Sequential entry and exit decisions with an ergodic performance criteria*

Anne Laure Bronstein† and Mihail Zervos‡
Department of Mathematics
King's College London
The Strand
London WC2R 2LS, UK

November 27, 2005

Abstract

We consider a variant of an optimisation problem involving sequential entry and exit decisions that has emerged in the economics literature as a real option model. The problem that we solve aims at maximising an ergodic, or long-term average, performance criterion in a pathwise as well as in an expected sense. Such a performance index is probably better suited to decision making within a sustainable economic environment. Our results include a complete characterisation of the optimal strategy, which can take qualitatively different forms depending on the problem's data, as well as explicit expressions for the maximal value of the associated performance criterion.

1 Introduction

We consider an investment project in a random economic environment that can be operated in two modes, say "active" and "passive". When it is in its "active" mode, the project yields payoff at a rate that depends on the value of an underlying random economic indicator, such as a given commodity's price or demand, which we model by a general one-dimensional ergodic Itô diffusion. When the project is in its "passive" mode, it incurs losses that reflect, for example, maintenance costs. The transition of the project from one mode to the other one can be realised immediately at certain fixed costs. The sequence of times at which the project's mode is changed constitutes a decision strategy that is determined by the

*Research supported by EPSRC grant no. GR/S22998/01.
†albronstein@hotmail.com
‡mihail.zervos@kcl.ac.uk

project's management. The objective of the resulting optimisation problem is to maximise a performance criterion that quantifies the payoff flow associated with each switching strategy over the set of all admissible such strategies. This type of a real option model has emerged in the economics literature (see, e.g., Brennan and Schwartz [BS85], Dixit [D89] and Dixit and Pindyck [DP94]). Similar models have been analysed in the mathematics literature by Brekke and Øksendal [BØ91, BØ94], Duckworth and Zervos [DZ01], Lumley and Zervos [LZ01], and Wang [W05].

To the best of our knowledge, all of the real option theory, including the references mentioned above, addresses optimisation problems involving *expected discounted* performance indices. Such performance criteria are justified by standard economics theory because they quantify the present value of the payoff flow that is expected from each admissible managerial decision policy. If payoffs resulting from decision making are of a "monetary" nature, then such an approach is the appropriate one. However, if decision making payoffs are of a "utility" nature, then the use of an expected discounted performance criterion is not ideal because, by their very nature, such indices attach higher values to payoffs arising in the shorter term time horizon. Indeed, the choice of the discounting rate that an investor uses in, e.g., Merton's classical utility maximisation problem with an infinite horizon can be interpreted as a quantification of the investor's impatience to consume. Plainly, apart from being associated with "unfairness" when one considers the utility derived from consumption by successive generations, the choice of a discounting rate seems rather arbitrary. As a matter of fact, its main purpose is to guarantee the convergence of the associated performance criteria and the finiteness of the associated value functions. With regard to these economic considerations, one novelty of this paper arises from the fact that we consider an *ergodic*, or *long-term average*, performance criterion that we maximise in a *pathwise* as well as in an *expected* sense. Such a type of an index is probably better suited to "utility" based decision making in the context of sustainable development because it assigns the same weighting to payoffs enjoyed by present and future generations.

The vast majority of the models in the real option theory that admit solutions of an explicit analytical form assume that the underlying economic indicator is modelled by a geometric Brownian motion. One major advantage of the ergodic criterion that we consider here arises from the fact that it allows for results of an equally explicit nature when the underlying economic indicator dynamics are modelled by a wide range of one-dimensional Itô diffusions. These include the exponential of an Ornstein-Uhlenbeck process, which appears in the Black-Karasinski interest rate model, and the family of constant elasticity of variance processes, such as the square root process appearing in the Cox-Ingersoll-Ross interest rate model. It is well documented in the economics literature that such mean-reverting diffusions present much more realistic models for a range of economic indicators, such as commodity prices, than the geometric Brownian motion. Therefore, the model that we study can provide a most valuable alternative when addressing practical applications.

The use of performance indices of an ergodic nature can be criticised on the grounds that they result in highly non-unique optimal strategies. In particular, any two decision strategies that differ on an arbitrarily long, but finite, time period are associated with the same value of the performance criterion. However, the idea that long-term average criteria should be considered in connection with sustainable development applications addresses this issue because, in the presence of a transparent decision making process, it rules out speculation from the perspective of the decision maker.

At this point, we should observe that ergodic stochastic optimal control currently has a well-developed body of theory. In particular, one should note major advances in the field that include, restricting attention to continuous-time models, Kushner [K78], Karatzas [K83], Gatarek and Stettner [GS90], Borkar and Ghosh [BG88], Bensoussan and Frehse [BF92], Menaldi, Robin and Taksar [MRT92], Duncan, Maslowski and Pasik-Duncan [DMP98], Kurtz and Stockbridge [KS98], Borkar [B99], Kruk [K00], Sadowy and Stettner [SS02], the references therein, and others. Also, ergodic stochastic control with a pathwise rather than an expected criterion has recently attracted considerable interest in the literature, e.g., see Rotar [R91], Presman, Rotar and Taksar [PRT93], Dai Pra, Di Masi and Trivellato [DDT01], Dai Pra, Runggaldier and Tolloti [DRT04], and the references therein.

The paper is organised as follows. Section 2 is concerned with the formulation of the investment project model that we study. In Section 3, we consider examples of stochastic dynamics for the underlying economic indicator that satisfy our assumptions, and we reformulate the optimisation problems that we solve to equivalent and simpler ones. In Section 4, we consider the associated dynamic programming equation, and we establish a verification theorem and an ergodic result that we use later. Finally, Section 5 is concerned with the solution to the optimisation problems considered.

2 Problem formulation

We consider an investment project that is operated within a random economic environment. We model this environment by means of a one-dimensional Itô diffusion. In particular, we assume that all randomness affecting the payoff flow resulting from the project's management is characterised by a state process X that satisfies the one-dimensional SDE

$$dX_t = b(X_t)\, dt + \sigma(X_t)\, dW_t, \quad X_0 = x \in \mathbb{R}, \tag{1}$$

where $b, \sigma : \mathbb{R} \to \mathbb{R}$ are given functions, and W is a standard one-dimensional Brownian motion. In practice, we can think of such an investment project as a unit that can produce a single commodity. In this context, the process X can be used to model an economic indicator, such as the commodity's demand, or the logarithm of the commodity's price.

We assume that the project can be operated in two distinct modes, say "active" and "passive". The sequence of times at which the project's operating mode is switched from "active" to "passive" and vice versa presents a sequence of decisions made by the project's management. We assume that, when decided, the project's transition from one of its operating modes to the other one is realised instantaneously. To model a switching strategy adopted by the project's management, we use an adapted, finite variation, left-continuous process Z with values in $\{0, 1\}$ and we denote Z_0 by z. In particular, a choice of such a switching process Z represents a strategy that keeps the investment in its "active" operating mode when $Z_t = 1$, and in its "passive" mode whenever $Z_t = 0$. Also, the times at which the jumps of Z occur represent the discretionary times at which the project's mode is changed. The variable $Z_0 = z \in \{0, 1\}$ indicates the project's operating mode at time 0.

Throughout our analysis, we adopt a *weak formulation* point of view.

Definition 1 Given an initial condition $(x, z) \in \mathbb{R} \times \{0, 1\}$, a *switching strategy in the random economic environment modelled by* (1) is any collection $\mathbb{C}_{x,z} = (\mathbb{S}_x, Z)$ such that $\mathbb{S}_x = (\Omega, \mathcal{F}, \mathcal{F}_t, \mathbb{P}_x, W, X)$ is a weak solution of (1) and Z is an $(\mathcal{F}_t)$-adapted, finite variation, càglàd process with values in $\{0, 1\}$ and with $Z_0 = z$. We denote by $\mathcal{C}_{x,z}$ the set of all such switching strategies.

For a switching strategy to be well-defined, we adopt the following assumption.

Assumption 1 The deterministic functions $b, \sigma : \mathbb{R} \to \mathbb{R}$ satisfy the following conditions:

$$\sigma^2(x) > 0, \quad \text{for all } x \in \mathbb{R}, \tag{2}$$

$$\text{for all } x \in \mathbb{R}, \ \text{there exists } \varepsilon > 0 \text{ such that } \int_{x-\varepsilon}^{x+\varepsilon} \frac{1 + |b(s)|}{\sigma^2(s)} \, ds < \infty. \tag{3}$$

Indeed, with regard to standard theory of one-dimensional diffusions (see Karatzas and Shreve [KS88] and Rogers and Williams [RW00]), (2) and (3) imply that (1) defines a regular one-dimensional diffusion. Moreover, the *scale function* p and the *speed measure* m given by

$$p(0) = 0 \quad \text{and} \quad p'(x) = \exp\left(-2 \int_0^x \frac{b(s)}{\sigma^2(s)} \, ds\right), \quad \text{for } x \in \mathbb{R}, \tag{4}$$

and

$$m(dx) = \frac{2}{\sigma^2(x) p'(x)} \, dx, \tag{5}$$

respectively, which characterise one-dimensional diffusions, such as the one associated with (1), are well-defined.

We also assume that the solution to (1) is *non-explosive* and *recurrent*. With regard to Proposition 5.5.22 in Karatzas and Shreve [KS88], we therefore impose the following assumption.

Assumption 2 The scale function p defined by (4) satisfies $\lim_{x \to -\infty} p(x) = -\infty$ and $\lim_{x \to \infty} p(x) = \infty$.

With each switching strategy $\mathbb{C}_{x,z} \in \mathcal{C}_{x,z}$, we associate the *pathwise* performance criterion

$$\begin{aligned}\tilde{J}^{\mathrm{P}}(\mathbb{C}_{x,z}) &\equiv \tilde{J}^{\mathrm{P}}(\mathbb{C}_{x,z}; h_1, h_0, K_1, K_0) \\ &:= \limsup_{T\to\infty} \frac{1}{T}\Bigg[\int_0^T \big[Z_t h_1(X_t) + (1-Z_t)h_0(X_t)\big]\,dt \\ &\qquad\qquad - \sum_{t\in[0,T[} \big[K_1 \mathbf{1}_{\{\Delta Z_t = 1\}} + K_0 \mathbf{1}_{\{\Delta Z_t = -1\}}\big]\Bigg], \end{aligned} \tag{6}$$

as well as the *expected* performance criterion

$$\begin{aligned}\tilde{J}^{\mathrm{E}}(\mathbb{C}_{x,z}) &\equiv \tilde{J}^{\mathrm{E}}(\mathbb{C}_{x,z}; h_1, h_0, K_1, K_0) \\ &:= \limsup_{T\to\infty} \frac{1}{T}\mathbb{E}_x\Bigg[\int_0^T \big[Z_t h_1(X_t) + (1-Z_t)h_0(X_t)\big]\,dt \\ &\qquad\qquad - \sum_{t\in[0,T[} \big[K_1 \mathbf{1}_{\{\Delta Z_t = 1\}} + K_0 \mathbf{1}_{\{\Delta Z_t = -1\}}\big]\Bigg], \end{aligned} \tag{7}$$

where $\Delta Z_t = Z_{t+} - Z_t$. Here, h_1 (resp., h_0) models the running payoff resulting from the project when this is operated in its "active" (resp., "passive") mode. Also, K_0 and K_1 are the fixed costs associated with each switching of the project's operating mode from "active" to "passive" and vice versa, respectively.

The first objective is to maximise $\tilde{J}^{\mathrm{P}}$ over $\mathcal{C}_{x,z}$ in a *pathwise* sense. In particular, we are going to prove that there exists a constant λ^* such that

$$\sup_{\mathbb{C}_{x,z}\in\mathcal{C}_{x,z}} \tilde{J}^{\mathrm{P}}(\mathbb{C}_{x,z}) = \lambda^*, \tag{8}$$

in the sense that, given any initial condition (x,z),

$$\text{for all } \mathbb{C}_{x,z} = (\mathbb{S}_x, Z) \in \mathcal{C}_{x,z},\ \lambda^* \geq \tilde{J}^{\mathrm{P}}(\mathbb{C}_{x,z}),\ \mathbb{P}_x\text{-a.s.}, \tag{9}$$

$$\text{and there exists } \mathbb{C}^*_{x,z} = (\mathbb{S}^*_x, Z^*) \in \mathcal{C}_{x,z} \text{ such that } \lambda^* = \tilde{J}(\mathbb{C}^*_{x,z}),\ \mathbb{P}^*_x\text{-a.s..} \tag{10}$$

The second objective is to maximise $\tilde{J}^{\mathrm{E}}$ over $\mathcal{C}_{x,z}$. In this case, we are going to show that

$$\sup_{\mathbb{C}_{x,z}\in\mathcal{C}_{x,z}} \tilde{J}^{\mathrm{E}}(\mathbb{C}_{x,z}) = \lambda^*, \tag{11}$$

where λ^* is the same constant as the one in (8). The following additional assumption ensures that the resulting optimisation problems are well-defined.

Assumption 3 The following conditions hold:

$$\sigma \text{ is locally bounded}, \tag{12}$$

$$\int_{-\infty}^{\infty} \big[1 + |h_1(s)| + |h_0(s)|\big]\, m(ds) < \infty, \tag{13}$$

$$h := h_1 - h_0 \text{ is strictly increasing}, \tag{14}$$

$$K := K_1 + K_0 > 0. \tag{15}$$

Assumption (12) is of a technical nature, and is satisfied in all cases of interest. Assumption (13) ensures the ergodicity of certain processes, such as the state process X, and is essential for the performance criteria that we consider to be well-defined and for the constant λ^* appearing in (8) and (11) to be a real number. With regard to an interpretation of the state process as an economic indicator, such as demand or a log-price, (14) is a natural assumption to make in practice. Indeed, increased demand/prices are plainly associated with increased running payoff values, which implies that the running payoff function h_1 associated with the "active" mode of the investment project should be an increasing function. On the other hand, it would be reasonable to assume that the running payoff function h_0 associated with the "passive" mode of the project is identically equal to a negative constant modelling running maintenance costs. These two observations provide the grounds for adopting (14) as an assumption. At this point, it is worth noting that the only reason for allowing h_0 to have a non-trivial dependence on the state process is because such a generalisation does not affect the complexity of our analysis, and can potentially be associated with other applications.

Finally, assumption (15) is essential for the well-posedness of the optimisation problem considered. Indeed, the possibility $K_1 + K_0 < 0$ is associated with arbitrarily large values of the performance criteria that can be achieved by a strategy involving sufficiently rapid changes of the project's operational mode. However, even though we interpret the constants K_1 and K_0 as switching costs, we allow for the possibility that one of them is negative. With regard to economics considerations, this presents a degree of freedom that can be used to model a situation such as the one arising when the cost of switching the project from its "passive" mode to its "active" one is not totally sunk, but can be partially recovered by realising the reverse switching.

3 Examples and problem simplifications

If we interpret the state process X given by (1) as a log-price, the geometric Brownian motion that is widely used in finance as well as in the theory of real options as an asset price is not compatible with the assumptions that we have adopted in the previous section because its speed measure has infinite mass and, therefore, (13) is not satisfied. However, a number of asset price processes that are better suited to the commodity markets modelling, and have emerged in the context of the interest rate theory satisfy the requirements of our assumption. The following two examples are concerned with diffusions that are associated with the Black-Karasinski and the Cox-Ingersoll-Ross short rate models.

Example 1 In the context of the Black-Karasinski short rate model, the logarithm of an asset's price identifies with the Ornstein-Uhlenbeck process X given by the SDE

$$dX_t = k(\theta - X_t)\, dt + \sigma\, dW_t,$$

where k, θ and σ are strictly positive constants. It is straightforward to calculate that the scale function p and the speed measure m of this diffusion are given by

$$p'(x) = \exp\left(-\frac{2k\theta}{\sigma^2}x + \frac{k}{\sigma^2}x^2\right),$$
$$m(dx) = \frac{2}{\sigma^2}\exp\left(\frac{k\theta^2}{\sigma^2}\right)\exp\left(-\frac{(x-\theta)^2}{\sigma^2/k}\right)dx,$$

respectively, and to verify that the corresponding requirements in Assumptions 1, 2 and 3 hold, provided that the functions h_0 and h_1 are suitably chosen.

Example 2 We can model the price of a given asset by means of the process e^X satisfying the SDE

$$de^{X_t} = k(\theta - e^{X_t})\,dt + \sigma\left(e^{X_t}\right)^l\,dW_t,$$

where k, θ, σ are strictly positive constants, and $l \in [\frac{1}{2}, 1]$, so that e^X is a so-called constant elasticity of variance (CEV) process. Note that, for $l = \frac{1}{2}$ and $k\theta - \frac{1}{2}\sigma^2 > 0$, e^X identifies with the short rate process in the Cox-Ingersoll-Ross model. With regard to Itô's formula, it is straightforward to check that

$$dX_t = \left(k\theta e^{-X_t} - \tfrac{1}{2}\sigma^2 e^{-2(1-l)X_t} - k\right)dt + \sigma e^{-(1-l)X_t}\,dW_t.$$

The scale function p and the speed measure m of this diffusion are given by

$$p'(x) = \exp\left(-\frac{2k\theta\left(e^{(1-2l)x}-1\right)}{\sigma^2(1-2l)} + \frac{k\left(e^{2(1-l)x}-1\right)}{\sigma^2(1-l)} + x\right),$$
$$m(dx) = \frac{2}{\sigma^2}\exp\left(\frac{2k\theta\left(e^{(1-2l)x}-1\right)}{\sigma^2(1-2l)} - \frac{k\left(e^{2(1-l)x}-1\right)}{\sigma^2(1-l)} + (1-2l)x\right)dx,$$

if $l \in]\frac{1}{2}, 1[$, by

$$p'(x) = \exp\left(\left[1 - \frac{2k\theta}{\sigma^2}\right]x + \frac{2k\left(e^x-1\right)}{\sigma^2}\right),$$
$$m(dx) = \frac{2}{\sigma^2}\exp\left(\frac{2k\theta}{\sigma^2}x - \frac{2k\left(e^x-1\right)}{\sigma^2}\right)dx,$$

if $l = \frac{1}{2}$, and by

$$p'(x) = \exp\left(\frac{2k\theta\left(e^{-x}-1\right)}{\sigma^2} + \left[1 + \frac{2k}{\sigma^2}\right]x\right),$$
$$m(dx) = \frac{2}{\sigma^2}\exp\left(-\frac{2k\theta\left(e^{-x}-1\right)}{\sigma^2} - \left[\frac{2k}{\sigma^2}+1\right]x\right)dx,$$

if $l = 1$. We can check that if $l \in]\frac{1}{2}, 1]$, or if $l = \frac{1}{2}$ and $k\theta - \frac{1}{2}\sigma^2 > 0$, then the requirements in Assumptions 1, 2 and 3 are all satisfied for appropriate choices of the functions h_0 and h_1.

We now consider simplifications of the control problems formulated in Section 2 that we are going to solve. Fix any initial condition (x, z) and any switching strategy $\mathbb{C}_{x,z} \in \mathcal{C}_{x,z}$. With reference to (13) in Assumption 3, the ergodic Theorems V.53.1 and V.54.5 in Rogers and Williams [RW00] imply

$$\lim_{T\to\infty} \frac{1}{T}\left[\int_0^T h_0(X_t)\, dt\right] = \lim_{T\to\infty} \frac{1}{T}\mathbb{E}_x\left[\int_0^T h_0(X_t)\, dt\right] = \frac{1}{m\,(\mathbb{R})}\int_{-\infty}^{\infty} h_0(s)\, m(ds) < \infty. \quad (16)$$

Also, it is straightforward to verify that

$$\lim_{T\to\infty} \frac{1}{T}\sum_{t\in[0,T[} K_0 \Delta Z_t = \lim_{T\to\infty} \frac{1}{T}\mathbb{E}_x\left[\sum_{t\in[0,T[} K_0 \Delta Z_t\right] = 0. \quad (17)$$

Combining these observations with the calculation

$$\begin{aligned}\tilde{J}^{\mathrm{P}}(\mathbb{C}_{x,z}; h_1, h_0, K_1, K_0) = \limsup_{T\to\infty} \frac{1}{T}\Bigg[\int_0^T Z_t\left[h_1(X_t) - h_0(X_t)\right] dt - \sum_{t\in[0,T[} (K_1 + K_0)\, \mathbf{1}_{\{\Delta Z_t = 1\}} \\ + \int_0^T h_0(X_t)\, dt + \sum_{t\in[0,T[} K_0 \Delta Z_t\Bigg],\end{aligned}$$

we can see that

$$\tilde{J}^{\mathrm{P}}(\mathbb{C}_{x,z}; h_1, h_0, K_1, K_0) = J^{\mathrm{P}}(\mathbb{C}_{x,z}; h_1 - h_0, K_1 + K_0) + \frac{1}{m(\mathbb{R})}\int_{-\infty}^{\infty} h_0(s)\, m(ds),$$

where

$$J^{\mathrm{P}}(\mathbb{C}_{x,z}) \equiv J^{\mathrm{P}}(\mathbb{C}_{x,z}; h, K) = \limsup_{T\to\infty} \frac{1}{T}\left[\int_0^T Z_t h(X_t)\, dt - \sum_{t\in[0,T[} K \mathbf{1}_{\{\Delta Z_t = 1\}}\right]. \quad (18)$$

Similarly, we can use (16) and (17) to show that

$$\tilde{J}^{\mathrm{E}}(\mathbb{C}_{x,z}; h_1, h_0, K_1, K_0) = J^{\mathrm{E}}(\mathbb{C}_{x,z}; h_1 - h_0, K_1 + K_0) + \frac{1}{m(\mathbb{R})}\int_{-\infty}^{\infty} h_0(s)\, m(ds),$$

where

$$J^{\mathrm{E}}(\mathbb{C}_{x,z}) \equiv J^{\mathrm{E}}(\mathbb{C}_{x,z}; h, K) = \limsup_{T\to\infty} \frac{1}{T}\mathbb{E}_x\left[\int_0^T Z_t h(X_t)\, dt - \sum_{t\in[0,T[} K \mathbf{1}_{\{\Delta Z_t = 1\}}\right]. \quad (19)$$

It follows that, given any initial condition (x, z), the problem of maximising the performance index $\tilde{J}^{\mathrm{P}}$ (resp., $\tilde{J}^{\mathrm{E}}$) over $\mathcal{C}_{x,z}$ is *equivalent* to maximising the performance criterion J^{P} (resp., J^{E}) over $\mathcal{C}_{x,z}$.

4 The dynamic programming equation

We now consider the problem of maximising the performance indices J^{P} and J^{E} defined by (18) and (19), respectively, over all admissible switching strategies. To discover the optimal strategy, we look for a solution (w_1, w_0) to the Hamilton-Jacobi-Bellman (HJB) equation that takes the form of the following pair of coupled quasi-variational inequalities

$$\max\left\{\tfrac{1}{2}\sigma^2(x)w_1''(x) + b(x)w_1'(x) + h(x),\ w_0(x) - w_1(x)\right\} = 0, \quad x \in \mathbb{R}, \tag{20}$$

$$\max\left\{\tfrac{1}{2}\sigma^2(x)w_0''(x) + b(x)w_0'(x),\ w_1(x) - w_0(x) - K\right\} = 0, \quad x \in \mathbb{R}. \tag{21}$$

With regard to standard theory of stochastic control, the structure of these equations is closely related with the following considerations. Assuming that, at a given time t, the project is in its "passive" mode and the state process X assumes the value x, the project's management is faced with two possible actions. The first one is to switch the project to its "active" mode and then continue optimally. Since the choice of such an action is not necessarily optimal, we can conclude that the value $\mathcal{V}_0(x)$ of the project in its "passive" mode is greater than or equal to the value $\mathcal{V}_1(x)$ of the project in its "active" mode minus the switching cost of K. This observation is associated with the inequality

$$\mathcal{V}_0(x) \geq \mathcal{V}_1(x) - K. \tag{22}$$

The second possible action is to leave the project in its "passive" mode, which is associated with a zero rate of payoff, over a short period of time, and then continue optimally. This second possibility, which may be suboptimal, is associated with the inequality

$$\tfrac{1}{2}\sigma^2(x)\mathcal{V}_0''(x) + b(x)\mathcal{V}_0'(x) \leq 0. \tag{23}$$

Since these are the only two actions that are available to the project's management, one has to be optimal, so one of (22) or (23) must be satisfied with equality. However, these arguments *suggest* the structure of (21). The structure of (20) can be explained in a similar way.

The considerations above explaining the structure of the HJB equation (20)–(21) will play an important role in our investigation that leads to the solution of the optimisation problem considered. However, these ideas *have to be used with care* because the functions w_1 and w_0 neither identify with the value function of the optimisation problem, which, as it turns out is identically equal to a constant, nor do they determine uniquely the optimal strategy. The latter observation is related with the fact that, due to the "average" nature of the performance criterion considered, a suboptimal behaviour over an arbitrarily long, but finite, time period does not affect optimality.

The following result provides conditions that are sufficient for a switching strategy to be optimal.

Theorem 1 *Fix any initial condition $(x, z) \in \mathbb{R}\times\{0, 1\}$, consider the problem of maximising the performance indices J^{P} and J^{E} defined by (18) and (19), respectively, over the class of all* *admissible switching strategies $\mathcal{C}_{x,z}$, and suppose that Assumptions 1, 2 and 3 hold. Suppose that the functions $w_1, w_0 \in W^{2,\infty}_{\mathrm{loc}}(\mathbb{R})$ satisfy (20)–(21), and there exists a constant C such that*

$$\sup_{x\in\mathbb{R}} |w_1(x) - w_0(x)| + \sup_{x\in\mathbb{R}} |\sigma(x)\,[w_1'(x) - w_0'(x)]|^2 < C. \tag{24}$$

Also, suppose that there exists a switching strategy $\mathbb{C}^*_{x,z} = (\mathbb{S}^*_x, Z^*)$ such that

$$\left[\tfrac{1}{2}\sigma^2(X^*_t)w_1''(X^*_t) + b(X^*_t)w_1'(X^*_t) + h(X^*_t)\right] Z^*_t = 0, \quad \textit{Leb-a.e., for all } t \geq 0, \tag{25}$$

$$\left[\tfrac{1}{2}\sigma^2(X^*_t)w_0''(X^*_t) + b(X^*_t)w_0'(X^*_t)\right] (1 - Z^*_t) = 0, \quad \textit{Leb-a.e., for all } t \geq 0, \tag{26}$$

$$[w_1(X^*_t) - w_0(X^*_t) - K]\,\mathbf{1}_{\{\Delta Z^*_t = 1\}} = 0, \quad \textit{for all } t \geq 0, \tag{27}$$

$$[w_0(X^*_t) - w_1(X^*_t)]\,\mathbf{1}_{\{\Delta Z^*_t = -1\}} = 0, \quad \textit{for all } t \geq 0, \tag{28}$$

$\mathbb{P}^*_x$-a.s.. Under these assumptions,

$$J^{\mathrm{P}}(\mathbb{C}_{x,z}) \leq \limsup_{T\to\infty} \frac{1}{T}\left[-w_1(X_T) + \int_0^T \sigma(X_t)w_1'(X_t)\,dW_t\right], \quad \mathbb{P}_x\text{-a.s.,} \tag{29}$$

for all $\mathbb{C}_{x,z} \in \mathcal{C}_{x,z}$, and

$$J^{\mathrm{P}}(\mathbb{C}^*_{x,z}) = \limsup_{T\to\infty} \frac{1}{T}\left[-w_1(X^*_T) + \int_0^T \sigma(X^*_t)w_1'(X^*_t)\,dW^*_t\right], \quad \mathbb{P}^*_x\text{-a.s..} \tag{30}$$

Also,

$$\sup_{\mathbb{C}_{x,z}\in\mathcal{C}_{x,z}} J^{\mathrm{E}}(\mathbb{C}_{x,z}) = J^{\mathrm{E}}(\mathbb{C}^*_{x,z}) = \limsup_{T\to\infty} \frac{1}{T}\mathbb{E}^*_x\left[-w_1(X^*_T) + \int_0^T \sigma(X^*_t)w_1'(X^*_t)\,dW^*_t\right]. \tag{31}$$

Proof. Fix any initial condition $(x,z) \in \mathbb{R}\times\{0,1\}$, and consider any switching strategy $\mathbb{C}_{x,z} = (\mathbb{S}_x, Z) \in \mathcal{C}_{x,z}$. Using the generalised Itô's formula that is applicable for functions $w \in W^{2,\infty}_{\mathrm{loc}}(\mathbb{R})$ (e.g., see Krylov [K80, Theorem 2.10.1]) and the integration by parts formula, we calculate

$$\begin{aligned} Z_T w_1(X_T) &= z w_1(x) + \int_0^T \left[\tfrac{1}{2}\sigma^2(X_t)w_1''(X_t) + b(X_t)w_1'(X_t)\right] Z_t\,dt \\ &\quad + \sum_{t\in[0,T[} w_1(X_t)\Delta Z_t + \int_0^T \sigma(X_t)w_1'(X_t)Z_t\,dW_t, \\ (1-Z_T)w_0(X_T) &= (1-z)w_0(x) + \int_0^T \left[\tfrac{1}{2}\sigma^2(X_t)w_0''(X_t) + b(X_t)w_0'(X_t)\right](1-Z_t)\,dt \\ &\quad - \sum_{t\in[0,T[} w_0(X_t)\Delta Z_t + \int_0^T \sigma(X_t)w_0'(X_t)(1-Z_t)\,dW_t. \end{aligned}$$

With regard to the definitions (18) and (19) of the performance indices J^{P} and J^{E}, these calculations imply

$$J^{\mathrm{P}}(\mathbb{C}_{x,z}) = \limsup_{T\to\infty} \frac{1}{T} \left[\mathcal{Q}_T^{(1)} + \mathcal{Q}_T^{(2)} + \mathcal{Q}_T^{(3)} + \mathcal{Q}_T^{(4)} + \mathcal{Q}_T^{(5)} \right], \tag{32}$$

$$J^{\mathrm{E}}(\mathbb{C}_{x,z}) = \limsup_{T\to\infty} \frac{1}{T} \mathbb{E}_x \left[\mathcal{Q}_T^{(1)} + \mathcal{Q}_T^{(2)} + \mathcal{Q}_T^{(3)} + \mathcal{Q}_T^{(4)} + \mathcal{Q}_T^{(5)} \right], \tag{33}$$

where

$$\begin{aligned}
\mathcal{Q}_T^{(1)} &= -w_1(X_T) + \int_0^T \sigma(X_t) w_1'(X_t)\, dW_t, \\
\mathcal{Q}_T^{(2)} &= z w_1(x) + (1-z) w_0(x) - (1 - Z_T)\left[w_0(X_T) - w_1(X_T)\right], \\
\mathcal{Q}_T^{(3)} &= \int_0^T \sigma(X_t)\left[w_0'(X_t) - w_1'(X_t)\right](1 - Z_t)\, dW_t, \\
\mathcal{Q}_T^{(4)} &= \int_0^T \left[\tfrac{1}{2}\sigma^2(X_t) w_1''(X_t) + b(X_t) w_1'(X_t) + h(X_t)\right] Z_t\, dt \\
&\quad + \int_0^T \left[\tfrac{1}{2}\sigma^2(X_t) w_0''(X_t) + b(X_t) w_0'(X_t)\right](1 - Z_t)\, dt, \\
\mathcal{Q}_T^{(5)} &= \sum_{t\in[0,T[} \left[w_1(X_t) - w_0(X_t) - K\right] \mathbf{1}_{\{\Delta Z_t = 1\}} + \sum_{t\in[0,T[} \left[w_0(X_t) - w_1(X_t)\right] \mathbf{1}_{\{\Delta Z_t = -1\}}.
\end{aligned}$$

Assumption (24) implies

$$\lim_{T\to\infty} \frac{1}{T} \mathcal{Q}_T^{(2)} = \lim_{T\to\infty} \frac{1}{T} \mathbb{E}_x\left[\mathcal{Q}_T^{(2)}\right] = 0, \quad \mathbb{P}_x\text{-a.s.}, \tag{34}$$

and

$$\lim_{T\to\infty} \frac{1}{T} \mathcal{Q}_T^{(3)} = \lim_{T\to\infty} \frac{1}{T} \mathbb{E}_x\left[\mathcal{Q}_T^{(3)}\right] = 0, \quad \mathbb{P}_x\text{-a.s.}. \tag{35}$$

The limits in (34) are indeed obvious. To see (35), we first observe that the quadratic variation of the local martingale $\mathcal{Q}^{(3)}$ satisfies

$$\left\langle \mathcal{Q}^{(3)} \right\rangle_T = \int_0^T \left[\sigma(X_t)\left[w_0'(X_t) - w_1'(X_t)\right](1 - Z_t)\right]^2 dt \le CT, \tag{36}$$

where $C > 0$ is the constant appearing in (24). It follows that the stochastic integral $\mathcal{Q}^{(3)}$ is a square-integrable martingale, so

$$\mathbb{E}_x\left[\mathcal{Q}_T^{(3)}\right] = 0, \quad \text{for all } T \ge 0. \tag{37}$$

Furthermore, with regard to the Dambis, Dubins & Schwarz theorem (e.g., see Karatzas and Shreve [KS88, Theorem 3.4.6]), there exists a standard, one-dimensional Brownian motion

B defined on a possible extension of $(\Omega, \mathcal{F}, \mathbb{P}_x)$ such that $\mathcal{Q}_T^{(3)} = B_{\langle \mathcal{Q}^{(3)} \rangle_T}$. In view of this representation, the fact that $\lim_{T\to\infty} B_T/T = 0$, $\mathbb{P}_x$-a.s., and (36), we calculate

$$\begin{aligned}
\lim_{T\to\infty} \frac{1}{T} \left| \mathcal{Q}_T^{(3)} \right| &\le \lim_{T\to\infty} \left[\frac{1}{T} \left| B_{\langle \mathcal{Q}^{(3)} \rangle_T} \right| \mathbf{1}_{\{\langle \mathcal{Q}^{(3)} \rangle_\infty < \infty\}} + \frac{C}{\langle \mathcal{Q}^{(3)} \rangle_T} \left| B_{\langle \mathcal{Q}^{(3)} \rangle_T} \right| \mathbf{1}_{\{\langle \mathcal{Q}^{(3)} \rangle_\infty = \infty\}} \right] \\
&\le \lim_{T\to\infty} \left[\frac{1}{T} \sup_{t\in[0, \langle \mathcal{Q}^{(3)} \rangle_\infty]} |B_t| \mathbf{1}_{\{\langle \mathcal{Q}^{(3)} \rangle_\infty < \infty\}} + \frac{C}{\langle \mathcal{Q}^{(3)} \rangle_T} \left| B_{\langle \mathcal{Q}^{(3)} \rangle_T} \right| \mathbf{1}_{\{\langle \mathcal{Q}^{(3)} \rangle_\infty = \infty\}} \right] \\
&= 0.
\end{aligned}$$

However, these inequalities and (37) imply (35).

To proceed further, we note that, since w_1, w_0 satisfy the HJB equation (20)–(21),

$$\mathcal{Q}_T^{(4)} + \mathcal{Q}_T^{(5)} \le 0, \quad \text{for all } T \ge 0. \tag{38}$$

In view of this inequality, we can see that (32)–(33) and (34)–(35) imply (29) as well as

$$J^{\mathrm{E}}(\mathbb{C}_{x,z}) \le \limsup_{T\to\infty} \frac{1}{T} \mathbb{E}_x \left[-w_1(X_T) + \int_0^T \sigma(X_s) w_1'(X_s)\, dW_s \right]. \tag{39}$$

Finally, if $\mathbb{C}_{x,z}^*$ satisfies (25)–(28), then we can see that (38) holds with equality. Therefore, $J^{\mathrm{P}}(\mathbb{C}_{x,z}^*)$ satisfies (30), while $J^{\mathrm{E}}(\mathbb{C}_{x,z}^*)$ satisfies (39) with equality, and the proof is complete. $\square$

As we are going to see, the expressions on the right hand sides of (30) and (31) are both equal to the same constant. To this end, we are going to use the following result.

Lemma 2 *Let $\mathbb{S}_x$ be a weak solution to the SDE (1), and let $f : \mathbb{R} \to \mathbb{R}$ be any function satisfying $\int_{-\infty}^{\infty} |f(s)|\, m(ds) < \infty$. Also, suppose that the function $u \in W_{\mathrm{loc}}^{2,\infty}(\mathbb{R})$ satisfies*

$$\tfrac{1}{2}\sigma^2(x) u''(x) + b(x) u'(x) + f(x) = 0. \tag{40}$$

Then

$$\begin{aligned}
\lim_{T\to\infty} \frac{1}{T} \left[-u(X_T) + \int_0^T \sigma(X_t) u'(X_t)\, dW_t \right] &= \lim_{T\to\infty} \frac{1}{T} \mathbb{E}_x \left[-u(X_T) + \int_0^T \sigma(X_t) u'(X_t)\, dW_t \right] \\
&= \frac{1}{m(\mathbb{R})} \int_{-\infty}^{\infty} f(s)\, m(ds).
\end{aligned} \tag{41}$$

Proof. With regard to Itô's formula,

$$u(X_T) = u(x) + \int_0^T \left[\tfrac{1}{2}\sigma^2(X_t) u''(X_t) + b(X_t) u'(X_t) \right] dt + \int_0^T \sigma(X_t) u'(X_t)\, dW_t.$$

Since u satisfies (40), it follows that

$$\lim_{T\to\infty} \frac{1}{T} \left[-u(X_T) + \int_0^T \sigma(X_t) u'(X_t)\, dW_t \right] = \lim_{T\to\infty} \frac{1}{T} \int_0^T f(X_t)\, dt.$$

With regard to the ergodic Theorems V.53.1 and V.54.5 in Rogers and Williams [RW00], the limit appearing on the right hand side of this identity exists and is equal to the last expression in (41). □

5 The solution to the control problem

We can now solve the optimisation problems considered. Up to a point in our analysis below, we are going to consider solutions to the HJB equation (20)–(21) that are associated with switching strategies that are suggested by intuitive economics considerations in connection with the dynamic programming ideas discussed at the beginning of Section 4.

A first possibility arises if the operation of the investment project in its "active" mode is very profitable, so that the optimal strategy should keep the project in its "active" mode at all times (for a pictorial representation, see Figure 1). In this case, the optimality ideas

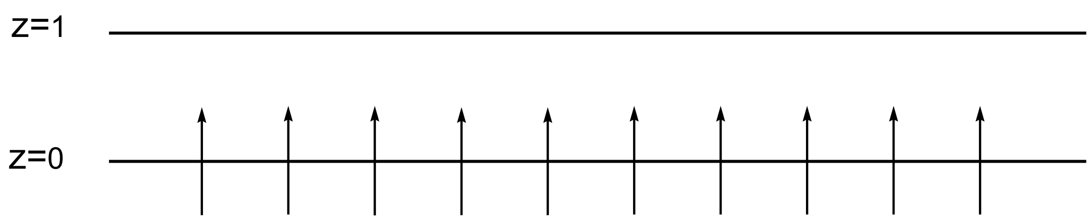

Figure 1: The case when it is optimal to keep the project in its "active" mode at all times.

discussed at the beginning of Section 4 suggest that we should look for a solution (w_1, w_0) to the HJB equation (20)–(21) that is characterised by

$$\tfrac{1}{2}\sigma^2(x) w_1''(x) + b(x) w_1'(x) + h(x) = 0, \quad \text{for all } x \in \mathbb{R},$$
$$w_0(x) = w_1(x) - K, \quad \text{for all } x \in \mathbb{R}.$$

It is straightforward to verify that every solution to these equations is given by

$$w_1(x) = w_0(x) + K = A + Bp(x) - \int_{x_0}^x p'(s) \int_{x_0}^s h(u)\, m(du)\, ds, \quad \text{for } x \in \mathbb{R}, \tag{42}$$

where $A, B \in \mathbb{R}$ and $x_0 \in \mathbb{R}$ are constants. Here, p and m are the scale function and the speed measure defined by (4) and (5), respectively. The following result is concerned with a necessary and sufficient condition for a choice of the functions w_1 and w_0 as in (42) to provide a solution to the HJB equation.

Lemma 3 *The functions w_1 and w_0 given by (42) satisfy the HJB equation (20)–(21) if and only if $h(x) \geq 0$, for all $x \in \mathbb{R}$.*

We collect in the Appendix the proofs of those results that are not developed in the text.

A similar case arises when it is optimal to always keep the project in its "passive" mode (see Figure 2). In this case, we look for a solution to (20)–(21) that satisfies

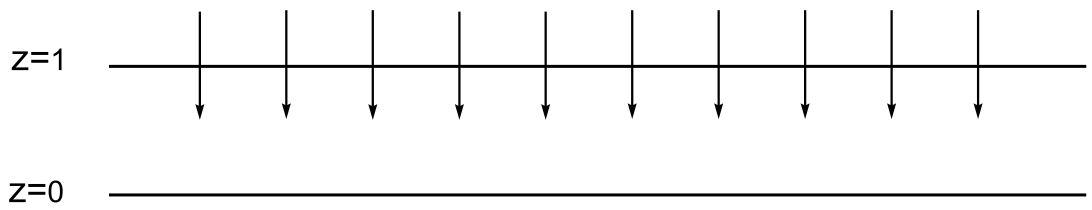

Figure 2: The case when it is optimal to keep the project in its "passive" mode at all times.

$$\tfrac{1}{2}\sigma^2(x)w_0''(x) + b(x)w_0'(x) = 0, \quad \text{for all } x \in \mathbb{R},$$
$$w_1(x) = w_0(x), \quad \text{for all } x \in \mathbb{R}.$$

Every such solution is given by

$$w_1(x) = w_0(x) = A + Bp(x), \quad \text{for } x \in \mathbb{R}, \tag{43}$$

for some constants $A, B \in \mathbb{R}$. A necessary and sufficient condition for these functions to satisfy the HJB equation is provided by the following result, the proof of which we omit because it is very similar to the proof of Lemma 3.

Lemma 4 *The functions w_1 and w_0 given by (43) satisfy the HJB equation (20)–(21) if and only if $h(x) \leq 0$, for all $x \in \mathbb{R}$.*

A more interesting case arises when the optimal strategy involves a sequence of switchings. In such a case, we can guess that the optimal strategy takes the form that can be depicted by Figure 3, and can be described as follows. Recalling that the running payoff function h is strictly increasing, we should keep the investment in its "active" mode for as long as the state process assumes sufficiently large values, and we should switch it to its "passive" mode as soon as the state process hits a given "low" level that we are going to denote by $\alpha \in \mathbb{R}$. On the other hand, we should keep the project in its "passive" mode for as long as the state process assumes sufficiently low values, and we should switch it to its "active" mode as soon as the state process rises to an appropriate "high" level that we denote by $\beta \in \mathbb{R}$. Of course,

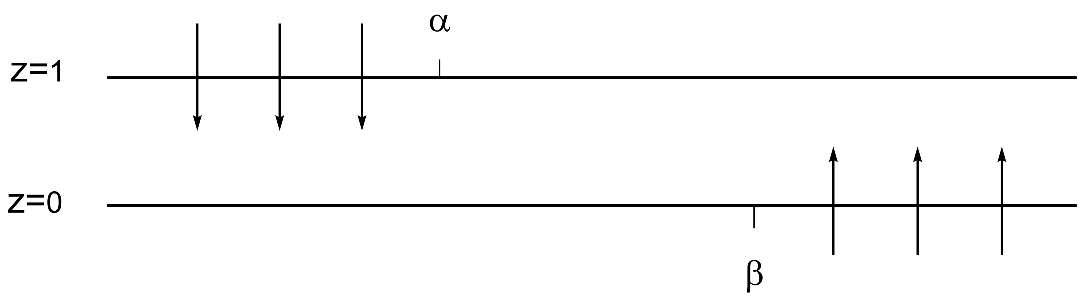

Figure 3: The case when it is optimal to switch sequentially.

for this strategy to be well-defined, we must have $\alpha < \beta$. In this case, we look for a solution to (20)–(21) that is characterised by

$$\begin{aligned}
w_0(x) - w_1(x) = 0, &\quad \text{for } x \in \,]-\infty, \alpha], && (44)\\
\tfrac{1}{2}\sigma^2(x) w_1''(x) + b(x) w_1'(x) + h(x) = 0, &\quad \text{for } x \in \,]\alpha, \infty[, && (45)\\
\tfrac{1}{2}\sigma^2(x) w_0''(x) + b(x) w_0'(x) = 0, &\quad \text{for } x \in \,]-\infty, \beta[, && (46)\\
w_1(x) - w_0(x) - K = 0, &\quad \text{for } x \in [\beta, \infty[. && (47)
\end{aligned}$$

To specify the parameters α and β, we appeal to the so-called *principle of smooth fit* that dictates that the functions w_1 and w_0 should be C^1 at the free boundary points α and β, respectively. To this end, we first observe that every solution to (46) is given by

$$w_0(x) = A + Bp(x), \quad \text{for } x \in \,]-\infty, \beta[, \tag{48}$$

where A and B are constants. Given such a solution, we can see that the only C^1 function w_1 satisfying (44)–(45) is given by

$$w_1(x) = \begin{cases} A + Bp(x), & \text{if } x \in \,]-\infty, \alpha], \\ A + Bp(x) - \int_\alpha^x p'(s) \int_\alpha^s h(u)\, m(du)\, ds, & \text{if } x \in \,]\alpha, \infty[. \end{cases} \tag{49}$$

Moreover, (48) and (47) imply that w_0 is given by

$$w_0(x) = \begin{cases} A + Bp(x), & \text{if } x \in \,]-\infty, \beta[, \\ A + Bp(x) - K - \int_\alpha^x p'(s) \int_\alpha^s h(u)\, m(du)\, ds, & \text{if } x \in [\beta, \infty[. \end{cases} \tag{50}$$

From this expression, we can see that w_0 will be C^1 if and only if the free boundary points $\alpha < \beta$ satisfy the system of equations

$$F(\alpha, \beta) = 0 \quad \text{and} \quad G(\alpha, \beta) = K, \tag{51}$$

where

$$F(\alpha,\beta) = \int_\alpha^\beta h(s)\, m(ds), \tag{52}$$

$$\begin{aligned} G(\alpha,\beta) &= -\int_\alpha^\beta p'(s) \int_\alpha^s h(u)\, m(du)\, ds \\ &= -\int_\alpha^\beta p'(s) F(\alpha,s)\, ds. \end{aligned} \tag{53}$$

For future reference, we note that

$$\begin{aligned} G(\alpha,\beta) &= -\int_\alpha^\beta \int_u^\beta p'(s) h(u)\, ds\, m(du) \\ &= \int_\alpha^\beta p(s) h(s)\, m(ds) - p(\beta) F(\alpha,\beta), \end{aligned} \tag{54}$$

the first identity following thanks to Fubini's theorem. In view of condition (13) in Assumption 3, $F(\alpha,\beta)$ is well-defined and finite for all choices of $\alpha, \beta \in [-\infty, \infty]$ such that $\alpha < \beta$. Also, $G(\alpha,\beta)$ is well-defined and finite for all $\alpha, \beta \in \mathbb{R}$ such that $\alpha < \beta$. However, we have to take care in all arguments involving limits such as $\lim_{\alpha \to -\infty} G(\alpha,\beta)$ or $\lim_{\beta \to \infty} G(\alpha,\beta)$ (see also the situation associated with Example 3 after Lemma 6 below).

Now, recalling that h is strictly increasing, we can see that there exist points $\alpha < \beta$ satisfying $F(\alpha,\beta) = 0$ only if

$$\lim_{x \to -\infty} h(x) < 0 < \lim_{x \to \infty} h(x), \tag{55}$$

which is a condition that complements the conditions required by the cases associated with Lemmas 3 and 4. For future reference, we also note that (55) and the assumption that h is strictly increasing imply that

$$\text{there exists a unique } \gamma \in \mathbb{R} \text{ such that } h(x) \begin{cases} < 0, & \text{for } x \in]-\infty, \gamma[, \\ > 0, & \text{for } x \in]\gamma, \infty[. \end{cases} \tag{56}$$

To proceed further, we define

$$\alpha^* = \sup \left\{ \alpha \in \mathbb{R} \mid \ F(\alpha, \infty) \equiv \int_\alpha^\infty h(s)\, m(ds) < 0 \right\} \tag{57}$$

and

$$\beta^* = \inf \left\{ \beta \in \mathbb{R} \mid F(-\infty, \beta) \equiv \int_{-\infty}^\beta h(s)\, m(ds) > 0 \right\}, \tag{58}$$

with the usual conventions $\sup \emptyset = -\infty$ and $\inf \emptyset = \infty$. In the presence of (56), we can see that

$$-\infty \le \alpha^* < \gamma < \beta^* \le \infty. \tag{59}$$

Moreover, given this definition and the monotonicity of h, we can check that

$$\int_{-\infty}^{\infty} h(s)\, m(ds) < 0 \quad \Leftrightarrow \quad \big(\alpha^* \in]-\infty, \gamma[\text{ and } \beta^* = \infty\big),$$
$$\int_{-\infty}^{\infty} h(s)\, m(ds) > 0 \quad \Leftrightarrow \quad \big(\alpha^* = -\infty \text{ and } \beta^* \in]\gamma, \infty[\big),$$

and

$$\int_{-\infty}^{\infty} h(s)\, m(ds) = 0 \quad \Leftrightarrow \quad \big(\alpha^* = -\infty \text{ and } \beta^* = \infty\big).$$

The following result provides a stepping stone for our subsequent analysis.

Lemma 5 *Suppose that (55) is true, and let γ, α^* and β^* be the points defined by (56), (57) and (58), respectively. There exists a unique, C^1 function $L :]\alpha^*, \gamma[\to]\gamma, \beta^*[$ such that $F(\alpha, L(\alpha)) = 0$. Moreover, this function satisfies*

$$\lim_{\alpha \downarrow \alpha^*} L(\alpha) = \beta^*, \quad \lim_{\alpha \uparrow \gamma} L(\alpha) = \gamma, \quad F(\alpha, x) \begin{cases} < 0, & \text{for all } x \in]\alpha, L(\alpha)[, \\ > 0, & \text{for all } x \in]L(\alpha), \infty[, \end{cases} \tag{60}$$

and

$$L'(\alpha) = \frac{\sigma^2(L(\alpha))p'(L(\alpha))h(\alpha)}{\sigma^2(\alpha)p'(\alpha)h(L(\alpha))} < 0. \tag{61}$$

We can now address the solvability of the system of equations given by (51).

Lemma 6 *The system of equations given by (51) has a solution (α, β) such that $-\infty < \alpha < \beta < \infty$ if and only if (55) is true and the constant G^* defined by*

$$G^* = \lim_{\alpha \downarrow \alpha^*} G(\alpha, L(\alpha)) > 0 \tag{62}$$

satisfies

$$G^* > K. \tag{63}$$

Under these conditions, the solution (α, β) is unique, and (49)–(50) define a solution (w_1, w_0) to the HJB equation (20)–(21) such that $w_1, w_0 \in W^{2,\infty}_{\mathrm{loc}}(\mathbb{R})$. Moreover, if we consider the constant $K > 0$ as a variable, then

$$\alpha = \alpha(K) \text{ is strictly decreasing}, \quad \lim_{K \downarrow 0} \alpha(K) = \gamma \quad \text{and} \quad \lim_{K \uparrow G^*} \alpha(K) = \alpha^*, \tag{64}$$
$$\beta = \beta(K) \text{ is strictly increasing}, \quad \lim_{K \downarrow 0} \beta(K) = \gamma \quad \text{and} \quad \lim_{K \uparrow G^*} \beta(K) = \beta^*. \tag{65}$$

In view of the fact that $\lim_{\alpha \downarrow \alpha^*} L(\alpha) = \beta^*$, it is tempting to replace (63) by $G(\alpha^*, \beta^*) > K$, which would result in a simpler restatement of Lemma 6. However, the following example shows that such a condition is not always well-defined.

Example 3 Suppose that

$$b(x) \equiv 0, \quad \sigma(x) = \sqrt{1+x^4} \quad \text{and} \quad h(x) = \begin{cases} -x^2, & \text{if } x \le 0, \\ x^2, & \text{if } x \ge 0. \end{cases}$$

For these choices of the problem's data, we can check that $p(x) = x$, all of the associated conditions in Assumptions 1, 2 and 3 hold, $\alpha^* = -\infty$ and $\beta^* = \infty$. Also, $\gamma = 0$, $L(\alpha) = -\alpha$, for all $\alpha < 0$, and, with regard to the expression for $G(\alpha, L(\alpha))$ provided by (73) in the Appendix,

$$G(\alpha, L(\alpha)) = 4 \int_0^{|\alpha|} \frac{s^3}{1+s^4}\, ds = \ln\left(1+\alpha^4\right), \quad \text{for } \alpha < \gamma = 0,$$

defines a strictly decreasing function. Since $\lim_{\alpha \to -\infty} G(\alpha, L(\alpha)) = \infty$, condition (63) is satisfied for any choice of the positive constant K. Now, with regard to (54), we calculate

$$G(\alpha, 0) = \tfrac{1}{2} \ln\left(1+\alpha^4\right) \quad \text{and} \quad G(0, \beta) = \tfrac{1}{2} \ln\left(1+\beta^4\right) - 2\beta \int_0^{\beta} \frac{s^2}{1+s^4}\, ds,$$

which show that

$$\lim_{\alpha \downarrow \alpha^* \equiv -\infty} G(\alpha, \beta) = \infty, \text{ for all } \beta \in \mathbb{R}, \quad \text{and} \quad \lim_{\beta \uparrow \beta^* \equiv \infty} G(\alpha, \beta) = -\infty, \text{ for all } \alpha \in \mathbb{R}.$$

However, these limits show that the expression $G(\alpha^*, \beta^*) \equiv G(-\infty, \infty)$ is *not* well-defined.

The cases considered up to now exhaust the range of candidates for the optimal strategy that arise from simple economic arguments (see Figures 1, 2 and 3). It is therefore tempting to assume that (63) is true for any positive value of K. However, the following example reveals that this is not in general the case.

Example 4 Suppose that

$$b(x) \equiv 0, \quad \sigma(x) = \sqrt{1+x^4} \quad \text{and} \quad h(x) = \begin{cases} \zeta(x-1), & \text{if } x \le 1, \\ x-1, & \text{if } x \ge 1, \end{cases}$$

where

$$\zeta = \frac{\int_1^\infty (s-1)(1+s^4)^{-1}\, ds}{\int_0^1 (1-s)(1+s^4)^{-1}\, ds}.$$

Plainly, all of the conditions in Assumptions 1, 2 and 3 are satisfied, $\alpha^* = 0$ and $\beta^* = \infty$. Furthermore, since

$$\lim_{\alpha \downarrow \alpha^*} G(\alpha, L(\alpha)) = 2\zeta \int_0^1 \frac{s(s-1)}{1+s^4}\, ds + 2 \int_1^\infty \frac{s(s-1)}{1+s^4}\, ds < \infty,$$

there exist values for K such that (63) is not satisfied.

When the assumptions of Lemmas 3, 4 and 6 are not satisfied, we cannot construct a solution to the HJB equation (20)–(21) that conforms with the heuristic considerations discussed at the beginning of Section 4 and that does not have a non-trivial transient nature. In this case, we indeed have to resort to a "variational" approach as in the proof of the following theorem that is our main result.

Theorem 7 *Fix any initial condition $(x, z) \in \mathbb{R} \times \{0, 1\}$, consider the problem of maximising the performance indices J^{P} and J^{E} defined by (18) and (19), respectively, over all admissible switching strategies in $\mathcal{C}_{x,z}$, and suppose that Assumptions 1, 2 and 3 are all satisfied. The following statements, in which, $\sup_{\mathbb{C}_{x,z} \in \mathcal{C}_{x,z}} J^{\mathrm{P}}(\mathbb{C}_{x,z})$ is understood as in (9)–(10), hold true:*

(i) *If $0 \leq h(x)$, for all $x \in \mathbb{R}$, then*

$$\begin{aligned} \sup_{\mathbb{C}_{x,z} \in \mathcal{C}_{x,z}} J^{\mathrm{P}}(\mathbb{C}_{x,z}; h, K) &= \sup_{\mathbb{C}_{x,z} \in \mathcal{C}_{x,z}} J^{\mathrm{E}}(\mathbb{C}_{x,z}; h, K) \\ &= \frac{1}{m(\mathbb{R})} \int_{-\infty}^{\infty} h(s)\, m(ds), \end{aligned} \tag{66}$$

and the switching strategy $(\mathbb{S}_x^*, Z^*)$, where $\mathbb{S}_x^*$ is a weak solution of (1) and Z^* is defined by $Z_t^* = z\mathbf{1}_{\{t=0\}} + \mathbf{1}_{\{t>0\}}$, is optimal.

(ii) *If $h(x) \leq 0$, for all $x \in \mathbb{R}$, then*

$$\sup_{\mathbb{C}_{x,z} \in \mathcal{C}_{x,z}} J^{\mathrm{P}}(\mathbb{C}_{x,z}; h, K) = \sup_{\mathbb{C}_{x,z} \in \mathcal{C}_{x,z}} J^{\mathrm{E}}(\mathbb{C}_{x,z}; h, K) = 0, \tag{67}$$

and the switching strategy $(\mathbb{S}_x^*, Z^*)$, where $\mathbb{S}_x^*$ is a weak solution of (1) and Z^* is defined by $Z_t^* = z\mathbf{1}_{\{t=0\}}$, is optimal.

(iii) *If $\lim_{x \to -\infty} h(x) < 0 < \lim_{x \to \infty} h(x)$ and (63) is true, then*

$$\begin{aligned} \sup_{\mathbb{C}_{x,z} \in \mathcal{C}_{x,z}} J^{\mathrm{P}}(\mathbb{C}_{x,z}; h, K) &= \sup_{\mathbb{C}_{x,z} \in \mathcal{C}_{x,z}} J^{\mathrm{E}}(\mathbb{C}_{x,z}; h, K) \\ &= \frac{1}{m(\mathbb{R})} \int_{\alpha}^{\infty} h(s)\, m(ds) \\ &= \frac{1}{m(\mathbb{R})} \int_{\beta}^{\infty} h(s)\, m(ds), \end{aligned} \tag{68}$$

where (α, β) is the unique solution of (51) derived in Lemma 6, and an optimal switching strategy can be constructed as in the proof below.

(iv) *If* $\lim_{x\to-\infty} h(x) < 0 < \lim_{x\to\infty} h(x)$ *and (63) is not true, then*

$$\sup_{\mathbb{C}_{x,z}\in\mathcal{C}_{x,z}} J^{\mathrm{P}}(\mathbb{C}_{x,z}; h, K) = \sup_{\mathbb{C}_{x,z}\in\mathcal{C}_{x,z}} J^{\mathrm{E}}(\mathbb{C}_{x,z}; h, K) = 0 \vee \frac{1}{m(\mathbb{R})}\int_{-\infty}^{\infty} h(s)\, m(ds). \tag{69}$$

In this case, if $\sup_{\mathbb{C}_{x,z}\in\mathcal{C}_{x,z}} J(\mathbb{C}_{x,z}; h, K) > 0$ *(resp.,* $\sup_{\mathbb{C}_{x,z}\in\mathcal{C}_{x,z}} J(\mathbb{C}_{x,z}; h, K) = 0$*), then a switching strategy that is optimal for case (i) (resp, case (ii)) above is optimal.*

Proof. In each of the cases (i)–(iii), $w_1, w_0 \in W^{2,\infty}_{\mathrm{loc}}(\mathbb{R})$. Assumption (24) in the verification Theorem 1 follows immediately in cases (i) and (ii), and can be verified in case (iii) by appealing to (12) in Assumption 3 and to the fact that w_1, w_0 are C^1. In cases (i)–(ii), the strategies postulated in the statement of the theorem clearly satisfy (25)–(28). With regard to case (iii), suppose that $z = 1$ and, given any initial condition $x \in \mathbb{R}$, let $\mathbb{S}^*_x = (\Omega^*, \mathcal{F}^*, \mathcal{F}^*_t, \mathbb{P}^*_x, W^*, X^*)$ be a weak solution of (1). If Z^* is the process defined by

$$Z^*_t = \mathbf{1}_{\{t=0\}} + \sum_{n=0}^{\infty} \mathbf{1}_{\{S_n<t\le T_n\}},$$

where $S_0 = 0$ and the $(\mathcal{F}^*_t)$-stopping times T_n and S_n, $n \in \mathbb{N}^*$, are defined recursively by

$$T_n = \inf\{t \ge S_n \mid X^*_t \le \alpha\}, \quad n = 0, 1, \ldots,$$
$$S_n = \inf\{t \ge T_n \mid X^*_t \ge \beta\}, \quad n = 1, 2, \ldots,$$

then we can check that $(\mathbb{S}^*_x, Z^*) \in \mathcal{C}_{x,1}$, and (25)–(28) are satisfied. If $z = 0$, an admissible switching strategy satisfying (25)–(28) can be constructed in a similar way. These observations show that all of the requirements of Theorem 1 are satisfied in cases (i)–(iii), which establishes (29), (30) and (31), as well as the optimality of the associated switching strategies.

Now, (66) and (67) follow immediately from (29)–(31) and Lemma 2. In case (iii), w_1 satisfies

$$\tfrac{1}{2}\sigma^2(x)w_1''(x) + b(x)w_1'(x) + \mathbf{1}_{]\alpha,\infty[}(x)h(x) = 0,$$

by construction. Combining this observation with (29)–(31) and Lemma 2, we can see that the second equality in (68) is true. The last equality in (68) follows from the first one and the fact that $F(\alpha, \beta) = 0$, where F is defined by (52).

To prove (iv), assume that $\lim_{x\to-\infty} h(x) < 0 < \lim_{x\to\infty} h(x)$ and that $G^* \in]0, \infty[$, where G^* is defined by (63). Also, fix any $K \ge G^*$, and denote by J either of the performance indices J^{P} or J^{E}. A simple inspection of (18) and (19) that define J^{P} and J^{E}, respectively, reveals that $J(\mathbb{C}_{x,z}; h, K_1) \le J(\mathbb{C}_{x,z}; h, K_2)$, for all $K_1 > K_2$, for all $\mathbb{C}_{x,z} \in \mathcal{C}_{x,z}$. It follows that, given any $\mathbb{C}_{x,z} \in \mathcal{C}_{x,z}$,

$$J(\mathbb{C}_{x,z}; h, K) \le J(\mathbb{C}_{x,z}; h, \hat{K}) \le \frac{1}{m(\mathbb{R})}\int_{\alpha(\hat{K})}^{\infty} h(s)\, m(ds), \quad \text{for all } \hat{K} \in]0, G^*[,$$

the second inequality following from case (iii) that we established above. In view of (64), the dominated convergence theorem, and the definition (57) of α^*, we can pass to the limit $\hat{K} \uparrow G^*$ in these inequalities to obtain

$$\begin{aligned} J(\mathbb{C}_{x,z}; h, K) &\leq \frac{1}{m(\mathbb{R})} \int_{\alpha^*}^{\infty} h(s)\, m(ds) \\ &= \begin{cases} \frac{1}{m(\mathbb{R})} \int_{-\infty}^{\infty} h(s)\, m(ds), & \text{if } \alpha^* = -\infty, \\ 0, & \text{if } \alpha^* > -\infty, \end{cases} \end{aligned} \tag{70}$$

for all $\mathbb{C}_{x,z} = (\mathbb{S}_x, Z)$. Now, if $\mathbb{C}^*_{x,z} = (\mathbb{S}^*_x, Z^*)$ is the optimal switching strategy considered in case (i) or case (ii) of this theorem, depending on whether $\alpha^* = -\infty$ or $\alpha^* > -\infty$, then

$$J(\mathbb{C}^*_{x,z}; h, K) = \begin{cases} \frac{1}{m(\mathbb{R})} \int_{-\infty}^{\infty} h(s)\, m(ds), & \text{if } \alpha^* = -\infty, \\ 0, & \text{if } \alpha^* > -\infty, \end{cases}$$

which, combined with (70), establishes all of the claims made in case (iv), and the proof is complete. □

Remark 1 It is worth noting that, although we have focused on conditions such as (63) that is expressed in terms of the point α^* defined by (57), we can indeed develop a totally symmetric and equivalent analysis based on conditions involving the point β^* defined by (58).

6 Appendix: Proofs of selected results

Proof of Lemma 3. With regard to their construction, w_1 and w_0 satisfy (20)–(21) if and only if

$$w_0(x) - w_1(x) \leq 0, \quad \text{for all } x \in \mathbb{R}, \tag{71}$$

$$\tfrac{1}{2}\sigma^2(x) w_0''(x) + b(x) w_0'(x) \leq 0, \quad \text{for all } x \in \mathbb{R}. \tag{72}$$

Plainly, (71) is equivalent to $K \geq 0$, which is implied by Assumption (15). Also, we can check that (72) is equivalent to $h(x) \geq 0$, for all $x \in \mathbb{R}$, which completes the proof. □

Proof of Lemma 5. Given $\alpha \in]\alpha^*, \gamma[$, we consider the function $\bar{F}_{[\alpha]} : [\alpha, \infty[\to \mathbb{R}$ that is defined by $\bar{F}_{[\alpha]}(\beta) = F(\alpha, \beta)$. The calculation

$$\bar{F}'_{[\alpha]}(\beta) = \frac{2h(\beta)}{\sigma^2(\beta) p'(\beta)} \begin{cases} < 0, & \text{if } \beta \in]\alpha, \gamma[, \\ > 0, & \text{if } \beta \in]\gamma, \infty[, \end{cases}$$

shows that $\bar{F}_{[\alpha]}$ is strictly decreasing in $]\alpha, \gamma[$ and strictly increasing in $]\gamma, \infty[$. Combining this observation with $\bar{F}_{[\alpha]}(\alpha) = 0$ and the definitions (57) and (58) of α^* and β^*, respectively, we can see that there exists a unique function $L :]\alpha^*, \gamma[\to]\gamma, \beta^*[$ such that $\bar{F}_{[\alpha]}(L(\alpha)) \equiv F(\alpha, L(\alpha)) = 0$ and (60) are true. Moreover, if $-\infty < \alpha^*$, then $F(\alpha, \beta) = 0$ has no solution $\beta \in]\alpha, \infty[$ if $\alpha \in]-\infty, \alpha^*]$. Finally, differentiation of $F(\alpha, L(\alpha)) = 0$ with respect to α yields (61), the inequality there following from (56) and the fact that $\alpha < \gamma < L(\alpha)$. □

Proof of Lemma 6. In view of Lemma 5, the system of equations given by (51) has a unique solution (α, β) such that $-\infty < \alpha < \beta < \infty$ if and only if the equation $G(\alpha, L(\alpha)) = K$ has a unique solution $\alpha \in]\alpha^*, \gamma[$. Now, with regard to (54), we can see that

$$G(\alpha, L(\alpha)) = \int_{\alpha}^{L(\alpha)} p(s)h(s)\, m(ds). \tag{73}$$

Recalling the definition (5) of the speed measure m, this expression and (61) imply

$$\frac{d}{d\alpha} G(\alpha, L(\alpha)) = \frac{2h(\alpha)\,[p(L(\alpha)) - p(\alpha)]}{\sigma^2(\alpha)p'(\alpha)} < 0, \quad \text{for all } \alpha \in]\alpha^*, \gamma[,$$

the inequality following because $h(\alpha) < 0$ for $\alpha < \gamma$, $L(\alpha) > \alpha$, and p is strictly increasing. However, this calculation shows that the function $G(\cdot, L(\cdot)) :]\alpha^*, \gamma[\to \mathbb{R}$ is strictly decreasing. Combining this observation with

$$\lim_{\alpha \uparrow \gamma} G(\alpha, L(\alpha)) = G(\gamma, \gamma) = 0,$$

which follows from (73) and the second limit in (60), we can conclude that the constant G^* defined by (62) is strictly positive and that the system of equations (51) has a unique solution of the required form if and only of (63) is true.

Now, in the presence of (55) and (63), suppose that the switching cost $K > 0$ is an independent variable, and consider the solution (α, β) of (51) as a function of K. The limits in (64) follow from the arguments that we used above in this proof to identify $\alpha = \alpha(K)$ with the solution to $G(\alpha, L(\alpha)) = K$. Also, the limits in (65) regarding $\beta = \beta(K) = L(\alpha(K))$ follow from the ones in (64) and the ones in (60). To establish the monotonicity properties of $\alpha(\cdot)$ and $\beta(\cdot)$, we first differentiate equation $F(\alpha(K), \beta(K)) = 0$ with respect to K to calculate

$$\beta'(K) = \frac{\sigma^2(\beta(K))p'(\beta(K))h(\alpha(K))}{\sigma^2(\alpha(K))p'(\alpha(K))h(\beta(K))}\alpha'(K). \tag{74}$$

Differentiating the equation $G(\alpha(K), \beta(K)) = K$ with respect to K, and using this expression, we obtain

$$\alpha'(K) = \frac{\sigma^2(\alpha(K))p'(\alpha(K))}{2h(\alpha(K))\,[p(\beta(K)) - p(\alpha(K))]} < 0,$$

the inequality following thanks to (56), and the facts that $\alpha(K) < \gamma < \beta(K)$ and p is strictly increasing. However, this inequality proves that $\alpha(\cdot)$ is strictly decreasing. Moreover, this calculation, combined with (74), implies $\beta'(K) > 0$, which proves that $\beta(\cdot)$ is strictly increasing.

To complete the proof, it remains to show that, assuming that (55) and (63) hold, (w_1, w_0) given by (49)–(50), where α and β are the unique solution to (51), solve the HJB equation (20)–(21), which amounts to proving that

$$\frac{1}{2}\sigma^2(x)w_1''(x) + b(x)w_1'(x) + h(x) \le 0, \quad \text{for } x < \alpha, \tag{75}$$

$$w_0(x) - w_1(x) \le 0, \quad \text{for } x \ge \alpha, \tag{76}$$

$$w_1(x) - w_0(x) - K \le 0, \quad \text{for } x \le \beta, \tag{77}$$

$$\frac{1}{2}\sigma^2(x)w_0''(x) + b(x)w_0'(x) \le 0, \quad \text{for } x > \beta. \tag{78}$$

By construction, (75) is equivalent to $h(x) \le 0$, for $x < \alpha$, which is true in the light of (56) and the fact that $\alpha < \gamma$. Similarly, (78) is equivalent to $h(x) \ge 0$, for $x > \beta$, which is implied by (56) and the fact that $\beta > \gamma$.

Either of (76) with $x \ge \beta$ or (77) with $x \le \alpha$ is equivalent to $-K \le 0$, which is implied by (15) in Assumption 3. In view of (49)–(50) and (53), we can see that (76) and (77) for $\alpha \le x \le \beta$ will follow if we show that

$$0 \le G(\alpha, x) \le K, \quad \text{for } x \in [\alpha, \beta]. \tag{79}$$

In the light of (53) and the last assertion in (60), we can see that

$$\frac{\partial}{\partial x}G(\alpha, x) = -p'(x)F(\alpha, x) > 0, \quad \text{for } x \in \,]\alpha, \beta[,$$

which shows that the function $G(\alpha, \cdot)$ is strictly increasing in $]\alpha, \beta[$. However, if we combine this observation with the equalities $G(\alpha, \alpha) = 0$ and $G(\alpha, \beta) = K$, we can see that (79) is true, and the proof is complete. □

Acknowledgement

We are grateful to the anonymous referee whose comments lead to a most significant improvement of the paper.

References

[BF92] A. Bensoussan and J. Frehse (1992), On Bellman equations of ergodic control in $\mathbb{R}^n$, *Journal für die Reine und Angewandte Mathematik*, vol. **429**, pp. 125–160.

[B99] V. S. Borkar (1999), The Value Function in Ergodic Control of Diffusion Processes with Partial Observations, *Stochastics and Stochastics Reports*, vol. **67**, pp. 255–266.

[BG88] V. S. Borkar and M. K. Ghosh (1988), Ergodic Control of Multidimensional Diffusions I: The Existence Results, *SIAM Journal on Control and Optimization*, vol. **26**, pp. 112–126.

[BØ91] K. A. Brekke and B. Øksendal (1991), The High Contact Principle as a Sufficiency Condition for Optimal Stopping, in *Stochastic Models and Option Values*, D. Lund and B. Øksendal, eds., North-Holland, pp. 187–208.

[BØ94] K. A. Brekke and B. Øksendal (1994), Optimal Switching in an Economic Activity under Uncertainty, *SIAM Journal on Control and Optimization*, **32**, pp. 1021–1036.

[BS85] M. J. Brennan and E. S. Schwartz (1985), Evaluating Natural Resource Investments, *Journal of Business*, **58**, pp. 135–157.

[DDT01] P. Dai Pra, G. B. Di Masi and B. Trivellato (2001), Pathwise Optimality in Stochastic Control, *SIAM Journal of Control and Optimization*, vol. **39**, pp. 1540–1557.

[DRT04] P. Dai Pra, W. J. Runggaldier and M. Tolotti (2004) Pathwise optimality for benchmark tracking, *IEEE Transactions on Automatic Control*, vol. **49**, pp. 386–395.

[D89] A. Dixit (1989), Entry and Exit Decisions under Uncertainty, *Journal of Political Economy*, **97**, pp. 620–638.

[DP94] A. K. Dixit and R. S. Pindyck (1994), *Investment under Uncertainty*, Princeton University Press.

[DZ01] K. Duckworth and M. Zervos (2001), A Model for Investment Decisions with Switching Costs, *The Annals of Applied Probability*, **11**, pp. 239–260.

[DMP98] T. E. Duncan, B. Maslowski, B. Pasik-Duncan, (1998), Ergodic boundary/point control of stochastic semilinear systems, *SIAM Journal on Control and Optimization* vol. **36**, pp. 1020–1047.

[GS90] D. Gatarek and L. Stettner (1990), On the compactness method in general ergodic impulsive control of Markov processes, *Stochastics & Stochastics Reports*, vol. **31**, pp. 15–25.

[K83] I. Karatzas (1983), A Class of Singular Stochastic Control Problems, *Advances in Applied Probability*, vol. **15**, pp. 225–254.

[KS88] I. Karatzas and S. E. Shreve (1988), *Brownian Motion and Stochastic Calculus*, Springer-Verlag.

[K00] L. Kruk (2000), Optimal policies for n-dimensional singular stochastic control problems. II. The radially symmetric case. Ergodic control, *SIAM Journal on control and Optimization*, vol. **39**, pp. 635–659.

[K80] N. V. Krylov (1980), *Controlled Diffusion Processes*, Springer-Verlag.

[KS98] T. G. Kurtz and R. H. Stockbridge (1998), Existence of Markov Controls and Characterization of Optimal Markov Controls, *SIAM Journal on Control and Optimization*, vol. **36**, pp. 609–653.

[K78] H. J. Kushner (1978), Optimality Conditions for the Average Cost per Unit Time Problem with a Diffusion Model, *SIAM Journal on Control and Optimization*, vol. **16**, pp. 330–346.

[LZ01] R. R. Lumley and M. Zervos (2001), A Model for Investments in the Natural Resource Industry with Switching Costs,*Mathematics of Operations Research*, **26**, pp. 637–653.

[MRT92] J. L. Menaldi, M. Robin and M. I. Taksar (1992), Singular ergodic control for multidimensional Gaussian processes, *Mathematics of Control, Signals, and Systems*, vol. **5**, pp. 93–114.

[PRT93] E. L. Presman, V. I. Rotar, M. Taksar (1993), Optimality in Probability and Almost Surely. The General Scheme and a Linear Quadratic Regulator Problem, *Stochastics and Stochastic Reports*, vol. **43**, pp. 127–137.

[RW00] L. C. G. Rogers and D. Williams (2000), *Diffusions, Markov Processes and Martingales*, volume 2, Cambridge University Press.

[R91] V. I. Rotar (1991), Connectivity property and optimality almost surely and in probability, in *New Trends in Probability and Statistics*, VSP Utrecht, pp. 528–539.

[SS02] R. Sadowy and L. Stettner (2002), On risk-sensitive ergodic impulsive control of Markov processes, *Applied Mathematics and Optimization*, vol. **45** , pp. 45–61.

[W05] H. Wang (2005), Optimal stopping with Forced exits, *Mathematics of Operations Research*, to appear.