Discussion Paper No. 159

Optimal Stochastic Development Strategy

by

S. D. Deshmukh* and S. D. Chikte+

July 1975

* Graduate School of Management, Northwestern University, Evanston, Illinois. This research was supported by the National Science Foundation. Grant GS 28287.

+ Department of Electrical Engineering, Polytechnic Institute of Brooklyn, Brooklyn, New York 11201

Abstract

A controlled random walk model is presented for managing a product development project which proceeds stochastically through intermediate states of progress. The problems of optimally undertaking, continuing and stopping the development process are integrated into the single problem of finding an optimal dynamic resource allocation strategy as a function of the worth of the product developed so far. The strategy is shown to be monotone in product worth and characterizes the convex optimal stopping region for not undertaking or discontinuing the project short of completion. Implications for the multiple project comparison and selection problems are indicated.

KEYWORDS: DYNAMIC RESOURCE ALLOCATION, PRODUCT DEVELOPMENT PROJECT, OPTIMAL STRATEGIES

OPTIMAL STOCHASTIC DEVELOPMENT STRATEGY

1. INTRODUCTION

Consider the following general framework encompassing a wide class of problems of dynamic resource allocation under uncertainty. We have an activity to which effort is allocated through time. A state variable (or a vector) represents the progress of that activity at any particular time. Often the activity is pursued with a view towards reaching a certain goal as specified for a given target state. The effect of an allocation of effort on the progress toward the target is stochastic due to inherent and environmental uncertainties. Thus allocating a large amount of effort costs more but also yields a greater progress on average, though perhaps at a decreasing marginal rate. On the other hand, allocating too small an effort, though economical in short run, may lead to negative progress due to stochastic deterioration and obsolescence. Once the allocation of effort is stopped, a reward is collected, whose value depends on the terminal state.

Three interrelated problems facing the manager in charge of such an activity are: (1) given the initial state of the activity, determine whether or not it is worthwhile undertaking its further development, (2) given that we choose to undertake the activity, determine an optimal strategy for allocating effort through time, and (3) given that the allocation is on its way, determine the optimal point in time at which the activity should be discontinued and the reward collected if the goal has been reached.

Some problems where the above framework applies can be briefly described as follows. Improvement in a product's market share to a desired level can be obtained through advertising expenditures, effect of which upon the buyer behavior is uncertain due to lack of brand loyalty, changes in consumer tastes, new competing products, etc. The problem of determining optimal advertising expenditure plan would then correspond to a stochastic version of the optimal control model considered by Sasieni [20], Sethi [21] and others. Similarly, the machine replacement model of Derman [7] could be generalized to provide another example for our framework. The state of a machine can be improved and maintained at a desired level of performance through maintenance expenditures, whose effectiveness is probabilistic in counteracting the stochastic deterioration due to wear out and aging. As a third example, in optimal stopping problems (e.g. see Breiman [3], Chow-Robbins-Siegmund [5]) the usual stop or continue actions may be generalized to include the amount of effort to be expended in searching (for a house, a job or a secretary) to attain a certain specified acceptable quality.

In this paper we shall consider in detail in the above framework the specific problem of optimal resource allocation to a research and development project. In section 2 we describe the model along with the assumptions used to prove subsequent results and show how it generalizes currently available models of the R and D project management. In section 3 we analyze the structure of the optimal allocation strategy and its implications. In the fourth section we indicate the implications of this model to the problem of comparison and resource allocation among different projects. The final section concludes the paper with a summary.

2. THE MODEL

Consider the problem of developing a product of a specified target quality before introducing it into the market. Often, such a development project evolves through a series of identifiable intermediate states of progress towards the target. The state of the project at any point in time may be summarized by the worth of the product quality developed so far, evaluated with respect to the existing environmental factors such as other similar products, competitor's actions, consumer tastes and other market conditions affecting the product demand and hence its price. The progress of the project is reviewed in stages at discrete points in time and its state at the n^{th} stage is denoted by $x_n \in [0,\bar{x}]$, $n=0,1,2...$, where $\bar{x}$ is the worth of the product having the desired target quality. Upon successful compleition of the project the product of worth $\bar{x}$ will be produced and then introduced into the market, yielding a given terminal reward R. For example, R may be the expected value of the net profit stream collected from the completion date onwards, discounted to the completion date. The development process requires allocation of resources at each stage, which may be aggregated into the monetary expenditure $a_n \in [0,B]$, $n=0,1,2,...$, where B denotes the maximum amount available. The terminal reward as well as the interim development expenditures are discounted at rate $\beta \in [0,1)$, i.e. β is the present value of a unit income to be earned in the next period. Inclusion of the discount factor provides, among other things, an incentive to reach the target at an early date.

During the development process the worth of the product changes stochastically from one stage to the next due to two types of uncertainties. Internal uncertainties form an integral feature of any research and development program, which involves activities of uncertain nature, such as generation of new ideas, formation of theories, their experimental verification and technological implementation. External uncertainties, on the other hand, stem from random changes in economic conditions such as introduction of new competing products, shifts in consumer tastes, changes in availability of factors of production, etc. At any stage of the development process, an expenditure of resources counteracts these uncertainties either successfully, thereby increasing the product worth, or unsuccessfully, resulting in its decrease. A higher expenditure in any stage will be assumed to result in a greater chance of success, though at a decreasing rate, thus reflecting decreasing marginal returns to scale. Let $p(a)$ be the probability of success in a stage as a result of an allocation a, where $p(\cdot)$ is a continuous, twice differentiable, concave and nondecreasing function with $p(0) = 0$, implying necessity of effort to achieve success in face of an unfavorable environment.

If success (failure) occurs at any stage, the magnitude of increase (decrease) in the worth of the product may depend upon its current worth x and will be denoted by $U(x)$ $(L(x))$, a continuous twice differentiable function on $(0,\bar{x})$. If the current product worth is too low, there is a large scope for its improvement and hence a success will result in a high increase in the product worth. On the other hand, a product of superior initial quality can not be further improved significantly, due to technological limitations and saturation effects. Similarly, loss due to an unsuccessful allocation is likely to be high in case of a product of high worth, because of losing a portion of a large market share to a competitor, say, while a low worth product has little to lose in spite of failure. Furthermore, it is reasonable to assume that improvement (reduction) in the worth is decreasing (increasing) at a decreasing rate due to enhanced saturation effects. Thus, $U(\cdot)$ ($L(\cdot)$) is assumed to be nonnegative, nonincreasing (nondecreasing) and convex (concave) on $(0,\bar{x})$. Also we assume that $x + U(x)$ and $x - L(x)$ are nondecreasing in x, so that starting a stage with a higher worth yields a higher worth at the end of the stage both in case of a success or a failure in that stage, i.e. $x_1 \geq x_2$ implies $x_1 + U(x_1) \geq x_2 + U(x_2)$ and $x_1 - L(x_1) \geq x_2 - L(x_2)$. Finally we assume that $U(x) + L(x) = [x + U(x)] - [x - L(x)]$ is nondecreasing in x. Thus, the opportunity loss in the product worth due to suffering a deterioration rather than an improvement can not decrease as we get closer to the target, so that the nearer we are to the target the more the success in each stage of the development counts. The components of the controlled random walk model and the intuitively meaningful assumptions described above are stated below and will be needed to obtain subsequent results.

For any $x_n \in [0,\bar{x}]$, $a_n \in [0,B]$, $n = 0,1,2,\ldots$,

(2.1) $$x_{n+1} = \begin{cases} x_n + U(x_n) & \text{with probability } p(a_n) \\ x_n - L(x_n) & \text{with probability } [1 - p(a_n)] \end{cases}$$

where

(2.2) $$0 \leq p(a_n) \leq 1,\ p(0) = 0,\ p'(a_n) \geq 0,\ p''(a_n) \leq 0$$

(2.3) $0 \leq U(x_n),\ 0 \leq L(x_n),\ 0 = U(0) = L(0) = U(\bar{x}) = L(\bar{x})$

(2.4) $0 \leq -U'(x_n) \leq L'(x_n) \leq 1, \qquad x_n \in (0,\bar{x})$

(2.5) $0 \leq U''(x_n),\ 0 \geq L''(x_n) \quad , \qquad x_n \in (0,\bar{x})$

An immediate consequence of these conditions is that, development here refers to improving upon a product of some positive worth and that starting with such a product and spending a positive amount in each stage of the development process, there is a positive probability that a product of the desired worth $\bar{x}$ will be developed in a finite number of stages, at the end of which the process terminates. A simple example where (2.1) - (2.5) are satisfied can be obtained by taking $p(a) = 1 - e^{-\lambda a}$, $U(x) = k(\bar{x} - x)$, and $L(x) = kx$, where $\lambda > 0$, $k \in (0,1)$, $x \in (0,\bar{x})$ and $U(0) = L(\bar{x}) = 0$.

An allocation strategy Π is a sequence of (possibly randomized) decision rules $\{\Pi_n : n = 0,1,2,\ldots\}$ which specifies an expenditure $a_n \in [0,B]$ in each stage n of the development process, as a Borel measurable function of its previous history $h_n = (x_o, a_o, \ldots x_{n-1}, a_{n-1}, x_n)$, $n = 0,1,2,\ldots$ i.e. $\Pi_n(\cdot \mid h_n)$ is a regular conditional probability measure on $[0,B]$. A strategy Π is said to be (non-randomized) <u>Markov</u> if $\Pi = \{\alpha_n : n=0,1,2,\ldots\}$ where $\alpha_n : [0,\bar{x}] \to [0,B]$ is Borel; such a policy specifies an allocation $\alpha_n(x_n)$ at the n^{th} stage if the product worth then is x_n. A (nonrandomized) <u>stationary</u> strategy is given by a single Borel map $\alpha : [0,\bar{x}] \to [0,B]$, so that at any stage of the development process if the current product worth is x, the allocation $\alpha(x)$ is specified by the strategy.

Starting in an initial state x_o and following an allocation strategy Π let

$$(2.6) \qquad N(x_o,\Pi) = \text{Inf}\{n : x_n = \bar{x},\ n=0,1,2,\ldots\}$$

be the random stopping time (possibly infinite) at which the development process successfully terminates, thereby yielding a reward R. The net expected discounted return, starting in x_o and following Π, may then be denoted by

$$(2.7) \qquad W(x_o,\Pi) = E[\beta^{N(x_o,\Pi)} R - \sum_{n=0}^{N(x_o,\Pi)-1} \beta^n a_n | x_n,\Pi].$$

Finally, denote the optimal value function

$$(2.8) \qquad V(x_o) = \sup_{\Pi} W(x_o,\Pi), \qquad x_o \in [0,\bar{x}]$$

From the above definitions and conditions it clearly follows that

$$(2.9) \qquad V(\bar{x}) = R, \qquad V(0) = 0.$$

A strategy Π^* is said to be <u>optimal</u> at x_o if $W(x_o,\Pi^*) = V(x_o)$ and is said to be optimal if it is optimal at x_o for all $x_o \in [0,\bar{x}]$.

The objective of a development manager is to select an optimal strategy. This problem can be analyzed naturally in the framework of Markovian decision processes with discounting (see, for example, Blackwell [2], Strauch [22] and Ross [19]. The optimal value function $V : [0,\bar{x}] \to \mathbb{R}$ is known to be (universally) measurable and satisfies the dynamic programming functional equation (see Strauch [22], Theorem 8.2), which in our case becomes

$$(2.10) \qquad V(x) = \sup_{a\in[0,B]} \{-a + \beta p(a)\, V(x + U(x)) + \beta[1-p(a)]\, V(x - L(x))\},$$

$$x \in [0,\bar{x}].$$

It can be easily verified that the topological and measurability conditions of the selection theorem of Maitra [17] are satisfied in our model, so that, according to his results, there exists an optimal stationary strategy $\alpha^* : [0,\bar{x}] \to [0,B]$, which, when in state x, chooses an action $\alpha^*(x)$ attaining the supremum in the functional equation (2.10). We shall characterize the structure of the optimal value function $V(\cdot)$ and the optimal stationary strategy $\alpha^*(\cdot)$ in the next sction. We close this section by indicating the relationship of our model to other related R and D resource allocation models in the literature.

In the area of dynamic resource allocation in an R and D project, Lucas [16] and Kamien and Schwartz [14] have obtained forms of optimal expenditures as functions of time, using control theory, while Hess [12] and Aldrich and Morton [1] have used the dynamic programming methodology. The corresponding problem of optimally distributing resources among several projects has been considered by Gittins [9,10,11] and Laska, Meisner and Siegel [15]. In these and other related models, the state of the project at any time is classified as either being (completely) successful or not, partial success during its progress being inconsequential. Success of the project may occur instantaneously at any time, the probability of success being a function of the total effort accumulated till that time, thereby reflecting the internal uncertainty regarding the total effort required for successful completion of the project. Such a model may be applicable in the development of, for example, a desired chemical compound or a cure for a disease, or, more frivolously, in solving a jig-saw puzzle.

Our model, on the other hand, may be applicable in the development of, for example, a more efficient energy saving device, a consumer product of improved quality, etc., where the interim progress (partial success) of the project can be meaningfully evaluated in terms of the current product quality developed and its worth under current economic conditions; reaching the target worth then corresponds to the total success. Thus, total success of the project is a cumulative result of a series of interim partial successes and failures. An interim success or a failure depends on both the internal and the external uncertainties and is partially controlled by current resource expenditure, the cumulative effort being of importance only insofar as its past effectiveness as reflected in the present worth. As a consequence, the optimal allocation strategy is a function only of the currently achieved progress, so that we have an adaptive feedback control strategy, rather than an expediture plan which may have to be revised at each stage of the project. Finally, it may be noted that our model may be specialized to the discrete time version of the above mentioned models by taking a binary valued state space, appropriately selecting $U(\cdot)$ and $L(\cdot)$ and modifying $p(\cdot)$ to be a function of the cumulative effort.

3. STRUCTURE OF THE OPTIMAL POLICY

As a finite horizon version of (2.10) consider the sequence of functions $\{V_n(\cdot)\}_{n=o}^{\infty}$ defined recursively by

$$V_o(x) = \begin{cases} 0 & \text{if } 0 \leq x < \bar{x} \\ R & \text{if } x = \bar{x} \end{cases}$$

(3.1)

$$V_{n+1}(x) = \max_{a \in [0,B]} \{-a + \beta\, p(a)\, V_n(x + U(x)) + \beta[1 - p(a)]\, V_n(x - L(x))\}$$

$V_n(x)$ may be interpreted as the maximum expected return from the project starting with the product of worth x if we are allowed to continue for no more than n periods. It follows (e.g. see Ross [19]) that $\lim_{n\to\infty} V_n(x) = V(x)$ uniformly in x. This fact will be used in the proof of the following proposition.

Proposition 1. The optimal value function is nonnegative, bounded, nondecreasing and convex for all $x \in [0,\bar{x}]$, i.e.

(3.2) (i) $R = V(\bar{x}) \geq V(x) \geq V(0) = 0$

(3.3) (ii) $x_1 \geq x_2$ implies $V(x_1) \geq V(x_2)$

(3.4) (iii) $V(\lambda x_1 + (1-\lambda)x_2) \leq \lambda\, V(x_1) + (1-\lambda)\, V(x_2)$ if $\lambda \in [0,1]$, $x_1, x_2 \in [0,x]$.

Proof: We prove monotonicity and convexity of each $V_n(\cdot)$ by induction on n, whereupon noting that these properties are preserved in the limit as $n \to \infty$, (ii) and (iii) follow, from which (i) easily follows upon using (2.9). Clearly from (3.1), $V_o(\cdot)$ is nondecreasing and convex. Suppose $V_n(\cdot)$ is nondecreasing and convex. Since $x + U(x)$ and $x - L(x)$ are nondecreasing and convex in x by assumptions (2.4) and (2.5), it can be checked that $V_n(x + U(x))$ and $V_n(x - L(x))$ are nondecreasing and convex in x. Hence the maximand in (3.1) is nondecreasing and convex in x for each a; hence $V_{n+1}(\cdot)$ is nondecreasing and convex, completing the induction argument. Q.E.D.

Thus, optimally developing the target quality never yields losses, while it is clearly best to have the desired quality product on hand to be marketed. Also starting the development with a product of higher work and continuing optimally will always yield a higher net return, and, moreover, this advantage increases as we get closer to the target. Next, we examine the form of an optimal stationary strategy $\alpha^*(\cdot)$ as a function of the state variable x.

Proposition 2. The optimal stationary strategy $\alpha^*(\cdot)$ is nondecreasing in $x \in (0,\bar{x})$ with $\alpha^*(0) = \alpha^*(\bar{x}) = 0$.

Proof: As remarked in section 2, an optimal allocation $\alpha^*(x)$ maximizes the right hand side of the functional equation (2.10) for each x. A necessary condition for $\alpha^*(x)$ in (0,B) to maximize this right hand side is $-1 + \beta\, p'(\alpha^*(x))\, V(x + U(x)) - \beta\, p'(\alpha^*(x))\, V(x - L(x)) = 0$, i.e.

$$(3.5) \qquad p'(\alpha^*(x)) = \frac{1}{\beta[V(x+U(x))-V(x-L(x))]}$$

Since, by Proposition 1, $V(\cdot)$ is nondecreasing and convex and $U(x) + L(x)$ is nondecreasing by assumption (2.4), it can be shown that the right hand side of (3.5) is nonincreasing in x. Now, using concavity of $p(\cdot)$, it follows that $\alpha(x)$ is nondecreasing in x. The maximand in

(2.10) is concave in a since $V(x + U(x)) \geq V(x - L(x))$ (by Proposition 1), hence (3.5) is also sufficient for an interior maximum. For a boundary solution of $\alpha^*(x) = 0$ to be optimal (3.5) is replaced by

$$p'(0) \leq \frac{1}{\beta[V(x+U(x))-V(x-L(x))]}$$

and if this holds for some x', then, again using Proposition 1 and that $U(x) + L(x)$ is nondecreasing, it also holds for all $x \leq x'$. Similarly for $\alpha^*(x) = B$ to be optimal, we have

$$p'(B) \geq \frac{1}{\beta[V(x+U(x))-V(x-L(x))]}$$

and if this holds for some x'', then it will also hold for all $x \geq x''$.

Define $x_* = \text{Sup}\,\{x' : \alpha^*(x') = 0\}$ and

$x^* = \text{Inf}\,\{x'' : \alpha(x'') = B\}$

Then $\alpha^*(x)$ is 0 for all $x \in [0,x_*]$ and B for all $x \in [x^*,\bar{x}]$ and is nondecreasing for intermediate values of x. Finally, upon reaching the worth 0 or the target $\bar{x}$ the process is absorbed, by assumptions (2.3), and yields 0 or R regardless of a, and hence $\alpha^*(\bar{x}) = \alpha^*(0) = 0$.

Q.E.D.

Thus, the closer we get to the target, the more optimistic our outlook becomes and the more we are encouraged to strive harder. On the other hand, if at some stage we fall too far away from the goal (i.e. below x_*), the terminal reward is not lucrative enough in relation to the expected effort required to attain it, hence it is not optimal to continue the development, regardless of the effort expended in the past. Thus, at any stage of the development we may stop either because we have completed the project (i.e. $x = \bar{x}$)

or because it is not worthwhile completing it (i.e. $x \leq x_*$). As a special case, if the initial worth of the product is $x_o \leq x_*$, it is optimal not to undertake the development at all, while if $x_o = \bar{x}$, then the development effort is unnecessary. Allowing for this discontinuation possibility is in contrast with the optimal "pursuit" and the "first passage" problems considered by Eaton and Zadeh [8] and Derman [7], where reaching the target is compulsory and is ensured, in the case of finite state and action spaces, by imposing conditions on the transition probabilities and interim costs. Also these models concentrate on the problems of existence of an optimal strategy and its computation, rather than on its form under specific structural assumptions.

A useful consequence of the results obtained so far is that, in search for an optimal development strategy under the stated assumptions one may restrict attention to the class $\mathcal{C}$ of stationary strategies which are monotone nondecreasing in the state variable. Theoretically, for any $\alpha \in \mathcal{C}$ we may, in principle, compute the net expected discounted return $V_\alpha(x)$ from following such a strategy by solving the system of equations

$$V_\alpha(x) = - \alpha(x) + \beta\, p(\alpha(x))\, V_\alpha(x + U(x)) + \beta[1-p(\alpha(x))]\, V_\alpha(x-L(x)), \quad x \in [0,x]$$

and then we may select $\alpha(\cdot)$ maximizing $V_\alpha(\cdot)$. If the state and action spaces are finite, the policy improvement routine (Howard [13]) can be used to yield the optimal strategy. Otherwise, from the practical point of view, one may select a set $\mathcal{C}'$ of strategies in $\mathcal{C}$ and, as in Marschak-Yahav [18] implement them simultaneously to obtain better estimates of $V_\alpha(\cdot)$, $\alpha \in \mathcal{C}'$, in each period. As the development proceeds, strategies with low estimates are sequentially abandoned, finally ending up with the most attractive strategy in $\mathcal{C}'$ with an estimate of $V(\cdot)$. The selection of the initial set $\mathcal{C}'$ in the Marschak-Yahav approach is thus simplified by knowing the form of the optimal strategy.

Monotonicity and nonnegativity of $\alpha^*(\cdot)$ and $V(\cdot)$ enable us to write

$$(3.6) \qquad [0,x_*] = \{x \in X : \alpha^*(x) = 0\}$$

$$(3.7) \qquad [0,x_{**}] = \{x \in X : V(x) = 0\}$$

The following proposition relates these two sets.

<u>Proposition 3</u>. $x_* \leq x_{**}$, with equality holding if $p''(0) < 0$.

<u>Proof</u>: Suppose $x \in [0,x_*]$, i.e. $\alpha^*(x) = 0$. Since $\alpha^*(\cdot)$ must satisfy the functional equation (2.10) and since $p(0) = 0$ we have $V(x) = \beta\, V(x - L(x))$ whenever $\alpha^*(x) = 0$. Since $V(\cdot)$ is nonnegative, nondecreasing and $L(x) \geq 0$ and $\beta < 1$, this equality is satisfied only if $V(x) = V(x - L(x)) = 0$, hence $x \leq x_{**}$. Next, suppose $V(x) = 0$ (and hence $V(x - L(x)) = 0$). The optimality equation (2.10) now becomes

$$(3.8) \qquad 0 = \max_{a \in [0,B]} \{-a + \beta\, p(a)\, V(x + U(x))\}.$$

Since the maximand is 0 at $a = 0$ and is concave in a, it is necessary and sufficient that it be nonincreasing in a at a = 0 in order that (3.8) be satisfied, i.e. $V(x + U(x)) \leq \frac{1}{\beta p'(0)}$. In that case the set of a's attaining the maximum in (3.8) constitute an interval $[0,\bar{a}]$ with $\alpha^*(x) \in[0,\bar{a}]$. This interval reduces to a singleton set $\{0\}$ if $p''(0) < 0$ (so that the maximand in (3.8) is strictly decreasing) in which case $\alpha^*(x) = 0$ and thus $x_* = x_{**}$. Q.E.D.

Thus, if we start in (or reach) a state in $[x_*,x_{**}]$ we may pursue the project, though the expected return will exactly balance the expected cost of attaining it. In this region, if we are a risk averter we may choose the sure thing $\alpha \equiv 0$ (i.e. abandon the project) yielding the same expected return of zero. On the other hand, if we are a risk lover, we may follow the non-null policy yielding zero net expected return. If $p''(0) < 0$ and the current quality is x_{**} or worse, the only optimal strategy is to discontinue. From now on we will assume this to be the case, so that, for a given project, the optimal stopping region becomes $[0,x_*] \cup \{\bar{x}\}$.

4. COMPARISON OF DEVELOPMENT PROJECTS

A development project $\mathcal{P}$ is characterized by the initial worth of the product x_o, the desired worth $\bar{x}$, the value R of completing the project, the inherent and environmental uncertainties in the development process as controlled by effort and summarized in $p(\cdot)$ and the maximum effort B available to be expended. Thus, we may denote the project by $\mathcal{P} = (x_o,\bar{x},R,p(\cdot),B)$, while its value $V_{\mathcal{P}}(\cdot)$ is the net expected discounted return function as a result of conducting this project according to the optimal stationary strategy $\alpha^*_{\mathcal{P}}(\cdot)$ as in section 3.

As in Proposition 1, $V_{\mathcal{P}}(x_o)$ is nondecreasing convex in x_o, so that, other things being equal, a project with a better initial quality product is more profitable in the long run as well.

Similarly, it easily follows from the definitions

$$(4.1) \qquad V_{\mathcal{P}}(x) = \underset{\Pi}{\text{Sup}}\ E\{R\cdot\beta^{N(x,\Pi,\bar{x})} - \sum_{n=1}^{N(x,\Pi,\bar{x})-1} \beta^n a_n \,|\, x,\Pi\}$$

$$(4.2) \qquad N(x,\Pi,\bar{x}) = \text{Inf}\{n : x_n = \bar{x} \,|\, x,\Pi\}$$

that, for all x, $V_{\mathcal{P}}(x)$ is nondecreasing and convex in the terminal return R for a given target state $\bar{x}$ so that it is better to undertake a project with a higher terminal reward. Analogously, for a given terminal reward R, we prefer a project with lower required target quality. Similarly, increasing the budget B enlarges the class of admissible strategies over which the supremium is taken in (4.1), so that $V_{\mathcal{P}}(x)$ is nondecreasing in B. Finally, if the effectiveness of effort p(a) is higher for all a, the corresponding stopping time in (4.2) will be lower (a.s.) under all policies and hence $V_{\mathcal{P}}(\cdot)$ will increase. These intuitively appealing observations regarding the value of a project are summarized as

Proposition 4. The value $V_{\mathcal{P}}(x_o)$ of a project $\mathcal{P}$ is convex and nondecreasing in x_o and R, nondecreasing in $p(\cdot)$ and B and nonincreasing in $\bar{x}$.

Combining Propositions 1 and 4 we get

Proposition 5. Let $x_*(\mathcal{P}) = \text{Sup}\{x : V_{\mathcal{P}}(x) = \alpha^*_{\mathcal{P}}(x) = 0\}$. Then $x_*(\mathcal{P})$ and hence the optimal stopping region $[0,x_*(\mathcal{P})] \cup \{\bar{x}\}$ is nonincreasing in R, $p(\cdot)$ and B, and nondecreasing in $\bar{x}$.

Thus a higher terminal reward or a lower desired quality provide a greater inducement for undertaking the development of a lower initial quality product.

The above remarks can serve as useful guidelines to the solution of the general problem of R and D project evaluation and selection treated extensively in the literature (see, for example, the survey paper by Cetron, Martino and Roepcke [4]). Thus, if a total budget B is to be distributed among n activities in each period, the allocation B_i i=1,2,...,n with $\sum_{i=1}^{n} B_i \leq B$ yields projects $\theta_1,\ldots,\theta_n$, where $\theta_i = (x_{o_i}, \bar{x}_i, R_i, p_i(\cdot), B_i)$, with values $V_{\theta_i}(x_{o_i})$, i=1,2,...,n. Conceptually this problem of selecting an optimal subset of projects then becomes

$$\text{Max} \sum_{i=1}^{n} V_{\theta_i}(x_{o_i})$$

$$\text{s.t.} \quad \sum_{i=1}^{n} B_i \leq B$$

$$B_i \geq 0 \qquad i=1,\ldots,n$$

which may be approximately solved if the estimates of $V_{\theta_i}(x_{o_i})$ are obtained as indicated in section 3.

Finally, we indicate the relevance of our model in the context of optimal dynamic resource allocation among multiple projects as considered by Gittins [9,10,11], Laska, Meisner and Siegel [15] and others. In our version of the model the states of progress of n (possibly interrelated) projects proceeding simultaneously may be described by an n-vector, the terminal reward being a function of the n-vector of the desired target levels. The optimal stationary strategy would then specify in each period the distribution of a total budget among these n activities as a function of the current progress state vector. This problem will be treated elsewhere.

5. SUMMARY

Thus, we have analyzed the problem of optimal dynamic resource allocation in management of a research and development project within the framework of Markov decision processes. We have characterized an optimal stochastic feedback control strategy which is time invariant and specifies the resource allocation in any stage of the project as a function of the current state of its progress. If the current product worth is too low, it prescribes terminating the project short of its successful completion. If the overall progress of the project so far has been satisfactory, resulting in a reasonably good product, then it is optimal to continue its further development; the better the current product the greater should be the effort to develop it further until its culmination in a successful conclusion. Thus, the problems of optimal starting, continuing and stopping of a development project, as indicated in section 1, have been integrated into a single problem of finding an optimal strategy.

References

[1] Aldrich, C. and Morton, T. E., "Optimal Funding Paths for a Class of Risky R&D Projects," Management Science, Vol. 21, No. 5, pp. 491-500 (1975).

[2] Blackwell, D., "Discounted Dynamic Programming," Annals of Mathematical Statistics, Vol. 36, pp. 226-235, (1965).

[3] Breiman, L., "Stopping Rule Problems," Ch. 10 in Applied Combinatorial Mathematics, E. F. Beckenbach (Editor), Wiley (1964).

[4] Cetron, M. J., Martino, J., and Roepeke, L., "The Selection of R&D Program Content: Survey of Quantitative Methods," IEEE Trans. Engineering Management EM-14, pp. 4-13 (1967).

[5] Chow, Y.S., Robbins, H., Siegmund, D., Great Expectations: The Theory of Optimal Stopping," Houghton Mifflin, 1971.

[6] Dean, B. V., (Ed.), Operations Research in Research and Development, Wiley (1963).

[7] Derman, C., Finite State Markovian Decision Processes, Academic Press (1970).

[8] Eaton, J. H., and Zadeh, L. A., "Optimal Pursuit Strategies in Discrete State Probabilistic Systems," Trans. ASME Ser. D., J. Basic Engineering, Vol. 84, pp. 23-29 (1962).

[9] Gittins, J. C., "Optimal Resource Allocation in Chemical Research," Advances in Applied Probability, Vol. 1, pp. 238-270 (1969).

[10] Gittins, J. C., "Some Problems of Stochastic Resource Allocation," Journal of Applied Probability, Vol. 9, pp. 360-369 (1972).

[11] Gittins, J. C., "Resource Allocation in Speculative Chemical Research," Journal of Applied Probability, Vol. 11, pp. 255-265 (1974).

[12] Hess, S. W., "A Dynamic Programming Approach to R&D Budgeting and Project Selection," IRE Transactions on Engineering Management, Vol. EM 9, pp. 170-178 (1962).

[13] Howard, R. A., Dynamic Programming and Markov Processes, Technology Press and Wiley, (1960).

[14] Kamien, M. I., and Schwartz, N. L., "Expenditure Patterns for Risky R&D Projects," Journal of Applied Probability, Vol. 8, pp. 60-73 (1971).

[15] Laska, E., Meinser, M., and Seigel, C., "Contributions to the Theory of Optimal Resource Allocation," Journal of Applied Probability, Vol. 9, pp. 337-359.

[16] Lucas, R. E., "Optimal Management of an R&D Project," Management Science, Vol. 17, pp. A679-A697, (1971).

[17] Maitra, A., "Discounted Dynamic Programming on Compact Metric Spaces," Sankhyā, series A, Vol. 30, part 2, pp. 211-216 (1968).

[18] Marschak, T. A., and Yahav, J. A., "The Sequential Selection of Approaches to a Task," Management Science, Vol. 12, pp. 627-647 (1966).

[19] Ross, S. M., Applied Probability Models with Optimization Applications, Holden Day, (1970).

[20] Sasieni, M., "Optimal Advertising Expenditure," Management Science, Vol. 18, pp. P64-P72 (1971).

[21] Sethi, S. P., "Optimal Control of the Vidale-Wolfe Advertising Model," Operations Research, Vol. 21, pp. 998-1013 (1973).

[22] Strauch, R. E., "Negative Dynamic Programming," Annals of Mathematical Statistics, Vol. 37, pp. 871-890 (1966).