapter 3

The Heat Equation

3.1 The Heat Equation

We begin straightforwardly[1].

Definition 3.1. *(The Heat Equation).* Let $x = (x_1, \dots, x_n) \in \mathbb{R}^n$, $D \subset \mathbb{R}^n$ open and bounded, and $T \in (0, \infty)$. Moreover, let $D_T = (0, T] \times D$ be the parabolic cylinder and $u \in C^{1,2}(D_T)$. Then the equation

$$\frac{\partial u}{\partial t}(t, x) - \Delta u(t, x) = 0,$$

where

$$\Delta = \sum_{i=1}^{n} \frac{\partial^2}{\partial x_i^2}$$

is the so-called Laplace operator, is called the Heat Equation.

That is, we mostly consider the heat equation on a bounded domain, more specifically inside the cylinder D_T. Most often, we define the initial condition on D. Moreover, for later use, let us denote the so-called parabolic boundary of D_T as

$$\Gamma_T = \overline{D}_T \setminus D_T,$$

where $\overline{D}_T$ denotes the union of the set D_T and its boundary[2]. Moreover, when we talk about boundary values, we understand a boundary value of $u(t, x)$ as a limit

$$\lim_{(s,y) \to (t,x)} u(s, y),$$

where the points $(s, y) \in (0, t) \times D_T$ and $(t, x) \in (s, T) \times \Gamma_T$.

[1] This chapter scratches briefly the theory for PDEs, namely the Heat Equation. For a deeper approach, one should refer to, for example, Evans 2010, namely section 2.3.

[2] Note that since D_T is open, $\partial D_T \not\subset D_T$.

In the study of PDE's, one often wishes to find solutions to the equation, which exists and are unique. We are no different, and in the following passages we seek for such solutions.

The following theorem is extremely useful when trying to understand the Heat equation and its solution.

Theorem 3.2. *(Maximum Principle for the Heat Equation).* Assume that $u \in C^{1,2}(D_T)$ is a solution of the heat equation in D_T, and that the solution is additionally attained at the boundary so that $u \in C^{1,2}(D_T) \cap C(\overline{D}_T)$. Then

(i) *(Weak Maximum Principle).*

$$\max_{\overline{D}_T} u = \max_{\Gamma_T} u,$$

and

(ii) *(Strong Maximum Principle)* if D is connected and there exists some $(t_0, x_0) \in D_T$ such that

$$u(t_0, x_0) = \max_{\overline{D}_T} u,$$

then u is constant in $\overline{D}_{t_0}$

Proof. Omitted. See the proof and discussion in Evans 2010, section 2.3.3. □

Remark 3.3. The statement of the theorem holds even if you replace "max" operator with "min" in all of the claims.

Essentially, the theorem tells that the maximum for the heat equation is attained at its boundary, and that the solution will be constant on a specific time interval until t_0, given that the initial and boundary conditions don't change.[3]

Direct consequence of the Maximum principle is the uniqueness of the solution. This is very handy, considering that uniqueness of a solution is one of the first-sought-after matters in the study of PDEs.

Theorem 3.4. *(Uniqueness of the solution to the Heat Equation on bounded domains).* Let $\Phi \in C(\Gamma_T)$, $\Psi \in C(D_T)$, and assume that $u \in C^{1,2}(D_T) \cap C(\overline{D}_T)$ solves

$$(*) \begin{cases} u_t - \Delta u = \Psi & \text{in } D_T, \\ u = \Phi & \text{on } \Gamma_T. \end{cases}$$

Then u is unique.

[3]It is also to be noted that the maximum principle property does not hold for general u, and instead it is special to the heat equation and more generally to the class of so-called harmonic functions.

Proof. Assume that u_1, u_2 both solve $(*)$. Then $w := \pm(u_1 - u_2)$ solves

$$\begin{cases} w_t - \Delta w = \Psi & \text{in } D_T, \\ w = \Phi & \text{on } \Gamma_T. \end{cases}$$

Then, by the Weak maximum principle

$$\max(u_1 - u_2) = \min(u_1 - u_2) = 0,$$

from which the result follows. □

The assumption of boundedness can be relaxed by limiting the growth of u.

Theorem 3.5. *(Cauchy Initial Value Problem (IVP)).* Let Ψ, Φ be continuous, and let u solve

$$\begin{cases} u_t - \Delta u = \Psi & x \in \mathbb{R}^n, \\ u(0, x) = \Phi(x). \end{cases}$$

Moreover, let $|u(t, x)| \leq Ae^{a|x|^2}$ for some $A, a > 0$. Then u is unique.

Proof. Identically as above with the maximum principle. Technically we also would have required to state the maximum principle for the Cauchy problem separately, which follow's from the original maximum principle but involves slightly more involved details, see e.g. Evans 2010 Section 2.3. theorem 6. □

The growth restriction is necessary for us in order for u to have a unique solution. If, in the above theorem, we instead let $\Phi = \Psi = 0$. Then the equation would have infinitely many solutions, which would be both annoying and impractical.

No matter, we have now shown that by the Maximum principle for the Heat equation, we have unique solutions. We still haven't confirmed that the solutions exist! For the sake of convenience and applications, we restrict our search to the Cauchy initial value problem.

At the very end of Chapter 2, we characterized the class of so-called harmonic functions. These turn out to be related to the fundamental solution of the Laplace equation $\Delta u = 0$. More generally, we now follow the suit and "guess" a solution for the heat equation which has the properties we desire.

Our guess begins with letting $u(t, x)$ solve the homogeneous Cauchy initial value problem, i.e. $u_t - \Delta u = 0$ with some initial condition. We know that a suitable solution should be scaled, which hints that our guess solution can be scaled so that $u(ax, a^2t)$ solves the Cauchy initial value problem for some constant a. Moreover, a good solution is scale invariant, which suggest a form of $u(t, x) = v(\frac{x}{t})$ for some v. In addition, nearing infinity, we wish to attain $u_x \to 0$ as $x \to \pm\infty$. One should also observe that

$$\frac{\partial}{\partial t} \int_{\mathbb{R}} u(t, x) \, \mathrm{d}x = 0$$

implies

$$\int_{\mathbb{R}} u(t,x)\,\mathrm{d}x = C$$

for some constant C. On the other hand, by change of variables we have

$$\int_R v\left(\frac{x}{t}\right)\,\mathrm{d}x = \sqrt{t}\int_{\mathbb{R}} v(y)\,\mathrm{d}y.$$

That is, the scale invariant solution v should be expanded so that $u(t,x) = \frac{1}{\sqrt{t}}v(\frac{x}{\sqrt{t}})$ for some v. This is starting to feel familiar? In any case, we are ready to define the fundamental solution to the Heat equation. (Unfortunately, the solution to the Cauchy initial value problem has to wait, but not for long).

Definition 3.6. *(Fundamental Solution to the Heat Equation, the Heat Kernel).* Let

$$\phi(x) = \frac{1}{\sqrt{2\pi}}e^{-\frac{1}{2}x^2}$$

for $t > 0, x \in R$. Then

$$g(t,x) := \frac{1}{\sqrt{2t}}\phi(\frac{x}{\sqrt{2t}})$$

is called the Heat kernel, or the fundamental solution to the (1-dimensional) Heat Equation.

Indeed, g is a solution to the Heat equation!

Proposition 3.7. Let $g(t,x)$ be as above in 3.6. Then

$$g_t - \Delta g = 0.$$

Proof. Left as an Exercise 302. ☐

A few nice properties follow.

Proposition 3.8. *(Properties of $g(t,x)$).* Let $g(t,x)$ be as above in 3.6. Then

(i) If $x \neq 0$, then $\lim_{t\to 0} g(t,x) = 0$,

(ii) If $x = 0$, then $\lim_{t\to 0} g(t,x) = \infty$,

(iii) $\int_{\mathbb{R}} g(t,x)\,\mathrm{d}x = 1$ for all $t > 0$, and

(iv) $g(t,x)$ is smooth, i.e. $g \in C^\infty$.

Proof. Properties (i)–(iii) are left as exercises 303 and 304. The proof for smoothness is slightly more involved, and an exquisite proof can be found in Evans 2010 section 2.3., Theorem 8. ☐

It is not surprising that the solution concept extends also to $\mathbb{R}^n$.

Definition 3.9. *(Heat Kernel in $\mathbb{R}^n$).* Let

$$\phi(x) = \frac{1}{(2\pi)^{n/2}} e^{-\frac{||x||^2}{2}}$$

for $x \in \mathbb{R}^n$. Then

$$g(t,x) = \frac{1}{(\sqrt{2t})^n} \phi(\frac{x}{\sqrt{2t}}) = \frac{1}{(4\pi t)^{n/2}} e^{-\frac{||x||^2}{4t}}$$

is called the Heat kernel, or the fundamental solution to the Heat equation in $\mathbb{R}^n$.

In order to progress to the solution of Cauchy initial value problem, which is transformed from the fundamental solution g, we require small bits of measure-theoretical machinery. The main reason for this is the fact that $\lim_{t\to 0} g(t,x)$ is not a well-behaved function near 0.

Definition 3.10. *(Dirac delta measure).* Let f be a continuous function with compact support. Then we call a singular measure $\delta(x)$ the Dirac delta if it satisfies[4]

$$\int_{\mathbb{R}} f(x)\delta(x)\,\mathrm{d}x = f(0).$$

Alternatively, the dirac delta δ can be characterized as a probability measure on $(\Omega, \mathcal{F})$, for which the distribution function can be characterized as[5]

$$H(x) = \begin{cases} 1, & \text{if } x \geq 0 \\ 0, & \text{if } x < 0. \end{cases}$$

We included the two definitions for clarity. Simply, the dirac delta is a measure or a function which has all of its mass centered at $x = 0$. Moreover, one can think that for some test function f: when convoluted with the dirac delta, the dirac delta maps the test function's value to their value at some x:

$$\int_{\mathbb{R}} \delta(x-y) f(y)\,\mathrm{d}y = f(x).$$

With this notion, we can make sense of the fundamental solution near 0, which is the initial value boundary we define for the Cauchy problem. That is, we say that g solves the Cauchy initial value problem

$$\begin{cases} g_t - \Delta g = 0 \\ \lim_{t\to 0} g(t,x) = \delta(x). \end{cases}$$

[4]Even though this integral is written as the standard Lebesgue integral, δ is not absolutely continuous with respect to the Lebesgue measure, and thus no "true" function satisfies this property.

[5]The reader should note that a proper Riemann-Stjieltjes can be constructed with respect to H:

$$\int_{\mathbb{R}} f(x)\delta(\,\mathrm{d}x) = \int_{\mathbb{R}} f(x)\,\mathrm{d}H(x).$$

Before we move on to characterizing the solution, we define the following operator, which proves to be useful[6].

Definition 3.11. *(Convolution operator).* Let f, g be integrable functions in $\mathbb{R}$. Their convolution is the integral operator $*$ for which

$$(*)\,(t) = \int_{\mathbb{R}} f(s)g(t-s)\,\mathrm{d}s.$$

Convolution is a handy operator. It essentially retains the smoothness of the functions, as it averages the values of f around some time point t with respect to g. It has some other properties, such as commutativity (check exercise 301).

Now, we construct the solution to the Cauchy problem from the fundamental solution of the HE in the following steps. We know that the solution is constant through $x \mapsto x-y$, and in addition for any bounded and continuous Φ, the heat equation is also solved by $g(t, x-y)\Phi(y)$. Recalling that an integral is merely a limit of linear combinations, we see that the solution is also retained through convolution. Finally, we can show the solution to the Cauchy initial value problem.

Theorem 3.12. *(Solution to the Cauchy Initial Value Problem).* Consider the problem introduced in 3.5. Let

$$u(t,x) := g(t,x) * \Phi(x) = \int_{\mathbb{R}} g(t, x-y)\Phi(y)\,\mathrm{d}y,$$

where g is the fundamental solution to the Heat equation and Φ is a bounded and continuous function. Then $u(t,x) \in C^{\infty}\left((0,\infty) \times \mathbb{R}\right)$, and it solves the homogeneous Cauhcy initial value problem

$$\begin{cases} u_t - \Delta u = 0 & x \in \mathbb{R}, \\ u(0,x) = \Phi(x). \end{cases}$$

Proof. The elementary direction is left as the exercise 306. The full proof is found, again, in Evans 2010 Section 2.3. Theorem 1. ☐

Similarly to the solution of the heat equation, the solution to the Cauchy initial value problem has some nice and welcomed properties.

Proposition 3.13. *(Properties of the solution to the Cauchy Problem).* Let u be a solution to

$$(*) \quad u_t(t,x) - u_{xx}(t,x) = 0,$$

where

$$0 < t < T, \quad T \in (0,\infty], \quad a < x < b, \quad a \in [-\infty, \infty), \quad b \in (-\infty, \infty].$$

Then the following hold:

[6] And the writer is surprised we haven't done this yet!

(i) If u, v both solve $(*)$, and $\alpha, \beta \in \mathbb{R}$, then $\alpha u+\beta v$ is also a solution to $(*)$ *(the solution is linear)*,

(ii) If u solves $(*)$, and $\alpha > 0, x_0 \in \mathbb{R}$, then $u(\alpha^2 t, \alpha x - x_0)$ solves $(*)$ for $t \in (0, \alpha^2 T)$ and $x \in (\alpha a + x_0, \alpha b + x_0)$ *(the solution retains under shift and scale)*,

(iii) If $u \in C^3$ and u solves $(*)$, then its derivatives u_t, u_x also solve $(*)$ *(the solution has a differential property)*,

(iv) If u solves $(*)$ and

$$v(t,x) = \int_a^x u(t,z)dz, x \in (a,b),$$

then v also solves $(*)$ given that

$$\lim_{z \to a} u_x(t,z) = 0$$

for each $t \in (0,T)$ *(the solution retains through integration)*, and

(v) If u solves $(*)$ and $f : \mathbb{R} \to \mathbb{R}$ is a function, then $(u * f)(t,x)$ also solves $(*)$ *(the solution retains through convolution)*.

Proof. Left as exercises 307–311. □

In this section we discussed the heat equation and the Cauchy equation. The reader is advised to be vigilant, as in many times we referred to the heat equation when we were discussing a problem constrained to a bounded domain. On the other hand, the word heat equation can just refer to the equation $u_t - \Delta u = 0$, which is also the main driver in Cauchy problems, which we specifically separated from the concept of HE's by generalizing the problem to $\mathbb{R}$ (with some additional constraints). However, above all else: Cauchy problems are also fundamentally problems of the heat equation, and thus some confusion is unpreventable.

3.2 Applications of the Cauchy Problem

Before we venture to examples and applications of the Cauchy initial value problem, we make our notational affairs slightly easier by letting I_0 denote the family of piecewise continuous functions, which growth is limited to at most exponential when $x \to \infty$.

For most parts, this subsection covers useful examples. The results, if not motivated, follow from the previous section and the theory of heat equations.

Example 3.1. Let u solve

$$(*) \begin{cases} u_t - \Delta u = 0, \quad x \in \mathbb{R}, \quad t > 0 \\ u(0,x) = \varphi(x) \end{cases}$$

for some $\varphi(x) \in I_0$. Then

$$u(t,x) = g * \varphi(x)$$

is its unique solution of $(*)$.

Example 3.2. *(Heat Equation with rate of diffusion k).* Let u solve

$$(*) \begin{cases} u_t - ku_{xx} = 0, \quad x \in \mathbb{R}, \quad t > 0, \quad k > 0 \\ u(0,x) = \delta(x). \end{cases}$$

Then

$$u(t,x) = g(t,x) = \frac{1}{\sqrt{2kt}} \phi\left(\frac{x}{\sqrt{2kt}}\right) = \frac{1}{2\sqrt{\pi kt}} e^{-\frac{x^2}{\sqrt{4kt}}}$$

is the fundamental solution of $(*)$.

We point out that if $k = \frac{1}{2}$, then this is equivalent to the so-called Kolmogorov Equation for the Brownian motion! We will return to this concept later. Before the next problem, we introduce yet another property for functions.

Definition 3.14. *(Odd and Even functions).* We say that a function $f : \mathbb{R} \to \mathbb{R}$ is odd relative to x if

$$f(x-y) = -f(x+y) \quad \text{for all } y \geq 0.$$

Similarly, we say that a function f is even relative to x if

$$f(x-y) = f(x+y) \quad \text{for all } y \geq 0.$$

For a related exercise to familiarize yourself with the concept, see exercise 312. The concept of odd and even functions is useful when deriving the next example.

Example 3.3. *(The Quarter-Plane Problem).* We review the so-called quarter-plane problem. Let u solve

$$\begin{cases} u_t - \Delta u = 0, \quad t > 0, x > 0 \\ u(t,0) = 0, \quad t > 0 \\ u(0,x) = \varphi, \quad x > 0, \varphi \in I_0. \end{cases}$$

Then

$$u(t,x) = \int_0^\infty \left(g(t,x-y) - g(t,x+y)\right) \varphi(y)\,\mathrm{d}y.$$

The method to tackle this problem is simple. We mirror φ to an odd function with respect to 0 in the "x sense", and then the solution comes similarly as in example 3.1.

Let

$$\tilde{\varphi}(x) = \begin{cases} \varphi(x), \quad x > 0 \\ -\varphi(-x), \quad x < 0, \end{cases}$$

which is an odd function by definition. Now by the properties of an odd function,

$$
\begin{aligned}
u(t,x) =& g(t,x) * \tilde{\varphi}(x) = \int_{\mathbb{R}} g(t,x-y)\tilde{\varphi}(y)\,\mathrm{d}y \\
&= \int_0^\infty g(t,x-y)\varphi(y)\,\mathrm{d}y - \int_{-\infty}^0 g(t,x-y)\varphi(-y)\,\mathrm{d}y \\
&= \int_0^\infty g(t,x-y)\varphi(y)\,\mathrm{d}y - \int_0^\infty g(t,x+z)\varphi(z)\,\mathrm{d}z \\
&= \int_0^\infty \left(g(t,x-y) - g(t,x+y)\right)\varphi(y)\,\mathrm{d}y.
\end{aligned}
$$

It is then easy to check that u indeed satisfies the PDE in the quarter-plane problem. In particular, since g is an even function and $\tilde{\varphi}$ is odd function, their convolution is odd with respect to 0:

$$
u(t,0) = \int_{\mathbb{R}} g(t,0-y)\tilde{\varphi}(y)\,\mathrm{d}y = 0.
$$

Few remarks might be in place regarding this example. First, the parity of $\tilde{\varphi}$ is transferred to u (which makes u an odd function). Secondly, the quarter-plane problem is often applied to Barrier option pricing. Thirdly, the following definition might be useful in life, the following examples, and for the exercise 314.

Definition 3.15. *(Green's Function for the quarter-plane problem).* Let g be the fundamental solution to the Heat equation. Then we call

$$
G(t,x,y) = g(t,x-y) - g(t,x+y)
$$

the Green's function for the quarter-plane problem.

Example 3.4. *(Time-varying Boundary problem).* Let w solve

$$
\begin{cases}
w_t - w_{xx} = 0, \quad t,x > 0 \\
w(t,0) = f(t) \quad x > 0 \\
w(0,x) = 0 \quad t > 0
\end{cases}
$$

for $f(t) \in I_0$. Then

$$
w(t,x) = \int_0^t \frac{\partial G}{\partial y}(t-s,x,0) f(s)\,\mathrm{d}s.
$$

Example 3.5. *(Initial value formulation for the quarter-plane problem).* Let v solve

$$
\begin{cases}
v_t - v_{xx} = 0, \quad t,x > 0 \\
v(t,0) = f(t) \quad x > 0 \\
v(0,x) = \varphi(x) \quad t > 0
\end{cases}
$$

for $f, \varphi \in I_0$ and $f(0) = \varphi(0)$. Then

$$v = u + w,$$

where u is the solution to the quarter-plane problem 3.3 and w is the solution to the time-varying boundary problem 3.4.

Example 3.6. *(Infinite Propagation Speed).* Consider the Cauchy problem introduced in 3.5 and 3.12. Let $\Phi \geq 0$ so that its positive at least somewhere. Then

$$u(t, x) = \int_{\mathbb{R}} g(t, x - y)\Phi(y)\,\mathrm{d}y > 0$$

for all $x \in \mathbb{R}, t > 0$.

The proof of the result is left as the exercise 315.

We now continue with similar approach, but instead restrict the family of examples to those of mathematical finance.

3.3 Financial Applications

In this section, we shortly review core examples of the Cauchy problem with applications in Finance. This section is targeted towards readers with already some knowledge in the pricing of financial derivatives, but we try to contain sources in sections where we omit necessary details, definitions, theorems, and assumptions.

The following example is technically an example, but for its importance and for the related result we call it a theorem instead.

Theorem 3.16. *(Black-Scholes PDE and Risk-Neutral Valuation).* Let S_t be a Geometric Brownian motion[7] such that

$$\mathrm{d}S_t = rS_t\,\mathrm{d}t + \sigma S_t\,\mathrm{d}W_t,$$

where W_t is a Brownian motion[8], r is the known risk-free rate, and σ is also known. Moreover, for some financial contract with a payoff function Φ, let its price process be Π_t. Then, if

$$\Pi_t = F(t, S_t)$$

[7] For its solution, see example 1.13.

[8] Technically, we say that W_t is a Brownian motion under the so-called Pricing measure $\mathbb{Q}$, also called the martingale measure. We denote the corresponding Brownian motion $W_t^{\mathbb{Q}}$ if S_t follows the dynamics

$$\begin{cases} \mathrm{d}S_u = rS_u\,\mathrm{d}u + \sigma S_u\,\mathrm{d}W_u^{\mathbb{Q}} \\ S_t = s \end{cases}$$

for some initial price S_t, $u > t$. We could, for example, then refer to a corresponding expectation under the $\mathbb{Q}$ as $\mathbb{E}^{\mathbb{Q}}$, i.e. the expectation is calculated based on the known $\mathbb{Q}$ dynamics.

for some smooth F, then, for some terminal time $T > 0$, F solves the Black-Scholes PDE

$$\begin{cases} F_t 0 rsF_s + \frac{1}{2}\sigma^2 s^2 F_{ss} - rF = 0 \\ F(T, s) = \Phi(s). \end{cases}$$

Moreover, by the Feynman-Kac theorem, we have

$$F(t, s) = e^{-r(T-t)} \mathbb{E}^{\mathbb{Q}}_{t,s} \left[\Phi(S_T) \right],$$

which is called the risk-neutral valuation formula.

Proof. To show the result, we let

$$\begin{cases} \tau = \frac{1}{2}\sigma^2 (T - t) \\ x = \log s. \end{cases}$$

Moreover, we assume that F is of form

$$F(t, s) = v(\tau, x).$$

Then

$$F_t = v_\tau \tau_t = -\frac{1}{2} v_\tau, \quad F_x = \frac{1}{s} v_x, \quad \text{and} F_{xx} = -\frac{1}{s^2} v_x + \frac{1}{s^2} vxx.$$

We plug these in to the Black-Scholes PDE and get

$$\begin{cases} v_\tau + \left(1 - \frac{2s}{\sigma^2}\right) v_x - v_{xx} + \frac{2r}{\sigma^2} v, \quad (\tau, x) \in \left[0, \frac{\sigma^2}{2} T\right] \times \mathbb{R} \\ v(0, x) = \Phi(e^x). \end{cases}$$

Next, let $u(\tau, x) = \exp\{-\alpha x - \beta\tau\}$ for some α and β. If we let

$$k = \frac{2r}{\sigma}, \quad \alpha = \frac{1}{2}(1 - k), \quad \text{and } \beta = -\frac{1}{4}(k + 1)^2,$$

then u solves (details left as Exercise 316)

$$\begin{cases} u_\tau - \Delta u = 0, \quad (\tau, x) \in [0, \frac{\sigma^2}{2} T] \times \mathbb{R} \\ u(0, x)) e^{-\alpha x} \Phi(e^x). \end{cases}$$

□

An avid reader notices that the Black-Scholes PDE is basically equivalent to the Cauchy initial value problem of the heat equation. The following result can be obtained in the so-called "Black-Scholes world", i.e. in a market where the assumptions of the Black-Scholes PDE hold. We don't focus on this result and don't derive it, but it is useful to know regarding some of our later results.

Theorem 3.17. *(Price of an European call option, the Black-Scholes formula).* Let $S_t = s, T > t$ and $r, \sigma > 0$ be known constants, and let k be the strike price. Moreover, let a stock price S_t follow the dynamics

$$\begin{cases} \mathrm{d}S_u = rS_u \,\mathrm{d}u + \sigma S_u \,\mathrm{d}W_u^{\mathbb{Q}} \\ S_t = s \end{cases}$$

for $t < u < T$. Then, the price of an European call option with a payoff function $\Phi(S_T) = (S_T - k)_+ := \max\{S_t - k, 0\}$ at terminal time T is

$$C(t, s) = s\mathrm{N}(d_1) - Ke^{-r(T-t)}\mathrm{N}(d_2),$$

where

$$\begin{cases} d_1 = \frac{\log \frac{s}{k} + \left(r + \frac{\sigma^2}{2}\right)(T-t)}{\sigma\sqrt{T-t}} \\ d_2 = d_1 - \sigma\sqrt{T-t} \end{cases}$$

and $\mathrm{N}(\cdot)$ is the cumulative density function of a standard normal distribution.

Proof. Omitted, see section 7.6. in Björk 2020. □

We recall that respectively to the European call option, the European put option is a contract with a payoff function $\Phi(S_T) = (k - S_T \vee 0)$.

Proposition 3.18. *(Price of an European put option, the Put-Call Parity).* Consider the notation and underlying items as in theorem 3.17. Then, the price of an European put option with a payoff function $\Phi(S_T) = (k - S_T \vee 0)$ at terminal time T is

$$p(t, s) = Ke^{-r(T-t)} + C(t, s) - s.$$

Proof. Follows from the so-called no-arbitrage argument[9]. See Björk 2020 Section 10. □

Next, we consider so-called Barrier options. We first define what they are and introduce some basic properties of these types of options, and then we proceed to price them using PDEs.

[9] We assume that we are operating in a market where there exist no arbitrage possibilities, i.e. that the prices of the financial instruments are arbitrage-free, i.e. that one can not make positive profit surely by changing from instrument to instrument. More specifically, for any (self-financed) portfolio V_t and $0 < t \leq T$, we don't have the following simultaneously:

$$\begin{cases} V_0 = 0, \\ \mathbb{P}(V_T \geq 0) = 1, \\ \mathbb{P}(V_T > 0) > 0. \end{cases}$$

For more discussion about arbitrage, see, for example, Björk 2020 Section 7

Definition 3.19. *(Barrier Options).* Consider $C(t,s)$, the pricing function of an European call option with strike price k and payoff function $\Phi(S_T)$. In this example, we consider the price function for the "down-and-out" (DO) version of a barrier option with respect to some barrier $L \in (0,k)$. The payoff at T for such option is

$$\tilde{\Phi}(S_T) = \begin{cases} \Phi(S_T), & \text{if } S_t > L \text{ for all } t \le T \\ 0, & \text{if } S_t \le L \text{ for some } t \le T, \end{cases}$$

or in other words

$$\tilde{\Phi}(S_T) = ((S_T - k)_+)\, \mathbb{1}_{\min_{t\le T} S_t > L}.$$

In addition, there are also the barrier types

- "down-and-in" (DI), $\tilde{\Phi}(S_T) = ((S_T - k)_+)\, \mathbb{1}_{\min_{t\le T} S_t \le L}$,
- "up-and-out" (UO), $\tilde{\Phi}(S_T) = ((S_T - k)_+)\, \mathbb{1}_{\max_{t\le T} S_t < L}$, and
- "up-and-in" (UI), $\tilde{\Phi}(S_T) = ((S_T - k)_+)\, \mathbb{1}_{\max_{t\le T} S_t \ge L}$.

The following proposition follows easily when one notes that the less likely one gets positive payoff, the cheaper the corresponding European option becomes.

Proposition 3.20. *(Parity for European-type options).* Consider the above European barrier options and $C(t,s)$, the pricing function of an European call option with strike price k and payoff function $\Phi(S_T)$. Then, for the prices of call options we have

$$\mathrm{DO}_{call} + \mathrm{DI}_{call} = \mathrm{UO}_{call} + \mathrm{UI}_{call} = C(t,s),$$

and correspondingly for the prices of put options we have

$$\mathrm{DO}_{put} + \mathrm{DI}_{put} = \mathrm{UO}_{put} + \mathrm{UI}_{put} = p(t,s),$$

Proof. Follows from definitions. □

The above parity is, sadly, not sufficient on its own to determine the prices of DO-options. In addition to the knowledge of the price of the call-option (which we know), we would need to know the probability density of the process $S_{\tau\wedge t}$ with

$$\tau = \inf\{t \ge 0 : S_t = L\}.$$

This price is discussed a lot, for example in the book Björk 2020, theorem 18.8. Instead of solely relying on their guidance, we present the price using our theory of PDE's.

Theorem 3.21. *(Price of a DO-call option).* Consider notation as above and let $F(t,s) := \mathrm{DO}_{call}$. Then

$$F(t,s) = C(t,s) - \left(\frac{s}{L}\right)^{1-\frac{2r}{\sigma^2}} C(t, \frac{L^2}{s}).$$

Proof. We begin by referring to the theorem 3.16, and recall that the connection of $F(t,s)$ of the BS-PDE and the $u(\tau,x)$ of the Heat equation lies in transformations

$$\tau = \frac{\sigma^2}{2}(T-t), \quad x = \log s,$$

so that

$$F(t,s) = e^{\alpha x + \beta\tau} u(t,x)$$

for

$$\alpha = \frac{1}{2}\left(1 - \frac{2r}{\sigma^2}\right), \quad \beta = -\frac{1}{4}\left(1 + \frac{2r}{\sigma}\right)^2.$$

In order to find a solution with our method, we need an initial point for the PDE to be uniquely solvable. We observe the following.

For the regular call option, we have

$$\Phi(x) := u(0,x) = e^{-\alpha x} F(T,s) = e^{-\alpha x}\left(e^x - k \vee 0\right),$$

and additionally for barrier options $S > L$ implies that $x > \log L$.

We further proceed similarly to the quarter-plane problem 314: we wish to extend Φ to make it odd relative to $x = \log L$. That is, we are looking for some $u(\tau,x)$ such that

$$u(\tau, 2\log L - x) = -u(\tau, x)$$

for all x.

Next, we extend Φ into some

$$h_0(x) := \Phi(x) - \Phi(2\log L - x).$$

Similarly as in example 3.1, now $u = g * h_0$ solves

$$\begin{cases} u_t - \Delta u = 0, \\ u(0,x) = h_0, \end{cases}$$

and thus u is a smooth function odd with respect to $\log L$. That is, $u(t, \log L) = 0$.

For the regular call option, we have

$$\left(g * \Phi\right)(\tau,x) = e^{-\alpha x - \beta\tau} C(t, e^x).$$

Moreover

$$\begin{aligned}\left(g * h_0\right)(\tau,x) &= \left(g * \Phi\right)(\tau,x) - \left(g * \Phi(2\log L - y)\right)(\tau,x) \\ &= \left(g * \Phi\right)(\tau,x) - e^{-\alpha(2\log L - x) - \beta\tau} C(t, e^{2\log L - x}\end{aligned}$$

for some arbitrary y. From this we see that

$$\begin{aligned} F(t,s) &= e^{\alpha x+\beta\tau}\left(g * h_0\right)(\tau,x) \\ &= C(t,s) - e^{\alpha x-\alpha(2\log L-x)}C(t,e^{2\log L-x}) \\ &= C(t,s) - \left(\frac{s}{L}\right)^{1-\frac{2r}{\sigma^2}} C(t,\frac{L^2}{s}) \end{aligned}$$

for some $s > L$. $\square$

For more discussion related to the proof of the above theorem, see Kohn 2014 Section 2.

3.4 Exercises for Chapter 3

Heat Equation and Heat Kernel

301. Show that $f * g = g * f$, where $*$ denotes the convolution operator.

302. Let $g(t,x)$ be the fundamental solution to the 1-dimensional heat equation. Show that $g_t - \Delta g = 0$.

303. Let $g(t,x)$ be the fundamental solution to the 1-dimensional heat equation. Show that
$$\lim_{t\searrow 0} g(t,x) = \begin{cases} 0, x \neq 0, \\ \infty, x = 0. \end{cases}$$

304. Let $g(t,x)$ be the fundamental solution to the 1-dimensional heat equation and fix $t > 0$. Show that
$$\int_{\mathbb{R}} g(t,x)dx = 1.$$

305. Let $g(t,x)$ be the fundamental solution to the heat equation in $\mathbb{R}^n$ and fix $t > 0$. Show that
$$\int_{\mathbb{R}^n} g(t,x)dx = 1.$$

For the following exercises, consider the one-dimensional heat equation (Cauchy initial value problem)

$$(*) \quad u_t(t,x) - u_{xx}(t,x) = 0,$$

where

$$0 < t < T, \quad T \in (0,\infty], \quad a < x < b, \quad a \in [-\infty,\infty), \quad b \in (-\infty,\infty].$$

The exercises serve as a proof to the properties 3.13 of the solution to the Cauchy problem.

306. Let $g(t,x)$ be the fundamental solution to the 1-dimensional heat equation. Check that g indeed satisfies the Cauchy initial problem via u.

307. Show that if u, v both solve $(*)$, and $\alpha, \beta \in \mathbb{R}$, then $\alpha u + \beta v$ is also a solution.

308. Show that if u solves $(*)$, and $\alpha > 0, x_0 \in \mathbb{R}$, then $u(\alpha^2 t, \alpha x - x_0)$ solves $(*)$ for $t \in (0, \alpha^2 T)$ and $x \in (\alpha a + x_0, \alpha b + x_0)$.

309. Show that if $u \in C^3$ and u solves $(*)$, then its derivatives u_t, u_x also solve $(*)$.

310. Show that if u solves $(*)$ and $v(t,x) = \int_a^x u(t,z)dz, x \in (a,b)$, then v also solves $(*)$ given that
$$\lim_{z \to a} u_x(t,z) = 0$$
for each $t \in (0,T)$.

311. Show that if u solves $(*)$ and $f : \mathbb{R} \to \mathbb{R}$ is a function, then $(u * f)(t,x)$ also solves $(*)$.

Cauchy IVP Problems and applications to the HE

312. Show that if f is even and h is odd relative to x_0, then $f * h$ is odd relative to x_0.

313. **(Homogeneous Neumann Condition).** Let $u(t,x)$ be a solution to the heat equation
$$u_t(t,x) - u_{xx}(t,x) = 0$$
on $\{(t,x) : t > 0, x > 0\}$ with $u(0,x) = u_0(x)$ for $x > 0$, and $\frac{\partial u}{\partial x}(t,0) = 0$ for $t > 0$. Show that
$$u(t,x) = \int_0^\infty u_0(y)h(t,x,y)dy$$
for some function $h(t,x,y)$.

314. **(Time varying boundary)**. Check that
$$u(t,x) = \int_0^t G_y(t-s,x,0)f(s)ds$$
solves the quarter-plane problem
$$\begin{cases} u_t - u_{xx} = 0, & t,x > 0, \\ w(t,0) = f(t), \\ w(0,x) = 0, \end{cases}$$
where $G(t,x,y)$ is the Green's function for the quarter-plane problem.

315. **(Infinite propagation speed)**. Show that in the Cauchy IVP problem (for example 3.12), if the initial data $\Phi(x) \geq 0, \Phi(x) \not\equiv 0$, then

$$u(t,x) = \int_{\mathbb{R}} g(t, x-y)\Phi(y)dy > 0$$

for all $x \in \mathbb{R}, t > 0$. (*Hint: Use a "min" version of the Strong maximum principle*).

316. **(Solving suitable transformations for the Black-Scholes-PDE, see 3.16)**Let $u(\tau, x) = \exp\{-\alpha x - \beta\tau\}$ for some α and β. Show that if

$$k = \frac{2r}{\sigma}, \quad \alpha = \frac{1}{2}(1-k), \quad \text{and } \beta = -\frac{1}{4}(k+1)^2,$$

then u solves

$$\begin{cases} u_\tau - \Delta u = 0, & (\tau, x) \in [0, \frac{\sigma^2}{2}T] \times \mathbb{R} \\ u(0,x))e^{-\alpha x}\Phi(e^x). \end{cases}$$

317. **(HE with positive rate of diffusion)**. Show that

$$g\,(t,x) \;= \frac{1}{\sqrt{2kt}}\varphi(\frac{x}{\sqrt{2kt}})$$

solves $g_t = kg_{xx}$ for some constant k.

Chapter 4

Markov Processes and the Kolmogorov Equations

4.1 Markov Processes

Sometimes arbitrary stochastic processes are difficult to solve, and we have to make limiting assumptions for their solvability. One such assumption is the so-called Markov property, which makes processes easier to manage and which is an intuitive concept.

Definition 4.1. *(Markov Property and Process).* Let $X_t \in \mathbb{R}^n$ be a stochastic process and $\mathcal{F}_t$ be its natural filtration. We say that X_t has the Markov property if, for any $s \in [0, t]$ and for any bounded Borel function f, we have

$$\mathbb{E}\Big[f(X_t)|\mathcal{F}_s\Big] = \mathbb{E}\Big[f(X_t)|X_s\Big].$$

If a process X_t has the Markov property, we say that it is a Markov process.

It is easy for the reader to see that an equivalent definition would be to define the Markov property for X_t if, for any Borel set $A \in \mathbb{R}^n$

$$\mathbb{P}\left(X_t \in A|\mathcal{F}_s\right) = \mathbb{P}\left(X_t \in A|X_s\right).$$

The intuition is clear: if X_t is a Markov process, our beliefs of its behaviour at some future time point t are the same regardless if we are observing only the latest realization X_s, or the whole "history" $\mathcal{F}_s$. A naive, or perhaps an underlining understanding about the assumption is that for Markov processes, the future depends on the past only through the present. Sounds unreasonable for many real-life processes, doesn't it?

The property also admits a stronger version.

Definition 4.2. *(Strong Markov Property).* Let τ be any stopping time and X_t, f be as in definition 4.1. If, on event $\{\tau < \infty\}$, X_t satisfies

$$\mathbb{E}\Big[f(X_{\tau+t})|\mathcal{F}_\tau\Big] = \mathbb{E}\Big[f(X_{\tau+t})|X_\tau\Big]$$

for each $t \geq 0$, then we say that X_t has the Strong Markov property.

By taking $\tau = t$, it is straightforward to see how the strong property implies the Markov property.

Naturally, these new definitions are to be understood with some key results.

Proposition 4.3. *(The Brownian motion is Markov).* Let W_t be the standard Brownian motion. Then

(i) W_t has a Strong Markov property, and

(ii) the Itô diffusion, for some standard drift and diffusion coefficients,

$$\mathrm{d}X_t = \mu(t, X_t)\,\mathrm{d}t + \sigma(t, X_t)\,\mathrm{d}W_t,$$

has a Strong Markov property.

Proof. Omitted, see Øksendal 2003 sections 7.1.2 and 7.2.4. □

Remark 4.4. The theorem above suggests that independent increments of a stochastic process imply a Markov property. This is true, but the reverse does not hold.

In literature, Markov processes induce a language of their own. We adapt some of it as the notation is efficient, and mostly understandable.

Definition 4.5. *(Transition Probability Measure and Density).* Let $X_t \in \mathbb{R}^n$ be a Markov process and $A \in \mathbb{R}^n$ a Borel set. We call the probability

$$P(A, t; x, s) := \begin{cases} \mathbb{P}\left(X_t \in A | X_s = x\right), & \text{if } X_t \text{ is a discrete stochastic process,} \\ \mathbb{E}_{s,x}\left[\mathbb{1}_A(X_t)\right], & \text{if } X_t \text{ is a continuous stochastic process,} \end{cases}$$

a transition probability measure at time t, from state x at time $s < t$. Moreover, if the measure P has a density, we denote the density by $p(y, t; x, s)$, the transition density function from x at s to y at t[1] as the conditional density

$$p(y, t; x, s) := f_{X_t|X_s}(y|x).$$

[1] In the literature, the pair (y, t) is often called the pair of forward variables and the pair (x, s) is called the pair of backward variables.

We note that splitting the definition of the transition probability measure is necessary, as for any continuous X_s we have $\mathbb{P}(X_s = x) = 0$. By the definition of conditional probability, the definition would not be well-defined.

It should also be noted that

$$P(A,t;x,s) = \int_A p(y,t;x,s)\,\mathrm{d}y,$$

and similarly if we consider the whole space $\mathbb{R}^n$, we would have

$$\int_{\mathbb{R}^n} p(y,t;x,s)\,\mathrm{d}y = 1,$$

which makes sense considering that we denote p a probability density function.

We now describe a distribution which doesn't change in time[2].

Definition 4.6. *(Stationary Distribution and Stationary Density).* Let X_t be a time-homogeneous Itô diffusion with $\mu = \mu(x)$ and $\sigma = \sigma(x)$. We say that μ is a Stationary Distribution or Invariant Measure if

$$\mathbb{P}\left(X_t \in A | X_0 \sim \mu(x)\right) = \mu(A)$$

for all measurable sets $A \in \mathbb{R}^n$.

Moreover, if the process has a density $\rho(t,x)$, then ρ is a stationary density if

$$\rho_t = 0 \Leftrightarrow \mathcal{L}^*\rho = 0.$$

It should be noted that for the stationarity density, similarly for other probability densities, we have that

$$\int_{\mathbb{R}} \rho_\infty\,\mathrm{d}x = 1.$$

For example, the concept of a probability density can be used to describe the so-called large-time statistics of a process, if such exists:

$$\rho_\infty(x) = \lim_{t\to\infty} \rho(t,x).$$

For this, see the Example 4.3. For other uses of a probability density, see Example 4.4.

Before moving to the Kolmogorov equations, we expand our knowledge regarding infinitesimal generators. Recall the definition 2.1 for the infitesimal generator and the result 2.3 on how its calculated for Itô diffusions. For further use, we define its adjoint.

[2]Recall the analogous concept for a Markov chain. We say that π is a stationary distribution, if, for transition matrix P we have

$$\pi = \pi P.$$

Definition 4.7. *(Adjoint of the Infinitesimal Generator).* Let $\mathcal{L}$ be the infinitesimal generator and $g, h \in C_o^2$ smooth functions. Provided that the inner products exists, the adjoint of the generator $\mathcal{L}$, denoted by $\mathcal{L}^*$, is defined so that

$$\langle \mathcal{L}g, h \rangle_{L^2} = \langle g, \mathcal{L}^* h \rangle_{L^2},$$

or in $\mathbb{R}$,

$$\int_{\mathbb{R}} h(x)(\mathcal{L}g)(x)\,\mathrm{d}x = \int_{\mathbb{R}} g(x)(\mathcal{L}^* h)(x)\,\mathrm{d}x.$$

The definition itself is non-productive as it doesn't show the form of the adjoint.

Theorem 4.8. *Adjoint generator for Itô Diffusion* Let $X_t \in \mathbb{R}^n$ be a d-dimensional Itô diffusion and $g \in C_o^2$ a smooth function. Then $\mathcal{L}^* g(x)$ exists, and

$$\mathcal{L}^* g(x) = -\sum_i \frac{\partial}{\partial x_i}(\mu_i g)(x) + \frac{1}{2}\sum_{i,j} \frac{\partial^2}{\partial x_i x_j}(C_{ij} g),$$

where $[C_{ij}] = \sigma\sigma^T$.

Proof. To show that $\mathcal{L}^* g(x)$ satisfies the criterion in the definition is left as an exercise 404. □

4.2 The Kolmogorov Equations

In this section, we present the most common Kolmogorov-equations, such as the Chapman-Kolmogorov equation, Kolmogorov Backward equation, and the Fokker-Planck equation.

First equation is motivated for a desire to have some transition-like properties for a Markov process. Within the context of the equation, we wish for the process to move from (x, s) to (y, t). We do this by choosing some in-between time u and integrate over all the possible locations z at the intermediate time u.

Theorem 4.9. *(Chapman-Kolmogorov Equation).* Let X_t be a Markov Process. Then

$$p(y, t; x, s) = \int_{\mathbb{R}^n} p(y, t; z, u)p(z, u; x, s)\,\mathrm{d}z$$

for $s \le u \le t$.

Proof. It suffices to prove the result in a $1-$dimensional case. Indeed, by direct calculations and the properties of joint densities, we have

$$\begin{aligned}
p(y,t;x,s) = f_{X_t|X_s}(y|x) &= \frac{f_{X_t,X_s}(y,x)}{f_{X_s}(x)} \\
&= \frac{\int_{\mathbb{R}^n} f_{X_t,X_u,X_s}(y,z,x)\,\mathrm{d}z}{f_{X_s}(x)} \\
&= \int_{\mathbb{R}^n} f_{X_t,X_u|X_s}(y,z|x)\,\mathrm{d}z \\
&= \int_{\mathbb{R}^n} f_{X_t|X_u,X_s}(y|z,x) f_{X_u|X_s}(z|x)\,\mathrm{d}z \\
&= \int_{\mathbb{R}^n} f_{X_t|X_u}(y|z) f_{X_u|X_s}(z|x)\,\mathrm{d}z \\
&= \int_{\mathbb{R}^n} p(y,t;z,u)p(z,u;x,s)\,\mathrm{d}z.
\end{aligned}$$

□

Let us now examine a familiar process, an Itô diffusion

$$\mathrm{d}X_t = \mu(t,X_t)\,\mathrm{d}t + \sigma(t,X_t)\,\mathrm{d}W_t$$

for Lipschitz-regular *mu* and σ. Recall that Itô diffusion has a strong Markov property, that is, the following result considers more specific process than the Chapman-Kolmogorov Equation. We present the result in two different forms, first for the transition measure P and the second for the density p.

Theorem 4.10. *(Kolmogorov Backward Equation, KBE 1st form).* Let P be a transition measure as a function of the backward variables (x,s). Let $A \subset \mathbb{R}^n$ and $t > 0$. Then

$$\begin{cases} \frac{\partial P}{\partial s}(A,t;x,s) + \mathcal{L}P(A,t;x,s) = 0, \quad (s,x) \in (0,t) \times \mathbb{R}^n \\ P(A,t;x,t) = \mathbb{1}_A(x). \end{cases}$$

Proof. The result follows from Feynman-Kac with the payoff function $\Phi = \mathbb{1}_A$. □

Theorem 4.11. *(Kolmogorov Backward Equation, KBE 2nd form).* Let p be the transition density as a function of the backward variables (x,s). Let $y \in \mathbb{R}^n$ and $t > 0$. Then

$$\begin{cases} \frac{\partial p}{\partial s}(y,t;x,s) + \mathcal{L}p(y,t;x,s) = 0, \quad (s,x) \in (0,t) \times \mathbb{R}^n \\ p(y,t;x,s) = \delta(x-y), \quad \text{as } s \to t. \end{cases}$$

Proof. Follows as a corollary from the first form of the Kolmogorov Backward equation by letting A shrink to a single point. □

We are now ready to proceed to our third Kolmogorov equation bearing two names. Similarly to the backward equation, the forward equation has two forms.

Theorem 4.12. *(Kolmogorov Forward Equation/Fokker-Planck/KFE, 1st form).* Let p be the transition density with respect to the forward variables (y,t). Let $s > 0$. Then

$$\begin{cases} \frac{\partial p}{\partial t}(y,t;x,s) - \mathcal{L}^* p(y,t;x,s) = 0, & (t,y) \in (s,\infty) \times \mathbb{R}^n \\ p(y,t;x,s) = \delta(x-y), & \text{as } t \to s. \end{cases}$$

Proof. We provide proof in the 1−dimensional case and assume that X_t has a transition probability density with initial value $X_s = x$. Let $h \in C_o^2$. Then, for $t > s$ by Itô's:

$$h(X_t) = h(x) + \int_s^t \mu h_x + \frac{1}{2}\sigma^2 h_{xx}\,\mathrm{d}u + \int_s^t h_x\,\mathrm{d}W_u.$$

Taking expectations, we get

$$\begin{aligned} \mathbb{E}_{s,x}\left[h(X_t)\right] &= h(x) + \mathbb{E}_{s,x}\left[\int_s^t (\mathcal{L}h)(X_u)\,\mathrm{d}u\right] \\ &= h(x) + \int_s^t \mathbb{E}_{s,x}\left[(\mathcal{L}h)(X_u)\right]\,\mathrm{d}u \\ &= h(x) + \int_s^t \int_{\mathbb{R}} (\mathcal{L}h)(y)p(y,u;x,s)\,\mathrm{d}y\,\mathrm{d}u. \end{aligned}$$

Noting that by definition

$$\mathbb{E}_{s,x}\left[h(X_t)\right] = \int_{\mathbb{R}} h(y)p(y,t;x,s)\,\mathrm{d}y$$

and differentiating both representations of the expectation with respect to t we get

$$\begin{aligned} \int_{\mathbb{R}} h(y)\frac{\partial p}{\partial t}(y,t;x,s)\,\mathrm{d}y &= \int_{\mathbb{R}} (\mathcal{L}h)(y)p(y,u;x,s)\,\mathrm{d}y \\ &= \int_{\mathbb{R}} h(y)(\mathcal{L}^* p)(y,u;x,s)\,\mathrm{d}y. \end{aligned}$$

Since h is arbitrary, we conclude that

$$\frac{\partial p}{\partial t} = \mathcal{L}^* p.$$

□

The second form of the theorem describes a probability distribution by solving a certain initial value problem.

Theorem 4.13. *(Kolmogorov Forward Equation/Fokker-Planck/KFE, 2nd form).* Let ρ be the probability density of X_t and $\rho_s(x)$ be the initial density at time s. Then

$$\begin{cases} \frac{\partial}{\partial t}\rho - \mathcal{L}^*\rho(t,x) = 0 \\ \rho(s,x) = \rho_s(x). \end{cases}$$

Proof. Omitted. □

Note that the Kolmogorov Forward equation is weaker result that the corresponding backward equation. We illustrate the equations with few examples.

Example 4.1. *(Kolmogorov Equations for the Brownian motion).* Let $\rho(t,x)$ be the density of a Brownian motion, $u(t,x) = \mathbb{E}[f(X_T)]$. Then the Kolmogorov Forward Equation/Fokker-Planck suggests that

$$\begin{cases} \rho_t - \frac{1}{2}\Delta\rho = 0, & t > s \\ \rho(0,t) = \delta_0. \end{cases}$$

Similarly, the Kolmogorov Backward Equation suggests that

$$\begin{cases} u_t - \frac{1}{2}\Delta u = 0, & t < T \\ u(T,x) = f(x). \end{cases}$$

Observe also that here $\mathcal{L} = \mathcal{L}^*$, that is, the two equations give the same results but only reversed in time.

Example 4.2. *(Kolmogorov Forward Equation for a Brownian Motion with a drift).* Now consider a process X_t with dynamics $\mathrm{d}X_t = \mu\,\mathrm{d}t + \mathrm{d}W_t$ with $X_s = 0$. Clearly, for $t > s$, we have

$$X_t = \mu(t-s) + (W_t - W_s).$$

Now we have $p(y,t;0,s) = \frac{\partial}{\partial y}\mathbb{P}(X_t \leq y | X_s = 0)$ with

$$\mathbb{P}(X_t \leq y | X_s = 0) = \mathbb{P}\left(\frac{W_t - W_s}{\sqrt{t-s}} \leq \frac{y - \mu(t-s)}{\sqrt{t-s}}\right) = \mathrm{N}(f(y)),$$

where $f(y) := \frac{y-\mu(t-s)}{\sqrt{t-s}}$ and $\mathrm{N}(\cdot)$ is the cumulative density function of the standard normal distribution. Moreover,

$$\mathrm{N}(f(y)) = \frac{1}{\sqrt{2\pi}} \exp\left\{-\frac{1}{2}f(y)^2\right\} \frac{\partial}{\partial y} f(y) = \frac{1}{\sqrt{2\pi(t-s)}} \exp\left\{-\frac{(y-\mu(t-s))^2}{2(t-s)}\right\}.$$

Thus, the Kolmogorov Forward equation is

$$\frac{\partial}{\partial t}p(y,t;0,s) + \mu\frac{\partial}{\partial y}p(y,t;+,s) - \frac{1}{2}\frac{\partial^2}{\partial y^2}p(y,t;0,s) = 0,$$

i.e. $\mathcal{L} \neq \mathcal{L}^*$, unlike in Example 4.1.

In addition, a remark might be in place.

Remark 4.14. We note that the Kolmogorov Forward equation is in general difficult to solve analytically. Instead, one often solves the problem considering the so-called large-time behaviour or statistics of a given Itô diffusion. Usually this is done with the help of stationary densities, if any exist.

4.3 Applications to the Kolmogorov Forward Equation

Example 4.3. *(Stationary Density for Ornstein-Uhlenbeck).* Consider the Ornstein-Uhlenbeck process $\mathrm{d}X_t = \mu X_t + \sigma\,\mathrm{d}W_t$ for some $\mu > 0$. Then, the Kolmogorov Backward Equation gives us

$$\frac{\partial \rho}{\partial t} - \left(-\frac{\partial}{\partial x}(\mu x \rho) + \frac{1}{2}\sigma^2 \frac{\partial^2 \rho}{\partial x^2}\right) = 0$$

for some $\rho = \rho(t, x)$. If ρ is the stationary density, i.e. if $\rho = \rho_\infty$, then the condition

$$\mathcal{L}^* \rho_\infty = 0$$

yields

$$-\frac{\partial}{\partial x}(\mu x \rho_\infty) + \frac{1}{2}\sigma^2 \frac{\partial^2 \rho_\infty}{\partial x^2}.$$

Integrating this with respect to x yields

$$-\mu x \rho_\infty + \frac{1}{2}\sigma^2 \frac{\partial \rho_\infty}{\partial x} = C$$

for some constant C. Moreover, since $\int_{\mathbb{R}} \rho_\infty \,\mathrm{d}x = 1$, we expect that $\rho_\infty \to 0$ faster than $\frac{1}{|x|} \to 0$ as $|x| \to \infty$. This suggests that $C = 0$. Reorganizing our previous calculations provides us with

$$\frac{\frac{\partial \rho_\infty}{\partial x}}{\rho_\infty} = \frac{2\mu x}{\sigma^2}.$$

Recalling the derivative of a natural logarithm, this leads to

$$\log \rho_\infty = \frac{\mu}{\sigma^2}x^2 + D \quad \Leftarrow \quad \rho_\infty = De^{\frac{\mu}{\sigma^2}x^2}$$

for some constant D. The constant can be solved by recalling that for densities we have

$$\rho_\infty(x) = \lim_{t\to\infty} \rho(t, x),$$

and thus solving for D yields

$$\rho_\infty(x) = \frac{1}{\sqrt{\frac{2\pi\sigma^2}{-2\mu}}} e^{-\frac{1}{2}\frac{x^2}{\frac{\sigma^2}{-2\mu}}}.$$

Recalling knowledge about normal distributions, we see that this is exactly the probability density function of a normal distributed random variable with expectation 0 and variance $\sigma^2/(-2\mu)$.

Remark 4.15. It should be remarked that most processes have no stationary distribution, i.e. the limit doesn't exist. For example, the Brownian motion has no stationary distribution. Actually, it turns out that the OU process is the only Markovian, Stationary, and Gaussian process.

Example 4.4. *(Moments of the Brownian motion revisited).* We wish to determine the kth moments of W_t. Let the kth moment be denoted by $\beta_k(t)$. Then

$$\beta_k(t) = \int_{\mathbb{R}} x^k \rho(t,x)\,\mathrm{d}x,$$

where $\rho(t,x)$ solves the PDE

$$\rho_t - \frac{1}{2}\rho_{xx} = 0.$$

Climbing up the integer ladder, the first two steps are easy. If $k = 0$, we get $\beta_0(t) = \int_{\mathbb{R}} x^0 \rho(t,x)\,\mathrm{d}x = 1$ as we integrate a density over the real line. Not similarly, but in an equally satisfying way, by noting that the integrand $x\rho(t,x)$ is odd, we get $\beta_1(t) = \int_{\mathbb{R}} x\rho(t,x)\,\mathrm{d}x = 0$.

If $k \geq 2$, things get more interesting. Differentiating $\beta_k(t)$ with respect to t yields by Kolmogorov Forward Equation (2nd form) and integration by parts (twice):

$$\begin{aligned}
\frac{\partial}{\partial t}\beta_k(t) &= \int_{\mathbb{R}} x^k \rho_t\,\mathrm{d}x \\
&= \frac{1}{2}\int_{\mathbb{R}} x^k \rho_{xx}\,\mathrm{d}x \\
&= \frac{1}{2}\left(\left[x^k\rho_x\right]_{-\infty}^{\infty} - \int_{\mathbb{R}} kx^{k-1}\rho_x\,\mathrm{d}x\right) \\
&= -\frac{1}{2}k\int_{\mathbb{R}} x^{k-1}\rho_x\,\mathrm{d}x \\
&= -\frac{1}{2}\left(\left[x^{k-1}\rho\right]_{-\infty}^{\infty} - \int_{\mathbb{R}} (k-1)x^{k-2}\rho\,\mathrm{d}x\right) \\
&= \frac{1}{2}k(k-1)\beta_{k-2}(t).
\end{aligned}$$

Thus for $k \geq 2$, $\beta_k(t)$ is recursively defined.

The above example could also be used to solve the moments of the Geometric Brownian motion.

4.4 Local Volatility and Dupire's Formula

For the remaining section of this chapter, we focus on financial applications. In the classic Black-Scholes world, one assumes the volatility σ to be constant. On the other hand, if the model assumes σ to be a stochastic process, the model is often called a *stochastic volatility model.* Moreover, if the stochastic process σ can be represented as a deterministic function of time t and the asset price S_t, the model is called a *local volatility model.*

Here we consider asset price S_t with the dynamics

$$\mathrm{d}S_t = rS_t\,\mathrm{d}t + \sigma(t,S_t)S_t\,\mathrm{d}W_t,$$

and let $\rho(t, s)$ be probability density of S_t.

The question of interest is: can we use the price of an European call option $c(T, k)$ at time 0 with a terminal time T and strike price k to derive a formula for the local volatility $\sigma(t, S_t)$? In other words, our aim is to derive some PDE satisfied by $c(T, k)$.[3]

For the remaining of this section, we impose two additional assumptions:

(A) $\lim_{s\to\infty} s^2\rho(T, s) = 0$,

(B) $\lim_{s\to\infty} s\frac{\partial}{\partial s}\left(\sigma(T, s)^2 s^2\rho(T, s)\right) = 0$.

Fokker-Planck for ρ with S_t gives

$$\frac{\partial}{\partial t}\rho + r\frac{\partial}{\partial s}(s\rho) - \frac{1}{2}\frac{\partial^2}{\partial s^2}\left(\sigma^2 s^2\rho\right) = 0.$$

Moreover, by risk-neutral valuation (see Theorem 3.16), we have

$$c(T, k) = E^{\mathbb{Q}}\left[e^{-rT}(S_T - k)_+\right] = \int_k^\infty e^{-rT}(s-k)\rho(T, s)\,\mathrm{d}s.$$

Utilizing this and differentiating $c(T, k)$ with respect to k yields:

$$\begin{aligned}
\frac{\partial c}{\partial k} &= e^{-rT}\frac{\partial}{\partial k}\int_k^\infty (s-k)\rho(T, s)\,\mathrm{d}s \\
&= e^{-rT}\frac{\partial}{\partial k}\left(\int_k^\infty s\rho(T, s)\,\mathrm{d}s - k\int_k^\infty \rho(T, s)\,\mathrm{d}s\right) \\
&= e^{-rT}\left(-k\rho(T, k) - \int_k^\infty \rho(T, s)\,\mathrm{d}s + k\rho(T, k)\right) \\
&= -e^{-rT}\int_k^\infty \rho(T, s)\,\mathrm{d}s.
\end{aligned}$$

Moreover,

$$\frac{\partial^2 c}{\partial k^2} = e^{-rT}\rho(T, k).$$

Similarly, we find the derivative of $c(T, k)$ with respect to T.

[3] Again, this section could be imposed as an Example, but since the calculations are somewhat lengthy, the topic reads better as a section with few results in the end.

$$
\begin{aligned}
\frac{\partial c}{\partial T} &= \frac{\partial}{\partial T}\left(e^{-rT}\int_k^\infty (s-k)\rho(T,s)\,\mathrm{d}s\right)\\
&= -re^{-rT}\int_k^\infty (s-k)\rho\,\mathrm{d}s + e^{-rT}\int_k^\infty (s-k)\rho_T(T,s)\,\mathrm{d}s\\
&= -rc + e^{-rT}\int_k^\infty (s-k)\rho_T(T,s)\,\mathrm{d}s\\
&= -rc + e^{-rT}\int_k^\infty (s-k)\left(-r\frac{\partial}{\partial s}(s\rho) + \frac{1}{2}\frac{\partial^2}{\partial s^2}\left(\sigma^2 s^2\rho\right)\right)\,\mathrm{d}s\\
&= -rc + e^{-rT}\int_k^\infty -r(s-k)\frac{\partial}{\partial s}(s\rho)\,\mathrm{d}s + e^{-rT}\int_k^\infty (s-k)\frac{1}{2}\frac{\partial^2}{\partial s^2}\left(\sigma^2 s^2\rho\right)\,\mathrm{d}s,
\end{aligned}
$$

where the second last equality follows from again using the Fokker-Planck. Moreover, we further deduce that

$$
\begin{aligned}
&e^{-rT}\int_k^\infty -r(s-k)\frac{\partial}{\partial s}(s\rho)\,\mathrm{d}s\\
&= -e^{-rT}\int_k^\infty (s-k)\,\mathrm{d}(s\rho)\\
&= -re^{-rT}\left([(s\rho)(s-k)]_k^\infty - \int_k^\infty s\rho\,\mathrm{d}s\right)\\
&= -re^{-rT}\int_k^\infty s\rho\,\mathrm{d}s\\
&= r\left(e^{-rT}\int_k^\infty (s-k)\rho\,\mathrm{d}s - k\left(-e^{-rT}\int_k^\infty \rho\,\mathrm{d}s\right)\right)\\
&= r(c - k\frac{\partial c}{\partial k}).
\end{aligned}
$$

Here in the third last equality we used the assumption (A)

$$
\lim_{s\to\infty} s^2\rho(T,s) = 0
$$

from above, and in the last equality we recalled that indeed

$$
c = e^{-rT}\int_k^\infty (s-k)\rho\,\mathrm{d}s, \quad \frac{\partial c}{\partial k} = -e^{-rT}\int_k^\infty \rho\,\mathrm{d}s.
$$

We proceed with

$$
\begin{aligned}
& e^{-rT}\int_k^{\infty}(s-k)\frac{1}{2}\frac{\partial^2}{\partial s^2}\left(\sigma^2 s^2\rho\right)\,\mathrm{d}s \\
&= \frac{1}{2}e^{-rT}\int_k^{\infty}(s-k)\,\mathrm{d}\left(\frac{\partial}{\partial s}(\sigma^2 s^2\rho)\right) \\
&= \frac{1}{2}e^{-rT}\left(\left[(s-k)\frac{\partial}{\partial s}(\sigma^2 s^2\rho)\right]_k^{\infty}-\int_k^{\infty}\frac{\partial}{\partial s}(\sigma^2 s^2\rho)\,\mathrm{d}s\right) \\
&= -\frac{1}{2}e^{-rT}\int_k^{\infty}\frac{\partial}{\partial s}(\sigma^2 s^2\rho)\,\mathrm{d}s \\
&= \frac{1}{2}e^{-rT}\sigma^2(T,k)k^2\rho(T,k) \\
&= \frac{1}{2}\sigma^2(T,k)k^2\frac{\partial^2 c}{\partial k^2},
\end{aligned}
$$

where we used the assumption (B):

$$
\lim_{s\to\infty} s\frac{\partial}{\partial s}\left(\sigma(T,s)^2 s^2\rho(T,s)\right)=0.
$$

Now, recalling that

$$
\frac{\partial c}{\partial T}=-rc+e^{-rT}\int_k^{\infty}-r(s-k)\frac{\partial}{\partial s}(s\rho)\,\mathrm{d}s+e^{-rT}\int_k^{\infty}(s-k)\frac{1}{2}\frac{\partial^2}{\partial s^2}\left(\sigma^2 s^2\rho\right)\,\mathrm{d}s,
$$

we combine our calculations into the so-called Dupire's PDE. Moreover, we present a so-called Dupire's formula, which is used to calculate the local volatility based on the PDE.

Theorem 4.16. *(Dupire's PDE and Dupire's Formula).* Let $c(T,k)$ be the price of an European call option and $r>0$. Moreover, assume that

(A) $\lim_{s\to\infty} s^2\rho(T,s)=0$,

(B) $\lim_{s\to\infty} s\frac{\partial}{\partial s}\left(\sigma(T,s)^2 s^2\rho(T,s)\right)=0$.

Then, the Dupire's PDE holds

$$
\frac{\partial c}{\partial T}(T,k)=-rk\frac{\partial c}{\partial k}+\frac{1}{2}\sigma^2(T,k)k^2\frac{\partial^2 c}{\partial k^2}.
$$

Moreover, if $\frac{\partial c}{\partial T}$, $\frac{\partial c}{\partial k}$, and $\frac{\partial^2 c}{\partial k^2}$ are known, we have the Dupire's formula for local volatility:

$$
\sigma^2(T,k)=2\frac{\frac{\partial c}{\partial T}+rk\frac{\partial c}{\partial k}}{k^2\frac{\partial^2 c}{\partial k^2}}.
$$

Proof. Sketched above. $\square$

One recalls that with *implied volatility* we mean a volatility σ, which is calculated from the Black-Scholes PDE, given that we have the market data available. That is, implied volatility gives us an injective mapping between a realized trajectory of a stock price S_t and the volatility: the calculated volatility holds for only a single realization of stock price and is thus difficult to generalize! On the other hand, local volatility provides us with an injective map from all option prices to σ, and in such is a stronger result. However, local volatility requires a Markovian representation of σ.

After the set of exercises, we move to discuss Stochastic Control.

4.5 Exercises for Chapter 4

Markov Processes and Kolmogorov Equations

401. Verify that $\rho(t,x) = \frac{1}{\sqrt{2\pi t}}\exp(-\frac{x^2}{2t})$, the probability density function of a normal distribution $\mathrm{N}(0,t)$ (which corresponds to the distribution of the increments of a Brownian motion from 0 to t), solves $\rho_t = \frac{1}{2}\rho_{xx}$.

402. Let $u(t,x)$ be the probability density function of some stochastic process at time t. In addition, u solves $u_t = u_{xx}$. What is this process?

403. Write down $\mathcal{L}^*$ for the following 1-dimensional Itô diffusions:

    (a) $\mathrm{d}X_t = \mathrm{d}W_t$,

    (b) $\mathrm{d}X_t = \mu\,\mathrm{d}t + \sigma\,\mathrm{d}W_t$,

    (c) $\mathrm{d}X_t = \mu X_t\,\mathrm{d}t + \sigma\,\mathrm{d}W_t$,

    (d) $\mathrm{d}X_t = \mu X_t\,\mathrm{d}t + \sigma X_t\,\mathrm{d}W_t$.

404. Show that $\mathcal{L}^*$ is the adjoint of $\mathcal{L}$ with respect to the quadratic inner product: let g, h be smooth and vanish at infinity (along with their derivatives), then

$$\int_{\mathbb{R}} h(x)(\mathcal{L}g)(x)\,\mathrm{d}x = \int_{\mathbb{R}} g(x)(\mathcal{L}^* h)(x)\,\mathrm{d}x.$$

405. Argue that the steady-state distribution is indeed the limiting distribution of the OU process. *(Hint: lift T to infinity).*

406. Write down the KFE for a standard BM, and

    (a) show that the stationary solution is of the form $\rho_\infty(x) = ax + b$,

    (b) look at the boundary condition at ∞ and conclude $\rho_\infty(x) = b$,

    (c) conclude that such a distribution cannot be normalised.

    Together, these imply that $\rho_\infty(x) = 0$.

407. **(BM in the unit interval)**. Consider the KFE of a 1-dimensional Brownian motion in the interval $(0, 1)$ with the boundary condition

$$\frac{\partial \rho_\infty}{\partial x}|_{x=0} = \frac{\partial \rho_\infty}{\partial x}|_{x=1} = 0.$$

Show that the steady-state distribution exists and is uniform.

(Remark. This homogeneous Neumann boundary condition is often called the reflecting boundary).

408. **(Moments of Brownian motion, revisited)**. Fix $t > 0$ and let $\beta_k(t) := \mathbb{E}[W_t^k]$. Recall that

(a) $\mathbb{E}[X^n] = \int_{\mathbb{R}} x^n f(x)\,\mathrm{d}x$ where f is the probability density function,

(b) $\rho(t, x)$ decays sufficiently fast at infinity.

Use the KFE of the standard BM to show that

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

409. **(Generalized Ornstein-Uhlenbeck)**. Let $a > 0$ and $V : \mathbb{R} \to \mathbb{R}$ such that

$$\int_{\mathbb{R}} e^{-V(x)}\,\mathrm{d}x < \infty.$$

Find the limiting density function for X_t that solves

$$\mathrm{d}X_t = -aV'(X_t)\,\mathrm{d}t + \sqrt{2a}\,\mathrm{d}W_t.$$

Chapter 5

Optimal Stochastic Control

5.1 Formulating Optimal Control

Before venturing into the world of stochastic optimal control, we briefly remind the reader of the deterministic or classical control. In deterministic control, one often has a system, where the time derivative y_t solves

$$y_t = f(y(t), \alpha(t))$$

for some function f and control α. Here, the goal is to optimize – in other words, to minimize or maximize f as a some function of both the state variable y and the control α.

In this chapter, we study the stochastic version of this problem, where we control a certain set of stochastic differential equations and where the optimization is done with respect to a expected value of some process.

Definition 5.1. *(Controlled Stochastic Differential Equation).* Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space with a d-dimensional Brownian motion W_t. In stochastic control theory, we consider a controlled Stochastic differential equation

$$\begin{cases} \mathrm{d}X_t^\alpha = \mu(t, X_t, \alpha_t)\,\mathrm{d}t + \sigma(t, X_t, \alpha_t)\,\mathrm{d}W_t, \\ X_0^\alpha = x, \end{cases}$$

where $\mu : \mathbb{R}_+ \times \mathbb{R}^n \times A \to \mathbb{R}^n$ and $\sigma : \mathbb{R}_+ \times \mathbb{R}^n \times A$, and where A is the control set, and $\alpha \in A$ is the control process.

In order to make the problem well-defined, we impose some limitations to the family A.

Definition 5.2. *(Admissible Strategy).* Let α_t be a control process. We call it an admissible strategy if

(i) α_t is $\mathcal{F}_t$-adapted,

(ii) $\alpha_t(\omega) \in A$ for every t and ω, and

(iii) the Controlled Stochastic Differential Equation has a unique strong solution.

Definition 5.3. *(Markov Strategy).* Let α_t be an admissible strategy. We call it further a Markov strategy (or a Markovian) if

$$\alpha_t = \alpha(t, X_t^\alpha)$$

for some function $\alpha : \mathbb{R}_+ \times \mathbb{R}^n \to A$.

The concept of Markov strategies is important for optimal control, as a Markovian α_t implies Markovian X_t^α. Moreover, methods for solving the control problems, are easy to implement for Markov strategies and moreover, the method we present requires the strategy to be Markov.

As we stated in the beginning of this chapter, we wish to optimize – maximize or minimize – expected values, more prominently an expectation of a payoff function, by controlling α.

Definition 5.4. *(Expected Payoff Function, Value Function, Optimal Control).* Let D_T be the cylinder set[1] $D_T = [0, T) \times D$ with a boundary Γ_T, where $D \subset \mathbb{R}^n$ is an open and bounded set. Moreover, let

$$\begin{cases} \Psi^\alpha : D_T \times A \to \mathbb{R}, & \text{(running payoff function)} \\ \Phi : \Gamma_T \to \mathbb{R} & \text{(final payoff function)}. \end{cases}$$

Then, we say that J^α is the finite time horizon expected payoff function

$$J^\alpha(t, x) = \mathbb{E}_{t,x}\left[\Phi(X_T^\alpha) + \int_t^T \Psi^\alpha(s, X_s^\alpha)\,\mathrm{d}s\right],$$

and that V is the value function

$$V(t, x) = \sup_{\alpha \in A} J^\alpha(t, x) = J^{\alpha^*}(t, x).$$

Moreover, we say that if α^* exists, it is the optimal control.

Shortly, the purpose of Optimal control as a field is to find V and α^*. Together, they answer to questions of ”how much do I get?” and ”how much should I control?”

In addition, there are alternative formulations for the optimal control problem, just like there were alternative formulations for the Feynman-Kac formula 2.5, for example for stopping times, see 2.9.

[1]Introduced for the first time in Chapter 3

Remark 5.5. For example, we can also consider two modifications to the problem:

(i) Time horizon determined by a stopping time. Consider

$$\tau := \min\{\inf\{s \geq t | X_s^\alpha \notin D\}, T\}$$

with an expected payoff function

$$J^\alpha(t,x) = \mathbb{E}_{t,x}\left[\Phi(\tau, X_\tau^\alpha) + \int_t^\tau \Psi^\alpha(s, X_s^\alpha)\,\mathrm{d}s\right].$$

(ii) Infinite time horizon, where $t \in [0,\infty]$. Consider a discounting factor r and an expected payoff function

$$J_r^\alpha(t,x) = \mathbb{E}_{t,x}\left[\int_t^\infty e^{-rs}\Psi^\alpha(s, X_s^\alpha)\,\mathrm{d}s\right].$$

Naturally, if one were to maximize $J_r^\alpha(t,x)$, it is necessary to implement discounting for non-trivial solutions.

5.2 Dynamic Programming Principle and Hamilton-Jacobi-Bellman Equation

When figuring out on how to solve the optimal control problem formally set up in Definition 5.4, two natural questions come to mind.

1. Does the optimal control exist, i.e. does any α^* satisfy

$$J^{\alpha^*}(t,x) \geq J^\alpha(t,x)$$

for all $\alpha \in A$?[2]

2. If yes, how to find such α^*?

Here we focus on the case 2. Our approach is something called *dynamic programming*, which is centered around a principle of optimality, often referred to as the dynamic programming principle.

Theorem 5.6. *(Bellman's Principle of Optimality).* Let α^* be optimal on an interval $[t,T]$. Then it is also optimal on every subinterval $[s,T]$ for any $s \in [t,T]$.

Proof. Omitted. Is proven by the law of iterated expectation. □

[2] If the problem considers minimizing, respectively $J^{\alpha^*}(t,x) \leq J^\alpha(t,x)$.

Originally Richard Bellman formulated the principle so that α^* being the optimal control requires that after an initial control, say, x at time t, the optimal control should lead to optimal strategies based only on the initial point. These two definitions are equal.

Next, we proceed in motivating the co-called Hamilton-Jacobi-Bellman Equation. In other words, we are, as the reader is by now used to, derive a PDE and show that solving for the optimal control problem is equal to finding a solution to this particular PDE. Our strategy is as follows. We

1. fix $(t, x) \in D_T$,

2. choose a small $h > 0$ so that $t + h < T$, and we

3. choose an arbitrary control $\alpha \in A$.

Examining the break-point $t + h$, it is natural to consider the following two cases:

(i) Denote $\alpha_1 = \alpha^*$ for all $s \in [t, T]$, and

(ii) denote

$$\alpha_2 = \begin{cases} \alpha, & s \leq t+h, \\ \alpha^*, & s > t+h. \end{cases}$$

The expected payoff in case (i) is straightforward. Indeed,

$$J^{\alpha_1}(t, x) = J^{\alpha^*}(t, x) = V(t, x).$$

On the other hand, for case (ii) we have

$$\begin{aligned} J^{\alpha_2}(t, x) &= \mathbb{E}_{t,x}\left[\Phi(X_T^{\alpha_2}) + \int_t^T \Psi(s, X_s^{\alpha_2})\,\mathrm{d}s\right] \\ &= \mathbb{E}_{t,x}\left[\Phi(X_T^{\alpha^*}) + \int_{t+h}^T \Psi(s, X_s^{\alpha^*})\,\mathrm{d}s + \int_t^{t+h} \Psi(s, X_s^{\alpha^*})\,\mathrm{d}s\right] \\ &= \mathbb{E}_{t,x}\left[\int_t^{t+h} \Psi(s, X_s^{\alpha})\,\mathrm{d}s + \mathbb{E}_{t+h, X_{t+h}^{\alpha}}\left[\Phi(X_T^{\alpha^*}) + \int_{t+h}^T \Psi(s, X_s^{\alpha^*})\,\mathrm{d}s + \int_t^{t+h} \Psi(s, X_s^{\alpha^*})\,\mathrm{d}s\right]\right] \\ &= \mathbb{E}_{t,x}\left[\int_t^{t+h} \Psi(s, X_s^{\alpha})\,\mathrm{d}s + V(t+h, X_{t+h}^{\alpha})\right]. \end{aligned}$$

Moreover, we now examine $V(t+h, X_{t+h}^{\alpha})$ and use Itô's formula:

$$V(t+h, X_{t+h}^{\alpha}) = V(t, x) + \int_t^{t+h} \left(\frac{\partial V}{\partial t} + \mathcal{L}^{\alpha} V\right)\,\mathrm{d}s + \int_t^T \ldots\,\mathrm{d}W_s,$$

and by taking expectation we arrive at

$$E_{t,x}\left[V(t+h, X_{t+h}^{\alpha})\right] = V(t, x) + \mathbb{E}_{t,x}\left[\int_t^{t+h} \left(\frac{\partial V}{\partial t} + \mathcal{L}^{\alpha} V\right)\,\mathrm{d}s\right].$$

We note that $V(t,x) \geq J^{\alpha_2}(t,x)$ as V is defined to be the supremum over different payoff functions J. Combining this with the expectation, we get

$$V(t,x) \geq J^{\alpha_2}(t,x) = \mathbb{E}_{t,x}\left[\int_t^{t+h} \Psi(s, X_s^\alpha)\,\mathrm{d}s\right] + V(t,x) + \mathbb{E}_{t,x}\left[\int_t^{t+h} \left(\frac{\partial V}{\partial t} + \mathcal{L}^\alpha V\right)\,\mathrm{d}s\right],$$

from which we deduce

$$\mathbb{E}_{t,x}\left[\int_t^{t+h} \Psi(s, X_s^\alpha)\,\mathrm{d}s\right] \leq -\left[\int_t^{t+h} \left(\frac{\partial V}{\partial t} + \mathcal{L}^\alpha V\right)\,\mathrm{d}s\right].$$

To formulate this inequality into a PDE, we combine the expectations, divide the terms by h and let $h \to 0$. We arrive at

$$\frac{\partial V}{\partial t}(t,x) + \mathcal{L}^\alpha V(t,x) + \Psi^\alpha(t,x) \leq 0.$$

Having formulated this PDE, consider the following.

Remark 5.7. Consider the inequality

$$\frac{\partial V}{\partial t}(t,x) + \mathcal{L}^\alpha V(t,x) + \Psi^\alpha(t,x) \leq 0.$$

For this inequality, one achieves equality if and only if $\alpha = \alpha^*$. Moreover, since the pair (t,x) is arbitrary, the inequality holds for all $(t,x) \in D_T$. Finally, we note that for the terminal value we have $V(T,x) = \Phi(x)$.

Now, if we note that $\frac{\partial V}{\partial t}$ is independent of α, we take supremum of the inequality to arrive at the Hamilton-Jacobi-Bellman equation.

Theorem 5.8. *(Hamilton-Jacobi-Bellman Equation).* Let value function $V \in C^{1,2}$ and consider the running payoff function $\Psi^\alpha : D_T \times A \to \mathbb{R}$ and the final payoff function $\Phi : \Gamma_T \to \mathbb{R}$. Moreover, let the optimal control α^* exist. Then V is the solution to the Hamilton-Jacobi-Bellman (HJB) equation:

$$\begin{cases} \frac{\partial V}{\partial t}(t,x) + \sup_{\alpha \in A}\{\Psi^\alpha(t,x) + \mathcal{L}^\alpha V(t,x)\} = 0, & (t,x) \in (0,T) \times \mathbb{R}^n, \\ V(T,x) = \Phi(x), \quad x \in \mathbb{R}^n. \end{cases}$$

Proof. Sketched above. □

The theorem states that, considering $V \in C^{1,2}$, then from the optimality of V and existence of α^* follow that V solves the HJB equation. In other words, this is a necessary condition. A natural question is, that if we have found V that solves the HJB equation, does this also imply the optimality of V and the existence of α^*? The answer is yes! This ”reverse-implication” is something that is proven via the so-called verification theorem.

Theorem 5.9. *(Verification Theorem).* Let $H \in C^{1,2}$ solve the HJB equation, g be an admissible strategy,and for each pair (t, x) let

$$\sup_{\alpha} \{\Psi^{\alpha}(t, x) + \mathcal{L}^{\alpha} H(t, x)\} = \Psi^{g}(t, x) + \mathcal{L}^{g} H(t, x).$$

Then $H \equiv V$ is the value function for the stochastic control problem and the optimal control $\alpha^* \equiv g$ exists.

Proof. We first show that $H \geq V$. We begin by choosing arbitrary $\alpha \in A$ and fix (t, x). By Itô's formula we have

$$H(T, X_T^{\alpha}) = H(t, x) + \int_t^T \left(\frac{\partial H}{\partial t} + \mathcal{L}^{\alpha} H \right) \mathrm{d}s + \int_t^T \dots \mathrm{d}W_s.$$

From the assumption that H solves HJB, we get that at arbitrary (s, X_s^{α}),

$$\frac{\partial H}{\partial t} + \mathcal{L}^{\alpha} H + \Psi^{\alpha} \leq 0 \Leftrightarrow \frac{\partial H}{\partial t} + \mathcal{L}^{\alpha} H \leq -\Psi^{\alpha}.$$

Using the boundary bondition $H(T, X_T^{\alpha}) = \Psi(X_T^{\alpha})$, we arrive at

$$\Phi(X_T^{\alpha}) \leq H(t, x) + \int_t^T -\Psi(s, X_s^{\alpha}) \, \mathrm{d}s + \int_t^T \dots \mathrm{d}W_s.$$

Taking expectation, we get

$$H(t, x) \geq E_{t,x} \left[\int_t^T \Psi^{\alpha}(s, X_s^{\alpha}) \, \mathrm{d}s + \Phi(X_T^{\alpha}) \right] = J^{\alpha}(t, x).$$

Finally, supremum provides us with

$$H(t, x) \geq \sup_{\alpha} J^{\alpha} = V(t, x).$$

Next, we finish the proof by showing that $H \leq V$. Instead of picking an arbitrary α, we let $\alpha = g$. Since

$$\frac{\partial H}{\partial t} + \mathcal{L}^{g} H + \Psi^{g} = 0,$$

using Itô's formula exactly as before we get

$$H(t, x) = \mathbb{E}_{t,x} \left[\int_t^T \Psi^{g}(s, X_s^{g}) \, \mathrm{d}s + \Phi(X_T^{g}) \right] = J^{g}(t, x).$$

Recalling how we define V to be the supremum over the payoff functions, we get

$$V(t, x) \geq J^{g}(t, x) = H(t, x),$$

from which the identity $V \equiv H$ follows. Moreover, g is the optimal control. $\square$

Remark 5.10. In applications and examples, we often denote $\hat{V} = H$ to highlight that H is used as a candidate for the "true" value function V.

Moreover, the following modifications are useful to state out loud.

Corollary 5.11. *(Hamilton-Jacobi-Bellman Equation for a Minimizing Problem).* Consider the assumptions in Theorem 5.8, but instead let

$$V(t,x) = \inf_{\alpha \in A} J^\alpha(t,x).$$

Then V solves

$$\begin{cases} \frac{\partial V}{\partial t}(t,x) + \inf_{\alpha \in A} \{\Psi^\alpha(t,x) + \mathcal{L}^\alpha V(t,x)\} = 0, & (t,x) \in (0,T) \times \mathbb{R}^n, \\ V(T,x) = \Phi(x), \quad x \in \mathbb{R}^n. \end{cases}$$

Remark 5.12. Let

$$\tau = \min \{\inf \{s \geq t | X_s^\alpha \in \Gamma_T\}, T\}.$$

Then the HJB generalizes to the indefinite time horizon version with τ as the hitting time of Γ_T.

Before showing an application to the HJB, we close by stating short instructions to solving an optimal control problem.

1. Write down the corresponding HJB equation.
2. Fix (t,x), and by single-variable optimization (finding where the differential attains zero), find where
$$\sup_{\alpha \in A} \{\Psi^\alpha(t,x) + \mathcal{L}^\alpha V(t,x)\}$$
is attained. This is then our best candidate to be α^*.
3. Plug the candidate α^* to the HJB equiation and solve the resulting PDE (often by a suitable guess/Ansatz).
4. Use the verification theorem to finish the problem (often by proving a corresponding version on your own).

5.3 Merton's Allocation Problem and Other Applications in Finance

In the previous section, we discussed the connection between HJB equation and stochastic optimal control problems. Now, we focus on financial applications.

Example 5.1. *(Merton's Allocation Problem).* In this example, we consider the so-called Black-Scholes world (see Theorem 3.16). Namely, we consider a risky asset S_t with dynamics

$$\mathrm{d}S_t = \mu S_t \, \mathrm{d}t + \sigma S_t \, \mathrm{d}W_t,$$

and a risk-free asset B_t with dynamics

$$\mathrm{d}B_t = r B_t \, \mathrm{d}t,$$

where $\mu > r > 0$[3]. We assume that an investor has wealth $X_t := X_t^\alpha$, of which they invest a fraction α_t in S_t and the remaining $(1-\alpha_t)$ in B_t. Assuming that $\alpha_t \in L^2$, we get the wealth dynamics

$$\mathrm{d}X_t^\alpha = \frac{\alpha_t X_t^\alpha}{S_t} \, \mathrm{d}S_t + \frac{(1-\alpha_t) X_t^\alpha}{B_t} \, \mathrm{d}B_t = \left(\mu\alpha_t + r(1-\alpha_t)\right) X_t^\alpha \, \mathrm{d}t + \sigma \alpha_t X_t^\alpha \, \mathrm{d}W_t.$$

In this example, our goal is to maximize, for some increasing and concave utility function Φ, the investors' expected utility at time T

$$V(t,x) = \sup_\alpha \mathbb{E}_{t,x}\left[\Phi(X_T^\alpha)\right].$$

More specifically, we are looking for a Markovian control $\alpha_t \geq 0$, that is, no borrowing of assets are allowed.

We solve for this problem using the step-by-step ingredients from the previous section.

1. First, we write down the corresponding HJB-equation (by noting that in the expected utility, there is no running payoff or discounting):

$$\begin{cases} \frac{\partial V}{\partial t} + \sup_\alpha \left\{\mathcal{L}^\alpha V + 0\right\} = 0, \\ V(T,x) = \Phi(x), \end{cases}$$

   where $\mathcal{L}^\alpha V = \left(\mu\alpha + r(1-\alpha)\right) x V_x + \frac{1}{2}\sigma^2\alpha^2 x^2 V_{xx}$. To solve for this equation, we consider an Ansatz $\hat{V}(t,x) = \lambda(t)\Phi(x)$ with $\lambda(T) = 1$, and $\Phi(x) = x^\lambda$ for some $\lambda \in (0,1)$[4].

2. Next, we solve for the optimal control α at which the supremum is attained.

   For this, we note that

$$\hat{V}_t = \lambda_t \Phi, \quad \hat{V}_x = \lambda \Phi_x, \quad \hat{V}_{xx} = \lambda \Phi_{xx}.$$

   That is, from the generator derived in step 1. and using the Ansatz we arrive with

$$\sup_\alpha \left\{\mathcal{L}^\alpha \hat{V}(t,x)\right\} = \sup_\alpha \left\{\left(\mu\alpha + r(1-\alpha)\right) \lambda^\gamma x^\gamma + \frac{1}{2}\sigma^2\alpha^2\gamma(\gamma-1)\lambda x^\gamma\right\}.$$

[3] This is a natural assumption, as the volatility of S_t need to be balanced against higher drift to avoid a trivial problem.

[4] This is often called the power utility function.

By First order condition, the supremum is attained at

$$\alpha^* := \frac{\gamma\lambda(\mu - r)x^\gamma}{-\gamma(\gamma-1)\sigma^2\lambda x^\lambda} = \frac{\mu - r}{(1-\gamma)\sigma^2}.$$

We note that α^* is a constant, i.e. doesn't depend on x. Moreover, as desired, it is non-negative.

3. Next, we plug our α^* in to the HJB equation and solve the following PDE:

$$\lambda_t x^\gamma + \lambda x^\gamma \left(\gamma((\mu - r)\alpha^* + r) + \frac{1}{2}\gamma(\gamma - 1)\sigma^2(\alpha^*)^2\right) = 0.$$

Denoting

$$k := \gamma((\mu - r)\alpha^* + r) + \frac{1}{2}\gamma(\gamma - 1)\sigma^2(\alpha^*)^2,$$

we arrive with an ODE

$$\begin{cases} \lambda_t = -k\lambda, \\ \lambda(T) = 1, \end{cases}$$

which is solved by $\lambda(t) = \exp\{k(T-t)\}$. Putting our so-far argumentation together, we have now arrived with

$$\hat{V}(t,x) = \exp\{k(T-t)\}x^\gamma.$$

4. Finally, the verification theorem 5.9 states that indeed $\hat{V} \equiv V$ and that α^* is the optimal control.

Remark 5.13. The following remarks are important additions and clarifications to the example:

(i) If $\gamma = 1$, then $\Phi(x) = x$, i.e. risk neutral utility function. With such utility function, we would end up with $\alpha^* = 1$.

(ii) Naturally, another constraint on short-selling, $\alpha(t) \leq 1$, would naturally end up with

$$\alpha^* = \frac{\mu - r}{(1-\gamma)\sigma^2} \wedge 1.$$

(iii) Another possible utility function would be to consider the so-called Kelly criterion $\Phi(x) = \log x$ with an Ansatz $\hat{V}(t,x) = \Phi(x) + \lambda(t)$ with $\lambda(T) = 0$. By similar methods as above, we end up with

$$\alpha^* = \frac{\mu - r}{\sigma^2}$$

and

$$\hat{V}(t,x) = \log x + k(T-t).$$

Here, the optimal control would also be easy to solve with Itô's formula.

Example 5.2. *(Portfolio Optimization with Two Assets).* For $i \in \{1, 2\}$, consider two assets S_i with dynamics

$$\mathrm{d}S_i = \mu_i S_i \,\mathrm{d}t + \sigma_i S_i \,\mathrm{d}W_t.$$

Then, an optimization problem

$$\sup_{\alpha} \mathbb{E}_{t,x} \left[\Phi(X_T^{\alpha}) \right],$$

with α being the share invested in S_1 and $\Phi(x) = x^{\gamma}$ for $\gamma \in (0, 1)$, can be solved entirely parallel to the previous example.

The following example motivates a more general extension to HJB, namely regarding discounting.

Example 5.3. *(Stochastic Control with Discounting).* In this example, we solve

$$V(t, x) = \sup_{\alpha} \mathbb{E}_{t,x} \left[e^{-\beta(T-t)} \Phi(X_T^{\alpha}) \right]$$

for some $\beta > 0$, where X follows

$$\mathrm{d}X_s^{\alpha} = \beta X_s^{\alpha} \,\mathrm{d}s + \alpha_s X_s^{\alpha} \,\mathrm{d}W_s.$$

We begin by setting up a dummy function

$$u(t, x) = \sup_{\alpha} \mathbb{E}_{t,x} \left[\Phi(X_T^{\alpha}) \right].$$

Now

$$V(t, x) = e^{-\beta(T-t)} u(t, x)$$

and

$$\begin{cases} u_t + \sup_{\alpha}\{\beta x u_x + \frac{1}{2}\alpha^2 x^2 u_{xx}\} = 0, \\ u(T, x) = \Phi(x). \end{cases}$$

Alternatively, we could have directly plugged in

$$u(t, x) = e^{\beta(T-t)} V(t, x)$$

to gain

$$\begin{cases} V_t + \sup_{\alpha}\{\beta x V_x + \frac{1}{2}\alpha^2 x^2 V_{xx}\} - \beta V = 0, \\ V(T, x) = \Phi(x). \end{cases}$$

These could then be solved very similarly to our previous examples.

Generalizing the previous example, we end up with the following corollary.

Corollary 5.14. *(Hamilton-Jacobi-Bellman Equation with Discounting).* Let

$$V(t,x) = \sup_{\alpha \in A} \mathbb{E}_{t,x}\Big[\int_t^T e^{-\int_t^s \beta_u \,\mathrm{d}u} \Psi_s^\alpha \,\mathrm{d}s + e^{-\int_t^T \beta_s \,\mathrm{d}s} \Phi(X_T^\alpha)\Big],$$

with $t < T$ and where $\beta_t = \beta(t, X_t^\alpha, \alpha_t)$ is the discounting process. Then V solves, omitting all arguments[5] but the terminal condition,

$$\begin{cases} V_t + \sup_{\alpha \in a} \{\mathcal{L}^\alpha V + \Psi^\alpha - \beta V\} = 0, \\ V(T,x) = \Phi(x). \end{cases}$$

Proof. Omitted (follows the suite of Example 5.3. □

Discounting gives a motivation for infinite time-horizon problems: without discounting, such problems would be reduced to trivialities.

Corollary 5.15. *(Hamilton-Jacobi-Bellman Equation with Infinite Time-Horizon).* Let $\beta > 0$ be a constant and consider the optimization problem

$$V(x) = \sup_{\alpha \in A} \mathbb{E}_{0,x}\Big[\int_0^\infty e^{-\beta t} \Psi(X_t^\alpha, \alpha) \,\mathrm{d}t\Big].$$

Then, Hamilton-Jacobi-Bellman equation implies

$$\sup_{\alpha \in A} \{\mathcal{L}^\alpha V + \Psi^\alpha - \beta V\} = 0.$$

Proof. Follows straightforwardly from HJB with discounting by noting that in the infinite time-horizon case, V must be time-homogeneous, resulting in x being the only argument and V_t vanishing. Moreover, since there is no boundary condition, for the respective Verification Theorem we would need

$$e^{-\beta T} V(X_T) \to_{T \to \infty} 0.$$

□

Consider the following example, which follows the corollaries nicely. In economic literature, the infinite time-horizon can be thought as an economic agent maximizing their profit throughout subsequent generations, but we might as well consider a person who never dies.

[5] And once again noting how the time-subscript in r_t could've been ambiguous with the time derivative so we omitted it, as well.

Example 5.4. *(Optimal consumption for an Immortal).* Consider the problem set-up as in Merton's problem 5.1, but set $T = \infty$ and allow the optimizer to consume (spending money not tied to assets) continuously. Let c_t be the rate of consumption, another control in addition to α_t, and Ψ be the running payoff function. We then consider the optimization problem

$$V(x) = \sup_{\alpha,c} \mathbb{E}_{0,x}\left[\int_0^\infty e^{-\beta s}\Psi(c_s, X_s^{\alpha,c})\,\mathrm{d}s\right],$$

where $\beta > 0$ and $c_t X_t^{\alpha,c}$ is the amount consumed at time t. Note how now the wealth process X depends on both controls, and how the problem attains a two-dimensional control input. Let $X_t := X_t^{\alpha,c}$, and then by Example 5.1 we have

$$\mathrm{d}X_t = (\alpha X_t\mu + (1-\alpha)X_t r)\;\mathrm{d}t + \alpha X_t\sigma\,\mathrm{d}W_t - c_t X_t\,\mathrm{d}t.$$

Followingly, the corresponding infinitesimal generator is

$$\mathcal{L}^{\alpha,c}V = (\alpha\mu + (1-\alpha)r - c)xV_x + \frac{1}{2}\sigma^2\alpha^2x^2V_{xx}.$$

Similarly to the earlier example, we again use the power utility $\Psi(x) = x^\gamma$ for $\gamma \in (0,1)$ and use the same steps we formulated earlier.

1. The HJB now becomes

$$\begin{aligned}
0 &= \sup_{\alpha,c}\{\mathcal{L}^{\alpha,c}V + \Psi(cx) - \beta V\} = 0\\
&= \sup_{\alpha,c}\left\{(\alpha\mu + (1-\alpha)r)xV_x + \frac{1}{2}\sigma^2\alpha^2x^2V_{xx} + \Psi(cx) - cxV_x\right\} - \beta V\\
&= \sup_{\alpha}\left\{(\alpha\mu + (1-\alpha)r)xV_x + \frac{1}{2}\sigma^2\alpha^2x^2V_{xx}\right\} + \sup_{c}\{\Psi(cx) - cxV_x\} - \beta V.
\end{aligned}$$

2. Now we solve for the optimal controls α^* and c^*. Let us use an Ansatz $\hat{V} = D\Psi(x)$ for some constant D. Plugging this into the HJB yields

$$\sup_{\alpha}\left\{(\alpha\mu + (1-\alpha)r)D\gamma x^\gamma + \frac{1}{2}\sigma^2\alpha^2x^2\gamma(\gamma-1)Dx^\gamma\right\} + \sup_{c}\{(cx)^\gamma - \gamma cDx^\gamma\}.$$

   From this, the First-Order conditions yield candidates for optimal controls

$$\alpha^* = \frac{\mu - r}{(1-\gamma)\sigma^2}, \quad c^* = D^{\frac{1}{\gamma-1}},$$

   which are both non-negative.

3. Plugging α^* and c^* into the PDE yields

$$Dx^\gamma\left(\gamma(\alpha^*\mu + (1-\alpha^*)r) + \frac{1}{2}\sigma^2(\alpha^*)^2\gamma(\gamma-1)\right) + x^\gamma\left((c^*)^\gamma - \gamma c^*D\right) - \beta Dx^\gamma = 0.$$

Denoting

$$k := \gamma(\alpha^*\mu + (1-\alpha^*)r) + \frac{1}{2}\sigma^2(\alpha^*)^2\gamma(\gamma-1),$$

this is equivalent with

$$(k-\beta)D + (1-\gamma)D^{\frac{\gamma}{\gamma-1}} = 0,$$

which is true when $D = \left(\frac{1-\gamma}{\beta-k}\right)^{1-\gamma}$. Thus

$$\hat{V}(x) = \left(\frac{1-\gamma}{r-k}\right)^{1-\gamma} x^\gamma.$$

4. It is to be noted that for this example the Verification theorem holds only if

$$\beta > k = \gamma r + \frac{1}{2}\sigma^2 \frac{(\mu-r)^2}{(1-\gamma)\sigma^2},$$

where we plugged in α^* to k. That is, if *beta* $> k$, then $\hat{V} \equiv V$, and α^* and c^* are the optimal controls.

The following example highlights that the optimal control might not exist.

Example 5.5. *(Optimizer Doesn't Exist).* Let

$$\mathrm{d}X_t = \alpha_t \,\mathrm{d}t + \,\mathrm{d}W_t$$

and

$$V(t,x) = \inf_{\alpha \in A} \mathbb{E}_{t,x}[X_T^2]$$

with $A = \mathbb{R}$, that is, consider a problem where the goal is to bring X_T as close to 0 as possible.

Consider $\alpha_t = -cX_t$ for some $c > 0$. Now X_t is the Ornstein-Uhlenbeck process with

$$X_T = e^{-c(T-t)} + \int_0^T e^{-c(T-s)} \,\mathrm{d}W_s.$$

(To see this, use Itô's formula with transformation $e^{-c(T-t)}$). By Itô isometry, we have

$$\mathbb{E}^c[X_T^2] = x^2 e^{-2c(T-t)} + \int_t^T e^{-2c(T-s)} \,\mathrm{d}s = \left(x^2 - \frac{1}{2c}\right) e^{-2c(T-t)} + \frac{1}{2c}.$$

Clearly, by design, $V(t,x) \geq 0$, and

$$V(t,x) \leq \inf_c \left\{ \mathbb{E}^c[X_T^2] \right\}.$$

Then

$$0 \le V(t,x) \le \inf_c \left\{ \left(x^2 - \frac{1}{2c} \right) e^{-2c(T-t)} + \frac{1}{2c} \right\}$$
$$\le \lim_{c\to\infty} \left(x^2 - \frac{1}{2c} \right) e^{-2c(T-t)} + \frac{1}{2c} = 0.$$

We have now deduced that $V(t,x) \equiv 0$. That is, if an optimal control would exist.

Indeed, assume that any α^* exists and let $X_t^* := X(t,\alpha^*)$. Then, by Itô's formula, we have

$$\mathrm{d}(X_t^*)^2 = 2X_t^*\alpha^*\,\mathrm{d}t + 2X_t^*\,\mathrm{d}W_t + \,\mathrm{d}t.$$

Then, using our previously found knowledge on V, we have

$$0 = V(t,x) = \mathbb{E}(\left[X_T^*)^2\right] = x^2 + \mathbb{E}\left[\int_t^T 2X_s^*\alpha^* + 1\,\mathrm{d}s\right] + \mathbb{E}\left[\int_t^T 2X_s^*\,\mathrm{d}W_s\right].$$

Noting that the last term is zero and taking the limit $t \to T$, by Fatou's lemma we have

$$-x^2 = \liminf_{t\to T} \mathbb{E}\left[\int_t^T 2X_s^*\alpha^* + 1\,\mathrm{d}s\right]$$
$$\ge \mathbb{E}\left[\liminf_{t\to T} \int_t^T 2X_s^*\alpha^* + 1\,\mathrm{d}s\right] = 0.$$

We have now deduced that for all $x \in \mathbb{R}$, $0 \le -x^2$, which is false, and thus α^* doesn't exist.

Alternatively, the same conclusion could be reached by writing down the corresponding HJB equation:

$$\begin{cases} V_t + \frac{1}{2}V_{xx} + \inf_{\alpha\in\mathbb{R}}\{\alpha V_x\} = 0, \\ V(T,x) = X^2. \end{cases}$$

Here, if $V_x = 0$, α can not be defined from the equation. On the other hand, if $V_x \neq 0$, then $\alpha = \pm\infty$. This implies that from a control perspective, α^* is unottainable.

Another problem a controller might face is that when α^* is not unique. Such problems are easy to come up with, and while not possibly affecting the profit of the controller, the non-uniqueness makes the mathematical problem ill-defined. For now, we cast these problems aside, and in the next Chapter, we consider a special type of stochastic control, namely, problems where the control is a stopping time.

5.4 Exercises for Chapter 5

Hamilton-Jacobi-Bellman and related equations

501. Write down the HJB equation for

$$V(0,x) = \sup_{\alpha\in A} \mathbb{E}_{0,x}[e^{-rT}g(X_T^\alpha)],$$

where

$$\mathrm{d}X_t^\alpha = rX_t^\alpha \,\mathrm{d}t + \alpha_t X_t^\alpha \,\mathrm{d}W_t,$$

with $X_0 = x$. *(Note: this can be interpreted as the pricing equation for an uncertain volatility model with constant interest rate r. The equation is called Black–Scholes–Barenblatt equation.)*

502. Write down the HJB equation for the following problem: consider

$$V(0,x) = \inf_{\alpha\in A} \mathbb{E}_{0,x}[X_T^2],$$

where

$$\mathrm{d}X_t^\alpha = \alpha_t \,\mathrm{d}t + \,\mathrm{d}W_t$$

with $X_0 = x$. *(Note: the goal in this problem is to bring the state process as close as possible to zero at the terminal time T. However, as defined above, there is no cost of actually controlling the system. This is a simplest example where there is no attainable optimizer.)*

503. A Bernoulli equation is an ODE of the form

$$x_t(t) + A(t)x(t) + B(t)x(t)^\alpha = 0,$$

where A, B are are deterministic functions of time and α is a constant. If $\alpha = 1$ then the equation is linear and easy to solve. Consider now that $\alpha \neq 1$ and introduce

$$y(t) = x(t)^{1-\alpha}.$$

Show that y satisfies the linear equation

$$y_t(t) + (1-\alpha)A(t)y(t) + (1-\alpha)B(t) = 0.$$

504. Sometimes we want to optimize the expected so-called exponential utility criterion:

$$V(t,x) = \mathbb{E}_{t,x}[\exp\{\int_t^T \Psi(s, X_s^\alpha, \alpha_s)\,\mathrm{d}s + \Phi(X_T^\alpha)\}].$$

show that the HJB equation for the expected exponential utility criterion is given by:

$$\begin{cases} \frac{\partial V}{\partial t}(t,x) + \sup_\alpha\{V(t,x)\Psi(t,x,\alpha) + \quad \mathcal{L}^\alpha V(t,x)\} = 0, \\ \qquad\qquad\qquad\qquad\qquad\qquad V(T,x) = e^{\Phi(x)}. \end{cases}$$

Stochastic Control and Applications

505. Consider the Merton's optimal consumption problem. Now instead of allowing the investor to be immortal, we let $T < \infty$ be the terminal time of its consumption. Solve the control problem

$$V(t,x) = \sup_{\alpha,c} \mathbb{E}_{t,x}[\int_t^T e^{-\beta t}\Psi(c_t)].$$

where $\beta > 0, \Phi(x) = x^\gamma$, α_t, c_t are as usual: the proportion invested in the risky asset and the consumption rate at time t. Use the ansatz

$$u(t,x) = g(t)x^\gamma.$$

You might need to use the conclusion from Exercise 503. *(Hint: in this problem the terminal condition is 0 since there's no terminal payoff.)*

506. As we discussed, problems like Exercise 505 can be generalized to stopping times. Consider

$$\tau = \min\{\inf\{t : X_t = 0\}, T\}$$

instead of T in the previous exercise. *(Hint: the result should not change.)*

507. Solve the problem

$$\sup_{u\in\mathbb{R}} \mathbb{E}[\int_0^T -u_t^2 \,\mathrm{d}t - X_T^2]$$

where $\mathrm{d}X_t = u_t\,\mathrm{d}t + \sigma X_t\,\mathrm{d}W_t$, $X_0 = x, \sigma > 0$. *(Hint: use the guess $V(t,x) = g(t)x^2$).*

508. Solve the problem

$$\sup_{u\in\mathbb{R}} \mathbb{E}_{t,x}[\int_t^T -u_s^2 \,\mathrm{d}s + X_T^2]$$

where $\mathrm{d}X_t = (u_t + \mu X_t)\,\mathrm{d}t + \sigma X_t\,\mathrm{d}W_t$, $X_0 = x, \sigma > 0$. *(Hint: use the guess $V(t,x) = \exp(f(t) + g(t)x^2)$).*

509. Solve the problem

$$\inf_u \mathbb{E}_{s,x}[\int_0^\infty e^{-\rho(s+t)}(X_t^2 + u_t^2)],$$

where $\mathrm{d}X_t = u_t\,\mathrm{d}t + \sigma X_t\,\mathrm{d}W_t$, $X_0 = x, \rho, \sigma > 0$. *(Hint: consider a two-dimensional process*

$$\mathrm{d}Y_t = \begin{bmatrix} \mathrm{d}t \\ \mathrm{d}X_t \end{bmatrix} = \begin{bmatrix} 1 \\ u_t \end{bmatrix} \mathrm{d}t + \begin{bmatrix} 0 \\ \sigma \end{bmatrix} \mathrm{d}W_t, \quad Y_0 = \begin{bmatrix} s \\ x \end{bmatrix}.$$

Consider then the guess $e^{-\rho s}(ax^2 + b)$ for some constant a, b).

Chapter 6

Optimal Stopping Theory and American Options

6.1 Optimal Stopping and Free-Boundary Problems

Motivation and Problem Formulation

Let us start by motivating using stopping times as the control in stochastic control.

Let $T > 0$ and assume that you hold a stock S_t[1]. If you needed to sell it before time T, what would be the optimal time to do so? The non-trivial aspect of the problem is clear: we can never know when a price is at its highest. A high-valued stock can go even higher, or alternatively come crashing down. More mathematically, we have

$$\{S_t = \max_{0 \leq s \leq T} S_s\} \notin \mathcal{F}_t.$$

Instead of chasing for the highest point, a wise strategy would be to decide on some boundary b and be determined to sell the stock once $S_t \geq b$ for the first time.

Such problems can be formulated as optimal stopping problems.

Definition 6.1. *(Optimal Stopping Problem).* Let $T > 0$, $\tau \in \mathcal{T}$ a stopping time in a set $\mathcal{T}$ of admissible stopping times, g a suitable gain function, and X_t be a stochastic process. An Optimal Stopping Problem is to find a value function V such that

$$V(t,x) = \sup_{\tau \in \mathcal{T}} \mathbb{E}_{t,x}\left[g(X_\tau)\right] = \mathbb{E}_{t,x}\left[g(X_{\tau^*})\right],$$

and to find the optimal stopping time τ^*.

Similarly as in regular control problems, the goal is two-fold. We wish to find V to find out how our gain structure is formed and how much we gain, and we wish to find τ^* in order to know when to stop.

[1] Which can, for example, live in the Black-Scholes world. See, once again, e.g. Theorem 3.16.

Remark 6.2. *(Simplifying Assumptions to Optimal Stopping).* Later on, we consider the following assumptions to simplify the problem:

(i) $T = \infty$, from which the time-homogeneity of V follows, and

(ii) X_t is a time-homogeneous Itô diffusion with dynamics

$$\mathrm{d}X_t = \mu(X_t)\,\mathrm{d}t + \sigma(X_t)\,\mathrm{d}W_t.$$

Before focusing in solving the problem, we highlight three important applications for optimal stopping.

1. **(Stochastic Analysis).** Consider a 1-dimensional Brownian motion W_t with $W_0 = x$, and a problem
$$V(x) = \sup_{\tau} \mathbb{E}_x \Big[e^{-r\tau} f(W_\tau) \Big]$$
for some function f. For example, let $f(x) = x^2$. Due to discounting, one assumes to stop after $|W_t|$ becomes sufficiently large.

   This example is left as an Exercise 604.

2. **(Sequential Analysis).** Now, introduce a drift to the Brownian motion so that we observe the realizations of X_t with
$$\mathrm{d}X_t = \mu\,\mathrm{d}t + \mathrm{d}W_t,$$
where μ is a Bernoulli distributed random variable with
$$\mathbb{P}(\mu = \mu_0) = p = 1 - \mathbb{P}(\mu = \mu_1).$$
Recalling basics from a statistical hypothesis testing, we could then test the hypotheses
$$H_0 : \mu = \mu_0, \qquad \text{and} \qquad H_1 : \mu = \mu_1.$$
We wish to minimize the probability of making a mistake while doing the decision as quickly as possible (waiting forever would make the problem trivial). Then, with some weights a, b and a decision function d, which attains value 0 if we accept H_0 and 1 otherwise, we could consider a problem
$$V(p) = \inf_{\tau, d} \mathbb{E}_p \Big[a\mathbb{1}_{d=0,\mu=\mu_1} + \mathbb{1}_{d=1,\mu=\mu_0} + \tau \Big].$$
Note that here the value function is a function of the a priori probability p. These types of problems are examined thoroughly in the book Peskir 2006.

3. **(Mathematical Finance).** A classical example in the realm of mathematical finance in pricing American options. For example, consider a put option, where one can exercise the option at any time before the terminal time T and get $g(S_\tau) = (k - S_\tau)^+$ for some strike price k. Naturally, here we want to maximize the payoff, so a natural problem set-up would be to consider

$$V(t, x) = \sup_{\tau < T} E^{\mathbb{Q}}_{t,x}\left[e^{-r(\tau - t)}(k - S_\tau)^+\right].$$

We will solve this problem explicitly in Section 6.2.

In the field of optimal stopping, the concept of free-boundary problem is inherently linked in solving the problem of stopping times, as finding the optimal stopping time τ^* is often interchangeable with finding some free-boundary b. Often optimal stopping problems arise from applications (as highlighted above), and thus one could summarize the process as

$$\text{Applications} \longleftrightarrow \text{Optimal Stopping} \longleftrightarrow \text{Free-Boundary Problems.}$$

Solving an Optimal Stopping Problem with Examples.

Recall that our goal was to find V and τ^*. We proceed with an example.

Example 6.1. *(When to Sell a Stock?)* Let X_t be a Geometric Brownian motion with $X_0 = x$ and

$$\mathrm{d}X_t = \mu X_t \,\mathrm{d}t + \sigma X_t \,\mathrm{d}W_t.$$

We make the further assumptions that

- Trading is not free, i.e. there exists a transaction cost $c > 0$, and that
- For the discount rate β we have that $\beta > \mu$.

The latter assumption could model a risk-free rate (see e.g. Example 5.1), tax, or inflation.

We proceed with similar steps as we did in Chapter 5 for solving the optimal control problems with Hamilton-Jacobi-Bellman equation.

1. First, we write down the value function V and characterize the continuation and stopping regions, as well as the optimal strategy. Our problem can be formulated as

$$V(x) = \sup_{\tau} \mathbb{E}_{0,x}\left[e^{-\beta\tau}(X_\tau - c)\right],$$

which is a time-homogeneous problem.

As mused in the motivation of the section, we expect some boundary $b > 0$ (and actually $b > c$, for which

- If $x \geq b$, we sell immediately and set $\tau^* = 0$, and
- If $x < b$, we wait until we hit the barrier.

That is, a suitable stopping time would be the first exit time

$$\tau^* := \inf\{t \geq 0 | X_t \geq b\}.$$

2. Second, we write down the *free-boundary problem* which characterizes the dynamics of V, this time coupled up with a slightly long discussion. Similarly as finding the Feynman-Kac and HJB, we ask ourselves: could V be characterized as a PDE?

   It turns out that optimizing for τ^* is equivalent with finding a suitable boundary b.

   If $x \geq b$, then $\tau^* = 0$.That is, one is immediately rewarded with

$$V(x) = \mathbb{E}_{0,x}\Big[e^{-\beta 0}(X_0 - c)\Big] = x - c.$$

   If, on the other hand, $x < b$, we have to work a bit harder. We note that since X is a continuous diffusion, starting from $x < b$, it has to stay below b for at least an arbitrarily short time period $t \in (0, h)$ for some sufficiently small h. That is, looking forward, at time 0 we are looking at

$$V(x) = e^{-\beta h}\mathbb{E}_{0,x}\Big[V(h)\Big] = x - c.$$

   By Feynman-Kac, V should then solve (recall that $V_t = 0$)

$$\begin{cases} \mathcal{L}V - \beta V = 0 \quad \text{ on } (0, b), \\ V(b) = b - c. \end{cases}$$

   Plugging the generator in to the PDE, noting that $V \in C^1$, and seeing how $V \to 0$ when $x \to 0$, we get the free-boundary problem

$$\begin{cases} \mu x V_x + \frac{1}{2}\sigma^2 x^2 V_{xx} - \beta V = 0 \quad \text{ on } (0, b), \\ V(b) = b - c, \\ V(x) = x - c \quad \text{ on } [b, \infty), \text{ (”Continuous fit”)} \\ V'(x) = 1 \quad \text{ at } b, \text{ (”Smooth fit”)} \\ V(0+) = 0. \end{cases}$$

   Often one omits the second and the last boundary conditions from the free-boundary problem. It is easy to see that at b, the second and the third conditions coincide, but we wished to highlight the behavior of V at the boundary b.

3. Thirdly, we use an Ansatz to solve for optimal strategies. We proceed with an Ansatz $\hat{V}(x) = x^\gamma$ for some γ. Plugging it in to the PDE yields

$$\mu\gamma x^\gamma + \frac{1}{2}\sigma^2\gamma(\gamma-1)x^\gamma - \beta x^\gamma = 0,$$

which holds if

$$\frac{1}{2}\gamma^2\sigma^2 + (\mu - \frac{1}{2}\sigma^2)\gamma - \beta = 0.$$

This quadratic equation has two unique roots, $\gamma_- < 0$ and $\gamma_+ > 0$, given by

$$\gamma_\pm = \frac{(\frac{1}{2}\sigma^2 - \mu) \pm \sqrt{(\mu - \frac{1}{2}\sigma^2) + 2\beta\sigma^2}}{\sigma^2}.$$

Thus, a general solution to the PDE is of type

$$\hat{V}(x) = C_1 x^{\gamma_+} + C_2 x^{\gamma_-}$$

for some constants C_1, C_2.

From the last boundary condition we find that $C_2 = 0$, and from the boundary condition at b we find that

$$C_1 = \frac{b-c}{b^{\gamma_+}}.$$

From these we deduce that

$$\hat{V}(x) = \begin{cases} \frac{b-c}{b^{\gamma_+}} x^{\gamma_+}, & x \in (0,b), \\ b-c, & x \geq b. \end{cases}$$

We note that b is still unsolved! For it we use the last remaining boundary condition, namely that

$$V'(x) = 1 \quad \text{at } b.$$

We call this condition the *smooth fit* condition, as it ensures that $V \in C^1$ even at b. That is, differentiating $\hat{V}$ yields that

$$b = \frac{c\gamma_+}{\gamma_+ - 1}.$$

4. Finally, we would need a verification theorem to ensure that indeed $\hat{V} \equiv V$ and that $\tau^* := \inf\{t \geq 0 | X_t \geq b\}$ is an optimal strategy. A suitable verification Theorem will be proved shortly.

(It can be also noted that the condition $\gamma_+ - 1 > 0$ implies $\beta > \mu$. This is left as an Exercise 603).

Above, we constructed a continuation region $C := (0, b)$. Considering the gain function g, it can be verified that on C we have $V > g$, i.e. not selling gives you higher value. Moreover, from above it is clear that by stopping on $D := [b, \infty)$ gives you $V = g$. That is, we could define the continuation region as

$$C = \{x \in \mathbb{R} | V(x) > g(x)\}$$

and the stopping region as

$$D = \{x \in \mathbb{R} | V(x) = g(x)\}.$$

To reiterate, we would need to prove a verification theorem for our result – at this point, the Ansatz was only a guess! While much of practical research regarding optimal stopping revolves around proving a verification theorem (and finding a candidate $\hat{V}$ for which verification is possible), we show one version of the theorem.

Theorem 6.3. *(Verification Theorem for Optimal Stopping).* Consider the problem

$$V(x) = \sup_{\tau} \mathbb{E}\left[g(X_\tau)\right]$$

for some function g and process X_t, and assume that $\hat{V} \in C^2$. Moreover, assume that $\hat{V}$ solves the free-boundary problem

$$\begin{cases} \mathcal{L}\hat{V} \leq 0, \\ \hat{V} \geq g \end{cases}$$

for all x and that there exists regions C and D such that

$$\hat{V} > g \quad \text{in } C, \quad \text{and} \quad \hat{V} = g \quad \text{in } D.$$

Then $\hat{V} \equiv V$, and $\tau^* := \inf\{t \geq 0 | X_t \in D\}$ is the optimal stopping time.

Moreover, assume that g is smooth in the neighbourhood of ∂C and the boundary ∂C is regular. Then the above free-boundary problem becomes

$$\begin{cases} \mathcal{L}\hat{V} = 0 \quad \text{in } C, \\ \hat{V} = g \quad \text{in } D, \\ \hat{V}_x = g_x \quad \text{at } \partial C \quad (\text{"smooth fit"}), \end{cases}$$

and $\hat{V} \equiv V$.

Proof. We sketch the outline of the proof for the first part of the claim.

We begin by showing that $\hat{V} \geq V$. Let $\theta = \tau \vee T$. Then By Itô's formula we have

$$\hat{V}(X_\theta) = \hat{V}(x) + \int_0^\theta \mathcal{L}\hat{V}(X_s)\,\mathrm{d}s + \int_0^\theta \ldots\,\mathrm{d}W_s.$$

From assumption we have $\mathcal{L}\hat{V} \leq 0$, and thus

$$\mathbb{E}\left[\hat{V}(X_\theta) - \hat{V}(x)\right] \leq 0 + \mathbb{E}\left[\int_0^\theta \dots \, \mathrm{d}W_s\right] = 0.$$

From this and noting that again by assumption $\hat{V} \geq g$,

$$\hat{V}(x) \geq \mathbb{E}\left[\hat{V}(X_\theta)\right] \geq \mathbb{E}\left[g(X_\theta)\right].$$

Assuming that $\tau < \infty$, we note that $\lim_{T\to\infty} \theta = \tau$, and thus we have

$$\hat{V}(x) \geq \mathbb{E}\left[g(X_\tau)\right]$$

almost surely for all $\tau < \infty$. That is,

$$\hat{V}(x) \geq \sup_\tau \mathbb{E}\left[g(X_\tau)\right] = V(x).$$

Other direction of the proof follows by letting $\tau = \tau^*$. Then $\mathcal{L}\hat{V} = 0$ in C, and

$$\mathbb{E}\left[\hat{V}(X_{\tau^*})\right] = \mathbb{E}\left[\hat{V}(x)\right] + \mathbb{E}\left[\int_0^{\tau^*} \mathcal{L}\hat{V}(X_s)\, \mathrm{d}s\right] = 0,$$

and by noting that $\hat{V} = g$ in D we finish the argument with

$$\hat{V}(x) = \mathbb{E}\left[\hat{V}(X_{\tau^*})\right] = \mathbb{E}\left[g(X_{\tau^*})\right] \leq \sup_\tau \mathbb{E}\left[g(X_\tau)\right].$$

That is, $\hat{V}(x) \equiv V(x)$ and τ^* is the optimal stopping time. □

Before venturing in to the examples with American options, we present some other formulations of optimal stopping as examples.

Example 6.2. *(Optimal Stopping with Discounting).* Consider

$$V(x) = \sup_\tau \mathbb{E}_x\left[e^{-\beta_\tau} g(X_\tau) + \int_0^\tau e^{-\beta_s}\Psi(X_s)\, \mathrm{d}s\right],$$

where the discounting process $\beta_t \in \mathbb{R}_+$ solves

$$\beta_t = \int_0^t \beta(X_s)\, \mathrm{d}s.$$

Then V solves the free-boundary problem

$$\begin{cases} \mathcal{L}V + \Psi - \beta V = 0 & \text{in } C, \\ V = g & \text{in } D, \\ V_x = g_x & \text{at } \partial C \quad \text{("smooth fit").} \end{cases}$$

Similarly as for the Hamilton-Jacobi-Bellman equation, a minimizing problem is also possible. Such problems often occur when minimizing a cost of some process.

Example 6.3. Consider the problem

$$V(x) = \inf_{\tau} \mathbb{E}\left[g(X_\tau)\right].$$

Solving for such problem is entirely parallel to the previous examples.

6.2 American Options

In this last section, we consider examples with American options.

Example 6.4. *(Perpetual American Put Option).* Consider an American Put option with a strike price $k > 0$, perpetual contract (i.e. it never expires, i.e. $T = \infty$), and with an underlying asset X_t with dynamics

$$\mathrm{d}X_t = (r - \delta)X_t \,\mathrm{d}t + \sigma \,\mathrm{d}W_t, \quad X_0 = x,$$

where $r > 0$ is the risk-free rate and $0 \le \delta < r$ is the dividend rate. Recall that an American option can be exercised at any time $t \ge 0$[2] and the put option has a payoff

$$(k - X_t)^+$$

at time t. We proceed with similar steps as in the previous Section.

1. First, we want to write the value function and characterize the optimal strategies. Indeed, now

$$V(x) = \sup^{\tau} \mathbb{E}_x\left[e^{-r\tau}(k - X_t)^+\right].$$

   Similarly as before, we characterize the continuation region C by expecting some kind of boundary which represents our view of "good enough" price to stop with. To understand the boundary, we note that for the initial stock price x, we have that

   - When $x \ge k$, we should wait, as the option is immediately worthless (note that discounting zero is still zero). This suggests $[k, \infty) \subset C$.
   - When $x < k$, we have $(k - X_t)^+ > 0$. However, we don't want to sell immediately since we could get a higher price, but we also don't wish to wait forever due to discounting.

[2]Compare this with European type options, see Theorem 3.17.

That is, we expect a threshold $b \in (0, k)$ for which

$$C = (b, \infty), \quad D = (0, b],$$

and an optimal strategy

$$\tau^* := \inf\{t \geq 0 | X_t \in D\} = \inf\{t \geq 0 | X_t \leq b\}.$$

In other words, we stop immediately after the stock price hits the boundary b.

2. Next, we write down the corresponding free-boundary problem, and assume that V solves

$$\begin{cases} (r-\delta)xV_x + \frac{1}{2}\sigma^2 x^2 V_{xx} - rV = 0, \quad x > b, \\ V(x) = k - x, \quad x \leq b \text{ (”Continuous fit”)} \\ V_x(b) = -1 \quad \text{(”Smooth fit”)}, \\ V(\infty-) = 0, \end{cases}$$

where the last boundary condition comes from the fact that if the stock price is high, the option will almost surely be worthless.

3. We make an Ansatz and solve the free-boundary problem. Since X_t is a Geometric Brownian motion, we expect an Ansatz $\hat{V}(x) = x^\gamma$ to be effective. Plug it in to the PDE to gain

$$(r-\delta)\gamma x^\gamma + \frac{1}{2}\sigma^2\gamma(\gamma-1)x^\gamma - rx^\gamma = 0,$$

which holds if

$$\frac{1}{2}\sigma^2\gamma^2 + (r - \delta - \frac{1}{2}\sigma^2)\gamma - r = 0.$$

This is a quadratic solution with two roots $\gamma_+ \geq 1$ and $\gamma_- < 0$, given by

$$\gamma_\pm = \frac{1}{\sigma^2}\left(-(r-\delta-\frac{1}{2}\sigma^2) \pm \sqrt{(r-\delta-\frac{1}{2}\sigma^2) + 2\sigma^2 r}\right).$$

Moreover, note that $\gamma_+ = 1$ when $\delta = 0$.

Very similarly as in our last example, the general solution for the candidate $\hat{V}$ is now

$$\hat{V}(x) = C_1 x^{\gamma_+} + C_2 x^{\gamma_-}$$

for some constants C_1, C_2. From the last boundary condition we see that $C_1 = 0$, and thus

$$\hat{V}(x) = C_2 x^{\gamma_-}.$$

From continuous fit condition, we get $C_2 b^{\gamma_-} = k - b$, and so

$$C_2 = \frac{k-b}{b^{\gamma_-}}.$$

Similarly, the smooth fit condition gives $C_2\gamma_- b^{\gamma_- - 1} = -1$, which yields

$$b = \frac{k\gamma_-}{\gamma_- - 1}.$$

That is, we have now reached to a conclusion that

$$\hat{V}(x) = \begin{cases} k - x, & x \in (0, b], \\ \frac{k-b}{b^{\gamma_-}} x^{\gamma_-}, & x > b. \end{cases}$$

4. By the Verification theorem, we indeed have $\hat{V} \equiv V$ and that τ^* is the optimal stopping time.

Similarly, we could solve a call option.

Example 6.5. *(Perpetual American Call Option).* Consider same underlying asset X_t as before, with

$$\mathrm{d}X_t = (r - \delta)X_t \,\mathrm{d}t + \sigma \,\mathrm{d}W_t, \quad X_0 = x.$$

Similarly, we have a strike price $k > 0$ and a payoff function $(x - k)^+$. We proceed exactly as before.

1. Now

$$V(x) = \sup_{\tau} \mathbb{E}_x \left[e^{-r\tau}(X_\tau - k)^+ \right].$$

We expect a boundary $b > k$, and we exercise the right to sell once X_t is big enough, i.e. when the price hits b. The corresponding continuation and stopping regions can be characterized as

$$C = (0, b), \quad D := [b, \infty),$$

and the optimal stopping time would then be

$$\tau^* = \inf\{t \geq 0 | X_t \in D\}.$$

2. V should solve the corresponding free-boundary problem

$$\begin{cases} (r - \delta)xV_x + \frac{1}{2}\sigma^2 x^2 V_{xx} - rV = 0, & x < b, \\ V(x) = x - k, \quad x \geq b \text{ (”Continuous fit”)} \\ V_x(b) = 1 \quad \text{(”Smooth fit”)}, \\ V(0+) = 0. \end{cases}$$

3. Using an Ansatz $\hat{V}(x) = x^\gamma$ and following similar procedure as above, we end up with

$$\hat{V}(x) = C_1 x^{\gamma_+}.$$

From the Continuous fit and Smooth fit conditions we get

$$C_1 = \frac{b-k}{b^{\gamma_+}}, \quad b = \frac{k\gamma^+}{\gamma_+ - 1},$$

latter of which is well defined if $\gamma_+ > 1$, i.e. if $\delta > 0$.

That is, if $\gamma_+ > 1$, we have arrived at

$$\hat{V}(x) = \begin{cases} \frac{b-k}{b^{\gamma_+}} x^{\gamma_+}, & x \in (0, b), \\ x - k, & x \geq b. \end{cases}$$

On the other hand, if $\gamma_+ = 1$ (or equivalently if $\delta = 0$), we let $b = \infty$, i.e. it would never be optimal to exercise the option!

4. Verification theorem asserts that indeed $\hat{V} \equiv V$, and that τ^* is optimal.

Moreover, let us examine a non-perpetual American call option.

Example 6.6. *(Stopping time for an American Call Option with finite T).* Consider the American Call option as in previous example, but this time let $T < \infty$. Before formulating the free-boundary problem, let us characterize τ^* for a moment.

Let $\tau < T$ be a stopping time (exercise time). Upon exercising, we receive $(X_\tau - k)^+$. The expected value at 0 is then

$$\mathbb{E}_x \left[e^{-r\tau}(X_\tau - k)^+ \right] \leq \mathbb{E}_x \left[\left(e^{-r\tau} X_\tau - k e^{-rT} \right)^+ \right].$$

Now recall that X_t is a Geometric Brownian motion, and thus $M_t := e^{-rt} X_t$ is a martingale. Then, for any convex function φ, we have by properties of conditional expectation, Jensen's inequality, and martingale property that

$$\mathbb{E}\left[\varphi(M_T)\right] = \mathbb{E}\left[\mathbb{E}\left[\varphi(M_T)|\mathcal{F}_\tau\right]\right] \geq \mathbb{E}\left[\varphi(\mathbb{E}\left[M_T|\mathcal{F}_\tau\right])\right] = \mathbb{E}\left[\varphi(M_\tau)\right].$$

By noting that $(x-c)^+$ is convex for all x, we have

$$\mathbb{E}_x \left[\left(e^{-r\tau} X_\tau - k e^{-rT} \right)^+ \right] \leq \mathbb{E}_x \left[\left(e^{-rT} X_\tau - k e^{-rT} \right)^+ \right] = \mathbb{E}_x \left[e^{-rT}(X_T - k)^+ \right],$$

that is, waiting until T is optimal:

$$\mathbb{E}_x \left[e^{-r\tau}(X_\tau - k)^+ \right] \leq \mathbb{E}_x \left[e^{-rT}(X_T - k)^+ \right].$$

A simple corollary is that the price of an American call option should equal the price of an European call.

As a final example, we highlight that the examples we have seen are quite remarkable in the sense that they are explicitly solvable. This is not always the case, as seen shortly.

Example 6.7. *(Value Function for American Options with finite T).* Here we wish to characterize the value function V for American options with $T < \infty$. (Note how we only discussed the existence of a stopping time in the previous example).

Let us consider a put option with payoff $(k-x)^+$[3].

Compared to the previous examples, following changes are now present:

(i) $V = V(t, x)$ is now time-dependent.

(ii) Consequently, the boundary $b = b(t)$ is now also time-dependent.

(iii) To ensure continuation, we must have $b(T) = k$.

The corresponding free-boundary problem becomes

$$\begin{cases} V_t + \mathcal{L}V - rV = 0, & \text{in } C, \\ V(t,x) > (k-x)^+ & \text{in } C, \\ V(t,x) = (k-x)^+ & \text{in } D, \\ V_x(t,x) = -1 & \text{on } x = b(t). \end{cases}$$

Explicit solution for this type of problem is then difficult to solve, and instead, we would need to study the structural properties of the problem. For example, one could study the boundary $b(t)$ and its monotonicity in efforts to show an explicit solution[4], or use numerics to characterize the boundary.

As seen in the previous example, it is fairly straightforward to come up with examples which are difficult to solve, but which are feasible and interesting from an applied perspective. Optimal stopping problems are thus an active and living subfield among the family of Stochastic Control problems, and in this field, often the most mathematically involved problems are inspired by applied foundations, where the application is characterized by optimal stopping problem, which in turn are solved as free-boundary problems:

$$\text{Applications} \longleftrightarrow \text{Optimal Stopping} \longleftrightarrow \text{Free-Boundary Problems.}$$

6.3 Exercises for Chapter 6

601. In the lectures we discussed a control problem

$$V(x) = \mathbb{E}[e^{-\beta\tau}(X_\tau - c)]$$

[3] Call option is almost equivalent.

[4] A similar problem reducing to a payoff function of American options is considered in a recent study by Ekström, Kitapbayev, Milazzo, T.: `https://arxiv.org/abs/2406.16493`.

regarding when to sell a stock, where X_t is a geometric Brownian motion. Now instead of the linear gain function, let $g(x) = h(x - c)$ and consider now

$$V(x) = \mathbb{E}[e^{-\beta\tau} h(X_\tau - c)].$$

Let $h(x) = x^\gamma, \gamma \in (0,1)$, write down the corresponding free-boundary problem satisfied by V.

602. In each of the optimal stopping problems below find the supremum V and an optimal stopping time τ^*, if any exist.

   (a) $V(x) = \sup_\tau \mathbb{E}[W_\tau^2]$,
   (b) $V(x) = \sup_\tau \mathbb{E}[W_\tau^p], \quad p > 0$,
   (c) $V(x) = \sup_\tau \mathbb{E}[\exp(-W_\tau^2)]$,
   (d) $V(x) = \sup_\tau \mathbb{E}[e^{-\rho(s+\tau)} \cosh W_\tau)], \quad \rho > 0$.

603. Consider the problem introduced in Exercise 601 with

$$V(x) = \mathbb{E}[e^{-\beta\tau}(X_\tau - c)].$$

Show that if $\mu > \beta$, then $V(x) = \infty$ and τ^* does not exist.

604. In Exercise 602 a), $V = \infty$ since waiting is not penalized. Now if we add a discounting factor then the problem becomes

$$V(x) = \sup_\tau \mathbb{E}[e^{-\beta\tau} W_\tau^2],$$

where $\beta > 0$. Solve this optimal stopping problem.

605. Solve the Exercise 601 with utility $h(x) = \log(x)$ instead.

606. Solve the optimal stopping problem

$$V(x) = \sup_\tau \mathbb{E}_x[e^{-\beta\tau} X_\tau^+],$$

where $\beta > 0$ and $X_t = x + \mu t + W_t$.

607. Solve the optimal stopping problem

$$V(x) = \sup_\tau \mathbb{E}_x[\int_0^\tau e^{-\beta t} W_t^2 dt + e^{-\beta\tau} W_\tau^2],$$

where $\beta > 1$. Find V, τ^* and some unsolvable system of b.

608. Let $0 < \beta \leq 1$ in Exercise 607. Discuss the implications of this change in parameter.

THANK YOU!

Bibliography

Bass, Richard F. (2013). *Stochastic Processes*. Cambridge University Press.

Björk, Tomas (2020). *Arbitrage theory in continuous time, Fourth edition*. Oxford University Press.

Evans, Lawrence C. (2010). *Partial Differential Equations, Second Edition*. American Mathematical Society.

Gut, Allan (2013). *Probability: A Graduate Course, Second Edition*. Springer.

Karatzas Ioannis; Shreve, Steven E. (1991). *Brownian motion and stochastic calculus, Second Edition*. Springer-Berlin.

Klimek, Maciej (2019). *Partial Differential Equations with Applications in Finance, Lecture Notes*. Department of Mathematics, Uppsala University.

Kohn, Robert (2014). *PDE for Finance, Lecture Notes*. Courant Institute of Mathematical Statistics, New York University. URL: `http://www.math.nyu.edu/faculty/kohn/pde_finance_2011.html`.

Øksendal, Bernt (2003). *Stochastic Differential Equations: An Introduction with Applications, Fifth Edition*. Springer-Verlag.

Peskir Goran; Shiryaev, Albert (2006). *Optimal Stopping and Free-Boundary Problems*. Birkhäuser Verlag.

Salamon, Dietmar A. (2020). *Measure and Integration*. ETH Zürich. URL: `http://www.math.nyu.edu/faculty/kohn/pde_finance_2011.html`.

Wang, Yuqiong (2023). *Partial Differential Equations with Applications in Finance, Lecture Notes*. Department of Mathematics, Uppsala University.

Williams, David (1991). *Probability with Martingales*. Cambridge University Press.

Index