Optimal Allocation of Instructional Effort to Interrelated Learning Strands[1]

VERNE G. CHANT AND RICHARD C. ATKINSON

Stanford University, Stanford, California 94305

The problem of allocating instructional effort to two interrelated blocks of learning material is studied. In many learning environments, the amount of material that has been mastered in one area of study affects the learning rate in another distinct but related area—for example, the curriculum subjects of mathematics and engineering. A model is developed that describes this phenomenon, and the Pontryagin Maximum Principle of control theory is applied to determine optimal instructional policies based on the model. The nature of this optimal solution is to allocate instructional effort so that the learner follows a maximal average learning rate "turnpike" path until near the end of the study period and then concentrates on only one strand. This strategy, when applied to a more realistic stochastic model, defines a closed-loop feedback controller that determines daily instructional allocation based on the best current estimate of how much the student has learned. This estimate is calculated by a multistage linear filter based on the Kalman filtering technique.

INTRODUCTION

The framework of control theory provides a useful structure for studying a number of problems in the theory of instruction. This framework is centered around a dynamic description of the essential characteristics of the problem, called state variables. This dynamic model, whether deterministic or probabilistic, describes the possible future results over the time period of interest for each decision or control policy. The decision variables usually are constrained to belong to some admissible or feasible set. Once the dynamic behavior of the system is known, the problem is to choose a control policy that, in some sense, is 'best'. In the framework of control theory an objective functional is formulated to determine whether one control policy is better than another one. This functional assigns a scalar value measuring the "goodness' of the control policy which

1

explicitly evaluates the trade-off between benefits and costs. In problems where this trade-off cannot be formulated, the objective can be simplified to maximizing benefits for a certain cost level or to minimizing cost for a certain minimum level of benefits.

In the theory of instruction the problem is to determine an instructional policy for a certain planning period that maximizes student achievement while maintaining or minimizing instructional costs. This problem fits naturally into the control theory framework provided the dynamic behavior of the interaction between student and teacher can be described. In certain highly simplified learning environments investigated by experimental psychologists the dynamic behavior of the student is described reasonably well by existing learning models. In other learning situations, especially with teacher–learner interaction, new models must be developed. Application of control theory to these problems shows promise of useful results.

A number of researchers have been interested in the application of optimization techniques to models of learning and instruction. Smallwood (1962) structured the interaction of learner and machine in a programmed instruction environment. Karush and Dear (1966) and Groen and Atkinson (1966) have applied optimization methods to mathematical learning models. Atkinson and Paulson (1972) outline the basic steps of rational decision making as applied to the theory of instruction and develop several applications. In the present paper, a model of a different kind of learning environment is considered and analyzed in order to derive optimal instructional policies.

The Stanford Reading Program (Atkinson and Fletcher, 1972) is a computer-assisted instructional program for teaching reading to students in grades 1–3. The program is organized around two basic curriculum strands, one devoted to instruction in sight-word identification and the other to instruction in phonics. Each strand is an ordered sequence of exercises that may be regarded, for our purposes, as unending. There are other components in the reading program, but each day the student spends some time on both the phonics and the sightword strands. It has been observed that the instantaneous learning rate on one strand depends on how far along the student is on the other strand. This phenomenon, of course, is not unique to the Stanford Reading Program or even to the general area of learning to read. Clearly, in most areas of learning that involve the simultaneous study of more than one curriculum, performance on one will be interrelated with the achievement level on the other. It is this interaction that is modeled and analyzed for optimal instructional policies in the present paper.

The basic problem is as follows: It is assumed that the learning rate for each of two strands depends on the difference between the achievement levels on the two strands. The dependence is such that if the student has progressed farther in the first strand than in the second, then the learning rate decreases for the first and increases for the second. This functional relationship is assumed to be known. The objective is to maximize the level of achievement for some weighted average of the two strands over a given number of days of instruction. The total daily period of instruction for the two

strands is fixed, but the relative amount of time devoted to each strand each day is a *control variable.*

The next section of the paper describes the basic features of the problem and the various formulations that will be analyzed in later sections. The first of two principal results is developed in the third section. Here the Maximum Principle of optimal control theory is applied which yields optimal instructional policies for one of the problem formulations. This optimal solution is characterized by a "turnpike path" in the state space that is part of any optimal trajectory. The turnpike path is also shown to be a maximal average learning rate path. The second result is developed in the fourth section where a solution is given for the case that includes a probabilistic description of the system. This solution makes use of the earlier result which applies to the deterministic case. The last section includes some comments and conclusions.

PROBLEM FORMULATION

Learning Rate Characteristics

The interdependence of the learning phenomenon of the two curriculum strands can be characterized as follows. Let the achievement levels of a student on the two strands be represented by x_1 and x_2. The interdependence between the strands is such that the instantaneous learning rate on a strand is a function of the difference, $x_1 - x_2$, in achievement levels. Typical learning rate characteristics are shown in Fig. 1. It is clear that the characteristics shown are *stable* in the sense that the larger the difference $x_1 - x_2$ is (i.e., the farther x_1 moves ahead of x_2) the smaller the learning rate on strand one and the larger the learning rate on strand two. Thus, the achievement levels tend to change in order to oppose the change in $x_1 - x_2$.
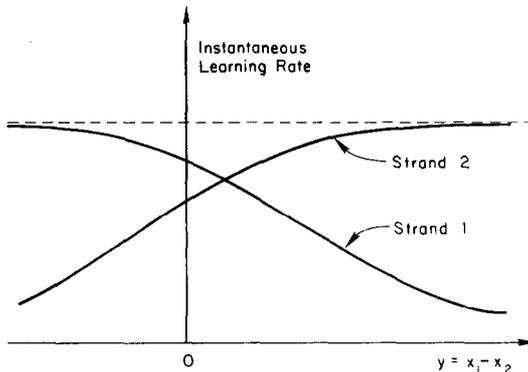

FIG. 1. Learning rate characteristics.

The long term, or steady state learning rates can be determined as follows. If the fraction of time allocated each day to each strand were constant over a long period, then the learning rates would approach the steady state values given by

$$\dot{x}_1 t_1 - \dot{x}_2 t_2 = 0, \tag{1}$$

where t_i is the time allocated to strand i per day and $\dot{x}_i = dx_i/dt$ is the learning rate on strand i. Equation 1 requires that the progress on each strand (learning rate multiplied by time) is the same each day, and hence the difference $x_1 - x_2$ does not change. Expressed differently, (1) becomes

$$\dot{x}_1/\dot{x}_2 = t_2/t_1 \,,$$

which shows that the ratio of the steady state learning rates is the reciprocal of the ratio of the times spent on each strand.

Discrete Time Model

The problem of interest is the determination of the optimal allocation of teaching effort (computer time in the Reading Program application) over an extended period of time. In order to do this, a dynamic model is required that describes how the learner progresses, given certain teaching inputs. If a time unit of one day (or one session at the computer terminal) is used, then the basic model is as follows: The achievement level at the end of the day is the achievement level at the beginning of the day plus the amount learned which is the product of learning rate and time. That is,

$$
\begin{aligned}
x_1(k + 1) &= x_1(k) + u(k)\, \phi_1(u(k), x_1(k) - x_2(k)), \\
x_2(k + 1) &= x_2(k) + [1 - u(k)]\, \phi_2(u(k), x_1(k) - x_2(k)),
\end{aligned}
\tag{2}
$$

where

$x_i(k)$ is the achievement level at the end of day k on strand i

$u(k)$ is the amount of time allocated to strand one (normalized so that $0 \leqslant u(k) \leqslant 1$)

$\phi_i(u(k), x_1(k) - x_2(k))$ is the average learning rate for day k on strand i which depends on the time allocation $u(k)$ and the difference, $x_1(k) - x_2(k)$, in achievement levels.

The functions ϕ_i are determined by the learning rate characteristic functions, denoted f_i, as illustrated in Fig. 1. As an approximation, the functions ϕ_i are the same as the f_i if rate variation during the session is ignored. If so, ϕ_i does not depend on $u(k)$, and we have

$$\phi_i(u(k), x_1(k) - x_2(k)) = f_i(x_1(k) - x_2(k)).$$

Although model (2) is quite general, a more realistic description of the learning process would include stochastic parameters. Assuming that the random behavior of this process may be represented by additive gaussian noise, model (2) becomes

$$x_1(k + 1) = x_1(k) + u(k)\,\phi_1(u(k), x_1(k) - x_2(k)) + n_1(k),$$

$$x_2(k + 1) = x_2(k) + [1 - u(k)]\,\phi_2(u(k), x_1(k) - x_2(k)) + n_2(k),$$

(3)

where the $n_i(k)$ represent noise. Our objective is to determine the optimal teaching effort allocation, $u(k)$, for all k. This allocation will depend on how much the student has learned up to the current moment. Thus knowledge of the state vector $x(k) = (x_1(k), x_2(k))^T$ during the learning process is essential. In most instructional situations, tests are used as a measure of the student's achievement. However, the test results are imperfect so that another noise component will enter the model. If we denote the measurements (test results) by $z(k)$, then the measurement model is

$$z(k) = x(k) + w(k),$$

(4)

where $z(k) = (z_1(k), z_2(k))^T$ represents measurements of the achievement levels on the two strands and $w(k) = (w_1(k), w_2(k))^T$ represents gaussian measurement noise.

The instructional objective is to maximize the final achievement levels of the student in the two strands. Let the relative importance of the two strands be represented by the nonnegative constants c_1 and c_2. Then the objective is to maximize

$$c_1 x_1(T) + c_2 x_2(T),$$

(5)

where $x_i(T)$ represents the final level of achievement on strand i after the fixed number of days T. In the case with random components included, the objective function is the expected value of (5).

The problem is to determine the teaching input variable or control variable $u(k)$ for $k = 0, 1, 2,..., T - 1$ such that

$$0 \leqslant u(k) \leqslant 1$$

(6)

that maximizes (5) subject to the system dynamics (3) with measurements (4).

Continuous Time Model

The problem as posed in the previous section would be impossible to solve analytically although numerical methods using simulation could be used for a particular case. In order to understand this process, we chose to simplify the model so that an analytic solution would be possible while keeping the essential characteristics of the problem. Later, these simplifications will be relaxed.

The first simplification is to look only at the deterministic problem, that is, the randomness introduced by the state and measurement noises will be ignored. The second simplification is to consider a continuous time model rather than the discrete time model formulated in the previous section. Although this will simplify the analysis, there is no loss in realism. The continuous time model will allow us to drop the dependence of the characteristic function ϕ on the control variable u and still have the model account for learning rate changes during the day.

The transition to the continuous model from the discrete model (2) can be thought of as dividing the study period into finer and finer equal partitions. In the limit, the continuous time control variable, $u(t)$, represents the relative amount of time spent on strand one. The continuous time model now expressed in differential equation form is

$$\dot{x}_1(t) = u(t) f_1(x_1(t) - x_2(t)),$$

$$\dot{x}_2(t) = [1 - u(t)] f_2(x_1(t) - x_2(t)),$$

(7)

with initial values $x_1(0)$, $x_2(0)$ given. The learning rate characteristic functions f_1 and f_2 are given by Fig. 1. The problem now is to maximize

$$c_1 x_1(T) + c_2 x_2(T),$$

(8)

where T is given and c_1 and c_2 are given nonnegative weights. This maximization is with respect to u subject to the constraint

$$0 \leqslant u(t) \leqslant 1$$

(9)

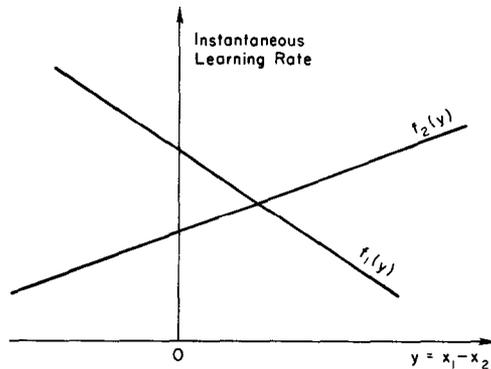for all t such that $0 \leqslant t \leqslant T$.

FIG. 2.   Linearized learning rate characteristics.

The third simplification is to consider only straight-line learning rate characteristics as shown in Fig. 2. Although this assumption may seem unrealistic we will see that the results from the analysis of the model may be reasonably transferred back to the original nonlinear model. Thus the form of the f_i functions is

$$f_1(x_1 - x_2) = b_1 - a_1(x_1 - x_2),$$
$$f_2(x_1 - x_2) = b_2 + a_2(x_1 - x_2),$$

(10)

where a_1, a_2, b_1, b_2 are positive constants and x_1 and x_2 are restricted to values such that these functions are positive.

With the continuous model expressed in this form this model is equivalent to the discrete model (3) if the following interpretations are given to the ϕ_i functions. Suppose that the control variable $u(t)$ is constant over day k at the value $u(k)$, i.e.,

$$u(t) = u(k), \qquad k \leqslant t < k + 1.$$

Then (7) and (10) show that

$$\dot{x}(t) = F(k)x(t) + G(k), \qquad k \leqslant t < k + 1,$$

(11)

where the system matrix $F(k)$ is given by

$$F(k) = \begin{bmatrix} -a_1 u(k) & a_1 u(k) \\ [1 - u(k)] a_2 & -[1 - u(k)] a_2 \end{bmatrix},$$

(12)

and

$$G(k) = \begin{bmatrix} u(k) b_1 \\ [1 - u(k)] b_2 \end{bmatrix}.$$

(13)

Integrating (11) from $t = k$ to $t = k + 1$ gives

$$x(k + 1) = e^{F(k)}x(k) + \int_k^{k+1} e^{F(k)(k+1-t)}G(k)\, dt.$$

(14)

If we define the discrete model transition 2×2 matrix $\phi(k)$ and vector $B(k)$ as

$$\phi(k) = e^{F(k)},$$

(15)

$$B(k) = \int_k^{k+1} e^{F(k)(k+1-t)}\, dt\, G(k),$$

(16)

then (14) becomes the apparent linear form

$$x(k + 1) = \phi(k)x(k) + B(k),$$

(17)

where $u(k)$ is included on $\phi(k)$.

ANALYSIS OF CONTINUOUS MODEL

Solution of Simplified Model Using Maximum Principle

In this section the Pontryagin Maximum Principle (Pontryagin *et al.*, 1962; Luenberger, 1969) will be used to obtain an analytic solution to the problem as stated in Eqs. 7–10. It will be shown that the optimal control input is initially to teach one strand only until its achievement level reaches a certain value. Then both strands should be taught at a rate to maintain the difference between the strands constant. Near the end of the study period T again only one strand should be taught depending on the ratio of the objective function weighting constants c_1 and c_2. The form of this solution is called a "turnpike"[2].

Necessary conditions for u to be the optimal solution for the maximization problem Eqs. 7–10 are given by the Maximum Principle. To begin this analysis, we examine the hamiltonian which is given by

$$
\begin{aligned}
H &= \lambda_1 \dot{x}_1 + \lambda_2 \dot{x}_2 \\
&= \lambda_1 u f_1(y) + \lambda_2 (1 - u) f_2(y) \\
&= \lambda_1 u (b_1 - a_1 y) + \lambda_2 (1 - u)(b_2 + a_2 y) \\
&= u[\lambda_1(b_1 - a_1 y) - \lambda_2(b_2 + a_2 y)] + \lambda_2(b_2 + a_2 y),
\end{aligned}
\tag{18}
$$

where y is defined as the difference $x_1 - x_2$ and where λ_1 and λ_2 are defined by the adjoint differential equations

$$
-\dot{\lambda} = \begin{bmatrix} -\dot{\lambda}_1 \\ -\dot{\lambda}_2 \end{bmatrix} = \begin{bmatrix} -a_1 u & a_2(1 - u) \\ a_1 u & -a_2(1 - u) \end{bmatrix} \cdot \lambda
\tag{19}
$$

with terminal condition $\lambda(T) = c$, the objective function coefficient vector. Necessary conditions for u to be optimal are that $u(t)$ maximize the hamiltonian H, for $0 \leqslant t \leqslant T$, subject to the constraint (9). Since H is linear in $u(t)$, the maximum will occur with $u(t)$ on the boundary of the constraint set, that is, either $u(t) = 0$ or $u(t) = 1$, unless the quantity multiplying $u(t)$, i.e.,

$$
\lambda_1(b_1 - a_1 y) - \lambda_2(b_2 + a_2 y)
\tag{20}
$$

is zero.

---

[2] The term "turnpike" has been used by Samuelson (1960) and other economists to describe a similar result in economic growth theory. The term comes from an analogy to a highway system where minimum time paths between towns often are not minimum distance paths but instead have the motorist detour to a turnpike for most of the journey and then leave the turnpike for his final destination. The turnpike path here differs from the maximal growth rate turnpike paths in economics in that the difference between two variables is constant in this result whereas ratios are constant in the economic growth results.

From (19) we have that

$$\dot{\lambda}_1 + \dot{\lambda}_2 = 0$$

and since $\lambda_1(T) = c_1$, $\lambda_2(T) = c_2$, we have

$$\lambda_2(t) = c_1 + c_2 - \lambda_1(t); \tag{21}$$

so expression (20), which will be denoted e, becomes

$$e = \lambda_1(b_1 - a_1 y) + \lambda_1(b_2 + a_2 y) - (c_1 + c_2)(b_2 + a_2 y). \tag{22}$$

And so

$$\begin{aligned}
\dot{y} &= \dot{x}_1 - \dot{x}_2 \\
&= ((a_2 - a_1)\, u - a_2)\, y + u(b_1 + b_2) - b_2\,,
\end{aligned} \tag{23}$$

where $y(0) = x_1(0) - x_2(0)$.

Now suppose that the optimal solution has an interval where the control variable $u(t)$ is not on the boundary (i.e., not zero or one).[3] In this interval e must be identically zero and so must the derivative of e. Looking at the derivative of e, we see that

$$\dot{e} = \dot{\lambda}_1(b_1 + b_2 + (a_2 - a_1)y) + \lambda_1(a_2 - a_1)\,\dot{y} - a_2(c_1 + c_2)\,\dot{y}.$$

Substituting (19) for $\dot{\lambda}_1$ and (23) for $\dot{y}$ and cancelling terms, yields

$$\dot{e} = (a_1 b_2 + a_2 b_1)\,\lambda_1 + a_1 a_2 (c_1 + c_2)\, y - a_2 b_1 (c_1 + c_2). \tag{24}$$

Imposing the condition $\dot{e} = 0$, we see that

$$\lambda_1 = k_2 - k_1 y, \tag{25}$$

where

$$k_1 = a_1 a_2 (c_1 + c_2)/(a_1 b_2 + a_2 b_1), \qquad k_2 = a_2 b_1 (c_1 + c_2)/(a_1 b_2 + a_2 b_1). \tag{26}$$

But (25) and $e = 0$ imply that y is constant and given by the quadratic equation

$$\begin{aligned}
(a_1 k_1 - a_2 k_1)\, y^2 &+ (-b_1 k_1 - a_1 k_2 + a_2 k_2 - b_2 k_1 - a_2(c_1 + c_2))\, y \\
&+ b_1 k_2 + b_2 k_2 - (c_1 + c_2)\, b_2 = 0,
\end{aligned} \tag{27}$$

or, if $k_1 \neq 0$,

$$\begin{aligned}
(a_1 - a_2)\, y^2 &- [b_1 + b_2 + (a_1 - a_2)(k_2/k_1) + a_2(c_1 + c_2)/k_1]\, y \\
&+ (k_2/k_1)(b_1 + b_2) - b_2(c_1 + c_2)/k_1 = 0.
\end{aligned}$$

[3] Problems with this characteristic are called *singular*. See Athans and Falb (1966), especially Section 6-21.

From (26) we have

$$k_2/k_1 = b_1/a_1, \qquad (c_1 + c_2)/k_1 = (a_1 b_2 + a_2 b_1)/a_1 a_2$$

and so (27) becomes

$$(a_1 - a_2)\, y^2 - 2(b_1 + b_2)y + (a_2 b_1{}^2 - a_1 b_2{}^2)/a_1 a_2 = 0. \qquad (28)$$

The solutions of (28) are

$$y = (\sqrt{a_2}\, b_1 \pm \sqrt{a_1}\, b_2)/((\sqrt{a_1} \mp \sqrt{a_2})\, \sqrt{a_1 a_2}), \qquad a_1 \neq a_2, \qquad (29)$$

where the $\pm$ and $\mp$ are taken respectively, and

$$y = (b_1 - b_2)/2a \qquad \text{for} \quad a_1 = a_2 = a.$$

By checking the functions f_1 and f_2, we see that the lower signs in (29) must be chosen in order to keep f_1 and f_2 positive (which is the only meaningful interpretation for this problem). Thus,

$$y^* = (\sqrt{a_2}b_1 - \sqrt{a_1}b_2)/(\sqrt{a_1} + \sqrt{a_2})\sqrt{a_1 a_2}, \qquad (30)$$

and from (23) with $\dot{y} = 0$ (which must be the case for y to be constant) we calculate that

$$u^* = f_2(y^*)/[f_1(y^*) + f_2(y^*)]$$
$$= \sqrt{a_2}/(\sqrt{a_1} + \sqrt{a_2}), \qquad (31)$$

which is also valid for $a_1 = a_2$. The asterisk (*) is used to identify values of variables along a "turnpike" solution.

This solution is shown in Fig. 3 which has been drawn for the particular values $a_1 = 0.1$, $a_2 = 0.05$, $b_1 = 2$, $b_2 = 1$ so that

$$u^* = (1 + \sqrt{2})^{-1}, \qquad y^* = 20(3 - 2\sqrt{2}).$$

The figure shows that y^* is less than y' the value such that $f_1(y') = f_2(y')$. If the solution were at y' then the control would be $u' = 1/2$ in order to keep y at this value. Since the optimal u^* is less than $1/2$, more time is spent on strand two and, consequently, the optimal y^* is less than y'. The y^* solution is better than y' since $a_1 > a_2$ and hence the increase in the learning rate on strand one more than compensates for the decrease on strand two. It is shown later that this solution actually maximizes $\dot{x}_1 + \dot{x}_2$ so that this is a *maximal average learning rate path*.
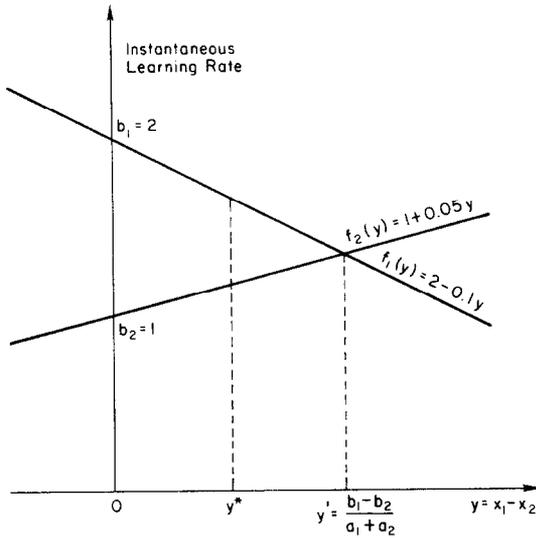
FIG. 3. Particular learning rate characteristics.

Returning to the basic problem, we must determine where the control variable is zero, one or given by the above solution. To simplify the notation of this development, the dependence of variables u and e on t will be omitted. The optimal value of u is determined by the sign of the quantity e defined by (22). To maximize H as given in (18), we see that

$$e > 0 \Rightarrow u = 1,$$

$$e < 0 \Rightarrow u = 0.$$

The solution to this problem will be shown to be of the form shown in Fig. 4. What is shown in this figure is the $x_1 x_2$ plane. At any point in time, the position of the student in terms of x_1 and x_2 can be shown in this plane. His progress is measured by movement to the right and up. From most initial points in the $x_1 x_2$ plane, the solution progresses as quickly as possible ($u = 0$ or $u = 1$) to the turnpike and stays on the turnpike for a period of time. Then near the end of time T, the solution leaves the turnpike with $u = 0$ or $u = 1$ depending on the slope of the objective function. The objective function can be represented in the $x_1 x_2$ plane by a straight line, i.e., a plot of $c_1 x_1 + c_2 x_2$ equal to a constant. The slope of this line is determined by the relative weights c_1, c_2 given to the two strands in the objective function.

The solution may not involve the turnpike path at all, especially for small values of final time T.
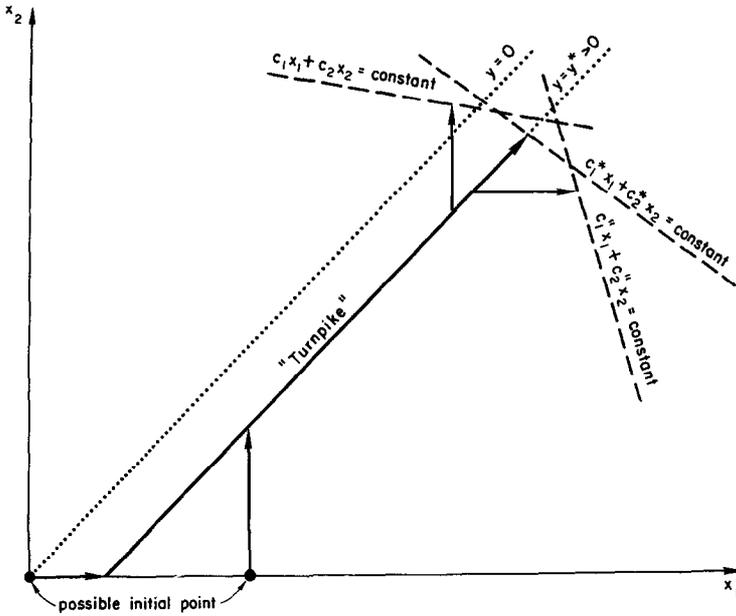
FIG. 4. Turnpike in x_1x_2 plane. Optimal trajectories shown for two possible initial points and three possible ratios of objective function weighting coefficients.

To complete the turnpike solution let us first consider the case of being on the turnpike and determine how to optimally proceed to the final time.

While the solution is on the turnpike, substitution of the value for y^* from (30) into (25) for the adjoint variable λ_1, we see that

$$\lambda_1^* = (c_1 + c_2)/(a_1b_2 + a_2b_1)\, a_2(b_1 - a_1 y^*)$$

$$= (c_1 + c_2)\, \sqrt{a_2}/(\sqrt{a_1} + \sqrt{a_2}).$$

Or, using (31) for u^*, we have

$$\lambda_1^* = (c_1 + c_2)\, u^* \tag{32}$$

and from (21)

$$\lambda_2^* = (c_1 + c_2)(1 - u^*). \tag{33}$$

Now consider the possibility of leaving the turnpike, i.e., the possibility that u switches from $u = u^*$ to $u = 0$ or $u = 1$. First consider the case $u = 1$. From (19)

$$\dot{\lambda}_1 = a_1 u \lambda_1 - a_2(1 - u)\, \lambda_2, \tag{34}$$

so that if u takes the value one, we have

$$\dot{\lambda}_1 = a_1 u \lambda_1 . \tag{35}$$

Since $\lambda_1 > 0$ on the turnpike (see (32)), (35) implies that λ_1 increases exponentially as long as $u = 1$.

From (23),

$$\dot{y} = - [a_1 u + a_2(1 - u)]y + u(b_1 + b_2) - b_2 ; \tag{36}$$

so if $u = 1$, then

$$\dot{y} = b_1 - a_1 y = f_1(y). \tag{37}$$

Now $f_1(y^*) > 0$ and $f_1(y)$ is positive in all cases of interest in this problem so that (37) implies that y increases as long as $u = 1$.

Finally, from (24), we see that (35) and (37) imply that $\dot{e}$ goes positive (from the value zero on the turnpike) as long as $u = 1$. If $\dot{e} > 0$, then $e > 0$, since $e = 0$ on the turnpike. Thus $e > 0$ and hence $u = 1$ is a possible optimal control. Note that if $u = 1$ is optimal on leaving the turnpike then u must stay at the value 1 until the end, $t = T$, since (35) and (37) do not change sign and so e can only increase in value (i.e., more positive).

Similarly for the case of $u = 0$ upon leaving the turnpike, we have from (34)

$$\dot{\lambda}_1 = -a_2 \lambda_2 ,$$

or, equivalently, by using (21),

$$\dot{\lambda}_1 = a_2 \lambda_1 - a_2(c_1 + c_2), \qquad \lambda_1(T) = c_1 , \tag{38}$$

which is negative (see (32)). Thus λ_1 decreases as long as $u = 0$. From (36), with $u = 0$, we have

$$\dot{y} = -a_2 y - b_2 = -f_2(y), \tag{39}$$

which implies that y decreases as long as $u = 0$. Thus (24) shows that e goes negative and, indeed, $u = 0$ is a possible optimal control. Note again that if $u = 0$ as the solution leaves the turnpike, that e cannot return to zero so u must stay at zero until the end.

The above arguments have proven the important conclusion that the optimal solution is such that if the trajectory in the $x_1 x_2$ plane ever follows the turnpike and subsequently leaves the turnpike, then the optimal trajectory never returns to the turnpike.

To determine the optimal value of u on leaving the turnpike (whether $u = 0$ or $u = 1$) we have the conditions from (19) that

$$\lambda_1(T) = c_1 ; \qquad \lambda_2(T) = c_2 . \tag{40}$$

Figure 5 shows the four possible situations with four typical values of the constants c_1 and c_2. The values of λ_1 and λ_2 while the solution is on the turnpike are given by (32) and (33). Since the final values are given by (40) then (34) and (21) are sufficient to determine the optimal time when the control variable switches from the turnpike value (u^*) to either zero or one. Whether the final value of control is zero or one depends on the relative values of c_1 and λ_1^* as shown in Fig. 5. The parameter values for Fig. 5 are the same as for Figs. 3 and 4.

Figure 5(a) is the case where

$$c_1/(c_1 + c_2) < u^* = \sqrt{a_2}/(\sqrt{a_1} + \sqrt{a_2}).$$
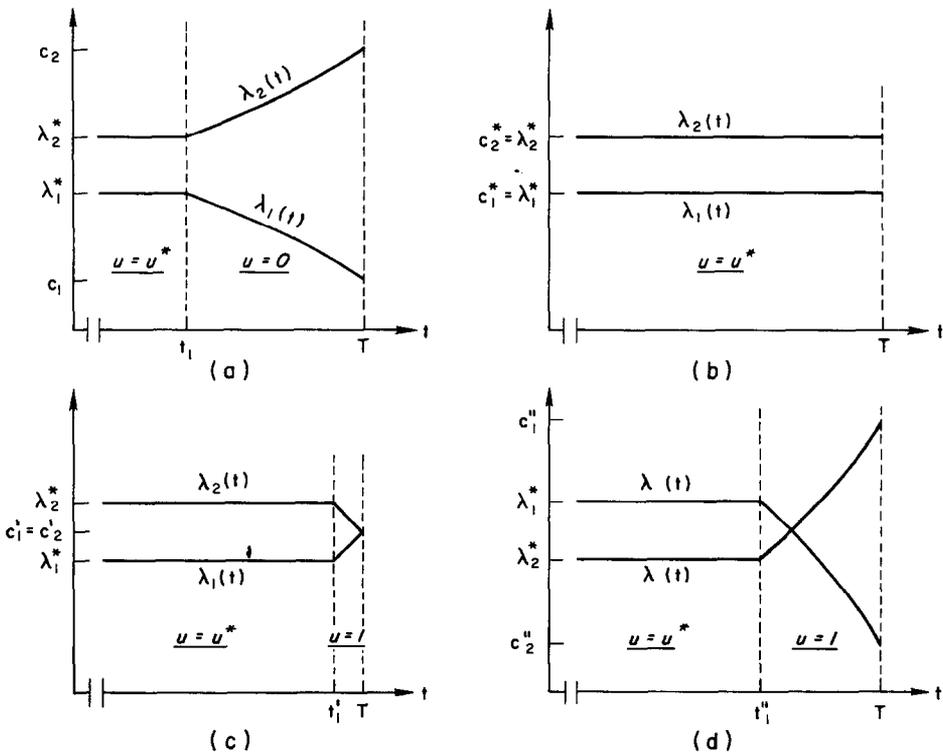

FIG. 5. Adjoint variables near end of study period for four different c_1/c_2 ratios.

In order for the λ's to reach their required terminal values, the control variable must switch to $u = 0$ at the time t_1 defined by the solution of (38).

Figure 5(b) is the case where the optimal solution stays on the turnpike right up to final time T. This occurs when

$$c_1^*/(c_1^* + c_2^*) = u^*$$

and not for $c_1 = c_2$ as might be intuitively assumed.

Figure 5(c) is the case where the objective coefficients c_1' and c_2' are equal and, as in Fig. 5(d), since

$$c_1''/(c_1'' + c_2'') > u^*,$$

the optimal control is $u = 1$ from the switching time t_1 to final time T. Again the value of t_1 is determined by (38).

The optimal path approaching the turnpike is determined simply by the relative values of y_0 and y^* where y_0 is the initial value of the difference of state values

$$y_0 = x_1(0) - x_2(0)$$

and y^* is this difference for states on the turnpike (see (30)). If y_0 is less than y^* then the path to the turnpike is defined by $u = 0$ and if y_0 is greater than y^* then $u = 1$.

For $u = 0$, the state variable x_1 does not change and x_2 increases so the path in the $x_1 x_2$ plane is parallel to the x_2 axis (see Fig. 4). The length of time for $u = 0$ can be determined by solving the differential equation (7) for x_2 using the given initial condition and the terminal condition that

$$x_1(t_0) - x_2(t_0) = y^*,$$

or, equivalently, since $x_1(t) = x_1(0)$,

$$x_2(t_0) = x_1(0) - y^*.$$

The solution for the time to the turnpike, t_0, is

$$e^{-a_2 t_0} = (b_2 + a_2 y^*)/(b_2 + a_2 y_0) = f_2(y^*)/f_2(y_0).$$

Also, during this initial period when $u = 0$, the solution for the adjoint variable $\lambda_1(t)$ is

$$\lambda_1(t) = (c_1 + c_2)[1 - (1 - u^*) e^{a_2(t - t_0)}]$$

for t in the interval $[0, t_0]$.

For $u = 1$, x_2 does not change but x_1 increases so the path is parallel to the x_1 axis (see Fig. 4). The time t_0' of this initial period is given by

$$e^{-a_1 t_0'} = (b_1 - a_1 y^*)/(b_1 - a_1 y_0) = f_1(y^*)/f_1(y_0).$$

The adjoint variable is given by

$$\lambda_1(t) = (c_1 + c_2) u^* e^{a_1(t - t_0')}, \tag{41}$$

for t in the interval $[0, t_0']$. Figure 6 shows the adjoint variables for the two particular initial state vectors shown in Fig. 4 for parameter values as used in Figs. 3–5.

It is possible that the optimal solution may not use the turnpike at all. This would be the case especially for small values of total learning time T. To determine if this is
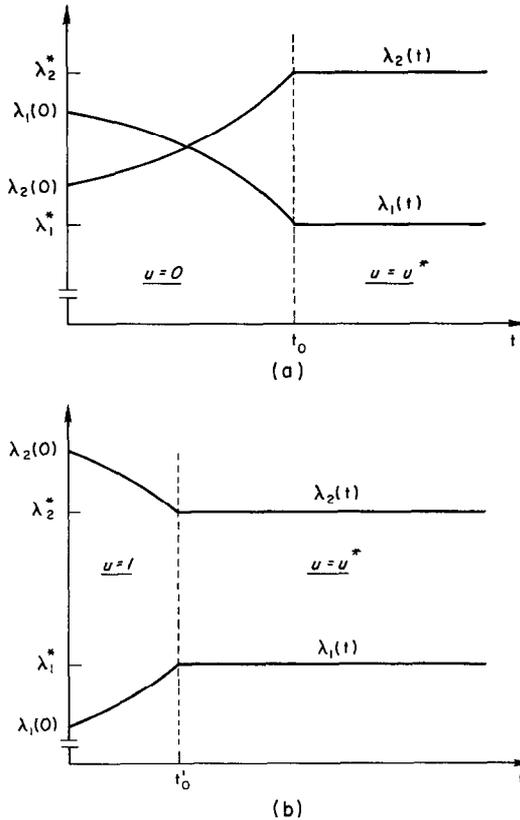

FIG. 6.   Adjoint variables near beginning of study period for two different initial points. Case (a) $y_0 > y^*$ and Case (b) $y_0 < y^*$.

indeed the optimal solution, the objective value must be calculated for each of the four possibilities of the control sequence, i.e., all zero, all one, zero then one and one then zero. There will be no more than one switch point for these solutions as shown by the previous analysis of the dynamics of the λ's and the quantity e which determine optimal u's.

Maximal Average Learning Rate Solution

Instead of using the optimal control approach to this problem, suppose we think of the problem as follows: If the optimal solution includes a path in the $x_1 x_2$ plane such that $y = x_1 - x_2$ is constant along the path, then that path should be the solution of the problem

$$\text{maximize}_{y} \; g(y) = \tfrac{1}{2}(\dot{x}_1 + \dot{x}_2) \tag{42}$$

subject to $x_1 - x_2 = y = $ constant. As justification for this conjecture, consider a comparison path that does not follow this maximal average learning rate path. Thus for all values of final time T larger than some particular value, say T', the time saved by following the turnpike would more than compensate for the time it would take to get to the turnpike from this comparison path plus the time to return to the comparison path.

The solution of (42) is straightforward. From (7) we have

$$g(y) = \tfrac{1}{2}[u f_1(y) + (1 - u) f_2(y)].$$

But to satisfy the constraint of (42) we have

$$\dot{x}_1 = \dot{x}_2$$

which implies that

$$u = f_2(y)/[f_1(y) + f_2(y)]. \tag{43}$$

Thus the function $g(y)$ becomes

$$g(y) = f_1(y) f_2(y)/[f_1(y) + f_2(y)], \tag{44}$$

where f_1 and f_2 are positive in the region of interest for this problem.

For y^* to maximize (44), it is necessary that

$$g'(y^*) = 0$$

i.e.,

$$f_1^2(y^*) f_2'(y^*) + f_1'(y^*) f_2^2(y^*) = 0$$

or

$$[f_1(y^*)/f_2(y^*)]^2 + [f_1'(y^*)/f_2'(y^*)] = 0. \tag{45}$$

For the problem in question f_1 and f_2 are linear varieties and are given by (10). Thus (45) becomes

$$[(b_1 - a_1 y^*)/(b_2 + a_2 y^*)]^2 - (a_1/a_2) = 0 \tag{46}$$

i.e.,

$$(b_1 - a_1 y^*)/(b_2 + a_2 y^*) = f_1(y^*)/f_2(y^*) = \pm(a_1/a_2)^{\frac{1}{2}},$$

where the positive sign must be chosen since f_1 and f_2 must be positive. Now (43) shows that

$$u = u^* = \sqrt{a_2}/(\sqrt{a_1} + \sqrt{a_2}), \tag{47}$$

which is the same result as before (see (31)). Equations 46 and 47 can be combined to show that the value of y^* is the same as the one derived in (30).

Solution of Discrete Model

Deterministic Model

The complete solution of the simplified continuous model was developed in the previous section. The optimal solution was such that the control variable $u(t)$ was a piecewise constant function with at most three values and two points of discontinuity. For example, as shown in Fig. 4 for a particular initial point and particular objective function weighting constants c_1 and c_2, the optimal solution is

$$u(t) = \begin{cases} 1 & 0 \leqslant t < t_0, \\ u^* & t_0 \leqslant t < t_1, \\ 0 & t_1 \leqslant t \leqslant T, \end{cases} \tag{48}$$

where u^* is the value of $u(t)$ to keep the trajectory on the turnpike.

The discrete model (15)–(17) was shown to be equivalent to the continuous model. Thus the optimal solution for the discrete model is also given by (48). However, in the discrete model, the control cannot change value within the interval so that if the optimal switching time t_0 or t_1 does not occur at the end of the discretized interval of the discrete model, then the controls cannot be made identical. One way to alleviate this difficulty is to make the discretized interval very small so that the error involved can be ignored. An exact equivalence can be made, however, if the interval containing switching is divided into two intervals at the switching point. The discrete model was set up with equal intervals but there is no mathematical difficulty in making intervals unequal. For

the two short intervals, separate calculations of the matrices $\phi(k)$ and $B(k)$ would have to be made. For all the other intervals where $u(k)$ does not change, the matrices $\phi(k)$ and $B(k)$ do not change so a one-time calculation is sufficient.

The matrices $\phi(k)$ and $B(k)$ will be required for the next section. Equations 15 and 16 may be solved explicitly for a given $u(k)$ by expressing the F matrix in modal form, i.e.,

$$F = U\Gamma U^{-1},$$

where U is the modal matrix of eigenvectors of F and Γ is the diagonal matrix of eigenvalues of F. To simplify the notation we make the definitions

$$a_1' = a_1 u(k),$$
$$a_2' = a_2[1 - u(k)],$$

(49)

so that (12) becomes

$$F = \begin{bmatrix} -a_1' & a_1' \\ a_2' & -a_2' \end{bmatrix}.$$

The eigenvalues of F are $-(a_1' + a_2')$ and zero. Now, with some manipulation, (15) and (16) yield

$$\phi(k) = \exp(-a_1' - a_2')M(k) + N(k),$$

(50)

$$B(k) = \left\{ \frac{1 - \exp(-a_1' - a_2')}{(a_1' + a_2')} M(k) + N(k) \right\} G(k),$$

(51)

where

$$M(k) = \frac{1}{a_1' + a_2'} \begin{bmatrix} a_1' & -a_1' \\ -a_2' & a_2' \end{bmatrix},$$

(52)

$$N(k) = \frac{1}{a_1' + a_2'} \begin{bmatrix} a_2' & a_1' \\ a_2' & a_1' \end{bmatrix}.$$

(53)

For intervals where $u(k) = 0$, then $a_1' = 0$ and $a_2' = a_2$ so that

$$\phi(k) = \begin{bmatrix} 1 & 0 \\ 1 - e^{-a_2} & e^{-a_2} \end{bmatrix},$$

and

$$B(k) = \begin{bmatrix} 0 \\ \frac{b_1}{a_2}(1 - e^{-a_2}) \end{bmatrix}.$$

Thus

$$x(k + 1) = \phi(k)x(k) + B(k)$$

shows that

$$x_1(k + 1) = x_1(k),$$

as required for $u(k) = 0$ and

$$x_2(k + 1) = e^{-a_2}x_2(k) + [x_1(k) + (b_2/a_2)](1 - e^{-a_2}).$$

Stochastic Model

In this section we will examine the discrete model with the random components included as originally posed in (3)–(6) except that the learning rate characteristics will still be approximated by the linearized version (Eq. 10 and Fig. 2). An optimal control policy for this problem would be very difficult to determine although, presumably, a computational procedure would find such a solution for a particular problem. However, with the optimal solution for the deterministic problem at hand, we can formulate a suboptimal control policy that approximates our previous solution.

The source of the difficulty at this point is the introduction of the random components. The randomness in the state Eq. 5 reflects uncertainty about how the student's achievement level changes from day to day even though the instructional input is known. The randomness in Eq. (6) reflects uncertainty about the state of the student even though a measurement (e.g., administering a test) has just been taken. Our solution for this problem is to calculate an estimate of the state vector at the end of each day and use this estimate to determine the instructional (control) input for the next day. The estimation procedure will make use of the Kalman filtering technique (Kalman, 1960) which yields the minimum variance estimate of the state taking into account all previous measurements, the history of control inputs and the dynamics of the system.

The state estimator and system controller configuration is shown in Fig. 7. Using the initial state estimate, $\hat{x}(0)$, the controller chooses an instructional input for the first day, $u(0)$. The achievement level of the student at the end of the first day is described by the state vector $x(1)$ which is the input to the measurement system. By a test performance, a measurement $z(1)$ of this state is made. The estimator takes this measurement and information from the controller about the instructional input and delivers a new estimate of the state of the student, $\hat{x}(1)$. The cycle now continues producing a multistage process.

The system dynamics is described by (3) with the linearized learning rate characteristics of (10). That is,

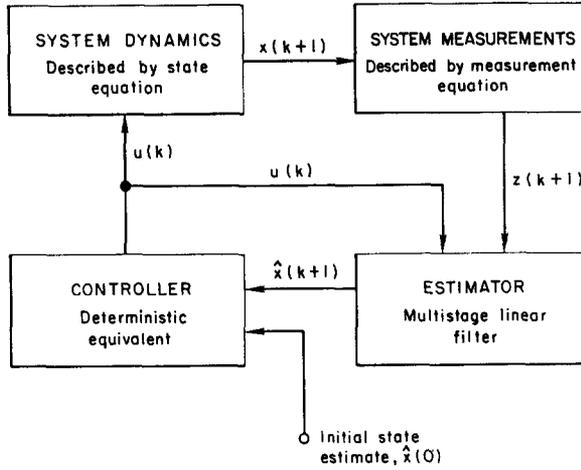$$x(k + 1) = \phi(k)x(k) + B(k) + n(k), \tag{54}$$

FIG. 7.   Block diagram of system with estimator and controller.

where

$\phi(k)$ is given by (50),

$B(k)$ is given by (51),

$n(k)$ is two component gaussion white noise vector with zero mean and
covariance matrix $E[n(k)\, n'(l)] = S\delta_{kl}$, where δ_{kl} is the Kronecker
delta.

The measurement system is described by (4) where the measurement noise $w(k)$ is
independent of $n(k)$, zero mean and with covariance matrix $E[w(k)\, w'(l)] = R\delta_{kl}$.

The estimator is described by the multistage linear filter equations (Bryson and
Ho, 1969),

$$\hat{x}(k + 1) = \bar{x}(k + 1) + P(k + 1)\, R^{-1}[z(k + 1) - \bar{x}(k + 1)], \qquad (55)$$

$$P(k + 1) = Q(k + 1) - Q(k + 1)[Q(k + 1) + R]^{-1}Q(k + 1), \qquad (56)$$

where

$$\bar{x}(k + 1) = \phi(k)\hat{x}(k) + B(k), \qquad (57)$$

$$Q(k + 1) = \phi(k)P(k)\phi(k)^{T} + S, \qquad (58)$$

$\bar{x}(k)$ is the best prediction of $x(k)$ given measurements up to time $k - 1$,

$P(k)$ is the covariance matrix of the estimate error $\hat{x}(k) - x(k)$,

$Q(k)$ is the covariance matrix of the error in the predictive estimate
$\bar{x}(k) - x(k)$.

Equations 55–58 may be interpreted as follows: The best estimate of $x(k + 1)$ after having the measurements $z(k + 1)$ is the prior predictive estimate $\bar{x}(k + 1)$, plus a weighted correction due to the "new" information in the measurements $z(k + 1) - \bar{x}(k + 1)$. The covariance of the new estimate $P(k + 1)$ decreases from the variance of the predictive estimate $Q(k + 1)$ by an amount dependent on the covariance of the measurement noise R. The predictive estimate $\bar{x}(k + 1)$ is the previous estimate modified by the system dynamics with its covariance being the modified previous estimate covariance plus the system noise covariance.

The controller system proposed here is based on the optimal solution to the deterministic model as given in the previous section. Specifically, the best estimate of the current state of the student as calculated by the estimator is treated as if it were the true state and the corresponding optimal control for the next period is determined.

The controller divides the $x_1 x_2$ plane into three regions. The Turnpike Region contains the turnpike path (see Fig. 4); it is defined as that region in the $x_1 x_2$ plane such that there exists an admissible control input that will transform the state to a point on the turnpike in one period (ignoring the noise component). Region Zero is below and to the right of the Turnpike Region, where Region One is above and to the left of the Turnpike Region. See Fig. 8.

If the state estimate is in Region Zero the controller assigns a control input of zero and hence the path of the system in the $x_1 x_2$ plane is parallel to the x_2 axis, toward the turnpike. If the state estimate is in Region One the controller assigns a control input of one, thus driving the system toward the turnpike as fast as possible. In the Turnpike Region, the controller calculates what value of control input will drive the system to the turnpike and assigns this value as the input for the next period.
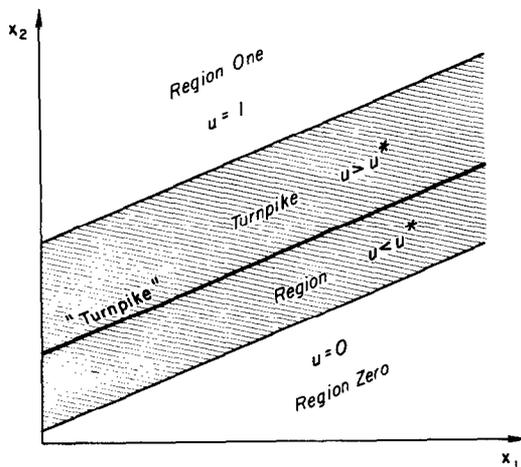

FIG. 8. Regions in $x_1 x_2$ plane defined by controller.

In order to determine the region boundaries and calculate the value of control for the Turnpike Region, the following problem must be solved: given the state $x(k)$ at time k, what value of $u(k)$ will put $x(k + 1)$ on the turnpike path (in expected value sense) and is this $u(k)$ admissible (i.e., between zero and one)? If $x(k + 1)$ is on the turnpike, then

$$x_1(k + 1) - x_2(k + 1) = y^*. \qquad (59)$$

Now from (17),

$$x_1(k + 1) - x_2(k + 1) = \phi_{11}x_1(k) + \phi_{12}x_2(k)$$

$$-\phi_{21}x_1(k) - \phi_{22}x_2(k) + B_1(k) - B_2(k)$$

and substituting from (50), (51) and denoting $x_1(k) - x_2(k)$ by $y(k)$

$$y(k + 1) = y(k)\exp(-a_1' - a_2') + [(b_1 + b_2)u(k) - b_2]\frac{\exp(-a_1' - a_2') - 1}{-a_1' - a_2'}. \qquad (60)$$

Since the a_i''s involve $u(k)$, this equation cannot be solved explicitly for $u(k)$ as a function of $y(k)$. For specific parameter values the relationship between $u(k)$ and $y(k)$ could be determined numerically. However, if the exponentials in (60) are approximated by linear terms only and the a_i' variables are expressed in terms of the original a_i variables, then

$$y(k + 1) = y(k)[1 + (a_2 - a_1)u(k) - a_2] + (b_1 + b_2)u(k) - b_2 .$$

Now (59) requires that

$$u(k) = \frac{f_2(y(k))}{f_1(y(k)) + f_2(y(k))} + \frac{y^* - y(k)}{f_1(y(k)) + f_2(y(k))}, \qquad (61)$$

where $f_1(y(k))$ and $f_2(y(k))$ are defined earlier in (10). Notice that the first term of (61) is the value of control that would maintain the system at the current "distance" from the turnpike since with this control $y(k + 1) = y(k)$. The second term adjusts the control input upward or downward depending on the sign of $y^* - y(k)$, i.e., depending on whether the current state is above or below the turnpike. If (61) gives a value of $u(k)$ less than zero or greater than one, then the turnpike cannot be reached by an admissible control and $u(k) < 0$ corresponds to Region Zero and $u(k) > 1$ corresponds to Region One.

The control near the final time T is either zero or one depending on the relative values of c_1 and c_2 as shown in Fig. 5. The time of switching from control by (61) to terminal control is given by the solution of (21), (34) and (40).

COMMENTS AND CONCLUSIONS

All of the results presented are for the case with the learning rate characteristics in linearized form. By the nature of the nonlinear characteristics it is intuitively clear that the optimal solution for the nonlinear case has the same "turnpike" form. Therefore, once the location of the turnpike path has been calculated, which is defined for the nonlinear case by (45), then the solution for this case is straightforward. If the initial estimate of the state is far removed from the Turnpike Region, then the estimation for the initial stages could use parameters derived from a local linearization of the non-linear characteristics in order to improve the estimator accuracy. Once the state estimate is "close" to the Turnpike Region, the parameters need not be updated throughout the remainder of the study period.

An extension of the results presented here to a learning environment with three or more curriculum strands should be straightforward. Although the results could not be viewed as simply as was possible in the present case (i.e., with a scalar difference variable $y(k)$ and two dimensional state space graphs) the concepts of the solution would transfer to the multistrand case. The turnpike path would be identified by finding a maximal growth path such that for some control policy the amount learned in each strand would be the same each day. In the multistrand case the linear filter estimator would make use of a linearized learning model about the turnpike path.

In the previous section the controller-estimator system made use of measurements after each learning session. The common measurement method of testing is costly in the sense of student time as well as instructor preparation and evaluation time. For this reason it may be advantageous to reduce the amount of testing by taking measurements only every other day or even less frequently. This possibility can be immediately implemented with the proposed system. The estimator described previously will provide minimum variance estimates even without new measurements but, of course, the variance of the estimate error will increase without new information. One attractive possibility would be to take frequent measurements during the initial stages until the state trajectory was near the turnpike path and the estimate error variance was small. Then the system could proceed with only occasional measurements as long as the error variance remained small. If a later measurement was farther from the predicted value than some predetermined threshold, more frequent measuring could be resumed. In learning environments where testing is extremely costly or impractical, then, an open loop control policy could be defined for the entire study period based only on the initial estimate of the state of the student.

This paper has presented an analytic solution to an optimization problem in the theory of multistrand learning. The form of this solution was used to define a sub-optimal controller for a more realistic learning model that included random components. The proposed controller was of a closed-loop design that made use of a separate estimator system based on linear filtering theory.

References

ATHANS, M., AND FALB, P. L. *Optimal control: An introduction to the theory and its application.* New York: McGraw–Hill, 1966.

ATKINSON, R. C., AND FLETCHER, J. D. Teaching children to read with a computer. *The Reading Teacher,* 1972, **25,** 319–327.

ATKINSON, R. C., AND PAULSON, J. A. An approach to the psychology of instruction. *Psychological Bulletin,* 1972, **78,** 49–61.

BRYSON, A. E., JR., AND HO, Y. C. *Applied optimal control.* Waltham, MA: Blaisdell, 1969.

GROEN, G. J., AND ATKINSON, R. C. Models for optimizing the learning process. *Psychological Bulletin,* 1966, **66,** 309–320.

KALMAN, R. E. A new approach to linear filtering and prediction problems. *Trans. ASME, Journal of Basic Engineering,* 1960, **82,** 35–45.

KARUSH, W., AND DEAR, R. E. Optimal stimulus presentation strategy for a stimulus sampling model of learning. *Journal of Mathematical Psychology,* 1966, **3,** 19–47.

LUENBERGER, D. G. *Optimization by vector space methods.* New York: Wiley, 1969.

PONTRYAGIN, L. S., BOLTYANSKII, V. G., GAMKRELIDZE, R. V., AND MISHCHENKO, E. F. *The mathematical theory of optimal processes.* (Translated by K. N. Trirogoff, edited by L. W. Neustadt.) New York: Wiley–Interscience, 1962.

SAMUELSON, P. A. Efficient paths of capital accumulation in terms of the calculus of variations. In Arrow, K. J., Karlin, S., Suppes, P. (Eds.), *Mathematical methods in the social sciences.* Stanford, CA: Stanford University Press, 1960.

SMALLWOOD, R. D. *A decision structure for teaching machines.* Cambridge, M.I.T. Press, 1962.