Research Report No. 23, 2010, *Probab. Statist. Group Manchester* (25 pp)

Quickest Detection of a Hidden Target and Extremal Surfaces

Goran Peskir

Let $Z = (Z_t)_{t \geq 0}$ be a regular diffusion process started at 0, let ℓ be an independent random variable with a strictly increasing and continuous distribution function F, and let $\tau_\ell = \inf\{\, t \geq 0 \mid Z_t = \ell \,\}$ be the first entry time of Z at the level ℓ. We show that the quickest detection problem

$$\inf_\tau \big[\, \mathsf{P}(\tau < \tau_\ell) + c\, \mathsf{E}(\tau - \tau_\ell)^+ \big]$$

is equivalent to the (three-dimensional) optimal stopping problem

$$\sup_\tau \mathsf{E}\Big[R_\tau - \int_0^\tau c(R_t)\, dt \Big]$$

where $R = S - I$ is the range process of $X = 2F(Z) - 1$ (i.e. the difference between the running maximum and the running minimum of X) and $c(r) = cr$ with $c > 0$. Solving the latter problem we find that the following stopping time is optimal

$$\tau_* = \inf\{\, t \geq 0 \mid f_*(I_t, S_t) \leq X_t \leq g_*(I_t, S_t) \,\}$$

where the surfaces f_* and g_* can be characterised as extremal solutions to a couple of first-order nonlinear PDEs expressed in terms of the infinitesimal characteristics of X and c. This is done by extending the arguments associated with the maximality principle [28] to the three-dimensional setting of the present problem and disclosing the general structure of the solution that is valid in all particular cases. The key arguments developed in the proof should be applicable in similar multi-dimensional settings.

1. Introduction

Imagine that you are observing a sample path $t \mapsto Z_t$ of the continuous process Z started at 0 and that you wish to detect when this sample path reaches a level ℓ that is not directly observable. Situations of this type occur naturally in many applied problems and there is a whole range of hypotheses that can be introduced to study various particular aspects of the problem. Assuming that Z and ℓ are independent, and denoting by τ_ℓ the first entry time of Z at ℓ, it was shown recently (see [32]) that the median/quantile rule minimises not only the spatial expectation $\mathsf{E}[(\ell - X_\tau)^+ + c\,(X_\tau - \ell)^+]$ (dating back to R. J. Boscovich 1711-1787) but also the *temporal* expectation $\mathsf{E}[(\tau_\ell - \tau)^+ + c(\tau - \tau_\ell)^+]$ over all stopping times τ of Z where

Mathematics Subject Classification 2010. Primary 60G35, 60G40, 60J60. Secondary 34A34, 35R35, 49J40.

Key words and phrases: Quickest detection, hidden target, optimal stopping, diffusion process, maximum process, minimum process, range process, excursion, the maximality principle, extremal surface, the principle of smooth fit, nonlinear differential equation.

c is a positive constant. Motivated by this development, and seeking for further insights and connections, in this paper we study the 'mixed' variational problem

$$\inf_\tau \big[\, \mathsf{P}(\tau < \tau_\ell) + c\, \mathsf{E}(\tau - \tau_\ell)^+ \big] \tag{1.1}$$

which appears in the classic formulation of quickest detection due to Shiryaev (see [34], [35], [33, Sections 22+24] and the references therein). The key difference between (1.1) and the classic formulation is that the unobservable time τ_ℓ in (1.1) is obtained through the uncertainty in the space domain (as the first entry time of Z at the unknown level ℓ), while the unobservable time in the classic formulation is obtained through the uncertainty in the time domain (as the unknown level itself). Unlike the classic formulation, however, we do not assume that the probabilistic characteristics of the observed process Z are being affected following τ_ℓ (quickest detection problems of this kind require a different treatment and will be studied elsewhere). Likewise, since the underlying loss processes $t \mapsto I(t < \tau_\ell)$ and $t \mapsto (t - \tau_\ell)^+$ are not adapted to the natural filtration generated by Z (or its usual augmentation) we see that the problem (1.1) belongs to the class of 'optimal prediction' problem (within optimal stopping). Similar optimal prediction problems have been studied in recent years by many authors (see e.g. [3], [4], [9]-[14], [18], [23], [27], [36]-[38]). It may be noted in this context that the non-adapted factor τ_ℓ in the optimal prediction problem (1.1) is not revealed at the 'end' of time (i.e. it is not measurable with respect to the σ-algebra generated by the process Z).

While the median/quantile rule was derived in [32] for general (continuous) processes, a closer analysis of the mixed variational problem (1.1) reveals that this generality can hardly be maintained. For this reason we restrict our attention to a smaller class of processes and assume that $Z = (Z_t)_{t \ge 0}$ is a one-dimensional diffusion starting at 0 and solving

$$dZ_t = a(Z_t)\, dt + b(Z_t)\, dB_t \tag{1.2}$$

where a and $b > 0$ are continuous functions and $B = (B_t)_{t \ge 0}$ is a standard Brownian motion. To gain tractability we also assume that the distribution function F of ℓ is strictly increasing and sufficiently smooth (although the results hold in greater generality as well). In the first step we show that the problem (1.1) is equivalent to the optimal stopping problem

$$\sup_\tau \mathsf{E}\Big[R_\tau - \int_0^\tau c(R_t)\, dt \Big] \tag{1.3}$$

where $R = S - I$ is the range process of $X = 2F(Z) - 1$ (i.e. the difference between the running maximum and the running minimum of X) and $c(r) = cr$. This problem is of independent interest and the appearance of the range process is novel in this context revealing also that the problem is fully three-dimensional. Two-dimensional versions of a related problem (when $I \equiv 0$ and c constant) were initially studied and solved in important special cases of diffusion processes in [7], [8] and [22]. The general solution to problems of this kind was derived in the form of the maximality principle in [28] (see also Section 13 and Chapter V in [33] and the other references therein). In these two-dimensional problems c was a function of X_t instead. More recent contributions and studies of related problems include [5], [15], [16], [19], [21], [24]-[26] (see also [1], [2], [20] and [29] for related results in optimal control theory). Close three-dimensional relatives of the problem (1.3) also appear in the recent papers [6] and [39] where the problems were effectively solved by guessing and finding the optimal stopping boundary in a closed form. These optimal stopping boundaries are still curves in the state space.

In this paper we show how the problem (1.3) can be solved when (i) no closed-form solution for the candidate stopping boundary is available and (ii) the optimal stopping boundaries are no longer curves in the state space. This is done by extending the arguments associated with the maximality principle [28] to the three-dimensional setting of the problem (1.3) and disclosing the general structure of the solution that is valid in all particular cases. In this way we find that that the optimal stopping boundary consists of two surfaces which can be characterised as extremal solutions to a couple of first-order nonlinear PDEs. More precisely, we show that the following stopping time is optimal

$$\tau_* = \inf\{\, t \ge 0 \mid f_*(I_t, S_t) \le X_t \le g_*(I_t, S_t)\,\} \tag{1.4}$$

where the surfaces f_* and g_* can be characterised as the minimal and maximal solutions to

$$\frac{\partial f}{\partial i}(i,s) = \frac{(\sigma^2/2)(f(i,s))\, L'(f(i,s))}{c(s-i)\,[L(f(i,s))-L(i)]}\left[1 + c'(s-i)\int_i^{f(i,s)} \frac{L(y)-L(i)}{(\sigma^2/2)(y)\, L'(y)}\, dy\right] \tag{1.5}$$

$$\frac{\partial g}{\partial s}(i,s) = \frac{(\sigma^2/2)(g(i,s))\, L'(g(i,s))}{c(s-i)\,[L(s)-L(g(i,s))]}\left[1 + c'(s-i)\int_{g(i,s)}^{s} \frac{L(s)-L(y)}{(\sigma^2/2)(y)\, L'(y)}\, dy\right] \tag{1.6}$$

staying above/below the lower/upper diagonal in the state space respectively. In these equations σ is the diffusion coefficient and L is the scale function of X. They can be expressed explicitly in terms of a, b and F. Recalling that the problems (1.1) and (1.3) are equivalent we see that this also yields the solution to the initial problem (1.1). A plain comparison with the median/quantile rule from [32] shows that the structure of the problem (1.1) is inherently more complicated and the optimal stopping time τ_* may be viewed as a nonlinear median/quantile rule. The optimal surfaces f_* and g_* combined with the excursions of X away from I and S exhibit interesting dynamics (not present in the two-dimensional setting) which we describe in fuller detail as we progress below. This dynamics can be combined with Lagrange multipliers to solve the constrained variant of the problem (1.1) where the probability error of early stopping is bounded from above (we do not pursue this in the present paper). It is also possible to see that swapping the order of τ and τ_ℓ is (1.1) leads to optimal stopping at the diagonal and thus corresponds to the linear median/quantile rule. The key arguments developed in the proof rely heavily upon the extremal properties of the optimal surfaces and should be applicable in similar multi-dimensional settings.

2. Quickest detection of a hidden target

In this section we will firstly formulate the quickest detection of a hidden target problem and then show that this problem is equivalent to an optimal stopping problem for the range process. The latter problem will be studied in the next section.

Let $Z = (Z_t)_{t \ge 0}$ be a one-dimensional diffusion process starting at 0 and solving

$$dZ_t = a(Z_t)\, dt + b(Z_t)\, dB_t \tag{2.1}$$

where a and $b > 0$ are continuous functions and $B = (B_t)_{t\ge 0}$ is a standard Brownian motion. Let ℓ be an independent random variable with values in $I\!R$ and let

$$\tau_\ell = \inf\{\, t \ge 0 \mid Z_t = \ell\,\} \tag{2.2}$$

be the first entry time of Z at the level ℓ. We consider the quickest detection problem

$$V_1 = \inf_{\tau} \big[\, \mathsf{P}(\tau < \tau_\ell) + c\, \mathsf{E}(\tau - \tau_\ell)^+ \big] \tag{2.3}$$

where the infimum is taken over all stopping times τ of Z (i.e. with respect to the natural filtration $(\mathcal{F}_t^Z)_{t \ge 0}$ generated by Z) and $c > 0$ is a given and fixed constant. Note that $\mathsf{P}(\tau < \tau_\ell)$ represents the probability of early stopping and $\mathsf{E}(\tau - \tau_\ell)^+$ represents the expectation of late stopping when a stopping time τ of Z is being applied. Our task therefore is to minimise the weighted sum of both errors over all stopping times τ of Z. Note that ℓ and τ_ℓ are not observable. Set

$$I_t^Z = \inf_{0 \le s \le t} Z_s \quad \& \quad S_t^Z = \sup_{0 \le s \le t} Z_s \tag{2.4}$$

for $t \ge 0$ and let F denote the distribution function of ℓ.

Proposition 1. *The problem* (2.3) *is equivalent to the optimal stopping problem*

$$V_2 = \sup_{\tau} \mathsf{E}\Big[F(S_\tau^Z) - F(I_\tau^Z -) - c \int_0^\tau \big[F(S_t^Z) - F(I_t^Z -) \big]\, dt \Big] \tag{2.5}$$

where the infimum is taken over all stopping times τ of Z.

Proof. Let a stopping time τ of Z be given and fixed. Firstly, using that ℓ and Z are independent, we find that

$$\begin{aligned} \mathsf{P}(\tau < \tau_\ell) &= 1 - \mathsf{P}(\tau \ge \tau_\ell) = 1 - \mathsf{P}(\tau \ge \tau_\ell\,,\, \ell > 0) - \mathsf{P}(\tau \ge \tau_\ell\,,\, \ell \le 0) \\ &= 1 - \mathsf{P}(S_\tau^Z \ge \ell > 0) - \mathsf{P}(I_\tau^Z \le \ell \le 0) = 1 - \mathsf{E} F(S_\tau^Z) + \mathsf{E} F(I_\tau^Z -) \\ &= 1 - \mathsf{E}\big[F(S_\tau^Z) - F(I_\tau^Z -) \big]\,. \end{aligned} \tag{2.6}$$

Secondly, using a well-known argument (see e.g. [33, p. 450]), it follows that

$$\begin{aligned} \mathsf{E}(\tau - \tau_\ell)^+ &= \mathsf{E} \int_0^\tau I(\tau_\ell \le t)\, dt = \mathsf{E} \int_0^\infty I(\tau_\ell \le t)\, I(t < \tau)\, dt \\ &= \int_0^\infty \mathsf{E}\big[\mathsf{E}\big(I(\tau_\ell \le t)\, I(t < \tau) \,|\, \mathcal{F}_t^Z \big) \big]\, dt \\ &= \int_0^\infty \mathsf{E}\big[I(t < \tau)\, \mathsf{E}\big(I(\tau_\ell \le t) \,|\, \mathcal{F}_t^Z \big) \big]\, dt \\ &= \mathsf{E} \int_0^\tau \mathsf{P}\big(\tau_\ell \le t \,|\, \mathcal{F}_t^Z \big)\, dt\,. \end{aligned} \tag{2.7}$$

Moreover, since ℓ and Z are independent, we see that

$$\begin{aligned} \mathsf{P}\big(\tau_\ell \le t \,|\, \mathcal{F}_t^Z \big) &= \mathsf{P}\big(\tau_\ell \le t\,,\, \ell > 0 \,|\, \mathcal{F}_t^Z \big) + \mathsf{P}\big(\tau_\ell \le t\,,\, \ell \le 0 \,|\, \mathcal{F}_t^Z \big) \\ &= \mathsf{P}(S_t^Z \ge \ell > 0 \,|\, \mathcal{F}_t^Z) + \mathsf{P}(I_t^Z \le \ell \le 0 \,|\, \mathcal{F}_t^Z) \\ &= F(S_t^Z) - F(I_t^Z -) \end{aligned} \tag{2.8}$$

for $t \geq 0$. Inserting (2.8) into (2.7) and combining it with (2.6), we find that $V_1 = 1 - V_2$ for any $c > 0$ and this completes the proof. $\square$

It follows from the previous proof that a stopping time τ of Z is optimal in (2.3) if and only if it is optimal in (2.5). To gain tractability when solving the optimal stopping problem (2.5) we will assume that the distribution function F of ℓ is strictly increasing and twice continuously differentiable (although the results derived below hold in greater generality as well). Then $F(Z)$ defines a regular diffusion process with values in $(0,1)$ and to gain symmetry and extend the state space to $(-1,1)$ we will rescale Z differently by setting

$$X = 2F(Z) - 1 . \tag{2.9}$$

Then X is a regular diffusion process starting at $2F(0)-1$ and solving

$$dX_t = \mu(X_t)\, dt + \sigma(X_t)\, dB_t \tag{2.10}$$

where the drift μ and the diffusion coefficient σ are given by

$$\mu(x) = \big(2aF' + b^2 F''\big)\big(F^{-1}\big(\tfrac{x+1}{2}\big)\big) \tag{2.11}$$

$$\sigma(x) = \big(2bF'\big)\big(F^{-1}\big(\tfrac{x+1}{2}\big)\big) \tag{2.12}$$

for $x \in (-1,1)$ as is easily verified by Itô's formula. Setting

$$I_t = \inf_{0 \leq s \leq t} X_s \quad \& \quad S_t = \sup_{0 \leq s \leq t} X_s \tag{2.13}$$

for $t \geq 0$ we see that the problem (2.5) is equivalent to the optimal stopping problem

$$V = \sup_{\tau} \mathsf{E}\Big[S_\tau - I_\tau - c \int_0^\tau \big(S_t - I_t\big)\, dt \Big] \tag{2.14}$$

where the infimum is taken over all stopping times τ of X . Note that $V = 2V_2 = 2(1-V_1)$ and there is a simple one-to-one correspondence between the optimal stopping times in (2.14) and (2.5) due to (2.9). We will therefore proceed by studying the problem (2.14).

For future reference let us note that the infinitesimal generator of X equals

$$\mathbb{L}_X = \mu(x)\, \frac{\partial}{\partial x} + \frac{\sigma^2(x)}{2}\, \frac{\partial^2}{\partial x^2} \tag{2.15}$$

and the scale function of X is given by

$$L(x) = \int_0^x \exp\Big(-\int_0^y \frac{\mu(z)}{(\sigma^2/2)(z)}\, dz \Big)\, dy \tag{2.16}$$

for $x \in (-1,1)$. Throughout we denote $\tau_a = \inf\{\, t \geq 0 \mid X_t = a \,\}$ and set $\tau_{a,b} = \tau_a \wedge \tau_b$ for $a < b$ in $(-1,1)$. Denoting by P_x the probability measure under which the process X starts at x , it is well known that

$$\mathsf{P}_x\big(X_{\tau_{a,b}} = a\big) = \frac{L(b) - L(x)}{L(b) - L(a)} \quad \& \quad \mathsf{P}_x\big(X_{\tau_{a,b}} = b\big) = \frac{L(x) - L(a)}{L(b) - L(a)} \tag{2.17}$$

for $a \le x \le b$ in $(-1,1)$. The speed measure of X is given by

$$m(dx) = \frac{dx}{L'(x)(\sigma^2/2)(x)} \tag{2.18}$$

and the Green function of X is given by

$$\begin{aligned} G_{a,b}(x,y) &= \frac{(L(b)-L(y))\,(L(x)-L(a))}{L(b)-L(a)} \quad \text{if} \quad a \le x \le y \le b \\ &= \frac{(L(b)-L(x))\,(L(y)-L(a))}{L(b)-L(a)} \quad \text{if} \quad a \le y \le x \le b\,. \end{aligned} \tag{2.19}$$

If $f : (-1,1) \to I\!\!R$ is a measurable function, then it is well known that

$$\mathsf{E}_x \int_0^{\tau_{a,b}} f(X_t)\,dt = \int_a^b f(y)\,G_{a,b}(x,y)\,m(dy) \tag{2.20}$$

for $a \le x \le b$ in $(-1,1)$. This identity holds in the sense that if one of the integrals exists so does the other one and they are equal.

3. Optimal stopping of the range process

It was shown in the previous section that the quickest detection problem (2.3) is equivalent to the optimal stopping problem (2.14). The purpose of this section is to present the solution to the latter problem. Using the fact that the two problems are equivalent this also leads to the solution of the former problem.

Let $X = (X_t)_{t \ge 0}$ be a one-dimensional diffusion process solving

$$dX_t = \mu(X_t)\,dt + \sigma(X_t)\,dB_t \tag{3.1}$$

where the drift μ and the diffusion coefficient $\sigma > 0$ are continuous functions and $B = (B_t)_{t \ge 0}$ is a standard Brownian motion. We will assume that the state space of X equals $(-1,1)$ as in the previous section, however, this hypothesis is not essential and the results below hold in greater generaility as well. By P_x we denote the probability measure under which X starts at $x \in (-1,1)$. For $i \le x \le s$ in $(-1,1)$ we set

$$I_t = i \wedge \inf_{0 \le s \le t} X_s \quad \& \quad S_t = s \vee \sup_{0 \le s \le t} X_s \tag{3.2}$$

for $t \ge 0$. These transformations enable the three-dimensional Markov process (I,X,S) to start at (i,x,s) under P_x and we will denote the resulting probability measure on the canonical space by $\mathsf{P}_{i,x,s}$. Thus under $\mathsf{P}_{i,x,s}$ the canonical process (I,X,S) starts at (i,x,s). The range process R of X is defined by

$$R_t = S_t - I_t \tag{3.3}$$

for $t \ge 0$. In this section we consider the optimal stopping problem

$$V(i,x,s) = \sup_\tau \mathsf{E}_{i,x,s}\Big[R_\tau - \int_0^\tau c(R_t)\,dt\Big] \tag{3.4}$$

for $i \le x \le s$ in $(-1,1)$ where the supremum is taken over all stopping times τ of X and the function c is increasing and continuous. To gain tractability we will also assume that c is continuously differentiable and we will see in the proof below that the case of $c(r) = c_2(s) - c_1(i)$ for $r = s - i$ where c_1 and c_2 are increasing and continuously differentiable functions leads to a closed-form expression for the value function in a joint continuation region. For any i given and fixed we will refer to $d_i = \{\, (x,s) \mid x = s \ge i \,\}$ as the *upper diagonal* in the state space, and for any s given and fixed we will refer to $d^s = \{\, (i,x) \mid i = x \le s \,\}$ as the *lower diagonal* in the state space. The main result of the paper may now be stated as follows.

Theorem 1. *The optimal stopping time in the problem* (3.4) *is given by*

$$\tau_* = \inf \{\, t \ge 0 \mid f_*(I_t, S_t) \le X_t \le g_*(I_t, S_t) \,\} \tag{3.5}$$

where the surfaces f_* *and* g_* *can be characterised as the minimal and maximal solutions to*

$$\frac{\partial f}{\partial i}(i,s) = \frac{(\sigma^2/2)(f(i,s))\, L'(f(i,s))}{c(s-i)\, [L(f(i,s)) - L(i)]} \left[1 + c'(s-i) \int_i^{f(i,s)} \frac{L(y) - L(i)}{(\sigma^2/2)(y)\, L'(y)}\, dy \right] \tag{3.6}$$

$$\frac{\partial g}{\partial s}(i,s) = \frac{(\sigma^2/2)(g(i,s))\, L'(g(i,s))}{c(s-i)\, [L(s) - L(g(i,s))]} \left[1 + c'(s-i) \int_{g(i,s)}^{s} \frac{L(s) - L(y)}{(\sigma^2/2)(y)\, L'(y)}\, dy \right] \tag{3.7}$$

staying above the lower diagonal d^s *and below the upper diagonal* d_i *respectively. This is true whenever the minimal and maximal solutions exist and if there is no minimal or maximal solution then there is no optimal stopping time (in the sense that it is optimal to continue forever). Explicit expressions for the value function* V *and further properties of the optimal surfaces* f_* *and* g_* *are exhibited in the proof below.*

Proof. The optimal stopping problem (3.4) is three-dimensional and the underlying Markov process equals (I,X,S). It is evident from the structure of the gain function in (3.4) that the excursions of X away from the running maximum S and the running minimum I play a key role in the analysis of the problem. A possible way to visualise the dynamics of these excursions is illustrated in Figure 1. Each excursion of X at an upper level s is mirror imaged with the excursion of X at a lower level i and vice versa. When the excursion returns to the upper diagonal, the process (X,S) receives an infinitesimal push upwards along the upper diagonal, and when the excursion returns to the lower diagonal, the process (I,X) receives an infinitesimal push downwards along the lower diagonal.

An important initial observation is that the process (I,X,S) can never be optimally stopped at the upper or lower diagonal. The analogous phenomenon is known to hold for optimal stopping of the maximum process (see [28, Proposition 2.1]) and the same arguments extend to the present case without major changes. Before we formalise this in the first step below let us recall that general theory of optimal stopping for Markov processes (see [33, Chapter 1]) implies that the continuation set in the problem (3.4) equals $C = \{\, (i,x,s) \mid V(i,x,s) > s - i \,\}$ and the stopping set equals $D = \{\, (i,x,s) \mid V(i,x,s) = s - i \,\}$. It means that the first entry time of (I,X,S) into D is optimal in the problem (3.4). To determine the sets C and D we will begin by formalising the initial observation above.

1. *The upper and lower diagonal* d_i *and* d^s *are always contained in* C. For this, take any $(s,s) \in d_i$ and consider $\tau_{l_n,r_n} = \inf \{\, t \ge 0 \mid X_t \notin (l_n,r_n) \,\}$ under $\mathsf{P}_{i,s,s}$ with $l_n = s - 1/n$

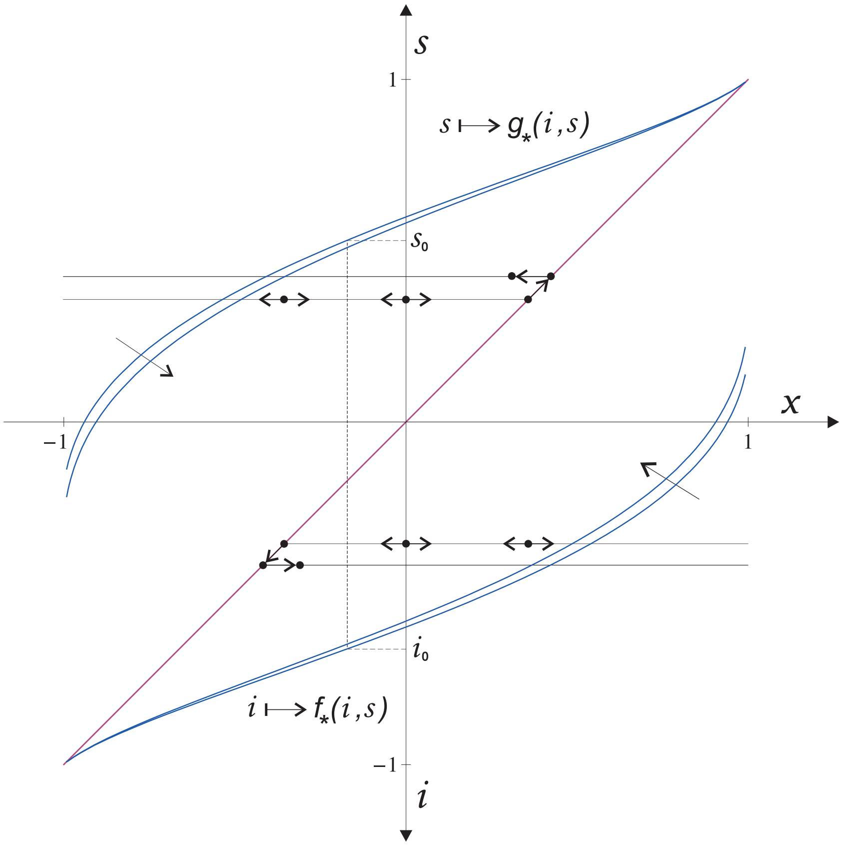

Figure 1. Excursions of X away from the running minimum I and the running maximum S combined with the dynamics of the optimal stopping surfaces f_* and g_* : (i) Return of X to the lower diagonal causes I to go down and forces g_* to go up; (ii) Return of X to the upper diagonal causes S to go up and forces f_* to go down; (iii) Even if X goes above f_* it may not be optimal to stop unless X is below g_* ; (iv) Even if X goes below g_* it may not be optimal to stop unless X is above f_* . The (movable) dotted vertical line marks the borderline levels i_0 and s_0 below and above which it is optimal to stop.

and $r_n = s+1/n$ for $n \geq 1$. Then (2.18)-(2.20) imply that $\mathsf{E}_{i,s,s} R_{\tau_{l_n,r_n}} \geq s-i+K/n$ and $\mathsf{E}_{i,s,s} \int_0^{\tau_{l_n,r_n}} c(R_t)\, dt \leq K/n^2$ for all $n \geq 1$ with some positive constant K (see the proof of Proposition 2.1 in [28] for details). Taking $n \geq 1$ large enough (to exploit the difference in the rates of the bounds) we see that (i,s,s) belongs to C . In exactly the same way one sees that if $(i,i) \in d^s$ then (i,i,s) belongs to C . This establishes the initial claim.

2. *Optimal stopping surfaces.* Assume now that the process (I,X,S) starts at (i,x,s) and consider the excursion of X away from the running maximum s with i given and fixed. In view of the fact that it is never optimal to stop at the upper diagonal d_i , and due to the existence of a strictly positive cost which is proportional to the duration of time in (3.4), we see that it is plausible to expect that there exists a point $g(i,s)$ (depending on both i and s) at/below which the process X should be stopped (should i remain constant). In exactly the same way, if we consider the excursion of X away from the running minimum i with s given and fixed, we see that it is plausible to expect that there exists a point $f(i,s)$ (depending on

both i and s) at/above which the process X should be stopped (should s remain constant).

The first complication in this reasoning comes from the fact that neither i nor s need to remain constant during the excursion of X away from the running maximum s or the running minimum i respectively. We will handle this difficulty implicitly by noting that if I is to decrease from i downwards then this will increase the rate of the cost in (3.4) which in turn will move the boundary point $g(i,s)$ upwards (it means that $i \mapsto g(i,s)$ is decreasing), and similarly if S is to increase from s upwards then this will increase the rate of the cost in (3.4) which in turn will move the boundary point $f(i,s)$ downwards (it means that $s \mapsto f(i,s)$ is decreasing), see Figure 1. Changes in either I or S therefore contribute to resetting i and s to new levels and starting from there afresh with the boundary points $f(i,s)$ and $g(i,s)$ adjusted. For these reasons it is not entirely surprising that the first complication will resolve itself after we describe the structure of the optimal surfaces f and g in fuller detail below.

The second complication comes from the fact that even if X is at/below $g(i,s)$ and normally (when i would not change) it would be optimal to stop, it may be that X is still below $f(i,s)$ and therefore the proximity of the lower diagonal d^s may be a valid incentive to continue. This incentive itself is further complicated by the fact that it may lead to a decrease of i and therefore the rate of the cost in (3.4) will also increase (as addressed in the first complication above). Likewise, even if X is at/above $f(i,s)$ and normally (when s would not change) it would be optimal to stop, it may be that X is still above $g(i,s)$ and therefore the proximity of the upper diagonal d_i may be a valid incentive to continue. This incentive itself is further complicated by the fact that it may lead to an increase of s and therefore the rate of the cost in (3.4) will also increase (as addressed in the first complication above).

Neither of these complications appear in the optimal stopping of the maximum process where g depends only on s (see [28] and the references therein) and our strategy in tackling the problem will be to extend the maximality principle [28] from the two-dimensional setting of the process (X,S) and the optimal stopping curves to the three-dimensional setting of the process (I,X,S) and the optimal stopping surfaces. This will enable us to resolve the second complication using the existence of the so-called 'bad-good' solutions (those hitting the upper or lower diagonal) which in turn will provide novel insights into the maximality/minimality principle in the three dimensions as will be seen below.

3. *Free-boundary problem.* Previous considerations suggest to seek the solution to (3.4) as the following stopping time

$$\tau_{f,g} = \inf\{\, t \ge 0 \mid f(I_t,S_t) \le X_t \le g(I_t,S_t)\,\} \tag{3.8}$$

where the surfaces f and g are to be found. The continuation set $C_{f,g}$ splits into

$$C^0_{f,g} = \{\, (i,x,s) \mid f(i,s) > g(i,s)\,\} \tag{3.9}$$

$$C^-_{f,g} = \{\, (i,x,s) \mid i \le x < f(i,s) \le g(i,s)\,\} \tag{3.10}$$

$$C^+_{f,g} = \{\, (i,x,s) \mid f(i,s) \le g(i,s) < x \le s\,\} \tag{3.11}$$

and we have $C_{f,g} = C^0_{f,g} \cup C^-_{f,g} \cup C^+_{f,g}$. To compute the value function V and determine the optimal surfaces f and g , we are led to formulate the free-boundary problem

$$(I\!\!L_X V)(i,x,s) = c(s-i) \quad \text{for } (i,x,s) \in C_{f,g} \tag{3.12}$$

$$V_i'(i,x,s)\Big|_{x=i+} = 0 \quad \text{(normal reflection)} \tag{3.13}$$

$$V_s'(i,x,s)\Big|_{x=s-} = 0 \quad \text{(normal reflection)} \tag{3.14}$$

$$V(i,x,s)\Big|_{x=f(i,s)-} = s-i \quad \text{for } f(i,s) \le g(i,s) \tag{3.15}$$

$$V(i,x,s)\Big|_{x=g(i,s)+} = s-i \quad \text{for } f(i,s) \le g(i,s) \tag{3.16}$$

$$V_x'(i,x,s)\Big|_{x=f(i,s)-} = 0 \quad \text{for } f(i,s) \le g(i,s) \text{ (smooth fit)} \tag{3.17}$$

$$V_x'(i,x,s)\Big|_{x=g(i,s)+} = 0 \quad \text{for } f(i,s) \le g(i,s) \text{ (smooth fit)} \tag{3.18}$$

where $I\!\!L_X$ is the infinitesimal generator of X given in (2.15) above. For the rationale and further details regarding free-boundary problems of this kind we refer to [33, Section 13] and the references therein (we note in addition that the conditions of normal reflection (3.13) and (3.14) date back to [17]).

4. *Nonlinear differential equations.* To solve the free-boundary problem (3.12)-(3.18) consider the resulting function

$$V_{f,g}(i,x,s) = \mathsf{E}_{i,x,s}\Big[R_{\tau_{f,g}} - \int_0^{\tau_{f,g}} c(R_t)\,dt\Big] \tag{3.19}$$

for $i \le x \le s$ in $(-1,1)$. Suppose that $f(i,s) \le s$ and consider $\tau_{i,f(i,s)} = \inf\{\, t \ge 0 \mid X_t \notin (i, f(i,s))\,\}$. Applying the strong Markov property of (I,X,S) at $\tau_{i,f(i,s)}$ and using (2.17)-(2.20) we find that

$$\begin{aligned} V_{f,g}(i,x,s) = (s-i)\,\frac{L(x)-L(i)}{L(f(i,s))-L(i)} + V_{f,g}(i,i,s)\,\frac{L(f(i,s))-L(x)}{L(f(i,s))-L(i)} \\ - c(s-i)\int_i^{f(i,s)} G_{i,f(i,s)}(x,y)\,m(dy)\,. \end{aligned} \tag{3.20}$$

It follows from (3.20) that

$$\begin{aligned} V_{f,g}(i,i,s) = s-i + \frac{L(f(i,s))-L(i)}{L(f(i,s))-L(x)}\Big[V_{f,g}(i,x,s) - (s-i) \\ + c(s-i)\int_i^{f(i,s)} G_{i,f(i,s)}(x,y)\,m(dy)\Big]\,. \end{aligned} \tag{3.21}$$

Dividing and multiplying through by $x-f(i,s)$ we find using (3.17) that

$$\lim_{x\uparrow f(i,s)} \frac{V_{f,g}(i,x,s)-(s-i)}{L(f(i,s))-L(x)} = -\frac{1}{L'(f(i,s))}\,\frac{\partial V_{f,g}}{\partial x}(i,x,s)\Big|_{x=f(i,s)-} = 0 \tag{3.22}$$

for $f(i,s) \le g(i,s)$. It is easily seen by (2.19) that

$$\lim_{x\uparrow f(i,s)} \frac{L(f(i,s))-L(i)}{L(f(i,s))-L(x)} \int_i^{f(i,s)} G_{i,f(i,s)}(x,y)\,m(dy) = \int_i^{f(i,s)} [L(y)-L(i)]\,m(dy)\,. \tag{3.23}$$

Combining (3.21)-(3.23) we find that

$$V_{f,g}(i,i,s) = s-i + c(s-i)\int_i^{f(i,s)} [L(y)-L(i)]\,m(dy) \tag{3.24}$$

for $f(i,s) \le g(i,s)$. Inserting this back into (3.20) and using (2.19)+(2.20) we conclude that

$$V_{f,g}(i,x,s) = s-i + c(s-i)\int_x^{f(i,s)} [L(y)-L(x)]\, m(dy) \tag{3.25}$$

for $x \le f(i,s) \le g(i,s)$. Finally, using (3.13) we find that

$$\frac{\partial f}{\partial i}(i,s) = \frac{(\sigma^2/2)(f(i,s))\, L'(f(i,s))}{c(s-i)\, [L(f(i,s))-L(i)]} \left[1 + c'(s-i)\int_i^{f(i,s)} [L(y)-L(i)]\, m(dy) \right] \tag{3.26}$$

for $f(i,s) \le g(i,s)$. By (2.18) we see that (3.26) coincides with (3.6) above. Similarly, suppose that $g(i,s) \ge i$ and consider $\tau_{g(i,s),s} = \inf\{\, t \ge 0 \mid X_t \notin (g(i,s),s)\,\}$. Applying the strong Markov property of (I,X,S) at $\tau_{g(i,s),s}$ and using (2.17)-(2.20) we find that

$$\begin{aligned} V_{f,g}(i,x,s) = (s-i)\, \frac{L(s)-L(x)}{L(s)-L(g(i,s))} + V_{f,g}(i,s,s)\, \frac{L(x)-L(g(i,s))}{L(s)-L(g(i,s))} \\ - c(s-i)\int_{g(i,s)}^{s} G_{g(i,s),s}(x,y)\, m(dy)\, . \end{aligned} \tag{3.27}$$

It follows from (3.27) that

$$\begin{aligned} V_{f,g}(i,s,s) = s-i + \frac{L(s)-L(g(i,s))}{L(x)-L(g(i,s))} \Big[V_{f,g}(i,x,s) - (s-i) \\ + c(s-i)\int_{g(i,s)}^{s} G_{g(i,s),s}(x,y)\, m(dy) \Big]\, . \end{aligned} \tag{3.28}$$

Dividing and multiplying through by $x-g(i,s)$ we find using (3.18) that

$$\lim_{x \downarrow g(i,s)} \frac{V_{f,g}(i,x,s)-(s-i)}{L(x)-L(g(i,s))} = \frac{1}{L'(g(i,s))}\, \frac{\partial V_{f,g}}{\partial x}(i,x,s)\Big|_{x=g(i,s)+} = 0 \tag{3.29}$$

for $g(i,s) \ge f(i,s)$. It is easily seen by (2.19) that

$$\lim_{x \downarrow g(i,s)} \frac{L(s)-L(g(i,s))}{L(x)-L(g(i,s))} \int_{g(i,s)}^{s} G_{g(i,s),s}(x,y)\, m(dy) = \int_{g(i,s)}^{s} [L(s)-L(y)]\, m(dy)\, . \tag{3.30}$$

Combining (3.28)-(3.30) we find that

$$V_{f,g}(i,s,s) = s-i + c(s-i)\int_{g(i,s)}^{s} [L(s)-L(y)]\, m(dy) \tag{3.31}$$

for $g(i,s) \ge f(i,s)$. Inserting this back into (3.27) and using (2.19)+(2.20) we conclude that

$$V_{f,g}(i,x,s) = s-i + c(s-i)\int_{g(i,s)}^{x} [L(x)-L(y)]\, m(dy) \tag{3.32}$$

for $x \ge g(i,s) \ge f(i,s)$. Finally, using (3.14) we find that

$$\frac{\partial g}{\partial s}(i,s) = \frac{(\sigma^2/2)(g(i,s))\, L'(g(i,s))}{c(s-i)\, [L(s)-L(g(i,s))]} \left[1 + c'(s-i)\int_{g(i,s)}^{s} [L(s)-L(y)]\, m(dy) \right] \tag{3.33}$$

for $g(i,s) \ge f(i,s)$. By (2.18) we see that (3.33) coincides with (3.7) above.

Summarising the preceding considerations we see that to each pair of surfaces f and g solving (3.6) and (3.7) there corresponds the function (3.25)+(3.32) solving the free-boundary problem (3.12)-(3.18) and admitting the probabilistic representation (3.19) associated with the stopping time (3.8) when the latter has finite expectation. The central question becomes how to select the optimal surfaces f and g among all admissible candidates solving (3.6) and (3.7). The answer is to invoke the superharmonic characterisation of the value function (see [33, Chapter 1]) for the four-dimensional Markov process (I,X,S,A) where $A_t = \int_0^t c(R_s)\,ds$ for $t \ge 0$. Fuller details of this argument will become clearer as we progress below.

5. *The minimal and maximal solution.* Motivated by the previous question we note from (3.25) and (3.32) that $f \mapsto V_{f,g}$ is increasing and $g \mapsto V_{f,g}$ is decreasing. Recalling also that it is not optimal to stop at the upper or lower diagonal, this motivates us to select solutions to (3.6) and (3.7) as far as possible from the upper and lower diagonal respectively (respecting also the meaning of (3.8) in (3.19) as well as the meaning of (3.19) itself). In the former case this means as small as possible below the upper diagonal, and in the latter case it means as large as possible above the lower diagonal.

To address the existence and uniqueness of solutions to these equations, denote the right-hand side of (3.6) by $\Phi(i,s,f(i,s))$ and denote the right-hand side of (3.7) by $\Psi(i,s,g(i,s))$. From general theory of nonlinear differential equations we know that if the direction fields $(i,f) \mapsto \Phi(i,s,f)$ and $(s,g) \mapsto \Psi(i,s,g)$ are (locally) continuous and (locally) Lipschitz in the second variable, then the equations (3.6) and (3.7) admit (locally) unique solutions. For instance, this will be the case if along a (local) continuity of $(i,f) \mapsto \Phi(i,s,f)$ and $(s,g) \mapsto \Psi(i,s,g)$ we also have a (local) continuity of $(i,f) \mapsto \Phi'_f(i,s,f)$ and $(s,g) \mapsto \Psi'_g(i,s,g)$. In particular, we see from the structure of Φ and Ψ that the equations (3.6) and (3.7) admit (locally) unique solutions whenever $x \mapsto \sigma^2(x)$ is (locally) continuously differentiable.

To construct the minimal solution to (3.6) staying above the lower diagonal d^s we can proceed as follows (see Figure 2). For any $i_n \in (-1,1)$ such that $i_n \downarrow -1$ as $n \to \infty$ let $i \mapsto f_n(i,s)$ denote the solution to (3.6) such that $f_n(i_n,s) = i_n$ for $n \ge 1$. Note that each solution $i \mapsto f(i,s)$ to (3.6) is singular at the lower diagonal d^s in the sense that $f'_i(i+,s) = +\infty$ for $f(i+,s) = i$, however, passing to the equivalent equation for the inverse of $i \mapsto f(i,s)$ (upon noting that each solution $i \mapsto f(i,s)$ to (3.6) is strictly increasing) we see that this singularity gets removed (note that the inverse of $i \mapsto f(i,s)$ has the derivative equal to zero at the lower diagonal d^s). By the uniqueness of the solution we know that the two curves $i \mapsto f_n(i,s)$ and $i \mapsto f_m(i,s)$ cannot intersect for $n \ne m$ and hence we see that $(f_n)_{n\ge1}$ is increasing. It follows therefore that $f_* := \lim_{n\to\infty} f_n$ exists. Passing to an integral equation equivalent to (3.6) (or its inverse) it is easily verified that $i \mapsto f_*(i,s)$ solves (3.6) whenever strictly larger than -1 . This f_* represents the minimal solution to (3.6) staying above the lower diagonal (for simplicity we will use the same symbol f below to denote it).

To construct the maximal solution to (3.7) staying below the upper diagonal d_i we can proceed similarly (see Figure 2). For any $s_n \in (-1,1)$ such that $s_n \uparrow 1$ as $n \to \infty$ let $s \mapsto g_n(i,s)$ denote the solution to (3.7) such that $g_n(s_n,i) = s_n$ for $n \ge 1$. Note that each solution $s \mapsto g(i,s)$ to (3.7) is singular at the upper diagonal d_i in the sense that $g'_s(i,s-) = +\infty$ for $g(i,s-) = s$, however, passing to the equivalent equation for the inverse of $s \mapsto g(i,s)$ (upon noting that each solution $s \mapsto g(i,s)$ to (3.7) is strictly increasing) we

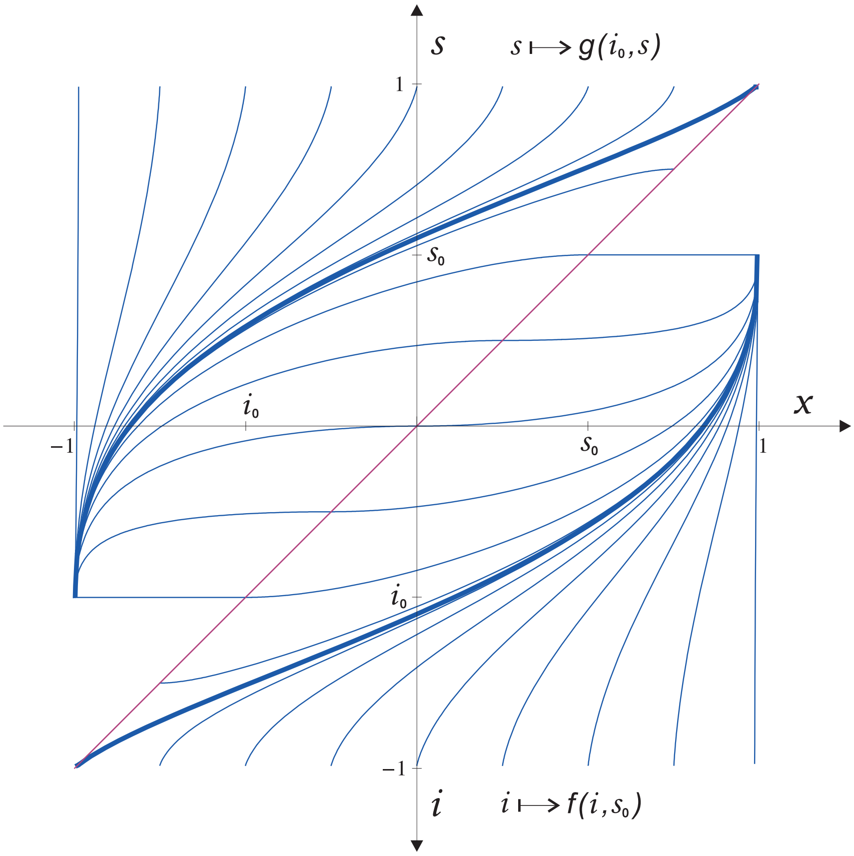

Figure 2. Smooth-fit solutions $i \mapsto f(i,s_0)$ and $s \mapsto g(i_0,s)$ to differential equations (3.6) and (3.7) for fixed s_0 and i_0 respectively. The minimal solution staying above the lower diagonal (bold f line) and the maximal solution staying below the upper diagonal (bold g line) are sections of the optimal stopping surfaces respectively.

see that this singularity gets removed (note that the inverse of $s \mapsto g(i,s)$ has the derivative equal to zero at the upper diagonal d_i). By the uniqueness of the solution we know that the two curves $s \mapsto g_n(i,s)$ and $s \mapsto g_m(i,s)$ cannot intersect for $n \neq m$ and hence we see that $(g_n)_{n\geq 1}$ is decreasing. It follows therefore that $g_* := \lim_{n\to\infty} g_n$ exists. Passing to an integral equation equivalent to (3.7) (or its inverse) it is easily verified that $s \mapsto g_*(i,s)$ solves (3.7) whenever strictly smaller than 1 . This g_* represents the maximal solution to (3.7) staying below the upper diagonal (for simplicity we will use the same symbol g below to denote it).

With the minimal and maximal solution f and g we can associate the stopping time (3.8) and the resulting function (3.19). Doing the same thing with f_n and g_n (noting that the stopping time (3.8) has finite expectation) the arguments above show that (3.25) and (3.32) hold for f_n and g_n for $n \geq 1$. Passing in these expressions to the limit as $n \to \infty$ we see that (3.25) and (3.32) remain valid for the minimal and maximal solution f and g . This establishes closed-form expressions for $V_{f,g}$ in terms of f and g on $C^+_{f,g}$ and $C^-_{f,g}$.

6. *Computing* $V_{f,g}$ *on* $C^0_{f,g}$. This calculation is technically more complicated and we will derive closed-form expressions for $V_{f,g}$ in terms of f and g on $C^0_{f,g}$ when $c(r) = c_2(s) - c_1(i)$ for $r = s - i$ where c_1 and c_2 are increasing and continuously differentiable functions. Note

that the latter decomposition is fulfilled in the setting in Section 2 above. Note also that these closed-form expressions are not needed to derive the optimality of f and g as it will be shown in the rest of the proof below.

We begin by noting that $V_{f,g}$ satisfies (3.12)-(3.14) on $C^0_{f,g}$. Recalling that a particular solution to $I\!\!L_X H = 1$ is given by

$$H(x) = \int_0^x [L(x) - L(y)]\, m(dy) \tag{3.34}$$

it follows from (3.12) that

$$V(i,x,s) = A(i,s)\, L(x) + B(i,s) + (c_2(s) - c_1(i))\, H(x) \tag{3.35}$$

for some unknown functions A and B to be found. By (3.13) and (3.14) we find that

$$A'_i(i,s)\, L(i) + B'_i(i,s) - c'_1(i)\, H(i) = 0 \tag{3.36}$$

$$A'_s(i,s)\, L(s) + B'_s(i,s) + c'_2(s)\, H(s) = 0\, . \tag{3.37}$$

Differentiating (3.36) with respect to s and (3.37) with respect to i (upon assuming that A and B are twice continuously differentiable) it follows by subtracting the resulting identities that $A''_{is}(i,s) = 0$ and hence $B''_{is}(i,s) = 0$ too. This implies that

$$A(i,s) = a_1(i) + a_2(s) \quad \& \quad B(i,s) = b_1(i) + b_2(s) \tag{3.38}$$

for some a_i and b_i to be found when $i = 1,2$. Inserting this back into (3.35)-(3.37) we obtain

$$V(i,x,s) = (a_1(i) + a_2(s))\, L(x) + b_1(i) + b_2(s) + (c_2(s) - c_1(i))\, H(x) \tag{3.39}$$

$$a'_1(i)\, L(i) + b'_1(i) - c'_1(i)\, H(i) = 0 \tag{3.40}$$

$$a'_2(s)\, L(s) + b'_2(s) + c'_2(s)\, H(s) = 0 \tag{3.41}$$

for $f(i,s) > g(i,s)$.

To determine a_i and b_i for $i = 1,2$ recall that $V_{f,g}$ is known at $C^-_{f,g}$ and $C^+_{f,g}$ so that it is also known at the boundary between $C^0_{f,g}$ and $C^-_{f,g}$ and the boundary between $C^0_{f,g}$ and $C^+_{f,g}$. This serves as a basic motivation for the introduction of the following functions. Given (i,s) such that $f(i,s) > g(i,s)$ there exist unique $i(s) < i$ and $s(i) > s$ such that

$$f(i(s),s) = g(i(s),s) \quad \& \quad f(i,s(i)) = g(i,s(i))\, . \tag{3.42}$$

The existence of $i(s)$ and $s(i)$ follows from the facts that $i \mapsto f(i,s)$ and $s \mapsto g(i,s)$ are strictly increasing and $s \mapsto f(i,s)$ and $i \mapsto g(i,s)$ are strictly decreasing (see Figure 3). More formally, the functions can be defined as follows

$$i(s) = (f(\cdot,s) - g(\cdot,s))^{-1}(0) \quad \& \quad s(i) = (f(i,\cdot) - g(i,\cdot))^{-1}(0) \tag{3.43}$$

for $f(i,s) > g(i,s)$. (Recall by construction that $f(-1+,s) = -1$ and $g(-1+,s) < 1$ as well as that $f(i,1-) > -1$ and $g(i,1-) = 1$ for $-1 < i < s < 1$.) Geometrically, moving from i down to $i(s)$ (with s fixed) corresponds to moving along the first coordinate from any (i,x,s)

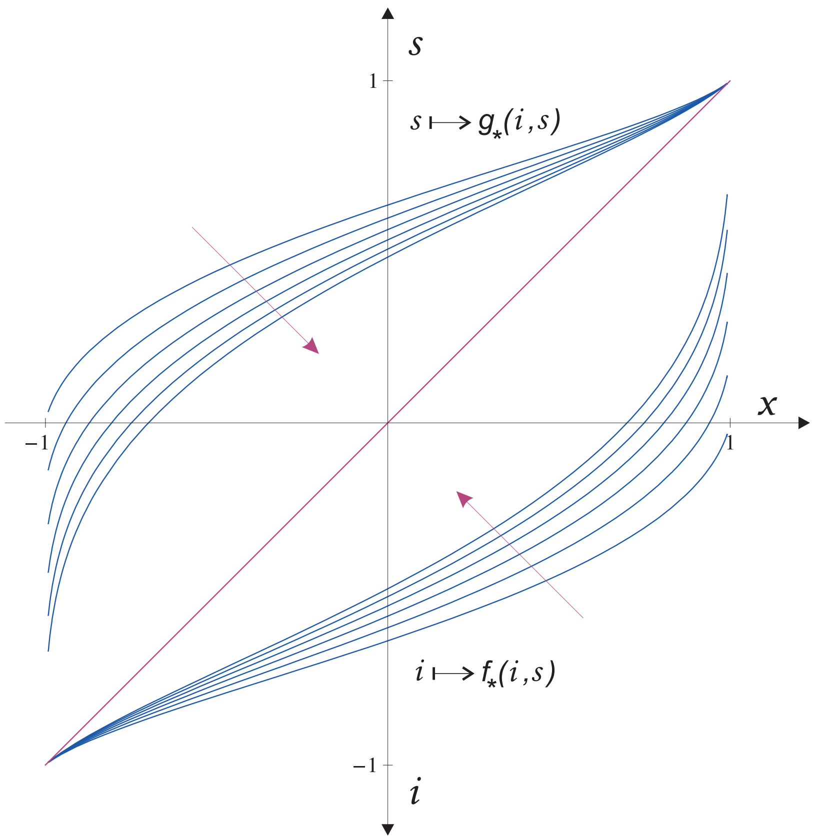

Figure 3. Movement and shape of sections $i \mapsto f_*(i,s)$ and $s \mapsto g_*(i,s)$ of the optimal surfaces f_* and g_* as the running maximum s increases and the running minimum i decreases respectively.

in $C^0_{f,g}$ to the closest point at the boundary between $C^0_{f,g}$ and $C^-_{f,g}$ if $x \le f(i(s),s)$ and to the closest point at the boundary between $C^0_{f,g}$ and $C^+_{f,g}$ if $x \ge f(i(s),s)$. Similarly, moving from s up to $s(i)$ (with i fixed) corresponds to moving along the third coordinate from any (i,x,s) in $C^0_{f,g}$ to the closest point at the boundary between $C^0_{f,g}$ and $C^+_{f,g}$ if $x \ge g(i,s(i))$ and to the closest point at the boundary between $C^0_{f,g}$ and $C^-_{f,g}$ if $x \le g(i,s(i))$.

Since $(i(s),x,s)$ with $x \le f(i(s),s)$ belongs to the boundary of $C^-_{f,g}$ we know that $V_{f,g}(i(s),x,s)$ is given by (3.25) above. Writing the integral from x to $f(i(s),s)$ in this expression as the integral from 0 to $f(i(s),s)$ minus the integral from 0 to x , it is easily seen that (3.25) reads as follows

$$(3.44)\qquad V(i(s),x,s) = s - i(s) + \big[c_2(s) - c_1(i(s))\big] \times \Big[H(x) - L(x) \int_0^{f(i(s),s)} m(dy) + \int_0^{f(i(s),s)} L(y)\, m(dy)\Big]$$

for $x \le f(i(s),s)$. Comparing (3.44) with (3.39) we can conclude that

$$(3.45)\qquad a_1(i(s)) + a_2(s) = -\big[c_2(s) - c_1(i(s))\big] \int_0^{f(i(s),s)} m(dy)$$

$$b_1(i(s)) + b_2(s) = s - i(s) + \big[c_2(s) - c_1(i(s))\big] \int_0^{f(i(s),s)} L(y)\, m(dy)\,. \tag{3.46}$$

Using (3.40)+(3.41) and (3.45)+(3.46) we can calculate $a_2'(s)$. Firstly, by (3.41) we can express $a_2'(s)$ in terms of $b_2'(s)$. Secondly, by (3.46) we can express $b_2'(s)$ in terms of $b_1'(i(s))$. Thirdly, by (3.40) we can express $b_1'(i(s))$ in terms of $a_1'(i(s))$. Fourthly, by (3.45) we can express $a_1'(i(s))$ in terms of $a_2'(s)$. This closes the loop and gives an equation for $a_2'(s)$. A lengthy calculation following these steps and making use of (3.6) above yields

$$\begin{aligned} a_2'(s) = -\frac{1}{L(s) - L(i(s))} \Bigg[\frac{f_s'(i(s),s)}{f_i'(i(s),s)} \Big[1 + c_1'(i(s)) \int_{i(s)}^{f(i(s),s)} [L(y) - L(i(s))]\, m(dy) \Big] \\ + 1 + c_2'(s) \Big[H(s) + \int_0^{f(i(s),s)} [L(y) - L(i(s))]\, m(dy) \Big] \Bigg]. \end{aligned} \tag{3.47}$$

Similarly, since $(i, x, s(i))$ with $x \geq g(i, s(i))$ belongs to the boundary of $C_{f,g}^+$ we know that $V_{f,g}(i, x, s(i))$ is given by (3.32) above. Writing the integral from $g(i, s(i))$ to x in this expression as the integral from 0 to x minus the integral from 0 to $g(i, s(i))$, it is easily seen that (3.32) reads as follows

$$\begin{aligned} V(i, x, s(i)) = s(i) - i + \big[c_2(s(i)) - c_1(i)\big] \\ \times \Big[H(x) - L(x) \int_0^{g(i,s(i))} m(dy) + \int_0^{g(i,s(i))} L(y)\, m(dy) \Big] \end{aligned} \tag{3.48}$$

for $x \geq g(i, s(i))$. Comparing (3.48) with (3.39) we can conclude that

$$a_1(i) + a_2(s(i)) = -\big[c_2(s(i)) - c_1(i)\big] \int_0^{g(i,s(i))} m(dy) \tag{3.49}$$

$$b_1(i) + b_2(s(i)) = s(i) - i + \big[c_2(s(i)) - c_1(i)\big] \int_0^{g(i,s(i))} L(y)\, m(dy)\,. \tag{3.50}$$

Using (3.40)+(3.41) and (3.49)+(3.50) we can calculate $a_1'(i)$. Firstly, by (3.40) we can express $a_1'(i)$ in terms of $b_1'(i)$. Secondly, by (3.50) we can express $b_1'(i)$ in terms of $b_1'(s(i))$. Thirdly, by (3.41) we can express $b_2'(s(i))$ in terms of $a_2'(s(i))$. Fourthly, by (3.49) we can express $a_2'(s(i))$ in terms of $a_1'(i)$. This closes the loop and gives an equation for $a_1'(i)$. A lengthy calculation following these steps and making use of (3.7) above yields

$$\begin{aligned} a_1'(i) = -\frac{1}{L(s(i)) - L(i)} \Bigg[\frac{g_i'(i,s(i))}{g_s'(i,s(i))} \Big[1 + c_2'(s(i)) \int_{g(i,s(i))}^{s(i)} [L(s(i)) - L(y)]\, m(dy) \Big] \\ + 1 + c_1'(i) \Big[H(i) - \int_0^{g(i,s(i))} [L(s(i)) - L(y)]\, m(dy) \Big] \Bigg]. \end{aligned} \tag{3.51}$$

We can now determine A and B in (3.35) using the closed-form expressions obtained. Firstly, note that by (3.45) we find that

$$A(i, s) = A(i(s), s) + \int_{i(s)}^{i} A_u'(u, s)\, du = a_1(i(s)) + a_2(s) + \int_{i(s)}^{i} a_1'(u)\, du \tag{3.52}$$

$$= -\big[c_2(s) - c_1(i(s))\big] \int_0^{f(i(s),s)} m(dy) + \int_{i(s)}^{i} a_1'(u)\, du$$

where $a_1'(u)$ is given by (3.51) above. Note also that by (3.49) we find that

$$A(i,s) = A(i,s(i)) - \int_s^{s(i)} A_v'(i,v)\, dv = a_1(i) + a_2(s(i)) - \int_s^{s(i)} a_2'(v)\, dv \tag{3.53}$$

$$= -\big[c_2(s(i)) - c_1(i)\big] \int_0^{g(i,s(i))} m(dy) - \int_s^{s(i)} a_2'(v)\, dv$$

where $a_2'(v)$ is given by (3.47) above. Secondly, observe that (3.40) and (3.41) yield

$$b_1'(i) = -a_1'(i) L(i) + c_1'(i) H(i) \tag{3.54}$$

$$b_2'(s) = -a_2'(s) L(s) - c_2'(s) H(s) \tag{3.55}$$

where $a_1'(i)$ and $a_2'(s)$ are given by (3.51) and (3.47) above. Note that by (3.46) we find that

$$B(i,s) = B(i(s),s) + \int_{i(s)}^{i} B_u'(u,s)\, du = b_1(i(s)) + b_2(s) + \int_{i(s)}^{i} b_1'(u)\, du \tag{3.56}$$

$$= s - i(s) + \big[c_2(s) - c_1(i(s))\big] \int_0^{f(i(s),s)} L(y)\, m(dy) + \int_{i(s)}^{i} b_1'(u)\, du$$

where $b_1'(u)$ is given by (3.54) above. Note also that by (3.50) we find that

$$B(i,s) = B(i,s(i)) - \int_s^{s(i)} B_v'(i,v)\, dv = b_1(i) + b_2(s(i)) - \int_s^{s(i)} b_2'(v)\, dv \tag{3.57}$$

$$= s - i(s) + \big[c_2(s(i)) - c_1(i)\big] \int_0^{g(i,s(i))} L(y)\, m(dy) - \int_s^{s(i)} b_2'(v)\, dv$$

where $b_2'(v)$ is given by (3.55) above.

Finally, inserting (3.52)+(3.56) and (3.53)+(3.57) into (3.35) we respectively obtain the following two closed-form expressions

$$V(i,x,s) = s - i(s) + \big[c_2(s) - c_1(i(s))\big] \int_0^{f(i(s),s)} [L(y) - L(x)]\, m(dy) \tag{3.58}$$

$$+ \big[c_2(s) - c_1(i)\big] H(x) + \int_{i(s)}^{i} \big([L(x) - L(u)] a_1'(u) + c_1'(u) H(u)\big)\, du$$

$$V(i,x,s) = s(i) - i + \big[c_2(s(i)) - c_1(i)\big] \int_0^{g(i,s(i))} [L(y) - L(x)]\, m(dy) \tag{3.59}$$

$$+ \big[c_2(s) - c_1(i)\big] H(x) + \int_s^{s(i)} \big([L(v) - L(x)] a_2'(v) + c_2'(v) H(v)\big)\, dv$$

for $f(i,s) > g(i,s)$ where $a_1'(u)$ and $a_2'(v)$ are given by (3.51) and (3.47) above. Observe that if $f(i,s) = g(i,s)$ then $i(s) = i$ and $s(i) = s$ so that the second integral in both (3.58) and (3.59) is zero and these expressions reduce to (3.25) and (3.32) respectively.

7. *Optimality of the minimal and maximal solution.* We will begin by disclosing the superharmonic characterisation of the value function in terms of the solutions to (3.6) and (3.7) staying above/below the lower/upper diagonal respectively. For this, let $i \mapsto f(i,s)$ be any solution to (3.6) satisfying $f(i,s) > i$ for all i, and let $s \mapsto g(i,s)$ be any solution to (3.7) satisfying $g(i,s) < s$ for all s. Consider the function $V_{f,g}$ defined by (3.25)+(3.32)+(3.58) (or (3.59)) on $C_{f,g}$ and set $V_{f,g}(i,x,s) = s-i$ for $(i,x,s) \notin C_{f,g}$. Recall that $V_{f,g}$ solves the free-boundary problem (3.12)-(3.18) on $C_{f,g}$. Due to the 'triple-deck' structure of $V_{f,g}$ we can apply the change-of-variable formula with local time on surfaces [30] which in view of (3.17) and (3.18) reduces to standard Itô's formula and gives

$$
\begin{aligned}
(3.60)\qquad V_{f,g}(I_t,X_t,S_t) &= V_{f,g}(i,x,s) + \int_0^t \frac{\partial V_{f,g}}{\partial i}(I_s,X_s,S_s)\,dI_s + \int_0^t \frac{\partial V_{f,g}}{\partial x}(I_s,X_s,S_s)\,dX_s \\
&\quad + \int_0^t \frac{\partial V_{f,g}}{\partial s}(I_s,X_s,S_s)\,dS_s + \frac{1}{2}\int_0^t \frac{\partial^2 V_{f,g}}{\partial x^2}(I_s,X_s,S_s)\,d\langle X,X\rangle_s \\
&= V_{f,g}(i,x,s) + \int_0^t \sigma(X_s)\,\frac{\partial V_{f,g}}{\partial x}(I_s,X_s,S_s)\,dB_s \\
&\quad + \int_0^t (I\!\!L_X V_{f,g})(I_s,X_s,S_s)\,ds
\end{aligned}
$$

where we also use (3.13) and (3.14) to conclude that the integrals with respect to dI_s and dS_s are equal to zero. The process $M = (M_t)_{t\ge 0}$ defined by

$$
(3.61)\qquad M_t = \int_0^t \sigma(X_s)\,\frac{\partial V_{f,g}}{\partial x}(I_s,X_s,S_s)\,dB_s
$$

is a continuous local martingale. Introducing the increasing process $P = (P_t)_{t\ge 0}$ by setting

$$
(3.62)\qquad P_t = \int_0^t c(S_s-I_s)\,I(f(I_s,S_s) \le X_s \le g(I_s,X_s))\,ds
$$

and using the fact that the set of all s for which X_s is either $f(I_s,S_s)$ or $g(I_s,S_s)$ is of Lebesgue measure zero, we see by (3.12) that (3.60) can be rewritten as follows

$$
(3.63)\qquad V_{f,g}(I_t,X_t,S_t) - \int_0^t c(S_s-I_s)\,ds = V_{f,g}(i,x,s) + M_t - P_t\,.
$$

From this representation we see that the process $V_{f,g}(I_t,X_t,S_t) - \int_0^t c(S_s-I_s)\,ds$ is a local supermartingale for $t \ge 0$.

Let τ be any stopping time of X. Choose a localisation sequence $(\sigma_n)_{n\ge 1}$ of bounded stopping times for M. Then for any (i,s) such that $f(i,s) \le g(i,s)$ we see from (3.25) and (3.32) that $V_{f,g}(i,x,s) \ge s-i$ for any $x \in [i,s]$. Note that each such (i,x,s) belongs to $C^-_{f,g} \cup C^+_{f,g} \cup D_{f,g}$ where $D_{f,g}$ denotes the complement of $C_{f,g}$. Moreover, it is important to notice that after starting at any point in the set $C^-_{f,g} \cup C^+_{f,g} \cup D_{f,g}$ the process (I,X,S) remains in the same set for the rest of time (i.e. it never enters the set $C^0_{f,g}$). For these reasons we can conclude from (3.63) using the optional sampling theorem that

$$
(3.64)\qquad \mathsf{E}_{i,x,s}\Big[S_{\tau\wedge\sigma_n} - I_{\tau\wedge\sigma_n} - \int_0^{\tau\wedge\sigma_n} c(S_t-I_t)\,dt\Big]
$$

$$\leq \mathsf{E}_{i,x,s}\Big[V_{f,g}(I_{\tau\wedge\sigma_n}, X_{\tau\wedge\sigma_n}, S_{\tau\wedge\sigma_n}) - \int_0^{\tau\wedge\sigma_n} c(S_t - I_t)\, dt\Big]$$
$$\leq V_{f,g}(i,x,s) + \mathsf{E}_{i,x,s}(M_{\tau\wedge\sigma_n}) = V_{f,g}(i,x,s)$$

for all $n \geq 1$. Letting $n \to \infty$ and using the monotone convergence theorem we find that

$$\mathsf{E}_{i,x,s}\Big[S_\tau - I_\tau - \int_0^\tau c(S_t - I_t)\, dt\Big] \leq V_{f,g}(i,x,s) \tag{3.65}$$

for all $(i,x,s) \in C^-_{f,g} \cup C^+_{f,g} \cup D_{f,g}$. Taking first the supremum over all τ , and then the infimum over all f and g , we conclude that

$$V(i,x,s) \leq \inf_{f,g} V_{f,g}(i,x,s) = V_{f_*,g_*}(i,x,s) \tag{3.66}$$

where f_* denotes the minimal solution to (3.6) staying above the lower diagonal and g_* denotes the maximal solution to (3.7) staying below the upper diagonal. Recalling that $f \mapsto V_{f,g}$ is increasing and $g \mapsto V_{f,g}$ is decreasing when $f \leq g$ we see that the infimum in (3.66) is attained over any sequence of solutions f_n and g_n to (3.6) and (3.7) such that $f_n \downarrow f_*$ and $g_n \uparrow g_*$ as $n \to \infty$. Since f_* and g_* are solutions themselves to which (3.65) applies, we see that (3.66) holds for all (i,x,s) in the set $C^-_{f_*,g_*} \cup C^+_{f_*,g_*} \cup D_{f_*,g_*}$ which is the increasing union of the sets $C^-_{f_n,g_n} \cup C^+_{f_n,g_n} \cup D_{f_n,g_n}$ for $n \geq 1$. From these considerations and (3.66) in particular it follows that the only possible candidate for the optimal stopping boundary are the minimal and maximal solution f_* and g_* . Note that (3.64) also implies that

$$\mathsf{E}_{i,x,s}\Big[V_{f,g}(I_\tau, X_\tau, S_\tau) - \int_0^\tau c(S_t - I_t)\, dt\Big] \leq V_{f,g}(i,x,s) \tag{3.67}$$

showing that the function $(i,x,s,a) \mapsto V_{f,g}(i,x,s) - a$ is superharmonic for the Markov process (I,X,S,A) on the set $C^-_{f,g} \cup C^+_{f,g} \cup D_{f,g}$ where $A_t = \int_0^t c(S_s - I_s)\, ds$ for $t \geq 0$. Recalling that $f \mapsto V_{f,g}$ is increasing and $g \mapsto V_{f,g}$ is decreasing when $f \leq g$, and that $V_{f,g}(i,x,s) \geq s - i$ for all $(i,x,s) \in C^-_{f,g} \cup C^+_{f,g} \cup D_{f,g}$, we see that selecting the minimal solution f_* staying above the lower diagonal and the maximal solution g_* staying below the upper diagonal is equivalent to invoking the superharmonic characterisation of the value function (according to which the value function is the smallest superharmonic function which dominates the gain function). For more details on the latter characterisation in a general setting we refer to [33, Chapter 1]. It is also useful to know that the superharmonic characterisation of the value function represents the dual problem to the primal problem (3.4) (for more details on the meaning of this claim including connections to the Legendre transform see [31]).

To prove that f_* and g_* are optimal, consider the stopping time τ_{f_n,g_n} defined in (3.8) where $i \mapsto f_n(i,s)$ is the solution to (3.6) such that $f_n(i_n,s) = i_n$ and $s \mapsto g_n(i,s)$ is the solution to (3.7) such that $g_n(i,s_n) = s_n$ for some $i_n \downarrow -1$ and $s_n \uparrow 1$ as $n \to \infty$. Consider the function V_{f_n,g_n} defined by (3.25)+(3.32) on $C^-_{f_n,g_n} \cup C^+_{f_n,g_n}$ and set $V_{f_n,g_n}(i,x,s) = s - i$ for $(i,x,s) \in D_{f_n,g_n}$ and $n \geq 1$. Recall that V_{f_n,g_n} solves the free-boundary problem (3.12)-(3.18) on C_{f_n,g_n} for $n \geq 1$. The same arguments as above yield the formula (3.60) with f_n and g_n in place of f and g for $n \geq 1$. Since σ and $\partial V_{f_n,g_n}/\partial x$ are bounded on $C^-_{f_n,g_n} \cup C^+_{f_n,g_n} \cup D_{f_n,g_n}$ and the process (I,X,S) never leaves the latter set we see that M

defined by (3.61) with f_n and g_n in place of f and g is a martingale under $\mathsf{P}_{i,x,s}$ and τ_{f_n,g_n} is optional for M whenever $(i,x,s) \in C^-_{f_n,g_n} \cup C^+_{f_n,g_n} \cup D_{f_n,g_n}$ and $n \geq 1$. The latter conclusion follows from the fact that $\tau_{f_n,g_n} \leq \tau_{i_n,s_n}$ implying also that $\mathsf{E}_{i,x,s} \int_0^{\tau_{f_n,g_n}} c(S_t - I_t)\, dt < \infty$ for $n \geq 1$. Since the process P defined by (3.62) with f_n and g_n in place of f and g satisfies $P_{\tau_{f_n,g_n}} = 0$, it follows from (3.63) using (3.15) and (3.16) that

$$V_{f_n,g_n}(i,x,s) = \mathsf{E}_{i,x,s}\Big[S_{\tau_{f_n,g_n}} - I_{\tau_{f_n,g_n}} - \int_0^{\tau_{f_n,g_n}} c(S_t - I_t)\, dt\Big] \tag{3.68}$$

for all $i \leq x \leq s$ such that $f_n(i,s) \leq g_n(i,s)$ with $n \geq 1$. Recalling that $f_n \downarrow f_*$ and $g_n \uparrow f_*$ as $n \to \infty$ we see that taking any (i,s) such that $f(i,s) < g(i,s)$ we can find $n_0 \geq 1$ such that $f_n(i,s) \leq g_n(i,s)$ for all $n \geq n_0$. Letting $n \to \infty$ in (3.68), noting that $\tau_{f_n,g_n} \uparrow \tau_{f_*,g_*}$ (since $[-1,1]^3$ is compact), and using the monotone convergence theorem we find that

$$V_{f_*,g_*}(i,x,s) = \mathsf{E}_{i,x,s}\Big[S_{\tau_{f_*,g_*}} - I_{\tau_{f_*,g_*}} - \int_0^{\tau_{f_*,g_*}} c(S_t - I_t)\, dt\Big] \tag{3.69}$$

for all $i \leq x \leq s$ such that $f_*(i,s) \leq g_*(i,s)$. This shows that we have equality in (3.66) and completes the proof of the optimality of τ_{f_*,g_*} on the set $C^-_{f_*,g_*} \cup C^+_{f_*,g_*} \cup D_{f_*,g_*}$.

To prove the optimality of τ_{f_*,g_*} on the set $C^0_{f_*,g_*}$, i.e. when $f_*(i,s) > g_*(i,s)$ for some (i,s) given and fixed, one could attempt to apply similar arguments to those in (3.64) above. For this, however, we would need to know that $V_{f,g}(i,x,s) \geq s - i$ not only for $f(i,s) \leq g(i,s)$ as follows from the closed-form expressions (3.25)+(3.32) above but also for $f(i,s) > g(i,s)$. A closer inspection of the latter case indicates that this verification may be problematic if it is to follow from similar closed-form expressions. Indeed, even in the special case of $c(r) = c_2(s) - c_1(i)$ we see from (3.58) and (3.59) that the conclusion is unclear since $a_1'(u)$ and $a_2'(v)$ appearing there could also (at least in principle) take negative values as well (see (3.47) and (3.51) above). To overcome this difficulty we will exploit the extremal properties of the candidate surfaces f_* and g_* in an essential way (in many ways this can be seen as a key argument in the proof showing the full power of the method). For this, take any point (i_0, x_0, s_0) in the state space such that $f_*(i_0,s_0) > g_*(i_0,s_0)$ and consider the cases when x belongs to either $[i_0, s_0 \wedge f_*(i_0,s_0))$ or $(i_0 \vee g_*(i_0,s_0), s_0]$ respectively. Take any d_0 in either $(x \vee g_*(i_0,s_0), s_0 \wedge f_*(i_0,s_0))$ or $(i_0 \vee g_*(i_0,s_0), x \wedge f_*(i_0,s_0))$ respectively and choose solutions $i \mapsto f_d(i,s_0)$ and $s \mapsto g_d(i_0,s)$ to (3.6) and (3.7) such that $f_d(i_0,s_0) = d_0$ and $g_d(i_0,s_0) = d_0$ respectively. Note that this is possible since d_0 lies strictly between i_0 and $f_*(i_0,s_0)$ in the first case and strictly between $g_*(i_0,s_0)$ and s_0 in the second case. Note also that $i \mapsto f_d(i,s_0)$ must hit the lower diagonal and $s \mapsto g_d(i_0,s)$ must hit the upper diagonal since $i \mapsto f_*(i,s_0)$ and $s \mapsto g_*(i_0,s)$ are the minimal and maximal solutions staying above/below the lower/upper diagonal respectively. Moreover, by the construction of f_d and g_d we see that (i_0,x_0,s_0) belongs to either $C^-_{f_d,g_d}$ or $C^+_{f_d,g_d}$ respectively, and after starting at (i_0,x_0,s_0) the process (I,X,S) remains in either $C^-_{f_d,g_d}$ or $C^+_{f_d,g_d}$ respectively before hitting D_{f_d,g_d} . Considering the stopping time τ_{f_d,g_d} defined in (3.8) we therefore see that the same arguments as those leading to (3.68) also show that

$$V_{f_d,g_d}(i_0,x_0,s_0) = \mathsf{E}_{i_0,x_0,s_0}\Big[S_{\tau_{f_d,g_d}} - I_{\tau_{f_d,g_d}} - \int_0^{\tau_{f_d,g_d}} c(S_t - I_t)\, dt\Big] \tag{3.70}$$

where V_{f_d,g_d} is given by either (3.25) or (3.32) respectively. From the latter closed-form expressions we see that $V_{f_d,g_d}(i_0,x_0,s_0) > s_0 - i_0$ and from (3.70) it therefore follows that (i_0,x_0,s_0) belongs to the continuation set C. Combining this conclusion with the description of the stopping set D outside $C^0_{f_*,g_*}$ derived above, we see that $C = C^0_{f_*,g_*} \cup C^-_{f_*,g_*} \cup C^+_{f_*,g_*}$. This proves the optimality of τ_* in (3.5) and completes the proof. □

Remark 2. The arguments used in the proof above can easily be extended to cover more general settings of the optimal stopping problem (3.4). As stated above it is not essential that the state space of the diffusion process X equals $(-1,1)$ and the result of Theorem 1 is valid for more general state spaces (including $\mathbb{R}$ and $\mathbb{R}_+$ in particular). This also includes various boundary behaviour of the process X at the end points of the state space (for example 0 when the state space equals $\mathbb{R}_+$). Similarly, the cost function c in (3.4) can be more general. For example, the arguments used in the proof above are also valid when $c(R_t)$ is replaced by $c(I_t,X_t,S_t)$ where $i \mapsto c(i,x,s)$ is decreasing and $s \mapsto c(i,x,s)$ is increasing (both being strictly positive). Omitting further details in this direction we briefly turn to some examples.

4. Examples

Combining the results of Proposition 1 and Theorem 1 we obtain the solution to the quickest detection problem (2.3). We illustrate various special cases of this correspondence through one particular example.

Example 1. Assume that the observed process Z is a standard Brownian motion B starting at 0, suppose that ℓ is a standard normal random variable independent from B, and consider the quickest detection problem (2.3) where $c > 0$ is a given and fixed constant. By the result of Proposition 1 we know that this problem is equivalent to the optimal stopping problem (2.14) where $X = 2F(Z) - 1$ solves (2.10) with μ and σ given by (2.11) and (2.12). From (2.1) we see that $a = 0$ and $b = 1$ so that

$$\mu(x) = -\Phi^{-1}\big(\tfrac{x+1}{2}\big)\, \varphi\big(\Phi^{-1}\big(\tfrac{x+1}{2}\big)\big) \tag{4.1}$$

$$\sigma(x) = 2\varphi\big(\Phi^{-1}\big(\tfrac{x+1}{2}\big)\big) \tag{4.2}$$

for $x \in (-1,1)$ where $\Phi(y) = (1/\sqrt{2\pi}) \int_{-\infty}^{y} e^{-z^2/2}\, dz$ is the standard normal distribution function and $\varphi(y) = (1/\sqrt{2\pi})\, e^{-y^2/2}$ is the standard normal density function for $y \in \mathbb{R}$. It is easily verified using (2.16) that the scale function of X can be taken as

$$L(x) = \int_0^x \exp\Big(\tfrac{1}{2}\big(\Phi^{-1}\big(\tfrac{y+1}{2}\big)\big)^2\Big)\, dy \tag{4.3}$$

for $x \in (-1,1)$. Note that the expected value of $\int_0^\infty c(S_t - I_t)\, dt$ is infinite so that it cannot be optimal to continue forever. By the result of Theorem 1 we therefore know that the following stopping time is optimal

$$\tau_* = \inf\, \{\, t \geq 0 \mid f_*(I_t,S_t) \leq X_t \leq g_*(I_t,S_t)\, \} \tag{4.4}$$

where the surfaces f_* and g_* are the minimal and maximal solutions to

$$
\frac{\partial f}{\partial i}(i,s) = \frac{2\varphi^2\big(\Phi^{-1}\big(\frac{f(i,s)+1}{2}\big)\big)\exp\Big(\frac{1}{2}\big(\Phi^{-1}\big(\frac{f(i,s)+1}{2}\big)\big)^2\Big)}{c(s-i)\int_i^{f(i,s)}\exp\Big(\frac{1}{2}\big(\Phi^{-1}\big(\frac{y+1}{2}\big)\big)^2\Big)\,dy} \times\left[1+c\int_i^{f(i,s)}\frac{\int_i^y \exp\Big(\frac{1}{2}\big(\Phi^{-1}\big(\frac{z+1}{2}\big)\big)^2\Big)\,dz}{2\varphi^2\big(\Phi^{-1}\big(\frac{y+1}{2}\big)\big)\exp\Big(\frac{1}{2}\big(\Phi^{-1}\big(\frac{y+1}{2}\big)\big)^2\Big)}\,dy\right] \tag{4.5}
$$

$$
\frac{\partial g}{\partial s}(i,s) = \frac{2\varphi^2\big(\Phi^{-1}\big(\frac{g(i,s)+1}{2}\big)\big)\exp\Big(\frac{1}{2}\big(\Phi^{-1}\big(\frac{g(i,s)+1}{2}\big)\big)^2\Big)}{c(s-i)\int_{g(i,s)}^{s}\exp\Big(\frac{1}{2}\big(\Phi^{-1}\big(\frac{y+1}{2}\big)\big)^2\Big)\,dy} \times\left[1+c\int_{g(i,s)}^{s}\frac{\int_y^s \exp\Big(\frac{1}{2}\big(\Phi^{-1}\big(\frac{z+1}{2}\big)\big)^2\Big)\,dz}{2\varphi^2\big(\Phi^{-1}\big(\frac{y+1}{2}\big)\big)\exp\Big(\frac{1}{2}\big(\Phi^{-1}\big(\frac{y+1}{2}\big)\big)^2\Big)}\,dy\right] \tag{4.6}
$$

staying above the lower diagonal d^s and below the upper diagonal d_i respectively. The equations (4.5) and (4.6) are singular at the lower and upper diagonal. Passing to the inverse equations $\partial i/\partial f$ and $\partial s/\partial g$ these singularities get removed and one can determine the minimal and maximal solution by approximating them with the solutions which hit the lower and upper diagonal respectively (as explained in the proof above). The results of these calculations are illustrated in Figures 1-3. Similar qualitative behaviour of the optimal surfaces can also be observed in other examples of diffusions and hidden levels.

The long list of examples is easily continued. Apart from the problems where the optimal stopping boundaries are surfaces, this also includes problems where the optimal stopping boundaries are curves. We illustrate this through one known example from stochastic analysis.

Example 2. Taking X to be a standard Brownian motion B and setting $c(r) \equiv c$, it is easily seen that the minimal and maximal solutions to (2.6) and (2.7) are given by

$$
f(i,s) = i + \tfrac{1}{2c} \quad \& \quad g(i,s) = s - \tfrac{1}{2c}\,. \tag{4.7}
$$

From (3.42) we see that $i(s) = s - \frac{1}{c}$ and $s(i) = i + \frac{1}{c}$. Since $f_s' \equiv 0$ we see from (3.47) that $a_2'(s) \equiv -c$. Inserting this into (3.59) we find that $V(0,0,0) = \frac{3}{4c}$. This shows that for any stopping time τ of B we have

$$
\mathsf{E}(S_\tau - I_\tau) \le c\,\mathsf{E}\tau + \tfrac{3}{4c}\,. \tag{4.8}
$$

Taking the infimum over all $c > 0$ we obtain the result of [6]:

$$
\mathsf{E}(S_\tau - I_\tau) \le \sqrt{3}\,\sqrt{\mathsf{E}\tau}\,. \tag{4.9}
$$

One can extract similar other inequalities/information from the proof above.

References

[1] Barles, G. Daher, C. and Romano, M. (1994). Optimal control on the L^∞ norm of a diffusion process. *SIAM J. Control Optim.* 32 (612–634).

[2] Barron, E. N. *and* Ishii, H. (1989). The Bellman equation for minimizing the maximum cost. *Nonlinear Anal.* 13 (1067–1090).

[3] Bernyk, V. Dalang, R. C. *and* Peskir, G. (2010). Predicting the ultimate supremum of a stable Lévy process with no negative jumps. To appear in *Ann. Probab.*

[4] Cohen, A. (2010). Examples of optimal prediction in the infinite horizon case. *Statist. Probab. Lett.* 80 (950–957).

[5] Cox, A. M. G. Hobson, D. *and* Obłój, J. (2008). Pathwise inequalities for local time: Applications to Skorokhod embeddings and optimal stopping. *Ann. Appl. Probab.* 18 (1870–1896).

[6] Dubins, L. E. Gilat, D. *and* Meilijson, I. (2009). On the expected diameter of an L_2-bounded martingale. *Ann. Probab.* 37 (393–402).

[7] Dubins, L. E. *and* Schwarz, G. (1988). A sharp inequality for sub-martingales and stopping times. *Astérisque* 157–158 (129–145).

[8] Dubins, L. E. Shepp, L. A. *and* Shiryaev, A. N. (1993). Optimal stopping rules and maximal inequalities for Bessel processes. *Theory Probab. Appl.* 38 (226–261).

[9] Du Toit, J. *and* Peskir, G. (2007). The trap of complacency in predicting the maximum. *Ann. Probab.* 35 (340–365).

[10] Du Toit, J. *and* Peskir, G. (2008). Predicting the time of the ultimate maximum for Brownian motion with drift. *Proc. Math. Control Theory Finance* (Lisbon 2007), Springer (95–112).

[11] Du Toit, J. *and* Peskir, G. (2009). Selling a stock at the ultimate maximum. *Ann. Appl. Probab.* 19 (983–1014).

[12] Du Toit, J. Peskir, G. *and* Shiryaev, A. N. (2008). Predicting the last zero of Brownian motion with drift. *Stochastics* 80 (229–245).

[13] Elie, R. *and* Espinosa, G-E. (2011). Optimal stopping of a mean reverting diffusion: minimizing the relative distance to the maximum. *Working Paper, Université Paris-Dauphine* (37 pp).

[14] Espinosa, G-E. *and* Touzi, N. (2010). Detecting the maximum of a mean-reverting scalar diffusion. *Working Paper, Université Paris-Dauphine* (37 pp).

[15] Gapeev, P. V. (2006). Discounted optimal stopping for maxima in diffusion models with finite horizon. *Electron. J. Probab.* 11 (1031–1048).

[16] Gapeev, P. V. (2007). Discounted optimal stopping for maxima of some jump-diffusion processes. *J. Appl. Probab.* 44 (713–731).

[17] Goldman, M. B. Sosin, H. B. *and* Gatto, M. A. (1979). Path dependent options: "Buy at the low, sell at the high". *J. Finance* 34 (1111–1127).

[18] Graversen, S. E. Peskir, G. *and* Shiryaev, A. N. (2001). Stopping Brownian motion without anticipation as close as possible to its ulimate maximum. *Theory Probab. Appl.* 45 (41–50).

[19] Guo, X. *and* Zervos, M. (2010). π options. *Stochastic Process. Appl.* 120 (1033–1059).

[20] Heinricher, A. C. and Stockbridge, R. H. (1991). Optimal control of the running max. *SIAM J. Control Optim.* 29 (936–953).

[21] Hobson, D. (2007). Optimal stopping of the maximum process: a converse to the results of Peskir. *Stochastics* 79 (85–102).

[22] Jacka, S. D. (1991). Optimal stopping and best constants for Doob-like inequalities I: The case $p = 1$. *Ann. Probab.* 19 (1798–1821).

[23] Novikov, A. A. *and* Shiryaev, A. N. (2008). On a stochastic version of the trading rule "buy and hold". *Statist. Decisions* 26 (289–302).

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

[25] Obłój, J. (2007). The maximality principle revisited: On certain optimal stopping problems. *Sem. Probab.* XL, *Lecture Notes in Math.* 1899, Springer (309–328).

[26] Pedersen, J. L. (2000). Discounted optimal stopping problems for the maximum process. *J. Appl. Probab.* 37 (972–983).

[27] Pedersen, J. L. (2003). Optimal prediction of the ultimate maximum of Brownian motion. *Stoch. Stoch. Rep.* 75 (205–219).

[28] Peskir, G. (1998). Optimal stopping of the maximum process: The maximality principle. *Ann. Probab.* 26 (1614–1640).

[29] Peskir, G. (2005). Maximum process problems in optimal control theory. *J. Appl. Math. Stoch. Anal.* (77–88).

[30] Peskir, G. (2007). A change-of-variable formula with local time on surfaces. *Sém. de Probab.* XL, *Lecture Notes in Math.* 1899, Springer (69–96).

[31] Peskir, G. (2010). A duality principle for the Legendre transform. *Research Report* No. 2, *Probab. Statist. Group Manchester* (21 pp). To appear in *J. Convex Anal.*

[32] Peskir, G. (2010). Optimal detection of a hidden target: The median rule. *Research Report* No. 8, *Probab. Statist. Group Manchester* (13 pp). Submitted.

[33] Peskir, G. *and* Shiryaev, A. N. (2006). *Optimal Stopping and Free-Boundary Problems.* Lectures in Mathematics, ETH Zürich, Birkhäuser.

[34] Shiryaev, A. N. (1961). The problem of the most rapid detection of a disturbance of a stationary regime. *Soviet Math. Dokl.* 2 (795–799).

[35] Shiryaev, A. N. (1963). On optimal methods in quickest detection problems. *Theory Probab. Appl.* 8 (22–46).

[36] Shiryaev, A. N. (2002). Quickest detection problems in the technical analysis of the financial data. *Proc. Math. Finance Bachelier Congress* (Paris, 2000), Springer (487–521).

[37] Shiryaev, A. N. (2009). On conditional-extremal problems of the quickest detection of nonpredictable times of the observable Brownian motion. *Theory Probab. Appl.* 53 (663–678).

[38] Urusov, M. A. (2005). On a property of the moment at which Brownian motion attains its maximum and some optimal stopping problems. *Theory Probab. Appl.* 49 (169–176).

[39] Zhitlukhin, M. (2009). A maximal inequality for skew Brownian motion. *Statist. Decisions* 27 (261–280).

Goran Peskir
School of Mathematics
The University of Manchester
Oxford Road
Manchester M13 9PL
United Kingdom
goran@maths.man.ac.uk