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. It is not hard to see that $\mathbb{E}[\tau] = \int x^2\,\mu(dx)$ for any solution of (SEP). As already demonstrated by Skorokhod [48] in the mid-1960's, it is always possible to construct solutions to the problem. Indeed, the survey article 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 the survey of Obłój [35] (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 of optimal transportation. Many of the previous developments are thus obtained as applications of one unifying principle (Theorem 1.2) and several difficult problems are rendered tractable.

1.1. **A motivating example — Root's construction.**

To illustrate our approach we introduce a solution that will serve as inspiration in the rest of the paper: Root's construction [41] which is one of the earliest solutions to (SEP). 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.

Date: September 13, 2014.

The authors thank Walter Schachermayer for many discussions and Nizar Touzi for helpful comments. The first author was supported by the FWF-grant P21209 and P26736, the third author by the CRC 1060.

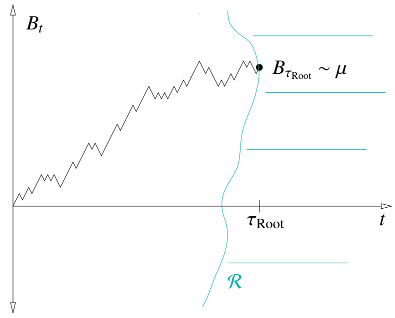

FIGURE 1. Root's solution of (SEP).

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

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

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 maximises $\mathbb{E}[\sqrt{\tau}]$. 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 functional

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

The optimal Skorokhod embedding problem is to construct a stopping time optimizing

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

We will usually assume that (OptSEP) is well-posed in the sense that $\mathbb{E}\big[\gamma((B_t)_{t\leq\tau}, \tau)\big]$ exists with values in $[-\infty, \infty)$ for all τ which solve (SEP) and is finite for one such τ.

As mentioned above, the Root stopping time solves (OptSEP) in the case where $\gamma(f, s) = \sqrt{s}$. Other examples where the solution is known include functionals depending on the running maximum $\gamma((f, s)) := \bar{f}(s) := \max_{t\leq s} f(t)$ or functionals 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 entirely obvious; for example, the optimality property of the Root embedding was first conjectured by Kiefer [28] and subsequently established by Rost [43].

In contrast, our starting point will be the optimization problem (OptSEP) and we seek a systematic method to construct the maximizer for a given functional γ. To develop a general theory for this optimization problem we interpret stopping times in terms of a transport plan from the Wiener space $(C(\mathbb{R}_+), \mathbb{W})$ 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. Instead we develop an analogous theory, which in particular needs to account for the adaptedness properties of stopping times. To this end, it is necessary to combine ideas and results from optimal transportation 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 Theorem 4.12 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 "Kantorovich-type" stopping times, 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) is compact in a natural sense. As a particular consequence we will establish:

Theorem 1.1. *Let $\gamma : S \to \mathbb{R}$ be upper semi-continuous and bounded from above in the sense*[1] *that for some constants $a, b, c \in \mathbb{R}_+$*

$$\gamma(f, s) \leq a + b \cdot s + c \cdot \max_{r \leq s} f(r)^2, \quad (f, s) \in S. \tag{1.3}$$

Then (OptSEP) *admits a maximizer $\hat{\tau}$. More precisely, there exists a Brownian motion B on some stochastic basis $\Omega = (\Omega, \mathcal{F}, (\mathcal{F}_t)_{t\geq 0}, \mathbb{P})$ and a stopping time $\hat{\tau}$ of B which attains* (OptSEP).

Here we can talk about the continuity properties of γ since S possesses a natural Polish topology (cf. (4.1)).

In terms of linear optimization, Theorem 1.1 is a primal problem. In Section 5 we will introduce the corresponding dual problem and establish that there is no duality gap.

1.3. **Geometric Characterization of Optimizers — Montonicity 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* which we recall in (3.4) below. Its relevance for the theory of optimal transport has been fully recognized by Gangbo and McCann [20], based on earlier work of Knott and Smith [30] and Rüschendorf [44, 45] 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.

Importantly, this transport-based approach readily admits a number of strong generalisations and extensions. With only minor changes our existence result, Theorem 1.1, and the monotonicity principle, Theorem 1.2, extend to general starting distributions and Brownian motion in $\mathbb{R}^d$ and more generally to sufficiently regular Markov processes; see Sections 6, 7 and 9. This is notable since previous constructions usually exploit rather specific properties of Brownian motion.

Theorem 1.2 (Monotonicity Principle). *Let $\gamma : S \to \mathbb{R}$ be Borel measurable, B be a Brownian motion on some stochastic basis $(\Omega, \mathcal{F}, (\mathcal{F}_t)_{t\geq 0}, \mathbb{P})$ and τ an optimizer of* (OptSEP). *Then, there exists a* γ-monotone *Borel set $\Gamma \subseteq S$ such that $\mathbb{P}$-a.s.*

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

[1]In terms of Brownian motion, Assumption (1.3) amounts to $\gamma((B_r)_{r\leq s}, s) \leq a + b \cdot s + c \cdot \max_{r\leq s} B_r^2$ and naturally holds in all our intended applications. Other conditions which guarantee uniform integrability of the positive part of γ w.r.t. solutions of (SEP) would suffices as well.

If (1.4) holds, we will loosely say that Γ *supports* τ. The significance of Theorem 1.2 is to link 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.

The precise definition of *γ-monotonicity* is intricate and we present it in its simplest and most strict form in this introductory section. (See Definition 7.1 for a more general version that leads to a stronger assertion in Theorem 1.2.) We observe that γ is only required to be Borel. This will be important when we apply our results.

The notion of γ-monotonicity expresses the following idea, which we first introduce informally. The set $\Gamma \subseteq S$ contains all the possible 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 $(f(r))_{r\in[0,t]}$. Corresponding to the set of stopped, or "killed" paths is the set of paths which we may observe, and at which we may not yet have stopped: these are the "living" paths. Since all paths must eventually be killed, we deduce that a path may be living if there is a longer, killed path which contains the living path as a sub-path. Specifically, if $(g, t) \in \Gamma$ is a killed path, then the sub-paths $(f, s) = ((g(r))_{t\in[0,s]}, s)$ are living for all $s < t$. We will write $\Gamma^<$ for the set of living paths corresponding to the killed paths Γ, so:

$$\Gamma^< := \{(f, s) : \exists (g, t) \in \Gamma, s < t \text{ and } f \equiv g \text{ on } [0, s]\} \tag{1.5}$$

We now consider a possible modification to a given stopping rule. Imagine that we have one "living" path $(f, s) \in \Gamma^<$ and a second "killed" path $(g, t) \in \Gamma$. If $f(s) = g(t)$, then we can imagine "killing" the path (f, s) at time s, and allowing (g, t) to live by transferring all paths which extend (f, s), the "remaining lifetime", onto the newly-living (g, t). If this guarantees an improved value of γ in total for any possible remaining lifetime, then we call $((f, s), (g, t))$ a *bad pair*, since our original rule would be improved by this swapping procedure. If the set Γ and its living sub-paths $\Gamma^<$ contain no bad pairs, we call the set Γ a γ-monotone set. Observe that the condition $f(s) = g(t)$ is necessary to guarantee that a modified stopping rule still embeds the measure μ. A pictorial representation of this process is given in Figure 2.

We formalise these ideas in the following definition:

Definition 1.3. *We say that $((f, s), (g, t)) \in S$ is a* bad pair *iff $f(s) = g(t)$ and for all $(h, u) \in S$ it holds that*

$$\gamma((f \oplus h, s + u)) + \gamma((g, t)) \quad < \quad \gamma((f, s)) + \gamma((g \oplus h, t + u)), \tag{1.6}$$

where $f \oplus h$ denotes the concatenation[2] *of the the two paths f and h. The set of bad pairs will be denoted by BP. Then a set $\Gamma \subseteq S$ is called γ-monotone iff*

$$\mathsf{BP} \cap (\Gamma^< \times \Gamma) = \emptyset.$$

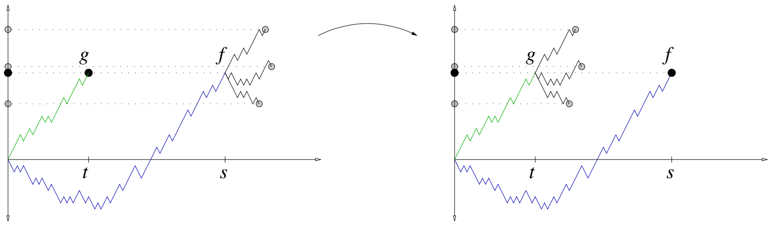

FIGURE 2. The left side shows a bad pair in the Root case $\gamma(f, s) = \sqrt{s}$. Here (1.6) asserts $\sqrt{s+u} + \sqrt{t} < \sqrt{s} + \sqrt{t+u}$. Intuitively, switching the roles of f and g improves the embedding since $s > t$.

[2]More precisely, $f \oplus h(t) = f(t)$ for $t \in [0, s]$ and $f \oplus h(t) = f(s) + h(t - s)$ for $t \in [s, s + u]$.

In Section 2 below we give a short teaser on how particular embeddings are obtained as a consequence of Theorem 1.2: there we establish the Root and the Rost solutions of (SEP), as well as a continuum of new embeddings which "interpolate" between them. It will become clear that the essence of the proof is already contained in Figure 2.

The monotonicity principle, Theorem 1.2, is the most complex part of this paper, and requires 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. Without the elements from stochastic analysis, the "classical" optimal transport version of Theorem 1.2 can be established through fairly direct arguments, at least in a reasonably regular setting, cf. [3, Thms. 3.2, 3.3] and [51, Ex. 2.38]. However, these approaches do not extend easily to our probabilistic setup. The argument given subsequently is more in the spirit of [5, 7] and requires a fusion of ideas from optimal transport and stochastic analysis. To achieve this, we will need to revisit a number of classical notions from the theory of stochastic processes within a novel framework.

1.4. **New Horizons.** The methods and 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 generalisation of the results in Section 4, where the specific structure of the space of continuous paths is exploited.
(2) **Multiple path-swapping:** In our monotonicity principle, Theorem 1.2, we consider the impact of swapping mass from a single unstopped path onto a single stopped path, and argue that if this improve the objective γ on average, then we cannot observe such behaviour under an optimiser. 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 occurs in the present setup.
(3) **Multiple marginals:** A natural generalisation 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 maximises $\mathbb{E}[\gamma((B_t)_{t\leq\tau_n}, \tau_1, \dots, \tau_n)]$ for a suitable functional γ. In this setup, it is natural to ask whether there exists a suitable monotonicity principle, corresponding to Theorem 1.2.
(4) **Constrained embedding problems:** In this paper, we consider classical embedding problems, where the optimisation 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 maximises some functional 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 generalisations of our duality results, and a corresponding monotonicity principle for such problems.

1.5. **Background.** Since the first solution to (SEP) by Skorokhod [48] 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. [41, 43, 4, 26, 50, 36]. The survey [35] 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. [31], 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. [22] have used solutions to the Skorokhod embedding problem to construct Peacocks; and Gassiat et. al. [21], 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 [23]: 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*, where the Skorokhod embedding problem has become a central tool; we refer the reader to the survey article [24] 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. The first papers to study such problems include [25, 6, 19, 18], 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 [19]; through a stochastic control problem they recover the Azèma-Yor solution of the Skorokhod embedding problem. Notably, their approach is very different from the one pursued in the present paper: the approach of this paper is instead to use an analogue of c-cyclical monotonicity from classical optimal transport in the martingale context.

1.6. **Organization of the Article.** In Section 2 we establish the Root and the Rost embedding as well as a family of new embeddings. In Section 3 we recall some required definitions and results from optimal transport. In Section 4 we consider randomized stopping times on the Wiener space and establish some basic properties. In Section 5 we develop a dual problem to (OptSEP) and prove our duality using classical duality results from optimal transport. In Sections 6 and 7 we will finally establish Theorem 1.2 by combining the duality theory with Choquet's capacity theorem. In Section 8 we use our results to establish the known solutions to (OptSEP) as well as further embeddings. In Section 9 we describe extensions to Feller processes under certain assumptions, which we are able to verify for a large class of processes.

2. Particular embeddings

In this section we explain how Theorem 1.2 can be used to derive particular solutions to the Skorokhod embedding problem, (SEP), using the optimisation problem (OptSEP). For much of the paper, we will 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 solves (SEP), τ is minimal if, for any solution τ' to (SEP), $\tau' \leq \tau$ a.s. implies $\tau' = \tau$ a.s.. In the case where μ has a second moment, minimality and $\mathbb{E}[\tau] < \infty$ are equivalent. The extension of our results to the more general case will be discussed in Section 9.

We first recall one of the key parts of Definition 1.3

Definition 2.1. *Write $(g \oplus h, t+u)$ for the path obtained from concatenating (g,t) and $(h,u) \in S$. Then the set of* bad pairs *for $\gamma : S \to \mathbb{R}$ is given by*

$$\mathsf{BP} = \Big\{((f,s),(g,t)) : f(s) = g(t), \ \textit{for all } (h,u) \in S, u > 0$$
$$\gamma((f \oplus h, s+u)) + \gamma((g,t)) \quad < \quad \gamma((f,s)) + \gamma((g \oplus h, t+u))\Big\}.$$

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 ([41]) from an optimisation problem. Let $\gamma(f,t) = h(t)$, where $h : \mathbb{R}_+ \to \mathbb{R}$ is a strictly concave function such that

$$\sup\{\mathbb{E}[h(\tau)] : \tau \text{ solves (SEP)}\} \tag{2.1}$$

is well posed and pick a maximizer $\hat\tau$ of (2.1) by Theorem 1.1. Then we have:

Theorem 2.2. *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* (1.1).

Proof. Pick, by Theorem 1.2, a γ-*monotone* set $\Gamma \subseteq S$ such that $\mathbb{P}(((B_t)_{t\leq\hat\tau}, \hat\tau) \in \Gamma) = 1$. Note that due to the concavity of h the set of bad pairs is given by (cf. Figure 2)

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

As Γ is γ-monotone, $(\Gamma^< \times \Gamma) \cap \mathsf{BP} = \emptyset$. Define a closed and an open barrier by

$$\mathcal{R}_{\text{CL}} := \{(s,x) : \exists (g,t) \in \Gamma, g(t) = x, t \leq s\}, \tag{2.2}$$
$$\mathcal{R}_{\text{OP}} := \{(s,x) : \exists (g,t) \in \Gamma, g(t) = x, t < s\}, \tag{2.3}$$

and denote the respective hitting times by τ_{CL} and τ_{OP}. We claim that $\tau_{\text{CL}} \leq \hat\tau \leq \tau_{\text{OP}}$ a.s..

Note that $\tau_{\text{CL}} \leq \hat\tau$ holds by definition of τ_{CL}. To show the other inequality pick ω satisfying $((B_t(\omega))_{t\leq\hat\tau(\omega)}, \hat\tau(\omega)) \in \Gamma$ and assume for contradiction that $\tau_{\text{OP}}(\omega) < \hat\tau(\omega)$. Then there exists $s < \hat\tau(\omega)$ such that $(s, B_s(\omega)) \in \mathcal{R}_{\text{OP}}$. By definition of the open barrier, this means that there is some $(g,t) \in \Gamma$ such that $t < s$ and $g(t) = B_s(\omega)$. But then $(f,s) := ((B_u(\omega))_{u\leq s}, s) \in \Gamma^<$, hence $\big((f,s),(g,t)\big) \in \mathsf{BP} \cap (\Gamma^< \times \Gamma)$ which is the desired contradiction. We finally observe that $\tau_{\text{CL}} = \tau_{\text{OP}}$ 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[3] of the Skorokhod embedding problem which maximizes (2.1). Assume that maximizers τ_1 and τ_2 are given. Then we can use an independent coin-flip to define a new maximizer $\bar\tau$ 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 $\tau_1 = \tau_2$.

In Section 8.4 we will prove generalisations of Theorem 2.2 which admit similar conclusions in $\mathbb{R}^d$ and for general initial distributions.

We also note that the above proof of Theorem 2.2 is based on a heuristic derivation of the optimality properties of the Root embedding given by Hobson in [24]. Indeed Hobson's approach was the starting point of the present paper.

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 [43] 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:

Let $\gamma(f,t) = h(t)$, where $h : \mathbb{R}_+ \to \mathbb{R}$ is a strictly *convex* function such that the problem to maximize $\mathbb{E}[h(\tau)]$ over all solutions to the Skorokhod-problem (SEP) is well posed. Pick, by Theorem 1.1, a maximizer $\hat\tau$. Then we have:

Theorem 2.3. *Suppose $\mu(\{0\}) = 0$. Then 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 problem can be solved by a hitting time of an inverse barrier.*

[3] This was first established in [43], together with the optimality property of Root's solution.

Proof. Pick, by Theorem 1.2, a γ-*monotone* set $\Gamma \subseteq S$ such that $\mathbb{P}(((B_t)_{t\le\hat\tau},\hat\tau)\in\Gamma)=1$. Note that due to the convexity of h the set of bad pairs is given by (cf. Figure 2)

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

As Γ is γ-monotone, $(\Gamma^< \times \Gamma) \cap \mathsf{BP} = \emptyset$. Define open and closed inverse barriers by

$$\mathcal{R}_{\text{OP}} := \{(s,x) : \exists (g,t) \in \Gamma, g(t) = x, s < t\}, \tag{2.4}$$

$$\mathcal{R}_{\text{CL}} := \{(s,x) : \exists (g,t) \in \Gamma, g(t) = x, s \le t\}, \tag{2.5}$$

and denote the respective hitting times by τ_{OP} and τ_{CL}. We claim that $\tau_{\text{CL}} \le \hat\tau \le \tau_{\text{OP}}$ a.s..

Note that $\tau_{\text{CL}} \le \hat\tau$ holds by definition of τ_{CL}. To show the other inequality pick ω satisfying $((B_t(\omega))_{t\le\hat\tau(\omega)},\hat\tau(\omega)) \in \Gamma$ and assume for contradiction that $\tau_{\text{OP}}(\omega) < \hat\tau(\omega)$. Then there exists $s < \hat\tau(\omega)$ such that $(s, B_s(\omega)) \in \mathcal{R}_{\text{OP}}$. By definition of the open inverse barrier, this means that there is some $(g,t) \in \Gamma$ such that $s < t$ and $g(t) = B_s(\omega)$. But then $(f,s) := ((B_u(\omega))_{u\le s}, s) \in \Gamma^<$, hence $\big((f,s),(g,t)\big) \in \mathsf{BP} \cap (\Gamma^< \times \Gamma)$ which is the desired contradiction.

Then there exists an increasing function $b(t) = \inf\{x > 0 : (t,x) \in \mathcal{R}_{\text{CL}}\}$ and a decreasing function $c(t) = \sup\{x < 0 : (t,x) \in \mathcal{R}_{\text{CL}}\}$ such that τ_{CL} is the first time $B_t \notin (c(t), b(t))$.

It then follows that τ_{OP} and τ_{CL} are almost surely equal — see for example equation (2.9) of [12]. □

In Section 8.4 we will give a generalisation of this result, which includes a more direct verification of the final part of this proof, which does not rely on the result of [12].

As in the case of the Root embedding we obtain that the maximizer 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 1.2. 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 $[0,t_0]$.

Let $\gamma((f,s)) = \varphi(s)$. Since φ is bounded, the problem to minimise $\mathbb{E}[\varphi(\tau)]$ over all solutions to (SEP) is well posed. Pick, by Theorem 1.1, a minimiser $\hat\tau$. Then we have:

Theorem 2.4 (Cave embedding)**.** *Suppose $\mu(\{0\}) = 0$. Then there exists a cave barrier $\mathcal{R}$ such that $\hat\tau = \inf\{t \ge 0 : (t,B_t) \in \mathcal{R}\}$. In particular the Skorokhod embedding problem can be solved by a hitting time of a cave barrier.*

Proof. Pick, by Theorem 1.2, a γ-*monotone* set $\Gamma \subseteq S$ such that $\mathbb{P}(((B_t)_{t\le\hat\tau},\hat\tau)\in\Gamma)=1$. We define an open cave barrier by

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

and $\mathcal{R}_{\text{OP}} = \mathcal{R}^0_{\text{OP}} \cup \mathcal{R}^1_{\text{OP}}$ (resp. a closed cave barrier where we allow $t \le s$ and $s \le t$ in $\mathcal{R}^0_{\text{CL}}$ and $\mathcal{R}^1_{\text{CL}}$ resp.). We denote the corresponding hitting time by $\tau_{\mathcal{R}_{\text{OP}}} = \tau_{R^0_{\text{OP}}} \wedge \tau_{R^1_{\text{OP}}}$ (resp. $\tau_{\mathcal{R}_{\text{CL}}}$). We claim that $\tau_{\mathcal{R}_{\text{CL}}} \le \hat\tau \le \tau_{\mathcal{R}_{\text{OP}}}$ $\mathbb{W}$-a.s.. The first inequality follows by construction. To show the second inequality, we argue by contradiction. Suppose that $\tau_{R^0_{\text{OP}}} < \hat\tau$ on a set of positive mass. Then there exist $(f,s) \in \Gamma^<$ and $(g,t) \in \Gamma$ with $g(t) = f(s)$ and $s < t \le t_0$. We claim that $((f,s),(g,t)) \in \mathsf{BP}$. Indeed, since we are minimising the objective, for 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 $t \le t_0, t + r > t_0$ this follows since φ' is strictly positive on $[0, t_0)$ and strictly negative on (t_0, ∞). Hence, $\tau_{R^0_{\text{OP}}} \ge \hat{\tau}$. Similarly it follows that $\tau_{R^1_{\text{OP}}} \ge \hat{\tau}$ and therefore $\tau_{\mathcal{R}_{\text{OP}}} \ge \tau$. As in Theorems 2.2 and 2.3 above, we can deduce that $\tau_{\mathcal{R}_{\text{CL}}} = \tau_{\mathcal{R}_{\text{OP}}}$ a.s., proving the claim. □

2.4. **Remarks.** The arguments given here only use the properties of one-dimensional Brownian motion to show that our candidate stopping times τ_{OP} and τ_{CL} are the same. In Section 8.4 we will show that these arguments can be adapted to prove the existence of Rost and Root embeddings in a more general setting. In fact, in Sections 8 and 9 we will show that the above 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 rather 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 $\tau_{\text{OP}} < \tau_{\text{CL}}$ 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 randomisation is necessary. In the Rost case, it is also straightforward to compute the independent randomisation which preserves the embedding property.

Other recent approaches to the Root and Rost embeddings can be found in [13, 34, 12, 14]. These papers largely exploit PDE techniques, and as a result, are able to produce more explicit descriptions of the barriers, but the methods tend to be highly specific to the problem under consideration.

3. The classical Transport Problem

We will shortly review here some notions of transport theory which are used below or which will serve as motivations for analogous concepts in our probabilistic setup.

In abstract terms the transport problem (cf. [51, 52]) can be stated as follows: For probabilities λ, μ on Polish spaces X, Y the set $\mathsf{Cpl}(\lambda, \mu)$ of *transport plans* consists of all *couplings* between λ and μ. These are all measures on $X \times Y$ with X-marginal λ and Y-marginal μ. Associated to a *cost function* $c : X \times Y \to [0, \infty]$ and $\pi \in \mathsf{Cpl}(\lambda, \mu)$ are the *transport costs* $\int_{X \times Y} c(x, y)\, d\pi(x, y)$. The Monge-Kantorovich problem is then to determine the value

$$\inf \Big\{ \int c\, d\pi : \pi \in \mathsf{Cpl}(\lambda, \mu) \Big\} \tag{3.1}$$

and to identify an *optimal* transport plan $\hat{\pi} \in \mathsf{Cpl}(\lambda, \mu)$, i.e. a minimizer of (3.1). Going back to Kantorovich, this is related to the following dual problem. Consider the set $\Phi(\lambda, \mu)$ of pairs (φ, ψ) of integrable functions $\varphi : X \to [-\infty, \infty)$ and $\psi : Y \to [-\infty, \infty)$ which satisfy $\varphi(x) + \psi(y) \le c(x, y)$ for all $(x, y) \in X \times Y$. The dual counterpart of the Monge-Kantorovich problem is then to maximize

$$J(\varphi, \psi) = \int_X \varphi\, d\lambda + \int_Y \psi\, d\mu \tag{3.2}$$

over $(\varphi, \psi) \in \Phi(\lambda, \mu)$. In the literature duality has been established under various conditions, see for instance [52, p. 98f] for a short overview.

Theorem 3.1 (Monge-Kantorovich Duality, [27, Theorem 2.2])**.** *Let $(X, \lambda), (Y, \mu)$ be Polish probability spaces and $c : X \times Y \to [0, \infty]$ be lower semi-continuous. Then*

$$\inf \Big\{ \int c\, d\pi : \pi \in \mathsf{Cpl}(\lambda, \mu) \Big\} = \sup \Big\{ J(\varphi, \psi) : (\varphi, \psi) \in \Phi(\lambda, \mu) \Big\}. \tag{3.3}$$

Moreover the duality relation pertains if the optimization in the dual problem is restricted to continuous and bounded functions φ, ψ.

A basic and important goal is to characterize minimizers through a tractable property of their support sets: a Borel set $\Gamma \subseteq X \times Y$ is *c-cyclically monotone* iff

$$c(x_1, y_2) - c(x_1, y_1) + \ldots + c(x_{n-1}, y_n) - c(x_{n-1}, y_{n-1}) + c(x_n, y_1) - c(x_n, y_n) \geq 0 \qquad (3.4)$$

whenever $(x_1, y_1), (x_2, y_2), \ldots, (x_n, y_n) \in \Gamma$. A transport plan π is *c-cyclically monotone* if it assigns full measure to some cyclically monotone set Γ.

Concerning the origins of c-cyclical monotonicity in convex analysis and the study of the relation to optimality we mention [39, 29, 46, 20]. Intuitively speaking, c-cyclically monotone transport plans resist improvement by means of cyclical rerouting and optimal transport plans are expected to have this property. Indeed we have:

Theorem 3.2. *Let $c : X \times Y \to \mathbb{R}_+$ be a lower semi-continuous cost function. Then a transport plan is optimal if and only if it is c-cyclically monotone.*

Even in the case where c is the squared Euclidean distance this is a non trivial result, posed as an open question by Villani in [51, Problem 2.25]. Following contributions of Ambrosio and Pratelli [3], this problem was resolved by Pratelli [37] and Schachermayer and Teichmann [47] who established the clear-cut characterization stated in Theorem 3.2. Lower semi-continuity of the cost function can also be relaxed, as shown in [5] and [7].

We will need the following straightforward corollary of Theorem 3.1. In this article, we usually consider maximisation problems. Therefore, we switch the sup and inf in the corollary as we will use it in this form in Section 5. However, by considering $\tilde{c} := -c$ this is clearly equivalent to the usual convention.

Corollary 3.3. *Let $\tilde{c} : X \times Y \times [0, t_0] \to \mathbb{R}$ be upper semi-continuous and bounded from above. Then*

$$\sup\Big\{\int \tilde{c}\, d\pi : \pi \in \mathcal{P}(X \times Y \times [0, t_0]), \mathrm{proj}_X(\pi) = \lambda, \mathrm{proj}_Y(\pi) = \mu\} \qquad (3.5)$$

$$= \inf\Big\{J(\varphi, \psi) : (\varphi, \psi) \in L^\infty(\lambda) \times L^\infty(\mu), \varphi(x) + \psi(y) \geq \tilde{c}(x, y, t)\Big\}. \qquad (3.6)$$

Again, the duality relation pertains if the optimization in the dual problem is restricted to continuous and bounded functions φ, ψ.

4. Preliminaries on stopping times and filtrations

4.1. **Spaces and Filtrations.** In this section we mainly discuss the formal aspects of filtrations, measure theory, etc., and how classical notions relate to properties of functions on the space S introduced above.

We consider the space $\Omega = C(\mathbb{R}_+)$ of continuous paths with the topology of uniform convergence on compact sets. The elements of Ω will be denoted by ω. We denote the canonical process on Ω by $(B_t)_{t\geq 0}$, i.e. $B_t(\omega) = \omega_t$. As explained above we consider the set S of all continuous functions defined on some initial segment $[0, s]$ of $\mathbb{R}_+$; 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 $g_{\upharpoonright [0,s]}$ of g to the interval $[0, s]$ equals f. In this case we write $(f, s) \prec (g, t)$. We consider S with the topology determined by the following metric: let $(f, s), (g, t) \in S$ and suppose $s \leq t$. We then say that (f, s) and (g, t) are ε-close if

$$d_S((f, s), (g, t)) := \max\Big(t - s, \sup_{0\leq u\leq s} |f(u) - g(u)|, \sup_{s\leq u\leq t} |g(u) - g(s)|\Big) < \varepsilon. \qquad (4.1)$$

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 $C(\mathbb{R}_+) \times \mathbb{R}_+$ and S. We therefore discuss the underlying filtrations in some detail.

We consider three different filtrations on the Wiener space $C(\mathbb{R}_+)$, the canonical or natural filtration $\mathcal{F}^0 = (\mathcal{F}^0_t)_{t\in\mathbb{R}_+}$, the right-continuous filtration $\mathcal{F}^+ = (\mathcal{F}^+_t)_{t\in\mathbb{R}_+}$, and the augmented filtration $\mathcal{F}^a = (\mathcal{F}^a_t)_{t\in\mathbb{R}_+}$ obtained from $(\mathcal{F}^0_t)_{t\in\mathbb{R}_+}$ by including all $\mathbb{W}$-null sets in $\mathcal{F}^0_0$. As Brownian motion is a continuous Feller process, $\mathcal{F}^a$ is automatically right-continuous, all $\mathcal{F}^a$-stopping times are predictable and all right-continuous $\mathcal{F}^a$-martingales are continuous. In particular, the $\mathcal{F}^a$-optional and the $\mathcal{F}^a$-predictable σ-algebras coincide (see e.g. [38, Corollary IV 5.7]). By [16, Thm. IV. 97, Rem. IV. 98] we also have that the $\mathcal{F}^0$-predictable,-optional and -progressive σ-algebras coincide because Ω is the set of *continuous* paths. Moreover, we will often use the following result.

Theorem 4.1 ([16, Theorem IV. 78])**.** *For every $\mathcal{F}^a$-predictable process $(X_t)_{t\in\mathbb{R}_+}$ there is an $\mathcal{F}^0$-predictable process $(X'_t)_{t\in\mathbb{R}_+}$ which is indistinguishable from $(X_t)_{t\in\mathbb{R}_+}$. If τ is an $\mathcal{F}^a$-stopping time, there exists an $\mathcal{F}^0$-stopping time τ' such that $\tau=\tau'$ a.s..*

Of course, every $\mathcal{F}^a$-martingale has a continuous version. Not so commonly used but entirely straightforward is the following: if M is an $\mathcal{F}^0$-martingale then there is a version M' of M which is an $\mathcal{F}^0$-martingale and almost all paths of M' are continuous.

The message of Proposition 4.4 below is that a process $(X_t)_{t\in\mathbb{R}_+}$ is $\mathcal{F}^0$-predictable iff $(X_t)_{t\in\mathbb{R}_+}$ is $\mathcal{F}^0$-optional iff $X_t(\omega)$ can be calculated from the restriction $\omega_{\upharpoonright[0,t]}$. We introduce the mapping

$$r : C(\mathbb{R}_+)\times\mathbb{R}_+ \to S, \quad r(\omega,t) = (\omega_{\upharpoonright[0,t]}, t). \tag{4.2}$$

Note that the topology on S introduced in (4.1) coincides with the final topology induced by the mapping r; in particular r is a continuous open mapping. The mapping r is not a closed mapping: it is easy to see that there exist closed sets in $C(\mathbb{R})\times\mathbb{R}$ with a non-closed image under r. However this does not happen for closed optional sets, see Proposition 4.4.

Remark 4.2. *In the following we will say that $H : S \to \mathbb{R}$ is* continuous / right-continuous / etc. *if the corresponding property holds for the process $H\circ r$. Similarly we say that $H_1, H_2 : S \to \mathbb{R}$ are* indistinguishable *if this holds for the processes $H_1\circ r, H_2\circ r$ w.r.t. Wiener measure. We will also often use the notation $H(\omega_{\upharpoonright[0,t]}) = H(\omega_{\upharpoonright[0,t]}, t)$.*

Definition 4.3. *We say that a process $X_t(\omega)$ is S- continuous if there exists a continuous function $h : S \to \mathbb{R}$ such that*

$$X_t(\omega) = h((\omega_{\upharpoonright[0,t]}, t))$$

for all $t\geq 0$, $\mathbb{W}$-a.s..

It is trivially true that an S- continuous process is $\mathcal{F}^0$-adapted, and continuous ($\mathbb{W}$-a.s.). The converse is not generally true — consider the case where X_t is the local time of the Brownian motion at a level x. This is a continuous, $\mathcal{F}^0$-adapted process, however the corresponding function h is not a continuous mapping from S to $\mathbb{R}$. (Indeed, any path which has strictly positive local time can be approximated uniformly by paths with both zero and infinite local time).

Proposition 4.4. *$\mathcal{F}^0$-optional sets / functions on $C(\mathbb{R}_+)\times\mathbb{R}_+$ correspond to Borel-measurable sets / functions on S. More precisely we have:*

1. *A set $D\subseteq C(\mathbb{R}_+)\times\mathbb{R}_+$ is $\mathcal{F}^0$-optional iff there is a Borel set $A\subseteq S$ with $D=r^{-1}(A)$.*
2. *A process $X=(X_t)_{t\in\mathbb{R}_+}$ is $\mathcal{F}^0$-optional iff there is a Borel measurable $H : S\to\mathbb{R}$ such that $X = H\circ r$.*

An $\mathcal{F}^0$-optional set $A\subseteq C(\mathbb{R}_+)\times\mathbb{R}_+$ is closed in $C(\mathbb{R}_+)\times\mathbb{R}_+$ iff the corresponding set $r(A)$ is closed in S. An $\mathcal{F}^0$-optional process $X = H\circ r$ is S- continuous iff $H : S\to\mathbb{R}$ is continuous.

For the proof of Proposition 4.4 we need another result from [16]. Write $a_t : \Omega\to\Omega$ for the stopping operation, i.e. $a_t(\omega)$ is the path which agrees with ω until t and stays constant afterwards.

Theorem 4.5 (cf. [16, Theorem IV. 97])**.** *Let $Z = (Z_t)_{t\in\mathbb{R}_+}$ be a measurable process on $\Omega = C(\mathbb{R}_+)$. Then Z is $\mathcal{F}^0$-optional iff $Z_t = Z_t\circ a_t$ for all $t\in\mathbb{R}_+$.*

Proof of Proposition 4.4. We will only prove the second assertion; the first one being an obvious consequence.

Set $\Omega' = \Omega \times \mathbb{R}_+$ and $a'(\omega, t) = (a_t(\omega), t)$. Then $Z_t = Z_t \circ a_t$ for all $t \in \mathbb{R}_+$ is equivalent to asserting that $Z = Z \circ a'$. Let S' be the set of all $(\omega, t) \in \Omega'$ for which $a'(\omega, t) = (\omega, t)$ (i.e. ω remains constant from t on). Note that r is a homeomorphism from S' to S and denote its inverse by r^{-1}.

Assume now that Z is an optional process. Then $Z = Z \circ a'$. Since $r = r \circ a'$ we have $Z = Z \circ r^{-1} \circ r \circ a' = (Z \circ r^{-1}) \circ r$. Hence we may take $H = Z \circ r^{-1}$ in Proposition 4.4.

Conversely, if $Z = H \circ r$, then we have $Z \circ a' = H \circ r \circ a' = H \circ r = Z$. Hence Z is optional.

The last assertion of the proposition follows from the identification of S with S'. □

Definition 4.6. *We call a set $D \subseteq S$* right complete *if $(g, t) \in D$ and $(g, t) \prec (f, s)$ implies $(f, s) \in D$. We say $D \subseteq S$ is* left complete *if $(g, t) \in D$ and $(g, t) \succ (f, s)$ implies $(f, s) \in D$.*

Subsequently we will be interested in the *stochastic intervals* $[\![0, \tau]\!]$ for stopping times τ. In particular, recall that $[\![0, \tau]\!] = \{(\omega, t) : t \in [0, \tau(\omega)]\} \subseteq C(\mathbb{R}_+) \times \mathbb{R}_+$. The following lemma connects characterisations of stopping times, sets in S, and stochastic intervals.

Lemma 4.7.
1. *Suppose τ is an $\mathcal{F}^0$-stopping time. Then the set $D = r([\![\tau, \infty[\![) \subseteq S$ satisfies*
   - *(a) D is Borel and right complete;*
   - *(b) if $(f, s) \in D$, the set $\{t : (f_{\restriction[0,t]}, t) \in D\}$ has a smallest element.*

   Moreover, given such a set D, there exists an $\mathcal{F}^0$-stopping time τ determined by $[\![\tau, \infty[\![= r^{-1}(D)$.
2. *Suppose τ is an $\mathcal{F}^+$-stopping time. Then the set $D = r(]\!]\tau, \infty[\![) \subseteq S$ satisfies*
   - *(a) D is Borel, and right complete;*
   - *(b) if $(f, s) \in D$, the set $\{t : (f_{\restriction[0,t]}, t) \in D\}$ has no smallest element.*

   Moreover, given such a set D, there exists an $\mathcal{F}^+$-stopping time τ determined by $]\!]\tau, \infty[\![= r^{-1}(D)$.

Proof.
1. First observe that if we set $\tau(\omega) = \inf\{t \geq 0 : (\omega_{\restriction[0,t]}, t) \in D\}$, it follows that τ is the required $\mathcal{F}^0$-stopping time. On the other hand, if τ is a $\mathcal{F}^0$-stopping time, then D is Borel (by Proposition 4.4), since $[\![\tau, \infty[\![$ is an optional set, and the other properties are straightforward.
2. Observe that if τ is a $\mathcal{F}^+$ stopping time, then $\tau_n = \tau + 1/n$ is a sequence of strictly decreasing $\mathcal{F}^0$-stopping times, and $]\!]\tau, \infty[\![= \bigcup_{n \in \mathbb{N}} [\![\tau_n, \infty[\![$. The conclusions follow from (1). □

By Proposition 4.4 we then have:

Corollary 4.8. *The map r leaves stochastic intervals of $\mathcal{F}^+$-stopping times invariant, i.e. for every $\mathcal{F}^+$-stopping time κ it holds that $r^{-1}(r([\![0, \kappa]\!])) = [\![0, \kappa]\!]$. If κ is an $\mathcal{F}^0$-stopping time then also $r^{-1}(r([\![0, \kappa[\![)) = [\![0, \kappa[\![$.*

Recalling Definition 4.3, we call a martingale $(X_t)_{t \in \mathbb{R}_+}$ a S*- continuous martingale* if it can be written as $X_t(\omega) = h((\omega_{\restriction[0,t]}, t))$ for some $h : S \to \mathbb{R}$, which is continuous.

Definition 4.9. *Let $X : C(\mathbb{R}_+) \to \mathbb{R}$ be a measurable function which is bounded or positive. Then we define $\mathbb{E}[X|\mathcal{F}^0_t]$ to be the unique $\mathcal{F}^0_t$-measurable function satisfying*

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

Proposition 4.10. *Let $X \in C_b(C(\mathbb{R}_+))$. Then $X_t(\omega) := \mathbb{E}[X|\mathcal{F}^0_t](\omega)$ defines a S- continuous martingale. We denote this martingale by X^M.*

Proof. Note that $(f_n, s_n) \to (f, s)$ in S implies $f_n \oplus \omega \to f \oplus \omega$ in $C(\mathbb{R}_+)$ for every $\omega \in C(\mathbb{R}_+)$, where $g \oplus h$ denotes concatenation of paths as usual. Hence, putting $X_g(\omega) := X(g \oplus \omega)$ the convergence $(f_n, s_n) \to (f, s)$ implies the pointwise convergence $X_{f_n}(\omega) \to X_f(\omega)$ for all $\omega \in C(\mathbb{R}_+)$ by continuity of X. Moreover, for $(f, s) \in S$

$$\int X_f(\omega)\, \mathbb{W}(d\omega) =: X^M(f, s)$$

is a function of (f, s). Since X is bounded, this allows to deduce using the dominated convergence theorem that

$$X^M(f_n, s_n) \to X^M(f, s).$$

This means that X^M is continuous on S, hence, S- continuous. □

Proposition 4.11. *Suppose X is a bounded lower semi-continuous function on S. Then there exists a continuous martingale ψ such that $X^M_\tau = \psi_\tau$ almost surely for every $\mathcal{F}^+$-stopping time τ.*

Proof. Since X is lower semi-continuous, we can approximate from below by (bounded) continuous functions. In particular, let $\varphi^n \uparrow X$, and then the corresponding martingales $\varphi^{M,n}$ are S- continuous. In addition, we know that there exists a version of the martingale $(\mathbb{E}[X|\mathcal{F}^0_t])_{t\in\mathbb{R}_+}$ denoted by $(\psi_t)_{t\in\mathbb{R}_+}$ whose paths are almost surely continuous. It follows that $\varphi^{M,n}_\tau \uparrow \psi_\tau$ almost surely, and the claimed result holds. □

4.2. **Randomized stopping times.** Working on the path space $C(\mathbb{R}_+)$, 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 ξ_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 been established previously in [32], and is closely connected to the class of *pseudo-stopping times*, which we will also exploit.

We consider the space

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

where $(\xi_\omega)_{\omega\in\Omega}$ is a disintegration of ξ in the first coordinate $\omega \in \Omega$. We equip M with the weak topology induced by the continuous bounded functions on $C(\mathbb{R}_+) \times \mathbb{R}_+$.

Recall that our principle interest is in the probability space $(\Omega, \mathcal{F}, \mathbb{P})$, where $\Omega = C(\mathbb{R}_+)$ and $\mathbb{P} = \mathbb{W}$. Sometimes we will also consider the associated, right-continuous and complete filtration $(\mathcal{F}^a_t)_{t\geq 0}$. In what follows, we will also use a natural extension of the filtered probability space denoted $(\bar\Omega, \bar{\mathcal{F}}, (\bar{\mathcal{F}}_t)_{t\geq 0}, \bar{\mathbb{P}})$, where we take $\bar\Omega = \Omega \times [0, 1]$, $\bar{\mathcal{F}} = \mathcal{F} \otimes \mathcal{B}([0, 1])$, $\bar{\mathbb{P}}(A \times B) = \mathbb{P}(A)\mathrm{Leb}(B)$, and set $\bar{\mathcal{F}}_t = \mathcal{F}^a_t \otimes \sigma([0, 1])$.

We have the following result characterising the class of *randomized stopping times*.

Theorem 4.12. *Let $\xi \in \mathsf{M}$. Then the following are equivalent:*

(1) *There is a Borel function $H : S \to [0, 1]$ such that H is right-continuous, decreasing and*

$$\xi_\omega([0, s]) := 1 - H(\omega_{\upharpoonright[0,s]}) \tag{4.3}$$

defines a disintegration of ξ w.r.t. to $\mathbb{W}$.

(2) *For every disintegration $(\xi_\omega)_{\omega\in\Omega}$ of ξ, for all $t \in \mathbb{R}_+$ and every Borel set $A \subseteq [0, t]$ the random variable*

$$X_t(\omega) = \xi_\omega(A)$$

is $\mathcal{F}^a_t$-measurable.

(3) *There is a disintegration* $(\xi_\omega)_{\omega\in\Omega}$ *of* ξ *such that for all* $t \in \mathbb{R}_+$ *and all* $f \in C_b(\mathbb{R}_+)$ *such that the support of* f *lies in* $[0,t]$ *the random variable*

$$X_t(\omega) = \xi_\omega(f)$$

is $\mathcal{F}_t^0$*-measurable.*

(4) *On the probability space* $(\bar\Omega, \bar{\mathcal{F}}, (\bar{\mathcal{F}}_t)_{t\geq 0}, \bar{\mathbb{P}})$*, the random time*

$$\rho(\omega, u) = \inf\{t \geq 0 : \xi_\omega([0,t]) \geq u\} \tag{4.4}$$

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

We call a measure $\xi \in \mathsf{M}$ *satisfying one of these conditions a randomized stopping time. We denote the subset of* M *containing the randomized stopping times* RST.

Proof. We show each of the later conditions is equivalent to (1).

We first establish that (1) implies (2). Let ξ_ω and ξ'_ω be disintegrations of ξ. From (1) it follows that there is some function H such that $\xi_\omega([0,s]) = 1 - H(\omega_{\upharpoonright[0,s]})$, $s \leq t$, and $t \mapsto 1 - H(\omega_{\upharpoonright[0,t]})$ is an increasing, càdlàg function for fixed ω. It follows that $\xi_\omega(A) = \int_A d(1 - H(\omega_{\upharpoonright[0,s]}))$ is $\mathcal{F}_t^0$-measurable, and hence, since $\xi'_\omega(A) = \xi_\omega(A)$ $\mathbb{W}$-a.s., $\xi'_\omega(A)$ is $\mathcal{F}_t^a$-measurable. We next establish that (2) implies (1). Consider a disintegration $(\xi_\omega)_{\omega\in\Omega}$ of ξ. Define a process $\bar H$ by

$$\bar H_t(\omega) := 1 - \xi_\omega([0,t]).$$

Then $\bar H$ is càdlàg and therefore $\mathcal{F}^a$-optional, hence $\mathcal{F}^a$-predictable in our setup. By Theorem 4.1 and Proposition 4.4 there exists a Borel function H on S such that $\bar H$ is indistinguishable from $H \circ r$. This function H is as required. Similar arguments establish the equivalence of (1) and (3).

Finally, it is straightforward to deduce (4) from (1). To see (1) given (4), observe that $\xi_t(\omega) = \int \mathbb{1}\{\rho(\omega,u) \leq t\}\, du$ and hence is $\mathcal{F}_t^a$-measurable, càdlàg, and hence (as above) we get the required function H. □

Remark 4.13. (1) *The function* H *in* (4.3) *is unique up to indistinguishability (cf. Remark 4.2). We will designate this function* H^ξ *in the following. This function has a natural interpretation.* $H^\xi(f,s)$ *is the probability that a particle is still alive at time* s *given that it has followed the path* f*. We call* H^ξ *the* survival function *associated to* ξ*.*

(2) *We will say* ξ *is a non-randomized stopping time iff there is a disintegration* $(\xi_\omega)_{\omega\in\Omega}$ *of* ξ *such that* ξ_ω *is 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* H^ξ *which only attains the values* 0 *and* 1.

Corollary 4.14. *The set* RST *is closed.*

Proof. We consider condition (2) resp. (3) in Theorem 4.12; the goal is to express measurability of $X_t(\omega) := \xi_\omega(f)$, $\operatorname{supp} f \subseteq [0,t]$ in a different fashion. Note that a bounded Borel function h is $\mathcal{F}_t^0$-measurable iff for all bounded Borel functions g

$$\mathbb{E}[hg] = \mathbb{E}[h\mathbb{E}[g|\mathcal{F}_t^0]],$$

of course this does not rely on our particular setup. By a functional monotone class argument, for $\mathcal{F}_t^0$-measurability of X_t it is sufficient to check that

$$\mathbb{E}[X_t(g - \mathbb{E}[g|\mathcal{F}_t^0])] = 0 \tag{4.5}$$

for all $g \in C_b(C(\mathbb{R}_+))$. In terms of ξ, (4.5) amounts to

$$0 = \mathbb{E}[X_t(g - \mathbb{E}[g|\mathcal{F}_t^0])] = \int \mathbb{W}(d\omega) \int \xi_\omega(ds) f(s)(g - \mathbb{E}[g|\mathcal{F}_t^0])(\omega) \tag{4.6}$$

$$= \int f(s)(g - \mathbb{E}[g|\mathcal{F}_t^0])(\omega)\, \xi(d\omega, ds) \tag{4.7}$$

which is a closed condition by Proposition 4.10. $\square$

Definition 4.15. *A randomized stopping time is* finite *iff* $\xi(C(\mathbb{R}_+) \times \mathbb{R}_+) = 1$. *The set of all finite randomized stopping times will be denoted by* RST^1.

Recall from (1.5) that $\Gamma^< = \{(f,s) : \exists (g,t) \in \Gamma, s < t, (f,s) = (g_{\upharpoonright[0,s]}, s)\}$ for $\Gamma \subseteq S$.

Lemma 4.16. *Let* $\xi \in \mathsf{RST}^1$. *Then there exists a Borel set* $\Gamma \subseteq S$ *with* $\xi(\Gamma) = 1$ *and* $\Gamma^< \cap \Gamma = \emptyset$ *iff* $\xi = \delta_\tau$ *for some* $\mathcal{F}^a$*-stopping time* τ.

Proof. Let τ be an $\mathcal{F}^a$-stopping time. By Theorem 4.1, there exists an $\mathcal{F}^0$-stopping time τ' with $\tau = \tau'$ $\mathbb{W}$-a.s.. Then $\Gamma = r(\omega, \tau'(\omega))$ satisfies $\Gamma^< \cap \Gamma = \emptyset$ and $\xi = \delta_\tau$ is concentrated on Γ. Here Γ is an analytic set and hence universally measurable. We may thus replace Γ with a Borel subset of full ξ-measure to obtain the desired conclusion.

Pick $\xi \in \mathsf{RST}^1$ and a set Γ on which ξ is concentrated. $\Gamma^< \cap \Gamma = \emptyset$ implies that for any ω the set $\{t : r(\omega,t) \in \Gamma\}$ is at most single-valued. Put $D := \{(g,t) : \exists (f,s) \in \Gamma, (f,s) \prec (g,t)\}$. By Lemma 4.7 this defines an $\mathcal{F}^0$-stopping time on a subset of full measure (recall that ξ is only concentrated on Γ) proving the result. $\square$

Given $\xi \in \mathsf{M}$ and $s \in \mathbb{R}_+$ we define the measure $\xi \wedge s \in \mathsf{M}$ to be the random time which is the minimum of ξ and s; formally this means that for $\omega \in \Omega$ and $A \subseteq \mathbb{R}_+$

$$(\xi \wedge s)_\omega(A) := \xi_\omega(A \cap [0,s)) + \delta_s(A)(1 - \xi_\omega([0,s))),$$

or equivalently, in terms of the survival function:

$$H^{\xi\wedge s}(f,t) = H^\xi(f,t)\mathbb{1}_{\{t<s\}}.$$

Assume that $(M_s)_{s\in\mathbb{R}_+}$ is a process on Ω. Then the stopped process $(M_s^\xi)_{s\in\mathbb{R}_+}$ is defined to be the *probability measure* on $\mathbb{R}$ such that for all bounded and measurable functions f

$$\int_{\mathbb{R}} f(x)\, M_s^\xi(dx) := \int f(M_t(\omega))\,(\xi \wedge s)(d\omega, dt).$$

Otherwise said M_s^ξ is the image measure of $\xi \wedge s$ under the map $M : C(\mathbb{R}_+) \times \mathbb{R}_+ \to \mathbb{R}, (\omega,t) \mapsto M_t(\omega)$. We write $\lim_{s\to\infty} M_s^\xi = M_\xi$ if it exists.

4.3. **Pseudo-randomized stopping times and dual optional projections.** We wish to characterise the subset of M corresponding to RST. A natural candidate for such a condition would be via the optional stopping theorem:

Definition 4.17. *Let* PRST *be the set of all* pseudo-randomized stopping times*, that is, the set of* $\xi \in \mathsf{M}$ *satisfying*

$$\int X(\omega)\,\mathbb{W}(d\omega) = \int X_s^M(\omega)\,(\xi \wedge t)(d\omega, ds), \tag{4.8}$$

for all $t \geq 0$ *and all* $X \in \mathcal{B}_b(C(\mathbb{R}_+))$*, the class of bounded Borel functions on* $C(\mathbb{R}_+)$.

Unfortunately RST is a proper subset of PRST; it is not hard to see this from [33]. By a functional monotone class argument it is sufficient to check (4.8) for all $X \in C_b(C(\mathbb{R}_+))$, in particular we have:

Proposition 4.18. *The set* PRST *is closed.*

Fortunately, the difference between RST and PRST is not seen by optional processes: given a pseudo-randomized stopping time ξ there always exists a randomized stopping time $\bar\xi$ such that for every optional bounded or positive process X we have $\int X\,d\xi = \int X\,d\bar\xi$.

Lemma 4.19. *Let* $\xi \in \mathsf{PRST}$. *Set* $A_t(\omega) := \xi_\omega([0,t])$. *Define* ξ^o *through* $\xi^o_\omega([0,t]) := A^o_t(\omega)$*, where* A^o *denotes the dual optional projection of* A. *Then* $\xi^o \in \mathsf{RST}$.

Proof. We prove this for a finite time ξ. By Theorem 4.12, we have to show that A^o_t is $\mathcal{F}^a_t$-measurable for every t, A^o is increasing nonnegative and bounded by 1. The only property that does not follow directly from the definition of dual optional projection is the boundedness by 1. As $\xi \in \mathsf{PRST}$ we have using $X \equiv 1$

$$\mathbb{E}[A^o_\infty] = 1.$$

Hence, it is sufficient to show that $A^o_\infty \leq 1$. To this end, assume that $D := \{A^o_\infty(\omega) > 1\}$ has positive mass. Then we have using (4.8) and $X = 1_D$

$$\mathbb{W}(D) = \mathbb{E}[X] = \mathbb{E}[X^M_0] = \mathbb{E}_\xi[X^M] = \mathbb{E}\int_0^\infty X^M_s\, dA_s = \mathbb{E}\int_0^\infty X^M_s\, dA^o_s = \mathbb{E}XA^o_\infty > \mathbb{W}(D),$$

implying that $\mathbb{W}(D) = 0$. Hence, $\xi^o \in \mathsf{RST}$. □

Clearly every pseudo-randomized stopping time $\xi \in \mathsf{PRST}$ can be represented as a positive random variable on $\bar{\Omega}$ in a similar manner to (4.4) by taking $\rho(\omega, u) = \inf\{t \geq 0 : \xi_\omega([0,t]) \geq u\}$. The message of the above result is that, for any such ξ, and any optional bounded or positive process X on Ω, there exists a stopping time ρ^o on the extended space $(\bar{\Omega}, \bar{\mathcal{F}}, \bar{\mathcal{F}}_t, \bar{\mathbb{P}})$ such that $\bar{\mathbb{E}}\left[\bar{X}_{\rho\wedge t}\right] = \mathbb{E}\left[X^\xi_t\right] = \mathbb{E}\left[X^{\xi^0}_t\right] = \bar{\mathbb{E}}\left[\bar{X}_{\rho^o\wedge t}\right]$ for any $t \geq 0$, where $\bar{X}(\omega, u) = X(\omega)$. Of course, we will eventually be interested in the subset of stopping times corresponding moreover to (SEP) — that is, we are specifically interested in the subset of PRST which both embed μ, and which satisfy a further natural criteria corresponding to the second condition in (SEP). However, by taking the optional processes $X_t = f(B_t)$ for bounded f and $X_t = t$, we immediately see that $\bar{B}_\rho \sim \bar{B}_{\rho^o}$ and $\mathbb{E}_\xi[T] = \mathbb{E}_{\xi^o}[T] = \bar{\mathbb{E}}[\rho^o]$, where we denote by T the projection

$$T : C(\mathbb{R}_+) \times \mathbb{R}_+ \to \mathbb{R}_+. \tag{4.9}$$

We observe also that to show the process B^ξ_t is uniformly integrable, we need to show $\lim_{R\to\infty} \sup_t \int_{|x|>R} |x|\, B^\xi_t(dx) = 0$. However, with the above definitions, we have $B^\xi_t(dx) = B^{\xi^o}(dx)$, and so B^ξ is uniformly integrable if and only if B^{ξ^o} is also.

From now on we make the assumption that the measure μ which we want to embed has mean 0 and finite second moment[4]

$$V := \int x^2\, \mu(dx) < \infty. \tag{4.10}$$

Then by the above arguments, and as a direct consequence of the same result for the stopping time ρ^o, we have:

Lemma 4.20. *Let $\xi \in \mathsf{PRST}$. Assume that $B_\#\xi = \mu$, i.e. $\bar{B}_\rho \sim \mu$, where ρ is the random time on $\bar{B}$ corresponding to ξ. Then the following are equivalent:*

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

Definition 4.21. *We denote by* $\mathsf{PRST}(\mu)$ *the set of all pseudo-randomized stopping times satisfying the conditions in Lemma* (4.20). *Similarly, we define* $\mathsf{RST}(\mu) = \mathsf{PRST}(\mu) \cap \mathsf{RST}$.

An immediate consequence is:

Corollary 4.22. *Let X_t be an optional process. Then for every $\xi \in \mathsf{PRST}(\mu)$, there exists $\xi^o \in \mathsf{RST}(\mu)$ with $\mathbb{E}[X_\xi] = \mathbb{E}[X_{\xi^o}]$.*

The main reason why we consider *randomized* stopping times and their pseudo-randomized counterparts is that they have the following property:

Theorem 4.23. *The set* $\mathsf{PRST}(\mu)$ *is compact.*

[4]This assumption is only made for ease of exposition. We refer to Section 9 and in particular Proposition 9.3 for the general case.

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

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

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

Hence, $\mathsf{PRST}(\mu)$ is tight.

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

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

Hence, we have $h(\xi) = \mu$. Moreover, the set $\{(\omega, t) : t \leq L\}$ is closed. Hence, by the Portmanteau theorem, for any $L \geq 0$

$$\limsup \xi_n(t \leq L) \leq \xi(t \leq L).$$

This readily implies that $\mathbb{E}_\xi[T] \leq \liminf \mathbb{E}_{\xi_n}[T] = V < \infty$. □

Since RST is a closed set we also have:

Corollary 4.24. *The set* $\mathsf{RST}(\mu)$ *of all randomized stopping times which embed* μ *is compact.*

4.4. **Joinings / Tagged Stopping Times.** We now add another dimension: assume that (Y, ν) is some Polish probability space. The set of all *tagged pseudo-randomized stopping times* or rather *joinings* $\mathsf{JOIN}(\mathbb{W}, \nu) = \mathsf{JOIN}(\nu)$ is given by

$$\Big\{\pi \in \mathcal{P}^{\leq 1}(C(\mathbb{R}_+) \times \mathbb{R}_+ \times Y), \mathsf{proj}_{C(\mathbb{R}_+)\times\mathbb{R}_+}(\pi_{\upharpoonright C(\mathbb{R}_+)\times\mathbb{R}_+\times B}) \in \mathsf{PRST}, B \in \mathcal{B}(Y), \mathsf{proj}_Y(\pi) \leq \nu\Big\}.$$

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

Remark 4.25. *Write* pred *for the* σ*-algebra of* $\mathcal{F}^0$*-predictable sets in* $C(\mathbb{R}_+) \times \mathbb{R}_+$.

We call a set $A \subseteq C(\mathbb{R}_+) \times \mathbb{R}_+ \times Y$ *predictable if it is an element of* $\mathsf{pred} \otimes \mathcal{B}(Y)$. *We will say that a function defined on* $C(\mathbb{R}_+) \times \mathbb{R}_+ \times Y$ *is predictable if it is measurable w.r.t.* $\mathsf{pred} \otimes \mathcal{B}(Y)$.

5. The Optimization Problem and Duality

5.1. **The Primal Problem.**

Recall that our aim is to maximise the value given by functional $\gamma : S \to \mathbb{R}$, where the maximization is taken over randomized stopping times. Each randomized stopping time ξ gives rise to the probability measure $\xi_S := r(\xi)$. Given an $\mathcal{F}^0$-predictable function $\tilde{\gamma}$ on $C(\mathbb{R}_+) \times \mathbb{R}_+$ we can find a Borel function γ on S such that $\gamma \circ r$ is indistinguishable from $\tilde{\gamma}$ and then

$$\int_{C(\mathbb{R}_+)\times\mathbb{R}_+} \tilde{\gamma}(\omega, t) \, \xi(d(\omega, t)) = \int_S \gamma(f, s) \, \xi_S(d(f, s)). \tag{5.1}$$

Where there is no danger of confusion, we will not distinguish between $\tilde{\gamma}$ and γ or ξ and ξ_S.

We assume that there exists at least one $\xi \in \mathsf{RST}(\mu)$ such that

$$\int \gamma(\omega, t) \, d\xi(\omega, t) > -\infty, \tag{5.2}$$

and that the integral in (5.2) is less than ∞ for all $\xi \in \mathsf{RST}(\mu)$. The maximization problem introduced in the introduction, see (OptSEP), can then also be written as

$$P_\gamma(\mathbb{W}, \mu) = \sup\left\{ \int \gamma(\omega, t)\, d\xi(\omega, t), \xi \in \mathsf{RST}(\mu) \right\}. \tag{5.3}$$

We make the important comment that the optimization problem is not altered if the set $\mathsf{RST}(\mu)$ is replaced by $\mathsf{PRST}(\mu)$ (cf. Corollary 4.22).

It is straightforward to see that the functional (5.1) is upper semi-continuous provided that $\gamma : S \to \mathbb{R}$ is (upper semi-) continuous and bounded from above by a constant. (This is spelled out in detail for instance in [52, Chapter 4] in the context of classical optimal transport.) In particular (5.3) then admits an optimizer according to the compactness properties derived above.

Next, assume that γ admits only the bound (1.3), i.e. $\gamma(f, s) \le a + b \cdot s + c \cdot \max_{r \le s} f(s)^2$. Using the pathwise Doob-inequality (see [1])

$$\max_{r \le s} f(r)^2 \le \underbrace{\int_0^s 4 \max_{t \le r} |f(t)|\, df(r)}_{=:M_s} + 4 f(s)^2$$

we obtain that

$$\gamma(f, s) \le a + bs + c \cdot (M_s + 4 f(s)^2).$$

It follows that $\tilde{\gamma}(f, s) := \gamma(f, s) - bs - c \cdot (M_s + 4 f(s)^2)$ is bounded from above and gives rise to the same optimization problem as γ. Hence, applying the above argument to the function $\tilde{\gamma}$, Theorem 1.1 follows also in the general case.

5.2. **The dual problem.**

Theorem 5.1. *Let $\gamma : S \to \mathbb{R}$ be S- upper semi-continuous, bounded from above in the sense of* (1.3)*, i.e. $\gamma(f, s) \le a + b \cdot s + c \cdot \max_{r \le s} f(s)^2$ for some constants $a, b, c \in \mathbb{R}_+$, and predictable. Put*

$$D_\gamma(\mathbb{W}, \mu) = \inf\left\{ \int \psi(y)\, d\mu(y) : \psi \in C_b(\mathbb{R}), \exists \varphi, \begin{array}{l} \varphi \text{ is a } S\text{- continuous martingale}, \varphi_0 = 0 \\ \varphi_t(\omega) + \psi(\omega(t)) \ge \gamma(\omega, t)\ \forall\ (\omega, t) \end{array} \right\}$$

where φ runs through all S- continuous $\mathcal{F}^0$-martingales with $|\varphi_t| \le at + bB_t^2 + c$ for some $a, b, c > 0$ depending on φ. Then we have the duality relation

$$P_\gamma(\mathbb{W}, \mu) = D_\gamma(\mathbb{W}, \mu). \tag{5.4}$$

We note that through Hobson's time change approach ([23, 24]) Theorem 5.1 can be interpreted as super-replication theorem for robust finance. In this sense, Theorem 5.1 is parallel to the work of Dolinsky and Soner [17]. Comparable duality results in a discrete time framework are established by Bouchard and Nutz [8] among others.

Using the same argument as above, we see that it suffices to establish Theorem 5.1 in the case where γ is bounded from above.

As usual the inequality $P_\gamma(\mathbb{W}, \mu) \le D_\gamma(\mathbb{W}, \mu)$ is straightforward to verify:

Lemma 5.2. *With the above notations and assumptions we have $P_\gamma(\mathbb{W}, \mu) \le D_\gamma(\mathbb{W}, \mu)$.*

Proof. Take (φ, ψ) satisfying the dual constraint and $\xi \in \mathsf{RST}(\mu)$. Then we have

$$\int \psi(y)\, \mu(dy) = \int \psi(\omega(t))\, \xi(d\omega, dt) + \int \varphi_t^M(\omega)\, \xi(d\omega, dt) \ge \int \gamma(\omega, t)\, \xi(d\omega, dt),$$

where the inequality holds by the dual constraint. $\square$

The key idea for the proof of Theorem 5.1 is to translate the embedding problem for μ into a transportation problem between the Wiener measure $\mathbb{W}$ and μ using the cost function

$$c(\omega, t, y) = \begin{cases} \gamma(\omega, t) & \text{if } \omega(t) = y \\ -\infty & \text{otherwise,} \end{cases}$$

for $(\omega, t, y) \in C(\mathbb{R}_+) \times \mathbb{R}_+ \times \mathbb{R}$. The result of choosing this special cost function is that $P_\gamma(\mathbb{W}, \mu) = P_c^{C(\mathbb{R}_+)\times\mathbb{R}_+\times\mathbb{R}}(\mathbb{W}, \mu)$, where

$$P_c^{C(\mathbb{R}_+)\times\mathbb{R}_+\times\mathbb{R}}(\mathbb{W}, \mu) = \sup\left\{\int c(\omega, t, y)\, \pi(d\omega, dt, dy), \pi \in \mathsf{JOIN}^1(\mathbb{W}, \mu), \mathbb{E}_\pi[T] \le V\right\}. \quad (5.5)$$

Here T is the projection on $\mathbb{R}_+$, $V = \int x^2 \, \mu(dx)$ and we used $Y = \mathbb{R}$ in the definition of $\mathsf{JOIN}^1(\mathbb{W}, \mu)$ (see Section 4.4).

To see this, define $p(\omega, t, y) = (\omega, t)$. If $\pi \in \mathsf{JOIN}^1(\mathbb{W}, \mu)$ is concentrated on $\{(\omega, t, y) : \omega(t) = y\}$ we have $\xi = p(\pi) \in \mathsf{PRST}(\mu)$ and $\int c \, d\pi = \int \gamma \, d\xi$.

On the other hand, let $h(\omega, t) = \omega(t)$. If $\xi \in \mathsf{PRST}(\mu)$ then $\pi = (id, h)(\xi) \in \mathsf{JOIN}^1(\mathbb{W}, \mu)$ and as before $\int c \, d\pi = \int \gamma \, d\xi$.

In Proposition 5.6 we will establish a dual problem corresponding to $P_c^{C(\mathbb{R}_+)\times\mathbb{R}_+\times\mathbb{R}}(\mathbb{W}, \mu)$ and Theorem 5.1 will then be a simple consequence. However we need some preparation before we can establish Proposition 5.6.

5.3. **A Non-Adapted (NA) Duality Result.**

We first prove a "non-adapted" version of the desired result and afterwards we use the min-max theorem (Theorem 5.4) to introduce adaptedness. To this end, put

$$\mathsf{TM}^V(\mathbb{W}, \mu) = \{\pi \in \mathcal{P}(C(\mathbb{R}_+) \times \mathbb{R}_+ \times \mathbb{R}) : \mathsf{proj}_{C(\mathbb{R}_+)}(\pi) = \mathbb{W}, \mathsf{proj}_{\mathbb{R}}(\pi) = \mu, \mathbb{E}_\pi[T] \le V\},$$

and

$$\mathsf{DC}^V_{NA}(c) = \left\{(\varphi, \psi) \in C_b(\Omega) \times C_b(\mathbb{R}) : \exists \alpha \ge 0, \begin{array}{l} \varphi(\omega) + \psi(y) + \alpha(t - V) \ge c(\omega, t, y) \\ \text{for all } \omega \in \Omega, y \in \mathbb{R}, t \ge 0 \end{array}\right\}.$$

Note that the set TM^V is compact as a consequence of Prohorov's theorem.

Proposition 5.3. *Let $c : C(\mathbb{R}_+) \times \mathbb{R}_+ \times \mathbb{R} \to \mathbb{R} \cup \{-\infty\}$ be upper semi-continuous and bounded from above. Then*

$$P_c^{NA} = \sup_{\pi \in \mathsf{TM}^V(\mathbb{W},\mu)} \int c \, d\pi = \inf_{(\varphi,\psi)\in \mathsf{DC}^V_{NA}(c)} \mathbb{W}(\varphi) + \mu(\psi) = D_c^{NA}. \quad (5.6)$$

Again it is easy to show that $D_c^{NA} \ge P_c^{NA}$ (cf. Lemma 5.2). To show the other inequality we first collect some ingredients which will also be useful later on. In particular, we will use the min-max theorem in the following form.

Theorem 5.4 (see e.g. [49, 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 concave on K for every $y \in L$,
(3) $F(x, \cdot)$ is convex on L for every $x \in K$

then

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

Lemma 5.5. *If* (5.6) *is valid for a sequence of continuous bounded functions $c_n, n \ge 1$ such that $c_n \downarrow c$ then* (5.6) *applies also to c.*

Proof. To keep track of the different cost functions we write

$$P_{c_n}^{NA} = \sup_{\mathsf{TM}^V(\mathbb{W},\mu)} \int c_n \, d\pi \quad \text{and} \quad D_{c_n}^{NA} = \inf_{(\varphi,\psi)\in\mathsf{DC}^V_{NA}(c_n)} \left(\mathbb{E}_{\mathbb{W}}[\varphi] + \mathbb{E}_\mu[\psi]\right),$$

where $\mathsf{DC}^V_{NA}(c_n)$ is to remind us on the dependence of the dual constraint set on c_n. P_c^{NA} and D_c^{NA} are defined analogously. We have to prove that $D_c^{NA} \le P_c^{NA}$. For each k let $\pi_k \in \mathsf{TM}^V(\mathbb{W}, \mu)$ be such that

$$P_{c_k}^{NA} \le \int c_k \, d\pi_k + 1/k.$$

By compactness of $\mathsf{TM}^V(\mathbb{W},\mu)$ there is a subsequence, still denoted by k, such that $(\pi_k)_k$ converges weakly to some $\pi \in \mathsf{TM}^V(\mathbb{W},\mu)$. Then by monotone convergence using the monotonicity of the sequence $(c_k)_{k\in\mathbb{N}}$ we have

$$P_c^{NA} \geq \int c\,d\pi = \lim_{m\to\infty}\int c_m\,d\pi = \lim_{m\to\infty}\left(\lim_{k\to\infty}\int c_m\,d\pi_k\right)$$
$$\geq \lim_{m\to\infty}\left(\lim_{k\to\infty}\int c_k\,d\pi_k\right) = \lim_{k\to\infty} P_{c_k}^{NA}.$$

Since, $c_k \geq c$ implies $P_{c_k}^{NA} \geq P_c^{NA}$ and $D_c^{NA} \leq D_{c_k}^{NA}$ this allows us to deduce that

$$D_c^{NA} \leq D_{c_k}^{NA} = P_{c_k}^{NA} \searrow P_c^{NA}. \qquad \square$$

Proof of Proposition 5.3. We may assume that c is bounded from above by zero. Hence, by Lemma 5.5 it is sufficient to establish (5.6) for bounded continuous functions whose support satisfies

$$\operatorname{supp} c \subseteq C(\mathbb{R}_+)\times[0,t_0]\times\mathbb{R} \tag{5.7}$$

for some $t_0 \in \mathbb{R}_+$. Put

$$\mathsf{TM}^V_{t_0}(\mathbb{W},\mu) = \{\pi \in \mathsf{TM}^V(\mathbb{W},\mu) : \operatorname{supp}\pi \subseteq C(\mathbb{R}_+)\times[0,t_0]\times\mathbb{R}\},$$

and

$$\mathsf{DC}^V_{NA,t_0}(c) = \left\{(\varphi,\psi)\in C_b(\Omega)\times C_b(\mathbb{R}) : \exists\alpha\geq 0,\ \begin{array}{l}\varphi(\omega)+\psi(y)+\alpha(t-V)\geq c(\omega,t,y)\\ \text{for all } \omega\in\Omega, y\in\mathbb{R}, t\leq t_0\end{array}\right\}.$$

Assume now that c satisfies (5.7) for some $t_0 \geq V$. We then have

$$\sup_{\pi\in\mathsf{TM}^V(\mathbb{W},\mu)}\int c\,d\pi = \sup_{\pi\in\mathsf{TM}^V_{t_0}(\mathbb{W},\mu)}\int c\,d\pi \quad\text{and} \tag{5.8}$$

$$\inf_{(\varphi,\psi)\in\mathsf{DC}^V_{NA}(c)}\mathbb{W}(\varphi)+\mu(\psi) = \inf_{(\varphi,\psi)\in\mathsf{DC}^V_{NA,t_0}(c)}\mathbb{W}(\varphi)+\mu(\psi). \tag{5.9}$$

Formally the conditions involving V disappear in $\mathsf{TM}^V_{t_0}(\mathbb{W},\mu)$ and $\mathsf{DC}^V_{NA,t_0}(c)$ if we put $V=\infty$, we therefore define $\mathsf{TM}^\infty_{t_0}(\mathbb{W},\mu)$ and $\mathsf{DC}^\infty_{NA,t_0}(c)$ through

$$\mathsf{TM}^\infty_{t_0}(\mathbb{W},\mu) := \{\pi : \mathsf{proj}_{C(\mathbb{R}_+)}(\pi)=\mathbb{W}, \mathsf{proj}_{\mathbb{R}}(\pi)=\mu, \operatorname{supp}\pi\subseteq C(\mathbb{R}_+)\times[0,t_0]\times\mathbb{R}\},$$
$$\mathsf{DC}^\infty_{NA,t_0}(c) := \{(\varphi,\psi)\in C_b(\Omega)\times C_b(\mathbb{R}) :\ \varphi(\omega)+\psi(y)\geq c(\omega,t,y) \text{ for } t\leq t_0, y\in\mathbb{R}\ \omega\in\Omega\}.$$

As a consequence of the classical Monge-Kantorovich duality, Theorem 3.1, we have (see Corollary 3.3)

$$\sup_{\pi\in\mathsf{TM}^\infty_{t_0}(\mathbb{W},\mu)}\int\tilde c\,d\pi = \inf_{(\varphi,\psi)\in\mathsf{DC}^\infty_{NA,t_0}(\tilde c)}\mathbb{W}(\varphi)+\mu(\psi) \tag{5.10}$$

for $\tilde c$ upper semi-continuous and bounded from above. Using the min-max theorem (Theorem 5.4) with the function

$$F(\pi,\alpha) = \int c-\alpha(t-V)\,d\pi$$

for $\pi\in\mathsf{TM}^\infty_{t_0}(\mathbb{W},\mu)$ and $\alpha\geq 0$ we thus obtain

$$\sup_{\pi\in\mathsf{TM}^V_{t_0}(\mathbb{W},\mu)}\int c\,d\pi = \sup_{\pi\in\mathsf{TM}^\infty_{t_0}(\mathbb{W},\mu)}\int c\,d\pi + \inf_{\alpha\geq 0}(-\alpha)\int t-V\,d\pi$$
$$= \inf_{\alpha\geq 0}\sup_{\pi\in\mathsf{TM}^\infty_{t_0}(\mathbb{W},\mu)}\int c-\alpha(t-V)\,d\pi \tag{5.11}$$
$$= \inf_{\alpha\geq 0}\inf_{(\varphi,\psi)\in\mathsf{DC}^\infty_{NA,t_0}(c-\alpha(t-V))}\mathbb{W}(\varphi)+\mu(\psi) \tag{5.12}$$
$$= \inf_{(\varphi,\psi)\in\mathsf{DC}^V_{NA,t_0}(c)}\mathbb{W}(\varphi)+\mu(\psi),$$

where we have applied (5.10) to the function $\tilde{c} = c - \alpha(t - V)$ to establish the equality between (5.11) and (5.12). This concludes the proof. □

5.4. **Introducing Adaptedness.**

Using the defining property of PRST, we are able to test the "adaptedness" of a measure $\pi \in \mathcal{P}(C(\mathbb{R}_+) \times \mathbb{R}_+ \times \mathbb{R})$ by testing against martingales: Set

$$\mathsf{JOIN}^V(\mathbb{W}, \mu) = \mathsf{JOIN}^1(\mathbb{W}, \mu) \cap \mathsf{TM}^V(\mathbb{W}, \mu).$$

For a continuous and bounded function $f : C(\mathbb{R}_+) \to \mathbb{R}_+$ we consider the S- *continuous* martingale f^M as in Proposition 4.10. Then $\pi \in \mathsf{TM}^V(\mathbb{W}, \mu)$ satisfies $\pi \in \mathsf{JOIN}^V(\mathbb{W}, \mu)$ if and only if for all continuous bounded functions $f : C(\mathbb{R}_+) \to \mathbb{R}, g : \mathbb{R} \to \mathbb{R}$

$$\int f g \, d\pi = \int f^M g \, d\pi. \tag{5.13}$$

This is a direct consequence of the definition of PRST and JOIN, see Definition 4.17 and Section 4.4.

Consider now the following set of dual candidates:

$$\mathsf{DC}^V(c) = \left\{ (\varphi, \psi) : \begin{array}{l} \varphi \text{ is an } S\text{- continuous bounded martingale, } \psi \in C_b(\mathbb{R}), \ \exists \alpha \geq 0, \\ \varphi_t(\omega) + \psi(y) + \alpha(t - V) \geq c(\omega, t, y), \ \text{for all } \omega \in \Omega, y \in \mathbb{R}, t \in \mathbb{R}_+ \end{array} \right\}.$$

Then we can derive the following, adapted version of Proposition 5.3.

Proposition 5.6. *Let $c : C(\mathbb{R}_+) \times \mathbb{R}_+ \times \mathbb{R} \to \mathbb{R} \cup \{-\infty\}$ be upper semi-continuous, predictable (cf. Remark 4.25) and bounded from above. Then,*

$$P_c^{C(\mathbb{R}_+)\times\mathbb{R}_+\times\mathbb{R}}(\mathbb{W}, \mu) = \sup_{\pi \in \mathsf{JOIN}^V(\mathbb{W},\mu)} \int c \, d\pi = \inf_{(\varphi,\psi)\in \mathsf{DC}^V(c)} \mathbb{W}(\varphi) + \mu(\psi) =: D_c^{C(\mathbb{R}_+)\times\mathbb{R}_+\times\mathbb{R}}(\mathbb{W}, \mu).$$

Proof. Let us start with the case that c is continuous and bounded. The general case will follow by approximation, cf. Lemma 5.5. For the S- continuous martingale induced by a continuous and bounded function f we recall the notation f_t^M introduced in Proposition 4.10. We want to use the min-max theorem, Theorem 5.4, with the function

$$F(\pi, h) = \int c + \bar{h} \, d\pi$$

for $\pi \in \mathsf{TM}^V(\mathbb{W}, \mu)$ and

$$h(\omega, y) = \sum_{i=1}^n f_i(\omega) g_i(y), \quad \bar{h}(\omega, t, y) = \sum_{i=1}^n (f_i(\omega) - f_i^M(\omega, t)) g_i(y), \tag{5.14}$$

where $n \in \mathbb{N}$, $f_i \in C_b(C(\mathbb{R}_+))$, $g_i \in C_b(\mathbb{R}_+)$.

The set $\mathsf{TM}^V(\mathbb{W}, \mu)$ is convex and compact by Prohorov's theorem and the set of all h of the form (5.14) is convex as well. Then we have

$$\begin{aligned} P_c^{C(\mathbb{R}_+)\times\mathbb{R}_+\times\mathbb{R}} &= \sup_{\pi \in \mathsf{JOIN}^V(\mathbb{W},\mu)} \int c \, d\pi \\ &= \sup_{\pi \in \mathsf{TM}^V(\mathbb{W},\mu)} \inf_h \left(\int c + \bar{h} \, d\pi \right) \\ &\overset{\text{Thm. 5.4}}{=} \inf_h \sup_{\pi \in \mathsf{TM}^V(\mathbb{W},\mu)} \left(\int c + \bar{h} \, d\pi \right) \\ &= \inf_h \inf_{(\varphi,\psi)\in \mathsf{DC}^V_{NA}(c+\bar{h})} (\mathbb{W}(\varphi) + \mu(\psi)), \end{aligned}$$

where the last equality holds by Proposition 5.3.

We write $c_h = c + \bar{h}$. For $(\varphi, \psi) \in \mathsf{DC}^V_{NA}(c_h)$ there is some $\alpha \geq 0$ such that

$$c_h(\omega, t, y) \leq \varphi(\omega) + \psi(y) + \alpha(t - V).$$

Taking conditional expectations w.r.t. $\mathcal{F}_t^0$ in the sense of Definition 4.9 we obtain

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

for all $\omega \in \Omega, t \in \mathbb{R}_+, y \in \mathbb{R}$ since c is predictable. This implies that $(\varphi_t^M, \psi) \in \mathsf{DC}^V(c)$. Because $\mathbb{W}(\varphi_t^M) = \mathbb{W}(\varphi)$ this implies that $\mathsf{DC}_{NA}^V(c_h) \subseteq \mathsf{DC}^V(c)$. Therefore, we have

$$\begin{aligned} P_c^{C(\mathbb{R}_+)\times\mathbb{R}_+\times\mathbb{R}} &= \inf_h \inf_{(\varphi,\psi)\in\mathsf{DC}_{NA}^V(c+\bar{h})} (\mathbb{W}(\varphi) + \mu(\psi)) \\ &\geq \inf_{(\varphi,\psi)\in\mathsf{DC}^V(c)} (\mathbb{W}(\varphi) + \mu(\psi)) = D_c^{C(\mathbb{R}_+)\times\mathbb{R}_+\times\mathbb{R}}. \end{aligned} \tag{5.15}$$

As usual, the other inequality is straightforward. □

Proof of Theorem 5.1. At the start of the section, we showed that $P_\gamma(\mathbb{W}, \mu) = P_c^{C(\mathbb{R}_+)\times\mathbb{R}_+\times\mathbb{R}}(\mathbb{W}, \mu)$ if we set

$$c(\omega, t, y) = \begin{cases} \gamma(\omega, t) & \text{if } \omega(t) = y \\ -\infty & \text{else.} \end{cases}$$

Moreover, as γ was assumed to be upper semi-continuous, also c is upper semi-continuous. Indeed, take any sequence (ω_n, t_n, y_n) converging to (ω, t, y). If $\limsup_n c(\omega_n, t_n, y_n) = -\infty$ there is nothing to prove. On the other hand, if $\limsup_n c(\omega_n, t_n, y_n) > -\infty$ there is a subsequence $(\omega_{n_k}, t_{n_k}, y_{n_k})$ with $\omega_{n_k}(t_{n_k}) = y_{n_k}$ converging to some (ω, t, y). Then we necessarily have that $\omega(t) = y$ because $|\omega(t) - y| \leq |\omega(t) - \omega_{n_k}(t_{n_k})| + |y_{n_k} - y|$. Thus the upper semi-continuity of c follows from the upper semi-continuity of γ.

Hence, by Proposition 5.6, to see that

$$P_\gamma(\mathbb{W}, \mu) \geq D_\gamma(\mathbb{W}, \mu)$$

it remains to show that $D_c^{C(\mathbb{R}_+)\times\mathbb{R}_+\times\mathbb{R}}(\mathbb{W}, \mu) \geq D_\gamma(\mathbb{W}, \mu)$. A bounded pair (φ, ψ) belongs to $\mathsf{DC}^V(c)$ iff there is $\alpha \geq 0$ such that for all $\omega \in \Omega, y \in \mathbb{R}, t \in \mathbb{R}_+$

$$\varphi_t(\omega) + \psi(y) + \alpha(t - V) \geq c(\omega, t, y)$$

which holds iff

$$\varphi_t(\omega) + \psi(\omega(t)) + \alpha(t - V) \geq \gamma(\omega, t).$$

This is trivially equivalent to

$$[\varphi_t(\omega) - \alpha(\omega(t)^2 - t)] + [\psi(\omega(t)) + \alpha\omega(t)^2 - \alpha V] \geq \gamma(\omega, t). \tag{5.16}$$

The alternative representation in (5.16) is useful to us since $\omega(t)^2 - t$ is an S- continuous martingale.

Putting

$$\bar{\varphi}_t(\omega) = \varphi_t(\omega) - \alpha(\omega(t)^2 - t) \quad \text{ and } \quad \bar{\psi}(y) = \psi(y) + \alpha y^2 - \alpha V,$$

we have $\bar{\varphi}_t(\omega) + \bar{\psi}(\omega(t)) \geq \gamma(\omega, t)$. This means that $(\bar{\varphi} - \bar{\varphi}_0, \bar{\psi} + \bar{\varphi}_0)$ satisfy the constraint in the dual problem in (5.4). Recalling that V was defined by $V = \int y^2 \mu(dy)$ we have $\int \bar{\psi}(y)\, \mu(dy) = \int \psi(y)\, \mu(dy)$. Therefore, we can conclude that

$$D_c^{C(\mathbb{R}_+)\times\mathbb{R}_+\times\mathbb{R}}(\mathbb{W}, \mu) \geq D_\gamma(\mathbb{W}, \mu).$$ □

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

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

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

$$P_\gamma(\mathbb{W}_\lambda, \mu) = \sup\left\{ \int \tilde{\gamma}(\omega, t)\xi(d\omega, dt), \xi \in \mathsf{RST}(\lambda, \mu) \right\}, \tag{5.17}$$

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

Theorem 5.7. *Let $\gamma : \tilde{S} \to \mathbb{R}$ be $\tilde{S}$- upper semi-continuous, bounded from above and predictable. Put*

$$D_\gamma(\mathbb{W}_\lambda, \mu) = \inf\left\{ \int \psi(y)\, d\mu(y) : \psi \in C_b(\mathbb{R}), \begin{array}{l} \exists\, \tilde{S}\text{- continuous mart. } \varphi, \mathbb{E}_{\mathbb{W}_\lambda}[\varphi_0] = 0, \\ \varphi_t(\omega) + \psi(\omega(t)) \geq \gamma(\omega, t), t \in \mathbb{R}_+, \omega \in \tilde{\Omega} \end{array} \right\}$$

where φ runs through all $\tilde{S}$- continuous $\mathcal{F}^0$-martingales with $|\varphi_t| \leq at + bB_t^2 + c$ for some $a, b, c > 0$ depending on φ. Then we have the duality relation

$$P_\gamma(\mathbb{W}_\lambda, \mu) = D_\gamma(\mathbb{W}_\lambda, \mu). \tag{5.18}$$

More generally, the result still holds if γ is only $\tilde{S}$- upper semi-continuous and predictable, and $D_\gamma(\mathbb{W}_\lambda, \mu) < \infty$.

The proof goes along the same lines as the proof of Theorem 5.1. The inequality $P_\gamma(\mathbb{W}_\lambda, \mu) \leq D_\gamma(\mathbb{W}_\lambda, \mu)$ is straightforward. For the other direction we can use the same argument as before. However, we have to replace $\mathbb{W}$ by $\mathbb{W}_\lambda$ and V by $\tilde{V} := V - V_\lambda$. Up to equation (5.16) everything can be copied line to line. Then we have to use the decomposition

$$\begin{aligned} &\varphi_t(\omega) + \psi(\omega(t)) + \alpha(t - V + V_\lambda) \\ &= [\,\varphi_t(\omega) - \alpha(\omega(t)^2 - t - V_\lambda)] + [\psi(\omega(t)) + \alpha(\omega(t)^2 - V)] \end{aligned}$$

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

6. Bad Pairs and Closed Stochastic Intervals

In Section 7 we will assume that τ is a non-randomized, bounded stopping time such that $[\![0, \tau]\!]$ is closed[5]. Then the set

$$\mathsf{M}_\tau := \{\xi \in \mathsf{M} : \xi(]\!]\tau, \infty]\!]) = 0\} \tag{6.1}$$

is compact as a consequence of Prohorov's theorem. We also let $\mathsf{RST}_\tau = \mathsf{RST} \cap \mathsf{M}_\tau$ and $\mathsf{PRST}_\tau = \mathsf{PRST} \cap \mathsf{M}_\tau$. Since RST_τ and PRST_τ are closed we have the following result:

Lemma 6.1. *Let τ be a bounded stopping time such that $[\![0, \tau]\!]$ is closed. Then RST_τ as well as PRST_τ is compact in the topology induced by the continuous bounded functions on $C(\mathbb{R}_+) \times \mathbb{R}_+$.*

Recall the definition of joinings in Section 4.4. Then the *joinings before τ* is the set

$$\mathsf{JOIN}(\tau, \nu) = \left\{\pi \in \mathsf{JOIN}(\mathbb{W}, \nu) : \mathsf{proj}_{C(\mathbb{R}_+)\times\mathbb{R}_+}(\pi) \in \mathsf{PRST}_\tau\right\}. \tag{6.2}$$

We make the important comment that all the results involving $\mathsf{JOIN}(\tau, \nu)$ do not change if we require $\mathsf{proj}_{C(\mathbb{R}_+)\times\mathbb{R}_+}(\pi) \in \mathsf{RST}_\tau$ in (6.2) by an application of Lemma 4.19.

We make a straightforward observation:

Lemma 6.2. *Under the above assumptions, the set $\mathsf{JOIN}(\tau, \nu)$ of tagged pseudo-random times / joinings before τ is compact with respect to the topology coming from the continuous bounded functions on $C(\mathbb{R}_+) \times \mathbb{R}_+ \times \mathbb{R}$.*

[5]We emphasize that this means that $[\![0, \tau]\!]$ is closed as a subset of $C(\mathbb{R}_+) \times \mathbb{R}_+$.

In the following, ν will always denote an optimizer of the optimization problem (5.3).

The notion BP introduced in Definition 2.1 requires that *all* possible extensions (h, u) are considered. In this section we consider also a relaxed notion which is sensitive to the stopping measure ν. To this end we introduce the *conditional randomized stopping time given* (f, s).

Definition 6.3. *Let $\xi \in \mathsf{RST}$ be given and consider the survival function H^ξ as in Remark 4.13. The conditional randomized stopping time of ξ, given $(f, s) \in S$, denoted by $\xi^{(f,s)}$, is defined to be*

$$\xi^{(f,s)}_\omega([0, t]) := \frac{1}{H^\xi(f, s)} \Big(H^\xi(f, s) - H^\xi(f \oplus \omega_{\restriction[0,t]}, s + t) \Big), \tag{6.3}$$

if $H^\xi(f, s) > 0$ and 0 otherwise.

This 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). Note that we can equivalently write

$$\xi^{(f,s)}_\omega([0, t]) = \frac{1}{H^\xi(f, s)} \Big(\xi_{f\oplus\omega}([0, t + s]) - \xi_{f\oplus\omega}([0, s]) \Big).$$

Definition 6.4. *The set of* bad pairs relative to ν *is defined by*

$$\mathsf{BP}_\nu = \Big\{ ((f, s), (g, t)) : f(s) = g(t), \tag{6.4}$$

$$\int \gamma(f \oplus \omega_{\restriction[0,r]}, s + r)\, d\nu^{(f,s)}(\omega, r) + \gamma(g, t) < \gamma(f, s) + \int \gamma(g \oplus \omega_{\restriction[0,r]}, t + r)\, d\nu^{(f,s)}(\omega, r) \Big\}.$$

The interpretation of BP_ν is that in average it is better to stop at (f, s), chop off the "bush" and transfer it onto the "stub" (g, t).

The following result constitutes an important intermediate step towards Theorem 1.2. In the formulation as well as in the proof we interpret the space $(C(\mathbb{R}_+)\times\mathbb{R}_+)\times(C(\mathbb{R}_+)\times\mathbb{R}_+)$ as a product $\mathcal{X} \times \mathcal{Y}$ so that we can make sense of the projections $\mathsf{proj}_\mathcal{X}$ and $\mathsf{proj}_\mathcal{Y}$.

Proposition 6.5. *Let ν be a randomized stopping time which maximizes* (5.3) *for a Borel measurable function $\gamma : S \to \mathbb{R}$. Then $(\mathcal{Y}, \nu) = (C(\mathbb{R}_+) \times \mathbb{R}_+, \nu)$ is a Polish probability space. Assume that $\pi \in \mathsf{JOIN}(\tau, \nu)$ (where τ can be arbitrary) satisfies*

$$H^{\mathsf{proj}_\mathcal{X}(\pi)}(f, s) > 0 \quad \Longrightarrow \quad H^\nu(f, s) > 0 \quad \textit{for } (f, s) \in S. \tag{6.5}$$

Then we have $\pi(\mathsf{BP}_\nu) = 0$.

The interpretation of (6.5) is that if a particle has a strictly positive chance to be alive under $\mathsf{proj}_\mathcal{X}(\pi)$ then the probability that this particle is still alive under ν is positive as well.

Proof. Note that, given $\nu' \in \mathsf{RST}(\mu')$ and $\nu'' \in \mathsf{RST}(\mu'')$, we have that $(\nu' + \nu'')/2 \in \mathsf{RST}((\mu' + \mu'')/2)$. The probabilistic interpretation of this easy fact goes by visualizing the random stopping time $(\nu' + \nu'')/2$ as flipping a coin at time $t = 0$ and subsequently either applying the randomized stopping rule ν' or ν''.

Working towards a contradiction we assume that there is $\pi \in \mathsf{JOIN}(\tau, \nu)$ such that $\pi(\mathsf{BP}_\nu) > 0$. By looking at $\bar\pi := \pi_{\restriction \mathsf{BP}_\nu}$ we can assume that π is concentrated on BP_ν. As BP_ν is predictable (recall Remark 4.25) we can assume that $\mathsf{proj}_\mathcal{X}(\pi) \in \mathsf{RST}_\tau$. Set $\nu_0 = \nu_1 := \nu$. We then use π to define two modifications ν_0^π and ν_1^π of ν such that the following hold true:

(1) The terminal distributions μ_0, μ_1 corresponding to ν_0^π and ν_1^π satisfy $(\mu_0+\mu_1)/2 = \mu$.
(2) ν_0^π stops paths earlier than $\nu_0 = \nu$ while ν_1^π stops later than $\nu_1 = \nu$.
(3) The cost of ν_0^π plus the cost of ν_1^π is bigger than twice the cost of ν, i.e.

$$\int \gamma(\omega_{\restriction[0,t]}, t)\, d\nu_0^\pi(\omega, t) + \int \gamma(\omega_{\restriction[0,t]}, t)\, d\nu_1^\pi(\omega, t) > 2 \int \gamma(\omega_{\restriction[0,t]}, t)\, d\nu(\omega, t).$$

More formally, (2) asserts that for every $s \geq 0$,

$$(\nu_0^\pi)_\omega[0,s] \geq \nu_\omega[0,s], \text{ a.s.} \tag{6.6}$$

$$\text{and } (\nu_1^\pi)_\omega[0,s] \leq \nu_\omega[0,s], \text{ a.s.}, \tag{6.7}$$

where $\nu_{\omega\in\Omega}, (\nu_0^\pi)_{\omega\in\Omega}, (\nu_1^\pi)_{\omega\in\Omega}$ are disintegrations of $\nu_0, \nu_0^\pi, \nu_1^\pi$ respectively w.r.t. $\mathbb{W}$.

If we are able to construct such a pair ν_0^π, ν_1^π, then $(\nu_0^\pi + \nu_1^\pi)/2$ is a randomized stopping time in $\mathsf{RST}(\mu)$ which is strictly better than ν and therefore yields the desired contradiction.

To define ν_0^π, we first consider $\rho_0 = \mathsf{proj}_X(\pi)$ which is a randomized stopping time. As in Remark 4.13 we can view ρ_0 as right-continuous decreasing survival function $H^{\rho_0} : S \to [0,1]$ which starts at 1. It is possible that ρ_0 does not decrease to 0 since we allow particles to survive until ∞.

We now define the randomized stopping time ν_0^π as the product

$$H^{\nu_0^\pi}(f,s) := H^{\rho_0}(f,s) \cdot H^\nu(f,s).$$

The probabilistic interpretation of this definition is that a particle is stopped by ν_0^π if it is stopped by ρ_0 or stopped by ν, where these events are taken to be conditionally independent given the path $\omega \in C(\mathbb{R}_+)$. Comparing ν_0 and ν_0^π the latter will stop some particles earlier than the first one. We note that this in particular implies that $\mathbb{E}_{\nu_0^\pi}[T] \leq \mathbb{E}_\nu[T] < \infty$, where T is the projection from $C(\mathbb{R}_+) \times \mathbb{R}_+$ onto $\mathbb{R}_+$ as defined in (4.9). Also clearly, $\nu_0^\pi \in \mathsf{RST}$, i.e. ν_0^π inherits adaptivity from ρ_0 and ν. Equivalently we can define ν_0^π by setting for $A \subseteq C(\mathbb{R}_+) \times \mathbb{R}_+$

$$\nu_0^\pi(A) = \int_A H^\nu(\omega,t)\, d\rho_0(\omega,t) + \int_A H^{\rho_0}(\omega,t)\, d\nu(\omega,t).$$

Let us now turn to the definition of ν_1^π. For $A \subseteq C(\mathbb{R}_+) \times \mathbb{R}_+$ we define

$$\rho_1(A) \;=\; \int_{S\times A} H^\nu(f,s)\, d\pi((f,s),(g,t)).$$

Fix an $\mathcal{F}^0$-measurable disintegration $(\nu_\omega)_{\omega\in C(\mathbb{R}_+)}$ of ν by (1) of Theorem 4.12. Given $(f,s) \in S$ and $(\omega,t) \in C(\mathbb{R}_+) \times \mathbb{R}_+$ we define a measure on $\mathbb{R}$ with support in $[t,\infty)$ by setting for $A \subseteq [t,\infty)$

$$\nu_{(f,s),(\omega,t)}(A) \;:=\; \nu_{f\oplus\theta_t(\omega)}(A-t+s) \;=\; H^\nu(f,s)\, \nu^{(f,s)}_{\theta_t(\omega)}(A-t+s), \tag{6.8}$$

where $\theta_t(\omega) = (\omega_{s+t} - \omega_t)_{s\geq 0}$. Note that this is a slight generalization of a conditional randomized stopping time, see (6.3). Here we additionally allow a shift of the time parameter and do not normalize (hence the additional factor $H^\nu(f,s)$). This is necessary as in the next step – for defining ν_1^π – we need to trim bushes; i.e. we need to cut some paths at time s and plant them on a stub at time t. Additionally, we can only move the mass that is present which accounts for the $H^\nu(f,s)$ appearing in (6.8) and the definition of ρ_1. Moreover, note that for a set $A \subseteq (t, t+u)$ given (f,s) the map

$$(\omega,t) \mapsto \nu_{(f,s),(\omega,t)}(A)$$

is $\sigma(\omega_l, t \leq l \leq u)$-measurable.

As discussed above, randomized stopping times can be represented either as probability measures on $C(\mathbb{R}_+)\times\mathbb{R}_+$ or as probability measures on S. Formally the tagged random time π is a measure on $(C(\mathbb{R}_+)\times\mathbb{R}_+) \times (C(\mathbb{R}_+)\times\mathbb{R}_+)$. However in defining ν_1^π we consider π as a probability on $S \times (C(\mathbb{R}_+)\times\mathbb{R}_+)$ which, formally, we could do by defining the measure $\tilde{\pi} := (r \times \mathrm{Id})(\pi)$, where r is the natural "projection" from $C(\mathbb{R}_+)\times\mathbb{R}_+$ to S.

We define the probability measure ν_1^π on $C(\mathbb{R}_+)\times\mathbb{R}_+$ by

$$\nu_1^\pi(B) = \nu_1(B) - \rho_1(B) + \int_{S\times B} \nu_{(f,s),(\omega,t)}(B_\omega)\, d\pi((f,s),(\omega,t)),$$

where $B_\omega = \{t \in \mathbb{R}_+ : (\omega,t) \in B\}$. The interpretation of this definition is the following. The support of the randomized stopping time ν can be thought as a tree. The joining π defines a plan how to trim the tree, i.e. cut a bush at position (f,s) and plant it on top of (g,t). Hence, we take the tree, ν, prepare the position where something will be newly planted, subtract ρ_1 which takes away some mass, and plant as much as possible on these stubs to end up with a tree of mass one again. Due to the measurability properties of $(\omega,t)\mapsto \nu^{(f,s)}_{(\omega,t)}$ we directly see that $\nu_1^\pi\in \mathsf{RST}$. Moreover, as $\mathrm{proj}_{\mathcal{Y}}(\pi)\le\nu$ by the definition of $\mathsf{JOIN}(\tau,\nu)$ we directly get $\mathbb{E}_{\nu_1^\pi}[T]\le 2\mathbb{E}_\nu[T]<\infty$.

Summing up we have constructed $\nu_0^\pi,\nu_1^\pi\in\mathsf{RST}$ such that $\mathbb{E}_{\nu_0^\pi}[T],\mathbb{E}_{\nu_1^\pi}[T]<\infty$. To prove the theorem we need to show that $\nu^\pi=\frac{1}{2}(\nu_0^\pi+\nu_1^\pi)\in\mathsf{RST}(\mu)$ and that $\int\gamma\,d\nu^\pi>\int\gamma\,d\nu$. To this end let us consider the constributions of ν_0^π and ν_1^π separately. For $A\subseteq\Omega\times\mathbb{R}_+$ it holds that

$$\nu_0^\pi(A)-\nu(A)=\int_A H^\nu(\omega,t)\,d\rho_0(\omega,t)-\int_A(\rho_0)_\omega([0,t])\,d\nu(\omega,t).$$

Furthermore,

$$\begin{aligned}
&\int_A(\rho_0)_\omega([0,t])\,d\nu(\omega,t)\\
&=\int_A\int_{\mathcal{Y}}\int_0^t\pi((\omega,du),d(g,s))\nu_\omega(dt)\,\mathbb{W}(d\omega)\\
&=\int_\Omega\int_{\mathcal{Y}}\int_{\mathbb{R}_+}\int_u^\infty\mathbb{1}_A(\omega,t)\,\nu_\omega(dt)\,\pi((\omega,du),d(g,s))\,\mathbb{W}(d\omega)\\
&=\int_\Omega\int_{\mathcal{Y}}\int_{\mathbb{R}_+}H^\nu(\omega,u)\nu^{(\omega,u)}_{\theta_u(\omega)}(A_\omega)\pi((\omega,du),d(g,s))\,\mathbb{W}(d\omega),
\end{aligned}$$

where $A_\omega=\{t\in\mathbb{R}_+:(\omega,t)\in A\}$. This yields

$$\int\gamma d(\nu_0^\pi-\nu)=\int\pi(d(f,s),d(g,t))\,H^\nu(f,s)\,\left[\gamma(f,s)-\int\gamma(f\oplus\omega_{\upharpoonright[0,u]},s+u)\nu^{(f,s)}(d(\omega,u))\right].\tag{6.9}$$

For ν_1^π we can compute

$$\int\gamma d(\nu_1^\pi-\nu)=\int\pi(d(f,s),d(g,t))H^\nu(f,s)\left[\int\gamma(g\oplus\omega_{\upharpoonright[0,u]},t+u)\nu^{(f,s)}(d(\omega,u))-\gamma(g,t)\right].\tag{6.10}$$

Putting this together yields

$$\begin{aligned}
&2\int\gamma\,d(\nu^\pi-\nu)\\
&=\int d\pi((f,s),(g,t))\Big(H^\nu(f,s)\Big[-\int\gamma(f\oplus\omega_{\upharpoonright[0,r]},s+r)\,d\nu^{(f,s)}(\omega,r)-\gamma(g,t)+\\
&\qquad\gamma(f,s)+\int\gamma(g\oplus\omega_{\upharpoonright[0,r]},t+r)\,d\nu^{(f,s)}(\omega,r)\Big]\Big),
\end{aligned}$$

which is strictly positive by the definition of bad pairs relative to ν and Assumption (6.5). Moreover, the last identity holds for any measurable function $F:C(\mathbb{R}_+)\times\mathbb{R}_+\to\mathbb{R}$ instead of γ. In particular, taking $F(\omega,t)=G(\omega(t))$ for some measurable $G:\mathbb{R}\to\mathbb{R}$ we get

$$\begin{aligned}
&2\int G\,d(\nu^\pi-\nu)\\
&=\int d\pi((f,s),(g,t))\Big(H^\nu(f,s)\Big[-\int G(f\oplus\omega_{\upharpoonright[0,r]}(s+r))\,d\nu^{(f,s)}(\omega,r)-G(g(t))+\\
&\qquad G(f(s))+\int G(g\oplus\omega_{\upharpoonright[0,r]}(t+r))\,d\nu^{(f,s)}(\omega,r)\Big]\Big)=0,
\end{aligned}$$

because $\{(f,s),(g,t)\} \in \mathsf{BP}_\nu$ implies that $f(s) = g(t)$ and π is concentrated on BP_ν. This proves $\nu^\pi \in \mathsf{RST}(\mu)$. Hence, we derive at a contradiction and the claim is proven. □

6.1. **Approximation by particular stopping times.**

Lemma 6.6. *Let τ be a non-randomized $\mathcal{F}^+$-stopping time. For any $\varepsilon, \eta > 0$ there is an $\mathcal{F}^+$-stopping time ρ such that*

(1) $\rho \le \tau$
(2) $\mathbb{W}(\tau - \rho \ge \varepsilon) \le \eta$
(3) $\mathbb{W}(\{\tau = \infty, \rho < \infty\}) \le \eta$
(4) $[\![0,\rho]\!]$ *is closed in* $C(\mathbb{R}_+) \times \mathbb{R}_+$ *and* S.

Recall from Lemma 4.7 that for an $\mathcal{F}^+$-stopping time, the stochastic interval $[\![0,\rho]\!]$ can be identified with the Borel subset $r([\![0,\rho]\!])$ of S and from Proposition 4.4 that $[\![0,\rho]\!]$ is closed in $C(\mathbb{R}_+) \times \mathbb{R}$ iff $r([\![0,\rho]\!])$ is closed in S.

Proof. Fix $\varepsilon, \eta > 0$. Assume first that

$$\tau(\omega) = \begin{cases} t & \omega \in A \\ \infty & \text{otherwise} \end{cases},$$

for some $\mathcal{F}^0_t$-measurable set A. If $t = 0$ we are done, so we assume $t > 0$. By Proposition 4.4, there is a Borel set $A_0 \subseteq C([0,t])$ such that $A = A_0 \oplus C((t,\infty))$. In this proof we will often use this kind of identification of $\mathcal{F}^0_t$-measurable events with measurable subsets of $C([0,t])$ without explicitly mentioning it. In particular, we will loosely write $\mathbb{W}(D)$ instead of $\mathbb{W}(D \oplus C((t,\infty)))$ or $\mathbb{W}_{\upharpoonright C([0,t])}(D)$ for some measurable $D \subseteq C([0,t])$.

By outer regularity of $\mathbb{W}$ there is an open set $O \subseteq C[0,t]$, $O \supseteq A_0$ such that $\mathbb{W}(O \setminus A_0) \le \eta/2$. Moreover, O can be written as a countable union of open sets $O_n, n \ge 1$, where for each n the set O_n is an open sausage corresponding to some continuous function $f_n : [0,t] \to \mathbb{R}$ and some $\eta_n > 0$, i.e. $O_n = \{g : \mathbb{R}_+ \to \mathbb{R}, \sup_{s \le t} |f_n(s) - g(s)| < \eta_n\}$. For all $n \ge 1$ there is $t - \varepsilon \le t_n < t$ such that the open sausage O'_n corresponding to η_n and the function f_n restricted to $[0,t_n]$ satisfies $\mathbb{W}(O'_n \setminus O_n) \le 2^{-(n+1)}\eta$. Put, $O' = \cup_{n\ge1} O'_n$. Then $O \subseteq O'$ and $\mathbb{W}(O' \setminus O) \le \eta/2$ and therefore $\mathbb{W}(O' \setminus A_0) \le \eta$. Set

$$\rho_n(\omega) = \begin{cases} t_n & \omega \in O'_n \\ \infty & \text{otherwise} \end{cases}.$$

Then, $]\!]\rho_n, \infty]\!]$ is open and $[\![0,\rho_n]\!]$ is closed. Put $U = \bigcup_n]\!]\rho_n,\infty]\!]$ and define

$$\rho(\omega) = \inf\{t : (\omega,t) \in U\}.$$

Then, we have

$$[\![0,\rho]\!] = \bigcap_n [\![0,\rho_n]\!],$$

which implies that $\rho(\omega) = \inf_n \rho_n(\omega)$. Hence, ρ is an $\mathcal{F}^+$ stopping time and $[\![0,\rho]\!]$ is closed. Moreover, because $t-\varepsilon \le t_n < t$ we have that for all $n \ge 1$ it holds that $t-\varepsilon \le \rho_n(\omega) \le \tau(\omega)$, and $\rho_n < \tau$ if $\rho_n < \infty$. Hence, it also holds that $t - \varepsilon \le \rho(\omega) \le \tau(\omega)$. Therefore, we can conclude that

$$\mathbb{W}(\{|\tau - \rho| > \varepsilon\}) = \mathbb{W}(\{\tau = \infty, \rho < \infty\}) = \mathbb{W}(O' \setminus A_0) \le \eta.$$

This proves the Lemma for the case that τ is an $\mathcal{F}^0$-stopping time which only takes the values t and ∞. From here it is straightforward to prove the Lemma for the case where τ takes values in a discrete subset of $\mathbb{R}_+$.

Assume now that τ is an arbitrary $\mathcal{F}^0$-stopping time. Pick a sequence of stopping times $(\tau_n)_{n\in\mathbb{N}}$ which for any n take only values in some countable discrete set such that $\tau_n \downarrow \tau$. Put $\eta_n = 2^{-n}\eta/2$. According to what we have proved above pick ρ_n which are very close (in terms of ε, η_n) to the τ_n and satisfy that $]\!]\rho_n,\infty]\!]$ is open. Then set

$$V := \bigcup_n]\!]\rho_n, \infty]\!]$$

and

$$\rho := \inf\{t : (\omega, t) \in V\}$$

such that $V =]\!]\rho, \infty]\!]$ is open. Note that $\rho = \inf_n \rho_n$. Hence, by construction $\rho \leq \tau$ satisfies the required properties. Indeed, we only have to check that $\mathbb{W}(\tau - \rho \geq \varepsilon) \leq \eta$. To this end, one easily checks that

$$\bigcap_n \{\tau_n - \rho_n < \varepsilon\} \subseteq \{\tau - \rho < \varepsilon\},$$

which directly yields the estimate.

If τ is an $\mathcal{F}^+$-stopping time, it can be represented as a decreasing limit of $\mathcal{F}^0$-stopping times and repeating the above argument yields the result also in this case. □

Remark 6.7. *To prove the previous lemma for a general starting distribution λ we need to make an additional approximation step for stopping times of the form*

$$\tau(\omega) = \begin{cases} 0 & \omega_0 \in A \\ \infty & else \end{cases}.$$

Just take an open set $O \supseteq A$ with $\lambda(O \setminus A) \leq \eta$. The rest of the argument stays the same.

Corollary 6.8. *Let τ be a non-randomized $\mathcal{F}^+$-stopping time. Then there is a sequence of $\mathcal{F}^+$-stopping times τ_n such that*

(1) $\tau_n \uparrow \tau$ $\mathbb{W}$-a.s.
(2) $\mathbb{W}(\{\tau = \infty\} \cap \{\tau_n < \infty\}) \to 0$.
(3) For each n the stochastic interval $[\![0, \tau_n]\!]$ is closed in $C(\mathbb{R}_+) \times \mathbb{R}_+$ and S.

Proof. For each n apply the previous lemma with $\varepsilon_n = \eta_n = 2^{-n}$. □

If τ is an $\mathcal{F}^a$-stopping time then the result still applies with a minor modification: we have to allow for an exceptional null set N.

7. A Filtered Kellerer-type Lemma and the Monotonicity Principle

Recall for a set $\Gamma \subseteq S$ the definition of $\Gamma^<$ from (1.5) and for $\nu \in \mathsf{RST}(\mu)$ the definition of BP_ν from Definition 6.4.

Definition 7.1. *Let $\nu \in \mathsf{RST}(\mu)$. Then a set $\Gamma \subseteq S$ is called γ-*monotone *iff*

$$\mathsf{BP}_\nu \cap (\Gamma^< \times \Gamma) = \emptyset.$$

The following theorem implies Theorem 1.2 stated in the introduction.

Theorem 7.2. *Assume that $\gamma : S \to \mathbb{R}$ is Borel-measurable, the optimization problem* (5.3) *is well-posed and that ν is an optimizer of* (OptSEP). *Then, ν is supported on a γ-monotone Borel set $\Gamma \subseteq S$.*

As an intermediate step towards the proof of Theorem 7.2 we will look for two different sets $\Gamma_L \subseteq S$ and $\Gamma_D \subseteq S$ where Γ_L (which roughly corresponds to $\Gamma^<$) represents the "still living pairs", while ν is concentrated on Γ_D which represents the paths which get killed by ν. Here Γ_L is a subset of all (f, s) which lie before the "death"-set Γ_D. The above condition on Γ then corresponds to: for $\big((f, s), (g, t)\big) \in \mathsf{BP}_\nu$, at least one of the following applies:

(1) $(f, s) \notin \Gamma_L$ ((f, s) is not living).
(2) $(g, t) \notin \Gamma_D$ ((g, t) is not dying).

This can equivalently be expressed as

$$\mathsf{BP}_\nu \ \cap \ (\Gamma_L \times \Gamma_D) = \emptyset. \tag{7.1}$$

Define a (non-randomized) stopping time τ_ν by

$$\tau_\nu(\omega) := \inf\{t : H^\nu \circ r(\omega, t) = 0\}.$$

Using Lemma 6.6 and Corollary 6.8, by choosing a suitable foretelling sequence of the stopping times $\inf\{t : H^\nu(r(\omega,t)) \le 1/n\}$, we can pick a sequence $\tau_n, n \ge 1$ of $\mathcal{F}^+$-stopping times such that

(1) $\tau_n \uparrow \tau_\nu$.
(2) $\tau_n \le n$.
(3) $[\![0,\tau_n]\!]$ is closed.
(4) $\tau_n < \inf\{t : H^\nu(r(\omega,t)) \le 1/n\}$ on $\{\omega : H^\nu(r(\omega,0)) > 1/n\}$.

Recall the definition of $\mathsf{JOIN}(\tau,\nu)$ from the beginning of Section 6. Fix n. Then every joining $\pi \in \mathsf{JOIN}(\tau_n,\nu)$ satisfies the assumptions of Proposition 6.5 and hence $\pi(\mathsf{BP}_\nu) = 0$.

For a non-randomized stopping time τ we write $\bar\tau$ for the random measure given through $\bar\tau(d\omega,dt) = \mathbb{W}(d\omega)\delta_{\tau(\omega)}(dt)$. Note that we can view $\bar\tau$ also as a measure on S, as in (5.1). We also recall from Definition 4.6 that a subset of S is left/right complete iff it is closed under forming restrictions/extensions of its paths. Subsequently we will prove:

Lemma 7.3 (Filtered Kellerer Lemma)**.** *Assume that τ is a (non-randomized) bounded $\mathcal{F}^+$-stopping time such that $[\![0,\tau]\!]$ is closed.*

Let $(\mathcal{Y},\nu)$ be a Polish probability space. Consider a ("bad") set $B \subseteq S\times\mathcal{Y}$. If $\pi(B) = 0$ for all $\pi \in \mathsf{JOIN}(\tau,\nu)$, then there exists a right complete set $D \subseteq S$ and a set $N \subseteq \mathcal{Y}$ such that $B \subseteq (D\times\mathcal{Y}) \cup (S\times N)$ and $\bar\tau(D) = \nu(N) = 0$.

Admitting Lemma 7.3 we set $\Gamma_L = S \setminus D$ and $\Gamma_D = \mathcal{Y}\setminus N$ to obtain (7.1).

Proof of Theorem 7.2 from Lemma 7.3. Specify $\mathcal{Y} = (S,\nu)$. We admit Lemma 7.3 and apply it to the stopping times τ_n defined above to find left complete sets $L_n := S\setminus D_n$ and sets $\Gamma_n := S\setminus N_n$ such that

$$\mathsf{BP}_\nu \cap (L_n \times \Gamma_n) = \emptyset$$

and $\bar\tau_n(L_n) = 1, \nu(\Gamma_n) = 1$. Setting $\tilde\Gamma := \bigcap_n \Gamma_n$ we have $\mathsf{BP}_\nu \cap (L_n\times\tilde\Gamma) = \emptyset$ for all n. With $L := \bigcup_n L_n$ we have

$$\mathsf{BP}_\nu \cap (L\times\tilde\Gamma) = \emptyset. \tag{7.2}$$

Put

$$L^+ := \{(g,t) \,:\, (g_{\restriction[0,s]}, s) \in L \text{ for all } s<t\}.$$

Note that L^+ is universally measurable: its complement $S\setminus L^+$ is given by

$$\{(g,t) : \exists s<t, (g_{\restriction[0,s]},s) \notin L\} \tag{7.3}$$

$$= \mathrm{proj}_S\,\{((g,t),s) \in S\times[0,t) : (g_{\restriction[0,s]},s)\notin L\} \tag{7.4}$$

and is hence analytic.

Because L is left complete we have $L \subseteq L^+$ and also $(L^+)^< \subseteq L$. Clearly $\bar\tau_n(L) = \bar\tau(L^+) = 1$ for all n. Moreover, the set $\{t : r(\omega,t)\in L^+\}$ is either empty or closed. Hence, the convergence $\tau_n \nearrow \tau_\nu$ almost surely implies that $\bar\tau_\nu(L^+) = 1$ and therefore $\nu(L^+) = 1$. Hence, setting $\bar\Gamma := L^+\cap\tilde\Gamma$ we have $\nu(\bar\Gamma) = 1$, $\bar\Gamma^< \subseteq (L^+)^< \subseteq L$ and by (7.2) we can conclude

$$\mathsf{BP}_\nu \cap (\bar\Gamma^< \times \bar\Gamma) = \emptyset.$$

Of course this pertains if we replace $\bar\Gamma$ by a Borel-subset of full measure. $\square$

It remains to establish Lemma 7.3 which we shall now do.

Important Convention. For the remainder of this section we fix a (finite) non-randomized stopping time τ such that $[\![0,\tau]\!]$ is closed and satisfies $\tau \le t_0$ for some $t_0 \in \mathbb{R}_+$.

7.1. **An auxiliary Optimization Problem.** We fix a Polish probability space $(\mathcal{Y}, \nu)$ which eventually will be taken to be (S, ν), where ν denotes an optimizer of the primal problem (5.3). Let $c : C(\mathbb{R}_+) \times \mathbb{R}_+ \times \mathcal{Y} \to \mathbb{R}_+$ be a predictable and S- upper semi-continuous function. We are interested in the following maximization problem

$$P^{\leq 1} = P_c^{\leq 1}(\mathbb{W}, \tau, \nu) = \sup_{\pi \in \mathsf{JOIN}(\tau,\nu)} \int_{C(\mathbb{R}_+)\times\mathbb{R}_+\times\mathcal{Y}} c\, d\pi \tag{7.5}$$

and its relation to the dual problem

$$D^{\leq 1} = D_c^{\leq 1}(\mathbb{W}, \tau, \nu) = \inf_{(\varphi,\psi)\in\mathsf{DC}} \Big(\mathbb{E}_{\mathbb{W}}[\varphi_\tau^M] + \mathbb{E}_\nu[\psi]\Big), \tag{7.6}$$

$$\mathsf{DC} = \{(\varphi, \psi) \in C_b(\Omega) \times C_b(\mathcal{Y}) : \varphi, \psi \geq 0, c(\omega, t, y) \leq \varphi_t^M(\omega) + \psi(y), \forall t \leq \tau, y \in \mathcal{Y}, \omega \in \Omega\},$$

where φ^M is an S- continuous martingale (cf. Proposition 4.10). To indicate the dependence of DC on the cost function c and the stopping time τ we sometimes write $\mathsf{DC}(c)$ or $\mathsf{DC}(c, \tau)$. Note that for integrable φ we always have $\mathbb{E}_{\mathbb{W}}[\varphi] = \mathbb{E}_{\mathbb{W}}[(\varphi_\tau^M)]$ by optional stopping. This result differs from the already established duality also in that we allow subprobability measures.

Observe that due to predictability of c the maximization problem is not altered if we replace PRST_τ by RST_τ in the definition of $\mathsf{JOIN}(\tau, \nu)$, cf. (6.2).

As in Lemma 5.2 it is easy to show $D^{\leq 1} \geq P^{\leq 1}$. We now consider the other inequality.

Theorem 7.4. *Let $c : C(\mathbb{R}_+) \times \mathbb{R}_+ \times \mathcal{Y} \to \mathbb{R}_+$ be predictable (in the sense of Remark 4.25), S- upper semi-continuous and bounded from above. Assume that τ is a bounded stopping time such that $[\![0, \tau]\!]$ is closed. Then*

$$P^{\leq 1} = \sup_{\pi \in \mathsf{JOIN}(\tau,\nu)} \int_{C(\mathbb{R}_+)\times\mathbb{R}_+\times\mathcal{Y}} c\, d\pi = \inf_{(\varphi,\psi)\in\mathsf{DC}} (\mathbb{E}_{\mathbb{W}}[\varphi] + \mathbb{E}_\nu[\psi]) = D^{\leq 1}.$$

We will first prove a version which applies to not necessarily predictable c. Afterwards, we will use the defining property of PRST, Equation (4.8), to derive the predictable version.

Theorem 7.5. *Let $c : C(\mathbb{R}_+) \times \mathbb{R}_+ \times \mathcal{Y} \to \mathbb{R}_+$ be (upper semi-) continuous and bounded from above. Then*

$$P^{\leq 1, NA} := \sup_{\pi \in \mathsf{TM}(\tau,\nu)} \int_{C(\mathbb{R}_+)\times\mathbb{R}_+\times\mathcal{Y}} c\, d\pi = \inf_{(\varphi,\psi)\in\widetilde{\mathsf{DC}}} (\mathbb{E}_{\mathbb{W}}[\varphi] + \mathbb{E}_\nu[\psi]) =: D^{\leq 1, NA},$$

where $\widetilde{\mathsf{DC}} = \{(\varphi, \psi) \in C_b(\Omega) \times C_b(\mathcal{Y}) : \varphi, \psi \geq 0, c(\omega, t, y) \leq \varphi(\omega) + \psi(y)$ *for all* $y \in \mathcal{Y}, t \leq \tau,\ \omega \in \Omega\}$.

Here the set of all *tagged random measures* is given by

$$\mathsf{TM}(\tau, \nu) := \Big\{\pi \in \mathcal{P}^{\leq 1}(C(\mathbb{R}_+) \times \mathbb{R}_+ \times \mathcal{Y}), \mathsf{proj}_{C(\mathbb{R}_+)\times\mathbb{R}_+}(\pi) \in \mathsf{M}_\tau, \mathsf{proj}_{\mathcal{Y}}(\pi) \leq \nu\Big\}.$$

Proof of Theorem 7.5. We reduce the theorem to the classical duality theorem in optimal transport. Put $\bar{c}(\omega, y) = \sup_{t \leq \tau(\omega)} c(\omega, t, y)$. As $[\![0, \tau]\!]$ is closed and bounded $\bar{c}$ is continuous.

Then the dual constraint set can be written as

$$\widetilde{\mathsf{DC}} = \{(\varphi, \psi) \in C_b(\Omega) \times C_b(\mathcal{Y}) : \varphi, \psi \geq 0, \bar{c}(\omega, y) \leq \varphi(\omega) + \psi(y) \text{ for all } y \in \mathcal{Y}, \omega \in \Omega\}.$$

From the classical duality theorem of optimal transport (3.1) we know that

$$\inf_{(\varphi,\psi)\in\widetilde{\mathsf{DC}}} \mathbb{W}(\varphi) + \nu(\psi) = \sup_{q \in \mathsf{Cpl}(\mathbb{W},\nu)} \int_{\Omega\times\mathcal{Y}} \bar{c}(\omega, y)\, q(d\omega, dy) =: \check{P}.$$

It remains to show that $\check{P} = P^{\leq 1, NA}$. From the definition of $\bar{c}$ and TM it is clear that we always have $P^{\leq 1, NA} \leq \check{P}$. To prove the other inequality fix $\varepsilon > 0$ and take $q \in \mathsf{Cpl}(\mathbb{W}, \nu)$.

For any (ω, y) there is $t(\omega, y) \leq \tau(\omega)$ such that $c(\omega, t(\omega, y), y) + \varepsilon \geq \bar{c}(\omega, y)$. Putting $\pi(d\omega, ds, dy) := q(d\omega, dy)\delta_{t(\omega,y)}(ds) \in \mathsf{TM}$ we get

$$\int_{C(\mathbb{R}_+)\times\mathbb{R}_+\times\mathcal{Y}} c(\omega, s, y)\, \pi(d\omega, ds, dy) + \varepsilon \geq \int_{C\times\mathcal{Y}} \bar{c}(\omega, y)\, q(d\omega, dy).$$

This implies that $P^{\leq 1,NA} + \varepsilon \geq \check{P}$. Letting ε go to zero we obtain the claim. □

Proof of Theorem 7.4. As c is bounded from above we have $P^{\leq 1} < \infty$. Arguing as in Lemma 5.5, we may assume that the cost function c is continuous.

We will now argue as in Proposition 5.6, i.e. we consider again the functions $h, \bar{h}$ as in (5.14) and we shall apply Theorem 5.4 to the function

$$F(\pi, h) = \int c + \bar{h}\, d\pi$$

for $\pi \in \mathsf{TM}(\tau, \nu)$. The set $\mathsf{TM}(\tau, \nu)$ is convex and compact by Prohorov's theorem and the set of h under consideration is convex as well. The function F is continuous by definition of TM.

This allows us to deduce

$$\begin{aligned} P^{\leq 1} &= \sup_{\pi\in\mathsf{JOIN}(\tau,\nu)} \int c\, d\pi \\ &= \sup_{\pi\in\mathsf{TM}(\tau,\nu)} \inf_h \left(\int c + \bar{h}\, d\pi\right) \\ &\overset{Thm5.4}{=} \inf_h \sup_{\pi\in\mathsf{TM}(\tau,\nu)} \left(\int (c + \bar{h})\, d\pi\right) \\ &= \inf_h \inf_{(\varphi,\psi)\in\widetilde{\mathsf{DC}}(c+\bar{h})} \left(\mathbb{W}(\varphi) + \nu(\psi)\right). \end{aligned}$$

The last equality holds by Theorem 7.5. Write

$$c_h(\omega, t, y) = c(\omega, t, y) + \sum_{i=1}^{n} (f_i(\omega) - f^M_{i,t}(\omega))g(y).$$

For $(\varphi, \psi) \in \widetilde{\mathsf{DC}}(c_h)$ it holds that

$$c_h(\omega, t, y) \leq \varphi(\omega) + \psi(y).$$

Taking conditional expectations w.r.t. $\mathcal{F}^0_t$ in the sense of Definition 4.9 we get using predictability of c

$$c(\omega, t, y) \leq \varphi^M_t(\omega) + \psi(y).$$

This implies that $(\varphi, \psi) \in \mathsf{DC}(c)$, hence $\widetilde{\mathsf{DC}}(c_h) \subseteq \mathsf{DC}(c)$. Because $\mathbb{W}(\varphi^M_t) = \mathbb{W}(\varphi)$ this implies that

$$\begin{aligned} P^{\leq 1} &= \inf_h \inf_{(\varphi,\psi)\in\widetilde{\mathsf{DC}}(c+\bar{h})} (\mathbb{W}(\varphi) + \nu(\psi)) \\ &\geq \inf_{(\varphi,\psi)\in\mathsf{DC}(c)} (\mathbb{W}(\varphi) + \nu(\psi)) = D^{\leq 1}. \end{aligned}$$

As also $D^{\leq 1} \geq P^{\leq 1}$ we can conclude $D^{\leq 1} = P^{\leq 1}$. □

7.2. **A Choquet argument.** Denote by $LSC_b(X)$ the set of bounded lower semi-continuous functions on X. The following lemma is a simple consequence of Theorem 7.4.

Lemma 7.6. *Let $c : C(\mathbb{R}_+)\times\mathbb{R}_+\times\mathcal{Y} \to [0,1]$ be predictable (in the sense of Remark 4.25) and S - upper semi-continuous. Assume that τ is a bounded stopping time such that $[\![0,\tau]\!]$ is closed. Then*

$$P^{\leq 1} = \sup_{\pi\in\mathsf{JOIN}(\tau,\nu)} \int_{C(\mathbb{R}_+)\times\mathbb{R}_+\times\mathcal{Y}} c\, d\pi = \inf_{(\varphi,\psi)\in\mathsf{DC}'} (\mathbb{E}_{\mathbb{W}}[\varphi] + \mathbb{E}_\nu[\psi]) = D',$$

where

$$\mathsf{DC}' = \{(\varphi,\psi) \in LSC_b(\Omega)\times LSC_b(\mathcal{Y}) : \varphi,\psi \in [0,1], c(\omega,t,y) \leq \varphi^M_t(\omega) + \psi(y), \forall\, t \leq \tau, y \in \mathcal{Y}, \omega \in \Omega\}. \tag{7.7}$$

Proof. Pick by Theorem 7.4 two continuous, bounded and non-negative functions $\varphi,\psi \in \mathsf{DC}$. Then $\bar\psi := \psi \wedge 1$ is still continuous and as $\varphi \geq 0$ we also have $(\varphi,\bar\psi) \in \mathsf{DC}$. Put $\rho = \inf\{t \geq 0 : \varphi^M_t > 1\}$. Due to continuity of φ^M the set $D := \{\varphi^M > 1\}$ is open. Hence also $\{\rho < \infty\} = \mathrm{proj}_\Omega D$ is open as projections are open maps. Then the map $\omega \mapsto \varphi^M_{\rho(\omega)}(\omega) =: \bar\varphi(\omega) \leq 1$ is lower semi-continuous. Clearly, $(\bar\varphi,\bar\psi) \in \mathsf{DC}'$ with $\mathbb{E}_{\mathbb{W}}[\bar\varphi] + \mathbb{E}_\nu[\bar\psi] \leq \mathbb{E}_{\mathbb{W}}[\varphi] + \mathbb{E}_\nu[\psi]$. □

Corollary 7.7. *Let $K \subseteq C(\mathbb{R}_+)\times\mathbb{R}_+\times\mathcal{Y}$ be predictable. Assume that τ is a bounded stopping time such that $[\![0,\tau]\!]$ is closed. Then*

$$P^{\leq 1} = \sup_{\pi\in\mathsf{JOIN}(\tau,\nu)} \int_{C(\mathbb{R}_+)\times\mathbb{R}_+\times\mathcal{Y}} \mathbb{1}_K\, d\pi = \inf_{(\varphi,\psi)\in\mathsf{DC}''} (\mathbb{E}_{\mathbb{W}}[\varphi] + \mathbb{E}_\nu[\psi]) = D'', \tag{7.8}$$

where DC'' is given by

$$\{(\varphi,\psi) \in LSC_b(\Omega)\times LSC_b(\mathcal{Y}) : \varphi,\psi \in [0,1], \mathbb{1}_K(\omega,t,y) \leq \varphi^M_t(\omega) + \psi(y),\ t \leq \tau, \omega \in \Omega, y \in \mathcal{Y}\}.$$

Proof. To indicate the dependence on the set K we write $D''(K)$ and for notational convenience we drop the two primes and simply write $D(K)$. The left hand side in (7.8) is clearly a capacity. To establish the claim, it is therefore sufficient to show that $D(K)$ is also a capacity on $S\times\mathcal{Y}$, because the indicator of a closed set is upper semi continuous and the result then follows from Lemma 7.6 and Choquet's theorem.

Hence, we need to show the three defining properties of a capacity, namely monotonicity, continuity from below and continuity from above for compact sets. The monotonicity is clear. Let us turn to the continuity from below.

Take an increasing sequence $A_1 \subseteq A_2 \subseteq \ldots \subseteq C(\mathbb{R}_+)\times\mathbb{R}_+\times\mathcal{Y}$ of Borel measurable predictable sets and put $A = \bigcup_n A_n$. For all n there are lower semi-continuous functions $\varphi_n : C(\mathbb{R}_+) \to [0,1]$ and $\psi_n : \mathcal{Y} \to [0,1]$ such that $\mathbb{1}_{A_n}(\omega,t,y) \leq \varphi^M_n(\omega,t) + \psi_n(y)$ for all ω,t,y and

$$\nu(\psi_n) + \mathbb{W}(\varphi_n) \leq D(A_n) + \frac{1}{n}.$$

Using a Komlos type lemma we can assume that some appropriate convex combinations of ψ_n and φ_n converge a.s. to functions ψ and φ. Let us be a little bit more precise here. By [9, Proposition 3.3], there exist convex coefficients $\alpha^n_n,\ldots,\alpha^n_{k_n}, n \geq 1, k_n < \infty$, and full measure subsets $\Omega_1 \subseteq \Omega$, $\mathcal{Y}_1 \subseteq \mathcal{Y}$ such that with $\tilde\varphi_n := \sum_{i=n}^{k_n} \alpha^n_i\varphi_i$, $\tilde\psi_n := \sum_{i=n}^{k_n}\alpha^n_i\psi_i$ we have that for all $\omega \in \Omega_1$ and all $y \in \mathcal{Y}_1$

$$\lim_{n\to\infty} \tilde\varphi_n(\omega) =: \varphi(\omega) \text{ and } \lim_{n\to\infty}\tilde\psi_n(y) =: \psi(y) \tag{7.9}$$

exist. Extend these functions to $\mathcal{X}$ and $\mathcal{Y}$, resp., through

$$\limsup_{n\to\infty} \tilde\varphi_n(\omega) =: \varphi(\omega) \text{ and } \limsup_{n\to\infty}\tilde\psi_n(y) =: \psi(y). \tag{7.10}$$

Due to the boundedness of $\tilde{\varphi}_n$ the same equalities hold with φ^M_t and $\tilde{\varphi}^M_n(\cdot, t)$ in place of φ and $\tilde{\varphi}_n$ for all $t \geq 0$. Given $m \leq n$ we have for all ω, t, y

$$\mathbb{1}_{A_m}(\omega, t, y) \leq \tilde{\varphi}^M_n(\omega, t) + \tilde{\psi}_n(y),$$

hence $\mathbb{1}_{A_m}(\omega, t, y) \leq \varphi^M_t(\omega) + \psi(y)$ and thus also

$$\mathbb{1}_A(\omega, t, y) \leq \varphi^M_t(\omega) + \psi(y).$$

Given $\varepsilon > 0$, we can find lower semi continuous functions $\varphi^\varepsilon \geq \varphi$ and $\psi^\varepsilon \geq \psi$ such that $\mathbb{W}(\varphi^\varepsilon) - \varepsilon/2 < \mathbb{W}(\varphi) = \lim \mathbb{W}(\tilde{\varphi}_n)$ and $\nu(\psi^\varepsilon) - \varepsilon/2 < \nu(\psi) = \lim \nu(\tilde{\psi}_n)$. Therefore we can conclude

$$D(A) \leq \limsup_n D(A_n) + \tfrac{1}{n} + \varepsilon.$$

To show continuity from above for compact sets, take a sequence $K_1 \supseteq K_2 \supseteq \ldots$ of compact and predictable sets in $\mathcal{X} \times \mathcal{Y}$ and put $K = \bigcap_n K_n$. Fix $\varepsilon > 0$. Then there is $(\varphi, \psi) \in \mathsf{DC}''$ s.t.

$$\int \psi \, d\nu + \int \varphi \, d\mathbb{W} \leq D(K) + \varepsilon.$$

As $(\varphi, \psi) \in \mathsf{DC}''$ it holds that $K \subseteq \{\varphi^M + \psi \geq 1\}$. On the cost of another ε we can find two lower semi-continuous functions $\varphi^\varepsilon \geq \varphi$ and $\psi^\varepsilon \geq \psi$ such that $\mathbb{E}\varphi^\varepsilon + \nu(\psi^\varepsilon) \leq \mathbb{E}\varphi + \nu(\psi) + \varepsilon$ and $K \subseteq \{(\varphi^\varepsilon)^M + \psi^\varepsilon > 1\}$. By lower semi-continuity, $\{(\varphi^\varepsilon)^M + \psi^\varepsilon > 1\}$ is open. Hence, there is an N such that for all $n \geq N$ we must have $\mathbb{1}_{K_n} \leq (\varphi^\varepsilon)^M + \psi^\varepsilon$. This implies that

$$D(K_n) \leq D(K) + 2\varepsilon,$$

proving the claim. □

Corollary 7.8. *Let $K \subseteq [\![0, \tau]\!] \times \mathcal{Y}$ be* predictable *(see Remark 4.25) and assume that $\sup_{\pi \in \mathsf{JOIN}(\tau,\nu)} \pi(K) < 1/2$. Then*

$$\frac{1}{2} \sup_{\pi \in \mathsf{JOIN}(\tau,\nu)} \pi(K) \leq \inf_{(\kappa,A) \in \mathsf{Cov}(K)} \Big(\mathbb{W}(\kappa < \tau) + \nu(A)\Big) \leq 2 \sup_{\pi \in \mathsf{JOIN}(\tau,\nu)} \pi(K),$$

where

$$\mathsf{Cov}(K) = \{\kappa \text{ is a } \mathcal{F}^+\text{-stopping time}, A \subseteq \mathcal{Y} : K \subseteq]\!]\kappa, \infty]\!] \times \mathcal{Y} \cup (C(\mathbb{R}_+) \times \mathbb{R}_+) \times A\}.$$

Proof. We want to apply the previous Corollary. Clearly, $P^{\leq 1}(\mathbb{1}_K) = \sup_{\pi \in \mathsf{JOIN}(\tau,\nu)} \pi(K)$. We have to show that

$$D(K) \approx \inf_{(\kappa,A) \in \mathsf{Cov}} \Big(\mathbb{W}(\kappa < \infty) + \nu(A)\Big).$$

To this end take $(\varphi, \psi) \in \mathsf{DC}''$. As the cost function is $\{0, 1\}$-valued, the dual constraint

$$\mathbb{1}_K((\omega, t), y) \leq \varphi^M_t(\omega) + \psi(y)$$

implies that

$$K \subseteq \{(\omega, t) : \varphi^M_t(\omega) \geq 1/2\} \times \mathcal{Y} \cup C(\mathbb{R}_+) \times \mathbb{R}_+ \times \{y : \psi(y) \geq 1/2\}.$$

Hence, on the cost of a factor 2 we can replace ψ by the indicator of a set $A \subseteq \mathcal{Y}$. We can just take $A = \{\psi \geq 1/2\}$. In particular, given ψ we can safely replace it by $\tilde{\psi} = 1_A \psi \wedge 1$ because $\tilde{\psi}$ has smaller expectation and at least as good covering properties as ψ. Obviously, $1/2\nu(A) \leq \mathbb{E}_\nu \tilde{\psi} \leq \nu(A)$.

Let us turn our attention to the set $E = \{(\omega, t) : \varphi^M_t(\omega) \geq \frac{1}{2}\}$. By the condition on π given in the corollary, we may assume that $\varphi^M_0 < \frac{1}{2}$, and choose $\frac{1}{2} - \varphi^M_0 > \varepsilon > 0$. Define the stopping time

$$\kappa(\omega) = \inf\{t \geq 0 : \varphi^M_t(\omega) > 1/2 - \varepsilon\},$$

with $\inf(\emptyset) = \infty$. As φ^M is S- lower semi-continuous, the set

$$]\!]\kappa, \infty[\![= \{(\omega, t) : \exists s < t, \varphi^M_s(\omega) > \frac{1}{2} - \varepsilon\} = \bigcup_s r^{-1}(\{(g, t) : t > s, \varphi^M(g_{\restriction [0,s]}) > \frac{1}{2} - \varepsilon\})$$

is open, and therefore defines an $\mathcal{F}^+$-stopping time by Lemma 4.7. Due to lower semi-continuity we have $E \subseteq \rrbracket\kappa, \infty\llbracket$. Therefore, $(\kappa, A) \in \mathsf{Cov}(K)$.

By Proposition 4.11, there is a continuous martingale ζ which almost surely equals φ^M at all stopping times. This allows us to deduce that

$$\mathbb{E}[\varphi_0^M] = \mathbb{E}[\varphi_\kappa^M] = \mathbb{E}[\zeta_\kappa] \geq \left(\frac{1}{2} - \varepsilon\right)\mathbb{W}(\kappa < \tau).$$

As $\varepsilon > 0$ is arbitrary, this implies

$$\inf_{(\kappa,A)\in\mathsf{Cov}(K)} \Big(\mathbb{W}(\kappa < \tau) + \nu(A)\Big) \leq 2 \sup_{\pi\in\mathsf{JOIN}(\tau,\nu)} \pi(K).$$

To show the other inequality, take an open set $O \supseteq \{\kappa < \tau\}$ satisfying $\mathbb{W}(O) \leq 2\mathbb{W}(\kappa < \tau)$. Define a martingale $\bar\varphi^M$ by putting

$$\bar\varphi(\omega) = \mathbb{1}_O(\omega).$$

As O is open $\bar\varphi$ is lower semi-continuous and $\bar\varphi^M$ is an S - lower semi-continuous martingale and we have $\mathbb{E}\bar\varphi \leq 2\mathbb{W}(\kappa < \tau)$. Take an open set $\bar A \supseteq A$ satisfying $\nu(\bar A) \leq 2\nu(A)$. Then, $(\bar\varphi, \mathbb{1}_{\bar A}) \in \mathsf{DC}''$ and we can conclude

$$\frac{1}{2} \sup_{\pi\in\mathsf{JOIN}(\tau,\nu)} \pi(K) \leq \inf_{(\kappa,A)\in\mathsf{Cov}(K)} \Big(\mathbb{W}(\kappa < \tau) + \nu(A)\Big).$$

□

7.3. **Proof of Lemma 7.3.** Recall that we assume that the stopping time τ is smaller than or equal to some number t_0. Moreover, recall that for an $\mathcal{F}^+$ stopping time κ we can identify the set $\rrbracket\kappa, \infty\llbracket$ with a right complete Borel set in S by Lemma 4.7.

Proof of Lemma 7.3. By Corollary 7.8, for each $\varepsilon > 0$ there exist an $\mathcal{F}^+$-stopping time κ and a set $N \subseteq \mathcal{Y}$ such that $B \subseteq (\rrbracket\kappa, \infty\llbracket \times\mathcal{Y}) \cup (S \times N)$ and $\mathbb{W}(\kappa < \tau) + \nu(N) \leq 2\varepsilon$. Put $D = \rrbracket\kappa, \infty\llbracket$ and note that $\bar\tau(D) + \nu(N) \leq 2\varepsilon$.

Fix $\eta > 0$ and pick for each k some right complete set $D_k \subseteq S$ and sets $N_k \subseteq \mathcal{Y}$ such that $B \subseteq (D_k \times \mathcal{Y}) \cup (S \times N_k)$ and $\bar\tau(D_k) + \nu(N_k) \leq \eta 2^{-k}$. Then setting $\tilde D = \bigcap_k D_k$ which is still right complete we get

$$\begin{aligned} B \;&\subseteq \textstyle\bigcap_k \Big((D_k \times \mathcal{Y}) \cup \Big(S \times \big(\bigcup_j N_j\big)\Big)\Big) \\ &\subseteq \Big(\tilde D \times \mathcal{Y}\Big) \cup \Big(S \times \big(\textstyle\bigcup_j N_j\big)\Big). \end{aligned}$$

This shows that D, N can be chosen so that $\bar\tau(D) = 0$ and $\nu(N) < \varepsilon$, for any $\varepsilon > 0$. Similarly, taking a sequence of such right complete sets D'_k and sets N'_k such that $\bar\tau(D'_k) = 0$ and $\nu(N'_k) < \eta 2^{-k}$, we see that

$$B \subseteq \bigcup_k \Big(D'_k \times \mathcal{Y}\Big) \cup \left(S \times \left(\bigcap_l N'_l\right)\right).$$

The desired conclusion follows upon taking the right complete set $D = \bigcup_k D'_k$ and $N = \bigcap_k N'_k$. □

8. Embeddings in abundance

In this section, we will show that all existing solutions to (OptSEP) can be established by Theorem 7.2. Moreover, we will give further examples to demonstrate how new embeddings as well as higher dimensional versions of classical embeddings can be constructed using the monotonicity principle.

8.1. **A secondary maximisation result.** In certain cases, it is useful to identify particular solutions as the solutions not only to a primary optimisation result, but to further identify the unique optimiser within this class of a second maximisation problem in order to resolve possible non-uniqueness of a maximiser. To this end, we begin by making the following definition: suppose $\gamma : S \to \mathbb{R}$ is a Borel-measurable function, we write $\mathsf{Opt}(\gamma,\mu)$ for the set of optimisers of $P_\gamma(\mathbb{W},\mu)$. Observe that, when $P_\gamma(\mathbb{W},\mu) < \infty$, and the map $\pi \mapsto \int \gamma \, d\pi$ is upper semi-continuous the set $\mathsf{Opt}(\gamma,\mu)$ is a closed subset of $\mathsf{RST}(\mathbb{W},\mu)$, and hence also compact.

Theorem 8.1. *Let γ, γ' be Borel measurable functions on S. Suppose that $\mathsf{Opt}(\gamma,\mu) \neq \emptyset$ and that $\nu \in \mathsf{Opt}(\gamma,\mu)$ is an optimiser of:*

$$P_{\gamma'|\gamma}(\mathbb{W},\mu) = \sup_{\tilde\nu \in \mathsf{Opt}(\gamma,\mu)} \int \gamma' d\tilde\nu. \tag{8.1}$$

Then there exists a Borel set $\Gamma \subseteq S$ such that $\nu(\Gamma) = 1$ and

$$(\mathsf{BP}_\nu \cup \mathsf{SBP}_\nu) \cap (\Gamma^< \times \Gamma) = \emptyset, \tag{8.2}$$

where

$$\mathsf{SBP}_\nu = \Big\{((f,s),(g,t)) : f(s) = g(t), \tag{8.3}$$
$$\int \gamma(f \oplus \omega_{\restriction[0,r]}, s+r)\, d\nu^{(f,s)}(\omega,r) + \gamma(g,t) = \gamma(f,s) + \int \gamma(g \oplus \omega_{\restriction[0,r]}, t+r)\, d\nu^{(f,s)}(\omega,r),$$
$$\int \gamma'(f \oplus \omega_{\restriction[0,r]}, s+r)\, d\nu^{(f,s)}(\omega,r) + \gamma'(g,t) < \gamma'(f,s) + \int \gamma'(g \oplus \omega_{\restriction[0,r]}, t+r)\, d\nu^{(f,s)}(\omega,r)\Big\}.$$

Proof. We will show that $\pi(\mathsf{BP}_\nu \cup \mathsf{SBP}_\nu) = 0$ for all $\pi \in \mathsf{JOIN}(\tau_n,\nu)$ with τ_n defined as in the proof of Theorem 7.2. Then the very same proof as for Theorem 7.2 applies again. Hence, the result follows by the following straighforward variant of Proposition 6.5. □

Proposition 8.2. *Let ν be a randomized stopping time which maximizes* (8.1). *Assume that $\pi \in \mathsf{JOIN}(\tau,\nu)$ (where τ can be arbitrary) satisfies*

$$H^{\mathsf{proj}_X(\pi)}(f,s) > 0 \quad \Longrightarrow \quad H^\nu(f,s) > 0 \quad \textit{for } (f,s) \in S.$$

Then we have $\pi(\mathsf{BP}_\nu \cup \mathsf{SBP}_\nu) = 0$.

Proof. As $\nu \in \mathsf{Opt}(\gamma,\mu)$ we only have to show that $\pi(\mathsf{SBP}_\nu) = 0$ by Proposition 6.5. However, by the very same construction as in the proof of Proposition 6.5 we can again argue by contradiction proving the claim. Indeed, we only have to evaluate the integrals of γ and γ' as in (6.9) and (6.10), sum them up, use the assumptions and derive a contradiction. □

Theorem 8.1 will be most useful if we combine it with the following Lemma which allows us to find a particularly nice set Γ reflecting many properties of Brownian motion. To this end take $\nu \in \mathsf{Opt}(\gamma,\mu)$ and put

$$U_0 := \Big\{(f,s) \in S : H^\nu(f,s) > 0, \int \omega(t)\, d\nu^{(f,s)}(\omega,t) \neq 0\Big\}$$
$$U_1 := \Big\{(f,s) \in S : H^\nu(f,s) > 0, \int \omega(t)^2\, d\nu^{(f,s)}(\omega,t) - \Big(\int \omega(t)\, d\nu^{(f,s)}(\omega,t)\Big)^2 = 0\Big\},$$

and $U := U_0 \cup U_1$.

Take the sequence of stopping times $(\tau_n)_{n\in\mathbb{N}}$ as in the proof of Theorem 7.2. Due to the defining martingale property of PRST we have:

Lemma 8.3. *For every $\pi \in \mathsf{PRST}_{\tau_n}$ it holds that $\pi(U) = 0$.*

Proof. We will prove that $\pi(U_i) = 0$ for $i = 0, 1$ seperately. Let us start with U_0. We will argue by contradiction. Put

$$U_0^+ := \left\{(f,s) \in S \,:\, H^\nu(f,s) > 0,\ \int \omega(t)\, d\nu^{(f,s)}(\omega,t) > 0\right\}$$

and assume that there is $\pi \in \mathsf{PRST}_{\tau_n}$ such that $\pi(U_0^+) > 0$. By Lemma 4.19, we can assume that $\pi \in \mathsf{RST}_{\tau_n}$. Let ρ be the representation of π given by Theorem 4.12 (4) and α the representation of ν. Then, there exists a set $A \in \bar{\mathcal{F}}_\rho$ with $\bar{\mathbb{P}}(A) > 0$ such that

$$A_S := \{r(\omega, \rho(\omega,u)), (\omega,u) \in A\} \subseteq U_0^+.$$

Define $M_t := \bar{B}_t\, \bar{\mathbb{E}}[\mathbb{1}_A | \bar{\mathcal{F}}_t]$ where $\bar{B}_t(\omega,u) = B_t(\omega)$. Then, $(M_{\rho+s})_{s\geq 0}$ is an $\bar{\mathcal{F}}_{\rho+s}$ martingale as $A \in \bar{\mathcal{F}}_\rho$. Define $\kappa := \rho \vee \alpha$. Applying optional stopping twice, we get $\bar{\mathbb{E}} M_\kappa = \bar{\mathbb{E}} M_\rho$. However,

$$\bar{\mathbb{E}}(M_\kappa - M_\rho) = \int_{A_S} H^\nu(f,s) \int \omega(t)\, d\nu^{(f,s)}(\omega,t)\, d\pi((f,s)) > 0,$$

by the choice of A and κ. This is a contradiction.

Let us turn to U_1. By the first part of the proof we can assume that $\int \omega(t)\, d\nu^{(f,s)}(\omega,t) = 0$. Also, as before we assume that there is $\pi \in \mathsf{RST}_{\tau_n}$ with $\pi(U_1) > 0$ and consider the representations ρ and α of π and ν respectively. We find again a set $A^1 \in \bar{\mathcal{F}}_\rho$ such that $\bar{\mathbb{P}}(A^1) > 0$ and

$$\{r(\omega, \rho(\omega,u)), (\omega,u) \in A^1\} \subseteq U_1.$$

Considering $M_t^1 := 1_{A^1}(\bar{B}_t^2 - t)$ with $\bar{B}_t$ as above and applying optional stopping twice, with ρ and $\kappa = \rho \vee \alpha$ we get $\bar{\mathbb{E}} M_\rho^1 = \bar{\mathbb{E}} M_\kappa^1$. However, as $\pi \leq \tau_n$ we also have using the definition of U_1

$$\bar{\mathbb{E}}[M_\kappa^1 - M_\rho^1] = \bar{\mathbb{E}}[\mathbb{1}_{A^1}(\rho - \kappa)] < 0,$$

deriving a contradiction. □

Then another application of Lemma 7.3 gives:

Proposition 8.4. *Assume we are in the situation of Theorem 8.1. Then there is a set* $\Gamma \subseteq S$ *such that* (8.2) *holds and*

$$U \cap \Gamma^< = \emptyset. \tag{8.4}$$

Proof. Either by Corollary 7.8 or by applying Lemma 7.3 with $\mathcal{Y} = \{y\}$ and probability measure δ_y for any n we get the existence of a left complete set L_n such that

$$U \cap L_n = \emptyset$$

and $\bar{\tau}_n(L_n) = 1$. Proceeding as in the proof of Theorem 7.2 yields the claim. □

8.2. **Recovering classical embeddings.**

Theorem 8.5 (The Azéma-Yor embedding, cf. [4])**.** *Let* $\varphi : \mathbb{R}_+ \to \mathbb{R}$ *be a bounded, strictly increasing right-continuous function. There exists a stopping time* τ_{AY} *which maximises*

$$\mathbb{E}\left[\varphi\left(\sup_{t\leq\tau} B_t\right)\right]$$

over the set $\mathsf{RST}(\mu)$*, and which is of the form* $\tau_{AY} = \inf\left\{t > 0 : B_t \leq \psi\left(\sup_{s\leq t} B_s\right)\right\}$ *a.s., for some increasing function* ψ*.*

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 $|f|^* = \sup_{r\leq s} |f(r)|$.

Proof. Fix another bounded and strictly increasing continuous function $\tilde{\varphi} : \mathbb{R}_+ \to \mathbb{R}$. Observe that we can define S- upper-semi-continuous functions $\gamma((f,s)) = \varphi(\bar{f})$ and $\gamma'((f,s)) = -\tilde{\varphi}(\bar{f})(f(s))^2$. By the assumptions on φ and the second moment condition on μ, both $P_\gamma(\mathbb{W},\mu)$ and $P_{\gamma'|\gamma}(\mathbb{W},\mu)$ are finite, and hence optimisers exist. Choose one such optimiser, which we denote τ_{AY} (although this may be in $\mathsf{RST}(\mu)$). We can therefore apply Proposition 8.4 to obtain the existence of a set $\Gamma \subseteq S$ such that (8.2) and (8.4) hold.

Define

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

and write $\tau_{\text{CL}}, \tau_{\text{OP}}$ for the first entrance times of the process $(\bar{B}_t(\omega), B_t(\omega))$ into 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. First observe that τ_{AY} is supported in Γ, so $(\bar{B}_{\tau_{AY}}, B_{\tau_{AY}}) \in \Gamma$ $\mathbb{W}$-a.s., and hence $\tau_{\text{CL}} \le \tau_{AY}$ $\mathbb{W}$-a.s..

We now suppose for a contradiction that $\tau_{AY} > \tau_{\text{OP}}$ with positive probability. In particular, we observe that there exist paths $(f,s) \in \Gamma^<$ such that $(\bar{f}, f(s)) \in \mathcal{R}_{\text{OP}}$. Consider first the case where $(\bar{f}, f(s)) \in \{(m,x) : \exists (g,t) \in \Gamma, \bar{g} < m, g(t) = x\}$. Then we can find $(g,t) \in \Gamma$ such that $g(t) = f(s)$ and $\bar{f} > \bar{g}$. In particular, consider the effect of transferring the bush $\tau_{\text{OP}}^{(f,s)}$ onto (g,t). This will certainly not worsen γ, and will strictly improve the average, unless no paths in the bush have a maximum greater than $\bar{g}$. Hence, $((f,s),(g,t)) \in \mathsf{BP}_{\tau_{AY}}$ unless no paths have a maximum above $\bar{g}$. However, in the latter case, it is immediate that we are in $\mathsf{SBP}_{\tau_{AY}}$.

So it remains to argue in the case where there exists $(f,s) \in \Gamma^<$ such that $(\bar{f}, f(s)) \in \{(m,m) : \exists (g,t) \in \Gamma, \bar{g} = m, g(t) = m\}$. We choose $(g,t) \in \Gamma$ such that $g(t) = \bar{g} = \bar{f}$. Since $(f,s) \in \Gamma^<$, we may assume $H^{\tau_{AY}}((f,s)) > 0$, and hence there is a $(f',s') \in \Gamma^<$ such that $(f',s') \succ (f,s)$, $f(s) = f'(s')$ and $\bar{f}' > \bar{f}$. It follows that $((f',s'),(g,t)) \in \mathsf{BP}_{\tau_{AY}}$, and therefore we deduce that $\tau_{\text{CL}} \le \tau_{AY} \le \tau_{\text{OP}}$ $\mathbb{W}$-a.s..

Finally, we define

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

It follows from the definition of $\mathcal{R}_{\text{CL}}$ that $\psi(m)$ is increasing, right-continuous, and $\tau_{\text{CL}} = \inf\{t \ge 0 : B_t \le \psi(\bar{B}_t)\}$. Moreover, we see that

$$\mathcal{R}_{\text{CL}} \setminus \mathcal{R}_{\text{OP}} \subseteq \{(m,x) : x \le \psi(m) < m, \psi(m-) \ne \psi(m)\} \cup \{(m,\psi(m)) : \psi(m) < m, \psi(m-) = \psi(m)\}.$$

However, $\{m : \psi(m-) \ne \psi(m)\}$ is countable, and hence by standard properties of Brownian motion, a Brownian motion never hits the first of these two sets. On the other hand, if $\psi(m) < m$, then a Brownian motion started at $\psi(m)$ immediately enters $(-\infty, \psi(m))$, and so $\tau_{\text{CL}} = \tau_{\text{OP}}$ a.s.. □

Theorem 8.6 (The Jacka Embedding, cf. [26])**.** *Let $\varphi : \mathbb{R}_+ \to \mathbb{R}$ be a bounded, strictly increasing right-continuous function. There exists a stopping time τ_J which maximises*

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

over the set $\mathsf{RST}(\mu)$, and which is of the form

$$\tau_J = \inf\left\{t > 0 : B_t \ge \gamma_-\left(\sup_{s \le \tau} |B_s|\right) \text{ or } B_t \le \gamma_+\left(\sup_{s \le \tau} |B_s|\right)\right\}$$

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

Proof. The proof runs along similar lines to the proof of Theorem 8.5, once we define

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

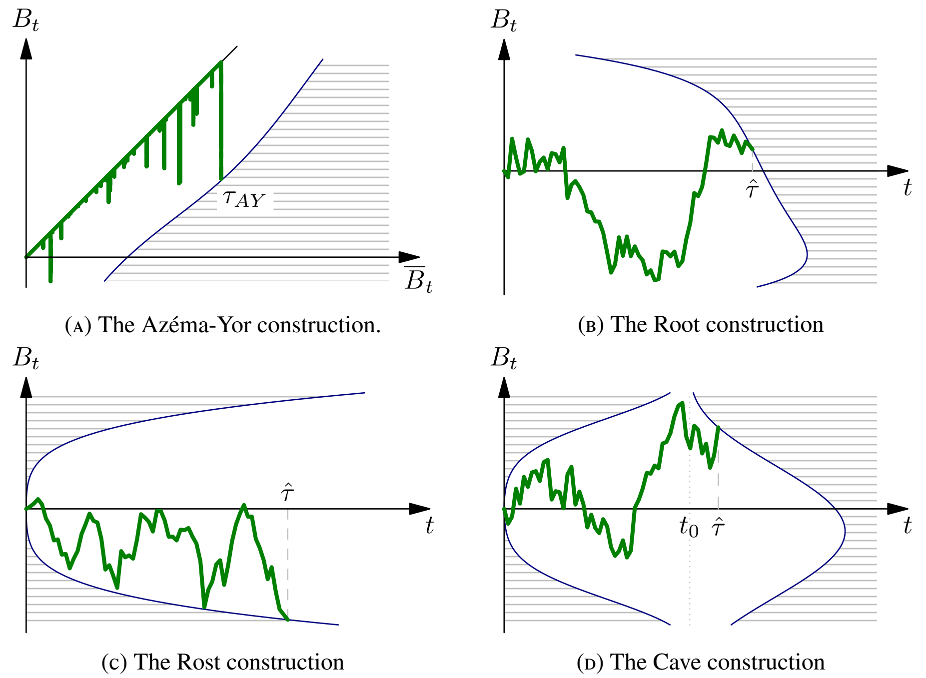

FIGURE 3. Graphical representations of the Azéma-Yor, Root, Rost and Cave constructions.

and then take

$$\gamma_-(m) = \inf\{x : \exists (m,x) \in \mathcal{R}_{\mathrm{CL}}\}$$
$$\gamma_+(m) = \sup\{x : \exists (m,x) \in \mathcal{R}_{\mathrm{CL}}\}.$$

□

Remark 8.7. *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 8.8 (The Perkins Embedding, cf. [36])**.** *Suppose $\mu(\{0\}) = 0$. Let $\varphi : \mathbb{R}_+^2 \to \mathbb{R}$ be a bounded function which is strictly increasing and left-continuous in both arguments. There exists a stopping time τ_P which minimises*

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

over the set $\mathsf{RST}(\mu)$*, and which is of the form $\tau_P = \inf\left\{t > 0 : B_t \notin \left(\gamma_+(\bar{B}_t), \gamma_-(\underline{B}_t)\right)\right\}$ a.s., for some decreasing functions $\gamma_+(m)$ and $\gamma_-(i)$ which are left- and right-continuous respectively.*

Proof. Our primary objective function will of course be to maximise $\int \gamma((f,s))\, d\xi$, where $\gamma((f,s)) = -\varphi(\bar{f}, -\underline{f})$; observe that this is an upper-semi-continuous function on S. We again introduce a secondary maximisation problem: specifically, we consider the functional $\gamma'((f,s)) = (f(s))^2\tilde{\varphi}(\bar{f}), -\underline{f})$ for some bounded continuous and strictly increasing function $\tilde{\varphi} : \mathbb{R}_+^2 \to \mathbb{R}$. As above, we observe that both the primary and secondary problems are well-posed, and hence admit maximisers, and we can apply Theorem 8.1. Take Γ to be the resulting set supporting an arbitrary optimiser which we denote τ_P.

Observe from the fact that $\mu(\{0\}) = 0$ that we can always restrict Γ to the set of points such that $\underline{f} < 0 < \bar{f}$. We first show that Γ can be assumed to contain no points of the type: $\{(f,s) : \underline{f} < f(s) < \bar{f}\}$. This is straightforward from the secondary monotonicity principle, since if there were such a point (f,s) with $f(s) = x$ say, then we must also have passed through x on the way to set the most recent extremum (i.e. either between setting the current minimum and the current maximum, or vice versa). Then there exists $(g,t) \in \Gamma^<$ such that $g(t) = x$ and $\varphi(\bar{g}, -\underline{g}) < \varphi(\bar{f}, -\underline{f})$, contradicting Theorem 8.1.

Now consider the sets:

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

and their respective hitting times, $\tau_{\text{CL}}, \tau_{\text{OP}}$. It is immediate that $\tau_{\text{CL}} \leq \tau_P$ a.s., and an essentially identical argument to that used in the proof of Theorem 8.5 gives $\tau_P \leq \tau_{\text{OP}}$ a.s..

We now set

$$\begin{aligned}
\gamma_+(m) &= \sup\{x < 0 : (m,x) \in \underline{\mathcal{R}}_{\text{CL}}\} \\
\gamma_-(i) &= \inf\{x > 0 : (i,x) \in \bar{\mathcal{R}}_{\text{CL}}\}.
\end{aligned}$$

Then the 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

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

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

Theorem 8.9 (Maximizing the range)**.** *Let $\varphi : \mathbb{R}_+^2 \to \mathbb{R}$ be a bounded function which is strictly increasing and right-continuous in both variables. There exists a stopping time τ_{xr} which maximises*

$$\mathbb{E}\left[\varphi\left(\sup_{t \leq \tau} B_t, -\inf_{t \leq \tau} B_t\right)\right]$$

over the set $\mathsf{RST}(\mu)$*, and which is of the form*

$$\tau_{xr} = \inf\left\{t > 0 : (\underline{B}_t, \bar{B}_t, B_t) \in \left(\gamma_-(\underline{B}_t, \bar{B}_t), \gamma_+(\underline{B}_t, \bar{B}_t)\right)\right\} \text{ a.s.},$$

for some right-continuous functions $\gamma_-(i,m)$ increasing in i decreasing in m and $\gamma_+(i,m)$ increasing in m decreasing in i.

Proof. We again introduce a secondary maximisation problem: specifically, we consider the functional $\gamma'((f,s)) = -\tilde{\varphi}(\bar{f}, \underline{f}) f(s)^2$ for some bounded continuous strictly increasing function $\tilde{\varphi} : \mathbb{R}_+^2 \to \mathbb{R}$. As above, we observe that both the primary and secondary problems are well-posed, and hence admit maximisers, and we can apply Theorem 8.1. Take Γ to be the resulting set supporting an arbitrary optimiser which we denote τ_{xr}.

We put $I(\underline{b}, \bar{b}) := \mathsf{conv}(\{g(t) : (g,t) \in \Gamma, \underline{g} = \underline{b}, \bar{g} = \bar{b}\})$, where conv denotes the convex hull. Moreover, we set $\gamma_-(\underline{b}, \bar{b}) = \min\{x : x \in I(\underline{b}, \bar{b})\}$ and $\gamma_+(\underline{b}, \bar{b}) = \max\{x : x \in I(\underline{b}, \bar{b})\}$. If $\gamma_-(\underline{b}, \bar{b}) = \underline{b}$ we set $\gamma_-(\underline{d}, \bar{d}) = \underline{d}$ for all $\underline{d} \leq \underline{b}$ and $\bar{d} \geq \bar{b}$ and anologously for γ_+.

We claim that γ_+ is increasing in $\bar{b}$ and decreasing in $\underline{b}$ and γ_- is decreasing in $\bar{b}$ and increasing in $\underline{b}$, i.e. $I(\underline{b}, \bar{b}) \subseteq I(\underline{d}, \bar{d})$ if $\underline{d} \leq \underline{b}$ and $\bar{b} \leq \bar{d}$. Assume the contrary, then with the notation as before there is $x \in I(\underline{b}, \bar{b}) \setminus I(\underline{d}, \bar{d})$. W.l.o.g. we can assume that $x = \gamma_-(\underline{b}, \bar{b}) > \underline{b}$. Then there is $(f,s) \in \Gamma^<, (g,t) \in \Gamma$ with $\underline{g} = \underline{b}, \bar{g} = \bar{b}, g(t) = \gamma_-(\underline{b}, \bar{b}) = f(s)$ and $\underline{f} = \underline{d}, \bar{f} = \bar{d}$.

Now consider the effect of transferring the bush $\tau_{xr}^{(f,s)}$ onto (g,t). As for the Azéma-Yor embedding, this will not worsen γ, and it will strictly improve the average unless no paths in the bush have a minimum which is less or equal to $\underline{g} = \underline{b}$. In the former case we have $((f,s),(g,t)) \in \mathsf{BP}_{\tau_{xr}}$ and in the latter case we have $((f,s),(g,t)) \in \mathsf{SBP}_{\tau_{xr}}$.

Putting

$$\mathcal{R}_{\text{CL}} = \{(\underline{b}, \bar{b}, b) : b \in I(\underline{b}, \bar{b})\}$$

and

$$\mathcal{R}_{\text{OP}} = \{(\underline{b}, \bar{b}, b) : \exists \underline{d} \geq \underline{b}, \bar{d} \leq \bar{b} \text{ one inequality being strict, s.t. } b \in I(\underline{d}, \bar{d})\}$$

with respective hitting times τ_{CL} and τ_{OP} we can deduce similarly as in the Azéma-Yor embedding that $\tau_{\text{CL}} \leq \tau_{rx} \leq \tau_{\text{OP}}$ a.s. and also that $\tau_{\text{CL}} = \tau_{\text{OP}}$ a.s. and the conclusion follows. □

Remark 8.10. *Considering the last argument we see that looking for minimizers of the range the picture turns inside out. The stopping region will be of the form* $(-\infty, a) \cup (b, \infty)$ *and we can directly deduce that we only stop in a minimum or a maximum, i.e. the three dimensional picture reduces to a two dimensional picture and we are back in the Perkins case. This symmetry is similar to the symmetry in the Root and Rost embedding.*

Remark 8.11. *We observe that, in the case of Theorem 8.9, the characterisation provided would not appear to be sufficient to identify the functions* γ_+, γ_- *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: consider the Azéma-Yor embedding: from Theorem 8.5 it is clear that if we stop with maximum above s, then we must hit s before stopping, and there exists* $y = \psi(s)$ *such that we stop at or above y if our maximum is above s. Moreover, we only stop above y if our maximum is at least s. We conclude that, conditional on hitting s before stopping, we must embed* μ *restricted to* $[y, \infty)$ *(with some more care needed if there are atoms of* μ*). However there is a unique y such that this distribution has mean s, and this y must then be* $\psi(s)$*.*

8.3. **The Vallois-embedding and optimizing functionals of local time.** In this section we shall be interested to determine the stopping rule which solves

$$\sup\{\mathbb{E}[h(L_\tau)] : \tau \text{ solves (SEP)}\}, \tag{8.5}$$

where L the denotes the local time of Brownian motion in 0 and h is a convex or concave function.

A large part of the argument is virtually identical to the argument which we used in the previous section. The most involved part will in fact be to show that the problem (8.5) admits a maximizer. As mentioned below Definition 4.3, L is not S- continuous. Nevertheless we will prove the following result.

Lemma 8.12. *Let* $\xi_n, n \geq 1, \xi \in \mathsf{RST}(\mu)$ *and assume that* $\xi_n \to \xi$ *weakly. Then* $L_{\xi_n} \to L_\xi$ *weakly. In fact, if* ρ_n, ρ *are the representations of* ξ_n, ξ *on* $\bar{\Omega}$*, then* $\bar{L}_{\rho_n} \to \bar{L}_\rho$ *in* $L^1(\bar{\Omega}, \mathbb{P} \otimes \lambda)$*.*

We first give a simple result on the connection between convergence of stopping times in $\mathsf{RST}(\mu)$ and their representations.

Proposition 8.13. *Let* ξ_n, ξ, ρ_n *and* ρ *be as in Lemma 8.12. Then* $\xi_n \to \xi$ *weakly iff* $\rho_n \to \rho$ *in probability.*

Proof. Let $X \in C_b(C(\mathbb{R}_+) \times \mathbb{R}_+)$. Recall that

$$\int X_t(\omega)\, d\xi(\omega, t) = \int \mathbb{W}(d\omega) \int \xi_\omega(dt) X_t(\omega) = \int \mathbb{W}(d\omega) \int X(\omega, t)\, dA_t^\xi(\omega) = \tag{8.6}$$

$$= \int \mathbb{W}(d\omega) \int \lambda(dx) X_{\rho(x,\omega)}(\omega) = \bar{\mathbb{E}} X_\rho, \tag{8.7}$$

and hence

$$\int X_t(\omega)\, d(\xi - \xi_n)(\omega, t) = \int \mathbb{W}(d\omega) \int \lambda(dx)\, [X_{\rho(x,\omega)}(\omega) - X_{\rho_n(x,\omega)}(\omega)]. \tag{8.8}$$

Considering processes which depend only on the time t but not ω, i.e. $X_t(\omega) = X_t$, we obtain that $\xi_n \to \xi$ weakly implies that $\rho_n \to \rho$ in probability. Conversely, if $\rho_n \to \rho$

in probability under $\bar{\mathbb{P}}$, then also $\rho_n \to \rho$ almost surely along some subsequence of every subsequence. By dominated convergence, $\xi_n \to \xi$ weakly. □

Proof of Lemma 8.12. As a consequence of Proposition 8.13 we have that $\rho_n \wedge \rho \to \rho, \rho_n \vee \rho \to \rho$. Note also that for every minimal embedding $\xi' \in \mathsf{RST}(\mu')$, $\mathbb{E}_{\xi'} L = \int |x|\, d\mu'(x)$. Write μ_n for the law embedded by $\rho_n \wedge \rho$. Then $\mu_n \to \mu$ weakly, and $L_{\rho_n \wedge \rho} \leq L_\rho$, so (using Lemma 4.20) $\bar{\mathbb{E}} L_{\rho_n \wedge \rho} = \int |x|\, d\mu_n \to \int |x|\, d\mu$ and hence $\bar{\mathbb{E}} L_{\rho_n \wedge \rho} \to \bar{\mathbb{E}} L_\rho$. This implies that $L_{\rho_n \wedge \rho} \to L_\rho$ in $L^1(\bar{\Omega}, \mathbb{P} \otimes \lambda)$. Since $L_{\rho_n \vee \rho} + L_{\rho_n \wedge \rho} = L_\rho + L_{\rho_n}$ we also find that $\bar{\mathbb{E}} L_{\rho_n \vee \rho} = \bar{\mathbb{E}}\left[L_{\rho_n} + (L_\rho - L_{\rho_n})_+\right] = \bar{\mathbb{E}} L_\rho + \bar{\mathbb{E}}\left[(L_\rho - L_{\rho_n})_+\right] \to \bar{\mathbb{E}} L_\rho$, where we used that $\xi_n, \xi \in \mathsf{RST}(\mu)$. Thus $L_{\rho_n \vee \rho} \to L_\rho$ in $L^1(\bar{\Omega}, \mathbb{P} \otimes \lambda)$. Combining these results, we see that $L_{\rho_n} \to L_\rho$ in $L^1(\bar{\Omega}, \mathbb{P} \otimes \lambda)$. □

With these tools, we are now able to show:

Theorem 8.14. *Let $h : [0, \infty) \to \mathbb{R}$ be a bounded, strictly concave function.*

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

$$\mathbb{E}\left[h\left(L_\tau\right)\right]$$

over the set $\mathsf{RST}(\mu)$, and which is of the form $\tau_{V-} = \inf\left\{t > 0 : B_t \notin \left(\varphi_-\left(L_t\right), \varphi_+\left(L_t\right)\right)\right\}$ a.s., for some decreasing, non-negative function φ_+ and increasing, non-positive function φ_-.

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

$$\mathbb{E}\left[h\left(L_\tau\right)\right]$$

over the set $\mathsf{RST}(\mu)$, and which is of the form $\inf\left\{t > 0 : B_t \notin \left(\varphi_-\left(L_t\right), \varphi_+\left(L_t\right)\right)\right\} \wedge Z$ a.s., for some increasing, non-negative function φ_+, decreasing, non-positive function φ_-, and an $\mathcal{F}_0$-measurable random variable Z with $\mathbb{P}(Z = 0) = \mu(0)$.

Proof. We consider the second case, the first case being slightly simpler. We will apply Theorem 8.1 to the optimisations corresponding to: $\gamma((\omega, t)) = -h(L_t(\omega))$ and $\gamma'((\omega, t)) = -e^{-L_t(\omega)} B_t^2(\omega)$.

By Lemma 8.12 we observe that, if $\xi_n \in \mathsf{RST}(\mu)$ is a sequence of stopping times such that $V^* = \lim \mathbb{E}[h(L_{\xi_n})] = \sup\{\mathbb{E}[h(L_\tau)] : \tau \in \mathsf{RST}(\mu)\}$, then (possibly passing to a subsequence) $\xi = \lim_n \xi_n$ satisfies $\mathbb{E}[h(L_\xi)] = V^*$. Hence, the set $\mathsf{Opt}(\gamma, \mu)$ is non-empty and closed and we can apply Theorem 8.1. Write $\tau_{V+} = \xi$.

By an application of Theorem 4.1 and Proposition 4.4, there is a function $L : S \to \mathbb{R}_+$ such that for $\mathbb{W}$-a.e. $\omega \in C(\mathbb{R}_+)$ it holds that $L_t(\omega) = L(\omega_{\restriction[0,t]}, t)$. Define the sets

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

It follows immediately that $\tau_{\text{CL}} \leq \tau_{V+}$. Moreover, by the monotonicity principle, we immediately observe that $(l, 0) \notin \mathcal{R}_{\text{OP}}$ for any $l \geq 0$, and $\tau_{V+} \leq \tau_{\text{OP}}$. It follows that $\tau_{V_+}(\Omega \times \{0\}) = \mu(\{0\})$. We consider τ_{V+} on $\{\tau_{V+} \geq \eta\}$, for $\eta > 0$. Write $\tau^\eta_{\text{OP}} = \inf\{t \geq \eta : (L_t, B_t) \in \mathcal{R}_{\text{OP}}\}$ and $\tau^\eta_{\text{CL}} = \inf\{t \geq \eta : (L_t, B_t) \in \mathcal{R}_{\text{CL}}\}$. Then on this set $\tau^\eta_{\text{CL}} \leq \tau_{V+} \leq \tau^\eta_{\text{OP}}$. Moreover, define $\varphi_+(l) = \inf\{x > 0 : (l, x) \in \mathcal{R}_{\text{OP}}\}$ and $\varphi_-(l) = \sup\{x < 0 : (l, x) \in \mathcal{R}_{\text{OP}}\}$. Observe that, since $\tau_{V+} \leq \tau_{\text{OP}}$, if $\mathbb{P}(\tau_{V+} > \eta) > 0$, then $|\varphi_+(\eta) - \varphi_-(\eta)| > 0$. Moreover, $\varphi_+(l)$ is clearly right-continuous, and increasing, so it must have at most countably many discontinuities, and similarly for $\varphi_-(l)$. Moreover, we can write

$$\inf\left\{t \geq \eta : B_t \notin \left(\varphi_-\left(L_t\right), \varphi_+\left(L_t\right)\right)\right\} \leq \tau^\eta_{\text{CL}} \leq \tau^\eta_{\text{OP}} \leq \inf\left\{t \geq \eta : B_t \notin \left[\varphi_-\left(L_t+\right), \varphi_+\left(L_t+\right)\right]\right\}$$

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 $\varphi_+(l)$ and $-\varphi_-(l)$ are bounded away from zero on $[\eta, \infty)$). It follows that $\tau_{V+} = \inf\left\{t \geq \eta : B_t \notin \left(\varphi_-\left(L_t\right), \varphi_+\left(L_t\right)\right)\right\}$ on $\{\tau_{V+} \geq \eta\}$, and since $\eta > 0$ was arbitrary, we get the desired behaviour. □

Remark 8.15. *The arguments above extend from local time at* 0 *to a general continuous additive functional A. Writing L^x for local time in x, A can be represented in the form $A_t := \int_0^t L_s^x \, dm_A(x)$. Let f be a convex function such that $f'' = m_A$ in the sense of distributions. If $\int f \, d\mu < \infty$, then Proposition 8.13 still holds with A in place of L; the above proof is easily adapted to the more general situation.*

In this manner, we deduce the existence of optimal solutions to (SEP) *for functionals depending on A. By analogy with Theorem 8.14 this can be used to generate (inverse-/cave-) barrier-type embeddings of various kinds. Other generalisations and variants may be considered in a similar manner. We leave specific examples as an exercise for the reader.*

Remark 8.16. *In Cox and Obłój* [15]*, embeddings are constructed which maximise certain double-exit probabilities: for example, to maximise the probability that both $\bar{B}_\tau \geq \bar{b}$ and $\underline{B}_t \leq \underline{b}$, 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 characterise 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 maximisation problem, to show that there exists an optimiser demonstrating the behaviour characterising the Cox-Obłój embeddings. (Specifically, if we write $H_b((f,s)) = \inf\{t \leq s : f(t) = b\}$, $\underline{H} = H_{\underline{b}} \wedge H_{\bar{b}}$ and $\bar{H} = H_{\underline{b}} \vee H_{\bar{b}}$ then the secondary maximisation problem $\gamma'((f,s)) = ((f(s) - \underline{H}((f,s)))^2 1_{\underline{H} \leq s}/2 - ((f(s) - \bar{H}((f,s)))^2 1_{\bar{H} \leq s}$ is sufficient to rederive the form of these embeddings.)*

8.4. **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 where the Brownian motion is in $\mathbb{R}^d$, for $d \geq 2$ and we have a general initial distribution. In $\mathbb{R}^d$, since the Brownian motion is transient, it is no longer straightforward to assert the existence of an embedding. In general, [42] 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.[6] There exists a stopping time τ such that $B_0 \sim \lambda$ and $B_\tau \sim \mu$ if and only if for all $x \in \mathbb{R}^d$

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

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

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

As a result, it is not straightforward to give simple criteria for the existence of a solution in $\mathsf{RST}(\mu)$. However, assuming that we do have such a solution, then we are able to state the following:

Theorem 8.17. *Suppose $\mathsf{RST}(\mu)$ is non-empty. If h is a strictly convex function and $\hat{\tau} \in \mathsf{RST}(\mu)$ maximises $\mathbb{E}[h(\tau)]$ 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 2.2, except we no longer show that $\tau_{\text{CL}} = \tau_{\text{OP}}$. Generally, 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}

[6]For $d = 2$, it would seem that the same condition, but with $u(x,y) = -\ln|x-y|$ would also be a natural condition for the existence of an embedding, however this is not so immediate from Rost's results, and we are unaware of an explicit proof of this result.

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 $\lambda(D) = 1 = \mu(\complement D)$.

Proof of Theorem 8.17. The first part of the proof proceeds similarly to the proof of Theorem 2.2. In particular, we define the sets $\mathcal{R}_{\text{CL}}, \mathcal{R}_{\text{OP}}$ and the stopping times $\tau_{\text{CL}}, \tau_{\text{OP}}$ as in the proof of this result. We fix $\delta > 0$, and consider the set $\{\hat{\tau} \geq \delta\}$.

For $\eta \geq 0$, we then define $B_t^{-\eta} = B_{t+\eta}$, for $t \geq -\eta$ and also set

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

Then $\tau_{\text{CL}}^{\eta,\delta} \geq \delta$, and for any $\varepsilon > 0$, we can choose $\eta > 0$ sufficiently small that

$$d_{TV}(B_\delta^{-\eta}, B_\delta) < \varepsilon$$

and hence from the Strong Markov property of Brownian motion, it follows that

$$d_{TV}(B^{-\eta}_{\tau_{\text{CL}}^{\eta,\delta}}, B_{\tau_{\text{CL}}^{0,\delta}}) < \varepsilon.$$

In particular, we have weak convergence of the law of $B^{-\eta}_{\tau_{\text{CL}}^{\eta,\delta}}$ to the law of $B_{\tau_{\text{CL}}^{0,\delta}}$ as $\eta \to 0$.

It also follows that

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

so $\tau_{\text{CL}}^{\eta,\delta} \geq \tau_R^{0,\delta}$, and moreover, $\tau_{\text{CL}}^{\eta,\delta} \to \tau_{\text{OP}}^{0,\delta}$ a.s. as $\eta \to 0$. Hence, $B^{-\eta}_{\tau_{\text{CL}}^{\eta,\delta}} \to B_{\tau_{\text{OP}}^{0,\delta}}$ in probability, as $\eta \to 0$, so we have weak convergence of the law of $B^{-\eta}_{\tau_{\text{CL}}^{\eta,\delta}}$ to the law of $B_{\tau_{\text{OP}}^{0,\delta}}$, and hence $B_{\tau_{\text{OP}}^{0,\delta}} \sim B_{\tau_{\text{CL}}^{0,\delta}}$.

We now observe that, by an essentially identical argument to that in the proof of Theorem 2.2, we must have: $\tau_{\text{CL}}^{0,\delta} \leq \hat{\tau} \leq \tau_{\text{OP}}^{0,\delta}$ on $\{\hat{\tau} \geq \delta\}$. However, in the argument above, we know that $\tau_{\text{CL}}^{0,\delta} \leq \hat{\tau} \leq \tau_{\text{OP}}^{0,\delta}$, and $\tau_{\text{CL}}^{\eta,\delta} \to_{\mathcal{D}} \tau_{\text{CL}}^{0,\delta}$ and $\tau_{\text{CL}}^{\eta,\delta} \to_{\mathcal{D}} \tau_{\text{OP}}^{0,\delta}$ as $\eta \to 0$ (where $\mathcal{D}$ denotes convergence in distribution). It follows that $\tau_{\text{CL}}^{0,\delta} =_{\mathcal{D}} \tau_{\text{OP}}^{0,\delta}$ and hence $\tau_{\text{CL}}^{0,\delta} = \tau_{\text{OP}}^{0,\delta}$ a.s. In particular, $B_{\tau_{\text{CL}}^{0,\delta}} = B_{\tau_{\text{OP}}^{0,\delta}} = B_{\hat{\tau}}$ on $\{\hat{\tau} \geq \delta\}$. Letting $\delta \to 0$ we observe that $\tau_{\text{OP}}^{0,\delta} \to \tau_{\text{OP}}$, and hence the required result holds on taking $\mathcal{R} = \mathcal{R}_{\text{OP}}$. □

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

Theorem 8.18. *Suppose λ, μ are measures in $\mathbb{R}^d$ and $\hat{\tau} \in \mathsf{RST}(\mu)$ maximises $\mathbb{E}[h(\tau)]$ 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 $\mathbb{P}(\hat{\tau} = 0, B_0 \in A) = (\lambda \wedge \mu)(A)$, 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 2.3 to recover 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 2.3 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 $\mathbb{P}(\hat{\tau} = 0, B_0 \in A) \leq (\lambda \wedge \mu)(A)$. If there is strict inequality, then there exist some paths which start at $x \in A$, and paths which stop at x at strictly positive time, violating the monotonicity principle. □

9. Embedding Feller processes

In this section we discuss which changes are needed to extablish the duality result, Theorem 5.1, as well as the monotonicity principle, Theorem 7.2, for continuous Feller processes. In fact, most of our arguments are abstract and do not use any specific structure of the Wiener measure. The relation between the spaces S and $\mathcal{X} = C(\mathbb{R}_+) \times \mathbb{R}_+$ as well as the approximation of stopping times rely on abstract theory of stochastic processes and topological properties of S and $\mathcal{X}$. The proof of Theorem 7.2 uses duality theory of optimal transport and Choquet's theorem. Proposition 4.10 is very valuable to identify certain hitting times as stopping times. To prove the duality statement we use again duality theory of optimal transport and — crucially — the compactness of $\mathsf{PRST}(\mu)$.

This last point, the compactness of $\mathsf{PRST}(\mu)$ and the characterization of minimal stopping times in terms of the expection $\mathbb{E}_\xi[T] = V < \infty$ is in fact the only point where we use specific properties of Brownian motion (apart from Section 8).

So we assume now that we are given a continuous Feller process $Z = (Z_t)_{t\geq 0}$. As usual we assume Z to be the canonical process on the space of continuous functions. We write $(\mathbb{P}_x)_{x\in\mathbb{R}}$ for the law of the Feller process started in x and $\mathbb{P}$ for the law of the process started with law λ.

We define the set PRST of pseudo-randomized stopping times as before with $\mathbb{P}$ replacing $\mathbb{W}$. Let $\mu \in \mathcal{P}(\mathbb{R})$. We say that $\xi \in \mathsf{PRST}$ is a minimal embedding of μ if ξ embeds μ in Z and if for any $\bar\xi \in \mathsf{PRST}$ with $\bar\xi \leq \xi$ $\mathbb{P}$-a.s. and also embedding μ it holds that $\bar\xi = \xi$ $\mathbb{P}$-a.s.

Definition 9.1. *For $\mu \in \mathcal{P}(\mathbb{R})$ we define $\mathsf{PRST}(\mu)$ to be the set of all minimal pseudo-randomized stopping times embedding the measure μ, and $\mathsf{RST}(\mu) = \mathsf{PRST}(\mu) \cap \mathsf{RST}$.*

We observe that, by taking optional projections, the set $\mathsf{RST}(\mu)$ is equivalently the subset of RST which embeds μ, and such that there is no strictly smaller $\bar\xi \in \mathsf{PRST}$, or in RST embedding μ.

Assumption 9.2. *From now on we* assume *that the set $\mathsf{PRST}(\mu)$ is non-empty, compact, and either:*

(1) That there exists an increasing, $\mathcal{F}^0$-optional process $\zeta : \mathcal{X} \to \mathbb{R}$ with $\zeta_s \to \infty$ $\mathbb{P}$-a.s. as $s \to \infty$ such that the following hold true:
- *For a finite $\xi \in \mathsf{PRST}$ with $Z_\xi \sim \mu$ we have $\mathbb{E}[\zeta_\xi] < \infty$ if and only if ξ is minimal.*
- *There is a corresponding S-continuous martingale $X_t = h(Z_t) - \zeta_t$ such that $X_{t\wedge\xi}$ is uniformly integrable for all $\xi \in \mathsf{RST}(\mu)$.*

or

(2) That $\xi \in \mathsf{PRST}$ and $B_\xi \sim \mu$ implies ξ is minimal (i.e. all embeddings are minimal).

Below we will verify that this assumption is satisfied in a number of natural examples. Note that compactness of $\mathsf{PRST}(\mu)$ is equivalent to the existence of an increasing and diverging function $G : \mathbb{R}_+ \to \mathbb{R}$ such that

$$\sup_{\xi\in\mathsf{PRST}(\mu)} \mathbb{E}_\xi[G(T)] =: V_G < \infty.$$

We first note that 9.2 (1) is also relevant in the usual Brownian setup, where it allows us to dispose of the second moment condition.

Proposition 9.3. *Let Z be Brownian motion and assume that λ and μ are in convex order. Then Assumption 9.2 (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

$$\zeta_t(\omega) = \frac{1}{2}\int_0^t F''(\omega_s)\,ds$$

and note that by Ito's formula and our conditions on F,

$$X_t = F(Z_t) - \frac{1}{2}\int_0^t F''(Z_s)\,ds = F(Z_t) - \zeta_t$$

is an S- continuous martingale.

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

Assume now that $Z_\xi \sim \mu$ and that $(Z_{t\wedge\xi})_{t\geq 0}$ is uniformly integrable. Then for each t, the law of $Z_{t\wedge\xi}$ is bounded by μ in the convex order and in particular $\mathbb{E}F(Z_{t\wedge\xi}) \leq V$. We obtain

$$\mathbb{E}\zeta_\xi = \lim_{t\to\infty}\mathbb{E}\zeta_{t\wedge\xi} = \lim_{t\to\infty} F(Z_{t\wedge\xi}) - \mathbb{E}F(Z_0) = V - \mathbb{E}F(Z_0) < \infty.$$

Next assume that $\mathbb{E}\zeta_\xi < \infty$. Then $\sup_{t\geq 0}\mathbb{E}F(Z_{t\wedge\xi}) < \infty$, hence $(Z_{t\wedge\xi})_{t\geq 0}$ is uniformly integrable.

To see that $\lim_{t\to\infty}\zeta_t = \infty$, note that $\mathbb{P}\Big(\int_0^\infty \mathbb{1}_{[-1,1]}(Z_t)\,dt = \infty\Big) = 1$ and that F'' is bounded away from 0 on $[-1,1]$.

Finally it remains to show that X is uniformly integrable. To see this we apply again the de la Vallée-Poussin theorem to obtain an increasing, super-linear, convex function $g : \mathbb{R}_+ \to \mathbb{R}_+$ such that $\int g\circ F\,d\mu < \infty$. Since $g\circ F$ is convex we obtain (as above) that $\sup_t \mathbb{E}[g\circ F(Z_{\xi\wedge t})] < \infty$. Thus $(F(Z_{\xi\wedge t}))_{t\geq 0}$ is uniformly integrable and this carries over to X. □

Definition 9.4. *Let $X : C(\mathbb{R}_+) \to \mathbb{R}$ be a measurable function which is bounded or positive. Then we define $\mathbb{E}[X|\mathcal{F}_t^0]$ to be the unique $\mathcal{F}_t^0$-measurable function satisfying*

$$\mathbb{E}[X|\mathcal{F}_t^0](\omega) = \int X((\omega_{|[0,t]})\otimes\omega')\,d\mathbb{P}_{\omega(t)}(\omega').$$

Then the natural analogue of Proposition 4.10 holds by the Feller property of X:

Proposition 9.5. *Let $X \in C_b(C(\mathbb{R}_+))$. Then $X_t(\omega) := \mathbb{E}[X|\mathcal{F}_t^0](\omega)$ defines an S-continuous martingale. We denote this martingale by X^M.*

Proof. By the Feller property, we have for any continuous and bounded function X and any sequence $x_n \to x$ that also $\int X\,d\mathbb{P}_{x_n} \to \int X\,d\mathbb{P}_x$. Combining this with the proof of Proposition 4.10 yields the result. □

This allows us to prove the following duality result:

Theorem 9.6. *Let $\gamma : S \to \mathbb{R}$ be S-upper semi-continuous and bounded from above. Suppose that Assumption 9.2 holds. Put*

$$P_\gamma(\mathbb{P},\mu) := \sup_{\xi\in\mathsf{PRST}(\mu)}\int\gamma\,d\xi = \sup_{\xi\in\mathsf{RST}(\mu)}\int\gamma\,d\xi.$$

Let $\mathsf{DC}(\gamma)$ be the set of all pairs (ψ,φ) where $\psi\in C_b(\mathbb{R})$ and φ is a $\mathbb{P}$-semimartingale with decomposition $\varphi = M^\varphi + A^\varphi$ where M^φ is an S-continuous and bounded $\mathbb{P}$-martingale starting at zero and A^φ is an increasing process satisfying $\sup_{\xi\in\mathsf{RST}(\mu)}\int A^\varphi\,d\xi \leq 0$. Put

$$D_\gamma(\mathbb{P},\mu) := \inf_{(\psi,\varphi)\in\mathsf{DC}(\gamma)}\int\psi\,d\mu.$$

Then, it holds that $P_\gamma(\mathbb{P},\mu) = D_\gamma(\mathbb{P},\mu)$. Moreover, in case (1) of Assumption 9.2, the process A^φ may be assumed to be zero at the expense of assuming that $M^\varphi_{t\wedge\xi}$ is only uniformly integrable for all $\xi\in\mathsf{RST}(\mu)$.

Proof. Consider first case (2) of Assumption 9.2. Let $G(t)$ be an increasing, diverging function such that $\sup_{\xi\in\mathsf{PRST}(\mu)}\mathbb{E}_\xi[G(T)] =: V_G < \infty$, and note that the set

$$\mathsf{TM}^V(\mathbb{P},\mu) := \{\pi\in\mathcal{P}(C(\mathbb{R}_+)\times\mathbb{R}_+\times\mathbb{R}) : \mathsf{proj}_{C(\mathbb{R}_+)}(\pi) = \mathbb{W}, \mathsf{proj}_{\mathbb{R}}(\pi) = \mu, \mathbb{E}_\pi[G(T)] \leq V_G\}$$

is compact. This allows us to establish the non-adapted duality result. Putting

$$\mathsf{DC}^V(c) := \left\{ (\psi, \varphi) : \begin{array}{l} \varphi \text{ is an } S\text{-continuous bounded } \mathbb{P}\text{-martingale}, \psi \in C_b(\mathbb{R}),\ \exists \alpha \geq 0, \\ \varphi_t(\omega) + \psi(y) + \alpha(G(t) - V) \geq c(\omega, t, y), \text{ for all } \omega \in \Omega, y \in \mathbb{R}, t \in \mathbb{R}_+ \end{array} \right\}$$

we can derive the corresponding version of Proposition 5.6. Finally, we have to note that $\alpha(G(t) - V)$ is an increasing process as claimed to deduce that

$$D_c^{C(\mathbb{R}_+)\times\mathbb{R}_+\times\mathbb{R}}(\mathbb{P}, \mu) \geq D_\gamma(\mathbb{P}, \mu)$$

proving the claim.

To show case (1) of Assumption 9.2, we argue along the lines above, noting that we can replace the condition $\mathbb{E}_\pi[G(T)] \leq V$ by $\mathbb{E}[\zeta_\pi] \leq V$, so

$$\mathsf{TM}^V(\mathbb{P}, \mu) := \{\pi \in \mathcal{P}(C(\mathbb{R}_+) \times \mathbb{R}_+ \times \mathbb{R}) : \mathsf{proj}_{C(\mathbb{R}_+)}(\pi) = \mathbb{W}, \mathsf{proj}_{\mathbb{R}}(\pi) = \mu, \mathbb{E}[\zeta_\pi] \leq V\}.$$

Observe that this set is compact: suppose $\varepsilon > 0$. Since $\zeta_t \to \infty$, $\mathbb{P}$-a.s., we deduce the existence of a $K > 0$ such that $\mathbb{P}(\zeta_K < 2V/\varepsilon) < \varepsilon/2$. Hence $\mathbb{E}[\zeta_\pi] \leq V$ implies $\pi(T \geq K) < \varepsilon$, and TM^V is indeed compact.

Finally, observe that we can use the S-continuous martingale $h(\omega(t)) - \zeta_t$ to replace $\omega(t)^2 - t$ in (5.16) to deduce that

$$D_c^{C(\mathbb{R}_+)\times\mathbb{R}_+\times\mathbb{R}}(\mathbb{P}, \mu) \geq D_\gamma(\mathbb{P}, \mu).$$

□

Apart from the abstract theory the main ingredients in the proof of Theorem 7.2 are Proposition 6.5 to apply Lemma 7.3 and Proposition 4.10 to identify certain hitting times as $\mathcal{F}^+$-stopping times in the proof of Lemma 7.3. In fact, Proposition 6.5 is still valid and we just proved the analogue of Proposition 4.10, namely Proposition 9.5. Then, the very same proof yields the following:

Theorem 9.7. *Assume that $\gamma : S \to \mathbb{R}$ is Borel-measurable, the optimization problem (5.3) is well-posed and that $\nu \in \mathsf{RST}(\mu)$ is an optimizer of $P_\gamma(\mathbb{P}, \mu)$. Then ν is supported by a γ-monotone set Γ.*

9.1. **Examples: One-dimensional Diffusions.** Let Z_t be a regular (time-homogenous) one-dimensional diffusion on an interval $I \subseteq \mathbb{R}$, with inaccessible or absorbing endpoints (see [40] for the relevant definitions and terminology) and $Z_0 \sim \lambda$, some $\lambda \in \mathcal{P}(I)$. In particular, Z_t is a Feller process ([40, Proposition V.50.1]). Then (on a possibly enlarged probability space) there exists a scale function $s(x)$ 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 stopping time $\xi \in \mathsf{RST}(\mu; Z)$ if and only if there exists a stopping time $\xi' \in \mathsf{RST}(s(\mu); B)$ such that $\xi'([\![0, \tau_{s(I)}]\!]) = 1$, where $\tau_{s(I)} = \inf\{t \geq 0 : B_t \notin s(I^\circ)\}$. Write A_t^{-1} to be the inverse of A_t, so $A_{A_t}^{-1} = t$. We now consider three cases:

- Suppose $s(I^\circ) = (a, b)$ for $a, b \in \mathbb{R}$. Then it follows from [10, Theorems 17 and 22] that $\mathsf{RST}(\mu; Z)$ is non-empty if and only if $s(\lambda)$ precedes $s(\mu)$ in convex order, and in fact, any $\xi \in \mathsf{RST}(Z)$ with $Z_\xi \sim \mu$ is minimal (so $\xi \in \mathsf{RST}(\mu, Z)$).
- Suppose $s(I^\circ) = (a, \infty)$ for $a \in \mathbb{R}$, and $s(\lambda), s(\mu)$ are integrable measures. Write $m_\lambda = \int s(y)\lambda(dy)$, and $m_\mu = \int s(y)\mu(dy)$. Then it follows from Theorems 17 and 22 and the discussion at the top of p. 245 of [10] that $\mathsf{RST}(\mu; Z)$ is non-empty if and only if $-\int |s(y) - x|\mu(dy) \leq -\int |s(y) - x|\lambda(dy) + (m_\lambda - m_\mu)$ for all $x \geq a$, or equivalently, that $\int (s(y) - x)_+\mu(dy) \leq \int (s(y) - x)_+\lambda(dy)$ for all $x \geq a$. Again, any $\xi \in \mathsf{RST}(Z)$ with $Z_\xi \sim \mu$ is minimal. By symmetry, similar results for the case where $s(I^\circ) = (-\infty, b)$ for $b \in \mathbb{R}$ can be given.
- Suppose $s(I^\circ) = (-\infty, \infty)$, and $\int (s(y))^2\lambda(dy), \int (s(y))^2\mu(dy) < \infty$. Then we are in the classical case, and a stopping time $\xi \in \mathsf{RST}(Z)$ is minimal if and only if $Z_\xi \sim \mu$, and $\mathbb{E}[A_\xi^{-1}] < \infty$. In particular, compactness of $\mathsf{RST}(\mu; Z)$ follows

directly from compactness of $\mathsf{RST}(s(\mu); B)$. Further, if the scale function s is suitably differentiable, one can show that $X_t = s(Z_t)^2 - A_t^{-1}$ is an S-continuous martingale, and under the condition that $\int (s(y))^2 \lambda(dy), \int (s(y))^2 \mu(dy) < \infty$, we deduce that $X_{t\wedge\xi}$ is uniformly integrable for all $\xi \in \mathsf{RST}(\mu; Z)$. In particular, (1) of Assumption 9.2 is satisfied.

More generally, when only the integrals $\int s(y)\lambda(dy)$ and/or $\int s(y)\mu(dy)$ are finite, we are in the situation of Proposition 9.3 and Assumption 9.2 (1) is satisfied.

Remark 9.8. *Observe that none of the constructions described in Section 8.2 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 one-dimensional diffusions described above.*

9.1.1. *Brownian motion with drift.* Let $Z_t = B_t + dt$ for some $d < 0$ with $Z_0 \sim \lambda$, and $I = (-\infty, \infty)$. Then a possible choice of the scale function is $s(x) = \exp(-2dx)$, and $s(I^\circ) = (0, \infty)$. Let $\lambda, \mu \in \mathbb{P}(\mathbb{R})$ be such that $s(\lambda), s(\mu)$ are integrable, and suppose

$$\int (\exp(-2dy) - x)_+ \mu(dy) \le \int (\exp(-2dy) - x)_+ \lambda(dy),$$

for all $x \ge 0$. By the arguments above, there exists an embedding and all stopping times embedding μ are minimal. Then the set $\mathsf{PRST}(\mu)$ is compact as can be seen by the following estimate inserted in the proof of Theorem 4.23. Fix $\varepsilon > 0$ and take $K > 0$ such that $\mu((-\infty, -K)) \le \varepsilon/4$. Then there is $R > 0$ such that

$$\mathbb{P}_\lambda(\{\exists\, R' \ge R : Z_{R'} \ge -K\}) \le \varepsilon/4.$$

Then $\xi \in \mathsf{PRST}(\mu)$ implies that $\xi(T > R) \le \varepsilon/2$.

9.1.2. *Geometric Brownian motion.* Let Z_t be a geometric Brownian motion and μ be concentrated on the *positive reals* $(0, \infty)$. Then the compactness of $\mathsf{PRST}(\mu)$ follows from the compactness in the case of Brownian motion with drift as $\exp : \mathbb{R} \to (0, \infty)$ is a homeomorphism. Similarly, conditions for the existence and minimality of $\xi \in \mathsf{PRST}(\mu)$ follow directly from the case of Brownian motion with drift, or more generally, from the observation that Z_t is a regular diffusion.

9.1.3. *Three-dimensional Bessel process.* Let $Z_t = \|B_t\|$ for a three-dimensional Brownian motion $(B_t)_{t\ge 0}$ (or d-dimensional with $d \ge 3$) with $Z_0 \sim \lambda$. Let $\mu \in \mathcal{P}((0, \infty))$ be such that there exists at least one embedding. Then any embedding is minimal and $\mathsf{PRST}(\mu)$ is compact. This can be seen by similar argument as in the case of Brownian motion with drift as B_t is transient in dimension three and higher. Indeed, fix $\varepsilon > 0$ and take $K > 0$ such that $\mu((K, \infty)) \le \varepsilon/4$. By the transience of B_t there is $R > 0$ such that

$$\mathbb{P}(\exists\, R' \ge R \;:\; Z_{R'} \le K) \le \varepsilon/4,$$

which implies that $\xi(T > R) \le \varepsilon/2$ implying the compactness of $\mathsf{PRST}(\mu)$ by a straightforward modification of Theorem 4.23.

9.1.4. *Ornstein-Uhlenbeck Processes.* Let Z_t be an Ornstein-Uhlenbeck process, given for example as the solution to the SDE $dZ_t = -Z_t\,dt + dW_t$. Then Z_t is a regular diffusion on $I = (-\infty, \infty)$ with scale function given (up to constants) by $s'(x) = \exp(x^2)$. Then $s(I^\circ) = (-\infty, \infty)$. Suppose λ, μ are measures on $\mathbb{R}$ with $s(\lambda), s(\mu)$ square integrable, and in convex order. Then $\mathsf{RST}(\mu; Z)$ is compact and $\xi \in \mathsf{RST}(\mu; Z)$ if and only if $Z_\xi \sim \mu$ and $\mathbb{E}[A_\xi^{-1}] < \infty$.

9.1.5. *The Hoeffding-Frechet Coupling as a very particular Root Solution.* Let Z_t be the deterministic process given by $dZ_t = dt$ started in $Z_0 \sim \mu$. Let ν be another probability and assume for simplicity that $\max \operatorname{supp} \mu \leq \min \operatorname{supp} \nu$. Then the problem to minimize $\mathbb{E}[\tau^2]$ is of course solved by the Root solution. 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 a most simple setup. I.e. 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 μ and ν.

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] Jacques Azéma and Marc Yor. Une solution simple au probleme de skorokhod. In *Séminaire de probabilités XIII*, pages 90–115. Springer, 1979.

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

[6] 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.

[7] 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.

[8] B. Bouchard and M. Nutz. Arbitrage and Duality in Nondominated Discrete-Time Models. *Ann. Appl. Probab., to appear*, May 2013.

[9] L. Campi and W. Schachermayer. A super-replication theorem in Kabanov's model of transaction costs. *Finance Stoch.*, 10(4):579–596, 2006.

[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. G. Hobson. Skorokhod embeddings, minimality and non-centred target distributions. *Probability Theory and Related Fields*, 135:395–414, September 2005.

[12] A. M. G. Cox and G. Peskir. Embedding Laws in Diffusions by Functions of Time. *ArXiv e-prints*, January 2012.

[13] A. M. G. Cox and J. Wang. Root's barrier: Construction, optimality and applications to variance options. *Ann. Appl. Probab., to appear*, April 2012.

[14] A. M. G. Cox and J. Wang. Optimal robust bounds for variance options. *ArXiv e-prints*, August 2013.

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

[16] Claude Dellacherie and Paul-André Meyer. *Probabilities and potential*, volume 29 of *North-Holland Mathematics Studies*. North-Holland Publishing Co., Amsterdam, 1978.

[17] Y. Dolinsky and H. Mete Soner. Robust Hedging with Proportional Transaction Costs. *Finance Stoch., to appear*, February 2013.

[18] Y. Dolinsky and H. M. Soner. Martingale optimal transport and robust hedging in continuous time. *Probab. Theory Relat. Fields, to appear*, pages 1–37, 2013.

[19] 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.

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

[21] P. Gassiat, A. Mijatovic, and H. Oberhauser. An integral equation for root's barrier and the generation of brownian increments. *Annals of Applied Probability*, To Appear.

[22] 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.

[23] D. Hobson. Robust hedging of the lookback option. *Finance and Stochastics*, 2:329–347, 1998. 10.1007/s007800050044.

[24] D. Hobson. The Skorokhod embedding problem and model-independent bounds for option prices. volume 2003 of *Lecture Notes in Math.*, pages 267–318. Springer, Berlin, 2011.

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

[26] SD Jacka. Doob's inequalities revisited: A maximal h¡ sup¿ 1¡/sup¿-embedding. *Stochastic processes and their applications*, 29(2):281–290, 1988.

[27] H. G. Kellerer. Duality theorems for marginal problems. *Z. Wahrsch. Verw. Gebiete*, 67(4):399–432, 1984.

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

[29] M. Knott and C. Smith. On Hoeffding-Fréchet bounds and cyclic monotone relations. *J. Multivariate Anal.*, 40(2):328–334, 1992.

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

[31] Günter Last, Peter Mörters, and Hermann Thorisson. Unbiased shifts of Brownian motion. *Ann. Probab.*, 42(2):431–463, 2014.

[32] 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.

[33] A. Nikeghbali and M. Yor. A definition and some characteristic properties of pseudo-stopping times. *The Annals of Probability*, 33(5):1804–1824, September 2005.

[34] H. Oberhauser and G. dos Reis. Root's barrier, viscosity solutions of obstacle problems and reflected FBS-DEs. *ArXiv e-prints*, January 2013.

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

[36] Edwin Perkins. The cereteli-davis solution to the h1-embedding problem and an optimal embedding in brownian motion. In *Seminar on stochastic processes, 1985*, pages 172–223. Springer, 1986.

[37] A. Pratelli. On the sufficiency of c-cyclical monotonicity for optimality of transport plans. *Math. Z.*, 258(3):677–690, 2008.

[38] 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.

[39] R.T. Rockafellar. Characterization of the subdifferentials of convex functions. *Pacific J. Math.*, 17:497–510, 1966.

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

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

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

[43] 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.

[44] 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.

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

[46] L. Rüschendorf. On c-optimal random variables. *Statist. Probab. Lett.*, 27(3):267–270, 1996.

[47] W. Schachermayer and J. Teichmann. Characterization of optimal transport plans for the Monge-Kantorovich problem. *Proc. Amer. Math. Soc.*, 137(2):519–529, 2009.

[48] 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.

[49] 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.

[50] P. Vallois. Le probleme de skorokhod sur ir: une approche avec le temps local. In *Séminaire de Probabilités XVII 1981/82*, pages 227–239. Springer, 1983.

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

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