OPTIMAL TRANSPORT AND SKOROKHOD EMBEDDING

MATHIAS BEIGLBÖCK, ALEXANDER M. G. COX, AND MARTIN HUESMANN

Abstract. The *Skorokhod embedding problem* is to represent a given probability as the distribution of Brownian motion at a chosen stopping time. Over the last 50 years this has become one of the important classical problems in probability theory and a number of authors have constructed solutions with particular optimality properties. These constructions employ a variety of techniques ranging from excursion theory to potential and PDE theory and have been used in many different branches of pure and applied probability.

We develop a new approach to Skorokhod embedding based on ideas and concepts from *optimal mass transport*. In analogy to the celebrated article of Gangbo and McCann on the geometry of optimal transport, we establish a geometric characterization of Skorokhod embeddings with desired optimality properties. This leads to a systematic method to construct optimal embeddings. It allows us, for the first time, to derive all known optimal Skorokhod embeddings as special cases of one unified construction and leads to a variety of new embeddings. While previous constructions typically used particular properties of Brownian motion, our approach applies to all sufficiently regular Markov processes.

Keywords: Optimal Transport, Skorokhod Embedding, cyclical monotonicity.
Mathematics Subject Classification (2010): Primary 60G42, 60G44; Secondary 91G20.

1. Introduction

Let B be a Brownian motion started in 0 and consider a probability μ on the real line which is centered and has second moment. The Skorokhod embedding problem is to construct a stopping time τ embedding μ into Brownian motion in the sense that

$$B_\tau \text{ is distributed according to } \mu, \quad \mathbb{E}[\tau] < \infty. \tag{SEP}$$

Here, the second condition is imposed to exclude certain undesirable solutions, and can be modified to extend to measures without a second moment. As already demonstrated by Skorokhod [53, 54] in the early 1960s, it is always possible to construct solutions to the problem. Indeed, the survey article [43] of Obłój classifies 21 distinct solutions to (SEP), although this list (from 2004) misses many more recent contributions. A common inspiration for many of these papers is to construct solutions to (SEP) that exhibit additional desirable properties or a distinct internal structure. These have found applications in different fields and various extensions of the original problem have been considered. We refer to [43] (and the 120+ references therein) for a comprehensive account of the field.

Our aim is to develop a new approach to (SEP) based on ideas from optimal transport. Many of the previous developments are thus obtained as applications of one unifying principle (Theorem 1.3) and several difficult problems are rendered tractable. Moreover, our methods can easily handle a number of more general versions of the problem: for example, integrable measures, general starting distributions, and $\mathbb{R}^d$-valued Feller processes.

1.1. A motivating example — Root's construction.

To illustrate our approach we introduce Root's construction, [48], which will serve as inspiration in the rest of the paper. Root's construction is one of the earliest solutions to (SEP), and it is prototypical for many further solutions to (SEP) in that it has a simple *geometric description* and possesses a certain *optimality property* in the class of all solutions.

The first author was supported by the FWF-grants P26736 and Y00782, the third author by the CRC 1060.

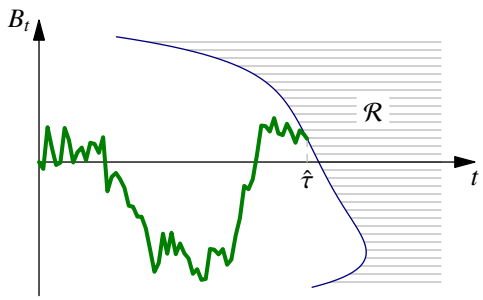


FIGURE 1. Root's solution of (SEP).

Root established that there exists a *barrier* $\mathcal{R}$ (which is essentially unique) such that the Skorokhod embedding problem is solved by the stopping time

$$\tau_{\text{Root}} = \inf\{t \geq 0 : (t, B_t) \in \mathcal{R}\}. \tag{1.1}$$

A barrier is a Borel set $\mathcal{R} \subseteq \mathbb{R}_+ \times \mathbb{R}$ such that $(s, x) \in \mathcal{R}$ and $s < t$ implies $(t, x) \in \mathcal{R}$. The Root construction is distinguished by the following optimality property: among all solutions to (SEP) for a fixed terminal distribution μ, it minimizes $\mathbb{E}[\tau^2]$. For us, the optimality property will be the starting point from which we deduce a geometric characterization of τ_{Root}. To this end, we now formalize the corresponding optimization problem.

1.2. **Optimal Skorokhod Embedding Problem.** We consider the set of *stopped paths*

$$S = \{(f, s) : f : [0, s] \to \mathbb{R} \text{ is continuous}, f(0) = 0\}. \tag{1.2}$$

Throughout the paper we consider a function

$$\gamma : S \to \mathbb{R}.$$

We fix a stochastic basis $\Omega = (\Omega, \mathcal{G}, (\mathcal{G}_t)_{t\geq 0}, \mathbb{P})$ which is sufficiently rich to support a Brownian motion B and a uniformly distributed $\mathcal{G}_0$-random variable, independent of B. The *optimal Skorokhod embedding problem* is to construct a stopping time τ on Ω which optimizes

$$P_\gamma = \inf\left\{\mathbb{E}[\gamma((B_t)_{t\leq\tau}, \tau)] : \ \tau \text{ solves (SEP)}\right\}. \tag{OptSEP}$$

We emphasize that (OptSEP) does not depend on the particular choice of the underlying basis as long as it is rich enough in the above sense, cf. Lemma 3.11 / Section 4.1. We will usually assume that (OptSEP) is *well posed* in the sense that $\mathbb{E}[\gamma((B_t)_{t\leq\tau}, \tau)]$ exists with values in $(-\infty, \infty]$ for all τ which solve (SEP) and is finite for one such τ.

The Root stopping time solves (OptSEP) in the case where $\gamma(f, s) = s^2$. Other examples where the solution is known include functions depending on the running maximum $\gamma((f, s)) := \bar{f}(s) := \max_{t\leq s} f(t)$ or functions of the local time at 0.

The solutions to (SEP) have their origins in many different branches of probability theory, and in many cases, the original derivation of the embedding occurred separately from the proof of the corresponding optimality properties. Moreover, the optimality of a given construction is often not immediate; for example, the optimality property of the Root embedding was first conjectured by Kiefer [36] and subsequently established by Rost [50].

In contrast to existing work, we will start with the optimization problem (OptSEP) and we seek a systematic method to determine the minimizer for a given function γ. To develop a general theory for this optimization problem we interpret stopping times in terms of a transport plan from the Wiener space $(C_0(\mathbb{R}_+), \mathbb{W})$ to the target measure μ, i.e. we want to think of a stopping time τ as transporting the mass of a trajectory $(B_t(\omega))_{t\in\mathbb{R}_+}$ to the point $B_{\tau(\omega)}(\omega) \in \mathbb{R}$. Note that this is *not a coupling* between $\mathbb{W}$ and μ in the usual sense and one cannot directly apply optimal transport theory. Nevertheless the transport perspective provides a powerful intuition that guides us to develop an analogous theory, which in

particular accounts for the adaptedness properties of stopping times. To this end, it is necessary to combine ideas and results from optimal transport with concepts and techniques from stochastic analysis.

As in optimal transport, it is crucial to consider (OptSEP) in a suitably relaxed form, i.e. in (OptSEP) we will optimize over *randomized stopping times* (see Definition 3.7 below). These can be viewed as usual stopping times on a possibly enlarged probability space but in our context it is more natural to interpret them as stopping times of 'Kantorovich-type' (in the sense of optimal transport), i.e. stopping times which terminate a given path not at a single deterministic time instance but according to a distribution.

This relaxation will allow us to transfer many of the convenient properties of classical transport theory to our probabilistic setup. Exactly as in classical transport theory, (OptSEP) can be viewed as a linear optimization problem. The set of couplings in mass transport is compact and similarly the set of all randomized stopping times solving (SEP) on Wiener space is compact in a natural sense. Under the standing assumption that B is defined on a sufficiently rich stochastic basis, these considerations allow us to prove:

Theorem 1.1. *Let $\gamma : S \to \mathbb{R}$ be lsc and bounded from below. Then* (OptSEP) *admits a minimizing stopping time τ.*

Here we can talk about the continuity properties of γ since S possesses a natural Polish topology (cf. (3.1)).

In the language of linear optimization, Theorem 1.1 is a primal problem. It is therefore natural to expect that there exists a corresponding dual problem, and our second main result concerns this duality:

Theorem 1.2. *Let $\gamma : S \to \mathbb{R}$ be lsc and bounded from below, and set*

$$D_\gamma = \sup\left\{ \int \psi(y)\, d\mu(y) : \psi \in C(\mathbb{R}), \exists M, \begin{array}{c} M \text{ is a continuous } \mathcal{G}\text{-martingale}, M_0 = 0 \\ \mathbb{P}-a.s., \forall t \geq 0, M_t + \psi(B_t) \leq \gamma((B_s)_{s\leq t}, t) \end{array} \right\}$$

where M, ψ satisfy $|M_t| \leq a + bt + cB_t^2$, $|\psi(y)| \leq a + by^2$ for some $a, b, c > 0$. Then we have the duality relation

$$P_\gamma = D_\gamma. \tag{1.3}$$

We will prove this result in Section 4, and variants of this result will prove to be important in establishing later results. Theorem 1.2 has close analogues in the literature. In particular, using Hobson's time change argument ([30, 31]), Theorem 1.2 is comparable to the work of Dolinsky and Soner [21, 20]. Similar duality results in a discrete time framework are established by Bouchard and Nutz [9] among others.

1.3. **Geometric Characterization of Optimizers — Monotonicity Principle.** A fundamental idea in optimal transport is that the optimality of a transport plan is reflected by the geometry of its support set. Often this is key to understanding the transport problem. On the level of support sets, the relevant notion is *c-cyclical monotonicity*. The relevance of this concept for the theory of optimal transport has been fully recognized by Gangbo and McCann [24], based on earlier work of Knott and Smith [37] and Rüschendorf [51, 52] among others.

Inspired by these results, we establish a *monotonicity principle* which links the optimality of a stopping time τ with 'geometric' properties of τ. Combined with Theorem 1.1, this principle will turn out to be surprisingly powerful. For the first time, *all* the known solutions to (SEP) with optimality properties can be established through one unifying principle. Moreover, the monotonicity principle allows us to treat the optimization problem (OptSEP) in a systematic manner, generating further embeddings as a by-product.

Our third main result states:

Theorem 1.3 (Monotonicity Principle)**.** *Let $\gamma : S \to \mathbb{R}$ be Borel measurable. Suppose that* (OptSEP) *is well posed and τ is an optimizer. Then there exists a γ-*monotone *(cf.*

Definition 1.5 below) Borel set $\Gamma \subseteq S$ *such that* $\mathbb{P}$*-a.s.*

$$((B_t)_{t\leq\tau}, \tau) \in \Gamma \ . \tag{1.4}$$

If (1.4) holds, we will loosely say that Γ *supports* τ. The significance of Theorem 1.3 is that it links the optimality of the stopping time τ with a particular property of the set Γ, i.e. γ-monotonicity. In applications, the latter turns out to be much more tangible. We emphasize that we do not require continuity assumptions on γ in this result. This will be important when we apply our results.

To link the optimality of a stopping time with properties of the set Γ we consider the minimization problem (OptSEP) on a pathwise level. Consider two paths $(f, s), (g, t) \in S$ which end at the same value, i.e. $f(s) = g(t)$. We want to determine which of the two paths should be *stopped* and which one should be allowed to *go on* further, bearing in mind that we try to minimize $\mathbb{E}[\gamma((B_s)_{s\leq\tau}, \tau)]$. To make this definition formal, we need to perform an operation at the level of individual paths. We will write $f \oplus h$ for the concatenation of the two paths $(f, s), (h, u) \in S$, specifically:

$$(f \oplus h)(r) := \begin{cases} f(r) & r \leq s \\ f(s) + h(r - s) & r \in (s, s + u] \end{cases} .$$

Then we set

$$\gamma^{(f,s)\oplus}(h, u) := \gamma(f \oplus h, s + u). \tag{1.5}$$

We will call $((f, s), (g, t))$ a *stop-go* pair if it is advantageous to *stop* (f, s) and to *go on* after (g, t) in the following sense:

Definition 1.4. *The pair* $((f, s), (g, t)) \in S \times S$ *is a* stop-go pair*, written* $((f, s), (g, t)) \in \mathsf{SG}$*, iff* $f(s) = g(t)$ *and*

$$\mathbb{E}\left[\gamma^{(f,s)\oplus}\left((B_u)_{u\leq\sigma}, \sigma\right)\right] + \gamma(g, t) \quad > \quad \gamma(f, s) + \mathbb{E}\left[\gamma^{(g,t)\oplus}\left((B_u)_{u\leq\sigma}, \sigma\right)\right] \tag{1.6}$$

for every $(\mathcal{F}_t^B)_{t\geq 0}$*-stopping time* σ *which satisfies* $0 < \mathbb{E}[\sigma] < \infty$ *and for which both sides of* (1.6) *are well defined and the left hand side is finite.*

Here $(\mathcal{F}_t^B)_{t\geq 0}$ denotes the natural filtration generated by the Brownian motion B. A consequence of considering only $(\mathcal{F}_t^B)_{t\geq 0}$-stopping times is that the set SG does not depend on the particular choice of the underlying stochastic basis.

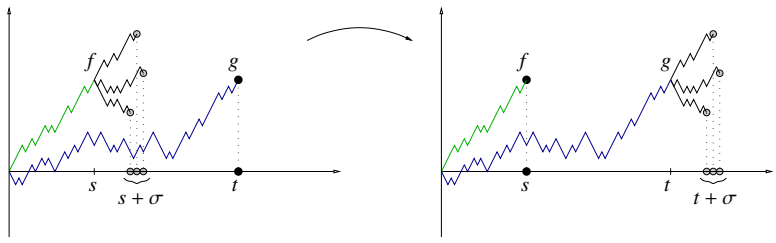


FIGURE 2. The left hand side of (1.6) corresponds to averaging the function γ over the stopped paths on the left picture; the right hand side to averaging the function γ over the stopped paths on the right picture.

We note that a swapping of paths (as illustrated in Figure 2) was used by Hobson [31, p 34] to provide a heuristic derivation of the optimality properties of the Root embedding. Indeed Hobson's approach was the starting point of the present paper.

Recalling (1.4), we see that the set $\Gamma \subseteq S$ contains the *stopped* paths: that is, a path (g, t) is in Γ if there is some possibility that the optimal stopping rule decides to stop at time t having observed the path $(g(u))_{u\in[0,t]}$. In addition, we need to consider those paths which we observe as the initial section of a longer, stopped, path: these are the *going* paths

$$\Gamma^< := \{(f, s) : \exists(\tilde{f}, \tilde{s}) \in \Gamma, s < \tilde{s} \text{ and } f \equiv \tilde{f} \text{ on } [0, s]\} \ . \tag{1.7}$$

We can now formally introduce γ-monotonicity.

Definition 1.5. *A set $\Gamma \subseteq S$ is called γ-monotone iff $\Gamma^< \times \Gamma$ contains no stop-go pairs, i.e.*

$$\mathsf{SG} \cap (\Gamma^< \times \Gamma) = \emptyset. \tag{1.8}$$

By the monotonicity principle, Theorem 1.3, an optimal stopping time is supported by a set Γ such that $\Gamma^< \times \Gamma$ contains no stop-go pair $((f,s),(g,t))$. Intuitively, such a pair gives rise to a possible modification, improving the given stopping rule: as $f(s) = g(t)$, we can imagine stopping the path (f,s) at time s, and allowing (g,t) to go on by transferring all paths which extend (f,s), the 'remaining lifetime', onto (g,t), which is now *going* (see Figure 2). By (1.6) this guarantees an improved value of P_γ, contradicting the optimality of our stopping rule. Observe that the condition $f(s) = g(t)$ is what guarantees that a modified stopping rule still embeds the measure μ. In Section 2 below we will briefly indicate how the monotonicity principle can be used to derive existing solutions to the Skorokhod embedding problem as well as a whole family of novel solutions to the Skorokhod embedding problem; many further examples will be provided in Section 6.

Importantly, the transport-based approach readily admits a number of strong generalizations and extensions. With only minor changes the existence result, Theorem 1.1, the duality result, Theorem 1.2, and the monotonicity principle, Theorem 1.3 below, extend to general starting distributions and Brownian motion in $\mathbb{R}^d$, and more generally to sufficiently regular Markov processes; see Sections 5 and 7. This is notable since previous constructions usually exploit rather specific properties of Brownian motion.

The monotonicity principle, Theorem 1.3, represents the culmination of the three main results, and the proof of this result will be the most complex part of this paper, requiring substantial preparation in order to combine the relevant concepts from stochastic analysis and optimal transport. The preparation and proof of this result will therefore comprise the majority of the paper. In fact the proof will automatically imply a stronger version (Theorem 5.7) of Theorem 1.3. For our applications, it will also be helpful to introduce a version of this result which incorporates a secondary optimization, Theorem 5.16.

The 'classical' optimal transport version of Theorem 1.3 can be established through fairly direct arguments, at least in a reasonably regular setting, cf. [3, Thms. 3.2, 3.3] and [57, p. 88f]. However, these approaches do not extend easily to our setup: stopping times are of course not couplings in the usual sense and there is no reason for particular combinatorial manipulations to carry over in a direct fashion. Another substantial difference is that the procedure of transferring paths described below Definition 1.5 necessarily refers to a *continuum* of paths while the classical notion of cyclical monotonicity is concerned with rearrangements along finite cycles. The argument given subsequently is more in the spirit of [6, 8] and requires a fusion of ideas from optimal transport and stochastic analysis.

1.4. **New Horizons.** The results presented in this paper are limited to the case of the classical Skorokhod embedding problem for Markov processes with continuous paths. However we believe that our methods are sufficiently general that a number of interesting and important extensions, which previously would have been intractable, may now be within reach:

(1) **Markov processes:** The results presented in this paper should extend to a more general class of Markov processes with càdlàg paths. The main technical issues this would present lie in the generalization of the results in Section 3, where the specific structure of the space of continuous paths is exploited.
(2) **Multiple path-swapping:** In our monotonicity principle, Theorem 1.3, we consider the impact of swapping mass from a single unstopped path onto a single stopped path, and argue that if this improves the objective γ on average, then the stopping time in question was not optimal. In classical optimal transport, it is known that single swapping is not sufficient to guarantee optimality; rather, one needs to consider the impact of allowing a finite 'cycle' of swaps to occur, and moreover, that this is both a necessary and sufficient condition for optimality. It is natural to conjecture that a similar result applies in the present setup.

(3) **Multiple marginals:** A natural generalization of the Skorokhod embedding problem is to consider the case where a sequence of measures, $\mu_1, \mu_2, \dots, \mu_n$ are given, and the aim is to find a sequence of stopping times $\tau_1 \leq \tau_2 \leq \cdots \leq \tau_n$ such that $B_{\tau_k} \sim \mu_k$, and such that the chosen sequence of stopping times minimizes $\mathbb{E}[\gamma((B_t)_{t\leq\tau_n}, \tau_1, \dots, \tau_n)]$ for a suitable function γ. In this setup, it is natural to ask whether there exists a suitable monotonicity principle, corresponding to Theorem 1.3.

(4) **Constrained embedding problems:** In this paper, we consider classical embedding problems, where the optimization is carried out over the class of solutions to (SEP). However, in many natural applications, one needs to further consider the class of *constrained* embedding problems: for example, where one minimizes some function over the class of embeddings which also satisfy a restriction on the probability of stopping after a given time. It would be natural to derive generalizations of our duality results, and a corresponding monotonicity principle for such problems.

1.5. **Background.** Since the first solution to (SEP) by Skorokhod [54] the embedding problem has received frequent attention in the literature, with new solutions appearing regularly, and exploiting a number of different mathematical tools. Many of these solutions also prove to be, by design or accident, solutions of (OptSEP) for a particular choice of γ, e.g. [48, 50, 4, 33, 56, 45]. The survey [43] is a comprehensive account of all the solutions to (SEP) up to 2004 and references many articles which use or develop solutions to the Skorokhod embedding problem. More recently, novel twists on the classical Skorokhod embedding problem have been investigated by: Last et. al. [38], who consider the closely related problem of finding unbiased shifts of Brownian motion (and where there are also natural connections to optimal transport); Hirsch et. al. [29], who have used solutions to the Skorokhod embedding problem to construct Peacocks; and Gassiat et. al. [25], who have exploited particular properties of Root's solution to construct efficient numerical schemes for SDEs.

The Skorokhod embedding problem has also recently received substantial attention from the mathematical finance community. This goes back to an idea of Hobson [30]: through the Dambis-Dubins-Schwarz Theorem, the optimization problems (OptSEP) are related to the pricing of financial derivatives, and in particular to the problem of *model-risk*. We refer the reader to the survey article [31] for further details.

Recently there has been much interest in optimal transport problems where the transport plan must satisfy additional martingale constraints. Such problems arise naturally in the financial context, but are also of independent mathematical interest, for example — mirroring classical optimal transport — they have important consequences for the study of martingale inequalities (see e.g. [9, 28, 44]). The first papers to study such problems include [32, 7, 23, 19], and this field is commonly referred to as *martingale optimal transport*. The Skorokhod embedding problem has been considered in this context by Galichon et. al. in [23]; through a stochastic control problem they recover the Azéma-Yor solution of the Skorokhod embedding problem. Notably, their approach is very different from the one pursued in the present paper.

1.6. **Outline of the Article.** In Section 2 we establish the Root and the Rost embeddings as a consequence of Theorems 1.1 and 1.3, as well as constructing a family of new embeddings. The results presented in this section are intended as a motivation for the rest of the paper. In the derivation of these embeddings we highlight the interplay between arguments of a probabilistic nature, and arguments relating to the pathwise space S introduced in (1.2). A major benefit of working in these two separate domains is that it is typically relatively easy to prove pointwise statements in the setup of the space S; on the other hand, the associated probabilistic arguments are usually straightforward. However neither set of arguments naturally transfers to the other setup.

The link between these distinct domains is provided by Theorems 1.1 and 1.2, and in particular the monotonicity principle Theorem 1.3 which we establish in Sections 3 to 5. In Section 3, we introduce a framework that allows us to view classical probabilistic concepts on the pathwise space S and establish a number of auxiliary results that will be needed later on. In Section 4 we prove our first two main results. As in the transport case, Theorem 1.1 will be a simple consequence of lower semi-continuity plus compactness of the set of solutions to the Skorokhod problem. To establish Theorem 1.2, we use classical duality results from optimal transport. In Section 5 we prove Theorem 1.3 based on a combination of arguments from optimal transport with Choquet's capacitability theorem and ingredients from stochastic analysis.

In Section 6 we use our results to establish all known solutions to (OptSEP) as well as further embeddings. We also give an example in which (OptSEP) admits only optimizers depending on additional randomization. For readers who are mainly interested in these applications, it should be possible to read this section immediately after Section 2.

In Section 7 we describe a number of extensions of our previous results. In particular we consider general starting distributions and show that our main results extend to continuous Feller processes under certain assumptions which we are able to verify for a large class of processes. As a special case of the results in this section, we also show that, as usual, the moment condition on μ can be dropped when the second condition in (SEP) is recast in terms of uniform integrability resp. minimality (cf. (2.1)).

1.7. **Frequently used notation.**

- The set of (sub-)probability measures on a space X is denoted by $\mathcal{P}(\mathsf{X})$ / $\mathcal{P}^{\leq 1}(\mathsf{X})$.
- For a measure ξ on X we write $f(\xi)$ for the push-forward of ξ under $f : \mathsf{X} \to \mathsf{Y}$.
- We use $\xi(f)$ as well as $\int f \, d\xi$ to denote the integral of a function f against a measure ξ.
- Stochastic processes are usually denoted by capital letters like X, Y, Z.
- $C_x(\mathbb{R}_+)$ denotes the continuous functions starting in x; $C(\mathbb{R}_+) = \bigcup_{x\in\mathbb{R}} C_x(\mathbb{R}_+)$.
- The set of stopped paths is $S = \{(f, s) : f : [0, s] \to \mathbb{R}$ is continuous, $f(0) = 0\}$ and we define $r : C_0(\mathbb{R}_+) \times \mathbb{R}_+ \to S$ by $r(\omega, t) := (\omega_{\restriction[0,t]}, t)$.
- For $\Gamma \subseteq S$ we set $\Gamma^< := \{(f, s) : \exists(\tilde{f}, \tilde{s}) \in \Gamma, s < \tilde{s}$ and $f \equiv \tilde{f}$ on $[0, s]\}$.
- For $(f, s) \in S$ we write $\bar{f} = \sup_{r\leq s} f(r)$, $\underline{f} = \inf_{r\leq s} f(r)$ and $f^* = \sup_{r\leq s} |f(r)|$.
- We use $\oplus$ for the concatenation of paths: depending on the context the arguments may be elements of S, $C_0(\mathbb{R}_+)$ or $C_0(\mathbb{R}_+) \times \mathbb{R}_+$.
- If F is a function on S resp. $C_0(\mathbb{R}_+) \times \mathbb{R}_+$ and $(f, s) \in S$ we set $F^{(f,s)\oplus}(y) := F((f, s) \oplus y)$, where y may be an element of S, $C_0(\mathbb{R}_+)$, or $C_0(\mathbb{R}_+) \times \mathbb{R}_+$.
- $\mathbb{W}$ denotes Wiener measure; $\mathcal{F}^0$ ($\mathcal{F}^a$) the natural (augmented) filtration on $C_0(\mathbb{R}_+)$.
- Two commonly used probability spaces are $(\Omega, \mathcal{G}, (\mathcal{G}_t)_{t\geq 0}, \mathbb{P})$, which is an arbitrary probability space, on which there exists a process B which is Brownian motion, and sometimes also a $\mathcal{G}_0$-random variable Y which is uniformly distributed on $[0, 1]$. On this space, the natural filtration generated by the process B is denoted by $(\mathcal{F}^B_t)_{t\geq 0}$ In addition, we sometimes refer to the space $(\overline{C}_0(\mathbb{R}_+), \bar{\mathcal{F}}, (\bar{\mathcal{F}}_t)_{t\geq 0}, \overline{\mathbb{W}})$, which is the product space $\overline{C}_0(\mathbb{R}_+) = C_0(\mathbb{R}_+) \times [0, 1]$ equipped with a suitable filtration (see the discussion above Theorem 3.8 for further details) and the product measure $\overline{\mathbb{W}} = \mathbb{W} \otimes \mathcal{L}$ of Wiener and Lebesgue measure.

Acknowledgement. The authors thank Julio Backhoff, Walter Schachermayer, and Nizar Touzi for useful discussions. We are particularly indebted to the anonymous referees and Manu Eder for many helpful suggestions which had a significant impact on the final version of this article.

2. Particular embeddings

In this section we explain how Theorem 1.3 can be used to derive particular solutions to the Skorokhod embedding problem, (SEP), using the optimization problem (OptSEP).

For much of the paper, we consider (SEP) for measures μ where $\int x^2\,\mu(dx) < \infty$. This constraint can be weakened to require only the first moment to be finite, subject to the restriction that the stopping time is *minimal*: that is, if τ is a stopping time such that $B_\tau \sim \mu$, then for any stopping time τ',

$$B_{\tau'} \sim \mu \text{ and } \tau' \leq \tau \text{ implies } \tau' = \tau \text{ a.s.} \tag{2.1}$$

In the case where μ has a second moment, minimality and $\mathbb{E}[\tau] < \infty$ are equivalent. We emphasize that, *mutatis mutandis*, all of our results are valid in this more general setup, see Section 7. Recall that we are working on a stochastic basis which is rich enough to support a Brownian motion and a uniformly distributed random variable.

2.1. **The Root embedding.** We recall the definition of the Root embedding, τ_{Root}, from (1.1), and we wish to recover Root's result ([48]) from an optimization problem. Remember that, according to Root's terminology, a (closed) set $\mathcal{R} \subseteq \mathbb{R}_+ \times \mathbb{R}$ is a *barrier* if $(s, x) \in \mathcal{R}$ implies $(t, x) \in \mathcal{R}$ whenever $t > s$. Then Root's construction of a solution to the Skorokhod embedding problem can be summarized as follows:

Theorem 2.1. *Let $\gamma(f, t) = h(t)$, where $h : \mathbb{R}_+ \to \mathbb{R}$ is a strictly convex function such that* (OptSEP) *is well posed. Then a minimizer of* (OptSEP) *exists, and moreover for any minimizer $\hat{\tau}$, there exists a barrier $\mathcal{R}$ such that $\hat{\tau} = \inf\{t \geq 0 : (t, B_t) \in \mathcal{R}\}$. In particular the Skorokhod embedding problem has a solution of barrier type as in* (1.1).

Proof. **Step 1.** We first pick — by Theorem 1.1 — a stopping time $\hat{\tau}$ which attains P_γ. By Theorem 1.3 there exists a set $\Gamma \subseteq S$ such that $((B_s)_{s\leq\hat{\tau}}, \hat{\tau}) \in \Gamma$ almost surely, and such that $(\Gamma^< \times \Gamma) \cap \mathsf{SG} = \emptyset$.
Step 2. Next, consider paths $(f, s), (g, t) \in S$ such that $f(s) = g(t)$. We consider when $((f, s), (g, t)) \in \mathsf{SG}$, i.e. under which conditions (f, s) should be stopped and Brownian motion should continue to go after (g, t). In the present case (1.6) amounts to

$$\mathbb{E}[h(s + \sigma)] + h(t) \quad > \quad h(s) + \mathbb{E}[h(t + \sigma)]. \tag{2.2}$$

Thus, by strict convexity of h, $((f, s), (g, t)) \in \mathsf{SG}$ iff $t < s$. We define two barriers by

$$\mathcal{R}_{\text{CL}} := \{(s, x) : \exists (g, t) \in \Gamma, g(t) = x, t \leq s\},$$
$$\mathcal{R}_{\text{OP}} := \{(s, x) : \exists (g, t) \in \Gamma, g(t) = x, t < s\}.$$

Fix $(g, t) \in \Gamma$. Then we have $(t, g(t)) \in \mathcal{R}_{\text{CL}}$. Suppose for contradiction that $\inf\{s \in [0, t] : (s, g(s)) \in \mathcal{R}_{\text{OP}}\} < t$. Then there exists $s < t$ such that $(f, s) := (g_{\restriction[0,s]}, s) \in \Gamma^<$ and $(s, f(s)) \in \mathcal{R}_{\text{OP}}$. By definition of $\mathcal{R}_{\text{OP}}$, it follows that there exists another path $(k, u) \in \Gamma$ such that $k(u) = f(s)$ and $u < s$. But then $((f, s), (k, u)) \in \mathsf{SG} \cap (\Gamma^< \times \Gamma)$ which cannot be the case. Hence,

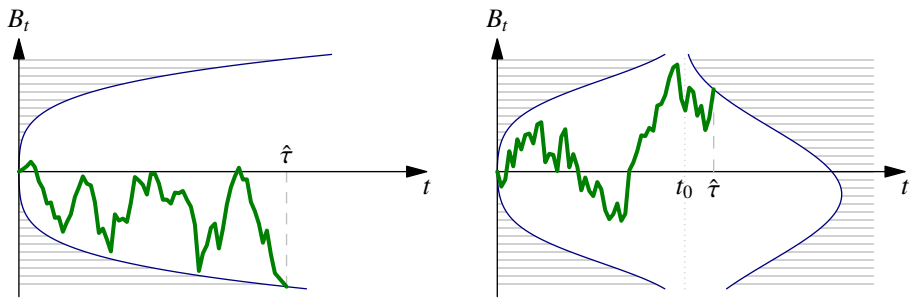
$$(g, t) \in \Gamma \implies \inf\{s \in [0, t] : (s, g(s)) \in \mathcal{R}_{\text{CL}}\} \leq t \leq \inf\{s \in [0, t] : (s, g(s)) \in \mathcal{R}_{\text{OP}}\}.$$

Step 3. Now consider $\omega \in \Omega$ such that $(g, t) = \left((B_s(\omega))_{s\leq\hat{\tau}(\omega)}, \hat{\tau}(\omega)\right) \in \Gamma$. Then it follows immediately that:

$$\tau_{\text{CL}}(\omega) := \inf\{s : (s, B_s(\omega)) \in \mathcal{R}_{\text{CL}}\} \leq \hat{\tau}(\omega) \leq \inf\{s : (s, B_s(\omega)) \in \mathcal{R}_{\text{OP}}\} =: \tau_{\text{OP}}(\omega). \tag{2.3}$$

We finally observe that $\tau_{\text{CL}} = \tau_{\text{OP}}$ a.s. by the strong Markov property, and the fact that one-dimensional Brownian motion immediately returns to its starting point. □

A consequence of this proof is that (on a given stochastic basis) there exists exactly one solution of the Skorokhod embedding problem which minimizes $\mathbb{E}[h(\tau)]$; this property was first established in [50], together with the optimality property of Root's solution. To see this, assume that minimizers τ_1 and τ_2 are given. Then we can use an independent coin-flip to define a new minimizer $\bar{\tau}$ which is with probability $1/2$ equal to τ_1 and with probability $1/2$ equal to τ_2. By Theorem 2.1, $\bar{\tau}$ is of barrier type and hence $\tau_1 = \tau_2$.



(A) The Rost construction (B) The Cave construction

FIGURE 3. The barriers corresponding to the Rost and Cave embeddings

Remark 2.2. We highlight here the nature of the proof of Theorem 2.1. The proof divides into three steps, two of these steps (Steps 1 and 3) being probabilistic in nature, making arguments about random variables on a particular probability space. The second step, however, is purely a pointwise argument about the properties of subsets of Γ in relation to the function γ which we look to optimize. The latter arguments are *not* probabilistic in nature.

Remark 2.3. The following argument, due to Loynes [39], can be used to argue that barriers are unique in the sense that if two barriers solve (SEP), then their hitting times must be equal. Suppose that $\mathcal{R}$ and $\mathcal{S}$ are both *closed* barriers which embed μ. Note that we can take the closed barriers without altering the stopping properties. Consider the barrier $\mathcal{R}\cup\mathcal{S}$: let $A \subseteq \Omega_{\mathcal{R}} := \{x : (t,x) \in \mathcal{S} \implies (t,x) \in \mathcal{R}\}$. Then $\mathbb{P}(B_{\tau_{\mathcal{R}\cup\mathcal{S}}} \in A) \leq \mathbb{P}(B_{\tau_{\mathcal{R}}} \in A) = \mu(A)$. Similarly, for $A' \subseteq \Omega_{\mathcal{S}} := \{x : (t,x) \in \mathcal{R} \implies (t,x) \in \mathcal{S}\}$, $\mathbb{P}(B_{\tau_{\mathcal{R}\cup\mathcal{S}}} \in A') \leq \mathbb{P}(B_{\tau_{\mathcal{S}}} \in A') = \mu(A')$. Since $\mu(\Omega_{\mathcal{R}} \cup \Omega_{\mathcal{S}}) = 1$, $\tau_{\mathcal{R}\cup\mathcal{S}}$ embeds μ.

It is known (see Monroe [41]) that, when μ has a second moment, the second condition in (SEP), $\mathbb{E}[\tau] < \infty$ is equivalent to minimality of the stopping time (recall (2.1)). It immediately follows from the argument above that if the barriers $\mathcal{R}$ and $\mathcal{S}$ solve (SEP), then $\tau_{\mathcal{R}} = \tau_{\mathcal{S}}$ a.s. With minor modifications the argument of Loynes also applies to the Rost solution discussed below as well as to a number of further classical embeddings presented in Section 6 below.

In Section 6.3 we will prove generalizations of Theorem 2.1 which admit similar conclusions in $\mathbb{R}^d$ and for general initial distributions.

2.2. **The Rost embedding.** A set $\mathcal{R} \subseteq \mathbb{R}_+ \times \mathbb{R}$ is an *inverse barrier* if $(s,x) \in \mathcal{R}$ and $s > t$ implies that $(t,x) \in \mathcal{R}$. It has been shown by Rost [50] that under the condition $\mu(\{0\}) = 0$ there exists an inverse barrier such that the corresponding hitting time (in the sense of (1.1)) solves the Skorokhod problem. It is not hard to see that without this condition some additional randomization is required. We derive this using an argument almost identical to the one above.

Theorem 2.4. *Suppose $\mu(\{0\}) = 0$. Let $\gamma(f,t) = h(t)$, where $h : \mathbb{R}_+ \to \mathbb{R}_+$ is a strictly* concave *function such that* (OptSEP) *is well posed. Then a minimizer $\hat{\tau}$ of* (OptSEP) *exists, and moreover for any minimizer $\hat{\tau}$, there exists an inverse barrier $\mathcal{R}$ such that $\hat{\tau} =$* $\inf\{t \geq 0 : (t, B_t) \in \mathcal{R}\}$. *In particular the Skorokhod embedding problem has a solution which is the hitting time of an inverse-barrier.*

Proof. Our proof follows closely the proof of Theorem 2.1. In particular, Steps 1 and 2 can be carried out almost verbatim to get an optimizer $\hat{\tau}$ and a *γ-monotone* set $\Gamma \subseteq S$ such that $\mathbb{P}(((B_t)_{t\leq\hat{\tau}}, \hat{\tau}) \in \Gamma) = 1$. By concavity of h, the set of stop-go pairs is now given by

$$\mathsf{SG} = \{((f,s),(g,t)) \in S \times S : f(s) = g(t), s < t\}.$$

We remove all paths (f, s) with $f(s) = 0$ from Γ, as $\mu(\{0\}) = 0$ this does not alter the full support property (or the γ-monotone property). Next we define inverse barriers by

$$\mathcal{R}_{\text{OP}} := \{(s, x) : \exists (g, t) \in \Gamma, g(t) = x, s < t\},$$
$$\mathcal{R}_{\text{CL}} := \{(s, x) : \exists (g, t) \in \Gamma, g(t) = x, s \leq t\}.$$

Denoting the respective hitting times by τ_{OP} and τ_{CL} the argument familiar from the Root case yields $\tau_{\text{CL}} \leq \hat{\tau} \leq \tau_{\text{OP}}$ a.s. and it remains to show $\tau_{\text{CL}} = \tau_{\text{OP}}$ a.s. The argument is slightly more involved than in the Root case but again entirely probabilistic:

We define $b(t) := \inf\{x > 0 : (t, x) \in \mathcal{R}_{\text{CL}}\}$, $c(t) := \sup\{x < 0 : (t, x) \in \mathcal{R}_{\text{CL}}\}$ and note that

$$\inf\{t > 0 : B_t \notin (c(t), b(t))\} \leq \tau_{\text{CL}} \leq \tau_{\text{OP}} \leq \inf\{t > 0 : B_t \notin [c(t), b(t)]\}.$$

Concentrating on the function b, we have for $\varepsilon > 0$

$$\underbrace{\inf\{t > 0 : B_t \geq b(t)\}}_{=:\sigma_b} \leq \underbrace{\inf\{t > 0 : B_t > b(t)\}}_{=:\sigma_b^+} \leq \underbrace{\inf\{t > 0 : B_t - \varepsilon t \geq b(t)\}}_{=:\sigma_b^\varepsilon}.$$

By Girsanov's Theorem, $\lim_{\varepsilon \to 0} \mathbb{P}(\sigma_b^\varepsilon \leq t) = \mathbb{P}(\sigma_b \leq t)$ for each $t \in \mathbb{R}_+$ hence $\sigma_b^+ = \sigma_b$ a.s.

Arguing likewise on c, we obtain $\tau_{\text{CL}} = \tau_{\text{OP}}$ a.s. □

As in the case of the Root embedding we obtain that the minimizer of $\mathbb{E}[h(\tau)]$ is unique.

2.3. **The cave embedding.** In this section we give an example of a new embedding that can be derived from Theorem 1.3. It can be seen as a unification of the Root and Rost embeddings. A set $\mathcal{R} \subseteq \mathbb{R}_+ \times \mathbb{R}$ is a *cave barrier* if there exists $t_0 \in \mathbb{R}_+$, an inverse barrier $\mathcal{R}^0 \subseteq [0, t_0] \times \mathbb{R}$ and a barrier $\mathcal{R}^1 \subseteq [t_0, \infty) \times \mathbb{R}$ such that $\mathcal{R} = \mathcal{R}^0 \cup \mathcal{R}^1$. We will show that there exists a cave barrier such that the corresponding hitting time (in the sense of (1.1)) solves the Skorokhod problem. We derive this using an argument similar to the one above:

Fix $t_0 \in \mathbb{R}$ and pick a continuous function $\varphi : \mathbb{R}_+ \to [0, 1]$ such that

- $\varphi(0) = 0, \lim_{t\to\infty} \varphi(t) = 0, \varphi(t_0) = 1$
- φ is strictly concave on $[0, t_0]$
- φ is strictly convex on $[t_0, \infty)$.

It follows that φ is strictly increasing on $[0, t_0]$ and strictly decreasing on $[t_0, \infty)$.

Theorem 2.5 (Cave embedding)**.** *Suppose $\mu(\{0\}) = 0$. Let $\gamma(f, t) = \varphi(t)$. Then a minimizer $\hat{\tau}$ of* (OptSEP) *exists, and moreover for any minimizer $\hat{\tau}$, there exists a cave barrier $\mathcal{R}$ such that $\hat{\tau} = \inf\{t \geq 0 : (t, B_t) \in \mathcal{R}\}$. In particular the Skorokhod embedding problem has a solution which is the hitting time of a cave barrier.*

Since this construction does not already appear in the literature, we emphasize that the result remains true for integrable (centered) measures μ (see Section 7).

Proof of Theorem 2.5. Note that since φ is bounded, the problem (OptSEP) is well posed. Following the steps of the proofs of Theorems 2.1 and 2.4, we find an optimizer $\hat{\tau}$ and a γ-monotone set $\Gamma \subseteq S$ such that $\mathbb{P}((B_t)_{t\leq\hat{\tau}}, \hat{\tau}) \in \Gamma) = 1$. The set of stop-go pairs is given by

$$\mathsf{SG} = \{((f, s), (g, t)) \in S \times S : f(s) = g(t); s < t \leq t_0 \text{ or } t_0 \leq t < s\}.$$

Indeed, for $s < t \leq t_0$ and any $(h, r) \in S$ we have

$$\begin{aligned} \gamma((f \oplus h, s + r)) + \gamma((g, t)) \ &> \ \gamma((f, s)) + \gamma((g \oplus h, t + r)) \\ \Leftrightarrow \quad \varphi(s + r) - \varphi(s) \ &> \ \varphi(t + r) - \varphi(t) \end{aligned}$$

which holds iff $t \mapsto \varphi(t + r) - \varphi(t)$ is strictly decreasing on $[0, t_0]$ for all $r > 0$. If $t + r, t \in [0, t_0]$ this follows from concavity of φ. In the case that $t \leq t_0, t + r > t_0$ this follows since φ' is strictly positive on $[0, t_0)$ and strictly negative on (t_0, ∞). The case $t_0 \leq t < s$ can be established similarly.

Then, we define an 'open' cave barrier by

$$\mathcal{R}^0_{\text{OP}} := \{(t, x) : \exists (f, s) \in \Gamma, t < s \leq t_0\}, \quad \mathcal{R}^1_{\text{OP}} := \{(t, x) : \exists (f, s) \in \Gamma, t_0 \leq s < t\}$$

and $\mathcal{R}_{\text{OP}} := \mathcal{R}^0_{\text{OP}} \cup \mathcal{R}^1_{\text{OP}}$ (resp. a 'closed' cave barrier where we allow $t \leq s$ and $s \leq t$ in $\mathcal{R}^0_{\text{CL}}$ and $\mathcal{R}^1_{\text{CL}}$ resp.). We denote the corresponding hitting time by $\tau_{\mathcal{R}_{\text{OP}}} = \tau_{\mathcal{R}^0_{\text{OP}}} \wedge \tau_{\mathcal{R}^1_{\text{OP}}}$ (resp. $\tau_{\mathcal{R}_{\text{CL}}}$).

By the same argument as for the Root and Rost embeddings it then follows that $\tau_{\mathcal{R}_{\text{CL}}} \leq \hat{\tau} \leq \tau_{\mathcal{R}_{\text{OP}}}$ a.s. and also that $\tau_{\mathcal{R}_{\text{CL}}} = \tau_{\mathcal{R}_{\text{OP}}}$ a.s., proving the claim. □

2.4. **Remarks.** In Section 6.3 we will show that the arguments above can be adapted to prove the existence of Rost and Root embeddings in a more general setting. Specifically, in Sections 6 and 7 we will show that this approach generalizes to a multi-dimensional setup and (sufficiently regular) Markov processes. In the case of the Root embedding it does not matter for the argument whether the starting distribution is a Dirac in 0 as in our setup or a more general distribution λ. For the Rost embedding a general starting distribution is slightly more difficult. In the case where λ and μ have common mass, then it may be the case that $\text{proj}_{\mathbb{R}_+}(\mathcal{R}_{\text{CL}} \cap (A \times \mathbb{R}_+)) = \{0\}$ for some set A — that is, all paths which stop at $x \in A$ do so at time zero. In this case it is possible that $\hat{\tau} < \tau_{\text{OP}}$ when the process starts in A, and in general, some proportion of the paths starting on A must be stopped instantly. As a result, in the case of general starting measures, independent randomization is necessary. In the Rost case, it is also straightforward to compute the independent randomization which preserves the embedding property.

Other recent approaches to the Root and Rost embeddings can be found in [25, 26, 14, 13]. These papers largely exploit PDE techniques, and as a consequence, are able to produce more explicit descriptions of the barriers, however the methods tend to be highly specific to the problem under consideration.

3. Preliminaries on stopping times and filtrations

A key feature of this article is that we are taking a non-standard perspective on stopping times; the main purpose of this section is to provide a convenient framework. To this end, we need to discuss connections between common notions defined on an arbitrary probability space and their related notions defined on the canonical path space $C_0(\mathbb{R}_+)$ and the space S. We then see (by Lemma 3.11, Theorem 3.8) that in the context of our optimization problem, rather than studying the class of all possible stopping times, we can equivalently focus on *randomized* stopping times on the canonical space. These can be characterized in various equivalent terms (cf. Theorem 3.8); e.g. viewing them as measures on $C_0(\mathbb{R}_+) \times \mathbb{R}_+$ is useful to establish compactness results while the representation through 'increasing' functions on S is necessary for the manipulations of stopping times which we need to consider in the proof of the monotonicity principle, Theorem 1.3, in Section 5. Finally, we shall consider the set of 'joinings' which can be interpreted as a type of coupling between a randomized stopping time and an abstract probability measure. This is an important ingredient in the proofs of Theorem 1.2 and Theorem 1.3.

3.1. **Spaces and Filtrations.** We will primarily consider the space $C_0(\mathbb{R}_+)$ of continuous functions on $\mathbb{R}_+$ starting at the value 0, with the topology of uniform convergence on compact sets. The elements of $C_0(\mathbb{R}_+)$ will be denoted by ω. We denote the canonical process on $C_0(\mathbb{R}_+)$ by $(B_t)_{t\geq 0}$, i.e. $B_t(\omega) = \omega_t$. We denote the Wiener measure by $\mathbb{W}$. As explained above we consider the set S of all continuous functions defined on some initial segment $[0, s]$ of $\mathbb{R}_+$ and starting with value 0; we will denote the elements of S by (f, s) and (g, t). The set S admits a natural partial ordering; we say that (g, t) extends (f, s) if $t \geq s$ and the restriction $g_{\restriction[0,s]}$ of g to the interval $[0, s]$ equals f. We consider S with the topology induced by the metric

$$d_S((f, s), (g, t)) := \max\left(t - s, \sup_{0\leq u\leq s} |f(u) - g(u)|, \sup_{s\leq u\leq t} |g(u) - f(s)|\right), \tag{3.1}$$

for $(f, s), (g, t) \in S$, $s \leq t$. Equipped with this topology, S is a Polish space.

For our arguments it will be important to be precise about the relationship between the sets $C_0(\mathbb{R}_+) \times \mathbb{R}_+$ and S. We therefore discuss the underlying filtrations in some detail.

We consider two different filtrations on the Wiener space $C_0(\mathbb{R}_+)$, the canonical or natural filtration $\mathcal{F}^0 = (\mathcal{F}^0_t)_{t\in\mathbb{R}_+}$ as well as its usual augmentation $\mathcal{F}^a = (\mathcal{F}^a_t)_{t\in\mathbb{R}_+}$. As Brownian

motion is a continuous Feller process, all right-continuous $\mathcal{F}^a$-martingales are continuous ([47, Theorem VI. 15.4]) and hence all $\mathcal{F}^a$-stopping times are predictable and the $\mathcal{F}^a$-optional and $\mathcal{F}^a$-predictable σ-algebras coincide ([46, Corollary IV 5.7]). By [16, Theorem IV. 97, Rem. IV. 98] we also have that the $\mathcal{F}^0$-predictable, $\mathcal{F}^0$-optional and $\mathcal{F}^0$-progressive σ-algebras coincide because $C_0(\mathbb{R}_+)$ is the set of *continuous* paths. Moreover, we will use the following result.

Theorem 3.1. *Let $(\Omega, \mathcal{G}, (\mathcal{G}_t)_{t\in\mathbb{R}_+}, \mathbb{P})$ be a filtered probability space and let $\mathcal{G}^a$ be the usual augmentation of the filtration $\mathcal{G}$.*

(1) If τ is a predictable time wrt $\mathcal{G}^a$, then there exists a predictable time τ' wrt $\mathcal{G}$ such that $\tau = \tau'$ a.s. For every $\mathcal{G}^a$-predictable process $(X_t)_{t\in\mathbb{R}_+}$ there is a $\mathcal{G}$-predictable process $(X'_t)_{t\in\mathbb{R}_+}$ which is indistinguishable from $(X_t)_{t\in\mathbb{R}_+}$.

(2) If $(A_t)_{t\in\mathbb{R}_+}$ is an increasing right-continuous $\mathcal{G}^a$-predictable process there is an increasing right-continuous $\mathcal{G}$-predictable process $(A'_t)_{t\in\mathbb{R}_+}$ (possibly assuming the value $+\infty$) which is indistinguishable from $(A_t)_{t\in\mathbb{R}_+}$.

Proof. For Statement (1) we refer to [16, Theorem IV. 78] and the comments directly afterwards. To prove statement (2), let $(A_t)_{t\in\mathbb{R}_+}$ be an increasing right-continuous $\mathcal{G}^a$-predictable process. Arguing on $(\frac{2}{\pi}\arctan(A_t - A_0))_{t\in\mathbb{R}_+}$, we may assume that A takes values in $[0, 1]$.

We use an extension of the filtered probability space denoted $(\bar{\Omega}, \bar{\mathcal{G}}, (\bar{\mathcal{G}}_t)_{t\geq 0}, \bar{\mathbb{P}})$, where we take $\bar{\Omega} = \Omega \times [0, 1]$, $\bar{\mathcal{G}} = \mathcal{G} \otimes \mathcal{B}([0, 1]), \bar{\mathbb{P}}(D_1 \times D_2) = \mathbb{P}(D_1)\mathcal{L}(D_2)$, and set $\bar{\mathcal{G}}_t = \mathcal{G}_t \otimes \mathcal{B}([0, 1])$ and let $\bar{\mathcal{G}}^a$ be its usual augmentation. Here, $\mathcal{L}$ denotes Lebesgue measure. Abusing notation we also write A for the mapping $(\omega, x, t) \mapsto A_t(\omega)$ on $\bar{\Omega} \times \mathbb{R}_+$.

Set $Y(\omega, x) := x$. Then $A - Y$ is $\bar{\mathcal{G}}^a$-predictable and right-continuous, hence

$$\rho(\omega, x) := \inf\{t \geq 0 : A_t(\omega) \geq x\} = \inf\{t \geq 0 : A_t(\omega) - Y(\omega, x) \geq 0\}$$

is a $\bar{\mathcal{G}}^a$-predictable stopping time by the (predictable) Debut theorem. Moreover

$$A_t(\omega) = \inf\{x \geq 0 : \rho(\omega, x) > t\} = 1 - \mathcal{L}\{x : \rho(\omega, x) > t\}.$$

Pick a $\bar{\mathcal{G}}$-predictable stopping time ρ' such that $\rho' = \rho$, $\bar{\mathbb{P}}$-a.s. and set

$$A'_t(\omega) := 1 - \mathcal{L}\{x : \rho'(\omega, x) > t\}.$$

Then $A'(\omega)$ is increasing and right-continuous for each ω. For each t

$$\mathcal{L}\{x : \rho'(\omega, x) > t\} = \mathcal{L}\{x : \rho(\omega, x) > t\}$$

for $\mathbb{P}$-a.a. ω, hence A' is a version of A. By right-continuity, A and A' are indistinguishable. Predictability of ρ' asserts that (using obvious abbreviations)

$$\{(\omega, x, t) : \rho'(\omega, x) > t\} \in \mathsf{pred}_{\bar{\mathcal{G}}} = \mathsf{pred}_{\mathcal{G}} \otimes \mathcal{B}_{[0,1]}.$$

Hence $(\omega, t) \mapsto A'_t(\omega)$ is $\mathsf{pred}_{\mathcal{G}}$-measurable. □

The message of Theorem 3.2 below is that a process $(X_t)_{t\in\mathbb{R}_+}$ is $\mathcal{F}^0$-optional (and hence also $\mathcal{F}^0$-predictable in our setup) iff $X_t(\omega)$ can be calculated from the restriction $\omega_{\restriction[0,t]}$. We introduce the mapping

$$r : C_0(\mathbb{R}_+) \times \mathbb{R}_+ \to S, \quad r(\omega, t) = (\omega_{\restriction[0,t]}, t). \tag{3.2}$$

We note that the topology on S introduced in (3.1) coincides with the final topology induced by the mapping r; moreover r is a continuous open mapping.

The following result is a particular case of [16, Theorem IV. 97] (in somewhat different notation).

Theorem 3.2. *$\mathcal{F}^0$-optional sets and functions on $C_0(\mathbb{R}_+) \times \mathbb{R}_+$ correspond to Borel measurable sets and functions on S. More precisely we have:*

(1) A set $D \subseteq C_0(\mathbb{R}_+) \times \mathbb{R}_+$ is $\mathcal{F}^0$-optional iff $D = r^{-1}(A)$ for some Borel set $A \subseteq S$.

(2) A process $X = (X_t)_{t\in\mathbb{R}_+}$ is $\mathcal{F}^0$-optional iff $X = H \circ r$ for some Borel measurable $H : S \to \mathbb{R}$.

The mapping r is not a closed mapping: it is easy to see that there exist closed sets in $C_0(\mathbb{R}_+) \times \mathbb{R}_+$ with a non-closed image under r. However this does not happen for closed optional sets: it is straightforward that an $\mathcal{F}^0$-optional set $A \subseteq C_0(\mathbb{R}_+) \times \mathbb{R}_+$ is closed iff the corresponding set $r(A)$ is closed in S.

Definition 3.3. *If X is an $\mathcal{F}^0$-optional process we write X^S for the unique function $S \to \mathbb{R}$ satisfying $X = X^S \circ r$. We say that an optional process X is S-continuous (resp. S-lsc) if the corresponding function $X^S : S \to \mathbb{R}$ is continuous (resp. lsc).*

It is trivially true that an S-continuous process is continuous in the usual pathwise sense. The converse is not generally true — consider the case where $X_t(\omega) = \operatorname{sign}(\omega(1))(t-2)_+$. This is a continuous, optional process, however the corresponding function X^S is not a continuous mapping from S to $\mathbb{R}$. Other examples arise from functions connected to the local time of Brownian motion, cf. Section 6.2.

Definition 3.4. *For a measurable $X : C_0(\mathbb{R}_+) \to \mathbb{R}$ which is bounded or positive we set*

$$\mathbb{E}[X|\mathcal{F}_t^0](\omega) := X_t^M(\omega) := \int X((\omega_{\restriction[0,t]}) \oplus \omega')\, d\mathbb{W}(\omega'). \tag{3.3}$$

Clearly, (3.3) defines an $\mathcal{F}_t^0$-measurable function which is a version of the classical conditional expectation; subsequently, it will be useful to have this function defined for *all* ω. In accordance with Definition 3.3 we write $X^{M,S}$ for the function satisfying $X^M = X^{M,S} \circ r$.

Proposition 3.5. *Let $X \in C_b(C_0(\mathbb{R}_+))$. Then X_t^M is an S-continuous martingale, $X_\infty^M = \lim_{t\to\infty} X_t^M$ exists and equals X.*

Proof. Note that $X^{M,S}(f,s) = \int X^{(f,s)\oplus}(\omega)\, \mathbb{W}(d\omega)$ for $(f,s) \in S$. Also, $(f_n, s_n) \to (f,s)$ implies $f_n \oplus \omega \to f \oplus \omega$ for $\omega \in C_0(\mathbb{R}_+)$ and, by continuity of X, $X^{(f_n,s)\oplus}(\omega) \to X^{(f,s)\oplus}(\omega)$. Since X is bounded, dominated convergence implies $X^{M,S}(f_n, s_n) \to X^{M,S}(f,s)$. □

For $X \in C_b(C_0(\mathbb{R}_+))$, X^M is a martingale with continuous paths and hence satisfies the optional stopping theorem. Using the functional monotone class theorem, we see that the optional stopping theorem holds for X^M for all bounded measurable $X : C_0(\mathbb{R}_+) \to \mathbb{R}$. Also one can prove that X^M has almost surely continuous paths, even if X itself was not continuous, but we will not use this fact.

3.2. **Randomized stopping times.** Working on the probability space $(C_0(\mathbb{R}_+), \mathbb{W})$, a stopping time τ is a mapping which assigns to each path ω the time $\tau(\omega)$ at which the path is stopped. If the stopping time depends on external randomization, then we may consider a path ω which is not stopped at a single point $\tau(\omega)$, but rather that there is a sub-probability measure τ_ω on $\mathbb{R}$ which represents the probability that the path ω is stopped at a given time, conditional on observing the path ω. The aim of this section is to make this idea precise, and to establish connections with related properties in the literature. Specifically, the notion of a *randomized stopping time* has previously appeared in e.g. [5, 40, 49].

Subsequently we will identify randomized stopping times as a subset of the well studied **P**-measures: A finite measure ξ on $C_0(\mathbb{R}_+) \times \mathbb{R}_+$ is a **P**-*measure* (wrt $\mathbb{W}$) if it does not charge any $\mathbb{W}$-evanescent set. A basic result of Doléans [18] is the following

Theorem 3.6 (cf. [17, Theorem VI 65])**.** *A finite measure ξ on $C_0(\mathbb{R}_+) \times \mathbb{R}_+$ is a* **P**-*measure iff there exists a right-continuous increasing process A, $\mathbb{E}[A_\infty] < \infty$ such that for all bounded and measurable processes X*

$$\xi(X) = \mathbb{E}[\int X_s\, dA_s].$$

Here the process A is unique up to evanescence.

We will be particularly interested in the following subset of **P**-measures:

$$\begin{aligned}\mathsf{M} &:= \{\xi \in \mathcal{P}^{\leq 1}(C_0(\mathbb{R}_+) \times \mathbb{R}_+) : \xi(d\omega, dt) = \xi_\omega(dt)\mathbb{W}(d\omega), \xi_\omega \in \mathcal{P}^{\leq 1}(\mathbb{R}_+) \text{ for } \mathbb{W}\text{-a.e. } \omega\} \\ &= \{\xi \in \mathcal{P}^{\leq 1}(C_0(\mathbb{R}_+) \times \mathbb{R}_+) : \operatorname{proj}_{C_0(\mathbb{R}_+)}(\xi) \leq \mathbb{W}\},\end{aligned}$$

where $(\xi_\omega)_{\omega\in C_0(\mathbb{R}_+)}$ is a disintegration of ξ in the first coordinate $\omega \in C_0(\mathbb{R}_+)$. We equip M with the weak topology induced by the continuous bounded functions on $C_0(\mathbb{R}_+) \times \mathbb{R}_+$. Clearly any $\xi \in \mathsf{M}$ is a $\mathbf{P}$-measure with corresponding increasing process $A^\xi_\omega(t) = \xi_\omega([0,t])$ being the cumulative distribution function of ξ_ω.

Definition 3.7 (Randomized stopping times)**.** *A measure $\xi \in \mathsf{M}$ is called a randomized stopping time, written $\xi \in \mathsf{RST}$, iff the associated increasing process A is optional.*

Below, it will sometimes be convenient to represent randomized stopping times on an extension of the space $(C_0(\mathbb{R}_+), \mathcal{F}^0, (\mathcal{F}^0_t)_{t\geq 0}, \mathbb{W})$: we will consider $(\overline{C}_0(\mathbb{R}_+), \bar{\mathcal{F}}, (\bar{\mathcal{F}}_t)_{t\geq 0}, \overline{\mathbb{W}})$, where $\overline{C}_0(\mathbb{R}_+) = C_0(\mathbb{R}_+) \times [0,1]$, $\overline{\mathbb{W}}(A_1 \times A_2) = \mathbb{W}(A_1)\mathcal{L}(A_2)$ (where $\mathcal{L}$ denotes Lebesgue measure), $\bar{\mathcal{F}}$ is the completion of $\mathcal{F}^0 \otimes \mathcal{B}([0,1])$, and $\bar{\mathcal{F}}_t$ the usual augmentation of $(\mathcal{F}^0_t \otimes \mathcal{B}([0,1]))_{t\geq 0}$. We will write $\bar{B} = (\bar{B}_t)_{t\geq 0}$ for the process given by $\bar{B}_t(\omega, u) = \omega_t$. Observe that if $Y_t(\omega, u) = u$, then $(\bar{B}_t, Y_t)$ is (trivially) a continuous Feller process, and hence by the same arguments as above, the $\bar{\mathcal{F}}$-predictable and $\bar{\mathcal{F}}$-optional σ-algebras coincide.

Randomized stopping times play a key role in this paper; depending on the respective context, the following different characterizations will be useful:

Theorem 3.8. *Let $\xi \in \mathsf{M}$. Then the following are equivalent:*

(1) There is a Borel function $A : S \to [0,1]$ such that the process $A \circ r$ is right-continuous increasing and

$$\xi_\omega([0,s]) := A \circ r(\omega, s) \tag{3.4}$$

defines a disintegration of ξ wrt to $\mathbb{W}$.

(2) We have $\xi \in \mathsf{RST}$, i.e. given a disintegration $(\xi_\omega)_{\omega\in C_0(\mathbb{R}_+)}$ of ξ, the random variable $\tilde{A}_t(\omega) = \xi_\omega([0,t])$ is $\mathcal{F}^a_t$-measurable for all $t \in \mathbb{R}_+$.

(3) For all $f \in C_b(\mathbb{R}_+)$ supported on some $[0,t]$, $t \geq 0$ and all $g \in C_b(C_0(\mathbb{R}_+))$

$$\int f(s)(g - \mathbb{E}[g|\mathcal{F}^0_t])(\omega)\, \xi(d\omega, ds) = 0 \tag{3.5}$$

(4) On the probability space $(\overline{C}_0(\mathbb{R}_+), \bar{\mathcal{F}}, (\bar{\mathcal{F}}_t)_{t\geq 0}, \overline{\mathbb{W}})$, the random time

$$\rho(\omega, u) := \inf\{t \geq 0 : \xi_\omega([0,t]) \geq u\} \tag{3.6}$$

defines an $\bar{\mathcal{F}}$-stopping time.

Proof. The equivalence of (1) and (2) follows directly from Theorems 3.1, 3.2 and 3.6.

It is straightforward to deduce (4) from (1). To see that (4) implies (2), consider for $t \geq 0, \omega \in C_0(\mathbb{R}_+)$

$$\tilde{A}(\omega, t) := \int_0^1 \mathbb{1}_{[0,t]}(\rho(\omega, u))\, du.$$

To show that (2) and (3) are equivalent, we first note that (2) is equivalent to requiring that $X_t(\omega) := \xi_\omega(f)$ is $\mathcal{F}^a_t$ measurable whenever $f \in C_b(\mathbb{R}_+)$ is supported on $[0,t]$. However we can express this measurability in a different fashion. Note that a bounded Borel function h is $\mathcal{F}^a_t$-measurable iff for all bounded Borel functions g

$$\mathbb{E}[h(\mathbb{E}[g|\mathcal{F}^a_t] - g)] = \mathbb{E}[h(\mathbb{E}[g|\mathcal{F}^0_t] - g)]$$

vanishes; of course this does not rely on our particular setup. By a functional monotone class argument, for $\mathcal{F}^a_t$-measurability of X_t it is sufficient to check that

$$\mathbb{E}[X_t(g - \mathbb{E}[g|\mathcal{F}^0_t])] = 0 \tag{3.7}$$

for all $g \in C_b(C_0(\mathbb{R}_+))$. In terms of ξ, (3.7) amounts to

$$\begin{aligned} 0 = \mathbb{E}[X_t(g - \mathbb{E}[g|\mathcal{F}^0_t])] &= \int \mathbb{W}(d\omega) \int \xi_\omega(ds) f(s)(g - \mathbb{E}[g|\mathcal{F}^0_t])(\omega) \\ &= \int f(s)(g - \mathbb{E}[g|\mathcal{F}^0_t])(\omega)\, \xi(d\omega, ds). \end{aligned}$$

□

Remark 3.9. (1) The function A in (3.4) is unique up to indistinguishability (cf. Theorem 3.6). We will denote this function by A^ξ.

(2) We will say $\xi \in \mathsf{RST}$ is a *non-randomized stopping time* iff there is a disintegration $(\xi_\omega)_{\omega \in C_0(\mathbb{R}_+)}$ of ξ such that ξ_ω is either null (corresponding to a path which is not stopped) or a Dirac-measure (of mass 1) for every ω. Clearly this means that $\xi_\omega = \delta_{\tau(\omega)}$ a.s. for some (non-randomized) stopping time τ. ξ is a non-randomized stopping time iff there is a version of A^ξ which only attains the values 0 and 1.

(3) We will say $\xi \in \mathsf{RST}$ is a *finite randomized stopping time* iff $\xi(C_0(\mathbb{R}_+) \times \mathbb{R}_+) = 1$.

An immediate consequence of Theorem 3.8 (3) is the following

Corollary 3.10. *The set* RST *is closed wrt the weak topology induced by the continuous bounded functions on* $C_0(\mathbb{R}_+) \times \mathbb{R}_+$.

The next lemma implies that optimizing over usual stopping times on a rich enough probability space in (OptSEP) is equivalent to optimizing over randomized stopping times on Wiener space.

Lemma 3.11. *Let B be a Brownian motion on some stochastic basis $(\Omega, \mathcal{G}, (\mathcal{G}_t)_{t\geq 0}, \mathbb{P})$, let τ be a $\mathcal{G}$-stopping time and consider*

$$\Phi : \Omega \to C_0(\mathbb{R}_+) \times \mathbb{R}_+, \bar{\omega} \mapsto ((B_t(\bar{\omega}))_{t\geq 0}, \tau(\bar{\omega})).$$

Then $\xi := \Phi(\mathbb{P})$ is a randomized stopping time and for any measurable $\gamma : S \to \mathbb{R}$ we have

$$\int \gamma((f,s))\, r(\xi)(d(f,s)) = \mathbb{E}_{\mathbb{P}}[\gamma((B_t)_{t\leq\tau}, \tau)]. \tag{3.8}$$

If Ω is sufficiently rich that it supports a uniformly distributed random variable which is $\mathcal{G}_0$-measurable then for any $\xi \in \mathsf{RST}$, we can find a $\mathcal{G}$-stopping time τ such that $\xi = \Phi(\mathbb{P})$ and (3.8) *holds.*

Proof. Clearly $\xi := \Phi(\mathbb{P}) \in \mathsf{M}$. Write $(\xi_\omega)_{\omega \in C_0(\mathbb{R}_+)}$ for a disintegration wrt Wiener measure. We need to show that $\xi_\omega([0,t])$ is $\mathcal{F}_t^a$-measurable. Let $g : C_0(\mathbb{R}_+) \to \mathbb{R}$ be a measurable function. If $h = \mathbb{E}_{\mathbb{W}}[g|\mathcal{F}_t^a]$, writing $\mathcal{G}_t^a$ for the usual augmentation of $\mathcal{G}$, and noting that $(B_t)_{t\geq 0}$ is also a $(\mathcal{G}_t^a)_{t\geq 0}$-Brownian motion, we have

$$\mathbb{E}_{\mathbb{P}}[g((B_r)_{r\geq 0})|\mathcal{G}_t^a] = h((B_r)_{r\geq 0}), \quad \mathbb{P}\text{ - a.s.}$$

It then follows that

$$\begin{aligned}\int g(\omega)\xi_\omega([0,t])\, \mathbb{W}(d\omega) &= \mathbb{E}_{\mathbb{P}}[g((B_r)_{r\geq 0})\mathbb{1}_{\{\tau\leq t\}}] \\ &= \mathbb{E}_{\mathbb{P}}[\mathbb{E}_{\mathbb{P}}[g((B_r)_{r\geq 0})|\mathcal{G}_t^a]\mathbb{1}_{\{\tau\leq t\}}] \\ &= \mathbb{E}_{\mathbb{P}}[h((B_r)_{r\geq 0})\mathbb{1}_{\{\tau\leq t\}}] = \int h(\omega)\xi_\omega([0,t])\, \mathbb{W}(d\omega).\end{aligned}$$

Hence $\xi_\omega([0,t])$ is $\mathcal{F}_t^a$-measurable as required.

To prove the second part, we observe that by Theorem 3.8 (4), there exists an $\bar{\mathcal{F}}$-stopping time ρ' representing ξ. Since ρ' is $\bar{\mathcal{F}}$-predictable, it follows from Theorem 3.1 that there exists an almost surely equal $(\mathcal{F}_t^0 \times \mathcal{B}([0,1]))_{t\geq 0}$-stopping time ρ. Then we can define a random time on Ω by $\rho((B_s)_{s\geq 0}, Y)$, where B is the Brownian motion, and Y the independent $\mathcal{G}_0$-measurable, uniform random variable. Consider the map $\bar{\Phi} : \Omega \to \overline{C}_0(\mathbb{R}_+), \bar{\omega} \mapsto ((B_t(\bar{\omega}))_{t\geq 0}, Y(\bar{\omega}))$. Since ρ is a $(\mathcal{F}_t^0 \times \mathcal{B}([0,1]))_{t\geq 0}$-stopping time and $\bar{\Phi}$ is measurable from $(\Omega, \mathcal{G}_t)$ to $(\overline{C}_0(\mathbb{R}_+), \mathcal{F}_t^0 \times \mathcal{B}([0,1]))$, $\rho \circ (B, Y)$ is a $\mathcal{G}$-stopping time. □

3.3. **Randomized stopping times solving the Skorokhod problem and compactness.** For a finite randomized stopping time ξ and optional $Y : C_0(\mathbb{R}_+) \times \mathbb{R}_+ \to \mathbb{R}$ which is bounded or positive, define Y_ξ as the push-forward of ξ under the mapping $(t, \omega) \mapsto Y_t(\omega)$ and denote $Y_t^\xi := Y_{\xi\wedge t}$ for $t \in \mathbb{R}_+$. Considering the representation ρ of ξ on the extended space $\overline{C}_0(\mathbb{R}_+)$ as in (3.6) and writing $\bar{Y}_t(\omega, u) = Y_t(\omega)$, we then have

$$\bar{Y}_\rho \sim Y_\xi \text{ and } \bar{Y}_t^\rho \sim Y_t^\xi. \tag{3.9}$$

Taking $Y_t = t$ we obtain $\xi(T) = \bar{\mathbb{E}}[\rho]$, where T denotes the projection

$$T : C_0(\mathbb{R}_+) \times \mathbb{R}_+ \to \mathbb{R}_+. \tag{3.10}$$

Recall that μ has mean 0 and finite second moment $\int x^2\,\mu(dx) =: V$. Then the following result follows directly from classical properties of stopping times (e.g. [31, Corollary 3.3]).

Lemma 3.12. *Let $\xi \in \mathsf{RST}^1$ with representation ρ on $\overline{C}_0(\mathbb{R}_+)$ as in* (3.6)*. Assume that $B_\xi = \mu$, i.e. $\bar{B}_\rho \sim \mu$. Then the following are equivalent:*

(1) $\xi(T) = \bar{\mathbb{E}}[\rho] < \infty$,
(2) $\xi(T) = \bar{\mathbb{E}}[\rho] = V$,
(3) $(\bar{B}_{\rho\wedge t})$ is uniformly integrable.

Definition 3.13. *We denote by $\mathsf{RST}(\mu)$ the set of all finite randomized stopping times satisfying the conditions in Lemma 3.12.*

For us it is crucial that *randomized* stopping times have the following property:

Theorem 3.14. *The set $\mathsf{RST}(\mu)$ is non-empty and compact wrt the weak topology induced by the continuous and bounded functions on $C_0(\mathbb{R}_+) \times \mathbb{R}_+$.*

Proof. If μ is a centered probability then it is not hard to establish that the Skorokhod embedding problem has a solution, e.g. one can use the external randomization $u \in [0, 1]$ to stop $(\bar{B}_t(\omega, u))_{t\geq 0}$ once it leaves $(a(u), b(u))$. Choosing a, b carefully we obtain a solution of (SEP), see e.g. [43, p332] for a detailed account.

By Prokhorov's theorem we have to show that $\mathsf{RST}(\mu)$ is tight and closed.

Tightness. Fix $\varepsilon > 0$ and take $R = 2V/\varepsilon$. Then, for any $\xi \in \mathsf{RST}(\mu)$ we have $\xi(T > R) \leq \varepsilon/2$. As $C_0(\mathbb{R}_+)$ is Polish there is a compact set $\tilde{K} \subseteq C_0(\mathbb{R}_+)$ such that $\mathbb{W}(\tilde{K}^c) \leq \varepsilon/2$. Set $K := \tilde{K} \times [0, R]$. Then K is compact and we have for any $\xi \in \mathsf{RST}(\mu)$

$$\xi(K^c) \leq \mathbb{W}(\tilde{K}^c) + \xi(T > R) \leq \varepsilon.$$

Closedness. Take a sequence $(\xi_n)_{n\in\mathbb{N}}$ in $\mathsf{RST}(\mu)$ converging to some ξ. Putting $h : C_0(\mathbb{R}_+) \times \mathbb{R}_+ \to \mathbb{R}, (\omega, t) \mapsto \omega(t)$ we have to show that $h(\xi) = \mu$ and that $\xi(T) < \infty$. Note that h is a continuous map. Take any $g \in C_b(\mathbb{R})$. Then $g \circ h \in C_b(C_0(\mathbb{R}_+) \times \mathbb{R}_+)$ and hence

$$\int g\,d\mu = \lim_n \int_{C_0(\mathbb{R}_+)\times\mathbb{R}_+} g \circ h\,d\xi_n = \int_{C_0(\mathbb{R}_+)\times\mathbb{R}_+} g \circ h\,d\xi = \int g\,dh(\xi),$$

thus $h(\xi) = \mu$. Moreover, $T \wedge N$ is continuous and bounded for each $N \in \mathbb{N}$, hence $\xi(T \wedge N) = \lim_n \xi_n(T \wedge N) \leq V$. As N was arbitrary, it follows that also $\xi(T) \leq V < \infty$. □

Our use of randomization to achieve compactness of a set of stopping times has similarities to the work of Baxter and Chacon [5]. However their setup is different, and their intended applications are not connected to Skorokhod embedding.

3.4. **Joinings.** We now add another dimension: we assume that (Y, ν) is some Polish probability space and consider randomized stopping times where each death of a particle is tagged by an element of Y. More precisely, the set of *joinings* $\mathsf{JOIN}(\nu)$ is given by

$$\Big\{\pi \in \mathcal{P}^{\leq 1}(C_0(\mathbb{R}_+) \times \mathbb{R}_+ \times \mathsf{Y}) : \mathsf{proj}_{C_0(\mathbb{R}_+)\times\mathbb{R}_+}(\pi_{\restriction C_0(\mathbb{R}_+)\times\mathbb{R}_+\times D}) \in \mathsf{RST}, D \in \mathcal{B}(\mathsf{Y}), \mathsf{proj}_{\mathsf{Y}}(\pi) \leq \nu\Big\}.$$

We shall also write $\mathsf{JOIN}^1(\nu)$ for the subset of $\pi \in \mathsf{JOIN}(\nu)$ having mass 1.

Remark 3.15. Write pred for the σ-algebra of $\mathcal{F}^0$-predictable sets in $C_0(\mathbb{R}_+) \times \mathbb{R}_+$. We call a set $A \subseteq C_0(\mathbb{R}_+) \times \mathbb{R}_+ \times \mathsf{Y}$ predictable if it is an element of $\mathsf{pred} \otimes \mathcal{B}(\mathsf{Y})$. We will say that a function defined on $C_0(\mathbb{R}_+) \times \mathbb{R}_+ \times \mathsf{Y}$ is predictable if it is measurable wrt $\mathsf{pred} \otimes \mathcal{B}(\mathsf{Y})$. As before, predictable subsets of $C_0(\mathbb{R}_+) \times \mathbb{R}_+ \times \mathsf{Y}$ correspond to measurable subsets of $S \times \mathsf{Y}$, and similarly for functions.

4. The Optimization Problem and Duality

4.1. **The Primal Problem.** As defined in (OptSEP) in the introduction, our primal problem is to minimize the value corresponding to a function $\gamma : S \to \mathbb{R}$, where the minimization is taken over stopping times of Brownian motion defined on a sufficiently rich probability space. By Lemma 3.11, we obtain an equivalent problem if we take B to be

the canonical process on Wiener space $C_0(\mathbb{R}_+)$ and minimize over all randomized stopping times, i.e. we have

$$P_\gamma = \inf\left\{\int \gamma \circ r(\omega,t)\,\xi(d\omega,dt) : \xi \in \mathsf{RST}(\mu)\right\}. \tag{4.1}$$

In the following we will mainly work with the technically convenient formulation given in (4.1). It immediately allows us to establish the existence of optimal stopping times:

Theorem 4.1. *Assume that $\gamma : S \to \mathbb{R}$ is lsc and bounded from below in the sense that for some constants $a, b, c \in \mathbb{R}_+$*

$$-\big(a + bs + c\max_{r\le s} B_r^2\big) \le \gamma((B_r)_{r\le s}, s) \tag{4.2}$$

holds on $C_0(\mathbb{R}_+)\times\mathbb{R}_+$. Then the functional

$$\xi \mapsto \int_{C_0(\mathbb{R}_+)\times\mathbb{R}_+} \gamma\circ r(\omega,t)\,\xi(d\omega,dt) \tag{4.3}$$

is lsc and (4.1) *admits a minimizer.*

By Lemma 3.11, Theorem 1.1 is a consequence of this result.

Proof of Theorem 4.1 / Theorem 1.1. By the Portmanteau theorem, the functional (4.3) is lsc if $\gamma : S \to \mathbb{R}$ is lsc and bounded from below by a constant.

For the general case we recall the pathwise version of Doob's inequality (see [1])

$$\max_{r\le s} B_r^2 \le \underbrace{\int_0^s 4\max_{t\le r}|B_t|\,dB_r}_{=:M_s} + 4B_s^2. \tag{4.4}$$

We emphasize that we can understand the integral defining M in a *pathwise* fashion. This is possible since $r \mapsto \max_{t\le r}|B_t|$ is increasing; we refer to [1] for details. In fact it is straightforward to show that M is an S-continuous martingale satisfying $|M_t| < 2\max_{r\le t} B_r^2$. It follows that $\tilde\gamma(f,s) := \gamma(f,s) + bs + c(M^S(f,s) + 4f(s)^2)$ is bounded from below and hence $\xi \mapsto \int \tilde\gamma\, d\xi$ is lsc. As the value of $\int bs + c(M_s(\omega) + 4B_s^2(\omega))\,d\xi(\omega,s)$ is the same for any $\xi \in \mathsf{RST}(\mu)$ the functional (4.3) is lsc as well. □

In Section 7 below we establish existence of a minimizing stopping time in the case where the measure μ does not necessarily admit a finite second moment. However we will then replace Assumption (4.2) by the requirement that γ is bounded from below.

4.2. **The dual problem.** The following result implies Theorem 1.2.

Theorem 4.2. *Let $\gamma : S \to \mathbb{R}$ be lsc and bounded from below in the sense of* (4.2)*. Set*

$$D_\gamma = \sup\left\{\int \psi(y)\,d\mu(y) : \psi \in C(\mathbb{R}),\ \begin{array}{l}\exists\varphi, \varphi \text{ is an } S\text{-continuous martingale}, \varphi_0 = 0\\ \varphi_t(\omega) + \psi(\omega(t)) \le \gamma\circ r(\omega,t),\ (\omega,t)\in C_0(\mathbb{R}_+)\times\mathbb{R}_+\end{array}\right\}$$

where φ, ψ satisfy $|\varphi_t| \le a + bt + cB_t^2$, $|\psi(y)| \le a + by^2$ for some $a, b, c > 0$. Then we have

$$P_\gamma = D_\gamma. \tag{4.5}$$

Using the same argument as in the proof of Theorem 4.1, we see that it suffices to establish Theorem 4.2 in the case where γ is bounded from below. As usual, one part of the duality relation is straightforward to verify: for (φ,ψ) satisfying the dual constraint and $\xi \in \mathsf{RST}(\mu)$ we have

$$\int \psi(y)\,\mu(dy) = \int \psi(\omega(t))\,\xi(d\omega,dt) + \int \varphi_t(\omega)\,\xi(d\omega,dt) \le \int \gamma\circ r(\omega,t)\,\xi(d\omega,dt),$$

hence $D_\gamma \le P_\gamma$.

We will establish Theorem 4.2 as a consequence of the following auxiliary duality result, where we write T for the projection map $C_0(\mathbb{R}_+)\times\mathbb{R}_+\times\mathbb{R} \to \mathbb{R}_+$, $T(\omega,t,y) = t$.

Proposition 4.3. *Let $c : C_0(\mathbb{R}_+) \times \mathbb{R}_+ \times \mathbb{R} \to \mathbb{R} \cup \{\infty\}$ be lsc, predictable (cf. Remark 3.15) and bounded from below. Write $V = \int x^2 \, \mu(dx)$. Then*

$$\inf_{\pi} \int c(\omega, t, y) \, d\pi(\omega, t, y) = \sup_{(\varphi,\psi)} \int \varphi \, d\mathbb{W} + \int \psi \, d\mu. \tag{$\star$}$$

where the infimum is taken over the set

$$\mathsf{JOIN}^{1,V}(\mu) := \{\pi \in \mathsf{JOIN}^1(\mu) : \pi(T) \leq V\}$$

and the supremum is taken over $\varphi \in C_b(C_0(\mathbb{R}_+))$, $\psi \in C_b(\mathbb{R})$ such that

$$\exists \alpha \geq 0 \text{ s.t. } \varphi_t^M(\omega) + \psi(y) - \alpha(t - V) \leq c(\omega, t, y) \text{ for } \omega \in C_0(\mathbb{R}_+), t \in \mathbb{R}_+, y \in \mathbb{R}. \quad (d^M[c, V])$$

Proposition 4.3 should be compared to the (formally) very similar classical duality theorem of optimal transport, see e.g. [58, Section 5] for a proof as well as for a discussion of its origin and related literature.

Theorem 4.4 (Monge-Kantorovich Duality)**.** *Let (X_i, μ_i), $i = 1, 2$ be Polish probability spaces and $c : \mathsf{X}_1 \times \mathsf{X}_2 \to \mathbb{R} \cup \{\infty\}$ a lsc and bounded from below* cost function*. Then*

$$\inf_{\pi} \int c(x_1, x_2) \, d\pi(x_1, x_2) = \sup_{(\varphi,\psi)} \Big(\int \varphi \, d\mu_1 + \int \psi \, d\mu_2 \Big), \tag{4.6}$$

where the inf *is taken over probabilities π on $\mathsf{X}_1 \times \mathsf{X}_2$ satisfying* $\mathrm{proj}_{\mathsf{X}_1}(\pi) = \mu_1, \mathrm{proj}_{\mathsf{X}_2}(\pi) = \mu_2$ *and the* sup *is taken over $\varphi \in C_b(\mathsf{X}_1)$, $\psi \in C_b(\mathsf{X}_2)$ satisfying for $x_1 \in \mathsf{X}_1, x_2 \in \mathsf{X}_2$*

$$\varphi(x_1) + \psi(x_2) \leq c(x_1, x_2).$$

The strategy of the proof of Proposition 4.3 is to establish the duality relation ($\star$) for π, resp. (φ, ψ) taken from certain larger candidate sets, in which case the duality relation follows from Theorem 4.4. Then we introduce additional constraints via a variational approach to obtain an improved duality through the following min-max theorem.

Theorem 4.5 (see e.g. [55, Thm. 45.8] or [2, Thm. 2.4.1])**.** *Let K, L be convex subsets of vector spaces H_1 resp. H_2, where H_1 is locally convex and let $F : K \times L \to \mathbb{R}$ be given. If*

(1) K is compact,
(2) $F(\cdot, y)$ is continuous and convex on K for every $y \in L$,
(3) $F(x, \cdot)$ is concave on L for every $x \in K$

then

$$\sup_{y \in L} \inf_{x \in K} F(x, y) = \inf_{x \in K} \sup_{y \in L} F(x, y).$$

Proof of Proposition 4.3. Fix $t_0 > 0$ and consider for a probability π on $C_0(\mathbb{R}_+) \times \mathbb{R}_+ \times \mathbb{R}$ and $(\varphi, \psi) \in C_b(C_0(\mathbb{R}_+)) \times C_b(\mathbb{R})$ the conditions

$$\operatorname{supp} \pi \subseteq C_0(\mathbb{R}_+) \times [0, t_0] \times \mathbb{R}, \mathrm{proj}_{C_0(\mathbb{R}_+)}(\pi) = \mathbb{W}, \mathrm{proj}_{\mathbb{R}}(\pi) = \mu \tag{$p[t_0]$}$$

$$\varphi(\omega) + \psi(y) \leq c(\omega, t, y), \text{ for } \omega \in C_0(\mathbb{R}_+), t \leq t_0, y \in \mathbb{R}. \tag{$d[c, t_0]$}$$

Using compactness of $[0, t_0]$ it is not hard to see that $\tilde{c}(\omega, y) = \inf_{t \leq t_0} c(\omega, t, y)$ is continuous. We may thus apply the Monge-Kantorovich duality (Theorem 4.4) to the cost $\tilde{c}$ and obtain:

Claim 1. Taking the inf over π satisfying $(p[t_0])$ and the sup over (φ, ψ) satisfying $(d[c, t_0])$, the duality relation ($\star$) holds for continuous bounded $c : C_0(\mathbb{R}_+) \times \mathbb{R}_+ \times \mathbb{R} \to \mathbb{R}$.

Next consider the constraints

$$\operatorname{supp} \pi \subseteq C_0(\mathbb{R}_+) \times [0, t_0] \times \mathbb{R}, \mathrm{proj}_{C_0(\mathbb{R}_+)}(\pi) = \mathbb{W}, \mathrm{proj}_{\mathbb{R}}(\pi) = \mu, \pi(T) \leq V \tag{$p[t_0, V]$}$$

$$\exists \alpha \geq 0 \text{ s.t. } \varphi(\omega) + \psi(y) - \alpha(t - V) \leq c(\omega, t, y) \text{ for } \omega \in C_0(\mathbb{R}_+), t \leq t_0, y \in \mathbb{R}. \quad (d[c, t_0, V])$$

Using the min-max theorem (Theorem 4.5) with the function $F(\pi,\alpha) = \int c + \alpha(T-V)\,d\pi$, the set of π satisfying $(p[t_0])$, and $\alpha \geq 0$ we thus obtain

$$\begin{aligned}\inf_{\pi \text{ sat. } (p[t_0,V])} \int c\,d\pi &= \inf_{\pi \text{ sat. } (p[t_0])} \Big(\int c\,d\pi + \sup_{\alpha\geq 0} \alpha \int T - V\,d\pi\Big) \\ &= \sup_{\alpha\geq 0} \inf_{\pi \text{ sat. } (p[t_0])} \int c + \alpha(T-V)\,d\pi \qquad (4.7)\\ &= \sup_{\alpha\geq 0} \sup_{(\varphi,\psi) \text{ sat. } (d[c+\alpha(T-V),t_0])} \mathbb{W}(\varphi) + \mu(\psi) \qquad (4.8)\\ &= \sup_{(\varphi,\psi) \text{ sat. } (d[c,t_0,V])} \mathbb{W}(\varphi) + \mu(\psi),\end{aligned}$$

where we applied Claim 1 to the function $\tilde{c} = c + \alpha(T-V)$ to establish the equality between (4.7) and (4.8). Hence we obtain:

Claim 2. Taking the inf over π satisfying $(p[t_0,V])$ and the sup over (φ,ψ) satisfying $(d[c,t_0,V])$, the duality relation $(\star)$ holds for continuous bounded $c : C_0(\mathbb{R}_+)\times\mathbb{R}_+\times\mathbb{R} \to \mathbb{R}$.

In the next step we will drop t_0 and consider the constraints

$$\begin{aligned}\text{proj}_{C_0(\mathbb{R}_+)}(\pi) = \mathbb{W}, \text{proj}_{\mathbb{R}}(\pi) = \mu, \pi(T) \leq V \qquad (p[V])\\ \exists \alpha \geq 0 \text{ s.t. } \varphi(\omega) + \psi(y) - \alpha(t-V) \leq c(\omega,t,y), \text{ for } \omega \in C_0(\mathbb{R}_+), t \in \mathbb{R}_+, y \in \mathbb{R}. \quad (d[c,V])\end{aligned}$$

Claim 3. Taking the inf over π satisfying $(p[V])$ and the sup over (φ,ψ) satisfying $(d[c,V])$, the duality relation $(\star)$ holds for $c : C_0(\mathbb{R}_+)\times\mathbb{R}_+\times\mathbb{R} \to \mathbb{R}$ lsc and bounded from below.

Given $c \geq 0$ lsc, $\text{supp}\, c \subseteq C_0(\mathbb{R}_+)\times[0,t_0]\times\mathbb{R}$ for some t_0 it is straightforward to verify

$$\begin{aligned}\inf_{\pi \text{ sat. } (p[t_0,V])} \int c\,d\pi &= \inf_{\pi \text{ sat. } (p[V])} \int c\,d\pi,\\ \sup_{(\varphi,\psi) \text{ sat. } (d[c,t_0,V])} \mathbb{W}(\varphi) + \mu(\psi) &= \sup_{(\varphi,\psi) \text{ sat. } (d[c,V])} \mathbb{W}(\varphi) + \mu(\psi).\end{aligned}$$

Such functions can be used to approximate any non-negative lsc function on $C_0(\mathbb{R}_+)\times\mathbb{R}_+\times\mathbb{R}$ from below. Using that the set of π satisfying $(p[V])$ is compact, a straightforward approximation argument (see e.g. [58, Proof of Theorem 5.10, Step 5] for details) yields Claim 3.

Recalling (3.5), $\pi \in \mathsf{JOIN}^{1,V}(\mu)$ if and only if

$$\begin{aligned}\text{proj}_{C_0(\mathbb{R}_+)}(\pi) = \mathbb{W}, \text{proj}_{\mathbb{R}}(\pi) = \mu, \pi(T) \leq V, \quad \text{and}\\ \int f(s)(g - \mathbb{E}[g|\mathcal{F}_t^0])(\omega)k(y)\,\pi(d\omega,ds,dy) = 0 \qquad (p^M[V])\\ \text{for } f \in C_b(\mathbb{R}_+), \text{supp}\, f \subseteq [0,t], t \geq 0, g \in C_b(C_0(\mathbb{R}_+)), k \in C_b(\mathbb{R});\end{aligned}$$

here, k enforces the condition that $\mathsf{proj}_{C_0(\mathbb{R}_+)\times\mathbb{R}_+}(\pi_{\restriction C_0(\mathbb{R}_+)\times\mathbb{R}_+\times D}) \in \mathsf{RST}$ for all Borel sets D. We will apply the min-max theorem to $F(\pi,h) = \int c + h\,d\pi$, where π satisfies $(p[V])$ and

$$h(\omega,s,y) = \textstyle\sum_{i=1}^n f_i(s)(g_i - \mathbb{E}[g_i|\mathcal{F}_{t_i}^0])(\omega)k_i(y), \qquad (4.9)$$

$n \in \mathbb{N}$, $f_i \in C_b(\mathbb{R}_+)$, $\text{supp}\, f_i \subseteq [0,t_i]$, $t_i \geq 0$, $g_i \in C_b(C_0(\mathbb{R}_+))$, $k_i \in C_b(\mathbb{R})$.

The set of π satisfying $(p[V])$ is convex and compact by Prokhorov's theorem and the set of all h of the form (4.9) is a vector space as well. Hence we obtain for c continuous and bounded

$$\begin{aligned}\inf_{\pi \text{ sat. } (p^M[V])} \int c\,d\pi &= \inf_{\pi \text{ sat. } (p[V])} \sup_h \int c + h\,d\pi\\ &\overset{\text{Thm. 4.5}}{=} \sup_h \inf_{\pi \text{ sat. } (p[V])} \int c + h\,d\pi \qquad (4.10)\\ &= \sup_h \sup_{(\varphi,\psi) \text{ sat. } (d[c+h,V])} \mathbb{W}(\varphi) + \mu(\psi),\end{aligned}$$

where the last equality holds by Claim 3. Assume now that c is also predictable. For (φ,ψ) satisfying $(d[c+h,V])$ there is some $\alpha \geq 0$ such that

$$\varphi(\omega) + \psi(y) - \alpha(t-V) \leq (c+h)(\omega,t,y). \qquad (4.11)$$

Fixing t and y, (4.11) can be read as an inequality between functions in ω. Taking conditional expectations wrt $\mathcal{F}_t^0$ in the sense of Definition 3.4 we obtain

$$\varphi_t^M(\omega) + \psi(y) - \alpha(t - V) \leq c(\omega, t, y)$$

for all $\omega \in C_0(\mathbb{R}_+), t \in \mathbb{R}_+, y \in \mathbb{R}$, where we have used that c is predictable and that $\mathbb{E}[f(t)(g - \mathbb{E}[g|\mathcal{F}_u^0])|\mathcal{F}_t^0] = 0$ whenever $\operatorname{supp} f \subseteq [0, u]$.

It follows that (φ, ψ) satisfy $(d^M[c, V])$. Thus (4.10) yields the non-trivial part of $(\star)$ for the constraints $(p^M[V])$, $(d^M[c, V])$ in the case of continuous bounded c. As above, the extension to lsc c is straightforward. □

Proof of Theorem 4.2. Consider the space $C_0(\mathbb{R}_+) \times \mathbb{R}_+ \times \mathbb{R}$ and the cost function

$$c(\omega, t, y) := \begin{cases} \gamma \circ r(\omega, t) & \text{if } \omega(t) = y \\ \infty & \text{otherwise} \end{cases}. \tag{4.12}$$

It is straightforward to see that c is lsc since γ was assumed to be lsc. Hence $(\star)$ holds by Proposition 4.3. It remains to show that

$$\sup_{(\varphi,\psi) \text{ sat. } (d^M[c,V])} \mathbb{W}(\varphi) + \mu(\psi) \leq D_\gamma \text{ and } P_\gamma \leq \inf_{\pi \in \mathsf{JOIN}^{1,V}(\mu)} \pi(c). \tag{4.13}$$

To prove the first inequality, consider a bounded pair (φ, ψ) satisfying $(d^M[c, V])$, i.e. there is $\alpha \geq 0$ such that $\varphi_t^M(\omega) + \psi(y) - \alpha(t - V) \leq c(\omega, t, y)$ for all $\omega \in C_0(\mathbb{R}_+), y \in \mathbb{R}, t \in \mathbb{R}_+$. But then

$$\varphi_t^M(\omega) + \psi(\omega(t)) - \alpha(t - V) \leq \gamma \circ r(\omega, t),$$

which we rewrite as

$$[\underbrace{\varphi_t^M(\omega) - \mathbb{W}(\varphi) + \alpha(\omega(t)^2 - t)}_{=:\bar{\varphi}_t(\omega)}] + [\underbrace{\psi(\omega(t)) + \mathbb{W}(\varphi) - \alpha\omega(t)^2 + \alpha V}_{=:\bar{\psi}(\omega(t))}] \leq \gamma \circ r(\omega, t). \tag{4.14}$$

Noting that $\alpha(\omega(t)^2 - t)$ is an S-continuous martingale starting in 0, we find that $(\bar{\varphi}, \bar{\psi})$ satisfies the constraint of the dual problem considered in Theorem 4.2. Since $V = \int y^2 \mu(dy)$ we have $\int \bar{\psi}(y)\, \mu(dy) = \int \psi(y)\, \mu(dy) + \mathbb{W}(\varphi)$, establishing the first part of (4.13).

To prove the latter inequality, note that each $\pi \in \mathsf{JOIN}^{1,V}(\mu)$ satisfying $\int c\, d\pi < \infty$ is concentrated on $\{(\omega, t, y) : \omega(t) = y\}$ and writing $p(\omega, t, y) := (\omega, t)$ we find $\xi := p(\pi) \in \mathsf{RST}(\mu)$, $\int c\, d\pi = \int \gamma\, d\xi$. □

4.3. **General starting distribution.** In this section we consider $C(\mathbb{R}_+)$, the set of all continuous functions on $\mathbb{R}_+$, and

$$S_{\mathbb{R}} = \{(f, s) : f : [0, s] \to \mathbb{R} \text{ is continuous}, f(0) \in \mathbb{R}\}.$$

Let λ be a probability measure on $\mathbb{R}$ prior to μ in convex order — i.e., $\int F(x)\, \lambda(dx) \leq \int F(x)\, \mu(dx)$ for any convex function F. In particular λ is centered and $V_\lambda = \int x^2\, \lambda(dx) \leq V < \infty$. This ensures the existence of solutions to the Skorokhod embedding problem with general starting distribution λ with finite first moment. Denote by $\mathbb{W}_x$ the law of Brownian motion starting in x and put $\mathbb{W}_\lambda(d\omega) = \int \mathbb{W}_x(d\omega)\lambda(dx)$ for $\omega \in C(\mathbb{R}_+)$, the law of Brownian motion starting at a random point according to the distribution λ. Given a function $\gamma : S_{\mathbb{R}} \to \mathbb{R}$ we are interested in the minimization problem

$$P_\gamma = \inf\{\int \gamma \circ r(\omega, t)\, \xi(d\omega, dt) : \xi \in \mathsf{RST}(\lambda, \mu)\}, \tag{4.15}$$

where $\mathsf{RST}(\lambda, \mu)$ is the set of all randomized stopping times ξ on $(C(\mathbb{R}_+), \mathbb{W}_\lambda)$ embedding μ and satisfying $\xi(T) = V - V_\lambda$; in particular $\operatorname{proj}_{C(\mathbb{R}_+)}(\xi) = \mathbb{W}_\lambda$ and $h(\xi) = \mu$ for the map $h : C(\mathbb{R}_+) \times \mathbb{R}_+ \to \mathbb{R}, (\omega, t) \mapsto \omega(t)$. We then have the following result:

Theorem 4.6. *Let $\gamma : S_{\mathbb{R}} \to \mathbb{R}$ be lsc and bounded from below as in* (4.2)*. Put*

$$D_\gamma = \sup\left\{\int \psi(y)\, d\mu(y) : \psi \in C(\mathbb{R}),\ \begin{array}{l} \exists \varphi, \varphi \text{ is a } S_{\mathbb{R}}\text{-continuous martingale}, \\ \mathbb{W}_\lambda[\varphi_0] = 0, \varphi_t(\omega) + \psi(\omega(t)) \leq \gamma \circ r(\omega, t) \end{array}\right\}$$

where φ, ψ satisfy $|\varphi_t| \leq a + bt + cB_t^2$, $|\psi(y)| \leq a + by^2$ for some $a, b, c > 0$. Then we have the duality relation $P_\gamma = D_\gamma$.

The proof goes along the same lines as the proof of Theorem 4.2. The inequality $D_\gamma(\lambda,\mu) \le P_\gamma(\lambda,\mu)$ is straightforward. For the other direction we can use the same argument as before, replacing $\mathbb{W}$ by $\mathbb{W}_\lambda$ and V by $\tilde{V} := V - V_\lambda$. Up to equation (4.14) everything can be copied verbatim. Then we rewrite $\varphi_t^M(\omega) + \psi(\omega(t)) - \alpha(t - V + V_\lambda)$ as

$$[\varphi_t^M(\omega) - \mathbb{W}_\lambda(\varphi) + \alpha(\omega(t)^2 - t - V_\lambda)] + [\psi(\omega(t)) + \mathbb{W}_\lambda(\varphi) - \alpha(\omega(t)^2 - V)]$$

and note that $\mathbb{W}_\lambda(\omega(t)^2) = t + V_\lambda$. The proof concludes as before.

5. The monotonicity principle

In this section we will establish the monotonicity principle: suppose $\xi \in \mathsf{RST}(\mu)$ is an optimal stopping rule for some function γ, then we will find a set Γ supporting ξ such that $\mathsf{SG} \cap (\Gamma^< \times \Gamma) = \emptyset$. The argument can be divided into two major steps:

(1) Consider an optimal stopping rule ξ and a stop-go pair $((f,s),(g,t)) \in \mathsf{SG}$ where (f,s) is still going according to the stopping rule ξ while (g,t) is stopped by ξ. Intuitively speaking, we can find an (infinitesimal) improvement of ξ by switching the roles of f and g. As ξ is optimal, there should only exist a few such pairs. We formalize this in Proposition 5.8 by showing that if $\pi(\mathsf{SG}) > 0$ for some $\pi \in \mathsf{JOIN}(r(\xi))$ we can explicitly construct a stopping rule with strictly lower 'cost'.

(2) Knowing that SG is negligible in the sense that it is not seen by the 'couplings' π just described, it remains to find a support Γ of ξ such that $\mathsf{SG} \cap (\Gamma^< \times \Gamma) = \emptyset$. The crucial step is the characterization of a set which is null wrt all $\pi \in \mathsf{JOIN}(r(\xi))$ which we establish in Proposition 5.9 based on Choquet's capacitability theorem and an auxiliary duality result.

Armed with Propositions 5.8 and 5.9, we will establish Theorem 1.3: If ξ is an optimal stopping time, then Proposition 5.8 implies that a certain *set of pairs of paths*, i.e. the set of stop-go pairs is negligible in a quasi-sure sense, i.e. almost surely null with respect to all $\pi \in \mathsf{JOIN}(r(\xi))$. Proposition 5.9 will then allow us to exclude a $r(\xi)$-*null set of paths* to obtain a support Γ of $r(\xi)$ such that $\Gamma^< \times \Gamma$ avoids all stop-go pairs.

In the first part of this section we will give a number of definitions and results that are needed to establish Theorem 1.3 (including the statements of Propositions 5.8 and 5.9); the respective proofs will be given subsequently.

The notion of stop-go pairs introduced in Definition 1.4 requires that *all* possible extensions σ are considered. However, to establish the monotonicity principle, it is actually more natural to prove a stronger result that appeals to a relaxed notion of stop-go pairs which are sensitive to the stopping measure ξ, or – more precisely – to a representation of ξ through a function A^ξ as in Theorem 3.8 (1).

Important Convention. *Throughout this section we will fix $\xi \in \mathsf{RST}(\mu)$, as well as the particular representation A^ξ.*

Definition 5.1. *For $(f,s) \in S$, the conditional randomized stopping time $\xi^{(f,s)}$ is given as*

$$\xi_\omega^{(f,s)}([0,t]) := \begin{cases} \frac{1}{1-A^\xi(f,s)}\left(A^\xi(f \oplus \omega_{\restriction[0,t]}, s+t) - A^\xi(f,s)\right) & \text{if } A^\xi(f,s) < 1 \\ 1 & \text{otherwise} \end{cases}. \tag{5.1}$$

The measure $\xi^{(f,s)}$ is the normalized stopping measure given that we followed the path f up to time s. In other words this is the normalized stopping measure of the 'bush' which follows the 'stub' (f,s). We note that $\xi^{(f,s)}$ depends measurably on $(f,s) \in S$.

Informally, the following lemma asserts that if ξ is a well-behaved stopping time, then the same holds for $\xi^{(f,s)}$ for typical $(f,s) \in S$. More precisely, we say that $V \subseteq S$ is *evanescent* if $r^{-1}(V)$ is an evanescent subset of $C_0(\mathbb{R}_+) \times \mathbb{R}_+$. Equivalently, V is evanescent if there is a Borel set $A \subseteq C_0(\mathbb{R}_+)$, $\mathbb{W}(A) = 1$ such that $r(A \times \mathbb{R}_+) \cap V = \emptyset$. Recall that T denotes the projection from $C_0(\mathbb{R}_+) \times \mathbb{R}_+$ onto $\mathbb{R}_+$.

Lemma 5.2. *The set* $\{(f,s) \in S : \xi^{(f,s)} \notin \mathsf{RST}^1\}$ *is evanescent. Moreover, if* $F : C_0(\mathbb{R}_+) \times \mathbb{R}_+ \to \mathbb{R}_+$ *is predictable and satisfies* $\xi(F) < \infty$ *then the set* $\{(f,s) \in S : \xi^{(f,s)}(F^{(f,s)\oplus}) = \infty\}$ *is evanescent. In particular,* $\{(f,s) \in S : \xi^{(f,s)}(T) = \infty\}$ *is evanescent, since* $\xi \in \mathsf{RST}(\mu)$.

Definition 5.3. *The set* SG^ξ *of* stop-go pairs relative to ξ *consists of all* $((f,s),(g,t)) \in S \times S$, $f(s) = g(t)$ *such that*

$$\int \gamma^{(f,s)\oplus}(r(\omega,u))\, d\xi^{(f,s)}(\omega,u) + \gamma(g,t) > \gamma(f,s) + \int \gamma^{(g,t)\oplus}(r(\omega,u))\, d\xi^{(f,s)}(\omega,u). \quad (5.2)$$

We define stop-go pairs in the wide sense *by* $\widehat{\mathsf{SG}}^\xi = \mathsf{SG}^\xi \cup \{(f,s) \in S : A^\xi(f,s) = 1\} \times S$.

In analogy to Definition 1.4 we agree that (5.2) holds in any of the following cases:

(1) $\int T\, d\xi^{(f,s)} = \infty$ or $\xi^{(f,s)}(C_0(\mathbb{R}_+) \times \mathbb{R}_+) < 1$;
(2) the integral on the left side equals ∞;
(3) either of the integrals is not defined.

We now discuss the relation between the set SG given in Definition 1.4 and the set $\widehat{\mathsf{SG}}^\xi$. Note that if $A^\xi(f,s) < 1$ and $((f,s),(g,t)) \in \mathsf{SG}$ for some $(g,t) \in S$ then we shall show below that $((f,s),(g,t)) \in \mathsf{SG}^\xi$. In contrast to this, whenever $A^\xi(f,s) = 1$, the left and right hand sides of (5.2) are identical and $((f,s),(g,t))$ cannot be a stop-go pair relative to ξ. However, in general $\mathsf{SG} \cap \{(f,s) \in S : A^\xi(f,s) = 1\} \times S$ may be non-empty. For this reason we are also interested in the set of stop-go pairs in the wide sense which satisfy:

Lemma 5.4. *Every stop-go pair is a stop-go pair in the wide sense, i.e.*

$$\mathsf{SG} \subseteq \widehat{\mathsf{SG}}^\xi. \quad (5.3)$$

Remark 5.5. Note that SG^ξ and $\widehat{\mathsf{SG}}^\xi$ are Borel subsets of $S \times S$ (corresponding to predictable subsets of $(C_0(\mathbb{R}_+) \times \mathbb{R}_+) \times (C_0(\mathbb{R}_+) \times \mathbb{R}_+) = (C_0(\mathbb{R}_+) \times \mathbb{R}_+) \times \mathsf{Y}$ in the sense of Remark 3.15). In contrast, SG is in general just co-analytic.

Definition 5.6. *A set* $\Gamma \subseteq S$ *is called* (γ,ξ)-monotone *iff*

$$\widehat{\mathsf{SG}}^\xi \cap (\Gamma^< \times \Gamma) = \emptyset.$$

Recall that we say that our optimization problem (OptSEP) is well posed if $\int \gamma\, d\xi$ exists with values in $(-\infty,\infty]$ for all $\xi \in \mathsf{RST}(\mu)$ and it is finite for one such ξ. Together with Lemma 5.4, the following result implies Theorem 1.3 stated in the introduction, and is itself a slightly stronger result.

Theorem 5.7. *Assume that* $\gamma : S \to \mathbb{R}$ *is Borel measurable, the optimization problem* (4.1) *is well posed and that* $\xi \in \mathsf{RST}(\mu)$ *is an optimizer. Then there exists a* (γ,ξ)*-monotone Borel set* $\Gamma \subseteq S$ *which supports* ξ *in the sense that* $r(\xi)(\Gamma) = 1$.

The proof of Theorem 5.7 relies on Proposition 5.8 and Proposition 5.9 below. The first result formalizes the heuristic idea that an optimizer cannot be improved on a large set of paths but at most on a small set of exceptional paths. The second result allows us to entirely exclude such an exceptional set of paths.

Given functions $F : \mathsf{X} \to \mathsf{X}', G : \mathsf{Y} \to \mathsf{Y}'$ we denote the product map by $F \otimes G : \mathsf{X} \times \mathsf{Y} \to \mathsf{X}' \times \mathsf{Y}'$. Given a probability ν on a Polish space Y, we defined the set $\mathsf{JOIN}(\nu)$ in Section 3.4. An element $\pi \in \mathsf{JOIN}(\nu)$ is a measure on $(C_0(\mathbb{R}_+) \times \mathbb{R}_+) \times \mathsf{Y}$, and we will commonly consider the push-forward measure $(F \otimes G)(\pi)$. Typically F will be the map $r : C_0(\mathbb{R}_+) \times \mathbb{R}_+ \to S$, and G will be r or the identity.

Proposition 5.8. *Assume that* $\gamma : S \to \mathbb{R}$ *is Borel measurable, the optimization problem* (4.1) *is well posed and that* $\xi \in \mathsf{RST}(\mu)$ *is an optimizer. Then* $(r \otimes \mathrm{Id})(\pi)(\mathsf{SG}^\xi) = 0$ *for any* $\pi \in \mathsf{JOIN}^1(r(\xi))$.

Below we apply Proposition 5.9 to $(\mathsf{Y},\nu) = (S, r(\xi))$, but this choice is not relevant for the proof of Proposition 5.9 and so we state it for an abstract Polish probability space (Y,ν).

Proposition 5.9. *Let* (Y, ν) *be a Polish probability space and* $E \subseteq S \times \mathsf{Y}$ *a Borel set. Then the following are equivalent:*

(1) $(r \otimes \mathrm{Id})(\pi)(E) = 0$ *for all* $\pi \in \mathsf{JOIN}^1(\nu)$.
(2) $E \subseteq (F \times \mathsf{Y}) \cup (S \times N)$ *for some evanescent set* $F \subseteq S$ *and a* ν*-null set* $N \subseteq \mathsf{Y}$.

Intuitively speaking, Proposition 5.9 characterizes when a predictable set $E \subseteq S \times \mathsf{Y}$ is 'negligible'. In this sense it relates to the classical (cross) section theorem, which implies the following characterization of negligible subsets of S.

Proposition 5.10. *Let* $E \subseteq S$ *be Borel. Then the following are equivalent:*

(1) $r(\alpha)(E) = 0$ *for all* $\alpha \in \mathsf{RST}$.
(2) E *is evanescent.*
(1') $\mathbb{W}(((B_s)_{s\le\tau}, \tau) \in E) = 0$ *for every* $\mathcal{F}^0$*-stopping time* τ.

Note that the equivalence of (1) and (2) in Proposition 5.10 corresponds precisely to Proposition 5.9 in the case where Y consists of a single element.

Proof of Theorem 5.7. By Proposition 5.8, $(r \otimes \mathrm{Id})(\pi)(\mathsf{SG}^\xi) = 0$ for any $\pi \in \mathsf{JOIN}^1(r(\xi))$. Applying Proposition 5.9 with $(\mathsf{Y}, \nu) = (S, r(\xi))$ we deduce that there exists an evanescent set $\tilde{F} \subseteq S$ and a set $N \subseteq S$ such that $r(\xi)(N) = 0$, and

$$\mathsf{SG}^\xi \subseteq (\tilde{F} \times S) \cup (S \times N).$$

Put $F := \{(g,t) \in S : \exists (f,s) \in \tilde{F}, t \ge s, g \equiv f \text{ on } [0,s]\}$. Then F is evanescent and satisfies

$$\mathsf{SG}^\xi \subseteq (F \times S) \cup (S \times N).$$

Setting $\Gamma_0 = S \setminus (F \cup N)$ we have $r(\xi)(\Gamma_0) = 1$ as well as $\mathsf{SG}^\xi \cap (\Gamma_0^< \times \Gamma_0) = \emptyset$. Next we define

$$\Gamma_1 := \Gamma_0 \cap \{(g,t) \in S : A^\xi(g_{\restriction[0,s]}, s) < 1 \text{ for all } s < t\}.$$

Then $r(\xi)(\Gamma_1) = 1$ and $\Gamma_1^< \cap \{(f,s) : A^\xi(f,s) = 1\} = \emptyset$ so that $\widehat{\mathsf{SG}}^\xi \cap (\Gamma_1^< \times \Gamma_1) = \emptyset$. Finally we take Γ to be a Borel subset of Γ_1 which has full measure. □

It remains to establish the auxiliary results stated above.

Proof of Lemma 5.2. Consider

$$\begin{aligned} U_1 &= \Big\{(f,s) \in S : A^\xi(f,s) < 1, \textstyle\int d\xi^{(f,s)}(\omega,t) < 1\Big\}, \\ U_2 &= \Big\{(f,s) \in S : A^\xi(f,s) < 1, \xi^{(f,s)}(F^{(f,s)\oplus}) = \infty\Big\}. \end{aligned}$$

Set $A^\xi(\omega) := \lim_{t\to\infty} A^\xi \circ r(\omega,t)$. Then $(f,s) \in U_1$ is equivalent to $\int A^\xi(f \oplus \omega)\, d\mathbb{W}(\omega) < 1$. Given an $\mathcal{F}^0$-stopping time τ, the strong Markov property implies

$$1 = \int d\mathbb{W}(\omega)\, A^\xi(\omega) = \int d\mathbb{W}(\omega) \left[\mathbb{1}_{\tau(\omega)=\infty} A^\xi(\omega) + \mathbb{1}_{\tau(\omega)<\infty} \int d\mathbb{W}(\omega')\, A^\xi(\omega_{\restriction[0,\tau(\omega)]} \oplus \omega')\right],$$

hence $\mathbb{W}(((B_s)_{s\le\tau}, \tau) \in U_1) = 0$.

Additionally, setting $\alpha(d\omega, dt) = \delta_{\tau(\omega)}(dt)\mathbb{W}(d\omega)$ we have

$$\infty > \xi(F) \ge \int_{U_2} dr(\alpha)(f,s)\,(1 - A^\xi(f,s)) \int F^{(f,s)\oplus}(\omega,t)\, d\xi^{(f,s)}(\omega,t),$$

which implies $r(\alpha)(U_2) = 0$. Summing up, we get $\mathbb{W}(((B_s)_{s\le\tau}, \tau) \in U_1 \cup U_2) = 0$ proving the claim in view of Proposition 5.10. □

Proof of Lemma 5.4. Suppose that $A^\xi(f,s) < 1$ and $((f,s),(g,t)) \notin \widehat{\mathsf{SG}}^\xi$ for some $(g,t) \in S$ with $g(t) = f(s)$. In particular, (5.2) fails for $\xi^{(f,s)}$, and conditions (1)–(3) above all fail. By Theorem 3.8 (4), and using the same argument as seen in the proof of Lemma 3.11, we can find a $(\mathcal{F}_t^0 \otimes \mathcal{B}([0,1]))_{t\ge0}$-stopping time ρ such that $\overline{\mathbb{W}}(\rho > 0) > 0$ and for any measurable and bounded or non-negative $X : C_0(\mathbb{R}_+) \times \mathbb{R}_+ \to \mathbb{R}$, we have $\int X_t(\omega)\, d\xi^{(f,s)}(\omega,t) = \int \mathcal{L}(du) \int \mathbb{W}(d\omega) X_{\rho(\omega,u)}(\omega)$. By the conditions below Definition 5.3, it follows that there exists $u_0 \in [0,1]$ such that

$$\int \gamma^{(f,s)\oplus}(B_{\rho(\omega,u_0)})\, d\mathbb{W}(\omega) + \gamma(g,t) \le \gamma(f,s) + \int \gamma^{(g,t)\oplus}(B_{\rho(\omega,u_0)})\, d\mathbb{W}(\omega), \tag{5.4}$$

and such that $\rho_0 : \omega \mapsto \rho(\omega, u_0)$ is an $(\mathcal{F}_t^0)_{t\geq 0}$-stopping time with $0 < \mathbb{W}(\rho_0) < \infty$, both sides of (5.4) are well defined, and the left hand side is finite. In particular, writing B^Ω for Brownian motion on the abstract probability space Ω, $\sigma = \rho_0 \circ B^\Omega$ defines an $\mathcal{F}^B$-stopping time, and (1.6) fails for this stopping time. Hence $((f,s),(g,t)) \notin \mathsf{SG}$. □

5.1. **Proof of Proposition 5.8.** Working towards a contradiction we assume that there is $\pi \in \mathsf{JOIN}(r(\xi))$ such that $(r \otimes \mathrm{Id})(\pi)(\mathsf{SG}^\xi) > 0$. Observe that $\pi \in \mathsf{JOIN}(r(\xi))$ implies that $\pi_{\restriction (r\otimes \mathrm{Id})^{-1}(E)} \in \mathsf{JOIN}(r(\xi))$ for any $E \subseteq S \times S$. Hence, considering $(r \otimes \mathrm{Id})(\pi)_{\restriction \mathsf{SG}^\xi}$, we can also assume that $(r \otimes \mathrm{Id})(\pi)$ is concentrated on SG^ξ and then $r(\mathrm{proj}_{\mathsf{X}}(\pi))(\{(f,s) : A^\xi(f,s) = 1\}) = 0$, where $\mathsf{X} := C_0(\mathbb{R}_+) \times \mathbb{R}_+$. Finally we also consider the representation of π on $(C_0(\mathbb{R}_+) \times \mathbb{R}_+) \times (C_0(\mathbb{R}_+) \times \mathbb{R}_+)$ defined through

$$\bar{\pi}(C \times D) := \int d\pi((\omega,s),(g,t)) \int d\mathbb{W}(\eta)\, \mathbb{1}_{C\times D}((\omega,s),((g,t)\oplus\eta)) \tag{5.5}$$

and note that $\pi = (\mathrm{Id} \otimes r)(\bar{\pi})$.

We will use π and $\bar{\pi}$ to define modifications $\xi_0^\pi \in \mathsf{RST}$ and $\xi_1^\pi \in \mathsf{RST}$ of ξ such that the following hold true:

(1) The terminal distributions μ_0, μ_1 corresponding to ξ_0^π and ξ_1^π satisfy $(\mu_0+\mu_1)/2 = \mu$.
(2) ξ_0^π stops paths earlier than ξ while ξ_1^π stops later than ξ.
(3) The cost of ξ_0^π plus the cost of ξ_1^π is less than twice the cost of ξ, i.e.

$$\int \gamma \circ r(\omega,t)\, d\xi_0^\pi(\omega,t) + \int \gamma \circ r(\omega,t)\, d\xi_1^\pi(\omega,t) < 2 \int \gamma \circ r(\omega,t)\, d\xi(\omega,t).$$

More formally, (2) asserts that for almost all ω, and every $s \geq 0$

$$(\xi_0^\pi)_\omega([0,s]) \geq \xi_\omega([0,s]) \quad \text{and} \quad (\xi_1^\pi)_\omega([0,s]) \leq \xi_\omega([0,s]),$$

where $(\xi_\omega)_{\omega\in C_0(\mathbb{R}_+)}$ is the disintegration of ξ wrt $\mathbb{W}$ induced by A^ξ and $((\xi_0^\pi)_\omega)_{\omega\in C_0(\mathbb{R}_+)}$, $((\xi_1^\pi)_\omega)_{\omega\in C_0(\mathbb{R}_+)}$ are disintegrations of ξ_0^π, ξ_1^π wrt $\mathbb{W}$.

If we are able to construct such a pair ξ_0^π, ξ_1^π, then $\xi^\pi := (\xi_0^\pi + \xi_1^\pi)/2 \in \mathsf{RST}(\mu)$ is strictly better than ξ and therefore yields the desired contradiction.

In the proof we will often use the following 'strong Markov property' of randomized stopping times: for $\alpha \in \mathsf{RST}$ and bounded measurable $F : C_0(\mathbb{R}_+) \times \mathbb{R}_+ \to \mathbb{R}$ we have

$$\iint F^{(g,t)\oplus}(\omega)\, d\mathbb{W}(\omega)\, dr(\alpha)(g,t) = \int F(\omega,t)\, d\alpha(\omega,t)\,. \tag{5.6}$$

To define ξ_0^π, let $\alpha_0 = \mathsf{proj}_{\mathsf{X}}(\pi) \in \mathsf{RST}$ and consider $A^{\alpha_0} : S \to [0,1]$ as in Theorem 3.8 (1). We define the randomized stopping time ξ_0^π via the product

$$(1 - A^{\xi_0^\pi})(f,s) := (1 - A^{\alpha_0})(f,s) \cdot (1 - A^\xi)(f,s).$$

The probabilistic interpretation of this definition is that a particle is stopped by ξ_0^π if it is stopped by α_0 or stopped by ξ, where these events are taken to be conditionally independent given the particle followed the path f until time s. Comparing ξ and ξ_0^π the latter will stop some particles earlier than the first one. Also, $\xi_0^\pi \in \mathsf{RST}$ by Theorem 3.8 (1). By partial integration, if $D \subseteq C_0(\mathbb{R}_+) \times \mathbb{R}_+$ then ξ_0^π satisfies

$$\xi_0^\pi(D) \;=\; \int_D (1 - A^\xi \circ r(\omega,t))\, d\alpha_0(\omega,t) + \int_D (1 - (A^{\alpha_0} \circ r)_-(\omega,t))\, d\xi(\omega,t),$$

where $(A^{\alpha_0} \circ r)_-$ denotes the left continuous version of $A^{\alpha_0} \circ r$.

Our next goal is to derive (in (5.9) below) a representation for the difference between ξ_0^π and ξ. For Borel $D \subseteq C_0(\mathbb{R}_+) \times \mathbb{R}_+$ we have

$$\xi_0^\pi(D) - \xi(D) = \int_D (1 - A^\xi \circ r(\omega,t))\, d\alpha_0(\omega,t) - \int_D (\alpha_0)_\omega([0,t))\, d\xi(\omega,t). \tag{5.7}$$

Furthermore, writing $D_\omega = \{t \in \mathbb{R}_+ : (\omega,t) \in D\}$ and $\theta_s(\omega) = (\omega_{t+s} - \omega_s)_{t\geq 0}$, we have

$$\begin{aligned}
\int_D (\alpha_0)_\omega([0,t))\, d\xi(\omega,t) &= \int_{C_0(\mathbb{R}_+)} d\mathbb{W}(\omega) \int_{\mathbb{R}_+} d\xi_\omega(t) \int_{\mathbb{R}_+} d(\alpha_0)_\omega(s)\, \mathbb{1}_D(\omega,t) \mathbb{1}_{[0,t)}(s) \\
&= \int_{C_0(\mathbb{R}_+)} d\mathbb{W}(\omega) \int_{\mathbb{R}_+} d(\alpha_0)_\omega(s) \int_{\mathbb{R}_+} d\xi_\omega(t)\, \mathbb{1}_{D_\omega}(t) \mathbb{1}_{(s,\infty)}(t) \\
&= \int_{C_0(\mathbb{R}_+)\times\mathbb{R}_+} d\alpha_0(\omega,s)\, (1 - A^\xi \circ r(\omega,s)) \xi_{\theta_s(\omega)}^{r(\omega,s)}(D_\omega - s).
\end{aligned} \tag{5.8}$$

Combining (5.7) and (5.8) we obtain for bounded measurable $F : S \to \mathbb{R}$ using (5.6)

$$\begin{aligned}&\int F \circ r \, d(\xi_0^\pi - \xi) = \\ &\int d\alpha_0(\omega, s)\,(1 - A^\xi \circ r(\omega, s))\Big[F \circ r(\omega, s) - \int F^{r(\omega,s)\oplus}(r(\tilde\omega, u))\, d\xi^{r(\omega,s)}(\tilde\omega, u)\Big] = \\ &\int d\bar\pi((\omega, s), (\eta, t))\,(1 - A^\xi \circ r(\omega, s))\Big[F \circ r(\omega, s) - \int F^{r(\omega,s)\oplus}(r(\tilde\omega, u))\, d\xi^{r(\omega,s)}(\tilde\omega, u)\Big].\end{aligned} \tag{5.9}$$

Let us now turn to the definition of ξ_1^π. For $D \subseteq C_0(\mathbb{R}_+) \times \mathbb{R}_+$ we define

$$\alpha_1(D) \;=\; \int_{\mathsf{X}\times D} (1 - A^\xi \circ r(\omega, s))\, d\bar\pi((\omega, s), (\eta, t)) \tag{5.10}$$

and observe that $\alpha_1 \in \mathsf{RST}$ by Theorem 3.8 (2) since $\eta \mapsto (\alpha_1)_\eta([0, t])$ is $\mathcal{F}_t^a$-measurable by (5.5). Then we define the probability measure ξ_1^π on $C_0(\mathbb{R}_+) \times \mathbb{R}_+$ by

$$\xi_1^\pi(D) := \xi(D) - \alpha_1(D) + \int (1 - A^\xi \circ r(\omega, s))\, \xi_{\theta_t(\eta)}^{r(\omega,s)}(D_\eta - t)\, d\bar\pi((\omega, s), (\eta, t)). \tag{5.11}$$

To motivate this definition, we note that the support of the randomized stopping time ξ can be viewed (informally) as a sub-tree of S. The joining π defines a plan how to trim this tree, i.e. to cut a bush at position $r(\omega, s)$ and to plant it on top of $r(\eta, t)$. Hence, we take the tree, ξ, prepare the position where something will be newly planted, subtract α_1 which takes away some mass, and plant as much as possible (accounting for the factor $(1 - A^\xi \circ r(\omega, s))$ in (5.10) and (5.11)) on these stubs to end up with a tree of mass one again.

As a consequence of Definition 5.1 for each u the map $\eta \mapsto \xi_{\theta_t(\eta)}^{(f,s)}([0, (u - t) \vee 0])$ is $\mathcal{F}_u^0$-measurable. Moreover, $(\xi_1^\pi)_\eta \in \mathcal{P}^{\le 1}(\mathbb{R}_+)$ and it follows that $\xi_1^\pi \in \mathsf{RST}$. From (5.10) and (5.11) it follows that for bounded measurable $F : S \to \mathbb{R}$ using (5.5)

$$\begin{aligned}&\int F \circ r \, d(\xi_1^\pi - \xi) = \\ &\int d\bar\pi((\omega, s), (\eta, t))\,(1 - A^\xi \circ r(\omega, s))\Big[\int F^{r(\eta,t)\oplus}(r(\tilde\omega, u))\, d\xi^{r(\omega,s)}(\tilde\omega, u) - F \circ r(\eta, t)\Big].\end{aligned} \tag{5.12}$$

Adding (5.9) and (5.12) and recalling $2\xi^\pi = \xi_0^\pi + \xi_1^\pi$, we obtain for bounded measurable $F : S \to \mathbb{R}$

$$\begin{aligned}2\int F \circ r \, d(\xi^\pi - \xi) &= \int d\bar\pi((\omega, s), (\eta, t))\,(1 - A^\xi \circ r(\omega, s)) \\ &\Big[F \circ r(\omega, s) + \int F^{r(\eta,t)\oplus}(r(\tilde\omega, u))\, d\xi^{r(\omega,s)}(\tilde\omega, u) - \int F^{r(\omega,s)\oplus}(r(\tilde\omega, u))\, d\xi^{r(\omega,s)}(\tilde\omega, u) - F \circ r(\eta, t)\Big].\end{aligned} \tag{5.13}$$

Next we show that (5.13) extends to non-negative functions $F : S \to \mathbb{R}_+$ satisfying $\xi(F) < \infty$. Put $X(\omega) := \int F(r(\omega, t))\, \xi_\omega(dt)$. Then $\mathbb{E}[X] = \xi(F) < \infty$. Moreover, recalling Definition 3.4 we have

$$X_s^M(\omega) = \int_0^s F(r(\omega, t))\, d\xi_\omega(t) + \int F^{r(\omega,s)\oplus}(r(\tilde\omega, u)) \cdot (1 - A^\xi(r(\omega, s))\, d\xi^{r(\omega,s)}(\tilde\omega, u)\,.$$

It then follows that

$$\begin{aligned}&\iint F^{r(\omega,s)\oplus}(r(\tilde\omega, u)) \cdot (1 - A^\xi(r(\omega, s))\, d\xi^{r(\omega,s)}(\tilde\omega, u)\, d\bar\pi((\omega, s), (\eta, t)) \\ &\le \int X_s^M(\omega)\, d\bar\pi((\omega, s), (\eta, t)) = \bar{\mathbb{E}}[\bar X_{\rho_0}^M \mathbb{1}_{\rho_0<\infty}] \le \mathbb{E}[X_\infty^M] = \mathbb{E}[X] < \infty\,,\end{aligned}$$

where ρ_0 denotes the representation of α_0 as in (3.6) and $\bar X_t^M(\omega, u) = X_t^M(\omega)$. This implies

$$\int d\bar\pi((\omega, s), (\eta, t))\,(1 - A^\xi \circ r(\omega, s))\Big[\int F^{r(\omega,s)\oplus}(r(\tilde\omega, u))\, d\xi^{r(\omega,s)}(\tilde\omega, u) + F \circ r(\eta, t)\Big] < \infty,$$

hence (5.13) holds also for such F. Applying this to $F(f, s) = s$ we find $\xi^\pi(T) = \xi(T) < \infty$. Taking $F(f, s) = G(f(s))$ for bounded measurable $G : \mathbb{R} \to \mathbb{R}$, the right hand side of (5.13) vanishes since $\bar\pi$ is concentrated on pairs $((\omega, s), (\eta, t))$ satisfying $\omega(s) = \eta(t)$. This implies that ξ and ξ^π embed the same distribution, i.e. $\xi^\pi \in \mathsf{RST}(\mu)$.

Arguing on the negative and positive part of γ and using that $\xi^\pi(\gamma^-), \xi(\gamma^-) < \infty$ we see that (5.13) applies to $F = \gamma$. By definition of SG^ξ,

$$(1 - A^\xi \circ r(\omega, s))\Big[\gamma \circ r(\omega, s) + \int \gamma^{r(\eta,t)\oplus} \circ r\, d\xi^{r(\omega,s)} - \int \gamma^{r(\omega,s)\oplus} \circ r\, d\xi^{r(\omega,s)} - \gamma \circ r(\eta, t)\Big]$$

is $\bar\pi$-a.s. strictly negative since $r(\mathrm{proj}_{\mathsf{X}}(\pi))(\{(f, s) : \xi^{(f,s)}(T) = \infty \text{ or } \xi^{(f,s)} \notin \mathsf{RST}^1\}) = 0$ by Lemma 5.2. Hence $\xi^\pi(\gamma) < \xi(\gamma)$, contradicting optimality of ξ. □

5.2. **Proof of Proposition 5.9.** Only the implication (1) ⇒ (2) of Proposition 5.9 is non-trivial. The proof is based on Choquet's capacitability theorem and the following auxiliary duality result which is closely related to Proposition 4.3. We fix $t_0 \in \mathbb{R}_+$ and set $S_{t_0} := \{(f,s) \in S : s \le t_0\}$.

Proposition 5.11. *Consider a Polish probability space* (Y, ν) *and let* $c : C_0(\mathbb{R}_+) \times \mathbb{R}_+ \times \mathsf{Y} \to \mathbb{R} \cup \{\infty\}$ *be lsc, predictable (cf. Remark 3.15) and bounded from below. Then*

$$\inf_{\pi} \int c(\omega, t, y)\, \pi(d\omega, dt, dy) = \sup_{(\varphi,\psi)} (\mathbb{W}(\varphi) + \nu(\psi)) \tag{$\star\star$}$$

where the infimum is taken over the set

$$\mathsf{JOIN}^1_{t_0}(\nu) = \{\pi \in \mathsf{JOIN}^1(\nu) : \operatorname{supp} \pi \subseteq C_0(\mathbb{R}_+) \times [0, t_0] \times \mathsf{Y}\}$$

and the supremum is taken over $\varphi \in C_b(C_0(\mathbb{R}_+))$, $\psi \in C_b(\mathsf{Y})$ *such that*

$$\varphi^M_t(\omega) + \psi(y) \le c(\omega, t, y), \text{ for } \omega \in C_0(\mathbb{R}_+), t \le t_0, y \in \mathsf{Y}.$$

Proof. As the arguments are almost identical to the ones from Proposition 4.3 we will only sketch the proof. By approximation it is sufficient to establish the result for continuous bounded c. As before the Monge-Kantorovich duality yields that ($\star\star$) holds provided that π and (φ, ψ), resp., satisfy

$$\operatorname{supp} \pi \subseteq C_0(\mathbb{R}_+) \times [0, t_0] \times \mathsf{Y}, \operatorname{proj}_{C_0(\mathbb{R}_+)}(\pi) = \mathbb{W}, \operatorname{proj}_{\mathsf{Y}}(\pi) = \nu$$
$$\varphi(\omega) + \psi(y) \le c(\omega, t, y), \text{ for } \omega \in C_0(\mathbb{R}_+), t \le t_0, y \in \mathsf{Y}.$$

If c is predictable, we can then argue as in the last step of Proposition 4.3 to obtain the assertion of Proposition 5.11 □

We now state several consequences of Proposition 5.11 in which we switch the roles of inf and sup to provide a more natural formulation.

Denote for Borel $K \subseteq S_{t_0} \times \mathsf{Y}$

$$D_{t_0}(K) := \inf_{(\varphi,\psi) \in \mathsf{DC}_{t_0}(K)} (\mathbb{W}(\varphi) + \nu(\psi)) \tag{5.14}$$

where $\mathsf{DC}_{t_0}(K)$ consists of all pairs of lsc φ, ψ on $C_0(\mathbb{R}_+)$ resp. Y satisfying

$$0 \le \varphi, \psi \le 1, \mathbb{1}_K((f,s), y) \le \varphi^{M,S}(f,s) + \psi(y), (f,s) \in S_{t_0}, y \in \mathsf{Y}, \tag{5.15}$$

where we recall the notation $\varphi^{M,S}$ from Definition 3.4.

Corollary 5.12. *Let* $K \subseteq S_{t_0} \times \mathsf{Y}$ *be closed. Then*

$$\sup_{\pi \in \mathsf{JOIN}^1_{t_0}(\nu)} (r \otimes \mathrm{Id})(\pi)(K) = D_{t_0}(K).$$

Proof. Fix $\varepsilon > 0$. Applying Proposition 5.11 to $c = -\mathbb{1}_{(r\otimes \mathrm{Id})^{-1}(K)}$, which is lsc due to the continuity of r, we obtain that there exist functions $\varphi \in C_b(C_0(\mathbb{R}_+))$, $\psi \in C_b(\mathsf{Y})$ such that

$$\mathbb{1}_K((f,s), y) \le \varphi^{M,S}(f,s) + \psi(y), \text{ for } (f,s) \in S_{t_0}, y \in \mathsf{Y}, \quad \text{and} \tag{5.16}$$
$$\sup_{\pi \in \mathsf{JOIN}^1_{t_0}(\nu)} \int \mathbb{1}_{(r\otimes \mathrm{Id})^{-1}(K)}\, d\pi = \sup_{\pi \in \mathsf{JOIN}^1_{t_0}(\nu)} (r \otimes \mathrm{Id})(\pi)(K) > \mathbb{W}(\varphi) + \nu(\psi) - \varepsilon.$$

It follows from (5.16) that $\varphi^{M,S}$ is bounded from below on S_{t_0} and wlog we may assume that $\varphi(\omega) = \varphi^M_{t_0}(\omega)$. Subtracting a constant from φ and adding it to ψ, we may assume that $\inf \varphi = 0$ (which implies $\psi \ge 0$). It follows that we can replace ψ with $\bar{\psi} = \psi \wedge 1$.

It suffices to consider the case $\mathbb{W}(\varphi) \le 1$. Put $\rho = \inf\{t \ge 0 : \varphi^M_t > 1\}$. Due to S-continuity of φ^M (by Proposition 3.5) the set $O := \{(\omega, t) : \varphi^{M,S} \circ r(\omega, t) > 1\}$ is open. Hence also $\{\rho < \infty\} = \operatorname{proj}_{C_0(\mathbb{R}_+)} O$ is open as projections are open mappings and the map $\omega \mapsto \varphi^M_{\rho(\omega)}(\omega) =: \bar{\varphi}(\omega) \le 1$ is lsc. Clearly, $(\bar{\varphi}, \bar{\psi})$ satisfies (5.15) and $\mathbb{W}(\bar{\varphi}) + \nu(\bar{\psi}) \le \mathbb{W}(\varphi) + \nu(\psi)$. □

Lemma 5.13. D_{t_0} *is a Choquet capacity on* $S \times \mathsf{Y}$.

Proof. We need to verify the defining properties of a capacity Ψ (cf. [35, Definition 30.1]):

(1) monotonicity: $A \subseteq B \Rightarrow \Psi(A) \le \Psi(B)$

(2) continuity from below: $A_1 \subseteq A_2 \subseteq ... \Rightarrow \Psi(A_n) \to \Psi(\bigcup_j A_j)$
(3) boundedness: $\Psi(K) < \infty$ for all compact K; if $\Psi(K) < u$ there exists open $U \supseteq K$ with $\Psi(U) < u$.

Moreover, it is sufficient to test these properties for Borel sets (see [35, Section 30B]). The monotonicity is immediate. Let us turn to the continuity from below.

Take an increasing sequence $A_1 \subseteq A_2 \subseteq \ldots \subseteq S \times \mathsf{Y}$ of Borel sets and put $A = \bigcup_n A_n$. For all n there are lsc functions $\varphi_n : C_0(\mathbb{R}_+) \to [0,1]$ (which give rise to S-lsc martingales) and $\psi_n : \mathsf{Y} \to [0,1]$ such that $\mathbb{1}_{A_n}((f,s),y) \le \varphi_n^{M,S}(f,s) + \psi_n(y)$ for all $(f,s) \in S_{t_0}, y \in \mathsf{Y}$ and

$$\nu(\psi_n) + \mathbb{W}(\varphi_n) \le D_{t_0}(A_n) + 1/n.$$

Using a Mazur/Komlos-type lemma (e.g. Lemma A1.1 in [15]) we can assume that some appropriate convex combinations of ψ_n and φ_n converge a.s. to functions ψ and φ. More precisely: there exist convex coefficients $\alpha_n^n, \ldots, \alpha_{k_n}^n, n \ge 1, k_n < \infty$, and full measure subsets $\Omega_1 \subseteq C_0(\mathbb{R}_+)$, $\mathsf{Y}_1 \subseteq \mathsf{Y}$ such that with $\tilde{\varphi}_n := \sum_{i=n}^{k_n} \alpha_i^n \varphi_i$, $\tilde{\psi}_n := \sum_{i=n}^{k_n} \alpha_i^n \psi_i$ we have that for all $\omega \in \Omega_1$ and all $y \in \mathsf{Y}_1$

$$\lim_{n\to\infty} \tilde{\varphi}_n(\omega) =: \varphi(\omega) \text{ and } \lim_{n\to\infty} \tilde{\psi}_n(y) =: \psi(y) \tag{5.17}$$

exist. Extend these functions to $C_0(\mathbb{R}_+)$ and Y, resp., through

$$\limsup_{n\to\infty} \tilde{\varphi}_n(\omega) =: \varphi(\omega) \text{ and } \limsup_{n\to\infty} \tilde{\psi}_n(y) =: \psi(y). \tag{5.18}$$

This implies for $(f,s) \in S$

$\limsup_{n\to\infty} \tilde{\varphi}_n^{M,S}(f,s) \le \int \limsup_{n\to\infty} \tilde{\varphi}_n(f \oplus \omega)\, \mathbb{W}(d\omega) = \int \varphi(f \oplus \omega)\, \mathbb{W}(d\omega) = \varphi^{M,S}(f,s).$

Given $m \le n$ we have for $(f,s) \in S_{t_0}, y \in \mathsf{Y}$

$$\mathbb{1}_{A_m}((f,s),y) \le \tilde{\varphi}_n^{M,S}(f,s) + \tilde{\psi}_n(y),$$

hence $\mathbb{1}_{A_m}((f,s),y) \le \varphi^{M,S}(f,s) + \psi(y)$ and thus also

$$\mathbb{1}_A((f,s),y) \le \varphi^{M,S}(f,s) + \psi(y).$$

Given $\varepsilon > 0$, we can find lsc functions $\varphi^\varepsilon \ge \varphi$ and $\psi^\varepsilon \ge \psi$ such that $\mathbb{W}(\varphi^\varepsilon) - \varepsilon/2 < \mathbb{W}(\varphi) = \lim \mathbb{W}(\tilde{\varphi}_n)$ and $\nu(\psi^\varepsilon) - \varepsilon/2 < \nu(\psi) = \lim \nu(\tilde{\psi}_n)$. It follows that

$$D_{t_0}(A) \le \limsup_n D_{t_0}(A_n) + 1/n + \varepsilon.$$

Let us turn to the third property. Trivially, $D_{t_0}(K) \le 1$, so take a compact set $K \subseteq S \times \mathsf{Y}$ and fix $\varepsilon > 0$. By Corollary 5.12 there is $(\varphi,\psi) \in \mathsf{DC}_{t_0}(K)$ such that

$$\nu(\psi) + \mathbb{W}(\varphi) \le D_{t_0}(K) + \varepsilon.$$

As $(\varphi,\psi) \in \mathsf{DC}_{t_0}(K)$ we have $K \subseteq \{((f,s),y) : \varphi^{M,S}(f,s) + \psi(y) \ge 1\}$. At the additional cost of $2\,\varepsilon$ we can find two lsc functions $\varphi^\varepsilon := (\varphi + \varepsilon) \wedge 1 \ge \varphi$ and $\psi^\varepsilon := (\psi + \varepsilon) \wedge 1 \ge \psi$ such that $\mathbb{W}(\varphi^\varepsilon) + \nu(\psi^\varepsilon) \le \mathbb{W}(\varphi) + \nu(\psi) + 2\varepsilon$ and $K \subseteq \{((f,s),y) : (\varphi^\varepsilon)^{M,S}(f,s) + \psi^\varepsilon(y) > 1\}$. By lower semi-continuity, $U := \{((f,s),y) : (\varphi^\varepsilon)^{M,S}(f,s) + \psi^\varepsilon(y) > 1\}$ is open. Hence, for every $\varepsilon > 0$ we have found an open $U \supseteq K$ such that $D_{t_0}(U) \le D_{t_0}(K) + 3\varepsilon$, proving the last claim. □

The next step is to show that up to a factor of 2 we can restrict ourselves to dual functions φ and ψ which are indicator functions. The simple reason is that if $1 \le a + b$ then $a > 1/2$ or $b \ge 1/2$. In the formulation of the next lemma and subsequently we use the notation

$$\mathrm{deb}_{t_0}(F) := \{\omega : \exists t < t_0, r(\omega,t) \in F\}.$$

Lemma 5.14. *Let $K \subseteq S_{t_0} \times \mathsf{Y}$ be Borel. Then*

$$\inf_{(F,A)\in\mathsf{Cov}(K)} \Big(\mathbb{W}(\mathrm{deb}_{t_0}(F))) + \nu(A)\Big) \le 2D_{t_0}(K), \tag{5.19}$$

where $\mathsf{Cov}(K) = \{F \subseteq S \text{ open}, A \subseteq \mathsf{Y} : K \subseteq (F \times \mathsf{Y}) \cup (S \times A)\}$.

Proof. We may assume $D_{t_0}(K) < 1/2$, otherwise simply take $A = \mathsf{Y}, F = \emptyset$.

Take $(\varphi, \psi) \in \mathsf{DC}_{t_0}(K)$. As the cost function is $\{0, 1\}$-valued, the dual constraint

$$\mathbb{1}_K((f, s), y) \leq \varphi^{M,S}(f, s) + \psi(y)$$

implies that

$$K \subseteq (\{(f, s) : \varphi^{M,S}(f, s) > 1/2\} \times \mathsf{Y}) \cup (S \times \{y : \psi(y) \geq 1/2\}).$$

Recalling that $0 \leq \psi \leq 1$ we set $A = \{\psi \geq 1/2\}$ and note that $\nu(A)/2 \leq \nu(\psi)$.

Let us turn our attention to the set $F = \{(f, s) : \varphi^{M,S}(f, s) > 1/2\}$. As $D_{t_0}(K) < 1/2$, we may assume that $\varphi^{M,S}(0, 0) < 1/2$. Given $\varepsilon > 0$ we apply the optional section theorem to $r^{-1}(F) \cap (C_0(\mathbb{R}_+) \times [0, t_0))$ to obtain a stopping time τ such that $\mathbb{W}(\tau < t_0) > \mathbb{W}(\mathrm{deb}_{t_0}(F)) - \varepsilon$ and $\varphi^M_\tau > 1/2$ on $\{\tau < t_0\}$. By optional stopping

$$\mathbb{E}[\varphi^M_0] = \mathbb{E}[\varphi^M_\tau] \geq \mathbb{W}(\tau < t_0)/2.$$

As $\varepsilon > 0$ was arbitrary, $\mathbb{W}(\mathrm{deb}_{t_0}(F)) + \nu(A) \leq 2(\mathbb{E}[\varphi^M_0] + \nu(\psi))$, establishing (5.19). □

Proof of Proposition 5.9. Assume first that $E \subseteq S_{t_0} \times \mathsf{Y}$. We have $\sup_{\pi \in \mathsf{JOIN}^1(\nu)} \pi(K) = 0$ for all compact $K \subseteq E$. By Corollary 5.12, this implies that $D_{t_0}(K) = 0$ for all compact $K \subseteq E$. By Choquet's capacitability theorem [35, Theorem 30.13] and Lemma 5.13 this in turn implies $D_{t_0}(E) = 0$.

Hence, by Lemma 5.14, for each $\varepsilon > 0$ there exist $F \subseteq S$ and a set $N \subseteq \mathsf{Y}$ such that $E \subseteq (F \times \mathsf{Y}) \cup (S \times N)$ and $\mathbb{W}(\mathrm{deb}_{t_0}(F)) + \nu(N) \leq 2\varepsilon$.

For each k, pick some set $F_k \subseteq S$ and a set $N_k \subseteq \mathsf{Y}$ such that $E \subseteq (F_k \times \mathsf{Y}) \cup (S \times N_k)$ and $\mathbb{W}(\mathrm{deb}_{t_0}(F_k)) + \nu(N_k) \leq 2^{-k}$. Setting $F = \limsup_k F_k$ and $N = \limsup_k N_k$ we get $\mathbb{W}(\mathrm{deb}_{t_0}(F)) = 0$, $\nu(N) = 0$ and

$$E \subseteq (F \times \mathsf{Y}) \cup (S \times N).$$

To establish the result in the case of general $E \subseteq S \times \mathsf{Y}$, for each $n \in \mathbb{N}$ pick sets $N_n \subseteq \mathsf{Y}, \nu(N_n) = 0, F_n \subseteq S, \mathbb{W}(\mathrm{deb}_n(F_n)) = 0$ such that $E \cap (S_n \times \mathsf{Y}) \subseteq (F_n \times \mathsf{Y}) \cup (S \times N_n)$. Then $N := \bigcup_{n \geq 1} N_n$ and $F := \bigcup_{n \geq 1} F_n$ are as required. □

5.3. **A secondary minimization result.** In certain cases, in order to resolve possible non-uniqueness of a minimizer, it will be useful to identify particular solutions as the solution not only to a primary optimization result, but also as the unique optimizer within this class of a second minimization problem. To this end, we begin by making the following definition: Supposing that $\gamma, \tilde{\gamma} : S \to \mathbb{R}$ are Borel measurable, we write Opt_γ for the set of optimizers of (4.1). If $\mathsf{Opt}_\gamma \neq \emptyset$, we consider the secondary optimization problem

$$P_{\tilde{\gamma}|\gamma} = \inf_{\xi \in \mathsf{Opt}_\gamma} \int \tilde{\gamma} \, d\xi. \tag{5.20}$$

We will say that (5.20) is *well posed* if the primary optimization problem (4.1) is well posed and $\int \tilde{\gamma} \, d\xi$ exists with values in $(-\infty, \infty]$ for all $\xi \in \mathsf{Opt}_\gamma$ and is finite for one such ξ. Observe that, when P_γ is finite and the map $\pi \mapsto \int \gamma \, d\pi$ is lsc the set Opt_γ is a closed subset of $\mathsf{RST}(\mu)$, and hence also compact.

We need an extended version of the stop-go pairs introduced in Definition 5.3.

Definition 5.15. *Let $\gamma, \tilde{\gamma} : S \to \mathbb{R}$ be Borel measurable. The set of secondary stop-go pairs SG^ξ_2 (relative to ξ) consists of all $((f, s), (g, t)) \in S \times S$, $f(s) = g(t)$ such that either $((f, s), (g, t)) \in \mathsf{SG}^\xi$, or*

$$\begin{aligned} &\int \gamma^{(f,s)\oplus} \circ r \, d\xi^{(f,s)} + \gamma(g, t) = \gamma(f, s) + \int \gamma^{(g,t)\oplus} \circ r \, d\xi^{(f,s)} \\ \text{and } &\int \tilde{\gamma}^{(f,s)\oplus} \circ r \, d\xi^{(f,s)} + \tilde{\gamma}(g, t) > \tilde{\gamma}(f, s) + \int \tilde{\gamma}^{(g,t)\oplus} \circ r \, d\xi^{(f,s)}. \end{aligned} \tag{5.21}$$

As before, we also say that (5.21) *holds if any of the integrals in the second equation are not defined, or the integral on the left-hand side equals ∞.*

We also define secondary stop-go pairs in the wide sense *by* $\widehat{\mathsf{SG}}^\xi_2 = \mathsf{SG}^\xi_2 \cup \{(f, s) \in S : A^\xi(f, s) = 1\} \times S$.

Then we have the following generalization of Theorem 5.7.

Theorem 5.16. *Let $\gamma, \tilde{\gamma}$ be Borel measurable functions on S. Suppose that $\mathsf{Opt}_\gamma \neq \emptyset$, and that the optimization problem* (5.20) *is well posed with optimizer $\xi \in \mathsf{Opt}_\gamma$. Then there exists a Borel set $\Gamma \subseteq S$ such that $r(\xi)(\Gamma) = 1$ and*

$$\widehat{\mathsf{SG}}_2^{\xi} \cap (\Gamma^< \times \Gamma) = \emptyset. \tag{5.22}$$

The proof given for Theorem 5.7 also applies in the present situation. Hence, the result follows immediately from the following straightforward variant of Proposition 5.8.

Proposition 5.17. *Assume that $\gamma, \tilde{\gamma} : S \to \mathbb{R}$ are measurable, the optimization problem* (5.20) *is well posed, and that $\xi \in \mathsf{RST}(\mu)$ is an optimizer. Then $(r \otimes \mathrm{Id})(\pi)(\mathsf{SG}_2^{\xi}) = 0$ for any $\pi \in \mathsf{JOIN}^1(r(\xi))$.*

Proof. As $\xi \in \mathsf{Opt}_\gamma$ we have to show that $(r \otimes \mathrm{Id})(\pi)(\mathsf{SG}_2^{\xi} \setminus \mathsf{SG}^{\xi}) = 0$, however this follows by considering the same construction as in the proof of Proposition 5.8. □

6. Embeddings in abundance

In the following we suppose that $(\Omega, \mathcal{G}, (\mathcal{G}_t)_{t\geq 0}, \mathbb{P})$ is a stochastic basis which is sufficiently rich to support a Brownian motion B and a uniformly distributed $\mathcal{G}_0$-random variable. We suppose that $\gamma : S \to \mathbb{R}$ is a Borel measurable function. In a slight abuse of notation we will also write $(\gamma_t)_{t\in\mathbb{R}_+}$ for the process given by

$$t \mapsto \gamma((B_s)_{s\leq t}, t).$$

In the previous section we have considered a secondary optimization problem and a version of the monotonicity principle (Theorem 5.16) accounting for this extension. We now give a brief summary in probabilistic terms.

Write Opt_γ for the set of $\mathcal{G}$-stopping times on Ω which are optimizers of (OptSEP) and consider another Borel function $\tilde{\gamma} : S \to \mathbb{R}$. We call $\hat{\tau} \in \mathsf{Opt}_\gamma$ a secondary minimizer if it solves

$$P_{\tilde{\gamma}|\gamma} = \inf\{\mathbb{E}\,[\tilde{\gamma}_\tau] : \tau \in \mathsf{Opt}_\gamma\}. \tag{OptSEP$_2$}$$

As in (5.20) we say that (OptSEP$_2$) is well posed if the primary optimization problem (OptSEP) is well posed and $\mathbb{E}\,[\tilde{\gamma}_\tau]$ exists with values in $(-\infty, \infty]$ for all $\tau \in \mathsf{Opt}_\gamma$ and is finite for one such τ. Then we have the following version of Theorems 1.1 and 4.1:

Theorem 6.1. *Let $\gamma, \tilde{\gamma} : S \to \mathbb{R}$ be lsc and bounded from below in the sense of* (4.2)*. Then* (OptSEP$_2$) *admits a minimizer $\hat{\tau}$.*

We now provide the appropriate generalizations of Definitions 1.4 and 1.5 and Theorem 1.3 for this case.

Definition 6.2. *The pair $((f, s), (g, t)) \in S \times S$ constitutes a* secondary stop-go pair*, written $((f, s), (g, t)) \in \mathsf{SG}_2$, iff $f(s) = g(t)$, and for* every *$(\mathcal{F}_t^B)_{t\geq 0}$-stopping time σ which satisfies $0 < \mathbb{E}[\sigma] < \infty$,*

$$\mathbb{E}[(\gamma^{(f,s)\oplus})_\sigma] \;+\; \gamma(g, t) \quad \geq \quad \gamma(f, s) \;+\; \mathbb{E}[(\gamma^{(g,t)\oplus})_\sigma], \tag{6.1}$$

whenever both sides are well defined, and the left-hand side is finite; and if

$$\mathbb{E}[(\gamma^{(f,s)\oplus})_\sigma] \;+\; \gamma(g, t) \quad = \quad \gamma(f, s) \;+\; \mathbb{E}[(\gamma^{(g,t)\oplus})_\sigma] \tag{6.2}$$

then

$$\mathbb{E}[(\tilde{\gamma}^{(f,s)\oplus})_\sigma] \;+\; \tilde{\gamma}(g, t) \quad > \quad \tilde{\gamma}(f, s) \;+\; \mathbb{E}[(\tilde{\gamma}^{(g,t)\oplus})_\sigma], \tag{6.3}$$

whenever both sides are well-defined and the left-hand side (of (6.3)*) is finite.*

Definition 6.3. *We say that Γ is $\tilde{\gamma}|\gamma$-monotone if*

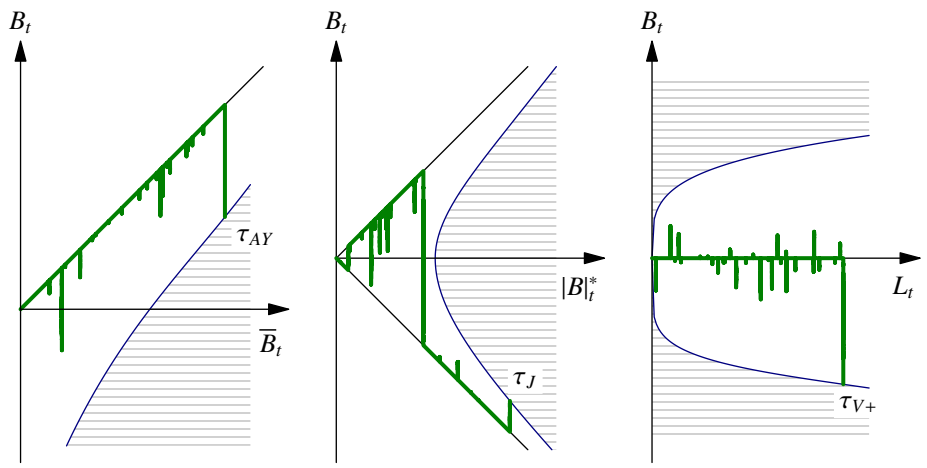
$$\mathsf{SG}_2 \cap (\Gamma^< \times \Gamma) = \emptyset. \tag{6.4}$$

From Theorem 5.16 together with a trivial modification of Lemma 5.4 we then obtain:

Theorem 6.4 (Monotonicity Principle II)**.** *Let $\gamma, \tilde{\gamma} : S \to \mathbb{R}$ be Borel measurable, suppose that* (OptSEP$_2$) *is well posed and that $\hat{\tau}$ is an optimizer. Then there exists a $\tilde{\gamma}|\gamma$-*monotone *Borel set $\Gamma \subseteq S$ such that $\mathbb{P}$-a.s.*

$$((B_t)_{t\leq\hat{\tau}}, \hat{\tau}) \in \Gamma\,. \tag{6.5}$$

6.1. **Recovering classical embeddings.** In this section we derive a number of classical embeddings as well as establish new embeddings. Figure 4 shows graphical representations of some of these constructions. We highlight the common feature of all these pictures: when plotted in an appropriate phase space, the stopping time is the hitting time of a barrier type set. Identifying the appropriate phase space, and determining the exact structure of the barrier will be the key step in deriving the solutions to (SEP) in this section.



(A) The Azéma-Yor construction. (B) The Jacka construction (C) The Vallois construction

FIGURE 4. Representations of the Azéma-Yor, Vallois and Jacka constructions.

For subsequent use, it will be helpful to write, for $(f, s) \in S$, $\bar{f} = \sup_{r\leq s} f(r)$, $\underline{f} = \inf_{r\leq s} f(r)$ and $|f|^* = \sup_{r\leq s} |f(r)|$.

Theorem 6.5 (The Azéma-Yor embedding, cf. [4])**.** *There exists a stopping time τ_{AY} which* maximizes

$$\mathbb{E}\Big[\sup_{t\leq\tau} B_t\Big]$$

over all solutions to (SEP) *and which is of the form $\tau_{AY} = \inf\{t > 0 : B_t \leq \psi(\sup_{s\leq t} B_s)\}$ a.s., for some increasing function ψ.*

Proof. Fix a bounded and strictly increasing continuous function $\varphi : \mathbb{R}_+ \to \mathbb{R}$ and consider the continuous functions $\gamma((f, s)) = -\bar{f}$ and $\tilde{\gamma}((f, s)) = \varphi(\bar{f})(f(s))^2$. Then (OptSEP$_2$) is well posed and by Theorem 6.1 there exists a minimizer τ_{AY}. By Theorem 6.4, pick a $\tilde{\gamma}|\gamma$-monotone set $\Gamma \subseteq S$ supporting τ_{AY}. We claim that

$$\mathsf{SG}_2 \supseteq \{((f, s), (g, t)) \in S \times S : g(t) = f(s), \bar{g} < \bar{f}\}. \tag{6.6}$$

This is represented graphically in Figure 5.

Indeed, pick $((f, s), (g, t)) \in S \times S$ with $f(s) = g(t)$ and $\bar{g} < \bar{f}$ and a stopping time σ with positive and finite expectation. Then (6.1) amounts to

$$\mathbb{E}[\bar{f} \vee (f(s) + \bar{B}_\sigma)] + \bar{g} \leq \bar{f} + \mathbb{E}[\bar{g} \vee (g(t) + \bar{B}_\sigma)]$$

with a strict inequality unless $\bar{g} \geq g(t) + \bar{B}_\sigma$ a.s. However in that case (6.2) is trivially satisfied and (6.3) amounts to

$$\mathbb{E}\Big[\varphi(\bar{f})(f(s) + B_\sigma)^2\Big] + \varphi(\bar{g})g(t)^2 > \varphi(\bar{f})f(s)^2 + \mathbb{E}\Big[\varphi(\bar{g})(g(t) + B_\sigma)^2\Big]$$

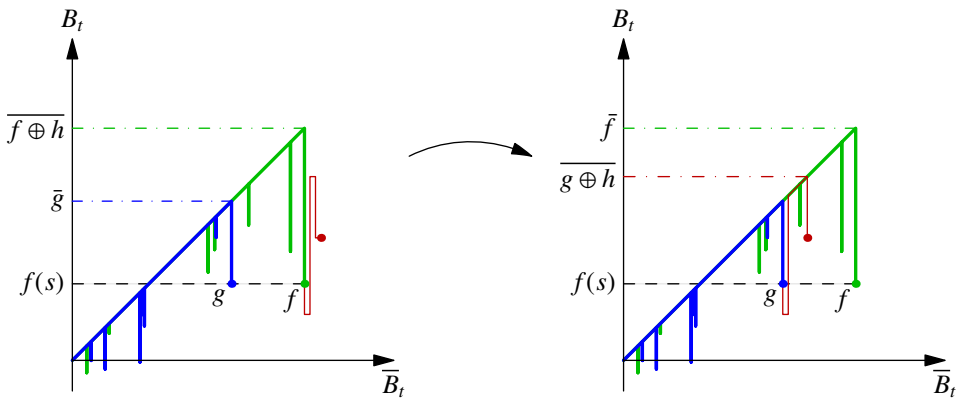


FIGURE 5. The stop-go pairs for the Azéma-Yor embedding. On the left, the blue path (g, t) is stopped, and the green path, (f, s), is allowed to continue; a possible continuation, h, being shown in red. On the right hand side we see the effect of allowing g to go and stopping f: the maximum of g is increased, but the maximum of f stays the same.

which holds since $g(t) = f(s)$. Summing up, $((f, s), (g, t)) \in \mathsf{SG} \subseteq \mathsf{SG}_2$ in the former case and $((f, s), (g, t)) \in \mathsf{SG}_2$ in the latter case, proving (6.6).

In complete analogy with the derivation of the Root embedding (Theorem 2.1) we define

$$\mathcal{R}_{\text{CL}} := \{(m, x) : \exists (g, t) \in \Gamma, \bar{g} \le m, g(t) = x\}, \ \mathcal{R}_{\text{OP}} := \{(m, x) : \exists (g, t) \in \Gamma, \bar{g} < m, g(t) = x\},$$

and write $\tau_{\text{CL}}, \tau_{\text{OP}}$ for the first times the process $(\bar{B}_t(\omega), B_t(\omega))$ hits the sets $\mathcal{R}_{\text{CL}}$ and $\mathcal{R}_{\text{OP}}$ respectively. Then we claim $\tau_{\text{CL}} \le \tau_{AY} \le \tau_{\text{OP}}$ a.s. Note that $\tau_{\text{CL}} \le \tau_{AY}$ holds by definition of τ_{CL}. To show $\tau_{AY} \le \tau_{\text{OP}}$, consider ω satisfying $((B_s(\omega))_{s \le \tau_{AY}(\omega)}, \tau_{AY}(\omega)) \in \Gamma$ and assume for contradiction that $\tau_{\text{OP}}(\omega) < \tau_{AY}(\omega)$. Then there exists $s \in [\tau_{\text{OP}}(\omega), \tau_{AY}(\omega))$ such that $f := (B_r(\omega))_{r \le s}$ satisfies $(\bar{f}, f(s)) \in \mathcal{R}_{\text{OP}}$. Since $s < \tau_{AY}(\omega)$ we have $(f, s) \in \Gamma^{<}$. By definition of $\mathcal{R}_{\text{OP}}$, there exists $(g, t) \in \Gamma$ such that $f(s) = g(t)$ and $\bar{g} < \bar{f}$, yielding a contradiction.

Finally, we define

$$\psi_0(m) = \sup\{x : (m, x) \in \mathcal{R}_{\text{CL}}\}.$$

It follows from the definition of $\mathcal{R}_{\text{CL}}$ that $\psi_0(m)$ is increasing, and we define the right-continuous function $\psi_+(m) = \psi_0(m+)$, and the left-continuous function $\psi_-(m) = \psi_0(m-)$. It follows from the definitions of τ_{OP} and τ_{CL} that:

$$\tau_+ := \inf\{t \ge 0 : B_t \le \psi_+(\bar{B}_t)\} \le \tau_{\text{CL}} \le \tau_{\text{OP}} \le \inf\{t \ge 0 : B_t < \psi_-(\bar{B}_t)\} =: \tau_-.$$

It is then easily checked that $\tau_- = \tau_+$ a.s., and the result follows on taking $\psi = \psi_+$. □

Theorem 6.6 (The Jacka Embedding, cf. [33]). *Let $\varphi : \mathbb{R}_+ \to \mathbb{R}$ be a bounded, strictly increasing right-continuous function. There exists a stopping time τ_J which* maximizes

$$\mathbb{E}\Big[\varphi\Big(\sup_{t \le \tau} |B_t|\Big)\Big]$$

over all solutions to (SEP), *and which is of the form*

$$\tau_J = \inf\Big\{t > 0 : B_t \ge \alpha_-\Big(\sup_{s \le t} |B_s|\Big) \text{ and } B_t \le \alpha_+\Big(\sup_{s \le t} |B_s|\Big)\Big\}$$

a.s., for some functions α_+, α_-, where α_+ is increasing, α_- is decreasing, and $\alpha_+(y) \ge \alpha_-(y)$ for all $y > y_0$, $\alpha_-(y) = -\alpha_+(y) = \infty$ for $y < y_0$, some $y_0 \ge 0$.

Proof. The proof runs along similar lines to the proof of Theorem 6.5, when we take $\gamma((f, s)) = -\varphi(|f|^*)$ and set $\tilde{\gamma}((f, s)) = \tilde{\varphi}(|f|^*)(f(s))^2$ for some bounded and strictly increasing, continuous function $\tilde{\varphi}$. Then the statement follows once we see

$$\mathsf{SG}_2 \supseteq \{((f, s), (g, t)) \in S \times S : f(s) = g(t), |f|^* > |g|^*\},$$

define

$$\mathcal{R}_{\text{CL}} := \{(m,x) : \exists (g,t) \in \Gamma, |g|^* \le m, g(t) = x\}$$
$$\mathcal{R}_{\text{OP}} := \{(m,x) : \exists (g,t) \in \Gamma, |g|^* < m, g(t) = x\},$$

and then take $\alpha_-(m) = \inf\{x : (m,x) \in \mathcal{R}_{\text{CL}}\}$ and $\alpha_+(m) = \sup\{x : (m,x) \in \mathcal{R}_{\text{CL}}\}$. □

Remark 6.7. We observe that both the results hold for one-dimensional Brownian motion with an arbitrary starting distribution λ satisfying the usual convex ordering condition.

Theorem 6.8 (The Perkins Embedding, cf. [45])**.** *Suppose $\mu(\{0\}) = 0$. Let $\varphi : \mathbb{R}_+^2 \to \mathbb{R}$ be a bounded function which is continuous and strictly increasing in both arguments. There exists a stopping time τ_P which* minimizes

$$\mathbb{E}\Big[\varphi\Big(\sup_{t\le\tau} B_t, -\inf_{t\le\tau} B_t\Big)\Big]$$

over all solutions to (SEP) *and which is of the form $\tau_P = \inf\{t > 0 : B_t \notin (\alpha_+(\bar{B}_t), \alpha_-(\underline{B}_t))\}$, for some decreasing functions α_+ and α_- which are left- and right-continuous respectively.*

Proof. Fix a bounded and strictly increasing continuous function $\tilde{\varphi} : \mathbb{R}_+^2 \to \mathbb{R}$ and consider the continuous functions $\gamma((f,s)) = \varphi(\bar{f}, -\underline{f})$ and $\tilde{\gamma}((f,s)) = -(f(s))^2\tilde{\varphi}(\bar{f}, -\underline{f})$. Then (OptSEP$_2$) is well posed and by Theorem 6.1 there exists a minimizer τ_P. By Theorem 6.4, pick a $\tilde{\gamma}|\gamma$-monotone set $\Gamma \subseteq S$ supporting τ_P. Note that we may assume that Γ only contains points such that $\underline{g} < 0 < \bar{g}$, since $\mu(\{0\}) = 0$.

By a similar argument to that given in the proof of Theorem 6.5 we can show

$$\mathsf{SG}_2 \supseteq \{((f,s),(g,t)) \in S \times S : f(s) = g(t), (\bar{f}, -\underline{f}) < (\bar{g}, -\underline{g})\},$$

where $(\bar{f}, -\underline{f}) < (\bar{g}, -\underline{g})$ iff $(\bar{f}, -\underline{f}) \le (\bar{g}, -\underline{g})$ but not $(\bar{f}, -\underline{f}) = (\bar{g}, -\underline{g})$ and$(\bar{f}, -\underline{f}) \le (\bar{g}, -\underline{g})$ refers to the partial order of $\mathbb{R}^2$.

In addition, consider a path $(g,t) \in S$ such that $\underline{g} < g(t) < \bar{g}$. Then there exists $(f,s) \in S$ such that $f(r) = g(r)$ for $r \le s$, and such that $f(s) = g(t)$, and exactly one of $\bar{f} = \bar{g}$, or $\underline{f} = \underline{g}$. This is true since there must exist a last time that $g(r) = x$ before setting the most recent extremum. In particular, $((f,s),(g,t)) \in \mathsf{SG}_2$. It follows that $\Gamma \cap \{(g,t) : \underline{g} < g(t) < \bar{g}\} = \emptyset$, that is, any stopped path must stop at a minimum or a maximum.

Now consider the sets:

$$\begin{aligned}\mathcal{R}_{\text{CL}} &= \Big\{(m,x) : \exists (g,t) \in \Gamma, g(t) = x = \underline{g}, \bar{g} \ge m\Big\} \cup \Big\{(x,i) : \exists (g,t) \in \Gamma, g(t) = x = \bar{g}, \underline{g} \le i\Big\} \\ &= \underline{\mathcal{R}}_{\text{CL}} \cup \bar{\mathcal{R}}_{\text{CL}} \\ \mathcal{R}_{\text{OP}} &= \Big\{(m,x) : \exists (g,t) \in \Gamma, g(t) = x = \underline{g}, \bar{g} > m\Big\} \cup \Big\{(x,i) : \exists (g,t) \in \Gamma, g(t) = x = \bar{g}, \underline{g} < i\Big\} \\ &= \underline{\mathcal{R}}_{\text{OP}} \cup \bar{\mathcal{R}}_{\text{OP}},\end{aligned}$$

and their respective hitting times by $(\bar{B}_t, \underline{B}_t)_{t\ge0}$, denoted $\tau_{\text{CL}}, \tau_{\text{OP}}$. Since $\Gamma \cap \{(g,t) : \underline{g} < g(t) < \bar{g}\} = \emptyset$, it follows that $\tau_{\text{CL}} \le \tau_P$ a.s. In addition, an essentially identical argument to that used in the proof of Theorem 6.5 gives $\tau_P \le \tau_{\text{OP}}$ a.s.

We now set $\alpha_+(m) = \sup\{x < 0 : (m,x) \in \underline{\mathcal{R}}_{\text{CL}}\}$, $\alpha_-(i) = \inf\{x > 0 : (x,i) \in \bar{\mathcal{R}}_{\text{CL}}\}$. Then these functions are both clearly decreasing and left- and right-continuous respectively, by definition of the respective sets $\underline{\mathcal{R}}_{\text{CL}}, \bar{\mathcal{R}}_{\text{CL}}$. Moreover, it is immediate that

$$\tau_{\text{CL}} = \inf\Big\{t > 0 : B_t \notin \Big(\alpha_+(\bar{B}_t), \alpha_-(\underline{B}_t)\Big)\Big\},$$

and we deduce that $\tau_{\text{CL}} = \tau_{\text{OP}}$ a.s. by standard properties of Brownian motion. The conclusion follows. □

Theorem 6.9 (Maximizing the range)**.** *Let $\varphi : \mathbb{R}_+^2 \to \mathbb{R}$ be a bounded function which is continuous and strictly increasing in both arguments. There exists a stopping time τ_{xr} which* maximizes

$$\mathbb{E}\Big[\varphi\Big(\sup_{t\le\tau} B_t, -\inf_{t\le\tau} B_t\Big)\Big]$$

over all solutions to (SEP)*, and which is of the form* $\tau_{xr} = \inf\{t > 0 : B_t \geq \alpha_-(\bar{B}_t, -\underline{B}_t)$ *or* $B_t \leq \alpha_+(\bar{B}_t, -\underline{B}_t)\}$ *for some right-continuous functions* $\alpha_-(m, i)$ *decreasing in both coordinates and* $\alpha_+(m, i)$ *increasing in both coordinates.*

Proof. Our primary objective will be to minimize $\gamma((f, s)) = -\varphi(\bar{f}, -\underline{f})$, which is a lsc function on S. We again introduce a secondary minimization problem: specifically, we consider the function $\tilde{\gamma}((f, s)) = (f(s))^2\tilde{\varphi}(\bar{f}, -\underline{f})$ for some bounded, continuous and strictly increasing function $\tilde{\varphi} : \mathbb{R}^2_+ \to \mathbb{R}$. Then (OptSEP$_2$) is well posed and by Theorem 6.1 there exists a minimizer τ_{xr}. By Theorem 6.4, pick a $\tilde{\gamma}|\gamma$-monotone set $\Gamma \subseteq S$ supporting τ_{xr}.

By a similar argument to that given in the proof of Theorem 6.5 we can show $\mathsf{SG}_2 \supseteq \{((f, s), (g, t)) \in S \times S : f(s) = g(t), (\bar{f}, -\underline{f}) > (\bar{g}, -\underline{g})\}$.

Let conv denote the convex hull, and write

$$I_{\text{CL}}(\bar{b}, -\underline{b}) := \text{conv}\left\{x : \exists (g, t) \in \Gamma, g(t) = x, (\bar{g}, -\underline{g}) \leq (\bar{b}, -\underline{b})\right\},$$
$$I_{\text{OP}}(\bar{b}, -\underline{b}) := \text{conv}\left\{x : \exists (g, t) \in \Gamma, g(t) = x, (\bar{g}, -\underline{g}) < (\bar{b}, -\underline{b})\right\}.$$

Then $I_{\text{CL}}, I_{\text{OP}}$ are both increasing in both coordinates, and $I_{\text{CL}} \supseteq I_{\text{OP}}$. Write $\tau_{\text{OP}} := \inf\{t \geq 0 : B_t \in I_{\text{OP}}(\bar{B}_t, -\underline{B}_t)\}$, and $\tau_{\text{CL}} := \inf\{t \geq 0 : B_t \in I_{\text{CL}}(\bar{B}_t, -\underline{B}_t)\}$. As previously, we deduce that $\tau_{\text{CL}} \leq \tau_{xr} \leq \tau_{\text{OP}}$. If, in addition, we define

$$\alpha_+(m, i) := \sup I_{\text{OP}}(m, i) \qquad \alpha_-(m, i) := \inf I_{\text{OP}}(m, i)$$
$$\alpha_{+,\text{CL}}(m, i) := \sup I_{\text{CL}}(m, i) \qquad \alpha_{-,\text{CL}}(m, i) := \inf I_{\text{CL}}(m, i)$$

then α_+, α_- satisfy the conditions of the theorem, and

$$\tau_{\text{OP}} = \inf\left\{t \geq 0 : B_t \geq \alpha_-(\bar{B}_t, -\underline{B}_t) \text{ or } B_t \leq \alpha_+(\bar{B}_t, -\underline{B}_t)\right\}$$
$$\tau_{\text{CL}} = \inf\left\{t \geq 0 : B_t \geq \alpha_{-,\text{CL}}(\bar{B}_t, -\underline{B}_t) \text{ or } B_t \leq \alpha_{+,\text{CL}}(\bar{B}_t, -\underline{B}_t)\right\}.$$

To conclude, we need to show that $\tau_{\text{OP}} = \tau_{\text{CL}}$. However, we first observe that $\tau_{\text{OP}} \geq \sigma$, and $\tau_{\text{CL}} \geq \sigma_{\text{CL}}$, where

$$\sigma := \inf\left\{t \geq 0 : \alpha_-(\bar{B}_t, -\underline{B}_t) < \infty \text{ or } \alpha_+(\bar{B}_t, -\underline{B}_t) > -\infty\right\}$$
$$\sigma_{\text{CL}} := \inf\left\{t \geq 0 : \alpha_{-,\text{CL}}(\bar{B}_t, -\underline{B}_t) < \infty \text{ or } \alpha_{+,\text{CL}}(\bar{B}_t, -\underline{B}_t) > -\infty\right\},$$

and in fact, $\sigma = \sigma_{\text{CL}}$ a.s. In addition, on $\{\sigma > 0\}$ we have $B_\sigma \in \{\bar{B}_\sigma, \underline{B}_\sigma\}$. On the set $\{B_\sigma = \bar{B}_\sigma\}$ say, then

$$\tau_{\text{OP}} = \inf\{t \geq \sigma : B_t \leq \alpha_+(\bar{B}_t, -\underline{B}_\sigma)\} = \inf\{t \geq \sigma : B_t \leq \alpha_{+,\text{CL}}(\bar{B}_t, -\underline{B}_\sigma)\} \quad a.s.$$

by the same argument as used at the end of the proof of Theorem 6.5, and the fact that $\alpha_+(m+, i) = \alpha_{+,\text{CL}}(m, i)$, by the definition of the sets $I_{\text{CL}}, I_{\text{OP}}$. □

Remark 6.10. We observe that, in the case of Theorem 6.9, the characterization provided would not appear to be sufficient to identify the functions α_+, α_- given the measure μ. This is in contrast to the constructions of Azéma-Yor, Perkins and Jacka, where knowledge of the form of the embedding is sufficient to identify the corresponding stopping rule.

On a more abstract level, uniqueness of barrier type embeddings in a two dimensional phase space can be seen as a consequence of Loynes' argument [39]. More precisely, let A_t be some continuous process and suppose that τ_1 and τ_2 denote the times when (A_t, B_t) hits a closed barrier type set R_1 resp. R_2. If $\mathbb{E}[\tau_1], \mathbb{E}[\tau_2] < \infty$ and both stopping times embed the same measure, the argument presented in Remark 2.3 shows that $\tau_1 = \tau_2$.

Remark 6.11. In Cox and Obłój [12], embeddings are constructed which maximize certain double-exit probabilities: for example, to maximize the probability that both $\bar{B}_\tau \geq \bar{b}$ and $\underline{B}_\tau \leq \underline{b}$, for given levels $\bar{b}$ and $\underline{b}$. In this case, the embedding is no longer naturally viewed as a barrier type construction; instead, it is natural to characterize the embedding in terms of where the paths with different crossing behaviour for the barriers finish (for example, the paths which only hit the upper level may end up above a certain value, or between two other values). However, it is possible, again using a suitable secondary maximization problem, to show that there exists an optimizer demonstrating the behaviour characterizing

the Cox-Obłój embeddings. (Specifically, if we write $H_b((f,s)) = \inf\{t \le s : f(t) = b\}$, $\underline{H} = H_{\underline{b}} \wedge H_{\bar{b}}$ and $\bar{H} = H_{\underline{b}} \vee H_{\bar{b}}$ then the secondary maximization problem

$$\tilde{\gamma}((f,s)) = 1/2((f(s) - \underline{H}((f,s)))^2 \mathbb{1}_{\underline{H} \le s} - ((f(s) - \bar{H}((f,s)))^2 \mathbb{1}_{\bar{H} \le s}$$

is sufficient to rederive the form of these embeddings.)

6.2. **The Vallois-embedding and optimizing functions of local time.** In this section we shall determine the stopping rule which solves

$$\inf\{\mathbb{E}[h(\mathfrak{L}_\tau)] : \tau \text{ solves (SEP)}\}, \tag{6.7}$$

where $\mathfrak{L}$ denotes the local time of Brownian motion at 0 and h is a convex or concave function. In many ways, the proof of this result will follow the arguments used in the previous section, however in contrast to the functions considered there, $h(\mathfrak{L})$ is not defined on S in a straightforward way and hence we need to apply some care in fixing our notions. Moreover local time does not have an S-continuous modification and hence some additional argument is needed to establish that (6.7) admits a minimizer.

We say that a $\mathcal{G}$-adapted process $\mathfrak{L}^x$ is a *local time in x* if it is a (right-continuous, increasing) compensator of $|B - x|$ and we suppress x in the case of local time at 0. This determines $\mathfrak{L}^x$ up to indistinguishability (and clearly the choice of $\mathfrak{L}^x$ is irrelevant for (6.7)).

For us it is convenient to allow local time to assume the value $+\infty$ on an evanescent set. Using this convention, Theorem 4.1 implies that there exists a Borel function $L^x : S \to [0,\infty]$ such that $L^x \circ r$ is a (right-continuous, increasing) $\mathcal{F}^0$-predictable local time on Wiener space. We will call such a process L^x a *raw local time* in x. We note that the value $+\infty$ cannot be avoided here, see [42].

Lemma 6.12. *Let L be a raw local time in* 0*. Then there exists a Borel set $A \subseteq C_0(\mathbb{R}_+)$, $\mathbb{W}(A) = 1$ such that for all*

$$(f,s) \in U = \{(f,s) \in S : \exists \omega \in A, f = (\omega_r)_{r \le s}\}$$

we have $L(f,s) < \infty$ and

$$(g,t) \mapsto L^{(f,s)}(g,t) := L(f \oplus g, s+t) - L(f,s) \tag{6.8}$$

is a raw local time in $-f(s)$.

Proof. Write V for the set of all (f,s) such that $L^{(f,s)}$ is not a raw local time. To understand whether $(f,s) \in V$ we need to check whether or not $(\omega,t) \mapsto |B_{s+t}(f \oplus \omega)| - L^{(f,s)}(r(\omega,t))$ defines a martingale. Since this is a Borel property, $V \subseteq S$ is Borel. Hence

$$\mathrm{deb}(V) := \{\omega : \exists t, r(\omega,t) \in V\}$$

is analytic and thus universally measurable. To prove that $\mathbb{W}(\mathrm{deb}(V)) = 0$ it is sufficient to show this for any given Borel subset of $\mathrm{deb}(V)$. Suppose for contradiction that $\mathbb{W}(E) > 0$ for some Borel set $E \subseteq \mathrm{deb}(V)$. By the optional section theorem this implies that there exists an $\mathcal{F}^a$-stopping time τ such that $\mathbb{W}(\tau < \infty) > 0$ and $(\omega, \tau(\omega)) \in r^{-1}(V)$ whenever $\tau(\omega) < \infty$. Upon requiring this only a.s. we may of course assume that τ is an $\mathcal{F}^0$-stopping time.

Given $H = G\mathbb{1}_{[\![\tau,\infty[\![}$ for bounded $\mathcal{F}^0_\tau$-measurable G it follows from usual properties of local time that

$$t \mapsto (H \cdot (|B| - L \circ r))_t = G\big[(|B| - L \circ r)_t - (|B| - L \circ r)_{\tau \wedge t}\big]$$

is a martingale. As G was arbitrary,

$$(\omega,t) \mapsto |B_{\tau(\omega')+t}(\omega'_{\restriction[0,\tau(\omega')]} \oplus \omega)| - L^{(\omega'_{\restriction[0,\tau(\omega')]},\tau(\omega'))}(r(\omega,t))$$

defines a martingale for almost all ω', $\tau(\omega') < \infty$, contradicting $\mathbb{W}(\mathrm{deb}(V)) > 0$.

It follows that $\mathbb{W}(\mathrm{deb}(V)) = 0$, hence we may pick a Borel set $A \subseteq \mathrm{deb}(V)^c$ with $\mathbb{W}(A) = 1$ such that (6.8) holds. □

Our next goal is to verify that (6.7) admits an optimizer.

Proposition 6.13. *Let $h : [0,\infty) \to \mathbb{R}$ be continuous and bounded. Then there exists an optimizer for* (6.7)*. Moreover, if $\tilde{\gamma}(f,s) = e^{-L(f,s)} f^2(s)$ or $\tilde{\gamma}(f,s) = -e^{-L(f,s)} f^2(s)$, the secondary minimization problem* (OptSEP_2) *also admits a solution.*

Proof. Let L be a raw local time. We first observe that $(\mathfrak{L}_t)_{t\geq 0} := (L \circ r((B_t)_{t\geq 0}, t))_{t\geq 0}$ is (indistinguishable from) the local time of $(B_t)_{t\geq 0}$ on $(\Omega, \mathcal{G}, (\mathcal{G}_t)_{t\geq 0}, \mathbb{P})$. By Lemma 3.11 there exists a sequence $\xi_1, \xi_2, \xi_3, \ldots \in \mathsf{RST}(\mu)$ such that

$$V^* = \lim \int h(\mathfrak{L}_t(\omega))\, \xi_n(d\omega, dt) = \inf\{\mathbb{E}[h(\mathfrak{L}_\tau)] : \tau \text{ solves (SEP)}\}.$$

Possibly passing to a subsequence there is $\xi \in \mathsf{RST}(\mu)$ such that $\xi = \lim_n \xi_n$.

A result of Jacod and Memin ([34, Corollary 2.9]) asserts

$$\int \varphi \, d\xi_n \to \int \varphi \, d\xi$$

for any bounded measurable function $\varphi : C_0(\mathbb{R}_+) \times \mathbb{R}_+$ such that $t \mapsto \varphi(\omega, t)$ is continuous for every $\omega \in C_0(\mathbb{R}_+)$. It follows that $\int h(\mathfrak{L}_t(\omega))\, \xi(d\omega, dt) = V^*$. Moreover (again by Lemma 3.11) there exists a $\mathcal{G}$-stopping time τ^* such that $\mathbb{E}[h(\mathfrak{L}_{\tau^*})] = \int h(\mathfrak{L}_t(\omega))\, \xi(d\omega, dt)$.

The second assertion follows from a similar reasoning, using an approximation argument to handle the unboundedness of $\tilde{\gamma}$. □

Note added in revision. Guo, Tan, and Touzi [27] were able to relax the continuity assumption in our existence and duality results Theorems 1.1 and 1.2. Based on the work of Jacod and Memin [34] they establish these results under the assumption that $t \mapsto \gamma \circ r(\omega, t)$ is lsc for every $\omega \in C_0(\mathbb{R}_+)$. In particular their results would imply a more general version of Proposition 6.13.

We are now able to show:

Theorem 6.14. *Let $h : [0,\infty] \to \mathbb{R}$ be a bounded, strictly concave function and $\mathfrak{L}$ the local time of B at* 0*.*

(1) There exists a stopping time τ_{V-} which maximizes

$$\mathbb{E}\left[h\left(\mathfrak{L}_\tau\right)\right]$$

over the set of all solutions to (SEP)*, and which is of the form*

$$\tau_{V-} = \inf\left\{t > 0 : B_t \notin \left(\alpha_-\left(\mathfrak{L}_t\right), \alpha_+\left(\mathfrak{L}_t\right)\right)\right\} \text{ a.s.,}$$

for some decreasing function $\alpha_+ \geq 0$ and increasing function $\alpha_- \leq 0$.

(2) There exists a stopping time τ_{V+} which minimizes

$$\mathbb{E}\left[h\left(\mathfrak{L}_\tau\right)\right]$$

over the set of all solutions to (SEP)*, and which is of the form*

$$\tau_{V+} = Z \wedge \inf\left\{t > 0 : B_t \notin \left(\alpha_-\left(\mathfrak{L}_t\right), \alpha_+\left(\mathfrak{L}_t\right)\right)\right\}, \text{ a.s.}$$

for some increasing function $\alpha_+ \geq 0$, and some decreasing function $\alpha_- \leq 0$, and a $\{0,\infty\}$-valued $\mathcal{G}_0$-measurable random variable Z.

Proof. We consider the second case, under the additional assumption that $0 < \mu(\{0\}) < 1$, the other cases being slightly simpler. As above, we let L be a raw local time and observe that $(\mathfrak{L}_t)_{t\geq 0} := (L \circ r((B_t)_{t\geq 0}, t))_{t\geq 0}$ is (indistinguishable from) the local time of $(B_t)_{t\geq 0}$ on $(\Omega, \mathcal{G}, (\mathcal{G}_t)_{t\geq 0}, \mathbb{P})$.

Applying Proposition 6.13 and Theorem 6.4 to the optimizations corresponding to $\gamma(\omega, t) = h(L \circ r(\omega, t))$ and $\tilde{\gamma}(\omega, t) = e^{-L\circ r(\omega,t)} \omega_t^2$ we obtain a minimizer τ_{V+} and a $\tilde{\gamma}|\gamma$-monotone set $\Gamma \subseteq S$ supporting τ_{V+}.

Recall the set $A \subseteq C_0(\mathbb{R}_+)$ given by Lemma 6.12. By projection the set

$$U = \{(f,s) \in S : \exists \omega \in A, f = (\omega_r)_{r\leq s}\}$$

is universally measurable and since τ_{V+} is a finite stopping time, $\mathbb{P}(((B_t)_{t\leq \tau_{V+}}, \tau_{V+}) \in U) = 1$. Passing to an appropriate subset if necessary, we may also assume that U is Borel. We may therefore assume $\Gamma \subseteq U$, and it then also follows that $\Gamma^< \subseteq U$.

By a similar argument to the previous cases we can show that

$$\mathsf{SG}_2 \supseteq \{((f,s),(g,t)) \in U \times U : f(s) = g(t), L(f,s) < L(g,t)\}, \tag{6.9}$$

where Lemma 6.12 guarantees that local time of paths is well-behaved following a path-swapping operation. In particular, since both f and g belong to U, it follows that (6.8) holds, and (6.9) is a direct consequence of this.

Define the sets

$$\mathcal{R}_{\text{OP}} := \{(l,x) : \exists (g,t) \in \Gamma, g(t) = x, L(g,t) > l\},$$
$$\mathcal{R}_{\text{CL}} := \{(l,x) : \exists (g,t) \in \Gamma, g(t) = x, L(g,t) \geq l\},$$

and the corresponding stopping times

$$\tau^*_{\text{OP}} := \inf\{t \geq 0 : \mathfrak{L}_t > 0, (\mathfrak{L}_t, B_t) \in \mathcal{R}_{\text{OP}}\}, \ \tau^*_{\text{CL}} := \inf\{t \geq 0 : \mathfrak{L}_t > 0, (\mathfrak{L}_t, B_t) \in \mathcal{R}_{\text{CL}}\}.$$

Strictly speaking, the random times on the right-hand side only define stopping times in the augmented filtration (by the Début Theorem), however by Theorem 3.1, this is sufficient to find almost surely equal $\mathcal{G}$-stopping times.

Since $(\Gamma^< \times \Gamma) \cap \mathsf{SG}_2 = \emptyset$ and $(0,0) \in \Gamma^<$ (Γ contains a non-trivial element since $\mu(\{0\}) < 1$) then $(l,0) \notin \Gamma$ for any $l \geq 0$. It follows that $\mathbb{P}(\tau_{V+} = 0) = \mu(\{0\})$.

We now consider τ_{V+} on $\{\tau_{V+} > 0\}$. Note that $\{\tau > 0\} = \{\mathfrak{L}_\tau > 0\}$ a.s., for any stopping time τ and hence in particular $\{\tau_{V+} > 0\} = \{\mathfrak{L}_{\tau_{V+}} > 0\}$ a.s. Then on $\{\tau_{V+} > 0\}$, $\tau^*_{\text{CL}} \leq \tau_{V+} \leq \tau^*_{\text{OP}}$ a.s., and hence $\mathbb{P}(\tau_{V+} \leq \tau^*_{\text{OP}}) = 1$. Define $\alpha_+(l) = \inf\{x > 0 : (l,x) \in \mathcal{R}_{\text{OP}}\}$ and $\alpha_-(l) = \sup\{x < 0 : (l,x) \in \mathcal{R}_{\text{OP}}\}$.

If either of $\alpha_-(\eta) = 0$ or $\alpha_+(\eta) = 0$ for some $\eta > 0$, then $\tau^*_{\text{OP}} = 0$ a.s. Since $\tau_{V+} \leq \tau^*_{\text{OP}}$ and $\mathbb{P}(\tau_{V+} > 0) > 0$ we must therefore have $\alpha_+(\eta) > 0, \alpha_-(\eta) < 0$ for $\eta > 0$. In addition, $\alpha_+(l)$ is clearly right-continuous and increasing, so it must have at most countably many discontinuities, and similarly for $\alpha_-(l)$. We can write

$$\inf\{t : \mathfrak{L}_t > 0, B_t \notin (\alpha_-(\mathfrak{L}_t-), \alpha_+(\mathfrak{L}_t-))\} \leq \tau^*_{\text{CL}} \leq \tau^*_{\text{OP}} \leq \inf\{t : \mathfrak{L}_t > 0, B_t \notin [\alpha_-(\mathfrak{L}_t), \alpha_+(\mathfrak{L}_t)]\}$$

and observe that (by standard properties of Brownian motion) the stopping times on the left and right are almost surely equal (since there are at most countably many discontinuities, and $\alpha_+(l)$ and $\alpha_-(l)$ are bounded away from zero on $[\eta, \infty)$ for $\eta > 0$). It follows that $\tau_{V+} = \inf\{t : \mathfrak{L}_t > 0, B_t \notin (\alpha_-(\mathfrak{L}_t), \alpha_+(\mathfrak{L}_t))\}$ on $\{\tau_{V+} > 0\}$, and we deduce that τ_{V+} is zero with probability $\mu(\{0\})$, and, conditional on being greater than zero, $\tau_{V+} = \inf\{t > 0 : B_t \notin (\alpha_-(\mathfrak{L}_t), \alpha_+(\mathfrak{L}_t))\}$ a.s. □

Remark 6.15. The arguments above extend from local time at 0 to a general continuous additive functional A. Recalling that $\mathfrak{L}^x$ denotes local time in x, A can be represented in the form $A_t := \int_0^t \mathfrak{L}^x_s \, dm_A(x)$. Let f be a convex function such that $f'' = m_A$ in the sense of distributions. If $\int f \, d\mu < \infty$ then the above proof is easily adapted to the more general situation.

In this manner, we deduce the existence of optimal solutions to (SEP) for functions depending on A. By analogy with Theorem 6.14 this can be used to generate (inverse-/cave-) barrier type embeddings of various kinds. Other generalizations and variants may be considered in a similar manner. We leave specific examples as an exercise for the reader.

6.3. **Root and Rost Embeddings in Higher Dimensions.** In this section we consider the Root and Rost constructions of Sections 2.1 and 2.2 in the case of d-dimensional Brownian motion with general initial distribution, for $d \geq 2$. In $\mathbb{R}^d$, since Brownian motion is transient, it is no longer straightforward to assert the existence of an embedding. In general, [49] gives necessary and sufficient conditions for the existence of an embedding, and without the additional condition that $\mathbb{E}[\tau] < \infty$. In the Brownian case, Rost's conditions for $d \geq 3$ can be written as follows. There exists a stopping time τ such that $B_0 \sim \lambda$ and $B_\tau \sim \mu$ if and only if for all $y \in \mathbb{R}^d$

$$\int u(x,y)\, \lambda(dx) \leq \int u(x,y)\, \mu(dx), \text{ where } u(x,y) = |x-y|^{2-d}. \tag{6.10}$$

However, it is not clear that such a stopping time will satisfy the condition

$$\mathbb{E}[\tau] = 1/d\left(\int |x|^2 (\mu - \lambda)(dx)\right). \tag{6.11}$$

As a result, it is not straightforward to give simple criteria for the existence of a solution in $\mathsf{RST}(\mu)$.

In the case $d = 2$ it follows from Falkner's results [22] that the Skorokhod problem admits a solution (i.e. $\mathsf{RST}(\mu) \neq \emptyset$) if (6.10) is satisfied for $u(x, y) = -\ln|x - y|$ and then (6.11) applies.

In either case, assuming that we do have a solution satisfying (6.11), then the existence result as well as the monotonicity principle carry over to the present setup (with identical proofs) and we are able to state the following:

Theorem 6.16. *Suppose* $\mathsf{RST}(\mu)$ *is non-empty. If* h *is a strictly convex function and* $\hat{\tau} \in \mathsf{RST}(\mu)$ *minimizes* $\mathbb{E}[h(\tau)]$ *over* $\tau \in \mathsf{RST}(\mu)$ *then there exists a barrier* $\mathcal{R}$ *such that* $\hat{\tau} = \inf\{t > 0 : (B_t, t) \in \mathcal{R}\}$ *on* $\{\hat{\tau} > 0\}$ *a.s.*

The proof of this result is much the same as that of Theorem 2.1, except we no longer show that $\tau_{\mathrm{CL}} = \tau_{\mathrm{OP}}$. In higher dimensions with general initial laws, it is easy to construct examples where there are common atoms of λ and μ, but where the size of the atom in λ is strictly larger than the atom of μ. By the transience of the process, it is clear that the optimal (indeed, only) behaviour is to stop mass starting at such a point immediately with a probability strictly between 0 and 1, however the stopping times τ_{CL} and τ_{OP} will always stop either all the mass, or none of this mass respectively. For this reason, we do not say anything about the behaviour of $\hat{\tau}$ when $\hat{\tau} = 0$. Trivially, the above result tells us that the solution of the optimal embedding problem is given by a barrier if there exists a set D such that $\lambda(D) = 1 = \mu(D^c)$.

Proof of Theorem 6.16. The first part of the proof proceeds similarly to the proof of Theorem 2.1. In particular, the set of stop-go pairs is given by

$$\mathsf{SG} \supseteq \{((f, s), (g, t)) \in S \times S : f(s) = g(t), s > t\}$$

and we define the sets $\mathcal{R}_{\mathrm{CL}}, \mathcal{R}_{\mathrm{OP}}$ and the stopping times $\tau_{\mathrm{CL}}, \tau_{\mathrm{OP}}$ as above. We then fix $\delta > 0$, and consider the set $\{\hat{\tau} \geq \delta\}$. Given $\eta \geq 0$, we define $B_t^{-\eta} = B_{t+\eta}$, for $t \geq -\eta$ and set

$$\tau_{\mathrm{CL}}^{\eta,\delta} := \inf\{t \geq \delta : (t, B_t^{-\eta}) \in \mathcal{R}_{\mathrm{CL}}\}.$$

Then $\tau_{\mathrm{CL}}^{\eta,\delta} \geq \delta$, and for any $\varepsilon > 0$, there exists $\eta > 0$ sufficiently small that $d_{TV}(B_\delta^{-\eta}, B_\delta) < \varepsilon$, where d_{TV} denotes the total variation distance. By the Strong Markov property of Brownian motion, it follows that $d_{TV}(B^{-\eta}_{\tau_{\mathrm{CL}}^{\eta,\delta}}, B_{\tau_{\mathrm{CL}}^{0,\delta}}) < \varepsilon$. In particular, the law of $B^{-\eta}_{\tau_{\mathrm{CL}}^{\eta,\delta}}$ converges weakly to the law of $B_{\tau_{\mathrm{CL}}^{0,\delta}}$ as $\eta \to 0$. Thus

$$\tau_{\mathrm{CL}}^{\eta,\delta} = \inf\{t \geq \eta + \delta : (t - \eta, B_t) \in \mathcal{R}_{\mathrm{CL}}\},$$

so $\tau_{\mathrm{CL}}^{\eta,\delta} \geq \tau_R^{0,\delta}$, and moreover, $\tau_{\mathrm{CL}}^{\eta,\delta} \to \tau_{\mathrm{OP}}^{0,\delta}$ a.s. as $\eta \to 0$. Hence, $B^{-\eta}_{\tau_{\mathrm{CL}}^{\eta,\delta}} \to B_{\tau_{\mathrm{OP}}^{0,\delta}}$ in probability, as $\eta \to 0$, so we have weak convergence of the law of $B^{-\eta}_{\tau_{\mathrm{CL}}^{\eta,\delta}}$ to the law of $B_{\tau_{\mathrm{OP}}^{0,\delta}}$, and hence $B_{\tau_{\mathrm{OP}}^{0,\delta}} \sim B_{\tau_{\mathrm{CL}}^{0,\delta}}$. We now observe that, by an essentially identical argument to that in the proof of Theorem 2.1, we must have $\tau_{\mathrm{CL}}^{0,\delta} \leq \hat{\tau} \leq \tau_{\mathrm{OP}}^{0,\delta}$ on $\{\hat{\tau} \geq \delta\}$. However, in the argument above, we know that $\tau_{\mathrm{CL}}^{0,\delta} \leq \hat{\tau} \leq \tau_{\mathrm{OP}}^{0,\delta}$, and $\tau_{\mathrm{CL}}^{\eta,\delta} \to_{\mathcal{D}} \tau_{\mathrm{CL}}^{0,\delta}$ and $\tau_{\mathrm{CL}}^{\eta,\delta} \to_{\mathcal{D}} \tau_{\mathrm{OP}}^{0,\delta}$ as $\eta \to 0$ (where $\mathcal{D}$ denotes convergence in distribution). It follows that $\tau_{\mathrm{CL}}^{0,\delta} =_{\mathcal{D}} \tau_{\mathrm{OP}}^{0,\delta}$ and hence $\tau_{\mathrm{CL}}^{0,\delta} = \tau_{\mathrm{OP}}^{0,\delta}$ a.s. In particular, $B_{\tau_{\mathrm{CL}}^{0,\delta}} = B_{\tau_{\mathrm{OP}}^{0,\delta}} = B_{\hat{\tau}}$ on $\{\hat{\tau} \geq \delta\}$. Letting $\delta \to 0$ we observe that $\tau_{\mathrm{OP}}^{0,\delta} \to \tau_{\mathrm{OP}}$, and hence the required result holds on taking $\mathcal{R} = \mathcal{R}_{\mathrm{OP}}$. □

We now consider the generalization of the Rost embedding. Recall that $(\min(\lambda, \mu))(A) := \inf_{B \subseteq A} (\lambda(B) + \mu(A \setminus B))$ defines a measure.

Theorem 6.17. *Suppose* λ, μ *are measures in* $\mathbb{R}^d$ *and* $\hat{\tau} \in \mathsf{RST}(\mu)$ *maximizes* $\mathbb{E}[h(\tau)]$ *over all stopping times in* $\mathsf{RST}(\mu)$*, for a convex function* $h : \mathbb{R}_+ \to \mathbb{R}$*, with* $\mathbb{E}[h(\tau)] < \infty$*. Then*

$\mathbb{P}(\hat{\tau} = 0, B_0 \in A) = (\min(\lambda, \mu))(A)$*, for* $A \in \mathcal{B}(\mathbb{R})$*, and on* $\{\hat{\tau} > 0\}$*,* $\hat{\tau}$ *is the first hitting time of an inverse barrier.*

Proof. We follow the proof of Theorem 2.4 to recover the set of stop-go pairs given by

$$\mathsf{SG} \supseteq \{((f, s), (g, t)) \in S \times S : f(s) = g(t), s < t\}$$

and the sets $\mathcal{R}_{\text{OP}}$ and $\mathcal{R}_{\text{CL}}$, and their corresponding hitting times $\tau_{\text{OP}}, \tau_{\text{CL}}$. For $0 \leq \eta \leq \delta$, we define in addition the stopping times

$$\tau_{\text{CL}}^{\eta,\delta} := \inf\{t \geq \delta : (t, B_t^\eta) \in \mathcal{R}_{\text{CL}}\},\ \tau_{\text{OP}}^{\eta,\delta} := \inf\{t \geq \delta : (t, B_t^\eta) \in \mathcal{R}_{\text{OP}}\},$$

where $B_t^\eta = B_{t-\eta}$, for $t \geq \eta$.

It follows from an identical argument to that in the proof of Theorem 2.4 that $\tau_{\text{CL}}^{0,\delta} \leq \hat{\tau} \leq \tau_{\text{OP}}^{0,\delta}$ on $\{\hat{\tau} \geq \delta\}$. However, by similar arguments to those used above, we deduce that $\tau_{\text{OP}}^{0,\delta}$ and $\tau_{\text{CL}}^{0,\delta}$ have the same law on $\{\hat{\tau} \geq \delta\}$, and hence that $\hat{\tau} = \tau_{\text{OP}}^{0,\delta}$ on this set, and then by taking $\delta \to 0$, we get $\hat{\tau} = \tau_{\text{OP}}$ on $\{\hat{\tau} > 0\}$.

To see the final claim, we note that trivially $\mathbb{P}(\hat{\tau} = 0, B_0 \in A) \leq (\min(\lambda, \mu))(A)$. If there is strict inequality, then there exist some paths in Γ which start at $x \in A$, and paths in Γ which stop at x at strictly positive time, constituting a stop-go pair and therefore violating the monotonicity principle. □

Remark 6.18. We observe that the arguments of Remark 2.3 can be applied again in this context. However, one needs to be a little more careful, since it is necessary to take the *fine closure* of the barriers with respect to the fine topology for the processes $(t, B_t)_{t\geq 0}$. With this modification in place, the argument of Loynes can be easily adapted to show that the (finely closed versions) of the barriers in Theorems 6.16 and 6.17 are unique in the sense of Remark 2.3.

6.4. **An optimal Skorokhod embedding problem which admits only randomized solutions.** By analogy with optimal transport, we might interpret a 'natural stopping time' (i.e. a stopping time wrt to the Brownian filtration) which solves (OptSEP) as a Monge-type solution whereas stopping times which depend on additional randomization are of Kantorovich-type. With the exception of the Rost solution, all optimal stopping times encountered in the previous section are natural stopping times, and in the Rost case external randomization is only needed at time 0. One might ask whether the optimal Skorokhod embedding problem always admits a solution τ which is natural on $\{\tau > 0\}$. We sketch an example, showing that this is not the case:

Example 6.19. There exist an absolutely continuous probability μ and a continuous adapted process $\gamma_t = \gamma((B_s)_{s\leq t})$ with values in $[0, 1]$ such that (OptSEP) admits only randomized solutions.

Proof. Define the stopping time $\sigma := \inf\{t \geq 0 : B_t^2 + t^2 \geq 1\}$, the first time the Brownian path leaves the right half of the unit disc. Write $(C(0, \sigma), \mathbb{W}_\sigma)$ for the space of continuous functions up to time σ, equipped with the corresponding projection of Wiener measure. Pick an isomorphism

$$l : (C(0, \sigma), \mathbb{W}_\sigma) \to ([2, 3], \mathcal{L})$$

of standard Borel probability spaces. Using some extra randomization (independent of $\mathcal{F}^B$) we define a stopping time τ such that

(1) $\tau = \sigma$ with probability $1/2$,
(2) otherwise τ stops the first time the Brownian path reaches the level $\pm l((B_s)_{s\leq\sigma})$.

We then define $\mu := \text{Law}(B_\tau)$ and pick γ to be a function which equals 0 on paths which are stopped by τ and is strictly positive otherwise; clearly we can do this in such a way that γ has continuous paths.

Write $\hat{\tau}$ for the randomized stopping time $\mathsf{RST}(\mu)$ corresponding to τ. It is then straightforward to see that $\hat{\tau}$ is the unique solution of (OptSEP). Thus, the optimal Skorokhod embedding problem admits no (non-randomized) solution in the natural filtration of B. □

In optimal transport it is a difficult and interesting problem to understand under which conditions transport problems admit solutions of Monge-type. An interesting subject for future research would be to understand when Monge-type solutions exist for the optimal Skorokhod embedding problem.

7. Skorokhod Embedding for Feller processes

In this section we discuss the extension of our results to the embedding problem for a continuous Feller process Z, with values in $\mathbb{R}^d$ and $Z_0 \sim \lambda$. Throughout we suppose that Z is defined on a stochastic basis $(\Omega, \mathcal{G}, (\mathcal{G}_t)_{t\geq 0}, \mathbb{P})$ which is sufficiently rich to support a uniformly distributed $\mathcal{G}_0$-random variable independent of Z. Given a probability $\mu \in \mathcal{P}(\mathbb{R}^d)$ the analogue of (SEP) is to construct a stopping time τ such that

$$Z_\tau \sim \mu, \quad \tau \text{ is minimal.} \tag{SEP^Z}$$

Recall from (2.1) that a stopping time τ is minimal iff for any stopping time τ' such that $Z_{\tau'} \sim Z_\tau$ then $\tau' \leq \tau$ implies $\tau' = \tau$ a.s. If Z is a one dimensional Brownian motion and μ has second moment, minimality of τ is equivalent to $\mathbb{E}[\tau] < \infty$. Working in higher dimensions with general starting law we redefine

$$S := \{(f, s) : f \in C([0, s], \mathbb{R}^d)\}.$$

Given a function $\gamma : S \to \mathbb{R}$ the optimal Skorokhod embedding problem for Z is to construct a stopping time optimizing

$$P^Z_\gamma := \inf\{\mathbb{E}[\gamma((Z_s)_{s\leq\tau}, \tau)] : \tau \text{ solves (SEP}^Z)\}. \tag{OptSEP^Z}$$

(As above, the value of P^Z_γ does not depend on the underlying stochastic basis provided it supports a uniformly distributed random variable independent of Z.)

Most of the arguments required to establish our main results are abstract and carry over to the present setup. In fact, only the parts building on the condition $\mathbb{E}[\tau] < \infty$ need to be adjusted to account for the more general condition of τ being minimal. Therefore, to establish Theorems 1.1, 1.2, and 1.3 in the general Feller setup, we need the crucial Assumption 7.1 below which we verify in a number of natural examples in Section 7.2.

Assumption 7.1. *From now on we* assume *that* (SEP^Z) *admits a solution and either*

(1) that there exist continuous functions $h : \mathbb{R}^n \to \mathbb{R}$ and $\zeta : S \to \mathbb{R}$ such that:
- *$\zeta_t := \zeta((Z_s)_{s\leq t}, t)$ is strictly increasing, $\zeta_0 = 0$, $\lim_{t\to\infty} \zeta_t = \infty$, $\mathbb{P}$-a.s. and*
- *$X_t := h(Z_t) - \zeta_t$ is a martingale and $(X^\tau_t)_{t\geq 0}$ is uniformly integrable for all τ solving* (SEP^Z)*, or*

(2) that whenever τ is a finite stopping time satisfying $Z_\tau \sim \mu$ then τ is minimal and there is an increasing function $G : \mathbb{R}_+ \to \mathbb{R}$, $\lim_{t\to\infty} G(t) = \infty$ which satisfies

$$\sup\{\mathbb{E}[G(\tau)] : \tau \textit{ solves } (\text{SEP}^Z)\} =: V < \infty. \tag{7.1}$$

The existence of a function G such that (7.1) holds is equivalent to

$$\{\tau : \tau \text{ solves (SEP}^Z)\} \text{ is bounded in probability.} \tag{7.2}$$

In fact, it is straightforward to see that we would arrive at an equivalent condition when replacing the deterministic function G by a stochastic process $(\zeta_t)_{t\geq 0}$ as in Case (1).

Note also that in Case (1) of Assumption 7.1, τ with $Z_\tau \sim \mu$ is minimal if and only if

$$\mathbb{E}[\zeta_\tau] = \int h\, d\mu - \int h\, d\lambda =: V \ \ (< \infty). \tag{7.3}$$

Under Assumption 7.1, our main results extend to continuous Feller processes:

Theorem 7.2. *If $\gamma : S \to \mathbb{R}$ is lsc and bounded from below,* (OptSEP^Z) *admits a minimizer.*

Theorem 7.3. *Let $\gamma : S \to \mathbb{R}$ be lsc and bounded from below. Then we have the duality relation $P^Z_\gamma = D^Z_\gamma$ for $D^Z_\gamma := \sup \int \psi(y)\, d\mu(y)$, where the supremum is taken over all continuous $\psi \in L^1(\mu)$ such that there exists a continuous bounded martingale M with $\mathbb{E}[M_0] = 0$*

and a decreasing process A with $\mathbb{E}[A_\tau] \geq 0$ for all solutions τ of (SEP^Z) *and almost surely for all $t \geq 0$*

$$M_t + A_t + \psi(Z_t) \leq \gamma((Z_s)_{s\leq t}, t). \tag{7.4}$$

Moreover, in Case (1) of Assumption 7.1, the process A may be assumed to be zero at the expense of assuming that $(M_{\tau\wedge t})_{t\geq 0}$ is only uniformly integrable for all τ solving (SEP^Z).

Theorem 7.4. *Let $\gamma : S \to \mathbb{R}$ be Borel measurable. If* (OptSEP^Z) *is well posed and τ is an optimizer, there exists a γ-monotone Borel set $\Gamma \subseteq S$ such that $\mathbb{P}$-a.s.*

$$((Z_t)_{t\leq\tau}, \tau) \in \Gamma.$$

Remark 7.5. (1) Of course, the analogues of the secondary optimization results, Theorems 6.1 (on existence of a minimizer) and 6.4 (monotonicity principle), carry over to the present setup with the obvious changes.
(2) The continuity of ζ on S which was imposed in Assumption 7.1 (1) is not required in Theorems 7.2 and 7.4.
(3) The condition $0 < \mathbb{E}[\sigma] < \infty$ in Definition 1.4 should be replaced by considering all stopping times with $0 < \mathbb{E}[\zeta((B_s)_{s\leq\sigma}, \sigma)] < \infty$ in case (1) of Assumption 7.1, or $0 < \mathbb{E}[G(\tau)] < \infty$ in case (2). In addition, the expectation should be taken over the law of the Feller process started at $f(s) = g(t)$.

7.1. **Sketch of proofs.** As in Section 3 we consider the canonical setup $(C(\mathbb{R}_+, \mathbb{R}^d), \mathcal{F}^0, \mathbb{Q})$ (where $\mathbb{Q}$ denotes the law of the Feller process) and we write Y for the canonical process. It follows from continuity of Y (resp. Z) and the Feller property that the $\mathcal{F}^a$-optional and the $\mathcal{F}^a$-predictable σ-algebra on the canonical space agree; similarly Proposition 3.5 on the definition of S-continuous martingales extends to the present context. We define RST, JOIN and related notions as before with $\mathbb{Q}$ replacing $\mathbb{W}$. We say that $\xi \in \mathsf{RST}$ is a minimal embedding of μ if the corresponding stopping time ρ (cf. (3.6)) on the enlarged probability space $(C(\mathbb{R}_+, \mathbb{R}^d) \times [0, 1], \bar{\mathbb{Q}})$ constitutes a minimal embedding. (Representing randomized stopping times as in Theorem 3.8 (1), the stopping time ξ constitutes a minimal embedding iff there is no randomized stopping time $\xi' \neq \xi$ embedding the same measure which satisfies $A^{\xi'} \geq A^\xi$.) For $\mu \in \mathcal{P}(\mathbb{R}^d)$ we define $\mathsf{RST}(\mu)$ to be the set of all minimal randomized stopping times embedding the measure μ.

Recalling the argument from Theorem 3.14, we see that the existence of a function $\zeta : S \to \mathbb{R}$ such that $\zeta \circ r$ increases to ∞ and $\sup_{\xi\in\mathsf{RST}(\mu)} \xi(\zeta \circ r) < \infty$ implies that $\mathsf{RST}(\mu)$ is compact. (Vice versa, if $\mathsf{RST}(\mu)$ is compact then such a function exists and can be chosen so that $\zeta \circ r$ is deterministic). Hence, by (7.3) resp. (7.1), $\mathsf{RST}(\mu)$ is compact.

Proof of Theorem 7.2. The argument follows the proof of Theorem 4.1 line by line. □

Proof of Theorem 7.3. We give the argument in the case $\lambda = \delta_0$ for ease of exposition. Setting $h = \zeta \circ r$ resp. $h = G \circ T$ (and using identical arguments as previously) we obtain the following extension of Proposition 4.3:

For $c : C_0(\mathbb{R}_+) \times \mathbb{R}_+ \times \mathbb{R} \to \mathbb{R} \cup \{\infty\}$ lsc, predictable and bounded from below

$$\inf_\pi \int c(\omega, t, y)\, d\pi(\omega, t, y) = \sup_{(\varphi,\psi)} \int \varphi\, d\mathbb{Q} + \int \psi\, d\mu, \tag{7.5}$$

where the infimum is taken over the set $\mathsf{JOIN}^{1,V}(\mu) = \{\pi \in \mathsf{JOIN}^1(\mu) : \pi(h) \leq V\}$ and the supremum is taken over $\varphi \in C_b(C_0(\mathbb{R}_+))$, $\psi \in C_b(\mathbb{R})$ for which

$$\exists \alpha \geq 0 \text{ s.t. } \varphi_t^M(\omega) + \psi(y) - \alpha(h_t - V) \leq c(\omega, t, y) \text{ for } \omega \in C_0(\mathbb{R}_+), t \in \mathbb{R}_+, y \in \mathbb{R}.$$

The argument used to derive Theorem 4.2 from Proposition 4.3 then implies the desired duality relation $P_\gamma^Z = D_\gamma^Z$, with a decreasing process (in (7.4)) of the form $A_t = -\alpha(h_t - V)$ for some $\alpha \geq 0$. In Case (1) of Assumption 7.1, A can be 'hidden' in M / ψ as in (4.14). □

Proof of Theorem 7.4. Apart from the abstract theory the ingredients of the proof of Theorem 5.7 are Proposition 5.8 and Proposition 5.9. The only stage where the proof of Proposition 5.8 has to be altered is when establishing that the randomized stopping time ξ^π is

minimal. Under Assumption (7.1) (1) this follows using the minimality characterization given in (7.3), under Assumption (7.1) (2) this is of course trivial.

Proposition 5.9 only uses transport duality, the Feller property to construct S-continuous martingales and Choquet's capacitability theorem. □

7.2. **Examples.** We now provide a list of Examples in which Assumption 7.1 is satisfied and Theorems 7.2, 7.3, and 7.4 apply.

7.2.1. *Let Z be a one-dimensional Brownian motion and assume that λ and μ have first moments and are in convex order. Then Assumption 7.1 (1) holds.*

Proof. By the de la Vallée-Poussin theorem (see e.g. [16, Thm. II 22]) there exists a positive, smooth and symmetric function $F : \mathbb{R} \to \mathbb{R}_+$ with strictly positive, bounded second derivative and $\lim_{x\to\infty} F(x)/x = \infty$ such that $V := \int F(x)\,\mu(dx) < \infty$. We set

$$\zeta_t := 1/2 \int_0^t F''(Z_s)\,ds$$

and note that ζ_t increases to ∞ since $\mathbb{Q}(\int_0^\infty \mathbb{1}_{[-1,1]}(Z_t)\,dt = \infty) = 1$ and F'' is bounded away from 0 on $[-1, 1]$. Using Itô's formula and our conditions on F we define the martingale

$$X_t := F(Z_t) - 1/2 \int_0^t F''(Z_s)\,ds = F(Z_t) - \zeta_t.$$

In the present Brownian case, it is known that the minimality of a finite stopping time τ is equivalent to $(Z_{\tau\wedge t})_{t\geq 0}$ being a uniformly integrable martingale. This follows (in the case of a general starting law) from Lemma 12 and Theorem 17 of [10].

If $Z_\tau \sim \mu$ and $(Z_t^\tau)_{t\geq 0}$ is uniformly integrable, then for each t, the law of $Z_{\tau\wedge t}$ is bounded by μ in the convex order and in particular $\mathbb{E}[F(Z_{\tau\wedge t})] \leq V, t \geq 0$. Uniform integrability of X then follows upon noting

$$\mathbb{E}[\zeta_\tau] = \lim_{t\to\infty} \mathbb{E}[\zeta_t^\tau] = \lim_{t\to\infty} \mathbb{E}[F(Z_t^\tau)] - \mathbb{E}[F(Z_0)] \leq V - \mathbb{E}[F(Z_0)] < \infty. \quad \square$$

7.2.2. *One-dimensional regular diffusions.* Let Z be a regular (time-homogeneous) one-dimensional diffusion on an interval $I \subseteq \mathbb{R}$, with inaccessible or absorbing endpoints (see [47] for the relevant definitions and terminology) and $Z_0 \sim \lambda$, $\lambda(I^\circ) = \mu(I^\circ) = 1$. In particular, Z is a continuous Feller process ([47, Proposition V.50.1]). Then (on a possibly enlarged probability space) there exists a scale function s and a continuous, strictly increasing time change A_t such that $B_t = s(Z_{A_t})$ is a Brownian motion up to the exit of $s(I^\circ)$. Recalling the discussion in [11, Section 5], with the obvious extension of our notation, it is clear that there exists a minimal stopping time τ embedding μ in Z if and only if there exists a stopping time τ' embedding $s(\mu)$ in B such that

$$\tau' \leq \tau_{s(I)} := \inf\{t \geq 0 : B_t \notin s(I^\circ)\}. \tag{7.6}$$

Moreover, write A_t^{-1} for the inverse of A_t, so $A_{A_t}^{-1} = t$. Since A and A^{-1} are continuous and strictly increasing τ is a minimal embedding of μ in Z if and only if $\tau' := A_\tau^{-1}$ is a minimal embedding of $s(\mu)$ in B.

We now consider three cases. In the first two we verify Assumption 7.1 (2) and in the last case we verify Assumption 7.1 (1) under some additional smoothness assumptions. Subsequently we give some concrete examples.

(i) Suppose $s(I^\circ) = (a, b)$ for $a, b \in \mathbb{R}$. Then it follows from [10, Theorems 17 and 22] that a solution to (SEP^Z) exists if and only if $s(\lambda)$ precedes $s(\mu)$ in convex order, and in fact, any finite τ with $Z_\tau \sim \mu$ is minimal.

Moreover we note that
- $\{\tau' : B_{\tau'} \sim s(\mu), \tau'$ is a minimal$\}$ is bounded in probability
- $A_t < \infty$ provided the path $(B_s)_{s\leq t}$ stays inside an interval $[c, d] \subseteq (a, b)$.
- Given $\varepsilon > 0$ there exists an interval $[c, d] \subseteq (a, b)$ such that $(B_s)_{s\leq\tau'}$ stays inside $[c, d]$ with probability $> 1 - \varepsilon$ for each minimal τ', $B_{\tau'} \sim s(\mu)$.

It follows that $\{A_{\tau'} : B_{\tau'} \sim s(\mu), \tau'$ is minimal$\}$ is bounded in probability, hence (7.2) and then Assumption 7.1 (2) holds.

(ii) Suppose $s(I^\circ) = (a, \infty)$ for $a \in \mathbb{R}$, and that $s(\lambda)$ and $s(\mu)$ are in convex order and that the moments $m_\lambda = \int s(y)\,\lambda(dy)$, $m_\mu = \int s(y)\,\mu(dy)$ exist. Then it follows from Theorems 17 and 22 and the discussion at the top of p. 245 of [10] that a solution to (SEP^Z) exists if and only if for all $x \geq a$,

$$-\int |s(y) - x|\,\mu(dy) \leq -\int |s(y) - x|\,\lambda(dy) + (m_\lambda - m_\mu) \tag{7.7}$$

Again, any finite τ with $Z_\tau \sim \mu$ is minimal and (7.2) follows as above.

An analogous result holds if $s(I^\circ) = (-\infty, b)$ for $b \in \mathbb{R}$.

(iii) Suppose $s(I^\circ) = (-\infty, \infty)$ and that $s(\lambda), s(\mu)$ are in convex order, $\int s(y)^2\,\mu(dy) < \infty$. Then we are in the classical case, and a stopping time τ with $Z_\tau \sim \mu$ is minimal if and only if $\mathbb{E}[A_\tau^{-1}] < \infty$. If the process Z is sufficiently well-behaved (as in the examples below) one can show that $X_t = s(Z_t)^2 - A_t^{-1}$ is a martingale and that A^{-1} depends continuously on the path $(Z_s)_{s\leq t}$. For all τ solving (SEP^Z), $\mathbb{E}[A_\tau^{-1}] < \infty$; hence $(X_t^\tau)_{t\geq 0}$ is uniformly integrable and Assumption 7.1 (1) is satisfied.

More generally, when only the integrals $\int s(y)\,\lambda(dy)$, $\int s(y)\,\mu(dy)$ are finite, (assuming sufficient regularity of the diffusion), Assumption 7.1 (1) follows as in Section 7.2.1.

Remark 7.6. Observe that none of the constructions described in Sections 6.1 and 6.2 rely on fine properties of Brownian motion — the main properties used are the continuity of paths, the strong Markov property, and the regularity and diffusive nature of paths (that the process started at x immediately returns to x, and immediately enters the sets (x, ∞) and $(-\infty, x)$). It follows that all the given constructions extend to the case of regular diffusions described above.

Example 7.7 (Brownian motion with drift). Let $Z_t = B_t + at$ for some $a < 0$ with $Z_0 \sim \lambda$, and $I = (-\infty, \infty)$. Then a possible choice of the scale function is $s(x) = \exp(-2ax)$. Let $\lambda, \mu \in \mathcal{P}(\mathbb{R})$ be such that $s(\lambda), s(\mu)$ are integrable and satisfy (7.7). Then Assumption 7.1 holds by (ii) above.

Example 7.8 (Geometric Brownian motion). Let Z be a geometric Brownian motion, given through the SDE $dZ_t = Z_t dB_t$, $Z_0 \sim \lambda$. A possible choice of scale function is $s(x) = x$. Let $\lambda, \mu \in \mathcal{P}(0, \infty)$ be such that $s(\lambda), s(\mu)$ are integrable and satisfy the corresponding version of (7.7). Then Assumption 7.1 holds by (ii) above. (More general versions of geometric Brownian motion can be treated similarly.)

Example 7.9 (Three-dimensional Bessel process). Let $Z = |B|$ for a three-dimensional Brownian motion $(B_t)_{t\geq 0}$ with $Z_0 \sim \lambda$. A possible choice of scale function is $s(x) = 1 - 1/x$, and $s(I^\circ) = (-\infty, 1)$. Let $\lambda, \mu \in \mathcal{P}(0, \infty)$ be such that $s(\lambda), s(\mu)$ are integrable and satisfy the corresponding version of (7.7). Then Assumption 7.1 holds by (ii) above. Similar results hold for d-dimensional Bessel processes, with $d > 2$.

Example 7.10 (Ornstein-Uhlenbeck process). Let Z be an Ornstein-Uhlenbeck process, given for example as the solution to the SDE $dZ_t = -Z_t\,dt + dW_t, Z_0 \sim \lambda$. Then Z_t is a regular diffusion on $I = (-\infty, \infty)$ with scale function given (up to constants) by $s'(x) = \exp(x^2)$, and $s(I^\circ) = (-\infty, \infty)$. Suppose λ, μ are measures on $\mathbb{R}$ such that $s(\lambda), s(\mu)$ are in convex order and $\int s(y)^2\,\mu(dy) < \infty$. Then $A_t^{-1} = \int_0^t \exp\{2Z_t^2\}\,ds$ is continuous as a function of $(Z_s)_{s\leq t}$, and hence Assumption 7.1 holds by (iii) above.

7.2.3. *The Hoeffding-Frechet coupling as a very particular Root solution.* Let Z be the deterministic process given by $dZ_t = dt$ started in $Z_0 \sim \lambda$. Z is not a regular diffusion, however Assumption 7.1 (2) is easily checked. Let μ be another probability and assume for simplicity that $\max \operatorname{supp} \lambda \leq \min \operatorname{supp} \mu$. Then the Root solution minimizes $\mathbb{E}[\tau^2]$. But note also that since $\tau = Z_\tau - Z_0$, this minimization problem corresponds precisely to finding the joint distribution (Z_0, Z_τ) which minimizes $\mathbb{E}[(Z_\tau - Z_0)^2]$: the classical transport problem in the most simple setup. Specifically, the Root solution for the particular case of the process Z corresponds precisely to the monotone (Hoeffding-Frechet) coupling. In the same fashion the Rost solution corresponds to the co-monotone coupling between λ and μ.

References

[1] B. Acciaio, M. Beiglböck, F. Penkner, W. Schachermayer, and J. Temme. A trajectorial interpretation of doob's martingale inequalities. *The Annals of Applied Probability*, 23(4):1494–1505, 2013.

[2] D. R. Adams and L. I. Hedberg. *Function spaces and potential theory*, volume 314 of *Grundlehren der Mathematischen Wissenschaften [Fundamental Principles of Mathematical Sciences]*. Springer-Verlag, Berlin, 1996.

[3] L. Ambrosio and A. Pratelli. Existence and stability results in the L^1 theory of optimal transportation. In *Optimal transportation and applications (Martina Franca, 2001)*, volume 1813 of *Lecture Notes in Math.*, pages 123–160. Springer, Berlin, 2003.

[4] J. Azéma and M. Yor. Une solution simple au problème de Skorokhod. In *Séminaire de Probabilités, XIII (Univ. Strasbourg, Strasbourg, 1977/78)*, volume 721 of *Lecture Notes in Math.*, pages 90–115. Springer, Berlin, 1979.

[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, M. Goldstern, G. Maresch, and W. Schachermayer. Optimal and better transport plans. *J. Funct. Anal.*, 256(6):1907–1927, 2009.

[7] M. Beiglböck, P. Henry-Labordère, and F. Penkner. Model-independent bounds for option prices: A mass transport approach. *Finance and Stochastics*, 17(3):477–501, 2013.

[8] S. Bianchini and L. Caravenna. On optimality of c-cyclically monotone transference plans. *C. R. Math. Acad. Sci. Paris*, 348(11-12):613–618, 2010.

[9] B. Bouchard and M. Nutz. Arbitrage and duality in nondominated discrete-time models. *The Annals of Applied Probability*, 25(2):823–859, 2015.

[10] A. M. G. Cox. Extending Chacon-Walsh: minimality and generalised starting distributions. In *Séminaire de probabilités XLI*, volume 1934 of *Lecture Notes in Math.*, pages 233–264. Springer, Berlin, 2008.

[11] A. M. G. Cox and D. Hobson. Skorokhod embeddings, minimality and non-centred target distributions. *Probability Theory and Related Fields*, 135:395–414, 2005.

[12] A. M. G. Cox and J. Obłój. Robust hedging of double touch barrier options. *SIAM J. Financial Math.*, 2:141–182, 2011.

[13] 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.

[14] A.M.G. Cox and G Peskir. Embedding laws in diffusions by functions of time. *The Annals of Probability*, 43(5):2481–2510, 2015.

[15] F. Delbaen and W. Schachermayer. A general version of the fundamental theorem of asset pricing. *Math. Ann.*, 300(3):463–520, 1994.

[16] C. Dellacherie and P.-A. Meyer. *Probabilities and potential*, volume 29 of *North-Holland Mathematics Studies*. North-Holland Publishing Co., Amsterdam, 1978.

[17] C. Dellacherie and P.-A. Meyer. *Probabilities and potential. B*, volume 72 of *North-Holland Mathematics Studies*. North-Holland Publishing Co., Amsterdam, 1982. Theory of martingales, Translated from the French by J. P. Wilson.

[18] C. Doléans. Existence du processus croissant naturel associé à un potentiel de la classe (d). *Probability Theory and Related Fields*, 9(4):309–314, 1968.

[19] Y. Dolinsky and H. M. Soner. Martingale optimal transport and robust hedging in continuous time. *Probab. Theory Relat. Fields*, 160(1-2):391–427, 2014.

[20] Y. Dolinsky and H. M. Soner. Martingale optimal transport in the Skorokhod space. *Stochastic Processes and their Applications*, 125(10):3893–3931, 2015.

[21] Y. Dolinsky and M. H. Soner. Robust hedging with proportional transaction costs. *Finance Stoch.*, 18(2):327–347, 2014.

[22] N. Falkner. The distribution of Brownian motion in $\mathbf{R}^n$ at a natural stopping time. *Adv. in Math.*, 40(2):97–127, 1981.

[23] 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.

[24] W. Gangbo and R. McCann. The geometry of optimal transportation. *Acta Math.*, 177(2):113–161, 1996.

[25] P. Gassiat, A. Mijatović, and H. Oberhauser. An integral equation for Root's barrier and the generation of Brownian increments. *Ann. Appl. Probab.*, 25(4):2039–2065, 2015.

[26] P. Gassiat, H. Oberhauser, and G. dos Reis. Root's barrier, viscosity solutions of obstacle problems and reflected FBSDEs. *Stochastic Processes and their Applications*, 125(12):4601–4631, 2015.

[27] G. Guo, X. Tan, and N. Touzi. Optimal Skorokhod embedding under finitely-many marginal constraints. *ArXiv e-prints*, June 2015.

[28] P. Henry-Labordère, J. Obłój, P. Spoida, and N. Touzi. The maximum maximum of a martingale with given n marginals. *Ann. Appl. Probab.*, 26(1):1–44, 2016.

[29] F. Hirsch, C. Profeta, B. Roynette, and M. Yor. *Peacocks and associated martingales, with explicit constructions*, volume 3 of *Bocconi & Springer Series*. Springer, Milan; Bocconi University Press, Milan, 2011.

[30] D. Hobson. Robust hedging of the lookback option. *Finance and Stochastics*, 2:329–347, 1998.

[31] 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.

[32] D. Hobson and A. Neuberger. Robust bounds for forward start options. *Mathematical Finance*, 22(1):31–56, 2012.

[33] S. Jacka. Doob's inequalities revisited: A maximal h^1-embedding. *Stochastic processes and their applications*, 29(2):281–290, 1988.

[34] 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.

[35] A. S. Kechris. *Classical descriptive set theory*, volume 156 of *Graduate Texts in Mathematics*. Springer-Verlag, New York, 1995.

[36] J. Kiefer. Skorohod embedding of multivariate RV's, and the sample DF. *Z. Wahrscheinlichkeitstheorie und Verw. Gebiete*, 24(1):1–35, 1972.

[37] M. Knott and C. S. Smith. On the optimal mapping of distributions. *J. Optim. Theory Appl.*, 43(1):39–49, 1984.

[38] G. Last, P. Mörters, and H. Thorisson. Unbiased shifts of Brownian motion. *Ann. Probab.*, 42(2):431–463, 2014.

[39] R. M. Loynes. Stopping times on Brownian motion: Some properties of Root's construction. *Z. Wahrscheinlichkeitstheorie und Verw. Gebiete*, 16:211–218, 1970.

[40] P.-A. Meyer. Convergence faible et compacité des temps d'arrêt, d'après baxter et chacon. In *Séminaire de probabilités*, volume XII of *Lecture Notes in Mathematics, 649*, pages 411–423. Springer Berlin Heidelberg, 1978.

[41] I. Monroe. On embedding right continuous martingales in Brownian motion. *Ann. Math. Statist.*, 43:1293–1311, 1972.

[42] J. Najnudel and A. Nikeghbali. A new kind of augmentation of filtrations. *ESAIM: Probability and Statistics*, 15:S39–S57, January 2011.

[43] J. Obłój. The Skorokhod embedding problem and its offspring. *Probab. Surv.*, 1:321–390, 2004.

[44] J. Obłój, P. Spoida, and N. Touzi. Martingale inequalities for the maximum via pathwise arguments. In *In Memoriam Marc Yor-Séminaire de Probabilités XLVII*, pages 227–247. Springer, 2015.

[45] E. Perkins. The Cereteli-Davis solution to the H^1-embedding problem and an optimal embedding in Brownian motion. In *Seminar on stochastic processes, 1985*, pages 172–223. Springer, 1986.

[46] D. Revuz and M. Yor. *Continuous martingales and Brownian motion*, volume 293 of *Grundlehren der Mathematischen Wissenschaften [Fundamental Principles of Mathematical Sciences]*. Springer-Verlag, Berlin, third edition, 1999.

[47] L. C. G. Rogers and D. Williams. *Diffusions, Markov processes and martingales: Volume 2, Itô calculus*, volume 2. Cambridge university press, 2000.

[48] D. H. Root. The existence of certain stopping times on Brownian motion. *Ann. Math. Statist.*, 40:715–718, 1969.

[49] H. Rost. The stopping distributions of a Markov Process. *Invent. Math.*, 14:1–16, 1971.

[50] H. Rost. Skorokhod stopping times of minimal variance. In *Séminaire de Probabilités, X (Première partie, Univ. Strasbourg, Strasbourg, année universitaire 1974/1975)*, pages 194–208. Lecture Notes in Math., Vol. 511. Springer, Berlin, 1976.

[51] L. Rüschendorf. Fréchet-bounds and their applications. In *Advances in probability distributions with given marginals (Rome, 1990)*, volume 67 of *Math. Appl.*, pages 151–187. Kluwer Acad. Publ., Dordrecht, 1991.

[52] L. Rüschendorf. Optimal solutions of multivariate coupling problems. *Appl. Math. (Warsaw)*, 23(3):325–338, 1995.

[53] A. V. Skorohod. *Issledovaniya po teorii sluchainykh protsessov (Stokhasticheskie differentsialnye uravneniya i predelnye teoremy dlya protsessov Markova)*. Izdat. Kiev. Univ., Kiev, 1961.

[54] 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.

[55] H. Strasser. *Mathematical theory of statistics*, volume 7 of *de Gruyter Studies in Mathematics*. Walter de Gruyter & Co., Berlin, 1985. Statistical experiments and asymptotic decision theory.

[56] P. Vallois. Le probleme de Skorokhod sur $\mathbb{R}$: une approche avec le temps local. In *Séminaire de Probabilités XVII 1981/82*, pages 227–239. Springer, 1983.

[57] C. Villani. *Topics in optimal transportation*, volume 58 of *Graduate Studies in Mathematics*. American Mathematical Society, Providence, RI, 2003.

[58] C. Villani. *Optimal Transport. Old and New*, volume 338 of *Grundlehren der mathematischen Wissenschaften*. Springer, 2009.