The Annals of Applied Probability
2014, Vol. 24, No. 6, 2340–2370
DOI: 10.1214/13-AAP979

QUICKEST DETECTION OF A HIDDEN TARGET AND EXTREMAL SURFACES

BY GORAN PESKIR

University of Manchester

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 $\tau_\ell = \inf\{t \geq 0 | Z_t = \ell\}$ 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

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

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:

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

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 [*Ann. Probab.* **26** (1998) 1614–1640] 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

Received January 2013.

MSC2010 subject classifications. 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.

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 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] and [33], Sections 22 and 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 Z change following τ_ℓ so that there is no learning about the position of ℓ through the observation of Z (quickest detection problems of this kind require a different treatment and will be studied elsewhere). Likewise, since the underlying loss processes $t \mapsto 1(t < \tau_\ell)$ and $t \mapsto 1(t - \tau_\ell)^+$ are not adapted to the natural filtration generated by Z (or its usual augmentation), we see that 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, 6–9, 13, 14, 17, 19, 27, 36–39]). It may be noted in this context that the nonadapted 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 $Z = (Z_t)_{t \geq 0}$ 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 $B = (B_t)_{t \geq 0}$ is a standard Brownian motion. To gain tractability we also assume that the distribution function F of ℓ is strictly increasing and twice continuously differentiable. In the first step we show that problem (1.1) is equivalent to the optimal stopping problem

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

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 [11, 12] and [23]. 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, 20, 22, 24–26]; see also [1, 2, 21] and [29] for related results in optimal control theory. Close three-dimensional relatives of the problem (1.3) also appear in the recent papers [10] and [40] 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 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, replacing $c(r)$ in problem (1.3) above with a more general function $c(i,x,s)$ specified below, we show that the following stopping time is optimal:

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

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

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

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

staying strictly above/below the lower/upper diagonal in the state space, respectively (Theorem 1). 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 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 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 may be combined with Lagrange multipliers to tackle 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 easily seen that swapping the order of τ and τ_ℓ in (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 first 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 $Z = (Z_t)_{t\geq 0}$ 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 $B = (B_t)_{t\geq 0}$ is a standard Brownian motion. To meet a sufficient condition used in the proof of Theorem 1 below we will also assume that b^2 is (locally) Lipschitz. Let ℓ be an independent random variable with values in $\mathbb{R}$, and let

$$\tau_\ell = \inf\{t \geq 0 \,|\, Z_t = \ell\} \tag{2.2}$$

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

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

where the infimum is taken over all stopping times τ of Z [i.e., with respect to the natural filtration $(\mathcal{F}_t^Z)_{t\geq 0}$ generated by Z], and $c > 0$ is a given and fixed constant (note that whenever we say a stopping time throughout we always mean a finite valued stopping time). 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

$$I_t^Z = \inf_{0\leq s\leq t} Z_s \quad \text{and} \quad S_t^Z = \sup_{0\leq s\leq t} Z_s \tag{2.4}$$

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

PROPOSITION 1. *Problem* (2.3) *is equivalent to the optimal stopping problem*

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

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

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

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

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

$$\begin{aligned}\mathsf{E}(\tau - \tau_\ell)^+ &= \mathsf{E}\int_0^\tau 1(\tau_\ell \le t)\,dt = \mathsf{E}\int_0^\infty 1(\tau_\ell \le t)1(t < \tau)\,dt\\ &= \int_0^\infty \mathsf{E}\big[\mathsf{E}\big(1(\tau_\ell \le t)1(t<\tau)|\mathcal{F}_t^Z\big)\big]\,dt\\ &= \int_0^\infty \mathsf{E}\big[1(t<\tau)\mathsf{E}\big(1(\tau_\ell \le t)|\mathcal{F}_t^Z\big)\big]\,dt\\ &= \mathsf{E}\int_0^\tau \mathsf{P}\big(\tau_\ell \le t|\mathcal{F}_t^Z\big)\,dt.\end{aligned} \tag{2.7}$$

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

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

for $t \ge 0$. 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. 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

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

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

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

$$I_t = \inf_{0\le s\le t} X_s \quad \text{and} \quad S_t = \sup_{0\le s\le t} X_s \tag{2.13}$$

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

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

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 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

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

for $x \in (-1, 1)$. Throughout we denote $\rho_a = \inf\{t \ge 0 \,|\, X_t = a\}$ and set $\rho_{a,b} = \rho_a \wedge \rho_b$ 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

$$\mathsf{P}_x(X_{\rho_{a,b}} = a) = \frac{L(b) - L(x)}{L(b) - L(a)} \quad \text{and} \quad \mathsf{P}_x(X_{\rho_{a,b}} = b) = \frac{L(x) - L(a)}{L(b) - L(a)} \tag{2.17}$$

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

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

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

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

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 in somewhat greater generality. 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. To meet a sufficient condition used in the proof below, we will also assume that σ^2 is (locally) Lipschitz. We will further assume that the state space of X equals $(-1, 1)$ as in the previous section; however, this hypothesis is not essential; see Remark 4 below. 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

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

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

$$V(i, x, s) = \sup_{\tau} \mathsf{E}_{i,x,s}\left[R_\tau - \int_0^\tau c(I_t, X_t, S_t)\,dt\right] \tag{3.4}$$

for $i \le x \le s$ in $(-1, 1)$ where the supremum is taken over all stopping times τ of X.

Regarding the cost function c in (3.4) we will assume that (i) $i \mapsto c(i, x, s)$ is decreasing and $s \mapsto c(i, x, s)$ is increasing with $c(i, x, s) > 0$ for $i \le x \le s$ in $(-1, 1)$. These conditions have a natural interpretation in the sense that any new increase in gain (when X reaches either S or I) is followed by a proportional increase in cost. To gain existence and tractability we will also assume that (ii) $(i, x, s) \mapsto c(i, x, s)$ is continuous, $x \mapsto c(i, x, s)$ is (locally) Lipschitz,

$(i,s) \mapsto c(i,x,s)$ is continuously differentiable. To gain monotonicity and joint continuity we will further assume that (iii) $i \mapsto \frac{\partial c}{\partial s}(i,x,s)$ and $s \mapsto \frac{\partial c}{\partial i}(i,x,s)$ are increasing and (locally) Lipschitz. Note that conditions (i)–(iii) are satisfied for $c(i,x,s) = c(s-i) > 0$ when c is increasing concave and continuously differentiable with c' (locally) Lipschitz. Note also that conditions (i)–(iii) are satisfied for $c(i,x,s) = c_2(s) - c_1(i) > 0$ when c_1 and c_2 are increasing and continuously differentiable functions. Note finally that conditions (i)–(iii) are satisfied for $c(i,x,s) = c(x) > 0$ when c is (locally) Lipschitz (in this case f_* and g_* below are no longer surfaces but curves as functions of i and s, respectively).

For any s given and fixed we will refer to $d^s = \{(i,x) | i = x \le s\}$ as the *lower diagonal* in the state space, and for any i given and fixed we will refer to $d_i = \{(x,s) | x = s \ge i\}$ as the *upper diagonal* in the state space. We will say that a function f *stays strictly above the lower diagonal* d^s if $f(i,s) > i$ for all $i < s$, and we will say that a function g *stays strictly below the upper diagonal* d_i if $g(i,s) < s$ for all $s > i$.

The main result of the paper may now be stated as follows.

THEOREM 1. *Under the hypotheses on X and c stated above, the optimal stopping time in problem* (3.4) *is given by*

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

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

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

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

staying strictly above the lower diagonal d^s and strictly below the upper diagonal d_i for $i < s$ in $(-1,1)$, respectively.

Explicit formulae for the value function V on the continuation sets (3.10) *and* (3.11) *below are given by* (3.25) *and* (3.32) *below for any cost function c satisfying* (i)–(iii) *above. Explicit formulae for the value function V on the continuation set* (3.9) *below are given by* (3.64) *and* (3.65) *below when $c(i,s) = c_2(s) - c_1(i) > 0$ where c_1 and c_2 are increasing and continuously differentiable functions. Outside these sets the value function V equals $s - i$ for $i < s$ in $(-1,1)$. The optimal surfaces f_* and g_* satisfy the additional properties* (3.34)–(3.39).

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 below. 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

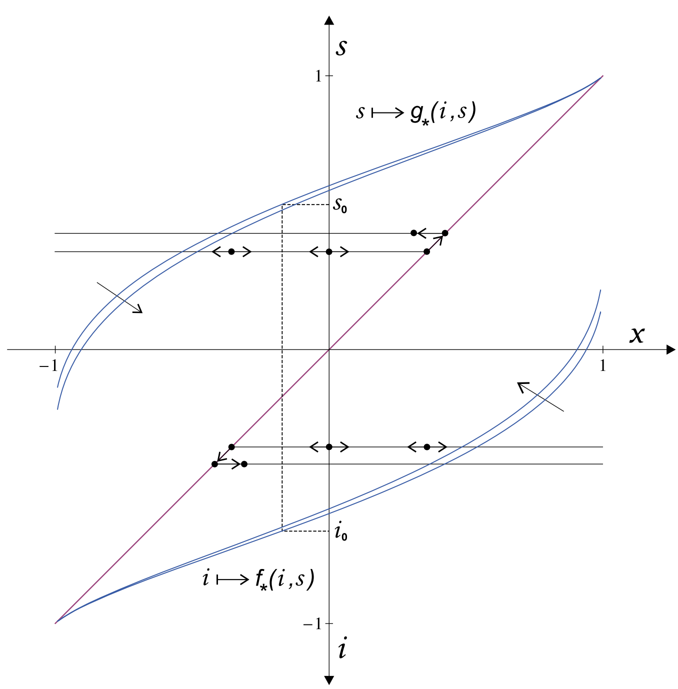

FIG. 1. *Excursions of X away from the running minimum I and the running maximum S combined with the dynamics of the optimal stopping surfaces f_* and g_**: (i) *return of X to the lower diagonal causes I to go down and forces g_* to go up*; (ii) *return of X to the upper diagonal causes S to go up and 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.*

optimal stopping for Markov processes (see [33], Chapter 1) implies that the continuation set in the problem (3.4) equals $C = \{(i, x, s) | V(i, x, s) > s - i\}$ and the stopping set equals $D = \{(i, x, s) | V(i, x, s) = s - i\}$. It means that the first entry time of (I, X, S) into D is optimal in 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 $\rho_{l_n,r_n} = \inf\{t \geq 0 | X_t \notin (l_n, r_n)\}$ under $\mathsf{P}_{i,s,s}$ with $l_n = s - 1/n$ and $r_n = s + 1/n$ for $n \geq 1$. Then (2.18)–(2.20) imply that $\mathsf{E}_{i,s,s} R_{\rho_{l_n,r_n}} \geq s - i + K/n$ and $\mathsf{E}_{i,s,s} \int_0^{\rho_{l_n,r_n}} c(I_t, X_t, S_t)\, dt \leq K/n^2$ 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]. To visualise these movements see Figure 1 above. 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:

$$\tau_{f,g} = \inf\{t \geq 0 \,|\, f(I_t, S_t) \leq X_t \leq g(I_t, S_t)\}, \tag{3.8}$$

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

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

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

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

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

$$(\mathbb{L}_X V)(i,x,s) = c(i,x,s) \qquad \text{for } (i,x,s) \in C_{f,g}, \tag{3.12}$$

$$V'_i(i,x,s)|_{x=i+} = 0 \qquad \text{(normal reflection)}, \tag{3.13}$$

$$V'_s(i,x,s)|_{x=s-} = 0 \qquad \text{(normal reflection)}, \tag{3.14}$$

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

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

$$V'_x(i,x,s)|_{x=f(i,s)-} = 0 \qquad \text{for } f(i,s) \leq g(i,s) \qquad \text{(smooth fit)}, \tag{3.17}$$

$$V'_x(i,x,s)|_{x=g(i,s)+} = 0 \qquad \text{for } f(i,s) \leq g(i,s) \qquad \text{(smooth fit)}, \tag{3.18}$$

where $\mathbb{L}_X$ 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 [18].

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

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

for $i \le x \le s$ in $(-1,1)$ upon assuming that $\mathsf{E}_{i,x,s}\tau_{f,g} < \infty$ with candidate surfaces f and g to be specified below. Suppose that $f(i,s) \le s$ and consider $\rho_{i,f(i,s)} = \inf\{t \ge 0 | X_t \notin (i, f(i,s))\}$ under $\mathsf{P}_{i,x,s}$ with $i < x < f(i,s)$ given and fixed. Applying the strong Markov property of (I, X, S) at $\rho_{i,f(i,s)}$ and using (2.17)–(2.20) we find that

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

It follows from (3.20) that

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

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

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

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

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

Combining (3.21)–(3.23) we find that

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

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

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

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

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

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 $\rho_{g(i,s),s} = \inf\{t \ge 0 | X_t \notin (g(i,s),s)\}$ under $\mathsf{P}_{i,x,s}$ with $g(i,s) < x < s$ given and fixed. Applying the strong Markov property of (I,X,S) at $\rho_{g(i,s),s}$ and using (2.17)–(2.20) we find that

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

It follows from (3.27) that

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

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

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

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

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

Combining (3.28)–(3.30) we find that

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

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

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

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

$$\begin{aligned} \frac{\partial g}{\partial s}(i,s) &= \frac{(\sigma^2/2)(g(i,s))L'(g(i,s))}{c(i,g(i,s),s)[L(s) - L(g(i,s))]} \\ &\quad \times \bigg[1 + \int_{g(i,s)}^{s} \frac{\partial c}{\partial s}(i,y,s)\big[L(s) - L(y)\big]m(dy)\bigg] \end{aligned} \tag{3.33}$$

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

Summarising the preceding considerations we can conclude that to each pair of the candidate surfaces f and g solving (3.6) and (3.7) there corresponds the function (3.25) and (3.32) on $C^-_{f,g} \cup C^+_{f,g}$ solving the free-boundary problem (3.12)–(3.18) on $C^-_{f,g} \cup C^+_{f,g}$ (this can be verified by direct differentiation) and admitting the probabilistic representation (3.19) on $C^-_{f,g} \cup C^+_{f,g}$ associated with the stopping time (3.8) when the latter has finite expectation [this will be formally proved for the surfaces of interest in (3.74) and (3.75) below].

The central question becomes how to select the optimal surfaces f and g among all admissible candidates solving (3.6) and (3.7). We will answer this question by invoking the superharmonic characterisation of the value function (see [33], Chapter 1) for the four-dimensional Markov process (I, X, S, A) where $A_t = \int_0^t c(I_s, X_s, S_s)\, ds$ for $t \geq 0$. 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. We ought to recall, however, that stopping time (3.8) needs to have finite expectation, and this will put a natural constraint on how small and large these solutions can be (this is a subtle point in the background of the argument).

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 equations (3.6) and (3.7) admit (locally) unique solutions. In particular, recalling that $(i, x, s) \mapsto c(i, x, s)$ is continuous we see from the structure of Φ and Ψ that equations (3.6) and (3.7) admit (locally) unique solutions since $x \mapsto \sigma^2(x)$ and $x \mapsto c(i, x, s)$ are (locally) Lipschitz.

To construct the minimal solution to (3.6) staying strictly above the lower diagonal d^s, we can proceed as follows; see Figure 2 above. 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 $n \geq 1$. 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 $n \neq m$, and hence

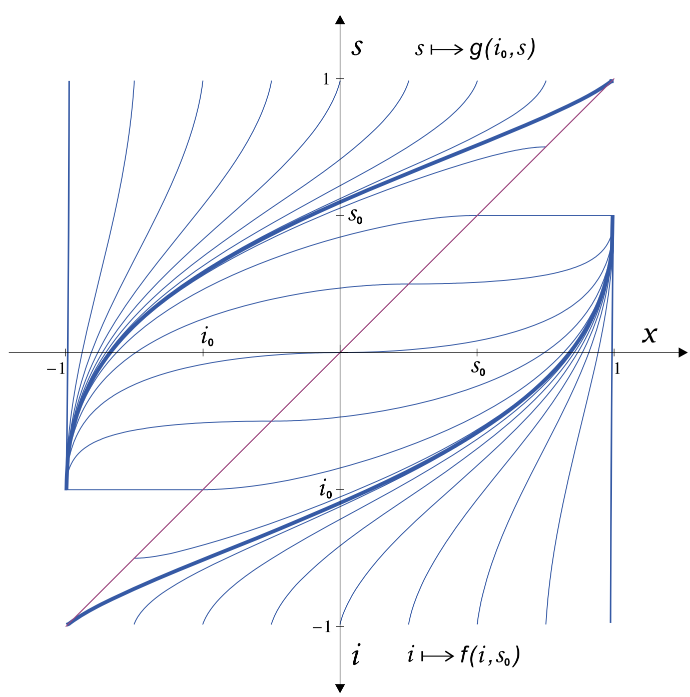

FIG. 2. *Smooth-fit solutions $i \mapsto f(i, s_0)$ and $s \mapsto g(i_0, s)$ to differential equations (3.6) and (3.7) for fixed s_0 and i_0, respectively. The minimal solution staying strictly above the lower diagonal (bold f line) and the maximal solution staying strictly below the upper diagonal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

we see that $(f_n)_{n\geq 1}$ 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 strictly above the lower diagonal. Since $i \mapsto c(i,x,s)$ is decreasing we see from (3.6) that

$$\begin{aligned} &i \mapsto f_n(i,s) \text{ and } i \mapsto f_*(i,s) \text{ are strictly increasing} \\ &\text{with } f_*(-1+,s) = -1 \end{aligned} \tag{3.34}$$

for $i < s$ in $(-1,1)$ and $n \geq 1$. Note further that the increase of $s \mapsto \frac{\partial c}{\partial i}(i,x,s)$ combined with the increase of $s \mapsto c(i,x,s)$ implies that $s \mapsto \Phi(i,s,f)$ is decreasing. Recalling that (3.6) is being solved forwards, this shows that

$$s \mapsto f_n(i,s) \text{ and } s \mapsto f_*(i,s) \text{ are decreasing} \tag{3.35}$$

for $i < s$ in $(-1,1)$ and $n \geq 1$; see Figure 3 below. Moreover, since $s \mapsto \frac{\partial c}{\partial i}(i,x,s)$ is (locally) Lipschitz we see that $s \mapsto \Phi(i,s,f)$ is (locally) Lipschitz from where we can easily deduce using Gronwall's inequality that

$$(i,s) \mapsto f_n(i,s) \text{ and } (i,s) \mapsto f_*(i,s) \text{ are continuous} \tag{3.36}$$

for $i < s$ in $(-1,1)$ and $n \geq 1$. To simplify the notation we will use the same symbol f below to denote the minimal solution f_* unless stated otherwise.

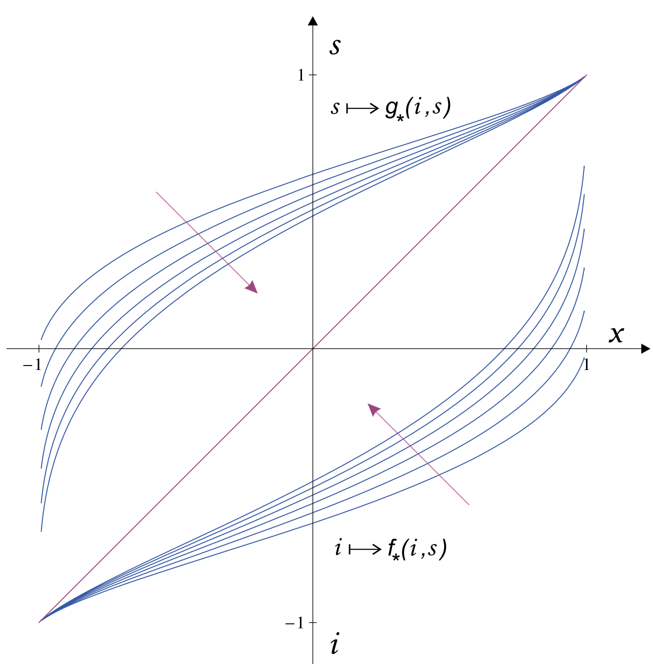

FIG. 3. *Movement and shape of sections $i \mapsto f_*(i,s)$ and $s \mapsto g_*(i,s)$ of the optimal surfaces f_* and g_* as the running maximum s increases and the running minimum i decreases, respectively.*

To construct the maximal solution to (3.7) staying strictly below the upper diagonal d_i, we can proceed similarly; see Figure 2 above. 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 $g_n(i, s_n) = s_n$ for $n \geq 1$. 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 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 strictly below the upper diagonal. Since $s \mapsto c(i, x, s)$ is increasing we see from (3.7) that

$$s \mapsto g_n(i, s) \text{ and } s \mapsto g_*(i, s) \text{ are strictly increasing with } g_*(i, 1-) = 1 \tag{3.37}$$

for $i < s$ in $(-1, 1)$ and $n \geq 1$. Note further that the increase of $i \mapsto \frac{\partial c}{\partial s}(i, x, s)$ combined with the decrease of $i \mapsto c(i, x, s)$ implies that $i \mapsto \Psi(i, s, f)$ is increasing. Recalling that (3.7) is being solved backwards, this shows that

$$i \mapsto g_n(i, s) \text{ and } i \mapsto g_*(i, s) \text{ are decreasing} \tag{3.38}$$

for $i < s$ in $(-1, 1)$ and $n \geq 1$; see Figure 3 above. Moreover, since $i \mapsto \frac{\partial c}{\partial s}(i, s)$ is (locally) Lipschitz we see that $i \mapsto \Psi(i, s, f)$ is (locally) Lipschitz from where we can easily deduce using Gronwall's inequality that

$$(i, s) \mapsto g_n(i, s) \text{ and } (i, s) \mapsto g_*(i, s) \text{ are continuous} \tag{3.39}$$

for $i < s$ in $(-1, 1)$ and $n \geq 1$. To simplify the notation we will use the same symbol g below to denote the maximal solution g_* unless stated otherwise.

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. The claims of the past two sentences will be formally verified in (3.74) and (3.75) below. 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 $C_{f,g}^0$*. This calculation is technically more complicated, and we will derive closed-form expressions for $V_{f,g}$ in terms of f and g on $C_{f,g}^0$ when $c(i, s) = c_2(s) - c_1(i) > 0$ 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 $V_{f,g}$ needs to satisfy (3.12)–(3.14) on $C^0_{f,g}$; see Remark 2 below. Recalling that a particular solution to $\mathbb{L}_X H = 1$ is given by

$$H(x) = \int_0^x [L(x) - L(y)] m(dy), \tag{3.40}$$

it follows from (3.12) that

$$V(i,x,s) = A(i,s)L(x) + B(i,s) + \big(c_2(s) - c_1(i)\big)H(x) \tag{3.41}$$

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

$$A'_i(i,s)L(i) + B'_i(i,s) - c'_1(i)H(i) = 0, \tag{3.42}$$

$$A'_s(i,s)L(s) + B'_s(i,s) + c'_2(s)H(s) = 0. \tag{3.43}$$

Differentiating (3.42) with respect to s and (3.43) 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

$$A(i,s) = a_1(i) + a_2(s) \quad \text{and} \quad B(i,s) = b_1(i) + b_2(s) \tag{3.44}$$

for some a_i and b_i to be found when $i = 1, 2$. Inserting this back into (3.41)–(3.43) we obtain

$$\begin{aligned} V(i,x,s) = {} & \big(a_1(i) + a_2(s)\big)L(x) \\ & + b_1(i) + b_2(s) + \big(c_2(s) - c_1(i)\big)H(x), \end{aligned} \tag{3.45}$$

$$a'_1(i)L(i) + b'_1(i) - c'_1(i)H(i) = 0, \tag{3.46}$$

$$a'_2(s)L(s) + b'_2(s) + c'_2(s)H(s) = 0 \tag{3.47}$$

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

$$f\big(i(s), s\big) = g\big(i(s), s\big) \quad \text{and} \quad f\big(i, s(i)\big) = g\big(i, s(i)\big). \tag{3.48}$$

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 above. More formally, the functions can be defined as follows:

$$i(s) = \big(f(\cdot,s) - g(\cdot,s)\big)^{-1}(0) \quad \text{and} \quad s(i) = \big(f(i,\cdot) - g(i,\cdot)\big)^{-1}(0) \tag{3.49}$$

for $f(i,s) > g(i,s)$. [Recall from (3.34) and (3.37) 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) 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:

$$
\begin{aligned}
V\big(i(s),x,s\big) &= s - i(s) \\
&\quad + \big[c_2(s) - c_1\big(i(s)\big)\big]\bigg[H(x) - L(x)\int_0^{f(i(s),s)} m(dy) \\
&\qquad + \int_0^{f(i(s),s)} L(y)m(dy)\bigg]
\end{aligned}
\tag{3.50}
$$

for $x \le f(i(s),s)$. Comparing (3.50) with (3.45), we can conclude that

$$
a_1\big(i(s)\big) + a_2(s) = -\big[c_2(s) - c_1\big(i(s)\big)\big]\int_0^{f(i(s),s)} m(dy), \tag{3.51}
$$

$$
b_1\big(i(s)\big) + b_2(s) = s - i(s) + \big[c_2(s) - c_1\big(i(s)\big)\big]\int_0^{f(i(s),s)} L(y)m(dy). \tag{3.52}
$$

Using (3.46)–(3.47) and (3.51)–(3.52) we can calculate $a_2'(s)$. First, by (3.47) we can express $a_2'(s)$ in terms of $b_2'(s)$. Second, by (3.52) we can express $b_2'(s)$ in terms of $b_1'(i(s))$. Third, by (3.46) we can express $b_1'(i(s))$ in terms of $a_1'(i(s))$. Fourth, by (3.51) 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

$$
\begin{aligned}
a_2'(s) &= -\frac{1}{L(s) - L(i(s))} \\
&\quad \times \bigg[\frac{f_s'(i(s),s)}{f_i'(i(s),s)}\bigg[1 + c_1'\big(i(s)\big)\int_{i(s)}^{f(i(s),s)} \big[L(y) - L\big(i(s)\big)\big]m(dy)\bigg] \\
&\qquad + 1 + c_2'(s)\bigg[H(s) + \int_0^{f(i(s),s)} \big[L(y) - L\big(i(s)\big)\big]m(dy)\bigg]\bigg].
\end{aligned}
\tag{3.53}
$$

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:

$$
\begin{aligned}
V\big(i,x,s(i)\big) &= s(i) - i \\
&\quad + \big[c_2\big(s(i)\big) - c_1(i)\big] \\
&\quad \times \bigg[H(x) - L(x)\int_0^{g(i,s(i))} m(dy) + \int_0^{g(i,s(i))} L(y)m(dy)\bigg]
\end{aligned}
\tag{3.54}
$$

for $x \geq g(i,s(i))$. Comparing (3.54) with (3.45) we can conclude that

$$
a_1(i) + a_2\big(s(i)\big) = -\big[c_2\big(s(i)\big) - c_1(i)\big]\int_0^{g(i,s(i))} m(dy), \tag{3.55}
$$

$$
b_1(i) + b_2\big(s(i)\big) = s(i) - i + \big[c_2\big(s(i)\big) - c_1(i)\big]\int_0^{g(i,s(i))} L(y)m(dy). \tag{3.56}
$$

Using (3.46)–(3.47) and (3.55)–(3.56) we can calculate $a'_1(i)$. First, by (3.46) we can express $a'_1(i)$ in terms of $b'_1(i)$. Second, by (3.56) we can express $b'_1(i)$ in terms of $b'_1(s(i))$. Third, by (3.47) we can express $b'_2(s(i))$ in terms of $a'_2(s(i))$. Fourth, by (3.55) 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

$$
\begin{aligned}
a'_1(i) &= -\frac{1}{L(s(i)) - L(i)} \\
&\quad \times \bigg[\frac{g'_i(i,s(i))}{g'_s(i,s(i))}\bigg[1 + c'_2\big(s(i)\big)\int_{g(i,s(i))}^{s(i)} \big[L\big(s(i)\big) - L(y)\big]m(dy)\bigg] \\
&\qquad + 1 + c'_1(i)\bigg[H(i) - \int_0^{g(i,s(i))} \big[L\big(s(i)\big) - L(y)\big]m(dy)\bigg]\bigg].
\end{aligned}
\tag{3.57}
$$

We can now determine A and B in (3.41) using the closed-form expressions obtained. First, note that by (3.51) we find that

$$
\begin{aligned}
A(i,s) &= A\big(i(s),s\big) + \int_{i(s)}^{i} A'_u(u,s)\,du \\
&= a_1\big(i(s)\big) + a_2(s) + \int_{i(s)}^{i} a'_1(u)\,du \\
&= -\big[c_2(s) - c_1\big(i(s)\big)\big]\int_0^{f(i(s),s)} m(dy) \\
&\quad + \int_{i(s)}^{i} a'_1(u)\,du,
\end{aligned}
\tag{3.58}
$$

where $a_1'(u)$ is given by (3.57) above. Note also that by (3.55) we find that

$$
\begin{aligned}
A(i,s) &= A\big(i,s(i)\big) - \int_s^{s(i)} A_v'(i,v)\,dv \\
&= a_1(i) + a_2\big(s(i)\big) - \int_s^{s(i)} a_2'(v)\,dv \qquad (3.59)\\
&= -\big[c_2\big(s(i)\big) - c_1(i)\big] \int_0^{g(i,s(i))} m(dy) - \int_s^{s(i)} a_2'(v)\,dv,
\end{aligned}
$$

where $a_2'(v)$ is given by (3.53) above. Second, observe that (3.46) and (3.47) yield

$$b_1'(i) = -a_1'(i)L(i) + c_1'(i)H(i), \tag{3.60}$$

$$b_2'(s) = -a_2'(s)L(s) - c_2'(s)H(s), \tag{3.61}$$

where $a_1'(i)$ and $a_2'(s)$ are given by (3.57) and (3.53) above. Note that by (3.52) we find that

$$
\begin{aligned}
B(i,s) &= B\big(i(s),s\big) + \int_{i(s)}^{i} B_u'(u,s)\,du \\
&= b_1\big(i(s)\big) + b_2(s) + \int_{i(s)}^{i} b_1'(u)\,du \qquad (3.62)\\
&= s - i(s) + \big[c_2(s) - c_1\big(i(s)\big)\big] \int_0^{f(i(s),s)} L(y)m(dy) + \int_{i(s)}^{i} b_1'(u)\,du,
\end{aligned}
$$

where $b_1'(u)$ is given by (3.60) above. Note also that by (3.56) we find that

$$
\begin{aligned}
B(i,s) &= B\big(i,s(i)\big) - \int_s^{s(i)} B_v'(i,v)\,dv \\
&= b_1(i) + b_2\big(s(i)\big) - \int_s^{s(i)} b_2'(v)\,dv \qquad (3.63)\\
&= s - i(s) + \big[c_2\big(s(i)\big) - c_1(i)\big] \int_0^{g(i,s(i))} L(y)m(dy) - \int_s^{s(i)} b_2'(v)\,dv,
\end{aligned}
$$

where $b_2'(v)$ is given by (3.61) above.

Finally, inserting (3.58), (3.62) and (3.59), (3.63) into (3.41) we, respectively, obtain the following two closed-form expressions:

$$
\begin{aligned}
V(i,x,s) = s - i(s) &+ \big[c_2(s) - c_1\big(i(s)\big)\big] \int_0^{f(i(s),s)} [L(y) - L(x)]m(dy) \\
&+ \big[c_2(s) - c_1(i)\big]H(x) \qquad (3.64)\\
&+ \int_{i(s)}^{i} \big([L(x) - L(u)]a_1'(u) + c_1'(u)H(u)\big)\,du,
\end{aligned}
$$

$$
\begin{aligned}
V(i,x,s) = s(i) - i + \big[c_2\big(s(i)\big) - c_1(i)\big] \int_0^{g(i,s(i))} \big[L(y) - L(x)\big] m(dy) \\
+ \big[c_2(s) - c_1(i)\big] H(x) \\
+ \int_s^{s(i)} \big(\big[L(v) - L(x)\big] a_2'(v) + c_2'(v) H(v)\big)\, dv
\end{aligned}
\tag{3.65}
$$

for $f(i,s) > g(i,s)$ where $a_1'(u)$ and $a_2'(v)$ are given by (3.57) and (3.53) above. A formal verification of (3.64) and (3.65) can be easily done by Itô's formula once we derive the optimality in the next step; see Remark 2 below. Observe that if $f(i,s) = g(i,s)$, then $i(s) = i$ and $s(i) = s$ so that the second integral in both (3.64) and (3.65) 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 strictly 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 with $f(-1+,s) \in [-1,1)$, and let $s \mapsto g(i,s)$ be any solution to (3.7) satisfying $g(i,s) < s$ for all s with $g(i,1-) \in (-1,1]$. Consider the function $V_{f,g}$ defined by (3.25) and (3.32) on $C_{f,g}^- \cup C_{f,g}^+$, and set $V_{f,g}(i,x,s) = s - i$ on $D_{f,g}$ which denotes the complement of $C_{f,g}$. Then the same arguments as in (3.35) and (3.38) above show that $s \mapsto f(i,s)$ and $i \mapsto g(i,s)$ are decreasing. This implies that after starting 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 $C_{f,g}^0$). Fix any point (i,x,s) such that $f(i,s) \le g(i,s)$ with $i \le x \le s$. Note that (i,x,s) belongs to $C_{f,g}^- \cup C_{f,g}^+ \cup D_{f,g}$, and consider the motion of (I,X,S) under $\mathsf{P}_{i,x,s}$. Recall that $V_{f,g}$ solves the free boundary problem (3.12)–(3.18) on $C_{f,g}^- \cup C_{f,g}^+$. 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) (note that these conditions can fail for the second derivatives) reduces to standard Itô's formula and gives

$$
\begin{aligned}
&V_{f,g}(I_t, X_t, S_t) \\
&= V_{f,g}(i,x,s) + \int_0^t \frac{\partial V_{f,g}}{\partial i}(I_s, X_s, S_s)\, dI_s + \int_0^t \frac{\partial V_{f,g}}{\partial x}(I_s, X_s, S_s)\, dX_s \\
&\quad + \int_0^t \frac{\partial V_{f,g}}{\partial s}(I_s, X_s, S_s)\, dS_s + \frac{1}{2}\int_0^t \frac{\partial^2 V_{f,g}}{\partial x^2}(I_s, X_s, S_s)\, d\langle X, X\rangle_s \\
&= V_{f,g}(i,x,s) + \int_0^t \sigma(X_s) \frac{\partial V_{f,g}}{\partial x}(I_s, X_s, S_s)\, dB_s \\
&\quad + \int_0^t (\mathbb{L}_X V_{f,g})(I_s, X_s, S_s)\, ds,
\end{aligned}
\tag{3.66}
$$

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 $M = (M_t)_{t\geq 0}$ defined by

$$M_t = \int_0^t \sigma(X_s) \frac{\partial V_{f,g}}{\partial x}(I_s, X_s, S_s)\, dB_s \tag{3.67}$$

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

$$P_t = \int_0^t c(I_s, X_t, S_s) 1\big(f(I_s, S_s) \leq X_s \leq g(I_s, X_s)\big)\, ds \tag{3.68}$$

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 (3.66) can be rewritten as follows:

$$V_{f,g}(I_t, X_t, S_t) - \int_0^t c(I_s, X_s, S_s)\, ds = V_{f,g}(i, x, s) + M_t - P_t. \tag{3.69}$$

From this representation we see that the process

$$V_{f,g}(I_t, X_t, S_t) - \int_0^t c(I_s, X_s, S_s)\, ds$$

is a local supermartingale for $t \geq 0$.

Let τ be any stopping time of X. Choose a localisation sequence $(\sigma_n)_{n\geq 1}$ of bounded stopping times for M. From (3.25) and (3.32) we see that $V_{f,g}(i, x, s) \geq s - i$ for all $(i, x, s) \in C^-_{f,g} \cup C^+_{f,g} \cup D_{f,g}$. Recalling that the process (I, X, S) remains in the latter set, we can conclude from (3.69) using the optional sampling theorem that

$$\begin{aligned} &\mathsf{E}_{i,x,s}\left[S_{\tau\wedge\sigma_n} - I_{\tau\wedge\sigma_n} - \int_0^{\tau\wedge\sigma_n} c(I_s, X_s, S_s)\, dt\right] \\ &\quad \leq \mathsf{E}_{i,x,s}\left[V_{f,g}(I_{\tau\wedge\sigma_n}, X_{\tau\wedge\sigma_n}, S_{\tau\wedge\sigma_n}) - \int_0^{\tau\wedge\sigma_n} c(I_s, X_s, S_s)\, dt\right] \\ &\quad \leq V_{f,g}(i, x, s) + \mathsf{E}_{i,x,s}(M_{\tau\wedge\sigma_n}) = V_{f,g}(i, x, s) \end{aligned} \tag{3.70}$$

for all $(i, x, s) \in C^-_{f,g} \cup C^+_{f,g} \cup D_{f,g}$ and all $n \geq 1$. Letting $n \to \infty$ and using the monotone convergence theorem we find that

$$\mathsf{E}_{i,x,s}\left[S_\tau - I_\tau - \int_0^\tau c(I_s, X_s, S_s)\, dt\right] \leq V_{f,g}(i, x, s) \tag{3.71}$$

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

$$V(i, x, s) \leq \inf_{f,g} V_{f,g}(i, x, s) = V_{f_*,g_*}(i, x, s) \tag{3.72}$$

for all $(i, x, s) \in C^-_{f,g} \cup C^+_{f,g} \cup D_{f,g}$ where f_* denotes the minimal solution to (3.6) staying strictly above the lower diagonal, and g_* denotes the maximal solution to (3.7) staying strictly below the upper diagonal. Recalling that $f \mapsto V_{f,g}$

is increasing and $g \mapsto V_{f,g}$ is decreasing when $f \le g$, we see that the infimum in (3.72) 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 (3.71) applies, we see that (3.72) 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 $n \ge 1$. From these considerations and (3.72) in particular, it follows that the only possible candidates for the optimal stopping boundary are the minimal and maximal solution f_* and g_*. Note that (3.70) also implies that

$$(3.73) \qquad \mathsf{E}_{i,x,s}\left[V_{f,g}(I_\tau, X_\tau, S_\tau) - \int_0^\tau c(I_s, X_s, S_s)\,dt\right] \le V_{f,g}(i,x,s)$$

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 $A_t = \int_0^t c(I_s,X_s,S_s)\,ds$ for $t \ge 0$. Recalling that $f \mapsto V_{f,g}$ is increasing and $g \mapsto V_{f,g}$ is decreasing when $f \le g$, and that $V_{f,g}(i,x,s) \ge s - i$ 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 strictly above the lower diagonal and the maximal solution g_* staying strictly 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; see also Remark 3 below.

To prove that f_* and g_* are optimal on $C^-_{f_*,g_*} \cup C^+_{f_*,g_*} \cup D_{f_*,g_*}$, 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) and (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 $n \ge 1$. Recall that V_{f_n,g_n} solves the free-boundary problem (3.12)–(3.18) on $C^-_{f_n,g_n} \cup C^+_{f_n,g_n}$ for $n \ge 1$. Fix any (i,x,s) in $C^-_{f_*,g_*} \cup C^+_{f_*,g_*} \cup D_{f_*,g_*}$, and note that this (i,x,s) belongs to $C^-_{f_n,g_n} \cup C^+_{f_n,g_n} \cup D_{f_n,g_n}$ since $f_n \le f_*$ and $g_* \le g_n$ for every $n \ge 1$. The same arguments as above yield the formula (3.66) with f_n and g_n in place of f and g for $n \ge 1$. Since σ and $\partial V_{f_n,g_n}/\partial x$ are bounded on $C^-_{f_n,g_n} \cup C^+_{f_n,g_n}$, we see that $(M_{t \wedge \tau_{f_n,g_n}})_{t \ge 0}$ defined by (3.67) with f_n and g_n in place of f and g is a martingale under $\mathsf{P}_{i,x,s}$. The latter conclusion follows from the fact that $\tau_{f_n,g_n} \le \rho_{i_n,s_n}$ with $\mathsf{E}_{i,x,s}\rho_{i_n,s_n} < \infty$ implying also that $\mathsf{E}_{i,x,s}\int_0^{\tau_{f_n,g_n}} c(I_s,X_s,S_s)\,dt < \infty$ for $n \ge 1$. Since the process P defined by (3.68) with f_n and g_n in place of f and g satisfies $P_{\tau_{f_n,g_n}} = 0$, it follows from (3.69) using (3.15) and (3.16) that

$$(3.74) \qquad V_{f_n,g_n}(i,x,s) = \mathsf{E}_{i,x,s}\left[S_{\tau_{f_n,g_n}} - I_{\tau_{f_n,g_n}} - \int_0^{\tau_{f_n,g_n}} c(I_s,X_s,S_s)\,dt\right]$$

for all $i \le x \le s$ such that $f_n(i,s) \le g_n(i,s)$ with $n \ge 1$. Letting $n \to \infty$ in (3.74), noting that $\tau_{f_n,g_n} \uparrow \tau_{f_*,g_*}$ (since $[-1,1]^3$ is compact), and using the monotone convergence theorem (recalling that $S_{\tau_{f_*,g_*}} - I_{\tau_{f_*,g_*}}$ is bounded by 2 and therefore integrable) we find that

$$V_{f_*,g_*}(i,x,s) = \mathsf{E}_{i,x,s}\left[S_{\tau_{f_*,g_*}} - I_{\tau_{f_*,g_*}} - \int_0^{\tau_{f_*,g_*}} c(I_s, X_s, S_s)\,dt\right] \tag{3.75}$$

for all $i \le x \le s$ such that $f_*(i,s) \le g_*(i,s)$. This shows that we have equality in (3.72) 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 $C^0_{f_*,g_*}$, that is, when $f_*(i,s) > g_*(i,s)$ for some (i,s) given and fixed, one could attempt to apply similar arguments to those in (3.70) above. For this, however, we would need to know that $V_{f,g}(i,x,s) \ge s - i$ not only for $f(i,s) \le g(i,s)$ as follows from the closed-form expressions (3.25) and (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 $c(i,s) = c_2(s) - c_1(i)$, we see from (3.64) and (3.65) 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 (3.53) and (3.57) 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 $f_*(i_0, s_0) > g_*(i_0, s_0)$ with $i_0 < x_0 < s_0$ and fix any $d_0 \in (i_0 \vee g_*(i_0, s_0), s_0 \wedge f_*(i_0, s_0)) \setminus \{x_0\}$. 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 strictly 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 $C^-_{f_d,g_d}$ if $x_0 < d_0$ or $C^+_{f_d,g_d}$ if $x_0 > d_0$, 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 (3.74) also show that

$$\begin{aligned} &V_{f_d,g_d}(i_0, x_0, s_0) \\ &\quad = \mathsf{E}_{i_0,x_0,s_0}\left[S_{\tau_{f_d,g_d}} - I_{\tau_{f_d,g_d}} - \int_0^{\tau_{f_d,g_d}} c(I_s, X_s, S_s)\,dt\right], \end{aligned} \tag{3.76}$$

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 (3.76) 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. $\square$

We conclude this section with a few remarks on the preceding result and proof.

REMARK 1. To describe the nature of the optimal stopping time τ_{f_*,g_*} from (3.5), assume that the process (I, X, S) starts at $(0, 0, 0)$. Then due to $g_*(0, 0) < 0 < f_*(0, 0)$ we see that it is not optimal to stop at once so that $t \mapsto I_t$ and $t \mapsto S_t$ will gradually start to decrease and increase whenever $t \mapsto X_t$ returns to the lower and upper diagonal, respectively. Due to (3.34)–(3.35) and (3.37)–(3.38) we see that $t \mapsto f_*(I_t, S_t)$ is decreasing and $t \mapsto g_*(I_t, S_t)$ is increasing. Since $I_t \downarrow -1$ and/or $S_t \uparrow 1$ as $t \uparrow \infty$ we see from (3.34) and (3.37) that the two sample paths $t \mapsto f_*(I_t, S_t)$ and $t \mapsto g_*(I_t, S_t)$ will meet at some random time which coincides with the first exit time of (I, X, S) from the set $C^0_{f_*,g_*}$ defined in (3.9). This can only happen either through the lower diagonal (when X is equal to I) or through the upper diagonal (when X is equal to S). In the former case the process (I, X, S) enters the set $C^-_{f_*,g_*}$ defined in (3.10) and in the latter case the process (I, X, S) enters the set $C^+_{f_*,g_*}$ defined in (3.11). After entering either $C^-_{f_*,g_*}$ or $C^+_{f_*,g_*}$ the process (I, X, S) remains in the same set until the first hitting of X to either $f(I, S)$ from below or $g(I, S)$ from above happens, respectively. This moment defines the optimal stopping time τ_{f_*,g_*}. Note that from the optimality derived in the proof of Theorem 1 [recall (3.74) and (3.75) in particular] we see that τ_{f_*,g_*} has finite expectation (since otherwise the value function would be equal to $-\infty$ and as such τ_{f_*,g_*} could not be optimal). Note that the analogous description of τ_{f_*,g_*} also holds for any starting point (i, x, s) of (I, X, S) in the state space. After starting in $C^0_{f_*,g_*}$ the process (I, X, S) enters either $C^-_{f_*,g_*}$ or $C^+_{f_*,g_*}$ to remain in the same set until τ_{f_*,g_*} happens. The latter fact also holds if (I, X, S) starts in either $C^-_{f_*,g_*}$ or $C^+_{f_*,g_*}$ directly. To visualise these movements, see Figure 1 above and note that i_0 and s_0 mark the borderline levels between $C^0_{f_*,g_*}$ and $C^-_{f_*,g_*} \cup C^+_{f_*,g_*}$ as described above.

REMARK 2. Although we do not make use of this fact in the proof of the optimality above, we note that in addition to the closed-form expressions (3.25) and (3.32) on $C^-_{f,g}$ and $C^+_{f,g}$, respectively, the probabilistic representation (3.19) itself can also be used to define the function $V_{f,g}$ on $C^0_{f,g}$ when the stopping time $\tau_{f,g}$ from (3.8) has finite expectation, and the resulting function will solve the free boundary problem (3.12)–(3.18) on $C_{f,g}$ for the surfaces f and g constructed in the proof above (those hitting the lower/upper diagonal at a single point and the minimal/maximal solutions staying above/below the lower/upper diagonal). Indeed, due to the monotonicity properties of f and g derived above, we see that after starting in $C^0_{f,g}$, the process (I, X, S) enters either the set $C^-_{f,g}$ or the set $C^+_{f,g}$ through the boundary $f = g$ to stay in the same set until $\tau_{f,g}$ happens. This shows that defining the function $V_{f,g}$ by (3.19) on $C^0_{f,g}$ corresponds to solving the Dirichlet problem stochastically where the value at the boundary $f = g$ is set to be either (3.25) at the lower diagonal or (3.32) at the upper diagonal, respectively. For standard arguments how this can be done including how the required smoothness of $V_{f,g}$ on $C^0_{f,g}$ can be derived; see, for example, [33], Sections 7.1–7.3.

REMARK 3. In addition to the facts used in the proof above 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].

REMARK 4. A closer look into the proof above indicates that the arguments developed and/or used should be applicable in more general settings of the optimal stopping problem (3.4) and its relatives. As stated above it is not essential that the state space of the diffusion process X equals $(-1, 1)$, and the result and methodology of Theorem 1 should be valid for more general state spaces (including $\mathbb{R}$ and $\mathbb{R}_+$ in particular). In this case we may need to take the supremum in (3.4) over all stopping times such that the expectation of the integral is finite, and although the stopping time τ_{f_*,g_*} may not belong to this class in some particular examples [so that the right-hand side of (3.4) may not even be well-defined], this stopping time should be approximately optimal in the sense that the approximate stopping times τ_{f_n,g_n} yield the value (3.4) in the limit as $n \to \infty$. These extensions also include various boundary behaviour of the process X at the endpoints of the state space (e.g., 0 when the state space equals $\mathbb{R}_+$). We leave precise formulations of these statements and proofs as informal conjectures open for future developments. We emphasise that these questions are best studied through examples, and each particular example may have specifics which are difficult to cover by any meta-theorem in advance. Omitting further details 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

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

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

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

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

for $x \in (-1, 1)$. By Theorem 1 we 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

$$\begin{aligned} &\frac{\partial f}{\partial i}(i, s) \\ &\quad = \frac{2\varphi^2(\Phi^{-1}((f(i,s)+1)/2)) \exp(1/2(\Phi^{-1}((f(i,s)+1)/2))^2)}{c(s-i) \int_i^{f(i,s)} \exp(1/2(\Phi^{-1}((y+1)/2))^2)\, dy} \\ &\qquad \times \Big[1 + c \int_i^{f(i,s)} \frac{\int_i^y \exp(1/2(\Phi^{-1}((z+1)/2))^2)\, dz}{2\varphi^2(\Phi^{-1}((y+1)/2)) \exp(1/2(\Phi^{-1}((y+1)/2))^2)}\, dy\Big], \end{aligned} \tag{4.5}$$

$$\begin{aligned} &\frac{\partial g}{\partial s}(i, s) \\ &\quad = \frac{2\varphi^2(\Phi^{-1}((g(i,s)+1)/2)) \exp(1/2(\Phi^{-1}((g(i,s)+1)/2))^2)}{c(s-i) \int_{g(i,s)}^{s} \exp(1/2(\Phi^{-1}((y+1)/2))^2)\, dy} \\ &\qquad \times \Big[1 + c \int_{g(i,s)}^{s} \frac{\int_y^s \exp(1/2(\Phi^{-1}((z+1)/2))^2)\, dz}{2\varphi^2(\Phi^{-1}((y+1)/2)) \exp(1/2(\Phi^{-1}((y+1)/2))^2)}\, dy\Big] \end{aligned} \tag{4.6}$$

staying strictly above the lower diagonal d^s and strictly below the upper diagonal d_i for $i < s$ in $(-1, 1)$, respectively. 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 list of examples can be continued by considering various diffusion processes Z and hidden targets ℓ. This leads to a classification of the laws of ℓ against the laws of Z (through the drift and diffusion coefficient) in terms of the optimal surfaces derived in Theorem 1. This classification can be used for calibration against observed performance (where either of the two laws is taken initially to be known, e.g.).

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 briefly 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

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

From (3.48) we see that $i(s) = s - \frac{1}{c}$ and $s(i) = i + \frac{1}{c}$. Since $f_s' \equiv 0$ we see from (3.53) that $a_2'(s) \equiv -c$. Inserting this into (3.65) we find that $V(0,0,0) = \frac{3}{4c}$; note that unboundedness of B presents no difficulty since the optimal stopping time has finite expectation. This shows that for any stopping time τ of B (with finite expectation) we have

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

Taking the infimum over all $c > 0$ we obtain the result of [10],

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

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. MR1269985

[2] BARRON, E. N. and ISHII, H. (1989). The Bellman equation for minimizing the maximum cost. *Nonlinear Anal.* **13** 1067–1090. MR1013311

[3] BERNYK, V., DALANG, R. C. and PESKIR, G. (2011). Predicting the ultimate supremum of a stable Lévy process with no negative jumps. *Ann. Probab.* **39** 2385–2423. MR2932671

[4] COHEN, A. (2010). Examples of optimal prediction in the infinite horizon case. *Statist. Probab. Lett.* **80** 950–957. MR2638963

[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. MR2462552

[6] DU TOIT, J. and PESKIR, G. (2007). The trap of complacency in predicting the maximum. *Ann. Probab.* **35** 340–365. MR2303953

[7] DU TOIT, J. and PESKIR, G. (2008). Predicting the time of the ultimate maximum for Brownian motion with drift. In *Mathematical Control Theory and Finance* 95–112. Springer, Berlin. MR2484106

[8] DU TOIT, J. and PESKIR, G. (2009). Selling a stock at the ultimate maximum. *Ann. Appl. Probab.* **19** 983–1014. MR2537196

[9] DU TOIT, J., PESKIR, G. and SHIRYAEV, A. N. (2008). Predicting the last zero of Brownian motion with drift. *Stochastics* **80** 229–245. MR2402166

[10] DUBINS, L. E., GILAT, D. and MEILIJSON, I. (2009). On the expected diameter of an L_2-bounded martingale. *Ann. Probab.* **37** 393–402. MR2489169

[11] DUBINS, L. E. and SCHWARZ, G. (1988). A sharp inequality for sub-martingales and stopping-times. *Astérisque* 129–145. MR0976216

[12] 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.

[13] ELIE, R. and ESPINOSA, G.-E. (2014). Optimal selling rules for monetary invariant criteria: Tracking the maximum of a portfolio with negative drift. *Math. Finance*. To appear.

[14] ESPINOSA, G.-E. and TOUZI, N. (2012). Detecting the maximum of a scalar diffusion with negative drift. *SIAM J. Control Optim.* **50** 2543–2572. MR3022078

[15] GAPEEV, P. V. (2006). Discounted optimal stopping for maxima in diffusion models with finite horizon. *Electron. J. Probab.* **11** 1031–1048 (electronic). MR2268535

[16] GAPEEV, P. V. (2007). Discounted optimal stopping for maxima of some jump-diffusion processes. *J. Appl. Probab.* **44** 713–731. MR2355587

[17] GLOVER, K., HULLEY, H. and PESKIR, G. (2013). Three-dimensional Brownian motion and the golden ratio rule. *Ann. Appl. Probab.* **23** 895–922. MR3076673

[18] 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.

[19] GRAVERSEN, S. E., PESKIR, G. and SHIRYAEV, A. N. (2001). Stopping Brownian motion without anticipation as close as possible to its ultimate maximum. *Theory Probab. Appl.* **45** 41–50.

[20] GUO, X. and ZERVOS, M. (2010). π options. *Stochastic Process. Appl.* **120** 1033–1059. MR2639737

[21] HEINRICHER, A. C. and STOCKBRIDGE, R. H. (1991). Optimal control of the running max. *SIAM J. Control Optim.* **29** 936–953. MR1111669

[22] HOBSON, D. (2007). Optimal stopping of the maximum process: A converse to the results of Peskir. *Stochastics* **79** 85–102. MR2290399

[23] JACKA, S. D. (1991). Optimal stopping and best constants for Doob-like inequalities. I. The case $p = 1$. *Ann. Probab.* **19** 1798–1821. MR1127729

[24] OBŁÓJ, J. (2004). The Skorokhod embedding problem and its offspring. *Probab. Surv.* **1** 321–390. MR2068476

[25] OBŁÓJ, J. (2007). The maximality principle revisited: On certain optimal stopping problems. In *Séminaire de Probabilités XL. Lecture Notes in Math.* **1899** 309–328. Springer, Berlin. MR2409013

[26] PEDERSEN, J. L. (2000). Discounted optimal stopping problems for the maximum process. *J. Appl. Probab.* **37** 972–983. MR1808862

[27] PEDERSEN, J. L. (2003). Optimal prediction of the ultimate maximum of Brownian motion. *Stoch. Stoch. Rep.* **75** 205–219. MR1994906

[28] PESKIR, G. (1998). Optimal stopping of the maximum process: The maximality principle. *Ann. Probab.* **26** 1614–1640. MR1675047

[29] PESKIR, G. (2005). Maximum process problems in optimal control theory. *J. Appl. Math. Stoch. Anal.* **1** 77–88. MR2140329

[30] PESKIR, G. (2007). A change-of-variable formula with local time on surfaces. In *Séminaire de Probabilités XL. Lecture Notes in Math.* **1899** 69–96. Springer, Berlin. MR2408999

[31] PESKIR, G. (2012). A duality principle for the Legendre transform. *J. Convex Anal*. **19** 609–630. MR3013752
[32] PESKIR, G. (2012). Optimal detection of a hidden target: The median rule. *Stochastic Process. Appl*. **122** 2249–2263. MR2921979
[33] PESKIR, G. and SHIRYAEV, A. (2006). *Optimal Stopping and Free-Boundary Problems*. *Lectures in Mathematics ETH Zürich*. Birkhäuser, Basel. MR2256030
[34] SHIRYAEV, A. N. (1961). The problem of the most rapid detection of a disturbance of a stationary regime. *Sov. 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. In *Mathematical Finance—Bachelier Congress*, 2000 (*Paris*). *Springer Finance* 487–521. Springer, Berlin. MR1960576
[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] SHIRYAEV, A. N. and NOVIKOV, A. A. (2008). On a stochastic version of the trading rule "buy and hold". *Statist. Decisions* **26** 289–302. MR2554513
[39] URUSOV, M. A. (2004). On a property of the moment at which Brownian motion attains its maximum and some optimal stopping problems. *Theory Probab. Appl*. **49** 169–176.
[40] ZHITLUKHIN, M. (2009). A maximal inequality for skew Brownian motion. *Statist. Decisions* **27** 261–280. MR2732522

SCHOOL OF MATHEMATICS
UNIVERSITY OF MANCHESTER
OXFORD ROAD
MANCHESTER M13 9PL
UNITED KINGDOM
E-MAIL: goran@maths.man.ac.uk