A CLASS OF SOLVABLE STOCHASTIC INVESTMENT PROBLEMS INVOLVING SINGULAR CONTROLS

by
T.Ø. Kobila*

1. Introduction

In this paper we shall study a class of stochastic control problems which admit very explicit solutions. The problems come from economy, and we shall first present them from an economic point of view.

Consider a company which wants to adjust its production capacity to a fluctuating market. As the market improves and demand increases, the company naturally wants to expand its capacity, but the problem is that the investments needed for the expansion are irreversible in the sense that if the market later fails, the company can not get the invested capital back by reducing the capacity. There are many examples of investments which are irreversible or nearly irreversible in this sense; e.g., investments in highly specialized production equipment or in an extended work force which can not easily be reduced for legal or humanitarian reasons (see Pindyck (1988, 1991a,b) for detailed discussions of irreversible investments from an economist's point of view). In its search for a strategy maximizing the long term profit, the company obviously has to balance its urge to expand in a good market with the fear of overinvesting and hence losing money if the market drops.

We shall be studying a very simple mathematical model for this quite general and complex economic problem. At any time, our company's financial situation will be described by two nonnegative, real parameters θ and k, where k is just the current production capacity, and θ is an economic indicator for the state of the market - intuitively, a high value of θ corresponds to a booming economy with high demand, while a low value for θ indicates a market with little activity and low demand. To specify the actual income and expenditure of the company, we introduce two functions

$$\Pi, \Gamma : [0,\infty) \times [0,\infty) \rightarrow [0,\infty)$$

with the following interpretation: $\Pi(\theta, k)$ is the net profit per unit time the company is making from its production in the economic situation described by (θ, k), while $\Gamma(\theta, k + \Delta k) - \Gamma(\theta, k)$ is the investment needed in order to increase the capacity from k to $k + \Delta k$.

To model the statistical development of the market, we shall use a one-dimensional, geometric Brownian motion

$$d\Theta_t = \alpha\Theta_t dt + \beta\Theta_t dB_t, \quad \Theta_0 = \theta, \tag{1.1}$$

---

* This is an acronym for Iulie Aslaksen, Olav Bjerkholt and Kjell Arne Brekke, all at the Central Bureau of Statistics, Box 8131 Dep, N-0033 Oslo, Norway, and Tom Lindstrøm and Bernt Øksendal both at the Department of Mathematics, University of Oslo, Box 1053, Blindern, N-0316 Oslo, Norway.

where α and β are two (fixed) real parameters, and B is a Brownian motion. Thus if the market is in state θ at time 0, its state at time t is given by the random variable

$$\Theta_t = \theta e^{(\alpha - \frac{\beta^2}{2})t + \beta B_t} \tag{1.2}$$

Roughly speaking, an expansion strategy for the company will be an increasing process K_t which is measurable with respect to the σ-algebra $\sigma\{\Theta_s : s \leq t\}$ generated by Θ up to time t (see Section 3 for the technical details), and the total, discounted profit the company makes from such a strategy is heuristically given by

$$J(K) = E^{\theta,k}[\int_0^\infty e^{-rt}(\Pi(\Theta_t, K_t)dt - \Gamma_k(\Theta_t, K_t)dK_t)], \tag{1.3}$$

where $r \in \mathbf{R}_+$ is a constant discount factor, Γ_k is the partial derivative of Γ w.r.t. the variable k, and $E^{\theta,k}$ denotes expectation w.r.t. the processes (Θ_t, K_t) started at (θ, k). (It turns out that when K_t is not absolutely continuous, we have to modify (1.3) somewhat (see Section 3), but this need not worry us for the time being). Our aim is to find the strategy K which maximizes the profit $J(K)$.

Let us try to approach the problem in an intuitive and, perhaps, slightly naive way. Assume first that we start in the situation indicated by (θ_1, k_1) in Figure I.1; i.e. in a situation where our capacity is large, but demand very low.

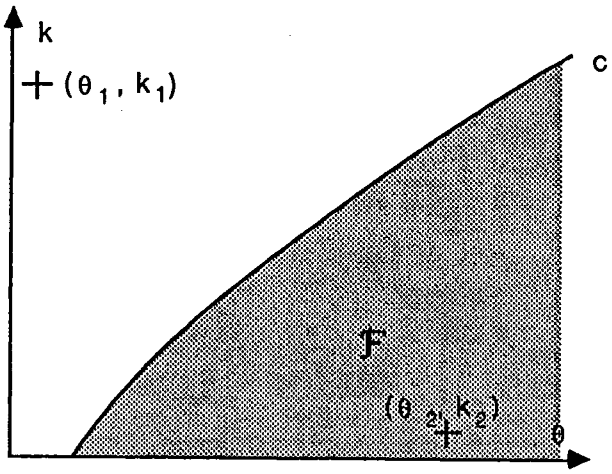

Figure I.1

Under normal circumstances, we would clearly not want to expand in such a state. If, on the other hand, we are in the situation indicated by (θ_2, k_2) where demand is high, but our capacity quite small, we would probably want to make a major investment as soon as possible. It's reasonable to assume that there will be a clear boundary between the points where we want to invest and those where we want to wait - perhaps something like the curve C in Figure I.1. With this picture in mind, it's easy to get an intuitive understanding of what the optimal policy must be: If we start at a point above C, we just wait till we hit C, and then invest just enough to always keep us on or above C. If, on the other hand, we start below C, we immediately invest just enough to get up to C, and then follow the strategy above.

As we shall see, this simple, intuitive picture is indeed correct, but the problem is to identify the "forbidden region" (under the curve C) which the process (Θ_t, K_t) is not allowed to enter. It turns out that (under some assumptions) this region has a very simple description in terms of two functions

$$\psi_1, \psi_2 : [0, \infty) \to [0, \infty]$$

(note that we allow ψ_1, ψ_2 to take the value ∞). Before we write down the definitions of these functions, we need to introduce the following notation. The differential operator A defined by

$$Ah(\theta) = \frac{\beta^2}{2}\theta^2 h''(\theta) + \alpha\theta h'(\theta) - rh(\theta) \tag{1.4}$$

is basically the infinitesimal generator

$$A_0 h(\theta) = \frac{\beta^2}{2}\theta^2 h''(\theta) + \alpha\theta h'(\theta) \tag{1.5}$$

of the geometric Brownian motion Θ with an extra term added to handle the discount factor e^{-rt}. To see the relationship between A and our problem more clearly, just observe that by Itô's formula

$$E[h(\Theta_t)e^{-rt}] = h(\Theta_0) + E(\int_0^t Ah(\Theta_s)e^{-rs}ds)$$

for all sufficiently regular functions h. Note that A is an Euler-operator with characteristic equation

$$\frac{\beta^2}{2}\gamma^2 + (\alpha - \frac{\beta^2}{2})\gamma - r = 0, \tag{1.6}$$

and let γ_1 be the positive and γ_2 the negative root of this equation. We now define ψ_1 and ψ_2 by

$$\psi_1(k) = \inf\{\theta : \int_0^\theta \frac{(\Pi_k + A\Gamma_k)(\eta, k)}{\eta^{\gamma_2+1}} d\eta > 0\} \tag{1.7}$$

$$\psi_2(k) = \sup\{\theta : \int_\theta^\infty \frac{(\Pi_k + A\Gamma_k)(\eta, k)}{\eta^{\gamma_1+1}} d\eta > 0\}, \tag{1.8}$$

(where $\prod_k$ and Γ_k denote the partial derivatives of $\prod$ and Γ w.r.t k) and note that for many choices of Π and Γ, we can compute ψ_1 and ψ_2 explicitly as the smallest/largest zero of an elementary integral. If we assume that the function $(\Pi_k + A\Gamma_k)\ (\eta, k)$ is decreasing in the k-variable - and, as we shall see in Sections 5 and 7, this is an assumption which has a clear economic interpretation and which is often, but not always, reasonable - then ψ_1 is increasing and ψ_2 is decreasing, and we typically get the picture in Figure I.2. The "forbidden region" $\mathcal{F}$ is just the area lying under both the curves ψ_1 and ψ_2; i.e.,

$$\mathcal{F} = \{(\theta, k) : \psi_1(k) \leq \theta \leq \psi_2(k)\}.$$

In many economical models, $\psi_2(k) = \infty$ for all k, and we have the situation in Figure I.3.a). Mathematically possible, but less plausible from an economic point of view, is the reverse situation in Figure I.3.b) where $\psi_1 \equiv 0$.

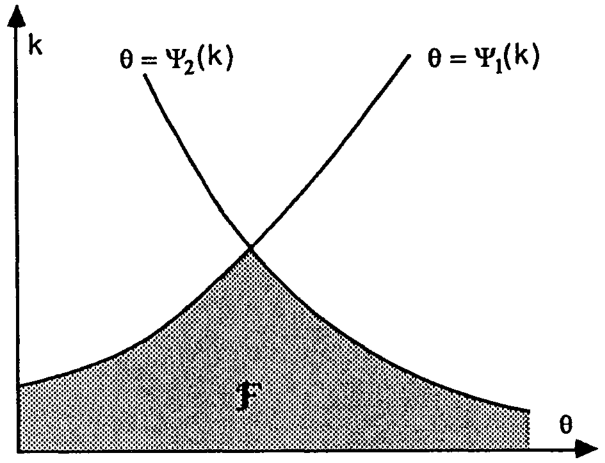

Figure I.2

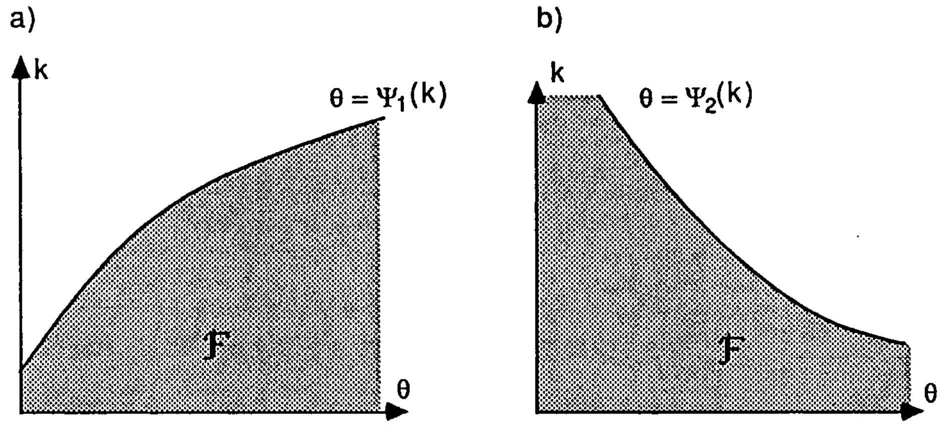

Figure I.3

A fourth possibility is the situation in Figure I.4 where ψ_1 and ψ_2 never meet; in this case there is no optimal strategy (unless we allow the capacity to become infinite in finite time), but we can still use our picture to produce almost optimal strategies. A situation which is ruled out by our technical conditions, but which can probably occur in a more general setting, is the one shown in Figure I.5.a) where the forbidden area has several camel-like bumps.

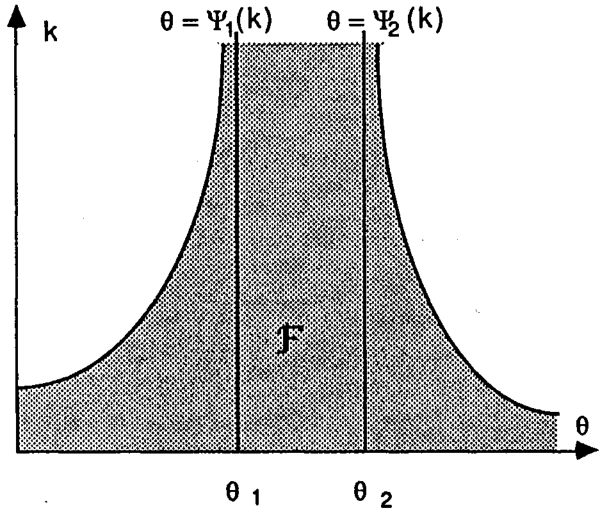

Figure I.4

On the other hand, if we remove the monotonicity condition on $(\Pi_k + A\Gamma_k)$, we may get bumps of the kind shown in Figure I.5.b), but in this case the problem becomes much more complicated, and the forbidden region can no longer be described directly in terms of ψ_1 and ψ_2. We shall study this situation in Section 7.

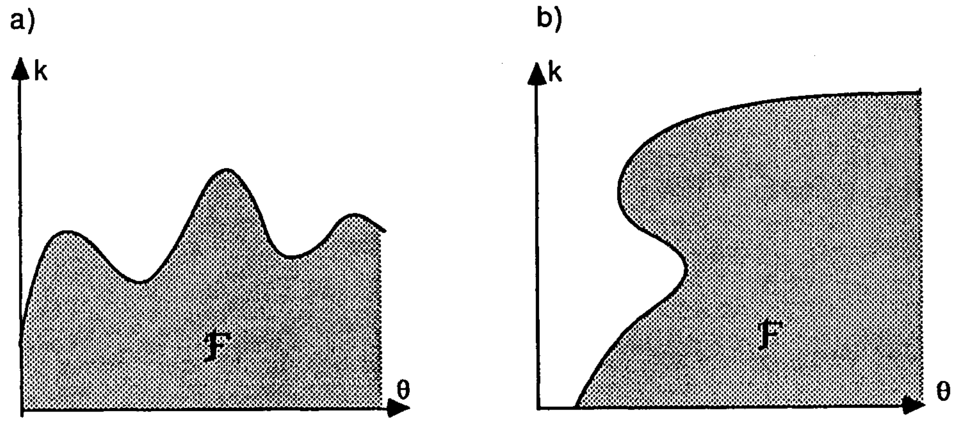

Figure I.5

The paper is organized as follows. In the next section, we present some preliminary results on the Green function of the operator A; this is quite standard material, but it will play such an important part throughout the paper that we find it practical to have it readily available on a form tailored to our needs. In Section 3, we briefly describe the controls we shall be working with, and also rewrite the criterion (1.3) on a form suitable for discontinuous controls. The long Section 4 is devoted to a heuristic derivation of the main results; we shall never try to justify the different steps in these calculations, but instead verify directly that the results they lead us to are, indeed, correct. In Section 5, we discuss briefly some of the conditions we shall have to impose in order to carry out this verification, and the verification itself you will find in Section 6; we first show that it suffices to prove that our proposed solution satisfies a certain variational inequality, and then check that this is really the case. In Section 7, we take a brief and informal look at what happens if we remove the monotonicity assumption on $\Pi_k + A\Gamma_k$, and in the final section we discuss the possibility of extending our results in various directions.

This paper has friends and relatives in both the economics and the mathematics literature. On the economics side, there are papers on irreversible investments by Brennan and Schwartz (1985), McDonald and Siegel (1986), and Pindyck (1988, 1991a,b) among others. There are other economic questions than the one we started with that are covered by the same mathematical formalism, and some of them are discussed in earlier papers by Kobila (1991, 1992 a,b); mathematically these papers deal with special cases of the general theory developed here. In the mathematics literature, there is by now a number of papers on singular controls, see, e.g., Beneš, Shepp, and Witsenhausen (1980), Karatzas and Shreve (1984, 1985, 1986), Gaver, Lehoczky, and Shreve (1984), Lehoczky and Shreve (1986), Sun (1987); and also the introductory paper by Shreve (1988) and the books by Harrison (1985) and Karatzas and Shreve (1988). The close connection between the Skorohod problem and optimal stopping discovered by El-Karoui and Karatzas (1991) in a finite horizon setting, transfers easily to the setting of the present paper and may turn out to be helpful in extending our results to more complicated and realistic models.

Before we begin, let us emphasize that our aim in this paper is to obtain as precise and explicit information as possible by quite elementary methods. Using more abstract machinery, it is undoubtedly possible to prove existence results in much more general settings, but this is not our purpose here.

2. The Green function.

Let

(2.1) $$d\Theta_t = \alpha\Theta_t dt + \beta\Theta_t dB_t, \quad \Theta_0 = \theta$$

be our one-dimensional, geometric Brownian motion. As should be clear from the introduction, we shall be interested in quantities of the form

(2.2) $$u(\theta) = E^\theta[\int_0^\infty f(\Theta_t)e^{-rt}dt],$$

where f is a given, continuous function. Standard arguments show that if the integrals in (2.2) converge, then u is a solution of the ordinary differential equation

(2.3) $$\frac{1}{2}\beta^2\theta^2u''(\theta) + \alpha\theta u'(\theta) - ru(\theta) = -f(\theta),$$

but the questions we want to ask in this section are when can we be sure that the integrals really do converge, what are the right boundary conditions to impose in (2.3), and what is the value of the integral (2.2)?

Before we begin, let us agree to write A for the differential operator in (2.3), i.e.

$$Au(\theta) = \frac{1}{2}\beta^2\theta^2 u''(\theta) + \alpha\theta u'(\theta) - ru(\theta) \tag{2.4}$$

We next observe that the homogeneous equation

$$Au = 0 \tag{2.5}$$

is just an Euler equation with general solution

$$u(\theta) = C_1\theta^{\gamma_1} + C_2\theta^{\gamma_2} \tag{2.6}$$

where γ_1 and γ_2 are the roots of the characteristic equation

$$\frac{\beta^2}{2}\gamma^2 + (\alpha - \frac{\beta^2}{2})\gamma - r = 0. \tag{2.7}$$

These two roots obviously have opposite signs, and we shall always take γ_1 to be the positive one.

We shall be interested in two function spaces associated with γ_1 and γ_2:

2.1 Definition.

a) $L^1_{\gamma_1,\gamma_2}$ is the space of all Lebesgue measurable functions $f : (0,\infty) \to \mathbf{R}$ such that the integrals

$$\int_0^a f(\eta)\eta^{-\gamma_2-1}d\eta \quad \text{and} \quad \int_a^\infty f(\eta)\eta^{-\gamma_1-1}d\eta \tag{2.8}$$

converge for all $a \in (0,\infty)$.

b) $C^2_{\gamma_1,\gamma_2}$ is the set of all twice continuously differentiable functions $u : (0,\infty) \to \mathbf{R}$ such that

$$\lim_{\theta\to 0}\theta^{-\gamma_2}u(\theta) = \lim_{\theta\to\infty}\theta^{-\gamma_1}u(\theta) = 0. \tag{2.9}$$

Using variation of parameters to find an expression for the solution of the inhomogeneous equation (2.3), we are led to the following result.

2.2 Proposition. Let f be a continuous function in $L^1_{\gamma_1,\gamma_2}$. Then the differential equation

$$Au = -f \tag{2.10}$$

has a unique solution in $C^2_{\gamma_1,\gamma_2}$, namely

$$u(\theta) = \frac{2}{(\gamma_1-\gamma_2)\beta^2}[\theta^{\gamma_2}\int_0^\theta f(\eta)\eta^{-\gamma_2-1}d\eta + \theta^{\gamma_1}\int_\theta^\infty f(\eta)\eta^{-\gamma_1-1}d\eta] \tag{2.11}$$

Proof. To check that the function in (2.11) really satisfies the differential equation is a trivial exercise which we leave to the reader.

Turning to the boundary condition at 0, we first observe that since

$$\theta^{-\gamma_2}\theta^{\gamma_2}\int_0^\theta f(\eta)\eta^{-\gamma_2-1}d\eta = \int_0^\theta f(\eta)\eta^{-\gamma_2-1}d\eta \to 0$$

by the integrability condition on f, we only have to check that

$$\theta^{\gamma_1-\gamma_2}\int_\theta^\infty f(\eta)\eta^{-\gamma_1-1}d\eta \to 0 \quad \text{as} \quad \theta \to 0.$$

But this is easy; given an $\epsilon > 0$, we first use the integrability condition to find a θ_0 such that

$$\int_0^{\theta_0} |f(\eta)|\eta^{-\gamma_2-1}d\eta < \frac{\epsilon}{2},$$

and then observe that

$$\begin{aligned}
&|\theta^{\gamma_1-\gamma_2}\int_\theta^\infty f(\eta)\eta^{-\gamma_1-1}d\eta| \\
&\leq \int_\theta^{\theta_0} |f(\eta)|\eta^{-\gamma_2-1}(\frac{\theta}{\eta})^{\gamma_1-\gamma_2}d\eta + \theta^{\gamma_1-\gamma_2}|\int_{\theta_0}^\infty f(\eta)\eta^{-\gamma_1-1}d\eta| \\
&\leq \frac{\epsilon}{2} + \theta^{\gamma_1-\gamma_2}|\int_{\theta_0}^\infty f(\eta)\eta^{-\gamma_1-1}d\eta|,
\end{aligned}$$

where we can get the last term less than $\frac{\epsilon}{2}$ by choosing θ small enough. The boundary condition at infinity is handled similarly.

Finally, to prove uniqueness we just observe that any other solution of (2.10) is of the form

$$u(\theta) + C_1\theta^{\gamma_1} + C_2\theta^{\gamma_2},$$

and hence has to violate at least one of the boundary conditions in (2.9).

□

It is often helpful to think of the result above in terms of Green functions. If we introduce the function

(2.12)
$$g(\theta,\eta) = \begin{cases} \frac{2}{(\gamma_1-\gamma_2)\beta^2}\theta^{\gamma_2}\eta^{-\gamma_2-1} & \text{if } \eta \le \theta \\ \frac{2}{(\gamma_1-\gamma_2)\beta^2}\theta^{\gamma_1}\eta^{-\gamma_1-1} & \text{if } \eta > \theta, \end{cases}$$

we can write (2.11) as

(2.13)
$$u(\theta) = \int_0^\infty g(\theta,\eta)f(\eta)d\eta,$$

and it's clear that g is the Green function of the operator A (or - more correctly - of $-A$).

Writing G for the operator

(2.14)
$$Gf(\theta) = \int_0^\infty g(\theta,\eta)f(\eta)d\eta =$$
$$= \frac{2}{(\gamma_1-\gamma_2)\beta^2}[\theta^{\gamma_2}\int_0^\theta f(\eta)\eta^{-\gamma_2-1}d\eta + \theta^{\gamma_1}\int_\theta^\infty f(\eta)\eta^{-\gamma_1-1}d\eta],$$

we can sum up the relationship between A and G in a simple corollary.

2.3 Corollary.

a) If f is a continuous function in $L^1_{\gamma_1,\gamma_2}$, then

(2.15)
$$AGf = -f$$

b) If $u \in C^1_{\gamma_1,\gamma_2}$ and $Au \in L^1_{\gamma_1,\gamma_2}$, then

(2.16)
$$GAu = -u.$$

Proof.

(a) By the proposition, $u = Gf$ is a solution of the equation $Au = -f$.

(b) Let $h = Au$ and $z = -Gh$. According to the proposition, $z \in C^2_{\gamma_1,\gamma_2}$ is a solution of the equation $Az = h$. Clearly, u is another such solution, and by the uniqueness part of the proposition, we must have $u = z$. □

It is now easy to answer the questions we asked at the beginning of this section.

2.4 Proposition. If $f \in L^1_{\gamma_1,\gamma_2}$, then

$$
\begin{aligned}
&E^\theta[\int_0^\infty f(\Theta_t)e^{-rt}dt] = \\
(2.17)\qquad & = \frac{2}{(\gamma_1-\gamma_2)\beta^2}[\theta^{\gamma_2}\int_0^\theta f(\eta)\eta^{-\gamma_2-1}d\eta + \theta^{\gamma_1}\int_\theta^\infty f(\eta)\eta^{-\gamma_1-1}d\eta]
\end{aligned}
$$

Proof: Assume first that f is bounded and continuous, and let

$$u(\theta) = E^\theta[\int_0^\infty f(\Theta_t)e^{-rt}dt].$$

There's a number of standard ways to see that u is a solution of the equation

$$Au = -f,$$

and since f is bounded, so is u. This means that $u \in C^2_{\gamma_1,\gamma_2}$, and hence

$$u(\theta) = \frac{2}{(\gamma_1-\gamma_2)\beta^2}[\theta^{\gamma_2}\int_0^\theta f(\eta)\eta^{-\gamma_2-1}d\eta + \theta^{\gamma_1}\int_\theta^\infty f(\eta)\eta^{-\gamma_1-1}d\eta]$$

by Proposition 2.2.

The rest of the proof is just an easy exercise in measure theory. If f is a bounded, measurable function, let $\{f_n\}$ be a bounded sequence of continuous function converging to f Lebesgue a.e. Observe that $f_n(\Theta_t) \to f(\Theta_t)$ a.s. for all $t > 0$. Since (2.17) holds for each f_n by what we have already shown, it also holds for f by the Dominated Convergence Theorem.

If f is a nonnegative, measurable and unbounded function in $L^1_{\gamma_1,\gamma_2}$, we approximate f by an increasing sequence $\{f_n\}$ of bounded measurable function. Since (2.17) holds for each f_n, the Monotone Convergence Theorem tells us that it also holds for f.

Finally, we extend the result to general functions in $L^1_{\gamma_1,\gamma_2}$ by treating the positive and negative parts separately. □

3. The controls and the criterion.

Let us first of all agree on what kinds of strategies we shall allow when we start at the point (θ, k) at time zero.

3.1 Definition. By a *(θ, k)-strategy* we shall mean a stochastic process

$$K : [0,\infty) \times \Omega \to [0,\infty)$$

with the following properties:

(i) K is non-decreasing and right continuous.
(ii) $K(0) \geq k$.
(iii) K_t is measurable with respect to the σ-algebra $\sigma\{\Theta_s : s \leq t\}$ generated by the process $\Theta_s = \theta e^{(\alpha-\beta^2/2)s+\beta B_s}$ up to time t.

Note that we only require that $K(0) \geq k$ and not that $K(0) = k$. If $K(0) > k$, we shall interpret this to mean that we are making an initial investment at time $t = 0$ in order to increase the capacity from k to $K(0)$ (another solution to this problem would, of course, be to use left continuous processes instead of right continuous ones, but we don't want to pay the price of having to work with an unconventional version of stochastic calculus).

When K has (absolutely) continuous paths and $K(0) = k$, it's clear from the economic formulation of our problem, that

$$J(K) = E^{(\theta,k)}\Big[\int_0^\infty e^{-rt}[\Pi(\Theta_t, K_t)dt - \Gamma_k(\Theta_t, K_t)dK_t]\Big] \tag{3.1}$$

is the intuitively correct expression for the total, discounted profit of the strategy K. When K has jumps, however, formula (3.1) does not give us the economically correct expression for the profit. To see why, consider the simple case where K is constant k until time t, then makes a jump to $k + \Delta k$ and remains constant ever after. According to the economic model, the total expansion costs of this strategy is

$$[\Gamma(k + \Delta k) - \Gamma(k)]e^{-rt},$$

but formula (3.1) yields

$$\Gamma_k(k + \Delta k)\Delta k e^{-rt}$$

To find a modification of (3.1) which works for all strategies, we proceed as follows. For a *continuous* strategy K with $K(0) = k$, we have by Itô's formula

$$\begin{aligned}
&\Gamma(\Theta_T, K_T)e^{-rT} - \Gamma(\theta, k) = \\
&= \int_0^T A\Gamma(\Theta_t, K_t)e^{-rt}dt + \int_0^T \beta\Theta_t\Gamma_\theta(\Theta_t, K_t)e^{-rt}dB_t + \\
&+ \int_0^T \Gamma_k(\Theta_t, K_t)e^{-rt}dK_t,
\end{aligned}$$

and hence

$$\begin{aligned}
&E^{\theta,k}\Big[\int_0^T \Gamma_k(\Theta_t, K_t)e^{-rt}dK_t\Big] = \\
&= -E^{\theta,k}\Big[\int_0^T A\Gamma(\Theta_t, K_t)e^{-rt}\Big]dt - \Gamma(\theta, k) + E^{\theta,k}[\Gamma(\Theta_T, K_T)e^{-rT}].
\end{aligned}$$

Under reasonable conditions, the last term will vanish as T goes to infinity, and we are left with

$$E^{\theta,k}[\int_0^\infty \Gamma_k(\Theta_t, K_t)e^{-rt}dK_t] = -E^{\theta,k}[\int_0^\infty A\Gamma(\Theta_t, K_t)e^{-rt}dt] - \Gamma(\theta, k).$$

Substituting this expression into (3.1), we get

$$J(K) = E^{(\theta,k)}[\int_0^\infty e^{-rt}(\Pi + A\Gamma)(\Theta_t, K_t)dt] + \Gamma(\theta, k) \tag{3.2}$$

Hence (3.1) and (3.2) are equivalent for continuous strategies K with $K(0) = k$, but the point is that (3.2) gives the economically correct value for the profit even when K has jumps. The easiest way to convince oneself that this is really the case, is probably by observing that if $\{K_n\}$ is a sequence of continuous strategies converging to K in a reasonable sense, then $J(K_n) \to J(K)$ if J is defined by (3.2) (but *not* if it is defined by (3.1)). Note also that if Γ is independent of θ, we have

$$-E^{(\theta,k)}[\int_0^\infty A\Gamma(K_t)e^{-rt}dt] - \Gamma(k) = \int_0^\infty r(\Gamma(K_t) - \Gamma(k))e^{-rt}dt$$

which corresponds to the cost of renting the extra capacity $K_t - k$ at a cost equal to the return of the investment cost.

We can now formulate the goal of this paper precisely.

3.2 Criterion. For each pair (θ, k) we want to find the (θ, k)-strategy which maximizes

$$J(K) = E^{(\theta,k)}[\int_0^\infty e^{-rt}(\Pi + A\Gamma)(\Theta_t, K_t)dt] + \Gamma(\theta, k) \tag{3.3}$$

(if it exists). We also want to compute the maximal profit

$$h(\theta, k) = \sup\{J(K) : K \quad \text{is a } (\theta, k)\text{-strategy}\} \tag{3.4}$$

□

Let us end this section by giving a more precise and formal description of what is going to be our optimal strategy. In the introduction we just described it as a non-decreasing process K which increased just enough to keep the pair (Θ_t, K_t) outside the forbidden region. But does such a process exist - and if so, what does it look like?

Let us assume that our forbidden region is of the form

$$\mathcal{F} = \{(\theta, k) : k \leq \phi(\theta)\}$$

for some continuous function $\phi : (0, \infty) \to [0, \infty)$. For each initial point (θ, k), what we really want is a process K with the following properties:

(i) $K_0 = k \vee \phi(\theta)$
(ii) $K_t \geq \phi(\Theta_t)$ for all t
(iii) K is a continuous, non-decreasing process which only increases at times when $K_t = \phi(\Theta_t)$; i.e., we want $\int (K_t - \phi(\Theta_t)) dK_t = 0$.

This is known as the *Skorohod problem* (see Skorohod (1961) and also the book by Karatzas and Shreve (1988)), and it has a very simple (and unique) solution; we just let

$$K(t) = k \vee \sup\{\phi(\Theta_s) : s \leq t\}. \tag{3.5}$$

This will be our candidate for the optimal strategy once the forbidden set $\mathcal{F}$ has been determined. Note that K is singular in the sense that it only increases on a set of $t's$ of Lebesgue measure zero. Note also that if f is a continuous function which is identically zero in the forbidden region $\mathcal{F}$, then

$$\int f(\Theta_t, K_t) dK_t = 0. \tag{3.6}$$

This will be an important property in Section 6.

4. Some heuristic calculations.

In this section we present a heuristic derivation of the results we described in the introduction. There are two reasons why we have to classify our arguments as heuristic; not only are we assuming that all the functions we meet are sufficiently smooth and regular for our calculations to be valid, but we are also assuming that the nature of the solution is really of the kind described in the introduction. It may be useful to know already at this stage that we shall never attempt to give a formal justification of our procedure; instead we shall verify by a direct argument that the results we arrive at are correct.

We shall assume that the optimal strategies can be described in terms of a "forbidden region" $\mathcal{F}$ as explained in the introduction. We also assume that $\mathcal{F}$ is on the form

$$\mathcal{F} = \{(\theta, k) : \psi_1(k) \leq \theta \leq \psi_2(k)\},$$

where ψ_1 is an increasing function and ψ_2 a decreasing one. Our task is determine ψ_1 and ψ_2.

It's often useful to describe the boundary of $\mathcal{F}$ as one function of θ instead of two functions of k, an hence we introduce

$$\phi(\theta) = \sup\{k : (\theta, k) \in \mathcal{F}\}, \tag{4.1}$$

(see Figure IV.1)

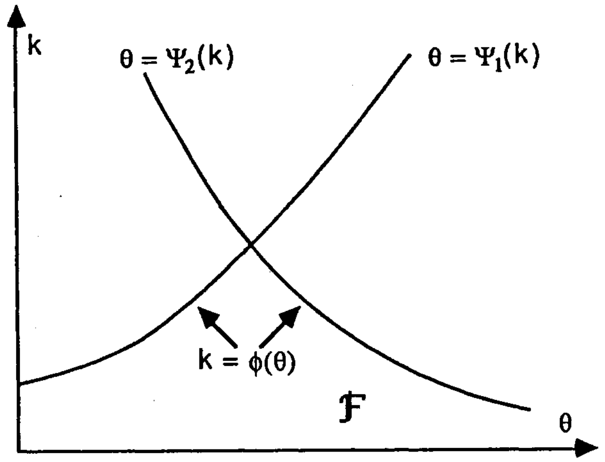

Figure IV.1

Our arguments are based on a simple analysis of what Ah and h_k are like inside and outside the forbidden region (recall that h is the optimal profit defined by (3.4)). Let us begin by starting the optimal process at two points (θ, k_1) and (θ, k_2) inside the forbidden region with the same first coordinate. The two processes will both start by jumping immediately to $(\theta, \phi(\theta))$ and from then on behave in exactly the same way. Since the difference in the profit is entirely due to the difference in the initial investment, we have

$$h(\theta, k_1) - h(\theta, k_2) = \Gamma(\theta, k_1) - \Gamma(\theta, k_2),$$

which means that

$$h_k(\theta, k) = \Gamma_k(\theta, k) \quad \text{when } (\theta, k) \in \mathcal{F}. \tag{4.2}$$

On the other hand, if we start at a point outside $\mathcal{F}$, we must have $h_k \leq \Gamma_k$ since we could otherwise improve the optimal profit by making an investment at time zero. Thus

$$h_k(\theta, k) \leq \Gamma_k(\theta, k) \quad \text{when } (\theta, k) \notin \mathcal{F}. \tag{4.3}$$

Also, if we start outside $\mathcal{F}$, we are not making any immediate investment, and hence

$$Ah(\theta, k) = -\Pi(\theta, k) \quad \text{when } (\theta, k) \notin \mathcal{F}. \tag{4.4}$$

Finally, we want to find $Ah(\theta, k)$ when (θ, k) is in the forbidden region. Since the optimal process jumps immediately to $(\theta, \phi(\theta))$, we clearly have

$$h(\theta, k) = h(\theta, \phi(\theta)) - \Gamma(\theta, \phi(\theta)) + \Gamma(\theta, k)$$

If we write v for the function $h - \Gamma$, we get

$$Ah(\theta, k) = \frac{\beta^2}{2}\theta^2 \frac{d^2}{d\theta^2} v(\theta, \phi(\theta)) + \alpha\theta \frac{d}{d\theta} v(\theta, \phi(\theta)) - rv(\theta, \phi(\theta)) + A\Gamma(\theta, k).$$

Observe next that

$$\frac{d}{d\theta}v(\theta,\phi(\theta)) = v_\theta(\theta,\phi(\theta)) + v_k(\theta,\phi(\theta))\phi'(\theta)$$
$$= v_\theta(\theta,\phi(\theta))$$

since $v_k(\theta,\phi(\theta)) = 0$ by (4.2). Similarly,

$$\frac{d^2}{d\theta^2}v(\theta,\phi(\theta)) = \frac{d}{d\theta}v_\theta(\theta,\phi(\theta)) = v_{\theta\theta}(\theta,\phi(\theta)),$$

and hence

$$Ah(\theta,k) = Av(\theta,\phi(\theta)) + A\Gamma(\theta,k) = Ah(\theta,\phi(\theta)) - A\Gamma(\theta,\phi(\theta)) + A\Gamma(\theta,k).$$

(our notation is a little ambiguous at this point; note that by $Av(\theta,\phi(\theta))$ we mean the function Av evaluated at a point (θ,k) which happens to be of the form $(\theta,\phi(\theta))$; we do *not* mean the operator A applied to the composite function $v(\theta,\phi(\theta))$. Similarly, of course, for $Ah(\theta,\phi(\theta))$ and $A\Gamma(\theta,\phi(\theta))$.)

By (4.4), $Ah(\theta,k) = -\Pi(\theta,k)$ outside $\mathcal{F}$, and since $(\theta,\phi(\theta))$ is on the boundary of $\mathcal{F}$, a weak continuity assumption leads to

$$Ah(\theta,\phi(\theta)) = -\Pi(\theta,\phi(\theta)),$$

and hence we get

$$Ah(\theta,k) = -(\Pi + A\Gamma)(\theta,\phi(\theta)) + A\Gamma(\theta,k) \quad \text{for } (\theta,k)\in\mathcal{F}. \tag{4.5}$$

To sum up, we have

$$h_k(\theta,k)\begin{cases} \le \Gamma_k(\theta,k) & \text{if } (\theta,k)\notin\mathcal{F} \\ = \Gamma_k(\theta,k) & \text{if } (\theta,k)\in\mathcal{F} \end{cases} \tag{4.6}$$

$$Ah(\theta,k) = \begin{cases} -\Pi(\theta,k) & \text{if } (\theta,k)\notin\mathcal{F} \\ -(\Pi + A\Gamma)(\theta,\phi(\theta)) + A\Gamma(\theta,k) & \text{if } (\theta,k)\in\mathcal{F}. \end{cases} \tag{4.7}$$

These equations become nicer and more symmetric if we express them in terms of the function

$$v(\theta,k) = h(\theta,k) - \Gamma(\theta,k) \tag{4.8}$$

instead of h:

$$v_k(\theta,k)\begin{cases} \le 0 & \text{if } (\theta,k)\notin\mathcal{F} \\ = 0 & \text{if } (\theta,k)\in\mathcal{F} \end{cases} \tag{4.9}$$

$$Av(\theta,k) = \begin{cases} -(\Pi + A\Gamma)(\theta,k) & \text{if } (\theta,k) \notin \mathcal{F} \\ -(\Pi + A\Gamma)(\theta,\phi(\theta)) & \text{if } (\theta,k) \in \mathcal{F}. \end{cases} \tag{4.10}$$

Remark: These formulas immediately bring the theory of variational inequalities to mind, and the problem may, in fact, be treated from that point of view. However, in this paper we prefer a more pedestrian and direct approach.

□

Combining (4.10) and (2.13), we can now express v in terms of ψ_1, ψ_2, and the Green function g in (2.12):

$$\begin{aligned} v(\theta,k) = &\int_0^{\psi_1(k)} g(\theta,\eta)(\Pi + A\Gamma)(\eta,k)d\eta + \\ &+ \int_{\psi_1(k)}^{\psi_2(k)} g(\theta,\eta)(\Pi + A\Gamma)(\eta,\phi(\eta))d\eta + \int_{\psi_2(k)}^{\infty} g(\theta,\eta)(\Pi + A\Gamma)(\eta,k)d\eta \end{aligned} \tag{4.11}$$

Differentiating with respect to k, we get

$$\begin{aligned} v_k(\theta,k) = &\int_0^{\psi_1(k)} g(\theta,\eta)(\Pi_k + A\Gamma_k)(\eta,k)d\eta + \\ &+ g(\theta,\psi_1(k))(\Pi + A\Gamma)(\psi_1(k),k)\psi_1'(k) \\ &+ g(\theta,\psi_2(k))(\Pi + A\Gamma)(\psi_2(k),k)\psi_2'(k) \\ &- g(\theta,\psi_1(k))(\Pi + A\Gamma)(\psi_1(k),k)\psi_1'(k) \\ &+ \int_{\psi_2(k)}^{\infty} g(\theta,\eta)(\Pi_k + A\Gamma_k)(\eta,k)d\eta \\ &- g(\theta,\psi_2(k))(\Pi + A\Gamma)(\psi_2(k),k)\psi_2'(k) \\ = &\int_0^{\psi_1(k)} g(\theta,\eta)(\Pi_k + A\Gamma_k)(\eta,k)d\eta + \int_{\psi_2(k)}^{\infty} g(\theta,\eta)(\Pi_k + A\Gamma_k)(\eta,k)d\eta \end{aligned} \tag{4.12}$$

If we now assume that (θ,k) is in the forbidden region, then $\psi_1(k) < \theta < \psi_2(k)$, and inserting the explicit expression (2.12) for g, we have

$$\begin{aligned} v_k(\theta,k) = &\frac{2}{(\gamma_1 - \gamma_2)\beta^2}[\theta^{\gamma_2} \int_0^{\psi_1(k)} \frac{(\Pi_k + A\Gamma_k)(\eta,k)}{\eta^{\gamma_2+1}}d\eta + \\ &+ \theta^{\gamma_1} \int_{\psi_2(k)}^{\infty} \frac{(\Pi_k + A\Gamma_k)(\eta,k)}{\eta^{\gamma_1+1}}d\eta] \end{aligned} \tag{4.13}$$

According to (4.9), this expression must be equal to zero for all θ between $\psi_1(k)$ and $\psi_2(k)$, and the only way to achieve this is to have

(4.14)
$$\int_0^{\psi_1(k)} \frac{(\Pi_k + A\Gamma_k)(\eta, k)}{\eta^{\gamma_2+1}} d\eta = 0$$

(4.15)
$$\int_{\psi_2(k)}^{\infty} \frac{(\Pi_k + A\Gamma_k)(\eta, k)}{\eta^{\gamma_1+1}} d\eta = 0.$$

If we also take into account that $v_k \leq 0$ outside the forbidden region, we are led to formulas (1.7) and (1.8) in the introduction, i.e.

(4.16)
$$\psi_1(k) = \inf\{\theta : \int_0^{\theta} \frac{(\Pi_k + A\Gamma_k)(\eta, k)}{\eta^{\gamma_2+1}} d\eta > 0\}$$

(4.16)
$$\psi_2(k) = \sup\{\theta : \int_{\theta}^{\infty} \frac{(\Pi_k + A\Gamma_k)(\eta, k)}{\eta^{\gamma_1+1}} d\eta > 0\}$$

Once we have identified ψ_1 and ψ_2, we can use (4.11) to compute the optimal profit:

(4.18)
$$h(\theta, k) = \Gamma(\theta, k) + \int_0^{\psi_1(k)} g(\theta, \eta)(\Pi + A\Gamma)(\eta, k) d\eta +$$
$$+ \int_{\psi_1(k)}^{\psi_2(k)} g(\theta, \eta)(\Pi + A\Gamma)(\eta, \phi(\theta)) d\eta + \int_{\psi_2(k)}^{\infty} g(\theta, \eta)(\Pi + A\Gamma)(\eta, k) d\eta,$$

or - changing variable in the second integral -

(4.19)
$$h(\theta, k) = \Gamma(\theta, k) + \int_0^{\psi_1(k)} g(\theta, \eta)(\Pi + A\Gamma)(\eta, k) d\eta +$$
$$+ \int_{\psi_2(k)}^{\infty} g(\theta, \eta)(\Pi + A\Gamma)(\eta, k) d\eta +$$
$$+ \int_k^{k\max} g(\theta, \psi_1(K))(\Pi + A\Gamma)(\psi_1(K), K)\psi_1'(K) dK -$$
$$- \int_k^{k\max} g(\theta, \psi_2(K))(\Pi + A\Gamma)(\psi_2(K), K)\psi_2'(K) dK,$$

where $k_{\max}$ is the largest k-value in the forbidden region.

It's quite informative at this stage to take a look at an example:

4.1 Example. We choose the parameters of the geometric Brownian motion to be $\beta = \sqrt{2}, \alpha = 1$, and we set the discount parameter r equal to 4. (Economists who find an interest rate of 400% per year slightly eccentric, may want to measure time in other units.) Then the characteristic polynomial (1.6) is

$$\frac{\beta^2}{2}\gamma^2 + (\alpha - \frac{\beta^2}{2})\gamma - r = \gamma^2 - 4 = (\gamma - 2)(\gamma + 2),$$

i.e. $\gamma_1 = 2, \gamma_2 = -2$. We also choose

$$\Pi(\theta, k) = 2k$$
$$\Gamma(\theta, k) = \frac{k^2}{2}(\frac{\theta}{3} + \frac{1}{\theta});$$

in economic terms this means that the profit from the production is proportional to the capacity, while the marginal investment costs

$$\Gamma_k(\theta, k) = k(\frac{\theta}{3} + \frac{1}{\theta})$$

are increasing as a function of k, but has a minimum at $\theta = \sqrt{3}$ as a function of θ. Since

$$(\Pi_k + A\Gamma_k)(\theta, k) = 2 - k\theta - 3k\theta^{-1},$$

(4.14) becomes

$$0 = \int_0^{\psi_1(k)} \frac{2 - k\eta - 3k\eta^{-1}}{\eta^{-1}} d\eta = -\frac{k}{3}\psi_1(k)^3 + \psi_1(k)^2 - 3k\psi_1(k)$$

and (4.15) turns into

$$0 = \int_{\psi_2(k)}^{\infty} \frac{2 - k\eta - 3k\eta^{-1}}{\eta^3} d\eta = -k\psi_2(k)^{-1} + \psi_2(k)^{-2} - k\psi_2(k)^{-3}$$

Solving these equations for $\psi_1(k)$ and $\psi_2(k)$, we get

$$\psi_1(k) = \begin{cases} 0 & \text{for } k \geq \frac{1}{2} \\ 0, \frac{3}{2k}(1 \pm \sqrt{1 - 4k^2}) & \text{for } k < \frac{1}{2} \end{cases}$$

and

$$\psi_2(k) = \frac{1}{2k}(1 \pm \sqrt{1 - 4k^2}) \quad \text{for } k \leq \frac{1}{2}$$

If we then take (4.16) and (4.17) into account, we see that the actual solutions are

$$\psi_1(k) = \begin{cases} \frac{3}{2k}(1 - \sqrt{1-4k^2}) & \text{for } k \leq \frac{1}{2} \\ \infty & \text{for } k \geq \frac{1}{2} \end{cases} \tag{4.20}$$

$$\psi_2(k) = \begin{cases} \frac{1}{2k}(1 + \sqrt{1-4k^2}) & \text{for } k < \frac{1}{2} \\ 0 & \text{for } k \geq \frac{1}{2} \end{cases} \tag{4.21}$$

Figure IV.2 shows the regions

$$\mathcal{F}_1 = \{(\theta, k) : \psi_1(k) \leq \theta\}$$
$$\mathcal{F}_2 = \{(\theta, k) : \theta \leq \psi_2(k)\},$$

whose intersection is $\mathcal{F}$.

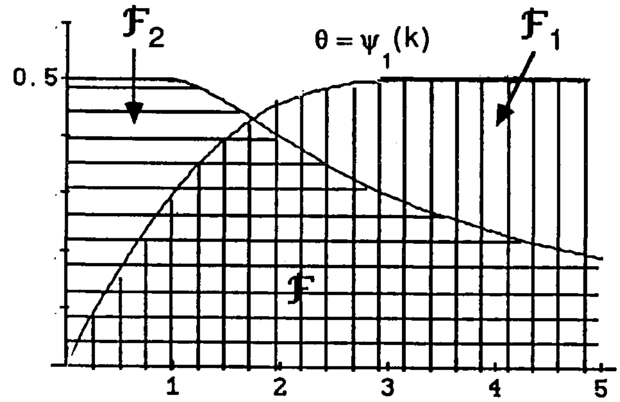

Figure IV.2

θ = ψ2(k)

Among other things, this example tells us that even for very simple choices of Π and Γ, we can not expect ψ_1 and ψ_2 to be continuous. Observe also that ψ_1 and ψ_2 intersect at $\theta = \sqrt{3}$, and hence the most favorable climate for expansion is when Θ_t is at $\sqrt{3}$ - as we would expect from our discussion of Γ.

□

In the remaining sections of the paper, we shall show that under reasonable assumptions, the heuristic arguments above do, in fact, give the correct solution to our problem.

5. The monotonicity condition.

Most of the conditions we shall have to impose in order to verify our solution will be quite innocent growth and differentiability assumptions on Π and Γ. But two of our conditions will be of a different nature, and in this section we shall discuss one of these - namely that

$$(\Pi_k + A\Gamma_k)(\theta, k) \quad \text{is strictly decreasing in } k. \tag{5.1}$$

This condition comes from a consistency problem in the calculations above; we started out by assuming that ψ_1 was increasing and ψ_2 decreasing, but the final answer (4.16), (4.17) shows that this is only true when (5.1) is satisfied. It turns out that this problem is more serious than it may seem at first glance; in fact, the entire argument in Section 4 breaks down when (5.1) is not satisfied (to be precise, the calculations in (4.12) are no longer valid). In Section 7, we shall sketch how our solution can be modified to hold even when (5.1) is not satisfied, but in the main body of the paper, we shall solve the problem simply by assuming (5.1).

But how reasonable is this assumption? Fortunately, it turns out that (5.1) has a quite intuitive economic interpretation. Assume that we have decided to expand our capacity by a small ("infinitesimal") amount Δk, and that we want to know whether it's better to make this investment immediately or to wait a short time Δt. If we invest immediately, the net profit from the period $[0, \Delta t]$ will be

$$E^\theta[\int_0^{\Delta t} \Pi(\Theta_t, k + \Delta k)e^{-rt}dt - \Gamma(\theta, k + \Delta k) + \Gamma(\theta, k)]$$

and if we wait, the net profit will be

$$E^\theta[\int_0^{\Delta t} \Pi(\Theta_t, k)e^{-rt}dt - e^{-r\Delta t}(\Gamma(\Theta_{\Delta t}, k + \Delta k) - \Gamma(\Theta_{\Delta t}, k))]$$

Subtracting the second quantity from the first, we see that the difference is of order of magnitude

$$(\Pi_k + A\Gamma_k)(\theta, k)\Delta k \Delta t.$$

Hence we may call $\Pi_k + A\Gamma_k$ the *local investment incentive* - it measures how profitable a small investment is on a short term basis. In most situations it is reasonable to assume that the local investment incentive is decreasing in k for two reasons; partly because as the market gets more and more saturated, it gets increasingly difficult to sell the extra products without lowering prices, and partly because as we go on expanding, we run out of cheap expansion alternatives, and have to turn to more expensive ones. On the other hand, it is not difficult to think of situations where the local investment incentive is *not* always decreasing in k; e.g., if a company is trying to break into a new market or to establish itself abroad, the first investment will usually be much more expensive than later ones.

Let us also take a brief look at another problem. From the interpretation of $(\Pi_k + A\Gamma_k)$ as the local investment incentive, it is easy to see that in many situations it will not be unnatural to assume that

$$(5.2) \qquad (\Pi_k + A\Gamma_k)(\theta, k) \quad \text{is non-decreasing in} \quad \theta,$$

but again we can think of exceptions - e.g., if the investment costs react more drastically to changes in θ than does the price of our product. From the mathematical point of view, (5.2) has the unpleasant effect of eliminating the need for the function ψ_2 (we would always be in the situation shown in Figure I.3.a)), and for that reason also we shall not adopt it. But we shall have to introduce a weaker, but related condition which will be described in the next section.

6. Verifying the results.

We have now reached the stage where we can verify that under reasonable conditions, the heuristic arguments in Section 4 lead to correct conclusions. Let us first give an abstract description of what we are looking for - it may be helpful to compare (iii) and (iv) below to formulas (4.8)-(4.10) in Section 4.

6.1 Proposition. Assume that there exist a region $\mathcal{F}$ and a bounded, continuous function v satisfying the following conditions:

(i) $\mathcal{F} = \{(\theta,k) : k \le \phi(\theta)\}$ for some continuous function $\phi : (0,\infty) \to [0,\infty)$.
(ii) The partial derivatives v_θ, v_k and Av exist and are continuous.
(iii)

$$Av(\theta,k) \begin{cases} = -(\Pi + A\Gamma)(\theta,k) & \text{when } (\theta,k) \notin \mathcal{F} \\ \le -(\Pi + A\Gamma)(\theta,k) & \text{when } (\theta,k) \in \mathcal{F} \end{cases}$$

(iv)

$$v_k(\theta,k) = \begin{cases} \le 0 & \text{when } (\theta,k) \notin \mathcal{F} \\ = 0 & \text{when } (\theta,k) \in \mathcal{F}. \end{cases}$$

Then for any initial point (θ,k) and any (θ,k)-strategy K

(6.1) $$J(K) \le v(\theta,k) - \Gamma(\theta,k).$$

If $K_t = k \vee \sup\{\phi(\Theta_s) : s \le t\}$, then equality holds in (6.1).

Proof: For any (θ,k)-strategy K, we have by Itô's formula

$$e^{-rT}v(\Theta_T,K_T) - v(\Theta_0,K_0) = \int_0^T Av(\Theta_t,K_t)e^{-rt}dt +$$

(6.2) $$+ \int_0^T \beta\Theta_t v_\theta(\Theta_t,K_t)e^{-rt}dB_t +$$

$$+ \int_0^T v_k(\Theta_t,K_t)e^{-rt}dK_t^c + \sum e^{-rt}\Delta v(\Theta_t,K_t),$$

where K^c is the continuous part of K, and the final sum is over all $K's$ jumps before time T. Note that by (iv), $v(\theta, k) \geq v(\Theta_0, K_0)$ and the two last terms in (6.2) are negative. Hence we get

$$(6.3) \qquad \begin{aligned} v(\theta, k) \geq & -\int_0^T Av(\Theta_t, K_t)e^{-rt}dt + e^{-rT}v(\Theta_T, K_T) - \\ & -\int_0^T \beta\Theta_t v_\theta(\Theta_t, K_t)e^{-rt}dB_t \end{aligned}$$

By (iii), $-Av(\Theta_t, K_t) \geq (\Pi + A\Gamma)(\Theta_t, K_t)$, and since the martingale term disappears when we take expectations (using a localization argument if necessary), we see that

$$(6.4) \qquad v(\theta, k) \geq E^{(\theta,k)}[\int_0^T (\Pi + A\Gamma)(\Theta_t, K_t)e^{-rt}dt + e^{-rT}v(\Theta_T, K_T)],$$

and letting T go to infinity, we get

$$\begin{aligned} v(\theta, k) &\geq E^{(\theta,k)}[\int_0^\infty (\Pi + A\Gamma)(\Theta_t, K_t)e^{-rt}dt] \\ &= J(K) - \Gamma(\theta, k), \end{aligned}$$

which proves the first part of the proposition.

For the second part, we let

$$K_t = k \vee \sup\{\phi(\Theta_s) : s \leq t\},$$

and check that in this case all the inequalities above are, in fact, equalities. Since (Θ_t, K_t) now never enters the forbidden region, we always have $-Av(\Theta_t, K_t) = (\Pi + A\Gamma)(\Theta_t, K_t)$. Moreover, since $v_k(\Theta_t, K_t) = 0$ whenever K increases, we always have

$$\int v_k(\Theta_t, K_t)e^{-rt}dK_t = 0,$$

and the sum in (6.2) is trivially zero since K does not have any jumps (recall that an investment at time 0 is not reflected in K). By the second half of (iv), $v(\Theta_0, K_0) = v(\theta, k)$. Hence (6.4) holds with equality, and it follows that

$$v(\theta, k) = J(K) - \Gamma(\theta, k).$$

□

We now let $\mathcal{F}$ and v be as in Section 4, i.e.,

$$\mathcal{F} = \{(\theta, k) : \psi_1(k) \leq \theta \leq \psi_2(k)\} \tag{6.5}$$

where

$$\psi_1(k) = \inf\{\theta : \int_0^\theta \frac{(\Pi_k + A\Gamma)(\eta, k)}{\eta^{\gamma_2+1}} d\eta > 0\} \tag{6.6}$$

$$\psi_2(k) = \sup\{\theta : \int_\theta^\infty \frac{(\Pi_k + A\Gamma)(\eta, k)}{\eta^{\gamma_1+1}} d\eta > 0\}, \tag{6.7}$$

and - recall formula (4.11) -

$$v(\theta, k) = \int_{\mathcal{F}_k^c} g(\theta, \eta)(\Pi + A\Gamma)(\eta, k) d\eta + \int_{\mathcal{F}_k} g(\theta, \eta)(\Pi + A\Gamma)(\eta, \phi(\eta)) d\eta \tag{6.8}$$

where $\mathcal{F}_k = \{\theta : \psi_1(k) \leq \theta \leq \psi_2(k)\}$ is the cross-section of $\mathcal{F}$ at k, and $\mathcal{F}_k^c = \mathbf{R}_+ \setminus \mathcal{F}_k$ is its complement. To show that $\mathcal{F}$ and v satisfies the assumptions of Proposition 6.1, we need to impose the following conditions.

6.2 Conditions.

(i) Π is nonnegative, Γ is non-decreasing in the k variable, and the partial derivatives $\frac{\partial \Pi}{\partial k}$ and $\frac{\partial^{n+m}\Gamma}{\partial \theta^n \partial k^m}$, $n = 0, 1, 2$, $m = 0, 1$, exist and are continuous.

(ii) The function $M(\theta) = \sup\{|(\Pi + A\Gamma)(\theta, k)| : k > 0\}$ belongs to $L^1_{\gamma_1,\gamma_2}$, and so does the function $N^{(k)}(\theta) = \sup\{|(\Pi_k + A\Gamma_k)(\theta, y)| : 0 < y \leq k\}$ for each k.

(iii) For each θ

$$\phi(\theta) = \sup\{k : (\theta, k) \in \mathcal{F}\}$$

is finite.

(iv) The function $(\Pi_k + A\Gamma_k)(\theta, k)$ is strictly decreasing in the second variable.

(v) For all $(\theta, k) \in \mathcal{F}$

$$(\Pi + A\Gamma)(\theta, k) \leq (\Pi + A\Gamma)(\theta, \phi(\theta))$$

□

The first two conditions above just guarantee that Π and Γ are sufficiently regular for our derivatives to exist and our integrals to converge, and the third condition rules out the situation in Figure I.4 where no optimal strategy exists. Condition (iv) is the monotonicity condition we discussed in great detail in Section 5, and which we shall return to in Section

7. The fifth condition is needed to pass from formula (4.10) to condition (iv) in Proposition 6.1. As condition (v) can be difficult to check directly, we prove the following simple, but quite useful criteria (note that the first of these follows from condition (5.2) discussed at the end of the preceding section, but in many situations it is more natural to think of it as a concavity assumption).

6.3 Lemma. Assume that conditions (i)-(iv) above are satisfied. If one of the following two assumptions hold, then (v) is also satisfied:

a) For each k, the set $I_k = \{\theta : (\Pi_k + A\Gamma_k)(\theta, k) \geq 0\}$ is connected.

b) Each point on the boundary of $\mathcal{F}$ is of the form $(\psi_1(k), k)$, $(\psi_2(k), k)$, $(0, k)$ or $(\theta, 0)$ (this means that ψ_1 and ψ_2 can not have jumps in the region where they make up the boundary of $\mathcal{F}$).

Proof: a) We first observe that by definition of ψ_1 and ψ_2,

$$(\Pi_k + A\Gamma_k)(\psi_i(k), k) \geq 0 \quad i = 1, 2, \tag{6.9}$$

and hence $\psi_1(k), \psi_2(k) \in I_k$ (note that this is even the case when $\psi_2(k) = \infty$ in the sense that then $\sup I_k = \infty$). If $(\theta, k) \in \mathcal{F}$, then $\psi_1(k) \leq \theta \leq \psi_2(k)$, and since I_k is connected, this means that $\theta \in I_k$. But this implies that $(\Pi_k + A\Gamma_k)(\theta, k) \geq 0$ for all $(\theta, k) \in \mathcal{F}$, and condition (v) follows.

b) For each $(\theta, k) \in \mathcal{F}$, it follows from (6.9) and the assumption that $(\Pi_k + A\Gamma_k)$ $(\theta, \phi(\theta)) \geq 0$. Since $(\Pi_k + A\Gamma_k)$ is decreasing in k, this means that $(\Pi_k + A\Gamma_k)$ $(\theta, k) \geq 0$, and condition (v) follows as above.

□

We can now begin to check that under the conditions in 6.2, the assumptions in Proposition 6.1 are satisfied for the function v in (6.8).

6.4 Lemma. v is bounded and has partial derivatives $v_\theta, v_{\theta\theta}$. Moreover,

$$Av(\theta, k) = \begin{cases} -(\Pi + A\Gamma)(\theta, k) & \text{when } (\theta, k) \notin \mathcal{F} \\ -(\Pi + A\Gamma)(\theta, \phi(\theta)) \leq -(\Pi + A\Gamma)(\theta, k) & \text{when } (\theta, k) \in \mathcal{F}. \end{cases} \tag{6.10}$$

Proof: By Condition 6.2(i), the function

$$H(\theta, k) = \begin{cases} (\Pi + A\Gamma)(\theta, k) & \text{if } (\theta, k) \notin \mathcal{F} \\ (\Pi + A\Gamma)(\theta, \phi(\theta)) & \text{if } (\theta, k) \in \mathcal{F} \end{cases}$$

is bounded by a function $M(\theta)$ which belongs to $L^1_{\gamma_1,\gamma_2}$, and hence $\theta \longmapsto H(\theta, k)$ belongs to $L^1_{\gamma_1,\gamma_2}$ for each k. Since

$$v(\theta, k) = \int_0^\infty g(\theta, \eta) H(\eta, k) d\eta,$$

v is bounded by

$$\int_0^\infty g(\theta,\eta)|M(\eta)|d\eta < \infty,$$

and by Proposition 2.2, v is twice continuously differentiable with respect to θ, and satisfies

$$Av(\theta,k) = -H(\theta,k).$$

An appeal to Condition 6.2(v) completes the proof.

□

Before we turn to differentiability in the k variable, it is convenient to prove a simple lemma about ψ_1 and ψ_2.

6.5 Lemma. ψ_1 is an increasing function which is continuous from above, while ψ_2 is a decreasing function which is continuous from below.

Proof: The monotonicity follows immediately from Condition 6.2(iv) and the definitions of ψ_1 and ψ_2. To prove that ψ_1 is continuous from above, fix a k and an $\epsilon > 0$. Pick θ such that $\psi_1(k) < \theta < \psi_1(k) + \epsilon$ and

$$\int_0^\theta \frac{(\Pi_k + A\Gamma_k)(\eta,k)}{\eta^{\gamma_2+1}} d\eta > 0. \tag{6.11}$$

It clearly suffices to show that if we choose $y > k$ sufficiently close to k, then

$$\int_0^\theta \frac{(\Pi_k + A\Gamma_k)(\eta,y)}{\eta^{\gamma_2+1}} d\eta > 0. \tag{6.12}$$

By the assumptions on Π and Γ, the integrand in (6.12) converges pointwise to the integrand in (6.11) as $y \downarrow k$, and by the second part of Condition 6.2(ii) all the integrands are dominated by an integrable function. By Dominated Convergence Theorem, the integrals in (6.12) converge to the integral in (6.11) and thus they eventually have to become positive.

That ψ_2 is continuous from below is proved analogously.

□

6.6 Lemma. v is differentiable with respect to k, and

$$v_k(\theta,k) = \int_{F_k^c} g(\theta,\eta)(\Pi_k + A\Gamma_k)(\eta,k)d\eta \tag{6.13}$$

Proof: If $k_2 > k_1$ and $\Delta k = k_2 - k_1$, then by the Mean Value Theorem there is a function $c(\eta)$ taking values between k_1 and k_2 such that

$$\frac{v(\theta,k_2)-v(\theta,k_1)}{\Delta k} = \Delta k^{-1}\{ \int_0^{\psi_1(k_1)} g(\theta,\eta)[(\Pi + A\Gamma)(\eta,k_2) - (\Pi + A\Gamma)(\eta,k_1)]d\eta +$$

$$+ \int_{\psi_1(k_1)}^{\psi_1(k_2)} g(\theta,\eta)[(\Pi + A\Gamma)(\eta,k_2) - (\Pi + A\Gamma)(\eta,\phi(\eta))]d\eta +$$

$$+ \int_{\psi_2(k_2)}^{\psi_2(k_1)} g(\theta,\eta)[(\Pi + A\Gamma)(\eta,k_2) - (\Pi + A\Gamma)(\eta,\phi(\eta))]d\eta +$$

$$+ \int_{\psi_2(k_1)}^{\infty} g(\theta,\eta)[(\Pi + A\Gamma)(\eta,k_2) - (\Pi + A\Gamma)(\eta,k_1)]d\eta\} =$$

$$= \int_0^{\psi_1(k_2)} g(\theta,\eta)(\Pi_k + A\Gamma_k)(\eta,c(\eta))d\eta + \int_{\psi_2(k_2)}^{\infty} g(\theta,\eta)(\Pi_k + A\Gamma_k)(\eta,c(\eta))d\eta$$

If we first fix k_2 and let $k_1 \uparrow k_2$, then the Dominated Convergence Theorem combined with the second half of Condition 6.2(ii), assures us that the left derivative v_k^- exists and equals the expression in (6.13). If we then fix k_1 and let $k_2 \downarrow k_1$, we can combine the argument we just gave with an appeal to Lemma 6.5 to see that the right derivative v_k^+ also exists and equals the right hand side of (6.13).

□

Before we show that assumption (iv) in Proposition 6.1 is satisfied, we make a simple observation.

6.7 Lemma. Let f be a continuous function such that

$$\int_\theta^a f(\eta)d\eta \geq 0 \quad \text{for all} \quad \theta \in (0,a).$$

Then

$$\int_\theta^a f(\eta)(1-(\frac{\theta}{\eta})^{\gamma_1-\gamma_2})d\eta \geq 0 \quad \text{for all} \quad \theta \in (0,a).$$

Proof: Define

$$g(\theta) = \theta^{\gamma_2-\gamma_1}\int_\theta^a f(\eta)d\eta - \int_\theta^a f(\eta)\eta^{\gamma_2-\gamma_1}d\eta$$

and observe that $g(a) = 0$ and

$$g'(\theta) = (\gamma_2 - \gamma_1)\theta^{\gamma_2-\gamma_1-1} \int_\theta^a f(\eta)d\eta - \theta^{\gamma_2-\gamma_1} f(\theta) + f(\theta)\theta^{\gamma_2-\gamma_1}$$

$$= (\gamma_2 - \gamma_1)\theta^{\gamma_2-\gamma_1-1} \int_\theta^a f(\eta)d\eta \leq 0,$$

and hence that $g(\theta) \geq 0$ for $\theta \in (0, a)$. Consequently,

$$0 \leq \theta^{\gamma_1-\gamma_2} g(\theta) = \int_\theta^a f(\eta)(1 - (\frac{\theta}{\eta})^{\gamma_1-\gamma_2})d\eta$$

□

6.8 Lemma.

(6.14)
$$v_k(\theta, k) \begin{cases} \leq 0 & \text{when } (\theta, k) \notin \mathcal{F} \\ = 0 & \text{when } (\theta, k) \in \mathcal{F} \end{cases}$$

Proof: Assume first that $(\theta, k) \in \mathcal{F}$. Then $\psi_1(k) \leq \theta \leq \psi_2(k)$, and formula (6.13) becomes

$$v_k(\theta, k) = \frac{2}{(\gamma_1 - \gamma_2)\beta^2} [\theta^{\gamma_2} \int_0^{\psi_1(k)} \frac{(\Pi_k + A\Gamma_k)(\eta, k)}{\eta^{\gamma_2+1}} d\eta +$$

$$+ \theta^{\gamma_1} \int_{\psi_2(k)}^{\infty} \frac{(\Pi_k + A\Gamma_k)(\eta, k)}{\eta^{\gamma_1+1}} d\eta]$$

which is zero by definition of ψ_1 and ψ_2.

We turn to the case where $(\theta, k) \notin \mathcal{F}$; say, $\theta \leq \psi_1(k)$ (the case $\theta \geq \psi_2(k)$ can be treated similarly). Formula (6.13) now becomes

$$v_k(\theta, k) = \frac{2}{(\gamma_1 - \gamma_2)\beta^2} [\theta^{\gamma_2} \int_0^{\theta} \frac{(\Pi_k + A\Gamma_k)(\eta, k)}{\eta^{\gamma_2+1}} d\eta +$$

$$+ \theta^{\gamma_1} \int_\theta^{\psi_1(k)} \frac{(\Pi_k + A\Gamma_k)(\eta, k)}{\eta^{\gamma_1+1}} d\eta + \theta^{\gamma_1} \int_{\psi_2(k)}^{\infty} \frac{(\Pi_k + A\Gamma_k)(\eta, k)}{\eta^{\gamma_1+1}} d\eta] =$$

$$= \frac{2}{(\gamma_1 - \gamma_2)\beta^2} [\theta^{\gamma_1} \int_\theta^{\psi_1(k)} \frac{(\Pi_k + A\Gamma_k)(\eta, k)}{\eta^{\gamma_1+1}} d\eta - \theta^{\gamma_2} \int_\theta^{\psi_1(k)} \frac{(\Pi_k + A\Gamma_k)(\eta, k)}{\eta^{\gamma_2+1}} d\eta],$$

where the last step uses the definition of both ψ_1 and ψ_2. If we rewrite this expression as

$$v_k(\theta,k) = -\frac{2}{(\gamma_1-\gamma_2)\beta^2}\theta^{\gamma_2}\int_{\theta}^{\psi_1(k)}\frac{(\Pi_k + A\Gamma_k)(\eta,k)}{\eta^{\gamma_2+1}}(1-(\frac{\theta}{\eta})^{\gamma_1-\gamma_2})d\eta,$$

and observe that by definition of ψ_1,

$$\int_{\theta}^{\psi_1(k)}\frac{(\Pi_k + A\Gamma_k)(\eta,k)}{\eta^{\gamma_2+1}}d\eta \geq 0,$$

we can use Lemma 6.7 to conclude that $v_k(\theta,k) \leq 0$.

□

We can now put all the pieces together and prove our main theorem.

6.9 Theorem. Given a geometric Brownian motion

$$d\Theta_t = \alpha\Theta_t dt + \beta\Theta_t dB_t \quad , \Theta_0 = \theta,$$

and a discount factor $r > 0$, let A be the operator

$$Af(\theta) = \frac{1}{2}\beta^2\theta^2 f''(\theta) + \alpha\theta f'(\theta) - rf(\theta).$$

For each initial point $(\theta,k) \in \mathbf{R}^2_+$, we seek the non-decreasing process K_t which is $\sigma\{\Theta_s : s \leq t\}$-measurable, satisfies $K_0 \geq k$, and maximizes the total discounted profit

$$J(K) = E^{(\theta,k)}[\int_0^\infty e^{-rt}(\Pi + A\Gamma)(\Theta_t, K_t)dt] + \Gamma(\theta,k).$$

If Conditions 6.2 are satisfied, such an optimal process K exists and can be described as a vertical deflection off a forbidden region $\mathcal{F}$. More precisely, if $\gamma_1 > 0 > \gamma_2$ are the two roots of the characteristic equation $\frac{1}{2}\beta^2\gamma^2 + (\alpha - \frac{\beta^2}{2})\gamma - r = 0$, and

$$\psi_1(k) = \inf\{\theta : \int_0^\theta \frac{(\Pi_k + A\Gamma_k)(\eta,k)}{\eta^{\gamma_2+1}}d\eta > 0\}$$

$$\psi_2(k) = \sup\{\theta : \int_\theta^\infty \frac{(\Pi_k + A\Gamma_k)(\eta,k)}{\eta^{\gamma_1+1}}d\eta > 0\}$$

$$\mathcal{F} = \{(\theta,k) : \psi_1(k) \leq \theta \leq \psi_2(k)\}$$

$$\phi(\theta) = \sup\{k : (\theta,k) \in \mathcal{F}\},$$

then

$$K_t = k \vee \sup\{\phi(\Theta_s) : s \leq t\},$$

and the optimal profit starting from (θ, k) is

$$h(\theta, k) = \Gamma(\theta, k) + \int_{\mathcal{F}_k^c} g(\theta, \eta)(\Pi + A\Gamma)(\eta, k)d\eta$$
$$+ \int_{\mathcal{F}_k} g(\theta, \eta)(\Pi + A\Gamma)(\eta, \phi(\eta))d\eta$$

where $\mathcal{F}_k = \{\theta : (\theta, k) \in \mathcal{F}\}$, and g is the Green function

$$g(\theta, \eta) = \begin{cases} \frac{2}{(\gamma_1 - \gamma_2)\beta^2} \theta^{\gamma_2} \eta^{-\gamma_2 - 1} & \text{if } \eta < \theta \\ \frac{2}{(\gamma_1 - \gamma_2)\beta^2} \theta^{\gamma_1} \eta^{-\gamma_1 - 1} & \text{if } \eta > \theta \end{cases}$$

Proof: We only have to check that $\mathcal{F}$ and v satisfies the assumptions of Proposition 6.1. That v is bounded and continuous with continuous partial derivatives v_θ, v_k and Av, follows from lemmas 6.4 and 6.5 and the explicit formulas (6.8), (6.10), and (6.13). That ϕ is continuous is an immediate consequence of the strict monotonicity of ψ_1 and ψ_2 proved in Lemma 6.5, and the main assumptions (iii) and (iv) of Proposition 6.1 were proved in lemmas 6.4 and 6.8.

□

In many applications to economics it is more convenient to use the following corollary where we are assuming that Condition 5.2 is satisfied.

6.10 Corollary. Assume that the following conditions are satisfied.

a) Π is nonnegative, Γ is non-decreasing in the k variable, and the partial derivatives $\frac{\partial \Pi}{\partial k}$ and $\frac{\partial^{n+m}\Gamma}{\partial \theta^n dk^m}$, $n = 0, 1, 2, m = 0, 1$, exist and are continuous.

b) The function $M(\theta) = \sup\{|(\Pi + A\Gamma)(\theta, k)| : k > 0\}$ belongs to $L^1_{\gamma_1, \gamma_2}$, and so does the function $N^{(k)}(\theta) = \sup\{|(\Pi_k + A\Gamma_k)(\theta, y)| : 0 < y \leq k\}$ for each k.

c) $\lim_{k \to \infty} \psi_1(k) = \infty$ (we allow ψ_1 to become infinite for finite values of k).

d) $(\Pi_k + A\Gamma_k)(\theta, k)$ is nondecreasing in θ and strictly decreasing in k.

For each initial value (θ, k), the optimal (θ, k)-strategy is

$$K_t = k \vee \sup\{\phi(\Theta_s) : s \leq t\},$$

where $\phi = \psi_1^{-1}$, and the maximum profit is

$$h(\theta, k) = \Gamma(\theta, k) + \int_0^{\psi_1(k)} g(\theta, \eta)(\Pi + A\Gamma)(\eta, k)d\eta + \tag{6.15}$$
$$+ \int_{\psi_1(k)}^{\infty} g(\theta, \eta)(\Pi + A\Gamma)(\eta, \phi(\eta))d\eta$$

Proof: It is easy to see that conditions a)-d) above imply Conditions 6.2(i)-(v): the first two conditions are identical; c) obviously implies (iii); (iv) is part of d); and (v) follows from d) and Lemma 6.3a). We also observe that since $(\Pi_k + A\Gamma_k)(\theta, k)$ is nondecreasing in $\theta, \psi_2(k)$ must be infinite whenever $\psi_1(k)$ is finite, and hence ψ_2 plays no rôle in the description of the forbidden region $\mathcal{F}$. The corollary now follows from the theorem.

□

7. Relaxing the monotonicity assumption.

From a practical point of view, there is something quite puzzling about the results above. Since $\psi_1(k)$ and $\psi_2(k)$ are given in terms of the functions $\Pi_k(\cdot, k)$ and $\Gamma_k(\cdot, k)$, the decision of whether to expand or not is determined solely by the marginal profit Π_k and the marginal expansion costs Γ_k at the current production capacity k. We all know that real life is not always that simple - usually, a sound expansion policy should not only depend on the effect of a small first investment, but also take the prospects of later expansions into account.

In this section, we shall show that the phenomenon we just described is initimately connected to our basic monotonicity assumption

$$(\Pi_k + A\Gamma_k)(\theta, k) \quad \text{is strictly decreasing in} \quad k. \tag{7.1}$$

Intuitively, this is not too surprising since what (7.1) basically says is that we make the most profitable expansions first and then turn to less profitable ones. Hence if the first investments are not attractive, later ones will certainly not tempt us.

The situation we shall be working with is the following. Although we shall *not* assume (7.1), we shall for simplicity adopt the other natural monotonicity assumption

$$(\Pi_k + A\Gamma_k)(\theta, k) \quad \text{is nondecreasing in} \quad \theta. \tag{7.2}$$

As we have seen, this means that we can forget about ψ_2 and concentrate on ψ_1. Assume now that if we compute ψ_1 by our usual formula

$$\psi_1(k) = \inf\{\theta : \int_0^{\theta} \frac{(\Pi_k + A\Gamma_k)(\eta, k)}{\eta^{\gamma_2+1}} d\eta > 0\}, \tag{7.3}$$

we get the solid curve in Figure VII.1.

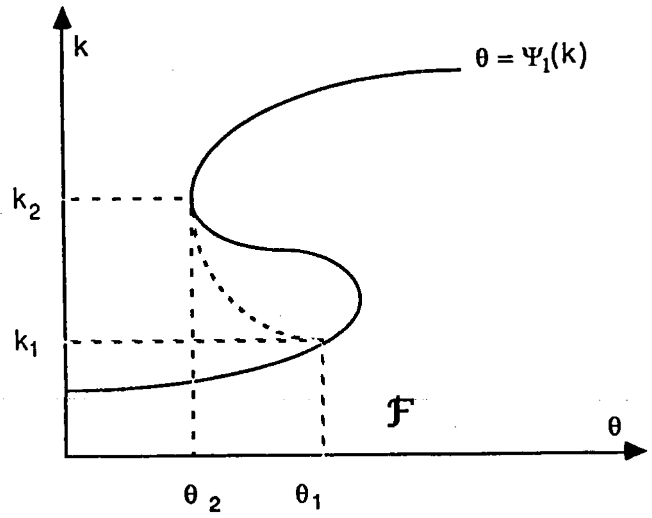

Figure VII.1

It turns out that this curve no longer gives the correct solution to our problem; we shall have to replace part of ψ_1 by the dotted curve connecting the points (θ_1, k_1) and (θ_2, k_2). Let us write $k = \phi(\theta)$ for the inverse function of $\theta = \psi_1(k)$ in the region $[k_2, \infty)$, and let us denote the (still unknown) dotted curve by $\theta = \zeta(k)$.

Since the dotted curve is the boundary of the forbidden region in the interval $[k_1, k_2]$, the function $v = h - \Gamma$ (where h still denotes the optimal profit) is given by

$$v(\theta, k) = \int_0^{\zeta(k)} g(\theta, \eta)(\Pi + A\Gamma)(\eta, k)d\eta +$$

$$+ \int_{\zeta(k)}^{\infty} g(\theta, \eta)(\Pi + A\Gamma)(\eta, \phi(\eta))d\eta$$

when $k_1 \leq k \leq k_2$ (compare formulas (4.11) and (6.15)). Differentiating with respect to k, we get

$$v_k(\theta, k) = \int_0^{\zeta(k)} g(\theta, \eta)(\Pi_k + A\Gamma_k)(\eta, k)d\eta -$$

$$- g(\theta, \zeta(k))\zeta'(k)[(\Pi + A\Gamma)(\zeta(k), y)]_{y=k}^{y=\phi(\theta)}$$

and since $v_k \equiv 0$ along the boundary of the forbidden region, we must have

$$0 = v_k(\zeta(k), k) = \int_0^{\zeta(k)} g(\zeta(k), \eta)(\Pi_k + A\Gamma_k)(\eta, k)d\eta - \tag{7.4}$$

$$- g(\zeta(k), \zeta(k))\zeta'(k)[(\Pi + A\Gamma)(\zeta(k), y)]_{y=k}^{y=\phi(\zeta(k))}$$

Now we see what is new in the present situation; since $\phi(\zeta(k)) \neq k$, the last term in (7.4) doesnt't vanish as before.

It turns out that we can still find a quite explicit expression for ζ. We begin by inserting the expression for the Green function g in (7.4):

$$0 = \frac{2}{(\gamma_1 - \gamma_2)\beta^2}\{\zeta(k)^{\gamma_2} \int_0^{\zeta(k)} \frac{(\Pi_k + A\Gamma_k)(\eta, k)}{\eta^{\gamma_2+1}} d\eta - $$
$$- [\zeta(k)^{\gamma_2} \frac{(\Pi + A\Gamma)(\zeta(k), y)}{\zeta(k)^{\gamma_2+1}} \zeta'(k)]_{y=k}^{y=\phi(\zeta(k))}\},$$

which leads to

$$0 = \int_0^{\zeta(k)} \frac{(\Pi_k + A\Gamma_k)(\eta, k)}{\eta^{\gamma_2+1}} d\eta - \zeta'(k)[\frac{(\Pi + A\Gamma)(\zeta(k), y)}{\zeta(k)^{\gamma_2+1}}]_{y=k}^{y=\phi(\zeta(k))}. \tag{7.5}$$

If we introduce the function

$$V(\theta, k) = \int_0^{\theta} \frac{(\Pi + A\Gamma)(\eta, k)}{\eta^{\gamma_2+1}} d\eta, \tag{7.6}$$

we can write (7.5) as

$$0 = V_k(\zeta(k), k) - \zeta'(k)[V_\theta(\zeta(k), y)]_{y=k}^{y=\phi(\zeta(k))}. \tag{7.7}$$

On the other hand, a simple chain rule calculation gives

$$\frac{d}{dk}\{V(\zeta(k), \phi(\zeta(k)) - V(\zeta(k), k)\} =$$
$$= V_\theta(\zeta(k), \phi(\zeta(k)))\zeta'(k) + V_k(\zeta(k), \phi(\zeta(k)))\phi'(\zeta(k))\zeta'(k) -$$
$$- V_\theta(\zeta(k), k)\zeta'(k) - V_k(\zeta(k), k) =$$
$$= V_k(\zeta(k), \phi(k)))\phi'(\zeta(k))\zeta'(k),$$

where the last step makes use of (7.7). Integrating both sides of this expression from k to k_2, and observing that $V(\zeta(k_2), \phi(\zeta(k_2))) - V(\zeta(k_2), k_2) = 0$, we get

$$- V(\zeta(k), \phi(\zeta(k))) + V(\zeta(k), k) =$$
$$= \int_k^{k_2} V_k(\zeta(y), \phi(\zeta(y)))\phi'(\zeta(y))\zeta'(y) dy,$$

and substituting $z = \phi(\zeta(y))$, we see that

$$(7.8)\qquad \begin{aligned} &- V(\zeta(k), \phi(\zeta(k))) + V(\zeta(k), k) = \\ &= \int_{\phi(\zeta(k))}^{k_2} V_k(\psi_1(z), z)dz \end{aligned}$$

At this stage, it might be useful to throw a glance at Figure VII.2 to see the relationship between the various variables. We now just observe that

$$V_k(\psi_1(z), z) = \int_0^{\psi_1(z)} \frac{(\Pi_k + A\Gamma_k)(\eta, z)}{z^{\gamma_2+1}} dz = 0$$

by definition of ψ_1, and hence (7.8) simply becomes

$$(7.9)\qquad V(\zeta(k), k) = V(\zeta(k), \phi(\zeta(k)))$$

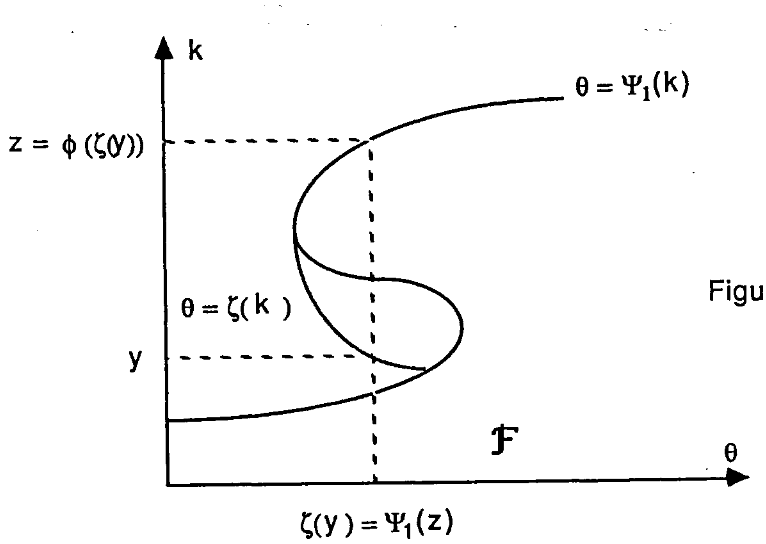

Figure VII.2

If we not only check where v_k is zero, but also take into account that $v_k \le 0$ outside the forbidden region, we get the following description of ζ:

$$(7.10)\qquad \begin{aligned} &\zeta(k) = \inf\{\theta : V(\theta, k) < V(\theta, \phi(\theta))\} = \\ &= \inf\{\theta : \int_0^{\theta} \frac{(\Pi + A\Gamma)(\eta, k)}{\eta^{\gamma_2+1}} d\eta < \int_0^{\theta} \frac{(\Pi + A\Gamma)(\eta, \phi(\theta))}{\eta^{\gamma_2+1}} d\eta\} \end{aligned}$$

There are two natural comments to make at this point; the first is that the definition of $\zeta(k)$ is "non-local" in the sense that it depends on how Π and Γ behaves for capacities larger than k, and the second is that (7.10) is a natural generalization of (7.3) if we replace differentials by finite differences.

To sum up, we can now say that the forbidden region $\mathcal{F}$ is given by

$$\begin{aligned} \mathcal{F} = \{(\theta,k) : \theta \geq \psi_1(k) \quad \text{and} \quad k \leq k_1, \quad \text{or} \quad k \geq k_2\} \cup \\ \cup \{(\theta,k) : \theta \geq \zeta(k), \quad \text{and} \quad k_1 \leq k \leq k_2\}, \end{aligned} \tag{7.11}$$

where k_2 is the "tip of the nose" of the curve $\theta = \psi_1(k)$, and k_1 is given by $\psi_1(k_1) = \zeta(k_1)$. The optimal strategy K_t is constant outside $\mathcal{F}$, is deflected upwards as before whenever (Θ_t, K_t) hits the part of $\mathcal{F}'s$ boundary given by ψ_1, and jumps immediately through $\mathcal{F}$ when it hits $\mathcal{F}'s$ boundary between k_1 and k_2.

The arguments we have given for these results are, of course, just as formal and heuristic as those we presented in Section 4. Under reasonable assumptions, it is possible to verify directly that the conclusions are correct, but as these arguments are quite similar to those in Section 6, we shall not carry them through here. Instead, we shall take a brief look at an example which illustrates the "non-locality" of ζ.

7.1 Example: Let

$$\Pi(\theta,k) = k\theta$$

$$\Gamma(\theta,k) = \begin{cases} 2k - \frac{k^2}{2} & \text{for } k \leq 1 \\ \frac{Dk^2}{2} + (1-D)k + \frac{D+1}{2} & \text{for } k \geq 1, \end{cases}$$

where D is a positive constant. Note that Γ and Γ_k are continuous at $k = 1$. Computing $\psi_1(k)$, we get

$$\psi_1(k) = \begin{cases} -\frac{r(1-\gamma_2)}{\gamma_2}(2-k) & \text{for } k \leq 1 \\ -\frac{r(1-\gamma_2)}{\gamma_2}[Dk + (1-D)] & \text{for } k \geq 1, \end{cases}$$

which shows that ψ_1 makes up the boundary of the forbidden region $\mathcal{F}$ for $k \geq 1$. For $k \leq 1$, we can compute $\zeta(k)$ explicitly from (7.9); long, tedious, but totally elementary calculations yield

$$\zeta(k) = \frac{r(1-\gamma_2)}{\gamma_2(1+\sqrt{1+1/D})}(k - 2 - \sqrt{1+1/D}).$$

The situation is shown graphically in Figure VII.3. Our point is that the shape of the forbidden region depends on D even in the interval $[0,1]$ where Π and Γ are independent of D. Hence local considerations can not determine the optimal strategy; we also need to know what happens above $k = 1$.

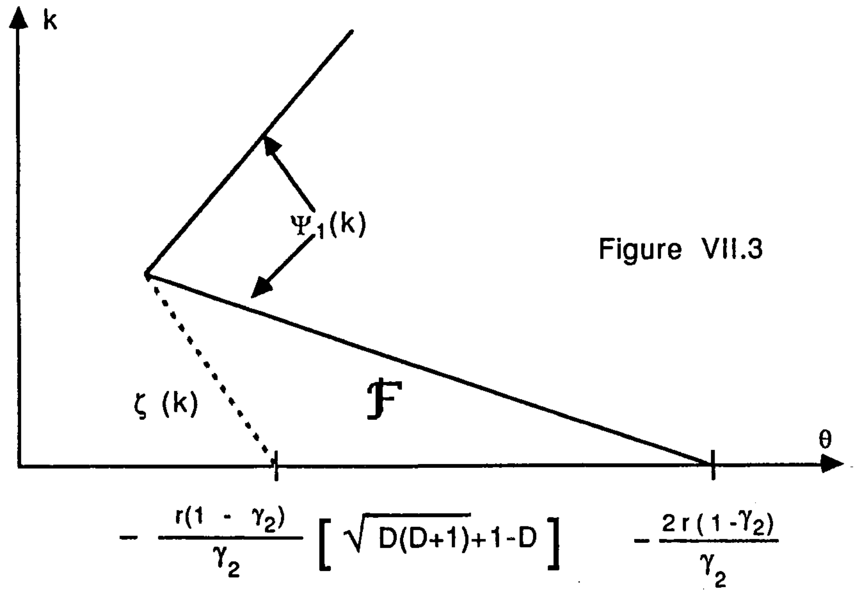

Figure VII.3

8. Discussion.

We have shown that the mathematical problem we set out to study can be solved very completely by quite elementary methods – the optimal strategy is described implicitly by an extremely simple relation which is trivial to solve numerically and which in some cases can even be solved analytically. In order to achieve such completeness and simplicity, we have had to work with a rather naive mathematical model of a complex economic phenomenon, and in this section we shall briefly indicate ways in which the model can be made more realistic. The most obvious modification is to allow more general diffusions than geometric Brownian motion, and as long as the Green function is relatively well behaved, this should not cause serious mathematical problems. A more interesting change is to allow Θ to be multidimensional; this would allow us to study randomly varying discount factors as well as the interplay between independent or nearly independent economic factors. From a mathematical point of view, the multidimensional problem is much harder than the one-dimensional one, and we must probably abandon the idea of finding very explicit solutions. On the other hand, using more abstract machinery it should be possible to obtain good results about the structure of the solution; e.g., that it can be described as a vertical deflection off a forbidden region given by, say, the solution of a variational inequality. Results of this kind would be very attractive from a numerical point of view, and they seem quite plausible at least as long as our basic monotonicity condition (5.1) is in force. If this condition fails, however, the associated variational inequality seems uncomfortably close to some very difficult problems in the theory of Hele Shaw flows (see, e.g., Gustafsson (1985a,b)).

A quite natural objection to the model we have been working with is that the Markovian nature of Θ makes it impossible to take trends into account. To some extent this problem is solved by allowing multidimensional processes as we can then incorporate derivatives into the state (at least as long as we can solve parabolic problems as well as elliptic ones), but a more drastic solution would be to work with general semimartingales instead of diffusions. We have not studied this problem at all, but it would be quite interesting to see what can be said in such a general and abstract setting.

References

V. E. Beneš, L. A. Shepp, and H. S. Witsenhausen (1980): Some solvable stochastic control problems, Stochastics **4**, 39-83.

M. J. Brennan and E. S. Schwartz (1985): Evaluating natural resource investment, J. of Business **58**, 135-157.

N. El Karoui and I. Karatzas (1991): A new approach to the Skorohod problem, and its applications, Stochastics and Stochastics Reports **34**, 57-82.

D. P. Gaver, J. P. Lehoczky, and S. E. Shreve (1984): Optimal consumption for general diffusions with absorbing and reflecting barriers, SIAM J. Control **22**, 55-75.

B. Gustafsson (1985a): Applications of variational inequalities to a moving boundary problem for Hele Shaw flows, SIAM J. Math. Anal. **16**, 279-300.

B. Gustafsson (1985b): Existence of weak backward solutions to a generalized Hele Shaw flow moving boundary problem, Nonlinear Anal. **9**, 203-215.

J. M. Harrison (1985): *Brownian Motion and Stochastic Flow Systems*, Wiley, New York.

I. Karatzas and S. E. Shreve (1984): Connections between optimal stopping and singular stochastic control I. Monotone follower problems. SIAM J. Control **22**, 856-877.

I. Karatzas and S. E. Shreve (1985): Connections between optimal stopping and singular stochastic control II. Reflected follower problems. SIAM J. Control **23**, 433-451.

I. Karatzas and S. E. Shreve (1986): Equivalent models for finite-fuel stochastic control, Stochastics **18**, 245-276.

I. Karatzas and S. E. Shreve (1988): *Brownian Motion and Stochastic Calculus*, Springer-Verlag, New York.

T. Ø. Kobila (1991): Partial investments under uncertainty, in D. Lund and B. Øksendal (eds.) *Stochastic Models and Option Values*, North-Holland, Amsterdam, 167-186.

T. Ø. Kobila (1992a): An application of reflected diffusions to the problem of choosing between hydro and thermal power generation, to appear in to Stoch. Proc. Appl.

T. Ø. Kobila (1992b): Equilibrium in real options, Preprint.

J. P. Lehoczky, and S. E. Shreve (1986): Absolutely continuous and singular stochastic control, Stochastics **17**, 91-109.

R. McDonald and D. R. Siegel (1986): The value of waiting to invest, Quarterly J. of Economics **101**, 707-728.

R. S. Pindyck (1988): Irreversible investments, capacity choice, and the value of the firm, American Economic Review **78**, 969-985.

R. S. Pindyck (1991a): Irreversibility and the explanation of investment behaviour, in D. Lund and B. Øksendal (eds.) *Stochastic Models and Option Values*, North-Holland, Amsterdam,

R. S. Pindyck (1991b): Irreversibility, uncertainty and investment, J. Economic Literature **29**, 1110-1148.

S. E. Shreve (1988): An introduction to singular stochastic control, in W. H. Fleming and P. L. Lions (eds.) *Stochastic Differential Systems, Stochastic Control Theory and Applications. IMA Volumes in Math. & Appl., Vol. 10*, Springer-Verlag, New York, 513-528.

A. V. Skorohod (1961): Stochastic equations for diffusion processes in a bounded domain, Theory Probab. Appl. **6**, 264-274.

M. Sun (1987): Singular control problems in bounded intervals, Stochastics **21**, 303-344.