Available online at www.sciencedirect.com

SCIENCE DIRECT®

ELSEVIER

Stochastic Processes and their Applications 111 (2004) 17–39

stochastic processes and their applications

www.elsevier.com/locate/spa

An optimal Skorokhod embedding for diffusions

A.M.G. Cox, D.G. Hobson*

Department of Mathematics, University of Bath, Bath BA2 7AY, UK

Received 2 December 2002; received in revised form 3 December 2003; accepted 21 January 2004

Abstract

Given a Brownian motion $(B_t)_{t\geqslant 0}$ and a general target law μ (not necessarily centered or even in $\mathscr{L}^1$) we show how to construct an embedding of μ in B. This embedding is an extension of an embedding due to Perkins, and is optimal in the sense that it simultaneously minimises the distribution of the maximum and maximises the distribution of the minimum among all embeddings of μ. The embedding is then applied to regular diffusions, and used to characterise the target laws for which a H^p-embedding may be found.

MSC: primary 60G40; 60J60; secondary 60G44; 60J65

Keywords: Brownian motion; Diffusion; Embedding; H^p-embedding; Stopping time; Skorokhod problem

1. Introduction

Let $(X_t)_{t\geqslant 0}$ be an adapted stochastic process with state space I, and let μ be a probability measure on I. Given X and μ, the Skorokhod embedding problem is to find a stopping time with the property that $X_T \sim \mu$. For a general stochastic process X, and an arbitrary measure μ, necessary and sufficient conditions for the existence of a solution to Skorokhod problem were given by Rost (1971). Hence attention switches to the construction of solutions.

When $(X_t)_{t\geqslant 0}$ is a one-dimensional Brownian motion started at 0 and μ is a zero-mean target distribution, many explicit constructions of stopping rules which embed μ are known, see for example Skorokhod (1965), Dubins (1968), Root (1969) and Chacon and Walsh (1976). For Brownian motion it is interesting to seek embeddings with additional optimality properties, such as the embedding which minimises the variance of the stopping time (Rost, 1976), the embedding which stochastically minimises

* Corresponding author.
E-mail addresses: mapamgc@maths.bath.ac.uk (A.M.G. Cox), dgh@maths.bath.ac.uk (D.G. Hobson).

0304-4149/$ - see front matter © 2004 Elsevier B.V. All rights reserved.
doi:10.1016/j.spa.2004.01.003

the law of the local time at zero (Vallois, 1992), or the embedding which maximises the law of the supremum of the stopped process amongst the class of stopping times for which the process $(X_{t\wedge T})_{t\geqslant 0}$ is uniformly integrable (Azéma and Yor, 1979a, b).

The first purpose of this article is to consider the embedding in Brownian motion of a target distribution which is not centered and may not even be integrable. Note that if the target distribution has finite mean m then one way to embed the law is to wait until the Brownian motion first hits the level m and then adopt a favourite embedding for a centered target distribution for the shifted process now starting at m. However, this cannot work if m is not well-defined and finite and even if m exists this construction may not share the optimality properties of the original embedding.

More generally, if X_t is a Brownian motion and μ is any target distribution, then there is a simple solution of the Skorokhod embedding problem given by

$$T_D = \inf\{t > 1\colon X_t = h(X_1)\}, \tag{1}$$

where h is chosen so that $h(X_1) \sim \mu$. This solution was pointed out by Doob (see also Revuz and Yor (1999, Exercise VI.5.7)) and can be adapted to any recurrent process. However there is a strong sense in which this simple solution is not a good solution. To be more precise, T_D is not a minimal stopping time (except in a few special cases), or in other words there exists a uniformly smaller stopping time $S \leqslant T_D$ which also embeds μ (see Cox and Hobson (2003)).

The concept of minimal stopping times was introduced by Monroe (1972). Monroe showed that when μ is a zero-mean target distribution for Brownian motion, a stopping time T is minimal if and only if $(X_{t\wedge T})$ is uniformly integrable. Note that, with μ centered and in $\mathscr{L}^1$, T_D defined in (1) is clearly not a uniformly integrable stopping time unless μ is normal with variance 1. Hence we know that there are 'better' constructions in this case at least (see for example the embeddings listed in the second paragraph).

Our aim is to give a 'better' stopping rule than the one given in (1), which applies to all target distributions whether integrable or not. We adapt an embedding which was first proposed by Perkins (1986). For zero mean target laws the Perkins embedding has the property that it simultaneously maximises the distribution of the minimum, and minimises the distribution of the maximum, amongst the class of all stopping times which embed μ. Our adaptation of the Perkins embedding extends to all target distributions and retains the optimality properties of the Perkins embedding. The fact that this embedding minimises the law of a functional of the Brownian path over all embeddings is our justification for the use of the word optimal in the title. This property also allows us to see that the embedding is minimal in the sense proposed by Monroe.

The second purpose of this article is to consider the embedding of μ in a one-dimensional diffusion. The main technique is to use a change of scale to reduce the problem to the Brownian case, and under this transformation it is completely natural for the target measure to have non-zero mean in the Brownian scale. We will see that our embedding is a natural one to use in this situation, and we are able to identify the cases where it is possible to embed a given target distribution, thus rederiving a result in Pedersen and Peskir (2001). We also identify some properties of the maximum and minimum of the processes in these cases. Our results in this direction can be seen as an extension of the results in Grandits and Falkner (2000) (for drifting Brownian motion) and Pedersen and Peskir (2001). In this last paper, the authors use an extension of the Azema–Yor embedding which may not be defined in certain cases of interest. Thus our construction of a Skorokhod embedding is both different to, and more general than, the embedding in Pedersen and Peskir (2001).

The remainder of the article is structured as follows. In Section 2 we consider the problem of embedding a general target measure in Brownian motion. We construct an embedding which is defined for all circumstances in which it is possible to find an embedding and with the property that the law of the maximum is stochastically as small as possible. In Section 3 we show this embedding can be applied to construct embeddings in regular diffusions and in Section 4 we answer the question of when it is possible to construct a H^p-embedding, i.e. given a diffusion process Y and a target law v when does there exists a stopping time T such that $Y_T \sim v$ and $\mathbb{E}\left[\left(\sup_t |Y_{t\wedge T}|\right)^p\right] < \infty$.

2. Embedding a general target measure in Brownian motion

Consider first the problem of embedding a target distribution μ in a one-dimensional local martingale $(M_t)_{t\geqslant 0}$, $M_0 = 0$ a.s. We make no assumptions on μ other than that $\mu(\mathbb{R}) = 1$, and that μ has no atom at 0. In fact this second assumption can be avoided by stopping immediately according to some independent randomisation with suitable probability, and then using the construction to embed the remaining mass of μ, conditional on not stopping at 0. Clearly such a construction is necessary in any stopping time that will minimise the maximum, and maximise the minimum.

For a general local martingale the above conditions are not sufficient to ensure that an embedding exists. However, a sufficient condition for the existence of an embedding for any μ is that our local martingale almost surely has infinite quadratic variation. Since any local martingale is simply a time change of Brownian motion, this just ensures that our time change does not stall.

We begin by defining a series of functions. Let

$$c(x) = \begin{cases} \displaystyle\int_{\{u\geqslant 0\}} (x\wedge u)\mu(\mathrm{d}u), & x \geqslant 0, \\ \displaystyle\int_{\{u<0\}} (|x|\wedge |u|)\mu(\mathrm{d}u), & x < 0. \end{cases} \tag{2}$$

Then $c(x)$ is increasing and concave on $\{x \geqslant 0\}$, decreasing and concave on $\{x \leqslant 0\}$ and continuous on $\mathbb{R}$ (see Figs. 1 and 2). It is also differentiable Lebesgue-almost-everywhere and

$$c'(x)_+ = \begin{cases} \mu((x,\infty)), & x \geqslant 0, \\ -\mu((-\infty,x]), & x < 0, \end{cases} \tag{3}$$

$$c'(x)_- = \begin{cases} \mu([x,\infty)), & x > 0, \\ -\mu((-\infty,x)), & x \leqslant 0, \end{cases} \tag{4}$$

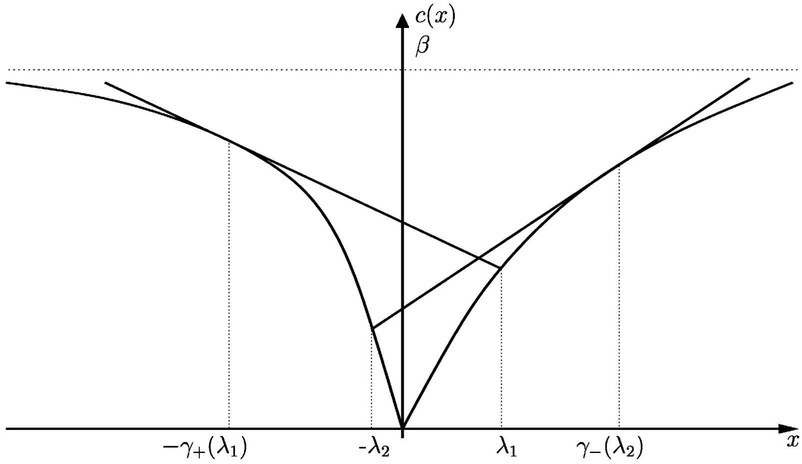

Fig. 1. $c(x)$ for a centered non-atomic measure. As $|x| \to \infty$, $c(x)$ is asymptotic to β, where $\beta = \int_{\{x \geqslant 0\}} x\mu(\mathrm{d}x)$.

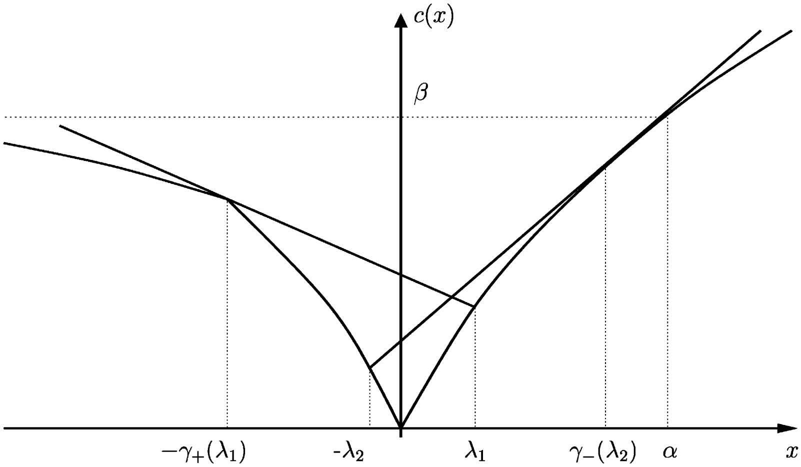

Fig. 2. $c(x)$ for a non-integrable measure with an atom at $-\gamma_+(\lambda_1)$. As $x \to \infty$, $c(x) \to \int_{\{x \geqslant 0\}} x\mu(\mathrm{d}x) = \infty$, while as $x \to -\infty$, $c(x)$ is asymptotic to the level $\beta = -\int_{\{x \leqslant 0\}} x\mu(\mathrm{d}x)$, which for this example is taken to be finite. The point α is such that $c(\alpha) = \beta$, and for all $\lambda > \alpha$, $\gamma_+(\lambda) = \infty$.

where $c'(x)_-$, $c'(x)_+$ are the left and right derivatives, respectively. In particular, the points at which $c(x)$ is not differentiable are precisely the atoms of out target distribution. We also note that $c(\infty) < \infty$ if and only if our target distribution satisfies $\int_{\{x \geqslant 0\}} x\mu(\mathrm{d}x) < \infty$, and $c(-\infty) = \int_{u<0} |u|\mu(\mathrm{d}u)$. Finally, we have $c(\infty) = c(-\infty) < \infty$ if and only if $\mu \in \mathscr{L}^1$ and μ is centered.

For $\lambda > 0$, define the following quantities:

$$\gamma_+(\lambda) = \operatorname*{argmin}_{x>0} \left\{ \frac{c(\lambda) - c(-x)}{\lambda - (-x)} \right\}, \tag{5}$$

$$\gamma_-(\lambda) = \operatorname*{argmax}_{x>0} \left\{ \frac{c(x) - c(-\lambda)}{x - (-\lambda)} \right\}, \tag{6}$$

$$\theta_+(\lambda) = -\inf_{x>0} \left\{ \frac{c(\lambda) - c(-x)}{\lambda - (-x)} \right\}, \tag{7}$$

$$\theta_-(\lambda) = \sup_{x>0} \left\{ \frac{c(x) - c(-\lambda)}{x - (-\lambda)} \right\}, \tag{8}$$

$$\begin{aligned} \mu_+(\lambda) &= \theta_+(\lambda) + \mu([\lambda, \infty)), \\ &= -\frac{c(\lambda) - c(-\gamma_+(\lambda))}{\lambda - (-\gamma_+(\lambda))} + c'(\lambda)_-, \end{aligned} \tag{9}$$

$$\begin{aligned} \mu_-(\lambda) &= \mu((-\infty, -\lambda]) + \theta_-(\lambda), \\ &= -c'(-\lambda)_+ + \frac{c(\gamma_-(\lambda)) - c(-\lambda)}{\gamma_-(\lambda) - (-\lambda)}. \end{aligned} \tag{10}$$

If the minimising (respectively maximising) x in (5) (resp. (6)) is not unique then we take the smallest such x. If there is no minimising x, then the function we are minimising is decreasing (resp. increasing) as $x \to \infty$, and we define $\gamma_+(\lambda) = \infty$ (resp. $\gamma_-(\lambda) = \infty$). In this case we also define $\theta_+(\lambda) = 0$ (resp. $\theta_-(\lambda) = 0$).

Remark 1. Although we have given formal definitions these quantities are best described pictorially. Given $\lambda > 0$, we consider points $(y, c(y))$ for $y < 0$ and more specifically the line segment joining $(y, c(y))$ with $(\lambda, c(\lambda))$. As y ranges over the negative reals we let $\theta_+(\lambda)$ be the steepest possible downward slope of this line segment, and we let $\gamma_+(\lambda)$ be the absolute value of the x-coordinate of the point where this maximum is attained. See Figs. 1 and 2.

The quantities $\theta_-(\lambda)$ and $\gamma_-(\lambda)$ are obtained by reflecting the picture. Alternatively, if we define the measure $\tilde{\mu}((-\infty, x]) = \mu([-x, \infty))$ then we obtain a correspondence between the pairs of definitions above—that is $\gamma_-^{\mu}(\lambda) = \gamma_+^{\tilde{\mu}}(\lambda)$, $\theta_-^{\mu}(\lambda) = \theta_+^{\tilde{\mu}}(\lambda)$ and $\mu_-(\lambda) = \tilde{\mu}_+(\lambda)$, with the obvious extension of the notation.

Remark 2. It is only possible to have $\gamma_+(\lambda) = \infty$ when $\int_{\{x \geqslant 0\}} x\mu(\mathrm{d}x) > \int_{\{x \leqslant 0\}} |x|\mu(\mathrm{d}x)$, see Fig. 2. If this is true, then $\gamma_+(\lambda) = \infty$ for all λ such that $c(\lambda) > \int_{\{x \leqslant 0\}} |x|\mu(\mathrm{d}x)$ (and if the support of μ is not bounded below, also when equality holds).

We take this opportunity to record some further relationships between the various quantities defined in (5)–(10). It follows from (5) and (6) that for $\lambda > 0$:

$$-c'(-\gamma_+(\lambda))_- \leqslant \theta_+(\lambda) \leqslant -c'(-\gamma_+(\lambda))_+, \tag{11}$$

$$c'(\gamma_-(\lambda))_+ \leqslant \theta_-(\lambda) \leqslant c'(\gamma_-(\lambda))_-, \tag{12}$$

so there is equality in (11) or (12) when there is no atom of μ at $-\gamma_+(\lambda)$ or $\gamma_-(\lambda)$, respectively. From Fig. 2 it is clear that if there is an atom of μ at $-\gamma_+(\lambda)$ then c has a kink there, and $-\theta_+(\lambda)$ is then the gradient of the line joining $c(-\gamma_+(\lambda))$ and $c(\lambda)$. Further, for $\lambda > 0$ such that $\gamma_+(\lambda), \gamma_-(\lambda) < \infty$, we have

$$c(\lambda) = c(-\gamma_+(\lambda)) - (\lambda + \gamma_+(\lambda))\theta_+(\lambda), \tag{13}$$

$$c(-\lambda) = c(\gamma_-(\lambda)) - (\lambda + \gamma_-(\lambda))\theta_-(\lambda). \tag{14}$$

Note that as a simple consequence of these equalities, $c(\lambda) \leqslant c(-\gamma_+(\lambda))$ and $c(-\lambda) \leqslant c(\gamma_-(\lambda))$.

Remark 3. By considering Figs. 1 and 2, we see that alternative definitions for $\gamma_+(\lambda)$, $\gamma_-(\lambda)$, $\theta_+(\lambda)$ and $\theta_-(\lambda)$ are

$$\gamma_+(\lambda) = -\sup\left\{x < 0 : \frac{c(\lambda) - c(x)}{\lambda - x} \leqslant c'(x)_+\right\}, \tag{15}$$

$$\gamma_-(\lambda) = \inf\left\{x > 0 : \frac{c(x) - c(-\lambda)}{x - (-\lambda)} \geqslant c'(x)_-\right\}, \tag{16}$$

$$\theta_+(\lambda) = -\frac{c(\lambda) - c(-\gamma_+(\lambda))}{\lambda - (-\gamma_+(\lambda))}, \tag{17}$$

$$\theta_-(\lambda) = \frac{c(\gamma_-(\lambda)) - c(-\lambda)}{\gamma_-(\lambda) - (-\lambda)}. \tag{18}$$

As a result it is easy to see that, in the case where μ is centered, these quantities are identical to the quantities defined in Perkins (1986), where the quantity $q_+(\lambda)$ defined in Perkins (1986) satisfies $\theta_+(\lambda) = q_+(\lambda) + \mu((-\infty, -\gamma_+(\lambda)))$.

Our first theorem shows that for any target measure μ there is an embedding which simultaneously stochastically maximises the distribution of the minimum, and minimises the distribution of the maximum.

Theorem 4. (1) *Let $(M_t)_{t\geqslant 0}$ be a continuous local martingale, vanishing at zero and with supremum process $S_t = \sup_{u\leqslant t} M_u$ and infimum process $J_t = -\inf_{u\leqslant t} M_u$, and let T be a stopping time such that $M_T \sim \mu$. Then, for all $\lambda \geqslant 0$, the following hold*:

$$\mathbb{P}(S_T \geqslant \lambda) \geqslant \mu_+(\lambda), \tag{19}$$

$$\mathbb{P}(J_T \geqslant \lambda) \geqslant \mu_-(\lambda). \tag{20}$$

(2) *For a continuous local martingale*, M_t, *vanishing at zero and such that* $\langle M\rangle_\infty = \infty$ *a.s.*, *with supremum process* $S_t = \sup_{u\leqslant t} M_u$ *and infimum process* $J_t = -\inf_{u\leqslant t} M_u$, *define the stopping time*

$$T = \inf\{t > 0\colon\ M_t \notin (-\gamma_+(S_t), \gamma_-(J_t))\}. \tag{21}$$

Then the stopped process M_T *has distribution* μ, *and equality holds in* (19) *and* (20).

Remark 5. When μ is centered, the fact that the quantities γ_+ and γ_- agree with those in Perkins (1986), and the fact that in this case T as defined in (21) is the Perkins stopping time, means that we know that T embeds μ. Moreover, we know that T minimises the law of the maximum, and maximises the law of the minimum. These results follows directly from Theorems 3.7 and 3.8 in Perkins (1986). The content of Theorem 4 is that these results can be extended to any choice of μ.

Remark 6. We may think of $\theta_+(\lambda)$ and $\theta_-(\lambda)$ as probabilities, and in particular, for the embedding defined in (21), $\theta_+(\lambda)$ is the probability that our process stops below $-\gamma_+(\lambda)$ but with a maximum above λ. If μ has no atom at $-\gamma_+(\lambda)$ then for this construction the maximum will be above λ if and only if our final value is above λ or below $-\gamma_+(\lambda)$. However, if there is an atom at $-\gamma_+(\lambda)$, the process may stop there without previously having reached λ. This event is represented graphically by the fact that there are multiple tangents to c at $-\gamma_+(\lambda)$. Also, when $\gamma_+(\lambda) = \infty$ for some λ, if the supremum of our process gets above λ before stopping, then our stopping rule becomes simply to wait until we reach some upper level, dependent on the infimum.

An alternative way to visualise the stopping time in (21) is shown in Fig. 3. We think of the process (J_t, S_t), and define the stopping time to be the first time it leaves the region defined via γ_+ and γ_- as shown.

The first half of the proof of Theorem 4 is a consequence of the following lemma.

Lemma 7. *Let* $(M_t)_{t\geqslant 0}$ *be a continuous local martingale. Suppose that* M *vanishes at zero*, M *converges a.s.*, *and that* $M_\infty \sim \mu$, *for some probability measure* μ *on* $\mathbb{R}$. *Then*, *for* $\lambda > 0$,

$$\mathbb{P}(S_\infty \geqslant \lambda) \geqslant \mu_+(\lambda), \tag{22}$$

$$\mathbb{P}(J_\infty \geqslant \lambda) \geqslant \mu_-(\lambda), \tag{23}$$

where $S_\infty = \sup_s M_s$, *and* $J_\infty = -\inf_s M_s$.

Proof. For $x < 0 < \lambda$, we define $H_\lambda = \inf\{t > 0\colon\ M_t \geqslant \lambda\}$, where we take $\inf\emptyset = \infty$. By examining on a case by case basis, we find that the following inequality holds:

$$\mathbf{1}_{\{S_\infty\geqslant\lambda\}} \geqslant \mathbf{1}_{\{M_\infty\geqslant\lambda\}} + \frac{1}{\lambda - x}[M_{H_\lambda} - (\lambda\wedge M_\infty)\mathbf{1}_{\{M_\infty\geqslant 0\}} + (|M_\infty|\wedge|x|)\mathbf{1}_{\{M_\infty<0\}}].$$

After taking expectations, this implies that

$$\mathbb{P}(S_\infty \geqslant \lambda) \geqslant c'(\lambda)_- + \frac{1}{\lambda - x}\,\mathbb{E}M_{H_\lambda} - \frac{c(\lambda) - c(x)}{\lambda - x}.$$

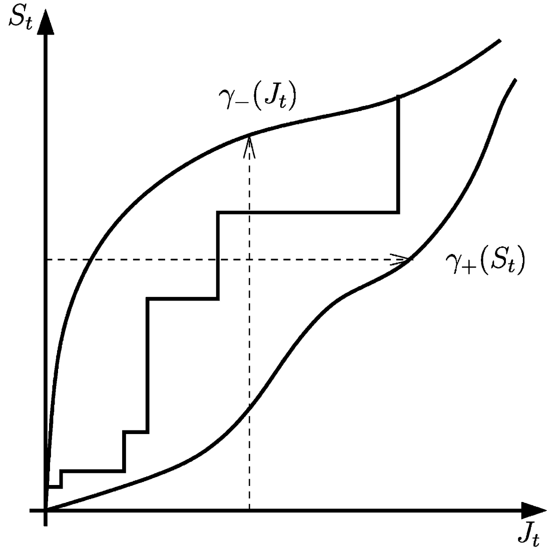

Fig. 3. The path of the process in the (J_t, S_t)-space. T is the first time this process leaves the region.

Now $M_{t\wedge H_\lambda}$ is a local martingale bounded above, and hence a submartingale, so $\mathbb{E}M_{H_\lambda} \geqslant M_0 = 0$. Substituting this in the above equation, we get

$$\mathbb{P}(S_\infty \geqslant \lambda) \geqslant c'(\lambda)_- - \frac{c(\lambda) - c(x)}{\lambda - x},$$

and since x is arbitrary,

$$\begin{aligned}\mathbb{P}(S_\infty \geqslant \lambda) &\geqslant c'(\lambda)_- + \sup_{x<0}\left\{\frac{c(x) - c(\lambda)}{\lambda - x}\right\}\\ &\geqslant \mu([\lambda,\infty)) + \theta_+(\lambda) = \mu_+(\lambda),\end{aligned}$$

which is (22).

We may deduce (23) using the correspondence $\mu \mapsto \tilde{\mu}$. □

Remark 8. In particular, for equality to hold for fixed λ in the above, we must have

(1) if $S_\infty \geqslant \lambda$, either $M_\infty \geqslant \lambda$ or $M_\infty \leqslant -\gamma_+(\lambda)$ a.s.,
(2) if $S_\infty < \lambda$, $M_\infty \geqslant -\gamma_+(\lambda)$ a.s.,
(3) $\mathbb{E}M_{H_\lambda} = 0$, so that $M_{t\wedge H_\lambda}$ is a true martingale.

It can be seen that these will hold simultaneously for all λ in the case where the stopping time is that given in Theorem 4, and that this is almost surely the only stopping time where (22) and (23) hold.

Proof of Theorem 4. We apply Lemma 7 to the process $(M_{T\wedge t})_{t\geqslant 0}$, which allows us to deduce (19) and (20).

For the second part of the theorem recall that if μ is centered then the Theorem follows from Theorems 3.7 and 3.8 in Perkins (1986). In the case when μ is not centered define

$$\xi_+^n = \inf\left\{x : \mu([x,\infty)) \leqslant \frac{1}{2n}\right\},$$

$$\xi_-^n = \sup\left\{x : \mu((-\infty,x]) \leqslant \frac{1}{2n}\right\},$$

and, for n sufficiently large, consider a sequence of measures μ^n satisfying

(i) $\mu^n((\alpha,\beta)) = \mu((\alpha,\beta)),\ \xi_-^n < \alpha \leqslant \beta < \xi_+^n$;
(ii) $\mu^n([\xi_-^n,\xi_+^n]) = \mu^n([(-n)\wedge\xi_-^n, n\vee\xi_+^n]) = \frac{n-1}{n}$;
(iii) $\mu^n(\{\xi_\pm^n\}) \leqslant \mu(\{\xi_\pm^n\})$;
(iv) $\int x\mu^n(\mathrm{d}x) = 0$;
(v) $\int |x|\mu^n(\mathrm{d}x) < \infty$.

We can construct such a sequence by redistributing the mass that lies in the tails of μ as follows: each μ^n agrees with μ on the interval (ξ_-^n,ξ_+^n), and mass is placed at the endpoints of this interval to satisfy (ii) and (iii) if there are atoms here; the remaining mass is then placed outside the interval $[(-n)\wedge\xi_-^n, n\vee\xi_+^n]$ in such a way as to ensure that (iv) and (v) hold.

For the rest of this section a superscript n will denote the fact that a quantity is calculated relative to the measure μ^n.

Note that if we can construct μ^n in such a way that $\mu^n(\mathbb{R}_-) = \mu(\mathbb{R}_-)$ then we find that $c^n(x) \equiv c(x)$ on $[\xi_-^n,\xi_+^n]$. However, it is not possible to construct μ^n with this additional property if $\mu(\mathbb{R}_-) = 0$ or 1, and in that case we need a more general argument.

Suppose $\mu^n(\mathbb{R}_-) - \mu(\mathbb{R}_-) = \psi_n$ for some number $\psi_n \in (-1/2n, 1/2n)$, then $c^n(x) = c(x) - \psi_n x$ for $x \in [\xi_-^n,\xi_+^n]$. If both λ and $\gamma_+(\lambda)$ lie in this interval then it is clear from (5) that $\gamma_+^n(\lambda) = \gamma_+(\lambda)$. Conversely if $\gamma_+(\lambda) = \infty$, then $\gamma_+^n(\lambda) \geqslant n$. Similar results hold for γ_-^n.

We define the stopping times associated with these measures,

$$T^n := \inf\{t > 0\colon M_t \notin (-\gamma_+^n(S_t), \gamma_-^n(J_t))\},$$

so that $M_{T^n} \sim \mu^n$. Note that if $M_{T^n} \in [(-n)\wedge\xi_-^n, n\vee\xi_+^n]$, then $T = T^n$ a.s. (see Fig. 4). However this implies that $\mathbb{P}(T = T^n) \to 1$, since these intervals are increasing to cover the whole of $\mathbb{R}$. Together with the fact that $\mu^n([\lambda,\infty)) \to \mu([\lambda,\infty))$, we conclude that $M_T \sim \mu$.

Finally, we need to show that our process attains equality in (19) and (20). Fix $\lambda > 0$. We know that

$$\mathbb{P}(S_{T^n} \geqslant \lambda) = \mu_+^n(\lambda) = \mu^n([\lambda,\infty)) + \theta_+^n(\lambda)$$

and since $\mathbb{P}(T^n = T) \geqslant (n-1)/n$, we have $\mathbb{P}(S_{T^n} \geqslant \lambda) \to \mathbb{P}(S_T \geqslant \lambda)$. Moreover, $\mu^n([\lambda,\infty)) \to \mu([\lambda,\infty))$ so that in order to prove

$$\mathbb{P}(S_T \geqslant \lambda) = \mu([\lambda,\infty)) + \theta_+^\mu(\lambda) = \mu_+(\lambda), \tag{24}$$

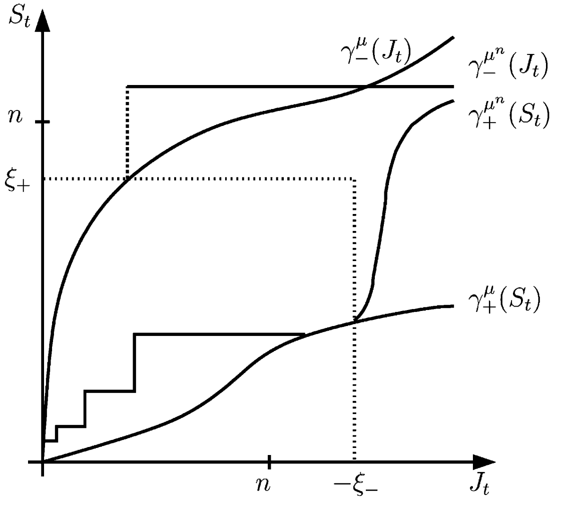

Fig. 4. The path of the process in the (J_t, S_t)-space, showing boundaries to embed both μ and μ^n. We have shown here a possible choice of μ^n in the case where $\xi_+ < n < (-\xi_-)$.

it is sufficient to show that $\theta_+^n(\lambda) \to \theta_+^\mu(\lambda)$ as $n \to \infty$. Now, when $x \in [\xi_-^n, \xi_+^n]$, we have $c^n(x) - c(x) = \psi_n x$ and for x outside this range $(c^n)' - c' \leqslant 1/n$. Hence

$$|c^n(x) - c(x)| \leqslant \frac{|x|}{n}$$

for all x. As a corollary, for $x < 0 < \lambda$,

$$\left| \frac{c^n(\lambda) - c^n(x)}{\lambda - x} - \frac{c(\lambda) - c(x)}{\lambda - x} \right| \leqslant \frac{1}{n},$$

from which it follows that

$$|\theta_+^{\mu^n}(\lambda) - \theta_+^\mu(\lambda)| \leqslant \frac{1}{n}.$$

using representation (17).

As before we can also show (20) holds by using the correspondence $\mu \mapsto \tilde{\mu}$. □

3. Applications to diffusions

We now work with the class of regular (time-homogeneous) diffusions (see Rogers and Williams (2000), V. 45) $(Y_t)_{t \geqslant 0}$ on an interval $I \subseteq \mathbb{R}$, with absorbing or inaccessible endpoints, and vanishing at zero. Consider the problem of determining when and how we may embed a distribution v on I° in the diffusion. Since the diffusion is regular, there exists a continuous, strictly increasing scale function $s : I \to \mathbb{R}$ such that $M_t = s(Y_t)$ is a diffusion on natural scale on $s(I)$. We may also choose s such that $s(0) = 0$. In particular, M_t is (up to exit from the interior of $s(I)$) a time change of a Brownian motion, with strictly positive speed measure.

If we now define the measure μ on $s(I)$ by

$$\mu(A) = \nu(s^{-1}(A)), \quad A \subseteq s(I), \text{ Borel},$$

then our problem is equivalent to that of embedding μ in a Brownian motion before it leaves $s(I)^\circ$. This is because M is a local martingale on $s(I)^\circ$, and hence a time change of a Brownian motion on $s(I)^\circ$, and if we construct a stopping time T such that $M_T = s(Y_T) \sim \mu$, then $Y_T \sim \nu$. In this context it makes sense to consider ν and μ as measures on $\mathbb{R}$ which place all their mass on I° and $s(I)^\circ$ respectively. Our approach will be to use the embedding we established in Theorem 4 to embed μ in the local martingale M, and our first step will be to transfer the framework of the previous section to our new setting.

An advantage of using the embedding we established in Section 2 in this situation is that, because we have a strictly increasing scale function, the properties of the maximum and the minimum are preserved. In particular, this transformed stopping time will maximise the distribution of the minimum, and minimise the distribution of the maximum of the process $(Y_{T \wedge t})$ among all stopping times of Y_t with $Y_T \sim \nu$.

The first question that it is necessary to ask is: when is it possible to embed a given target law? This is exactly the question considered by Rost (1971) using potentials, but we want a more direct criterion. In the diffusion case it is no longer possible to embed all target laws, as can be witnessed by considering the problem of embedding unit mass at -1 in Brownian motion with positive drift. The result we need was first proved in Pedersen and Peskir (2001).

Lemma 9 (Pedersen and Peskir (2001), Theorem 2.1). *There are three different cases*:

(1) $s(I)^\circ = \mathbb{R}$, *in which case the diffusion is recurrent*, *and we can embed any distribution* ν *on* I° *in* Y,
(2) $s(I)^\circ = (-\infty, \alpha)$ (*respectively* (α, ∞)) *for some* $\alpha \in \mathbb{R}$. *Then we may embed* ν *in* Y *if and only if* $m = \int_I s(y)\nu(\mathrm{d}y)$ *exists*, *and* $m \geqslant 0$ (*resp.* $m \leqslant 0$).
(3) $s(I)^\circ = (\alpha, \beta)$, $\alpha, \beta \in \mathbb{R}$. *Then we may embed* ν *in* Y *if and only if* $m = 0$.

The statement of the result in Pedersen and Peskir (2001) has the additional assumption in Case (1) that $\int_I |s(y)|\nu(\mathrm{d}y) < \infty$. This can be dropped since in Case (1) the diffusion is recurrent so that either the simple stopping time defined the introduction, or the extension of the Perkins embedding we introduced in the previous section, can be used to embed μ.

For the precise details of the proof of Lemma 9 we refer the reader to Pedersen and Peskir (2001). However, we can provide a sketch of the proof using the modified Perkins embedding. For t less than the first exit time of the diffusion from the interior of $s(I)$ we have $M_t = s(Y_t) = X_{\tau_t}$ for some time-change τ and Brownian motion X. If $\langle M \rangle_\infty = \tau_\infty < \infty$ we may extend the time domain on which X_{τ_t} is defined to all positive times by continuing the Brownian motion beyond τ_∞. In this way, we may drop the assumption of Theorem 4 that the process M_t has infinite variation. We deduce that we may embed our distribution on $s(I)^\circ$ if and only if, when we consider the problem of embedding μ in Brownian motion, our process remains on $s(I)^\circ$. However the transformed target distribution has support concentrated only on this interval, so when we consider the stopping time T defined in (21) and the form of $\gamma_+(\lambda)$ and $\gamma_-(\lambda)$ in the martingale scale, we see that problems can only occur if $\gamma_+(\lambda)=\infty$ or $\gamma_-(\lambda)=\infty$ for some λ. Further examination shows that this is only possible when μ is not integrable, or not centered—see Remark 2—and the three cases of Lemma 9 all follow.

Our aim in the remainder of this section is to look at some of the properties of the construction, and of embeddings in general. Our principal question is (cf. Perkins (1986) and Jacka (1988), where the law of $\sup|Y_t|$ in the Brownian case with centered target distribution is considered),

> given a diffusion Y_t, and a law v, when does there exists an embedding for which the law of the maximum modulus of the process, $\sup_t |Y_{T\wedge t}|$, lies in the space $\mathscr{L}^p$ of random variables with finite pth moment?

Before answering this question we show how the results of the previous section can be used to define an embedding of a target law in a diffusion.

Given v and $(Y_t)_{t\geqslant 0}$ define μ and $M=s(Y)$ as above. As before, for M on $s(I)$ we can define

$$c_M(x)=\begin{cases}\displaystyle\int_{\{u\geqslant 0\}}(x\wedge u)\mu(\mathrm{d}u), & x\geqslant 0,\\ \displaystyle\int_{\{u<0\}}(|x|\wedge |u|)\mu(\mathrm{d}u), & x<0,\end{cases}$$

together with the quantities defined in (5)–(10). Write

$$c_Y(y)=c_M(s(y))=\begin{cases}\displaystyle\int_{\{w\geqslant 0\}}(s(y)\wedge s(w))v(\mathrm{d}w), & y\geqslant 0,\\ \displaystyle\int_{\{w<0\}}(|s(y)|\wedge |s(w)|)v(\mathrm{d}w), & y<0.\end{cases}$$

and, for $z>0$, define the quantities

$$\rho_+(z)=\operatorname*{argmin}_{y>0}\left\{\frac{c_Y(z)-c_Y(-y)}{s(z)-s(-y)}\right\}, \tag{25}$$

$$\rho_-(z)=\operatorname*{argmax}_{y>0}\left\{\frac{c_Y(y)-c_Y(-z)}{s(y)-s(-z)}\right\}, \tag{26}$$

$$\zeta_+(z)=-\inf_{y>0}\left\{\frac{c_Y(z)-c_Y(-y)}{s(z)-s(-y)}\right\}, \tag{27}$$

$$\zeta_-(z)=\sup_{y>0}\left\{\frac{c_Y(y)-c_Y(-z)}{s(y)-s(-z)}\right\}, \tag{28}$$

$$\nu_+(z) = \zeta_+(z) + \nu([z,\infty)), \tag{29}$$

$$\nu_-(z) = \nu((-\infty,-z]) + \zeta_-(z). \tag{30}$$

By convention, if $\rho_+(z)$ or $\rho_-(z)$ is not uniquely defined then we take the smallest solution.

Now define a stopping time for Y_t by

$$\begin{aligned} T &= \inf\{t > 0:\ Y_t \notin (-\rho_+(S_t^Y), \rho_-(J_t^Y))\} \\ &= \inf\{t > 0:\ M_t \notin (-\gamma_+(S_t^M), \gamma_-(J_t^M))\}, \end{aligned} \tag{31}$$

where we write $S_t^Y = \sup_{s\leqslant t} Y_s$, $J_t^Y = -\inf_{s\leqslant t} Y_s$, $S_t^M = \sup_{s\leqslant t} M_s$ and $J_t^M = -\inf_{s\leqslant t} M_s$. The two alternative characterisations of T are equivalent because of the identities

$$s(-\rho_+(z)) = -\gamma_+(s(z)),$$

$$s(\rho_-(z)) = \gamma_-(-s(-z)).$$

We also have that $\zeta_+(z) = \theta_+(s(z))$, and $\zeta_-(z) = \theta_-(-s(-z))$. It follows that T embeds μ in $(M_t)_{t\geqslant 0}$, and hence ν in $(Y_t)_{t\geqslant 0}$. Also ν_+ and ν_- are the laws of the supremum and infimum, respectively, of $Y_{T\wedge t}$. Consequently, we may restate Theorem 4 in the diffusion context.

Theorem 10. *Let $(Y_t)_{t\geqslant 0}$ be a regular, time-homogeneous diffusion, vanishing at zero and with supremum process S_t^Y and infimum process J_t^Y, and let T be a stopping time such that $Y_T \sim \nu$. Then, for all $\lambda \geqslant 0$, the following hold:*

$$\mathbb{P}(S_T^Y \geqslant \lambda) \geqslant \nu_+(\lambda), \tag{32}$$

$$\mathbb{P}(J_T^Y \geqslant \lambda) \geqslant \nu_-(\lambda). \tag{33}$$

If there exists an embedding, the stopping time T defined in (31) is an embedding and is optimal in the sense that it attains equality in (32) and (33).

We are interested in the measure ν_* where ν_* is the law of $\sup_{t\leqslant T} |Y_t|$. Trivially, for $z \geqslant 0$,

$$\max(\nu_+(z), \nu_-(z)) \leqslant \nu_*([z,\infty)) \leqslant \nu_+(z) + \nu_-(z), \tag{34}$$

and it follows that $\nu_* \in \mathscr{L}^p$ if and only both ν_+ and ν_- are elements of $\mathscr{L}^p$.

The next two lemmas give upper and lower bounds on ν_+ and ν_-. We give proofs in the case of ν_+; the corresponding results for ν_- can be deduced using the transformation $\mu \mapsto \tilde{\mu}$.

Lemma 11. *For all $z > 0$, we have*

$$\begin{aligned} \nu_+(z) \leqslant {} & \frac{1}{s(z)}\,[c_Y(-z) - c_Y(z) - |s(-z)|\nu((-\infty,-z])]_+ \mathbf{1}_{\{z > \rho_+(z)\}} \\ & + \nu(\{|y| \geqslant z\}), \end{aligned} \tag{35}$$

$$\nu_-(z) \leqslant \frac{1}{|s(-z)|}[c_Y(z) - c_Y(-z) - s(z)\nu([z,\infty))]_+ \mathbf{1}_{\{z>\rho_-(z)\}} \\ + \nu(\{|y| \geqslant z\}). \tag{36}$$

Proof. Suppose first that $z > \rho_+(z)$, or equivalently $s(-z) < -\gamma_+(s(z))$. Then by the convexity of c_M on $\mathbb{R}_-$,

$$c_M(-\gamma_+(s(z))) - \gamma_+(s(z))\theta_+(s(z)) \leqslant c_M(s(-z)) + s(-z)\nu((-\infty,-z]),$$

which translates to

$$c_Y(-\rho_+(z)) + s(-\rho_+(z))\zeta_+(z) \leqslant c_Y(-z) + s(-z)\nu((-\infty,-z]).$$

Substituting this inequality into (27) we deduce that

$$s(z)\zeta_+(z) = s(-\rho_+(z))\zeta_+(z) + c_Y(-\rho_+(z)) - c_Y(z) \\ \leqslant c_Y(-z) - c_Y(z) + s(-z)\nu((-\infty,-z]).$$

Conversely, if $z \leqslant \rho_+(z)$, then

$$\zeta_+(z) \leqslant \nu((-\infty,-\rho_+(z)]) \leqslant \nu((-\infty,-z]).$$

Given that $\nu_+(z) = \nu([z,\infty)) + \zeta_+(z)$, these two bounds lead directly to (35). □

Lemma 12. *For all* $z > 0$, *we have*

$$\nu_+(z) \geqslant \frac{[c_Y(-z) - c_Y(z)]_+}{s(z) + |s(-z)|} + \nu([z,\infty)), \tag{37}$$

$$\nu_-(z) \geqslant \frac{[c_Y(z) - c_Y(-z)]_+}{s(z) + |s(-z)|} + \nu((-\infty,-z]). \tag{38}$$

Proof. By (27), for $z > 0$,

$$\zeta_+(z) \geqslant \frac{c_Y(-z) - c_Y(z)}{s(z) + |s(-z)|}.$$

Since also $\zeta_+(z) \geqslant 0$ the result follows easily from the identity $\nu_+(z) = \nu([z,\infty)) + \zeta_+(z)$. □

Corollary 13. *For* $z > 0$, *we have*

$$\left(\frac{1}{s(z)} + \frac{1}{|s(-z)|}\right)|c_Y(z) - c_Y(-z)| + 2\nu(\{|y| \geqslant z\}) \\ \geqslant \nu_+(z) + \nu_-(z) \geqslant \frac{|c_Y(z) - c_Y(-z)|}{s(z) + |s(-z)|} + \nu(\{|y| \geqslant z\}).$$

Let T' be an embedding of ν in Y. For $p > 0$ we say this embedding is a H^p-embedding if $\sup_t |Y_{t\wedge T'}|$ is in $\mathscr{L}^p$. We may ask when does there exist a solution of the Skorokhod problem which is a H^p-embedding, and when is every solution of the Skorokhod problem a H^p-embedding? In this paper we are interested in the first of these questions. By the extremality properties of our embedding T it is clear that there exists a H^p-embedding if and only if T is a H^p-embedding.

Corollary 13 can be used to give necessary and sufficient conditions for v_* to be an element of $\mathscr{L}^p$. In particular, the following result follows easily from Corollary 13 and (34).

Theorem 14. *Let Y_t be a regular diffusion and suppose that v can be embedded in Y. Consider the embedding T of v given in* (31). *A sufficient condition for T to be a H^p-embedding is that $v \in \mathscr{L}^p$ and*

$$\int^{\infty} z^{p-1}\left(\frac{1}{s(z)}+\frac{1}{|s(-z)|}\right)|c_Y(z)-c_Y(-z)|\,\mathrm{d}z<\infty. \tag{39}$$

Necessary conditions are that $v \in \mathscr{L}^p$ and

$$\int_0^{\infty} z^{p-1}\frac{|c_Y(z)-c_Y(-z)|}{s(z)+|s(-z)|}\,\mathrm{d}z<\infty. \tag{40}$$

Remark 15. Note that in the symmetric case where $s(z)=-s(-z)$ then (39) and (40) are equivalent and Theorem 14 gives a necessary and sufficient condition for T to be a H^p-embedding.

We return to the problem of the existence of a H^p-embedding in the next section, and close this section with a further observation about the optimality of the embedding T.

Remark 16. Fix a measurable function $f:\mathbb{R}\to\mathbb{R}$ and let $(Y_t)_{t\geqslant 0}$ be a regular diffusion with $Y_0=0$ and v a probability measure on $\mathbb{R}$. Then the embedding defined in (31) minimises the distribution of $\sup_{t\geqslant 0} f(Y_{t\wedge T'})$ over all stopping times T' such that $Y_{T'}\sim v$.

In particular the minimising choice of stopping time does not depend on the function f. This is in contrast with the problem of finding the Skorokhod embedding which maximises the law of $\sup_{t\geqslant 0} f(Y_{t\wedge T'})$. In that case the optimal embedding will in general depend on f.

4. H^p embeddings for diffusions

Our goal in this section is to investigate further conditions on whether T is a H^p-embedding in the cases when $s(I)^\circ=(-\infty,\alpha),(\alpha,\infty),(\beta,\alpha)$ and $\mathbb{R}$. The first two cases are equivalent up to the map $x\mapsto -x$ and we consider them first.

4.1. Diffusions transient to $+\infty$

Theorem 17. *Let Y_t be a diffusion on I with scale function $s(z)$, such that $s(0)=0$, $\sup_{z\in I} s(z)=\alpha<\infty$, and $\inf_{z\in I} s(z)=-\infty$. We may embed a law v in Y if and only if $\int_I |s(z)|v(\mathrm{d}z)<\infty$ and $m=\int_I s(z)v(\mathrm{d}z)\geqslant 0$.*

Under these conditions:

- *if* $m>0$, *then a necessary and sufficient condition for* $\mathbb{E}\sup_t |Y_{T\wedge t}|^p<\infty$ *is that*

$$\int^{\infty} \frac{z^{p-1}}{|s(-z)|}\,\mathrm{d}z<\infty \quad \textit{and } \nu\in\mathscr{L}^p, \tag{41}$$

- *if* $m=0$, *this is also a sufficient condition. A necessary and sufficient condition is*

$$\int^{\infty} \frac{z^{p-1}}{|s(-z)|}\,|c_Y(z)-c_Y(-z)|\,\mathrm{d}z<\infty \quad \textit{and } \nu\in\mathscr{L}^p. \tag{42}$$

Proof. The first part of this Theorem is a restatement of Lemma 9(2) (or equivalently Pedersen and Peskir (2001) (Theorem 2.1)). For the second part assume $m\geqslant 0$, where $m=\int_0^\infty s(y)\nu(\mathrm{d}y)-\int_{-\infty}^0 |s(y)|\nu(\mathrm{d}y)$. For $z\geqslant 0$,

$$\begin{aligned} c_Y(-z)-c_Y(z) &= -\int_{\{y<-z\}} |s(y)|\nu(\mathrm{d}y)+\int_{\{y>z\}} s(y)\nu(\mathrm{d}y)-m \\ &\quad +\int_{\{y\leqslant -z\}} |s(-z)|\nu(\mathrm{d}y)-\int_{\{y\geqslant z\}} s(z)\nu(\mathrm{d}y) \\ &\leqslant \int_{\{y>z\}} s(y)\nu(\mathrm{d}y)+\int_{\{y\leqslant -z\}} |s(-z)|\nu(\mathrm{d}y), \end{aligned}$$

so by Lemma 11,

$$\begin{aligned} \nu_+(z) &\leqslant \frac{1}{s(z)}\,[c_Y(-z)-c_Y(z)-|s(-z)|\nu((-\infty,-z])]_+\mathbf{1}_{\{z>\rho_+(z)\}} \\ &\quad +\nu(\{|y|\geqslant z\}) \\ &\leqslant \int_{\{y>z\}} \frac{s(y)}{s(z)}\nu(\mathrm{d}y)+\nu(\{|y|\geqslant z\}) \\ &\leqslant \frac{\alpha}{s(z)}\nu(\{|y|\geqslant z\}). \end{aligned}$$

Since $\alpha/s(z)<2$ for sufficiently large z it follows that $\nu\in\mathscr{L}^p$ is a necessary and sufficient condition for $\nu_+\in\mathscr{L}^p$.

Now consider $\nu_-(z)$. We note that given $\varepsilon>0$, for sufficiently large z,

$$m-\varepsilon\leqslant c_Y(z)-c_Y(-z)\leqslant m+\varepsilon,$$

and so by Lemma 11,

$$\nu_-(z) \leqslant \frac{1}{|s(-z)|}(m+\varepsilon) + \nu(\{|y| \geqslant z\}).$$

As a result (41) is a sufficient condition for $\nu_- \in \mathscr{L}^p$ when $m \geqslant 0$.

Conversely, if $m > 0$ Lemma 12 implies that for sufficiently large z,

$$\nu_-(z) \geqslant \frac{1}{2|s(-z)|}(m-\varepsilon),$$

and so (41) is also necessary.

Now suppose $m = 0$. By (36),

$$\nu_-(z) \leqslant \frac{1}{|s(-z)|}[c_Y(z) - c_Y(-z)]_+ + \nu(\{|y| \geqslant z\}),$$

so (42) is a sufficient condition for $\nu_- \in \mathscr{L}^p$. By Corollary 13, for sufficiently large z,

$$\nu_+(z) + \nu_-(z) \geqslant \frac{|c_Y(z) - c_Y(-z)|}{2|s(-z)|} + \nu(\{|y| \geqslant z\}).$$

If $\nu_* \in \mathscr{L}^p$ then both ν_+ and ν_- lie in $\mathscr{L}^p$, and so (42) is a necessary condition. □

Example 18 (Drifting Brownian motion). Suppose Y is drifting Brownian motion on $\mathbb{R}$,

$$Y_t = B_t + \kappa t,$$

for $t \geqslant 0$ and $\kappa > 0$. Then $s(y) = 1 - \mathrm{e}^{-2\kappa y}$ is the scale function for Y, so $\sup_y s(y) = 1$. If $\int_{\mathbb{R}} s(y)\nu(\mathrm{d}y) < 0$, then it is not possible to embed ν in Y. If $\int_{\mathbb{R}} s(y)\nu(\mathrm{d}y) \geqslant 0$, we may embed ν in Y, and since

$$\int^{\infty} \frac{y^{p-1}}{|s(-y)|}\,\mathrm{d}y = \int^{\infty} \frac{y^{p-1}}{\mathrm{e}^{2\kappa y} - 1}\,\mathrm{d}y < \infty,$$

if follows that if $\nu \in \mathscr{L}^p$, then $\sup_t |Y_{T\wedge t}|$ is too.

These conclusions should be compared with those in Grandits and Falkner (2000). Grandits and Falkner conclude that if Y is drifting Brownian motion, and if T' is any embedding of ν in Y, then $T' \in H^1$.

Example 19 (Bessel d process). In Hambly et al. (2003) the authors consider a Skorokhod embedding for the BES(3) process. For $d > 2$ let Y solve

$$\mathrm{d}Y_t = \mathrm{d}B_t + \frac{d-1}{2Y_t}\,\mathrm{d}t, \qquad Y_0 = 1.$$

Then $I = (0, \infty)$ and $s(y) = -y^{2-d}$. We do not have $Y_0 = 0$, nor $s(0) = 0$ but the modifications to the theory are trivial. We can embed ν in Y if and only if $\int_0^\infty y^{2-d}\nu(\mathrm{d}y) < 1$. Furthermore Y is only defined on the positive reals, so in deciding whether $\nu_* \in \mathscr{L}^p$ we need only consider ν_+. But, provided we may embed ν in Y, it follows from the proof of Theorem 17 that a necessary and sufficient condition for $\nu_+ \in \mathscr{L}^p$ is $\nu \in \mathscr{L}^p$.

4.2. Recurrent diffusions

The general case is covered by Theorem 14. If we have some control on the scale function then we are able to make the results more explicit.

Theorem 20. *Suppose for $|y| \geqslant 1$ there exists $k, K > 0$ such that*

$$k|y|^r \leqslant |s(y)| \leqslant K|y|^q, \quad \textit{for some } q \geqslant r \geqslant 0. \tag{43}$$

Then for $p > 0$,

(1) *if $p > q$,*

$$m = 0 \textit{ and } \nu \in \mathscr{L}^{p+q-r} \Rightarrow \nu_* \in \mathscr{L}^p \Rightarrow \nu \in \mathscr{L}^p \textit{ and } m = 0;$$

(2) *if $p < r$,*

$$\nu \in \mathscr{L}^{p+q-r} \Rightarrow \nu_* \in \mathscr{L}^p \Rightarrow \nu \in \mathscr{L}^p;$$

(3) *if $r \leqslant p \leqslant q$,*

$$\int_1^\infty y^{p-r-1}|c_Y(y) - c_Y(-y)|\,\mathrm{d}y < \infty \textit{ and } \nu \in \mathscr{L}^p$$

$$\Rightarrow \nu_* \in \mathscr{L}^p \tag{44}$$

$$\Rightarrow \nu \in \mathscr{L}^p \textit{ and } \int_0^\infty y^{p-q-1}|c_Y(y) - c_Y(-y)|\,\mathrm{d}y < \infty. \tag{45}$$

In particular, if $r = q$, the three cases each become if and only if statements.

Remark 21. The case where the diffusion is in natural scale, so that $s(y) = y$, is the case considered by Perkins (1986). Here the Cases (1) and (2) are dealt with in his introduction, while in Case (3) he shows that $\nu \in \mathscr{L}^1$, $m = 0$ and $H(\mu) < \infty$, where

$$H(\mu) = \int_0^\infty y^{-1} \left| \int_{-\infty}^\infty x \mathbf{1}_{\{|x| \geqslant y\}} \mu(\mathrm{d}x) \right| \mathrm{d}y,$$

are necessary and sufficient conditions for $\nu_* \in \mathscr{L}^1$. It is not hard to see that this condition is equivalent to (45).

Proof. (1) Suppose $p > q$. If $\nu \in \mathscr{L}^q$ then since $|s(y)| \leqslant K|y|^q$ for $|y| \geqslant 1$, we have $\int |s(y)|\nu(\mathrm{d}y) < \infty$, so m exists.

Now suppose $m = 0$ and $\nu \in \mathscr{L}^{p+q-r}$. By Theorem 14 it is sufficient to show

$$\int_1^\infty y^{p-1} \left(\frac{1}{s(y)} + \frac{1}{|s(-y)|} \right) |c_Y(y) - c_Y(-y)|\,\mathrm{d}y < \infty.$$

For $y>0$,

$$\begin{aligned}
&c_Y(y)-c_Y(-y)\\
&\quad=\int_{\{|w|\leqslant y\}} s(w)\nu(\mathrm{d}w)+\int_{\{w>y\}} s(y)\nu(\mathrm{d}w)-\int_{\{w<-y\}} |s(-y)|\nu(\mathrm{d}w)\\
&\quad=-\int_{\{|w|>y\}} s(w)\nu(\mathrm{d}w)+s(y)\nu(\{w>y\})-|s(-y)|\nu(\{w<-y\}),
\end{aligned}$$

where we have used the fact that $m=0$. By assumption

$$\left(\frac{1}{s(y)}+\frac{1}{|s(-y)|}\right)\leqslant \frac{2}{ky^r}\quad \text{for } y\geqslant 1,$$

so that

$$\begin{aligned}
&\int_1^\infty y^{p-1}\left(\frac{1}{s(y)}+\frac{1}{|s(-y)|}\right)|c_Y(y)-c_Y(-y)|\,\mathrm{d}y\\
&\quad\leqslant \frac{2}{k}\int_1^\infty y^{p-r-1}\left[Ky^q\nu((y,\infty))+Ky^q\nu((-\infty,-y))\right.\\
&\qquad\left.+\int_{\{|w|>y\}}|s(w)|\nu(\mathrm{d}w)\right]\mathrm{d}y.
\end{aligned}$$

The first two terms in the bracket will be finite upon integration since $\nu\in\mathscr{L}^{p+q-r}$. Also, by Fubini,

$$\begin{aligned}
\int_1^\infty y^{p-r-1}\left[\int_{\{w>y\}} s(w)\nu(\mathrm{d}w)\right]\mathrm{d}y &= \int_{\{w>1\}}\left[\int_1^w y^{p-r-1}s(w)\,\mathrm{d}y\right]\nu(\mathrm{d}w)\\
&\leqslant K\int_{\{w>1\}}\frac{w^{q+p-r}}{p+q-r}\nu(\mathrm{d}w)<\infty.
\end{aligned}$$

We can show a similar result for the integral over $\{w<0\}$ and it follows that $\nu_*\in\mathscr{L}^p$.

Now suppose that $\nu_*\in\mathscr{L}^p$. Then clearly $\nu\in\mathscr{L}^p$, and

$$\mathbb{E}\sup_t|s(Y_{T\wedge t})|\leqslant K\mathbb{E}\left(\sup_t|Y_{T\wedge t}|^q+1\right)\leqslant K\mathbb{E}\left(\sup_t|Y_{T\wedge t}|^p\right)+K<\infty.$$

Furthermore $s(Y_t)$ is a local martingale, so, since $\mathbb{E}\sup_t|s(Y_{T\wedge t})|<\infty$, $s(Y_{T\wedge t})$ is a UI martingale, and hence

$$m=\mathbb{E}(s(Y_T))=0.$$

(2) Suppose now $p<r$, and $\nu\in\mathscr{L}^{p+q-r}$. Then as before, by Theorem 14 it is sufficient to show

$$\int_1^\infty y^{p-1}\left(\frac{1}{s(y)}+\frac{1}{|s(-y)|}\right)|c_Y(y)-c_Y(-y)|\,\mathrm{d}y<\infty.$$

A simple inequality gives

$$|c_Y(y) - c_Y(-y)| \leqslant c_Y(y) + c_Y(-y)$$
$$= \int_{\{|w| \leqslant y\}} |s(w)|\nu(\mathrm{d}w) + s(y)\nu(\{w > y\}) + |s(-y)|\nu(\{w < -y\}),$$

and so

$$\int_1^\infty \left(\frac{1}{s(y)} + \frac{1}{|s(-y)|}\right) |c_Y(y) - c_Y(-y)|\,\mathrm{d}y$$
$$\leqslant \frac{2}{k}\int_1^\infty y^{p-r-1}\left[K y^q \nu(\{|w| > y\}) + \int_{\{|w| \leqslant y\}} |s(w)|\nu(\mathrm{d}y)\right]\mathrm{d}y,$$

where, as before, the first term is finite upon integration. For the final term

$$\int_1^\infty y^{p-r-1}\left[\int_{\{0<w\leqslant y\}} s(w)\nu(\mathrm{d}w)\right]\mathrm{d}y$$
$$= \int_{\{w>0\}} s(w)\left[\int_{w\vee 1}^\infty y^{p-r-1}\,\mathrm{d}y\right]\nu(\mathrm{d}w)$$
$$\leqslant \int_{\{w>0\}} \frac{(w\vee 1)^{p-r}}{r-p} s(w)\nu(\mathrm{d}w)$$
$$\leqslant \int_0^1 \frac{s(w)}{r-p}\nu(\mathrm{d}w) + \frac{K}{r-p}\int_{\{w>1\}} |w|^{p+q-r}\nu(\mathrm{d}w),$$

which is finite by assumption since $\nu \in \mathscr{L}^{p+q-r}$. The corresponding result also holds over $\{w < 0\}$. So we have shown $\nu \in \mathscr{L}^{p+q-r} \Rightarrow \nu_* \in \mathscr{L}^p$. The second implication $\nu_* \in \mathscr{L}^p \Rightarrow \nu \in \mathscr{L}^p$ is clear.

(3) This case is a trivial application of (43) to Theorem 14. □

For the integral condition in (44) to hold, a necessary condition is that $|c_Y(z) - c_Y(-z)| \to 0$ as $z \to \infty$. However this occurs if and only if $m = 0$, provided m exists. So if m exists, if $r = p = q$ and if $\nu \in \mathscr{L}^p$, then $m = 0$ is a necessary condition for $\nu_* \in \mathscr{L}^p$. We show in Example 22 that this condition is not sufficient.

Note that it is not necessary for m to exist for the integral condition in (40) to be satisfied, and for ν_* to be an element of $\mathscr{L}^p$. For example, suppose that both the scale function and the target measure are symmetric about 0, i.e. suppose $s(z) = -s(-z)$ and $\nu(\mathrm{d}z) = \nu(\mathrm{d}(-z))$. Then $c_Y(z) = c_Y(-z)$ and (40) is trivially satisfied. If s and ν are symmetric then $\nu_* \in \mathscr{L}^p$ if and only if $\nu \in \mathscr{L}^p$.

Example 22. We now consider a diffusion on $\mathbb{R}$ with behaviour specified by

$$\mathrm{d}Y_t = 2\sqrt{|Y|_t}\mathrm{d}B_t + \alpha\,\mathrm{sign}(Y_t)\,\mathrm{d}t,$$

where $Y_0 = 0$, and $\alpha \in (0,2)$. The solution to this SDE is not unique in law, but we make it so by assuming the law of the process is symmetric about 0, and that the process does not wait at 0. In particular, $|Y_t|$ is a Bessel process of dimension α. Such a process is recurrent, and we can construct the process Y_t from $|Y_t|$ by assigning to each excursion away from 0 an independent random variable with value either 1 or -1. Alternatively we may define the process by its scale function

$$s(y) = (|y|^{1-\alpha/2})\,\mathrm{sign}(y),$$

and write $Y_t = s(W_{A_t})$, for a Brownian motion W_t and a suitable time change A_t. Since $(Y_t)_{t\geqslant 0}$ is recurrent on $\mathbb{R}$ we may embed any target distribution. We may apply Theorem 20 to this process for some target distribution v and examine the behaviour of $\sup_t |Y_{T\wedge t}|$, for our embedding T. We note that, using the notation of Theorem 20, $r = q = 1 - \alpha/2$, so the statements in the theorem each become if and only if statements. We can consider each case separately:

(1) In the case, where $p > 1 - \alpha/2$, $v \in \mathscr{L}^p$ guarantees that m exists, and a necessary and sufficient condition for $\sup_t |Y_{T\wedge t}| \in \mathscr{L}^p$ is that $m = 0$.

(2) If $p < 1 - \alpha/2$, $v \in \mathscr{L}^p$ is both necessary and sufficient for $\sup_t |Y_{T\wedge t}| \in \mathscr{L}^p$.

(3) Suppose now that $p = 1 - \alpha/2$. If $m \neq 0$ then $\sup_t |Y_{T\wedge t}| \notin \mathscr{L}^p$. However, we now show that $m = 0$ is not a sufficient condition for $\sup_t |Y_{T\wedge t}| \in \mathscr{L}^p$.

We embed the probability measure v defined by

$$v(\mathrm{d}y) = \frac{y^{-p-1}}{(\log y)^2}\,\mathrm{d}y \quad \text{for } y \geqslant \mathrm{e},$$

with the rest of the mass placed at $-b$. Here b is chosen such that $\int s(y)v(\mathrm{d}y) = 0$. It can be checked that $v \in \mathscr{L}^p$. Then, provided $z > \max(\mathrm{e}, -s^{-1}(-b))$,

$$\begin{aligned} |c_Y(z) - c_Y(-z)| &= \int_z^\infty \frac{1}{y(\log y)^2}\,\mathrm{d}y - z^p v((z,\infty)) \\ &= \frac{1}{\log z} - z^p v([z,\infty)). \end{aligned}$$

Consequently, because $v \in \mathscr{L}^p$ and $\int_z^\infty 1/(y\log(y))\,\mathrm{d}y = \infty$,

$$\int^\infty y^{-1}|c_Y(y) - c_Y(-y)|\,\mathrm{d}y = \infty.$$

So $m = 0$ is not sufficient to ensure that $\sup_t |Y_{T\wedge t}| \in \mathscr{L}^p$.

4.3. Diffusions which in natural scale have state space consisting of a finite interval

Theorem 23. *Let Y_t be a diffusion on I with scale function $s(z)$, such that $s(0) = 0$, $\sup_{z\in I} s(z) = \alpha < \infty$, and $\inf_{z\in I} s(z) = \beta > -\infty$. We may embed a law v in Y if and only if $\int_I |s(z)|v(\mathrm{d}z) < \infty$ and $m = \int_I s(z)v(\mathrm{d}z) = 0$.*

Furthermore $v_* \in \mathscr{L}^p$ if and only if $v \in \mathscr{L}^p$.

Proof. The first part of this result follows from Lemma 9(3) (or equivalently Pedersen and Peskir (2001) (Theorem 2.1)). The remaining part follows from Theorem 20. In our setting the scale function s is bounded—so we have $q=r=0$, $p>0$ and we are in case (1). In particular, m exists, and $v_* \in \mathscr{L}^p$ if and only if $m=0$ and $v \in \mathscr{L}^p$. However, we have already noted that in order to be able to embed in this case we must have $m=0$, so our condition is essentially $v_* \in \mathscr{L}^p \Leftrightarrow v \in \mathscr{L}^p$. □

Acknowledgements

We would like to thank the referees, and especially the Associate Editor, for their comments on the first draft of this paper. These comments helped the authors to much improve the clarity of the presentation.

References

Azéma, J., Yor, M., 1979a. Le problème de Skorokhod: compléments à "Une solution simple au problème de Skorokhod". In: Séminaire de Probabilités, Vol. XIII (University of Strasbourg, Strasbourg, 1977/78). Springer, Berlin, pp. 625–633.

Azéma, J., Yor, M., 1979b. Une solution simple au problème de Skorokhod. In: Séminaire de Probabilités, Vol. XIII (University of Strasbourg, Strasbourg, 1977/78). Springer, Berlin, pp. 90–115.

Chacon, R.V., Walsh, J.B., 1976. One-dimensional potential embedding. In: Séminaire de Probabilités, Vol. X (Prèmiere partie, University of Strasbourg, Strasbourg, année universitaire 1974/1975), Lecture Notes in Mathematics, Vol. 511. Springer, Berlin, pp. 19–23.

Cox, A.M.G., Hobson, D.G., 2003. Skorokhod embeddings, minimality and non-centered target distributions. University of Bath Mathematics, Preprint 03/11, http://www.maths.bath.ac.uk/MATHEMATICS/preprints.html.

Dubins, L.E., 1968. On a theorem of Skorohod. Ann. Math. Statist. 39, 2094–2097.

Grandits, P., Falkner, N., 2000. Embedding in Brownian motion with drift and the Azéma–Yor construction. Stochastic Process. Appl. 85 (2), 249–254.

Hambly, B., Kersting, G., Kyprianou, A., 2003. Law of the iterated logarithm for oscillating random walks conditioned to stay non-negative. Stochastic Process. Appl. 108 (2), 327–343.

Jacka, S.D., 1988. Doob's inequalities revisited: a maximal H^1-embedding. Stochastic Process. Appl. 29 (2), 281–290.

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

Pedersen, J.L., Peskir, G., 2001. The Azéma–Yor embedding in non-singular diffusions. Stochastic Process. Appl. 96 (2), 305–312.

Perkins, E., 1986. The Cereteli–Davis solution to the H^1-embedding problem and an optimal embedding in Brownian motion. In: Seminar on stochastic processes, 1985 (Gainesville, FL., 1985). Birkhäuser, Boston, MA, pp. 172–223.

Revuz, D., Yor., M., 1999. Continuous martingales and Brownian motion. Grundlehren der Mathematischen Wissenschaften (Fundamental Principles of Mathematical Sciences), 3rd Edition, Vol. 293. Springer, Berlin.

Rogers, L.C.G., Williams, D., 2000. Diffusions, Markov processes, and martingales. Itô Calculus, Vol. 2. Cambridge University Press, Cambridge (reprint of the second 1994 edition).

Root, D.H., 1969. The existence of certain stopping times on Brownian motion. Ann. Math. Statist. 4, 715–718.

Rost, H., 1971. The stopping distributions of a Markov Process. Invent. Math. 14, 1–16.

Rost, H., 1976. Skorokhod stopping times of minimal variance. In: Séminaire de Probabilités, Vol. X (Première partie, University of Strasbourg, Strasbourg, année universitaire 1974/1975), Lecture Notes in Mathematics, Vol. 511. Springer, Berlin, pp. 194–208.

Skorokhod, A.V., 1965. Studies in the theory of random processes. Translated from the Russian by Scripta Technica, Inc. Addison-Wesley Publishing Co., Inc., Reading, MA.

Vallois, P., 1992. Quelques inégalités avec le temps local en zero du mouvement brownien. Stochastic Process. Appl. 41 (1), 117–155.