AD-A109 811 STANFORD UNIV CA F/G 12/2

IMPULSE CONTROL OF BROWNIAN MOTION.(U)

NOV 81 J M HARRISON, T M SELLKE, A J TAYLOR N00014-75-C-0561

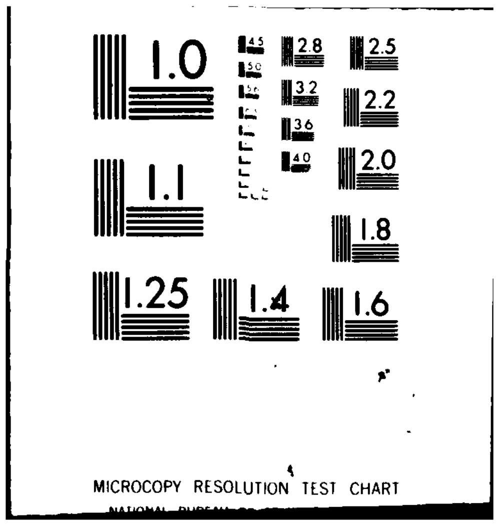
UNCLASSIFIED TR-202 NL

END

DATE FILMED

2-82

DTIC


1.0
2.8
2.5
3.2
2.2
3.6
4.0
2.0
1.1
1.8
1.25
1.4
1.6
MICROCOPY RESOLUTION TEST CHART

AD A109811

(12) LEVEL

IMPULSE CONTROL OF BROWNIAN MOTION

by

J. MICHAEL HARRISON
THOMAS M. SELLKE
and
ALLISON J. TAYLOR

TECHNICAL REPORT NO. 202

November 1981

SUPPORTED UNDER CONTRACT N00014-75-C-0561 (NR-047-200)
WITH THE OFFICE OF NAVAL RESEARCH

Gerald J. Lieberman, Project Director

DEPARTMENT OF OPERATIONS RESEARCH
AND
DEPARTMENT OF STATISTICS
STANFORD UNIVERSITY
STANFORD, CALIFORNIA

IMPULSE CONTROL OF BROWNIAN MOTION

J. Michael Harrison, Stanford University
Thomas M. Sellke, Stanford University
Allison J. Taylor, Queen's University

Abstract

Consider a storage system, such as an inventory or cash fund, whose content fluctuates as a (μ, σ^2) Brownian motion in the absence of control. Holding costs are continuously incurred at a rate proportional to the storage level, and we may cause the storage level to jump by any desired amount at any time except that the content must be kept nonnegative. Both positive and negative jumps entail fixed plus proportional costs, and our objective is to minimize expected discounted costs over an infinite planning horizon. A control band policy is one that enforces an upward jump to q whenever level zero is hit, and enforces a downward jump to Q whenever level S is hit $(0 < q < Q < S)$. We prove the existence of an optimal control band policy and calculate explicitly the optimal values of the critical numbers (q,Q,S).

Key Words and Phrases

Brownian Motion, Stochastic Control, Jump Boundaries, Inventory and Production Control, Impulse Control, Stochastic Cash Management.

IMPULSE CONTROL OF BROWNIAN MOTION

J. Michael Harrison, Stanford University
Thomas M. Sellke, Stanford University
Allison J. Taylor, Queen's University

1. Introduction and Summary

Consider a controller who continuously monitors the content, or state, of a storage system. In the absence of control, the content process $Z = \{Z_t, t \geq 0\}$ fluctuates as a Brownian motion with drift μ and variance σ^2. The controller can at any time increase or decrease the content of the system by any amount desired, but he is obliged to keep $Z_t \geq 0$, and there are three types of cost to be considered.

(1.1) In order to effect an increase from level x to level $x+\delta$, the controller must pay a fixed charge K plus a proportional charge $k\delta$.

(1.2) Similarly, it costs $L+\ell\delta$ to effect a decrease from level x to level $x-\delta$.

(1.3) Inventory holding costs are continuously incurred at rate hZ_t.

Thus we have linear holding costs and fixed plus proportional costs of control. We seek a policy that will minimize, subject to the constraint $Z_t \geq 0$, the expected present value of holding costs and control costs incurred over an infinite planning horizon, where future costs are continuously discounted at interest rate $\gamma > 0$.

For a concrete application, one may consider the so-called stochastic cash management problem. Here Z_t represents the content at time t of a cash fund, into which a certain amount of income or revenue is automatically channelled and out of which operating disbursements are made. Interpret h as the opportunity loss rate for cash held within the fund, meaning that h is the amount of income per period that could have been earned by a dollar of cash if it had been invested in securities. When the content of the cash fund gets too large, the controller may choose to convert some of his cash into securities, and for this he pays a fixed transaction cost K plus a proportional cost of k times the transaction size. On the other hand, he may at any time convert securities into cash, this too involving fixed plus proportional transaction costs.

It is more or less obvious that there exists for this problem an optimal policy of the type pictured in Figure 1. Using the language of the stochastic cash balance problem, this control band policy may be described as follows. First, it is characterized by three parameters (q, Q, S) satisfying $0 < q < Q < S$, and for future reference we define

(1.4) $$s \equiv S-Q \quad \text{and} \quad \Delta \equiv Q-q \ .$$

Whenever the content Z of the cash fund hits zero, the controller liquidates q dollars of securities, incurring a transaction cost of $K+kq$. (He never liquidates securities except when it is necessary to maintain a positive cash balance.) On the other hand, whenever the content of the fund reaches an upper limit S, the controller buys s dollars of securities, thus reducing his cash balance to Q and incurring a total transaction cost of $L+\ell s$. (If the initial cash balance exceeds S, the controller immediately buys enough securities to reduce this balance to Q.)

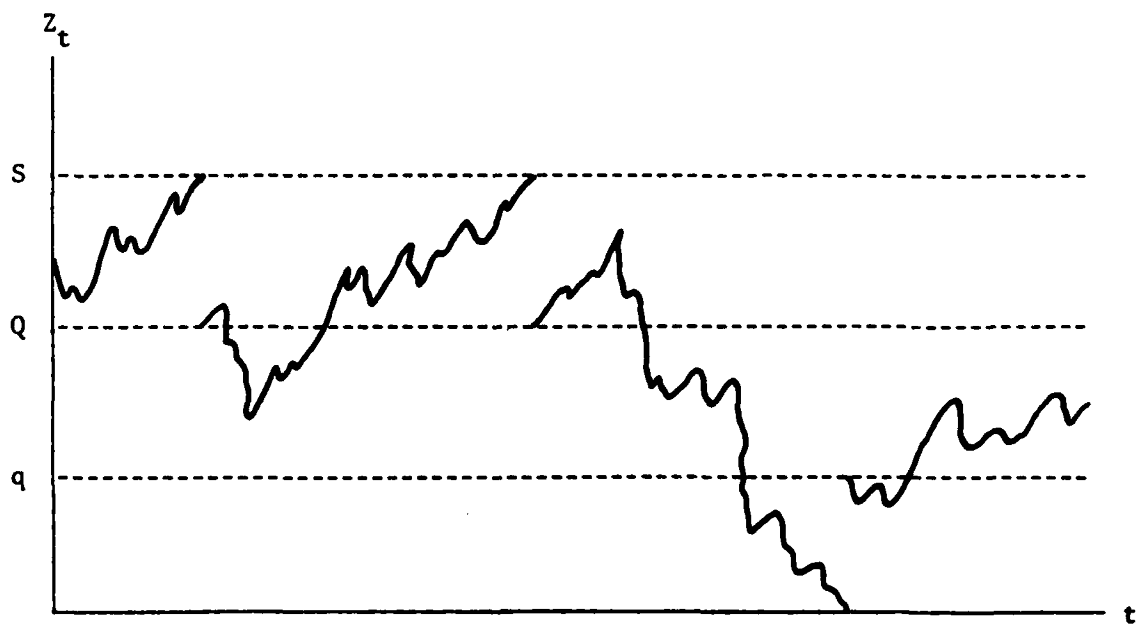


Figure 1. A Control Band Policy

Assuming that K and L are strictly positive, as we shall do throughout, our problem is one of impulse control [4]. This means that the controller exerts his influence through lump sum displacements effected at isolated points in time. The impulse control problem is quite easy to formulate in precise mathematical terms, and we shall do this shortly. Harrison and Taylor [6] have studied the analagous problem with proportional control costs only $(K = L = 0)$, which requires a more subtle formulation but is easier to solve explicitly. With this cost structure, it was shown that the optimal control policy enforces an upper reflecting barrier at Q and a lower reflecting barrier at zero, where Q is the unique solution of a certain transcendental equation. Roughly speaking, this barrier policy is the limit, as s and q both approach zero, of this control band policy pictured in Figure 1. The controller exerts influence at uncountably many time points, but the total amount of upward or downward displacement effected during any finite period is finite. The controller can obviously effect instantaneous state changes in this problem, so the state constraint $Z_t \geq 0$ makes sense, and yet policies cannot be described through a discrete sequence of intervention times. Harrison and Taksar [7] have coined the term instantaneous control to describe that state of affairs, and the interested reader may see [7] and [3] for analyses of other such problems.

Harrison and Taylor [6] also considered the case where $K > 0$ and $L = 0$, obtaining an optimal policy that imposes an upper reflecting barrier at S and enforces an upward jump to q whenever level zero is hit $(0 < q < S)$. Such a policy can be obtained by letting $s \to 0$ in the control band policy of Figure 1. Finally, Constantinides and Richard [5] have studied a Brownian impulse control problem more general than ours. (To be more precise, our problem can be obtained by letting a certain cost parameter approach ∞ in their formulation.) They prove the existence of a structured optimal policy but do not show how to compute its critical numbers except for certain simple special cases.

In this paper we show that an optimal control band policy exists for the impulse control problem, and we determine explicitly the optimal policy parameters (q, Q, S). The optimal policies of Harrison and Taylor [6] can be obtained by letting one or both of the fixed control costs approach zero in our formulas. In addition, our mathematical development is cleaner and more nearly self-contained than that in [6], and we give better economic and probabilistic interpretations for our results. To briefly summarize those results, let us first define

$$(1.4) \qquad c = h/\gamma + k \qquad \text{and} \qquad r = h/\gamma - \ell \ .$$

It will be shown that the original problem is completely equivalent to another impulse control problem with the following cost structure.

(1.5) When an upward jump of size δ is effected, the controller incurs a fixed cost K plus a proportional cost $c\delta$.

(1.6) When a downward jump of size δ is effected, the controller incurs a fixed cost L but earns a proportional reward $r\delta$.

(1.7) There are no holding costs.

To understand this equivalence, note first that h/γ is the discounted cost of holding one unit of stock in inventory forever. Under the cost structure (1.5)-(1.7), our controller is charged this full infinite-horizon holding cost for each unit of stock that he introduces into the system, he is credited with a refund of equal size each time he removes a unit of stock from the system, and no holding costs are incurred in the interim. Except for certain uncontrollable terms, this cost structure is found to be identical to the original one, where holding costs are charged continuously according to the current stock on hand. To avoid uninteresting degeneracies, we assume throughout that

$$0 < r < c < \infty \,. \tag{1.8}$$

For any choice of policy parameters satisfying $0 < q < Q < S$, there exists a unique function π on $[0,S]$ satisfying

$$\frac{1}{2}\sigma^2 \pi''(x) + \mu\pi'(x) - \gamma\pi(x) = 0 \,, \qquad 0 \leq x \leq S \,, \tag{1.9}$$

$$\pi(Q) = \pi(S) = r \ , \tag{1.10}$$

and one can furthermore write out an explicit and relatively simple formula for π in terms of the policy parameters Q and S. The function π is strictly convex, with a minimum between Q and S, and there is exactly one choice of the policy parameters (q, Q, S) such that

$$\int_Q^S [r-\pi(x)]dx = L \ , \tag{1.11}$$

$$\pi(q) = c \ , \tag{1.12}$$

and

$$\int_0^q [\pi(x)-c]dx = K \ , \tag{1.13}$$

as depicted in Figure 2 below. These are the parameters of the optimal control band policy, and the associated function π is the derivative of the optimal value function. It will be shown that (1.11) alone determines $s \equiv S-Q$, after which (1.12) determines $\Delta \equiv Q-q$, and then (1.13) determines q. This three-step algorithm for determination of the optimal parameters will be written out in algebraic form, and interpretations of the conditions (1.11)-(1.13) will be given.

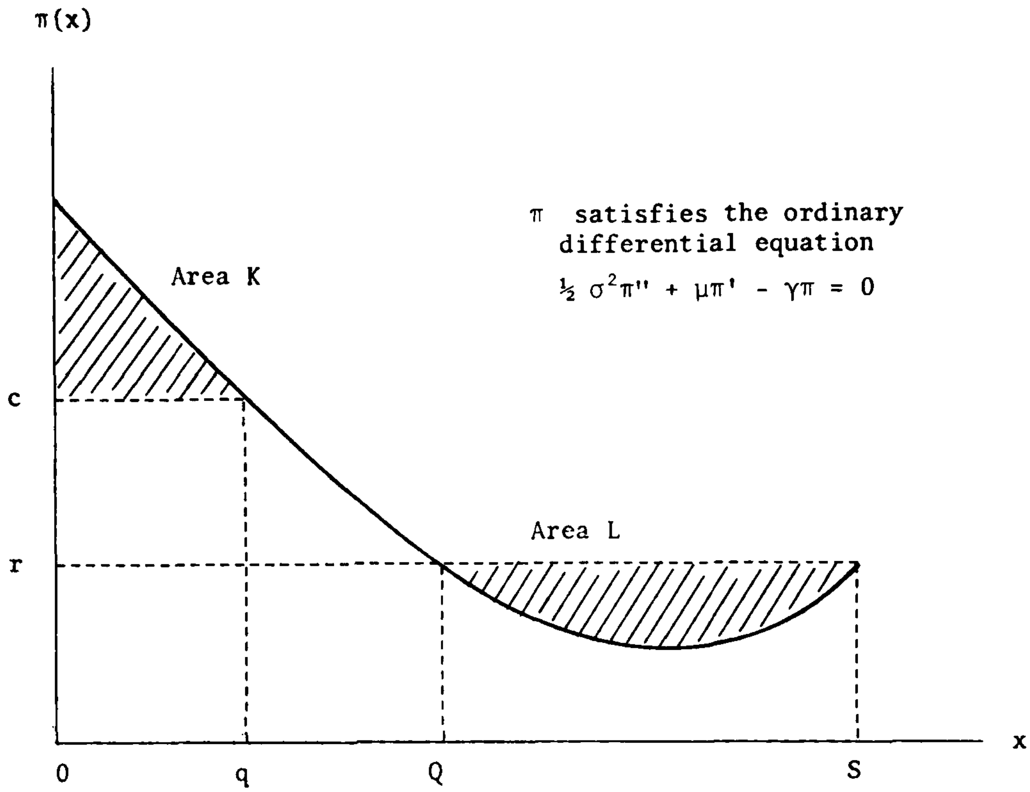


Figure 2: Optimal Policy Parameters

The paper is organized as follows. In §2 we give a precise formulation of our impulse control problem, prove equivalence of the cost structures (1.1)-(1.3) and (1.5)-(1.7), and lay out some other preliminary propositions. Section 3 is devoted to characterization of control band policies. In §4 we show that there exists a unique set of policy parameters (q,Q,S) satisfying a certain set of conditions, and we rigorously prove the optimality of the corresponding control band policy. Finally, §5 develops interpretations for the optimality conditions taken as primitive in §4.

2. Problem Formulation and Preliminaries

The data for our problem are a drift parameter μ, a variance parameter $\sigma^2 > 0$, fixed control costs $K > 0$ and $L > 0$, proportional control cost rates k and ℓ, a holding cost rate h, and an interest rate $\gamma > 0$. Defining $c = h/\gamma + k$ and $r = h/\gamma - \ell$, we assume throughout that $0 < r < c$.

Let Ω be the space of all continuous functions $\omega : [0,\infty) \to R$ (the real line). For $t \geq 0$ let $X_t : \Omega \to R$ be the coordinate projection map $X_t(\omega) = \omega(t)$. Then $X = (X_t, t \geq 0)$ is simply the identity map $\Omega \to \Omega$. Let $F = \sigma(X_t, t \geq 0)$ denote the smallest σ-field such that X_t is F-measurable for each $t \geq 0$, and similarly let $F_t = \sigma(X_s, 0 \leq s \leq t)$ for $t \geq 0$. Hereafer, when we speak of adapted processes and stopping times, the underlying information structure (filtration) is understood to be $(F_t, t \geq 0)$. Finally, for each $x \in R$ let P_x be the unique probability measure on (Ω, F) such that X is a Brownian motion with drift μ, variance σ^2 and starting state x under P_x. Let E_x be the associated expectation operator.

A policy consists of a sequence of stopping times $\{T_0, T_1, \ldots\}$ and a sequence of random variables $\{\xi_0, \xi_1, \ldots\}$ such that

(2.1) $\quad P_x(0 = T_0 < T_1 < \cdots \to \infty) = 1 \,, \quad$ for all $x \in R$,

(2.2) $\quad \xi_n \in F_{T_n} \,, \quad$ for all $n = 0, 1, \ldots$.

Interpret T_n as the n^{th} time at which the controller enforces a jump in the state of the system, with ξ_n the size of the jump (either positive or negative) enforced. The convention $T_0 = 0$ will prove to be convenient, but then we must of course allow $\xi_0 = 0$. We associate with a policy $\{(T_n, \xi_n)\}$ the processes

$$N(t) = \sup\{n \geq 0 : T_n \leq t\}, \qquad t \geq 0,$$

$$Y_t = \xi_1 + \cdots + \xi_{N(t)}, \qquad t \geq 0,$$

$$Z_t = X_t + Y_t, \qquad t \geq 0.$$

(The time parameter of a given process may be written either as a subscript or as a functional argument, depending on which is more convenient.) Note that N, Y and Z are all adapted and right continuous with left limits. The policy $\{(T_n,\xi_n)\}$ is said to be feasible if

$$\text{(2.3)} \qquad P_x(Z_t \geq 0 \text{ for all } t \geq 0) = 1, \qquad \text{for all } x \in R,$$

$$\text{(2.4)} \qquad E_x\Big(\sum_{n=0}^{\infty} (1+|\xi_n|)\, e^{-\gamma T_n}\Big) < \infty, \qquad \text{for all } x \in R.$$

Setting

$$\text{(2.5)} \qquad \phi(\xi) = \begin{cases} K + k\xi, & \text{if } \xi > 0, \\ 0, & \text{if } \xi = 0, \\ L - \ell\xi, & \text{if } \xi < 0, \end{cases}$$

we define the cost function for a feasible policy $\{(T_n, \xi_n)\}$ by

$$C(x) = E_x\Big[h \int_0^\infty e^{-\gamma t} Z_t \, dt + \sum_{n=0}^\infty e^{-\gamma T_n} \phi(\xi_n)\Big] \quad . \tag{2.6}$$

for all $x \in R$. From (2.4) it follows that $C(x)$ is both well defined and finite for all $x \in R$. We say that this policy is <u>optimal</u> if it minimizes $C(x)$, over all feasible policies, for each $x \in R$. Now let

$$\psi(\xi) = \begin{cases} -K - c\xi \ , & \text{if} \quad \xi > 0 \ , \\ 0 \ , & \text{if} \quad \xi = 0 \ , \\ -L - r\xi \ , & \text{if} \quad \xi < 0 \end{cases} \tag{2.7}$$

so that $\psi(\xi) = -\phi(\xi) - (h/\gamma)\xi$. For each feasible policy $\{(T_n,\xi_n)\}$ define the <u>value function</u>

$$V(x) = E_x\Big\{\sum_{n=0}^\infty e^{-\gamma T_n} \psi(\xi_n)\Big\} \ , \qquad x \in R \quad . \tag{2.8}$$

Obviously $C(x)$ is the expected present value of total costs, starting in state x, under our original cost structure (1.1)-(1.3), while $V(x)$ is the expected present value of net rewards under the alternate cost structure (1.5)-(1.7).

(2.9) <u>Proposition</u>: For each feasible policy $\{(T_n,\xi_n)\}$ we have $C(x) = hx/\gamma + h\mu/\gamma^2 - V(x)$ for all $x \in R$. Thus a feasible policy is optimal if and only if it maximizes $V(x)$, over all feasible policies, for each $x \in R$.

Remark: Hereafter we shall deal exclusively with the equivalent maximization problem. This equivalence was used in [6] and is essentially due to Bell [2].

Proof. From Fubini's Theorem we have

$$(2.10) \qquad E_x\left(\int_0^\infty e^{-\gamma t} Z_t\,dt\right) = E_x\left[\int_0^\infty e^{-\gamma t}(X_t+Y_t)\,dt\right]$$

$$= \int_0^\infty e^{-\gamma t}(x+\mu t)\,dt + E_x\left(\int_0^\infty e^{-\gamma t}\, Y_t\,dt\right)$$

$$= x/\gamma + \mu/\gamma^2 + E_x\left(\int_0^\infty e^{-\gamma t}\, Y_t\,dt\right)$$

and

$$(2.11) \qquad \int_0^\infty e^{-\gamma t}\, Y_t\,dt = \int_0^\infty e^{-\gamma t}\left[\sum_{n=0}^\infty \xi_n\, 1_{\{T_n \le t\}}\right]dt$$

$$= \sum_{n=0}^\infty \xi_n \int_0^\infty e^{-\gamma t}\, 1_{\{T_n \le t\}}\,dt$$

$$= \sum_{n=0}^\infty \xi_n \int_{T_n}^\infty e^{-\gamma t}\,dt = \sum_{n=0}^\infty \xi_n\, \frac{1}{\gamma}\, e^{-\gamma T_n} \quad .$$

Combining (2.5)-(2.8) with (2.10) and (2.11) gives

$$(2.12)\qquad C(x) = hE_x\{\int_0^\infty e^{-\gamma t} Z_t dt\} + E_x\{\sum_{n=0}^\infty e^{-\gamma T_n} \phi(\xi_n)\}$$

$$= hx/\gamma + h\mu/\gamma^2 + E_x\{\sum_{n=0}^\infty e^{-\gamma T_n}[\frac{h}{\gamma}\xi_n + \phi(\xi_n)]\}$$

$$= hx/\gamma + h\mu/\gamma^2 - E_x\{\sum_{n=0}^\infty e^{-\gamma T_n} \psi(\xi_n)\}$$

$$= hx/\gamma + h\mu/\gamma^2 - V(x) \ .$$

(2.13) Proposition. Suppose that $f : [0,\infty) \to R$ is continuously differentiable, has a bounded derivative, and has a continuous second derivative at all but a finite number of points. Then for any $T > 0$, any $x \in R$ and any feasible policy we have

$$(2.14)\qquad E_x[e^{-\gamma T} f(Z_T)] = E_x[f(Z_0)]$$

$$+ E_x[\int_0^T e^{-\gamma T}(\Gamma f-\gamma f)(Z_t)dt] + E_x[\sum_{n=1}^{N(T)} \theta_n e^{-\gamma T_n}] \quad ,$$

where

$$\theta_n \equiv f(Z(T_n)) - f(Z(T_n-)) \ , \qquad \text{for } n = 1, 2, \ldots$$

and

(2.15) $$\Gamma f = \frac{1}{2}\sigma^2 f'' + \mu f' \quad .$$

Remark. We may define $f''(y)$ arbitrarily at those points y where the second derivative does not exist, becasue $\{t \geq 0 : Z_t = y\}$ has zero Lebesgue measure almost surely under each P_x.

Proof. This is almost identical to Proposition (4.2) of [7], so we shall merely sketch the proof. Fix $x \in R$, and represent X in the form $X_t = X_0 + \sigma W_t + \mu t$, where W is a standard (zero drift and unit variance) Brownian motion (under P_x) with $W_0 = 0$. If f is twice continuously differentiable, then we may apply the one-dimensional change of variable formula (or generalized Ito formula) for semimartingales, which appears on page 301 of Meyer [8], to obtain

(2.16) $$f(Z_T) = f(Z_0) + \sigma \int_0^T f'(Z_t)\, dW_t + \int_0^T \Gamma f(Z_t)dt + \sum_{n=1}^{N(T)} \theta_n \quad .$$

In fact, this is making things a bit more difficult than is really necessary, since the same result can be obtained by applying the ordinary one-dimensional Ito formula over each of the intervals $[T_n, T_{n+1})$ and then summing up over $n = 0, 1, \ldots, N(T)$. Furthermore, it is well known that the Ito formula (or the general change of variable formula) remains valid even when f is not twice continuously differentiable, provided that it has an absolutely continuous derivative f' and f'' is chosen as any density of f', cf. [1]. Thus (2.16) is valid with our hypotheses. Now using (2.16) and the integration by parts formula for semimartingales, which appears on page 303 of Meyer [8], we obtain

$$(2.17) \quad e^{-\gamma T} f(Z_T) = f(Z_0) + \int_0^T e^{-\gamma t}\, df(Z)_t - \int_0^T f(Z_{t-})\, \gamma e^{-\gamma t}\, dt$$

$$= f(Z_0) + \sigma \int_0^T e^{-\gamma t} f'(Z_t)\, dW_t + \int_0^T e^{-\gamma t}\, \Gamma f(Z_t) dt$$

$$+ \sum_{0 < t \leq T} e^{-\gamma t}\, \Delta f(Z)_t - \int_0^T f(Z_t)\, \gamma e^{-\gamma t} dt$$

$$= f(Z_0) + \sigma \int_0^T e^{-\gamma t} f'(Z_t)\, dW_t$$

$$+ \int_0^T e^{-\gamma t} (\Gamma f - \gamma f)(Z_t) dt + \sum_{n=1}^{N(T)} e^{-\gamma T_n}\, \theta_n \quad .$$

To get the desired result, we take E_x of both sides in (2.17) and observe that the expectation of the Ito integral vanishes, becasue its integrand is bounded by hypothesis.

(2.18) Proposition. Suppose that $f : [0,\infty) \to R$ satisfies all the hypotheses of (2.13) plus

$$\text{(2.19)} \qquad \Gamma f(x) - \gamma f(x) \leq 0 \qquad \text{for almost all} \quad x \geq 0 \ ,$$

$$\text{(2.20)} \qquad f(x) \geq f(y)-K-(y-x)c \qquad \text{for} \qquad 0 \leq x < y \ ,$$

$$\text{(2.21)} \qquad f(x) \geq f(y)-L+(x-y)r \qquad \text{for} \qquad 0 \leq y < x \ .$$

Then $f(x) \geq V(x)$ for any feasible policy and any $x \geq 0$.

Proof. Using the definition (2.7) of $\phi(\cdot)$, we see that (2.20) and (2.21) together give us $f(x)-f(y) \geq \phi(y-x)$, which means that

$$\text{(2.22)} \qquad -\theta_n \geq \phi(Z(T_n)-Z(T_n-)) = \phi(\xi_n) \ , \quad \text{for} \ n = 1, 2, \dots ,$$

where θ_n is defined as in Proposition (2.13). Putting (2.19) and (2.22) into (2.14) and rearranging terms, we have

$$E_x[f(Z_0)] \geq E_x\left[\sum_{n=1}^{N(T)} \phi(\xi_n)\, e^{-\gamma T_n}\right] + E_x[e^{-\gamma t} f(Z_T)] \ .$$

From (2.4) and the boundedness of f' it follows that $E_x[\exp(-\gamma T) f(Z_T)] \to 0$ as $T \to \infty$, so we have

$$(2.23) \qquad E_x[f(Z_0)] \geq E_x[\sum_{n=1}^{\infty} \phi(\xi_n) e^{-\gamma T_n}] .$$

Finally, since $Z_0 = X_0+\xi_0$, another application of (2.20)-(2.21) gives

$$(2.24) \qquad f(X_0) \geq f(Z_0) + \phi(\xi_0) .$$

Of course $E_x[f(X_0)] = f(x)$, so by combining (2.23) and (2.24) we have the desired result,

$$f(x) \geq E_x[\sum_{n=0}^{\infty} \phi(\xi_n) e^{-\gamma T_n}] \equiv V(x) .$$

3. Control Band Policies

Let us now consider a control band policy with parameters (q,Q,S) satisfying $0 < q < Q < S$. Remembering that $T_0 = 0$ by definition, we take

$$(3.1) \qquad \xi_0 = \begin{cases} q-X_0 , & \text{if } 0 \geq X_0 , \\ 0 , & \text{if } 0 < X_0 < S , \\ Q-X_0 , & \text{if } X_0 \geq S \end{cases}$$

and of course $Z_0 = X_0+\xi_0$. Assuming that it's clear from the verbal description given in §1 how $T_1, T_2, \cdots$ and $\xi_1, \xi_2, \cdots$ are recursively constructed, we shall not write out their formal definitions. The relevant properties of the control band policy

$\{(T_n, \xi_n)\}$ are the following. With $Y = \{Y_t, t \geq 0\}$ defined as in §1, $Z \equiv X+Y$, and $s \equiv S-Q$ as before, we have

$$(3.2) \qquad Z(T_n-) \in \{0,S\} \qquad \text{for all } n = 1, 2, \ldots ,$$

$$(3.3) \qquad \xi_n = \begin{cases} q, & \text{if } Z(T_n-) = 0 \\ -s, & \text{if } Z(T_n-) = S \end{cases} .$$

We now want to compute explicitly the value function V for this control band policy. With the differential operator Γ defined by (2.15), it will ultimately be seen that V is twice continuously differentiable on $[0,S]$ and uniquely satisfies

$$(3.4) \qquad \Gamma V(x) - \gamma V(x) = 0 , \qquad 0 \leq x \leq S ,$$

subject to the auxiliary conditions

$$(3.5) \qquad V(0) = V(q) + \phi(q) = V(q)-K-cq ,$$

$$(3.6) \qquad V(S) = V(Q) + \phi(-s) = V(Q)-L+rs .$$

To extend V to a function on all of R, we write (3.5) and (3.6) in the more general form

$$(3.7) \qquad V(x) = V(q)-K-c(q-x) , \qquad \text{for } x \leq 0 ,$$

$$(3.8) \qquad V(x) = V(Q)-L+r(x-Q) , \qquad \text{for } x \geq S .$$

Now let

$$(3.9) \qquad \alpha \equiv [(\mu^2 + 2\gamma\sigma^2)^{1/2} - \mu]/\sigma^2 > 0 \ ,$$

and

$$(3.10) \qquad \beta \equiv [(\mu^2 + 2\gamma\sigma^2)^{1/2} + \mu]/\sigma^2 > 0 \ ,$$

so that $z = \alpha$ and $z = -\beta$ are the two solutions of the quadratic equation $1/2\ \sigma^2 z^2 + \mu z - \gamma = 0$. The general solution of the ordinary differential equation (ODE) (3.4) is

$$(3.11) \qquad V(x) = A\ e^{\alpha x} + B\ e^{-\beta x} \ , \qquad 0 \leq x \leq S \ ,$$

and in our case the constants A and B must be chosen so as to satisfy (3.5) and (3.6).

(3.12) Proposition. Let V be defined on $[0,S]$ by (3.11), with A and B chosen so as to satisfy (3.5) and (3.6), and extend V to a continuous function on all of R by (3.7) and (3.8). Then V is the value function for the (q,Q,S) control band policy.

Proof. We shall use the fact that this function V satisfies (3.4), which is easy to verify. The central step in the proof is an application of Proposition (2.13), with V in place of f. Since $0 \leq Z_t \leq S$ for all $t \geq 0$, it is sufficient for this application that V be twice continuously differentiable on $[0,S]$. With θ_n defined as in (2.13), we have from (3.3), (3.5) and (3.6) that

$$(3.13) \qquad \theta_n = \begin{cases} L-rs\ , & \text{if } Z(T_n-) = S \\ K+cq\ , & \text{if } Z(T_n-) = 0 \end{cases} ,$$

which means simply that $\theta_n = -\psi(\xi_n)$. Furthermore, $\Gamma V(Z_t) - \gamma V(Z_y) = 0$ for all $t \geq 0$ by (3.4). Combining these facts with (2.13) gives

$$(3.14) \qquad E_x[e^{-\gamma t}\ V(Z_T)] = E_x[V(Z_0)] - E_x[\sum_{n=1}^{N(T)} e^{-\gamma T_n}\ \psi(\xi_n)]$$

for any $T > 0$ and $x \in R$. Next, (3.1), (3.7) and (3.8) give

$$(3.15) \qquad V(Z_0) = V(X_0+\xi_0) = V(X_0)-\psi(\xi_0)\ ,$$

so we can rewrite (3.14) as

$$(3.16) \qquad E_x[e^{-\gamma t}\ V(Z_T)] = E_x[V(X_0)] - E_x[\sum_{n=0}^{N(T)} e^{-\gamma T_n}\ \psi(\xi_n)]\ .$$

Of course $E_x[V(X_0)] = V(x)$, and the left side of (3.16) vanishes as $T \to \infty$ because $V(Z_y)$ is bounded, so we obtain the desired result by letting $T \to \infty$ in (3.16).

4. Optimal Policy Parameters

Continuing the discussion of control band policies, it will be convenient to define

$$(4.1) \qquad \pi(x) = V'(x) , \qquad \text{for } x \in R .$$

Actually, the left and right derivatives of V need not agree at $x =$ and $x = S$, so (4.1) is ambiguous at those points. To resolve that ambiguity, let us agree to <u>define</u> π <u>so that it is continuous on</u> $[0,S]$. From (3.11) we have

$$(4.2) \qquad \pi(x) = \alpha A\ e^{\alpha x} - \beta B\ e^{-\beta x} , \qquad 0 \leq x \leq S ,$$

and the conditions (3.5)-(3.6) determining A and B may be rewritten in terms of π as

$$(4.3) \qquad \int_0^q [\pi(x)-c]dx = K ,$$

and

$$(4.4) \qquad \int_Q^S [r-\pi(x)]dx = L .$$

In this section it will be shown that there is exactly one choice of the control band policy parameters (q,Q,S) such that

$$\pi(q) = c \tag{4.5}$$

and

$$\pi(Q) = \pi(S) = r \ , \tag{4.6}$$

and the corresponding control band policy is optimal. Interpretations for (4.5)-(4.6), and the conditions to be derived from them shortly, will be offered in the next section. For arbitrary $s > 0$ we define

$$a(s) = (1 - e^{-\beta s})/(e^{\alpha s} - e^{-\beta s}) > 0 \ , \tag{4.7}$$

$$b(s) = (e^{\alpha s} - 1)/(e^{\alpha s} - e^{-\beta s}) > 0 \ , \tag{4.8}$$

and

$$f_s(y) = r[a(s)\, e^{\alpha y} + b(s)\, e^{-\beta y}] \ , \quad y \in R \ , \tag{4.9}$$

so that f_s uniquely satisfies $\Gamma f_s - \gamma f_s = 0$ subject to $f_s(0) = f_s(s) = r$. Obviously $f_s'' > 0$ and thus we have the following:

(4.10) For any $s > 0$ the function $f_s(\cdot)$ is strictly convex on R and has a minimum in $(0,s)$.

This situation is pictured in Figure 3 below. From (4.2) we have $\Gamma\pi(x) - \gamma\pi(x) = 0$ for $0 \leq x \leq S$, and hence π can only satisfy (4.6) if

$$\pi(x) = f_s(x-Q) , \qquad 0 \leq x \leq S , \tag{4.11}$$

where $s \equiv S-Q$ as usual. Then (4.4) demands that

$$\begin{aligned} L &= \int_Q^S [r - f_s(x-Q)]dx = \int_0^s [r - f_s(y)]dy \\ &= rs - r(\frac{1}{\alpha} + \frac{1}{\beta})(e^{\alpha s}-1)(1 - e^{-\beta s})/(e^{\alpha s} - e^{-\beta s}) \end{aligned} \tag{4.12}$$

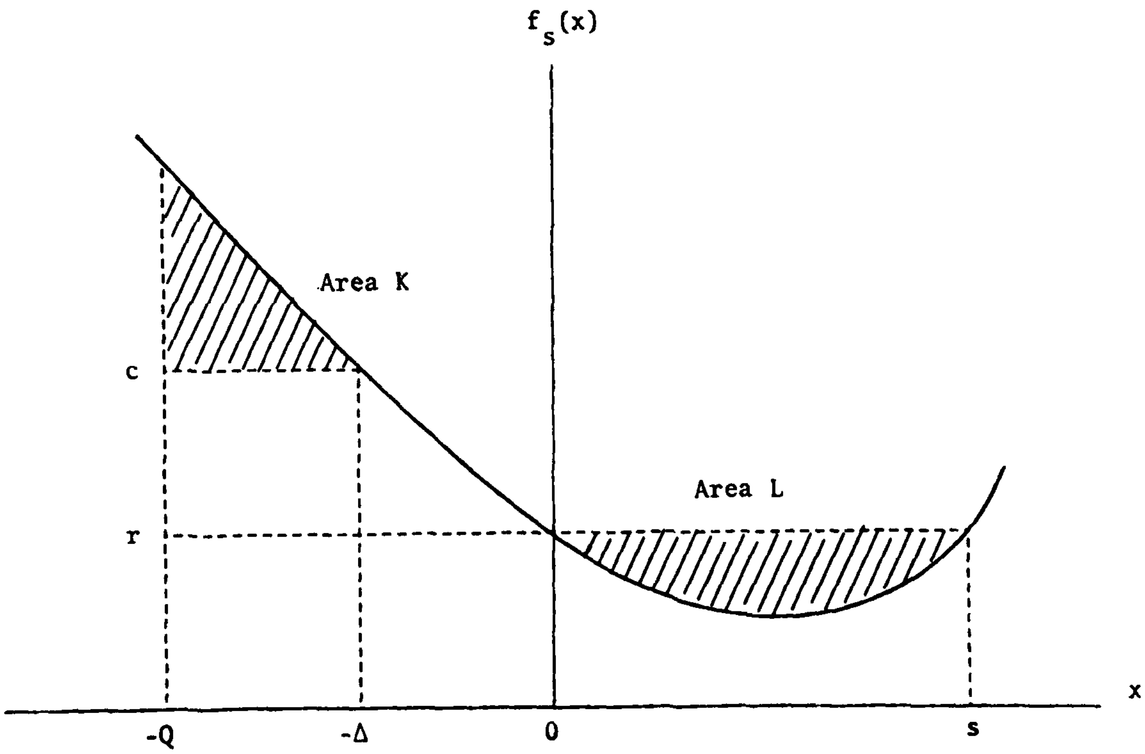


Figure 3. Determining the Optimal Parameters

In §5 we shall give a probabilistic interpretation for $f_s(\cdot)$ that makes the following proposition obvious, but it can also be verified analytically, and this is left as an exercise.

(4.13) Proposition. The right side of (4.12) increases continuously from 0 to ∞ as s increases from 0 to ∞, so there is a unique $s > 0$ satisfying (4.12).

Hereafter we assume that $s \equiv S-Q$ has been chosen to satisfy (4.12). Next, setting $\Delta \equiv Q-q$ as in §1, we see that (4.5) and (4.11) require

$$(4.14) \qquad c = \pi(q) = f_x(-\Delta) = r[a(s)\, e^{-\alpha\Delta} + b(s)\, e^{\beta\Delta}]$$

as shown in Figure 3. It is immediate from (4.10) that there exists a unique $\Delta > 0$ satisfying (4.13), because $c > r$ by assumption, and we assume hereafter that Δ has been chosen in this way. Finally, (4.3) and (4.11) demand that $Q > \Delta$ satisfy

$$(4.15) \qquad K = \int_0^q [\pi(x) - c]dx = \int_{-Q}^{-\Delta} [f_s(y)-c]dy$$

$$= r[\frac{a(s)}{\alpha} (e^{-\alpha\Delta} - e^{-\alpha Q}) + \frac{b(s)}{\beta} (e^{\beta Q} - e^{\beta\Delta})] - c(Q-\Delta) .$$

It is again immediate from (4.10) that there exists a unique $Q > \Delta$ satisfying this condition, as shown in Figure 3.

(4.16) Proposition. Let $s > 0$, $\Delta > 0$ and $Q > \Delta$ be chosen to satisfy (4.12), (4.14) and (4.15) respectively, and set $q \equiv Q-\Delta$ and $S \equiv Q+s$. Then the (q,Q,S) control band policy is optimal.

Proof. Our first task is to construct the value function V by reversing the logic of this section. With (q,Q,S) chosen in the indicated way, we define $f_s(\cdot)$ by (4.7)-(4.9) and set $\pi(x) = f_s(x-Q)$ for $0 \le x \le s$, as in (4.11). Now let

$$V(0) = \frac{1}{\gamma}\left[\frac{1}{2}\sigma^2 \pi'(0) + \mu\pi(0)\right] , \tag{4.17}$$

and

$$V(x) = V(0) + \int_0^x \pi(y)dy . \tag{4.18}$$

The value of $V(0)$ has been set so that $V(\cdot)$, as defined by (4.18), will satisfy $(\Gamma V-\gamma V)(0) = 0$. Then (4.18) insures that $(\Gamma V-\gamma V)(x) = 0$ for all $x \in [0,S]$ because π satisfies this same ODE. Now extend V to a function on all of R by (3.7) and (3.8). Because π satisfies (4.3) and (4.4) by construction, we see that V satisfies all the hypotheses of Proposition (3.12). Thus V is the value function for the (q,Q,S) control band policy, as desired.

Our next task is to show that V satisfies the hypotheses of Proposition (2.18) and hence provides an upper bound for the value function of any other feasible policy. First note that V is continuously differentiable on $[0,\infty)$ because $V'(S+) = r$ by (3.8), while $V'(S-) = \pi(S-) = f_s(s-) = r$ by construction. Next, it must be established that

$$(4.19) \qquad (\Gamma V - \gamma V)(x) \leq 0\ , \qquad \text{for all } x \geq 0\ .$$

Of course (4.19) holds with equality on $[0,S]$. As we pass through S from the left, both V and $V' = \pi$ are continuous, while $V''(\cdot) \equiv \pi'(\cdot)$ jumps from the positive value at $S-$ pictured in Figure 1 to a zero value. (Remember that V is linear with slope r to the right of S.) Thus $(\Gamma V - \gamma V)(S+) < 0$. Finally, ΓV is constant to the right of S, while V is increasing linearly, so $\Gamma V - \gamma V$ becomes even more negative as we move right from S, and (4.19) is confirmed. To verify the remaining hypotheses (2.20) and (2.21) of Proposition (2.18), one needs little more than the picture of $\pi \equiv V'$ given in Figure 1, and we shall leave this as an exercise. Then (2.18) gives us $V(x) \geq V^*(x)$ for all $x \geq 0$, where V^* is the value function for any other feasible policy. It remains only to show that this same inequality holds for $x < 0$, which we also leave as an exercise.

5. Interpretations

In this section we seek to interpret the conditions (4.5) and (4.6) that were implosed at the beginning of the previous section, and to elaborate on the relationships (4.12), (4.14) and (4.15) that were ultimately found to determine the optimal policy parameters. For this purpose we fix an arbitrary control band policy, hereafter called the nominal policy or candidate policy, with parameters $0 < q < Q < S$. Let $Z = X+Y$ be the associated controlled process, let $V(x)$ be the associated value function, and set $\pi(x) = V'(x)$ as in §4. Using policy improvement logic, we shall derive three plausible necessary conditions for optimality of the nominal policy.

If the controller starts in state S, immediately jumps downward to level x, and thereafter follows the control band policy (q,Q,S), his total expected discounted reward will be $\phi(x-S)+V(x) = V(x)-L+(S-x)r$. If the candidate policy is to be optimal, then it must be that this expression is maximized by taking $x = Q$, which obviously demands

$$(5.1) \qquad \pi(Q) = r .$$

In exactly the same way, by considering the various points x to which the controller could jump from zero, we obtain the optimality condition

$$(5.2) \qquad \pi(q) = c .$$

To complete the motivation of (4.5)-(4.6), we need to argue that a necessary condition for optimality is $\pi(Q) = \pi(S)$. One can actually obtain a much more stringent and enlightening optimality condition by the following argument. First define

$$T(y) \equiv \inf\{t \geq 0 : Z_t = y\}, \qquad 0 \leq y \leq S$$

$$\theta(x,y) \equiv E_x[e^{-\gamma T(y)}], \qquad 0 \leq x, y \leq S$$

Suppose that our controller, following the candidate control band policy, starts in state x, and let y be another state such that $0 < y < x < S$ and $y < Q$. The expected present value of his total net reward over $[0,\infty)$ is of course $V(x)$, and we define

$U(x,y) \equiv$ expected present value, when starting in state x and following the nominal policy, of net rewards earned over the period $[0, T(y)]$.

From the strong Markov property of X and the stationary character of control band policies, it is apparent that $V(x) = U(x,y)+\theta(x,y)\, V(y)$, so we have

$$U(x,y) = V(x) - \theta(x,y)\, V(y) \,. \tag{5.3}$$

Now fix x and y satisfying $0 < y < x < S$ and $y < Q$. Let ε be a perturbation, either positive or negative, small enoguh that $0 < y+\varepsilon < Q$. Let the starting state be $x+\varepsilon$, and consider the <u>alternate strategy</u> where one follows a control band policy with parameters $(q, Q+\varepsilon, S+\varepsilon)$ up until the first time $T^*(y+\varepsilon)$ at which level $y+\varepsilon$ is hit, and then reverts to usage of the nominal policy ever afterward. Let

$V^*(x+\varepsilon) \equiv$ expected present value, starting in state $x+\varepsilon$, of net rewards earned under the alternate strategy over $[0,\infty)$.

From the spatial homogeneity of Brownian motion we obtain

$$\theta(x,y) \equiv E_x[e^{-\gamma T(y)}] = E_{x+\varepsilon}[e^{-\gamma T^*(y+\varepsilon)}] \quad , \tag{5.4}$$

and similarly

(5.5) $U(x,y)$ = expected present value, when starting in state $x+\varepsilon$ and following the alternate strategy, of net rewards earned over the period $[0, T^*(y+\varepsilon)]$.

Thus, as a precise analog to (5.3), we have

$$\begin{aligned} V^*(x+\varepsilon) &= U(x,y) + \theta(x,y)\, V(y+\varepsilon) \\ &= V(x) + \theta(x,y)\, [V(y+\varepsilon) - V(y)] \quad . \end{aligned} \tag{5.6}$$

The last equality is obtained by substitution of (5.3). Subtracting $V(x+\varepsilon)$ from (5.6), we see that the <u>improvement</u> effected by the alternate strategy is

$$V^*(x+\varepsilon) - V(x+\varepsilon) = \theta(x,y)\, [V(y+\varepsilon) - V(y)] - [V(x+\varepsilon) - V(x)] \tag{5.7}$$

If the nominal policy is to be optimal, this expression must have a local minimum at $\varepsilon = 0$, which obviously requiries $0 = \theta(x,y)\,V'(y) - V'(x)$, or equivalently $\pi(x) = \theta(x,y)\,\pi(y)$. We have derived this for $0 < y < x < S$ and $y < Q$, and then by continuity we arrive at our final optimality condition

$$\pi(x) = \theta(x,y)\,\pi(y)\ , \quad \text{for} \quad 0 \le y \le x \le S \quad \text{and} \quad y \le Q\ . \tag{5.8}$$

Because Z jumps immediately to Q upon hitting S, we have $\theta(S,Q)$ = so (5.8) implies

$$\pi(Q) = \pi(S)\ , \tag{5.9}$$

and this completes our justification for the conditions (4.5)-(4.6) that were imposed earlier. To get more insight from (5.8), set $y = Q$ and invoke the condition $\pi(Q) = r$ derived above. This gives $\pi(x) = r\theta(x,Q)$ for $Q \le x \le S$, and then the basic identity (3.6) demands that

$$\begin{aligned} L = r(S-Q) - [V(S)-V(Q)] &= \int_Q^S [r - \pi(x)]dx \\ &= r\int_Q^S [1-\theta(x,Q)]dx = r\int_0^s E_x[1-e^{-\gamma T(Q)}]dx \ . \end{aligned} \tag{5.10}$$

After a bit of reflection, one realizes that the right-hand side of (5.10) depends only on $s \equiv S-Q$ and that it increases continuously from 0 to ∞ as s increases from 0 to ∞. Thus (5.10) uniquely determines the value of s for an optimal control band policy, and it is just the probabilistic articulation of the analytical condition

(4.12) derived earlier. From the definitive analytical properties of the function f_s introduced in §4, one can easily verify the interpretation

$$f_s(x) = rE_{Q+x}[e^{-\gamma T(Q)}] ,$$

which establishes the equivalence of (4.12) and (5.10). To determine the policy parameter $\Delta \equiv Q-q$ from (5.8), set $x = Q$ and $y = q$, and use the fact that $\pi(q) = c$ by (5.2) while $\pi(Q) = r$ by (5.1). Then (5.8) gives

$$r = c\theta(Q,q) = cE_Q[e^{-\gamma T(q)}] \quad . \tag{5.11}$$

With s already determined, the right-hand side of (5.11) depends only on $\Delta \equiv Q-q$, and it decreases continuously from c to 0 as Δ increases from 0 to ∞. Thus (5.11) uniquely determines Δ for the optimal control band policy, and one can easily show that it is equivalent to the analytical condition (4.14) derived earlier. Once s and Δ have been set, we can determine q from (5.8) as follows. With $x = q$ and $0 \le y \le q$, (5.8) reduces to $c = \theta(q,y)\,\pi(y)$ because $\pi(q) = c$ by (5.2). Then the basic identity (4.19) requires that

$$\begin{aligned} K &= \int^{q} \pi(y)dy - cq = c \int_0^{q} [1/\theta(q,y) - 1]dy \\ &= c \int_0^{q} \{E_q[1 - e^{-\gamma T(y)}]/E_q[e^{-\gamma T(y)}]\}dy \ , \end{aligned} \tag{5.12}$$

which is the probabilistic articulation of our analytical condition (4.15).

References

[1] J. Azema and M. Yor (1978), Temps Locaux, Asterisque, No. 52-53, Soc. Math. France.

[2] C. Bell (1971), Characterization and Computation of Optimal Policies for Operating an M/G/1 Queuing System with Removable Server, Opns. Rsch., Vol. 19, 208-218.

[3] V.E. Benes, L.A. Shepp and H.S. Witsenhausern (1980), Some Solvable Stochastic Control Problems, Stochastics, Vol. 4, 39-83.

[4] A. Bensoussan and J.L. Lions (1975), Nouvelles Methods en Controle Impulsionnel, Appl. Math. Opt., Vol. 1, 289-312.

[5] G.M. Constantinides and S.F. Richard (1978), Existence of Optimal Simple Policies for Discounted-Cost Inventory and Cash Management in Continuous Time, Ops. Resch., Vol. 26, 620-636.

[6] J.M. Harrison and A.J. Taylor (1978), Optimal Control of a Brownian Storage System, Stoch. Proc. Appl., Vol. 6, 179-194.

[7] J.M. Harrison and M.I. Taksar (1981), Instantaneous Control of Brownian Motion, Technical Report, Department of Operations Research, Stanford University (submitted for publication).

[8] P.-A. Meyer (1976), Un Cours sur les Integrales Stochastiques, in Sem. de Prob. X, Lecture Notes in Mathematics 511, Springer-Verlag, Berlin.

UNCLASSIFIED

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

REPORT DOCUMENTATION PAGE		READ INSTRUCTIONS BEFORE COMPLETING FORM
1. REPORT NUMBER 202	2. GOVT ACCESSION NO. AD-A109 811	3. RECIPIENT'S CATALOG NUMBER
4. TITLE (and Subtitle) IMPULSE CONTROL OF BROWNIAN MOTION		5. TYPE OF REPORT & PERIOD COVERED TECHNICAL REPORT
		6. PERFORMING ORG. REPORT NUMBER
7. AUTHOR(s) J. Michael Harrison, Thomas M. Sellke and Allison J. Taylor		8. CONTRACT OR GRANT NUMBER(s) N00014-75-C-0561
9. PERFORMING ORGANIZATION NAME AND ADDRESS DEPT'S OF OPERATIONS RESEARCH & STATISTICS STANFORD UNIVERSITY STANFORD, CALIFORNIA 94305		10. PROGRAM ELEMENT, PROJECT, TASK AREA & WORK UNIT NUMBERS (NR-047-200)
11. CONTROLLING OFFICE NAME AND ADDRESS OPERATIONS RESEARCH, CODE 434 OFFICE OF NAVAL RESEARCH ARLINGTON, VIRGINIA 22217		12. REPORT DATE NOVEMBER 1981
		13. NUMBER OF PAGES 32
14. MONITORING AGENCY NAME & ADDRESS(if different from Controlling Office)		15. SECURITY CLASS. (of this report) UNCLASSIFIED
		15a. DECLASSIFICATION/DOWNGRADING SCHEDULE

16. DISTRIBUTION STATEMENT (of this Report)

APPROVED FOR PUBLIC RELEASE: DISTRIBUTION IS UNLIMITED

17. DISTRIBUTION STATEMENT (of the abstract entered in Block 20, if different from Report)

18. SUPPLEMENTARY NOTES

This research has been partially supported by National Science Foundation Grant ECS-80-17867, and issued as Technical Report No. 64 - Department of Operations Research, Stanford University.

19. KEY WORDS (Continue on reverse side if necessary and identify by block number)

BROWNIAN MOTION   STOCHASTIC CONTROL   JUMP BOUNDARIES

INVENTORY AND PRODUCTION CONTROL   IMPULSE CONTROL

STOCHASTIC CASH MANAGEMENT

20. ABSTRACT (Continue on reverse side if necessary and identify by block number)

PLEASE SEE REVERSE SIDE

UNCLASSIFIED

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

ABSTRACT: REPORT NO. 202 (mu and sigma sq)

Consider a storage system, such as an inventory or cash fund, whose content fluctuates as a (μ,σ^2) Brownian motion in the absence of control. Holding costs are continuously incurred at a rate proportional to the storage level, and we may cause the storage level to jump by any desired amount at any time except that the content must be kept nonnegative. Both positive and negative jumps entail fixed plus proportional costs, and our objective is to minimize expected discounted costs over an infinite planning horizon. A control band policy is one that enforces an upward jump to q whenever level zero is hit, and enforces a downward jump to Q whenever level S is hit $(0 < q < Q < S)$. We prove the existence of an optimal control band policy and calculate explicitly the optimal values of the critical numbers (q,Q,S).

UNCLASSIFIED

SECURITY CLASSIFICATION OF THIS PAGE(When Data Entered)