An Actively Adaptive Control for Linear Systems with Random Parameters via the Dual Control Approach

EDISON TSE AND YAAKOV BAR-SHALOM

Abstract—A new method is presented for controlling a discrete-time linear system with possibly time-varying random parameters in the presence of input and output noise. The cost is assumed to be quadratic in the state and control.

Previous algorithms for the above problem when the system had both zeros and poles unknown were of the open-loop feedback type, i.e., they did not take into account that future observations will be made. Therefore, even though these schemes were adaptive, their learning was "accidental." In contrast to this, the new approach uses an expression of the optimal cost-to-go that exhibits the dual purpose of the control, i.e., learning and control.

The effect of the present control on the future estimation ("learning") appears explicitly in the cost used in the stochastic dynamic programming equation. The resulting sequence of controls, which is of the closed-loop type, is shown via simulations to appropriately divide its energy between the learning and the control purposes. Therefore, this control is called actively adaptive because it regulates the speed and amount of learning as required by the performance index. The simulations on a third-order system with six unknown parameters also demonstrate the computational feasibility of the proposed algorithm.

I. Introduction

THE CONTROL of linear systems with unknown parameters is a problem of major theoretical and practical importance. The development of adaptive control for this class of problems has been an area of extensive research. Conceptually, the problem can be solved exactly if one can solve the stochastic dynamic programming equation associated with the problem [4]; unfortunately, a numerical solution for this is prohibited by the curse of dimensionality. Thus, different approaches have been suggested in treating this class of problems.

One widely used approach is via the certainty equivalence [1], [9], [14]. The advantage of this approach is the simplicity of the control law; however, it ignores the confidence level of the parameter estimates in deriving the adaptive control scheme; one would expect that such a control scheme would result in a control system that is extremely sensitive to stochastic variations.

If the design of adaptive systems takes not only the instantaneous parameter estimates but also the associated confidence levels into account, it would surely result in a "better" system. One such method is the open-loop feedback approach [8], [5], [18], [7], [1], [12]. In this open-loop feedback approach, the fact that the estimated parameter may not be exact is therefore taken into consideration, but the knowledge of the *future observation program* is completely ignored.

According to the theory of dual control, introduced by Feldbaum [9], the control signal has two purposes which might be in conflict with each other: 1) to help learn about any unknown parameters and/or state of the system (estimation); 2) to control. In view of this, one can see that the open-loop feedback control is, from the estimation point of view, passive, since it does not take into account that learning is possible in the future. In contrast to this, a dual control is active, not only for the control purpose but also for the estimation purpose because the performance depends also on the "quality" of the estimates. Therefore, the dual control can be called *actively adaptive* since it regulates its adaptation (learning) in an optimal manner.

The objective of this paper is to obtain a control strategy for linear systems with unknown parameters with quadratic cost, based on the dual nature of the control. In the paper by Tse, Bar-Shalom, and Meier [20] a wide-sense adaptive dual control formulation for a general nonlinear discrete-time system has been obtained. This formulation leads to a control that has the closed-loop property of being a function of the past observations and the *future observation program and the associated statistics*. Note the difference between this and the controls that are a function of only the past observations, the latter being open-loop feedback-type controls.

In this paper these results will be applied to obtain an actively adaptive control for linear systems with unknown parameters. A specific algorithm is derived which seems to be appropriate in terms of computational feasibility for this class of problems. Some nontrivial examples are presented that provide deeper understanding of the differences between active and passive learning control strategies.

The structure of the paper is as follows. In Section II, the control problem is stated. In Sections III and IV, the algorithm that yields an actively adaptive control sequence is described in detail. The simulation results presented in Section V demonstrate the computational feasibility and performance level of the new algorithm and also provide

Manuscript received July 1, 1972; revised October 31, 1972. Paper recommended by G. Saridis, Chairman of the IEEE S-CS Adaptive and Learning Systems, Pattern Recognition Committee. This work was supported by the Air Force Office of Scientific Research under Contract F44620-71-C-0077.

The authors are with Systems Control, Inc., Palo Alto, Calif. 94306.

more insight into the actively adaptive feature of the control.

II. Problem Statement

Consider a discrete-time linear system described by

$$x(k+1) = A[k,\theta(k)]x(k) + b[k,\theta(k)]u(k) + \xi(k), \qquad k = 0,1,\cdots$$
$$y(k) = C[k,\theta(k)]x(k) + n(k), \qquad k = 1,2,\cdots \tag{2.1}$$

where $x(k) \in R^n$, $y(k) \in R^m$, $\theta(k) \in R^s$, and $u(k)$ is a scalar control.[1] It is assumed that $\theta(k)$ is a Markov process satisfying

$$\theta(k+1) = D(k)\theta(k) + \gamma(k), \qquad k = 0,1,\cdots \tag{2.2}$$

where $D(k)$ is a known matrix.[2] The vectors $\{x(0),\theta(0),\xi(k),n(k+1),\gamma(k),\ k = 0,1,\cdots\}$ are assumed to be mutually independent Gaussian random variables with known statistical laws

$$x(0) \sim \mathcal{G}[\hat{x}(0),\Sigma^{xx}(0)]; \qquad \theta(0) \sim \mathcal{G}[\hat{\theta}(0),\Sigma^{\theta\theta}(0)];$$
$$\xi(k) \sim \mathcal{G}[0,Q(k)]; \qquad n(k) \sim \mathcal{G}[0,R(k)];$$
$$\gamma(k) \sim \mathcal{G}[0,G(k)] \tag{2.3}$$

with $\Sigma^{xx}(0) > 0$, $\Sigma^{\theta\theta}(0) > 0$, $R(k) > 0$, $Q(k) \geq 0$, $G(k) \geq 0$. The notation $v \sim \mathcal{G}(a,B)$ is used to denote that the random vector v is Gaussian with mean a and covariance B. Furthermore, we assume that the unknown parameter $\theta(k)$ enters linearly in $A(k,\cdot)$, $b(k,\cdot)$, and $C(k,\cdot)$.

A control is admissible if it is nonanticipative; i.e.,

$$u(k) = u(k,Y^k,U^{k-1}); \qquad Y^k \triangleq \{y(1),\cdots,y(k)\};$$
$$U^{k-1} \triangleq \{u(1),\cdots,u(k-1)\}. \tag{2.4}$$

Our objective is to find an admissible control sequence U^{N-1} such that the cost functional

$$J = \tfrac{1}{2}E\Big\{[x(N) - \varrho(N)]'W(N)[x(N) - \varrho(N)] + \sum_{k=0}^{N-1} [x(k) - \varrho(k)]'W(k)[x(k) - \varrho(k)] + \lambda(k)u^2(k)\Big\} \tag{2.5}$$

is minimized subject to the dynamic constraints (2.1) and (2.2). The expectation in (2.5) is over all the underlying random quantities $x(0)$, $\theta(0)$, $\{\xi(k),n(k+1),\gamma(k),\ k = 0,1,\cdots,N-1\}$. It will be assumed that $W(k) \geq 0$, $\lambda(k) > 0$, and $\{\varrho(k),\ k = 0,1,\cdots,N\}$ is given *a priori*.

III. The Approximate Expected Optimal Cost-To-Go and the Dual Effect

In this section, we shall briefly describe the wide-sense adaptive dual control approach introduced earlier [20]. Then we shall obtain all the equations, relevant to a one-step optimization problem, that must be solved every time a new observation is obtained.

Let the present time be denoted by k. Consider an arbitrary control $u(k)$ applied at time k, and the resulting predicted augmented state and covariance will be $\hat{z}(k+1|k) \triangleq [\hat{x}'(k+1|k),\hat{\theta}'(k+1|k)]'$ and $\Sigma(k+1|k)$, respectively.[3] Associated with $u(k)$ is a future (fictitious) nominal control sequence $U_o[k+1,N-1;u(k)]$. The choice of the nominal is quite flexible in the approach described in [20]; in this paper 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 $U_o[k+1,N+1;u(k)]$ is chosen by minimizing

$$J_o(k+1) = \tfrac{1}{2}[x_o(N) - \varrho(N)]'W(N)[x_o(N) - \varrho(N)] + \tfrac{1}{2}\sum_{j=k+1}^{N-1}\{[x_o(j) - \varrho(j)]'W(j)[x_o(j) - \varrho(j)] + \lambda(j)[u_o(j)]^2\} \tag{3.1}$$

subject to the constraints

$$x_o(j+1) = A[j;\theta_o(j)]x_o(j) + b[j;\theta_o(j)]u_o(j); \qquad x_o(k+1) = \hat{x}(k+1|k) \tag{3.2}$$

$$\theta_o(j+1) = D(j)\theta_o(j); \qquad \theta_o(k+1) = \hat{\theta}(k+1|k) \tag{3.3}$$

where $\hat{x}(k+1|k)$ is the predicted state if $u(k)$ is applied. Note that $\theta_o(j)$, $j = k+1,\cdots,N$, can be computed independently of how the control $u_o(j)$ is selected. The solution for this optimization problem can be obtained easily [1]. The optimal control $u_o{}^*(j)$ is given by

$$u_o{}^*(j) = -\mu_o(j)b_o'(j)[\tilde{K}_o(j+1)A_o(j)x_o(j) + \tilde{p}_o(j+1)] \tag{3.4}$$

where

$$\mu_o(j) \triangleq [\lambda(j) + b_o'(j)\tilde{K}_o(j+1)b_o(j)]^{-1} \tag{3.5}$$

and $\tilde{K}_o(j+1),\tilde{p}_o(j+1)$ satisfy[4]

$$\tilde{K}_o(j) = A_o'(j)[I - \mu_o(j)\tilde{K}_o(j+1)b_o(j)b_o'(j)] \cdot \tilde{K}_o(j+1)A_o(j) + W(j); \qquad \tilde{K}_o(N) = W(N) \tag{3.6}$$

$$\tilde{p}_o(j) = A_o'(j)[I - \mu_o(j)\tilde{K}_o(j+1)b_o(j)b_o'(j)]\tilde{p}_o(j+1) - W(j)\varrho(j); \qquad \tilde{p}_o(N) = -W(N)\varrho(n). \tag{3.7}$$

The sole purpose of this is to obtain an approximate value of the optimal cost-to-go associated with $\{\hat{z}(k+1|k+1),\Sigma(k+1|k+1)\}$, the "information state" at the next stage. A second-order perturbation analysis is carried out about this nominal trajectory and control. In this manner, an approximate optimal cost-to-go, $I_d{}^*[\hat{z}(k+1|k+1),\Sigma(k+1|k+1),k+1]$, that explicitly reflects the future learning and control performance is obtained using the results of [20]. Therefore, the "dual

[1] For simplicity, we shall discuss only the scalar input case. The results can be extended to the multi-input case. See Section IV.

[2] The approach can be extended to the case where D is a function of x also.

[3] These are functions of $u(k)$, but, to avoid further complication of the notations, $u(k)$ does not appear as argument.

[4] Here the tilde denotes quantities related to the certainty equivalence control, which determines the nominal trajectory from $k+1$ to N.

cost" of applying $u(k)$ is given by

$$J_d[u(k)] = \tfrac{1}{2}\lambda(k)u^2(k) + E\{I_d^*[\hat{z}(k+1|k+1), \boldsymbol{\Sigma}(k+1|k+1), k+1] | Y^k\}. \quad (3.8)$$

The optimization problem is reduced to that of finding the $u^*(k)$ that minimizes $J_d(\cdot)$, as given by (3.8). After $u^*(k)$ is applied to the system and a new observation $\boldsymbol{y}(k+1)$ is obtained, one updates the estimate of $\boldsymbol{z}(k+1)$ and its error covariance. The real-time updating can be done by using extended Kalman filter [16], [11], second-order filter [2], or optimal filters [6], [19], [3]. Then, with this as a starting point, the same procedure is repeated to obtain $u^*(k+1)$.

In the following, the procedure to compute the dual cost $J_d[u(k)]$ is described. The derivation is straightforward and will be omitted because of lack of space. The basic steps are as follows.

Step 1: Consider the augmented system

$$\boldsymbol{z}_o(j+1) \triangleq \begin{bmatrix} \boldsymbol{x}_o(j+1) \\ \boldsymbol{\theta}_o(j+1) \end{bmatrix} = \boldsymbol{f}_o(j) \triangleq \begin{bmatrix} \boldsymbol{f}_o^x(j) \\ \boldsymbol{f}_o^\theta(j) \end{bmatrix}$$
$$\triangleq \begin{bmatrix} \boldsymbol{A}_o(j)\boldsymbol{x}_o(j) + \boldsymbol{b}_o(j)u_o(j) \\ \boldsymbol{D}(j)\boldsymbol{\theta}_o(j) \end{bmatrix}, \qquad j = k+1,\cdots,N-1 \quad (3.9)$$

where superscripts denote matrix partitions and

$$\boldsymbol{A}_o(j) \triangleq \boldsymbol{A}[j,\boldsymbol{\theta}_o(j)] \qquad \boldsymbol{b}_o(j) \triangleq \boldsymbol{b}[j,\boldsymbol{\theta}_o(j)] \quad (3.10)$$

and with measurement vector

$$\boldsymbol{h}(j) = [\boldsymbol{C}(j,\boldsymbol{\theta}(j))|\boldsymbol{0}]\boldsymbol{z}(j). \quad (3.11)$$

This, with the nominal controls obtained in (3.4), defines the nominal trajectory associated with $u(k)$.

Step 2: Then apply the results from [20] to this augmented system; the future estimation is assumed to be done by a linearized filter about the nominal trajectory. After some algebraic manipulations, we have,

$$\begin{aligned} J_d[u(k)] = {} & \tfrac{1}{2}\lambda(k)u^2(k) + \tfrac{1}{2}\hat{\boldsymbol{x}}'(k+1|k)\tilde{\boldsymbol{K}}_o(k+1) \\ & \cdot \hat{\boldsymbol{x}}(k+1|k) + \tilde{\boldsymbol{p}}_o'(k+1)\hat{\boldsymbol{x}}(k+1|k) \\ & + \tfrac{1}{2}\,\mathrm{tr}\Bigg\{\sum_{j=k+1}^{N} \mathcal{W}(j)\boldsymbol{\Sigma}_o(j|j) + [\boldsymbol{\Sigma}(k+1|k) \\ & - \boldsymbol{\Sigma}_o(k+1|k+1)]\boldsymbol{K}_o(k+1) \\ & + \sum_{j=k+1}^{N-1} [\boldsymbol{\Sigma}_o(j+1|j) \\ & - \boldsymbol{\Sigma}_o(j+1|j+1)]\boldsymbol{K}_o(j+1)\Bigg\}. \end{aligned} \quad (3.12)$$

The matrices $\boldsymbol{K}_o(j+1)$, $\mathcal{W}(j)$ are given by

$$\boldsymbol{K}_o(j) \triangleq \begin{bmatrix} \boldsymbol{K}_o^{xx}(j) & \boldsymbol{K}_o^{\theta x\prime}(j) \\ \boldsymbol{K}_o^{\theta x}(j) & \boldsymbol{K}_o^{\theta\theta}(j) \end{bmatrix} \quad (3.13)$$

$$\boldsymbol{K}_o^{xx}(j) = \tilde{\boldsymbol{K}}_o(j) \quad (3.14)$$

$$\begin{aligned} \boldsymbol{K}_o^{\theta x}(j) = {} & [\boldsymbol{f}_{o,\boldsymbol{\theta}}^{x\prime}(j)\boldsymbol{K}_o^{xx}(j+1) + \boldsymbol{D}'(j)\boldsymbol{K}_o^{\theta x}(j+1)]\boldsymbol{A}_o(j) \\ & - \mu_o(j)\Bigg\{[\boldsymbol{f}_{o,\boldsymbol{\theta}}^{x\prime}(j)\boldsymbol{K}_o^{xx}(j+1) + \boldsymbol{D}'(j)\boldsymbol{K}_o^{\theta x} \\ & \cdot (j+1)]\boldsymbol{b}_o(j) + \left[\sum_{i=1}^{n} \boldsymbol{e}_i'\boldsymbol{p}_0^x(j+1)b_{\boldsymbol{\theta}}^i(j)\right]'\Bigg\} \\ & \cdot\{\boldsymbol{b}_o'(j)\boldsymbol{K}_o^{xx}(j+1)\boldsymbol{A}_o(j)\}; \qquad \boldsymbol{K}_o^{\theta x}(N) = \boldsymbol{0} \end{aligned} \quad (3.15)$$

$$\begin{aligned} \boldsymbol{K}_o^{\theta\theta}(j) = {} & \boldsymbol{f}_{o,\boldsymbol{\theta}}^{x\prime}(j)\boldsymbol{K}_o^{xx}(j+1)\boldsymbol{f}_{o,\boldsymbol{\theta}}^{x}(j) \\ & + \boldsymbol{D}'(j)\boldsymbol{K}_o^{\theta x}(j+1)\boldsymbol{f}_{o,\boldsymbol{\theta}}^{x}(j) \\ & + \boldsymbol{f}_{o,\boldsymbol{\theta}}^{x\prime}(j)\boldsymbol{K}_o^{x\theta}(j+1)\boldsymbol{D}(j) + \boldsymbol{D}'(j)\boldsymbol{K}_o^{\theta\theta}(j+1) \\ & \cdot\boldsymbol{D}(j) - \mu_o(j)\Bigg\{\boldsymbol{b}'(j)[\boldsymbol{K}_o^{xx}(j+1)\boldsymbol{f}_{o,\boldsymbol{\theta}}^{x}(j) \\ & + \boldsymbol{K}_o^{x\theta}(j+1)\boldsymbol{D}(j)] + \sum_{i=1}^{n} \boldsymbol{e}_i'\boldsymbol{p}_o^x(j+1)b_{\boldsymbol{\theta}}^i(j)\Bigg\}' \\ & \cdot\Bigg\{\boldsymbol{b}'(j)[\boldsymbol{K}_o^{xx}(j+1)\,\boldsymbol{f}_{o,\boldsymbol{\theta}}^{x}(j) + \boldsymbol{K}_o^{x\theta}(j+1)\boldsymbol{D}(j)] \\ & + \sum_{i=1}^{n} \boldsymbol{e}_i'\boldsymbol{p}_0^x(j+1)\boldsymbol{b}_{\boldsymbol{\theta}}^i(j)\Bigg\}; \boldsymbol{K}_o^{\theta\theta}(N) = 0 \end{aligned} \quad (3.16)$$

$$\mathcal{W}(j) \triangleq \left[\begin{array}{c:c} \boldsymbol{W}(j) & \displaystyle\sum_{i=1}^{n} \boldsymbol{e}_i'\boldsymbol{p}_o^x(j+1)\boldsymbol{a}_{\boldsymbol{\theta}}^i(j) \\ \hdashline \displaystyle\sum_{i=1}^{n} \boldsymbol{e}_i'\boldsymbol{p}_o^x(j+1)\boldsymbol{a}_{\boldsymbol{\theta}}^{i\prime}(j) & \boldsymbol{0} \end{array}\right] \quad (3.17)$$

$$\boldsymbol{p}_o^x(j) = \tilde{\boldsymbol{K}}_o(j)\boldsymbol{x}_o(j) + \tilde{\boldsymbol{p}}_o(j) \quad (3.18)$$

where $\boldsymbol{x}_o(j)$ is the future nominal state obtained by applying the future nominal control sequence $U_o[k+1,N;u(k)]$ obtained from (3.4); $\boldsymbol{\Sigma}_o(j|j)$, $\boldsymbol{\Sigma}_o(j+1|j)$ are the updated and predicted error covariances of the augmented state, respectively. They are obtained by the extended Kalman filter equation linearized about the future nominal control and trajectory, initializing at $j = k$ with $\boldsymbol{\Sigma}_o(k+1|k) = \boldsymbol{\Sigma}(k+1|k)$, the prediction covariance obtained after applying $u(k)$ to the system. The notation $\boldsymbol{f}_{o,\boldsymbol{z}}$ stands for the Jacobian of $\boldsymbol{f}$ with respect to $\boldsymbol{z}$ evaluated along the nominal.

In the next section an actively adaptive control algorithm, which is the main result of this paper, will be described in detail.

IV. Actively Adaptive Control Algorithm

Let the present time be denoted by k. We shall assume that the updated state and its covariance are obtainable through a real-time estimation algorithm. The following procedure is carried out to obtain $u^*(k)$.

A. Initialization of the Search

Generate $\boldsymbol{\theta}_o(j)$, $j \geq k$, via

$$\boldsymbol{\theta}_o(j+1) = \boldsymbol{D}(j)\boldsymbol{\theta}_o(j); \qquad \boldsymbol{\theta}_o(k) = \hat{\boldsymbol{\theta}}(k|k), \quad (4.1)$$

and store these values. Compute $\tilde{\boldsymbol{K}}_o(j)$, $\tilde{\boldsymbol{p}}_o(j)$, $j = k+1$,

$\cdots,N$, using (3.6) and (3.7), and store these values. Note that these equations are a function of $\hat{\boldsymbol{\theta}}(k|k)$ only. Set

$$u(k) = u_o^*(k) \tag{4.2}$$

as given by (3.4) with $j = k$.

B. Evaluation of $J_d[u(k)]$

1) Extrapolation:

$$\hat{\boldsymbol{x}}(k+1|k) = \boldsymbol{A}[k;\hat{\boldsymbol{\theta}}(k|k)]\,\hat{\boldsymbol{x}}(k|k) + \boldsymbol{b}[k;\hat{\boldsymbol{\theta}}(k|k)]u(k) + \tfrac{1}{2}\sum_{i=1}^{n} \boldsymbol{e}_i \operatorname{tr}\{f_{xx}{}^i[\hat{\boldsymbol{x}}(k|k),u(k)]\boldsymbol{\Sigma}^{xx}(k|k)\} \tag{4.3}$$

$$\boldsymbol{\Sigma}_o(k+1|k) = \boldsymbol{f}_z(k)\boldsymbol{\Sigma}(k|k)\boldsymbol{f}_z{}'(k) + \mathcal{Q}(k) + \tfrac{1}{2}\sum_{i=1}^{n+s}\sum_{j=1}^{n+s} \boldsymbol{e}_i\boldsymbol{e}_j{}' \operatorname{tr}[f_{zz}{}^i(k)\boldsymbol{\Sigma}(k|k)f_{zz}{}^j(k)\boldsymbol{\Sigma}(k|k)] \tag{4.4}$$

where $f_{zz}{}^i$ is the Hessian of the ith component of $\boldsymbol{f}$ with respect to $\boldsymbol{z}$ and

$$\mathcal{Q}(j) = \begin{bmatrix} Q(j) & 0 \\ 0 & G(j) \end{bmatrix}. \tag{4.5}$$

Note that $\hat{\boldsymbol{\theta}}(k+1|k)$ is already available.

2) Computation of the (Fictitious) Future Nominal: Generate $\{\boldsymbol{x}_o(j)\}_{j=k+1}^{N}$ and $\{u_o(j)\}_{j=k+1}^{N-1}$ using (3.2) where

$$u_o(j) \triangleq u_o^*(j), \qquad j = k+1,\cdots,N-1, \tag{4.6}$$

as given by (3.4)

3) Computation of the Dual Cost:

a) Compute $\boldsymbol{K}_o{}^{\theta x}(j)$ and $\boldsymbol{K}_o{}^{\theta\theta}(j)$, $j = k+1,\cdots,N$, by (3.15) and (3.16). These are backward equations.

b) Form the matrix $\boldsymbol{K}_o(j)$, $k+1,\cdots,N$, using (3.13) and (3.14).

c) Compute $\boldsymbol{\Sigma}_o(j+1|j+1)$, $j = k,\cdots,N-1$, $\boldsymbol{\Sigma}_o(j+1|j)$, $j = k+1,\cdots,N-1$, using the extended Kalman filter covariance equations along the nominal.

d) Obtain the dual cost by (3.12).

C. The Search

After having evaluated J_d at the value of the certainty equivalence control, one proceeds with a line search (e.g., of Fibonacci type combined with quadratic interpolation [13]) along the $u(k)$-axis to obtain $u^*(k)$, which will minimize $J_d(\cdot)$. Although such a procedure does not necessarily yield the absolute minimum of the function $J_d(\cdot)$, it will, however, result in a value $u^*(k)$ that is at least as good as the first value at which the evaluation was made.

D. Control and Observation

Apply $u^*(k)$ and observe $\boldsymbol{y}(k+1)$. Use the real-time estimation algorithm to update the state and its covariance. Repeat procedures A–C with $k+1 \rightarrow k$ until $k = N-1$.

An outline of the algorithm, in the form of a flow chart, is given in Fig. 1.

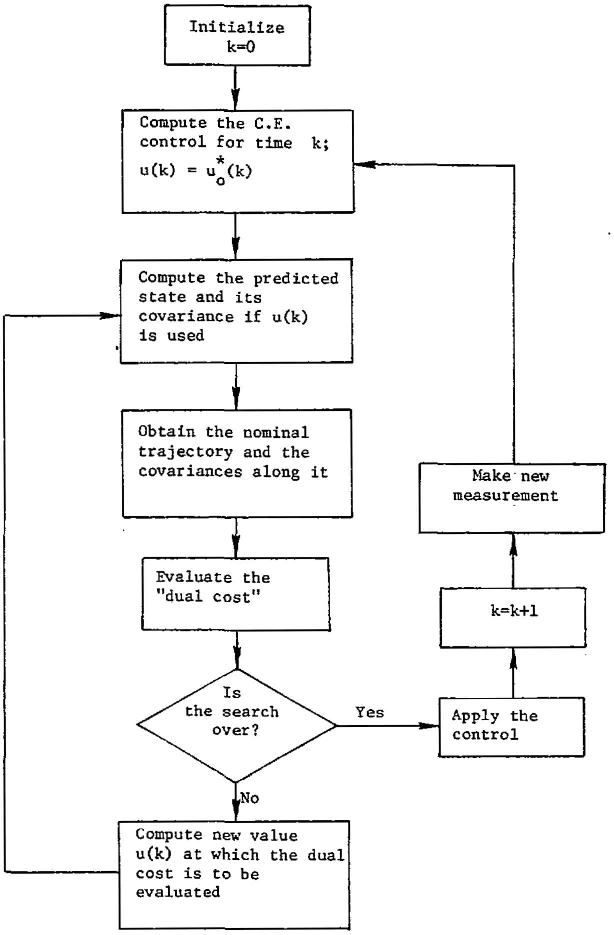
Initialize
k=0
Compute the C.E. control for time k;
u(k) = u*o(k)
Compute the predicted state and its covariance if u(k) is used
Obtain the nominal trajectory and the covariances along it
Evaluate the "dual cost"
Is the search over?
Yes
No
Compute new value u(k) at which the dual cost is to be evaluated
Apply the control
k=k+1
Make new measurement

Fig. 1. Flow chart of the dual control algorithm.

E. Remarks

1) The dual property of the control is revealed in (3.12), where the cost to be minimized includes both control and estimation cost.

2) If we have a high confidence on the parameter estimate, i.e., $\boldsymbol{\Sigma}^{\theta\theta} = \boldsymbol{0}$, then it can be easily seen from (3.12) that the dual effect disappears and we can assume that separation holds.

3) Note how the dual cost reflects also the effect of the future observation program. For example, if it is known *a priori* that during the interval $l < j \leq N$, $l \geq k$, no observations will be made, then we would have $\boldsymbol{\Sigma}_o(j|j-1) = \boldsymbol{\Sigma}_o(j|j)$. In this case, (3.12) becomes

$$\begin{aligned} J_d[u(k)] = {} & \tfrac{1}{2}\lambda(k)u^2(k) + \tfrac{1}{2}\hat{\boldsymbol{x}}'(k+1|k)\tilde{\boldsymbol{K}}_o(k+1) \\ & \cdot\hat{\boldsymbol{x}}(k+1|k) + \tilde{\boldsymbol{p}}_o{}'(k+1)\hat{\boldsymbol{x}}(k+1|k) \\ & + \tfrac{1}{2}\operatorname{tr}\Bigg\{\sum_{j=k+1}^{l-1} \mathcal{W}(j)\boldsymbol{\Sigma}_o(j|j) \\ & + \sum_{j=l}^{N} \mathcal{W}(j)\boldsymbol{\Sigma}_o(j|l) + \boldsymbol{K}_o(k+1)[\boldsymbol{\Sigma}(k+1|k) \\ & - \boldsymbol{\Sigma}_o(k+1|k+1)] + \sum_{j=k+1}^{l} \boldsymbol{K}_o(j+1) \\ & [\boldsymbol{\Sigma}_o(j+1|j) - \boldsymbol{\Sigma}_o(j+1|j+1)]\Bigg\}. \end{aligned} \tag{4.7}$$

Therefore, the knowledge that future observations will or will not be taken would change the present control strategy. If future learning will not take place, the present control tries to minimize the average control performance, whereas, if future observation will take place, the present control will invest some of its energy to help the future learning (see [20] for an illustrative example). It is in this way that the dual control regulates its future learning under some control objective. Because of this "active learning" characteristic we call this control strategy an *actively adaptive control.*

4) The estimation cost in (3.12) is also a function of time-to-go. In the beginning of the control interval, the estimation cost is relatively high. The dual control must therefore be selected so that it compromises between the control and estimation purposes. When k is approaching $N - 1$, the estimation cost becomes smaller, and thus the dual control will give less weight to the estimation part and will finally concentrate on the control purpose.

5) For the case where the control $\boldsymbol{u}(k)$ is a vector rather than a scalar value, one can obtain similar equations as above, except that now $\mu_o(j)$ are matrices. In the vector control case, the search for the actively adaptive control is more complicated since one has to search over a volume rather than a line.

V. Simulation Studies

In this section, two examples of controlling a third-order 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 control and the optimal control with the known parameters. The latter will serve as an unachievable lower bound. In both examples, a second-order filter is used for real-time estimation. A discussion of the actively adaptive feature of the control algorithm and its computational feasibility is also presented.

Consider the third-order system

$$\begin{aligned} \boldsymbol{x}(k+1) &= \boldsymbol{A}(\theta_1,\theta_2,\theta_3)\boldsymbol{x}(k) + \boldsymbol{B}(\theta_4,\theta_5,\theta_6)u(k) + \boldsymbol{\xi}(k) \\ y(k) &= [0 \quad 0 \quad 1]\boldsymbol{x}(k) + \eta(k) \end{aligned} \tag{5.1}$$

where

$$\boldsymbol{A}(\theta_1,\theta_2,\theta_3) = \begin{bmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ \theta_1 & \theta_2 & \theta_3 \end{bmatrix}; \qquad \boldsymbol{B}(\theta_4,\theta_5,\theta_6) = \begin{bmatrix} \theta_4 \\ \theta_5 \\ \theta_6 \end{bmatrix}; \tag{5.2}$$

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

$$\hat{\boldsymbol{\theta}}(0|0) = [1, -0.6, 0.3, 0.1, 0.7, 1.5]' \tag{5.3}$$

$$\boldsymbol{\Sigma}^{\theta\theta}(0|0) = \text{diag } (0.1, 0.1, 0.01, 0.01, 0.01, 0.1). \tag{5.4}$$

The true parameters are

$$\boldsymbol{\theta} = [1.8, -1.01, 0.58, 0.3, 0.5, 1]'. \tag{5.5}$$

The initial state is assumed to be known as

$$\hat{\boldsymbol{x}}(0|0) = \boldsymbol{x}(0) = \boldsymbol{0}. \tag{5.6}$$

A. Interception-Type Example

In this case the objective is to bring the third component of the state to a desired value. This is expressed by the cost

$$J = \tfrac{1}{2}E\left\{[x_3(N) - \rho]^2 + \sum_{i=0}^{N-1} \lambda u^2(i)\right\} \tag{5.7}$$

where ρ is some value and λ is chosen to be small. In this example, $\rho = 20$ and λ is chosen to be 10^{-3}. The noises $\{\xi_i(k)\}_{i=1}^{3}$ and $\eta(k+1)$ are assumed to be independent and are normally distributed with zero mean and unit variance. If we interpret x_3 as the position of an object, then this example corresponds to an interception problem, i.e., the guidance of an object to reach a certain point, without constraints on the velocity and acceleration of the object when it reaches that point. The difficulty lies in the fact that the *poles and zeros* of the system are both unknown. The initial condition (5.6) represents the fact that the system is initially at rest.

Twenty Monte Carlo runs were performed on the interception example and average performances are summarized in Table I and in Figs. 2–4. As shown in Table I, the dual control performance is an order of magnitude better than the certainty equivalence (CE) control. The second and third rows indicate that the dual control performance is much more predictable than the CE control. Note that the dual control uses only about twice the energy of the CE control, at the same time achieving a dramatic improvement in the miss distance squared over the CE control. This indicates that the dual control does use control energy at appropriate times to improve learning, thus achieving satisfactorily the control objective.

As seen in Fig. 4, the dual control invested considerable energy in learning at the beginning. The effect of this is revealed in Figs. 2 and 3, where the average error squared for the parameters' estimates is displayed.

Note that the CE control provides fairly good learning in θ_1, θ_2, and θ_3, but practically no learning in θ_4, θ_5, and θ_6. The learning in θ_1, θ_2, and θ_3 is mainly due to the process noise, which serves as a random input that excites the modes of the system. Thus, in this case, the learning of θ_1, θ_2, and θ_3 is accidental; also, because this learning is too slow, it is of little use in achieving the control objective.

B. Soft Landing-Type Example

In this case, instead of bringing only the third component of the state to a desired value, the objective is to bring the final state to a certain point in the state space. This is expressed by

$$J = \tfrac{1}{2}E\left\{[\boldsymbol{x}(N) - \boldsymbol{\varrho}]'[\boldsymbol{x}(N) - \boldsymbol{\varrho}] + \sum_{i=0}^{N-1} \lambda u^2(i)\right\} \tag{5.8}$$

where $\boldsymbol{\varrho}$ is a point in R^3 and λ is as before. This may be interpreted as a soft landing problem by selecting the terminal desired state to be

TABLE I
SUMMARY OF RESULTS FOR THE INTERCEPTION EXAMPLE

Control Policy	Optimal Control with Known Parameters	CE Control with Unknown Parameters	Dual Control with Unknown Parameters
Average cost	6	114	14
Maximum cost in a sample of 20 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

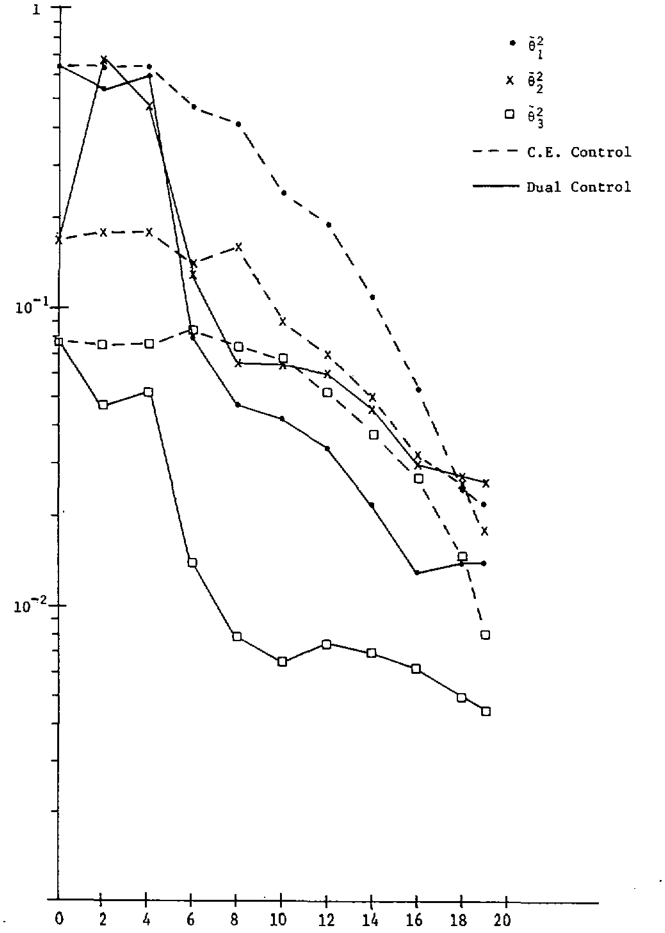

Fig. 2. Average estimation error squared in θ_1, θ_2, θ_3 for the interception example.

$$\varrho = \begin{bmatrix} 0 \\ 0 \\ 20 \end{bmatrix}. \tag{5.9}$$

Comparing the results of this problem to those obtained in Section V-A will provide more insight into the dual nature of the control. Twenty Monte Carlo runs were carried out for the CE control, the dual control, and the optimal control with known parameters. The results are summarized in Table II and Figs. 5-7.

Conceptually, the soft landing is a "harder" problem than the intercept problem. Here, the aim is to "hit" a

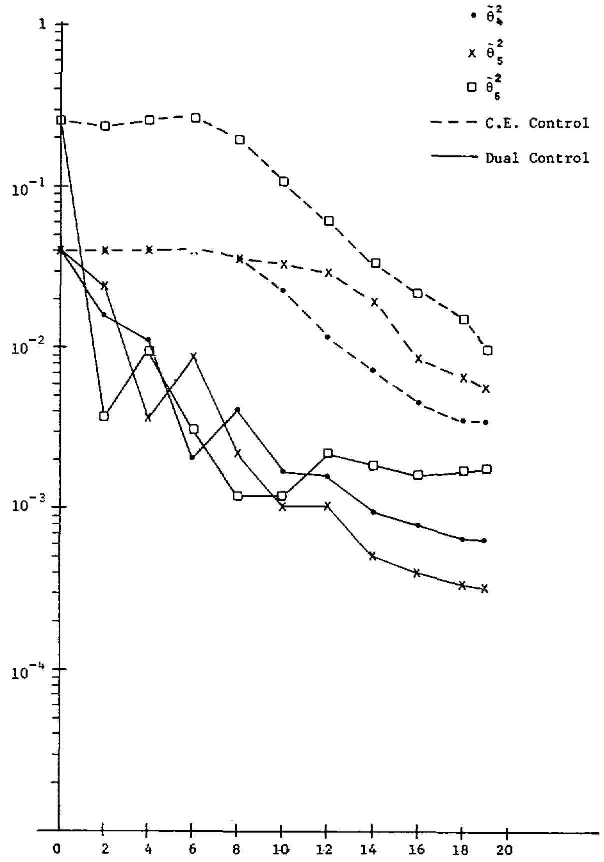

Fig. 3. Average estimation error squared in θ_4, θ_5, θ_6 for the interception example.

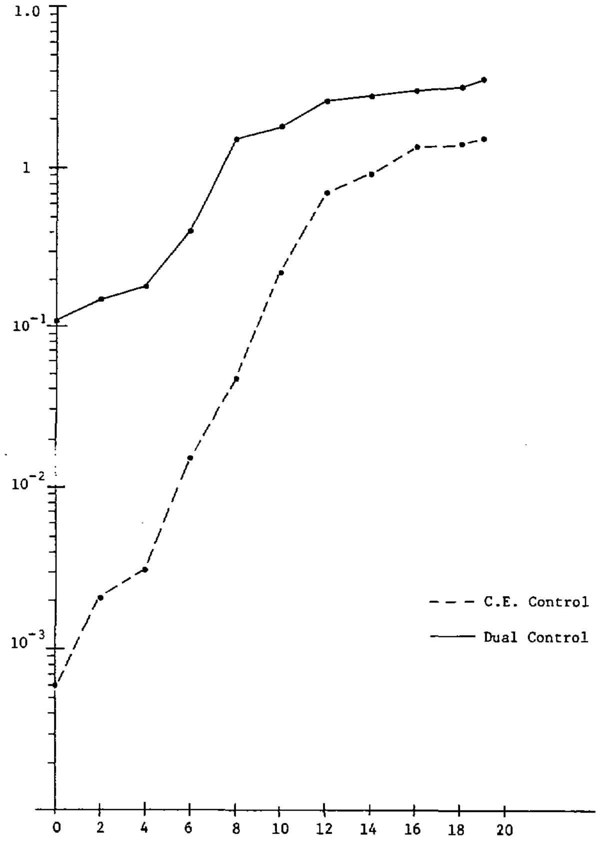

Fig. 4. Average cumulative control energy for the interception example.

TABLE II
SUMMARY OF RESULTS FOR THE SOFT LANDING EXAMPLE

Control Policy	Optimal Control with Known Parameters	CE Control with Unknown Parameters	Dual Control with Unknown Parameters
Average cost	15	104	28
Maximum cost in a sample of 20 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

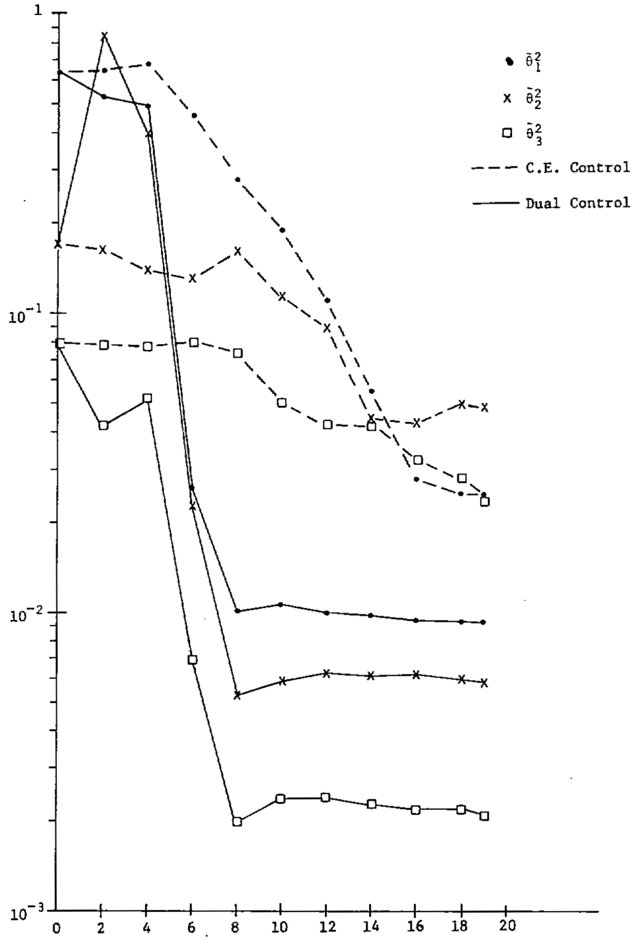

Fig. 5. Average estimation error squared in θ_1, θ_2, θ_3 for the soft landing example.

point in the state space, while the aim before was to hit a surface. Therefore, it should be expected that the average cost is higher than in the previous example. This is seen to hold true, as shown in Tables I and II, for the dual control and the optimal control with known parameters. However, for CE control, it does not hold true. This may look strange at first sight, but careful analysis of the simulation will offer an explanation for this.

Table II indicates the improvement of dual control over CE control, both in average performance and reliability. The terminal miss distance squared for the dual control is

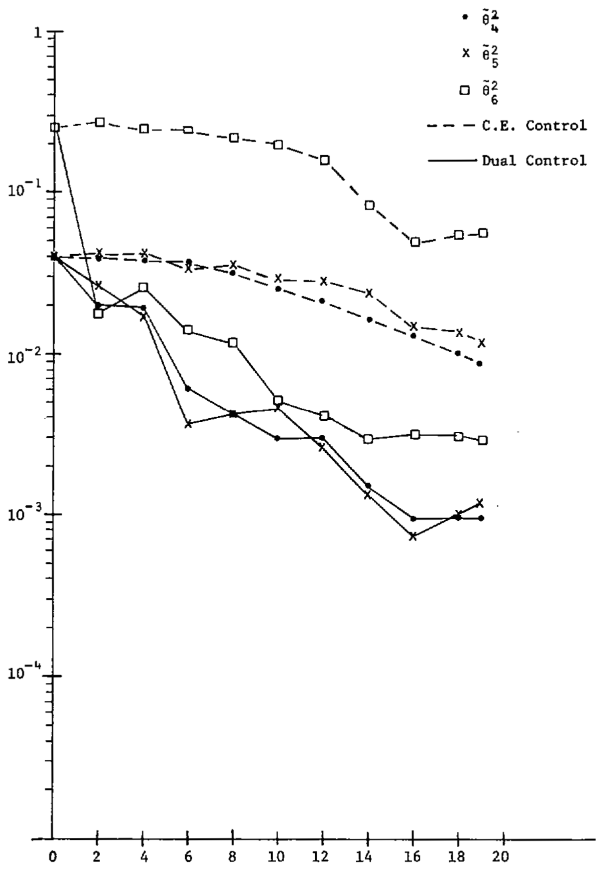

Fig. 6. Average estimation error squared in θ_4, θ_5, θ_6 for the soft landing example.

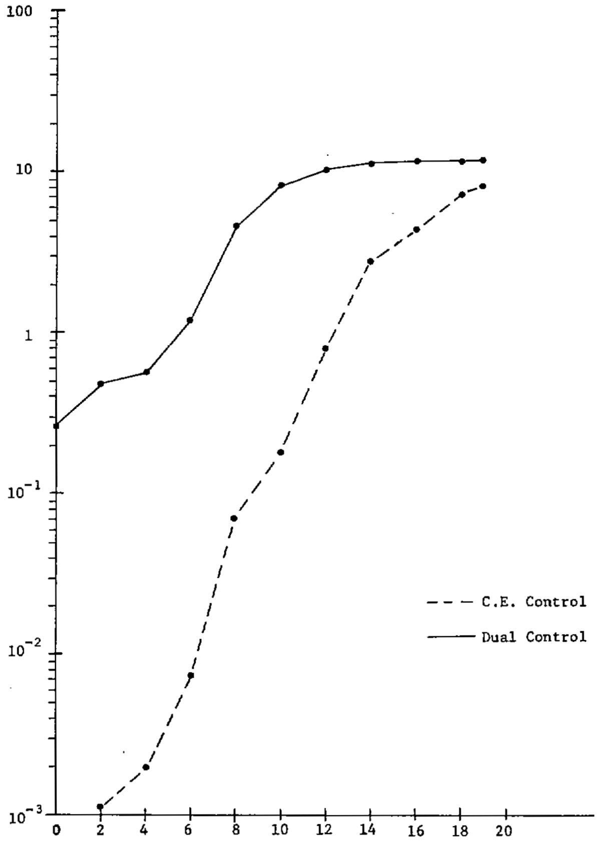

Fig. 7. Average cumulative control energy for the soft landing example.

very close to the unachievable lower bound given by the optimal control with known parameters. To achieve this small miss distance, the dual control invests considerable energy for learning purposes. This can be seen in Fig. 7, where it is shown that a large amount of energy is invested at the initial time to promote future learning. As a result, the parameters are adequately learned, and the dual control smoothly hits the final point $\underline{\rho}$ (see Fig. 7).

On the other hand, the CE control, being only passive in learning, learns more slowly, with the result that the terminal error is an order of magnitude higher than that of the dual control. As a consequence, the miss distance squared is substantially larger than that of the dual control.

To understand the passive and active learning of the CE and dual control, the results of the soft landing example and the previous example will be compared. First, compare the two CE controls. Note that the CE control energy used in the soft landing example (we shall call this the second example) is much more than that used in the interception example (we shall call this the first example). Note from Figs. 4 and 7 that, up to about $k = 12$, the CE 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 CE control is still applying little energy after time 12. 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 12 increases very quickly for the second example. This results in a much better estimation on the gain parameters. Since the learning in the first example is poorer than in the second example for the CE 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.

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 achieving a satisfactory final objective. Note the "speed" of learning in the second example compared with the first example (see Figs. 2, 3, 5, 6). The dual control regulates its energy in learning. In the first example, where learning is less important, it does not insist on learning by applying large controls in the beginning; in the second example, the learning is much more important, and thus 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 soft landing example is primarily due to the increase in accumulative input energy. This demonstrates the active learning characteristic of the dual control.

C. Remarks

1) A comparison of the computation time required by the dual control with that for CE control gives some idea of the computational feasibility of the proposed algorithm. The optimum control with known parameters took 3 s on a Univac 1108, while the CE required 6 s for one run. The time required for the dual control was 45 s (with a program that was not optimized). However, judging from the improvement over the CE control, the extra computation time seems worthwhile.

2) The active learning feature of this algorithm distinguishes it from the other approaches in the literature. The examples not only demonstrate that the dual control gives good performance, but, more importantly, it illustrates *why* it gives good performance.

VI. Conclusion

This work has presented a new algorithm for the control of noisy-input noisy-output linear systems with random parameters and with quadratic cost. The main feature of this algorithm is the fact that it is actively adaptive, i.e., the control plans the future learning of the system parameters as needed by the overall performance.

This control is obtained by using the stochastic dynamic programming equation in which the dual effect of the control appears explicitly. The algorithm yields a closed-loop control that takes into account not only the past observations but also the future observation program and the associated statistics.

A detailed description of the algorithm is given and examples on a three-dimensional system with unknown poles and zeros are presented. These simulations point out the importance of the active learning.

Acknowledgment

The authors wish to thank Dr. R. E. Larson, Dr. A. J. Tether, and Dr. L. Meier, all of Systems Control, Inc., for helpful discussions.

References

[1] A. Aoki, *Optimization of Stochastic Systems.* New York: Academic, 1967.

[2] M. Athans, R. P. Wishner, and A. Bertolini, "Suboptimal state estimation for continuous-time nonlinear systems from discrete noisy measurements," *IEEE Trans. Automat. Contr.*, vol. AC-13, pp. 504–514, Oct. 1968.

[3] D. L. Alspach and H. W. Sorenson, "Approximation of density function by a sum of Gaussian for nonlinear Bayesian estimation," in *Proc. Symp. Nonlinear Estimation Theory and Its Application* (San Diego, Calif.), 1970.

[4] R. Bellman, *Adaptive Control Processes: A Guided Tour.* Princeton, N.J.: Princeton Univ. Press, 1961.

[5] Y. Bar-Shalom and R. Sivan, "On the optimal control of discrete-time linear systems with random parameters," *IEEE Trans. Automat. Contr.*, vol. AC-14, pp. 3–8, Feb. 1969.

[6] R. S. Bucy and K. D. Senne, "Realization of optimum discrete-time nonlinear estimators," in *Proc. Symp. Nonlinear Estimation Theory and Its Application* (San Diego, Calif.), 1970.
[7] R. E. Curry, "A new algorithm for suboptimal stochastic control," *IEEE Trans. Automat. Contr.* (Short Papers), vol. AC-14, pp. 533–536, Oct. 1969.
[8] S. Dreyfus, "Some types of optimal control of stochastic systems," *SIAM J. Contr.*, vol. 2, no. 1, pp. 120–134, 1964.
[9] J. B. Farison, R. E. Graham, and R. C. Shelton, Jr., "Identification and control of linear discrete systems," *IEEE Trans. Automat. Contr.* (Short Papers), vol. AC-12, pp. 438–442, Aug. 1967.
[10] A. A. Feldbaum, *Optimal Control Systems*. New York: Academic, 1965.
[11] A. Jazwinski, *Stochastic Processes and Filtering Theory*. New York: Academic, 1970.
[12] D. G. Lainiotis, T. N. Upadhyay, and J. G. Deshpande, "Optimal adaptive control of linear systems," in *Proc. 1971 IEEE Conf. Decision and Control.*
[13] D. G. Luenberger, *Introduction to Linear and Nonlinear Programming*. Reading, Mass.: Addison-Wesley, 1973.
[14] G. Saridis and R. N. Lobbia, "Parameter identification, and control of linear discrete-time systems," in *1971 Joint Automatic Control Conf., Preprints.*
[15] H. A. Spang, "Optimum control of an unknown linear plant using Bayesian estimation of the error," *IEEE Trans. Automat. Contr.* (Short Papers), vol. AC-10, pp. 80–83, Jan. 1965.
[16] H. W. Sorenson, "Kalman filtering techniques," in *Advances in Control Systems*, vol. 3, C. T. Leondès, Ed. New York: Academic, 1966.
[17] G. Stein and G. N. Saridis, "A parameter-adaptive control technique," *Automatica*, vol. 5, pp. 731–740, Nov. 1969.
[18] E. Tse and M. Athans, "Adaptive stochastic control for a class of linear systems," *IEEE Trans. Automat. Contr.*, vol. AC-17, pp. 38–52, Feb. 1972.
[19] E. Tse, "Parallel computation of the conditional mean state estimate for nonlinear systems," in *Proc. 2nd Symp. Nonlinear Estimation Theory*, (San Diego, Calif.), Sept. 1971.
[20] E. Tse, Y. Bar-Shalom, and L. Meier, III, "Wide-sense adaptive dual control for nonlinear stochastic systems," this issue, pp. 98–108.

Edison Tse (M'70), for a photograph and biography see page 16 of the February 1973 issue of this TRANSACTIONS.

Yaakov Bar-Shalom (S'68–M'70), for a photograph and biography see this issue, page 108.

Sufficiently Informative Functions and the Minimax Feedback Control of Uncertain Dynamic Systems

DIMITRI P. BERTSEKAS AND IAN B. RHODES

Abstract—The problem of optimal feedback control of uncertain discrete-time dynamic systems is considered where the uncertain quantities do not have a stochastic description but instead are known to belong to given sets. The problem is converted to a sequential minimax problem and dynamic programming is suggested as a general method for its solution. The notion of a sufficiently informative function, which parallels the notion of a sufficient statistic of stochastic optimal control, is introduced, and conditions under which the optimal controller decomposes into an estimator and an actuator are identified. A limited class of problems for which this decomposition simplifies the computation and implementation of the optimal controller is delineated.

I. INTRODUCTION

THIS PAPER is concerned with the optimal feedback control of a discrete-time dynamic system in the presence of uncertainty. The traditional treatment of this problem has been to assign probability distributions to the uncertain quantities and to formulate the optimization problem as one of minimizing the expected value of a suitable cost functional. In this paper, a nonprobabilistic description of the uncertainty is adopted, where, instead of being modeled as random vectors with given probability distributions, the uncertainties are considered to be unknown except for the fact that they belong to given subsets of appropriate vector spaces. The optimization problem is then cast as one of finding the feedback controller within a prescribed admissible class that minimizes the maximum value (over all possible values of the uncertain quantities) of a suitable cost functional. This worst case approach to the optimal control of uncertain dynamic systems is applicable to problems where a set-membership description of the uncertain quantities is more natural or more readily available than a probabilistic one, or when specified tolerances must be met with certainty.

The modeling of uncertainties as quantities that are unknown except that they belong to prescribed sets and the adoption of a worst case viewpoint in the context of the problem of feedback control of a dynamic system was first considered by Witsenhausen [1], [2] and received further attention in [4]–[10]. In this paper a general minimax feedback control problem which involves a

Manuscript received May 1, 1972; revised October 9, 1972. Paper recommended by D. Sworder, Chairman of the IEEE S-CS Stochastic Control Committee. This work was supported in part by the Air Force Office of Scientific Research under Grants AFOSR-69-1724 and F-44620-69-C-0116 and in part by NASA under Grant NGL-22-009 (124).

D. P. Bertsekas is with the Department of Engineering–Economic Systems, Stanford University, Stanford, Calif. 94305.
I. B. Rhodes is with the Laboratory of Control Systems Science and Engineering, Washington University, St. Louis, Mo. 63130.