Wide-Sense Adaptive Dual Control for Nonlinear Stochastic Systems

EDISON TSE, YAAKOV BAR-SHALOM, AND LEWIS MEIER, III

Abstract—A new approach is presented for the problem of stochastic control of nonlinear systems. It is well known that, except for the linear-quadratic problem, the optimal stochastic controller cannot be obtained in practice. In general it is the curse of dimensionality that makes the strict application of the principle of optimality infeasible. The two subproblems of stochastic control, estimation and control proper, are, except for the linear-quadratic case, intercoupled. As pointed out by Feldbaum, in addition to its effects on the state of the system, the control also affects the estimation performance. In this paper, the control problem is formulated such that this dual property of the control appears explicitly. The resulting control sequence exhibits the closed-loop property, i.e., it takes into account the past observations and also the future observation program. Thus, in addition to being adaptive, this control also plans its future learning according to the control objective. Some preliminary simulation results illustrate these properties of the control.

I. Introduction

IN MANY control problems the uncertainties regarding the plant and the measurements can be modeled as stochastic processes. The procedure for obtaining the optimal (closed-loop) stochastic control is application of the principle of optimality, which leads to a stochastic dynamic programming equation [1], [5]. In the literature, one approximation method used for stochastic control is linearization of the plant about the deterministic optimal trajectory and application of the well-known separation theorem to the resulting perturbation equations. However, this may not give good performance if the system is very nonlinear and the noise level is high because, with the linearization approach, the control action is corrected only after it has been discovered that the trajectory has deviated from the nominal. But, in fact, if it is known that a disturbance will occur in the future, the control should be modified before as well as after the disturbance occurs in order to minimize its effects. Therefore, if linearization is to be used, some nominal trajectory other than the deterministic optimal trajectory should be used. Some authors [10], [21], [34] have considered the problem of choosing a nominal path to minimize a certain cost criterion obtained by using second-order perturbation analysis along the nominal path. The advantage of these approaches is the simplicity of the resulting control law; the main drawback is the validity of assuming that the system will remain close to the nominal trajectory. The open-loop feedback approach [12], [27], [6], [1], [18], [9], [23] suffers from the drawback that the resulting control is passive in learning—the control action does not anticipate the fact that future learning is possible.

In 1960, Feldbaum, in a series of three papers, introduced the dual control theory [13]. He pointed out that the control signal has two purposes that might be conflicting: one is to help learning about any unknown parameters and/or the state of the system (estimation); the other is to achieve the control objective. Thus the best control must have the characteristic of appropriately distributing its energy between the estimation and control purposes. Even though some further investigations have been reported [4], [16], no general development of algorithms that implement these ideas appears in the literature. The main purpose of this paper is to provide an understanding of dual control theory for complex dynamical systems and to develop an approach toward obtaining a near optimal dual control that can be implemented.

In Section II, optimal stochastic control theory is reviewed and the practical difficulties in computing and realizing the optimal control law are pointed out, serving as a motivation for the development of the later sections. In Sections III and IV the stochastic control problem is reformulated in light of the dual nature of the control, and a "wide-sense" adaptive dual control strategy which possesses an active learning characteristic is obtained. This result is new and is entirely different from the other suboptimal approaches reported in the literature. Since the derived algorithm is rather complicated, an illustrative example is presented to provide understanding of the dual nature of the resulting control strategy. In Section V, a simple scalar example is presented to demonstrate: 1) the computational feasibility of the new algorithm; 2) the performance level of the new algorithm; and 3) to provide some insight into the dual control theory.

II. Optimal Stochastic Control

In this section, the formulation and solution of the optimal stochastic control problem for discrete-time systems is discussed, as are the difficulties associated with the solution procedures. These difficulties motivate the wide-sense adaptive dual control approach presented in Section III.

Consider a discrete-time nonlinear stochastic system described by

$$x(k+1) = f[k,x(k),u(k)] + \xi(k); \qquad k = 0,1,\cdots,N-1$$

Manuscript received May 29, 1972; revised November 2, 1972. Paper recommended by D. Sworder, Chairman of the IEEE S-CS Stochastic Control 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.

$$y(k) = h[k,x(k)] + n(k); \qquad k = 1,\cdots,N \tag{2.1}$$

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

$$E\{x(0)\} = \hat{x}(0|0); \qquad \mathrm{cov}\,\{x(0)\} = \Sigma(0|0) \tag{2.2}$$

$$E\{\xi(k)\} = 0; \qquad \mathrm{cov}\,\{\xi(k)\} = Q(k) \tag{2.3}$$

$$E\{n(k+1)\} = 0; \qquad \mathrm{cov}\,\{n(k+1)\} = R(k+1). \tag{2.4}$$

Consider further the performance measure

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

where the expectation $E\{\cdot\}$ is taken over all underlying random quantities. Finally, consider admissible controls of the feedback type

$$u(k) = u(k,Y^k,U^{k-1}) \qquad Y^k \triangleq \{y(1),\cdots,y(k)\}$$
$$U^{k-1} \triangleq \{u(0),\cdots,u(k-1)\}. \tag{2.6}$$

The goal is to find the optimal control sequence $\{u^*(k)\}_{k=0}^{N-1}$ that is of the form (2.6) and minimizes the cost (2.5) subject to the dynamic constraint (2.1).

To solve the optimal control problem stated above, Bayes' rule and dynamic programming are used. A complete derivation for the optimal solution is given by Meier [20]; therefore, we shall only outline the derivation and summarize the results below.

An important concept is the *information state*. This can be viewed as a quantity that is equivalent to the observation process Y^k *and* all *a priori* knowledge of the system *and* U^{k-1} in describing the future evolution of the system [24]. Thus, an information state will summarize all the information content conveyed by the observation process Y^k and past control sequence U^{k-1}. Clearly, the combined sequence

$$\mathcal{P}_k^1 = (Y^k,U^{k-1}) \tag{2.7}$$

is an information state. Another such information state is the conditional density

$$\mathcal{P}_k^2 \triangleq p[x(k)\,|\,Y^k,U^{k-1}]. \tag{2.8}$$

Using Bayes' rule, a recursive equation for the conditional density is given by [1], [20] as

$$\mathcal{P}_{k+1}^2[\mathcal{P}_k^2,u(k)] = \frac{1}{C_k}\, p[y(k+1)\,|\,x(k+1)] \cdot \int p[x(k+1)\,|\,x(k),u(k)]\mathcal{P}_k^2\, dx(k) \tag{2.9}$$

where C_k is a normalizing constant.

Using the principle of optimality in the information state space yields the stochastic dynamic programming equation

$$I^*\{\mathcal{P}_k,k\} = \min_{u(k)} E\{\mathcal{L}[x(k),u(k),k] + I^*\{\mathcal{P}_{k+1}[\mathcal{P}_k,u(k)],k+1\}\,|\,Y^k,U^{k-1}\} \tag{2.10}$$

where $u(k)$ is a *deterministic* quantity, $\mathcal{P}_k$ is an information state (can be either $\mathcal{P}_k^1$ or $\mathcal{P}_k^2$), and $I^*\{\cdot,k\}$ denotes the optimal cost-to-go associated with the information state at time k. If we use $\mathcal{P}_k^1$ as an information state, then the optimal control can be obtained by solving (2.10) where an optimal feedback table $\{u^*(Y^k,U^{k-1})\}_{k=1}^{N-1}$ is constructed for *all possible* pairs (Y^k,U^{k-1}), $k = 1,2,\cdots,N-1$. On the other hand, if we use $\mathcal{P}_k^2$ as an information state, then the optimal control can be solved by the following separate procedures.

1) Control: The optimum control law is found as a function of the conditional density $p[x(k)\,|\,Y^k,U^{k-1}]$ by solving the stochastic dynamic programming equation (2.10). In general, this can be an off-line procedure.

2) Estimation: The conditional density is updated by use of the recursion relation (2.9), and the optimum input is obtained from the optimum control law. The updating of the conditional density must be done in real time.

In principle, the optimal control problem has been solved when (2.9) and (2.10) are derived; however, in practice, the problem only begins with these equations. In the following, we discuss the difficulties associated with the solution procedures using either $\mathcal{P}_k^1$ or $\mathcal{P}_k^2$ as the information state. This will motivate our development in the next section.

From (2.7), note that the dimension of $\mathcal{P}_k^1$ grows linearly in k. Thus, even with appropriate quantizing, the number of quantization points for the feedback table, which grows in time, will soon become too large to be handled by a computer of any size. Note that the expectation in (2.10) requires the availability of the conditional density $p[x(k)\,|\,Y^k,U^{k-1}]$, which is usually infinite dimensional. In general, the optimal cost-to-go-function $I^*[\cdot,\cdot]$ *cannot* be expressed as an explicit function of the information state. Thus, direct solution of (2.10) becomes practically impossible for any computer. We face a similar kind of difficulty even if we use $\mathcal{P}_k^2$ as the information state. In this case the information state $\mathcal{P}_k^2$ is usually of infinite dimension for all $k \geq 1$. One may attempt to approximate the solution for (2.10). However, even if this can be done, it still does not solve the dimensionality problem, since, in general, the approximate optimal control law is nonlinear in the information state and can only be expressed as a table look-up type of function of the information state. This prohibits functional realization of the optimal control law, and thus real-time generation of the optimum control value is practically impossible for most problems.

Note that the basic difficulty is in the control rather than in the estimation procedure. The updating of the density, although a difficult problem in itself, can be reasonably approximated by using parallel estimation procedures. (Some recent results [7], [29], [3], [17] indicate the feasibility of parallel estimation.) We should emphasize the fact that the capability of approximating the conditional density does not solve half the problem because the difficulty in obtaining the optimal control in real time is not so much due to the estimation procedure as to the curse of dimensionality [5] and to the fact that, even if an optimal control law is obtained, the extremely large (perhaps infinite) number of possible information states will prevent it from being realizable.

In the special case where the system (2.1) is linear, the conditional density $p[x(k) \mid Y^k, U^{k-1}]$ is equivalent to the conditional mean estimate $\hat{x}(k \mid k)$ and the corresponding covariance generated by the Kalman filter [24], [20], [28]. If, in addition, the cost is quadratic, then the optimal cost-to-go $I^*[\cdot, k]$ can be expressed *explicitly* as a function of $\hat{x}(k \mid k)$, so that (2.10) can be solved exactly to yield a realizable linear feedback law [14], [22], [24], [30]. This result is known as the separation theorem or certainty equivalence principle. In this case the optimal stochastic (closed-loop) control does not depend on the future observation program because the covariances appearing in the expression of the optimal cost-to-go are independent of the control.

In the next section, a new method will be presented that is based on the principle of optimality using an approximation of the information state $\mathcal{P}_k$[2] and the concept of dual control. In contrast to the previous approaches, this method will not only take into account the past observation information but also the *future observation program and its associated statistics.*

III. Wide-Sense Adaptive Control and the Dual Cost

In Section II, it was noted that the main difficulties in implementing the optimal control law are: 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; 3) storage of the control value associated with each information state at time k, $k = 0, \cdots, N-1$, is practically impossible due to the large dimensionality. Thus, a reasonable suboptimal approach would be to: 1) approximate the information state such that the dimension of the information state space stays constant for all time; 2) approximate the optimal cost-to-go associated with each information state at time k, $k = 0, \cdots, N-1$; and 3) compute the control value on-line rather than obtain the feedback law off-line and store the whole "feedback table." Each of these procedures is discussed in detail in the following. To simplify the discussion, assume that the cost is of the form

$$\mathcal{L}[x(k), u(k), k] = L[x(k), k] + \phi[u(k), k]. \tag{3.1}$$

The extension to the more general cost is straightforward (see Section IV, Remark 3).

As discussed in Section II, the information state $p[x(k) \mid Y^k, U^{k-1}]$ is generally of infinite dimension. One approach to reduce this dimension is to use the wide-sense property [11]; in this approach the controller is restricted to the form

$$u(k) = u[k, \hat{x}(k \mid k), \Sigma(k \mid k)] \tag{3.2}$$

where

$$\begin{aligned} \hat{x}(k \mid k) &= E\{x(k) \mid Y^k, U^{k-1}\} \\ \Sigma(k \mid k) &= \operatorname{cov}\{x(k) \mid Y^k, U^{k-1}\}. \end{aligned} \tag{3.3}$$

Such a control scheme is referred to as the wide-sense adaptive control law. The computation of $\hat{x}(k \mid k), \Sigma(k \mid k)$ can be done by one of the following approximate methods: 1) extended Kalman filter [25], [15]; 2) adaptive filter with tuning [15], [31]; 3) second-order filter [2]; or by using the optimum estimator [7], [29], [3]. Depending on the specific problem under investigation, one of these methods may be more appropriate than the others.

Before going into the new approximation procedure, the perturbation control problem is considered first, and a cost is obtained that exhibits the dual property of the control. As will be seen later, the result from this section will provide a reasonable approximation of the cost-to-go to be used in the stochastic dynamic programming equation.

The present time is indexed by k. Let us assume that U^{k-1} has been applied to the system and that the observation sequence Y^k has been obtained. The conditional mean $\hat{x}(k \mid k)$ and covariance $\Sigma(k \mid k)$ are assumed available from an estimator. Consider a nominal open-loop control sequence $U_o(k, N-1) \triangleq \{u_o(j)\}_{j=k}^{N-1}$ and the associated nominal path

$$x_o(j+1) = f[j, x_o(j), u_o(j)]; \qquad j = k, \cdots, N-1 \tag{3.4}$$

with initial condition $x_o(k)$. Let $\delta x(j)$ be a small perturbation about the nominal path due to the disturbance $\xi(j)$ and a perturbation control $\delta u(j)$. The true trajectory and control are given by

$$x(j) = x_o(j) + \delta x(j) \qquad u(j) = u_o(j) + \delta u(j) \tag{3.5}$$

with $x(j)$, $u(j)$ satisfying (2.1). Since $\delta x(j)$, $\delta u(j)$ are assumed to be small, we can approximate the cost-to-go by expanding it up to second order, i.e.,

$$\begin{aligned} J(k) &\triangleq E\left\{\psi[x(N)] + \sum_{j=k}^{N-1} [L[x(j), j] + \phi[u(j), j]] \,\Big|\, Y^k\right\} \\ &\approx J_o(k) + E\left\{\psi_{o,x}{}'\delta x(N) + \frac{1}{2}\,\delta x'(N)\psi_{o,xx}\delta x(N)\right. \\ &\quad + \sum_{j=k}^{N-1}\left[L_{o,x}{}'(j)\delta x(j) + \frac{1}{2}\,\delta x'(j)L_{o,xx}(j)\delta x(j)\right. \\ &\quad \left.\left. + \phi_{o,u}{}'(j)\delta u(j) + \frac{1}{2}\,\delta u'(j)\phi_{o,uu}(j)\delta u(j)\right] \Big|\, Y^k\right\} \end{aligned} \tag{3.6}$$

where

$$J_o(k) = \psi[x_o(N)] + \sum_{j=k}^{N-1} L[x_o(j), j] + \phi[u_o(j), j]. \tag{3.7}$$

The quantities $\psi_{o,x}$ and $\psi_{o,xx}$ are, respectively, the gradient and Hessian of $\psi(\cdot)$ with respect to x evaluated along the nominal trajectory. For a fixed nominal, by choosing $\delta u_o(j)$, $j = k,\cdots,N-1$, to minimize the incremental cost $\Delta J(k) \triangleq J(k) - J_o(k)$, one obtains a cost $J^*[k,U_o(k,N-1)]$ associated with the nominal control $U_o(k,N-1)$.

Let us consider the perturbation control problem. From (3.6) we have

$$\Delta J(k) \triangleq J(k) - J_o(k) = E\left\{\psi_{o,x}'\delta\hat{x}(N|N) + \frac{1}{2}\,\delta\hat{x}'(N|N)\psi_{o,xx}\delta\hat{x}(N|N) + \sum_{j=k}^{N-1}\left[\left(L_{o,x}'(j)\delta\hat{x}(j|j) + \frac{1}{2}\,\delta\hat{x}'(j|j)L_{o,xx}\delta\hat{x}(j|j) + \phi_{o,u}'(j)\delta u(j) + \frac{1}{2}\,\delta u'(j)\phi_{o,uu}(j)\delta u\right)\right]\middle| Y^k\right\} + \frac{1}{2}\,\mathrm{tr}\left\{\psi_{o,xx}\Sigma_o(N|N) + \sum_{j=k}^{N-1} L_{o,xx}(j)\Sigma_o(j|j)\right\} \tag{3.8}$$

where $\delta\hat{x}(j|j) \triangleq E\{\delta x(j)\,|\,Y^j\}$ and $\Sigma_o(j|j) \triangleq \mathrm{cov}\,\{\delta x(j)\,|\,Y^j\}$ is the future error covariance, which is assumed to be generated by the extended Kalman filter linearized about the nominal control and trajectory [25], and therefore is deterministic. The problem is to minimize $\Delta J(k)$, subject to the dynamic constraints of the second-order incremental process.

Application of dynamic programming with retention of up to second-order terms yields the following (the derivations of (3.9)–(3.17) can be found in the Appendix):

$$\delta u_o^*(j) = -[H_{o,uu}(j) + f_{o,u}'(j)K_o(j+1)f_{o,u}(j)]^{-1} \cdot \{[f_{o,u}'(j)K_o(j+1)f_{o,x}(j) + H_{o,ux}(j)]\delta\hat{x}(j|j) + H_{o,u}(j)\} \tag{3.9}$$

where

$$H_o(j) \triangleq L[x_o(j),j] + \phi[u_o(j),j] + p_o'(j+1)f_o(j); \qquad f_o(j) \triangleq f[j,x_o(j),u_o(j)] \tag{3.10}$$

$$p_o(j) = H_{o,x}(j) - [f_{o,u}'(j)K_o(j+1)f_{o,x}(j) + H_{o,ux}(j)]' \cdot [H_{o,uu}(j) + f_{o,u}'(j)K_o(j+1)f_{o,u}j)]^{-1}H_{o,u}(j); \qquad p_0(N) = \psi_{o,x} \tag{3.11[1]}$$

$$K_o(j) = f_{o,x}'(j)K_o(j+1)f_{o,x}(j) - [f_{o,u}'(j)K_o(j+1)f_{o,x}(j) + H_{o,ux}(j)]'[H_{o,uu}(j) + f_{o,u}'(j)K_o(j+1)f_{o,u}(j)]^{-1} \cdot [f_{o,u}'(j)K_o(j+1)f_{o,x}(j) + H_{o,ux}(j)] + H_{o,xx}(j); \qquad K_o(N) = \psi_{o,xx} \tag{3.12}$$

and the optimal cost associated with the nominal $U_o(k,N-1)$ is given by

$$J^*[k,U_o(k,N-1)] = J_o(k) + g_o(k) + \frac{1}{2}\,\mathrm{tr}\left\{\psi_{o,xx}\Sigma_o(N|N) + \sum_{j=k}^{N-1}\{H_{o,xx}(j)\Sigma_o(j|j) + [\Sigma_o(j+1|j) - \Sigma_o(j+1|j+1)]K_o(j+1)\}\right\} + p_o'[\hat{x}(k|k) - x_o(k)] + \frac{1}{2}\,[\hat{x}(k|k) - x_o(k)]' \cdot K_o(j+1)[\hat{x}(k|k) - x_o(k)] \tag{3.13}$$

with $g_o(j)$ satisfying

$$g_o(j) = g_o(j+1) - \frac{1}{2}\,H_{o,u}(j)[H_{o,uu}(j) + f_{o,u}'(j) \cdot K_o(j+1)f_{o,u}(j)]^{-1}H_{o,u}(j); \qquad g_o(N) = 0. \tag{3.14}$$

Notice that the updated and one-step prediction error covariances $\Sigma_o(j+1|j+1)$ and $\Sigma_o(j+1|j)$ are dependent on the choice of the nominal control $U_o(k,N-1)$. The cost $J^*[k,U_o(k,N-1)]$ associated with $U_o(k,N-1)$ involves: 1) control cost $J_o(k)$; and 2) estimation cost—the remaining terms involving nonnegative weightings of error covariances. For this reason, $J^*[k,U_o(k,N-1)]$ will be called the dual cost associated with the nominal $U_o(k,N-1)$.

The cost (3.13) incorporates the future observation program and the associated statistics via the covariances of the future updated states.

IV. Description of the Adaptive Dual Control Method

The outline of this method, which is the main result of this paper, is summarized as follows. It is assumed that at the present time k, one can apply an *arbitrary* control $u(k)$. From time k to $k+1$, a second-order extrapolation is performed to obtain the predicted state and covariance. A cost is associated with the predicted state and covariance that approximately represents the optimum future learning and control (closed-loop cost) if $u(k)$ is applied. This approximate optimal cost-to-go is obtained by considering perturbation analysis about some nominal trajectory for the future time, $j \geq k+1$. By assuming that perturbation control will be applied in addition to the nominal from time $k+1$ to the end of the process, one obtains an expression of the cost-to-go that includes the future estimation and control performance. As will be seen later, the future estimation performance depends upon the present control $u(k)$. By adding the cost of the present control $\phi[u(k)]$ and the resulting approximate optimal cost-to-go associated with the predicted state and covariance, one obtains a cost which includes both control performance and estimation performance. The method is to choose $u(k)$ to minimize this overall cost. Once the minimizing value control $u^*(k)$

[1] The inverse is assumed to exist; if it does not exist, replace it by the pseudoinverse.

is obtained, it is applied to the system. One of the estimators discussed in Section III can be used to update the estimate and covariance, and the whole procedure is repeated for time $k+1$, and so on, until the end of the control period.

As will be seen later, this approach provides an active learning characteristic in the resulting adaptive control method that distinguishes it from all the previous suboptimal approaches.

Consider a control $u(k)$ to be applied at time k. Expanding the function $f[k,\cdot,u(k)]$ about $\hat{x}(k|k)$ up to second-order terms, we obtain the predicted state and covariance

$$\hat{x}(k+1|k) = f[k,\hat{x}(k|k),u(k)] + \frac{1}{2}\sum_{i=1}^{n} e_i \cdot \mathrm{tr}\,\{f_{xx}{}^i[\hat{x}(k|k),u(k)]\Sigma(k|k)\} \quad (4.1)$$

$$\begin{aligned}\Sigma(k+1|k) &= f_x[\hat{x}(k|k),u(k)]\Sigma(k|k)f_x{}'[\hat{x}(k|k),u(k)] \\ &\quad + Q(k) + \frac{1}{2}\sum_{i=1}^{n}\sum_{j=1}^{n} e_i e_j{}' \\ &\quad \cdot \mathrm{tr}\,\{f_{xx}{}^i[\hat{x}(k|k),u(k)]\Sigma(k|k) \\ &\quad \cdot f_{xx}{}^j[\hat{x}(k|k),u(k)]\Sigma(k|k)\} \qquad (4.2)\end{aligned}$$

where $f_{xx}{}^i$ denotes the Hessian of the ith component of f with respect to x and $\{e_i\}_{i=1}^{n}$ is the natural base in R^n. The updated error covariance for the incremental state estimate is

$$\Sigma(k+1|k+1) = \{I - V(k+1)h_x[k+1,\hat{x}(k+1|k)]\} \cdot \Sigma(k+1|k) \quad (4.3)$$

$$\begin{aligned}V(k+1) &= \Sigma(k+1|k)h_x[k+1,\hat{x}(k+1|k)]' \\ &\quad \cdot \Big\{h_x[k+1,\hat{x}(k+1|k)]\Sigma(k+1|k) \\ &\quad \cdot h_x{}'[k+1,\hat{x}(k+1|k)] + R(k+1) \\ &\quad + \frac{1}{2}\sum_{i=1}^{m}\sum_{j=1}^{m} e_i e_j{}'\,\mathrm{tr}\,\{h_{xx}{}^i[k+1, \\ &\quad \hat{x}(k+1|k)]\Sigma(k+1|k)h_{xx}{}^j[k+1, \\ &\quad \hat{x}(k+1|k)]\Sigma(k+1|k)\}\Big\}^{-1}. \qquad (4.4)\end{aligned}$$

With each control $u(k)$, associate a point in the state space and a future nominal control sequence

$$\{x_o[k+1;u(k)],\ U_o[k+1,N-1;u(k)]\}$$

that defines a nominal trajectory from $k+1$ to N as in (3.4), with initial state

$$x_o(k+1) = x_o[k+1;u(k)]. \quad (4.5)$$

The future analysis (from $k+1$ to N) will be based on perturbation about this nominal to obtain the approximate optimal cost-to-go that includes the dual effect.

Since we assume that, for $j \geq k+1$, only perturbation analysis will be carried out along this nominal defined above, the optimal cost-to-go at time $k+1$ can be written, based upon the results of the previous section, as follows (see the Appendix for the derivation):

$$\begin{aligned}&I^*[\hat{x}(k+1|k+1),\Sigma(k+1|k+1),k+1] \\ &= J_o(k+1) + g_o(k+1) + \tfrac{1}{2}\,\mathrm{tr}\,\Big\{\psi_{o,xx}\Sigma_o(N|N) \\ &\quad + \sum_{j=k+1}^{N-1}\{H_{o,xx}(j)\Sigma_o(j|j) + [\Sigma_o(j+1|j) \\ &\quad - \Sigma_o(j+1|j+1)]K_o(j+1)\}\Big\} \\ &\quad + p_o{}'(k+1)\tilde{x}_o(k+1|k+1) \\ &\quad + \tfrac{1}{2}\tilde{x}_o{}'(k+1|k+1)K_o(k+1)\tilde{x}_o(k+1|k+1) \qquad (4.6)\end{aligned}$$

where

$$\tilde{x}_o(k+1|k+1) \triangleq \hat{x}(k+1|k+1) - x_o(k+1) \quad (4.7)$$

and the error covariances $\Sigma_o(j+1|j)$, $\Sigma_o(j+1|j+1)$ are given by extended Kalman filter equations [15] with initial condition $\Sigma(k+1|k+1)$ given by (4.3). Therefore, the cost of applying $u(k)$ can be approximated as follows:

$$\begin{aligned}I_d[u(k)] &= E\{\phi[u(k),k] + L[x(k),k] \\ &\quad + I^*[\hat{x}(k+1|k+1),\Sigma(k+1|k+1),k+1]\,|\,Y^k\}. \qquad (4.8)\end{aligned}$$

The above is the key equation of the wide-sense adaptive control procedure developed in this paper. Its two main features are the following.

1) It is of the *closed-loop* type. The cost-to-go is a function of the future observations that are averaged out in order to obtain a nonanticipative control at the present time; however, this control will be a function of the *future observation program and the associated statistics*; this is the most important feature of the closed-loop control (see, e.g., [1], [6]).

2) The approximation is in the value of the cost-to-go. The use of a future nominal and perturbation is to make this computation feasible.

Substituting (4.7) into (4.8) and noting that $E\{L[x(k),k]\,|\,Y^k\}$ is independent of $u(k)$, the cost to be minimized is given by

$$\begin{aligned}J_d[u(k)] &= J_o(k+1) + \phi[u(k),k] + g_o(k+1) \\ &\quad + p_o{}'(k+1)[\hat{x}(k+1|k) - x_o(k+1)] \\ &\quad + [\hat{x}(k+1|k) - x_o(k+1)]'K_o(k+1) \\ &\quad \cdot [\hat{x}(k+1|k) - x_o(k+1)] \\ &\quad + \tfrac{1}{2}\,\mathrm{tr}\,\Big\{[\Sigma(k+1|k) - \Sigma(k+1|k+1)] \\ &\quad \cdot K_o(k+1) + \psi_{o,xx}\Sigma_o(N|N) \\ &\quad + \sum_{j=k+1}^{N-1} H_{o,xx}(j)\Sigma_o(j|j) + [\Sigma_o(j+1|j) \\ &\quad - \Sigma_o(j+1|j+1)]K_o(j+1)\Big\} \qquad (4.9)\end{aligned}$$

subject to the constraints (4.1)–(4.5).

If

$$x_o[k+1;u(k)] = \hat{x}(k+1|k) \tag{4.10}$$

where $\hat{x}(k+1|k)$ is given by (4.1), then the dual cost becomes

$$\begin{aligned} J_d[u(k)] = J_o(k+1) + \phi[u(k),k] + g_o(k+1) \\ + \tfrac{1}{2}\operatorname{tr}\Big\{[\Sigma(k+1|k) - \Sigma(k+1|k+1)] \\ \cdot K_o(k+1) + \psi_{o,xx}\Sigma_o(N|N) \\ + \sum_{j=k+1}^{N-1} H_{o,xx}(j)\Sigma_o(j|j) + [\Sigma_o(j+1|j) \\ - \Sigma_o(j+1|j+1)]K_o(j+1)\Big\}. \end{aligned} \tag{4.11}$$

Alternately, rather than associating a nominal to each $u(k)$ in order to evaluate the corresponding cost-to-go, one can use a finite set of precomputed nominals. In this case, for a given $u(k)$, one can choose the nominal whose state at time $k+1$ is the nearest (e.g., using Euclidean distance) to the predicted state at $k+1$ yielded by $u(k)$. For more details, see [32].

Denote the optimal solution for the minimization of $J_d[u(k)]$ by $u^*(k)$. When $u^*(k)$ is applied to the system and a new observation $y(k+1)$ is obtained, the estimate of $x(k+1)$ and its error covariance are updated using any one of the estimation methods discussed in the beginning of Section III, and the same procedure is repeated to obtain $u^*(k+1)$. Starting with $k=0$ to $k=N-1$, we obtain a sequence of controls $\{u^*(k)\}_{k=0}^{N-1}$. We shall call this sequence the *wide-sense adaptive dual control.*

Note that, in the above development, the choice of a future nominal control is "fictitious" in the sense that it is not actually applied in the future, i.e., it is only used to approximate the optimal cost-to-go. Therefore, its choice is quite flexible and is dependent upon the problem under consideration. In Section V, we indicate how the nominal can be selected for an example problem.

Remark 1: Because of the rather complicated dependence of J_d on $u(k)$, one has to search to find the minimizing $u(k)$ that will be applied to the system. Search methods appropriate for finding $u(k)$ are those of local variations or, if the control is a scalar, a line search, e.g., Fibonacci. To obtain $u^*(k)$, start the search at $u_{ce}(k)$, the first of the sequence of controls obtained by assuming certainty equivalence (i.e., the separation theorem) to be valid. Then determine which direction J_d decreases; next, the "box" in which the minimum lies; and then narrow it down to a certain predetermined size; and finally make a quadratic interpolation from the last three points; the result is taken as $u^*(k)$. For different search methods, see, e.g., [19].

Remark 2: Let us comment on the dual nature of the control. The estimation purpose of the control is reflected by the covariances appearing in (4.9). If the predicted and updated error covariances are independent of the control, the dual property will disappear. This would be the case if the system is linear with known parameters. In general, this dual property of the control is important.

Remark 3: Similar results hold even when the cost has the more general form (2.5). The only change one needs to make is to replace (3.10) by

$$H_o(j) \triangleq \mathcal{L}[x_o(j),u_o(j)] + p_o'(j+1)f[j,x_o(j),u_o(j)] \tag{4.12}$$

and the term $\phi[u(k)]$ in (4.8) by

$$\begin{aligned} E\{\mathcal{L}[x(k),u(k),k]|Y^k\} \approx \mathcal{L}[\hat{x}(k|k),u(k),k] \\ + \operatorname{tr}\{\mathcal{L}_{xx}[\hat{x}(k|k),u(k),k]\Sigma(k|k)\}. \end{aligned} \tag{4.13}$$

V. Simulation Studies

In this section, we shall study an example problem via simulation. The purposes of this study are: 1) to indicate how the nominal can be selected for a specific problem; 2) to investigate the computational feasibility of the adaptive dual control algorithm; 3) to compare this algorithm with another widely used suboptimal algorithm—the certainty equivalence algorithm; and 4) to understand the dual nature of the proposed algorithm; in particular, to understand the active learning feature of the control. Since there is no simulation study on dual control in the literature, and since the equations developed in the previous section look quite complicated, we shall consider a very simple example in this section. The simplicity of this example will allow one to understand the implications more clearly. More complicated examples will appear elsewhere [33].

Consider a scalar linear system

$$\begin{aligned} x(k+1) &= ax(k) + bu(k) + \xi(k) \\ y(k) &= x(k) + \eta(k) \end{aligned} \tag{5.1}$$

where a, b are unknown constants and $\xi(k)$, $\eta(k)$ are independent zero-mean white noises with covariances q and r, respectively. The problem is to find a control sequence $\{u^*(k)\}_{k=0}^{N-1}$ to minimize the performance

$$J = E\left\{\frac{c}{2}[x(N)-\rho]^2 + \frac{1}{2}\sum_{k=0}^{N-1}u^2(k)\right\}; \qquad c>0. \tag{5.2}$$

By augmenting the state

$$z(k) \triangleq [x(k),a,b]', \tag{5.3}$$

one converts the problem into the form considered in the preceding sections. For this specific example, a future nominal control $u_o[j;\hat{z}(k+1|k)]$ is associated with each predicted state $\hat{z}(k+1|k)$, and $u_o[j;\hat{z}(k+1|k)]$ is obtained by assuming $\hat{a}(k|k)$, $\hat{b}(k|k)$ to be the true parameters and $\hat{x}(k+1|k)$ to be the true state and by solving the deterministic version of the above control problem.

In the following, the certainty equivalence (CE) control strategy is compared to the adaptive dual control developed in this paper. Two cases are considered [for both cases, $N=20$, $\rho=5$, $C=100$, $x(0)=0$, $a=0.8$, $b=0.5$, $q=0.25$, $r=0.04$; the initial estimates are $\hat{x}(0|0)=0.13$, $\hat{a}(0|0)=1.2$, $\hat{b}(0|0)=0.3$ with initial error covariance $\Sigma(0|0) = \operatorname{diag}(0.25, 0.04, 0.01)$]: 1) the observations are available for all $k=1,2,\cdots,19$; 2) the observations are available at $k=1,\cdots,14$, but for $k \geq 15$, no observa-

TABLE I
COMPARISON OF DUAL CONTROL WITH CE CONTROL FOR THE SCALAR EXAMPLE (CASE 1)

	Average Miss Distance Squared	Average Terminal Error Squared in a	Average Terminal Error Squared in b	Average Performance	Range of Performance	Standard Deviation of Performance
Optimum	0.0653	0	0	20.7	15.71–36.34	6
CE	0.311	0.126	0.233	34.7	22.11–70.43	17
Dual	0.219	0.125	0.228	32.0	22.04–48.40	10

TABLE II
COMPARISON OF DUAL CONTROL WITH CE CONTROL FOR THE SCALAR EXAMPLE (CASE 2)

	Average Miss Distance Squared	Average Terminal Error Squared in a	Average Terminal Error Squared in b	Average Performance	Range of Performance	Standard Deviation of Performance
Optimum	0.097	0	0	22.3	17.66– 43.26	7
CE	7.353	1.609	0.650	308.6	39.31–1882.5	561
Dual	0.143	0.282	0.258	74.5	66.6 – 113.36	13

tion is available. It is impossible to see how close the dual control strategy performance is to that of the truly optimum control strategy since the truly optimum control strategy is very difficult to obtain. To give an idea of the performance level of the dual control strategy, we shall include the results for the optimal control when the parameters are all known. The performance for this will serve as a lower bound. It must be kept in mind that this lower bound is *not* achievable even by the truly optimal stochastic control for our problem. Ten Monte Carlo runs were performed for both cases, the results of which are shown in Tables I and II.[2] The first row shows the results for the optimum control when the parameters are known.

From Table I, we see that, on the average, the dual control is better than the CE control, though not by much. An important fact here is that the dual control performance has a relatively small deviation from its average performance compared with that of the CE control. This property indicates that the dual control is more reliable than the CE control under stochastic effects.

From the estimation performance, both the dual control and CE control perform well in terms of estimation at the terminal time, as shown in Figs. 1 and 2, with the dual control being slightly better. In Fig. 1, the control histories are plotted for one particular sample run. In this sample run, the noise sequences are the same for the optimal with known parameters, CE, and dual controls. If the parameters are known exactly, learning is obviously not required, and thus the control action will have only the control objective. Notice that, to achieve this objective, the control energy should be kept small in the beginning and become larger toward the terminal time. In general, the CE control has this characteristic (the overshoots at about 10 and 13 are due to stochastic effects). However, the dual control acts quite differently; namely, at the initial time,

[2] This small number of Monte Carlo runs do not make possible inferences with high confidence; however, as pointed out at the beginning of Section V, the main purpose of these simulations is to provide understanding of the dual nature of the control.

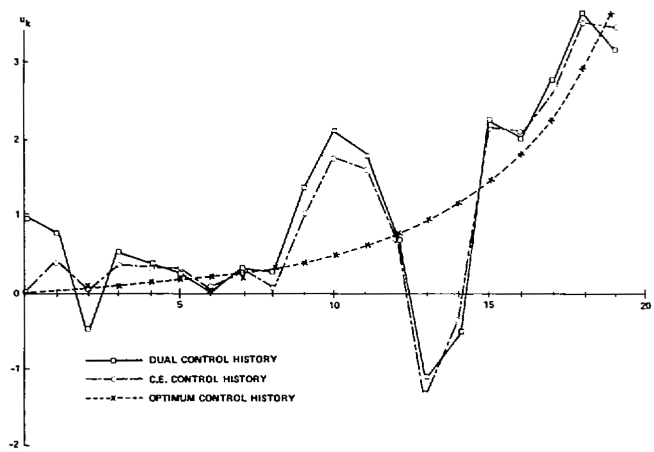


Fig. 1. Control history (one sample run for case 1).

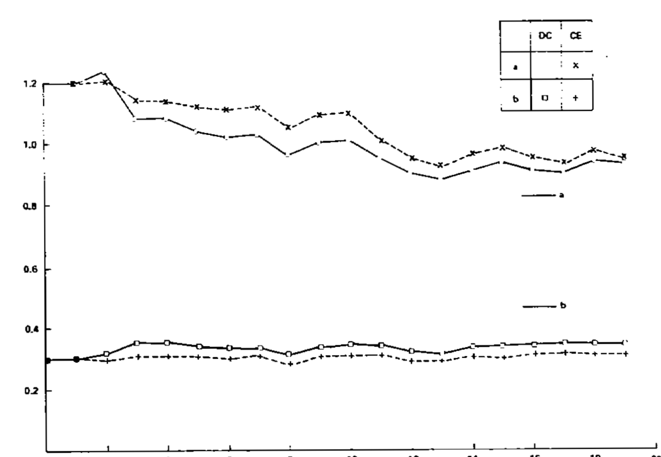


Fig. 2. Learning history for dual control and CE control (one sample run for case 1).

the control value is quite far from zero. Thus, the dual control allocates some energy specifically intended for the learning objective in the beginning; this indicates that, in the initial period, achieving the control objective and learning are in conflict. Later, the control energy is build-

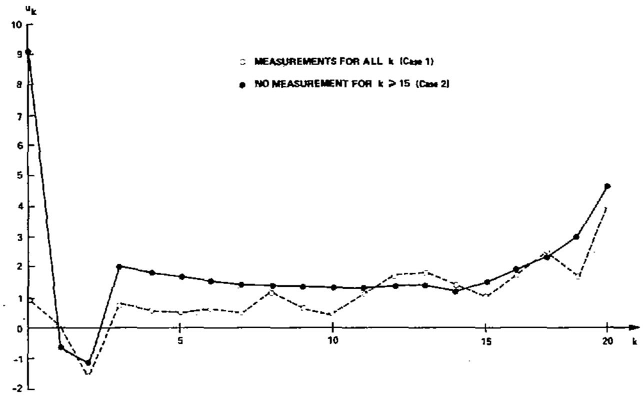


Fig. 3. Dual control history for case 1 and case 2 (one sample run).

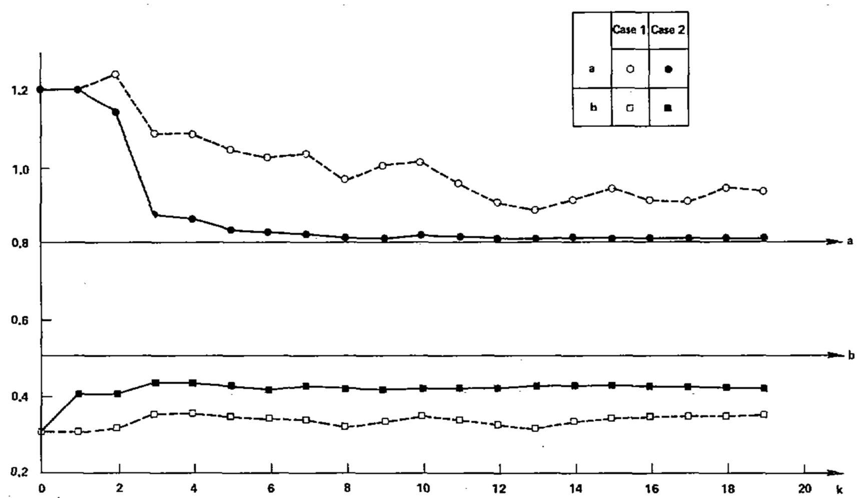


Fig. 4. Evolution of the parameter estimates obtained by applying the dual control (case 1 and case 2; one sample run).

ing up in order to achieve the control objective. Since large control energy will excite the modes and improve the signal-to-noise ratio, it will promote learning. Thus, in this period, learning *and* controlling are not in conflict. This explains why the CE control does have good estimates at the terminal time. In this case, learning is "accidental."

To illustrate this point, we shall see what happens if learning is not possible in the final period. This is shown in case 2. Table II shows the results of artificially terminating the final learning period. The CE control does very poorly in estimation, and, consequently, very poorly in achieving the control objective. This is reflected by the very large average cost and its standard deviation. The dual control, on the other hand, still performs reasonably well due to anticipation of the open-loop period at the end.

In Fig. 3, the control histories for another sample run with identical noise sequence for dual control case 1 and case 2 are plotted. Note that, at the initial time, more energy is allocated to learning in case 2 than in case 1. This is so because, in case 2, the derivation of the dual control takes into account that no learning is possible for $k \geq 15$, and thus any large control at the end will not help in learning; therefore, in order to achieve good control performance, a large amount of energy must be invested for pure learning purposes during the initial period to

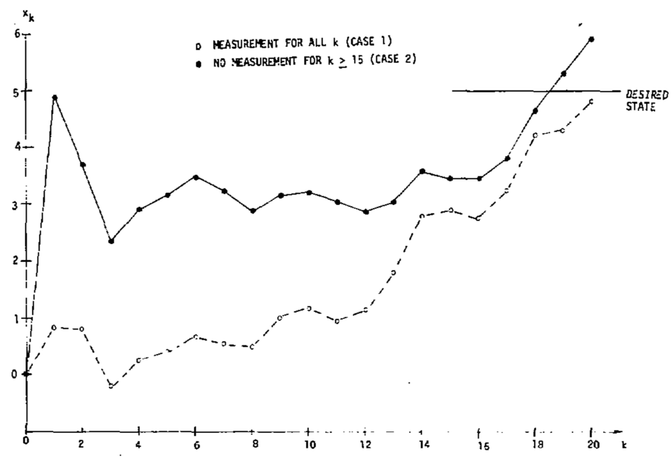


Fig. 5. Evolution of the state obtained by applying the dual control (case 1 and case 2; one sample run).

excite the system and to improve the signal-to-noise ratio. This is illustrated in Figs. 3–5. As a result, the dual control achieves a much lower average cost and, at the same time, a much more reliable control strategy than does the CE control.

This active learning characteristic is a distinguishing feature of the dual control strategy, which depends not only on past observation information but also on the future observation program; therefore, the control value will differ, depending on whether or not future observations will be made. Note that such a feature is not possessed by any of the existing suboptimal schemes suggested in the literature.

In the example, a second-order filter [2] is used for on-line estimation of state and parameters. This estimation scheme is quite effective. Clearly, one may expect better performance if one uses a more sophisticated estimation algorithm; e.g., via optimum filters [7], [29], [3].

One important point to be stressed is that the dual control strategy tries to improve the performance by considering what should be done *before* as well as *after* the parameters are identified; whereas the CE control strategy only tells what should be done after the parameters are identified.

Remark 1: First, some comments will be made on the computational requirements of the wide-sense adaptive dual control algorithm. This is highly dependent on the problem under consideration and also on the search algorithm being used. In the above example, a quadratic search method was used, and it was noted that more search was performed in the beginning than in the final period. The time required by one run of the example discussed above on a Univac 1108 was 0.5 s for the optimal control with known parameters, 1 s for the CE control, and 3 s for the dual control.[3]

Remark 2: The CE control law is a very crude suboptimal method. Other more sophisticated suboptimal control methods reported in the literature (e.g., [10], [27], [6], [9], [23], [26], [18]) may give a better performance than the CE control. We do not intend to compare the wide-sense adaptive control law with all the other suboptimal methods quantitatively, but we would like to provide a reasonable argument for why we would expect the dual control method to perform better. The distinguishing characteristic of the present control method is the "active" learning feature, which the other control strategies lack. The wide-sense adaptive dual control regulates its learning as required by the control objective. As noted in case 2, the improvement of the performance when using the dual control method is due to the fact that the knowledge of the future observation program has propagated to the initial time, and this causes a completely different control action. This closed-loop feature does not appear in the other suboptimal control methods mentioned above.

Remark 3: The example here provides an understanding of the dual control concept. It is believed that the approach described in this paper will eventually lead to an implementable control scheme that is a close approximation of the *optimal closed-loop* control law.

VI. Conclusion

This paper describes an approach for obtaining a control algorithm that exhibits the dual characteristic of appropriately distributing the control energy for learning and control purposes. The approach is an approximation based on the principle of optimality that retains the closed-loop feature of the control. A wide-sense adaptive dual control is derived that possesses the distinguishing characteristic of regulating its learning as required by the control objective. A simple example is used to demonstrate the computational feasibility of the algorithm and its performance level when applied to a specific problem, and to provide some insight into the dual control theory.

[3] The algorithms were not coded in the most efficient manner; these figures are meant only to give some idea of the computational requirements of the dual control algorithm.

Appendix

The Optimal Perturbation Control

Denote the optimal incremental cost-to-go by

$$\Delta J_o^*(Y^j,j) = \min_{\delta u(j)} E\Big\{ \min_{\delta u(j+1)} E\Big[\cdots \min_{\delta u(N-1)} E \cdot (\Delta J_o(Y^N,N)\,|\,Y^{N-1})\,|\cdots|\,Y^{j+1}\Big]Y^j\Big\}. \quad \text{(A.1)}$$

The alternating minimizations and expectations in the above reflect the closed-loop property of the control [6]. The principle of optimality leads to

$$\begin{aligned}\Delta J_o^*(Y^j,j) = \min_{\delta u(j)} \{&L_{o,x}'(j)\delta\hat{x}(j|j) \\ &+ \tfrac{1}{2}\delta\hat{x}'(j|j)L_{o,xx}(j)\delta\hat{x}(j|j) + \phi_{o,u}'(j)\delta\boldsymbol{u}(j) \\ &+ \tfrac{1}{2}\delta\boldsymbol{u}'(j)\phi_{o,uu}(j)\delta\boldsymbol{u}(j) + \tfrac{1}{2}\operatorname{tr}[L_{o,xx}(j)\boldsymbol{\Sigma}_o(j|j)] \\ &+ E[\Delta J_o^*(Y^{j+1},j+1)\,|\,Y^j]\}. \end{aligned} \quad \text{(A.2)}$$

The covariance $\boldsymbol{\Sigma}_o(j|j)$ is propagated, independently of the perturbation control, according to the extended Kalman filter equation (see, e.g., [15]).

Take $\Delta J_o^*(Y^j,j)$ of the form

$$\Delta J_o^*(Y^j,j) = \hat{g}_o(j) + \boldsymbol{p}_o'(j)\delta\hat{x}(j|j) + \tfrac{1}{2}\delta\hat{x}'(j|j)\boldsymbol{K}_o(j)\delta\hat{x}(j|j). \quad \text{(A.3)}$$

Substituting (A.3) into (A.2), the minimization of the right-hand side of (A.2) is obtained by letting

$$\begin{aligned}\frac{\partial}{\partial \boldsymbol{u}}\,[&\phi_{o,u}'(k) + \tfrac{1}{2}\delta\boldsymbol{u}'(k)\phi_{o,uu}(k)\delta\boldsymbol{u}(k) \\ &+ \boldsymbol{p}_o'(k+1)\delta\hat{x}(k+1|k) \\ &+ \tfrac{1}{2}\delta\hat{x}'(k+1|k)\boldsymbol{K}_o(k+1)\delta\hat{x}(k+1|k)] = \boldsymbol{0}; \end{aligned} \quad \text{(A.4)}$$

where, to second order,

$$\begin{aligned}\delta\hat{x}(j+1|j) = &f_{o,x}(j)\delta\hat{x}(j|j) + f_{o,u}(j)\delta\boldsymbol{u}(j) \\ &+ \tfrac{1}{2}\sum_{i=1}^{n} \boldsymbol{e}_i \operatorname{tr}\{f_{o,xx}{}^i(j)[\boldsymbol{\Sigma}_o(j|j) \\ &+ \delta\hat{x}(j|j)\delta\hat{x}'(j|j)]\} + \sum_{i=1}^{n} \boldsymbol{e}_i\delta\boldsymbol{u}'(j)f_{o,ux}{}^i(j) \\ &\cdot\delta\hat{x}(j|j) + \tfrac{1}{2}\sum_{i=1}^{n}\boldsymbol{e}_i\delta\boldsymbol{u}'(j)f_{o,uu}{}^i(j)\delta\boldsymbol{u}(j). \end{aligned} \quad \text{(A.5)}$$

Substitute (A.5) into (A.4) and retain terms only up to second order; the optimum $\delta\boldsymbol{u}_o^*(j)$ is given by

$$\begin{aligned}\delta u_o^*(j) = -\Bigg[\phi_{o,uu}(j) + f_{o,u}'(j)K_o(j+1)f_{o,u}(j) \\ + \sum_{i=1}^{n} p_o'(j+1)e_i f_{o,uu}{}^i(j)\Bigg]^{-1}\Bigg\{\Bigg[f_{o,u}'(j)K_o(j+1) \\ \cdot f_{o,x}(j) + \sum_{i=1}^{n} p_o'(j+1)e_i f_{o,ux}{}^i(j)\Bigg]\delta\hat{x}(j|j) \\ + f_{o,u}'(j)p_o(j+1) + \phi_{o,u}(j)\Bigg\}, \qquad (A.6)\end{aligned}$$

which yields (3.9).

Substituting (A.6) and (A.3) into (A.2) (keeping only up to second-order terms) and equating terms in zeroth-, first-, and second-order of $\delta\hat{x}(k|k)$, one has, by using the definition of $H_o(j)$, (3.10), the equations for $\hat{g}_o(j)$, $p_o(j)$, and $K_o(j)$:

$$\begin{aligned}\hat{g}_o(j) = \hat{g}_o(j+1) - \tfrac{1}{2}H_{o,u}'(j)[H_{o,uu}(j) \\ + f_{o,u}'(j)K_o(j+1)f_{o,u}(j)]^{-1}H_{o,u}(j) \\ + \tfrac{1}{2}\operatorname{tr}\{H_{o,xx}(j)\Sigma_o(j|j) \\ + [\Sigma_o(j+1|j) - \Sigma_o(j+1|j+1)]K_o(j+1)\}; \\ \hat{g}_o(N) = \tfrac{1}{2}\operatorname{tr}[\psi_{o,xx}\Sigma_o(N|N)] \qquad (A.7)\end{aligned}$$

$$\begin{aligned}p_o(j) = H_{o,x}(j) - [f_{o,u}'(j)K_o(j+1)f_{o,x}(j) + H_{o,ux}(j)]' \\ \cdot [H_{o,uu}(j) + f_{o,u}'(j)K_o(j+1)f_{o,u}(j)]^{-1}H_{o,u}(j); \\ p_o(N) = \psi_{o,x} \qquad (A.8)\end{aligned}$$

$$\begin{aligned}K_o(j) = f_{o,x}'(j)K_o(j+1)f_{o,x}(j) - [f_{o,u}'(j)K_o(j+1)f_{o,x}(j) \\ + H_{o,ux}(j)]'[H_{o,uu}(j) + f_{o,u}'(j)K_o(j+1)f_{o,u}(j)]^{-1} \\ \cdot [f_{o,u}'(j)K_o(j+1)f_{o,x}(j) + H_{o,ux}(j)] + H_{o,xx}(j); \\ K_o(N) = \psi_{o,xx}. \qquad (A.9)\end{aligned}$$

The resulting optimum cost if $U_o(k,N-1)$ is selected is thus given by

$$\begin{aligned}J^*[k,U_o(k,N-1)] = J_o(k) + \Delta J_o^*(Y^k,k) = J_o(k) \\ + \hat{g}_o(k) + p_o'(k)\delta\hat{x}(k|k) \\ + \tfrac{1}{2}\delta\hat{x}'(k|k)K_o(k+1)\delta\hat{x}(k|k). \qquad (A.10)\end{aligned}$$

To stress the estimation performance reflected in $J^*[k,U_o(k,N-1)]$, define $g_o(j)$, $j = k+1,\cdots,N$, according to

$$\begin{aligned}g_o(j) = g_o(j+1) - \tfrac{1}{2}H_{o,u}(j)[H_{o,uu}(j) + f_{o,u}'(j) \\ \cdot K_o(j+1)f_{o,u}(j)]^{-1}H_{o,u}(j); \qquad g_o(N) = 0. \qquad (A.11)\end{aligned}$$

Then, by (A.7), (A.10), and (A.11), $J^*[k,U_o(k,N-1)]$ can be expressed alternatively as

$$\begin{aligned}J^*[k,U_o(k,N-1)] \\ = J_o(k) + g_o(k) + \tfrac{1}{2}\operatorname{tr}\Bigg\{\psi_{o,xx}\Sigma_o(N|N) \\ + \sum_{j=k}^{N-1}\{H_{o,xx}(j)\Sigma_o(j|j) + [\Sigma_o(j+1|j) \\ - \Sigma_o(j+1|j+1)]K_o(j+1)\}\Bigg\} + p_o'(k)\delta\hat{x}(k|k) \\ + \tfrac{1}{2}\delta\hat{x}'(k|k)K_o(k+1)\delta\hat{x}(k|k), \qquad (A.12)\end{aligned}$$

which is (3.13).

In the wide-sense adaptive dual control consideration, for $j \geq k+1$, only perturbation analysis will be carried out along the nominal associated with $\hat{x}(k+1|k)$; thus the cost of applying $u(k)$ can be approximated by

$$\begin{aligned}I_d[u(k)] \simeq E\{\phi[u(k),k] + L[x(k),k] + J_o(k+1) \\ + g_o(k+1) + p_o'(k+1)[\hat{x}(k+1|k+1) \\ - x_o(k+1)] + \tfrac{1}{2}[\hat{x}(k+1|k+1) \\ - x_o(k+1)]'K_o(k+1)[\hat{x}(k+1|k+1) \\ - x_o(k+1)]|Y^k\}. \qquad (A.13)\end{aligned}$$

Equation (4.8) can now be obtained by noting that

$$\begin{aligned}E\{\phi[u(k),k] + J_o(k+1) + g_o(k+1)|Y^k\} \\ = \phi[u(k),k] + J_o(k+1) + g_o(k+1) \\ E\{p_o'(k+1)[\hat{x}(k+1|k+1) - x_o(k+1)]|Y^k\} \\ = p_o'(k+1)[\hat{x}(k+1|k) - x_o(k+1)] \\ E\{[\hat{x}(k+1|k+1) - x_o(k+1)]'K_o(k+1) \\ \cdot [\hat{x}(k+1|k+1) - x_o(k+1)]|Y^k\} \\ = [\hat{x}(k+1|k) - x_o(k+1)]K_o(k+1)[\hat{x}(k+1|k) \\ - x_o(k+1)] + \operatorname{tr}\{K_o(k+1)[\Sigma(k+1|k) \\ - \Sigma(k+1|k+1)]\}. \qquad (A.14)\end{aligned}$$

Acknowledgment

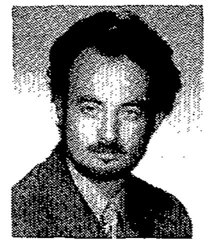
The authors wish to thank Dr. R. E. Larson and Dr. A. J. Tether, of Systems Control, Inc., for useful discussions and for their continuing support.

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] D. L. Alspach, "Dual control based on approximate *a posteriori* density functions," in *1972 Joint Automatic Control Conf., Preprints.*
[5] R. Bellman, *Adaptive Control Processes: A Guided Tour.* Princeton, N.J.: Princeton Univ. Press, 1961.
[6] 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.
[7] 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.
[8] A. E. Bryson and Y. C. Ho, *Applied Optimal Control.* Waltham, Mass.: Blaisdell, 1969.
[9] R. E. Curry, "A new algorithm for suboptimal stochastic control," *IEEE Trans. Automat. Contr.* (Short Papers), vol. AC-14, pp. 533–536, Oct. 1969.
[10] W. Denham, "Choosing the nominal path for a dynamic system with random forcing function to optimize statistical performance," Div. Eng. Appl. Phys., Harvard Univ., Cambridge, Mass., Tech. Rep. 449, 1964.
[11] J. L. Doob, *Stochastic Processes.* New York: Wiley, 1953.
[12] S. Dreyfus, "Some types of optimal control of stochastic systems," *SIAM J. Contr.*, vol. 2, no. 1, pp. 120–134, 1964.
[13] A. A. Feldbaum, *Optimal Control Systems.* New York: Academic, 1965.
[14] P. D. Joseph and J. T. Tou, "On linear control theory," *AIEE Trans.* (Appl. Ind.), vol. 80, pp. 193–196, Sept. 1961.
[15] A. Jazwinski, *Stochastic Processes and Filtering Theory.* New York: Academic, 1970.
[16] O. L. R. Jacobs and J. W. Patchell, "Caution and probing in stochastic control," *Int. J. Contr.*, 1972.
[17] R. E. Larson and E. Tse, "Modal trajectory estimation and

parallel computers," in *Proc. 2nd Symp. Nonlinear Estimation Theory*, 1971.
[18] D. G. Lainiotis, T. N. Upadhyay, and J. G. Deshpande, "Optimal adaptive control of linear systems," in *Proc. 1971 IEEE Conf. Decision and Control.*
[19] D. G. Luenberger, *Introduction to Linear and Nonlinear Programming.* Reading, Mass.: Addison-Wesley, 1973.
[20] L. Meier, "Combined optimal control and estimation," in *Proc. 3rd Annu. Allerton Conf. Circuit and System Theory*, 1965.
[21] ——, "Some stochastic control theory," in *Information Requirements for Supersonic Transport Operation*, NASA, Contractor Rep. CR-1570, pp. 91–95, July 1970.
[22] L. Meier, III, R. E. Larson, and A. J. Tether, "Dynamic programming for stochastic control of discrete systems," *IEEE Trans. Automat. Contr. (Special Issue on Linear-Quadratic-Gaussian Problem)*, vol. AC-16, pp. 767–775, Dec. 1971.
[23] 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.
[24] C. Streibel, "Sufficient statistics in the optimum control of stochastic systems," *J. Math. Anal. Appl.*, vol. 12, no. 3, pp. 576–592, 1965.
[25] H. W. Sorenson, "Kalman filtering techniques," in *Advances in Control Systems*, vol. 3, C. T. Leondes, Ed. New York: Academic, 1966.
[26] G. Stein and G. N. Saridis, "A parameter-adaptive control technique," *Automatica*, vol. 5, pp. 731–740, Nov. 1969.
[27] 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.
[28] E. Tse, "On the optimal control of stochastic linear systems," *IEEE Trans. Automat. Contr. (Special Issue on Linear-Quadratic-Gaussian Problem)*, vol. AC-16, pp. 776–785, Dec. 1971.
[29] ——, "Parallel computation of the conditional mean state estimate for nonlinear systems," in *Proc. 2nd Symp. Nonlinear Estimation Theory* (San Diego, Calif.), 1971.
[30] ——, "On the optimal control of linear systems with incomplete information," Electron. Syst. Lab., Mass. Inst. Technol., Cambridge, Rep. ESL-R-412, Jan. 1970.
[31] E. Tse, R. Dressler, and Y. Bar-Shalom, "Application of adaptive tuning of filters to exoatmospheric target tracking," in *Proc. 3rd Symp. Nonlinear Estimation Theory* (San Diego, Calif.), 1972.
[32] E. Tse, L. Meier, and Y. Bar-Shalom, "Dual control of stochastic nonlinear systems," in *Proc. 1971 IEEE Conf. Decision and Control*, 1971.
[33] E. Tse and Y. Bar-Shalom, "An actively adaptive control for linear systems with random parameters via the dual control approach," this issue, pp. 109–117.
[34] D. R. Vander Stoep, "Trajectory shaping for the minimization of state-variable estimation errors," *IEEE Trans. Automat. Contr.* (Short Papers), vol. AC-13, pp. 284–286, June 1968.

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) was born on May 11, 1941. He received the B.S. and M.S. degrees from the Technion—Israel Institute of Technology, Haifa, Israel, in 1963 and 1967, respectively, and the Ph.D. degree from Princeton University, Princeton, N.J., in 1970, all in electrical engineering.

During 1963–1966 he served in the Israel Defense Army as an Electronics Engineer. From 1966 to 1967 he was a Teaching Assistant in the Department of Electrical Engineering, Technion. Between 1967 and 1970 he was at Princeton University, first as Research Assistant, and then as Research Associate. During the academic year 1968–1969 he was appointed Charles Grosvenor Osgood Fellow in recognition of outstanding performance. Since 1970 he has been with Systems Control, Inc., Palo Alto, Calif. His research interests are in the areas of estimation theory, stochastic adaptive control, modeling, analysis, and control of real-life systems.

Dr. Bar-Shalom is a member of Eta Kappa Nu.

Lewis Meier, III (S'61–M'65) was born in Indianapolis, Ind., on April 11, 1938. He received the B.S. degree in engineering sciences with highest honors and the M.S. degree, both from Purdue University, Lafayette, Ind., in 1960 and 1961, respectively, and the Ph.D. degree in electrical engineering from Stanford University, Stanford, Calif., in 1965.

From 1958 to 1961 he studied the analysis and simulation of inertial guidance systems at the U.S. Naval Avionic Facility, Indianapolis, Ind. From 1962 to 1964 he investigated system reliability concepts and mission analysis techniques at the Jet Propulsion Laboratory, Pasadena, Calif. He was a Research Engineer at the Stanford Research Institute from 1965 to 1971, where he worked on a variety of projects involving applications of estimation theory, stochastic control theory, mathematical optimization theory, and differential game theory. Presently he is a Senior Research Engineer with Systems Control, Inc., Palo Alto, Calif., where he is engaged in the application of modern control theory to the solution of system analysis problems.

Dr. Meier is an Associate Editor of *Automatica* and a member of Tau Beta Pi and Sigma Xi.