Volume Title: Annals of Economic and Social Measurement, Volume 5, number 3

Volume Author/Editor: Sanford V. Berg, editor

Volume Publisher: NBER

Volume URL: http://www.nber.org/books/aesm76-3

Publication Date: July 1976

Chapter Title: Caution, Probing, and the Value of Information in the Control of Uncertain Systems

Chapter Author: Yaakov Bar-Shalom, Edison Tse

Chapter URL: http://www.nber.org/chapters/c10483

Chapter pages in book: (p. 323 - 337)

Annals of Economic and Social Measurement, 5/3, 1976

CAUTION, PROBING, AND THE VALUE OF INFORMATION IN THE CONTROL OF UNCERTAIN SYSTEMS*

BY YAAKOV BAR-SHALOM AND EDISON TSE

This paper discusses the control of nonlinear stochastic systems and, in particular, linear systems with unknown parameters. It is shown how the optimal policy utilizes preposterior analysis to obtain the control values. The stochastic nature of the problem leads to the probing and caution properties of the control. Explicit expressions of the probing and caution terms in a stochastic control problem are presented. These terms are obtained by a closed-loop approximation of the stochastic dynamic programming equation. An approximate value of information can be evaluated and the benefit to be derived from probing (experimentation) can be traded off against its cost. The interplay between caution and probing is illustrated by an example. The performance of the closed-loop control obtained from the above approximation of the stochastic dynamic programming is compared with several other suboptimal controls as well as the optimal one.

1. INTRODUCTION

Many models used in economic decision making are assumed to be linear but it is generally accepted that the parameters of these models are imperfectly known and sometimes time-varying. The decision (or control) problem in linear systems with unknown parameters is actually a nonlinear stochastic control problem. The optimal solution of all but a few stochastic control problems is not known and cannot be obtained numerically because of the dimensionality associated with the numerical solutions [B1]. The notable few exceptions are the linear-quadratic problem [S3, T5, A1], and the exponential-linear-quadratic problem [S2]. The optimal stochastic control is obtained by applying the principle of optimality [B1] and, unless a closed-form solution can be guessed and verified, as in the above two problems, suboptimal solutions are usually sought. Since one has to give up on the optimality, it is desirable to obtain a solution that has, at least, the features (structural characteristics) of the optimal solution. Therefore, it is important to be able to define the structure of the optimal solution, i.e., what types of information can be available to the controller and how are they used.

It is well known that, in stochastic control problems, utilizing observations improves the performance over the open-loop controls because the utilization of measurements on the system reduces the uncertainty (see, e.g., [M1]). A non-anticipative control cannot, obviously, use future observations that can benefit the performance of the control; however, the probabilistic description of these observations can reveal the "value of the future information" before they are actually taken. Therefore, while being non-anticipative, a control can still "look into the future" and utilize what is presently known about the information to be obtained later. This is called "preposterior analysis" [R1].

This type of control, called *closed-loop* because it "anticipates" that the loop will be closed in the future, is to be distinguished from the *feedback* type (see [B3, B4] for details). The latter utilizes the past measurements, but it does not

* Research supported by NSF under Grant GS32271.

"anticipate" (via a probabilistic description) the future measurements. Since most of the existing suboptimal algorithms (where the optimal is not known) for many problems of practical importance are of the feedback type, it is of interest to see how closed-loop type algorithms can be derived and what improvements can be obtained in the performance over existing feedback algorithms.

It is in light of this property of anticipating the value of future information by using the statistics of the future measurements that a closed loop stochastic control algorithm is examined. Namely, the algorithm developed in [T1], [T2] for a large class of systems that includes linear systems with unknown parameters is discussed in Section 2 and a rigorous derivation of it is given. It is shown how the control can carry out experiments, i.e., it can probe the system in anticipation of the value of the information to be derived from future observations. This probing of the system is done by utilizing Feldbaum's "dual effect" [F1]. A control is said to have a dual effect if, in addition to its effect on the state, it can affect the estimation—then it can be used for "active information storage" [F1], or, in other words, it can actively learn. An explicit expression of the value of future information which weights the future uncertainty in the state is derived. The other aspect of controlling a stochastic system is the need for caution [F1, J1]. A caution term is obtained which includes the uncertainties that cannot be affected by the control but their weightings in the cost depend on it.

The conflict between caution and probing is illustrated in a numerical example in Section 3. Comparisons of the closed-loop control algorithm with the certainty equivalence algorithm, the adaptive algorithms of MacRae [M2] and Chow [C2], the open-loop-optimal feedback and the optimal one are presented. The optimal control was obtained by extensive Monte Carlo simulation such that the results of the comparison can be stated with very high confidence.

2. A Stochastic Closed-loop Control Algorithm for Nonlinear Systems

Feldbaum [F1], Aoki [A1] and Dreyfus [D1] stressed the importance of closed-loop controllers. They showed in several examples the improved performance one can achieve due to the closed-loop property. In this section, the suboptimal stochastic closed loop control algorithm ("dual control") developed in [T1, T2] is discussed in light of the concepts presented in [B3] and a rigorous derivation is presented.

Consider the system whose state, an n-vector, evolves according to the equation

$$\mathbf{x}(k+1)=\mathbf{f}[k,\mathbf{x}(k),\mathbf{u}(k)]+\mathbf{v}(k) \qquad k=0,1,\ldots,N-1 \tag{2.1}$$

and with observations (m-vector)

$$\mathbf{y}(k)=\mathbf{h}[k,\mathbf{x}(k)]+\mathbf{w}(k) \qquad k=1,\ldots,N \tag{2.2}$$

where $\mathbf{x}(0)$ is the initial condition, a random variable with mean $\hat{\mathbf{x}}(0|0)$ and covariance $\boldsymbol{\Sigma}(0|0)$; $\{\mathbf{v}(k)\}$ and $\{\mathbf{w}(k)\}$ are the sequences of process and measurement noises, respectively, mutually independent, white and with known statistics up to second order. For simplicity we shall assume they are zero-mean. For the

purpose of the control algorithm to be derived no assumptions about the distributions of these random variables are needed.

The cost function is taken as

$$C(N) = \psi[\mathbf{x}(N)] + \sum_{k=0}^{N-1} L[\mathbf{x}(k), k] + \phi(\mathbf{u}(k), k] \tag{2.3}$$

and the performance index is

$$J(N) = E\{C(N)\}. \tag{2.4}$$

In the case of a linear system with unknown parameters, $\mathbf{x}$ is the "augmented" state, a stacked vector that includes the unknown parameters.

Rather than using the exact information state $\{Y^k, U^{k-1}\}$ the following approximate "wide-sense" information state is used.

$$\mathcal{P}^k = \{\hat{\mathbf{x}}(k|k), \boldsymbol{\Sigma}(k|k)\} \tag{2.5}$$

i.e., the conditional mean and covariance of $\mathbf{x}(k)$. The computation of $\mathcal{P}^k$ can be done by a number of approximate methods, e.g., extended Kalman filter, second order filter, or non-linear filter.

Assume now that the system is at time k and a closed-loop control [B3] is to be computed using $\mathcal{P}^k$ *and* the present knowledge (statistical) about the future observations.

The cost-to-go for the last $N-k$ steps is

$$C(N-k) = \psi[\mathbf{x}(N)] + \sum_{j=k}^{N-1} L[\mathbf{x}(j), j] + \phi(\mathbf{u}(j), j]. \tag{2.6}$$

The principle of optimality [B1] with the information state (2.5) yields the following stochastic dynamic programming equation for the closed-loop-optimal expected cost-to-go at time k

$$J^*(N-k) = \min_{\mathbf{u}(k)} E\{L[\mathbf{x}(k), k] + \phi[\mathbf{u}(k), k] + J^*(N-k-1)|\mathcal{P}^k\}. \tag{2.7}$$

The main problem is to obtain an approximate expression for $J^*(N-k-1)$ while preserving its closed-loop feature, i.e., this expression should incorporate the "value" of the future observations. Note that $J^*(N-k-1)$ is obtained by the closed-loop minimization [B1, B3] of $C(N-k-1)$. In order to find an explicit solution to this minimization, the cost-to-go C for the last $N-k-1$ steps is expanded about a nominal trajectory as follows. Let the nominal trajectory be

$$\mathbf{x}_0(j+1) = \mathbf{f}[j, \mathbf{x}_0(j), \mathbf{u}_0(j)] \qquad j = k+1, \ldots, N-1 \tag{2.8}$$

where $\mathbf{u}_0(j)$, $j = k+1, \ldots, N-1$ is a sequence of nominal controls (to be discussed later) and the initial condition for this nominal trajectory is taken as

$$\mathbf{x}_0(k+1) = \hat{\mathbf{x}}[k+1|k; \mathbf{u}(k)] \tag{2.9}$$

i.e., the predicted value of the state at $k+1$ given $\mathcal{P}^k$ and the control (yet to be found) $\mathbf{u}(k)$. The expansion of the cost-to-go (2.6) with k replaced by $k+1$ is, with terms up to second order,

$$C(N-k-1) = C_0(N-k-1) + \Delta C_0(N-k-1) \tag{2.10}$$

where

$$C_0(N-k-1) \triangleq \psi[\mathbf{x}_0[N]] + \sum_{j=k+1}^{N-1} L[\mathbf{x}_0(j), j] + \phi[\mathbf{u}_0(j), j] \tag{2.11}$$

is the cost along the nominal and

$$\begin{aligned}\Delta C_0(N-k-1) \triangleq{}& \psi'_{0,\mathbf{x}}\,\delta\mathbf{x}(N) + \tfrac{1}{2}\delta\mathbf{x}'(N)\psi_{0,\mathbf{xx}}\,\delta\mathbf{x}(N) + \sum_{j=k+1}^{N-1} [L'_{0,\mathbf{x}}(j)\,\delta\mathbf{x}(j) \\ &+ \tfrac{1}{2}\delta\mathbf{x}'(j)L_{0,\mathbf{xx}}(j)\,\delta\mathbf{x}(j) + \phi'_{0,\mathbf{u}}(j)\,\delta\mathbf{u}(j) \\ &+ \tfrac{1}{2}\delta\mathbf{u}'(j)\phi_{0,\mathbf{uu}}(j)\,\delta\mathbf{u}(j)]\end{aligned} \tag{2.12}$$

is the variation of the cost about the nominal. The notations $L_{0,\mathbf{x}}$, $L_{0,\mathbf{xx}}$ stand for the gradient and Hessian of L w.r.t. $\mathbf{x}$ evaluated along the nominal trajectory and

$$\delta\mathbf{x}(j) = \mathbf{x}(j) - \mathbf{x}_0(j) \tag{2.13a}$$

$$\delta\mathbf{u}(j) = \mathbf{u}(j) - \mathbf{u}_0(j) \tag{2.13b}$$

are the perturbed state and control, respectively.

The approximation of the closed-loop-optimal expected cost-to-go for the last $N-k-1$ steps is done now as follows:

$$\begin{aligned}J^*(N-k-1) &= \min_{\mathbf{u}(k+1)} E\{\ldots \min_{\mathbf{u}(N-1)} E[C(N-k-1)|\mathcal{P}^{N-1}]\ldots|\mathcal{P}^{k+1}\} \\ &= J_0(N-k-1) + \Delta J_0^*(N-k-1)\end{aligned} \tag{2.14}$$

where

$$J_0(N-k-1) = C_0(N-k-1) \tag{2.15}$$

$$\Delta J_0^*(N-k-1) = \min_{\delta\mathbf{u}(k+1)} E\{\ldots \min_{\delta\mathbf{u}(N-1)} E[\Delta C_0(N-k-1)|\mathcal{P}^{N-1}]\ldots|\mathcal{P}^{k+1}\} \tag{2.16}$$

Note that the closed-loop minimization of (2.16) is over a cost quadratic in $\delta\mathbf{x}(i+1)$, $\delta\mathbf{u}(i)$, $i = k+1, \ldots, N-1$ as can be seen from (2.12). Furthermore, from the definition of the nominal trajectory (2.8) and the dynamics of the system (2.1), the perturbations (2.13) obey the following dynamic equation (with terms up to second order; $f^i_{0,\mathbf{xx}}$ denotes the Hessian of the ith component of $\mathbf{f}$, $i = 1, \ldots, n$).

$$\begin{aligned}\delta\mathbf{x}(j+1) = {}& \mathbf{f}'_{0,\mathbf{x}}(j)\,\delta\mathbf{x}(j) + \mathbf{f}'_{0,\mathbf{u}}(j)\,\delta\mathbf{u}(j) \\ &+ \sum_{i=1}^{n} \mathbf{e}_i[\tfrac{1}{2}\delta\mathbf{x}'(j)f^i_{0,\mathbf{xx}}(j)\,\delta\mathbf{x}(j) \\ &+ \delta\mathbf{u}'(j)f^i_{0,\mathbf{ux}}(j)\,\delta\mathbf{x}(j) \\ &+ \tfrac{1}{2}\delta\mathbf{u}'(j)f^i_{0,\mathbf{uu}}(j)\,\delta\mathbf{u}(j)] + \mathbf{v}(j) \qquad j = k+1, \ldots, N-1\end{aligned} \tag{2.17}$$

with initial condition

$$\delta\mathbf{x}(k+1) = \mathbf{x}(k+1) - \mathbf{x}_0(k+1). \tag{2.18}$$

Thus, the problem defined by (2.16) consists of the minimization of the quadratic cost (2.12) for the quadratic system (2.17) and is somewhat similar to

the linear-quadratic problem. Up to terms of second order, the solution of this problem can be assumed to be of the form

$$\Delta J_0^*(N-k-1) = g_0(k+1) + E\{\mathbf{p}_0'(k+1)\,\delta\mathbf{x}(k+1) + \tfrac{1}{2}\,\delta\mathbf{x}'(k+1)\mathbf{K}_0(k+1)\,\delta\mathbf{x}(k+1)|\mathcal{P}^{k+1}\}. \tag{2.19}$$

The proof by induction of the above is given in the Appendix. The algorithm for control of linear unknown systems with learning developed by Chow [C2], which is also of the closed-loop type, approximates the entire optimal expected cost-to-go by a second order expansion about a tentative path using numerical derivatives. This expansion is a function of the state, which is assumed to be observed perfectly. In contradistinction to this, the procedure discussed here utilizes a quadratic expansion of the cost prior to taking the expectations and the resulting perturbation cost ΔC_0 can be minimized explicitly; a numerical minimization is required once at every period when obtaining the present control as will be seen later.

To emphasize the close-loop property of ΔJ_0^*, i.e., the manner in which it is a function of the future uncertainties, it is rewritten as follows (the detailed derivations are presented in the Appendix).

$$\begin{aligned}\Delta J_0^*(N-k-1) = \gamma_0(k+1) + E\{&\mathbf{p}_0'(k+1)\,\delta\mathbf{x}(k+1) \\ &+ \tfrac{1}{2}\,\delta\mathbf{x}'(k+1)\mathbf{K}_0(k+1)\,\delta\mathbf{x}(k+1)|\mathcal{P}^{k+1}\} \\ &+ \tfrac{1}{2}\sum_{j=k+1}^{N-1} \operatorname{tr}[\mathbf{K}_0(j+1)\mathbf{Q}(j) + \mathcal{A}_{0,\mathbf{xx}}(j)\boldsymbol{\Sigma}_0(j|j)]\end{aligned} \tag{2.20}$$

where $\boldsymbol{\Sigma}_0(j|j)$ is the covariance of the state along the nominal trajectory and Q is the process noise covariance. The existence and uniqueness of the above solution is discussed in the Appendix.

The recursions that yield $\gamma_0(k+1)$, $\mathbf{p}(k+1)$, $\mathbf{K}(k+1)$ as well as the definition of $\mathcal{A}_{0,\mathbf{xx}}$ can also be found in the Appendix (see (A.15)–(A.17) and (A.3), (A.5)–(A.8)).

Combining (2.20) with (2.14), the stochastic dynamic programming equation (2.7) that will yield $\mathbf{u}(k)$ becomes

$$\begin{aligned}J^*(N-k) = \min_{\mathbf{u}(k)}\{&E\{L[\mathbf{x}(k), k] + \phi[\mathbf{u}(k), k] + C_0(N-k-1) + \gamma_0(k+1) \\ &+ \mathbf{p}_0'(k+1)\,\delta\mathbf{x}(k+1) + \tfrac{1}{2}\,\delta\mathbf{x}'(k+1)\mathbf{K}_0(k+1)\,\delta\mathbf{x}(k+1)|\mathcal{P}^k\} \\ &+ \tfrac{1}{2}\sum_{j=k+1}^{N-1} \operatorname{tr}[\mathbf{K}_0(j+1)\mathbf{Q}(j) + \mathcal{A}_{0,\mathbf{xx}}(j)\boldsymbol{\Sigma}_0(j|j)]\}.\end{aligned} \tag{2.21}$$

From (2.9) and 2.18) it follows that

$$E[\delta\mathbf{x}(k+1)|\mathcal{P}^k] = E[\tilde{\mathbf{x}}(k+1|k)\mathcal{P}^k] = 0 \tag{2.22}$$

and

$$E[\delta\mathbf{x}'(k+1)\mathbf{K}_0(k+1)\,\delta\mathbf{x}(k+1)|\mathcal{P}^k] = \operatorname{tr}[\mathbf{K}_0(k+1)\boldsymbol{\Sigma}(k+1|k)]. \tag{2.23}$$

Finally, dropping from (2.21) the first term which does not depend on $\mathbf{u}(k)$, and using (2.22), (2.23) the closed-loop control is obtained as

$$\mathbf{u}^{CL}(k) = \arg\min J^{CL}(N-k) \tag{2.24}$$

$$J^{CL}(N-k) \triangleq J_D(N-k) + J_C(N-k) + J_P(N-k) \tag{2.25}$$

where

$$J_D(N-k) \triangleq \phi[\mathbf{u}(k), k] + C_0(N-k-1) + \gamma_0(k+1)$$

is the deterministic part of the cost and

$$J_C(N-k) \triangleq \tfrac{1}{2}\operatorname{tr}[\mathbf{K}_0(k+1)\boldsymbol{\Sigma}(k+1|k)] + \tfrac{1}{2}\sum_{j=k+1}^{N-1}\operatorname{tr}[\mathbf{K}_0(j+1)\mathbf{Q}(j)] \tag{2.37}$$

$$J_P(N-k) \triangleq \tfrac{1}{2}\sum_{j=k+1}^{N-1}\operatorname{tr}[\mathscr{A}_{0,\mathbf{xx}}(j)\boldsymbol{\Sigma}_0(j|j)] \tag{2.38}$$

at the stochastic terms in the cost.

The first stochastic term, (2.27), reflects the effect of the uncertainty at time k and subsequent process noises on the cost. These uncertainties cannot be affected by $\mathbf{u}(k)$ but their weightings do depend on it. The effect of these uncontrollable uncertainties on the cost should be minimized by the control; this term indicates the need for the control to be cautious and thus is called *caution term*. The second stochastic term, (2.28) accounts for the effect of uncertainties when subsequent decisions will be made. As discussed in the Appendix, if the perturbation problem has a solution, then the weighting of these future uncertainties is non-negative ($\mathscr{A}_{0,\mathbf{xx}}$ is positive semidefinite). If the control can reduce by probing (experimentation) the future updated covariances, it can thus reduce the cost. The weighting matrix $\mathscr{A}_{0,\mathbf{xx}}$ yields approximately the *value of future information* for the problem under consideration. Therefore this is called the probing term.

The benefit of probing is weighted by its cost and a compromise is chosen such as to minimize the sum of the deterministic, caution and probing terms. The minimization of J^{CL} will also achieve a tradeoff between the present and future actions according to the information available at the time the corresponding decisions are made. A sufficient condition for the existence and uniqueness of the solution to the perturbation problem is that the sequence of nominal controls will minimize (2.11) subject to (2.8). In this case one also has

$$\gamma_0(k+1) = 0 \qquad k = 0, \ldots, N-1 \tag{2.29}$$

The preposterior analysis can be seen as appearing explicitly in the decision on the present control which is to be done using the ("prior") estimate $\boldsymbol{\Sigma}_0(j|j)$ of the future updated ("posterior") covariance $\boldsymbol{\Sigma}(j|j)$ of the state.

To find the closed-loop control $\mathbf{u}(k)$, the minimization of (2.25) is performed using a search procedure. At every k to each control $\mathbf{u}(k)$ for which (2.25) is evaluated during the search there corresponds a predicted state (2.9) and to this prediced state a sequence of deterministic controls is attached that defines the nominal trajectory. The cost-to-go is then evaluated by expansion about this nominal and its variation (up to second order) is minimized in a closed-loop fashion. This leads to (2.25) where the possible benefit from probing (active

learning) as well as the need for caution appear explicitly. The only use of the nominals and perturbations is to make possible the evaluation of the cost-to-go optimized in a closed-loop manner. This procedure is repeated at every time a new control is obtained. While the computational requirements of this algorithm are significantly higher than those of the certainty equivalence algorithm, problems of modest size can be handled; for example, a 20 period problem for a system with a three dimensional state and six unknown parameters took approximately 45 s for a complete run on a Univac 1108 [T2].

The performance of this algorithm, which is of the "performance adaptive" type [S1], is examined in the next section for a simple example of a linear system with an unknown parameter. Its usefulness is, however, not limited for this class of problems (see [T4] for its application to a nonlinear system which had no unknown parameters).

3. Example

Similarly to MacRae [M2], we shall consider the problem of controlling the system

$$x(k+1) = ax(k) + bu(k) + c + v(k) \qquad k = 0, 1. \tag{3.1}$$

with perfect observations of x. The parameters a and c will be assumed known. The input gain will be unknown, time-invariant, with prior mean $\hat{b}(0)$ and variance $\Sigma^{bb}(0)$. The noise sequence $v(k)$ is zero-mean, white, and with variance Q. The cost function whose expected value is to be minimized is taken as

$$C(2) = \tfrac{1}{2} \sum_{k=1}^{2} q(k)x^2(k) + ru^2(k-1) \tag{3.2}*$$

The initial state is $x(0) = 0$.

The augmented system with the state including the unknown parameter is

$$\mathbf{x}(k) = \begin{bmatrix} x_1(k) \\ x_2(k) \end{bmatrix} \triangleq \begin{bmatrix} x(k) \\ b(k) \end{bmatrix} \tag{3.3}$$

and obeys the nonlinear stochastic difference equation

$$\mathbf{x}(k+1) = \mathbf{f}[\mathbf{x}(k)] + \mathbf{v}(k) \tag{3.4}$$

where

$$\mathbf{f}[\mathbf{x}(k)] = \begin{bmatrix} ax_1(k) + x_2(k)u(k) + c \\ x_2(k) \end{bmatrix} \tag{3.5}$$

$$\mathbf{v}(k) = \begin{bmatrix} v_1(k) \\ v_2(k) \end{bmatrix} = \begin{bmatrix} v(k) \\ 0 \end{bmatrix} \tag{3.6}$$

The covariance matrix of the augmented state (3.3) is denoted by $\boldsymbol{\Sigma}$.

* Only one case where the "goals" are zero is considered for simplicity.

The closed-loop cost $J^{CL}(2)$ to be minimized by the first period control $u(0)$ is, following the discussion of Section 2, as follows

$$J^{CL}(2) = J_D(2) + J_C(2) + J_P(2) \tag{3.7}$$

where

$$J_D(2) = \tfrac{1}{2} r u^2(0) + C_0(1) \tag{3.8}$$

is the deterministic part of the total cost, and the stochastic part of the cost is divided into the "caution" and "probing" components given respectively, by

$$J_C(2) = \tfrac{1}{2} \operatorname{tr} \{\mathbf{K}_0(1)\boldsymbol{\Sigma}(1|0) + \mathbf{K}_0(2)\mathbf{Q}(1)\} \tag{3.9}$$

$$J_P(2) = \tfrac{1}{2} \operatorname{tr} \{\mathscr{A}_{0,xx}(1)\boldsymbol{\Sigma}_0(1|1)\} \tag{3.10}$$

Simulations were done for several cases. The values of the parameters that define each case are given in Table 1. The weighting of the final state was taken as $q(2) = 1$ in each case.

TABLE 1.
CASES SIMULATED

Case \ Parameters	a	$\hat{b}(0)$	$\Sigma^{bb}(0)$	c	Q	$q(1)$	r
I	0.7	−0.5	0.5	3.5	0.2	1	1
II	0.7	−0.5	0.5	3.5	0.2	1	0.2
III	0.7	−0.5	0.5	3.5	1	1	0.2
IV	0.7	0	0.5	3.5	0.2	1	0.2
V	0.7	−0.5	0.5	1	0.5	1	0.2
VI	0.25	−0.5	0.5	1	0.5	1	0.2
VII	0.25	−0.5	0.5	1	0.2	0	0.2

Figure 1 presents, for Case I, the plot of the closed-loop cost from the initial time and its three components as defined above. The deterministic component (obtained by using the CE control as nominal) attains its minimum at the value $u^{CE}(0) = 2.53$. The caution part J_C is a monotonically increasing function of u(0). This is due to the fact that a large input applied through the imperfectly known gain b will lead to a large uncertainty. The probing term J_P is monotonically decreasing in the first period control—the larger this control, the more accurate estimates will be available subsequently.

The minimum of the closed-loop cost J^{CL} is attained at $u^{CL}(0) = 1.33$. This value is below u^{CE} because the benefit from probing (decrease of J_P) is much less than its cost, due mainly to the steep increase of J_C. Thus the problem under consideration the caution dominates the probing.

An interesting and important question is what is the value of the *optimal* first period control and how does the (suboptimal) u^{CL} relate to it. Also the comparison with the open-loop optimal feedback control u^{OLOF}, the adaptive controls of MacRae [M2] and Chow [C2], denoted, respectively, as u^{AM} and u^{AC} is of interest.* The three learning algorithms AM, AC and CL used CE controls as nominal.

* For the problem simulated here the algorithm of Rausser and Freebairn [R2] coincides with the one of MacRae [M2].

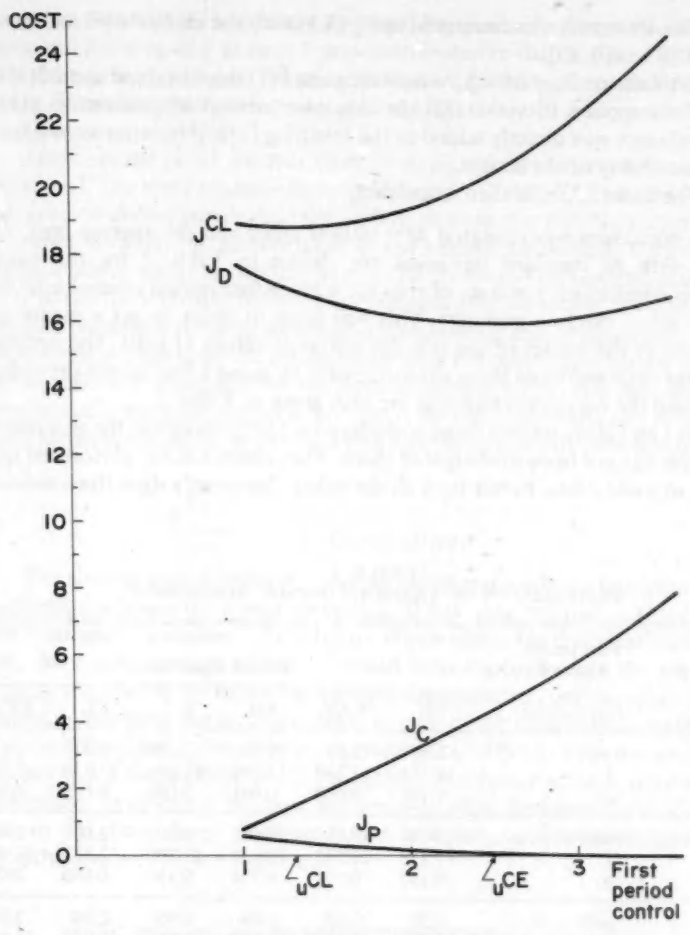

Figure 1 The Closed-Loop Approximate Cost and Its Components

Since, for the problem under consideration the optimal control cannot be obtained exactly, a Monte Carlo simulation was carried out. Note that the optimal control prior to the last stage, $u(1)$, is of the open-loop type and is given by

$$u^*(1) = -[r + (\hat{b}(1))^2 + \Sigma^{bb}(1|1)]^{-1}\hat{b}(1)[ax(1) + c] \tag{3.11}$$

where $\hat{b}(1)$ is the estimate of b after $x(1)$ has been observed.

The expected value of the cost (3.2) for a given value of $u(0)$ was evaluated by going through the following steps in each run

1. The true b was obtained from a random number generator as $\mathcal{N}[\hat{b}(0), \Sigma^{bb}(0)]$, as implied by our Bayesian assumption.

2. The state $x(1)$ was computed using (3.1) with the chosen $u(0)$ and a noise $v(0) \sim \mathcal{N}[0, Q]$.
3. With the realization $x(1)$ a new estimate $\hat{b}(1)$ was obtained and $u^*(1)$ was then applied to yield $x(2)$. In this case instead of generating $v(1)$ its variance was directly added to the resulting $[x(2)]^2$ in order to reduce the variability of the result.
4. The cost (3.2) was then calculated.

This procedure was repeated $M = 10{,}000$ times and the average cost, $\bar{J}(2)$, together with its standard deviation are shown in Table 2 for the control algorithms considered. Each set of runs for a given first period control $u(0)$ used the same set of random numbers. This was done in order to get a meaningful comparison of the values of the cost for different values of $u(0)$. The optimum control was obtained from these extensive runs by using a line search procedure; its value and the corresponding cost are also given in Table 2.

Cases I and II are among those considered in [M2]; however, the goodness of the controls has not been investigated there. The reason OLOF performed quite well and, in some cases, better than all the other "learning" algorithms seems to

TABLE 2
PERFORMANCE OF VARIOUS CONTROL ALGORITHMS

Case	First period Control and Performance	Control algorithm					
		CE	OLOF	AM	AC	CL	OPT
I	$u(0)$	2.53	1.71	1.74	1.998	1.33	1.63
	$\bar{J}$	18.158	17.239	17.246	17.396	17.350	17.230
	σ_J	0.105	0.091	0.091	0.096	0.085	0.089
II	$u(0)$	5.30	2.69	2.68	3.30	2.00	2.46
	$\bar{J}$	17.178	12.533	12.530	13.100	12.652	12.497
	σ_J	0.135	0.101	0.101	0.110	0.096	0.099
III	$u(0)$	5.30	2.69	2.49	2.96	2.04	2.85
	$\bar{J}$	18.474	14.554	14.633	14.550	15.070	14.538
	σ_J	0.148	0.122	0.122	0.123	0.122	0.122
IV	$u(0)$	0	0	0	0	2.40	1.67
	$\bar{J}$	24.10	24.10	24.10	24.10	18.70	17.60
	σ_J	0.078	0.078	0.078	0.078	0.22	0.20
V	$u(0)$	1.51	0.77	0.78	0.62	0.58	0.89
	$\bar{J}$	2.114	1.889	1.888	1.921	1.933	1.881
	σ_J	0.018	0.017	0.017	0.017	0.017	0.017
VI	$u(0)$	1.25	0.62	0.69	0.33	0.55	0.70
	$\bar{J}$	1.563	1.414	1.411	1.467	1.422	1.411
	σ_J	0.012	0.011	0.011	0.011	0.011	0.011
VII	$u(0)$	0.34	0.49	−0.10	0.32	0.08	0.85
	$\bar{J}$	0.656	0.633	0.680	0.659	0.681	0.605
	σ_J	0.004	0.004	0.005	0.004	0.005	0.004

stem from the short horizon of the problem. Furthermore, except for case IV, the accidental learning of b at time 1 also contributed to this. In case IV, if $u(0) = 0$ then there is no accidental learning and active probing is needed. In all cases except IV caution dominates probing. In the last case, u^{CE} is second best while in the other cases it is worst.

These results point out that there is no superiority between the algorithms considered. The more sophisticated algorithms (AM, AC, CL), which are suboptimal, are not always better than the OLOF or even the CE. Note also the small difference between the performance of the various algorithms. Further work is required where these algorithms have to be compared on realistic problems. In order to see the interplay between caution and probing longer horizon problems should be considered. Since meaningful comparisons can be made only by Monte Carlo methods, the required computation time for results that can be stated with high significance appears to be the main stumbling block. Another aspect to be investigated is the trade-off between computational complexity and performance. Some preliminary results in this direction have been recently obtained by Norman [N1].

5. Conclusion

The structural properties of the closed-loop and feedback type controllers for stochastic problems have been discussed. It has been pointed out that a closed-loop controller "anticipates" the future observations via their statistical description. As a consequence of this, a closed-loop controller has the capability of probing the system to reduce the existing uncertainties. On the other hand, this controller will have to exercise caution in view of the uncertainties in the system. When minimizing the closed-loop approximation of the cost, such a controller will also achieve a trade-off between the present and future actions according to the information available at the time the corresponding decisions are made. Simulation results have been presented that compare several suboptimal control algorithms as well as the optimum.

Acknowledgement

Thanks are due to Professor Ya. Z. Tsypkin for stimulating comments in an exchange of letters that led to the derivation of the algorithm presented in Section 2. The very careful reading of the paper by Professors D. Kendrick and A. L. Norman and their ensuing comments are gratefully acknowledged.

Systems Control, Inc.
Standford University

Appendix

The Closed Loop Optimization of the Cost-to-go

Rewriting (2.16) in the stochastic dynamic programming form with the assumed closed-loop-optimal expected cost-to-go of the form (2.19) yields

$$\text{(A.1)}\quad \Delta J_0^*(N-j)=\min_{\delta\mathbf{u}(j)} E\{L'_{0,\mathbf{x}}(j)\,\delta\mathbf{x}(j)+\tfrac{1}{2}\,\delta\mathbf{x}'(j)L_{0,\mathbf{xx}}\,\delta\mathbf{x}(j)$$
$$+\phi'_{0,\mathbf{u}}(j)\,\delta\mathbf{u}(j)+\tfrac{1}{2}\,\delta\mathbf{u}'(j)\,\phi_{0,\mathbf{uu}}(j)\,\delta\mathbf{u}(j)$$
$$+g_0(j+1)+E[\mathbf{p}'_0(j+1)\,\delta\mathbf{x}(j+1)$$
$$+\tfrac{1}{2}\,\delta\mathbf{x}'(j+1)\mathbf{K}_0(j+1)\,\delta\mathbf{x}(j+1)|\mathscr{P}^{j+1}]|\mathscr{P}^j\}$$
$$j=k+1,\ldots,N-1.$$

Using (2.17) and retaining terms up to second order only, the above becomes

$$\text{(A.2)}\quad \Delta J_0^*(N-j)=\min_{\delta\mathbf{u}(j)}\{E\{L'_{0,\mathbf{x}}(j)\,\delta\mathbf{x}(j)+\tfrac{1}{2}\,\delta\mathbf{x}'(j)\,L_{0,\mathbf{xx}}(j)\,\delta\mathbf{x}(j)$$
$$+\phi'_{0,\mathbf{u}}(j)\,\delta\mathbf{u}(j)+\tfrac{1}{2}\,\delta\mathbf{u}'(j)\phi_{0,\mathbf{uu}}(j)\,\delta\mathbf{u}(j)+g_0(j+1)$$
$$+\mathbf{p}'_0(j+1)\mathbf{f}_{0,\mathbf{x}}(j)\,\delta\mathbf{x}(j)+\mathbf{p}'_0(j+1)\mathbf{f}_{0,\mathbf{u}}(j)\,\delta\mathbf{u}(j)$$
$$+\mathbf{p}'_0(j+1)\sum_{i=1}^{M}\mathbf{e}_i[\tfrac{1}{2}\,\delta\mathbf{x}'(j)f^i_{0,\mathbf{xx}}(j)\,\delta\mathbf{x}(j)+\delta\mathbf{u}'(j)f^i_{0,\mathbf{ux}}(j)\,\delta\mathbf{x}(j)$$
$$+\tfrac{1}{2}\,\delta\mathbf{u}'(j)f^i_{0,\mathbf{uu}}(j)\,\delta\mathbf{u}(j)]+\tfrac{1}{2}\,\delta\mathbf{x}'(j)\mathbf{f}'_{0,\mathbf{x}}(j)\mathbf{K}_0(j+1)\mathbf{f}_{0,\mathbf{x}}(j)\,\delta\mathbf{x}(j)$$
$$+\delta\mathbf{u}'(j)\mathbf{f}'_{0,\mathbf{u}}(j)\mathbf{K}_0(j+1)\mathbf{f}_{0,\mathbf{x}}(j)\,\delta\mathbf{x}(j)$$
$$+\tfrac{1}{2}\,\delta\mathbf{u}'(j)\mathbf{f}'_{0,\mathbf{u}}(j)\mathbf{K}_0(j+1)\mathbf{f}_{0,\mathbf{u}}(j)\,\delta\mathbf{u}(j)|\mathscr{P}^j\}$$
$$+\tfrac{1}{2}\,\mathrm{tr}\,[\mathbf{K}_0(j+1)\mathbf{Q}(j)]\}$$

where $\mathbf{Q}(j)$ is the covariance of $\mathbf{v}(j)$.

Denoting

$$\text{(A.3)}\qquad H_0(j)\triangleq L_0(j)+\phi_0(j)+\mathbf{p}'_0(j+1)\mathbf{f}_0(j)$$

and rearranging the terms in (A.2) it becomes

$$\text{(A.4)}\quad \Delta J_0^*(N-j)=\min_{\delta\mathbf{u}(j)} E\{H'_{0,\mathbf{x}}(j)\,\delta\mathbf{x}(j)+H'_{0,\mathbf{u}}(j)\,\delta\mathbf{u}(j)$$
$$+\tfrac{1}{2}\,\delta\mathbf{x}'(j)[H_{0,\mathbf{xx}}(j)+\mathbf{f}'_{0,\mathbf{x}}(j)\mathbf{K}_0(j+1)\mathbf{f}_{0,\mathbf{x}}(j)]\,\delta\mathbf{x}(j)$$
$$+\delta\mathbf{u}'(j)[H_{0,\mathbf{ux}}(j)+\mathbf{f}'_{0,\mathbf{u}}(j)\mathbf{K}_0(j+1)\mathbf{f}_{0,\mathbf{x}}(j)]\,\delta\mathbf{x}(j)$$
$$+\tfrac{1}{2}\,\delta\mathbf{u}'(j)[H_{0,\mathbf{uu}}(j)+\mathbf{f}'_{0,\mathbf{u}}(j)\mathbf{K}_0(j+1)\mathbf{f}_{0,\mathbf{u}}(j)]\,\delta\mathbf{u}(j)$$
$$+g_0(j+1)+\tfrac{1}{2}\,\mathrm{tr}\,[\mathbf{K}_0(j+1)\mathbf{Q}(j)]|\mathscr{P}^j\}.$$

Denote

$$\text{(A.5)}\qquad \mathscr{H}_{0,\mathbf{xx}}(j)\triangleq H_{0,\mathbf{xx}}(j)+\mathbf{f}'_{0,\mathbf{x}}(j)\mathbf{K}_0(j+1)\mathbf{f}_{0,\mathbf{x}}(j)$$

$$\text{(A.6)}\qquad \mathscr{H}_{0,\mathbf{ux}}(j)\triangleq H_{0,\mathbf{ux}}(j)+\mathbf{f}'_{0,\mathbf{u}}(j)\mathbf{K}_0(j+1)\mathbf{f}_{0,\mathbf{x}}(j)$$

$$\text{(A.7)}\qquad \mathscr{H}_{0,\mathbf{uu}}(j)\triangleq H_{0,\mathbf{uu}}(j)+\mathbf{f}'_{0,\mathbf{u}}(j)\mathbf{K}_0(j+1)\mathbf{f}_{0,\mathbf{u}}(j)$$

$$\text{(A.8)}\qquad \mathscr{A}_{0,\mathbf{xx}}(j)\triangleq \mathscr{H}'_{0,\mathbf{ux}}(j)\mathscr{H}^{-1}_{0,\mathbf{uu}}(j)\mathscr{H}_{0,\mathbf{ux}}(j).$$

With these notations, the optimal perturbation control resulting from (A.4) is

$$\text{(A.9)}\qquad \delta\mathbf{u}^*(j)=-\mathscr{H}^{-1}_{0,\mathbf{uu}}(j)[\mathscr{H}_{0,\mathbf{ux}}(j)\,\delta\hat{\mathbf{x}}(j|j)+H_{0,\mathbf{u}}(j)]$$

where

$$\delta\hat{\mathbf{x}}(j|j) = E[\delta\mathbf{x}(j)|\mathcal{P}^j]. \tag{A.10}$$

A necessary and sufficient condition for the existence and uniqueness of the solution to the perturbation problem is that (A.7) be positive definite. Note that this is guaranteed if the nominal is a local minimum for the deterministic problem. In this case (A.8) will be positive semidefinite. Reinserting (A.9) into (A.4) yields

$$\begin{aligned}\Delta J_0^*(N-j) = E\{&H'_{0,\mathbf{x}}(j)\,\delta\mathbf{x}(j) - H'_{0,\mathbf{u}}(j)\mathcal{H}^{-1}_{0,\mathbf{uu}}(j)H_{0,\mathbf{u}}(j)\\ &- H'_{0,\mathbf{u}}(j)\mathcal{H}^{-1}_{0,\mathbf{uu}}(j)\mathcal{H}_{0,\mathbf{ux}}(j)\,\delta\hat{\mathbf{x}}(j|j)\\ &+\tfrac{1}{2}\,\delta\mathbf{x}'(j)\mathcal{H}_{0,\mathbf{xx}}(j)\,\delta\mathbf{x}(j) - \delta\hat{\mathbf{x}}(j|j)\mathcal{A}_{0,\mathbf{xx}}(j)\,\delta\mathbf{x}(j)\\ &- H'_{0,\mathbf{u}}(j)\mathcal{H}^{-1}_{0,\mathbf{uu}}(j)\mathcal{H}_{0,\mathbf{ux}}(j)\,\delta\mathbf{x}(j)\\ &+\tfrac{1}{2}\,\delta\hat{\mathbf{x}}'(j|j)\mathcal{A}_{0,\mathbf{xx}}\,\delta\hat{\mathbf{x}}(j|j)\\ &+ H'_{0,\mathbf{u}}(j)\mathcal{H}^{-1}_{0,\mathbf{uu}}(j)\mathcal{H}_{0,\mathbf{ux}}(j)\,\delta\hat{\mathbf{x}}(j|j)\\ &+\tfrac{1}{2}H'_{0,\mathbf{u}}(j)\mathcal{H}^{-1}_{0,\mathbf{uu}}(j)H_{0,\mathbf{u}}(j)|\mathcal{P}^j\} + g_0(j+1)\\ &+\tfrac{1}{2}\operatorname{tr}[\mathbf{K}_0(j+1)\mathbf{Q}(j)].\end{aligned} \tag{A.11}$$

Notice that

$$\begin{aligned}E[\delta\mathbf{x}'(j)\mathcal{A}_{0,\mathbf{xx}}(j)\,\delta\mathbf{x}(j)|\mathcal{P}^j]\\ = \delta\hat{\mathbf{x}}'(j|j)\mathcal{A}_{0,\mathbf{xx}}(j)\,\delta\hat{\mathbf{x}}(j|j) + \operatorname{tr}[\mathcal{A}_{0,\mathbf{xx}}(j)\boldsymbol{\Sigma}_0(j|j)]\end{aligned} \tag{A.12}$$

where $\boldsymbol{\Sigma}_0(j|j)$, is the covariance of the (future) updated state, along the nominal. With this, (A.11) can be rewritten as follows:

$$\begin{aligned}\Delta J_0^*(N-j) = g_0(j+1) &-\tfrac{1}{2}H'_{0,\mathbf{u}}(j)\mathcal{H}^{-1}_{0,\mathbf{uu}}(j)H_{0,\mathbf{u}}(j)\\ &+\tfrac{1}{2}\operatorname{tr}[\mathbf{K}_0(j+1)\mathbf{Q}(j) + \mathcal{A}_{0,\mathbf{xx}}(j)\boldsymbol{\Sigma}_0(j|j)]\\ &+ E\{[H_{0,\mathbf{x}}(j) - \mathcal{H}'_{0,\mathbf{ux}}(j)\mathcal{H}^{-1}_{0,\mathbf{uu}}(j)H_{0,\mathbf{u}}(j)]'\,\delta\mathbf{x}(j)\\ &+\tfrac{1}{2}\,\delta\mathbf{x}'(j)[\mathcal{H}_{0,\mathbf{xx}}(j) - \mathcal{A}_{0,\mathbf{xx}}(j)]\,\delta\mathbf{x}(j)|\mathcal{P}^j\}.\end{aligned} \tag{A.13}$$

Thus, it can be seen that (A.13) is indeed the assumed quadratic form of (2.19) and the recursions for g_0, $\mathbf{p}_0$ and $\mathbf{K}_0$ are, using notations (A.5)–(A.8)

$$\begin{aligned}g_0(j) = g_0(j+1) &-\tfrac{1}{2}H'_{0,\mathbf{u}}(j)\mathcal{H}^{-1}_{0,\mathbf{uu}}(j)H_{0,\mathbf{u}}(j) + \tfrac{1}{2}\operatorname{tr}[\mathbf{K}_0(j+1)\mathbf{Q}(j)\\ &+\mathcal{A}_{0,\mathbf{xx}}(j)\boldsymbol{\Sigma}_0(j|j)]\quad j = N-1,\ldots,k+1;\ g_0(N) = 0\end{aligned} \tag{A.14}$$

$$\begin{aligned}\mathbf{p}_0(j) = H_{0,\mathbf{x}}(j) - \mathcal{H}'_{0,\mathbf{ux}}(j)\mathcal{H}^{-1}_{0,\mathbf{uu}}(j)H_{0,\mathbf{u}}(j)\\ j = N-1,\ldots,k+1;\ \mathbf{p}_0(N) = \psi_{0,\mathbf{x}}\end{aligned} \tag{A.15}$$

$$\begin{aligned}\mathbf{K}_0(j) = \mathcal{H}_{0,\mathbf{xx}}(j) - \mathcal{A}_{0,\mathbf{xx}}(j)\\ j = N-1,\ldots,k+1;\ \mathbf{K}_0(N) = \psi_{0,\mathbf{xx}}.\end{aligned} \tag{A.16}$$

In order to separate the stochastic effects in the expected cost, introduce

$$\text{(A.17)} \qquad \gamma_0(j) = \gamma_0(j+1) - \tfrac{1}{2} H'_{0,\mathbf{u}}(j) \mathcal{H}^{-1}_{0,\mathbf{uu}}(j) H_{0,\mathbf{u}}(j)$$
$$j = N-1, \ldots, k+1; \ \gamma_0(N) = 0.$$

Then

$$\text{(A.18)} \quad g_0(k+1) = \gamma_0(k+1) + \tfrac{1}{2} \sum_{j=k+1}^{N-1} \operatorname{tr} [\mathbf{K}_0(j+1)\mathbf{Q}(j) + \mathcal{A}_{0,\mathbf{xx}}(j)\boldsymbol{\Sigma}_0(j|j)]$$

This completes the proof of (2.20). If the sequence of nominal controls is optimal for the deterministic problem, then the Hamiltonian (A.3) achieves its minimum and $H_{0,\mathbf{u}} = \mathbf{0}$ (unless one has a constrained optimization and the minimum occurs at the boundary). In this case $\gamma_0(j) = 0$ for all j.

References

[A1] M. Aoki, *Optimization of Stochastic Systems*, Academic Press, 1967.

[A2] K. J. Åström, *Introduction to Stochastic Control Theory*, Academic Press, 1970.

[A3] A. B. Abel, "A Comparison of Three Control Algorithms as Applied to the Monetarist Fiscalist Debate," *Annals of Economic and Social Measurement*, Vol. 4, No. 2, pp. 239–252, 1975.

[B1] R. Bellman, *Adaptive Control Processes: A Guided Tour*, Princeton University Press, 1961.

[B2] Y. Bar-Shalom and R. Sivan, "On the Optimal Control of Discrete-Time Linear Systems with Random Parameters," *IEEE Trans. Automatic Control*, Vol. AC-14, pp. 3–8, February 1969.

[B3] Y. Bar-Shalom and E. Tse, "Dual Effect, Certainty Equivalence and Separation in Stochastic Control," *IEEE Trans. Automatic Control*, Vol. AC-19, pp. 494–500, Oct. 1974.

[B4] Y. Bar-Shalom and E. Tse, "Concepts and Methods in Stochastic Control" in *Control and Dynamic Systems: Advances in Theory and Applications*, C. T. Leondes, ed., Academic Press, 1975.

[C1] R. E. Curry, "A New Algorithm for Suboptimal Stochastic Control," *IEEE Trans. Auto. Control*, Vol. AC-14, pp. 533–536, Oct. 1969.

[C2] G. Chow, *Analysis and Control of Dynamic Economic Systems*, Wiley, 1975.

[C3] G. Chow, "On the Control of Nonlinear Systems with Unknown Parameters," Econometric Research Program Report No. 175, Princeton University, March 1975.

[D1] S. E. Dreyfus, *Dynamic Programming and the Calculus of Variations*, Academic Press, 1965.

[D2] J. G. Deshpande, T. N. Upadhyay and D. G. Lainiotis, "Adaptive Control of Linear Stochastic Systems," *Automatica*, Vol. 9, pp. 107–115, 1972.

[F1] A. A. Feldbaum, *Optimal Control Systems*, Academic Press, 1965.

[J1] O. L. R. Jacobs and J. W. Patchell, "Caution and Probing in Stochastic Control," *International Journal of Control*, Vol. 15, pp. 189–199, 1972.

[M1] L. Meier, R. E. Larson, and A. J. Tether, "Dynamic Programming for Stochastic Control of Discrete Systems," *IEEE Trans. Auto. Control*, Vol. AC-16, pp. 767–775, Dec. 1971 (Special Issue on Linear Quadratic Gaussian Problem).

[M2] E. C. MacRae, "Linear Decision with Experimentation," *Annals of Economic and Social Measurement*, Vol. 1, pp. 437–447, 1972.

[N1] A. L. Norman, "First Order Dual Control," *Annals of Economic and Social Measurement*, 1976.

[R1] H. Raiffa and R. Schlaifer, *Applied Statistical Decision Theory*, M.I.T. Press, 1972.

[R2] G. C. Rausser and J. W. Freebairn, "Approximate Adaptive Control Solutions to U.S. Beef Trade Policy," *Annals of Economic and Social Measurement*, Vol. 3, No. 1, pp. 177–203, 1974.

[S1] G. N. Saridis, *Self-Organizing Control Stochastic Systems*, M. Dekker, 1976.

[S2] J. Speyer, J. Deyst, and D. Jacobson, "Optimization of Stochastic Linear Systems with Additive Measurement and Process Noise Using Exponential Performance Criteria," *IEEE Trans. Auto Control*, Vol. AC-19, pp. 358–366, August 1974.

[S3] H. Simon, "Dynamic Programming under Uncertainty with a Quadratic Function," *Econometrica*, Vol. 24, pp. 74–81, 1956.

[T1] E. Tse, Y. Bar-Shalom, and L. Meier, "Wide-Sense Adaptive Dual Control of Stochastic Nonlinear Systems," *IEEE Trans. Automatic Control*, Vol. AC-18, pp. 98–108, April 1973.

[T2] E. Tse, and Y. Bar-Shalom, "An Actively Adaptive Control for Discrete-Time Systems with Random Parameters," *IEEE Trans. Auto. Control*, Vol. AC-18, pp. 109-117, April 1973.

[T3] E. Tse and M. Athans, "Adaptive Stochastic Control for a Class of Linear Systems," *IEEE Trans. Automatic Control*, Vol. AC-17, pp. 38–51, February 1972.

[T4] E. Tse and Y. Bar-Shalom, "Adaptive Dual Control for Stochastic Nonlinear Systems with Free End-Time," *IEEE Transactions Auto. Control*, Vol. AC-20, pp. 600–605, October 1975.
[T5] H. Theil, "A Note on Certainty Equivalence in Dynamic Planning," *Econometrica*, Vol. 25, pp. 346–349, 1957.