

## Stopping with expectation constraints: 3 points suffice

Stefan Ankirchner\*    Nabil Kazi-Tani<sup>†</sup>    Maike Klein<sup>‡</sup>  
 Thomas Kruse<sup>§</sup>

### Abstract

We consider the problem of optimally stopping a one-dimensional regular continuous strong Markov process with a stopping time satisfying an expectation constraint. We show that it is sufficient to consider only stopping times such that the law of the process at the stopping time is a weighted sum of 3 Dirac measures. The proof uses recent results on Skorokhod embeddings in order to reduce the stopping problem to a linear optimization problem over a convex set of probability measures.

**Keywords:** optimal stopping; expectation constraint; Skorokhod embedding problem; one-dimensional strong Markov processes; extreme points of sets of probability measures.

**AMS MSC 2010:** 60G40; 60J60; 60B05.

Submitted to EJP on September 23, 2018, final version accepted on April 29, 2019.

## 1 Introduction

Let  $(Y_t)_{t \in \mathbb{R}_{\geq 0}}$  be a one-dimensional regular continuous strong Markov process with respect to a right-continuous filtration  $(\mathcal{F}_t)$ . In the sequel we use the term “general diffusions” as a synonym for these processes. Let  $f: \mathbb{R} \rightarrow \mathbb{R}$  be measurable and denote by  $\mathcal{T}(T)$  the set of  $(\mathcal{F}_t)$ -stopping times such that  $E[\tau] \leq T \in \mathbb{R}_{\geq 0}$ . In the following we consider the optimal stopping problem

$$\text{maximize } E[f(Y_\tau)] \quad \text{subject to } \tau \in \mathcal{T}(T). \tag{1.1}$$

The problem (1.1) arises whenever an average time constraint applies for any stopping rule. If a process has to be stopped repeatedly and independently of the previous stopping times, then it is reasonable to impose an average time constraint instead of a sharp constraint of the type  $\tau \leq T$ , a.s. For example think of the question of when to stop searching for a parking space. If you face this question every morning when driving

\*University of Jena, Germany. E-mail: [s.ankirchner@uni-jena.de](mailto:s.ankirchner@uni-jena.de)

<sup>†</sup>Université Lyon 1, France. E-mail: [nabil.kazi-tani@univ-lyon1.fr](mailto:nabil.kazi-tani@univ-lyon1.fr)

<sup>‡</sup>TU Wien, Austria. E-mail: [maike.klein@fam.tuwien.ac.at](mailto:maike.klein@fam.tuwien.ac.at)

<sup>§</sup>University of Gießen, Germany. E-mail: [thomas.kruse@math.uni-giessen.de](mailto:thomas.kruse@math.uni-giessen.de)

to your work, it is more likely that you impose an average constraint on your searching time than just a sharp upper bound.

Notice that there is no deterministic dependence of the constraint on time. For solving the stopping problem (1.1) one needs to turn the expectation constraint into a path-dependent constraint.

In this article we show that for the stopping problem (1.1) it is sufficient to consider only stopping times  $\tau$  such that the law of  $Y_\tau$  is a weighted sum of at most 3 Dirac measures. Any such stopping time can be interpreted as a composition of exit times from intervals.

We also show that in general a reduction to weighted sums of 2 Dirac measures is not possible. In particular, one can not split the state space into a deterministic stopping and continuation region. This is in contrast to stopping problems with a sharp bound on the stopping time and to stopping problems with infinite time horizon and discounting.

Our idea for proving a reduction to 3 Dirac measures is to rewrite the stopping problem (1.1) as a linear optimization problem over a set of probability measures. To this end we use recent results on the Skorokhod embedding problem characterizing the set  $\mathcal{A}(T)$  of probability distributions that can be embedded into  $Y$  with stopping times having expectation smaller than or equal to  $T$  (see [1] and [13]). As for standard linear problems the maximal value of the optimization is attained by extreme points. The extreme points of  $\mathcal{A}(T)$  turn out to be contained in the set of probability measures that can be written as weighted sums of at most 3 Dirac measures.

To the best of our knowledge, the idea of using Skorokhod embeddings to directly solve optimal stopping problems first appeared in [30], where the authors deal with an optimal stopping problem for the geometric Brownian motion, under the Choquet integral, and where the only condition imposed on the stopping times is that they are almost surely finite. Connections between specific optimal stopping problems and the Skorokhod embedding problem have already been observed and examined earlier, see e.g. [17, 21, 19, 12]. When it comes to optimal stopping problems with constraints on the stopping time distribution, the literature is rather scarce: The seminal book by Shiryaev [27] discusses in Section 4.3 and 4.4 versions of the quickest detection problem with probability constraints. Kennedy [15] solves an optimal stopping problem with an expectation constraint for a discrete time process. In [27] and in [15] the authors use Lagrangian techniques to reduce the constrained problems to unconstrained ones. Within a continuous time setting, the article [2] formulates a dynamic programming principle for stopping problems with expectation constraints and derives a verification theorem. Bayraktar and Yao [5] provide a proof of the dynamic programming principle and characterize the value function of the stopping problem as the unique viscosity solution of the associated fully non-linear Hamilton-Jacobi-Bellman equation. Different constraints have been recently studied: Bayraktar and Miller [4] consider the problem of optimally stopping the Brownian motion with a stopping time whose distribution is atomic with finitely many points of mass. Miller [18] analyzes stopping problems with time inconsistent constraints. In [6], the authors use optimal transport techniques to treat the problem of optimally stopping the Brownian motion with a stopping time having a fixed specified distribution. Further stopping problems with an expectation constraint on the stopping time have been solved by Urusov [28]. Let  $\theta \in [0, 1]$  be the moment at which a standard Brownian motion attains its maximal value on  $[0, 1]$  and let  $\alpha \geq 0$ . Then Urusov [28] characterizes the stopping time that minimizes  $E[(\tau - \theta)^+]$  over all stopping times  $\tau$  satisfying the expectation constraint  $E[(\tau - \theta)^-] \leq \alpha$ . Shiryaev [26] determines the stopping time  $\tau$  minimizing  $E[(\tau - \theta)^+]$  among all stopping times satisfying the probability constraint  $P[\tau < \theta] \leq \alpha$ . Likewise, Shiryaev [26] solves a variant of this stopping problem where  $\theta$  is replaced by the last time before 1 when the

Brownian motion visits zero.

The article is organized as follows. In Section 2 we describe the precise setting of the stopping problem considered. In Section 3 we show how one can reduce the stopping problem to an optimization over the set of probability measures that are weighted sums of 3 Dirac measures. Finally, in Section 4 we provide sufficient conditions for the existence of an optimal stopping time.

## 2 Stopping after consecutive exit times

In this section we rigorously set the framework for the optimal stopping problem. The process to stop is assumed to be a one-dimensional regular continuous strong Markov process (general diffusion). Let the state space  $J \subseteq \mathbb{R}$  be an open, half-open or closed interval and denote by  $\mathring{J} := (l, r)$  the interior of  $J$ , where  $-\infty \leq l < r \leq \infty$ . Moreover, denote by  $\bar{J}$  the closure of  $J$  in  $\mathbb{R}$ . Let  $\Omega = \mathcal{C}([0, \infty), J)$  be the space of all continuous  $J$ -valued functions and  $(Y_t)_{t \in \mathbb{R}_{\geq 0}}$  be the coordinate process, i.e.  $Y_t(\omega) = \omega(t)$ ,  $t \in \mathbb{R}_{\geq 0}$ ,  $\omega \in \Omega$ . Let  $\mathcal{F}_t^0$  be the  $\sigma$ -algebra generated by  $(Y_s)_{s \leq t}$  and  $\mathcal{F}^0 := \mathcal{F}_\infty^0 := \bigvee_{t \in \mathbb{R}_{\geq 0}} \mathcal{F}_t^0$ . Denote by  $(\theta_t)_{t \in \mathbb{R}_{\geq 0}}$  the family of shift operators on  $\Omega$  defined by  $(\theta_t \omega)(s) = \omega(t + s)$ ,  $s \in \mathbb{R}_{\geq 0}$ . Let  $(P^x)_{x \in J}$  be a family of probability measures on  $(\Omega, \mathcal{F}^0)$  that is a regular diffusion in the sense of [25, Chapter V.45]. In particular, we have  $P^x[Y_0 = x] = 1$  for all  $x \in J$ . Regularity means that for every  $y \in \mathring{J}$  and  $x \in J$  we have that  $P^y[\tau_x < \infty] > 0$ , where  $\tau_x = \inf\{t \in \mathbb{R}_{\geq 0} : Y_t = x\}$ . Here and in the sequel we use the convention that  $\inf \emptyset = \infty$ .

For a probability measure  $\nu$  on  $(J, \mathcal{B}(J))$  let

$$P^\nu(A) := \int P^x(A) \nu(dx), \quad A \in \mathcal{F}^0.$$

Let  $\mathcal{F}^\nu$  be the completion of  $\mathcal{F}^0$  with respect to  $P^\nu$  and set  $\mathcal{F}_t^\nu = \sigma(\mathcal{F}_t^0, \mathcal{N})$ ,  $t \in \mathbb{R}_{\geq 0}$ , where  $\mathcal{N}$  denotes the collection of  $P^\nu$ -null sets in  $\mathcal{F}^\nu$ . One can show that  $(\Omega, \mathcal{F}^\nu, (\mathcal{F}_t^\nu), P^\nu)$  satisfies the usual conditions. We set  $\mathcal{F}_t = \bigcap_{\nu} \mathcal{F}_t^\nu$  and  $\mathcal{F} = \bigcap_{\nu} \mathcal{F}^\nu$ . Observe that  $(\mathcal{F}_t)$  is right-continuous, but that in general  $(\Omega, \mathcal{F}, (\mathcal{F}_t), P^\nu)$  does not satisfy the usual conditions. The process  $(Y_t)_{t \in \mathbb{R}_{\geq 0}}$  fulfills the strong Markov property (cf. Theorem 9.4, Chapter III, in [24]): For any bounded  $\mathcal{F}$ -measurable mapping  $\eta$  and any finite  $(\mathcal{F}_t)$ -stopping time  $\tau$  we have

$$E^\nu[\eta \circ \theta_\tau | \mathcal{F}_\tau] = E^{Y_\tau}[\eta], \quad P^\nu - a.s.$$

Let  $m$  be the speed measure of the diffusion  $(P^x)_{x \in J}$  on  $J$  (see Theorem 3.6 and Definition 3.7 in Chapter VII of [23]). Since  $Y$  is regular we have for all  $a < b \in \mathring{J}$

$$0 < m([a, b]) < \infty.$$

Throughout we assume that the diffusion  $Y$  is in natural scale. If  $Y$  is not in natural scale, then there exists a strictly increasing continuous function  $s: J \rightarrow \mathbb{R}$ , the so-called scale function, such that  $s(Y_t)$ ,  $t \in \mathbb{R}_{\geq 0}$ , is in natural scale. In Remark 4.5 below we show how to reduce the case where  $Y$  is not in natural scale to the case where it is in natural scale.

In addition, we assume that if one of the endpoints  $l$  and  $r$  is accessible, then it is absorbing. This implies  $Y$  is a local martingale (see Corollary 46.15 in [25]).

For  $y \in \mathring{J}$  we define  $q_y: \bar{J} \rightarrow [0, \infty]$ ,

$$q_y(x) = \frac{1}{2} m(\{y\}) |x - y| + \int_y^x m((y, u)) du, \quad x \in \bar{J}, \tag{2.1}$$

with the convention that  $m((y, u)) = -m((u, y))$  whenever  $u < y$ . Moreover, we set  $q_y(r) := \lim_{x \uparrow r} q_y(x) = \infty$  if  $r = \infty$  and  $q_y(l) := \lim_{x \downarrow l} q_y(x) = \infty$  if  $l = -\infty$ .

Let  $\tau_{l,r} = \inf\{t \in \mathbb{R}_{\geq 0} : Y_t \notin (l, r)\}$ . One can show that  $q_y(Y_t) - (t \wedge \tau_{l,r})$ ,  $t \in \mathbb{R}_{\geq 0}$ , is a local martingale with respect to  $P^y$  and  $(\mathcal{F}_t)$  (see Theorem 2.1 in [3]). Moreover, the behavior of  $q_y$  at  $l$  and  $r$  determines whether the process attains the boundary points with a positive probability or not.

**Lemma 2.1** (see Theorem 3.3 in [3]). *Let  $y \in \overset{\circ}{J}$ . We have  $q_y(r) < \infty$  if and only if  $r \in J$ . Similarly,  $q_y(l) < \infty$  if and only if  $l \in J$ .*

**Remark 2.2.** Observe that the case where the process to stop is described by a homogeneous stochastic differential equation (SDE) driven by a Brownian motion  $W$  is a special case of the general framework that we set up above. Indeed, let  $b, \eta: \mathbb{R} \rightarrow \mathbb{R}$  be Borel-measurable functions that satisfy  $\eta(x) \neq 0$  for all  $x \in (l, r)$ ,  $\eta(x) = 0$  for all  $x \in \mathbb{R} \setminus (l, r)$  and  $\frac{1+|b|}{\eta^2} \in L^1_{loc}((l, r))$ . Then for all  $y \in (l, r)$  the SDE

$$dY_t = b(Y_t)dt + \eta(Y_t)dW_t, \quad Y_0 = y, \tag{2.2}$$

possesses a weak solution  $(Y, W)$  that is unique in law (see e.g. Theorem 2.11 in [9] or Section 5.5 C in [14]). If  $b \equiv 0$ , then  $Y$  is in natural scale and the speed measure of  $Y$  is given by  $m(dx) = \frac{2}{\eta^2(x)}dx$ . For all  $y \in (l, r)$  the function  $q_y$  satisfies

$$q_y(x) = \int_y^x \int_y^z \frac{2}{\eta^2(u)} du dz, \quad x \in \overset{\circ}{J}.$$

The case where the SDE (2.2) contains a non-zero drift component  $b$  is a special case of the setting considered in Remark 4.5.

Let  $f: J \rightarrow \mathbb{R}$  be a Borel-measurable function determining the payoff of the stopping problem. Throughout we make the following assumption on  $f$ :

**Assumption (A).** For every  $y \in \overset{\circ}{J}$  there exists  $C(y) \in \mathbb{R}_{\geq 0}$  such that

$$f(x) \geq -C(y)(1 + q_y(x)), \quad \forall x \in J. \tag{2.3}$$

For any  $T \in \mathbb{R}_{\geq 0}$ , let  $\mathcal{T}(T, y)$  be the set of all  $(\mathcal{F}_t)$ -stopping times  $\tau$  with  $E^y[\tau] \leq T$ .

**Remark 2.3.** Assumption (A) ensures that the expectation  $E^y[f(Y_\tau)]$  exists for all  $y \in \overset{\circ}{J}$ ,  $T \in \mathbb{R}_{\geq 0}$  and  $\tau \in \mathcal{T}(T, y)$ . Indeed, for an appropriately chosen localizing sequence of stopping times  $(\tau_n)$ , it holds that

$$\begin{aligned} E^y[\{f(Y_\tau)\}^-] &\leq E^y[C(y)(1 + q_y(Y_\tau))] = C(y)(1 + E^y[\liminf_{n \rightarrow \infty} q_y(Y_{\tau \wedge \tau_n})]) \\ &\leq C(y)(1 + \liminf_{n \rightarrow \infty} E^y[q_y(Y_{\tau \wedge \tau_n})]) = C(y)(1 + \liminf_{n \rightarrow \infty} E^y[\tau \wedge \tau_n]) \\ &\leq C(y)(1 + T). \end{aligned}$$

We consider the problem of finding the stopping time in  $\mathcal{T}(T, y)$  that maximizes the expected payoff  $E^y[f(Y_\tau)]$ . The value function is defined by

$$v(T, y) = \sup_{\tau \in \mathcal{T}(T, y)} E^y[f(Y_\tau)] \tag{2.4}$$

for all  $T \geq 0$  and  $y \in J$ . Observe that  $v(0, y) = f(y)$  for all  $y \in J$ . Moreover, for  $y \in J \setminus \overset{\circ}{J}$  it holds true that  $v(T, y) = f(y)$  for all  $T \in \mathbb{R}_{\geq 0}$ , because accessible endpoints are assumed to be absorbing. Therefore, we assume throughout this article that  $y \in \overset{\circ}{J}$ .

**Remark 2.4.** If Assumption (A) is replaced by the stronger assumption that there exists  $C(y) \in \mathbb{R}_{\geq 0}$  such that

$$|f(x)| \leq C(y)(1 + q_y(x)), \quad x \in J, \tag{2.5}$$

then the value function  $v(T, y)$  is finite. Indeed, it follows by using similar arguments as in Remark 2.3 that  $\sup_{\tau \in \mathcal{T}(T, y)} E^y[f(Y_\tau)] \leq C(y)(1 + T)$  in this case. The following example shows that in general one can not dispense with condition (2.5) if we want to guarantee that  $v$  is finite. For a Brownian motion  $Y = W$  we have  $q_0(x) = x^2$ . Let  $f(x) = |x|^{2+\varepsilon}$ ,  $\varepsilon > 0$ , be the payoff function. The first time  $\tau_{-\frac{T}{a}, a}$ ,  $a \in \mathbb{R}_{>0}$ , when  $W$  hits  $a$  or  $-T/a$  has expectation  $T$  under  $P^0$ . Hence,

$$v(T, 0) \geq \sup_{a \in \mathbb{R}_{>0}} E^0[f(W_{\tau_{-\frac{T}{a}, a}})] = \sup_{a \in \mathbb{R}_{>0}} \left\{ a^{2+\varepsilon} \frac{T}{a^2 + T} + \frac{a^2}{a^{2+\varepsilon}} \frac{T^{2+\varepsilon}}{a^2 + T} \right\} = \infty.$$

For stopping problems *without* an expectation constraint an optimal stopping time is given by the exit time of the continuation region (see Corollary 2.9, Chapter I in [22]). In particular, for solving unconstrained stopping problems it is enough to consider exit times from intervals. For constrained stopping problems a reduction to simple exit times is not possible. We show, however, that it is enough to consider at most three consecutive exit times.

To give a precise statement, we denote for  $a, b \in \mathbb{R}$  with  $a \leq b$  the first hitting time of  $a$  by  $\tau_a = \inf\{t \in \mathbb{R}_{\geq 0} : Y_t = a\}$  and the first exit time from the interval  $(a, b)$  after time  $r \in \mathbb{R}_{\geq 0}$  by  $\tau_{a,b}(r) = \inf\{t \in [r, \infty) : Y_t \notin (a, b)\}$ . Moreover, we write  $\mathcal{T}_3(T, y)$  for the collection of stopping times  $\tau \in \mathcal{T}(T, y)$  for which there exist  $p_1, p_2, p_3 \in [0, 1]$  with  $p_1 + p_2 + p_3 = 1$  and  $a, c, d \in \mathbb{R}$  with  $a \leq c \leq d$  such that

$$\tau = \begin{cases} \tau_{a,b}(\tau_{\bar{\mu}}) + \mathbf{1}_{\{Y_{\tau_{a,b}(\tau_{\bar{\mu}})} = b\}} \inf\{t \in \mathbb{R}_{\geq 0} : Y_{t+\tau_{a,b}(\tau_{\bar{\mu}})} \in \{c, d\}\}, & \text{if } \bar{\mu} > a, \\ \tau_a, & \text{if } \bar{\mu} = a, \end{cases}$$

where  $\bar{\mu} = p_1 a + p_2 c + p_3 d$  and  $b = \frac{p_2 c + p_3 d}{1 - p_1}$  if  $\bar{\mu} > a$ . Notice that  $b \in (\max\{c, \bar{\mu}\}, d)$  for  $p_1, p_2, p_3 \in (0, 1)$  and  $c < d$ .

One of our main results is that the stopping problem (2.4) can be simplified to the set  $\mathcal{T}_3(T, y)$ .

**Theorem 2.5.** *We have*

$$v(T, y) = \sup_{\tau \in \mathcal{T}_3(T, y)} E^y[f(Y_\tau)]. \tag{2.6}$$

We prove Theorem 2.5 in the following section. We do so by reducing problem (2.4) to an optimization over a set of probability measures.

Theorem 2.5 brings up the question whether the supremum is attained in  $\mathcal{T}_3(T, y)$ . In Section 4 below we provide sufficient conditions guaranteeing the existence of an optimal stopping time in  $\mathcal{T}_3(T, y)$ .

### 3 Optimal stopping as a measure optimization

In this section we first explain how one can reduce the stopping problem (2.4) to a linear optimization problem over a set of probability measures satisfying some integrability constraints. The linear nature of the measure optimization allows us then to conclude that the maximum values are attained by extreme points, which here are weighted sums of three Dirac measures.

We denote by  $\mathcal{M} = \mathcal{M}(J)$  the set of all probability measures on  $\mathbb{R}$  with support in  $J$  and by  $\mathcal{M}^1$  the set of all measures  $\mu$  in  $\mathcal{M}$  with finite first moment  $\bar{\mu} = \int x \mu(dx)$ . For  $y \in \overset{\circ}{J}$  let  $\mathcal{A}(T, y)$  be the set of measures  $\mu \in \mathcal{M}^1$  satisfying the following properties:

1. (a) If  $l > -\infty$ , then  $\bar{\mu} \leq y$ .
- (b) If  $r < \infty$ , then  $\bar{\mu} \geq y$ .

2.  $\mu$  integrates  $q_y$  such that

$$\int q_y(x)\mu(dx) \leq T - H(y, \bar{\mu}), \tag{3.1}$$

where

$$H(y, \bar{\mu}) = \begin{cases} (y - \bar{\mu})(m((y, \infty)) + \frac{1}{2}m(\{y\})), & \bar{\mu} < y, \\ 0, & \bar{\mu} = y, \\ (\bar{\mu} - y)(m((-\infty, y)) + \frac{1}{2}m(\{y\})), & \bar{\mu} > y. \end{cases}$$

**Remark 3.1.** Observe that the following consequences of the definition of  $\mathcal{A}(T, y)$  hold true. If there exists  $\mu \in \mathcal{A}(T, y)$  such that  $\bar{\mu} > y$ , then it follows that  $l = -\infty$  and that  $m((-\infty, y)) < \infty$ . Indeed, the fact that  $l = -\infty$  follows directly from Condition 1. (a) in this case. For the second claim, suppose on the contrary that  $m((-\infty, y)) = \infty$ . Then it follows from the definition of  $H$  that  $H(y, \bar{\mu}) = \infty$  and hence (3.1) can not be satisfied by  $\mu$ . Consequently,  $m((-\infty, y)) < \infty$ . Similarly, it holds that  $r = \infty$  and that  $m((y, \infty)) < \infty$  if there exists  $\mu \in \mathcal{A}(T, y)$  such that  $\bar{\mu} < y$ .

Results from [13] on the Skorokhod embedding problem for diffusions (and from [1] for processes described in terms of SDEs) imply that  $\mathcal{A}(T, y)$  coincides with the set of probability measures that can be embedded into  $Y$  under  $P^y$  with stopping times  $\tau$  satisfying  $E^y[\tau] \leq T$ . More precisely, we have the following:

**Proposition 3.2.** *Let  $\mu \in \mathcal{M}$ . There exists a stopping time  $\tau \in \mathcal{T}(T, y)$  with  $Y_\tau \sim \mu$  under  $P^y$  if and only if  $\mu \in \mathcal{A}(T, y)$ .*

*Proof.* Let  $\tau \in \mathcal{T}(T, y)$  be an embedding of  $\mu$  in  $Y$  under  $P^y$ , i.e. let  $Y_\tau$  have the distribution  $\mu$  under  $P^y$ . Then [10] and [20] imply that

- if  $l > -\infty$ , then  $\bar{\mu} \leq y$ ,
- if  $r < \infty$ , then  $\bar{\mu} \geq y$ .

Thus,  $\mu \in \mathcal{M}^1$  whenever  $r$  or  $l$  is finite. Section 3.5 in [13] shows that if  $J = (-\infty, \infty)$  and  $\tau$  is an integrable embedding for  $\mu$ , then  $\mu \in \mathcal{M}^1$ . If  $\bar{\mu} = y$ , then it follows from Theorem 2.4. in [13] that

$$\int q_y(x)\mu(dx) \leq E^y[\tau] \leq T = T - H(y, \bar{\mu}).$$

If  $\bar{\mu} < y$ , then we conclude from Theorem 3.6 in [13] that

$$\int q_y(x)\mu(dx) + (y - \bar{\mu}) \left( m((y, \infty)) + \frac{1}{2}m(\{y\}) \right) \leq T,$$

which implies that Property 2 holds true. If  $\bar{\mu} > y$ , we again apply Theorem 3.6 in [13] to obtain  $\int q_y(x)\mu(dx) \leq T - H(y, \bar{\mu})$ .

For the reverse direction let  $\mu \in \mathcal{A}(T, y)$  and assume first that  $\mu$  is centered around  $y$ . Then  $\mu$  can be embedded in  $Y$  under  $P^y$  for  $-\infty \leq l < r \leq \infty$  by [10] and [20]. It follows from Theorem 3.4 in [13] that there exists a minimal stopping time  $\tau$  with  $Y_\tau \sim \mu$  under  $P^y$  and

$$E^y[\tau] = \int q_y(x)\mu(dx) \leq T.$$

Hence,  $\tau \in \mathcal{T}(T, y)$ .

Now let  $\mu \in \mathcal{A}(T, y)$  with  $\bar{\mu} < y$ . Then we have  $r = \infty$ . Theorem 3.6 in [13] shows the existence of a minimal embedding  $\tau$  of  $\mu$  in  $Y$  under  $P^y$  with

$$E^y[\tau] = \int q_y(x)\mu(dx) + (y - \bar{\mu}) \left( m((y, \infty)) + \frac{1}{2}m(\{y\}) \right) \leq T,$$

where the last inequality follows from the second property of  $\mu$ . Hence,  $\tau \in \mathcal{T}(T, y)$ .

Finally, for  $\mu \in \mathcal{A}(T, y)$  with  $\bar{\mu} > y$ , using similar arguments, one can show that there exists a stopping time  $\tau$  with  $Y_\tau \sim \mu$  under  $P^y$  and  $E^y[\tau] \leq T$ .  $\square$

**Remark 3.3.** The function  $q_y$  appearing in the definition of the set of measures  $\mathcal{A}(T, y)$  plays for the Markov process  $Y$  the same role as the function  $x \mapsto x^2$  plays for the Brownian motion. Indeed, we know that when  $Y$  is a Brownian motion starting in  $y = 0$ , we can find an embedding of  $\mu$  with an integrable stopping time if and only if  $\mu$  is centered and in  $L^2$ . The papers [13] and [1] identify the function  $q_y$  as the counterpart of the second-order moment when  $Y$  is a general diffusion.

**Remark 3.4.** When  $\mu \in \mathcal{A}(T, y)$  is not centered around  $y$  (i.e.  $\bar{\mu} \neq y$ ), the function  $H$  does not vanish in the constraint 2 of  $\mathcal{A}(T, y)$ . In this case, the measure  $\mu$  can be embedded by the following stopping rule  $\tau$ : First wait until  $\tau_{\bar{\mu}} = \inf\{t \in \mathbb{R}_{\geq 0} : Y_t = \bar{\mu}\}$  and then embed  $\mu$  in  $Y$ , started at  $\bar{\mu}$ . To prove this, note that

$$q_z(x) = q_y(x) - q_y(z) - \frac{1}{2}(x - z) \left( \frac{\partial^+ q_y}{\partial x}(z) + \frac{\partial^- q_y}{\partial x}(z) \right), \quad x \in J, \quad (3.2)$$

where  $\frac{\partial^+ q_y}{\partial x}$  and  $\frac{\partial^- q_y}{\partial x}$  denote the right-hand side and left-hand side derivative of  $q_y$ , respectively. Let  $a_n = -n$  if  $\bar{\mu} > y$  (i.e.  $l = -\infty$  by Remark 3.1) and  $a_n = n$  otherwise. Define  $\tau_{\bar{\mu}, a_n} = \inf\{t \geq 0 : Y_t \notin (\bar{\mu} \wedge a_n, \bar{\mu} \vee a_n)\}$ . Monotone convergence and Lemma 2.2 in [3] imply

$$\begin{aligned} E^y[\tau_{\bar{\mu}}] &= \lim_{n \rightarrow \infty} E^y[\tau_{\bar{\mu}, a_n}] = \lim_{n \rightarrow \infty} E^y[q_y(Y_{\tau_{\bar{\mu}, a_n}})] \\ &= q_y(\bar{\mu}) + \frac{1}{2}m(\{y\})|y - \bar{\mu}| + \mathbf{1}_{\{\bar{\mu} < y\}}(y - \bar{\mu})m((y, \infty)) \\ &\quad + \mathbf{1}_{\{\bar{\mu} > y\}}(\bar{\mu} - y)m((-\infty, y)) \\ &= q_y(\bar{\mu}) + H(y, \bar{\mu}). \end{aligned} \quad (3.3)$$

(3.3) together with (3.2) yields that

$$\begin{aligned} E^y[\tau] &= E^y[\tau_{\bar{\mu}}] + \int q_{\bar{\mu}}(x)\mu(dx) = q_y(\bar{\mu}) + H(y, \bar{\mu}) + \int q_{\bar{\mu}}(x)\mu(dx) \\ &= \int q_y(x)\mu(dx) + H(y, \bar{\mu}). \end{aligned}$$

Proposition 3.2 allows to reformulate the stopping problem (2.4) as a linear problem on  $\mathcal{M}$ .

**Corollary 3.5.** We have

$$v(T, y) = \sup_{\mu \in \mathcal{A}(T, y)} \int f(x)\mu(dx) \quad (3.4)$$

and for any optimal  $\mu \in \mathcal{A}(T, y)$  there exists an optimal stopping time  $\tau \in \mathcal{T}(T, y)$  in (2.4) with  $Y_\tau \sim \mu$  under  $P^y$ .

Notice that the functional  $\mu \mapsto \int f(x)\mu(dx)$  is linear on  $\mathcal{A}(T, y)$ . We have thus obtained a linear problem over a set of probability measures  $\mu$  with some integrability constraints. Recall that for standard linear problems the maximum value is attained by extreme

points. We have a similar result for an optimization problem  $\int g d\mu$  over measures  $\mu \in \mathcal{M}$  satisfying moment constraints of the form  $\int f_i d\mu \leq c_i$ ,  $g$  and  $f_i$  measurable,  $c_i \in \mathbb{R}$ ,  $i \in \{1, \dots, n\}$ . The maximum value of  $\int g d\mu$  is also attained in the set of extreme points, see [29]. Furthermore, the extreme points are contained in the set of all weighted Dirac measures with at most  $n + 1$  mass points satisfying the moment constraints.

In the following we denote the extreme points of a convex set  $A \subseteq \mathcal{M}$  by  $\mathcal{E}(A)$  and for any  $M \subseteq \mathcal{M}$  we denote by  $M_3$  the set of all measures in  $M$  which are a weighted sum of at most 3 Dirac measures.

Now we reduce the optimization problem (3.4) to an optimization problem over weighted sums of Dirac measures.

**Theorem 3.6.** *We have*

$$v(T, y) = \sup_{\mu \in \mathcal{A}_3(T, y)} \int f(x) \mu(dx). \tag{3.5}$$

*Proof.* We consider two cases. In the first case we assume that all measures  $\mu$  in  $\mathcal{A}(T, y)$  are centered around  $y$ , i.e.  $\bar{\mu} = y$ . Observe that  $\bar{\mu} = y$  for all  $\mu \in \mathcal{A}(T, y)$  if and only if one of the following four cases is satisfied: 1.  $J$  is bounded, 2.  $l > -\infty$ ,  $r = \infty$  and  $m((y, \infty)) = \infty$ , 3.  $l = -\infty$ ,  $r < \infty$  and  $m((-\infty, y)) = \infty$  and 4.  $J = \mathbb{R}$ ,  $m((y, \infty)) = \infty$  and  $m((-\infty, y)) = \infty$ . The optimization problem (3.4) can be rewritten as

$$v(T, y) = \sup_{t \in [0, T]} \sup_{\mu \in \mathcal{D}(t, y)} \int f(x) \mu(dx),$$

where  $\mathcal{D}(t, y) = \{\mu \in \mathcal{M}^1 : \bar{\mu} = y \text{ and } \int q_y(x) \mu(dx) = t\}$ ,  $0 \leq t \leq T$ . Theorem 2.1(b), Proposition 3.1 and Theorem 3.2 in [29] imply that

$$\sup_{\mu \in \mathcal{D}(t, y)} \int f(x) \mu(dx) = \sup_{\mu \in \mathcal{E}(\mathcal{D}(t, y))} \int f(x) \mu(dx) = \sup_{\mu \in \mathcal{D}_3(t, y)} \int f(x) \mu(dx)$$

because  $\mathcal{D}_3(t, y)$  coincides with  $\mathcal{E}(\mathcal{D}(t, y))$ . For all  $t \in [0, T]$  we have  $\mathcal{D}_3(t, y) \subseteq \mathcal{A}_3(T, y)$ . Therefore,

$$\begin{aligned} v(T, y) &= \sup_{t \in [0, T]} \sup_{\mu \in \mathcal{D}_3(t, y)} \int f(x) \mu(dx) \leq \sup_{\mu \in \mathcal{A}_3(T, y)} \int f(x) \mu(dx) \\ &\leq \sup_{\mu \in \mathcal{A}(T, y)} \int f(x) \mu(dx) = v(T, y). \end{aligned}$$

This proves (3.5) in the first case.

In the second case the set  $\mathcal{A}(T, y)$  also contains uncentered measures. We define

$$\begin{aligned} \mathcal{A}^+(T, y) &= \begin{cases} \{\mu \in \mathcal{A}(T, y) : \bar{\mu} \geq y\}, & \text{if } \exists \mu \in \mathcal{A}(T, y) \text{ with } \bar{\mu} > y, \\ \emptyset, & \text{if } \bar{\mu} \leq y \text{ for all } \mu \in \mathcal{A}(T, y), \end{cases} \\ \mathcal{A}^-(T, y) &= \begin{cases} \{\mu \in \mathcal{A}(T, y) : \bar{\mu} \leq y\}, & \text{if } \exists \mu \in \mathcal{A}(T, y) \text{ with } \bar{\mu} < y, \\ \emptyset, & \text{if } \bar{\mu} \geq y \text{ for all } \mu \in \mathcal{A}(T, y). \end{cases} \end{aligned}$$

Observe that at least one of the sets  $\mathcal{A}^+(T, y)$  or  $\mathcal{A}^-(T, y)$  is nonempty and that (3.4) can be reduced to the two optimization problems  $\sup_{\mu \in \mathcal{A}^+(T, y)} \int f(x) \mu(dx)$  and  $\sup_{\mu \in \mathcal{A}^-(T, y)} \int f(x) \mu(dx)$ , where we follow the convention that the supremum over the empty set is equal to  $-\infty$ . If  $\mathcal{A}^+(T, y)$  is nonempty, then

$$\begin{aligned} \mathcal{A}^+(T, y) &= \left\{ \mu \in \mathcal{M}^1 : \bar{\mu} \geq y, \int q_y(x) \mu(dx) \leq T - H(y, \bar{\mu}) \right\} \\ &= \left\{ \mu \in \mathcal{M}^1 : \int -x \mu(dx) \leq -y, \int (q_y(x) + Cx) \mu(dx) \leq T + Cy \right\}, \end{aligned}$$



where  $C = m((-\infty, y)) + \frac{1}{2}m(\{y\}) < \infty$ . Therefore, Proposition 3.1 and Theorem 3.2 in [29] imply that

$$\sup_{\mu \in \mathcal{A}^+(T, y)} \int f(x)\mu(dx) = \sup_{\mu \in \mathcal{E}(\mathcal{A}^+(T, y))} \int f(x)\mu(dx).$$

By Theorem 2.1(a) in [29] we have  $\mathcal{E}(\mathcal{A}^+(T, y)) \subseteq \mathcal{A}_3^+(T, y)$ . Thus,

$$\sup_{\mu \in \mathcal{A}^+(T, y)} \int f(x)\mu(dx) = \sup_{\mu \in \mathcal{A}_3^+(T, y)} \int f(x)\mu(dx). \tag{3.6}$$

If  $\mathcal{A}^-(T, y)$  is nonempty, similar arguments show that (3.6) holds with  $\mathcal{A}^+(T, y)$  and  $\mathcal{A}_3^+(T, y)$  replaced by  $\mathcal{A}^-(T, y)$  and  $\mathcal{A}_3^-(T, y)$ , respectively. Since  $\mathcal{A}_3(T, y) = \mathcal{A}_3^-(T, y) \cup \mathcal{A}_3^+(T, y)$  we conclude that

$$v(T, y) = \sup_{\mu \in \mathcal{A}_3(T, y)} \int f(x)\mu(dx). \quad \square$$

With Theorem 3.6 we can prove Theorem 2.5.

*Proof of Theorem 2.5.* Let  $\mu \in \mathcal{A}_3(T, y)$  with exactly three mass points  $a < c < d$ . First observe that we can assume that  $\mu$  is centered around  $y$ . Otherwise the first hitting time of  $\bar{\mu}$  is integrable with respect to  $P^y$  (Theorem 2.4 in [13]) and we wait until  $Y$  hits  $\bar{\mu}$  and then continue as in the centered case (cf. Remark 3.4).

We next use the balayage method, developed by Chacon and Walsh in [8] for Brownian motion and extended in [11] to general starting and target distributions, in order to construct a stopping time that embeds  $\mu$  into  $Y$ . More precisely, we define consecutive exit times for the diffusion  $Y$  as follows:

$$\begin{aligned} \tau_1 &= \inf\{t \in \mathbb{R}_{\geq 0} : Y_t \notin (a, b)\} \\ \tau_2 &= \tau_1 + \inf\{t \in \mathbb{R}_{\geq 0} : Y_t \notin (c, d)\} \circ \theta_{\tau_1}, \end{aligned}$$

where  $b = (\mu(\{c\})c + \mu(\{d\})d)/(1 - \mu(\{a\}))$ . Notice that  $b \in (c, d)$ . Moreover, since  $\mu$  is centered around  $y$  and  $a < y$ , it holds that  $b = (y - \mu(\{a\})a)/(1 - \mu(\{a\})) > y$ . Thus,  $b \in (\max\{c, y\}, d)$ . The stopping time  $\tau_2$  is an embedding of  $\mu$  into  $Y$  under  $P^y$ . By using that  $q_y(Y_t) - (t \wedge \tau_{1,r})$  is a local martingale, one can show  $E^y[\tau_2] = E^y[q_y(Y_{\tau_2})] \leq T$ ; hence  $\tau_2 \in \mathcal{T}_3(T, y)$ .

If  $\mu$  has two mass points  $a < c$ , then  $\tau = \inf\{t \in \mathbb{R}_{\geq 0} : Y_t \notin (a, c)\} \in \mathcal{T}_3(T, y)$ . And similarly, if  $\mu = \delta_a$ , then  $\tau = \inf\{t \in \mathbb{R}_{\geq 0} : Y_t = a\} \in \mathcal{T}_3(T, y)$ .  $\square$

The following example shows that in general a reduction to  $\mathcal{A}_2(T, y)$ , the set of probability measures in  $\mathcal{A}(T, y)$  that are weighted sums of at most 2 Dirac measures, is not possible.

**Example 3.7.** Let  $(Y_t)_{t \in \mathbb{R}_{\geq 0}}$  be a Brownian motion starting in 0 and let  $f(x) = 1_{\{|x| \geq 1\}}$ ,  $x \in \mathbb{R}$ , be the payoff function. According to Remark 2.2 the speed measure is in this case given by  $m(dx) = 2 dx$  and the function  $q_0$  satisfies  $q_0(x) = x^2$  for all  $x \in \mathbb{R}$ . We claim that

$$v(T, 0) = \sup_{\mu \in \mathcal{A}(T, 0)} \int f(x)\mu(dx) = T \wedge 1,$$

and

$$\tilde{v}(T, 0) = \sup_{\mu \in \mathcal{A}_2(T, 0)} \int f(x)\mu(dx) = \begin{cases} \frac{T}{1+T}, & T < 1, \\ 1, & T \geq 1. \end{cases} \tag{3.7}$$

To show this, first observe that the second constraint in the definition of  $\mathcal{A}(T, 0)$  ensures that all measures in  $\mathcal{A}(T, 0)$  are centered around 0. If  $T \geq 1$ , the measure  $\mu^*$  given by

$$\mu^* = \frac{1}{2}\delta_{-1} + \frac{1}{2}\delta_1$$

satisfies  $\mu^* \in \mathcal{A}_2(T, 0) \subseteq \mathcal{A}(T, 0)$ . Since  $f$  attains its maximum at  $-1$  and  $1$  it follows that  $v(T, 0) = \tilde{v}(T, 0) = \int f(x)\mu^*(dx) = 1$  in this case.

In the sequel assume that  $T < 1$ . Observe that for every measure  $\mu \in \mathcal{A}_2(T, 0)$  at least one mass point is contained in  $(-1, 1)$ . Due to the symmetry of the optimization problem in (3.7) and the form of  $f$ , we can restrict ourselves to measures of the form

$$\mu^S = \frac{1}{1+S}\delta_{-S} + \frac{S}{1+S}\delta_1 \in \mathcal{A}_2(T, 0),$$

where  $S \in (0, T]$ . Then we obtain

$$\tilde{v}(T, 0) = \sup_{\mu \in \mathcal{A}_2(T, 0)} \int f(x)\mu(dx) = \sup_{S \in (0, T]} \int f(x)\mu^S(dx) = \frac{T}{1+T}.$$

Theorem 3.6 implies that in the maximization problem for  $v$  it is sufficient to consider measures  $\mu \in \mathcal{A}_3(T, 0)$ . Moreover, since  $f$  is constant and maximal on  $\mathbb{R} \setminus (-1, 1)$  and symmetric, we can restrict ourselves to mass points  $-1, c$  and  $1$  for  $c \in [0, 1)$ . The class of all centered probability measures  $\mu^{S,c}$  with these three mass points and  $\int q_0 d\mu^{S,c} = S$  is given by

$$\mu^{S,c} = \frac{S+c}{2(1+c)}\delta_{-1} + \frac{1-S}{1-c^2}\delta_c + \frac{S-c}{2(1-c)}\delta_1, \quad c \in [0, 1), S \in [c, 1].$$

Hence  $\mu^{S,c} \in \mathcal{A}_3(T, 0)$  if and only if  $c \in [0, T]$  and  $S \in [c, T]$ . We have

$$\int f(x)\mu^{S,c}(dx) = \frac{S-c^2}{1-c^2} = 1 - \frac{1-S}{1-c^2},$$

which is maximized for  $c = 0$  and  $S = T$ . Hence we obtain,

$$v(T, 0) = \sup_{\mu \in \mathcal{A}(T, 0)} \int f(x)\mu(dx) = \sup_{\mu \in \mathcal{A}_3(T, 0)} \int f(x)\mu(dx) = \int f(x)\mu^*(dx) = T$$

with

$$\mu^* = \frac{T}{2}\delta_{-1} + (1-T)\delta_0 + \frac{T}{2}\delta_1.$$

The proof of Theorem 2.5 yields that the corresponding optimal stopping time is given by

$$\tau = \tau_{-1,b}(0) + \mathbf{1}_{\{Y_{\tau_{-1,b}(0)} = b\}} \inf \{t \in \mathbb{R}_{\geq 0} : Y_{t+\tau_{-1,b}(0)} \in \{0, 1\}\},$$

where  $b = \frac{T}{2-T}$ .

**Example 3.8.** The framework of Section 2 allows to solve stopping problems where the process to stop is not necessarily characterized as solution of an SDE. One such example is Brownian motion on  $\mathbb{R}$  sticky at 0. This process evolves like a Brownian motion outside 0 but spends a Lebesgue-positive amount of time at zero without having intervals of zeros. More formally, let  $Y$  be a general diffusion in natural scale with state space  $J = \mathbb{R}$  and speed measure

$$m(dx) = 2dx + 2\kappa\delta_0(dx),$$

where  $\kappa \in [0, \infty)$ . It follows that the function  $q_0$  satisfies

$$q_0(x) = x^2 + \kappa|x|, \quad \forall x \in \mathbb{R}.$$

One can generalize the results of Example 3.7 to the sticky case. Let  $f(x) = 1_{\{|x| \geq 1\}}$ ,  $x \in \mathbb{R}$ , be the payoff function. Then similar calculations as in Example 3.7 show that

$$v(T, 0) = \sup_{\mu \in \mathcal{A}(T, 0)} \int f(x) \mu(dx) = \frac{T}{1 + \kappa} \wedge 1,$$

with optimal measure

$$\mu^* = \frac{1}{2} \left( \frac{T}{1 + \kappa} \wedge 1 \right) \delta_{-1} + \left( 1 - \frac{T}{1 + \kappa} \right)^+ \delta_0 + \frac{1}{2} \left( \frac{T}{1 + \kappa} \wedge 1 \right) \delta_1.$$

Moreover, a straight-forward calculation shows that for  $T < 1 + \kappa$  the supremum over  $\mathcal{A}_2(T, 0)$  is strictly smaller than the value function  $v(T, 0)$ .

The parameter  $\kappa$  controls the amount of time spent at zero by the sticky Brownian motion. For large values of  $\kappa$ , the process is held longer in zero, and the optimal value  $v(T, 0)$ , giving the probability of stopping the process  $Y$  outside the interval  $(-1, 1)$  is small. Also, remark that when  $\kappa = 0$ , the optimal values given above for the sticky Brownian motion coincide with the results of Example 3.7 for the Brownian motion.

#### 4 Existence of an optimizer

The next example shows that the supremum in (3.5) is not always attained.

**Example 4.1.** Let  $f_1(x) = x^2 \frac{|x|}{1+|x|}$ ,  $x \in \mathbb{R}$ , and  $Y$  be a Brownian motion starting in 0 under  $P^0$ . In this case there does not exist an optimal stopping time. To prove this let  $v_1 := \sup_{\tau \in \mathcal{T}(T, 0)} E^0[f_1(Y_\tau)]$ . Moreover, consider the second payoff function  $f_2(x) = x^2$ . Note that for any integrable stopping time  $\tau$  we have  $E^0[Y_\tau^2] = E^0[\tau]$ . Therefore,  $v_2 := \sup_{\tau \in \mathcal{T}(T, 0)} E^0[f_2(Y_\tau)] = T$ .

One can show that  $v_1 = v_2$ . Indeed, on the one hand it must hold that  $v_1 \leq v_2$  since  $f_1 \leq f_2$ . On the other hand, for the stopping times  $\tau_n = \tau_{-1/n, nT}$  we have  $E^0[\tau_n] = T$  and

$$E^0[f_1(Y_{\tau_n})] = \frac{nT}{1/n + nT} \frac{1}{n^2} \frac{1/n}{1 + 1/n} + \frac{1/n}{1/n + nT} n^2 T^2 \frac{nT}{1 + nT} \rightarrow T,$$

as  $n \rightarrow \infty$ , and hence  $v_1 \geq v_2$ .

From  $v_1 = v_2$  we can deduce that the supremum can not be attained in  $v_1$ , because for any stopping time  $\tau \neq 0$  with  $E^0[\tau] < \infty$  we have  $P^0[f_1(Y_\tau) < f_2(Y_\tau)] > 0$ .

We now establish the existence of an optimal measure in  $\mathcal{A}_3(T, y)$  in (3.5) under mild conditions on the payoff function  $f$ .

**Theorem 4.2.** Assume that  $f: J \rightarrow \mathbb{R}$  is upper semi-continuous with  $\limsup_{x \uparrow r} \frac{f(x)}{q_y(x)} \leq 0$  if  $r \notin J$  and  $\limsup_{x \downarrow l} \frac{f(x)}{q_y(x)} \leq 0$  if  $l \notin J$ . Then there exists an optimal measure in  $\mathcal{A}_3(T, y)$  for (3.5) and an optimal stopping time in  $\mathcal{T}_3(T, y)$  for (2.6).

**Remark 4.3.** In the special case where  $J$  is a compact interval and  $f: J \rightarrow \mathbb{R}$  is upper semi-continuous, the existence of an optimal measure in  $\mathcal{A}_3(T, y)$  for (3.5) can be shown using Prokhorov's theorem. Indeed, if  $(\mu_n)_{n \in \mathbb{N}}$  is a sequence in  $\mathcal{A}_3(T, y)$  such that  $\lim_{n \rightarrow \infty} \int f d\mu_n = v(T, y)$ , then compactness of  $J$  ensures that  $(\mu_n)_{n \in \mathbb{N}}$  is tight. By Prokhorov's theorem,  $(\mu_n)_{n \in \mathbb{N}}$  converges weakly along a subsequence to a probability measure  $\mu$ . It follows that  $\mu \in \mathcal{A}_3(T, y)$  and, moreover, the Portmanteau theorem and the fact that  $f$  is bounded from above by compactness of  $J$  ensure that  $v(T, y) = \lim_{n \rightarrow \infty} \int f d\mu_n \leq \int f d\mu$ . Hence,  $\mu \in \mathcal{A}_3(T, y)$  is optimal in (3.5). For the more general setting of Theorem 4.2 we provide a more elementary proof.

*Proof of Theorem 4.2.* Throughout the proof we denote by  $C_+, C_- \in [0, \infty]$  the extended real numbers given by  $C_+ = m((y, \infty)) + \frac{1}{2}m(\{y\})$  and  $C_- = m((-\infty, y)) + \frac{1}{2}m(\{y\})$ . Let  $\mu_n = \sum_{j=1}^3 p_n^j \delta_{x_n^j} \in \mathcal{A}(T, y)$ ,  $n \in \mathbb{N}$ , be a sequence of measures such that

$$\lim_{n \rightarrow \infty} \int_{\mathbb{R}} f d\mu_n = v(T, y).$$

If the sequence  $(x_n^1)_{n \in \mathbb{N}}$  is unbounded, choose a subsequence, also denoted by  $(x_n^1)_{n \in \mathbb{N}}$ , such that either  $\lim_{n \rightarrow \infty} x_n^1 = -\infty =: x^1$  or  $\lim_{n \rightarrow \infty} x_n^1 = \infty =: x^1$ . If  $(x_n^1)_{n \in \mathbb{N}}$  is bounded, extract a subsequence such that  $\lim_{n \rightarrow \infty} x_n^1 = x^1 \in \bar{J}$ . By extracting further subsequences, proceed in the same way with  $(x_n^2)_{n \in \mathbb{N}}$  and  $(x_n^3)_{n \in \mathbb{N}}$ . Refine once again the sequence to obtain that  $(p_n^1, p_n^2, p_n^3) \rightarrow (p^1, p^2, p^3) \in [0, 1]^3$  as  $n \rightarrow \infty$ . Overall we obtain for  $n \rightarrow \infty$  that

$$(x_n^1, x_n^2, x_n^3, p_n^1, p_n^2, p_n^3) \rightarrow (x^1, x^2, x^3, p^1, p^2, p^3) \in (\bar{J} \cup \{-\infty\} \cup \{\infty\})^3 \times [0, 1]^3.$$

Recall that  $\bar{J}$  denotes the closure of  $J$  in  $\mathbb{R}$  with respect to the Euclidean metric. Note that  $x^j = \infty$  and  $x^j = -\infty$  are only possible if  $r = \infty$  and  $l = -\infty$ , respectively.

Let  $K = \{j \in \{1, 2, 3\} \mid x^j \in J\}$ ,  $\bar{K}^+ = \{j \in \{1, 2, 3\} \mid x^j \notin J, x^j = r\}$  and  $\bar{K}^- = \{j \in \{1, 2, 3\} \mid x^j \notin J, x^j = l\}$ . Define  $\mu = \sum_{k \in K} p^k \delta_{x^k}$ . We show that  $\mu$  is an optimizer for (3.5).

From the fact that for all  $i \in \{1, 2, 3\}$  it holds

$$0 \leq p_n^i q_y(x_n^i) \leq \sum_{j=1}^3 p_n^j q_y(x_n^j) = \int_{\mathbb{R}} q_y(x) \mu_n(dx) \leq T \tag{4.1}$$

and that  $\lim_{n \rightarrow \infty} q_y(x_n^i) = \infty$  for all  $i \in \{1, 2, 3\} \setminus K$  by Lemma 2.1, it follows for all  $i \in \{1, 2, 3\} \setminus K$  that

$$\lim_{n \rightarrow \infty} p_n^i = \lim_{n \rightarrow \infty} \frac{1}{q_y(x_n^i)} p_n^i q_y(x_n^i) = 0. \tag{4.2}$$

We conclude from (4.2) that

$$\mu(J) = \sum_{k \in K} p^k = \lim_{n \rightarrow \infty} \sum_{k \in K} p_n^k = \lim_{n \rightarrow \infty} \sum_{j=1}^3 p_n^j = \lim_{n \rightarrow \infty} \mu_n(J) = 1.$$

Thus,  $\mu \in \mathcal{M}^1$ . Next we show that  $\mu \in \mathcal{A}_3(T, y)$ . To this end we distinguish four cases.

**1.**  $l > -\infty, r < \infty$ . Observe that in this case we have  $\bar{\mu}_n = y$  for all  $n \in \mathbb{N}$ . This together with (4.2) ensures that

$$y = \lim_{n \rightarrow \infty} \bar{\mu}_n = \lim_{n \rightarrow \infty} \left( \sum_{k \in K} p_n^k x_n^k + \sum_{i \in \{1, 2, 3\} \setminus K} p_n^i x_n^i \right) = \sum_{k \in K} p^k x^k = \bar{\mu}. \tag{4.3}$$

Moreover, continuity and nonnegativity of  $q_y$  on  $J$  imply that

$$\int_{\mathbb{R}} q_y(x) \mu(dx) = \lim_{n \rightarrow \infty} \sum_{k \in K} p_n^k q_y(x_n^k) \leq \limsup_{n \rightarrow \infty} \sum_{i=1}^3 p_n^i q_y(x_n^i) \leq T = T - H(y, \bar{\mu}). \tag{4.4}$$

This proves that  $\mu \in \mathcal{A}_3(T, y)$ .

**2.**  $l > -\infty, r = \infty$ . In this case we know that  $\bar{\mu}_n \leq y$  for all  $n \in \mathbb{N}$ . Let us first assume that  $m((y, \infty)) = \infty$ . Then it holds that  $\bar{\mu}_n = y$  for all  $n \in \mathbb{N}$ . Moreover, we have for all  $i \in \bar{K}^+$  that  $\lim_{n \rightarrow \infty} \frac{q_y(x_n^i)}{x_n^i} = \frac{1}{2}m(\{y\}) + m((y, \infty)) = \infty$  and hence, with (4.1),

$$\lim_{n \rightarrow \infty} p_n^i x_n^i = \lim_{n \rightarrow \infty} p_n^i q_y(x_n^i) \frac{x_n^i}{q_y(x_n^i)} = 0. \tag{4.5}$$

This and (4.2) show that

$$y = \lim_{n \rightarrow \infty} \bar{\mu}_n = \lim_{n \rightarrow \infty} \left( \sum_{k \in K} p_n^k x_n^k + \sum_{i \in \bar{K}^+} p_n^i x_n^i + \sum_{i \in \bar{K}^-} p_n^i x_n^i \right) = \sum_{k \in K} p^k x^k = \bar{\mu}.$$

Then the same reasoning as in (4.4) demonstrates that  $\int q_y(x) \mu_n(dx) \leq T - H(y, \bar{\mu})$  and hence  $\mu \in \mathcal{A}_3(T, y)$ .

Let us now assume that  $m((y, \infty)) < \infty$ . Equation (4.2) implies that

$$\begin{aligned} y &\geq \limsup_{n \rightarrow \infty} \bar{\mu}_n = \limsup_{n \rightarrow \infty} \left( \sum_{k \in K} p_n^k x_n^k + \sum_{i \in \bar{K}^+} p_n^i x_n^i + \sum_{i \in \bar{K}^-} p_n^i x_n^i \right) \\ &\geq \sum_{k \in K} p^k x^k + \limsup_{n \rightarrow \infty} \sum_{i \in \bar{K}^-} p_n^i x_n^i = \bar{\mu}. \end{aligned} \tag{4.6}$$

Moreover, it holds for all  $n \in \mathbb{N}$  that

$$\begin{aligned} T &\geq \int_{\mathbb{R}} q_y(x) \mu_n(dx) + H(y, \bar{\mu}_n) = \sum_{i=1}^3 p_n^i q_y(x_n^i) + C_+ \left( y - \sum_{i=1}^3 p_n^i x_n^i \right) \\ &= \sum_{i=1}^3 p_n^i \left( q_y(x_n^i) - C_+(x_n^i - y) \right). \end{aligned} \tag{4.7}$$

It follows with (4.6) that

$$\begin{aligned} \lim_{n \rightarrow \infty} \sum_{k \in K} p_n^k \left( q_y(x_n^k) - C_+(x_n^k - y) \right) &= \sum_{k \in K} p^k \left( q_y(x^k) - C_+(x^k - y) \right) \\ &= \int_{\mathbb{R}} q_y(x) \mu(dx) + H(y, \bar{\mu}). \end{aligned} \tag{4.8}$$

Combining (4.1) and  $\lim_{x \rightarrow \infty} \frac{q_y(x)}{x-y} = C_+$  yields that

$$\lim_{n \rightarrow \infty} \sum_{i \in \bar{K}^+} p_n^i \left( q_y(x_n^i) - C_+(x_n^i - y) \right) = 0. \tag{4.9}$$

Moreover, the nonnegativity of  $q_y$  and (4.2) imply that

$$\liminf_{n \rightarrow \infty} \sum_{i \in \bar{K}^-} p_n^i \left( q_y(x_n^i) - C_+(x_n^i - y) \right) \geq \liminf_{n \rightarrow \infty} \sum_{i \in \bar{K}^-} C_+ p_n^i (y - x_n^i) = 0. \tag{4.10}$$

Combining (4.7), (4.8), (4.9) and (4.10) proves that  $\mu \in \mathcal{A}_3(T, y)$ .

**3.**  $l = -\infty, r < \infty$ . This case is analog to the case  $l > -\infty, r = \infty$ .

**4.**  $l = -\infty, r = \infty$ . In this case no conditions on  $\bar{\mu}$  have to be verified. Assume first that  $m((y, \infty)) = \infty$  and  $m((-\infty, y)) = \infty$ . As in (4.5) it follows in this case that  $\lim_{n \rightarrow \infty} p_n^i x_n^i = 0$  for all  $i \in \bar{K}^+ \cup \bar{K}^-$ . In addition, it holds that  $\bar{\mu}_n = y$  for all  $n \in \mathbb{N}$ . Hence, we conclude that  $y = \lim_{n \rightarrow \infty} \bar{\mu}_n = \bar{\mu}$ . As in (4.4) we obtain that  $\mu \in \mathcal{A}_3(T, y)$ . Next, assume that  $m((y, \infty)) < \infty$  and  $m((-\infty, y)) = \infty$ . In this case we obtain as in (4.5) that  $\lim_{n \rightarrow \infty} p_n^i x_n^i = 0$  for all  $i \in \bar{K}^-$ . This, together with the fact that  $y \geq \bar{\mu}_n$  for all  $n \in \mathbb{N}$ , proves that  $y \geq \bar{\mu}$  (see also (4.6)). Since  $\lim_{n \rightarrow \infty} x_n^i = -\infty$  for all  $i \in \bar{K}^-$  we conclude that

$$\liminf_{n \rightarrow \infty} \sum_{i \in \bar{K}^-} p_n^i \left( q_y(x_n^i) - C_+(x_n^i - y) \right) \geq \liminf_{n \rightarrow \infty} \sum_{i \in \bar{K}^-} C_+ p_n^i (y - x_n^i) \geq 0. \tag{4.11}$$

Then proceeding exactly as in (4.7), (4.8) and (4.9) shows that  $\mu \in \mathcal{A}_3(T, y)$ . The case  $m((y, \infty)) = \infty$  and  $m((-\infty, y)) < \infty$  can be treated analogously. Finally, we assume that  $m((y, \infty)) < \infty$  and  $m((-\infty, y)) < \infty$ . Without loss of generality we also assume that  $\bar{\mu} \leq y$ . In this case we obtain for all  $n \in \mathbb{N}$  that

$$\begin{aligned} T &\geq \int_{\mathbb{R}} q_y(x) \mu_n(dx) + H(y, \bar{\mu}_n) \\ &= \sum_{i=1}^3 p_n^i q_y(x_n^i) + C_+ \left( y - \sum_{i=1}^3 p_n^i x_n^i \right) + \mathbb{1}_{(y, \infty)}(\bar{\mu}_n) (C_+ + C_-) (\bar{\mu}_n - y) \\ &\geq \sum_{i=1}^3 p_n^i \left( q_y(x_n^i) - C_+(x_n^i - y) \right). \end{aligned} \tag{4.12}$$

Proceeding as in (4.8), (4.9) and (4.11) proves that  $\mu \in \mathcal{A}_3(T, y)$ .

To summarize, we have shown that  $\mu \in \mathcal{A}_3(T, y)$  in any possible case. It remains to show the optimality of  $\mu$ . First note that  $v(T, y) \geq \int_{\mathbb{R}} f d\mu$ . The assumptions  $\limsup_{x \uparrow r} \frac{f(x)}{q_y(x)} \leq 0$  if  $r \notin J$  and  $\limsup_{x \downarrow l} \frac{f(x)}{q_y(x)} \leq 0$  if  $l \notin J$  together with (4.1) imply that for  $i \notin K$

$$\limsup_{n \rightarrow \infty} p_n^i f(x_n^i) = \limsup_{n \rightarrow \infty} p_n^i q_y(x_n^i) \frac{f(x_n^i)}{q_y(x_n^i)} \leq 0. \tag{4.13}$$

Finally, the upper semi-continuity of  $f$  and (4.13) result in

$$\begin{aligned} \int_{\mathbb{R}} f(x) \mu(dx) &= \sum_{k \in K} p^k f(x^k) \geq \limsup_{n \rightarrow \infty} \sum_{k \in K} p_n^k f(x_n^k) \\ &\geq \limsup_{n \rightarrow \infty} \sum_{k \in K} p_n^k f(x_n^k) + \limsup_{n \rightarrow \infty} \sum_{i \notin K} p_n^i f(x_n^i) \\ &\geq \limsup_{n \rightarrow \infty} \sum_{i=1}^3 p_n^i f(x_n^i) = \lim_{n \rightarrow \infty} \int_{\mathbb{R}} f(x) \mu_n(dx) = v(T, y). \end{aligned}$$

Therefore we conclude that  $v(T, y) = \int_{\mathbb{R}} f d\mu$ . Moreover, the proof of Theorem 2.5 allows to construct a stopping time in  $\mathcal{T}_3(T, y)$  which is optimal in (2.6).  $\square$

**Remark 4.4.** Example 4.1 shows that the condition that  $\limsup_{x \uparrow r} \frac{f(x)}{q_y(x)} \leq 0$  if  $r \notin J$  and  $\limsup_{x \downarrow l} \frac{f(x)}{q_y(x)} \leq 0$  if  $l \notin J$  in Theorem 4.2 can not be weakened in general.

Finally, to complete the article, we explain how to deal with the optimal stopping problem (2.4) if  $Y$  is not in natural scale.

**Remark 4.5.** Let  $Y$  be a general diffusion and suppose that  $Y$  satisfies all the properties of Section 2 apart from being in natural scale. Let  $s$  be the scale function of  $Y$ . Then  $Z_t = s(Y_t)$ ,  $t \in \mathbb{R}_{\geq 0}$ , is a diffusion in natural scale on  $s(J)$ , see Theorem 46.12, in Chapter V, [25] or Theorem 2.1 in [7]. Hence we can convert the optimal stopping problem with reward function  $f$  for the process  $Y$  under  $P^y$  into an optimal stopping problem with reward function  $f \circ s^{-1}$  for  $Z$  under a measure  $Q^{s(y)}$ .

If  $f \circ s^{-1}$  satisfies Assumption (A), where  $y \in s(\overset{\circ}{J})$  and  $q_y$  is defined in (2.1) using the speed measure of  $Z$ , then all results of Section 2–4 apply.

For more details see Section III.7 in [16].

## References

- [1] S. Ankirchner, D. Hobson, and P. Strack, *Finite, integrable and bounded time embeddings for diffusions*, Bernoulli **21** (2015), no. 2, 1067–1088. MR-3338657.

- [2] S. Ankirchner, M. Klein, and T. Kruse, *A Verification Theorem for Optimal Stopping Problems with Expectation Constraints*, *Appl. Math. Optim.* **79** (2019), no. 1, 145–177. MR-3903781.
- [3] S. Ankirchner, M. Klein, T. Kruse, and M. Urusov, *On a certain local martingale in a general diffusion setting*, <https://hal.archives-ouvertes.fr/hal-01700656>, 2018.
- [4] E. Bayraktar and C. W. Miller, *Distribution-constrained optimal stopping*, *Math. Finance* **29** (2019), no. 1, 368–406. MR-3905747.
- [5] E. Bayraktar and S. Yao, *Dynamic programming principles for optimal stopping with expectation constraint*, arXiv preprint arXiv:1708.02192 (2017).
- [6] M. Beiglböck, M. Eder, C. Elgert, and U. Schmock, *Geometry of distribution-constrained optimal stopping problems*, *Probab. Theory Related Fields* **172** (2018), no. 1-2, 71–101. MR-3851830.
- [7] J. K. Brooks, *One-dimensional diffusions and their convergence in distribution*, *Handbook of measure theory*, Vol. I, II, North-Holland, Amsterdam, 2002, pp. 295–343. MR-1954617.
- [8] R. V. Chacon and J. B. Walsh, *One-dimensional potential embedding*, *Séminaire de Probabilités, X (Première partie, Univ. Strasbourg, Strasbourg, année universitaire 1974/1975)*, Springer, Berlin, 1976, pp. 19–23. *Lecture Notes in Math.*, Vol. 511. MR-0445598.
- [9] A. S. Cherny and H.-J. Engelbert, *Singular stochastic differential equations*, *Lecture Notes in Mathematics*, vol. 1858, Springer-Verlag, Berlin, 2005. MR-2112227.
- [10] A. Cox and D. G. Hobson, *An optimal Skorokhod embedding for diffusions*, *Stochastic Process. Appl.* **111** (2004), no. 1, 17–39. MR-2049567.
- [11] A. M. G. Cox, *Extending Chacon-Walsh: Minimality and generalised starting distributions*, *Séminaire de probabilités XLI, Lecture Notes in Math.*, vol. 1934, Springer, Berlin, 2008, pp. 233–264. MR-2483735.
- [12] A. M. G. Cox, D. Hobson, and J. Oblój, *Pathwise inequalities for local time: applications to Skorokhod embeddings and optimal stopping*, *Ann. Appl. Probab.* **18** (2008), no. 5, 1870–1896. MR-2462552.
- [13] D. Hobson, *Integrability of solutions of the Skorokhod embedding problem for diffusions*, *Electron. J. Probab.* **20** (2015), no. 83, 1–26. MR-3383567.
- [14] I. Karatzas and S. E. Shreve, *Brownian motion and stochastic calculus*, second ed., *Graduate Texts in Mathematics*, vol. 113, Springer-Verlag, New York, 1991. MR-1121940.
- [15] D. P. Kennedy, *On a constrained optimal stopping problem*, *J. Appl. Probab.* **19** (1982), no. 3, 631–641. MR-0664846.
- [16] M. Klein, *Optimal Stopping Problems with Expectation Constraints*, Ph.D. thesis, Friedrich Schiller University Jena, 2018, <https://doi.org/10.22032/dbt.35299>.
- [17] T. R. McConnell, *The two-sided Stefan problem with a spatially dependent latent heat*, *Trans. Amer. Math. Soc.* **326** (1991), no. 2, 669–699. MR-1008699.
- [18] C. W. Miller, *Non-linear PDE approach to time-inconsistent optimal stopping*, *SIAM J. Control Optim.* **55** (2017), no. 1, 557–573. MR-3614676.
- [19] J. Oblój, *The maximality principle revisited: on certain optimal stopping problems*, *Séminaire de Probabilités XL, Lecture Notes in Math.*, vol. 1899, Springer, Berlin, 2007, pp. 309–328. MR-2409013.
- [20] J. L. Pedersen and G. Peskir, *The Azéma-Yor embedding in non-singular diffusions*, *Stochastic Process. Appl.* **96** (2001), no. 2, 305–312. MR-1865760.
- [21] G. Peskir, *Designing options given the risk: the optimal Skorokhod-embedding problem*, *Stochastic Process. Appl.* **81** (1999), no. 1, 25–38. MR-1680505.
- [22] G. Peskir and A. Shiryaev, *Optimal stopping and free-boundary problems*, *Lectures in Mathematics ETH Zürich*, Birkhäuser Verlag, Basel, 2006. MR-2256030.
- [23] D. Revuz and M. Yor, *Continuous martingales and Brownian motion*, third ed., *Grundlehren der Mathematischen Wissenschaften [Fundamental Principles of Mathematical Sciences]*, vol. 293, Springer-Verlag, Berlin, 1999. MR-1725357.
- [24] L. C. G. Rogers and D. Williams, *Diffusions, Markov processes, and Martingales. Vol. 1: Foundations*, second ed., *Cambridge Mathematical Library*, Cambridge University Press, 2000. MR-1796539.

- [25] L. C. G. Rogers and D. Williams, *Diffusions, Markov processes, and Martingales. Vol. 2: Itô calculus*, second ed., Cambridge Mathematical Library, Cambridge University Press, 2000. MR-1780932.
- [26] A. N. Shiryaev, *On conditional-extremal problems of the quickest detection of nonpredictable times of the observable Brownian motion*, *Theory Probab. Appl.* **53** (2009), no. 4, 663–678. MR-2766144.
- [27] A. N. Shiryaev, *Optimal stopping rules*, Springer-Verlag, New York-Heidelberg, 1978, Translated from the Russian by A. B. Aries, *Applications of Mathematics*, Vol. 8. MR-0468067.
- [28] M. A. Urusov, *On a property of the moment at which Brownian motion attains its maximum and some optimal stopping problems*, *Theory Probab. Appl.* **49** (2005), no. 1, 169–176. MR-2141339.
- [29] G. Winkler, *Extreme points of moment sets*, *Math. Oper. Res.* **13** (1988), no. 4, 581–587. MR-971911.
- [30] Z. Q. Xu and X. Y. Zhou, *Optimal stopping under probability distortion*, *Ann. Appl. Probab.* **23** (2013), no. 1, 251–282. MR-3059235.

**Acknowledgments.** We are grateful to Tiziano De Angelis, Goran Peskir and Mikhail Urusov for helpful comments. We thank an anonymous referee for carefully reading the manuscript and highly appreciate her/his comments, which contributed to improve our paper. Parts of the research results presented in this article have been obtained while S.A. and N. K.-T. were visiting the Vietnam Institute for Advanced Study in Mathematics (VIASM) in August and September 2016 and while M.K. was visiting the School of Mathematics at the University of Manchester in October and November 2016. Financial support from the VIASM and the DAAD (German Academic Exchange Service) is gratefully acknowledged. M.K. was partially supported by the Austrian Science Fund (FWF) under grant P30864.