An Ergodic Impulse Control Model with Applications Including the Control of an Exchange Rate

DIMITRIS MELAS† and MIHAIL ZERVOS‡
Department of Mathematics
London School of Economics
Houghton Street
London WC2A 2AE, UK

October 15, 2007

Abstract

We address the problem of controlling a general one-dimensional Itô diffusion within an externally specified bounded interval \([A, B]\) by means of an impulse control process. We minimise a long-term average criterion that penalises deviations of the state process from a given nominal point within this region as well as the use of impulsive control effort. We solve the resulting optimisation problem and we provide an explicit optimal control strategy under general assumptions. The model that we study is motivated by several applications, including the problem of determining an optimal central bank intervention strategy aiming at the control of an exchange rate or an inflation rate, as well as the problem of designing optimal contribution policies in defined benefit pension plans.

1 Introduction

The dynamics of the stochastic system that we study are described by the controlled, one-dimensional Itô diffusion

\[
dX_t = b(X_t) \, dt + \sigma(X_t) \, dW_t, \quad X_0 = x \in [A, B],
\]

where \(W\) is a standard, one-dimensional Brownian motion, and \(-\infty < A < B < \infty\). As long as the state process \(X\) is inside the open interval \([A, B]\), the controller should take no action.

---

*Research supported by EPSRC grant no. GR/S22998/01
†E-mail: d.melas@lse.ac.uk
‡Corresponding author. E-mail: m.zervos@lse.ac.uk
Impulses should be applied whenever the state process hits the boundaries of the interval $[A,B]$. In particular, whenever $X$ assumes the value $x = A$ or the value $x = B$, control should be exercised to “push” it instantaneously inside the band $[A,B]$. Accordingly, the associated controlled process $Z$ has piece-wise constant sample paths, and its jumps occur at the times when the system’s controller intervenes to reposition the system’s state from the boundary of the interval $[A,B]$ to its interior. The objective of the control problem is to minimise the pathwise long-term average performance criterion defined by

$$J^P(C_x) = \limsup_{T \to \infty} \frac{1}{T} \left[ \int_0^T h(X_t) \, dt + \sum_{t \in [0,T]} \left( C_A + \int_A^{A+\Delta Z_t} K_A(s) \, ds \right) 1_{\{\Delta Z_t > 0\}} + \sum_{t \in [0,T]} \left( C_B + \int_{B-\Delta Z_t}^B K_B(s) \, ds \right) 1_{\{\Delta Z_t < 0\}} \right],$$

(1)

as well as the expected long-term average criterion

$$J^E(C_x) = \limsup_{T \to \infty} \frac{1}{T} \mathbb{E}_x \left[ \int_0^T h(X_t) \, dt + \sum_{t \in [0,T]} \left( C_A + \int_A^{A+\Delta Z_t} K_A(s) \, ds \right) 1_{\{\Delta Z_t > 0\}} + \sum_{t \in [0,T]} \left( C_B + \int_{B-\Delta Z_t}^B K_B(s) \, ds \right) 1_{\{\Delta Z_t < 0\}} \right],$$

(2)

over all admissible control strategies $C_x$. Here, $h : [A,B] \to [0,\infty]$ is a given function that models the running costs resulting from the system’s operation. In particular, this function can be used to penalise deviations of the state process $X$ from a given nominal operating point in the interior of $[A,B]$. The constants $C_A, C_B > 0$ and the functions $K_A, K_B > 0$ are used to penalise control expenditure.

This control problem has been motivated by several applications. The first one arises in the context of intervening to control an exchange rate that must be maintained within a specific trading range. This trading range may be externally imposed, as was the case for European currencies that were subject to the Exchange Rate Mechanism (ERM) during the late 1980’s and early 1990’s, or self-imposed, as is the case for several Asian currencies that are pegged to the US dollar at present. In this context, the state process $X$ represents the targeted exchange rate, while the control process $Z$ models central bank intervention efforts.

The problem of controlling exchange rate dynamics within a specified target zone has attracted considerable interest in the literature. Major theoretical contributions and empirical studies on the subject include Bekaert and Gray [BG98], Cadenillas and Zapatero [CZ99, CZ00], Flood and P. Garber [FG91], Froot and Obstfeld [FO91], Jeanblanc-Picqué [JP93] Krugman [K91], Miller and Zhang [MZ96], Mundaca and Øksendal [MO98], and others. In these references, the optimisation problems that are analysed all involve expected discounted performance criteria. The significance of the associated discounting rates is not fully justified by the relevant economic theory. Indeed, an exchange rate is not
an asset, and central banks typically attempt to control exchange rates by influencing the
supply and demand of the foreign currency relative to the domestic currency through the
purchase and sale of foreign government securities in the open market. Therefore, a long-
term average criterion may be more appropriate for assessing the efficacy of such central
bank interventions.

A second application of the problem we study is associated with controlling inflation
through monetary policy. Historically, the primary function of central banks in free market
economies was to promote stability in the banking system, while political authorities were
responsible for monetary policy. More recently, central banks have been given responsibility
for controlling inflation through the exercise of monetary policy that is independent from
other government policies and objectives. In this new regime, most central banks target
explicit or implicit inflation levels and may face statutory penalties if inflation falls below or
(more often) rises above certain thresholds. In this context, the state process $X$ can be used
to represent the targeted inflation rate while the control process $Z$ can be used to model
central bank intervention efforts. A singular stochastic control model with an expected dis-
counted performance criterion addressing this application has been developed and studied by
Chiarolla and Haussmann [CH98]. In their task of controlling inflation, central banks typi-
cally attempt to control the rate either by changing the domestic interest rate or by altering
money supply through the issuance or withdrawal of domestic government securities. Once
again, a long-term average criterion seems more appropriate than an expected discounted
one for assessing the efficacy of such interventions.

A further potential application of the stochastic control problem that we study arises in
the context of defined benefit pension fund management. In this context, the state process
$X$ models the present value of the surplus or deficit of the fund, while the impulse control
process $Z$ models the present value of the contributions that the sponsoring organisation
would be required to make in order to maintain the funding position of the plan within an
appropriate range. In this case, a long-term average criterion would be more appropriate
for assessing contribution policy from an accounting, regulatory, and solvency perspective
because contributions are effectively subtracted from retained earnings and therefore affect
directly the book value of equity, which only changes through new stock issuance or through
earnings retention over time.

Controlling the price of energy commodities (such as oil or natural gas) through coordi-
nated action on supply undertaken by members of the Organisation of Petroleum Exporting
Countries (OPEC) presents another possible application of the optimal control problem that
we study. In this context, the state process is used to model the price in real terms of the
targeted commodity, while the control process reflects coordinated changes in the supply of
the targeted commodity.

With respect to the applications discussed above, the time lag between a controlled action
and the resulting state process response presents a potential shortcoming. This is a concern
relative to applications such as the control of an inflation rate, in which case, it is commonly
accepted that the state process responds to impulsive action with a lag of nine to twelve
months. However, this is less of an issue in exchange rate and commodity price control applications, where an announcement or even the suspicion of an intervention is sufficient to move the targeted rate or price in the desired direction. In any case, we view the stochastic control problem that we solve as an approximation to the specific practical applications that we have discussed.

Stochastic control problems with asymptotic or ergodic performance criteria have recently attracted significant interest in the literature. In the context of portfolio optimisation, a number of models with asymptotic performance criteria have been studied by Akian, Sulem and Taksar [AST01], Bielecki and Pliska [BP99, BP00], Bielecki, Pliska and Sheu [BPS05], Fleming and Sheu [FS02], Irle and Sass [IS06a, IS06b], and others. From a control theoretic perspective, the stochastic control problems solved by Jack and Zervos [JZ06a], and Irle and Sass [IS06b], in which references, the specification of the band \([A, B]\) is part of the problem, are the closest to the problem that we solve. These authors make different assumptions (for instance, half of the special cases that we discuss in Examples 1 and 2 do not satisfy the assumptions of Jack and Zervos [JZ06a], who also make the additional assumptions that the functions \(K_A\) and \(K_B\) appearing in (1)–(2) are constant and that the function \(h\) has a “cup” shape), and their analysis of the associated free-boundary problems turns out to be of a surprisingly different nature. Other related problems involving singular rather than impulse control were solved by Karatzas [K83], Weerasinghe [W02], and Jack and Zervos [JZ06b]. Further notable contributions to the theory of continuous time stochastic control with an ergodic criterion include Borkar [B99], Borkar and Ghosh [BG88], Kurtz and Stockbridge [KS98], Kushner [K78], and several others.

2 Problem formulation

The state process \(X\) of the stochastic system that we study is driven by a Brownian motion \(W\) and a controlled process \(Z\) that affects the system’s dynamics in an impulsive way. The associated one-dimensional SDE that describes the dynamics of this system is

\[
dX_t = b(X_t) \, dt + dZ_t + \sigma(X_t) \, dW_t, \quad X_0 = x \in [A, B],
\]

(3)

where \(b, \sigma : [A, B] \to \mathbb{R}\) are given functions, \(W\) is a standard, one-dimensional Brownian motion, and \(-\infty < A < B < \infty\). The impulse control process \(Z\) is a finite variation, càglàd, piece-wise constant process. The time evolution of this process is determined by the system’s controller, who is given the two boundary points \(A, B \in \mathbb{R}\) and has the task of applying a strictly negative impulse whenever the state process \(X\) assumes the value \(x = B\) and a strictly positive impulse when the state process \(X\) assumes the value \(x = A\) so as to keep \(X\) inside the band \([A, B]\) at all times. Such a process can also be described by the collection

\[
Z = (\tau_1, \tau_2, \ldots, \tau_n, \ldots; \Delta Z_{\tau_1}, \Delta Z_{\tau_2}, \ldots, \Delta Z_{\tau_n}, \ldots),
\]

where \(\tau_n\) is the random time at which the \(n\)-th jump of \(Z\) occurs and \(\Delta Z_{\tau_n} := Z_{\tau_n} - Z_{\tau_n^-}\) is the size of the corresponding jump. In view of this characterisation, an admissible choice of
Z should satisfy
\[
X_{\tau_n} \mathbb{1}_{\{\Delta Z_{\tau_n} > 0\}} = A, \quad X_{\tau_n} \mathbb{1}_{\{\Delta Z_{\tau_n} < 0\}} = B \quad \text{and} \quad |\Delta Z_{\tau_n}| < B - A, \quad \text{for all } n \geq 1. \quad (4)
\]

The controlled process Z affects the system’s state process only by causing a jump of size \(\Delta X_{\tau_n} = \Delta Z_{\tau_n}\) at each of the times \(\tau_n\). Indeed, the evolution of the state process between any two consecutive times at which Z has a discretionary jump is governed by the uncontrolled SDE
\[
dX_t = b(X_t) \, dt + \sigma(X_t) \, dW_t, \quad X_0 = x \in [A, B]. \quad (5)
\]

We make the following assumption.

**Assumption 1** The function \(\sigma^2\) is bounded, \(\sigma^2(x) > 0\), for all \(x \in [A, B]\), and
\[
\int_A^B \frac{1 + |b(s)|}{\sigma^2(s)} \, ds < \infty. \quad (6)
\]

In the presence of this assumption both of conditions \((\text{ND})'\) and \((\text{LI})'\) in Section 5.5 of Karatzas and Shreve [KS88] are satisfied, and, therefore, \((5)\) has a weak solution up to the exit time from \([A, B]\) that is unique in the sense of distribution law. Also, the scale function and the speed measure that characterise a one-dimensional diffusion such as the one in \((5)\), given by
\[
p_\gamma(\gamma) = 0, \quad p_\gamma'(x) = \exp \left(- \int_\gamma^x \frac{2b(s)}{\sigma^2(s)} \, ds\right), \quad \text{for } x \in [A, B], \quad (7)
\]
and
\[
m_\gamma(dx) = \frac{2}{\sigma^2(x)p_\gamma'(x)} \, dx, \quad (8)
\]
respectively, are well-defined for any given choice of \(\gamma \in [A, B]\). At this point, we make the following remark that we will need in our analysis.

**Remark 1** In light of the integrability condition \((6)\), we can see that
\[
0 < \exp \left(- \int_A^B \frac{2|b(s)|}{\sigma^2(s)} \, ds\right) \leq p_\gamma'(x) \leq \exp \left(\int_A^B \frac{2|b(s)|}{\sigma^2(s)} \, ds\right) < \infty,
\]
which implies that there exists a constant \(C_1 > 0\) such that
\[
\frac{1}{C_1} \leq p_\gamma'(x) \leq C_1, \quad \text{for all } \gamma, x \in [A, B]. \quad (9)
\]
Combining these inequalities with (6), we can calculate
\[ \int_{x_1}^{x_2} m_\gamma(ds) \leq 2C_1 \int_{x_1}^{x_2} \sigma^{-2}(s) ds \]
\[ \leq 2C_1 \int_A^B \sigma^{-2}(s) ds \]
\[ < \infty, \quad \text{for all } \gamma \in [A, B] \text{ and } A \leq x_1 < x_2 \leq B. \] (10)

In particular, \( m_\gamma \) is a finite positive measure on \( ([A, B], \mathcal{B}([A, B])) \), where \( \mathcal{B}([A, B]) \) is the Borel \( \sigma \)-algebra on \([A, B]\), and
\[ \lim_{\gamma \uparrow B} m_\gamma ([\gamma, B]) = 0. \] (11)

Furthermore, we note that
\[ \int_A^B \sigma^2(s) m_\gamma(ds) \leq 2C_1(B - A) < \infty, \] (12)
and that
\[ \int_{x_1}^{x_2} b(s) m_\gamma(ds) = \frac{1}{p_\gamma(x_2)} - \frac{1}{p_\gamma(x_1)} \in \mathbb{R}, \quad \text{for all } \gamma, x_1, x_2 \in [A, B]. \] (13)

We adopt a weak formulation of the control problem that we study.

**Definition 1** Given an initial condition \( x \in [A, B] \), an admissible impulse control of the stochastic system under consideration is any seven-tuple \( C_x = (\Omega, \mathcal{F}, \mathcal{F}_t, P, W, Z, X) \) such that

- \((\Omega, \mathcal{F}, \mathcal{F}_t, P)\) is a filtered probability space satisfying the usual conditions,
- \( W \) is a standard one-dimensional \((\mathcal{F}_t)\)-Brownian motion,
- \( Z \) is a finite variation, piece-wise constant, càglàd, \((\mathcal{F}_t)\)-adapted process,
- \( X \) is a càglàd, \((\mathcal{F}_t)\)-adapted process with values in \([A, B]\), and both of (3) and (4) hold true.

We define \( C_x \) to be the family of all admissible controls \( C_x \).

With each control \( C_x \in C_x \), we associate the long-term average *pathwise* performance criterion \( J^P(C_x) \) as well as the long term average *expected* criterion \( J^E(C_x) \), which are defined by (1) and (2) in the introduction, respectively. The objective of the optimisation problem is to minimise these performance criteria over all control strategies \( C_x \in C_x \). For the resulting control problem to be well-posed, we make the following additional assumption. (Note that the equivalence in (14) follows immediately from the definition (8) of the speed measure \( m_\gamma \), and and (9).)
Assumption 2 The function \( h : ]A, B[ \rightarrow [0, \infty[ \) is Lebesgue-measurable and satisfies the integrability condition
\[
\int_{A}^{B} h(s) m_{\gamma}(ds) < \infty \quad \iff \quad \int_{A}^{B} \frac{h(s)}{\sigma^{2}(s)} ds < \infty, \quad (14)
\]
\( C_{A}, C_{B} > 0 \), and the functions \( K_{A}, K_{B} : [A, B] \rightarrow ]0, \infty[ \) are absolutely continuous. Furthermore, there exist constants \( \mu^{A}, \nu^{A}, \mu^{B}, \nu^{B} \in [A, B] \) such that
\[
A \leq \mu^{A} \leq \nu^{A} \leq B, \quad A \leq \mu^{B} \leq \nu^{B} \leq B, \quad (15)
\]
\[
-\frac{1}{2} \sigma^{2}(x) K'_{A}(x) - b(x) K_{A}(x) + h(x) \left\{ \begin{array}{ll}
& \text{is positive and strictly decreasing in } ]A, \mu^{A}], \\
& \text{is negative in } ]\mu^{A}, \nu^{A}], \\
& \text{is positive and strictly increasing in } ]\nu^{A}, B[,
\end{array} \right. \quad (16)
\]
and
\[
\frac{1}{2} \sigma^{2}(x) K'_{B}(x) + b(x) K_{B}(x) + h(x) \left\{ \begin{array}{ll}
& \text{is positive and strictly decreasing in } ]A, \mu^{B}], \\
& \text{is negative in } ]\mu^{B}, \nu^{B}], \\
& \text{is positive and strictly increasing in } ]\nu^{B}, B[.
\end{array} \right. \quad (17)
\]

Remark 2 In view of the fact that some of the inequalities in (15) are not strict, (16) or (17) are understood in the sense that, if a condition is associated with an empty interval of the form \( ]a, a[ \), then it is omitted from the corresponding list. Also, it is worth noting that we allow for the possibility that
\[
\lim_{x \downarrow A} h(x) = \infty \quad \text{and/or} \quad \lim_{x \uparrow B} h(x) = \infty.
\]
Furthermore, the boundedness of the interval \( [A, B] \) and the continuity of the strictly positive functions \( K_{A} \) and \( K_{B} \) imply that there exists a constant \( C_{2} > 0 \) such that
\[
\frac{1}{C_{2}} \leq K_{A}(x), K_{B}(x) \leq C_{2}, \quad \text{for all } x \in [A, B].
\]

The following examples, which have been motivated by the existing literature on the subject, show that our assumptions are rather general.

Example 1 Suppose that \( X \) is a geometric Brownian motion, so that \( b(x) = \tilde{b} x \) and \( \sigma(x) = \tilde{\sigma} x \), for some constants \( \tilde{b} \) and \( \tilde{\sigma} \), and that \( 0 < A < B < \infty \). In this case, the scale function of \( X \) is given by \( p'_{\gamma}(x) = (\gamma/x)^{2\tilde{b}/\tilde{\sigma}^{2}} \), for \( \gamma, x > 0 \).
In line with Jeanblanc-Picqué [JP93], suppose also that
\[ h(x) = 0, \quad C_A = C_B = C \quad \text{and} \quad K_A(x) = K_B(x) = K, \]
for some constants \( C, K > 0 \). Noting that, in this case,
\[ -\frac{1}{2} \sigma^2(x) K_A'(x) - b(x) K_A(x) + h(x) = -\tilde{b} K x \]
and
\[ \frac{1}{2} \sigma^2(x) K_B'(x) + b(x) K_B(x) + h(x) = \tilde{b} K x, \]
we can see that both of (16) and (17) are satisfied for \( \mu^A = \nu^A = A, \mu^B = A \) and \( \nu^B = B \)
if \( \tilde{b} < 0 \), for \( \mu^A = A, \nu^A = B \) and \( \mu^B = \nu^B = B \) if \( \tilde{b} > 0 \), and for \( \mu^A = \mu^B = A \) and \( \nu^A = \nu^B = B \) if \( \tilde{b} = 0 \). In light of this observation, we can see that all of our assumptions are satisfied.

Now, with reference to Cadenillas and Zapatero [CZ99], suppose that
\[ h(x) = (x - \rho)^2, \quad C_A = C_B = C \quad \text{and} \quad K_A(x) = K_B(x) = K, \]
for some constants \( \rho \in ]A, B[ \) and \( C, K > 0 \). In this case,
\[ -\frac{1}{2} \sigma^2(x) K_A'(x) - b(x) K_A(x) + h(x) = x^2 - (2\rho + \tilde{b} K)x + \rho^2. \]
If the quadratic appearing on the right hand side of this expression has no real roots, which is the case if and only if \( \tilde{b} < 0 \), then (16) is satisfied for
\[ \mu^A = \nu^A = \begin{cases} A, & \text{if } \rho + \frac{1}{2} \tilde{b} K < A, \\ \rho + \frac{1}{2} \tilde{b} K, & \text{if } \rho + \frac{1}{2} \tilde{b} K \in ]A, B[, \\ B, & \text{if } \rho + \frac{1}{2} \tilde{b} K < B. \end{cases} \]

On the other hand, if the quadratic has real solutions, say \( x_1 < x_2 \), then (16) is satisfied for \( \mu^A = A \lor x_1 \) and \( \nu^A = B \land x_2 \). Similarly, we can verify that (17) holds true for appropriate values of \( \mu^B \) and \( \nu^B \), and conclude that all of our assumptions are satisfied. \( \square \)

**Example 2** Suppose that \( X \) is a mean-reverting Ornstein-Uhlenbeck process as in Vasicek’s model, so that \( b(x) = \kappa(\theta - x) \) and \( \sigma(x) = \tilde{\sigma} \), for some constants \( \kappa, \theta > 0 \) and \( \tilde{\sigma} \), or a mean-reverting square-root process as in the Cox-Ingersoll-Ross model, so that \( b(x) = \kappa(\theta - x) \) and \( \sigma(x) = \tilde{\sigma} \sqrt{x} \), for some constants \( \kappa, \theta > 0 \) and \( \tilde{\sigma} \) such that \( 2\kappa\theta/\tilde{\sigma}^2 > 0 \), and that \( 0 < A < B < \infty \). The scale functions of the associated diffusions are given by
\[ p^*_\gamma(x) = \exp \left( \frac{\kappa}{\tilde{\sigma}^2} \left( x^2 - 2\theta x + 2\gamma \gamma - \gamma^2 \right) \right), \quad \text{for } \gamma, x > 0, \]
in the first case, and by
\[ p'_1(x) = \left( \frac{2}{x} \right)^{2\theta/\sigma^2} \exp \left( \frac{2\kappa}{\sigma^2}(x - \gamma) \right), \quad \text{for } \gamma, x > 0, \]
in the second case.

If we choose
\[ h(x) = 0, \quad C_A = C_B = C \quad \text{and} \quad K_A(x) = K_B(x) = K, \]
for some constants \( C, K > 0 \), then the identities
\[ -\frac{1}{2}\sigma^2(x)K'_A(x) - b(x)K_A(x) + h(x) = \kappa K(x - \theta) \]
and
\[ \frac{1}{2}\sigma^2(x)K'_B(x) + b(x)K_B(x) + h(x) = -\kappa K(x - \theta), \]
which are associated with either of the two possibilities for the underlying diffusion \( X \) considered here, reveal that both of (16) and (17) are satisfied if we let \( \mu^A = A, \nu^A = (A \lor \theta) \land B, \mu^B = A \lor (\theta \land B) \) and \( \nu^B = B \). Thus, we can see that all of our assumptions are satisfied.

We can also show that all of our assumptions are satisfied if we choose
\[ h(x) = (x - \rho)^2, \quad C_A = C_B = C \quad \text{and} \quad K_A(x) = K_B(x) = K, \]
for some constants \( \rho \in ]A, B[ \) and \( C, K > 0 \); in this case, verifying (16) and (17) gives rise to the same situation as the one corresponding to the same choices for \( h, C_A, C_B, K_A \) and \( K_B \) in Example 1.

3 The solution of the control problem

We solve the control problem formulated in the previous section by finding a sufficiently smooth, for an application of Itô’s formula, function \( w \) and a constant \( \lambda^* \) satisfying the ODE
\[ \frac{1}{2}\sigma^2(x)w''(x) + b(x)w'(x) + h(x) - \lambda^* = 0, \quad x \in ]A, B[, \]
with the boundary conditions
\[ \inf_{x \in ]A, B[} \mathcal{W}_A(x) = 0 \quad \text{and} \quad \sup_{x \in ]A, B[} \mathcal{W}_B(x) = 0, \]
where the functions \( \mathcal{W}_A \) and \( \mathcal{W}_B \) are defined by
\[ \mathcal{W}_A(x) = \int_A^x [w'(s) + K_A(s)] ds + C_A \]
with the boundary conditions
\[ \inf_{x \in ]A, B[} \mathcal{W}_A(x) = 0 \quad \text{and} \quad \sup_{x \in ]A, B[} \mathcal{W}_B(x) = 0, \]
\[ W_B(x) = \int_x^B [w'(s) - K_B(s)] ds - C_B, \quad (21) \]

respectively. If such a pair \((w, \lambda^*)\) exists, then we expect the following statements to be true. Given any initial condition \(x \in [A, B]\),

\[ \lambda^* = \inf_{C_x \in \mathcal{C}_x} J^P(C_x) = \inf_{C_x \in \mathcal{C}_x} J^E(C_x), \]

which reflects the fact that the optimal value of the performance criterion is independent of the system’s initial condition. Equation (18) is associated with the requirement that the controller should take no action as long as the system’s state process assumes values in the interval \([A, B]\). Also, the boundary conditions (19) arise from the requirement that the controller should exercise immediate impulsive action so as to appropriately reposition the state process inside the interval \([A, B]\) whenever this hits the boundary points \(A\) and \(B\).

We conjecture that an optimal strategy is characterised by two free-boundary points \(\alpha, \beta\) such that \(A < \alpha < \beta < B\), and takes the form that can be described as follows. Whenever the state process \(X\) assumes the value \(x = A\) (resp., \(x = B\)), control is exercised to “push” it instantaneously to the level \(\alpha\) (resp., \(\beta\)). As long as the state process is inside the interval \([A, B]\), the controller should take no action. Accordingly, we look for two points \(\alpha < \beta\) in \([A, B]\) and a solution \((w, \lambda^*)\) to (18)–(19) such that

\[ W_A(\alpha) = W_B(\beta) = 0. \quad (22) \]

Assuming that this strategy is indeed optimal, we need an appropriate system of equations to determine the free boundary points \(\alpha, \beta\) and the constant \(\lambda^*\). To this end, we note that (19) and (22) imply that \(\alpha\) (resp., \(\beta\)) is a local minimum (resp., maximum) of the function \(W_A\) (resp., \(W_B\)). It follows that the equations

\[ w'(\alpha) = -K_A(\alpha) \quad \text{and} \quad w'(\beta) = K_B(\beta) \]

should be satisfied. Furthermore, these equations, the strict positivity of \(K_A, K_B\), and the continuity of \(w'\) imply that there exists a point \(\gamma^* \in ]\alpha, \beta[\) such that \(w'(\gamma^*) = 0\).

Now, it is straightforward to see that the solution to the ODE (18) that satisfies the boundary condition \(w'(\gamma^*) = 0\) is given by

\[ w'(x) = p_{\gamma^*}(x) \int_{\gamma^*}^x [\lambda^* - h(s)] m_{\gamma^*}(ds), \quad x \in [A, B], \quad (23) \]

where \(p_{\gamma^*}\) and \(m_{\gamma^*}\) are the scale function and the speed measure of the uncontrolled diffusion (5), defined by (7) and (8), respectively. It follows that, in order to determine the four
unknown parameters $\alpha < \gamma^* < \beta$ and $\lambda^*$, we have to solve the system of non-linear algebraic equations

\begin{align*}
g(\alpha, \lambda^*, \gamma^*) + K_A(\alpha) &= 0, \\
g(\beta, \lambda^*, \gamma^*) - K_B(\beta) &= 0,
\end{align*}

where the function $g$ is defined by

\begin{equation}
g(x, \lambda, \gamma) = p_\gamma'(x) \int_\gamma^x [\lambda - h(s)] m_\gamma(ds), \quad \text{for } x, \gamma \in [A, B] \text{ and } \lambda \in \mathbb{R},
\end{equation}

and the functions $Q_A$, $Q_B$ are defined by

\begin{align*}
Q_A(x, \lambda, \gamma) &= \int_A^x [g(s, \lambda, \gamma) + K_A(s)] ds + C_A, \quad \text{for } x, \gamma \in [A, B] \text{ and } \lambda \in \mathbb{R}, \\
Q_B(x, \lambda, \gamma) &= \int_x^B [g(s, \lambda, \gamma) - K_B(s)] ds - C_B, \quad \text{for } x, \gamma \in [A, B] \text{ and } \lambda \in \mathbb{R},
\end{align*}

respectively. Furthermore, if (24)–(25) have a solution $(\alpha, \gamma^*, \beta, \lambda^*)$ such that $A < \alpha < \gamma^* < \beta < B$, then the pair $(w, \lambda^*)$, where $w$ is a function satisfying

\begin{equation}
w'(x) = g(x, \lambda^*, \gamma^*), \quad \text{for all } x \in ]A, B[,
\end{equation}

is the required solution to (18)–(19) provided that the associated functions $W_A$ and $W_B$ given by

\begin{equation}
W_A(x) = Q_A(x, \lambda^*, \gamma^*) \quad \text{and} \quad W_B(x) = Q_B(x, \lambda^*, \gamma^*),
\end{equation}

respectively, satisfy (19).

The following result is concerned with showing that the system of equations (24)–(25) has a solution of the required form. To preserve the continuity of the presentation, we prove it in the Appendix.

**Lemma 1** Suppose that Assumptions 1 and 2 hold. The system of equations (24)–(25), where $g$, $Q_A$, $Q_B$ are defined by (26)–(28), has a solution $(\alpha, \gamma^*, \beta, \lambda^*)$ such that $A < \alpha < \gamma^* < \beta < B$. If $w$ is a function satisfying (29), then $w$ is twice differentiable in the classical sense inside $]A, B[$, both of $w$ and $w'$ are bounded in $]A, B[$, and the pair $(w, \lambda^*)$ satisfies (18)–(19). Furthermore, the triplet $(\alpha, \gamma^*, \beta)$ is uniquely defined if

\begin{equation}
h \text{ is strictly decreasing in } ]A, x^\dagger[ \text{ and strictly increasing in } ]x^\dagger, B[,
\end{equation}

for some $x^\dagger \in [A, B]$.

We can now prove the main result of the paper.
Theorem 2 Suppose that Assumptions 1 and 2 hold, let \((\alpha, \gamma^*, \beta, \lambda^*)\) be a solution to the system of equations (24)–(25), where \(g, Q_A, Q_B\) are defined by (26)–(28), and let \((w, \lambda^*)\) be the associated solution to (18)–(19). Given any initial condition \(x \in [A, B]\),

\[
\lambda^* = \inf_{C_x \in \mathcal{C}_x} J^P(C_x) = \inf_{C_x \in \mathcal{C}_x} J^E(C_x),
\]

and the points \(\alpha, \beta\) determine an optimal strategy that has the qualitative nature discussed above.

Proof. Throughout this proof, we fix a solution \((w, \lambda^*)\) to (18)–(19) that is constructed as in Lemma 1, and any initial condition \(x \in [A, B]\). Given an admissible control strategy \(C_x \in \mathcal{C}_x\), we can use Itô’s formula to obtain

\[
w(X_{T+}) = w(x) + \int_0^T \left[ \frac{1}{2} \sigma^2(X_t) w''(X_t) + b(x) w'(X_t) \right] dt + M_T + \sum_{t \leq T} \Delta w(X_t), \tag{32}
\]

where \(\Delta w(X_t) = w(X_{t+}) - w(X_t)\) and \(M\) is the stochastic integral defined by

\[
M_t = \int_0^t \sigma(X_s) w'(X_s) dW_s.
\]

In view of the fact that the pair \((w, \lambda^*)\) satisfies (18), the conditions in (4) that characterise an admissible control strategy, and the definitions (20) and (21) of the functions \(W_A\) and \(W_B\), we can see that (32) implies

\[
\frac{1}{T} \left[ \int_0^T h(X_t) dt + \sum_{t \in [0,T]} \left( C_A + \int_{A+\Delta Z_t} C_B + \int_{B-|\Delta Z_t|} K_A(s) ds \right) \mathbf{1}_{\{\Delta Z_t > 0\}} \right] + \sum_{t \in [0,T]} \left( C_B + \int_{B-|\Delta Z_t|} K_B(s) ds \right) \mathbf{1}_{\{\Delta Z_t < 0\}} \right] \\
= - \frac{w(X_{T+})}{T} + \frac{w(x)}{T} + \lambda^* + \frac{M_T}{T} \\
+ \frac{1}{T} \sum_{t \in [0,T]} W_A(A + \Delta Z_t) \mathbf{1}_{\{\Delta Z_t > 0\}} - \frac{1}{T} \sum_{t \in [0,T]} W_B(B + \Delta Z_t) \mathbf{1}_{\{\Delta Z_t < 0\}} \geq - \frac{w(X_{T+})}{T} + \frac{w(x)}{T} + \lambda^* + \frac{M_T}{T}, \tag{33}
\]

the inequality following thanks to (19). Now, since both of the functions \(\sigma\) and \(w'\) are bounded, the quadratic variation \(\langle M \rangle\) of \(M\) satisfies

\[
\langle M \rangle_T \leq \sup_{s \in [A,B]} [\sigma(s)w'(s)]^2 T < \infty, \text{ for all } T > 0,
\]

12
\[ E[M_T] = 0 \quad \text{and} \quad \lim_{T \to \infty} \frac{M_T}{T} = \infty. \]  

(34)

In light of this observation, we can take expectations, depending on the case considered, and pass to the limit \( T \to \infty \) in (33) to obtain

\[ J^P(C_x) \geq \lambda^* \quad \text{and} \quad J^E(C_x) \geq \lambda^*. \]

To prove the reverse inequality, let \( C_o \in C_x \) be such that

\[ \Delta Z^o \mathbf{1}_{\{ \Delta Z^o > 0 \}} = (\alpha - A) \mathbf{1}_{\{X^o_t = A\}} \quad \text{and} \quad \Delta Z^o \mathbf{1}_{\{ \Delta Z^o < 0 \}} = (\beta - B) \mathbf{1}_{\{X^o_t = B\}}. \]

Such an admissible control policy can be constructed rigorously as in the proof of Theorem 2 in Jack and Zervos [JZ06a]. For this control strategy, the identities in (22) imply that (33) holds with equality. In view of this observation and (34), we can pass to the limit \( T \to \infty \) to obtain \( J^P(C^o) = \lambda^* \) and \( J^E(C^o) = \lambda^* \), and the proof is complete. \( \square \)

**Appendix: Proof of Lemma 1**

We first note that, in light of (10), (12) and (13) in Remark 1 and the integrability condition (14), our assumptions imply that the functions \( g(\cdot, \cdot, \lambda), Q_A(\cdot, \cdot, \lambda) \) and \( Q_B(\cdot, \cdot, \lambda) \) defined by (26)–(28) as well as the functions \( G_A(\cdot, \cdot, \lambda) \) and \( G_B(\cdot, \cdot, \lambda) \) defined by (39) and (41) below, are all bounded, for all \( \lambda \in \mathbb{R} \). In particular, if \( w \) is a function satisfying (29), for some \( \lambda^* \in \mathbb{R} \) and \( \gamma^* \in [A, B] \), then \( w' \) is bounded in \( [A, B] \), which, combined with the fact that the interval \( [A, B] \) is bounded, implies that \( w \) is bounded in \( [A, B] \).

To assist the reader, we split the proof that we develop in four main steps. At this point, we also note that we adopt the convention \( a, a[ = \emptyset \), for all \( a \in \mathbb{R} \), throughout the proof.

**Step 1. Preliminary calculations.** Recalling the definitions (7) and (8) of the scale function \( p_{\gamma} \) and the speed measure \( m_{\gamma} \), respectively, we can calculate

\[
\frac{\partial g}{\partial x}(x, \lambda, \gamma) = -\frac{2}{\sigma^2(x)} [b(x)g(x, \lambda, \gamma) + h(x) - \lambda] \quad (35)
\]

and

\[
\frac{\partial g}{\partial \lambda}(x, \lambda, \gamma) = p'_{\gamma}(x) \int_{\gamma}^{x} m_{\gamma}(ds) \begin{cases} < 0, & \text{if } x < \gamma, \\ > 0, & \text{if } x > \gamma, \end{cases} \quad (36)
\]

and

\[
\frac{\partial g}{\partial \gamma}(x, \lambda, \gamma) = \frac{\partial}{\partial \gamma} \int_{\gamma}^{x} \frac{2[\lambda - h(s)]}{\sigma^2(s)} \exp \left( -\int_{s}^{x} \frac{2b(u)}{\sigma^2(u)} \, du \right) \, ds \\
= -\frac{2[\lambda - h(\gamma)]}{\sigma^2(\gamma)} p'_{\gamma}(x). \quad (37)
\]
Also, noting that the function $K_A$ is a solution $v$ to the ODE
\[
\frac{1}{2}\sigma^2(x)v'(x) + b(x)v(x) - \left[\frac{1}{2}\sigma^2(x)K'_A(x) + b(x)K_A(x)\right] = 0,
\]
we can see that
\[
K_A(x) = p'_\gamma(x) \left( K_A(\gamma) + \int^{x}_{\gamma} \left[\frac{1}{2}\sigma^2(s)K'_A(s) + b(s)K_A(s)\right] m_\gamma(ds) \right).
\]
Combining this observation with the definition (26) of the function $g$, we obtain
\[
g(x, \lambda, \gamma) + K_A(x) = p'_\gamma(x)G_A(x, \lambda, \gamma),
\]
where
\[
G_A(x, \lambda, \gamma) = K_A(\gamma) + \int^{x}_{\gamma} \left[\frac{1}{2}\sigma^2(s)K'_A(s) + b(s)K_A(s) + h(s) - \lambda\right] m_\gamma(ds).
\]
Similarly, we can see that
\[
g(x, \lambda, \gamma) - K_B(x) = -p'_\gamma(x)G_B(x, \lambda, \gamma),
\]
where
\[
G_B(x, \lambda, \gamma) = K_B(\gamma) + \int^{x}_{\gamma} \left[\frac{1}{2}\sigma^2(s)K'_B(s) + b(s)K_B(s) + h(s) - \lambda\right] m_\gamma(ds).
\]

**Step 2. Study of the function $g$.** We note that (36) implies that
\[
\lim_{\lambda \to \infty} g(x, \lambda, \gamma) = \begin{cases} -\infty, & \text{if } x < \gamma, \\ \infty, & \text{if } x > \gamma, \end{cases}
\]
and that the positivity of the function $h$ implies that
\[
g(x, 0, \gamma) = -p'_\gamma(x) \int^{x}_{\gamma} h(s) m_\gamma(ds) \begin{cases} > 0, & \text{if } x < \gamma, \\ < 0, & \text{if } x > \gamma. \end{cases}
\]
Combining these two observations with the fact that the continuous functions $K_A, K_B : [A, B] \to ]0, \infty[$ both are uniformly bounded away from 0, we can see that
\[
\lambda_A(\gamma) := \inf \left\{ \lambda \in \mathbb{R} \mid \min_{x \in [A, \gamma]} [g(x, \lambda, \gamma) + K_A(x)] \leq 0 \right\} \in ]0, \infty[, \text{ for all } \gamma \in [A, B], \quad (42)
\]
\[
\lambda_B(\gamma) := \inf \left\{ \lambda \in \mathbb{R} \mid \max_{x \in [\gamma, B]} [g(x, \lambda, \gamma) - K_B(x)] \geq 0 \right\} \in ]0, \infty[, \text{ for all } \gamma \in [A, B], \quad (43)
\]
where we adopt the usual convention \( \inf \emptyset = \infty \). In view of these definitions, it follows that, given any \( \gamma \in ]A, B[ \) and any \( \lambda > \lambda_A(\gamma) \), the equation

\[
G_A(x, \lambda, \gamma) = 0 \iff g(x, \lambda, \gamma) + K_A(x) = 0
\]

has at least one solution in \( ]A, \gamma[ \), and that, given any \( \gamma \in [A, B[ \) and any \( \lambda > \lambda_B(\gamma) \), the equation

\[
G_B(x, \lambda, \gamma) = 0 \iff g(x, \lambda, \gamma) - K_B(x) = 0
\]

has at least one solution in \( ]\gamma, B[ \). Also, the definition (26) of \( g \), the positivity of the function \( h \), the continuity of the strictly positive function \( K_A \), and (9) in Remark 1 imply that, given any \( \lambda \in \mathbb{R} \) and \( \gamma \in ]A, B[ \),

\[
g(x, \lambda, \gamma) + K_A(x) \geq p'_\gamma(x)m_\gamma([x, \gamma]) \left[ \frac{K_A}{p'_\gamma(x)m_\gamma([x, \gamma])} - \lambda \right] \\
\geq p'_\gamma(x)m_\gamma([x, \gamma]) \left[ \frac{K_A}{C_1m_\gamma([A, \gamma])} - \lambda \right], \quad \text{for all } x \in ]A, \gamma[,
\]

where \( K_A = \min_{s \in [A, B]} K_A(s) > 0 \). Similarly, we can show that, given any \( \lambda \in \mathbb{R} \) and \( \gamma \in ]A, B[ \),

\[
g(x, \lambda, \gamma) - K_B(x) \leq p'_\gamma(x)m_\gamma([x, \gamma]) \left[ \lambda - \frac{K_B}{C_1m_\gamma([\gamma, B])} \right], \quad \text{for all } x \in ]\gamma, B[,
\]

where \( K_B = \min_{s \in [A, B]} K_B(s) > 0 \). However, these inequalities, (11) in Remark 1, and the definitions (42), (43) of \( \lambda_A, \lambda_B \) imply that

\[
\lim_{\gamma \uparrow A} \lambda_A(\gamma) = \lim_{\gamma \uparrow B} \lambda_B(\gamma) = \infty. \tag{46}
\]

To proceed further, we fix any \( \gamma \in ]A, B[ \) and any \( \lambda > \lambda_A(\gamma) \), and we consider the equation

\[
\frac{\partial G_A}{\partial x}(x, \lambda, \gamma) = -\frac{1}{2\sigma^2(x)K_A'(x) - b(x)K_A(x) + h(x) - \lambda}{\frac{1}{2}\sigma^2(x)p'_\gamma(x)} = 0.
\]

In view of (16) in Assumption 2, this equation can have at most two solutions in \( ]A, B[ \), say \( x_1 = x_1(\lambda, \gamma) \) and \( x_2 = x_2(\lambda, \gamma) \). In fact,

\[
x_1 \in ]A, \mu_A[ \text{ exists if and only if } A < \mu_A \text{ and }
\]

\[
\lim_{x \downarrow A} \left[ -\frac{1}{2\sigma^2(x)K_A'(x) - b(x)K_A(x) + h(x) - \lambda} \right] > 0,
\]

15
and

$$x_2 \in ]\nu^A, B[ \text{ exists if and only if } \nu^B < B \text{ and }$$

$$\lim_{x \to B} \left[ -\frac{1}{2} \sigma^2(x) K_A'(x) - b(x) K_A(x) + h(x) - \lambda \right] > 0.$$  

Accordingly, if we define $\chi_1 = \chi_1(\lambda, \gamma)$ and $\chi_2 = \chi_2(\lambda, \gamma)$ by

$$\chi_1 = \begin{cases} A, & \text{if } x_1 \text{ does not exist,} \\ x_1, & \text{if } x_1 \in ]A, \mu^A[ \text{ exists,} \end{cases}$$

and

$$\chi_2 = \begin{cases} x_2, & \text{if } x_2 \in ]\nu^A, B[ \text{ exists,} \\ B, & \text{if } x_2 \text{ does not exist,} \end{cases}$$

respectively, then

$$\left\{ \begin{array}{lcl} \frac{\partial G_A}{\partial x}(x, \lambda, \gamma) > 0, & \text{for } x \in ]\chi_1, \chi_2[; \\ < 0, & \text{for } x \in ]A, \chi_1 \cup ]\chi_2, B[. \end{array} \right.$$  \hspace{1cm} (47)

Furthermore, these inequalities imply that

$$\chi_1(\lambda, \gamma) < \gamma, \text{ for all } \gamma \in ]A, B[ \text{ and } \lambda > \lambda_A(\gamma),$$

because the fact that equation (44) has a solution in $]A, \gamma[$ and the fact that $G_A(\gamma, \lambda, \gamma) = K_A(\gamma) > 0$ imply that there exists $\tilde{x} \in ]A, \gamma[$ such that $\frac{\partial G_A}{\partial x}(\tilde{x}, \lambda, \gamma) > 0$.

Now, if we combine the fact that equation (44) has a solution in $]A, \gamma[$ with the observation that $G_A(\gamma, \lambda, \gamma) = K_A(\gamma) > 0$ and (47), we can see that there exists a unique $A(\lambda, \gamma) \in ]\chi_1, \gamma \land \chi_2[ \text{ such that }$

$$G_A(A(\lambda, \gamma), \lambda, \gamma) = 0 \iff g(A(\lambda, \gamma), \lambda, \gamma) + K_A(A(\lambda, \gamma)) = 0, \hspace{1cm} (48)$$

the equation $G_A(x, \lambda, \gamma) = 0$ can have at most one solution $\zeta_1 = \zeta_1(\lambda, \gamma) \in ]A, \chi_1[ \text{ if } \chi_1 > A$ and $G_A(A, \lambda, \gamma) > 0$, and at most one solution $\zeta_2 = \zeta_2(\lambda, \gamma) \in ]\chi_2 \lor \gamma, B[ \text{ if } \gamma < B, \chi_2 < B$ and $G_A(B, \lambda, \gamma) < 0$. Furthermore,

$$G_A(x, \lambda, \gamma) \left\{ \begin{array}{lcl} > 0, & \text{for } x \in ]A, \xi_1^A \cup ]A, \xi_2^A[; \\ < 0, & \text{for } x \in ]\xi_1^A, A \cup ]\xi_2^A, B[, \end{array} \right. \hspace{1cm} (49)$$

where $\xi_1^A = \xi_1^A(\lambda, \gamma)$ and $\xi_2^A = \xi_2^A(\lambda, \gamma)$ are given by

$$\xi_1^A = \begin{cases} A, & \text{if } \zeta_1^A \text{ does not exist,} \\ x_1, & \text{if } \zeta_1^A \text{ exists,} \end{cases}$$

and

$$\xi_2^A = \begin{cases} \zeta_2^A, & \text{if } \zeta_2 \text{ exists,} \\ B, & \text{if } \zeta_2^A \text{ does not exist.} \end{cases}$$

In view of (38), it follows that, given any $\gamma \in ]A, B[ \text{ and any } \lambda > \lambda_A(\gamma)$, there exist unique points $A(\lambda, \gamma), \xi_1^A(\lambda, \gamma)$ and $\xi_2^A(\lambda, \gamma)$ such that

$$A \leq \xi_1^A(\lambda, \gamma) < A(\lambda, \gamma) < \gamma \leq \xi_2^A(\lambda, \gamma) \leq B,$$

16
and
\[
g(x, \lambda, \gamma) + K_A(x) \begin{cases} > 0, & \text{for } x \in ]A, \xi_1^A[ \cup ]A, \xi_2^A[, \\ < 0, & \text{for } x \in ]\xi_1^A, A[ \cup ]A, \xi_2^A, B[. \end{cases} \tag{50}
\]
Additionally, if \( \inf_{x < \gamma} [h(x) - \lambda] > 0 \), then the definition of \( g \) in (26) and the strict positivity of \( K_A \) imply that
\[
g(x, \lambda, \gamma) + K_A(x) = p'_\gamma(x) \int_x^\gamma [h(s) - \lambda] m_\gamma(ds) + K_A(x) > 0, \quad \text{for all } x < \gamma,
\]
which contradicts the existence of \( A(\lambda, \gamma) \in ]A, \gamma[ \) satisfying (48) and (50). Therefore,
\[
\inf_{x \in ]A, \gamma[} [h(x) - \lambda] < 0, \quad \text{for all } \gamma \in ]A, B[ \text{ and } \lambda > \lambda_A(\gamma). \tag{51}
\]
Using similar arguments, we can also show that, given any \( \gamma \in ]A, B[ \) and any \( \lambda > \lambda_B(\gamma) \), there exist unique \( B(\lambda, \gamma), \xi_1^B(\lambda, \gamma) \) and \( \xi_2^B(\lambda, \gamma) \) such that
\[
A \leq \xi_1^B(\lambda, \gamma) \leq \gamma < B(\lambda, \gamma) < \xi_2^B(\lambda, \gamma) \leq B,
\]
\[
g(x, \lambda, \gamma) - K_B(x) \begin{cases} > 0, & \text{for } x \in ]A, \xi_1^B[ \cup ]B, \xi_2^B[, \\ < 0, & \text{for } x \in ]\xi_1^B, B[ \cup ]B, \xi_2^B[, \end{cases} \tag{52}
\]
and
\[
\inf_{x \in ]\gamma, B[} [h(x) - \lambda] < 0, \quad \text{for all } \gamma \in ]A, B[ \text{ and } \lambda > \lambda_B(\gamma). \tag{53}
\]
For future reference, we note that (50) implies that
\[
\frac{\partial g}{\partial x}(A(\lambda, \gamma), \lambda, \gamma) + K'_A(A(\lambda, \gamma)) > 0, \quad \text{for all } \gamma \in ]A, B[ \text{ and } \lambda > \lambda_A(\gamma).
\]
Recalling that \( A(\lambda, \gamma) < \gamma \), we can differentiate the identity (48) with respect to \( \lambda \), and use this inequality and (36) to obtain
\[
\frac{\partial A}{\partial \lambda}(\lambda, \gamma) = -\frac{\partial g}{\partial x}(A(\lambda, \gamma), \lambda, \gamma) + K'_A(A(\lambda, \gamma)) > 0, \quad \text{for all } \gamma \in ]A, B[ \text{ and } \lambda > \lambda_B(\gamma).
\]
However, this calculation and a similar one for \( B \) imply that
\[
A(\cdot, \gamma) \text{ is strictly increasing in } ]\lambda_A(\gamma), \infty[, \text{ for all } \gamma \in ]A, B[,
\tag{54}
\]
and
\[
B(\cdot, \gamma) \text{ is strictly decreasing in } ]\lambda_B(\gamma), \infty[, \text{ for all } \gamma \in ]A, B[.
\tag{55}
\]
Step 3. The solution to the free-boundary problem. We define

\[ Q_A(\lambda, \gamma) = Q_A(A(\lambda, \gamma), \lambda, \gamma), \quad \text{for} \ \gamma \in [A, B] \text{ and } \lambda > \lambda_A(\gamma), \]

\[ Q_B(\lambda, \gamma) = Q_B(B(\lambda, \gamma), \lambda, \gamma), \quad \text{for} \ \gamma \in [A, B] \text{ and } \lambda > \lambda_B(\gamma), \]

where the functions \( Q_A \) and \( Q_B \) are given by (27) and (28). Using (36), (37) and the fact that \( A(\lambda, \gamma) < \gamma \) satisfies (48), we can calculate

\[
\frac{\partial Q_A}{\partial \lambda}(\lambda, \gamma) = \frac{\partial Q_A}{\partial \lambda}(A(\lambda, \gamma), \lambda, \gamma) \frac{\partial A}{\partial \lambda}(\lambda, \gamma) + \frac{\partial Q_A}{\partial \lambda}(A(\lambda, \gamma), \lambda, \gamma)
\]

\[
= \int_A^{A(\lambda,\gamma)} \frac{\partial g}{\partial \lambda}(s, \lambda, \gamma) \, ds
\]

\[
< 0, \tag{56}
\]

and

\[
\frac{\partial Q_A}{\partial \gamma}(\lambda, \gamma) = \frac{\partial Q_A}{\partial \gamma}(A(\lambda, \gamma), \lambda, \gamma) \frac{\partial A}{\partial \gamma}(\lambda, \gamma) + \frac{\partial Q_A}{\partial \gamma}(A(\lambda, \gamma), \lambda, \gamma)
\]

\[
= -\frac{2}[\lambda - h(\gamma)] \int_A^{A(\lambda,\gamma)} p_\gamma(s) \, ds, \tag{57}
\]

for \( \gamma \in [A, B] \) and \( \lambda > \lambda_A(\gamma) \). In light of (56) and (54), we can see that, given any \( \gamma \in [A, B] \) and any \( \lambda_1 > \lambda_A(\gamma) \),

\[
\frac{\partial Q_A}{\partial \lambda}(\lambda, \gamma) < \frac{\partial Q_A}{\partial \lambda}(\lambda_1, \gamma), \quad \text{for all } \lambda > \lambda_1.
\]

Since the right hand side of this inequality is a strictly negative constant, independent of \( \lambda \), it follows that \( \lim_{\lambda \to \infty} Q_A(\lambda, \gamma) = -\infty \), for all \( \gamma \in [A, B] \). However, this limit, the strict inequality in (56), and the observation that

\[
\lim_{\lambda \to \lambda_A(\gamma)} Q_A(\lambda, \gamma) \geq C_A > 0, \tag{58}
\]

which follows from the definition of \( \lambda_A \) in (42), imply that there exists a unique, continuous function \( \Lambda_A : [A, B] \to [0, \infty[ \) such that \( \Lambda_A > \lambda_A \) and

\[ Q_A(\Lambda_A(\gamma), \gamma) \equiv Q_A(A(\Lambda_A(\gamma), \gamma), \Lambda_A(\gamma), \gamma) = 0, \quad \text{for all } \gamma \in [A, B]. \tag{59} \]

We can use symmetric arguments to show that there exists a unique, continuous function \( \Lambda_B : [A, B] \to [0, \infty[ \) such that \( \Lambda_B > \lambda_B \) and

\[ Q_B(\Lambda_B(\gamma), \gamma) \equiv Q_B(B(\Lambda_B(\gamma), \gamma), \Lambda_B(\gamma), \gamma) = 0, \quad \text{for all } \gamma \in [A, B]. \tag{60} \]
The continuity of the positive functions $\Lambda_A$ and $\Lambda_B$, the fact that $\Lambda_A(B), \Lambda_B(A) < \infty$ and (46) imply that the equation $\Lambda_A(\gamma) = \Lambda_B(\gamma)$ has a solution $\gamma^* \in ]A, B[$. Furthermore, in view of (50), (52), (59) and (60), we can see that, if $\gamma^*$ is such a solution, and we let

$$
\lambda^* = \Lambda_A(\gamma^*) \equiv \Lambda_B(\gamma^*), \quad \alpha = \mathcal{A}(\lambda^*, \gamma^*) \quad \text{and} \quad \beta = \mathcal{B}(\lambda^*, \gamma^*),
$$

then $A < \alpha < \gamma^* < \beta < B$ and the collection $(\alpha, \gamma^*, \beta, \lambda^*)$ satisfies the system of equations (24)–(25).

Now, suppose that $h$ satisfies (31). If $\gamma^*$ is a solution to the equation $\Lambda_A(\gamma) = \Lambda_B(\gamma)$, then we can combine the inequalities $\Lambda_A > \lambda_A$ and $\Lambda_B > \lambda_B$ with (31), (51) and (53) to obtain

$$
h(\gamma^*) < \Lambda_A(\gamma^*) \equiv \Lambda_B(\gamma^*).
$$

However, this observation, (56)–(57), and the corresponding expressions for the derivatives of $\mathcal{Q}_B$, imply that

$$
\Lambda_A'(\gamma^*) - \Lambda_B'(\gamma^*) = - \frac{\partial \mathcal{Q}_A}{\partial x}(\Lambda_A(\gamma^*), \gamma^*) + \frac{\partial \mathcal{Q}_B}{\partial x}(\Lambda_B(\gamma^*), \gamma^*) < 0,
$$

which implies that $\gamma^*$ is unique, and establishes the uniqueness claim of the lemma.

**Step 4. Concluding arguments.** To complete the proof, we need to show that the functions appearing in (30) satisfy (19). Taking note of the calculation

$$
\frac{\partial \mathcal{Q}_A}{\partial x}(x, \lambda^*, \gamma^*) = p'_{\gamma^*}(x)\mathcal{G}_A(x, \lambda^*, \gamma^*),
$$

which follows from the definition (27) of the function $\mathcal{Q}_A$ and (38), we can see that the identities $\mathcal{Q}_A(A, \lambda^*, \gamma^*) = C_A$ and $\mathcal{Q}_A(\mathcal{A}(\lambda^*, \gamma^*), \lambda^*, \gamma^*) = 0$ (see (27), (59) and (61)), and the inequalities in (49) imply that the condition $\inf_{x \in [A, B]} \mathcal{W}_A(x) = 0$ in (19) holds true if and only if $\mathcal{Q}_A(B, \lambda^*, \gamma^*) \geq 0$. To prove this inequality, we note that

$$
g(\mathcal{B}(\lambda^*, \gamma^*), \lambda^*, \gamma^*) + K_A(\mathcal{B}(\lambda^*, \gamma^*)) > g(\mathcal{B}(\lambda^*, \gamma^*), \lambda^*, \gamma^*) - K_B(\mathcal{B}(\lambda^*, \gamma^*)) = 0,
$$

which, combined with (50) and the fact that $\mathcal{A}(\lambda^*, \gamma^*) < \mathcal{B}(\lambda^*, \gamma^*)$, implies that

$$
g(x, \lambda^*, \gamma^*) + K_A(x) > 0, \quad \text{for all } x \in ]\mathcal{A}(\lambda^*, \gamma^*), \mathcal{B}(\lambda^*, \gamma^*)[.
$$

However, this observation, the calculation

$$
\int_B^{B(\lambda^*, \gamma^*)} [g(s, \lambda^*, \gamma^*) + K_A(s)] ds
$$

$$
= \mathcal{Q}_B(\mathcal{B}(\lambda^*, \gamma^*), \lambda^*, \gamma^*) + C_B + \int_{B(\lambda^*, \gamma^*)}^{B} [K_A(s) + K_B(s)] ds
$$

$$
= C_B + \int_{B(\lambda^*, \gamma^*)}^{B} [K_A(s) + K_B(s)] ds
$$

$$
> 0,
$$

19
and the fact that $Q_A(\mathcal{A}(\lambda^*, \gamma^*), \lambda^*, \gamma^*) = 0$ imply that

$$Q_A(B, \lambda^*, \gamma^*) = Q_A(\mathcal{A}(\lambda^*, \gamma^*), \lambda^*, \gamma^*) + \int_B^{\mathcal{A}(\lambda^*, \gamma^*)} [g(s, \lambda^*, \gamma^*) + K_A(s)] ds > 0,$$

which establishes the required inequality. Finally, we can use symmetric arguments to show that $\sup_{x \in [A,B]} W_B(x) = 0$ holds true, and the proof is complete. □

References


