# MODEL-INDEPENDENT BOUNDS FOR ASIAN OPTIONS: A DYNAMIC PROGRAMMING APPROACH

ALEXANDER M. G. COX AND SIGRID KÄLLBLAD

ABSTRACT. We consider the problem of finding model-independent bounds on the price of an Asian option, when the call prices at the maturity date of the option are known. Our methods differ from most approaches to model-independent pricing in that we consider the problem as a dynamic programming problem, where the controlled process is the conditional distribution of the asset at the maturity date. By formulating the problem in this manner, we are able to determine the model-independent price through a PDE formulation. Notably, this approach does not require specific constraints on the payoff function (e.g. convexity), and would appear to generalise to many related problems.

## 1. INTRODUCTION

Since the seminal paper of Hobson [36], there has been substantial interest in questions of the following form: given an asset with price $(S_t)_{t \in [0,T]}$, a derivative whose payoff, $X_T$, depends on the path of the asset, and the prices of call options at maturity time $T$, find a static portfolio of calls, and a dynamic trading strategy in the asset which superhedges the derivative at time $T$, under (essentially) *any* model for the asset. The class of models considered are usually very large (for example, all models with continuous paths), and so the resulting price is usually called the model-independent superhedging price.

The problem of finding the model-independent superhedging price is closely related to the problem of identifying the largest model-based price: specifically, in a classical setting, one would expect the prices of all options to be given as the expected value under some risk-neutral measure,[1] and by specifying the call prices at time 0, the distribution of $S_T$ under this risk-neutral measure is determined. It is therefore natural to conjecture that the model-independent superhedging price is equal to $\sup_{\mathbb{Q}} \mathbb{E}^{\mathbb{Q}}[X_T]$, where the supremum is taken over all probability measures $\mathbb{Q}$ such that $(S_t)_{t \in [0,T]}$ is a martingale, and $S_T$ has the distribution determined by the call options. Recently a number of papers, starting with Beiglböck, Henry-Labordère, and Penkner [8] in discrete time, and followed up by Dolinsky and Soner [24] in continuous time (see also [1, 7, 10, 11, 25, 40]), have made this result explicit under a variety of technical conditions. Note that in this formulation it is very natural to consider the supremum over the set of probability measure as a

[1]For ease of presentation, we largely assume that the interest rate is zero; this has no substantial effect on our main results.

primal problem, and the infimum over the class of super-hedging strategies as the corresponding dual problem.

An alternative approach to these problems, following Hobson [36], is to use a time-change argument to reformulate the primal problem in terms of an optimisation over solutions to the Skorokhod embedding problem (SEP): that is, to argue that, up to an unknown time change $\tau_t$, the martingale $S_t = B_{\tau_t}$ is a time change of a Brownian motion. For a number of important quantities (maximum, quadratic variation, local time, . . . ) the values of these quantities for the asset price up to time $T$ and for the Brownian motion up to the (stopping time) $\tau_T$ agree. It can often then be argued that the choice of a model for $S_t$ with given law, and the choice of a stopping time $\tau_T$ are equivalent provided $B_{\tau_T}$ has the required distribution (and satisfies an integrability constraint). The latter problem is well known as the Skorokhod embedding problem. A common approach to solving the model-independent superhedging problem is then to consider the corresponding Skorokhod embedding problem. If an optimal solution to this problem can be found, then it is often possible to guess the correct solution to the corresponding dual problem, and interpret this in terms of a superhedging strategy. This approach has been used in e.g. [14, 16, 17, 18, 19, 20, 34, 37, 38, 42]; see also the survey article of Hobson [35].

Of note in all of the known optimal solutions to the SEP is that some underlying structure is required on the form of the option payoff: for example, if we write $M_t := \sup_{u \leq t} B_u$ for the maximum process, then the optimal constructions are known to maximise $\mathbb{E}\left[F(M_\tau)\right]$ over solutions to the SEP, provided that $F$ is monotonic. To the best of our knowledge, the optimal construction when $F$ is not monotonic is not known. Similarly, in the case of variance options or the local time, the function $F$ must be concave/convex in order to have a known optimal solution. In Beiglböck, Cox, and Huesmann [9], this behaviour was explained in terms of a natural convexity property which holds when a path-swapping operation is performed. It follows from this operation that many constructions of solutions to the SEP are optimal when the payoff to be optimised has such a convexity property. However without the corresponding convexity, a 'nice' description of the optimal solution seems impossible. One of the key strengths of the results described in this paper is that our methods are *not* constrained by such a convexity assumption on the payoff, and therefore will work for very general payoffs.

The main results in this paper concern the case where the option described above is an Asian option, that is, $X_T = F(A_T)$, where $A_t = \int_0^t S_u \, du$ (we omit the usual scaling factor, $\frac{1}{t}$ for notational ease), and we consider the primal version of the problem, that is, we look to maximise $\mathbb{E}\left[F(A_T)\right]$ over all price processes $(S_u)_{u \in [0,T]}$ which are martingales, and which satisfy a constraint on the terminal law, $S_T \sim \mu$. Notably, the Asian example already falls outside the case of payoffs which can easily be handled by SEP methods, since the whole time-change $(\tau_u)_{u \in [0,T]}$, and not just the final time, $\tau_T$, is already important in determining the value of $A_T$. However, in the case where the function $F$ is convex, the optimal model is still easily determined: essentially, the asset will jump to its terminal distribution immediately, and the manner in which this is done (the 'embedding component') turns out to be irrelevant. This result was first given in Stebegg [45], which, to the best of our knowledge, is the first paper to characterise optimality in a setting where the SEP approach fails, or more generally to consider a problem of this form in continuous time without using the SEP approach. The standing assumptions in [45] are slightly different to ours — essentially, [45] allows a slightly more general setup (general starting measures, and discrete and continuously monitored payoffs are included) at the cost of considering only convex payoff functions (see also Section 4.1). At a heuristic level, this restriction to convex functions in [45] appears comparable to

the convexity constraint described above for the SEP in determining the 'simple' set of optimisers. We also observe that there is a long history of considering model-independent bounds for the prices of Asian options (e.g. [3, 4, 15, 22, 23, 31]), although we note that, in contrast to the case considered in this paper, existing results tend to consider discretely monitored Asian options, often when call options on the underlying are traded at some or all intermediate maturities.

The novelty of our approach relates to the manner in which we formulate the problem as a dynamic programming problem. In particular, we include the conditional law of the final value of the asset price in the formulation of our problem. The condition that the process is a martingale with this conditional law is then formulated in terms of the behaviour of the conditional law. Specifically, we require the conditional law $\xi_t$ to be a *measure-valued martingale*, by which we mean that $(\xi_t(A))_{t\in[0,T]}$ is required to be a martingale for any (Borel) set $A$. We will show in Section 2 that this condition is equivalent to the original formulation. In particular, by requiring $\xi_0 = \mu$ and requiring $\xi_T$ to be singular, we enforce the condition that the terminal law of $S_t = \int x\,\xi_t(\mathrm{d}x)$ is $\mu$. The concept of a measure-valued martingale is classical, (see e.g. Dawson [21] and Horowitz [39]; in this literature, the rather confusing terminology 'martingale measure' is also common), and has appeared in the context of the SEP in Eldan [29]. A key result for our purposes is that we are able to show that our value function is continuous in $\xi$, where the space of measures is equipped with the Wasserstein topology. This allows us to approximate $\xi$ by atomic measures, which enables us to reduce the whole problem to a finite-dimensional problem, at which point classical methods can be used (Section 3). We note that, in this discrete formulation, our problem could be compared to (a special case of) the problems considered in Bouchard and Nutz [13], El Karoui and Tan [28], and Žitković [46], although we prove our results via more direct, classical methods. We also remark that Galichon, Henry-Labordère, and Touzi [32] have also previously used a stochastic control approach to solve a similar problem, but in a rather different manner to the approach of this paper. In Section 4 we are able to use these results to provide concrete solutions to certain problems.

We believe that the methods and ideas we describe in this paper can be applied far beyond the case of Asian options. However, the Asian option setting does provide us with some useful structure which we are able to exploit in the construction and formulation of our problem. In particular, it is easy to show that 'small' changes in the conditional terminal law result in small changes in the value function for the problem, the increase in the average, $\mathrm{d}A_t$, is easy to write in terms of the current conditional law, and also our underlying problem is not strongly affected by jumps in the process: particularly, the value function for the problem where the path is assumed to be continuous, and the problem where the path is assumed to be càdlàg are identical (although optimisers may exist in the càdlàg formulation, and not in the continuous formulation). In Section 5 we discuss further extensions.

## 2. Problem formulation using measure-valued martingales

Consider the following problem: we have an asset $(S_t)_{t\in[0,T]}$ in a market with a riskless bank account and a time-horizon $T$, and we wish to find a model-independent super-hedge of an option which pays the holder $F(A_T)$, where $\frac{1}{T}A_T = \frac{1}{T}\int_0^T S_t\,\mathrm{d}t$ is the running average[2]. We will consider the problem where the law of

---

[2]We use the slightly unconventional notation $A_T = \int_0^T S_t\,\mathrm{d}t$ to avoid an unnecessary number of terms of the form $\frac{1}{T}$ in all our calculations; it is clear that this is just a scaling factor and can be removed.

the underlying asset at maturity, $S_T$, is given at time 0, and we consider the primal optimisation problem: that is, to find the law of the process which maximises $\mathbb{E}\left[F(A_T)\right]$ subject to $S_T \sim \mu$. Here, we consider the case where the interest rate $\rho = 0$, although the extension to constant interest rates is straightforward.

Our basic approach is to consider the problem as a dynamic programming problem where the current state includes the conditional distribution of the process at maturity. Specifically, we assume $S_T \in \mathbb{R}_+$, and with $\mathcal{M}(\mathbb{R}_+)$ the set of Borel measures on $\mathbb{R}_+$, we write

$$(2.1) \qquad \mathcal{P}^1 := \{\mu \in \mathcal{M}(\mathbb{R}_+) : \mu(\mathbb{R}_+) = 1, \int |x|\, \mu(\mathrm{d}x) < \infty\}.$$

Our aim is to set the problem up as a dynamic programming problem. We suppose that the problem evolves on an artificial time horizon, $r \geq 0$, on which a measure-valued process $(\xi_r)_{r \geq 0}, \xi_r \in \mathcal{P}^1$ evolves. We let $(T_r)_{r \geq 0}$ be an increasing process in $[0, T]$. Our interpretation of this quantity is that $T_r$ represents the 'real' time at the artificial time $r$. Roughly, the slower $T_r$ increases, the higher 'volatility' we see in the real-time scale. We set the problem up in this way, since we wish to allow a substantial change in the $r$ time-scale to happen instantaneously in real time, which may correspond to jumps in the asset price. However, we wish to maintain a 'continuous' evolution of the measure-valued process over its natural time-scale (we do not wish to deal with jumps in the measure-valued process). The choice of the increasing processes $(T_r)$ will form part of the control of the problem — specifically, we optimise over $\lambda_r \in [0, 1]$ and define

$$(2.2) \qquad T_r = \int_0^r \lambda_s \, \mathrm{d}s.$$

The second part of the control will be the choice of the measure-valued process $\xi$. This process will determine the conditional distribution of the asset $(S_t)$. Specifically, the initial value is $\xi_0 = \mu$, where $\mu$ is the terminal law of the asset at time 0, and over time we suppose that $(\xi_r)$ evolves in a manner that ensures that $(S_t)$ remains a martingale.

**Definition 2.1.** We say that an adapted process $(\xi_r)_{r \geq 0}$ with $\xi_r \in \mathcal{P}^1$ is a *measure-valued martingale* if, for any $f \in C_b(\mathbb{R}_+)$, $\xi.(f) := \int f(x)\, \xi.(\mathrm{d}x)$ is a martingale.

Note trivially that, if $f \in C_b(\mathbb{R}_+)$, then $\xi_r(f)$ is bounded for each $t$, and hence the martingale $\xi.(f)$ is uniformly integrable, with well defined limit $\xi_\infty(f)$ (in particular, $\xi_\infty$ is a measure; see [39, Proposition 2.1]).

*Remark* 2.2. An adapted process $(\xi_r)$ with $\xi_r \in \mathcal{P}^1$ is a measure-valued martingale if and only if $\xi.(A)$ is a martingale for any $A \in \mathcal{B}(\mathbb{R})$. Indeed, the indicator function over an interval of $\mathbb{R}$ may be approximated by continuous functions, and an application of the monotone class theorem yields that the claim holds for any $A \in \mathcal{B}(\mathbb{R})$ (see Lemma 2.12 for a similar argument). Conversely, any $f \in C_b(\mathbb{R}_+)$ may be approximated from below by simple functions. In fact, by the same argument, we see that $\xi.(f)$ is a martingale for any (non-negative) measurable function.

*Remark* 2.3. Our underlying probability spaces will generally be assumed to satisfy the usual conditions. Under this assumption, of course, for every $f \in C_b(\mathbb{R}_+)$, the martingale $\xi.(f)$ has a càdlàg version. More pertinently, we can choose a version of $(\xi_r)$ such that $\xi.(f)$ is càdlàg for every bounded Borel function $f$, see [39, Theorem 2.5]. In what follows, we will assume that we always take this version of $(\xi_r)$.

We will think of measure-valued martingales as processes, evolving in time. We emphasise that the support of the measure-valued martingale can only ever decrease: if $\xi_{r_0}(A) = 0$ then $\xi_r(A) = 0$ for all $r \geq r_0$. In the particularly nice case

that $\xi_{r_0}$ is an atomic measure, then for all $r \geq r_0$, the measure will also be atomic, and supported on the same set of points. In particular, the spatial distribution of such a measure will not change, only the weights attributed to each point. Since the weight associated to each point is a martingale and constrained to lie in the interval $[0, 1]$, it follows that in the limit as $r \to \infty$, the weight assigned to each point must converge to a limit; commonly, this limit will be assumed to be either 0 or 1, and this motivates the following definitions. Consider the set of singular measures on $\mathbb{R}_+$, $\mathcal{P}^s = \{\mu \in \mathcal{M}(\mathbb{R}_+) : \mu = \delta_y, y \in \mathbb{R}_+\}$, then:

**Definition 2.4.** We say that a measure-valued martingale $(\xi_r)$ is *terminating* if $\xi_r \to \xi_\infty \in \mathcal{P}^s$ a.s. as $r \to \infty$, where the convergence is in the sense of weak convergence of measures. It is *finitely terminating* if $\tau_s := \inf\{r \geq 0 : \xi_r \in \mathcal{P}^s\}$ is almost surely finite.

**Lemma 2.5.** *Suppose $(\xi_r)$ is a terminating measure-valued martingale with $\xi_0 = \mu$. Then $X_\cdot = \int x \, \xi_\cdot(\mathrm{d}x)$ is a non-negative UI martingale with $X_\infty \sim \mu$.*

*Proof.* The martingale property follows from Remark 2.2. In particular, for $y \in \mathbb{R}_+$, we have that

$$\mathbb{E}\left[(y - X_\infty)_+\right] = \mathbb{E}\left[\int (y - x)_+ \, \xi_\infty(\mathrm{d}x)\right] = \int (y - x)_+ \, \mu(\mathrm{d}x).$$

Since $\mathbb{E}\left[(y - X_\infty)_+\right]$ characterises the law of $X_\infty$ uniquely, $X_\infty \sim \mu$. As $X_0 = \int x \, \mu(dx) < \infty$, it also follows that $\mathbb{E}\left[|X_r|\right] = \mathbb{E}\left[X_r\right] < \infty$. Finally, we observe that $X_r = \mathbb{E}\left[X_\infty | \mathcal{F}_r\right]$ and $X_\infty \sim \mu \in \mathcal{P}^1$ imply $X$ is a UI martingale. □

**Corollary 2.6.** *If $(\xi_r)$ is a terminating measure-valued martingale with $\xi_0 = \mu$, then for every 1-Lipschitz function $f$, $X^f_\cdot := \xi_\cdot(f) = \int f(x)\xi_\cdot(\mathrm{d}x)$ is a uniformly integrable martingale with $X^f_0 = \int f \, \mathrm{d}\mu$ and $X^f_\infty \sim f(\mu)$.*

We also wish to discuss the continuity of the process $(\xi_r)$. In order to do this, we make the following definition:

**Definition 2.7.** We say that a measure-valued martingale is *continuous* if, for any 1-Lipschitz function $f$, $X^f_\cdot = \int f(x)\, \xi_\cdot(\mathrm{d}x)$ is almost surely continuous.

It immediately follows that $X_\cdot = \int x \, \xi_\cdot(\mathrm{d}x)$ is a continuous process whenever $(\xi_r)$ is continuous. This is also equivalent to requiring (almost sure) continuity of $r \mapsto \xi_r$ in the topology of $\mathcal{W}_1$, the first Wasserstein metric, by the duality of the Wasserstein distance [5, Theorem 6.1.1].

Having introduced these concepts, we will take the second control in our problem to be the choice of a process $(\xi_r)$, subject to the constraint that $(\xi_r)$ is a terminating, continuous measure-valued martingale with $\xi_0 = \mu$.

Observe that, once we have chosen a process $(\xi_r)$, the 'asset price' at time $T_r$ is given by $\int x \, \xi_r(\mathrm{d}x)$. Since the process $T_r$ is non-decreasing, there exists a right-continuous inverse, $T_t^{-1} = \inf\{r > 0 : T_r > t\}$, and introduce $T_t^{-1,*} = \inf\{r > 0 : T_r \wedge T > t\}$; moreover, there can be only countably many jumps in $T_t^{-1}$. We therefore define the càdlàg process

$$(2.3) \qquad S_t = \int x \, \xi_{T_t^{-1,*}}(\mathrm{d}x), \qquad t \leq T,$$

and note that $S_T = \int x \, \xi_\infty(\mathrm{d}x)$. The average process is then given by

$$(2.4) \qquad A_t = \int_0^t S_s \, \mathrm{d}s = \int_0^t \int x \, \xi_{T_s^{-1,*}}(\mathrm{d}x) \, \mathrm{d}s.$$

Then the main problem we wish to solve is the following:

**Problem 2.8** (Basic optimisation problem)**.** *Given an integrable probability measure $\mu$ on $\mathbb{R}_+$ and a function $F : \mathbb{R}_+ \to \mathbb{R}_+$, we want to find a probability space $(\Omega, \mathcal{H}, (\mathcal{H}_t)_{t\in[0,T]}, \mathbb{P})$ satisfying the usual conditions, and a càdlàg UI martingale $(S_t)_{t\in[0,T]}$ on this space with $S_T \sim \mu$ which maximises $\mathbb{E}[F(A_T)]$ over the class of such probability spaces and processes.*

***Remark*** 2.9. Since we do not require $\mathcal{H}_0$ to be trivial, $S_0$ need not be a constant. However, for the Asian option, it holds for any probability space and càdlàg martingale $S_t$ as given in Problem 2.8, that one may construct a sequence of càdlàg martingales $(S^n)$ such that $S_0^n = s_0 \in \mathbb{R}$, $S_T^n \sim \mu$, and

$$(2.5) \qquad\qquad \lim_{n\to\infty} \mathbb{E}[F(A_T^n)] = \mathbb{E}[F(A_T)].$$

Hence, the value of Problem 2.8 remains the same under the additional assumption that $S_0 = s_0$, and for any optimiser to the former problem an approximately optimal sequence may be constructed for the latter; we refer to Lemma 5.1 in Stebegg [45] for a precise argument (see also Assumption 3.9 in Guo, Tan, and Touzi [33] and the proof of Lemma 4.1 in Dolinsky and Soner [25] for related arguments). We argue in the proof of Lemma 2.11 below that the value of Problem 2.8 remains the same if restricting to martingales which are piecewise constant over arbitrary but finite partitions. Hence, a similar argument yields that the value of the problem also remains the same if we restrict to continuous martingales.

To formalise this remark, and since we generally wish to work with probability spaces satisfying the usual conditions, we extend our framework slightly: given a complete probability space with a right-continuous filtration $(\mathcal{G}_t)_{t\geq0}$, we can always extend the filtration to $(-\varepsilon, \infty)$, for some $\varepsilon > 0$, by taking $\mathcal{G}_t$ to be the (completion of the) trivial $\sigma$-algebra for $t < 0$. Similarly, a càdlàg process $Z_t$ on $[0, \infty)$ can be extended to a càdlàg process on $(-\varepsilon, \infty)$ by setting $Z_t$ to be some constant value for $t < 0$. Since this is constant we may write $Z_{0-}$ for this value without confusion. Similarly, to avoid the excessive use of $\varepsilon$'s, we write $(\mathcal{G}_t)_{t\in[0-,\infty)}$ to denote a filtration extended in this manner. All other terminology (e.g. martingales) are then to be understood in the obvious way.

Our first claim is that Problem 2.8 is equivalent to the following formulation:

**Problem 2.10** (Measure-valued martingale formulation)**.** *Given an integrable probability measure $\mu$ on $\mathbb{R}_+$ and a function $F : \mathbb{R}_+ \to \mathbb{R}_+$, we want to find a probability space $(\Omega, \mathcal{G}, (\mathcal{G}_r)_{r\in[0-,\infty)}, \mathbb{P})$ satisfying the usual conditions, a progressively measurable process $\lambda_r \in [0,1]$, such that $\int_0^r \lambda_s \, \mathrm{d}s \to \infty$ a.s. as $r \to \infty$, and a finitely terminating measure-valued $(\mathcal{G}_r)_{r\in[0-,\infty)}$-martingale $(\xi_r)_{r\in[0-,\infty]}$ with $\xi_{0-} = \mu$ and $\int x \, \xi_r(\mathrm{d}x)$ continuous a.s., which maximises $\mathbb{E}[F(A_T)]$ with $A_T$ given by (2.4).*

**Lemma 2.11.** *Problems 2.8 and 2.10 are equivalent, in the sense that the values coincide and if there exists an optimiser in Problem 2.8, then we can construct an optimiser for Problem 2.10, and vice-versa; if the supremum for the problem can only be approximated, then equivalent approximating sequences can be found.*

*Moreover, if $F$ is continuous, then the value of the problem remains the same if we restrict Problem 2.10 to probability spaces and processes such that the filtration $\mathcal{G}_r$ is the usual augmentation of the natural filtration of a $(\mathcal{G}_r)_{r\geq0}$-Brownian motion and $(\xi_r)_{r\geq0}$ is continuous in the sense of Definitions 2.4 and 2.7.*

As a consequence, if the restricted measure-valued martingale problem admits a solution, then a corresponding optimiser may be constructed also for the Basic optimisation problem. Before proving this result, we give an auxiliary lemma.

**Lemma 2.12.** *Suppose* $(X_r)_{r\in[0,\infty]}$ *is a martingale on* $(\Omega,\mathcal{G},(\mathcal{G}_r)_{r\in[0,\infty]},\mathbb{P})$ *such that* $\mathbb{E}[|X_\infty|]<\infty$. *Then there exists a terminating measure-valued martingale,* $(\xi_r)_{r\in[0,\infty]}$, *such that* $X_r=\int x\,\xi_r(\mathrm{d}x)$, *a.s. for all* $r\in[0,\infty]$.

*Proof.* Define the $\mathcal{G}_\infty$-measurable random measure $\xi_\infty(\mathrm{d}x):=\delta_{X_\infty}(\mathrm{d}x)$. Then, $\xi_\infty\in\mathcal{P}^1$, a.s. Further, let $\mathcal{A}$ be a countable Boolean algebra generating $\mathcal{B}(\mathbb{R})$ and define the $\mathcal{G}_r$-measurable set function $\xi_r$ by

$$(2.6)\qquad\qquad \xi_r(A):=\mathbb{E}[\xi_\infty(A)|\mathcal{G}_r],\quad A\in\mathcal{A}.$$

Since $\xi_\infty$ is countably additive a.s., so is $\xi_r$. Indeed, for $A_n\in\mathcal{A}$, $n\in\mathbb{N}$, such that $\cup A_n\in\mathcal{A}$ and $A_i\cap A_j=\emptyset$, $i\neq j$,

$$\xi_r(\cup A_n)=\mathbb{E}[\xi_\infty(\cup A_n)|\mathcal{G}_r]=\sum_{n=1}^\infty\mathbb{E}[\xi_\infty(A_n)|\mathcal{G}_r]=\sum_{n=1}^\infty\xi_r(A_n).$$

Since $\xi_r$ is also finite, it follows that $(2.6)$ uniquely defines a $\mathcal{G}_r$-measurable measure on $\mathcal{B}(\mathbb{R})$, up to a null set, on which we arbitrarily take $\xi_t=\delta_0$. Next, let $\mathcal{O}:=\{A\in\mathcal{B}(\mathbb{R}):\xi_r(A)$ is a martingale on $[0,\infty]\}$. Since, for any $r\in[0,\infty]$, $\xi_r$ is a measure and thus continuous from below, it follows that $\mathcal{O}$ is a monotone class. Indeed, for $A_n\in\mathcal{O}$, $n\in\mathbb{N}$, with $A_n\subset A_{n+1}\subset...$,

$$\mathbb{E}[\xi_\infty(\cup A_n)|\mathcal{G}_r]=\lim_{n\to\infty}\mathbb{E}[\xi_\infty(A_n)|\mathcal{G}_r]=\lim_{n\to\infty}\xi_r(A_n)=\xi_r(\cup A_n).$$

Since $\mathcal{A}\subset\mathcal{O}$, we have by the monotone class theorem that $\xi_r(A)$ is a martingale for all $A\in\mathcal{B}(\mathbb{R})$. Since $\mathbb{E}[\int|x|\,\xi_\infty(\mathrm{d}x)]=\mathbb{E}[|X_\infty|]<\infty$, this yields in particular that $\xi_r\in\mathcal{P}^1$, for $r\in[0,\infty]$. According to Remark 2.2, $(\xi_r)$ is thus a measure-valued martingale. It is terminating by definition. It therefore follows directly from Lemma 2.5 that $\int x\,\xi_r(\mathrm{d}x)=X_r$, a.s. for $r\in[0,\infty]$. $\qquad\square$

**Remark** 2.13. The above result can be partially found in [39, Theorem 1.3], on taking $(\xi_r)_{r\in[0,\infty]}$ as the optional projection of the random measure $\delta_{X_\infty}$.

*Proof of Lemma 2.11.* We show that every candidate solution to Problem 2.8 gives rise to a candidate solution to Problem 2.10, and *vice-versa*. The claim about optimisers follows.

We first suppose that we have a solution to Problem 2.8. By Monroe [43, Theorem 11], there exists a probability space $(\Omega,\mathcal{G}',(\mathcal{G}'_s)_{s\in[0,\infty)},\mathbb{P})$, a $(\mathcal{G}'_s)$-Brownian motion $(W_s)$ with $W_0=\int x\,\mu(\mathrm{d}x)$, and a right-continuous $(\mathcal{G}'_s)$-time change $(\tau_t)_{t\in[0,T]}$, such that $(S_t)$ and $(W_{\tau_t})$ are equal in law, $\tau_T$ is almost surely finite, and $W_{.\wedge\tau_T}$ is a UI martingale. We then define $\xi'_s$ to be the law of $W_{\tau_T}$ conditional on $\mathcal{G}'_s$. That is, we apply Lemma 2.12 to the process $W_{.\wedge\tau_T}$ to obtain a terminating measure-valued martingale $(\xi'_s)_{s\in[0,\infty]}$, such that $\int x\,\xi'_s(\mathrm{d}x)=W_{s\wedge\tau_T}$, a.s. Note that the properties of $\xi'_s$ are preserved by defining $\xi'_{0-}=\mu$, and that $\int x\,\xi'_s(\mathrm{d}x)$ must be continuous.

We now need to construct a measurable process $\lambda_r$ giving rise to a time-change $T_r$ via $(2.2)$ such that the process $(S_t)$ given by $(2.3)$ is the required process. Note that by construction, $(S_t)_{t\geq0}$ and $\left(\int x\,\xi'_{\tau_t}(\mathrm{d}x)\right)_{t\geq0}=(W_{\tau_t})_{t\geq0}$ are equal in law, and therefore they both give rise to the same value of $\mathbb{E}[F(A_T)]$. We will now modify the time-change and deduce that this gives rise to the correct process. Specifically, we recall that $\tau_T$ is finite a.s., let

$$(2.7)\qquad\qquad T_t^{-1}=\begin{cases}\tau_t+\frac{t}{T} & t\leq T\\\tau_T+1+\frac{(t-T)}{T} & t>T\end{cases},$$

and, in turn, define $T_r:=\sup\{t\geq0:T_t^{-1}\leq r\}$. From $(2.7)$ we immediately see that $T_t^{-1}$ is strictly increasing, with $T_t^{-1}-T_s^{-1}\geq\frac{t-s}{T}$ for $t>s$, so that $T_r$ is non-decreasing and $\frac{1}{T}$-Lipschitz. In particular, $T_{\tau_T+1+r}=T+r$ for $r\geq0$ so that

$T_t^{-1}$ given by (2.7) is indeed the right-continuous inverse of $T_r$. Further, with $R_r = r - \frac{T_r \wedge T}{T}$ and $\xi_r := \xi'_{R_r}$, $r \in [0, \infty]$, it follows that $R_{T_t^{-1}}(\omega) = \tau_t(\omega)$, $t < T$, and, thus, $(S_t)_{t<T}$ and $\left( \int x \, \xi_{T_t^{-1}}(\mathrm{d}x) \right)_{t<T}$ are equal in law. Indeed, $\xi'_r \in \mathcal{P}^s$ for $r \geq \tau_T$. Therefore let $(\mathcal{G}_r)$ be the (right-continuous) time-changed filtration given by $\mathcal{G}_r = \mathcal{G}'_{R_r}$, $r \in [0, \infty)$. Then $\xi_r$ is a finitely terminating measure-valued $(\mathcal{G}_r)$-martingale. Further, $T_t^{-1} \in \mathcal{G}'_{\tau_t} = \mathcal{G}'_{R(T_t^{-1})} = \mathcal{G}_{T_t^{-1}}$ and, thus, $T_r \in \mathcal{G}_r$. Recalling the properties of $T_r$, we deduce that there exists a process $\lambda_r \in [0, 1]$ which is $\mathcal{G}_r$ measurable and such that $T_r = T \int_0^r \lambda_s \, \mathrm{d}s$. Hence (possibly by taking a modification), $\lambda_r$ can be assumed to be progressively measurable, and it is immediate that $T_r \to \infty$ as $r \to \infty$.

To see the converse, suppose we are given a solution to Problem 2.10. From Lemma 2.5 it follows immediately that $S_. = \int x \, \xi_{T_.^{-1},*}(\mathrm{d}x)$ is the required process.

It remains to argue the second part of the lemma. Indeed, in general, the time-change granted by Monroe [43] may not necessarily be measurable with respect to the Brownian filtration. However, for any probability space $(\Omega, \mathcal{H}, (\mathcal{H}_t)_{t \in [0,T]}, \mathbb{P})$ and càdlàg martingale $(S_t)_{t \in [0,T]}$, we may define a sequence $(S_t^n)_{t \in [0,T]}$ by

$$S_t^n = S_{[nt/T]T/n}, \quad n \geq 1.$$

Then the $(S_t^n)$ are still martingales with $S_T \sim \mu$. Further, since $F$ is continuous, $F(A_T^n)$ converges a.s. to $F(A_T)$, and an application of Fatou's Lemma gives that $\mathbb{E}[F(A_T)] \leq \liminf_{n \to \infty} \mathbb{E}[F(A_T^n)]$. In consequence, the value of Problem 2.8 remains the same if restricting to martingales which are piecewise constant over arbitrary but finite partitions. Since any discrete martingale may be embedded in a Brownian motion with stopping times measurable with respect to the Brownian filtration (c.f. e.g. [26]), it follows that we may restrict to Brownian filtrations $(\mathcal{G}_r)$ in Problem 2.10. By the Martingale Representation Theorem, any $(\mathcal{G}_r)$-martingale is continuous. In consequence, recalling Remark 2.3, the $(\xi_r)$ defined above can be assumed to be continuous in the sense of Definition 2.7. The fact that the resulting measure-valued martingale is finitely terminating, and that the first time the integral of $\lambda_s$ reaches $T$ is finite also follow immediately from this embedding procedure. □

**Remark** 2.14. We note that the embedding of a process $(S_t)_{t \in [0,1]}$ into the pair of a continuous measure-valued martingale $(\xi_r)_{r \in [0,\infty]}$ and time-change $(\lambda_r)_{r \in [0,\infty]}$, is not unique. In particular, choosing $T_t^{-1} := 1 - e^{-\tau_t} + \frac{t}{T}$ (c.f. (2.7)) renders $T_T^{-1} \leq 2$ a.s. and the problem might be viewed as evolving on the *finite* time-scale $r \in [0, 2]$. In Lemma 3.4 below, we will consider yet another scaling which yields a specific relation between the evolution of the $\xi$ and the $\lambda$.

**Remark** 2.15. We observe in fact that, from the proof of the lemma, if $\lambda_r = 1$ for $r \in [u, v)$, for some interval $[u, v)$, then $\xi_r = \xi_u$ for all $r \in [u, v)$. In particular, $\lambda = 1$ corresponds to a constant $\xi$ and, thus, $(S_t)$ is constant on $t \in [T_u, T_v)$.

## 3. The Dynamic Programming Problem

3.1. **Problem formulation and continuity.** We want to write the optimisation problem as a 'Markovian' optimisation problem: we suppose that at time $r$, we have 'real' time $T_r = t$, current law $\xi_r = \xi \in \mathcal{P}^1$, running average $A_{T_r} = a$, and we wish to find:

$$(3.1) \qquad U(r, t, \xi, a) = \sup \mathbb{E}\left[ F(A_T) | T_r = t, \xi_r = \xi, A_{T_r} = a \right],$$

where the supremum is taken over all time-change determining processes $(\lambda_u)_{u \in [r,\infty)}$ and measure-valued martingales $(\xi_u)_{u \in [r,\infty)}$ satisfying the conditions of Problem

2.10. By Lemma 2.11 it follows that $U(0, 0, \mu, 0)$ will be the value of the Asian option under the optimal model. At this stage, we directly define the function in (3.1) to equal the value of Problem 2.10 when the law to be embedded is given by $\xi$, the horizon by $T - t$, and the payoff function by $F(a + \cdot)$. Then we have:

**Lemma 3.1.** *Suppose $F$ is a non-negative, Lipschitz function. The function $U$ : $\mathbb{R}_+ \times [0, T] \times \mathcal{P}^1 \times \mathbb{R}_+ \to \mathbb{R}$ is continuous (here the topology on $\mathcal{P}^1$ is the topology derived from the Wasserstein-1 metric), and independent of $r$.*

*Proof.* We begin by observing that continuity in all the variables except $\xi$ follows immediately from the continuity properties of $F$: any small change in $a$ will result in a direct shift in the final value of $A_T$, while small changes in $t$ can be handled by computing the average of the same model over the modified time-horizon. In addition, the independence of the problem on the value of the 'measure-valued' time-scale, $r$, is immediate.

We consider finally the continuity in the measure, $\xi$. Consider a given probability space $(\Omega, \mathcal{G}, (\mathcal{G}_s)_{s \geq r}, \mathbb{P})$ and a measure-valued martingale $(\xi_s)_{s \geq r}$. Recall that $\mathcal{W}_1$ is the Wasserstein-1 metric space, and write $d_{\mathcal{W}_1}$ for the metric on this space. Let $\xi' \in \mathcal{P}^1$. We will first show that, if $d_{\mathcal{W}_1}(\xi_r, \xi') < \varepsilon$, then there is a measure-valued martingale $(\xi'_s)_{s \geq r}$ such that $\xi'_r = \xi'$ and $\mathbb{E}\left[\left|\int x\, \xi_s(\mathrm{d}x) - \int x\, \xi'_s(\mathrm{d}x)\right| \big| \mathcal{G}_r\right] < \varepsilon$ for all $s \in [r, \infty)$. Recall that $d_{\mathcal{W}_1}(\xi_r, \xi') < \varepsilon$ implies that there exists a transport plan, $\Gamma \in \mathcal{M}(\mathbb{R}_+ \times \mathbb{R}_+)$ such that $\xi'(\mathrm{d}y) = \Gamma(\mathbb{R}_+, \mathrm{d}y)$, $\xi_r(\mathrm{d}x) = \Gamma(\mathrm{d}x, \mathbb{R}_+)$ and $\int\int |x - y|\, \Gamma(\mathrm{d}x, \mathrm{d}y) < \varepsilon$. First, by disintegration (e.g. [5, Theorem 5.3.1]) there exists a Borel family of probability measures, $m(x, \mathrm{d}y)$ such that $\Gamma(\mathrm{d}x, \mathrm{d}y) = \xi_r(\mathrm{d}x)\, m(x, \mathrm{d}y)$.

Now define the process

$$\xi'_s(\mathrm{d}y) := \int \xi_s(\mathrm{d}x)\, m(x, \mathrm{d}y), \quad s \geq r.$$

Then $\xi'_s \in \mathcal{P}^1$ and $\xi'_r = \xi'$. Further, for any $A \in \mathcal{B}(\mathbb{R})$, since $m(\cdot, A)$ is measurable,

$$\mathbb{E}\left[\xi'_u(A) | \mathcal{G}_s\right] = \mathbb{E}\left[\int m(x, A)\xi_u(\mathrm{d}x) | \mathcal{G}_s\right] = \int m(x, A)\xi_s(\mathrm{d}x) = \xi'_s(A), \quad s \leq u,$$

and, thus, $\xi'_s$, $s \geq r$, is a measure-valued martingale. Next, note that

$$\left|\int x\, \xi_s(\mathrm{d}x) - \int x\, \xi'_s(\mathrm{d}x)\right| = \left|\int\int (x - y)\, \xi_s(\mathrm{d}x)m(x, \mathrm{d}y)\right|$$

$$\leq \int\int |x - y|\, \xi_s(\mathrm{d}x)m(x, \mathrm{d}y).$$

Hence

$$\mathbb{E}\left[\left|\int x\, \xi_s(\mathrm{d}x) - \int x\, \xi'_s(\mathrm{d}x)\right| \big| \mathcal{G}_r\right] \leq \int\int |x - y|\, \Gamma(\mathrm{d}x, \mathrm{d}y).$$

By the definition of the metric on $\mathcal{W}_1$, since $d_{\mathcal{W}_1}(\xi_r, \xi'_r) < \varepsilon$, we can find a transport plan $\Gamma$ with the desired marginals and $\int\int |x - y|\, \Gamma(\mathrm{d}x, \mathrm{d}y) < \varepsilon$. Fix some process $(\lambda_s)_{s \geq r}$, and write $A_t^{\xi, \lambda}$ for the average process corresponding to the measure-valued process $\xi$ and the time-change process $\lambda$, conditional on $\mathcal{F}_r$. Recalling (2.4) we have

$$\mathbb{E}\left[\left|A_T^{\xi, \lambda} - A_T^{\xi', \lambda}\right| \big| \mathcal{G}_r\right] = \mathbb{E}\left[\left|\int_{T_r}^T \int x\, \xi_{T_s^{-1}}(\mathrm{d}x)\, \mathrm{d}s - \int_{T_r}^T \int x\, \xi'_{T_s^{-1}}(\mathrm{d}x)\, \mathrm{d}s\right| \big| \mathcal{G}_r\right]$$

$$\leq \mathbb{E}\left[\int_{T_r}^T \left|\int x\, \xi_{T_s^{-1}}(\mathrm{d}x) - \int x\, \xi'_{T_s^{-1}}(\mathrm{d}x)\right|\, \mathrm{d}s \big| \mathcal{G}_r\right]$$

$$\leq \varepsilon(T - T_r).$$

Now fix $\varepsilon' > 0$ and consider $\xi, \xi' \in \mathcal{P}^1$ such that $d_{\mathcal{W}_1}(\xi, \xi') < \varepsilon'/(2T\zeta)$, where $\zeta$ is the Lipschitz constant of $F$. Then there exists $(\xi_s, \lambda_s)_{s \geq r}$, such that $\xi_r = \xi$ and $U(r, t, \xi, a) \leq \mathbb{E}\left[F\left(A_T^{\xi, \lambda}\right) \Big| \mathcal{G}_r\right] + \varepsilon'/2$. Using the estimate above, and by the Lipschitz property of $F$, we can moreover find $(\xi'_s)_{s \geq r}$, such that $\xi'_r = \xi'$ and $\mathbb{E}\left[\left|F\left(A_T^{\xi, \lambda}\right) - F\left(A_T^{\xi', \lambda}\right)\right| \Big| \mathcal{G}_r\right] \leq \varepsilon'/2$. It follows that

$$U(r, t, \xi, a) \leq \mathbb{E}\left[F\left(A_T^{\xi, \lambda}\right) \Big| \mathcal{G}_r\right] + \varepsilon'/2 \leq \mathbb{E}\left[F\left(A_T^{\xi', \lambda}\right) \Big| \mathcal{G}_r\right] + \varepsilon' \leq U(r, t, \xi', a) + \varepsilon'.$$

By symmetry, $|U(r, t, \xi', a) - U(r, t, \xi', a)| \leq \varepsilon'$. Finally, we note that joint continuity follows as a simple adaptation of this argument combined with the arguments for the other parameters. □

Since the function $U(r, t, \xi, a)$ is independent of the parameter $r$, we will often write $U(t, \xi, a)$ where there is no confusion.

**Remark** 3.2. Continuity of the primal problem as a function of $\mu$ was proven by alternative methods in Dolinsky and Soner [25, Theorem 4.1]. As demonstrated in Guo, Tan, and Touzi [33, Proposition 4.3], upper semi-continuity can be proven by yet an alternative method. We now recall their argument in the present context. To this end, consider the space of all càdlàg paths on $[0, 1]$ and let the filtration be the one generated by the canonical process $(S_t)$. Problem 2.8 can then be formulated as maximizing $\mathbb{E}[F(A_T)]$ over martingale measures satisfying the constraint $S_T \sim^{\mathbb{P}} \mu$. Given a sequence of probability measures $(\mu_n)$ on $\mathbb{R}_+$ converging in $d_{\mathcal{W}_1}$ to $\mu$, let $(\mathbb{P}_n)$ be a sequence of martingale measures such that $S_T \sim^{\mathbb{P}_n} \mu_n$, and

$$(3.2) \qquad \lim_{n \to \infty} \mathbb{E}^{\mathbb{P}_n}[F(A_T)] = \limsup_{n \to \infty} U(0, \mu_n, 0).$$

According to Jakubowski [41], there exists a sub-sequence $(\mathbb{P}_{n_k})_{k \geq 1}$ which is weakly convergent with respect to the so-called $S$-topology on the set of càdlàg paths. Let $\mathbb{P}_0$ be the limiting measure. According to Guo, Tan, and Touzi [33], $\mathbb{P}_0$ is then a martingale measure and $S_T \sim^{\mathbb{P}_0} \mu$. Since the mapping $\omega \mapsto A_T(\omega)$ is $S$-continuous (c.f. Corollary 2.11 in [41]) it follows that

$$U(0, \mu, 0) \geq \mathbb{E}^{\mathbb{P}_0}[F(A_T)] \geq \lim_{n \to \infty} \mathbb{E}^{\mathbb{P}_{n_k}}[F(A_T)],$$

which combined with (3.2) yields the upper semi-continuity.

3.2. **Reduction to a finite dimensional problem.** Our aim now is to provide a more concrete description of the function $U$. However, because the function $U$ is continuous in $\xi$, we can restrict ourselves to a nicer class of problems: specifically, we can approximate our object of primary interest, $U(t, \xi, a)$ by a sequence $U(t, \xi^N, a)$, where $\xi^N$ can be chosen to have nice properties. For our purposes, a natural simplifying assumption is to assume that the measures $\xi^N$ are atomic measures. In this case, as we shall see, the problem becomes much more tractable via classical methods. As a consequence of this reduction, we will be able to deduce that a Dynamic Programming Principle holds by standard results from the literature. However the more theoretical question of whether a DPP holds for the original formulation is proved in the appendix; this result will not be used elsewhere in the paper.

To do this, we let $\mathbb{X}_N = \{x_0, x_1, \ldots, x_N\}$, where $0 \leq x_0 < x_1 < \cdots < x_N$, and write $\mathcal{P}^1(\mathbb{X}_N) = \mathcal{P}^1 \cap \mathcal{M}(\mathbb{X}_N)$ and $\mathcal{P}^s(\mathbb{X}_N) = \mathcal{P}^s \cap \mathcal{M}(\mathbb{X}_N)$. Observe that if $(\xi_r)$ is a terminating measure-valued martingale and $\xi_0 \in \mathcal{P}^1(\mathbb{X}_N)$ then $\xi_r \in \mathcal{P}^1(\mathbb{X}_N)$ a.s. for all $r \geq 0$ and $\xi_\infty = \delta_{x_i}$ for some $x_i \in \mathbb{X}_N$. Further, write $\alpha \subseteq \{0, 1, \ldots, N\}$, $\mathbb{X}_\alpha = \{x_i : i \in \alpha\}$, and $\mathcal{P}^1(\mathbb{X}_\alpha), \mathcal{P}^s(\mathbb{X}_\alpha)$ etc. as above. In particular, $\mathbb{X}_N = \mathbb{X}_{\{0,1,\ldots,N\}}$.

We then have the following characterisation:

**Lemma 3.3.** *Suppose $\mu \in \mathcal{P}^1(\mathbb{X}_N)$. Then $(\xi_r)$ is a measure-valued martingale with $\xi_0 = \mu$ if and only if $\xi_r^n := \xi_r(\{x_n\})$ is a non-negative martingale for each $n$ and $\sum_{i=0}^N \xi_r^n = 1$. Moreover, $(\xi_r)$ is terminating if and only if $\xi_\infty^n = 0$ for all but one $n \in \{0, 1, 2, \ldots, N\}$, almost surely, and $(\xi_r)$ is continuous if and only if $\xi_r^n$ is continuous for each $n$.*

It is clear that there are similar statements where $\mathcal{P}^1(\mathbb{X}_N)$ is replaced by $\mathcal{P}^1(\mathbb{X}_\alpha)$.

Then we consider the further consequence of Lemma 2.11: by the Martingale Representation Theorem, working on the probability space granted by Lemma 2.11, we can assume that the dynamics of $(\xi_r)$ are given by a controlled Brownian motion, in a recursive formulation. For fixed $N \geq 1$, we suppose that $(\xi_r)$ solves the SDE

$$(3.3) \qquad \mathrm{d}\xi_r^n = w_r^n \mathrm{d}W_r,$$

for $(W_r)$ a standard Brownian motion, and where $\mathbf{w}_r = (w_r^1, \ldots, w_r^N) \in \mathbb{R}^N$, $w_r^0 = -\sum_{n=1}^N w_r^n$, and $\xi_r^n \in \{0, 1\}$ implies $w_r^n = 0$, $n \in \{0, \ldots, N\}$, — that is, as soon as one of the atoms disappears, it can never be resurrected.

Next, we show that $(\xi_r)$ and $(\lambda_r)$ may be chosen so that a specific relation holds between $\mathbf{w}_r$ and $\lambda_r$.

**Lemma 3.4.** *Let $\mu \in \mathcal{P}^1(\mathbb{X}_N)$ and consider a martingale $(S_t)_{t \in [0,T]}$ represented via (2.3) by processes $(\lambda_r, \xi_r)_{r \in [0,\infty)}$ given on a probability space $(\Omega, \mathcal{G}, (\mathcal{G}_r)_{r \in [0,\infty)}, \mathbb{P})$, such that $\lambda_r \in [0, 1]$ is a progressively measurable process and $(\xi_r)$ is a measure-valued martingale with $\xi_0 = \mu$. Suppose further that $(\mathcal{G}_r)$ is the natural filtration of a Brownian motion $(W_r)$, $\inf\{r \geq 0 : \int_0^r \lambda_s \, \mathrm{d}s = T\} < \infty$ a.s., and $(\xi_r)$ is continuous and finitely terminating. Then, w.l.o.g., we may assume that*

$$(3.4) \qquad ||\mathbf{w}_u||^2 + \lambda_u = 1 - \mathbf{1}_{\{\xi_u \in \mathcal{P}^s\}} \mathbf{1}_{\{T_u = T\}}.$$

*That is, we can always choose a multiple $(\Omega, \mathcal{G}, (\mathcal{G}_r), \mathbb{P}, (\lambda_r, \xi_r))$ which represents $(S_t)$ via (2.3), and which satisfies the above properties as well as (3.4).*

*Proof.* Suppose $(\Omega, \mathcal{G}, (\mathcal{G}_r), \mathbb{P}, (\lambda_r, \xi_r))$ satisfy the assumptions of the lemma (apart from (3.4)). We aim to construct a time-change such that the time-changed filtration $(\overline{\mathcal{G}}_u)$ and time-changed processes $(\bar{\lambda}_u, \bar{\xi}_u)_{u \geq 0}$ satisfy the assertions. To this end, recall that $T_r$ is given by (2.2) and let $\tau := \inf\{r : \xi_r \in \mathcal{P}^s \text{ and } T_r = T\}$; since $\xi_r$ is finitely terminating, $\tau$ is finite a.s. Let $\phi : \Omega \times [0, \infty) \to \mathbb{R}_+$ be given by $\phi(u) = \int_0^u \eta_s^2 \mathrm{d}s$ for some positive, adapted process $\eta_u$ such that $\phi(\infty) \geq \tau$. Then $\phi$ is continuous and increasing in $u$, and its inverse $\phi^{-1}$ is well-defined and continuous on $[0, \tau]$. We define

$$(3.5) \qquad \bar{\xi}_u := \xi_{\phi(u)} \quad \text{and} \quad \overline{T}_u := T_{\phi(u)},$$

and let $(\overline{\mathcal{G}}_u)_{u \in [0,\infty)}$ the time-changed filtration with $\overline{\mathcal{G}}_u = \mathcal{G}_{\phi(u)}$. Note that $\bar{\xi}_u$ is a continuous measure-valued $(\overline{\mathcal{G}}_u)$-martingale. Moreover, (3.5) implies that $\overline{T}_t^{-1} = \phi^{-1}(T_t^{-1})$, $t < T$ (recall that $\overline{T}_T^{-1} = \infty$). Hence, $S_t$ is given by (2.3) defined with respect to $\bar{\xi}_u$ and $\overline{T}_u$. It remains to argue that $\eta_u$ can be chosen such that

$$(3.6) \qquad ||\bar{\mathbf{w}}_u||^2 + \bar{\lambda}_u = 1 - \mathbf{1}_{\{\bar{\xi}_u \in \mathcal{P}^s\}} \mathbf{1}_{\{\overline{T}_u = T\}}.$$

First, note that $\phi^{-1}(\tau) = \inf\{u : \bar{\xi}_u \in \mathcal{P}^s \text{ and } \overline{T}_u = T\}$. By the Martingale Representation Theorem, we know that $\xi_r$ is given by (3.3) for some process $(\mathbf{w}_r) \in \mathbb{R}^N$. Since there is a $(\overline{\mathcal{G}}_u)$-Brownian motion $(B_u)$ such that $\mathrm{d}W_{\phi(u)} = \eta_u \mathrm{d}B_u$, it follows from (3.5) that

$$(3.7) \qquad \mathrm{d}\bar{\xi}_u^n = w_{\phi(u)}^n \eta_u \mathrm{d}B_u \quad \text{and} \quad \bar{\lambda}_u = \eta_u^2 \lambda_{\phi(u)}.$$

Let $\bar{\mathbf{w}}_u = \mathbf{w}_{\phi(u)}\eta_u$. Then (3.6) holds for $u \le \phi^{-1}(\tau)$, if $\eta$ satisfies

$$\eta_u^2 \;=\; \frac{1}{||\mathbf{w}_{\phi(u)}||^2 + \lambda_{\phi(u)}}, \quad u \le \phi^{-1}(\tau).$$

We therefore proceed by defining $\phi^{-1}$ via

$$\mathrm{d}\phi^{-1}(r) := \left(||\mathbf{w}_r||^2 + \lambda_r\right)\mathrm{d}r, \qquad r \le \tau.$$

It follows from the construction of $(\xi_r)_{r\in[0,\infty)}$ and $(\lambda_r)_{r\in[0,\infty)}$, that for $r \le \tau$, $||\mathbf{w}_r|| = 0$ if and only if $\lambda_r = 1$ (c.f. Remark 2.15, and note that $||\mathbf{w}|| = 0$, if and only if, $||(w^0, \mathbf{w})|| = 0$). In consequence, $\phi^{-1}$ is well-defined, continuous and strictly increasing on $[0, \tau]$. In particular,

$$\phi^{-1}(\tau) \;=\; \int_0^\tau \left(||\mathbf{w}_r||^2 + \lambda_r\right)\mathrm{d}r \;=\; T + \int_0^\tau ||\mathbf{w}_r||^2 \,\mathrm{d}r,$$

and we observe that, as the quadratic variation process of a bounded martingale (in $\mathbb{R}^N$), $\int_0^\tau ||\mathbf{w}_r||^2 \,\mathrm{d}r$ is almost surely finite. It follows that $\phi_u$ and $\eta_u$ are well-defined, for $u \le \phi^{-1}(\tau)$, and that $\phi_u$ and $\phi_r^{-1}$ are adapted with respect to the respective filtrations $(\bar{\mathcal{G}}_u)$ and $(\mathcal{G}_r)$. In consequence, $\bar{\xi}_u$ and $\bar{\lambda}_u$ are well-defined via (3.5) for $u \le \phi^{-1}(\tau)$ and, according to (3.7), $\bar{\lambda}_u = \eta_u^2\lambda_{\phi(u)} \in [0, 1]$, and can therefore be assumed to be progressively measurable (possibly after taking a modification). For $u > \phi^{-1}(\tau)$, we take $\mathbf{w}_u = 0 = \lambda_u$. $\qquad\square$

**Remark** 3.5. When embedding piecewise constant martingales as was done in the second part of the proof of Lemma 2.11 it follows that $\lambda_r \in \{0, 1\}$. However, the solution to Problem 2.10 (which in general is not unique), and thus to the basic optimisation problem, may be attained for more general processes $\lambda_u \in [0, 1]$; c.f. the non-convex example considered in Section 4.2. Hence, we do not further restrict the set of $\lambda$'s even though the value of the problem would remain unaffected.

Given the above dynamics of the stochastic factors, we note that $U(r, t, \xi, a)$ in (3.1) is now well-defined as the value-function corresponding to a class of dynamic control problems. In particular, w.l.o.g., we may optimise over controlled processes defined on a fixed reference probability space; see e.g. [30]. The following result is now an immediate consequence of the lemma above. Let $\mathcal{A}_{u_0}^N = \{(\mathbf{w}_u)_{u \ge u_0} : \mathbf{w}_u$ prog. meas., $\mathbf{w}_u = (w_u^1, \ldots, w_u^N) \in \mathbb{R}^N$ and $||\mathbf{w}_u|| \le 1\}$ the set of admissible controls[3].

**Corollary 3.6.** *For each* $\mathbf{w} \in \mathcal{A}_{u_0}^N$, *define* $(\lambda_u)_{u \ge u_0}$ *by (3.4), and* $(\xi_u)_{u \ge u_0}$ *by (3.3) with* $w^0 = -\sum_{i=1}^N w^i$. *Then, for* $\mu \in \mathcal{P}(\mathbb{X}_N)$, *Problem 2.10 in its restricted form (c.f. Lemma 2.11) is equivalent to finding a process* $\mathbf{w} \in \mathcal{A}_0^N$ *such that* $\xi_u^n \in \{0, 1\}$ *implies* $w_s^n = 0$, $s \ge u$, *for* $n \in \{0, ..., N\}$, *and such that* $\mathbf{w}$ *maximises* $\mathbb{E}[F(A_T)]$ *over the class of such processes where*

(3.8) $$\mathrm{d}A_{T_u} = (x_0, ..., x_N) \cdot (\xi_u^0, ..., \xi_u^N)\,\lambda_u T \,\mathrm{d}u.$$

*Moreover, for all* $\varepsilon > 0$, *there exists* $u^* = u^*(\varepsilon) > 0$ *such that, for any* $\mu, t, a$,

$$\inf_{\mathbf{w} \in \mathcal{A}_{u_0}} \mathbb{P}(\xi_{u_0+u^*} \in \mathcal{P}^s(\mathbb{X}_N)|\xi_{u_0} = \mu, T_{u_0} = t, A_{T_{u_0}} = a) > 1 - \varepsilon.$$

*Proof.* The only part of the first half of the corollary that does not follow immediately from the previous result is that the process $(\xi_u)_{u \ge 0}$ which results from a given choice $\mathbf{w} \in \mathcal{A}_{u_0}$ is terminating, and this follows once we show the second half of the result.

---

[3]Strictly speaking, we should also include here the set of possible probability spaces, as in [30]; for ease of presentation, we omit this detail, which does not affect our arguments.

To see the second half of the result, note that it is sufficient to show that there is a similar bound for the first time that $\xi_u^n = 0$ for some $n \in \{0, 1, \ldots, N\}$. Consider the process at time $u^* \gg u_0$, and suppose that the measure $\xi$ has not already become singular at this time, so we have

$$\int_{u_0}^{u^*} (\|\mathbf{w}_u\| + \lambda_u) \, du = u^* - u \implies \int_{u_0}^{u^*} \|\mathbf{w}_u\| \, du \geq u^* - u - T.$$

In particular, we can ensure that $\max_k \left\{ \int_{u_0}^{u^*} (w_u^k)^2 \, du \right\}$ is arbitrarily large by choosing $u^*$ sufficiently large. Let $k^*$ be the maximising component; it follows immediately from the fact that $\xi_u^{k^*}$ is a $[0,1]$-valued martingale with quadratic variation process $\int_{u_0}^{u^*} (w_u^{k^*})^2 \, du$, that with high probability at least one component must have become zero or one. $\qquad\square$

Between Lemma 3.1, and Corollary 3.6, we have shown that the problem (3.1) is equivalent to choosing the variable $\mathbf{w}$ in the problem above. Moreover, we can break the problem up into a sequence of independent problems by considering the process only up to the first time that one of the current atoms becomes zero. In particular, for $\xi \in \mathcal{P}(\mathbb{X}^N)$, we can consider the problem:

(3.9)
$$\tilde{V}_N(u, t, \xi, a) = \sup_{\mathbf{w} \in \mathcal{A}_u^N} \mathbb{E}\Big[ \tilde{V}_{N-1}(\sigma, T_\sigma, \xi_\sigma, A_{T_\sigma}) \mathbf{1}_{\{T_\sigma < T\}}$$
$$+ F(A_{T_\sigma}) \mathbf{1}_{\{T_\sigma = T\}} | A_{T_u} = a, \xi_u = \xi, T_u = t \Big],$$

where $\sigma = \inf\{s \geq u : \xi_s^n \notin (0,1) \text{ some } n \in \{0, \ldots, N\} \text{ or } T_s = T\}$. We also have the boundary conditions $\tilde{V}_0(u, t, \xi, a) = F(a + (1-t)x)$, where $\xi = \delta_x$. Here, the function $\tilde{V}_k(u, t, \xi, a)$ is determined for $\xi \in \mathcal{P}^1(\mathbb{X}_\alpha)$ with $|\alpha| = k+1$. Specifically, for $\xi \in \mathcal{P}^1(\mathbb{X}_\alpha)$ with $|\alpha| = k+1$

$$\tilde{V}_k(u, t, \xi, a) = \sup_{\mathbf{w} \in \mathcal{A}_u^k} \mathbb{E}\Big[ \tilde{V}_{k-1}(\sigma, T_\sigma, \xi_\sigma, A_{T_\sigma}) \mathbf{1}_{\{T_\sigma < T\}}$$
$$+ F(A_{T_\sigma}) \mathbf{1}_{\{T_\sigma = T\}} | A_{T_u} = a, \xi_u = \xi, T_u = t \Big],$$

where we set $\sigma = \inf\{s \geq u : \xi_s \in \mathcal{P}^1(\mathbb{X}_\alpha) \text{ some } \alpha, |\alpha| \leq k \text{ or } T_r = T\}$. Observe in particular that each $\xi \in \mathcal{P}^1(\mathbb{X}_N)$ determines a unique set $\alpha$ such that $\xi \in \mathcal{P}^1(\mathbb{X}_{\alpha'})$ implies $\alpha' \supset \alpha$. Write $\alpha(\xi)$ for this unique subset, and we observe that we have the consistency conditions:

$$\tilde{V}_{|\alpha(\xi)|-1}(u, t, \xi, a) = \tilde{V}_k(u, t, \xi, a), \quad \text{for all } k \geq |\alpha(\xi)| - 1.$$

Finally, fix $\xi$ with $|\alpha(\xi)| = k+1$. We can make the identification between the probability measure $\xi = \sum_{j=0}^k \xi^{i_j} \delta_{x_{i_j}}$ (on $\mathbb{X}_\alpha$), where $i_0, i_1, \ldots, i_k$ are the ordered elements of $\alpha$ and the vector $\boldsymbol{\xi}^\alpha = (\xi^{i_1}, \ldots, \xi^{i_k}) \in \Delta^k := \{\mathbf{z} \in \mathbb{R}_+^k : \sum z_i \leq 1\}$. Specifically, $\xi^{i_0} = 1 - \mathbf{1} \cdot \boldsymbol{\xi}^\alpha$. With this identification, we define:

$$V_\alpha(u, t, \boldsymbol{\xi}^\alpha, a) = \tilde{V}_k(u, t, \xi, a).$$

We write $\mathbf{x}^\alpha = (x_{i_0}, x_{i_1}, \ldots, x_{i_k})$, and $\mathbb{S}^k = \{\mathbf{z} \in \mathbb{R}^k : \|z\| = 1\}$ for the unit sphere in $\mathbb{R}^k$. Finally, note that for $|\alpha| = 1$, $V_\alpha(t, a) = \tilde{V}_0(t, \xi, a) = F(a + (T-t)x_{i_0})$, and we then define $(1 - \mathbf{1} \cdot \boldsymbol{\xi}^\alpha, \boldsymbol{\xi}^\alpha) := \xi^{i_0} = 1$. We also use the convention $\mathbb{S}^0 := \emptyset$ and $\sup \emptyset := -\infty$.

**Theorem 3.7.** *Suppose $F(a)$ is continuous and non-negative. Fix $\alpha \subseteq \{0, \ldots, N\}$, with $|\alpha| \geq 1$, and write $k = |\alpha| - 1$. The function $V_\alpha : \mathbb{R}_+ \times [0,T] \times \Delta^k \times \mathbb{R}_+ \to \mathbb{R}$*

*is independent of $u$, and is the unique non-negative viscosity solution bounded by $F(a + (T - t)x_N)$, to*

$$(3.10) \qquad \max\left\{ \frac{\partial V_\alpha}{\partial t} + \mathbf{x}^\alpha \cdot (1 - \mathbf{1} \cdot \boldsymbol{\xi}^\alpha, \boldsymbol{\xi}^\alpha) \frac{\partial V_\alpha}{\partial a}, \sup_{\mathbf{w} \in \mathbb{S}^k} \left[ \mathrm{tr}(\mathbf{w}\mathbf{w}^\mathrm{T} D^2_{\boldsymbol{\xi}} V_\alpha) \right] \right\} = 0$$

*for $\boldsymbol{\xi}^\alpha \in (\Delta^k)^\circ$, and $t < T$, with the boundary conditions*

$$(3.11) \qquad \begin{array}{rcl} V_\alpha(u, T, \boldsymbol{\xi}^\alpha, a) &=& F(a) \\ V_\alpha(u, t, \boldsymbol{\xi}^{\alpha'}, a) &=& V_{\alpha'}(u, t, \boldsymbol{\xi}^{\alpha'}, a) \end{array}$$

*where the second equation is taken when $\boldsymbol{\xi}^\alpha \in \partial\Delta^k$. Here $\alpha'$ is the subset of $\alpha$ corresponding to non-zero entries of $(1 - \mathbf{1} \cdot \boldsymbol{\xi}^\alpha, \boldsymbol{\xi}^\alpha)$, and $\boldsymbol{\xi}^{\alpha'}$ is the vector identifying the corresponding probability measure.*

*Proof.* We work by induction; suppose the problem has been solved for $k' < k$, to give continuous value functions, independent of time. The case where $k = 0$ is trivial. The first step is to approximate by a problem with a finite time-horizon. To this end, we fix a sequence $K \nearrow \infty$, and consider the following two problems: For given $(u, t, \xi, a)$ with $\xi \in \mathcal{P}^1(\mathbb{X}_\alpha)$ and $|\alpha| = k + 1$, we set $\sigma^K = \sigma \wedge (K + 1)$ and define the functions $\tilde{V}_k^{\overline{K}}$ and $\tilde{V}_k^{\underline{K}}$ by

$$\tilde{V}_k^{\overline{K}}(u, t, \xi, a) = \sup_{\mathbf{w} \in \mathcal{A}_u^{k,K}} \mathbb{E}\Big[ \big( F(a + (T - t)x_N)(\sigma^K - K)_+ \big) \vee \Big( F(A_{T_{\sigma^K}})\mathbf{1}_{\{T_{\sigma^K} = T\}}$$
$$+ \tilde{V}_{k-1}(\sigma^K, T_{\sigma^K}, \xi_{\sigma^K}, A_{T_{\sigma^K}})\mathbf{1}_{\{T_{\sigma^K} < T\}} \Big) \Big| T_u = t, \xi_u = \xi, A_{T_u} = a \Big],$$

and

$$\tilde{V}_k^{\underline{K}}(u, t, \xi, a) = \sup_{\mathbf{w} \in \mathcal{A}_u^{k,K}} \mathbb{E}\Big[ \big( F(a + (T - t)x_N)(K + 1 - \sigma^K)_+ \big) \wedge \Big( F(A_{T_{\sigma^K}})\mathbf{1}_{\{T_{\sigma^K} = T\}}$$
$$+ \tilde{V}_{k-1}(\sigma^K, T_{\sigma^K}, \xi_{\sigma^K}, A_{T_{\sigma^K}})\mathbf{1}_{\{T_{\sigma^K} < T\}} \Big) \Big| T_u = t, \xi_u = \xi, A_{T_u} = a \Big],$$

where $\mathcal{A}_u^{k,K} = \big\{ (\mathbf{w}_s)_{s \in [u, K+1]} : \text{prog. meas. with } \mathbf{w}_s \in \mathbb{R}^k \text{ and } \|\mathbf{w}\| \le 1 \big\}$ and, as previously, $\sigma = \inf\big\{ s \ge u : \xi_s^n \notin (0, 1) \text{ some } n \in \{0, ..., k\} \text{ or } T_s = T \big\}$. With the same identification as above, we define $V_\alpha^{\overline{K}}$ and $V_\alpha^{\underline{K}}$ by

$$V_\alpha^{\overline{K}}(u, t, \boldsymbol{\xi}^\alpha, a) = \tilde{V}_{|\alpha(\xi)|-1}^{\overline{K}}(u, t, \xi, a) \quad \text{and} \quad V_\alpha^{\underline{K}}(u, t, \boldsymbol{\xi}^\alpha, a) = \tilde{V}_{|\alpha(\xi)|-1}^{\underline{K}}(u, t, \xi, a).$$

Recall that the dynamics of the involved factors is governed by (2.2), (3.3) and (3.8), with $\lambda_s$ given by (3.4). Note that without loss of generality, we may write $\mathcal{A}_u^{k,K} = \big\{ (\lambda_s, \mathbf{w}_s)_{s \in [u, K+1]} : (\lambda_s, \mathbf{w}_s) \text{ prog. meas. with } \mathbf{w}_s \in \mathbb{S}^k \text{ and } \lambda_s \in [0, 1] \big\}$. It follows from [30, Corollary V.3.1] that on the domain $[0, K + 1] \times [0, T] \times \mathbb{R}^k \times \mathbb{R}$, $V_\alpha^{\overline{K}}$ and $V_\alpha^{\underline{K}}$ are both viscosity solutions to

$$(3.12)$$
$$\frac{\partial V_\alpha}{\partial u} - \sup_{\substack{\mathbf{w} \in \mathbb{S}^k, \\ \lambda \in [0,1]}} \left[ \frac{1}{2}(1 - \lambda)\,\mathrm{tr}(\mathbf{w}\mathbf{w}^T D^2_{\boldsymbol{\xi}} V_\alpha) + \lambda \left( \frac{\partial V_\alpha}{\partial t} + \mathbf{x}^\alpha \cdot (1 - \mathbf{1} \cdot \boldsymbol{\xi}^\alpha, \boldsymbol{\xi}^\alpha) \frac{\partial V_\alpha}{\partial a} \right) \right] = 0,$$

equipped with the boundary conditions

$$(3.13) \qquad \left\{ \begin{array}{rcl} V_\alpha(u, T, \boldsymbol{\xi}^\alpha, a) &=& F(a) \\ V_\alpha(u, t, \boldsymbol{\xi}^{\alpha'}, a) &=& V_{\alpha'}(u, t, \boldsymbol{\xi}^{\alpha'}, a) \end{array} \right.$$

for $u < K$, and either increasing to $F(a + (T - t)x_N)$ for $u \in [K, K + 1]$ in the first case, or decreasing to 0 in $[K, K + 1]$ in the second case. In both cases, we have a viscosity equation with controls in a compact set, and with continuous boundary data on a compact domain. It follows that both equations have unique, continuous viscosity solutions, and the viscosity solutions to both equations correspond to the value functions of the corresponding optimal control problems. In particular, we

see immediately that $V_\alpha^{\overline{K}}(u,t,\boldsymbol{\xi},a) \geq V_\alpha(u,t,\boldsymbol{\xi},a) \geq V_\alpha^K(u,t,\boldsymbol{\xi},a)$ for $u \leq K+1$. Moreover, from Lemma 3.1, identifying $U$ and $V_\alpha$, we know the function $V_\alpha$ is continuous, and moreover, from Corollary 3.6, we know that $V_\alpha^{\overline{K}}(u,t,\boldsymbol{\xi},a)$ will decrease pointwise to $V_\alpha(u,t,\boldsymbol{\xi},a)$ as $K \to \infty$, and $V_\alpha^K(u,t,\boldsymbol{\xi},a)$ will increase pointwise to the same limit. We conclude that $V_\alpha$ is a viscosity solution to the given equation (see e.g. Barles and Souganidis [6]).

Now suppose that $W$ is another viscosity solution to the same equation, also nonnegative and bounded by $F(a + (T-t)x_N)$. By the comparison principle, for any $K$, $V_\alpha^{\overline{K}}(u,t,\boldsymbol{\xi},a) \geq W(u,t,\boldsymbol{\xi},a) \geq V_\alpha^K(u,t,\boldsymbol{\xi},a)$, for $u \leq K$. Hence $V_\alpha \geq W \geq V_\alpha$; that is, $W = V_\alpha$. Finally, we observe that the solution $V_\alpha$ is independent of $u$, by Lemma 3.1, so $\frac{\partial V_\alpha}{\partial u} = 0$, and optimising over $\lambda$ immediately gives the equivalent formulation. $\qquad \square$

**Remark** 3.8. We note that some obvious generalisations of this setup can easily be handled. For example, consider Asian options with non-constant weighting, so $\tilde{A}_T = \int_0^T f(t) S_t \, \mathrm{d}t$, for some (possibly signed) continuous function $f : [0,T] \to \mathbb{R}$. A simple modification to the arguments above gives the same result with the corresponding PDE:

$$\max\left\{\frac{\partial V_\alpha}{\partial t} + f(t)\mathbf{x}^\alpha \cdot (1 - \mathbf{1} \cdot \boldsymbol{\xi}^\alpha, \boldsymbol{\xi}^\alpha) \frac{\partial V_\alpha}{\partial a}, \sup_{\mathbf{w} \in \mathbb{S}^k} \left[\mathrm{tr}(\mathbf{w}\mathbf{w}^\mathrm{T} D_{\boldsymbol{\xi}}^2 V_\alpha)\right]\right\} = 0.$$

## 4. Examples and Superhedging

In this section we consider some simple cases where explicit solutions to the viscosity equations in Theorem 3.7 can be given. We also give some arguments regarding the construction of superhedging strategies. A number of the results in this section can be compared to the recent work of Stebegg [45], but we emphasise that our results extend beyond the case where $F$ is convex, and we will consider such an example below.

### 4.1. **Convex payoff functions.**

**Lemma 4.1.** *Suppose the function $F$ is convex and Lipschitz. Then for all $\xi \in \mathcal{P}^1(\mathbb{R}_+)$:*

$$U(t, \xi, a) = \int F(a + (T-t)x) \, \xi(\mathrm{d}x).$$

*Moreover, an optimal model is given by:*

$$S_{0-} = \int x \, \xi(\mathrm{d}x)$$
$$S_t = S_T, \qquad t \geq 0,$$

*where $S_T \sim \xi$.*

In terms of the class of models considered in Corollary 3.6, this corresponds to a model which takes $\lambda_u = 0$ until the measure $\xi_u \in \mathcal{P}^s$, and then setting $\lambda_u = 1$ until $T_u = T$.

*Proof.* By continuity, we are only required to check that (3.10) holds for atomic $\xi$. However, if we write $\overline{\xi} = \int x \, \xi(\mathrm{d}x)$, then

$$\frac{\partial U}{\partial t} + \overline{\xi} \frac{\partial U}{\partial a} = \int F'(a + (T-t)x)(\overline{\xi} - x) \, \xi(\mathrm{d}x)$$

$$\leq \int F'(a + (T-t)\overline{\xi})(\overline{\xi} - x) \, \xi(\mathrm{d}x) = 0.$$

Moreover, if $t = T$ or $\xi \in \mathcal{P}^s$ then we have equality.

Recalling the notation of Theorem 3.7, we have

$$U(t, \xi, a) = \sum_{j=0}^{|\alpha(\xi)|-1} F\left(a + (T-t)x_{i_j}\right)\xi^{i_j},$$

and computing the second derivatives, we have $D_{\boldsymbol{\xi}}^2 U = 0$. Hence (3.10) holds. $\quad\square$

In this convex case, we are easily able to provide a martingale inequality interpretation of this result. Indeed, this has already appeared in [45]. Since this will help our intuition, we provide an alternative approach to [45], which will enable us to gain insight into the optimal strategies for the non-convex case. We restrict first to the case where $F(a) = (a - K)_+$, for some $K > 0$, and we write $Y_t = A_t + (T-t)S_t$. We suppose also that $(S_t)$ is a continuous semi-martingale (although a pathwise analogue of this argument is possible, where $S_t$ is assumed just to have continuous paths). From the definition of local time, we get:

$$(A_T - K)_+ = (Y_T - K)_+ = (Y_0 - K)_+ + M_T + L_T^{Y,K},$$

where $M_T$ is a local martingale, and $L_T^{Y,K}$ is the local time of $Y$ at the level $K$. It follows from the definition of $Y$, that we have:

$$L_T^{Y,K} = \int_0^T (T-t)\, \mathrm{d}L_t^{S,K_t}, \quad \text{where } K_t = \frac{Y_t - A_t}{T - t},$$

so $L^{S,K_t}$ is the local time of the asset price along the curve $K_t$. That is, we have a local time contribution coming from the crossing of the curve $K_t$ by the asset price. However, for a given distribution of $S_T$, the local time at each value of $x$ is fixed, and decreases as $|x - S_0|$ increases. We now argue that $L_T^{Y,K}$ is maximised by trying to accumulate all the local time on the curve $K_t$ as close as possible to time zero: that is, all crossings of $S_t = K_t$ should happen as close to time zero as possible. This happens since if $S_t \neq K_t$, then $|S_t - K_t|$ is increasing, and there will be less local time which can later be accumulated at $K_t$ since the process needs to accumulate the local time at a (relatively) more distant point. In addition, the factor $(T - t)$ which appears in the integral only makes the weight of local time accumulated at later times smaller.

It follows (and again, this can be made rigorous) that the optimal model should make all crossings of $K_t$ necessary to embed in a short time interval. After this time, it is irrelevant how the process behaves, so long as it either remains above or below $K_t$.

**Remark** 4.2. The cases where there is a positive interest rate can be handled similarly (the process $Y_t = A_t + \frac{S_t}{\rho}\left(e^{\rho(T-t)} - 1\right)$ should be used instead). In addition, by adding constraints, one can extend to general convex functions, with the model which crosses each relevant curve $K_t$ corresponding to a convexity point of $F$ immediately being the optimal choice.

4.2. **A non-convex example.** In this section, we provide a solution to the problem for a non-convex example. Specifically, we use the intuition from the convex case established above to try and find a solution to the problem for a payoff function of the form:

(4.1) $\qquad F(A_T) = (A_T - K_1)_+ - (A_T - K_2)_+, \qquad K_1 < K_2.$

Given certain additional assumptions on the measure we wish to embed, we will then verify that an optimal model may be determined through the use of Theorem 3.7. We observe that the results of this paper simply verify the existence of an optimal model. Given the existence of an optimal model, the existence of a super-hedging strategy follows from general results (e.g. Dolinsky and Soner [24]).

The intuition established above suggests that we wish to gain the benefit of the convexity at $K_1$ immediately, while leaving the concavity at $K_2$ until as late as possible. However there is a trade-off, since the process may sacrifice some of the convexity at $K_1$ by waiting at $K_2$. To specify this, note that since the payoff is constant for $A_T \geq K_2$, it must be suboptimal to have positive support on events for which $Y_t > K_2$, $t \in (0, T]$. Indeed, the payoff will not be improved by this but the martingale property of $Y$ implies that more mass must then be put on events yielding an average strictly less than $K_2$. In consequence, at least for some values of $K_1, K_2$, it is natural to conjecture that at time 0, $S$ will either run to $K_2$, or to some lower level; at the lower level, the paths will behave as indicated by Lemma 4.1.

For a measure $\mu$ with continuous support, we therefore define the level $\eta$ by

$$(4.2) \qquad \eta := \inf \left\{ x \in \mathbb{R} : \int_\eta^\infty x \mu(\mathrm{d}x) \geq K_2 \right\}.$$

We then expect to accumulate all mass above $x = \eta$ into a branch of the underlying taking the value $S_t = K_2$, $t \in (0, T)$, and embedding $\mathbf{1}_{x \geq \eta} \mu(\mathrm{d}x)$ at $t = T$. As for the mass to be embedded on $[0, \eta)$, we expect the same optimal behaviour as detected for the convex case in Section 4.1. Put differently, at $u = T_0^{-1}$, with probability $\int_\eta^\infty \mu(\mathrm{d}x)$ we expect the measure-valued martingale $\xi_u$ to take the value $\frac{\mathbf{1}_{x \geq \eta} \mu(\mathrm{d}x)}{\mu([\eta, \infty))}$ and stay constant until $T_T^{-1}$, and with probability $\int_0^\eta \mu(\mathrm{d}x)$ we expect to recover the structure of Lemma 4.1.

To specify this, we restrict to a certain class of measures $\mu$. Specifically, we consider the problem at time $t \in [0, T]$ with current average $A_t = a$ when we take $\mathbf{x}^\alpha = (-1, 0, 1)$, so $|\alpha| = 3$, and consider the terminal distribution

$$(4.3) \qquad \mu = (1 - \beta - \gamma)\delta_{-1} + \beta \delta_0 + \gamma \delta_1, \qquad \beta, \gamma \in (0, 1).$$

That is, $\boldsymbol{\xi}^\alpha = (\beta, \gamma)$, and we write $V(t, a; \beta, \gamma) = V_\alpha(t, a; \boldsymbol{\xi}^\alpha)$. Further, we let $K_1 \in (-1, 1)$ and $K_2 \in (0, 1)$. Suppose now that $a + \frac{\gamma}{\gamma+\beta}(T-t) < K_2 \leq a + (T-t)$. That is, the expected averages considering the mass at both $x = 0$ and $x = 1$, and at $x = 1$ only, are, respectively, smaller and greater than $K_2$. Following the reasoning above, at $u = T_t^{-1}$, we then expect to have split $\xi_u$ into the two measures:

$$(4.4) \qquad \xi^1 = \frac{\bar{\eta}\delta_0 + \gamma \delta_1}{\bar{\eta} + \gamma} \quad \text{and} \quad \xi^2 = \frac{(1 - \gamma - \beta)\delta_{-1} + (\beta - \bar{\eta})\delta_0}{1 - \gamma - \bar{\eta}},$$

where (c.f. (4.2)) $\bar{\eta}$ is given by

$$\bar{\eta} = \sup \left\{ y : \frac{\gamma}{\gamma + y} \geq \frac{K_2 - a}{T - t} \right\} = \gamma \left( \frac{T - t}{K_2 - a} - 1 \right).$$

If $a < K_1$, this yields $V(t, a; \beta, \gamma) = (\gamma + \bar{\eta})(K_2 - K_1)$. However, if $a - (T - t) < K_1 \leq a$, the result for the convex case guides us to further split the measure $\xi^2$ into $\delta_{-1}$ and $\delta_0$ (equivalently, all mass at $x = 0$ and $x = 1$ might be accumulated in one measure; see further discussion below) and it follows that $V(t, a; \beta, \gamma) = (\gamma + \bar{\eta})(K_2 - K_1) + (\beta - \bar{\eta})(a - K_1)$. Similar considerations for the other cases guides us to define the following candidate value function:

$(4.5)$
$$V(t, a; \beta, \gamma) := \begin{cases} K_2 - K_1 & \text{(i)} & K_2 \leq a^{-101} \\ (2\gamma + \beta - 1)(T - t) + a - K_1 & \text{(ii)} & K_1 \leq a^{-1}, a^{-101} < K_2 \\ \frac{2\gamma + \beta}{1 + \frac{K_2 - a}{T - t}}(K_2 - K_1) & \text{(iii)} & a^{-1} < K_1, a^{-101} < K_2 \leq a^{01} \\ \gamma(T - t) - (\gamma + \beta)(K_1 - a) & \text{(iv)} & a^{-1} < K_1 \leq a^0, a^{01} < K_2 \\ \gamma \frac{T - t}{K_2 - a}(K_2 - K_1) & \text{(v)} & a^0 < K_1, a^{01} < K_2 \leq a^1 \\ \gamma(T - t - (K_1 - a)) & \text{(vi)} & a^0 < K_1 \leq a^1 < K_2 \\ 0 & \text{(vii)} & a^1 < K_1 \end{cases}$$

where we used the notation $a^i = a + s^i(T-t)$, with $s^i = i$, $i \in \{-1, 0, 1\}$, $s^{01} = \frac{\gamma}{\gamma+\beta}$ (with the convention $\frac{\gamma}{\gamma+\beta} = K_2$ when $\gamma+\beta = 0$) and $s^{-101} = s = 2\gamma + \beta - 1$ — i.e. the expected average taking the mass at various atoms into account. The function is depicted in Figure 1, together with a candidate sample path.
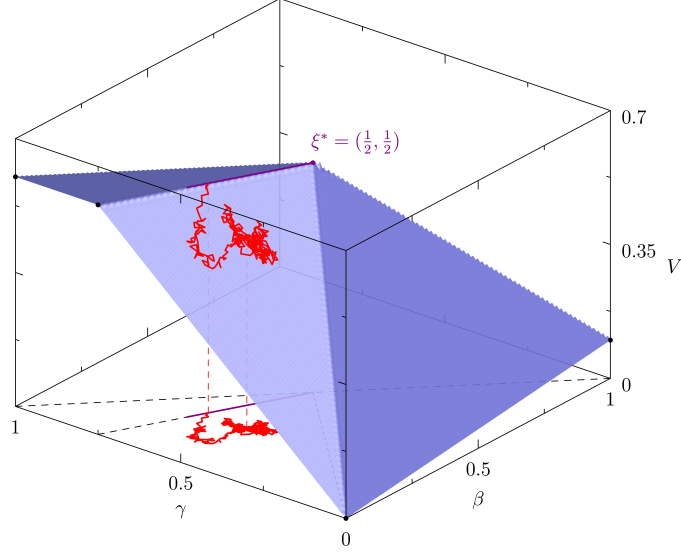


FIGURE 1. The value function graphed as a function of $\beta, \gamma$. Here $t = a = 0, T = 1$, $K_1 = -0.1, K_2 = 0.5$. Also shown (in red) is a possible path of $(\xi_r)$, starting from $(\beta, \gamma) = (\frac{1}{4}, \frac{1}{2})$. The planar regions in the diagram correspond to the regions (i), (iii) and (iv) given in (4.5). The process starts in region (iii), and runs until it hits the boundary of this region and region (i). The continuing path is then shown running along the boundary of regions (i) and (iii). In an optimal model, this behaviour happens at time 0, although note that there are many possible solutions: for example, the movement along the boundary between regions (i) and (iii) could happen at an time between $t = 0$ and $t = T$. On reaching the point $\xi^*$, the process is unable to move any further before the time $t = T$ without being suboptimal.

*Example* 4.3. Observe that sending $K_2 \to \infty$, $V(t, a; \beta, \gamma)$ reduces to the value-function for the (convex) payoff function $F(A_T) = (A_T - K_1)_+$ given in Section 4.1 with $\mu$ given by (4.3). Alternatively, let $\beta = 0$, $K_1 = 0$ and $K_2 = \frac{1}{2}$. Then, $V(t, a; \gamma)$ reduces in the following way:

$$(4.6) \qquad V(t, a; \gamma) = \begin{cases} \frac{1}{2}, & 2\gamma - 1 > \frac{1/2-a}{T-t} \\ \gamma \frac{1}{1 + \frac{1/2-a}{T-t}}, & 2\gamma - 1 \le \frac{1/2-a}{T-t} \end{cases}.$$

Note that for $t = a = 0$ and $2\gamma - 1 \le 1/2$, $V_t + sV_a = 0$ if and only if $\gamma = 0$ or $\gamma = 3/4$. Therefore the optimal model $(S_t)$ will jump to either $-1$ or $1/2$ at time $t = 0$, and then stay constant until time $t = T$ where $\mu$ is embedded.

It can be verified by elementary calculations that $V(t, a; \beta, \gamma)$ is continuous. The next result verifies that it is a (bounded) solution to equation (3.10) with $F(a)$ and $\mu$ given by (4.1) and (4.3). Hence, according to Theorem 3.7, $V(t, a; \beta, \gamma)$ is indeed the associated value function.

**Proposition 4.4.** *The function $V(t, a; \beta, \gamma)$ defined in (4.5) is the unique non-negative viscosity solution bounded by $K_2 - K_1$, to the equation (3.10) equipped with the boundary condition (3.11). In particular, it is the value-function associated with the payoff (4.1) and $\mu$ given by (4.3).*

*Proof.* Differentiating $V(t, a; \beta, \gamma)$ within the respective regions, we obtain that $V_t(t, a; \beta, \gamma) + s V_a(t, a; \beta, \gamma) = 0$ in regions (i), (ii) and (vii), and

$$(4.7) \qquad V_t + sV_a = \begin{cases} \frac{(K_2-K_1)(2\gamma+\beta)}{\left(1+\frac{K_2-a}{T-t}\right)^2} \left(-\frac{K_2-a}{(T-t)^2} + \frac{2\gamma+\beta-1}{T-t}\right) & \text{in (iii)} \\ -\gamma + (2\gamma+\beta-1)(\gamma+\beta) & \text{in (iv)} \\ -\gamma \frac{K_2-K_1}{(K_2-a)^2}\left(K_2 - a - (2\gamma+\beta-1)(T-t)\right) & \text{in (v)} \\ 2\gamma\left(\gamma + \frac{\beta}{2} - 1\right) & \text{in (vi)} \end{cases},$$

where $s = \mathbf{x}^\alpha \cdot (1-\beta-\gamma, \beta, \gamma) = 2\gamma + \beta - 1$. Using that $2\gamma + \beta - 1 \le \frac{K_2-a}{T-t}$ in regions (iii) and (v), and that $\gamma + \beta \le 1$ implies $(2\gamma + \beta - 1)\frac{\gamma+\beta}{\gamma} \le 1$ for $\gamma > 0$, it follows from (4.7) that $V_t + sV_a \le 0$ within the respective regions. In consequence, with $V_\mathbf{v}$ denoting the directional derivative in the direction of $\mathbf{v}$, it holds on $\{\gamma + \beta \le 1\}$ that

$$(4.8) \qquad\qquad V_\mathbf{v} \le 0, \qquad \mathbf{v} = (1, 2\gamma + \beta - 1, 0, 0).$$

Recall that $V(t, a; \boldsymbol{\xi})$ is a viscosity super (resp. sub) solution to (3.10) if for each $\varphi \in \mathcal{C}^{1,1,2}$, and at each point $(\bar{t}, \bar{a}, \bar{\boldsymbol{\xi}})$ minimizing (resp. maximizing) $V - \varphi$,

$$(4.9) \qquad \max\left\{\varphi_t + (2\bar{\gamma} + \bar{\beta} - 1)\varphi_a, \frac{1}{2}\sup_{\mathbf{w} \in \mathbb{S}^2}\left[\text{tr}\left(\mathbf{ww}^\mathrm{T} D_{\boldsymbol{\xi}}^2 \varphi\right)\right]\right\} \le 0, \quad (\text{resp. } \ge 0).$$

We first argue that $V$ is a sub solution. To this end, let $\varphi \in \mathcal{C}^{1,1,2.2}$ and $\bar{\mathbf{z}} = (\bar{t}, \bar{a}, \bar{\beta}, \bar{\gamma})$ such that $\bar{\mathbf{z}}$ maximises $V - \varphi$. Note that there exists $\bar{\mathbf{w}} \in \mathbb{S}^2$ such that the directional derivative at $\bar{\mathbf{z}}$ along $\bar{\mathbf{w}}$ (keeping $a$ and $t$ constant) satisfies $V_{\bar{\mathbf{w}}\bar{\mathbf{w}}} = 0$; if $\bar{\mathbf{z}} \in \{2\gamma + \beta - 1 = \frac{K_2-a}{T-t}\}$ or $\bar{\mathbf{z}} \in \{\frac{\gamma}{\gamma+\beta} = \frac{K_2-a}{T-t}\}$, let $\bar{\mathbf{w}}$ in the direction of that line. Since $(V - \varphi)_{\bar{\mathbf{w}}\bar{\mathbf{w}}} \le 0$, it follows that $\text{tr}\left(\bar{\mathbf{w}}\bar{\mathbf{w}}^\mathrm{T} D_{\boldsymbol{\xi}}^2 \varphi\right) \ge 0$ (note that $\text{tr}\left(\mathbf{ww}^\mathrm{T} D_{\boldsymbol{\xi}}^2 \varphi\right) = \mathbf{w}^\mathrm{T} D_{\boldsymbol{\xi}}^2 \varphi \, \mathbf{w} = \varphi_{\mathbf{ww}}$). In consequence,

$$(4.10) \qquad\qquad \sup_{\mathbf{w} \in \mathbb{S}^2}\left[\text{tr}\left(\mathbf{ww}^\mathrm{T} D_{\boldsymbol{\xi}}^2 \varphi\right)\right] \ge 0.$$

In order to show that $V$ is a super solution, let $\varphi \in \mathcal{C}^{1,1,2,2}$ and let $\bar{\mathbf{z}} = (\bar{t}, \bar{a}, \bar{\beta}, \bar{\gamma})$ be a minimiser to $V - \varphi$. Due to the concavity of $V$ as a function of $\beta$ and $\gamma$ and the differentiability of $\varphi$, $\bar{\mathbf{z}}$ must lie strictly within one of the regions given in (4.7). Hence, for all $\mathbf{w} \in \mathbb{S}^2$, $\text{tr}\left(\mathbf{ww}^\mathrm{T} D_{\boldsymbol{\xi}}^2 (V - \varphi)\right) \ge 0$ and $\text{tr}\left(\mathbf{ww}^\mathrm{T} D_{\boldsymbol{\xi}}^2 V\right) = 0$. In consequence,

$$(4.11) \qquad\qquad \sup_{\mathbf{w} \in \mathbb{S}^2}\left[\text{tr}\left(\mathbf{ww}^\mathrm{T} D_{\boldsymbol{\xi}}^2 \varphi\right)\right] \le 0.$$

Let now $\mathbf{v} = (1, 2\bar{\gamma} + \bar{\beta} - 1, 0, 0)$. Since $\bar{\mathbf{z}}$ minimises $V - \varphi$, it follows that $\varphi_\mathbf{v} \le V_\mathbf{v}$. According to (4.8), we thus obtain

$$\varphi_t + (2\bar{\gamma} + \bar{\beta} - 1)\,\varphi_a \ \le \ V_\mathbf{v} \ \le \ 0,$$

which combined with (4.11) renders (4.9).

It remains to argue the boundary conditions (3.11). Note that for $t = T$, the only possible regions are (i), (ii) and (vii) (for $a \ge K_2$, $a \in [K_1, K_2)$ and $a < K_1$) for which $V(t, a; \beta, \gamma)$ is given, respectively, by $K_2 - K_1$, $a - K_1$ and $0$. Hence, the terminal condition is satisfied. Next, note that for $\gamma = 0$ and $\beta = 0$ or $\gamma + \beta = 1$, the problem reduces, respectively, to the convex case and the case presented in Example 4.3. This verifies the second boundary condition and we conclude. $\square$

We now discuss the optimal control associated with the value function (4.5) and the corresponding solution to the basic optimisation problem, Problem 2.8. Indeed, recall Lemma 2.11, which says that if Problem 2.10 admits an optimal solution, this solution corresponds to a solution of Problem 2.8. Naturally, the solution coincides with the conjectured optimal model used to deduce the form of $V(t, a; \beta, \gamma)$. However, our aim below is to illustrate how it may be deduced directly from the value function and, in consequence, from the dynamic programming equation (3.12)–(3.13) and to show that it is non-unique, and also non-trivial since it necessarily has a jump $t = T$ as well as $t = 0$. We let $T = 1$ and split the behaviour into three parts.

(I) *Real time is kept constant while the measure-valued martingale evolves ($\lambda_u = 0$ and $(\beta_u, \gamma_u)$ diffuses): $(S_t)$ jumps to certain points at time $t = 0$.*

Depending on the parameters of the problem, the starting point $(0, 0; \beta, \gamma)$ lies in one of the regions (i), (iii), (iv) or (v). It follows from the DPP equation (3.10), that the model can evolve in real time only if $V_t + sV_a = 0$. According to (4.7), while $V_t + sV_a = 0$ for all $(\beta, \gamma)$ in regions (i), (ii) and (vii), it holds for the remaining regions that $V_t + sV_a = 0$, if and only if,

$$(4.12) \quad \begin{cases} 2\gamma + \beta - 1 = \frac{K_2 - a}{T - t} \quad \text{or} \quad (\beta, \gamma) = (0, 0) & \text{in (iii)} \\ \gamma + \beta = 1 \quad \text{or} \quad (\beta, \gamma) = (0, 0) & \text{in (iv)} \\ 2\gamma + \beta - 1 = \frac{K_2 - a}{T - t} \quad \text{or} \quad \gamma = 0 & \text{in (v)} \\ \gamma = 0 \quad \text{or} \quad (\beta, \gamma) = (0, 1) & \text{in (vi)} \end{cases} \; .$$

In consequence, if starting in region (i), one may immediately evolve in (real) time. However, if starting in regions (iii)–(v), (real) time must be kept constant while $(\beta_u, \gamma_u)$ are allowed to diffuse until the above boundaries are reached: that is, $\lambda_u = 0$ until the measure-valued martingale $\xi_. = (\beta_., \gamma_.)$ satisfies (4.12). Note that since $V(0, 0; \beta_u, \gamma_u)$ is a martingale, if at the line $\frac{\gamma}{\gamma + \beta} = \frac{K_2 - a}{T - t}$, diffusion will take place only in the direction of that line and the region remains the same until the boundaries are reached. This implies that the associated price process $(S_t)$ jumps to certain points at time $t = 0$.

(II) *Progress in real time only ($\lambda_u = 1$): $(S_t)$ is kept constant.*

Once the measure-valued martingale satisfies (4.12), (real) time might start to evolve ($\lambda_u > 0$). In particular, one might let $\lambda_u = 1$ which implies that $(S_t)$ is kept constant. On a case by case basis, it can be verified that once at a point where $V_t + sV_a = 0$, this remains the case. For example, consider sitting at $\{\gamma + \beta = 1\}$ in region (iv). With a slight abuse of notation, we see that at this line

$$(4.13) \quad \frac{\partial}{\partial t} a^i(t) = \begin{cases} \frac{\partial}{\partial t} a(t) = 2\gamma + \beta - 1 \geq 0, & i = 0 \\ \frac{\partial}{\partial t}\big(a(t) - (T - t)\big) = 2\gamma + \beta \geq 0, & i = -1 \\ \frac{\partial}{\partial t}\big(a(t) + \frac{\gamma}{\gamma + \beta}(T - t)\big) = 0, & i = 1 \end{cases} \; .$$

Hence, when evolving in (real) time, and with no change in $\beta, \gamma$, $a^{-1}(t) = a(t) - (T - t)$ may move above $K_1$ and so the point $(t, a(t); \beta, \gamma)$ moves to region (ii). Since $V_t + sV_a = 0$ within region (ii), the claim holds for this case. Similar arguments apply to the other cases.

(III) *When $T_u = T$, the measure-valued martingale $\xi_. = (\beta_., \gamma_.)$ terminates: $(S_t)$ jumps and embeds $\mu$ at $t = T$.*

Real time may run until $T_u = T$. Thereafter $\lambda_u = 0$ and $(\beta, \gamma)$ diffuses until $\xi_u$ terminates; that is, until $\xi_. = (\beta_., \gamma_.)$ reaches $(0, 0)$, $(0, 1)$ or $(1, 0)$. As expected, $V(T, a(T); \beta, \gamma)$ stays constant during this process as it is independent of $\gamma$ and $\beta$. This step corresponds to $S$ embedding $\mu$ via a jump at $t = T$.

The evolution in time and measure described in (II) and (III) could, partially, happen simultaneously or in the reverse order. This implies that the optimal model

is not unique. For example, having reached the line $\gamma + \beta = 1$ in region (iv), one might let $(\beta_u, \gamma_u)$ continue to diffuse until reaching either of the points $(1 - \frac{K_2 - A_u}{T - u}, \frac{K_2 - A_u}{T - u})$ or $(1, 0)$, before letting (real) time evolve. This corresponds to the behaviour used to deduce $V(t, a; \beta, \gamma)$: i.e. $(S_t)$ jumps to one of the values $-1$, $0$ or $K_2$ at time $t = 0$. Alternatively, by letting $\lambda_u \in (0, 1)$, and supposing that $A_t \geq K_1$, one may let time and measure evolve simultaneously, which corresponds to $Y_u = A_u + (T - u)S_u$, $u \in (t, 1)$ being either constantly equal to $-1$ or moving (as a continuous martingale) between the values $0$ and $K_2$. Observe that this behaviour may result in a different distribution to $A_T$ in comparison with the case where all the diffusion happens immediately. Similar behaviour can be observed in the regions (i), (ii) and (v), although the distribution of $F(A_T)$ then remains the same.

While the optimal model is not unique, we note that it has certain characteristics: the model necessarily has a jump at both $t = 0$ and $t = T$. Indeed, there is a certain amount of mass which is 'locked in', and cannot be embedded until the terminal time $t = T$. This to ensure that $S_t = K_2$, $t \in (0, T)$, with a certain probability (e.g. in regions (iii) and (v)). On the other hand, sending $K_2 \to \infty$ and thus isolating the behaviour at the convex kink $K_1$, we see that the mass terminating above/below $K_1$ must already at time $t = 0$ be accumulated above/below $K_1$. Although affected by the presence of $K_2$, this feature is present also for the general case (e.g. regions (iv) and (vi)).

## 5. CONCLUSIONS AND FURTHER WORK

In this paper we have considered the model-independent pricing problem for Asian options using a novel approach based on measure-valued martingales. While this paper concentrated on the case of Asian options, the main ideas should generalise to other cases, and may provide insights beyond the existing literature. Moreover, there are a number of natural questions which arise from our work:

- Is it possible to generalise the results in this paper to the case of a general starting law? Financially, this has the interpretation of pricing a forward starting option at time 0, where $0 < T_0 < T_1$, the option pays the holder the amount $F\left(\int_{T_0}^{T_1} S_u \, du\right)$ at time $T_1$ and the price of call options are known at times $T_0$ and $T_1$. Write $\lambda$ for the implied law of $S_{T_0}$ and $\mu$ for the implied law of $S_{T_1}$. It follows immediately from the results of this paper that the problem is equivalent to finding a function $m : \mathbb{R} \to \mathcal{P}^1, x \mapsto m_x$ which maximises $\int U(0, m_x, 0) \, \lambda(\mathrm{d}x)$ over all functions $m$ such that $\int m_x(A) \, \lambda(\mathrm{d}x) = \mu(A)$, for all Borel sets $A$ and $x = \int y \, m_x(\mathrm{d}y)$. However, it would be interesting to have a dynamic formulation of the problem, similar to the simple case.
- The PDE (3.10) is formulated for the case of atomic measures. Is there a similar formulation that holds when $\xi$ is only assumed to be measure-valued?
- What is the corresponding formulation for (3.10) in the case of (say) options on variance?
- Do the methods described above extend to related problems in higher dimensions? If the formulation is given in terms of a measure $\mu$ on $\mathbb{R}^d$, one might hope that a very similar approach would be possible. Is this also true of (the financially more meaningful) case where $S_t \in \mathbb{R}^d$, and the marginal distribution of each component of $S_T$ is specified?

## Appendix A. A Formal Dynamic Programming Principle

In this section, we formally derive a Dynamic Programming Principle (DPP) for the pricing problem in its weak form given in Definition 2.10. We note that our previous results do not make use of this DPP, but we believe that this result is of independent interest. We choose to follow closely the setup in [46]; see however also [27] and [44] for similar arguments.

We denote by $\mathbb{D}$ the set of càdlàg paths on $[0, \infty)$ taking values in $E := \mathcal{P}(\mathbb{R}) \times [0, T] \times \mathbb{R}$, where we equip $\mathcal{P}(\mathbb{R})$ with the topology induced by the $\mathcal{W}_1$-metric and $E$ with the product topology; in particular, this renders $E$ a Polish space, and using the Skorokhod topology on $\mathbb{D}$ it is a Polish space too. For $x, x' \in E$, we write $d(x, x') := \mathcal{W}_1(\xi, \xi') \vee |t - t'| \vee |a - a'|$. A generic path in $\mathbb{D}$ is denoted by $\omega$ and we use $X = (\xi, T, A)$ for the co-ordinate process: $X_r(\omega) = (\xi_r, T_r, A_r)(\omega) = \omega(r)$.

The set of all probability measures on $\mathcal{B}(\mathbb{D})$ is denoted by $\mathfrak{P}$. A map $\nu : E \times \mathcal{B}(\mathbb{D}) \to [0, 1]$ is called a (universally) measurable kernel if i) $\nu(x, \cdot) \in \mathfrak{P}$ for all $x \in E$, and ii) $E \ni x \to \nu(x, A)$ is (universally) measurable for all $A \in \mathcal{B}(\mathbb{D})$; recall that the universal $\sigma$-algebra is the intersection of the completions of the Borel $\sigma$-algebra over all probability measures on the space, and that universally measurable functions are integrable with respect to any such probability measure. We write $\nu_x$ for the probability measure $\nu(x, \cdot)$ and interpret $\nu$ as a (universally) measurable map $E \to \mathfrak{P}$.

A Borel-measurable map from $\mathbb{D}$ to $[0, \infty)$ is called a random time. For any random time $\tau$, we define the shift-operator $\theta_\tau$ on $\mathbb{D}$ by $X_r(\theta_\tau(\omega)) = X_{\tau(\omega)+r}(\omega)$. Further, for a random time $\tau$ and any two paths $\omega, \omega' \in \mathbb{D}$ such that $X_\tau(\omega) = X_0(\omega')$, the concatenation $\omega *_\tau \omega'$ is an element of $\mathbb{D}$ specified by

$$X_t(\omega *_\tau \omega') = \mathbf{1}_{\{t < \tau(\omega)\}} X_t(\omega) + \mathbf{1}_{\{t \geq \tau(\omega)\}} X_{t-\tau(\omega)}(\omega').$$

For a random time $\tau$, a probability measure $\mu \in \mathfrak{P}$ and a universally measurable kernel $\nu$, we then define the concatenation $\mu *_\tau \nu$ as the probability measure in $\mathfrak{P}$ given by

$$(\mu *_\tau \nu)(A) = \iint \mathbf{1}_A(\omega *_\tau \omega') \nu_{X_\tau(\omega)}(\mathrm{d}\omega') \mu(\mathrm{d}\omega), \qquad A \in \mathcal{B}(\mathbb{D}).$$

We let $\mathbb{F}^0 = \{\mathcal{F}_r^0\}_{r \in [0, \infty)}$ denote the filtration generated by the co-ordinate process $X$, and let $\mathbb{F} = \{\mathcal{F}_r\}_{r \in [0, \infty)}$ be its right-continuous hull; i.e. $\mathcal{F}_r = \cap_{s > r} \mathcal{F}_s^0$, for $r \geq 0$. For $x = (\xi, t, a) \in E$, we denote by $\mathcal{P}_x$ the set of probability measures in $\mathfrak{P}$ for which:

 (i) $X_0 = x$ a.s.,
 (ii) $\xi_r$ is a measure-valued $\mathbb{F}$-martingale,
 (iii) $T_r$ is non-decreasing with $\lim_{r \to \infty} T_r = \infty$, a.s.,
 (iv) $A_r = a + \int_0^{r \wedge \tau_0} \bar{\xi}_{u-} \mathrm{d} T_u$ a.s., where $\bar{\xi}_\cdot = \int x \xi_\cdot(\mathrm{d}x)$ and $\tau_0 = \inf\{r : T_r \geq T\}$.

Finally, we note that according to Lemma 3.12 in [46], there exists a measurable functional $\bar{X} = (\bar{\xi}, \bar{T}, \bar{A}) : \mathbb{D} \to E$ such that $\bar{X}(\omega) = \lim_{t \to \infty} X_t(\omega)$ whenever the limit exists and $\bar{X}(\theta_t(\omega)) = \bar{X}(\omega)$ for all $t \geq 0$. We let $G(\omega) := F(\bar{A}(\omega))$; for any $\mu \in \mathcal{P}_x$, $x \in E$, we then have that $G = \lim_{t \to \infty} F(A_t)$ a.s. We define the problem:

$$(A.1) \qquad\qquad v(x) = \sup_{\mu \in \mathcal{P}_x} \mathbb{E}^\mu[G].$$

**Lemma A.1.** *The value function defined in* (A.1) *coincides with the value function given in* (3.1). *In particular,* $x \mapsto v(x)$ *is continuous.*

*Proof.* Let $(\Omega, \mathcal{G}, (\mathcal{G}_r), \mathbb{P}, (\xi_r), (\lambda_r))$ be a multiple as specified in Problem 2.10. Without loss of generality, let $x = (\mu, 0, 0)$ (Note that in Problem 2.10 it was assumed only that $\xi_{0-} = x$; by considering a time transformation $t \mapsto \frac{(t-\varepsilon)_+}{T-\varepsilon} T$,

this difference can be seen to be irrelevant). Since any martingale is a martingale in its own filtration, it follows that any such multiple induces on the canonical space $\mathbb{D}$ a measure $\mu \in \mathcal{P}_x$. Conversely, any probability measures $\mu \in \mathcal{P}_x$ together with the space $(\mathbb{D}, \mathcal{B}(\mathbb{D}), \mathbb{F})$ and the canonical process $(\xi, T)$ produces such a multiple. Indeed, the fact that one may, without loss of generality, assume that $T$. is absolutely continuous a.s. follows as in the proof of Lemma 2.11. Moreover, since $\tau_0 < \infty$ a.s., for any pair $(T_r, \xi_r)$ with $\xi$ a measure-valued martingale, one may construct a terminating measure valued martingale $\tilde{\xi}_r$ such that $(T_r, \tilde{\xi}_r)$ yields the same value of the payoff. The continuity is then an immediate consequence of Lemma 3.1. $\qquad \square$

**Remark** A.2. Let $\tilde{\mathbb{F}}^0$ be the filtration generated by $(T, \bar{\xi})$, where $\bar{\xi}. = \int x \xi.(\mathrm{d}x)$, and let $\tilde{\mathbb{F}}$ be its right-continuous hull. Further, let $\tilde{\mathcal{P}}_x$ denote the set of measures in $\mathfrak{P}$ which satisfy properties (i) to (iv) above with the difference that $\xi$ is only assumed to be a measure-valued $\tilde{\mathbb{F}}$-martingale. We then have that

$$(A.2) \qquad v(x) = \sup_{\mu \in \tilde{\mathcal{P}}_x} \mathbb{E}^\mu [G].$$

Indeed, this follows from the proof of Lemma 2.11, where the constructed measure-valued martingales are indeed adapted to the filtration generated by $T$. and $\bar{\xi}$..

We are now ready to state the DPP. For simplicity we provide it here for bounded payoff functions. We denote by $\mathcal{T}$ the set of finite $\mathbb{F}$-stopping times.

**Theorem A.3.** *Let $F : \mathbb{R}_+ \to \mathbb{R}_+$ be bounded and Lipschitz. Then, for all $x \in E$ and $\tau \in \mathcal{T}$,*

$$v(x) = \sup_{\mu \in \mathcal{P}_x} \mathbb{E}\left[ v(X_\tau) \right].$$

*Proof.* Given $\varepsilon > 0$, $x \in E$ and $\tau \in \mathcal{T}$, take $\mu \in \mathcal{P}_x$ such that $v(x) - \varepsilon \leq \mathbb{E}^\mu[G]$. Let $\nu_x$ be the regular conditional probability distribution of $\theta_\tau$ under $\mu$ given $X_\tau = x$; since $\mathbb{D}$ is Polish it exists $\mu \circ X_\tau^{-1}$-a.s. Recall that for any $f \in C_b(\mathbb{R}_+)$, $\xi.(f)$ is a bounded $\mu$-martingale. By use of the same argument as given in the proof of Proposition 3.11 in [46], we may then conclude that $\xi.(f)$ is a $\nu_x$-martingale for $\mu \circ X_\tau^{-1}$-almost all $x \in E$. It follows that $\nu_x \in \mathcal{P}_x$ for $\mu \circ X_\tau^{-1}$-almost all $x \in E$. As argued in the proof of Proposition 2.5 in [46], we may further pick a universally measurable version of $\nu_x$ such that $\nu_x \in \mathcal{P}_x$ for all $x \in E$. Now, note that $G(\omega') = G(\omega *_\tau \omega')$ for all $\omega, \omega' \in \mathbb{D}$ with $X_\tau(\omega) = X_0(\omega')$, and thus

$$\iint G(\omega') \nu_{X_\tau(\omega)}(\mathrm{d}\omega') \mu(\mathrm{d}\omega) = \iint G(\omega *_\tau \omega') \nu_{X_\tau(\omega)}(\mathrm{d}\omega') \mu(\mathrm{d}\omega).$$

By use of the properties of the r.c.p.d., we thus obtain the following line of equalities:

$$\mathbb{E}^\mu[G] = \mathbb{E}^\mu[G \circ \theta_\tau] = \mathbb{E}^{\mu *_\tau \nu}[G] = \mathbb{E}^\mu[g(X_\tau)],$$

where $g(x) = \mathbb{E}^{\nu_x}[G] = \int G(\omega')\nu_x(\mathrm{d}\omega')$. Hence, $v(x) - \varepsilon \leq \mathbb{E}^\mu[v(X_\tau)]$ for some $\mu \in \mathcal{P}_x$, and since $\varepsilon$ was chosen arbitrarily we obtain $v(x) \leq \sup_{\mu \in \mathcal{P}_x} \mathbb{E}\left[ v(X_\tau) \right]$.

In order to argue the reverse inequality, for any $\varepsilon > 0$, we first argue the existence of a measurable kernel $\nu$ with $\nu_x \in \mathcal{P}_x$ and $\mathbb{E}^{\nu_x}[G] \geq v(x) - \varepsilon$, for each $x \in E$. To this end, we define a mapping $E \times \mathbb{D} \ni (x, \bar{\omega}) \mapsto \omega^{x, \bar{\omega}} \in \mathbb{D}$ such that for each $x = (\xi, t, a) \in E$, the mapping $\bar{\omega} = (\bar{\xi}., \bar{t}., \bar{a}.) \mapsto \alpha^x(\bar{\omega}) := \omega^{x, \bar{\omega}} = (\xi, t, a.)$ modifies the path $\bar{\omega}$ as follows:

$$(A.3) \qquad \begin{cases} \xi_r(\mathrm{d}y) &= \int \bar{\xi}_r(\mathrm{d}x) m^\xi(x, \mathrm{d}y), \\ t_r &= \bar{t}_r + t - \bar{t}_0, \\ a_r &= a + \int_0^{r \wedge \tau_0} \int x \xi_{u-}(\mathrm{d}x) \mathrm{d}t_u, \end{cases}$$

where the family $m^\xi(\cdot, \mathrm{d}y)$ satisfies $\mathcal{W}_1(\bar{\xi}_0, \xi) = \iint |x - y|\bar{\xi}_0(\mathrm{d}x)m^\xi(x, \mathrm{d}y)$, and $\tau_0 = \inf\{r : t_r = T\}$. Then $(x, \bar{\omega}) \mapsto \omega^{x,\bar{\omega}}$ is $\mathcal{B}(E) \times \mathcal{B}(\mathbb{D})$ measurable. Hence, for any $\bar{\mathbb{P}} \in \cup_{x \in E}\mathcal{P}_x$, defining $\bar{\nu}_x := \bar{\mathbb{P}} \circ (\alpha^x)^{-1}$, $x \in E$, yields a measurable kernel $\bar{\nu}$ with $\bar{\nu}_x \in \mathcal{P}_x$. Indeed, the martingale property of $\xi$ under $\bar{\nu}_x$ follows as in the proof of Lemma 3.1. Further, from (A.3) we have that

$$|\bar{a}_\infty - a_\infty| \leq \int_{\bar{t}_0}^T \left| \int x\bar{\xi}_{\bar{t}_s^{-1}}(\mathrm{d}x) - \int x\xi_{\bar{t}_s^{-1}}(\mathrm{d}x) \right| \mathrm{d}s$$
$$+ \int_{T-(t-\bar{t}_0)}^T \left( \int x\bar{\xi}_{\bar{t}_s^{-1}}(\mathrm{d}x) + \int x\xi_{\bar{t}_s^{-1}}(\mathrm{d}x) \right) \mathrm{d}s + |\bar{a}_0 - a|.$$

Proceeding as in the proof of Lemma 3.1, for any $\bar{\mathbb{P}} \in \mathcal{P}_{\bar{x}}$ with $\bar{x} = (\bar{\xi}, \bar{t}, \bar{a})$, we then have that

$$\mathbb{E}^{\bar{\mathbb{P}}}\big[\big|A_\infty(\bar{\omega}) - A_\infty(\alpha^x \circ \bar{\omega})\big|\big] \leq T\,\mathcal{W}_1(\bar{\xi}, \xi) + (t - \bar{t})\int x\bar{\xi}(\mathrm{d}x) \vee \int x\xi(\mathrm{d}x) + |\bar{a} - a|,$$

and with $\bar{\nu}_x = \bar{\mathbb{P}} \circ (\alpha^x)^{-1}$, the Lipschitz property of $F$ thus yields $|\mathbb{E}^{\bar{\mathbb{P}}}[G] - \mathbb{E}^{\bar{\nu}_x}[G]| \leq \delta^{\bar{\xi}}(d(\bar{x}, x))$ for some modulus of continuity $\delta^{\bar{\xi}}$. Now, let $\varepsilon > 0$, and let $\{x^n\}_{n \in \mathbb{N}}$ be a countable dense subset of $E$. For each $n$, let $\mathbb{P}_n \in \mathcal{P}_{x^n}$ such that $\mathbb{E}^{\mathbb{P}_n}[G] \geq v(x^n) - \frac{\varepsilon}{3}$. Further, for each $x^n$, let $r_n$ such that for all $x \in B^n := \{x \in E : d(x, x^n) \leq r_n\}$, it holds that $v(x^n) \geq v(x) - \frac{\varepsilon}{3}$ and $\big|\mathbb{E}^{\mathbb{P}_n}[G] - \mathbb{E}^{\nu_x^n}[G]\big| \leq \frac{\varepsilon}{3}$ with $\nu_x^n := \mathbb{P}_n \circ (\alpha^x)^{-1}$; the existence of such $r_n$, $n \in \mathbb{N}$, follows from the above and Lemma 3.1. We then define the measurable kernel $(\nu_x)_{x \in E}$ by

$$(\text{A.4}) \qquad \nu_x := \sum_{n \in \mathbb{N}} \mathbf{1}_{C^n}(x)\, \mathbb{P}_n \circ (\alpha^x)^{-1}, \quad \text{where} \quad C^n = B^n \setminus \bigcup_{k-1}^{n-1} B^k.$$

By construction, for $x \in C^n$, $n \in \mathbb{N}$, we then have that

$$\mathbb{E}^{\nu_x}[G] \geq \mathbb{E}^{\mathbb{P}_n}[G] - \frac{1}{3}\varepsilon \geq v(x^n) - \frac{2}{3}\varepsilon \geq v(x) - \varepsilon.$$

Hence, $\nu$ is a measurable kernel with $\nu_x \in \mathcal{P}_x$ and $\mathbb{E}^{\nu_x}[G] \geq v(x) - \varepsilon$, for $x \in E$.

In order to conclude, we take $x_0 \in E$, $\mu \in \mathcal{P}_{x_0}$, $\tau \in \mathcal{T}$ and $\nu$ as constructed in (A.4). Since $\xi.(f)$ is a bounded $\mu$-martingale for any $f \in C_b(\mathbb{R}_+)$, we may use the same arguments as in the proof of Proposition 3.10 in [46] to deduce that $\xi.(f)$ is also a $\mu *_\tau \nu$-martingale. We may thus conclude that $\mu *_\tau \nu \in \mathcal{P}_{x_0}$. Letting $g(x) = \mathbb{E}^{\nu_x}[G]$ and noticing that $g$ is measurable, we thus obtain

$$v(x_0) \geq \mathbb{E}^{\mu *_\tau \nu}[G] = \mathbb{E}^\mu[g(X_\tau)] \geq \mathbb{E}^\mu[v(X_\tau)] - \varepsilon.$$

Since $\varepsilon$ and $\mu \in \mathcal{P}_{x_0}$ were both chosen arbitrarily, we obtain $v(x) \geq \sup_{\mu \in \mathcal{P}_x} \mathbb{E}\left[v(X_\tau)\right]$ and conclude. $\qquad\square$

The above proof exploits the continuity properties of our problem in order to construct an approximately optimal measurable kernel; see [2] and [12] for similar approaches.

## References

[1] B. Acciaio, M. Beiglböck, F. Penkner, and W. Schachermayer. "A model-free version of the fundamental theorem of asset pricing and the super-replication theorem". *Mathematical Finance* 26.2 (2016), pp. 233–251. DOI: 10.1111/mafi.12060.

[2] A. Aksamit, Z. Hou, and J. Obłój. "Robust framework for quantifying the value of information in pricing and hedging". arXiv: 1605.02539. 2016. URL: http://arxiv.org/abs/1605.02539.

[3] H. Albrecher, P. A. Mayer, and W. Schoutens. "General Lower Bounds for Arithmetic Asian Option Prices". *Applied Mathematical Finance* 15.2 (2008), pp. 123–149. DOI: 10.1080/13527260701356633.

[4] H. Albrecher, J. Dhaene, M. Goovaerts, and W. Schoutens. "Static Hedging of Asian Options under Lévy Models". *The Journal of Derivatives* 12.3 (2005), pp. 63–72. DOI: 10.3905/jod.2005.479381.

[5] L. Ambrosio, N. Gigli, and G. Savaré. *Gradient Flows - In Metric Spaces and in the Space of Probability Measures*. Second Edition. Birkhäuser, 2008.

[6] G. Barles and P. Souganidis. "Convergence of approximation schemes for fully nonlinear second order equations". *Asymptotic Analysis* 4.3 (1991), pp. 271–283. DOI: 10.3233/ASY-1991-4305.

[7] E. Bayraktar and Z. Zhou. "On Arbitrage and Duality Under Model Uncertainty and Portfolio Constraints". *Mathematical Finance* (2015), n/a–n/a. DOI: 10.1111/mafi.12104.

[8] M. Beiglböck, P. Henry-Labordère, and F. Penkner. "Model-independent Bounds for Option Prices: A Mass Transport Approach". *Finance Stoch.* 17.3 (2013), pp. 477–501.

[9] M. Beiglböck, A. M. G. Cox, and M. Huesmann. "Optimal Transport and Skorokhod Embedding". arXiv: 1307.3656. 2013. URL: http://arxiv.org/abs/1307.3656.

[10] M. Beiglböck, A. M. G. Cox, M. Huesmann, N. Perkowski, and D. J. Prömel. "Pathwise super-replication via Vovk's outer measure". arXiv: 1504.03644. 2015. URL: http://arxiv.org/abs/1504.03644.

[11] S. Biagini, B. Bouchard, C. Kardaras, and M. Nutz. "Robust Fundamental Theorem for Continuous Processes". *Mathematical Finance* (2015). DOI: 10.1111/mafi.12110.

[12] B. Bouchard and N. Touzi. "Weak Dynamic Programming Principle for Viscosity Solutions". *SIAM Journal on Control and Optimization* 49.3 (2011), pp. 948–962. DOI: 10.1137/090752328.

[13] B. Bouchard and M. Nutz. "Stochastic Target Games and Dynamic Programming via Regularized Viscosity Solutions". *Mathematics of Operations Research* 41.1 (2015), pp. 109–124. DOI: 10.1287/moor.2015.0718.

[14] H. Brown, D. Hobson, and L. C. G. Rogers. "The maximum maximum of a martingale constrained by an intermediate law". *Probability Theory and Related Fields* 119.4 (2001), pp. 558–578. DOI: 10.1007/PL00008771.

[15] X. Chen, G. Deelstra, J. Dhaene, and M. Vanmaele. "Static super-replicating strategies for a class of exotic options". *Insurance: Mathematics and Economics* 42.3 (2008), pp. 1067–1085. DOI: 10.1016/j.insmatheco.2008.01.002.

[16] A. M. G. Cox, D. Hobson, and J. Obłój. "Pathwise inequalities for local time: Applications to Skorokhod embeddings and optimal stopping". *The Annals of Applied Probability* 18.5 (2008), pp. 1870–1896. DOI: 10.1214/07-AAP507.

[17] A. M. G. Cox and J. Obłój. "Robust Hedging of Double Touch Barrier Options". *SIAM Journal on Financial Mathematics* 2.1 (2011), pp. 141–182. DOI: 10.1137/090777487.

[18] A. M. G. Cox, J. Obłój, and N. Touzi. "The Root solution to the multi-marginal embedding problem: an optimal stopping and time-reversal approach". arXiv: 1505.03169. 2015. URL: http://arxiv.org/abs/1505.03169.

[19] A. M. G. Cox and J. Wang. "Optimal robust bounds for variance options". arXiv: 1308.4363. 2013. URL: http://arxiv.org/abs/1308.4363.

[20] A. M. G. Cox and J. Wang. "Root's barrier: Construction, optimality and applications to variance options". *The Annals of Applied Probability* 23.3 (2013), pp. 859–894. DOI: 10.1214/12-AAP857.

[21]   D. Dawson. "Measure-valued Markov processes". *Ecole d'Eté de Probabilités de Saint-Flour XXI - 1991*. 1541. 1993, pp. 1–260.

[22]   G. Deelstra, I. Diallo, and M. Vanmaele. "Bounds for Asian basket options". *Journal of Computational and Applied Mathematics* 218.2 (2008), pp. 215–228. DOI: 10.1016/j.cam.2006.12.017.

[23]   J. Dhaene, M. Denuit, M. J. Goovaerts, R. Kaas, and D. Vyncke. "The concept of comonotonicity in actuarial science and finance: applications". *Insurance: Mathematics and Economics* 31.2 (2002), pp. 133–161. DOI: 10.1016/S0167-6687(02)00135-X.

[24]   Y. Dolinsky and H. M. Soner. "Martingale optimal transport and robust hedging in continuous time". *Probability Theory and Related Fields* 160.1-2 (2013), pp. 391–427. DOI: 10.1007/s00440-013-0531-y.

[25]   Y. Dolinsky and H. M. Soner. "Martingale optimal transport in the Skorokhod space". *Stochastic Processes and their Applications* 125.10 (2015), pp. 3893–3931. DOI: 10.1016/j.spa.2015.05.009.

[26]   L. E. Dubins. "On a Theorem of Skorohod". *The Annals of Mathematical Statistics* 39.6 (1968), pp. 2094–2097. DOI: 10.2307/2239305.

[27]   N. El Karoui and X. Tan. "Capacities, Measurable Selection and Dynamic Programming Part I: Abstract Framework". arXiv: 1310.3363. 2013. URL: http://arxiv.org/abs/1310.3363.

[28]   N. El Karoui and X. Tan. "Capacities, Measurable Selection and Dynamic Programming Part II: Application in Stochastic Control Problems". arXiv: 1310.3364. 2013. URL: http://arxiv.org/abs/1310.3364.

[29]   R. Eldan. "Skorokhod Embeddings via Stochastic Flows on the Space of Measures". arXiv: 1303.3315. 2013. URL: http://arxiv.org/abs/1303.3315.

[30]   W. H. Fleming and H. M. Soner. *Controlled Markov processes and viscosity solutions*. Second. Vol. 25. Springer, 2006, pp. xviii+429.

[31]   M. Forde and A. Jacquier. "Robust Approximations for Pricing Asian Options and Volatility Swaps Under Stochastic Volatility". *Applied Mathematical Finance* 17.3 (2010), pp. 241–259. DOI: 10.1080/13504860903335348.

[32]   A. Galichon, P. Henry-Labordère, and N. Touzi. "A stochastic control approach to no-arbitrage bounds given marginals, with an application to lookback options". *The Annals of Applied Probability* 24.1 (2014), pp. 312–336. DOI: 10.1214/13-AAP925.

[33]   G. Guo, X. Tan, and N. Touzi. "Tightness and duality of martingale transport on the Skorokhod space". arXiv: 1507.01125. 2015. URL: http://arxiv.org/abs/1507.01125.

[34]   P. Henry-Labordère, J. Obłój, P. Spoida, and N. Touzi. "The maximum maximum of a martingale with given **n** marginals". *The Annals of Applied Probability* 26.1 (2016), pp. 1–44. DOI: 10.1214/14-AAP1084.

[35]   D. Hobson. "The Skorokhod Embedding Problem and Model-Independent Bounds for Option Prices". *Paris-Princeton Lectures on Mathematical Finance 2010*. Vol. 2003. 2011, pp. 267–318.

[36]   D. G. Hobson. "Robust hedging of the lookback option". *Finance and Stochastics* 2.4 (1998), pp. 329–347. DOI: 10.1007/s007800050044.

[37]   D. G. Hobson and J. L. Pedersen. "The Minimum Maximum of a Continuous Martingale with Given Initial and Terminal Laws". *The Annals of Probability* 30.2 (2002), pp. 978–999. DOI: 10.2307/1558824.

[38]   D. Hobson and M. Klimmek. "Maximizing functionals of the maximum in the Skorokhod embedding problem and an application to variance swaps". *The Annals of Applied Probability* 23.5 (2013), pp. 2020–2052. DOI: 10.1214/12-AAP893.

[39] J. Horowitz. "Measure-valued random processes". *Zeitschrift für Wahrschein-lichkeitstheorie und Verwandte Gebiete* 70.2 (1985), pp. 213–236. DOI: 10.1007/BF02451429.

[40] Z. Hou and J. Obłój. "On robust pricing–hedging duality in continuous time". arXiv: 1503.02822. 2015. URL: http://arxiv.org/abs/1503.02822.

[41] A. Jakubowski. "A Non-Skorohod Topology on the Skorohod Space". *Electronic Journal of Probability* 2 (1997). DOI: 10.1214/EJP.v2-18.

[42] S. Källblad, X. Tan, and N. Touzi. "Optimal Skorokhod embedding given full marginals and Azema-Yor peacocks". arXiv: 1503.00500. 2015. URL: http://arxiv.org/abs/1503.00500.

[43] I. Monroe. "On embedding right continuous martingales in Brownian motion". *Ann. Math. Statist.* 43 (1972), pp. 1293–1311.

[44] M. Nutz and R. van Handel. "Constructing sublinear expectations on path space". *Stochastic Processes and their Applications* 123.8 (2013), pp. 3100–3121. DOI: 10.1016/j.spa.2013.03.022.

[45] F. Stebegg. "Model-Independent Pricing of Asian Options via Optimal Martingale Transport". arXiv: 1412.1429. 2014. URL: http://arxiv.org/abs/1412.1429.

[46] G. Žitković. "Dynamic Programming for Controlled Markov Families: Abstractly and over Martingale Measures". *SIAM Journal on Control and Optimization* 52.3 (2014), pp. 1597–1621. DOI: 10.1137/130926481.