

Optimal stopping with nonlinear expectation: geometric and algorithmic solutions

Tomasz Kosmala

*Department of Mathematics, Imperial College London and
School of Mathematical Sciences, Queen Mary University of London*

John Moriarty*

School of Mathematical Sciences, Queen Mary University of London

11 June 2025

Abstract

We use the geometry of suitably generalised potentials to solve risk-sensitive Markovian optimal stopping problems. As in the linear case due to Dynkin and Yushkievich (1967), the value function is the pointwise infimum of those functions which dominate the gain function. An emphasis is placed on geometric and pathwise arguments, rather than exploiting convexity, positive homogeneity or related analytical properties. An algorithm is provided to construct the value function at the computational cost of a two-dimensional search.

Key words: Optimal stopping, nonlinear expectation, risk measures, geometric method, Markov property.

MSC2010 Classification: 60G40, 91B08, 91B06, 60J25.

JEL Classification: C61, D81.

Convexity and associated properties have been used as main tools in the global solution of both stochastic and deterministic optimisation problems. This paper concerns maximising a Markovian performance criterion when stopping a diffusion, where the relationship between concavity and excessivity has been used since the seminal work of Dynkin [20] (see also [16]). However this concavity arises from the linearity of the expectation operator, and there is increasing recent interest in the use of nonlinear, including nonconcave, expectations. In this paper we study the lower envelope (that is, the pointwise infimum) of suitably generalised potentials dominating the gain function (Figure 1), to obtain constructive formulae for the value function and optimal stopping policy in the Markovian context. The formulae are thus analogues to those of Dynkin and Yushkievich [21] for the linear case.

Interest in such nonlinear criteria has arisen in numerous domains. A few examples include the theory of preferences, see for example [22, Ch. 2] and [44], optimisation under ambiguity [8, 9, 13], and even reinforcement learning [24, 31]. The celebrated *prospect theory* of Kahneman and Tversky [26] describes how the same agent can switch between risk averse (concave) and

*Corresponding author. Email: j.moriarty@qmul.ac.uk

risk seeking (convex) criteria, depending on whether losses or gains are considered. This helps to motivate methods based on other properties.

In the absence of convexity or related global analytical properties it is natural to seek local solutions. Accordingly we study the lower envelope of a constrained set of generalised potentials to solve the problem locally. A pathwise argument then confirms this local solution to be optimal.

Our main result is that, under sufficient regularity, the value of the nonlinear optimal stopping problem is the pointwise infimum over those functions in $\mathcal{H}$ which dominate the gain function determining the reward or payoff. It is optimal to stop the first time this value function coincides with the gain function. Thus the value function and optimal stopping region are constructed from the nonconvex geometry of the generalised potentials and gain function. In the linear case, this result was obtained for the Wiener process in [21, Ch. III.7], and extended to one-dimensional regular diffusion processes (with exponential discounting of the reward) in [16] by appropriately generalising the notion of concavity. For κ -ambiguity, which is a positively homogeneous expectation, a result of this type was obtained in [13] using ideas from [6].

A common solution technique in optimal stopping problems is to generate a heuristic solution and then apply a verification argument (see for example [9, Th. 3.4]). Although the assumptions on the nonlinear expectation are strong (see for example [30]), in Section 1.1 we confirm that the solution is correct for the worst-case risk mapping, which is less regular. If used as a heuristic then, as illustrated by Figure 2, the construction differs from the well-known principle of smooth fit or smooth pasting (see for example [37]). As a corollary, no smooth transformation of Figure 2 could reduce the solution to the classical one of [21, Ch. III.7].

To the best of our knowledge the method of proof is different to existing approaches using linearity, generalised concavity, positive homogeneity or related analytical properties, and is novel even in the linear case. This may be of independent interest, particularly as few explicit solutions are known for optimal stopping of Brownian motion in dimension greater than one [14], although the topic is outside the scope of this paper.

A simple algorithmic implementation of the construction returns the value function and optimal stopping region, at the cost of a two-dimensional nonconvex optimisation. The algorithm is presented under regularity assumptions which may be compared to those made in [15], where the authors use analytic methods to provide an algorithmic solution in the linear case.

1 Optimal stopping with nonlinear expectation

The nonlinear optimal stopping problem we consider is

$$V(x) := \sup_{\tau \in \mathcal{T}} \rho(g(X_\tau^x)), \quad (1)$$

where ρ is a nonlinear expectation (more precisely a *risk mapping* with a reference probability measure $\mathbb{P}$, as defined in Section 1.3 below), $g: [0, 1] \rightarrow \mathbb{R}_+$ is a *gain function*, X^x is a sufficiently regular diffusion process which starts at $x \in [0, 1]$ and is absorbed on $\{0, 1\}$, $\mathcal{T}$ is the set of $\mathbb{P}$ -almost surely finite stopping times, and V is the *value function*. Our approach is via the elementary geometry of functions such as $x \mapsto h_{\beta, \gamma}^{y, z}(x)$, where

$$h_{\beta, \gamma}^{y, z}(x) := \begin{cases} \rho(\beta \mathbb{1}_{\{\tau_y^x < \tau_z^x\}} + \gamma \mathbb{1}_{\{\tau_y^x > \tau_z^x\}}), & 0 \leq y < x < z \leq 1, \\ \infty, & \text{otherwise,} \end{cases}$$

and τ_a^x denotes the first hitting time by X^x of the point $a \in [0, 1]$. (Where no confusion arises, we may omit the superscript x and write τ_a .) By analogy with the case of linear expectation,

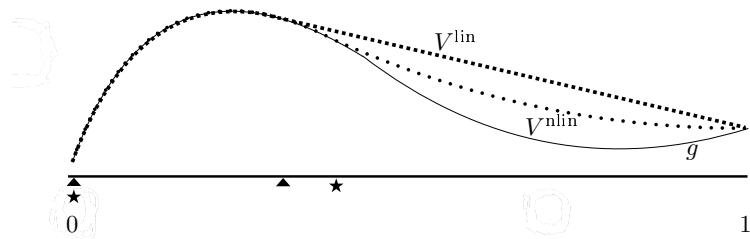


Figure 1: Solutions for an example gain function g (solid curve) when X^x is Brownian motion. The value function V^{lin} in the linear case $\rho = \mathbb{E}$ is the smallest concave majorant of g (square markers). The value function V^{nlm} for an indicative nonlinear ρ is a nonconcave majorant of g (circular markers). The optimal stopping regions are respectively the intervals marked by triangles and stars. In the differentiable case, the value and gain functions meet smoothly (at the rightmost triangles and stars respectively, see Corollary 2.20); for a non-smooth case see Figure 2.

where they are solutions to the Dirichlet problem, we refer to the functions $h_{\beta,\gamma}^{y,z}$ as *generalised potentials*. In Section 2 we obtain the solution

$$V(x) = \inf\{h(x) : h \in \mathcal{H}, h \geq g \text{ on } [0, 1]\}, \quad (2)$$

where the set $\mathcal{H}$ is given by

$$\mathcal{H} = \left\{ h_{\beta,\gamma}^{y,z} : ((y = 0, z = 1) \text{ or } (y \in (0, 1), \beta = \bar{g} \text{ and } z = 1) \text{ or } (y = 0, z \in (0, 1) \text{ and } \gamma = \bar{g})) \right\}, \quad (3)$$

where $\bar{g} = \max_{x \in [0,1]} g(x)$. In the classical case [21, Ch. III.7] (Figure 1, square markers), the expectation is $\rho = \mathbb{E}$ and the functions $h_{\beta,\gamma}^{y,z}$ are linear. In the present setting (Figure 1, circular markers), the functions $h_{\beta,\gamma}^{y,z}$ are nonlinear and defined to be finite only on a subinterval of $[0, 1]$. This choice for $\mathcal{H}$ is motivated by the example in Section 1.1.

Remark 1.1. Definition (3) is used for simplicity. In fact it is sufficient to use a set of functions smaller than $\mathcal{H}$ in the construction, by adding to (3) the constraint that $\beta \vee \gamma := \max\{\beta, \gamma\} \leq \bar{g}$, since the infimum in (2) then remains unchanged. This will be an easy consequence of the monotonicity of ρ (Definition 1.3), the ϱ -martingale property (Lemma 2.4), and the continuity of $\mathcal{H}$ (Lemma 2.6)).

1.1 A motivating example

A key idea in this work is to define the generalised potentials not only on $[0, 1]$, but also on some subdomains of $[0, 1]$ (that is, to allow either $y > 0$ or $z < 1$) in (3). Since this contrasts with the linear case (see eg. (10) below), the following example provides useful motivation. For bounded random variables Z consider the worst-case risk mapping ρ given by

$$\rho(Z) = \mathbb{P} - \text{essinf}(Z).$$

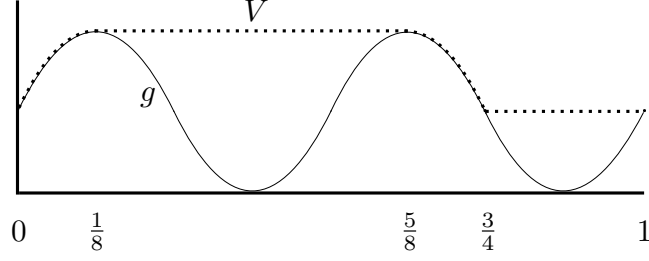


Figure 2: Solution for the gain function $g(x) = 1 + \sin(4\pi x)$ under the worst-case risk mapping. Although this nonlinear expectation has less regularity than assumed below, the value function (square markers) is that of Theorem 2.10, namely $V(x) = \inf\{h(x) : h \in \mathcal{H}, h \geq g \text{ on } [0, 1]\}$. The value and gain functions meet smoothly at $x = \frac{1}{8}$ and $x = \frac{5}{8}$, but not at $x = \frac{3}{4}$. No smooth transformation of this figure could reduce the solution to the classical one illustrated in Figure 1.

With the above setup, the generalised potentials are

$$h_{\beta, \gamma}^{y, z}(x) = \begin{cases} \beta, & x = y, \\ \gamma, & x = z, \\ \beta \wedge \gamma, & x \in (y, z), \\ \infty, & \text{otherwise.} \end{cases} \quad (4)$$

Considering the constraints which define $\mathcal{H}$ (cf. (3)), for example by a suitable sketch, it is easy to see that

$$\inf\{h(x) : h \in \mathcal{H}, h \geq g \text{ on } [0, 1]\} = \min \left\{ \sup_{y \in [0, x]} g(y), \sup_{y \in [x, 1]} g(y) \right\}, \quad (5)$$

see for example Figure 2. We may now verify that (5) is the value function. Taking $x_1 \in \arg \sup_{y \in [0, x]} g(y)$ and $x_2 \in \arg \sup_{y \in [x, 1]} g(y)$, the stopping time $\tau_{x_1, x_2} = \min\{\tau_{x_1}, \tau_{x_2}\}$ gives

$$V(x) \geq \rho \left(g \left(X_{\tau_{x_1, x_2}}^x \right) \right) = \min \left\{ \sup_{y \in [0, x]} g(y), \sup_{y \in [x, 1]} g(y) \right\}.$$

For the reverse inequality take any stopping time $\tau \in \mathcal{T}$ and, writing S for the scale function of X^x (see e.g. [41, V.28]), note that $X_\tau^x \in [0, x]$ with positive probability since $S(X^x)$ is a martingale. It follows that $\mathbb{P} - \text{essinf}(g(X_\tau^x)) \leq \sup_{y \in [0, x]} g(y)$. Since τ was arbitrary, the symmetric argument on $[x, 1]$ gives

$$V(x) \leq \min \left\{ \sup_{y \in [0, x]} g(y), \sup_{y \in [x, 1]} g(y) \right\},$$

and so the solution (2) holds.

1.2 Probabilistic setting

As noted in [21, Ch. III.7], problem (1) would be trivial if X^x were defined on the real line, since it then has probability one of hitting any point. This motivates the choice of X^x as a diffusion in the interval $[0, 1]$ with absorption at its endpoints.

Take the classical Wiener space $\Omega = C(\mathbb{R}_+; \mathbb{R})$ with $B_t(\omega) = \omega(t)$ for $\omega \in \Omega$ and $t \geq 0$, taking $\mathbb{F} = (\mathcal{F}_t)_{t \geq 0}$ to be the smallest filtration satisfying the usual conditions constructed from the natural filtration $\{\sigma(B(s) : s \in [0, t])\}_{t \geq 0}$, and $\mathcal{F} = \sigma(\cup_{t \geq 0} \mathcal{F}_t)$. Let $\mathbb{P}$ be a probability measure on $(\Omega, \mathcal{F})$ under which $B = (B_t)_{t \geq 0}$ is a standard Brownian motion on $\mathbb{R}$ with $B_0 = 0$ almost surely (we will refer to $\mathbb{P}$ as the *reference probability*). Let μ, σ be real measurable functions with σ taking strictly positive values such that, for each $x \in \mathbb{R}$, there exists a solution $Y^x := (Y_s^x(\omega))_{s \geq 0}$ to the stochastic differential equation

$$dY_s^x = \mu(Y_s^x)ds + \sigma(Y_s^x)dB_s, \quad s \geq 0, \quad Y_0^x = x \quad \mathbb{P}\text{-a.s.}, \quad (6)$$

and such that, for each $t > 0$ and almost every $\omega \in \Omega$, the path $Y_{0,t}^x := (Y_s^x(\omega))_{0 \leq s \leq t}$ depends continuously on x in the uniform topology. Set

$$X_t^x := Y_{t \wedge \tau_{0,1}}^x, \quad t \geq 0, \quad (7)$$

where, for $a \in [0, 1]$, we define $\tau_a^x = \inf\{t \geq 0 : Y_t^x = a\}$, and for $0 \leq a < b \leq 1$ set $\tau_{a,b}^x = \inf\{t \geq 0 : Y_t^x \notin (a, b)\}$, and write τ_a and $\tau_{a,b}$ respectively if no confusion arises. That is, the process $X^x = (X_t^x)_{t \geq 0}$ starts at x and is absorbed at the boundary $\{0, 1\}$. For $t \geq 0$ let θ_t be the shift operator $\theta_t : \Omega \rightarrow \Omega$ given by

$$\theta_t(\omega)(s) = \omega(t + s), \quad s \geq 0. \quad (8)$$

Remark 1.2. This continuous dependence of the path $Y_{0,t}^x$ on x in the uniform topology is trivial for the Brownian motion $Y^x := B + x$. For more general diffusions, a simple sufficient condition is that the coefficients are Lipschitz, see e.g. [27, Lem. 21.5 and p. 417].

1.3 Risk mappings

Although the space of all nonlinear performance criteria is large, a number of basic properties are desirable when studying optimisation problems, see for example [9]. We adopt the following commonly used general dynamic framework, which (modulo possible sign changes) has been called a conditional risk measure [22], dynamic risk measure [30] and dynamic conditional risk mapping [29]. Let $b\mathcal{F}$, $b\mathcal{F}_t$ denote respectively the spaces of bounded random variables and bounded $\mathcal{F}_t$ -measurable random variables on $(\Omega, \mathcal{F})$.

Definition 1.3. A *dynamic risk mapping* is a family $\varrho = \{\rho_t : t \geq 0\}$ with $\rho := \rho_0$ of functions $\rho_t : b\mathcal{F} \rightarrow b\mathcal{F}_t$ for $t > 0$ and $\rho : b\mathcal{F} \rightarrow \mathbb{R}$ such that

(i) ρ is a risk mapping, that is it satisfies:

- normalisation: $\rho(0) = 0$,
- monotonicity: for all $Y, Z \in b\mathcal{F}$ we have $Y \leq Z$ $\mathbb{P}$ -a.s. $\implies \rho(Y) \leq \rho(Z)$,
- translation invariance: $Y \in b\mathcal{F}, c \in \mathbb{R} \implies \rho(Y + c) = \rho(Y) + c$,

(ii) for each $t > 0$ ρ_t is a conditional risk mapping, that is it satisfies:

- normalisation: $\rho_t(0) = 0$, $\mathbb{P}$ -a.s.

- *monotonicity*: for all $Y, Z \in b\mathcal{F}$ we have $Y \leq Z$ $\mathbb{P}$ -a.s. $\implies \rho_t(Y) \leq \rho_t(Z)$ $\mathbb{P}$ -a.s.,
- *conditional translation invariance*: $Y \in b\mathcal{F}, Z \in b\mathcal{F}_t \implies \rho_t(Y + Z) = \rho_t(Y) + Z$ $\mathbb{P}$ -a.s..

A consequence is *conditional locality* (see for example [10, Prop. 3.3]): for every Z, Z' in $b\mathcal{F}$ and $A \in \mathcal{F}_t$,

$$\rho_t(\mathbb{1}_A Z + \mathbb{1}_{A^c} Z') = \mathbb{1}_A \rho_t(Z) + \mathbb{1}_{A^c} \rho_t(Z').$$

Definition 1.4. The dynamic risk mapping ϱ is

- *continuous* if for each bounded sequence $\{Y_n\}_{n \geq 1} \subset b\mathcal{F}$ with $Y_n \rightarrow Y$ $\mathbb{P}$ -a.s. we have

$$\rho_t(Y_n) \rightarrow \rho_t(Y) \quad \mathbb{P}\text{-a.s. for all } t \geq 0,$$

- *time consistent* if for all bounded stopping times $\tau_1 \leq \tau_2$ we have

$$\rho_{\tau_1} = \rho_{\tau_1} \circ \rho_{\tau_2}, \tag{9}$$

(cf. [7, Def. 2]),

- *concave* if for all $t \geq 0, \lambda \in [0, 1]$ and $Y, Z \in b\mathcal{F}$ we have

$$\rho_t(\lambda Y + (1 - \lambda)Z) \geq \lambda \rho_t(Y) + (1 - \lambda) \rho_t(Z),$$

(see for example [17, 38]),

- *positively homogeneous* if $\rho_t(\lambda Z) = \lambda \rho_t(Z)$ for all $t \geq 0$ and $\lambda \geq 0$,
- *law invariant* if for each $t \geq 0$ we have $\rho_t(Y) = \rho_t(Z)$ if $Y, Z \in b\mathcal{F}$ have the same distribution under $\mathbb{P}$.

Recalling that for each $\omega \in \Omega$ we have $X^x(\omega) := (X_t^x)_{t \geq 0}(\omega) \in C(\mathbb{R}_+; \mathbb{R})$, we will appeal to the probabilistic Markov property introduced in [29].

Definition 1.5. The dynamic risk mapping ϱ is *Markov* if for all $x \in [0, 1], t > 0$ and $Z = Z(X^x) \in b\mathcal{F}$ we have

$$\rho_t \left(Z \left(X^{X_t^x} \right) \circ \theta_t \right) = \rho \left(Z \left(X^{X_t^x} \right) \right) \quad \mathbb{P}\text{-a.s.},$$

and *strong Markov* if for each $\tau \in \mathcal{T}, x \in [0, 1]$ and $Z \in b\mathcal{F}$ we have

$$\rho_\tau \left(Z \left(X^{X_\tau^x} \right) \circ \theta_\tau \right) = \rho \left(Z \left(X^{X_\tau^x} \right) \right) \quad \mathbb{P}\text{-a.s..}$$

1.4 Related work

Several studies have been made of *robust* optimal stopping, where multiple possible linear expectations are considered. The stopping policy is chosen assuming the worst case, making these robust criteria concave. Related work in the Markovian setting includes [1], and [9] in higher dimension, where solutions are characterised using the Bellman equation. In the non-Markovian framework (also in higher dimension), martingale methods are applied to characterise solutions in terms of Snell envelopes, see for example [4, 5, 25, 33]. Such problems are also closely related to stochastic games of control and stopping, see for example [28].

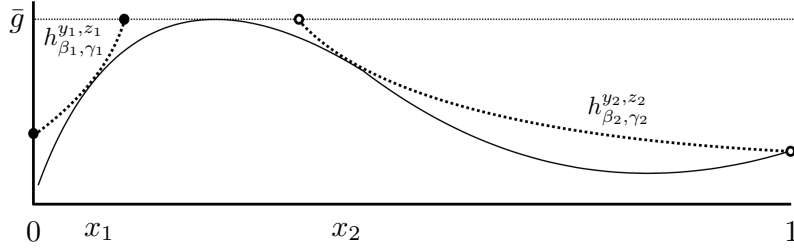


Figure 3: Two generalised potentials $h^{y_i, z_i}_{\beta_i, \gamma_i}$, $i = 1, 2$, satisfying $h^{y_i, z_i}_{\beta_i, \gamma_i} \geq g$. Both lie in $\mathcal{H}$ (cf. Definition 2.1), since their endpoints satisfy (i) $y_1 = 0$, $\gamma_1 = \bar{g}$ (filled circular markers) and (ii) $z_1 = 1$, $\beta_1 = \bar{g}$ (open markers). When evaluated at the point $x = x_1$, $h^{y_1, z_1}_{\beta_1, \gamma_1}$ achieves the infimum in (2), and $h^{y_2, z_2}_{\beta_2, \gamma_2}$ achieves the corresponding infimum at $x = x_2$.

To the best of our knowledge, algorithmic solutions have thus far only been provided in the linear case. In addition to [15] there is a series of studies [11, 23, 42] which embed the problem within an infinite dimensional linear program over a space of measures and examine its dual, yielding a semi-infinite linear program and a nonlinear optimisation problem which facilitate the construction of the value function. Our algorithm has some similarity to the latter approach, in that it begins by seeking local solutions.

The remainder of the paper is organised as follows. Section 2.2 states assumptions and provides preliminary results, Section 2.3 states and proves the main result, Section 2.4 establishes properties of the stopping set and the regularity of the value function, Section 2.5 develops the solution algorithm, and Section 3 provides examples.

2 Geometric and algorithmic solutions

Motivated by the remarkable constructive result of [21], our aim in this work is to identify a family $\mathcal{H}$ of generalised potentials whose geometry encodes the solution to (1) in similar fashion. In the former case, one takes the set $\mathcal{H}^\dagger$ of all affine functions on $[0, 1]$:

$$\mathcal{H}^\dagger := \left\{ h^{0,1}_{\beta, \gamma} : \beta, \gamma \in \mathbb{R} \right\}, \quad (10)$$

and the value function is the pointwise infimum over those $h \in \mathcal{H}^\dagger$ which dominate the gain function. Although the structure of the set $\mathcal{H}^\dagger$ is attractively simple, the counterexamples of Section 3 will demonstrate that its use does not generate the correct solution in general.

In the classical proof of [21, Ch. III.7], the problem is solved globally using the equivalence of excessive and nonnegative concave functions, or the ‘fundamental property of concave functions’ of [21, Ch. III.8]. Our method of proof is instead local. Given $x = X_0$, we identify the first time at which it is optimal to stop. Under this local approach, as noted in [11, 23], it is more natural to consider functions $h^{y, z}_{\beta, \gamma}$ with $0 \leq y \leq x \leq z \leq 1$. The set $\mathcal{H}$ will include some, although not all such functions, as follows.

The argument below requires that at each $v \in \{y, z\}$, either the process is absorbed (i.e. $v \in \{0, 1\}$) or $h(v) \geq \bar{g} = \max_{x \in [0, 1]} g(x)$ (see the proof of Lemma 2.12). This suggests the following construction for $\mathcal{H}$ (which is illustrated in Figure 3):

Definition 2.1. (i) (The set $\mathcal{H}$) Write

$$\mathcal{H} = \left\{ h_{\beta, \gamma}^{y, z} : ((y = 0, z = 1) \text{ or } (y \in (0, 1), \beta = \bar{g} \text{ and } z = 1) \right. \\ \left. \text{or } (y = 0, z \in (0, 1) \text{ and } \gamma = \bar{g})) \right\}$$

where

$$h_{\beta, \gamma}^{y, z}(x) := \begin{cases} \rho(\beta \mathbb{1}_{\{\tau_y^x < \tau_z^x\}} + \gamma \mathbb{1}_{\{\tau_y^x > \tau_z^x\}}), & 0 \leq y < x < z \leq 1, \\ \infty, & \text{otherwise,} \end{cases}$$

(ii) (Domains) Write $d(h_{\beta, \gamma}^{y, z}) := (y, z)$ and $\overline{d(h_{\beta, \gamma}^{y, z})} := [y, z]$,

(iii) (Δ -Equicontinuity) The set $\mathcal{H}$ will be called Δ -equicontinuous if for each $\Delta > 0$, the functions

$$\{h_{\beta, \gamma}^{y, z} : 0 \leq y < z \leq 1, z - y > \Delta, \beta, \gamma \in [0, \bar{g} + 1]\}$$

are equicontinuous.

That is: for each $\Delta > 0$ and each such function $h = h_{\beta, \gamma}^{y, z}$, there exists $\delta > 0$ such that

$$|u - v| \leq \delta \implies |h(u) - h(v)| \leq \varepsilon,$$

(iv) (ϱ -martingales) A process $M = (M_t)_{t \geq 0}$ is a ϱ -martingale if M is bounded and for each $t \geq s$ we have $\rho_s(M_t) = M_s$, $\mathbb{P}$ -a.s..

Remark 2.2. On each compact domain the properties of pointwise equicontinuity and uniform equicontinuity coincide [39, Th. 36, p. 261], and we do not distinguish between the two. Also, we will say ‘ $\mathcal{H}$ is differentiable’ if each $h \in \mathcal{H}$ is differentiable.

For $y < z \in [0, 1]$ and $\beta, \gamma \in [0, \bar{g}]$ define

$$g_{\beta, \gamma}^{y, z}(x) = \beta \mathbb{1}_{\{y\}}(x) + \gamma \mathbb{1}_{\{z\}}(x), \quad x \in [0, 1].$$

2.1 Assumptions

The standing assumption made in this paper is the following:

Assumption 2.3. The dynamic risk mapping ϱ is strongly Markov, time consistent and continuous, $\mathcal{H}$ is Δ -equicontinuous, and g is continuous and nonnegative.

The Δ -equicontinuity of $\mathcal{H}$ is assumed in order to obtain the continuity of the lower envelope

$$w(x) := \inf\{h(x) : h \in \mathcal{H}, h \geq g \text{ on } [0, 1]\}, \quad (11)$$

(see Figure 1). In line with the classical case (cf. (10)), each $h \in \mathcal{H}$ will be continuous (cf. part (i) of Lemma 2.6 below), so the lower envelope w (as the infimum of a family of upper semicontinuous functions) is upper semicontinuous.

Note that by translation invariance, the assumption that g is nonnegative is made without loss of generality. It is also interesting to note that the worst-case risk mapping of Section 1.1 is outside the scope of Assumption 2.3, since $\mathcal{H}$ contains discontinuous functions in that case.

2.2 Preliminary results

Henceforth making Assumption 2.3, we obtain the following results:

Lemma 2.4. Writing $M_t^x = h_{\beta,\gamma}^{y,z}(X_{t \wedge \tau_{y,z}}^x)$, the process $(M_t^x)_{t \geq 0}$ is a ϱ -martingale, where $0 \leq y \leq x \leq z \leq 1$. In particular $h_{\beta,\gamma}^{y,z}(x) = \rho(h_{\beta,\gamma}^{y,z}(X_{\tau_{y,z}}^x))$.

Proof. Setting $g = g_{\beta,\gamma}^{y,z}$ and $h = h_{\beta,\gamma}^{y,z}$, we have by Lemma 2.4 and the strong Markov property for ϱ that, for $s \leq t$,

$$\begin{aligned} \rho_s(M_t^x) &= \rho_s\left(h\left(X_{t \wedge \tau_{y,z}}^x\right)\right) = \rho_s\left(\rho\left(g\left(X_{\tau_{y,z}}^{X_{t \wedge \tau_{y,z}}^x}\right)\right)\right) \\ &= \rho_s\left(\rho_{t \wedge \tau_{y,z}}\left(g\left(X_{\tau_{y,z}}^{X_{t \wedge \tau_{y,z}}^x} \circ \theta_{t \wedge \tau_{y,z}}\right)\right)\right) \\ &= \rho_{s \wedge \tau_{y,z}}\left(g\left(X_{\tau_{y,z}}^{X_{s \wedge \tau_{y,z}}^x} \circ \theta_{s \wedge \tau_{y,z}}\right)\right) = \rho_{s \wedge \tau_{y,z}}\left(g\left(X_{\tau_{y,z}}^{X_{s \wedge \tau_{y,z}}^x} \circ \theta_{s \wedge \tau_{y,z}}\right)\right) \quad (*) \\ &= \rho\left(g\left(X_{\tau_{y,z}}^{X_{s \wedge \tau_{y,z}}^x}\right)\right) = h\left(X_{s \wedge \tau_{y,z}}^x\right) = M_s^x, \end{aligned}$$

where in step (*) we use the following properties: firstly, for all $s \geq 0$ we have

$$X_{\tau_{y,z}}^x = X_{\tau_{y,z}}^{X_{s \wedge \tau_{y,z}}^x} \circ \theta_{s \wedge \tau_{y,z}},$$

and secondly, if τ_1 and τ_2 are two stopping times, then

$$\rho_{\tau_1} \circ \rho_{\tau_2} = \rho_{\tau_1 \wedge \tau_2}.$$

Indeed by conditional translation invariance, conditional locality and time consistency, for each $Z \in b\mathcal{F}$ we have

$$\begin{aligned} \rho_{\tau_1}(\rho_{\tau_2}(Z)) &= \rho_{\tau_1}(\rho_{\tau_2}(Z) \mathbb{1}_{\{\tau_1 \leq \tau_2\}} + \rho_{\tau_2}(Z) \mathbb{1}_{\{\tau_2 < \tau_1\}}) \\ &= \rho_{\tau_1}(\rho_{\tau_1 \vee \tau_2}(Z) \mathbb{1}_{\{\tau_1 \leq \tau_2\}}) + \rho_{\tau_2}(Z) \mathbb{1}_{\{\tau_2 < \tau_1\}} = \rho_{\tau_1 \wedge \tau_2}(Z). \end{aligned}$$

The second claim follows by taking $s = 0$, $t \rightarrow \infty$ and appealing to the a.s. finiteness of $\tau_{y,z}$ and the continuity of ϱ . $\square$

We will use the following version of the optional sampling theorem (a related result in the case of positively homogeneous nonlinear expectation appears in [32, Th. 4.10]):

Proposition 2.5. If M is a càdlàg ϱ -martingale then for all $\tau \in \mathcal{T}$ we have $\rho(M_\tau) = \rho(M_0)$.

Proof. Supposing first that time is discrete, we show that $\rho_n(M_{\tau \wedge (n+1)}) = M_{\tau \wedge n}$. Indeed, since $M_{\tau \wedge n}$ is $\mathcal{F}_n$ -measurable and $\{\tau > n\} \in \mathcal{F}_n$ we have

$$\begin{aligned} \rho_n(M_{\tau \wedge (n+1)}) - M_{\tau \wedge n} &= \rho_n(M_{\tau \wedge (n+1)} - M_{\tau \wedge n}) \\ &= \rho_n((M_{\tau \wedge (n+1)} - M_{\tau \wedge n}) \mathbb{1}_{\{\tau > n\}}) = \mathbb{1}_{\{\tau > n\}} \rho_n(M_{n+1} - M_n) = 0. \end{aligned}$$

It follows by time consistency that

$$\rho(M_{\tau \wedge (n+1)}) = \rho(M_{\tau \wedge n}) = \dots = \rho(M_0).$$

Taking $n \rightarrow \infty$ it follows by the continuity of ρ that $\rho(M_\tau) = \rho(M_0)$.

In continuous time there exists a sequence $(\tau_n)_{n \geq 1}$ of stopping times attaining countably many values and decreasing almost surely to τ [40, Th. 77.1, p. 188]. Taking the limit in $\rho(M_{\tau_n}) = \rho(M_0)$ and using the fact that M is càdlàg and ρ is continuous, we obtain the desired result. $\square$

Lemma 2.6 (Properties of $\mathcal{H}$). Writing $h = h_{\beta,\gamma}^{y,z}$, for $0 \leq y \leq x \leq z \leq 1$ and $\beta, \gamma \in [0, \bar{g}]$ we have:

- (i) (Continuity in x) the function $x \mapsto h_{\beta,\gamma}^{y,z}(x)$ is continuous on $[y, z]$,
- (ii) (Continuity in (β, γ)) the function $(\beta, \gamma) \mapsto h_{\beta,\gamma}^{y,z}(x)$ is continuous,
- (iii) (Maximum principle)
$$\beta \wedge \gamma \leq h(x) \leq \beta \vee \gamma, \quad x \in [y, z],$$
- (iv) (Monotonicity) if $\beta < \gamma$ then h is nondecreasing and if $\beta > \gamma$ then h is nonincreasing,
- (v) (Translation invariance) we have

$$h + \alpha = h_{\alpha+\beta, \alpha+\gamma}^{y,z},$$

- (vi) (Monotone derivatives) suppose that the functions h and $\tilde{h} = h_{\tilde{\beta}, \gamma}^{y,z} \in \mathcal{H}$ both have left derivatives at z . Then if $\tilde{\beta} > \beta$ we have

$$h'^-(z) \geq \tilde{h}'^-(z),$$

where $'^-$ denotes the left derivative,

- (vii) (Bounding) For each pair $y, u \in [0, 1]$ with $y < u$ there exists $\delta > 0$ such that

$$h_{0, \bar{g}+1}^{y,z}(u') > \bar{g}, \quad \forall u' \in [u, 1],$$

where $z = \min\{u' + \delta, 1\}$. Similarly, for each pair $u, z \in (0, 1]$ with $u < z$ there exists $\delta > 0$ such that

$$h_{\bar{g}+1, 0}^{y,z}(u') > \bar{g}, \quad \forall u' \in (0, u],$$

where $y = \max\{u' - \delta, 0\}$.

- (viii) (Scaling) in the case when the underlying diffusion Y^x is standard Brownian motion, taking $h := h_{0,\gamma}^{0,z}$ and $\hat{h} := h_{0,\gamma}^{0,\hat{z}}$ where $\hat{z} \in [z, 1]$, for each $\delta \in (0, z)$ we have

$$h(z - \delta) \leq \hat{h}(\hat{z} - \delta),$$

and thus if h and $\hat{h}$ are differentiable then $h'^-(z) \geq \hat{h}'^-(\hat{z})$. Conversely, if $h := h_{\beta,0}^{0,z}$ and $\hat{h} := h_{\beta,0}^{0,\hat{z}}$ where $\hat{z} \in [z, 1]$, and if h and $\hat{h}$ are differentiable then $h'^-(z) \leq \hat{h}'^-(\hat{z})$.

Proof. For part (i), we will establish continuity at each $x \in (0, 1)$. One-sided continuity at 0 and 1 can then be proved analogously by considering the underlying diffusion Y^x instead of X^x (cf. Section 1.2), since Y is not absorbed at these boundary points. Recalling that we write S for the scale function of X^x , the process $(M_t)_{t \geq 0} := (S(Y_t^x))_{t \geq 0}$ is a continuous local martingale and by the Dubins-Schwarz Theorem [41, Th. 34.1] there exists a Brownian motion W such that $M_t = W_{\langle M \rangle_t}$, where $\langle M \rangle$ denotes the quadratic variation of M .

Recall from Section 1.2 that, for each $t > 0$ and almost every $\omega \in \Omega$, the path $X_{0,t}^x := (X_s^x(\omega))_{0 \leq s \leq t}$ depends continuously on x in the uniform topology. Also, by the Law of the Iterated Logarithm at $t = 0$ (see e.g. [37, p. 97]), for almost all $\omega \in \Omega$ there exists $\delta > 0$ such that for all $t \in [0, \delta]$ we have

$$\min(W_{0,\delta}) < S(x) < \max(W_{0,\delta}) < S(z), \text{ giving } \min(X_{0,\delta'}^x) < x < \max(X_{0,\delta'}^x) < z, \quad (12)$$

where $\langle M \rangle_{\delta'} = \delta$. Thus almost surely, for all sufficiently small $\varepsilon_n > 0$ we have $\mathbb{1}_{\{\tau_x^{x+\varepsilon_n} < \tau_z^{x+\varepsilon_n}\}} = 1$.

Take $\varepsilon_n \downarrow 0$. Then for $0 \leq y \leq x \leq z \leq 1$ and $x + \varepsilon_n \in [y, z]$, we have from the strong Markov property for ϱ that

$$\begin{aligned} \rho_{\tau_{x,z}^{x+\varepsilon_n}} \left(h \left(X_{\tau_{y,z}}^{x+\varepsilon_n} \right) \right) &= \rho_{\tau_{x,z}^{x+\varepsilon_n}} \left(h \left(X_{\tau_{y,z}}^{X_{\tau_{y,z}}^{x+\varepsilon_n}} \right) \circ \theta_{\tau_{x,z}} \right) = \rho \left(h \left(X_{\tau_{y,z}}^{X_{\tau_{y,z}}^{x+\varepsilon_n}} \right) \right) \\ &= \mathbb{1}_{\{\tau_x^{x+\varepsilon_n} < \tau_z^{x+\varepsilon_n}\}} \rho \left(h \left(X_{\tau_{y,z}}^x \right) \right) + \mathbb{1}_{\{\tau_x^{x+\varepsilon_n} > \tau_z^{x+\varepsilon_n}\}} \rho \left(h \left(X_{\tau_{y,z}}^z \right) \right), \end{aligned}$$

and time consistency gives

$$\rho \left(h \left(X_{\tau_{y,z}}^{x+\varepsilon_n} \right) \right) = \rho \left(\mathbb{1}_{\{\tau_x^{x+\varepsilon_n} < \tau_z^{x+\varepsilon_n}\}} \rho \left(h \left(X_{\tau_{y,z}}^x \right) \right) + \mathbb{1}_{\{\tau_x^{x+\varepsilon_n} > \tau_z^{x+\varepsilon_n}\}} \rho \left(h \left(X_{\tau_{y,z}}^z \right) \right) \right).$$

We conclude from the continuity of ρ that

$$h_{\beta,\gamma}^{y,z}(x + \varepsilon_n) = \rho(h(X_{\tau_{y,z}}^{x+\varepsilon_n})) \rightarrow \rho(h(X_{\tau_{y,z}}^x)) = \rho(h(X_{\tau_{y,z}}^x)) = h_{\beta,\gamma}^{y,z}(x),$$

establishing the right continuity of $x \mapsto h_{\beta,\gamma}^{y,z}(x)$ on $[y, z]$, while left continuity is established via symmetric arguments.

For part (ii) take a sequence $(\beta_n, \gamma_n) \rightarrow (\beta, \gamma)$, define the functions $f_n(u) := \beta_n + \frac{u-y}{z-y}(\gamma_n - \beta_n)$, $f(u) := \beta + \frac{u-y}{z-y}(\gamma - \beta)$ and set $Y_n := f_n(X_{\tau_{y,z}}^x)$. Then a.s. we have $Y_n \rightarrow Y := f(X_{\tau_{y,z}}^x)$, so the continuity of ϱ gives

$$h_{\beta_n,\gamma_n}^{y,z}(x) = \rho(Y_n) \rightarrow \rho(Y) = h_{\beta,\gamma}^{y,z}(x).$$

Part (iii) follows from the monotonicity of ϱ .

For part (iv), note that if $\beta < \gamma$ and $y \leq x' \leq x \leq z$ then we have from part (iii) that $\beta \leq h(x)$, so the ϱ -martingale property (Lemma 2.4) and monotonicity of ϱ give

$$h(x') = \rho(h(X_{\tau_{y,x}}^{x'})) \leq h(y) \vee h(x) = \beta \vee h(x) = h(x).$$

Part (v) follows from the translation invariance of ϱ and part (vi) follows from the monotonicity of ϱ , since for $v \in [y, z]$ we have

$$\frac{h(z) - h(v)}{z - v} = \frac{\tilde{h}(z) - h(v)}{z - v} \geq \frac{\tilde{h}(z) - \tilde{h}(v)}{z - v}.$$

For part (vii), suppose that there exist $y, u \in [0, 1)$ with $y < u$ such that

$$M := \sup_{\delta > 0} \inf_{u' \in [u, 1)} h_{0,\bar{g}+1}^{y,\min\{u'+\delta,1\}}(u') \leq \bar{g}. \quad (13)$$

In the definition of Δ -equicontinuity (Definition 2.1), take $\Delta = u - y$, let $\varepsilon \in (0, 1)$, and let $\delta > 0$ be the resulting constant. Then for each $u' \in [u, 1)$ we have by Δ -equicontinuity that

$$h_{0,\bar{g}+1}^{y,\min\{u'+\delta,1\}}(u') \geq h_{0,\bar{g}+1}^{y,\min\{u'+\delta,1\}}(\min\{u' + \delta, 1\}) - \varepsilon = \bar{g} + 1 - \varepsilon,$$

and taking the infimum over $u' \in [u, 1)$ and supremum over $\delta > 0$ gives $M \geq \bar{g} + 1$, contradicting (13). The remaining part is analogous.

For part (viii) we have $\mathbb{1}_{\{\tau_z^{z-\delta} < \tau_0^{z-\delta}\}} = \mathbb{1}_{\{\tau_{\hat{z}}^{z-\delta} < \tau_{\hat{z}-z}^{z-\delta}\}}$ by the spatial homogeneity of the Wiener process (and the fact that $z - \delta, \hat{z} - \delta \in (0, 1)$ and $z, \hat{z}, \hat{z} - z \in [0, 1]$). Then if $h := h_{0,\gamma}^{0,z}$ and $\hat{h} := h_{0,\gamma}^{0,\hat{z}}$,

$$h(z - \delta) = \rho(\gamma \mathbb{1}_{\{\tau_z^{z-\delta} < \tau_0^{z-\delta}\}}) = \rho(\gamma \mathbb{1}_{\{\tau_{\hat{z}}^{z-\delta} < \tau_{\hat{z}-z}^{z-\delta}\}}) \leq \rho(\gamma \mathbb{1}_{\{\tau_{\hat{z}}^{\hat{z}-\delta} < \tau_0^{\hat{z}-\delta}\}}) = \hat{h}(\hat{z} - \delta),$$

where the inequality follows from the monotonicity of ϱ and the fact that $\mathbb{1}_{\{\tau_{\hat{z}}^{z-\delta} < \tau_{\hat{z}-z}^{z-\delta}\}} \leq \mathbb{1}_{\{\tau_{\hat{z}}^{\hat{z}-\delta} < \tau_0^{\hat{z}-\delta}\}}$, $\mathbb{P}$ -a.s.. $\square$

Remark 2.7. In financial and economic applications one may wish to account for the time value of money in problem (1) via exponential discounting, by setting $V^\delta(x) := \sup_{\tau \in \mathcal{T}} \rho(e^{-\delta\tau} g(X_\tau^x))$. It was shown in [16] that under linear expectation, V^δ is the envelope of discounted potentials of the form $\mathbb{E}(e^{-\delta\tau_{y,z}} (\beta \mathbb{1}_{\{\tau_y < \tau_z\}} + \gamma \mathbb{1}_{\{\tau_y > \tau_z\}}))$. Nonetheless the linear expectation has the property of positive homogeneity (cf. Definition 1.4), which would enable discounting to be included in Lemma 2.4. In contrast our nonlinear setting is not positively homogeneous (cf. Section 1.3), hence there is no obvious discounted analogue of Lemma 2.4. More precisely, if we instead define $M_t = e^{-\alpha t \wedge \tau_{y,z}} h_{\beta,\gamma}^{y,z}(X_{t \wedge \tau_{y,z}}^x)$ in Lemma 2.4 then its first line reads

$$\rho_s(M_t) = \rho_s \left(e^{-\alpha t \wedge \tau_{y,z}} h \left(X_{t \wedge \tau_{y,z}}^x \right) \right) = \rho_s \left(e^{-\alpha t \wedge \tau_{y,z}} \rho \left(e^{-\alpha \tau_{y,z}} g \left(X_{\tau_{y,z}}^{X_{t \wedge \tau_{y,z}}^x} \right) \right) \right).$$

Without a suitable multiplicative property, the term $e^{-\alpha t \wedge \tau_{y,z}}$ cannot be brought inside the inner expectation and, thus, the strong Markov property cannot be applied in order to proceed with the proof. We thus leave consideration of discounting as an interesting open question.

Lemma 2.8. If the underlying diffusion Y is standard Brownian motion and if $\mathcal{H}$ is differentiable then $\mathcal{H}$ is Δ -equicontinuous.

Proof. Take $h = h_{\beta,\gamma}^{y,z}$ with $z - y > \Delta$ and $\beta, \gamma \in [0, \bar{g} + 1]$. By the spatial homogeneity of the Wiener process we may suppose without loss of generality that $y = 0$. We will show that there exists a constant $M > 0$ independent of h such that $|h'(u)| < M$ for all $u \in (0, z)$, since this is sufficient to establish the required result. For this, note that for all such u , by the ϱ -martingale property and parts (v)–(viii) of Lemma 2.6, setting $m := \bar{g}$ we have

$$\begin{aligned} h'(u) &= h'^-(u) = (h_{h(0),h(u)}^{0,u})'^-(u) = (m - h(u) + h_{h(0),h(u)}^{0,u})'^-(u) \\ &= (h_{m-h(u)+h(0),m}^{0,u})'^-(u) \leq (h_{0,m}^{0,u})'^-(u) \leq (h_{0,m}^{0,\Delta})'^-(\Delta), \end{aligned}$$

and similarly $h'(u) \geq (h_{m,0}^{0,\Delta})'^-(\Delta)$. Thus $|h'(u)| \leq M := (h_{0,m}^{0,\Delta})'^-(\Delta) \vee (-(h_{m,0}^{0,\Delta})'^-(\Delta))$, a bound which is independent both of h and of the particular choice of u . $\square$

2.3 Representation of the value function

Next we use these properties of $\mathcal{H}$ to establish the structure of the value function (cf. Figure 1).

Lemma 2.9. The lower envelope w of (11) satisfies $w(0) - g(0) = w(1) - g(1) = 0$.

Proof. If $g(0) = \bar{g}$ then trivially $w(0) = g(0) = \bar{g}$. If $g(0) < \bar{g}$ then let $\varepsilon \in (0, \bar{g} - g(0))$ and let $\delta > 0$ be such that $g < g(0) + \varepsilon$ on $[0, \delta]$. Then by the monotonicity of ϱ the function $h = h_{g(0)+\varepsilon,\bar{g}}^{0,\delta} \in \mathcal{H}$ satisfies $h \geq g$, giving $w(0) \leq h(0) = g(0) + \varepsilon$ and completing the claim. $\square$

Theorem 2.10. Under Assumption 2.3 we have $V = w$.

Remark 2.11. The proof of Theorem 2.10 proceeds by first localising the problem to a minimal interval $[x^-, x^+] \ni x$ such that $V = g$ at $x^\pm$ (Proposition 2.13), and obtaining the corresponding solution h° (Lemma 2.14). Proposition 2.16 then uses the strong Markov property for ϱ (Lemma 2.15) to show that the local and global solutions coincide.

Lemma 2.12. For $x \in (0, 1)$ such that $g(x) = w(x)$, we have $w(x) = g(x) = V(x)$.

Proof. Let $\tau \in \mathcal{T}$. For each fixed $\Delta > 0$, by (11) there exists $h = h_{\beta, \gamma}^{y, z} \in \mathcal{H}$ with $h \geq g$ and $w(x) \geq h(x) - \Delta$. In particular, $x \in [y, z]$ by finiteness. By the ϱ -martingale property and optional sampling (Proposition 2.5) we have

$$g(x) = w(x) \geq h(x) - \Delta = \rho(h(X_{\tau \wedge \tau_{y, z}}^x)) - \Delta \geq \rho(g(X_\tau^x)) - \Delta, \quad (14)$$

where the last inequality can be verified case-by-case, as follows. If $\tau < \tau_{y, z}$ we have $h(X_{\tau \wedge \tau_{y, z}}^x) = h(X_\tau^x) \geq g(X_\tau^x)$. If $\tau_{y, z} \leq \tau$ with $y > 0$ and $z = 1$, then in the case $\tau_{y, z} = \tau_y$ we have $h(X_{\tau \wedge \tau_{y, z}}^x) = h(y) = \bar{g} \geq g(X_\tau^x)$ by the construction of $\mathcal{H}$ in Definition 2.1, while in the case $\tau_{y, z} = \tau_z$ we have $h(X_{\tau \wedge \tau_{y, z}}^x) = h(1) \geq g(1) = g(X_\tau^x)$ due to absorption at 1. We conclude from (14) that $g(x) \geq V(x)$. The reverse inequality is obvious and we have $g(x) = w(x) = V(x)$. $\square$

Now, fix $x \in (0, 1)$ such that $w(x) > g(x)$. By Lemma 2.9 we may define

$$x^- = \sup\{u \in [0, x) : w(u) = g(u)\}, \quad (15)$$

$$x^+ = \inf\{u \in (x, 1] : w(u) = g(u)\}, \quad (16)$$

$$\tau^* = \tau_{x^-, x^+} \leq \tau_{0, 1}. \quad (17)$$

Proposition 2.13. $w(x^\pm) = g(x^\pm) = V(x^\pm)$.

Proof. Suppose for a contradiction that $w(x^-) - g(x^-) = \Delta > 0$. Then if $x^- = 0$, the required contradiction follows from Lemma 2.9. If $x^- > 0$ then by construction, for each $\varepsilon > 0$ there exists $u \in (x^- - \varepsilon, x^-]$ and $h^\varepsilon \in \mathcal{H}$ with $w(u) = g(u)$, $h^\varepsilon \geq g$ and $h^\varepsilon(u) \leq w(u) + \frac{\Delta}{3}$, and by the definition of w we have $h^\varepsilon(x^-) \geq w(x^-) = g(x^-) + \Delta$. By continuity of g , for sufficiently small ε we have $w(u) = g(u) < g(x^-) + \frac{\Delta}{3}$ and hence

$$h^\varepsilon(x^-) - h^\varepsilon(u) \geq g(x^-) + \Delta - h^\varepsilon(u) \geq g(x^-) - w(u) + \frac{2\Delta}{3} > \frac{\Delta}{3}. \quad (18)$$

However each h^ε lies in the set $\mathcal{H}$ (Definition (2.1)), so may be written $h^\varepsilon = h_{\beta, \gamma}^{y, z}$ with $[y, z] \ni x$ and either $y = 0$ or $z = 1$. Thus by the Δ -equicontinuity of $\mathcal{H}$, the functions h^ε appearing in (18) are equicontinuous, which is a contradiction. We conclude that $w(x^-) = g(x^-) = V(x^-)$ (cf. Lemma 2.12) and similarly $w(x^+) = g(x^+) = V(x^+)$. $\square$

Lemma 2.14. If $g(x) < w(x)$ then there exists $h^\circ \in \mathcal{H}$ with $h^\circ(x^-) = g(x^-)$, $h^\circ(x^+) = g(x^+)$, and $h^\circ \geq g$.

Proof. We exclude the trivial case $g(x^-) = g(x^+) = \bar{g}$, when h° may simply be taken as the constant function $h_{\bar{g}, \bar{g}}^{0, 1} \in \mathcal{H}$.

The idea of the proof is to extend the function $h_{g(x^-), g(x^+)}^{x^-, x^+}$ to a function $h^\circ \in \mathcal{H}$ in such a way that $h^\circ \geq g$. There will be four intermediate steps:

- (i) Extend the function $h_0 := h_{g(x^-), g(x^+)}^{x^-, x^+}$ to the right, obtaining a function $h_{n'} := h_{g(x^-), \Gamma}^{x^-, Z}$ satisfying either $Z = 1$ or $\Gamma \geq \bar{g}$,
- (ii) extend the function $h_{n'}$ to the left, obtaining a function $\bar{h}_N := h_{B, \Gamma}^{W, Z}$ satisfying either $W = 0$ or $B \geq \bar{g}$,
- (iii) verify that $\bar{h}_N$ dominates g on $[0, x^-] \cup [x^+, 1]$,
- (iv) restrict the function $\bar{h}_N$ to obtain an element $h^\circ = h_{B', \Gamma'}^{W', Z'}$ with $h^\circ \in \mathcal{H}$, and finally verify that $h^\circ \geq g$.

Step (i). We first extend $h_0 := h_{g(x^-), g(x^+)}^{x^-, x^+}$ to the right. If $g(x^+) = \bar{g}$ then set $\Gamma := \bar{g}$ and $Z = x^+$, and step (i) is complete. Otherwise we have $g(x^+) < \bar{g}$. Note from Proposition 2.13 that $g(x) < w(x)$ implies $x^- < x^+$. Then take $y = x^-$, $u = x^+$ in the bounding property (part (vii) of Lemma 2.6) to obtain the corresponding δ . Set $z_0 := x^+$ and $z_1 = \min\{x^+ + \delta, 1\}$.

Suppose for a contradiction that $\tilde{h}(x^+) > g(x^+)$, where $\tilde{h} := h_{g(x^-), 0}^{x^-, z_1}$. Then since $w(x^+) = g(x^+) < \tilde{h}(x^+)$, by the definition of w there exists $h \in \mathcal{H}$ with

$$h \geq g, \quad h(x^+) < \tilde{h}(x^+), \quad h(x^-) \geq g(x^-) = \tilde{h}(x^-),$$

so that by the continuity of $\mathcal{H}$ we have $h(u) = \tilde{h}(u)$ for some $u \in [x^-, x^+]$. Next note that $h(z_1) \geq 0 = \tilde{h}(z_1)$, so by the continuity of $\mathcal{H}$ we have $h(v) = \tilde{h}(v)$ for some $v \in (x^+, z_1]$. Then since h and $\tilde{h}$ coincide at $u \leq x^+$ and at $v > x^+$, the ϱ -martingale property gives $h(x^+) = \tilde{h}(x^+)$, a contradiction. We conclude that $\tilde{h}(x^+) = h_{g(x^-), 0}^{x^-, z_1}(x^+) \leq g(x^+)$.

Then recalling the constant δ obtained above from the bounding property with $y = x^-$ and $u = x^+$, we set $u' = x^+$ to obtain $h_{0, \bar{g}+1}^{x^-, z_1}(x^+) > \bar{g}$. Further, the monotonicity of ϱ gives $h_{g(x^-), \bar{g}+1}^{x^-, z_1}(x^+) > \bar{g}$. Then, from ‘continuity in (β, γ) ’ (part (ii) of Lemma 2.6) and the intermediate value theorem, there exists $\gamma \in [0, \bar{g} + 1)$ such that $h_{g(x^-), \gamma}^{x^-, z_1}(x^+) = g(x^+)$. Set

$$\begin{aligned} \gamma_1 &:= \sup\{\gamma \geq 0 : h_{g(x^-), \gamma}^{x^-, z_1}(x^+) = g(x^+)\}, \\ h_1 &:= h_{g(x^-), \gamma_1}^{x^-, z_1}, \end{aligned}$$

noting by continuity that $h_1(x^+) = g(x^+)$. If either $z_1 = 1$ or $\gamma_1 \geq \bar{g}$ then set $\Gamma := \gamma_1$ and $Z = z_1$, and step (i) is complete. Otherwise, if $z_1 < 1$ and $\gamma_1 < \bar{g}$ then set $z_2 = \min\{x^+ + 2\delta, 1\}$. As before we have $h_{g(x^-), 0}^{x^-, z_2}(x^+) \leq g(x^+)$. The definition of γ_1 , ϱ -martingale property and monotonicity of ϱ then give $h_{g(x^-), 0}^{x^-, z_2}(z_1) \leq \gamma_1$. Also taking $u' = z_1$ in the bounding property gives $h_{0, \bar{g}+1}^{x^-, z_2}(z_1) > \bar{g}$ and the monotonicity of ϱ gives $h_{g(x^-), \bar{g}+1}^{x^-, z_2}(z_1) > \bar{g}$ and, as before, we may write

$$\begin{aligned} \gamma_2 &:= \sup\{\gamma \geq 0 : h_{g(x^-), \gamma}^{x^-, z_2}(z_1) = \gamma_1\}, \\ h_2 &:= h_{g(x^-), \gamma_2}^{x^-, z_2}. \end{aligned}$$

Proceeding similarly, for some $n' \in \mathbb{N}$ we have either $z_{n'} = 1$ or $\gamma_{n'} \geq \bar{g}$, and set $\Gamma = \gamma_{n'}$ and $Z = z_{n'}$, and step (i) is complete.

Step (ii). In Step (i) the function $h_0 := h_{g(x^-), g(x^+)}^{x^-, x^+}$ was extended to the right, obtaining a function $h_{n'} := h_{g(x^-), \Gamma}^{x^-, Z}$ satisfying either $Z = 1$ or $\Gamma \geq \bar{g}$. Indeed it can be verified by construction and the ϱ -martingale property that for each $n \geq 1$ we have $h_n = h_{n-1}$ on $[x^-, z_{n-1}]$, giving

$$h_n(v) = h_0(v) = \rho(g(X_{\tau^*}^v)), \quad \forall v \in [x^-, x^+], \quad (19)$$

where τ^* was defined in (17) as τ_{x^-, x^+} .

Next we extend $h_{n'} := h_{g(x^-), \Gamma}^{x^-, Z}$ to the left. Setting $w_1 = \max\{0, x^- - \delta\}$ (for a suitable $\delta > 0$) we may define

$$\beta_1 = \sup\{\beta \geq 0 : h_{\beta, g(x^+)}^{w_1, x^+}(x^-) = g(x^-)\}, \quad (20)$$

$$\bar{h}_1 = h_{\beta_1, \Gamma}^{w_1, Z}.$$

By an iterative extension analogous to the previous one, we obtain $B \geq 0$ and $W \in [0, x^-]$ such that either $B \geq \bar{g}$ or $W = 0$. Let N be the index at which the iteration stops. As in (19) we have $\bar{h}_N = \bar{h}_1 = h_0$ on $[x^-, x^+]$.

Step (iii). In Step (ii) the function $h_{n'}$ was extended to the left, obtaining a function $\bar{h}_N := h_{B, \Gamma}^{W, Z}$ satisfying either $W = 0$ or $B \geq \bar{g}$. Next we verify that $\bar{h}_N \geq g$ on $[0, x^-]$. Suppose for a contradiction that $\bar{h}_N(w) < g(w)$ for some $w \in [0, x^-]$. Since then $\bar{h}_N(w) < \infty$, we have $w_N \leq w$ and there exists $\tilde{n}$ with $1 \leq \tilde{n} \leq N$ such that $w \in [w_{\tilde{n}}, w_{\tilde{n}-1}]$. Defining the function $h_w := h_{g(w), g(x^+)}^{w, x^+}$, we have $h_w(w) = g(w) \leq \bar{g}$ and so the above procedure may be applied (with the same value for δ) to extend the function h_w to the left, yielding $h_w^* := h_{\gamma', g(x^+)}^{w', x^+}$ where $w' = \max\{w - \delta, 0\} \leq w_{\tilde{n}}$. By the definition of V and of h_w , and recalling Proposition 2.13, suboptimality gives $h_w(x^-) \leq V(x^-) = g(x^-)$. On the other hand, the ϱ -martingale property and monotonicity of ϱ give $h_w(x^-) \geq \bar{h}_N(x^-) = g(x^-)$, so $h_w(x^-) = g(x^-)$. The suprema taken during the extension process (cf. (20)) therefore give $h_w^*(w_k) \leq \bar{h}_N(w_k)$ for $k = 1, 2, \dots, \tilde{n}$. The ϱ -martingale property then gives $g(w) = h_w^*(w) \leq \rho \left(\bar{h}_N \left(X_{\tau_{w_{\tilde{n}-1}, w_{\tilde{n}}}}^w \right) \right) = \bar{h}_N(w)$, which is the required contradiction. An analogous argument establishes that $\bar{h}_N \geq g$ on $[x^+, 1]$.

Step (iv). In Step (iii) it was verified that the function $\bar{h}_N = h_{B, \Gamma}^{W, Z}$ dominates g on $[0, x^-] \cup [x^+, 1]$. Since $\bar{h}_N$ does not necessarily satisfy the definition (3) of an element of $\mathcal{H}$, in this step we first identify a ‘restriction’ h° of the function $\bar{h}_N$ with $h^\circ \in \mathcal{H}$. This is achieved by modifying the constants W, Z, B and Γ , as follows. If $\Gamma \geq \bar{g}$, by continuity in x (Part (i) of Lemma 2.6) there exists $Z' \in (z_{n'-1}, Z]$ with $h_{n'}(Z') = \bar{g}$, and we then define $\Gamma' = \bar{g}$. Otherwise, if $\Gamma < \bar{g}$ then by the above procedure we must have $Z = 1$; set $Z' = Z = 1$ and $\Gamma' = \Gamma$. Similarly, if $B \geq \bar{g}$ there exists $W' \in [W, w_{N-1})$ with $\bar{h}_N(W') = \bar{g}$, and we then define $B' = \bar{g}$; otherwise, if $B < \bar{g}$ then $W = 0$ and we set $W' = W = 0$ and $B' = B$.

Writing $h^\circ := h_{B', \Gamma'}^{W', Z'}$, we observe that either $W = 0$ or $Z = 1$. (This follows since if both $W' > 0$ and $Z' < 1$ then $B = \Gamma = \bar{g}$, and the monotonicity of ϱ then forces $g(x^-) = g(x^+) = \bar{g}$, which was excluded at the beginning of the proof.) It is now immediate from Definition 2.1 that the function $h^\circ := h_{B', \Gamma'}^{W', Z'}$ is an element of $\mathcal{H}$. It is also immediate from Definition 2.1 that

$$h^\circ(x) = \begin{cases} \bar{h}_N(x), & x \in [W', Z'], \\ \infty, & x \in [0, W') \cup (Z', 1], \end{cases}$$

so the above results give $h^\circ \geq \bar{h}_N \geq g$ on $[0, x^-] \cup [x^+, 1]$.

To complete the proof it remains only to establish that $h^\circ \geq g$ on $[x^-, x^+]$. For this, let

$$\varepsilon = \sup\{g(u) - h^\circ(u) : u \in [x^-, x^+]\}, \quad (21)$$

and suppose for a contradiction that $\varepsilon > 0$. Then the function $h^\circ + \varepsilon$ dominates g on $[x^-, x^+]$ and thus on $[0, 1]$. Recalling the translation invariance property of $\mathcal{H}$ (part (v) of Lemma 2.6), the function $h^\circ + \varepsilon = h_{B'+\varepsilon, \Gamma'+\varepsilon}^{W', Z'}$ can be ‘restricted’ to obtain an element $h_\varepsilon^\circ = h_{B'', \Gamma''}^{W'', Z''} \in \mathcal{H}$ which also dominates g on $[0, 1]$. Then by continuity the supremum in (21) is attained on $[x^-, x^+]$ and by the construction of h_ε° it must be attained at some $u \in (x^-, x^+)$. Then we have

$$g(u) \leq w(u) \leq h_\varepsilon^\circ(u) = h^\circ(u) + \varepsilon = g(u),$$

thus $g(u) = w(u)$, contradicting (15)–(16). We conclude that $h^\circ \geq g$ on $[x^-, x^+]$ and thus on $[0, 1]$. $\square$

Lemma 2.15. Let $\tau_1, \tau_2 \in \mathcal{T}$ with $0 \leq \tau_1 \leq \tau_2$ a.s.. Then there exists a (random) stopping time $\tau_3 \in \mathcal{T}$ such that almost surely we have

$$\rho_{\tau_1}(g(X_{\tau_2}^x)) = \begin{cases} \rho\left(g\left(X_{\tau_3}^{X_{\tau_1}^x}\right)\right), & \{\tau_1 < \tau_2\}, \\ g(X_{\tau_1}^x), & \{\tau_1 = \tau_2\}. \end{cases}$$

Proof. We divide the proof into two steps.

Step (i). Fix $\omega \in \Omega$ and consider the map $G^\omega: \Omega \rightarrow \Omega$ given by $\bar{\omega} \mapsto G^\omega(\bar{\omega})$, which constructs a sample path $G^\omega(\bar{\omega})$ by (a) following ω from time 0 until time $\tau_1(\omega)$, then (b) following the increments of $\bar{\omega}$ from time 0 onwards. That is,

$$G^\omega(\bar{\omega})(t) = \begin{cases} \omega(t), & t \leq \tau_1(\omega), \\ \omega(\tau_1(\omega)) + \bar{\omega}(t - \tau_1(\omega)) - \bar{\omega}(0), & t > \tau_1(\omega). \end{cases}$$

For any stopping time $\tau \in \mathcal{T}$ consider the random time $\tau^\omega(\bar{\omega})$ obtained by applying the function τ to this sample path:

$$\bar{\omega} \mapsto \tau^\omega(\bar{\omega}) := \tau(\bar{\omega}), \quad \text{where} \quad \bar{\omega} := G^\omega(\bar{\omega}).$$

Then since $G^\omega(\theta_{\tau_1(\omega)}(\omega)) = \omega$, we have $\tau^\omega(\theta_{\tau_1(\omega)}(\omega)) = \tau(\omega)$ and so $\tau_i^\omega \circ \theta_{\tau_1(\omega)}(\omega) = \tau_i(\omega)$ for $i = 1, 2$. Thus

$$\tau_2(\omega) = \tau_1(\omega) + (\tau_2^\omega - \tau_1^\omega) \circ \theta_{\tau_1(\omega)}(\omega),$$

and recalling that $\tau_1 \leq \tau_2$ almost surely we have

$$g\left(X_{\tau_2(\omega)}^x\right) = g\left(X_{\tau_2^\omega - \tau_1^\omega}^{X_{\tau_1(\omega)}^x}\right) \circ \theta_{\tau_1(\omega)}(\omega). \quad (22)$$

Define an auxiliary bounded random variable $Z^{\tau_1(\omega)}$ by

$$\begin{aligned} Z^{\tau_1(\omega)}: \Omega &\rightarrow \mathbb{R}, \\ Z^{\tau_1(\omega)}(\bar{\omega}) &= g\left(X_{\tau_2^\omega - \tau_1^\omega}^{X_{\tau_1(\omega)}^x}(\bar{\omega})\right). \end{aligned} \quad (23)$$

Now suppose that $\bar{\omega}$ is an independent random draw from $(\Omega, \mathcal{F}, \mathbb{P})$. Putting the above together we have that almost surely

$$\rho_{\tau_1(\omega)}(g(X_{\tau_2(\omega)}^x)) = \rho_{\tau_1(\omega)}(Z^{\tau_1(\omega)} \circ \theta_{\tau_1(\omega)}(\omega)) = \rho_{\tau_1(\bar{\omega})}(Z^{\tau_1(\omega)} \circ \theta_{\tau_1(\bar{\omega})}(\bar{\omega})), \quad (24)$$

where the first equality follows from (22)–(23). For the second equality, note the identity $\tau_1(\omega) = \tau_1(\bar{\omega})$, and also that the Brownian increments $(\omega(\tau_1(\omega) + t) - \omega(\tau_1(\omega)))_{t \geq 0}$ are independent of $\mathcal{F}_{\tau_1(\omega)}$ by the strong Markov property. Then since $\bar{\omega}$ also has Brownian increments independent of $\mathcal{F}_{\tau_1(\omega)}$, we may replace $\theta_{\tau_1(\omega)}(\omega)$ by $\theta_{\tau_1(\bar{\omega})}(\bar{\omega})$ in (24).

Step (ii). Next we identify $\tau_3^\omega := \tau_2^\omega - \tau_1^\omega$ with an $\mathbb{F}$ -stopping time. For this, note that the filtration $\mathbb{F}$ can be identified with the smallest filtration satisfying the usual conditions constructed from the natural filtration of $(\bar{\omega}(t))_{t \geq 0}$. Thus writing $\tilde{\mathbb{F}} = (\tilde{\mathcal{F}}_t)_{t \geq 0}$ for the latter, the $\mathbb{F}$ -stopping time τ_2 can be identified with an $\tilde{\mathbb{F}}$ -stopping time. Then since $\tau_1(\bar{\omega}) = \tau_1(\omega)$, for each $\bar{\omega} \in \Omega$ we have

$$\{(\tau_2^\omega - \tau_1^\omega)(\bar{\omega}) \leq t\} = \{\tau_2(\bar{\omega}) \leq \tau_1(\omega) + t\} \in \tilde{\mathcal{F}}_{\tau_1(\omega) + t} =: \mathcal{F}_t^{\tau_1(\omega)},$$

so that τ_3^ω is measurable in the filtration $\mathbb{F}^{\tau_1(\omega)} := (\mathcal{F}_t^{\tau_1(\omega)})_{t \geq 0}$.

Next, fix ω and note that the filtration $\mathbb{F}^{\tau_1(\omega)}$ is then simply that generated by the increments of the process $(\tilde{\omega}(t))_{t \geq 0}$. In particular $\mathbb{F}^{\tau_1(\omega)}$ is coarser than $\mathbb{F}$, so any $\mathbb{F}^{\tau_1(\omega)}$ -stopping time can be identified with an $\mathbb{F}$ -stopping time. By the strong Markov property for Brownian motion we have $\tau_2^\omega < \infty$ a.s., giving $0 \leq \tau_3^\omega < \infty$ a.s.. We conclude that τ_3^ω can be identified with an $\mathbb{F}$ -stopping time.

Combining (23) and (24) and then applying the strong Markov property of ρ gives almost surely

$$\begin{aligned} \rho_{\tau_1(\omega)}(g(X_{\tau_2(\omega)}^x)) &= \rho_{\tau_1(\tilde{\omega})} \left(g \left(X_{\tau_3^\omega}^{X_{\tau_1(\tilde{\omega})}^x} \right) \circ \theta_{\tau_1(\tilde{\omega})}(\tilde{\omega}) \right) \\ &= \rho \left(g \left(X_{\tau_3^\omega}^{X_{\tau_1(\tilde{\omega})}^x} \right) \right) = \rho \left(g \left(X_{\tau_3^\omega}^{X_{\tau_1(\omega)}^x} \right) \right), \end{aligned}$$

completing the proof. $\square$

Proposition 2.16. For $x \in [0, 1]$ and τ^* defined in (17) we have

$$V(x) = \check{V}(x) := \sup_{\tau \in \mathcal{T}, \tau \leq \tau^*} \rho(g(X_\tau^x)).$$

Proof. Let $\tau \in \mathcal{T}$. Setting $\tau_1 = \tau \wedge \tau^*$ and $\tau_2 = \tau$ in Lemma 2.15 gives that for some stopping time τ_3 ,

$$\begin{aligned} \rho(g(X_\tau^x)) &= \rho(\rho_{\tau \wedge \tau^*}(g(X_\tau^x))) = \rho \left(\mathbb{1}_{\{\tau < \tau^*\}} g(X_\tau^x) + \mathbb{1}_{\{\tau \geq \tau^*\}} \rho \left(g \left(X_{\tau_3}^{X_{\tau^*}^x} \right) \right) \right) \\ &\leq \rho(\mathbb{1}_{\{\tau < \tau^*\}} g(X_\tau^x) + \mathbb{1}_{\{\tau \geq \tau^*\}} g(X_{\tau^*}^x)) = \rho(g(X_{\tau \wedge \tau^*}^x)), \end{aligned}$$

where the first equality follows from time consistency and the inequality follows from the fact that $g(X_{\tau^*}^x) = V(X_{\tau^*}^x)$ and optimality. This gives $V \leq \check{V}$ and hence $V = \check{V}$. $\square$

Thus to evaluate $V(x)$ it suffices to study $\check{V}$ on $[x^-, x^+]$. For this, the above analysis may be repeated with the interval $[0, 1]$ replaced by $[x^-, x^+]$. Denote the resulting analogues of all objects with a check mark so that, for example,

$$\check{w}(x) = \inf \{ \check{h}(x) : \check{h} \in \check{\mathcal{H}}, \check{h} \geq g \text{ on } [x^-, x^+] \}.$$

Proof of Theorem 2.10. From Lemma 2.14 we have $h^\circ(x^-) = g(x^-)$ and $h^\circ(x^+) = g(x^+)$, so the optional sampling theorem (Proposition 2.5) gives

$$h^\circ(x) = \rho(h^\circ(X_{\tau^*}^x)) = \rho(g(X_{\tau^*}^x)).$$

Then since $g \leq h^\circ \in \mathcal{H}$, we have

$$w(x) \leq h^\circ(x) = \rho(g(X_{\tau^*}^x)) \leq \check{V}(x).$$

Conversely, if $\tau \in \mathcal{T}$ and $\tau \leq \tau^*$ then $h^\circ(x) = \rho(h^\circ(X_\tau^x)) \geq \rho(g(X_\tau^x))$, giving $h^\circ(x) = \check{V}(x) = V(x)$ by Proposition 2.16. To complete the proof note that for any $\check{h} = h_{\beta, \gamma}^{y, z} \in \check{\mathcal{H}}$ with $x \in [y, z]$ and $\check{h} \geq g$ we have

$$\check{h}(x) = \rho(\check{h}(X_{\tau^* \wedge \tau_{y, z}}^x)) \geq \rho(g(X_{\tau^*}^x)) = h^\circ(x),$$

(cf. the last inequality in (14)) so that $w(x) \geq h^\circ(x)$ and hence $w(x) = h^\circ(x) = V(x)$. $\square$

2.4 Stopping set and regularity

Since the proof of Theorem 2.10 concludes with $h^\circ(x) = V(x)$, we have

Corollary 2.17. The stopping policy

$$\tau^* = \tau_{x^-, x^+} = \inf\{t \geq 0 : g(X_t^x) = V(X_t^x)\}$$

is optimal in (1).

In this section we characterise the *stopping set*, the set of points $S = \{x \in [0, 1] : V(x) = g(x)\}$ at which stopping is optimal, or equivalently the *continuation set* $C = [0, 1] \setminus S$. We establish the continuity of the value function V (Proposition 2.18) and use the characterisation of S (Lemma 2.19) to show in Corollary 2.20 that differentiability of g and $\mathcal{H}$ are sufficient for the differentiability of V .

Proposition 2.18. V is continuous.

Proof. The upper semicontinuity of V was noted at the beginning of Section 2.2. If $V(x) = g(x)$ then the lower semicontinuity of V follows from the continuity of g and the fact that $V \geq g$. If $V(x) > g(x)$ then since $g(x^\pm) = V(x^\pm)$ (Proposition 2.13) we have $x^- < x < x^+$. Since $V = h^\circ$ on $[x^-, x^+]$ it follows that V is continuous on (x^-, x^+) . $\square$

Lemma 2.19. Suppose that g and $\mathcal{H}$ are differentiable. Then x lies in the stopping set if and only if at least one of the following holds:

(i) There exists $h \in \mathcal{H}$ with:

$$\begin{cases} h(0) < \infty \\ h(x) = g(x) \\ h'(x) = g'(x) \quad \text{if } x > 0 \\ h \geq g \text{ on } [0, 1], \end{cases} \quad (25)$$

(ii) There exists $h \in \mathcal{H}$ with:

$$\begin{cases} h(1) < \infty \\ h(x) = g(x) \\ h'(x) = g'(x) \quad \text{if } x < 1 \\ h \geq g \text{ on } [0, 1]. \end{cases} \quad (26)$$

Proof. Condition (i) is sufficient to imply $x \in S$, as is condition (ii), so we assume that $x \in S$ and seek to establish the necessity of either condition (i) or (ii).

We first construct $h_* \in \mathcal{H}$ with $h_* \geq g$ and $h_*(x) = g(x)$. Let $(\varepsilon_n)_{n \in \mathbb{N}} \downarrow 0$ with $\varepsilon_0 < \frac{1}{2}$. Since $x \in S$ implies that $w(x) = g(x)$, for each $n \in \mathbb{N}$ there exists $h_n \in \mathcal{H}$ with $h_n \geq g$ and $h_n(x) < g(x) + \varepsilon_n$. By the construction of $\mathcal{H}$ (cf. Definition 2.1) and the monotonicity of ϱ , either $h_n(0) \leq g(x) + \varepsilon_n$ infinitely often or $h_n(1) \leq g(x) + \varepsilon_n$ infinitely often. Without loss of generality, by the Bolzano-Weierstrass theorem (and choosing an appropriate subsequence in $\mathbb{N}$ if necessary) we may suppose that the sequence $(h_n(0))_n$ converges, so that $\beta := \lim_n h_n(0) \leq g(x)$. If $x = 1$ then, by ‘continuity in (β, γ) ’, the functions $h_{h_n(0), h_n(1)}^{0,1}$ converge pointwise to $h_* := h_{\beta, g(1)}^{0,1} \in \mathcal{H}$ with $h_* \geq g$ and $h_*(1) = g(1)$.

Suppose therefore that $x < 1$ and, with the convention $\inf \emptyset = \infty$, set

$$z_n := \inf\{z \in (x, 1] : h_n(z) > \bar{g} + 1\} \wedge 1.$$

By again choosing a suitable subsequence we may assume that $(z_n)_n$ converges to some $z' \leq 1$. Noting that $h_n(x) < \bar{g} + \frac{1}{2}$, we may appeal to the Δ -equicontinuity of $\mathcal{H}$ as in the proof of Proposition 2.13 to conclude that $z' > x$. Appealing again to Δ -equicontinuity, there exists $z \in (x, z')$ such that $h_n(z) > \bar{g} + \frac{1}{2}$ for all n . Then by ‘continuity in (β, γ) ’, the functions $h_{h_n(0), h_n(z)}^{0, z}$ converge along some subsequence of $\mathbb{N}$ to $h_* := h_{\beta, \gamma}^{0, z}$ for some $\gamma \in [\bar{g} + \frac{1}{2}, \bar{g} + 1]$ with $h_*(x) = g(x)$. If $z < 1$ then h_* may be extended to an element of $\mathcal{H}$ (cf. the proof of Lemma 2.14). Assume that this extension has been performed. Then h_* lies in $\mathcal{H}$ and monotonicity gives $h_* \geq g$ on $[0, 1]$. If $x \in (0, 1)$, this tangency implies that $h'_*(x) = g'(x)$, so that condition (i) of the lemma is satisfied. An analogous proof shows that if $\sup_n h_n(1) \leq g(x)$ then condition (ii) of the lemma is satisfied. $\square$

Corollary 2.20 (Smooth fit principle). If g and $\mathcal{H}$ are differentiable then V is differentiable on $(0, 1)$. In particular $V'(x) = g'(x)$ for $x \in S \cap (0, 1)$.

Proof. For each $x \in S \cap (0, 1)$, taking h as in Lemma 2.19 with $h(x) = g(x)$, $h'(x) = g'(x)$ and $h \geq g$ we have $h(x) = g(x) = V(x)$, so Theorem 2.10 gives $g(x) \leq V(x) = w(x) \leq h(x)$ and

$$\begin{aligned} g'(x) &= \lim_{\varepsilon \rightarrow 0} \frac{g(x + \varepsilon) - g(x)}{\varepsilon} \leq \lim_{\varepsilon \rightarrow 0} \frac{V(x + \varepsilon) - V(x)}{\varepsilon} \\ &\leq \lim_{\varepsilon \rightarrow 0} \frac{h(x + \varepsilon) - h(x)}{\varepsilon} = h'(x) = g'(x), \end{aligned}$$

so that $V'(x) = g'(x)$.

For each connected component C_* of C and each $x_* \in C_*$, let x_*^-, x_*^+ be defined as in (15)–(16). Then $g(x_*^\pm) = V(x_*^\pm)$ (Lemma 2.12) and for any $x_*^- < x < x_*^+$ we have $x^\pm = x_*^\pm$. Then for h° defined as in Lemma 2.14, we have $V(x) = h^\circ(x) = h_{g(x^-), g(x^+)}^{x^-, x^+}(x) = h_{g(x_*^-), g(x_*^+)}^{x_*^-, x_*^+}(x) =: h_*(x)$. Thus $V = h_*^\circ$ on C_* , hence the differentiability of V on C_* follows immediately from that of $\mathcal{H}$ (Assumption 2.21). $\square$

2.5 Solution algorithm

The following algorithm aims to construct the value function globally by returning the set of continuation regions - that is, the connected components of the continuation region C . It exploits the fact, due to the smooth fit principle (Corollary 2.20) and the proof of Theorem 2.10, that for each $x \in C$ the value function is tangent to the gain function at both x^- and x^+ . Then since $V = h_{g(x^-), g(x^+)}^{x^-, x^+}$ on $[x^-, x^+]$, this double tangency can be used to search for x^- and x^+ . The computational cost is thus on the order of a search over $[0, 1]^2$. In order to terminate in finite time with the correct result, we make the following additional assumptions:

- Assumption 2.21.** (i) $\delta > 0$ is strictly smaller than the minimum length of the connected components of C ,
- (ii) the gain function g and $\mathcal{H}$ are differentiable,
- (iii) (strong monotonicity) if $X, Y \in b\mathcal{F}$ with $X \leq Y$ a.s. and $\mathbb{P}(X < Y) > 0$ then $\rho(X) < \rho(Y)$.

Recall from the proof of Theorem 2.10 that if $x \in C$ then there exist $0 \leq x^- < x < x^+ \leq 1$ such that $V = h_{g(x^-), g(x^+)}^{x^-, x^+}$ on $[x^-, x^+]$. Denoting by h'^+ , h'^- the right and left derivatives of h respectively, we obtain the following lemma.

Lemma 2.22. Setting

$$\begin{aligned} F(x) &:= \{(y, z) : 0 \leq y < x < z \leq 1, h = h_{g(y), g(z)}^{y, z} \text{ satisfies} \\ &\quad y(h'^+(y) - g'(y)) = (1 - z)(h'^-(z) - g'(z)) = 0\}, \\ G(x) &= \sup\{h_{g(y), g(z)}^{y, z}(x) : (y, z) \in F(x)\}, \\ \tilde{F}(x) &= \arg \max\{h_{g(y), g(z)}^{y, z}(x) : (y, z) \in F(x)\}, \end{aligned}$$

(with the convention that $\sup \emptyset = -\infty$) we have:

- (i) $x \in S$ iff $G(x) \leq g(x)$,
- (ii) $x \in C$ iff $G(x) > g(x)$ and then

$$\begin{aligned} x^- &= \sup\{y : (y, z) \in \tilde{F}(x)\}, \quad x^+ = \inf\{z : (y, z) \in \tilde{F}(x)\}, \\ V &= h_{g(x^-), g(x^+)}^{x^-, x^+} \text{ on } [x^-, x^+]. \end{aligned}$$

Proof. For part (i) note that if $x \in C$ then $G(x) = V(x) > g(x)$, giving the implication $G(x) \leq g(x) \implies x \in S$. Conversely, if $x \in S$ then by optimality we have $G(x) \leq g(x) = V(x)$.

For part (ii), by the proof of Theorem 2.10 and the smooth fit principle we have $(x^-, x^+) \in \tilde{F}(x)$, and it is straightforward to show that $x^- = \sup\{y : (y, z) \in \tilde{F}(x)\}$ and $x^+ = \inf\{z : (y, z) \in \tilde{F}(x)\}$, as follows (we show only the latter). Suppose that $(y, z) \in \tilde{F}(x)$. Then if $z < x^+$ we have either $y > x^-$ or $y \leq x^-$.

If $x^- < y < x < z < x^+$, by construction we have $\min\{V(y) - g(y), V(z) - g(z)\} > 0$ and monotonicity and the ϱ -martingale property give

$$h_{g(y), g(z)}^{y, z}(x) < h_{V(y), V(z)}^{y, z}(x) = V(x) = h_{g(x^-), g(x^+)}^{x^-, x^+}(x),$$

contradicting the definition of $\tilde{F}(x)$.

Supposing alternatively that $y \leq x^- < x < z < x^+$, we again have $V(z) > g(z) = h_{g(y), g(z)}^{y, z}(z)$. Now if $h_{g(y), g(z)}^{y, z}(x^-) \leq V(x^-)$ then monotonicity and part (ii) of Assumption 2.21 give $h_{g(y), g(z)}^{y, z}(x) < V(x)$, again contradicting the definition of $\tilde{F}(x)$. Thus $h_{g(y), g(z)}^{y, z}(x^-) > V(x^-) = g(x^-)$, which contradicts the definition of V . We conclude that $z \geq x^+$. $\square$

Lemma 2.22 is the basis for Algorithm 1.

Remark 2.23. Part (i) of Assumption 2.21 ensures that the continuation region C has finitely many (at most $\lfloor 1/\delta \rfloor$) connected components and, thus, that Algorithm 1 terminates in finite time. Since the assumptions made in this section are strong, it is worth remarking on the output of Algorithm 1 when Assumptions 2.3 and 2.21 do not hold. The algorithm identifies the points x^+, x^- defined in (15)–(16) using the principle of smooth fit. It may be that for certain initial points x_0 , the principle of smooth fit applies in a sufficiently large neighbourhood of x_0 . An example is provided in Section 3.2 below using the worst-case risk mapping. In this case it is not difficult to see that Algorithm 1 correctly returns the value $V(x_0)$ and the optimal stopping

Algorithm 1 Under Assumptions 2.3 and 2.21 this algorithm solves the optimal stopping problem by returning the set of connected components of the continuation region C .

Evaluate

$$F := \{(y, z) : 0 \leq y < z \leq 1, h = h_{g(y), g(z)}^{y, z} \text{ satisfies} \\ y(h'^+(y) - g'(y)) = (1 - z)(h'^-(z) - g'(z)) = 0\},$$

Set $x = 0$, $B = \emptyset$

while $x \leq 1$ **do**

Set

$$F(x) := \{(y, z) \in F : y < x < z\}, \\ G(x) = \sup\{h_{g(y), g(z)}^{y, z}(x) : (y, z) \in F(x)\}, \\ \tilde{F}(x) = \arg \max\{h_{g(y), g(z)}^{y, z}(x) : (y, z) \in F(x)\}, \\ x^- = \sup\{y : (y, z) \in \tilde{F}(x)\}, \\ x^+ = \inf\{z : (y, z) \in \tilde{F}(x)\},$$

if $G(x) > g(x)$ **then**

Append the element (x^-, x^+) to B

Set $x = x^+$

end if

Increase x by δ

end while

return B

time τ^* for initial points $x_0 \in (\frac{1}{8}, \frac{5}{8})$. However for initial points $x_0 \in (\frac{3}{4}, 1)$, the lack of smooth fit at $x = \frac{3}{4}$ prevents the component $(\frac{1}{8}, \frac{5}{8})$ of C from being identified.

Nevertheless the characterisation (15)–(16) of the points x^+, x^- uses a weaker property than smooth fit, namely *continuous fit* between g and V . Since formula (2) correctly constructs V under the worst-case risk mapping, it seems possible to modify Algorithm 1 to use continuous fit instead to identify the points x^+, x^- . While outside the scope of the present paper, such a modification would enable the algorithm to solve the example of Section 3.2 for all $x \in C$.

3 Examples

In this section we consider optimal stopping under three different risk mappings ρ . For a unified framework, consider the construction

$$\rho(Z) := \inf_{Q \ll \mathbb{P}} (\mathbb{E}^Q[Z] + \alpha(Q)), \quad Z \in b\mathcal{F}, \quad (27)$$

where for each $Q \ll \mathbb{P}$, $\alpha(Q) \in [0, \infty]$ is a ‘penalty’ associated with Q such that $\inf_{Q \ll \mathbb{P}} \alpha(Q) = 0$, so that ρ is indeed a risk mapping (which is also concave, see e.g. [22, p. 207]). Thus (27) may be used directly to construct a risk mapping by specifying the penalty function α , and this is the approach which will be taken below. It is interesting to note that further examples of risk mappings of the form (27) can be constructed from the solutions to backwards stochastic differential equations (BSDEs, see for example [2, 12]), and such examples are time consistent by construction.

Section 3.1 recovers the classical linear solution of [21, Ch. III.7] and Section 3.2 revisits the worst-case risk mapping of Section 1.1, while Section 3.3 addresses the entropic risk mapping. Unless otherwise stated, we take the underlying diffusion Y to be standard Brownian motion.

3.1 Linear case

Setting $\alpha(Q) = 0$ for $Q = \mathbb{P}$ and $\alpha(Q) = +\infty$ otherwise gives the linear expectation $\rho(Z) = \mathbb{E}[Z]$, which has the corresponding dynamic risk mapping

$$\rho_t(Z) = \mathbb{E}[Z|\mathcal{F}_t], \quad t \geq 0.$$

In this case, $\mathcal{H}$ consists of line segments:

$$h_{\beta,\gamma}^{y,z}(x) = \beta\mathbb{P}(\tau_y^x < \tau_z^x) + \gamma\mathbb{P}(\tau_z^x < \tau_y^x) = \beta\frac{z-x}{z-y} + \gamma\frac{x-y}{z-y}, \quad x \in [y, z], \quad (28)$$

(see for example [43, Th. 30]). Thus $\mathcal{H}$ is differentiable, and Δ -equicontinuous by Lemma 2.8. Also, ϱ is continuous by the conditional bounded convergence theorem for $\mathbb{E}$ and time consistent by its conditional tower property. The strong Markov property (Definition 1.5) is trivially satisfied as Brownian motion is a strong Markov process under $\mathbb{E}$.

Thus Assumption 2.3 holds, and Theorem 2.10 gives that if g is continuous then $V = w$ (cf. (11)). It is straightforward to check that w is then equal to the pointwise infimum over those $h \in \mathcal{H}^\dagger$ (cf. (10)) which dominate g or, equivalently, that w is the smallest concave majorant of g , and we recover the classical geometric method for optimal stopping problems [21, Ch. III.7].

More generally if X^x is a one-dimensional diffusion, stopped as in Section 1.2, then we recover the result of Dayanik and Karatzas for undiscounted optimal stopping on a bounded interval (see [16]). To see this write $Z_t^{S(x)} = S(X_t^x)$, where S is the scale function of X^x , and note from part (i) of Definition 2.1 that the set $\mathcal{H}$ of generalised potentials depends implicitly on the law of the process X^x . Thus, for clarity we may denote by $\mathcal{H}_X$ and $\mathcal{H}_Z$ the sets of generalised potentials induced respectively by the processes X^x and $Z^{S(x)} = S(X^x)$. According to [16, Prop. 2.2], the elements of $\mathcal{H}_X$ take the form

$$h_{\beta,\gamma}^{y,z}(x) = \beta\frac{S(z) - S(x)}{S(z) - S(y)} + \gamma\frac{S(x) - S(y)}{S(z) - S(y)}, \quad x \in [y, z]. \quad (29)$$

Then since $X^x = S^{-1}(Z^{S(x)})$, it follows from the definition of $h_{\beta,\gamma}^{y,z}$ via hitting probabilities (Definition 2.1) that the elements of $\mathcal{H}_Z$ take the form

$$h_{\beta,\gamma}^{y,z}(x) = \beta\frac{z-x}{z-y} + \gamma\frac{x-y}{z-y}, \quad x \in [y, z],$$

which is the same as (28), and the conditions of Theorem 2.10 are again satisfied. Rewriting problem (1) as $V(x) = \sup_{\tau \in \mathcal{T}} \mathbb{E} \left[g \circ S^{-1} \left(Z_\tau^{S(x)} \right) \right]$, we conclude that its solution is the smallest concave majorant of $g \circ S^{-1}$, which is [16, Prop. 3.3].

3.2 Worst-case risk mapping

As discussed for example in [19], where further references can be found, many ‘worst case’ risk mappings are naturally associated with settings involving ambiguity: for example, when the drift function μ or the volatility function σ is ambiguous. These respective settings can be treated in two frameworks due to Peng: the former using the g -expectation [34] (see also [8]), and the latter

with G -Brownian motion and G -expectation [36, 35]. In this section we consider a definition of ‘worst case’ which is not related to ambiguity: namely, when the drift and volatility functions are known and the essential infimum of the payoff is considered (this was the nonlinear expectation considered in the example of Section 1.1). This can be constructed by setting $\alpha(Q) = 0$ in (27) for all $Q \ll \mathbb{P}$, giving $\rho(Z) = \mathbb{P}\text{-essinf}(Z)$ [3, Remark 3.7]. Although the worst-case risk mapping does not satisfy Assumption 2.3, it was observed in Section 1.1 that formula (2) generates the correct solution and, as such, is a useful heuristic. This heuristic is novel to the best of our knowledge: for example, it differs from the use of the simpler set $\mathcal{H}^\dagger$ (cf. (10)). Indeed it is easy to see that if $\mathcal{H}$ is replaced by $\mathcal{H}^\dagger$ in (11), instead of w we obtain the constant function $w^\dagger(x) \equiv \bar{g}$, and for all $x \notin \arg \sup_{y \in [0,1]} g(y)$ we have

$$V(x) = w(x) = \min\left\{\sup_{y \in [0,x]} g(y), \sup_{y \in [x,1]} g(y)\right\} \neq w^\dagger(x) = \bar{g}.$$

Formula (2) also differs from the heuristic use of the principle of smooth fit (cf. Corollary 2.20). This can be seen in Figure 2, where the gain function is $g(x) = 1 + \sin(4\pi x)$ and, as observed in Remark 2.23, the use of smooth fit fails for initial points $x_0 \in (\frac{3}{4}, 1)$.

3.3 Entropic case

Setting $Y := \frac{dQ}{d\mathbb{P}}$ and $\alpha(Q) = \mathbb{E}^Q[\ln Y]$ in (27), one obtains the entropic risk mapping [22, Ex. 4.34]:

$$\rho(Z) = -\ln \mathbb{E}[e^{-Z}], \quad Z \in b\mathcal{F},$$

which has the corresponding dynamic conditional risk mapping

$$\rho_t(Z) = -\ln \mathbb{E}[e^{-Z} | \mathcal{F}_t], \quad t \geq 0.$$

(For an alternative construction of the entropic risk mapping via quadratic BSDEs see, for example, [2, Proposition 6.4], [12, 18]). From (28), for $x \in [y, z]$ we have $h_{\beta, \gamma}^{y,z}(x) = -\ln\left(e^{-\beta \frac{z-x}{z-y}} + e^{-\gamma \frac{x-y}{z-y}}\right)$, so $\mathcal{H}$ is again differentiable, and Δ -equicontinuous by Lemma 2.8. The continuity, strong Markov property and time consistency of ρ follow from the corresponding properties for $\mathbb{E}$ as in Section 3.1, so Assumption 2.3 holds and Theorem 2.10 gives that if g is continuous then $V = w$.

In this case it is a priori equivalent to rescale the problem and instead maximise the linear expectation $\mathbb{E}[e^{-g(X_\tau^\tau)}]$, for which the classical construction using $\mathcal{H}^\dagger$ suffices. Interestingly, though, the direct construction of the present paper is qualitatively different, in the sense that the set $\mathcal{H}$ must again be used rather than $\mathcal{H}^\dagger$. Indeed, fix $x \in (0, 1)$ and let g be any gain function with $g(0) = 1$ and $g(x) > 2 - \ln(1 - x)$. We have $h_{0, \gamma}^{0,1}(x) \uparrow -\ln(1 - x)$ as $\gamma \uparrow \infty$. Then if $\mathcal{H}^\dagger$ is used to construct $w^\dagger$ (rather than using $\mathcal{H}$ to construct w), by part (v) of Lemma 2.6 (translation invariance) we would have $w^\dagger(0) \geq 2 > 1 = w(0) = V(0) = g(0)$, so $w^\dagger$ cannot be the value function.

3.4 A sufficient condition for Algorithm 1

We conclude by providing an example problem for which Assumption 2.21 holds and, thus, Algorithm 1 may be applied. Our focus is on part (i) of Assumption 2.21, since the remaining parts are straightforward.

Suppose that each $h \in \mathcal{H}$ is a convex function on $d(h)$. Examples include the linear case of Section 3.1, the entropic case of Section 3.3, and further examples which can be constructed using

the dual representation (27), as follows. From (29), the potentials $h \in \mathcal{H}_X$ are convex precisely when the scale function S of X^x is convex. (This holds, for example, when X^x has negative drift.) Let $\mathcal{Q}$ be a finite set of measures $Q \ll \mathbb{P}$, each corresponding to a regular diffusion with a convex scale function. If the penalty function α is finite on $\mathcal{Q}$, and is set to $+\infty$ otherwise, the resulting concave risk mapping ρ also has convex potentials $\mathcal{H}$.

Let $C_0 = (x_0, y_0)$ be a connected component of the continuation region C . Then since V dominates g and is convex on C_0 with $V = g$ at its endpoints, g cannot be concave on the whole of C_0 (otherwise $V = g$ on C_0 , a contradiction). Thus g is convex on some subinterval of C_0 . Hence a sufficient condition for part (i) of Assumption 2.21 is that $\mathcal{A}$ is finite, where $\mathcal{A}$ is the set of (maximal) intervals on which g is convex, and then we may take $\delta < \min_{[x_1, y_1] \in \mathcal{A}} (y_1 - x_1)$. Note that δ can be found from knowledge of g only.

Acknowledgements. The authors would like to thank Giorgio Ferrari, Ben Hambly, Vicky Henderson, David Hobson and Jan Palczewski for comments on earlier versions of this work, and also two anonymous referees whose feedback significantly improved the manuscript.

Funding. This work was supported by the Lloyd’s Register Foundation-Alan Turing Institute programme on Data-Centric Engineering under the LRF grant G0095.

References

- [1] L. Alvarez. Knightian uncertainty, k-ignorance, and optimal timing. Discussion Papers 25, Aboa Centre for Economics, 2007.
- [2] Pauline Barrieu and Nicole El Karoui. Pricing, hedging, and designing derivatives with risk measures. *Indifference pricing: Theory and applications*, pages 77–146, 2009.
- [3] Emmanuel N Barron, Pierre Cardaliaguet, and Robert R Jensen. Radon—nikodym theorem in l^∞ . *Applied Mathematics and Optimization*, 42:103–126, 2000.
- [4] E. Bayraktar and S. Yao. Optimal stopping for non-linear expectations—Part I. *Stochastic Process. Appl.*, 121(2):185–211, 2011.
- [5] E. Bayraktar and S. Yao. Optimal stopping for non-linear expectations—Part II. *Stochastic Process. Appl.*, 121(2):212–264, 2011.
- [6] Martin Beibel and Hans Rudolf Lerche. A new look at optimal stopping problems related to mathematical finance. *Statistica Sinica*, 7(1):93–108, 1997.
- [7] J. Bion-Nadal. Time consistent dynamic risk processes. *Stochastic Process. Appl.*, 119(2):633–654, 2009.
- [8] Z. Chen and L. Epstein. Ambiguity, risk, and asset returns in continuous time. *Econometrica*, 70(4):1403–1443, 2002.
- [9] X. Cheng and F. Riedel. Optimal stopping under ambiguity in continuous time. *Math. Financ. Econ.*, 7(1):29–68, 2013.

- [10] P. Cheridito, F. Delbaen, and M. Kupper. Dynamic monetary risk measures for bounded discrete-time processes. *Electron. J. Probab.*, 11:no. 3, 57–106, 2006.
- [11] M. J. Cho and R. H. Stockbridge. Linear programming formulation for optimal stopping problems. *SIAM Journal on Control and Optimization*, 40(6):1965–1982, 2002.
- [12] Wing Fung Chong, Ying Hu, Gechun Liang, and Thaleia Zariphopoulou. An ergodic bsde approach to forward entropic risk measures: representation and large-maturity behavior. *Finance and Stochastics*, 23(1):239–273, 2019.
- [13] S. Christensen. Optimal decision under ambiguity for diffusion processes. *Math. Methods Oper. Res.*, 77(2):207–226, 2013.
- [14] S. Christensen, F. Croce, E. Mordecki, and P. Salminen. On optimal stopping of multidimensional diffusions. *Stochastic Process. Appl.*, 129(7):2561–2581, 2019.
- [15] F. Croce and E. Mordecki. An algorithm to solve optimal stopping problems for one-dimensional diffusions. *ALEA Lat. Am. J. Probab. Math. Stat.*, 19(2):1353–1375, 2022.
- [16] S. Dayanik and I. Karatzas. On the optimal stopping problem for one-dimensional diffusions. *Stochastic Process. Appl.*, 107(2):173–212, 2003.
- [17] F. Delbaen, S. Peng, and E. Rosazza Gianin. Representation of the penalty term of dynamic concave utilities. *Finance Stoch.*, 14(3):449–472, 2010.
- [18] Freddy Delbaen, Ying Hu, and Adrien Richou. On the uniqueness of solutions to quadratic bsdes with convex generators and unbounded terminal conditions. In *Annales de l’IHP Probabilités et statistiques*, volume 47, pages 559–574, 2011.
- [19] Laurent Denis, Mingshang Hu, and Shige Peng. Function spaces and capacity related to a sublinear expectation: application to g-brownian motion paths. *Potential analysis*, 34:139–161, 2011.
- [20] E. B. Dynkin. The optimum choice of the instant for stopping a markov process. *Soviet Mathematics*, 4:627–629, 1963.
- [21] E. B. Dynkin and A. A. Yushkevich. *Markov processes: Theorems and problems*. Nauka Press, Moscow, 1967.
- [22] H. Föllmer and A. Schied. *Stochastic finance. An introduction in discrete time*. De Gruyter Graduate. De Gruyter, Berlin, 2016. Fourth revised and extended edition.
- [23] K. Helmes and R. H Stockbridge. Construction of the value function and optimal rules in optimal stopping of one-dimensional diffusions. *Advances in Applied Probability*, 42(1):158–182, 2010.
- [24] W. Huang and W. B. Haskell. Risk-aware q-learning for markov decision processes. In *2017 IEEE 56th Annual Conference on Decision and Control (CDC)*, pages 4928–4933, 2017.
- [25] D. Jelito, M. Pitera, and Ł. Stettner. Risk sensitive optimal stopping. *Stochastic Process. Appl.*, 136:125–144, 2021.
- [26] D. Kahneman and A. Tversky. Prospect theory: An analysis of decision under risk. *Econometrica*, 47(2):263–292, 1979.

- [27] O. Kallenberg. *Foundations of modern probability*. Probability and its Applications. Springer-Verlag, New York, second edition, 2002.
- [28] I. Karatzas and I.-M. Zamfirescu. Martingale approach to stochastic differential games of control and stopping. *Ann. Probab.*, 36(4):1495–1527, 2008.
- [29] T. Kosmala, R. Martyr, and J. Moriarty. Markov risk mappings and risk-sensitive optimal prediction. *Math. Methods Oper. Res.*, 97(1):91–116, 2023.
- [30] M. Kupper and W. Schachermayer. Representation results for law invariant time consistent functions. *Mathematics and Financial Economics*, 2:189–210, 2009.
- [31] T. Lam, A. Verma, B. Kian Hsiang Low, and P. Jaillet. Risk-aware reinforcement learning with coherent risk measures and non-linear function approximation. In *The Eleventh International Conference on Learning Representations*, 2023.
- [32] M. Nutz and H. M. Soner. Superhedging and dynamic risk measures under volatility uncertainty. *SIAM J. Control Optim.*, 50(4):2065–2089, 2012.
- [33] M. Nutz and J. Zhang. Optimal stopping under adverse nonlinear expectation and related games. *Ann. Appl. Probab.*, 25(5):2503–2534, 2015.
- [34] Shige Peng. Backward sde and related g-expectation. *Pitman research notes in mathematics series*, pages 141–160, 1997.
- [35] Shige Peng. G-brownian motion and dynamic risk measure under volatility uncertainty. *arXiv preprint arXiv:0711.2834*, 2007.
- [36] Shige Peng. G-expectation, g-brownian motion and related stochastic calculus of itô type. In *Stochastic Analysis and Applications: The Abel Symposium 2005*, pages 541–567. Springer, 2007.
- [37] G. Peskir and A. N. Shiryaev. *Optimal Stopping and Free-Boundary Problems*. Lectures in Mathematics. ETH Zürich. Birkhäuser Basel, Basel, 2006.
- [38] G. Ch. Pflug and W. Römisch. *Modeling, measuring and managing risk*. World Scientific Publishing Co. Pte. Ltd., Hackensack, NJ, 2007.
- [39] C. C. Pugh. *Real mathematical analysis*. Undergraduate Texts Math. Cham: Springer, 2nd ed. edition, 2015.
- [40] L. C. G. Rogers and David Williams. *Diffusions, Markov processes and martingales. Vol. 1: Foundations*. Cambridge: Cambridge University Press, 2nd ed. edition, 2000.
- [41] L. C. G. Rogers and David Williams. *Diffusions, Markov processes, and martingales. Vol. 2: Itô calculus*. Cambridge: Cambridge University Press, 2nd ed. edition, 2000.
- [42] Stefan Röhl. *Ein linearer Programmierungsansatz zur Lösung von Stopp- und Steuerungsproblemen*. PhD thesis, Humboldt-Universität zu Berlin, Wirtschaftswissenschaftliche Fakultät, 2001.
- [43] R. Serfozo. *Basics of applied stochastic processes*. Probability and its Applications (New York). Springer-Verlag, Berlin, 2009.
- [44] W. Wang and H. Xu. Preference robust distortion risk measure and its application. *Math. Finance*, 33(2):389–434, 2023.