WARWICK
THE UNIVERSITY OF WARWICK

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

Permanent WRAP URL:

http://wrap.warwick.ac.uk/142722

Copyright and reuse:
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.

For more information, please contact the WRAP Team at: wrap@warwick.ac.uk

warwick.ac.uk/lib-publications

Techniques and approaches for pricing American options

by

Dominykas Norgilas

Thesis

Submitted to the University of Warwick

for the degree of

Doctor of Philosophy

Department of Statistics

July 2019

Contents

List of Figures

Acknowledgments

It is a great pleasure to thank David Hobson and Saul D. Jacka for their excellent mentorship during my PhD study.

I also wish to acknowledge the financial support from a EPSRC Doctoral Training Partnerships grant EP/M508184/1 which made my study at Warwick possible.

Special thanks go to my family.

Declarations

This thesis is submitted to the University of Warwick in support of my application for the degree of Doctor of Philosophy. It has been composed by myself and has not been submitted in any previous application for any degree. Work based on collaborative research is declared as follows:

Chapter 1 is based on a joint work with Saul Jacka, 'On the compensator in the Doob-Meyer decomposition of the Snell envelope', *SIAM Journal on Control and Optimization*, Volume 57(3), pp 1869-1889, 2019.

Chapter 2 is adapted from a joint work with Mathias Beiglböck and David Hobson, 'Geometry of the shadow', 2019 (working paper).

Chapter 3 is based on a joint work with David Hobson, 'Robust bounds for the American put', *Finance and Stochastics*, Volume 23(2), pp 359-395, 2019.

Chapter 4 is adapted from a joint work with David Hobson, 'The left-curtain martingale coupling in the presence of atoms', *Annals of Applied Probability*, Volume 29(3), pp 1904-1928, 2019.

0.1 Introduction

This thesis is concerned with the theory of optimal stopping and martingale optimal transport, and its applications to the pricing and hedging of American-type contingent claims.

In the first chapter we revisit the classical optimal stopping problem in continuous time and explore a delicate connection between semimartingale and Markovian formulations of the problem. More specifically, in the Markovian setting we are motivated by the question of whether the value function, corresponding to the optimal stopping problem, belongs to a certain class of functions (i.e. the domain of the extended or martingale generator) associated to the underlying Markov process. We show that the answer follows naturally from the fundamental property of the value process in a more general, semimartingale setting. We investigate applications of these results to the dual formulation of the optimal stopping problem and the classical smooth fit principle.

The goal of the second chapter is to study the problem in martingale optimal transport, which is to move mass from a starting law (on $\mathbb{R}$) to a terminal law (on $\mathbb{R}$) in a way which respects the martingale property. One method is the 'shadow embedding' of Beiglböck and Juillet [10]. Using the potential functions of the starting and terminal laws, we show how to explicitly construct the associated shadow measure. We also discuss the properties of the left-curtain martingale coupling, which is a coupling that arises (via shadow measure) from a certain parametrisation of the marginals. This coupling turns out to be optimal for the novel *optimal martingale transport with stopping* problem studied in the third chapter.

The third chapter studies the problem of finding the highest robust or model-independent price of the American put option given the prices of liquid European options, in a simple (but non-trivial) two time period setting. Combining ideas from the theory of optimal stopping and martingale optimal transport, we find, under some simplifying but still general conditions on the given data, the optimal model and the optimal stopping time. We also explicitly calculate the cheapest superhedging trading strategy.

In the fourth chapter our goal is to find a specific geometric description of the left-curtain martingale coupling, which can be viewed as a martingale counterpart of the monotone Hoeffding-Fréchet coupling in the classical optimal transport. While this is of independent interest, we also show that this generalised martingale coupling maximises the price of the American put option (studied in the third chapter under some simplifying assumptions).

0.1.1 Classical optimal stopping

Consider the optimal stopping problem where we observe a stochastic (payoff or gains) process and seek for the time instance at which the expected value of this process is maximised. Two natural questions arise: can we characterise an optimal stopping rule and what is the maximal value?

In the formulation of optimal stopping problems where the underlying payoff process is given by a sequence of random variables, Snell [103] discovered a crucial supermartingale characterisation of the associated value process and showed that stopping at the first instance when the value process is equal to the gains process is optimal. In continuous time, El Karoui [38] studied the problem when the payoff is given by a progressive process. In this case it is not clear *a priori* that the value of the game at every time instance can be identified with a progressive value process, and, therefore, an important step is to find such a characteristic process. This problem is solved by employing some delicate results from the General Theory of Processes (the corresponding supermartingale process is historically called the Snell envelope of the reward process). In the same setting, a penalisation method introduced by Maingueneau [79] is used to show, under some right regularity conditions on the reward process, the existence of ϵ-optimal stopping times. If, in addition, one imposes left regularity conditions on the gains process, the minimal optimal stopping time is indeed the first time the gains and value processes coincide. Moreover, if the gains process is sufficiently integrable, the value process admits a Doob-Meyer decomposition (as a difference of a uniformly integrable martingale and a non-decreasing process of integrable variation). This decomposition is further used to characterise the maximal optimal stopping time as the first time the increasing component of the Doob-Meyer decomposition of the value process is positive. For the theory of the optimal stopping problems in a more general framework, where the gains are given by an admissible family of random variables, we refer to Kobylanksi and Quenez [72] and references therein.

The relationship between the supermartingale characterisation of the value process, the maximal stopping time and the reward process is further investigated in the first chapter. In particular, from the general results on optimal stopping mentioned above (see El Karoui [38]), the value process is a martingale (and thus the non-decreasing finite variation part in its Doob-Meyer decomposition must be zero) up until the maximal optimal stopping time. Now, in addition, suppose the gains process is also a semimartingale (a sum of a local martingale and a process of finite variation). Then (just as in the Hahn-Jordan decomposition for measures) we can further decompose the finite variation part into (mutually singular) decreasing and increasing components. The intuitive but crucial observation is that off the support of the decreasing process, the reward process is (locally) a submartingale, so that it is non-decreasing in (conditional) expectation and thus it is suboptimal to stop. In this case we, therefore, again expect the value process to be (locally) a martingale. This suggests that the finite variation part in the Doob-Meyer decomposition of the value process increases only if the decreasing component of the finite variation part of the reward process decreases. We prove the following fundamental result:

> *The finite-variation process in the Doob-Meyer decomposition of the value process is absolutely continuous with respect to the decreasing part of the corresponding finite-variation process in the semimartingale decomposition of the gains process.*

This being a very natural conjecture, it is not surprising that some variants of it have already been considered. More specifically, several versions of this result were established in the literature on reflected BSDEs under various assumptions on the gains process, see El Karoui et. al. [39] (gains process is a continuous semimartingale), Crepéy and Matoussi [28] (gains process is a càdlàg quasi-martingale), Hamadéne and Ouknine [48] (gains process is a limiting process of a sequence of sufficiently regular semimartingales). These results (except Hamadéne and Ouknine [48], where the assumed regularity of the reward process is exploited) are proved essentially by using (or appropriately extending) the related (but different) result established in Jacka [64]. There, under the assumption that the reward and corresponding value processes are both continuous and sufficiently integrable semimartingales, the author shows that a local time of the difference of these two processes at zero is absolutely continuous with respect to the decreasing part of the finite-variation process in the semimartingale decomposition of the reward process. Our approach in proving the absolute continuity result is different. We use the semimartingale decomposition of relevant processes and exploit the classical methods establishing the Doob-Meyer decomposition of a supermartingale.

In practice we often do not model the gains process directly. In particular, in the (slightly less general) Markovian formulation of the problem, one starts with an underlying Markov process and then applies a given payoff function to it, to produce the corresponding gains process. One advantage of such a formulation is that the value process is then also given in a functional form and the goal is to construct a value function (or at least to deduce its properties) on the state space of the underlying Markov process (see El Karoui et al. [40]). It was established by Dynkin [36] that the proposed (by Snell [103]) supermartingale characterisation of the value process of an optimal stopping problem is equivalent to the superharmonic characterisation of the value function in a Markovian setting. This resulted in further development of the field producing more concrete results. We refer to the monograph of Shiryaev [100] for an overview of optimal stopping theory for both discrete and continuous-time Markov processes.

From the results of Snell [103] and Dynkin [36] it is clear that there is a deep connection between the two formulations of the optimal stopping problem. In particular, the properties of the value process (proved in a more general setting of processes) carries over, in some sense, into the Markovian formulation of the problem, and allows to deduce particular characteristics of the corresponding value function. With this in mind, in chapter one we try to answer the following canonical question of interest:

> *When does the value function of the optimal stopping problem belong to the domain of the extended (martingale) generator of the underlying Markov process?*

Trying to deduce such a property for the value function, it is reasonable to assume that the payoff function also belongs to the domain of the martingale generator of the underlying Markov process. Then, given the absolute continuity result in a semimartingale setting, the answer to the motivating question follows naturally. We show that (see Theorem 1.2.13) under very general assumptions on the underlying Markov process, if the payoff function belongs to the domain of the martingale generator, so does the value function of the optimal stopping problem.

Another advantage of the Markovian formulation of the optimal stopping problem is that the decision whether to stop, at any given time instant, can be based only on the location (or the value at that particular moment) of the underlying Markov process, and not the whole history of the process. In particular, the state space of the observed Markov process can be split into continuation and stopping regions. The optimal stopping time is proved to be the first exit (entrance) time from (into) the continuation (stopping) region, and the crux of the problem is then to simultaneously find the value function and to characterise the unknown boundary separating continuation and stopping regions. Consequently, a solution of the free-boundary (or obstacle) problem for differential operators (for example, Stefan's ice-melting problem in mathematical physics) directly relates to the original optimal stopping problem. Imagine a slab of ice (at temperature G, which corresponds to the gain function) immersed in water (at temperature V, which corresponds to the value function). Then the ice-water interface (as a function of time and space) will coincide with the optimal boundary (surface). This illustrates a basic link between optimal stopping and the Stefan's problem.

In order to select the unique solution of the free-boundary problem, which will eventually (by standard verification arguments such as extended Itô's formula and/or Doob's optional sampling theorem) turn out to be the solution of the initial optimal stopping problem, one usually imposes non-trivial boundary conditions. In particular, one expects that the value and payoff functions (the so-called continuous fit condition) and their first derivatives (the so-called smooth pasting condition) coincide at the (unknown and therefore free) optimal boundary points. The smooth fit principle was first applied by Mikhalevich [80] for concrete problems in sequential analysis and later by Chernoff [22] and Lindley [77]. McKean [78] applied the principle to the American option problem. See also Grigelionis and Shiryaev [45] and van Moerbeke [107]. On the other hand, the continuous fit, as a key ingredient of the solution, was recognised by Peskir and Shiryaev [88, 89]. The book by Peskir and Shiryaev [90] provides many explicitly solved problems (via a free-boundary approach) for various gains functionals. An extensive bibliography on the subject can also be found in Shiryaev [100].

While one expects the smooth fit principle to hold at the boundary of the continuation region, a priori it is not clear whether the value function associated to the optimal stopping problem is even differentiable. The work by Peskir [86] in the diffusion setting shows that, for the above to hold, in general, the differentiability of the payoff function is not enough, and some sort of 'smoothness' of the underlying Markov process is required. In the first chapter we investigate this further. On one hand, the minimal concave characterisation of the value function by Dynkin [36] (see also Dayanik and Karatzas [30]) and the extended Itô's formula for concave (or convex functions) provides a particular decomposition of the associated value process. On the other hand, if the value function belongs to the domain of an extended generator of the underlying Markov process, the value process also admits the Doob-Meyer decomposition with absolutely continuous (with respect to Lebesgue measure on the time axis) finite variation part. Comparing these two, in a sense, canonical decompositions, under some additional but general assumptions on the Markov process, we show that the value function is nearly **twice** continuously differentiable (i.e. it has absolutely continuous first derivative).

Solving optimal stopping problems can be quite straightforward in low dimensions. However, many problems arising in practice are high-dimensional, and one has to resort to numerical approximations of the solutions. In particular, the rate of convergence of these approximations is of great importance. On the other hand, embedding optimal stopping problems into stochastic control problems allows one to use all the available machinery for stochastic control. Therefore a first step in this direction is to identify a suitable stochastic control problem corresponding to the initial optimal stopping problem.

It is known (see Krylov [75]) that optimal stopping problems for controlled diffusion processes can be transformed into optimal control problems by means of randomised stopping. More recently, Gyöngy et al. [47] showed that this transformation is possible even in the case when the coefficients of the diffusions and the functions defining the payoff are unbounded functions of the control parameter. See also Peskir [87] for the duality principle (in terms of Legendre transform) of the optimal stopping games. An alternative approach (which we adopt in the first chapter) is due to Davis and Karatzas [29], Rogers [96], and Haugh and Kogan [49]. In particular, given any martingale, the Snell envelope process (the value process associated to the optimal stopping problem) is dominated by the expected value of the pathwise supremum of the difference of the reward process and a chosen martingale. Then the *dual* problem is to find a martingale that minimises this quantity. Given that the value process admits a Doob-Meyer decomposition, Rogers [96] shows that the martingale part of this decomposition is optimal for the dual problem, and, in particular, *strong duality* holds. The restatement of this result in a Markovian setting is provided in chapter one.

Unfortunately, even though the characterisation of the optimal martingale is clear, since the value process (or the corresponding value function in the Markovian setting) is unknown, finding an optimal martingale is, in principle, no easier than exhibiting an optimal stopping time in the primal problem. On the other hand, choosing an arbitrary martingale in the dual problem, produces an upper bound for the value of the original optimal stopping problem. Therefore an important challenge is to find or construct martingales with good approximating properties (see e.g. Andersen and Broadie [2], Kolodko and Schoenmakers [73], Glasserman and Yu [44], Belomestny et al. [12], Belomestny [13] Desai et al. [32]). Our contribution to this subject is the following. If the value function lies in the domain of a martingale generator of the underlying Markov process, the search for the optimal martingale in the dual problem can be restricted to this particular set of functions. As a consequence, the dual can be viewed as a stochastic control problem for a controlled Markov process where the set of admissible controls is the domain of a martingale generator, and thus is amenable to standard theory (see Fleming and Soner [41]) related to such problems.

0.1.2 Martingale optimal transport (MOT)

Suppose we are given a pile of sand and a ditch (of the same volume as the pile) that we have to completely fill up with the sand. As moving the sand needs some effort, we suppose we are also given a (measurable) cost function c, such that $c(x, y)$ tells how much it costs to transport one unit of mass from location x to location y. Then basic question in the theory of classical optimal transport (OT) is *how to realise the transportation at minimal cost?*

Mathematically, the problem can be formulated (following Kantorovich [69]) as follows. Given two probability measures μ (corresponding to the pile of sand) and ν (corresponding to the ditch) on some measurable spaces $\mathcal{X}$ and $\mathcal{Y}$ (for our applications in Chapters 2 to 4 we take $\mathcal{X} = \mathbb{R} = \mathcal{Y}$), respectively, consider a probability distribution π (also called transport plan or coupling) on the product space $\mathcal{X} \times \mathcal{Y}$ with given first and second marginals μ and ν, respectively. Probabilistically, π represents a joint distribution of random variables $X \sim \mu$ and $Y \sim \nu$. Informally, $\pi(dx, dy)$ measures the amount of mass transferred from location x to location y. Then given a cost/payoff function $c : \mathcal{X} \times \mathcal{Y} \mapsto \mathbb{R}$, the goal is to construct (or at least to characterise) a joint distribution π which minimises the total expected cost $\mathbb{E}^\pi[c(X, Y)]$. We refer to Villani's [108] excellent exposition of the theory of the classical optimal transportation.

A similar (and actually harder) formulation of the OT problem was already considered in 1781 (a way before Kantorovich's one) . Monge [82] studied the same problem (of finding a minimal transport cost) without allowing the mass to be split, so that all the mass from location x has to end up in a unique location y. In other words, this restricts to the case when the random variable $Y \sim \nu$ is a function of a random variable $X \sim \mu$ (i.e. $Y = T(X)$, for some $T : \mathcal{X} \mapsto \mathcal{Y}$). Mathematically, Monge's goal was to find a measurable function T (called a transport map) such that the initial measure μ is *pushed forward* to the target measure ν through T (i.e. $\mu(\{x \in \mathcal{X} : T(x) \in A\}) = \nu(A)$, for any Borel set A) and the total expected cost $\mathbb{E}^\mu[c(X, T(X))]$ is minimal. In terms of transport plans, this means that we restrict to joint distributions π (of $X \sim \mu$ and $Y \sim \nu$) which are of the special form $\pi(dx, dy) = \mu(dx)\delta_{T(x)}(dy)$.

The fundamental question is under what conditions the Kantorovich's and Monge's problems coincide. Put differently, when can the optimal transport plan π be represented by a transport (push-forward) map T. A cornerstone result in this direction is the so-called Brenier's [17] theorem (see also Rüschendorf and Rachev [93]) which treats a particular case $c(x, y) = |y - x|^2$: for sufficiently regular initial distribution μ, the optimal transport plan is unique and supported by the graph of the gradient of some convex function (a monotonically increasing function in one dimension, so that in this case the optimal transport plan is realised by a mapping). More specifically, when $c(x,y) = h(y-x)$, where $h : \mathbb{R} \mapsto \mathbb{R}$ is a strictly convex function, the monotone Hoeffding-Fréchet coupling $\pi_{HF} = (G_\mu \otimes G_\nu)_\# \mathrm{Leb}_{[0,1]}$ is optimal and, when the initial measure μ is atomless, the optimal transport map is given by $T(x) = (G_\nu \circ F_\mu)(x)$ (here, for a measure χ, F_χ denotes the corresponding cumulative distribution function, while G_χ is a generalised inverse of F_χ).

In this thesis we will study a particular variant of the classical OT problem. Beiglböck et al. [8] and Galichon et al. [42] (see also Dolinksy and Soner [34]) introduced a martingale version of the transportation problem (and related it to the problem of finding model-independent bounds of exotic derivatives in mathematical finance, see Section 0.1.3). Given μ and ν in convex order, the basic problem of martingale optimal transport (MOT) is to construct a martingale M, with $M_1 \sim \mu, M_2 \sim \nu$, which minimises $\mathbb{E}[c(M_1, M_2)]$. In this setting a martingale transport plan or coupling can be identified with a measure π on $\mathbb{R}^2$ with univariate marginals μ and ν, and such that additional martingale constraint $(\int_{y\in\mathbb{R}} \int_{x\in A}(y-x)\pi(dx,dy) = 0$ for all Borel sets A) holds. In the context of mathematical finance this problem was first studied in Hobson and Neuberger [58] for the payoff $c(x,y) = -|y-x|$.

The MOT theory was developed in parallel to the classical one, and thus it is not surprising that it shares some similar properties. It is worth noting that both, the set of joint distributions (with given marginals) and its subset enjoying additional martingale requirement, are both compact with respect to weak convergence of measures (a consequence of Prokhorov's [91] theorem). Moreover, if the cost function is lower semi-continuous, the same regularity property holds for $\pi \mapsto \mathbb{E}^\pi[c(X,Y)]$, for both sets of joints distributions. From these observations it follows that the optimal couplings exists in both OT and MOT, formulations of the problem.

On the other hand there are several important differences. While the set of classical transport plans is always non-empty (the product measure belongs to this set), the same conclusion does not hold for the set of martingale couplings. In particular, the additional requirement that μ and ν are in convex order (see Section 2.1) is necessary. Moreover, in OT problems the important tool in establishing the optimality of a candidate coupling is the corresponding dual problem. Suppose $c(x,y) \geq a(x) + b(y)$ for all x and y, for some functions $a(\cdot)$ and $b(\cdot)$. Then, $\mathbb{E}^\pi[c(X,Y)] \geq \mathbb{E}^\mu[a(X)] + \mathbb{E}^\nu[b(Y)]$. Note that the left hand side of inequality is independent of functions a and b, while the right hand side is independent of the coupling π. Therefore, taking the supremum of the right hand side over the functions $a(\cdot)$ and $b(\cdot)$ satisfying path-wise inequality (this maximisation problem is the corresponding *dual* problem) and the infimum of the left hand side over joint distributions π (this corresponds to the *primal* problem), preserves the inequality between expectations. Then, if we can find $\pi, a(\cdot)$ and $b(\cdot)$ such that $\mathbb{E}^{\pi}[c(X,Y)] = \mathbb{E}^{\mu}[a(X)] + \mathbb{E}^{\nu}[b(Y)]$, we must have that π is optimal (i.e. minimises the total expected cost). On the other hand, if π is a martingale coupling and $f(\cdot)$ a given function, using a tower-property for expectations and a martingale condition we have that $\mathbb{E}^{\pi}[f(X)(Y-X)] = 0$. Therefore, in MOT problems, if we can find a triple $a(\cdot), b(\cdot), f(\cdot)$ satisfying $c(x,y) \geq a(x) + b(y) + f(x)(y-x)$ for all x and y, and such that $\mathbb{E}^{\pi}[c(X,Y)] = \mathbb{E}^{\mu}[a(X)] + \mathbb{E}^{\nu}[b(Y)]$, then a martingale coupling π must be optimal. Finally, for the quadratic cost functions in the classical OT theory, MOT problems turn out to be trivial (see Remark 2.1.5).

Beiglböck and Juillet [10] introduced the notion of the shadow embedding, which gives rise to a family of martingale couplings. See Section 2.2 for the definition and the properties of the corresponding shadow measure. In particular, a specific parametrisation of the initial law μ produces, via shadow measure, the left-monotone martingale coupling. Beiglböck and Juillet [10] established that for (arbitrary) fixed marginals μ and ν in convex order there exists a unique such coupling (called the left-curtain martingale coupling and denoted by π_{lc}). The left-curtain martingale coupling may be viewed as a martingale analogue to the monotone Hoeffding-Fréchet coupling in classical optimal transport. The authors also proved the optimality of π_{lc} for a specific class of payoff functions. Henry-Labordère and Touzi [50] extended the results of Beiglböck and Juillet [10] and showed optimality for a wider class of payoff functions. Beiglböck et al. [9] analysed the left-curtain coupling further and gave a simplified proof (using a Skorokhod embedding argument) of uniqueness under the additional assumption that μ is continuous. Juillet [67] proved that $(\mu, \nu) \mapsto \pi_{lc}$ is continuous (with respect to its marginals), and thus, for general distributions, it can be approximated by the left-curtain couplings corresponding to 'nice' (e.g. finitely supported or continuous) initial and/or target laws. A number of further articles investigate the properties and extensions of π_{lc} (e.g. multi-marginal case and connections with Skorokhod embedding problem), see Beiglböck et al. [7, 9], Nutz et al. [84, 85].

Beiglböck and Juillet [10] also established a martingale version of the fundamental Brenier's theorem. In particular, the authors showed that under the assumption that the initial law μ is continuous, the left-curtain martingale coupling is supported by the graphs of lower and upper functions T_d and T_u, respectively, so that $M_2 \in \{T_d(M_1), T_u(M_1)\}$. Henry-Labordère and Touzi [50] gave an explicit construction of T_d and T_u using differential equations. However, when μ has an atom at x the element $\pi^x_{lc}(\cdot)$ in the disintegration $\pi_{lc}(dx, dy) = \mu(dx)\pi^x_{lc}(dy)$ becomes a measure with support on non-trivial subsets of $\mathbb{R}$ and not just on a two point set. Then we cannot construct functions (T_d, T_u), unless we allow them to be multi-valued.

In Chapters 2 and 4 we study the shadow embedding and the left-curtain martingale coupling, respectively. While Beiglböck and Juillet [10] proved the existence and uniqueness, Chapter 2 provides an explicit construction (via associated potential functions) of the shadow measure. In Chapter 4 we concentrate on the left-curtain martingale coupling. There the goal is to show how (for general initial law μ, with or without atoms) by changing our viewpoint we can again recover the property that M_2 takes values in a two-point set. The idea is to write $M_1 = h(Z)$ for a continuous random variable Z (in fact we take $Z \equiv U \sim U(0,1)$) and then to find $f_{Z,h}$ and $g_{Z,h}$ such that $M_2 \in \{f_{Z,h}(Z), g_{Z,h}(Z)\}$. Then, although there is uniqueness at the level of martingale couplings π, when μ contains atoms there are many possible choices of $(f_{Z,h}, g_{Z,h})$, even for fixed Z and monotonic increasing h. Nonetheless, we show that amongst this set there is an essentially unique choice $(f_{Z,h}, g_{Z,h})$ with a special monotonicity property.

The motivation for the extension of the left-curtain martingale coupling in Chapter 4 comes from mathematical finance. The study of American put options in Chapter 3 highlights the role of the left-curtain martingale coupling in finding the model-independent upper bound on the price of the American put. When μ is continuous we show how the optimal martingale coupling and the optimal stopping time can be obtained from the functions $f = T_d$ and $g = T_u$ which arise in the construction of the left-curtain coupling. In particular, for the optimal model there is a Borel subset of $\mathbb{R}$, say B, such that it is optimal to stop at time-1 if $M_1 \in B$, and at time-2 otherwise. Moreover, by considering the corresponding dual problem, the structure of f and g allows us to identify the cheapest superhedging strategy that supports the price of the American put.

If μ has atoms then the situation becomes more delicate, essentially because we must allow for a wider range of possible candidates for exercise determining sets B. On atoms of μ we may want to sometimes stop and sometimes continue, although we must still take stopping decisions which do not violate the martingale property. As the stopping decision in the continuous case is based on the natural filtration of the martingale M, if M_1 ends up at the atom of μ, then it is not clear, using only the structure of f and g, what part of mass at time-1 should be stopped and what part should be allowed to continue. This is the reason why we must extend the notion of the left-curtain martingale coupling.

We show that the extended left-curtain coupling constructed in Chapter 4 is again characterised by lower and upper functions, R and S, respectively. However, while f and g are multi-valued on the atoms of μ, R and S remain well-defined. Then our second achievement is to show how the structure of R and S can be used to characterise the model and stopping rule which achieves the highest possible price for the American put, and the cheapest superhedge. This generalises results of Chapter 3: for arbitrary μ and ν, the highest model based price of the American put is equal to the cost of the cheapest superhedge.

0.1.3 Model-independent approach to option pricing

The standard approach in pricing of financial contracts (derivatives or contingent claims) is to start by postulating a model for the price process of the underlying risky asset (i.e. a stochastic process living on some fixed filtered probability space). Then the price of the contingent claim is calculated as a discounted expected value (with respect to equivalent martingale measure, under which a discounted price process is a martingale) of the payoff at the maturity. When the market is complete (so that there exists only one equivalent martingale measure) this pricing rationale is supported by the hedging or replication strategy. In particular, in this situation we can construct a hedge that perfectly replicates the payoff of the derivative, given that the model provides an exact description of reality. For example, the setting of Chapter 1 belongs to this framework, and the value of the optimal stopping problem considered in there, in financial context, is equivalent to the price of the American type derivate contract, under the fixed probabilistic model.

However, if an agent confines herself in the situation described above, this leaves her facing the Knightian uncertainty (i.e. the model risk). For this reason, in Chapters 3 and 4 we employ an alternative approach, where our starting point is not the model itself, but rather a financial data available in the market. The idea is to take option prices (of liquidly traded securities) as exogenously given by the market and to use those to extract the stochastic properties of the underlying price process of the risky asset that we are trying to model. Then the next step is to consider only those models that perfectly calibrate to the market data, so that the prices of vanilla options, calculated as expectations under the chosen model, match the prices observed in the market. (For example, Krylov [76], Dupire [35] and Gyöngy [46] showed that if we know the prices of European puts (or calls) for all strikes and maturities (which is a rather ideal situation), then there exists a unique diffusion process under which prices of these vanilla options match the prices quoted by the market.) Potentially, this approach still gives a large set of models to choose from, and while, under any of these models, the prices of vanilla options are the same, the prices of exotic derivatives are different. The goal is then to look for the 'extremal' models, i.e. the models that produce the highest/lowest price of a given exotic derivative contract.

This notion of model-independent, or robust, bounds on the prices of exotic options was introduced in Hobson [54] in the context of lookback options, and has been applied several times since, see Brown et al. [18] (barrier options), Cox and Obłój [26] (no-touch options), Hobson and Neuberger [58] and Hobson and Klimmek [57] (forward-start straddles), Carr and Lee [19] and Cox and Wang [27] (variance options), Stebegg [104] (Asian options) and the survey article Hobson [56]. The principal idea is that the prices of the vanilla European puts or call, by arguments of Breeden and Litzenberger [16], determine the marginal distributions of the price process at the traded maturities (but not the joint distributions) and that these distributional requirements, coupled with the martingale property, place meaningful and useful restrictions on the class of consistent models. These restrictions lead to bounds on the expected payoffs of path-dependent functionals, or equivalently, to bounds on the prices of exotic options.

In addition to the pricing problem there is a related dual or hedging problem. In the dual problem the aim is to construct a static portfolio of European put options and a dynamic discrete (or continuous) time hedge in the underlying which combine to form a superhedge (pathwise over a suitable class of candidate price paths) for the exotic option. The value of the dual problem is the cost of the cheapest superhedge. There is a growing literature, beginning with Beiglböck et al. [8] for discrete-time problems, and Galichon et al. [42] in continuous time, which aims to explain how to formulate the problem in such a way that there is no duality gap, i.e. the highest model-based price is equal to the cheapest superhedge, either for specific derivatives, or in general.

Many of the early papers on robust hedging exploited a link with the Skorokhod embedding problem (Skorokhod [102]). For example, in the study of the lookback option in Hobson [54] the consistent model which achieves the highest lookback price is constructed from the Azéma-Yor [5] solution of the Skorokhod embedding problem. More recently, Beiglböck et al. [8] (see also Dolinsky and Soner [34] and Touzi [106]) have championed the connection between robust hedging problems and martingale optimal transport. In Chapter 3 we will make use of the left-curtain martingale coupling introduced by Beiglböck and Juillet [10], and developed by Henry-Labordère and Touzi [50] and Beiglböck et al. [9]. The generalised construction of the left-curtain martingale coupling is provided in Chapter 4.

In Chapters 3 and 4 we are motivated by an attempt to understand the range of possible prices of an American put in a robust, or model-independent, framework. In our interpretation this means that we assume we are given today's prices of a family of European-style vanilla puts (for a continuum of strikes and for a discrete set of maturities). The goal is to find the consistent model for the underlying, for which the American put has the highest price, where by definition a model is consistent if the discounted price process is a martingale and if the model-based discounted expected values of European-put payoffs match the given prices of European puts.

The study of American style claims in the robust framework was initiated by Neuberger [83], see also Hobson and Neuberger [60], Bayraktar and Zhou [6] and Aksamit et al. [1]. (There is also a paper by Cox and Hoeggerl [25] which asks about the possible shapes of the price of an American put, considered as a function of strike, given the prices of co-maturing European puts.) The main innovation of Chapter 3 is that rather than focussing on general American payoffs and proving that the pricing (primal) problem and the dual (hedging) problem have the same value, we focus explicitly on American puts and try to say as much as possible about the structure of the consistent price process for which the model-based American put price is maximised, and the structure of the cheapest superhedge.

Our problem can be cast as follows. Let $M = (M_0 = \bar{\mu}, M_1 = X, M_2 = Y)$ represent the discounted price of an underlying asset, where $\bar{\mu}$ is a known constant. The laws of X and Y are presumed to be given and $\mathcal{L}(X) = \mu$ and $\mathcal{L}(Y) = \nu$, where μ and ν are (integrable) probability measures on $\mathbb{R}$ with mean $\bar{\mu}$. Given a martingale model (a filtered probability space, supporting a stochastic process M which is a martingale) we consider an American put on M with strike K. The option may only be exercised at time 1 or time 2: if the put is exercised at time 1 the payoff is $(K_1 - X)^+$; if the put is exercised at time 2 the payoff is $(K_2 - Y)^+$. Here K_1 and K_2 represent the discounted strikes of the put. For any martingale model, the model-based price of the American put is then given by the expected value of the payoff calculated under the best available stopping time (defined with respect to the filtration associated to the given model, and taking values in time 1 or time 2). Our primal problem is to find the highest possible model-based price of the American put, i.e. the highest expected payoff, where expectations are calculated under the probability measure of a consistent model (a model under which M is a martingale, and has the given laws at times 1 and 2).

There is a corresponding dual or hedging problem of finding the cheapest superhedge based on static portfolios of European puts and a piecewise constant holding of the underlying asset, see Section 3.1.2.

Our main achievement is as follows:

> *Suppose μ is continuous. The highest model-based expected payoff of the American put is equal to the cheapest superhedging price. Moreover, the highest model-based expected payoff is attained by the model associated with the left-curtain martingale coupling of Beiglböck and Juillet [10] (and a judiciously chosen stopping rule). Further, we can characterise the cheapest super-hedging strategy and it is one of four possible types.*

For fixed μ, ν and $K_1 > K_2$ there is typically a family of optimal models. Fixing μ and ν but varying K_1 and K_2 it turns out that there is a model which is optimal for all K_1 and K_2 simultaneously. This model is related to the left-curtain coupling of Beiglböck and Juillet [10] (see also Chapters 2 and 4).

Chapter 1

Properties of the Doob-Meyer decomposition of the Snell envelope and its applications in Markovian setting

The main result of this chapter (in a Markovian formulation of the optimal stopping problem) is Theorem 1.2.13: if the payoff function belongs to the domain of a martingale generator of the underlying Markov process, so does the value function. This is a consequence of Theorem 1.1.15, a fundamental result relating semimartingale decomposition of the gains process and the Doob-Meyer decomposition of the value (Snell envelope) processes. Several applications to the dual problem (Theorem 1.2.16) and to the smoothness of the value function (Theorems 1.2.18 and 1.2.23) are investigated.

1.1 General framework

1.1.1 Preliminaries

Fix a time horizon $T \in (0, \infty]$. Let G be an adapted, càdlàg gains process on $(\Omega, \mathcal{F}, \mathbb{F} = (\mathcal{F}_t)_{0 \le t \le T}, \mathbb{P})$, where $\mathbb{F}$ is a right-continuous and complete filtration. We suppose that $\mathcal{F}_0$ is trivial. In the case $T = \infty$, we interpret $\mathcal{F}_\infty = \sigma\Big(\cup_{0 \le t < \infty} \mathcal{F}_t \Big)$ and $G_\infty = \liminf_{t \to \infty} G_t$. For two $\mathbb{F}$-stopping times σ_1, σ_1 with $\sigma_1 \le \sigma_2$ $\mathbb{P}$-a.s., by $\mathcal{T}_{\sigma_1, \sigma_2}$ we denote the set of all $\mathbb{F}$-stopping times τ such that $\mathbb{P}(\sigma_1 \le \tau \le \sigma_2) = 1$. We will assume that the following condition is satisfied:

$$\mathbb{E}\Big[\sup_{0\le t\le T}|G_t|\Big] < \infty, \tag{1.1}$$

and let

$\bar{\mathbb{G}}$ be the space of all adapted, càdlàg processes such that (1.1) holds.

The *optimal stopping problem* is to compute the maximal expected reward

$$v_0 := \sup_{\tau\in\mathcal{T}_{0,T}} \mathbb{E}[G_\tau].$$

Remark 1.1.1. *First note that by* (1.1)*, $\mathbb{E}[G_\tau] < \infty$ for all $\tau \in \mathcal{T}_{0,T}$, and thus v_0 is finite. Moreover, most of the general results regarding optimal stopping problems are proved under the assumption that G is a non-negative (hence the gains) process. However, under* (1.1)*, $N = (N_t)_{0\le t\le T}$ given by $N_t = \mathbb{E}[\sup_{0\le s\le T}|G_s||\mathcal{F}_t]$ is a uniformly integrable martingale, while $\hat{G} := N + G$ defines a non-negative process (even if G is allowed to take negative values). Then*

$$\hat{v}_0 := \sup_{\tau\in\mathcal{T}_{0,T}} \mathbb{E}[N_\tau + G_\tau] = \mathbb{E}\Big[\sup_{0\le t\le T}|G_t|\Big] + \sup_{\tau\in\mathcal{T}_{0,T}} \mathbb{E}[G_\tau],$$

and finding $\hat{v}_0$ is the same as finding v_0. Hence we may, and shall, assume without loss of generality that $G \ge 0$.

The key to our study is provided by the family $\{v_\sigma\}_{\sigma\in\mathcal{T}_{0,T}}$ of random variables

$$v_\sigma := \operatorname*{ess\,sup}_{\tau\in\mathcal{T}_{\sigma,T}} \mathbb{E}[G_\tau|\mathcal{F}_\sigma], \quad \sigma \in \mathcal{T}_{0,T}. \tag{1.2}$$

The random variable v_σ is the optimal conditional expected reward for stopping at time σ or later. Note that, since each deterministic time $t \in [0, T]$ is also a stopping time, (1.2) defines an adapted *value* process $(v_t)_{0\le t\le T}$.

Let $\sigma \in \mathcal{T}_{0,T}$ and $\tau \in \mathcal{T}_{\sigma,T}$. Then, by Lemma 1.3.1, the family $\{\mathbb{E}[G_\rho|\mathcal{F}_\sigma]\}_{\rho\in\mathcal{T}_{\tau,T}}$ is *directed upwards*. Therefore, from the properties of essential supremum (see, for example, Lemma 1.3 in Peskir and Shiryaev [90]), there exists a sequence $\{\rho_n\}_{n\ge1}$ of stopping times in $\mathcal{T}_{\tau,T}$ such that the sequence $\{\mathbb{E}[G_{\rho_n}|\mathcal{F}_\sigma]\}_{n\ge1}$ is non-decreasing and

$$\operatorname*{ess\,sup}_{\rho\in\mathcal{T}_{\tau,T}} \mathbb{E}[G_\rho|\mathcal{F}_\sigma] = \lim_{n\to\infty} \mathbb{E}[G_{\rho_n}|\mathcal{F}_\sigma] \quad \mathbb{P}\text{-a.s.} \tag{1.3}$$

Then the definition of v_τ, (1.3) and the monotone convergence theorem for conditional expectations give

$$\mathbb{E}[v_\tau|\mathcal{F}_\sigma] \leq \operatorname*{ess\,sup}_{\rho\in\mathcal{T}_{\tau,T}} \mathbb{E}[G_\rho|\mathcal{F}_\sigma] \quad \mathbb{P}\text{-a.s.}$$

On the other hand, the reverse inequality also holds. To see this note that, since, for any $\rho \in \mathcal{T}_{\tau,T}$, $v_\tau \geq \mathbb{E}[G_\rho|\mathcal{F}_\tau]$ $\mathbb{P}$-a.s., taking conditional expectations yields $\mathbb{E}[v_\tau|\mathcal{F}_\sigma] \geq \mathbb{E}[G_\rho|\mathcal{F}_\sigma]$ $\mathbb{P}$-a.s., and therefore $\mathbb{E}[v_\tau|\mathcal{F}_\sigma] \geq \operatorname{ess\,sup}_{\rho\in\mathcal{T}_{\tau,T}} \mathbb{E}[G_\rho|\mathcal{F}_\sigma]$ $\mathbb{P}$-a.s. This implies that

$$\mathbb{E}[v_\tau|\mathcal{F}_\sigma] = \operatorname*{ess\,sup}_{\rho\in\mathcal{T}_{\tau,T}} \mathbb{E}[G_\rho|\mathcal{F}_\sigma] \quad \mathbb{P}\text{-a.s.} \tag{1.4}$$

Several implications of (1.4) follow. Since $\tau \in \mathcal{T}_{\sigma,T}$, $\mathcal{T}_{\tau,T} \subseteq \mathcal{T}_{\sigma,T}$, and therefore $\operatorname{ess\,sup}_{\rho\in\mathcal{T}_{\tau,T}} \mathbb{E}[G_\rho|\mathcal{F}_\sigma] \leq \operatorname{ess\,sup}_{\rho\in\mathcal{T}_{\sigma,T}} \mathbb{E}[G_\rho|\mathcal{F}_\sigma]$ $\mathbb{P}$-a.s. Hence, $\mathbb{E}[v_\tau|\mathcal{F}_\sigma] \leq v_\sigma$ $\mathbb{P}$-a.s., which proves that the value process $(v_t)_{0\leq t\leq T}$ is a supermartingale. Moreover, setting $\sigma = 0$ in (1.4) and using supermartingale property of $(v_t)_{0\leq t\leq T}$, we have that $\mathbb{E}[v_\tau] = \sup_{\rho\in\mathcal{T}_{\tau,T}} \mathbb{E}[G_\rho] \leq v_0 < \infty$.

For $\sigma \in \mathcal{T}_{0,T}$, it is tempting to regard v_σ as the process $(v_t)_{0\leq t\leq T}$ evaluated at the stopping time σ. It turns out that there is indeed a modification $(S_t)_{0\leq t\leq T}$ of the process $(v_t)_{0\leq t\leq T}$ that aggregates the family $\{v_\sigma\}_{\sigma\in\mathcal{T}_{0,T}}$ at each stopping time σ, i.e. $v_\sigma(\omega) = S_{\sigma(\omega)}(\omega)$ for $\mathbb{P}$-a.e. $\omega \in \Omega$. In particular, one can show (using the right-continuity of G and supermartingale property of $(v_t)_{0\leq t\leq T}$) that $t \mapsto \mathbb{E}[v_t]$ is right-continuous (see Theorem 2.2 in Peskir and Shiryaev [90] or Proposition D.3 in Karatzas and Shreve [71] for the details). It is well-known (see, for example, Theorem 1.3.13 in Karatzas and Shreve [70]) that this is necessary and sufficient for the existence of an adapted càdlàg modification of $(v_t)_{0\leq t\leq T}$, which we denote by $S = (S_t)_{0\leq t\leq T}$. This process S is the Snell envelope of G. From the Optional Sampling theorem and right-continuity of G and S, we have that S is the smallest supermartingale dominating G and, in addition, it aggregates the supermartingale $(v_t)_{0\leq t\leq T}$ (see Theorem D.7 in Karatzas and Shreve [71]):

Theorem 1.1.2 (Characterisation of S)**.** *Let $G \in \bar{\mathbb{G}}$. For every $\sigma \in \mathcal{T}_{0,T}$, the Snell envelope process S of G satisfies*

$$S_\sigma = \operatorname*{ess\,sup}_{\tau\in\mathcal{T}_{\sigma,T}} \mathbb{E}[G_\tau|\mathcal{F}_\sigma] \quad \mathbb{P}\text{-}a.s. \tag{1.5}$$

Moreover, S is the minimal càdlàg supermartingale that dominates G.

For the proof of Theorem 1.1.2 under slightly more general assumptions on the gains process G (i.e. G is assumed to be *optional* process) consult Appendix I in Dellacherie and Meyer [31] or Proposition 2.26 in El Karoui [38].

Remark 1.1.3. *The existence of a finite Snell envelope does not require condition (1.1), however so-called prophet inequalities (see, for example, Hill and Kertz [52] and Assaf, Goldstein and Samuel-Cahn [4]) show that the gap may be small. It is also easy to prove, using the Optional Section Theorem and Markov's Inequality that if G is a non-negative optional process and S is finite then $\sup_t G_t \in L^p$ for every $p < 1$.*

If $G \in \bar{\mathbb{G}}$, it is clear that G is a uniformly integrable process. In particular, it is also of class (D), i.e. the family of random variables $\{G_\tau 1_{\{\tau<\infty\}} : \tau \in \mathcal{T}_{0,T}\}$ is uniformly integrable. On the other hand, a right-continuous adapted process Z belongs to the class (D) if there exists a uniformly integrable martingale $\hat{N}$, such that, for all $t \in [0,T]$, $|Z_t| \leq \hat{N}_t$ $\mathbb{P}$-a.s. (see Appendix I in Dellacherie and Meyer [31] and references therein, or, alternatively, Theorem D.13 in Karatzas and Shreve [71]). In our case, by (1.5) and using the conditional version of Jensen's inequality, for $t \in [0,T]$, we have

$$|S_t| \leq \mathbb{E}\Big[\sup_{0\leq s\leq T} |G_s| \Big| \mathcal{F}_t\Big] := N_t \quad \mathbb{P}\text{-a.s.}$$

But, since $G \in \bar{\mathbb{G}}$, N is a uniformly integrable martingale, which proves the following

Lemma 1.1.4. *Suppose $G \in \bar{\mathbb{G}}$. Then S is of class (D).*

Let $\mathcal{M}_0$ denote the set of right-continuous martingales started at zero. Let $\mathcal{M}_{0,loc}$ and $\mathcal{M}_{0,UI}$ denote the spaces of local and uniformly integrable martingales (started at zero), respectively. Similarly, the adapted processes of finite and integrable variation will be denoted by FV and IV, respectively.

It is well-known that a right-continuous (local) supermartingale P has a unique decomposition $P = B - I$ where $B \in \mathcal{M}_{0,loc}$ and I is an increasing (FV) process which is predictable (see Theorem 16 in Protter [92] (p.116)). This can be regarded as the general Doob-Meyer decomposition of a supermartingale. Specialising to class (D) supermartingales we have a stronger result (see, for example, Theorem 11 in Protter [92] (p.112)):

Theorem 1.1.5 (Doob-Meyer decomposition)**.** *Let $G \in \bar{\mathbb{G}}$. Then the Snell envelope process S admits a unique decomposition*

$$S = M^* - A, \tag{1.6}$$

where $M^* \in \mathcal{M}_{0,UI}$, and A is a predictable, increasing IV process.

Remark 1.1.6. *It is normal to assume that the process A in the Doob-Meyer decomposition of S is started at zero. The duality result alluded to in the introduction is one reason why we do not do so here.*

An immediate consequence of Theorem 1.1.5 is that S is a semimartingale. In addition, we also assume that G is a semimartingale with the following decomposition:

$$G = N + D, \tag{1.7}$$

where $N \in \mathcal{M}_{0,loc}$ and D is a FV process. Unfortunately, the decomposition (1.7) is not, in general, unique:

Example 1.1.7. *Let G be a semimartingale with the following decomposition: $G_t = G_0 + N_t + D_t$ with $N_0 = 0 = D_0$. If the underlying filtered probability space $(\Omega, \mathcal{F}, (\mathcal{F}_t)_{0 \le t \le T}, \mathbb{P})$ supports a Poisson process $P = (P_t)_{0 \le t \le T}$ (which is independent of G), then*

$$G_t = G_0 + [N_t + P_t - t] + [A_t + t - P_t]$$

is another decomposition of G.

On the other hand, uniqueness *is* obtained by requiring the FV term to also be predictable, at the cost of restricting only to locally integrable processes. If there exists a decomposition of a semimartingale X with a predictable FV process, then we say that X is *special*. For a special semimartingale we always choose to work with its *canonical* decomposition (so that a FV process is predictable). Let

$$\mathbb{G} \text{ be the space of semimartingales in } \bar{\mathbb{G}}.$$

Due to the integrability condition (1.1), we have the following (see Theorems 36 and 37 in Protter [92] (p.132))

Lemma 1.1.8. *Suppose $G \in \mathbb{G}$. Then G is a special semimartingale.*

The following lemma provides a further decomposition of a semimartingale (see Proposition 3.3 in Jacod and Shiryaev [65] (p.27)). In particular, the FV term of a special semimartingale can be uniquely (up to initial values) decomposed in a predictable way, into the difference of two increasing, mutually singular FV processes.

Lemma 1.1.9. *Suppose that K is a càdlàg, adapted process such that $K \in FV$. Then there exists a unique pair (K^+, K^-) of adapted increasing processes such that $K - K_0 = K^+ - K^-$ and $\int |dK_s| = K^+ + K^-$. Moreover, if K is predictable, then K^+, K^- and $\int |dK_s|$ are also predictable.*

1.1.2 Main results

The assumption that $G \in \mathbb{G}$ (i.e. G is a semimartingale with integrable supremum and $G = N + D$ is its canonical decomposition), neither ensures that $N \in \mathcal{M}_0$, nor that D is an IV process, the latter, it turns out, being sufficient for the main result of this section to hold. In order to prove Theorem 1.1.15 we will need a stronger integrability condition on G.

For any adapted càdlàg process H, define

$$H^* = \sup_{0 \le t \le T} |H_t| \tag{1.8}$$

and

$$||H||_{\mathcal{S}^p} = ||H^*||_{L^p} := \mathbb{E}\big[|H^*|^p\big]^{1/p}, \quad 1 \le p \le \infty. \tag{1.9}$$

Remark 1.1.10. *Note that $\bar{\mathbb{G}} = \mathcal{S}^1$, so that under the current conditions we have that $G \in \mathcal{S}^1$.*

For a special semimartingale X with canonical decomposition $X = \bar{B} + \bar{I}$, where $\bar{B} \in \mathcal{M}_{0,loc}$ and $\bar{I}$ is a predictable FV process (with $I_0 = X_0$), define the $\mathcal{H}^p$ norm, for $1 \le p \le \infty$, by

$$||X||_{\mathcal{H}^p} = ||\bar{B}||_{\mathcal{S}^p} + \Big|\Big| \int_0^T |d\bar{I}_s| \Big|\Big|_{L^p} + ||I_0||_{L^p}, \tag{1.10}$$

and, as usual, write $X \in \mathcal{H}^p$ if $||X||_{\mathcal{H}^p} < \infty$.

Remark 1.1.11. *A more standard definition of the $\mathcal{H}^p$ norm is with $||\bar{B}||_{\mathcal{S}^p}$ replaced by $||[\bar{B}, \bar{B}]_T^{1/2}||_{L^p}$. However, the Burkholder-Davis-Gundy inequalities (see, for example, Theorem 48 in Protter [92] (p.195) and references therein) imply the equivalence of these norms.*

The following lemma follows from the fact that $\bar{I}^* \le \int_0^T |d\bar{I}_s| + |I_0|$, $\mathbb{P}$-a.s:

Lemma 1.1.12. *On the space of special semimartingales, the $\mathcal{H}^p$ norm is stronger than $\mathcal{S}^p$ for $1 \le p < \infty$, i.e. convergence in $\mathcal{H}^p$ implies convergence in $\mathcal{S}^p$.*

In general, it is challenging to check whether a given process belongs to $\mathcal{H}^1$, and thus the assumption that $G \in \mathcal{H}^1$ might be too stringent. On the other hand, under the assumptions in the Markov setting (see Assumption 1.2.9 in Section 1.2.2), we will have that G is *locally* in $\mathcal{H}^1$. Recall that a semimartingale X belongs to $\mathcal{H}^p_{loc}$, for $1 \le p \le \infty$, if there exists a sequence of stopping times $\{\sigma_n\}_{n \in \mathbb{N}}$, increasing to infinity almost surely, such that for each $n \ge 1$, the stopped process X^{σ_n} belongs to $\mathcal{H}^p$. Hence, the main assumption in this section is the following:

Assumption 1.1.13. *G is a semimartingale in both $\mathcal{S}^1$ and $\mathcal{H}^1_{loc}$.*

Remark 1.1.14. *Let G be an adapted, càdlàg semimartingale with the following decomposition*

$$G = N + D, \tag{1.11}$$

where N is a local martingale (started at zero) and D is an adapted FV process.

Assuming that $G \in \mathcal{H}^1$, Lemma 1.1.12 implies that Assumption 1.1.13 is satisfied, and thus all the results of Section 1.1.1 hold. Moreover, we then have that $N \in \mathcal{M}_{0,UI}$ and D is a predictable IV process.

On the other hand, under Assumption 1.1.13, N and D are only locally uniformly integrable martingale (started at zero) and the process of integrable variation, respectively, i.e. $N^{\sigma_n} \in \mathcal{M}_{0,UI}$ and I^{σ_n} is a predictable IV process, where $\{\sigma_n\}_{n\geq 1}$ is a localising sequence.

We finally arrive to the main result of this section:

Theorem 1.1.15. *Suppose Assumption 1.1.13 holds. Let D be a predictable FV process in the decomposition* (1.11) *of the gains process G. Let D^- (D^+) denote the decreasing (increasing) components of D, as in Lemma 1.1.9. Let A be a predictable, increasing IV process in the decomposition of the Snell envelope S (of G), as in Theorem 1.1.5.*

Then A is, as a measure, absolutely continuous with respect to D^- almost surely on $[0, T]$, and μ, defined by

$$\mu_t := \frac{dA_t}{dD^-_t}, \quad 0 \leq t \leq T,$$

has a version that satisfies $0 \leq \mu_t \leq 1$ almost surely.

Remark 1.1.16. *As is usual in semimartingale calculus, we treat a process of bounded variation and its corresponding Lebesgue-Stiltjes signed measure as synonymous.*

The proof of Theorem 1.1.15 is based on the discrete-time approximation of the predictable FV processes in the decompositions of S (1.6) and G (1.7). In particular, let $\mathcal{P}_n = \{0 = t^n_0 < t^n_1 < t^n_2 < ... < t^n_{k_n} = T\}$, $n = 1, 2, ...$, be an increasing sequence of partitions of $[0, T]$ (i.e. if, for $n < m$, $\mathcal{P}_n$ and $\mathcal{P}_m$ are two partitions of $[0, T]$, then we have that $k_n < k_m$ and for each $i \in \{0, ..., k_n\}$ there exists $j \in \{0, ..., k_m\}$ such that $t^n_i = t^m_j$) with $\max_{1\leq k\leq k_n} t^n_k - t^n_{k-1} \to 0$ as $n \to \infty$. Note that here $T < \infty$ is fixed, but arbitrary. Let $S^n_t = S_{t^n_k}$ if $t^n_k \leq t < t^n_{k+1}$ and

$S^n_T = S_T$ define the discretizations of S, and set

$$A^n_t = 0 \quad \text{if } 0 \le t < t^n_1,$$

$$A^n_t = \sum_{j=1}^{k} \mathbb{E}[S_{t^n_{j-1}} - S_{t^n_j} | \mathcal{F}_{t^n_{j-1}}] \quad \text{if } t^n_k \le t < t^n_{k+1},\ k = 1, 2, ..., k_n - 1,$$

$$A^n_T = \sum_{j=1}^{k_n} \mathbb{E}[S_{t^n_{j-1}} - S_{t^n_j} | \mathcal{F}_{t^n_{j-1}}].$$

If S is *regular*, in the sense that for every stopping time τ and nondecreasing sequence $(\tau_n)_{n\in\mathbb{N}}$ of stopping times with $\tau = \lim_{n\to\infty} \tau_n$ we have $\lim_{n\to\infty} \mathbb{E}[S_{\tau_n}] = \mathbb{E}[S_\tau]$, or equivalently, if A is continuous, Doléans [33] showed that $A^n_t \to A_t$ uniformly in L^1 as $n \to \infty$ (see also Theorem 31.2 (VI.31) in Rogers and Williams [97]). Hence, assuming that S is regular, we can extract a subsequence $\{A^{n_l}_t\}$, such that $\lim_{l\to\infty} A^{n_l}_t = A_t$ a.s. On the other hand, it is enough for G to be regular:

Lemma 1.1.17. *Suppose $G \in \bar{\mathbb{G}}$ is a regular gains process. Then so is its Snell envelope process S.*

See Section 1.3.2 for the proof of Lemma 1.1.17.

Remark 1.1.18. *If it is not known that G is regular, Kobylanski and Quenez [72], in a slightly more general setting, showed that S is still regular,* provided *that G is upper semicontinuous in expectation along stopping times, i.e. for all $\tau \in \mathcal{T}_{0,T}$ and for all sequences of stopping times $(\tau_n)_{n\ge1}$ such that $\tau_n \uparrow \tau$, we have*

$$\mathbb{E}[G_\tau] \ge \limsup_{n\to\infty} \mathbb{E}[G_{\tau_n}].$$

The case where S is not regular is more subtle. In his classical paper Rao [94] utilised the Dunford-Pettis compactness criterion and showed that, in general, $A^n_t \to A_t$ only *weakly* in L^1 as $n \to \infty$ (where a sequence $(X_n)_{n\in\mathbb{N}}$ of random variables in L^1 converges weakly in L^1 to X if for every bounded random variable Y we have that $\mathbb{E}[X_nY] \to \mathbb{E}[XY]$ as $n \to \infty$).

Recall that *weak* convergence in L^1 does not imply convergence in probability, and therefore, we cannot immediately deduce an almost sure convergence along a subsequence. (For example, let $([0,1], \mathcal{B}([0,1]), \lambda_{[0,1]})$ were λ is the Lebesgue measure, and take the following sequence of r.v.s $X_n(\omega) = sin(2\pi n\omega)$, $\omega \in [0,1]$, $n \ge 1$.) However, it turns out that by modifying the sequence of approximating random variables, the required convergence can be achieved. This has been done in recent improvements of the Doob-Meyer decomposition (see Jakubowski [66] and

Beiglböck et al. [11]. Also, Siorpaes [101] showed that there is a subsequence that works for all $(t,\omega) \in [0,T] \times \Omega$ simultaneously). In particular, Jakubowski proceeds as Rao, but then uses Komlós's theorem [74] (which can be viewed as a substitute for the Bolzano-Weierstrass Theorem in infinite dimensional spaces) and proves the following (Jakubowski [66], Theorem 3 and Remark 1):

Theorem 1.1.19. *There exists a subsequence $\{n_l\}$ such that for $t \in \cup_{n=1}^{\infty} \mathcal{P}_n$ and as $L \to \infty$*

$$\frac{1}{L}\Big(\sum_{l=1}^{L} A_t^{n_l}\Big) \to A_t, \quad \mathbb{P}\text{-}a.s. \text{ and in } L^1. \tag{1.12}$$

In particular, in any subsequence we can find a further subsequence such that (1.12) *holds.*

We are now ready to prove Theorem 1.1.15.

Proof of Theorem 1.1.15. Let $(\sigma_n)_{n\geq 1}$ be a localising sequence for G such that, for each $n \geq 1$, $G^{\sigma_n} = (G_{t\wedge\sigma_n})_{0\leq t\leq T}$ is in $\mathcal{H}^1$. Similarly, set $S^{\sigma_n} = (S_{t\wedge\sigma_n})_{0\leq t\leq T}$ for a fixed $n \geq 1$. We need to prove that

$$0 \leq A_t^{\sigma_n} - A_s^{\sigma_n} \leq (D^-)_t^{\sigma_n} - (D^-)_s^{\sigma_n} \quad \mathbb{P}\text{-a.s.}, \tag{1.13}$$

since then, as $\sigma_n \uparrow \infty$ $\mathbb{P}$-almost surely (as $n \to \infty$), and by uniqueness of A and D^-, the result follows. In particular, since A is increasing, the first inequality in (1.13) is immediate, and thus we only need to prove the second one.

After localisation we assume that $G \in \mathcal{H}^1$. For any $0 \leq t \leq T$ and $0 \leq \epsilon \leq T - t$ we have that

$$\begin{aligned}
\mathbb{E}[S_{t+\epsilon}|\mathcal{F}_t] &= \mathbb{E}\Big[\operatorname*{ess\,sup}_{\tau\in\mathcal{T}_{t+\epsilon,T}} \mathbb{E}[G_\tau|\mathcal{F}_{t+\epsilon}]\Big|\mathcal{F}_t\Big] \\
&\geq \mathbb{E}\Big[\mathbb{E}[G_\tau|\mathcal{F}_{t+\epsilon}]\Big|\mathcal{F}_t\Big] \\
&= \mathbb{E}[G_\tau|\mathcal{F}_t] \quad \mathbb{P}\text{-a.s.},
\end{aligned}$$

where $\tau \in \mathcal{T}_{t+\epsilon,T}$ is arbitrary. Therefore

$$\mathbb{E}[S_{t+\epsilon}|\mathcal{F}_t] \geq \operatorname*{ess\,sup}_{\tau\in\mathcal{T}_{t+\epsilon,T}} \mathbb{E}[G_\tau|\mathcal{F}_t] \quad \mathbb{P}\text{-a.s.} \tag{1.14}$$

Then by (1.5) and using (1.14) together with the properties of the *essential supremum* (see also Lemma 1.3.1) we obtain

$$\begin{aligned}
\mathbb{E}[S_t - S_{t+\epsilon}|\mathcal{F}_t] &\le \operatorname*{ess\,sup}_{\tau\in\mathcal{T}_{t,T}} \mathbb{E}[G_\tau|\mathcal{F}_t] - \operatorname*{ess\,sup}_{\tau\in\mathcal{T}_{t+\epsilon,T}} \mathbb{E}[G_\tau|\mathcal{F}_t] \\
&\le \operatorname*{ess\,sup}_{\tau\in\mathcal{T}_{t,T}} \mathbb{E}[G_\tau - G_{\tau\vee(t+\epsilon)}|\mathcal{F}_t] \\
&= \operatorname*{ess\,sup}_{\tau\in\mathcal{T}_{t,t+\epsilon}} \mathbb{E}[G_\tau - G_{\tau\vee(t+\epsilon)}|\mathcal{F}_t] \qquad (1.15) \\
&= \operatorname*{ess\,sup}_{\tau\in\mathcal{T}_{t,t+\epsilon}} \mathbb{E}[G_\tau - G_{t+\epsilon}|\mathcal{F}_t] \quad \mathbb{P}\text{-a.s.}
\end{aligned}$$

(1.15) follows by noting that $\mathcal{T}_{t+\epsilon,T} \subset \mathcal{T}_{t,T}$, and that for any $\tau \in \mathcal{T}_{t+\epsilon,T}$ the term inside the expectation vanishes. Using the decomposition of G and by observing that, for all $\tau \in \mathcal{T}_{t,t+\epsilon}$, $(D^+_\tau - D^+_{t+\epsilon}) \le 0$ $\mathbb{P}$-a.s. (since $\tau \le t+\epsilon$ and D^+ is non-decreasing, $\mathbb{P}$-a.s.), while N is a uniformly integrable martingale, we obtain

$$\begin{aligned}
\mathbb{E}[S_t - S_{t+\epsilon}|\mathcal{F}_t] &\le \operatorname*{ess\,sup}_{\tau\in\mathcal{T}_{t,t+\epsilon}} \mathbb{E}[D^-_{t+\epsilon} - D^-_\tau|\mathcal{F}_t] \\
&= \mathbb{E}[D^-_{t+\epsilon} - D^-_t|\mathcal{F}_t] \quad \mathbb{P}\text{-a.s.} \qquad (1.16)
\end{aligned}$$

Finally, for $0 \le s < t \le T$, applying Theorem 1.1.19 to A together with (1.16) gives

$$\begin{aligned}
A_t - A_s &= \lim_{L\to\infty} \frac{1}{L}\Big(\sum_{l=1}^{L}\sum_{j=k'}^{k} \mathbb{E}[S_{t^{n_l}_{j-1}} - S_{t^{n_l}_j}|\mathcal{F}_{t^{n_l}_{j-1}}]\Big) \\
&\le \lim_{L\to\infty} \frac{1}{L}\Big(\sum_{l=1}^{L}\sum_{j=k'}^{k} \mathbb{E}[D^-_{t^{n_l}_j} - D^-_{t^{n_l}_{j-1}}|\mathcal{F}_{t^{n_l}_{j-1}}]\Big) \quad \mathbb{P}\text{-a.s.,} \qquad (1.17)
\end{aligned}$$

where $k' \le k$ are such that $t^{n_l}_{k'} \le s < t^{n_l}_{k'+1}$ and $t^{n_l}_k \le t < t^{n_l}_{k+1}$. Note that D^- is also the predictable, increasing IV process in the Doob-Meyer decomposition of the class (D) supermartingale $(G - D^+)$. Therefore we can approximate it in the same way as A, so that $(D^-_t - D^-_s)$ is the almost sure limit along, possibly, a further subsequence $\{n_{l_k}\}$ of $\{n_l\}$, of the right hand side of (1.17). □

We finish this section with a lemma that gives an easy test as to whether the given process belongs to $\mathcal{H}^1_{loc}$ (consult Section 1.3.2 for the proof). It will be relevant in the Markovian treatment of the optimal stopping problem considered in the following section.

Lemma 1.1.20. *Let $X \in \mathbb{G}$ with a canonical decomposition $X = L + K$, where $L \in \mathcal{M}_{0,loc}$ and K is a predictable FV process. If the jumps of K are uniformly* *bounded by some finite constant $c > 0$, then $X \in \mathcal{H}^1_{loc}$.*

1.2 Markovian setting

1.2.1 Preliminaries

The Markov process Let $(E, \mathcal{E})$ be a metrizable Lusin space endowed with the σ-field of Borel subsets of E. Let $X = (\Omega, \mathcal{G}, \mathcal{G}_t, X_t, \theta_t, \mathbb{P}_x : x \in E, t \in \mathbb{R}_+)$ be a Markov process taking values in $(E, \mathcal{E})$. We assume that a sample space Ω is such that the usual semi-group of shift operators $(\theta_t)_{t \geq 0}$ is well-defined (which is the case, for example, if $\Omega = E^{[0,\infty)}$ is the canonical path space). If the corresponding semigroup of X, (P_t), is the primary object of study, then we say that X is a realisation of a Markov semigroup (P_t). In the case of (P_t) being sub-Markovian, i.e. $P_t 1_E \leq 1_E$, we extend it to a Markovian semigroup over $E^\Delta = E \cup \{\Delta\}$, where Δ is a coffin-state. We also denote by $\mathcal{C}(X) = (\Omega, \mathcal{F}, \mathcal{F}_t, X_t, \theta_t, \mathbb{P}_x : x \in E, t \in \mathbb{R}_+)$ the *canonical realisation* associated with X, defined on Ω with the filtration $(\mathcal{F}_t)$ deduced from $\mathcal{F}^0_t = \sigma(X_s : s \leq t)$ by standard regularisation procedures (universal completeness, with respect to all $\mathbb{P}_x, x \in E$, and right-continuity).

In this section our standing assumption is that the underlying Markov process X is a *right process* (consult Getoor [43], Sharpe [99] for the general theory). Essentially, right processes are the processes satisfying Meyer's regularity hypotheses (*hypothèses droites*) HD1 and HD2. If a given Markov semigroup (P_t) satisfies HD1 and μ is an arbitrary probability measure on $(E, \mathcal{E})$, then there exists a homogeneous E-valued Markov process X with transition semigroup (P_t) and initial law μ. Moreover, a realisation of such (P_t) is right-continuous (Sharpe [99], Theorem 2.7). Under the second fundamental hypothesis, HD2, $t \mapsto f(X_t)$ is right-continuous for every α-excessive function f. Recall, for $\alpha > 0$, a universally measurable function $f : E \mapsto \mathbb{R}$ is α-super-median if $e^{-\alpha t} P_t f \leq f$ for all $t \geq 0$, and α-excessive if it is α-super-median and $e^{-\alpha t} P_t f \to f$ as $t \to 0$. If (P_t) satisfies HD1 and HD2 then the corresponding realisation X is strong Markov (Getoor [43], Theorem 9.4 and Blumenthal and Getoor [15], Theorem 8.11).

Remark 1.2.1. *One has the following inclusions among classes of Markov processes:*

$$\text{(Feller)} \subset \text{(Hunt)} \subset \text{(right)}$$

Let $\mathcal{L}$ be a given extended infinitesimal (martingale) generator of X with a domain $\mathbb{D}(\mathcal{L})$, i.e. we say a Borel function $f : E \mapsto \mathbb{R}$ belongs to $\mathbb{D}(\mathcal{L})$ if there exists a Borel function $h : E \mapsto \mathbb{R}$, such that $\int_0^t |h(X_s)| ds < \infty$, $\forall t \geq 0$, $\mathbb{P}_x$-a.s. for each x

and the process $M^f = (M^f_t)_{t\geq 0}$, given by

$$M^f_t := f(X_t) - f(x) - \int_0^t h(X_s)ds, \quad t \geq 0,\ x \in E, \tag{1.18}$$

is a local martingale under each $\mathbb{P}_x$ (see Revuz and Yor [95] p.285), and then we write $h = \mathcal{L}f$.

Remark 1.2.2. *Note that if $A \in \mathcal{E}$ and $\mathbb{P}_x(\lambda(\{t : \ X_t \in A\} = 0) = 1$ for each $x \in E$, where λ is Lebesgue measure, then h may be altered on A without affecting the validity of* (1.18)*, so that, in general, the map $f \mapsto h$ is not unique. This is why we refer to* a *martingale generator.*

Optimal stopping problem Let $X = (\Omega, \mathcal{G}, \mathcal{G}_t, X_t, \theta_t, \mathbb{P}_x : x \in E, t \in \mathbb{R}_+)$ be a right process. Given a function $g : E \mapsto \mathbb{R}$ (with $g(\Delta) = 0$), $\alpha \geq 0$ and $T \in \mathbb{R}_+ \cup \{\infty\}$ define a corresponding gains process G^α (we simply write G if $\alpha = 0$) by $G^\alpha_t = e^{-\alpha t} g(X_t)$ for $t \in [0, T]$. In the case of $T = \infty$, we make a convention that $G^\alpha_\infty = \liminf_{t\to\infty} G^\alpha_t$. Let $\mathcal{E}^e, \mathcal{E}^u$ be the σ-algebras on E generated by α-excessive functions and universally measurable sets, respectively (recall that $\mathcal{E} \subset \mathcal{E}^e \subset \mathcal{E}^u$). We write

$$g \in \mathcal{Y}, \text{ given that } g(\cdot) \text{ is } \mathcal{E}^e\text{-measurable and } G^\alpha \text{ is of class (D).}$$

For a filtration $(\hat{\mathcal{G}}_t)$, and $(\hat{\mathcal{G}}_t)$ - stopping times σ_1 and σ_2, with $\mathbb{P}_x[0 \leq \sigma_1 \leq \sigma_2 \leq T] = 1$, $x \in E$, let $\mathcal{T}_{\sigma_1,\sigma_2}(\hat{\mathcal{G}})$ be the set of $(\hat{\mathcal{G}}_t)$ - stopping times τ with $\mathbb{P}_x[\sigma_1 \leq \tau \leq \sigma_2] = 1$. Consider the following optimal stopping problem:

$$V(x) = \sup_{\tau\in\mathcal{T}_{0,T}(\mathcal{G})} \mathbb{E}_x[e^{-\alpha\tau} g(X_\tau)], \quad x \in E.$$

By convention we set $V(\Delta) = g(\Delta)$. The following result is due to (Theorems 1.7 and 3.4 in) El Karoui et al. [40].

Theorem 1.2.3. *Let $X = (\Omega, \mathcal{G}, \mathcal{G}_t, X_t, \theta_t, \mathbb{P}_x : x \in E, t \in \mathbb{R}_+)$ be a right process with canonical filtration $(\mathcal{F}_t)$. If $g \in \mathcal{Y}$, then*

$$V(x) = \sup_{\tau\in\mathcal{T}_{0,T}(\mathcal{F})} \mathbb{E}_x[e^{-\alpha\tau} g(X_\tau)], \quad x \in E,$$

and $(e^{-\alpha t} V(X_t))$ is a Snell envelope of G^α, i.e. for all $x \in E$ and $\tau \in \mathcal{T}_{0,T}(\mathcal{F})$

$$e^{-\alpha\tau} V(X_\tau) = \operatorname*{ess\,sup}_{\sigma\in\mathcal{T}_{\tau,T}(\mathcal{F})} \mathbb{E}_x[G^\alpha_\sigma | \mathcal{F}_\tau] \quad \mathbb{P}_x\text{-}a.s.$$

The first important consequence of the theorem is that we can (and will) work with the canonical realisation $\mathcal{C}(X)$. The second one provides a crucial link between the Snell envelope process in the general setting and the value function in the Markovian framework.

Remark 1.2.4. *The restriction to gains processes of the form $G = g(X)$ (or G^α if $\alpha > 0$) is* much less restrictive than might appear. *Given that we work on the canonical path space with θ being the usual shift operator, we can expand the state-space of X by appending an adapted functional F, taking values in the space $(E', \mathcal{E}')$, with the property that*

$$\{F_{t+s} \in A\} \in \sigma(F_s) \cup \sigma(\theta_s \circ X_u : \ 0 \leq u \leq t), \quad \textit{for all } A \in \mathcal{E}'. \tag{1.19}$$

This allows us to deal with time-dependent problems, running rewards and other path-functionals of the underlying Markov process.

Lemma 1.2.5. *Suppose X is a canonical Markov process taking values in the space $(E, \mathcal{E})$ where E is a locally compact, countably based Hausdorff space and $\mathcal{E}$ is its Borel σ-algebra. Suppose also that F is a path functional of X satisfying* (1.19) *and taking values in the space $(E', \mathcal{E}')$ where E' is a locally compact, countably based Hausdorff space with Borel σ-algebra $\mathcal{E}'$, then, defining $Y = (X, F)$, Y is still Markovian. If X is a strong Markov process and F is right-continuous, then Y is strong Markov. If X is a Feller process and F is right-continuous , then Y is strong Markov, has a càdlàg modification and the completion of the natural filtration of X, $\mathbb{F}$, is right-continuous and quasi-left continuous, and thus Y is a right process.*

The proof is reported in Section 1.3.1.

Example 1.2.6. *If X is a one-dimensional Brownian motion and $\alpha, f : \mathbb{R} \mapsto \mathbb{R}$ Borel functions, then Y, defined by*

$$Y_t = \left(X_t, L_t^0, \sup_{0 \leq s \leq t} X_s, \int_0^t \exp(-\int_0^s \alpha(X_u) du) f(X_s) ds \right), \quad t \geq 0,$$

where L^0 is the local time of X at 0, is a Feller process on the filtration of X.

1.2.2 Main results

In the rest of the section (and the chapter) we consider the following optimal stopping problem:

$$V(x) = \sup_{\tau \in \mathcal{T}_{0,T}} \mathbb{E}_x[g(X_\tau)], \quad x \in E, \tag{1.20}$$

for a measurable function $g : E \mapsto \mathbb{R}$ and a Markov process X satisfying the following set of assumptions:

Assumption 1.2.7. *X is a right process.*

Assumption 1.2.8. $\sup_{0\leq t\leq T}|g(X_t)|\in L^1(\mathbb{P}_x)$, *$x \in E$.*

Assumption 1.2.9. *$g \in \mathbb{D}(\mathcal{L})$, i.e. $g(\cdot)$ belongs to the domain of a martingale generator of X.*

Remark 1.2.10. *Lemma 1.2.5 tells us that if X is Feller and F is an adapted path-functional of the form given in* (1.19) *then (a modification of) (X, F) satisfies Assumption 1.2.7.*

Example 1.2.11. *Let $X = (X_t)_{t\geq 0}$ be a Markov process and let $\mathbb{D}(\hat{\mathcal{L}})$ be the domain of a classical infinitesimal generator of X, i.e. the set of measurable functions $f : E \mapsto \mathbb{R}$, such that $\lim_{t\to 0}(\mathbb{E}_x[f(X_t)] - f(x))/t$ exists. Then $\mathbb{D}(\hat{\mathcal{L}}) \subset \mathbb{D}(\mathcal{L})$. In particular,*

1. *if $X = (X_t)_{t\geq 0}$ is a solution of an SDE driven by a Brownian motion in $\mathbb{R}^d$, then $C_b^2(\mathbb{R}^d, \mathbb{R}) \subset \mathbb{D}(\hat{\mathcal{L}})$;*
2. *if the state space E is finite (so that X is a continuous time Markov chain), then any measurable and bounded $f : E \mapsto \mathbb{R}$ belongs to $\mathbb{D}(\hat{\mathcal{L}})$*
3. *if X is a Lévy process on $\mathbb{R}^d$ with finite variance increments then $C_b^2(\mathbb{R}^d, \mathbb{R}) \subset \mathbb{D}(\hat{\mathcal{L}})$*

Note that the gains process is of the form $G = g(X)$, while by Theorem 1.2.3, the corresponding Snell envelope is given by

$$S_t^T := \begin{cases} V(X_t) : t < T, \\ g(X_T) : t \geq T. \end{cases}$$

In a similar fashion to that in the general setting, Assumption 1.2.8 ensures the class (D) property for the gains and Snell envelope processes. Moreover, under Assumption 1.2.9,

$$g(X_t) = g(x) + M_t^g + \int_0^t \mathcal{L}g(X_s)ds, \quad 0 \leq t \leq T,\, x \in E, \tag{1.21}$$

and the FV process in the semimartingale decomposition of $G = g(X)$ is absolutely continuous with respect to Lebesgue measure, and therefore predictable, so that (1.21) is a canonical semimartingale decomposition of $G = g(X)$. Then, by Assumption 1.2.8, and using Lemma 1.1.20, we also deduce that $g(X) \in \mathcal{H}^1_{loc}$.

Remark 1.2.12. *When $T < \infty$, the optimal stopping problem, in general, is time-inhomogeneous, and we need to replace the process X_t by the process $Z_t = (t, X_t)$, $t \in [0, T]$, so that* (1.20) *reads*

$$\tilde{V}(t, x) = \sup_{\tau \in \mathcal{T}_{0,T-t}} \mathbb{E}_{t,x}[\tilde{g}(t + \tau, X_{t+\tau})], \quad x \in E, \tag{1.22}$$

where $\tilde{g} : [0, T] \times E \mapsto \mathbb{R}$ is a new payoff function (consult Peskir and Shiryaev [90] for examples). In this case, Assumption 1.2.9 should be replaced by a requirement that there exists a measurable function $\tilde{h} : [0, T] \times E \mapsto \mathbb{R}$ such that $M_t^{\tilde{g}} := \tilde{g}(Z_t) - \tilde{g}(0, x) - \int_0^t \tilde{h}(Z_s)ds$ defines a local martingale.

The crucial result of this section is the following:

Theorem 1.2.13. *Suppose Assumptions 1.2.7 to 1.2.9 hold. Then $V \in \mathbb{D}(\mathcal{L})$.*

Proof. In order to be consistent with the notation in the general framework, let

$$D_t := g(X_0) + \int_0^t \mathcal{L}g(X_s)ds, \quad 0 \le t \le T.$$

Recall Lemma 1.1.9. Then D^+ and D^- are explicitly given (up to initial values) by

$$D_t^+ := \int_0^t \mathcal{L}g(X_s)^+ ds,$$
$$D_t^- := \int_0^t \mathcal{L}g(X_s)^- ds.$$

In particular, D^- is, as a measure, absolutely continuous with respect to Lebesgue measure. By applying Theorem 1.1.15, we deduce that

$$V(X_t) = V(x) + M_t^* - \int_0^t \mu_s \mathcal{L}g(X_s)^- ds, \quad 0 \le t \le T, \ x \in \mathbb{R}, \tag{1.23}$$

where μ is a non-negative Radon-Nikodym derivative with $0 \le \mu_s \le 1$. Then we also have that $\int_0^t |\mu_s \mathcal{L}g(X_s)^-| ds < \infty$, for every $0 \le t \le T$.

In order to finish the proof we are left to show that there exists a suitable measurable function $\lambda : E \mapsto \mathbb{R}$ such that $A_t = \int_0^t \mu_s \mathcal{L}g(X_s)^- ds = \int_0^t \lambda(X_s)ds$ $\mathbb{P}$-a.s., for all $t \in [0, T]$. For this, recall that a process Z (on $(\Omega, \mathcal{G}, \mathcal{G}_t, X_t, \theta_t, \mathbb{P}_x : x \in E, t \in \mathbb{R}_+)$ or just on $\mathcal{C}(X)$) is *additive* if $Z_0 = 0$ $\mathbb{P}$-a.s. and $Z_{t+s} = Z_t + Z_s \circ \theta_t$ $\mathbb{P}$-a.s., for all $s, t \in [0, T]$. Then, for any measurable function $f : E \mapsto \mathbb{R}$, $Z_t^f = f(X_t) - f(x)$ defines an additive process. (Çinlar et al. [23] gives necessary and sufficient conditions for Z^f to be a semimartingale.) More importantly, if Z^f is a semimartingale, then the martingale and FV processes in the decomposition of Z^f are also additive, see Theorem 3.18 in Çinlar et al. [23].

Set $K_t = \liminf_{s\downarrow 0, s\in\mathbb{Q}}(A_{t+s} - A_t)/s$ and $\beta(x) = \mathbb{E}_x[K_0]$, $x \in E$. We have that $A_t = \int_0^t \mu_s \mathcal{L}g(X_s)^- ds$, $t \in [0,T]$, is an increasing additive process which is absolutely continuous with respect to Lebesgue measure, $\mathbb{P}_x$-a.s., $x \in E$. For every ω such that $dA_t(\omega) \ll dt$ we then have $A_t(\omega) = \int_0^t K_t(\omega) ds$ for all $0 \le t \le T$. Now, from the additivity of A we have that $K_t = K_0 \circ \theta_t$, $\mathbb{P}_x$-a.s. Therefore, using the Markov property we have that

$$\beta(X_t) = \mathbb{E}_x[K_0 \circ \theta_t | \mathcal{F}_t] = K_t, \quad \mathbb{P}_x\text{-a.s., for all } 0 \le t \le T. \tag{1.24}$$

Define $J = \{(\omega, t) : K_t(\omega) \neq \beta(X_t(\omega))\}$, $J_\omega = \{t : (\omega, t) \in J\}$ and $J_t = \{\omega : (\omega, t) \in J\}$. Then, by (1.24), we have that $\mathbb{P}_x[J_t] = 0$, for all $0 \le t \le T$.

Finally, Proposition 3.56 in Çinlar et al. [23] shows that J is measurable with respect to ($d\mathbb{P}_x \otimes dt$-completion of) $\mathcal{F} \times [0,T]$, and therefore we can apply Fubini's theorem:

$$\int_0^T \mathbb{P}_x[J_t] dt = 0 = \int_\Omega \Big(\int_0^T I_{J_\omega}(t) dt \Big) d\mathbb{P}_x(\omega).$$

We conclude that, for $\mathbb{P}_x-$almost every ω, J_ω is a set of Lebesgue measure zero, and therefore, for $t \in [0,T]$, $A_t = \int_0^t \beta(X_s) ds$ $\mathbb{P}_x$-a.s. for each $x \in E$. □

Remark 1.2.14. *In some specific examples it is possible to relax Assumption 1.2.9. Let $\mathcal{S} := \{x \in E : V(x) = g(x)\}$ be the stopping region. It is well-known that $S = V(X)$ is a martingale on the go region $\mathcal{S}^c$, i.e. M^c given by*

$$M_t^c \stackrel{def}{=} \int_0^t 1_{(X_{s-} \in \mathcal{S}^c)} dS_s$$

is a martingale (see Lemma 1.3.2). This implies that $\int_0^t 1_{(X_{s-}\in\mathcal{S}^c)} dA_s = 0$, and therefore we note that in order for $V \in \mathbb{D}(\mathcal{L})$, we need D to be absolutely continuous with respect to Lebesgue measure λ only on the stopping region, i.e. that $\int_0^\cdot 1_{(X_{s-}\in\mathcal{S})} dD_s \ll \lambda$.

For example, let $E = \mathbb{R}$, fix $K \in \mathbb{R}_+$ and consider the optimal stopping problem (1.36), with $g(\cdot)$ given by $g(x) = (K - x)^+$, $x \in E$. Then, in general, $g \notin \mathbb{D}(\mathcal{L})$ (e.g. a consequence of the Tanaka's formula in the Brownian case). On the other hand, we can easily show, under very weak conditions, that $\mathcal{S} \subset [0, K)$. Therefore, for V to belong to $\mathbb{D}(\mathcal{L})$, we need only have that $\int_0^\cdot 1_{(X_{s-}<K)} dD_s$ is absolutely continuous (for example, this holds if X is a Brownian motion). However, the last condition depends on the underlying process and needs to be verified case-by-case.

1.2.3 Application: duality

Let $x \in E$ be fixed. As before, let $\mathcal{M}^x_{0,UI}$ denote all the right-continuous uniformly integrable càdlàg martingales (started at zero) on the filtered space $(\Omega, \mathcal{F}, \mathbb{F}, \mathbb{P}_x)$. Let $S = V(X)$ be the Snell envelope of $G = g(X)$. The main result of Rogers [96] in the Markovian setting reads:

Theorem 1.2.15. *Suppose Assumptions 1.2.7 and 1.2.8 hold. Then*

$$V(x) = \sup_{\tau \in \mathcal{T}_{0,T}} \mathbb{E}_x[G_\tau] = \inf_{M \in \mathcal{M}^x_{0,UI}} \mathbb{E}_x\Big[\sup_{0 \le t \le T}\Big(G_t - M_t\Big)\Big], \quad x \in E. \tag{1.25}$$

Proof. Fix $x \in E$. For any $M \in \mathcal{M}^x_{0,UI}$ we have

$$V(x) = \sup_{\tau \in \mathcal{T}_{0,T}} \mathbb{E}_x[G_\tau - M_\tau] \le \mathbb{E}_x\Big[\sup_{0 \le t \le T}\Big(G_t - M_t\Big)\Big]$$

and optimising over $\mathcal{M}^x_{0,UI}$ gives

$$V(x) \le \inf_{M \in \mathcal{M}^x_{0,UI}} \mathbb{E}_x\Big[\sup_{0 \le t \le T}\Big(G_t - M_t\Big)\Big], \quad x \in E.$$

On the other hand, suppose $S = V(X)$ has the following decomposition

$$V(X_t) = V(x) + M^*_t - A^*_t,$$

where $M^* \in \mathcal{M}^x_{0,UI}$ and A^* is an adapted, increasing process of integrable variation, with $A^*_0 = 0$. Then $G_t \le V(x) + M^*_t - A^*_t$, $\mathbb{P}_x$-a.s., and therefore

$$\begin{aligned}
\inf_{M \in \mathcal{M}^x_{0,UI}} \mathbb{E}_x\Big[\sup_{0 \le t \le T}\Big(G_t - M_t\Big)\Big] &\le \mathbb{E}_x\Big[\sup_{0 \le t \le T}\Big(G_t - M^*_t\Big)\Big] \\
&\le \mathbb{E}_x\Big[\sup_{0 \le t \le T}\Big(V(x) - A^*_t\Big)\Big] \\
&= V(x), \quad x \in E.
\end{aligned}$$

□

We call the right hand side of (1.25) the *dual* of the optimal stopping problem. In particular, the right hand side of (1.25) is a 'generalised stochastic control problem of Girsanov type', where a controller is allowed to choose a martingale from $\mathcal{M}^x_{0,UI}$, $x \in E$. Note that an optimal martingale for the dual is M^*, the martingale appearing in the Doob-Meyer decomposition of S, while any other martingale in $\mathcal{M}^x_{0,UI}$ gives an upper bound of $V(x)$. We already showed that $M^* = M^V$, which means that, when solving the dual problem, one can search only over martingales of the form M^f, for $f \in \mathbb{D}(\mathcal{L})$, or equivalently over the functions $f \in \mathbb{D}(\mathcal{L})$. We can further define $\mathcal{D}_{\mathcal{M}_{0,UI}} \subset \mathbb{D}(\mathcal{L})$ by

$$\mathcal{D}_{\mathcal{M}_{0,UI}} := \{f \in \mathbb{D}(\mathcal{L}) : f \geq g, f \text{ is superharmonic}, M^f \in \mathcal{M}_{0,UI}\}.$$

To conclude that $V \in \mathcal{D}_{\mathcal{M}_{0,UI}}$ we need to show that V is superharmonic, i.e. for all stopping times $\sigma \in \mathcal{T}_{0,T}$ and all $x \in E$, $\mathbb{E}_x[V(X_\sigma)] \leq V(x)$. But this follows immediately from the Optional Sampling theorem, since $S = V(X)$ is a uniformly integrable supermartingale. Hence, as expected, we can restrict our search for the best minimising martingale to the set $\mathcal{D}_{\mathcal{M}_{0,UI}}$.

Theorem 1.2.16. *Suppose that $G = g(X)$ and the assumptions of Theorem 1.2.13 hold. Let $\mathcal{D}_{\mathcal{M}_{0,UI}}$ be the set of admissible controls. Then the dual problem, i.e. the right hand side of* (1.25)*, is a stochastic control problem for a controlled Markov process (X, Y^f, Z^f), $f \in \mathcal{D}_{\mathcal{M}_{0,UI}}$ (defined by* (1.26) *and* (1.27)*), with a value function $\hat{V}$ given by* (1.28)

Proof. For any $f \in \mathcal{D}_{\mathcal{M}^x_{0,UI}}$, $x \in E$ and $y, z \in \mathbb{R}$, define processes Y^f and Z^f via

$$Y^f_t := y + \int_0^t \mathcal{L}f(X_s)ds, \quad 0 \leq t \leq T, \tag{1.26}$$

$$Z^f_{s,t} := \sup_{s \leq r \leq t} \Big(f(x) + g(X_r) - f(X_r) + Y^f_r \Big), \quad 0 \leq s \leq t \leq T, \tag{1.27}$$

and to allow arbitrary starting positions, set $Z^f_t = Z^f_{0,t} \vee z$, for $z \geq g(x) + y$. Note that, for any $f \in \mathbb{D}(\mathcal{L})$, Y^f is an additive functional of X. Lemma 1.2.5 implies that if $f \in \mathcal{D}_{\mathcal{M}_{0,UI}}$ then (X, Y^f, Z^f) is a Markov process.

Define $\hat{V} : E \times \mathbb{R}^2 \mapsto \mathbb{R}$ by

$$\hat{V}(x, y, z) = \inf_{f \in \mathcal{D}_{\mathcal{M}^x_{0,UI}}} \mathbb{E}_{x,y,z}[Z^f_T], \quad (x, y, z) \in E \times \mathbb{R} \times \mathbb{R}. \tag{1.28}$$

It is clear that this is a stochastic control problem for the controlled Markov process (X, Y^f, Z^f), where the admissible controls are functions in $\mathcal{D}_{\mathcal{M}_{0,UI}}$. Moreover, since $V \in \mathcal{D}_{\mathcal{M}_{0,UI}}$, by virtue of Theorem 1.2.15, and adjusting initial conditions as necessary, we have

$$V(x) = \hat{V}(x, 0, g(x)) = \mathbb{E}_{x,0,g(x)}[Z^V_T], \quad x \in E.$$

□

1.2.4 Application: *smooth pasting* condition

We will now discuss the implications of Theorem 1.2.13 for the smoothness of the value function $V(\cdot)$ of the optimal stopping problem given in (1.20).

Remark 1.2.17. *While in Theorem 1.2.18 (resp. Theorem 1.2.23) we essentially recover (a small improvement of) Theorem 2.3 in Peskir [86] (resp. Theorem 2.3 in Samee [98]), the novelty is that we prove the results by means of stochastic calculus, as opposed to the analytic approach in [86] (resp. [98]).*

In addition to Assumption 1.2.8 and Assumption 1.2.9, we now assume that X is a one-dimensional diffusion in the Itô-McKean [63] sense, so that X is a strong Markov process with continuous sample paths. We also assume that the state space $E \subset \mathbb{R}$ is an interval with endpoints $-\infty \leq a \leq b \leq +\infty$. Note that the diffusion assumption implies Assumption 1.2.7. Finally, we assume that X is *regular*: for any $x, y \in \text{int}(E)$, $\mathbb{P}_x[\tau_y < \infty] > 0$, where $\tau_y = \min\{t \geq 0 : X_t = y\}$. Let $\alpha \geq 0$ be fixed; α corresponds to a killing rate of the sample paths of X.

The case without killing: $\alpha = 0$ Let $s(\cdot)$ denote a scale function of X, i.e. a continuous, strictly increasing function on E such that for l, r, $x \in E$, with $a \leq l < x < r \leq b$, we have

$$\mathbb{P}_x(\tau_r < \tau_l) = \frac{s(x) - s(l)}{s(r) - s(l)}, \tag{1.29}$$

see Proposition 3.2 in Revuz and Yor [95] (p.301) for the proof of existence and properties of such a function.

From (1.29), using regularity of X and that $V(X)$ is a supermartingale of class (D) we have that $V(\cdot)$ is s-concave:

$$V(x) \geq V(l)\frac{s(r) - s(x)}{s(r) - s(l)} + V(r)\frac{s(x) - s(l)}{s(r) - s(l)}, \quad x \in [l, r]. \tag{1.30}$$

Theorem 1.2.18. *Suppose the assumptions of Theorem 1.2.13 are satisfied, so that $V \in \mathbb{D}(\mathcal{L})$. Further assume that X is a regular, strong Markov process with continuous sample paths. Let $Y = s(X)$, where $s(\cdot)$ is a scale function of X.*

1. *Assume that for each $y \in [s(a), s(b)]$, the local time of Y at y, L^y, is singular with respect to Lebesgue measure. Then, if $s \in C^1$, $V(\cdot)$, given by (1.20), belongs to C^1.*

2. *Assume that $([Y, Y]_t)_{t \geq 0}$ is, as a measure, absolutely continuous with respect*

to Lebesgue measure. If $s'(\cdot)$ is absolutely continuous, then $V \in C^1$ and $V'(\cdot)$ is also absolutely continuous.

Remark 1.2.19. *If $\mathcal{G}$ is the filtration of a Brownian motion, B, then $Y = s(X)$ is a stochastic integral with respect to B (a consequence of martingale representation):*

$$Y_t = Y_0 + \int_0^t \sigma_s dB_s. \tag{1.31}$$

Moreover, Proposition 3.56 in Çinlar et al. [23] ensures that $\sigma_t = \sigma(Y_t)$ for a suitably measurable function σ and

$$[Y, Y]_t = \int_0^t \sigma^2(Y_s) ds.$$

In this case, both, the singularity of the local time of Y and absolute continuity of $[Y, Y]$ (with respect to Lebesgue measure), are inherited from those of Brownian motion. On the other hand, if X is a regular diffusion (not necessarily a solution to an SDE driven by a Brownian motion), absolute continuity of $[Y, Y]$ still holds, if the speed measure of X is absolutely continuous (with respect to Lebesgue measure).

Proof. Note that $Y = s(X)$ is a Markov process, and let $\mathcal{K}$ denote its martingale generator. Moreover, $V(x) = W(s(x))$ (see Lemma 1.2.21 and the following remark), where, on the interval $[s(a), s(b)]$, $W(\cdot)$ is the smallest nonnegative concave majorant of the function $\hat{g}(y) = g \circ s^{-1}(y)$. Then, since $V \in \mathbb{D}(\mathcal{L})$,

$$V(X_t) = V(x) + M_t^V + \int_0^t \mathcal{L}V(X_u) du, \quad 0 \leq t \leq T,$$

and thus

$$W(Y_t) = W(y) + M_t^V + \int_0^t (\mathcal{L}V) \circ s^{-1}(Y_u) du, \quad 0 \leq t \leq T.$$

Therefore, $W \in \mathbb{D}(\mathcal{K})$, since

$$W(Y_t) = W(y) + M_t^V + \int_0^t \mathcal{K}W(Y_u) du, \tag{1.32}$$

for $y \in [s(a), s(b)]$, $0 \leq t \leq T$, with $\mathcal{K}W = \mathcal{L}V \circ s^{-1} \leq 0$.

On the other hand, using the generalised Itô formula for concave/convex functions (see, for example, Theorem 1.5 in Revuz and Yor [95] (p.223)) we have

$$W(Y_t) = W(y) + \int_0^t W'_+(Y_u)dY_u - \int_{s(a)}^{s(b)} L_t^z \nu(dz),$$

for $y \in [s(a), s(b)]$, $0 \le t \le T$, where L_t^z is the local time of Y_t at z, and ν is a non-negative σ-finite measure corresponding to the second derivative of $-W$ in the sense of distributions. Then, by the uniqueness of the decomposition of a special semimartingale, we have that, for $t \in [0, T]$,

$$-\int_0^t \mathcal{K}W(Y_u)du = \int_{s(a)}^{s(b)} L_t^z \nu(dz) \quad \text{a.s.} \tag{1.33}$$

We prove the first claim by contradiction. Suppose that $\nu(\{z_0\}) > 0$ for some $z_0 \in (s(a), s(b))$. Then, using (1.33) we have that

$$-\int_0^t \mathcal{K}W(Y_u)du = L_t^{z_0}\nu(\{z_0\}) + \int_{s(a)}^{s(b)} 1_{\{z \neq z_0\}} L_t^z \nu(dz) \quad \text{a.s.} \tag{1.34}$$

Since $L_t^{z_0}$ is positive with positive probability and, by assumption, L^y, $y \in [s(a), s(b)]$, is singular with respect to Lebesgue measure, the process on the right hand side of (1.34) is not absolutely continuous with respect to Lebesgue measure, which contradicts absolute continuity of the left hand side. Therefore, $\nu(\{z_0\}) = 0$, and since z_0 was arbitrary, we have that ν does not charge points. It follows that $W \in C^1$. Since $s \in C^1$ by assumption, we conclude that $V \in C^1$.

We now prove the second claim. By assumption, $[Y, Y]$ is absolutely continuous with respect to Lebesgue measure (on the time axis). Invoking Proposition 3.56 in Çinlar et al. [23] again, we have that

$$[Y, Y]_t = \int_0^t \sigma^2(Y_u)du$$

(as in Remark 1.2.19). A time-change argument allows us to conclude that Y is a time-change of a BM and that we may neglect the set $\{t : \sigma^2(Y_t) = 0\}$ in the representation (1.32). Thus

$$W(Y_t) = W(Y_0) + \int_0^t 1_{N^c}(Y_u)dM_u^V + \int_0^t 1_{N^c}(Y_u)\mathcal{K}W(Y_u)du$$

where N is the zero set of σ. Then, using the occupation time formula (see, for example, Corollary 1.6 in Revuz and Yor [95] (p.224)) we have that

$$-\int_0^t \mathcal{K}W(Y_u)du = \int_0^t f(Y_u)d[Y,Y]_u = \int_{s(b)}^{s(b)} f(z)L_t^z dz \quad \text{a.s.},$$

where $f : [s(a), s(b)] \mapsto \mathbb{R}$ is given by $f : y \mapsto -\frac{\mathcal{K}W}{\sigma^2} 1_{N^c}(y)$.

Now observe that, for $0 \le r \le t \le T$,

$$\eta([r,t]) := \int_{s(a)}^{s(b)} f(z)\Big(L_t^z - L_r^z\Big)dz \quad \text{and} \quad \pi([r,t]) := \int_{s(a)}^{s(b)} \Big(L_t^z - L_r^z\Big)\nu(dz)$$

define measures on the time axis, which, by virtue of (1.33), are equal (and thus both are absolutely continuous with respect to Lebesgue measure). Now define

$$T^{\underline{l},\bar{l}} := \{t : Y_t \in [\underline{l}, \bar{l}]\}, \quad s(a) \le \underline{l} \le \bar{l} \le s(b).$$

Then the restrictions of η and π to $T^{\underline{l},\bar{l}}$, $\eta|_{T^{\underline{l},\bar{l}}}$ and $\pi|_{T^{\underline{l},\bar{l}}}$, are also equal. Moreover, since Y is a local martingale, it is also a semimartingale. Therefore, for every $0 \le t \le T$, L_t^z is carried by the set $\{t : Y_t = z\}$ (see Theorem 69 in Protter [92] (p.217)). Hence, for each $t \in [0,T]$,

$$\eta|_{T^{\underline{l},\bar{l}}}([0,t]) = \int_{\underline{l}}^{\bar{l}} L_t^z f(z)dz = \int_{\underline{l}}^{\bar{l}} L_t^z \nu(dz) = \pi|_{T^{\underline{l},\bar{l}}}([0,t]), \tag{1.35}$$

and, since $\underline{l}$ and $\bar{l}$ are arbitrary, the left and right hand sides of (1.35) define measures on $[s(a), s(b)] \subseteq \mathbb{R}$, which are equal. It follows that ν is absolutely continuous with respect to Lebesgue measure on $[s(a), s(b)]$ and $f(z)dz = \nu(dz)$. This proves that $W \in C^1$ and $W'(\cdot)$ is absolutely continuous on $[s(a), s(b)]$ with Radon-Nikodym derivative f. Since the product and composition of absolutely continuous functions are absolutely continuous, we conclude that $V'(\cdot)$ is absolutely continuous (since $s'(\cdot)$ is, by assumption). □

Remark 1.2.20. *We note that for a smooth fit principle to hold, it is not necessary that $s \in C^1$. Given that all the other conditions of Theorem 1.2.18 hold, it is sufficient that $s(\cdot)$ is differentiable at the boundary of the continuation region. On the other hand, if $g \in \mathbb{D}(\mathcal{L})$, $V \in C^1$, even if $g \notin C^1$.*

Moreover, since $V = g$ on the stopping region, Theorem 1.2.18 tells us that $g \in C^1$ on the interior of the stopping region. However, the question whether this stems already from the assumption that $g \in \mathbb{D}(\mathcal{L})$ is more subtle. For example, if $g \in \mathbb{D}(\mathcal{L})$ and g is a difference of two convex functions, then by the generalised Itô *formula and the local time argument (similarly to the proof of Theorem 1.2.18) we could conclude that $g \in C^1$ on the whole state space E.*

Case with killing: $\alpha > 0$ We now generalise the results of the Theorem 1.2.18 in the presence of a non-trivial killing rate. Consider the following optimal stopping problem

$$V(x) = \sup_{\tau \in \mathcal{T}_{0,T}} \mathbb{E}_x[e^{-\alpha\tau} g(X_\tau)], \quad x \in E. \tag{1.36}$$

Note that, since $\alpha > 0$, using the regularity of X together with the supermartingale property of $V(X)$ we have that

$$V(x) \geq V(l)\mathbb{E}_x[e^{-\alpha\tau_l} 1_{\tau_l < \tau_r}] + V(r)\mathbb{E}_x[e^{-\alpha\tau_r} 1_{\tau_r < \tau_l}], \quad x \in [l, r] \subseteq E. \tag{1.37}$$

Define increasing and decreasing functions $\psi, \phi : E \mapsto \mathbb{R}$, respectively, by

$$\psi(x) = \begin{cases} \mathbb{E}_x[e^{-\alpha\tau_c}], & \text{if } x \leq c \\ 1/\mathbb{E}_c[e^{-\alpha\tau_x}], & \text{if } x > c \end{cases} \qquad \phi(x) = \begin{cases} 1/\mathbb{E}_c[e^{-\alpha\tau_x}], & \text{if } x \leq c \\ \mathbb{E}_x[e^{-\alpha\tau_c}], & \text{if } x > c \end{cases} \tag{1.38}$$

where $c \in E$ is arbitrary. Then, $(\Psi_t)_{0 \leq t \leq T}$ and $(\Phi_t)_{0 \leq t \leq T}$, given by

$$\Psi_t = e^{-\alpha t}\psi(X_t), \quad \Phi_t = e^{-\alpha t}\phi(X_t), \quad 0 \leq t \leq T,$$

respectively, are local martingales (and also supermartingales, since ψ, ϕ are non-negative); see Dynkin [37] and Itô and McKean [63].

Let $p_1, p_2 : [l, r] \mapsto [0, 1]$ (where $[l, r] \subseteq E$) be given by

$$p_1(x) = \mathbb{E}_x[e^{-\alpha\tau_l} 1_{\tau_l < \tau_r}], \quad p_2(x) = \mathbb{E}_x[e^{-\alpha\tau_r} 1_{\tau_r < \tau_l}].$$

Continuity of paths of X implies that $p_i(\cdot), i = 1, 2$, are both continuous (the proof of continuity of the scale function in (1.29) can be adapted for a killed process). In terms of the functions $\psi(\cdot)$, $\phi(\cdot)$ of (1.38), using appropriate boundary conditions, one calculates

$$p_1(x) = \frac{\psi(x)\phi(r) - \psi(r)\phi(x)}{\psi(l)\phi(r) - \psi(r)\phi(l)}, \quad p_2(x) = \frac{\psi(l)\phi(x) - \psi(x)\phi(l)}{\psi(l)\phi(r) - \psi(r)\phi(l)}, \quad x \in [l, r]. \tag{1.39}$$

Let $\tilde{s} : E \mapsto \mathbb{R}_+$ be the continuous increasing function defined by $\tilde{s}(x) = \psi(x)/\phi(x)$. Substituting (1.39) into (1.37) and then dividing both sides by $\phi(x)$ we get

$$\frac{V(x)}{\phi(x)} \geq \frac{V(l)}{\phi(l)} \cdot \frac{\tilde{s}(r) - \tilde{s}(x)}{\tilde{s}(r) - \tilde{s}(l)} + \frac{V(r)}{\phi(r)} \cdot \frac{\tilde{s}(x) - \tilde{s}(l)}{\tilde{s}(r) - \tilde{s}(l)}, \quad x \in [l, r] \subseteq E,$$

so that $V(\cdot)/\phi(\cdot)$ is $\tilde{s}$-concave.

Recall that (1.37) essentially follows from $V(\cdot)$ being α-superharmonic, so that it satisfies $\mathbb{E}_x[e^{-\alpha\tau}V(X_\tau)] \leq V(x)$ for $x \in E$ and any stopping time τ. Since Φ and Ψ are local martingales, it follows that the converse is also true, i.e. given a measurable function $f : E \mapsto \mathbb{R}$, $f(\cdot)/\phi(\cdot)$ is $\tilde{s}$-concave if and only if $f(\cdot)$ is α-superharmonic (Proposition 4.1 in Dayanik and Karatzas [30]). This shows that a value function $V(\cdot)$ is the minimal majorant of $g(\cdot)$ such that $V(\cdot)/\phi(\cdot)$ is $\tilde{s}$-concave.

Lemma 1.2.21. *Suppose $[l, r] \subseteq E$ and let $W(\cdot)$ be the smallest nonnegative concave majorant of $\tilde{g} := (g/\phi) \circ \tilde{s}^{-1}$ on $[\tilde{s}(l), \tilde{s}(r)]$, where $\tilde{s}^{-1}$ is the inverse of $\tilde{s}$. Then $V(x) = \phi(x)W(\tilde{s}(x))$ on $[l, r]$.*

Proof. Define $\hat{V}(x) = \phi(x)W(\tilde{s}(x))$ on $[l, r]$. Then, trivially, $\hat{V}(\cdot)$ majorizes $g(\cdot)$ and $\hat{V}(\cdot)/\phi(\cdot)$ is $\tilde{s}$-concave. Therefore $V(x) \leq \hat{V}(x)$ on $[l, r]$.

On the other hand, let $\hat{W}(y) = (V/\phi)(\tilde{s}^{-1}(y))$ on $[\tilde{s}(l), \tilde{s}(r)]$. Since $V(x) \geq g(x)$ and $(V/\phi)(\cdot)$ is $\tilde{s}$-concave on $[l, r]$, $\hat{W}(\cdot)$ is concave and majorizes $(g/\phi) \circ \tilde{s}^{-1}(\cdot)$ on $[\tilde{s}(l), \tilde{s}(r)]$. Hence, $W(y) \leq \hat{W}(y)$ on $[\tilde{s}(l), \tilde{s}(r)]$.

Finally, $(V/\phi)(x) \leq (\hat{V}/\phi)(x) = W(\tilde{s}(x)) \leq \hat{W}(\tilde{s}(x)) = (V/\phi)(x)$ on $[l, r]$. $\square$

Remark 1.2.22. *When $\alpha = 0$, let $(\psi, \phi) = (s, 1)$. Then Lemma 1.2.21 is just Proposition 4.3. in Dayanik and Karatzas [30].*

With the help of Lemma 1.2.21 and using parallel arguments to those in the proof of Theorem 1.2.18 we can formulate sufficient conditions for V to be in C^1 and have absolutely continuous derivative.

Theorem 1.2.23. *Suppose the assumptions of (1.2.13) are satisfied, so that $V \in \mathbb{D}(\mathcal{L})$. Further assume that X is a regular Markov process with continuous sample paths. Let $\psi(\cdot), \phi(\cdot)$ be as in (1.38) and consider the process $Y = \tilde{s}(X)$.*

1. *Assume that, for each $y \in [\tilde{s}(a), \tilde{s}(b)]$, the local time of Y at $y \in [\tilde{s}(a), \tilde{s}(b)]$, $\hat{L}^y$, is singular with respect to Lebesgue measure. Then if $\psi, \phi \in C^1$, $V(\cdot)$, given by (1.36), belongs to C^1.*

2. *Assume that $[Y, Y]$ is, as a measure, absolutely continuous with respect to Lebesgue measure. If $\psi'(\cdot), \phi'(\cdot)$ are both absolutely continuous, then $V'(\cdot)$ is also absolutely continuous.*

Proof. First note that Y is not necessarily a local martingale, while ΦY is. Indeed, $\Phi Y = \Psi$. Hence

$$(N_t)_{0 \leq t \leq T} := \Big(\int_0^t \Phi_t dY_t + [\Phi, Y]_t \Big)_{0 \leq t \leq T}$$

is the difference of two local martingales, and thus is a local martingale itself. Using the generalised Itô formula for concave/convex functions, we have

$$\Phi_t W(Y_t) = \Phi_0 W(y) + \int_0^t W(Y_s) d\Phi_s + \int_0^t W_+^{'}(Y_s) dN_s - \int_{\tilde{s}(l)}^{\tilde{s}(r)} \Phi_t \hat{L}_t^z \nu(dz), \quad (1.40)$$

for $y \in [\tilde{s}(l), \tilde{s}(r)]$, $0 \leq t \leq T$, where $\hat{L}_t^z$ is the local time of Y_t at z, and ν is a non-negative σ-finite measure corresponding to the derivative $W^{''}$ in the sense of distributions.

On the other hand, if $g \in \mathbb{D}(\mathcal{L})$, then $V \in \mathbb{D}(\mathcal{L})$. Therefore,

$$e^{-\alpha t} V(X_t) = V(x) + \int_0^t e^{-\alpha s} dM_s^V + \int_0^t e^{-\alpha s} \{\mathcal{L} - \alpha\} V(X_s) ds, \qquad 0 \leq t \leq T. \quad (1.41)$$

Then, similarly to before, from the uniqueness of the decomposition of the Snell envelope, we have that the martingale and FV terms in (1.40) and (1.41) coincide. Hence, for $t \in [0, T]$,

$$\int_{\tilde{s}(l)}^{\tilde{s}(r)} e^{-\alpha t} \phi(X_t) \hat{L}_t^z \nu(dz) = - \int_0^t e^{-\alpha s} \{\mathcal{L} - \alpha\} V(X_s) ds \quad \text{a.s.}$$

Using the same arguments as in the proof of Theorem 1.2.18 we can show that both statements of this theorem hold. The details are left to the reader. □

The following example shows that the smooth pasting may fail even if $g \in C^\infty$.

Example 1.2.24. *Suppose $X_t = (Z_t, L_t^0, t)$, $t \geq 0$, where Z is a one-dimensional Itô diffusion and L^0 is its local time at 0. Fix $\alpha > 0$ and let*

$$g : (z, l, t) \mapsto -e^{-\alpha t} l$$

be the payoff function and G, given by $G_t = g(Z_t, L_t^0, t)$, a corresponding gains process. Let $\psi : z \mapsto \mathbb{E}_z[e^{-\alpha \tau_0}]$, where τ_0 is the first hitting time of 0 by Z.

Now, define

$$v : (z, l, t) \mapsto -\psi(z) e^{-\alpha t} l,$$

and note that V, given by $V_t = v(Z_t, L_t^0, t)$, is the conditional expected payoff obtained by stopping at the first hit of 0 by Z after time t. It follows that $G \leq V \leq S$, where S is the Snell envelope of G. Conversely, since ψ is bounded and L^0 is continuous and only increases when Z is at 0, $(V_{t \wedge \tau_0})_{t \geq 0}$ is a uniformly integrable martingale. Then, since ψ is positive and L^0 is non-decreasing, it follows that V is *a supermartingale. Consequently $V = S$ and, in particular, v is the value function of the optimal stopping problem (with payoff function g).*

Notice that, in general, ψ (and hence v) is not in C^1. For example, taking Z to be a Brownian motion and $\alpha = 1/2$ gives

$$\psi : z \mapsto e^{-|z|}.$$

1.3 Auxiliary results and proofs

In this section we retain the notation of Sections 1.1.1 and 1.2.1.

Lemma 1.3.1. *Let $G \in \bar{\mathbb{G}}$, $\sigma \in \mathcal{T}_{0,T}$ and $\tau \in \mathcal{T}_{\sigma,T}$. Then the family of random variables $\{\mathbb{E}[G_\rho|\mathcal{F}_\sigma] : \rho \in \mathcal{T}_{\tau,T}\}$ is directed upwards, i.e. for any ρ_1, $\rho_2 \in \mathcal{T}_{\tau,T}$, there exists $\rho_3 \in \mathcal{T}_{\tau,T}$, such that*

$$\mathbb{E}[G_{\rho_1}|\mathcal{F}_\sigma] \vee \mathbb{E}[G_{\rho_2}|\mathcal{F}_\sigma] \leq \mathbb{E}[G_{\rho_3}|\mathcal{F}_\sigma] \quad \mathbb{P}\text{-}a.s.$$

Proof. Fix $\tau \in \mathcal{T}_{\sigma,T}$. Suppose $\rho_1, \rho_2 \in \mathcal{T}_{\tau,T}$ and define $A := \{\mathbb{E}[G_{\rho_1}|\mathcal{F}_\sigma] \geq \mathbb{E}[G_{\rho_2}|\mathcal{F}_\sigma]\}$. Set $\rho_3 := \rho_1 1_A + \rho_2 1_{A^c}$. Note that $\rho_3 \in \mathcal{T}_{\tau,T}$. Using $\mathcal{F}_\sigma$-measurability of A, we have

$$\begin{aligned}\mathbb{E}[G_{\rho_3}|\mathcal{F}_\sigma] &= 1_A\mathbb{E}[G_{\rho_1}|\mathcal{F}_\sigma] + 1_{A^c}\mathbb{E}[G_{\rho_2}|\mathcal{F}_\sigma] \\ &= \mathbb{E}[G_{\rho_1}|\mathcal{F}_\sigma] \vee \mathbb{E}[G_{\rho_2}|\mathcal{F}_\sigma] \quad \mathbb{P}\text{-a.s.},\end{aligned}$$

which proves the claim. □

Lemma 1.3.2. *Let $G \in \bar{\mathbb{G}}$ and S be its Snell envelope with decomposition $S = M^* - A$, where $M^* \in \mathcal{M}_{0,UI}$ and A is an adapted, non-decreasing process of integrable variation. For $0 \leq t \leq T$ and $\epsilon > 0$, define*

$$K_t^\epsilon = \inf\{s \geq t : G_s \geq S_s - \epsilon\}. \tag{1.42}$$

Then $A_{K_t^\epsilon} = A_t$ a.s. and the processes $(A_{K_t^\epsilon})$ and A are indistinguishable.

Proof. From the directed upwards property (Lemma 1.3.1) we know that $\mathbb{E}[S_t] = \sup_{\tau\in\mathcal{T}_{t,T}} \mathbb{E}[G_\tau]$. Then for a sequence $(\tau_n)_{n\in\mathbb{N}}$ of stopping times in $\mathcal{T}_{t,T}$, such that $\lim_{n\to\infty} \mathbb{E}[G_{\tau_n}] = \mathbb{E}[S_t]$, we have

$$\mathbb{E}[G_{\tau_n}] \leq \mathbb{E}[S_{\tau_n}] = \mathbb{E}[M^*_{\tau_n} - A_{\tau_n}] = \mathbb{E}[S_t] - \mathbb{E}[A_{\tau_n} - A_t],$$

since M^* is uniformly integrable. Hence, since A is non-decreasing,

$$0 \leq \lim_{n\to\infty} \mathbb{E}[S_{\tau_n} - G_{\tau_n}] = -\lim_{n\to\infty} \mathbb{E}[A_{\tau_n} - A_t] \leq 0,$$

and thus we have equalities throughout. By passing to a sub-sequence we can assume that

$$\lim_{n\to\infty}(S_{\tau_n} - G_{\tau_n}) = 0 = \lim_{n\to\infty}(A_{\tau_n} - A_t) \quad \text{a.s.} \tag{1.43}$$

The first equality in (1.43) implies that $K_t^\epsilon \leq \tau_{n_0}$ a.s., for some large enough $n_0 \in \mathbb{N}$, and thus $A_{K_t^\epsilon} \leq A_{\tau_n}$, for all $n_0 \leq n$. Since A is non-decreasing, we also have that $0 \leq A_{K_t^\epsilon} - A_t \leq A_{\tau_n} - A_t$ a.s., $n_0 \leq n$, and from the second equality in (1.43) we conclude that $A_{K_t^\epsilon} = A_t$ a.s. The indistinguishability follows from the right-continuity of G and S. □

1.3.1 Proofs of results in Section 1.2

Proof of Lemma 1.2.5. The completed filtration generated by a Feller process satisfies the *usual assumptions*, in particular, it is both right-continuous and quasi-left-continuous. The latter means that for any predictable stopping time σ, $\mathcal{F}_{\sigma-} = \mathcal{F}_\sigma$. Moreover, every càdlàg Feller process is left-continuous over stopping times and satisfies the strong Markov property. On the other hand, every Feller process admits a càdlàg modification (these are standard results and can be found, for example, in Revuz and Yor [95] or Rogers and Williams [97]). All that remains is to show that the addition of the functional F leaves (X, F) strong Markov. This is elementary from (1.19). □

1.3.2 Proofs of results in Section 1.1

Proof of Lemma 1.1.17. Let $(\tau_n)_{n\in\mathbb{N}}$ be a nondecreasing sequence of stopping times with $\lim_{n\to\infty}\tau_n = \tau$, for some fixed $\tau \in \mathcal{T}_{0,T}$. Since S is a supermartingale, $\mathbb{E}[S_{\tau_n}] \geq \mathbb{E}[S_\tau]$, for every $n \in \mathbb{N}$. For a fixed $\epsilon > 0$, $K_{\tau_n}^\epsilon$ (defined by (1.42)) is a stopping time, and by Lemma 1.3.2, $A_{K_{\tau_n}^\epsilon} = A_{\tau_n}$ a.s. Therefore, since M^* is uniformly integrable,

$$\mathbb{E}[S_{K_{\tau_n}^\epsilon}] = \mathbb{E}[M^*_{K_{\tau_n}^\epsilon} - A_{K_{\tau_n}^\epsilon}] = \mathbb{E}[M^*_{\tau_n} - A_{\tau_n}] = \mathbb{E}[S_{\tau_n}].$$

Thus, by the definition of $K_{\tau_n}^\epsilon$,

$$\mathbb{E}[G_{K_{\tau_n}^\epsilon}] \geq \mathbb{E}[S_{K_{\tau_n}^\epsilon}] - \epsilon = \mathbb{E}[S_{\tau_n}] - \epsilon.$$

Let $\hat{\tau} := \lim_{n\to\infty} K^{\epsilon}_{\tau_n}$. Note that the sequence $(K^{\epsilon}_{\tau_n})_{n\in\mathbb{N}}$ is non-decreasing and dominated by K^{ϵ}_{τ}. Hence $\tau \leq \hat{\tau} \leq K^{\epsilon}_{\tau}$ a.s. Finally, using the regularity of G we obtain

$$\mathbb{E}[S_{\tau}] \geq \mathbb{E}[S_{\hat{\tau}}] \geq \mathbb{E}[G_{\hat{\tau}}] = \lim_{n\to\infty} \mathbb{E}[G_{K^{\epsilon}_{\tau_n}}] \geq \lim_{n\to\infty} \mathbb{E}[S_{\tau_n}] - \epsilon.$$

Since ϵ is arbitrary, the result follows. □

Proof of Lemma 1.1.20. For $n \geq 1$, define

$$\tau_n := \inf\{t \geq 0 : \int_0^t |dK_s| \geq n\}.$$

Clearly $\tau_n \uparrow \infty$ as $n \to \infty$. Then for each $n \geq 1$

$$\begin{aligned}\mathbb{E}[\int_0^{t\wedge\tau_n} |dK_s|] &\leq \mathbb{E}[\int_0^{\tau_n} |dK_s|] \\ &= \mathbb{E}[\int_0^{\tau_n -} |dK_s|] + |\Delta K_{\tau_n}|] \\ &\leq n + c.\end{aligned}$$

Therefore, since $X \in \mathbb{G}$,

$$||L^{\tau_n}||_{\mathcal{S}^1} \leq ||X^{\tau_n}||_{\mathcal{S}^1} + \mathbb{E}[\int_0^{\tau_n} |dK_s|] < \infty,$$

and thus, $||X^{\tau_n}||_{\mathcal{H}^1} < \infty$, for all $n \geq 1$. □

Chapter 2

Shadow coupling

In this chapter we investigate the so-called shadow embedding between two measures, introduced by Beiglböck and Juillet [10]. The main result, Theorem 2.2.2, provides an explicit construction (via associated potential functions) of the shadow measure. We also discuss the left-curtain martingale coupling, which is a martingale transport between two probability measures that arises, via shadow measure, from a particular parametrisation of the initial law. This martingale transport plan will be extensively used in studying the robust upper bound of the American put option in Chapter 3, while in Chapter 4 we will provide an explicit construction of the generalised left-curtain coupling. For this reason, the current chapter also serves as a prerequisite chapter for Chapters 3 and 4.

2.1 Measures and Convex order

Let $\mathcal{M}$ (respectively $\mathcal{P}$) be the set of measures (respectively probability measures) on $\mathbb{R}$ with finite total mass and finite first moment, i.e. if $\eta \in \mathcal{M}$, then $\eta(\mathbb{R}) < \infty$ and $\int_{\mathbb{R}} |x| \eta(dx) < \infty$. Given a measure $\eta \in \mathcal{M}$ (not necessarily a probability measure), define $\bar{\eta} = \frac{\int_{\mathbb{R}} x\eta(dx)}{\int_{\mathbb{R}} \eta(dx)}$ to be the barycentre of η. Let $\mathcal{I}_\eta$ be the smallest interval containing the support of η, and let $\{\ell_\eta, r_\eta\}$ be the endpoints of $\mathcal{I}_\eta$ (if η has an atom at ℓ_η then ℓ_η is included in $\mathcal{I}$, and similarly for r_η).

For $\eta \in \mathcal{P}$ let F_η be the distribution function of η and let $G_\eta : (0,1) \mapsto \mathbb{R}$ be the quantile function of η, which is taken to be left-continuous unless otherwise stated.

For any real numbers $c < d$ and a measure $\eta \in \mathcal{M}$, let $\eta_{c,d}$ be the measure given by $\eta_{c,d}(A) = \eta(A \cap (c,d))$, $A \in \mathcal{B}(\mathbb{R})$. Let $\tilde{\eta}_{c,d} = \eta - \eta_{c,d}$.

For $\eta \in \mathcal{M}$, define the function $P_\eta : \mathbb{R} \mapsto \mathbb{R}^+$ by

$$P_\eta(k) := \int_{\mathbb{R}} (k-x)^+ \eta(dx), \quad k \in \mathbb{R}.$$

Then P_η is convex and non-decreasing, $\lim_{z\downarrow -\infty} P_\eta(z) = 0$, $\lim_{z\uparrow\infty} P_\eta(z) - \eta(\mathbb{R})(z - \bar{\eta})^+ = 0$ and $\{k : P_\eta(k) > \eta(\mathbb{R})(k-\bar{\eta})^+\} \subseteq \mathcal{I}_\eta$. Conversely, if h is a non-negative, non-decreasing and convex function satisfying $\lim_{z\downarrow -\infty} h(z) = 0$ and $\lim_{z\uparrow\infty} h(z) - k_m(z-k_b)^+ = 0$, for some numbers $k_m, k_b \in \mathbb{R}$, then there exists a unique measure $\eta \in \mathcal{M}$, with a total mass $\eta(\mathbb{R}) = k_m$ and a barycentre $\bar{\eta}(\mathbb{R}) = k_b$, such that $h = P_\eta$ (see, for example, Proposition 2.1 in Hirsch et al. [53]). In particular, η is uniquely identified by the second derivative of h in the sense of distributions. (In financial context, P_η represents the discounted European put-price, expressed as a function of strike, if the discounted underlying has law η at maturity.) Additional properties of P_η can be found in Chacon [20], and Chacon and Walsh [21]. Note that P_η is related to the potential U_η, defined by

$$U_\eta(k) := \int_{\mathbb{R}} |k-x| \eta(dx), \quad k \in \mathbb{R},$$

by $P_\eta(k) = \frac{1}{2}(U_\eta(k) + (k-\bar{\eta})\eta(\mathbb{R}))$.

For $\eta, \chi \in \mathcal{M}$, we write $\eta \leq \chi$ if

$$\int f d\eta \leq \int f d\chi, \quad \text{for all } f : \mathbb{R} \mapsto \mathbb{R}_+.$$

Then $\eta \leq \chi$ if and only if $P_\chi - P_\eta$ is convex, i.e. P_η has a smaller curvature than P_χ.

Two measures $\eta, \chi \in \mathcal{M}$ are in convex order, and we write $\eta \leq_{cx} \chi$, if

$$\int_{\mathbb{R}} f d\eta \leq \int_{\mathbb{R}} f d\chi, \quad \text{for all convex } f : \mathbb{R} \mapsto \mathbb{R}. \tag{2.1}$$

Then, if $\eta \leq_{cx} \chi$, η and χ have the same total mass ($\eta(\mathbb{R}) = \chi(\mathbb{R})$) and the same barycentre ($\bar{\eta} = \bar{\chi}$). Indeed, since we can apply (2.1) to all affine functions, choosing $f(x) = \pm 1$ and $f(x) = \pm x$ gives a desired result. Moreover, necessarily we must have $\ell_\chi \leq \ell_\eta \leq r_\eta \leq r_\chi$. From simple approximation arguments (see Hirsch et al. [53]) we also have that, if η and χ have the same total mass and the same barycentre, then $\eta \leq_{cx} \chi$ if and only if $P_\eta(k) \leq P_\chi(k)$, $k \in \mathbb{R}$.

Example 2.1.1. *If $\mu = \alpha\delta_x$, for $\alpha > 0$ and $x \in \mathbb{R}$, then any $\nu \in \mathcal{M}$, with $\nu(\mathbb{R}) = \alpha$ and $\bar{\nu} = x$, satisfies $\mu \leq_{cx} \nu$.*

If $\mu_i \leq_{cx} \nu_i$ for $i = 1, ..., n$ then $\sum_{i=1}^n \mu_i \leq_{cx} \sum_{i=1}^n \nu_i$.

If $\mu(\mathbb{R}) = \nu(\mathbb{R})$ and $\bar{\mu} = \bar{\nu}$, μ is concentrated on $[a, b] \subset \mathbb{R}$ and ν is concentrated on $\mathbb{R} \setminus (a, b)$, then $\mu \leq_{cx} \nu$.

If $\mu(\mathbb{R}) = \nu(\mathbb{R})$ and $\bar{\mu} = \bar{\nu}$, $\mu - (\mu \wedge \nu)$ is concentrated on $[a, b] \subset \mathbb{R}$ and $\nu - (\mu \wedge \nu)$ is concentrated on $\mathbb{R} \setminus (a, b)$, then $\mu \leq_{cx} \nu$ (see also Dispersion Assumption 3.2.3 in Chapter 3).

For our purposes in Section 2.2 we need a generalisation of the convex order of two measures. We say $\eta, \chi \in \mathcal{M}$ are in an *extended* convex order, and write $\eta \leq_E \chi$, if

$$\int f d\eta \leq \int f d\chi, \quad \text{for all convex } f : \mathbb{R} \mapsto \mathbb{R}_+.$$

The partial order $\leq_E$ generalises $\leq_{cx}$ in a sense that it preserves the old inequalities and gives rise to new ones. If $\eta \leq_{cx} \chi$ then also $\eta \leq_E \chi$ (since non-negative convex functions are convex), while if $\eta \leq \chi$, we also have that $\eta \leq_E \chi$ (since non-negative convex functions are non-negative). Note that, if $\eta \leq_E \chi$, then $\eta(\mathbb{R}) \leq \chi(\mathbb{R})$ (apply the non-negative convex function $\phi(x) = 1$, $x \in \mathbb{R}$, in the definition of $\leq_E$). It is also easy to prove that, if $\eta(\mathbb{R}) = \chi(\mathbb{R})$, then $\eta \leq_E \chi$ is equivalent to $\eta \leq_{cx} \chi$.

Example 2.1.2. *Let $0 \leq \mu' \leq \mu$ and $\mu \leq_{cx} \nu$. Then, for any convex $f : \mathbb{R} \mapsto \mathbb{R}_+$,*

$$\int f d\mu' \leq \int f d\mu \leq \int f d\nu,$$

and thus $\mu' \leq_E \nu$.

Let $(\eta_n)_{n \geq 1}$ be a sequence of probability measures in $\mathcal{P}$. For $\eta \in \mathcal{P}$, we write $\eta_n \xrightarrow{w} \eta$, and say η_n converges weakly to η, if $\lim_{n \to \infty} \int f d\eta_n = \int f d\eta$ for all bounded and continuous functions $f : \mathbb{R} \mapsto \mathbb{R}$ (see Billingsley [14]). If $\eta_n \xrightarrow{w} \eta$, if $\eta_n \leq_{cx} \eta$ and if $(\eta_n)_{n \geq 1}$ is increasing in convex order, i.e. $\eta_n \leq_{cx} \eta_{n+1}$ for each n, then we write $\eta_n \uparrow_{cx} \eta$.

Lemma 2.1.3. *Suppose $\mu \in \mathcal{P}$. Then there exists a sequence $(\mu_n)_{n \geq 1}$ of finitely supported integrable measures in $\mathcal{P}$ such that $\mu_n \uparrow_{cx} \mu$.*

Proof. Recall that for any $\eta \in \mathcal{P}(\mathbb{R})$, U_η is convex, linear on each interval $I \subset \mathbb{R}$ with $\eta(I) = 0$, $U_\eta(x) \geq |\bar{\eta} - x| = U_{\delta_{\bar{\eta}}}(x)$ on $\mathbb{R}$, and $\lim_{|x| \to \infty} U_\eta(x) - |\bar{\eta} - x| = 0$. Moreover, $\eta_n \uparrow_{cx} \eta$ if and only if $U_{\eta_n} \uparrow U_\eta$ pointwise, see Chacon [20].

Let $\mathcal{U}_\mu$ be a set of piecewise linear convex functions $\tilde{U} : \mathbb{R} \mapsto \mathbb{R}_+$ such that $U_{\delta_{\bar{\mu}}}(x) \leq \tilde{U}(x) \leq U_\mu(x)$. Then each $\tilde{U} \in \mathcal{U}_\mu$ corresponds to a finitely supported integrable probability measure $\tilde{\mu}$ on $\mathbb{R}$ such that $\delta_{\bar{\mu}} \leq_{cx} \tilde{\mu} \leq_{cx} \mu$. Finally, Chacon and Walsh [21] provide a sequence of functions $(\tilde{U}_n)_{n\geq 1}$ in $\mathcal{U}_\mu$, such that $\tilde{U}_n \uparrow U_\mu$ pointwise, proving our claim. $\square$

Let $\bar{\Pi}(\eta,\chi)$ be the set of probability measures on $\mathbb{R}^2$ with the first marginal η and second marginal χ. Let $\Pi(\eta,\chi)$ be the set of martingale couplings of η and χ. Then

$$\Pi(\eta,\chi) = \left\{\pi \in \bar{\Pi}(\eta,\chi) : (2.2) \text{ holds}\right\}$$

where (2.2) is the martingale condition

$$\int_{x\in B}\int_{y\in\mathbb{R}} y\pi(dx,dy) = \int_{x\in B}\int_{y\in\mathbb{R}} x\pi(dx,dy) = \int_B x\eta(dx) \quad \forall \text{ Borel } B \subseteq \mathbb{R}. \quad (2.2)$$

Equivalently, $\Pi(\eta,\chi)$ consists of all transport plans π (i.e. elements of $\bar{\Pi}(\eta,\chi)$) such that the disintegration in probability measures $(\pi_x)_{x\in\mathbb{R}}$ with respect to η satisfies $\int_{\mathbb{R}} y\pi_x(dy) = x$ for η-almost every x. Then, $\pi \in \Pi(\eta,\chi)$ if and only if $\int_{\mathbb{R}^2} h(x)(y-x)\pi(dx,dy) = 0$ for all bounded and measurable $h : \mathbb{R} \mapsto \mathbb{R}$.

If we ignore the martingale requirement (2.2), it is easy to see that the set of probability measures with given marginals is non-empty, i.e. $\bar{\Pi}(\eta,\chi) \neq \emptyset$ (consider the product measure $\eta \otimes \chi$). However, the fundamental question whether, for given η and χ, the set of martingale couplings $\Pi(\eta,\chi)$ is non-empty, is more delicate. For any $\pi \in \Pi(\eta,\chi)$ and convex $f : \mathbb{R} \mapsto \mathbb{R}$, by conditional Jensen's inequality, we have that

$$\int_{\mathbb{R}} f(x)\eta(dx) \leq \int_{\mathbb{R}}\int_{\mathbb{R}} f(y)\pi_x(dy)\eta(dx) = \int_{\mathbb{R}} f(y)\pi(\mathbb{R},dy) = \int_{\mathbb{R}} f(y)\chi(dy),$$

so that $\eta \leq_{cx} \chi$. On the other hand, Strassen [105] showed that a converse is also true (i.e. $\eta \leq_{cx} \chi$ implies that $\Pi(\eta,\chi) \neq \emptyset$), so that $\Pi(\eta,\chi)$ is non-empty if and only if $\eta \leq_{cx} \chi$.

It is worth noting that $\Pi(\eta,\chi)$ is compact (with respect to weak convergence of measures). To see this, observe that, since $\Pi(\eta,\chi) \subseteq \bar{\Pi}(\eta,\chi)$ (and $\bar{\Pi}(\eta,\chi)$ is compact by Prokhorov's [91] theorem), it is enough to show that $\Pi(\eta,\chi)$ is closed. However, by standard approximation arguments, (2.2) is equivalent to $\int_{\mathbb{R}^2} \tilde{h}(x)(y-x)\pi(dx,dy) = 0$ for all continuous and bounded $\tilde{h} : \mathbb{R} \mapsto \mathbb{R}$, while, by Lemma 2.2 in Beiglböck and Juillet [10], $\pi \mapsto \int \tilde{c}d\pi$ (with $\tilde{c} : (x,y) \mapsto \tilde{h}(x)(y-x)$) is continuous (with respect to weak convergence of measures).

Remark 2.1.4. *Suppose we are given $\eta \leq_{cx} \chi$ and want to show that $\Pi(\eta,\chi) \neq \emptyset$. The idea is to first show that the assertion holds when η is finitely supported (see also Section 4.1.2 in Chapter 4). Then, by Lemma 2.1.3, there exists a sequence* *of measures $\{\eta_n\}_{n\geq 1}$ (where each η_n consists of n atoms) such that $\eta_n \leq_{cx} \eta$ and $\eta_n \xrightarrow{w} \eta$. For each n, consider $\pi_n \in \Pi(\eta_n, \chi)$. Relying on the compactness of $\tilde{\Pi} := \Pi(\eta, \chi) \cup \big(\cup_{n=1}^{\infty} \Pi(\eta_n, \chi)\big)$ we have that $\{\pi_n\}_{n\geq 1}$ has an accumulation point $\pi_\infty \in \tilde{\Pi}$, which thus has a desired (marginal and martingale) properties. (See Section 2 in Beiglböck and Juillet [10] for more details.)*

Remark 2.1.5. *The main problem in the theory of the classical optimal transport is, for a given cost function $c : \mathbb{R}^2 \mapsto \mathbb{R}$, to find a probability measure π on $\mathbb{R}^2$ (with given marginals η and χ) that minimises the total expected cost $\int_{\mathbb{R}^2} c(x, y)\pi(dx, dy)$. Arguably the most important cost function is $c(x, y) = (x-y)^2$, for which Hoeffding-Fréchet coupling can be shown to be optimal.*

However, if we are interested in minimising $\pi \mapsto \int_{\mathbb{R}^2}(x - y)^2\pi(dx, dy)$ over martingale couplings $\Pi(\eta, \chi)$, the problem becomes trivial. Indeed, for any $\pi \in \Pi(\eta, \chi)$,

$$\begin{aligned}
\int_{\mathbb{R}^2} (x-y)^2\pi(dx,dy) &= \int_{\mathbb{R}} x^2\eta(dx) - 2\int_{\mathbb{R}^2} xy\pi(dx,dy) + \int_{\mathbb{R}} y^2\chi(dy) \\
&= \int_{\mathbb{R}} x^2\eta(dx) - 2\int_{\mathbb{R}} x\int_{\mathbb{R}} y\pi_x(dy)\eta(dx) + \int_{\mathbb{R}} y^2\chi(dy) \\
&= \int_{\mathbb{R}} x^2\eta(dx) - 2\int_{\mathbb{R}} x^2\eta(dx) + \int_{\mathbb{R}} y^2\chi(dy) \\
&= \int_{\mathbb{R}} y^2\chi(dy) - \int_{\mathbb{R}} x^2\eta(dx),
\end{aligned}$$

so that the total expected cost does depend on the marginals, but not on a particular choice of a martingale coupling (i.e. any $\pi \in \Pi(\eta, \chi)$ is optimal).

For a pair of measures η, χ on $\mathbb{R}$, let the function $D = D_{\eta,\chi} : \mathbb{R} \mapsto \mathbb{R}^+$ be defined by $D_{\eta,\chi}(k) = P_\chi(k) - P_\eta(k)$. Note that if η, χ have equal mass and equal barycentre then $\eta \leq_{cx} \chi$ is equivalent to $D \geq 0$ on $\mathbb{R}$. Let $\mathcal{I}_D = [\ell_D, r_D]$ be the smallest closed interval containing $\{k : D_{\eta,\chi}(k) > 0\}$. If $\mathcal{I}_D$ is such that $\mathcal{I}_D \subset \mathcal{I}_\chi$ then we must have $\eta = \chi$ on $[\ell_\chi, \ell_D) \cup (r_D, r_\chi]$.

The following result, first observed by Hobson [55] (see also Beiglböck and Juillet [10], Section A.1), tells us that, if $D_{\eta,\chi}(x) = 0$ for some x, then in any martingale coupling of η and χ no mass can cross x.

Lemma 2.1.6. *Suppose η and χ are probability measures with $\eta \leq_{cx} \chi$. Suppose that $D(x) = 0$. If $\pi \in \Pi(\eta, \chi)$, then we have $\pi((-\infty, x), (x, \infty)) + \pi((x, \infty), (-\infty, x)) = 0$.*

Proof. If $(\bar{\eta}, X, Y)$ is a martingale with $X \sim \eta$ and $Y \sim \chi$ then

$$P_\chi(x) = \mathbb{E}[(x-Y)^+] \geq \mathbb{E}[(x-Y)^+; X \leq x] \geq \mathbb{E}[(x-Y); X \leq x] \\ = \mathbb{E}[(x-X); X \leq x] = P_\eta(x). \quad (2.3)$$

If $D(x) = 0$ we must have equality in the first two inequalities. From the fact that the first inequality is an equality we conclude that $(Y < x) \subseteq (X \leq x)$ and $\pi((x,\infty),(-\infty,x)) = 0$. Observe that (2.3) also holds if we replace $(X \leq x)$ with $(X < x)$. Again, if $D(x) = 0$ there is equality in the first two inequalities. This time from the second inequality we conclude $(Y > x) \subseteq (X < x)^c$ and $\pi((-\infty,x),(x,\infty)) = 0$. $\square$

It follows from Lemma 2.1.6 that, if there is a point x in the interior of the interval $\mathcal{I}_\eta$ such that $D_{\eta,\chi}(x) = 0$, then we can separate the problem of constructing martingale couplings of η to χ into a pair of subproblems involving mass to the left and right of x, respectively, always taking care to allocate mass of χ at x appropriately. Indeed, if there are multiple $\{x_j\}_{j\geq 1}$ with $D_{\eta,\chi}(x_j) = 0$, then we can divide the problem into a sequence of 'irreducible' problems[1], each taking place on an interval $\mathcal{I}_i$ such that $D > 0$ on the interior of $\mathcal{I}_i$ and $D = 0$ at the endpoints. All mass starting in a given interval is transported to a point in the same interval.

2.2 The shadow measure

In this section we study the shadow embedding introduced by Beiglböck and Juillet [10], which induces a family of martingale couplings.

Let $\mu', \nu \in \mathcal{M}$ with $\mu' \leq_E \nu$. Define $\mathcal{S}_{\mu',\nu} := \{\theta \in \mathcal{M} : \theta \leq \nu, \mu' \leq_{cx} \theta\}$. Then $\mathcal{S}_{\mu',\nu}$ is non-empty (see Proposition 4.4 in Beiglböck and Juillet [10]) and has a minimal element in convex order:

Lemma 2.2.1 (Beiglböck and Juillet [10], Lemma 4.6 and Theorem 4.8)**.** *Let $\mu', \nu \in \mathcal{M}$ and assume $\mu' \leq_E \nu$. Then there exists a (unique) measure $S^\nu(\mu')$, called the shadow of μ' in ν, such that*

1. *$S^\nu(\mu') \leq \nu$.*
2. *$\mu' \leq_{cx} S^\nu(\mu')$.*
3. *If η is another measure satisfying Item 1 and Item 2, then $S^\nu(\mu') \leq_{cx} \eta$.*

[1] The terminology 'irreducible' is due to Beiglböck and Juillet [10] although the idea of splitting a problem into separate components is also present in the earlier papers of Hobson [55] and Cox [24].

Moreover, let $\mu_1, \mu_2, \nu \in \mathcal{M}$ and suppose $\mu' = \mu_1 + \mu_2 \leq_E \nu$. Then, $\mu_2 \leq_E \nu - S^\nu(\mu_1)$, $\mu_1 \leq_E \nu - S^\nu(\mu_2)$ and the following associativity property holds

$$S^\nu(\mu_1 + \mu_2) = S^\nu(\mu_1) + S^{\nu - S^\nu(\mu_1)}(\mu_2) \tag{2.4}$$

$$= S^\nu(\mu_2) + S^{\nu - S^\nu(\mu_2)}(\mu_1). \tag{2.5}$$

In Theorem 2.2.2 below we will provide an explicit construction of the shadow measure in terms of its potential function. On the other hand, the associativity property (2.4) (or equivalently (2.5)) can be proved by approximation arguments, similarly as in Remark 2.1.4.

Fix $\mu', \nu \in \mathcal{M}$ such that $\mu' \leq_E \nu$. Our goal is to construct the shadow measure $S^\nu(\mu')$. On the other hand, given $P_{S^\nu(\mu')}$, $S^\nu(\mu')$ can be identified as the second derivative of $P_{S^\nu(\mu')}$ in the sense of distributions. Moreover, consider the measure $\hat{S}^\nu(\mu') := \nu - S^\nu(\mu')$. Then $P_{S^\nu(\mu')} = P_\nu - P_{\hat{S}^\nu(\mu')}$ on $\mathbb{R}$, and hence, to find $P_{S^\nu(\mu')}$ it is enough to find $P_{\hat{S}^\nu(\mu')}$.

For a measurable function $h : \mathbb{R} \mapsto \mathbb{R}$, let h^c denote its convex hull, i.e. the greatest convex minorant of h. The main result of this chapter (which relies on the auxiliary Lemma 2.5.1) is the following

Theorem 2.2.2. *Let $\mu', \nu \in \mathcal{M}$ with $\mu' \leq_E \nu$. Then*

$$P_{\hat{S}^\nu(\mu')}(x) = (P_\nu - P_{\mu'})^c(x), \quad x \in \mathbb{R}. \tag{2.6}$$

In particular, $P_{S^\nu(\mu')} = P_\nu - (P_\nu - P_{\mu'})^c$ on $\mathbb{R}$.

Proof of Theorem 2.2.2. We first restate the defining properties of the shadow measure, Items 1 to 3, in Lemma 2.2.1 in terms of the potential function $P_{\hat{S}^\nu(\mu')}$.

From Item 1 we have that $0 \leq \hat{S}^\nu(\mu') \leq \nu$, and therefore

$$P_{\hat{S}^\nu(\mu')} \text{ and } P_{\nu - \hat{S}^\nu(\mu')} = P_\nu - P_{\hat{S}^\nu(\mu')} \text{ are both convex on } \mathbb{R}. \tag{2.7}$$

Item 2 is equivalent to

$$\int f d\hat{S}^\nu(\mu') \leq \int f d(\nu - \mu'), \quad \text{for all convex } f : \mathbb{R} \mapsto \mathbb{R},$$

and therefore

$$P_{\hat{S}^\nu(\mu')}(k) \leq P_\nu(k) - P_{\mu'}(k), \quad k \in \mathbb{R}. \tag{2.8}$$

Finally, if η is another measure satisfying Items 1 and 2 in Lemma 2.2.1, then

Item 3 is equivalent to

$$\int f d(\nu - \eta) \leq \int f d\hat{S}^{\nu}(\mu'), \quad \text{for all convex } f : \mathbb{R} \mapsto \mathbb{R},$$

and thus $P_\nu - P_\eta \leq P_{\hat{S}^\nu(\mu')}$ on $\mathbb{R}$. In particular,

$$P_{\hat{S}^\nu(\mu')} \quad \text{is pointwise maximal with respect to (2.7) and (2.8).} \tag{2.9}$$

Our goal is to show that $P_{\hat{S}^\nu(\mu')} = (P_\nu - P_{\mu'})^c$.

Since $(P_\nu - P_{\mu'})^c$ is the largest convex function dominated by $P_\nu - P_{\mu'}$ (while $P_{\hat{S}^\nu(\mu')}$ is convex and satisfies (2.8)), we have both, $(P_\nu - P_{\mu'})^c$ is convex and $P_{\hat{S}^\nu(\mu')} \leq (P_\nu - P_{\mu'})^c \leq P_\nu - P_{\mu'}$ on $\mathbb{R}$.

Now, note that P_ν and $P_{\mu'}$ are both convex (on $\mathbb{R}$). Therefore, by Lemma 2.5.1, we have that $P_\nu - (P_\nu - P_{\mu'})^c$ is also convex (on $\mathbb{R}$). Hence, $(P_\nu - P_{\mu'})^c$ satisfies (2.7) and (2.8). But, by (2.9), $P_{\hat{S}^\nu(\mu')}$ is maximal with respect to (2.7) and (2.8). Therefore, $(P_\nu - P_{\mu'})^c \leq P_{\hat{S}^\nu(\mu')}$ on $\mathbb{R}$. It follows that $(P_\nu - P_{\mu'})^c = P_{\hat{S}^\nu(\mu')}$. □

2.3 The left-curtain coupling π_{lc}

The left-curtain coupling (or martingale transport) was introduced by Beiglböck and Juillet [10] (and further studied by Henry-Labordère and Touzi [50] and Beiglböck et al. [9]).

The left-curtain martingale coupling, denoted by π_{lc}, is a martingale coupling that arises, from a particular parametrisation of the initial law, via shadow measure. More specifically (see Theorem 4.18 in Beiglböck and Juillet [10]), there exists a (unique) measure $\pi_{lc} \in \Pi(\mu, \nu)$ that transports $\mu|_{(-\infty,x]}$ to the shadow $S^\nu(\mu|_{(-\infty,x]})$, $x \in \mathbb{R}$. In other words, the first and second marginals of $\pi_{lc}|_{(-\infty,x]\times\mathbb{R}}$ are $\mu|_{(-\infty,x]}$ and $S^\nu(\mu|_{(-\infty,x]})$, respectively.

Definition 2.3.1. *A transport plan $\pi \in \Pi(\mu, \nu)$ is said to be left-monotone if there exists $\Gamma \in \mathcal{B}(\mathbb{R}^2)$ with $\pi(\Gamma) = 1$ and such that, if $(x, y^-), (x, y^+), (x', y') \in \Gamma$ we cannot have $x < x'$ and $y^- < y' < y^+$.*

Using the minimality of the shadow measure with respect to convex order, i.e. Item 3 in Lemma 2.2.1, it can be shown that π_{lc} is also a unique left-monotone martingale coupling (see Theorems 4.21 and 5.3 in Beiglböck and Juillet [10]).

The following additional property of π_{lc} is of particular importance in Chapters 3 and 4. When the initial law μ is continuous, the left-curtain coupling has a rather simple representation. In particular, the element $\pi^x_{lc}(\cdot)$ in the disintegration $\pi_{lc}(dx, dy) = \mu(dx)\pi^x_{lc}(dy)$ is a measure supported on a set of at most two points. For real numbers c, d with $c \leq x \leq d$ define the probability measure $\chi_{c,x,d}$ by $\chi_{c,x,d} = \frac{d-x}{d-c}\delta_c + \frac{x-c}{d-c}\delta_d$ with $\chi_{c,x,d} = \delta_x$ if $(d-x)(x-c) = 0$. Note that $\chi_{c,x,d}$ has mean x. $\chi_{c,x,d}$ is the law of a Brownian motion started at x evaluated on the first exit from (c, d).

Lemma 2.3.2 (Beiglböck and Juillet[10], Corollary 1.6)**.** *Let μ, ν be probability measures in convex order and assume that μ is continuous. Then there exists a pair of measurable functions $T_d : \mathbb{R} \mapsto \mathbb{R}$ and $T_u : \mathbb{R} \mapsto \mathbb{R}$ such that $T_d(x) \leq x \leq T_u(x)$, such that for all $x < x'$ we have $T_u(x) \leq T_u(x')$ and $T_d(x') \notin (T_d(x), T_u(x))$, and such that, if we define $\hat{\pi}(dx, dy) = \mu(dx)\chi_{T_d(x),x,T_u(x)}(dy)$, then $\hat{\pi} \in \Pi(\mu, \nu)$ and $\hat{\pi} = \pi_{lc}$.*

Note that there is no claim of uniqueness of the functions T_d, T_u in Lemma 2.3.2. Indeed, T_d, T_u are unique only μ-a.s. (see Beiglböck et al. [9], Proposition 3.4). For example, the definitions of T_d and T_u are immaterial outside $[\ell_\mu, r_\mu]$. Further, if T_u has a (necessarily upward) jump at x' then it does not matter what value we take for $T_u(x')$ provided $T_u(x') \in [T_u(x'-), T_u(x'+)]$. (Since we are assuming μ is continuous, the probability that we choose an x-coordinate value of x' is zero.) More importantly, if (T_d, T_u) satisfy the properties of Lemma 2.3.2 and if $T_u(x) = x$ on an interval $[\underline{x}, \overline{x})$ then we can modify the definition of T_d on $[\underline{x}, \overline{x})$ to either $T_d(x) = x$ or $T_d(x) = T_d(\underline{x}-)$ and still satisfy the relevant monotonicity properties. Henry-Labordère and Touzi [50] resolve this indeterminacy by setting $T_d(x) = x$ on the set $T_u(x) = x$ and also taking T_u and T_d to be right-continuous.

In a sequel we follow Henry-Labordère and Touzi [50] by taking $T_d(x) = x$ on the set $T_u(x) = x$ but we do not make right-continuity assumptions on T_d and T_u. Also we write (f, g) in place of (T_d, T_u). Our functions f and g will eventually be defined on $\mathbb{R}$, see Section 3.2.3, but for now we define them just on $[\ell_\mu, r_\mu]$.

Lemma 2.3.3. *Let (T_d, T_u) be a pair of functions satisfying the monotonicity properties listed in Lemma 2.3.2. Suppose they lead to a solution $\pi_{lc} \in \Pi(\mu, \nu)$.*
On $[\ell_\mu, r_\mu]$ set $g(x) = T_u(x)$, on $g(x) > x$ set $f(x) = T_d(x)$ and on $g(x) = x$ set $f(x) = x$. Then (f, g) are such that $f(x) \leq x \leq g(x)$ and for all $x' > x$ we have $g(x') \geq g(x)$ and $f(x') \notin (f(x), g(x))$. Moreover, $\mu(dx)\chi_{f(x),x,g(x)}(dy) = \mu(dx)\chi_{T_d(x),x,T_u(x)}(dy)$.

Proof. The property $f(x) \leq x \leq g(x)$ is immediate so we simply need to check that for $x' > x$ we have $g(x') \geq g(x)$ and $f(x') \notin (f(x), g(x))$. Monotonicity of g is inherited from monotonicity of T_u. If $g(x) = x$ then $f(x) = x$ and $f(x') \notin (f(x), g(x)) = \emptyset$. If $g(x) > x$ and $g(x') > x'$ then $f(x') = T_d(x') \notin (T_d(x), T_u(x)) =$

$(f(x), g(x))$. Finally, if $g(x) > x$ and $g(x') = x'$ then $f(x') = x' \notin (f(x), x' = g(x')) \supseteq (f(x), g(x))$. $\square$

Figure 2.1 gives a stylised representation of f and g in the case where ν has no atoms. (Atoms of ν lead to horizontal sections of f and g, see Section 3.2.6.) In the figure, the set $\{g(x) > x\}$ is a finite union of intervals whereas in general it may be a countable union of intervals. Similarly, in the figure f has finitely many downward jumps, whereas in general it may have countably many jumps. Nonetheless Figure 2.1 captures the essential behaviour of f and g.

Remark 2.3.4. *The left-curtain martingale coupling can be identified with Figure 2.1 in the following way: choose an x-coordinate according to μ; then if $g(x) = x$ set $Y = x = X$ so the pair (X, Y) lies on the diagonal; otherwise if $g(x) > x$ then $f(x) < x$ and we set the y-coordinate to be $g(x)$ with probability $\frac{x-f(x)}{g(x)-f(x)}$ and $f(x)$ with probability $\frac{g(x)-x}{g(x)-f(x)}$. Then the coordinates (x, y) represent the realised values of (X, Y).*

For a horizontal level y there are two cases. Either, $g(y) > y$ and then the value of y arises from a choice according to μ of $x = g^{-1}(y)$ for which $g(x)$ is chosen rather than f(x); or $g(y) = y$ and the value y arises either from a choice according to μ of $x = y$, or from a choice according to μ of $f^{-1}(y)$ combined with a choice of y-coordinate of $f(f^{-1}(y)) = y$.

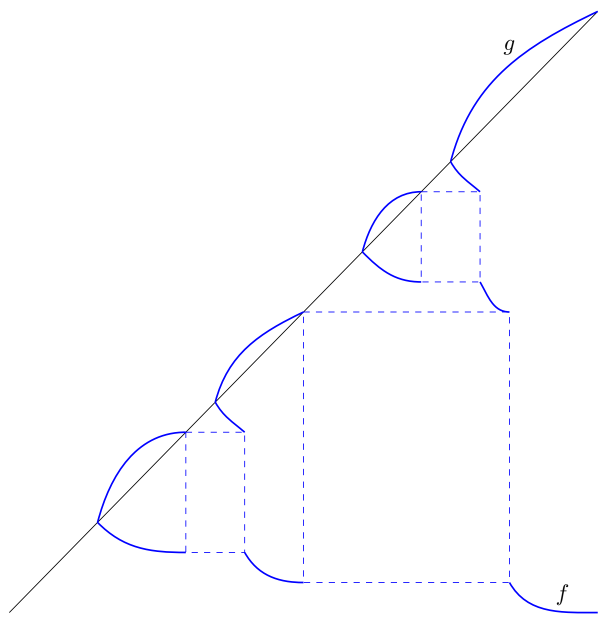

Figure 2.1: Stylised plot of the functions f and g in the general case (with no atoms). Note that on the set $g(x) = x$ we have $f(x) = x$.

Suppose ν is also continuous and fix x. Under the left-curtain martingale coupling mass in the interval $(f(x), x)$ at time 1 is mapped to the interval $(f(x), g(x))$ at time 2. Thus $\{f(x), g(x)\}$ with $f(x) \leq x \leq g(x)$ are solutions to

$$\int_f^x \mu(dz) = \int_f^g \nu(dz), \tag{2.10}$$

$$\int_f^x z\mu(dz) = \int_f^g z\nu(dz). \tag{2.11}$$

Essentially, (2.10) is preservation of mass condition and (2.11) is preservation of mean and the martingale property. If ν has atoms then (2.10) and (2.11) become

$$\int_f^x \mu(dz) = \int_{(f,g)} \nu(dz) + \lambda_f + \lambda_g, \tag{2.12}$$

$$\int_f^x z\mu(dz) = \int_{(f,g)} z\nu(dz) + f\lambda_f + g\lambda_g, \tag{2.13}$$

respectively, where $0 \leq \lambda_f \leq \nu(\{f\})$ and $0 \leq \lambda_g \leq \nu(\{g\})$.

Returning to the case of continuous μ and ν, for fixed x there can be multiple solutions to (2.10) and (2.11). If, however, we consider f and g as functions of x and impose the additional monotonicity properties of Lemma 2.3.2 (for $x < x'$, $g(x) \leq g(x')$ and $f(x') \notin (f(x), g(x))$), then typically, for almost all x there is a unique solution to (2.10) and (2.11). However, there are exceptional x at which f jumps and at which there are multiple solutions, see Section 3.2.3.

Remark 2.3.5. *We note that mass and mean equations* (2.10) *and* (2.11) *are crucial in Chapter 3. If* $g(x) = T_u(x) = x = f(x) > T_d(x)$ *then we typically do not have* $\int_{T_d(x)}^x \mu(dz) = \int_{T_d(x)}^{T_u(x)} \nu(dz)$ *and* $\int_{T_d(x)}^x z\mu(dz) = \int_{T_d(x)}^{T_u(x)} z\nu(dz)$. *However, we trivially have* $\int_{f(x)}^x \mu(dz) = \int_{f(x)}^{g(x)} \nu(dz)$ *and* $\int_{f(x)}^x z\mu(dz) = \int_{f(x)}^{g(x)} z\nu(dz)$. *This explains our choice of* $f(x)$ *when* $g(x) = x$.

Remark 2.3.6. *In a related problem, Hobson and Klimmek [57] show how under natural simplifying assumptions, upper and lower functions can be characterised as solutions of a pair of coupled differential equations. For example, under the Dispersion Assumption 3.2.3,* (f, g) *solve a pair of coupled differential equations on* $[e_-, r_\mu)$ *obtained from differentiating* (2.10) *and* (2.11)*:*

$$\begin{aligned} \frac{df}{dx} &= -\frac{g-x}{g-f}\frac{\rho(x)}{\eta(f)-\rho(f)}, \\ \frac{dg}{dx} &= \frac{x-f}{g-f}\frac{\rho(x)}{\eta(g)}, \end{aligned}$$

with the initial condition $f(e_-) = e_- = g(e_-)$. In addition, see Henry-Labordère and Touzi [50, Equations (3.10) and (3.9)], where the construction of T_d and T_u are exactly based on the resolution of (2.10) *and* (2.11).

There are many pairs (μ, ν) which lead to the same pair of functions (f, g). Moreover, given a pair $\mu \leq_{cx} \nu$ it may be difficult to determine the properties of (f, g) which define the left-curtain coupling, beyond the fact that (f, g) satisfy monotonicity properties as in Lemma 2.3.2. (For example, it may be difficult to ascertain the number of downward jumps of f without calculating f and g everywhere.) However, if we want to construct examples for which (f, g) have additional properties (such as no downward jump) then we can start with an appropriate pair (f, g), take arbitrary (continuous) initial law μ with support on the interval where f is defined, and then define ν via (2.14). This observation underpins our analysis in Sections 3.2.2 and 3.2.3.

Lemma 2.3.7. *Let $\mathcal{I}_1 \subseteq \mathcal{I}_2 \subseteq \mathbb{R}$ be intervals and define*

$$\Xi^{\mathcal{I}_1,\mathcal{I}_2} = \{(f,g) : g : \mathcal{I}_1 \to \mathcal{I}_2, g(x) \geq x, f : \mathcal{I}_1 \to \mathcal{I}_2, f(x) \leq x\},$$

$$\Xi^{\mathcal{I}_1,\mathcal{I}_2}_{Mon} = \{(f,g) \in \Xi^{\mathcal{I}_1,\mathcal{I}_2} : g \text{ increasing, } f(x) = x \text{ on } g(x) = x, \text{ for } x' > x \ f(x') \notin (f(x), g(x))\},$$

$$\Xi = \cup_{\mathcal{I}_1 \subseteq \mathcal{I}_2} \Xi^{\mathcal{I}_1,\mathcal{I}_2}, \qquad \Xi_{Mon} = \cup_{\mathcal{I}_1 \subseteq \mathcal{I}_2}.\Xi^{\mathcal{I}_1,\mathcal{I}_2}_{Mon}.$$

Suppose $(f, g) \in \Xi$ and μ is any continuous and integrable measure with support in $\mathcal{I}_1$ and define π via $\pi(dx, dy) = \mu(dx)\chi_{f(x),x,g(x)}(dy)$ and ν via

$$\nu(dy) = \int_x \mu(dx)\chi_{f(x),x,g(x)}(dy). \tag{2.14}$$

Then, subject to integrability condition $\int \mu(dx)\frac{(g(x)-x)(x-f(x))}{g(x)-f(x)} 1_{\{g(x)>f(x)\}} < \infty$, ν is integrable, $\mu \leq_{cx} \nu$ and $\pi \in \Pi(\mu, \nu)$. In addition, if $(f, g) \in \Xi_{Mon}$ then $\pi = \pi_{lc}$.

Proof. Let $X \sim \mu$ and $Y \sim \nu$, and note that the first and second marginals of π are μ and ν, respectively. Therefore

$$\mathbb{E}^\nu[|Y|] - \mathbb{E}^\mu[|X|] \leq \mathbb{E}^\pi[|Y - X|] \leq \mathbb{E}^\nu[|Y|] + \mathbb{E}^\mu[|X|],$$

and, since μ is integrable by assumption, $\mathbb{E}^\pi[|Y - X|] < \infty$ if and only if ν is integrable. Using the definition of π we have that

$$\begin{aligned}
\mathbb{E}^{\pi}[|Y-X|] &= \int_{\mathbb{R}^2} |y-x|\mu(dx)\chi_{f(x),x,g(x)}(dy) \\
&= \int_{\mathbb{R}} \left[\frac{g(x)-x}{g(x)-f(x)}|f(x)-x| + \frac{x-f(x)}{g(x)-f(x)}|g(x)-x|\right] 1_{\{g(x)>f(x)\}}\mu(dx) \\
&= \int_{\mathbb{R}} 2\frac{(g(x)-x)(x-f(x))}{g(x)-f(x)} 1_{\{g(x)>f(x)\}}\mu(dx) < \infty.
\end{aligned}$$

Moreover, a direct calculation shows that the martingale property is also satisfied:

$$\int_{\mathbb{R}} y\chi_{f(x),x,g(x)}(dy) = \frac{g(x)-x}{g(x)-f(x)}f(x) + \frac{x-f(x)}{g(x)-f(x)}g(x) = x, \quad \text{for } \mu\text{-almost every } x.$$

It follows that $\pi \in \Pi(\mu,\nu)$ and $\mu \leq_{cx} \nu$.

Finally, since $(f,g) \in \Xi_{Mon}$, it is immediate to verify that π is left-monotone in a sense of Definition 2.3.1. Therefore, by uniqueness, $\pi = \pi_{lc}$. □

2.4 Extensions

The fundamental question related to the left-curtain martingale coupling $\pi_{lc} \in \Pi(\mu,\nu)$ is how, given $x \in \mathcal{I}_\mu$, to find two points $f(x), g(x)$ such that (2.10) and (2.11) hold. Put differently, how to explicitly construct upper and lower functions characterising π_{lc}. One such construction when μ is atomless is provided by Henry-Labordère and Touzi [50] via differential equations. For general μ, in Chapter 4 we construct the upper and lower functions that give rise to the extended left-curtain martingale coupling. We conjecture, however, that the relevant functions can be read off directly from Theorem 2.2.2, and (2.6) in particular.

Suppose $\mu \leq_{cx} \nu$ and (for simplicity) that both measures are atomless (the following arguments can be easily extended to general measures). Fix $x \in \mathbb{R}$ and consider the measure μ_x defined by

$$\mu_x(A) := \mu((-\infty,x] \cap A), \quad \text{for all Borel sets } A \subseteq \mathbb{R}.$$

Then $P_{\mu_x}(k) = P_\mu(k)$ for $k \leq x$, while $P_{\mu_x}(k) \leq P_\mu(k)$ for $k > x$. In particular, $P_{\mu_x}(\cdot)$ is linear on $[x,\infty)$ and $P'_{\mu_x}(k) = P'_\mu(x)$ for $k \geq x$. Recall the definition of $D(k) = P_\nu(k) - P_\mu(k)$, $k \in \mathbb{R}$.

Define $\mathcal{E}_u : \mathbb{R} \mapsto \mathbb{R}^+$ by

$$\begin{aligned}\mathcal{E}_x(k) &= P_\nu(k) - P_{\mu_x}(k)\\ &= D(k) + P_\mu(k) - P_{\mu_x}(k), \quad k \in \mathbb{R}.\end{aligned}$$

Note that $\mathcal{E}_x(k) = D(k)$ for $k \leq x$. Moreover, $P_\mu - P_{\mu_x}$ is non-negative and convex on $\mathbb{R}$, and hence $\mathcal{E}_x(k) \geq D(k)$ for $k > x$. Let $\mathcal{E}_x^c$ denote the convex hull of $\mathcal{E}_x$. Then, by Theorem 2.2.2, we have that

$$P_{S^\nu(\mu_x)}(k) = P_\nu(k) - \mathcal{E}_x^c(k), \quad k \in \mathbb{R}.$$

Recall that $S^\nu(\mu_x)$ is the second marginal of $\pi_{lc}|_{(-\infty,x]\times\mathbb{R}}$.

Set

$$\begin{aligned}Z_= &:= \{k \in \mathbb{R} : \mathcal{E}_x^c(k) = \mathcal{E}_x(k)\}.\\ Z_x^< &:= \{k < x : \mathcal{E}_x^c(y) < \mathcal{E}_x(y) \text{ for all } y \in (k,x)\},\\ Z_x^> &:= \{k > x : \mathcal{E}_x^c(y) < \mathcal{E}_x(y) \text{ for all } y \in (x,k)\}\end{aligned}$$

and define $L, U : \mathbb{R} \mapsto \mathbb{R}$ by $L(x) = x = U(x)$ if $x \in Z_=$, and

$$L(x) = Z_x^< \cap Z_= \quad \text{and} \quad U(x) = Z_x^> \cap Z_= \quad \text{if } x \notin Z_=.$$

Conjecture. *Suppose $\mu \leq_{cx} \nu$. Then $\hat{\pi}$, defined by*

$$\hat{\pi}(dx, dy) = \mu(dx)\chi_{L(x),x,U(x)}(dy),$$

is the left-curtain martingale coupling $\pi_{lc} \in \Pi(\mu, \nu)$.

The intuition behind the conjecture above is that, for each $x \in \mathbb{R}$, $L(x)$ and $U(x)$ satisfy the mass and mean conditions, (2.10) and (2.11), respectively. To see this, suppose $x \notin Z_=$, so that $L(x) < x < U(x)$. See Fig. 2.2. Since $\mathcal{E}_x^c(\cdot)$ is linear on $[L(x), U(x)]$, the slopes of $\mathcal{E}_x^c(\cdot)$ at $L(x)$ and $U(x)$ are equal:

$$D'(L(x)) = P_\nu'(U(x)) - P_{\mu_x}'(U(x)), \tag{2.15}$$

which is equivalent to

$$\int_{-\infty}^{L(x)} d(\nu - \mu) = \int_{-\infty}^{U(x)} d\nu - \int_{-\infty}^{x} d\mu. \tag{2.16}$$

Rearranging (2.16) shows that L and U satisfy the mass condition (2.10):

$$\int_{L(x)}^{x} \mu(dz) = \int_{L(x)}^{U(x)} \nu(dz). \tag{2.17}$$

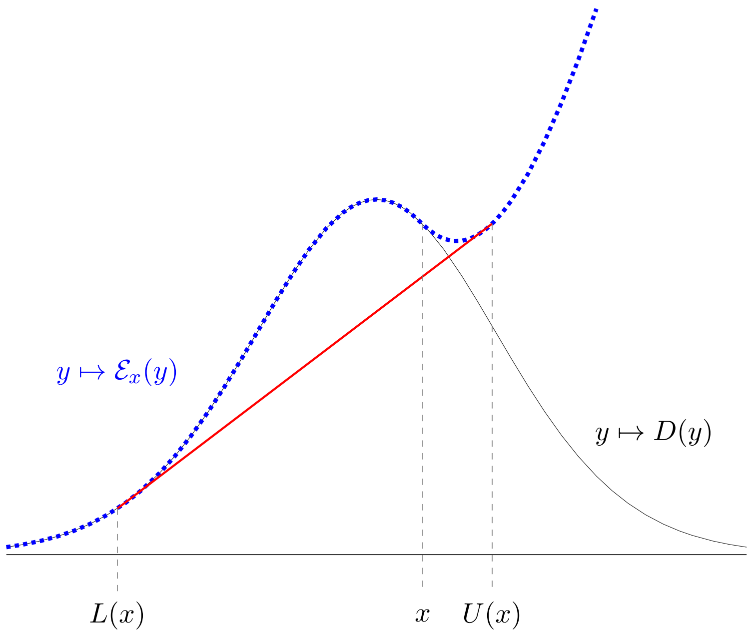

Figure 2.2: Plot of locations of $L(x) < x < U(x)$, when $x \notin Z_=$. Dotted curve corresponds to the graph of $\mathcal{E}_x$. The solid line below $\mathcal{E}_x$ is tangent to $\mathcal{E}_x$ at $L(x)$ and $U(x)$. In particular, it corresponds to the linear section of $\mathcal{E}_x^c$ on $[L(x), U(x)]$.

Moreover, note that $\mathcal{E}_x^c(L(x)) = D(L(x))$ and $\mathcal{E}_x^c(U(x)) = P_\nu(U(x)) - P_{\mu_x}(U(x))$. Then, using linearity of $\mathcal{E}_x^c(\cdot)$ on $[L(x), U(x)]$ again, we have that

$$\begin{aligned} D(L(x)) + D'(L(x))(U(x) - L(x)) &= P_\nu(U(x)) - P_{\mu_x}(U(x)) \\ &= P_\nu(U(x)) - P_\mu(x) - (U(x) - x)P'_\mu(x), \end{aligned}$$

which is equivalent to

$$\int_{-\infty}^{L(x)} (U(x) - z)[\nu(dz) - \mu(dz)] = \int_{-\infty}^{U(x)} (U(x) - z)\nu(dz) - \int_{-\infty}^{x} (U(x) - z)\mu(dz). \tag{2.18}$$

Rearranging (2.18) and using (2.16) we have that L and U also satisfy the mean condition (2.11):

$$\int_{L(x)}^{x} z\mu(dz) = \int_{L(x)}^{U(x)} z\nu(dz). \tag{2.19}$$

On the other hand, when $x \in Z_=$, $L(x) = x = U(x)$. Then the mass and mean conditions, (2.17) and (2.19), respectively, are trivially satisfied.

In order to prove the conjecture, we have to show that the second marginal of $\hat{\pi}$ is ν, i.e. $\int_x \mu(dx)\chi_{L(x),x,U(x)}(dy) = \nu(dy)$ (which we expect to follow from the mass and mean equations, (2.17) and (2.19)), and that $(L,U) \in \Xi_{Mon}$. This is left for future research.

2.5 Auxiliary results

Lemma 2.5.1. *Let $f, g : \mathbb{R} \mapsto \mathbb{R}$ be two convex functions. Set $h = g - f$ on $\mathbb{R}$, and let h^c be the largest convex minorant of h.*

Set $\psi = g - h^c$ on $\mathbb{R}$. Then, ψ is convex.

Proof. First note that, since $h^c(x) \leq h(x)$, $\psi(x) = g(x) - h^c(x) \geq f(x)$, $x \in \mathbb{R}$.

Define

$$\mathcal{A}_= := \{y : h(y) = h^c(y)\} \quad \text{and} \quad \mathcal{A}_> := \{y : h(y) > h^c(y)\}.$$

Then $\psi = f$ on $\mathcal{A}_=$, while $\psi > f$ on $\mathcal{A}_>$.

Recall that $\phi : \mathbb{R} \mapsto \mathbb{R}$ is convex if for all $x, y, z \in \mathbb{R}$, with $x < y < z$,

$$\phi(y) \leq \lambda\phi(x) + (1-\lambda)\phi(z), \quad \text{where} \quad \lambda = \frac{z-y}{z-x}.$$

Suppose $y \in \mathcal{A}_=$. Then

$$\psi(y) = f(y) \leq \lambda f(x) + (1-\lambda)f(z) \leq \lambda\psi(x) + (1-\lambda)\psi(z),$$

where the first inequality follows from convexity of f, while the second one holds since ψ dominates f on $\mathbb{R}$.

In the rest of the proof we take $y \in \mathcal{A}_>$. In this case y belongs to the linear section of h^c. In particular, let $y_-, y_+ \in Z_=$ be such that $y_- < y < y_+$ and $h^c(k) < h(k)$, $k \in (y_-, y_+)$. Then

$$h^c(y) = \gamma h(y_-) + (1-\gamma)h(y_+), \quad \text{where} \quad \gamma = \frac{y_+ - y}{y_+ - y_-}.$$

Suppose $x < y_- < y_+ < z$. Then

$$
\begin{aligned}
\psi(y) &= g(y) - h^c(y) \\
&\le [\gamma g(y_-) + (1-\gamma) g(y_+)] - [\gamma h(y_-) + (1-\gamma) h(y_+)] && (2.20) \\
&= \gamma f(y_-) + (1-\gamma) f(y_+) \\
&\le \gamma \left[\frac{z - y_-}{z - x} f(x) + \frac{y_- - x}{z - x} f(z) \right] + (1-\gamma) \left[\frac{z - y_+}{z - x} f(x) + \frac{y_+ - x}{z - x} f(z) \right] && (2.21) \\
&= \lambda f(x) + \lambda f(z) \\
&\le \lambda \psi(x) + \lambda \psi(z). && (2.22)
\end{aligned}
$$

(2.20) follows from convexity of g on $\mathbb{R}$ and linearity of h^c on $[y_-, y_+]$. (2.21) is a consequence of convexity of f on $\mathbb{R}$. Finally, (2.22) holds since $f \le \psi$ on $\mathbb{R}$.

Suppose $y_- \le x < y < z \le y_+$. Then h^c is also linear on $[x, z]$, and therefore $h^c(y) = \lambda h^c(x) + (1-\lambda) h^c(z)$. Using convexity of g on $\mathbb{R}$ we conclude that

$$
\begin{aligned}
\psi(y) &= g(y) - h^c(y) \\
&\le [\lambda g(x) + (1-\lambda) g(z)] - [\lambda h^c(x) + (1-\lambda) h^c(z)] \\
&= \lambda \psi(x) + (1-\lambda) \psi(z).
\end{aligned}
$$

Suppose $y_- \le x < y < y_+ < z$. (The case $x < y_- < y < z \le y_+$ follows by symmetry.) In this case $h^c(\cdot)$ is also linear on $[x, y_+]$ and therefore

$$
\begin{aligned}
h^c(y) &= \delta h^c(x) + (1-\delta) h^c(y_+) \\
&= \delta h^c(x) + (1-\delta) h(y_+), \quad \text{where} \quad \delta = \frac{y_+ - y}{y_+ - x}.
\end{aligned}
$$

Then

$$
\begin{aligned}
\psi(y) &= g(y) - h^c(y) \\
&\leq [\delta g(x) + (1-\delta) g(y_+)] - [\delta h^c(x) + (1-\delta) h(y_+)] && (2.23) \\
&= \delta[g(x) - h^c(x)] + (1-\delta)[g(y_+) - h(y_+)] \\
&= \delta\psi(x) + (1-\delta) f(y_+) \\
&\leq \delta\psi(x) + (1-\delta)\left[\frac{z-y_+}{z-x} f(x) + \frac{y_+ - x}{z-x} f(z)\right] && (2.24) \\
&\leq \delta\psi(x) + (1-\delta)\left[\frac{z-y_+}{z-x} \psi(x) + \frac{y_+ - x}{z-x} \psi(z)\right] && (2.25) \\
&= \lambda\psi(x) + \lambda\psi(z).
\end{aligned}
$$

(2.23) follows from convexity of g on $\mathbb{R}$ and linearity of h^c on $[x, y_+]$. (2.24) is a consequence of convexity of f on $\mathbb{R}$. Finally, (2.25) holds since $f \leq \psi$ on $\mathbb{R}$. This finishes the proof. $\square$

Chapter 3

Robust bounds for the American put

This chapter is structured as follows. In the next section we formulate precisely our problem of finding the robust, model-independent price of an American put. We also explain how the pricing problem is related to the dual problem of constructing the cheapest superhedge. In Section 3.2 we assume that a starting law is continuous, transform the primal pricing problem into a martingale optimal transport (MOT) problem and show by studying a series of ever more complicated set-ups how to determine the best model and hedge. The constructions in this section make use of results on the left-curtain coupling of Beiglböck and Juillet [10] and Henry-Labordère and Touzi [50] (see Chapter 2).

By weak duality the highest model price is bounded above by the cost of the cheapest superhedge. Hence, if on the one hand we can identify a consistent model and stopping rule, and on the other a superhedge, such that the expected payoff in that model with that stopping rule is equal to the cost of the superhedge, then we must have identified an optimal model and an optimal stopping rule together with an optimal hedging strategy. Moreover, there is no duality gap. This is the strategy of our proofs. One feature of our analysis is that wherever possible we provide pictorial explanations and derivations of our results. In our view this approach helps to bring insights which may be hidden under calculus-based approaches.

3.1 Preliminaries and set-up

For this chapter, Section 2.1 serves as a prerequisite section on probability measures, convex order, martingale couplings and, in particular, the left-curtain martingale coupling.

3.1.1 The financial model and model-based prices for American puts

Suppose time is discrete and let $\tilde{M} = (\tilde{M}_{T_i})_{i=0,1,2}$ be the price of a financial security which pays no dividends, where $T_0 = 0$ is today's date. (In this section a superscript $\sim$ denotes an undiscounted quantity.) Suppose interest rates are non-stochastic. Let one unit of cash invested at time T_0 in a bank account paying the riskless rate be worth $\tilde{B}_{T_i}$ at time T_i for $i = 0, 1, 2$. Then $\tilde{B}_0 = 1$. Define $M = (M_i)_{i=0,1,2}$ by $M_i = \tilde{M}_{T_i}/\tilde{B}_{T_i}$, so that M is the discounted asset price (with a simplified time-index $i = 0, 1, 2$) which we expect to be a martingale under a pricing measure. We assume that M_0 is known at time 0.

Let Σ be the set of stopping rules taking values in $\{T_1, T_2\}$ and let $\mathcal{T}$ be the set of stopping rules taking values in $\{1, 2\}$. We are interested in pricing an American put with strike $\tilde{K}$ and maturity T_2, and which may be exercised at T_1 or T_2 only. (Note that we do not allow exercise at $t = 0$.) Define $K_i = \tilde{K}/\tilde{B}_{T_i}$ and note that $K_1 > K_2$ provided interest rates are strictly positive, which we now assume without further comment. Under a fixed model the expected payoff of an American put under an exercise (stopping) rule σ taking values in $\{T_1, T_2\}$ is given by $\mathbb{E}[\frac{1}{\tilde{B}_\sigma}(\tilde{K} - \tilde{M}_\sigma)^+]$ and the price of the American option (assuming exercise is only allowed at T_1 or T_2) is

$$\sup_{\sigma\in\Sigma} \mathbb{E}\left[\frac{1}{B_\sigma}(\tilde{K} - \tilde{M}_\sigma)^+\right] = \sup_{\tau\in\mathcal{T}} \mathbb{E}[(K_\tau - M_\tau)^+].$$

We suppose we are given European put prices $\{\tilde{P}_{T_i}(\tilde{k})\}_{\tilde{k}\geq 0}$ for $i = 1, 2$ for a continuum of strikes $\tilde{k}$. If the call prices have come from a (market) model for which the discounted price process is a martingale, then

$$\tilde{P}_{T_i}(\tilde{k}) = \frac{1}{\tilde{B}_{T_i}}\mathbb{E}[(\tilde{k} - \tilde{M}_{T_i})^+] = \mathbb{E}\left[\left(\frac{\tilde{k}}{\tilde{B}_{T_i}} - M_i\right)^+\right] =: P_i\left(\frac{\tilde{k}}{\tilde{B}_{T_i}}\right).$$

Then for fixed i we have $P_i(k) = \tilde{P}_{T_i}(k\tilde{B}_{T_i})$, and if we are given European put prices with maturity T_i then, by classical arguments (e.g. Breeden and Litzenberger [16]), it is possible to infer the laws of the price of the asset, and hence the laws of the discounted asset price:

$$\mathbb{P}(M_i < k) = P_i'(k-) = \frac{\partial}{\partial k}\tilde{P}_{T_i}(k\tilde{B}_{T_i}-).$$

Henceforth we assume we work in a discounted setting and with time-index in the set $i = 0, 1, 2$. In this setting the American put has payoff $(K_1 - M_1)^+$ at time

1 and payoff $(K_2 - M_2)^+$ at time 2. Denote the law of $X = M_1$ by μ and the law of $Y = M_2$ by ν. It follows from Jensen's inequality that, if μ and ν have arisen from sets of European put options, μ and ν are in convex order and we write $\mu \leq_{cx} \nu$ (see Section 2.1 for a further discussion of the properties of μ and ν).

Definition 3.1.1 (Hobson and Neuberger [59])**.** *Suppose $\mu \leq_{cx} \nu$.*

Let $\mathcal{S} = (\Omega, \mathcal{F}, \mathbb{P}, \mathbb{F} = \{\mathcal{F}_0, \mathcal{F}_1, \mathcal{F}_2\})$ be a filtered probability space. We say $M = (M_0, M_1, M_2) = (\bar{\mu}, X, Y)$ is a $(\mathcal{S}, \mu, \nu)$-consistent stochastic process and we write $M \in \mathcal{M}(\mathcal{S}, \mu, \nu)$ if

1. *M is a $\mathcal{S}$-martingale,*
2. *$\mathcal{L}(M_1) = \mu$ and $\mathcal{L}(M_2) = \nu$.*

We say $(\mathcal{S}, M)$ is a (μ, ν)-consistent model if $\mathcal{S}$ is a filtered probability space and M is a $(\mathcal{S}, \mu, \nu)$-consistent stochastic process. Where μ and ν are clear from the context this is sometimes abbreviated to a consistent model.

Let $B_1 \in \mathcal{F}_1$. Define the stopping time τ_{B_1} by $\tau_{B_1} = 1$ on B_1 and $\tau_{B_1} = 2$ on B_1^c. (Conversely, any stopping rule taking values in $\{1, 2\}$ has a representation of this form.) Suppose $(\mathcal{S}, M)$ is a (μ, ν)-consistent model. The $(\mathcal{S}, M)$ model-based expected payoff ($MBEP$) of the American put under stopping rule τ_{B_1} is

$$\mathcal{A}(B_1, M, \mathcal{S}) = \mathbb{E}[(K_{\tau_{B_1}} - M_{\tau_{B_1}})^+].$$

Then, optimising over stopping rules under the model $(\mathcal{S}, M)$, the model-based price of the American put is $\mathcal{A}(M, \mathcal{S}) = \sup_{B_1 \in \mathcal{F}_1} \mathcal{A}(B_1, M, \mathcal{S})$. The highest model-based expected payoff for the American put is

$$\mathcal{P} = \mathcal{P}(\mu, \nu) = \sup_{\mathcal{S}} \sup_{M \in \mathcal{M}(\mathcal{S}, \mu, \nu)} \mathcal{A}(M, \mathcal{S}).$$

Amongst the class of consistent models there is a natural and important class of models which we call the class of canonical models. Although in the finance context we typically expect non-negative prices, in this definition and in the mathematical analysis which follows, we allow for measures μ and ν supported on $\mathbb{R}$.

Definition 3.1.2. *Suppose $\mu \leq_{cx} \nu$.*

We say $(\hat{\mathcal{S}} = (\hat{\Omega}, \hat{\mathcal{F}}, \hat{\mathbb{P}}, \hat{\mathbb{F}} = \{\hat{\mathcal{F}}_0, \hat{\mathcal{F}}_1, \hat{\mathcal{F}}_2\}), \hat{M})$ is a canonical (μ, ν)-consistent model if $(\hat{\mathcal{S}}, \hat{M})$ is a (μ, ν)-consistent model such that $\hat{\Omega} = \mathbb{R} \times \mathbb{R}$, $\hat{\mathbb{F}} = \mathcal{B}(\hat{\Omega})$, $\hat{M}_1(\omega_1, \omega_2) = \omega_1$, $\hat{M}_2(\omega_1, \omega_2) = \omega_2$ and such that $\hat{\mathcal{F}}_0$ is trivial, $\hat{\mathcal{F}}_1 = \sigma(\hat{M}_1)$ and

$\hat{\mathcal{F}}_2 = \sigma(\hat{M}_1, \hat{M}_2)$. *Then* $\hat{B}_1 \in \mathcal{F}_1$ *can be identified with an element* $\hat{B}$ *in* $\mathcal{B}(\mathbb{R})$ *via* $\hat{B}_1 = \hat{B} \times \mathbb{R}$.

In the canonical setting different models (consistent or not) can be parametrised by a probability measure π on $\mathbb{R}^2$. To simplify the notation we shall write $\hat{M}_\pi$ for the canonical model $(\hat{\mathcal{S}}_\pi = (\hat{\Omega}, \hat{\mathcal{F}}, \hat{\mathbb{P}}_\pi, \hat{\mathbb{F}}), \hat{M})$, where $\hat{\mathbb{P}}_\pi$ is the probability measure such that $\hat{\mathbb{P}}_\pi(X \in dx, Y \in dy) = \pi(dx, dy)$. Note that there is a 1-1 correspondence between canonical, (μ, ν)-consistent models and martingale couplings $\pi \in \Pi(\mu, \nu)$ (recall Section 2.1), given by $\hat{\mathbb{P}}(\hat{M}_1 \in dx, \hat{M}_2 \in dy) = \pi(dx, dy)$.

We say π maps $A \subseteq \mathbb{R}$ to $B \subseteq \mathbb{R}$ if $\pi(A \times \mathbb{R}) = \pi(A \times B)$ (or equivalently if under the canonical model $\hat{M}_1 \in A$ implies $\hat{M}_2 \in B$ almost surely). We say π maps A onto B if $\pi(A \times \mathbb{R}) = \pi(A \times B) = \pi(\mathbb{R} \times B)$ (or equivalently if under the canonical model $\hat{M}_1 \in A$ if and only if $\hat{M}_2 \in B$ almost surely). Finally, we say π is constant on A if $\pi(C \times \mathbb{R}) = \pi(C \times C)$ for all $C \subseteq A$ or equivalently if $\hat{M}_2 = \hat{M}_1$ almost surely on $\hat{M}_1 \in A$.

Define $\hat{\mathcal{P}} = \sup \sup_{\hat{B} \in \mathcal{B}(\mathbb{R})} \mathbb{E}[(K_{\tau_{\hat{B}}} - M_{\tau_{\hat{B}}})_+]$, where the first supremum is taken over canonical (μ, ν)-consistent models and $\tau_{\hat{B}}(\omega) = 1$ if $X(\omega) \in \hat{B}$ and $\tau_{\hat{B}}(\omega) = 2$ otherwise. Clearly, since the set of canonical consistent models is a subset of the set of all consistent models, we have $\hat{\mathcal{P}} \leq \mathcal{P}$.

In this chapter we will concentrate on the case where μ is continuous. In that case we will show that $\hat{\mathcal{P}} = \mathcal{P}$. The main insight of this chapter is that the structure of functions f, g that characterise the left-curtain coupling (recall Section 2.3) allows us to determine the optimal stopping time and the optimal superhedging strategy. However, if μ has atoms, then the situation becomes more delicate, as pointed out in Hobson and Neuberger [60], see also Hobson and Neuberger [59], Bayraktar and Zhou [6] and Aksamit et al. [1]. In particular, it is sometimes possible to achieve a higher model price if we work on a richer probability space. (In the financial context, the choice of probability space is typically not specified. Instead the choice of probability space is a modelling issue, and it seems unreasonable to restrict attention to a sub-class of models without good reason, especially if this sub-class does not include the optimum.) On the one hand, we must allow for a wider range of possible candidates for exercise determining sets B_1. On atoms of X we may want to sometimes stop and sometimes continue, although we must still take stopping decisions which do not violate the martingale property of future price movements. On the other hand, the functions f, g that characterise the left-curtain coupling (see Section 2.3) become ill-defined on the points where μ has atoms. Then it is not clear how the optimal model can be identified. For these reasons we must extend our notion of a martingale coupling and generalise, in a useful fashion, the left-curtain martingale coupling of Beiglböck and Juillet [10] to the case with atoms. The appropriate extension of the left-curtain coupling to the case with atoms in μ is discussed in Chapter 4; in this chapter we focus on the financial aspects of our results, namely the application to the robust hedging of American puts.

3.1.2 Superhedging

The following notion of a robust superhedge for an American option was first introduced by Neuberger [83], see also Bayraktar and Zhou [6] and Hobson and Neuberger [60].

We work in discounted units over two time-points. Consider a general American-style option with payoff a if exercised at time 1, and payoff b if exercised at time 2, where $a : \mathbb{R} \mapsto \mathbb{R}_+$ and $b : \mathbb{R} \mapsto \mathbb{R}_+$ are positive functions.

Definition 3.1.3. $(\phi, \psi, \{\theta_i\}_{i=1,2})$ *is a superhedge for* (a, b) *if*

$$a(x) \leq \phi(x) + \psi(y) + \theta_1(x)(y - x), \tag{3.1}$$

$$b(y) \leq \phi(x) + \psi(y) + \theta_2(x)(y - x). \tag{3.2}$$

The hedging cost (HC) associated with the superhedge $(\phi, \psi, \{\theta_i\}_{i=1,2})$ *is given by*

$$\mathcal{C} = \mathcal{C}(\phi, \psi, \{\theta_i\}_{i=1,2}; \mu, \nu) = \int \phi(x)\mu(dx) + \int \psi(y)\nu(dy),$$

where we set $\mathcal{C} = \infty$ *if* $\int \phi(x)^+\mu(dx) + \int \psi(y)^+\nu(dy) = \infty$. *We let* $\mathcal{H}(a, b)$ *be the set of superhedging strategies* $(\phi, \psi, \{\theta_i\}_{i=1,2})$.

The idea behind the definition is that the hedger purchases a portfolio of maturity-1 European puts (and calls) with payoff ϕ and a portfolio of maturity-2 European puts (and calls) with payoff ψ. (The fact that this can be done and has cost $\mathcal{C}$ follows from arguments of Breeden and Litzenberger [16].) In addition, if the American option is exercised at time 1 the hedger holds θ_1 units of the underlying between times 1 and 2; otherwise the hedger holds θ_2 units of the underlying over this time-period. In the former case, (3.1) implies that the strategy superhedges the American option payout; in the later case (3.2) implies the same.

Remark 3.1.4. *We could extend the definition and allow a holding of* θ_0 *units of discounted asset over the time-period* $[0, 1)$. *Then the RHS of* (3.1) *would be*

$$\phi(x) + \psi(y) + \theta_0(x - M_0) + \theta_1(x)(y - x). \tag{3.3}$$

However, after a relabelling $\phi(x)+\theta_0(x-M_0) \mapsto \phi(x)$, (3.3) reduces to (3.1). (Note that $\int \theta_0(x-M_0)\mu(dx) = 0$ by the martingale property so that $\mathcal{C}$ is unchanged.) Similarly for (3.2). Hence there is no gain in generality by allowing non-zero strategies between times 0 and 1.

The dual (superhedging) problem is to find

$$\mathcal{D} = \mathcal{D}(a, b; \mu, \nu) = \inf_{(\phi,\psi,\{\theta_i\}_{i=1,2})\in\mathcal{H}(a,b)} \mathcal{C}(\phi, \psi, \{\theta_i\}_{i=1,2}; \mu, \nu).$$

Potentially the space $\mathcal{H} = \mathcal{H}(a, b)$ could be very large and it is extremely useful to be able to search over a smaller space. The next lemma shows that any convex ψ, with $\psi \geq b$, can be used to generate a superhedge $(\phi, \psi, \{\theta_i\}_{i=1,2})$.

For a convex function χ let χ'_+ denote the right-derivative of χ.

Lemma 3.1.5. *Suppose $\psi \geq b$ with ψ convex. Define $\phi = (a - \psi)^+$ and set $\theta_2 = 0$ and $\theta_1 = -\psi'_+$. Then $(\phi, \psi, \{\theta_i\}_{i=1,2})$ is a superhedge.*

Proof. We have

$$b(y) \leq \psi(y) \leq \phi(x) + \psi(y) = \phi(x) + \psi(y) + \theta_2(x)(y - x)$$

and (3.2) follows. Also, by the convexity of ψ, $\psi(x) \leq \psi(y) - \psi'_+(x)(y - x)$ and

$$a(x) \leq \big(a(x) - \psi(x)\big)^+ + \psi(x) \leq \phi(x) + \psi(y) + \theta_1(x)(y - x).$$

Hence (3.1) follows. □

Let $\breve{\mathcal{H}} = \breve{\mathcal{H}}(b)$ be the set of convex functions ψ with $\psi \geq b$. For $\psi \in \breve{\mathcal{H}}$ we can define the associated hedging cost $\breve{\mathcal{C}}(\psi; \mu, \nu)$ by

$$\begin{aligned}\breve{\mathcal{C}}(\psi; \mu, \nu) &= \mathcal{C}\left((a - \psi)^+, \psi, \theta_1 = -\psi'_+, \theta_2 = 0; \mu, \nu\right)\\ &= \int \big(a(x) - \psi(x)\big)^+ \mu(dx) + \int \psi(y)\nu(dy).\end{aligned}$$

The reduced dual hedging problem restricts attention to superhedges generated from $\psi \in \breve{\mathcal{H}}$ and is to find

$$\breve{\mathcal{D}} = \breve{\mathcal{D}}(a, b; \mu, \nu) = \inf_{\psi\in\breve{\mathcal{H}}(b)} \breve{\mathcal{C}}(\psi; \mu, \nu).$$

Clearly we have $\mathcal{D} \leq \breve{\mathcal{D}}$: we will show that $\mathcal{D} = \breve{\mathcal{D}}$ for the American put.

3.1.3 Weak and Strong Duality

Let $(\mathcal{S}, M)$ be a (μ,ν)-consistent model and let τ be an arbitrary stopping time in this framework. The expected payoff of the American put under this stopping rule is $\mathbb{E}[(K_\tau - M_\tau)^+]$. Conversely, let ψ be any convex function with $\psi(y) \geq (K_2 - y)^+$ and let $\phi(x) = [(K_1 - x)^+ - \psi(x)]^+$ and $\theta_i(x) = -\psi'_+(x)1_{\{i=1\}}$. Then for any $i \in \{1,2\}$ we have $(K_i - M_i)^+ \leq \psi(M_2) + \phi(M_1) + \theta_i(M_1)(M_2 - M_1)$ and hence for any random time τ taking values in $\{1,2\}$, $(K_\tau - M_\tau)^+ \leq \psi(M_2) + \phi(M_1) + \theta_\tau(M_1)(M_2 - M_1)$. Then $\mathbb{E}[(K_\tau - M_\tau)^+] \leq \mathbb{E}^{X\sim\mu, Y\sim\nu}[\phi(X) + \psi(Y)]$ and we have weak duality $\mathcal{P} \leq \mathcal{D}$.

In Section 3.2 we will show that we can find $(\hat{\mathcal{S}}^*, \hat{M}^*, \hat{B}^*)$ with $(\hat{\mathcal{S}}^*, \hat{M}^*)$ a canonical (μ,ν)-consistent model and $\hat{B}^* \subseteq \mathcal{B}(\mathbb{R})$, and $\psi^* \in \breve{H}$ such that

$$\mathcal{A}(\hat{B}^* \times \mathbb{R}, \hat{M}^*, \hat{\mathcal{S}}^*) = \breve{\mathcal{C}}(\psi^*; \mu, \nu).$$

Then $\mathcal{A}(\hat{B}^* \times \mathbb{R}, \hat{M}^*, \hat{\mathcal{S}}^*) \leq \hat{\mathcal{P}} \leq \mathcal{P} \leq \mathcal{D} \leq \breve{\mathcal{D}} \leq \breve{\mathcal{C}}(\psi^*; \mu, \nu)$ but since the two outer terms are equal we have $\mathcal{P} = \mathcal{D}$ and strong duality. Moreover, $(\hat{\mathcal{S}}^*, \hat{M}^*)$ is a canonical, consistent model which generates the highest price for the American put (and τ^* given by $\tau^* = 1$ if and only if $X \in \hat{B}^*$ is the optimal exercise rule) and ψ^* generates the cheapest superhedge.

3.2 Robust bounds for American puts when μ is atom-free

3.2.1 Problem formulation

Our goal in this section is to derive the highest consistent model price for the American put. We begin by giving a concise reformulation of the primal problem (recall Section 3.1.1) as a problem of martingale optimal transport (MOT), and stating the main theorem (Theorem 3.2.1). Then we first study the problem in a simple special case, second generalise to a case which exhibits all the main features and third present the analysis in the general case.

Recall the definition of the canonical (μ,ν)-consistent model (abbreviated to $\hat{M}_\pi$) for which $\hat{\mathbb{P}}(\hat{M}_1 \in dx, \hat{M}_2 \in dy) = \pi(dx, dy)$ where $\pi \in \Pi(\mu,\nu)$. For a pair of fixed constants K_1 and K_2 the problem we consider is to find

$$\hat{\mathcal{P}} := \sup_{\pi\in\Pi(\mu,\nu)} \sup_{B\in\mathcal{B}(\mathbb{R})} \mathbb{E}^{\mathcal{L}(X,Y)\sim\pi}\big[(K_1 - X)^+ 1_{\{X\in B\}} + (K_2 - Y)^+ 1_{\{X\notin B\}}\big]. \tag{3.4}$$

Note that $\hat{\mathcal{P}}$ corresponds to the highest model-based price of the American put over the specific subset of consistent models, and therefore $\hat{\mathcal{P}} \leq \mathcal{P}$. By weak duality (recall Section 3.1.3) it follows that $\hat{\mathcal{P}} \leq \mathcal{P} \leq \mathcal{D} \leq \breve{\mathcal{D}}$.

Throughout this chapter we assume that μ has no atoms. The same assumption is made in (parts of) Beiglböck and Juillet [10], Henry-Labordère and Touzi [50] and Beiglböck et al. [9]. (The case with atoms requires an extension of the left-curtain martingale coupling, which is the subject of Chapter 4.)

Theorem 3.2.1. *Suppose μ has no atoms. Then $\hat{\mathcal{P}} = \mathcal{P} = \mathcal{D} = \breve{\mathcal{D}}$.*

We begin by considering a couple of degenerate cases.

We say the put is in-the-money at time 1 (respectively time 2) if $X < K_1$ (respectively $Y < K_2$). If the inequality is reversed then the put is out-of-the-money. Recall that $\{\ell_\mu, r_\mu\}$ (resp. $\{\ell_\nu, r_\nu\}$) are the endpoints of $\mathcal{I}_\mu$ (resp. $\mathcal{I}_\nu$), the smallest interval containing the support of μ (resp. ν). If $K_1 \leq \ell_\mu$ then the American put is always out-of-the-money at time 1, and the American put is equivalent to the European put with strike K_2 and maturity 2. Since puts with strike K and maturity 1 are costless for $K \leq \ell_\mu$, a simple superhedging strategy is to purchase one European put with strike K_1 and sell one European put with strike K_2, both with maturity 1, and also purchase one European put with strike K_2 and maturity 2. (This strategy is of the form discussed in Lemma 3.1.5 and is generated by $\Psi(y) = (K_2 - y)^+$.) The cost of this hedge is $P_\nu(K_2)$, this is also the model-based expected payoff of the American put under *any* consistent model.

If $K_1 \leq K_2$ then $\mathbb{E}[(K_2 - Y)^+|X] \geq (K_2 - X)^+ \geq (K_1 - X)^+$ and $\tau = 2$ is optimal. Again, the American put is equivalent to the European put with strike K_2 and maturity 2. In this case, for a superhedge it is sufficient to purchase one European put with strike K_2 and maturity 2. By Lemma 3.1.5 (with $\psi(y) = (K_2 - y)^+$ and $\phi = 0$) this generates a superhedge with cost $P_\nu(K_2)$. Again, this is the the model-based expected payoff of the American put under *any* consistent model.

For the remainder of the chapter we make

Standing Assumption. $K_1 > \max\{\ell_\mu, K_2\}$.

Remark 3.2.2. *Recall Lemma 2.1.6 in Chapter 2. In the present setting, in addition to specifying a model (or equivalently a martingale coupling) we also need to specify a stopping rule, and this needs to be defined across all irreducible components simultaneously. For this reason, when looking for an optimal martingale coupling $\pi \in \Pi(\mu, \nu)$, we do not insist that $D > 0$ on the interior of $\mathcal{I}_\nu$, although this will be the case in the simple settings in which we build our solution.*

3.2.2 American puts under the dispersion assumption

The left-curtain coupling

The goal in this section is to present the theory in a simple special case, and to illustrate the main features and solution techniques of our approach unencumbered by technical issues or the consideration of exceptional cases. The following assumption is a small modification of one introduced by Hobson and Klimmek [57], see also Henry-Labordère and Touzi [50]. See Figure 3.1.

Assumption 3.2.3 (Dispersion Assumption). *μ and ν are absolutely continuous with continuous densities ρ and η, respectively. ν has support on $(\ell_\nu, r_\nu) \subseteq (-\infty, \infty)$ and $\eta > 0$ on (ℓ_ν, r_ν). μ has support on $(\ell_\mu, r_\mu) \subseteq (\ell_\nu, r_\nu)$ and $\rho > 0$ on (ℓ_μ, r_μ). In addition:*
$(\mu - \nu)^+$ is concentrated on an interval $E = (e_-, e_+)$ and $\rho > \eta$ on E;
$(\nu - \mu)^+$ is concentrated on $(\ell_\nu, r_\nu) \setminus E$ and $\eta > \rho$ on $(\ell_\nu, e_-) \cup (e_+, r_\nu)$.

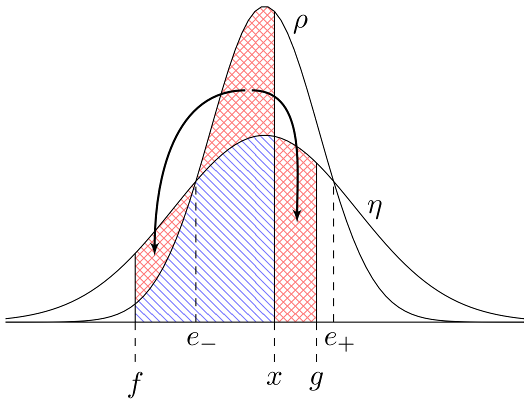

Figure 3.1: Sketch of the densities ρ and η and, for given $x > e_-$, the locations of the functions f, g that characterise the left-curtain martingale coupling, evaluated at x. Time-1 mass in the interval (f, x) stays in the same place if possible. Mass which cannot stay constant is mapped to (f, e_-) or (x, g) in a way which respects the martingale property.

Example 3.2.4. *If $\mu \leq_{cx} \nu$ are centred normal distributions with different variances or distinct lognormal random variables with common mean, then Assumption 3.2.3 is satisfied.*

Under the Dispersion Assumption $\{k : D_{\mu,\nu}(k) > 0\}$ is an interval and $D = D_{\mu,\nu}$ is convex to the left of e_-, concave on (e_-, e_+) and again convex above e_+.

Lemma 3.2.5 (Henry-Labordère and Touzi [50], Section 3.4)**.** *Suppose Assumption 3.2.3 holds. For all $x \in (e_-, r_\mu)$, there exist f, g with $f < e_- < x < g$ such that* (2.10) *and* (2.11) *hold. Moreover, if we consider f and g as functions of x on (e_-, r_μ) then f and g are continuous, f is strictly decreasing and g is strictly increasing,* $\lim_{x \downarrow e_-} f(x) = e_- = \lim_{x \downarrow e_-} g(x)$, $\lim_{x \uparrow r_\mu} f(x) = \ell_\nu$ *and* $\lim_{x \uparrow r_\mu} g(x) = r_\nu$*. Finally, if we extend the domain of f and g to $[\ell_\mu, r_\mu]$ by setting $f(x) = x = g(x)$ on $[\ell_\mu, e_-]$ and $f(r_\mu) = \ell_\nu$ and $g(r_\mu) = r_\nu$ then $(f, g) \in \Xi_{Mon}^{[\ell_\mu, r_\mu], [\ell_\nu, r_\nu]}$.*

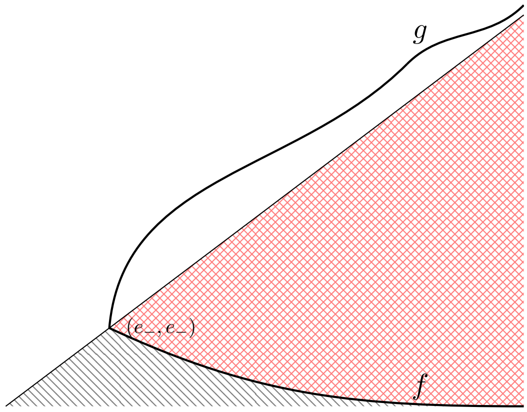

Figure 3.2: Sketch of functions f and g under the Dispersion Assumption, with the regions $K_2 < f(K_1)$ and $K_2 > f(K_1)$ shaded. This is a simple special case of Figure 2.1.

Remark 3.2.6. *As discussed in Lemma 2.3.7 and the paragraph above it, for the purposes of the analysis of this section it is not the fact that the measures μ and ν satisfy the Dispersion Assumption which is important, but rather that π_{lc} is so simple, and $\{k : g(k) > k\}$ is a single interval on which f is a monotone decreasing function.*

Starting with monotonic f and g, letting μ be continuous and defining ν by $\nu(dy) = \int_x \mu(dx)\chi_{f(x),x,g(x)}(dy)$ and π_{lc} by

$$\pi_{lc}(dx, dy) = \mu(dx)\delta_x(dy)1_{\{x \le e_-\}} + \mu(dx)\chi_{f(x),x,g(x)}(dy)1_{\{x > e_-\}}, \qquad (3.5)$$

the pair (μ, ν) may or may not satisfy Assumption 3.2.3 but nonetheless, a candidate optimal model, stopping time and hedge can be constructed exactly as described in this section, and can be proved to be optimal by the methods of this section.

Since our analysis depends on the pair (μ, ν) only through the functions (f, g) we may take as our starting point $(f, g) \in \Xi_{Mon}$.

The principle behind the left-curtain martingale coupling in Beiglböck-Juillet [10] is that they determine where to map mass at x at time 1 sequentially working from left to right. In our current setting there is an interval $(\ell_\mu, e_-]$ on which mass can remain unmoved between times 1 and 2. To the right of e_- we can define f, g in such a way that mass is moved as little as possible. This leads to the ODEs in Remark 2.3.6.

The American put

Suppose $K_1 \in (e_-, r_\mu]$ and suppose f and g are constructed as in Lemma 3.2.5. Define $\Lambda : [g^{-1}(K_1), K_1] \mapsto \mathbb{R}$ by

$$\Lambda(x) = \frac{\big(K_2 - f(x)\big) - (K_1 - x)}{x - f(x)} - \frac{(K_1 - x)}{g(x) - x} = \frac{\big(g(x) - K_1\big)}{g(x) - x} - \frac{(K_1 - K_2)}{x - f(x)}. \tag{3.6}$$

Pictorially Λ is the difference in slope of the two dashed lines in Figure 3.3.

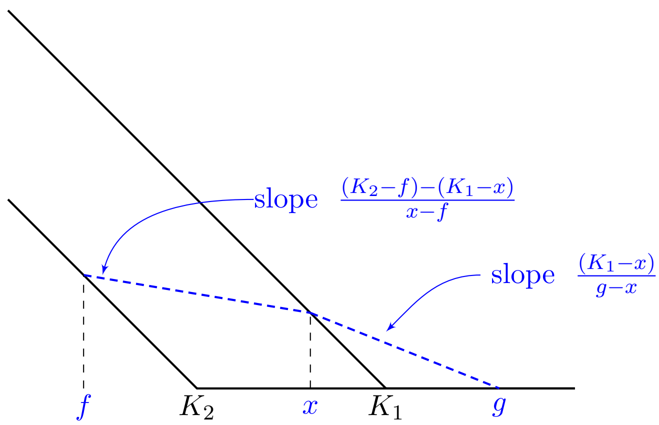

Figure 3.3: Sketch of put payoffs with points x, f and g marked. $\Lambda(x)$ is the difference in slope of the two dashed lines.

Lemma 3.2.7. *Suppose $K_1 \in (e_-, r_\mu]$ and $f(K_1) < K_2$. Then there is a unique scalar $x^* = x^*(\mu, \nu; K_1, K_2) \in (g^{-1}(K_1), K_1)$ such that $\Lambda(x^*) = 0$. Moreover $f(x^*) < K_2$ and*

$$\frac{\big(K_2 - f(x^*)\big)}{g(x^*) - f(x^*)} = \frac{(K_1 - x^*)}{g(x^*) - x^*} = \frac{\big(x^* - f(x^*)\big) - (K_1 - K_2)}{x^* - f(x^*)} = 1 - \frac{(K_1 - K_2)}{x^* - f(x^*)}. \tag{3.7}$$

Proof. First note that, from the continuity and monotonicity properties of f and g, we have that (see Figure 3.3) Λ is continuous and strictly increasing. Moreover,

$\Lambda(g^{-1}(K_1)) = -\frac{(K_1-K_2)}{g^{-1}(K_1)-(f\circ g^{-1})(K_1)} < 0$ and $\Lambda(K_1) = \frac{K_2-f(K_1)}{K_1-f(K_1)} > 0$ by hypothesis. Hence there is a unique root to $\Lambda = 0$. At this root the equalities in (3.7) hold. $\square$

Suppose $K_1 > e_-$ and $f(K_1) < K_2$ and that $x^* = x^*(\mu, \nu; K_1, K_2) \in (e_-, K_1)$ is such that $\Lambda(x^*) = 0$. It is easy to find a martingale coupling π of μ and ν such that π maps $(f(x^*), x^*)$ onto $(f(x^*), g(x^*))$, and π is constant on $(-\infty, f^*)$. For example, we may take $\hat{\pi} = \pi_{lc} = \pi_{lc}(\mu, \nu)$, the left-curtain martingale coupling of Beiglböck and Juillet [10]. More generally, let $\pi_{x^*} \in \Pi(\mu, \nu)$ be any martingale coupling such that π_{x^*} maps $(-\infty, f(x^*))$ to itself, maps $(f(x^*), x^*)$ onto $(f(x^*), g(x^*))$ and maps (x^*, ∞) to $(-\infty, f(x^*)) \cup (g(x^*), \infty)$. The martingale coupling represented in Figure 3.2 has this property.

Consider a canonical (μ, ν)-consistent model $\hat{M}_{\pi_{x^*}}$, under which the corresponding probability measure $\hat{\mathbb{P}}$ is given by $\hat{\mathbb{P}}(X \in dx, Y \in dy) = \pi_{x^*}(dx, dy)$. Let τ^* be the stopping time such that $\tau^* = 1$ on $(-\infty, x^*)$ and $\tau^* = 2$ otherwise. Our claim in Theorem 3.2.8 below is that $\hat{M}_{\pi_{x^*}}$ and the stopping time τ^* are such that the model-based price of the American put under this stopping time is the highest possible, over all consistent models.

Continue to suppose $K_1 > e_-$ and $f(K_1) < K_2$. Now we define a superhedge of the American put. Let ψ^* be the function

$$\psi^*(z) = \begin{cases} (K_2 - z) & z \le f(x^*); \\ \frac{\left(g(x^*)-z\right)\left(K_2-f(x^*)\right)}{g(x^*)-f(x^*)} & f(x^*) < z \le g(x^*); \\ 0 & z > g(x^*). \end{cases} \tag{3.8}$$

Note that by construction and by (3.7), $\frac{K_2-f(x^*)}{g(x^*)-f(x^*)} = \frac{K_1-x^*}{g(x^*)-x^*}$. Therefore, we have that $\psi^*(x^*) = K_1 - x^*$. Moreover, ψ^* is convex and satisfies $\psi^*(z) \ge (K_2 - z)^+$. Hence by Lemma 3.1.5, ψ^* can be used to generate a superhedge $(\psi^*, \phi^*, \theta^*_{1,2})$.

In the following theorem we will assume the American put is not always strictly in-the-money at time 1 (or equivalently, $K_1 \le r_\mu$). Discussion of the case $K_1 > r_\mu$ is postponed until Section 3.2.3 below.

Theorem 3.2.8. *Suppose Assumption 3.2.3 holds and $K_1 \le r_\mu$.*

1. *Suppose $K_1 \in (e_-, r_\mu]$ and $f(K_1) < K_2$. The model $\hat{M}_{\pi_{x^*}}$ described in the previous paragraphs is a canonical (μ, ν)-consistent model for which the price of the American option is the highest. The stopping time τ^* is the optimal exercise time. The function ψ^* defined in* (3.8) *defines the cheapest superhedge. Moreover, the highest model-based price is equal to the cost of the cheapest superhedge.*

2. *Suppose either that Case A: $K_1 \leq e_-$ or that Case B: $K_1 \in (e_-, r_\mu]$ together with $f(K_1) \geq K_2$. Then there exists is a canonical, consistent model under which*

$$\{Y < K_2\} = \{X < K_2\} \cup \{X > K_1, Y < K_2\}$$

and any model with this property, with the stopping rule $\tau = 1$ if $X < K_1$ and $\tau = 2$ otherwise, attains the highest consistent model price. The cheapest superhedge is generated from $\psi(x) = (K_2 - x)^+$ and the highest model-based price is equal to the cost of the cheapest hedge.

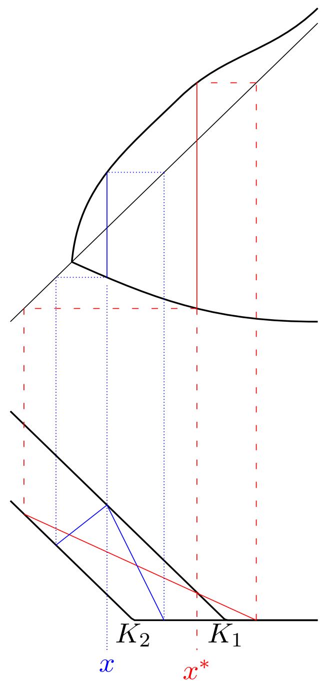

Figure 3.4: A combination of Figures 3.2 and 3.3, showing how jointly they define the best model and best hedge. By adjusting x we can find x^* such that $\Lambda(x^*) = 0$. Together the quantities $(f(x^*), x^*, g(x^*))$ define the optimal model, stopping time and hedge.

Remark 3.2.9. *In Part 2 of the Theorem 3.2.8, the left-curtain coupling generates a model for which $\{Y < K_2\} = \{X < K_2\} \cup \{X > K_1, Y < K_2\}$, and hence when associated with the stopping rule of the theorem, attains the highest consistent model price.*

Proof. 1. Suppose $K_1 > e_-$ and $f(K_1) < K_2$. Then by Lemma 3.2.7 there is a unique $x^* \in (g^{-1}(K_1), K_1)$ such that $\Lambda(x^*) = 0$. For this x^* we can find $f^* = f(x^*)$ and

$g^* = g(x^*)$ with $f^* < K_2$ and $K_1 < g^*$ such that $\frac{K_2 - f^*}{g^* - f^*} = \frac{K_1 - x^*}{g^* - x^*}$. For typographical reasons we abbreviate this (x^*, f^*, g^*) to (x, f, g) for the rest of this proof.

Since ν is continuous we have that f, x, g solve (2.10) and (2.11). The elements f, x, g can be used to define a model using the construction after Lemma 3.2.7 above. For this model we can calculate the expected payoff of the American put. At the same time we can use (f, x, g) to define a superhedge. The remaining task is to show that the cost of the superhedge equals that of the model-based expected payoff. Then by the discussion in Section 3.1.3 we have found an optimal model and a cheapest superhedge.

The model-based expected payoff ($MBEP$) of the American put (for this model and stopping rule) is

$$\begin{aligned} MBEP &= \int_{-\infty}^{x} (K_1 - w)\mu(dw) + \int_{-\infty}^{f} (K_2 - w)(\nu - \mu)(dw) \\ &= P_\mu(x) + (K_1 - x)P'_\mu(x) + D(f) + (K_2 - f)D'(f). \end{aligned}$$

Now we consider the hedging cost (HC). Set $\Theta = \frac{K_2 - f}{g - f} \in (0,1)$. Note that, since x is such that $\Lambda(x) = 0$, we have $\Theta = \frac{K_1 - x}{g - x}$. Recall the definition of ψ^* in (3.8). Then

$$\psi^*(y) = \Theta(g - y)^+ + (1 - \Theta)(f - y)^+.$$

Following Lemma 3.1.5 we can use ψ^* to generate a superhedging strategy. The hedging cost (HC) of this strategy is

$$HC = \Theta P_\nu(g) + (1 - \Theta)P_\nu(f) + (1 - \Theta)\big(P_\mu(x) - P_\mu(f)\big), \tag{3.9}$$

where the first two terms arise from the purchase of the static time-2 portfolio ψ^* and the third comes from the purchase of the time-1 portfolio $((K_1 - w)^+ - \psi^*(w))^+$. The expression in (3.9) can be rewritten as

$$P_\mu(x) + D(f) + \Theta\big(P_\nu(g) - P_\nu(f) - P_\mu(x) + P_\mu(f)\big).$$

Now we consider the difference between the hedging cost and the model-based expected payoff. First recall that $P_\chi(k) = \int_{-\infty}^{k} (k - x)\chi(dx)$, $\chi \in \{\mu, \nu\}$, and that $D(k) = D_{\mu,\nu}(k) = P_\nu(k) - P_\mu(k)$. Then (2.10) and (2.11) can be rewritten as

$$P'_\mu(x) - P'_\mu(f) = P'_\nu(g) - P'_\nu(f), \tag{3.10}$$

$$\big(xP'_\mu(x) - P_\mu(x)\big) - \big(fP'_\mu(f) - P_\mu(f)\big) = \big(gP'_\nu(g) - P_\nu(g)\big) - \big(fP'_\nu(f) - P_\nu(f)\big). \tag{3.11}$$

We find

$$\begin{aligned}
HC - MBEP &= \Theta\big(P_\nu(g) - P_\nu(f) - P_\mu(x) + P_\mu(f)\big) \\
&\qquad\qquad - (K_1 - x)P'_\mu(x) - (K_2 - f)D'(f) \\
&= \Theta\big(gP'_\nu(g) - xP'_\mu(x) - fD'(f)\big) - (K_1 - x)P'_\mu(x) - (K_2 - f)D'(f) \\
&= \Theta\big((g - x)P'_\mu(x) + (g - f)D'(f)\big) - (K_1 - x)P'_\mu(x) - (K_2 - f)D'(f) \\
&= P'_\mu(x)\big(\Theta(g - x) - (K_1 - x)\big) + D'(f)\big(\Theta(g - f) - (K_2 - f)\big) \\
&= 0,
\end{aligned}$$

where we use (3.11), (3.10) and the definition of Θ, respectively. Optimality of the model, stopping rule and hedge now follows.

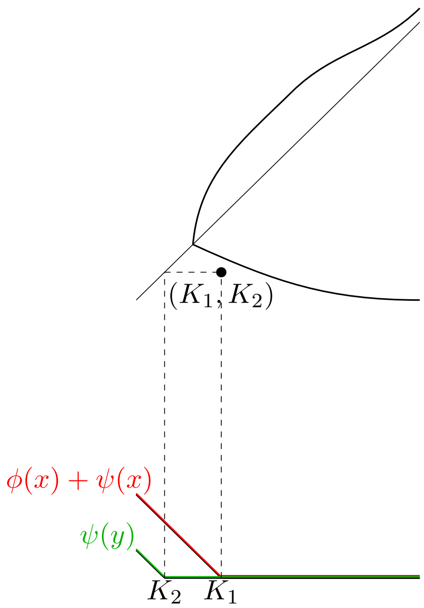

Figure 3.5: Sketch of put payoffs with $\psi(y) = (K_2 - y)^+$ and $\phi(x) = (K_1 - x)^+ - (K_2 - x)^+$.

2. Now suppose $K_1 \leq e_-$. Consider an exercise rule in which the American put is exercised at time 1 if it is in-the-money, otherwise it is exercised at time 2, and a model in which mass below K_1 at time 1 stays constant between times 1 and 2. (This is possible since $\mu \leq \nu$ on $(-\infty, e_-)$ and $K_1 \leq e_-$.) The expected payoff of the American put is

$$\int_{-\infty}^{K_1} (K_1 - x)\mu(dx) + \int_{-\infty}^{K_2} (K_2 - y)(\nu - \mu)(dy) = P_\mu(K_1) + P_\nu(K_2) - P_\mu(K_2). \quad (3.12)$$

Alternatively, suppose $K_1 > e_-$ and $f(K_1) \geq K_2$. Then under the left-curtain martingale coupling mass below K_2 at time 1 stays constant between times 1 and 2 (note that $K_2 \leq f(K_1) \leq e_-$), and mass between K_2 and K_1 at time 1 is mapped to (K_2, ∞). Then, mass which is below K_2 at time 2 was either below K_2 at time 1, or above K_1 at time 1. The expected payoff under this model (using a strategy of exercising at time 1 if the American put is in-the-money) is again given by (3.12).

Now consider the hedging cost. Let $\psi(y) = (K_2 - y)^+$. Defining ϕ as in Lemma 3.1.5 we find $\phi(x) = (K_1 - x)^+ - (K_2 - x)^+ = (K_1 - (x \vee K_2))^+$ and the superhedging cost is

$$HC = P_\nu(K_2) + P_\mu(K_1) - P_\mu(K_2).$$

Hence the model-based expected payoff equals the hedging cost. □

3.2.3 Two intervals of $g > x$ and one downward jump in f

We now relax the Dispersion Assumption to the case where f is not monotone. The simplest situation when this may arise is when there are two intervals on which $g(x) > x$. We do not contend that there are many natural examples which fall into this situation, but rather that this intermediate case illustrates phenomena which are to be found in the general case but which were not to be found under the Dispersion Assumption.

Assumption 3.2.10 (Single Jump Assumption). *μ and ν are absolutely continuous with continuous densities ρ and η, respectively. ν has support on $(\ell_\nu, r_\nu) \subseteq (-\infty, \infty)$ and $\eta > 0$ on (ℓ_ν, r_ν). μ has support on $(\ell_\mu, r_\mu) \subseteq (\ell_\nu, r_\nu)$ and $\rho > 0$ on (ℓ_μ, r_μ). In addition:*
$(\mu - \nu)^+$ is concentrated on $E = (e^1_-, e^1_+) \cup (e^2_-, e^2_+)$ with $e^1_+ < e^2_-$ and $\rho > \eta$ on E;
$(\nu - \mu)^+$ is concentrated on $(\ell_\nu, r_\nu) \setminus E$ and $\eta > \rho$ on $(\ell_\nu, e^1_-) \cup (e^1_+, e^2_-) \cup (e^2_+, r_\nu)$;
there exists $f' < e^1_-$ and $x' \in (e^1_+, e^2_-)$ such that

$$\int_{f'}^{x'} \mu(dz) = \int_{f'}^{x'} \nu(dz) \qquad \textit{and} \qquad \int_{f'}^{x'} z\mu(dz) = \int_{f'}^{x'} z\nu(dz). \tag{3.13}$$

Under Assumption 3.2.10 it is possible to find functions $g : (\ell_\mu, r_\mu) \to (\ell_\nu, r_\nu)$ and $f : (\ell_\mu, r_\mu) \to (\ell_\nu, r_\nu)$ with the properties (see the lower part of Figure 3.6):

1. $g(x) = x$ on $(\ell_\mu, e^1_-] \cup [x', e^2_-]$;

2. $g(x) > x$ on $(e^1_-, x') \cup (e^2_-, r_\mu)$;

3. g is continuous and strictly increasing;

4. $f(x) = x$ on $(\ell_\mu, e^1_-] \cup [x', e^2_-]$;

5. $f : (e^1_-, x') \mapsto (f', x')$ is continuous and strictly decreasing;

6. $f : (e^2_-, r_\mu) \mapsto (\ell_\nu, e^2_-) \setminus (f', x')$ is strictly decreasing;

7. there exists $x'' \in (e^2_-, r_\mu)$ such that f jumps at x'' and $f(x''-) = x' > f' = f(x''+)$. Away from x'', f is continuous on (e^2_-, r_μ).

By construction we have that

$$\int_{x'}^{x''} \mu(dz) = \int_{x'}^{g(x'')} \nu(dz); \qquad \int_{x'}^{x''} z\mu(dz) = \int_{x'}^{g(x'')} z\nu(dz), \tag{3.14}$$

so that if mass in (x', x'') at time 1 is mapped to $(x', g(x''))$ at time 2 then total mass and mean are preserved. Further, given that (f', x') satisfy (3.13), we also have that $\int_{f'}^{x''} \mu(dz) = \int_{f'}^{g(x'')} \nu(dz)$ and $\int_{f'}^{x''} z\mu(dz) = \int_{f'}^{g(x'')} z\nu(dz)$. In particular, given (3.13) and (3.14), the pair of equations

$$\int_{f}^{x''} \mu(dz) = \int_{f}^{g(x'')} \nu(dz); \qquad \int_{f}^{x''} z\mu(dz) = \int_{f}^{g(x'')} z\nu(dz)$$

has two solutions for f, namely $f = x'$ and $f = f'$. Hence, in defining the left-curtain martingale coupling there are two choices for f at x'': we may take $f(x'') = x'$ or $f(x'') = f'$. Rather than assuming one of these choices (for example by requiring left-continuity of f) it is convenient to allow f to be multi-valued. Then, for each x, such that $g(x) > x$, let $\aleph(x) = \{f : (f, x, g(x)) \text{ solves (2.10) and (2.11)}\}$. Then we have that, in the setting of Assumption 3.2.10, for $x > e_-$, $|\aleph(x)| = 1$ except at x'' and there $\aleph(x'') = \{f(x''+), f(x''-)\} = \{f', x'\}$.

Remark 3.2.11. *As discussed in Lemma 2.3.7 and a paragraph above it, when constructing examples which fit with the analysis of this section, we may begin with f, g as presented in the bottom half of Fig. 3.6. Given μ with support (ℓ_μ, r_μ) we* *can define ν via $\nu(dy) = \int \mu(dx)\chi_{f(x),x,g(x)}(dy)$. Then the pair (μ,ν) satisfy the hypothesis of Assumption 3.2.10.*

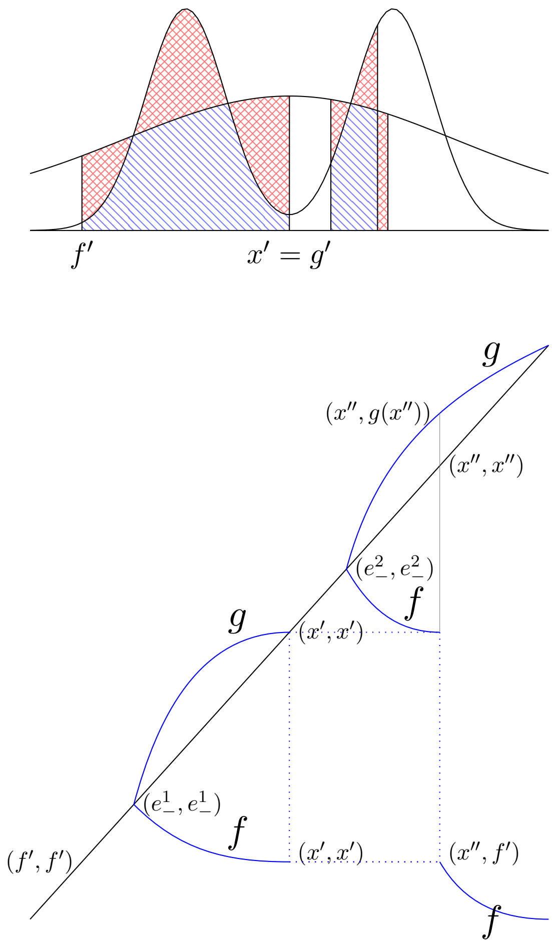

Figure 3.6: Picture of f and g under Assumption 3.2.10.

Remark 3.2.12. *Recall Remark 2.3.6 and the principle that quantities in the left-curtain coupling are determined working from left to right. Given that μ and ν have continuous densities and given that $\eta > \rho$ on (ℓ_μ, e^1_-) we can set $f = g = x$ on this interval. To the right of e^1_- we have $\rho > \eta$ and we can define f and g using the differential equations in Remark 2.3.6. There are two cases, either $g(x) > x$ for all $x \in (e^1_-, r_\mu)$ (in which case we can define (f,g) on (e^1_-, r_μ) with the properties described in Lemma 3.2.5) or there is some point at which g first hits the diagonal line $y = x$ again. This point is exactly x'.*

If x' exists it must satisfy $x' \in (e_+^1, e_-^2)$. Then we set $g(x) = x$ on (x', e_-^2) and let $f = g$ solve the same coupled differential equations as in Remark 2.3.6 but with a new starting point $g(e_-^2) = e_-^2 = f(e_-^2)$. The ODEs determine f and g until f first reaches x'. This happens at x'', and at x'' f jumps down to f' (and g is continuous). To the right of x'', f and g solve the differential equations again subject to initial conditions $f(x'') = f', g(x'') = g(x''-)$.

Recall the definition $\Lambda(x) = \frac{g(x)-K_1}{g(x)-x} - \frac{(K_1-K_2)}{x-f(x)}$. If f is multi-valued, then Λ will also be multi-valued. In Section 3.2.2, one of our main steps was to find x such that $\Lambda(x) = 0$, and our aim is similar here.

Introduce $\Upsilon = \Upsilon_{K_1,K_2}(f, x, g)$ which is defined for $f \leq K_2, x \leq K_1 \leq g$ by

$$\Upsilon(f, x, g) = \frac{(K_2 - f) - (K_1 - x)}{x - f} - \frac{K_1 - x}{g - x} = \frac{g - K_1}{g - x} - \frac{(K_1 - K_2)}{x - f}.$$

Instead of seeking x which is a root of $\Lambda(x) = 0$ our goal is to find (f, x, g) with $g = g(x)$ and $f \in \aleph(x)$ such that $\Upsilon(f, x, g) = 0$.

For a fixed K_1, the value of K_2 such that $\Upsilon(f' = f(x''+), x'', g(x'')) = 0$ is given by $K_2 = f' + (K_1 - x'')\frac{g(x'')-f'}{g(x'')-x''}$. On the other hand, setting $K_2 = x' + (K_1 - x'')\frac{g(x'')-x'}{g(x'')-x''}$ gives $\Upsilon(x' = f(x''-), x'', g(x'')) = 0$. This motivates the introduction of the linear increasing functions L_u, $L_d : [x'', g(x'')] \mapsto \mathbb{R}$ defined by

$$L_u(x) = x' + (x - x'')\frac{g(x'') - x'}{g(x'') - x''},$$
$$L_d(x) = f' + (x - x'')\frac{g(x'') - f'}{g(x'') - x''}.$$

Pictorially, L_d and L_u are the lower and upper boundaries, respectively, of the dotted triangular area $\mathcal{G}$ in Figure 3.7.

From Figure 3.7 we identify four regions (and various subregions) on which four different hedging strategies will be needed in order to find the cheapest super-hedge for the American put. (Compare this with two regimes under the Dispersion Assumption in Figure 3.2.)

Define

$$\mathcal{R}_1 = \{(k_1, k_2) : e_-^1 < k_1 < x', f(k_1) < k_2 < k_1\},$$

which we write more compactly as $\mathcal{R}_1 = \{e_-^1 < k_1 < x', f(k_1) < k_2 < k_1\}$. Using the same compact notation define

$$
\begin{aligned}
\mathcal{R}_2 &= \{e^2_- < k_1 < x'', f(k_1) < k_2 < k_1\} \cup \{k_1 = x'', x' < k_2 < k_1\};\\
\mathcal{R}_3 &= \{x'' < k_1 < g(x''), L_u(k_1) \le k_2 < k_1\};\\
\mathcal{R}_4 &= \{x'' < k_1 < g(x''), f(k_1) < k_2 \le L_d(k_1)\};\\
\mathcal{R}_5 &= \{g(x'') \le k_1 \le r_\mu, f(k_1) < k_2 < k_1\};\\
\mathcal{B}_1 &= \{\ell_\mu \le k_1 \le e^1_-, k_2 < k_1\} \cup \{e^1_- < k_1 < x', k_2 \le f(k_1)\}\\
&\qquad\qquad\qquad \cup \{x'' < k_1 \le r_\mu, k_2 \le f(k_1)\};\\
\mathcal{B}_2 &= \{x' \le k_1 \le x'', k_2 \le f'\};\\
\mathcal{B}_3 &= \{x' \le k_1 \le e^2_-, x' \le k_2 < k_1\} \cup \{e^2_- < k_1 \le x'', x' \le k_2 \le f(k_1)\};\\
\mathcal{G} &= \{x'' < k_1 < g(x''), L_d(k_1) < k_2 < L_u(k_1)\};\\
\mathcal{W} &= \{x' \le k_1 \le x'', f' < k_2 < x'\};
\end{aligned}
$$

and set $\mathcal{R} = \cup_{i=1}^5 \mathcal{R}_i$ and $\mathcal{B} = \cup_{i=1}^3 \mathcal{B}_i$. In general, on the boundaries between the regions the boundaries could be allocated to either region. However, we allocate points on the boundary to the region where the hedge is simplest.

Note that $\mathcal{R} \cup \mathcal{B} \cup \mathcal{G} \cup \mathcal{W} = \{(k_1, k_2) : \ell_\mu \le k_1 \le r_\mu, k_2 < k_1\}$.

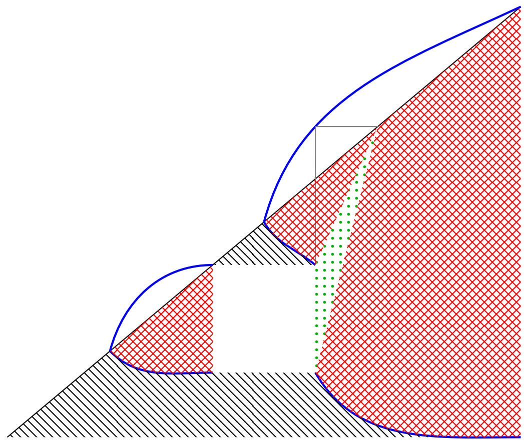

Figure 3.7: Picture of f and g in the single jump case, now with 4 regions shaded (cross-hatched, diagonally, dotted and blank).

Case $(K_1, K_2) \in \mathcal{R}$**.**

Lemma 3.2.13. *Suppose that* $(K_1, K_2) \in \mathcal{R}$.
Then there exists a unique $x^* = x^*(\mu, \nu; K_1, K_2) \in (g^{-1}(K_1), K_1)$ *and* $f^* \in \aleph(x^*)$ *such that* $\Upsilon(f^*, x^*, g^* = g(x^*)) = 0$.

Proof. Suppose that $(K_1, K_2) \in \mathcal{R}_1 \cup \mathcal{R}_2 \cup \mathcal{R}_5$. Consider $\Lambda : [g^{-1}(K_1), K_1] \mapsto \mathbb{R}$ defined by (3.6). Note that for this choice of (K_1, K_2), f and g are both continuous on $[g^{-1}(K_1), K_1]$, see Figure 3.6. Hence $\Lambda(x) = \Upsilon(f(x), x, g(x))$ is also continuous. Then the same argument as in the proof of Lemma 3.2.7 shows that there exists a unique $x^* = x^*(\mu, \nu; K_1, K_2) \in (g^{-1}(K_1), K_1)$ such that $\Lambda(x^*) = 0$.

Now suppose $(K_1, K_2) \in \mathcal{R}_3 \cup \mathcal{R}_4$ and consider Λ as before. Recall that Λ is increasing, $\Lambda(g^{-1}(K_1)) < 0$ and $\Lambda(K_1) > 0$. On the other hand, $g^{-1}(K_1) < x''$ and hence Λ has an upward jump at x'' (since f has a downward jump at x''). There are two cases depending on whether $(K_1, K_2) \in \mathcal{R}_3$ or $\mathcal{R}_4$.

1. Suppose that $K_2 > L_u(K_1)$. Then $\Lambda(x''-) > 0$. Moreover, since $\Lambda(g^{-1}(K_1)) < 0$, the continuity of Λ on $(g^{-1}(K_1), x'')$ ensures that there exists a unique scalar $x^* = x^*(\mu, \nu; K_1, K_2) \in (g^{-1}(K_1), x'')$ such that $\Lambda(x^*) = 0$. If $K_2 = L_u(K_1)$ then $\Upsilon(x', x''g(x'')) = 0$ and we take $x^* = x''$, $g^* = g(x'')$ and $f^* = f(x''-) = x'$.

2. Suppose that $K_2 < L_d(K_1)$. Then $\Lambda(x''+) < 0$. Further, since $\Lambda(K_1) > 0$ there exists a unique $x^* = x^*(\mu, \nu; K_1, K_2) \in (x'', K_1)$ such that $\Lambda(x^*) = 0$. If $K_2 = L_d(K_1)$ then $\Upsilon(f', x'', g(x'')) = 0$ and we take $x^* = x''$, $g^* = g(x'')$ and $f^* = f(x''+) = f'$.

$\square$

By Lemma 3.2.13, for $(K_1, K_2) \in \mathcal{R}$ there exists $\{f^* \in \aleph(x^*), x^*, g^* = g(x^*)\}$ such that $\Upsilon(f^*, x^*, g^*) = 0$. Suppose $(K_1, K_2) \in \mathcal{R}_1 \cup \mathcal{R}_4 \cup \mathcal{R}_5$. In this case we let $\hat{M}_{\pi_{x^*}}$ be a canonical (μ, ν)-consistent model (recall that $\hat{M}_\pi$ is the abbreviated notation for the canonical model $(\hat{\mathcal{S}}_\pi, \hat{M})$ for which $\hat{\mathbb{P}}(X \in dx, Y \in dy) = \pi(dx, dy)$). Here $\pi_{x^*} \in \Pi(\mu, \nu)$ is a martingale coupling that is constant on $(-\infty, f^*)$, maps (f^*, x^*) onto (f^*, g^*) and (g^*, ∞) to $(-\infty, f^*) \cup (g^*, \infty)$.

Recall the proof of Theorem 3.2.8. There, to show that $MBEP = HC$, we used the fact that under canonical model $\hat{M}_{\pi_{x^*}}$, π_{x^*} is constant on $(-\infty, f^*)$ and maps (f^*, x^*) onto (f^*, g^*). In fact, the equality $MBEP = HC$ will hold for any canonical model for which the associated martingale coupling has the same property. Then, the mass that is 'unexercised' at time 1 and is in-the-money at time 2 is given by $(\nu - \mu)|_{(-\infty, f^*)}$ where $f^* < e_-$. When $f(x') < e^1_-$ (as is the case when $(K_1, K_2) \in \mathcal{R}_1 \cup \mathcal{R}_4 \cup \mathcal{R}_5$), the same proof applies, so that $MBEP = HC$ and we have optimality. On the other hand, if $(K_1, K_2) \in \mathcal{R}_2 \cup \mathcal{R}_3$, then it is not the case that $f^* < e^1_-$ and thus, in order to specify the optimal model, we need to impose additional structure on the coupling $\tilde{\mu}_{f^*, x^*} \mapsto \tilde{\nu}_{f^*, g^*}$.

Suppose that $(K_1, K_2) \in \mathcal{R}_2 \cup \mathcal{R}_3$. Then $x' < f^*$, so that $(f', x') \cap (f^*, g^*) = \emptyset$. From the defining properties of f' and x' we see that there exists a martingale coupling, which we term $\pi_{x',x^*} \in \Pi(\mu, \nu)$, which is constant on $(-\infty, f')$ and (x', f^*), which maps (f', x') onto itself and (f^*, x^*) onto (f^*, g^*) and which maps (x^*, ∞) to $(-\infty, f') \cup (x', f^*) \cup (g^*, \infty)$.

If $(K_1, K_2) \in \mathcal{R}_1 \cup \mathcal{R}_4 \cup \mathcal{R}_5$ we have the canonical model $\hat{M}_{\pi_{x^*}}$ and in the case when $(K_1, K_2) \in \mathcal{R}_2 \cap \mathcal{R}_3$ we have $\hat{M}_{\pi_{x',x^*}}$. For both models we consider a candidate stopping time $\tau^* = 1$ if $X < x^*$ and $\tau^* = 2$ otherwise, and a candidate superhedge $(\psi^*, \phi^*, \theta^*_{1,2})$ generated by the function ψ^* defined in (3.8).

Theorem 3.2.14. *Suppose Assumption 3.2.10 holds and $(K_1, K_2) \in \mathcal{R}$. Then, depending on whether $(K_1, K_2) \in \mathcal{R}_1 \cup \mathcal{R}_4 \cup \mathcal{R}_5$ or $\mathcal{R}_2 \cup \mathcal{R}_3$, the models $\hat{M}_{\pi_{x^*}}$ and $\hat{M}_{\pi_{x',x^*}}$ and the stopping time τ^* are the consistent models for which the price of the American option is the highest. The function ψ^* defined in (3.8) defines the cheapest superhedge. Moreover, the highest model-based price is equal to the cost of the cheapest superhedge.*

Proof. If $(K_1, K_2) \in \mathcal{R}_1 \cup \mathcal{R}_4 \cup \mathcal{R}_5$ then the proof is essentially the same as the proof of the first case in Theorem 3.2.8. We repeat the main steps for convenience. First find $x^* \in (g^{-1}(K_1), K_1)$ and $f^* \in \aleph(x^*)$ such that $\Upsilon(f^*, x^* g^* = g(x^*)) = 0$. If $x^* = x''$ we find $f^* = f(x''+) = f'$. Under the candidate model $\hat{M}_{\pi_{x^*}}$ mass below f^* at time 1 is mapped to the same point at time 2 (which is possible since $f^* < e^1_-$), and mass in (f^*, x^*) is mapped onto (f^*, g^*), while mass above x^* is either mapped to below f^* or to above g^*. Then under the candidate stopping rule τ^* the model-based expected payoff is equal to the cost of the hedging strategy generated by ψ^*:

$$
\begin{aligned}
MBEP &= \int_{-\infty}^{x^*} (K_1 - w)^+ \mu(dw) + \int_{-\infty}^{f^*} (K_2 - w)^+ (\nu - \mu)(dw) \\
&= P_\mu(x^*) + (K_1 - x^*) P'_\mu(x^*) + D(f^*) + (K_2 - f^*) D'(f^*) \\
&= HC.
\end{aligned}
$$

Now suppose that $(K_1, K_2) \in \mathcal{R}_2 \cup \mathcal{R}_3$. Then, by Lemma 3.2.13, there exists a unique $x^* \in (g^{-1}(K_1), x'']$ and $f^* \in \aleph(x^*)$ such that $\Upsilon(f^*, x^*, g^* = g(x^*)) = 0$. If $x^* = x''$ then we have $f^* = f(x''-) = x'$. Then, since ν is continuous we have that f^*, x^*, g^* solve (2.10) and (2.11). Note, however, that $x' \le f^* < e^2_-$.

Under the candidate model $\hat{M}_{\pi_{x',x^*}}$ mass in (f', x') at time 1 is mapped onto the same interval at time 2. Also, mass below f' and mass in (x', f^*) at time 1 is mapped to the same point at time 2, and mass in (f^*, x^*) is mapped onto (f^*, g^*).

Mass above x^* is either mapped to below f', to (x', f^*), or to above g^*. In particular $(\nu - \mu)|_{(-\infty, f') \cup (x', f^*)}$ is the mass that was not 'exercised' at time 1 and is 'exercised' in-the-money at time 2. In other words, $(\nu - \mu)|_{(-\infty, f') \cup (x', f^*)}$ is the probability under $\hat{M}_{\pi_{x', x^*}}$ that $\{X > x^*, Y < K_2\}$. From (3.13) we have $\int_{x'}^{f'} (K_2 - w)(\nu - \mu)(dw) = 0$. Then

$$
\begin{aligned}
MBEP &= \int_{-\infty}^{x^*} (K_1 - w)\mu(dw) + \int_{-\infty}^{f'} (K_2 - w)(\nu - \mu)(dw) \\
&\qquad\qquad + \int_{x'}^{f^*} (K_2 - w)(\nu - \mu)(dw) \\
&= \int_{-\infty}^{x^*} (K_1 - w)\mu(dw) + \int_{-\infty}^{f^*} (K_2 - w)(\nu - \mu)(dw) \\
&\qquad\qquad - \int_{f'}^{x'} (K_2 - w)(\nu - \mu)(dw) \\
&= \int_{-\infty}^{x^*} (K_1 - w)\mu(dw) + \int_{-\infty}^{f^*} (K_2 - w)(\nu - \mu)(dw) \\
&= P_\mu(x^*) + (K_1 - x^*)P'_\mu(x^*) + D(f^*) + (K_2 - f^*)D'(f^*) \\
&= HC.
\end{aligned}
$$

□

Case $(K_1, K_2) \in \mathcal{B} = \mathcal{B}_1 \cup \mathcal{B}_2 \cup \mathcal{B}_3$

Theorem 3.2.15. *Suppose that Assumption 3.2.10 holds and $(K_1, K_2) \in \mathcal{B}$. Then there is a consistent model for which $\{Y < K_2\} = \{X < K_2\} \cup \{X > K_1, Y < K_2\}$ and, if $x' < K_2$, $\{f' < X < x'\} = \{f' < Y < x'\}$. Then any model with these properties with the stopping rule $\tau = 1$ if $X < K_1$ and $\tau = 2$ otherwise attains the highest consistent model price. The cheapest superhedge is generated from $\psi(x) = (K_2 - x)^+$ and the highest model-based price is equal to the cost of the cheapest hedge.*

Proof. Let $\psi(y) = (K_2 - y)^+$. Now, as in Lemma 3.1.5, define a corresponding ϕ. We find that $\phi(x) = (K_1 - x)^+ - (K_2 - x)^+$ and the superhedging cost (which is the same for all the cases) is

$$HC = P_\nu(K_2) + P_\mu(K_1) - P_\mu(K_2).$$

Suppose $(K_1, K_2) \in \mathcal{B}_1$. Then using the properties of f and g and the left-curtain coupling we see that the proof that the model-based expected payoff is equal to the hedging cost is the same as in the second case of Theorem 3.2.8. In particular,

$$\begin{aligned} MBEP &= \int_{-\infty}^{K_1}(K_1-x)\mu(dx) + \int_{-\infty}^{K_2}(K_2-y)(\nu-\mu)(dy) \\ &= P_\mu(K_1) + P_\nu(K_2) - P_\mu(K_2). \end{aligned}$$

Now suppose $(K_1, K_2) \in \mathcal{B}_2$. Then under the left-curtain coupling mass from (f', x') at time 1 is mapped onto the same interval at time 2. Therefore mass which is below K_2 at time 2 was either below K_2 at time 1, or above x' at time 1. Therefore, we again have

$$MBEP = \int_{-\infty}^{K_1}(K_1-x)\mu(dx) + \int_{-\infty}^{K_2}(K_2-y)(\nu-\mu)(dy).$$

Finally, suppose $(K_1, K_2) \in \mathcal{B}_3$. We again utilise the fact that under the left-curtain coupling, mass from (f', x') at time 1 is mapped onto the same interval at time 2. In both cases, the mass which is below K_2 at time 2 was either below K_2 at time 1, or above K_1 at time 1. In particular, mass that can be 'exercised' at time 2 is given by $(\nu-\mu)|_{(-\infty,f')\cup(x',K_2)}$. Then using $\int_{f'}^{x'}(K_2-z)(\nu-\mu)(dz) = 0$ we again have

$$\begin{aligned} MBEP &= \int_{-\infty}^{K_1}(K_1-x)\mu(dx) + \int_{-\infty}^{f'}(K_2-y)(\nu-\mu)(dy) + \int_{x'}^{K_2}(K_2-y)(\nu-\mu)(dy) \\ &= \int_{-\infty}^{K_1}(K_1-x)\mu(dx) + \int_{-\infty}^{K_2}(K_2-y)(\nu-\mu)(dy), \end{aligned}$$

which ends the proof. $\square$

Case $(K_1, K_2) \in \mathcal{W}$

Suppose $(K_1, K_2) \in \mathcal{W}$. For this case we associate the following superhedge: let $\psi^{x'}$ be given by

$$\psi^{x'}(z) = \begin{cases} (K_2 - z) & z \le f'; \\ (K_2 - f') - (z - f')\frac{K_2 - f'}{x' - f'} & f' < z \le x'; \\ 0 & z > x', \end{cases} \tag{3.15}$$

see Figure 3.8. Since $\psi^{x'}$ is convex and $\psi^{x'}(z) \ge (K_2 - z)^+$, we can use Lemma 3.1.5 to generate a corresponding superhedging strategy $(\psi^{x'}, \phi^{x'}, \theta^{x'}_{1,2})$.

Theorem 3.2.16. *Suppose Assumption 3.2.10 holds and $(K_1, K_2) \in \mathcal{W}$. Then there is a consistent model for which $\{f' < X < x'\} = \{f' < Y < x'\}$ and any* *model with this property with the stopping rule $\tau = 1$ if $X < K_1$ and $\tau = 2$ otherwise attains the highest consistent model price. The cheapest superhedge is generated from $\psi^{x'}$ defined in (3.15) and the highest model-based price is equal to the cost of the cheapest hedge.*

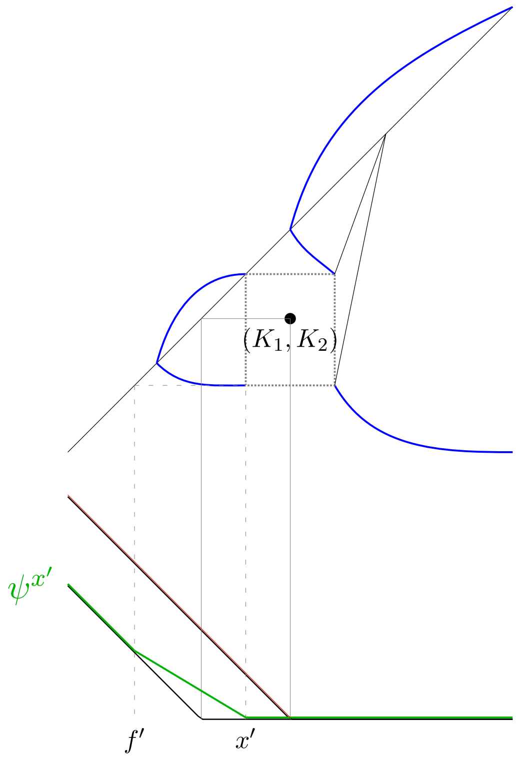

Figure 3.8: Picture of f and g along with superhedge for the blank region $\mathcal{W}$.

Proof. First note that

$$\psi^{x'}(z) = \Theta(x' - z)^+ + (1 - \Theta)(f' - z)^+,$$

where $\Theta = \frac{K_2 - f'}{x' - f'}$. Since $x' < K_1$ we have

$$\begin{aligned}
\phi^{x'}(w) + \psi^{x'}(z) &= (K_1 - w)^+ - \psi^{x'}(w) + \psi^{x'}(z) \\
&= (K_1 - w)^+ + \Theta\big((x' - z)^+ - (x' - w)^+\big) \\
&\qquad + (1 - \Theta)\big((f' - z)^+ - (f' - w)^+\big).
\end{aligned}$$

It follows that $HC = P_\mu(K_1) + \Theta D(x') + (1 - \Theta)D(f')$ is the cost of this strategy (under any consistent model).

Now consider the model-based expected payoff. From (3.13) it follows that

$\mu_{f',x'}$ and $\nu_{f',x'}$ have the same mean and mass, and are in convex order. Moreover, the same holds for $\tilde{\mu}_{f',x'}$ and $\tilde{\nu}_{f',x'}$. Therefore, there exists a martingale coupling, which we term $\pi_{x'} \in \Pi(\mu,\nu)$, which is constant on $(-\infty, f')$ and maps (f', x') onto itself. It follows that under this model the only mass that can be 'exercised' at time 2 is given by $(\nu-\mu)|_{(-\infty,f')}$.

Note that, since f' and x' satisfy (3.13), and hence $\int_{f'}^{x'}(x'-w)(\nu-\mu)(dw) = 0$,

$$\begin{aligned}
D(x') - D(f') &= \int_{-\infty}^{x'} (x'-w)^+(\nu-\mu)(dw) - \int_{-\infty}^{f'} (f'-w)^+(\nu-\mu)(dw) \\
&= \int_{-\infty}^{f'} (x'-f')(\nu-\mu)(dw) + \int_{f'}^{x'} (x'-w)(\nu-\mu)(dw) \\
&= (x'-f') \int_{-\infty}^{f'} (\nu-\mu)(dw).
\end{aligned}$$

Then given that we stop at time 1 if $X < K_1$ and at time 2 otherwise we have

$$\begin{aligned}
MBEP &= \int_{-\infty}^{K_1} (K_1-w)^+\mu(dw) + \int_{-\infty}^{f'} (K_2-w)^+(\nu-\mu)(dw) \\
&= \int_{-\infty}^{K_1} (K_1-w)\mu(dw) + \int_{-\infty}^{f'} (f'-w)(\nu-\mu)(dw) \\
&\qquad\qquad + (K_2-f') \int_{-\infty}^{f'} (\nu-\mu)(dw) \\
&= P_\mu(K_1) + D(f') + \Theta\big(D(x') - D(f')\big) \\
&= P_\mu(K_1) + \Theta D(x') + (1-\Theta)D(f') = HC
\end{aligned}$$

as required. □

Case $(K_1, K_2) \in \mathcal{G}$

Recall the construction of L_u and L_d. For $K_1 \in (x'', g(x''))$ and $K_2 \in (L_d(K_1), L_u(K_1))$ there does not exist $x^* \in (g^{-1}(K_1), K_1)$ such that $\Lambda(x^*) = 0$; instead we have that $\Lambda(x''-) < 0 < \Lambda(x''+)$. On the other hand, from (3.14) we have that there exists a martingale coupling of $\mu_{x',x''}$ and $\nu_{x',g(x'')}$. Moreover, note that the restrictions of $\tilde{\mu}_{f',x'}$ to (x', x'') and $\tilde{\nu}_{f',x'}$ to $(x', g(x''))$ are equal to $\mu_{x',x''}$ and $\nu_{x',g(x'')}$, respectively. Then we define a martingale coupling $\pi_{x',x''} \in \Pi(\mu,\nu)$ which is constant on $(-\infty, f')$, maps (f', x') onto itself, (x', x'') onto $(x', g(x''))$ and (x'', ∞) to $(-\infty, f') \cup (g(x''), \infty)$. Let $\hat{M}_{\pi_{x',x''}}$ be the canonical model under which $\hat{\mathbb{P}}(X \in dx, Y \in dy) = \pi_{x',x''}(dx, dy)$. Note that the model $\hat{M}_{\pi_{x',x''}}$ is a refinement of $\hat{M}_{\pi_{x'}}$ used in the proof of Theorem

3.2.16.

Given x', and thus also x'', we define the superhedge as follows. First define linear functions $\Delta_1 : [f', x'] \mapsto \mathbb{R}$ and $\Delta_2 : [x', g(x'')] \mapsto \mathbb{R}$ by

$$\Delta_1(x) = (K_2 - f') - (x - f')\frac{(K_2 - f') - \Delta_2(x')}{x' - f'}; \qquad \Delta_2(x) = \big(g(x'') - x\big)\frac{K_1 - x''}{g(x'') - x''}.$$

Then $\Delta_1(f') = (K_2 - f')$, $\Delta_1(x') = \Delta_2(x')$, $\Delta_2(x'') = K_1 - x''$ and $\Delta_2(g(x'')) = 0$. Moreover, direct calculation shows that $-1 < \Delta_1'(x) < \Delta_2'(x) < 0$. Now define a function $\psi^{x',x''}$ by

$$\psi^{x',x''}(z) = \begin{cases} (K_2 - z) & z \leq f'; \\ \Delta_1(z) & f' < z \leq x'; \\ \Delta_2(z) & x' < z \leq g(x''); \\ 0 & z > g(x''). \end{cases} \tag{3.16}$$

By construction $\psi^{x',x''}$ is convex and $\psi^{x',x''}(z) \geq (K_2 - z)^+$ (see Figure 3.9), and thus by Lemma 3.1.5 it can be used to construct a superhedge $(\psi^{x',x''}, \phi^{x',x''}, \theta_{1,2}^{x',x''})$.

Theorem 3.2.17. *Suppose Assumption 3.2.10 holds and $(K_1, K_2) \in \mathcal{G}$. The model $\hat{M}_{\pi_{x',x''}}$ and the stopping time $\tau = 1$ if $X < x''$ and $\tau = 2$ otherwise attains the highest consistent model price. Moreover, $\psi^{x',x''}$ defined in* (3.16) *generates the cheapest superhedge and the highest model-based price is equal to the cost of the cheapest superhedge.*

Proof. The candidate canonical model is associated with the martingale coupling $\pi_{x',x''}$ which is constant on $(-\infty, f')$, maps (f', x') onto itself, maps (x', x'') onto $(x', g(x''))$, and (x'', ∞) to $(-\infty, f') \cup (g(x''), \infty)$. Then under the candidate stopping time (exercise at time 1 if $X < x''$ and at time 2 otherwise) we have that the law of Y (under $\hat{M}_{\pi_{x',x''}}$), on the event that the option was not exercised at time 1, is given by $(\nu - \mu)|_{(-\infty,f')} + \nu|_{(g(x''),\infty)}$. Therefore

$$\begin{aligned} MBEP &= \int_{-\infty}^{x''} (K_1 - w)^+ \mu(dw) + \int_{-\infty}^{f'} (K_2 - w)^+ (\nu - \mu)(dw) \\ &= P_\mu(x'') + (K_1 - x'')P_\mu'(x'') + D(f') + (K_2 - f')D'(f'). \end{aligned}$$

Now consider the hedging cost generated by $\psi^{x',x''}$. Let $\Theta_1 = \frac{K_2 - f' - \Delta_2(x')}{x' - f'} = -\Delta_1'$ and $\Theta_2 = \frac{K_1 - x''}{g(x'') - x''} = -\Delta_2'$. Note that we can rewrite (3.16) as

$$\psi^{x',x''}(z) = \Theta_2\big(g(x'') - z\big)^+ + (\Theta_1 - \Theta_2)(x' - z)^+ + (1 - \Theta_1)(f' - z)^+.$$

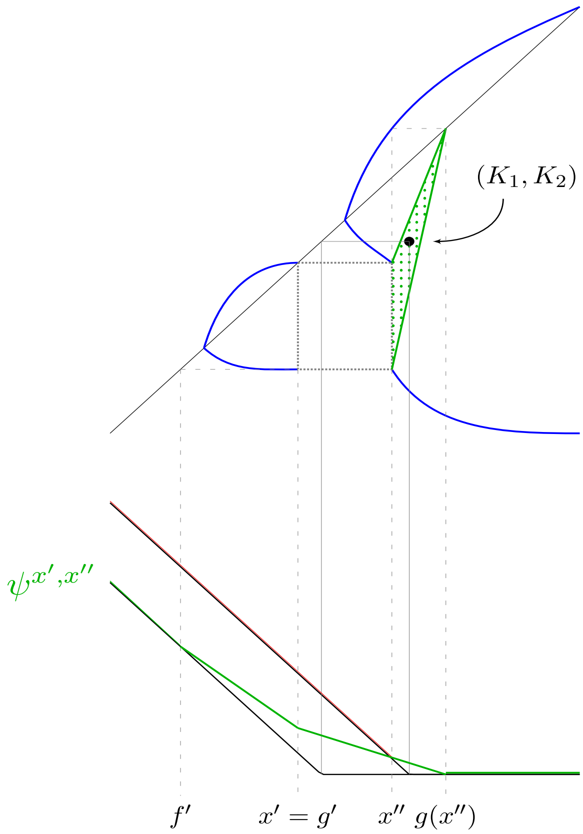

Figure 3.9: Picture of f and g along with superhedge for the dotted region $\mathcal{G}$. The hedge function $\psi^{x',x''}$ has a kink at x'.

Then

$$\phi(z) = (1 - \Theta_1)\big((x' - z)^+ - (f' - z)^+\big) + (1 - \Theta_2)\big((x'' - z)^+ - (x' - z)^+\big),$$

and thus the hedging cost is

$$\begin{aligned} HC &= \Theta_2 P_\nu\big(g(x'')\big) + (1 - \Theta_1)D(f') + (1 - \Theta_2)P_\mu(x'') + (\Theta_1 - \Theta_2)D(x') \\ &= P_\mu(x'') + D(f') + \Theta_1\big(D(x') - D(f')\big) \\ &\qquad\qquad + \Theta_2\Big(P_\nu\big(g(x'')\big) - P_\nu(x') - P_\mu(x'') + P_\mu(x')\Big). \end{aligned}$$

Now using (3.13) and the fact that $g(x') = x'$ we have that $D'(f') = D'(x')$ and $f'D'(f') - D(f') = x'D'(x') - D(x')$. Hence

$$\Theta_1\big(D(x') - D(f')\big) = (K_2 - f')D'(f') - \Delta_2(x')D'(f'); \tag{3.17}$$

moreover, (3.14) gives that

$$\begin{aligned}
\Theta_2\Big(P_\nu\big(g(x'')\big) - P_\nu(x') - P_\mu(x'') + P_\mu(x')\Big) \\
&= \Theta_2\Big(g(x'')P'_\nu\big(g(x'')\big) - x''P'_\mu(x'') - x'D'(x')\Big) \\
&= \Theta_2\Big(\big(g(x'') - x''\big)P'_\mu(x'') + \big(g(x'') - x'\big)D'(x')\Big) \\
&= (K_1 - x'')P'_\mu(x'') + \Delta_2(x')D'(f'). \qquad (3.18)
\end{aligned}$$

Then, combining (3.17) and (3.18) we conclude that $HC = MBEP$. $\square$

Case $K_1 > r_\mu$

In Lemma 3.2.5, and under the Dispersion Assumption, we constructed f and g but only on the interval $(e_-, r_\mu]$. More generally, when μ is continuous the arguments of Beiglböck and Juillet [10] and Henry-Labordère and Touzi [50] allow us to construct $T_d = f$ and $T_u = g$ on $[\ell_\mu, r_\mu]$ for arbitrary laws $\mu \leq_{cx} \nu$. For their purposes the definitions of f and g outside the range of μ are not important since they have no impact on the construction of the left-curtain martingale coupling.

Nonetheless, we can extend the definitions of f and g to $\mathbb{R}$ in a way which respects the conditions in Lemma 2.3.2, by setting

$$\begin{aligned}
f(x) = x = g(x), &\qquad -\infty < x \leq \ell_\mu; \\
f(x) = \ell_\nu, \qquad g(x) = r_\nu, &\qquad r_\mu < x < r_\nu; \\
f(x) = x = g(x), &\qquad r_\nu \leq x < \infty.
\end{aligned}$$

We will show that with these definitions for f and g the analysis of the previous sections extends to the case $K_1 > r_\mu$.

Suppose that $r_\nu > r_\mu$ and $r_\mu < K_1 < r_\nu$. Then $\Lambda(r_\mu) = \frac{r_\nu - K_1}{r_\nu - r_\mu} - \frac{(K_1 - K_2)}{r_\mu - \ell_\nu}$ and $\Lambda(r_\nu -) = \infty$. If $\Lambda(r_\mu) \geq 0$ and Λ is continuous then there exists $x^* \in [\ell_\mu, r_\mu]$ such that $g(x^*) > x^*$ and $\Lambda(x^*) = 0$. Then, exactly as in Section 3.2.2 we can construct a model, stopping time and superhedge such that the model-based expected payoff equals the hedging cost, and hence the model, stopping time and hedge are all optimal. The model could be based on the left-curtain coupling, and the optimal exercise rule is to exercise the American put at time 1 if $X < x^*$. Even if Λ is not continuous, there may exist x^* such that $\Lambda(x^*) = 0$ and the same arguments apply (see Section 3.2.3). If not, then we are in the setting of Section 3.2.3, but again we can identify the optimal model and hedge. Essentially, the case $\Lambda(r_\mu) \geq 0$ is covered by a direct extension of existing arguments. Note that $\Lambda(r_\mu) \geq 0$ is equivalent to

$$K_2 \geq K_1 - \frac{(r_\mu - \ell_\nu)(r_\nu - K_1)}{r_\nu - r_\mu}.$$

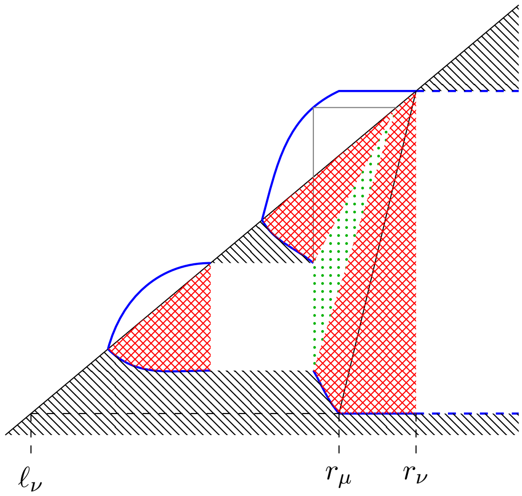

Figure 3.10: The various cases for $K_1 > r_\nu$ in the setting of Section 3.2.3.

Now suppose $r_\mu < K_1 < r_\nu$ and $K_2 < K_1 - \frac{(r_\mu - \ell_\nu)(r_\nu - K_1)}{r_\nu - r_\mu}$. Then $\Lambda(r_\mu) < 0$ and since $\Lambda(r_\nu -) = \infty$ and Λ is continuous on $[r_\mu, r_\nu]$ (note that we have defined f and g to be constants on this range) there must exist $x^* \in (r_\mu, K_1)$ such that $\Lambda(x^*) = 0$. It is always optimal to exercise at time 1 and any martingale coupling can be used to generate a model which attains the highest model based price of $P_\mu(K_1) = (K_1 - \overline{\mu})$. A cheapest superhedge is generated by

$$\psi(y) = \frac{K_2 - \ell_\nu}{r_\nu - \ell_\nu}(r_\nu - y)^+ + \frac{r_\nu - K_2}{r_\nu - \ell_\nu}(\ell_\nu - y)^+. \tag{3.19}$$

The cost of this hedge is

$$\begin{aligned} &\frac{K_2 - \ell_\nu}{r_\nu - \ell_\nu} P_\nu(r_\nu) + \frac{r_\nu - K_2}{r_\nu - \ell_\nu} P_\nu(\ell_\nu) + P_\mu(K_1) - \frac{K_2 - \ell_\nu}{r_\nu - \ell_\nu} P_\mu(r_\nu) - \frac{r_\nu - K_2}{r_\nu - \ell_\nu} P_\mu(\ell_\nu) \\ &= \frac{K_2 - \ell_\nu}{r_\nu - \ell_\nu}(r_\nu - \bar{\mu}) + (K_1 - \bar{\mu}) - \frac{K_2 - \ell_\nu}{r_\nu - \ell_\nu}(r_\nu - \bar{\mu}) = (K_1 - \bar{\mu}). \end{aligned}$$

Finally suppose $K_1 > r_\nu$. Then $Y < K_1$ almost surely under any consistent model and

$$\mathbb{E}[(K_2 - Y)^+ | \mathcal{F}_1] \leq \mathbb{E}[(K_1 - Y)^+ | \mathcal{F}_1] = \mathbb{E}[(K_1 - Y) | \mathcal{F}_1] = (K_1 - X).$$

Therefore, it is always optimal to exercise the American put at time 1. If $K_2 > r_\nu$ or $K_2 < \ell_\nu$ then we are in the case studied in Section 3.2.3 and the cheapest hedge is generated by a time 2 payoff $\psi(y) = (K_2 - y)^+$. If $K_2 \in [\ell_\nu, r_\nu]$ then we are in the case studied in Section 3.2.3 and the cheapest superhedge is generated by $\psi = \psi(y)$ where ψ is given by (3.19). In either case the highest model-based expected payoff is $P_\mu(K_1) = (K_1 - \bar{\mu})$ and this is also the cost of the superhedge.

3.2.4 Intervals where ν has no mass, or $\nu = \mu$.

The definition of the left-curtain martingale coupling (recall Lemma 3.2.5) only requires that $g = T_u$ is increasing, and not that it is continuous. In general g may have jumps; such jumps occur when there is an interval on which ν places no mass.

If g has a jump then we need to adapt the superhedge. Suppose g has a jump at $\hat{x}$ (which has to be upwards since g is increasing) and f is continuous at $\hat{x}$. Suppose further that K_1 is such that $\hat{x} \in (g^{-1}(K_1), K_1)$. Then as before, we would like to find $x^* \in (g^{-1}(K_1), K_1)$ such that $\Lambda(x^*) = 0$. Recall that Λ is increasing and suppose $\Lambda(g^{-1}(K_1)) < 0 < \Lambda(K_1)$. If $\Lambda(\hat{x}-) < 0$ and $\Lambda(\hat{x}+) > 0$, then there will be no solution to $\Lambda = 0$. However, by keeping $x = \hat{x}$, $\hat{f} = f(\hat{x})$ fixed in (3.6), and varying g only, we can find $\hat{g} \in (g(\hat{x}-), g(\hat{x}+))$, such that $(\hat{g} - K_1)/(\hat{g} - \hat{x}) = (K_1 - K_2)/(\hat{x} - \hat{f})$ so that $\Upsilon(f(\hat{x}), \hat{x}, \hat{g}) = 0$. Then, the candidate (and indeed optimal) superhedging strategy is generated by ψ^*, given in (3.8), with $(f^*, x^*, g^*) = (\hat{f}, \hat{x}, \hat{g})$, see Figure 3.11. Moreover, since ν does not charge $(g(\hat{x}-), g(\hat{x}+))$, the triple $(\hat{f}, \hat{x}, \hat{g})$ solves the mass and mean equations (2.10) and (2.11). The strong duality between the model-based expected payoff and the hedging cost follows as before.

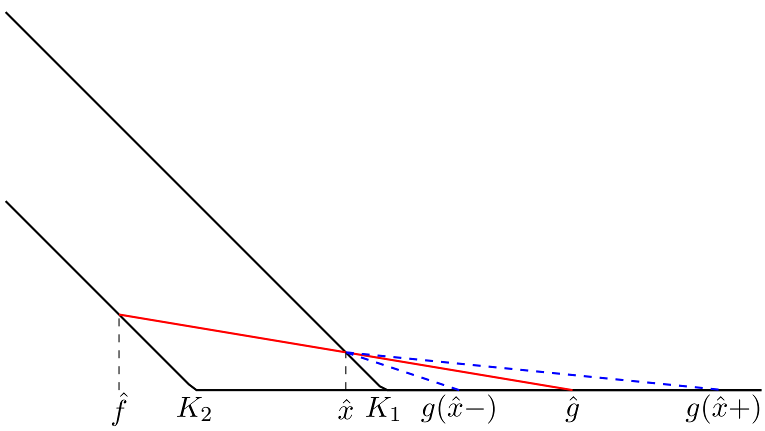

Figure 3.11: Sketch of put payoffs with points $\hat{x}$, $\hat{f}$ and $\hat{g}$ marked.

Alternatively, suppose f has a downward jump at $\bar{x}$. This can happen if $\nu = \mu$ on $(f(\bar{x}+), f(\bar{x}-))$. Suppose that K_1 is such that $\bar{x} \in (g^{-1}(K_1), K_1)$ and $\Lambda(\bar{x}-) < 0$ and $\Lambda(\bar{x}+) > 0$, so that again we cannot find $x \in (g^{-1}(K_1), K_1)$ with $\Lambda(x) = 0$. We can deal with this similarly as in the case of discontinuity in g: choose $\bar{f} \in (f(\bar{x}+), f(\bar{x}-))$ such that $\Upsilon(\bar{f}, \bar{x}, g(\bar{x})) = 0$, then consider a hedging strategy generated by ψ^* with $(f^*, x^*, g^*) = (\bar{f}, \bar{x}, g(\bar{x}))$. Note that $\mu = \nu$ on $(f(\bar{x}+), f(\bar{x}-))$ and so if (2.10) and (2.11) hold for some $f \in [f(\bar{x}+), f(\bar{x}-)]$ (with $\bar{x}, \bar{g}$) then they hold for all f in this interval. It follows that we can construct a coupling in which $(\bar{f}, \bar{x})$ is mapped to $(\bar{f}, \bar{g})$ and strong duality holds.

In the case of f and g jumping simultaneously, we have a pictorial representation of the regions of pairs (K_1, K_2) which lead to a hedging strategy which has to be adapted as above, see Figure 3.13. If g has a jump at $\hat{x}$, then $\Lambda(\hat{x}-) < 0$ and $\Lambda(\hat{x}+) > 0$ is equivalent to point (K_1, K_2) lying in the interior of a triangle with vertices $\{(g(\hat{x}-), g(\hat{x}-)), (g(\hat{x}+), g(\hat{x}+)), (\hat{x}, f(\hat{x}))\}$. On the other hand, if f jumps downwards at $\bar{x}$, then $\Lambda(\bar{x}-) < 0$ and $\Lambda(\bar{x}+) > 0$ is equivalent to point K_1, K_2 lying in the interior of a triangle with vertices $\{(\bar{x}, f(\bar{x}-)), (\bar{x}, f(\bar{x}+)), (g(\bar{x}), g(\bar{x}))\}$ (compare this with a region $\mathcal{G}$).

Exceptionally we may have simultaneous jumps in g and f at $\check{x}$. Then the set of (K_1, K_2) for which these arguments are needed is a quadrilateral with vertices $(\check{x}, f(\check{x}-))$, $(\check{x}, f(\check{x}+))$, $(g(\check{x}+), g(\check{x}+))$ and $(g(\check{x}-), g(\check{x}-))$. In particular, then there are multiple pairs $(\check{f}, \check{g})$ with $\check{f} \in (f(\check{x}+), f(\check{x}-))$ and $\check{g} \in (g(\check{x}-), g(\check{x}+))$ such that $\Upsilon(\check{f}, \check{x}, \check{g}) = 0$, so that an optimal hedging strategy is not unique.

3.2.5 The general case for continuous ν

In the previous sections we showed how the left-curtain coupling can be used to find an optimal model, exercise strategy and a superhedge, under the assumption that both μ and ν are continuous together with further regularity and simplifying assumptions which we labelled the Dispersion Assumption and the Single Jump Assumption. Under the latter assumption, the existence of points that solve (3.13) led us to identify two further types of hedging strategy that were not present under the dispersion assumption, making four in total.

If we relax the assumptions further and require only that both μ and ν are continuous, then we expect that in some cases there may exist multiple pairs (f_i', x_i'), $i = 1, 2, 3, ...$, that solve (3.13). Note that from the monotonicity of g we can write $\{x : g(x) > x\}$ as a countable union of intervals, and on each such interval f is decreasing. f jumps over the intervals (f_i', x_i') identified above (at least those with x' to the left of the current value of x). In particular, f has only countably many downward jumps. Figure 2.1 is a stylised representation of the general left-curtain martingale coupling, not least because in the figure f has only finitely many jumps. Starting from Figure 2.1 and using the constructions in Section 3.2.3 we can divide $(K_1, K_2 < K_1)$ into four regions, see Figure 3.12. They key point is that these four regions are characterised exactly as in the cases described in Section 3.2.3. For given (K_1, K_2) we can determine which of the types of hedging strategy is a candidate optimal superhedge, and determine a candidate optimal stopping rule. (We can always use the model associated with the left-curtain martingale coupling π_{lc}.) The fact that these candidates are indeed optimal can be proved using exactly analogous techniques to those used in Section 3.2.3.

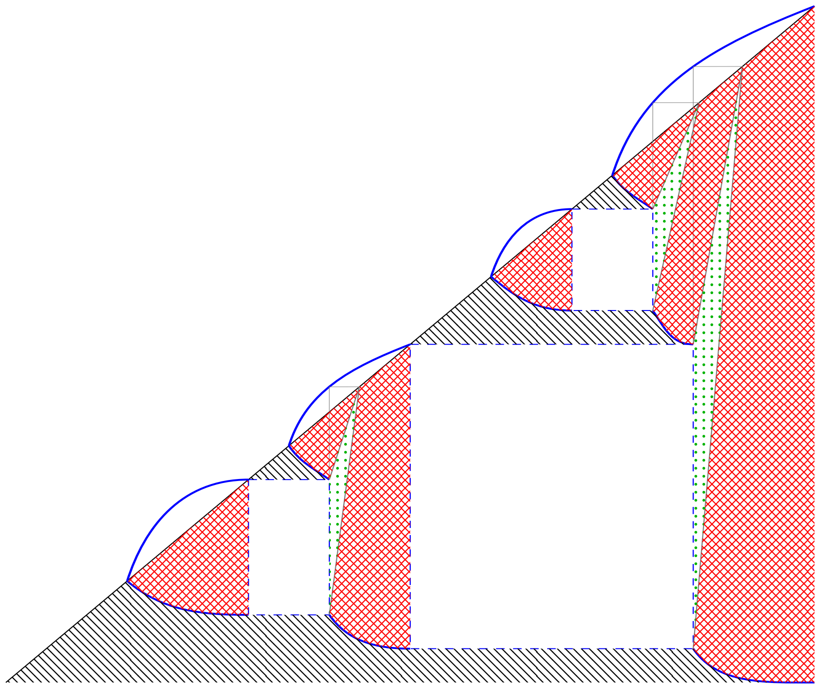

Figure 3.12: General picture of f, g with shading of regions. There remain 4 types of shading corresponding to 4 forms of optimal hedge.

More specifically, we can divide $\{(k_1, k_2) : k_2 < k_1\}$ into two disjoint regions, $\{(k_1, k_2) : k_2 \leq f(k_1)\}$ and $\{(k_1, k_2) : f(k_1) < k_2 < k_1\}$. We can divide the former into two further regions $\mathcal{W} = \{(k_1, k_2) : k_2 < k_1, \exists x \leq k_1 \text{such that } f(x) < k_2 < g(x)\}$ and $\mathcal{B} = \{(k_1, k_2) : k_2 \leq f(k_1)\} \setminus \mathcal{W}$. The latter we again divide into two regions $\mathcal{G}$ and $\mathcal{R} = \{(k_1, k_2) : f(k_1) < k_2 < k_1\} \setminus \mathcal{G}$. Here we can write $\mathcal{G} = \cup_{x: f(x-) > f(x+)} \Delta(x)$ where $\Delta(x)$ is a triangle with vertices $(x, f(x+))$, $(x, f(x-))$ and $(g(x), g(x))$. Then on each of the regions $\mathcal{W}$, $\mathcal{B}$, $\mathcal{G}$ and $\mathcal{R}$ we have a superhedge exactly as described in Section 3.2.3. Moreover, again by the arguments of Section 3.2.3, we can show that the hedging cost associated with the super-hedging strategy is precisely the model-based expected payoff of the American put under the martingale coupling π_{lc} (and candidate stopping rule) thus proving the optimality of the hedge and of the model/exercise rule.

For example, suppose $(K_1, K_2) \in \mathcal{W}$. (The cases for $(K_1, K_2) \in \mathcal{R} \cup \mathcal{B}$ are generally even simpler, and for $(K_1, K_2) \in \mathcal{G}$ the story is roughly equally involved.) Recall that under Single Jump Assumption 3.2.10, in order to show that $MBEP = HC$, we used the existence of $\bar{x}$ and $\bar{f}$ satisfying (3.13) together with the fact that

$$\int_y \int_{x>K_1} (K_2 - y)^+ \pi_{lc}(dx, dy) = \int_{-\infty}^{\bar{f}} (K_2 - y)(\nu - \mu)(dy). \tag{3.20}$$

Then for general probability measures μ and ν, provided we can find $\bar{x}$, $\bar{f}$ satisfying (3.13) and (3.20), the proof that $MBEP = HC$ and hence of optimality follows exactly as in Theorem 3.2.16.

Lemma 3.2.18. *Suppose* $(K_1, K_2) \in \mathcal{W}$. *Then there exists* $\bar{x}, \bar{f}$ *such that*

$$\bar{f} < K_2 < \bar{x} \leq K_1, \qquad \int_{\bar{f}}^{\bar{x}} z^i \mu(dz) = \int_{\bar{f}}^{\bar{x}} z^i \nu(dz), \quad i \in \{1, 2\}, \tag{3.21}$$

and such that under the left-curtain coupling

$$\{X > K_1, Y \leq K_2\} = \{X > K_1, Y \leq \bar{f}\} = \{Y \leq \bar{f}) \setminus (X \leq \bar{f}\},$$

so that (3.20) *holds.*

Proof. Define $\mathcal{X} = \mathcal{X}_{K_1,K_2} = \{x : x \leq K_1, f(x) < K_2 < g(x)\}$. Since $(K_1, K_2) \in \mathcal{W}$, $\mathcal{X}_{K_1,K_2}$ is non-empty. Define $\hat{x} = \sup\{x : x \in \mathcal{X}\}$. We show that $\hat{x}$ and a suitably defined $\hat{f}$ are such that (3.20) and (3.21) hold.

First suppose that $\hat{x} < K_1$. Suppose further that $g(\hat{x}) > \hat{x}$. Take $\tilde{x} \in (\hat{x}, g(\hat{x}) \wedge K_1)$. Then $g(\tilde{x}) \geq g(\hat{x}) > \tilde{x}$. Also $f(\tilde{x}) \notin (f(\hat{x}), g(\hat{x}))$ and if $f(\tilde{x}) \geq g(\hat{x})$ then we have that $f(\tilde{x}) \geq g(\hat{x}) > \tilde{x}$, which is a contradiction. Hence $f(\tilde{x}) \leq f(\hat{x}) < K_2 < g(\hat{x}) \leq g(\tilde{x})$. Then $\tilde{x} \in \mathcal{X}$ contradicting the maximality of $\hat{x}$. Hence $g(\hat{x}) \leq \hat{x}$ (and thus $g(\hat{x}) = \hat{x}$). But then $f(\hat{x}) = \hat{x}$ and $\hat{x} \notin \mathcal{X}$.

Hence there exists $(x_n)_{n\geq 1}$ such that $x_n \in \mathcal{X}$ and $x_n \uparrow \hat{x}$. Let $g(\hat{x}-) = \lim g(x_n)$. By the same argument as above we cannot have $g(\hat{x}-) > \hat{x}$. Hence $\hat{x} = g(\hat{x}-) > K_2$.

Now suppose $\hat{x} = K_1 > K_2$. Then $K_1 \notin \mathcal{X}$ since we cannot have both $K_2 \leq f(K_1)$ and $f(K_1) < K_2 < g(K_1)$. Hence there exists $(x_n)_{n\geq 1}$ such that $x_n \in \mathcal{X}$ and $x_n \uparrow \hat{x}$. Let $g_n = g(x_n)$ and $f_n = f(x_n)$. If $g(K_1) > K_1$ then there exists n_0, such that for all $n \geq n_0$, $g_n > K_1$. Then $f(K_1) \notin (f_n, g_n)$ and therefore $f(K_1) \leq f_n < K_2$ contradicting $K_2 \leq f(K_1)$. Hence $g(\hat{x}-) = g(K_1-) = K_1$.

In either case $\hat{x} \notin \mathcal{X}$ and there exists $(x_n)_{n\geq 1}$ such that $x_n \in \mathcal{X}$, $x_n \uparrow \hat{x}$ and $(f_n)_{n\geq 1}$ is a decreasing sequence while $(g_n)_{n\geq 1}$ is such that $g_n \uparrow \hat{x}$. Let $\hat{f} = \lim_{n\to\infty} f_n$. Then

$$\int_{f_n}^{x_n} z^i \mu(dz) = \int_{f_n}^{g_n} z^i \nu(dz), \quad i \in \{1, 2\}$$

and by taking limits we have that $\hat{x}$ and $\hat{f}$ solve (3.21). Note also that $\hat{x} > K_2$.

We are left to show that $\hat{x}$ and $\hat{f}$ solve (3.20). This follows from the fact that $\hat{f} < K_2$, together with the set identifications

$$\{X > K_1, Y \leq K_2\} = \{X > K_1, Y \leq \hat{f}\} = \{X > \hat{f}, Y \leq \hat{f}\} = \{Y \leq \bar{f}\} \setminus \{X \leq \bar{f}\}.$$

□

Remark 3.2.19. *The set $\{x : g(x) > x\}$ is a collection of intervals and we let I_+ denote the set of right-endpoints of these intervals. As remarked above, Figure 3.12 is drawn in the case of 'finite complexity' in the sense that the set I_+ contains a finite number of elements. The results extend easily to countable I_+ provided I_+ contains no accumulation points.*

In general I_+ may contain an accumulation point, and, as discussed in Henry-Labordère and Touzi [50], care is needed in the construction of the left-curtain mappings (T_d, T_u) in this case. However, from our perspective such subtleties do not cause a problem. The reason for this is we do not aim to derive the left-curtain coupling, but rather take the left-curtain coupling as a given, and use it to solve the put pricing problem.

Our construction of the best model and the cheapest hedge is local in the sense that when in Figure 3.12 we look at in which region the point (K_1, K_2) lies, the fine detail of the picture in other parts of (k_1, k_2)-space is not important. So, the existence of accumulation points can only be an issue if K_1 is equal to one of those accumulation points.

Let x_∞ be such an accumulation point in I_+ and suppose $K_1 = x_\infty$. Depending on the value of K_2 then either there exists (x', f') with $f' < K_2 < x'$ such that (3.13) *holds or not. In the former case we can follow the analysis of Section 3.2.3, and in the latter Section 3.2.3: in either case we construct a model and hedge such that the model price and hedging cost agree, thus proving optimality of both.*

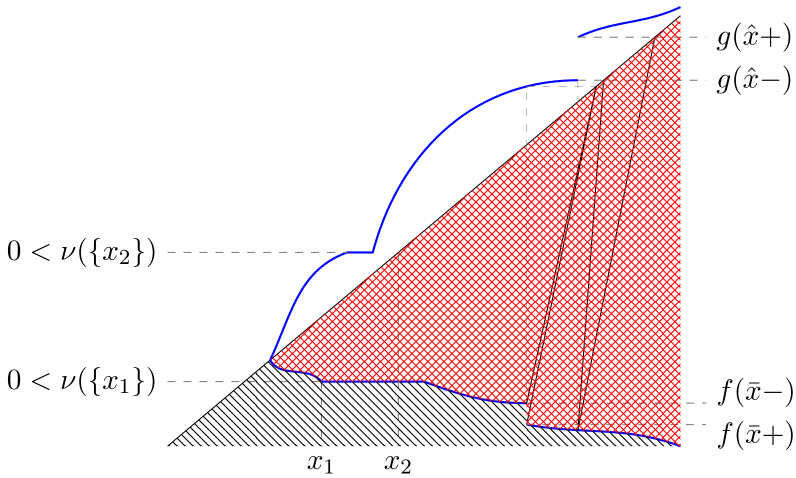

Figure 3.13: Atoms of ν correspond to flat sections in f and g. Regions of no mass of ν correspond to jumps of f and g.

3.2.6 Atoms in the target law

When ν has atoms, the preservation of mass and mean conditions become (2.12) and (2.13), respectively. In particular, atoms of ν correspond to the flat sections in f or g. See Figure 3.13. In this case we still can find all the optimal quantities as before. In particular, $\Lambda(x) := \frac{g(x)-K_1}{g(x)-x} - \frac{(K_1-K_2)}{x-f(x)}$ is strictly increasing in x, even if f and/or g is constant. Hence we can find solutions to $\Lambda = 0$ (more generally solutions $x, f \in \aleph(x)$ to $\Upsilon(f, x, g = g(x)) = 0$) exactly as before. The superhedge is unchanged. A little care is needed in constructing the optimal model, but under the associated martingale coupling mass in $(f(x^*), x^*)$ is mapped onto $(f(x^*), g(x^*))$ together with (potentially) atoms at $f(x^*)$ or $g(x^*)$. Specifically, given f^*, x^*, g^* we can find λ^*_f and λ^*_g such that (2.12) and (2.13) hold. Then, in any optimal canonical model $\hat{M}_\pi$, π is constant on $(-\infty, f^*)$, and the law of $\hat{M}_2$ on the event $\hat{M}_1 \in (f^*, x^*)$ is ν_{x^*} which is defined to be $\nu_{x^*} = \nu|_{(f^*,g^*)} + \lambda^*_{f^*}\delta_{f^*} + \lambda^*_{g^*}\delta_{g^*}$. We also find the law of $\hat{M}_2$ on the event $\hat{M}_1 \in (x^*, \infty)$ is $\nu - \nu_{x^*} - \mu|_{(-\infty,f^*)}$.

3.3 Discussion and extensions

3.3.1 The role of the left-curtain coupling

For any pair of strikes (K_1, K_2) the left-curtain model attains the highest expected payoff for the American put. However, although it optimises simultaneously across all pairs of strikes it is not (in general) optimal for linear combinations of American puts. For example, if we consider a generalised American option with payoff a if exercised at time 1 and b if exercised at time 2, where $a(x) = \sum_{j=1}^{J}(K_1^j - x)^+$ and $b(y) = \sum_{j=1}^{J}(K_2^j - y)^+$ (with $K_2^j \leq K_1^j$ for each j), then the model associated with the left-curtain coupling is typically not optimal. The reason is that a model $(\mathcal{S}, M)$ is only optimal when it is combined with the best stopping rule, and the optimal stopping rule *does* depend on (K_1, K_2).

Conversely, although the model associated with the left-curtain coupling is optimal (simultaneously across all pairs K_1, K_2), we do not need the full power of this coupling when we work with fixed (K_1, K_2). In the dispersion assumption case all we need is a coupling in which $(f(x^*), x^*)$ is mapped onto $(f(x^*), g(x^*))$ where x^* is such that $\Lambda(x^*) = 0$, and $(-\infty, f^*)$ is mapped to itself, but not necessarily in a constant fashion. There are many martingale couplings which have this property.

The intuition behind the optimality of the left-curtain coupling is as follows. With American puts there is a tension between the time-decay of the option payout promoting early exercise, and the convexity of the payoff function promoting delay. If the aim is to maximise the payoff of the option then any paths which are in-the-money at time 1, and will remain in-the-money, are best exercised at time 1. However, once a path has been exercised, any further volatility is irrelevant. In particular, when designing a candidate optimal model we should try to keep paths which are exercised at time 1 constant (or near constant) whenever possible. Thus the probability space should be split into two regions: one region where the put is in-the-money at time 1 and is exercised, and thereafter paths move little, and a second region where the put is out-of-the-money at time 1 (and sometimes just in-the-money, but left unexercised at time 1) and then paths move a long way between times 1 and 2. The left-curtain coupling has this property.

3.3.2 Multiple exercise times

It is natural to ask if it is possible to extend the analysis to American puts which can be exercised at multiple dates $(T_1, T_2, \dots T_N)$ where $N > 2$, or equivalently to martingales $M = (M_n)_{0 \leq n \leq N}$ with marginals (μ_n) where μ_1 has mean $M_0 = \bar{\mu}$ and $\mu_n \leq_{cx} \mu_{n+1}$ for $1 \leq n \leq N-1$. It is clear that many of the ideas extend naturally to the multi-marginal case. However, the number of types of hedging strategy may grow exponentially with N. This is left as future work.

3.3.3 General convex payoffs

We now discuss the generalisation of problem (3.4). Let $a, b : \mathbb{R} \mapsto \mathbb{R}$ and consider the problem of finding the highest model-independent upper bound of the American option, that pays $a(x)$ if exercised at time 1 and $b(y)$ if exercised at time 2:

$$\tilde{\mathcal{P}} := \sup_{\pi\in\Pi(\mu,\nu)} \sup_{B\in\mathcal{B}(\mathbb{R})} \mathbb{E}^{\mathcal{L}(X,Y)\sim\pi}\big[a(X)1_{\{X\in B\}} + b(Y)1_{\{X\notin B\}}\big]. \tag{3.22}$$

Consider the optimal stopping problem (with respect to a fixed canonical (μ,ν)-consistent model) given by the inner supremum in (3.22). A very ambitious task would be to find a set $B^* \in \mathcal{B}(\mathbb{R})$ that is optimal for all $\pi \in \Pi(\mu,\nu)$, i.e. $\tilde{\mathcal{P}} = \mathbb{E}^{\mathcal{L}(X,Y)\sim\pi}[a(X)1_{\{X\in B^*\}} + b(Y)1_{\{X\notin B^*\}}]$, for all $\pi \in \Pi(\mu,\nu)$. We expect this to be true in situations when the expected payoff does not depend on the coupling π (e.g. when one of the functions a, b are equal to zero).

On the other hand, observe that in the canonical setting the filtration does not depend on a particular choice of a martingale coupling $\pi \in \Pi(\mu,\nu)$. Therefore we can interchange two supremums in (3.22):

$$\tilde{\mathcal{P}} = \sup_{B\in\mathcal{B}(\mathbb{R})} \sup_{\pi\in\Pi(\mu,\nu)} \mathbb{E}^{\mathcal{L}(X,Y)\sim\pi}\big[a(X)1_{\{X\in B\}} + b(Y)1_{\{X\notin B\}}\big]. \tag{3.23}$$

Let, for $B \in \mathcal{B}(\mathbb{R})$, $\tilde{\mathcal{P}}_B$ be defined by

$$\tilde{\mathcal{P}}_B := \sup_{\pi\in\Pi(\mu,\nu)} \mathbb{E}^{\mathcal{L}(X,Y)\sim\pi}\big[a(X)1_{\{X\in B\}} + b(Y)1_{\{X\notin B\}}\big], \tag{3.24}$$

so that $\tilde{\mathcal{P}} = \sup_{B\in\mathcal{B}(\mathbb{R})} \tilde{\mathcal{P}}_B$. Note that $\tilde{\mathcal{P}}_B$ can be treated as the MOT problem with a (non-standard) payoff function $c(x,y,B) = a(x)I_B(x) + b(y)1_{B^c}(x)$.

The main advantage of considering (3.23) instead of (3.22) is the following:

Lemma 3.3.1. *Suppose $a, b : \mathbb{R} \mapsto \mathbb{R}$ with $b(\cdot)$ convex. Then, for any $B \in \mathcal{B}(\mathbb{R})$, the shadow embedding induced by the restriction $\mu|_B$, $S^\nu(\mu|_B)$, is optimal in* (3.24). *In particular,*

$$\tilde{\mathcal{P}} = \int_{\mathbb{R}} b(y)\nu(dy) + \sup_{B\in\mathcal{B}(\mathbb{R})} \Big\{ \int_B a(x)\mu(dx) - \int_{\mathbb{R}} b(y)S^\nu(\mu|_B)(dy) \Big\}.$$

Proof. For any $\pi \in \Pi(\mu,\nu)$ and $B \in \mathcal{B}(\mathbb{R})$, we have that

$$\begin{aligned}
&\mathbb{E}^{\mathcal{L}(X,Y)\sim\pi}\big[a(X)1_{\{X\in B\}} + b(Y)1_{\{X\notin B\}}\big] \\
&\qquad = \mathbb{E}^{X\sim\mu}\big[a(X)1_{\{X\in B\}}\big] + \mathbb{E}^{\mathcal{L}(X,Y)\sim\pi}\big[b(Y)1_{\{X\notin B\}}\big] \\
&\qquad = \mathbb{E}^{X\sim\mu}\big[a(X)1_{\{X\in B\}}\big] + \mathbb{E}^{Y\sim\nu}\big[b(Y)\big] - \mathbb{E}^{\mathcal{L}(X,Y)\sim\pi}\big[b(Y)1_{\{X\in B\}}\big].
\end{aligned}$$

Taking supremum over $\Pi(\mu,\nu)$ gives

$$\begin{aligned}\tilde{\mathcal{P}}_B &= \mathbb{E}^{X\sim\mu}\big[a(X)1_{\{X\in B\}}\big] + \mathbb{E}^{Y\sim\nu}\big[b(Y)\big] - \inf_{\pi\in\Pi(\mu,\nu)} \mathbb{E}^{\mathcal{L}(X,Y)\sim\pi}\big[b(Y)1_{\{X\in B\}}\big] \\ &= \int_B a(x)\mu(dx) + \int_{\mathbb{R}} b(y)\nu(dy) - \inf_{\pi\in\Pi(\mu,\nu)} \int_{\mathbb{R}} b(y)\pi(B,dy),\end{aligned}$$

for any $B\in\mathcal{B}(\mathbb{R})$. Since $\mu|_B\leq_E \nu$ and $b(\cdot)$ is convex, by Lemma 2.2.1 we have that

$$\inf_{\pi\in\Pi(\mu,\nu)} \int_{\mathbb{R}} b(y)\pi(B,dy) = \int_{\mathbb{R}} b(y)S^\nu(\mu|_B)(dy), \quad B\in\mathcal{B}(\mathbb{R}),$$

which ends the proof. $\square$

We hope that an explicit construction of the shadow (see Theorem 2.2.2) coupled with a particular structure of time 1 payoff $a(\cdot)$ (e.g. assuming it is also convex) allows to identify the Borel set B that maximises $\tilde{\mathcal{P}}_B$. This is left for future research.

Remark 3.3.2. *If the goal is to find the lowest model-independent price of the American type option, i.e. to calculate*

$$\bar{\mathcal{P}} := \inf_{\pi\in\Pi(\mu,\nu)} \sup_{B\in\mathcal{B}(\mathbb{R})} \mathbb{E}^{\mathcal{L}(X,Y)\sim\pi}\big[a(X)1_{\{X\in B\}} + b(Y)1_{\{X\notin B\}}\big],$$

then, similarly as in Lemma 3.3.1, we have that

$$\bar{\mathcal{P}} = \sup_{B\in\mathcal{B}(\mathbb{R})} \Big\{ \int_B a(x)\mu(dx) + \int_{\mathbb{R}} b(y)S^\nu(\mu|_{B^c})(dy).$$

Chapter 4

The left-curtain martingale coupling in the presence of atoms

The main effort in this chapter is in proving Theorem 4.1.1 which extends the left-curtain martingale coupling to the presence of atoms in the starting law μ.

For this chapter, Section 2.1 again serves as a prerequisite section on probability measures, convex order, martingale couplings, and the left-curtain martingale coupling in particular. Moreover, for two probability measures μ, ν in convex order, recall the definition of a (μ, ν)-consistent model from Chapter 3 (see Definition 3.1.1): $(\mathcal{S}, M)$ is a (μ, ν)-consistent model if $\mathcal{S}$ is a filtered probability space and M is a $(\mathcal{S}, \mu, \nu)$ consistent stochastic process, i.e. M is an $\mathcal{S}$-martingale and $M_1 \sim \mu, M_2 \sim \nu$.

4.1 An extension of the left-curtain mapping to the general case

In this section we construct a new representation of the left-curtain martingale coupling of Beiglböck and Juillet [10]. Our approach is to construct (X, Y) from a pair of independent uniform $U(0,1)$ random variables U and V. The construction of X is straightforward: we set $X = G_\mu(U)$ (where G_μ is the left-continuous quantile function of a random variable with law μ).

It remains to construct Y. First we consider the case of a point mass at w, $\mu = \delta_w$, and show how to construct functions $R = R_{\mu,\nu}$ and $S = S_{\mu,\nu}$ with $R_{\mu,\nu}(u) \leq G_\mu(u) \leq S_{\mu,\nu}(u)$, such that if $X = G_{\delta_w}(U) = w$ and $Y \in \{R(U), S(U)\}$

with $\mathbb{P}(Y = R(u)|U = u) = \frac{S(u)-G(u)}{S(u)-R(u)}$ then Y has law ν. In particular, conditional on $U = u$, Y takes values in $\{R(u), S(u)\}$ and satisfies $\mathbb{E}[Y|U = u] = G_\mu(u)$. Second, we show how this result extends to the case of a measure μ consisting of finitely many atoms. Third, for the case of general μ we construct an approximation $(\mu_n)_{n\geq 1}$ of μ and associated functions $(R_n, G_n, S_n)_{n\geq 1}$ where each μ_n is finitely supported. We show that we can define limits (R, G, S) such that (R, G, S) can be used to construct a martingale $M = (M_0 = \bar{\mu}, M_1 = X, M_2 = Y)$ with the property that $X = G(U)$ and $Y \in \{R(U), S(U)\}$ and such that $\mathcal{L}(X) = \mu$ and $\mathcal{L}(Y) = \nu$. The functions $R, S : (0, 1) \mapsto \mathbb{R}$ we define have the properties

$$R(u) \leq G(u) \leq S(u); \qquad S \text{ is increasing}; \qquad \text{for } 0 < u < v < 1,\ R(v) \notin (R(u), S(u)). \tag{4.1}$$

We suppose $\mu \leq_{cx} \nu$ are fixed and given and we abbreviate the quantile function G_μ by G. The aim of this section is to prove the following theorem:

Theorem 4.1.1. *There exist functions $R, S : (0, 1) \mapsto \mathbb{R}$ satisfying (4.1) such that if we define $X(u, v) = X(u) = G(u)$ and $Y(u, v) \in \{R(u), S(u)\}$ by $Y(u, v) = G(u)$ on $G(u) = S(u)$ and*

$$Y(u, v) = R(u)1_{\{v\leq \frac{S(u)-G(u)}{S(u)-R(u)}\}} + S(u)1_{\{v> \frac{S(u)-G(u)}{S(u)-R(u)}\}} \tag{4.2}$$

otherwise, and if U and V are independent $U(0, 1)$ random variables then $M = (\bar{\mu}, X(U), Y(U, V))$ is a $\mathbb{F} = (\mathcal{F}_0 = \{\emptyset, \Omega\}, \mathcal{F}_1 = \sigma(U), \mathcal{F}_2 = \sigma(U, V)\})$-martingale for which $\mathcal{L}(X) = \mu$ and $\mathcal{L}(Y) = \nu$.

In particular, if $\Omega = (0, 1) \times (0, 1)$, $\mathcal{F} = \mathcal{B}(\Omega)$, $\mathbb{P} = Leb(\Omega)$, if $\mathbb{F}$ and M are defined as above and if $\mathcal{S} = (\Omega, \mathcal{F}, \mathbb{F}, \mathbb{P})$ then $(\mathcal{S}, M)$ is a (μ, ν)-consistent model.

Remark 4.1.2. *For $n \geq 1$, let π^n_{lc} be the left-curtain coupling of the initial law μ_n (consisting of n atoms) and target law ν. Juillet [67] proved that if $(\mu_n)_{n\geq 1}$ converges weakly to μ then $(\pi^n_{lc})_{n\geq 1}$ converges weakly to the left-curtain coupling of μ and ν.*

Here we argue differently. We use the fact that π^n_{lc} can be represented by an explicitly constructed triple (S_n, G_n, R_n). Then, by sending $n \to +\infty$, we show that the limiting functions give rise to the left-monotone martingale coupling, and thus also to π_{lc}, of μ and ν.

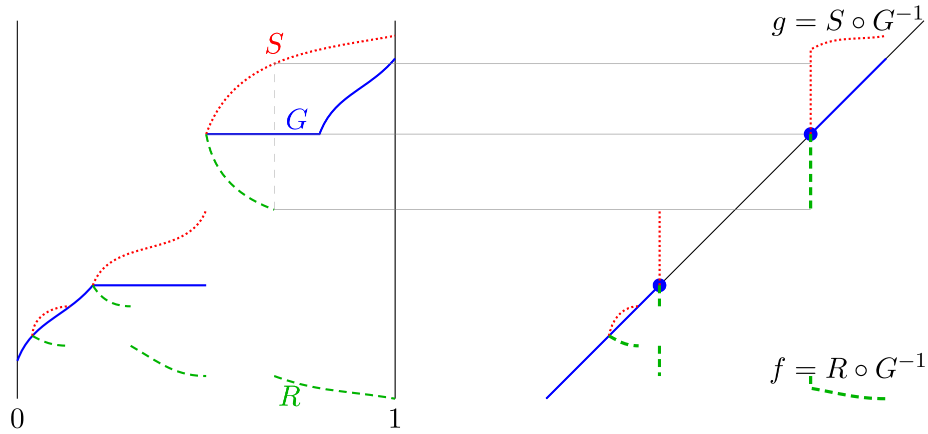

Figure 4.1: Sketch of R, G, S and the corresponding f and g. On the atoms of μ, G is flat, and f and g are multi-valued, but R and S remain well-defined.

When μ is continuous and f and g are well-defined the construction of this section is related to that of Beiglböck and Juillet [10] (see also Henry-Labordère and Touzi [50]) via the relationships $S = g \circ G_\mu$ and $R = f \circ G_\mu$. Suppose ν is also continuous and fix x. Then under the left-curtain martingale coupling $\{f(x), g(x)\}$ with $f(x) \leq x \leq g(x)$ are solutions to the mass and mean conditions

$$\int_f^x \mu(dz) = \int_f^g \nu(dz), \tag{4.3}$$

$$\int_f^x z\mu(dz) = \int_f^g z\nu(dz). \tag{4.4}$$

When μ has atoms, G_μ has intervals of constancy and f and g are multi-valued, but R and S remain well-defined. See Figure 4.1. Then, for general μ and ν, the appropriate generalisations of (4.3) and (4.4) are

$$\int_{(R(u),G(u))} \mu(dz) + \overline{\lambda}^\mu_u = \int_{(R(u),S(u))} \nu(dz) + \underline{\lambda}^\nu_u + \overline{\lambda}^\nu_u, \tag{4.5}$$

$$\int_{(R(u),G(u))} z\mu(dz) + \overline{\lambda}^\mu_u G(u) = \int_{(R(u),S(u))} z\nu(dz) + \underline{\lambda}^\nu_u R(u) + \overline{\lambda}^\nu_u S(u), \tag{4.6}$$

respectively, where the quantities $0 \leq \overline{\lambda}^\mu_u \leq \mu(\{G(u)\})$, $0 \leq \underline{\lambda}^\nu_u \leq (\nu - \mu)(\{R(u)\})$, $0 \leq \overline{\lambda}^\nu_u \leq \nu(\{S(u)\})$ are uniquely determined by the triple (R, G, S). Essentially, (4.5) is preservation of mass condition and (4.6) is preservation of mean condition. Together they give the martingale property.

4.1.1 The case where μ is a point mass

The goal in this section is to prove Theorem 4.1.1 in the special case where μ is a point mass. We assume that μ is a unit atom at w and ν is centred at w. Then $\mu = \delta_w \leq_{cx} \nu$.

Let $P(k) = P_\nu(k) = \int_{-\infty}^{\infty}(k-z)^+\nu(dz)$. Then $P(k) \geq (k-w)^+$. For $p \in [0, P(w)]$ define $\alpha : [0, P(w)] \mapsto [w, \infty]$ and $\beta : [0, P(w)] \mapsto [-\infty, w]$ by

$$\alpha(p) = \operatorname{arginf}_{k>w}\left\{\frac{P(k)-p}{k-w}\right\}; \qquad \beta(p) = \operatorname{argsup}_{k<w}\left\{\frac{p-P(k)}{w-k}\right\}, \quad (4.7)$$

see Figure 4.2. Then α is decreasing and β is increasing. Since the arginf and argsup may not be uniquely defined (this happens when ν has intervals with no mass) we avoid indeterminacy by assuming that α and β are right-continuous. (We also set $\alpha(P(w)) = \inf\{z > w : F_\nu(z) > F_\nu(w)\}$ and $\beta(P(w)) = \sup\{z < w : F_\nu(z) < F_\nu(w-)\}$. Note that $\alpha(0) = r_\nu$ and $\beta(0) = \ell_\nu$.) If ν has atoms then α and β may fail to be strictly monotonic.

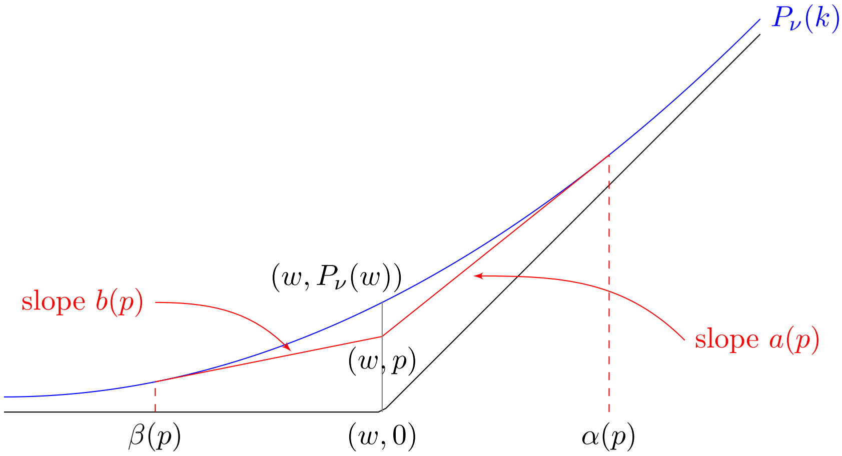

Figure 4.2: The definitions of α, β, a and b. $\Upsilon(p)$ is the difference in the slopes of the tangents to $P_\nu(k)$ which pass through (w, p).

For $p \in (0, P(w))$ define also

$$a(p) = \inf_{k>w}\frac{P(k)-p}{k-w} = \frac{P(\alpha(p))-p}{\alpha(p)-w}; \qquad b(p) = \sup_{k<w}\frac{p-P(k)}{w-k} = \frac{p-P(\beta(p))}{w-\beta(p)}. \quad (4.8)$$

Extend the representations to $[0, P(w)]$ by taking limits. Then $a : [0, P(w)] \mapsto [P'(w+), 1]$ is decreasing and $b : [0, P(w)] \mapsto [0, P'(w-)]$ is increasing. We have the representations

$$a(p) = 1 - \int_0^p \frac{dq}{\alpha(q) - w}; \qquad\qquad b(p) = \int_0^p \frac{dq}{w - \beta(q)}.$$

Let $\Upsilon : [0, P(w)] \mapsto [0,1]$ be given by $\Upsilon(p) = a(p) - b(p)$. Then $\Upsilon(0) = 1$ and $\Upsilon(P(w)) = \nu(\{w\})$. Υ is a decreasing, concave function which is absolutely continuous on $[0, P(w))$. We can define an inverse $\Upsilon^{-1} : [0,1] \to [0, P(w)]$ provided we set $\Upsilon^{-1}(q) = 1$ for $q \le \nu(\{w\})$. Where α and β are continuous we have $\Upsilon'(p) = -\frac{1}{\alpha(p)-w} - \frac{1}{w-\beta(p)}$.

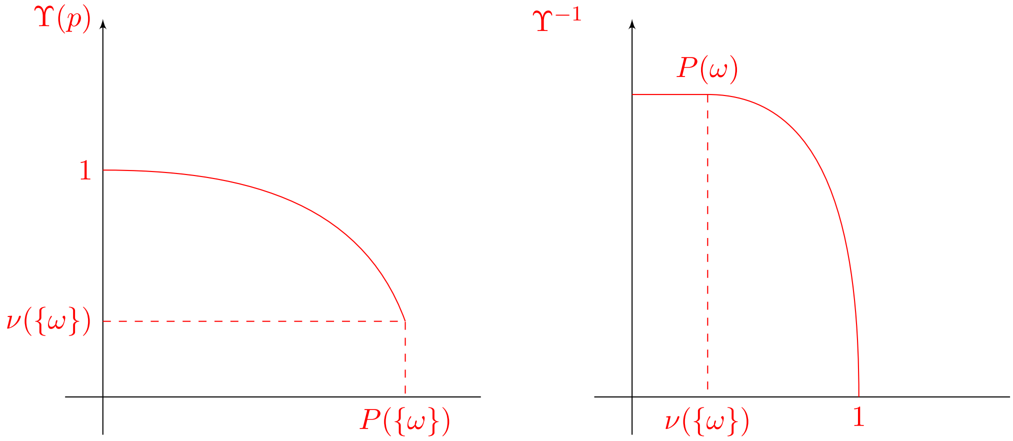

Figure 4.3: Sketch of Υ and Υ^{-1}.

Define $S : (0,1) \mapsto \mathbb{R}$ by $S(u) = (\alpha \circ \Upsilon^{-1})(u)$ and $R : (0,1) \mapsto \mathbb{R}$ by $R(u) = (\beta \circ \Upsilon^{-1})(u)$.

Remark 4.1.3. *If ν does not charge an open interval $A \subset (w, \infty)$, then P is linear on A. Then α jumps over this set and S does not take values in A. Similarly if ν does not charge an open interval $B \subset (-\infty, w)$ then R jumps over this interval.*

Remark 4.1.4. *By construction, α and β are both right-continuous. Since Υ^{-1} is continuous and decreasing, it follows that R and S are left-continuous. Moreover, $\lim_{u \to 1} R(u) = \ell_\nu$ and $\lim_{u \to 1} S(u) = r_\nu$.*

Let Y be defined by (4.2) in Theorem 4.1.1. Note that since μ is a point mass $G(u) = w$ for all $u \in (0,1)$.

Lemma 4.1.5. *Suppose U, V are independent uniform random variables. Then $Y(U,V)$ has law ν.*

Proof. Let ϕ be a test function: a continuously differentiable function with support contained in $[w+\epsilon, w+\epsilon^{-1}]$ for some $\epsilon \in (0,1)$. We will show that $\mathbb{E}[\phi(Y)] = \int \phi(y)\nu(dy)$. We can prove a similar result for test functions ψ with support in $[w-\epsilon^{-1}, w-\epsilon]$. It follows that $\mathcal{L}(Y) = \nu$.

By construction

$$\begin{aligned}
\mathbb{E}[\phi(Y)] &= \int_0^1 du \frac{w - R(u)}{S(u)-R(u)}\phi(S(u)) \\
&= \int_0^1 du \frac{w - \beta\circ\Upsilon^{-1}(u)}{\alpha\circ\Upsilon^{-1}(u) - \beta\circ\Upsilon^{-1}(u)}\phi(\alpha\circ\Upsilon^{-1}(u)) \\
&= \int_0^{P(w)} dp|\Upsilon'(p)|\frac{w-\beta(p)}{\alpha(p)-\beta(p)}\phi(\alpha(p)).
\end{aligned}$$

But $\Upsilon'(p) = -\frac{\alpha(p)-\beta(p)}{(\alpha(p)-w)(w-\beta(p))}$. Thus, writing $\psi(y) = \frac{\phi(y)}{(y-w)}$ and using the fact that $\alpha^{-1}(y) = P(y) - (y-w)P'(y)$ except at the countably many points where α^{-1} is multi-valued,

$$\begin{aligned}
\mathbb{E}[\phi(Y)] &= \int_0^{P(w)} dp \frac{\phi(\alpha(p))}{\alpha(p)-w} = -\int_w^\infty d(\alpha^{-1}(y))\psi(y) = \int_w^\infty [P(y) - (y-w)P'(y)]\psi'(y)dy \\
&= -\int_w^\infty P'(y)[\psi(y) + (y-w)\psi'(y)]dy = -\int_w^\infty P'(y)\phi'(y)dy = \int \phi(y)\nu(dy).
\end{aligned}$$

Hence $\mathbb{E}[\phi(y)] = \int \phi(y)\nu(dy)$. □

Remark 4.1.6. *If α and β are strictly monotonic at $\Upsilon^{-1}(u)$, then conditional on $U \le u$, Y has law ν conditioned to take values in $[\beta\circ\Upsilon^{-1}(u), \alpha\circ\Upsilon^{-1}(u)]$. Necessarily, $\nu([\beta\circ\Upsilon^{-1}(u), \alpha\circ\Upsilon^{-1}(u)]) = u$.*

If there is an atom of ν at $\beta\circ\Upsilon^{-1}(u)$ or $\alpha\circ\Upsilon^{-1}(u)$ then we can choose appropriate masses $\underline{\lambda}_u$ and $\overline{\lambda}_u$ such that $\nu((\beta\circ\Upsilon^{-1}(u), \alpha\circ\Upsilon^{-1}(u))) + \underline{\lambda}_u\delta_{\beta\circ\Upsilon^{-1}(u)} + \overline{\lambda}_u\delta_{\alpha\circ\Upsilon^{-1}(u)}$ has total mass u and mean w. We must have $0 \le \underline{\lambda}_u \le \nu(\{\beta\circ\Upsilon^{-1}(u)\})$ and $0 \le \overline{\lambda}_u \le \nu(\{\alpha\circ\Upsilon^{-1}(u)\})$.

On $U \le u_1$ let $Y^{u_1} = Y^{u_1}(U,V)$ be constructed as in (4.2). On $U > u_1$, let Y^{u_1} be in a graveyard state Δ. Then $\mathcal{L}(Y^{u_1}) = \nu_{u_1} + (1-u_1)\delta_\Delta$ where ν_{u_1} is a measure on $[R(u_1), S(u_1)]$ with total mass u_1 and mean w. In particular, $\nu_{u_1} = \nu$ on $(R(u_1), S(u_1))$, $\nu_{u_1} \le \nu$ on $\{R(u_1), S(u_1)\}$ and $\nu_{u_1} = 0$ on $[R(u_1), S(u_1)]^C$.

4.1.2 The case where μ consists of a finite number of atoms

Suppose $\mu = \sum_{i=1}^N \lambda_i\delta_{x_i}$ where $x_1 < x_2 \ldots < x_N$ with $\lambda_i > 0$ and $\sum_{i=1}^N \lambda_i = 1$. Suppose ν is an arbitrary probability measure satisfying the convex order condition $\mu \le_{cx} \nu$.

For $0 \leq p \leq P_\nu(x_1)$ we can construct α, β, a and b as in (4.7) and (4.8) (but relative to x_1 rather than the mean w) and set $\Upsilon = a - b$. For example, $\alpha(p) = \mathrm{arginf}_{k>x_1} \frac{P_\nu(k)-p}{k-x_1}$ and $a(p) = \inf_{k>x_1} \frac{P_\nu(k)-p}{k-x_1}$. Note that $\Upsilon(0) = \Lambda_1 := \inf_{x>x_1} \frac{P_\nu(x)}{x-x_1}$ and since $P_\nu(x) \geq P_\mu(x) \geq \lambda_1(x - x_1)$ we have $\Lambda_1 \geq \lambda_1$. The inverse Υ^{-1} can be defined on $[0, \Lambda_1]$, but we are only interested in Υ^{-1} over the interval $[0, \lambda_1]$. Using Υ^{-1} and the construction of the previous section we can define $S = \alpha \circ \Upsilon^{-1} : (0, \lambda_1] \mapsto [x_1, \infty)$ and $R = \beta \circ \Upsilon^{-1} : (0, \lambda_1] \mapsto (-\infty, x_1]$ with S increasing and R decreasing.

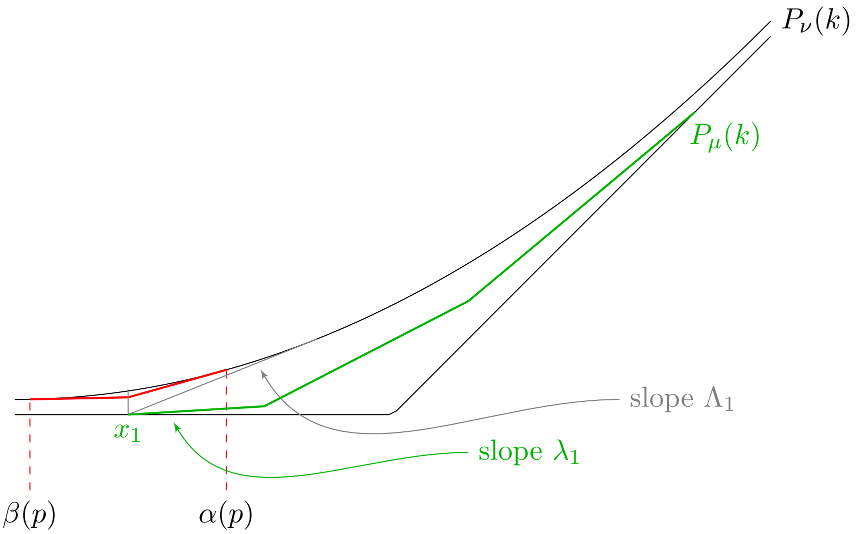

Figure 4.4: Calculation of α, β, a and b in this case

By the final comments in Remark 4.1.6, the construction of R and S on $(0, \lambda_1]$ is such that if Y is constructed as in (4.2), then on $U \leq \lambda_1$ we find Y has law ν_{λ_1}, where $\nu_{\lambda_1} = \nu$ on $(R(\lambda_1), S(\lambda_1))$ and $\nu_{\lambda_1} \leq \nu$ on $\{R(\lambda_1), S(\lambda_1)\}$.

We now claim that $\tilde{\mu}_1 := \mu - \lambda_1 \delta_{x_1} = \sum_{i=2}^N \lambda_i \delta_{x_i}$ and $\tilde{\nu}_1 = \nu - \nu_{\lambda_1}$ are in convex order. By construction ν_{λ_1} has mass λ_1 and barycentre x_1. Hence $\tilde{\mu}_1$ and $\tilde{\nu}_1$ also have the same total mass and barycentre.

Lemma 4.1.7. $\tilde{\mu}_1 \leq_{cx} \tilde{\nu}_1$.

Proof. Let $\hat{\nu} = \lambda_1 \delta_{x_1} + \tilde{\nu}_1$. Since $\lambda_1 \delta_{x_1} \leq_{cx} \nu_{\lambda_1}$ we have $\hat{\nu} \leq_{cx} \nu$. Also $P_\mu(k) \leq P_{\hat{\nu}}(k)$. To see this note that $P_{\hat{\nu}}$ is continuous everywhere and linear on intervals $[R(\lambda_1), x_1]$ and $[x_1, S(\lambda_1)]$, whereas P_μ is continuous and convex on $[R(\lambda_1), S(\lambda_1)]$. Moreover, $P_\mu(R(\lambda_1)) = 0 \leq P_{\hat{\nu}}(R(\lambda_1))$, $P_\mu(x_1) = 0 \leq P_{\hat{\nu}}(R(x_1))$ and $P_\mu(S(\lambda_1)) \leq P_\nu(S(\lambda_1)) = P_{\hat{\nu}}(S(\lambda_1))$. Hence $P_{\tilde{\mu}}(k) + \lambda_1(x_1 - k)^+ = P_\mu(k) \leq P_{\hat{\nu}}(k) = P_{\tilde{\nu}_1}(k) + \lambda_1(x_1 - k)^+$ and it follows that $P_{\tilde{\mu}}(k) \leq P_{\tilde{\nu}_1}(k)$ as required. $\square$

We have constructed (R,S) on $(0,\lambda_1]$ with S increasing and R decreasing in such a way that the point mass at x_1 is mapped to ν_{λ_1}. It remains to embed $\tilde{\nu}_1$ starting from $\tilde{\mu}_1$. Note that by Remark 4.1.6, $\tilde{\nu}_1$ places no mass on $(R(\lambda_1), S(\lambda_1))$.

As a next step we embed the atom $\lambda_2\delta_{x_2}$ of $\tilde{\mu}_1$ in $\tilde{\nu}_1$. x_2 is the lowest location of an atom in $\tilde{\mu}_1$ so we can use the same algorithm as before. In this way, for $\lambda_1 < u \le \lambda_1 + \lambda_2$ we construct S increasing with $S(\lambda_1+) \ge S(\lambda_1-) \vee x_2$ and R decreasing with $R(\lambda_1+) \le x_2$. By Remark 4.1.3, R jumps over the interval $(R(\lambda_1), S(\lambda_1))$. We conclude that for $0 < u < v < \lambda_1 + \lambda_2$, $R(v) \notin (R(u), S(u))$.

Thereafter, we proceed inductively on the number of atoms which have been embedded. The initial law is a sub-probability $\tilde{\mu}_k = \sum_{k+1}^N \lambda_i\delta_{x_i}$ which we want to map to a target law $\tilde{\nu}_k$ where $\tilde{\mu}_k \le_{cx} \tilde{\nu}_k$ and $\tilde{\nu}_k \le \nu$. Since μ consists of a finite number of atoms the construction terminates. Moreover the random variable Y we construct in this way has law ν and R and S have the properties in (4.1). It follows that we have proved Theorem 4.1.1 in the case where μ consists of a finite number of atoms.

4.1.3 The martingale coupling and its inverse as maps

Given ν centred, (and $\mu = \delta_0$) we saw in Section 4.1.1 how to construct $R : (0,1) \mapsto \mathbb{R}_-$ and $S : (0,1) \mapsto \mathbb{R}_+$ such that $Y = Y(U,V)$ has law ν where Y is given by $Y(u,v) = 0$ if $S(u) = 0$ and

$$Y(u,v) = R(u)1_{\{v \le \frac{S(u)}{S(u)-R(u)}\}} + S(u)1_{\{v > \frac{S(u)}{S(u)-R(u)}\}} \tag{4.9}$$

otherwise.

Let $\mathcal{P}^0$ denote the set of centred probability measures on $\mathbb{R}$. Let $\mathcal{V}^1$ denote the set of pairs of functions R,S with $R : (0,1) \to \mathbb{R}_-$ and $S : (0,1) \to \mathbb{R}_+$, let $\mathcal{V}^1_{Mon}$ denote the subset of $\mathcal{V}^1$ for which R is decreasing and S is increasing, and let $\mathcal{V}^1_{Int}$ denote the subset of $\mathcal{V}^1$ such that $I(R,S) < \infty$ where

$$I(f,g) = \int_0^1 du \frac{|f(u)|g(u)}{g(u)-f(u)} 1_{\{g(u)>0\}}.$$

Finally, let $\mathcal{V}^1_{Mon,Int} = \mathcal{V}^1_{Mon} \cap \mathcal{V}^1_{Int}$.

The construction in Section 4.1.1 can be considered as a pair of maps

$$\mathcal{Q}^1 : \mathcal{P}^0(\mathbb{R}) \mapsto \mathcal{V}^1_{Mon,Int}$$
$$\mathcal{R}^1 : \mathcal{V}^1_{Mon,Int} \mapsto \mathcal{P}^0(\mathbb{R})$$

Note that $\mathbb{E}[|Y|] = 2I(R,S)$ which can be shown using the ideas in the proof of Lemma 4.1.5 to be equal to $2P_\nu(0)$. Moreover, under $I(R,S) < \infty$ we have $\mathbb{E}[Y] = 0$.

Note that if we take $(R,S) \in \mathcal{V}^1_{Mon} \setminus \mathcal{V}^1_{Mon,Int}$ then we can still define Y via (4.9) but $\mathcal{L}(Y)$ will not be integrable. Then M given by $M_1 = 0$, $M_2 = Y$ is a local martingale, but not a martingale.

Section 4.1.2 extends these results from initial laws which consist of a single atom to finite combinations of atoms. Let $\mathcal{P}^0_F$ be the subset of $\mathcal{P}^0$ for which the measure consists of a finite set of atoms and let $\mathcal{C}_F = \{(\zeta,\chi) : \zeta \in \mathcal{P}^0_F, \chi \in \mathcal{P}^0; \zeta \leq_{cx} \chi\}$. Let

$$\mathcal{V} = \{(R,G,S); R : (0,1) \to \mathbb{R}, G : (0,1) \to \mathbb{R}, S : (0,1) \to \mathbb{R}; R(u) \leq G(u) \leq S(u); \int_0^1 |G(u)|du < \infty, \int_0^1 G(u)du = 0\} \quad .$$

Consider now the subsets

$$\begin{aligned}
\mathcal{V}_F &= \{(R,G,S) \in \mathcal{V} : G \text{ non-decreasing and takes only finitely many values}\}, \\
\mathcal{V}_{Mon} &= \{(R,G,S) \in \mathcal{V} : (4.1) \text{ holds}\}, \\
\mathcal{V}_{Int} &= \{(R,G,S) \in \mathcal{V} : I(R,G,S) < \infty\};
\end{aligned}$$

where $I(R,G,S) = \int_0^1 du \frac{(S(u)-G(u))(G(u)-R(u))}{S(u)-R(u)} 1_{\{S(u)>G(u)\}}$, and consider also intersections of these subsets, for example $\mathcal{V}_{Mon,Int} = \mathcal{V}_{Mon} \cap \mathcal{V}_{Int}$. In Section 4.1.2 we constructed a map $Q : \mathcal{C}_F \to \mathcal{V}_{F,Mon}$ which we write as $Q(\zeta,\chi) = (R_{(\zeta,\chi)}, G_\zeta, S_{(\zeta,\chi)})$. Indeed, since $\chi \in L^1$ and since $\mathbb{E}[|Y - X|] \leq \mathbb{E}[|X|] + \mathbb{E}[|Y|] < \infty$ we have that $\mathbb{E}[|Y - X|] = 2I(R_{(\zeta,\chi)}, G_\zeta, S_{(\zeta,\chi)})$, so that we actually have a map $Q : \mathcal{C}_F \to \mathcal{V}_{F,Mon,Int}$. Conversely, the arguments after Lemma 4.1.7 show that (4.2) defines a inverse map $\mathcal{R} : \mathcal{V}_{F,Mon,Int} \to \mathcal{C}_F$.

Note that given any element (R,G,S) of $\mathcal{V}$ we can define the map $\mathcal{R} : \mathcal{V} \to \mathcal{P}^0 \times \mathcal{P}$ via $\mathcal{R}(R,G,S) = (\mathcal{L}(X(U)), \mathcal{L}(Y(U,V)))$ where $Y(U,V)$ is as given in the statement of Theorem 4.1.1. We will make no further use of this idea, but different properties of (R,G,S) will lead to different (local)-martingale couplings. The embedding of Hobson and Neuberger [58] is of this type. In the Hobson and Neuberger embedding R and S are both increasing.

4.1.4 The case of general integrable μ

We assume μ is centred at zero, but the general case follows by translation.

Our goal in this section is to extend the map $\mathcal{Q} : \mathcal{C}_F \to \mathcal{V}_{F,Mon,Int}$ with inverse $\mathcal{R}$ to a map $\mathcal{Q} : \mathcal{C} \to \mathcal{V}_{Mon,Int}$ where $\mathcal{C} = \{(\zeta,\chi) : \zeta \in \mathcal{P}^0, \chi \in \mathcal{P}^0; \zeta \leq_{cx} \chi\}$.

For μ a general centred probability measure and ν a centred target measure with $\mu \leq_{cx} \nu$ we construct a sequence $(\mu_n)_{n\geq 1}$ of approximations of μ by elements of $\mathcal{P}^0_F$. For each μ_n we can construct a triple (R_n, G_n, S_n). We show that $(R_n, G_n, S_n)_{n\geq 1}$ converge to a limit (R, G, S) first on the rationals and then (almost surely) on $(0, 1)$. Convergence of G_n and S_n is straightforward, but convergence of R_n is more subtle, and indeed we only have convergence on $\{u : S(u) > G(u)\}$. Finally we show that $\mathcal{R}(R, G, S) = (\mu, \nu)$ so that the trio (R, G, S) defines a martingale coupling between μ and ν.

Let $\{q_1, q_2 \ldots\}$ be an enumeration of $\mathbb{Q} \cap (0, 1)$. Then $\{S_n(q_1)\}_{n\geq 1}$ converges down a subsequence n_{k_1} to a limit $S_\infty(q_1) := \lim_{k_1\uparrow\infty} S_{n_{k_1}}(q_1)$. Down a further subsequence if necessary we have that $S_{n_{k_2}}(q_2)$ converges to $S_\infty(q_2)$. Proceeding inductively, we have by a diagonal argument (see, for example, Billingsley [14]) that there is a subsequence $(m_1, m_2, \ldots)$ such that $\{S_{m_k}\}_{k\geq 1}$ converges to S_∞ at every rational $q \in \mathbb{Q} \cap (0, 1)$. This limit is non-decreasing.

Our first result shows that any limit of S_n is finite valued. Since the ideas behind the proof are not relevant to the arguments of this section the proof is postponed to Section 4.3.

Lemma 4.1.8. *Let $\mu_n \uparrow_{cx} \mu$. Then $\limsup S_n(u) \leq J(u)$ for some function $J = J_{\mu,\nu} : (0, 1) \mapsto (-\infty, \infty)$.*

We want to extend the domain from the rationals to $(0, 1)$. To this end define $S(u) = \lim_{q_j\uparrow u} S_\infty(q_j)$. This limit is well defined (and non-decreasing) by the monotonicity of S_∞. Then from the monotonicity of S we conclude that S has only countably many discontinuities. Note that, by definition, S is left-continuous.

We can construct G from $\{G_n\}$ in an identical fashion. In this case the finiteness of the limit follows from the tightness of the singleton $\{G\}$. Moreover, since $G_n \leq S_n$ by construction, we have $G_\infty \leq S_\infty$ and $G \leq S$. Again, the increasing limit G has at most countably many discontinuities and is left-continuous.

Define $\mathcal{N}_S = \{u : S_n(u) \not\to S(u)\}$ and $\mathcal{N}_G = \{u : G_n(u) \not\to G(u)\}$ where the subscript n refers to a subsequence down which S_n and G_n converge on rationals. Define also $\mathcal{N}^\Delta_S = \{u : S(u+) > S(u-)\}$ and $\mathcal{N}^\Delta_G = \{u : G(+) > G(u-)\}$.

Lemma 4.1.9. *$\mathcal{N}_S \subseteq \mathcal{N}^\Delta_S$ and $\mathcal{N}_G \subseteq \mathcal{N}^\Delta_G$. Moreover, Leb$(\mathcal{N}_S \cup \mathcal{N}_G) = 0$.*

Proof. Suppose u is a continuity point of S. Suppose further that there is a subsequence $(n_j)_{j\geq 1}$ along which $S_{n_j}(u) > S(u) + \epsilon$. Using the continuity of S at u we may pick $q > u$ such that $S(q) < S(u) + \epsilon/2$. Take $q_k \in (u, q)$ with $q_k \downarrow u$. Then $S_{n_j}(q_k) \geq S_{n_j}(u) > S(u) + \epsilon > S(q_k) + \epsilon/2$. Letting $j \uparrow \infty$, $S_\infty(q_k) > S(q_k) + \epsilon/2$. Letting $k \uparrow \infty$, $S(u) \geq S(u) + \epsilon/2$ which is a contradiction.

A similar argument (without the need of continuity at u) shows that down any subsequence $\lim_j S_{n_j}(u) > S(u) - \epsilon$. Hence, if $S(u) = S(u+)$ then $S(u) = \lim S_n(u)$. Since the set of points for which $S(u+) > S(u)$ is countable we conclude that $\mathrm{Leb}(\mathcal{N}_S) = 0$.

An identical argument gives that $G(u) = \lim_n G_n(u)$ on $G(u+) = G(u)$ and $\mathrm{Leb}(\mathcal{N}_G) = 0$. $\square$

Now consider $(R_n)_{n\geq 1}$ and the existence of a possible limit R. By the same diagonal argument as above we can define $R_\infty : \mathbb{Q} \cap (0,1) \to \mathbb{R}$ such that on a subsequence $R_{n_k}(q) \to R_\infty(q) \in [-\infty, \infty]$ for every q. (From now on we work on a subsequence indexed n such that $\{S_n\}_n$, $\{G_n\}_n$ and $\{R_n\}_n$ converge for every $q \in \mathbb{Q} \cap (0,1)$.) We want to construct R from R_∞, but unlike in the case of S or G we do not have monotonicity. Note that for $q' > q$ we have $R_n(q') \notin (R_n(q), S_n(q))$ for each n and this implies $R_\infty(q') \notin (R_\infty(q), S_\infty(q))$.

The following lemma shows that R_∞ is finite valued, at least for q such that $G(u+) < S(u)$.

Lemma 4.1.10. *Let $\mu_n \uparrow_{cx} \mu$. Then $\liminf R_n(u) \geq j(u)$ on $G(u+) < S(u)$ for some function $j = j_{\mu,\nu} : (0,1) \mapsto (-\infty, \infty)$.*

Let $\mathcal{A} = \{u \in (0,1) : G(u+) < S(u)\}$. By the above lemma $R_\infty(q) > j(q) > -\infty$ for $q \in \mathcal{A}$. If $u \in \mathcal{A}$ then the left continuity of S implies that there exists an interval $(u-\epsilon, u] \subseteq \mathcal{A}$; since every such interval must contain a rational we have that $\mathcal{A}$ is a countable union of intervals.

We now show that R_∞ is decreasing on each such interval. Suppose not. Then there exists $q < q'$ in the same interval I with $R_\infty(q') > R_\infty(q)$. Let $v = \inf_{q'' \in \mathbb{Q}\cap I}\{q'' : R_\infty(q'') > R_\infty(q)\}$. Choose $\tilde{q}_m \uparrow v$ with $\tilde{q}_m \geq q$ and $\hat{q}_n \downarrow v$ with $R_\infty(\hat{q}_n) > R_\infty(q)$. Then $R_\infty(\hat{q}_n) \notin (R_\infty(\tilde{q}_m), S_\infty(\tilde{q}_m))$, and since $R_\infty(\hat{q}_n) > R_\infty(q) \geq R_\infty(\tilde{q}_m)$ we conclude $R_\infty(\hat{q}_n) \geq S_\infty(\tilde{q}_m)$. Letting n tend to infinity we conclude $\liminf R_\infty(\hat{q}_n) \geq S_\infty(\tilde{q}_m)$, and letting m tend to infinity $\liminf_{n\uparrow\infty} R_\infty(\hat{q}_n) \geq S(v)$. However, $R_\infty(\hat{q}_n) \leq G_\infty(\hat{q}_n)$ and hence $\limsup_{n\uparrow\infty} R_\infty(\hat{q}_n) \leq G(v+) < S(v)$. These two statements are inconsistent, and hence R_∞ must be decreasing on each interval of $\mathcal{A}$.

Given that R_∞ is decreasing on each interval of $\mathcal{A}$, we can define R on $\mathcal{A}$ by $R(u) = \lim_{q\uparrow u} R_\infty(q)$. Then the function R is decreasing and therefore has only countably many discontinuities in any interval of $\mathcal{A}$. Away from these discontinuities, we have $R_n(u) \to R(u)$ by an argument similar to that in Lemma 4.1.9.

Define $\mathcal{B}_= = \{u \in (0,1) : G(u) = S(u)\}$ and $\mathcal{B}_< = \{u \in (0,1) : G(u) < S(u)\}$. Then $\mathcal{B}_< = \mathcal{A} \cup \mathcal{C}$ where $\mathcal{C} = \{u \in (0,1) : G(u) < S(u) \leq G(u+)\}$. Since

$\mathcal{C} \subseteq \mathcal{N}_G^{\Delta}$, we have that $\mathcal{B}_<$ and $\mathcal{A}$ differ by a set of measure zero and we conclude:

Lemma 4.1.11. $1_{\{u \in \mathcal{B}_<\}}(R_n(u) - R(u)) \to 0$, *except on a set of measure zero.*

Note that we cannot expect $R_n(u)$ to converge on $\mathcal{B}_=$.

It remains to define R on $\mathcal{B}_=$ and $\mathcal{C}$ in such a way that R satisfies (4.1). On $\mathcal{B}_=$ we set $R(u) = G(u) = S(u)$. For $u \in \mathcal{C}$ we have by the left continuity of S that there exists $\epsilon > 0$ such that $I = (u - \epsilon, u) \subset \mathcal{A}$. By the same arguments as before we conclude that R_∞ is decreasing on I and we set $R(u) = \lim_{q \uparrow u} R_\infty(q)$. Indeed, for $u \in \mathcal{B}_<$ we have $R(u) = \lim_{q \uparrow u} R_\infty(q)$. Note that for $u \in \mathcal{C}$ we may have that $R(u+) > R(u)$ and it is not true in general that R is decreasing on intervals contained in $\mathcal{B}_<$.

Fix $u < v$. If u or v is in $\mathcal{B}_=$ then since we have defined $R(w) = G(w) = S(w)$ on $\mathcal{B}_=$ we trivially have $R(v) \notin (R(u), S(u))$. For $u, v \in \mathcal{B}_<$ choose sequences $\{q_m\}_m$ with $q_m < u$ and $q_m \uparrow u$ and $\{q_l\}_l$ with $q_l \in (u, v)$ and $q_l \uparrow v$. Then $R_n(q_l) \notin (R_n(q_m), S_n(q_m))$ and hence $R_\infty(q_l) \notin (R_\infty(q_m), S_\infty(q_m))$. Letting $l \uparrow \infty$ we have $R(v) \notin (R_\infty(q_m), S_\infty(q_m))$ and letting $m \uparrow \infty$ we have $R(v) \notin (R(u), S(u))$. Hence, (R, G, S) satisfy (4.1).

On the space $\{(r, g, s); r \leq g \leq s\} \subseteq \mathbb{R}^3$ define $\Theta^x = \Theta^x(r, g, s)$ by $\Theta^x(r, g, s) = 1_{\{r \leq x < s\}} \frac{s-g}{s-r}$ with the convention that $\Theta^x(r, g, s) = 0$ for $g = s$. In particular, $\Theta^x(g, g, g) = 0$.

Proposition 4.1.12. *If x is such that* $Leb(\{u : S(u) = x\} \cup \{u : R(u) = x; S(u) > G(u)\}) = 0$, *then we have*

$$\int_0^1 du \left\{1_{\{S_n(u) \leq x\}} + \Theta^x(R_n(u), G_n(u), S_n(u))\right\} \to \int_0^1 du \left\{1_{\{S(u) \leq x\}} + \Theta^x(R(u), G(u), S(u))\right\} \tag{4.10}$$

Proof. Since $S_n(u) \to S(u)$ almost surely and since $\int_0^1 du 1_{\{S(u)=x\}} = 0$ by hypothesis, we have $\int_0^1 du 1_{\{S_n(u) \leq x\}} \to \int_0^1 du 1_{\{S(u) \leq x\}}$ by bounded convergence.

Let $\Omega_< = \{u : S_n(u) \to S(u), G_n(u) \to G(u), R_n(u) \to R(u), G(u) < S(u)\}$ and $\Omega_= = \{u : S_n(u) \to S(u), G_n(u) \to G(u), G(u) = S(u)\}$. By Lemmas 4.1.9 and 4.1.11, $\mathrm{Leb}(\Omega_< \cup \Omega_=) = 1$.

Now let $\Omega^x_< = \{u : S_n(u) \to S(u) \neq x, G_n(u) \to G(u), R_n(u) \to R(u) \neq x, G(u) < S(u)\}$ and $\Omega^x_= = \{u : S_n(u) \to S(u) \neq x, G_n(u) \to G(u), G(u) = S(u)\}$. By the hypothesis on x we still have that $\mathrm{Leb}(\Omega^x_< \cup \Omega^x_=) = 1$, and by bounded convergence the result of the proposition will follow if we can show that $\Theta^x(R_n, G_n, S_n) \to \Theta^x(R, G, S)$ on $\Omega^x_< \cup \Omega^x_=$.

This is immediate on $\Omega^x_<$. On $\Omega^x_=$ we need only note that,

$$\Theta^x(R_n, G_n, S_n) = 1_{\{R_n \le x < S_n\}} \frac{(S_n - G_n)}{(S_n - R_n)} \le \frac{(S_n - G_n)}{(S_n - x)} 1_{\{S_n > x\}} \to 0 = \Theta^x(R, G, S).$$

□

Proof of Theorem 4.1.1. All that remains to show is that (R, G, S) embeds ν.

There are at most countably many x for which $\mathrm{Leb}(\{u : S(u) = x\}) + \mathrm{Leb}(\{u : R(u) = x; S(u) > G(u)\}) > 0$. Hence it is sufficient to prove that $\int_0^1 du \left\{ 1_{\{S(u) \le x\}} + 1_{\{R(u) \le x < S(u)\}} \frac{S(u)-G(u)}{S(u)-R(u)} \right\} = \nu((-\infty, x])$ outside this set. For such an x, (4.10) holds. Then, since (R_n, G_n, S_n) embeds ν from μ_n,

$$\begin{aligned}
&\int_0^1 du \left\{ 1_{\{S(u) \le x\}} + 1_{\{R(u) \le x < S(u)\}} \frac{S(u) - G(u)}{S(u) - R(u)} \right\} \\
&= \lim_n \left\{ \int_0^1 du \left\{ 1_{\{S_n(u) \le x\}} + 1_{\{R_n(u) \le x < S_n(u)\}} \frac{S_n(u) - G_n(u)}{S_n(u) - R_n(u)} \right\} \right\} \\
&= \lim_n \nu((-\infty, x]) = \nu((-\infty, x])
\end{aligned}$$

as required. □

Remark 4.1.13 (Alternative construction)**.** *Let $(\pi^x_{lc})_{x \in \mathbb{R}}$ be the disintegration of π_{lc} with respect to μ, so that $\pi_{lc}(dx, dy) = \mu(dx)\pi^x_{lc}(dy)$. It follows that for any $\mu' \le \mu$, $\pi'(dx, dy) := \mu'(dx)\pi^x_{lc}(dy)$ is again a left-curtain coupling. Decompose $\mu = \mu_c + \sum_n \alpha_n \delta_{x_n}$ into continuous and discrete parts, respectively. The desired representation of π_{lc} through graphs of functions can then be obtained by pasting together the representations of $\pi_c(dx, dy) := \mu_c(dx)\pi^x_{lc}(dy)$ and $\pi_d(dx, dy) := \sum_n \alpha_n \delta_{x_n}(dx)\pi^{x_n}_{lc}(dy)$. Note that in the case of π_c, the result of Theorem 4.1.1 follows from the original theorem of Beiglböck and Juillet [10], while the case of π_d follows from the arguments given in Section 4.1.*

4.2 Robust bounds for the American put

Our motivation for the study of the left-curtain mapping came from a connection with the robust pricing of American puts. Recall that in robust or model-independent pricing (Hobson [54, 56]) the idea is that instead of writing down a model for the asset price (for example, geometric Brownian motion or a stochastic volatility model) we consider the class of all models for which the discounted asset price is a martingale and which are consistent with the prices of traded vanilla options. Then, given an exotic option which we would like to price, we search over this class of models to find the range of feasible model-based prices.

Typically the set of traded vanilla options is taken to be the set of European-style puts and calls. Given a family of European puts and calls for a fixed maturity and a continuum of strikes we can infer the law of the asset price at that maturity (under the market measure used for pricing). Given the prices of puts and calls for a sequence of maturities we can infer the marginal distributions of the asset price, but not the joint distributions. Then, working under the bond-price numeraire, the class of asset price processes which are consistent with the prices of traded vanilla options can be identified with the class of martingales with given marginals. The problem of finding the robust upper bound on the price of an American-style option becomes a search over consistent martingale models of the model-based price of the American option, see Neuberger [83], Hobson and Neuberger [60] and Bayraktar et al. [6]. Crucially, the primal pricing problem can be identified with a dual hedging problem.

When the American-style option is an American put and the number of candidate exercise dates is two, in Chapter 3 we found the robust upper bound under an assumption that the law of the underlying at the first exercise date is continuous. It turns out that the consistent model for which the American put has highest price is the model associated with the left-curtain coupling of Beiglböck and Juillet [10]. Here we briefly explain how the results of Chapter 3 extend to the atomic case, and why the atomic case is important. There is a subtlety in the case with atoms which is not present when there are no atoms, and to deal with this subtlety we need the extension of the left-curtain coupling to the atomic case as constructed in this chapter.

We are interested in pricing the American put which, in discounted units has strike K_1 at maturity 1 and strike K_2 at maturity 2, with $K_2 < K_1$, see Chapter 3 for further details. The expected payoff arising from a given joint law $\pi \in \Pi(\mu, \nu)$ and a given stopping rule τ taking values in $\{1, 2\}$ is

$$\phi_\pi(\tau) = \mathbb{E}^{\mathcal{L}(X_1,X_2)\sim\pi}\left[(K_1 - X_1)^+ 1_{\{\tau=1\}} + (K_2 - X_2)^+ 1_{\{\tau=2\}}\right]$$

Here X represents the discounted asset price, and is a martingale with joint law π.

For a Borel set B we can let τ_B be the stopping rule $\tau_B = 1$ if $X_1 \in B$ and $\tau_B = 2$ otherwise. Then the payoff under the stopping rule τ_B is $\Phi_\pi(B) := \phi_\pi(\tau_B)$ and the American put price under the model is $\overline{\Phi}_\pi = \sup_B \Phi_\pi(B)$.

Bayraktar et al. [6][1] define the upper bound on the price of the American

[1] [6] contains many interesting and important results and this is just a small element of the paper

put to be

$$\mathcal{P}_{BHZ} = \sup_{\pi \in \Pi(\mu,\nu)} \overline{\Phi}_\pi = \sup_{\pi \in \Pi(\mu,\nu)} \sup_B \Phi_\pi(B).$$

The definition of the model-independent upper bound on the price of the American put given by Neuberger [83] and Hobson and Neuberger [60] is different. Suppose $(\mathcal{S} = (\Omega, \mathcal{F}, \mathbb{F}, \mathbb{P}), X = (X_0, X_1, X_2))$ is a (μ,ν)-consistent model. The model-based price of the American put is

$$\mathcal{A}(\mathcal{S}, X) = \sup_{\tau \in \mathcal{T}_{1,2}(\mathcal{S})} \mathbb{E}^{\mathcal{S},X}[(K_\tau - X_\tau)^+]$$

where $\mathcal{T}_{1,2}(\mathcal{S})$ is the set of all $\mathbb{F}$-stopping times taking values in $\{1,2\}$. Then (Neuberger [83], Hobson and Neuberger [60]) the highest model-based price is

$$\mathcal{P}_N = \sup_{\mathcal{S},X} \mathcal{A}(\mathcal{S}, X) \tag{4.11}$$

where the supremum is taken over (μ,ν)-consistent models.

Set $\overline{\Omega} = \mathbb{R} \times \mathbb{R} = \{\omega = (\omega_1, \omega_2)\}$, $\overline{\mathcal{F}} = \mathcal{B}(\Omega)$ and $(X_1(\omega), X_2(\omega)) = (\omega_1, \omega_2)$, and let $\overline{\mathbb{P}}$ be such that $\mathcal{L}(X_1) = \mu$ and $\mathcal{L}(X_2) = \nu$. Let $\overline{\mathcal{F}}_0 = \{\emptyset, \Omega\}$, $\overline{\mathcal{F}}_1 = \sigma(X_1)$ and $\overline{\mathcal{F}}_2 = \sigma(X_1, X_2)$. If $\overline{\mathcal{S}} = (\overline{\Omega}, \overline{\mathcal{F}}, \overline{\mathbb{F}}, \overline{\mathbb{P}})$ then $(\overline{S}, \overline{X})$ is a (μ,ν)-consistent model.

Consistent models of the form $(\overline{S}, \overline{X})$ can be identified with martingale couplings π. It follows that $\mathcal{P}_{BHZ} \leq \mathcal{P}_N$, the inequality following from the fact that in principle we could work on a richer probability space. It follows from our work in Chapter 3 that if μ is continuous then the martingale coupling associated with the optimiser for either $\mathcal{P}_{BHZ}$ or $\mathcal{P}_N$ is the left-curtain coupling and $\mathcal{P}_{BHZ} = \mathcal{P}_N$. Our interest in extending the left-curtain mapping arose from the fact that when μ has atoms we may have $\mathcal{P}_{BHZ} < \mathcal{P}_N$. Then, in order to construct the optimiser for $\mathcal{P}_N$ we need an appropriate extension of the left-curtain coupling.

4.2.1 The trivial law for μ

The difference between the modelling approaches of Bayraktar et al. [6] and Hobson and Neuberger [60] can be illustrated most simply when $\mu = \delta_w$. Also for simplicity we assume ν has a continuous law with mean w.

In the framework of Bayraktar et al. [6], since the filtration generated by X is still trivial at time 1, the only choices facing the holder of the American put are either to always stop at time 1, or to never stop at time 1. The expected payoff of the American put does not depend on the martingale coupling and thus

$$\begin{aligned}\mathcal{P}_{BHZ} = \sup_{\pi\in\Pi(\mu,\nu)} \max\{\Phi_\pi(\Omega),\Phi_\pi(\emptyset)\} &= \sup_{\pi\in\Pi(\mu,\nu)} \max\{\phi_\pi(1),\phi_\pi(2)\} \\ &= \max\left\{(K_1-w)^+, \int (K_2-z)^+\nu(dz)\right\}.\end{aligned}$$

On the other hand we can construct a richer model which is (δ_w,ν)-consistent. Set $\Omega=(0,1)\times(0,1)$ and let $\mathbb{P}$ be Lebesgue measure on Ω. Let (U,V) be a pair of independent uniform random variables, let $(\mathcal{F}_0=\{\emptyset,\Omega\},\mathcal{F}_1=\sigma(U),\mathcal{F}_2=\sigma(U,V))$ and let $X_0=X_1=w$ and $X_2=Y$, where $Y=Y(U,V)$ is as given in (4.2) with $G(u)\equiv w$. Here (R,S) are a pair of monotonic functions with

$$u=\int_{R(u)}^{S(u)}\nu(dz), \qquad\qquad 0=\int_{R(u)}^{S(u)}(z-w)\nu(dz). \tag{4.12}$$

In this way we construct a (μ,ν)-consistent model.

Under this model the value $A(u)$ of the American put under the stopping rule τ_u where $\tau_u=1$ if $U<u$ and $\tau_u=2$ otherwise is

$$\begin{aligned}A(u) &= \mathbb{E}[(K_1-X_1)^+1_{\{\tau_u=1\}}+(K_2-X_2)^+1_{\{\tau_u=1\}}] \\ &= (K_1-w)u+\int_{-\infty}^{R(u)}(K_2-z)^+\nu(dz)+\int_{S(u)}^{\infty}(K_2-z)^+\nu(dz).\end{aligned}$$

It follows that $\mathcal{P}_N\geq\sup_{u\in[0,1]}A(u)$. (In the next section we will argue that there is equality here.) Note that $\mathcal{P}_{BHZ}=A(0)\vee A(1)$, so that $\mathcal{P}_N>\mathcal{P}_{BHZ}$ will follow if $\sup_{u\in[0,1]}A(u)>A(0)\vee A(1)$.

For a simple example, suppose $w=1$ and $\nu=U[0,2]$; suppose $K_1=\frac{5}{4}$ and $K_2=1$. Then $R(u)=1-u$ and $S(u)=1+u$. We have

$$A(u)=\frac{u}{4}+\int_0^{1-u}(1-z)\frac{dz}{2}=\frac{1+u-u^2}{4}.$$

Then $P_N\geq\max_{u\in[0,1]}A(u)=\frac{5}{16}>\frac{1}{4}=A(0)\vee A(1)=\mathcal{P}_{BHZ}$.

Remark 4.2.1. *In our set-up there are two possible exercise times for the American put, denoted 1 and 2, and we construct a martingale $(X_0=w,X_1,X_2)$ to match the marginals at these times. But if $\mathcal{L}(X_1)=\delta_{X_0}$ the problem can be recast as a problem for a stochastic process $\tilde{X}=(\tilde{X}_0,\tilde{X}_1)$ where $\tilde{X}_0=X_0=X_1$ and $\tilde{X}_1=X_2$. We also set $\tilde{\tau}=\tau-1$; then $\tilde{\tau}\in\{0,1\}$ and $\tilde{\tau}=0$ corresponds to immediate exercise. Put another way, one way to allow for immediate exercise of the American put, is to introduce an additional point (labelled 1) into the time-indexing set and to require*

$\mathcal{L}(X_1) = \delta_{X_0}$. *For this reason it is very natural for* μ *to have a trivial law, if we want to allow immediate exercise.*

4.2.2 Tightness of the bound for a trivial law μ

Our goal in this section is to show that $\mathcal{P}_N = \sup_{u \in [0,1]} A(u)$. We do this by finding an upper bound on the American put pricing problem and then showing that this bound is equal to $\sup_{u \in [0,1]} A(u)$.

Let ψ be a convex function with $\psi(z) \geq (K_2 - z)^+$. Let $\phi(z) = ((K_1 - z)^+ - \psi(z))^+$ and let $\theta(z) = -\psi'_+(z)$, where ψ'_+ is the right derivative. Then, for all x_1 and x_2

$$\begin{aligned} (K_1 - x_1)^+ &\leq \phi(x_1) + \psi(x_2) + (x_2 - x_1)\theta(x_1), \\ (K_2 - x_2)^+ &\leq \phi(x_1) + \psi(x_2). \end{aligned}$$

It follows that for any set $B \in \mathcal{F}$ and for every ω,

$$(K_1 - X_1)^+ I_B + (K_2 - X_2)^+ 1_{B^C} \leq \phi(X_1) + \psi(X_2) + (X_2 - X_1)\theta(X_1) 1_B.$$

In particular, if we think of B as the set of scenarios on which the put is exercised at time 1 then we have that the payoff of the American put is bounded above by the sum of the European-style payoffs ϕ and ψ and the gains from trade from a strategy which involves holding $\theta(X_1)$ units of the underlying over the time-interval $(1, 2]$, provided the put was exercised at time 1. Then, for $B \in \mathcal{F}_1$

$$\begin{aligned} \mathbb{E}[(K_{\tau_B} - X_{\tau_B})^+] &\leq \mathbb{E}[\phi(X_1)] + \mathbb{E}[\psi(X_2)] \\ &= \int ((K_1 - x)^+ - \psi(x))^+ \mu(dx) + \int \psi(y)\nu(dy). \end{aligned}$$

In our context with $\mu = \delta_w$ this simplifies to $((K_1 - w)^+ - \psi(w))^+ + \int \psi(y)\nu(dy) =: \mathcal{D}(\psi)$. Let $\mathcal{D} = \inf_\psi \mathcal{D}(\psi)$ (where the infimum is taken over convex ψ with $\psi(z) \geq (K_2 - z)^+$). $\mathcal{D}$ forms an upper bound for the price of the American option under any consistent model and hence $\mathcal{P}_N \leq \mathcal{D}$.

Let R and S be defined as in Section 4.1.1. Let $P_\nu(z) = \int (z - x)^+ \nu(dx)$. Then (4.12) can be rewritten as $u = P'_\nu(S(u)) - P'_\nu(R(u))$ together with

$$(S(u) - w)P'_\nu(S(u)) - P_\nu(S(u)) = (R - w)P'_\nu(R(u)) - P_\nu(R(u)). \tag{4.13}$$

Fix $K_2 < K_1$ with $K_1 > w$ and define $\Lambda_w : (-\infty, K_2 \wedge w) \times (K_1, \infty) \mapsto \mathbb{R}$ by

$$\Lambda_w(r,s) = \frac{K_1 - w}{s - w} - \frac{(K_2 - r) - (K_1 - w)}{w - r}.$$

Since ν is continuous by assumption, R and S are strictly decreasing and strictly increasing, respectively. Define $u_w = \inf\{u \in (0,1) : R(u) < K_2 \text{ and } S(u) > K_1\}$, and for $u \in (u_w, 1)$ set $\bar{\Lambda}_w(u) = \Lambda_w(R(u), S(u))$. It follows that $\bar{\Lambda}_w$ is strictly decreasing.

Suppose that the smallest closed interval containing the support of ν, $\mathcal{I}_\nu = [\ell_\nu, r_\nu]$, is such that $\frac{K_1-w}{r_\nu-w} < \frac{(K_2-\ell_\nu)-(K_1-w)}{w-\ell_\nu}$ (this will follow if $0 = \ell_\nu < w < r_\nu = \infty$ and $K_2 > K_1 - w$, for example). This assumption is sufficient to guarantee that there exists $u^* \in (u_w, 1)$ such that $\bar{\Lambda}_w(u^*) = 0$. Then $S^* := S(u^*) > K_1 > K_2 > R(u^*) =: R^*$. Also $\bar{\Lambda}_w(u^*) = 0$ implies $\frac{K_1-w}{S^*-w} = \frac{K_2-R^*}{S^*-R^*}$. For the model constructed in Section 4.1.1 we have

$$\begin{aligned}
\sup_{u\in[0,1]} A(u) \geq A(u^*) &= (K_1 - w)^+ u^* + \int_{-\infty}^{R^*} (K_2 - z)^+ \nu(dz) \\
&= (K_1 - w)[P_\nu'(S^*) - P_\nu'(R^*)] + P_\nu(R^*) + (K_2 - R^*)P_\nu'(R^*).
\end{aligned}$$

Conversely, let $\Theta = \frac{K_1-w}{S^*-w} = \frac{K_2-R^*}{S^*-R^*} = \frac{(K_2-R^*)-(K_1-w)}{w-R^*} \in (0,1)$ and let $\psi^*(x) = \Theta(S^*-x)^+ + (1-\Theta)(R^*-x)^+$. Note that by design $\psi^*(R^*) = \Theta(S^*-R^*) = (K_2 - R^*)$ so that $\psi^*(z) \geq (K_2 - z)^+$. Further, $\psi^*(w) = \Theta(S^* - w) = (K_1 - w)$ so that $\phi^*(w) = 0$ where $\phi^*(z) = ((K_1 - z)^+ - \psi^*(z))^+$. Then $\mathcal{D} \leq \Theta P_\nu(S^*) + (1 - \Theta)P_\nu(R^*) = \mathcal{D}(\psi^*)$.

Now consider $\mathcal{D}(\psi^*) - A(u^*)$. Using (4.13) for the second equality and the alternative characterisations of Θ for the third we have

$$\begin{aligned}
\mathcal{D}(\psi^*) - A(u^*) &= \Theta(P_\nu(S^*) - P_\nu(R^*)) - (K_1 - w)[P_\nu'(S^*) - P_\nu'(R^*)] - (K_2 - R^*)P_\nu'(R^*) \\
&= P_\nu'(S^*)[\Theta(S^* - w) - (K_1 - w)] - P_\nu'(R^*)[\Theta(w - R^*) - (K_1 - w) + (K_2 - R^*)] \\
&= 0.
\end{aligned}$$

Then $\mathcal{D}(\psi^*) = A(u^*) \leq \sup_{u\in[0,1]} A(u) \leq \mathcal{P}_N \leq \mathcal{D} \leq \mathcal{D}(\psi^*)$. It follows that this chain of inequalities is in fact a chain of equalities and $\mathcal{P}_N = \sup_{u\in[0,1]} A(u)$. Moreover, we have identified an optimal model and an optimal stopping rule. The model which yields the highest price for the American put is our extension of the left-curtain coupling.

4.2.3 American puts with a general time-1 law

We seek to generalise the arguments of the previous section to allow for non-trivial initial laws. Define $\Lambda = \Lambda(r, g, s)$ via

$$\Lambda(r, g, s) = \frac{K_1 - g}{s - g} - \frac{(K_2 - r) - (K_1 - g)}{g - r}$$

Suppose we are in the case of continuous μ. Define $\hat{\Lambda}(x) = \Lambda(f(x), x, g(x))$ where f and g are the lower and upper functions which arise in the Beiglböck-Juillet [10] characterisation of the left-curtain martingale coupling. In our notation this can be written as $\hat{\Lambda}(x) = \Lambda((R \circ G^{-1})(x), x, (S \circ G^{-1})(x))$. The fundamental insight in Hobson and Norgilas [61] is that, in the case of continuous μ, the cheapest superhedge can be described in terms of a simple portfolio of European-style puts whose strikes depend on quantities which arise from looking for the root x^*, if any, of $\hat{\Lambda}(\cdot) = 0$. Moreover the most expensive model is the model described by the left-curtain coupling, and an optimal exercise rule is to exercise at time-1 if and only if $X_1 < x^*$. Hobson and Norgilas [61] identify four archetypes of hedging portfolios. The first two cases correspond to when there is a root to $\hat{\Lambda} = 0$ and when $\hat{\Lambda} < 0$ for all x. (The remaining cases correspond to cases where $\hat{\Lambda}$ is discontinuous, and jumps downwards over the value 0.)

In the case with atoms in μ we cannot use $\hat{\Lambda}$ directly since G^{-1} has jumps. Instead, following the analysis in Section 4.2.1 we define $\bar{\Lambda}(u) = \Lambda(R(u), u, S(u))$, and look for solutions, if any, to $\bar{\Lambda}(\cdot) = 0$. We may still have the cases where $\bar{\Lambda} < 0$ for all $u \in (0, 1)$ or where $\bar{\Lambda}(\cdot)$ jumps over zero, but these cases can be dealt with as in [61]. The new case is when the root u^* of $\bar{\Lambda} = 0$ occurs in an interval $(\underline{u}, \overline{u}]$ over which G is constant. This means that there is an atom of μ at $G(u^*)$. See Figure 4.5. A model which maximises the price of the American put is the extended left-curtain martingale coupling model, and the optimal stopping rule is to exercise at time-1 whenever $X_1 < G(u)$ and to sometimes exercise when $X_1 = G(u)$. When $X_1 = G(u)$ the optimal stopping rule is to exercise precisely when $U \in (\underline{u}, u^*]$ and to wait if $U \in (u^*, \overline{u}]$. Because R and S are monotonic over $(\underline{u}, \overline{u}]$ paths with low future variability are exercised at time-1 whereas on paths with high future variability exercise is delayed to time-2.

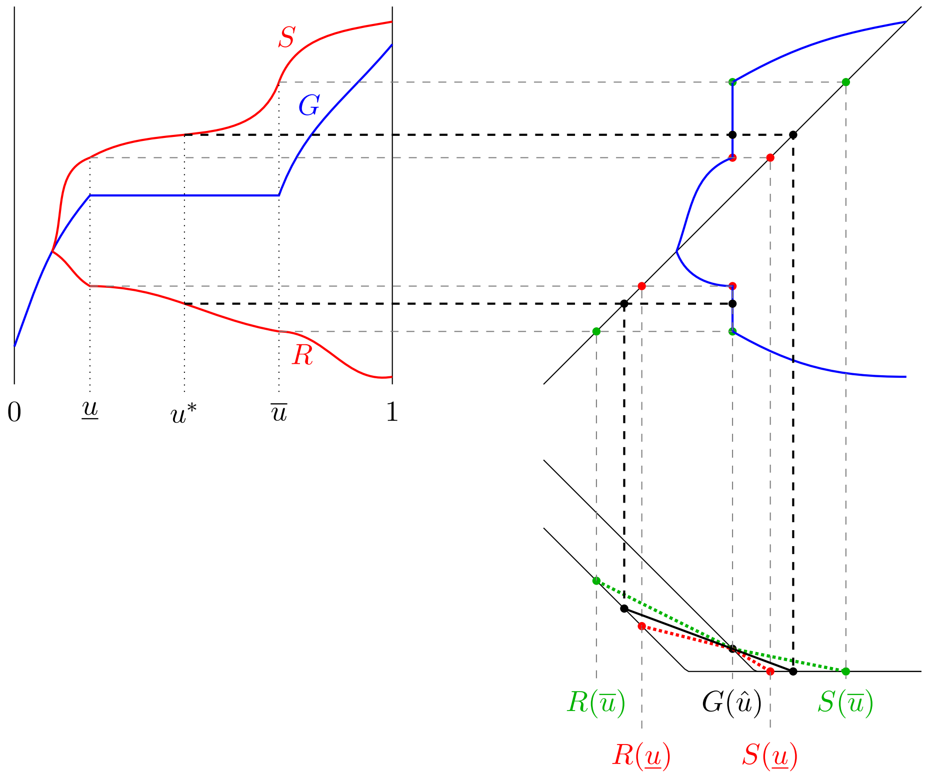

Figure 4.5: Finding the optimal hedge for general measures. The initial law μ has an atom of size $\overline{u} - \underline{u}$. Moreover, the piecewise linear curve joining $(R(\underline{u}), K_2 - R(\underline{u}))$, $(G(\hat{u}), K_1 - G(\hat{u}))$ and $(S(\underline{u}), 0)$ is concave (where $\hat{u}$ is any element of $(\underline{u}, \overline{u}])$, whereas the piecewise linear curve joining $(R(\overline{u}), K_2 - R(\overline{u}))$, $(G(\hat{u}), K_1 - G(\hat{u}))$ and $(S(\overline{u}), 0)$ is convex. There exists $u^* \in (\underline{u}, \overline{u}]$ such that $(R(u^*), K_2 - R(u^*))$, $(G(\hat{u}), K_1 - G(\hat{u}))$ and $(S(u^*), 0)$ all lie on a straight line. The figure describes the optimal coupling (via (U, V) and (4.2)) and the optimal exercise strategy for the American put is to exercise at time-1 if $U \leq u^*$.

4.3 Proofs

Proof of Lemma 4.1.8. We begin our study of the upper bound on S_n by considering the case of a single starting measure μ and fixed target law ν. First we assume that μ and ν are regular (no atoms and no intervals within the support with no mass), before extending to the general case. Then we consider what happens when we consider $\mu_n \uparrow_{cx} \mu$.

Suppose μ and ν have no atoms and no intervals within the support with no mass. Then G_μ is continuous and strictly increasing. Fix $u \in (0, 1)$ and let $\ell_1 \equiv \ell_1^u$ be the tangent to P_μ with slope u. See Figure 4.6. By construction this tangent meets P_μ at $G = G_\mu(u)$. Let $H = H(u)$ be the point where the tangent crosses the x-axis. Let $\ell_2 \equiv \ell_2^u$ be the tangent to P_ν with slope greater than u which passes through $(G, P_\mu(G))$; this tangent meets P_ν at the x-coordinate $J = J(u) = J_{\mu,\nu}(u)$.

We now show that $S(u) \leq J$.

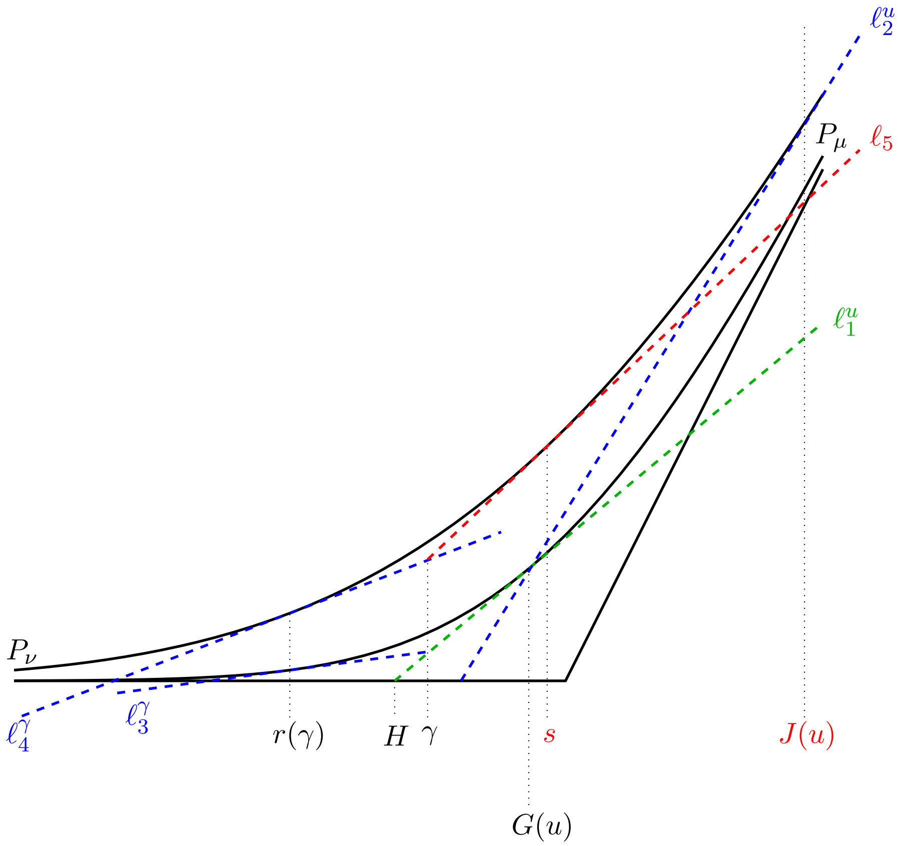

Figure 4.6: Construction of function J that bounds the upper function S on $(0, 1)$.

Choose $\gamma \in [H, G)$. Let ℓ_3^γ be the tangent to P_μ which passes through $(\gamma, \ell_1(\gamma))$ and has slope less than u. Suppose this tangent meets P_μ at $r = r(\gamma)$; the slope of the tangent is $P'_\mu(r)$. Let ℓ_4^γ be the tangent to P_ν at r. Finally, let ℓ_5^γ be the line passing through $(\gamma, \ell_4^r(\gamma))$ with slope $u + P'_\nu(r) - P'_\mu(r)$.

If there exists γ such that ℓ_5^γ is a tangent to P_ν (meeting P_ν at s say), then (r, G, s) satisfy

$$\int_r^G w^i \mu(dw) = \int_r^s w^i \nu(dw) \qquad i = 0, 1 \tag{4.14}$$

(and moreover $\gamma = \int_r^G w\mu(dw) / \int_r^G \mu(dw) = \int_r^s w\nu(dw) / \int_r^s \nu(dw)$ is the barycentre of the measures $\mu|_{(r,G)}$ and $\nu|_{(r,s)}$).

For each u there may be multiple γ which lead to a triple (r, G, s) which satisfies (4.14). We show that in each case $s \leq J$. It follows that $S(u) \leq J$.

Suppose $\ell_4^\gamma(\gamma) \leq \ell_3^\gamma(\gamma) = \ell_1(\gamma)$. Then necessarily $P'_\nu(r) < P'_\mu(r)$ and ℓ_5^γ lies below ℓ_1 to the right of γ; in particular ℓ_5^γ stays below P_μ to the right of γ and cannot be a tangent to P_ν. Hence if (r, G, s) satisfies (4.14) we must have $\ell_4^\gamma(\gamma) > \ell_1(\gamma)$. Then, if ℓ_5^γ is a tangent to P_ν we must have that the point of tangency is below J.

In the above we used the regularity assumptions on μ and ν to conclude that there was a unique tangent to $P_\cdot \in \{P_\mu, P_\nu\}$ at a given point, and that there was a unique point at which $P_\cdot$ had a given slope. If μ or ν is not regular then, for fixed u, there may be multiple quintiles G, multiple points r and multiple tangents to P_ν at r. The point is that although there are multiple versions of the construction in this case each candidate triple (r, G, s) satisfying (4.14) has $s \leq J$ where J is defined using an arbitrary point $G \in [G_\mu(u), G_\mu(u+)]$. We define $J_- = J_-(u)$ to be the smallest x-coordinate at which the tangent to P_ν with slope greater than u passing through $(G(u), P_\mu(G(u)))$ meets P_ν and $J_+ = J_+(u)$ to be the largest x-coordinate at which the tangent to P_ν with slope greater than u passing through $(G(u+), P_\mu(G(u)+))$ meets P_ν. We have $S(u) \leq J_-(u) \leq J_+(u)$.

Finally, we want to show that if we approximate μ by μ_n (with $\mu_n \uparrow_{cx} \mu$) then the bound $\limsup S_n \leq J_+$ remains valid, where J_+ is constructed from μ and ν.

Define $K(k) = \overline{\text{argsup}}_\kappa \frac{P_\nu(\kappa) - P_\mu(k)}{\kappa - k}$. The notation $\overline{\text{argsup}}$ is used to indicate that where there are multiple elements in the argsup we choose the largest one. Then K is increasing and right continuous in k. Note that $J_+(u) = K(G(u+))$. In a similar fashion we can define K_n and J_n using P_{μ_n} in place of P_μ. (The target law is assumed fixed throughout.) Since $P_{\mu_n}(k) \uparrow P_\mu(k)$ and K is right-continuous we have $K_n(k) \downarrow K(k)$. Then, for $\epsilon > 0$,

$$\limsup_n J_n(u) = \limsup_n K_n(G_n(u+)) \leq \limsup_n K_n(G(u+) + \epsilon) \leq K(G(u+) + \epsilon).$$

Since ϵ is arbitrary and K is right continuous, $\limsup S_n(u) \leq \limsup_n J_n(u) \leq J_+(u)$.

□

Proof of Lemma 4.1.10. As for the proof of Lemma 4.1.8 we begin by considering a single initial law μ, and supposing that μ and ν are regular.

Fix $u \in (0, 1)$ and let ℓ_1 be the tangent to P_μ with slope u. Let $H = H(u)$ be the point where this tangent crosses the x-axis. Suppose that ℓ_1 is not a tangent to P_ν. Then ℓ_1 must lie strictly below P_ν. There exists $\epsilon = \epsilon(u) > 0$ such that the line passing through (H, ϵ) with slope $u + \epsilon$ lies below P_ν. Now choose $j = j(u)$ such that the tangents to P_μ and P_ν at j both have slope less than ϵ and both cross the line $y = x$ below ϵ. Then $R(u) \geq j$.

To see this let γ be the x-coordinate of the point where the tangent to P_μ at j crosses ℓ_1. Then if ℓ_4 is the tangent to P_ν at j then $\ell_4(\gamma) < \epsilon$; if ℓ_5 is the line passing through $(\gamma, \ell_4(\gamma))$ with slope $u + P'_\nu(j) - P'_\mu(j) < u + \epsilon$, then by our defining assumption on ϵ, ℓ_5 lies below P_ν. Hence $R(u) > j$.

We can extend the result to irregular measures, and to $\liminf R_n(u)$ by similar techniques as for S. The only extra issue that arises is our assumption that ℓ^1 is not a tangent to P_ν. But, if for each n, ℓ^1 is a tangent to P_ν, then the same is certainly true in the limit. Then there must exist x such that $\ell_1(x) = P_\mu(x) = P_\nu(x)$ and then $S(u) \leq x \leq G(u+)$. This case is excluded by hypothesis. $\square$

Bibliography

[1] Aksamit A., Deng S., Obłój J., Tan X.: *Robust pricing-hedging duality for American options in discrete time financial markets.* Mathematical Finance, 29:861–897, (2019).

[2] Andersen L., Broadie M.: *Primal-dual simulation algorithm for pricing multidimensional American options.* Management Science, 50(9):1222–1234, (2004).

[3] Ankirchner S., Klein M., Kruse T.: *A verification theorem for optimal stopping problems with expectation constraints.* Applied Mathematics & Optimization, 1–33, (2015).

[4] Assaf D., Goldstein L., Samuel-Cahn E.: *Ratio prophet inequalities when the mortal has several choices.* The Annals of Applied Probability, 12(3):972–984, (2002).

[5] Azéma J., Yor M.: *Une solution simple au problème de Skorokhod.* In Séminaire de probabilités XIII, vol. 721 of Lecture Notes in Mathematics, 90–115, Springer, Berlin, (1979).

[6] Bayraktar E., Huang Y. J., Zhou Z.: *On hedging American options under model uncertainty.* SIAM Journal on Financial Mathematics, 6(1):425–447, (2015).

[7] Beiglböck, M., Cox A.M.G., Huesmann, M.: *The geometry of multi-marginal Skorokhod Embedding.* Available online at: arXiv:1705.09505, (2017).

[8] Beiglböck M., Henry-Labordère P., Penkner F.: *Model-independent bounds for option prices—mass transport approach.* Finance and Stochastics, 17(3):477–501, (2013).

[9] Beiglböck M., Henry-Labordère P., Touzi N.: *Monotone martingale transport plans and Skorokhod embedding.* Stochastic Processes and their Applications, 127(9):3005–3013, (2017).

[10] Beiglböck M., Juillet N.: *On a problem of optimal transport under marginal martingale constraints.* The Annals of Probability, 44(1):42–106, (2016).

[11] Beiglböck M., Schachermayer W., Veliyev B.: *A short proof of the Doob-Meyer theorem.* Stochastic Processes and their Applications, 122(4):1204–1209, (2012).

[12] Belomestny D., Bender C., Schoenmakers J.: *True upper bounds for Bermudan products via non-nested Monte Carlo.* Mathematical Finance, 19(1):53–71, (2009).

[13] Belomestny D.: *Solving optimal stopping problems via empirical dual optimization.* The Annals of Applied Probability, 23(5):1988–2019, (2013).

[14] Billingsley P.: *Convergence of probability measures.* John Wiley & Sons, (2013).

[15] Blumenthal R. M., Getoor R. K.: *Markov processes and potential theory.* Courier Corporation, (2007).

[16] Breeden D.T., Litzenberger R.H.: *Prices of state-contingent claims implicit in option prices.* Journal of Business, 621–651, (1978).

[17] Brenier Y.: *Décomposition polaire et réarrangement monotone des champs de vecteurs.* Comptes Rendus de l'Académie des Sciences, Series I - Mathematics, Paris, 305(19): 805-808, (1987).

[18] Brown H., Hobson D.G., Rogers L.C.G.: *Robust hedging of barrier options.* Mathematical Finance, 11(3):285–314, (2001).

[19] Carr P., Lee R.: *Hedging variance options on continuous semi-martingales.* Finance and Stochastics, 14(2):179–207, (2010).

[20] Chacon, R.V.: *Potential processes.* Transactions of the American Mathematical Society, 226:39–58, (1977).

[21] Chacon, R.V., Walsh J.B.: *One-dimensional potential embedding.* In Séminaire de Probabilités X, vol. 511 of Lecture Notes in Mathematics, 19–23, Springer, Berlin, (1976).

[22] CHERNOFF H.: *Sequential tests for the mean of a normal distribution.* In Proceedings of the Fourth Berkeley Symposium on Mathematical Statistics and Probability, University of California Press, 1:79–91, (1961).

[23] ÇINLAR E., JACOD J., PROTTER P., SHARPE M. J.: *Semimartingales and Markov processes.* Zeitschrift für Wahrscheinlichkeitstheorie und verwandte Gebiete, 54(2):161–219, (1980).

[24] COX A.M.G.: *Extending Chacon-Walsh: minimality and generalised starting distributions.* In Séminaire de probabilités XLI, vol. 1934 of Lecture Notes in Mathematics, 233–264, Springer, Berlin, (2008).

[25] COX A.M.G., HOEGGERL C.: *Model-independent no-arbitrage conditions on American put options.* Mathematical Finance, 26(2):431–458, (2016).

[26] COX A.M.G., OBŁÓJ J.: *Robust pricing and hedging of double no-touch options.* Finance and Stochastics, 15(3):573–605, (2011).

[27] COX A.M.G., WANG J.: *Root's barrier: construction, optimality and applications to variance options.* The Annals of Applied Probability, 23(3):859–894, (2013).

[28] CREPÉY S., MATOUSSI A.: *Reflected and doubly reflected BSDEs with jumps: a priori estimates and comparison.* The Annals of Applied Probability, 18(5):2041-2069, (2008).

[29] DAVIS M., KARATZAS I.: *A deterministic approach to optimal stopping.* Probability, Statistics and Optimisation (ed. FP Kelly), NewYork Chichester: John Wiley & Sons Ltd, 455–466, (1994).

[30] DAYANIK S., KARATZAS I.:, *On the optimal stopping problem for one-dimensional diffusions.* Stochastic Processes and their Applications, 107(2):173–212, (2003).

[31] DELLACHERIE C., MEYER P.-A.: *Probabilities and potential, B.* Volume 72 of North-Holland Mathematics Studies, (1982).

[32] DESAI V. V., FARIAS V. F., MOALLEMI C. C.: *Pathwise optimization for optimal stopping problems.* Management Science, 58(12):2292–2308, (2012).

[33] DOLÉANS C.: *Existence du processus croissant naturel associé à un potentiel de la classe (D).* Probability Theory and Related Fields, 9(4):309–314, (1968).

[34] Dolinsky Y., Soner M.: *Martingale optimal transport and robust hedging in continuous time.* Probability Theory and Related Fields, 160(1-2):391–427, (2014).

[35] Dupire, B.: *Pricing with a smile.* Risk Magazine, 18–20, (1994).

[36] Dynkin E. B.: *The optimum choice of the instant for stopping a Markov process.* Soviet Mathematical Doklady, 4:627–629, (1963).

[37] Dynkin E. B.: *Markov processes.* Vol. 2, Springer, (1965).

[38] El Karoui N.: *Les aspects probabilistes du contrôle stochastique.* In Ecole d'Eté de Probabilités de Saint-Flour IX-1979, 73–238, Springer, Berlin, Heidelberg, (1981).

[39] El Karoui N., Kapoudjian C., Pardoux E., Peng S., Quenez M. C.: *Reflected solutions of backward SDE's, and related obstacle problems for PDE's.* The Annals of Probability, 25(2):702–737, (1997).

[40] El Karoui N., Lepeltier J. P., Millet A.: *A probabilistic approach to the réduite in optimal stopping.* Probability and Mathematical Statistics, 13(1):97–121, (1992).

[41] Fleming W. H., Soner H. M.: *Controlled Markov processes and viscosity solutions.* Vol. 25, Springer Science & Business Media, (2006).

[42] Galcihon A., Henry-Labordère P., Touzi N.: *A stochastic control approach to no-arbitrage bounds given marginals with an application to lookback options.* The Annals of Applied Probability, 24(1):313–336, (2014).

[43] Getoor R. K.: *Markov processes: Ray processes and right processes.* Vol. 440, Springer, (2006).

[44] Glasserman P., Yu B.: *Number of paths versus number of basis functions in American option pricing.* The Annals of Applied Probability, 14(4):2090–2119, (2004).

[45] Grigelionis B. I., Shiryaev A. N.: *On Stefan?s problem and optimal stopping rules for Markov processes.* Theory of Probability & Its Applications, 11(4):541–558, (1966).

[46] Gyöngy I.: *Mimicking the one-dimensional marginal distributions of processes having an Itô differential.* Probability Theory and Related Fields, 71:501–516, (1986).

[47] Gyöngy I., Šiška D.: *On randomized stopping.* Bernoulli, 14(2):352–361, (2008).

[48] Hamadène S., Ouknine Y.: *Reflected backward stochastic differential equation with jumps and random obstacle.* Electronic Journal of Probability, 8, (1993).

[49] Haugh M. B., Kogan L.: *Pricing American options: a duality approach.* Operations Research, 52(2):258–270, (2004).

[50] Henry-Labordère P., Touzi N.: *An explicit martingale version of the one-dimensional Brenier theorem.* Finance and Stochastics, 20(3):635–668, (2016).

[51] Herrmann S., Stebegg F.: *Robust pricing and hedging around the globe.* Available online at: arXiv:1707.08545, (2017).

[52] Hill T.: *Prophet inequalities and order selection in optimal stopping problems.* Proceedings of the American Mathematical Society, 88(1):131–137, (1983).

[53] Hirsch F., Roynette B.: *A new proof of Kellerer's theorem.* ESAIM: Probability and Statistics, 16:48–60, (2012).

[54] Hobson D.G.: *Robust hedging of the lookback option.* Finance and Stochastics, 2(4):329–347, (1998).

[55] Hobson D.G.: *The maximum maximum of a martingale.* In Séminaire de probabilités XXXII, vol. 1686 of Lecture notes in Mathematics, 250–263, Springer, Berlin, (1998).

[56] Hobson D.G.: *The Skorokhod embedding problem and model-independent bounds for option prices.* In Paris-Princeton Lectures on Mathematical Finance 2010, 267–318, Springer, Berlin, (2011).

[57] Hobson D.G., Klimmek M.: *Robust price bounds for the forward starting straddle.* Finance and Stochastics, 19(1):189–214, (2015).

[58] Hobson D.G., Neuberger A.: *Robust bounds for forward start options.* Mathematical Finance, 22(1):31–56, (2012).

[59] Hobson D.G., Neuberger A.: *On the value of being American.* Finance and Stochastics, 21(1):285–329, (2017).

[60] Hobson D.G., Neuberger A.: *More on hedging American options under model uncertainty.* Available online at: arXiv:1604.02274, (2016).

[61] Hobson D.G., Norgilas D.: *Robust bounds for the American Put.* Finance and Stochastics, 23(2):359–395, (2019).

[62] Hobson D.G., Norgilas D.: *The left-curtain martingale coupling in the presence of atoms.* The Annals of Applied Probability, 29(3):1904–1928, (2019).

[63] Itô K., McKean Jr. H. P.: *Diffusion Processes and Their Sample Paths.* Grundlehren der Mathematischen Wissenschaften, 125, (1965).

[64] Jacka S. D.: *Local times, optimal stopping and semimartingales.* The Annals of Probability, 21(1):329–339, (1993).

[65] Jacod J., Shiryaev A. N.:, *Limit theorems for stochastic processes.* Vol. 288, Springer Science & Business Media, (2013).

[66] Jakubowski A.: *An almost sure approximation for the predictable process in the Doob-Meyer decomposition theorem.* In Séminaire de Probabilités XXXVIII, vol. 1857 of Lecture Notes in Mathematics, 158–164, Springer, Berlin, (2005).

[67] Juillet N.: *Stability of the shadow projection and the left-curtain coupling.* Annales de l'Institut Henri Poincaré, Probabilités et Statistiques, 52(4):1823–1843, (2016).

[68] Kallenberg O. *Foundations of modern probability.* Springer Science & Business Media, (2006).

[69] Kantorovich L. V.: *On the translocation of masses.* In Doklady Akademii Nauk SSSR, 37(7-8):227–229, (1942).

[70] Karatzas I., Shreve S. E.: *Brownian Motion and Stochastic Calculus.* New York: Springer, (1998).

[71] Karatzas I., Shreve S. E.: *Methods of mathematical finance.* Vol 39, New York: Springer, (1998).

[72] Kobylanski M., Quenez M.-C.:, *Optimal stopping time problem in a general framework.* Electronic Journal of Probability, 17(72):1–28, (2012).

[73] Kolodko A., Schoenmakers J. *Upper bounds for Bermudan style derivatives.* Monte Carlo Methods and Applications, 10(3-4):331–343, (2004).

[74] Komlós J. *A generalization of a problem of Steinhaus.* Acta Mathematica Hungarica, 18(1-2):217–229, (1967).

[75] Krylov N. V.: *Controlled diffusion processes.* Vol. 14, Springer Science & Business Media, (2008).

[76] Krylov, N.V.: *On the relation between differential operators of second order and the solutions of differential equations.* Steklov Seminar 1984, Ed N.V. Krylov et al., Inc. Publications Division, New York, 214–229, (1985).

[77] Lindley D.: *Dynamic programming and decision theory.* Journal of the Royal Statistical Society: Series C (Applied Statistics), 10(1): 39–51, (1961).

[78] McKean H. P.: *A free boundary problem for the heat equation arising from a problem of mathematical economics.* Industrial Management Review, 6:32–39, (1965).

[79] Maingueneau M. A.: *Temps d'arrêt optimaux et théorie générale.* In Séminaire de Probabilités XII, vol. 649 of Lecture Notes in Mathematics, 457–467, Springer, Berlin, (1978).

[80] Mikhalevich V. S.: *Bayesian choice between two hypotheses for the mean value of a normal process.* Visn. Kiiv. Univ. Ser. Fiz.-Mat. Nauki 1:101–104, (1958).

[81] Miller C. W.: *Nonlinear pde approach to time-inconsistent optimal stopping.* SIAM Journal on Control and Optimization, 55(1):557–573, (2017).

[82] Monge G.: *Mémoire sur la théorie des déblais et des remblais.* Histoire de l'Académie royale des sciences de Paris, (1781).

[83] Neuberger A.: *Bounds on the American option.* Available online at: SSRN:966333, (2007).

[84] Nutz M., Stebegg F.: *Canonical supermartingale couplings.* The Annals of Probability, 46(6):3351–3398, (2018).

[85] Nutz M., Stebegg F., Tan X.: *Multiperiod martingale transport.* Stochastic Processes and their Applications (in press), available online at: https://doi.org/10.1016/j.spa.2019.05.010, (2019).

[86] Peskir G.: *Principle of smooth fit and diffusions with angles.* Stochastics An International Journal of Probability and Stochastic Processes, 79(3-4):293–302, (2007).

[87] Peskir G.: *A duality principle for the Legendre transform.* Journal of Convex Analysis, 19(3):609–630, (2012).

[88] Peskir G., Shiryaev A. N.: *Sequential testing problems for Poisson processes.* The Annals of Statistics, 28(3):837–859, (2000).

[89] Peskir G., Shiryaev A. N.: *Solving the Poisson disorder problem.* Advances in Finance and Stochastics. Essays in Honour of Dieter Sondermann, 295–312, Springer, Berlin, (2002).

[90] Peskir G., Shiryaev A. N.: *Optimal stopping and free-boundary problems.* Birkhäuser, Basel, (2006).

[91] Prokhorov, Y. V.: *Convergence of random processes and limit theorems in probability theory.* Theory of Probability & Its Applications, 1(2):157–214, (1956).

[92] Protter P. E.: *Stochastic integration and differential equations.* Springer, (2005).

[93] Rüschendorf L., Rachev S.T.: *A characterization of random variables with minimum L^2-distance.* JJournal of Multivariate Analysis, 32:48–54, (1990).

[94] Rao K. M.: *On decomposition theorems of Meyer.* Mathematica Scandinavica, 24(1):66–78, (1969).

[95] Revuz D., Yor M.: *Continuous martingales and Brownian motion.* Vol. 293, Springer Science & Business Media, (2013).

[96] Rogers L. C. G.: *Monte Carlo valuation of American options.* Mathematical Finance, 12(3):271–286, (2002).

[97] Rogers L. C. G., Williams D.: *Diffusions, Markov processes and martingales: Volume 2, Itô Calculus.* Cambridge university press, (2000).

[98] Samee F.: *On the principle of smooth fit for killed diffusions.* Electronic Communications in Probability, 15:89–98, (2010).

[99] Sharpe M.: *General theory of Markov processes.* Vol. 133, Academic press, (1988).

[100] Shiryaev A. N.: *Optimal stopping rules.* Springer, (1978).

[101] Siorpaes P.: *On a dyadic approximation of predictable processes of finite variation.* Electronic Communications in Probability, 19(22):1–12, (2014).

[102] Skorokhod A.V.: *Studies in the theory of random processes.* Translated from the Russian by Scripta Technica, Inc. Addison-Wesley Publishing Co., Reading, Massachusetts, (1965).

[103] Snell J. L.: *Applications of martingale system theorems.* Transactions of the American Mathematical Society, 73(2):293–312 (1952).

[104] Stebegg F.: *Model-independent pricing of Asian options via optimal martingale transport.* Available online at: arXiv:1412.1429, (2014).

[105] Strassen V.: *The existence of probability measures with given marginals.* The Annals of Mathematical Statistics, 36(2):423-439, (1965).

[106] Touzi N.: *Martingale inequalities, optimal martingale transport, and robust superhedging.* ESAIM: Proceedings and Surveys, 45:32–47, (2014).

[107] van Moerbeke P.: *Optimal stopping and free boundary problems.* Rocky Mountain Journal of Mathematics, 4:539–578, (1974).

[108] Villani C.: *Optimal transport: old and new.* Vol. 338, Springer Science & Business Media, (2008).