SJÄLVSTÄNDIGA ARBETEN I MATEMATIK

MATEMATISKA INSTITUTIONEN, STOCKHOLMS UNIVERSITET

Optimal Stopping in American Put Options:
A Mathematical Analysis

av

Oscar Sjöstrand

2024 - No K23

MATEMATISKA INSTITUTIONEN, STOCKHOLMS UNIVERSITET, 106 91 STOCKHOLM

Optimal Stopping in American Put Options: A Mathematical Analysis

Oscar Sjöstrand

Självständigt arbete i matematik 15 högskolepoäng, grundnivå

Handledare: Kristoffer Lindensjö

2024

Abstract

In this thesis, we explore the mathematical results used to find a solution to the optimal stopping problem for American put options. We begin with an overview of probability theory, stochastic processes, and stochastic calculus to build a foundation for understanding the optimal stopping problem. We then apply these mathematical concepts to derive and prove a solution for the perpetual American put option. Additionally, we analyze the reasonableness of our results in the context of the characteristics of the underlying asset on which the option is based.

Abstract

I denna uppsats utforskar vi de matematiska resultat som används för att fastställa när man ska sälja en amerikansk säljoption för att maximera den förväntade avkastningen. Inledningsvis ger vi en grund i sannolikhetsteori, stokastiska processer och stokastisk differentialkalkyl för att bygga upp och introducera teorin gällande optimal stoppning. Vi kommer att använda denna teori för att presentera matematiska resonemang som löser och bevisar den optimala tidpunkten för att sälja en evig amerikansk säljoption. Därefter kommer vi att tolka och resonera kring de resultat vi har uppnått med avseende på de egenskaperna hos den underliggande tillgången optionen är baserad på.

Contents

1 Introduction

When should one stop to obtain success? Should one take the first best option or wait for a better opportunity? Mathematicians who tried to answer these questions developed the theory of optimal stopping.

When solving the optimal stopping problem for American put options, we aim to determine the optimal time to exercise the option in order to maximize the expected return. By optimizing the expected return, we ensure that the potential profit is not only substantial but also has a high probability of being realized.

Let us outline the structure of the paper. In Section 2 we will introduce the probability space, random variables and some results, such as conditional probability and conditional expectation. Section 2.1 explains what a stochastic process is, introduces filtrations and gives examples of some stochastic processes, such as Martingales and Markov processes. Section 2.1.1 introduces symmetric random walks as an intuition of Brownian motion, while Section 2.1.2 formally defines Brownian Motion. Section 3 delves into stochastic calculus, where Section 3.1 introduces Itô's integral and formula, and Section 3.2 stochastic differential equations. Section 4 introduces stopping times, optimal stopping times and the Dirichlet problem as a method of solving a optimal stopping problems. Section 5 presents American put options and how the optimal stopping problem for the perpetual American put is solved. Finally in Section 6 we analyse the results found in Section 5 and examine what happens to the solution when the parameters value changes.

2 Probability Theory

Let us begin by introducing some probability theory, to establish a foundation upon which the latter sections will rest. The definitions presented in this section take inspiration mainly from the introductory Chapter in *An Intermediate Course in Probability* by [Gut, 1998] and Chapter 1 *Stochastic Calculus for Finance II*, by [Shreve, 2004].

The basis for probability theory is the probability space. We can think of the probability space as a structure built by three components, the sample space Ω, which contains all possible events ω. From the sample space, we can find a σ-algebra $\mathcal{F}$, which equals the collection of measurable subsets of Ω. Probability spaces are also equipped by a probability measure $\mathbb{P}$, which gives us the probability that an event $A \in \mathcal{F}$ occurs or not. We can formally define the components as

Definition 2.1. Let Ω be a nonempty sample space, and $\mathcal{F}$ be a collection of subsets of Ω. $\mathcal{F}$ is said to be a *σ-algebra* if the following holds:

i) the empty set $\emptyset$ belongs to $\mathcal{F}$,

ii) if the set A belongs to $\mathcal{F}$, then A:s complement A^c will also belong to $\mathcal{F}$,

iii) if the sets $A_1, A_2, ...$ belongs to $\mathcal{F}$, then their union $\cup_{n=1}^{\infty} A_n$ belongs to $\mathcal{F}$ as well.

Definition 2.2. Consider the non-empty set Ω and let $\mathcal{F}$ be a σ-algebra of Ω. A *probability measure* $\mathbb{P}$ is a function that, for every event A in $\mathcal{F}$, $\mathbb{P}$ maps A onto $[0, 1]$.

The following must hold:

i) $\mathbb{P}(\Omega) = 1$,

ii) whenever $A_1, A_2, ..$ is a selection of disjoint sets in $\mathcal{F}$ then,

$$\mathbb{P}(\bigcup_{n=1}^{\infty} A_n) = \sum_{n=1}^{\infty} \mathbb{P}(A_n). \tag{1}$$

Definition 2.3. The triple $(\Omega, \mathcal{F}, \mathbb{P})$ is a *probability space*.

Example 2.4. Consider a symmetric six-sided die and an experiment where you toss the die once. The sample space Ω is then $\{1, 2, 3, 4, 5, 6\}$. A σ-algebra $\mathcal{F}$ is the set of all subsets of Ω, including the empty set and Ω. The maximal σ-algebra for a single die-cast consists of $2^{|\Omega|} = 2^6 = 64$ elements. First, we have the trivial cases

$$\{\emptyset\}, \{1\}, \{2\}, \{3\}, \{4\}, \{5\}, \{6\},$$

and their complements

$$\{\Omega\}, \{2, 3, 4, 5, 6\}, \{1, 3, 4, 5, 6\}, \{1, 2, 4, 5, 6\}, \{1, 2, 4, 5, 6\}, \{1, 2, 3, 5, 6\},$$
$$\{1, 2, 3, 4, 6\}, \{1, 2, 3, 4, 5\}.$$

Let us call the event when you rolled a one, $A = \{1\}$. The complement (not A) is then $A^c = \{2, 3, 4, 5, 6\}$, i.e. you did not roll a one, or you rolled either a two, three, four, five or six. We can expand this reasoning to all possible combinations of the elements in the sample space, $B = \{1, 2\}$ is the event where you roll either a one or a two, and so on. The probability measure $\mathbb{P}$ given the symmetric die, for the event A is given by $\mathbb{P}(A) = \mathbb{P}(\{1\}) = \frac{1}{6}$, and by the nature of the symmetric die it is as likely to roll a one as any other possible number. By definition of probability measure $\mathbb{P}(B) = \mathbb{P}(\{1, 2\}) = \mathbb{P}(\{1\} \cup \{2\}) = \frac{1}{6} + \frac{1}{6} = \frac{1}{3}$.

Definition 2.5. Given a probability space $(\Omega, \mathcal{F}, \mathbb{P})$, an event $A \in \mathcal{F}$ that satisfies $\mathbb{P}(A) = 1$ is said to occur *almost surely.*

The following definitions are borrowed from Chapter 2 in *Stochastic Differential Equations*, by [Øksendal, 2013]. We will use some analytical concepts, which we will not expand on in this paper. However, if you are interested, we refer to *Principles Of Mathematical Analysis*, by [Rudin, 1976], which will give you a substantial understanding of the concepts below.

Definition 2.6. Let $\mathcal{U}$ be the collection of all open subsets of Ω, given a probability space $(\Omega, \mathcal{F}, \mathbb{P})$. The σ-algebra generated by $\mathcal{U}$ is called a *Borel σ-algebra* $\mathcal{B}$ on Ω and the sets $B \in \mathcal{B}$ are called *Borel sets.*

Remark 2.7. By generated by $\mathcal{U}$, we mean that since $\mathcal{U}$ are a collection of subsets of Ω, there exists a smallest σ-algebra $\mathcal{F}_{\mathcal{U}}$ which contains $\mathcal{U}$. We say that the σ-algebra $\mathcal{F}_{\mathcal{U}}$ is generated by $\mathcal{U}$.

Definition 2.8. Consider a probability space $(\Omega, \mathcal{F}, \mathbb{P})$. We define a function $Y : \Omega \to \mathbb{R}^n$ as *$\mathcal{F}$-measurable* if the preimage of all open set $U \in \mathbb{R}^n$,

$$Y^{-1}(U) := \{\omega \in \Omega; Y(\omega) \in U\} \in \mathcal{F}.$$

Remark 2.9. If U is a Borel set, we call the Y Borel-measurable.

Definition 2.10. Consider a probability space $(\Omega, \mathcal{F}, \mathbb{P})$. A *Stochastic variable* X is a Borel-measurable real-valued function, which maps Ω onto $\mathbb{R}^n$.

Remark 2.11. We sometimes refer to a stochastic variable as a random variable without changing the meaning.

Example 2.12. Consider the die and probability space we examined in Example 2.4. We might play a game where we win a coin if we roll a one, two or three and lose a coin if we roll a four, five or six. We can now define a stochastic variable

$$X(\omega) = \begin{cases} 1 & \text{if } \omega \in \{1, 2, 3\} \\ -1 & \text{if } \omega \in \{4, 5, 6\}. \end{cases}$$

We can calculate the probability to win a coin $\mathbb{P}(X = 1) = \mathbb{P}(\{1, 2, 3\}) = \mathbb{P}(\{1\} \cup \{2\} \cup \{3\}) = \frac{1}{2}$.

In Example 2.12, we describe a random variable where all outcomes have the same probability. This is rarely the case. Consider the random variable, which is the sum of two symmetric dice. We see that it is more likely for the sum to equal 7 than 12, since there is only one way for the sum to equal 12, namely two sixes. However, there are six ways to construct seven. We call this phenomenon the stochastic variables distribution, and we will now formally define it.

Definition 2.13. Consider a discrete random variable X on $(\Omega, \mathcal{F}, \mathbb{P})$. We define the *Distribution function* of X as

$$F_X(x) = \mathbb{P}(X \leq x),$$

which is equal to

$$F_X(x) = \sum_{y \leq x} p_X(y),$$

where $p_X(x) = \mathbb{P}(X = x)$. We call $p_X(x)$ the *Probability function.* When X is continuous we get that

$$F_X(x) = \int_{-\infty}^{x} f_X(y)dy.$$

and we call $f_X(x)$ the *probability density function.*

Remark 2.14. When considering continuous random variables, we note that the function $X : \Omega \to \mathbb{R}$ is typically assumed to have a continuous distribution. However, there are cases where X does not meet the criteria for density functions, resulting in cases when the density function does not exist. We will not explore these exceptions further in this paper.

Definition 2.15. The *expected value* of a random variable X is given by

$$E[X] := \begin{cases} \sum_{k=1}^{\infty} x_k p_X(x_k) & \text{if } X \text{ is discrete,} \\ \int_{-\infty}^{\infty} x f_X(x)dx & \text{if } X \text{ is continuous.} \end{cases}$$

We say that the expected value exists if $E[|X|] < \infty$.

Definition 2.16. Consider the random variable X. We define the *variance* of X as

$$Var(X) = E[X^2] - E[X]^2.$$

The general definitions above can be explained by an example.

Example 2.17. The game described in Example 2.12 gives us the random variable X. The distribution of X is called symmetric Bernoulli or $Be(\frac{1}{2})$ since we have two outcomes $+1$ or -1 with equal probability $p = \frac{1}{2}$. The expected value is given by the weighted average of all possibilities,

$$E[X] := p(-1) + p(1) = \frac{1}{2}(-1 + 1) = 0.$$

We interpret $E[X] = 0$ as when you play many iterations of the game, the aggregate of the winnings and losses would equal 0 coins. The variance

$$Var(X) = E[X^2] - E[X]^2 = E[X^2] = \frac{1}{2}(1^2 + (-1)^2) = 1.$$

Definition 2.18. Consider a stochastic variable X defined on the probability space

$(\Omega, \mathcal{F}, \mathbb{P})$. The *distribution measure* of X is defined as the probability measure $\mu_X(B) = \mathbb{P}(X \in B)$, where B is each Borel subset of Ω.

Definition 2.19. The *conditional probability* of two events is given by the formula

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

assuming $\mathbb{P}(B) \neq 0$, which means the probability of event A given B has already happened. If

$$\mathbb{P}(A|B) = \mathbb{P}(A) \iff \mathbb{P}(A \cap B) = \mathbb{P}(A)\mathbb{P}(B),$$

we call the events *independent*. This means that the events do not affect each other's probability.

Definition 2.20. Consider two random variables X and Y defined on the probability space $(\Omega, \mathcal{F}, \mathbb{P})$. Their *joint distribution function* is given by

$$F_{X,Y}(x, y) = \mathbb{P}(X \leq x, Y \leq y),$$

for $-\infty < x, y < \infty$. When X and Y are discrete their *joint probability function* is given by

$$p_{X,Y}(x, y) = \mathbb{P}(X = x, Y = y),$$

and when X and Y continuous the *joint density function*

$$f_{X,Y}(x, y) = \frac{\partial^2 F_{X,Y}(x, y)}{\partial x \partial y}.$$

Remark 2.21. When applying the concepts defined in Definition 2.19 to stochastic variables, let us follow the reasoning given in [Gut, 1998] page 8 and 9. We say that the stochastic variables X and Y are independent if

$$F_{X,Y}(x, y) = F_X(x) \cdot F_y(y)$$

for all x and y. In the discrete case

$$p_{X,Y}(x, y) = p_X(x) \cdot p_Y(y)$$

and the continuous case

$$f_{X,y}(x,y) = f_X(x) \cdot f_Y(y).$$

The formula for conditional probability is given by in the discrete case

$$p_{X|Y}(x,y) = \frac{p_{X,Y}(x,y)}{p_Y(y)}.$$

In the continuous case, we then get

$$f_{X|Y}(x,y) = \frac{f_{X,Y}(x,y)}{f_Y(y)}.$$

Definition 2.22. Let two random variables X and Y have a joint distribution. The *conditional expectation* of Y given $X = x$ is given by

$$E[Y|X=x] = \begin{cases} \sum_y y p_{Y|X=x}(y) & \text{discrete case,} \\ \int_{-\infty}^{\infty} y f_{Y|X=x}(y) dy & \text{continuous case.} \end{cases}$$

2.1 Stochastic processes

We will now introduce the notion of stochastic processes. This section based upon Chapter 2 in [Shreve, 2004].

Definition 2.23. Given some probability space $(\Omega, \mathcal{F}, \mathbb{P})$ a *stochastic process* $(X(t))_{t\geq 0}$ is a collection of stochastic variables,

$$\{X(t) : t \in T\} \tag{2}$$

where T is some indexed set.

Remark 2.24. A couple of examples of T are $[0, 1, 2, 3, ...]$ or $[0, \infty)$.

Definition 2.25. A *filtered probability space* $(\Omega, \mathcal{F}, (\mathcal{F}_t)_{t\geq 0}, \mathbb{P})$ is a probability space equipped with an additional structure of the σ-algebras called a filtration $(\mathcal{F}_t)_{t\geq 0}$. The filtration is defined as the the sequence of σ-algebras $\mathcal{F}_s \subseteq \mathcal{F}_t \subseteq \mathcal{F}$ where $s \leq t$.

Definition 2.26. Let $(\Omega, \mathcal{F}, (\mathcal{F}_t)_{t\geq 0}, \mathbb{P})$ be a filtered probability space. A stochastic process $(X(t))_{t\geq 0}$ is called *adapted* to the filtration $(\mathcal{F}_t)_{t\geq 0}$ if for every $t \geq 0$, $X(t)$ is $\mathcal{F}_t$-measurable.

Let us put a filtration in context by an example inspired from the thread [GEdgar, 2017].

Example 2.27. Consider our earlier example with the symmetric die. If we cast the die once more, we now get the filtered probability space $(\Omega_2, \mathcal{F}, (\mathcal{F}_t)_{t\geq 0}, \mathbb{P})$, where $\Omega_2 = \Omega \times \Omega = \{1,2,3,4,5,6\} \times \{1,2,3,4,5,6\}$, leading to a event possibly looking like $(\{1\},\{2\})$ if you roll a one in the first cast, and a two in the second. Remember the σ-algebra from Example 2.4 these are the possible outcomes of the fist cast, which gives us $\mathcal{F}_1$. When the first cast is thrown, and we know the result, let us call the event A. We can generate a collection subsets of Ω which all share the element given by the first cast. The σ-algebra for the second cast is the Cartesian product between A and $\mathcal{F}_1$, i.e. $\mathcal{F}_2 = A \times \mathcal{F}_1$. Giving us a filtration.

A purpose of filtered probability spaces is to condition on the filtration. Given a stochastic process $(X(t))_{t\geq 0}$ and a filtration $(F_t)_{t\geq 0}$. An intuition of the conditional expected value $E[X(t)|\mathcal{F}_s]$ is that it is the expectation of $X(t)$ depending on the information given in $\mathcal{F}_s$, where $s \leq t$. Refer to pages 68 and 69 in [Shreve, 2004], for a more formal definition.

We will now use the conditional expected value to define two important stochastic processes, the Martingale and the Markov Process.

Definition 2.28. Given a filtered probability space $(\Omega, \mathcal{F}, (\mathcal{F}_t)_{t\geq 0}, \mathbb{P})$ a stochastic process $(X(t))_{t\geq 0}$ is called a *martingale* if

i) $X(t)$ is $\mathcal{F}_t$-measurable for $t \geq 0$, i.e. $((X(t))_{t\geq 0}$ is adapted to $(\mathcal{F}_t)_{t\geq 0})$,

ii) $E|X(t)| < \infty$ for $t \geq 0$,

iii) $E[X(t)|\mathcal{F}_s] = X(t)$ for $s \leq t$ (called a *super-martingale* if $E[X(t)|\mathcal{F}_s] \leq X(t)$ and *sub-martingale* i $E[X(t)|\mathcal{F}_s] \geq X(t)$).

An intuition for the Markov process can be a stochastic process that continuously restarts itself. Consider a function f and a random process $(X(t))_{t\geq 0}$. If we take the conditional expectation $f(X(t))$ on a filtration $(\mathcal{F}_s)_{s\leq t}$ to a time $s \leq t$, i.e. $E[f(X(t))|\mathcal{F}_s]$, it is equivalent to starting the process at time s. This gives us a formal definition.

Definition 2.29. Consider a filtered probability space $(\Omega, \mathcal{F}, (\mathcal{F}_t)_{t\geq 0}, \mathbb{P})$ and an adapted stochastic process $(X(t))_{t\geq 0}$. We say that $(X(t))_{t\geq 0}$ has the *Markov property* if, for $0 \le s \le t$ and for every non-negative, Borel-measurable function f, there exists a second Borel-measurable function g such that

$$E[f(X(t))|\mathcal{F}_s] = g(X(s)). \tag{3}$$

2.1.1 Random walks

We will make use of a famous and useful stochastic process later in this paper called Brownian Motion. To build an intuition of Brownian Motion as a stochastic process, we begin with a more straightforward concept: symmetric random walks. This section of the paper follows the reasoning given in [Shreve, 2004] Chapter 3.2.

Example 2.30. Suppose you have a fair coin you repeatedly toss. Note that the tosses are independent. A count starts at 0, and if the coin shows heads (H), you add 1, and if the coin shows tails (T), subtract 1. Let us denote the tosses as ω and

$$X = \begin{cases} 1 & \text{if} \quad \omega = H, \\ -1 & \text{if} \quad \omega = T, \end{cases} \tag{4}$$

We define $M(0) = 0$ (the start), and the process

$$M(k) = \sum_{j=1}^{k} X(j), \quad k = 1, 2, ... \tag{5}$$

The stochastic process $M(k)$ is an example of a *symmetric random walk.*

Proposition 2.31. *A symmetric random walk has independent increments.*

Proof. Consider a partition $0 = k_0 < k_1 < ... < k_m$ where all k_i for $i = 0, 1, ..., m$ are integers. The stochastic variables

$$M_{k_1} = (M_{k_1} - M_{k_0}), (M_{k_2} - M_{k_1}), (M_{k_m} - M_{k_{m-1}}) \tag{6}$$

are independent. We call $M_{k_{i+1}} - M_{k_i} = \sum_{j=k_i+1}^{k_{i+1}} X(j)$ an increment of the random walk and depends on different coin tosses. □

Corollary 2.32. *We see* $E[M_{k_{i+1}} - M_{k_i}] = 0$ *since the expected value of the random* *variable* $X(j)$ *is* 0*. The variance of the increment*

$$Var(M_{k_{i+1}} - M_{k_i}) = \sum_{j=k_i+1}^{k_{i+1}} Var(X(j)) = \sum_{j=k_i+1}^{k_{i+1}} 1 = k_{i+1} - k_i.$$

Lemma 2.33. *A symmetric random walk is a martingale.*

Proof. If we choose two non-negative integers $k < l$ and remember the definition of a martingale,

$$\begin{aligned} E[M(l)|\mathcal{F}_k] &= E[(M(l) - M(k)) + M(k)|\mathcal{F}_k] && (7) \\ &= E[M(l) - M(k)|\mathcal{F}_k] + E[M(k)|\mathcal{F}_k] && (8) \\ &= E[M(l) - M(k)|\mathcal{F}_k] + M(k) && (9) \\ &= E[M(l) - M(k)] + M(k) = M(k). && (10) \end{aligned}$$

Step (8) uses the linearity of the conditional expectation, and step (9) $M(k)$ is $\mathcal{F}_k$-measurable and only depend on the first k tosses. Step (10) follows from $M(l)-M(k)$ independence of $\mathcal{F}_k$. □

Definition 2.34. Let $M(k)$ be a stochastic process. We define quadratic variation as

$$[M, M](k) = \sum_{j=1}^{k} (M(j) - M(j-1))^2. \quad (11)$$

Lemma 2.35. *The quadratic variation of a symmetric random walk up to time* k *equals* k*.*

Proof. We see that $M(j) - M(j-1) = X(j)$, which equals either 1 or -1. This gives us that $(X(j))^2 = 1$ and the sum $\sum_{j=1}^{k} 1 = k$. □

2.2 Brownian Motion

Brownian motion was discovered by and named after the botanist Robert Brown. While looking at pollen immersed in water through a microscope, Brown discovered that the pollen kept moving in a random pattern, which later was explained to derive from the water molecules bouncing off the pollen, making it move. We will now see how to model this motion as a stochastic process.

We can think of the Brownian motion as a continuous symmetric random walk. However, the symmetric random walk defined in the previous section is discrete, which makes this argument imprecise. The following reasoning is taken from Chapter 3 in [Shreve, 2004] and gives a good intuition of a Brownian motion before giving a formal definition. We now introduce the scaled symmetric random walk

$$W^{(n)}(t) = \frac{1}{\sqrt{n}} M(nt), \tag{12}$$

where $M(nt)$ is a symmetric random walk, n is a positive integer, and nt is also a positive integer. Since t is not necessarily an integer we might get cases where nt is not an integer. Then, we choose s and u, which are the closest integers of the form $s < t < u$ and we define $W^{(n)}(t)$ as the interpolation between $\frac{1}{\sqrt{n}} M(ns)$ and $\frac{1}{\sqrt{n}} M(nu)$. As with the regular symmetric random walk, the scaled symmetric random walk has independent increments. Consider the partition $0 = t_0 < t_1 < ... < t_m$ where nt_j is and integer for all j. Then we have

$$(W^{(n)}(t_1) - W^{(n)}(t_0)), (W^{(n)}(t_2) - W^{(n)}(t_1)), ..., (W^{(n)}(t_m) - W^{(n)}(t_{m-1}))$$

are independent. When considering the walk presented in Example 2.30, we see that $W^{(100)}(0.10) - W^{(100)}(0)$ depend on the first 10 coin tosses and $W^{(100)}(0.90) - W^{(100)}(0.30)$ depends on the 60 tosses starting at toss 30. Using this reasoning, we can conclude that

$$E[W^{(n)}(t) - W^{(n)}(s)] = 0 \quad \text{and } Var(W^{(n)}(t) - W^{(n)}(s)) = t - s$$

since $W^{(n)}(t) - W^{(n)}(s)$ is the sum of $n(t-s)$ independent stochastic variables with expected value 0 and variance $\frac{1}{n}$. The intuition of the Brownian Motion is given by the limit when $n \to \infty$ of the scaled symmetric random walk. When letting n go to infinity and observing a single increment, we get by the Central Limit Theorem (see page 89 in [Shreve, 2004]) that each increment is normally distributed.

The formal definition of Brownian motion, we use is borrowed from [Øksendal, 2013] pages 13 to 15.

Definition 2.36. Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space, a stochastic process $(W(t))_{t \geq 0}$ is called a *Wiener process* or *Brownian Motion* if it satisfies the following properties

i) The increments $W(t)-W(s)$, for $s \leq t$, are normally distributed with $E[W(t)-W(s)] = 0$ and $Var(W(t)-W(s)) = t-s$,

ii) $(W(t))_{t\geq 0}$ has independent increments, i.e. if $r < s < t < u$ then the increments $W(s)-W(r)$ and $W(u)-W(t)$ are independent,

iii) $(W(t))_{t\geq 0}$ paths are almost surely continuous.

We have defined the Brownian Motion, but how can we be certain that the process exists? To show this, we introduce some results from pages 11 to 14 in [Øksendal, 2013], and page 49 and 50 in *Brownian Motion and Stochastic Calculus* by [Karatzas and Shreve, 1998], which we will use to prove the existence of the Wiener process as we formally defined it.

Definition 2.37. Consider a stochastic process $X = (X(t))_{t\geq 0}$. We define the *finite dimensional distribution* of X as the distribution measures $\mu_{t_1,t_2,...,t_k}$ on $\mathbb{R}^k$, for $k = 1, 2, ...$, satisfying

$$\mu_{t_1,...,t_k}(F_1 \times F_2 \times \dots F_k) = \mathbb{P}(X_{t_1} \in F_1, ..., X_{t_k} \in F_k),$$

where $F_1, ..., F_k$ denotes Borel sets in $\mathbb{R}$.

Let us now introduce a result that gives us a framework how to construct the Brownian motion.

Theorem 2.38. ***(Kolmogorov's extension theorem)*** *Consider a stochastic process $(X(t))_{t\geq 0}$ and consider the partition $t_1, t_2, ..., t_k \in \mathbb{R}^k_{\geq 0}$ where k is a natural number. Then, for all t_i let μ_{t_i} be distribution measures for $i = 1, 2, ..., k$ meaning that μ_{t_1} denotes the probability distribution of the element $X(t_1)$, and so on. We want the distribution measures to satisfy:*

$$\mu_{t_{\sigma(1)},...,t_{\sigma(k)}}(F_1 \times \dots \times F_k) = \mu_{t_1,...,t_k}(F_{\sigma^{-1}(1)}, \times \dots \times F_{\sigma^{-1}(k)}) \tag{13}$$

for all permutations σ on $\{1, 2, ..., k\}$ and all Borel sets F_i for $i \in \{1, 2, ..., k\}$. We also want the measures to satisfy:

$$\mu_{t_1,...,t_k}(F_1 \times \dots \times F_k) = \mu_{t_1,...,t_k,t_{k+1},...,t_{k+m}}(F_1 \times \dots \times F_k \times \mathbb{R} \times \dots \times \mathbb{R}) \tag{14}$$

for all $m \in \mathbb{N}$. If (13) and (14) are satisfied, there exist a probability space $(\Omega, \mathcal{F}, \mathbb{P})$ and a stochastic process $(X(t))_{t\geq 0}$ where $X(t) : \Omega \to \mathbb{R}$ such that

$$\mu_{t_1,\dots,t_k}(F_1 \times \dots \times F_k) = \mathbb{P}(X(t_1) \in F_1, \dots X(t_k) \in F_k) \tag{15}$$

for all $t_i \in R_{\geq 0}$ and $k \in \mathbb{N}$ and all Borel sets F_i.

Theorem 2.39. ***(Kolmogorov's continuity theorem)*** *Let $(X(t))_{t\geq 0}$ be a stochastic process. If for all $T > 0$ there exist positive constants α, β, D such that*

$$E[|X(t) - X(s)|^{\alpha}] \leq D|t - s|^{1+\beta} \quad 0 \leq s, t, \leq T.$$

Then there exists a continuous version $\tilde{X}$ of $(X(t))_{t\geq 0}$ such that $\tilde{X}$ is a continuous function for almost all possible paths.

Let us now prove that the Brownian motion as we have defined it in Definition 2.36 exists.

Proof. To show that the Wiener process defined in Definition 2.36 exists, we construct a process in accordance with Kolmogorov's Extension Theorem. The reasoning will closely follow Chapter 2.2 in [Øksendal, 2013]. We specify a family of probability measures $\mu_{t_1,\dots,t_k}$ satisfying (13) and (14) and define a probability density function

$$p(t, x, y) = \frac{1}{\sqrt{2\pi t}} \exp\{-\frac{(x - y)^2}{2t}\}, \tag{16}$$

for $x, y \in \mathbb{R}$ and $t > 0$. For $0 \leq t_1 \leq ... \leq t_k$ we define a probability measure $\mu_{t_1,\dots,t_k}$ on $\mathbb{R}^k$. From Kolmogorov's Extension Theorem, we see that

$$\mu_{t_1,\dots,t_k}(F_1 \times \dots \times F_k) = \tag{17}$$

$$\int_{F_1 \times \dots \times F_k} p(t_1, x, x_1)p(t_2 - t_1, x_1, x_2) \dots p(t_k - t_{k-1}, x_{k-1}, x_k)dx_1dx_2 \dots dx_k = \tag{18}$$

$$\mathbb{P}(W(t_1) \in F_1, \dots, W(t_k) \in F_k). \tag{19}$$

To further confirm that the Wiener process exists, we confirm each of the conditions given in Definition 2.36.

(i) Since $(W(t))_{t\geq 0}$ has normally distributed increments, we know that for all $0 \leq t_1 \leq ... \leq t_k$ the vector $Z = (W(t_1), W(t_2), ..., W(t_k)) \in \mathbb{R}^k$ has a multivariate normal distribution. This means that there exists an expected value vector $E[Z] = \underline{u} \in \mathbb{R}^k$ and a covariance matrix $\Lambda \in \mathbb{R}^{k\times k}$. Note that we observe a one-dimensional Wiener process. Using the characteristic function (see page 71 [Gut, 1998]), we can confirm that the Wiener process is normal. From page 123 in [Gut, 1998] and page 13 in [Øksendal, 2013], we get that

$$\varphi_Z(t) = \exp\{i\sum_j t_j u_j - \frac{1}{2}\sum_{j,m} t_j \lambda_{jm} t_m\} \tag{20}$$

where $t = (t_1, ..., t_k) \in \mathbb{R}^k$ and $\lambda_{jm} = E[(Z_j - u_j)(Z_m - u_m)]$ by definition of covariance. We get (20) by setting $E[W(t)] = x$ for all $t \geq 0$ and, we get $E[(W(t)-x)^2] = t, E[(W(t) - x)(W(s) - x)] = \min(s, t)$. We also get $E[(W(t) - W(s))^2] = (t - s)$ for $t \geq s$ by simple calculation.

(ii) To confirm that $(W(t))_{t\geq 0}$ has independent increments, we use the result from page 130 in [Gut, 1998], that the components in a multivariate normally distributed vector are independent if and only if they are uncorrelated. We want to prove

$$E[(W(t_i) - W(t_{i-1}))(W(t_j) - W(t_{j-1})] = 0$$

for $t_i < t_j$. From (step 1) we get that

$$\begin{aligned} &E[W(t_i)W(t_j) - W(t_{i-1})W(t_j) - W(t_i)W(t_{j-1}) + W(t_{i-1})W(t_{j-1})] \\ &= t_i - t_{i-1} - t_i + t_{i-1} = 0. \end{aligned}$$

(iii) Using Kolmogorov's Continuity Theorem, we can see, using the result, that the n :th derivative of the moment generating function (see page 64 [Gut, 1998]) $\psi_Z^{(n)}(t) = E[Z^n]$, we get that

$$E[|W(t) - W(s)|^4] = 3|t - s|^2,$$

which fulfills the continuity theorem. □

Let us see some results from pages 102 to 108 in [Shreve, 2004].

Theorem 2.40. *Let* $(W(t))_{t\geq 0}$ *be a Wiener process, then the quadratic variation* $[W,W](T) = T$ *for all* $T \geq 0$.

Proof. Consider a partition $\mathcal{P} = \{t_0, t_1, ..., t_n\}$ of $[0,T]$, and define the quadratic variation corresponding to $\mathcal{P}$ as

$$Q_{\mathcal{P}} = \sum_{j=0}^{n-1} (W(t_{j+1}) - W(t_j))^2.$$

We can prove the theorem by showing that $Q_{\mathcal{P}}$ converges to T as $\Delta t \to 0$, where Δt is the distance between t_{i+1} and t_i for all $t_i \in \mathcal{P}$. Remember that the expected value of the increments of The Wiener process equals 0, and we get,

$$E[(W(t_{j+1}) - W(t_j))^2] = Var(W(t_{j+1}) - W(t_j)) = t_{j+1} - t_j, \tag{21}$$

which in turn gives us that,

$$E[Q_{\mathcal{P}}] = \sum_{j=0}^{n-1} E[(W(t_{j+1}) - W(t_j))^2] = \sum_{j=0}^{n-1} (t_{j+1} - t_j) = T.$$

We also have that,

$$\begin{aligned} Var\left((W(t_{j+1}) - W(t_j))^2\right) &= E\left[\left((W(t_{j+1}) - W(t_j))^2 - (t_{j+1} - t_j)\right)^2\right] \\ &= E\left[(W(t_{j+1}) - W(t_j))^4\right] - 2(t_{j+1} - t_j)E\left[(W(t_{j+1}) - W(t_j))^2\right] \\ &+ (t_{j+1} - t_j)^2. \end{aligned}$$

Using the moment generating function $\psi_X(t)$ and the fact that $E[X^n] = \psi_X^{(n)}(0)$ (n:th derivative) for proof, see Chapter 3.3 in [Gut, 1998]; we get that

$$\begin{aligned} E\left[(W(t_{j+1}) - W(t_j))^4\right] &= 3(t_{j+1} - t_j)^2 \\ Var\left((W(t_{j+1}) - W(t_j))^2\right) &= 3(t_{j+1})^2 - 2(t_{j+1} - t_j)^2 + (t_{j+1} - t_j)^2 \\ &= 2(t_{j+1} - t_j)^2. \end{aligned}$$

This, in turn, gives us that,

$$
\begin{aligned}
Var(Q_{\mathcal{P}}) &= \sum_{j=0}^{n-1} Var\left((W(t_{j+1}) - W(t_j))^2\right) = \sum_{j=0}^{n-1} 2(t_{j+1} - t_j)^2 \\
&\leq \sum_{j=0}^{n-1} 2\Delta t(t_{j+1} - t_j) = 2\Delta t T.
\end{aligned}
$$

To conclude, $Var(Q_{\mathcal{P}}) \to 0$ as $\Delta t \to 0$, and we get that $\lim_{\Delta t \to 0} Q_{\mathcal{P}} = E[Q_{\mathcal{P}}] = T$. □

Theorem 2.41. *The Wiener process is a Martingale.*

The proof is similar to the proof for symmetric random walks. Let us introduce an important result from pages 107 and 108 in [Shreve, 2004].

Theorem 2.42. *Let $(\Omega, \mathcal{F}, (\mathcal{F}_t)_{t\geq 0}, \mathbb{P})$ be a filtered probability space and let $W = (W(t))_{t\geq 0}$ be a Wiener process with filtration $(\mathcal{F}_t)_{t\geq 0}$. Then, W is a Markov process.*

Proof. Consider the Markov property defined in Definition 2.29. We want to show that

$$
E[f(W(t))|\mathcal{F}_s] = g(W(s)),
$$

where f and g are defined as in Definition 2.29. We begin by setting

$$
E[f(W(t))|\mathcal{F}_s] = E[f((W(t) - W(s)) + W(s))|\mathcal{F}_s],
$$

and note that the increment $W(t) - W(s)$ is independent of $\mathcal{F}_s$, and $W(s)$ is $\mathcal{F}_s$-measurable. By the Independence lemma, Chapter 2 [Shreve, 2004], and replace $W(s)$ with a dummy variable x and get $g(x) = E[(f(W(t) - W(s)) + x)]$. By property of the Wiener process we know that $W(t) - W(s)$ is $N(0, t-s)$-distributed which gives us that

$$
g(x) = \frac{1}{\sqrt{2\pi(t-s)}} \int_{-\infty}^{\infty} f(w + x) e^{-\frac{w^2}{2(t-s)}} dw.
$$

We now replace x with $W(s)$, and the equation holds. □

3 Stochastic Calculus

Calculus is a powerful tool for studying and understanding change, and is widely used in the physical sciences, such as physics and engineering. However, the regular rules for calculus do not hold for the Wiener process. This, leads us to Stochastic calculus, which will let us apply the benefits of calculus to the stochastic processes in which, we are interested.

3.1 Itô's Integral

To build an understanding of stochastic calculus, we will study Itô's integral. We will begin with simple integrands $\phi(t)$, which we can then generalize to nonsimple integrands. The reasoning presented in this section is mostly based on Chapter 3 in [Øksendal, 2013] and Chapter 4 in [Shreve, 2004].

Consider a partition $0 < t_1 < ... < t_{n-1} < T$ of $[0, T]$ and define a simple process $\phi(t)$ to be constant on each subintervall $[t_j, t_{j+1})$. Note that the interval does not contain t_{j+1} but approaches it. The value $\phi(t_j)$ of each element in the partition t_j depends on a path ω generated by the Wiener process. We also know that $\phi(t)$ can only depend on the information given at time t, so $\phi(0)$ has to be the same for all paths. We can conclude that the first step of $\phi(t)$ for $0 \leq t < t_1$, t always start at 0. However, the second step on the interval $[t_1, t_2)$ depends on the information from $[0, t_1)$ as well as ω. We define each step of the partition

$$\int_0^t \phi(s)dW(s) = \phi(0)[W(t) - W(0)] = \phi(0)W(t), \quad \text{for} \quad 0 \leq t \leq t_1$$
$$\int_{t_1}^t \phi(s)dW(s) = \phi(t_1)[W(t) - W(t_1)], \quad \text{for} \quad t_1 \leq t \leq t_2$$

and so on. In general, if $t_k \leq t \leq t_{k+1}$, then we get the definition:

Definition 3.1. Consider the partition $0 < t_1 < ... < t_{n-1} < t$ and $t_k \leq t \leq t_{k+1}$ for some integer k, the *Ito integral* for simple integrands is of the form

$$\int_0^t \phi(s)dW(s) = \sum_{j=0}^{k-1} \phi(t_j)[W(t_{j+1}) - W(t_j)] + \phi(t_k)[W(t) - W(t_k)]. \quad (22)$$

Before we generalise the Itô integral for simple integrands to non simple, let us explore some of its properties.

Theorem 3.2. *The Itô integral for simple integrands is a martingale.*

We refer to pages 128 and 129 in [Shreve, 2004] for the proof. The theorem and proof below is inspired by page 26 in [Øksendal, 2013].

Theorem 3.3. *(Itô isometry) The Itô integral for simple integrands satisfy*

$$E[(\int_0^t \phi(s)dW(s))^2] = E\int_0^t \phi^2(s)ds \tag{23}$$

Proof. Let the $\int_0^t \phi(s)dW(s) = \sum_j e_j(W(t_{j+1}) - W(t_j))$, where e_j is some $\mathcal{F}_{t_j}$-measurable function. By ease of notation let the increment $W(t_{j+1}) - W(t_j) = \Delta W_j$. We then get

$$E[e_i e_j \Delta W_i \Delta W_j] = \begin{cases} 0 & \text{if } i \neq j, \\ E[e_j^2](t_{j+1} - t_j) & \text{if } i = j, \end{cases}$$

from the fact that $e_i e_j \Delta W_i$ is independent from ΔW_j if $i < j$. This gives us that

$$\begin{aligned} E[(\int_0^t \phi(s)dW(s))^2] &= \sum_{i,j} e_i e_j \Delta W_i \Delta W_j \\ &= \sum_j E[e_j^2](t_{j+1} - t_j) \\ &= E[\int_0^t \phi^2(s)ds]. \end{aligned}$$

□

Theorem 3.4. *Let the Itô integral $\int_0^t \phi(s)dW(s)$ be denoted as $I(t)$. The quadratic variation of $I(t)$ on the interval $[0, t]$ is*

$$[I, I](t) = \int_0^t \phi^2(s)ds \tag{24}$$

Proof. Consider the Itô integral for a subinterval $[t_j, t_{j+1}]$ where $\phi(s)$ is constant. Choose a partition such that $t_j = s_0 < s_1 < ... < s_m = t_{j+1}$ and consider the sum

$$\begin{aligned} \sum_{i=0}^{m-1} [I(s_{i+1}) - I(s_i)]^2 &= \sum_{i=0}^{m-1} [\phi(t_j)(W(s_{i+1}) - W(s_i)]^2 \\ &= \phi^2(t_j) \sum_{i=0}^{m-1} [W(s_{i+1}) - W(s_i)]^2. \end{aligned}$$

As the step sizes approach zero and m infinity, the sum approaches the quadratic variation for the Wiener process. This gives us that

$$\phi^2(t_j)(t_{j+1} - t_j) = \int_{t_j}^{t_{j+1}} \phi^2(s)ds \tag{25}$$

since $\phi(s)$ is constant on the interval $[t_j, t_{j+1}]$. We use the same reasoning for the interval $[t_k, t]$. Finally, we can add all the integrals over all the intervals to finish the proof. □

Let us now extend the definition of Itô integral for simple integrands to the general case. The intuition is to construct a sequence of simple processes $(\phi_n)_{n\geq 0}$ such that $\int_0^T \phi_n(t)dW(t) \to \int_0^T f(t)dW(t)$, for a general stochastic process f. [Øksendal, 2013] shows in depth how to choose such a stochastic process on pages 27 and 28.

Definition 3.5. To define Ito's integral for general integrands, we want to choose a $\mathcal{B} \times \mathcal{F}$-measurable function f, where $\mathcal{B}$ is the Borel σ-algebra on $[0, \infty)$, the function is also $(\mathcal{F}_t)_{t\geq 0}$ adapted (see Remark 3.15) and $E[\int_0^t f(s,\omega)^2 ds] < \infty$. Choose also a sequence of simple processes ϕ_n such that

$$E[\int_0^t (f - \phi_n)^2 ds] \to 0$$

as $n \to \infty$. Then, we can define

$$\int_0^t f(s)dW(s) = \lim_{n\to\infty} \int_0^t \phi_n(s)dW(s).$$

Note that the limit is an L^2-limit (see Remark 3.6), since $(\int_0^t \phi_n(s)dW(s))$ is a Cauchy sequence (see Remark 3.6).

Remark 3.6. Consider the probability space $(\Omega, \mathcal{F}, \mathbb{P})$ and the random variable $X : \Omega \to \mathbb{R}^n$. The L^2-norm is given by

$$||X||_2 = ||X||_{L^2(P)} = (\int_\Omega |X(\omega)|^2 dP(\omega))^{\frac{1}{2}}.$$

The L^2-space is defined by

$$L^2(P) = L^2(\Omega) = \{X : \Omega \to \mathbb{R}; ||X||_2 < \infty\}.$$

In [Øksendal, 2013] page 10, we are given that the L^2-space is a complete inner product space, with the inner product defined as

$$(X, Y)_{L^2(\mathbb{P})} := E[X \cdot Y]$$

for $X, Y \in L^2(\mathbb{P})$. The definition of complete gives us that all Cauchy sequences (see page 52 in [Rudin, 1976]) converge on L^2. We now have to show that ϕ_n is a Cauchy sequence. We can do this by following the reasoning given on page 4 in *Lecture notes in Introduction and Techniques in Financial Mathematics* [Ortiz-Latorre, 2015]. Let us use the Itô isometry to show that

$$\begin{aligned}
E[(\int_0^t \phi_n dW(s) - \int_0^t \phi_m dW(s))^2] &= E[(\int_0^t \phi_n - \phi_m dW(s))^2] \\
&= E[\int_0^t |\phi_n - \phi_m|^2 ds] \\
&= E[\int_0^t |\phi_n - f + f - \phi_m|^2 ds] \\
&\leq 2E[\int_0^t |f - \phi_n|^2 ds] + 2E[\int_0^t |f - \phi_m|^2 ds].
\end{aligned}$$

Since $(\phi_n)_{n \geq 1}$ is a sequence of simple processes approximating f as $n \to \infty$, we see that the right-hand side of the inequality converges to zero. To conclude, we see that $\int_0^t \phi_n dW(s)$ is a Cauchy sequence that converges in L^2.

Remark 3.7. Theorems 3.2, 3.3 and 3.4 also hold for the general Itô integral see page 137 in [Shreve, 2004].

To show how to use the Itô integral, we look at an example from pages 134 to 136 in [Shreve, 2004].

Example 3.8. Compute the integral $\int_0^t W(s)dW(s)$.

We begin by observing

$$\begin{aligned}
\int_0^t W(s)dW(s) &= \lim_{n\to\infty} \phi_n(s)dW(s) \\
&= \lim_{\Delta t_j \to 0} \sum_{j=0} W(t_j)[W(t_{j+1}) - W(t_j)],
\end{aligned}$$

where $\Delta t_j = t_{j+1} - t_j$. We can expand the statement inside the sum to find

$$\begin{aligned}
W(t_j)[W(t_{j+1}) - W(t_j)] &= W(t_{j+1})W(t_j) - W(t_j)^2 \\
&= W(t_{j+1})W(t_j) - \frac{1}{2}W(t_j)^2 - \frac{1}{2}W(t_j)^2 \\
&= -\frac{1}{2}W(t_{j+1}) + W(t_{j+1})W(t_j) + \frac{1}{2}W(t_j)^2 - \frac{1}{2}W(t_j)^2 + \frac{1}{2}W(t_{j+1}) \\
&= -\frac{1}{2}[W(t_{j+1}^2 - 2W(t_{j+1})W(t_j) + W(t_j)^2] + \frac{1}{2}[W(t_{j+1})^2 - W(t_j)^2] \\
&= \frac{1}{2}[W(t_{j+1})^2 - W(t_j)^2] - \frac{1}{2}[W(t_{j+1}) - W(t_j)]^2
\end{aligned}$$

We separate the terms and observe let us observe the first term. Note that

$$\begin{aligned}
&\frac{1}{2}\sum_{j}^{n-1}[W(t_{j+1})^2 - W(t_j)^2] \\
&= \frac{1}{2}[W(t_n)^2 - W(t_{n-1})^2 + W(t_{n-1})^2 - \ldots - W(t_0)^2]
\end{aligned}$$

where all of the terms cancel except for $t_0 = 0$ and $t_n = t$ which gives us that the first term is $\frac{1}{2}W(t)^2$. Observe that the second term is the quadratic variation of the Wiener Process (see Theorem 2.40), which gives us that the integral

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

We have seen the Itô integral, and we now want to expand our understanding of Stochastic calculus to include derivation. Let us start by defining a new process from page 44 in [Øksendal, 2013], which will help us understand the stochastic derivative.

Definition 3.9. Consider the probability space $(\Omega, \mathcal{F}, (\mathcal{F}_t)_{t\geq 0}, \mathbb{P})$ and the Wiener process $(W(t))_{t\geq 0}$. Let μ, σ be $\mathcal{F}_t$-adapted stochastic processes where the filtration is generated by $(W(t))_{t\geq 0}$ (see Remark 3.15). Assume

$$\mathbb{P}\left(\int_0^t \sigma(s)^2 ds < \infty; \text{for all } t > 0\right) = 1$$

and

$$\mathbb{P}\left(\int_0^t |\mu(s)| ds < \infty; \text{for all } t > 0\right) = 1.$$

We call $(X(t))_{t\geq 0}$ an *Itô process* if it is of the form

$$X(t) = X(0) + \int_0^t \mu(s)ds + \int_0^t \sigma(s)dW(s). \tag{26}$$

Calculating the integral in Example 3.8 using regular calculus, we would use the chain rule followed by the Fundamental Theorem of Calculus. However, since we do not have access to differentiation yet, we have to find another way. Using the Itô process, we will introduce the stochastic calculus version of the chain rule. We take inspiration from the theorem on page 44 in [Øksendal, 2013].

Theorem 3.10. *(Itô's fomula/Itô's Lemma) Let $g(t,x)$ be a function defined on $[0,\infty)\times\mathbb{R}$ and be twice continuously differentiable. Let $(X(t))_{t\geq 0}$ be a Itô process defined in Definition 3.9. Then $Y(t) = g(t, X(t))$ is also an Itô process where,*

$$dY(t) = \frac{\partial g}{\partial t}(t, X(t))dt + \frac{\partial g}{\partial x}(t, X(t))dX(t) + \frac{1}{2}\frac{\partial^2 g}{\partial x^2}(t, X(t))(dX(t))^2. \tag{27}$$

The following rules hold

1. $dt \cdot dt = dt \cdot dW(t) = dW(t) \cdot dt = 0,$
2. $dW(t)^2 = dW(t) \cdot dW(t) = dt.$

Remark 3.11. We can express Itô's formula in integral form as

$$g(t, X(t)) = g(0, X(0)) + \int_0^t (\frac{\partial g}{\partial t} + \mu\frac{\partial g}{\partial x} + \frac{1}{2}\sigma^2\frac{\partial^2 g}{\partial x^2})ds + \int_0^t \sigma\frac{\partial g}{\partial x}dW(s)$$

Let us show the theorem using an example inspired by Example 4.1.3 on page 45 in [Øksendal, 2013]. For a general proof, we refer to pages 46 to 48 in [Øksendal, 2013].

Example 3.12. Remember the integral from Example 3.8. Using Itô's Formula, we want to show that we can choose $X(t) = W(t)$ and $g(t,x) = \frac{1}{2}x^2$ to solve the stochastic integral $\int_0^t W(s)dW(s)$. We set $Y(t) = g(t, W(t)) = \frac{1}{2}W^2(t)$ and note that $\frac{\partial g}{\partial x} = x$ and $\frac{\partial^2 g}{\partial x^2} = 1$ by regular derivation. This gives us that

$$dY(t) = W(t)dW(t) + \frac{1}{2}dt.$$

By integrating both sides, we get

$$\frac{1}{2}W^2(t) = \int_0^t W(s)dW(s) + \frac{1}{2}ds.$$

Rearranging gives us $\int_0^t W(s)dW(s) = \frac{1}{2}W^2(t) - \frac{1}{2}t$, which is the same answer as in Example 3.8.

Remark 3.13. This is the 1-dimensional version of the theorem. We refer to page 49 in [Øksendal, 2013], for the general formula. In this paper, we will also denote the partial derivative $\frac{\partial g}{\partial x}$ as g_x.

3.2 Stochastic Differential Equations

A powerful tool that calculus gives us access to is differential equations. These are equations that include both one or more functions and their derivatives. Since we have introduced stochastic calculus in the previous section; It is time to introduce stochastic differential equations (SDE:s). This section is based on Chapter 5 in [Øksendal, 2013] and Chapter 5 in *Arbitrage Theory in Continuous Time* by [Björk, 2004].

Definition 3.14. Let $(W(t))_{t\geq 0}$ be the Wiener process, $\mu : \mathbb{R}_+ \times \mathbb{R} \to \mathbb{R}$ and $\sigma : \mathbb{R}_+ \times \mathbb{R} \to \mathbb{R}$ and a constant $x_0 \in \mathbb{R}$. We define a *Stochastic Differential Equations* (SDE) as equations of the form

$$dX(t) = \mu(t, X(t))dt + \sigma(t, X(t))dW(t), \tag{28}$$

$$X(0) = x_0. \tag{29}$$

The solution $X(t)$ is called an Itô Diffusion and is an Itô process that satisfy

$$X(t) = X(0) + \int_0^t \mu(s, X(s))ds + \int_0^t \sigma(s, X(s))dW(s). \tag{30}$$

We call the functions μ and σ the drift and diffusion.

Remark 3.15. An Itô diffusion $X = (X(t))_{t\geq 0}$ is adapted to the filtration $(\mathcal{F}_t)_{t\geq 0}$, which in turn is generated by the Wiener process. Using the definition on page 39 in [Björk, 2004], we get that for each $t \geq 0$, the stochastic variable $X(t)$ is a function depending on the Wiener trajectory (the path generated by the Wiener process) on the interval $[0, t]$.

Let us now introduce an existence theorem for the (SDE) inspired by the theorem on page 70 in [Øksendal, 2013].

Theorem 3.16. *Let $T > 0$, μ and σ be defined as in Definition 3.9 and satisfying*

$$|\mu(t,x)| + |\sigma(t,x)| \leq C(1+|x|), \tag{31}$$

for $x \in \mathbb{R}$, $t \in [0,T]$ and C is a constant. They also satisfy

$$|\mu(t,x) - \mu(t,y)| + |\sigma(t,x) - \sigma(t,y)| \leq D|x-y|, \tag{32}$$

where $x, y \in \mathbb{R}, t \in [0,T]$ and D is a constant. For some σ-algebra $\mathcal{F}_s$ generated by $W(s)$ for $s \geq 0$, the SDE

$$dX(t) = \mu(t, X(t))dt + \sigma(t, X(t))dW(t), \tag{33}$$

for $0 \leq t \leq T$ and $X(0) = x_0$ has a unique t-continuous solution $X(t)$.

The solution is equipped with the property that $X(t)$ is adapted to the filtration $(\mathcal{F}_t)_{t\geq 0}$ generated by $W(s)$ for $s \leq t$ and

$$E[\int_0^t |X(u)|^2 du] < \infty.$$

Example 3.17. An example of an SDE is the Geometric Brownian Motion (GBM), often used to model a stock price. It is of the form

$$dS(t) = rS(t)dt + \sigma S(t)dW(t). \tag{34}$$

Let r be the μ defined above. To solve the GBM we introduce a function $g(S(t)) = \ln(S(t))$. By applying Ito's formula on g, we get

$$dg(S(t)) = \frac{\partial g}{\partial t}dt + \frac{\partial g}{\partial S(t)}dS(t) + \frac{1}{2}\frac{\partial^2 g}{\partial S(t)^2}dS(t)^2, \tag{35}$$

By regular derivative rules, we get

$$\frac{\partial g}{\partial t} = 0, \quad \frac{\partial g}{\partial S(t)} = \frac{1}{S(t)}, \quad \frac{\partial^2 g}{\partial S(t)^2} = -\frac{1}{S(t)^2}. \tag{36}$$

By substituting the results into Ito's Formula, we get

$$\begin{aligned}dg(S(t)) &= 0 + \frac{1}{S(t)}(rS(t)dt + \sigma S(t)dW(t)) - \frac{1}{2}\sigma^2 dt\\&= rdt + \sigma dW(t) - \frac{1}{2}\sigma^2 dt\\&= (r - \frac{1}{2}\sigma^2)dt + \sigma dW(t).\end{aligned}$$

Let us integrate both sides,

$$\begin{aligned}\int_0^t dg(S(t)) &= \int_0^t (r - \frac{1}{2}\sigma^2)dt + \int_0^t \sigma dW(t),\\g(S(t)) - g(S(0)) &= (r - \frac{1}{2}\sigma^2)(t-0) + \sigma(W(t) - W(0)).\end{aligned}$$

By the logarithm laws, we get

$$\ln(\frac{S(t)}{S(0)}) = (r - \frac{1}{2}\sigma^2)t + \sigma W(t)$$

Finally, we get that the (GBM:s) solution is of the form

$$S(t) = S(0)\exp\left\{\sigma W(t) + (r - \frac{1}{2}\sigma^2)t\right\}. \tag{37}$$

Now we introduce some notation. Let the expected value $E_x[X(t)]$ denote that the initial value of the process $(X(t))_{t\geq 0}$ is $X(0) = x \in \mathbb{R}$, and we can write

$$E_x[f(X(t))] = E[f(X_x(t))],$$

for all bounded Borel functions f and E denote the expected value with respect to the probability law $\mathbb{P}_0$, meaning the process starts at 0.

Theorem 3.18. *(The Markov Property for Itô diffusions) Let f be a bounded Borel-measurable function from $\mathbb{R}$ to $\mathbb{R}$. Then, for $t, h \geq 0$*

$$E_x[f(X(t+h)|\mathcal{F}_t] = E_{X(t)}[f(X(h))]. \tag{38}$$

Let us refer to the proof in [Øksendal, 2013] pages 117 and 118.

Definition 3.19. Let $X(t)$ be a Itô diffusion in $\mathbb{R}$. The generator $\mathbb{L}$ of $X(t)$ is defined by

$$\mathbb{L}_X f(x) = \lim_{t \downarrow 0} \frac{E_x[f(X(t))] - f(x)}{t}; \quad x \in \mathbb{R}^n. \tag{39}$$

The set of functions $f : \mathbb{R} \to \mathbb{R}$ such that the limit exists at x is denoted by $\mathcal{D}_{\mathbb{L}}(x)$, while $\mathcal{D}_{\mathbb{L}}$ denotes the set of functions for which the limit exists for all $x \in \mathbb{R}^n$.

Theorem 3.20. *Consider the Ito diffusion $X(t)$ of*

$$dX(t) = \mu(X(t))dt + \sigma(X(t))dW(t). \tag{40}$$

If f is a twice continuously differentiable function, then f is in $\mathcal{D}_{\mathbb{L}}$, and

$$\mathbb{L}_X f(x) = \mu(x)\frac{\partial f}{\partial x} + \frac{1}{2}\sigma(x)^2\frac{\partial^2 f}{\partial x^2}. \tag{41}$$

The proof follows from Itô's Lemma, and we refer to a formal proof in [Øksendal, 2013] pages 124 and 125.

A relation between stochastic differential equations and stochastic processes is useful in solving SDE:s. To examine this relation, introduce the Feynman-Kac formula, the theorem and proof are inspired by [Björk, 2004] pages 68 and 69.

Theorem 3.21. *(Feynman-Kac formula) Assume that X solves the SDE*

$$\begin{aligned} dX(s) &= \mu(s, X(s))ds + \sigma(s, X(s))dW(s), \\ X(t) &= x, \end{aligned}$$

and consider F as a solution to the boundary value problem

$$\begin{aligned} \frac{\partial F}{\partial t}(t,x) + \mu(t,x)\frac{\partial F}{\partial x} + \frac{1}{2}\sigma^2(t,x)\frac{\partial^2 F}{\partial x^2} &= 0 \\ F(T,x) &= \Phi(x). \end{aligned}$$

Assume as well that the process $\sigma(s, X(s)\frac{\partial F}{\partial x}(s, X(s))$ is in L^2. Then, we can represent F as

$$F(t,x) = E_{t,x}[\Phi(X(T)]$$

Proof. (Sketch) Consider the infinitesimal generator

$$\mathbb{L}_X = \mu(t,x)\frac{\partial}{\partial x} + \frac{1}{2}\sigma^2(t,x)\frac{\partial^2}{\partial x^2}.$$

We can now rewrite the BVP as

$$\begin{gathered}\frac{\partial F}{\partial t}(t,x) + \mathbb{L}_X F(t,x) = 0\\ F(T,x) = \Phi(x)\end{gathered}$$

By applying Itô's formula to the process $F(s, X(s))$ we get that

$$\begin{aligned}F(T,X(T)) = F(t,X(t)) + \int_t^T \left(\frac{\partial F}{\partial t}(s,X(s)) + \mathbb{L}F(s,X(s))\right) ds\\ + \int_t^T \sigma(s,X(s))\frac{\partial F}{\partial x}(s,X(s))dW(s).\end{aligned}$$

If the process $\sigma(s, X(s))(\frac{\partial F}{\partial x})(s, X)$ is sufficiently integrable and when the expected value is taken, the stochastic integral vanishes by property of the Itô integral (see page 30 in [Øksendal, 2013]). The boundary condition $F(T, x) = \Phi(x)$ and the initial value $X(t) = x$ gives us the formula

$$F(t,x) = E_{t,x}[\Phi(X(T))],$$

which proves the theorem. □

4 Optimal Stopping Problem

The Optimal Stopping Problem studies when to stop a stochastic process to obtain success. This could be to maximize an expected gain function or when to stop searching for an employee. A classic example of this is the Secretary problem.

Example 4.1. In this scenario, you aim to hire the best secretary from n applicants. The applicants are interviewed in a random order; you can rank the current applicant relative to those previously interviewed. At the end of each interview, you must immediately decide whether to hire that applicant.

Consider a strategy where you first interview a number of the applicants, let us call this number k. We will reject all k of these applicants, and then hire the next applicant who is better than the best of the first k. This is one strategy to solve the secretary problem, and we will show this strategy in the paragraph below. There are instances where the best possible applicant is one of the k applicants, or the $k+s$ applicant is better than the k first, but worse the the $k+t$ applicant, where $s < t$. Then, the strategy does not yield the optimal result. However, it turns out that the strategy results in the highest probability of success. To solve the problem, we want to find the k that maximize the probability of success.

This solution is based upon the reasoning given in the video [Numberphile2, 2014] and is a simplification of the real solution. However, it gives a good understanding of the Optimal Stopping Problem and how one can attempt to solve it. Assume that the best secretary is as likely to be any of the n applicants. We can then conclude that the probability of the best applicant to be in a specific place is $\frac{1}{n}$. If the best candidate is in the $k+1$ place, the probability of picking the best candidate is 1 because of our strategy. What if the best candidate is in the $k+2$ place? What is the probability that we pick them? This problem is equivalent to the probability that the best candidate is not in the $k+1$ place, i.e. $1 - \frac{1}{k+1} = \frac{k}{k+1}$. If we carry on this reasoning, we find that the probability of picking the best candidate if they are in the $k+3$ place is $\frac{k}{k+2}$, $k+4$ is $\frac{k}{k+3}$, and so on. Together, the probability for the best applicant to be in a specific place $\frac{1}{n}$ and the probability of picking the best applicant given its place, yields the probability function $p(k)$, describing the probability of picking the best applicant out of the n, given our strategy of rejecting the k first. We formulate the probability function as:

$$\begin{aligned}p(k) &= \frac{1}{n}(1 + \frac{k}{k+1} + \frac{k}{k+2} ... + \frac{k}{n-1}\\&= \frac{k}{n}(\frac{1}{k} + \frac{1}{k+1} + ... + \frac{1}{n-1}).\end{aligned}$$

We ignore the k first applicants since the probability of picking them is 0. The series inside the brackets can be approximated by the integral of the function $f(x) = \frac{1}{x}$, we assume this without giving proof. This gives us an approximation

$$\tilde{p}(k) = \frac{k}{n}\int_k^n \frac{1}{x}dx.$$

Using regular calculus gives us $\tilde{p}(k) = \frac{k}{n}(\ln(n) - \ln(k))$ and by the laws of logarithm we get $\tilde{p}(k) = \frac{k}{n}\ln(\frac{n}{k})$. If we now substitute $\frac{k}{n} = x$, we get a function $g(x) = x\ln(\frac{1}{x})$ and by the laws of logarithm, $g(x) = -x\ln(x)$. If we plot $g(x)$ we see that the function increases to a maximum point, and after the max is attained, it decreases. To find the max, we differentiate using regular calculus and solve the equation where $g'(x) = 0$. We get that $g'(x) = -\ln(x) - x\frac{1}{x} = 0$, which gives us that g attains max when $x = \frac{1}{e}$. We maximize $\tilde{p}(k)$ when $k = \frac{n}{e}$ and, we have solved the problem. To summarize, our strategy described above yields the result that we should reject the $\frac{n}{e}$ first applicants. As a percentage this is $\frac{1}{e} \approx 0.37$ of the applicants. If we choose the next best applicant, the probability of success can be derived from the statement $k = \frac{n}{e}$, and results in $\frac{k}{n} = \frac{1}{e}$. To conclude if you reject the first 37% of applicants, then you have a 37% likelihood of picking the best applicant.

Let us give an outline for the structure of this section. We begin by formulating the optimal stopping problem, then show a strategy to derive a solution. Finally, we will introduce some results that will help us verify that the solution derived, is actually optimal.

Using the theory of stochastic processes and calculus, we will formalize the stopping time and some ways to solve the optimal stopping problem. To begin, we will follow the reasoning given in Chapter 1 in *Optimal Stopping and Free Boundary Problem* by [Peskir and Shiryaev, 2006].

Definition 4.2. Consider a filtered probability space $(\Omega, \mathcal{F}, (\mathcal{F}_t)_{t\geq 0}, \mathbb{P})$ and a random variable $\tau : \Omega \to [0, \infty]$, is called a stopping time if $\{\tau \leq t\} \in \mathcal{F}_t$ for all $t \geq 0$.

Example 4.3. An example of a stopping time is

$$\tau = \inf\{t \geq 0 : X(t) \in B\}. \tag{42}$$

for some predetermined set B. We see that τ is the first time the random process $(X(t))_{t\geq 0}$ enters B. This is a stopping time since we know B, and we can check for each $t \geq 0$ if $X(t) \in B$.

Example 4.4. A non-example of a stopping time is quite simple: consider

$$\tau = \sup\{t \geq 0 : X(t) \in B\}.$$

Since τ is the largest t where $X(t)$ is in B we cannot know when this is since we cannot see the future. This means that τ is not a stopping time.

To find the optimal stopping time, we will closely follow the reasoning of the Markovian approach, is given in Chapter 1.2 in [Peskir and Shiryaev, 2006].

From now on we will denote $X = (X(t))_{t\geq 0}$, and let X be a Markov process, which takes values on $E \subset \mathbb{R}$ and maps onto $\mathbb{R}$.

We define the optimal stopping problem as.

Definition 4.5. Consider a stochastic process X, and E denote all possible values of $X(t)$. Let $G : E \to \mathbb{R}$ be a Borel-measurable function satisfying that $E_x[\sup_{0\leq t\leq T} |G(X(t))|] < \infty$ for all $x \in E$. The *optimal stopping problem* is then given by,

$$V(x) = \sup_{0\leq\tau\leq T} E_x[G(X(\tau))], \tag{43}$$

for a stopping time $\tau < \infty$.

We call $V(x)$ the *value function* and G the *gain function*. To solve the optimal stopping problem we want to first evaluate V, then we will introduce a method to check if the stopping time is optimal.

The intuition of the Markovian approach uses the memoryless property of the Markov process. Consider a stopped process, we have to choose if we want to continue the process or accept the result. So we have to find a stopping set $\mathcal{S}$ such that when the process is inside of $\mathcal{S}$, the conditions for optimal stopping are fulfilled. We will follow the reasoning on pages 12 and 13 in [Peskir and Shiryaev, 2006].

Definition 4.6. We define the *stopping set* $\mathcal{S}$, and the *continuation set* $\mathcal{C}$ formally as,

$$\mathcal{S} = \{x \in E : V(x) \leq G(x)\},$$
$$\mathcal{C} = \{x \in E : V(x) > G(x)\}$$

Definition 4.7. Define the first entry time of X into $\mathcal{S}$ as

$$\tau_{\mathcal{S}} = \inf\{t \geq 0 : X(t) \in \mathcal{S}\} \tag{44}$$

Remark 4.8. We will only handle continuous functions in this paper, which gives us that $\mathcal{S}$ is closed.

An important concept for understanding the Markovian approach to solving the optimal stopping problem is the Strong Markov property. To give an intuition of the Strong Markov property, we remember the Markov Property in Definition 2.29. Instead of determining a particular time point s, we now have a random stopping time τ, in which the process can restart. We can formalise this in a definition.

Definition 4.9. Consider a stochastic process X. We can express *the strong Markov property* of X as

$$E_x[F \circ \theta_\tau | \mathcal{F}_\tau] = E_{X(\tau)}[F] \tag{45}$$

for a bounded function F and stopping time τ.

Remark 4.10. We say that $\theta : \Omega \to \Omega$ is a shift operator. The shift operator acts on a stochastic process X like

$$\theta_n \circ X(t) = X(n + t).$$

For a formal definition see page 122 in [Øksendal, 2013].

Using some of the reasoning given in [Peskir and Shiryaev, 2006] pages 77 and 78, we can examine some results from the shift operator.

Corollary 4.11. *Consider a stopping time $\sigma \leq \tau$ where τ is an entry time (the value of t where the process X enters a set). Then*

$$X(\tau) \circ \theta_\sigma = X(\tau \circ \theta_\sigma + \sigma). \tag{46}$$

This means that we can shift the process with finite stopping time. We keep the θ_σ term since the stopping time σ is a random variable. If $\sigma = n$ where $n \in \mathbb{R}$ we get that $X(\tau) \circ \theta_n = X(\tau + n)$.

Definition 4.12. Consider a strong Markov process X on a filtered probability space $(\Omega, \mathcal{F}, (\mathcal{F}_t)_{t\geq 0}, \mathbb{P})$. Let X start at x, where $x \in E$ and consider a stopping time τ. We call optimal stopping problem *discounted* when it is of the form:

$$V = \sup_\tau E[e^{-r\tau} G(X(\tau))]$$

where $r \in \mathbb{R}_+$ is the discount factor and $\tau < \infty$ a stopping time. Consider reducing the problem presented above to the original optimal stopping problem

$$V = \sup_\tau E[G(\tilde{X}(\tau))],$$

where $\tilde{X}$ is the Markov process corresponding to the "discounting" of X by the same rate as r. The infinitesimal generator $\mathbb{L}_{\tilde{X}}$ is given by, $\mathbb{L}_{\tilde{X}} = \mathbb{L}_X - rI$, where I is the identity operator.

Remark 4.13. Let us refer to page 127 in [Peskir and Shiryaev, 2006] if you want to see a more general definition of the discounted problem.

Let us now introduce an important result using the Dirichlet problem or, first boundary problem (for a formal formulation of the Dirichlet problem, see page 84 in [Peskir and Shiryaev, 2006]). We will use the Dirichlet problem together with Theorem 4.15 to create a system of equations that will help us derive a solution to the optimal stopping problem. The theorem below is inspired by the theorem given on page 130 in [Peskir and Shiryaev, 2006].

Theorem 4.14. *Consider a continuous function $G : \partial\mathcal{C} \to \mathbb{R}$, let*

$$V(x) = E_x[e^{-r\tau_\mathcal{S}} G(X(\tau_\mathcal{S}))]$$

where $\tau_\mathcal{S}$ is defined in Definition 4.7, r is the discount factor and $\partial\mathcal{C}$ is the boundary *of the continuation set $\mathcal{C}$. The function V solves the discounted Dirichlet Problem:*

$$\mathbb{L}_X V = rV \quad \text{in } \mathcal{C},$$
$$V|_{\partial\mathcal{C}} = G.$$

Proof. (Sketch) From Definition 4.12, we see that we can represent

$$E_x[e^{-r\tau} G(X(\tau_{\mathcal{S}}))] = E_x[G(\tilde{X}(\tau_{\mathcal{S}}))].$$

The infinitesimal generator $\mathbb{L}_{\tilde{X}} = \mathbb{L}_X - rI$ applied on V. By following the proof for the regular Dirichlet problem (see page 130 in [Peskir and Shiryaev, 2006]) we get that $\mathbb{L}_{\tilde{X}} V = 0$, which leads to the first equality. The second equality follows from V evaluated at the optimal stopping time. □

Finally, let us introduce a result which helps us study the value function V is defined at the optimal stopping point.

Theorem 4.15. *(Smooth fit) Assume that the stopping set is $[0, b]$ and the continuation set (b, ∞), note that b is the optimal stopping point. If V is differentiable at b then $V'(b) = G'(b)$.*

For a more general formulation and proof, see page 150 in [Peskir and Shiryaev, 2006].

Using the theorems above, we can derive a system of equations which solved will gives us a candidate for a solution to an optimal stopping problem. We introduce some results from [Peskir and Shiryaev, 2006] page 37 and page 214 in [Øksendal, 2013], that will help us prove that the solution is optimal.

Definition 4.16. Consider a measurable function $F : E \to \mathbb{R}$. We call F superharmonic if it satisfies

$$f(x) \geq E_x[f(X(\tau))],$$

for all stopping times τ and x in E.

Proposition 4.17. *A measurable function F is superharmonic if and only if the process $(F(X(t))_{t\geq 0}$ is a supermartingale.*

Proof. [$\Longleftarrow$] Consider a supermartingale $(F(X(t)))_{t\geq 0}$. By the Markov property, we get

$$E_x[F(X(t))|\mathcal{F}_s] = E_{X_s}[F(X(t-s))] \leq F(X(s))$$

for all $t \geq 0$ and we have shown the supermartingale property of $(F(X(t)))_{t\geq 0}$.

[$\Longrightarrow$] Consider the superharmonic function F, for all $x \in E$. Then, we can follow the reasoning above backwards, which proves the right direction as well. □

Theorem 4.18. *Suppose there exists an optimal stopping time τ_* such that*

$$V(x) = E_x[G(X(\tau^*))] \tag{47}$$

for all $x \in E$. Then, V is the smallest superharmonic function, which dominates the function G on E.

Proof. We want to show that $V(x)$ is a superharmonic function. Using the strong Markov property, we see that

$$\begin{aligned} E_x[V(X_\sigma)] &= E_x\left[E_{X_\sigma}[G(X(\tau^*))]\right] = E_x\left[E_x[G(X(\tau^*)) \circ \theta_\sigma|\mathcal{F}_\sigma]\right] \\ &= E_x[G(X(\tau^* \circ \theta_\sigma + \sigma))] \leq \sup_\tau E_x[G(X(\tau))] = V(x). \end{aligned}$$

Where σ is some stopping time and θ is a shift operator. Consider F a superharmonic function that dominates G on the set E. We then have

$$E_x[G(X(\tau))] \leq E_x[F(X(\tau))] \leq F(x), \tag{48}$$

for all τ and all x in E. Let us take

$$\sup_{0\leq\tau\leq T} E_x[G(X(\tau))] = V(x) \leq F(x) \quad \text{for all} \quad x \in E. \tag{49}$$

Since $V(x)$ is a superharmonic function, we have now proven the statement. □

Theorem 4.19. *Consider the optimal stopping problem*

$$V(x) = \sup_{0\leq\tau\leq T} E_x[G(X(\tau)] \tag{50}$$

and suppose that $E_x[\sup_{0\leq t\leq T} |G(X(t))|] < \infty$ *is fulfilled. Let* $\hat{V}$ *be the smallest superharmonic function that dominates* G *on* E*. Let* $\mathcal{S}$ *and* $\tau_{\mathcal{S}}$ *be defined as in Definition 4.6 and 4.7. Then*

i) If $\mathbb{P}_x(\tau_{\mathcal{S}} < \infty) = 1$ *for all* $x \in E$*, then* $\hat{V} = V$ *and* $\tau_{\mathcal{S}}$ *is optimal.*

ii) If $\mathbb{P}(\tau_{\mathcal{S}} < \infty) < 1$ *for some* $x \in E$*, then there exists no optimal stopping time.*

Let us refer to the proof in [Peskir and Shiryaev, 2006] pages 41,42 and 43.

The optimal sampling theorem, which we will borrow from the formulation on page 342 in [Shreve, 2004], is a helpful fact we will use to prove that the optimal stopping times are optimal.

Theorem 4.20. *(Optional sampling) Consider a martingale* $(X(t))_{t\geq 0}$ *and a stopping time* τ*. The stopped process* $(X(t \wedge \tau))_{t\geq 0}$ *is also a martingale. Where* $t \wedge \tau$ *is the minimum.*

5 American Put Options

An option is a financial derivative where the parties agree upon a premium that the holder (the party going long the option) will pay for the option, but not the obligation, to buy or sell an asset from or to the writer (the party going short the option), at an agreed-upon price, called the strike price, on a predetermined future date. We call the right to sell the asset in the future, for the put option, and the name put comes from the expression "put up for sale". Let us give an example of how and why you would buy (go long) a put option:

Example 5.1. Consider a stock that costs 100 dollars, and you believe that the stock will go down in value. You now buy a put option and pay a premium of 10 dollars, and you agree to an exercise date in the future and the strike price, which, in our case, is the same price as the underlying stock, i.e., 100 dollars. There are a couple of possible outcomes. If the stock's value goes up or stays the same at the exercise date, you decide not to exercise, and lose the premium. The stock price can also fall; for example if the stock's price is 80 dollars at the exercise date, you can now exercise the put (Buy a stock at 80 dollars and sell the stock at the strike price of 100 dollars). This gives a gain of 10 dollars since $100-80-10 = 10$. The transaction only becomes profitable when the price of the asset falls below the strike price with more than the premium.

The option described in example 5.1 is called a *European Put Option*, meaning you can only exercise the option at the exercise date. An *American Put Option*, in which we are interested, can be exercised at any time, including the exercise date.

To determine the return of a put option. Consider the stock price modelled by a geometric Brownian motion $S(t)$ at time t. Suppose we exercise an American put option at this time, with a strike price K. The gain from exercising the put option would be $K - S(t)$ dollars, or since we do not exercise if we lose money, the gain is $\max(0, K - S(t))$, we will denote this as $G(t) = (K - S(t))^+$ from now on. However, it is crucial to recognize that we also assume a risk by investing in the option. Consequently, we need to account for the gains we could have been achieved by investing the same amount in risk-free assets over the same period. These risk-free gains are given by the compounding interest over the same time as the option is held, i.e. from $t = 0$ to $t = t$. We can express this discount factor as e^{-rt}, where r represents the risk-free interest rate. We now have a formula for the value of an

American put,

$$e^{-rt}G(t) = e^{-rt}(K - S(t))^+. \tag{51}$$

We aim to find a stopping time τ, which maximizes $e^{-r\tau}G(\tau)$.

Remark 5.2. We will assume that all prices are arbitrage-free. Arbitrage occurs when there is an opportunity for risk-free gain, i.e. you can buy the asset or option at a price and sell at a gain directly afterwards. We also call an arbitrage-free priced asset a fair-priced asset.

5.1 Perpetual Put

A perpetual put is a conceptual derivative. However, it allows us an excellent opportunity to solve the optimal stopping problem explicitly. You can derive from the perpetual puts name that it lacks an exercise date, giving the possibility that $t \to \infty$. The optimal stopping problem for the perpetual put addresses the question of when to exercise the option to maximize the expected return. We will use the approach described in Section 4 to derive a guess for the solution and then prove that the guess is correct. The intuition is that we will derive from the method described in Section 4, a lower boundary for the underlying asset, which maximizes the expected return. If and when the asset reaches that valuation, we exercise the option.

We will assume that the expected discounted return at a stopping time gives the value of the perpetual American put option. The reasoning presented in this section follows pages 375 to 378 in [Peskir and Shiryaev, 2006] mixed with pages 349 to 356 in [Shreve, 2004].

Definition 5.3. Let τ be a stopping time and $S(t)$ be the price of an underlying asset generated by a GBM with $S(0) = x$. The value of the perpetual put is given by

$$V(x) = \sup_{\tau} E[e^{-r\tau}(K - S(\tau))^+]. \tag{52}$$

Consider the process $(S(t))_{t\geq 0} = S$ which solves the GBM presented in Example 3.17. Remember that $S(0) = x$, the interest rate $r > 0$, K is the strike price, and

$\sigma > 0$ is the volatility coefficient. From Example 3.17, we get that the solution is

$$S(t) = x\exp\{\sigma W(t) + (r - x\frac{\sigma^2}{2})t\}$$

for $t \geq 0$. Note that the process S is equipped with the strong Markov property and the infinitesimal generator

$$\mathbb{L} = rx\frac{\partial}{\partial x} + \frac{\sigma^2}{2}x^2\frac{\partial^2}{\partial x^2}.$$

To solve the optimal stopping problem for a perpetual put, our approach is to find a $b \in (0, K)$ such that the stopping time

$$\tau_b = \inf\{t \geq 0 : S(t) \leq b\} \tag{53}$$

is optimal. Note that b is then the boundary $\partial\mathcal{C}$ of the continuation set $\mathcal{C}$.

Using Theorem 4.14 and Theorem 4.15, we can reduce (52) to the free boundary problem:

$$\mathbb{L}V = rV \quad \text{for } x > b, \tag{54}$$

$$V(x) = (K - x)^+ \quad \text{for } 0 < x \leq b, \tag{55}$$

$$V'(x) = -1 \quad \text{for } x = b, \tag{56}$$

$$V(x) > (K - x)^+ \quad \text{for } x > b. \tag{57}$$

The solution to the free-boundary problem we note that the equation (54) can be expanded to

$$\frac{\sigma^2}{2}x^2V''(x) + rxV'(x) - rV = 0. \tag{58}$$

The equation in (58) is a so-called Cauchy-Euler equation. To solve such an equation, we seek a solution of the form $V(x) = x^p$, and by insertion, we get

$$p^2 - (1 - r\frac{2}{\sigma^2})p - r\frac{2}{\sigma^2} = 0.$$

Solving the equation results in the roots $p_1 = 1$ and $p_2 = -r\frac{2}{\sigma^2}$, which gives us the general solution

$$V(x) = C_1 x + C_2 x^{-r\frac{2}{\sigma^2}},$$

where C_1 and C_2 are constants. From the conditions given in (55) and (56), we see that $C_1 = 0$ and by setting $x = b$ and solving the system

$$\begin{cases} C_2 b^{-r\frac{2}{\sigma^2}} = K - b, \\ r\frac{2}{\sigma^2} C_2 b^{-r\frac{2}{\sigma^2}-1} = 1, \end{cases}$$

we get that $b = \frac{K}{1+\frac{\sigma^2}{2r}}$ and $C_2 = \frac{\sigma^2}{2r}(\frac{K}{1+\frac{\sigma^2}{2r}})^{1+r\frac{2}{\sigma^2}}$. Finally, we get that

$$V(x) = \begin{cases} \frac{\sigma^2}{2r}(\frac{K}{1+\frac{\sigma^2}{2r}})^{1+r\frac{2}{\sigma^2}} x^{-\frac{2r}{\sigma^2}} & \text{if } x \in [b, \infty), \\ (K-x)^+ & \text{if } x \in (0, b], \end{cases} \tag{59}$$

We have now reached a possible solution to the optimal stopping problem for the perpetual put. This means we exercise the option when the price $S(t) = b$. However, we want to show that this is indeed the optimal solution. We will do this by presenting and proving the following theorem from page 377 in [Peskir and Shiryaev, 2006].

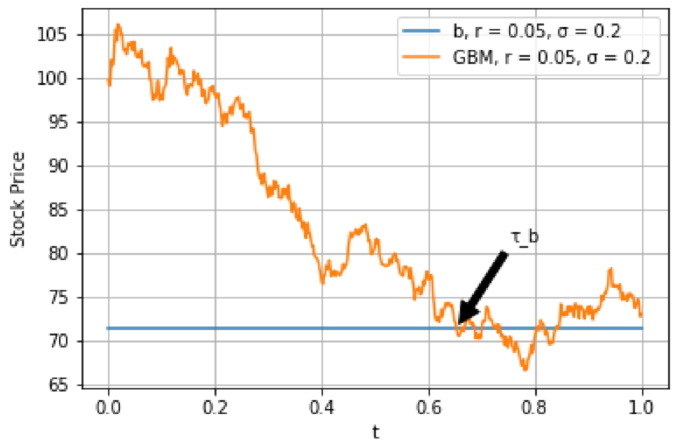

Figure 1: A visualisation of the strategy

Theorem 5.4. *Consider the arbitrage-free price $V(x)$ given by Definition 5.3. The value function is given by (59). The stopping time τ_b presented in (53) where $b = \frac{K}{1+\frac{\sigma^2}{2r}}$ is optimal in (52).*

Proof. For ease of notation, we let $V_*(x)$ be the function given in Definition 5.3, and we want to show $V_*(x) = V(x)$ for all $x > 0$, given that $V(x)$ is from (59). We begin by applying Itô's formula (see Remark 3.11) on the function $e^{-rt}V(S(t))$, which gives us

$$e^{-rt}V(S(t)) = V(x) + \int_0^t e^{-rs}(\mathbb{L}V - rV)(S(s))\mathbb{I}(S(s) \neq b)ds + \int_0^t e^{-rs}\sigma S(s)V'(S(s))dW(s), \tag{60}$$

Where $\mathbb{I}$ is the indicator function that is defined

$$\mathbb{I}(S(t) \neq b) = \begin{cases} 1 & \text{if } S(t) \neq b, \\ 0 & \text{else.} \end{cases}$$

We use the indicator function since we can see from (56) that $V''(x)$ is undefined in $x = b$. Note that by setting $G(x) = (K - x)$ we get $(\mathbb{L}G - rG)(x) = -rK$ and since $\mathbb{L}V - rV = 0$ for $x > b$ we get that

$$\mathbb{L}_X V - rV \leq 0 \tag{61}$$

except when $x = b$. From (55) and (57), we get that $V(x) \geq (K - x)^+$ for all x and together with (60) and (61), we get that

$$e^{-rt}(K - S(t))^+ \leq e^{-rt}V(S(t)) \leq V(S(t)) + M(t),$$

where $M = (M(t))_{t \geq 0}$ is given by

$$M(t) = \int_0^t e^{-rs}\sigma S(s)V'(S(s))dW(s).$$

Note that M is a martingale (see Remark 3.7), since $|V'(x)| \leq 1$ for all $x > 0$. Consider a sequence of stopping times $(\tau_n)_{n \geq 0}$ for M. Then we have that for every stopping time τ of S, we know

$$e^{-r(\tau \wedge \tau_n)}(K - S(\tau \wedge \tau_n))^+ \leq V(x) + M(\tau \wedge \tau_n),$$

for all $n \geq 1$, by Theorem 4.20. Taking the expected value over x, we get that $E_x[M(\tau \wedge \tau_n)] = 0$ for all n. By taking the limit $n \to \infty$ and using Fatou's Lemma (see Theorem A.21 in [Björk, 2004]) we get that

$$E_x[e^{-r\tau}(K - S(\tau))^+] \leq V(x).$$

If we take the supremum over all stopping times τ of S, we get that $V_*(x) \leq V(x)$ for all $x > 0$.

We also want to prove that $V_*(x) \geq V(x)$ before we can conclude that $V_*(x) = V(x)$. We do this by using equation (60) together with the condition in (55) and by Theorem 4.20, we get that

$$E_x[e^{-r(\tau_b \wedge \tau_n)} V(S(\tau_b \wedge \tau))] = V(x)$$

for all $n \geq 1$. If we look at the limit when $n \to \infty$ and using the fact that $e^{-r\tau_b} V(S(\tau_b)) = e^{-r\tau_b}(K - S(\tau_b))^+$ (note that both left and right-hand sides tend to 0 in $\tau_b \to \infty$), we find that

$$E_x[e^{-r\tau_b}(K - S(\tau_b))^+] = V(x),$$

which concludes the proof that τ_b is optimal. Thus showing that $V_*(x) = V(x)$ for all $x > 0$. □

6 Sensitivity Analysis

We have found the solution to the optimal stopping problem for perpetual American put. In this section, we will explore what the solution tells us and what happens to the solution when we change the values of the constants. A visualisation of the strategy is shown in Figure 1: wait until the stock price reaches b to exercise. Consider the solution

$$b = \frac{K}{1 + \frac{\sigma^2}{2r}},$$

with strike price K, σ is the volatility, and r is the risk-free interest rate. We are interested in observing the changes in b while using different values of σ and r.

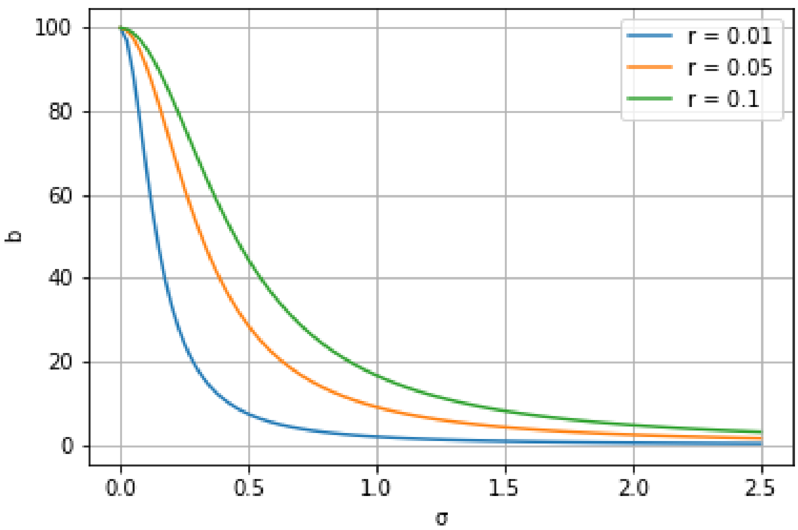

Figure 2: The difference in optimal exercise level depending on σ.

Figure 2 shows two interesting results: The higher the volatility σ, the lower the optimal exercise boundary b. Also, when the interest rate is low, we see that b gets lower. To explain the results, we can look at the statement for the Geometric Brownian Motion (34) used to simulate a stock price. Note that the second term $\sigma S(t)dW(t)$ denotes the random price changes where the volatility constant σ is a scalar factor, i.e. higher volatility leads to more significant price changes. Larger price changes lead to a higher possibility of a lower stock price and therefore lower boundary. However, we are only interested in asset prices lower than the strike price.

This reasoning confirms the results seen in Figure 2 regarding σ. The first term in (34) $rS(t)dt$ is the drift rate of the asset. This means that the underlying asset can be expected to rise in value equal to the interest rate. A higher interest rate leads to higher b since we can expect the underlying asset to rise in value, when $r > 0$. This is following the results seen in Figure 2.

When considering the value of the American put option, its intrinsic value is given by $K - x$, where K is the strike price and x is the stock price. The intrinsic value represents the payoff when exercising the put option at price x. As shown in (59) and Figure 3, the value of the American put equals its intrinsic value at $0 \leq x \leq b$. For $x \geq b$, the option's value exceeds its intrinsic value, indicating that it is more profitable for the holder not to exercise the option until the optimal exercise boundary b. We see that the green line is the value of the perpetual American put corresponding to the boundary $b2$, meaning the underlying asset has interest rate $r = 0.05$ and volatility $\sigma = 0.3$, and the red line corresponding to $b1$. This shows us that the value of different perpetual puts equals their intrinsic value at different points.

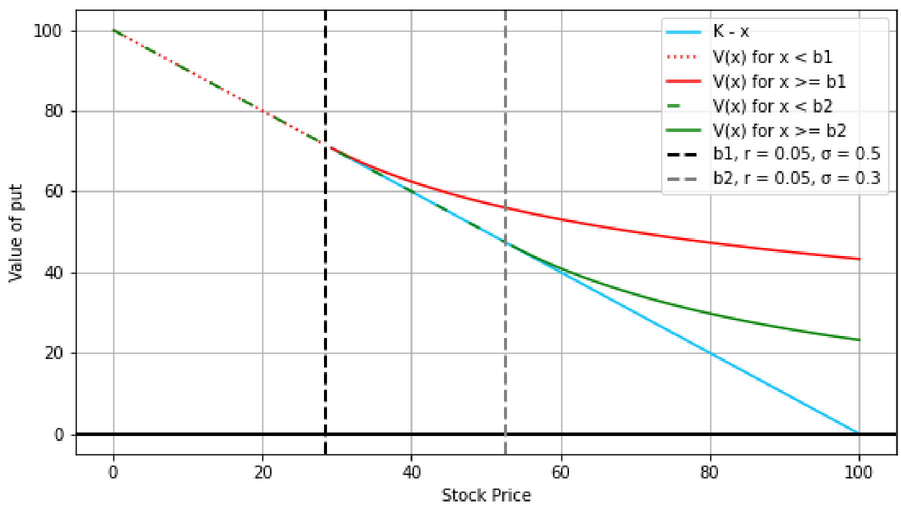

Figure 3: The value of the put depending on stock price.

Furthermore, when analysing the option's value at different volatility levels, we see that higher volatility leads to a lower optimal exercise boundary b.

To conclude, when the price of the underlying asset reaches the optimal boundary b, the value of the perpetual American put equals that of its intrinsic value. If the price of the underlying asset is higher then the optimal boundary, the value of the option is higher then its intrinsic value and it is more favorable for the holder of the option to keep it. If the price of the underlying asset is lower then the optimal boundary, then the value of the perpetual American put equals its intrinsic value as well. The boundary depends on the interest rate, represented as the drift of the underlying asset r and the volatility σ which indicates the scale of the price changes. An increase in interest rate leads to a higher expected value of the underlying asset, leading to a higher exercise boundary. When the volatility σ increases we see that the changes in price increases, resulting to a lower exercise boundary.

References

[Björk, 2004] Björk, T. (2004). *Arbitrage Theory in Continuous Time*. Oxford University Press, second edition.

[GEdgar, 2017] GEdgar (2017). https://math.stackexchange.com/questions/2279205/example-of-filtration-in-probability-theory.

[Gut, 1998] Gut, A. (1998). *An Intermediate Course in Probability*. Springer, second edition.

[Karatzas and Shreve, 1998] Karatzas, I. and Shreve, S. E. (1998). *Brownian Motion and Stochastic Calculus*. Springer Verlag, second edition.

[Numberphile2, 2014] Numberphile2 (2014). Choosing toilets (mathematical extended ending). https://www.youtube.com/watch?v=XIOoCKO-ybQ.

[Ortiz-Latorre, 2015] Ortiz-Latorre, S. (2015). Lecture notes on Introduction and Techniques in Financial Mathematics. https://www.uio.no/studier/emner/matnat/math/nedlagte-emner/STK4510/h15/lectures/l6-the-it%C3%B4-integral.pdf.

[Peskir and Shiryaev, 2006] Peskir, G. and Shiryaev, A. (2006). *Optimal Stopping and Free-Boundary Problem*. Birkhäuser Verlag.

[Rudin, 1976] Rudin, W. (1976). *Principles of Mathematical Analysis*. McGraw-Hill.

[Shreve, 2004] Shreve, S. E. (2004). *Stochastic Calculus for Finance II: Continuous-Time Models*. Springer Finance, second edition.

[Øksendal, 2013] Øksendal, B. (2013). *Stochastic Differential Equations: An introduction with Applications*. Springer, sixth edition.