JOURNAL OF MATHEMATICAL ANALYSIS AND APPLICATIONS 10, 174-205 (1965)

Optimal Control of Markov Processes with Incomplete State Information

K. J. ÅSTRÖM

IBM Nordic Laboratory, Sweden

Submitted by Richard Bellman

I. INTRODUCTION

The problem discussed in this paper has grown out of an attempt to construct a theoretical framework which is suitable for the treatment of some of the problems arising in the design of control systems. The fundamental problem is to find a suitable control law, i.e., a relationship between the observed output and the control signal. In the earlier stages of the development of control theory it was customary to postulate a certain structure for the control law which had a few unknown parameters to be determined.

In more recent developments, the trend is to replace this postulate by the postulate that the purpose of the control system is to minimize a cost function [1, 2].

The design problem is then reduced to a variational problem, whose solution will yield the control law. In this approach the control law expresses the control signal as a function of the state variables. Hence there is no dynamic element in the feedback. It should also be noted that with an approach of this type there is no analytical difference between a *control law* and a *control schedule*, or in other words, between a *closed loop system* and an *open loop system.*

The control law being a function of the state variables implies that all state variables must be measured. If only measurements of a few state variables are available and if the system is observable, the other state variables can easily be reconstructed by differentiation. Hence there is no natural way to introduce, as a limitation, the fact that only a few state variables can be measured. Thus, the treatment of the control problem as a deterministic variational problem is not a completely satisfactory approach.

It is clear that this inadequacy arises when disturbances are neglected. One way to remedy this definiency is to introduce disturbances as random functions. If the control problem is still formulated so as to minimize functionals of the trajectories of the system, we are led to a stochastic variational problem. Such a problem is, in general, very difficult to solve. There is, however, one special case which can be solved: linear systems with quadratic criteria. See [3–6]. This solution is the foundation of linear control theory which is complete in the sense that many essential problems, such as stability, sampling rate, etc., can be solved. See [7]. The solution of the linear problem has an interesting structure. The feedback law obtained can be considered as consisting of two parts:—the estimation of the state of the system from the observed measurements, and the calculation of the control variable from the estimated state. The first part is reduced to the solution of a differential equation. The second part is simply the evaluation of a linear function. This function can be obtained as the solution of a deterministic control problem. Both the differential equation and the linear function contain parameters which can be precomputed from the a priori data of the problem and stored in tables. By doing this, the amount of real time computations required in the implementation of the optimal system is significantly reduced. It should also be noted that the linear stochastic control theory provides an interesting approach to a class of adaptive systems: a typical situation is the case where the disturbances have constant but unknown averages. It is easily seen that this case is transformed to the standard linear problem by introducing the unknown averages as new state variables. See [7].

It would indeed be interesting to pursue the same idea for stationary systems with unknown parameters. It is easily seen that this problem can be transformed to a nonlinear stochastic variational problem where the unknown parameters are considered as state variables. Hence, a generalization of the linear stochastic control problem will lead directly to a stochastic variational problem. The solution of such a problem will provide an approach to a theory of adaptive control systems.

In developing a theory for a stochastic variational problem it is natural to rely on Markovian theory by making assumptions which will guarantee that the solutions of the differential equations with random disturbances which describe the system are Markov processes. The main reason for this approach is that the transition probabilities of Markov processes are governed by linear equations even if the original stochastic differential equations are highly nonlinear.

The control problem can be stated as follows. The system is described by a stochastic differential equation containing certain parameters, called *control variables*. The trajectories of the system can be influenced by the choice of these control variables. There is incomplete state information, i.e., only a few coordinates can be observed and the measurements of these coordinates are affected by disturbances. The performance of the system is characterized by a functional of the trajectories. The problem is to find values of the control variables such that the mathematical expectation of the functional is as small as possible, when the value of the control variable at time t may depend on all measurements prior to t. When trying to solve such a problem one is soon faced with great mathematical difficulties. However, by quantizing both the state and time variables of the problem formulated above, we arrive at a problem of essentially the same structure where some of the mathematical difficulties are eliminated. Such a problem is studied in this paper. When a solution to the quantized problem is obtained we may return to the original continuous problem by various limit processes. The quantized problem has, however, a value of its own in the sense that it represents a situation which arises when a digital computer is used to implement the optimal solution. In such a case the quantization in time enters naturally into the problem as sampling, and the quantization of the state variables is obtained from the analog to digital conversion.

A precise statement of the quantized problem is given in Section II. The solution of the problem is presented in Section III. The solution is obtained using Dynamic Programming, and the result is given in terms of a functional equation. To obtain the solution, some elementary results of the theory of conditional Markov processes are required. The functional equation obtained is an analog of the Hamilton-Jacobi equation in classical calculus of variations. In Section III we also present an inverse result which shows that if the basic functional equation has a solution, then the maximum exists. This is the analog of a theorem of Caratheodory in the classical calculus of variations [8, p. 200]. In Section IV we give some interpretations of the results obtained. It is shown that the optimal control law can be expressed as $u = u(w, t)$ where w is a function of the observed outputs. The function $u = u(w, t)$ can be calculated a priori, without any knowledge of the actual outputs of the system, as the solution of an associated problem with complete state information. The function w which is a function of the observations must obviously be calculated in real time as the outputs are observed. Dynamic feedback is obtained through this computation. We can thus divide the problem in the same way as is done in the linear quadratic case. This division is of great importance for the practical implementation of the solution, since computation of the function $u = u(w, t)$ is complicated and time consuming. Precomputing and storing this function greatly reduces requirements for real time computations and is a considerable simplification in the realization of the optimal system.

In Section V we compare the solution of our problem with two associated problems, namely, those of complete state information and of no state information at all. The latter problems are easier to solve and the solutions will provide bounds for the solution of our problem. The comparison will also make it possible to associate cost with state information. This provides an interesting connection between information theory and the theory of stochastic optimal control, which has not been explored. Finally, in Section VI we present two examples.

Although the problem has arisen from study of control systems, its solution may have applications in other fields. In queuing theory, to mention one example, it will thus be possible to treat queues so that service is made to depend on the past status of the queue.

II. Statement of the Problem

Let $\{x_t, t = 0, 1, ...\}$ be a Markov process with finite state space and discrete time. The states are labeled by the positive integers 1, 2, ..., n. The initial probability distribution is

$$p_i^0 = P\{x_0 = i\}$$

The row vector formed by $(p_1^0, p_2^0, ... p_n^0)$ is denoted by p^0. Let $P(u, t)$ be the matrix of the transition probabilities of the process. The ij component of P, $p_{ij}(u, t)$ is defined by

$$p_{ij}(u, t) = P\{x_t = j \mid x_{t-1} = i\} \tag{2.1}$$

where

$$p_{ij}(u, t) \geqslant 0, \qquad \sum_j p_{ij}(u, t) = 1$$

The transition probabilities may depend on time t, and on a set of parameters $u_1, ..., u_k$ which are combined to form a column vector u, and called *control variables* or *decision variables*, thereby reflecting the fact that the process x_t can be influenced by the choice of these parameters. It is assumed that $u(t)$ for each t belongs to a closed compact set U, which is called the *set of admissible controls at fixed times*, and that the transition probabilities are continuous functions of u. Further, let $\{y_t, t = 1, 2, ...\}$ be a discrete time random process which is related to the x process in the following way

$$y_t = f(x_t, e_t)$$

where $\{e_t, t = 1, 2, ...\}$ is a sequence of independent random variables, and the range of f is the integers 1, ..., m.

The realizations of the process $\{y_t, t = 1, 2, ...\}$ represents the results of the physical measurements of the process $\{x_t, t = 0, 1, ...\}$ and the process $\{y_t, t = 1, 2, ...\}$ is therefore referred to as the *output* of the system or the *observable*.

The function $f(x, e)$, which represents the characteristics of the measuring instruments, and the random variables $\{e_t, t = 1, 2, ...\}$ are specified by

$$q_{ij} = P\{y_t = j \mid x_t = i\} \tag{2.2}$$

where

$$q_{ij} \geqslant 0$$
$$\sum_j q_{ij} = 1$$

A particular realization of the process y_t or equivalently a particular outcome of the measurements is denoted by $\eta_1, \ldots, \eta_t$ and these numbers are grouped together to form the vector.

$$\eta(t) = \text{col}\,(\eta_1, \ldots, \eta_t) \tag{2.3}$$

The matrix Q formed by the $q_{ij}: s$ and defined by Eq. (2.2) will thus reflect measurement errors. Notice that Q is not necessarily a square matrix, the number of possible output states may differ from the number of states of the x process. When Q equals the unit matrix, we have *complete state information*, *i.e.*, each measurement gives the exact state with probability one.

When controlling the process $\{x_t, t = 0, 1, \ldots\}$ we want to determine $u(t)$ both as a function of the outputs observed up to time t and as a function of the previous control variables $u(1), \ldots, u(t-1)$. This is to be done in such a way that the behavior of the controlled process is optimal in some sense.

Let the observed outputs up to time t be $y_1 = \eta_1, \ldots, y_t = \eta_t$. The relation between the control variable $u(t)$, the observed outputs $\eta_1 \ldots, \eta_t$, and the previous control variables $u(1), \ldots, u(t-1)$ is expressed as

$$u(t) = c'(\eta_1, \ldots, \eta_t, u(1), \ldots, u(t), t) \qquad t = 1, \ldots, N$$

By successive substitutions we can immediately eliminate $u(1), \ldots, u(t-1)$ from the right-hand member and we get

$$u(t) = c(\eta_1, \ldots, \eta_t, t), \qquad t = 1, \ldots, N \tag{2.4}$$

The set of functions $C = \{c(\eta_1, \ldots, \eta_t, t), t = 1, \ldots, N\}$ is referred to as *strategy* or a *control law*. A control law C is admissible if $c(\eta_1, \ldots, \eta_t, t) \in U$ for all t and all possible η_i. As the control law C gives a relation between the measured outputs and the control signals it will also define *feedback*.

The object of controlling the process is specified in the following way. Let $g(u, x, t)$ be a scalar function of u, x and t. It is assumed that the dependence on u is continuous. The function g is called the *instantaneous cost function* and it gives the cost associated with the outcome $x_t = x$ and the control u. Further, the *total cost* of the process is defined as

$$L = \sum_{t=1}^{N} g(u(t), x_t, t) \tag{2.5}$$

The mathematical expectation of L is denoted

$$EL = E \sum_{t=1}^{N} g(u(t), x_t, t) = E \sum_{t=1}^{N} g(c(\eta_1, ..., \eta_t, t), x_t, t) \qquad (2.6)$$

where E denotes expectation with respect to the distributions of x_t and y_t.

We will now formulate the following problem.

*P.*1 find an admissible control signal whose value at time t is a function of the outputs observed up to that time and are such that the expected value of the total cost is minimal.

An alternative formulation is:

*P.*1′ find an admissible control law such that the expected value of the total cost (L) is minimal.

Remark. In the statement of the problem it is postulated that $u(t)$ is a function of $\eta_1, ..., \eta_t$. As $u(t)$ is allowed to be a function of η_t this implies that there are no delays in measurements, and that the time required to calculate the control signal from the measurements is negligible. There will be no essential change in the arguments if we instead postulate that $u(t)$ is a function of $\eta_1, ..., \eta_{t-s}$ thereby allowing for a delay of the measurements and the control computations of s units of time. The delay s may also be a function of time. In this way we can get a hierarchy of problems.

One particular case which deserves special attention is when $s(t) = t$. This means that u is just a function of time (and of the a priori information) and that no measurements are used. The control function $u(t)$ obtained in this way is called a *control schedule* and the system obtained is called an *open loop system*, as there is no feedback from the measurements. These variations in the formulation of the problem are both of practical and theoretical interest; they give us tools to analyze the influence of delays in measurements and to form estimates of minimal loss. This is of importance when analyzing different schemes for implementing a system, and for discussions of convergence, etc.

III. Solution of the Problem

A. Dynamic Programming

Leaving questions concerning existence and uniqueness of the solution aside for a moment, we will now postulate that the problem has a solution, and we will characterize this solution by a functional equation. We will then go back to find conditions which ensure existence and uniqueness. The function to be minimized is the mathematical expectation of the total cost. We will thus have to find

$$\min_{u(1)} \dots \min_{u(N)} E \sum_{t=1}^{N} g(u(t), x_t, t) \tag{3.1}$$

where expectation is taken with respect to the distributions of $\{x_t, t = 0, 1, ...\}$ and $\{y_t, t = 1, 2, ...\}$ and where $u(t)$ has to be a function of the outputs observed up to time t, i.e., of $\eta_1 = y_1, ..., \eta_t = y_t$. The technique of Dynamic Programming will be used to solve the problem.

Let us first consider the situation at the last step, i.e., $t = N$. The outputs $y_1 = \eta_1, ..., y_N = \eta_N$ have been observed and the problem is to determine $u(N)$ as a function of these. We notice that the only term of the sum in expression (3.1) for the total cost that depends on $u(N)$ is the last one, i.e., $g(u(N), x_N, N)$. The control signal $u(N)$ must therefore be chosen so as to minimize the quantity.

$$Eg(u(N), x_N, N) \tag{3.2}$$

Again E denotes expectation with respect to the distributions of the processes $\{x_t, t = 0, 1, ...\}$ and $\{y_t = 1, 2, ...\}$. The quantity (3.2) has to be minimized with respect to all $u(N)$ which are functions of $\eta_1, ..., \eta_N$. To perform the minimization we will first rewrite (3.2) so that the dependence of $\eta_1, ..., \eta_N$ is explicit. Using the definition of conditional expectation, we get

$$Eg(u, x_N, N) = \underset{\eta(N)}{E} [\underset{|\eta(N)}{E} g(u, x_N, N)] \tag{3.3}$$

where $E_{|\eta(N)}$ denotes mathematical expectation with respect to the conditional distribution of x_N, given $\eta(N)$ and $E_{\eta(N)}$ denotes the mathematical expectation with respect to the distribution of $\eta(N)$. The expression of the right member of (3.3) which is within brackets, is a function only of u, η, ..., η_N and we can thus perform the minimization. The minimal cost of the last step is thus

$$\underset{\eta(N)}{E} \min_{u} \underset{|\eta(N)}{E} g(u, x_N, N) \tag{3.4}$$

Let V_N be defined by

$$V_N = \min_{u} \underset{|\eta(N)}{E} g(u, x_N, N) \tag{3.5}$$

We notice that V_N is a function of $\eta_1, ..., \eta_N$ and N, but that the dependence of V_N on $\eta_1, ..., \eta_N$ only enters through the conditional distribution of x_N given $\eta(N)$. To emphasize this we introduce

$$w_i(N) = P\{x_N = i \mid y_1 = \eta_1, ..., y_N = \eta_N\} \tag{3.6}$$

and

$$w(N) = (w_1(N), w_2(N), ...) \tag{3.7}$$

and we write

$$V_N = V_N(w(N)) \tag{3.8}$$

Summarizing our findings so far, we find that the minimal cost during the final step is

$$\underset{\eta(N)}{E} V_N(w(N)) \tag{3.9}$$

We now proceed recursively to show that the minimal cost of the $N - k$ last steps can be written as

$$\min_{u(k+1)} ... \min_{u(N)} E \sum_{t=k+1}^{N} g(u(t), x_t , t) = \underset{\eta(k+1)}{E} V_{k+1}(w(k + 1)) \tag{3.10}$$

where we have introduced

$$V_k(w(k)) = \min_{u(k)} \underset{|\eta(k)}{E} ... \min_{u(N)} \underset{|\eta(N)}{E} \sum_{t=k}^{N} g(u(t), x_t , t)$$

$$= \underset{|\eta(k)}{E} \sum_{t=k}^{N} g(c(\eta(t), t), x_t , t) \tag{3.11}$$

in analogy to (3.5). To obtain this result and a recursive equation for V_k we will use Dynamic Programming and proceed by induction.

We assume that the statement is true for the $N - k$ last steps and we will show that it is also true for the last $N - k + 1$ steps. Consider the situation at time $t = k$. The situation is this: the output signals $y_1 = \eta_1 , ..., y_k = \eta_k$ have been observed, and the control signal $u(k)$ is to be determined. We notice that only the last $N - k$ terms of the cost function are affected by the choice of $u(k)$. The control signal $u(k)$ must therefore be chosen so as to minimize the sum

$$E \sum_{t=k}^{N} g(u(t), x_t , t) \tag{3.12}$$

Due to assumption (3.10) we have

$$\min_{u(k)} ... \min_{u(N)} E \sum_{t=k}^{N} g(u(t), x_t , t)$$

$$= \min_{u(k)} E[g(u k), x_t , k) + \underset{\eta(k+1)}{E} V_{k+1}(w(k + 1))] \tag{3.13}$$

By using hypothesis (3.10), we are thus left with only one minimization. The control variable $u(k)$ thus has to be chosen as a function of $\eta_1, ..., \eta_k$ so as to minimize (3.13). To perform this minimization we rewrite (3.13) in such a way that the dependence of $u(k)$ on $\eta_1, ..., \eta_k$ is explicit. We get from the definition of conditional expectation:

$$\begin{aligned} &E[g(u(k), x_k, k) + \underset{\eta(k+1)}{E} V_{k+1}(w(k+1))] \\ &= \underset{\eta(k)}{E} [\underset{|\eta(k)}{E} g(u(k), x_k, k) + \underset{|\eta(k)}{E} V_{k+1}(w(k+1))] \qquad (3.14) \\ &= \underset{\eta(k)}{E} \left[\int g(u(k), \xi_k, k)\, dF(\xi_k \mid \eta(k)) + \int V_{k+1}(w(k+1))\, dF(\eta_{k+1} \mid \eta(k))\right] \end{aligned}$$

where $F(\xi_k \mid \eta(k))$ and $F(\eta_{k+1} \mid \eta(k))$ are the conditional distribution functions of $x(k)$ and $y(k+1)$, given $\eta(k)$. To evaluate the last integral of (3.14), it is necessary to exhibit explicitly the dependence of $w(k+1)$ on η_{k+1}. To do this we make a digression.

B. A Recursive Equation for the Conditional Distributions

In order to obtain the relation between $w(t)$ and η_t we will express $w(t)$ as a function of $\eta_1, ..., \eta_t$ in terms of a recursive equation. To do this we consider the probability

$$p(\xi_t, \eta_1, ..., \eta_t) = P[x_t = \xi_t, y_1 = \eta_1, ..., y_t = \eta_t] \qquad (3.15)$$

If

$$p(\eta_1, ..., \eta_t) \neq 0$$

it follows from the multiplication rule for conditional expectations that

$$p(\xi_t \mid \eta(t)) = \frac{p(\xi_t, \eta_t \mid \eta(t-1))}{p(\eta_t \mid \eta(t-1))} \qquad (3.16)$$

But

$$\begin{aligned} p(\xi_t, \eta_t \mid \eta(t-1)) &= \sum_{\xi_{t-1}} p(\xi_t, \eta_t, \xi_{t-1} \mid \eta(t-1)) \\ &= \sum_{\xi_{t-1}} p(\xi_t, \eta_t \mid \xi_{t-1}, \eta(t-1)) p(\xi_{t-1} \mid \eta(t-1)) \qquad (3.17) \end{aligned}$$

We have further

$$\begin{aligned} p(\xi_t, \eta_t \mid \xi_{t-1}, \eta(t-1)) &= p(\xi_t \mid \xi_{t-1}, \eta(t-1) p(\eta_t \mid \xi_t, \xi_{t-1}, \eta(t-1)) \\ &= p(\xi_t \mid \xi_{t-1}) p(\eta_t \mid \xi_t) \qquad (3.18) \end{aligned}$$

where the last equality follows from (2.2) and the fact that $x(t, w)$ is a Markov process. We now get from (3.16), (3.17), and (3.18)

$$p(\xi_t \mid \eta(t)) = \frac{\sum\limits_{\xi_{t-1}} p(\eta_t \mid \xi_t) p(\xi_t \mid \xi_{t-1}) p(\xi_{t-1} \mid \eta(t-1))}{\sum\limits_{\xi_t} \sum\limits_{\xi_{t-1}} p(\eta_t \mid \xi_t) p(\xi_t \mid \xi_{t-1}) p(\xi_{t-1} \mid \eta(t-1))} \tag{3.19}$$

Now introducing p, q and $w(t)$ from the equations (2.1), (2.2), and (3.6) we get the following recursive equation for $w_i(t)$.

$$w_i(t+1) = \frac{\sum_s q_{ij} p_{si}(u) w_s(t)}{\sum_s \sum_i q_{ij} p_{si}(u) w_s(t)} \tag{3.20}$$

where

$$\eta_{t+1} = j \tag{3.21}$$

Introduce the notation

$$z_{ij}(u, w(t)) = \sum_s q_{ij} p_{si}(u) w_s(t) \tag{3.22}$$

Notice that z_{ij} are all nonnegative and that a second index of z_{ij} refers to the outcome of the measurement. Introduce the vector

$$z^j = \text{col}\,[z_{ij}, \ldots, z_{nj}] \tag{3.23}$$

and define the norm

$$\| z^j \| = \sum_i | z_{ij} | \tag{3.24}$$

The Equation (3.20) then becomes

$$w(t+1) = \frac{z^i(u, w(t))}{\| z^i(u, w(t)) \|} \tag{3.25}$$

Notice that the norm $\| z^j \|$ has a physical interpretation as the conditional probability

$$\| z^j \| = P[y(t+1) = j \mid y_1 = \eta_1, \ldots, y_t = \eta_t]$$

C. *Results*

Having obtained the desired recursive equation for $w(t)$, we will now return to Eq. (3.14). We get from (3.14) and (3.25)

$$\begin{aligned} &E[g(u(k), x(k), k) + \underset{\eta(k+1)}{E} V_{k+1}(w(k+1))] \\ &= \underset{\eta(k)}{E} \left\{ \sum_i g(u(k), i, k) w_i(k) + \sum_j V_{k+1} \left(\frac{z^j(u, w(t))}{\| z^j(u, w(t)) \|} \right) \| z^j(u, w(t)) \| \right\} \end{aligned} \tag{3.26}$$

Hence

$$\min_{u(k)} \dots \min_{u(N)} E \sum_{t=k} g(u(t), x(t), t) = \mathop{E}_{\eta(k)} [V_k(w(k))] \qquad (3.27)$$

where

$$V_k(w(k)) = \min_u \left\{ \sum_i g(u, i, k) w_i(k) + \sum_j V_{k+1} \left(\frac{z^j(u, w(k))}{\| z^j(u, w(k)) \|} \right) \| z^j(u, w(k)) \| \right\} \qquad (3.28)$$

The minimal cost of the $N - k + 1$ last steps is thus of the form (3.10)· Hence, from the assumption that the minimal cost of the last $N - k$ steps is of the form (3.10), it follows that the minimal cost of the last $N - k + 1$ steps is also of the same form. Further, it was shown in Section III, A that the minimal cost of the last step has the form (3.4). We have thus completed the induction and have achieved the desired result.

Summarizing, we get

THEOREM 1. *Let the control law* $C^0 = \{c^0(w(t), t), t = 1, \dots, N\}$ *minimize the functional* (2.6) *and let*

$$\begin{aligned} V_k(w(k)) &= \min_{u(k)} \mathop{E}_{|\eta(k)} \dots \min_{u(N)} \mathop{E}_{|\eta(N)} \sum_{t=k}^{N} g(u(t), x_t, t) \\ &= \mathop{E}_{|\eta(k)} \sum_{t=k}^{N} g(c^0(w(t), t), x_t, t) \qquad (3.11) \end{aligned}$$

where

$$[w(t)]_i = P[x(t) = i \mid \eta(t)] \qquad (3.6)$$

Then

$$\begin{aligned} V_k(w(k)) &= \min_u \left\{ \sum_i g(u, i, k) w_i(k) + \sum_j V_{k+1} \left(\frac{z^j(u, w(k))}{\| z^j(u, w(k)) \|} \right) \| z^j(u, w(k)) \| \right\} \\ &= \sum_i g(c^0(w(k), k), i, k) w_i(k) \\ &\quad + \sum_j V_{k+1} \left(\frac{z^j(c^0(w(k), k), w(k))}{\| z^j(c^0(w(k), k), w(k)) \|} \right) \cdot \| z^j(c^0(w(k), k), w(k) \| \\ &= \sum_{t=k}^{N} \sum_i g(c^0(w(t), t), i, t) w_i(t) \qquad (3.28) \end{aligned}$$

where

$$[z^j(u, w(t))]_i = \sum_s q_{ij} p_{si}(u) w_s(t) \qquad (3.22)$$

and

$$\| x \| = \sum_i | x_i |$$

We have thus obtained a necessary condition. We will also give a sufficient condition.

THEOREM 2. *Let the functional equation* (3.28) *have a solution* $V_t(w(t))$, *then the problem* $P.1$ *has a solution, the control law* C^0 *minimizes the functional* (2.6), *and the minimal value of* (2.6) *is*

$$\underset{\eta_1}{E}\, V_1(w(1)) \tag{3.29}$$

PROOF: Let $C = \{c(\eta(t), t),\ t = 1, ..., N\}$ be an admissible control law. Introduce

$$W_k(w(k), \eta(k)) = \sum_{t=k}^{N} \sum_i g(c(\eta(t), t), i, t) w_i(t) \tag{3.30}$$

The quantity $W_k(w(k),\ \eta(k))$ is the expected loss over the time interval $[k, N]$ given that the control law C is used and given that at time k $\eta(k)$ is observed. If the control law C is used the expected cost of the last $N - k$ steps is thus

$$\underset{\eta(k)}{E}\, W_k(w(k), \eta(k)) \tag{3.31}$$

and the value of the functional (2.6) is

$$\underset{\eta_1}{E} W_1(w(1), \eta_1)$$

Notice that $w(k)$ is a function of $\eta(k)$; it is, however, advantageous to separate the dependence of W_k on $w(k)$ and $\eta(k)$ as is done in (3.30).

The function $W_k(w(k),\ \eta(k))$ satisfies the equation

$$\begin{aligned} W_k(w(k), \eta(k)) &= \sum_i g(c(\eta(k), k), i, k) w_i(k) \\ &\quad + \sum_j W_{k+1}\left(\frac{z^j(c(\eta(k), k), w(k))}{\| z^j(c(\eta(k), k), w(k))\|}, \eta(k+1)\right) \| z^j(c(\eta(k), k), w(k))\| \end{aligned} \tag{3.32}$$

where

$$\eta_{k+1} = j$$

We will now show that

$$W_t(w(t), \eta(t)) \geqslant V_t(w(t)) \qquad \text{for all } t \tag{3.33}$$

The statement is obviously true for $t = N$. We will now show by induction that it holds for all t.

Assuming that (3.33) is true for $t = k + 1$, we get from (3.28) and (3.33)

$$W_k(w(k)) = \sum_i g(c(\eta(k), k), i, k)w_i(k)$$
$$+ \sum_j W_{k+1}\left(\frac{z^j(c(\eta(k), k), w(k))}{\| z^j(c(\eta(k), k), w(k)) \|}, \eta(k+1)\right) \| z^j(c(\eta(k), k), w(k)) \|$$
$$\geqslant \sum_i g(c(\eta(k), k), i, k)w_i(k)$$
$$+ \sum_j V_{k+1}\left(\frac{z^j(c(\eta(k), k), w(k))}{\| z^j(c(\eta(k), k), wk)) \|}\right) \| z^j(c(\eta(k), k), w(k)) \|$$
$$\geqslant V_k(w(k)) \tag{3.34}$$

where the first inequality follows from the assumption (3.33) with $t = k + 1$, and the second inequality follows from (3.28). We have thus shown that (3.33) with $t = k + 1$ implies (3.34), thereby completing the induction. Now we put $k = 1$ in (3.33) and take mathematical expectation with respect to the distribution of η_1, hence

$$\underset{\eta_1}{E} W(w(1), \eta_1) \geqslant \underset{\eta_1}{E} V(w(1)) \tag{3.35}$$

Further, the continuity of $g(u, x, t)$ and $p_{ij}(u)$ implies that if (3.28) has a solution V, then this solution is continuous in w, which implies that $C = C^0$ gives equality in (3.35). Q.E.D.

IV. Discussion of the Results

We will now draw some conclusions from the results of Section III. The functional equation (3.28) can be solved a priori, knowing only the instantaneous cost function $g(u, x, t)$, the transition matrix P, and the observation matrix Q, and without any knowledge of the actual values of the observed output y. A typical element of the control law expresses the control variable $u(t)$ as a function of the outputs observed up to time t, that is

$$u = u(\eta(t), t) = u(w(t, \eta(t)), t)$$

Notice in particular that for the optimal control law the dependence of u on $\eta(t)$ only enters via the conditional distributions $w(t)$. The function $u = u(w, t)$ is obtained directly from the solution of Eq. (3.28). This function can thus be calculated off-line without any knowledge of the actual output signal.

The function $w = w(\eta(t), t)$ which expresses the conditional distributions of the state $w(t)$ as functions of the measured output signals $\eta(t)$, is given recursively by the equation (3.25). This function must obviously be evaluated in real time, as the outputs are observed.

We now observe

LEMMA 1. *For a control law such that* $u(t) = c(w(t), t)$ *the set of conditional probabilities* $\{w(t); t = 0, 1, 2, ...\}$ *is a Markov process. For fixed* $t \in [0, 1, ...]$ $w(t)$ *takes its values in the positive orthant in* R^n. *The transition probabilities of the w-process are given by*

$$P(y, \Gamma, u) = P[w(t+1) \in \Gamma \mid w(t) = y] = \sum_{k \in K} \| z^k(u, y) \| \qquad (4.1)$$

where $z^k(u, y)$ *is given by Eqs.* (3.22), (3.23) *and*

$$K = \left\{ k; \frac{z^k(u, y)}{\| z^k(u, y) \|} \in \Gamma \right\} \qquad (4.2)$$

Initially, $w(0)$ *equals* p^0 *with probability one.*

PROOF: For the optimal control law $u(t)$ is a function of $w(t)$ and the equation (3.25) gives

$$P[w(t+1) \mid w(t), w(t-1), ..., w(1)] = P[w(t+1) \mid w(t)]$$

which implies that w is a Markov process. The formula (4.1) for the transition probability now follows from (3.25). Q.E.D.

Notice that the transition probability for the w-process has its mass concentrated in m points. Also, notice that as the transition probabilities (2.1) depend on u, the w-process can be influenced by the choice of control variables. We will now consider a variational problem relative to the w-process.

Let $g(u, t)$ denote the vector

$$g(u, t) = \text{col}\, [g(u, i, t), ..., g(u, n, t)] \qquad (4.3)$$

where $g(u, i, t)$ is the instantaneous cost function introduced in Section II. Introduce the functional

$$E \sum_{t=1}^{N} (g(u, t), w(t)) \qquad (4.4)$$

where (a, b) denotes the scalar product of the vectors a and b and E denotes mathematical expectation with respect to the distribution of $w(0), ..., w(N)$.

Now consider the following problem:

P.2 find a sequence of admissible control variables $u(t)$, $t = 1, ..., N$ such that (4.4) is minimal. The value of u at time t may depend on $w(0)$, $w(1), ..., w(t)$.

We have the following result:

THEOREM 3. *The problems P.1 and P.2 are equivalent in the sense that if one of the problems has a solution then the other problem also has a solution. Furthermore, the optimal control law is*

$$u(t) = c^0(w(t), t)$$

in both cases where c^0 is given by Theorem 1.

PROOF: In problem *P*.2 there is complete state information and the solution is thus well-known. See [9]. Assume that *P*.2 has a solution and introduce

$$V_t = \min_u E \left\{ \sum_{i=t}^{N} (g(u, i), w(i)) \mid w(t), w(t-1), ..., w(0) \right\} \tag{4.5}$$

This implies that at each time t the optimal value of the control variable is a function of $w(1), ..., w(t)$. The minimal value of (4.4) is

$$\underset{w(1)}{E} V_1 \tag{4.6}$$

But w is a Markov process, hence

$$V_t = V_t(w(t)) = \min_u E \left\{ \sum_{i=t}^{N} (g(u, i), w(i)) \mid w(t) \right\} \tag{4.7}$$

The Markovian property of w thus implies that the optimal control variable $u(t)$ is a function of $w(t)$ only. Using the standard argument of Dynamic Programming we obtain the following functional equation for $V_t(w(t))$

$$V_t(w(t)) = \min_u \{(g(u, t), w(t)) + E[V_{t+1}(w(t+1)) \mid w(t)]\} \tag{4.8}$$

Hence

$$V_t(w(t)) = \min_u \left\{ (g(u, t), w(t)) + \int V_{t+1}(y) P(w(t), dy, u) \right\} \tag{4.9}$$

where $P(x, \Gamma, u)$ is the transition probability of the Markov porcess The equation (4.1) now implies that Eq. (4.9) is identical to (3.28).Hence , if $P.2$ has a solution then (3.28) holds and Theorem 2 then implies that the problem $P.1$ also has a solution. The reverse statement is proved in the same way, using the equivalent of the Theorem 2 for problem $P.2$. Q.E.D.

Theorem 3 thus implies that the problem of optimal control of a Markov process with incomplete state information can be transformed to a problem of optimal control of a process with complete state information. Notice that the state space of the associated process with complete state information is the space of probability distributions over the states of the original problem. Also, notice that the transition probability distribution of the associated problem has its mass concentrated at most m points, where m is the number of possible outcomes of a single measurement.

The problem of controlling a Markov process with incomplete state information can thus be subdivided into two parts:

1. The solution of the functional equation (3.28), which is equivalent to solving a variational problem for an associated Markov process w with complete state information. This will give $u = u(w, t)$.

2. The calculation of the conditional probability distributions $w(t)$ of the states of the associated Markov process from the measured output signals $\eta(t)$.

This subdivision is a generalization of a well-known theorem for linear systems with a quadratic loss function [3, 4, 6, 9].

In the theory of linear systems with quadratic criteria the states of the associated problem are simply the conditional means of the original states, while in the problem studied in this paper the states are probability distributions on the state space of the original Markov process $\{x_t, t = 0, 1, ...\}$.

The possibility of separating the problem in this way is of great importance for the realization of optimal systems. The fact that the first part of the problem can be solved off-line means a great reduction of the requirements for real time computations.

V. Bounds for Optimal Returns

In this section we will give some bounds on the solution of the functional equation (3.28). We will obtain these bounds by modifying the amount of data which is available for the choice of the control variables. Two particular cases will be considered, namely, the case of complete state information and the case when control is based only on a priori information and no measurements are used (open-loop system, control schedule). The results will enable us to assign a value to the information which is available for making a decision.

A. Complete State Information

Consider the particular case when the measurements are exact, i.e., the measured output y will coincide with the state x with probability one. Hence

$$\underset{\sim}{Q} = I \tag{5.1}$$

Let m denote the measured output at time k and we get

$$w_i(k) = \delta_{im} \tag{5.2}$$

with probability one. Equations (5.1), (5.2), and (3.22) now give

$$z_{ij} = \delta_{ji} p_{mi}$$

$$\| z^j \| = p_{mj}$$

Introducing this into (3.26) we get

$$V_k(w(k)) = \sum_i S_k(i) w_i(k) \tag{5.3}$$

where

$$S_k(m) = \min_u \left[g(u, m, k) + \sum_i S_{k+1}(i) p_{mi}(u) \right] \tag{5.4}$$

In case of perfect state information $V_k(w)$ is thus a linear function of $w(k)$. We also notice that the functional equation (5.4) is the equivalent of the Hamilton-Jacobi equation for the following variational problem. Let $x(t)$ be a Markov process with the transition probability $P(u)$. Find a control $u(t)$ which is a function of $x(t)$ such that the functional

$$E \sum_{t=1}^{N} g(u(t), x_t, t)$$

is minimal.

This is easily verified by applying Dynamic Programming to the problem.

B. Open-Loop System

In this section we will consider the other extreme case, namely, the case when no a posteriori state information is obtainable. We assume

$$q_{ij} = C = \text{constant for all } i \text{ and } j \tag{5.5}$$

Introducing this into (3.22), (3.24), and (3.25) we get

$$\begin{aligned} z_{ij} &= C \cdot \sum_s p_{si}(u) w_s(t) \\ \sum_i z_{ij} &= C \qquad\qquad (5.6) \\ w_i(t+1) &= \sum_s p_{si}(u) w_s(t) \end{aligned}$$

The conditional probabilities are thus independent of the outcome of the measurements, which means that the measurements do not contain any information of use for the calculation of w. Using (5.6), the equation (3.28) reduces to

$$V_k(w(k)) = \min_u \left\{ \sum_i g(u, i, k) w_i(k) + V_{k+1}(w(k)P(u)) \right\} \qquad (5.7)$$

Notice that Eq. (5.7) is the equivalent of the Hamilton-Jacobi equation for the following variational problem.

Consider the difference equation

$$w(t+1) = w(t)P(u) \qquad (5.8)$$

Find an admissible control u which minimizes the functional

$$\sum_{t=1}^{N} \sum_i g(u(t), i, t) w_i(t) \qquad (5.9)$$

Also, notice that as the conditional distributions of the state are independent of the actual observations, the solution of (5.8) will give the cost associated with the best control schedule. Compare Section II.

The functional equations (5.4) and (5.7) are considerably simpler than the equation (3.28). We will now show that the solution of (3.28) is bounded from below by the solution of (5.4) and from above by the solution of (5.7).

Theorem 4. *Let the solution of* (3.28) *by* $V_k(w)$ *and that of* (5.3), (5.4) be $V_k'(w)$ *then*

$$V_k'(w) \leqslant V_k(w) \qquad (5.10)$$

Proof: We will obtain the results by going through the steps of the proof of Theorem 1 and using the following inequality at each step.

$$\min_x \int f(x, y)\, dy \geqslant \int \min_x f(x, y)\, dy$$

Consider Eqs. (3.4) and (3.5). We get

$$V_N(w(N)) = \min_u \underset{|\eta(N)}{E}\, g(u, x_N), N) \geqslant \underset{|\eta(N)}{E} \min_u g(u, x_N, N)$$

$$= \sum_i S_N(i) w_i(N) = V(w(N))$$

We will now show by induction that

$$V_t'(w(t)) \leqslant V_t(w(t)) \qquad \text{for all } t$$

Assuming that the inequality holds for $t = k + 1$ we get

$$V_{t+1}(w(t)) = \min_u \{ \underset{x_t|\eta(t)}{E} g(u, x_t, t) + \underset{w(t+1)|\eta(t)}{E} V_{t+1}(w(t+1)\}$$

$$\geqslant \min_u \{ \underset{x_t|\eta(t)}{E} g(u, x_t, t) + \underset{w(t+1)|\eta(t)}{E} V_{t+1}'(w(t+1))\}$$

$$\geqslant \underset{x_t|\eta(t)}{E} \min_u \left\{ g(u, x_t, t) + \sum_i S_{k+1}(i) p_{x_t i}(u) \right\}$$

$$= \sum S_k(i) w_i(k) = V_k'(w(k))$$

and the theorem now follows by complete induction.

We also have

THEOREM 5. *Let the solution of the functional equation* (3.28) *be* $V_k(w(k))$ *and let that of* (5.7) *be* $V_k''(w)$, *then*

$$V_k(w(k)) \leqslant V_k''(w(k)) \tag{5.11}$$

PROOF: Equation (3.11) gives

$$V_k(w(k)) = \min_{u(k)} \underset{|\eta(k)}{E} \cdots \min_{u(N)} \underset{|\eta(N)}{E} \sum_{t=k}^{N} g(u(t), x(t), t)$$

$$\leqslant \min_{u(k)} \cdots \min_{u(N)} \underset{|\eta(k)}{E} \sum_{t=k}^{N} g(u(t), x(t), t)$$

$$= \min_{u(k)} \cdots \min_{u(N)} \sum_{t=k}^{N} \left(\sum_i g(u(t), i, t)(w(k) P^{t-k}(u))_i \right)$$

$$= V_k''(w(k))$$

Q.E.D.

The Theorems 4 and 5 thus imply that the minimal value of the loss function is bounded by

$$\underset{\eta_1}{E} V_1'(w(1)) \leqslant \underset{\eta_1}{E} V_1(w(1)) \leqslant \underset{\eta_1}{E} V_1''(w(1)) \tag{5.12}$$

where the left-hand member represents the minimal cost in the case of perfect measurements and the right-hand member represents the minimal cost in the case of no measurements at all. The difference

$$\underset{\eta_1}{E}[V_1''(w(1)) - V_1'(w(1))] \tag{5.13}$$

is thus the value of perfect state information, and the difference

$$\underset{\eta_1}{E}[V_1(w(1)) - V_1'(w(1))] \tag{5.14}$$

is the value of incomplete state information.

VI. Examples

In this section we will consider some examples.

Example 1. Let the transition matrix be

$$P = \begin{pmatrix} 1-u & u \\ u & 1-u \end{pmatrix} \tag{6.1}$$

The set of admissible controls is $U = [0, 1]$. Further, let the observation matrix be

$$Q = \begin{pmatrix} q & 1-q \\ 1-q & q \end{pmatrix} \tag{6.2}$$

This means that the probability of getting a correct measurement is q. Further, let the cost functions be

$$g(u(t), x_t, t) = \begin{cases} 1 & x(4) = 2, \quad t = 0 \\ 0 & \text{all other cases} \end{cases} \tag{6.3}$$

This implies that the total cost equals the probability of being in state $i = 2$ at the final step of the four step process.

We get from (3.22)

$$
\begin{aligned}
z_{11} &= \sum_s q_{11} p_{s1} w_s = q[(1 \ - u)w_1 + uw_2] \\
z_{12} &= \sum_s q_{12} p_{s1} w_s = (1 - q)[(1 - u)w_1 + uw_2] \\
z_{21} &= \sum_s q_{21} p_{s2} w_s = (1 - q)[uw_1 + (1 - u)w_2] \\
z_{22} &= \sum_s q_{22} p_{s2} w_s = q[uw_1 + (1 - u)w_2]
\end{aligned} \tag{6.4}
$$

Consider the functional equation (3.28). We get

$$V_4(w) = w_2 \tag{6.5}$$

$$
\begin{aligned}
V_3(w) &= \min_u \left(\frac{z_{12}}{\| z^1 \|} \cdot \| z^1 \| + \frac{z_{22}}{\| z^2 \|} \cdot \| z^2 \| \right) = \min_u (z_{12} + z_{22}) \\
&= \min_u [w_2 + u(w_1 - w_2)] = \min (w_1 , w_2)
\end{aligned} \tag{6.6}
$$

The minimum occurs for the strategy

$$u = \begin{cases} 1 & w_1 < 0.5 \\ 0 & w_1 > 0.5 \end{cases}$$

Now consider the next step. We get from (3.28)

$$
\begin{aligned}
V_2(w) &= \min_u \left\{ \| z^1 \| \min \left(\frac{z_{11}}{\| z^1 \|} , \frac{z_{12}}{\| z^1 \|} \right) + \| z^2 \| \min \left(\frac{z_{12}}{\| z^2 \|} , \frac{z_{22}}{\| z^2 \|} \right) \right\} \\
&= \min_u [\min (z_{11} , z_{21}) + \min (z_{12} , z_{22})]
\end{aligned}
$$

We have four cases

I.	$z_{11} > z_{21}$	$z_{12} > z_{22}$
II.	$z_{11} > z_{21}$	$z_{12} < z_{22}$
III.	$z_{11} < z_{21}$	$z_{12} > z_{22}$
IV.	$z_{11} < z_{21}$	$z_{12} < z_{22}$

We find that the costs and the optimal strategies of the various cases are

$$\begin{array}{lll}
\text{I.} & V = \min(w_1, w_2), & u = \begin{cases} 1 & w_1 < w_2 \\ 0 & w_1 > w_2 \end{cases} \\
\text{II.} & V = 1 - q & u \ \text{arbitrary} \\
\text{III.} & V = q & u \ \text{arbitrary} \\
\text{IV.} & V = \min(w_1, w_2) & u = \begin{cases} 0 & w_1 < w_2 \\ 1 & w_1 > w_2 \end{cases}
\end{array}$$

We find that case II is not possible when $q < 0.5$ and that case III is not possible when $q > 0.5$. The minimal cost is thus

$$V_2(w) = \min(w_1, w_2, q_0) \tag{6.7}$$

where

$$q_0 = \min(q, 1 - q) \tag{6.8}$$

Notice that the strategy yielding the minimal cost is not unique. It is easily verified that either strategy I or II will give the minimal cost. There are also other strategies for which this occurs. For example

$$u = \begin{cases} 1 & 0 \leqslant w_1 < q_0 \\ \alpha & q_0 \leqslant w_1 < 1 - q_0, \qquad 0 \leqslant \alpha \leqslant 1 \\ 0 & 1 - q_0 \leqslant w_1 \leqslant 1 \end{cases}$$

We thus have the equivalents of conjugate points in the classical calculus of variations. Now consider step 1. We get from Eqs. (3.28) and (6.7)

$$V_1 = \min_u \{\min(z_{11}, z_{21}, q_0 \| z^1 \|) + \min(z_{12}, z_{22}, q_0 \| z^2 \|)\}$$

We have now 9 cases

	$\min(z_{11}, z_{21}, q_0\| z^1 \|)$	$\min(z_{12}, z_{22}, q_0\| z^2 \|)$
I.	z_{11}	z_{12}
II.	z_{21}	z_{12}
III.	$q_0\| z^1 \|$	z_{12}
IV.	z_{11}	z_{22}
V.	z_{21}	z_{22}
VI.	$q_0\| z^1 \|$	z_{22}
VII.	z_{11}	$q_0\| z^2 \|$
VIII.	z_{21}	$q_0\| z^2 \|$
IX.	$q_0\| z^1 \|$	$q_0\| z^2 \|$

The minimal costs and the optimal strategies of the various cases are

I. $V_1(w) = \min(w_1, w_2)$

II. $V_1(w) = 1 - q$

III. $$V_1(w) = \begin{cases} \alpha w_1 + \beta w_2 & w_1 < w_2 \\ \alpha w_2 + \beta w_1 & w_1 > w_2 \end{cases}$$

where

$$\alpha = 1 - q + qq_0$$
$$\beta = q_0 - qq_0$$

IV. $V_1(w) = q$

V. $V_1(w) = \min(w_1, w_2)$

VI. $$V_1(w) = \begin{cases} \alpha w_1 + \beta w_2 & w_1 \leqslant w_2 \\ \alpha w_2 + \beta w_1 & w_1 > w_2 \end{cases}$$

VII. $$V_1(w) = \begin{cases} \alpha w_1 + \beta w_2 & w_1 \leqslant w_2 \\ \alpha w_2 + \beta w_1 & w_1 > w_2 \end{cases}$$

where

$$\alpha = q + q_0 - qq_0$$
$$\beta = qq_0$$

VIII. $$V_1(w) = \begin{cases} \alpha w_1 + \beta w_2 & w_1 \leqslant w_2 \\ \alpha w_2 + \beta w_1 & w_1 > w \end{cases}$$

where

$$\alpha = 1 - q + qq_0$$
$$\beta = q_0 - qq_0$$

IX. $V_1(w) = q_0$

We find that if $q > 0.5$ cases IV, VI, and VII are not possible, and similarly if $q < 0.5$, cases II, III, and VIII are not possible. We find that

$$V_1(w) = \min(w_1, w_2, \alpha_0 w_1 + \beta_0 w_2, \alpha_0 w_2 + \beta_0 w_1) \tag{6.9}$$

where

$$\alpha_0 = 2q_0 - q_0^2$$
$$\beta_0 = q_0^2 \tag{6.10}$$

In Fig. 1 the functions $V_1(w)$. $V_2(w)$, and $V_3(w)$ are graphed for $q = 0.8$.

We have thus obtained optimal control strategies, i.e., the optimal values of the decision variables have been expressed as functions of the conditional probabilities of state, $w_i(t)$. To complete the solution it is now necessary to relate the conditional probabilities of the state to the measured outputs of the system. We get from Eq. (3.20)

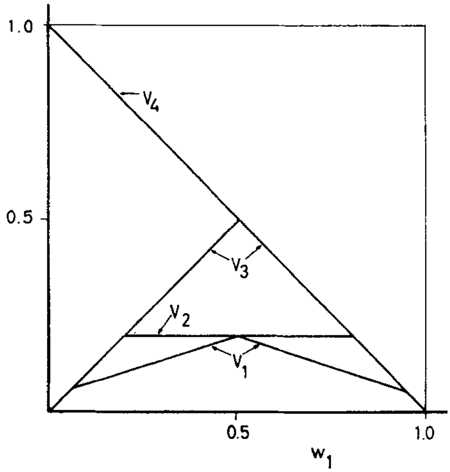

FIG. 1. Optimal returns for the four-stage process of Example 1. $q = 0.8$.

$$w_1(t+1) = \begin{cases} \dfrac{w_1(t)[q - u(t)q] + w_2(t)\cdot q \cdot u(t)}{w_1(t)[q + u(t)(1-2q)] + w_2(t)[1 - q + u(t)(2q-1)]} & \text{if } y(t+1) = 1 \\[2ex] \dfrac{w_1(t)[1 - q + u(t)(q-1)] + w_2(t)u(t)(1-q)}{w_1(t)[1 - u(t) + u(t)(2q-1)] + w_2(t)[q + u(t)(1-2q)]} & \text{if } y(t+1) = 2 \end{cases}$$

$$w_2(t+1) = \begin{cases} \dfrac{w_1(t)[1 - q + u(t)(q-1)] + w_2(t)\cdot u(t)(1-q)}{w_1(t)[q + u(t)(1-2q)] + w_2(t)[1 - q + u(t)(2q-1)]} & \text{if } y(t+1) = 1 \\[2ex] \dfrac{w_1(t)\cdot u(t)\cdot q + w_2(t)[q - qu(t)]}{w_1(t)[1 - u(t) + u(t)(2q-1)] + w_2(t)[q + u(t)(1-2q)]} & \text{if } y(t+1) = 2 \end{cases}$$

which completes the solution of the problem.

It is of interest to compare these results with those obtained in cases of very accurate and very inaccurate measurements. In the case of complete state information we get from Eq. (5.3) and (5.4)

$$\begin{aligned} V_4''(w) &= w_2 \\ V_3'(w) &= V_2'(w) = V_1'(w) = 0 \end{aligned} \tag{6.11}$$

If the measurements are not used at all, that is, a control schedule is used, we get from Eq. (5.7)

$$\begin{aligned} V_4''(w) &= w_2 \\ V_3''(w) &= V_2(w) = V_1(w) = \min (w_1 , w_2) \end{aligned} \tag{6.12}$$

From Theorems 4 and 5 it now follows that

$$0 \leqslant V_i(w) \leqslant \min(w_1 , 1 - w_1) \qquad i = 1, 2, 3 \tag{6.13}$$

An examination of Eqs. (6.5), (6.6), (6.7), (6.9), (6.11), and (6.12) will also give the cost associated with the state information. See also Fig. 1.

Example 2. As a second example we will consider a case where the set of admissible controls is a finite discrete set. Theorems 1 and 2 still hold in this case.

The transition matrix of the problem is given by

u	p_{11}	p_{12}	p_{21}	p_{22}
1	0.5	0.5	0.4	0.6
2	0.5	0.5	0.7	0.3
3	0.8	0.2	0.4	0.6
4	0.8	0.2	0.7	0.3

(6.14)

The instantaneous cost function $g(u, x, t)$ is independent of t and is given by

u x	1	2	3	4
1	20	17	10	7
2	-5	-8	-15	-18

(6.15)

The observation matrix Q is given by

$$Q = \begin{pmatrix} 0.8 & 0.2 \\ 0.3 & 0.7 \end{pmatrix} \tag{6.16}$$

The transition matrix of this example is taken from the toymakers example of Howard [10, p. 28]. Howard uses the two-state Markov process as an idealized model for a manufacturing process. State $x = 1$ is associated with the production of a successful toy and state $x = 2$ is associated with the production of an unsuccessful toy. The four possible decisions represent the following actions:

$u = 1$ no advertising and no research
$u = 2$ no advertising, but research
$u = 3$ advertising, but no research
$u = 4$ advertising and research

The payoff matrix is different from Howards example.

The inclusion of uncertainty in the state information would correspond to the case that the manufacturer does not know whether the toy currently being produced is going to be successful or not. The problem we consider is to maximize the profit over four steps.

$$\max \sum_{t=1}^{4} g(x_t, u(t)) \tag{6.17}$$

Theorems 1 and 2 are easily modified to handle maximization instead of minimization. From (3.5) we get

$$V_4(w) = \max Eg(u, x_4)) = 20w_1(4) - 5w_2(4)$$

We will now proceed recursively and solve Eq. (3.28). We get from (3.22)

u	z_{11}	z_{21}	z_{12}	z_{22}
1	$0.08w_1 + 0.32$	$-0.03w_1 + 0.18$	$0.02w_1 + 0.08$	$-0.07w_1 + 0.42$
2	$-0.16w_1 + 0.56$	$0.06w_1 + 0.09$	$-0.04w_1 + 0.14$	$0.14w_1 + 0.21$
3	$0.32w_1 + 0.32$	$-0.12w_1 + 0.18$	$0.08w_1 + 0.08$	$-0.28w_1 + 0.42$
4	$0.08w_1 + 0.56$	$-0.03w_1 + 0.09$	$0.02w_1 + 0.14$	$-0.07w_1 + 0.21$

Hence, for $t = 3$

u	$\sum_i g(u, i)\, w_i + \sum_j V\left(\frac{z^j}{\|z^j\|}\right) \|z^j\|$
1	$27.5w_1$
2	$20.0w_1 + 4.5$
3	$35.0w_1 - 10.0$
4	$27.5w_1 - 4.5$

and we get

$$V_3(w) = \max\,(27.5w_1\,,\, 20.0w_1 + 4.5) = \begin{cases} 20.0w_1 + 4.5 & w_1 \leqslant 0.6 \\ 27.5w_1 & w_1 > 0.6 \end{cases}$$

Proceeding in the same way we get for $t = 2$

u	$\sum_i g(u, i) + \sum_j V\left(\frac{z^j}{\|z^j\|}\right) \|z^j\|$
1	$27.375w_1 + 7.650$
2	$20.250w_1 + 11.775$
3	$34.500w_1 - 2.350$
4	$27.750w_1 + 1.250$

hence

$$\begin{aligned} V_2(w) &= \max\,(27{,}375w_1 + 7.650,\, 20.250w_1 + 11.775) \\ &= \begin{cases} 20.250w_1 + 11.775 & w_1 \leqslant 0.5789 \\ 27.375w_1 + 7.650 & w_1 > 0.5789 \end{cases} \end{aligned}$$

Similarly, we get for $t = 1$

u	$\sum_i g(u, i)\, w_i + \sum_j V\left(\frac{z^j}{\|z^j\|}\right) \|z^j\|$
1	$27.389w_1 + 15.092$
2	$20.222w_1 + 19.259$
3	$34.555w_1 + 4.092$
4	$27.389w_1 + 9.259$

hence

$$V_1(w) = \max(27.389w_1 + 15.092, 20.222w_1 + 19.259)$$
$$= \begin{cases} 20.222w_1 + 19.259 & w_1 \leqslant 0.5814 \\ 27.389w_1 + 15.092 & w_1 > 0.5814 \end{cases}$$

We now compare the results for the case of incomplete state information with those obtained in the case of perfect state information and in the case of a control schedule.

Let us first consider the case of perfect state information. The cost table at $t = 4$ is

x \ u	1	2	3	4
1	20	17	10	7
2	−5	−8	−15	−18

Hence

$$V_4'(w) = 20w_1 - 5w_2 = 25w_1 - 5$$

At $t = 3$ we get from (4.4)

x \ u	1	2	3	4
1	27.5	24.5	25	22
2	0	4.5	−10	−5.5

Hence

$$V_3'(w) = 27.5w_1 + 4.5w_2 = 23w_1 + 4.5$$

Further, for $t = 2$ we find

x \ u	1	2	3	4
1	36	33	32.9	29.9
2	8.7	12.6	−1.3	2.6

Hence

$$V_2'(w) = 36w_1 + 12.6w_2 = 23.4w_1 + 12.6$$

Finally for $t = 1$ we get

$x \backslash u$	1	2	3	4
1	44.30	41.30	41.32	38.32
2	16.96	20.98	6.96	10.98

Hence

$$V_1'(w) = 44.3w_1 + 20.98w_2 = 23.32w_1 + 20.98$$

Now consider the case of a control schedule. We get

$$V_4''(w) = 20w_1 - 5w_2$$

We get the following cost table for $t = 3$

u	$\sum_i g(u, i)\, w_i + V_{k+1}(wP)$
1	$27.5w_1$
2	$20.0w_1 + 4.5$
3	$35.0w_1 - 10$
4	$27.5w_1 - 4.5$

The equation (4.1) now gives

$$V_3''(w) = \max(27.5w_1, 20.0w_1 + 4.5) = \begin{cases} 20.0w_1 + 4.5 & w_1 \leqslant 0.6 \\ 27.5w_1 & w_1 > 0.6 \end{cases}$$

Similarly, we get for $t = 2$ the following cost table

u	$\sum_i g(u, i)\, w_i + V_{k+1}(wP)$
1	$27w_1 + 7.5$
2	$\max(19.5w_1 + 11.25,\ 21.0w_1 + 10.5)$
3	$\max(33w_1 - 2.5,\ 36w_1 - 4)$
4	$27.75w_1 + 1.25$

Hence

$$V_2''(w) = \max(19.5w_1 + 11.25, 27.0w_1 + 7.5) = \begin{cases} 19.5w_1 + 11.25 & w_1 \leqslant 0.5 \\ 27.0w_1 + 7.5 & w_1 > 0.5 \end{cases}$$

Finally we get for $t = 1$

u	$\sum_i g(u, i)\, w_i + V_{k+1}(wP)$
1	$26.5w_1 + 14.05$
2	$19.6w_1 + 18.4$
3	$\max(32.8w_1 + 4.05, \ 35.8w_1 + 3.30)$
4	$27.7w_1 + 8.4$

Hence

$$V_1''(w) = \max(26.5w_1 + 14.05, 19.6w_1 + 18.4) = \begin{cases} 19.6w_1 + 18.4 & w_1 \leqslant 0.63 \\ 26.5w_1 + 14.05 & w_1 > 0.63 \end{cases}$$

In Fig. 2 we have graphed the optimal value of the cost function for problem 2. The shaded areas in the graph represent the bounds obtained on $V_i(w)$

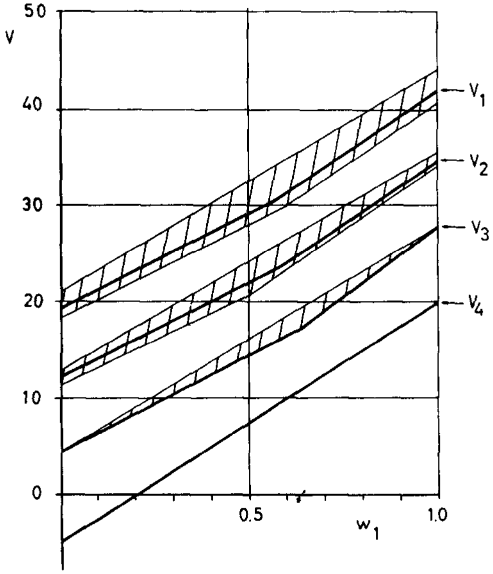

FIG. 2. Optimal returns for the four-stage process of Example 2. The lower limit indicates the maximum return for an open-loop system. The upper limit of the shaded area indicates the optimal return for a system with complete state information. Lines V_1 to V_4 show maximum return for the system with incomplete state information.

from Theorems 4 and 5. The upper limit of the shaded region is thus the maximal gain in the case of complete state information and the lower boundary represents the maximal gain when no measurements are made. The difference in the ordinates of the curves limiting the shaded region will thus represent the value of having complete state information.

VII. Notes

The foundations of the stochastic variational calculus have essentially been laid by Bellman [9, 11, 12], who first developed the basic tool, used in this paper, Dynamic Programming. Bellman has strongly emphasized the use of Markovian models for control problems; this is also done by Feldbaum [13], Florentin [14], Kolmogorov [15], Krassovskii [16], and Pontryagin [2, chap. VII].

The case of complete state information is extensively treated. The case of continuous time continuous state Markov process is discussed by Fleming [17], Florentin [14], Krassovskii [16]. The case of Markov chains with complete state information is treated by Zachrisson [18–20], who considers the game situation. Results on Markov chains are given by Bellman [9] and Howard [10].

Apart from the linear quadratic case [3–6, 16, 21], the case of incomplete state information is not well-known. The Theorems 1 to 5 of this paper are believed to be new. The concept of conditional Markov processes, in particular the equation (3.19) of Section III, B is from Stratonovich [22].

References

1. Bellman, R., Glicksberg, I. and Gross, O. Mathematical Aspects of Control Theory. Report R-313, The RAND Corporation, 1956.
2. Pontryagin, L. S., Boltyansky, V. G., Gamkrelidze, R. V. and Mishchenko, E. F. "The Mathematical Theory of Optimal Process." Interscience, New York, 1962.
3. Gunckel, T. L. and Franklin, G. F. A general solution for linear sampled data control. *ASME J. Basic Eng.* **85**, 197-203 (1963).
4. Joseph, P. D. and Tou, J. T. On linear control theory. *AIEE Trans. (Appl. Ind.)* 193-6 (1961).
5. Tou, J. "Optimum Design of Digital Control Systems." Academic Press, New York, 1963.
6. Åström, K. J., Koepcke, R. W. and Tung, F. On the Control of Linear Discrete Dynamic Systems with Quadratic Loss. Rept. RJ-222, IBM Research Laboratory, San José, California, Sept. 10, 1962.
7. Åström, K. J. Användning av Siffermaskiner för Syntes och Realisering av Regleringssystem (Application of digital computers for the analysis and synthesis

of control systems). Nord SAM, Helsinki, Aug. 1963, to appear in the Congress Proceedings (in Swedish).

8. Carathéodory, C. "Variationsrechnung und Partielle Differentialgleichungen Erster Ordnung." Teubner, Berlin, 1935.
9. Bellman, R. "Adaptive Control Process." Princeton Univ. Press, Princeton, New Jersey, 1961.
10. Howard, R. A. "Dynamic Programming and Markov Process." MIT Tech. Press and Wiley, New York, 1960.
11. Bellman, R. "Dynamic Programming." Princeton Univ. Press, Princeton, New Jersey, 1957.
12. Bellman, R. On the foundations of a theory of stochastic variational problems. *Proc. Symp. Appl. Math. Am. Math. Soc.* **13**, 275-86 (1962).
13. Feldbaum, A. A. On optimal control of Markov objects. *Autom. Remote Control* **23**, 993-1007 (1962).
14. Florentin, J. J. Optimal control of continuous time, Markov stochastic systems. *J. Electron. Control* **10**, 413-88 (1961).
15. Kolmogorov, A. N., Mischenko, E. F. and Pontryagin, L. S. On a probability problem of optimal control. *Sov. Math.* **3**, 1143-5 (1962).
16. Krassovskii, N. N. and Lidskii, E. A. Analytical design of controllers in systems with random attributes, I, II and III. *Autom. Remote Control* **22**, 1021-5 ,1141-6, 1289-94 (1962).
17. Fleming, W. H. Some Markovian optimization problems. *J. Math. Mech.* **12**, 131-40 (1963).
18. Zachrisson, L. E. En Stridsvagnsduell med Spelteoretiska Konsekvenser (A Tank Duel with Game Theoretic Implications). Rept. 2355-2990, Research Inst. Nat. Defence, Stockholm, May 1955 (in Swedish).
19. Zachrisson, L. E. En Stridsvagnsduell med Spelteoretiska Konsekvenser (A Tank Duel with Game Theoretic Implications). *Artilleritidskr.* **84**, 112-21 (1955).
20. Zachrisson, L. E. Markov games. *In* "Contributions to the Theory of Games." Princeton Univ. Press, Princeton, New Jersey, 1964.
21. Wonham, W. M. Stochastic Problems in Optimal Control. Tech. Rept. 63-14, Center for Control Theory RIAS, Baltimore, Maryland, May 1963.
22. Stratonovich, R. L. Conditional Markov process. *Theory Prob. Appl.* **5**, 156-78 (1960).
23. Feller, W. "An Introduction to Probability Theory and Its Applications." Vol. I, 2nd ed. Wiley, New York, 1958.
24. Karlin, S. The mathematical theory of inventory process, *in* Beckenback, ed., "Modern Mathematics for the Engineer," 2nd ser. McGraw-Hill, New York, 1961.
25. Kushner, H. On the Stochastic Maximum Principle: Fixed Time of Control. Tech Rept. 63-24, Center of Control Theory RIAS, Baltimore, Maryland, Dec. 1963.
26. Pugachev, V. S. Statistical methods in automatic control. *Second Intern. Congr. IFAC, Basle, 1963* (Butterworth-Oldenbourg, London, Munich).
27. Stratonovich, R. L. On optimal control theory. Sufficient coordinates. *Autom. Remote Control* **23**, 847-54 (1962).
28. Stratonovich, R. L. Most recent development of dynamic programming techniques and their application to optimal system design. *Second Intern. Congr. IFAC, Basle, 1963* (Butterworth-Oldenbourg, London, Munich).
29. Truxal, J. G. Adaptive control. *Second Intern. Congr. IFAC, Basle, 1963* (Butterworth-Oldenbourg, London, Munich).