Imperial College London

Imperial College London

Department of Mathematics

Optimal Stopping Problems: Autonomous Trading Over an Infinite Time Horizon

Author: Viraj Shah (CID: 01191054)

A thesis submitted for the degree of

MSc in Mathematics and Finance, 2019-2020

Declaration

The work contained in this thesis is my own work unless otherwise stated.

Acknowledgements

I would like to take this opportunity to thank Dr. Alex Tse for being a wonderful supervisor throughout my thesis project. His understanding in this area of financial mathematics has been brilliant and he has pushed me to achieve my potential. I have thoroughly enjoyed working with Dr. Tse and will carry forward the lessons learned from this experience. Without his judgement and continual encouragement, this project would not have been possible.

I am extremely grateful for the opportunity I have had to study at Imperial College London and would like to express my gratitude and appreciation to all those who have helped me to this point. I would like to recognise the support and companionship my friends have provided me with, as well as the guidance I have received from numerous academic staff.

Lastly, I am indebted to my family and want to honour their unconditional support towards me. I am truly thankful for their consistent confidence in me and for everything that they have given me.

Abstract

Statistical arbitrage (StatArb) has taken off ever since its advent in the 1980's and is being increasingly used by hedge funds and investment banks to produce profitable, quantitative trading strategies. With financial firms pushing for a larger amount of automation, we are driven to investigate trading methods that decide optimally on our behalf. At the core of StatArb strategies lies mean-reverting pairs trading models that work by exploiting market pricing discrepancies.

This thesis is devoted to the study of an optimal double-stopping problem characterised by the optimal entry and exit times. We consider a model for both a long and short trading position and then combine both strategies into one joint optimal stopping problem. The theory is idealised for the Ornstein-Uhlenbeck process but has been framed in a general way to open the methodology to other pricing processes. Analysis is given on finding the optimal execution levels, optimal strategy timing and a case study of the Ornstein-Uhlenbeck process concludes the study.

Contents

Figures

Acronyms

a.s.	almost surely
ABM	Arithmetic Brownian motion
BM	Brownian motion
DPP	dynamic programming principle
FPT	first passage time
GBM	Geometric Brownian motion
IVT	intermediate value theorem
KBE	Kolmogorov's backward equation
OU	Ornstein-Uhlenbeck
PDE	partial differential equation
SDE	stochastic differential equation
StatArb	statistical arbitrage

Chapter 1

Introduction

In all walks of life, optimisation of a task is a challenge that everyone will have faced. For a trader, this will mean executing an order at the perfect time to realise the maximum profit per transaction. Stochastic control theory encapsulates optimal trading problems like this, providing a quantitative solution to the control problem. One of the main pillars of research within this area is algorithmic trading; we will delve into optimal liquidation and order placement control problems. Assume a trader is considering a basket of assets which can be bought or sold at any time. The issue they face is the following question: when is it best to enter a position on this collection of securities and what position should I take?

StatArb strategies dominate a lot of the algorithmic trading that occurs within hedge funds and investment banks. This class of trading strategies emerged from the basic pairs trading model; a market neutral strategy that involves going long in one asset and short in another that are highly correlated. A classic example is a pair trade on Coca-Cola (KO) and Pepsi (PEP). Both companies produce almost identical fizzy drinks products so a pairs trade on these two companies should be feasible. If one out of the pair of stocks is under-performing it is bought; the other stock is short sold at a premium price. The strategy works by the logic that both companies should be equally valued given that they produce similar products, so the stock prices should eventually converge. However, this strategy is not risk-free as it heavily relies on the assumption that the pair of assets will return to its historical mean, which need not be the case. An extreme example of a pairs trading failure was seen by the hedge fund Long-Term Capital Management, who accrued losses of $286 million in traded equity. Furthermore, an important rule of the markets is that past behaviour is not indicative of future activity so one must be careful in implementing such a strategy.

This thesis begins by introducing many mathematical constructs in Chapter 2, most notably results such as Kolmogorov's backward equation, Dirichlet's problem and the Laplace transform. The theory is deliberately poised towards studying Itô diffusions as these processes have nice features that we can make use of. The most popular stochastic processes used in finance are often diffusion processes as well. Chapter 3 is where the optimal double-stopping problem is described. The first stage characterises an exit performance criterion and the second represents the entry performance criterion in terms of the exit problem. We solve these problems to obtain the optimal liquidation intervals and times at which to execute entry or exit orders for both the long and short problems. These two individual problems are then combined into one joint problem to reflect most realistic problem a trader would encounter. Lastly, Chapter 4 applies the results to the Ornstein-Uhlenbeck process to have a running example of the work.

Another aim in this thesis is to extend the theory already available for the double-stopping problems in Chapter 3 to any diffusion process. The upshot of doing this is that it allows the reader to evaluate the optimal stopping problems for a more general price process and collection of assets. To do so, we have proposed assumptions on the properties needed which can be found dotted within the chapter. The appendix also contains useful information on the minimal assumptions required to generalise the theory. Surprisingly, we show that the assumptions made can be verified for the arithmetic and geometric Brownian motion processes and consider the extension worthwhile for this reason.

Chapter 2

Diffusion Processes: Preliminary Definitions and Results

This chapter forms the foundations of the thesis. It will introduce and propose properties of diffusion processes in preparation for Chapter 4, where the Ornstein-Uhlenbeck process will be studied closely. The notion of stationarity will be explored along with what it means for a process to be Markov. Intuitively, a process is stationary if it looks the same at any point in time. It is Markov if the probabilistic structure given its history is the same as conditioning on the most recent piece of information. We then discuss transition functions for diffusion processes and construct the framework needed to obtain Kolmogorov's backward equation, which is expressed in terms of the infinitesimal generator. The solution to this partial differential equation can be conveniently expressed as a conditional expectation, which is extremely useful in many financial applications. This relationship between conditional expectation and differential equation is attractive as it opens up new avenues to computing expectations and gives a different method to solve PDEs. The extensive result of Feynman-Kac's theorem is also alluded to to show the wide scope of the theory. The remaining sections introduce the idea of a hitting time considering both one-sided and two-sided barriers, time-independent PDEs falling under the class of Sturm-Liouville boundary value problems and the concept of Laplace transformations. The last idea is widely used as a mathematical tool which we explain, along with the parallel relationship shared by Laplace transforms and solutions to the Dirichlet problem.

Our theory is built over the real line; however it can easily be extended to higher dimensions with the appropriate generalisations. We avoid this extension here for clarity of exposition and for the pure fact that we don't need it.

2.1 Existence, Uniqueness, Stationarity and Markovity

We now sketch the properties of diffusion processes which are necessary for the mathematical constructs we wish to use further on. We start by describing a general class of stochastic processes named after Itô. It is assumed throughout that all processes take values in the measurable space $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$, unless specified otherwise.

Take the probability space $(\Omega, \mathcal{F}, \mathbb{P})$, under which $(B_t)_{t\geq 0}$ is a Brownian motion under the measure $\mathbb{P}$. We consider a stochastic process $X = (X_t : \Omega \to \mathbb{R})_{t\geq 0}$ defined via the SDE

$$dX_t = b(t, X_t)\, dt + \sigma(t, X_t)\, dB_t, \tag{2.1.1}$$

with $\mathbb{F} \equiv \{\mathcal{F}_t : t \geq 0,\, \mathcal{F}_t \subseteq \mathcal{F}\}$ the collection of increasing σ-fields generated by X. A process that satisfies this SDE is called an Itô process. The term b is called the drift coefficient and σ the diffusion coefficient.

Theorem 2.1.1 (Existence and uniqueness)**.** *An Itô process $X = (X_t)_{0\leq t\leq T}$ satisfying SDE*

$$dX_t = b(t, X_t)\, dt + \sigma(t, X_t)\, dB_t, \qquad X_0 = Z,$$

with $T > 0$, where Z is independent of the Brownian motion and $\mathbb{E}|Z|^2 < \infty$ and

$$\begin{aligned} |b(t,x)| + |\sigma(t,x)| &\leq C(1+|x|), \\ |b(t,x) - b(t,y)| + |\sigma(t,x) - \sigma(t,y)| &\leq D|x-y|, \qquad x, y \in \mathbb{R}, \end{aligned} \tag{2.1.2}$$

for positive real-valued constants C and D has a unique strong solution with continuous paths almost surely.

Proof. See [Øksendal, 2013, page 66] for ideas on proving this theorem. The proof of existence is similar to the argument used to prove existence of a solution for ODEs. Uniqueness follows by an application of Gronwall's inequality. □

The two conditions in (2.1.2) ensure the stochastic process has global linear growth and global Lipschitz continuity (Appendix Definition B.1.2) respectively. This ensures that any solution does not disproportionately blow up. The class of stochastic processes with time-independent drift and diffusion coefficients, ie. $b(t,x) \equiv b(x)$ and $\sigma(t,x) \equiv \sigma(x)$ for all t, are called Itô diffusions if a unique strong solution exists.

Definition 2.1.2 (Itô diffusion)**.** Consider a stochastic process $X = (X_t)_{t \geq 0}$ satisfying SDE

$$dX_t = b(X_t)\,dt + \sigma(X_t)\,dB_t. \tag{2.1.3}$$

For the solution to uniquely exist for all $t \in [0, \infty)$, its coefficients must satisfy the conditions of Theorem 2.1.1 which reduces to the single condition of Lipschitz continuity:

$$|b(x) - b(y)| + |\sigma(x) - \sigma(y)| \leq D|x-y|, \qquad x, y \in \mathbb{R}. \tag{2.1.4}$$

We will interchangeably call a process satisfying both SDE (2.1.3) and condition (2.1.4) an Itô diffusion or more simply, a diffusion process.

The difference between an Itô process and an Itô diffusion process concerns the form of their drift and diffusion coefficients; the former has an explicit time structure in its coefficients whilst the latter has time-independent coefficients. In what follows, we consider strongly stationary diffusion processes, which will also be referred to as homogeneous (in time) or temporally homogeneous diffusions.

Definition 2.1.3 (Stationarity)**.** A process $X = (X_t)_{t \geq 0}$ is strongly stationary if for any integer $n \geq 1$, for any $t_1, t_2, \ldots, t_n \geq 0$ and for any $\tau \geq 0$, the joint distribution of $(X_{t_1}, X_{t_2}, \ldots, X_{t_n})$ is equal to that of $(X_{t_1+\tau}, X_{t_2+\tau}, \ldots, X_{t_n+\tau})$. One way of characterising this is via their joint distribution functions, namely that

$$F_{t_1,t_2,\ldots,t_n}(\boldsymbol{x}) = F_{t_1+\tau,t_2+\tau,\ldots,t_n+\tau}(\boldsymbol{x}), \qquad \boldsymbol{x} \in \mathbb{R}^n,$$

where $F_{i_1,i_2,\ldots,i_n}$ is the cumulative distribution function for $(X_{i_1}, X_{i_2}, \ldots, X_{i_n})$.

Proposition 2.1.4 (Time homogeneity)**.** *Itô diffusions are time homogeneous processes.*

Proof. [Øksendal, 2013, page 108] showed that Itô diffusions are strongly stationary, so we follow their proof.

Fix $s \geq 0$. Let $\left(X_t^{0,x}\right)_{t \geq 0}$ and $\left(X_{s+t}^{s,x}\right)_{t \geq 0}$ be two stochastic processes, where $X_t^{s,x}$ means the process X_t started at $X_s = x$ for $s \leq t$. Suppose they are both Itô diffusions satisfying SDE

$$dX_t = b(X_t)\,dt + \sigma(X_t)\,dB_t. \tag{2.1.5}$$

For each respective process, we can integrate to get

$$X_t^{0,x} = x + \int_0^t b(X_u^{0,x})\,du + \int_0^t \sigma(X_u^{0,x})\,dB_u,$$

for the first diffusion and

$$X^{s,x}_{s+t} = x + \int_s^{s+t} b(X^{s,x}_u)\,du + \int_s^{s+t} \sigma(X^{s,x}_u)\,dB_u,$$
$$= x + \int_0^t b(X^{s,x}_{s+v})\,dv + \int_0^t \sigma(X^{s,x}_{s+v})\,d\tilde{B}_v, \qquad (u = s+v),$$

for the second diffusion, where $\tilde{B}_v := B_{s+v} - B_s$ is a Brownian motion by Levy's characterisation for Brownian motions [Gan et al., 2014, Theorem 37.2, page 504]. The last line means equality in distribution, using the idea that the integrators $dB_{s+v} = d\tilde{B}_v$ are loosely equal. As both $(X^{0,x}_t)_{t\geq 0}$ and $(X^{s,x}_{s+t})_{t\geq 0}$ are solutions to SDE (2.1.5), it follows from the definition of a weak solution [Klebaner, 2005, Definition 5.8, page 136] that both processes are equal in distribution. Thus, we have time homogeneity. □

Definition 2.1.5 (Markov property)**.** A process $X = (X_t)_{t\geq 0}$ has the Markov property if for any $f : \mathbb{R} \to \mathbb{R}$, bounded and measurable,

$$\mathbb{E}\big[f(X_t)\,|\,\mathcal{F}_s\big] = \mathbb{E}\big[f(X_t)\,|\,X_s\big], \qquad 0 \leq s \leq t.$$

More formally, we should write this as

$$\mathbb{E}^x\big[f(X_t)\,|\,\mathcal{F}_s\big] = \mathbb{E}\big[f\big(X^{0,x}_t\big)\,|\,\mathcal{F}_s\big] \overset{\text{Markov}}{=} \mathbb{E}\big[f\big(X^{0,x}_t\big)\,|\,X_s\big] = \mathbb{E}^x\big[f\big(X_t\big)\,|\,X_s\big].$$

Equality means equal almost surely in this instance. Here, $\mathbb{E}^x[\,\cdot\,]$ represents the expectation with respect to the measure conditional on the process starting at $X_0 = x$, call it $\mathbb{P}^x$. For example, we write $\mathbb{E}^x[X_t]$ as an expectation with respect to $\mathbb{P}^x$. This is the same as $\mathbb{E}[X^{0,x}_t]$, where $\mathbb{E}[\,\cdot\,]$ is the expectation with respect to the underlying measure $\mathbb{P}$. One must be careful not to confuse $\mathbb{E}\big[X^{s,X_s}_t\big] \neq \mathbb{E}^{X_s}[X_t] = \mathbb{E}\big[X^{0,X_s}_t\big]$ as they are in general not equal. It is also common to write

$$\mathbb{E}^x\big[f\big(X_t\big)\,|\,X_s\big] = \mathbb{E}\big[f\big(X^{s,X_s}_t\big)\big], \tag{2.1.6}$$

which can be formally proved with the properties of conditional expectation, see [Øksendal, 2013, page 110]. We now state some results on Markov processes.

Lemma 2.1.6. *Let $X = (X_t)_{t\geq 0}$ be a time homogeneous Markov process started at $X_0 = x$. Then the Markov property can be expressed as*

$$\mathbb{E}^x\big[f(X_{s+t})\,|\,\mathcal{F}_s\big] = \mathbb{E}^{X_s}\big[f(X_t)\big], \qquad 0 \leq s \leq t.$$

Proof. As X is a stationary process, it should look the same after applying a time-shift. Therefore, it must be that X^{0,X_s}_t and X^{s,X_s}_{s+t} are equal in distribution for all times t. The proof then follows naturally. By definition of the Markov property in equation (2.1.6),

$$\begin{aligned}\mathbb{E}^x\big[f(X_{s+t})\,|\,\mathcal{F}_s\big] &= \mathbb{E}^x\big[f(X_{s+t})\,|\,X_s\big]\\ &= \mathbb{E}\big[f\big(X^{s,X_s}_{s+t}\big)\big]\\ &= \mathbb{E}\big[f\big(X^{0,X_s}_t\big)\big]\\ &= \mathbb{E}^{X_s}\big[f\big(X_t\big)\big]\end{aligned}$$

where the third line makes use of homogeneity. □

Corollary 2.1.7. *A process $X = (X_t)_{t\geq 0}$ has the Markov property if and only if for any $A \in \mathcal{B}(\mathbb{R})$,*

$$\mathbb{P}(X_t \in A\,|\,\mathcal{F}_s) = \mathbb{P}(X_t \in A\,|\,X_s), \qquad 0 \leq s \leq t.$$

Proof. To show the equivalence of definitions, we use for arbitrary σ-field $\mathcal{G}$ that

$$\mathbb{E}\big[\mathbb{1}_A(X_t)\,|\,\mathcal{G}\big] = \mathbb{P}\big(X_t \in A\,|\,\mathcal{G}\big).$$

The only if direction can be derived by choosing $f = \mathbb{1}_A$ for any $A \in \mathcal{B}(\mathbb{R})$. The converse direction can be proved by finding a sequence of simple functions $f_n \to f$, applying the dominated convergence theorem to interchange limit and expectation. □

Corollary 2.1.8. *Suppose $X = (X_t)_{t\geq 0}$ is a Markov process. Let $n \geq 1$ be an integer and take $0 \leq t_1 \leq t_2 \leq \cdots \leq t_n \leq s \leq t$ to be a sequence of non-decreasing times. Then, for any $A \in \mathcal{B}(\mathbb{R})$.*

$$\mathbb{P}(X_t \in A \,|\, X_{t_1}, X_{t_2}, \ldots, X_{t_n}, X_s) = \mathbb{P}(X_t \in A \,|\, X_s).$$

Proof. As $\sigma(X_{t_1}, X_{t_2}, \ldots, X_{t_n}, X_s) \subseteq \mathcal{F}_s$, it follows by the law of total expectation (tower rule) that,

$$\begin{aligned}\mathbb{E}\big[X_t \,|\, X_{t_1}, X_{t_2}, \ldots, X_{t_n}, X_s\big] &= \mathbb{E}\Big[\mathbb{E}\big[X_t \,|\, \mathcal{F}_s\big] \;\Big|\; X_{t_1}, X_{t_2}, \ldots, X_{t_n}, X_s\Big] \\ &= \mathbb{E}\Big[\mathbb{E}\big[X_t \,|\, X_s\big] \;\Big|\; X_{t_1}, X_{t_2}, \ldots, X_{t_n}, X_s\Big] \\ &= \mathbb{E}\big[X_t \,|\, X_s\big]\end{aligned}$$

where the final two lines follow by the Markov property from Definition 2.1.5 and by conditional expectation being $\sigma(X_s)$-measurable. Applying the same idea to $\mathbb{1}_A(X_t)$ produces the desired identity. □

We call a process which satisfies the Markov property a Markov process. Definition 2.1.5 is a popular formulation of the Markov property, however, we shall also be interested in using the auxiliary formulation in Corollary 2.1.8. This frames the definition of a process being Markov as not caring about the past. So far we have discussed that Itô processes have solutions and that they are time homogeneous. Another remarkable feature is that they are Markovian.

Theorem 2.1.9. *Itô diffusions have the Markov property.*

Proof. See [Øksendal, 2013, Thereom 7.1.2, page 109] for the proof; a basic outline is given now. Recall from the proof of homogeneity that

$$\begin{aligned}X_{s+t}^{s,X_s} &= X_s + \int_s^{s+t} b(X_u^{s,X_s})\,du + \int_s^{s+t} \sigma(X_u^{s,X_s})\,dB_u \\ &= X_s + \int_0^t b(X_{s+v}^{s,X_s})\,dv + \int_0^t \sigma(X_{s+v}^{s,X_s})\,d(B_{s+v} - B_s).\end{aligned}$$

This suggests that we can write $X_{s+t}^{s,X_s} = F(t, X_s, (B_{s+v} - B_s)_{0\leq v\leq t})$ as a function of three components. Then, for f bounded and measurable,

$$\begin{aligned}\mathbb{E}\big[f(X_{s+t}^{s,X_s}) \,|\, \mathcal{F}_s\big] &= \mathbb{E}\big[f\big(F\big(t, X_s, (B_{s+v} - B_s)_{0\leq v\leq t}\big)\big) \,|\, \mathcal{F}_s\big] \\ &= g(X_s)\end{aligned}$$

where

$$g(y) = \mathbb{E}\big[f\big(F\big(t, y, (B_{s+v} - B_s)_{0\leq v\leq t}\big)\big)\big]$$

by an application of the independence lemma (Appendix Lemma B.1.5). Here, X_s is $\mathcal{F}_s$-measurable and the BM increments $(B_{s+v} - B_s)_{0\leq v\leq t}$ are independent of $\mathcal{F}_s$. By uniqueness of solution to the SDE (2.1.3), we have a.s.

$$X_{s+t}^{0,x} \equiv X_{s+t}^{s,X_s}, \qquad t \geq 0,$$

and so

$$\mathbb{E}^x\big[f(X_{s+t}) \,|\, \mathcal{F}_s\big] = \mathbb{E}\big[f(X_{s+t}^{s,X_s}) \,|\, \mathcal{F}_s\big] = g(X_s).$$

The conditional expectation is equal to a function of X_s, which happens to be an alternative formulation of Markovity by Appendix Lemma B.2.2. This is enough to prove that Itô diffusions are Markov. □

The upshot of this theorem is that Lemma 2.1.6, Corollary 2.1.7 and Corollary 2.1.8 all apply for Itô diffusions. The Markov property as we have stated it is for continuous time and continuous state space processes, however, the appearance of this definition varies depending on whether we are in a continuous or discrete state and time space. There are numerous ways to devise the definition of a process being Markov and another useful variant of the definition is called the strong Markov property; it is an extension that applies the Markov property to stopping times. A precise formulation can be found in Appendix Definition B.2.3.

2.2 Transition Probabilities, Generators and Kolmogorov's Equations

Itô diffusions are a class of stochastic process that contain many imperative stochastic processes, such as the arithmetic and geometric Brownian motion, the Ornstein-Uhlenbeck process and the Cox-Ingersoll–Ross process. In light of the Markov property, one natural way to study these processes is via their transition probabilities. The theory is covered in much greater generality in, for example, [Rogers and Williams, 2000, Chapter III.1] or [Ethier and Kurtz, 2009, Chapter 4]. This section aims to build a rigorous and methodical derivation of the Kolmogorov backward equation through means of transition probabilities and the infinitesimal generator. We establish explicit expressions for the generator and give an extensive treatment to the KBE, talking about its different representations in literature.

Definition 2.2.1 (Transition functions)**.** A function $P(x,t\,|\,y,u)$ is called a transition function provided that for $x, y \in \mathbb{R}$ and $0 \le t \le s \le u$,

1. $P(x,0\,|\,\cdot\,,0) = \delta_x(\cdot)$ is a Dirac delta function,
2. $P(x,t\,|\,\cdot\,,u)$ is a probability measure,
3. $P(\,\cdot\,,t\,|\,y,u)$ is a measurable function,
4. $P(x,t\,|\,y,u) = \int_{\mathbb{R}} P(z,s\,|\,y,u)\, d_z P(x,t\,|\,z,s)$.

A transition function $P(x\,|\,y,t)$ is a temporally homogeneous transition function if the structure is time invariant, ie.

$$P(x\,|\,y,t) := P(x,s\,|\,y,s+t) = P(x,0\,|\,y,t), \qquad x,y \in \mathbb{R},\ s,t \ge 0.$$

As we are interested in studying stationary processes, we will examine temporally homogeneous transition functions and from this point onward will consider stochastic processes which are stationary Markov processes. Bearing this in mind, property 4 of Definition 2.2.1 can be rewritten for stationary processes as

$$P(x\,|\,y,s+t) = \int_{\mathbb{R}} P(z\,|\,y,s)\, d_z P(x\,|\,z,t), \qquad x,y \in \mathbb{R},\ s,t \ge 0. \tag{2.2.1}$$

This equation is known as the Chapman-Kolmogorov equation. A transition function is a time homogeneous transition function for a stationary Markov process X if for any $s \ge 0$,

$$P(x\,|\,y,t) = \mathbb{P}(X_{t+s} < y\,|\,X_s = x), \qquad x,y \in \mathbb{R}\,, t \ge 0,$$

with homogeneity coming as a consequence of the homogeneity of X. It now makes sense to interchange between calling P the transition probability or transition function for X. We also assume the existence of a transition density defined as

$$p(x\,|\,y,t) = \frac{\partial}{\partial y} P(x\,|\,y,t), \qquad x,y \in \mathbb{R}\,, t \ge 0. \tag{2.2.2}$$

Notice that the transition function P has no dependence on the time variable s. The variable x represents the initial state of the system; the second pair of variables (y,t) describe the final state of the process. These three parameters describe the probability of finding a future state of the process X inside the interval $(-\infty, y)$ after time t has elapsed, given that the process is currently at position x.

Since transition functions can naturally be associated to transition probabilities, the assumptions made in Definition 2.2.1 should appear to be somewhat rational. The requirement of property 4 can be justified with the following argument:

$$\begin{aligned} P(x\,|\,y,\underbrace{s+t}_{:=\tau}) &= \mathbb{P}(X_{\tau+u} < y\,|\,X_u = x) \\ &= \mathbb{P}(X_\tau < y\,|\,X_0 = x) \\ &= \int_{\mathbb{R}} \mathbb{P}(X_\tau < y,\, X_t = z\,|\,X_0 = x)\, dz \end{aligned}$$

$$
\begin{aligned}
&= \int_{\mathbb{R}} \mathbb{P}(X_\tau < y, \,|\, X_t = z \,, X_0 = x) \, \mathbb{P}(X_t = z \,|\, X_0 = x) \, dz \\
&= \int_{\mathbb{R}} \mathbb{P}(X_\tau < y, \,|\, X_t = z) \, \mathbb{P}(X_t = z \,|\, X_0 = x) \, dz \\
&= \int_{\mathbb{R}} P(z \,|\, y, s) \, d_z P(x \,|\, z, t)
\end{aligned}
$$

where we make use of time homogeneity and the Markov property as in Corollary 2.1.8. To ease the notation but keep clarity, we have written what should be the density of X_t as a probability.

A further construct we require is the infinitesimal generator, which is like a generalisation of the derivative. The generator appears when formulating PDEs in $(x, t) \mapsto P(x \,|\, y, t)$, which when solved, will produce expressions for the transition probabilities. It also acts as the connection between transition probabilities and Markov processes in the following sense: we have sketched how to construct transition functions given a Markov process; [Ethier and Kurtz, 2009, Theorem 1.1, page 157] show the converse, that we can construct a Markov process given a transition function and an initial distribution for the process. This presents a convenient coupling of concepts.

Definition 2.2.2 (Generator)**.** For Itô diffusion $X = (X_t)_{t \geq 0}$, the infinitesimal generator $\mathcal{A}$ acting on function $f : \mathbb{R} \to \mathbb{R}$ is defined as

$$
\mathcal{A} f(x) = \lim_{t \downarrow 0} \frac{\mathbb{E}^x[f(X_t)] - f(x)}{t}, \qquad x \in \mathbb{R},
$$

for suitable functions f such that the limit exists.

Here, defining what 'suitable' means is to a certain degree laborious. In many pieces of literature, the class of $C_0^2(\mathbb{R})$ functions (twice differentiable functions with continuous second derivative and compact support) is taken to guarantee existence of the generator.

Proposition 2.2.3. *Let X be an Itô diffusion starting at x and $f : \mathbb{R} \to \mathbb{R}$. If $f \in C_0^2(\mathbb{R})$, then the infinitesimal generator $\mathcal{A}$ applied to f exists and takes the form*

$$
\mathcal{A} f(x) = b(x) \, \frac{\partial f(x)}{\partial x} + \frac{1}{2} \, \sigma^2(x) \, \frac{\partial^2 f(x)}{\partial x^2}, \qquad x \in \mathbb{R}.
$$

Proof. By Itô's lemma, we have

$$
\begin{aligned}
df(X_t) =& \frac{\partial f(X_t)}{\partial x} \, dX_t + \frac{1}{2} \frac{\partial^2 f(X_t)}{\partial x^2} \, (dX_t)^2 \\
=& \left\{ b(X_t) \, \frac{\partial f(X_t)}{\partial x} + \frac{1}{2} \, \sigma^2(X_t) \, \frac{\partial^2 f(X_t)}{\partial x^2} \right\} dt + \sigma(X_t) \, \frac{\partial f(X_t)}{\partial x} \, dB_t.
\end{aligned}
$$

Integrating the SDE and taking the conditional expectation gives

$$
\begin{aligned}
\mathbb{E}^x[f(X_t)] - f(X_0) =& \mathbb{E}^x \left[\int_0^t \left\{ b(X_s) \, \frac{\partial f(X_s)}{\partial x} + \frac{1}{2} \, \sigma^2(X_s) \, \frac{\partial^2 f(X_s)}{\partial x^2} \right\} ds \right] \\
&+ \mathbb{E}^x \left[\int_0^t \sigma(X_s) \, \frac{\partial f(X_s)}{\partial x} \, dB_s \right].
\end{aligned}
$$

As f has continuous derivative and acts on a compact support, the derivative of f must be bounded. Furthermore, $\sigma(x)$ has at most linear growth. By properties of the stochastic integral [Cohen, 2015, Lemma 12.1.4, page 260], as $\sigma(X_t) \, \dfrac{\partial f(X_t)}{\partial x}$ is square-integrable, the final term is martingale. Taking the initial condition $X_0 = x$,

$$
\begin{aligned}
\lim_{t \to 0^+} \frac{\mathbb{E}^x[f(X_t)] - f(x)}{t} &= \lim_{t \to 0^+} \frac{1}{t} \, \mathbb{E}^x \left[\int_0^t \left\{ b(X_s) \, \frac{\partial f(X_s)}{\partial x} + \frac{1}{2} \, \sigma^2(X_s) \, \frac{\partial^2 f(X_s)}{\partial x^2} \right\} ds \right] \\
&= \mathbb{E}^x \left[\lim_{t \to 0^+} \frac{1}{t} \int_0^t \left\{ b(X_s) \, \frac{\partial f(X_s)}{\partial x} + \frac{1}{2} \, \sigma^2(X_s) \, \frac{\partial^2 f(X_s)}{\partial x^2} \right\} ds \right] \\
&= b(x) \, \frac{\partial f(x)}{\partial x} + \frac{1}{2} \, \sigma^2(x) \, \frac{\partial^2 f(x)}{\partial x^2}
\end{aligned}
$$

where the penultimate line follows by an application of Fubini's theorem and the final line by the fundamental theorem of calculus. □

The following theorem introduces a handy link between solving a PDE and calculating a conditional expectation. It provides two vastly different methods to attack the same problem, which is rather remarkable. Just as with the Markov property, there are multiple configurations of this concept which we will detail in time. We call the differential equation in (2.2.3) the Kolmogorov backward equation.

Theorem 2.2.4 (Kolmogorov's backward equation). *Let $f : \mathbb{R} \to \mathbb{R}$ be a suitable function with $f \in C_0^2(\mathbb{R})$. Define $u(t, x) = \mathbb{E}^x[f(X_t)]$ for an Itô diffusion X. Then the function $u(t, x)$, with infinitesimal generator $\mathcal{A}$ applied to the map $x \mapsto u(t, \cdot)(x) = u(t, x)$, satisfies*

$$\begin{aligned} \frac{\partial u}{\partial t}(t, x) &= \mathcal{A}u(t, x), \qquad && x \in \mathbb{R},\, t \geq 0, \\ u(0, x) &= f(x), && x \in \mathbb{R}. \end{aligned} \tag{2.2.3}$$

Furthermore, if $v(t, x)$ satisfies (2.2.3), then the solution is unique and $v(t, x) \equiv u(t, x)$ must hold, where $u(t, x) = \mathbb{E}^x[f(X_t)]$.

Proof. Our proof of existence and uniqueness follows the lines of [Øksendal, 2013, page 131]. To prove that $u(t, x)$ is a solution, define $w(x) = u(t, x)$. By the Markov property in Lemma 2.1.6,

$$\begin{aligned} \frac{\mathbb{E}^x[g(X_s)] - g(x)}{s} &= \frac{1}{s}\, \mathbb{E}^x\Big[\mathbb{E}^{X_s}\big[f(X_t)\big] - \mathbb{E}^x\big[f(X_t)\big]\Big] \\ &= \frac{1}{s}\, \mathbb{E}^x\Big[\mathbb{E}^x\big[f(X_{s+t}) \,|\, \mathcal{F}_s\big] - \mathbb{E}^x\big[f(X_t) \,|\, \mathcal{F}_s\big]\Big] \\ &= \frac{1}{s}\, \mathbb{E}^x\big[f(X_{s+t}) - f(X_t)\big] \\ &= \frac{u(s+t, x) - u(t, x)}{s} \end{aligned}$$

where differentiability with respect to t follows by Dynkin's formula (Appendix Proposition B.2.5) choosing $\tau = t$. Finally, taking the limit,

$$\mathcal{A}u(t, x) = \lim_{t \downarrow 0} \frac{\mathbb{E}^x[g(X_s)] - g(x)}{s} = \lim_{t \downarrow 0} \frac{u(s+t, x) - u(t, x)}{s} = \frac{\partial u}{\partial t}$$

and so $u(t, x)$ satisfies the PDE. To show uniqueness, suppose that $v(t, x)$ is a solution to system (2.2.3). For $0 \leq s \leq t$, define $w(s, x) = v(t - s, x)$. Then

$$0 = -\frac{\partial v}{\partial s}(t - s, x) + \mathcal{A}v(t - s, x) = \frac{\partial w}{\partial s}(s, x) + \mathcal{A}w(s, x) \tag{2.2.4}$$

and the initial condition converts to the terminal condition $w(t, x) = v(0, x) = f(x)$. An application of Itô's formula gives

$$\begin{aligned} w(t, X_t) - w(0, X_0) &= \int_0^t \left\{ \frac{\partial w(s, X_s)}{\partial s} + b(X_s)\, \frac{\partial w(s, X_s)}{\partial x} + \frac{1}{2}\, \sigma^2(X_s)\, \frac{\partial^2 w(s, X_s)}{\partial x^2} \right\} ds \\ &\quad + \int_0^t \sigma(X_s)\, \frac{\partial w(s, X_s)}{\partial x}\, dB_s \\ &= \int_0^t \sigma(X_s)\, \frac{\partial w(s, X_s)}{\partial x}\, dB_s \end{aligned}$$

using (2.2.4). Since $f \in C_0^2(\mathbb{R})$, the partial derivative of f is bounded. As the final term is square-integrable, by properties of the stochastic integral the final term is a true martingale. Taking the conditional expectation, we get

$$v(t, x) = w(0, x) = \mathbb{E}^x[w(0, X_0)] = \mathbb{E}^x[w(t, X_t)] = \mathbb{E}^x[f(X_s)].$$

□

The above proof gives an alternative formulation of the KBE for free. Define the time-shifted or time-reversed solution

$$v(t,x) := u(T-t,x) = \mathbb{E}^x[f(X_{T-t})].$$

for $0 \le t \le T$. Then $v(t,x)$ is the unique solution to the PDE system

$$\begin{aligned} &\frac{\partial v}{\partial t}(t,x) + \mathcal{A}v(t,x) = 0, && x \in \mathbb{R},\, t \ge 0, \\ &v(T,x) = f(x), && x \in \mathbb{R}, \end{aligned} \tag{2.2.5}$$

upon an application of the chain rule, using that $\frac{\partial}{\partial(T-t)}(t) = -1$. Comparing partial differential equation systems (2.2.3) and (2.2.5), the former is the initial condition formulation of the KBE and the latter is the terminal condition formulation.

Remark 2.2.5. A legitimate question to ask is whether we can extend the time-reversed system (2.2.5) to a second-order linear PDE that is non-homogeneous. The Feynman-Kac theorem (Appendix Theorem B.3.2) does this, giving a stochastic representation of the solution to the parabolic PDE. There also exists an analogue to the backward equation for transition densities called the Kolmogorov forward equation or Fokker-Planck equation, which is the adjoint of Kolmogorov's backward equation. Together, these two equations are known as Kolmogorov equations (see Appendix Theorem B.3.1).

It is tempting to write that if we chose $f(x) = \mathbb{1}_B(x)$ for all x, with $B = (-\infty, y) \subset \mathbb{R}$, then we realise a PDE for transition probabilities. Indeed, it would appear that

$$u(t,x) = \mathbb{E}^x[f(X_t)] = \mathbb{E}^x\big[\mathbb{1}_B(X_t)\big] = \mathbb{P}(X_t \in B \,|\, X_0 = x) = \mathbb{P}(X_t < y \,|\, X_0 = x) = P(x \,|\, y, t),$$

and similarly for v. But Theorem 2.2.4 is only for $f \in C_0^2(\mathbb{R})$ (as we know the generator definitely exists for such functions), thus the argument is incorrect. Nevertheless, we can still realise a PDE for transition densities as follows: assuming continuity of the transition density and its derivatives in x and t, we can exchange the order of differentiation and integration (Leibniz's rule) in (2.2.3) to give

$$\int_{\mathbb{R}} f(y) \Big\{ -\frac{\partial p}{\partial t}(x \,|\, y, t) + \mathcal{A}p(x \,|\, y, t) \Big\} dy = 0.$$

Furthermore, the initial density must be a Dirac delta function as it needs to satisfy

$$f(x) = u(0,x) = \int_{\mathbb{R}} f(y)\, p(x \,|\, y, 0)\, dy.$$

By arbitrariness of $f \in C_0^2(\mathbb{R})$, it follows that

$$\begin{aligned} &\frac{\partial p}{\partial t}(x \,|\, y, t) = \mathcal{A}p(x \,|\, y, t), && x, y \in \mathbb{R},\, t \ge 0, \\ &p(x \,|\, y, 0) = \delta_y(x), && x, y \in \mathbb{R}. \end{aligned} \tag{2.2.6}$$

This is Kolmogorov's backwards equation in terms of the transition density functions. The etymology behind the term 'backward equation' comes from the fact that the derivatives in the generator are with respect to the initial variable x. Conversely, the Fokker-Planck equation features derivatives of the future variable y and is dubbed the 'forward equation'.

Remark 2.2.6. Assuming continuity of the density function in y, we can integrate equation (2.2.6) and use Leibniz's rule again to realise a PDE in the transition probabilities. So given the transition density functions are sufficiently smooth, it is still possible to superficially pick $f = \mathbb{1}_B$ and also get the KBE in terms of transition probabilities. This suggests the KBE is defined for a wider class of functions than we have postulated. The theory of Markov processes defines the infinitesimal generator more generally for functions f that are bounded and measurable. For such functions, the linear operator P_t acting on f is defined as

$$P_t f(x) := \int_{\mathbb{R}} f(y) p(x \,|\, t, dy) = \mathbb{E}^x[f(X_t)] = u(t,x).$$

With this, one can define the KBE for f bounded and measurable, assuming the generator $\mathcal{A}$ applied to u exists. Note that this does not guarantee existence of the generator applied to f, but only for u.

2.3 First Passage Times, Laplace Transforms and the Dirichlet Problem

This section continues the theme of using Itô diffusions $X = (X_t)_{t\geq 0}$. We start by introducing the definition of a two-sided hitting time to set up the probabilistic framework involving first passage times for Section 2.4, motivated by the analysis in [Darling et al., 1953]. After this, we introduce two important mathematical constructs; an integral transform called the Laplace transform and the theory of Sturm-Liouville second-order linear differential equations. Within these PDE problems, there exists a classification based on the type of initial conditions; either Dirichlet, Neumann or Cauchy. We will study Dirichlet type problems with boundary value conditions. The study of this PDE system should appear familiar and one should notice the similarities of the results from discussing Kolmogorov's backward equation. This similarity is no coincidence, as we explain towards the end of the section.

We commence by talking about a stopping time, which is a random variable that captures the (random) time at which an action is done by the stochastic process it models. These constructs are a good way of creating time variables that are both $\mathbb{F}$-measurable and random. We then introduce hitting times that characterise the first time a process exits a domain $D \subseteq \mathbb{R}$. For this reason, hitting times are regularly referred to as first exit or entry times, or first passage times. Note that a hitting time is simply a special case of a stopping time, as shown in [Øksendal, 2013, Example 7.2.2, page 111].

Definition 2.3.1 (Stopping time)**.** Let $\mathbb{G} = \{\mathcal{G}_t\}_{t\geq 0}$ be a family of increasing σ-fields. A function $\tau : \Omega \to [0, \infty]$ is called a stopping time with respect to $\mathbb{G}$ if

$$\{\omega : \tau(\omega) \leq t\} \in \mathcal{G}_t, \qquad t \geq 0.$$

Lemma 2.3.2. *For two stopping times τ and ς, denote the minimum of these two random variables by $\tau \wedge \varsigma$. Then the minimum is a stopping time itself.*

Proof. The stopping times are measurable with respect to the family of increasing σ-fields $\{\mathcal{G}_t\}_{t\geq 0}$, so both

$$\{\omega : \tau(\omega) \leq t\} \in \mathcal{G}_t \text{ and } \{\omega : \varsigma(\omega) \leq t\} \in \mathcal{G}_t.$$

To prove that $\tau \wedge \varsigma$ is a stopping time, we need to show that

$$\{\omega : \tau(\omega) \wedge \varsigma(\omega) \leq t\} = \{\omega : \tau(\omega) \leq t\} \cup \{\omega : \varsigma(\omega) \leq t\} \in \mathcal{G}_t.$$

As σ-algebras are closed under union, the above set is contained in $\mathcal{G}_t$ and the proof is complete. □

Definition 2.3.3 (Hitting times)**.** The double barrier first passage time for a process $X = (X_t)_{t\geq 0}$ starting at $X_0 = x \in (a, b) \subseteq \mathbb{R}$, is defined as the random variable

$$T_{ab}(x) = \inf\Big\{t \geq 0 \,|\, X_t \notin (a, b)\Big\}.$$

Wherever it is clear, we omit the dependence on the starting value x to avoid duplication of notation. For example $\mathbb{E}^x\big[T_c\big]$ instead of $\mathbb{E}^x\big[T_c(x)\big]$. The distribution function and density for $T_{ab}(x)$ are defined as

$$F_{ab}(x, t) = \mathbb{P}(T_{ab}(x) < t),$$
$$f_{ab}(x, t) = \frac{\partial}{\partial t} F_{ab}(x, t),$$

where $\mathbb{P}(T_{ab}(x) < t) \equiv \mathbb{P}^x(T_{ab} < t)$.

The first passage time describes the earliest time at which the process leaves the interval (a, b), given that it started inside this interval. We can extend the definition to attain a one-sided barrier by either taking $a = -\infty$ or $b = \infty$. In that case, the one-sided hitting time is defined as

$$T_c(x) = \begin{cases} T_{-\infty,c}(x), & x < c, \\ T_{c,\infty}(x), & x > c, \end{cases}$$

and the analogous distribution and density functions as $F_c(x,t)$ and $f_c(x,t)$. We will also need the distribution of the double barrier hitting time given that it hits the lower barrier a or the upper barrier b first. We define their distribution functions respectively as

$$F_{ab}^{-}(x,t) := \mathbb{P}\Big(T_{ab}(x) < t,\, T_{ab}(x) = T_a(x)\Big),$$
$$F_{ab}^{+}(x,t) := \mathbb{P}\Big(T_{ab}(x) < t,\, T_{ab}(x) = T_b(x)\Big).$$

Their densities are correspondingly defined as $f_{ab}^{-}(x,t)$ and $f_{ab}^{+}(x,t)$ in the usual way. By the law of total probability, the three distribution functions can be related via

$$F_{ab}(x,t) = F_{ab}^{-}(x,t) + F_{ab}^{+}(x,t).$$

Note that although f_{ab}^{-} and f_{ab}^{+} are not necessarily probability densities (as their distribution functions may not integrate to one), f_{ab} is a true probability density function. Hence, the tag 'probability density function' is strictly reserved for f_{ab} whilst f_{ab}^{-} and f_{ab}^{+} are just called densities.

Definition 2.3.4 (Laplace transform)**.** For a function $f : \mathbb{R}_{\geq 0} \to \mathbb{R}$, the (unilateral) Laplace transform is defined as

$$\hat{f}(s) = \int_0^{\infty} f(t)\, \mathrm{e}^{-st}\, dt, \qquad s \in \mathbb{C},$$

whenever it exists.

The Laplace transform is an integral transform that maps a function from one space to another. It provides a way of solving complex differential equation systems in the domain space by converting them into simpler systems on the transformed space. Laplace transforms are more general than Fourier transforms, as they transform over the complex domain. We usually call the transformed space the 'complex frequency domain'. Much work has been done on this area of mathematics and so it is common to find tables of Laplace transforms for standard functions, for example as in [Bateman, 1954, page 129]. One useful formula we will use is the transform of a convolution. Let

$$(f * g)(t) = \int_0^{t} f(u)g(t-u)\, du.$$

Then the Laplace transform of $(f * g)$ is

$$\mathcal{L}\big\{(f * g)(t)\big\} = \hat{f}(s)\hat{g}(s), \tag{2.3.1}$$

where $\mathcal{L}\{\cdot\}$ represents the Laplace transform. So the Laplace transform of a convolution of two functions is their product. Another practical identity we can identify to ensure the finiteness of stopping times is the Laplace transform of $f_c(x,t)$. Its transform evaluated at $s = 0$ is

$$\hat{f}_c(x,0) = \int_0^{\infty} f_c(x,t)\, dt = \mathbb{P}(T_c(x) < \infty). \tag{2.3.2}$$

So the stopping time $T_c(x)$ is finite almost surely if and only if $\hat{f}_c(x,0) = 1$. In the subsequent chapters we will solve the transformed KBE in the complex frequency domain; to find this PDE, we will have to take the Laplace transform of the probability transition functions p. The time domain transform $\hat{p}$ is defined as

$$\hat{p}(x \,|\, y, s) = \int_0^{\infty} p(x \,|\, y, t)\, \mathrm{e}^{-st}\, dt.$$

The resulting differential equation will fall under the class of Sturm-Liouville problems, which are generally of the form

$$\mathcal{A}u(x) = r(x)u(x)$$

with appropriate boundary conditions. Notice the time dependence in the solution has been removed: the new PDE is much simpler and can be readily solved using standard techniques.

Remark 2.3.5. The Laplace transform of a continuous probability density function f for non-negative random variable X is close to its moment generating function. For $s \in \mathbb{C}$,

$$\hat{f}(-s) = \int_0^\infty f(t)\, \mathrm{e}^{st}\, dt = \mathbb{E}\big[\mathrm{e}^{sX}\big].$$

Typically $\hat{f}$ is called the Laplace transform of the random variable X itself, rather than the transform of its density function f.

In mathematical finance, the expression for the Laplace transform of a random variable is regularly encountered. When $s = r \in \mathbb{R}$ and $X = \tau$ is a random stopping time, then $\hat{f}(r) = \mathbb{E}\big[\mathrm{e}^{-r\tau}\big]$ often resembles a discount factor, discounting back from time τ. More commonly, problems are posed with the aim of calculating the discounted expectation for some payoff function f, ie. $\mathbb{E}\big[\mathrm{e}^{-r\tau} f(X_\tau)\big]$. To get a closed-form expression for this, we look at the solution to an elliptic PDE system.

Theorem 2.3.6 (Dirichlet Problem)**.** *Suppose that*

(i) $D \subset \mathbb{R}$ be open and bounded,

(ii) r is a non-negative Hölder continuous function on the closure $\bar{D}$,

(iii) f is continuous on the boundary ∂D,

(iv) the boundary is sufficiently smooth (eg. ∂D is C^2).

Let $\tau_D = \inf\{t \geq 0 : X_t \notin D\}$ be the first exit time of the process from D, where we consider $X = (X_t)_{t \geq 0}$ as an Itô diffusion started at $X_0 = x$. Define

$$u(x) = \mathbb{E}^x \left[\exp\left(- \int_0^{\tau_D} r(X_s)\, ds \right) f(X_{\tau_D}) \right]. \tag{2.3.3}$$

Then the function $u(x)$ satisfies

$$\begin{aligned} \mathcal{A}u(x) &= r(x)u(x), && x \in D, \\ u(x) &= f(x), && x \in \partial D. \end{aligned} \tag{2.3.4}$$

Furthermore, if $v(x)$ satisfies (2.3.4)*, then the solution is unique and $v(x) \equiv u(x)$ must hold.*

Proof. In the theory of partial differential equations, it is known that a unique solution exists for this problem. The existence of a solution is given in [Friedman, 1975, Theorem 2.4, page 134]. We do not attempt to derive a proof here and take this result for granted.

To show the unique stochastic representation, we outline the steps in [Friedman, 1975, page 145]. By Itô's formula and the product rule applied to

$$Y_t = u(X_t) \exp\left(- \int_0^t r(X_s)\, ds \right),$$

we get for $T > 0$

$$u(x) = \mathbb{E}^x \left[\exp\left(- \int_0^{\tau_D \wedge T} r(X_s)\, ds \right) f(X_{\tau_D \wedge T}) \right].$$

which is well-defined as $\tau_D \wedge T$ is a bounded stopping time. Assuming $\mathbb{E}^x[\tau_D] < \infty$ (also shown in the aforementioned reference) implies that $\tau_D < \infty$. Then we can let $T \to \infty$ to get the desired result. □

Remark 2.3.7. By taking the function $r \geq 0$ to be constant in Theorem 2.3.6, the payoff function $\mathbb{E}^x\big[\mathrm{e}^{-r\tau} f(X_\tau)\big]$ we aimed for is achieved.

Remark 2.3.8. As before, we can extend PDE (2.3.4) to a non-homogeneous system which is given in Appendix Theorem B.3.3. There also exists an extension that relaxes the constraint of D being bounded, which can be read in [Øksendal, 2013, Exercise 9.12, page 192]. Typically, we will be interested in taking $D = (a, b) \subseteq \mathbb{R}$ to be an open interval and so this result will be useful when D is unbounded. All in all, there are varying formulations of these PDE problems with different constraints which we don't place too much emphasis on. The conditions should not be the focus; more the impressive stochastic representation results of the PDE solutions.

As with many features of Markov processes, we often discover a property for non-random times and then are able to extend it to random stopping times. It is worth realising the same has been accomplished here. The Feynman-Kac theorem provides a means to work out discounted payoffs for deterministic times and the extension of this to stopping times can be addressed by the generalised Dirichlet problem, Appendix Theorem B.3.3. The reader may also detect the similarity between the Feynman-Kac theorem and the Dirichlet problem in Theorem 2.3.6. Let us inspect this relationship in reverse, ie. we will try to get a time-dependent version of the Feynman-Kac theorem for free from the Dirichlet problem.

Let $y = (t, x) \in \mathbb{R}^2$. Assuming X is an Itô process starting at x as usual, define the process $Y = (Y_s)_{s\geq 0}$ by $Y_s = (t + s, X_s) \in \mathbb{R}^2$ starting at $Y_0 = (t, x)$. Then

$$u(y) := u(t,x) = \mathbb{E}^{t,x}\left[\exp\left(-\int_0^{\tau_{\tilde{D}}} \tilde{r}(Y_s)\,ds\right)\tilde{f}(Y_{\tau_{\tilde{D}}})\right]$$

is the unique solution to the PDE system

$$\begin{aligned} \tilde{\mathcal{A}}u(y) &= \tilde{r}(y)u(y), & y \in \tilde{D}, \\ u(y) &= \tilde{f}(y), & y \in \partial\tilde{D}, \end{aligned} \tag{2.3.5}$$

for the appropriate two-dimensional version of problem (2.3.4). Here, $\tilde{\mathcal{A}}$ is the generator applied to $u \in \mathbb{R}^2$ and $\tilde{D} \subseteq \mathbb{R}^2$ is the augmented two-dimensional space to D. If we take $\tilde{D}$ to be the region $[0, T) \times \mathbb{R} \subset \mathbb{R}^2$, then $\tau_{\tilde{D}}$ describes the first exit time of Y_s from the time interval $[0, T)$.

Writing Y in differential form gives

$$dY_s = \begin{bmatrix} dY_{s,1} \\ dY_{s,2} \end{bmatrix} = \begin{bmatrix} 1 \\ b(X_s) \end{bmatrix} ds + \begin{bmatrix} 0 \\ \sigma(X_s) \end{bmatrix} dB_s, \qquad Y_0 = \begin{bmatrix} t \\ x \end{bmatrix}.$$

The infinitesimal generator for u can be written as

$$\begin{aligned} \tilde{\mathcal{A}}u(y) &= \frac{\partial u}{\partial t}(t,x) + b(x)\frac{\partial u}{\partial x}(t,x) + \frac{1}{2}\sigma^2(x)\frac{\partial^2 u}{\partial x^2}(t,x) \\ &= \frac{\partial u}{\partial t}(t,x) + \mathcal{A}u(t,x), \end{aligned}$$

where $\mathcal{A}$ is the generator applied to the map $x \mapsto u(t,\cdot)(x) = u(t,x)$. We also assume that $\tilde{r}(t,x) \equiv r(x)$ is homogeneous in time. With these three additions, system (2.3.5) becomes

$$\begin{aligned} &\frac{\partial u}{\partial t}(t,x) + \mathcal{A}u(t,x) - r(x)u(t,x) = 0, && (t,x) \in \tilde{D}, \\ &u(T,x) = \tilde{f}(T,x) =: f(x), && (T,x) \in \partial\tilde{D}. \end{aligned} \tag{2.3.6}$$

The stochastic representation of the solution to this PDE system is

$$u(t,x) = \mathbb{E}^{t,x}\left[\exp\left(-\int_0^{T-t} r(X_s)\,ds\right) f(X_{T-t})\right],$$

as $\tau_{\tilde{D}} = T - t$ on the boundary and $\tilde{f}(Y_{T-t}) = \tilde{f}(T, X_{T-t}) = f(X_{T-t})$. Here, $\mathbb{E}^{t,x}[\,\cdot\,]$ means conditional on $Y_0 = (Y_{0,1}, Y_{0,2}) = y = (t, x)$. By homogeneity of X, we can apply a time-shift to see that

$$\begin{aligned} u(t,x) &= \mathbb{E}\left[\exp\left(-\int_0^{T-t} r(X_{s+t})\,ds\right) f(X_T)\,\middle|\, X_t = x\right] \\ &= \mathbb{E}\left[\exp\left(-\int_t^{T} r(X_v)\,dv\right) f(X_T)\,\middle|\, X_t = x\right], \qquad (v = s+t). \end{aligned}$$

The new PDE system (2.3.6) together with solution $u(t, x)$ form the basic elements of the Feynman-Kac theorem. The Kolmogorov equation also comes for free: compare (2.2.5) to the above derivation with $r \equiv 0$.

The upshot from this analysis is that elliptic partial differential equations are worth studying first: we can frequently realise the time-dependent PDEs as an instance of the more general stopping time problems.

2.4 Laplace Transforms of Transition and First Passage Time Densities

We continue to work with Itô diffusions and closely follow the work of [Darling et al., 1953] in this part. Our goal is to determine Laplace transformations of the density functions laid out in Section 2.3; these identities will be used in the remaining chapters to tackle the optimal stopping problems. Derivations of the transforms are rather mechanical once equipped with the renewal-style equations. We also explore the interesting duality between conditional expectations and Laplace transforms. Previously, we had seen that conditional expectations were the solution to some second-order linear PDEs; the converse is also true and these solutions are nothing other than Laplace transformations.

The first identity we want to derive is for probability density functions p as defined in (2.2.2). The identity looks similar to the renewal equation from renewal theory, so it acquires its name as such.

Lemma 2.4.1 (Renewal principle)**.**

$$p(x\,|\,y,t) = \int_0^t f_c(x,s)\,p(c\,|\,y,t-s)\,ds, \qquad x < c < y. \tag{2.4.1}$$

Proof. We start via the transition probabilities using the law of total probability. By exhausting all possible outcomes, a direct enumeration of the joint events gives

$$\begin{aligned} P(x\,|\,y,t) &= \mathbb{P}(X_t < y\,|\,X_0 = x) \\ &= \mathbb{P}(X_t < y \cap T_c < t\,|\,X_0 = x) + \mathbb{P}(X_t < y \cap T_c \geq t\,|\,X_0 = x) \\ &= \mathbb{P}(X_t < y\,|\,T_c < t, X_0 = x)\,\mathbb{P}(T_c < t\,|\,X_0 = x) + \mathbb{P}(T_c \geq t\,|\,X_0 = x) \end{aligned}$$

where we assume $\tau_c < \infty$ almost surely. This happens if and only if $\hat{f}_c(x,0) = 1$, as seen in condition (2.3.2). The last line follows as the event $\{X_t < y \cap T_c \geq t\} = \{T_c \geq t\}$. Continuing,

$$\begin{aligned} P(x\,|\,y,t) &= \int_0^t \mathbb{P}(X_t < y \cap T_c = s\,|\,X_0 = x)\,ds + \mathbb{P}(T_c \geq t\,|\,X_0 = x) \\ &= \int_0^t \mathbb{P}(X_t < y\,|\,T_c = s, X_0 = x)\,\mathbb{P}(T_c = s\,|\,X_0 = x)\,ds + \mathbb{P}(T_c \geq t\,|\,X_0 = x) \\ &= \int_0^t \mathbb{P}(X_t < y\,|\,X_s = c, X_u < c \ \ \forall u \in (0,s), X_0 = x)\,f_c(x,s)\,ds + (1 - F_c(x,t)) \\ &= \int_0^t \mathbb{P}(X_t < y\,|\,X_s = c)\,f_c(x,s)\,ds + (1 - F_c(x,t)) \end{aligned}$$

by definition of conditional probability and the Markov property. Assuming continuity of the functions, we can exchange the derivative and integral to attain the transition density function as

$$\begin{aligned} p(x\,|\,y,t) &= \int_0^t \partial_y \mathbb{P}(X_t < y\,|\,X_s = c)\,f_c(x,s)\,ds \\ &= \int_0^t \partial_y \mathbb{P}(X_{t-s} < y\,|\,X_0 = c)\,f_c(x,s)\,ds \\ &= \int_0^t p(c\,|\,y,t-s)\,f_c(x,s)\,ds \end{aligned}$$

□

It is worth realising that the above proof also holds when $y < c < x$. This will be useful when we want to work out the form of $\hat{p}$ and $\hat{f}_c$. The second identity we need to carefully derive is similar to identity (2.4.1), except for the distributions of T_a and T_b. The basis of the proof is as before; we enumerate over all possible outcomes and reach an identity via homogeneity and Markovity.

Lemma 2.4.2. *For $a < x < b$,*

$$\begin{aligned} f_a(x,t) &= f_{ab}^{-}(x,t) + \int_0^t f_{ab}^{+}(x,s)\,f_a(b,t-s)\,ds, \\ f_b(x,t) &= f_{ab}^{+}(x,t) + \int_0^t f_{ab}^{-}(x,s)\,f_b(a,t-s)\,ds. \end{aligned} \tag{2.4.2}$$

Proof. We will prove the second identity, with the proof of the first being similar. Since

$$F_b(x,t) = \mathbb{P}\Big(T_b(x) < t, T_{ab}(x) = T_a(x)\Big) + \mathbb{P}\Big(T_b(x) < t, T_{ab}(x) = T_b(x)\Big),$$

we have that

$$f_b(x,t) = \mathbb{P}\Big(T_b(x) = t, T_{ab}(x) = T_a(x)\Big) + f_{ab}^{+}(x,t)$$

by using an abuse of notation, where the middle term is really a density function. Expanding over all possible times that the process can hit the lower barrier a, we get

$$\begin{aligned}
&\mathbb{P}\Big(T_b(x) = t, T_{ab}(x) = T_a(x)\Big) \\
&= \int_0^t \mathbb{P}\Big(T_b(x) = t, T_{ab}(x) = T_a(x), T_{ab}(x) = s\Big)\, ds \\
&= \int_0^t \mathbb{P}\Big(T_b(x) = t \,\Big|\, T_{ab}(x) = T_a(x), T_{ab}(x) = s\Big)\, \mathbb{P}\Big(T_{ab}(x) = T_a(x), T_{ab}(x) = s\Big)\, ds \\
&= \int_0^t \mathbb{P}\big(T_b(x) = t \,|\, T_a(x) = s\big)\, f_{ab}^{-}(x,s)\, ds \\
&= \int_0^t \mathbb{P}\big(T_b(a) = t - s \,|\, T_a(a) = 0\big)\, f_{ab}^{-}(x,s)\, ds \\
&= \int_0^t f_b(a, t-s)\, f_{ab}^{-}(x,s)\, ds
\end{aligned}$$

where $\mathbb{P}\big(T_b(x) = t \,\big|\, T_a(x) = s\big) = \mathbb{P}\big(T_b(a) = t - s \,\big|\, T_a(a) = 0\big)$. This is true by stationarity and the Markov property of the process X, for example $T_b(x) = t \iff X_u \in (-\infty, b)$ for $0 \le u < t$ and $X_t = b$ for deterministic time t. Upon rearranging, this completes the proof. ☐

Now we can find the Laplace transformations of the hitting time and transition function densities. The transform of a convolution from (2.3.1) will be useful, as both Lemmas 2.4.1 and 2.4.2 contain a convolution of functions. The proofs below follow the insightful work of [Darling et al., 1953, page 5] again.

Proposition 2.4.3. *For fixed $s \in \mathbb{C}$, the Laplace transform of the transition density p is a product*

$$\hat{p}(x \,|\, y, s) = \begin{cases} u(x)\tilde{u}(y), & x < y, \\ v(x)\tilde{v}(y), & x > y, \end{cases}$$

and the Laplace transform of the FPT density f_c is a ratio

$$\hat{f}_c(x,s) = \begin{cases} \dfrac{u(x)}{u(c)}, & x < c, \\[2ex] \dfrac{v(x)}{v(c)}, & x > c. \end{cases}$$

for $x, y, c \in \mathbb{R}$. The functions $u, \tilde{u}, v, \tilde{v}$ are unique up to scaling factors of s.

Proof. For $x < c < y$, the Laplace transform of the renewal principle equation (2.4.1) is

$$\hat{p}(x \,|\, y, s) = \hat{f}_c(x,s)\, \hat{p}(c \,|\, y, s) \tag{2.4.3}$$

For fixed $s \in \mathbb{C}$, the transformed density $\hat{p}(x \,|\, y, s)$ can be seen as a function of x multiplied by a function of y (both depending on c). Let

$$\hat{p}(x \,|\, y, s) = u(x)\tilde{u}(y) \equiv u(x;s)\tilde{u}(y;s), \tag{2.4.4}$$

where we suppress the dependence on s to simplify the notation. Upon rearranging, it follows that

$$\hat{f}_c(x,s) = \frac{\hat{p}(x \,|\, y, s)}{\hat{p}(c \,|\, y, s)} = \frac{u(x)}{u(c)}.$$

The proof of the alternate case is identical for $y < c < x$ once one spots that $\hat{p}(x \,|\, y, s) = v(x)\tilde{v}(y)$ is a function of x times a function of y. The functions v and $\tilde{v}$ are in general different from u and $\tilde{u}$ as the value c in (2.4.3) will differ. ☐

Proposition 2.4.4. *Let the functions u and v be as in Proposition 2.4.3 and $a < x < b$. Then*

$$\hat{f}^+_{ab}(x,s) = \frac{u(a)v(x) - u(x)v(a)}{u(a)v(b) - u(b)v(a)},$$

$$\hat{f}^-_{ab}(x,s) = \frac{u(x)v(b) - u(b)v(x)}{u(a)v(b) - u(b)v(a)},$$

$$\hat{f}_{ab}(x,s) = \frac{v(x)\big(u(a) - u(b)\big) - u(x)\big(v(a) - v(b)\big)}{u(a)v(b) - u(b)v(a)}.$$

Proof. We begin as in the last proof. Taking the Laplace transform of the equations in identity (2.4.2) provides a set of simultaneous equations for $\hat{f}^+_{ab}$ and $\hat{f}^-_{ab}$. Thus,

$$\hat{f}_a(x,s) = \hat{f}^-_{ab}(x,s) + \hat{f}^+_{ab}(x,s)\hat{f}_a(b,s),$$
$$\hat{f}_b(x,s) = \hat{f}^+_{ab}(x,s) + \hat{f}^-_{ab}(x,s)\hat{f}_b(a,s).$$

Using the expressions for $\hat{f}_c$ in Proposition 2.4.3, and upon solving for $\hat{f}^-_{ab}$ and $\hat{f}^+_{ab}$, the first two identities are obtained. We give a brief derivation of the identity for density $\hat{f}^+_{ab}$. Solving the pair of simultaneous equations gives

$$\begin{aligned}
\hat{f}^+_{ab}(x,s) &= \frac{\hat{f}_b(x,s) - \hat{f}_a(x,s)\hat{f}_b(a,s)}{1 - \hat{f}_a(b,s)\hat{f}_b(a,s)} \\
&= \frac{\dfrac{v(x)}{v(b)} - \dfrac{u(x)}{u(a)}\dfrac{v(a)}{v(b)}}{1 - \dfrac{u(b)}{u(a)}\dfrac{v(a)}{v(b)}} \\
&= \frac{u(a)v(x) - u(x)v(a)}{u(a)v(b) - u(b)v(a)}.
\end{aligned}$$

The final identity is shown by considering the relation between the three densities:

$$\begin{aligned}
\hat{f}_{ab}(x,s) &= \hat{f}^+_{ab}(x,s) + \hat{f}^-_{ab}(x,s) \\
&= \frac{\big(u(a)v(x) - v(a)u(x)\big) + \big(v(b)u(x) - u(b)v(x)\big)}{u(a)v(b) - u(b)v(a)} \\
&= \frac{v(x)\big(u(a) - u(b)\big) - u(x)\big(v(a) - v(b)\big)}{u(a)v(b) - u(b)v(a)}.
\end{aligned}$$ □

Proposition 2.4.4 will be especially useful when finding the solution to the value function in Chapter 3. Next, we explore the equation which $\hat{p}$ satisfies when taking the Laplace transform with respect to the time variable. We have already shown that the transition density p can be found as the solution to Kolmogorov's backward equation; see system (2.2.6). It turns out that we can transform the KBE to get another PDE system for the Laplace transform $\hat{p}$. The infinitesimal generator terms can be transformed by exchanging the order of differentiation with respect to x and integration with respect to t by continuity of p to yield terms in $\hat{p}$. Hence, the remaining object we need to transform is the time derivative of p. [Bateman, 1954, page 129] provides a formula for this, claiming that

$$\mathcal{L}\left\{\frac{\partial f(t)}{\partial t}\right\} = s\hat{f}(s) - f(0).$$

Consider the same system as in (2.2.6) with two boundary conditions on the x-domain. For $x, y \in \mathbb{R}$ and $t \geq 0$, suppose that

$$\begin{aligned}
\frac{\partial p}{\partial t}(x\,|\,y,t) &= \mathcal{A}p(x\,|\,y,t), & p(-\infty\,|\,y,t) &= 0, \\
p(x\,|\,y,0) &= \delta_y(x), & p(\infty\,|\,y,t) &= 0,
\end{aligned} \tag{2.4.5}$$

where we assume that the transition probability density function decays if the process starts very far away from the y-coordinate. These boundary conditions may not always be true, but they certainly hold for an arithmetic Brownian motion, geometric Brownian motion and the Ornstein-Uhlenbeck; see Appendix Section A.1. Taking $s \in \mathbb{R}$, the Laplace transform of system (2.4.5) for $x \neq y$ becomes

$$s\hat{p}(x\,|\,y,s) = \mathcal{A}\hat{p}(x\,|\,y,s), \qquad \begin{aligned} \hat{p}(-\infty\,|\,y,s) &= 0, \\ \hat{p}(\infty\,|\,y,s) &= 0, \end{aligned} \tag{2.4.6}$$

because for $x \neq y$,

$$\mathcal{L}\left\{\frac{\partial p(x\,|\,y,t)}{\partial t}\right\} = s\hat{p}(x\,|\,y,s) - (x\,|\,y,0) = s\hat{p}(x\,|\,y,s).$$

By definition of the Laplace transform, the two boundary conditions are also zero in the complex frequency domain. The final initial condition $p(x\,|\,y,0) = \delta_y(x)$ cannot be transformed, since the transform is in the time variable. Instead, this condition becomes encoded in the new problem via the transform of the time derivative.

The result from Proposition 2.4.3 allow us to translate these boundary conditions into conditions on u and v. For the case $x < y$, we had $\hat{p}(x\,|\,y,s) = u(x)\tilde{u}(y)$. As $x \to -\infty$, by arbitrariness of $\tilde{u}(y)$ it must be that $u(-\infty) = 0$. Similarly, as $x \to \infty$, the condition $v(\infty) = 0$ must hold. By equation (2.4.3), as

$$\hat{p}(x\,|\,y,s) = \hat{f}_c(x,s)\,\hat{p}(c\,|\,y,s)$$

is a function of x multiplied by a function of y and the derivatives in PDE system (2.4.6) are with respect to x, we can factor out the dependence on y to also get a PDE system for $\hat{f}_c$. The domain for f_c is either $(-\infty, c)$ or (c, ∞) depending on the starting value x and the function is degenerate otherwise. We now claim that the boundary condition at barrier level $x = c$ is given by $\hat{f}_c(c,s) = 1$. If the process X starts at $x = c$, then

$$\mathbb{P}(T_c(c) \geq t) = \begin{cases} 1, & t = 0, \\ 0, & t > 0, \end{cases}$$

by definition, since the process has left its open domain and thus $T_c(c) = 0$. Loosely speaking, the FPT density at the boundary is

$$-\delta(t) = \frac{\partial}{\partial t}\mathbb{P}(T_c(c) \geq t) = \frac{\partial}{\partial t}\Big(1 - \mathbb{P}(T_c(c) < t)\Big) = -f_c(c,t),$$

where δ is the Dirac delta distribution. Hence, the boundary condition for the Laplace transformed density at $x = c$ is

$$\hat{f}_c(c,s) = \int_0^\infty \delta(t)\,\mathrm{e}^{-st}\,dt = 1,$$

as required. By Proposition 2.4.3, one can also deduce that $\hat{f}_c(-\infty,s) = 0$ for $x < c$ or $\hat{f}_c(\infty,s) = 0$ for $x < c$. Putting this all together, the PDE system for $x < c$ is

$$s\hat{f}_c(x,s) = \mathcal{A}\hat{f}_c(x,s), \qquad \begin{aligned} \hat{f}_c(c,s) &= 1, \\ \hat{f}_c(-\infty,s) &= 0, \end{aligned} \tag{2.4.7}$$

and the corresponding system for $x > c$ can be found in the same way. Although not all the conditions of Theorem 2.3.6 are met, there is a close similarity and, it follows by plugging in the above components into formula (2.3.3) that

$$\hat{f}_c(x,s) = \mathcal{L}\big\{f_c(x,t)\big\} = \mathbb{E}^x\left[\mathrm{e}^{-sT_c}\right] \tag{2.4.8}$$

is the unique solution to the Dirichlet problem with $r(x) \equiv s$ and $f \equiv \hat{f}_c$. Alternatively, by Remark 2.3.5, we can view the expectation in (2.4.8) as the Laplace transform of a random variable $T_c(x)$, giving us two ways of viewing the conditional expectation.

Proceeding in a similar fashion, let us now consider the solution to Dirichlet's problem for $\hat{f}_{ab}^-$:

$$\begin{aligned} s\hat{f}_{ab}^-(x,s) &= \mathcal{A}\hat{f}_{ab}^-(x,s), & x &\in D, \\ \hat{f}_{ab}^-(x,s) &= \mathbb{1}_{\{T_a(x) < T_b(x)\}}, & x &\in \partial D, \end{aligned} \tag{2.4.9}$$

with $D = (a, b)$ being an open, bounded interval and the boundary $\partial D = \{a, b\}$. On the boundary, we have $\hat{f}_{ab}^{-}(a, s) = 1$ and $\hat{f}_{ab}^{-}(b, s) = 0$. The solution to the PDE is

$$\hat{f}_{ab}^{-}(x, s) = \mathbb{E}^{x}\left[\mathrm{e}^{-s\,T_a}\,\mathbb{1}_{\{T_a < T_b\}}\right] \tag{2.4.10}$$

as given by the stochastic solution to the Dirichlet problem. It is not too hard to convince oneself that the right-hand side of equation (2.4.10) is also a Laplace transform. Take the density $f_{ab}^{-}(x, t) = \mathbb{P}\big(T_{ab}(x) = t, T_{ab}(x) = T_a(x)\big)$ which is a joint density for $\big(T_{ab}(x), T_a(x)\big)$. The expected value can be recasted as

$$\begin{aligned}
\mathbb{E}^{x}\left[\mathrm{e}^{-s\,T_a}\,\mathbb{1}_{\{T_a < T_b\}}\right] &= \mathbb{E}^{x}\left[\mathrm{e}^{-s\,T_{ab}}\,\mathbb{1}_{\{T_{ab} = T_a\}}\right] \\
&= \int_0^\infty \int_0^\infty \mathrm{e}^{-sy}\,\mathbb{1}_{\{y=z\}}\mathbb{P}(T_{ab} = y, T_a = z \mid X_0 = x)\,dy\,dz \\
&= \int_0^\infty \mathrm{e}^{-sy}\,\mathbb{P}(T_{ab} = y, T_a = y \mid X_0 = x)\,dy \\
&= \mathcal{L}\big\{f_{ab}^{-}(x, t)\big\}.
\end{aligned}$$

Again, the nice connection on the conditional expectation exists between Dirichlet's PDE and Laplace transformations. The same analysis can be carried out for $\hat{f}_{ab}^{+}$ by a repeat argument.

Given the significance of the functions f_c and $f_{a,b}$, it is essential to know how to calculate them. Expressions in terms of the functions u and v have been formulated in this section, but the final part we need to address is how to find u and v. Based on the Laplace transform discussion of the KBE and identity (2.4.4), both functions are independent solutions of the Sturm-Liouville problem with Dirichlet boundary conditions. For example, $\hat{f}_c(x, s) \propto u(x)$ for all $x < c$ and so u must also satisfy Dirichlet's PDE. More details on this can be found in [Darling et al., 1953, Theorem 4.1, page 7].

Chapter 3

Optimal Trade Execution and Timing Problems

This chapter is the crux of the thesis and relates our mathematical theory to financial StatArb strategies. An optimal double-stopping problem is constructed on a price process tracking the value of a portfolio and the problem is split into two halves. The first half assesses when it is optimal to enter a trade; the second decides when it is best to close the trade. There are two perspectives a trader can take on the optimal execution problem. They can either pursue a long position to buy a portfolio or hold a short position to sell a portfolio. These problems are treated separately initially, with the hope of tying them together into a joint problem that determines whether to take a long or short position. Sections 3.1 and 3.2 deal with the restrictive single-sided strategies, whilst Section 3.3 addresses the flexible joint problem. We also mention how an exit stop-loss can be incorporated into the problem to make the model more realistic.

Our work has been motivated from the paper by [Leung and Li, 2015], who provide extensive coverage of the long optimal double-stopping problem for the Ornstein-Uhlenbeck process. We believe this chapter covers their theoretical framework in greater generality, allowing an application to more general Itô price processes. The benefit of extending their framework comes at a cost in the form of assumptions; these must be substantiated on a case-by-case basis for the chosen price process. However, we provide evidence in the thesis that the proposed properties hold at least for the ABM and GBM as well as the OU process.

3.1 One-Sided Optimal Stopping Problem

Let $X = (X_t)_{t\geq 0}$ be a price process modelling the value of a basket of assets and assume that it is a diffusion process satisfying the conditions from Section 2.1. In addition, let $c \in \mathbb{R}$, $r > 0$ and $\hat{c} \in \mathbb{R}$, $\hat{r} > 0$ be the transaction costs incurred and interest rate parameters for the exit and entry problems respectively. Depending on the duration of a trade from start to end, these parameters have the scope to change and so it may be that $c \neq \hat{c}$ and $r \neq \hat{r}$. We can include this market feature in our model but have to impose that $c + \hat{c} > 0$ and $0 < \hat{r} \leq r$.

We now look at the optimal trading problem, first describing a long position for an investor, and then the opposite short position. As we will see, it is natural to set out the exit problem first and use it to write down the entry problem.

3.1.1 The Initial Problem: Long Position

Assume an investor has already entered a trade, taking a long position in the selected assets. If they decide to liquidate their holdings at time τ, then their portfolio will be worth X_τ. The fee for the trade will also cost c assuming instantaneous execution. Hence, the time-zero value of the collection of assets X is given by

$$V_L^{(\tau)}(x) = \mathbb{E}^x\left[\mathrm{e}^{-r\tau}(X_\tau - c)\right],$$

under continuous discounting at exit rate r. We call this the exit performance criteria of the investor. For all $\mathbb{F}$-measurable stopping times $\tau \geq 0$, the exit value function is

$$V_L(x) = \sup_{\tau} \mathbb{E}^x \left[e^{-r\tau} (X_\tau - c) \right], \tag{3.1.1}$$

which gives the best possible profit for exiting a long position on the securities with current value x. Note by stationarity of diffusion processes, the performance function and hence value function are both independent of time. We refer to (3.1.1) as the optimal exit problem for a long portfolio position.

Next, we present the entry problem which is a function of the exit problem V. If we decide to enter the trade at time ν, then the assets will cost X_ν and a penalty fee of $\hat{c}$ will be paid. The expected payoff for the portfolio at this time will be $V(X_\nu)$. Therefore, the entry performance criteria of the investor is

$$J_L^{(\nu)}(x) = \mathbb{E}^x \left[e^{-\hat{r}\nu} \big(V_L(X_\nu) - X_\nu - \hat{c} \big) \right],$$

discounted at entry rate $\hat{r}$. For all $\mathbb{F}$-measurable stopping times $\nu \geq 0$, the entry value function is

$$J_L(x) = \sup_{\nu} \mathbb{E}^x \left[e^{-\hat{r}\nu} \big(V_L(X_\nu) - X_\nu - \hat{c} \big) \right], \tag{3.1.2}$$

which gives the optimal time-zero expected profit for the whole round-trip trade. By the set-up of the problem, the optimal entry time always precedes the optimal exit time and so there is no overlap of time in the two problems, ie. we always enter the trade before exiting. Equation (3.1.2) is referred to as the optimal entry problem for a long portfolio position.

To integrate the stop-loss component, we define the stopping time τ_K which is the time when the process X hits a predetermined level $K \in \mathbb{R}$. Recall that this is a true stopping time, as discussed in Section 2.3. The first passage time is given by

$$\tau_K = \inf\{t \geq 0 : X_t \leq K\}.$$

The stop-loss is an upper bound on the process X because we want to avoid a severe drop in value of our portfolio. Under the addition of a stop-loss constraint, the optimal entry and exit problems are altered in the following way:

$$\begin{aligned} J_L^*(x) &= \sup_{\nu} \mathbb{E}^x \left[e^{-\hat{r}\nu} \big(V_L^*(X_\nu) - X_\nu - \hat{c} \big) \right], \\ V_L^*(x) &= \sup_{\tau} \mathbb{E}^x \left[e^{-r(\tau \wedge \tau_K)} (X_{(\tau \wedge \tau_K)} - c) \right]. \end{aligned} \tag{3.1.3}$$

Before attempting to solve for each value function, it is worth identifying properties of the solution. As noted in [Leung and Li, 2015, page 6], both V_L and J_L are non-negative since $\tau = \nu = \infty$ is a valid stopping time. This could be the case for a bounded price process, where it would never be optimal to enter a trade. Also, the exit stop-loss constraint restricts the set of possible maximal stopping times. As a result, our value functions satisfy $x - c \leq V_L^*(x) \leq V_L(x)$ and $0 \leq J_L^*(x) \leq J_L(x)$.

3.1.2 The Reverse Problem: Short Position

In light of the previous subsection, the treatment of the flipped problem will be brief. We consider an investor holding a short position in a basket of assets. Under the same set up as before, the exit performance criteria becomes

$$V_S^{(\tau)}(x) = \mathbb{E}^x \left[e^{-r\tau} (-X_\tau - c) \right],$$

and the exit value function is

$$V_S(x) = \sup_{\tau} \mathbb{E}^x \left[e^{-r\tau} (-X_\tau - c) \right], \tag{3.1.4}$$

as the investor will acquire the portfolio when closing a position and pay a fee of c. We refer to (3.1.4) as the optimal exit problem for a short portfolio position. Similarly, the entry performance criteria becomes

$$J_S^{(\nu)}(x) = \mathbb{E}^x \left[e^{-\hat{r}\nu} \big(V_S(X_\nu) + X_\nu - \hat{c} \big) \right],$$

and the entry value function is

$$J_S(x) = \sup_{\nu} \mathbb{E}^x\big[\mathrm{e}^{-\hat{r}\nu}\big(V_S(X_\nu) + X_\nu - \hat{c}\big)\big], \tag{3.1.5}$$

as the investor receives money for selling the portfolio, pays a transaction fee $\hat{c}$ for the trade and expects to incur the cost $V(X_\nu)$ when purchasing the portfolio at exit time ν. Equation (3.1.5) is the optimal entry problem for a short portfolio position. The corresponding problems with an exit stop-loss are:

$$\begin{aligned} J_S^*(x) &= \sup_{\nu} \mathbb{E}^x\big[\mathrm{e}^{-\hat{r}\nu}\big(V_S^*(X_\nu) + X_\nu - \hat{c}\big)\big], \\ V_S^*(x) &= \sup_{\tau} \mathbb{E}^x\big[\mathrm{e}^{-r(\tau\wedge\tau_K)}(-X_{(\tau\wedge\tau_K)} - c)\big], \end{aligned} \tag{3.1.6}$$

where the stop-loss time at level K is now redefined as

$$\tau_K = \inf\{t \geq 0 : X_t \geq K\}.$$

This is because an investor holding a short position would want to avoid a large increase in the value of the portfolio. In the spirit of consistency with the previous subsection, we now see that $0 \leq V_S^*(x) \leq V_S(x)$ and $x - \hat{c} \leq J_S^*(x) \leq J_S(x)$.

Remark 3.1.1. From the outset, it appears as if not much separates the long and short optimal stopping problems. The difference between the two problems comes down to a couple of minus sign changes. For a process with symmetric probability distribution, the connection between both problems is enhanced and this creates convenient shortcuts.

Imagine that X is a process with symmetric probability distribution about zero and define another process Y such that $X_t \overset{d}{=} -X_t := Y_t$. Then the long exit payoff function for X equals the short exit payoff function for Y as

$$X_t - c \overset{d}{=} -X_t - c = Y_t - c$$

and the same is true of the entry payoff function. We hypothesise that it is enough to study only one optimal stopping problem for a centred, symmetric price process; eg. the OU process. For example, one can obtain the short problem by replacing every X term by $Y = -X$ (as they are equal in distribution) in the long problem. The stop-loss constraint would also change from level K to $-K$ when converting between long and short problems.

3.2 Solution to the One-Sided Problems

The traditional procedure to find the solution to the value functions in Section 3.1 would involve solving variational inequalities from the dynamic programming principle, producing a verification argument to ensure the solution is maximal. Here we avoid this laborious route, instead maximising the double-stopping problem over all open intervals (a, b) as suggested by [Leung and Li, 2015]. One bonus of using this method is that it reveals the form of the value function, unlike the DPP set up. Analysis for the stop-loss is omitted but can be found [Leung and Li, 2015, Section 5, page 13].

3.2.1 The Initial Problem Solution: Long Position

Consider the Sturm-Liouville problem defined as

$$\mathcal{A}w(x) = rw(x), \qquad x \in \mathbb{R} \tag{3.2.1}$$

where we take the function $r(x) \equiv r \geq 0$ to be the exit interest rate and the infinitesimal generator, where we assume it exists, for $X = (X_t)_{t\geq 0}$ acting on w to be

$$\mathcal{A}w(x) = b(x)\,\frac{\partial w(x)}{\partial x} + \frac{1}{2}\,\sigma^2(x)\,\frac{\partial^2 w(x)}{\partial x^2}.$$

As explained at the end of Section 2.4, we assume that two linearly independent solutions to the Sturm-Liouville problem exist, call them $u(x) \equiv u(x; r)$ and $v(x) \equiv v(x; r)$. These solutions take the form of the functions in identity (2.4.4), with Dirichlet boundary conditions $u(-\infty) = 0$ and $v(\infty) = 0$. In addition, we make the following assumptions about the solutions to the partial differential equation.

Assumption 3.2.1 (Properties of u and v).

(i) u and v are both positive and convex

(ii) u is strictly increasing whilst v is strictly decreasing

(iii) u and v are at least twice differentiable and continuous

These assumptions and the ones that follow are motivated by the characteristics of the OU solutions. We provide formulae in Appendix Section A.1 to confirm that these properties hold for the ABM, GBM and OU processes.

The first identity we can verify is the probabilistic relation for functions u and v. By combining the results in Proposition 2.4.3 and equation (2.4.8), we get that

$$\mathbb{E}^x\left[\mathrm{e}^{-rT_k}\right] = \begin{cases} \dfrac{u(x)}{u(k)}, & x \le k, \\ \dfrac{v(x)}{v(k)}, & x \ge k, \end{cases} \tag{3.2.2}$$

as described in [Leung and Li, 2015, Equation 3.5, page 7]. This is a well-known expression for the expected value of a discount factor at rate r and agrees with the classical reference [Itô, 1996, Section 4.6, page 128]. The aim is to take this a step further to try and find an analytic expression when attaching a cost or payoff function to the discount factor.

For process X starting at $X_0 = x$, take $T_a \wedge T_b := T_{a,\infty} \wedge T_{-\infty,b}$ to be its exit time from the interval (a, b), where the interval end points could be infinite, ie. $-\infty \le a < x < b \le \infty$. When $a = -\infty$ or $b = \infty$, what we are describing is the process hitting a single level b or a respectively. Also, introduce the function $h(x) := x - c$, which specifies the payoff function when exiting a long trade. We will now attempt to find an explicit expression for the long exit value function by looking at the interval that contains the initial portfolio price x.

The discounted payoff under expectation is:

$$\begin{aligned} \mathbb{E}^x\left[\mathrm{e}^{-r(T_a\wedge T_b)}h\big(X_{T_a\wedge T_b}\big)\right] &= h(a)\mathbb{E}^x\left[\mathrm{e}^{-r\,T_a}\,\mathbf{1}_{\{T_a< T_b\}}\right] + h(b)\mathbb{E}^x\left[\mathrm{e}^{-r\,T_b}\,\mathbf{1}_{\{T_a>T_b\}}\right] \\ &= h(a)\hat{f}^-_{ab}(x,r) + h(b)\hat{f}^+_{ab}(x,r) \\ &= h(a)\frac{u(x)v(b)-u(b)v(x)}{u(a)v(b)-u(b)v(a)} + h(b)\frac{u(a)v(x)-u(x)v(a)}{u(a)v(b)-u(b)v(a)} \end{aligned} \tag{3.2.3}$$

by expanding over whether $T_a \wedge T_b$ hits barrier level a or b first. The second line follows by equation (2.4.10). We are now ready to maximise the discounted payoff over all intervals (a, b), however, we propose three transformations which will help simplify this computation. Step one is to consider the transform:

$$\psi(x) := \frac{u(x)}{v(x)}, \qquad x \in \mathbb{R}, \tag{3.2.4}$$

which is well-defined as v is always positive. By Assumption 3.2.1 and by properties of a quotient of functions, it follows that ψ is continuous, positive and strictly increasing, so we can safely assume it has an inverse and is a bijective mapping. Applying the transformation, we get

$$\mathbb{E}^x\left[\mathrm{e}^{-r(T_a\wedge T_b)}h\big(X_{T_a\wedge T_b}\big)\right] = v(x)\left[\frac{h(a)}{v(a)}\frac{\psi(b)-\psi(x)}{\psi(b)-\psi(a)} + \frac{h(b)}{v(b)}\frac{\psi(x)-\psi(a)}{\psi(b)-\psi(a)}\right].$$

Steps two and three are intrinsically a relabelling exercise. Take $y = \psi(x)$ and define the more compact mapping

$$H_L(y) := \begin{cases} \dfrac{h}{v} \circ \psi^{-1}(y), & y > 0, \\ \lim\limits_{x\to-\infty} \dfrac{\big(h(x)\big)^+}{v(x)}, & y = 0, \end{cases} \tag{3.2.5}$$

which is also well-defined for all y. The second case ($y = 0$) may appear to be slightly arbitrary and in fact unnecessary as ψ is always positive. However, we will need to extend the function H

to be continuous at zero later on. Under these two transformations with $\psi(x) > 0$ for all x, we have

$$\mathbb{E}^x\left[\mathrm{e}^{-r(T_a\wedge T_b)}h\big(X_{T_a\wedge T_b}\big)\right] = v\big(\psi^{-1}(y)\big)\left[H_L(y_a)\frac{y_b-y}{y_b-y_a} + H_L(y_b)\frac{y-y_a}{y_b-y_a}\right],$$

where $y_a = \psi(a)$ and $y_b = \psi(b)$. To find the optimal interval such that this expectation is maximised, we need to take the supremum over all pairs $\{(a,b) : -\infty \le a < x < b \le \infty\}$. By the strict monotonicity of ψ, this is equivalent to maximising over all pairs $\{(y_a, y_b) : -\infty \le y_a < y < y_b \le \infty\}$. Define the function

$$W_L(y) := \sup_{\{(y_a,y_b):y_a<y<y_b\}}\left[H_L(y_a)\frac{y_b-y}{y_b-y_a} + H_L(y_b)\frac{y-y_a}{y_b-y_a}\right], \qquad y > 0, \tag{3.2.6}$$

which is the smallest concave majorant of H (see Appendix Lemma B.4.3). Then the expression for the optimal expected value is

$$\sup_{\{(a,b):a<x<b\}}\mathbb{E}^x\left[\mathrm{e}^{-r(T_a\wedge T_b)}h\big(X_{T_a\wedge T_b}\big)\right] = v(x)W_L\big(\psi(x)\big). \tag{3.2.7}$$

We claim that equation (3.2.7) is the exit value function for the long problem, that is,

$$V_L(x) = v(x)W_L\big(\psi(x)\big). \tag{3.2.8}$$

Therefore, it would appear that optimising the exit payoff function over all stopping times is equivalent to maximising the exit payoff function over all intervals (a, b). Equation (3.2.7) suggests the best time to liquidate one's holdings is when the price process X leaves the optimal exit interval (a^*, b^*) (more on this later).

Relation (3.2.8) will be verified in Theorem 3.2.2, but before doing so, we will apply the same treatment to the long entry value function J_L. The following text repeats the above derivation with the only difference being that we 'hat' the quantities whilst transferring them to the entry problem variant. One may question the decision to repeat the above arguments as opposed to dealing with both cases in one hit. Though this would be possible initially, we would run into notational issues further on and so it is in our best interests to treat the entry and exit problems separately with two sets of notations.

Suppose instead that two linearly independent solutions $\hat{u}(x) \equiv \hat{u}(x;\hat{r})$ and $\hat{v}(x) \equiv \hat{v}(x;\hat{r})$ exist to the Sturm-Liouville problem (3.2.1) with the discount rate r replaced by its entry problem counterpart $\hat{r}$. These solutions also share the properties of u and v from Assumption 3.2.1. In addition, define the cost function $\hat{h}(x) := V_L(x) - x - \hat{c}$ and the two transformations as in the long exit problem as

$$\hat{\psi}(x) := \frac{\hat{u}(x)}{\hat{v}(x)}, \qquad x \in \mathbb{R}, \tag{3.2.9}$$

which is continuous, positive and strictly increasing and the tidier mapping

$$\hat{H}_L(y) := \begin{cases} \dfrac{\hat{h}}{\hat{v}} \circ \hat{\psi}^{-1}(y), & y > 0, \\ \lim\limits_{x\to-\infty}\dfrac{\big(\hat{h}(x)\big)^+}{\hat{v}(x)}, & y = 0. \end{cases} \tag{3.2.10}$$

With these two transformations in place, by an identical argument to the long exit problem with u, v, ψ and H_L swapped for $\hat{u}, \hat{v}, \hat{\psi}$ and $\hat{H}_L$, we have for $y = \hat{\psi}(x)$ that

$$\mathbb{E}^x\left[\mathrm{e}^{-\hat{r}(T_{\hat{a}}\wedge T_{\hat{b}})}\hat{h}\big(X_{T_{\hat{a}}\wedge T_{\hat{b}}}\big)\right] = \hat{v}\big(\hat{\psi}^{-1}(y)\big)\left[\hat{H}_L(y_{\hat{a}})\frac{y_{\hat{b}}-y}{y_{\hat{b}}-y_{\hat{a}}} + \hat{H}_L(y_{\hat{b}})\frac{y-y_{\hat{a}}}{y_{\hat{b}}-y_{\hat{a}}}\right],$$

where $y_{\hat{a}} = \hat{\psi}(\hat{a})$ and $y_{\hat{b}} = \hat{\psi}(\hat{b})$. Defining the function

$$\hat{W}_L(y) := \sup_{\{(y_{\hat{a}},y_{\hat{b}}):y_{\hat{a}}<y<y_{\hat{b}}\}}\left[\hat{H}_L(y_{\hat{a}})\frac{y_{\hat{b}}-y}{y_{\hat{b}}-y_{\hat{a}}} + \hat{H}_L(y_{\hat{b}})\frac{y-y_{\hat{a}}}{y_{\hat{b}}-y_{\hat{a}}}\right], \qquad y > 0, \tag{3.2.11}$$

it then follows that

$$\sup_{\{(\hat{a},\hat{b}):\hat{a}<x<\hat{b}\}} \mathbb{E}^x\left[\mathrm{e}^{-\hat{r}(T_{\hat{a}}\wedge T_{\hat{b}})}\hat{h}\big(X_{T_{\hat{a}}\wedge T_{\hat{b}}}\big)\right] = \hat{v}(x)\hat{W}_L\big(\hat{\psi}(x)\big), \tag{3.2.12}$$

where $\hat{W}_L$ is the smallest concave majorant for $\hat{H}_L$ (see Appendix Lemma B.4.4). The silver lining is again that we make a likewise claim: the entry value function for the long problem is given by (3.2.12) and

$$J_L(x) = \hat{v}(x)\hat{W}_L\big(\hat{\psi}(x)\big). \tag{3.2.13}$$

So finding the candidate strategy for entering a trade over all possible stopping times is the same as finding the time at which the price process X leave the optimal entry interval $(\hat{a}^*, \hat{b}^*)$. This is when the purchase of assets will be triggered (more on this later). All that is left to do is to verify our claims in equations (3.2.8) and (3.2.13). This is what the next result describes.

Theorem 3.2.2 (Optimality of value functions)**.** *The optimal time to exit the long problem* (3.1.1) *is characterised by finding an optimal interval* (a^*, b^*) *such that when the price process* X *leaves this interval, the investor should sell. Consequently, we claim that*

$$V_L(x) = v(x)W_L\big(\psi(x)\big), \qquad x \in \mathbb{R}.$$

The optimal time to enter the long problem (3.1.2) *is characterised by finding an optimal interval* $(\hat{a}^*, \hat{b}^*)$ *such that when the price process* X *leaves this interval, the investor should buy. Consequently, we claim that*

$$J_L(x) = \hat{v}(x)\hat{W}_L\big(\hat{\psi}(x)\big), \qquad x \in \mathbb{R}.$$

Proof. We only prove the first identity. The proof of the second identity is identical by hatting the quantities. By Lemma 2.3.2, we know that $T_a \wedge T_b$ is a stopping time and so

$$V_L(x) = \sup_{\tau} \mathbb{E}^x\left[\mathrm{e}^{-r\tau}h(X_\tau)\right] \geq \sup_{\{(a,b):a<x<b\}} \mathbb{E}^x\left[\mathrm{e}^{-r(T_a\wedge T_b)}h\big(X_{T_a\wedge T_b}\big)\right] = v(x)W_L\big(\psi(x)\big).$$

The proof of the reverse inequality is a little more involved. W_L is concave so by Appendix Lemma B.4.2, for any $y_0 > 0$ there exists an affine majorant F that supports W_L at point y_0. In mathematical terms, this means

(majorant)	$W_L(y) \leq F(y),$	$y > 0,$
(supporting)	$W_L(y_0) = F(y_0),$	$y_0 > 0,$
(affine)	$F(y) = ay + b,$	$a, b \in \mathbb{R},$

where it is important to note that the function $F := F_{y_0}$ and coefficients $a := a_{y_0}$ and $b := b_{y_0}$ depend on the choice of y_0. Take any stopping time $\nu \in \mathbb{F}$ and $t \in \mathbb{R}_{\geq 0}$ and define $\tau := t \wedge \nu$. Also, let $y_0 := \psi(x)$ be the supporting boundary point of W_L. For payoff function h,

$$\begin{aligned} h(X_\tau) &= v(X_\tau)H_L\big(\psi(X_\tau)\big) \\ &\leq v(X_\tau)W_L\big(\psi(X_\tau)\big) \\ &\leq v(X_\tau)F\big(\psi(X_\tau)\big) \\ &= a\, v(X_\tau)\psi(X_\tau) + b\, v(X_\tau) \\ &= a\, u(X_\tau) + b\, v(X_\tau). \end{aligned}$$

Taking the discounted expectation, continuing on from the above expression gives

$$\begin{aligned} \mathbb{E}^x\left[\mathrm{e}^{-r\tau}h(X_\tau)\right] &\leq a\,\mathbb{E}^x\left[\mathrm{e}^{-r\tau}u(X_\tau)\right] + b\,\mathbb{E}^x\left[\mathrm{e}^{-r\tau}v(X_\tau)\right] \\ &= au(x) + bv(x) \qquad (3.2.14) \\ &= v(x)\left[a\frac{u(x)}{v(x)} + b\right] \\ &= v(x)F\big(\psi(x)\big) \\ &= v(x)W_L\big(\psi(x)\big), \end{aligned}$$

where line (3.2.14) is explained by an application of the optional stopping theorem (as τ is a bounded stopping time) and that $\big(e^{-rt}u(X_t)\big)_{t\geq 0}$ and $\big(e^{-rt}v(X_t)\big)_{t\geq 0}$ are both martingales. Taking the supremum over all times t and then ν, we get that $V_L(x) \leq v(x)W_L\big(\psi(x)\big)$ and so we have equality.

Verifying the martingale claim is consigned to Appendix Lemma B.3.4, but boils down to an application of Itô's lemma. An interesting argument can also be found in [Kessler et al., 1999, Section 5, page 312], which also includes information about the eigenfunction characterisation of solutions to the Sturm-Liouville problem. □

So far we have established the alternative light that the value function problems can be viewed in but have not yet found an explicit expression for them. We will now lay out some further assumptions relating to the properties of H_L and $\hat{H}_L$ that will allow us to derive exact expressions for both value functions. This is where the treatment of both problems begins to diverge and so we deal with the exit value function V_L first, since the entry value function J_L is posed in terms of V_L.

Assumption 3.2.3 (Properties of H_L)**.**

(i) H_L is continuous on $[0,\infty)$ and twice differentiable on $(0,\infty)$.

(ii) $H_L(0) = 0$ and for exit transactional cost c,

$$H_L(y) = \begin{cases} < 0 & \text{if } y \in (0, \psi(c)), \\ > 0 & \text{if } y \in (\psi(c), \infty), \end{cases}$$

with $H_L\big(\psi(c)\big) = 0$.

(iii) Given x^* as the solution to $v(x) - (x-c)v'(x) = 0$, then

$$H_L'(y) = \begin{cases} < 0 & \text{if } y \in (0, \psi(x^*)), \\ > 0 & \text{if } y \in (\psi(x^*), \infty), \end{cases}$$

where $x^* < c$. In addition, $H_L'\big(\psi(x^*)\big) = 0$ and $H_L'(y) \to 0$ as $y \to \infty$.

(iv) There exists a constant L^* such that

$$H_L''(y) = \begin{cases} \geq 0 & \text{if } y \in (0, \psi(L^*)], \\ \leq 0 & \text{if } y \in [\psi(L^*), \infty). \end{cases}$$

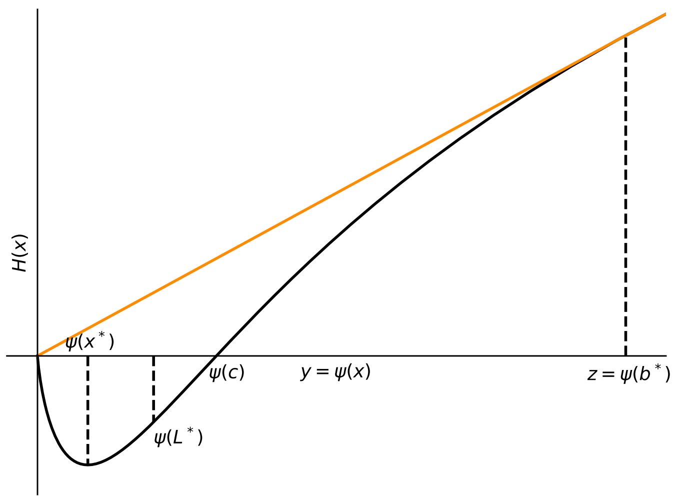

Figure 3.1: Properties of H_L

We pose these properties as an assumption rather than a lemma or proposition as they may not always hold for arbitrary Itô diffusions. Our attempt to justify these assumptions in the general case can be found in Appendix Section B.5.1. Thankfully for our select three diffusion processes, these properties can be verified. A proof is constructed for an ABM in Appendix Section B.5.3, the OU process is covered in the paper by [Leung and Li, 2015] and the GBM is left as an exercise with the steps mirroring those in the ABM example. Armed with the attributes of H_L, we can now find the exit value function V_L.

Theorem 3.2.4. *The optimal exit problem* (3.1.1) *for a long position is given by*

$$V_L(x) = \begin{cases} (b^* - c)\dfrac{u(x)}{u(b^*)}, & x \in (-\infty, b^*), \\ x - c, & x \in [b^*, \infty), \end{cases} \tag{3.2.15}$$

where the constant $b^* > c \wedge L^*$, which is the optimal liquidation level, is the solution to the equation

$$u(b) = (b - c)u'(b) \tag{3.2.16}$$

and the optimal liquidation time is

$$\tau^* = \inf\{t \geq 0 : X_t \geq b^*\}.$$

Proof. The presentation of this proof is not entirely rigorous and the derivation for the form of W_L may appear to be abstract. We do this to avoid complicated geometrical arguments, hoping that Figure 3.1 is enough to convince the reader.

From the properties of H_L and the sketch, we conclude that the smallest concave majorant of H_L must be a two part function. The two sections will meet at some value $y = z$, where $z > \psi(c) \wedge \psi(L^*)$. The first part of this function will be a line segment connecting the coordinates $(0, H_L(0))$ and $(z, H_L(z))$ and the second part must trace H_L exactly, since H_L is already concave in this region, ie.

$$W_L(y) = \begin{cases} y\dfrac{H_L(z)}{z}, & y \leq z, \\ H_L(y), & y \geq z. \end{cases} \tag{3.2.17}$$

The connecting point z must lie beyond both of $y = \psi(L^*)$ and $y = \psi(c)$, otherwise a contradiction can be found to W_L being concave. Since $\lim_{y\to\infty} H_L'(y) = 0$, the line segment $y\frac{H_L(z)}{z}$ must be tangential to H_L at $y = z$. Therefore, z must satisfy

$$\frac{H_L(z) - H_L(0)}{z - 0} = \frac{H_L(z)}{z} = H_L'(z). \tag{3.2.18}$$

We want to make use of Theorem 3.2.2 which is stated in terms of the initial starting value x. Therefore, we transform variables from the y to x domain. Taking $z := \psi(b^*)$, the left-hand side of equation (3.2.18) becomes

$$\frac{H_L(z)}{z} = \frac{H_L\big(\psi(b^*)\big)}{\psi(b^*)} = \frac{h(b^*)}{\psi(b^*)v(b^*)} = \frac{b^* - c}{u(b^*)}.$$

The right-hand side can be expressed as

$$H_L'(z) = \frac{v(b^*) - (b^* - c)v'(b^*)}{u'(b^*)v(b^*) - u(b^*)v'(b^*)},$$

where the derivative of H_L is known from Appendix equation (B.5.2). Equating the two expressions leads to the identity

$$u(b^*) = (b^* - c)u'(b^*)$$

for b^* after further cancelling. Moreover, taking $y = \psi(x)$, the smallest concave majorant W_L can be expressed in terms of x as

$$W_L\big(\psi(x)\big) = \begin{cases} \dfrac{u(x)}{v(x)}\dfrac{b^* - c}{u(b^*)}, & x \leq b^*, \\ \dfrac{x - c}{v(x)}, & x \geq b^*. \end{cases}$$

The value function then follows by identity (3.2.8), where $V_L(x) = v(x)W_L\big(\psi(x)\big)$. □

Finally, the long exit value function V_L has been found. An explication of the result is now given to relate the solution to the actions that a trader would make based on the value of the selected securities. Reminding ourselves of equation (3.2.3), for $x \in (a, b)$ we had

$$\mathbb{E}^x\left[\mathrm{e}^{-r(T_a \wedge T_b)} h\big(X_{T_a \wedge T_b}\big)\right] = h(a)\mathbb{E}^x\left[\mathrm{e}^{-r\,T_a}\,\mathbb{1}_{\{T_a < T_b\}}\right] + h(b)\mathbb{E}^x\left[\mathrm{e}^{-r\,T_b}\,\mathbb{1}_{\{T_a > T_b\}}\right].$$

Choosing $a = -\infty$ and $b = b^*$, where b^* is the solution to equation (3.2.16), our optimal interval is $\mathcal{I}_V := (-\infty, b^*)$ and we get

$$\mathbb{E}^x\left[\mathrm{e}^{-r(T_{-\infty} \wedge T_{b^*})} h\big(X_{T_{-\infty} \wedge T_{b^*}}\big)\right] = \mathbb{E}^x\left[\mathrm{e}^{-r\,T_{b^*}}\,h\big(X_{T_{b^*}}\big)\right] = h(b^*)\mathbb{E}^x\left[\mathrm{e}^{-r\,T_{b^*}}\right].$$

If $x < b^*$, then $x \in \mathcal{I}_V$ and $h(b^*) = b^* - c$. We can calculate the expectation to be

$$\mathbb{E}^x\left[\mathrm{e}^{-r\,T_{b^*}}\,h\big(X_{T_{b^*}}\big)\right] = (b^* - c)\frac{u(x)}{u(b^*)},$$

where we make use of formula (3.2.2) to compute the expected value of the discount factor. On the other hand, if $x \geq b^*$ then $x \notin \mathcal{I}_V$ and

$$\mathbb{E}^x\left[\mathrm{e}^{-r(T_{-\infty} \wedge T_{b^*})} h\big(X_{T_{-\infty} \wedge T_{b^*}}\big)\right] = x - c$$

because the process X started outside of $\mathcal{I}_V$, so $T_{-\infty} \wedge T_{b^*} = T_{b^*} := T_{-\infty, b^*} = 0$. The interpretation in the context of optimal trading is now elucidated:

If the portfolio value lies inside the optimal interval, ie. $x \in \mathcal{I}_V$, then the trader should liquidate their holdings only when the portfolio value increases to level b^*. Premature liquidation would be suboptimal. On the other hand, if the value of the traded assets are in excess of price b^*, ie. $x \notin \mathcal{I}_V$, then it is optimal to sell immediately. This explanation characterises b^* as the optimal liquidation level and $T_{b^*} = T_{-\infty, b^*}$ as the optimal liquidation time.

The following result analyses the impact of the transaction cost on the problem. The impact of the other parameters is not so clear cut but a plot of the dependence structure can be produced to evaluate this.

Proposition 3.2.5. *The value function V_L is decreasing in the exit transaction cost c and the optimal barrier level b^* is increasing in c.*

Proof. To show that V_L is a decreasing function of c, it is enough to realise that the payoff function

$$\mathbb{E}^x[\mathrm{e}^{-r\tau}(X_\tau - c)] = \mathbb{E}^x[\mathrm{e}^{-r\tau}X_\tau] - c\,\mathbb{E}^x[\mathrm{e}^{-r\tau}]$$

is decreasing in c. So if $c_1 < c_2$, then

$$V_L^{(\tau)}(x; c_1) := \mathbb{E}^x[\mathrm{e}^{-r\tau}(X_\tau - c_1)] \geq \mathbb{E}^x[\mathrm{e}^{-r\tau}(X_\tau - c_2)] =: V_L^{(\tau)}(x; c_2).$$

By taking the supremum over all stopping times, we see that $V_L(x; c_1) \geq V_L(x; c_2)$ and so V_L is decreasing in c. To show that b^* is increasing in c, we assess the derivative of $b^*(c)$. By differentiating identity (3.2.16) with respect to c, we obtain that

$$(b^*)'(c) = \frac{u'(b^*)}{(b^* - c)u''(b^*)} > 0$$

which proves the result. □

This result confirms the obvious for a more expensive trade: a smaller profit margin is made and the sell barrier level increases. The entry problem is tackled next with a brief but alike derivation.

Assumption 3.2.6 (Properties of $\hat{H}$)**.** Let V_L, b^* and L^* be as in Theorem 3.2.4. Then we assume

(i) $\hat{H}_L$ is continuous on $[0, \infty)$, differentiable on $(0, \infty)$ and twice differentiable on $(0, \hat{\psi}(b^*)) \cup (\hat{\psi}(b^*), \infty)$.

(ii) $\hat{H}_L(0) = 0$ and for constant $k_* < b^*$, where k_* is the unique solution to $\hat{h}(x) = 0$,

$$\hat{H}_L(y) = \begin{cases} > 0 & \text{if } y \in (0, \hat{\psi}(k_*)), \\ < 0 & \text{if } y \in (\hat{\psi}(k_*), \infty), \end{cases}$$

with $\hat{H}_L\big(\hat{\psi}(k_*)\big) = 0$.

(iii) $\hat{H}_L$ is strictly decreasing for $y \in (\hat{\psi}(b^*), \infty)$.

(iv) There exists a constant $L_* < b^*$ such that

$$\hat{H}_L''(y) = \begin{cases} \leq 0 & \text{if } y \in (0, \hat{\psi}(L_*)], \\ \geq 0 & \text{if } y \in [\hat{\psi}(L_*), \infty). \end{cases}$$

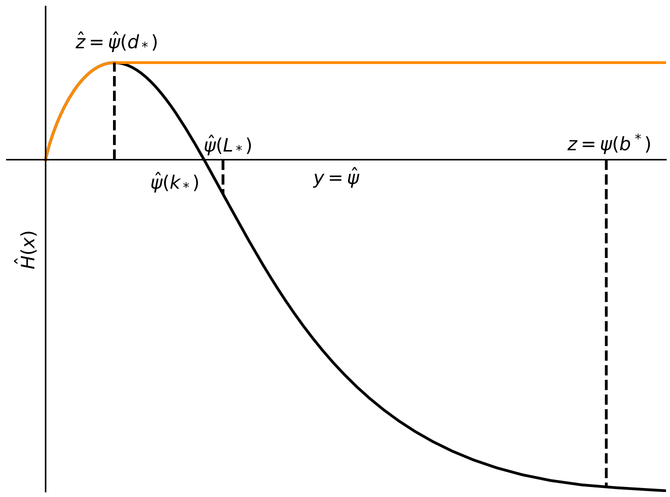

Figure 3.2: Properties of $\hat{H}_L$

In an identical manner, these properties are treated as an assumption rather than fact. Our attempt to justify the properties can be found in Appendix Section B.5.2. The case specific application to the ABM resides in Appendix Section B.5.3. We can now go on to find the exit value function J_L.

Theorem 3.2.7. *The optimal entry problem* (3.1.2) *for a long position is given by*

$$J_L(x) = \begin{cases} V_L(x) - x - \hat{c}, & x \in (-\infty, d_*], \\ \dfrac{V_L(d_*) - d_* - \hat{c}}{\hat{v}(d_*)}\,\hat{v}(x), & x \in (d_*, \infty), \end{cases} \tag{3.2.19}$$

where the constant $d_* < k_* \wedge L_*$, which is the optimal entry level, is the solution to the equation

$$\hat{v}(d)\big(V_L'(d) - 1\big) = \hat{v}'(d)\big(V_L(d) - d - \hat{c}\big) \tag{3.2.20}$$

and the optimal liquidation time is

$$\nu_* = \inf\{t \geq 0 : X_t \leq d_*\}.$$

Proof. As previously discussed, the presentation of this proof is best supplemented with the plot in Figure 3.2. A precise proof would be too cumbersome given the target of this thesis.

From the properties of $\hat{H}_L$ and the sketch, the smallest concave majorant of $\hat{H}_L$ must be a two part function. These two sections will meet at some value $y = \hat{z}$, a global maxima of $\hat{H}_L$. The first part will trace the concave part of $\hat{H}_L$ up to the maximum value $\hat{z}$ and the second part is a horizontal line segment, ie.

$$\hat{W}_L(y) = \begin{cases} \hat{H}_L(y), & y \leq \hat{z}, \\ \hat{H}_L(\hat{z}), & y \geq \hat{z}. \end{cases} \tag{3.2.21}$$

By definition of being a global maximum, $\hat{z} < \hat{\psi}(k_*) < \hat{\psi}(b^*)$. In fact, $\hat{z}$ must be bounded above by both $y = \hat{\psi}(k_*)$ and $y = \hat{\psi}(L_*)$. If not, a contradiction can be deduced to $\hat{W}_L$ being concave. A key property of being a maxima is that

$$\hat{H}_L'(\hat{z}) = 0.$$

Suppose $\hat{z} = \hat{\psi}(d_*)$. Then d_* must satisfy

$$\hat{H}_L'(\hat{z}) = \frac{\hat{h}'(d_*)\hat{v}(d_*) - \hat{h}(d_*)\hat{v}'(d_*)}{\hat{u}'(d_*)\hat{v}(d_*) - \hat{u}(d_*)\hat{v}'(d_*)} = 0,$$

by referring to the expression for H_L' from Appendix equation (B.5.1). By definition of $\hat{h}$, this identity becomes

$$\frac{(V_L'(d_*) - 1)\hat{v}(d_*) - (V_L(d_*) - d_* - \hat{c})\hat{v}'(d_*)}{\hat{u}'(d_*)\hat{v}(d_*) - \hat{u}(d_*)\hat{v}'(d_*)} = 0.$$

The denominator is always positive and so equating the numerator to zero yields equation (3.2.20), which can be solved to find the candidate level d_*. Moreover, taking $y = \hat{\psi}(x)$, the smallest concave majorant $\hat{W}_L$ can be displayed in terms of x as

$$\hat{W}_L\big(\hat{\psi}(x)\big) = \begin{cases} \dfrac{V_L(x) - x - \hat{c}}{\hat{v}(x)}, & x \leq d_*, \\ \dfrac{V_L(d_*) - d_* - \hat{c}}{\hat{v}(d_*)}, & x \geq d_*, \end{cases}$$

by definition of $\hat{h}$. The statement of the theorem then follows by an application of identity (3.2.13) with $J_L(x) = \hat{v}(x)\hat{W}_L\big(\hat{\psi}(x)\big)$. □

This completes the task of finding the long entry value function J_L. Like before, we can choose the optimal interval to be $\mathcal{I}_J := (d_*, \infty)$, where d_* is the solution to equation (3.2.20). Then for $x \in \mathcal{I}_J$,

$$\mathbb{E}^x\Big[e^{-\hat{r}(T_{d_*} \wedge T_\infty)}\hat{h}\big(X_{T_{d_*} \wedge T_\infty}\big)\Big] = \mathbb{E}^x\Big[e^{-\hat{r}\,T_{d_*}}\hat{h}\big(X_{T_{d_*}}\big)\Big] = \hat{h}(d_*)\mathbb{E}^x\Big[e^{-\hat{r}\,T_{d_*}}\Big] = \frac{V_L(d_*) - d_* - \hat{c}}{\hat{v}(d_*)}\,\hat{v}(x)$$

and conversely for $x \notin \mathcal{I}_J$,

$$\mathbb{E}^x\Big[e^{-\hat{r}(T_{d_*} \wedge T_\infty)}\hat{h}\big(X_{T_{d_*} \wedge T_\infty}\big)\Big] = V_L(x) - x - \hat{c}.$$

The strategy behind entering a long trade in an optimal fashion is now explained:

By the set up of the problem, $d_* < b^*$ so the strategy is well-defined and non-overlapping. A trader will only enter a long position once the initial portfolio value x goes below the level d_*, ie. $x \notin \mathcal{I}_J$. In this case, the trader should instantly buy and hold their selected assets modelled by price process X. On the contrary, if the initial portfolio value is contained inside the optimal interval, ie. $x \in \mathcal{I}_J$, then the portfolio is over-valued. When this occurs, the trader should wait until the portfolio value falls below the barrier level d_* before purchasing any assets. If and when the portfolio rises in value to b^*, the portfolio should be sold to cash in the profit.

Lastly, the impact of the transaction cost $\hat{c}$ is assessed in the next result. Instinctively, a higher fee would mean the trader would want to pay a lower cost to enter the trade and so we should expect d_* to decrease with increasing $\hat{c}$.

Proposition 3.2.8. *The optimal barrier level d_* is decreasing in the entry transaction cost $\hat{c}$.*

Proof. The method of proof relies on differentiating equation (3.2.20). We refer the reader to [Leung and Li, 2015, Proposition 4.6, page 13] for the proof. □

3.2.2 The Reverse Problem Solution: Short Position

Now we tackle the short problem and extend the work of [Leung and Li, 2015]. A large amount of the theory from Section 3.2.1 can be reshaped and reused to solve this problem. For this reason our analysis is kept brief and we highlight the similarities and differences in the build-up to the reverse problem solution. The notation is unchanged from the long problem for consistency of reading. However, the functions will be intrinsically different because of the adjusted payoff functions h and $\hat{h}$. Note that if the price process was a symmetric diffusion process, then this subsection would be redundant by Remark 3.1.1.

We begin with the short exit trading problem as the entry problem is defined in terms of the exit problem:

Assume that two linearly independent solutions u and v exist to the Sturm-Liouville problem (3.2.1) with Dirichlet boundary conditions. These solutions should share all the properties of Assumption 3.2.1. To reflect the new problem and payoff function in the short exit value function (3.1.4), we redefine

$$h(x) := -x - c.$$

From here on, the long problem framework can be copied with ψ, H_S and W_S being defined as in equations (3.2.4), (3.2.5) and (3.2.6). ψ is positive, continuous and strictly increasing and W_S is the smallest concave majorant of H_S (see Appendix Lemma B.4.4 and remove the hats).

Moving onto the short problem, the alternate formulation is described:

Assume that two analogous solutions $\hat{u}$ and $\hat{v}$ exist. The payoff function is altered to mimic the short entry value function (3.1.5) and is redefined in terms of the exit value function as

$$\hat{h} := V_S(x) + x - \hat{c}.$$

Apart from this, the rest of the framework is unchanged. $\hat{\psi}, \hat{H}_S$ and $\hat{W}_S$ are defined equivalently as in (3.2.9), (3.2.10) and (3.2.11). Likewise, $\hat{\psi}$ is positive, continuous and strictly increasing and $\hat{W}_S$ is the smallest concave majorant for $\hat{H}_S$ (see Appendix Lemma B.4.3 and hat the quantities).

Theorem 3.2.9 (Optimality of value functions)**.** *The optimal time to exit the short problem* (3.1.4) *is characterised by finding an optimal interval* (a^*, b^*) *such that when the price process* X *leaves this interval, the investor should buy. Consequently, we claim that*

$$V_S(x) = v(x) W_S\big(\psi(x)\big), \qquad x \in \mathbb{R}.$$

The optimal time to enter the short problem (3.1.5) *is characterised by finding an optimal interval* $(\hat{a}^*, \hat{b}^*)$ *such that when the price process* X *leaves this interval, the investor should sell. Consequently, we claim that*

$$J_S(x) = \hat{v}(x) \hat{W}_S\big(\hat{\psi}(x)\big), \qquad x \in \mathbb{R}.$$

Proof. See the proof of Theorem 3.2.2, which treats h and $\hat{h}$ as unspecified payoff functions. □

This theorem highlights that the short value functions can also be characterised in terms of optimal intervals. Our end goal is to solve for the value functions and Theorem 3.2.9 provides an easier route to find V_S and J_S. We begin with the short exit value function, assuming properties of H_S to help describe the shape of W_S.

Assumption 3.2.10 (Properties of H_S)**.**

(i) H_S is continuous on $[0, \infty)$ and twice differentiable on $(0, \infty)$.

(ii) $H_S(0) = 0$ and for exit transactional cost c,

$$H_S(y) = \begin{cases} > 0 & \text{if } y \in (0, \psi(-c)), \\ < 0 & \text{if } y \in (\psi(-c), \infty), \end{cases}$$

with $H_S\big(\psi(-c)\big) = 0$.

(iii) Given b^* as the solution to $v(b) - (b+c)v'(b) = 0$, then

$$H_S'(y) = \begin{cases} > 0 & \text{if } y \in (0, \psi(b^*)), \\ < 0 & \text{if } y \in (\psi(b^*), \infty), \end{cases}$$

where $b^* < -c$. Also, $H_S'\big(\psi(b^*)\big) = 0$.

(iv) There exists a constant L^* such that

$$H_S''(y) = \begin{cases} \leq 0 & \text{if } y \in (0, \psi(L^*)], \\ \geq 0 & \text{if } y \in [\psi(L^*), \infty). \end{cases}$$

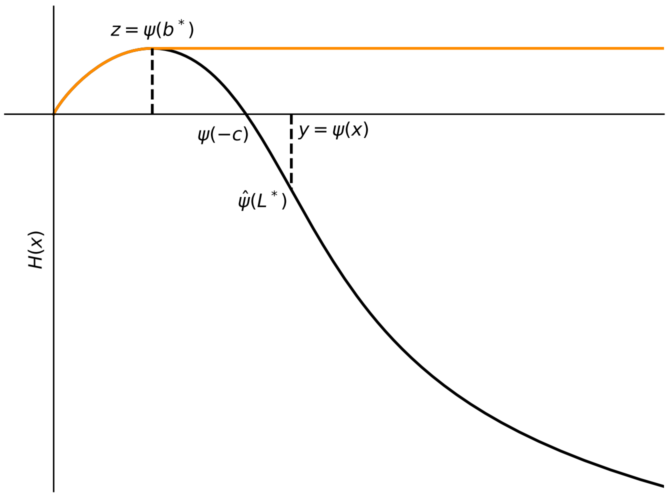

Figure 3.3: Properties of H_S

Like for the long problem, these properties are assumed as not all of them can be substantiated for a general diffusion process. We show as many characteristics as is possible in Appendix Section B.6.1. A specific verification on the properties of H_S for the OU process can be found in Appendix Section B.6.3. The ABM and GBM cases will be very similar.

Theorem 3.2.11. *The optimal exit problem* (3.1.4) *for a short position is given by*

$$V_S(x) = \begin{cases} -x - c, & x \in (-\infty, b^*], \\ (-b^* - c)\dfrac{v(x)}{v(b^*)}, & x \in (b^*, \infty), \end{cases} \tag{3.2.22}$$

where the constant $b^* < (-c) \wedge L^*$, which is the optimal exit level, is the solution to the equation

$$v(b) = (b+c)v'(b) \tag{3.2.23}$$

and the optimal liquidation time is

$$\tau^* = \inf\{t \geq 0 : X_t \leq b^*\}.$$

Proof. See Figure 3.3 for the supplementary plot to support the method of proof. As the properties of H_S for the short problem are close to those of $\hat{H}_L$ from the long problem, the proof of this theorem follows the lines of Theorem 3.2.7.

By the properties of H_S and the sketch, it should be apparent that the smallest concave majorant must be a two part function. The two sections will meet at some value $y = z$, a global maximum of

H_S. The first part will trace the concave part of H_S up till the maximum value z and the second part is a horizontal line segment, ie.

$$W_S(y) = \begin{cases} H_S(y), & y \leq z, \\ H_S(z), & y \geq z. \end{cases} \tag{3.2.24}$$

The connecting point z is a maximum with $H_S'(z) = 0$. By Assumption 3.2.10, the value $z = \psi(b^*)$, where b^* is the solution to equation (3.2.23). Recall that H_S is defined as

$$H_S(y) := \begin{cases} \frac{h}{v} \circ \psi^{-1}(y), & y > 0, \\ \lim_{x \to -\infty} \frac{\big(h(x)\big)^+}{v(x)}, & y = 0, \end{cases}$$

with $h = -x - c$. Taking $y = \psi(x)$, the smallest concave majorant W_S can be formulated in terms of x as

$$W_S(\psi(x)) = \begin{cases} \frac{-x-c}{v(x)}, & x \leq b^*, \\ \frac{-b^*-c}{v(b^*)}, & x \geq b^*. \end{cases}$$

By applying the identity in Theorem 3.2.9, $V_S(x) = v(x)W_S(\psi(x))$ and the result follows. The condition $b^* < (-c) \wedge L^*$ can be verified by producing a contradiction if not true. □

Notice the similarity to the solution for the long exit value function (3.2.15), supporting the comments in Remark 3.1.1. Next, the entry problem will be solved and a combined analysis of the solutions will be given afterwards.

Assumption 3.2.12 (Properties of $\hat{H}_S$)**.** Let V_S, b^* and L^* be as in Theorem 3.2.11. Then the following properties of $\hat{H}_S$ are assumed:

(i) $\hat{H}_S$ is continuous on $[0, \infty)$, differentiable on $(0, \infty)$ and twice differentiable on $(0, \hat{\psi}(b^*)) \cup (\hat{\psi}(b^*), \infty)$.

(ii) $\hat{H}_S(0) = 0$ and for constant $k_* > b^*$, where k_* is the unique solution to $\hat{h}(x) = 0$,

$$\hat{H}_S(y) = \begin{cases} < 0 & \text{if } y \in (0, \hat{\psi}(k_*)), \\ > 0 & \text{if } y \in (\hat{\psi}(k_*), \infty), \end{cases}$$

with $\hat{H}_S\big(\hat{\psi}(k_*)\big) = 0$.

(iii) $\hat{H}_S$ is strictly decreasing for $y \in (0, \hat{\psi}(b^*))$ and $\hat{H}_S'(y) \to 0$ as $y \to \infty$.

(iv) There exists a constant $L_* > b^*$ such that

$$\hat{H}_S''(y) = \begin{cases} \geq 0 & \text{if } y \in (0, \hat{\psi}(L_*)], \\ \leq 0 & \text{if } y \in [\hat{\psi}(L_*), \infty). \end{cases}$$

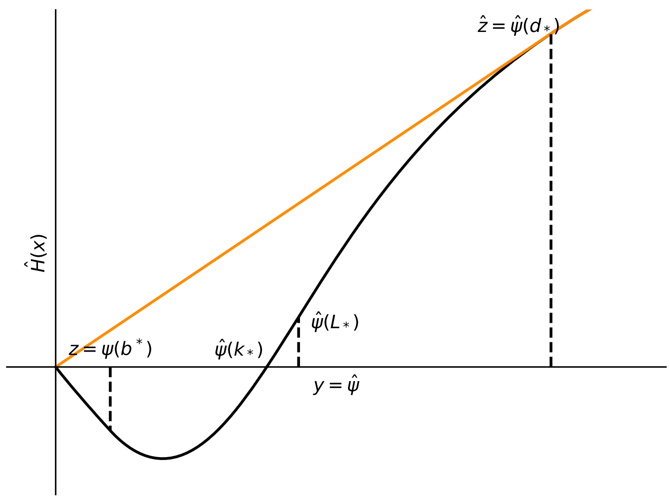

Figure 3.4: Properties of $\hat{H}_S$

Again, these properties are treated as an assumption for now. Some characteristics of $\hat{H}_S$ are established and can be found in Appendix Section B.6.2. The case-specific verification for the OU process can be found in Appendix Section B.6.3.

Theorem 3.2.13. *The optimal entry problem* (3.1.5) *for a short position is given by*

$$J_S(x) = \begin{cases} \dfrac{V_S(d_*) + d_* - \hat{c}}{\hat{u}(d_*)}\,\hat{u}(x), & x \in (-\infty, d_*), \\ V_S(x) + x - \hat{c}, & x \in [d_*, \infty), \end{cases} \tag{3.2.25}$$

where the constant $d_* > k_* \wedge L_*$, which is the optimal entry level, is the solution to the equation

$$\hat{u}(d)\big(V_S'(d) + 1\big) = \hat{u}'(d)\big(V_S(d) + d - \hat{c}\big) \tag{3.2.26}$$

and the optimal liquidation time is

$$\nu_* = \inf\{t \geq 0 : X_t \geq d_*\}.$$

Proof. See Figure 3.4 for the supplementary plot to support the method of proof. As the properties of $\hat{H}_S$ for the short problem are close to those of H_L from the long problem, the proof of this theorem follows the lines of Theorem 3.2.4.

By the properties of $\hat{H}_S$ and the sketch, it is evident that the smallest concave majorant must be a two part function. The sections will meet at some value $y = \hat{z}$, where $\hat{z} > \hat{\psi}(k_*) \wedge \hat{\psi}(L_*)$. The first part will be a line segment joining the coordinates $(0, \hat{H}_S(0))$ and $(\hat{z}, \hat{H}_S(\hat{z}))$ and the second part must trace $\hat{H}_S$ exactly as $\hat{H}_S$ is already concave in this region, ie.

$$W_S(y) = \begin{cases} y\dfrac{\hat{H}_S(\hat{z})}{\hat{z}}, & y \leq \hat{z}, \\ \hat{H}_S(y), & y \geq \hat{z}. \end{cases}$$

By Assumption 3.2.12, we have that $\lim_{y\to\infty} \hat{H}_S'(y) = 0$ and so the line segment $y\frac{\hat{H}_S(\hat{z})}{\hat{z}}$ must be tangential to $\hat{H}_S$ at $y = \hat{z}$. Therefore, the following condition holds:

$$\frac{\hat{H}_S(\hat{z}) - \hat{H}_S(0)}{\hat{z} - 0} = \frac{\hat{H}_S(\hat{z})}{\hat{z}} = \hat{H}_S'(\hat{z}). \tag{3.2.27}$$

Letting $\hat{z} = \hat{\psi}(d_*)$, the left-hand side of equation (3.2.27) becomes

$$\frac{\hat{H}_S(\hat{z})}{\hat{z}} = \frac{\hat{H}_S(\hat{\psi}(d_*))}{\hat{\psi}(d_*)} = \frac{\hat{h}(d_*)}{\hat{\psi}(d_*)\hat{v}(d_*)} = \frac{V_S(d_*) + d_* - \hat{c}}{\hat{u}(d_*)}$$

and the right-hand side is

$$\frac{(V_S'(d_*) + 1)\hat{v}(d_*) - (V_S(d_*) + d_* - \hat{c})\hat{v}'(d_*)}{\hat{u}'(d_*)\hat{v}(d_*) - \hat{u}(d_*)\hat{v}'(d_*)},$$

where expression (B.5.1) is used to calculate the derivative of $\hat{H}_S$. By equating the two sides of equation (3.2.27), one can realise that d_* solves

$$\hat{u}(d_*)(V_S'(d_*) + 1) = \hat{u}'(d_*)(V_S(d_*) + d_* - \hat{c})$$

after some cancellations. Moreover, taking $y = \hat{\psi}(x)$, the smallest concave majorant $\hat{W}$ in terms of x is

$$\hat{W}_S(\hat{\psi}(x)) = \begin{cases} \dfrac{\hat{u}(x)}{\hat{v}(x)} \dfrac{V_S(d_*) + d_* - \hat{c}}{\hat{u}(d_*)}, & x \leq d_*, \\ \dfrac{V_S(x) + x - \hat{c}}{\hat{v}(x)}, & x \geq d_*. \end{cases}$$

Taking the identity from Theorem 3.2.9, the value function is given by $J_S(x) = \hat{v}(x)\hat{W}_S(\hat{\psi}(x))$ and the result follows. Lastly, the connecting point $\hat{z}$ must be larger than both $y = \hat{\psi}(k_*)$ and $y = \hat{\psi}(L_*)$, otherwise a contradiction to the concavity of $\hat{W}_S$ would arise. □

The arduous task of finding the short entry and exit value problems has been completed. Observe that the methods used to find the value functions in the long and short problem were complementary and that no new techniques were needed. This time, the optimal interval for the short exit problem is $\mathcal{I}_V := (b^*, \infty)$, where b^* solves equation (3.2.23). With $a = b^*$ and $b = \infty$ in equation (3.2.3), the expectation calculation for $x \in \mathcal{I}_V$ becomes

$$\mathbb{E}^x\left[\mathrm{e}^{-r(T_{b^*} \wedge T_\infty)} h\big(X_{T_{b^*} \wedge T_\infty}\big)\right] = \mathbb{E}^x\left[\mathrm{e}^{-r\,T_{b^*}} h\big(X_{T_{b^*}}\big)\right] = h(b^*)\mathbb{E}^x\left[\mathrm{e}^{-r\,T_{b^*}}\right] = (-b^* - c)\frac{v(x)}{v(b^*)}$$

by the formula in identity (3.2.2). Conversely, for $x \notin \mathcal{I}_V$,

$$\mathbb{E}^x\left[\mathrm{e}^{-r(T_{b^*} \wedge T_\infty)} h\big(X_{T_{b^*} \wedge T_\infty}\big)\right] = -x - c$$

as $T_{b^*} \wedge T_\infty = 0$. Similarly, the short entry problem optimal interval is $\mathcal{I}_J := (-\infty, d_*)$, where d_* is the solution to equation (3.2.26). Taking $\hat{a} = -\infty$ and $\hat{b} = d_*$, for $x < d_*$ the expectation is

$$\mathbb{E}^x\left[\mathrm{e}^{-\hat{r}(T_{-\infty} \wedge T_{d_*})} \hat{h}\big(X_{T_{-\infty} \wedge T_{d_*}}\big)\right] = \mathbb{E}^x\left[\mathrm{e}^{-\hat{r}\,T_{d_*}} \hat{h}\big(X_{T_{d_*}}\big)\right] = \hat{h}(d_*)\mathbb{E}^x\left[\mathrm{e}^{-\hat{r}\,T_{d_*}}\right] = \frac{V_S(d_*) + d_* - \hat{c}}{\hat{u}(d_*)}\,\hat{u}(x)$$

and for $x \leq d_*$,

$$\mathbb{E}^x\left[\mathrm{e}^{-\hat{r}(T_{-\infty} \wedge T_{d_*})} \hat{h}\big(X_{T_{-\infty} \wedge T_{d_*}}\big)\right] = V_S(x) + x - \hat{c}.$$

Taking into account the long problem, the optimal trading strategy for the short problem should not be too surprising:

In this instance, $b^* < d_*$ and so the entry and exit problems do not conflict each other. The trader should make a market order for a short position if their price process exceeds value d_*, ie. $x \notin \mathcal{I}_J$, as the portfolio would be over-valued. Otherwise, they should do nothing and wait until the portfolio value rises. Once a position has been entered, the trader must buy back their short position if and when the portfolio value falls to b^*.

To conclude, a quick result on the impact of the transaction costs is presented.

Proposition 3.2.14. *The value function V_S is decreasing in the exit transaction cost c. The optimal barrier level b^* is decreasing in c, whilst d_* is increasing in $\hat{c}$.*

Proof. Refer to Proposition 3.2.5 and Proposition 3.2.8 for ideas on the proof. □

Remark 3.2.15. After repeating the arguments to calculate conditional expectations of the form

$$\mathbb{E}^x\Big[\mathrm{e}^{-r(T_a\wedge T_b)}h\big(X_{T_a\wedge T_b}\big)\Big]=h(a)\mathbb{E}^x\Big[\mathrm{e}^{-r\,T_a}\,\mathbb{1}_{\{T_a<T_b\}}\Big]+h(b)\mathbb{E}^x\Big[\mathrm{e}^{-r\,T_b}\,\mathbb{1}_{\{T_a>T_b\}}\Big]$$

multiple times, the reader should have an awareness of the ease at which this can be done. By taking an infinite end point, the expression simplifies massively and the calculation becomes routine. The difficult and lengthy part of this section was not the computation of the value functions, but rather finding the equations for the optimal levels b^* and d_*.

3.3 Two-Sided Optimal Stopping Problem

This section aims to glue together the two trading strategies into one overarching problem. We formulate two variants of the joint optimal stopping problems: the first considers a maximal profit scenario and the second considers the earliest trading signal to enter a position.

To formulate this optimal stopping problem, we consider the two exit value functions for the long and short problems

$$V_L(x)=\sup_\tau\,\mathbb{E}^x\big[\mathrm{e}^{-r\tau}(X_\tau-c)\big],$$
$$V_S(x)=\sup_\tau\,\mathbb{E}^x\big[\mathrm{e}^{-r\tau}(-X_\tau-c)\big].$$

We will also use the entry payoff functions h for both strategies, which we now denote by

$$\hat{h}_L(x)=V_L(x)-x-\hat{c}\text{ and }\hat{h}_S(x)=V_S(x)+x-\hat{c}.$$

3.3.1 The Joint Problem: Maximal Value Strategy

This strategy is written down for a purely illustrative purpose; we do not solve the problem and discuss why not. One possible way of combining the single-sided problems is to trade off the strategy that produces the largest expected profit at entry. More precisely, consider the entry performance criteria based on the maximum value of long and short strategies. Then we get

$$\begin{aligned}J^{(\nu)}(x)&=\mathbb{E}^x\Big[\mathrm{e}^{-\hat{r}\nu}\cdot\max\Big(V_L(X_\nu)-X_\nu-\hat{c},V_S(X_\nu)+X_\nu-\hat{c}\Big)\Big]\\&=\mathbb{E}^x\Big[\mathrm{e}^{-\hat{r}\nu}\cdot\max\Big(\hat{h}_L(X_\nu),\hat{h}_S(X_\nu)\Big)\Big]\end{aligned}$$

as the entry performance criteria for the joint trading problem. The joint entry value function is defined as

$$J(x)=\sup_\nu\,\mathbb{E}^x\Big[\mathrm{e}^{-\hat{r}\nu}\cdot\max\Big(\hat{h}_L(X_\nu),\hat{h}_S(X_\nu)\Big)\Big].\tag{3.3.1}$$

The best we can say about the combined value function J is that

$$\max\big[J_L(x),J_S(x)\big]\leq J(x)$$

by convexity of the maximum function. To make further progress with this two-sided stopping problem, we would need to know more about the maximal distribution. Another way to go about tackling this problem is by brute force, using Monte Carlo simulation to simulate the value functions.

3.3.2 The Joint Problem: First Execution Strategy

The strategy in this case is designed to pick the market order signal that occurs the earliest. We can formulate this problem independently of time as follows:

$$\begin{aligned}J^{(\nu)}(x)&=\mathbb{E}^x\Big[\mathrm{e}^{-\hat{r}\,\alpha}\Big\{V_L(X_\alpha)-X_\alpha-\hat{c}\Big\}\mathbb{1}_{\{\alpha\leq\upsilon\}}+\Big\{\mathrm{e}^{-\hat{r}\,\upsilon}V_S(X_\upsilon)+X_\upsilon-\hat{c}\Big\}\mathbb{1}_{\{\alpha\geq\upsilon\}}\Big]\\&=\mathbb{E}^x\big[\mathrm{e}^{-\hat{r}\,\alpha}\hat{h}_L(X_\alpha)\,\mathbb{1}_{\{\alpha\leq\upsilon\}}\big]+\mathbb{E}^x\big[\mathrm{e}^{-\hat{r}\,\upsilon}\hat{h}_S(X_\upsilon)\,\mathbb{1}_{\{\alpha\geq\upsilon\}}\big]\end{aligned}$$

is the performance criteria of the investor, where $\nu = \alpha \wedge \upsilon$ is the minimum of two random stopping times. Then we can define the joint problem entry value function as

$$J(x) = \sup_{\nu} \left\{ \mathbb{E}^x \left[\mathrm{e}^{-\hat{r}\,\alpha} \hat{h}_L(X_\alpha) \mathbb{1}_{\{\alpha \le \upsilon\}} \right] + \mathbb{E}^x \left[\mathrm{e}^{-\hat{r}\,\upsilon} \hat{h}_S(X_\upsilon) \mathbb{1}_{\{\alpha \ge \upsilon\}} \right] \right\}.$$

Note that by the set up, the maximal value trading strategy (3.3.1) will always be larger than the first execution strategy. From the single-sided problems, we know for $y = \hat{\psi}(x)$ that

$$\begin{aligned}
&\mathbb{E}^x \left[\mathrm{e}^{-\hat{r}\,T_{\hat{a}}} \hat{h}_L(X_{T_{\hat{a}}}) \mathbb{1}_{\{T_{\hat{a}} \le T_{\hat{b}}\}} \right] + \mathbb{E}^x \left[\mathrm{e}^{-\hat{r}\,T_{\hat{b}}} \hat{h}_S(X_{T_{\hat{b}}}) \mathbb{1}_{\{T_{\hat{a}} \ge T_{\hat{b}}\}} \right] \\
&= \hat{h}_L(\hat{a}) \mathbb{E}^x \left[\mathrm{e}^{-\hat{r}\,T_{\hat{a}}} \mathbb{1}_{\{T_{\hat{a}} < T_{\hat{b}}\}} \right] + \hat{h}_S(\hat{b}) \mathbb{E}^x \left[\mathrm{e}^{-\hat{r}\,T_{\hat{b}}} \mathbb{1}_{\{T_{\hat{a}} > T_{\hat{b}}\}} \right] \\
&= \hat{h}_L(\hat{a}) \frac{u(x)v(\hat{b}) - u(\hat{b})v(x)}{u(\hat{a})v(\hat{b}) - u(\hat{b})v(\hat{a})} + \hat{h}_S(\hat{b}) \frac{u(\hat{a})v(x) - u(x)v(\hat{a})}{u(\hat{a})v(\hat{b}) - u(\hat{b})v(\hat{a})} \\
&= \hat{v}\big(\hat{\psi}(y)\big) \left[\hat{H}_L(\hat{a}) \frac{y_{\hat{b}} - y}{y_{\hat{b}} - y_{\hat{a}}} + \hat{H}_S(\hat{b}) \frac{y - y_{\hat{a}}}{y_{\hat{b}} - y_{\hat{a}}} \right].
\end{aligned}$$

with $y_{\hat{a}} = \hat{\psi}(\hat{a})$ and $y_{\hat{b}} = \hat{\psi}(\hat{b})$. In the spirit of the single-sided problems, we define

$$\hat{W}(y) := \sup_{\{(y_{\hat{a}}, y_{\hat{b}}) : y_{\hat{a}} < y < y_{\hat{b}}\}} \left[\hat{H}_L(y_{\hat{a}}) \frac{y_{\hat{b}} - y}{y_{\hat{b}} - y_{\hat{a}}} + \hat{H}_S(y_{\hat{b}}) \frac{y - y_{\hat{a}}}{y_{\hat{b}} - y_{\hat{a}}} \right], \qquad y > 0,$$

and make the claim that

$$J(x) = \hat{v}(x) \hat{W}\big(\hat{\psi}(x)\big), \tag{3.3.2}$$

where $\hat{W}$ is now the smallest concave majorant of $\hat{H}_L$ and $\hat{H}_S$ together. Clearly $\hat{W}$ dominates both $\hat{H}_L$ and $\hat{H}_S$ by definition so $\hat{W}$ should be at least as large as $\max(\hat{H}_L, \hat{H}_S)$. We plot the maximum here, to get an idea for the form of $\hat{W}$.

The parameters used for the OU are: $\theta = 0.8, \mu = 0.1, \sigma = 1.25$. The parameters for the trading framework are: $r = 0.2, \hat{r} = 0.2, , c = 1, \hat{c} = 1$.

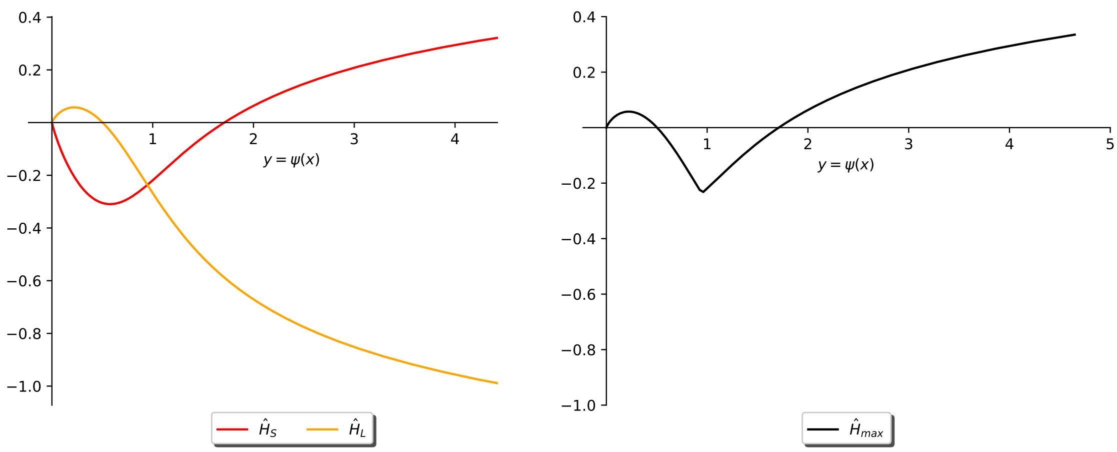

Figure 3.5: The Joint Problem: $\hat{H}_L$,$\hat{H}_S$ and $\hat{H}_{\max}$

Based on this plot, we propose that $\hat{W}$ is a three part function that traces $\hat{H}_L$ initially and traces $\hat{H}_S$ for larger y values. In between, $\hat{W}$ will be a line segment that is tangential to both $\hat{H}_L$ and $\hat{H}_S$ at their respective points. The next figure uses the same parameters as the previous figure and demonstrates the form of the smallest concave majorant $\hat{W}$.

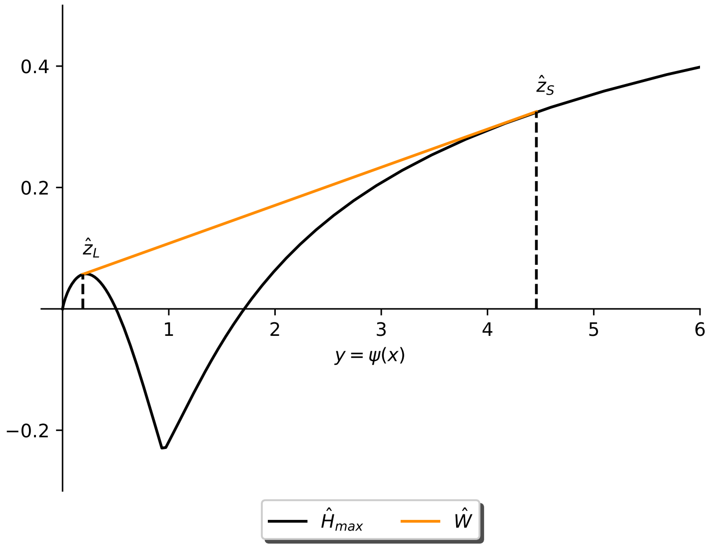

Figure 3.6: The Joint Problem: $\hat{W}$

So the entry value function will also be a three part function by identity (3.3.2). We now give an interpretation of the trading strategy:

Suppose that

$$\hat{z}_L = \hat{\psi}(d_L) \text{ and } \hat{z}_S = \hat{\psi}(d_S),$$

where d_L and d_S are the optimal barriers that can be found by solving for the tangential points to the respective $\hat{H}$-functions. There are three regions where the initial price of the portfolio x can be. By the single-sided problems, the optimal entry intervals for both problems will now be $\mathcal{I}_S := (-\infty, d_S)$ and $\mathcal{I}_L := (d_L, \infty)$ for the short and long problems respectively. Note that $d_L < d_S$ by the set up of the problems (refer to the single-sided problems to see this).

If $x < d_L$, then $x \notin \mathcal{I}_L$ so the optimal strategy in this case would be to enter a long position. On the other hand, if $x > d_S$, then $x \notin \mathcal{I}_S$ so the optimal strategy is to enter a short position. Finally, if $d_L < x < d_S$, then $x \in \mathcal{I}_L$ and $x \in \mathcal{I}_S$ and so the optimal strategy is to just wait until the process breaches one of the levels d_L or d_S. Summarising this,

$$\text{optimal strategy} = \begin{cases} x < d_L, & \text{enter a long position,} \\ d_L < x < d_s, & \text{wait,} \\ x > d_S, & \text{enter a short position.} \end{cases} \tag{3.3.3}$$

The exit trading strategy is unchanged from the single-sided strategies. If a short position is entered into, then the short exit level will be used for the optimal buy back level. If a long position is entered into, then the long exit level will be used to determine the optimal sell level.

Chapter 4

An Application of the Optimal Stopping Problems

This chapter is motivated by the practise of pairs trading in finance. We feel obliged to give at least one working example with the Ornstein-Uhlenbeck process given its popularity in industry. In addition, the assumptions postulated in Chapter 3 were made with the OU process in mind, so it is only fair to have an application here. We will explore the equivalence between the long and short problems and also provide a pictorial example for the long optimal stopping problem. Expressions for the transition density and other probabilistic constructs are also mentioned in this part.

4.1 The Ornstein-Uhlenbeck Process

Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space under which $(B_t)_{t\geq 0}$ is a Brownian motion. Consider the stochastic process $X = (X_t : \Omega \to \mathbb{R})_{t\geq 0}$ defined via the SDE

$$dX_t = \theta\,(\mu - X_t)\,dt + \sigma\,dB_t, \qquad X_0 \in \mathbb{R}, \tag{4.1.1}$$

where $\theta, \sigma > 0$ and $\mu \in \mathbb{R}$ are constants. A process satisfying this SDE is called an Ornstein-Uhlenbeck process. The term $\theta\,(\mu - X_t)$ is known as the drift coefficient and σ is the diffusion coefficient. Often, the OU process is defined with $\mu = 0$, in which case the long-term mean of the process is zero. Below is an illustration of the OU process for $\mu = 0$, $\theta = 0.2$ and $\sigma = 0.2$ starting at $X_0 = 0$.

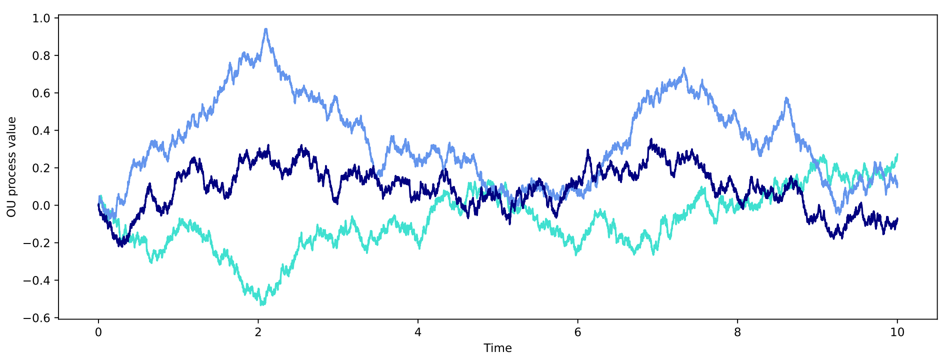

Figure 4.1: The Ornstein-Uhlenbeck Process

The process is often referred to as mean-reverting, meaning it has a tendency to fall back to its mean value. This behaviour arises due to the drift coefficient, which encourages the process to always recenter itself. We can analyse this via the SDE: if the process X grows, then $\theta\,(\mu - X_t)\,dt$

will contract and so the change in value dX_t will decrease. Conversely if the process X shrinks in value, then dX_t will grow and the process will rise. Therefore, the mean-reverting behaviour is an evident feature of this stochastic model.

The OU process is typically used to model the price spread on a pair of assets and is the most famous example of a mean-reverting model. A standard price process would be set up like this: suppose $Y^{(1)}$ and $Y^{(2)}$ are processes tracking the price of two highly correlated assets. The spread on the assets is measured by $X := Y^{(1))} - \alpha Y^{(2))}$, where $\alpha \in \mathbb{R}_{\geq 0}$ rebalances the asset weights appropriately. Assuming it is possible to statistically validate that the observed market data and model are homologous, one can begin to fit the OU model parameters to the observed data set. A standard method to do this is maximum likelihood estimation, which is not too difficult for the OU process because of its Gaussian transition densities. Given that all of this can be done, the optimal stopping problems can then be used to find the optimal barrier levels b^* and d_* to give a StatArb trading strategy.

4.1.1 Fundamentals

In this part we scrutinise the OU process and confirm the results that have been developed in the previous chapters. We also analyse the process and talk about its evolution and features.

Theorem 4.1.1. *The Ornstein-Uhlenbeck process has a unique strong solution with continuous paths almost surely. In addition, it falls into the class of diffusion processes.*

Proof. For the OU process, we can pick the constant $D = \theta$ such that the Lipschitz condition in inequality (2.1.4) is met:

$$|\theta\,(\mu - x) - \theta\,(\mu - y)| + |\sigma - \sigma| = \theta\,|x - y| \leq D\,|x - y|.$$

Independence of the initial condition is guaranteed by choosing the initial value to be a constant or a normal random variable independent of the driving BM. By Theorem 2.1.1, as the OU process is an Itô diffusion, it follows that a unique strong solution exists with continuous paths almost surely. □

By a simple application of Itô's lemma to $f(X_t, t) = X_t\, e^{\theta t}$, it is not too hard to show that the unique, strong solution to (4.1.1) in terms of the initial value X_0 is

$$X_t = X_0\, e^{-\theta t} + \mu\,(1 - e^{-\theta t}) + \sigma \int_0^t e^{-\theta(t-u)}\, dB_u, \qquad t \geq 0. \tag{4.1.2}$$

The solution can also be formulated in terms of an initial starting point X_s as

$$X_t = X_s\, e^{-\theta(t-s)} + \mu\,(1 - e^{-\theta(t-s)}) + \sigma \int_s^t e^{-\theta(t-u)}\, dB_u, \qquad t \geq s \geq 0. \tag{4.1.3}$$

The only source of randomness arises through the last term in both identities. It is widely known that the stochastic integral of a deterministic function of time with respect to Brownian motion is a centralised Gaussian process; see eg. [Shreve, 2004, Theorem 4.4.9, page 149]. Under the assumption that the initial condition $X_s \sim \mathcal{N}(\mu, \frac{\sigma^2}{2\theta})$ independently of $(B_u)_{s \leq u \leq t}$, the state of the OU process X_t itself is normally distributed. Here, $\mathcal{N}$ signifies the Gaussian distribution.

By the results in Chapter 2, one can conclude that the OU process is strongly stationary. An alternative proof of this is given to explore the mean and covariance of the process.

Theorem 4.1.2. *The Ornstein-Uhlenbeck process is a strongly stationary process.*

Proof. Strong stationarity holds for a Gaussian process if and only if the first moment is constant and the joint second moment only depends on the difference between the time indices [Percival and Walden, 1993, page 38]. In other words, for process X

$$\begin{aligned} \mu &:= \mathbb{E}[X_t] = \mathbb{E}[X_0], \\ \gamma(\tau) &:= \mathrm{Cov}[X_t, X_{t+\tau}] = \mathrm{Cov}[X_0, X_\tau], \qquad t, \tau \in \mathbb{R}_{\geq 0}, \end{aligned} \tag{4.1.4}$$

where μ and $\gamma(\tau)$ are both finite. By symmetry of covariance, it follows that $\gamma(\tau) \equiv \gamma(-\tau)$ and so we only need to consider γ as a function of $\tau \geq 0$. Any process satisfying condition (4.1.4) is usually called a weakly stationary process.

By identity (4.1.2), using that the stochastic integral is a zero-mean martingale,

$$\mathbb{E}[X_t] = \mathbb{E}[X_0]\,\mathrm{e}^{-\theta t} + \mu\,(1 - \mathrm{e}^{-\theta t}) = \mu, \qquad t \in \mathbb{R}_{\geq 0},$$

and by (4.1.3), using the independent increments property of a Brownian motion and that X_t depends on BM increments up to time t, it follows that

$$\begin{aligned}\mathbb{E}\big[X_t X_{t+\tau}\big] &= \mathbb{E}\left[X_t\left(X_t\,\mathrm{e}^{-\theta\tau} + \mu\,(1-\mathrm{e}^{-\theta\tau}) + \sigma\int_t^{t+\tau}\mathrm{e}^{-\theta(t+\tau-u)}\,dB_u\right)\right]\\ &= \mathbb{E}\big[X_t^2\big]\,\mathrm{e}^{-\theta\tau} + \mathbb{E}[X_t]\,\mu\,(1-\mathrm{e}^{-\theta\tau}) + \mathbb{E}[X_t]\,\mathbb{E}\left[\sigma\int_t^{t+\tau}\mathrm{e}^{-\theta(t+\tau-u)}\,dB_u\right]\\ &= \left(\frac{\sigma^2}{2\theta} + \mu^2\right)\mathrm{e}^{-\theta\tau} + \mu^2\,(1-\mathrm{e}^{-\theta\tau}).\end{aligned}$$

$$\begin{aligned}\mathrm{Cov}[X_t, X_{t+\tau}] &= \mathbb{E}\big[X_t X_{t+\tau}\big] - \mu^2\\ &= \frac{\sigma^2}{2\theta}\,\mathrm{e}^{-\theta\tau}, \qquad\qquad t,\tau \in \mathbb{R}_{\geq 0}.\end{aligned}$$

So the OU process has a constant and finite expectation for all times. Similarly, the covariance is finite and is only a function of the time difference τ; the other parameters are constants. Hence, the process is Gaussian and weakly stationary which implies it is also strongly stationary. □

The impact of each parameter can now be deduced: the long-term mean of the process is controlled by μ and this is confirmed via identity (4.1.2) by taking $t \to \infty$, the variance is given by $\mathrm{Var}[X_t] = \frac{\sigma^2}{2\theta}$ so σ influences the volatility and the decay rate of the covariance is commanded by θ.

Remark 4.1.3. By Section 2.1, we also know the OU process is Markovian. It is rather remarkable to hear the OU process is the only Gaussian, Markov and stationary process that is continuous in probability, up to a translation of a constant, as mentioned in [Breiman, 1968, page 350].

4.1.2 Further Properties

We now look at some properties of the OU process from Sections 2.2 and 2.3. First, we present a result of great significance due to its application in some PDE proofs and optimal stopping problems. The relevance of this result is underplayed in this thesis, but simply put, the optimal stopping problems would not make much sense otherwise.

Proposition 4.1.4. *The hitting time for an OU process is almost surely finite. That is, for hitting time*

$$\tau_K = \inf\{t \geq 0 \,|\, X_t = K\},$$

$\tau_K < \infty$ *a.s..*

Proof. By Levy's characterisation of Brownian motion, the OU process in (4.1.2) can be shown to be equal in distribution to the process $Y = (Y_t)_{t\geq 0}$, $Y_0 = X_0$ defined as

$$Y_t = \mathrm{e}^{-\theta t}\Big[X_0 + \mu\,(\mathrm{e}^{\theta t} - 1) + \sigma W_{u(t)}\Big], \tag{4.1.5}$$

with $W = (W_t)_{t\geq 0}$ a standard BM independent of everything and $u(t) = \frac{1}{2\theta}(\mathrm{e}^{2\theta t} - 1)$. This result is also known as Doob's transformation of the OU process. [Herrmann and Tanré, 2016, page 19] then shows the hitting time for an OU process is equal in distribution to $u^{-1}(\tau)$, where τ is a hitting time for W. The statement then follows as the FPT for a BM is finite almost surely. □

Next we detail some of the quantities introduced in the earlier chapters for reference and completeness. The transition density function is

$$p(x\,|\,y,t) = \sqrt{\frac{\theta}{\pi\sigma^2(1-\mathrm{e}^{-2\theta t})}}\exp\Bigg(-\frac{\theta}{\sigma^2}\frac{\Big((y-\mu)-(x-\mu)\mathrm{e}^{-\theta t}\Big)^2}{1-\mathrm{e}^{-2\theta t}}\Bigg),$$

which can be found by solving the KBE in system (2.4.5). One can also check that p satisfies the PDE and initial conditions that we proposed. A contrasting but simpler method to find p is to notice that the OU process is normally distributed, so the result can be deduced by transforming a Gaussian density function. By Proposition 2.2.3, the infinitesimal generator for the Ornstein-Uhlenbeck process is

$$\mathcal{A}f(x) = \theta\,(\mu-x)\,\frac{\partial f(x)}{\partial x} + \frac{1}{2}\,\sigma^2\,\frac{\partial^2 f(x)}{\partial x^2},$$

for suitable f. We will now use this to solve for the functions u and v in Chapter 3, motivated by [Capocelli and Ricciardi, 1971, Section 6, page 7]. Consider the highly non-trivial Sturm-Liouville PDE problem

$$\theta(\mu-x)\frac{\partial w(x)}{\partial x} + \frac{1}{2}\sigma^2\frac{\partial^2 w(x)}{\partial x^2} = rw(x).$$

where w is the eigenfunction of the equation with eigenvalue r. Note that $w(x) := w(x,r)$, but we suppress the dependence here for ease of notation. Rewriting the PDE, we have

$$\frac{\partial^2 w(x)}{\partial x^2} + \frac{2\theta\,(\mu-x)}{\sigma^2}\,\frac{\partial w(x)}{\partial x} - \frac{2r}{\sigma^2}\,w(x) = 0 \tag{4.1.6}$$

Let $x-\mu = \left(\frac{1}{\theta}\sigma^2 z\right)^{\frac{1}{2}}$ so that $(x-\mu)^2 = \frac{1}{\theta}\sigma^2 z := kz$. Then under this substitution,

$$\frac{\partial z}{\partial x} = \frac{2(x-\mu)}{k} \text{ and } \frac{\partial^2 z}{\partial x^2} = \frac{2}{k}.$$

By the chain rule, we have

$$\frac{\partial w}{\partial x} = \frac{\partial w}{\partial z}\frac{\partial z}{\partial x} = \frac{2(x-\mu)}{k}\frac{\partial w}{\partial z},$$

$$\frac{\partial^2 w}{\partial x^2} = \frac{\partial^2 w}{\partial z^2}\left(\frac{\partial z}{\partial x}\right)^2 + \frac{\partial w}{\partial z}\frac{\partial^2 z}{\partial x^2} = \frac{4(x-\mu)^2}{k^2}\frac{\partial^2 w}{\partial z^2} + \frac{2}{k}\frac{\partial w}{\partial z}.$$

Inputting these derivatives into equation (4.1.6), the differential equation becomes

$$\frac{4(x-\mu)^2}{k^2}\frac{\partial^2 w}{\partial z^2} + \left(\frac{2}{k} - \frac{4(x-\mu)^2}{k^2}\right)\frac{\partial w}{\partial z} - \frac{2r}{\sigma^2}\,w = 0. \qquad \left(k = \frac{1}{\theta}\sigma^2\right)$$

There is currently a mismatch of the terms with the differential equation being expressed in terms of x and z but we can remedy this to get, upon cancelling of some constants,

$$z\frac{\partial^2 w}{\partial z^2} + \left(\frac{1}{2} - z\right)\frac{\partial w}{\partial z} - \frac{r}{2\theta}\,w = 0. \qquad \left(x-\mu = (kz)^{\frac{1}{2}}\right)$$

Here, the function w is now a function of z and not x. We make an abuse of notation by writing $w(z)$, which should really be written as $w\big((\mu+(kz)^{\frac{1}{2}}\big)$. One final set of transformations are required to achieve a familiar PDE. Let

$$w(z) = \mathrm{e}^{\frac{z}{2}}u(z) \text{ and } z = \frac{\xi^2}{2}.$$

Then again by the chain rule,

$$\begin{aligned}\frac{\partial w}{\partial z} &= \frac{\partial w}{\partial \xi}\frac{\partial \xi}{\partial z} = \frac{1}{\xi}\frac{\partial w}{\partial \xi},\\ \frac{\partial^2 w}{\partial z^2} &= \frac{\partial^2 w}{\partial \xi^2}\left(\frac{\partial \xi}{\partial z}\right)^2 + \frac{\partial w}{\partial \xi}\frac{\partial^2 \xi}{\partial z^2} = \frac{1}{\xi^2}\frac{\partial^2 w}{\partial \xi^2} - \frac{1}{\xi^3}\frac{\partial w}{\partial \xi},\end{aligned} \tag{4.1.7}$$

and so

$$\frac{\partial^2 w}{\partial \xi^2} - \xi \frac{\partial w}{\partial \xi} - \frac{r}{\theta}\, w = 0. \qquad \left(z = \frac{\xi^2}{2}\right)$$

Again, we make an abuse of notation by writing $u(\xi)$ in place of $u\left(\frac{\xi^2}{2}\right)$. Reverse engineering the set of equations (4.1.7) in terms of our transformations gives

$$\begin{aligned}\frac{\partial w}{\partial \xi} &= \left[\frac{\xi}{2}\, u + \frac{\partial u}{\partial \xi}\right] \mathrm{e}^{\frac{\xi^2}{4}},\\ \frac{\partial^2 w}{\partial \xi^2} &= \left[\left(\frac{u}{2} + \frac{\xi}{2}\frac{\partial u}{\partial \xi} + \frac{\partial^2 u}{\partial \xi^2}\right) + \frac{\xi}{2}\left(\frac{\xi}{2}\, u + \frac{\partial u}{\partial \xi}\right)\right] \mathrm{e}^{\frac{\xi^2}{4}},\end{aligned}$$

and thus we are left with the final partial differential equation

$$\frac{\partial^2 u}{\partial \xi^2} + \left(\frac{1}{2} - \frac{\xi^2}{4} - \frac{r}{\theta}\right) u = 0, \qquad \left(w(z) = \mathrm{e}^{\frac{z}{2}} u(z)\right).$$

This is called the Weber equation and has known solutions in terms of special parabolic cylinder or Weber-Hermite functions. More details can be found in [Bateman, 1953, Section 8.2, page 116]. Undoing all the transformations leads to the solution in terms of our initial variable x.

The solution is presented in two differing forms that are convenient. [Leung and Li, 2015, Section 3, page 6] quote the two linearly independent solutions to be

$$\begin{aligned} u(x) &= \int_0^\infty y^{\frac{a}{\theta}-1} \mathrm{e}^{\sqrt{\frac{2\theta}{\sigma^2}}(x-\mu)y - \frac{y^2}{2}}\, dy,\\ v(x) &= \int_0^\infty y^{\frac{a}{\theta}-1} \mathrm{e}^{\sqrt{\frac{2\theta}{\sigma^2}}(\mu-x)y - \frac{y^2}{2}}\, dy. \end{aligned} \qquad (4.1.8)$$

To show the second solution form, we present the FPT for the OU process, where the numerator of the hitting time formula is a solution to the Sturm-Liouville problem by equation (3.2.2). [Ricciardi and Sato, 1988, Equations 1(a) and 1(b), page 3] calculate the FPT; we extend their work to include an OU process with non-zero drift to get

$$\mathbb{E}^x\left[\mathrm{e}^{-a\tau_c}\right] = \begin{cases} \exp\left[\frac{\kappa}{2}\left\{(x-\mu)^2 - (c-\mu)^2\right\}\right] \dfrac{\mathcal{D}_{-\frac{a}{\theta}}\left(\sqrt{2\kappa}(\mu - x)\right)}{\mathcal{D}_{-\frac{a}{\theta}}\left(\sqrt{2\kappa}(\mu - c)\right)}, & x \le c,\\ \exp\left[\frac{\kappa}{2}\left\{(x-\mu)^2 - (c-\mu)^2\right\}\right] \dfrac{\mathcal{D}_{-\frac{a}{\theta}}\left(\sqrt{2\kappa}(x-\mu)\right)}{\mathcal{D}_{-\frac{a}{\theta}}\left(\sqrt{2\kappa}(c-\mu)\right)}, & x \ge c, \end{cases}$$

where $\kappa = \frac{\theta}{\sigma^2}$ and $\mathcal{D}_a$ is a parabolic cylinder function [Borodin, 2002, Appendix 2, Section 9, page 639]. The other formulation of the solution in terms of Hermite functions can be found in [Ditlevsen, 2007, page 5].

4.1.3 Analytical Results

Assumption 3.2.1 can be verified for this u and v. For example considering the integral solution u, we can show it is strictly increasing and convex:

$$\begin{aligned} u'(x) &= \sqrt{\tfrac{2\theta}{\sigma^2}} \int_0^\infty y^{\frac{a}{\theta}} \mathrm{e}^{\sqrt{\frac{2\theta}{\sigma^2}}(x-\mu)y - \frac{y^2}{2}}\, dy > 0,\\ u''(x) &= \tfrac{2\theta}{\sigma^2} \int_0^\infty y^{\frac{a}{\theta}+1} \mathrm{e}^{\sqrt{\frac{2\theta}{\sigma^2}}(x-\mu)y - \frac{y^2}{2}}\, dy > 0, \end{aligned}$$

by an application of Leibniz's rule. Furthermore, the integrands are non-negative, continuous functions and the integral preserves these properties. We let the reader check that v is strictly decreasing and convex. Therefore, the work in Chapter 3 completely holds and the framework applies to the OU process. We plot the value functions below for the long and short problems to visualise the expected payoffs and confirm the properties of $H, \hat{H}, W$ and $\hat{W}$.

The parameters used for the OU are: $\theta = 0.3, \mu = 0.8, \sigma = 2$. The parameters for the trading framework are: $r = 0.2, \hat{r} = 0.2, , c = 1, \hat{c} = 1$.

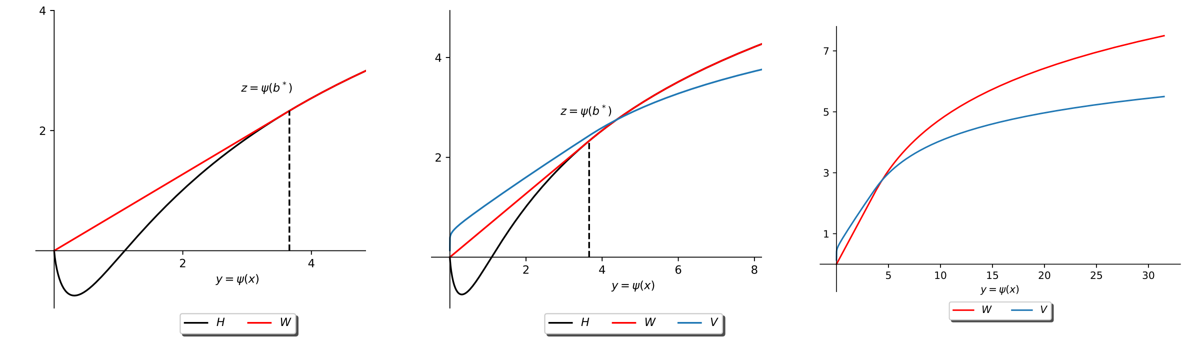

Figure 4.2: The Ornstein-Uhlenbeck Process: Long Exit Problem

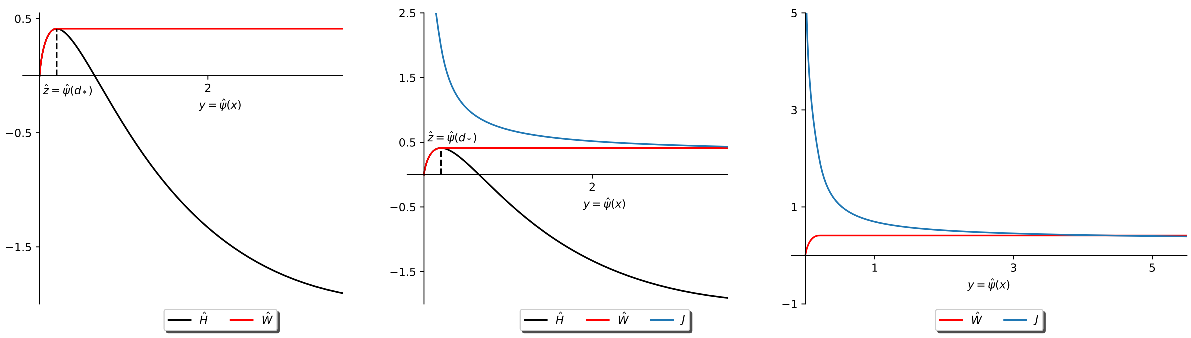

Figure 4.3: The Ornstein-Uhlenbeck Process: Long Entry Problem

Here we see that the long entry value function J_L is decreasing in $\hat{\psi}(x)$ and thus in the portfolio value x. This is in line with what we would expect: if the traders' portfolio of assets becomes over-valued, the return they can make by taking a long position diminishes.

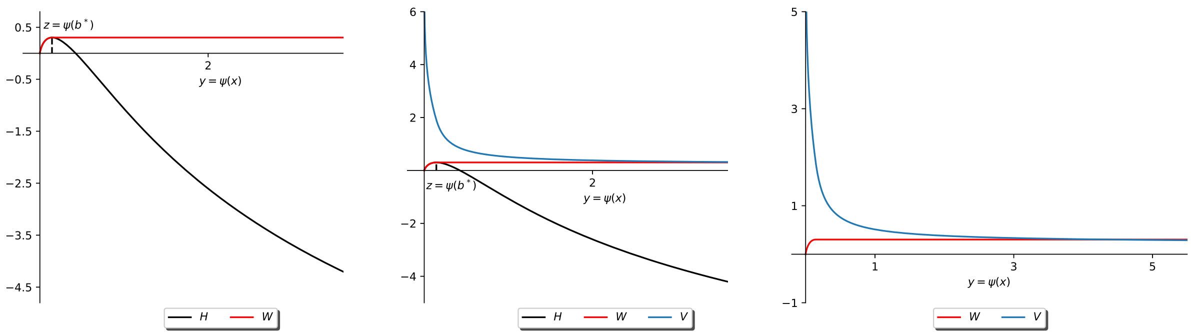

Figure 4.4: The Ornstein-Uhlenbeck Process: Short Exit Problem

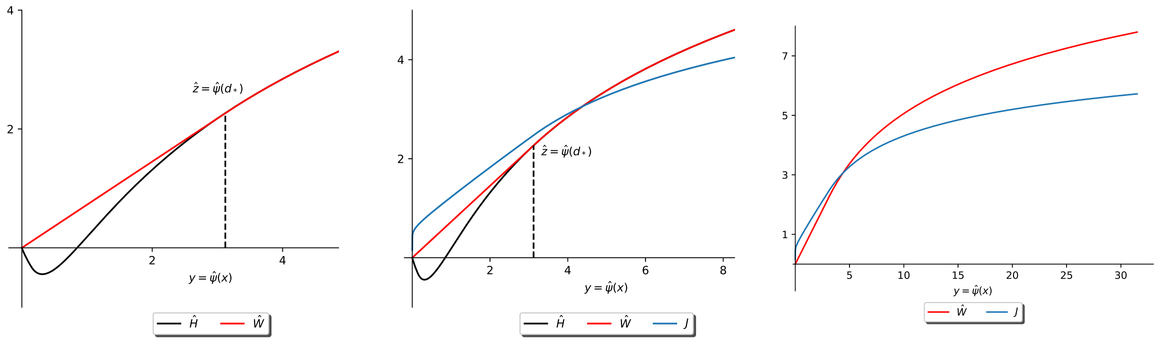

Figure 4.5: The Ornstein-Uhlenbeck Process: Short Entry Problem

Likewise, we see the opposite for the short entry problem: the value function J_S increases with the value of assets. This is because the trader will enter a short position with an overpriced portfolio.

One unique feature of the Ornstein-Uhlenbeck process is its symmetry, which does not hold for most other stochastic processes. Suppose X follows an OU process as postulated in SDE (4.1.1). Define the mirror image process $Y = \{Y_t := -X_t\}_{t\geq 0}$. Then

$$\begin{aligned} dY_t &= -dX_t \\ &= -\theta\,(\mu - X_t)\,dt - \sigma\,dB_t \\ &= -\theta\,(\mu + Y_t)\,dt + \sigma\,d\tilde{B}_t \\ &= -\theta\,(-\tilde{\mu} + Y_t)\,dt + \sigma\,d\tilde{B}_t \\ &= \theta\,(\tilde{\mu} - Y_t)\,dt + \sigma\,d\tilde{B}_t. \end{aligned}$$

Therefore, Y is an OU process with mean $\tilde{\mu} = -\mu$ being driven by a Brownian motion $\tilde{B}_t = -B_t$. Observe in Figures 4.2-4.5 the symmetry that can be seen between the long and short problems. We now describe why this harmony between both problems is to be expected. The long exit performance criteria was stated as

$$V_L^{(\tau)}(x) = \mathbb{E}^x\left[\mathrm{e}^{-r\tau}(X_\tau - c)\right],$$

whilst the short exit performance criteria was specified by

$$V_S^{(\tau)}(x) = \mathbb{E}^x\left[\mathrm{e}^{-r\tau}(-X_\tau - c)\right].$$

Given the mirror process Y which starts at $Y_0 = y = -x$, we have that

$$V_L^{(\tau)}(x,\mu) = \mathbb{E}^x\left[\mathrm{e}^{-r\tau}(X_\tau - c)\right] = \mathbb{E}^{-y}\left[\mathrm{e}^{-r\tau}(-Y_\tau - c)\right] = V_S^{(\tau)}(-y,\tilde{\mu}),$$

where the dependence on the long-term mean is highlighted. Subsequently, $V_L(x,\mu) = V_S(-y,\tilde{\mu})$, which is what Theorems 3.2.4 and 3.2.11 say. To go from the long exit problem to the short one, the pair (x,μ) is changed to $(-x,-\mu)$ and the solution u changes accordingly to v (see the definition of u and v in identity (4.1.8)). Correspondingly, the symmetry of the exit value function translates to the entry value functions and

$$\begin{aligned} J_S^{(\nu)}(x,\mu) &= \mathbb{E}^x\left[\mathrm{e}^{-\hat{r}\nu}\left(V_S(X_\nu) + X_\nu - \hat{c}\right)\right] \\ &= \mathbb{E}^{-y}\left[\mathrm{e}^{-\hat{r}\nu}\left(V_S(-Y_\nu) - Y_\nu - \hat{c}\right)\right] \\ &= \mathbb{E}^{-y}\left[\mathrm{e}^{-\hat{r}\nu}\left(V_L(Y_\nu) - Y_\nu - \hat{c}\right)\right] \\ &= J_L^{(\nu)}(-y,\tilde{\mu}). \end{aligned}$$

This yields that $J_S(x,\mu) = J_L(-y,\tilde{\mu})$. Based on this analysis, studying the long problem is identical to studying a variant of the short problem and so the need to study both problems is unnecessary.

Finally, we provide an example of the long trading strategy. We use the OU model parameters $\theta = 0.8, \mu = 0.1, \sigma = 1$ and $r = 0.2, \hat{r} = 0.2, c = 1, \hat{c} = 1$. The optimal levels are also calculated, with $d_* - 1.45$ and $b^* = 1.62$.

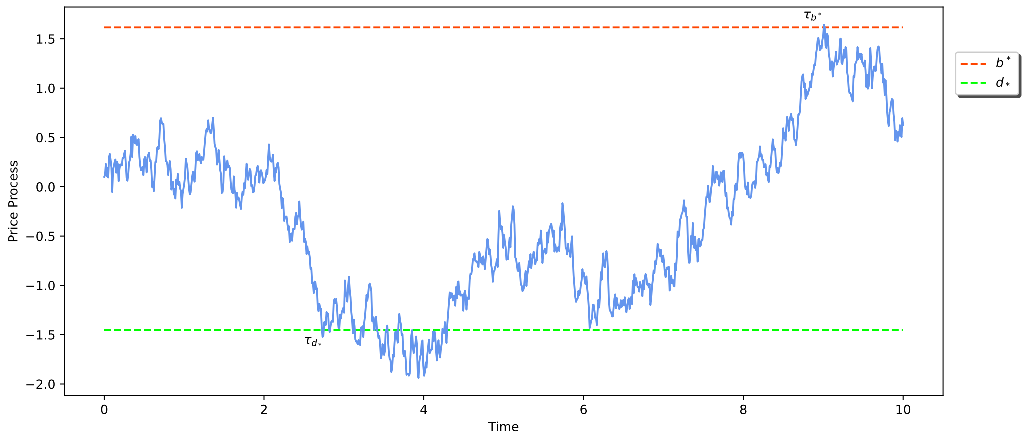

Figure 4.6: The Ornstein-Uhlenbeck Process: Long Trading Example

The picture illustrates that a trade is entered at about $t = 2.7$ and is exited at around $t = 9$. A market order for a long position is placed once the price process falls below level d_*. Then, the strategy is to hold the basket of securities until they exceed the value b^*.

Chapter 5

Conclusion

While many stochastic control problems are relatively easy to write down, finding a solution to them is a highly non-trivial exercise requiring some heavy machinery to be built up. Of course, one can work around this using computational methods, but obtaining a closed form solution is much more satisfactory. Nevertheless, we have managed to find the solution to the single-sided optimal stopping problems for both a long and short strategy. The optimal execution levels can be found computationally for the OU process to yield the best trading strategy. Although we did not analyse this, by the simplicity of the solution to the Sturm-Liouville problems, the optimal execution levels for the ABM and GBM should be able to be found explicitly.

Our efforts to generalise the framework to arbitrary diffusion processes had its advantages and disadvantages. The obvious benefit to this approach was that the optimal stopping problems could be applied to any price process meeting the assumptions made. An example where this extension could be useful is in the study of American perpetual option. The drawback of this approach was that a lot of assumptions began to creep into the formulation of the solution. There is a fine balance between having too many assumptions and not enough generality; we believe that this level of assumption was enough but not too overloading.

We also gained a better understanding of the duality that exists between a lot of the concepts described. For example, the relation between Laplace transformations and solutions to the PDE problems in Chapter 2 or the equivalence of the value functions for processes that have a mirror analogue in Chapter 4. In addition, we did all of this whilst maintaining mathematical rigour in the thesis. A common theme in papers and books describing optimal stopping problems is that a lot of the understanding in the build up to the problem is lost, mainly due to the sheer depth of the theory. We attempted to break this trend to give a more complete picture to the reader.

The progress on the joint problem is something that should be investigated as a future problem. It would have been nice to be able to have found the optimal entry equations to determine the parameters d_L and d_S. Although our analysis on the joint problem was minimal, the framework behind it has been provided so completing the study of the problem should be reasonable. Another area worth looking into is the effect of the problem and diffusion process parameters on the value function. This would really help in understanding the forms of $H, \hat{H}, W$ and $\hat{W}$. Another path of research is to investigate the properties of the optimal stopping time solution. It would be nice to have some idea of when we should expect to enter to exit the trade. However, the distributions of stopping times are notoriously complicated so this might be challenging.

Appendix A

Supplementary Examples

A.1 Examples of Diffusion Processes and Their Properties

This section of the Appendix considers two other popular stochastic processes: the arithmetic and geometric Brownian motions. We detail their defining stochastic differential equations, the solutions to their SDEs, expressions for their transition probability densities and first hitting times and further properties. Note that all of these processes have a unique strong solution and continuous paths a.s. (Theorem 2.1.1) and are Itô diffusions (Definition 2.1.2).

In Section 2.4, we claimed that two boundary conditions in equation (2.4.5) for the transition density function were reasonable assumptions. These boundary conditions are true for the trio of stochastic processes, which are in essence a monotone transform of the normal distribution that decays at $\pm\infty$. The transition density formulas can be determined by also leveraging the normal distribution relationship. An alternative method, albeit slightly more challenging, is to solve the Fokker-Planck equation to realise expressions for p.

Denote by $X = (X_t)_{t\geq 0}$ to be a stochastic process and τ_c to be the first time that X hits the level c, ie. τ_c is a stopping time defined as

$$\tau_c = \inf\{t \geq 0 \,|\, X_t = c\}.$$

Then, the formula to calculate $\mathbb{E}^x[\mathrm{e}^{-a\tau_c}]$ can be found from equation (3.2.2).

A.1.1 Arithmetic Brownian Motion

An arithmetic Brownian motion is defined as a process satisfying SDE

$$dX_t = \mu\, dt + \sigma\, dB_t,$$

where $\mu \in \mathbb{R}$ and $\sigma > 0$. The solution is

$$X_t = X_0 + \mu t + \sigma B_t$$

with initial condition $X_0 \in \mathbb{R}$. Its transition density function is

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

and the first hitting time for an ABM is

$$\mathbb{E}^x\left[\mathrm{e}^{-a\tau_c}\right] = \begin{cases} \exp\left\{\left(\sqrt{\mu^2 + 2a\sigma^2} - \mu\right)\dfrac{(x-c)}{\sigma^2}\right\}, & x \leq c, \\ \exp\left\{\left(\sqrt{\mu^2 + 2a\sigma^2} + \mu\right)\dfrac{(c-x)}{\sigma^2}\right\}, & x \geq c, \end{cases}$$

which can be verified by finding the solutions to the second-order linear Sturm-Liouville PDE

$$\mu\frac{\partial w(x)}{\partial x} + \frac{\sigma^2}{2}\frac{\partial^2 w(x)}{\partial x^2} = aw(x).$$

Alternatively, [Borodin, 2002, Part II, Section 2.2, page 295] provides this formula for a simplified ABM which can be transformed to get this result. We now check the properties of the solution to the PDE. Let us check them for

$$u(x) = \exp\left\{ \left(\sqrt{\mu^2 + 2a\sigma^2} - \mu \right) \frac{x}{\sigma^2} \right\}.$$

The function is evidently continuous and positive. Furthermore, so long as $a > 0$

$$u'(x) = \frac{1}{\sigma^2} \left(\sqrt{\mu^2 + 2a\sigma^2} - \mu \right) \exp\left\{ \left(\sqrt{\mu^2 + 2a\sigma^2} - \mu \right) \frac{x}{\sigma^2} \right\} > 0,$$

$$u''(x) = \frac{1}{\sigma^4} \left(\sqrt{\mu^2 + 2a\sigma^2} - \mu \right)^2 \exp\left\{ \left(\sqrt{\mu^2 + 2a\sigma^2} - \mu \right) \frac{x}{\sigma^2} \right\} > 0,$$

and so we have that u is strictly increasing and convex. These properties can also be checked for v assuming $a > 0$. Finally, one can observe that choosing $\mu = 0$ and $\sigma = 1$, we obtain formulae for Brownian motion. Going forward, we will only consider ABMs for $a > 0$.

A.1.2 Geometric Brownian Motion

The geometric Brownian motion is one natural extension of an ABM and is defined by the SDE

$$dX_t = \mu X_t \, dt + \sigma X_t \, dB_t$$

for $\mu \in \mathbb{R}$ and $\sigma > 0$. The solution to this SDE is obtained using Itô's lemma and is

$$X_t = X_0 \exp\left((\mu - \frac{\sigma^2}{2})t + \sigma B_t \right),$$

with initial value $X_0 > 0$. Unlike the arithmetic Brownian motion, the geometric counterpart is a strictly positive process. For this reason, the analysis we have done needs to be shifted over to the positive real line. The boundary points on this domain now reside at $x = 0$ and $x = \infty$. The transition function density for a GBM is

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

and the first hitting time is

$$\mathbb{E}^x\left[\mathrm{e}^{-a\tau_c}\right] = \begin{cases} \left(\frac{x}{c}\right)^{\sqrt{\nu^2+2a/\sigma^2}-\nu}, & x \le c, \\ \left(\frac{c}{x}\right)^{\sqrt{\nu^2+2a/\sigma^2}+\nu}, & x \ge c, \end{cases} \tag{A.1.1}$$

where $\mu = \frac{\sigma^2}{2} + \sigma^2\nu$, which can be confirmed by finding the solutions to the Sturm-Liouville/Euler-Cauchy PDE

$$\mu x \frac{\partial w(x)}{\partial x} + \frac{1}{2}\sigma^2 x^2 \frac{\partial^2 w(x)}{\partial x^2} = aw(x).$$

[Borodin, 2002, Part II, Section 9.2, page 622] quotes this formula for a GBM starting at $X_0 = 1$. Their result can be transformed to an arbitrary GBM starting at $X_0 > 0$ by using the rescaled process $Y = (Y_t := \frac{X_t}{X_0})_{t \ge 0}$ to yield equation (A.1.1). We now check the properties of the solution to the PDE. Let us take the solution

$$u(x) = x^{\sqrt{\nu^2+2a/\sigma^2}-\nu}.$$

This function is well-defined for $x > 0$ and is continuous and positive in this region. Assuming $a > \max(2\sigma^2(1 + \nu), 0)$, the derivatives are

$$u'(x) = \left(\sqrt{\nu^2 + 2a/\sigma^2} - \nu \right) x^{\sqrt{\nu^2+2a/\sigma^2}-\nu-1} > 0,$$

$$u''(x) = \left(\sqrt{\nu^2 + 2a/\sigma^2} - \nu \right) \left(\sqrt{\nu^2 + 2a/\sigma^2} - \nu - 1 \right) x^{\sqrt{\nu^2+2a/\sigma^2}-\nu-2} > 0,$$

and so u is strictly increasing and convex. This analysis can be repeated for the solution

$$v(x) = \left(\frac{1}{x}\right)^{\sqrt{\nu^2+2a/\sigma^2}+\nu},$$

where v is strictly decreasing and convex so long as $a > \max(2\sigma^2(1-\nu), 0)$. Thus, we can conclude that the properties of Assumption 3.2.1 are true given that $a > 2\sigma^2 \cdot \max(1-\nu, 1+\nu)$. Going forward, we will assume that a satisfies this inequality.

A.2 Additional Plots

Auxiliary plots are given for an arithmetic Brownian motion. The parameters used for the ABM are: $\mu = -0.2\,, \sigma = 2$ and for the trading framework: $r = 0.4\,, \hat{r} = 0.4\,, c = 0.1\,, \hat{c} = 0.1$.

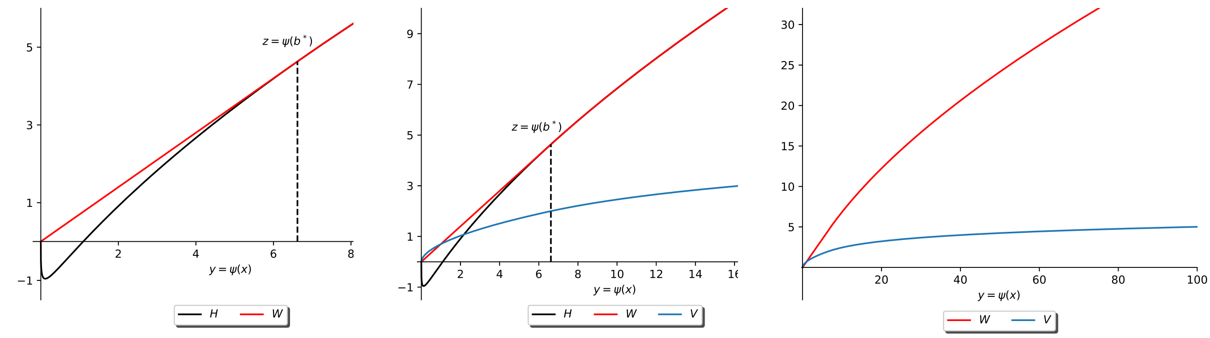

Figure A.1: Arithmetic Brownian Motion: Long Exit Problem

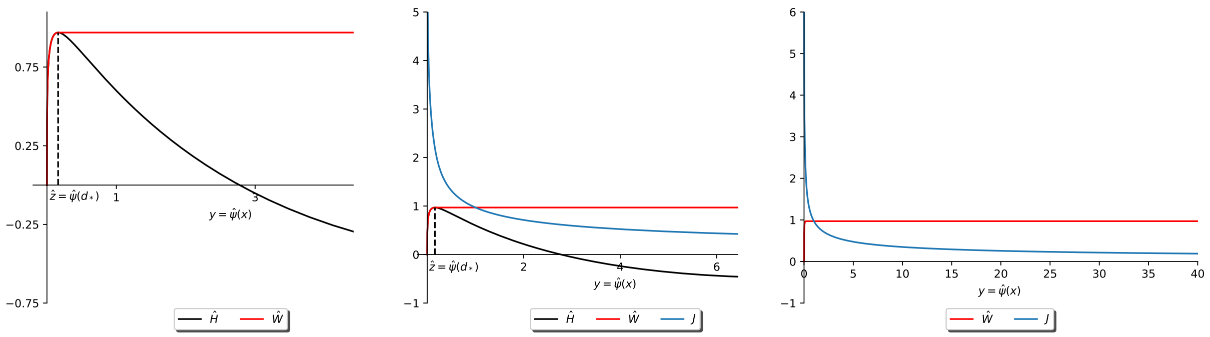

Figure A.2: Arithmetic Brownian Motion: Long Entry Problem

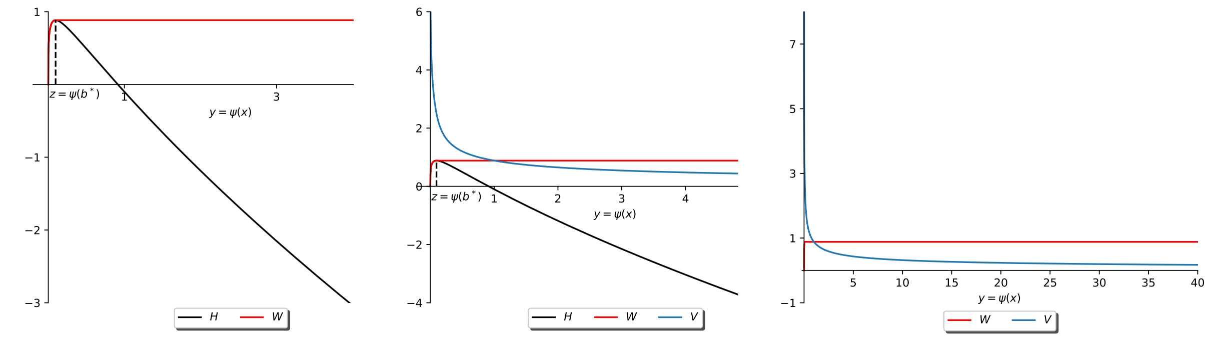

Figure A.3: Arithmetic Brownian Motion: Short Exit Problem

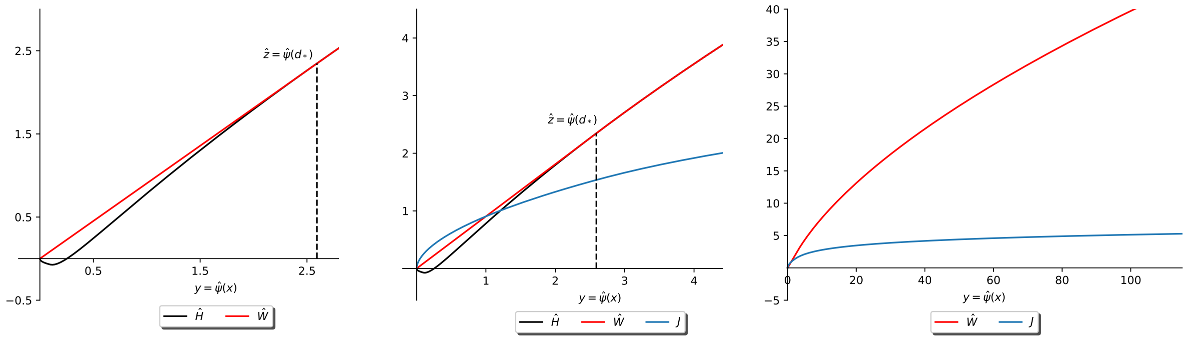

Figure A.4: Arithmetic Brownian Motion: Short Entry Problem

These plots should be seen as additional evidence to support the general framework developed in Chapter 3 regarding the solutions to the optimal stopping problems. The main difference between the OU plots in Figures 4.2-4.5 and these plots are the scale, where the ABM value functions dwarf the OU value functions. Like with the OU process, the symmetry between the long and short problems is also observable and is explained by a similar argument. Given that $Y = -X$, and that $dX_t = \mu\, dt + \sigma\, dB_t$,

$$\begin{aligned} dY_t &= -\mu\, dt - \sigma\, dB_t \\ &= \tilde{\mu}\, dt + \sigma\, d\tilde{B}_t \\ &= \tilde{\mu}\, dt + \sigma\, d\tilde{B}_t \end{aligned}$$

for $\tilde{\mu} = -\mu$ and $\tilde{B}_t = -B_t$. Hence, the same arguments used in Section 4.1.3 applies.

Appendix B

Mathematical Results

This chapter deals with the more technical details of proofs that are omitted from the main text and definitions which have been excluded or assumed to be known by the reader. The results used within the main text are detailed and extensions to these results are laid out to describe more general theorems.

B.1 Real Analysis and Probability Theory

The definitions below concern the variation of a function, with Lipschitz continuity being the most restrictive definition. Any function satisfying either of the two conditions below is automatically uniformly continuous by definition.

Definition B.1.1 (Hölder continuity)**.** A function $f : \mathbb{R} \mapsto \mathbb{R}$ is Hölder continuous or satisfies the Hölder condition if for constants $K, \alpha \in \mathbb{R}_{>0}$

$$|f(x) - f(y)| \leq K|x - y|^{\alpha}, \qquad x, y \in \mathbb{R}.$$

If we choose $\alpha = 1$, then we get Lipschitz continuity. Therefore, any Lipschitz continuous function must also be Hölder continuous.

Definition B.1.2 (Lipschitz continuity)**.** A function $f : \mathbb{R} \mapsto \mathbb{R}$ is Lipschitz continuous if for constant $K \in \mathbb{R}_{>0}$

$$|f(x) - f(y)| \leq K|x - y|, \qquad x, y \in \mathbb{R}.$$

The next two results concern functional properties used in the thesis.

Lemma B.1.3 (Inverse derivative)**.** *Let $f : \mathbb{R} \mapsto \mathbb{R}$ be a differentiable function and suppose that its inverse function f^{-1} exists and is well-defined. Then*

$$\frac{d}{dy} f^{-1}(y) = \frac{1}{f'(f^{-1}(y))}.$$

Proof. Suppose $y = f(x)$. Then $\frac{dy}{dx} = f'(x)$ and

$$\frac{d}{dy} f^{-1}(y) = \frac{d}{dx} f^{-1}(f(x)) \cdot \frac{dx}{dy} = \frac{dx}{dx} \cdot \frac{1}{\frac{dy}{dx}} = \frac{1}{f'(x)} = \frac{1}{f'(f^{-1}(y))}. \qquad \square$$

Lemma B.1.4 (Generalised L'Ĥopital's Rule)**.** *For real-valued functions f and g which are differentiable on some open interval,*

$$\liminf_{x \to a} \frac{f'(x)}{g'(x)} \leq \liminf_{x \to a} \frac{f(x)}{g(x)} \leq \limsup_{x \to a} \frac{f(x)}{g(x)} \leq \limsup_{x \to a} \frac{f'(x)}{g'(x)}$$

with $g(x) \neq 0$ close to a.

Proof. The proof of this fact comes from Cauchy's mean value theorem applied to some parts of the proof to L'Ĥopital's rule. By using this, one can show that, for example

$$\limsup_{x\to a} \frac{f(x)}{g(x)} \le \lim_{x\to a} \frac{f'(x)}{g'(x)}$$

and similarly for the infimum counterpart. The inequalities can then be achieved by using the definitions of inf and sup. □

Another theorem implemented a few times is Leibniz's rule; we recommend the reader understands how to apply this rule. Lastly, we describe the nifty independence lemma. Often, when given a conditional expectation which cannot be simplified, this lemma comes to hand to untangle calculations.

Lemma B.1.5 (Independence lemma)**.** *Given a sub-σ-algebra $\mathcal{G} \subseteq \mathcal{F}$ and two random variables, X independent of $\mathcal{G}$ and Y a $\mathcal{G}$-measurable random variable, then for arbitrary function f*

$$\mathbb{E}[f(X,Y)\,|\,\mathcal{G}] = g(Y),$$

where $g(y) := \mathbb{E}[f(X,y)]$ is an unconditional expectation.

Proof. The proof is not provided, as it is not in [Shreve, 2004, Lemma 2.3.4, page 73] either. However, the statement of the lemma in more generality can at least be found in the text. □

B.2 Stochastic Analysis and the Markov Property

This part states some fundamental results within stochastic analysis and Markov theory. Some other results not mentioned here that we use include the optional stopping theorem and the martingale property of the Itô integral for square-integrable functions.

Doob's lemma is a useful instrument that relates the σ-field inclusivity of two functions f and h to their functional relationship. The result is also known as the Doob-Dynkin lemma.

Lemma B.2.1 (Doob's lemma)**.** *Suppose f and h are measurable functions between the measurable spaces $(\Omega, \mathcal{F})$ and $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$. Then the following are equivalent:*

1. *$\sigma(f) \subseteq \sigma(h)$,*
2. *there exists a measurable function $g : (\mathbb{R}, \mathcal{B}(\mathbb{R})) \to (\mathbb{R}, \mathcal{B}(\mathbb{R}))$ such that $f = g \circ h$.*

Proof. See [Rao, 2006, Propostion 3, page 8] for a comprehensive proof. □

As seen in Chapter 2, there are many different formulations of the Markov property; having some formulations are more convenient than others. Doob's lemma will be used to derive another formulation of the Markov property.

Lemma B.2.2 (Markov property)**.** *Let $f : \mathbb{R} \to \mathbb{R}$ be bounded and measurable. A stochastic process $X = (X_t)_{t\ge 0}$ is Markov if and only if for arbitrary function g,*

$$\mathbb{E}\big[f(X_t)\,|\,\mathcal{F}_s\big] = g(X_s), \qquad s \le t.$$

Proof. We start by proving the only if direction. Suppose X is Markov and satisfies Definition 2.1.5. By properties of conditional expectation, we know that $\mathbb{E}[f(X_t)\,|\,X_s]$ is measurable with respect to $\sigma(X_s)$. So by an application of Doob's lemma, this holds if and only if there exists a measurable function g such that

$$\mathbb{E}\big[f(X_t)\,|\,X_s\big] = g(X_s).$$

Hence, we get the statement. To prove the if direction, by taking the conditional expectation with respect to X_s, we get

$$\mathbb{E}\big[f(X_t)\,|\,X_s\big] = \mathbb{E}\Big[\mathbb{E}\big[f(X_t)\,|\,\mathcal{F}_s\big]\;\Big|\;X_s\Big] = \mathbb{E}\big[g(X_s)\,|\,X_s\big] = g(X_s)$$

and by definition of $g(X_s)$ in the statement, the proof is complete. □

The next result introduced is the strong Markov property, an extension of the standard Markov property. The result is identical in all aspects apart from the time index being replaced with a stopping time. Amazingly, all Itô diffusions satisfy the strong Markov property.

Definition B.2.3 (Strong Markov property)**.** A process $X = (X_t)_{t\geq 0}$ has the strong Markov property if for any $f : \mathbb{R} \to \mathbb{R}$, bounded and measurable and stopping time $\tau < \infty$ a.s.,

$$\mathbb{E}\big[f(X_{\tau+s})\,|\,\mathcal{F}_\tau\big] = \mathbb{E}^{X_\tau}\big[f(X_{\tau+s})\big], \qquad 0 \leq s,$$

where convention dictates that as in Definition 2.1.5, we usually write that $\mathbb{E}^{X_\tau}\big[f(X_{\tau+s})\big] = \mathbb{E}\big[f(X_{\tau+s})\,|\,X_\tau\big]$. Details on the σ-algebra can be found in [Øksendal, 2013, Definition 7.2.3, page 111].

Theorem B.2.4 (Strong Markov property)**.** *Itô diffusions have the strong Markov property.*

Proof. See [Øksendal, 2013, Theorem 7.2.4, page 111] as the proof is beyond the scope of this thesis. □

Another useful formula to have for stopping times is Dynkin's formula. The result is extremely useful as it involves an integral evaluated at a stopping time.

Proposition B.2.5 (Dynkin's formula)**.** *Let X be an Itô diffusion starting at x and $f : \mathbb{R} \to \mathbb{R}$. If $f \in C_0^2(\mathbb{R})$ and $\mathbb{E}^x[\tau] < \infty$, then*

$$\mathbb{E}^x[f(X_\tau)] = f(x) + \mathbb{E}^x\left[\int_0^\tau \mathcal{A}f(X_s)\,ds\right].$$

Proof. By the proof of Proposition 2.2.3,

$$\begin{aligned} M_t^x(f) &= \mathbb{E}^x[f(X_t)] - f(X_0) - \mathbb{E}^x\left[\int_0^t \left(b(X_s)\,\frac{\partial f(X_s)}{\partial x} + \frac{1}{2}\,\sigma^2(X_s)\,\frac{\partial^2 f(X_s)}{\partial x^2}\right) ds\right] \\ &= \mathbb{E}^x\left[\int_0^t \sigma(X_s)\,\frac{\partial f(X_s)}{\partial x}\,dB_s\right] \\ &= 0 \end{aligned}$$

is a martingale for $t \in \mathbb{R}_{\geq 0}$, by square-integrability of the final integral. By the optional stopping theorem, we can also argue that the stopped integral formula

$$\int_0^\tau \sigma(X_s)\,\frac{\partial f(X_s)}{\partial x}\,dB_s = \int_0^\infty \mathbb{1}_{\{s<\tau\}}\,\sigma(X_s)\,\frac{\partial f(X_s)}{\partial x}\,dB_s, \tag{B.2.1}$$

is martingale for all stopping times τ. Hence, $M_\tau^x(f)$ still martingale under stopping time τ and

$$\mathbb{E}^x[f(X_\tau)] - f(X_0) - \mathbb{E}^x\left[\int_0^\tau \left(b(X_s)\,\frac{\partial f(X_s)}{\partial x} + \frac{1}{2}\,\sigma^2(X_s)\,\frac{\partial^2 f(X_s)}{\partial x^2}\right) ds\right] = 0.$$

Realising the expression for a generator, the proposition is proved. □

B.3 Partial Differential Equations and Their Stochastic Representation

The theorems here are mostly extensions to the PDE results from the main text. The PDEs considered here include non-homogeneous terms and also linear terms (without a derivative) in the solution variable u. The first result is an extension of Kolmogorov's equation and the second result extends Dirichlet's problem.

Theorem B.3.1 (Kolmogorov's equations)**.** *Suppose $X = (X_t)_{t\geq 0}$ is an Itô process satisfying SDE*

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

Let $P(x, t \,|\, y, u)$ be a transition function for X specified from Definition 2.2.1 as

$$P(x, t \,|\, y, u) = \mathbb{P}(X_u < y \,|\, X_t = x)$$

and assume that it is smooth enough such that its transition density function $p(x, t \,|\, y, u) = \frac{\partial}{\partial y} P(x, t \,|\, y, u)$ exists. Then the density p satisfies both the Kolmogorov forward differential equation in the forward variables (u, y)

$$\frac{\partial p}{\partial u} = -\frac{\partial}{\partial y}\big(b(u, y)p\big) + \frac{1}{2}\frac{\partial^2}{\partial y^2}\big(\sigma^2(u, y)p\big), \qquad p(t, x \,|\, t, y) = \delta_y(x),$$

and the Kolmogorov backward differential equation in the backward variables (t, x)

$$-\frac{\partial p}{\partial t} = b(t, x)\frac{\partial p}{\partial x} + \frac{1}{2}\sigma^2(t, x)\frac{\partial^2 p}{\partial x^2}, \qquad p(t, x \,|\, t, y) = \delta_y(x).$$

Proof. The proof of Kolmogorov's backward equation is shown in Section 2.2 using a time homogeneous transition function. To adapt the proof to a non-stationary transition function should be possible so this is left as an exercise. The proof of Kolmogorov's forward equation is not given, but is extensively covered in [Pavliotis, 2014, Chapter 4, page 48]. □

One may wonder why the backward equation (2.2.6) differs by a minus sign to the KBE. Since we were discussing time homogeneous processes, our definition of the transition function was $P(x, t \,|\, y, u) =: P(x \,|\, y, u - t)$ so the minus sign crept in via the derivative of $u - t$.

Theorem B.3.2 (Feynman-Kac)**.** *Let $T > 0$ and $X = (X_t)_{t\geq 0}$ be an Itô process started at $X_s = x$ satisfying the PDE*

$$dX_t = b(t, X_t)\, dB_t + \sigma(t, X_t)\, ds, \qquad t \in [s, T].$$

Given that the functions r, f, ψ meet the conditions of [Mao, 2008, Problem (iii), page 82], for $(t, x) \in [0, T] \times \mathbb{R}$ define

$$u(t, x) := \mathbb{E}^x\left[\exp\left(-\int_s^T r(t, X_t)\, dt\right)\psi(X_T)\right] + \mathbb{E}^x\left[\int_s^T \exp\left(-\int_s^t r(v, X_v)\, dv\right) f(t, X_t)\, dt\right].$$

Then the function $u(t, x)$ satisfies

$$\begin{aligned} \frac{\partial u(t, x)}{\partial t} + \mathcal{A}_t u(t, x) - r(t, x)u(t, x) + f(t, x) &= 0, \qquad && (t, x) \in (0, T) \times \mathbb{R}, \\ u(T, x) &= \psi(x), && x \in \mathbb{R}, \end{aligned} \tag{B.3.1}$$

where $\mathcal{A}_t$ is a time-dependent infinitesimal generator for X given by

$$\mathcal{A}_t f(x) = b(t, x)\,\frac{\partial f(x)}{\partial x} + \frac{1}{2}\,\sigma^2(t, x)\,\frac{\partial^2 f(x)}{\partial x^2}, \qquad x \in \mathbb{R},$$

whenever it exists for suitable f. Furthermore, if $w(t, x)$ satisfies (B.3.1)*, then the solution is unique and $w(t, x) \equiv u(t, x)$ holds.*

Proof. The proof that $u(t, x)$ is a solution can be found in [Øksendal, 2013, Thereom 8.2.1, page 135]. To show uniqueness of the solution, simply apply Itô's formula to the process

$$Y_\tau := \exp\left(-\int_s^\tau r(t, X_t)\, dt\right) u(\tau, X_\tau) + \int_s^\tau \exp\left(-\int_s^t r(v, X_v)\, dv\right) f(t, X_t)\, dt. \qquad \square$$

Presented next is the general form of Dirichlet's problem, still under the class of Sturm-Liouville partial differential equations. Recall from Section 2.3 that an argument was constructed to show how the PDE problems involving time could be constructed from the Dirichlet problem without time. The same argument applies here and so Theorem B.3.2 can be seen as a consequence of Theorem B.3.3.

Theorem B.3.3 (Generalised Dirichlet problem)**.** *Suppose that*

(i) $D \subset \mathbb{R}$ be open and bounded,

(ii) r is a non-negative Hölder continuous function on the closure $\bar{D}$,

(iii) ψ is continuous on the boundary ∂D,

(iv) f is Hölder continuous on the closure $\bar{D}$,

(v) the boundary is sufficiently smooth (eg. ∂D is C^2).

Let $\tau_D = \inf\{t \geq 0 : X_t \notin D\}$ be the first exit time of the process from D, where we consider $X = (X_t)_{t\geq 0}$ as an Itô diffusion started at $X_0 = x$. Define

$$u(x) = \mathbb{E}^x\left[\exp\left(-\int_0^{\tau_D} r(X_s)\,ds\right)\psi(X_{\tau_D})\right] + \mathbb{E}^x\left[\int_0^{\tau_D}\exp\left(-\int_0^t r(X_s)\,ds\right)f(X_t)\,dt\right]. \quad \text{(B.3.2)}$$

Then the function $u(x)$ satisfies

$$\begin{aligned} \mathcal{A}u(x) - r(x)u(x) + f(x) &= 0, && x \in D, \\ u(x) &= \psi(x), && x \in \partial D. \end{aligned} \quad \text{(B.3.3)}$$

Furthermore, if $v(x)$ satisfies (B.3.3)*, then the solution is unique and $v(x) \equiv u(x)$ must hold.*

Proof. The proof that $u(x)$ is a solution to (B.3.2) is not shown, but can be found in texts such as [Friedman, 1975], although the details are non-trivial to show and require a lot of PDE theory. To show uniqueness of the solution, the reader can refer to [Mao, 2008, Theorem 8.1, page 79]. The idea is to apply Itô's lemma in a similar way to the proof of uniqueness in Theorem B.3.2, the Feynman-Kac theorem. □

The final result included here proves the martingale property for solutions to the Sturm-Liouville problem for constant eigenfunction parameter r.

Lemma B.3.4. *Let u and v be solutions to the Sturm-Liouville problem as in Section 3.2.1. Then the stochastic processes*

$$\left\{\mathrm{e}^{-rt}u(X_t)\right\}_{t\geq 0} \quad \textit{and} \quad \left\{\mathrm{e}^{-rt}v(X_t)\right\}_{t\geq 0}$$

are martingales, where $X = (X_t)_{t\geq 0}$ is a diffusion process started at $X_0 = x$. Furthermore, u and v are eigenfunctions with eigenvalue r.

Proof. Notice that the solutions to the Sturm-Liouville problem resemble an eigenvalue equation. We call a solution w an eigenvalue solution with eigenvalue r if

$$\mathcal{A}w(x) = rw(x). \quad \text{(B.3.4)}$$

We claimed in Section 2.2 that the definition of an infinitesimal generator was a generalisation of a derivative for random processes. By the proof of Dynkin's formula in Appendix Proposition B.2.5, for the function defined as

$$\Psi_t(w)(x) := \mathbb{E}^x\big[w(X_t)\big],$$

it follows that

$$\frac{\partial}{\partial t}\Psi_t(w)(x) = \mathcal{A}\Psi_t(w)(x) = r\Psi_t(w)(x).$$

This is a standard first order differential equation, which we can solve to get

$$\mathbb{E}^x\big[w(X_t)\big] = \Psi_t(w)(x) = \mathrm{e}^{rt}w(x)$$

To prove the martingale property, consider an application of Itô's lemma to the process $Y_t := \mathrm{e}^{-rt}w(X_t)$. We have that

$$\begin{aligned} Y_t &= Y_0 + \int_0^t \mathrm{e}^{-rs}\Big[\mathcal{A}w(X_s) - rw(X_s)\Big]ds + \int_0^t \mathrm{e}^{-rs}\,\frac{\partial w(X_s)}{\partial x} \\ &= Y_0 + \int_0^t \mathrm{e}^{-rs}\,\frac{\partial w(X_s)}{\partial x}. \end{aligned}$$

By properties of the stochastic integral, as $\mathrm{e}^{-rs}\frac{\partial w(X_s)}{\partial x}$ is square-integrable, the last term is a martingale and so Y is a martingale process. Hence, for any solution w to the Sturm-Liouville problem, $\mathrm{e}^{-rt}w(X_t)$ must be martingale. □

B.4 Convex Analysis

The convex analysis results used in these proofs are highlighted here.

Definition B.4.1 (Epigraph and hypograph)**.** The epigraph of a function $f : \mathbb{R}^n \to \mathbb{R}$ is a subset of $\mathbb{R}^n \times \mathbb{R}$ and is defined as

$$\mathrm{epi}(f) = \{(x, a) \in \mathbb{R}^n \times \mathbb{R} : a \geq f(x)\}.$$

The hyograph of f is defined as

$$\mathrm{hypo}(f) = \{(x, a) \in \mathbb{R}^n \times \mathbb{R} : a \leq f(x)\}.$$

Lemma B.4.2 (Supporting hyperplane)**.** *Any convex function $f : \mathbb{R}^n \to \mathbb{R}$ has an affine minorant. In particular, for any $x_0 \in \mathbb{R}^n$ on the boundary of $\mathrm{epi}(f)$, there exists an affine minorant g which touches f at x_0, ie.*

$$f(x) \geq g(x) \text{ and } f(x_0) = g(x_0), \qquad x, x_0 \in \mathbb{R}^n.$$

Similarly, any concave function $\tilde{f} : \mathbb{R}^n \to \mathbb{R}$ has an affine majorant. In particular, for any $y_0 \in \mathbb{R}^n$ on the boundary of $\mathrm{hypo}(\tilde{f})$, there exists an affine majorant $\tilde{g}$ which touches $\tilde{f}$ at y_0, ie.

$$\tilde{f}(y) \leq \tilde{g}(y) \text{ and } \tilde{f}(y_0) = \tilde{g}(y_0), \qquad y, y_0 \in \mathbb{R}^n.$$

Proof. The proof for the inverse case of a convex minorant can be found in [Krantz, 2014, Proposition 3.23, page 71]. By using the fact the f is convex if and only if $-f$ is concave, we get the result. □

The supporting hyperplane lemma says that for any concave function, there exists a tangent to its curve that always lies above the function. This result will be the basis of the smallest concave majorant proofs. Note the logic to this point: we assumed W and $\hat{W}$ were concave and figured out their forms with this information. Now we are just checking that these two functions are actually concave.

Lemma B.4.3. *The function W_L given by*

$$W(y) = \sup_{\{(y_a, y_b): y_a < y < y_b\}} \left[H_L(y_a)\frac{y_b - y}{y_b - y_a} + H_L(y_b)\frac{y - y_a}{y_b - y_a} \right], \qquad y > 0 \tag{B.4.1}$$

is the smallest concave majorant of $H_L = \frac{h}{v} \circ \psi^{-1}$ with h, v, ψ exactly as defined in Section 3.2.1.

Proof. Recall the definition of W_L from equation (3.2.17). For $y > 0$, the function is defined piecewise as

$$W_L(y) = \begin{cases} y\dfrac{H_L(z)}{z}, & y \leq z, \\ H_L(y), & y \geq z, \end{cases}$$

where $z = \psi(b^*)$. Notice that W_L is a strictly increasing function for all y, since the first segment is a line with positive gradient and the second segment is strictly increasing as H_L is strictly increasing for $y \geq z \geq \psi(x^*)$. W_L is also greater than H_L at all points and is a dominating function. Let

$$J_L(y; y_a, y_b) := \left[H_L(y_a)\frac{y_b - y}{y_b - y_a} + H_L(y_b)\frac{y - y_a}{y_b - y_a} \right].$$

By continuity of H_L, we can take the limit in y_a to get

$$J_L(y; y, y_b) = \lim_{y_a \uparrow y} J_L(y; y_a, y_b) = \lim_{y_a \uparrow y} \left[H_L(y_a)\frac{y_b - y}{y_b - y_a} + H_L(y_b)\frac{y - y_a}{y_b - y_a} \right] = H_L(y).$$

We maintain that $y_a < y < y_b$ is upheld since $y_a \uparrow y$ but $y \neq y_a$ at all. Therefore, $W_L(y) \geq H_L(y)$ for all $y > 0$. To prove that W_L is concave, we use an exhaustive approach. There are two cases we need to address, either when $y \leq z$ or $y > z$.

The second case can easily be dispatched as follows. H_L is concave for $y > \psi(L^*)$ and $y > z = \psi(b^*) \geq \psi(L^*)$. So for any $y > z$, $W_L(y) = H_L(y)$ is a concave function. Clearly W_L dominates (and is equal to) H_L in this instance, so it must be the smallest concave majorant over this region. To prove the other case, we use proof by contradiction. W_L is a concave majorant of H_L, so the only contradicting assumption left to make is that there is a smaller concave majorant to H_L. Let φ be the smallest concave majorant. Then as φ dominates H_L,

$$H_L(y) \leq \varphi(y), \qquad y \leq z$$

and as φ is the smallest concave majorant to H_L, there must exist a point $y_0 < z$ such that

$$H_L(y_0) \leq \varphi(y_0) < W_L(y_0),$$

with strict inequality and $\varphi(y) \leq W_L(y)$ more generally for $y \leq z$. By Lemma B.4.2, there exists an affine majorant to φ at $(y_0, \varphi(y_0))$ that touches the curve on the boundary of its hypograph. Thus, for gradient $m \in \mathbb{R}$,

$$H_L(y) \leq \varphi(y) \leq \varphi(y_0) + m(y - y_0). \tag{B.4.2}$$

Suppose $m \leq 0$. Choosing $y = z$ from equation (B.4.2), we get that

$$\varphi(y_0) < W_L(y_0) = y_0 \frac{H_L(z)}{z} < H_L(z) \leq \varphi(z) \leq \varphi(y_0) + m(z - y_0) \leq \varphi(y_0), \tag{B.4.3}$$

which is a contradiction. Suppose instead that $m < 0$. Then by similar reasoning to equation (B.4.3), we have that

$$H_L(y) \leq \varphi(y) \leq \varphi(y_0) + m(y - y_0) < W(y_0) + m(y - y_0) = y_0 \frac{H_L(z)}{z} + m(y - y_0).$$

Taking $y = z$ again gives that

$$H_L(z) < y_0 \frac{H_L(z)}{z} + m(z - y_0) \iff H_L(z) < mz.$$

On the other hand by continuity of H_L, taking $y \downarrow 0$ in equation (B.4.3), we get that

$$0 = H_L(0) = \lim_{y \downarrow 0} H_L(y) \leq y_0 \frac{H_L(z)}{z} - my_0 \iff mz \leq H_L(z),$$

a contradiction. Therefore, we have proved that W_L is the smallest concave majorant of H_L over two collectively exhaustive disjoint regions covering all y so the statement of the lemma follows. □

Lemma B.4.4. *The function $\hat{W}_L$ given by*

$$\hat{W}_L(y) = \sup_{\{(y_{\hat{a}}, y_{\hat{b}}): y_{\hat{a}} < y < y_{\hat{b}}\}} \left[\hat{H}_L(y_{\hat{a}}) \frac{y_{\hat{b}} - y}{y_{\hat{b}} - y_{\hat{a}}} + \hat{H}_L(y_{\hat{b}}) \frac{y - y_{\hat{a}}}{y_{\hat{b}} - y_{\hat{a}}} \right], \qquad y > 0 \tag{B.4.4}$$

is smallest concave majorant of $\hat{H}_L = \frac{\hat{h}}{\hat{v}} \circ \hat{\psi}^{-1}$ with the corresponding functions $\hat{h}, \hat{v}, \hat{\psi}$ exactly as defined in Section 3.2.1.

Proof. The proof of this lemma follows the lines of Appendix Lemma B.4.3, where we split the problem up into two cases.

Recall that $\hat{W}_L$ is formulated in equation (3.2.21) and is defined for $y > 0$ as

$$\hat{W}_L(y) = \begin{cases} \hat{H}_L(y), & y \leq \hat{z}, \\ \hat{H}_L(\hat{z}), & y \geq \hat{z}, \end{cases}$$

where $\hat{z} = \hat{\psi}(d_*)$. For the same reasons, $\hat{W}_L$ also dominates $\hat{H}_L$. To prove that $\hat{W}_L$ is concave we consider two cases again, either when $y < \hat{z}$ or $y \geq \hat{z}$.

For the first case, $\hat{H}_L$ is concave for $y < \hat{\psi}(L_*)$ and $y < \hat{z} = \hat{\psi}(d_*) < \hat{\psi}(L_*)$. Thus, for any $y \leq \hat{z}$, $\hat{W}_L(y) = \hat{H}_L(y)$ is a concave function and it must be the smallest concave majorant in this region. We now attempt to show concavity for the second case via proof by contradiction. Suppose there exists a function $\hat{\varphi}$ that is the smallest concave majorant instead. Then as $\hat{\varphi}$ dominates $\hat{H}_L$,

$$\hat{H}_L(y) \leq \hat{\varphi}(y), \qquad y \geq \hat{z}.$$

By definition of $\hat{\varphi}$, there must also exist a point $\hat{y}_0 > \hat{z}$ such that

$$\hat{H}_L(\hat{y}_0) \leq \hat{\varphi}(\hat{y}_0) < \hat{W}_L(\hat{y}_0), \tag{B.4.5}$$

with strict inequality and $\hat{\varphi}(y) \leq \hat{W}_L(y)$ holding more generally for $y \geq \hat{z}$. Again, by Lemma B.4.2 we can find a tangent to the curve with gradient $\hat{m} \in \mathbb{R}$. It follows that

$$\hat{H}_L(y) \leq \hat{\varphi}(y) \leq \hat{\varphi}(\hat{y}_0) + \hat{m}(y - \hat{y}_0). \tag{B.4.6}$$

Now suppose that $\hat{m} \geq 0$. By equation (B.4.6), we have for $y = \hat{z}$ that

$$\hat{H}_L(\hat{z}) \leq \hat{\varphi}(\hat{z}) \leq \hat{\varphi}(\hat{y}_0) + \hat{m}(\hat{z} - \hat{y}_0) \leq \hat{\varphi}(\hat{y}_0) < \hat{W}_L(\hat{y}_0) = \hat{H}_L(\hat{z})$$

which is a contradiction. Conversely, for $\hat{m} < 0$ we construct a new argument.

$$\hat{H}_L(\hat{z}) \leq \hat{\varphi}(\hat{z}) \leq \hat{W}_L(\hat{z}) = \hat{H}_L(\hat{z})$$

and so $\hat{\varphi}(\hat{z}) = \hat{H}_L(\hat{z})$. Pick two points y and $\tilde{y}$ such that $\hat{z} < \hat{y}_0 < y < \tilde{y}$ and let $\theta := \frac{\tilde{y}-y}{\tilde{y}-\hat{z}} \in (0,1)$. Then $(1-\theta) = \frac{y-\hat{z}}{\tilde{y}-\hat{z}}$ and

$$\begin{aligned} \theta\hat{H}_L(\hat{z}) + (1-\theta)\hat{H}_L(\tilde{y}) &\leq \theta\hat{\varphi}(\hat{z}) + (1-\theta)\hat{\varphi}(\tilde{y}) \\ &\leq \hat{\varphi}\left(\theta\hat{z} + (1-\theta)\tilde{y}\right) \\ &\leq \hat{\varphi}(\hat{y}_0) + \hat{m}\left(\theta\hat{z} + (1-\theta)\tilde{y} - \hat{y}_0\right). \end{aligned} \tag{B.4.7}$$

In the limit, the individual terms can be calculated as

$$\lim_{\tilde{y}\to\infty} \theta = 1, \quad \lim_{\tilde{y}\to\infty} (1-\theta)\tilde{y} = y - \hat{z}, \quad \lim_{\tilde{y}\to\infty} (1-\theta)\hat{H}_L(\tilde{y}) = 0,$$

where the last limit holds since $\hat{H}_L$ decays at a slow rate, as characterised by the rate of increase of V. Taking the limit as $\tilde{y} \to \infty$ in (B.4.7) leads to the following inequality:

$$\hat{H}_L(\hat{z}) \leq \hat{\varphi}(\hat{y}_0) + \hat{m}\left(y - \hat{y}_0\right) < \hat{\varphi}(\hat{y}_0).$$

Finally, applying the definition of $\hat{W}_L$, we arrive at our contradiction:

$$\hat{W}(\hat{y}_0) = \hat{H}_L(\hat{z}) \leq \hat{\varphi}(\hat{y}_0) + \hat{m}\left(y - \hat{y}_0\right) < \hat{\varphi}(\hat{y}_0),$$

which contradicts the assertion in (B.4.5). Therefore, $\hat{W}_L$ must be the smallest concave majorant to $\hat{H}_L$ for all values of y. □

B.5 Justification of Assumptions: The Initial Problem

The attempts of proofs in this section are dedicated to providing some grounding to the assumptions made in Assumption 3.2.3 and Assumption 3.2.6 regarding the properties of H_L and $\hat{H}_L$ respectively for the long trading problem. The basis for these assumptions comes from the work by [Leung and Li, 2015, Appendix A.2 & A.3, pages 21 and 22] who prove these properties of H_L and $\hat{H}_L$ for the Ornstein-Uhlenbeck process.

Our aim is to extend their work to arbitrary Itô diffusions, allowing the reader to apply the problem to their preferred choice of stochastic process. This treatment is preferential as a trader may not necessarily have a portfolio of assets that exhibits mean-reverting behaviour like the OU process. Inevitably, by having this generalisation we will not be able to provide complete proofs for every property. We will highlight any asymptotic properties of u and v that we need for the arguments to work and will also apply the results to the arithmetic Brownian motion. This should convince the reader that the generalisation was worthwhile.

B.5.1 Assumption 3.2.3

We require the following asymptotic properties:

$$\lim_{x\to-\infty} u(x) = 0, \lim_{x\to-\infty} v(x) = \infty, \lim_{x\to-\infty} v'(x) = -\infty$$

and that the growth rate of $\left|\frac{v(x)}{v'(x)}\right| < Kx^\alpha$.

Recall the definition of H_L in (3.2.5). By continuity and differentiability of h, u, v, ψ and hence ψ^{-1}, assumption (i) follows over the domain $(0, \infty)$. Furthermore, one can verify that

$$H_L(0) = \lim_{x\to-\infty} \frac{(x-c)^+}{v(x)} = 0$$

since the numerator is zero whilst $v(x)$ is always positive, so this proves the first assumption. Continuity of H_L at $y = 0$ follows since $y = \psi(x) \to 0$ as $x \to -\infty$, so

$$\lim_{y\to 0} H_L(y) = \lim_{x\to-\infty} \frac{x-c}{v(x)} = \lim_{x\to-\infty} \frac{1}{v'(x)} = 0$$

by L'Hôpital's rule. The remainder of assumption (ii) explains whether H_L is positive or negative over different regions. Since $v > 0$, the sign of H_L only depends on the sign of $h\big(\psi^{-1}(y)\big)$. As explained in Section 3.2.1, ψ^{-1} is strictly increasing and positive so if $y \in (0, \psi(c))$, then

$$h\big(\psi^{-1}(y)\big) < h\big(\psi^{-1}(\psi(c))\big) = c - c = 0.$$

Similarly, we can verify that $h\big(\psi^{-1}(y)\big) > 0$ for $y \in (\psi(c), \infty)$. Assumption (ii) then follows directly. Next we attempt to justify assumption (iii). For $y = \psi(x)$,

$$\begin{aligned} H_L'(y) &= \left(\frac{h}{v}\right)' (\psi^{-1}(y)) \frac{1}{\psi'(\psi^{-1}(y))} \\ &= \left(\frac{h}{v}\right)' (x) \frac{1}{\psi'(x)} \\ &= \frac{h'(x)v(x) - h(x)v'(x)}{u'(x)v(x) - u(x)v'(x)} && \text{(B.5.1)} \\ &= \frac{v(x) - (x-c)v'(x)}{u'(x)v(x) - u(x)v'(x)}, && \text{(B.5.2)} \end{aligned}$$

where the derivative of ψ^{-1} is given by Appendix Lemma B.1.3. Since the denominator is always positive, the sign of the derivative of H_L depends on the sign of $v(x) - (x-c)v'(x)$. To assess this, we define the function

$$g(x) := (x-c) - \frac{v(x)}{v'(x)}.$$

As $x \to -\infty$, the function $g(x) < 0$. Conversely, for $x = c$,

$$g(c) = -\frac{v(c)}{v'(c)} > 0.$$

Since v is positive, strictly decreasing and convex, $-\frac{v(x)}{v'(x)}$ is strictly increasing and positive. Consequently, g must be strictly increasing. By the intermediate value theorem, there exists a unique root x^* such that $g(x^*) = 0$ if and only if x^* is the solution to $v(x) - (x-c)v'(x) = 0$. This confirms that $H_L'(\psi(x^*)) = 0$. Checking that $H_L'(y) \to 0$ as $y \to \infty$ needs to be done on a case-by-case basis. That $x^* < c$ is valid, since g is strictly increasing.

Finally, by differentiating H_L' again, we can find an expression for the second derivative of H_L. However, there is not much we can say about this quantity without having explicit formulae for u and v. For this reason, assumption (iv) is treated as a pure assumption.

B.5.2 Assumption 3.2.6

We require the following asymptotic properties:

$$\lim_{x\to-\infty} u(x) = 0, \ \lim_{x\to-\infty} u'(x) = 0,$$
$$\lim_{x\to-\infty} \hat{u}(x) = 0, \ \lim_{x\to-\infty} \hat{u}'(x) = 0, \ \lim_{x\to-\infty} \hat{v}(x) = \infty, \ \lim_{x\to-\infty} \hat{v}'(x) = -\infty.$$

The proofs for the properties of $\hat{H}_L$ share a lot of commonality with the arguments for H_L. That $\hat{H}_L$ is continuous on $(0,\infty)$ follows by continuity of all of its components. To assess differentiability of $\hat{H}_L$, we need to investigate the differentiability of $\hat{h}$ and in turn V_L. By equation (3.2.15), it follows that

$$V_L'(x) = \begin{cases} \dfrac{u'(x)}{u'(b^*)} & x \in (-\infty, b^*), \\ 1 & x \in (b^*, \infty), \end{cases}$$

since b^* satisfies $u(b^*) = (b^* - c)u'(b^*)$. The derivative at $x = b^*$ also exists and $V_L'(b^*) = 1$ as the left and right derivatives agree. Moreover, the second derivative is

$$V_L''(x) = \begin{cases} \dfrac{u''(x)}{u'(b^*)} & x \in (-\infty, b^*), \\ 0 & x \in (b^*, \infty). \end{cases}$$

However, the second derivative at $x = b^*$ does not exist for V_L. Differentiability of V_L translates into the differentiability of $\hat{h}$ and as a result into $\hat{H}_L$. Lastly, continuity of $\hat{H}_L$ at $y = 0$ follows because

$$0 \le \hat{H}_L(0) = \lim_{x\to-\infty} \frac{\left(\hat{h}(x)\right)^+}{\hat{v}(x)} = \limsup_{x\to-\infty} \frac{\frac{b^*-c}{u(b^*)} u(x) - x - \hat{c}}{\hat{v}(x)} \le \limsup_{x\to-\infty} \frac{\frac{b^*-c}{u(b^*)} u'(x) - 1}{\hat{v}'(x)} = 0$$

and so $\hat{H}_L(0) = 0$. The inequality is explained in Appendix Lemma B.1.4 and is a variant of L'Ĥopital's rule. In addition,

$$\lim_{y\to 0} \hat{H}_L(y) = \lim_{x\to-\infty} \frac{V_L(x) - x - \hat{c}}{\hat{v}(x)} = \lim_{x\to-\infty} \frac{V_L'(x) - 1}{\hat{v}'(x)} = 0$$

again by L'Ĥopital and so $\hat{H}_L$ is continuous at zero. This proves assumption (i). To assess the sign of $\hat{H}_L$, we need to first investigate properties of $\hat{h}$. For $x \in [b^*, \infty)$, the function is constant with

$$\hat{h}(x) = V_L(x) - x - \hat{c} = x - c - x - \hat{c} = -(c + \hat{c}) < 0$$

whilst for x small enough, $\hat{h}(x) > 0$ by virtue of $\lim_{x\to-\infty} \hat{h}(x) = \infty$. Since the derivative of $\hat{h}$ is given by

$$\hat{h}'(x) = V_L'(x) - 1 = \frac{u'(x)}{u'(b^*)} - 1 < \frac{u'(b^*)}{u'(b^*)} - 1 = 0,$$

the function $\hat{h}$ is strictly decreasing for $x < b^*$. By the IVT and the fact that $\hat{h}$ is strictly monotone, there exists a unique value $k_* \in (-\infty, b^*)$ such that $\hat{h}(k_*) = 0$. As a result,

$$\hat{h}(x) = \begin{cases} > 0 & \text{if } x \in (-\infty, k_*), \\ < 0 & \text{if } x \in (k_*, \infty). \end{cases}$$

By realising that $\hat{\psi}^{-1}$ is a strictly increasing mapping and that $\hat{v}(x) > 0$ for all $x \in \mathbb{R}$, assumption (ii) can be deduced. Assumption (iii) can be verified by calculating the derivative of $\hat{H}_L$, which can be found in Appendix equation (B.5.1), by hatting the quantities. We get for $y = \hat{\psi}(x)$ and $x > b^*$

$$\hat{H}_L'(y) = \frac{\hat{h}'(x)\hat{v}(x) - \hat{h}(x)\hat{v}'(x)}{\hat{u}'(x)\hat{v}(x) - \hat{u}(x)\hat{v}'(x)} = \frac{(c + \hat{c})\hat{v}'(x)}{\hat{u}'(x)\hat{v}(x) - \hat{u}(x)\hat{v}'(x)} < 0, \tag{B.5.3}$$

since the denominator is positive and $\hat{v}'$ is negative. So $\hat{H}_L$ is strictly decreasing for $y > \psi(\hat{b}^*)$. For $x < b^*$, the sign of $\hat{H}_L'$ depends on the sign of

$$\begin{aligned}\hat{h}'(x)\hat{v}(x) - \hat{h}(x)\hat{v}'(x) &= (V_L'(x) - 1)\,\hat{v}(x) - (V_L(x) - x - \hat{c})\,\hat{v}'(x) \\ &= \left(\frac{u'(x)}{u'(b^*)} - 1\right)\hat{v}(x) - \left(\frac{u(x)}{u'(b^*)} - x - \hat{c}\right)\hat{v}'(x),\end{aligned} \tag{B.5.4}$$

but no further progress can be made here without explicit expressions for u and v. As before the final assumption has less of an actual basis for general stochastic processes so we cannot provide supporting ideas to justify its presence within Assumption 3.2.6. This property would need to be checked on a case-by-case basis instead.

B.5.3 Application to the Arithmetic Brownian Motion

Since the properties of H_L and $\hat{H}_L$ have been well-described in the paper by [Leung and Li, 2015] for the OU process, we will instead verify them for an ABM. The procedure is identical for a GBM due to the similar structure of its solutions to the Sturm-Liouville problem. The majority of Assumption 3.2.3 and Assumption 3.2.6 has been proved in Section B.5.1 and Section B.5.2, so we do not need to review everything again.

Recall that the two linearly independent solutions to the Sturm-Liouville problem were:

$$\begin{aligned}u(x) &= \exp\left\{\left(\sqrt{\mu^2 + 2a\sigma^2} - \mu\right)\frac{x}{\sigma^2}\right\}, \\ v(x) &= \exp\left\{\left(\sqrt{\mu^2 + 2a\sigma^2} + \mu\right)\frac{-x}{\sigma^2}\right\}.\end{aligned}$$

One can check the asymptotic properties of the solution to be

$$\begin{aligned}&\lim_{x\to-\infty} u(x) = 0, &&\lim_{x\to-\infty} u'(x) = 0, &&\lim_{x\to-\infty} u''(x) = 0, \\ &\lim_{x\to-\infty} v(x) = \infty, &&\lim_{x\to-\infty} v'(x) = -\infty, &&\lim_{x\to-\infty} v''(x) = \infty, \\ &\lim_{x\to\infty} u(x) = \infty, &&\lim_{x\to\infty} u'(x) = \infty, &&\lim_{x\to\infty} u''(x) = \infty, \\ &\lim_{x\to\infty} v(x) = 0, &&\lim_{x\to\infty} v'(x) = 0, &&\lim_{x\to\infty} v''(x) = 0.\end{aligned}$$

The above results are the same when dealing with the hatted quantities and one just needs to map the terms by $u \mapsto \hat{u}$, $v \mapsto \hat{v}$ and $a \mapsto \hat{a}$.

B.5.3.1 Properties of H_L

Differentiability and continuity of H_L follows by the smooth and continuous nature of u and v. By the asymptotics of the two solutions u and v, all the properties needed to prove assumption (ii) have been verified. For assumption (iii), the growth rate of v is

$$\left|\frac{v(x)}{v'(x)}\right| = \frac{\sigma^2}{\sqrt{\mu^2 + 2a\sigma^2} + \mu}$$

and so the function $g(x) < 0$ as $x \to -\infty$. That $H_L'(y) \to 0$ as $y \to \infty$ follows using identity (B.5.2). For $y = \psi(x)$ with $y \to \infty$ as $x \to \infty$,

$$H_L'(y) = H_L'(\psi(x)) = \frac{v(x) - (x-c)v'(x)}{u'(x)v(x) - u(x)v'(x)} = \frac{\sigma^2 + (x-c)(\mu+\kappa)}{2\kappa}\exp\left\{(\mu - \kappa)\frac{x}{\sigma^2}\right\} \to 0,$$

with $\kappa = \sqrt{\mu^2 + 2a\sigma^2}$. To check the second derivative of H_L, we differentiate again to get

$$H_L''(y) = H_L''(\psi(x)) = \frac{\big(\sigma^2 + (x-c)(\mu+\kappa)\big)(\mu-\kappa) + \sigma^2(\mu+\kappa)}{4\kappa^2}\exp\left\{(\mu - 3\kappa)\frac{x}{\sigma^2}\right\}$$

by an application of the chain rule and through the definition of ψ. The sign of H_L'' depends only on the sign of

$$m(x) := \big(\sigma^2 + (x-c)(\mu+\kappa)\big)(\mu-\kappa) + \sigma^2(\mu+\kappa) = 2\sigma^2(\mu - a(x-c))$$

since the other terms are all positive. Solving for $m(x)=0$, we calculate $L^*=\frac{\mu+ac}{a}$ and that H_L is convex for $y\in(0,\psi(L^*)]$ and concave otherwise. Lastly, we can find the connection between c, x^* and L^*. We can calculate that

$$g(L^*)=(L^*-c)-\frac{v(L^*)}{v'(L^*)}=\frac{\mu}{a}+\frac{\mu+\kappa}{\sigma^2},$$

which depends on the value of μ. As g is a strictly increasing function,

$$\begin{aligned} L^*<x^*<c, &\qquad \text{if } \mu<-\frac{a\kappa}{\sigma^2+a},\\ L^*=x^*<c, &\qquad \text{if } \mu=-\frac{a\kappa}{\sigma^2+a},\\ x^*<c\wedge L^*, &\qquad \text{if } \mu>-\frac{a\kappa}{\sigma^2+a}. \end{aligned}$$

B.5.3.2 Properties of $\hat{H}_L$

Assumption (i) follows by continuity, differentiability and the asymptotic results of u, $\hat{u}$ and $\hat{v}$. Assumptions (ii) and (iii) are proved in full generality, so no further works needs to be done here. We can look into the sign of $\hat{H}_L'$ for $x<b^*$ using equation (B.5.4) to get

$$\hat{h}'(x)\hat{v}(x)-\hat{h}(x)\hat{v}'(x)=\left[\frac{2\hat{\kappa}}{\hat{\kappa}-\mu}\left\{\mathrm{e}^{(\hat{\kappa}-\mu)(x-b^*)/\sigma^2}-\frac{\hat{a}}{\hat{\kappa}}(x+\hat{c})\right\}-1\right]\hat{v}(x), \tag{B.5.5}$$

with $\hat{\kappa}=\sqrt{\mu^2+2\hat{a}\sigma^2}$. As $\hat{v}$ is positive, the sign of $\hat{H}_L'$ depends only on the inner bracketed expression. A simple plot of this function will determine the sign of $\hat{H}_L'$ from here.

Lastly, to determine the convexity of $\hat{H}_L$, we calculate its second derivative. The derivative inherits the two part property from the value function V_L. For $x>b^*$, we can differentiate equation (B.5.3) to achieve

$$\hat{H}_L''(y)=\hat{H}_L''(\hat{\psi}(x))=\frac{\hat{a}\sigma^2}{2\hat{\kappa}^2}(c+\hat{c})\exp\left\{(\mu-3\kappa)\frac{x}{\sigma^2}\right\}>0.$$

For $x<b^*$, one can show that

$$\hat{H}_L'(y)=\hat{H}_L'(\hat{\psi}(x))=\sigma^2\left[\frac{1}{\hat{\kappa}-\mu}\left\{\mathrm{e}^{(\hat{\kappa}-\mu)(x-b^*)/\sigma^2}-\frac{\hat{a}}{\hat{\kappa}}(x+\hat{c})\right\}-\frac{1}{2\hat{\kappa}}\right]\exp\left\{(\mu-\hat{\kappa})\frac{x}{\sigma^2}\right\}$$

using equation (B.5.5) and thus

$$\hat{H}_L''(y)=\hat{H}_L''(\hat{\psi}(x))=\frac{\sigma^2}{2\hat{\kappa}^2}\Big(\hat{a}(x+\hat{c})-\mu\Big)\exp\left\{(\mu-3\hat{\kappa})\frac{x}{\sigma^2}\right\}.$$

The sign of $\hat{H}_L''$ only depends on $\hat{a}(x+\hat{c})-\mu$. Define the function

$$\hat{m}(x)=\hat{a}(x+\hat{c})-\mu$$

The unique solution to $\hat{m}(x)=0$ is $L_*=\frac{\mu-\hat{a}\hat{c}}{\hat{a}}$. For $x\neq b^*$, if $x<L_*$ then $\hat{H}_L$ is concave. Otherwise, for $x>L_*$, the function $\hat{H}_L$ is convex. Convexity at b^* follows by continuity of $\hat{H}_L$ at this point. The condition $L_*<b^*$ also holds here.

B.6 Justification of Assumptions: The Reverse Problem

This section builds on the work from Section B.5. We attempt to verify as many of the properties of H_S and $\hat{H}_S$ as possible for the short trading problem. The last subsection confirms these properties hold for the OU process, giving the reader a complete picture of the long and short optimal stopping problems when combining the work by [Leung and Li, 2015].

B.6.1 Assumption 3.2.10

We require the following asymptotic properties:

$$\lim_{x\to-\infty} u(x) = 0,\ \lim_{x\to-\infty} v(x) = \infty,\ \lim_{x\to-\infty} v'(x) = -\infty$$

and that the growth rate of $\left|\frac{v(x)}{v'(x)}\right| < Kx^{\alpha}$ for $\alpha \in (0,1), K > 0$.

Continuity and differentiability of H_S is straightforward to show. One can verify that

$$H_S(0) = \lim_{x\to-\infty} \frac{\big(-x-c\big)^+}{v(x)} = \lim_{x\to-\infty} \frac{-x-c}{v(x)} = \lim_{x\to-\infty} \frac{-1}{v'(x)} = 0$$

by L'Ĥopital's rule. Continuity of H_S at $y = 0$ follows by the same reasoning since

$$\lim_{y\to 0} H_S(y) = \lim_{x\to-\infty} \frac{-x-c}{v(x)} = \lim_{x\to-\infty} \frac{-1}{v'(x)} = 0,$$

where $y = \psi(x) \to 0$ as $x \to -\infty$. The sign of $H_S\big(\psi^{-1}(y)\big)$ depends only on the sign of $h(x) = -x-c$. This function is positive is $x < c$ and negative for $x > c$. The gradient of H_S can be determined by working out the derivative. For $y = \psi(x)$,

$$\begin{aligned} H_S'(y) &= \frac{h'(x)v(x) - h(x)v'(x)}{u'(x)v(x) - u(x)v'(x)} \\ &= \frac{-v(x) + (x+c)v'(x)}{u'(x)v(x) - u(x)v'(x)}. \end{aligned}$$

Since the denominator is always positive, the sign of the derivative of H_S depends on the sign of $-v(x) + (x+c)v'(x)$. To assess this, we define the function

$$g(x) := -(x+c) + \frac{v(x)}{v'(x)}.$$

As $x \to -\infty$, we have that $g(x) > 0$. Conversely, for $x = -c$,

$$g(-c) = \frac{v(-c)}{v'(-c)} < 0.$$

As g is a strictly decreasing function, there exists a unique root x^* such that $g(x^*) = 0$ if and only if x^* is the solution to $v(x)-(x+c)v'(x) = 0$. Consequently, $H_S'(\psi(x^*)) = 0$ holds. That $x^* < -cc$ is valid, since g is strictly decreasing. We will denote $b^* := x^*$ as it turns out these two points are equal in Theorem 3.2.11. Assumption (iv) is taken for granted without a specific process to work on.

B.6.2 Assumption 3.2.12

We require the following asymptotic properties:

$$\begin{gathered} \lim_{x\to-\infty} u(x) = 0,\ \lim_{x\to-\infty} u'(x) = 0, \\ \lim_{x\to-\infty} \hat{u}(x) = 0,\ \lim_{x\to-\infty} \hat{u}'(x) = 0,\ \lim_{x\to-\infty} \hat{v}(x) = \infty,\ \lim_{x\to-\infty} \hat{v}'(x) = -\infty, \\ \lim_{x\to\infty} v(x) = 0,\ \lim_{x\to\infty} v'(x) = 0, \\ \lim_{x\to\infty} \hat{v}(x) = 0,\ \lim_{x\to\infty} \hat{v}'(x) = 0 \end{gathered}$$

and that $x\hat{v}'(x) \to 0$ as $x \to \infty$.

Differentiability and continuity of $\hat{H}_S$ is easy to show. We highlight the derivatives of V_S briefly. The first derivative is

$$V_S'(x) = \begin{cases} -1 & x \in (-\infty, b^*), \\ -\dfrac{v'(x)}{v'(b^*)} & x \in (b^*, \infty). \end{cases}$$

The derivative also exists at $x = b^*$. The second derivative is

$$V_S''(x) = \begin{cases} 0 & x \in (-\infty, b^*), \\ -\dfrac{v''(x)}{v'(b^*)} & x \in (b^*, \infty), \end{cases}$$

but does not exist for at $x = b^*$. Differentiability of V_S translates into the differentiability of $\hat{h}$ and as a result into $\hat{H}_S$. Continuity of $\hat{H}_S$ at $y = 0$ also holds as

$$\hat{H}_S(0) = \lim_{x\to-\infty} \frac{\left(\hat{h}(x)\right)^+}{\hat{v}(x)} = \lim_{x\to-\infty} \frac{\left(V_S(x) + x - \hat{c}\right)^+}{\hat{v}(x)} = \lim_{x\to-\infty} \frac{-c - \hat{c}}{\hat{v}(x)} = 0.$$

In addition, $\hat{H}_S$ is continuous at zero as

$$\lim_{y\to 0} \hat{H}_S(y) = \lim_{x\to-\infty} \frac{V_S(x) + x - \hat{c}}{\hat{v}(x)} = \lim_{x\to-\infty} \frac{V_S'(x) + 1}{\hat{v}'(x)} = 0,$$

again by L'Ĥopital. To assess the sign of $\hat{H}_S$, we need to first investigate properties of $\hat{h}$. For $x \in (-\infty, b^*]$, the function is constant with

$$\hat{h}(x) = V(x) + x - \hat{c} = -x - c + x - \hat{c} = -(c + \hat{c}) < 0$$

whilst for x large enough, $\hat{h}(x) > 0$. Since the derivative of $\hat{h}$ is given by

$$\hat{h}'(x) = V'(x) + 1 = -\frac{v'(x)}{v'(b^*)} + 1 > -\frac{v'(b^*)}{v'(b^*)} + 1 = 0,$$

the function $\hat{h}$ is strictly increasing for $x > b^*$. By the IVT and the fact that $\hat{h}$ is strictly monotone, there exists a unique value $k_* \in (b^*, \infty)$ such that $\hat{h}(k_*) = 0$. As a result,

$$\hat{h}(x) = \begin{cases} < 0 & \text{if } x \in (-\infty, k_*), \\ > 0 & \text{if } x \in (k_*, \infty). \end{cases}$$

Assumption (ii) follows. Assumption (iii) can be verified by calculating the derivative of $\hat{H}_S$. For $y = \hat{\psi}(x)$ and $x < b^*$,

$$\hat{H}_S'(y) = \frac{\hat{h}'(x)\hat{v}(x) - \hat{h}(x)\hat{v}'(x)}{\hat{u}'(x)\hat{v}(x) - \hat{u}(x)\hat{v}'(x)} = \frac{(c + \hat{c})\hat{v}'(x)}{\hat{u}'(x)\hat{v}(x) - \hat{u}(x)\hat{v}'(x)} < 0.$$

So $\hat{H}_S$ is strictly decreasing for $y < \psi(\hat{b}^*)$. For $x > b^*$, the sign of $\hat{H}_S'$ depends on the sign of

$$\begin{aligned} \hat{h}'(x)\hat{v}(x) - \hat{h}(x)\hat{v}'(x) &= \left(V_S'(x) - 1\right)\hat{v}(x) - \left(V_S(x) - x - \hat{c}\right)\hat{v}'(x) \\ &= \left(1 - \frac{v'(x)}{v'(b^*)}\right)\hat{v}(x) - \left(-\frac{v(x)}{v'(b^*)} + x - \hat{c}\right)\hat{v}'(x), \end{aligned}$$

but no further progress can be made here without explicit expressions for u and v. However, we can check that $\hat{H}_S'(y) \to 0$ as $y = \hat{\psi}(x) \to \infty$. The final assumption would need to be checked on a case-by-case basis like before.

B.6.3 Application to the Ornstein-Uhlenbeck Process

Our aim is to check the properties of H_S and $\hat{H}_S$ for the OU process and finish off the work by [Leung and Li, 2015]. We have given a flavour of what the solution to the ABM looks like and invite the reader to continue it. Since many of the properties have been checked in the general formulation, we only verify the case-specific properties.

Recall that the two linearly independent solutions to the Sturm-Liouville problem were:

$$u(x) = \int_0^\infty y^{\frac{a}{\theta}-1} e^{\sqrt{\frac{2\theta}{\sigma^2}}(x-\mu)y - \frac{y^2}{2}}\, dy,$$
$$v(x) = \int_0^\infty y^{\frac{a}{\theta}-1} e^{\sqrt{\frac{2\theta}{\sigma^2}}(\mu-x)y - \frac{y^2}{2}}\, dy.$$

The solutions u and v share the asymptotic properties mentioned in Section B.5.3 and the same goes for the hatted quantities.

B.6.3.1 Properties of H_S

Assumptions (i) and (ii) hold for any diffusion process. Assumption (iii) has already been determined provided that the growth rate condition of $\frac{v(x)}{v'(x)}$ is true. It is complicated to validate this so we suggest plotting the result instead to gauge the validity of the statement. The main property we have to verify is assumption (iv). From the paper by [Leung and Li, 2015], they calculate

$$H_S''(y) = H_S''(\psi(x)) = \frac{2}{\sigma^2 v(x)(\psi'(x))^2}(\mathcal{A} - a)h(x). \tag{B.6.1}$$

Therefore, the sign of the second derivative only depends on the sign of

$$(\mathcal{A} - a)h(x) = -\theta(\mu - x) - a(-x - c) = (ac - \theta\mu) + (\theta + a)x = \begin{cases} \leq 0 & x \in (-\infty, L^*], \\ \geq 0 & x \in [L^*, \infty), \end{cases}$$

where

$$L^* := \frac{\theta\mu - ac}{\theta + a}.$$

So we get that H_S is concave for $y \in (0, \psi(L^*)]$ and convex outside of this region. Encouragingly, this is close to the expression derived for L_* in the initial long problem for the arithmetic Brownian motion.

B.6.3.2 Properties of $\hat{H}_S$

As for H_S, the majority of assumptions (i)-(iii) have already been proved. The only thing we need to verify is that $\hat{H}_S'(y) \to 0$ as $y = \hat{\psi}(x) \to \infty$. Given the nature of the solution, it is complicated to substantiate this claim, so we also suggest to plot this to convince oneself. Assumption (iv) can be checked using equation (B.6.1) to reveal that

$$\hat{H}_S''(y) = \hat{H}_S''(\hat{\psi}(x)) = \frac{2}{\sigma^2 \hat{v}(x)(\hat{\psi}'(x))^2}(\mathcal{A} - \hat{a})\hat{h}(x),$$

where the sign of $\hat{H}_S''$ depends on $(\mathcal{A} - \hat{a})\hat{h}(x)$. Analysing this produces

$$\begin{aligned}(\mathcal{A} - \hat{a})\hat{h}(x) &= \left[\frac{\sigma^2}{2}V_S''(x) + \theta(\mu - x)V_S'(x)\right] + \theta(\mu - x) - \hat{a}(V_S(x) + x - \hat{c}) \\ &= \begin{cases} \hat{a}(c + \hat{c}) > 0, & x < b^*, \\ (a - \hat{a})V_S(x) - (\theta + \hat{a})x + \theta\mu + \hat{a}\hat{c}, & x > b^*. \end{cases}\end{aligned}$$

Thus, we only need to assess the second derivative's sign for $x > b^*$. Note that V_S is a decreasing function in this range and $a \geq \hat{a}$ by assumption. As a result, $(\mathcal{A} - \hat{a})\hat{h}(x)$ is decreasing for $x > b^*$. Also, for $x \in (b^*, L^*]$,

$$\begin{aligned}(\mathcal{A} - \hat{a})\hat{h}(x) &= (a - \hat{a})V_S(x) - (\theta + \hat{a})x + \theta\mu + \hat{a}\hat{c} \\ &\geq (a - \hat{a})(-x - c) - (\theta + \hat{a})x + \theta\mu + \hat{a}\hat{c} \\ &= -(\theta + a)x + (\theta\mu - ac) + \hat{a}(c + \hat{c}) \\ &\geq -(\theta + a)L^* + (\theta\mu - ac) + \hat{a}(c + \hat{c}) \\ &= \hat{a}(c + \hat{c}) > 0\end{aligned}$$

and $(\mathcal{A} - \hat{a})\hat{h}(x) \to -\infty$ as $x \to -\infty$ since V_S grows at a slow rate. By the IVT, there exists a unique L_* such that $(\mathcal{A} - \hat{a})\hat{h}(L_*) = 0$ with $L_* > b^*$. This completes the verification of assumption (iv).

Bibliography

[Bateman, 1953] Bateman, H. (1953). *Higher transcendental functions [volumes i-iii]*, volume 2. McGraw-Hill Book Company.

[Bateman, 1954] Bateman, H. (1954). *Tables of integral transforms [volumes I & II]*, volume 1. McGraw-Hill Book Company.

[Borodin, 2002] Borodin, A. N. (2002). *Handbook of Brownian motion : facts and formulae*. Probability and its applications. Birkhauser, Basel, 2nd ed. edition.

[Breiman, 1968] Breiman, L. (1968). Probability. *prob*.

[Capocelli and Ricciardi, 1971] Capocelli, R. and Ricciardi, L. (1971). Diffusion approximation and first passage time problem for a model neuron. *Kybernetik*, 8(6):214–223.

[Cohen, 2015] Cohen, S. N. a. (2015). Stochastic calculus and applications.

[Darling et al., 1953] Darling, D. A., Siegert, A., et al. (1953). The first passage problem for a continuous markov process. *The Annals of Mathematical Statistics*, 24(4):624–639.

[Ditlevsen, 2007] Ditlevsen, S. (2007). A result on the first-passage time of an ornstein–uhlenbeck process. *Statistics & probability letters*, 77(18):1744–1749.

[Ethier and Kurtz, 2009] Ethier, S. N. and Kurtz, T. G. (2009). *Markov processes: characterization and convergence*, volume 282. John Wiley & Sons.

[Friedman, 1975] Friedman, A. (1975). *Stochastic differential equations and applications / Vol.1*. Probability and mathematical statistics ; vol.28. Academic Press, New York ; London.

[Gan et al., 2014] Gan, G., Ma, C., and Xie, H. (2014). *Measure, Probability, and Mathematical Finance: A Problem-oriented Approach*. John Wiley & Sons.

[Herrmann and Tanré, 2016] Herrmann, S. and Tanré, E. (2016). The first-passage time of the brownian motion to a curved boundary: an algorithmic approach. *SIAM Journal on Scientific Computing*, 38(1):A196–A215.

[Itô, 1996] Itô, Kiyosi, . (1996). *Diffusion processes and their sample paths*. Classics in mathematics. Springer, Berlin ; London.

[Kessler et al., 1999] Kessler, M., Sørensen, M., et al. (1999). Estimating equations based on eigenfunctions for a discretely observed diffusion process. *Bernoulli*, 5(2):299–314.

[Klebaner, 2005] Klebaner, F. C. (2005). *Introduction to stochastic calculus with applications*. World Scientific Publishing Company.

[Krantz, 2014] Krantz, S. (2014). *Convex analysis*. CRC Press, Boca Raton.

[Leung and Li, 2015] Leung, T. and Li, X. (2015). Optimal mean reversion trading with transaction costs and stop-loss exit. *International Journal of Theoretical and Applied Finance*, 18(03):1550020.

[Mao, 2008] Mao, X. (2008). *Stochastic differential equations and applications*. Horwood, Chichester, 2nd ed. edition.

[Pavliotis, 2014] Pavliotis, G. A. (2014). *Stochastic Processes and Applications: Diffusion Processes, the Fokker-Planck and Langevin Equations*, volume 60 of *Texts in Applied Mathematics*. Springer, New York, NY, 2014 edition.

[Percival and Walden, 1993] Percival, D. B. and Walden, A. T. (1993). *Spectral Analysis for Physical Applications*. Cambridge University Press.

[Rao, 2006] Rao, M. M. a. (2006). Probability theory with applications.

[Ricciardi and Sato, 1988] Ricciardi, L. M. and Sato, S. (1988). First-passage-time density and moments of the ornstein-uhlenbeck process. *Journal of Applied Probability*, pages 43–57.

[Rogers and Williams, 2000] Rogers, L. C. G. and Williams, D. (2000). *Diffusions, Markov Processes, and Martingales*. Cambridge University Press.

[Shreve, 2004] Shreve, S. E. (2004). *Stochastic calculus for finance II: Continuous-time models*, volume 11. Springer Science & Business Media.

[Øksendal, 2013] Øksendal, B. (2013). *Stochastic differential equations: an introduction with applications*. Springer Science & Business Media.