Volume Title: Annals of Economic and Social Measurement, Volume
3, number 1

Volume Author/Editor: Sanford V. Berg, editor

Volume Publisher: NBER

Volume URL: http://www.nber.org/books/aesm74-1

Publication Date: 1974

Chapter Title: Adaptive Dual Control Methods

Chapter Author: Edison Tse

Chapter URL: http://www.nber.org/chapters/c9995

Chapter pages in book: (p. 65 - 83)

ADAPTIVE DUAL CONTROL METHODS*

EDISON TSE†

A new approach is discussed for the problem of stochastic control of nonlinear systems with noisy observations. The approach is based on the concept of dual control and the principle of optimality. The resulting control sequence exhibits the closed-loop property, i.e., it anticipates how future learning will be accomplished and how it can be fully utilized. Thus, in addition to being adaptive, this control also plans its future learning according to the control objective. Some simulation results illustrate these properties and demonstrate the computational feasibility of the adaptive dual control algorithm.

1. INTRODUCTION

In many processes arising in engineering, economic and biological systems, the problem of decision making (or control) under various sources of uncertainties is inherent. These uncertainties prevent exact determination of the effect of all present and future actions, and therefore deterministic control theory is not applicable. If the effect of these uncertainties is small, one can still use optimal control theory to obtain a feedback control law based on deterministic considerations. The feedback nature of the control would tend to reduce the sensitivity to uncertainties but would require the state of the system to be measured exactly. Again, this assumption is good only when the measurement error is small in comparison with the signal being measured.

In many cases, the phenomena of uncertainty (including measurement error) can be appropriately modelled as stochastic processes, allowing them to be considered via stochastic optimal control theory. A very important concept in stochastic control is the *information pattern* available to a controller at specific time, for the purpose of decision making. As the process unfolds, additional information becomes available to the controller. This information may come about accidentally through past control actions, or as a result of active probing which itself is a possible control policy. Thus "learning" is present, whether it is "accidental" or "deliberate." The information pattern available to the controller indicates not only what type of learning is possible at each instant of time, but, more importantly, whether future learning can be anticipated and how it could be influenced by present action; i.e., whether probing would be helpful in future learning. Since more learning may improve overall control performance, the probing signal may indirectly help in controlling the stochastic system. On the other hand, excessive probing should not be allowed even though it may promote learning because it is "expensive" in the sense that it will, in general, increase the

expected cost performance of the system. This interplay between learning and control is the key issue of stochastic control theory.

This "dual" purposes of the control was pointed out by Fel'dbaum [1] using the stochastic dynamic programming approach [2]. Unfortunately the solution cannot be obtained numerically in most situations. Some simple examples were worked out later indicating the dual role of the control proper and of probing or identification [3], [4]. Recently, a dual control algorithm was developed by Tse *et al.* [5], [6] which is applicable to a fairly large class of nonlinear stochastic systems. In this paper, the basic concepts involved in the development of the dual control methods as described in [5], [6] are discussed in detail. Hopefully, through this discussion, the interplay between learning and control will be brought out more clearly. Some simulation results as reported in [6] will also be presented to provide deeper understanding of the differences between active and passive learning control strategies.

2. PROBLEM STATEMENT

The class of nonlinear stochastic systems to be considered in this paper is described by

$$(2.1) \qquad \mathbf{x}(k + 1) = \mathbf{f}[k, \mathbf{x}(k), \mathbf{u}(k)] + \xi(k);$$

$$\mathbf{y}(k) = \mathbf{h}[k, \mathbf{x}(k)] + \eta(k), \qquad k = 0, 1, \ldots, N - 1$$

where $\mathbf{x}(k) \in R^n$, $\mathbf{u}(k) \in R^r$, and $\mathbf{y}(k) \in R^m$. It is assumed that $\mathbf{x}(0)$, $\{\xi(k), \eta(k + 1)\}_{k=0}^{N-1}$ are independent Gaussian vectors with statistics:

$$(2.2) \qquad E\{\mathbf{x}(0)\} = \hat{\mathbf{x}}(0|0); \qquad \text{cov}\{\mathbf{x}(0); \mathbf{x}(0)\} = \Sigma(0|0)$$

$$(2.3) \qquad E\{\xi(k)\} = \mathbf{0}; \qquad \text{cov}\{\xi(k); \xi(k)\} = \mathbf{Q}(k) \geq \mathbf{0}$$

$$(2.4)^1 \qquad E\{\eta(k + 1)\} = \mathbf{0}; \qquad \text{cov}\{\eta(k + 1); \eta(k + 1)\} = \mathbf{R}(k + 1) \geq \mathbf{0}.$$

The performance measure is given by

$$(2.5) \qquad J = E\left\{\psi[\mathbf{x}(N)] + \sum_{k=0}^{N-1} \mathscr{L}[\mathbf{x}(k), \mathbf{u}(k), k]\right\}$$

where the expectation $E\{\cdot\}$ is taken over all underlying random quantities. To complete the formulation, one has to specify the class of admissible control laws to be considered. In order to emphasize the interplay between learning and control, we shall distinguish the difference between a *feedback control law* and a *closed-loop control law*. Such a distinction has not been made in the literature; as a matter of fact, their usage has been interchanged quite frequently. However, in order to get further insight into the dual characteristic of the control, such a distinction should be stressed.

In the control engineering literature, a control law is defined as a mapping from the information state (see Section 3 and [8]) space to the control space. Within the class of control laws, we shall make fine distinction between feedback

---

[1] For perfect observation, we have $\mathbf{R}(k + 1) \equiv \mathbf{0}$ for all k. The discussions in this paper *include* this special case which is of interest to economists.

control law and closed-loop control law via the *information pattern* available to the controller at each instant of time. The information pattern indicates what type of knowledge is available to the controller so as to construct the mapping from the information state space to the control space. A feedback law is defined as one in which the structure of mapping is dependent on the system dynamic, the *past* measurement program and the *past* measurement statistics. For the feedback controller, the future observation program and future observation statistics are not available to the controller, and therefore the controller cannot anticipate how future learning will be utilized. Feedback control systems will ignore the possibility of future learning and perform control action in a cautious manner. Thus for a feedback system, learning is "accidental." A closed-loop controller is defined as one in which the structure of the mapping depends on the system dynamic, the *past and future* measurement program as well as the *past and future* measurement statistics. The closed-loop controller can therefore take into account the possibility of future learning and have it regulated according to the control objective. To express these concepts in mathematical terms, let us denote by $\mathscr{D}^k$ the information about the system dynamics

$$\mathscr{D}^k \triangleq \{\mathbf{f}[i, \cdot, \cdot]\}_{i=0}^k$$

$\mathscr{M}^k$ the information about the measurement program up to time k

$$\mathscr{M}^k \triangleq \{\mathbf{h}[i, \cdot]\}_{i=1}^k$$

and $\mathscr{S}^k$ the information about the statistics of the initial state, the process noise up to time $N - 1$ and the observation noise up to time k:

$$\mathscr{S}^k \triangleq \{\hat{\mathbf{x}}(0|0), \boldsymbol{\Sigma}(0|0), \mathbf{Q}(0), \ldots, \mathbf{Q}(N-1), \mathbf{R}(1), \ldots, \mathbf{R}(k)\}$$

and

$$\mathscr{S}^0 \triangleq \{\hat{\mathbf{x}}(0|0), \boldsymbol{\Sigma}(0|0), \mathbf{Q}(0), \ldots, \mathbf{Q}(N-1)\}.$$

A control law is said to be of feedback type if

$$(2.6) \qquad \mathbf{u}^{FB}(k) = \mathbf{u}^{FB}(k, Y^k, U^{k-1}; \mathscr{D}^{N-1}, \mathscr{M}^k, \mathscr{S}^k)$$

where $Y^k \triangleq \{\mathbf{y}(1), \ldots, \mathbf{y}(k)\}$, $U^{k-1} \triangleq \{\mathbf{u}(0)\}, \ldots, \mathbf{u}(k-1)\}$. A control law is said to be of closed-loop type if

$$(2.7) \qquad \mathbf{u}^{CL}(k) = \mathbf{u}^{CL}(k, Y^k, U^{k-1}; \mathscr{D}^{N-1}, \mathscr{M}^{N-1}, \mathscr{S}^{N-1}).$$

From (2.6) and (2.7) it is clear that we have the inclusion relation as described by Figure 2.1.

CLASS OF CONTROL LAWS

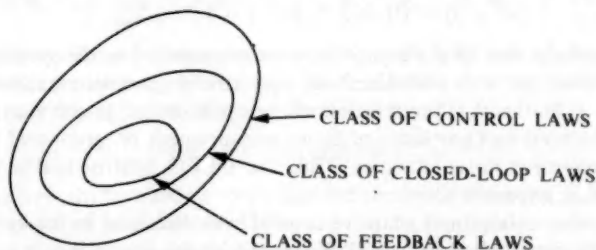CLASS OF CLOSED-LOOP LAWS

CLASS OF FEEDBACK LAWS

Figure 2.1 Inclusion of different control laws

Now a stochastic control problem can be formulated as follows:

Stochastic Control Problem

- Find a closed-loop control law which will minimize the average cost (2.5) subject to the dynamic and observation constraints (2.1).

Before going into the solution for optimal closed-loop control law, let us consider several suboptimal adaptive control laws which are frequently used in the literature; and see which subclass they belong to.

1. *Certainty Equivalence* [7], [31]

$$\mathbf{u}^{CE}(k) = \boldsymbol{\phi}(k, \hat{\mathbf{x}}(k|k))$$

where $\boldsymbol{\phi}(k, \cdot)$ is the optimal control law for the corresponding deterministic control problem (and thus does *not* depend on $\mathscr{S}^{N-1}$ and $\mathscr{M}^{N-1}$), and $\hat{\mathbf{x}}(k|k)$ is the optimum state estimate (and thus depends on $\mathscr{D}^k$, $\mathscr{M}^k$, $\mathscr{S}^k$). It is easily seen that $\mathbf{u}^{CE}(k)$ is within the class of feedback.

2. *Separation* [5], [32]

The control is a function of the conditional mean state estimate

$$\mathbf{u}^s(k) = \boldsymbol{\psi}(k, \hat{\mathbf{x}}(k|k))$$

where $\boldsymbol{\psi}(k, \cdot)$ can be different from the deterministic optimum control law $\boldsymbol{\phi}(k, \cdot)$. If $\boldsymbol{\psi}(k, \cdot)$ is dependent on $\{\mathscr{D}^{N-1}, \mathscr{M}^{N-1}, \mathscr{S}^{N-1}\}$, it is a closed-loop law; otherwise it is a feedback law.

3. *Open-Loop Feedback Optimal (OLFO)* [22], [29], [30]

At any time k, the problem of choosing a *deterministic* sequence

$$\{\mathbf{u}(k), \ldots, \mathbf{u}(N-1)\}$$

so as to minimize the conditional average cost

$$J_k = E\left\{\psi(\mathbf{x}(N)) + \sum_{i=k}^{N-1} \mathscr{L}[\mathbf{x}(i), u(i), i] \,|\, Y^k, U^{k-1}\right\}$$

subject to the dynamic constraint

$$\mathbf{x}(i+1) = \mathbf{f}(i, \mathbf{x}(i)) + \boldsymbol{\xi}(i); \quad i = k, \ldots, N-1$$

is solved; and the first in the control sequence is applied to the system. When a new observation $\mathbf{y}(k+1)$ is obtained, the optimization problem is repeated again at time $k+1$. Notice that the solution of the optimization problem at each time k is not influenced by knowledge of *future* measurement program and associated *future* measurement noise statistics. Thus the OLFO control law is within the class of feedback control laws.

Many other suboptimal adaptive control laws discussed in the literature are within the feedback class [21]–[25]. From the above discussion, we see that a feedback control law does not have the capability of anticipating future uncertainty,

and thus, in general. it may not give satisfactory performance when the time horizon to be considered is relatively short. It is only in some special cases that feedback law give satisfactory or even optimum performance [5]–[11], [28].

In the next section, a method is described for obtaining a closed-loop control law which appropriately regulates learning for the purpose of control.

3. OPTIMAL STOCHASTIC CONTROL

From the above discussion, we note that a closed-loop control law will have the capability of anticipating how future learning will be carried out. *The "best" or optimum closed-loop control law would take into account not only how future observations will be made but also how they will be utilized in an optimum manner.* This is an important aspect of the principle of optimality. Therefore a natural approach to the stochastic control problem is via stochastic dynamic programming. The derivation of the stochastic dynamic programming can be found in many different places [2], [8] and therefore will not be repeated here. What we would like to present in this section are:

1. The basic ingredient of stochastic dynamic programming, and
2. The basic difficulties involved in the solution.

These discussions will not only help us to appreciate the formulation, they also serve as motivation for future development.

There are three basic ingredients in stochastic dynamic programming:

1. The concept of *information state* [8] at time k which is sufficient to represent the past behavior of the system up to time k. This is analogous to the vector state in the deterministic case. We shall denote the information state by $\mathscr{P}_k$. The combined sequence (Y^k, U^{k-1}) can be an information state, and so is the conditional density $p(\mathbf{x}(k)|Y^k, U^{k-1})$.

2. An *optimal-cost-go* associated with each information state at time $k + 1$ which expresses how future observations will be made *and* they will be utilized by the controller in an optimum manner. It will be denoted by $I^*\{\mathscr{P}_{k+1}, k + 1\}$.

3. The conversion of a multistage stochastic optimization problem into a sequence of single stage optimization problems which can be performed sequentially.

The stochastic dynamic programming equation expresses how $I^*\{\mathscr{P}_k, k\}$ can be computed, at least in principle, recursively by

$$(3.1) \quad I^*\{\mathscr{P}_k, k\} = \min_{\mathbf{u}(k)} E\{\mathscr{L}[\mathbf{x}(k), \mathbf{u}(k), k] + I^*\{\mathscr{P}_{k+1}[\mathscr{P}_k, \mathbf{u}(k)], k + 1\}|Y^k, U^{k-1}\},$$

where $\mathbf{u}(k)$ is a *deterministic* quantity and $\mathscr{P}_{k+1}[\mathscr{P}_k, \mathbf{u}(k)]$ represents the evolution of the information state. Notice that from (3.1), the optimum $\mathbf{u}(k)$ will depend, among other things, on $\mathscr{P}_k$. The end condition for $I^*\{\cdot, \cdot\}$ is

$$(3.2) \quad I^*\{\mathscr{P}_N, N\} = E\{\psi(\mathbf{x}(N))|Y^N, U^{N-1}\}.$$

It is quite straight forward to verify that the control law obtained via (3.1)–(3.2) is a *closed-loop* type as defined by (2.7).

Theoretically, the optimal control problem has been solved when equations (3.1) and (3.2) are derived; however, in practice, the problem only begins with

these equations. Some of the major difficulties are discussed in Tse, Bar-Shalom and Meier [5]. In this paper we shall summarize the difficulties as follows:

1. The information state is either infinite dimensional or finite but grows with time.
2. The optimal cost-to-go associated with the information state is generally not an explicit function. In general, the optimal cost-to-go can only be expressed as a table-look-up type function of the information state.
3. Storage of the control value associated with each information state at time k, $k = 0, \ldots, N - 1$ is practically impossible due to the large dimensionality.

There is a very special class of problems, known as the LQG (Linear–Quadratic–Gaussian) [20] problems, in which (3.1), (3.2) can be solved exactly, and the optimal closed-loop control is a feedback law. This is the case when

$$(3.3) \qquad \mathbf{f}(k, \mathbf{x}(k), \mathbf{u}(k)) = \mathbf{A}(k)\mathbf{x}(k) + \mathbf{B}(k)\mathbf{u}(k)$$

$$(3.4) \qquad \mathbf{h}(k, \mathbf{x}(k)) = \mathbf{C}(k)\mathbf{x}(k)$$

$$(3.5) \qquad \boldsymbol{\psi}(\mathbf{x}(N)) = \tfrac{1}{2}\mathbf{x}'(N)\mathbf{F}\mathbf{x}(N)$$

$$(3.6) \qquad \mathscr{L}(\mathbf{x}(k), \mathbf{u}(k), k) = \tfrac{1}{2}[\mathbf{x}'(k)\mathbf{W}(k)\mathbf{x}(k) + \mathbf{u}'(k)\mathbf{N}(k)\mathbf{u}(k)]$$

with $\mathbf{F} \geq \mathbf{0}$, $\mathbf{W}(k) \geq \mathbf{0}$, $\mathbf{N}(k) > \mathbf{0}$. In this case, $\hat{\mathbf{x}}(k|k) \triangleq E\{\mathbf{x}(k)|Y^k, U^{k-1}\}$ is an information state and [21] the optimal cost-to-go has a closed-form expression of the information state.

$$(3.7) \qquad I^*\{\hat{\mathbf{x}}(k|k), k\} = \tfrac{1}{2}\hat{\mathbf{x}}'(k|k)\mathbf{K}(k)\hat{\mathbf{x}}(k|k) + \tfrac{1}{2}\operatorname{tr}\left\{ \sum_{i=k}^{N-1} [\mathbf{W}(i)\boldsymbol{\Sigma}(i|i) \right.$$
$$\left. + (\boldsymbol{\Sigma}(i + 1|i) - \boldsymbol{\Sigma}(i + 1|i + 1))\mathbf{K}(i + 1)] + \mathbf{F}\boldsymbol{\Sigma}(N|N) \right\}$$

where $\mathbf{K}(k)$ satisfies a Riccati equation which can be precomputed once $\mathbf{A}(k)$, $\mathbf{B}(k)$, $\mathbf{W}(k)$, $\mathbf{N}(k)$ and $\mathbf{F}$ are known; $\boldsymbol{\Sigma}(i|i) = \operatorname{cov}\{\mathbf{x}(i)|Y^i, U^{i-1}\}$, $\boldsymbol{\Sigma}(i + 1|i) = \operatorname{cov}\{\mathbf{x}(i + 1)|Y^i, U^{i-1}\}$ which are *independent* of control. Note that the future updated error covariances, which express how future learning will be possible, are included in $I^*\{\hat{\mathbf{x}}(k|k), k\}$. However, since these covariances will not be *influenced* by the control action, only caution but no probing should be exercised by the optimum control. For the particular cost criterion (Quadratic), the optimum control law is a certainty equivalence law [6], [9], [10], [28], which is a feedback law (see Section 2).

4. Dual Control Methods

In this section, we shall describe the dual control methods as developed by Tse *et al.* [5], [6]. The detailed equations can be found in [5], [6] and therefore will not be repeated here. The purpose of this section is to provide a basic understanding of the method.

As was noted before, the solution of (3.1)–(3.2) is practically impossible due to the large dimensionality. Instead of carrying out numerical approximation to the stochastic dynamic programming equation, we shall carry out "conceptual"

approximation to the principle of optimality. The procedures in the approximation are as follows:

1. Approximate the information state by keeping only the first two moments of the state estimate; i.e., consider a close-loop control of the type

(4.1) $\mathbf{u}_a^{CL}(k, \hat{\mathbf{x}}(k|k), \Sigma(k|k); \mathcal{D}^{N-1}, \mathcal{M}^{N-1}, \mathcal{S}^{N-1})$.

The computation of $\hat{\mathbf{x}}(k|k)$, $\Sigma(k|k)$ can be done by any one of the following methods: Extended Kalman filter [12], [13], adaptive filter [13], [14], second order filter [14], [15] and optimum filter [16], [17], [18]. For perfect measurement, we use observer-estimator [14], [26], [27] to obtain the state estimate.

2. Approximate the optimal cost-to-go associated with the approximated "information state," $\{(\hat{\mathbf{x}}(k + 1|k + 1), \Sigma(k + 1|k + 1)\}$, at time $k + 1$. Let us associate with each predicted state $\hat{\mathbf{x}}(k + 1|k)$ a nominal control sequence $U_0[k + 1, N - 1; \hat{\mathbf{x}}(k + 1|k)]$. Usually, this nominal control sequence represents optimum (or near optimum) decision sequence *if* no noise and no state uncertainties are present. Using this nominal control sequence, a nominal state trajectory sequence is also generated via state equation (2.1) *with all the noise terms set to zero*. Perturbation analysis is carried out around these nominals, approximate optimal cost-to-go,

$$I_d^*[\hat{\mathbf{x}}(k + 1|k + 1), \Sigma(k + 1|k + 1), k + 1],$$

that explicitly reflects the future learning and control performance can be obtained. Detail equations for the approximate optimal cost-to-go can be found in [5], we only remark here that $I_d^*[\cdot, \cdot, \cdot]$ is quadratic in $\hat{\mathbf{x}}(k + 1|k + 1)$.

3. At each time $k = 0, 1, \ldots, N - 1$, we shall solve an optimization problem in real-time. Using the concept of principle of optimality, the total cost of applying the control $\mathbf{u}(k)$ is

(4.2) $I_d[\mathbf{u}(k)] = E\{\mathcal{L}(\mathbf{x}(k), \mathbf{u}(k), k)$

$$+ I_d^*[\hat{\mathbf{x}}(k + 1|k + 1; \mathbf{u}(k)), \Sigma(k + 1|k + 1; \mathbf{u}(k)), k + 1]|Y^k, U^{k-1}\}$$

where $\hat{\mathbf{x}}(k + 1|k + 1; \mathbf{u}(k))$, $\Sigma(k + 1|k + 1; \mathbf{u}(k))$ is the updated state estimate and covariance when $\mathbf{u}(k)$ is used. Since $I_d^*[\cdot, \cdot, \cdot]$ is quadratic in $\hat{\mathbf{x}}(k + 1|k + 1)$, the right-hand side of (4.2) can be simplified to have the form

$$I_d[\mathbf{u}(k)] = \text{Tr}\{\mathcal{L}_{xx}[\hat{\mathbf{x}}(k|k), \mathbf{u}(k), k]\Sigma(k|k)\}$$

$$+ \tilde{I}[\hat{\mathbf{x}}(k + 1|k; \mathbf{u}(k)), \Sigma(k + 1|k; \mathbf{u}(k)), k + 1]$$

where $\hat{\mathbf{x}}(k + 1|k; \mathbf{u}(k))$, $\Sigma(k + 1|k; \mathbf{u}(k))$ is the predicted state and covariance when $\mathbf{u}(k)$ is applied (see [5] for the details on $\tilde{I}([\cdot, \cdot, \cdot])$). The optimization problem to be solved at time k is to find $\mathbf{u}(k)$ which will minimize $I_d[\mathbf{u}(k)]$. This is usually accomplished via search methods [5], [6], [19]. Once the minimizing value $\mathbf{u}^*(k)$ is obtained, it is applied to the system. Then $\hat{\mathbf{x}}(k + 1|k + 1)$ and $\Sigma(k + 1|k + 1)$ are updated using one of the estimation

71

methods mentioned above. The whole procedure is repeated for time $k + 1$ and so on until the end of the control period. An outline of the method, in the form of a flow-chart, is given in Figure 4.1. Note that the resulting control law is a closed-loop law.

5. LINEAR SYSTEMS WITH RANDOM PARAMETERS

In this section, we shall describe an explicit dual control algorithm for the class of problems of controlling linear systems with random parameters vector. Consider a discrete-time linear system described by

$(5.1)^2$ $\mathbf{x}(k + 1) = \mathbf{A}[k, \boldsymbol{\theta}(k)]\mathbf{x}(k) + \mathbf{b}[k, \boldsymbol{\theta}(k)]u(k) + \boldsymbol{\xi}(k);$ $k = 0, 1, \ldots$

$\qquad\quad \mathbf{y}(k) = \mathbf{C}[k, \boldsymbol{\theta}(k)]\mathbf{x}(k) + \boldsymbol{\eta}(k);$ $k = 1, 2, \ldots$

where $\mathbf{x}(k) \in R^n$, $\mathbf{y}(k) \in R^m$, $\boldsymbol{\theta}(k) \in R^s$ and $u(k)$ is a scalar control.[3] It is assumed that $\boldsymbol{\theta}(k)$ is a vector Markov process satisfying

(5.2) $\boldsymbol{\theta}(k + 1) = \mathbf{D}(k)\boldsymbol{\theta}(k) + \boldsymbol{\gamma}(k)$ $k = 0, 1, \ldots$

where $\mathbf{D}(k)$ is a known matrix.[4] The vectors $\{\mathbf{x}(0), \boldsymbol{\theta}(0), \boldsymbol{\xi}(k), \boldsymbol{\eta}(k + 1), \boldsymbol{\gamma}(k), k = 0, 1, \ldots\}$ are assumed to be mutually independent Gaussian random variables with known mean and covariances. The cost functional is quadratic in nature

(5.3) $J = \frac{1}{2}E\{[\mathbf{x}(N) - \boldsymbol{\rho}(N)]'\mathbf{W}(N)[\mathbf{x}(N) - \boldsymbol{\rho}(N)]$

$\qquad\qquad + \sum_{k=0}^{N-1} [\mathbf{x}(k) - \boldsymbol{\rho}(k)]' \cdot \mathbf{W}(k)[\mathbf{x}(k) - \boldsymbol{\rho}(k)] + \lambda(k)u^2(k)\}$

where it is assumed that $\mathbf{W}(k) \geq 0$, $\lambda(k) > 0$, and $\{\boldsymbol{\rho}(k), k = 0, \ldots, N\}$ is given *a priori*.

We can transform this problem to the form discussed in the previous sections by augmenting the parameters to form a new stage $\mathbf{z}'(k) \triangleq [\mathbf{x}'(k) : \boldsymbol{\theta}'(k)]$. In here, we shall specify a procedure to choose the nominals that results in an explicit algorithm for the class of problems discussed. The nominal control sequence $U_0[k + 1, N + 1; u(k)]$ is chosen by minimizing

$J_0(k + 1) = \frac{1}{2}[\mathbf{x}_0(N) - \boldsymbol{\rho}(N)]'\mathbf{W}(N)[\mathbf{x}_0(N) - \boldsymbol{\rho}(N)]$

$\qquad\qquad + \frac{1}{2} \sum_{j=k+1}^{N-1} \{[\mathbf{x}_0(j) - \boldsymbol{\rho}(j)]'\mathbf{W}(j)[\mathbf{x}_0(j) - \boldsymbol{\rho}(j)] + \lambda(j)[u_0(j)]^2\}$

subject to the constraints:

$\qquad \mathbf{x}_0(j + 1) = \mathbf{A}[j; \boldsymbol{\theta}_0(j)]\mathbf{x}_0(j) + \mathbf{b}[j; \boldsymbol{\theta}_0(j)]u_0(j); \mathbf{x}_0(k + 1) = \hat{\mathbf{x}}(k + 1|k)$

$\qquad \boldsymbol{\theta}_0(j + 1) = \mathbf{D}(j)\boldsymbol{\theta}_0(j); \boldsymbol{\theta}_0(k + 1) = \hat{\boldsymbol{\theta}}(k + 1|k)$

where $\hat{\mathbf{x}}(k + 1|k)$ is the predicted state if $u(k)$ is applied. Note that $\boldsymbol{\theta}_0(j), j = k + 1, \ldots, N$ can be computed independently of how the control $u_0(j)$ is selected. The

[2] If perfect measurement is available, we have $\mathbf{C}[k, \theta(k)] = \mathbf{I}_n$ and $\boldsymbol{\eta}(k) \equiv \mathbf{0}$, this would imply that the covariance for $\boldsymbol{\eta}(k)$ is zero.

[3] For simplicity, we shall discuss only the scalar input case. The results can be extended to the multi-input case. See Tse and Bar-Shalom [6].

[4] The approach can be extended to the case where $\mathbf{D}$ is a function of $\mathbf{x}$ also.
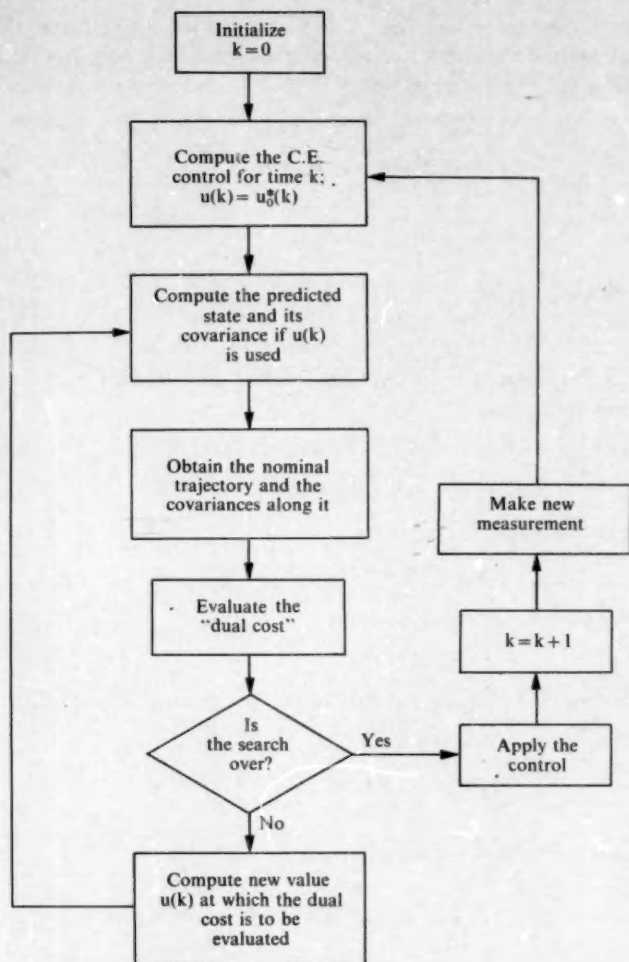
Figure 4.1   Flowchart of the dual control algorithm

solution for this optimization problem can be obtained easily [11]. For the complete set of equations, relevant to a one-step optimization problem, see Tse and Bar-Shalom [6].

6. SIMULATION STUDIES

In this section, an example of controlling a third order time invariant linear system with six unknown parameters will be presented. The performance of the actively adaptive dual control algorithm will be compared to those of the certainty equivalence (C.E.) control and the optimal control with the known parameters. The latter will serve as an unachievable lower bound. In both examples, a second

73

order filter is used for estimation. A discussion of the actively adaptive feature of the dual control algorithm and its computational feasibility is also presented.

Consider the third-order system

(6.1) $$\mathbf{x}(k+1) = \mathbf{A}(\theta_1, \theta_2, \theta_3)\mathbf{x}(k) + \mathbf{B}(\theta_4, \theta_5, \theta_6)u(k) + \boldsymbol{\xi}(k)$$

$$y(k) = [0 \ 0 \ 1]\mathbf{x}(k) + \eta(k)$$

where

(6.2) $$\mathbf{A}(\theta_1, \theta_2, \theta_3) = \begin{bmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ \theta_1 & \theta_2 & \theta_3 \end{bmatrix}: \quad \mathbf{B}(\theta_4, \theta_5, \theta_6) = \begin{bmatrix} \theta_4 \\ \theta_5 \\ \theta_6 \end{bmatrix}$$

and $\{\theta_i\}_{i=1}^6$ are unknown constant parameters with normal *a priori* statistics having mean and variance

$$\hat{\boldsymbol{\theta}}(0|0) = [1.0, -0.6, 0.3, 0.1, 0.7, 1.5]',$$

$$\Sigma^{\theta\theta}(0|0) = \text{diag}(0.1, 0.1, 0.01, 0.01, 0.01, 0.1).$$

The true parameters are

$$\boldsymbol{\theta} = [1.8, -1.01, 0.58, 0.3, 0.5, 1.0]'.$$

The initial state is assumed to be known:

$$\hat{\mathbf{x}}(0|0) = \mathbf{x}(0) = \mathbf{0}.$$

Two examples will be considered. In the first example, the cost performance is expressed by

(6.3) $$J = \tfrac{1}{2}E\left\{[x_3(N) - \rho]^2 + \sum_{i=1}^{N-1} \lambda u^2(i)\right\}.$$

In the second example, the cost performance is given by

(6.4) $$J = \tfrac{1}{2}E\left\{[\mathbf{x}(N) - \boldsymbol{\rho}]'[\mathbf{x}(N) - \boldsymbol{\rho}] + \sum_{i=2}^{N-1} \lambda u^2(i)\right\}$$

where $\lambda = 10^{-3}$, $\rho = 20$ and $\rho' = [0, 0, 20]$.

Twenty Monte Carlo runs were performed for both examples (with the same noise samples) and their average performances are summarized in Tables 6.1 and 6.2). It is noted that in both examples, the dual control algorithm gives a substantial improvement over the C.E. control, both in average performance and reliability. The terminal miss for the dual control is also much better than the C.E. control in both cases.

To understand the interplay between learning and control, and the distinction between active and passive learning, we shall take a closer look at the two examples.

Conceptually, the second example is a "harder" problem than the first example since in the first example, the aim is to "hit" a surface while in the second example, the aim is to "hit" a point in the state space. Therefore, it should be expected that the average cost would be higher in the second example than that

TABLE 6.1
SUMMARY OF RESULTS FOR THE FIRST EXAMPLE

Control Policy	Optimal Control with Known Parameters	C.E. Control with Unknown Parameters	Dual Control with Unknown Parameters
Average cost	6	114	14
Maximum cost in a sample of twenty runs	20	458	53
Standard deviation of the cost	6	140	16
Average miss distance squared	12	225	22
Weighted cumulative control energy prior to final stage	0.1	1.4	3.2

TABLE 6.2

SUMMARY OF RESULTS FOR THE SECOND EXAMPLE

Control Policy	Optimal Control with Known Parameters	C.E. Control with Unknown Parameters	Dual Control with Unknown Parameters
Average cost	15	104	28
Maximum cost in a sample of twenty runs	35	445	62
Standard deviation of the cost	9	114	11
Average miss distance squared	28	192	32
Weighted cumulative control energy prior to final stage	1	7	12

in the first case. This is seen to hold true, as shown in Tables 6.1 and 6.2, for the dual control and the optimal control with known parameters. However, for C.E. control, it does not hold true. This may seem strange, but careful analysis of the simulation will offer an explanation for this.

Let us compare the C.E. controls for the two examples. Note that the control energy used in the second case is much more than that used in the first example. Note from Figures 6.3 and 6.6 that up to about $k = 12$, the C.E. control uses about the same cumulative energy for the two examples. The fact that the final mission is different has not yet become important enough to change the control strategy. As a consequence, the learning for both cases is almost the same up to this time. In the first example, since the final destination is a surface, the controller can wait almost until the final time to apply a control to achieve the control objective, and therefore the C.E. control is still applying little energy after time
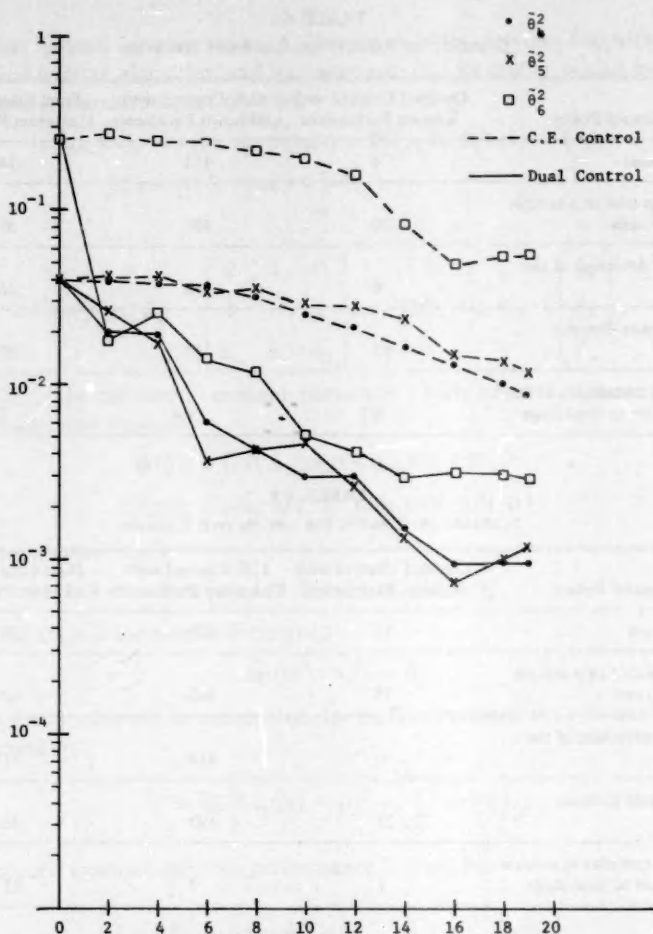
Figure 6.1 Average estimation error squared in $\theta_4, \theta_5, \theta_6$ for the first example

twelve. The learning of the parameters θ_4, θ_5, and θ_6 is only slightly improved. However, for the second example, since the final destination is a point in the state space, the control must work "harder" to achieve its objective (transferring from one point to another arbitrary point requires three time units). Therefore, the control energy after time twelve increases very quickly for the second example. This results in a much better estimation of the gain parameters. Since the learning in the first example is poorer than in the second example for the C.E. control, a higher cost is accrued in the first example than in the second. Note that even though the second example is a "harder" problem, a better performance value is obtained. This is primarily because "accidental" learning is enhanced by the difficulty of achieving the final mission.
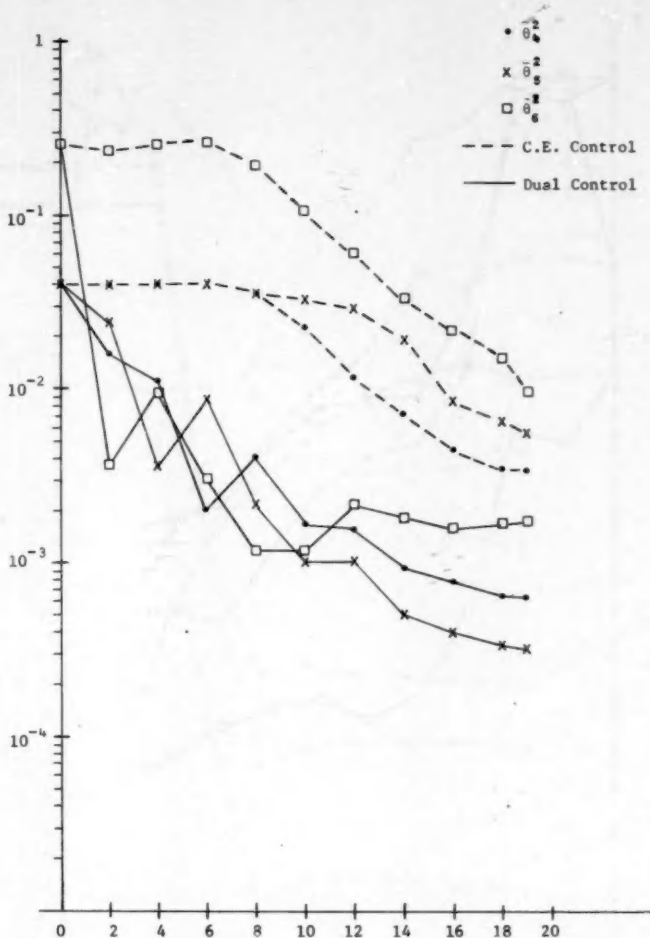
Figure 6.2   Average estimation error squared in θ_4, θ_5, θ_6 for the second example

For the dual control, quite a different control strategy at the beginning rather than at the end of the control interval can be noticed. The fact that a different end condition has to be fulfilled is propagated from the final time to the initial time. For the second example, the dual controller, realizing that the final mission is much more difficult to achieve, decides to invest more energy in the beginning, because learning is very important in this case to achieve a satisfactory final objective. Note the "speed" of learning in the second example compared with the first example (see Figures 6.1, 6.2, 6.4, 6.5). The dual control regulates its energy in learning: in the first example where learning is less important, it does not insist on learning which would involve the application of large controls in the beginning; in the second example, the learning is much more important and thus
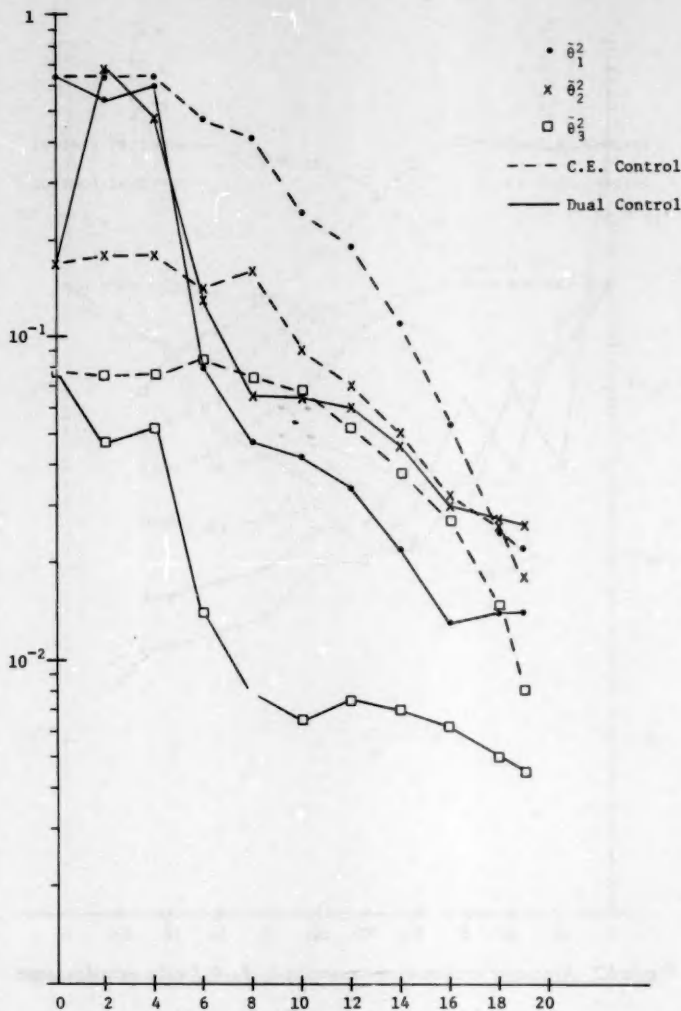
Figure 6.3   Average estimation error squared in θ_1, θ_2, θ_3 for the first example

more energy is utilized for the learning purpose. For both examples, the expected miss distances squared are comparable, thus, the increase in cost in the second example is primarily due to the increase in cumulative input energy. This demonstrates the active learning characteristic of the dual control.

Finally, we shall remark on the computation time required by the dual control and compare it with that for C.E. control to give some idea of the computational feasibility of the dual control algorithm. The optimum control with known parameters took 3 sec on an UNIVAC 1108 while the C.E. required 6 sec for one
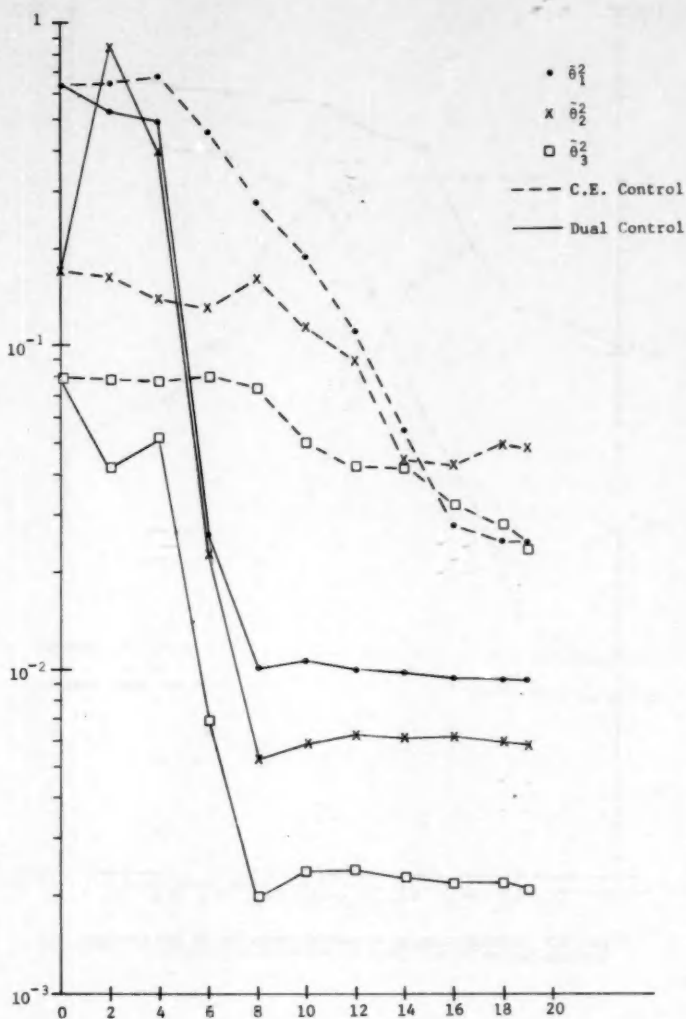
Figure 6.4   Average estimation error squared in θ_1, θ_2, θ_3 for the second example

run. The time required for the dual control was 45 sec (with a program that was not optimized). However, judging from the improvement over the C.E. control, the extra computation time seems worthwhile.

7. CONCLUDING REMARKS

This paper describes an approach for obtaining a control algorithm that exhibits the dual characteristic of appropriately distributing the control energy for
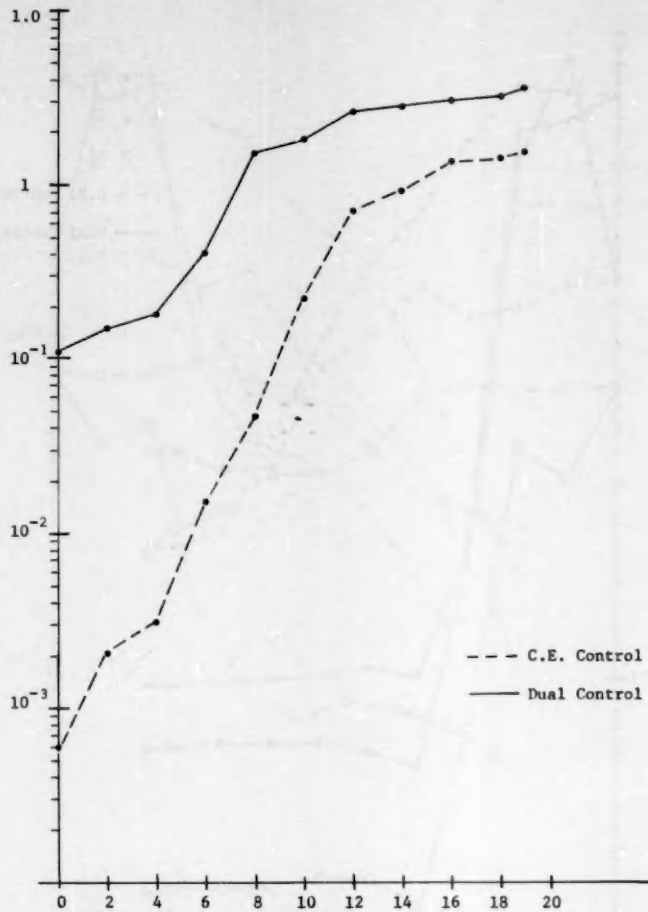
79

Figure 6.5   Average cumulative control energy for the first example

learning and control purposes. The approach is an approximation based on the principle of optimality that retains the closed-loop feature of the control. An adaptive dual control is described that possesses the distinguishing characteristic of regulating its learning as required by the control objective. Such an "active" learning feature is not present in most of the feedback control methods reported in the literature [21]–[25]. For those classes of problems where the interplay between learning and control is crucial for obtaining good system performance, the dual control method described in this paper can be expected to provide a better performance than the "passive" feedback control methods. One such class of problems is described in Sections 4 and 5.
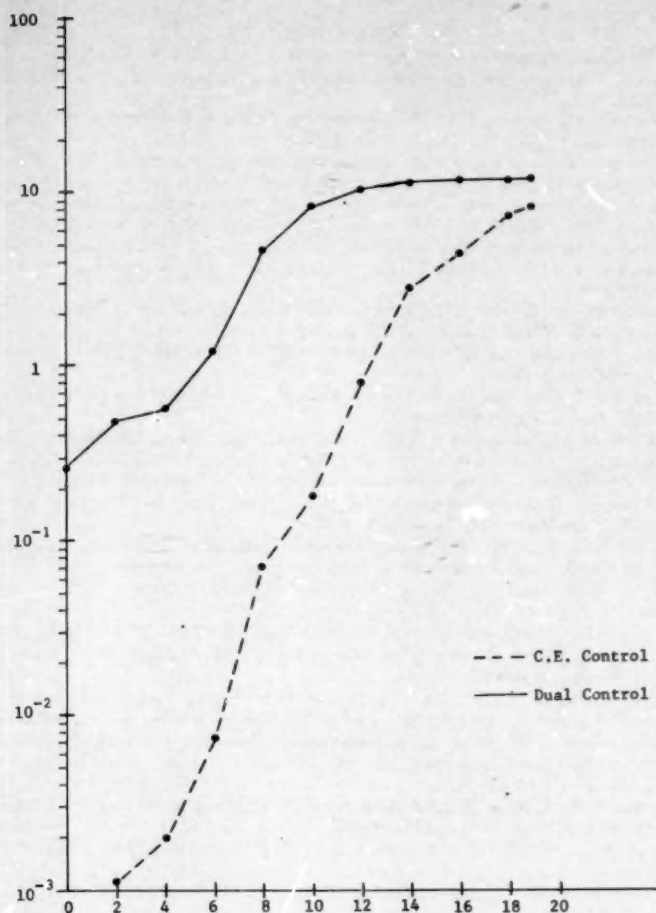
Figure 6.6   Average cumulative control energy for the second example

The applicability of the dual control algorithm is particularly well suited to problems when the physical sampling interval is on the order of hours and days (e.g., problems in economics). On the other hand, for those problems where "real time" is on the order of micro-seconds or seconds, more work on reducing the computational requirement of the dual control algorithm is needed. The potential of dual control is so promising that it is felt that further work should be continued.

Systems Control, Inc.
Palo Alto, California

REFERENCES

[1] A. A. Fel'dbaum, *Optimal Control Systems*, Academic Press, New York, 1965.

[2] R. Bellman, *Adaptive Control Processes: A Guided Tour*, Princeton University Press, Princeton, New Jersey, 1961.

[3] J. J. Florentin, "Optimal, Probing, Adaptive Control of a Simple Bayesian System," *J. Electronics and Control*, Ser. 1, 1962, Vol. 13, No. 2, pp. 165–177.

[4] K. J. Åstrom, "Optimal Control of Markov Processes with Incomplete State Information," *Journal of Mathematical Analysis and Applications*, Vol. 10, No. 1, February 1965.

[5] E. Tse, Y. Bar-Shalom, and L. Meier, "Wide-Sense Adaptive Dual Control for Nonlinear Stochastic Systems," *IEEE Trans. on Automatic Control*, April 1972, pp. 98–108.

[6] E. Tse and Y. Bar-Shalom, "An Actively Adaptive Control for Linear Systems with Random Parameters via the Dual Control Approach," *IEEE Trans. on Automatic Control*, April 1972, pp. 109–116.

[7] Y. Bar-Shalom and E. Tse, "Dual Effect and Certainty Equivalence in Stochastic Control," submitted to *SIAM Control*, also to appear in *Preprints JACC*, 1974.

[8] L. Meier, "Combined Optimal Control and Estimation," *Proc. Third Allerton Conference on Systems and Circuits*, 1965.

[9] P. Joseph and J. Tou, "On Linear Control Theory," *AIEE Trans. Applications and Industry*, Vol. 80, pp. 193–196, September 1961.

[10] L. Meier, R. E. Larson, and A. J. Tether, "Dynamic Programming for Stochastic Control of Discrete Systems," *IEEE Trans. on Automatic Control*, December 1971, pp. 767–775.

[11] M. Aoki, *Optimization of Stochastic Systems*, Academic Press, New York, 1967.

[12] H. W. Sorenson, "Kalman Filtering Techniques," in *Advances in Control Systems*, Vol. 3, C. T. Leondes, Ed., Academic Press, New York, 1966.

[13] A. Jazwinski, *Stochastic Processes and Filtering Theory*, Academic Press, New York, 1970.

[14] E. Tse, R. Dressler, and Y. Bar-Shalom, "Application of Adaptive Tuning of Filters to Exo-atmospheric Target Tracking," *Proc. 3rd Symposium on Nonlinear Estimation Theory*, San Diego, California, 1972.

[15] M. Athans, R. P. Wishner, and A. Bertolini, "Suboptimal State Estimation for Continuous-Time Nonlinear Systems from Discrete Noisy Measurements," *IEEE Trans. on Automatic Control*, October 1968, pp. 505–514.

[16] R. S. Bucy and K. D. Senne, "Realization of Optimum Discrete-Time Nonlinear Estimators," *Proc. Symp. Nonlinear Estimation Theory and Its Applications*, San Diego, California, 1970.

[17] D. L. Alspach and H. W. Sorenson, "Approximation of Density Function by a Sum of Gaussian for Nonlinear Bayesian Estimation," *Proc. Symp. Nonlinear Estimation Theory and Its Applications*, San Diego, California, 1970.

[18] E. Tse and R. E. Larson, "Parallel Algorithms for Optimum Nonlinear State Estimation," *Preprints JACC*, Columbus, Ohio, June 1973.

[19] D. G. Luenberger, *Introduction to Linear and Nonlinear Programming*, Addison-Wesley, Reading, Mass., 1973.

[20] Special Issue on Linear-Quadratic-Gaussian Problems, *IEEE Trans. Aut. Control*, Vol. AC-16, December 1971.

[21] E. Tse, "On the Optimal Control of Linear Systems with Incomplete Information," Electron. Syst. Lab., MIT, Cambridge, Rep. ESL-R-412, January 1970.

[22] E. Tse and M. Athans, "Adaptive Stochastic Control for a Class of Linear Systems," *IEEE Trans. Automatic Control*, Vol. AC-17, pp. 38–52, February 1972.

[23] D. G. Lainiotis, T. N. Upadhyay, and J. G. Deshpande, "Optimal Adaptive Control of Linear Systems," *Proc. 1971 IEEE Decision and Control*, Miami Beach, Florida, December 1971.

[24] G. Stein and G. N. Saridis, "A Parameter-Adaptive Control Technique," *Automatica*, Vol. 5, pp. 731–740, November 1969.

[25] R. Ku and M. Athans, "On the Adaptive Control of Linear Systems Using the Open-Loop-Feedback-Optimal Approach," *Proceedings of the 1972 IEEE Decision and Control*, New Orleans, Louisiana, December 1972.

[26] E. Tse and M. Athans, "Optimal Minimal-order Observer-Estimators for Discrete Linear Time-Varying Systems," *IEEE Trans. Automatic Control*, Vol. AC-15, pp. 416–426, August 1970.

[27] E. Tse, "Observer-Estimators for Discrete-Time Systems," *IEEE Trans. on Automatic Control*, Vol. AC-18, pp. 10–16, February 1973.

[28] M. Athans, "The Discrete Time Linear-Quadratic Gaussian Stochastic Control Problems," *Annals of Economic and Social Measurement*, Vol. I, No. 4, 1972.

[29] S. E. Dreyfus, "Some Types of Optimal Control of Stochastic Systems," *SIAM Journal of Control*, Vol. 2, pp. 120–134, 1964.

[30] Y. Bar-Shalom and R. Sivan, "The Optimal Control of Discrete Time Systems with Random Parameters," *IEEE Trans. Auto. Control*, Vol. AC-14, pp. 3–8, 1969.

[31] H. Theil, "A Note on Certainty Equivalence in Dynamic Planning," *Econometrica*, Vol. 25, pp. 346–349, 1957.

[32] W. M. Wonham, "On the Separation Theorem of Stochastic Control," *SIAM J. Control*, Vol. 6, No. 2, pp. 312–326, 1968.