

# Fine Properties of the Optimal Skorokhod Embedding Problem\*

Mathias Beiglböck<sup>†</sup>    Marcel Nutz<sup>‡</sup>    Florian Stebegg<sup>§</sup>

April 13, 2020

## Abstract

We study the problem of stopping a Brownian motion at a given distribution  $\nu$  while optimizing a reward function that depends on the (possibly randomized) stopping time and the Brownian motion. Our first result establishes that the set  $\mathcal{T}(\nu)$  of stopping times embedding  $\nu$  is weakly dense in the set  $\mathcal{R}(\nu)$  of randomized embeddings. In particular, the optimal Skorokhod embedding problem over  $\mathcal{T}(\nu)$  has the same value as the relaxed one over  $\mathcal{R}(\nu)$  when the reward function is semi-continuous, which parallels a fundamental result about Monge maps and Kantorovich couplings in optimal transport. A second part studies the dual optimization in the sense of linear programming. While existence of a dual solution failed in previous formulations, we introduce a relaxation of the dual problem that exploits a novel compactness property and yields existence of solutions as well as absence of a duality gap, even for irregular reward functions. This leads to a monotonicity principle which complements the key theorem of Beiglböck, Cox and Huesmann [Optimal transport and Skorokhod embedding, *Invent. Math.*, 208:327–400, 2017]. We show that these results can be applied to characterize the geometry of optimal embeddings through a variational condition.

*Keywords* Skorokhod Embedding; Randomized Stopping Time; Duality

*AMS 2010 Subject Classification* 60G40; 60G44; 90C08

---

\*The authors thank two anonymous referees for their detailed comments.

<sup>†</sup>University of Vienna, Department of Mathematics, mathias.beiglboeck@univie.ac.at. Research supported by FWF Grant Y-782.

<sup>‡</sup>Columbia University, Departments of Statistics and Mathematics, mnutz@columbia.edu. Research supported by an Alfred P. Sloan Fellowship and NSF Grants DMS-1512900 and DMS-1812661. MN is grateful to Alex Cox and Johannes Ruf for helpful discussions regarding Sections 6 and 7.1, respectively.

<sup>§</sup>Columbia University, Department of Statistics, florian.stebegg@columbia.edu.

# 1 Introduction

Given a centered and suitably integrable probability distribution  $\nu$  and a Brownian motion  $B$ , the Skorokhod embedding problem [79] consists in finding a stopping time  $\tau$  which embeds  $\nu$ ; that is,  $B_\tau$  has distribution  $\nu$ . A number of solutions exist and we denote the set of all such  $\tau$  by  $\mathcal{T}(\nu)$ . Examples include the classical Root [77] and Rost [78] embeddings; see [71] for a survey of various solutions. The *optimal* Skorokhod embedding problem is to maximize (or minimize) the expectation  $E[G_\tau]$  over  $\tau \in \mathcal{T}(\nu)$ , where  $G_t = G((B_s)_{s \leq t}, t)$  is an adapted functional. For instance, the Root embedding minimizes  $E[\tau^2]$  and the Rost embedding maximizes it; cf. [78]. Some early works related to optimal Skorokhod embeddings are [16, 47, 53, 61]. More recently, connections to numerous questions in probability, analysis and finance as well as extensions such as the multi-marginal case [7, 24, 42] have emerged and led to substantial activity; we refer to [48] for a survey with many more references.

A perspective pioneered by [6] is to see Skorokhod embeddings along the lines of optimal transport theory: optimal stopping times are analogues of Monge solutions to an optimal transport problem between the Wiener measure and  $\nu$ . A more general formulation of the embedding problem allows for a randomized stopping time; this can be interpreted as using an enlarged filtration or allowing for an external randomization (see Definition 2.1). The corresponding set is denoted  $\mathcal{R}(\nu)$  and gives rise to a relaxed formulation of the optimal Skorokhod embedding problem. Continuing the analogy, randomized stopping times correspond to transports in the sense of Kantorovich. We refer to [4, 74, 75, 83, 84] for background on optimal transport.

In the existing literature on the optimal Skorokhod embedding problem, a number of optimal embeddings have been found for specific reward functionals  $G$ ; e.g., [23, 47, 50, 51, 52]. In these examples, optimal embeddings are often unique and belong to the class  $\mathcal{T}(\nu)$  of stopping times even if one allows for randomized stopping times a priori. On the other hand, results concerning the general structure of the optimal Skorokhod embedding problem such as [6, 41, 42] use the formulation with randomized stopping times. Thus, an obvious—but not previously addressed—question is how to bridge this gap: when can the supremum value over randomized stopping times be achieved with stopping times? More generally, can randomized stopping times be approximated by stopping times in a suitable sense?

The analogy to classical optimal transport theory is apparent. While the Kantorovich relaxation is crucial to develop the theory, most examples of specific interest lead to transport maps in the sense of Monge. For instance,

Brenier's theorem states that the optimal transport for the quadratic cost (or reward, after changing the sign) is given by the gradient of a convex function if the first marginal measure is absolutely continuous (or, more generally, regular [62]). When the first marginal is atomless, it was shown in [3] that Monge transports form a dense subset of Kantorovich couplings and that the values of the Monge and Kantorovich transport problems agree for bounded continuous reward functions. This result was extended to unbounded continuous functions in [73]; see also [59] for a survey of related density properties.

In the first part of this paper, we provide comparable results for the optimal Skorokhod embedding problem. In Theorem 2.3 we show that the optimal embedding problems over  $\mathcal{T}(\nu)$  and  $\mathcal{R}(\nu)$  have the same value whenever the reward functional  $G$  is lower semicontinuous in time. This assertion can fail when  $G$  is not lower semicontinuous (Example 2.4), and that failure highlights a contrast with classical approximation results for randomized stopping times that hold without regularity conditions [5, 27, 34, 37]: the constraint given by the fixed embedding target  $\nu$  is not compatible with the classical results and techniques. In Theorem 3.1 we establish the more general result that  $\mathcal{T}(\nu) \subseteq \mathcal{R}(\nu)$  is dense for weak convergence. Our proof is constructive and gives insight into why the first result can fail when  $G$  is irregular. In a nutshell, the idea is to use a short initial segment of the Brownian path as a randomization device for the rest of the problem. In fact, we show that for randomized stopping times  $\xi \in \mathcal{R}(\nu)$  and reward functions  $G$  that do not depend on an initial segment of the paths (in a sense to be made precise), the expectation of  $G$  stopped at  $\xi$  can be exactly replicated by a stopping time  $\tau \in \mathcal{T}(\nu)$ , without any need for approximation (cf. Proposition 3.9). These ideas seem to be novel in the literature.

The optimal Skorokhod embedding problem over  $\mathcal{R}(\nu)$  is a linear programming problem with constraints and thus has a dual programming problem. Formally, the domain of the dual problem is the set of all pairs  $(M, \psi)$  where  $M$  is a martingale with  $M_0 = 0$  and  $\psi : \mathbb{R} \rightarrow \mathbb{R}$  is a function such that  $M_t + \psi(B_t) \geq G_t$  for  $t \geq 0$ . The dual problem then consists in minimizing  $\nu(\psi) := \int \psi d\nu$  over all such pairs  $(M, \psi)$ . More specifically, [6] uses martingales  $M$  that satisfy a quadratic growth condition relative to  $B$  and functions  $\psi$  that are continuous and satisfy a growth condition, or [42] works with similar functions  $\psi$  and supermartingales  $M$ . Such a dual problem has been used in numerous examples to help determine specific optimal Skorokhod embeddings; e.g., [22, 23, 25, 44, 47, 50, 51]. Moreover, it plays a vital role in [6] (see also [41]) in deriving a general monotonicity principle that describes the barriers representing optimal Skorokhod embeddings

through their hitting times. While dual solutions have been found in those specific examples, it has been observed in [6] that the dual problem fails to have a solution in general; that is, the minimum is not attained. We refer to the survey [48] for further references.

The second part of this paper introduces a novel relaxation of the dual problem and establishes the existence of its solution as well as the absence of a duality gap. Both of these results are obtained without continuity conditions for  $G$ , thus paralleling the generality of Kellerer’s theorem [58] in optimal transport and improving the results on the absence of a duality gap in [6, 42]. There are no previous existence results in our setting. We can mention [39] for a PDE approach with attainment in a different dual problem. Here the reward  $G$  is given by the integral of a continuous, finite-dimensional Lagrangian with exponential decay and the marginals are absolutely continuous with compact support. Remarkably, [39] allows for multidimensional Brownian motion.

Even if  $G$  is continuous and bounded, our main issue is the lack of compactness for the martingale component. Broadly speaking, for a given minimizing sequence  $(M^n, \psi^n)$  the functions  $\psi^n$  may have large positive values, so that the inequality  $M_t^n + \psi^n(B_t) \geq -\|G^-\|_\infty$  does not immediately result in a lower bound for  $M^n$ . On the other hand, the limit of a sequence of continuous-time (super-)martingales may fail to be a supermartingale in the absence of a lower bound. A crucial feature of our relaxation is to work with local martingales that have uniform lower bounds on sets where the Brownian motion is bounded and bounded away from the extremes of the support of  $\nu$ . It turns out that on such sets we can obtain enough compactness, while preserving the “weak” side of the duality. Less surprisingly, our functions  $\psi$  are merely in  $L^1(\nu)$  rather than continuous. We provide counterexamples showing that our positive results on duality can fail if one were to insist on true (super-)martingales or continuous functions; cf. Section 7.

An analogous duality result was obtained in [12] (see also [11]) for the so-called martingale optimal transport problem in a single period. The main compactness issue sketched above does not arise in that setting; basically, limits of martingales remain martingales in the discrete-time setting. A softer part of our proof does overlap with [12], namely the use of capacity theory to generalize from continuous to measurable reward functionals. Martingale optimal transport refers to a transport problem where the transports are constrained to be martingales. Originally motivated by model uncertainty in financial applications [47], a rich literature has emerged around this subject; see [48, 71, 82] for surveys and, e.g., [1, 8, 15, 9, 18, 19, 31, 35, 38, 40, 54, 67, 69, 70, 85] for models in discrete time and [14, 20, 23, 32, 33, 36, 44,

45, 46, 49, 56, 66, 68, 80, 10, 81] for continuous time. Many of these works exploit connections to the Skorokhod embedding problem. In particular, the optimal Skorokhod embedding problem can be related, by a time change, to the continuous-time martingale transport between two marginals. While a duality theory with dual existence has been elusive for the latter transport problem, the arguments in the present paper are expected to lead to such a result since the compactness issues are similar. This will be investigated in future work.

Our results on duality allow us to derive a monotonicity principle in the spirit of [12], under an integrability condition on  $G$ . Namely, there exists a universal support  $\Gamma$  that characterizes all optimal embeddings:  $\xi \in \mathcal{R}(\nu)$  is optimal if and only if  $\xi(\Gamma) = 1$ . This complements the monotonicity principle of [6] which gives a geometric condition on the support that is necessary for optimality, but not sufficient. By contrast, our result yields a necessary and sufficient condition. However, the geometry of the support is merely described in a weaker form, through the construction of  $\Gamma$  as the set where a dual optimizer equals the reward functional (we exemplify in Section 6 how this can be utilized to obtain more specific geometric statements). It is an interesting question for future research how to unify these results, though the answer does not seem to be within reach with present concepts and knowledge. It is noteworthy that the integrability condition is crucial for any monotonicity principle to hold; indeed, we provide a surprising example showing that for general  $G$ , the optimality of embeddings cannot be characterized in terms of the support. This contrasts with cyclical monotonicity properties in classical optimal transport [84] as well as the result of [12] which suggest that optimality of transports can be characterized by their geometry in great generality.

In Section 6 we illustrate how our result on dual existence can be exceptionally useful to describe the geometry of optimal Skorokhod embeddings in a concrete case. Namely, we specialize to a particular class of convex-concave reward functions  $G_t = g(t)$  which give rise to embeddings that can be represented as hitting times of sets consisting of both a left and a right barrier in the  $(t, x)$ -plane. This class of “cave” embeddings with a double boundary, unifying the ones of Root and Rost, was introduced in [6]. In contrast to Root’s and Rost’s, such barriers are not determined by  $\nu$  alone but depend on the details of the function  $g$ , and the arguments in [6] do not lead to a characterization of the optimal barriers. We provide such a characterization through a variational condition, very much inspired by [22] which studies a different class of reward functions. The condition is related to the principle of smooth fit for free boundary problems, and thus it is no surprise that the

proof of sufficiency for optimality takes the form of a verification argument. In [22], the proof of necessity is an impressive tour de force through a discretization that is carried out in the separate paper [21]. On the strength of our result on dual existence, we can provide a much more direct proof. First, we show that when the reward function is Markovian (that is, a deterministic function of time and current state), the abstract martingale  $M$  in the dual can be replaced by a function of two variables. Then, we can apply relatively soft probabilistic arguments to derive the variational characterization. Importantly, our proof reveals that similar conditions should extend to much more general classes of embeddings, and also suggests that regularity results for the stopping boundaries can be obtained through the dual maximizer. These aspects will be investigated in separate work.

The remainder of this paper is organized as follows. Section 2 details the optimal Skorokhod embedding problem and states the equality of the formulations with randomized and non-randomized stopping times for regular reward functions. The proof is given in Section 3 where it is shown more generally how randomized stopping times can be approximated by non-randomized ones. Section 4 introduces the relaxed dual problem and provides the existence of a solution. The absence of a duality gap is proved in Section 5 where we also state the monotonicity principle. In Section 6 we discuss cave embeddings and characterize the optimal barriers by exploiting our abstract results. Counterexamples regarding the formulation of the dual problem and the monotonicity principle are gathered in Section 7. For simplicity of exposition, we use a second moment condition on  $\nu$  throughout the paper. Appendix A explains how this can be replaced by a finite first moment without much effort.

## 2 The Primal Problem

Let  $C_0(I)$  be the set of continuous real-valued functions  $\omega = (\omega_t)_{t \in I}$  with  $\omega_0 = 0$ , for any interval  $0 \in I \subseteq \mathbb{R}$ . We denote by  $S$  the space of stopped continuous paths; that is, pairs  $(\omega, t)$  with  $t \in \mathbb{R}_+$  and  $\omega \in C_0([0, t])$ . The set  $S$  is a Polish space under the topology induced by the metric

$$d((f, s), (g, t)) = |t - s| \vee \sup_{u \geq 0} |f_{u \wedge s} - g_{u \wedge t}|.$$

We note that any  $(\omega, t) \in C_0(\mathbb{R}_+) \times \mathbb{R}_+$  projects to an element  $(\omega|_{[0, t]}, t)$  of  $S$ . Conversely, we can embed  $S$  in  $C_0(\mathbb{R}_+) \times \mathbb{R}_+$ , say by continuing any stopped path in a constant fashion. This identification will sometimes be used implicitly.

We fix a reward function  $G : S \rightarrow [0, \infty]$ . Any such  $G$  can be seen as an “adapted” process (in the sense of Galmarino’s test) on the canonical space  $C_0(\mathbb{R}_+)$  in that  $G_t(\omega) := G(\omega, t)$  depends only on  $\omega|_{[0, t]}$ . If  $G$  is Borel measurable, it can be identified with an optional process on  $C_0(\mathbb{R}_+)$  when the latter is equipped with the (raw) canonical filtration  $\mathbb{F} = (\mathcal{F}_t)_{t \geq 0}$  generated by the coordinate-mapping process  $B$ ; i.e.,  $\mathcal{F}_t = \sigma(B_s, s \leq t)$  where  $B_t(\omega) = \omega_t$  for  $(\omega, t) \in C_0(\mathbb{R}_+) \times \mathbb{R}_+$ . Conversely, any adapted (optional) process induces a function (Borel function) on  $S$ . We recall that with respect to  $\mathbb{F}$ , any measurable adapted process is already optional (and even predictable) and refer to [29, Nos. IV.94–103, pp.145–152] or [6, Section 3] for further background on path spaces and their filtrations.

We equip  $C_0(\mathbb{R}_+)$  with the Wiener measure  $\mathbb{W}$  so that  $B$  is a standard Brownian motion with initial distribution  $B_0 \sim \delta_0$ . In what follows, probabilistic notions generally refer to the Wiener measure and the canonical filtration unless a different context is given. All metric spaces are equipped with their Borel  $\sigma$ -fields, (in)equalities of processes are to be understood up to evanescence (meaning that the projection of the exceptional set is  $\mathbb{W}$ -null) and (in)equalities of random variables are in the almost-sure sense.

Consider a centered distribution  $\nu$  on  $\mathbb{R}$  with finite second moment (but see Appendix A for a generalization to finite first moment). Let  $\mathcal{T}$  be the set of all a.s. finite  $\mathbb{F}$ -stopping times and let  $\mathcal{T}(\nu)$  be the subset of all  $\tau \in \mathcal{T}$  such that  $E[\tau] < \infty$  and  $B_\tau \sim \nu$ . The set  $\mathcal{T}(\nu)$  is nonempty; a number of classical embeddings  $\tau \in \mathcal{T}(\nu)$  can be found in [71]. The optimal Skorokhod embedding problem with respect to stopping times is

$$\mathbf{S}_{\mathcal{T}}(G) = \sup_{\tau \in \mathcal{T}(\nu)} E[G_\tau].$$

Here and below, outer integrals are used whenever the integrand is not measurable. The optimal Skorokhod embedding problem is often discussed with respect to randomized stopping times, defined as follows.

**Definition 2.1.** A probability measure  $\xi$  on  $C_0(\mathbb{R}_+) \times \mathbb{R}_+$  with disintegration  $\xi(d\omega, dt) = \mathbb{W}(d\omega)\xi_\omega(dt)$  is a *randomized stopping time* if  $\xi_\omega(\mathbb{R}_+) = 1$  for almost all  $\omega$  and  $\omega \mapsto \xi_\omega([0, t])$  is  $\mathcal{F}_t$ -measurable for all  $t \geq 0$ . We denote the set of randomized stopping times by  $\mathcal{R}$ .

We emphasize that our randomized stopping times are defined to stop in finite time. We can embed  $\mathcal{T}$  in  $\mathcal{R}$  in a canonical way:  $\tau$  is mapped to the randomized stopping time  $\xi^\tau$  with kernel  $\xi_\omega^\tau := \delta_{\tau(\omega)}$ . The image of  $\mathcal{T}$  under this embedding is denoted by  $\mathcal{R}_{\mathcal{T}}$ ; we will refer to its elements as *non-randomized stopping times*. We note the analogy between  $\mathcal{R}$  and Kantorovich transports on the one hand versus  $\mathcal{R}_{\mathcal{T}}$  and Monge transports on the other.

**Definition 2.2.** The set  $\mathcal{R}(\nu)$  consists of all randomized stopping times  $\xi \in \mathcal{R}$  such that  $\xi(\mathbf{t}) < \infty$  and  $\xi \circ B^{-1} = \nu$ . We write  $\mathcal{R}_{\mathcal{T}}(\nu)$  for the subset of non-randomized stopping times.

Here  $\mathbf{t}$  is the projection given by  $\mathbf{t}(\omega, t) = t$  and the two conditions correspond to a finite first moment and the marginal constraint  $\nu$ . In particular, if  $\xi = \xi^\tau$  represents a stopping time, then  $\xi(\mathbf{t}) = E[\tau]$  and  $\xi \circ B^{-1}$  is the law of  $B_\tau$ . The optimal Skorokhod embedding problem is then given by

$$\mathbf{S}(G) = \sup_{\xi \in \mathcal{R}(\nu)} \xi(G)$$

where again  $\xi(G) := \int G d\xi$ . More briefly, we will also call this the *primal problem*.

It is an obvious question, not previously addressed in the literature, to give a general condition under which the two formulations of the optimal Skorokhod embedding problem have the same value. The answer we provide is: whenever the reward function  $G$  has sufficiently regular paths.

**Theorem 2.3.** *Let  $G : C_0(\mathbb{R}_+) \times \mathbb{R}_+ \rightarrow [0, \infty)$  be Borel and adapted, and let  $t \mapsto G_t(\omega)$  be lower semicontinuous for all  $\omega \in C_0(\mathbb{R}_+)$ . Then*

$$\sup_{\xi \in \mathcal{R}(\nu)} \xi(G) = \sup_{\tau \in \mathcal{T}(\nu)} E[G_\tau].$$

The proof is stated in the next section; it is a consequence of Theorem 3.1 below and the proof of the latter will also help to understand where and how the regularity of  $G$  comes into play. Theorem 2.3 should be contrasted with results in unconstrained optimal stopping where value functions over stopping times and randomized stopping times agree for general measurable reward functions; see e.g. [43, Theorem 2.1]. In particular, the following (well-known) example shows that lower semicontinuity of  $G$  is an important assumption in Theorem 2.3: the nonstandard constraint given by the marginal  $\nu$  markedly changes the character of the question, as will be discussed in more detail in Section 3.

**Example 2.4.** Let  $\nu$  have an atom of mass  $a \in (0, 1)$  at the origin and let  $G$  be the bounded upper semicontinuous function  $G(\omega, s) = \mathbf{1}_{\{0\}}(s)$ . Then

$$\sup_{\xi \in \mathcal{R}(\nu)} \xi(G) = a > 0 = \sup_{\tau \in \mathcal{T}(\nu)} E[G_\tau].$$

Indeed, any  $\tau \in \mathcal{T}$  satisfies  $\mathbb{W}\{\tau = 0\} \in \{0, 1\}$  by the Blumenthal 0-1 law. Hence, any  $\tau \in \mathcal{T}(\nu)$  must be strictly positive a.s. which entails that  $E[G_\tau] = 0$ . On the other hand, we can find  $\xi \in \mathcal{R}(\nu)$  with  $\xi(\{0\}) = a$  and any such  $\xi$  attains the supremum.

### 3 Approximation of Randomized Stopping Times with Fixed Marginal

The main aim of this section is a density result with respect to the weak topology on  $C_0(\mathbb{R}_+) \times \mathbb{R}_+$  (induced by the continuous bounded functions on that space, as usual). We recall that  $\mathcal{R}_{\mathcal{T}}(\nu)$  is the embedding of  $\mathcal{T}(\nu)$  in  $\mathcal{R}$ .

**Theorem 3.1.** *Let  $\nu$  be a centered probability on  $\mathbb{R}$  with finite second moment. Then  $\mathcal{R}_{\mathcal{T}}(\nu) \subseteq \mathcal{R}(\nu)$  is dense for the topology of weak convergence.*

Again, this should be compared with classical results on the convergence of (convex combinations of) stopping times to randomized ones, such as [5, 27, 34, 37]. Such approximations may not, in general, respect the constraint given by the marginal  $\nu$ . In fact, in contrast to the unconstrained setting, the extreme points of  $\mathcal{R}(\nu)$  are not necessarily contained in  $\mathcal{R}_{\mathcal{T}}(\nu)$  (a counterexample can be deduced from [6, Example 6.19]).

*Proof of Theorem 2.3.* When  $G$  is bounded, the claim is a direct consequence of Theorem 3.1 and the characterization of weak convergence in [55, Corollary 2.9 and Proposition 2.11]. (The main point of [55] is that since we are dealing with measures that all have the same first marginal  $\mathbb{W}$ , weak convergence implies convergence under bounded test functions that are continuous in  $t$  but merely measurable in  $\omega$ .) The result for general  $G$  now follows by monotone approximation.  $\square$

Our proof of Theorem 3.1, presented in the remainder of this section, is constructive and gives direct insight why a singularity at the origin (as in Example 2.4) is an obstruction. One result to be established as part of the proof is that for randomized stopping times  $\xi$  and reward functions  $G$  that do not depend on an initial segment of the paths (in a sense to be made precise), the expectation  $\xi(G)$  can be exactly replicated by a stopping time  $\tau \in \mathcal{T}(\nu)$ , without any need for approximation (cf. Proposition 3.9 below).

#### 3.1 Proof of Theorem 3.1

Our first aim is to formalize and show that any  $\xi \in \mathcal{R}(\nu)$  can be approximated by randomized stopping times that do not stop right after time 0. We denote by  $|\cdot|$  the Euclidean norm in any dimension.

**Definition 3.2.** Let  $\eta > 0$  and  $\tau_\eta = \inf\{t : |(t, \omega_t)| \geq \eta\}$ . We say that a randomized stopping time  $\xi$  is *bounded away from 0 with lower bound  $\eta > 0$*

if

$$\xi_\omega([\tau_\eta(\omega), \infty)) = 1 \quad \text{for almost all } \omega \in C_0(\mathbb{R}_+).$$

The set of all such  $\xi$  is denoted  $\mathcal{R}^\eta$ . We also set  $\mathcal{R}^+ = \cup_{\eta>0} \mathcal{R}^\eta$ ; any  $\xi \in \mathcal{R}^+$  is called *bounded away from 0*. Finally,  $\mathcal{R}^\eta(\nu) := \mathcal{R}^\eta \cap \mathcal{R}(\nu)$  and  $\mathcal{R}^+(\nu) := \mathcal{R}^+ \cap \mathcal{R}(\nu)$ .

The terminology of lower bound is convenient but slightly abusive: intuitively,  $\xi \in \mathcal{R}^\eta$  above means that  $\xi$  happens after time  $\tau_\eta$  (but the time  $\tau_\eta$  is not bounded away from zero in the usual sense). Some more notation will be useful.

**Definition 3.3.** Let  $\omega, \omega' \in C_0(\mathbb{R}_+)$ ,  $t \in \mathbb{R}_+$ ,  $\tau \in \mathcal{T}$  and  $\xi \in \mathcal{R}$ .

(i) The path of  $\omega$  and  $\omega'$  *glued at time  $t$*  is

$$(\omega \oplus_t \omega')(s) := \omega(s \wedge t) + \omega'(s - t \vee 0), \quad s \geq 0.$$

(ii) The path of  $\omega$  *after time  $t$*  is

$$\omega^{t \rightarrow}(s) := \omega(t + s) - \omega(t), \quad s \geq 0.$$

(iii) The randomized stopping time  $\xi$  *shifted by  $\tau$* , denoted  $\tau \oplus \xi$ , is defined by its kernel

$$(\tau \oplus \xi)_\omega([0, t]) := \mathbf{1}_{\tau(\omega) < t} \xi_{\omega^{\tau(\omega) \rightarrow}}([0, t - \tau(\omega)]), \quad t \geq 0.$$

The definition in (iii) can be understood as the randomized stopping time  $\xi$  applied to the Brownian motion started at  $(\tau, B_\tau)$ . For instance, if  $\tau = t_0$  is deterministic and  $\xi = \xi^\sigma$  corresponds to a nonrandomized stopping time  $\sigma > 0$ , then  $\tau \oplus \xi$  corresponds to the stopping time  $(t_0 \oplus \sigma)(\omega) = t_0 + \sigma(\omega^{t_0 \rightarrow})$ .

**Lemma 3.4.** For  $0 < \eta < 1$ , define

$$\rho'_\eta(\omega) := \inf\{t : |\omega_t| = \eta\}, \quad \rho_\eta(\omega) := \inf\{t \geq \rho'_\eta(\omega) : |\omega_t| \in \{0, \sqrt{\eta}\}\}.$$

For almost all  $\omega \in C_0(\mathbb{R}_+)$ , we have

(i)  $\rho_\eta(\omega) \rightarrow 0$  as  $\eta \rightarrow 0$ ,

(ii)  $d((\omega^{\rho_\eta(\omega) \rightarrow}, t), (\omega, t)) \rightarrow 0$  as  $\eta \rightarrow 0$ , for all  $t \geq 0$ .

*Proof.* We show that (i), (ii) hold on the set  $I$  of all paths  $\omega \in C_0(\mathbb{R}_+)$  that are not initially constant; i.e.,  $\omega|_{[0,\varepsilon]} \not\equiv 0$  for all  $\varepsilon > 0$ . Notice that  $\mathbb{W}(I) = 1$ .

(i) If  $\omega \in C_0(\mathbb{R}_+)$  is such that  $\rho_\eta(\omega)$  does not converge to 0, we can find  $\varepsilon > 0$  and a sequence  $\eta_n \rightarrow 0$  such that  $\rho_{\eta_n}(\omega) \geq \varepsilon$  for all  $n$ . In particular, this implies that  $\sup_{s \leq \varepsilon} |\omega(s)| \leq \sqrt{\eta_n}$  for all  $n$  and therefore  $\omega|_{[0,\varepsilon]} \equiv 0$ ; that is,  $\omega \notin I$ .

(ii) Fix  $\omega \in I$  and  $t > 0$ . Let  $\eta > 0$  be small enough so that  $\rho_\eta(\omega) < t$  and consider some  $0 < s < t$ . For  $s \leq \rho_\eta(\omega)$  we clearly have

$$|\omega^{\rho_\eta(\omega) \mapsto}(s) - \omega(s)| \leq \sup_{u \leq 2\rho_\eta(\omega)} 3|\omega(u)|.$$

Whereas for  $\rho_\eta(\omega) < s \leq t$  we have

$$|\omega^{\rho_\eta(\omega) \mapsto}(s) - \omega(s)| \leq \sqrt{\eta} + \sup_{s \leq t} |\omega(\rho_\eta(\omega) + s) - \omega(s)|.$$

Combining these two inequalities, we obtain that

$$\sup_{s \leq t} |\omega^{\rho_\eta(\omega) \mapsto}(s) - \omega(s)| \leq \sup_{u \leq 2\rho_\eta(\omega)} 3|\omega(u)| + \sqrt{\eta} + \sup_{s \leq t} |\omega(\rho_\eta(\omega) + s) - \omega(s)|.$$

Since  $\omega$  is uniformly continuous on compact intervals, the right-hand side tends to 0 when  $\rho_\eta(\omega) \rightarrow 0$ , and the latter holds by (i) as  $\eta \rightarrow 0$ .  $\square$

**Remark 3.5.** If  $G \in C_b(C_0(\mathbb{R}_+) \times \mathbb{R}_+)$  and  $\xi \in \mathcal{R}$ , then  $\xi(G) = \xi(G(B_{\cdot \wedge t}, t))$  by the adaptedness property of  $\xi \in \mathcal{R}$ . As a result, the weak convergence on  $\mathcal{R}$  is also induced by the subset of adapted functions  $G$  which, in turn, can be identified with  $C_b(S)$ .

We can now show that any  $\xi \in \mathcal{R}(\nu)$  can be approximated with embeddings that are bounded away from 0—except in the trivial case  $\nu = \delta_0$  where the assertion clearly fails.

**Proposition 3.6.** *Let  $\nu \neq \delta_0$ . Then  $\mathcal{R}^+(\nu)$  is weakly dense in  $\mathcal{R}(\nu)$ .*

*Proof.* As  $\nu \neq \delta_0$ , its potential function  $x \mapsto u_\nu(x) := \int |x - y| \nu(dy)$  satisfies  $u_\nu(0) > 0$ , and then by continuity of  $u_\nu$  we even have  $\min_{y \in [-\sqrt{\eta}, \sqrt{\eta}]} u_\nu(y) \geq \sqrt{\eta}$  for  $\eta > 0$  small enough. This shows that

$$\mu_\eta := \frac{1}{2}(\delta_{-\sqrt{\eta}} + \delta_{\sqrt{\eta}}) \leq_c \nu$$

where  $\leq_c$  denotes the convex order. As a result, for  $\eta > 0$  small enough we can embed  $\nu$  in a Brownian motion with initial distribution  $\mu_\eta$  by a stopping time  $\chi_\eta$  with  $E[\chi_\eta] < \infty$ . (See [71] for the relevant background.)

Given  $\xi \in \mathcal{R}(\nu)$ , let  $\rho_\eta$  be as in Lemma 3.4 and define a family of randomized stopping times  $\xi^\eta \in \mathcal{R}$  via their disintegration,

$$\xi_\omega^\eta := \mathbf{1}_{\{\omega(\rho_\eta(\omega))=0\}}(\rho_\eta \oplus \xi)_\omega + \mathbf{1}_{\{|\omega(\rho_\eta(\omega))|=\sqrt{\eta}\}}(\rho_\eta \oplus \chi_\eta)_\omega.$$

Then we have  $\xi^\eta \in \mathcal{R}^\eta(\nu)$  by construction. Next, we show that  $\xi^\eta \rightarrow \xi$  weakly as  $\eta \rightarrow 0$ . Let  $G : C_0(\mathbb{R}_+) \times \mathbb{R}_+ \rightarrow \mathbb{R}$  be bounded, continuous and adapted, then

$$\begin{aligned} \xi(G) &= \int_{C_0(\mathbb{R}_+)} \int_{\mathbb{R}_+} G(\omega, s) \xi_\omega(ds) \mathbb{W}(d\omega) \\ &= \int_{C_0(\mathbb{R}_+)} \left[ \int_{C_0(\mathbb{R}_+)} \int_{\mathbb{R}_+} G(\omega, s) \xi_\omega(ds) \mathbb{W}(d\omega) \right] \mathbb{W}(d\omega'). \end{aligned}$$

Using the stationarity and independence of Brownian increments, we can similarly write the expectation  $\xi^\eta(G)$  as

$$\begin{aligned} &\int_{C_0(\mathbb{R}_+)} \left[ \mathbf{1}_{\{\omega'(\rho_\eta(\omega'))=0\}} \int_{C_0(\mathbb{R}_+)} \int_{\mathbb{R}_+} G(\omega' \oplus_{\rho_\eta(\omega')} \omega, s + \rho_\eta(\omega')) \xi_\omega(ds) \mathbb{W}(d\omega) \right. \\ &\quad \left. + \mathbf{1}_{\{\omega'(\rho_\eta(\omega')) \neq 0\}} \int_{C_0(\mathbb{R}_+) \times \mathbb{R}_+} G(\omega' \oplus_{\rho_\eta(\omega')} \omega, s + \rho_\eta(\omega')) \chi_\eta(d\omega, ds) \right] \mathbb{W}(d\omega'). \end{aligned}$$

Therefore,

$$\begin{aligned} &|\xi(G) - \xi^\eta(G)| \\ &\leq \int_{C_0(\mathbb{R}_+)} \left[ \int_{C_0(\mathbb{R}_+)} \int_{\mathbb{R}_+} |G(\omega, s) - G(\omega' \oplus_{\rho_\eta(\omega')} \omega, s + \rho_\eta(\omega'))| \times \right. \\ &\quad \left. \xi_\omega(ds) \mathbb{W}(d\omega) \right] \mathbb{W}(d\omega') + 2\|G\|_\infty \mathbb{W}\{\omega' : \omega'(\rho_\eta(\omega')) \neq 0\}. \end{aligned}$$

We note that  $\mathbb{W}\{\omega' : \omega'(\rho_\eta(\omega')) \neq 0\} = \sqrt{\eta} \rightarrow 0$  as a consequence of the martingale property.

On the other hand, Lemma 3.4 yields that  $\rho_\eta(\omega') \rightarrow 0$  for all  $\omega'$  outside a nullset, and for such  $\omega'$  the second part of that lemma and the continuity of  $G$  imply that

$$|G(\omega, s) - G(\omega' \oplus_{\rho_\eta(\omega')} \omega, s + \rho_\eta(\omega'))| \rightarrow 0$$

for all  $\omega$ . By bounded convergence, we conclude that  $|\xi^\eta(G) - \xi(G)| \rightarrow 0$  for  $\eta \rightarrow 0$  as desired.  $\square$

For brevity, let us write  $C_b$  for the set of bounded continuous functions on  $C_0(\mathbb{R}_+) \times \mathbb{R}_+$ . Next, we introduce the subset of test functions which are independent of the initial segment of the path, or more precisely, their value depends on the path  $\omega$  before time  $\tau_\eta$  only through the level  $\omega_{\tau_\eta(\omega)}$  at that time.

**Definition 3.7.** Given  $\eta > 0$ , we write  $C_b^\eta$  for the set of all  $G \in C_b$  with the property that if  $\omega, \omega' \in C_0(\mathbb{R}_+)$  satisfy

$$\tau_\eta(\omega) = \tau_\eta(\omega') =: t_0 \text{ and } \omega|_{[t_0, \infty)} = \omega'|_{[t_0, \infty)}, \quad \text{then } G(\omega, t) = G(\omega', t)$$

for all  $t \geq 0$ . We set  $C_b^+ := \cup_{\eta > 0} C_b^\eta$ .

This definition is chosen so that the monotonicity property

$$C_b^\eta \supseteq C_b^{\eta'} \quad \text{for } \eta \leq \eta' \tag{3.1}$$

holds. In particular,  $C_b^+$  is the increasing limit of  $C_b^\eta$  as  $\eta \rightarrow 0$ .

**Lemma 3.8.** *The weak topology on  $\mathcal{R}(\nu)$  induced by  $C_b^+$  coincides with the usual weak topology.*

*Proof.* Let  $G \in C_b$ ; then  $G_k(\omega, t) := G(\omega^{\tau_{1/k}(\omega) \mapsto}, t)$  defines a function belonging to  $C_b^{1/k} \subseteq C_b^+$ . We have  $d((\omega^{\tau_{1/k}(\omega) \mapsto}, t), (\omega, t)) \rightarrow 0$  similarly as in Lemma 3.4 and as  $(G_k)$  is uniformly bounded this implies that  $\xi(G_k) \rightarrow \xi(G)$  for all  $\xi \in \mathcal{R}(\nu)$ . As a result,  $C_b^+$  separates the points of  $\mathcal{R}(\nu)$ ; i.e., the induced weak topology  $\mathcal{T}^+$  is Hausdorff. Since  $\mathcal{R}(\nu)$  is compact in the usual weak topology  $\mathcal{T}$  (see for instance [6, Theorem 3.14]) and  $\mathcal{T}^+ \subseteq \mathcal{T}$ , this already implies the result. Indeed, the identity map from  $\mathcal{T}$  to  $\mathcal{T}^+$  is continuous and maps compacts to compacts, hence is a homeomorphism.  $\square$

A key insight is that for  $G \in C_b^+$ , the expectation under a randomized stopping time that is suitably bounded away from zero is (exactly) replicated by a non-randomized stopping time. In particular, *the necessity for approximation when  $G \in C_b$  can be attributed to the initial portion of the paths.*

**Proposition 3.9.** *Let  $\eta > 0$  and  $\xi \in \mathcal{R}^\eta(\nu)$ . Then there exists a non-randomized stopping time  $\bar{\xi} \in \mathcal{R}^\eta(\nu)$  such that  $\bar{\xi}(G) = \xi(G)$  for all  $G \in C_b^\eta$ .*

Before stating the proof, let us show how to combine the above results to conclude Theorem 3.1.

*Proof of Theorem 3.1.* The case  $\nu = \delta_0$  is trivial; we assume that  $\nu \neq \delta_0$ . Fix  $\xi \in \mathcal{R}(\nu)$ , then by Proposition 3.6 we can find  $\xi^n \in \mathcal{R}^{1/n}(\nu)$  such that  $\xi^n \rightarrow \xi$ . Next, we use Proposition 3.9 to find corresponding non-randomized stopping times  $\bar{\xi}^n \in \mathcal{R}^{1/n}(\nu)$ . Let  $G \in C_b^+$ , then  $G \in C_b^\eta$  for some  $\eta > 0$  and for all  $n \geq 1/\eta$  we have by (3.1) that  $\bar{\xi}^n(G) = \xi^n(G)$ . In particular, Lemma 3.8 implies that  $(\bar{\xi}^n)$  and  $(\xi^n)$  have the same weak limit; that is,  $\bar{\xi}^n \rightarrow \xi$ .  $\square$

## 3.2 Proof of Proposition 3.9

The basic idea for this proof is to use the initial segments of the paths as a randomization device which, since we are only testing with  $G \in C_b^\eta$ , does not affect the evaluation of  $G$ . We first state two auxiliary results.

Let  $\lambda$  be the Lebesgue measure on  $[0, 1]$ . We consider the product space  $\bar{C}_0(\mathbb{R}_+) := C_0(\mathbb{R}_+) \times [0, 1]$  equipped with the product  $\sigma$ -field  $\bar{\mathcal{F}} = \mathcal{F} \otimes \mathcal{B}([0, 1])$ , the product measure  $\bar{\mathbb{W}} = \mathbb{W} \otimes \lambda$  and the product filtration  $\bar{\mathbb{F}}$ . Let  $\bar{B}$  be the process defined by  $t \mapsto \bar{B}_t(\omega, u) = \omega_t$  for  $(\omega, u) \in \bar{C}_0(\mathbb{R}_+)$ ; note that  $\bar{B}$  is a Brownian motion under  $\bar{\mathbb{W}}$ . The following is well known (see for instance the proof of [6, Theorem 3.8] and the subsequent lemma).

**Lemma 3.10.** *Let  $\xi \in \mathcal{R}$  have disintegration  $\xi = \mathbb{W}(d\omega)\xi_\omega(ds)$ . There exists an  $\bar{\mathbb{F}}$ -stopping time  $\rho$  such that*

$$\rho(\omega, u) = \inf\{t \geq 0 : \xi_\omega([0, t]) \geq u\} \quad \text{for a.e. } (\omega, u) \in \bar{C}_0(\mathbb{R}_+) \quad (3.2)$$

and  $\bar{\mathbb{W}} \circ (\bar{B}, \rho)^{-1} = \xi$ ; that is,  $E^{\bar{\mathbb{W}}}[G(\bar{B}, \rho)] = \xi(G)$  for every  $G \in C_b$ .

The second auxiliary result concerns the ‘‘internal’’ randomization device that will be used in lieu of the external randomization  $([0, 1], \lambda)$  as in the preceding lemma. This is somewhat involved because the randomization needs to be implemented conditionally on the level  $B_{\tau^\eta}(\omega)$ —indeed,  $G \in C_b^\eta$  is independent of how a path reached this level at time  $\tau^\eta(\omega)$ , but of course not of the level itself.

Fix  $\eta \geq 0$ . We note that the level  $h(\omega) := B_{\tau^\eta}(\omega)$  of a path  $\omega$  at the (almost surely finite) time  $\tau^\eta$  satisfies  $h(\omega) \in (-\sqrt{\eta}, \sqrt{\eta})$ . Moreover,

$$\tau^\eta(\omega) = \sqrt{\eta^2 - h(\omega)^2}$$

depends on  $\omega$  only through  $h = h(\omega)$ . Given  $h \in \mathbb{R}$ , we introduce the set

$$C_h := \{\omega \in C_0(\mathbb{R}_+) : \tau^\eta(\omega) < \infty, B_{\tau^\eta}(\omega) = h\};$$

we think of  $C_h$  as a set of initial segments of paths (since the path after  $\tau_\eta$  will not be used). Given  $f \in C_h$  and  $\omega \in C_0(\mathbb{R}_+)$ , we set

$$f \oplus \omega := f \oplus_{\tau_\eta(f)} \omega$$

for brevity. We also denote by  $\mathbb{W}_h$  the conditional law of  $B$  given  $B_{\tau_\eta} = h$ . That is,  $\mathbb{W}_h$  is a stochastic kernel on  $(-\sqrt{\eta}, \sqrt{\eta}) \times C_0(\mathbb{R}_+)$  such that

$$\mathbb{W}[A|B_{\tau_\eta} = h] = \mathbb{W}_h(A) = \mathbb{W}_h(A \cap C_h)$$

for  $h \in (-\sqrt{\eta}, \sqrt{\eta})$  and  $A \in \mathcal{B}(C_0(\mathbb{R}_+))$ . In particular,

$$\mathbb{W}(A) = \int_{\mathbb{R}} \mathbb{W}_h(A) \mu(dh), \quad \mu := \text{Law}(B_{\tau_\eta}).$$

**Lemma 3.11.** *The measure  $\mathbb{W}_h$  is atomless for all  $h \in (-\sqrt{\eta}, \sqrt{\eta})$ .*

*Proof.* Consider the map  $\Phi : C_0(\mathbb{R}_+) \rightarrow C_0(\mathbb{R}_+)$  given by  $\Phi(\omega) = \omega^{\tau_\eta(\omega) \mapsto \cdot}$ . By the stationarity and independence of Brownian increments, the pushforward  $\mathbb{W}_h \circ \Phi^{-1}$  is the Wiener measure and in particular atomless. As a consequence,  $\mathbb{W}_h$  is atomless, for if  $\mathbb{W}_h$  had an atom then any pushforward would also have an atom.  $\square$

**Lemma 3.12.** *Let  $(P_y(dz))$  be a stochastic kernel  $(Y, \mathcal{Y}) \times (Z, \mathcal{B}(Z)) \rightarrow [0, 1]$  where  $Z$  is a Polish space and  $(Y, \mathcal{Y})$  is a measurable space. If  $P_y$  is atomless for all  $y \in Y$ , there exists a jointly measurable map  $(y, z) \mapsto \phi_y(z) \in [0, 1]$  such that  $P_y \circ \phi_y^{-1} = \lambda$  for all  $y \in Y$ .*

*Proof.* Recall that any two Polish spaces of uncountable cardinality are homeomorphic as Borel spaces; cf. [72, Theorem 2.12, p.14]. As atomless measures can exist only on uncountable spaces, we deduce that there is a Borel homeomorphism  $\Phi : (Z, \mathcal{B}(Z)) \rightarrow ([0, 1], \mathcal{B}([0, 1]))$ . Consider  $Q_y = P_y \circ \Phi^{-1}$ ; then  $(Q_y)$  are atomless probability measures; i.e., their c.d.f.'s  $F_y(x) := Q_y((-\infty, x])$  are continuous in  $x$ . By construction, they are also measurable in  $y$ . In particular, as Caratheodory functions, they are jointly measurable in  $(x, y)$ ; cf. [2, Lemma 4.51, p.153]. Finally, recall that if  $F$  is the c.d.f. of a random variable  $X$  with continuous distribution, then  $F(X) \sim \lambda$ . As a result,  $\phi_y = F_y \circ \Phi$  satisfies the requirement of the lemma.  $\square$

Combining the two preceding lemmas, we obtain the following.

**Corollary 3.13.** *There exists a jointly Borel measurable map*

$$C_0(\mathbb{R}_+) \times (-\sqrt{\eta}, \sqrt{\eta}) \rightarrow [0, 1], \quad (\omega, h) \mapsto \phi_h(\omega)$$

such that  $\mathbb{W}_h \circ \phi_h^{-1} = \lambda$  for each  $h$ .

We can now provide the proof of the proposition.

*Proof of Proposition 3.9.* Let  $\xi = \mathbb{W}(d\omega)\xi_\omega(ds)$  be a disintegration of the given randomized stopping time  $\xi \in \mathcal{R}^\eta(\nu)$  and define  $\bar{\xi} = \mathbb{W}(d\omega)\bar{\xi}_\omega(ds)$  through

$$\bar{\xi}_{f \oplus \omega} := \int_{C_h(f)} \xi_{g \oplus \omega} \mathbb{W}_{h(f)}(dg), \quad f, \omega \in C_0(\mathbb{R}_+).$$

Clearly  $\bar{\xi}_{f \oplus \omega}$  depends on  $f$  only through  $h(f)$ . We show below that  $\bar{\xi} \in \mathcal{R}^\eta(\nu)$  and  $\bar{\xi}(G) = \xi(G)$  for all  $G \in C_b^\eta$ . Admitting this for the moment, it remains to construct a non-randomized stopping time with the same law as  $\bar{\xi}$ . Following Lemma 3.10, we can associate a stopping time  $(\omega, u) \mapsto \rho(\omega, u)$  on the probability space  $(C_0(\mathbb{R}_+) \times [0, 1], \mathbb{W} \otimes \lambda)$  with  $\bar{\xi}$ , and we can choose a version of  $\rho$  such that  $\rho(f \oplus \omega, u)$  depends on  $f$  only through  $h(f)$ . Finally, let  $\phi_h$  be as in Lemma 3.13, then by construction,  $\tau(\omega) := \rho(\omega, \phi_{h(\omega)}(\omega))$  is a stopping time on  $C_0(\mathbb{R}_+)$  such that  $\mathbb{W} \circ (B, \tau)^{-1} = \bar{\mathbb{W}} \circ (\bar{B}, \rho)^{-1} = \bar{\xi}$ . Thus, its embedding  $\xi^\tau \in \mathcal{R}_\tau$  is the required non-randomized stopping time.

It remains to verify that  $\bar{\xi} \in \mathcal{R}^\eta(\nu)$  and  $\bar{\xi}(G) = \xi(G)$  for all  $G \in C_b^\eta$ . Let  $f \in C_h$  and  $\omega \in C_0(\mathbb{R}_+)$ ; then  $\bar{\xi}_{f \oplus \omega}$  is adapted and concentrated on  $[\tau_\eta(f), \infty)$  since  $\xi_{g \oplus \omega}$  has these properties for all  $g \in C_h$  (recall that  $\tau_\eta$  is constant on  $C_h$ ). This yields that  $\bar{\xi} \in \mathcal{R}^\eta$ . Next, let  $G \in C_b^\eta$ . Then  $\bar{\xi}(G)$  is equal to

$$\begin{aligned} & \int_{C_0(\mathbb{R}_+)} \int_{\mathbb{R}_+} G(\omega, s) \bar{\xi}_\omega(ds) \mathbb{W}(d\omega) \\ &= \int_{\mathbb{R}} \int_{C_0(\mathbb{R}_+)} \int_{C_h} \int_{\mathbb{R}_+} G(f \oplus \omega, s) \bar{\xi}_{f \oplus \omega}(ds) \mathbb{W}_h(df) \mathbb{W}(d\omega) \mu(dh) \\ &= \int_{\mathbb{R}} \int_{C_0(\mathbb{R}_+)} \int_{C_h} \int_{C_h} \int_{\mathbb{R}_+} G(f \oplus \omega, s) \xi_{g \oplus \omega}(ds) \mathbb{W}_h(dg) \mathbb{W}_h(df) \mathbb{W}(d\omega) \mu(dh) \\ &= \int_{\mathbb{R}} \int_{C_0(\mathbb{R}_+)} \int_{C_h} \int_{C_h} \int_{\mathbb{R}_+} G(g \oplus \omega, s) \xi_{g \oplus \omega}(ds) \mathbb{W}_h(dg) \mathbb{W}_h(df) \mathbb{W}(d\omega) \mu(dh) \\ &= \int_{C_0(\mathbb{R}_+)} \int_{\mathbb{R}_+} G(\omega, s) \xi_\omega(ds) \mathbb{W}(d\omega) = \xi(G). \end{aligned}$$

This also implies that  $\bar{\xi}(\mathbf{t}) = \xi(\mathbf{t}) < \infty$ , since  $\mathbf{t} \wedge n \in C_b^\eta$  for all  $n \in \mathbb{N}$ . Finally, to see that  $\xi$  and  $\bar{\xi}$  embed the same distribution  $\nu$ , we show that

$\bar{\xi}(\phi(B)) = \xi(\phi(B))$  for  $\phi \in C_b(\mathbb{R})$ . Indeed, consider  $G := \phi(B_{t \vee \tau_\eta}) \in C_b^\eta$  (recall that adaptedness was not required). Since we already know that  $\xi, \bar{\xi} \in \mathcal{R}^\eta$ , we have  $\xi(\phi(B)) = \xi(G) = \bar{\xi}(G) = \bar{\xi}(\phi(B))$  and the proof is complete.  $\square$

## 4 The Dual Problem

We first introduce the domain of the dual problem. To that end, recall that  $\nu$  is a centered distribution on  $\mathbb{R}$  with finite second moment and let  $J \subseteq \mathbb{R}$  be the smallest convex set with  $\nu(J) = 1$ . Thus, a boundary point of  $J$  is contained in  $J$  if and only if it is an atom of  $\nu$ . We fix<sup>1</sup> an increasing sequence  $(K_n)$  of compact intervals  $0 \in K_n \subseteq J$  whose union is  $J$  and let

$$T_n = \inf\{t \geq 0 : B_t \in \partial K_n\}$$

be the first hitting time of the boundary  $\partial K_n$ . Recall that probabilistic notions are understood with respect to the canonical space  $C_0(\mathbb{R}_+)$  and that  $S$  can be embedded in  $C_0(\mathbb{R}_+) \times \mathbb{R}$ . We fix a (not necessarily measurable) function  $G : S \rightarrow [0, \infty]$  and introduce the following.

**Definition 4.1.** Let  $\mathcal{D}(G)$  be the set of all pairs  $(M, \psi)$  where the Borel function  $\psi : J \rightarrow \mathbb{R} \cup \{\infty\}$  is in  $L^1(\nu)$  and  $M$  is a (continuous) local  $(\mathbb{W}, \mathbb{F})$ -martingale with  $M_0 = 0$  such that

$$M + \psi(B) \geq G \quad \text{on} \quad \cup_n [0, T_n] \quad (\text{up to evanescence})$$

and  $M_{\cdot \wedge T_n}$  is bounded below for all  $n$ . The *dual problem* is

$$\mathbf{I}(G) = \inf_{(M, \psi) \in \mathcal{D}(G)} \nu(\psi).$$

The continuity of  $M$  refers to its paths being a.s. continuous. We recall that Brownian local martingales always admit continuous versions and assume implicitly that all local martingales are continuous in what follows. We also note that  $\cup_n [0, T_n] = \mathbb{R}_+ \times \Omega$  if  $J = \mathbb{R}$ . In Section 7 it will be shown that the value of the dual problem  $\mathbf{I}(G)$  can change if  $M$  is restricted to true martingales or if  $\psi$  is restricted to continuous functions. As the relaxations in Definition 4.1 are novel, we detail some technical observations.

<sup>1</sup>The results below do not depend on the choice of  $K_n$ . When  $\nu$  has bounded support with atoms at both endpoints, we can simply take  $K_n = J$  for all  $n$ .

**Remark 4.2.** Setting  $T = \lim T_n = \inf\{t \geq 0 : B_t \in \partial J\}$ , we have  $\cup_n[0, T_n] = [0, T]$  if  $J$  is closed and  $\cup_n[0, T_n] = [0, T)$  if  $J$  is open, but in general either of the inclusions  $[0, T) \subseteq \cup_n[0, T_n] \subseteq [0, T]$  can be strict and the value of  $\mathbf{I}(G)$  can change if  $\cup_n[0, T_n]$  is replaced by  $[0, T)$  in Definition 4.1. One can, however, replace  $\cup_n[0, T_n]$  by  $[0, T]$  without altering the value of  $\mathbf{I}(G)$ , since given  $(M, \psi) \in \mathcal{D}(G)$  we may extend  $\psi$  by setting  $\psi = \infty$  on  $\partial J \setminus J$  without affecting  $\nu(\psi)$ . (We have nevertheless found it less confusing to write  $\cup_n[0, T_n]$  everywhere.)

**Remark 4.3.** Suppose that  $G$  is Borel measurable; i.e.,  $G$  can be seen as an optional process. Then the inequality  $M + \psi(B) \geq G$  on  $\cup_n[0, T_n]$  up to evanescence is equivalent to the almost-sure inequality  $M_\tau + \psi(B_\tau) \geq G_\tau$  for all stopping times  $\tau$  with  $\tau \leq T_n$  for some  $n$ . This follows from the optional cross-section theorem; cf. [29, Theorem IV.84, p. 137].

**Remark 4.4.** Using Fatou's lemma, boundedness from below of  $M_{\cdot \wedge T_n}$  implies that for all stopping times  $\sigma \leq \tau \leq T_n$ , the random variables  $M_\sigma, M_\tau$  are integrable and satisfy the optional sampling property  $E[M_\tau | \mathcal{F}_\sigma] \leq M_\sigma$ .

**Lemma 4.5.** *Let  $(M, \psi) \in \mathcal{D}(G)$ . (i) We have  $\mathbb{W}\{\tau \leq T_n\} \rightarrow 1$  and  $M + \psi(B) \geq G$  on  $[0, \tau]$  for all  $\tau \in \mathcal{T}(\nu)$ . (ii) We have  $M + \psi(B) \geq G$   $\xi$ -a.s. for all  $\xi \in \mathcal{R}(\nu)$ .*

*Proof.* (i) Let  $\tau \in \mathcal{T}(\nu)$ ; in particular,  $E[\tau] < \infty$ . Since  $B_{\cdot \wedge \tau}$  is a uniformly integrable martingale and  $B_\tau$  is supported on  $J$ , we have  $\mathbb{W}\{\tau \leq T_n\} \rightarrow 1$  and hence  $[0, \tau] \subseteq \cup_n[0, T_n]$ . In particular,  $M + \psi(B) \geq G$  on  $[0, \tau]$  up to evanescence.

(ii) Consider the stopping time  $\rho$  associated with  $\xi$  on the extended space  $\bar{C}_0(\mathbb{R}_+)$ ; cf. Lemma 3.10. As  $\bar{\mathbb{W}} \circ (\bar{B}, \rho)^{-1} = \xi$ , we can deduce the claim by applying the arguments for (i) to  $\rho$  and  $\bar{B}$ .  $\square$

**Lemma 4.6.** *Let  $(M, \psi) \in \mathcal{D}(0)$  and write  $\psi^{**}$  for the convex envelope on  $J$ . Then  $(M, \psi^{**}) \in \mathcal{D}(0)$ ; i.e.,*

$$M + \psi^{**}(B) \geq 0 \quad \text{on} \quad \cup_n [0, T_n], \quad (4.1)$$

and in particular  $\psi^{**}(0) \geq 0$ .

*Proof.* We first observe that for  $x \in J$ ,

$$\psi^{**}(x) = \inf \left\{ \frac{b}{a+b} \psi(x-a) + \frac{a}{a+b} \psi(x+b) : a, b \in \mathbb{R}_+, [x-a, x+b] \subseteq J \right\}$$

where the sum is understood as  $\psi(x)$  if  $a = b = 0$ . Suppose that (4.1) fails, then by the optional cross-section theorem there exist  $n \geq 1$  and a stopping time  $\sigma \leq T_n$  such that

$$\mathbb{W}\{M_\sigma + \psi^{**}(B_\sigma) < 0\} > 0.$$

By the above formula for  $\psi^{**}(x)$  and a measurable selection argument, it follows that there exist  $m \geq n$  and  $\mathcal{F}_\sigma$ -measurable random variables  $a, b \geq 0$  with  $[B_\sigma - a, B_\sigma + b] \subseteq K_m$  such that

$$\mathbb{W}\left\{M_\sigma + \frac{b}{a+b}\psi(B_\sigma - a) + \frac{a}{a+b}\psi(B_\sigma + b) < 0\right\} > 0. \quad (4.2)$$

Let  $\tau = \inf\{t \geq \sigma : B_t - B_\sigma \notin (-a, b)\}$  and note that  $\tau \leq T_m$ . Then optional sampling (cf. Remark 4.4) implies

$$M_\sigma + \frac{b}{a+b}\psi(B_\sigma - a) + \frac{a}{a+b}\psi(B_\sigma + b) \geq E[M_\tau + \psi(B_\tau)|\mathcal{F}_\sigma].$$

But  $(M, \psi) \in \mathcal{D}(0)$  implies that  $M_\tau + \psi(B_\tau) \geq 0$ , and now a contradiction to (4.2) ensues. In particular,  $\psi^{**}(0) \geq 0$ . Hence, the convex function  $\psi^{**}$  is bounded from below by a linear function and from above by  $\psi$ , so that  $\psi^{**} \in L^1(\nu)$ .  $\square$

The following normalization will be used repeatedly below.

**Remark 4.7.** Let  $(M, \psi) \in \mathcal{D}(G)$  and  $c \in \mathbb{R}$ . Define  $M' = M + cB$  and  $\psi'(x) = \psi(x) - cx$ . Then  $(M', \psi') \in \mathcal{D}(G)$  since  $M' + \psi'(B) = M + \psi(B)$  and  $B$  is a bounded martingale on  $[0, T_n]$  with  $B_0 = 0$ .

Using this with the (left, say) derivative  $c := \partial^- \psi^{**}(0)$  and recalling that  $\psi^{**}(0) \geq 0$  by Lemma 4.6, we see that any  $(M, \psi) \in \mathcal{D}(G)$  can be normalized such that  $\psi^{**} \geq 0$  and hence also  $\psi \geq 0$ .

Next, we establish the key inequality for the “weak” duality and in particular that our definition of the dual domain is rigid enough despite the relaxations.

**Lemma 4.8.** *Let  $(M, \psi) \in \mathcal{D}(0)$ . (i) We have  $E[M_\tau] \leq 0$  for all  $\tau \in \mathcal{T}(\nu)$ . (ii) We have  $\xi(M) \leq 0$  for all  $\xi \in \mathcal{R}(\nu)$ .*

*Proof.* (i) We prove the claim for any stopping time  $\tau$  such that  $B_{\cdot \wedge \tau}$  is uniformly integrable,  $E[\psi(B_\tau)] < \infty$  and  $[0, \tau] \subseteq \cup_n [0, T_n]$ ; in particular, these are satisfied for  $\tau \in \mathcal{T}(\nu)$ . Indeed, by Remark 4.7, we can assume without loss of generality that  $\psi \geq 0$  and then  $0 \leq \psi^{**} \leq \psi$ . Since  $\psi^{**}$  is convex and  $\psi^{**}(B_\tau)$  is integrable, it follows that  $\psi^{**}(B_{\cdot \wedge \tau})$  is a nonnegative

uniformly integrable submartingale, hence of class (D). On the other hand, Lemma 4.6 yields that  $M \geq -\psi^{**}(B)$  on  $[0, \tau]$ , so we conclude that  $M_{\cdot \wedge \tau}^-$  is of class (D). Fatou's lemma (in its version with a uniformly integrable lower bound) now implies that  $E[M_\tau] \leq 0$ .

(ii) Let  $\xi \in \mathcal{R}(\nu)$  and let  $\rho$  be the stopping time associated with  $\xi$  on the extended space  $\bar{C}_0(\mathbb{R}_+)$ ; cf. Lemma 3.10. Then  $\xi \in \mathcal{R}(\nu)$  implies that  $\rho$  is an embedding of  $\nu$  into  $\bar{B}$  and the argument of (i) applied on  $\bar{C}_0(\mathbb{R}_+)$  yields  $\xi(M) = E^{\bar{\mathbb{W}}}[M(\bar{B})_\rho] \leq 0$ .  $\square$

Definition 4.1 is tailored to imply the following closedness property; it will be used to derive the absence of a duality gap in the next section and immediately implies the existence of a dual optimizer.

**Proposition 4.9.** *Consider  $G^k, G : S \rightarrow [0, \infty]$  and  $(M^k, \psi_k) \in \mathcal{D}(G^k)$  for  $k \geq 1$ . Assume that  $G^k \rightarrow G$  pointwise and  $\sup_k \nu(\psi_k) < \infty$ . Then there exist  $(M, \psi) \in \mathcal{D}(G)$  such that  $\nu(\psi) \leq \liminf \nu(\psi_k)$ .*

*Proof.* Let  $c = \sup \nu(\psi_k)$ . By passing to a subsequence we may assume that  $\lim \nu(\psi_k)$  exists. By Remark 4.7, we may normalize  $\psi_k$  such that  $\psi_k^{**} \geq 0$ .

As  $\psi_k \geq \psi_k^{**} \geq 0$ , Komlos' lemma (see [28, Lemma A1.1] and its subsequent remark) allows us to find convex combinations  $\phi_k$  of  $(\psi_k, \psi_{k+1}, \dots)$  which converge  $\nu$ -a.s. Let  $\psi := \limsup \phi_n$  and note that

$$0 \leq \nu(\psi) = \nu(\liminf \phi_n) \leq \lim \nu(\phi_k) = \lim \nu(\psi_k)$$

by Fatou's lemma. After replacing  $M_k$  with the corresponding convex combinations, we may assume that  $\phi_k = \psi_k$ .

We have  $0 \leq \nu(\psi_k^{**}) \leq \nu(\psi_k) \leq c$ . As in the proof of [12, Proposition 5.5], this implies a uniform (in  $k$ ) bound for the Lipschitz constant of  $\psi_k^{**}$  on each compact subset of  $J$ , and hence a uniform bound  $0 \leq \psi_k^{**} \leq c_n$  on  $K_n$ , independent of  $k$ . By Lemma 4.6, it follows that

$$M^k \geq -\psi_k^{**}(B) \geq -c_n \quad \text{on} \quad [0, T_n].$$

This guarantees that  $(M^k)$  admit a limiting supermartingale  $Z$  in a suitable sense. More precisely, [26, Theorem 2.7] and a diagonal argument show that after taking suitable convex combinations  $N^k$  of  $(M^k, M^{k+1}, \dots)$ , there exists an optional<sup>2</sup> process  $Z$  which is a strong supermartingale (as defined

<sup>2</sup> To be completely precise, [26] assumes the usual conditions for the filtration. We can, for instance, apply their result in the  $\mathbb{W}$ -augmented filtration  $\mathbb{F}^\alpha$  and then pass back to  $\mathbb{F}$  at the end of this paragraph: the local martingale  $M$  is necessarily continuous and hence  $\mathbb{F}^\alpha$ -predictable, but then we can choose an  $\mathbb{F}$ -predictable version (up to evanescence) of  $M$  by applying [30, Appendix I.7, p. 399].

in [30, Appendix I]) on  $[0, T_n]$  for all  $n$  and

$$N_\tau^k \rightarrow Z_\tau \quad \text{in probability}$$

for any stopping time  $\tau$  such that  $\tau \leq T_n$  for some  $n$ . We may assume that  $N^k = M^k$ . As  $Z_0 = 0$ , the process  $Z$  admits a Mertens decomposition  $Z = M - A$  where  $A$  is nondecreasing,  $M$  is a local martingale and  $A_0 = M_0 = 0$ ; cf. [30, Appendix I.20, p. 414].

It remains to show that  $(M, \psi) \in \mathcal{D}(G)$ . Set  $H = \limsup_k (M^k + \psi_k(B))$ ; then  $H$  is optional and  $H \geq G$  on  $\cup_n [0, T_n]$  up to evanescence. Thus, it suffices to show that  $(M, \psi) \in \mathcal{D}(H)$ . Indeed, we have

$$M_\tau \geq Z_\tau \geq -c_n \quad \text{and} \quad M_\tau + \psi(B_\tau) \geq Z_\tau + \psi(B_\tau) \geq H_\tau$$

for all  $\tau \leq T_n$ . As  $H$  is optional, we conclude that  $(M, \psi) \in \mathcal{D}(H)$  by using Remark 4.3.  $\square$

Applying Proposition 4.9 to a constant sequence, we deduce that dual existence holds for general reward functions, in contrast to previous formulations of the dual problem [6, 42] where existence can fail even for more regular reward functions.

**Corollary 4.10.** *Let  $G : S \rightarrow [0, \infty]$ . If  $\mathbf{I}(G) < \infty$ , there exists a dual optimizer  $(M, \psi) \in \mathcal{D}(G)$ .*

## 5 Duality

In this section we combine the closedness result of Proposition 4.9 with capacity theory and facts about optimal Skorokhod embeddings to establish the absence of a duality gap. We first state the weak duality.

**Lemma 5.1.** *Let  $G : S \rightarrow [0, \infty]$ . Then  $\mathbf{S}(G) \leq \mathbf{I}(G)$ .*

*Proof.* We need to show that  $\nu(\psi) \geq \xi(G)$  whenever  $(M, \psi) \in \mathcal{D}(G)$  and  $\xi \in \mathcal{R}(\nu)$ . Indeed, by Lemma 4.5 we have  $M + \psi(B) \geq G$   $\xi$ -a.s. Taking expectations under  $\xi$  and recalling that  $\xi(M) \leq 0$  by Lemma 4.8, the claim follows by the monotonicity of the (outer) integral.  $\square$

Next, we state a duality result for semicontinuous functions. The following is a consequence of [6, Theorem 1.2] as well as of [42, Theorem 2.4] after noting that their dual domain is a subset of ours. We provide a sketch of proof for the convenience of the reader.

**Proposition 5.2.** *Let  $G : S \rightarrow \mathbb{R}$  be bounded and upper semicontinuous. Then  $\mathbf{S}(G) = \mathbf{I}(G)$ .*

*Sketch of Proof.* The inequality  $\mathbf{S}(G) \leq \mathbf{I}(G)$  holds by Lemma 5.1. We sketch an argument for the reverse inequality following [42] (and refer to the latter for details). The key idea is to dualize the constraint  $\nu$  and use the Fenchel–Moreau theorem. Indeed, let  $\mathcal{V}$  be the set of all centered probability measures  $\nu$  with finite first moment, equipped with the 1-Wasserstein metric. One verifies that if  $\nu_n \rightarrow \nu$  in  $\mathcal{V}$  and  $\xi_n \in \mathcal{R}(\nu_n)$ , then there exists  $\xi \in \mathcal{R}(\nu)$  which is a weak limit of a subsequence  $(\xi_{n_k})$ . Let  $\nu \in \mathcal{V}$  and let  $\mathbf{S}(\nu) = \mathbf{S}(G, \nu)$  be the corresponding primal problem. Taking  $\xi_n \in \mathcal{R}(\nu_n)$  to be a  $(1/n)$ -optimizer for the primal problem  $\mathbf{S}(\nu_n)$ ; i.e.,

$$\xi_n(G) \geq \sup_{\xi' \in \mathcal{R}(\nu_n)} \xi'(G) - 1/n,$$

it follows that  $\mathbf{S}(\nu) \geq \xi(G) \geq \limsup \mathbf{S}(\nu_n)$ . In brief,  $\nu \mapsto \mathbf{S}(\nu)$  is upper semicontinuous on  $\mathcal{V}$ , and clearly it is also concave and finite-valued. The space  $\mathcal{V}$  can be seen as a closed convex subspace of a Hausdorff locally convex vector space  $V$  of signed measures and  $\nu \mapsto \mathbf{S}(\nu)$  can be extended to  $V$  by assigning the value  $-\infty$  outside of  $\mathcal{V}$ . The topological dual is  $V^* = \mathcal{C}^1$ , the space of continuous functions  $\psi : \mathbb{R} \rightarrow \mathbb{R}$  with linear growth. The Fenchel–Moreau theorem then shows that  $\mathbf{S}(\nu)$  is equal to its biconjugate,

$$\mathbf{S}(\nu) = \mathbf{S}^{**}(\nu) = \inf_{\psi \in \mathcal{C}^1} \left[ \sup_{\nu' \in \mathcal{V}} \mathbf{S}(\nu') - \nu'(\psi) \right] + \nu(\psi). \quad (5.1)$$

Let us fix  $\psi \in \mathcal{C}^1$  and focus on the inner optimization,

$$\sup_{\nu' \in \mathcal{V}} \mathbf{S}(\nu') - \nu'(\psi) = \sup_{\nu' \in \mathcal{V}} \sup_{\xi \in \mathcal{R}(\nu')} \xi(G - \psi(B)).$$

Seeing the functional  $Y := G - \psi(B)$  as the reward function of an optimal stopping problem, one can check that

$$v_0(\psi) := \sup_{\nu' \in \mathcal{V}} \sup_{\xi \in \mathcal{R}(\nu')} \xi(Y) = \sup_{\xi \in \mathcal{R}} \xi(Y) = \sup_{\tau \in \mathcal{T}} E[Y_\tau]$$

is simply the value function of the associated standard optimal stopping problem. In particular,  $v_0 = v_0(\psi)$  is the initial value of the associated Snell envelope  $Z$ ; i.e., the minimal supermartingale dominating  $Y$ . We write its Doob–Meyer decomposition as  $Z = v_0 + M - A$  where  $M$  is a local martingale,  $A$  is increasing and  $A_0 = M_0 = 0$ . Then

$$v_0 + M \geq Y \quad \text{or equivalently} \quad v_0 + M + \psi(B) \geq G.$$

Since  $\psi$  has linear growth and  $G$  is bounded,  $v_0 + M + \psi(B) \geq G$  implies that  $M \wedge T_n$  is bounded below for all  $n$ . As a result,  $(M, \bar{\psi}) \in \mathcal{D}(G)$  for the function  $\bar{\psi} := v_0 + \psi$ . Since this holds for all  $\psi \in \mathcal{C}^1$ , (5.1) yields that

$$\mathbf{S}(\nu) = \mathbf{S}^{**}(\nu) = \inf_{\psi \in \mathcal{C}^1} v_0(\psi) + \nu(\psi) \geq \inf_{(\bar{\psi}, M) \in \mathcal{D}(G)} \nu(\bar{\psi}) = \mathbf{I}(G)$$

as desired.  $\square$

On the strength of the closedness property in Proposition 4.9, we can use capacity theory to extend the duality to measurable functions. Let  $[0, \infty]^S$  be the set of all functions  $G : S \rightarrow [0, \infty]$ , let  $\text{USA}_+$  be the sublattice of upper semianalytic<sup>3</sup> functions and let  $\mathcal{U}$  be the sublattice of bounded upper semicontinuous functions; note that  $\mathcal{U}$  is stable with respect to countable infima. A mapping  $\mathbf{C} : [0, \infty]^S \rightarrow [0, \infty]$  is called a  $\mathcal{U}$ -capacity if it is monotone, sequentially continuous upwards on  $[0, \infty]^S$  and sequentially continuous downwards on  $\mathcal{U}$ .

**Lemma 5.3.** *The mapping  $\mathbf{S} : [0, \infty]^S \rightarrow [0, \infty]$  is a  $\mathcal{U}$ -capacity.*

*Proof.* As  $\xi(\mathbf{t}) = \int x^2 d\nu$  for all  $\xi \in \mathcal{R}(\nu)$ , the set  $\mathcal{R}(\nu)$  is a nonempty compact space of probability measures on  $S$ ; see for instance [6, Theorem 3.14]. This implies that the associated sublinear map  $G \mapsto \sup_{\xi \in \mathcal{R}(\nu)} \xi(G) \equiv \mathbf{S}(G)$  is a capacity. Indeed, continuity upwards is immediate by commuting two suprema. Let  $G_n \in \mathcal{U}$  decrease to  $G \in \mathcal{U}$ , then there are  $\xi_n \in \mathcal{R}(\nu)$  such that  $\xi_n(G_n) \geq \mathbf{S}(G_n) - 1/n$ . After passing to a subsequence,  $\xi_n \rightarrow \xi$  weakly for some  $\xi \in \mathcal{R}(\nu)$ . Then  $\mathbf{S}(G) = \lim_m \xi(G_m)$  and for each  $m$  we have  $\xi(G_m) \geq \limsup_n \xi_n(G_m) \geq \mathbf{S}(G_m)$ ; thus  $\mathbf{S}(G) \geq \limsup_m \mathbf{S}(G_m)$ . The reverse holds by monotonicity.  $\square$

**Lemma 5.4.** *The mapping  $\mathbf{I} : [0, \infty]^S \rightarrow [0, \infty]$  is a  $\mathcal{U}$ -capacity.*

*Proof.* As  $\mathbf{I} = \mathbf{S}$  on  $\mathcal{U}$  by Proposition 5.2, Lemma 5.3 already shows that  $\mathbf{I}$  is sequentially continuous downwards on  $\mathcal{U}$ . Let  $G, G_n \in [0, \infty]^S$  be such that  $G_n$  increases to  $G$  pointwise; we need to show that  $\mathbf{I}(G_n) \rightarrow \mathbf{I}(G)$ . It is clear that  $\mathbf{I}$  is monotone; in particular,  $\mathbf{I}(G) \geq \limsup \mathbf{I}(G_n)$ , and  $\mathbf{I}(G_n) \rightarrow \mathbf{I}(G)$  if  $\sup_n \mathbf{I}(G_n) = \infty$ .

---

<sup>3</sup>The function  $G$  is called upper semianalytic if the sets  $\{G \geq c\}$  are analytic for all  $c \in \mathbb{R}$ , where a subset of  $S$  is called analytic if it is the image of a Borel subset of a Polish space under a Borel mapping. Any Borel function is upper semianalytic and any upper semianalytic function is universally measurable. See, e.g., [13, Chapter 7] for background.

Hence, we only need to show  $\mathbf{I}(G) \leq \liminf \mathbf{I}(G_n)$  under the condition that  $\sup_n \mathbf{I}(G_n) < \infty$ . Indeed, by the definition of  $\mathbf{I}(G_n)$  there exist  $(M^n, \psi_n) \in \mathcal{D}(G_n)$  with  $\nu(\psi_n) \leq \mathbf{I}(G_n) + 1/n$ . Proposition 4.9 then yields  $(M, \psi) \in \mathcal{D}(G)$  with  $\nu(\psi) \leq \liminf[\mathbf{I}(G_n) + 1/n]$ , showing that  $\mathbf{I}(G) \leq \liminf \mathbf{I}(G_n)$ .  $\square$

We can now prove the main duality result. (Recall that dual attainment was already stated in Corollary 4.10, without measurability assumptions.)

**Theorem 5.5.** *Let  $G : S \rightarrow [0, \infty]$  be upper semianalytic. Then there is no duality gap:  $\mathbf{S}(G) = \mathbf{I}(G) \in [0, \infty]$ .*

*Proof.* In view of Lemma 5.3, Choquet's capacitability theorem (see for instance [58, Proposition 2.11]) shows that

$$\mathbf{S}(G) = \sup\{\mathbf{S}(G') : G' \in \mathcal{U}, G' \leq G\}, \quad G \in \text{USA}_+.$$

By Lemma 5.4, the analogue holds for  $\mathbf{I}$ , and hence the fact that  $\mathbf{S} = \mathbf{I}$  on  $\mathcal{U}$  by Proposition 5.2 already implies that  $\mathbf{S} = \mathbf{I}$  on  $\text{USA}_+$ .  $\square$

We deduce the following characterization of primal and dual optimizers.

**Corollary 5.6.** *Let  $G : S \rightarrow [0, \infty]$  be upper semianalytic and  $\mathbf{S}(G) < \infty$ . Given  $(M, \psi) \in \mathcal{D}(G)$  and  $\xi \in \mathcal{R}(\nu)$ , the following are equivalent:*

- (i)  $(M, \psi)$  is optimal for  $\mathbf{I}(G)$  and  $\xi$  is optimal for  $\mathbf{S}(G)$ ,
- (ii)  $M + \psi(B) = G$   $\xi$ -a.s. and  $\xi(M) = 0$ ,
- (iii)  $M + \psi(B) = G$   $\xi$ -a.s. and  $\xi(M) \geq 0$ .

*Proof.* Theorem 5.5 shows that

$$\xi(\psi(B)) = \nu(\psi) \geq \mathbf{I}(G) = \mathbf{S}(G) \geq \xi(G).$$

Given (iii), we have  $\xi(\psi(B)) \leq \xi(G)$  and thus the above must be equalities; that is, (i) holds. Given (i), these are again all equalities and then (ii) follows after recalling that  $M + \psi(B) \geq G$   $\xi$ -a.s. and  $\xi(M) \leq 0$ ; cf. Lemma 4.5 and Lemma 4.8. The implication from (ii) to (iii) is trivial.  $\square$

**Remark 5.7.** The lower bound on  $G$  in our main duality results can be relaxed to the following condition: there exist  $\psi \in L^1(\nu)$  and a local martingale  $M$  which is bounded on  $[0, T_n]$  for all  $n$  such that

$$G \geq -M - \psi(B).$$

Indeed, the stated results can then be applied to  $G' := G + M + \psi(B) \geq 0$  to derive the corresponding assertions for  $G$ .

Next, we provide a monotonicity principle in the spirit of [12, Corollary 7.8] which provides a universal support  $\Gamma$  characterizing all optimal embeddings, under an integrability condition on  $G$ . As mentioned in the Introduction, this complements the monotonicity principle of [6] which gives a geometric condition on the support that is necessary for optimality, but not sufficient. The following condition is necessary and sufficient. However, the geometry of  $\Gamma$  is merely described in a weaker form, through the construction via a suitable dual optimizer in (5.2). We will exemplify in Section 6 how to exploit such a description. For the statement, note that while  $\xi \in \mathcal{R}$  is defined as a measure on  $C_0(\mathbb{R}_+) \times \mathbb{R}_+$ , it naturally induces a measure on  $S$ : for  $\Gamma \in \mathcal{B}(S)$ , we set  $\xi(\Gamma) = \xi\{(\omega, t) \in C_0(\mathbb{R}_+) \times \mathbb{R}_+ : (\omega|_{[0,t]}, t) \in \Gamma\}$ .

**Corollary 5.8.** *Let  $G : S \rightarrow [0, \infty]$  be Borel and of class (D). There exists a Borel set  $\Gamma \subseteq S$  such that a randomized stopping time  $\xi \in \mathcal{R}(\nu)$  is optimal for  $\mathbf{S}(G)$  if and only if it is concentrated on  $\Gamma$ ; i.e.,  $\xi(\Gamma) = 1$ .*

*Proof.* Since  $G$  is of class (D), there exists a class (D) martingale  $N$  such that  $G \leq N$ ; cf. [30, Appendix I.24, p. 419]. In particular,  $\xi(G) \leq \xi(N) = N_0$  for all  $\xi \in \mathcal{R}(\nu)$ , showing that  $\mathbf{S}(G) < \infty$ . Let  $(M', \psi) \in \mathcal{D}(G)$  be a dual optimizer as guaranteed by Corollary 4.10, normalized such that  $\psi \geq 0$ . Define

$$M = M' \mathbf{1}_{[0, \tau]} + N \mathbf{1}_{[\tau, \infty)} \quad \text{where} \quad \tau = \inf\{t \geq 0 : M'_t = N_t\}.$$

Then  $(M, \psi) \in \mathcal{D}(G)$  is again a dual optimizer and as  $M^+$  is of class (D), Fatou's lemma (in its version with a uniformly integrable bound) yields that  $\xi(M) \geq 0$  for any  $\xi \in \mathcal{R}$ . We set

$$\Gamma := \{M + \psi(B) = G\} \subseteq S, \tag{5.2}$$

then the equivalence of (i) and (iii) in Corollary 5.6 shows that  $\xi \in \mathcal{R}(\nu)$  is optimal if and only if  $\xi(\Gamma) = 1$ .  $\square$

In the above proof, the integrability of  $G$  is used to infer that there exists a dual optimizer  $(M, \psi)$  such that  $\xi(M) = 0$  for all  $\xi \in \mathcal{R}(\nu)$ . We will see in Section 7.3 that this need not hold when  $G$  is not of class (D), even if  $\mathbf{S}(G) < \infty$ . More surprisingly, Corollary 5.8 and even the very essence of the monotonicity principle may fail: the optimality of an optimal embedding *cannot be described through its support* in that case.

## 6 Optimal Cave Embeddings

In this section, we consider an optimal Skorokhod embedding problem introduced in [6] where the (unique) optimal embedding is the hitting time of a set that is cave-shaped; that is, consists of a left and a right barrier. There are typically infinitely many such caves that embed a given distribution  $\nu$ , leading to the question of how to characterize the optimal one. Our main purpose here is to exemplify how our result on dual existence can be useful in deriving a variational characterization. The characterization itself is similar to the one provided by Cox and Kinsley in [22] for a different class of reward functions. However, thanks to dual existence, our argument here is much more direct than the one of Cox and Kinsley who develop a discretization approach [21] in order to reduce to finite-dimensional linear programming problems and obtain their result through tedious and delicate limiting arguments. More importantly, the proof given here suggests that *the variational condition holds for much more general classes of embeddings*, an issue to be addressed in future work.

We first fix some terminology (see also [6, 22, 71]). Let  $(\mathbb{R}_+ \times \mathbb{R})^* = (\mathbb{R}_+ \times \mathbb{R}) \setminus \{(0, 0)\}$  be the punctured half-plane. Following Root's embedding, a *right barrier* is a closed set  $\mathcal{R} \subseteq (\mathbb{R}_+ \times \mathbb{R})^*$  which is closed to the right; that is,  $(s, x) \in \mathcal{R}$  and  $t \geq s$  imply  $(t, x) \in \mathcal{R}$ . Such a barrier is characterized by a function  $x \mapsto r(x) \in [0, \infty]$  which traces its left boundary,

$$r(x) = \inf\{t \geq 0 : (t, x) \in \mathcal{R}\}, \quad \inf \emptyset := +\infty.$$

Similarly, in the spirit of Rost, a *left barrier*  $\mathcal{L}$  is a closed subset of  $(\mathbb{R}_+ \times \mathbb{R})^*$  which is closed to the left. It is characterized by a function  $x \mapsto l(x) \in \{-1\} \cup [0, \infty)$  where we now set

$$l(x) = \sup\{t \geq 0 : (t, x) \in \mathcal{L}\}, \quad \sup \emptyset := -1.$$

Note that we use the value  $-1$  for gaps in the left barrier.

**Definition 6.1.** Given  $t_p \in \mathbb{R}_+$ , a *cave barrier with parting  $t_p$*  is a set  $\mathcal{L} \cup \mathcal{R}$  where  $\mathcal{L} \subseteq [0, t_p] \times \mathbb{R}$  is a left barrier and  $\mathcal{R} \subseteq [t_p, \infty) \times \mathbb{R}$  is a right barrier.

Clearly, a cave barrier  $\mathcal{L} \cup \mathcal{R}$  is characterized by two functions  $l \leq t_p \leq r$ . We denote by  $D$  the (open) complement  $(\mathbb{R}_+ \times \mathbb{R})^* \setminus (\mathcal{L} \cup \mathcal{R})$  and refer to  $D$  as the *continuation region* of the cave barrier (or just as the cave, when there is no ambiguity). Let  $\tau = \inf\{t \geq 0 : (t, B_t) \notin D\}$  be the first exit time of  $D$ . This is the minimum of the two stopping times

$$\begin{aligned} \tau_l &= \inf\{t \geq 0 : B_t \in \mathcal{L}\} \in [0, t_p] \cup \{\infty\}, \\ \tau_r &= \inf\{t_p \leq t < \tau_l : B_t \in \mathcal{R}\} \in [t_p, \infty]. \end{aligned}$$

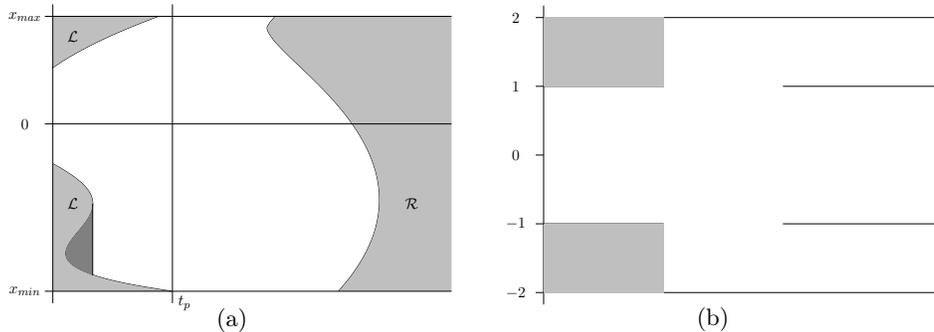


Figure 1: (a) The dark area is added to the left barrier  $\mathcal{L}$  to make it monotone. (b) Example of Remark 6.2.

Similarly, if  $\nu = \text{Law}(B_\tau)$  is the measure embedded by the cave barrier, then  $\nu = \nu_l + \nu_r$  is decomposed into the subprobabilities corresponding to mass absorbed at the left and right barrier; or more precisely,  $\nu_l = \text{Law}(B_{\tau_l})$  and  $\nu_r = \text{Law}(B_{\tau_r})$  with the convention that  $B_\infty$  is valued in an external cemetery state and the laws are restricted to  $\mathbb{R}$ .

Two different cave barriers may have the same hitting time. First, consider a left barrier  $\mathcal{L}$  with corresponding function  $l$ . If  $l$  has a “sink” on  $(0, \infty)$ , say, then  $(t, B_t)$  cannot hit that part of the boundary since the time coordinate always runs forward (see Figure 1). If we replace  $l$  by the increasing envelope of  $x \mapsto l(x)$  on  $(0, \infty)$  and its decreasing envelope on  $(-\infty, 0)$ , the new barrier has the same hitting time. We call  $\mathcal{L}$  *monotone* if  $l$  is already equal to this envelope. Note that given  $\mathcal{L}$ , there exists a minimal monotone left barrier containing  $\mathcal{L}$ . Next, consider a cave barrier and note that  $(t, B_t)$  can only hit the boundary of the component of  $D$  that contains  $(0, 0)$ . Thus, we say that the cave is *connected* if  $D$  is. For brevity we say that a cave barrier is *regular* if  $\mathcal{L}$  is monotone and  $D$  is connected, and note that every cave barrier has a minimal regular cave barrier containing it. This notion is important due to the following fact: the first hitting times of two regular caves are a.s. equal if and only if the caves are equal. We omit the details and refer instead to [71, p. 365] and the end of the proof of [6, Theorem 2.4] for analogous and detailed discussions.

It is also useful to note that an interval of constancy of  $l$  corresponds to an interval where  $\nu_l$  has no mass. On the other hand,  $\nu_r$  has no mass on an interval where  $r = \infty$ . Finally, an atom in  $\nu$  is generated by a horizontal portion in the boundary of  $\mathcal{L}$  or  $\mathcal{R}$ ; that is, a discontinuity of  $l$  or  $r$ .

**Remark 6.2.** Cave embeddings for a given distribution  $\nu$  are trivially non-unique because any cave can be regularized without changing the embedded distribution. However, in contrast to Root and Rost embeddings, *cave embeddings are non-unique even if this regularity is imposed.*

In essence, this ambiguity arises because a given “piece” of  $\nu$  can be absorbed either by the left or the right boundary. A simple example can be generated by taking  $\nu = (\delta_{-2} + \delta_{-1} + \delta_1 + \delta_2)/4$ , so that the left and right barriers are (the envelope of) horizontal spikes at these four locations; cf. Figure 1. One can shorten the spikes on the right at  $\pm 1$  and suitably enlarge the ones on the left, thus changing  $\nu_l$  and  $\nu_r$  without altering the embedded distribution  $\nu = \nu_l + \nu_r$ .

Next, we turn to optimal Skorokhod embeddings for a reward function  $G_t = g(t)$  that is deterministic and depends only on the time variable. More specifically, we assume the following.

**Assumption 6.3.** The reward function  $G_t = g(t)$  is given by a bounded, Lipschitz continuous function  $g : \mathbb{R}_+ \rightarrow \mathbb{R}$  of time such that for some  $t_p \geq 0$ ,  $g$  is differentiable on  $(t_p, \infty)$  and<sup>4</sup>

$$\begin{aligned} g \text{ is strictly convex and strictly decreasing} & \quad \text{on } (0, t_p), \\ g \text{ is strictly concave and strictly increasing} & \quad \text{on } (t_p, \infty). \end{aligned}$$

We also define  $g(\infty)$  as the obvious limit. Such reward functions give rise to cave embeddings as follows.

**Proposition 6.4.** *Suppose that  $\nu(\{0\}) = 0$  and that  $g$  satisfies Assumption 6.3. There exists an optimal stopping time  $\tau \in \mathcal{T}(\nu)$ ; that is, we have  $E[g(\tau)] = \sup_{\xi \in \mathcal{R}(\nu)} \xi(g(\mathbf{t}))$ . Moreover,  $\tau$  is the unique optimizer within  $\mathcal{R}(\nu)$  and given by the first hitting time of a regular cave barrier with parting  $t_p$ . The cave is unique among all regular caves. Finally, we have  $P\{\tau = t\} = 0$  for all  $t \in [0, t_p]$ .*

*Proof.* The first two assertions are stated in [6, Theorem 2.5] which is itself a direct consequence of the monotonicity principle in [6, Theorem 1.3]. The latter states that if  $\tau \in \mathcal{T}(\nu)$  is optimal, then  $(B_{\cdot \wedge \tau}, \tau) \in \Gamma$   $\mathbb{W}$ -a.s. for a Borel set  $\Gamma \subseteq S$  satisfying

$$\text{SG} \cap (\Gamma^< \times \Gamma) = \emptyset, \tag{6.1}$$

---

<sup>4</sup>The decrease/increase can be replaced by:  $\partial^+ g(s) < g'(t)$  for all  $0 \leq s < t_p < t$ .

where  $\Gamma^< \subseteq S$  is defined by

$$\Gamma^< = \{(f', t') \in S : t' < t \text{ and } f' = f \text{ on } [0, t'] \text{ for some } (f, t) \in \Gamma\}$$

and  $\text{SG}$  is the set of so-called stop-go pairs; cf. [6, Definition 1.4]. In the present setting, due to the convexity and monotonicity properties of  $g$ ,

$$\text{SG} = \{((f, t), (f', t')) \in S \times S : f(t) = f'(t'), t' < t \leq t_p \text{ or } t_p \leq t < t'\}. \quad (6.2)$$

(In [6] there are further assumptions on  $g$  which are, however, not used in the proofs.) [6, Theorem 2.5] also states that every optimal  $\xi \in \mathcal{R}(\nu)$  is the hitting time of a cave, and that implies uniqueness of  $\tau$  similarly as in [60]. Indeed, if  $\tau_1$  and  $\tau_2$  are optimizers we can define a randomized stopping time  $\tau'$  by independently picking  $\tau_1$  or  $\tau_2$  with probability 1/2. But then  $\tau'$  is optimal and it follows that  $\tau'$  is non-randomized and thus  $\tau_1 = \tau_2$  a.s. The uniqueness of the cave holds because regular caves are in one-to-one correspondence with their hitting times, as noted above.

For  $t < t_p$  we have  $P\{\tau = t\} = 0$  since the left barrier cannot give rise to an atom in  $\tau$ . To see that  $P\{\tau = t_p\} = 0$ , consider a measurable set  $\Gamma \subseteq S$  with  $P\{(B_{\wedge\tau}, \tau) \in \Gamma\} = 1$  and suppose that  $P\{\tau = t_p\} > 0$ . Then  $t_p > 0$  and  $\Gamma$  must contain a path  $f$  which is stopped at  $t_p$  and satisfies  $x := f(t_p) \in (x_{min}, x_{max})$ . Thus,  $l(x) < t_p$ , and we have  $(t', x) \in D$  for all  $l(x) < t' < t_p$ . It follows that there exists a stopped path  $(f', t') \in \Gamma^<$  with  $f'(t') = x$  and  $t' < t_p$ , but then  $(f, t_p)$  and  $(f', t')$  form a stop-go pair by (6.2) and now (6.1) yields a contradiction.  $\square$

The previous proposition leaves open how to *characterize* the functions  $l, r$  corresponding to the optimal barrier. Intuitively, the non-uniqueness of caves embedding  $\nu$  stems from the fact that a given piece of  $\nu$  can be absorbed at either of the two barriers. However, transferring mass from one to the other changes the reward  $E[g(\tau)]$ , and that is the basis of a variational characterization. Consider an optimal cave for  $g$  at a point  $x \in \mathbb{R}$  where both barriers absorb mass, or more precisely,  $x \in \text{supp } \nu_l \cap \text{supp } \nu_r$ . Intuitively, deforming  $l$  and  $r$  locally around  $x$  corresponds to transferring mass from one boundary to the other, and if the cave is optimal, the derivative corresponding to this variation should vanish. Of course we cannot absorb a negative mass, so that if  $x \in \text{supp } \nu_l \setminus \text{supp } \nu_r$  or vice versa, only one-sided variations are possible. Thus, the precise statement in Theorem 6.5 below will consist of inequality conditions for each of the supports, amounting to an equality only on the intersection.

It seems difficult to find a tractable parametrization for all variations of a cave that preserve the embedded distributions. Instead, we shall utilize the dual problem; a formal derivation runs as follows. Consider for simplicity the case  $x \in \text{supp } \nu_l \cap \text{supp } \nu_r$  and recall that the dual problem admits a solution  $(M, \psi)$ . Since  $g$  is Markovian (a function of time and state), one may expect that the martingale  $M$  can also be chosen of the Markovian form  $M_t = m(t, B_t)$ . Moreover, the dual solution satisfies  $M_\tau + \psi(B_\tau) = g(\tau)$  where  $\tau$  is the exit time of  $D$ , which roughly translates to  $m(t, x) = g(t) - \psi(x)$  for  $t \in \{l(x), r(x)\}$ . Assuming a *smooth fit* at the boundary, formally taking derivatives yields  $\partial_t m(t, x) = g'(t)$  for  $t \in \{l(x), r(x)\}$ . Since  $M$  is a martingale,  $m$  is a solution of the heat equation and then so is  $\partial_t m$ . A version of the Feynman–Kac formula now yields that  $\partial_t m(t, x) = E_{t,x}[g'(\tau)]$  for  $l(x) \leq t \leq r(x)$ . Thus, the difference

$$\begin{aligned} g(r(x)) - g(l(x)) &= m(r(x), x) + \psi(x) - m(l(x), x) - \psi(x) \\ &= m(r(x), x) - m(l(x), x) \end{aligned}$$

can be expressed as  $\int_{l(x)}^{r(x)} \partial_t m(t, x) dt = \int_{l(x)}^{r(x)} E_{t,x}[g'(\tau)] dt$ . The identity

$$g(r(x)) - g(l(x)) = \int_{l(x)}^{r(x)} E_{t,x}[g'(\tau)] dt$$

no longer refers to the dual solution. As mentioned above, this equation needs to be weakened to an inequality if  $x$  is not in the support of both measures  $\nu_l, \nu_r$ .

In what follows we assume throughout that  $\nu$  is a centered distribution with finite second moment and  $\nu(\{0\}) = 0$ . Some technical aspects of the proof depend on whether  $\nu$  has atoms at the endpoints of its support. We focus on the case where  $\nu$  has atoms at two endpoints:  $\nu$  is concentrated on a compact interval  $J = [x_{min}, x_{max}]$  with  $\nu(x_{min}) > 0$  and  $\nu(x_{max}) > 0$ . It is worth noting that given these properties of  $\nu$ , the regular caves under discussion satisfy  $l(x) = t_p = r(x)$  at the two endpoints and the necessary jumps of  $l$  and  $r$  imply horizontal portions of  $\partial D$  along  $\{x = x_{min}\}$  and  $\{x = x_{max}\}$ ; flat portions of floor and ceiling containing  $t_p$ , so to speak. While  $D$  is of bounded height, it can be unbounded to the right since  $r(x) = \infty$  is a possible value.

In the following result we use the right derivative  $\partial^+ g$ , but the same holds for the left derivative.

**Theorem 6.5.** *Let  $D$  be a regular cave with parting  $t_p$  embedding  $\nu$  and let  $l, r$  be the functions delimiting  $D$ . Then  $\tau = \inf\{t \geq 0 : (t, B_t) \notin D\}$  is the*

unique optimizer for  $g$  in  $\mathcal{R}(\nu)$  if and only if

$$\begin{aligned} \int_{l(x)}^{r(x)} E_{t,x}[\partial^+ g(\tau)] dt &\geq g(r(x)) - g(l(x)) \quad \text{for } x \in \text{supp } \nu_l, \\ \int_{l(x)}^{r(x)} E_{t,x}[\partial^+ g(\tau)] dt &\leq g(r(x)) - g(l(x)) \quad \text{for } x \in \text{supp } \nu_r, \end{aligned}$$

where  $E_{t,x}[\partial^+ g(\tau)] := E[\partial^+ g(\tau_{t,x})]$  for  $\tau_{t,x} := \inf\{s \geq t : (s, x + B_{s-t}) \notin D\}$ .

We use our setup for the dual problem from Section 4 with

$$T_n = T := \inf\{t \geq 0 : B_t \in \partial J\}, \quad n \geq 1.$$

Since in this setting we are not interested in the dual martingale  $M$  beyond time  $T$ , we redefine  $\mathcal{D}(G)$  slightly in the sense that  $M$  is only defined up to  $T$ . Clearly, this does not affect the previous results, as one can trivially extend  $M$  beyond  $T$  in a constant fashion to retrieve a dual element in the previous sense.

Next, we consider the optimal cave  $D$  and focus on the necessity of the variational condition. The first step in our proof is to construct the function  $m(t, x)$ . Importantly, we represent  $m$  through an optimal stopping problem that will be used to derive the crucial relationship for  $\partial_t m$ . We write  $I$  for the interior of  $J$ ; that is,  $I = (x_{\min}, x_{\max})$ .

**Proposition 6.6.** *There exists a dual optimizer  $(M, \psi) \in \mathcal{D}(G)$  with  $M_t = m(t, B_t)$  on  $[0, \tau]$  for a universally measurable function  $m : \mathbb{R}_+ \times J \rightarrow \mathbb{R} \cup \{-\infty\}$  which is  $C^\infty$  on  $D$ . Moreover,  $m$  can be taken to be the value function of the optimal stopping problem with reward  $g(t) - \psi(x)$ .*

*Proof.* Let  $(\tilde{M}, \psi) \in \mathcal{D}(G)$  be any dual optimizer; cf. Corollary 4.10. By Remark 4.7 we may assume that  $\psi \geq 0$ . Let  $\mathcal{T}_T = \{\sigma \in \mathcal{T} : \sigma \leq T\}$  and consider the optimal stopping problem

$$\sup_{\sigma \in \mathcal{T}_T} E[G_\sigma^\psi], \quad G_t^\psi := g^\psi(t, B_t), \quad g^\psi(t, x) := g(t) - \psi(x). \quad (6.3)$$

Below, it will be useful to see this as a Markovian optimal stopping problem in the sense of Mertens (we use the completed Brownian filtration throughout this proof). Indeed, setting  $Y_t = (t, B_t)$  we can define a process  $\tilde{Y}$  as the process  $Y$  with absorption on the complement of  $\mathbb{R}_+ \times I$  in  $\mathbb{R}_+ \times \mathbb{R}$ . Then (6.3) is equivalent to an infinite-horizon optimal stopping problem for the right-continuous Markov process  $\tilde{Y}$ .

Next, we show that we can truncate  $G^\psi$  from below without changing the value of (6.3). Since  $g$  is bounded we may assume that  $g \geq 0$ . We then have  $g(t) - \psi(x) \geq -\psi(x)$  and the value of  $\sup_{\sigma \in \mathcal{T}_T} E[-\psi(B_\sigma)]$  with initial condition  $B_t = x \in I$  is given by  $-\psi^{**}(x)$ , where  $\psi^{**} \geq 0$  is the convex hull of  $\psi$  on the interval  $J$ . Thus, the value of the optimal stopping problem

$$\sup_{\sigma \in \mathcal{T}_T} E[\bar{G}_\sigma^\psi], \quad \bar{G}_t^\psi := \bar{g}^\psi(t, B_t), \quad \bar{g}^\psi(t, x) := (g(t) - \psi(x)) \vee (-\psi^{**}(x)) \quad (6.4)$$

is the same as (6.3). We have  $\nu(\psi^{**}) \leq \nu(\psi) < \infty$  by the definition of  $\mathcal{D}(G)$  and due to the assumed form of  $\nu$  this implies that  $\psi^{**}$  is finite at the endpoints of the compact interval  $J$ ; that is,  $\psi^{**}$  is a bounded function and thus  $\bar{g}^\psi(t, x)$  is bounded from below. As a result, the reward function  $\bar{G}^\psi$  in (6.4) is of class (D) and optional (but not necessarily right-continuous), putting us in the setting of [63, 64, 34].

Consider the Snell envelope  $S$  of (6.4) as in [63, Theorem T4] or [34]; that is,  $S$  is the minimal strong supermartingale satisfying  $S \geq \bar{G}^\psi$  on  $[0, T]$ , and  $S$  has the property that  $S_0 = E[S_0] = \sup_{\tau \in \mathcal{T}_T} E[\bar{G}_\tau^\psi]$ . (We use the symbol  $S$  as in the cited references; there should be no possibility of confusion with the space of stopped paths.) Noting that  $\bar{G}^\psi$  is bounded, it follows that  $S_{\cdot \wedge T}$  is bounded. As  $(\tilde{M}, \psi) \in \mathcal{D}(G)$ , the local martingale  $\tilde{M}$  is another supermartingale satisfying  $\tilde{M} \geq G^\psi \geq \bar{G}^\psi$ , so that  $S \leq \tilde{M}$  by minimality and in particular  $S_0 \leq \tilde{M}_0 = 0$ . On the other hand, let  $\tau$  be the primal optimizer. Then

$$S_0 \geq E[\bar{G}_\tau^\psi] \geq E[G_\tau - \psi(B_\tau)] = E[G_\tau] - \nu(\psi) = 0$$

since  $\tau \in \mathcal{T}_T$  and there is no duality gap; cf. Theorem 5.5. As a result,  $S_0 = 0$  and  $\tau$  is an optimal stopping time for (6.4) and (6.3). In particular,  $S$  is a martingale on  $[0, \tau]$ .

The strong supermartingale  $S$  has a Mertens decomposition

$$S = M - A$$

where  $M$  is a local martingale,  $M_{\cdot \wedge T}$  is of class (D),  $A$  is a predictable increasing process and  $M_0 = A_0 = 0$ ; cf. [30, Appendix I.20, p.414]. We observe that  $(M, \psi) \in \mathcal{D}(G)$  is a dual optimizer. Indeed, the defining property of the Snell envelope shows that  $M_t \geq S_t \geq G_t - \psi(B_t)$  on  $[0, T]$  and  $M$  is uniformly bounded from below on  $[0, T]$  since  $M_t \geq S_t \geq -\psi^{**}(B_t)$  as above. Finally, the optimality property depends only on  $\psi$ .

The optimal stopping problem (6.4) is of a Markovian form as considered in [34, 64]. More precisely, [34, Theorem 3.4] or [64, Theorem 3] show that the

Snell envelope is given by  $S_t = v(t, B_t)$  where  $v$  is a universally measurable function, the smallest supermedian-valued function exceeding  $\bar{g}^\psi$  (as defined below [64, Theorem 1]), and that  $v$  coincides with the value function of the optimal stopping problem in the Markovian sense. Since  $S$  is a martingale on  $[0, \tau]$ , we have  $M = S = v(\mathbf{t}, B)$  on  $[0, \tau]$  and thus  $m := v$  satisfies  $m(t, B_t) = M_t$  on  $[0, \tau]$ .

The smoothness of  $m$  on  $D$  follows from the martingale property of  $v$  on  $D$  that is implied by the optimal stopping problem. Indeed, for  $(t, x) \in D$ , consider an open rectangle  $R$  around  $(t, x)$  whose closure is contained in  $D$ . Then  $v|_R$  is smooth as it is the convolution with a smooth kernel, namely,  $v(t, x)$  is the convolution of  $v|_{\partial R}$  with the exit distribution of  $(\mathbf{t}, B)$  when  $B$  is started at  $(t, x)$ .  $\square$

We can observe that the preceding arguments extend to more general functions  $g(t, x)$  with a spatial dependence.

The following result connects  $\partial_t m$  and  $\partial^+ g$  without going through a delicate smooth fit condition as in the sketch above. Instead, it exploits the representation of  $m$  through optimal stopping, and that is crucial in view of the barrier being non-smooth (for general  $\nu$ ).

**Lemma 6.7.** *Let  $x \in I$ . Then*

$$\partial_t m(t, x) = E_{t,x}[\partial^+ g(\tau)] = E_{t,x}[\partial^- g(\tau)] \quad \text{for } l(x) < t < r(x).$$

*Proof.* Let  $x \in I$  and  $l(x) < t < t_1 < r(x)$ . Following the proof of Proposition 6.6,  $m(t, x)$  is the value function of the stopping problem started at  $(t, x)$  and the first exit time  $\tau$  from  $D$  is optimal. Define a stopping time  $\tau_1$  for the initial condition  $(t_1, x)$  as the first exit time of the region  $D_1 := D + \{(t_1 - t, 0)\}$ ; that is, the region  $D$  translated to the right by  $t_1 - t$ . Then  $E_{t_1,x}[\psi(B_{\tau_1})] = E_{t,x}[\psi(B_\tau)]$  and hence

$$\begin{aligned} m(t_1, x) - m(t, x) &\geq E_{t_1,x}[g(\tau_1) - \psi(B_{\tau_1})] - E_{t,x}[g(\tau) - \psi(B_\tau)] \\ &= E_{t_1,x}[g(\tau_1)] - E_{t,x}[g(\tau)] \\ &= E_{t,x}[g(\tau + t_1 - t)] - E_{t,x}[g(\tau)] \\ &= E_{t,x}[g(\tau + h) - g(\tau)] \end{aligned}$$

where  $h = |t_1 - t|$ . Similarly, using  $t_1 < t$  yields  $m(t, x) - m(t - h, x) \leq E_{t,x}[g(\tau) - g(\tau - h)]$ . Recall that  $m$  is differentiable. Dividing by  $h$  and letting  $h \downarrow 0$ , dominated convergence then yields

$$\partial_t m(t, x) \leq E_{t,x}[\partial^- g(\tau)] \leq E_{t,x}[\partial^+ g(\tau)] \leq \partial_t m(t, x)$$

where the middle inequality holds due to  $\partial^- g \leq \partial^+ g$ ; cf. Assumption 6.3. The claim follows. We remark that  $E_{t,x}[\partial^- g(\tau)] = E_{t,x}[\partial^+ g(\tau)]$  can also be deduced directly from the properties of  $\tau$ ; cf. Proposition 6.4.  $\square$

*Proof of Theorem 6.5, Necessity.* Let  $D$  be optimal. For  $x \in \{x_{min}, x_{max}\}$  the result is trivial as  $l(x) = r(x)$ . Given the properties of  $l, r$  and  $g$ , it then suffices to show that

$$\begin{aligned} \int_{l(x)}^{r(x)} E_{t,x}[\partial^+ g(\tau)] dt &\geq g(r(x)) - g(l(x)) \quad \text{for } \nu_l\text{-a.e. } x \in I, \\ \int_{l(x)}^{r(x)} E_{t,x}[\partial^+ g(\tau)] dt &\leq g(r(x)) - g(l(x)) \quad \text{for } \nu_r\text{-a.e. } x \in I. \end{aligned}$$

Let  $x \in I$  and  $l(x) \leq l \leq r \leq r(x)$  with  $r < \infty$ . In view of Lemma 6.7 and the fundamental theorem of calculus,

$$\int_l^r E_{t,x}[\partial^+ g(\tau)] dt = \int_l^r \partial_t m(t, x) dt = m(r, x) - m(l, x).$$

By the construction of  $m$  we have  $m(t, x) \geq g(t, x) - \psi(x)$  on  $D$ , while for  $t = l(x)$  this inequality holds with equality for  $\nu_l$ -a.e.  $x$  and for  $t = r(x)$  it holds with equality for  $\nu_r$ -a.e.  $x$ . Focusing on the first case, we obtain  $\nu_l$ -a.e. that

$$\int_{l(x)}^{r_n} E_{t,x}[\partial^+ g(\tau)] dt \geq g(r_n) - g(l(x))$$

for a sequence  $r_n \uparrow r(x)$ , and the claim follows after recalling that  $g$  is continuous on  $[0, \infty]$  and  $E_{t,x}[\partial^+ g(\tau)] \geq 0$  for  $t \geq t_p$ . The  $\nu_r$ -a.e. inequality follows similarly.  $\square$

*Proof of Theorem 6.5, Sufficiency.* We only provide a sketch of the argument; the details are similar to the proof of [22, Theorem 4.1]. Suppose that the stated inequalities hold for  $l, r$ . Our goal is to construct a pair  $(M, \psi) \in \mathcal{D}(G)$  such that  $M_\tau + \psi(B_\tau) = g(\tau)$  and  $E[M_\tau] \leq 0$  for the exit time  $\tau$  defined by  $D$ ; this will imply the optimality of  $\tau$  by Corollary 5.6. For notational convenience, let

$$h(t, x) = E_{t,x}[\partial^+ g(\tau)], \quad \Gamma(x) = g(l(x)) - g(r(x)) + \int_{l(x)}^{r(x)} h(s, x) ds.$$

As a first step, we consider the function

$$H(t, x) = g(r(x)) - \int_t^{r(x)} h(s, x) ds + \Gamma(x)^+$$

and show that

$$g(t) \leq H(t, x) \quad \text{for } t \geq 0, \quad x \in J, \quad (6.5)$$

$$g(l(x)) = H(l(x), x) \quad \text{for } x \in \text{supp } \nu_l, \quad (6.6)$$

$$g(r(x)) = H(r(x), x) \quad \text{for } x \in \text{supp } \nu_r. \quad (6.7)$$

Indeed, (6.5) follows directly from the definitions. For  $t \geq t_p$  we have that  $g'(t)$  is decreasing and hence

$$g(t) = g(r(x)) - \int_t^{r(x)} \partial^+ g(s) ds \leq g(r(x)) - \int_t^{r(x)} h(s, x) ds + \Gamma(x)^+ = H(t, x)$$

whereas for  $0 \leq t < t_p$  we have  $\partial^+ g(s) \leq \partial^+ g(u)$  for  $t \leq s \leq u$  and hence

$$g(t) = g(l(x)) + \int_{l(x)}^t \partial^+ g(s) ds \leq g(l(x)) + \int_{l(x)}^t h(s, x) ds \leq H(t, x).$$

Moreover, the assumption in Theorem 6.5 states that  $\Gamma(x) \geq 0$  for  $x \in \text{supp } \nu_l$  and  $\Gamma(x) \leq 0$  for  $x \in \text{supp } \nu_r$ . This yields (6.6) and (6.7).

The rest of the proof consists in showing that  $H(t, B_t) = M_t + \psi(B_t)$  on  $[0, \tau]$  and  $H(t, B_t) \leq M_t + \psi(B_t)$  on  $[0, T]$  for a local martingale  $M$  with  $M_0 = 0$  which is bounded below and a function  $\psi \in L^1(\nu)$ . Once that is achieved, (6.5)–(6.7) show that  $(M, \psi) \in \mathcal{D}(G)$  have the desired properties.

To obtain the decomposition we consider the function  $\bar{m}(t, x) := g(r(x)) - \int_t^{r(x)} h(s, x) ds$ . One can show as in [22] that  $\bar{m}(t, B_t)$  is a submartingale on  $[0, \tau]$  and conclude that there is a unique increasing predictable process  $A$  on  $[0, \tau]$  such that  $\bar{m}(t, B_t) - A_t$  is a martingale on  $[0, \tau]$ . One can further show that  $A$  agrees with a continuous additive functional on  $[0, \tau]$  such that  $\bar{m}(t, B_t) - A_t$  is well-defined on  $[0, T]$  and a supermartingale. Using a representation result for additive functionals, one can therefore write  $A_t = z(B_t) - z(B_0) - \int_0^t z'_-(B_s) dB_s$  for a convex function  $z$  and conclude that  $\bar{m}(t, B_t) - z(B_t)$  is still a martingale on  $[0, \tau]$  and a supermartingale on  $[0, T]$ . We can choose  $z$  such that  $z \geq z(0) = \bar{m}(0, 0)$ .

Set  $\psi(x) = z(x) + \Gamma(x)^+$ , so that

$$H(t, x) = \bar{m}(t, x) - z(x) + \psi(x).$$

Moreover, let  $M$  be the martingale part of the supermartingale  $\bar{m}(t, B_t) - z(B_t)$ . To see that  $(M, \psi)$  is indeed a dual element, we show that  $\bar{m}$  and  $z$  are bounded. Then it follows that  $z(x) + \Gamma(x)^+$  is bounded (as  $\Gamma$  is bounded

under our assumptions) and that  $\bar{m}(t, x) - z(x)$ , and hence  $M$ , is bounded from below. Indeed, boundedness of  $\bar{m}$  follows from the identity

$$\bar{m}(t, x) = g(r(x)) - \int_t^{t_p} h(s, x) \mathbf{1}_{s < t_p} ds - \int_{t_p}^{r(x)} h(s, x) \mathbf{1}_{s > t} ds.$$

The first two terms are trivially bounded. For the last term, observe that on the domain of integration we have  $0 \leq h(s, x) \leq \partial^+ g(s)$  which implies boundedness. Next, suppose that  $z$  is unbounded, then we must have  $z(x_{min}) = +\infty$  or  $z(x_{max}) = +\infty$  as  $z$  is convex and bounded from below. Note that  $B_\tau = x_{min}$  and  $B_\tau = x_{max}$  with positive probability, but  $E[\bar{m}(\tau, B_\tau) - z(B_\tau)] = m(0, 0) - z(0) = 0$  by the martingale property. Therefore,  $z$  must be bounded and the proof is complete.  $\square$

## 7 Counterexamples

In this section we demonstrate that relaxing the regularity in the dual domain is necessary to achieve a complete duality theory for general reward functions. It is also shown that the monotonicity principle fails without an integrability condition.

### 7.1 Local Martingale Property of $M$

We construct an example showing that it is crucial to use local martingales  $M$  rather than true martingales as in previous works. More precisely, we construct a continuous reward function  $G$  such that for a wide class of marginals  $\nu$ , any dual optimizer  $(M, \psi) \in \mathcal{D}(G)$  fails to have the true martingale property; in fact,  $E[M_t] > 0$  for all  $t > 0$ .

Let  $\nu$  be a centered distribution which is equivalent to the Lebesgue measure on  $\mathbb{R}$  and satisfies  $\nu(f) < \infty$ , where

$$f(x) = \exp(x^4). \tag{7.1}$$

We note that  $J = \mathbb{R}$  and hence  $\cup_n [0, T_n] = C_0(\mathbb{R}_+) \times \mathbb{R}_+$  in Definition 4.1. An important ingredient for our reward function is the process

$$L_t := \exp \left( B_t^4 - 2 \int_0^t (3B_s^2 + 4B_s^6) ds \right), \quad t \geq 0 \tag{7.2}$$

which can also be written as the stochastic exponential  $L_t = \mathcal{E}_t \left( \int_0^t 4B_s^3 dB_s \right)$ . We are grateful to Johannes Ruf for indicating this remarkably simple example of a strict local martingale to us.

**Lemma 7.1** (J. Ruf). *The stochastic exponential  $L_t = \mathcal{E}_t\left(\int_0^t 4B_s^3 dB_s\right)$  is a positive local martingale with  $E[L_t] < 1$  for all  $t > 0$ . In particular,  $L$  is not a martingale on  $[0, t]$  for any  $t \in (0, \infty)$ .*

We defer the proof to the end of this subsection. As our payoff function, we then choose

$$G = 1 - L + f(B);$$

Note that  $G$  is a continuous function on  $S$ . Moreover, it follows directly from (7.1) and (7.2) that  $f(B_t) \geq L_t$  and hence  $G \geq 1$ . As  $L \geq 0$  and  $f \in L^1(\nu)$ , a particularly simple dual element is  $(M, \psi) := (0, 1 + f) \in \mathcal{D}(G)$ ; therefore,  $\mathcal{D}(G) \neq \emptyset$  and  $\mathbf{I}(G) \leq 1 + \nu(f) < \infty$ , so that the conditions of our main results in Sections 4 and 5 are all satisfied. This dual element features a true martingale; however, it is not optimal.

**Proposition 7.2.** *(i)  $(M, \psi) := (1 - L, f) \in \mathcal{D}(G)$  is optimal for the dual problem  $\mathbf{I}(G)$ . Moreover, any  $\xi \in \mathcal{R}(\nu)$  is optimal for  $\mathbf{S}(G)$ .*

*(ii) If  $(M, \psi) \in \mathcal{D}(G)$  is any optimizer for  $\mathbf{I}(G)$ , then  $E[M_t] > 0$  for all  $t > 0$ , so that  $M$  cannot be a martingale.*

*Proof.* (i) Let  $\tau \in \mathcal{T}(\nu)$  and  $(M, \psi) := (1 - L, f) \in \mathcal{D}(G)$ . It is clear that  $M + \psi(B) = G$   $\xi$ -a.s., and  $\xi(M) \geq 0$  as  $L$  is a nonnegative supermartingale with  $L_0 = 0$ . The claim now follows from Corollary 5.6.

(ii) Let  $(M, \psi) \in \mathcal{D}(G)$ . We first prove that there exists  $c \in \mathbb{R}$  such that

$$\psi(x) + cx \geq f(x), \quad x \in \mathbb{R}. \quad (7.3)$$

Indeed, let  $a, b \geq 0$  and  $\sigma = \inf\{t \geq 0 : B_t \notin (-a, b)\}$ . Note that the local martingale  $L_{\cdot \wedge \sigma}$  is bounded and hence a uniformly integrable martingale. On the other hand, as  $\sigma \leq T_n$  for  $n$  large enough,  $M_{\cdot \wedge \sigma}$  must be bounded below (Definition 4.1) and hence a supermartingale. In particular,  $E[L_\sigma] = 1$  and  $E[M_\sigma] \leq 0$ , so that  $M_\sigma + \psi(B_\sigma) \geq G_\sigma = 1 - L_\sigma + f(B_\sigma)$  implies

$$E[\psi(B_\sigma)] \geq E[f(B_\sigma)].$$

As  $a, b$  were arbitrary, this yields as in the proof of Lemma 4.6 that the convex hull satisfies  $(\psi - f)^{**}(0) \geq 0$ . Taking  $c = \partial^-(\psi - f)^{**}(0)$ , we have  $(\bar{\psi} - f)^{**} \geq 0$  for  $\bar{\psi}(x) = \psi(x) + cx$  and the claim follows.

In view of (7.3) and Remark 4.7 we may assume that  $\psi \geq f$ . If  $(M, \psi)$  is optimal, then  $\nu(\psi) = \mathbf{I}(G) = \nu(f)$  by (i) and it follows that  $\psi = f$   $\nu$ -a.s. and hence Lebesgue-a.e. But now  $M + \psi(B) \geq 1 - L + f(B)$  implies  $M_t \geq 1 - L_t$   $\mathbb{W}$ -a.s. and in particular  $E[M_t] \geq 1 - E[L_t]$  for all  $t > 0$ , so  $E[M_t] > 0$  by Lemma 7.1.  $\square$

*Proof of Lemma 7.1.* (i) We first provide an auxiliary result. Let  $W$  be a Brownian motion on a filtered probability space with measure  $Q$ . Then the SDE

$$dX_t = 4X_t^3 dt + dW_t, \quad X_0 = 0$$

has a unique strong solution  $X$  up to its explosion time  $\tau$  and  $Q\{\tau < T\} > 0$  for all  $T \in (0, \infty)$ . Indeed, existence and uniqueness of  $X$  follow from the local Lipschitz continuity of the coefficients (see, e.g., [76, Exercise 2.10, p. 383]). The scale function of  $X$  is  $p(x) = \int_0^x e^{-2y^4} dy$  and the speed measure is  $m(dx) = 2e^{2x^4} dx$ . Thus

$$v(x) := \int_0^x (p(x) - p(y)) m(dy) = \int_0^x \int_y^x e^{2(y^4 - z^4)} dz dy$$

and in particular  $v(\infty) := \lim_{x \rightarrow \infty} v(x)$  is given by

$$\int_0^\infty \int_0^\infty e^{2(y^4 - (y+u)^4)} du dy = \int_0^\infty \int_0^\infty e^{-2(u^4 + 4u^3y + 6u^2y^2 + 4uy^3)} du dy.$$

Comparison with a Gaussian integral shows that this quantity is finite. In view of the symmetry  $v(x) = v(-x)$ , the same holds for  $v(-\infty)$ . Thus, Feller's test implies that both boundaries  $\pm\infty$  are limit points of  $X$  in finite time with positive probability; in fact, the explosion time  $\tau$  even satisfies  $\tau < \infty$   $Q$ -a.s. by [57, Proposition 5.5.32, p. 350]. Using the homogeneity of  $X$ , this already implies that  $Q\{\tau < T\} > 0$  for all  $T \in (0, \infty)$ ; see, e.g., [17, Theorem 1.1] for an elegant argument.

(ii) We can now prove the lemma. As an exponential of a continuous local martingale, it is clear that  $L$  is a local martingale and strictly positive, hence a supermartingale. Let  $T \in (0, \infty)$  and suppose for contradiction that  $E[L_T] = 1$ , or equivalently, that  $L$  is a martingale on  $[0, T]$ . Then we can introduce the equivalent probability  $Q$  on  $\mathcal{F}_T$  via  $dQ/d\mathbb{W} = L_T$  and Girsanov's theorem shows that the process  $W_t := B_t - 4 \int_0^t B_s^3 ds$  is a  $Q$ -Brownian motion on  $[0, T]$ . Moreover,  $B$  satisfies  $dB_t = 4B_t^3 dt + dW_t$  and  $B_0 = 0$  under  $Q$ . As shown in (i), this implies that  $Q\{\tau < T\} > 0$  for the explosion time  $\tau$  of  $B$ , contradicting that the Brownian motion  $B$  under  $\mathbb{W}$  is non-explosive.  $\square$

## 7.2 Regularity of $\psi$

The following example shows that a duality gap can arise if the functions  $\psi$  in the dual domain  $\mathcal{D}(G)$  are required to be continuous. The reward  $G = \mathbf{1}_{\mathbb{Q}}(\mathbf{t})$  was previously used in [42] to illustrate that their duality result can fail when

the reward function is irregular in time. In our framework, duality holds by Theorem 5.5. Nevertheless, it is instructive to detail the optimizers as this highlights the mechanics of our definitions.

**Example 7.3.** Let  $G = \mathbf{1}_{\mathbb{Q}}(\mathbf{t})$  and  $\nu = (\delta_{-1} + \delta_1)/2$ .

(i) We have  $\mathbf{S}(G) = \mathbf{I}(G) = 0$ , a primal optimizer is given by the stopping time  $\tau = \inf\{t \geq 0 : |B_t| = 1\}$  and a dual optimizer is given by  $M \equiv 0$  and  $\psi = \mathbf{1}_{(-1,1)}$ .

To see this, let  $\tau, M, \psi$  be as above and note that  $J = [-1, 1]$ . We choose  $K_n = J$  and thus  $T := T_n = \tau$  for all  $n$ . We claim that  $\psi(B) \geq G$  on  $[0, T]$  up to evanescence. Indeed,  $\psi(B) = 1$  on  $[0, T]$ , thus  $\{\psi(B) < G\}$  is contained in the graph of  $T$  and of course also in  $\{G = 1\}$ . But since  $T$  has a continuous distribution,  $\mathbb{W}\{T \in \mathbb{Q}\} = 0$  and hence  $[T] \cap \{G = 1\} = \{T \in \mathbb{Q}\} \times \mathbb{R}_+$  is indeed negligible up to evanescence. As a result,  $(M, \psi) \in \mathcal{D}(G)$ . In view of  $\nu(\psi) = 0$  and  $E[G_\tau] = 0$ , the optimality of  $(M, \psi)$  and  $\tau$  now follows from Corollary 5.6.

(ii) If the dual domain is restricted to continuous functions  $\psi$ , a duality gap arises: the dual problem over continuous  $\psi$  has value 1 instead of 0. Indeed, let  $(M, \psi) \in \mathcal{D}(G)$  be such that  $\psi$  is continuous. We claim that there exists  $c \in \mathbb{R}$  such that  $\psi(x) \geq 1 + cx$  for all  $x \in [-1, 1]$ ; in particular, this will imply that  $\nu(\psi) \geq 1$ . Let  $\sigma < T$  be a stopping time, let  $\sigma'_n = \inf\{t \geq \sigma : t \in 2^{-n}\mathbb{N}\}$  be the usual dyadic approximation  $\sigma'_n \downarrow \sigma$  and let  $\sigma_n = \sigma'_n \wedge T$ . Then we have

$$M_{\sigma_n} + \psi(B_{\sigma_n}) \geq G_{\sigma_n} = 1 \quad \text{on} \quad \{\sigma_n < T\}$$

since  $\sigma_n$  has rational values on  $\{\sigma_n < T\}$ . As  $\mathbb{W}\{\sigma_n < T\} \rightarrow 1$ , the continuity of  $\psi$  and  $M$  yields that  $M_\sigma + \psi(B_\sigma) \geq 1$  and hence  $E[\psi(B_\sigma)] \geq 1$ . Since this holds in particular for the hitting time  $\sigma$  of any set  $\{-a, b\}$  where  $-1 < -a \leq 0 \leq b < 1$ , it follows that  $\psi^{**}(0) \geq 0$  where  $\psi^{**}$  is the convex hull on  $(-1, 1)$ . Let  $c = \partial^- \psi^{**}(0)$ , then using again the continuity it follows that  $\psi(x) \geq 1 + cx$  for all  $x \in [-1, 1]$ , as claimed.

### 7.3 Monotonicity Principle

In this section we show that the monotonicity principle of Corollary 5.8 does not hold without an integrability condition. Indeed, we have the following.

**Proposition 7.4.** *There exist a Borel function  $G : S \rightarrow [0, \infty)$ , a centered distribution  $\nu$  on  $\mathbb{R}$  with all moments finite and  $\mathbf{S}(G) < \infty$ , and randomized stopping times  $\xi^1, \xi^2 \in \mathcal{R}(\nu)$  which are equivalent as measures on  $S$ , such that  $\xi^1$  is optimal for  $\mathbf{S}(G)$  and  $\xi^2$  is not optimal for  $\mathbf{S}(G)$ .*

As  $\xi^1(\Gamma) = 1$  is equivalent to  $\xi^2(\Gamma) = 1$ , for any Borel set  $\Gamma \subseteq S$ , it follows that *optimality of  $\xi \in \mathcal{R}(\nu)$  cannot be determined by its support*.

We start with some preliminary results that will be used in the construction. Recall that  $\xi \in \mathcal{R}$  is defined as a measure on  $C_0(\mathbb{R}_+) \times \mathbb{R}_+$  and induces a measure on  $\mathcal{B}(S)$  via  $\xi(\Gamma) := \xi\{(\omega, t) \in C_0(\mathbb{R}_+) \times \mathbb{R}_+ : (\omega|_{[0,t]}, t) \in \Gamma\}$ ; in fact,  $\xi$  is completely characterized by the latter. A product measure  $\mathbb{W} \otimes \lambda$  on  $C_0(\mathbb{R}_+) \times \mathbb{R}_+$  induces a measure on  $S$  in the same fashion. In what follows, we set  $f(x) = \exp(x^4)$  and denote by  $L$  the strict local martingale defined in (7.2).

**Lemma 7.5.** *There exists  $\hat{\xi}^{\text{int}} \in \mathcal{R}$  such that  $\hat{\xi}^{\text{int}}(f(B)) + \hat{\xi}^{\text{int}}(\mathbf{t}) < \infty$  and  $\hat{\xi}^{\text{int}} \gg \mathbb{W} \otimes \lambda$  on  $S$ , where  $\lambda$  is the Lebesgue measure.*

*Proof.* For  $n \geq 1$  we define  $\tau^n = \inf\{t : |B_t| \geq n\}$  and

$$S^n = \{(\omega, t) \in S : \sup_{s \leq t} |\omega_s| < n\} = \{(\omega, t) \in S : t < \tau^n(\omega)\}.$$

We first construct  $\xi^n \in \mathcal{R}$  such that

$$\xi^n[0, \tau^n] = 1 \quad \text{and} \quad \xi^n \gg \mathbb{W} \otimes \lambda \quad \text{on} \quad S^n. \quad (7.4)$$

Consider the adapted, increasing process  $A^n$  defined by

$$A_t^n := 1 - e^{-t} \mathbf{1}_{t < \tau^n};$$

it is strictly increasing and differentiable up to  $\tau^n$  and then jumps to the value 1. Thus, the kernel

$$\xi_\omega^n(dt) = dA_t^n(\omega) = e^{-t} \mathbf{1}_{t < \tau^n(\omega)} dt + e^{-\tau^n(\omega)} \delta_{\tau^n(\omega)}(dt)$$

defines a randomized stopping time via  $\xi^n = \mathbb{W}(d\omega) \xi_\omega^n(dt)$  which satisfies (7.4).

Clearly (7.4) implies  $\xi^n(f(B)) \leq \exp(n^4)$  and  $\xi^n(\mathbf{t}) \leq E[\tau^n] = n^2 \leq \exp(n^4)$ . Let  $(a_n)_{n \geq 1}$  be a sequence in  $(0, 1)$  such that  $\sum_{n \geq 1} a_n = 1$  and  $\sum_{n \geq 1} a_n \exp(n^4) < \infty$ . We define  $\xi := \sum_{n \geq 1} a_n \xi^n$ ; then  $\xi \in \mathcal{R}$  satisfies  $\xi^n(f(B)) < \infty$  and  $\xi^n(\mathbf{t}) < \infty$ . Since every stopped path is bounded, we have  $\cup_{n \geq 1} S^n = S$  and thus (7.4) implies  $\xi \gg \mathbb{W} \otimes \lambda$  on  $S$  as desired.  $\square$

**Lemma 7.6.** *Let  $\sigma = \inf\{t : t^2 + B_t^2 = 1\}$ . There exists an  $\mathcal{F}_\sigma$ -measurable Bernoulli random variable  $X$  independent of  $B_\sigma$ .*

*Proof.* Define  $\sigma' = \inf\{t : t^2 + B_t^2 = 1/2\}$ . Then  $B_{\sigma'}$  is  $\mathcal{F}_\sigma$ -measurable and its conditional distribution given  $B_\sigma$  is atomless. In particular, there exists a conditional median  $m(x)$  given  $B_\sigma = x$ ; that is,  $P[B_{\sigma'} \geq m(x) | B_\sigma = x] = 1/2$ . By construction,  $X := \mathbf{1}_{B_{\sigma'} \geq m(B_\sigma)}$  has a Bernoulli distribution and is independent of  $B_\sigma$ .  $\square$

**Lemma 7.7.** *We have  $\xi(L) \leq 1$  for all  $\xi \in \mathcal{R}$ . If  $\xi$  embeds a distribution  $\nu$  with  $\nu(f) < \infty$ , then  $\xi(L) = 1$ .*

*Proof.* As  $L$  is a positive local martingale with  $L_0 = 1$ , we have  $\xi(L) \leq 1$  by Fatou's lemma. Suppose that  $\xi \in \mathcal{R}(\nu)$  where  $\nu(f) < \infty$  and recall that  $0 \leq L_t \leq f(B_t)$ . As  $f$  is convex,  $f(B_t)$  is a positive submartingale up to  $\xi$  and hence of class (D), where we may use the representation of  $\xi$  as nonrandomized stopping time in the enlarged filtration (cf. Lemma 3.10) to apply the standard results of stochastic analysis. This implies that  $L$  is a martingale of class (D) up to  $\xi$  and in particular  $\xi(L) = 1$ .  $\square$

*Proof of Proposition 7.4.* Let  $G \geq 0$  be the payoff defined by

$$G_t = \mathbf{1}_{t > \sigma, X=0} + L_{t-\sigma}(B^{\sigma \mapsto}) \mathbf{1}_{t > \sigma, X=1}$$

where  $\sigma = \inf\{t : t^2 + B_t^2 = 1\}$ , the random variable  $X$  is as in Lemma 7.6 and, with the notation of Definition 3.3,

$$L_{t-\sigma}(B^{\sigma \mapsto}) = \exp\left(\left(B_t - B_\sigma\right)^4 - 2 \int_\sigma^t (3(B_s - B_\sigma)^2 + 4(B_s - B_\sigma)^6) ds\right).$$

Let  $\xi^{\text{int}} = \sigma \oplus \hat{\xi}^{\text{int}}$  be the randomized stopping time obtained by shifting  $\hat{\xi}^{\text{int}}$  of Lemma 7.5 by  $\sigma$  (cf. Definition 3.3); that is, if  $\hat{\xi}^{\text{int}} = \mathbb{W}(d\omega) \hat{\xi}_\omega^{\text{int}}(ds)$ , then  $\xi^{\text{int}} = \mathbb{W}(d\omega) \xi_\omega^{\text{int}}(ds)$  where

$$\xi_\omega^{\text{int}}[0, t] = \mathbf{1}_{\sigma(\omega) < t} \hat{\xi}_{\omega, -\omega(\omega)}^{\text{int}}[0, t - \sigma(\omega)].$$

Using Lemma 7.7, we then have  $\xi^{\text{int}}(G) = \hat{\xi}^{\text{int}}(L) = 1$ .

Next, let  $\hat{\xi}^{\text{exp}}$  be an exponential random time, defined by its kernel  $\hat{\xi}_\omega^{\text{exp}}(ds) = e^{-s} ds$ . As  $E[L_t] < 1$  for  $t > 0$  by Lemma 7.1, we have  $\hat{\xi}^{\text{exp}}(L) = \int_0^\infty e^{-t} E[L_t] dt < 1$ . Moreover,  $\hat{\xi}^{\text{exp}}$  is clearly equivalent to  $\mathbb{W} \otimes \lambda$  on  $S$ , so that  $\hat{\xi}^{\text{exp}} \ll \hat{\xi}^{\text{int}}$ . The reverse is not true, but if we set

$$\hat{\xi}^{\text{avg}} = (\hat{\xi}^{\text{exp}} + \hat{\xi}^{\text{int}})/2,$$

then  $\hat{\xi}^{\text{avg}}$  is equivalent to  $\hat{\xi}^{\text{int}}$  on  $S$  and we also have  $\hat{\xi}^{\text{avg}}(L) < 1$ .

Define  $\xi^{\text{avg}} = \sigma \oplus \hat{\xi}^{\text{avg}}$ . Then  $\xi^{\text{avg}}$  and  $\xi^{\text{int}}$  already have properties close to the desired ones, but they do not embed the same distribution yet. To achieve that, we randomly mix the two stopping times using  $X$ . Indeed, we define the randomized stopping times  $\xi^1$  and  $\xi^2$  through their kernels

$$\xi_\omega^1 := \xi_\omega^{\text{avg}} \mathbf{1}_{X(\omega)=0} + \xi_\omega^{\text{int}} \mathbf{1}_{X(\omega)=1}, \quad \xi_\omega^2 := \xi_\omega^{\text{avg}} \mathbf{1}_{X(\omega)=1} + \xi_\omega^{\text{int}} \mathbf{1}_{X(\omega)=0}.$$

Then

$$\xi^1(G) = \frac{1}{2} \hat{\xi}^{\text{avg}}(1) + \frac{1}{2} \hat{\xi}^{\text{int}}(L) = 1$$

whereas

$$\xi^2(G) = \frac{1}{2} \hat{\xi}^{\text{avg}}(L) + \frac{1}{2} \hat{\xi}^{\text{int}}(1) < 1.$$

By construction,  $\xi^1$  and  $\xi^2$  are equivalent on  $S$  and embed the same distribution  $\nu$ . The integrability properties of  $\hat{\xi}^{\text{exp}}$  and  $\hat{\xi}^{\text{int}}$  entail that  $\nu$  has all moments (even some exponential moments) and that  $\xi^1(\mathbf{t}) = \xi^2(\mathbf{t}) < \infty$ .  $\square$

## A Appendix: Extension to Finite First Moment

In the body of this article we have assumed that the embedded measure  $\nu$  has a finite second moment, but the results can be extended to the case of a finite first moment by using well-known facts, at the expense of a slightly more convoluted definition. The crucial observation is that under the second moment condition,  $E[\tau] < \infty$  for  $\tau \in \mathcal{T}(\nu)$  is equivalent to  $\tau$  being minimal (or uniformly integrable), and minimality can still be used in the first moment case. For a randomized stopping time, the simplest definition is obtained by referring to its stopping time representation in a larger filtration.

For the remainder of this section,  $\nu$  is a centered distribution on  $\mathbb{R}$  with finite first moment.

**Definition A.1.** Let  $\tau$  be a finite stopping time such that  $B_\tau \sim \nu$ . Then  $\tau$  is called *minimal* if there exists no smaller embedding; that is, if  $\sigma$  is another stopping time such that  $B_\sigma \sim \nu$  and  $\sigma \leq \tau$  a.s., then  $\tau = \sigma$  a.s. We denote by  $\mathcal{T}(\nu)$  the set of all such  $\tau$ .

Let  $\xi$  be a randomized stopping time such that  $\xi \circ B^{-1} = \nu$  and let  $\rho$  be the associated  $\bar{\mathbb{F}}$ -stopping time; cf. Lemma 3.10. Then  $\xi$  is called *minimal* if  $\rho$  is minimal in the above sense. We denote by  $\mathcal{R}(\nu)$  the set of all such  $\xi$ .

We can now state the announced extension.

**Remark A.2.** The results in Sections 3–5 continue to hold under the first moment condition.

*Proof.* The modifications are as follows. (i) In the proof of Lemma 3.8 we used that  $\mathcal{R}(\nu)$  is weakly compact. This still holds in the present setting; cf. [6, Section 7.1]. (ii) A stopping time  $\tau$  is minimal if and only if  $B_{\cdot \wedge \tau}$  is uniformly integrable; cf. [65, Theorem 3]. Thus, in the proofs of Lemmas 4.5 and 4.8, we can use directly that  $B_{\cdot \wedge \tau}$  is uniformly integrable instead of deriving this from  $E[\tau] < \infty$ . (iii) Proposition 5.2 still holds, e.g., by [6, Theorem 7.2 and Example 7.2.1]. (iii) The proof of Lemma 5.3 again used that  $\mathcal{R}(\nu)$  is weakly compact; cf. (i). The other proofs did not use directly that  $\nu$  has finite second moment.  $\square$

## 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. *Math. Finance*, 26(2):233–251, 2016.
- [2] C. D. Aliprantis and K. C. Border. *Infinite Dimensional Analysis: A Hitchhiker’s Guide*. Springer, Berlin, 3rd edition, 2006.
- [3] L. Ambrosio. Lecture notes on optimal transport problems. In *Mathematical aspects of evolving interfaces (Funchal, 2000)*, volume 1812 of *Lecture Notes in Math.*, pages 1–52. Springer, Berlin, 2003.
- [4] L. Ambrosio and N. Gigli. A user’s guide to optimal transport. In *Modelling and optimisation of flows on networks*, volume 2062 of *Lecture Notes in Math.*, pages 1–155. Springer, Heidelberg, 2013.
- [5] J. R. Baxter and R. V. Chacon. Compactness of stopping times. *Z. Wahrscheinlichkeitstheorie und Verw. Gebiete*, 40(3):169–181, 1977.
- [6] M. Beiglböck, A. M. G. Cox, and M. Huesmann. Optimal transport and Skorokhod embedding. *Invent. Math.*, 208(2):327–400, 2017.
- [7] M. Beiglböck, A. M. G. Cox, and M. Huesmann. The geometry of multi-marginal Skorokhod embedding. *To appear in Probab. Theory Related Fields*, 2019.
- [8] M. Beiglböck, P. Henry-Labordère, and F. Penkner. Model-independent bounds for option prices: a mass transport approach. *Finance Stoch.*, 17(3):477–501, 2013.
- [9] M. Beiglböck, P. Henry-Labordère, and N. Touzi. Monotone martingale transport plans and Skorokhod embedding. *Stochastic Process. Appl.*, 127(9):3005–3013, 2017.
- [10] M. Beiglböck, M. Huesmann, and F. Stebegg. Root to Kellerer. In *Séminaire de probabilités XLVIII*, volume 2168 of *Lecture Notes in Math.*, pages 1–12. Springer, Berlin, 2016.
- [11] M. Beiglböck, T. Lim, and J. Obłój. Dual attainment for the martingale transport problem. *Bernoulli*, 25(3):1640–1658, 2019.
- [12] M. Beiglböck, M. Nutz, and N. Touzi. Complete duality for martingale optimal transport on the line. *Ann. Probab.*, 45(5):3038–3074, 2017.

- [13] D. P. Bertsekas and S. E. Shreve. *Stochastic Optimal Control. The Discrete-Time Case*. Academic Press, New York, 1978.
- [14] S. Biagini, B. Bouchard, C. Kardaras, and M. Nutz. Robust fundamental theorem for continuous processes. *Math. Finance*, 27(4):963–987, 2017.
- [15] B. Bouchard and M. Nutz. Arbitrage and duality in nondominated discrete-time models. *Ann. Appl. Probab.*, 25(2):823–859, 2015.
- [16] H. Brown, D. Hobson, and L. C. G. Rogers. Robust hedging of barrier options. *Math. Finance*, 11(3):285–314, 2001.
- [17] C. Bruggeman and J. Ruf. A one-dimensional diffusion hits points fast. *Electron. Commun. Probab.*, 21(22):1–7, 2016.
- [18] M. Burzoni, M. Frittelli, and M. Maggis. Model-free superhedging duality. *Ann. Appl. Probab.*, 27(3):1452–1477, 2017.
- [19] P. Cheridito, M. Kupper, and L. Tangpi. Representation of increasing convex functionals with countably additive measures. *Preprint arXiv:1502.05763v1*, 2015.
- [20] A. M. G. Cox, Z. Hou, and J. Oblój. Robust pricing and hedging under trading restrictions and the emergence of local martingale models. *Finance Stoch.*, 20(3):669–704, 2016.
- [21] A. M. G. Cox and S. M. Kinsley. Discretisation and duality of optimal skorhod embedding problems. *Stochastic Process. Appl.*, 129:2376–2405, 2019.
- [22] A. M. G. Cox and S. M. Kinsley. Robust hedging of options on a leveraged exchange traded fund. *Ann. Appl. Probab.*, 29(1):531–576, 2019.
- [23] A. M. G. Cox and J. Oblój. Robust pricing and hedging of double no-touch options. *Finance Stoch.*, 15(3):573–605, 2011.
- [24] A. M. G. Cox, J. Oblój, and N. Touzi. The Root solution to the multi-marginal embedding problem: an optimal stopping and time-reversal approach. *Probab. Theory Related Fields*, 173(1-2):211–259, 2019.
- [25] A. M. G. Cox and J. Wang. Root’s barrier: construction, optimality and applications to variance options. *Ann. Appl. Probab.*, 23(3):859–894, 2013.
- [26] C. Czichowsky and W. Schachermayer. Strong supermartingales and limits of nonnegative martingales. *Ann. Probab.*, 44(1):171–205, 2016.
- [27] R. C. Dalang. Sur l’arrêt optimal de processus à temps multidimensionnel continu. In *Seminar on probability, XVIII*, volume 1059 of *Lecture Notes in Math.*, pages 379–390. Springer, Berlin, 1984.
- [28] F. Delbaen and W. Schachermayer. A general version of the fundamental theorem of asset pricing. *Math. Ann.*, 300:463–520, 1994.
- [29] C. Dellacherie and P. A. Meyer. *Probabilities and Potential A*. North Holland, Amsterdam, 1978.
- [30] C. Dellacherie and P. A. Meyer. *Probabilities and Potential B*. North Holland, Amsterdam, 1982.
- [31] S. De Marco and P. Henry-Labordère. Linking vanillas and VIX options: a constrained martingale optimal transport problem. *SIAM J. Financial Math.*, 6(1):1171–1194, 2015.
- [32] Y. Dolinsky and H. M. Soner. Martingale optimal transport and robust hedging in continuous time. *Probab. Theory Related Fields*, 160(1-2):391–427, 2014.

- [33] Y. Dolinsky and H. M. Soner. Martingale optimal transport in the Skorokhod space. *Stochastic Process. Appl.*, 125(10):3893–3931, 2015.
- [34] N. El Karoui, J.-P. Lepeltier, and A. Millet. A probabilistic approach to the reduite in optimal stopping. *Probab. Math. Statist.*, 13(1):97–121, 1992.
- [35] A. Fahim and Y.-J. Huang. Model-independent superhedging under portfolio constraints. *Finance Stoch.*, 20(1):51–81, 2016.
- [36] A. Galichon, P. Henry-Labordère, and N. Touzi. A stochastic control approach to no-arbitrage bounds given marginals, with an application to lookback options. *Ann. Appl. Probab.*, 24(1):312–336, 2014.
- [37] N. Ghoussoub. An integral representation of randomized probabilities and its applications. In *Seminar on Probability, XVI*, volume 920 of *Lecture Notes in Math.*, pages 519–543. Springer, Berlin-New York, 1982.
- [38] N. Ghoussoub, Y.-H. Kim, and T. Lim. Structure of optimal martingale transport in general dimensions. *Ann. Probab.*, 47(1):109–164, 2019.
- [39] N. Ghoussoub, Y.-H. Kim, and A. Z. Palmer. PDE methods for optimal Skorokhod embeddings. *Calc. Var.*, 58(113), 2019.
- [40] N. Gozlan, C. Roberto, P.-M. Samson, and P. Tetali. Kantorovich duality for general transport costs and applications. *J. Funct. Anal.*, 273(11):3327–3405, 2017.
- [41] G. Guo, X. Tan, and N. Touzi. On the monotonicity principle of optimal Skorokhod embedding problem. *SIAM J. Control Optim.*, 54(5):2478–2489, 2016.
- [42] G. Guo, X. Tan, and N. Touzi. Optimal Skorokhod embedding under finitely many marginal constraints. *SIAM J. Control Optim.*, 54(4):2174–2201, 2016.
- [43] I. Gyöngy and D. Šiška. On randomized stopping. *Bernoulli*, 14(2):352–361, 05 2008.
- [44] P. Henry-Labordère, J. Oblój, P. Spoida, and N. Touzi. Maximum maximum of martingales given marginals. *Ann. Appl. Probab.*, 26(1):1–44, 2016.
- [45] P. Henry-Labordère, X. Tan, and N. Touzi. An explicit version of the one-dimensional Brenier’s theorem with full marginals constraint. *Stochastic Process. Appl.*, 126(9):2800–2834, 2016.
- [46] F. Hirsch, C. Profeta, B. Roynette, and M. Yor. *Peacocks and Associated Martingales, with Explicit Constructions*. Springer, Milan, 2011.
- [47] D. Hobson. Robust hedging of the lookback option. *Finance Stoch.*, 2(4):329–347, 1998.
- [48] D. Hobson. The Skorokhod embedding problem and model-independent bounds for option prices. In *Paris-Princeton Lectures on Mathematical Finance 2010*, volume 2003 of *Lecture Notes in Math.*, pages 267–318. Springer, Berlin, 2011.
- [49] D. Hobson. Mimicking martingales. *Ann. Appl. Probab.*, 26(4):2273–2303, 2016.
- [50] D. Hobson and M. Klimmek. Model-independent hedging strategies for variance swaps. *Finance Stoch.*, 16(4):611–649, 2012.
- [51] D. Hobson and M. Klimmek. Robust price bounds for the forward starting straddle. *Finance Stoch.*, 19(1):189–214, 2015.
- [52] D. Hobson and A. Neuberger. Robust bounds for forward start options. *Math.*

- Finance*, 22(1):31–56, 2012.
- [53] D. G. Hobson and J. L. Pedersen. The minimum maximum of a continuous martingale with given initial and terminal laws. *Ann. Probab.*, 30(2):978–999, 2002.
  - [54] M. Huesmann and F. Stebegg. Monotonicity preserving transformations of MOT and SEP. *Stochastic Process. Appl.*, 128(4):1114–1134, 2018.
  - [55] J. Jacod and J. Mémin. Sur un type de convergence intermédiaire entre la convergence en loi et la convergence en probabilité. In *Seminar on Probability, XV (Univ. Strasbourg, Strasbourg, 1979/1980) (French)*, volume 850 of *Lecture Notes in Math.*, pages 529–546. Springer, Berlin-New York, 1981.
  - [56] S. Källblad, X. Tan, and N. Touzi. Optimal Skorokhod embedding given full marginals and Azéma–Yor peacocks. *Ann. Appl. Probab.*, 27(2):686–719, 2017.
  - [57] I. Karatzas and S. E. Shreve. *Brownian Motion and Stochastic Calculus*. Springer, New York, 2nd edition, 1991.
  - [58] H. G. Kellerer. Duality theorems for marginal problems. *Z. Wahrsch. Verw. Gebiete*, 67(4):399–432, 1984.
  - [59] D. Lacker. Dense sets of joint distributions appearing in filtration enlargements, stochastic control, and causal optimal transport. *Preprint arXiv:1805.03185v1*, 2018.
  - [60] R. M. Loynes. Stopping times on Brownian motion: Some properties of Root’s construction. *Z. Wahrscheinlichkeitstheorie und Verw. Gebiete*, 16:211–218, 1970.
  - [61] D. B. Madan and M. Yor. Making Markov martingales meet marginals: with explicit constructions. *Bernoulli*, 8(4):509–536, 2002.
  - [62] R. J. McCann. Existence and uniqueness of monotone measure-preserving maps. *Duke Math. J.*, 80(2):309–323, 1995.
  - [63] J.-F. Mertens. Théorie des processus stochastiques généraux applications aux surmartingales. *Z. Wahrscheinlichkeitstheorie und Verw. Gebiete*, 22:45–68, 1972.
  - [64] J.-F. Mertens. Strongly supermedian functions and optimal stopping. *Z. Wahrscheinlichkeitstheorie und Verw. Gebiete*, 26:119–139, 1973.
  - [65] I. Monroe. On embedding right continuous martingales in Brownian motion. *Ann. Math. Statist.*, 43:1293–1311, 1972.
  - [66] A. Neufeld and M. Nutz. Superreplication under volatility uncertainty for measurable claims. *Electron. J. Probab.*, 18(48):1–14, 2013.
  - [67] M. Nutz. Superreplication under model uncertainty in discrete time. *Finance Stoch.*, 18(4):791–803, 2014.
  - [68] M. Nutz. Robust superhedging with jumps and diffusion. *Stochastic Process. Appl.*, 125(12):4543–4555, 2015.
  - [69] M. Nutz and F. Stebegg. Canonical supermartingale couplings. *Ann. Probab.*, 46(6):3351–3398, 2018.
  - [70] M. Nutz, F. Stebegg, and X. Tan. Multiperiod martingale transport. *Stochastic Process. Appl.*, 130(3):1568–1615, 2020.
  - [71] J. Oblój. The Skorokhod embedding problem and its offspring. *Probab. Surv.*, 1:321–390, 2004.

- [72] K. R. Parthasarathy. *Probability measures on metric spaces*. Probability and Mathematical Statistics, No. 3. Academic Press, New York, 1967.
- [73] A. Pratelli. On the equality between Monge’s infimum and Kantorovich’s minimum in optimal mass transportation. *Ann. Inst. H. Poincaré Probab. Statist.*, 43(1):1–13, 2007.
- [74] S. T. Rachev and L. Rüschendorf. *Mass transportation problems. Vol. I*. Probability and its Applications (New York). Springer-Verlag, New York, 1998. Theory.
- [75] S. T. Rachev and L. Rüschendorf. *Mass transportation problems. Vol. II*. Probability and its Applications (New York). Springer-Verlag, New York, 1998. Applications.
- [76] D. Revuz and M. Yor. *Continuous martingales and Brownian motion*, volume 293 of *Grundlehren der Mathematischen Wissenschaften*. Springer-Verlag, Berlin, third edition, 1999.
- [77] D. H. Root. The existence of certain stopping times on Brownian motion. *Ann. Math. Statist.*, 40:715–718, 1969.
- [78] H. Rost. Skorokhod stopping times of minimal variance. In *Séminaire de Probabilités, X (Première partie, Univ. Strasbourg, Strasbourg, année universitaire 1974/1975)*, volume 511 of *Lecture Notes in Math.*, pages 194–208. Springer, Berlin, 1976.
- [79] A. V. Skorokhod. *Studies in the theory of random processes*. Translated from the Russian by Scripta Technica, Inc. Addison-Wesley Publishing Co., Inc., Reading, Mass., 1965.
- [80] F. Stebegg. Model-independent pricing of Asian options via optimal martingale transport. *Preprint arXiv:1412.1429v1*, 2014.
- [81] X. Tan and N. Touzi. Optimal transportation under controlled stochastic dynamics. *Ann. Probab.*, 41(5):3201–3240, 2013.
- [82] N. Touzi. Martingale inequalities, optimal martingale transport, and robust superhedging. In *Congrès SMAI 2013*, volume 45 of *ESAIM Proc. Surveys*, pages 32–47. EDP Sci., Les Ulis, 2014.
- [83] C. Villani. *Topics in optimal transportation*, volume 58 of *Graduate Studies in Mathematics*. American Mathematical Society, Providence, RI, 2003.
- [84] C. Villani. *Optimal transport, old and new*, volume 338 of *Grundlehren der Mathematischen Wissenschaften*. Springer-Verlag, Berlin, 2009.
- [85] D. Zaev. On the Monge–Kantorovich problem with additional linear constraints. *Math. Notes*, 98(5):725–741, 2015.