THE UNIVERSITY OF
WARWICK

University of Warwick institutional repository: http://go.warwick.ac.uk/wrap

A Thesis Submitted for the Degree of PhD at the University of Warwick

http://go.warwick.ac.uk/wrap/55821

This thesis is made available online and is protected by original copyright.

Please scroll down to view the document itself.

Please refer to the repository record for this item for information to help you to cite it. Our policy information is available from the repository home page.

On inverse problems in mathematical finance

by

Martin Klimmek

Thesis

Submitted to the University of Warwick

for the degree of

Doctor of Philosophy

Statistics

March 2012

THE UNIVERSITY OF
WARWICK

Contents

THE UNIVERSITY OF WARWICK
S
LIBRARY

[res]
DIS
2012
116

Acknowledgments

It is a great pleasure to thank David G. Hobson for a very generous apprenticeship in mathematics.

Dedicated to Jennifer Fox

Declarations

Chapter 1 is joint work with David Hobson, 'Constructing time-homogeneous generalized diffusions consistent with optimal stopping values', Stochastics, Volume 83:4-6, 2011, pp. 477-503. Chapter 3 is joint work with David Hobson, 'Maximising functionals of the joint law of the maximum and terminal value in the Skorokhod embedding problem', ArXiv 1012.3909, 2012. Chapter 4 is joint work with David Hobson, 'Model independent hedging strategies for variance swaps', ArXiv 1104.4010, 2012.

0.1 Introduction

We consider two inverse problems motivated by questions in mathematical finance. In the first two chapters (Part 1) we recover processes consistent with given perpetual American option prices. In the third and fourth chapters (Part 2) we construct model-independent bounds for prices of contracts based on the realized variance of an asset price process. The two parts are linked by the question of how to recover information about asset price dynamics from option prices: in part one we assume knowledge of perpetual American option prices while in the second part we will assume knowledge of European call and put option prices. Mathematically, the first part of the thesis presents a framework for constructing generalised diffusions consistent with optimal stopping values. The second part aims at constructing bounds for path-dependent functionals of martingales given their terminal distribution.

Consider the idealized situation in which we can obtain prices for call options with a fixed expiry T for all strikes. By a well known argument due to Breeden and Litzenberger [13] this knowledge is equivalent to knowledge of the (market-implied) marginal law of the underlying asset price process at time T. Now suppose that we have a specific view about the variance of the asset price process. Then we may wish to buy or sell a contract called a variance swap which swaps a fixed payment (the price) for a variable payment based on the realized variance. In this thesis we will show how a static position in liquidly traded call options and a dynamic trading strategy in the underlying asset can super-replicate the value of variance swaps for any price path and calculate optimal no-arbitrage price bounds. We thereby solve the inverse problem of recovering information about asset price variability, a property of the realized price trajectory, from call option prices. The tightness of the variance swap bounds that can be constructed using the methods we will describe is a measure of the amount of information about the variability of the asset price process contained in call options.

Dupire [28] introduced the idea of recovering (continuous diffusion) models for the dynamics of an asset price process from European option prices. However, unless we are in Dupire's setting, where call prices are known for all expiries and all strikes and we are willing to restrict the inverse problem to the class of continuous diffusion models, there will be an infinite range of solutions to the inverse problem. Let us consider swapping the idealised situation in which call prices are known for all maturities and strikes for an alternative idealised situation in which perpetual American call options are traded. Perpetual American options are options which can be exercised at any time in perpetuity. In contrast to the prices for European options, perpetual American option prices contain information about the entire expected future dynamics of the asset price process. We will show how to recover exact (and in some cases unique) dynamics of an asset price process consistent with perpetual American option prices.

Inverse problems are practically relevant in finance. The Chicago Board Options Exchange Volatility Index (VIX index) exemplifies the questions we are interested in. The value of the VIX index represents the value of a portfolio of put and call options with a short-term expiry. The portfolio is such that the value of the VIX mimics the square root of the value of $-2\log$ contracts and is said to be a measure of the 'expected future volatility of the S&P 500 index'[1]. But why should the value of a static portfolio of put and call options be a measure of future volatility or variance? Both a one-jump model and a continuous model for the price path can be made consistent with market prices for call options (and thus with a given value of the VIX index). Indeed we will see later that it is easy to construct one-jump models for the asset price process consistent with a finite value for the VIX and an infinite fair value for a variance swap. A continuity assumption is an important caveat in any claim that the value of the VIX is related to realized variance and it is natural to ask how the price for $-2\log$ contracts relates to the range of no-arbitrage prices for a variance swap in a general setting without the continuity assumption. This is a question which falls within an area of mathematical finance concerned with solving the inverse problem of bounding the no-arbitrage prices for derivatives given call prices. In the first article to consider a question of this type, Hobson [41] calculated the range of no-arbitrage values for a bet on the maximum of a price path achieved up to a time T (called the lookback option) consistent with call option prices.

Before considering the case when only European call option prices are known, our first aim is to consider the case when we are given prices for perpetual American options. We can think of the given prices as a value function for a family of perpetual horizon optimal stopping problems parameterised by the strike: given the value function for the stopping problems we wish to recover consistent diffusions. The aim is analogous to the aim of Dupire [28] in an idealised setting where perpetual American options are traded. Infinite horizon stopping problems are of particular interest in real options theory as models for investment decisions. After considering the inverse stopping problem for fixed terminal payoffs we will see how the same theory can be modified to solve inverse stopping problems when there is not only a reward upon stopping (e.g. capital gains), but also a running reward (e.g. dividend payments).

In contrast to perpetual options, vanilla European option prices depend only on the distribution of the price process at a fixed time. Rather than trying to recover exact dynamics from European option prices we will ask how option prices restrict the fair price for contracts based on realized variance. Techniques from the Skorokhod embedding theory are crucial in constructing the bounds introduced in [41]. To construct bounds for the values of variance swaps we prove new optimality results for two well known embeddings, the Perkins and the Azéma-

[1] See for instance 'CBOE Research Notes Volume 1, Issue 2 - VIX - Fact and Fiction', http://www.cboe.com/Products/researchnotes.aspx

Yor embeddings. We then show how these two embeddings are related to the construction of model-independent super- and sub-replication strategies for variance swaps given call option prices.

0.1.1 Links to optimal transportation theory

The two parts of this thesis are linked by their connections to optimal transportation theory. We will briefly outline the link here before returning to discuss it in Chapter 5 based on the ideas presented in Chapters 1-4.

The first two chapters rely on theory from generalised convex analysis which plays a prominent role in the theory of optimal transport. For example: the measure preserving map between two measures μ and ν on $\mathbb{R}$ that minimises the cost of transportation (where the cost of transporting a unit of mass from x to y is e.g. $c(x, y) = (y - x)^2$) can be expressed in terms of a c-convex function and its c-subdifferential, see Gangbo and McCann [36]. Transportation plans of this type are simply monotone functions from the support of one measure to the support of the other. In the first chapter we represent optimal stopping thresholds for perpetual horizon stopping problems as monotone functions of the problem-parameter (which is the strike in the financial context). From the perspective of the underlying theory of generalised convex analysis, the monotone stopping thresholds are related to the monotone transportation plans in the classical deterministic optimal transportation problem.

In Chapter 3 we relate the problem of finding model independent bounds for variance swaps given call prices to a Skorokhod embedding problem. The Skorokhod embedding problem we consider is to minimise (or maximise) a path-dependent functional over all square integrable martingales started at a fixed point and with a given time T marginal distribution. The problem can be interpreted as an optimal transportation problem (from an atomic mass at the starting point to a given terminal distribution) with the constraint that the (stochastic) optimal transportation plan is a square integrable martingale. The idea of 'stochastic mass transportation' originates in a stochastic interpretation of the Schrödinger equation due to Nelson (see Villani [74] for an overview), where the interpolating stochastic processes are diffusions.

0.1.2 Models and prices; a brief survey of the theory

Breeden and Litzenberger [13] show that any sufficiently regular derivative contract depending on the asset price at a fixed point in time (e.g. $-2\log$ contracts) can be replicated by a portfolio of call options. Knowledge of call option prices with a given expiry for a continuum of strikes is equivalent to knowledge of the marginal law of the asset price at the expiry time. A well known example of a model which generates call prices from given data is the Black-Scholes model, see e.g. [10]. Taking the inverse view, Dupire showed how to derive models from call prices [28]. If vanilla option prices can be obtained not only for all strikes at a fixed time but for all possible maturities and strikes then under some regularity assumptions it is possible to recover a unique diffusion model for the asset price process.

Suppose we assume that the asset (forward) price process is continuous. Neuberger [58] and Dupire [27] showed that a continuously monitored variance swap is perfectly replicated by the following strategy: synthesize $-2\log$ contracts using put and call options and trade continuously in the asset to hold a number of shares equal to twice the reciprocal of the current asset price at all times. (We will call this hedge the standard hedge for variance swaps.) By the arguments of Breeden and Litzenberger [13], $-2\log$ contracts can be replicated with a portfolio of puts and calls. Thus if asset prices are continuous, call and put options contain complete information about the market expectation for the realised variance of the asset. We recognise this situation as the one upon which the VIX index is based. If, however, the continuity assumption is violated then the standard hedge for variance swaps fails. The first analysis of the discrepancy between realized variance and the value replicated by the standard hedge in the presence of jumps is due to Demeterfi et. al. [24]. Let us consider a situation in which the solution to the inverse problem is trivial under a continuity assumption but non-trivial without it: suppose that call options imply that the log-transformed asset price process at every time $t > 0$ is normally distributed with variance t. If we assume that the consistent process is continuous then following Dupire [28] we would recover geometric Brownian motion as the unique model consistent with option prices. According to the arguments behind the standard hedge the fair value of the variance swap would be the expected value of $-2\log$ contracts on a log-normal random variable. Now suppose that we drop the continuity assumption. There exist alternative processes that are not Brownian motion which have the same marginals, known as fake Brownian Motions - Hamza and Klebaner [38] construct a pure-jump fake Brownian Motion. The exponent of this pure jump process will be consistent with the same observed call prices but its realized variance will not be hedged by following the standard hedge for variance swaps. (See Section 4.2 for a detailed discussion of variance swap values in the presence of jumps.)

The no-arbitrage bounds introduced by Hobson rely on fewer assumptions than Dupire's results. Apart from assuming that prices can be expressed as the discounted expected payoff under an equivalent martingale measure, the only other assumptions are zero transaction costs and knowledge of call prices at the expiry time of the option whose price we wish to bound. (In practice, if only a finite number of call option prices are available, then we can nevertheless interpolate between the known prices to satisfy our assumption.) The construction of bounds given call option prices relies strongly on the optimality properties of some key Skorokhod embeddings, see Hobson [42] for a survey of bounds and the corresponding Skorokhod embedding theory. Apart from the lookback option, bounds have been constructed for barrier options ([14], [22]), forward starting straddles ([40]) and variance options ([16] and more recently, [20]).

Motivated by Dupire's approach to recovering price dynamics from European calls and puts, a first attempt at recovering a diffusion process from perpetual American call and put options was made by Alfonsi and Jourdain [1, 2, 3]. Hobson and Ekström [30] considered the problem in a less restrictive setting where the underlying proceses are generalized diffusions. While the results by Alfonsi and Jourdain rely on solving differential equations, the results in [30] are simplified by convex analysis. More recently, Lu [52] employs the techniques developed in [30] to recover diffusions from a finite set of put option prices.

0.1.3 Overview

The inverse perpetual optimal stopping problem is introduced in Chapter 1. To solve it we require a theory of parameter dependence in the standard forward problem. We make extensive use of generalized convex analysis to solve both the forward and the inverse problem. In Chapter 2 we generalize to the situation when the stopping problem may include a running reward. An interpretation of the Gittins index as the inverse of the stopping threshold allows us to make sense of inverse problems in this setting.

In Chapter 3 we introduce the problem of bounding the fair value for bets on realized variance. We introduce the Perkins [61] and the Azéma-Yor [7] solutions to the Skorokhod embedding problem and derive new optimality properties. These optimality properties are used to construct model-independent bounds for a continuously monitored variance swap based on squared returns. Chapter 4 is based on an effort to explain and understand recent work by Kahalé [47] which derives model-independent lower bounds for variance swaps based on the squared log-returns. We extend his results to derive model-independent super and sub-replication strategies for a family of variance swaps. In the continuous time limit, the corresponding price bounds are related to the optimality properties of the Skorokhod embeddings in Chapter 3.

In Chapter 5 we provide an overview of the links between the classical optimal transportation problem and the material presented in this thesis, and outline some ongoing projects and further work.

Chapter 1

Recovering one-dimensional generalised diffusions from perpetual option prices

Given a discount parameter, an objective function and a time-homogeneous diffusion started at a fixed point, a classical optimal stopping problem is to maximise the discounted payoff over all stopping times for the diffusion. We will call this problem the forward optimal stopping problem, and the expected payoff under the optimal stopping rule the (forward) problem value. Suppose that we are given the problem value for a one-parameter family of objective functions, for example perpetual American call option prices for a continuum of strikes. The inverse problem is to recover a time-homogeneous diffusion consistent with the given problem values (prices).

1.1 The Forward and the Inverse Problems

Let $\mathcal{X}$ be a class of diffusion processes, let ρ be a discount parameter, and let $\mathcal{G} = \{G(x,\theta); \theta \in \Theta\}$ be a family of non-negative objective functions, parameterised by a real parameter θ which lies in an interval Θ. The forward problem, which is standard in optimal stopping, is for a given $X \in \mathcal{X}$, to calculate for each $\theta \in \Theta$, the problem value

$$V(\theta) \equiv V_X(\theta) = \sup_{\tau} \mathbf{E}_0[e^{-\rho\tau} G(X_\tau, \theta)], \tag{1.1.1}$$

where the supremum is taken over finite stopping times τ, and $\mathbf{E}_0$ denotes the fact that $X_0 = 0$. The inverse problem is, given a fixed ρ and the family $\mathcal{G}$, to determine whether $V \equiv \{V(\theta) : \theta \in \Theta\}$ could have arisen as a solution to the family of problems (1.1.1) and if so, to characterise those elements $X \in \mathcal{X}$ which would lead to the value function V. The inverse problem, which is the main object of our analysis, is much less standard than the forward problem, but has recently been the subject of some studies ([3, 2, 30]) in the context of perpetual American options. In these papers the space of candidate diffusions is $\mathcal{X}_{stock}$, where $\mathcal{X}_{stock}$ is the set of price processes which, when discounted, are martingales and $G(x,\theta) = (\theta - x)^+$ is the put option payoff (slightly more general payoffs are considered in [2]). The aim is to find a stochastic model which is consistent with an observed continuum of perpetual put prices.

In fact it will be convenient to extend the set $\mathcal{X}$ to include the set of generalised diffusions in the sense of Itô and McKean [44]. These diffusions are generalised in the sense that the speed measure may include atoms, or regions with zero or infinite mass. Generalised diffusions are characterised by a speed measure and a strictly increasing and continuous scale function and can be constructed via time-changes of Brownian Motion, see [44], [51], [67], [31] and for a setup related to the one considered here, [30].

Let m be a non-negative, non-zero Borel measure on $\mathbb{R}$ and let $I = supp(m)$. Let $s : I \to \mathbb{R}$ be a strictly increasing and continuous function. Let $B = (B_t)_{t \geq 0}$ be a Brownian Motion started at $B_0 = s(X_0)$ supported on a filtration $\mathbb{F}^B = (\mathcal{F}^B_u)_{u \geq 0}$ with local time process $\{L^z_u; u \geq 0, z \in \mathbb{R}\}$. Define Γ to be the left-continuous, increasing, additive functional

$$\Gamma_u = \int_{\mathbb{R}} L^z_u m(dz),$$

and define its right-continuous inverse by

$$A_t = \inf\{u : \Gamma_u > t\}.$$

If $X_t = s^{-1}(B(A_t))$ then $X = (X_t)_{t \geq 0}$ is a one-dimensional regular diffusion with 'data' m and s and $X_t \in I$ almost surely for all $t \geq 0$.

Let $H_x = \inf\{u : X_u = x\}$. Then (see Borodin and Salminen [12] Chapter I. or Rogers and Williams [67] Chapter V.),

$$\xi_x(y) = \mathbb{E}_x[e^{-\rho H_y}] = \left\{ \begin{array}{cc} \frac{\varphi(x)}{\varphi(y)} & x \leq y \\ \frac{\phi(x)}{\phi(y)} & x \geq y \end{array} \right\}. \tag{1.1.2}$$

where φ and ϕ are respectively a strictly increasing and a strictly decreasing solution to the differential equation

$$\frac{1}{2}\frac{d}{dm}\frac{d}{ds}f = \rho f. \tag{1.1.3}$$

1.1.1 The class $\mathcal{X}_0$

In this chapter we will concentrate on the set of generalised diffusions started and reflected at 0, which are local martingales (at least when away from zero). We denote this class $\mathcal{X}_0$. (Alternatively we can think of an element X as the modulus of a local martingale Y whose characteristics are symmetric about the initial point zero.) We assume a natural right boundary but we do not exclude the possibility that it is absorbing. Away from zero the process is in natural scale and can be characterised by its speed measure, and in the case of a classical diffusion by the diffusion coefficient σ. In that case we may consider $X \in \mathcal{X}_0$ to be a solution of the SDE (with reflection)

$$dX_t = \sigma(X_t)dB_t + dL_t \qquad\qquad X_0 = 0,$$

where L is the local time at zero.

In Chapter 2 we will consider a more general setup with non-trivial scale functions and a running reward function. The twin reasons for focusing on $\mathcal{X}_0$ rather than $\mathcal{X}$ in this chapter, are that the optimal stopping problem is guaranteed to become one-sided rather than two-sided, and that within $\mathcal{X}_0$ there is some hope of finding a unique solution to the inverse problem. The former reason is more fundamental (we will comment in Section 1.5.2 below on other plausible choices of subsets of $\mathcal{X}$ for which a similar approach is equally fruitful).

A generalised diffusion $X \in \mathcal{X}_0$ is identified solely by its speed measure m. Let m be a non-negative, non-decreasing and right-continuous function which defines a measure on $\mathbb{R}^+$, and let m be identically zero on $\mathbb{R}^-$. We call x a point of growth of m if $m(x_1) < m(x_2)$ whenever $x_1 < x < x_2$ and denote the closed set of points of growth by E. Then m may assign mass to 0 or not, but in either case we assume $0 \in E$. We also assume that if $\xi = \sup\{x : x \in E\}$ then $\xi + m(\xi+) = \infty$. If $\xi < \infty$ then either ξ is an absorbing endpoint, or X does not reach ξ in finite time. The diffusion X with speed measure m is defined on $[0, \xi)$ and is constructed via a time-change of Brownian motion.

For a given diffusion $X \in \mathcal{X}_0$, $\varphi(x) \equiv \varphi_X(x)$ is defined via $\varphi_X(x) = (\mathbb{E}_0[e^{-\rho H_x}])^{-1}$. It is well known (see for example [67, V.50] and [29, pp 147-152]) that φ_X is the unique increasing, convex solution to the differential equation

$$\frac{1}{2}\frac{d^2 f}{dm dx} = \rho f; \qquad f(0) = 1, \qquad f'(0-) = 0. \tag{1.1.4}$$

Conversely, given an increasing convex function φ with $\varphi(0) = 1$ and $\varphi'(0+) \geq 0$, (1.1.4) can be used to define a measure m which in turn is the speed measure of a generalised diffusion $X \in \mathcal{X}_0$.

If $m(\{x\}) > 0$ then the process X spends a positive amount of time at x. If $x \in E$ is an isolated point, then there is a positive holding time at x, conversely, if for each neighbourhood N_x of x, m also assigns positive mass to $N_x \setminus \{x\}$, then x is a sticky point.

If $X \in \mathcal{X}_0$ and m has a density, then $m(dx) = \sigma(x)^{-2}dx$ where σ is the diffusion coefficient of X and the differential equation (1.1.4) becomes

$$\frac{1}{2}\sigma(x)^2 f''(x) - \rho f(x) = 0. \tag{1.1.5}$$

In this case, depending on the smoothness of g, v will also inherit smoothness properties. Conversely, 'nice' v will be associated with processes solving (1.1.5) for a smooth σ. However, rather than pursuing issues of regularity, we prefer to work with generalised diffusions.

We return to the (forward) optimal stopping problem: For fixed X define $\varphi(x) = \varphi_X(x) = \mathbb{E}_0[e^{-\rho H_x}]^{-1}$, where H_x is the first hitting time of level x. Let

$$\hat{V}(\theta) = \sup_{x:\varphi(x)<\infty} \left[G(x,\theta)\, \mathbb{E}_0[e^{-\rho H_x}]\right] = \sup_{x:\varphi(x)<\infty} \left[\frac{G(x,\theta)}{\varphi(x)}\right]. \tag{1.1.6}$$

Clearly $V \geq \hat{V}$. Indeed, as the following lemma shows, there is equality and for the forward problem (1.1.1), the search over all stopping times can be reduced to a search over first hitting times.

Lemma 1.1.1. *V and $\hat{V}$ coincide.*

Proof. Clearly $V \geq \hat{V}$, since the supremum over first hitting times must be less than or equal to the supremum over all stopping times.

Conversely, by (1.1.6), $\varphi(x) \geq \frac{G(x,\theta)}{\hat{V}(\theta)}$. Moreover, (1.1.4) implies that $e^{-\rho t}\varphi(X_t)$ is a non-negative local martingale and hence a supermartingale. Thus for stopping times τ we have

$$1 \geq \mathbb{E}_0[e^{-\rho\tau}\varphi(X_\tau)] \geq \mathbb{E}_0[e^{-\rho\tau}G(X_\tau,\theta)/\hat{V}(\theta)]$$

and hence $\hat{V}(\theta) \geq \sup_\tau E_0[e^{-\rho\tau}G(x_\tau,\theta)]$. □

The first step in our approach will be to take logarithms which converts a multiplicative problem into an additive one. Introduce the notation

$$\begin{aligned} v(\theta) &= \log(V(\theta)), \\ g(x,\theta) &= \log(G(x,\theta)), \\ \psi(x) &= \log(\mathbb{E}_0[e^{-\rho H_x}]^{-1}) = \log\varphi(x). \end{aligned}$$

Then the equivalent log-transformed problem (compare (1.1.6)) is

$$v(\theta) = \sup_x [g(x, \theta) - \psi(x)], \tag{1.1.7}$$

where the supremum is taken over those x for which $\psi(x)$ is finite. To each of these quantities we may attach the superscript X if we wish to associate the solution of the forward problem to a particular diffusion. We call φ_X the eigenfunction (and ψ_X the log-eigenfunction) associated with X.

In the case where $g(x, \theta) = \theta x$, v and ψ are convex duals. More generally the relationship between v and ψ is that of u-convexity ([69], [62], [75]). In Section 1.2 we give the definition of the u-convex dual f^u of a function f, and derive those properties that we will need. For our setting, and under mild regularity assumptions on the functions g, see Assumption 1.2.6 below, we will show that there is a duality relation between v and ψ via the log-payoff function g which can be exploited to solve both the forward and inverse problems. In particular our main results (see Proposition 1.3.4 and Theorems 1.4.1 and 1.4.4 for precise statements) include:

Forward Problem: Given a diffusion $X \in \mathcal{X}_0$, let $\varphi_X(x) = (\mathbb{E}_0[e^{-\rho H_x}])^{-1}$ and $\psi_X(x) = \log(\varphi_X(x))$. Set $\psi^g(\theta) = \sup_x\{g(x, \theta) - \psi(x)\}$. Then the solution to the forward problem is given by $V(\theta) = \exp(\psi^g(\theta))$, at least for those θ for which there is an optimal, finite stopping rule. We also find that V is locally Lipschitz over the same range of θ.

Inverse Problem: For $v = \{v(\theta) : \theta \in \Theta = [\theta_-, \theta_+]\}$ to be logarithm of the solution of (1.1.1) for some $X \in \mathcal{X}_0$ it is sufficient that the g-convex dual (given by $v^g(x) = \sup_\theta\{g(x, \theta) - v(\theta)\}$) satisfies $v^g(0) = 0$, $e^{v^g(x)}$ is convex and increasing, and $v^g(x) > \{g(x, \theta_-) - g(0, \theta_-)\}$ for all $x > 0$.

Note that in stating the result for the inverse problem we have assumed that Θ contains its endpoints, but this is not necessary, and our theory will allow for Θ to be open and/or unbounded at either end.

If X is a solution of the inverse problem then we will say that X is consistent with $\{V(\theta); \theta \in \Theta\}$. By abuse of notation we will say that φ_X (or ψ_X) is consistent with V (or $v = \log V$) if, when solving the optimal stopping problem (1.1.1) for the diffusion with eigenfunction φ_X, we obtain the problem values $V(\theta)$ for each $\theta \in \Theta$.

The main technique in the proofs of these results is to exploit (1.1.7) to relate the fundamental solution φ with V. Then there is a second part of the problem which is to relate φ to an element of $\mathcal{X}$. In the case where we restrict attention to $\mathcal{X}_0$, each increasing convex φ with $\varphi(0) = 1$ is associated with a unique generalised diffusion $X \in \mathcal{X}_0$. Other choices of subclasses of $\mathcal{X}$ may or may not have this uniqueness property. See the discussion in Section 1.4.6.

The following examples give an idea of the scope of the problem:

Example 1.1.2. *Forward Problem: Suppose $G(x,\theta) = e^{x\theta}$. Let $m > 1$ and suppose that $X \in \mathcal{X}_0$ solves $dX = \sigma(X)dW + dL$ for $\sigma(x)^{-2} = (x^{2(m-1)} + (m-1)x^{m-2})/(2\rho)$. For such a diffusion $\varphi(x) = \exp(\frac{1}{m}x^m), x \geq 0$. Then for $\theta \in \Theta = (0,\infty)$, $V(\theta) = \exp(\frac{m-1}{m}\theta^{\frac{m}{m-1}})$.*

Example 1.1.3. *Forward Problem: Let X be reflecting Brownian Motion on the positive half-line with a natural boundary at ∞. Then $\varphi(x) = \cosh(x\sqrt{2\rho})$. Let $g(x,\theta) = \theta x$ so that g-convexity is standard convexity, and suppose $\Theta = (0,\infty)$. Then*

$$v(\theta) = \sup_x [\theta x - \log(\cosh(x\sqrt{2\rho}))].$$

It is easy to ascertain that the supremum is attained at $x = x^*(\theta)$ where

$$x^*(\theta) = \frac{1}{\sqrt{2\rho}} \tanh^{-1}\left(\frac{\theta}{\sqrt{2\rho}}\right) \tag{1.1.8}$$

for $\theta \in [0, \sqrt{2\rho})$. Hence, for $\theta \in (0, \sqrt{2\rho})$

$$\begin{aligned} v(\theta) &= \frac{\theta}{\sqrt{2\rho}} \tanh^{-1}\left(\frac{\theta}{\sqrt{2\rho}}\right) - \log\left(\cosh \tanh^{-1}\left(\frac{\theta}{\sqrt{2\rho}}\right)\right) \\ &= \frac{\theta}{\sqrt{2\rho}} \tanh^{-1}\left(\frac{\theta}{\sqrt{2\rho}}\right) + \frac{1}{2}\log\left(1 - \frac{\theta^2}{2\rho}\right), \end{aligned}$$

with limits $v(0) = 0$ and $v(\sqrt{2\rho}) = \log 2$. For $\theta > \sqrt{2\rho}$ we have $v(\theta) = \infty$.

Example 1.1.4. *Inverse Problem: Suppose that $g(x,\theta) = \theta x$ and $\Theta = (0, \sqrt{2\rho})$. Suppose also that for $\theta \in \Theta$*

$$V(\theta) = \exp\left(\frac{\theta}{\sqrt{2\rho}} \tanh^{-1}\left(\frac{\theta}{\sqrt{2\rho}}\right) + \frac{1}{2}\log\left(1 - \frac{\theta^2}{2\rho}\right)\right).$$

Then X is reflecting Brownian Motion.

Note that $X \in \mathcal{X}_0$ is uniquely determined, and its diffusion coefficient is specified on $\mathbb{R}^+$. In particular, if we expand the domain of definition of Θ to $(0,\infty)$ then for consistency we must have $V(\Theta) = \infty$ for $\theta > \sqrt{2\rho}$.

Example 1.1.5. *Inverse Problem: Suppose $G(x,\theta) = x^\theta$ and $V(\theta) = \{\frac{\theta^{\frac{\theta}{2}}(2-\theta)^{\frac{2-\theta}{2}}}{2} : \theta \in (1,2)\}$. Then $\varphi(x) = 1 + x^2$ for $x > 1$ and, at least whilst $X_t > 1$, X solves the SDE $dX = \rho(1+X)^2 dW$. In particular, V does not contain enough information to determine a unique consistent diffusion in $\mathcal{X}_0$ since there is some indeterminacy of the diffusion co-efficient on $(0,1)$.*

Example 1.1.6. *Inverse Problem: Suppose $g(x,\theta) = -\theta^2/(2\{1+x\})$, $\Theta = [1,\infty)$ and $v(\theta) = \{-1/2 - \log\theta : \theta \geq 1\}$. Then the g-dual of v is given by $v^g(x) = \log(1+x)/2$, $x \geq 0$ and is a candidate for ψ. However $e^{v^g(x)} = \sqrt{1+x}$ is not convex. There is no diffusion in $\mathcal{X}_0$ consistent with V.*

Example 1.1.7. *Forward and Inverse Problem: In special cases, the optimal strategy in the forward problem may be to 'stop at the first hitting time of infinity' or to 'wait forever'. Nonetheless, it is possible to solve the forward and inverse problems.*

Let h be an increasing, differentiable function on $[0, \infty)$ with $h(0) = 1$, such that e^h is convex; let f be a positive, increasing, differentiable function on $[0, \infty)$ such that $\lim_{x\to\infty} f(x) = 1$; and let $w(\theta)$ be a non-negative, increasing and differentiable function on $\Theta = [\theta_-, \theta_+]$ with $w(\theta_-) = 0$.

Suppose that

$$g(x, \theta) = h(x) + f(x)w(\theta).$$

Note that the cross-derivative $g_{x\theta}(x, \theta) = f'(x)w'(\theta)$ is non-negative.

Consider the forward problem. Suppose we are given a diffusion in $\mathcal{X}_0$ with log-eigenfunction $\psi = h$. Then the log-problem value v is given by

$$v(\theta) = \psi^g(\theta) = \sup_{x\geq 0}\{g(x, \theta) - \psi(x)\} = \limsup_{x\to\infty}\{f(x)w(\theta)\} = w(\theta).$$

Conversely, suppose we are given the value function $V = e^w$ on Θ. Then

$$w^g(x) = \sup_{\theta\in\Theta}\{g(x, \theta) - w(\theta)\} = \sup_{\theta\in\Theta}\{h(x) + (f(x) - 1)w(\theta)\} = h(x)$$

is the log-eigenfunction of a diffusion $X \in \mathcal{X}_0$ which solves the inverse problem.

1.2 u-convex Analysis

In the following we will consider u-convex functions for $u = u(y, z)$ a function of two variables y and z. There will be complete symmetry in role between y and z so that although we will discuss u-convexity for functions of y, the same ideas apply immediately to u-convexity in the variable z. Then, in the sequel we will apply these results for the function g, and we will apply them for g-convex functions of both x and θ.

For a more detailed development of u-convexity, see in particular Rachev and Rüschendorf [62] and the references contained therein.

Let D_y and D_z be sub-intervals of $\mathbb{R}$. We suppose that $u : D_y \times D_z \mapsto \bar{\mathbb{R}}$ is well defined, though possibly infinite valued.

Definition 1.2.1. $f : D_y \to \mathbb{R}^+$ is u-convex iff there exists a non-empty $S \subset D_z \times \mathbb{R}$ such that for all $y \in D_y$

$$f(y) = \sup_{(z,a)\in S}[u(y, z) + a].$$

Definition 1.2.2. The u-dual of f is the u-convex function on D_z given by

$$f^u(z) = \sup_{y \in D_y} [u(y, z) - f(y)].$$

The u-dual of a function on D_z is defined analogously.

A fundamental fact from the theory of u-convexity is the following ([62], Proposition 3.3.5):

Lemma 1.2.3. *A function f is u-convex iff $(f^u)^u = f$.*

The function $(f^u)^u$ (the u-convexification of f) is the greatest u-convex minorant of f. The condition $(f^u)^u = f$ provides an alternative definition of a u-convex function, and is often preferred; checking whether $(f^u)^u = f$ is usually more natural than trying to identify the set S.

Diagrammatically (see Figure 1.1), we can think of $-(f^u)(z) = \inf_y [f(y) - u(y, z)]$ as the vertical distance between f and $u(., z)$. Thus $f^u(z) \leq 0$ when $f(y) \geq u(y, z)$ for all $y \in D_y$.

The following description found in Villani [75] is helpful in visualising what is going on: f is u-convex if at every point y we can find a parameter z so that we can *caress* f from below with $u(., z)$. A development of this description and its application to the Monge-Kantorovich problem can be found in Rüschendorf and Uckelmann [70].

The definition of the u-dual implies a generalised version of the Young inequality (familiar from convex analysis, e.g. [64]),

$$f(y) + f^u(z) \geq u(y, z)$$

for all $(y, z) \in D_y \times D_z$. Equality holds at pairs (y, z) where the supremum

$$\sup_z [u(y, z) - f^u(z)]$$

is achieved, see also [62] (Proposition 3.3.3).

Definition 1.2.4. The u-subdifferential of f at y is defined by

$$\partial^u f(y) = \{z \in D_z : f(y) + f^u(z) = u(y, z)\},$$

or equivalently

$$\partial^u f(y) = \{z \in D_z : u(y, z) - f(y) \geq u(\hat{y}, z) - f(\hat{y}), \forall \hat{y} \in D_y\}.$$

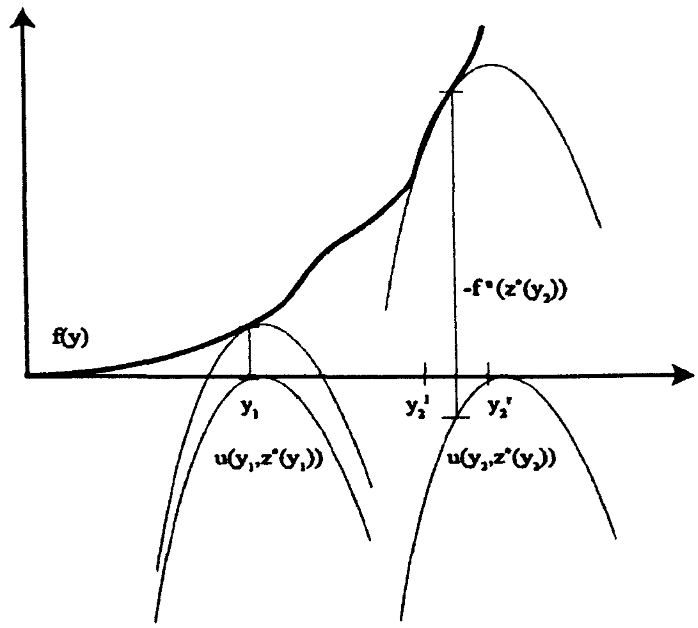

Figure 1.1: f is u-subdifferentiable. $\partial^u f(y_1) = z^*(y_1)$ and $\partial^u(y_2) = z^*(y_2)$ for $y_2 \in (y_2^l, y_2^r)$. The distance between $u(.,z)$ and f is equal to -$f^u(z)$. Note that the u-subdifferential is constant over the interval (y_2^l, y_2^r)

If U is a subset of D_y then we define $\partial^u f(U)$ to be the union of u-subdifferentials of f over all points in U.

Definition 1.2.5. f is u-subdifferentiable at y if $\partial^u f(y) \neq \emptyset$. f is u-subdifferentiable on U if it is u-subdifferentiable for all $y \in U$, and f is u-subdifferentiable if it is u-subdifferentiable on $U = D_y$.

In what follows it will be assumed that the function $u(y,z)$ is satisfies the following 'regularity conditions'.

Assumption 1.2.6.

(a) $u(y,z)$ is continuously twice differentiable.

(b) $u_y(y,z) = \frac{\partial}{\partial y}u(y,z)$ as a function of z, and $u_z(y,z) = \frac{\partial}{\partial z}u(y,z)$ as a function of y, are strictly increasing.

Remark 1.2.7. We will see below that by assuming 1.2.6(a) irregularities in the value function (1.1.1) can be identified with extremal behaviour of the diffusion.

Remark 1.2.8. Condition 1.2.6(b) is known as the single crossing property and as the Spence-Mirrlees condition ([15]). If instead we have the 'Reverse Spence-Mirrlees condition':
(bb) $u_y(y,z)$ as a function of z, and $u_z(y,z)$ as a function of y, are strictly decreasing, then there is a parallel theory, see Remark 1.2.12.

The following results from u-convex analysis will be fundamental in our application of u-convex analysis to finding the solutions of the forward and inverse problems. The idea, which goes back to Rüschendorf [68] (Equation 73), is to match the gradients of $u(y, z)$ and u-convex functions $f(y)$, whenever $z \in \partial^u f(y)$. The approach was also developed in Gangbo and McCann [36] and for applications in Economics by Carlier [15]. We refer to [15] (Lemma 4) for proofs of Lemma 1.2.9 and Proposition 1.2.11.

Lemma 1.2.9. *Suppose f is u-subdifferentiable, and u satisfies Assumption 1.2.6. Then $\partial^u f$ is monotone in the following sense:*
Let $y, \hat{y} \in D_y$, $\hat{y} > y$. Suppose $\hat{z} \in \partial^u f(\hat{y})$ and $z \in \partial^u f(y)$. Then $\hat{z} \geq z$.

Definition 1.2.10. We say that a function is strictly u-convex, when its u-subdifferential is strictly monotone.

Proposition 1.2.11. *Suppose that u satisfies Assumption 1.2.6.*

Suppose f is a.e differentiable and u-subdifferentiable. Then there exists a map $z^* : D_y \to D_z$ such that if f is differentiable at y then $f(y) = u(y, z^*(y)) - f^u(z^*(y))$ and

$$f'(y) = u_y(y, z^*(y)). \tag{1.2.1}$$

Moreover, z^* is such that $z^*(y)$ is non-decreasing.

Conversely, suppose that f is a.e differentiable and equal to the integral of its derivative. If (1.2.1) holds for a non-decreasing function $z^*(y)$, then f is u-convex and u-subdifferentiable with $f(y) = u(y, z^*(y)) - f^u(z^*(y))$.

Note that the subdifferential $\partial^u f(y)$ may be an interval in which case $z^*(y)$ may be taken to be any element in that interval. Under Assumption 1.2.6, $z^*(y)$ is non-decreasing

We observe that since $u(y, z^*(y)) = f(y) + f^u(z^*(y))$ we have $u(y^*(z), z) = f(y^*(z)) + f(z)$ and $y^*(z) \in \partial^u f^u(z)$ so that y^* may be defined directly as an element of $\partial^u f^u$. If z^* is strictly increasing then y^* is just the inverse of z^*.

Remark 1.2.12. If u satisfies the 'Reverse Spence-Mirrlees' condition, the conclusion of Lemma 1.2.9 is unchanged except that now '$z \geq \hat{z}$'. Similarly, Proposition 1.2.11 remains true, except that $z^*(y)$ and $y^*(z)$ are non-increasing.

Proposition 1.2.13. *Suppose that u satisfies Assumption 1.2.6.*

Suppose f is u-subdifferentiable in a neighbourhood of y. Then f is continuously differentiable at y if and only if z^* is continuous at y.

Proof. Suppose f is u-subdifferentiable in a neighbourhood of y. Then for small enough ϵ,

$$f(y+\epsilon) - f(y) \geq u(y+\epsilon, z^*(y)) - u(y, z^*(y))$$

and $\lim_{\epsilon\downarrow 0}\{f(y+\epsilon) - f(y)\}/\epsilon \geq u_y(y, z^*(y))$.

For the reverse inequality, if z^* is continuous at y then for ϵ small enough so that $z^*(y+\epsilon) < z^*(y) + \delta$ we have

$$f(y+\epsilon) - f(y) \leq u(y+\epsilon, z^*(y+\epsilon)) - u(y, z^*(y+\epsilon)) \leq u(y+\epsilon, z^*(y)+\delta) - u(y, z^*(y)+\delta)$$

and $\lim_{\epsilon\downarrow 0}\{f(y+\epsilon) - f(y)\}/\epsilon \leq \lim_{\delta\downarrow 0} u_y(y, z^*(y)+\delta) = u_y(y, z^*(y))$.

Inequalities for the left-derivative follow similarly, and then $f'(y) = u_y(y, z^*(y))$ which is continuous.

Conversely, if $\partial^u f$ is multi-valued at y so that z^* is discontinuous at y, then

$$\lim_{\epsilon\downarrow 0}\{f(y+\epsilon) - f(y)\}/\epsilon \geq u_y(y, z^*(y)+) > u_y(y, z^*(y)-) \geq \lim_{\epsilon\downarrow 0}\{f(y) - f(y-\epsilon)\}/\epsilon$$

where the strict middle inequality follows immediately from Assumption 1.2.6.

□

1.3 Application of u-convex analysis to the Forward Problems

Now we return to the context of the family of optimal control problems (1.1.1) and the representation (1.1.7).

Lemma 1.3.1. *Let $X \in \mathcal{X}_0$ be a diffusion in natural scale reflected at the origin with a finite or infinite right boundary point ξ. Then the increasing log-eigenfunction of the generator*

$$\psi_X(x) = -\log(\mathbb{E}[e^{-\rho H_x}]^{-1})$$

is locally Lipschitz continuous on $(0, \xi)$.

Proof. $\varphi_X(x)$ is increasing, convex and finite and therefore locally Lipschitz on $(0, \xi)$. $\varphi(0) = 1$, and since log is locally Lipschitz on $[1, \infty)$, $\psi = \log(\varphi)$ is locally Lipschitz on $(0, \xi)$. □

Henceforth we assume that g satisfies Assumption 1.2.6, so that g is twice differentiable and satisfies the Spence-Mirrlees condition. We assume further that $G(x, \theta)$ is non-decreasing in x. Note that this is without loss of generality since it can never be optimal to stop at $x' > x$

if $G(x',\theta) < G(x,\theta)$, since to wait until the first hitting time of x' involves greater discounting and a lower payoff.

Consider the forward problem. Suppose the aim is to solve (1.1.7) for a given $X \in \mathcal{X}_0$ with associated log-eigenfunction $\psi(x) = \psi_X(x) = -\log \mathbb{E}_0[e^{-\rho H_x}]$ for the family of objective functions $\{G(x,\theta) : \theta \in \Theta\}$. Here Θ is assumed to be an interval with endpoints θ_- and θ_+, such that $\Theta \subseteq D_\theta$.

Now let

$$v(\theta) = \sup_{x:\psi(x)<\infty} [g(x,\theta) - \psi(x).] \tag{1.3.1}$$

Then $v = \psi^g$ is the g-convex dual of ψ.

By definition $\partial^g v(\theta) = \{x : v(\theta) = g(x,\theta) - \psi(x)\}$ is the (set of) level(s) at which it is optimal to stop for the problem parameterised by θ. If $\partial^g v(\theta)$ is empty then there is no optimal stopping strategy in the sense that for any finite stopping rule there is another which involves waiting longer and gives a higher problem value.

Let θ_R be the infimum of those values of $\theta \in \Theta$ such that

$\partial^g v(\theta) = \emptyset$. If v is nowhere g-subdifferentiable then we set $\theta_R = \theta_-$.

Lemma 1.3.2. *The set where v is g-subdifferentiable forms an interval with endpoints θ_- and θ_R.*

Proof. Suppose v is g-subdifferentiable at $\hat{\theta}$, and suppose $\theta \in (\theta_-, \hat{\theta})$. We claim that v is g-subdifferentiable at θ.

Fix $\hat{x} \in \partial^g v(\hat{\theta})$. Then $v(\hat{\theta}) = g(\hat{x},\hat{\theta}) - \psi(\hat{x})$ and

$$g(\hat{x},\hat{\theta}) - \psi(\hat{x}) \geq g(x,\hat{\theta}) - \psi(x), \qquad \forall x < \xi, \tag{1.3.2}$$

and for $x = \xi$ if $\xi < \infty$. We write the remainder of the proof as if we are in the case $\xi < \infty$; the case $\xi = \infty$ involves replacing $x \leq \xi$ with $x < \xi$.

Fix $\theta < \hat{\theta}$. We want to show

$$g(\hat{x},\theta) - \psi(\hat{x}) \geq g(x,\theta) - \psi(x), \qquad \forall x \in (\hat{x},\xi], \tag{1.3.3}$$

for then

$$\sup_{x\leq\xi}\{g(x,\theta) - \psi(x)\} = \sup_{x\leq\hat{x}}\{g(x,\theta) - \psi(x)\},$$

and since $g(x,\theta) - \psi(x)$ is continuous in x the supremum is attained.

By assumption, $g_\theta(x,t)$ is increasing in x, and so for $x \in (\hat{x}, \xi]$

$$\int_\theta^{\hat{\theta}} [g_\theta(\hat{x},t) - g_\theta(x,t)]dt \le 0$$

or equivalently,

$$g(\hat{x},\hat{\theta}) - g(\hat{x},\theta) \le g(x,\hat{\theta}) - g(x,\theta). \tag{1.3.4}$$

Subtracting (1.3.4) from (1.3.2) gives (1.3.3). □

Lemma 1.3.3. *v is locally Lipschitz on (θ_-, θ_R).*

Proof. On (θ_-, θ_R) $v(\theta)$ is g-convex, g-subdifferentiable and $x^*(\theta)$ is monotone increasing.

Fix θ', θ'' such that $\theta_- < \theta'' < \theta' < \theta_R$. Choose $x' \in \partial^g v(\theta')$ and $x'' \in \partial^g v(\theta'')$ and suppose g has Lipschitz constant K' (with respect to θ) in a neighbourhood of (x', θ').

Then $v(\theta') = g(x',\theta') - \psi(x')$ and $v(\theta'') \ge g(x',\theta'') - \psi(x')$ so that

$$v(\theta') - v(\theta'') \le g(x',\theta') - g(x',\theta'') \le K'(\theta' - \theta'')$$

and a reverse inequality follows from considering $v(\theta'') = g(x'',\theta'') - \psi(x'')$. □

Note that it is not possible under our assumptions to date (g satisfying Assumption 1.2.6, and g monotonic in x) to conclude that v is continuous at θ_-, or even that $v(\theta_-)$ exists. Monotonicity guarantees that even if $\theta_- \notin \Theta$ we can still define $x^*(\theta_-) := \lim_{\theta \downarrow \theta_-} x^*(\theta)$. For example, suppose $\Theta = (0, \infty)$ and for $\epsilon \in (0,1)$ let $g_\epsilon(x,\theta) = g(x,\theta) + \epsilon f(\theta)$. Then if $v_\epsilon(\theta)$ is the g_ϵ-convex dual of ψ we have $v_\epsilon(\theta) = v(\theta) + \epsilon f(\theta)$, where $v(\theta) = v_0(\theta)$. If g and ψ are such that $\lim_{\theta \downarrow 0} v(\theta)$ exists and is finite, then choosing any bounded f for which $\lim_{\theta \downarrow 0} f(\theta)$ does not exist gives an example for which $\lim_{\theta \downarrow 0} v_\epsilon(\theta)$ does not exist. It is even easier to construct modified examples such that $v(\theta_-)$ is infinite.

Denote $\Sigma(\theta, \xi) = \limsup_{x \uparrow \xi} \{g(x,\theta) - \psi(x)\}$. Then for $\theta_R < \theta < \theta_+$, $\psi^g(\theta) = \Sigma(\theta, \xi)$. We have shown:

Proposition 1.3.4. *If g satisfies Assumption 1.2.6, g is increasing in x and if X is a reflecting diffusion in natural scale then the solution to the forward problem is $V(\theta) = \exp(\psi^g(\theta))$.*

Remark 1.3.5. Suppose now that $g_x(x,.)$ is strictly decreasing (the reverse Spence-Mirrlees condition). The arguments above apply with the obvious modifications. Let θ_L be the supremum of those values $\theta \in \Theta$ such that $x^*(\theta) = \emptyset$. Then the analogues to Lemmas 1.3.2 and 1.3.3 show that v is g-subdifferentiable and locally Lipschitz on (θ_L, θ_+) and that for $\theta_- < \theta < \theta_L$

$$V(\theta) = \exp(\Sigma(\theta, \xi)).$$

We close this section with some examples.

Example 1.3.6. *Recall Example 1.1.5, but note that in that example θ was restricted to take values in $\Theta = (1,2)$. Suppose $\Theta = [0,\infty)$, $g(x,\theta) = \theta \log x$ and $\psi(x) = \log(1+x^2)$. Then $\theta_R = 2$ and for $\theta < \theta_R$, $x^*(\theta) = (\theta/(2-\theta))^{1/2}$. Further, for $\theta \le 2$*

$$v(\theta) = \frac{\theta}{2}\log(\theta) + \frac{2-\theta}{2}\log(2-\theta) - \log 2,$$

and $v(\theta) = \infty$ for $\theta > 2$.

Note that v is continuous on $[0,\theta_R]$, but not on Θ.

Example 1.3.7. *Suppose $g(x,\theta) = x\theta$ and $\Theta = (0,\infty)$. Suppose X is a diffusion on $[0,1)$, with 1 a natural boundary and diffusion coefficient $\sigma(x)^2 = \frac{\rho(1-x^2)^2}{1+x^2}$. Then $\varphi(x) = \frac{1}{1-x^2}$ and*

$$v(\theta) = \sup_{x<1}[\theta x + \log(1-x^2)].$$

It is straightforward to calculate that $x^*(\theta) = \sqrt{1+\theta^{-2}} - 1/\theta$ and then that $v(\theta) : (0,\infty) \to \mathbb{R}$ is given by

$$v(\theta) = \sqrt{1+\theta^2} - 1 - \log\left(\frac{\theta^2}{2(\sqrt{1+\theta^2}-1)}\right). \tag{1.3.5}$$

1.4 Application of u-convex analysis to the Inverse Problem

Given an interval $\Theta \subseteq \overline{\mathbb{R}}$ with endpoints θ_- and θ_+ and a value function V defined on Θ we now discuss how to determine whether or not there exists a diffusion in $\mathcal{X}_0$ that solves the inverse problem for V. Theorem 1.4.1 gives a necessary and sufficient condition for existence. This condition is rather indirect, so in Theorem 1.4.4 we give some sufficient conditions in terms of the g-convex dual v^g and associated objects.

Then, given existence, a supplementary question is whether $\{V(\theta) : \theta \in \Theta\}$ contains enough information to determine the diffusion uniquely. In Sections 1.4.3, 1.4.4 and 1.4.5 we consider three different phenomena which lead to non-uniqueness. Finally in Section 1.4.6 we give a simple sufficient condition for uniqueness.

Two key quantities in this section are the lower and upper bound for the range of the g-subdifferential of v on Θ. Recall that we are assuming that the Spence-Mirrlees condition holds so that x^* is increasing on Θ. Then, if v is somewhere g-subdifferentiable we set $x_- = \sup\{x \in \partial^g v(\theta_-)\}$, or if $\theta_- \notin \Theta$, $x_- = \lim_{\theta \downarrow \theta_-} x^*(\theta)$. Similarly, we define $x_+ = \inf\{x \in \partial^g v(\theta_+)\}$, or if $\theta_+ \notin \Theta$, $x_+ = \lim_{\theta \uparrow \theta_+} x^*(\theta)$, and $x_R = \lim_{\theta \uparrow \theta_R} x^*(\theta)$. If v is nowhere g-subdifferentiable then we set $x_- = x_R = x_+ = \infty$.

1.4.1 Existence

In the following we assume that v is g-convex on $\mathbb{R}_+ \times \Theta$, which means that for all $\theta \in \Theta$,

$$v(\theta) = v^{gg}(\theta) = \sup_{x \geq 0} \{g(x, \theta) - v^g(x)\}.$$

Trivially this is a necessary condition for the existence of a diffusion such that the solution of the optimal stopping problems are given by V. Recall that we are also assuming that g is increasing in x and that it satisfies Assumption 1.2.6.

The following fundamental theorem provides necessary and sufficient conditions for existence of a consistent diffusion.

Theorem 1.4.1. *There exists $X \in \mathcal{X}_0$ such that $V_X = V$ if and only if there exists $\phi : [0, \infty) \to [1, \infty]$ such that $\phi(0) = 1$, ϕ is increasing and convex and ϕ is such that $(\log \phi)^g = v$ on Θ.*

Proof. If $X \in \mathcal{X}_0$ then $\phi_X(0) = 1$ and ϕ_X is increasing and convex. Set $\psi_X = \log \phi_X$. If $V_X = V$ then

$$v(\theta) = v_X(\theta) = \sup_x \{g(x, \theta) - \psi_X(x)\} = \psi_X^g.$$

Conversely, suppose ϕ satisfies the conditions of the theorem, and set $\psi = \log \phi$. Let $\xi = \sup\{x : \phi(x) < \infty\}$. Note that if $\xi < \infty$ then

$$(\log \phi)^g(\theta) = \sup_{x \geq 0} \{g(x, \theta) - \psi(x)\} = \sup_{x \leq \xi} \{g(x, \theta) - \psi(x)\}$$

and the maximiser $x^*(\theta)$ satisfies $x^*(\theta) \leq \xi$.

For $0 \leq x \leq \xi$ define a measure m via

$$m(dx) = \frac{1}{2\rho} \frac{\phi''(x)}{\phi(x)} dx = \frac{\psi''(x) + (\psi'(x))^2}{2\rho} dx. \tag{1.4.1}$$

Let $m(dx) = 0$ for $x < 0$, and, if ξ is finite $m(dx) = \infty$ for $x > \xi$. We interpret (1.4.1) in a distributional sense whenever ϕ has a discontinuous derivative. In the language of strings ξ is the length of the string with mass distribution m. We assume that $\xi > 0$. The case $\xi = 0$ is a degenerate case which can be covered by a direct argument.

Let B be a Brownian motion started at 0 with local time process L_u^z and define $(\Gamma_u)_{u \geq 0}$ via

$$\Gamma_u = \int_{\mathbb{R}} m(dz) L_u^z = \int_0^t \frac{1}{2\rho} \left(\psi''(B_s) + (\psi'(B_s))^2\right) ds.$$

Let A be the right-continuous inverse to Γ. Now set $X_t = B_{A_t}$. Then X is a local martingale (whilst away from zero) such that $d\langle X \rangle_t / dt = dA_t/dt = (dm/dx|_{x=X_t})^{-1}$. When $m(dx) =$

$\sigma(x)^{-2}dx$, we have $d\langle X\rangle_t = \sigma(X_t)^2 dt$.

We want to conclude that $\mathbb{E}[e^{-\rho H_x}] = \exp(-\psi(x))$. Now, $\varphi_X(x) = (\mathbb{E}[e^{-\rho H_x}])^{-1}$ is the unique increasing solution to

$$\frac{1}{2}\frac{d^2 f}{dm dx} = \rho f$$

with the boundary conditions $f'(0-) = 0$ and $f(0) = 1$. Equivalently, for all $x, y \in (0, \xi)$ with $x < y$, φ_X solves

$$f'(y-) - f'(x-) = \int_{[x,y)} 2\rho f(z) m(dz).$$

By the definition of m above it is easily verified that $\exp(\psi(x))$ is a solution to this equation. Hence $\phi = \varphi_X$ and our candidate process solves the inverse problem. □

Remark 1.4.2. Since v is g-convex a natural candidate for ϕ is $e^{v^g(x)}$, at least if $v^g(0) = 0$ and e^{v^g} is convex. Then ϕ is the eigenfunction φ_X of a diffusion $X \in \mathcal{X}_0$.

Our next example is one where $\phi(x) = e^{v^g(x)}$ is convex but not twice differentiable, and in consequence the consistent diffusion has a sticky point. This illustrates the need to work with generalised diffusions. For related examples in a different context see Ekström and Hobson [30].

Example 1.4.3. *Let $\Theta = \mathbb{R}_+$ and let the objective function be $g(x, \theta) = \exp(\theta x)$. Suppose*

$$V(\theta) = \begin{cases} \exp(\frac{1}{4}\theta^2) & 0 \le \theta \le 2, \\ \exp(\theta - 1) & 2 < \theta \le 3, \\ \exp(\frac{2}{3\sqrt{3}}\theta^{3/2}) & 3 < \theta. \end{cases}$$

Writing $\varphi = e^{v^g}$ we calculate

$$\varphi(x) = \begin{cases} \exp(x^2) & 0 \le x \le 1, \\ \exp(x^3) & 1 < x. \end{cases}$$

Note that φ is increasing and convex, and $\varphi(0) = 1$. Then φ' jumps at 1 and since

$$\varphi(1) = \varphi'(1+) - \varphi'(1-) = 2\rho\varphi(1)m(\{1\})$$

we conclude that $m(\{1\}) = \frac{1}{2\rho}$. Then Γ_u includes a multiple of the local time at 1 and the diffusion X is sticky there.

Theorem 1.4.1 converts a question about existence of a consistent diffusion into a question about existence of a log-eigenfunction with particular properties including $(\log \phi)^g = v$. We would like to have conditions which apply more directly to the value function $V(\cdot)$. The conditions we derive depend on the value of x_-.

As stated in Remark 1.4.2, a natural candidate for ϕ is $e^{v^g(x)}$. As we prove below, if $x_- = 0$ this candidate leads to a consistent diffusion provided $v^g(0) = 0$ and $e^{v^g(x)}$ is convex and strictly increasing. If $x_- > 0$ then the sufficient conditions are slightly different, and e^{v^g} need not be globally convex.

Theorem 1.4.4. *Assume v is g-convex. Each of the following is a sufficient condition for there to exist a consistent diffusion:*

1. *$x_- = 0$, $v^g(0) = 0$ and $e^{v^g(x)}$ is convex and increasing on $[0, x_+)$.*
2. *$0 < x_- < \infty$, $v^g(x_-) > 0$, $e^{v^g(x)}$ is convex and increasing on $[x_-, x_+)$, and on $[0, x_-)$, $v^g(x) \leq f(x) = \log(F(x))$ where*

$$F(x) = 1 + x\frac{\exp(v^g(x_-)) - 1}{x_-}$$

is the straight line connecting the points $(0, 1)$ and $(x_-, e^{v^g(x_-)})$.

3. *$x_- = \infty$ and there exists a convex, increasing function F with $\log(F(0)) = 0$ such that $f(x) \geq v^g(x)$ for all $x \geq 0$ and*

$$\lim_{x\to\infty}\{f(x) - v^g(x)\} = 0,$$

where $f = \log F$.

Proof. We treat each of the conditions in turn. If $x_- = 0$ then Theorem 1.4.1 applies directly on taking $\phi(x) = e^{v^g(x)}$, with $\phi(x) = \infty$ for $x > x_+$ (we use the fact that v is g-convex and so $v^{gg} = v$).

Suppose $0 < x_- < \infty$. The condition $e^{v^g(x)} \leq F(x)$ on $[0, x_-)$ implies $F'(x) = (e^{v^g(x_-)} - 1)/x_- \leq (e^{v^g(x_- -)})'$. Although the left-derivative $v^g(x-)'$ need not equal the right-derivative $v^g(x+)'$, by Proposition 1.2.11 $v^g(x-)' \leq v^g(x+)'$. This implies that the function

$$\phi_F(x) = \begin{cases} F(x) & x < x_- \\ \exp(v^g(x)) & x_- \leq x < x_+ \end{cases}$$

is convex at x_- and hence convex and increasing on $[0, x_+)$.

Setting $\phi_F(x_+) = \lim_{x\uparrow x_+}\phi_F(x)$ and $\phi_F = \infty$ for $x > x_+$ we have a candidate for the function in Theorem 1.4.1.

It remains to show that $(\log\phi_F)^g = v$ on Θ. We now check that ϕ_F is consistent with V on Θ, which follows if the g-convex dual of $\psi = \log(\phi_F)$ is equal to v on Θ.

Since $\psi \geq v^g$ we have $\psi^g \leq v$. We aim to prove the reverse inequality. By definition, we have for $\theta \in \Theta$

$$\psi^g(\theta) = \left(\sup_{x<x_-} \{g(x,\theta) - f(x)\} \right) \vee \left(\sup_{x_- \leq x \leq x_+} \{g(x,\theta) - v^g(x)\} \right). \tag{1.4.2}$$

Now fix $x \in [0, x_-)$. For $\theta < \theta_R$ we have by the definition of the g-subdifferential

$$g(x^*(\theta), \theta) - v^g(x^*(\theta)) \geq g(x,\theta) - v^g(x).$$

Hence $v(\theta) = \sup_{x\geq 0}\{g(x,\theta) - v^g(x)\} = \sup_{x\geq x_-}\{g(x,\theta) - v^g(x)\} \leq \psi^g(\theta)$.

Similarly, if $\theta \geq \theta_R$ we have for all $x' \in [0, x_-)$,

$$\limsup_{x\to\infty} g(x,\theta) - v^g(x) \geq g(x',\theta) - v^g(x').$$

and $v(\theta) = \limsup_x\{g(x,\theta) - v^g(x)\} = \sup_{x\geq x_-}\{g(x,\theta) - v^g(x)\} \leq \psi^g(\theta)$.

Finally, suppose $x_- = \infty$. By the definition of f^g and the condition $f \geq v^g$ we get

$$\begin{aligned} f^g(\theta) &= \sup_{x\geq 0}\{g(x,\theta) - f(x)\} \\ &\leq \sup_{x\geq 0}\{g(x,\theta) - v^g(x)\} \\ &= v(\theta). \end{aligned}$$

On the other hand

$$\begin{aligned} v(\theta) &= \limsup_{x\to\infty}\{g(x,\theta) - f(x) + f(x) - v^g(x)\} \\ &\leq \limsup_{x\to\infty}\{g(x,\theta) - f(x)\} + \lim_{x\to\infty}\{f(x) - v^g(x)\} \leq f^g(\theta). \end{aligned}$$

Hence $v(\theta) = f^g(\theta)$ on Θ. □

Remark 1.4.5. Case 1 of the Theorem gives the sufficient condition mentioned in the paragraph headed Inverse Problem in Section 1.1. If $\theta_- \in \Theta$ then $x_- = 0$ if and only if for all $x > 0$, $g(x,\theta_-) - v^g(x) < g(0,\theta_-)$, where we use the fact that, by supposition, $v^g(0) = 0$.

1.4.2 Non-Uniqueness

Given existence of a diffusion X which is consistent with the values $V(\theta)$, the aim of the next few sections is to determine whether such a diffusion is unique.

Fundamentally, there are two natural ways in which uniqueness may fail. Firstly, the domain Θ may be too small (in extreme cases Θ might contain a single element). Roughly speaking the g-convex duality is only sufficient to determine v^g (and hence the candidate ϕ) over (x_-, x_+) and there can be many different convex extensions of ϕ to the real line, for each of which $\psi^g = v$. Secondly, even when $x_- = 0$ and $x_+ = \infty$, if $x^*(\theta)$ is discontinuous then there can be circumstances in which there are a multitude of convex functions ϕ with $(\log \phi)^g = v$. In that case, if there are no θ for which it is optimal to stop in an interval I, then it is only possible to infer a limited amount about the speed measure of the diffusion over that interval.

In the following lemma we do not assume that ψ is g-convex.

Lemma 1.4.6. *Suppose v is g-convex and $\psi^g = v$ on Θ. Let $A(\theta) = \{x : g(x,\theta) - \psi(x) = \psi^g(\theta)\}$. Then, for each θ, $A(\theta) \subseteq \partial^g\psi^g(\theta) \equiv \partial^g v(\theta)$, and for $x \in A(\theta)$, $\psi(x) = \psi^{gg}(x) = v^g(x)$. Further, for $\theta \in (\theta_-, \theta_R)$ we have $A(\theta) \neq \emptyset$.*

Proof. Note that if ψ is any function, with $\psi^g = v$ then $\psi \geq \psi^{gg} = v^g$.

If $\hat{x} \in A(\theta)$ then

$$\psi^g(\theta) = g(\hat{x}, \theta) - \psi(\hat{x}) \leq g(\hat{x}, \theta) - v^g(\hat{x}) \leq v(\theta).$$

Hence there is equality throughout, so $\hat{x} \in \partial^g v(\theta)$ and $\psi(\hat{x}) = v^g(\hat{x}) = \psi^{gg}(\hat{x})$.

For the final part, suppose $\theta < \theta_R$ and fix $\tilde{\theta} \in (\theta, \theta_R)$. From the Spence-Mirrlees condition, if $x > \tilde{x} := x^*(\tilde{\theta})$,

$$g(x, \theta) - g(\tilde{x}, \theta) < g(x, \tilde{\theta}) - g(\tilde{x}, \tilde{\theta}),$$

and hence

$$\{g(x,\theta) - \psi(x)\} - \{g(\tilde{x},\theta) - \psi(\tilde{x})\} < \{g(x,\tilde{\theta}) - \psi(x)\} - \{g(\tilde{x},\tilde{\theta}) - \psi(\tilde{x})\} \leq 0.$$

In particular, for $x > \tilde{x}$, $g(x,\theta) - \psi(x) < g(\tilde{x},\theta) - \psi(\tilde{x})$ and

$$\sup_{x \geq 0} g(x,\theta) - \psi(x) = \sup_{0 \leq x \leq \tilde{x}} g(x,\theta) - \psi(x).$$

This last supremum is attained so that $A(\theta)$ is non-empty.

□

1.4.3 Left extensions

In the case where $x_- > 0$ and there exists a diffusion consistent with V then it is generally possible to construct many diffusions consistent with V. Typically V contains insufficient information to characterise the behaviour of the diffusion near zero.

Suppose that $0 < x_- < \infty$. Recall the definition of the straight line F from Theorem 1.4.4.

Lemma 1.4.7. *Suppose that $0 < x_- < \infty$ and that there exists $X \in \mathcal{X}_0$ consistent with V.*

Suppose that $\theta_R > \theta_-$ and that v^g and is continuous and differentiable to the right at x_-. Suppose further that $x^*(\theta) > x_-$ for each $\theta > \theta_-$.

Then, unless either $v^g(x) = f(x)$ for some $x \in [0, x_-)$ or $(v^g)'(x_-+) = f'(x_-)$, there are many diffusions consistent with V.

Proof. Let ϕ be the log-eigenfunction of a diffusion $X \in \mathcal{X}_0$ which is consistent with V

If $\theta_- \in \Theta$ then $v^g(x_-) = \psi(x_-)$ by Lemma 1.4.6. Otherwise the same conclusion holds on taking limits, since the convexity of ϕ necessitates continuity of ψ.

Moreover, taking a sequence $\theta_n \downarrow \theta_-$, and using $\hat{x}(\theta_n) > x^*(\theta_n-) > x_-$ we have

$$\psi'(x_-+) = \lim_{\theta_n \downarrow \theta_-} \frac{\psi(\hat{x}(\theta_n)) - \psi(x_-)}{\hat{x}(\theta_n) - x_-} = \lim_{\theta_n \downarrow \theta_-} \frac{v^g(\hat{x}(\theta_n)) - v^g(x_-)}{\hat{x}(\theta_n) - x_-} = (v^g)'(x_-+)$$

In particular, the conditions on v^g translate directly into conditions about ϕ.

Since the straight line F is the largest convex function with $F(0) = 1$ and $F(x_-) = e^{v^g(x_-)}$ we must have $\phi \leq F$.

Then if $\phi(x) = F(x)$ for some $x \in (0, x_-)$ or $\phi'(x_-+) = F'(x_-)$, then convexity of ϕ guarantees $\phi = F$ on $[0, x_-]$.

Otherwise there is a family of convex, increasing $\tilde{\phi}$ with $\tilde{\phi}(0) = 1$ and such that $v^g(x) \leq \log\tilde{\phi}(x) \leq F(x)$ for $x < x_-$ and $\tilde{\phi}(x) = \phi(x)$ for $x \geq x_-$.

For such a $\tilde{\phi}$, then by the arguments of Case 2 of Theorem 1.4.1 we have $(\log\phi_F)^g = v$ and then $v^g \leq \log\tilde{\phi} \leq \phi_F$ implies $v \geq (\log\tilde{\phi})^g \geq (\log\phi_F)^g = v$.

Hence each of $\tilde{\phi}$ is the eigenfunction of a diffusion which is consistent with V. □

Example 1.4.8. *Recall Example 1.1.5, in which we have $x_- = 1$, $\varphi'(1) = 2$ and $\varphi(1) = 2$. We can extend φ to $x \in [0, 1)$ by (for example) drawing the straight line between $(0, 1)$ and $(1, 2)$ (so that for $x \leq 1$, $\varphi(x) = 1 + x$). With this choice the resulting extended function will be convex, thus defining a consistent diffusion on $\mathbb{R}^+$. Note that any convex extension of φ (i.e. any function $\hat{\varphi}$ such that $\hat{\varphi}(0) = 1$ and $\hat{\varphi}'(0-) = 0$, $\hat{\varphi}(x) = \varphi(x)$ for $x > 1$) solves the inverse problem, (since necessarily $\hat{\varphi}(x) \geq 2x = e^{v^g(x)}$ on $(0, 1)$). The most natural choice is, perhaps, $\varphi(x) = 1 + x^2$ for $x \in (0, 1)$.*

Our next lemma covers the degenerate case where there is no optimal stopping rule, and for all θ it is never optimal to stop. Nevertheless, as Example 1.4.10 below shows, the theory of u-convexity as developed in this chapter still applies.

Lemma 1.4.9. *Suppose $x_- = \infty$, and that there exists a convex increasing function F with $F(0) = 1$ and such that $\log F(x) \geq v^g(x)$ and $\lim_{x\to\infty}\{\log F(x) - v^g(x)\} = 0$.*

Suppose that $\lim_{x\to\infty} e^{v^g(x)}/x$ exists in $(0,\infty]$ and write $\kappa = \lim_{x\to\infty} e^{v^g(x)}/x$. If $\kappa < \infty$ then X is the unique diffusion consistent with V if and only if $e^{v^g(x')} = 1 + \kappa x'$ for some $x' > 0$ or $\limsup_{x\uparrow\infty}(1 + \kappa x) - e^{v^g(x)} = 0$. If $\kappa = \infty$ then there exist uncountably many diffusions consistent with V.

Proof. The first case follows similar reasoning as Lemma 1.4.7 above. Note that $x \mapsto 1 + \kappa x$ is the largest convex function F on $[0,\infty)$ such that $F(0) = 1$ and $\lim_{x\to\infty} \frac{F(x)}{x} = \kappa$.

If $e^{v^g(x')} = 1+\kappa x'$ for any $x' > 0$, or if $\limsup_{x\uparrow\infty}(1+\kappa x) - e^{v^g(x)} = 0$ then there does not exist any convex function lying between $1+\kappa x$ and $e^{v^g(x)}$ on $[0,\infty)$. In particular $\phi(x) = 1+\kappa x$ is the unique eigenfunction consistent with V.

Conversely, if e^{v^g} lies strictly below the straight line $1+\kappa x$, and if $\limsup_{x\uparrow\infty}(1+\kappa x) - e^{v^g(x)} > 0$ then it is easy to verify that we can find other increasing convex functions with initial value 1, satisfying the same limit condition and lying between e^{v^g} and the line.

In the second case define $F_\alpha(x) = F(x) + \alpha x$ for $\alpha > 0$. Then since $\lim_{x\to\infty} e^{v^g(x)}/x = \infty$ we have

$$\lim_{x\to\infty} \frac{F_\alpha(x)}{e^{v^g(x)}} = \frac{F(x)}{e^{v^g(x)}} = 1$$

Hence F_α is the eigenfunction of another diffusion which is consistent with V. We conclude that there exist uncountably many consistent diffusions. $\square$

Example 1.4.10. *Suppose $g(x,\theta) = x^2 + \theta \tanh x$ and $v(\theta) = \theta$ on $\Theta = \mathbb{R}^+$. For this example we have that v is nowhere g-subdifferentiable and $x_- = \infty$. Then $v^g(x) = x^2$ and each of $\varphi(x) = e^{x^2}$,*

$$\bar{\varphi}(x) = \begin{cases} 1 + (e-1)x & 0 \leq x < 1 \\ e^{x^2} & 1 \leq x, \end{cases}$$

and $\varphi_\alpha(x) = \varphi(x) + \alpha x$ for any $\alpha \in \mathbb{R}_+$ is an eigenfunction consistent with V.

1.4.4 Right extensions

The case of $x_+ < \infty$ is very similar to the case $x_- > 0$, and typically if there exists one diffusion $X \in \mathcal{X}_0$ which is consistent with V, then there exist many such diffusions. Given X consistent with V, the idea is to produce modifications of the eigenfunction φ_X which agree with φ_X on $[0, x_+]$, but which are different on (x_+, ∞).

Lemma 1.4.11. *Suppose $x_+ < \infty$. Suppose there exists a diffusion $X \in \mathcal{X}_0$ such that V_X agrees with V on Θ. If $v^g(x_+) + (v^g)'(x_+ +) < \infty$ then there are infinitely many diffusions in*

$\mathcal{X}_0$ *which are consistent with* V.

Proof. It is sufficient to prove that given convex increasing $\hat{\phi}$ defined on $[0, x_+)$ with $\hat{\phi}(0) = 1$ and $(\log \hat{\phi})^g = v$ on Θ, then there are many increasing, convex ϕ with defined on $[0, \infty)$ with $\phi(0) = 1$ for which $(\log \phi)^g = v$.

The proof is similar to that of Lemma 1.4.7. □

Example 1.4.12. *Let* $G(x, \theta) = \theta x/(\theta + x)$, *and* $\Theta = (0, \infty)$.

Consider the forward problem when X *is a reflecting Brownian motion, so that the eigenfunction is given by* $\varphi(x) = \cosh(x\sqrt{2\rho})$. *Suppose* $\rho = 1/2$.

Then $\{g(x, \theta) - \log(\cosh x)\}$ *attains its maximum at the solution* $x = x^*(\theta)$ *to*

$$\theta = \frac{x^2 \tanh x}{1 - x \tanh x}. \tag{1.4.3}$$

It follows that $x_- = 0$ *but* $x_+ = \lim_{\theta\uparrow\infty} x^*(\theta) = \hat{\lambda}$ *where* $\hat{\lambda}$ *is the positive root of* $\mathcal{L}(\lambda) = 0$ *and* $\mathcal{L}(\lambda) = 1 - \lambda \tanh \lambda$.

Now consider an inverse problem. Let G *and* Θ *be as above, and suppose* $\rho = 1/2$. *Let* $x^*(\theta)$ *be the solution to (1.4.3) and let* $v(\theta) = g(x^*(\theta), \theta) - \log(\cosh x^*(\theta))$. *Then the diffusion with speed measure* $m(dx) = dx$ *(reflecting Brownian motion), is an element of* $\mathcal{X}_0$ *which is consistent with* $\{V(\theta) : \theta \in (0, \infty)\}$. *However, this solution of the inverse problem is not unique, and any convex function* φ *with* $\varphi(x) = \cosh x$ *for* $x \leq \hat{\lambda}$ *is the eigenfunction of a consistent diffusion. To see this note that for* $x > x_+$, $v^g(x) = \lim_{\theta\uparrow\infty}\{g(x, \theta) - v(\theta)\} = \log(x \cosh(x_+)/x_+)$ *so that any convex* φ *with* $\varphi(x) = \cosh x$ *for* $x \leq x_+$ *satisfies* $\varphi \geq e^{v^g}$.

Remark 1.4.13. If $x_+ + v^g(x_+) + (v^g)'(x_+ +) < \infty$ then one admissible choice is to take $\phi = \infty$. This was the implicit choice in the proof of Theorem 1.4.1.

Example 1.4.14. *The following example is 'dual' to Example 1.1.3.*

Suppose $\rho = 1/2$, $g(x, \theta) = \theta x$, $\Theta = (0, \infty)$ *and* $v(\theta) = \log(\cosh \theta)$. *Then* $v^g(x) = x \tanh^{-1}(x) + \frac{1}{2}\log(1 - x^2)$, *for* $x \leq 1$. *For* $x > 1$ *we have that* v^g *is infinite. Since* v *is convex, and* g*-duality is convex duality, we conclude that* v *is* g*-convex. Moreover,* v^g *is convex. Setting* $\psi = v^g$ *we have that* $\psi(0) = 0$, $\varphi = e^{\psi}$ *is convex and* $\psi^g = v^{gg} = v$. *Hence* ψ *is associated with a diffusion consistent with* V, *and this diffusion has an absorbing boundary at* $\xi \equiv 1$.

For this example we have $x_+ = 1$ *and* $v^g(x_+) = \log 2$, *but the left-derivative of* v^g *is infinite at* x_+ *and* v^g *is infinite to the right of* x_+. *Thus there is a unique diffusion in* $\mathcal{X}_0$ *which is consistent with* V.

1.4.5 Non-Uniqueness on $[x_-, x_+)$

Even if $[x_-, x_+)$ is the positive real line, then if $x^*(\theta)$ fails to be continuous it is possible that there are multiple diffusions consistent with V.

Lemma 1.4.15. *Suppose there exists a diffusion $X \in \mathcal{X}_0$ which is consistent with $\{V(\theta) : \theta \in \Theta\}$.*

Suppose the g-subdifferential of v is multivalued, or more generally that $x^*(\theta)$ is not continuous on Θ. Then there exists an interval $I \subset (x_-, x_+)$ where the g-subdifferential of $\psi = v^g$ is constant, so that $\theta^*(x) = \bar{\theta}$, $\forall x \in I$. If $G(x, \bar{\theta}) = e^{g(x,\bar{\theta})}$ is strictly convex in x on some subinterval of I_0 of I then the diffusion X is not the unique element of $\mathcal{X}_0$ which is consistent with V.

Proof. First note that if $x^*(\theta)$, is continuous then $\theta^* = x^{*-1}$ is nowhere constant and hence strictly monotone and thus $\psi = v^g$ is strictly g-convex (recall 1.2.10).

Suppose $\bar{G}(x) := G(x, \bar{\theta})$ is strictly convex on $I_0 \subseteq I$. Then we can choose $\hat{G}$ such that

- $\hat{G} = \bar{G}$ on I_0^c,
- $\hat{G}$ is linear on I_0,
- $\hat{G}$ is continuous.

Then $\hat{G}(x) \geq G(x, \bar{\theta})$.

By definition we have

$$\psi(x) = g(x, \bar{\theta}) - \psi^g(\bar{\theta}) \qquad x \in I.$$

Then $\varphi_X(x) = G(x, \bar{\theta})/V(\bar{\theta})$ on I.

Let $\hat{\varphi}$ be given by

$$\hat{\varphi}(x) = \begin{cases} \varphi_X(x) & \text{on } I_0^c\,, \\ \frac{\hat{G}(x)}{V(\bar{\theta})} & \text{on } I_0 \end{cases}$$

Then $\hat{\varphi}$ is convex and $\hat{\varphi} \geq \varphi$, so that they are associated with different elements of $\mathcal{X}_0$. Let $\hat{\psi} = \ln \hat{\varphi}$.

It remains to show that $\hat{v} := \hat{\psi}^g = \psi^g = v$. It is immediate from $\hat{\psi} \geq \psi$ that $\hat{\psi}^g \leq \psi^g$. For the converse, observe that

$$\begin{aligned}
v(\theta) &= \left(\sup_{x\in I_0}\{g(x,\theta)-\psi(x)\}\right)\vee\left(\sup_{x\in I_0^c}\{g(x,\theta)-\psi(x)\}\right)\\
&= \sup_{x\in I_0^c}\{g(x,\theta)-\psi(x)\}\\
&= \sup_{x\in I_0^c}\{g(x,\theta)-\hat{\psi}(x)\}\\
&\le \hat{v}(\theta).
\end{aligned}$$

□

Example 1.4.16. *Suppose $G(x,\theta)=e^{\theta x}$, and $\Theta=(0,\infty)$. Suppose that X is such that ψ is given by*

$$\psi(x)=\begin{cases}\frac{x^2}{4} & x<2\\ (x-1) & 2\le x<3\\ \frac{x^2}{6}+\frac{1}{2} & 3\le x\end{cases}$$

It follows that

$$v(\theta)=\begin{cases}\theta^2 & \theta\le 1\\ \frac{3\theta^2}{2}-\frac{1}{2} & 1<\theta\end{cases}$$

Then $\partial^g v(1)$ is multivalued, and there are a family of diffusions $\tilde{X}\in\mathcal{X}_0$ which give the same value functions as X.

In particular we can take

$$\hat{\psi}(x)=\begin{cases}\frac{x^2}{4} & x<2\\ \log((e^2-e)x+3e-2e^2) & 2\le x<3\\ \frac{x^2}{6}+\frac{1}{2} & 3\le x\end{cases}$$

Then $\hat{\psi}^g=v(\theta)$ and $\hat{\psi}$ is a log-eigenfunction.

1.4.6 Uniqueness of diffusions consistent with V

Proposition 1.4.17. *Suppose V is such that $x^*(\theta)$ is continuous on Θ, with range the positive real line.*

Then there exists at most one diffusion $X\in\mathcal{X}_0$ consistent with V.

Proof. The idea is to show that $v^g(x)$ is the only function with g-convex dual v. Suppose ψ is such that $\psi^g=v$ on Θ. For each x there is a θ with $x^*(\theta)=x$, and moreover $\partial^g v(\theta)=\{x\}$. Then by Lemma 1.4.6, $A(\theta)=\{x\}$ and $\psi(x)=v^g(x)$. □

Recall that we define $\theta_R = \sup_\theta \partial^g v(\theta) \neq \emptyset$ and if $\theta_R > \theta_-$ set $x_R = \lim_{\theta \uparrow \theta_R} x^*(\theta)$.

Theorem 1.4.18. *Suppose V is such that v is continuously differentiable on (θ_-, θ_R) and that $x_- = 0$ and $x_R = \infty$.*

Then there exists at most one diffusion $X \in \mathcal{X}_0$ consistent with V.

Proof. The condition on the range of $x^*(\theta)$ translates into the conditions on x_- and x_R, so it is sufficient to show that $x^*(\theta)$ is continuous at $\theta \in (\theta_-, \theta_R)$ if and only if v is differentiable there. This follows from Proposition 1.2.13. □

Corollary 1.4.19. *If the conditions of the Theorem hold but either $e^{v^g(x)}$ is not convex or $v^g(0) \neq 0$, then there is no diffusion $X \in \mathcal{X}_0$ which is consistent with $\{V(\theta), \theta \in \Theta\}$.*

Example 1.4.20. *Recall Example 1.1.6. For this example we have $x^*(\theta) = \theta^2 - 1$, which on $\Theta = (1, \infty)$ is continuous and strictly increasing. Then $e^{v^g(x)} = \sqrt{1+x}$ and by the above corollary there is no diffusion consistent with v.*

Remark 1.4.21. A more general but less succinct sufficient conditions for uniqueness can be deduced from Lemma 1.4.7 or Lemma 1.4.11. For example, if $0 < x_- < x_+ = \infty$, but $(v^g)'(x_-) = (1 - e^{-v^g(x_-)})/x_-$ then there is at most one $X \in \mathcal{X}_0$ which is consistent with V.

1.5 Further examples and remarks

1.5.1 Birth-Death processes

We now return to $\mathcal{X}_0$ and consider the case when E is made up of isolated points only; whence X is a birth-death process on points $x_n \in E$ indexed by $n \in \mathbf{N}_0$, with associated exponential holding times λ_n. We assume $x_0 = 0$, x_n is increasing, and write $x_\infty = \lim_n x_n$.

For a birth-death process the transition probabilities are given by

$$\mathbb{P}_{n,n+1}(t) = p_n \lambda_n t + o(t),$$
$$\mathbb{P}_{n,n-1}(t) = q_n \lambda_n t + o(t),$$

where of course $q_n = 1 - p_n$, with $p_0 = 1$. By our assumption that, away from zero, $(X_t)_{t \geq 0}$ is a martingale, we must have $p_n = \frac{x_{n+1} - x_n}{x_{n+1} - x_{n-1}}$. Then we can write $x_n = x_{n-1} + \frac{\prod_{i=1}^{n-1} q_i}{\prod_{i=0}^{n-1} p_i}$. Let

$$m(x_n) = \frac{1}{\lambda_n} \frac{p_0 p_1 \dots p_{n-1}}{q_1 q_2 \dots q_n}.$$

Then it is easy to verify, but see [33], that (1.1.4) can be expressed in terms of a second-order difference operator

$$\frac{1}{m(x_n)}\left[\frac{f(x_{n+1})-f(x_n)}{x_{n+1}-x_n}-\frac{f(x_n)-f(x_{n-1})}{x_n-x_{n-1}}\right]-\rho f(x_n)=0, \tag{1.5.1}$$

with boundary conditions $f(0)=1$ and $f'(0-)=0$.

Let $M(x)=\sum_{x_n<x} m(x)$. In the language of strings, the pair $(M,[0,x_\infty))$ is known as the Stieltjes String. If $x_\infty+M(x_\infty)<\infty$ the string is called regular and x_∞ is a regular boundary point, while otherwise the string is called singular, in which case we assume that x_∞ is natural (see Kac [46]).

In this section we consider the call option payoff, $G(x,\theta)=(x-\theta)^+$ defined for $\theta\in\Theta=[\theta_0,\infty)$. (Note that $g=\log(G)$ satisfies Assumption 1.2.6 on the set $\{(x,\theta);x>\theta\}$ which is all that we require since it is never optimal to stop for no reward.) The objective function G is straight-forward to analyse since the g-duality corresponds to straight lines in the original coordinates. It follows that for the forward problem V is decreasing and convex in θ. V is easily seen to be piecewise linear.

Our focus is on the inverse problem. Note that the solution of this problem involves finding the space E and the jump rates λ_n. Suppose that V is decreasing, convex and piecewise linear. Let $(\theta_n)_{n\in\mathbb{N}_0}$ be a sequence of increasing real valued parameters with $\theta_0<0$ and θ_n increasing to infinity, and suppose that V has negative slope s_i on each interval (θ_i,θ_{i+1}). Then s_i is increasing in i and

$$V(\theta)=V(\theta_n)+(\theta-\theta_n)s_n \qquad \text{for } \theta_n\le\theta<\theta_{n+1}. \tag{1.5.2}$$

We assume that $s_0=\frac{\theta_0}{V(\theta_0)}<0$.

Since V is convex, v is $\log((x-\theta)^+)$ convex. Let $\varphi(x)=\exp(v^g(x))$. By Proposition 1.2.11, for $\theta\in[\theta_n,\theta_{n+1})$

$$\frac{-1}{x^*(\theta)-\theta}=g_\theta(x^*(\theta),\theta)=\frac{s_n}{V(\theta_n)+(\theta-\theta_n)s_n}$$

so that $x_n:=x^*(\theta_n)=\theta_n-V(\theta_n)/s_n$. Note that $x^*(\theta)$ is constant on $[\theta_n,\theta_{n+1})$. We find that for $\theta\in[\theta_n,\theta_{n+1})$

$$\psi(x^*(\theta))=\log(\theta_n-\theta-V(\theta_n)/s_n)-v(\theta),$$

and hence $\varphi(x^*(\theta))=\frac{-1}{s_n}$. Then, for $x\in[x^*(\theta_n),x^*(\theta_{n+1}))$,

$$\psi(x)=\log(x-\theta_n)-v(\theta_n). \tag{1.5.3}$$

We proceed by determining the Q-matrix for the process on $[x^*(0) = 0, \xi)$. For each n, let p_n denote the probability of jumping to state x_{n+1} and q_n the probability of jumping to x_{n-1}. Then p_n and q_n are determined by the martingale property (and $p_0 = 1$). Further λ_n is determined either through (1.5.1) or from a standard recurrence relation for first hitting times of birth-death processes:

$$\lambda_n = \frac{\rho\varphi(x_n)}{p_n\varphi(x_{n+1}) + (1-p_n)\varphi(x_{n-1}) - \varphi(x_n)}, \qquad n \geq 1.$$

Example 1.5.1. *Suppose that* $\theta_n = n + 2^{-n} - 2$ *so that* $\theta_0 = -1$, *and* $V(\theta_n) = 2^{-n}$. *It follows that* $s_n = -(2^{n+1}-1)^{-1}$. *We find* $x_n = n$ *(this example has been crafted to ensure that the birth-death process has the integers as state space, and this is not a general result). Also* $\varphi(n) = 2^{n+1} - 1$ *(*φ *is piecewise linear with kinks at the integers) and the holding time at* x_n *is exponential with rate* $\lambda_n = 4\rho(1 - 2^{-(n+1)})$.

1.5.2 Subsets of $\mathcal{X}$ and uniqueness

So far in this chapter we have concentrated on the class $\mathcal{X}_0$. However, the methods and ideas translate to other classes of diffusions.

Let $\mathcal{X}^0_{m,s}$ denote the set of all diffusions reflected at 0. Here m denotes the speed measure, and s the scale function. With the boundary conditions as in (1.1.4), $\varphi(x) \equiv \varphi_X(x)$ is the increasing, but not necessarily convex solution to

$$\frac{1}{2}\frac{d^2 f}{dm ds} = \rho f. \tag{1.5.4}$$

In the smooth case, when m has a density $m(dx) = \nu(x)dx$ and s'' is continuous, (1.1.3) is equivalent to

$$\frac{1}{2}\sigma^2(x)f''(x) + \mu(x)f'(x) = \rho f(x), \tag{1.5.5}$$

where

$$\nu(x) = \sigma^{-2}(x)e^{M(x)}, \quad s'(x) = e^{-M(x)}, \quad M(x) = \int_{0-}^{x} 2\sigma^{-2}(z)\mu(z)dz,$$

see [12].

Now suppose $V \equiv \{V(\theta) : \theta \in \Theta\}$ is given such that $v^u(0) = 0$, $(v^u)'(0) = 0$ and v^u is increasing, then we will be able to find several pairs (σ, μ) such that $\exp(v^u)$ solves (1.5.5) so that there is a family of diffusions rather than a unique diffusion in $\mathcal{X}^0_{m,s}$ consistent with v.

It is only by considering subsets of $\mathcal{X}^0_{m,s}$, such as taking $s(x) = x$ as in the majority of this chapter, or perhaps by setting the diffusion co-efficient equal to unity, that we can hope to find a *unique* solution to the inverse problem.

Example 1.5.2. *Consider Example* 1.1.6 *where we found* $\psi(x) = \sqrt{1+x}$. *Let* $\mathcal{X}^0_{1,s}$ *be the set of diffusions with unit variance and scale function* s *(which are reflected at 0). Then there exists a unique diffusion in* $\mathcal{X}^0_{1,s}$ *consistent with* V. *The drift is given by*

$$\mu(x) = \frac{1/4 + 2\rho(1+x)^2}{1+x}.$$

1.6 Applications to Finance

1.6.1 Applications to Finance

Let $\mathcal{X}_{stock}$ be the set of diffusions with the representation

$$dX_t = (\rho - \delta)X_t dt + \eta(X_t)X_t dW_t.$$

In finance this SDE is often used to model a stock price process, with the interpretation that ρ is the interest rate, δ is the proportional dividend, and η is the level dependent volatility. Let x_0 denote the starting level of the diffusion and suppose that 0 is an absorbing barrier.

Our goal is to recover the underlying model, assumed to be an element of $\mathcal{X}_{stock}$, given a set of perpetual American option prices, parameterised by θ. The canonical example is when θ is the strike, and $G(x,\theta) = (\theta - x)^+$, and then, as discussed in Section 1.5.2, the fundamental ideas pass over from $\mathcal{X}_0$ to $\mathcal{X}_{stock}$. We suppose ρ and δ are given and aim to recover the volatility η.

Let φ be the convex and decreasing solution to the differential equation

$$\frac{1}{2}\eta(x)^2 x^2 f_{xx} + (\rho - \delta)x f_x - \rho f = 0. \tag{1.6.1}$$

(The fact that we now work with decreasing φ does not invalidate the method, though it is now appropriate to use payoffs G which are monotonic decreasing in x.) Then η is determined by the Black-Scholes equation

$$\eta(x)^2 = 2\frac{\rho\varphi(x) - (\rho - \delta)x\varphi'(x)}{x^2\varphi''(x)}. \tag{1.6.2}$$

Let $G \equiv G(x,\theta)$ be a family of payoff functions satisfying assumption 1.2.6. Under the additional assumption that G is decreasing in x (for example, the put payoff) Lemma 1.1.1 shows that the optimal stopping problem reduces to searching over first hitting times of levels $x < x_0$. Suppose that $\{V(\theta); \theta \in \Theta\}$ is used to determine a smooth, convex $\varphi = \exp(\psi)$ on $[0, \xi)$

via the g-convex duality

$$\psi(x) = v^g(x) = \sup_{\theta \in \Theta}[g(x,\theta) - v(\theta)].$$

Then the inverse problem is solved by the diffusion with volatility given by the solution of (1.6.2) above. Similarly, given a diffusion $X \in \mathcal{X}_{stock}$ such that $\psi = \log(\varphi)$ is g-convex on $[0,\xi)$, then the value function for the optimal stopping problem is given exactly as in Proposition 4.4. See Ekström and Hobson [30] for more details.

Chapter 2

Inverse perpetual-horizon stopping problems and allocation indices

The aim of this chapter is to extend the work on inverse stopping problems in Chapter 1 to a more general setup with running rewards and non-trivial scale functions. The two main aspects of our analysis are a theory of parameter dependence for forward problems and the use of so-called allocation (or Gittins) indices to parameterise solutions to inverse problems. Our main assumptions are that the underlying process is a generalized one-dimensional diffusion and that optimal stopping rules are threshold strategies. Threshold strategies are both a natural and a tractable class of optimal stopping times and focusing on them allows us to develop an approach to inverse problems based on the allocation index which we will interpret as a measure of investment preference.

Allocation indices are well known in the theory of multi-armed bandits and dynamic allocation problems (see for instance Whittle [76] and Karatzas [49]); here we give them an economic interpretation based on their role in parameterising solutions to inverse problems. Consider an investment yielding a running dividend and a taxed capital gain upon liquidation. Suppose that for each level of the underlying risky asset there is a critical capital gains tax below which we prefer to liquidate (and above which we remain invested), then we will call this critical rate an allocation index. The index may be seen as summarising our preferences with respect to receiving running dividends versus 'cashing-in' for a taxed terminal payoff. Now suppose the value of the investment as a function of the parameter (the tax rate) is known. For instance, we may be able to calculate the price at which we would be willing to sell our position to another investor as a function of the parameter. Then the question we seek to answer in this chapter is whether there exist consistent investment preferences (an allocation index) and a consistent diffusion model for the underlying asset price process. The example of deciding between capital gains and a running dividend in the presence of taxes will be used throughout to illustrate ideas in this chapter and is motivated by the work of Samuelson [71] on tax neutrality and the work of Klimmek [50] on taxation and risk preferences.

Along the way to solving inverse stopping problems we derive results about parameter dependence in forward problems. Techniques from monotone comparative statics which are well known in the mathematical economics literature (see for example Milgrom [54, 53] and Athey [5]) are shown to apply in our setting of optimal stopping. An 'envelope theorem' characterises the dependence of the value function on the parameter and a 'supermodularity' condition implies monotonicity of the threshold strategy in the parameter.

2.1 The forward and the inverse problems

Recall the construction of generalised one-dimensional diffusions in Chapter 1. Let $X = (X_t)_{t\geq 0}$ be a generalised diffusion process on an interval I, let ρ be a discount parameter. Let $G = \{G(x,\theta); \theta \in \Theta\}$ be a family of terminal reward functions and $c = \{c(x,\theta); \theta \in \Theta\}$ a family of running reward functions, both parameterised by a real parameter θ lying in an interval Θ with end-points θ_- and θ_+. The classical approach in optimal stopping problems is to fix the parameter, i.e. $\Theta = \{\theta\}$, and calculate

$$V(x) = \sup_{\tau} \mathbf{E}_x \left[\int_0^{\tau} e^{-\rho t} c(X_t,\theta)dt + e^{-\rho\tau} G(X_\tau,\theta)\right]$$

for $x \in \text{int}(I)$ using variational techniques, see for instance Bensoussan and Lions [9].

In contrast, we are interested in the case when the starting value is fixed and the parameter varies. Then the *forward problem* is to calculate $V \equiv \{V(\theta)\ ;\ \theta \in \Theta\}$ where

$$V(\theta) = \sup_{\tau} \mathbf{E}_{X_0} \left[\int_0^{\tau} e^{-\rho t} c(X_t,\theta)dt + e^{-\rho\tau} G(X_\tau,\theta)\right]. \tag{2.1.1}$$

We will assume that the process underlying the stopping problem is a regular one-dimensional diffusion processes characterised by a speed measure and a strictly increasing and continuous scale function (see Chapter 1). We will make the following assumption about the boundary behaviour of X.

Assumption 2.1.1.

Either the boundary of I is non-reflecting (absorbing or killing) or

X is started at a reflecting end-point and the other end-point is non-reflecting.

For a fixed diffusion with a fixed starting point we will scale φ and ϕ so that $\varphi(X_0) = \phi(X_0) = 1$. If m has a strictly positive density and s is differentiable then setting $Q(x) = -\log(s'(x))$, $\sigma^2(x) = \frac{e^{Q(x)}dx}{m(dx)}$ and $\mu(x) = \frac{2\sigma^2(x)Q(dx)}{dx}$, equation (1.1.3) is equivalent to

$$\frac{1}{2}\sigma^2(x)f''(x) + \mu(x)f'(x) = \rho f(x). \tag{2.1.2}$$

The boundary conditions of the differential equation (1.1.3) depend on whether the end-points of I are inaccessible, absorbing or reflecting, see Borodin and Salminen [12] for details. We will denote by $\text{int}(I)$ the interior of I and its accessible boundary points.

For $\theta \in \Theta$, let

$$R(x,\theta) = \mathbb{E}_x\left[\int_0^\infty e^{-\rho t}c(X_t,\theta)dt\right]. \tag{2.1.3}$$

Define $U : I\times\Theta \to \mathbb{R}$ by $U(x,\theta) = G(x,\theta) - R(x,\theta)$ and for all $\theta \in \Theta$ and $x \in I$ let $c^\theta(x) = c(x,\theta)$ and $R^\theta(x) = R(x,\theta)$.

Assumption 2.1.2. $\mathbb{E}_x\left[\int_0^\infty e^{-\rho s}|c^\theta(X_s)|ds\right] < \infty$ for all $x \in \text{int}(I)$ and $\theta \in \Theta$.

Under our assumptions it is well-known (see for instance Alvarez [4]) that $R^\theta : \text{int}(I) \to \mathbb{R}$ solves the differential equation

$$\frac{1}{2}\frac{d}{dm}\frac{d}{ds}f = \rho f - c^\theta. \tag{2.1.4}$$

Example 2.1.3. *In some cases R^θ can be calculated directly. Let $\mu < \rho$ and let $dX_t = \sigma X_t dB_t + \mu X_t dt$ and $c(x,\theta) = x\theta$. Then $\mathbb{E}_x\left[\int_0^\infty e^{-\rho t}X_t\theta dt\right]dt = x\theta\int_0^\infty e^{(\mu-\rho)t}dt = \frac{x\theta}{\rho-\mu}$.*

Example 2.1.4. *Suppose $m(dx) = 2x^2dx$ and $s(x) = -1/x$. Then X is known as the three-dimensional Bessel process and solves the SDE; $dX_t = dB_t + dt/X_t$. Let $c : \mathbb{R}^2 \to \mathbb{R}$ be defined $c(x,\theta) = \theta\cos(x)$ and $\rho = 1/2$. Then R^θ solves $\frac{1}{2}f''(x) + f'(x)/x - \frac{1}{2}f(x) = -\theta\cos(x)$ with $f(0) = 0$. The solution is $R^\theta(x) = \theta\left(\cos(x) - \frac{\sin(x)}{x}\right)$.*

In order to rule out the case of negative value functions we also make the following assumption.

Assumption 2.1.5. For all $\theta \in \Theta$, $x \to U(x,\theta)$ there exists $\hat{x} \in \text{int}(I)$ such that $U(\hat{x},\theta) > 0$.

Our main result for the forward problem can be summarised as follows.

Solution to the forward problem: Given a generalised diffusion X, if $U(x,\theta) = G(x,\theta) - R(x)$ is log-supermodular then a threshold strategy is optimal on an interval (θ_-,θ_R) and an optimal finite stopping rule does not exist for $\theta > \theta_R$. Furthermore, if U is sufficiently regular and V is differentiable at $\theta \in (\theta_-,\theta_R)$ then

$$V'(\theta) = \frac{U_\theta(x^*(\theta),\theta)}{\varphi(x^*(\theta))},$$

where $x^* : \Theta \to I$ is a monotone increasing function such that $\tau = H_{x^*(\theta)}$ is the optimal stopping rule.

Now suppose that we are given $V = \{V(\theta)\ ;\ \theta \in \Theta\}$ and $G = \{G(x,\theta)\ ;\ x \in \mathbb{R},\ \theta \in \Theta\}$, $c = \{c(x)\ ;\ x \in \mathbb{R}\}$ and X_0. Then the *inverse problem* is to construct a diffusion X such that $V_X = V$ is the value function corresponding to an optimal threshold strategy. As we will see, an important quantity in the inverse problem is the allocation index: suppose $X_0 = x \in \text{int}(I)$, then the allocation index at x is the critical parameter in Θ for which it is optimal to stop immediately. Allocation indices occur naturally in the theory of multi-armed bandits, where they are also known as Gittins indices. The main contribution of this chapter is to establish a connection between allocation indices and solutions to inverse optimal stopping problems. We also show that there is a natural economic interpretation for the allocation index: in the context of inverse stopping problems and real-option theory, allocation indices specify our investment preferences with respect to liquidating for capital gains or remaining invested for future returns. Depending on how we value our investment, we will show how to recover diffusion models for the underlying risky asset consistent with given preferences (allocation indices).

Solution to the inverse problem: Solutions to the inverse problem are parameterised by a choice of allocation index $\theta^* : I \to \Theta$: The functions φ and R defined

$$\varphi(x) = \frac{G_\theta(x, \theta^*(x))}{V'(\theta^*(x))}, \quad R(x) = G(x, \theta^*(x)) - \varphi(x)V(\theta^*(x)),$$

determine the speed measure and scale function of the solution through equations (1.1.3) and (2.1.4).

Parameter dependence of stopping problems is a common theme in the literature on multi-armed bandits in which a special case of the general forward problem, which we will call the *standard problem*, is studied.

Definition 2.1.6. If $G(x,\theta) = G(\theta)$ and $c(x,\theta) \equiv c(x)$ then the problem forward problem (2.1.1) is called the *standard (forward) problem.*

2.1.1 Threshold strategies

Threshold strategies are a natural class of candidates for the optimal stopping time in the forward problem. Our first aim is to establish necessary and sufficient conditions for the optimality of a threshold strategy.

By the strong Markov property of one-dimensional diffusions the value function for the optimal stopping problem can be decomposed into the reward from running the diffusion forever and an *early stopping reward.*

$$V(x,\theta) = R(x,\theta) + \sup_{\tau} \mathbb{E}_x[e^{-\rho\tau}(G(X_\tau,\theta) - R(X_\tau,\theta))]. \tag{2.1.5}$$

We will let $E(x,\theta) = V(x,\theta) - R(x,\theta)$ denote the *optimal early stopping reward* and let $U(x,\theta) = G(x,\theta) - R(x,\theta)$ denote the *early stopping reward function.*

Lemma 2.1.7. *Stopping at the first hitting time of $z \geq X_0$, $z \in \text{int}(I)$ is optimal if and only if $\frac{U(y,\theta)}{\varphi(y)}$ attains its global maximum on $\text{int}(I)$ at z.*

Proof. Suppose that the global maximum is achieved at $z \geq X_0$. Let

$$\hat{E}(\theta) = \frac{U(z,\theta)}{\varphi(z)}.$$

We will show that $E(X_0,\theta) = \hat{E}(\theta)$. On the one hand, $E(X_0,\theta) \geq \hat{E}(\theta)$ since the supremum over all stopping times is larger than the value of stopping upon hitting a given threshold. Moreover $e^{-\rho t}\varphi(X_t)$ is a non-negative local martingale hence a super-martingale. We have that for all stopping times τ,

$$1 \geq \mathbb{E}_{X_0}\left[e^{-\rho\tau}\varphi(X_\tau)\right] \geq \mathbb{E}_{X_0}\left[e^{-\rho\tau}\frac{U(X_\tau,\theta)}{\hat{E}(\theta)}\right],$$

and hence $\hat{E}(\theta) \geq \mathbb{E}_{X_0}[e^{-\rho\tau}(G(X_\tau,\theta) - R(X_\tau,\theta))]$ for all stopping times τ. Hence H_z is optimal.

For the converse, suppose that there exists an $z' \in \text{int}(I)$, $z' \neq z$ such that $\frac{U(z',\theta)}{\varphi(z')} > \frac{U(z,\theta)}{\varphi(z)}$. We will show that there exists a stopping time which is better than H_z. First, if $z' \geq X_0$ then stopping at $\tau = H_{z'}$ is a better strategy than stopping at $\tau = H_z$. Now suppose $z' < X_0$. Then

$$\begin{aligned}
U(z,\theta)\,\mathbb{E}_{X_0}[e^{-\rho H_z}] &= U(z,\theta)\,\mathbb{E}_{X_0}[e^{-\rho H_z}1_{H_z<H_{z'}}] + U(z,\theta)\,\mathbb{E}_{X_0}[e^{-\rho H_{z'}}1_{H_{z'}<H_z}]\,\mathbb{E}_{z'}[e^{-\rho H_z}] \\
&= U(z,\theta)\,\mathbb{E}_{X_0}[e^{-\rho H_z}1_{H_z<H_{z'}}] + U(z',\theta)\,\mathbb{E}_{X_0}[e^{-\rho H_{z'}}1_{H_{z'}<H_z}]\frac{U(z,\theta)/\varphi(z)}{U(z',\theta)/\varphi(z')} \\
&< U(z,\theta)\,\mathbb{E}_{X_0}[e^{-\rho H_z}1_{H_z<H_{z'}}] + U(z',\theta)\,\mathbb{E}_{X_0}[e^{-\rho H_{z'}}1_{H_{z'}<H_z}],
\end{aligned}$$

so stopping at $H_{(z',z)}$ is better than stopping at H_z. ☐

Remark 2.1.8. There is a parallel result for stopping at a threshold below X_0. A threshold below X_0 is optimal if and only if $\frac{U}{\phi}$ attains a global maximum below X_0.

Example 2.1.9. *Recall Example 2.1.3 and let X be a Geometric Brownian Motion started at 1 with volatility parameter σ and drift parameter $\mu < \rho$. Suppose $\Theta = \mathbb{R}^+$, $G(\theta) = \theta$ and*

$c(x,\theta) = x$. *Then* $U(x,\theta) = G(x,\theta) - R(x,\theta) = \theta - x/(\rho - \mu)$. $U(x,\theta)$ *is decreasing so we look for a stopping threshold below* 1. $\phi(x) = x^{-\sqrt{\nu^2+2\rho/\sigma^2}-\nu}$ *for* $0 < x \le 1$, *where* $\nu = \mu/\sigma^2 - 1/2$. *Let* $c_- = \sqrt{\nu^2 + 2\rho/\sigma^2} + \nu$ *and* $x(\theta) = \frac{c_- \theta(\rho-\mu)}{1+c_-}$. *If* $0 < x(\theta) \le 1$ *then* $x(\theta)$ *is the optimal stopping threshold. If* $x(\theta) = 0$ *then it is optimal to 'wait forever'. If* $x(\theta) > 1$ *then it is optimal to stop immediately.*

The following Lemma shows that if a threshold strategy is optimal then the optimal threshold is either above or below the starting point. This rules out the case that both an upper threshold and a lower threshold are optimal for a fixed parameter.

Lemma 2.1.10. *For a fixed parameter* θ, *let* $U(s) = U(s,\theta)$. *Let* $\Delta_- = \{z : z \in argmax_s[U(s)/\phi(s)]\}$ *and* $\Delta_+ = \{z : z \in argmax_s[U(s)/\varphi(s)]\}$. *If* $x \in \Delta_+$ *and* $y \in \Delta_-$ *then* $x \le y$.

Proof. Suppose that $y < x$. It follows that

$$\frac{\varphi(y)}{\phi(y)} = \frac{G(y,\theta)/\phi(y)}{G(y,\theta)/\varphi(y)} > \frac{G(x,\theta)/\phi(x)}{G(x,\theta)/\varphi(x)} = \frac{\varphi(x)}{\phi(x)},$$

contradicting the fact that $\frac{\varphi}{\phi}$ is strictly increasing. □

Example 2.1.11. *Let* X *be Brownian Motion on* $[0, 2\pi]$ *killed at* 0 *and at* 2π. *Let* $c \equiv 0$ *and* $\Theta = \mathbb{R}^+$ *and* $G(x,\theta) = \theta|\sinh(x\sin(x))|$ *and suppose* $\rho = 1/2$. *Then* $\varphi(x) = \sinh(x)$ *and* $\phi(x) = \sinh(2\pi - x)$. *Now fix* $\theta = 1$ *and define* Δ_+ *and* Δ_- *as in Lemma 2.1.10. We calculate* $\Delta_+ = \{\pi/2, 3\pi/2\}$ *and* $\Delta_- \approx \{5.14\}$. *If* X_0 *lies to the left (right) of an element in* Δ_+ (Δ_-) *then an upper (lower) threshold is optimal. If* X_0 *lies between the largest element in* Δ_+ *and the smallest element in* Δ_- *then a threshold strategy is not optimal.*

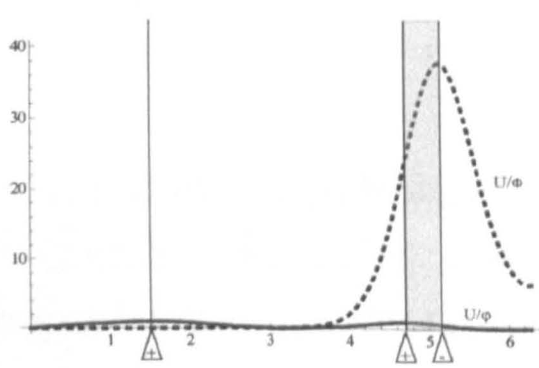

Figure 2.1: Picture for $\theta = 1$. $\frac{U}{\phi}$ is represented by the dashed line and Δ_- is a singleton. $\frac{U}{\varphi}$ is represented by the solid line and Δ_+ consists of two points. There is no optimal threshold strategy if X_0 lies in the shaded region.

In general, given a family of forward problems over an interval Θ, we may find that threshold stopping is optimal on the whole interval Θ, on a subset of Θ or nowhere on Θ.

We will temporarily assume that the forward problem (2.1.1) is such that a threshold strategy is optimal on the whole parameter space. Later, in Section 2.2 we will see how to relax the assumption.

Assumption 2.1.12. For all $\theta \in \Theta$ it is optimal to stop at a threshold above X_0.

There is, as will always be the case, a parallel theory when the optimal thresholds are below X_0, compare Remark 2.1.8.

2.1.2 The envelope theorem

We will now derive our main result for the parameter dependence of the value function through an envelope theorem. The aim is to derive an expression for the derivative of V.

For a fixed parameter θ let $X^*(\theta) = \text{argmax}_{x\in \text{int}(I)} \left[\frac{U(x,\theta)}{\varphi(x)}\right]$. Then $X^*(\theta)$ is the set of possible threshold strategies for a fixed parameter θ. We will let $X^*(\Theta)$ denote the collection of all threshold strategies for the parameter space. Letting $x_+ = \sup\{x : x \in X^*(\Theta)\}$, we have that $X^*(\Theta) \subseteq [X_0, x_+]$. Recall the definition of the *early stopping reward.* We abuse the notation slightly by setting $E(\theta) = V(X_0,\theta) - R(X_0,\theta)$, making the dependence on the starting value implicit. Let us also set $\eta(\theta) = \log(E(\theta))$. The following Proposition follows from an envelope theorem, see Corollary 4 in Segal and Milgrom, [54].

Proposition 2.1.13. *If $[X_0, x_+] \subseteq \text{int}(I)$, $U(x,\theta)$ is upper-semicontinuous in x and $U_\theta(x,\theta)$ is continuous on $[X_0,x_+]\times\Theta$ then V is Lipschitz continuous on (θ_-,θ_+) and the one-sided derivatives are given by*

$$\begin{aligned}
E'(\theta-) &= \min_{x(\theta)\in X^*(\theta)} \frac{U_\theta(x(\theta),\theta)}{\varphi(x(\theta))} \\
E'(\theta+) &= \max_{x(\theta)\in X^*(\theta)} \frac{U_\theta(x(\theta),\theta)}{\varphi(x(\theta))}.
\end{aligned}$$

E is differentiable at θ if and only if $\left\{\frac{U_\theta(x,\theta)}{\varphi(x)} : x(\theta) \in X^*(\theta)\right\}$ is a singleton. In particular we then have

$$\frac{d}{d\theta}\eta(\theta) = u_\theta(x(\theta),\theta), \tag{2.1.6}$$

for $x(\theta) \in X^*(\theta)$ where $u(x,\theta) = \log(U(x,\theta))$.

Remark 2.1.14. Equation (2.1.6) follows by combining the equations $E'(\theta) = \frac{U_\theta(x(\theta),\theta)}{\varphi(x(\theta))}$ (a consequence of the envelope theorem in Milgrom [54]) and $E(\theta) = \frac{U(x(\theta),\theta)}{\varphi(x(\theta))}$ (Lemma 2.1.7).

Remark 2.1.15. The condition $[X_0, x_+] \subseteq \text{int}(I)$ is satisfied if the boundary points of I are accessible.

Corollary 2.1.16. *If the conditions in Proposition 2.1.13 are satisfied then for any $\theta, \theta' \in \Theta$,*

$$E(\theta) - E(\theta') = \int_{\theta'}^{\theta} \frac{U_\theta(x(s), s)}{\varphi(x(s))} ds,$$

where $x(s)$ is a selection from $X^*(s)$.

Corollary 2.1.17. *Suppose $G(x, \theta) \equiv G(\theta)$ is continuously differentiable and $c(x, \theta) = c(x)$. If V is differentiable at θ then*

$$V'(\theta) = G'(\theta)\, \mathbb{E}_{X_0}[e^{-\rho H_{x(\theta)}}],$$

for $x(\theta) \in X^*(\theta)$.

Example 2.1.18. *In Example 2.1.9 if $\mu < \rho$,*

$$V'(\theta) = \left\{ \begin{array}{ll} \left(\frac{\theta c_- (\rho-\mu)}{c_- + 1}\right)^{c_-} & 0 < \theta < \frac{1+c_-}{(\rho-\mu)c_-} \\ 1 & \theta \geq \frac{1+c_-}{(\rho-\mu)c_-} \end{array} \right\}.$$

The preceding Corollary 2.1.17 is the analogue in a diffusion setting of Lemma 2 in Whittle [76]. As in [76] our setup allows for points of non-differentiability and for the possibility of multiple optimal thresholds above the starting point. In contrast, existing results in the diffusion setting, see for instance Karatzas [49] (Lemma 4.1) make strong assumptions on the diffusion and on c which ensure that $X^*(\theta)$ is single valued and that the value function is differentiable in the parameter.

In general, the optimal stopping thresholds for a parameter are given by a set-valued map $X^* : \Theta \rightarrow I$. We will now define the inverse map from the domain of the diffusion to the parameter space.

Definition 2.1.19. $\Theta^*(x)$, *the allocation index at* x, is the set of parameters $\theta \in \Theta$ for which it is optimal to stop immediately when $X_0 = x$.

The definition of the allocation index Θ^* as the critical parameter(s) for which immediate stopping is optimal generalises the definition common in the theory of multi-armed bandits: while we make few assumptions on the reward functions, the multi-armed bandit or dynamic allocation literature is restricted to the standard problem ($c(x, \theta) = c(x)$ and $G(x, \theta) = \theta$), see for instance Gittins and Glazebrook [37], Whittle [76] and for a diffusion setting closer to the setting of this chapter, Karatzas [49] and Alvarez [4].

The following example illustrates our approach to parameter dependent stopping problems and the idea of calculating critical parameter values. Although we have focused on the case when the forward problem is indexed by a single parameter, the analysis of forward problems parameterised by several parameters is analogous.

Example 2.1.20. A toy model for tax effects *Suppose X is a model for the profits of a firm: $X_t = \sigma B_t + x_0$, where B is a standard Brownian Motion and $\sigma > 0$. In a tax-free environment a model for the value V of the firm is*

$$V(\rho, \delta, \sigma) = \sup_{\tau} \mathbb{E}_{X_0} \left[\int_0^{\tau} e^{-\rho t} X_t dt + e^{-\rho \tau} \delta \right],$$

where δ is the salvage value of the firm. $U(x, \delta) = \delta - x/\rho$ is decreasing in x and we look for an optimal stopping threshold below X_0. We have $\phi(x) = e^{-\sqrt{2\rho}x/\sigma}$ and $R(x) = \frac{x}{\rho}$. Let x_1^* be the optimal threshold (investment decision) in the tax-free environment above. We calculate $x_1^* = x_1^*(\rho, \delta, \sigma) = \min\{\delta\rho - \frac{\sigma}{\sqrt{2\rho}}, X_0\}$.

Now consider what happens to the value of the firm if profits are taxed. Suppose that profits are taxed at a rate θ, and that the tax-base at time t is $X_t - d$, where d represents a tax-deductible depreciation expense (or some other adjustment to the tax base). The post-tax profit of the firm is $Y_t = X_t - \theta(X_t - d)$. The decreasing solution to (2.1.2) is $\phi^Y(x) = \exp\left(\frac{-x\sqrt{2\rho}}{(1-\theta)\sigma}\right)$ while $R^Y(x) = R((1-\theta)x + \theta d) = \frac{(1-\theta)x+\theta d}{\rho}$. The optimal threshold (x_2^*) for the after-tax investment problem is

$$x_2^*(\theta, \rho, m, \sigma, d) = \min\left\{ \frac{\rho\delta - \theta d}{1-\theta} - \frac{\sigma(1-\theta)}{\sqrt{2\rho}}, X_0 \right\}.$$

In taxation theory, a tax-rate is neutral if it does not change investment decisions. It is sometimes considered desirable for taxes to be neutral, see for instance Samuelson [71]. Let θ_N denote the neutral tax rate in this problem. To compute θ_N we solve $x_1^*(\rho, \delta, \sigma) = x_2^*(\theta_N, \rho, \delta, \sigma, d)$ for θ_N, to find

$$1 - \theta_N(\rho, \delta, \sigma, d) = \sqrt{2\rho}\, \frac{d - \delta\rho}{\sigma}.$$

Finally we check that $\theta_N \in (0, 1)$ if and only if $0 < d - \delta p < \sigma/\sqrt{2\rho}$. Similarly, given a tax rate θ we could calculate the depreciation adjustment $d^*(\rho, \delta, \sigma, \theta)$ so that the investment decision is unchanged, which is the idea in Samuelson [71].

In Example 2.1.20, the optimal thresholds are monotone in one or more of the parameters. In the next section we will derive natural conditions for the monotonicity of threshold strategies X^*. We will see that if X^* is monotone then we can relax Assumption 2.1.12.

2.2 Monotonicity of the optimal stopping threshold in the parameter value

We will say that X^* is increasing if $x \in X^*(\theta)$ and $x' \in X^*(\theta')$ with $\theta \leq \theta'$ implies $x \leq x'$.

Definition 2.2.1.

i) A function $f : \mathbb{R}^2 \to \mathbb{R}$ is supermodular in (y, z) if for all $y' > y$, $f(y', z) - f(y, z)$ is increasing in z and if for all $z' > z$, $f(y, z') - f(y, z)$ is increasing in y. Equivalently, f is supermodular if $f(\max\{y', y\}, \max\{z', z\}) + f(\min\{y', y\}, \min\{z', z\}) \geq f(y, z) + f(y', z')$ for all (y, z).

ii) If the inequalities in i) are strict then f is called strictly supermodular

iii) If $-f$ is (strictly) supermodular, f is called (strictly) submodular.

iv) f is (strictly) log-supermodular if $\log(f)$ is (strictly) supermodular.

Remark 2.2.2. Note that if f is twice differentiable then f is supermodular in (y, z) if and only if $f_{yz}(y, z) \geq 0$ for all y and z.

The following Lemma is an important result in this chapter. It follows from a straightforward application of standard techniques in monotone comparative statics to the setting of optimal stopping, see for instance Athey [5].

Lemma 2.2.3. *Suppose that* $U(x, \theta) = G(x, \theta) - R(x, \theta) > 0$ *on* $\mathrm{int}(I) \times \Theta$. *If* U *is* log-*supermodular then* X^* *is increasing in* θ.

Proof. Suppose that $\theta > \hat{\theta}$. $X^*(\theta)$ and $X^*(\hat{\theta})$ are non-empty by Assumption 2.1.12. Define a function f via $f(x, \theta) = u(x, \theta) - \psi(x)$, where $\psi(x) = \log(\varphi(x))$ (recall the definition of φ, (1.1.2)). Then f is also supermodular. Now for any $x(\theta) \in X^*(\theta)$ and $x(\hat{\theta}) \in X^*(\hat{\theta})$ we have

$$0 \geq f(\max\{x(\theta), x(\hat{\theta})\}, \theta) - f(x(\theta), \theta) \geq f(x(\hat{\theta}), \hat{\theta}) - f(\min\{x(\theta), x(\hat{\theta})\}, \hat{\theta}) \geq 0.$$

The first inequality follows by definition of $X^*(\theta)$ the second by supermodularity and the last inequality by definition of $X^*(\hat{\theta})$. Hence there is equality throughout and $\max\{x(\theta), x(\hat{\theta})\} \in X^*(\theta)$ and $\min\{x(\theta), x(\hat{\theta})\} \in X^*(\hat{\theta})$. It follows that $X^*(\theta)$ is increasing in θ. □

Corollary 2.2.4. *If* U *is* log-*submodular then* $X^*(\theta)$ *is decreasing in* θ.

Remark 2.2.5. It may be the case that $U(x, \theta)$ takes both strictly positive and negative values on $\mathrm{int}(I) \times \Theta$. In this case it is never optimal to stop at x' if $U(x', \theta) \leq 0$ and so we need only check supermodularity on the set $\{(x, \theta) : U(x, \theta) > 0)\}$.

The following assumption will ensure that X^* is increasing. There will be a parallel set of results when X^* is decreasing.

Assumption 2.2.6. $U(x, \theta) > 0$ and $\log(U)$ is supermodular.

In general, if we remove Assumption 2.1.12, a threshold strategy may never be optimal or be optimal on some subset of parameters in Θ. In the following we will show that if $U(x,\theta)$ is log-supermodular then a threshold strategy will be optimal for all parameters in a sub-interval of Θ. Let θ_R be the infimum of those values in Θ for which $X^*(\theta) = \emptyset$. If $X^*(\theta) = \emptyset$ for all $\theta \in \Theta$ then we set $\theta_R = \theta_-$.

Lemma 2.2.7. *The set of $\theta \in \Theta$ where $X^*(\theta)$ is non-empty (threshold stopping is optimal) forms an interval with end-points θ_- and θ_R.*

Proof. Let ϱ denote the right end-point of I. Suppose $X^*(\hat{\theta}) \neq \emptyset$ and $\theta \in (\theta_-, \hat{\theta})$. We claim that $X^*(\theta) \neq \emptyset$.

Fix $\hat{x} \in X^*(\hat{\theta})$. Then $E(X_0, \hat{\theta}) = u(\hat{x}, \hat{\theta}) - \psi(\hat{x})$ and

$$u(\hat{x}, \hat{\theta}) - \psi(\hat{x}) \geq u(x, \hat{\theta}) - \psi(x), \qquad \forall x < \varrho, \tag{2.2.1}$$

and for $x = \varrho$ if $\varrho \in \text{int}(I)$. We write the remainder of the proof as if we are in the case $\varrho \in \text{int}(I)$; the case when $\varrho \notin \text{int}(I)$ involves replacing $x \leq \varrho$ with $x < \varrho$.

Fix $\theta < \hat{\theta}$. We want to show

$$u(\hat{x}, \theta) - \psi(\hat{x}) \geq u(x, \theta) - \psi(x), \qquad \forall x \in (\hat{x}, \varrho], \tag{2.2.2}$$

for then

$$\sup_{x \leq \varrho}\{u(x,\theta) - \psi(x)\} = \sup_{x \leq \hat{x}}\{u(x,\theta) - \psi(x)\},$$

and since $u(x,\theta) - \psi(x)$ is continuous in x the supremum is attained.

Since u is supermodular by assumption we have for $x \in (\hat{x}, \varrho]$

$$u(\hat{x}, \hat{\theta}) - u(\hat{x}, \theta) \leq u(x, \hat{\theta}) - u(x, \theta). \tag{2.2.3}$$

Subtracting (2.2.3) from (2.2.1) gives (2.2.2). □

In the *standard case*, determining whether $u(x,\theta) = \log(G(x,\theta) - R(x,\theta))$ is supermodular is simplified by the following result.

Lemma 2.2.8. *Suppose that the boundary points of X are inaccessible.*

If $G \equiv 0$ then $X^*(\theta)$ is increasing in θ if and only if $-c(x,\theta)$ is log*-supermodular.*

In the standard case $G(\theta) - R(x)$ is log*-supermodular if and only if $Q(x,\theta)$ is* log*-supermodular where $Q : I \times \Theta \to \mathbb{R}$, $Q(x,\theta) = \rho G(\theta) - c(x)$.*

Proof. Athey [5] and Jewitt [45] prove that $f : \mathbb{R}^2 \to \mathbb{R}^+$ and $h : \mathbb{R}^2 \to \mathbb{R}^+$ are log-supermodular if and only if $\int_s f(x,s)h(s,\theta)ds$ is log-supermodular. The first statement now follows from the fact (e.g Alvarez [4] or Rogers [67], V.50) that $R(x,\theta) = \int_I r(x,y)c(y,\theta)m(dy)$, where $r(x,y)$ is a product of two single-variate functions and hence log-supermodular.

For the second statement note that $\mathbb{E}_x\left[\int_0^\infty e^{-\rho t}(\rho G(\theta) - c(X_t))dt\right] = G(\theta) - R(x)$. By the result of Athey and Jewitt, $G(\theta) - R(x)$ is log-supermodular if and only if $Q(x,\theta)$ is log-supermodular. $\square$

Example 2.2.9. *Recall Example 2.1.4. Let X be a three-dimensional Bessel process started at 1, $\rho = 1/2$, $c(x,\theta) = \theta\cos(x)$ and $G \equiv 0$. We have $\varphi(x) = \frac{\sinh(x)}{\sinh(1)x}$. Note that $c(x,\theta)$ is both* log*-supermodular and* log*-submodular. Suppose $\theta > 0$. $\frac{-R(x,\theta)}{\varphi(x)}$ attains its maximum at $\hat{x}$ where $\hat{x} \approx 2$ is the smallest solution to the equation* $\coth(x)(x\cos(x) - \sin(x)) + x\sin(x) = 0$. *For $\theta < 0$, the maximum is attained at the second smallest root of the same equation and $\hat{x} \approx 5.4$. Hence we find that $X^*(\theta)$ is decreasing. This does not contradict Lemma 2.2.8 because*

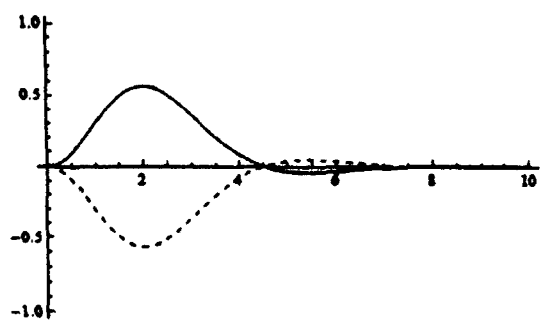

Figure 2.2: $\frac{-R(x,\theta)}{\phi(x)}$ for $\theta = 1$ (solid line) and $\theta = -1$ (dashed line).

Assumption 2.2.6 is violated: The set of points where $-R(x,\theta)$ is positive when $\theta > 0$ and stopping is feasible coincides with the set of points where $-R(x,\theta)$ is negative (and stopping is therefore not feasible) when $\theta < 0$. Compare Remark 2.2.5.

2.3 Inverse optimal stopping problems

In this section our aim is to recover diffusions consistent with a given value function for a stopping problem. We recall that when $c \equiv 0$ and $G(x,\theta) = (x-\theta)^+$, the problem has the interpretation of recovering price-processes consistent with perpetual American call option prices. Now suppose that we hold a dividend bearing stock which we may liquidate at any time for capital gains which depend on a parameter such as the tax rate. Suppose that we have calculated our value for the investment as a function of the tax rate. Then the allocation index has the natural interpretation of being the minimum tax rate at which we remain invested for each level of the stock price and the question we are interested in is how to recover allocation indices and models for the stock price process consistent with a given value function.

As before, let Θ be an interval with end-points θ_- and θ_+. Let us assume that we are given $V = \{V(\theta)\ ;\ \theta \in \Theta\}$, $G = \{G(x,\theta)\ ;\ x \in \mathbb{R},\ \theta \in \Theta\}$, $c = \{c(x)\ ;\ x \in \mathbb{R}\}$ and X_0.

Inverse Problem: Find a generalised diffusion X such that $V_X = V$ is consistent with one-sided stopping above X_0.

We will make the following regularity assumption.

Assumption 2.3.1. $V : \Theta \to \mathbb{R}$ is differentiable and $(x,\theta) \to G(x,\theta)$ is twice continuously differentiable.

We begin by introducing the notation we will use for our inverse stopping problem framework. The main change over the previous section is that we will highlight dependence on the (unknown) speed measure m and scale function s.

We wish to recover a speed measure m and scale function s to construct a diffusion $X^{m,s} = (X_t^{m,s})_{t\geq 0}$, supported on a domain $I^m \subseteq \mathbb{R}$ such that $V_{X^{m,s}} = V$. Our approach to solving this problem is to recover solutions $\varphi_{m,s}$ and $R_{m,s}$ to the differential equations (1.1.3) and (2.1.4) from V and to solve the two equations 'in reverse' to recover the speed measure and the scale function. Let $\psi_{m,s} = \log(\varphi_{m,s})$, where $\varphi_{m,s}$ is the increasing solution to (1.1.3) with $\varphi_{m,s}(X_0) = 1$ and let $R_{m,s}(x) = \mathbb{E}_x[\int_0^\infty e^{-\rho t} c(X_t^{m,s})dt]$. We will say that functions $R_{m,s}$, $\varphi_{m,s}$, X^*, etc. are *consistent* with the inverse problem if there exists a diffusion $X^{m,s}$ such that $V_{X^{m,s}} = V$. Our approach involves establishing $\psi_{m,s}$ and $\eta_{m,s}(\theta) = \log(V(\theta) - R_{m,s}(X_0))$ as $u_{m,s}$-convex dual functions, where $u_{m,s}(x,\theta) = \log(G(x,\theta) - R_{m,s}(x))$ (recall the definition of u-convexity in Chapter 1). Denote the $u_{m,s}$-convex duals of $\varphi_{m,s}$ and $\eta_{m,s}$ by $\varphi_{m,s}^u$ and $\eta_{m,s}^u$ respectively.

We will now see that in the context of the forward stopping problem, Proposition 1.2.11 is a version of Proposition 2.1.13. Suppose $V = V_{X^{m,s}}$ and that $u_{m,s}(x,\theta)$ is strictly log-supermodular and twice continuously differentiable. Then $\eta_{m,s}(\theta)$ is $u_{m,s}$-convex with $u_{m,s}$-subdifferential X^*, i.e. $\eta_{m,s}(\theta) = \sup_{x\in int(I^m)}[u_{m,s}(x,\theta) - \psi_{m,s}(x)] = u_{m,s}(x^*(\theta),\theta) - \psi_{m,s}(x^*(\theta))$, for some optimal stopping threshold $x^*(\theta) \in X^*(\theta)$. Hence by Proposition 1.2.11, we have

$$\frac{V'(\theta)}{V(\theta) - R_{m,s}(X_0)} = \frac{G_\theta(x^*(\theta))}{G(x^*(\theta),\theta) - R_{m,s}(x^*(\theta))}.$$

Substituting for $\varphi_{m,s}(x^*(\theta))$ we find

$$V'(\theta) = \frac{G_\theta(x^*(\theta), \theta)}{\varphi_{m,s}(x^*(\theta))},$$

which is the expression in Proposition 2.1.13 when V is differentiable.

In the following, whenever $u_{m,s}$ is strictly supermodular and $V_{X^{m,s}} = V$ we will let x^* denote the non-decreasing function satisfying $\eta'_{m,s}(\theta) = \frac{\partial}{\partial\theta}u_{m,s}(x^*(\theta), \theta)$. Then X^* is the set of points on the graph of x^*. We will let θ^* denote the inverse of the function x^*.

Corollary 2.3.2. *Suppose $u_{m,s}$ is strictly supermodular and twice continuously differentiable. If $\psi_{m,s}$ is $u_{m,s}$-subdifferentiable then the allocation index at $x \in int(I^m)$ satisfies $\psi'_{m,s}(x) = \frac{\partial}{\partial x}u_{m,s}(x, \theta^*(x))$. Moreover, θ^* is non-decreasing.*

2.3.1 Recovering consistent diffusions

The following theorem provides necessary and sufficient conditions for a diffusion $X^{m,s}$ to be the solution to the inverse stopping problem.

Proposition 2.3.3. *$X^{m,s}$ solves the inverse problem if and only if $\varphi_{m,s}$ and $R_{m,s}$ satisfy the following two conditions.*

i) For all $\theta \in \Theta$, $\psi^u_{m,s}(\theta) = \sup_{x\in int(I^m),\ x\geq X_0} [u_{m,s}(x,\theta) - \psi_{m,s}(x)]$,

ii) $\psi^u_{m,s}(\theta) = \log(V(\theta) - R_{m,s}(X_0))$.

Proof. If $X^{m,s}$ is consistent with V, G, X_0 and one-sided stopping above X_0 then

$$\begin{aligned} \log(V(\theta) - R_{m,s}(X_0)) &= \log(V_{X^{m,s}}(\theta) - R(X_0^{m,s})) \\ &= \sup_{x\in int(I^m),\ x\geq X_0} [u_{m,s}(x,\theta) - \psi_{m,s}(x)] \\ &= \sup_{x\in int(I^m)} [u_{m,s}(x,\theta) - \psi_{m,s}(x)] \\ &= \psi^u_{m,s}(\theta). \end{aligned}$$

On the other hand, if the two conditions are satisfied then we can construct a diffusion $X^{m,s}$ with starting point X_0. The first condition implies that one-sided stopping above X_0 is optimal while the second condition ensures that $V_{X^{m,s}} = V$. □

It is intuitively clear that a value function contains information about the dynamics of a consistent diffusions above the starting point. If $x \geq X_0$, and the allocation index function θ^* is known then the solution to (1.1.3) must satisfy $\varphi(x) = \frac{U(x,\theta^*(x))}{V(\theta^*(x)) - R(X_0)}$. Thus if we can calculate

$R(X_0)$ and the allocation index for all $x \geq X_0$ then we can calculate φ above the starting point. We can then recover pairs of scale functions and speed measures consistent with the solution above X_0 through (1.1.3).

On the other hand, for $x < X_0$, the only information we have is that $\frac{U(x,\theta^*(x))}{\varphi(x)}$ does not attain a maximum below the starting point, otherwise V would not be consistent with one-sided stopping above X_0. Thus, while we may attempt to specify (unique) diffusion dynamics above X_0, we expect there to be a variety of consistent specifications of the diffusion dynamics below the starting point. This is analogous to the situation in Hobson and Ekström [30] where a unique consistent volatility co-efficient is derived below the starting point but there is freedom of choice above the starting point.

The following two examples illustrate the ideas involved in recovering a consistent diffusion in the simplified setting when consistent diffusions are assumed to be either martingales (Example 2.3.4) or in natural scale and with additional information about the early stopping reward (Example 2.3.5).

Example 2.3.4. *Let $\Theta = \left(0, \frac{k+1}{k}\right]$ for some positive constant k. Suppose $V(\theta) = (\frac{k\theta}{k+1})^k \frac{\theta}{k+1} + 1$, $G(x,\theta) = \theta$, $c(x) = \rho x$ and $X_0 = 1$. Suppose the inverse problem is restricted to the class of diffusions that are also martingales. Then $s(x) = x$ and $R_{m,s}(x) = x$. We have $u_{m,s}(x,\theta) = \log(\theta - x)$ and calculate $\eta^u_{m,s}(x) = \sup_\theta[\log(\theta - x) - \log(V(\theta) - 1)] = \log(x^{-k})$, where the maximum is attained at $\theta^*(x) = \frac{x(k+1)}{k}$. To recover a consistent martingale diffusion on $\mathbb{R}^+$, let us extend the parameter space to $\bar{\Theta} = (0,\infty)$ and set $\theta^*(x) = \frac{x(k+1)}{k}$ on $(0,\infty)$. Then we find that $\phi_{m,s}(x) = x^{-k}$ on $(0,\infty)$ is a consistent eigenfunction. It follows that*

$$dX_t = \sigma X_t dB_t, \quad X_0 = 1$$

is consistent with V where σ satisfies $\sigma^2 = \frac{2\rho}{(k+1/2)^2 - 1/4}$.

Example 2.3.5. *Let $\Theta = [1,\infty)$. Recall the decomposition of the forward problem by the strong Markov property (2.1.5). Suppose we are given the optimal early stopping reward $E(\theta) = e^{\theta^2/2}$ and the early stopping reward function $U(x,\theta) = e^{\theta x}$ and that $X_0 = 0$. In this example, $\eta(\theta) = \log(E(\theta))$ is known, so we suppress the subscripts m and s. We calculate $\sup_\theta[u(x,\theta) - \eta(\theta)] = x^2/2$ where the maximum is attained at $\theta^*(x) = x$. Let us suppose that $s(x) = x$ and aim at recovering a (local)-martingale diffusion. On $X^*(\Theta) = [1,\infty)$, the candidate eigenfunction for the diffusion is $\varphi_{m,s}(x) = e^{\eta^u(x)} = e^{x^2/2}$. Solving for σ in (2.1.2) we obtain $\sigma(x) = \frac{2\rho}{1+x^2}$ for $x \in [1,\infty)$. We will recover a consistent diffusion by extrapolating the allocation index. Let $\bar{\Theta} = (0,\infty)$ and set $\theta^*(x) = x$ for $0 \leq x \leq 1$. By Proposition 1.2.11, $\psi_{m,s}$ is u-convex on $X^*(\bar{\Theta})$ if $\frac{\varphi'_{m,s}(x)}{\varphi_{m,s}(x)} = \theta^*(x) = x$. Thus by setting $\varphi_{m,s}(x) = e^{x^2/2}$ for $x \in \mathbb{R}^+$ we find that the diffusion* *with dynamics*

$$dX_t = \sigma(x)dB_t + dL_t, \quad X_0 = 0, \quad \sigma^2(x) = \frac{2\rho}{1+x^2}$$

is consistent with V, where L is the local time at 0.

In general we can choose any increasing function θ^* on $[0,1)$ with $\theta^*(1-) = 1$ as long as the recovered function $\varphi_{m,s}$ is an eigenfunction for a consistent diffusion. For instance, the choice $\theta^*(x) = \frac{3}{4}x^{1/2}$ for $0 \le x < 1$ leads to $\varphi_{m,s}(x) = \exp(\frac{x^{3/2}}{2})$ for $0 \le x < 1$. For this choice of extension and again setting $s(x) = x$, the consistent diffusion co-efficient is

$$\sigma^2(x) = \left\{ \begin{array}{ll} \frac{32\rho x^{1/2}}{6+9x^{3/2}} & 0 \le x < 1 \\ \frac{2\rho}{1+x^2} & x \ge 1 \end{array} \right\}.$$

Note that for this extension $\varphi'_{m,s}$ jumps at 1 *and since*

$$\varphi'_{m,s}(1+) - \varphi'_{m,s}(1-) = 2\rho\varphi_{m,s}(1)m(\{1\}),$$

we have $m(\{1\}) = \frac{1}{8\rho}$ (compare Example 1.4.3). Hence the increasing additive functional Γ_u includes a multiple of the local time at 1 *and the diffusion X^m is 'sticky' at* 1.

For the general case, the main difficulty over the previous simplified examples of inverse problems is having to recover both a speed measure and a non-trivial scale function. This means that we must recover $R_{m,s}$ as well as $\varphi_{m,s}$ to obtain two equations (1.1.3), (2.1.4) for the two unknown quantities.

2.3.2 Recovering diffusions through a consistent allocation index

If a monotone allocation index $\theta^* : I^m \to \Theta$ is consistent with an inverse problem then there exists a diffusion $X^{m,s}$ such that the inverse function x^* is the optimal threshold strategy for the forward problem. Then by Proposition 2.1.13,

$$V'(\theta^*(x)) = \frac{G_\theta(x, \theta^*(x))}{\varphi_{m,s}(x)}. \tag{2.3.1}$$

Lemma 2.3.6. *If θ^* is consistent with the inverse problem then all consistent diffusions $X^{m,s}$ satisfy*

$$\mathbf{E}_x\left[\int_0^\infty e^{-\rho t}c(X_t^{m,s})dt\right] = G(x, \theta^*(x)) + \varphi_{m,s}(x)(R_{m,s}(X_0) - V(\theta^*(x)))$$

for all $x \in int(I^m)$.

Proof. By the definition of θ^*, $x \in X^*(\theta^*(x))$. It follows by (2.1.6) that

$$\frac{V'(\theta^*(x))}{V(\theta^*(x)) - R_{m,s}(X_0)} = \frac{G_\theta(x, \theta^*(x))}{G(x, \theta^*(x)) - R_{m,s}(X_0)}$$

for all $x \in X^*(\Theta)$. Combining this equation with (2.3.1) we have

$$R_{m,s}(x) = G(x, \theta^*(x)) + \varphi_{m,s}(x)(R_{m,s}(X_0) - V(\theta^*(x))). \tag{2.3.2}$$

□

Let $\hat{R}_{m,s}(x)$ be the function on $X^*(\Theta)$ defined $\hat{R}_{m,s}(x) = G(x, \theta^*(x)) - \varphi_{m,s}(x)V(\theta^*(x))$.

Lemma 2.3.7. *If $R_{m,s}(x)$ given by equation (2.3.2) solves (2.1.4) then so does $\hat{R}_{m,s}$.*

Proof. Follows from the fact that $\varphi_{m,s}$ is a solution to the homogeneous equation (1.1.3). □

Given an inverse problem there will in general be many speed measures and scale functions satisfying the conditions in Proposition 2.3.3. Each solution corresponds to an optimal threshold strategy X^*. By definition, choosing a consistent allocation index is equivalent to choosing a consistent threshold strategy. Thus, rather than searching over all solutions $X^{m,s}$ satisfying the conditions in Proposition 2.3.3, we can solve inverse problems by specifying a candidate allocation index. The following verification result provides a set of easily verifiable conditions for $X^{m,s}$ to solve the inverse problem.

Proposition 2.3.8. *$X^{m,s}$ is a solution to an inverse problem if the following conditions are satisfied.*

i) $u_{m,s}$ is strictly supermodular and twice continuously differentiable and $\varphi_{m,s}$ is differentiable almost everywhere,

ii) there exists a monotone function $x^* : \bar{\Theta} \to I^m$ with inverse θ^* such that $\Theta \subseteq \bar{\Theta}$, $x^*(\theta) \geq X_0$ for $\theta \in \Theta$ and such that whenever $\psi_{m,s}$ is differentiable

$$\psi'_{m,s}(x) = \frac{\partial}{\partial x} u_{m,s}(x, \theta^*(x)),$$

iii) $\eta_{m,s} = \psi^u_{m,s}$.

Proof. By Proposition 1.2.11 and conditions i) and ii) $\psi_{m,s}$ is $u_{m,s}$-convex. It follows from iii) and Proposition 2.3.3 that $V_{X^{m,s}} = V$. □

Theorem 2.3.9. *A consistent monotone allocation index determines a unique solution to the inverse problem on $X^*(\Theta)$.*

Proof. Suppose θ^* is a consistent monotone allocation index. Define $f : X^*(\Theta) \to \mathbb{R}$, $f(x) = \frac{G_\theta(x,\theta^*(x))}{V'(\theta^*(x))}$ and $g : X^*(\Theta) \to \mathbb{R}$, $g(x) = G(x,\theta^*(x)) - f(x)V(\theta^*(x)))$. By (2.3.1), $f = \varphi_{m,s}$ on $X^*(\Theta)$ for a consistent diffusion $X^{m,s}$, hence f is a solution to (1.1.3) on $X^*(\theta)$. Similarly by Lemma 2.3.7, g is a solution to (2.1.4) on $X^*(\Theta)$. Solving the two equations for m and s we recover the (unique) dynamics of a consistent diffusion on $X^*(\Theta)$. □

Example 2.3.10. *Suppose $\Theta = (0, \frac{k+1}{k}]$ and for a positive constant k, $V(\theta) = \left(\frac{k\theta}{k+1}\right)^k \frac{\theta}{k+1} + 1$, $X_0 = 1$, $G(x,\theta) = \theta$ and $c(x) = \gamma x$ where γ is another positive constant. We define a family of allocation indices on $(0,1]$ parameterised by $\alpha > 0$ via $\theta^*_\alpha(x) = \frac{x^\alpha(k+1)}{k}$.*

We will calculate candidate diffusions using Proposition 2.3.8. By (2.3.1) we have that for $x \in X^*_\alpha(\Theta) = (0,1]$ a candidate solution to (2.1.2) corresponding to the allocation index θ^*_α is $\phi^\alpha(x) = x^{-\alpha k}$. Similarly by (2.3.2) we have $R^\alpha(x) = x^\alpha + \phi^\alpha(x)(R^\alpha_{m,s}(1) - 1)$ and so $\hat{R}^\alpha_{m,s}(x) = x^\alpha$ is a candidate solution to (2.1.4). Then, by equations (2.1.2) and (2.1.4), the corresponding consistent diffusion co-efficients on $X^*_\alpha(\Theta) = (0,1]$ are

$$\begin{aligned}
\sigma^2_\alpha(x) &= \frac{2(\rho(1+k)x^2 - k\gamma x^{3-\alpha})}{k(1+k)\alpha^2} \\
\mu_\alpha(x) &= (1+\alpha k)\frac{\rho x(1+k) - k\gamma x^{2-\alpha}}{k(1+k)\alpha^2} - \frac{\rho x}{\alpha k}.
\end{aligned}$$

Note that $\sigma^2_\alpha(x) \geq 0$ on $(0,1]$ if and only if $x^{1-\alpha} \leq \frac{\rho(1+k)}{k\gamma}$ and hence for a consistent diffusion to exist on $(0,1]$ the problem parameters must satisfy $\alpha \leq 1$ and $\rho + k(\rho - \gamma) \geq 0$.

To specify a diffusion on $(0,\infty)$ consistent with a given $\alpha \leq 1$ on $(0,1]$, we let

$$\theta^*(x) = \left\{ \begin{array}{ll} \theta^*_\alpha(x) & x \in X^*(\Theta) = (0,1] \\ \theta^*_3(x) & x > 1 \end{array} \right\}.$$

The corresponding diffusion X^α is given by

$$dX^\alpha_t = \sigma_\alpha(X_t)dB_t + \mu_\alpha(X_t)dt, \quad X_0 = 1$$

where

$$\sigma^2_\alpha(x) = \left\{ \begin{array}{ll} \frac{2(\rho(1+k)x^2 - k\gamma x^{3-\alpha})}{k(1+k)\alpha^2} & 0 < x \leq 1 \\ \frac{2(\rho(1+k)x^2 - k\gamma)}{9k(1+k)} & x > 1 \end{array} \right\}$$

and

$$\mu_\alpha(x) = \left\{ \begin{array}{ll} (1+\alpha k)\frac{\rho x(1+k) - k\gamma x^{2-\alpha}}{k(1+k)\alpha^2} - \frac{\rho x}{\alpha k} & 0 < x \leq 1 \\ (1+3k)\frac{\rho x(1+k) - k\gamma x^{-1}}{9k(1+k)} - \frac{\rho x}{3k} & x > 1 \end{array} \right\}.$$

The particular choice of θ^* on $(1,\infty)$ is convenient because it ensures that the trivial condition $\sigma^2_\alpha(x) \geq 0$ is satisfied for any choice of ρ, k and γ satisfying $\rho + k(\rho-\gamma) \geq 0$.

Since both boundary points are inaccessible we have $R^\alpha = \hat{R}^\alpha$ or equivalently $R^\alpha(1) = 1$. Note that if we set $\gamma = \rho$ and $\alpha = 1$ then we recover Example 2.3.4.

In addition to the problem data V, G, c and X_0, suppose we are given the allocation index θ^* representing our investment preferences. It is then interesting to ask whether there exists a diffusion model for the process consistent with the value function and our allocation strategy. A consistent diffusion, if it exists, will be uniquely determined on $X^*(\Theta)$. The following example illustrates this situation.

Example 2.3.11. *Suppose $\Theta = (1,3)$, $V(\theta) = 1+\frac{\theta}{3}(\sqrt{3\theta}-\sqrt{\theta/3})$, $G(x,\theta) = \theta x$, $c(x) = 1/x$ and $X_0 = 1$. Furthermore suppose we are given $\theta^*(x) = 3/x^2$ for $x \in (0,\infty)$. Then $x^*(\theta) = \sqrt{3/\theta}$ and $X^*(\Theta) = (1,\sqrt{3})$. By (2.3.1) we have $\varphi_{m,s}(x) = x^2$ and $R_{m,s}(x) = 1/x + x^2(R_{m,s}(1)-1)$ so that $\hat{R}_{m,s}(x) = 1/x$ is a candidate solution to equation (2.3.2). The differential equations (1.1.3) and (2.1.4) lead to the following simultaneous equations.*

$$\begin{aligned}
\sigma^2(x) + 2\mu(x)x &= \rho x^2, \\
\sigma^2(x) - x\mu(x) &= \rho x^2 - x^2.
\end{aligned}$$

We calculate $\sigma^2(x) = x\mu(x) + \rho x^2 - x^2 = \rho x^2 - 2\mu(x)x$ so that $\mu(x) = x/3$ and $\sigma^2(x) = x^2(\rho - 2/3)$. It follows that we must have $\rho > 2/3$ for a solution to the inverse problem to exist. Provided this condition is satisfied, the (unique) solution to the inverse problem is

$$dX_t = \sqrt{\rho - \frac{2}{3}}X_t dB_t + \frac{X_t}{3}dt, \quad X_0 = 1.$$

2.4 Concluding remarks

The main contribution of this chapter is to provide a new interpretation for the allocation (Gittins) index based on its role in solving inverse stopping problems. In the context of the forward problem we showed that the idea of an allocation index can be extended naturally from the 'standard case' to a general class of optimal stopping problems and that there are natural conditions under which the index is monotone. Furthermore, the allocation index parameterises solutions to the inverse problem. When an investment can be modelled as a perpetual horizon stopping problem, the index has the natural economic interpretation of representing investment preferences with respect to liquidating for a terminal reward or remaining invested for a running reward and the option to liquidate later.

Chapter 3

Model-independent bounds for variance swaps and optimality properties of the Perkins and Azéma-Yor embeddings

The main purpose of this chapter is to establish a connection between model-independent bounds for variance swaps and Skorokhod embedding theory. We will begin by introducing a particular definition of a variance swap which leads to a related Skorokhod embedding problem through a series of simple inequalities. In Chapter 4 we will build on the solution to this Skorokhod embedding problem to construct model independent bounds and hedging strategies for a general family of variance swaps.

3.1 Motivation: A variance swap on squared returns

Let $X = (X_t)_{0\leq t\leq T}$ represent the discounted price of a financial asset. Under the assumption of no-arbitrage, there exists a measure under which X is a (local)-martingale. We may suppose that there exists a filtered probability space $(\Omega, \mathcal{F}, \mathbb{F}, \mathbb{P})$ such that B is a $\mathbb{F}$-Brownian motion and such that $X_t = B_{A_t}$ for a (possibly discontinuous) time-change $t \to A_t$, null at 0. (If X is continuous then the existence of such a time-change is guaranteed by the Dambis-Dubins-Schwarz Theorem and in general the existence is guaranteed by Monroe [57], Theorem 2.) Since X is a non-negative price process we suppose it has starting value $X_0 = B_0 = x_0 > 0$.

Now suppose that we know the prices of put and call options with maturity T. Knowledge of put and call option prices with expiry time T is equivalent to knowledge of the marginal law of the process at time T (see Breeden and Litzenberger [13]). Suppose that $X_T \sim \mu$ and that μ is centred at x_0, and has support in $\mathbb{R}^+$ (then X is a true martingale and not a strict local martingale). We will determine bounds for the fair value of a variance swap given the terminal law μ. Note that if $X_T \sim \mu$ then A_T is a solution of the Skorokhod embedding problem for μ in B.

Following Demeterfi et al. [24] we define the pay-out $V = V((X_s)_{0 \le s \le T})$ of an idealised variance swap as

$$V_T = \int_0^T \frac{d[X,X]_t}{(X_{t-})^2} = \int_0^T \left(\frac{dX_t^c}{X_{t-}}\right)^2 + \sum_{0 \le t \le T} \left(\frac{\Delta X_t}{X_{t-}}\right)^2$$

where $\Delta X_t = X_t - X_{t-}$, and X^c is the continuous part of X. We refer to Chapter 4 for a general introduction to variance swaps where we will see that the idealised variance swap introduced in this chapter arises as the continuous time limit of a particular discrete-time definition of the variance swap.

Let A^c be the continuous part of A. Note that $dA_t^c = (dX_t^c)^2 = d[X,X]_t^c$. Let $S^X = (S_t^X)_{t \ge 0}$ (respectively S) be the process of the running maximum of X (respectively B), and let I^X (respectively I) denote the corresponding infimum. Then we have $X_t \le S_t^X \le S_{A_t}$ and it follows that path-by-path with $\Delta B_{A_t} = B_{A_t} - B_{A_{t-}}$ that

$$\begin{aligned} V_T &\ge \int_0^T \frac{d[X,X]_t^c}{(S_{t-}^X)^2} + \sum_{0 \le t \le T} \left(\frac{\Delta X_t}{S_{t-}^X}\right)^2 \qquad (3.1.1) \\ &\ge \int_0^T \frac{dA_t^c}{(S_{A_{t-}})^2} + \sum_{0 \le t \le T} \left(\frac{\Delta B_{A_t}}{S_{A_{t-}}}\right)^2. \end{aligned}$$

We suppose that X has a second moment. Then $(X_t)_{0 \le t \le T}$ is a square-integrable martingale and we find that,

$$\begin{aligned} \mathbb{E}\left[\int_0^T \frac{dA_t^c}{(S_{A_{t-}})^2} + \sum_{0 \le t \le T} \left(\frac{\Delta B_{A_t}}{S_{A_{t-}}}\right)^2\right] &= \mathbb{E}\left[\int_0^T \frac{dA_t^c + \Delta A_t}{(S_{A_{t-}})^2}\right] \qquad (3.1.2) \\ &= \mathbb{E}\left[\int_0^T \frac{dA_t}{(S_{A_{t-}})^2}\right] \\ &\ge \mathbb{E}\left[\int_0^{A_T} \frac{du}{(S_u)^2}\right]. \end{aligned}$$

We say that τ is an embedding of μ if τ is a stopping time for which B_τ has law μ (we write $B_\tau \sim \mu$ or $\mu = \mathcal{L}(B_\tau)$). Let $\mathcal{S} \equiv \mathcal{S}(B,\mu)$ be the set of stopping times which embed μ and let $\mathcal{S}_{UI} = \mathcal{S}_{UI}(B,\mu)$ be the subset of $\mathcal{S}(B,\mu)$ for which $(B_{t \wedge \tau})_{t \ge 0}$ is uniformly integrable. The inequalities above imply that the fair value of V_T is bounded below by

$$\inf_{\tau \in S_{UI}(B,\mu)} \mathbb{E}\left[\int_0^\tau \frac{du}{S_u^2}\right]. \tag{3.1.3}$$

Similarly, using the inequality $I_{A_t} \leq I_t^X \leq X_t$ we find that the fair value of V_T is bounded above by

$$\sup_{\tau \in S_{UI}(B,\mu)} \mathbb{E}\left[\int_0^\tau \frac{du}{I_u^2}\right]. \tag{3.1.4}$$

Now let $G(b,s) = \frac{(s-b)^2}{s^2}$ then by Itô's Lemma,

$$G(B_\tau, S_\tau) = G(0,0) + \int_0^\tau \frac{du}{S_u^2} - \int_0^\tau \frac{2(S_u - B_u)}{S_u^2} dB_u.$$

It follows that if $\int_0^{\tau \wedge t} \frac{2(S_u - B_u)}{S_u^2} dB_u$ is a uniformly integrable martingale then

$$\mathbb{E}\left[\int_0^\tau \frac{du}{S_u^2}\right] = \mathbb{E}\left[\frac{(S_\tau - B_\tau)^2}{S_\tau^2}\right],$$

and the question of bounding the fair value of V_T is transformed into a question of maximising or minimising expressions of the form $\mathbb{E}[F(B_\tau, S_\tau)]$ over embeddings of μ. We return to the calculation of variance swap bounds in Section 3.7.1.

In general, the Skorokhod embedding problem (SEP) (Skorokhod [72]) is to find a stopping time τ such that the stopped process satisfies $B_\tau \sim \mu$. There are many classical solutions to this problem (for a survey listing twenty-one, see Obloj [60]), and further solutions continue to appear in the literature, including most recently Hirsch et al [39].

Given the variety of solutions to the SEP, it is natural to search for embeddings with additional optimality properties. In particular, if Ψ is a functional of the stopped Brownian path $(B_t)_{0 \leq t \leq \tau}$, then these constructions aim to maximise Ψ over (a suitable subclass of) embeddings of μ. For example, if F is an increasing function, and $S_t = \sup_{s \leq t} B_s$, then the Azéma-Yor solution [7] maximises $\mathbb{E}[F(S_\tau)]$ over uniformly integrable embeddings, and the Perkins embedding [61] minimises the same quantity.

In this chapter we extend this result to functions $F = F(B_\tau, S_\tau)$. As we have seen, expressions for the upper and lower bounds for the fair value V_T can be recast in this form.

3.2 Preliminaries

Let $(\Omega, \mathcal{F}, \mathbf{F}, \mathbb{P})$ be a filtered probability space satisfying the usual conditions and supporting a Brownian motion $W = (W_t)_{t \geq 0}$ with $W_0 = 0$ (which is a martingale with respect to $\mathbf{F}$), and sufficiently rich that $\mathcal{F}_0$ contains a further uniform random variable which is independent of W. Let μ be a centred probability measure. To exclude trivialities we assume that μ is not δ_0, the unit mass at 0. We say that τ is an embedding of μ if τ is a stopping time for which W_τ has law μ (we write $W_\tau \sim \mu$ or $\mu = \mathcal{L}(W_\tau)$) and we say that τ is uniformly integrable if the family $(W_{t\wedge\tau})_{t\geq 0}$ is uniformly integrable.

Let $\mathcal{S} \equiv \mathcal{S}(W,\mu)$ be the set of stopping times which embed μ, and let $\mathcal{S}_{UI} \equiv \mathcal{S}_{UI}(W,\mu)$ be the subset of $\mathcal{S}(W,\mu)$ consisting of uniformly integrable stopping times. For $\mathcal{S}_{UI}(W,\mu)$ to be non-empty we must have that μ is centred (i.e. $\int_{\mathbb{R}} |x|\mu(dx) < \infty$ and $\int_{\mathbb{R}} x\mu(dx) = 0$). In this context (Brownian motion and centred target laws) a result of Monroe [56] gives that a stopping time is uniformly integrable if and only if it is minimal (in the sense that if τ is minimal and $\sigma \leq \tau$ with $W_\sigma \sim W_\tau$, then $\sigma \equiv \tau$ almost surely). The class of minimal stopping times is a natural class of 'good' (in the sense of small) stopping times.

For the Brownian motion W, started at 0, we write H_x for the first hitting time of x, and for a set A, $H_A = \inf\{u \geq 0 : W_u \in A\}$.

For a process $(Y_t)_{t\geq 0}$ and a stopping time σ we write $Y^\sigma = (Y_t^\sigma)_{t\geq 0}$ for the stopped process $Y_t^\sigma = Y_{\sigma\wedge t}$.

Given a centred probability measure μ, let X_μ be a random variable with law μ and define $C(x) \equiv C_\mu(x) = \mathbb{E}[(X_\mu - x)^+]$ and $P(x) \equiv P_\mu(x) = \mathbb{E}[(x - X_\mu)^+]$. Then C and P are monotonic convex functions with $c(0) = p(0)$. Then $U(x) = U_\mu(x) = \mathbb{E}[|X_\mu - x|] = C(x) + P(x)$ is (minus) the potential associated with μ. Conversely any convex function U with $\lim_{x\to\pm\infty}(U(x) - |x|) = 0$ is the potential of some centred probability measure μ (Chacon [18]).

If μ has an atom at zero then we write μ^* for the measure obtained by omitting the atom at 0, and then rescaling to get a probability measure. Thus $\mu^*(A) = \mu(A \setminus \{0\})/(1 - \mu(A))$. Finally, we write $\hat{x} = \hat{x}_\mu$ for the upper limit on the support of μ (so $\hat{x}_\mu = \sup\{x : C_\mu(x) > 0\}$) and $\check{x} = \check{x}_\mu$ for the corresponding lower limit $\check{x}_\mu = \inf\{x : P_\mu(x) > 0\}$.

3.2.1 The Azéma-Yor solution

For $x \geq 0$, up to the upper limit on the support of μ, define $\beta = \beta_\mu$ by

$$\beta(x) = \text{argmin}_{y<x} \frac{C_\mu(y)}{x - y}. \tag{3.2.1}$$

Then β is an increasing function with $\beta(x) < x$. Where the argmin is not uniquely defined it is not important which value we choose. However, we fix one by insisting that β is right continuous, or equivalently by choosing the largest value for which the minimum is attained. Observe that at $x = 0$, β takes the value of the infimum of the support of μ. For x equal to, or to the right of, the upper limit on the support of μ we set $\beta(x) = x$.

For an increasing function $\beta : \mathbb{R}^+ \mapsto \mathbb{R}$ with $\beta(x) \leq x$ let τ_β be given by

$$\tau_\beta = \inf\{u : W_u \leq \beta(S_u)\}.$$

Then $\tau^{AY} \equiv \tau_\mu^{AY}$, the Azéma-Yor stopping time for μ, is given by $\tau_\mu^{AY} \equiv \tau_{\beta_\mu}$. Thus we have $\tau_{\beta_\mu} \in \mathcal{S}_{UI}(W,\mu)$, and moreover, for F increasing, τ_{β_μ} maximises $\mathbb{E}[F(S_\tau)]$ over $\tau \in \mathcal{S}_{UI}(W,\mu)$ (Azéma-Yor [7, 6], Rogers [65]).

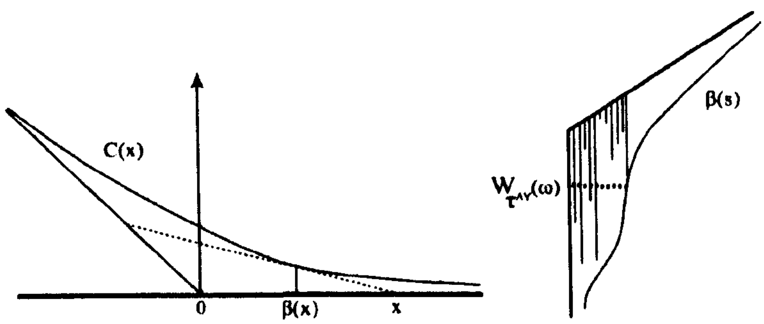

Figure 3.1: For each x, the value of $\beta(x)$ is determined by finding the tangent line to C_μ originating at x: $\beta(x)$ is the horizontal co-ordinate of the point of contact between the tangent line and C_μ. (If C_μ includes a straight line section then this point of contact may not be uniquely defined in which case we take $\beta(x)$ to be the largest value of the horizontal co-ordinate at which contact occurs.) The stopping time τ_β associated to this construction is given by the first time that an excursion from the maximum crosses below β.

Note that τ_{β_μ} does not maximise this quantity over *all* embeddings, but it does give the maximum over uniformly integrable (i.e. minimal) embeddings.

Let $b \equiv b_\mu$ be the right-continuous inverse to β. Then b is the barycentre function and for $x < \hat{x}_\mu$, $b(x)$ is given by

$$b(x) = \mathbb{E}[X_\mu | X_\mu \geq x]. \tag{3.2.2}$$

The barycentre $b(x)$ is defined up to the upper limit of the support of μ and is a non-negative, non-decreasing function with $b(x) \geq x$. We set $b(x) = x$ for $x \geq \hat{x}_\mu$. (The reverse barycentre $\check{b}(x) = \mathbb{E}[X|X \leq x]$ is defined analogously to the barycentre.)

It is more standard to define the barycentre function as in (3.2.2) and to set β to be the inverse barycentre function, but the two approaches are equivalent, and our approach via potentials allows for a unified treatment with the Perkins construction in the next section.

If μ has an interval with no mass, then b is constant over that interval and β has a jump. If μ has an atom at x_0 then b has a jump at x_0 (unless the atom is at the upper limit $\hat{x}$ of the support of μ in which case $b(\hat{x}) = \hat{x}$) and β is constant over a range of s. From the definition of τ_β see (3.2.1), and excursion theory (see Rogers [66], Equation 2.13)

$$\exp\left(-\int_0^s \frac{dr}{r-\beta(r)}\right) = \mathbb{P}(S_{\tau_\beta} \geq s) = \mathbb{P}(W_{\tau_\beta} \geq \beta(s)) = \mu(\beta(s), \infty) \tag{3.2.3}$$

and then also $\mathbb{P}(S_{\tau_\beta} \geq s) = \mathbb{P}(W_{\tau_\beta} \geq \beta(s)) = \mu(\beta(s), \infty)$. Note that it does not matter which convention we use for $\beta(s)$ here since μ places no mass on $(\beta(s-), \beta(s+))$.

Example 3.2.1. *If $\mu = U[-1,1]$, then $C_\mu(x) = (x-1)^2/4$ and $P_\mu(x) = (x+1)^2/4$ (at least for $-1 = \check{x}_\mu \leq x \leq \hat{x} = 1$). Then the barycentre function is given by $b(x) = (x+1)/2$ for $-1 \leq x \leq 1$ and hence $\beta(s) = 2s - 1$ for $0 \leq s \leq 1$. It follows that $S_{\tau_\mu^{AY}} \equiv b(W_{\tau_\mu^{AY}})$ is uniformly distributed on $U[0,1]$.*

Lemma 3.2.2. *If μ places mass on (x, ∞) then $(r - \beta(r))^{-1}$ is integrable over $[0, x]$.*

Proof. This follows immediately from (3.2.3) and $\mathbb{P}(S_{\tau_\beta} \geq x) \geq \mathbb{P}(W_{\tau_\beta} \geq x) > 0$. □

3.2.2 The Perkins solution

For $x > 0$ define $\alpha_\mu^+ = \alpha^+ : \mathbb{R}_+ \to \mathbb{R}_-$ by

$$\alpha^+(x) = \operatorname{argmin}_{y<0} \frac{C_\mu(x) - P_\mu(y)}{x - y}, \tag{3.2.4}$$

and for $x < 0$ define $\alpha_\mu^- = \alpha^- : \mathbb{R}_- \to \mathbb{R}_+$ by

$$\alpha^-(x) = \operatorname{argmax}_{y>0} \frac{P_\mu(x) - C_\mu(y)}{y - x}. \tag{3.2.5}$$

Then $\alpha^\pm$ are monotonic functions. If the argmin (or the argmax) is not uniquely defined we take the largest value (in modulus) for which the minimum is attained; in this way $\alpha^\pm : \mathbb{R}_\pm \mapsto \mathbb{R}_\mp$ is right-continuous. Again, none of the subsequent analysis will depend on this convention. For convenience we will sometimes write α instead of $\alpha^\pm$. If P_μ (respectively C_μ) is differentiable at $\alpha^+(x)$ (respectively $\alpha^-(x)$), then $\alpha^+(x)$ (respectively $\alpha^-(x)$) satisfies

$$\frac{C_\mu(x) - P_\mu(\alpha^+(x))}{x - \alpha^+(x)} = P'_\mu(\alpha^+(x)) \tag{3.2.6}$$

respectively $P_\mu(x) - C_\mu(\alpha^-(x)) = C'_\mu(\alpha^-(x))(x - \alpha^-(x))$.

Let $a^\pm$ be the inverse to $\alpha^\pm$ and let $\bar{a}(w) = w$ for $w > 0$ and $\bar{a}(w) = a^+(w)$ for $w < 0$. Let I be the infimum process for W so that $I_t = \inf_{s \leq t} W_s$.

For a pair of monotonic functions $\alpha^+ : \mathbb{R}_+ \mapsto \mathbb{R}_-$ (non-increasing) and $\alpha^- : \mathbb{R}_- \mapsto \mathbb{R}_+$

(non-decreasing) define the stopping time

$$\tau_\alpha = \inf\{u : W_u \le \alpha^+(S_u)\} \wedge \inf\{u : W_u \ge \alpha^-(I_u)\}.$$

Suppose μ does not have an atom at zero. Then the Perkins [61] embedding $\tau^P \equiv \tau_\mu^P \equiv \tau^P(\mu)$ is given by $\tau_\mu^P = \tau_{\alpha_\mu}$.

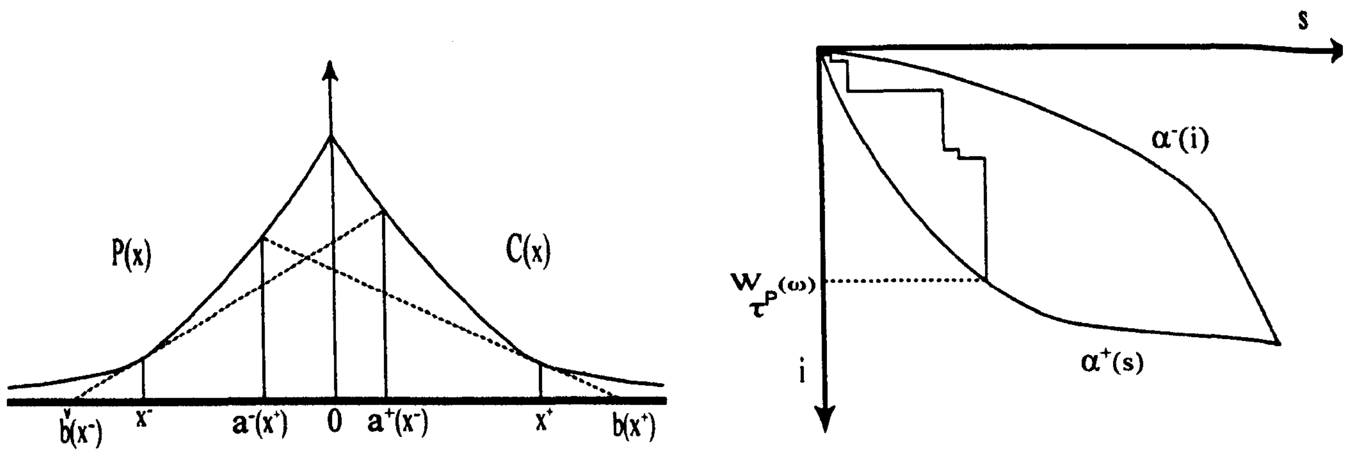

Figure 3.2: For $x^+ > 0$, $a^-(x^+)$ is the horizontal co-ordinate of the point where the tangent line to C emanating from $b(x^+)$ intersects with P. $a^+(x^-)$ is found similarly, by drawing tangents to P emanating from the reverse barycentre function evaluated at x^- and determining intersection points with C. The stopping rule associated with this construction is to stop the Brownian motion when its running maximum or minimum exit the region determined by the inverses to those functions α^+ and α^-.

If μ has an atom at zero, then we use independent randomisation to set $\tau^P = 0$ with probability $\mu(\{0\})$; and otherwise $\tau^P = \tau_{\alpha_\mu}$. More precisely, in the case where μ has an atom at zero we set the Perkins embedding to be

$$\tau^P = \begin{cases} 0 & \text{if } Z \le \mu(0), \\ \tau_{\alpha_\mu} & \text{if } Z > \mu(0), \end{cases}$$

where Z is a uniform random variable which is measurable with respect to $\mathcal{F}_0$. Here $\alpha_\mu^\pm$ are the quantities defined in (3.2.4) and (3.2.5) for μ. Note that if μ^* is obtained from μ by removing any mass at zero, and rescaling to give a probability measure, then although C_{μ^*} and P_{μ^*} are scalar multiples of C_μ and P_μ respectively, nonetheless we have $\alpha_{\mu^*}^\pm \equiv \alpha_\mu^\pm$.

Note that if μ has an atom at zero then we need $\mathcal{F}_0$ to be non-trivial in order to be able to define the Perkins embedding. Note further that since there are potentially many uniform random variables Z which are measurable with respect to $\mathcal{F}_0$, if $\mu(\{0\}) > 0$, the Perkins embedding is not unique. Sometimes it is convenient to think about the Perkins embedding associated with an identified $\mathcal{F}_0$ random variable Z, in which case we write $\tau_\mu^{P,Z}$ instead of just τ_μ^P.

The results of Perkins [61] show that $\tau_\mu^P \in \mathcal{S}_{UI}(W,\mu)$ and moreover, for F increasing, τ^P minimises $\mathbb{E}[F(S_\tau)]$ over $\tau \in \mathcal{S}(W,\mu)$, and not just $\mathcal{S}_{UI}(W,\mu)$ (Perkins [61], although the representation via (3.2.4) and (3.2.5) is due to Hobson and Pedersen [43]).

Example 3.2.3. *If $\mu = U[-1,1]$ then $P = P_\mu$ and $C = C_\mu$ are as given in Example 3.2.1 and $\alpha^+(s)$ is the unique root of the equation $P'(\alpha)(s-\alpha) = C(s) - P(\alpha)$. It is easily verified that this root is given by $\alpha^+(s) = s - 2\sqrt{s}$. Similarly, $\alpha^-(i) = i + 2\sqrt{|i|}$. It can be shown that $\mathbb{P}(S_{\tau_\alpha} \geq s) = \mathbb{P}(W_{\tau_\alpha} \geq s) = \mathbb{P}(W_{\tau_\alpha} \leq s - 2\sqrt{s}) = 1 - \sqrt{s})$.*

Example 3.2.4. *Notwithstanding the above example, in general it is difficult to derive an explicit form for the stopping boundary associated with the Perkins stopping time. Here we give a second example where analytic expressions, albeit complicated ones, can be derived.*

Suppose the target law is a centred Pareto distribution with support $[-1,\infty)$ and density function $f(x) = 2(x+2)^{-3}$. Then for $k \geq -1$, $C(k) = (2+k)^{-1}$ and $P(k) = k + (2+k)^{-1}$, and for $k < -1$, $C(k) = -k$, $P(k) = 0$.

Then, for the Azéma-Yor embedding, β solves $C(\beta) = (s-\beta)|C'(\beta)|$ and $\beta(s) = (s/2)-1$.

For the Perkins embedding, $\alpha^+(s)$ solves $p'(\alpha^+) = (c(s) - p(\alpha^+))/(s-\alpha^+)$ and we have (after some algebra)

$$\alpha^+(s) = \frac{-2s^2 - 5s + \sqrt{s^4 + 6s^3 + 12s^2 + 8s}}{2s - 1 + s^2}.$$

The expression for α^- is, $\alpha^-(i) = \frac{-3i-2i^2+\sqrt{-(i^4+6i^3+12i^2+8i)}}{2i+1+i^2}$.

If μ has an interval in $\mathbb{R}_+$ (respectively $\mathbb{R}_-$) with no mass, then α^- (respectively α_+) has a jump (unless that interval is contiguous with zero, in which case $\alpha^\pm$ starts at a non-zero value). If μ has an atom in $(0,\infty)$ (respectively $(-\infty,0)$) then α^- (respectively α_+) is constant over a range of values.

Lemma 3.2.5. *Suppose $x > 0$. If μ places mass on $[x,\infty)$ then $(r-\alpha^+(r))^{-1}$ is integrable over $(0,x)$.*

Proof: We have $(W_u \geq \alpha^+(S_u); \forall u \leq H_x) \supseteq (\tau_\alpha \geq H_x) \supseteq (W_\tau \geq x)$ and then by excursion theory, recall (3.2.3),

$$\exp\left(-\int_0^x \frac{dr}{r-\alpha^+(r)}\right) = \mathbb{P}(W_u \geq \alpha^+(S_u); \forall u \leq H_x) \geq \mu([x,\infty)) > 0.$$

□

3.3 Convergence of measures and convergence of embeddings

Let $(\mu_n)_{n\geq 1}$ be a sequence of measures and write U_n, β_n and α_n as shorthand for U_{μ_n}, β_{μ_n} and α_{μ_n}, with a similar convention for other functionals.

Suppose that, for each n, μ_n is centred and that $(\mu_n)_{n\geq 1}$ converges weakly to μ, where μ is also centred. Then it does not follow that $U_n \to U_\mu$, nor that $\beta_n \to \beta_\mu$, nor that $\alpha_n \to \alpha_\mu$. However, with the correct additional hypotheses, then these types of convergence are equivalent.

Our first key result is the following.

Proposition 3.3.1. *Let (μ_n) be a sequence of measures such that $\mu_n \Rightarrow \mu$ and $\mathbb{E}[|X_{\mu_n}|] \to \mathbb{E}[|X_\mu|]$. Then $b_n(x) \to b(x)$ at continuity points $x < \hat{x}$ of b.*

Proof. Chacon [18] shows that if $\mu_n \Rightarrow \mu$ and $U_n(0) \to U(0)$ then $U_n \to U$ pointwise.

It follows trivially that $C_n \to C$ pointwise, where $C_n(x) = C_{\mu_n}(x)$ and $C(x) = C_\mu(x)$. Recall that x is a discontinuity point of b if and only if there is an atom of μ at x. Suppose $x < \hat{x}$ is a continuity point of b. Then (3.2.2) gives $b(x) = x + \frac{C(x)}{\mu([x,\infty))}$ and

$$b_n(x) = x + \frac{C_n(x)}{\mu_n([x,\infty))} \to x + \frac{C(x)}{\mu([x,\infty))} = b(x).$$

□

Corollary 3.3.2. *Let (μ_n) be a sequence of measures such that $\mu_n \Rightarrow \mu$ and $\mathbb{E}[|X_{\mu_n}|] \to \mathbb{E}[|X_\mu|]$. Then $\beta_n(s) \to \beta(s)$ at continuity points $s < \hat{x}$ of β. Moreover, if $\hat{x} < \infty$ then for each $z > \hat{x}$,* $\liminf \beta_n(z) \geq \hat{x}$.

Proof. Since $b_n(\hat{x} - \epsilon) < \hat{x} + \epsilon$ for sufficiently large n we have for these same n that $\beta_n(\hat{x} + \epsilon) \geq \hat{x} - \epsilon$ □

Corollary 3.3.3. *Under the assumptions of Proposition 3.3.1, $\tau_{\beta_n} \to \tau_\beta$ almost surely.*

Proof. Let D be the set of discontinuity points of β. If $S_{\tau_\beta} \notin D$ then $W_{\tau_\beta} = \beta(S_{\tau_\beta})$ and it follows that

$$(\omega : \tau_{\beta_n} \not\to \tau_\beta) \subseteq \left(\omega : S_{\tau_\beta} \in D\right) \cup \left(\omega : S_{\tau_\beta} \notin D, W_{\tau_\beta} = \beta(S_{\tau_\beta}), \tau_{\beta_n} \not\to \tau_\beta\right).$$

For any stopping time σ write Let $H_x^\sigma = \inf\{u \geq \sigma : W_u = x\}$.

Case 1: $\hat{x} = \infty$.

Note that since β is increasing, D is countable and $\mathbb{P}(S_{\tau_\beta} \in D) = 0$.

First we argue that on $(\omega : S_{\tau_\beta} = x)$ we have that for sufficiently large n, $S_{\tau_{\beta_n}} \geq x$: since there are only countably many values of $s < x$ on which the value of W_u gets below $S_u = s$, and on each of these excursions W stays above $\beta(S)$, for sufficiently large n, W must stay above $\beta_n(S)$ also.

Hence $\liminf_n S_{\tau_{\beta_n}} \geq S_{\tau_\beta}$ almost surely. Then on $\{\omega : S_{\tau_\beta} = x \notin D, W_{\tau_\beta} = \beta(x)\}$, we have $\tau_{\beta_n}(\omega) \to \tau_\beta(\omega)$ unless $\inf\{W_u : \tau_\beta \leq u \leq H^{\tau_\beta}_{S_{\tau_\beta}}\} = W_{\tau_\beta} = \beta(x)$ and $\beta_n(x) < \beta(x)$ But, almost surely, on any interval of positive length Brownian motion goes below its starting value. In particular, the set $(\omega : S_{\tau_\beta} \notin D, W_{\tau_\beta} = \beta(S_{\tau_\beta}), \tau_{\beta_n} \not\to \tau_\beta)$ has probability zero.

Case 2: $\hat{x} < \infty$ and $\mu(\{\hat{x}\}) = 0$.

The only paths for which issues of convergence might be different to the previous case are those for which $S_{\tau_\beta} = \hat{x}$. But since μ has no atom at $\hat{x}$, $\mathbb{P}(S_{\tau_\beta} = \hat{x}) = \mathbb{P}(W_{\tau_\beta} = \hat{x}) = 0$ and $\tau_{\beta_n} \to \tau_\beta$ almost surely.

Case 3: $\hat{x} < \infty$ and $\mu(\{\hat{x}\}) > 0$.

In this case $\beta(\hat{x}-) := \lim_{y\uparrow\hat{x}} \beta(y) < \beta(\hat{x}) = \hat{x}$. We show that on the set $(S_{\tau_\beta} = \hat{x})$ we have $\lim \tau_{\beta_n} = \tau_\beta$, almost surely. Off the set $(S_{\tau_\beta} = \hat{x})$ convergence follows exactly as in the previous cases.

First we argue that $\limsup_n S_{\tau_{\beta_n}} \leq \hat{x}$ almost surely. Fix $z > \hat{x}$, then given $0 < \epsilon < z - \hat{x}$, there exists N such that for $n \geq N$, $\beta_n(\hat{x} + \epsilon) > \hat{x} - \epsilon$. Hence, for sufficiently large n,

$$\left(\omega : S_{\tau_{\beta_n}}(\omega) \geq z\right) \subseteq \left(\omega : \inf\{W_u : H_{\hat{x}+\epsilon} \leq u \leq H_z\} \geq \hat{x} - \epsilon\right).$$

But

$$\mathbb{P}\left(\inf\{W_u : H_{\hat{x}+\epsilon} \leq u \leq H_z\} \geq \hat{x} - \epsilon\right) \leq \exp\left(-\int_{\hat{x}+\epsilon}^{z} \frac{dy}{y - (\hat{x} - \epsilon)}\right) = \frac{2\epsilon}{z - \hat{x} + \epsilon}.$$

By choosing ϵ small compared with $(z - \hat{x})$ we deduce that $\limsup_n S_{\tau_{\beta_n}} \leq z$ almost surely for any $z > \hat{x}$.

Now we argue that on $S_{\tau_\beta} = \hat{x}$ we have $\liminf W_{\tau_{\beta_n}} \geq \hat{x}$ almost surely. Coupled with the result from the previous paragraph we can then conclude that on $W_{\tau_\beta} = \hat{x}$ we have $\tau_{\beta_n} \to H_{\hat{x}} = \tau_\beta$.

Given δ and $\epsilon < \hat{x} - \beta(\hat{x}-) - \delta$, there exists N such that for all $n > N$, $\beta_n(\hat{x} - \epsilon) < \beta(\hat{x}-) + \epsilon < \hat{x} - \delta$. Then

$$\begin{aligned}\left(\omega : W_{\tau_{\beta_n}}(\omega) < \hat{x} - \delta, S_{\tau_\beta}(\omega) = \hat{x}\right) \subseteq\ & \left(\omega : \inf\{W_u : H_{\hat{x}-\epsilon} \leq u \leq H_{\hat{x}}\} \leq \hat{x} - \delta\right) \\ & \cup\left(\omega : S_{\tau_{\beta_n}} < \hat{x} - \epsilon, S_{\tau_\beta} = \hat{x}\right).\end{aligned}$$

By similar arguments to those in Case 1 we can prove that the final event has small probability. Moreover, using that the fact that the probability that an event occurs is smaller than the expected number of times that it occurs,

$$\mathbb{P}(\omega : \inf\{W_u : H_{\hat{x}-\epsilon} \leq u \leq H_{\hat{x}}\} \leq \hat{x} - \delta) \leq \int_{\hat{x}-\epsilon}^{\hat{x}} \frac{dy}{y - (\hat{x} - \delta)} = \ln(\delta/(\delta - \epsilon)).$$

By choosing ϵ compared to δ this probability can be made arbitrarily small. □

Note that if $\tau_{\beta_n} \to \tau_\beta$ almost surely, then by the continuity of Brownian motion $W_{\tau_{\beta_n}} \to W_{\tau_\beta}$ almost surely and $\mu_n \Rightarrow \mu$.

We can summarise the results as follows:

Proposition 3.3.4. *Suppose that $(\mu_n)_{n\geq 1}$ and μ are centred and that $\mathbb{E}[|X_{\mu_n}| \to \mathbb{E}[|X_\mu|]$. Then the following are equivalent:*

(i) $\mu_n \Rightarrow \mu$ and $\mathbb{E}[|X_{\mu_n}|] \to \mathbb{E}[|X_\mu|]$;

(ii) $U_n(x) \to U_\mu(x)$ for each $x \in \mathbb{R}$;

(iii) $\beta_n \to \beta$ at continuity points s of β, provided s is less than or equal to the upper limit on the support of μ;

(iv) $\tau_{\beta_n} \xrightarrow{a.s.} \tau_\beta$;

(v) $W_{\tau_{\beta_n}} \xrightarrow{a.s.} W_{\tau_\beta}$.

Now we want to prove a similar result for the Perkins embedding.

Lemma 3.3.5. *Let $(\mu_n)_{n\geq 1}$ be a sequence of centred probability measures such that $\mu_n \Rightarrow \mu$ and $\mathbb{E}[|X_n||] \to \mathbb{E}[|X_\mu|]$. Then $a_n^\pm(x) \to a^\pm(x)$ at continuity points $x \in (\check{x}, \hat{x}) \setminus \{0\}$ of a.*

Proof. We prove the result for (a_n^+, a^+), the other case being similar.

Again we have that $x < 0$ is a discontinuity point of a^+ if and only if there is an atom of μ at x. Suppose that x is not an atom of x. Then $a^+(x)$ is the unique solution of $P(x) + P'(x)(z - x) = C(z)$. Moreover, for any $\hat{a}_n(x) \in (a_n^+(x+), a_n^+(x-))$

$$\begin{aligned} P_n(x) + P_n'(x+)(\hat{a}_n(x) - x) &\geq C_n(\hat{a}_n(x)), \\ P_n(x) + P_n'(x-)(\hat{a}_n(x) - x) &\leq C_n(\hat{a}_n(x)). \end{aligned}$$

Suppose $a_n^+(x) \to \gamma$ (down a subsequence if necessary). Then, since $P_n(x+) \to P(x)$ and $P_n'(x-) \to P(x)$,

$$P(x) + P'(x)(\gamma - x) \geq C(\gamma) \geq P(x) + P'(x)(\gamma - x).$$

Hence $\gamma = a^+(x)$ and $a_n^+(x) \to a(x)$. □

Proposition 3.3.6. *Suppose that $(\mu_n)_{n\geq 1}$ and μ are centred and that $\mathbb{E}[|X_{\mu_n}|\to \mathbb{E}[|X_\mu|]$.*

(a) Suppose there exists an open interval I containing 0 such that $\mu_n(I) = \mu(I) = 0$. Then the following are equivalent:

(i) $\mu_n \Rightarrow \mu$ and $\mathbb{E}[|X_{\mu_n}|] \to \mathbb{E}[|X_\mu|]$;

(ii) $U_n(x) \to U_\mu(x)$ for each $x \in \mathbb{R}$;

(iii) $\alpha_n^\pm \to \alpha^\pm$ at continuity points of $\alpha^\pm$ which lie within the range of the support of μ;

(iv) $\tau^P_{\mu_n} \xrightarrow{a.s.} \tau^P_\mu$;

(v) $W_{\tau^P_{\mu_n}} \xrightarrow{a.s.} W_{\tau^P_\mu}$.

(b) More generally, suppose $\mu_n \Rightarrow \mu$, $\mathbb{E}[|X_{\mu_n}|] \to \mathbb{E}[|X_\mu|]$. Then, $\alpha_n^\pm \to \alpha^\pm$ at continuity points of $\alpha^\pm$ which lie within the range of the support of μ.

Suppose further that $\mu_n(\{0\}) \to \mu(\{0\})$. Then there exists a sequence of Perkins embeddings of μ_n such that $\tau^P_{\mu_n}$ converges in probability to a Perkins embedding τ^P_μ of μ. In particular, if Z_n converges in probability to Z then the Perkins embedding $(\tau^{P,Z_n}_{\mu_n})_{n\geq 1}$ converges in probability to the Perkins embedding $\tau^{P,Z}_\mu$ of μ.

Thus, if $\mu_n \Rightarrow \mu$, $\mathbb{E}[|X_{\mu_n}|] \to \mathbb{E}[|X_\mu|]$ and $\mu_n(\{0\}) \to \mu(\{0\})$ then if $(\tau^{P,Z_n}_{\mu_n})_{n\geq 1}$ is a sequence of Perkins embeddings of $(\mu_n)_{n\geq 1}$ then there exists a subsequence n_k along which $\lim \tau^{P,Z_{n_k}}_{\mu_{n_k}}$ exists almost surely and is a Perkins embedding of μ.

Proof. For Part (a) the equivalence of *(i)* and *(ii)* follows as before. Lemma 3.3.5 gives that *(ii)* implies *(iii)*. It follows from the pathwise construction of τ_{α_n} (and the existence of the interval I which is not charged by μ_n so that $\tau^P_{\mu_n} \equiv \tau_{\alpha_n}$) that $\tau^P_{\mu_n} \to \tau^P_\mu$ almost surely and hence we have *(iii)* implies *(iv)*. The continuity of Brownian motion allows us to deduce *(v)*, from which *(i)* follows immediately.

For Part (b) the statement about the convergence of $\alpha_n^\pm$ follows as before. For the other results, suppose first that $\mu(\{0\}) = 0$ and $\mu_n(\{0\}) = 0$ for all sufficiently large n. Recall that $\tau_\alpha = \inf\{u : W_u \leq \alpha^+(S_u) \text{ or } W_u \geq \alpha^-(I_u)\}$ and for $\eta > 0$ define the stopping time

$$\rho_{\alpha,\eta} = \tau_{\alpha_\eta}$$

where $\alpha^+_\eta(s) = \min\{\alpha^+(s), -\eta\}$ and $\alpha^-_\eta(i) = \max\{\alpha^-(i), \eta\}$.

We have that $\alpha_n \to \alpha$ at continuity points. Then, with $\alpha^\pm_{n,\eta} = \mp\max\{\mp\alpha_n^\pm(s), \eta\}$, $\alpha^\pm_{n,\eta} \to \alpha^\pm_\eta$ at continuity points and by the results of Part (a), $\rho_{\alpha_n,\eta} \to \rho_{\alpha,\eta}$ almost surely. In particular, given $\delta, \epsilon > 0$ there exists N_0 such that for all $n \geq N_0$

$$\mathbb{P}(|\rho_{\alpha_n,\eta} - \rho_{\alpha,\eta}| > \epsilon) < \delta/2.$$

Note that on $|W_{\tau_\alpha}| > \eta$ we have $\rho_{\alpha,\eta} = \tau_\alpha$ with a similar statement for α_n. We can choose $\eta > 0$ so that $\mu([-2\eta, 2\eta]) < \delta/6$ and then N_1 so that for $n \geq N_1$, $\mu_n([-\eta,\eta]) < \delta/3$. Then

$$(|\tau_{\alpha_n} - \tau_\alpha| > \epsilon) \subseteq (|W_{\tau_\alpha}| \leq \eta) \cup (|W_{\tau_{\alpha_n}}| \leq \eta) \cup (|\tau_{\alpha_n} - \tau_\alpha| > \epsilon, |W_{\tau_\alpha}| > \eta, |W_{\tau_{\alpha_n}}| > \eta)$$

and the set $(|\tau_{\alpha_n} - \tau_\alpha| > \epsilon)$ has probability at most δ.

It follows that $\tau_{\alpha_n} \to \tau_\alpha$ in probability, and hence that there is almost sure convergence down a subsequence. Furthermore, down the same subsequence $W_{\tau_{\alpha_n}} \to W_{\tau_\alpha}$ almost surely.

Now suppose that $\mu(\{0\}) = 0$ and that $\lim \mu_n(\{0\}) = 0$. Recall the definition of μ_n^* as the measure μ_n with probability mass at zero removed, and then rescaled to be a probability measure, and note that $\alpha_{\mu_n^*} \equiv \alpha_\mu$. Then also $\mu_n^* \Rightarrow \mu$ and $U_{\mu_n^*} \to U_\mu$ pointwise.

Then, $\tau_{\mu_n}^{P,Z_n} = 0$ for $Z_n \leq \mu_n(\{0\})$ and $\tau_{\mu_n}^{P,Z_n} = \tau_{\alpha_n}$ otherwise, so that $\tau_{\mu_n}^{P,Z_n} \to \tau_\alpha$ in probability. Moreover, down a subsequence, $\tau_{\mu_n}^{P,Z_n} \to \tau_\alpha$ almost surely.

It remains to consider the case where $\mu(\{0\}) > 0$. For $\epsilon < 1$, writing $A_n = (Z_n \leq \mu_n(\{0\}), Z > \mu(\{0\}))$ and $B_n = (Z_n > \mu_n(\{0\}), Z < \mu(\{0\}))$,

$$(|\tau_{\mu_n}^{P,Z_n} - \tau_\mu^{P,Z}| > \epsilon) \subseteq A_n \cup B_n \cup (Z_n > \mu_n(\{0\}), Z > \mu(\{0\}), |\tau_{\alpha_n} - \tau_\alpha| > \epsilon)$$

and $\tau_{\mu_n}^{P,Z_n} \to \tau_\mu^{P,Z}$ in probability. As before, there is almost sure convergence down a subsequence. □

Remark 3.3.7. One easy and natural way to guarantee that $Z_n \to Z$ is to take $Z_n = Z$ with probability one, or in other words to use the same independent randomisation variable for each embedding.

Remark 3.3.8. Suppose that μ is less than or equal to ν in convex order (we write $\mu \leq_{cx} \nu$). Then $U_\mu \leq U_\nu$. However, it does not follow that $\beta_\mu \geq \beta_\nu$, and so it does not follow that $\tau_\mu^{AY} \leq \tau_\nu^{AY}$. Similarly, we do not have that $|\alpha_\mu^\pm| \leq |\alpha_\nu^\pm|$ nor $\tau_\mu^P \leq \tau_\nu^P$.

Nonetheless, given μ it is possible to choose μ_n increasing to μ in convex order and such that the barycentres are decreasing and hence the stopping times $\tau_{\mu_n}^{AY}$ are monotonically increasing and converge to μ. This idea is used extensively in Azéma and Yor [7], see also Revuz and Yor [63, Section VI.5], and also below in the proof of Theorem 3.6.1.

Similar remarks apply for the Perkins embedding.

Example 3.3.9. *In Proposition 3.3.4 it does not hold that $\beta_n(s) \to \beta(s)$ for s beyond the upper limit on the support of μ.*

Suppose $\mu = \frac{1}{2}(\delta_1+\delta_{-1})$ *and* $\mu_n = (1-n^{-2})\frac{1}{2}(\delta_1+\delta_{-1})+n^{-2}\frac{1}{2}(\delta_n+\delta_{-n})$. *Then* $U_\mu(0) = 1$ *and* $U_n(0) = 1 + n^{-1} - n^{-2} \to 1$.

We have b_n *is piecewise constant, and* $b_n(x) = 0$ *for* $x < -n$ $b_n(x) = n/(2n^2 - 1)$ *for* $-n \le x < -1$, $b_n(x) = 1 + n^{-1} - n^{-2}$ *for* $-1 \le x < 1$ *and* $b_n(x) = n$ *for* $1 \le x < n$. *Then* $\beta_n(s) \to \beta_\infty(s)$ *where* $\beta_\infty(s) = -1$ *for* $s \le 1$ *and* $\beta_\infty(s) = 1$ *for* $s > 1$. *In contrast,* $\beta(s) = -1$ *for* $s < 1$ *and* $\beta(s) = s$ *for* $s \ge 1$.

Example 3.3.10. *If* $\alpha_n \to \alpha_\mu$, *but* $U_n(0) \not\to U_\mu(0)$ *then in general* $\mu_n \not\Rightarrow \mu$.

Suppose $\mu = p(\delta_1 + \delta_{-1}) + (1 - 2p)\delta_0$ *and* $\mu_n = q(\delta_1 + \delta_{-1}) + (1 - 2q)\delta_0$. *Then* $\alpha_n \equiv \alpha_\mu$ *but* $\mu_n \not\Rightarrow \mu$ *unless* $p = q$.

Example 3.3.11. *Suppose* $\alpha_n \to \alpha_\mu$ *at continuity points of* α_μ *and suppose* $U_n(0) \to U_\mu(0)$. *Then it does not follow that* τ_{α_n} *converges almost surely, although even then we may still have* $\mu_n \Rightarrow \mu$.

Suppose $\mu = \frac{1}{4}(\delta_1+\delta_{-1})+\frac{1}{2}\delta_0$ *and* $\mu_n = \frac{1}{4}(\delta_1+\delta_{-1})+\frac{1}{4}(\delta_{1/n}+\delta_{-1/n})$. *Then* $\alpha^{\pm}(x) = \mp 1$ *and* $\alpha_n^{\pm}(x) = \mp 1/n$ *for* $0 < |x| \le 1/(3n)$ *and* $\alpha_n^{\pm}(x) = \mp 1$ *for* $|x| > 1/(3n)$. *Further,* $\tau_\alpha = \inf\{u : |W_u| = 1\}$ *and*

$$\tau_{\alpha_n} = \begin{cases} H_{-1/n} & \text{if } H_{-1/n} < H_{1/3n}; \\ H_{1/n} & \text{if } H_{1/n} < H_{-1/3n}; \\ H_{-1} & \text{if } H_{-1/n} > H_{1/3n},\ H_{1/n} > H_{-1/3n} \text{ and } H_{-1} < H_1; \\ H_1 & \text{if } H_{-1/n} > H_{1/3n},\ H_{1/n} > H_{-1/3n} \text{ and } H_1 < H_{-1}. \end{cases}$$

Then, for almost every ω, $\tau_{\alpha_n}(\omega)$ *fails to converge, and there is both a subsequence converging to* 0, *and another subsequence converging to* $(H_{-1} \wedge H_1)(\omega)$. *(See Figure 3.3.)*

3.4 Target laws with bounded support

Our goal is to prove that for a suitable class of bivariate functions $F(w, s)$, the Azéma-Yor and Perkins embeddings, which are well known to maximise and minimise $\mathbb{E}[F(W_\tau, S_\tau)]$ in the special case where F does not depend on w and F is increasing in s, continue to maximise this quantity even when there is non-trivial dependence on w.

We are interested in functions F which are monotonic in the following sense (note our terminology does not require a function to be strictly increasing for it to be called increasing).

Assumption 3.4.1. Throughout we assume that $F : \{(w, s) \in \mathbb{R} \times \mathbb{R}_+; w \le s\} \to \mathbb{R}_+$ is a continuous function which is bounded on compact sets. We further assume that the partial derivative F_s exists and is bounded on compact sets.

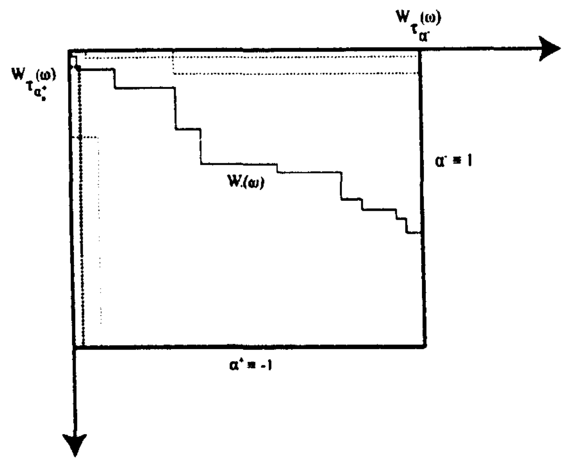

Figure 3.3: In Example 3.3.11, for almost every ω, $\tau_{\alpha_n}(\omega)$ fails to converge. For each ω, there are subsequences $n(k)$ for which the path is stopped at either ± 1 (for the ω in the path, they are stopped at $+1$) and there are other subsequences $\tilde{n}(k)$ for which the path is stopped when $|W|$ equals $1/\tilde{n}$, and so down these subsequences the stopping times converge to zero.

Definition 3.4.2.

F-MON↑ $F_s(w,s)/(s-w)$ is monotonic increasing in w.

F-MON↓ $F_s(w,s)/(s-w)$ is monotonic decreasing in w.

For $r \le \hat{x} \le \infty$ and $\eta \in \{\beta, \alpha^+\}$ define

$$\lambda_\eta(r) = \frac{F_s(\eta(r), r)}{r - \eta(r)},$$

$\Lambda_\eta(s) = \int_0^s \lambda_\eta(r)dr$ and $\Lambda_\eta^{(1)}(s) = \int_0^s r\lambda_\eta(r)dr$. Set $\bar{\Lambda}_\eta = \sup_{s<\hat{x}} |\Lambda_\eta(s)|$. Define $\Phi_\eta(w,s) = \int_0^s \lambda_\eta(r)(r-w)dr$; whence $\Phi_\eta(w,s) = \Lambda_\eta^{(1)}(s) - w\Lambda_\eta(s)$. Finally, define $\xi_\beta(w)$ by

$$\xi_\beta(w) = F(w, b(w)) - \Phi_\beta(w, b(w))$$

and $\xi_{\alpha^+}(w)$ by

$$\xi_{\alpha^+}(w) = F(w, \bar{a}(w)) - \Phi_\alpha(w, \bar{a}(w))$$

where $\bar{a}(w) = w$ for $w \ge 0$ and $\bar{a}(w) = a(w)$ for $w < 0$. Note that $\xi_\beta(w)$ (respectively $\xi_{\alpha^+}(w)$) does not depend on the convention chosen for $b(w)$ (respectively $a^+(w)$).

In this section we suppose μ has bounded support so that $\check{x}$ and $\hat{x}$ are finite. This assumption will be relaxed in the next section.

Both $\bar{\Lambda}_\beta$ and $\bar{\Lambda}_{\alpha^+}$ depend on the combination of μ and F. By Lemma 3.2.2 $(r-\alpha^+(r))^{-1}$ is integrable near zero so the fact that μ has bounded support is sufficient for $\bar{\Lambda}_{\alpha^+} < \infty$.

Furthermore, if μ has an atom at $\hat{x}$ then $(r-\beta(r))^{-1}$ is integrable near zero by Lemma 3.2.2 whence $r-\beta(r)$ is bounded below for $r < \hat{x}$. Hence, the fact that μ has bounded support, coupled with an atom at $\hat{x}$ is sufficient for $\bar{\Lambda}_\beta < \infty$.

Theorem 3.4.3. *Suppose F-MON↑ holds. Then*

$$\sup_{\tau \in \mathcal{S}_{UI}(W,\mu)} \mathbb{E}[F(W_\tau, S_\tau)] = \mathbb{E}[F(W_{\tau_\mu^{AY}}, S_{\tau_\mu^{AY}})], \tag{3.4.1}$$

$$\inf_{\tau \in \mathcal{S}_{UI}(W,\mu)} \mathbb{E}[F(W_\tau, S_\tau)] = \mathbb{E}[F(W_{\tau_\mu^P}, S_{\tau_\mu^P})]. \tag{3.4.2}$$

Remark 3.4.4. In the case where μ has no atoms (so that the argmin in (3.2.1) is strictly increasing and $\mathbb{E}[X|X \geq x] = \mathbb{E}[X|X > x]$) then we can write

$$\mathbb{E}[F(W_{\tau_\beta}, S_{\tau_\beta})] = \int_{\mathbb{R}} F(w, b_\mu(w))\mu(dw). \tag{3.4.3}$$

This formula need not hold if μ has atoms.

In cases where μ has a strictly positive density ρ on $(\check{x}, \hat{x})$ and β is differentiable, the expression in (3.4.3) can be rewritten as

$$\mathbb{E}[F(W_{\tau_\beta}, S_{\tau_\beta})] = \int_{\mathbb{R}} F(\beta(s), s)\,\mathbb{P}(S_{\tau_\beta} \in ds) = \int_{\mathbb{R}} F(\beta(s), s)\rho(\beta(s))\beta'(s)ds \tag{3.4.4}$$

where we use the fact that in the atom-free case

$$\mu([\beta(s), \infty)) = \mathbb{P}(W_{\tau_\beta} \geq \beta(s)) = \mathbb{P}(S_{\tau_\beta} \geq s).$$

A similar remark applies to $\mathbb{E}[F(W_{\tau_\mu^P}, S_{\tau_\mu^P})] = \int_{\mathbb{R}} F(w, \bar{a}(w))\mu(dw)$.

Remark 3.4.5. The requirement that the infimum in (3.4.2) is taken over $\tau \in \mathcal{S}_{UI}(W,\mu)$ (and not over all embeddings) is necessary, as can be seen by considering $F(w,s) = -(s-w)^3$. However, if we restrict attention to functions F which are increasing in s, then we may also replace the infimum in (3.4.2) with an infimum over all embeddings.

The key to the proof of the Theorem is the following lemma.

Lemma 3.4.6. *Suppose F satisfies F-MON↑. Then, for all $w \leq s$*

$$\xi_{\alpha^+}(w) + \Phi_{\alpha^+}(w,s) \leq F(w,s) \leq \xi_\beta(w) + \Phi_\beta(w,s)$$

with equality on the left at $w = s$ and $w = \alpha^+(w)$ and equality on the right at $w = \beta(s)$.

Proof:

For $\eta \in \{\beta, \alpha^+\}$ define

$$L_\eta(w,s) = \left[F(w,s) - \xi_\eta(w) - \int_0^s \lambda_\eta(r)(r-w)dr\right]. \tag{3.4.5}$$

We will show that $L_{\alpha^+}(w,s) \geq 0$ with equality at $w = s$ and $w = \alpha^+(s)$, and $L_\beta(w,s) \leq 0$ with equality at $w = \beta(s)$.

Consider the latter inequality first:

$$\begin{aligned}
L_\beta(w,s) &= F(w,s) - \xi_\beta(w) - \int_0^s \lambda_\beta(r)(r-w)dz \\
&= F(w,s) - F(w,b(w)) + \int_0^{b(w)} dr F_s(\beta(r),r)\frac{r-w}{r-\beta(r)} - \int_0^s dr F_s(\beta(r),r)\frac{r-w}{r-\beta(r)} \\
&= \int_{b(w)}^s \left\{\frac{F_s(w,r)}{r-w} - \frac{F_s(\beta(r),r)}{r-\beta(r)}\right\}(r-w)dr.
\end{aligned}$$

If $b(w) < r < s$, then since β is increasing, $w < \beta(r)$ and by F-MON$\uparrow$ the integrand is negative. If $s < r < b(w)$ then $w > b(r)$ and the integrand is positive. Thus $L_\beta(w,s) \leq 0$ as required. Clearly, there is equality at $s = b(w)$.

For L_{α^+} a similar calculation to the one above shows that

$$L_{\alpha^+}(w,s) = \int_{\bar{a}(w)}^s \left\{\frac{F_s(w,r)}{r-w} - \frac{F_s(\alpha^+(r),r)}{r-\alpha^+(r)}\right\}(r-w)dr.$$

To see that $L_{\alpha^+}(w,s) \geq 0$, consider $w \geq 0$ and $w < 0$ separately. For $w \geq 0$, $\bar{a}(w) = w$ and for $w < r < s$, $\alpha^+(r) \leq \alpha^+(w) \leq w$ so that the integrand is positive and $L_{\alpha^+}(w,s) \geq 0$. For $w < 0$, $\bar{a}(w) = a(w)$, and then if $a(w) < r < s$, we have $w > \alpha^+(r)$ and the integrand is positive. Otherwise if $s < r < a(w)$, $w < \alpha^+(r)$ and the integrand is negative. In either case, allowing for the limits on the integral, $L_{\alpha^+}(w,s) \geq 0$. Equality holds at $w = s$ and $w = \alpha^+(s)$. □

Remark 3.4.7. Essentially, the idea behind Lemma 3.4.6 and the proof of Theorem 3.4.3 is to interpret the embedding property and Doob's (in)-equality for the martingale W as linear constraints on the possible joint laws of (W_τ, S_τ), with associated Lagrange multipliers. Thus, if the joint law is given by $\nu(dw,ds)$, then $\int_{s\geq r}(w-r)\nu(dw,ds) = 0$ (which is equivalent to (3.2) in Rogers [66]). There is an identity of this form for each r and when they are integrated against a family of Lagrange multipliers $\lambda_\eta(r)$ we obtain

$$0 = \int_0^\infty \lambda_\eta(r) \int_{s\geq r}(w-r)\nu(dw,ds) = \int \nu(dw,ds) \int_{0\leq r\leq s} \lambda_\eta(r)(w-r)dr.$$

The integrand of this last expression appears as the last term in (3.4.5).

It remains to prove Theorem 3.4.3. The main idea for the proof of the Theorem is that for $\tau \in \mathcal{S}_{UI}(W,\mu)$ both $(\Phi_\alpha(W_t^\tau, S_t^\tau))_{t\geq 0}$ and $(\Phi_\beta(W_t^\tau, S_t^\tau))_{t\geq 0}$ are uniformly integrable martingales. (By Itô's formula, $d\Phi_\eta(W_t,S_t) = -\Lambda_\eta(S_t)dW_t$ since the finite variation term involves the product $(S_t - W_t)dS_t$ and when S is increasing we must also have $S_t - W_t = 0$.) It follows that $\mathbb{E}[\Phi_\beta(W_\tau, S_\tau)] = 0$ and

$$\mathbb{E}[\xi_\alpha(W_\tau)] \leq \mathbb{E}[F(W_\tau, S_\tau)] \leq \mathbb{E}[\xi_\beta(W_\tau)]$$

which, given the forms of ξ_α and ξ_β leads to the first result given in the introduction.

Remark 3.4.8. Both $(\Phi_\alpha(W_t^\tau, S_t^\tau))_{t\geq 0}$ and $(\Phi_\beta(W_t^\tau, S_t^\tau))_{t\geq 0}$ belong to the class of so-called Azéma martingales. A martingale $M = (M_t)_{t\geq 0}$ is an Azéma (or Azéma-Yor) martingale if $M_t = G(S_t^X) - (S_t^X - X_t)g(S_t)$ for X a martingale and $G' = g$, see [7].

Proof of Theorem 3.4.3. Consider the first bound associated with the Azéma-Yor embedding and suppose that μ has an atom at $\hat{x}$. Since $\tau \in \mathcal{S}_{UI}(W,\mu)$ implies $(W_t^\tau)_{t\geq 0}$ is bounded, and since $\Lambda_\beta(s)$ and $\Lambda^{(1)}(s)$ are bounded, we have that $\Phi_\beta(W_t^\tau, S_t^\tau)$ is a bounded local martingale and hence $\mathbb{E}[\Phi_\beta(W_t^\tau, S_t^\tau)] = 0$, which can be re-expressed as $\mathbb{E}[\Lambda_\beta^{(1)}(S_\tau)] = \mathbb{E}[W_\tau \Lambda_\beta(S_\tau)]$.

In view of Lemma 3.4.6 we have

$$F(W_\tau, S_\tau) \leq \xi_\beta(W_\tau) + \Phi_\beta(W_\tau, S_\tau). \tag{3.4.6}$$

Thus

$$\mathbb{E}[F(W_\tau, S_\tau)] \leq \int \xi_\beta(w)\mu(dw).$$

Note that for $\tau = \tau_\beta$ we have equality in (3.4.6) and hence equality in this last expression.

Now suppose there is no atom at $\hat{x}$. Fix $\tau \in \mathcal{S}_{UI}(W,\mu)$ and let $\sigma_n = \tau \wedge H_{\hat{x}-1/n}$ and $\mu_n = \mathcal{L}(W_{\sigma_n})$. Then $U_{\mu_n} \to U_\mu$ for each x and by bounded convergence we have both

$$\mathbb{E}[F(W_\tau, S_\tau)] = \mathbb{E}[\lim F(W_{\sigma_n}, S_{\sigma_n})] = \lim \mathbb{E}[F(W_{\sigma_n}, S_{\sigma_n})]$$

and

$$\mathbb{E}[F(W_{\tau_\mu^{AY}}, S_{\tau_\mu^{AY}})] = \mathbb{E}[\lim F(W_{\tau_{\mu_n}^{AY}}, S_{\tau_{\mu_n}^{AY}})] = \lim \mathbb{E}[F(W_{\tau_{\mu_n}^{AY}}, S_{\tau_{\mu_n}^{AY}})].$$

The result follows from the previous case on comparing σ_n with $\tau_{\mu_n}^{AY}$. The proof for the lower bound is identical except there is no need to treat the case where there is an atom at $\hat{x}$ separately since $(r - \alpha^+(r))^{-1}$ is integrable near zero regardless. □

There are a parallel pair of results based on F-MON↓, the proofs of which are very similar.

Lemma 3.4.9. *Suppose F satisfies F-MON$\downarrow$. Then, for all $w \leq s$*

$$\xi_\beta(w) + \Phi_\beta(w,s) \leq F(w,s) \leq \xi_\alpha(w) + \Phi_\alpha(w,s)$$

with equality on the right at $s = w$ and $s = a(w)$ and equality on the left at $s = b(w)$.

Theorem 3.4.10. *Suppose F-MON$\downarrow$ holds. Then*

$$\inf_{\tau \in S_{UI}(W,\mu)} \mathbb{E}[F(W_\tau, S_\tau)] = \mathbb{E}[F(W_{\tau_\mu^{AY}}, S_{\tau_\mu^{AY}})],$$

$$\sup_{\tau \in S_{UI}(W,\mu)} \mathbb{E}[F(W_\tau, S_\tau)] = \mathbb{E}[F(W_{\tau_\mu^P}, S_{\tau_\mu^P})].$$

Example 3.4.11. *Suppose $\mu = U[-1,1]$ and $F(w,s) = (s-w)^c$ for $c > -1$ (with $c \neq 0$). Then for $c \geq 2$ F-MON$\downarrow$ holds, for $0 < c \leq 2$ F-MON$\uparrow$ holds, and for $-1 < c < 0$, F-MON$\downarrow$ holds again. Note that Assumption 3.4.1 is not satisfied for $-1 < c < 1$. For c in this range and $\epsilon > 0$ let $F_\epsilon(w,s) = ((s-w) \vee \epsilon)^c$. The arguments of Theorem 3.4.3 provide the upper and lower bounds for F_ϵ and letting $\epsilon \downarrow 0$ we obtain the bounds for F.*

Write B^{AY} and B^P for the bounds associated with the Azéma-Yor and Perkins embeddings.

Recall the expressions for β and α from Examples 3.2.1 and 3.2.3. For the Azéma-Yor embedding, $\beta(s) = 2s - 1$ and the law of the S_{τ_β} is a uniform on $[0,1]$. The associated bound (as a function of the parameter c) is given by

$$B^{AY}(c) = \mathbb{E}[F(W_{\tau_\mu^{AY}}, S_{\tau_m^{AY}u})] = \int_{-1}^{1} (b(w)-w)^c \frac{dw}{2} = \int_0^1 (s-\beta(s))^c ds = \int_0^1 (1-s)^c ds = \frac{1}{c+1}.$$

For the Perkins bound, note that for $c < 0$, $F(s,s) = \infty$, and so $B^P(c) = 0$. For $c > 0$, $F(s,s) = 0$ and using the substitution $w = \alpha^+(s) = s - 2\sqrt{s}$,

$$\begin{aligned} B^P(c) = \mathbb{E}[F(W_{\tau_\mu^P}, S_{\tau_\mu^P})] &= \int_{-1}^{0} (a^+(w) - w)^c \frac{dw}{2} \\ &= \frac{2^c}{(c+1)(c+2)}. \end{aligned}$$

Note that for $c = 2$, $B^{AY}(2) = B^P(2) = 1/3$ and all uniformly integrable embeddings for the terminal law are consistent with the same expected payoff. The reason for this will become clear in Section 3.6 and will correspond to the choice $g \equiv 1$. (See Figure 3.4.)

Example 3.4.12. *Suppose again that $\mu = U[-1,1]$. Let $F(w,s) = \frac{(s-w)^2}{s^c}$. Note that for each c either F-MON$\uparrow$ or F-MON$\downarrow$ (or both) holds, so that the Azéma-Yor and Perkins embeddings*

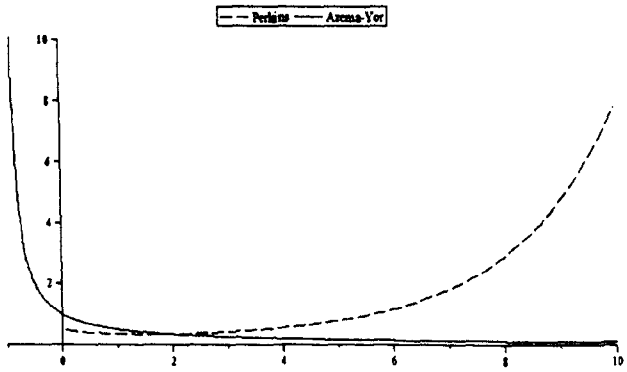

Figure 3.4: All uniformly integrable embeddings have the same expected value when $c = 2$. Note the reversal of the bounds at $c = 2$: for $0 < c < 2$ Theorem (3.4.3) applies while for $c > 2$ Theorem (3.4.10) applies. For $c < 0$, the Perkins bound is infinite and the Azéma-Yor bound is finite. The Perkins bound as a function of c is discontinuous at $c = 0$.

give extremal values for $\mathbb{E}[F(W_T, S_T)]$. *Consider the Azéma-Yor bound as a function of the parameter* c *(defined for* $c < 1$*):*

$$B^{AY}(c) = \int_{-1}^{1} \frac{(b(w) - w)^2}{b(w)^c} \frac{dw}{2} = \int_0^1 \frac{(s - \beta(s))^2}{s^c} ds = \int_0^1 \frac{(s-1)^2}{s^c} ds = \frac{2}{(1-c)(2-c)(3-c)}.$$

For the Perkins bound we have (for $c < 3/2$*)*

$$\begin{aligned} B^P(c) &= \int_{-1}^{0} \frac{(a^+(w) - w)^2}{a^+(w)} \frac{dw}{2} \\ &= \int_0^1 \frac{2\sqrt{s}}{s^c}(1 - \sqrt{s}) ds \\ &= \frac{1}{(3/2 - c)(2 - c)}. \end{aligned}$$

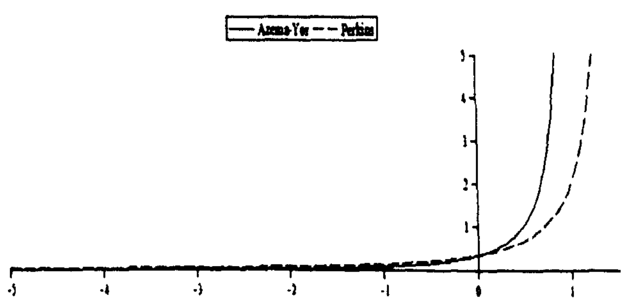

Figure 3.5: For $1 < c < 3/2$ the Azéma-Yor upper bound is infinite while the Perkins lower bound is finite.

3.5 General centred target measures

Theorem 3.5.1. *Fix $\tau \in \mathcal{S}_{UI}(W,\mu)$. Suppose, in addition to Assumption 3.4.1, that $F \geq 0$, that*

$$\mathbf{E}\left[F(W_{H_{\pm n}}, S_{H_{\pm n}}); \tau \geq H_{\pm n}\right] \to 0, \tag{3.5.1}$$

and that if $(\mu_n)_{n\geq 1}$ is any sequence of measures which is increasing in convex order for which $\mu_n \Rightarrow \mu$, $U_{\mu_n}(0) \to U_\mu(0)$ and $\mu_n(\{0\}) \to \mu(\{0\})$ then both

$$\mathbf{E}\left[F(W_{\tau_{\mu_n}^{AY}}, S_{\tau_{\mu_n}^{AY}})\right] \to \mathbf{E}\left[F(W_{\tau_{\mu}^{AY}}, S_{\tau_{\mu}^{AY}})\right] \text{ and} \tag{3.5.2}$$

$$\mathbf{E}\left[F(W_{\tau_{\mu_n}^{P}}, S_{\tau_{\mu_n}^{P}})\right] \to \mathbf{E}\left[F(W_{\tau_{\mu}^{P}}, S_{\tau_{\mu}^{P}})\right]. \tag{3.5.3}$$

Then, if F satisfies F-MON$\uparrow$

$$\mathbf{E}\left[F(W_{\tau_{\mu}^{P}}, S_{\tau_{\mu}^{P}})\right] \leq \mathbf{E}\left[F(W_\tau, S_\tau)\right] \leq \mathbf{E}\left[F(W_{\tau_{\mu}^{AY}}, S_{\tau_{\mu}^{AY}})\right],$$

whereas, if F satisfies F-MON$\downarrow$ then

$$\mathbf{E}\left[F(W_{\tau_{\mu}^{AY}}, S_{\tau_{\mu}^{AY}})\right] \leq \mathbf{E}\left[F(W_\tau, S_\tau)\right] \leq \mathbf{E}\left[F(W_{\tau_{\mu}^{P}}, S_{\tau_{\mu}^{P}})\right].$$

Proof Suppose F-MON$\uparrow$ holds (the proof for F-MON$\downarrow$ is similar). Given $\tau \in \mathcal{S}_{UI}(W,\mu)$, let $\sigma_n = \tau \wedge H_{\pm n}$, $\mu_n = \mathcal{L}(W_{\sigma_n})$ and define $\tau_{\mu_n}^{AY}$ and $\tau_{\mu_n}^{P}$ to be the Azéma-Yor and Perkins stopping times associated with μ_n.

We have, using monotone convergence, (3.5.1), Theorem 3.4.3 and finally (3.5.2),

$$\begin{aligned}
\mathbf{E}\left[F(W_\tau, S_\tau)\right] &= \mathbf{E}\left[\lim F(W_{\sigma_n}, S_{\sigma_n}); \sigma_n = \tau \leq H_{\pm n}\right] \\
&= \lim \mathbf{E}\left[F(W_{\sigma_n}, S_{\sigma_n}) I_{\{\tau \leq H_{\pm n}\}}\right] \\
&= \lim \mathbf{E}\left[F(W_{\sigma_n}, S_{\sigma_n}) I_{\{\tau < H_{\pm n}\}} + F(W_{H_{\pm n}}, S_{H_{\pm n}}) I_{\{\tau \geq H_{\pm n}\}}\right] \\
&= \lim \mathbf{E}\left[F(W_{\sigma_n}, S_{\sigma_n})\right] \\
&\leq \lim \mathbf{E}\left[F(W_{\tau_{\mu_n}^{AY}}, S_{\tau_{\mu_n}^{AY}})\right] = \mathbf{E}\left[F(W_{\tau_{\mu}^{AY}}, S_{\tau_{\mu}^{AY}})\right].
\end{aligned}$$

Similarly

$$\lim \mathbf{E}\left[F(W_{\sigma_n}, S_{\sigma_n})\right] \geq \lim \mathbf{E}\left[F(W_{\tau_{\mu_n}^{P}}, S_{\tau_{\mu_n}^{P}})\right] = \mathbf{E}\left[F(W_{\tau_{\mu}^{P}}, S_{\tau_{\mu}^{P}})\right].$$

□

Corollary 3.5.2. *Suppose that $F(w,s) \leq A(1 + |w|^k + s^k)$ for $k \geq 1$ and that μ has finite $k + \epsilon$*

moment, for some positive ϵ. Then the hypotheses (3.5.1), (3.5.2) *and* (3.5.3) *are all satisfied, and the conclusions of Theorem 3.5.1 hold.*

Proof By Doob's submartingale inequality for $(|W_{t\wedge\tau}|^{k+\epsilon})_{t\geq 0}$, for any $\tau \in \mathcal{S}_{UI}(W,\mu)$

$$m^{k+\epsilon}\,\mathbb{P}(\tau > H_{\pm m}) < \mathbb{E}[|W_\tau|^{k+\epsilon}] < \infty.$$

Then

$$\mathbb{E}\left[F(W_{H_{\pm n}}, S_{H_{\pm n}}); \tau \geq H_{\pm n}\right] \leq A(1+2n^k)\,\mathbb{P}(\tau > H_{\pm n}) \to 0.$$

For (3.5.2) we have that $\tau_{\beta_n} \to \tau_\beta$ almost surely. Moreover, since $\mu_n \leq_{cx} \mu$ there exists a stopping time (ρ_n say) with $\rho_n \geq \tau_{\beta_n}$ and $\rho_n \in \mathcal{S}_{UI}(W,\mu)$. For such a ρ_n, $\mathbb{E}[|W_{\rho_n}|^{k+\epsilon}] = \int_{\mathbb{R}} |x|^{k+\epsilon}\mu(dx) < \infty$ by hypothesis, and then (letting W^* denote the running maximum of $|W|$) by Doob's L^p inequality $\mathbb{E}[(W^*_{\rho_n})^{k+\epsilon}] \leq D < \infty$ for some constant D, independent of n.

Set $F_n = F(W_{\tau^{AY}_{\mu_n}}, S_{\tau^{AY}_{\mu_n}})$ and $F = F(W_{\tau^{AY}_{\mu}}, S_{\tau^{AY}_{\mu}})$, then $F_n \to F$ almost surely. The goal is to show that $\mathbb{E}[F_n] \to \mathbb{E}[F]$ which will follow if $\sup_n \mathbb{E}[(F_n)^p] < \infty$ for then $(F_n)_{n\geq 1}$ is uniformly integrable. We have that if $|w| \leq x$ and $s \leq x$ then with $p = 1 + k/\epsilon$,

$$F(w,s)^p \leq A^p(1+2x^k)^p \leq A^p 3^p(1+x^{kp}).$$

Hence

$$\mathbb{E}\left[F_n^p\right] \leq A^p 3^p(1+\mathbb{E}[(W^*_{\tau_n})^{kp}]) \leq A^p 3^p(1+\mathbb{E}[(W^*_{\rho_n})^{kp}]) \leq A^p 3^p(1+D) < \infty.$$

For (3.5.3), consider a subsequence $n(k)$. Then down a further subsequence $\tau^P_{\mu_n} \to \tau^P_\mu$ almost surely and down this subsequence (3.5.3) holds by identical arguments as in the case for the Azéma-Yor embedding. Hence (3.5.3) holds. □

3.6 Objective functions as running costs

Our original aim in studying functions $F(w,s)$ was as an aid in the analysis of the expected values of integrals of the form $\int_0^\tau g(S_t)dt$. Motivated by a problem in mathematical finance we asked:

> Given g and μ, what is the range of possible values of $\mathbb{E}[\int_0^\tau g(S_u)du]$ over embeddings τ of μ in Brownian motion.

Our aim is to reduce this problem to the case previously considered, but to use the extra structure to prove more powerful results under weaker hypotheses.

The expected value of $\int_0^\tau g(S_u)du$ is intimately related to the value of $\mathbb{E}[G(W_\tau, S_\tau)]$ where $G(w,s) = (s-w)^2 g(s)$. Indeed, if g is differentiable, then by Itô's Lemma,

$$G(W_\tau, S_\tau) = G(0,0) + \int_0^\tau g(S_u)du - \int_0^\tau 2(S_u - W_u)g(S_u)dW_u, \tag{3.6.1}$$

so that if $g(0)$ is finite (and then $G(0,0) = 0$), and if $(\int_0^{\tau\wedge t} 2(S_u - W_u)g(S_u)dW_u)_{t\geq 0}$ is a uniformly integrable martingale, then $\mathbb{E}[\int_0^\tau g(S_u)du] = \mathbb{E}[G(W_\tau, S_\tau)]$.

If g is increasing (respectively decreasing) then G satisfies G-MON$\downarrow$ (respectively G-MON$\uparrow$) and we can apply the results of previous sections to deduce that the Azéma-Yor and Perkins solutions give bounds $\mathbb{E}[\int_0^\tau g(S_u)du]$ over embeddings τ of μ.

Theorem 3.6.1. *Suppose $g : \mathbb{R}_+ \mapsto \mathbb{R}_+$ is a positive function and that μ is centred.*

1. *Suppose g is increasing. Then,*

$$\inf_{\tau\in\mathcal{S}(W,\mu)} \mathbf{E}\left[\int_0^\tau g(S_u)du\right] = \mathbf{E}\left[\int_0^{\tau_\mu^{AY}} g(S_u)du\right]$$

and

$$\sup_{\tau\in\mathcal{S}_{UI}(W,\mu)} \mathbf{E}\left[\int_0^\tau g(S_u)du\right] = \mathbf{E}\left[\int_0^{\tau_\mu^{P}} g(S_u)du\right].$$

2. *Suppose g is decreasing. Then,*

$$\inf_{\tau\in\mathcal{S}(W,\mu)} \mathbf{E}\left[\int_0^\tau g(S_u)du\right] = \mathbf{E}\left[\int_0^{\tau_\mu^{P}} g(S_u)du\right]$$

and

$$\sup_{\tau\in\mathcal{S}_{UI}(W,\mu)} \mathbf{E}\left[\int_0^\tau g(S_u)du\right] = \mathbf{E}\left[\int_0^{\tau_\mu^{AY}} g(S_u)du\right].$$

Remark 3.6.2. As we remarked in the introduction, at first sight this result is counter-intuitive. Given increasing g, the Azéma-Yor stopping time maximises $\mathbf{E}[g(S_\tau)]$ over $\tau \in \mathcal{S}_{UI}(W,\mu)$, and it seems plausible that it might also maximise $\mathbf{E}[\int_0^\tau g(S_u)du]$. In fact the exact opposite is true. The explanation is that for the Azéma-Yor embedding there is co-monotonicity between S_τ and W_τ, and conditional on $S_\tau \geq s$, the stopping time occurs quite soon (and certainly before W drops below $\beta(s)$), whereas for the Perkins embedding, conditional on $S_\tau \geq s$, there are paths which will only be stopped when W goes below $\alpha^+(s)$. Thus, for increasing g when we wish to maximise the time (before τ) for which S is large, this is best achieved by the Perkins embedding: although relatively few paths will have large S (most will have already been stopped) those with a large maximum will spend a long time after first hitting s before being stopped.

Example 3.6.3. *Recall Example 3.4.11 Suppose $\mu = U[-1,1]$ and $g(s) = s^{-c}$. Then, for $c < 0$, $(1-c)^{-1}(2-c)^{-1}(3-c)^{-1} \leq \mathbb{E}[\int_0^\tau S_u^{-c} du] \leq (2-c)^{-1}(3/2-c)^{-1}$.*

For $0 < c < 1$, $(2-c)^{-1}(3/2-c)^{-1} \leq \mathbb{E}[\int_0^\tau S_u^{-c} du] \leq (1-c)^{-1}(2-c)^{-1}(3-c)^{-1}$, for $1 \leq c < 3/2$, $(2-c)^{-1}(3/2-c)^{-1} \leq \mathbb{E}[\int_0^\tau S_u^{-c} du] \leq \infty$, and for $c \geq 3/2$, $\mathbb{E}[\int_0^\tau S_u^{-c} du] = \infty$ for all embeddings τ.

Note that for $c = 0$, $\mathbb{E}[\tau]$ is independent of τ and equal to the variance of μ.

Example 3.6.4. *Recall the calculations from Example 3.2.4. Let the target law μ with support $[-1,\infty)$ satisfy $\mu(dx) = \frac{2}{(x+2)^3}dx$. Let $g(s) = \frac{1}{c+s}$ for $c > 0$ which is decreasing in s.*

The Azéma-Yor upper bound can be calculated explicitly to be

$$\begin{aligned} B^{AY}(c) &= \int_{-1}^{\infty} \frac{(b(w)-w)^2}{b(w)+c} \frac{2}{(w+2)^3} dw \\ &= \frac{2(\log(c)-\log(2))}{c-2}. \end{aligned}$$

The expression for the Perkins lower bound is given by

$$B^P(c) = \int_{-1}^{\infty} \frac{(a^+(w)-w)^2}{a^+(w)+c} \frac{2}{(w+2)^3}.dw$$

The expression for α^+ is too complicated for the expression above to have an analytic representation. However, the values can be computed numerically for different c.

The rest of this section is devoted to a proof of Theorem 3.6.1. We split the proof into four separate parts.

Proof of Theorem 3.6.1(i): Lower bound

Suppose first that g is monotonic increasing and that we are interested in minimising the quantity $\mathbb{E}[\int_0^\tau g(S_u)du]$ over embeddings τ of μ in W. Note that it is sufficient to restrict attention to $\mathcal{S}_{UI}(W,\mu)$: for non-minimal $\tau \in \mathcal{S}(W,\mu)$ there exists $\tilde{\tau} \leq \tau$ with $\tilde{\tau} \in \mathcal{S}_{UI}(W,\mu)$, and then $\int_0^\tau g(S_u)du \geq \int_0^{\tilde{\tau}} g(S_u)du$ for each $\omega \in \Omega$.

Suppose temporarily that g is bounded and continuously differentiable. Later we will relax this assumption. Then $G(w,s) = (s-w)^2 g(s)$ satisfies G-MON$\downarrow$.

Let F_μ denote the distribution function of μ and let $k_n = F_\mu^{-1}(1-1/n)$. For $\tau \in \mathcal{S}_{UI}(W,\mu)$ let $\sigma_n = \tau \wedge H_{\pm k_n}$, let $\mu_n = \mathcal{L}(W_{\sigma_n})$, β_n be the inverse barycentre of μ_n and finally let $\tau_{\mu_n}^{AY}$ be the Azéma-Yor stopping rule associated with the law μ_n so that $\tau_{\mu_n}^{AY} = \tau_{\beta_n} = \inf\{u : W_u \leq \beta_n(S_u)\}$. Then, by Proposition 3.3.4, since $U_{\mu_n} \uparrow U_\mu$, $\tau_{\beta_n} \to \tau_\beta$ almost surely.

If a stopping time ρ is such that $\rho \leq H_{\pm k_n}$ then $\mathbb{E}[\rho] < \infty$ and for $u \leq \rho$, $(S_u - W_u)g(S_u)$ is bounded. Then, if $M_t = \int_0^t (S_u - W_u)g(S_u)dW_u$ we have that $(M_t^\rho)_{t\geq 0}$ is an L^2 bounded martingale for which

$$\mathbf{E}[M^{\rho}_{\infty}] = \mathbf{E}\left[\int_0^{\rho} (S_u - W_u) g(S_u) dW_u\right] = 0. \tag{3.6.2}$$

It follows that

$$\begin{aligned}
\mathbf{E}\left[\int_0^{\sigma_n} g(S_u)du\right] &= \mathbf{E}\left[(S_{\sigma_n} - W_{\sigma_n})^2 g(S_{\sigma_n})\right] \\
&\geq \mathbf{E}\left[(S_{\tau_{\beta_n}} - W_{\tau_{\beta_n}})^2 g(S_{\tau_{\beta_n}})\right] \\
&= \mathbf{E}\left[\int_0^{\tau_{\beta_n}} g(S_u)du\right],
\end{aligned}$$

where we have used (3.6.1) and (3.6.2) twice and Theorem 3.4.10. Then it follows from the Fatou Lemma that

$$\lim \mathbf{E}\left[\int_0^{\sigma_n} g(S_u)du\right] \geq \lim \mathbf{E}\left[\int_0^{\tau_{\beta_n}} g(S_u)du\right] \geq \mathbf{E}\left[\liminf \int_0^{\tau_{\beta_n}} g(S_u)du\right] \tag{3.6.3}$$

and by monotone convergence and the fact that $\tau_{\beta_n} \to \tau_\beta$ almost surely, $\mathbf{E}\left[\int_0^{\tau} g(S_u)du\right] \geq \mathbf{E}\left[\int_0^{\tau_\beta} g(S_u)du\right]$ as required.

Finally we remove the temporary assumptions on g. Given g is monotonic increasing we can find an increasing sequence of bounded continuously differentiable (increasing) functions g_m which approximate g from below. Then, by monotone convergence

$$\mathbf{E}\left[\int_0^{\tau} g(S_u)du\right] = \lim_m \mathbf{E}\left[\int_0^{\tau} g_m(S_u)du\right] \geq \lim_m \mathbf{E}\left[\int_0^{\tau_\beta} g_m(S_u)du\right] = \mathbf{E}\left[\int_0^{\tau_\beta} g(S_u)du\right].$$

Note that this same argument will apply in all four parts of Theorem 3.6.1, and henceforth without loss of generality we will assume that g is continuously differentiable and bounded by $\bar{g}$.

Proof of Theorem 3.6.1(ii): Lower bound

Case 1: There exists an open interval $I \subseteq [-1,1]$ containing 0 with $\mu(I) = 0$.

Given $\tau \in \mathcal{S}(W,\mu)$, let $\sigma_m = \tau \wedge H_{\pm m}$. Let $\mu_m = \mathcal{L}(W_{\sigma_m})$. Write τ^P_m for the Perkins embedding of μ_m. Note that $\mu_m \Rightarrow \mu$, $U_{\mu_m}(0) \to U_\mu(0)$ and $\mu_m(I) = 0$. Then, $\tau^P_m = \tau_{\alpha_m}$ and by Proposition 3.3.6(a), $\tau_{\alpha_m} \to \tau_\alpha$ almost surely. Then exactly as in (3.6.3), but now using Theorem 3.4.3 to give that the lower bound is attained by the Perkins embedding, we conclude that $\mathbf{E}\left[\int_0^{\tau} g(S_u)du\right] \geq \mathbf{E}\left[\int_0^{\tau^P_\mu} g(S_u)du\right]$.

Case 2: General μ.

Given any subsequence, by Proposition 3.3.6(b) we may take a further subsequence down which $\tau^P_m \to \tau^P$ almost surely. Then down this subsequence the result holds, as in Case 1. Since the first subsequence was arbitrary we are done.

Proof of Theorem 3.6.1(ii): Upper bound

Now consider the upper bound in Theorem 3.6.1(ii). Rather than attempting to find a dominating random variable which will allow us to use the Reverse Fatou Lemma in place of the Fatou Lemma above we will use a slightly different approach based on defining a sequence of intermediate stopping times.

Let τ be any element of $\mathcal{S}_{UI}(W,\mu)$. Suppose g is bounded, continuously differentiable and monotonic decreasing, and that μ has support in a bounded interval $[\check{x},\hat{x}]$. Then, as above, $\mathbb{E}[\int_0^\tau g(S_u)du] = \mathbb{E}[G(W_\tau, S_\tau)]$. Moreover, we can conclude from Theorem 3.4.3 that

$$\sup_{\tau\in\mathcal{S}_{UI}(W,\mu)} \mathbb{E}\left[\int_0^\tau g(S_u)du\right] = \mathbb{E}\left[\int_0^{\tau_\beta} g(S_u)du\right].$$

It remains to remove the assumptions on μ.

Given ϵ, let $U_\epsilon(x) = \max\{U_\mu(x)-\epsilon, |x|\}$ and let $\check{x}_\epsilon$ and $\hat{x}_\epsilon$ be the left and right-hand endpoints of the interval $I_\epsilon = \{x : U_\epsilon(x) > |x|\}$.

Let $\sigma_\epsilon = \tau \wedge \inf\{u : W_u \notin I_\epsilon\}$. Let $\bar{\mu}_\epsilon$ be the law of W_{σ_ϵ} and let $\bar{U}_\epsilon$ be the associated potential. Then $\bar{U}_\epsilon = U_\epsilon$ on I_ϵ^c and $U_\epsilon \le \bar{U}_\epsilon \le U_\mu$.

Now let $\tilde{U}_\epsilon$ be the largest convex function such that $\tilde{U}_\epsilon(x) = |x|$ on I_ϵ^c and $\tilde{U}_\epsilon \le U_\mu$. It follows that $\tilde{U}_\epsilon$ is actually equal to U on an interval $\tilde{I}_\epsilon = [\tilde{c}_\epsilon, \tilde{d}_\epsilon]$. If ϵ is small enough then $0 \in \tilde{I}_\epsilon$. See Figure 3.6. Further, $U_\epsilon \le \bar{U}_\epsilon \le \tilde{U}_\epsilon \le U$ and in terms of the associated measures $\mu_\epsilon \le_{cx} \bar{\mu}_\epsilon \le_{cx} \tilde{\mu}_\epsilon \le_{cx} \mu$, where $\tilde{\mu}_\epsilon$ is such that $U_{\tilde{\mu}_\epsilon} = \tilde{U}_\epsilon$ and we recall that $\le_{cx}$ denotes 'less than or equal to in convex order'. Then, by a theorem of Strassen [73] (or for a more explicit construction in our context, Chacon and Walsh [19]), given σ_ϵ there exists a stopping time $\tilde{\sigma}_\epsilon$ such that $\sigma_\epsilon \le \tilde{\sigma}_\epsilon$ almost surely, and $\tilde{\mu}_\epsilon = \mathcal{L}(W_{\tilde{\sigma}_\epsilon})$.

Now consider a sequence ϵ_n decreasing to zero. Let $\tilde{\beta}_{\epsilon_n}$ be the inverse barycentre associated with $\tilde{\mu}_{\epsilon_n}$ and let $\tilde{\tau}_n$ be the Azéma-Yor stopping time associated with $\tilde{\beta}_{\epsilon_n}$. The introduction of the stopping times $\tilde{\sigma}_{\epsilon_n}$ gives extra structure which means that not only do the barycentres converge (as in Proposition 3.3.4), but also that they converge monotonically.

Lemma 3.6.5. *$\tilde{\beta}_n \downarrow \beta$ and $\tilde{\tau}_n \uparrow \tau_\beta$ almost surely.*

Proof: Write $\check{x}_n$ (respectively $\hat{x}_n, c_n, d_n$) for $\check{x}_{\epsilon_n}$ (respectively $\hat{x}_{\epsilon_n}, c_{\epsilon_n}, d_{\epsilon_n}$).

Then, for $s \le b(c_n)$, $\tilde{\beta}_n(s) = \check{x}_n \ge \beta(s)$, for $b(\tilde{c}_n) < s < \hat{x}_n$, $\tilde{\beta}_n(s) = \beta(s)$ and for $s \ge \hat{x}_n$, $\tilde{\beta}_n(s) = s \ge \beta(s)$.

Monotonicity in n of $\tilde{\tau}_n$ follows immediately. □

It follows from the results for bounded target distributions that

$$\mathbb{E}\left[\int_0^{\sigma_n} g(S_u)du\right] \le \mathbb{E}\left[\int_0^{\tilde{\sigma}_n} g(S_u)du\right] = \mathbb{E}\left[G(W_{\tilde{\sigma}_n}, S_{\tilde{\sigma}_n})\right] \le \mathbb{E}\left[G(W_{\tilde{\tau}_n}, S_{\tilde{\tau}_n})\right] = \mathbb{E}\left[\int_0^{\tilde{\tau}_n} g(S_u)du\right].$$

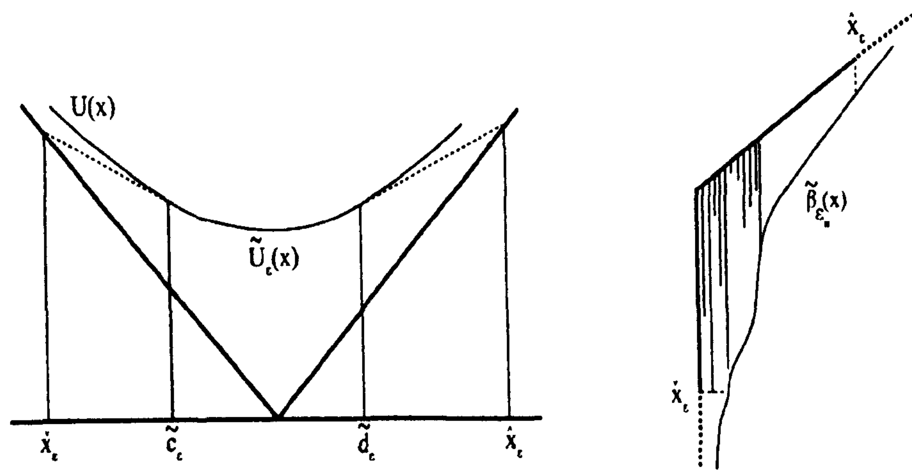

Figure 3.6: The potentials $\tilde{U}_\epsilon$ increase monotonically as ϵ decreases. Moreover, over a range of x, depending on ϵ_n, we have $\tilde{\beta}_{\epsilon_n}(x) \equiv \beta(x)$, and hence, the inverse barycentre functions converge monotonically.

We have that the integral inside the first expectation converges monotonically to $\int_0^\tau g(S_u)du$, whereas the integral inside the final expression converges monotonically to $\int_0^{\tau_\beta} g(S_u)du$. Hence $\mathbf{E}[\int_0^\tau g(S_u)du] \le \mathbf{E}[\int_0^{\tau_\beta} g(S_u)du]$ as required.

Proof of Theorem 3.6.1(i): Upper bound

The final element of Theorem 3.6.1 is the upper bound in the case of monotonically increasing g. Recall that we suppose that g is continuously differentiable, and bounded by $\bar{g}$.

If μ has bounded support then Theorem 3.4.10 applies directly, so we assume that the support of μ is unbounded.

If $\mu \notin L^2$ then for each $\tau \in \mathcal{S}(W,\mu)$ we have $\mathbf{E}[\tau] = \infty$ and using the fact that $\mathbf{E}[H_\epsilon \wedge \tau_\mu^P] \le \mathbf{E}[H_\epsilon \wedge H_{\alpha^+(\epsilon)}] < \infty$ we have that $\mathbf{E}[\int_0^{\tau_\mu^P} g(S_u)du] \ge g(\epsilon)\,\mathbf{E}[\int_{H_\epsilon \wedge \tau_\mu^P}^{\tau_\mu^P} du] = \infty$, and there is nothing to prove.

So suppose $\mu \in L^2$. Then the area between the curves $U_\mu(x)$ and $|x|$ is finite.

Let $U_\epsilon(x) = \max\{U_\mu(x) - \epsilon, |x|\}$ and related quantities be defined as above.

This time, since $\tilde{U}_\epsilon \equiv U_\mu$ on $\tilde{I}_\epsilon$ we have that $\alpha_{\tilde{\mu}_\epsilon} = \alpha_\mu$ on some sub-interval $\acute{I}_\epsilon \subseteq \tilde{I}_\epsilon$ of the form $\acute{I}_\epsilon = [\acute{c}_\epsilon, \acute{d}_\epsilon]$, and as $\epsilon \downarrow 0$, $\acute{I}_\epsilon$ increases to the support of μ.

Now

$$\mathbf{E}\left[\int_0^\tau g(S_u)du\right] = \lim_{\epsilon \downarrow 0} \mathbf{E}\left[\int_0^{\sigma_\epsilon} g(S_u)du\right]$$

and

$$\mathbf{E}\left[\int_0^{\sigma_\epsilon} g(S_u)du\right] \le \mathbf{E}\left[\int_0^{\tilde{\sigma}_\epsilon} g(S_u)du\right] \le \mathbf{E}\left[\int_0^{\tau^P(\tilde{\mu}_\epsilon)} g(S_u)du\right].$$

But

$$\mathbb{E}\left[\int_0^{\tau^P(\tilde{\mu}_\epsilon)} g(S_u)du\right] = \mathbb{E}\left[\int_0^{\tau^P(\tilde{\mu}_\epsilon)\wedge H_{\acute{c}_\epsilon}\wedge H_{\acute{d}_\epsilon}} g(S_u)du\right] + \mathbb{E}\left[\int_{\tau^P(\tilde{\mu}_\epsilon)\wedge H_{\acute{c}_\epsilon}\wedge H_{\acute{d}_\epsilon}}^{\tau^P(\tilde{\mu}_\epsilon)} g(S_u)du\right].$$

Since $\alpha_{\tilde{\mu}_\epsilon} = \alpha_\mu$ on $\acute{I}_\epsilon$ and we have that $\tau^P(\tilde{\mu}_\epsilon) \wedge H_{\acute{c}_\epsilon} \wedge H_{\acute{d}_\epsilon}$ is monotonically increasing as $\epsilon \downarrow 0$ and hence the first term on the right-hand-side converges to $\mathbb{E}\left[\int_0^{\tau^P(\mu)} g(S_u)du\right]$. Meanwhile, the second term is bounded by $\bar{g}\,\mathbb{E}[\tau^P(\tilde{\mu}_\epsilon) - \tau^P(\tilde{\mu}_\epsilon) \wedge H_{\acute{c}_\epsilon} \wedge H_{\acute{d}_\epsilon}]$. This last quantity is at most $\bar{g}$ multiplied by the area between the potentials U_μ and $U_{\acute{\mu}_\epsilon}$ where $\acute{\mu}_\epsilon = \mathcal{L}(W_{\tau^P(\tilde{\mu}_\epsilon)\wedge H_{\acute{c}}\wedge H_{\acute{d}}})$. However, as ϵ tends to zero this area tends to zero. Hence $\mathbb{E}\left[\int_0^T g(S_u)du\right] \leq \mathbb{E}\left[\int_0^{\tau^P(\mu)} g(S_u)du\right]$.

3.7 An application and extensions

3.7.1 Variance swap on the sum of squared returns

We now return to the question which originally motivated the work in this chapter which was to find model-independent bounds for variance swaps given the terminal law of the underlying asset price process or equivalently, call prices with expiry T for all strikes. We will show how to bound the idealised variance swap based on squared returns, introduced in Section 3.1. The relationship between variance swap bounds and the Skorokhod embedding problem solved in Section 3.6 is a crucial insight which we will exploit in the next chapter to generalise our analysis to a general class of variance swaps monitored in discrete or continuous time.

As in Section 3.1, let $X = (X_t)_{0\leq t\leq T}$ be a square-integrable martingale started at $X_0 = x_0$ with $X_T \sim \mu$, where μ is centred at x_0 and supported on $\mathbb{R}^+$. We recall the definition for the payoff of an idealised variance swap $V_T = V((X_s)_{0\leq s\leq T}) = \int_0^T \frac{d[X,X]_t}{(X_t)^2}$. By (3.1.3) and (3.1.4) we have

$$\inf_{\tau\in\mathcal{S}_{UI}(B,\mu)} \mathbb{E}\left[\int_0^\tau \frac{du}{S_u^2}\right] \leq \mathbb{E}[V_T] \leq \sup_{\tau\in\mathcal{S}_{UI}(B,\mu)} \mathbb{E}\left[\int_0^\tau \frac{du}{I_u^2}\right].$$

Let $\tilde{\mu}$ be the measure μ reflected around 0, so that $\tilde{\mu}$ is a measure on $\mathbb{R}_-$ and observe that

$$\sup_{\tau\in\mathcal{S}_{UI}(B,\mu)} \mathbb{E}\left[\int_0^\tau \frac{du}{I_u^2}\right] = \inf_{\tau\in\mathcal{S}_{UI}(B,\tilde{\mu})} \mathbb{E}\left[\int_0^\tau \frac{du}{\tilde{S}_u^2}\right]$$

where $\tilde{B}$ is a Brownian motion started at $-x_0$, with maximum process $\tilde{S}$. Now we apply Theorem 3.6.1 to see that

$$\mathbb{E}\left[\int_0^{\tau_\mu^P} \frac{du}{S_u^2}\right] \leq \mathbb{E}[V_T] \leq \mathbb{E}\left[\int_0^{\tau_{\tilde{\mu}}^P} \frac{du}{\tilde{S}_u^2}\right].$$

Note that the Perkins embedding for $\tau_{\tilde{\mu}}$ is determined by the monotonic functions $\alpha_{\tilde{\mu}}^{\pm}$ where

$\alpha_{\bar{\mu}}^{\pm}(x) = -\alpha_{\bar{\mu}}^{\mp}(-x)$.

Example 3.7.1. *Suppose that $X_0 = 1$ and $\mu = U[0,2]$. Shifting the quantities calculated in Example 3.2.1 to allow for the starting value $X_0 = 1$ it is clear that $\alpha_\mu^+ : [1,2] \to [0,1]$ is defined $\alpha_\mu^+(s) = s - 2\sqrt{s-1}$ and $\alpha_{\bar{\mu}}^- : [0,1] \to [1,2]$ is defined $\alpha_\mu^-(i) = i + \sqrt{1-i}$. Hence the lower bound can be calculated;*

$$\mathbf{E}\left[\int_0^{\tau_\mu^P} \frac{du}{S_u^2}\right] = \mathbf{E}\left[\left(1 - \frac{B_{\tau_\mu^P}}{S_{\tau_\mu^P}}\right)^2\right] = \int_0^1 \left(1 - \frac{x}{\alpha_\mu^+(x)}\right)^2 \frac{dx}{2} = \frac{\pi}{2} - 2\log 2.$$

For the upper bound; first considering $g_\epsilon(s) = s^{-2} \wedge \epsilon^{-2}$ and then letting $\epsilon \downarrow 0$,

$$\mathbf{E}\left[\int_0^{\tau_\mu^P} \frac{du}{\tilde{S}_u^2}\right] = \mathbf{E}\left[\left(1 - \frac{\tilde{B}_{\tilde{\tau}_\mu^P}}{\tilde{S}_{\tilde{\tau}_\mu^P}}\right)^2\right] = \int_0^1 \left(1 - \frac{x}{a_{\bar{\mu}}^-(x)}\right)^2 \frac{dx}{2} = \infty.$$

3.7.2 Extension to diffusions

Suppose that $(X_t)_{t\geq 0}$ is a time-homogeneous diffusion on $I \subseteq \mathbb{R}$. More specifically, let $\sigma : I \to (0,\infty)$ and $b : I \to \mathbb{R}$ be Lipschitz functions and define $(X_t)_{t\geq 0}$ to be the solution to

$$dX_t = \sigma(X_t)dB_t + b(X_t)dt, \quad X_0 = x_0,$$

where $(B_t)_{t\geq 0}$ is a Brownian motion.

Let $s : I \to \mathbb{R}$ be the strictly increasing and C^2 scale function of X,

$$s(x_0) = 0, \qquad s'(x) = \exp\left(-\int_0^x 2\frac{b(u)}{\sigma(u)^2}du\right),$$

and let $h = s^{-1}$.

Consider the problem of maximising (or minimising) $\mathbf{E}[F(X_\tau, S_\tau^X)]$ over minimal embeddings τ of μ. Since $M_t = s(X_t)$ is a local martingale it follows that it can be represented as $M_t = W_{A(t)}$, for some (continuous) time-change $t \to A(t)$. Define the measure ν by $\nu(G) = \mu(s^{-1}(G))$ for Borel sets $G \subseteq s(I)$. Notice that σ is a minimal embedding of ν in W if and only if $\tau = A^{-1}(\sigma)$ is a minimal embedding of ν in M and hence a minimal embedding of μ in X.

Define the function $\hat{F}$ by $\hat{F}(w,s) = F(h(w), h(s))$. Then,

$$F(X_\tau, S_\tau^X) = \mathbf{F}(h(W_{A_\tau}), h(S_{A_\tau})) = \hat{F}(W_{A_\tau}, S_{A_\tau}). \tag{3.7.1}$$

Lemma 3.7.2. *Suppose F satisfies $F - MON\uparrow$. Then $\hat{F}$ satisfies $\hat{F} - MON\uparrow$ if $F_s < 0$ and*

h is concave or if $F_s > 0$ and h is convex.

Similarly, suppose F satisfies $F - MON \downarrow$. Then $\hat{F}$ satisfies $\hat{F} - MON \downarrow$ if $F_s < 0$ and h is convex or if $F_s > 0$ and h is concave.

Proof. The result follows from the expression

$$\frac{\hat{F}_s(x,s)}{s-x} = \frac{h'(s)F_s(h(x),h(s))}{h(s)-h(x)}\frac{h(s)-h(x)}{s-x}. \tag{3.7.2}$$

□

Note that h is convex (concave) when s is concave (convex), and since $2s''(x)/s'(x) = -\sigma(x)^2/b(x)$, the scale function is concave if $b(x) > 0$ for all x.

Proposition 3.7.3. *Suppose $\nu = \mu \circ h$ is centred about zero and suppose $b > 0$. Suppose F satisfies $F - MON \uparrow$ and is increasing in s. Then*

$$\begin{aligned} \sup_{\tau \in S_{UI}(X,\mu)} \mathbb{E}[F(X_\tau, S_\tau^X)] &= \mathbb{E}[\hat{F}(W_{\tau_\nu^{AY}}, S_{\tau_\nu^{AY}})] \\ \inf_{\tau \in S_{UI}(X,\mu)} \mathbb{E}[F(X_\tau, S_\tau^X)] &= \mathbb{E}[\hat{F}(W_{\tau_\nu^{P}}, S_{\tau_\nu^{P}})]. \end{aligned}$$

Remark 3.7.4. Whilst necessary to apply the results of the Brownian setting, the assumption that $\nu \equiv \mu \circ h$ is centred is not as innocuous as might first appear, and in the setting of a transient diffusion it is natural to wish to consider embeddings for target laws which, after transformation by the scale function, are not centred. For example, let X be a three-dimensional Bessel process, started at one. Then $s(x) = -1/x + 1$ and $h(m) = 1/(1-m)$. Now let μ be any probability measure on $\mathbb{R}^+$ with $\int_{\mathbb{R}^+} x^{-1}\mu(dx) \leq 1$. Then, there exists a minimal embedding of μ in X, but only if $\int_{\mathbb{R}^+} x^{-1}\mu(dx) = 1$ does this embedding correspond to a uniformly integrable embedding of $M \equiv 1 - X^{-1}$.

See Cox and Hobson [21] (and the references therein) for a further discussion of this issue, and of the construction of embeddings in Brownian motion of non-centred target laws.

Chapter 4

Model independent hedging strategies for variance swaps

4.1 Variance Swap Kernels and Model-Independent Hedging

Motivated by the results for the idealised continuously monitored variance swap based on squared returns derived in Chapter 3 our aim in this chapter is to derive model independent bounds and hedging strategies for discretely monitored variance swaps. The ideas contained in this chapter grew out of an attempt to relate results obtained by Kahalé in a recent paper [47] to the results of Chapter 3. The contribution of the work contained in this chapter is to extend the work in [47], which focuses on a specific definition of the variance swap based on squared log returns to a wide class of alternative definitions of the variance swap. In the case of the definition for realised variance based on squared returns there is a direct link between the Skorokhod embedding problem considered in Chapter 3 and the results of this chapter.

4.1.1 Variation swaps

We begin by defining the payoff of a variance swap on a path-wise basis. The payoff will depend on a kernel, on the times at which the kernel is evaluated and on the asset price at these times.

Definition 4.1.1.

(i) A *variation swap kernel* is a continuously differentiable bi-variate function $H:(0,\infty)\times(0,\infty)\to[0,\infty)$ such that for all $x\in(0,\infty)$, $H(x,x)=0=H_y(x,x)$. We say that the swap kernel is *regular* if it is twice continuously differentiable. *A variance swap kernel* is a regular variation swap kernel H such that $H_{yy}(x,x)=2x^{-2}$.

(ii) A *partition* P on $[0,T]$ is a set of times $0=t_0<t_1<...<t_N=T$. A partition is *uniform*

if $t_k = \frac{kT}{N}$, $k = 0, 1, \dots N$. A sequence of partitions $\mathcal{P} = (P^{(n)})_{n \geq 1} = (\{t_k^{(n)}; 0 \leq k \leq N^{(n)}\})_{n \geq 1}$ is *dense* if $\lim_{n \uparrow \infty} \sup_{k \in \{0, \dots, N^{(n)}-1\}} |t_{k+1}^{(n)} - t_k^{(n)}| = 0$.

(iii) A *price realisation* $f = (f(t))_{0 \leq t \leq T}$ is a càdlàg function $f : [0, T] \to (0, \infty)$.

(iv) The *payoff of a variation swap* with kernel H for a partition P and a price realisation f is

$$V_H(f, P) = \sum_{k=0}^{N-1} H(f(t_k), f(t_{k+1})). \tag{4.1.1}$$

(v) Let $\mathcal{P} = (P^{(n)})_{n \geq 1}$ be a dense sequence of partitions. If $\lim_{n \uparrow \infty} V_H(f, P^{(n)})$ exists then the limit is denoted $V_H(f, P_\infty)$ and is called the continuous time limit of $V_H(f, P^{(n)})$ on $\mathcal{P}$.

Remark 4.1.2.

(i) Our main focus in this chapter is on variance swap kernels but we will discuss variation swap kernels $H^S(x, y) = (y - x)^3$ and $H^Q(x, y) = (y - x)^2$ briefly, see Remark 4.2.1 and Example 4.5.10. (Strictly speaking H^S is not a variation swap kernel since it is not non-negative, but most of our analysis still applies in this case.) A regular variation swap kernel is a variance swap kernel if $H(x, x(1 + \delta)) = \delta^2 + o(\delta^2)$ for δ small. Examples of variance swap kernels include $H^R(x, y) = \left(\frac{y-x}{x}\right)^2$, $H^L(x, y) = (\log(y) - \log(x))^2$ and $H^B(x, y) = -2\left(\log(y/x) - \left(\frac{y-x}{x}\right)\right)$.

(ii) The price realisations f should be interpreted as realisations of the forward price of the asset with maturity T. Later we will extend the analysis to cover undiscounted price processes, rather than forward prices.

(iii) Large parts of the subsequent analysis can be extended to allow for price processes which can take the value zero, provided we also define $H(0, 0) = 0$, or equivalently truncate the sum in (4.1.1) at the first time in the partition that f hits 0. In this case we must have that zero is absorbing, so that if $f(s) = 0$, then $f(t) = 0$ for all $s \leq t \leq T$.

(iv) In practice the variance swap contract is an exchange of the quantity $V = V_H(f, P)$ for a fixed amount K. However, since there is no optionality to the contract, and since the contract paying K can trivially be priced and hedged, we concentrate solely on the floating leg.

(v) In many of the earliest academic articles, and in particular in Demeterfi et. al [24, 25], but also in some very recent articles, e.g. Zhu and Lian [77], the variance swap is defined in terms of the kernel H^R. However, it has become market practice to trade variance swaps based on the kernel H^L. Nonetheless these contracts are traded over-the-counter and in principle it is possible to agree any reasonable definition for the kernel. Variance swaps defined using the variance kernel H^B were introduced by Bondarenko [11], see also Neuberger [59]. As we shall see, the contract based on this kernel has in its favour various desirable features. For continuous paths, in the limit of a dense partition the contracts based on either of these three kernels are equivalent, see Example 4.5.10 and Lemma 4.5.9, but this is not the case in general.

(vi) The labels $\{S, Q, R, L, B\}$ on the variation swap kernels denote {Skew, Quadratic, Returns, Logarithmic returns, Bondarenko} respectively.

An important concept will be the quadratic variation of a path. For a dense sequence of partitions $\mathcal{P}$, the *quadratic variation* $[f]$ of f on $\mathcal{P}$ is defined to be $[f]_t = \lim_{n\uparrow\infty} \sum_{t_k^{(n)} \leq t} (f(t_{k+1}^{(n)}) - f(t_k^{(n)}))^2$, provided the limit exists. We split the function into its continuous and discontinuous parts, $[f]_t = [f]_t^c + \sum_{u \leq t} (\Delta f(u))^2$. Later we will relate this definition to that introduced by Föllmer [34], which is used to develop a path-wise version of Itô calculus.

4.1.2 Model independent pricing

Our goal is to discuss how to price the variance swap contract, or more generally any path-dependent claim, under an assumption that European call and put (vanilla) options with maturity T are traded and can be used for hedging, but without any assumption that a proposed model is a true reflection of the real dynamics. In this sense the strategies and prices we derive are model independent and robust.

Let call prices for maturity T be given by $C(K)$, written as a function of strike and expressed in units of cash at time T. We assume that a continuum of calls are traded, and to preclude arbitrage we assume that C is a decreasing convex function such that $C(0) = f(0)$, $C(K) \geq (f(0) - K)^+$ and $\lim_{K\uparrow\infty} C(K) = 0$, see e.g. Davis and Hobson [23]. We exclude the case where $C(f(0)) = 0$ for then $C(K) = (f(0) - K)^+$ and the situation is degenerate: the forward price must remain constant and upper and lower bounds on the price of the variance swap are zero. Although we assume that calls are traded today (time 0), we do not make any assumption on how call prices will behave over time, except that they will respect no-arbitrage conditions and that on expiry they will be worth the intrinsic value.

Definition 4.1.3. A *synthesisable payoff* is a function $\psi : (0, \infty) \mapsto \mathbb{R}$ which can be represented as the difference of two convex functions (so that $\psi''(x)$ exists as a measure).

Let $\Psi = \{\psi : \psi \in \Psi\}$ be the set of synthesisable payoffs $\psi : (0, \infty) \mapsto \mathbb{R}$. Then the left- and right- derivatives $\psi'_\pm$ (or $\psi'(x\pm)$) exist and we have

$$\psi(f) = \psi(f(0)) + \psi'_+(f(0))(f - f(0)) + \int_{(0,f(0)]} (x - f)^+ \psi''(x)dx + \int_{(f(0),\infty)} (f - x)^+ \psi''(x)dx. \tag{4.1.2}$$

Thus we can represent the payoff of any sufficiently regular European contingent claim as a constant plus the gains from trade from holding a fixed quantity of forwards, plus the payoff of a static portfolio of vanilla calls and puts.

Let $D[0, t]$ denote the space of cádlág functions on $[0, t]$.

Definition 4.1.4. A *dynamic strategy* for a fixed partition P is a collection of functions $\Delta = (\delta_{t_0}, \ldots, \delta_{t_{N-1}})$, where $\delta_{t_j} : D[0, t_j] \to \mathbb{R}$. The *payoff of a dynamic strategy* along a price realisation f is

$$\sum_{k=0}^{N-1} \delta_{t_k}((f(t))_{0 \le t \le t_k})(f(t_{k+1}) - f(t_k)). \tag{4.1.3}$$

Let $\bar{\Delta}(P)$ be the set of dynamic strategies.

Definition 4.1.5. $\Delta = \bar{\Delta}(P)$ is a *Markov dynamic strategy* if $\delta_{t_j}(f(t)_{0 \le t \le t_j}) = \delta_{t_j}(f(t_j))$ for all j. A Markov dynamic strategy is a *time homogeneous Markov dynamic strategy* (THMD-strategy) if $\delta_{t_j}(f(t_j)) = \delta(f(t_j))$ for all j.

The quantity δ_{t_j} represents the quantity of forwards to be held over the interval $(t_j, t_{j+1}]$. In principle this quantity may depend on the current time and on the price history $(f(t))_{0 \le t \le t_j}$. However, as we shall see, for our purposes it is sufficient to work with a much simpler set of strategies where the quantity does not explicitly depend on time, nor on the price history except through the current value. We call this the Markov property, but note there are no probabilities involved here yet.

Definition 4.1.6. A *semi-static hedging strategy* (ψ, Δ) is a function $\psi \in \Psi$ and a dynamic strategy $\Delta \in \bar{\Delta}(P)$. The terminal payoff of a semi-static hedging strategy for a price realisation f is

$$\psi(f(T)) + \sum_{k=0}^{N-1} \delta_{t_k}((f(t))_{0 \le t \le t_k})(f(t_{k+1}) - f(t_k)). \tag{4.1.4}$$

Without loss of generality we may assume that $\psi'(f(0)+) = 0$. If not then we simply adjust each δ_{t_k} by the quantity $\psi'(f(0)+)$ and the payoff in (4.1.4) is unchanged. In the sequel, we will concentrate on the case when Δ is a THMD strategy. Then we identify $\Delta \in \bar{\Delta}(P)$ with $\delta : (0, \infty) \to \mathbb{R}$ and write (ψ, δ) instead of (ψ, Δ).

Given that investments in the forward market may be assumed to be costless, the dynamic strategy has zero price. Thus, in order to define the price of a semi-static hedging strategy it is sufficient to focus on the price associated with the payoff function ψ. The last two terms in (4.1.2) are expressed in terms of the payoffs of calls and puts. Thus we can identify the price of $\psi(f(T))$ with the price of a corresponding portfolio of vanilla objects. We also use put-call parity[1] to express the cost of the penultimate term in (4.1.2) in terms of call prices. Let $\Psi_0 = \{\psi \in \Psi : \psi'_+(f(0)) = 0\}$, and let $\Psi_c \subseteq \Psi_0$ be the subset of Ψ_0 consisting of the continuously differentiable functions.

Definition 4.1.7. The *price of a semi-static hedging strategy* ($\psi \in \Psi_0, \Delta \in \bar{\Delta}(P)$) is

$$\psi(f(0)) + \int_{(0,f(0)]} \psi''(x)(C(x) - f(0) + x)dx + \int_{(f(0),\infty)} \psi''(x)C(x)dx.$$

The idea we wish to capture is that the agent holds a static position in calls together with a dynamic position in the underlying such that in combination they provide sub- and super-hedges for the claim.

Definition 4.1.8. Let $G = G((f(t_k))_{k=0,\dots,N})$ be the payoff of a path-dependent option. Suppose that there exists a semi-static hedging strategy (ψ, Δ) such that on the partition P

$$G \leq \text{(respectively } \geq\text{)} \;\; \psi(f(T)) + \sum_{k=0}^{N-1} \delta_{t_k}((f(t))_{0\leq t\leq t_k})(f(t_{k+1}) - f(t_k)).$$

Then (ψ, Δ) is called a *semi-static super-hedge* (respectively *semi-static sub-hedge*) for G.

Given a semi-static sub-hedge (respectively super-hedge) we say that the price of the sub-hedge (respectively super-hedge) is a *model independent lower (respectively upper) bound* on the price of the path-dependent claim G.

4.1.3 Consistent models

The aim of the agent is to construct a hedge which works path-wise, and does not depend on an underlying model. Nonetheless, sometimes it is convenient to introduce a probabilistic model and a stochastic process, and to interpret $f(t)$ as a realisation of that stochastic process. Then we work with a probability space $(\Omega, \mathcal{F}, \mathbf{F}, \mathbf{P})$ supporting the stochastic process $X = (X_t)_{0\leq t\leq T}$.

[1]This means that we do not need to introduce a notation for the put price, which is convenient since P is already in use for the partition. Put-call parity for the forward says that the price of a put with strike x is the price of a call with the same strike minus $f(0) - x$

Definition 4.1.9. A model $(\Omega, \mathcal{F}, \mathbb{F}, \mathbb{P})$ and associated stochastic process $X = (X_t)_{0 \le t \le T}$ is *consistent* with the call prices $(C(K))_{K \ge 0}$ if $(X_t)_{t \ge 0}$ is a non-negative $(\mathbb{F}, \mathbb{P})$-martingale and if $\mathbb{E}[(X_T - K)^+] = C(K)$ for all $K > 0$.

In the setting of a stochastic model $V_H(X, P) : \Omega \to \mathbb{R}^+$ is a random variable, and for $\omega \in \Omega$, $V_H(X(\omega), P)$ is a realised value of a variance swap. From a pricing perspective we are interested in getting upper and lower bounds on $\mathbb{E}[V_H(X(\omega), P)]$ as we range over consistent models. Knowledge of call prices is equivalent to knowledge of the marginal law of X_T under a consistent model (Breeden and Litzenberger [13]). If we write μ for the law of X_T and if $C_\mu(K) = \mathbb{E}[(Z_\mu - K)^+]$ where Z_μ is a random variable with law μ, then X is consistent for the call prices C if $C_\mu(K) = C(K)$. We write $m = \int_0^\infty x\mu(dx)$ and we assume, using the martingale property, that $f(0) = m$. Then the problem of characterising consistent models is equivalent to the problem of characterising all martingales with a given distribution at time T.

4.2 Motivation

4.2.1 The continuous case

In the situation where both the monitoring and the price-realisations are continuous the theory for the pricing of variance swaps is complete and elegant. We will use this setting to develop intuition for the jump case.

Suppose that the price realisation f is continuous, and possesses a quadratic variation $[f] : [0, T] \to \mathbb{R}^+$ on a dense sequence of partitions $\mathcal{P}$. Dupire [27] and Neuberger [58] independently made the observation that the continuity assumption implies that a variance swap with payoff $\int_0^T f(t)^{-2} d[f]_t$ can be replicated perfectly by holding a static portfolio of log contracts and trading dynamically in the underlying asset. Both Dupire and Neuberger assume $f \equiv X$ is a realisation of a continuous semi-martingale, but in our setting, the observation follows from a path-wise application of Itô's formula in the sense of Föllmer [34], see Section 4.5. Applying Itô's formula to $-2\log(f(t))$ we have

$$-2\log(f(T)) + 2\log(f(0)) = -2\int_0^T \frac{1}{f(t)} df(t) + \int_0^T \frac{1}{f(t)^2} d[f]_t. \tag{4.2.1}$$

Then, as we show in Section 4.5 below, down a dense sequence of partitions

$$V_H(f, P_\infty) = \int_0^T \frac{1}{f(t)^2} d[f]_t = -2\log(f(T)) + 2\log(f(0)) + \int_0^T \frac{2}{f(t)} df(t). \tag{4.2.2}$$

Provided it is possible to trade continuously and without transaction costs, the right-hand-side of this identity has a clear interpretation as the sum of a European contingent claim with maturity T and payoff $-2\log(f(T)/f(0))$ and the gains from trade from a dynamic investment of $2/f(t)$ in the underlying. Alternatively, the right-hand-side of (4.2.2) can be viewed as the payoff of a semi-static hedging strategy in the continuous time limit for the choice $\psi(x) = -2\log(x/f(0)) + 2(x - f(0))/f(0)$ and $\Delta = (\delta_t)_{0\leq t\leq T}$ where $\delta_t((f(u))_{0\leq u\leq t}) = (2/f(t)) - (2/f(0))$. Note that there is equality in (4.2.2) so that (ψ, δ) is both a sub- and super-hedge for $V_H(f, P_\infty)$. In particular, under a price continuity assumption, the variance swap has a model-independent price and an associated riskless hedge.

4.2.2 The effect of jumps on hedging with the classical continuous hedge

Even if the continuity assumption cannot be justified, the associated replication strategy is nevertheless a reasonable candidate for a hedging strategy in the general case. Let us focus on the discrepancy between the payoff of the variance swap and the gains from trade resulting from using the hedge derived in the continuous case. The path-by-path Itô formula continues to apply in the case with jumps, see [34] and Section 4.5 below. Hence

$$\begin{aligned} -2\log(f(T)) + 2\log(f(0)) &= -2\int_0^T \frac{1}{f(t-)}df(t) + \int_0^T \frac{1}{f(t-)^2}d[f]^c_t \\ &\quad + \sum_{0\leq t\leq T} 2\left\{\left(\frac{\Delta f(t)}{f(t-)}\right) - \log\left(1+\frac{\Delta f(t)}{f(t-)}\right)\right\}. \end{aligned}$$

Note that $d[\log(f)]_t = d[f]^c_t/f(t-)^2 + (\Delta\log(f(t)))^2$. By adding and subtracting the discontinuous part of the quadratic variation of $\log(f)$ on the right-hand-side of the above expression, we find

$$-2\log(f(T)) + 2\log f(0) = -2\int_0^T \frac{1}{f(t-)}df(t) + [\log(f)]_T - \sum_{0\leq t\leq T} J_L(\Delta f(t)/f(t-)) \quad (4.2.3)$$

where

$$J_L(\eta) = -2\eta + 2\log(1+\eta) + \log(1+\eta)^2.$$

It is intuitively clear, but see also Corollary 4.5.5, that $V_{HL}(f, P_\infty) \equiv [\log(f)]_T$. Then it follows by re-arrangement of equation (4.2.3) that the discrepancy between the realised value of the variance swap $V_{HL}(f, P_\infty)$ and the return generated by the classical continuous hedging strategy can be represented as the sum of the jump contributions:

$$V_{HL}(f, P_\infty) - \left(-2\log(f(T)) + 2\log f(0) + 2\int_0^T \frac{1}{f(t-)}df(t)\right) = \sum_{0\leq t\leq T} J_L\left(\frac{\Delta f(t)}{f(t-)}\right).$$

We call this the hedging error with the convention that if the hedge sub-replicates the variance swap then the hedging error is positive.

Now consider the kernel H^R and define $V_{H^R}(f, P_\infty) = \int_0^T d[f]_t/f(t-)^2$, again, see Corollary 4.5.5 for justification. By a similar analysis, but adding and subtracting $\left(\frac{\Delta f(t)}{f(t-)}\right)^2$ instead of the discontinuous part of the quadratic variation of $\log(f)$, we have

$$V_{H^R}(f, P_\infty) - \left(-2\log(f(T)) + 2\log(f(0)) + 2\int_0^T \frac{1}{f(t-)} df(t)\right) = \sum_{0 \le t \le T} J_R\left(\frac{\Delta f(t)}{f(t-)}\right).$$

where

$$J_R(\eta) = -2\eta + 2\log(1+\eta) + \eta^2.$$

In the continuous case, under some mild regularity conditions on f and $\mathcal{P}$, the variance swap value is independent of the chosen kernel. In contrast, the value of a variance swap in the general case is highly dependent on the chosen kernel.

To see that this is the case, and to examine the impact of jumps on the hedging error for the kernels H^L and H^R we consider the shapes of the functions J_R and J_L, see Figure 1. For the kernel H^L, a downward jump results in a positive contribution to the hedging error. Thus, if all jumps are downwards, then the classical continuous hedging strategy sub-replicates $V_{H^L}(f, P_\infty)$. Conversely, upward jumps result in a negative contribution to the hedging error. The story is reversed for the kernel H^R.

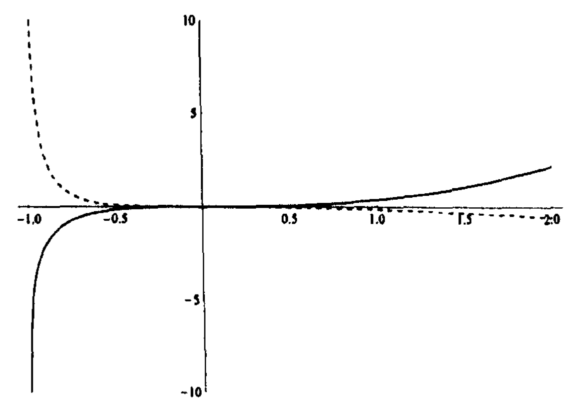

Figure 4.1: J_L (as represented by the dashed line) is convex decreasing for $x \le 0$ and concave decreasing for $x \ge 0$. In contrast J_R (solid line) is first concave increasing and then convex increasing. The different shapes of these two curves explains the different nature of the dependence of the payoff of the variance swap on upward and downward jumps for different kernels.

It follows from the argument in the previous paragraph that for the kernel H^L the hedging error will be maximised under scenarios for which the price realisation has downward jumps, but no upward jumps. Paths with this feature might arise as realisations of $-N$ where $N = (N_t)_{t\geq 0}$ is a compensated Poisson process. Moreover, from the convexity of J_L on $(-1,0)$, it is plausible that the scenarios in which the hedging error is maximised are those in which price realisations have a single large downward jump, rather than a series of small jumps. Again if we wish to minimise the hedging error we should expect a single large upward jump, and the story is reversed for the kernel H^R.

In summary, we find that, under a continuity assumption on f, and for a dense sequence of partitions, the value of a variance swap is independent of the kernel and can be replicated with a static hedge in a forward contract and a dynamic hedging strategy. In the presence of jumps, however, the value of the variance swap depends on the kernel. An agent who holds a variance swap and hedges under the assumption of continuity, may super-replicate or sub-replicate the payoff depending on the form of the jumps. For example, for the kernel H^L an agent who acts as if the price realisation can be assumed to be continuous will sub-replicate the variance swap if there are downward jumps and no upward jumps. Such an agent will underprice the swap.

We will use the analysis of this section to give us intuition about the extremal models which will lead to the price bounds on variance swaps derived in the Section 4.3. The bounds will depend crucially on the kernel. Models under which the variance swap with kernel H^L has highest price (assuming consistency with a given set of call prices) will be characterised by a single downward jump and no upward jumps.

Remark 4.2.1. We will see later that the model which minimises the price for variance swaps with kernel H^R also minimises the price for variation swaps with kernel H^S. If f has a quadratic variation, then in the continuous limit $V_{H^S}(f,P_\infty) = \sum_{0<t\leq T}(\Delta f(t))^3$. This payoff will be smallest if all jumps are downwards and we will see that if the call prices are given for expiry time T, then the model that produces the lowest price is one under which the price path has a single downward jump.

4.2.3 The related Skorokhod embedding problem

In this section we relate the problem of finding extremal prices for the variance swap to the Skorokhod embedding problem of Chapter 3. The aim is to transfer the results to gain intuition which will guide the derivation of the optimal model-free hedges in the next section.

Let μ be a measure on $\mathbf{R}^+$ with mean m and let $(\Omega,\mathcal{F},\mathbf{F},\mathbf{P})$ be a filtered probability space supporting a right-continuous martingale $X = (X_t)_{0\leq t\leq T}$ such that $X_0 = m$ and $X_T \sim \mu$. Suppose there exists Brownian motion B started at m and a time-change $t \to A_t$, null at 0, such that $X_t = B_{A_t}$. Here B is defined relative to a filtration $\mathbf{G} = (\mathcal{G}_u)_{0\leq u\leq A_T}$ and $\mathcal{F}_t \subseteq \mathcal{G}_{A_t}$. Let A^c be the continuous part of A. Note that $dA^c_t = (dX^c_t)^2 = d[X]^c_t$. Let $S^X = (S^X_t)_{t\geq 0}$ (respectively

S) be the process of the running maximum of X (respectively B) so that $S_t^X = \sup_{u\le t} X_u$. Note that $X_t \le S_t^X \le S_{A_t}$. If we suppose for the moment that μ has a second moment then $(X_t)_{0\le t\le T}$ is a square-integrable martingale and recalling Equations (3.1.1) and (3.1.1) we have

$$\mathbb{E}[V_{H^R}(X, P_\infty)] = \mathbb{E}\left[\int_0^T \frac{d[X]_t^c}{(X_{t-})^2} + \sum_{0\le t\le T}\left(\frac{\Delta X_t}{X_{t-}}\right)^2\right] \ge \mathbb{E}\left[\int_0^{A_T} \frac{du}{(S_u)^2}\right] \tag{4.2.4}$$

In Chapter 3 this inequality was the motivation to solve the Skorokhod embedding problem

$$\min_{\tau\in S_{UI}(B,\mu)} \mathbb{E}\left[\int_0^\tau \frac{du}{S_u^2}\right], \tag{4.2.5}$$

We showed that the minimum over uniformly integrable embeddings of μ is attained by the Perkins embedding, τ_μ^P.

Let $I = (I_t)_{t\ge 0}$ denote the infimum process $I_t = \inf_{u\le t} B_u$. We briefly recall the Perkins embedding from Chaper 3.

Theorem 4.2.2. *[Perkins [61], Hobson and Pedersen [43]] Given ν a probability measure with support on $\mathbb{R}^+$, with mean m let Z_ν denote a random variable with law ν and define $C_\nu(z) = \mathbb{E}[(Z_\nu - z)^+]$ and $P_\nu(z) = \mathbb{E}[(z - Z_\nu)^+]$. Define also $\alpha^+ = \alpha_\nu^+ : (m,\infty) \mapsto [0,m)$ and $\alpha^- = \alpha_\nu^- : [0,m) \mapsto (m,\infty)$ by*

$$\alpha^+(z) = \arg\min_{y<m} \frac{C_\nu(z) - P_\nu(y)}{z-y}, \qquad \alpha^-(z) = \arg\min_{y>m} \frac{P_\nu(z) - C_\nu(y)}{y-z}. \tag{4.2.6}$$

Let B be Brownian motion started at m, with maximum process S and minimum process I. Suppose μ has no atom at m. Then $\tau_\nu^P := \inf\{u > 0 : B_u < \alpha_\nu^+(S_u) \text{ or } B_u > \alpha_\nu^-(I_u)\}$ solves the Skorokhod embedding problem for ν in the sense that $B_{\tau_\nu^P} \sim \nu$ and $(B_{t\wedge\tau_\nu^P})_{t\ge 0}$ is uniformly integrable.

If ν has an atom at m then we assume $\mathcal{F}_0$ is sufficiently rich as to support a uniform random variable $\tilde{Z}_U$, which is independent of B. Then

$$\tau_\nu^P := \begin{cases} 0 & \tilde{Z}_U \le \nu(\{m\}) \\ \inf\{u > 0 : B_u < \alpha_\nu^+(S_u) \text{ or } B_u > \alpha_\nu^-(I_u)\} & \tilde{Z}_U > \nu(\{m\}) \end{cases}$$

solves the Skorokhod embedding for ν.

The Perkins embedding has the following minimality property: for increasing functions F it minimises $\mathbb{E}[F(S_\tau)]$ over uniformly integrable embeddings τ of μ. Moreover as we showed in Chapter 3 it also minimises the expected value of functionals of the joint law of the running maximum and terminal value $F(B_\tau, S_\tau)$ over stopping times τ in $S_{UI}(B,\mu)$, provided F satisfies a monotonicity condition. The salient characteristic of the Perkins embedding which results in optimality is that either $B_{\tau_\mu^P} = S_{\tau_\mu^P}$ or $B_{\tau_\mu^P} = \alpha_\nu^+(S_{\tau_\mu^P})$.

Now consider the problem of finding the consistent model for which $V_{HR}(X, P_\infty)$ has lowest possible price, and recall that knowledge of call prices is equivalent to knowledge of the marginal law μ of X_T. To obtain the lowest possible price we might expect equality in each of (3.1.1) and (3.1.2), and thus that just before a jump, the process is at its current maximum. Moreover, the model should be related to the Perkins embedding.

Lemma 4.2.3. *Let B be Brownian motion started at m. Let $\overline{H}_b = \inf\{u \geq 0 : B_u = b\}$ be the first hitting time of level b by Brownian motion. Let $\Lambda(t)$ be a strictly increasing, continuous function such that $\Lambda(0) = m$ and $\lim_{t\uparrow T} \Lambda(t)$ is infinite.*

Define the process $\tilde{Q}^\mu = (\tilde{Q}_t^\mu)_{0\leq t\leq T}$ by

$$\tilde{Q}_t^\mu = B_{\overline{H}_{\Lambda(t)}\wedge\tau_\mu^P}, \tag{4.2.7}$$

and let Q^μ be the right-continuous modification of $\tilde{Q}^\mu$.

Then, Q^μ is a martingale such that $Q_T^\mu \sim \mu$. Moreover, the paths of Q^μ are continuous and increasing, except possibly at a single jump time. Finally, either $Q_T^\mu \equiv B_{\tau_\mu^P} = S_{\tau_\mu^P}$ or $Q_T^\mu \equiv B_{\tau_\mu^P} = \alpha_\mu(S_{\tau_\mu^P})$.

Proof. Since τ_μ^P is finite almost surely we have that $Q_T^\mu \equiv B_{\tau_\mu^P} \sim \mu$. Moreover, for $\Lambda(t) < \tau_\mu^P$, $Q_t^\mu = \Lambda(t) = B_{\overline{H}_{\Lambda(t)}} = S_{\overline{H}_{\Lambda(t)}}$. □

The martingale Q^μ will be used in Section 4.5 to show that in the continuous-time limit, the bounds we obtain are tight. The martingale Q^μ is related to the Perkins embedding in the same way that the Dubins-Gilat [26] martingale is related to the Azéma-Yor [7] embedding.

We can also consider a reflected version of the martingale Q^μ based on the infimum process rather than the maximum process.

Lemma 4.2.4. *Let $\lambda(t)$ be a strictly decreasing, continuous function such that $\lambda(0) = m$ and $\lim_{t\uparrow T} \lambda(t)$ is zero.*

Define the process $\tilde{R}^\mu = (\tilde{R}_t^\mu)_{0\leq t\leq T}$ by

$$\tilde{R}_t^\mu = B_{\overline{H}_{\lambda(t)}\wedge\tau_\mu^P}, \tag{4.2.8}$$

and let R^μ be the right-continuous modification of $\tilde{R}^\mu$.

Then, R^μ is a martingale such that $R_T^\mu \sim \mu$. Moreover, the paths of R^μ are continuous and decreasing, except possibly at a single jump time. Finally, either $R_T^\mu \equiv B_{\tau_\mu^P} = I_{\tau_\mu^P}$ or $R_T^\mu \equiv B_{\tau_\mu^P} = \alpha_\mu^-(I_{\tau_\mu^P})$.

Remark 4.2.5. In this section we have exploited a connection between the problem of finding bounds on the prices of variance swaps and the Skorokhod embedding problem. This link is one of the recurring themes of the literature on the model-independent bounds, see Hobson [42]. We exhibit this link for the kernel H^R, and in this sense at least, it seems that variance swaps defined via H^R are the more natural mathematical object. Nonetheless, the intuition developed via H^R and the Skorokhod embedding problem is valid more widely.

4.3 Path-wise Bounds for Variance Swaps

We now begin the construction of path-wise hedging strategies based on the intuition developed previously. To construct a sub-hedge for a variation swap with kernel H for any price realisation f, suppose that there exists a pair of functions (ψ, δ) such that for $x, y \in \mathbb{R}$

$$H(x,y) \geq \psi(y) - \psi(x) + \delta(x)(y-x). \tag{4.3.1}$$

Then we may interpret (ψ, δ) as a semi-static hedging strategy (for a Markov and time-homogeneous dynamic strategy) and then for any price realisation f and partition P,

$$V_H(f,P) \geq \psi(f(T)) - \psi(f(0)) + \sum_k \delta(f(t_k))(f(t_{k+1}) - f(t_k)).$$

By Definition 4.1.8 we have constructed a sub-hedge for the variation swap with kernel H.

If (ψ, δ) satisfies (4.3.1) then so does $(\psi + a + by, \delta - b)$ for any constants a, b. Earlier we argued that without loss of generality for a semi-static hedging strategy we could assume $\psi'_+(f(0)) = 0$. Now we may restrict attention further to ψ with $\psi(f(0)) = 0$. Let $\Psi_{0,0} = \{\psi \in \Psi : \psi(f(0)) = 0 = \psi'(f(0)+)\}$.

Suppose now that H is a variance swap kernel, and that ψ is differentiable. Recall that $H_y(x,x) = 0$. Dividing both sides of (4.3.1) by $y - x$ and letting $y \downarrow x$, we find that $\delta(x) \leq -\psi'(x)$. Similarly letting $y \uparrow x$, $\delta(x) \geq -\psi'(x)$. Thus if (4.3.1) is to hold we must have that $\delta \equiv -\psi'$ and our search for pairs of functions satisfying (4.3.1) is reduced to finding differentiable functions ψ satisfying

$$H(x,y) \geq \psi(y) - \psi(x) - \psi'(x)(y-x). \tag{4.3.2}$$

or equivalently, $\psi(y) \leq H(x,y) + \psi(x) + \psi'(x)(y-x)$. Note that there is equality in this last expression at $y = x$.

It remains to show how to choose ψ solving (4.3.2). Using the intuition developed in the previous section for the kernel H^R we expect optimal sub-hedging strategies to be associated with the martingale Q defined in (4.2.7). For realisations of Q, either the path has no jump, or there is a single jump, and if the jump occurs when the process is at x then the jump is to $\alpha^+(x)$.

With this in mind let $\kappa : [f(0), \infty) \to (0, f(0)]$ be a decreasing function with a continuously differentiable inverse k. Fix $y > f(0)$ and consider varying x over $x < f(0)$. We want there to be equality in (4.3.2) at $x = k(y)$, and then also the x-derivatives of both sides of both sides of (4.3.2) must match. Then ψ must satisfy

$$\psi(y) = H(k(y), y) + \psi(k(y)) + \psi'(k(y))(y - k(y)), \tag{4.3.3}$$

and moreover, if ψ' is differentiable, we must have $H_x(k(y), y) + \psi''(k(y))(y - k(y)) = 0$ or equivalently

$$\psi''(x) = H_x(x, \kappa(x))/(x - \kappa(x)). \tag{4.3.4}$$

This suggests that we can define candidate sub-hedge payoffs ψ via (4.3.4) on $(f(0), \infty)$ and via (4.3.3) on $(0, f(0))$.

Now we wish to extend these arguments to the case when ψ and κ need not be regular. Suppose that left- and right-derivatives of ψ exist. By the arguments above we find that if (4.3.1) is to hold then $-\psi'(x-) \leq \delta(x) \leq -\psi'(x+)$. It does not matter which δ we choose in this interval, but for definiteness we take $\delta = -\psi'_+$.

Definition 4.3.1. $\psi \in \Psi_0$ is a *candidate sub-hedge payoff* if for all $y \in (0, \infty)$,

$$\psi(y) = \inf_x \left\{ H(x, y) + \psi'(x)(y - x) + \psi(x) \right\}. \tag{4.3.5}$$

Given a candidate sub-hedge payoff ψ we can generate a candidate semi-static hedge (ψ, δ) by taking $\delta = -\psi'$. We will say that ψ is the root of the semi-static sub-hedge $(\psi, -\psi')$.

Remark 4.3.2. See Remark 4.3.7 for the connection between Definition 4.3.1 and generalised convex duality.

Let $\mathcal{K} = \mathcal{K}(f(0))$ be the set of monotone decreasing right-continuous functions κ : $[f(0), \infty) \to (0, f(0)]$, with $\kappa(f(0)) = f(0)$ and let k denote the right-continuous inverse to κ. Define $\Phi(u, y) = H_x(u, y)/(u - y)$. Write $\Phi^R(u, y) = H^R_x(u, y)/(u - y)$, and similarly for other kernels.

Definition 4.3.3. For $\kappa \in \mathcal{K}$ with inverse k, define $\psi_{\kappa,H} \equiv \psi_\kappa : (0, \infty) \mapsto \mathbf{R}^+$, by $\psi_\kappa(f(0)) = 0$

and

$$\psi_\kappa = \left\{ \begin{array}{c} \psi_\kappa(x) \\ \psi_\kappa(z) \end{array} \right\} = \left\{ \begin{array}{ll} \int_{f(0)}^x (x-u)\Phi(u,\kappa(u))du & x > f(0) \\ \psi_\kappa(k(z)) + \psi_\kappa'(k(z))(z-k(z)) + H(k(z),z) & z < f(0) \end{array} \right.$$

We call such a function a *candidate payoff of Class $\mathcal{K}$.*

By convention we use the variable x on $(f(0),\infty)$ and z on $(0,f(0))$, to reflect the fact that ψ is defined explicitly on the former set, but only implicitly on the latter.

Remark 4.3.4. For $x > f(0)$ we have $\psi_\kappa'(x) = \int_{f(0)}^x \Phi(u,\kappa(u))du$. Note that ψ_κ' is continuous on $[f(0),\infty)$. For $z < f(0)$ we have $\psi_\kappa'(z+) = \psi_\kappa'(k(z)) + H_y(k(z),z)$ so that ψ_κ' is continuous at z if k is continuous there. If H is a regular variation kernel then it is straightforward to show that ψ_κ defined via (4.3.3) is the difference of two convex functions, and therefore that $\psi_\kappa \in \Psi_{0,0}$.

For the present we fix κ and we write simply ψ for ψ_κ. Note that the value of $\psi(x)$ does not depend on the right-continuity assumption for κ. Further, as we now argue, it does not depend on the right-continuity assumption of the inverse k. Observe that if κ is not injective and there is an interval $A_z \equiv \{x : \kappa(x) = z\} \subseteq (m,\infty)$ over which κ takes the value z then k has a jump at z. Nonetheless, the value of $\psi(z)$ does not depend on the choice of $k(z)$. To see this, for $x \in A_z$ consider $\Psi(x) := \psi(x) + \psi'(x)(z-x) + H(x,z)$. Then, on A_z, $d\Psi/dx = \psi''(x)(z-x) + H_x(x,z) \equiv 0$, using (4.3.4).

Motivated by the results of Section 4.4.3 we have defined ψ relative to the set of decreasing functions $\mathcal{K}$ with the aim of constructing a sub-hedge. However, there are analogous definitions based on constructing super-hedges or using the martingale R or both.

Definition 4.3.5. $\psi : (0,\infty) \to (0,\infty)$ is a *candidate super-hedge payoff* if for all $y \in (0,\infty)$,

$$\psi(y) = \sup_x \left\{ H(x,y) + \psi'(x)(y-x) + \psi(x) \right\}. \tag{4.3.6}$$

Define $\mathcal{L} = \mathcal{L}(f(0))$ be the set of monotone decreasing functions $\ell : (0,f(0)) \to (f(0),\infty)$, with $\ell(f(0)) = f(0)$. Let l be inverse to ℓ.

Definition 4.3.6. For $\ell \in \mathcal{L}$ with inverse l, define $\psi_\ell : (0,\infty) \mapsto \mathbb{R}^+$, *the candidate payoff of Class $\mathcal{L}$* by $\psi_\ell(f(0)) = 0$ and

$$\psi_\ell = \left\{ \begin{array}{c} \psi_\ell(x) \\ \psi_\ell(z) \end{array} \right\} = \left\{ \begin{array}{ll} \int_x^{f(0)} (u-x)\Phi(u,\ell(u))du & x < f(0) \\ \psi_\ell(l(z)) + \psi_\ell'(l(z))(z-l(z)) + H(l(z),z) & z > f(0) \end{array} \right.$$

Remark 4.3.7. Given $\psi_\ell : (0,\infty) \to \mathbb{R}^+$, define a function $m : (0,f(0)] \to \mathbb{R}$, $m(z) = -\psi_\ell(z)$ and $n : [f(0),\infty) \to \mathbb{R}$, $n(z) = \psi_\ell(z)$ and let $G(x,y) = H(x,y) + \psi_\ell'(x)(y-x)$. Then by (4.3.6) and the construction of ψ_ℓ, m and n are G-convex duals, i.e $n = m^G$ (see Definition 1.2.3) so that

$$n(y) = \sup_{0<x\leq f(0)} \{G(x,y) - m(x)\}.$$

For $y \in [f(0),\infty)$, $l(y) \in \partial^G n(y)$, where $\partial^G n(y)$ is the G-subdifferential of n at y (Definition 1.2.4). The connection between generalised convexity, super-hedging strategies and classical mass transport is the subject of ongoing work. An overview of the underlying questions is presented in Chapter 5.

Our next aim is to give conditions which guarantee that the semi-static strategy $(\psi, -\psi')$ satisfies equation (4.3.1).

Definition 4.3.8. A variation swap kernel H is an *increasing* (a *decreasing*) kernel if it is a regular variation swap kernel and

(i) $\Phi(u,y)$ is monotone increasing (decreasing) in y,

(ii) $H(a,b) + H_y(a,b)(c-b) \geq (\leq) H(a,c) - H(b,c)$ for all $a > b$.

Remark 4.3.9. A sufficient condition for the second condition in Definition 4.3.8 is that $H_{yy}(x,y)$ is decreasing (increasing) in its first argument.

Remark 4.3.10. Recall the definitions of F-MON↑ and F-MON↓ (Definition 3.4.2). If $H(u,y) = F(y,u)$ then $\Phi(u,y)$ is monotone increasing (decreasing) if F satisfies F-MON↑ (F-MON↓).

Example 4.3.11. *H^R and H^S are increasing kernels and H^L is a decreasing kernel. The kernels H^B and H^Q are simultaneously both increasing and decreasing since $\Phi^B(u,y) = 2u^{-2}$ and $\Phi^Q(u,y) = 2$ do not depend on y and Condition (ii) in Definition 4.3.8 is satisfied with equality in both cases.*

Example 4.3.12. *Consider the kernels $H^{G-}(u,y) = uH^R(u,y)$ and $H^{G+}(u,y) = yH^R(u,y)$. In the first case, variance is weighted by the pre-jump value of the price realisation and in the second case the variance is weighted by the post-jump value. Swaps of this type are known as Gamma swaps, see, for example, Carr and Lee [17]. Both H_{G-} and H_{G+} are increasing kernels.*

Theorem 4.3.13.

(i) (a) If H is an increasing kernel then every candidate payoff of Class $\mathcal{K}$ is the root of a semi-static sub-hedge for the kernel H.

(b) If H is an increasing kernel then every candidate payoff of Class $\mathcal{L}$ is the root of a semi-static super-hedge for the kernel H.

(ii) *(a) If H is a decreasing kernel then every candidate payoff of Class $\mathcal{L}$ is the root of a semi-static sub-hedge for the kernel H.*

(b) If H is a decreasing kernel then every candidate payoff of Class $\mathcal{K}$ is the root of a semi-static super-hedge for the kernel H.

Proof. We will prove the theorem in the case (i)(a). The proofs in the other cases are similar.

Fix $\kappa \in \mathcal{K}$ let $L_\kappa(x,y) = \psi_\kappa(x) + \psi'_\kappa(x)(y-x) + H(x,y) - \psi_\kappa(y)$. The result will follow if we can show that $L_\kappa(x,y) \geq 0$ for all $(x,y) \in (0,\infty)^2$. Since κ is fixed we drop the subscript κ in what follows.

Suppose that $x, z > f(0)$ and $y \in (0,\infty)$. Since $\psi(x) + \psi'(x)(y-x) = \int_{f(0)}^{x}(y-u)\Phi(u,\kappa(u))du$ we have that

$$\begin{aligned}
L(x,y) - L(z,y) &= \psi(x) + \psi'(x)(y-x) + H(x,y) - \psi(z) - \psi'(z)(y-z) - H(z,y) \\
&= \int_z^x \{(y-u)\Phi(u,\kappa(u)) + H_x(u,y)\}\, du \\
&= \int_z^x \{\Phi(u,y) - \Phi(u,\kappa(u))\}(u-y)du.
\end{aligned}$$

If $y \geq f(0)$, then set $z = y$ to find that

$$L(x,y) = \int_y^x \{\Phi(u,y) - \Phi(u,\kappa(u))\}(u-y)du.$$

Since $y \geq f(0) \geq \kappa(u)$, $\Phi(u,y) \geq \Phi(u,\kappa(u))$ for all u. Hence $L(x,y) \geq 0$ with equality at $y = x$.

If $y < f(0)$ and k is continuous at y set $z = k(y)$. Otherwise, for definiteness set $z = k(y+)$. Then $L(k(y+),y) = 0$ and

$$L(x,y) = \int_{k(y+)}^x \{\Phi(u,y) - \Phi(u,\kappa(u))\}(u-y)du.$$

If $k(y+) \leq x$ then $y \geq \hat{x}$, for all $\hat{x} \in [\kappa(x+),\kappa(x-)]$. Then for $u \in (k(y+),x)$, $\kappa(u) \leq y$ and since $\Phi(u,z)$ is increasing in z, the integrand is positive.

If $x < k(y+)$, then $y < \hat{x}$ for all $\hat{x} \in [\kappa(x+),\kappa(x-)]$. Then for $u \in (x,k(y+))$ we have $\kappa(u) > y$. Then again $L(x,y) \geq 0$.

Finally, we show that $L(x,y) \geq 0$ when $x < f(0)$. Note that since, by what we have shown above, $L(k(x),y) \geq 0$ it will suffice to show that $L(x,y) \geq L(k(x),y)$. But,

$$\begin{aligned}
L(x,y) - L(k(x),y) &= \psi(x) + \psi'(x)(y-x) + H(x,y) \\
&\quad -\psi(k(x)) - \psi'(k(x))(y-k(x)) - H(k(x),y) \\
&= \psi(k(x)) + \psi'(k(x))(x-k(x)) + H(k(x),x) + \psi'(k(x))(y-x) \\
&\quad +H_y(k(x),x)(y-x) + H(x,y) - \psi(k(x)) - \psi'(k(x))(y-k(x)) - H(k(x),y) \\
&= H(k(x),x) + H(x,y) + H_y(k(x),x)(y-x) - H(k(x),y) \\
&\geq 0,
\end{aligned}$$

where the last inequality follows from Definition (4.3.8). □

Remark 4.3.14. The proof of Theorem 4.3.13 proceeds exactly like the proof of Lemma 3.4.6 in Chapter 3, apart from the final step to show that $L(x,y) > 0$ when $x < f(0)$. This step is necessary here to ensure that the inequality (4.3.1) holds for all price paths, i.e for all pairs $(x,y) \in \mathbb{R}$. The results in Chapter 3 on the other hand are only concerned with the terminal value of a stopped Brownian motion and its running maximum and the corresponding inequality need only hold for pairs (w,s) with $w \leq s$.

4.4 The most expensive sub-hedge

In the next three sections we concentrate on lower bounds and increasing variance kernels, but there are equivalent results for upper bounds and/or decreasing variance kernels.

In this section we fix the call prices and attempt to identify the most expensive sub-hedge from the set of sub-hedges generated by candidate payoffs of Class $\mathcal{K}$. The price of this sub-hedge provides a highest model-independent lower bound on the price of the variance swap in a sense which we will explain in the section on continuous limits.

Associated with the set of call prices $C(k)$ (and put prices $C(k) - f(0) + k$ given by put-call parity) there is a measure μ on $\mathbb{R}^+$ with mean m. Since f is a forward price we must have $f(0) = m$. Write $C = C_\mu$ to emphasise the connection between these quantities. Then $C(k) = C_\mu(k) = \int_k^\infty (x-k)\mu(dx)$. Recall that C_μ is convex so that $\mu(dx) = C''_\mu(x)dx$ with the right-hand-side to be interpreted in a distributional sense as necessary. We wish to calculate the cost of the European claim which forms part of the semi-static sub-hedge. By construction this is equal to $\int_{\mathbb{R}^+} \psi(x)\mu(dx) = \int_0^m \psi''(z)(C_\mu(z) - m + z)dz + \int_m^\infty \psi''(x)C_\mu(x)dx$.

Proposition 4.4.1. *For H a variance swap kernel and $\kappa \in \mathcal{K}(m)$,*

$$\int_0^\infty \psi_\kappa(x)\mu(dx) = \int_{[0,m)} \mu(dz)H(m,z) + \int_m^\infty du\Sigma_\mu^{(u)}(\kappa(u)) \tag{4.4.1}$$

where, for $v < m < u$,

$$\Sigma_\mu^{(u)}(v) = \Phi(u,v)C_\mu(u) + \int_{[0,v)} \mu(dz)(u-z)\left\{\Phi(u,z) - \Phi(u,v)\right\}.$$

Proof. Let $\psi = \psi_\kappa$. Note that by definition $\psi(m) = 0$, so there is no contribution from mass at m and we can divide the integral on the left of (4.4.1) into intervals $[0,m)$ and (m,∞). For the latter,

$$\begin{aligned}\int_m^\infty \psi(x)\mu(dx) &= \int_m^\infty \mu(dx)\int_m^x (x-u)\Phi(u,\kappa(u))du \\ &= \int_{u=m}^\infty du\Phi(u,\kappa(u))\int_u^\infty (x-u)\mu(dx) \\ &= \int_{u=m}^\infty du\Phi(u,\kappa(u))C_\mu(u) =: I_1.\end{aligned}$$

Now consider $\int_{[0,m)}\psi(z)\mu(dz)$. For this, using $H(k,z) = H(m,z) + \int_m^k H_x(u,z)du$ and $\psi(x) + \psi'(x)(z-x) = \int_m^x du(z-u)\Phi(u,\kappa(u))$ we have

$$\begin{aligned}\int_{[0,m)} \psi(z)\mu(dz) &= \int_{[0,m)} \mu(dz)H(m,z) + \int_{[0,m)} \mu(dz)\int_m^{k(z)} du(u-z)\left\{\Phi(u,z) - \Phi(u,\kappa(u))\right\} \\ &=: I_2 + I_3\end{aligned}$$

Note that I_2 depends on H but not on κ. Moreover, I_3 does not depend on the particular values chosen for the inverse taken over intervals of constancy of κ. (If $x < \tilde{x}$ are a pair of possible values for $k(z)$ then $\int_x^{\tilde{x}} du(u-z)\{\Phi(u,z) - \Phi(u,\kappa(u))\} = 0$ since over this range $\kappa(u) = z$.) Changing the order of integration we have

$$I_3 = \int_m^\infty du \int_{[0,\kappa(u))} \mu(dz)(u-z)\left\{\Phi(u,z) - \Phi(u,\kappa(u))\right\},$$

and then $I_1 + I_3 = \int_m^\infty du\Sigma_\mu^{(u)}(\kappa(u))$. □

Our goal is to maximise the expression (4.4.1) over decreasing functions $\kappa \in \mathcal{K}$. As noted above, I_2 is independent of κ, and to maximise $\int_m^\infty du\Sigma_\mu^{(u)}(\kappa(u))$ we can maximise $\Sigma_\mu^{(u)}(\kappa)$ separately for each $u > m$, and then check that the maximiser is a decreasing function of u.

Proposition 4.4.2. *Suppose H is an increasing variance swap kernel. Then $\int_0^\infty \psi_\kappa(x)\mu(dx)$*

is maximised over $\kappa \in \mathcal{K}$ by $\kappa = \alpha^+$ where α^+ is the quantity which arises in (4.2.6) in the definition of the Perkins solution to the Skorokhod embedding problem.

Proof. For $u > m$ consider $\Theta_\mu^{(u)}(v) := C_\mu(v) - \int_{[0,v)} \mu(dz)(u-z)$ defined for $v \in [0, m]$. Then for each u, $\Theta_\mu^{(u)}$ is a strictly decreasing right-continuous function taking both positive and negative values on $[0, m]$, $\Theta_\mu^{(u)}(0) = C_\mu(u) \geq 0$ and $\Theta_\mu^{(u)}(m) = m - u + \int_m^\infty (u-m)\mu(dz) < 0$. Let $\overline{\kappa} = \overline{\kappa}(u) = \sup\{v : \Theta_\mu^{(u)}(v) \geq 0\}$. We have $\Theta_\mu^{(u)}(\overline{\kappa}(u-)) \geq 0 \geq \Theta_\mu^{(u)}(\overline{\kappa}(u+))$.

Suppose H is an increasing variance swap kernel so that $\Phi(u, y)$ is increasing in y. We want to show that $\Sigma_\mu^{(u)}(v)$ is maximised by $v = \overline{\kappa}(u)$.

Suppose $m > v > \overline{\kappa}(u)$. We aim to show that for all $\kappa \in (\overline{\kappa}(u), v)$ we have $\Sigma_\mu^{(u)}(v) \leq \Sigma_\mu^{(u)}(\kappa)$. We have

$$\begin{aligned}
\Sigma_\mu^{(u)}(v) - \Sigma_\mu^{(u)}(\kappa) &= \Phi(u,v)C_\mu(u) + \int_{[0,v)} \mu(dz)(u-z)\{\Phi(u,z) - \Phi(u,v)\} \\
&\quad -\Phi(u,\kappa)C_\mu(u) - \int_{[0,\kappa)} \mu(dz)(u-z)\{\Phi(u,z) - \Phi(u,\kappa)\} \\
&= \int_{[\kappa,v)} \mu(dz)(u-z)\{\Phi(u,z) - \Phi(u,v)\} + [\Phi(u,v) - \Phi(u,\kappa)]\,\Theta_\mu^{(u)}(\kappa).
\end{aligned}$$

Since H is an increasing variance kernel, for $z \in (\kappa, v)$, $\Phi(u,z) \leq \Phi(u,v)$, and the first integral is non-positive. Furthermore, $\Phi(u,v) \geq \Phi(u,\kappa)$ and $\Theta^{(u)}(\kappa) < 0$. Hence we conclude that $\Sigma_\mu^{(u)}(v) \leq \Sigma_\mu^{(u)}(\kappa)$.

Similar arguments show that if $v < \overline{\kappa}(u)$ then $\Sigma_\mu^{(u)}(v) \leq \Sigma_\mu^{(u)}(\kappa)$ for any $\kappa \in (v, \overline{\kappa}(u))$, and it follows that $\kappa = \overline{\kappa}(u)$ is a maximiser of $\Sigma_\mu^{(u)}(v)$.

Note that $\overline{\kappa}(u)$ is precisely the quantity α^+ which arises in the Perkins construction. Hence $\overline{\kappa}$ is a decreasing function. Moreover, the definition $\overline{\kappa}(u) = \sup\{v : \Theta_\mu^{(u)}(v) \geq 0\}$ ensures that $\overline{\kappa}$ is right continuous. □

Corollary 4.4.3. *Suppose $\kappa_n(x)$ is a sequence of elements of $\mathcal{K}$ with $\kappa_n(x) \downarrow \overline{\kappa}(x)$. Then $\int_{[0,\infty)} \psi_{\kappa_n}(x)\mu(dx)$ converges monotonically to $\int_{[0,\infty)} \psi_{\overline{\kappa}}(x)\mu(dx)$.*

Proof. Recall that $\int_{[0,\infty)} \psi_\kappa(x)\mu(dx) = \int_0^1 \mu(dz)H(1,z) + \int_1^\infty du\Sigma_\mu^{(u)}(\kappa(u))$. By the above arguments we have that $\Sigma_\mu^{(u)}(z)$ is increasing in z for $z > \overline{\kappa}(u)$. Hence the result follows by monotone convergence. □

Example 4.4.4. *Let $H = H^R$, an increasing variance kernel. Let $\mu = U[0,2]$ and let $\kappa : [1,2] \to [0,1]$ be given by $\kappa(x) = \alpha_\mu^+(x) = x - 2\sqrt{x-1}$. Similarly we define $\ell : [0,1] \to [1,2]$ by $\ell(x) = \alpha_\mu^-(x) = x + 2\sqrt{1-x}$. Then $(\psi_\kappa, -\psi_\kappa')$ is the most expensive sub-hedge of class $\mathcal{K}$ and $(\psi_\ell, -\psi_\ell')$ is the cheapest super-hedge of class $\mathcal{L}$. Although we cannot calculate the functions ψ_κ, ψ_ℓ explicitly, they can be evaluated numerically, see the left hand side of Figure 2. Now* *suppose $H = H^L$. The roles of ψ_κ and ψ_ℓ are reversed (see the right hand side of Figure 2) and ψ_κ is the root of a semi-static super-hedge and ψ_ℓ is the root of a semi-static sub-hedge.*

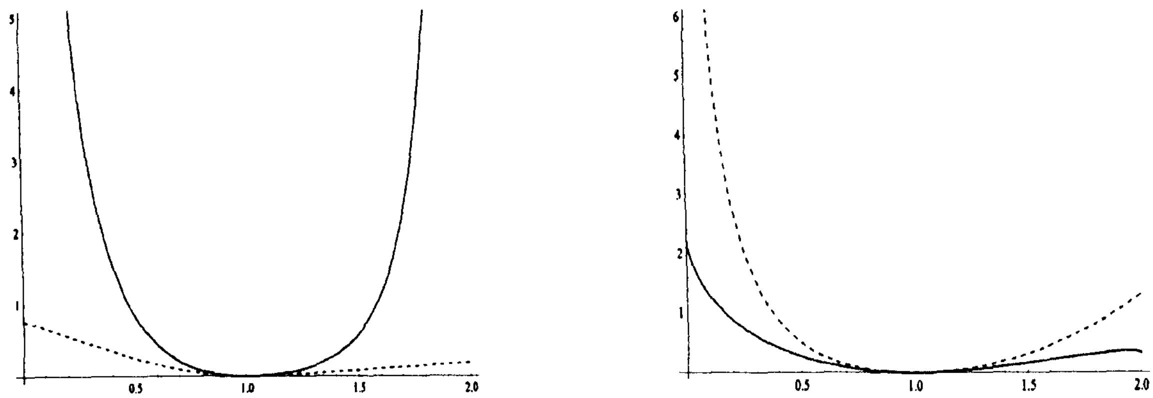

Figure 4.2: For the two kernels ψ_κ is shown as a dashed line and ψ_ℓ is shown as a solid line. For the kernel H^R (left-hand-side), ψ_κ is associated with a lower bound on the price of the variance swap. For the kernel H^L (right-hand-side) ψ_κ is associated with an upper bound.

4.5 Continuous limits and the tightness of the bound

The bounds we have constructed based on the functions ψ_κ hold simultaneously across all paths and all partitions. The purpose of this section is to consider the limit as the partition becomes finer. It will turn out that in the continuous limit there is a stochastic model which is consistent with the observed call prices and for which there is equality in the inequality (4.3.1) from which we derive the lower bound. In this sense the model-free bound is optimal, and can be attained.

The analysis of this section justifies restricting attention to candidate payoffs of classes $\mathcal{K}$ and $\mathcal{L}$. Hedges of this type either sub-replicate or super-replicate the payoff of the variance swap depending on the form of the kernel, but there could be other sub- and super-replicating strategies which do not take this form. In principle, for a given partition one of these other sub-hedges could give a tighter model-independent bound than we can derive from our analysis. (As an extreme example, suppose the partition is trivial ($0 = t_0 < t_1 = T$). Then $V_H(f, P) = H(f(0), f(T))$ which can be replicated exactly using call options.) However, in the continuous limit our bound is best possible, so that when the partition is finite, but the mesh size is small we expect our hedge to be close to best possible and relatively simple to implement.

Recalling the construction of the functions ψ_κ for an increasing kernel H we have for a finite partition $P^{(n)}$ in the dense sequence $\mathcal{P} = (P^{(n)})_{n\geq 1}$ that

$$V_H(f, P^{(n)}) = \sum_{k=0}^{N^{(n)}-1} H(f(t_k), f(t_{k+1})) \geq \psi(f(T)) - \psi(f(0)) - \sum_{k=0}^{N^{(n)}-1} \psi'(f(t_k))(f(t_{k+1}) - f(t_k)). \tag{4.5.1}$$

We want to conclude that the limits $V_H(f, P_\infty) = \lim_n V_H(f, P^{(n)})$ and

$$\lim_n \sum_{k=0}^{N^{(n)}-1} \psi'(f(t_k))(f(t_{k+1}) - f(t_k)) = \int_0^T \psi'(f(t-))df(t) \tag{4.5.2}$$

exist for each path under consideration. Our analysis follows the development of a path-wise Itô's formula in Föllmer [34]. Let ϵ_t denote a point mass at t.

Definition 4.5.1. A path realisation f has a quadratic variation on a dense sequence of partitions $\mathcal{P} = (P^{(n)})_{n\geq 1}$ if, when we define the measure

$$\zeta_n = \sum_{k=0,\ t_k\in P^{(n)}}^{N^{(n)}-1} (f(t_{k+1}) - f(t_k))^2 \epsilon_{t_k},$$

then the sequence ζ_n converges weakly to a Radon measure ζ on $[0, T]$. Then $([f]_t)_{t\geq 0}$ is given by $[f]_t = \zeta([0, t])$.

The atomic part of ζ is given by squared jumps of f. Moreover the quadratic variation $([f]_t)_{t\geq 0}$ is simply the cumulative mass function of ζ.

Theorem 4.5.2. *(Föllmer [34]) Suppose the price realisation f has a quadratic variation along $\mathcal{P} = (P^{(n)})_{n\geq 1}$ and G is a twice continuously differentiable function from $\mathbb{R}^+$ to $\mathbb{R}$, then*

$$\int_0^T G'(f(t-))df(t) = \lim_{n\uparrow\infty} \sum_{t=0}^{N^{(n)}-1} G'(f(t_k))(f(t_{k+1}) - f(t_k))$$

exists and

$$\begin{aligned} G(f(T)) - G(f(0)) \quad = \quad & \int_0^T G'(f(s-))df(s) + \frac{1}{2}\int_{(0,T]} G''(f(s))d[f]_s^c \\ & + \sum_{s\leq T} \left[G(f(s)) - G(f(s-)) - G'(f(s-))\Delta f(s)\right], \end{aligned}$$

and the series of jump terms is absolutely convergent.

Hence, provided ψ is twice continuously differentiable on the support of f and f has a quadratic variation along $\mathcal{P}$, it follows immediately that the limit in (4.5.2) exists. In our setting

$\psi_\kappa''(u) = \Phi(u, \kappa(u))$ for $u > 1$, so that a sufficient condition for $\psi_\kappa''(u)$ to be continuous on $(1, \infty)$ is that κ is continuous. Further, on $u < 1$, provided $k \equiv \kappa^{-1}$ is differentiable and H_y exists, we have $\psi'(z) = \psi'(k(z)) + H_y(k(z), z)$. Hence, sufficient conditions for ψ to be twice continuously differentiable on $(0,1)$ are that k is continuously differentiable, κ is continuous and H_{xy} and H_{yy} are continuous. Let $\mathcal{K}_c$ be the class of decreasing functions $\kappa : (f(0), \infty) \to (0, f(0))$ which are continuous and have an inverse k which is continuously differentiable.

Corollary 4.5.3. *Suppose that H is an increasing variance kernel, and that f has a quadratic variation along a dense sequence of partitions $\mathcal{P} = (P^{(n)})_{n\geq 1}$. Suppose $\kappa \in \mathcal{K}_c$ and $\psi = \psi_\kappa$. Then the limit in (4.5.2) exists.*

Now we want to consider $V_H(f, P_\infty) = \lim_n V_H(f, P^{(n)})$.

Lemma 4.5.4. *Suppose H is a variance swap kernel. If $\mathcal{P} = (P^{(n)})_{n\geq 1}$ is a dense sequence of partitions, and f has a quadratic variation along $\mathcal{P}$, then $\lim_{n\uparrow\infty} V_H(f, P^{(n)})$ exists and satisfies*

$$V_H(f, P_\infty) = \int_{(0,T]} \frac{1}{f(t-)^2} d[f]_t + \sum_{0<t\leq T} H(f(t-), f(t)) - \sum_{0<t\leq T} \frac{1}{f(t-)^2} (\Delta f(t))^2. \qquad (4.5.3)$$

Proof. Our proof follows Föllmer [34]. Fix $\epsilon > 0$. Partition $[0,T]$ into two classes: a finite class $C_1 = C_1(\epsilon)$ of jump times and a class $C_2 = C_2(\epsilon)$ such that

$$\sum_{s\in[0,T],\ s\in C_2(\epsilon)} (\Delta f(s))^2 \leq \epsilon^2. \qquad (4.5.4)$$

Then $\sum_{k=0}^{N^{(n)}-1} H(f(t_k), f(t_{k+1})) = \sum_1 H(f(t_k), f(t_{k+1})) + \sum_2 H(f(t_k), f(t_{k+1}))$, where $\sum_1$ indicates a sum over those $0 \leq k \leq N^{(n)} - 1$ for which $(t_k, t_{k+1}]$ contains a jump of class C_1. It follows that

$$\lim_{n\uparrow\infty} \sum_1 H(f(t_k), f(t_{k+1})) = \sum_{t\in C_1(\epsilon)} H(f(t-), f(t)). \qquad (4.5.5)$$

On the other hand, using the properties $H(x,x) = 0$, $H_y(x,x) = 0$ we have from Taylor's formula that $H(x,y) = \frac{1}{2} H_{yy}(x,x)(y-x)^2 + r(x,y)$. Using the fact that $(f(t))_{0\leq t\leq T}$ is a compact subset of $(0,\infty)$ we may assume that the remainder term satisfies $|r(x,y)| \leq R(|y-x|)(y-x)^2$ where

R is an increasing function on $[0,\infty)$ such that $R(c) \to 0$ as $c \to 0$. Then

$$\begin{aligned}
\sum_2 H(f(t_k), f(t_{k+1})) &= \frac{1}{2}\sum_2 H_{yy}(f(t_k), f(t_k))(f(t_{k+1}) - f(t_k))^2 + \sum_2 r(f(t_k), f(t_{k+1})) \\
&= \frac{1}{2}\sum_2 H_{yy}(f(t_k), f(t_k))(f(t_{k+1}) - f(t_k))^2 \\
&\quad -\frac{1}{2}\sum_1 H_{yy}(f(t_k), f(t_k))(f(t_{k+1}) - f(t_k))^2 \\
&\quad +\sum_2 r(f(t_k), f(t_{k+1})).
\end{aligned} \tag{4.5.6}$$

Since $H_{yy}(f, f) = 2/f^2$ is uniformly continuous over the bounded set of values $(f(t))_{0\leq t\leq T}$, by (9) in Föllmer [34], the first term in (4.5.6) converges to $\int_{(0,T]} \frac{1}{f(t-)^2} d[f]_t$ and the second term converges to $-\sum_{s\in C_1} \frac{1}{f(t-)^2}(\Delta f(t))^2$. Using (4.5.4) and the fact that the remainder term satisfies $|r(x,y)| \leq R(|y-x|)(y-x)^2$ we have that the last term is bounded by $R(\epsilon)[f]_T$. Finally, letting $\epsilon \downarrow 0$ we conclude that $V_H(f, P_\infty) = \lim_n V_H(f, P^{(n)})$ exists and (4.5.3) follows. □

Corollary 4.5.5. *$V_{HR}(f, P_\infty) = \int_{(0,T]} f(t-)^{-2} d[f]_t$ and $V_{HL}(f, P_\infty) = [\log f]_T$.*

Combining (4.5.1) with Theorem 4.5.2 and Lemma 4.5.4 it follows that for a path of finite quadratic variation and ψ a twice-continuously differentiable function with $\psi(f(0)) = 0$,

$$V_H(f, P_\infty) \geq \psi(f(T)) - \int_0^T \psi'(f(t-))df(t). \tag{4.5.7}$$

The left hand side is the payoff of the variance swap in the continuous limit. The expression on the right can be interpreted as the payoff of a semi-static hedging strategy $(\psi, -\psi')$ under continuous trading. From Definition 4.1.7 for each of the partitions in the sequence we have that the price of the semi-static hedge is

$$\int_0^\infty \psi(x)\mu(dx) = \int_{f(0)}^\infty \psi''(x)C_\mu(x)dx + \int_0^{f(0)} \psi''(z)(C_\mu(z) + f(0) - z)dz. \tag{4.5.8}$$

Since this value does not depend on the partition, in the continuous-time setting we define the price of sub-hedge $(\psi, -\psi')$ to also be the expression given in (4.5.8).

Corollary 4.5.6. *Suppose H is an increasing variance swap kernel. A model-independent lower bound on the price of the continuous time limit of the variance swap with payoff $V_H(f)$ is*

$$\sup_\kappa \int_0^\infty \psi_\kappa(x)\mu(dx) = \int_0^\infty \psi_{\alpha_\mu^+}(x)\mu(dx) \tag{4.5.9}$$

where α_μ^+ is the quantity arises in the Perkins embedding (Theorem 4.2.2).

Proof. For any decreasing function $\kappa \in \mathcal{K}_c$ we can construct ψ_κ such that $\int_0^\infty \psi_\kappa(x)\mu(dx)$ is the price of a sub-hedge for V_H for any partition, and this continues to hold in the continuous-time limit. Moreover, by optimising over κ we obtain a bound $\int_0^\infty \psi_{\alpha_\mu^+}(x)\mu(dx)$ which is the best bound of this form by Proposition 4.4.2. Note that even if α_μ^+ is not in class $\mathcal{K}_c$, by Corollary 4.4.3 we can approximate it from above by a sequence of elements of class $\mathcal{K}_c$ such that in the limit we obtain the price $\int_0^\infty \psi_{\alpha_\mu^+}(x)\mu(dx)$ as a bound. □

Our goal now is to show that this is a best bound in general and not just an optimal bound based on inequalities such as (4.5.1) for $\psi \equiv \psi_\kappa$ and κ a decreasing function. We do this by showing that there is a consistent model for which the price of the continuously monitored variance swap is equal to $\int_0^\infty \psi_{\alpha_\mu^+}(x)\mu(dx)$.

Theorem 4.5.7. *There exists a consistent model such that*

$$V_H((X_t)_{0\leq t\leq T}, P_\infty) = \psi_{\alpha_\mu}(X_T) - \int_0^T \psi'_{\alpha_\mu}(X_{s-})dX_s. \tag{4.5.10}$$

Proof. Recall Definition 4.1.9 and note that we are given a set of call prices and that in constructing a consistent model we are free to design an appropriate probability space $(\Omega, \mathcal{F}, \mathbb{F} = (\mathcal{F}_t)_{0\leq t\leq T}, \mathbb{P})$ as well as a stochastic process $(X_t)_{t\geq 0}$.

Suppose we are given call prices $C(x) = C_\mu(x)$ for some μ. Let $(\Omega, \mathcal{G}, \mathbb{G} = (\mathcal{G}_t)_{0\leq t\leq T}, \mathbb{P})$ support a Brownian motion $(W_u)_{u\geq 0}$ with initial value $W_0 = f(0) = \int_{\mathbb{R}^+} x\mu(dx)$ and suppose $\mathcal{G}_0$ contains a $U[0,1]$ random variable which is independent of W. (This last condition is necessary purely to ensure that the Perkins embedding of μ can be defined when μ has an atom at $f(0)$. If μ has no atom at $f(0)$ then we may take $\mathcal{G}_0$ to be trivial.)

Let τ_μ^P be the Perkins embedding of μ in W. Write S for the maximum process of W so that $S_u = \max_{v\leq u} W_v$. Write $\overline{H}_x$ for the first hitting time by W of x. Let $(\Lambda(t))_{0\leq t\leq T}$ be a strictly increasing continuous function with $\Lambda(0) = f(0)$ and $\lim_{t\uparrow T}\Lambda(t) = \infty$. Now define the left-continuous process $\tilde{X} = (\tilde{X}_t)_{0\leq t\leq T}$ via

$$\tilde{X}_t = \begin{cases} \Lambda(t) & \overline{H}_{\Lambda(t)} \leq \tau_\mu^P \\ W_{\tau_\mu^P} & \tau_\mu^P < \overline{H}_{\Lambda(t)}. \end{cases}$$

Note that the condition $\overline{H}_{\Lambda(t)} \leq \tau_\mu^P$ can be re-written as $\Lambda(t) \leq S_{\tau_\mu^P}$ or equivalently $t \leq \Lambda^{-1}(S_{\tau_\mu^P})$. Define also $\tilde{\mathcal{F}}_t = \mathcal{G}_{\overline{H}_{\Lambda(t)}}$. Then $\tilde{X}$ is adapted to the filtration $\tilde{\mathbb{F}} = (\tilde{\mathcal{F}}_t)_{0\leq t\leq T}$ and $\tilde{X}$ is a $\tilde{\mathbb{F}}$-martingale for which $\tilde{X}_T = W_{\tau_\mu^P} \sim \mu$.

In order to construct a right-continuous martingale with the same properties, for $t < T$ we set $\mathcal{F}_t = \cap_{u>t}\tilde{\mathcal{F}}_t$ and $X_t = \lim_{u\downarrow t}\tilde{X}_u$, and for $t = T$ we set $\mathcal{F}_T = \tilde{\mathcal{F}}_t$ and $X_T = \tilde{X}_T$. Then X

is a right-continuous $\mathcal{F}$-martingale such that $(\Omega, \mathcal{F}, \mathbb{F} = (\mathcal{F}_t)_{0 \leq t \leq T}, \mathbb{P})$ is a consistent model.

Now we want to show that for this model (4.5.10) holds path-wise. Writing ψ for ψ_{α_μ}, and X_t as shorthand for each $X_t(\omega)$ we have for each ω

$$\begin{aligned}
\psi(X_T) - \int_0^T \psi'(X_{t-})dX_t &= \psi(W_{\tau_\mu^P}) - \int_{t=0}^{\Lambda^{-1}(S_{\tau_\mu^P})} \psi'(\Lambda(t))d\Lambda(t) - \psi'(S_{\tau_\mu^P})(W_{\tau_\mu^P} - S_{\tau_\mu^P}) \\
&= \psi(W_{\tau_\mu^P}) - \int_{f(0)}^{S_{\tau_\mu^P}} \psi'(u)du - \psi'(S_{\tau_\mu^P})(W_{\tau_\mu^P} - S_{\tau_\mu^P}) \\
&= \psi(W_{\tau_\mu^P}) - \psi(S_{\tau_\mu^P}) - \psi'(S_{\tau_\mu^P})(W_{\tau_\mu^P} - S_{\tau_\mu^P}).
\end{aligned}$$

There are two cases. Either $W_{\tau_\mu^P} = S_{\tau_\mu^P}$, in which case this expression is equal to 0 or, $W_{\tau_\mu^P} = \alpha_\mu(S_{\tau_\mu^P})$ and then the expression becomes

$$\psi(\alpha_\mu(s)) - \psi(s) - \psi'(s)(\alpha_\mu(s) - s) \equiv H(s, \alpha(s))$$

at $s = S_{\tau_\mu^P}$, using Definition 4.3.3. In either case the right hand side of (4.5.10) is $H(S_{\tau_\mu^P}, W_{\tau_\mu^P})$. For the left hand side of (4.5.10), $[X]_T^c = 0$ and $(\Delta X_u)^2 = (S_{\tau_\mu^P} - W_{\tau_\mu^P})^2 1_{\{u = \Lambda^{-1}(S_{\tau_\mu^P})\}} 1_{\{W_{\tau_\mu^P} \neq S_{\tau_\mu^P}\}}$ so that from (4.5.3), $V_H(f, P_\infty) = H(S_{\tau_\mu^P}, W_{\tau_\mu^P})$. Hence (4.5.10) holds path-wise.

□

Corollary 4.5.8. *Suppose H is an increasing variance swap kernel. Then the highest model independent lower bound on the price of a variance swap which is valid across all partitions is given by the expression in (4.5.9).*

Corollary 4.5.9. *If $\Phi(u, y)$ does not depend on y then the corresponding variance swap is perfectly replicable by $(\psi, -\psi')$. For all consistent models the variation swap has price $\int_{\mathbb{R}^+} \psi(x)\mu(dx)$.*

Example 4.5.10. *Recall the definitions of the kernels H^B and H^Q and Example 4.3.11. $\Phi^B(u, y) = 2u^{-2}$ and so $\psi'(u) = -2/u$ and $\psi(u) = -2\log(u)$. Thus $H^B(x, y) = \psi(y) - \psi(x) - \psi'(x)(y - x)$ and the strategy $(\psi, -\psi')$ replicates the payoff perfectly for any price realisation. The observation that H^B has one model-independent price was first made by Bondarenko in [11]. Similarly, $H^Q(x, y) = \psi(y) - \psi(x) - \psi'(x)(y - x)$, where $\psi(x) = x^2$. An alternative analysis of these two payoffs is due to Neuberger [59]. Neuberger introduces the aggregation property. Translated into the notation of our setting, a kernel enjoys the aggregation property if $\mathbb{E}[V_H(X, P^{(n)})] = \mathbb{E}[H(X_T - X_0)]$. Both Bondarenko [11] and Neuberger [59] advocate the use of H^B due to the fact that its price is not sensitive to the price path, but only to the value of X_T.*

4.6 Non-zero interest rates

To date we have worked with forward prices. This has the implication that the dynamic part of a hedging strategy has zero cost. In this section we outline how our analysis can be extended to non-zero, but deterministic, interest rates.

Suppose that interest rates are deterministic. Let $D_t = D_t(T)$ be the discount factor over $[t,T]$ so that the asset price realisation ($s = (s_t)_{0\leq t\leq T}$) and the forward price realisation are related by $s(t) = D_t f(t)$. In the case of constant interest rates $D_t(T) = e^{-r(T-t)}$ so that $s(t) = e^{-r(T-t)}f(t)$.

Let P be a partition of $[0,T]$. For $k \in \{0,1,...,N-1\}$ write $s_k = s(t_k)$, $f_k = f(t_k)$ and $D_k = D_{t_k}(T)$. Set $D_{k,k+1} = D_{k+1}/D_k$. Note that if interest rates are non-negative then $D_{k,k+1} \geq 1$.

Let G be the kernel of a variation swap and write $G_k(x,y) = G(D_k x, D_k y)$. Then the payoff of the variance swap is given by

$$V_G(s,P) = \sum_{k=0}^{N-1} G(D_k f_k, D_{k+1} f_{k+1}) = \sum_{k=0}^{N-1} G_k(f_k, D_{k,k+1} f_{k+1}).$$

Proposition 4.6.1. *Suppose that there exists a variation swap kernel H, functions η, ϵ, B and a constant $A \in \mathbb{R}$ such that for all $D > 0$*

$$G_k(x, yD) \geq AH(x,y) + \eta(y) - \eta(x) + \epsilon(x,k,D)(y-x) + B(k,D). \qquad (4.6.1)$$

Without loss of generality we may take $\eta(f(0)) = 0$.

Suppose that there exists a semi-static sub-hedging strategy (ψ, Δ) for the variation swap with kernel H. Then

$$V_G(s,P) \geq (A\psi + \eta)(f(T)) + \sum_k [\epsilon(f_k, k, D_{k,k+1}) + \delta_{t_k}(f(t)_{t\leq t_k})](f_{k+1} - f_k) + \sum_k B(k, D_{k,k+1}),$$

and there is a model-independent sub-hedge and price lower bound for V_G.

Proof. We have

$$\begin{aligned}
V_G(s,P) &= \sum_{k=0}^{N-1} G_k(f_k, D_{k,k+1}f_{k+1}) \\
&\geq \sum_k [AH(f_k,f_{k+1}) + \eta(f_{k+1}) - \eta(f_k) + \epsilon(f_k,k,D_{k,k+1})(f_{k+1}-f_k) + B(k,D_{k,k+1})] \\
&\geq A[\psi(f(T)) + \sum_k \delta_{t_k}(f(t)_{t\leq t_k})(f_{k+1}-f_k)] + \eta(f(T)) \\
&\qquad + \sum_k \epsilon(f_k,k,D_{k,k+1})(f_{k+1}-f_k) + \sum_k B(k,D_{k,k+1})
\end{aligned}$$

□

Remark 4.6.2. If we assume that interest rates are non-negative then we only need (4.6.1) to hold for $D \geq 1$.

Remark 4.6.3. The price for the floating leg associated with the hedge is the price of the static vanilla portfolio with payoff $(A\psi + \eta)(f(T))$ plus the constant $\sum_{k=0}^{N-1} B(k, D_{k,k+1})$.

Corollary 4.6.4. *Suppose H is an increasing variance kernel, and ψ is of Class $\mathcal{K}$. If (4.6.1) holds then we have a path-wise sub-hedge and a model independent bound on the price of V_G.*

In the setting of increasing or decreasing variance kernels the bound in (4.6.2) will be tight provided $(\psi, -\psi')$ is a tight semi-static hedge for $V_H(f,P)$ and there is equality in Equation (4.6.1).

Example 4.6.5. *Suppose $G(x,y) = H^R(x,y) = \frac{(y-x)^2}{x^2}$. Then $G_k(x,y) = G(x,y)$, so that $\epsilon(x,k,D)$ and $B(k,D)$ will not depend on k. Moreover,*

$$\begin{aligned}
G(x,yD) &= \frac{1}{x^2}(Dy - Dx + Dx - x)^2 \\
&= D^2\left(\frac{y-x}{x}\right)^2 + D\frac{(D-1)}{x}(y-x) + (D-1)^2
\end{aligned}$$

Suppose that interest rates are non-negative so that $D_{k,k+1} \geq 1$. Then (4.6.1) holds for $A = 1$, $\eta = 0$, $\epsilon(x,D) = D(D-1)/x$ and $B(D) = (D-1)^2$.

Note that there is an inequality in (4.6.1) for $A = 1$. If $D_{k,k+1}$ is independent of k (the natural example is to assume that interest rates are constant and the partition is uniform, in which case $d = \log D_{k,k+1} = rT/N$) then we can have equality by taking $A = e^{2rT/N}$. In that case we have an improved bound, but the improvement becomes negligible in the limit $N \uparrow \infty$.

Example 4.6.6. *Suppose $G(x,y) = H^L(x,y) = (\log(y) - \log(x))^2$. Then $G_k(x,y) = G(x,y)$ and $G(x,yD) = (\log D + \log y - \log x)^2 = H^L(x,y) + 2\log D(\log y - \log x) + (\log D)^2$.*

Suppose now that the partition is such that $D_{k,k+1}$ is independent of k, and set $d = \log D_{k,k+1}$. Then Equation (4.6.1) holds with equality for $A = 1$, $\eta(y) = 2d \log y$, $\epsilon = 0$ and $B(D) = d^2$.

Example 4.6.7. *Suppose $G(x, y) = H^B(x, y) = -2(\log y - \log x) - (y/x - 1))$. Then $G_k(x, y) = G(x, y)$ and*

$$\begin{aligned} G(x, yD) &= -2(\log y - \log x + \log D) + 2D(y - x) + 2(D - 1) \\ &= H^B(x, y) + 2(D - 1)(y/x - 1) + H^B(1, D). \end{aligned}$$

Then Equation (4.6.1) holds with equality for $A = 1$, $\eta(y) = 0$, $\epsilon(x, D) = 2(D - 1)/x$, $B(D) = H^B(1, D)$.

We can consider the limit as the partition becomes dense, in which case the bounds for the variance swap become tight. For definiteness we will assume that we have a sequence of uniform partitions with mesh size tending to zero, and that interest rates are constant, though this can be weakened for the squared return and Bondarenko kernels.

Then, for each of the three examples above we have that $\sum_{k=0}^{N-1} B(k, D_{k,k+1}) = NB(e^{rT/N}) \to 0$. Further, in each case $\eta(y) \to 0$, and $A = 1$. Then in the limit the lower bound on the price of the variance swap based on the price realisation s is the same as the upper and lower bounds for the variance swap defined relative to the forward price f. Thus, for variance swaps based on frequent monitoring, the bounds we have calculated in earlier sections based on the forward price may also be used for undiscounted price processes.

4.6.1 Super-hedges and upper bounds

Corollary 4.6.8. *Suppose there exists H, η, ϵ, B, and A such that*

$$G_k(x, yD) \leq AH(x, y) + \eta(y) - \eta(x) + \epsilon(x, k, D)(y - x) + B(k, D), \tag{4.6.2}$$

and suppose that there exists a semi-static super-hedging strategy (ψ, Δ) for the variation swap with kernel H. Then there is a corresponding model-independent super-hedge and price upper bound for V_G.

The analysis of the kernels H^R, H^L, H^B and upper bounds is similar to that in Examples 4.6.5—4.6.7 above. For the kernel H^B, the choices listed in Example 4.6.7 give equality in (4.6.2) and can be used equally for upper bounds. Provided that we have an upper bound for $D_{k,k+1}$, so that $D_{k,k+1} \leq \bar{D}$ uniformly in k, for the kernel H^R we may take $A = \bar{D}^2$, $\eta = 0$,

$\epsilon(x,D) = D(D-1)/x$ and $B(D) = (D-1)^2$. Finally, for H^L, provided interest rates are non-negative, we can write

$$G(x,yD) = H^L(x,y) + 2\log D(\log y - \log x) + (\log D)^2 \leq H^L(x,y) + 2\frac{\log D}{x}(y-x) + (\log D)^2$$

so that (4.6.2) holds for $A=1$, $\eta=0$, $\epsilon(x,D) = 2(\log D)/x$ and $B(D) = (\log D)^2$. Note that, unlike for the lower bound in Example 4.6.6, for the upper bound we do not need to assume that $D_{k,k+1}$ is independent of k.

Remark 4.6.9. In his analysis of lower bounds for the kernel H^L, Kahalé [47] does not need to assume the partition is uniform and that interest rates are constant (or more generally that $D_{k,k+1}$ is constant), and can allow for arbitrary finite partitions and deterministic interest rates. Our results complement his results nicely. Although we need the assumption that $D_{k,k+1}$ is constant to recover Kahalé's result in the setting of lower bounds and the kernel H^L, in all other cases of study (upper bounds for V_{H^L} and upper and lower bounds for V_{H^R} and V_{H^B}) our methods also allow for arbitrary partitions and non-constant but deterministic interest rates.

4.7 Numerical Results

Given a continuum of call prices, it is possible to calculate the model independent bounds for the prices of variance swaps. When the implied terminal distribution of the asset price is sufficiently simple it is sometimes possible to calculate the monotone functions associated with the Perkins embedding explicitly (see Example 5.4) and to obtain a closed form integral expression for the model independent upper and lower bounds. For more realistic and complex target laws, the monotone functions and bounds can still be calculated numerically. The case when the terminal law is lognormally distributed is of particular practical interest.

A standard time frame for a volatility swap is 30 days or one month ($T = 1/12$), which is the time frame used for the widely quoted VIX index. Figure 4.3 plots the upper and lower bounds for the prices of variance swaps based on the kernels H_R and H_L relative to the cost of $-2\log$ contracts (the Neuberger/Dupire price of the standard hedge or 'VIX price') against the volatility parameter of the lognormal (terminal) distribution centred at 1. More precisely, the bounds are plots of

$$\sigma \rightarrow \mathbf{E}[\psi_{\kappa,H}(X_{\sigma/\sqrt{12}})]/\mathbf{E}[-2\log X_{\sigma/\sqrt{12}}], \qquad \text{and} \qquad \sigma \rightarrow \mathbf{E}[\psi_{\ell,H}(X_{\sigma/\sqrt{12}})]/\mathbf{E}[-2\log X_{\sigma/\sqrt{12}}],$$

where $X_\sigma \equiv e^{\sigma N - \sigma^2/2}$ is the lognormal random variable with volatility parameter σ and $H = H^R$ or H^L. Here, $\psi_{K,H}$ is the function given in Definition 4.3.3 and κ is chosen according to

Proposition 4.4.2 (with ℓ chosen similarly). Thus the upper bound for the kernel H_L and the lower bound for the kernel H_R correspond to the decreasing function κ associated with the Perkins embedding, while the other two bounds are constructed with the decreasing function ℓ associated with the reversed Perkins embedding.

Note that the price of a variance swap in the Black-Scholes model ($\mathbb{E}[-2\log X_{\sigma\sqrt{T}}] = \sigma^2 T$) is an increasing function of volatility. The upper and lower bounds are also increasing functions of volatility, and, as can be seen in the figure, they also become wider as volatility increases, when expressed as a ratio against the no-jump case. For reasonable values of volatility, and for both kernels, the impact of jumps is to affect the price by a factor of less than two, and for the kernel H^L the bounds are even tighter. The observation that the bounds for the kernel H_R are wider than those for the kernel H_L is partly explained by considering the leading term in the expansion of the hedging error (see Section 3.2). We have $J_R(x) \approx 2x^3/3$ whereas $J_L(x) \approx -x^3/3$ so that the magnitude of the leading error term for H_R is twice that of the leading error term for H_L. Note that for the optimal martingales the jumps are not local, so this approximation becomes less relevant as σ increases.

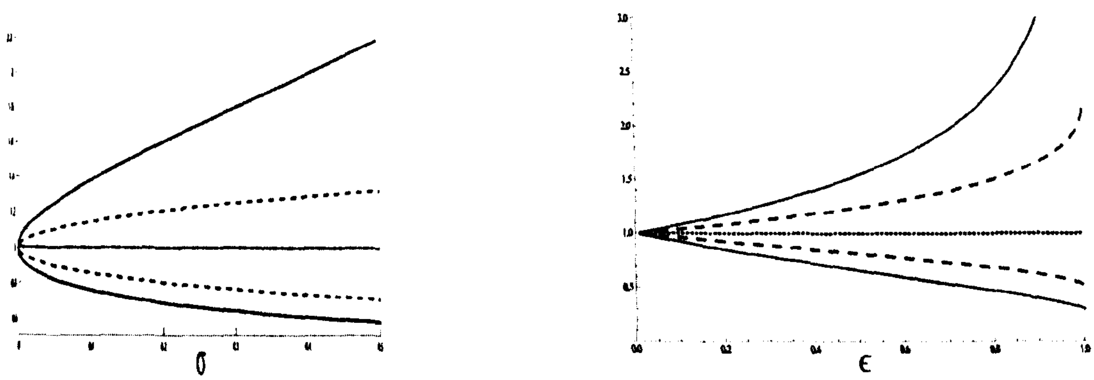

Figure 4.3: Model independent upper and lower bounds for the prices of variance swaps based on the kernels H_R (solid lines) and on H_L (dashed lines) relative to the price of $-2\log$ contracts (dotted line). There are two cases: on the left where the terminal distribution is lognormal with volatility between 0 and 0.5, (and $T = 1/12$), and on the right where the terminal distribution is uniform on $[1-\epsilon, 1+\epsilon]$, as ϵ ranges between 0 and 1. Here we work with variance swaps on forward prices

4.8 Further observations and remarks

4.8.1 Choice of kernel

One of the key contributions of the work presented in this chapter is that we investigate a variety of kernels. The ability to consider general kernels in the definition of the variance swap allows us to emphasise the dependence of the payoff on the presence and character of jumps, and to show that the nature of the dependence is strongly influenced by the form of the kernel.

Bondarenko [11] and Neuberger [59] argue that the finance industry should specialise to the kernel H^B as then variance swaps can be replicated perfectly, even in the presence of jumps, recall Example 4.5.10. The argument against this is that variance swaps are a useful instrument precisely because they are not redundant in this way. Sophisticated investors may want to take a position on the likely presence and direction of jumps. This is possible if the variance swap is defined using the kernel H^R or H^L, but not if it is defined using H^B.

4.8.2 Connections with the paper of Kahalé [47]

In his recent preprint Kahalé [47] investigates the same question that we discuss here, in the special case of the kernel H^L, and for lower bounds and sub-replicating strategies. Kahalé introduces the class of V-convex functions which have the property that each such function gives a lower bound on the price of the variance swap, and an associated sub-hedge. He then proceeds to show that functions ψ of Class $\mathcal{L}$ (in our notation) are V-convex. In this way he can deduce a lower bound on the price of a variance swap. Further, for a particular choice of decreasing function he can show that this lower bound can be attained in the continuous time limit under a well-chosen stochastic model — hence the bound he attains must be a best bound.

In contrast to Kahalé who works with the stock price, we being by considering contracts based on the forward price. This simplifies the analysis significantly and reduces the search for candidate sub-hedge payoffs to a search for functions satisfying (4.3.5). The condition (4.3.5) is simpler than the corresponding condition for V-convexity in Kahalé [47, Equation (3.1)]. The transparent representation of the key property allows us to find candidate super-hedge payoffs directly and to extend the analysis to general variation swap kernels provided they have a monotonicity property. The general framework presented here makes it possible to construct upper bounds to complement the lower bounds. Most importantly, we showed how Kahalé's contribution fits within an existing literature in which model-independent bounds are identified with Skorokhod embeddings.

Chapter 5

Links to classical optimal transport

The purpose of this chapter is to provide an overview of connections between the ideas in this thesis and in the theory surrounding the classical problem of optimal mass transport. These connections raise interesting questions for further work.

The body of literature surrounding mass transport is vast and we refer to the survey by Rüschendorf [69], and to the work of Evans and Gangbo [32] and Villani [75] for rigorous expositions of the classical problem. Our focus is restricted to the links between the classical problem in one dimension and the problem of calculating bounds for derivative values given marginal laws of the underlying asset price process.

5.1 Monge's earthwork problem

Mathematically, the problem of mass transport originates in a two-hundred year old 'earthwork' problem due to Monge [55]. The question is how to transport earth from a given area or deposit (the déblai) to another given area or target (the remblai) in a way that minimises the cost of carriage. Let us formulate the one-dimensional version of this problem. Denote the initial distribution of the 'earth deposit' (the starting law) by μ and the desired distribution at the target site by ν. Both μ and ν will be measures on $\mathbb{R}$. For a Borel map $s : \Omega \subseteq \mathbb{R} \to \mathbb{R}$, let $s_*\mu$ be the push-forward of μ through s defined by $s_*\mu(U) = \mu(s^{-1}(U))$ for Borel sets $U \subseteq \mathbb{R}$. The problem formulated by Monge was to calculate

$$\inf_{s: s_*\mu=\nu} \int_{\mathbb{R}} |x - s(x)| d\mu(x). \tag{5.1.1}$$

Here the expression under the integral is the cost of carriage of one unit of mass. In general we may be interested in a variety of cost functions $(x, y) \to c(x, y)$, but in this overview we will focus on Monge's original formulation, where $c(x, y) = |x - y|$. Due to the non-linearity and intractability of the space of push-forward maps, the problem in its original formulation has been solved only recently (under regularity assumptions on the two measures and for the general n-dimensional case) by Evans and Gangbo, see [32].

5.2 The Kantorovich relaxation and duality

Let us denote by $C(\mu,\nu)$ the set of all pairs of random variables (X,Y) with $X\sim\mu$ and $Y\sim\nu$, i.e.

$$C(\mu,\nu)=\{(X,Y):X\sim\mu,Y\sim\nu\}.$$

In a ground-breaking insight, Kantorovich [48] (who was not interested in 'earthwork' but rather in economic optimization problems arising in a planned economy[1]) reformulated Monge's problem in terms of joint measures rather than push-forward measures:

$$\inf_{(X,Y)\in C(\mu,\nu)} \mathbf{E}[c(X,Y)]. \tag{5.2.1}$$

Kantorovich also formulated the dual problem,

$$\sup_{\Psi,\Phi}\left\{\int_{\mathbf{R}}\Psi(x)\mu(dx)+\int_{\mathbf{R}}\Phi(y)\nu(dy)\ ;\ \Psi(x)+\Phi(y)\leq c(x,y)\right\}, \tag{5.2.2}$$

which he showed to be equivalent under regularity assumptions on μ, ν and c (see the references [69], [75], [32]). In fact (and most importantly in the context of this thesis), the optimal Ψ and Φ turn out to be c-convex functions.

For our purposes, the most important development based on the Kantorovich formulation of the optimal transportation problem is due to Rüschendorf [68, 69] and Gangbo and McCann [36], who showed that for $c(x,y)=f(|y-x|)$ (f strictly convex), the optimal coupling solving (5.2.1) is given by $(X,Y)=(X,s(X))$ where s is the generalized subdifferential of a c-convex function Ψ satisfying

$$\Psi'(x)=c_x(x,s(x)).$$

We recall Proposition 1.2.11 in Chapter 1 where we derived exactly this type of relationship between parameters and optimal stopping thresholds in perpetual horizon stopping problems. Similarly in Chapter 2, we showed how to interpret the Gittins index as a generalised subdifferential, compare (2.1.17). At least from a technical point of view, the approach to stopping problems based on generalised convex analysis is related to the classical transport problem.

[1]Kantorovich, L.V, Prize Lecture - Lecture to the memory of Alfred Nobel: Mathematics in Economics: Achievements, Difficulties, Perspectives., Dec 11, 1975. `http://www.nobelprize.org/nobel_prizes/economics/laureates/1975/kantorovich-lecture.html`

5.3 Skorokhod embedding as stochastic mass transport with a martingale constraint

Let us now denote by $\mathcal{M}(\mu,\nu)$ the space of all martingales $(M_t)_{0\leq t\leq T}$ with $M_0 \sim \mu$ and $M_T \sim \nu$. Then we may consider the following stochastic versions of Monge's original problem.

$$\sup_{X\in\mathcal{M}(\mu,\nu)} \mathbb{E}[|X_T - X_0|], \tag{5.3.1}$$

$$\inf_{X\in\mathcal{M}(\mu,\nu)} \mathbb{E}[|X_T - X_0|]. \tag{5.3.2}$$

These problems have the interpretation of finding the least upper and greatest lower bounds for a derivative called the forward starting straddle. Hobson and Neuberger [40] show how to solve the first of these two problems by re-casting it as the Skorokhod embedding problem,

$$\sup_{\tau\in S(\mu,\nu)} \mathbb{E}[|B_\tau - B_0|], \tag{5.3.3}$$

where $B=(B_t)_{t\geq 0}$ is Brownian motion started randomly with $B_0 \sim \mu$ and $S(\mu,\nu)$ is the set of stopping times such that $B_\tau \sim \nu$ and $B_{t\wedge\tau}$ is uniformly integrable. The idea in [40] is to use Lagrangian methods - to find α, β and γ such that $L(x,y)\leq 0$ for all x,y, where

$$L(x,y) = |y-x| - \alpha(x) - \beta(y) - \gamma(x)(x-y). \tag{5.3.4}$$

Then for any sample path,

$$|X_T(\omega) - X_0(\omega)| \leq \alpha(X_0(\omega)) + \beta(X_T(\omega)) + \gamma(X_0(\omega))(X_0(\omega) - X_T(\omega)).$$

Using similar methods, ongoing work not presented in this thesis shows how to solve the second of the two problems, i.e. find the greatest lower bound for the forward starting straddle.

Now we note that the problem of calculating the optimal super-hedge (i.e the functions α, β, γ) is the dual to the primal problem of calculating the least upper bound for the value of the forward starting straddle. If call prices with expiry times T_0 and T_1 are traded liquidly and imply that the marginal law of the forward price process is μ at T_0 and ν at T_1, then by the standard argument due to Breeden and Litzenberger [13], $\alpha(X_0)$ and $\beta(X_T)$ represent portfolios in call options. If we are able to sell the forward starting straddle while purchasing the two portfolios and going short $\gamma(X_0)$ forwards over the period $[T_0, T_1]$ then we will always profit. There is duality in the sense that the cost of the hedging strategy should be equivalent to the least upper price bound for the derivative. In spirit, the duality between pricing and hedging is analogous to the dual formulation of the deterministic transport problem due to Kantorovich (5.2.2), see also Section 5.5.2 below.

5.4 Path dependent costs in 'martingale transport'

While the preceding discussion focused on different formulations of the transport problem with Monge's 'original' cost function $c(x, y) = |y - x|$, we may now ask questions about optimal martingale transport with path-dependent cost functions. Based on classic questions about the law of the maximum of a martingale with a given target law, Hobson and Pedersen [43] show how to solve the problem,

$$\inf_{X \in \mathcal{M}(\mu,\nu)} \mathbf{E}[S_T^X],$$

where $S^X = (S_t^X)_{0 \leq t \leq T}$ denotes the running maximum of a martingale X.

Similarly, in this thesis we showed how to calculate least upper bounds and greatest lower bounds for variance swaps when the starting measure is atomic, $\mu = \delta_{X_0}$. In Chapter 3, we calculated bounds for an idealised variance swap $V_T = \int_0^T \frac{d[X,X]_t}{X_t^2}$. In Chapter 4 we used a dual Lagrangian formulation to calculate a super-hedging strategy $(\psi, -\psi')$ satisfying,

$$H(x, y) \leq \psi(y) - \psi(x) - \psi'(x)(y - x)$$

for a variance swap kernel H. We note the similarity to the Lagrangian formulation (5.3.4). We found that hedging strategies $(\psi, -\psi')$ can be parameterised by monotone functions, and that the optimal hedging strategy corresponds to the monotone function arising in the Perkins embedding of ν. Similarly in Hobson and Neuberger [40], The functions α, β and γ are constructed from increasing and decreasing functions which are hitting boundaries that define the solution to the corresponding Skorokhod embedding problem (5.3.3).

A natural extension to the work in this thesis is to construct model-independent hedging strategies and price bounds for forward starting variance swaps, i.e. replacing δ_{X_0} with a general starting measure μ. This is the subject of ongoing work.

5.5 Open problems and the Kantorovich duality in Finance

5.5.1 Continuous marginals

An interesting open problem is to ask about the situation when the marginals are given for all times $0 \leq t \leq T$. Then we can let $\mathcal{M}_{[0,T]}$ be the collection of all martingales $M = (M)_{0 \leq t \leq T}$

with the given marginals and attempt to calculate

$$\inf_{X \in \mathcal{M}_{[0,T]}} \mathbb{E}[P((X_t)_{0 \le t \le T})].$$

for some path-dependent payoff P. For example, what are the bounds for variance swaps in this case, and what are the corresponding extremal processes for which the bounds are tight?

5.5.2 Duality framework

Recent attention has been focused on the link between the classical Kantorovich duality and the duality between super-hedges and price bounds in financial mathematics. For instance, Galichon et. al. [35] re-consider the problem of bounding lookback options (first considered by Hobson and solved using Skorokhod embedding techniques [41]) by re-framing the question in terms of the problem of 'martingale transport along continuous martingales' and using a stochastic control approach. Related work by Beiglböck et. al. [8] extends the classical Kantorovich duality to the case of stochastic optimal transportation. This line of investigation promises a better understanding of the relationship between extremal models (optimal transportation plans), the corresponding Skorokhod embeddings and super-hedging strategies in terms of the classical framework.

Existing literature, the work presented in this thesis and further work in progress show that optimal price bounds and corresponding super-hedging strategies can be constructed for a variety of financial contracts including lookback options, straddles and variance swaps. The duality between hedges and price bounds can be understood case by case by relating the price bounds and the construction of super-hedging strategies to the monotone hitting boundaries of an underlying Skorokhod embedding. A general theory of duality between price bounds and hedging strategies and of the role played by monotone functions in their parameterisation would improve our understanding of the links between Skorokhod embeddings, super-hedging, model-independent bounds and duality principles in optimal transport.

Bibliography

[1] A. Alfonsi and B. Jourdain. A Call-Put Duality for Perpetual American options. *Preprint, ArXiv:0612648*, 2006.

[2] A. Alfonsi and B. Jourdain. General duality for perpetual American options. *International Journal of Theoretical and Applied Finance*, 11(6):545–566, 2008.

[3] A. Alfonsi and B. Jourdain. Exact volatility calibration based on a Dupire-type call-put duality for perpetual American options. *Nonlinear Differ. Equa. Appl*, 16(4):523–554, 2009.

[4] L.H.R Alvarez. Reward functionals, salvage values, and optimal stopping. *Mathematical Methods of Operations Research*, 54(2):315–337, 2001.

[5] S. Athey. Comparative statics under uncertainty: Single crossing properties and Log-Supermodularity. *Working paper, Department of Economics, MIT*, 1996.

[6] J. Azéma and M. Yor. Le problème de Skorokhod: compléments à "Une solution simple au problème de Skorokhod". In *Séminaire de Probabilités, XIII (Univ. Strasbourg, Strasbourg, 1977/78)*, volume 721 of *Lecture Notes in Math.*, pages 625–633. Springer, Berlin, 1979.

[7] J. Azéma and M. Yor. Une solution simple au problème de Skorokhod. In *Séminaire de Probabilités, XIII (Univ. Strasbourg, Strasbourg, 1977/78)*, volume 721 of *Lecture Notes in Math.*, pages 90–115. Springer, Berlin, 1979.

[8] M. Beiglböck, P. Henry-Labordère, and F. Penkner. Model-independent bounds for option prices: A mass transport approach. *Preprint, ArXiv:1106.5929*, 2011.

[9] A. Bensoussan and J.L. Lions. *Applications des Inéqualites Variationelles en Contrôle Stochastique.* Dunod, 1978.

[10] F. Black. How to use the holes in Black-Scholes. *Journal of Applied Corporate Finance*, 1(4):67–73, 1989.

[11] O. Bondarenko. Variance trading and market price of variance risk. Working paper, 2007.

[12] A.N. Borodin and P. Salminen. *Handbook of Brownian Motion - Facts and Formulae.* Birkhäuser, 2nd edition, 2002.

[13] D.T. Breeden and R.H. Litzenberger. Prices of state-contingent claims implicit in option prices. *J. Business*, 51:621–651, 1978.

[14] H. Brown, D.G. Hobson, and L.C.G Rogers. Robust hedging of barrier options. *Math. Finance*, 11(3):285–314, 2001.

[15] G. Carlier. Duality and existence for a class of mass transportation problems and economic applications. *Adv. in Mathematical economy*, 5:1–22, 2003.

[16] P. Carr and R. Lee. Hedging variance options on semi-martingales. *Preprint*, 2008.

[17] P. Carr and R. Lee. Variation and share-weighted variation swaps on time-changed Lévy processes. Preprint, 2010.

[18] R. V. Chacon. Potential processes. *Trans. Amer. Math. Soc.*, 226:39–58, 1977.

[19] R. V. Chacon and J. B. Walsh. One-dimensional potential embedding. In *Séminaire de Probabilités, X (Prèmiere partie, Univ. Strasbourg, Strasbourg, année universitaire 1974/1975)*, pages 19–23. Lecture Notes in Math., Vol. 511. Springer, Berlin, 1976.

[20] A. Cox and J. Wang. Root's barrier: construction, optimality and applications to variance options. Preprint, 2011.

[21] A.M.G Cox and D.G. Hobson. Skorokhod embeddings, minimality and non-centred target distributions. *Probability Theory and Related Fields, 2006*, 135(3):395–414, 2006.

[22] A.M.G Cox and J. Obloj. Robust hedging of double touch barrier options. *SIAM J. Finan. Math*, 2:141–182, 2011.

[23] M.H.A. Davis and D.G. Hobson. The range of traded option prices. *Mathematical Finance*, 17(1):1–14, 2007.

[24] K. Demeterfi, E. Derman, M. Kamal, and J. Zou. A guide to volatility and variance swaps. *The Journal of Derivatives*, 6(4):9–32, 1999.

[25] K. Demeterfi, E. Derman, M. Kamal, and J. Zou. More than you ever wanted to know about volatility swaps. *Goldman-Sachs Quantitive Strategies Research Notes*, 1999.

[26] L.E. Dubins and D. Gilat. On the distribution of maxima of martingales. *Proceedings of the American Mathematical Society*, 68:337–338, 1978.

[27] B. Dupire. Arbitrage pricing with stochastic volatility. *Société Générale, Options Division, Paris*, 1992.

[28] B. Dupire. Model-art. *Risk*, 6:118–124, 1993.

[29] H. Dym and H.P. McKean. *Gaussian Processes, Function Theory, and the Inverse Spectral Problem.* Academic Press, Inc., 1976.

[30] E. Ekström and D. Hobson. Recovering a time-homogeneous stock price process from perpetual option prices. *Ann. Appl. Probab.*, 21(3):1102–1135, 2011.

[31] E. Ekström, D. Hobson, and S. Janson. Can time-homogeneous diffusions produce any distribution? *ArXiv Preprint.*, 1103.4371, 2011.

[32] L.C. Evans and W. Gangbo. Differential equations methods for the Monge-Kantorovich Mass Transfer Problem. volume 137 of *Memoirs of the American Mathematical Society.* American Mathematical Society, 1999.

[33] W. Feller. The birth and death processes as diffusion processes. *J. Math. Pures Appl.*, 38:301–345, 1959.

[34] H. Föllmer. Calcul d'Itô sans probabilités. In *Séminaire de Probabilités, XV (Univ. Strasbourg, Strasbourg, 1981)*, volume 850 of *Lecture Notes in Math.*, pages 143–150. Springer, Berlin, 1981.

[35] A. Galichon, P. Henry-Labordère, and N. Touzi. A stochastic control approach to no-arbitrage bounds given marginals, with an application to Lookback options. *Preprint*, 2011.

[36] W. Gangbo and R.J McCann. The geometry of optimal transportation. *Acta Mathematica*, 177(2):113–161, 1996.

[37] J.C. Gittins and K.D. Glazebrook. On Bayesian models in stochastic scheduling. *Journal of Applied Probability*, 14:556–565, 1977.

[38] K. Hamza and F.C. Klebaner. A family of non-gaussian martingales with gaussian marginals. *J. Appl. Math. and Stochastic Analysis*, 1993. Article ID 9272.

[39] F. Hirsch, C. Profeta, B. Roynette, and M. Yor. Constructing self-similar martingales via two Skorokhod embeddings. In *Séminaire de Probabilités, XLIII*, pages 451–503. Lecture Notes in Math., Vol. 2006/2011. Springer, 2011.

[40] D. Hobson and A. Neuberger. Robust bounds for forward-start options. *Mathematical Finance*, 22(1):31–56, 2008.

[41] D.G Hobson. Robust hedging of the lookback option. *Finance and Stochastics*, 2:329–347, 1998.

[42] D.G. Hobson. The Skorokhod embedding problem and model independent bounds for option prices. In *Paris-Princeton Lecture Notes on Mathematical Finance*. Springer, 2010.

[43] D.G. Hobson and J. L. Pedersen. The minimum maximum of a continuous martingale with given initial and terminal laws. *Ann. Probab.*, 30(2):978–999, 2002.

[44] K. Itô and H.P McKean. *Diffusion Processes and their Sample Paths.* Springer-Verlag, 1974.

[45] I. Jewitt. Risk Aversion and the Choice Between Risky Prospects: The Preservation of Comparative Statics Results. *The Review of Economic Studies*, 54(1):73–85, 1987.

[46] I.S. Kac. Pathological Birth-and-Death Processes and the Spectral Theory of Strings. *Functional Analysis and Its Applications*, 39(2):144–147, 2005.

[47] N. Kahalé. Model-independent lower bound on variance swaps. Preprint, 2011.

[48] L. Kantorovich. On the translocation of masses. *Management Science*, 5(1), 1958.

[49] I. Karatzas. Gittins indices in the dynamic allocation problem for diffusion processes. *The Annals of Probability*, 12(1):173–192, 1984.

[50] R. Klimmek. *Risikoneigung und Besteuerung.* V. Florentz, 1986. Hochschulschriften zur Betriebswirtschafslehre, Bd. 37.

[51] S. Kotani and S. Watanabe. Krein's spectral theory of strings and generalized diffusion processes. *Functional Analysis in Markov Processes, (ed. M. Fukushima), Lecture Notes in Math.*, 923:235–259, 1982.

[52] B. Lu. Recovering a piecewise constant volatility from perpetual put option prices. *Journal of Applied Probability*, 47(3):680–692, 2010.

[53] P. Milgrom. *Putting Auction Theory to Work.* Cambridge University Press, 2003.

[54] P. Milgrom and I. Segal. Envelope theorems for arbitrary choice sets. *Econometrica*, 70(2):583–601, 2002.

[55] G. Monge. Mémoire sur la théorie des déblais et de remblais. *Histoire de l'Académie Royale des Sciences de Paris, avec les Mémoires de Mathématique et de Physique pour la même année*, pages 666–704, 1781.

[56] I. Monroe. On embedding right continuous martingales in Brownian motion. *Ann. Math. Statist.*, 43:1293–1311, 1972.

[57] I. Monroe. Processes that can be embedded in Brownian Motion. *The Annals of Probability*, 6(1):42–56, 1978.

[58] A. Neuberger. The Log Contract. *Journal of Portfolio Management*, 20(2):74–80, 1994.

[59] A. Neuberger. Realized skewness. Working paper, 2010.

[60] Jan Obłój. The Skorokhod embedding problem and its offspring. *Probab. Surv.*, 1:321–390 (electronic), 2004.

[61] E. Perkins. The Cereteli-Davis solution to the H^1-embedding problem and an optimal embedding in Brownian motion. In *Seminar on Stochastic Processes, 1985 (Gainesville, Fla., 1985)*, pages 172–223. Birkhäuser Boston, Boston, MA, 1986.

[62] S.T Rachev and L. Rüschendorf. *Mass Transportation Problems*, volume 1. Springer-Verlag, 1998.

[63] D. Revuz and M. Yor. *Continuous martingales and Brownian motion.* Springer-Verlag, Berlin, 1999.

[64] R. T. Rockafellar. *Convex Analysis.* Princeton University Press, 1970.

[65] L. C. G. Rogers. A guided tour through excursions. *Bull. London Math. Soc.*, 21(4):305–341, 1989.

[66] L.C.G Rogers. The joint law of the maximum and terminal value of a martingale. *Probability Theory and Related Fields*, 95:451–466, 1993.

[67] L.C.G. Rogers and D. Williams. *Diffusions, Markov Processes and Martingales*, volume 2. Cambridge University Press, 2000.

[68] L. Rüschendorf. Fréchet-bounds and their applications. In G. Dall'Aglio et al., editor, *Advances in Probability Distributions with Given Marginals : beyond the copulas*, pages 151–187. Kluwer Academic Publishers, 1991.

[69] L. Rüschendorf. Monge-Kantorovich transportation problem and optimal couplings. *Jahresbericht der DMV*, 109:113–137, 2007.

[70] L. Rüschendorf and L. Uckelmann. Numerical and analytical results for the transportation problem of Monge-Kantorovich. *Metrika*, 51(3):245–258, 2000.

[71] P. Samuelson. Tax deductibility of economic depreciation to insure invariant valuations. *Journal of Political Economy*, 72(6):604–606, 1964.

[72] A. V. Skorokhod. *Studies in the theory of random processes*. Translated from the Russian by Scripta Technica, Inc. Addison-Wesley Publishing Co., Inc., Reading, Mass., 1965.

[73] V. Strassen. Almost sure behavior of sums of independent random variables and martingales. In *Proc. Fifth Berkeley Sympos. Math. Statist. and Probability (Berkeley, Calif., 1965/66)*, pages Vol. II: Contributions to Probability Theory, Part 1, pp. 315–343. Univ. California Press, Berkeley, Calif., 1967.

[74] C. Villani. Current Trends in Optimal Transportation - A tribute to Ed Nelson. In F.G Faris, editor, *Diffusion, Quantum Theory, and Radically Elementary Mathematics*, volume 47 of *Mathematical Notes*, pages 141–156. Princeton University Press, 2006.

[75] C. Villani. *Optimal Transport: Old and New*. Springer-Verlag, 2009.

[76] P. Whittle. Multi-armed bandits and the Gittins index. *J. Roy Statist. Soc.*, Ser. B(42):143–149, 1980.

[77] S. Zhu and G-H. Lian. A closed-form exact solution for pricing variance swaps with stochastic volatility. *Math. Finance*, 11:233–256, 2011.