have been able to explicitly and completely solve and furnish a certainty equivalent separation theorem for the optimal solution, for various values of the parameters.

It is hoped that the methods used here will find application in other dynamic programming problems with no known explicit solutions.

REFERENCES

[1] W. R. Thompson, "On the likelihood that one unknown probability exceeds another in view of the evidence of two samples," *Biometrika*, vol. 25, pp. 285–294, 1933.

[2] R. Bellman, "A problem in the sequential design of experiments," *Sankhya*, vol. 16, pp. 221–229, 1956.

[3] R. N. Bradt, S. M. Johnson, and S. Karlin, "On sequential designs for maximizing the sum of *n* observations," *Ann. Math. Statist.*, vol. 27, pp. 1060–1074, 1956.

[4] W. Vogel, "A sequential design for the two-armed bandit," *Ann. Math. Statist.*, vol 31, pp. 430–443, 1960.

[5] D. A. Berry, "A Bernoulli two-armed bandit," *Ann. Math. Statist.*, vol. 43, pp. 871–897, 1972.

[6] L. Brown, "Optimal policies for a sequential decision process," *J. SIAM*, vol. 13, pp. 37–46, 1965.

[7] M. Rothschild, "A two-armed bandit theory of market pricing," *J. Econ. Theory*, vol. 9, pp. 185–202, 1974.

[8] D. Blackwell, "Discounted dynamic programming," *Ann. Math. Statist.*, vol. 36, pp. 226–235, 1965.

[9] D. P. Bertsekas and S. E. Shreve, *Stochastic Optimal Control: The Discrete Time Case*, New York: Academic, 1978.

[10] P. R. Kumar and T. H. Shiau, "Zero sum dynamic games," in *Advances in Dynamic Systems Control*, C. T. Leondes, Ed. New York: Academic, 1981.

[11] M. H. DeGroot, "Optimal statistical decision," New York: McGraw-Hill, 1970.

[12] R. Bellman, *Dynamic Programming*. Princeton, NJ: Princeton Univ. Press, 1957.

P. R. Kumar was born in Nagpur, India on April 21, 1952. He received the B.Tech. degree in electrical engineering from the Indian Institute of Technology, Madras, India, in 1973, and the M.S. and D.Sc. degrees in systems science and mathematics from Washington University, St. Louis, MO, in 1975 and 1977, respectively.

Since September 1977, he has been an Assistant Professor at the University of Maryland, Baltimore County. His current research interests are in adaptive control, dynamic programming theory, estimation and control of stochastic systems and zero-sum dynamic games.

Dr. Kumar is a member of Sigma Xi and the Society for Industrial and Applied Mathematics.

Thomas I. Seidman was born in New York in 1935. He received the Ph.D. degree in mathematics from New York University in 1959.

He has both worked in industry and taught at various universities. Currently, he is on leave from the University of Maryland, Baltimore County, as a Visiting Professor at the Université de Nice, Nice, France. His research interests are focused in three main areas: distributed parameter control theory, nonlinear partial differential equations, and computational methods for ill-posed problems.

Stochastic Dynamic Programming: Caution and Probing

YAAKOV BAR-SHALOM, SENIOR MEMBER, IEEE

***Abstract*—The purpose of this paper is to unify the concepts of caution and probing put forth by Feldbaum [14] with the mathematical technique of stochastic dynamic programming originated by Bellman [5]. The decomposition of the expected cost in a stochastic control problem, recently developed in [8], is used to assess quantitatively the caution and probing effects of the system uncertainties on the control. It is shown how in some problems, because of the uncertainties, the control becomes cautious (less aggressive) while in other problems it will probe (by becoming more aggressive) in order to enhance the estimation/identification while controlling the system. Following this a classification of stochastic control problems according to the dominant effect is discussed. This is then used to point out which are the stochastic control problems where substantial improvements can be expected from using a sophisticated algorithm versus a simple one.**

Manuscript received April 6, 1980; revised January 14, 1981 and March 20, 1981. This work was supported in part by the National Science Foundation under Grant GS-32271 and in part by the Air Force Office of Scientific Research under Grant 80-0098.

The author is with the Department of Electrical Engineering and Computer Science, University of Connecticut, Storrs, CT 06268.

I. INTRODUCTION

THIS PAPER reviews recent work in the area of stochastic control and shows how the concepts of caution and probing, originated by Feldbaum [14], can be unified with Bellman's dynamic programming technique [5], [6]. The concepts of caution and probing, developed by Feldbaum [14] about 20 years ago and also discussed in [16], dealt from an intuitive point of view with some phenomena peculiar to stochastic control problems or decision under uncertainty.

0018-9286/81/1000-1184$00.75 ©1981 IEEE

In the presence of uncertainty, modeled by random variables or stochastic processes, there is usually a deterioration of the system performance, which can be measured by an increase in the (expected) loss function compared to the deterministic case. In order to reduce the increase in the loss function the controller will tend to be "cautious," a property known in the decision theory literature as "risk aversion" [12]. This phenomenon occurs for convex loss functions that the decision maker (controller) wants to minimize, like in most control problems. On the other hand, in multistage problems where observations are made on the system at each stage, the controller might be able to carry out what has been called "active information gathering" or "probing" of the system for estimation enhancement. This is possible when the controller affects not only the state of the system but also the quality of the estimation process, i.e., has the so-called "dual effect."

This paper intends to provide a tutorial on these aspects of stochastic control by a suitable presentation of the basic concepts embodied in the stochastic dynamic programming. When the caution and probing phenomena are present in the multistage problems, the optimal solution is not known. In view of this, the insight is provided by considering a suboptimal algorithm that has the features of the optimal one.

Section II discusses the information state in the multistage control problem of a stochastic system. The formulation of the principle of optimality for stochastic systems and the resulting stochastic dynamic programming equation for additive cost functions are discussed in Section III. It is pointed out how the "preposterior analysis" technique is a direct consequence of the principle of optimality. The definition of the dual effect and the types of approximate solutions of the stochastic dynamic programming are the topic of Section IV. The "closed-loop" approximation of the stochastic dynamic programming using the "wide-sense" information state [8], [29], [30] is shown in Section V to lead to a decomposition of the expected cost into three terms. Two of these terms can be associated directly with the caution and probing effects discussed earlier giving thus a quantitative measure of these effects. It is shown in Section VI how one can classify stochastic control problems according to the dominant term in the cost decomposition. This is then illustrated via a number of examples where stochastic control problems that are probing-dominated, caution-dominated, and essentially deterministic are presented. The effect of various state weightings in the cost function and the anticipated future learning are also discussed. Conclusions are presented in Section VII.

II. The Information State in a Stochastic Control Problem

The principle of optimality of Bellman [5] can be stated as follows for stochastic problems: at any time, whatever the present information and past decisions, the remaining decisions must constitute an optimal policy with regard to the current information set.

In the deterministic case the information set is the state of the system. This, together with the controller's subsequent decisions fully determines the future evolution of the system. In the stochastic case the information set is, loosely, what the controller knows about the system. This will be discussed in more detail next.

Consider the following general stochastic control problem. The state x evolves according to the equation

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

where u is the control and v is the process noise. The measurements are described by

$$y(k)=h[k,x(k),w(k)] \qquad k=1,\cdots \tag{2.2}$$

where w is the measurement noise. The information set at time k is assumed to be the past measurements and controls

$$I^k \triangleq \{Y^k, U^{k-1}\} \supset I^{k-1} \tag{2.3}$$

where

$$Y^k=\{y(j)\}_{j=0}^{k}, \quad U_i^k \triangleq \{u(j)\}_{j=i}^{k} \tag{2.4}$$

and subscript $i=0$ is omitted. The inclusion property in (2.3) points to the fact that the sequence of information as assumed here is nested—each contains its predecessor.

Since (2.3) is growing with k it is of interest when a (nongrowing) information state can replace (2.3).

Note that $x(k)$ is a state only in the deterministic context when, together with U_k^{j-1}, it fully determines $x(j)$, $\forall j>k$, i.e., $x(k)$ summarizes the past of the system. The stochastic counterpart of this is the "information state."

The information state is defined as a vector-valued variable or a function that summarizes the past (i.e., it can replace I^k) when we want to characterize (probabilistically) the future evolution of the system. This is more general than the "informative statistic" of Striebel [26] which is, roughly, what the optimal controller (for the problem under consideration) needs from the past data (2.3).

It is assumed in the sequel that all the pertinent probability densities exist. Discrete-valued random variables will have a probability density function (pdf) with Dirac delta functions at the locations of the point masses.

If both sequences of process and measurement noises are white and mutually independent, then at time k the conditional probability density function of the vector $x(k)$

$$\mathcal{S}^k=p[x(k)|I^k]^{1} \tag{2.5}$$

is an information state. This can be seen from the following. The conditional density of $x(k+1)$ can be written using Bayes' rule

$$\mathcal{S}^{k+1}=p[x(k+1)|I^{k+1}]=\frac{1}{c}p[y(k+1)|x(k+1),I^k,u(k)] \cdot p[x(k+1)|I^k,u(k)] \tag{2.6}$$

where c is a normalization constant.

[1] Rigorously, the conditional density should be written $p[\cdot|Y^k;U^{k-1}]$ because this is conditioned on the sigma-algebra generated by the measurements but it is not well-defined unless the values of past controls or control functions are indicated [26]. For $k=0$ this is the prior density of the state.

If the measurement noise is white ($w(k+1)$ conditioned on $x(k+1)$ has to be independent of $w(j), j\leq k$, i.e., state dependent measurement noise is allowed), then

$$p[y(k+1)|x(k+1),I^k,u(k)]=p[y(k+1)|x(k+1)] \tag{2.7}$$

(the control is anyway irrelevant in the conditioning).

For an arbitrary value of the control at k one has

$$p[x(k+1)|I^k,u(k)]=\int p[x(k+1)|x(k),I^k,u(k)] \cdot p[x(k)|I^k,u(k)]\,dx(k). \tag{2.8}$$

If the process noise sequence is white and independent of the measurement noises ($v(k)$ conditioned on $x(k)$ has to be independent of $v(j-1), w(j), j\leq k$, i.e., state dependent process noise is allowed), then

$$p[x(k+1)|x(k),I^k,u(k)]=p[x(k+1)|x(k),u(k)] \tag{2.9}$$

and, since

$$p[x(k)|I^k,u(k)]=p[x(k)|I^k]=\mathbb{S}^k \tag{2.10}$$

then, inserting (2.9) and (2.10) into (2.8) it follows that

$$p[x(k+1)|I^k,u(k)]=\phi[k+1,\mathbb{S}^k;u(k)]. \tag{2.11}$$

Now, using (2.7) and (2.11) in (2.6) one has

$$\mathbb{S}^{k+1}=\psi[k+1,\mathbb{S}^k,y(k+1),u(k)], \tag{2.12}$$

i.e., I^k is summarized by $\mathbb{S}^k$. Equation (2.12) is the recursion for the information state.

From the smoothing property of expectations it also follows that, for $j>k$,

$$\begin{aligned} p[x(j)|I^k,U_k^{j-1}] &= E[\mathbb{S}^j|I^k,U_k^{j-1}] \\ &= \int p[x(j)|x(k),I^k,U_k^{j-1}]\,p[x(k)|I^k]\,dx(k) \\ &= \int p[x(j)|x(k),U_k^{j-1}]\,\mathbb{S}^k\,dx(k) \\ &= \mu[j,\mathbb{S}^k,U_k^{j-1}] \end{aligned} \tag{2.13}$$

where the whiteness of the process noise sequence and its independence from the measurement noises has been used again.

Therefore, the whiteness and mutual independence of the two noise sequences is a sufficient condition for $\mathbb{S}^k$ to be an information state. It should be emphasized that the whiteness is the crucial assumption. This is equivalent to the requirement that $x(k)$ be an incompletely observed Markov process. If, for example, the process noise sequence is not white it is obvious that $\mathbb{S}^k$ does not summarize the past data. In this case the vector x is not a state anymore and it has to be augmented (see, e.g., [31]). This discussion points out the reason why the formulation of stochastic control problems is done with white noise sequences.

III. From the Principle of Optimality to Stochastic Dynamic Programming

Consider the problem where the number N of time steps is finite and deterministic. In general, the terminal time can be a random variable, possibly depending on the state or a decision variable. The present discussion is limited to the fixed terminal time problems. See, e.g., [11], [18] for discussions on the free end-time problem. Denote the (scalar) cost function of the problem as

$$C=C(X^N,U^{N-1}). \tag{3.1}$$

Since this is a random variable, the minimization (in general, extremization) is done in the Bayesian approach on the expected cost

$$J=E\{C\}. \tag{3.2}$$

We assume here that the minimum and, therefore, an optimal solution (policy) exist. Otherwise, the infimum of (3.2) is to be obtained and then only an ϵ-optimal policy exists (see, e.g., [11, p. 42]). Other approaches, like min-max and worst distribution, are also used sometimes but they are usually more difficult.

In order for (3.2) to be a well-defined criterion, the expectation must exist, i.e., all the variables entering into the cost must be either deterministic or random (with suitable moment conditions that guarantee the existence of the expected cost). No "unknown constants" can be used in formulating stochastic control problems with the Bayesian approach.

If there are unknown system parameters, they have to be modeled as random variables with *a priori* pdf. If these parameters are time invariant, then one has a single realization from the prior pdf, i.e., an unknown system model generated by a probabilistic mechanism before the start of the process. In this case the minimization of the expected cost implies that we want to find the optimal policy

1) over all possible initial conditions (as specified by their pdf);

2) over all possible values of the unknown parameters (whose realization is according to the corresponding pdf)—the ensemble of systems perceived by the controller in view of its uncertainty;

3) over all possible disturbance sequences.

When there are unknown time-invariant or slowly varying system parameters the stochastic controller can then be adaptive, i.e., it will (hopefully) "learn" the system parameters during the control period.

The causality condition is that any decision function must depend only on the information set available at the time it has to be computed, i.e.,

$$u(k)=u(k,I^k) \qquad k=0,1,\cdots,N-1. \tag{3.3}$$

Since the principle of optimality states that every end part of the decision process must be optimal, the multistage optimization has to be started from the last stage. The last decision, $u(N-1)$, must be optimal with regard to the information set available when it has to be computed, i.e., it will be obtained from the functional minimization

$$\min_{u(N-1)} E\left(C|I^{N-1}\right) \tag{3.4}$$

where C is the cost for the entire problem.

The next to the last decision, $u(N-2)$

1) must be optimal with respect to (w.r.t.) I^{N-2} and

2) is to be made knowing that the remaining decision $u(N-1)$ will be optimal w.r.t. $I^{N-1} \supset I^{N-2}$.

Thus, the (functional) minimization that yields the decision function at $N-2$ is

$$\min_{u(N-2)} E\left[\min_{u(N-1)} E\left(C|I^{N-1}\right)|I^{N-2}\right] \tag{3.5}$$

and it uses the result of the functional minimization (3.4).

Note that the outside averaging in (3.5) is over $y(N-1)$ using the conditional density

$$p\left[y(N-1)|I^{N-2}, u(N-2)\right] \tag{3.6}$$

parameterized by the control at $N-2$. Since this measurement is not yet available when $u(N-2)$ is to be computed but it will be available for $u(N-1)$ it is "averaged out" in (3.5).

The above-described last two steps are entirely similar to the "preposterior analysis" technique from the operations research literature discussed, e.g., in [22]. This technique is usually formulated in the following context. The first decision [here $u(N-2)$] is for information gathering by an experiment from which a posterior information will result [here $y(N-1)$] that will be used to make the last decision [here $u(N-1)$]. The prior (to the experiment) probability density of the (posterior) result of the experiment is called the "preposterior density" and in the present problem this is (3.6). Thus, one can say that preposterior analysis, which is "anticipation" (in a statistical sense, i.e., causal) of future information is a consequence of the principle of optimality.

From the above discussion it can be seen that the principle of optimality's statement that, at every stage, "the remaining decisions must constitute an optimal policy with regard to the current information set" implies the following: every decision has to use the available "hard" information (2.3) and "soft" information (3.6) about the subsequent hard information. This can be paraphrased as the optimal controller has to know how to use what it knows as well as what it knows about what it shall know.

The extension of (3.5) to the full N-stage process yields the optimal expected cost starting from the initial time as

$$J^*(0,I^0) = \min_{u(0)} E\left\{\cdots \min_{u(N-2)} E\left[\min_{u(N-1)} E\left(C|I^{N-1}\right)|I^{N-2}\right]\cdots|I^0\right\} \tag{3.7}$$

where I^0 is the initial information. Note that this equation does not assume any particular form for the cost function C.

For the additive cost given by

$$C(k)=c\left[N,x(N)\right]+\sum_{j=k}^{N-1} c\left[j,x(j),u(j)\right] \tag{3.8}$$

the minimization (3.7) of $C(0)$, the cost starting from the initial time 0 yields the discrete-time stochastic dynamic programming equation. Dynamic programming can be applied only to the so-called class of "decomposable" cost functions, as pointed out in [21], [23]. The additive cost (3.8) belongs to this class.

Since

$$C^i \triangleq \sum_{j=0}^{i} c\left[j,x(j),u(j)\right] \tag{3.9}$$

is independent of U_{i+1}^{N-1} and using the smoothing property of the expectation operator, i.e.,

$$E\left[E(\cdot|I^j)|I^k\right]=E\left[\cdot|I^k\right] \qquad \forall j>k \tag{3.10}$$

one has from (3.7)

$$\begin{aligned}
&J^*(0,I^0)\\
&=\min_{u(0)} E\left\{\cdots \min_{u(N-2)} E\left[\min_{u(N-1)} E\left[c(N)\right.\right.\right.\\
&\qquad \left.\left.\left.+\sum_{j=0}^{N-1} c(j)|I^{N-1}\right]|I^{N-2}\right]\cdots|I^0\right\}\\
&=\min_{u(0)} E\left\{\cdots \min_{u(N-2)} E\left[C^{N-2}+\min_{u(N-1)} E\left[c(N)\right.\right.\right.\\
&\qquad \left.\left.\left.+c(N-1)|I^{N-1}\right]|I^{N-2}\right]\cdots|I^0\right\}\\
&=\min_{u(0)} E\left\{c(0)+\min_{u(1)} E\left\{c(1)+\cdots \min_{u(N-2)} E\left[c(N-2)\right.\right.\right.\\
&\qquad \left.\left.\left.+\min_{u(N-1)} E\left[c(N-1)+c(N)|I^{N-1}\right]|I^{N-2}\right]\cdots|I^1\right\}|I^0\right\}.
\end{aligned} \tag{3.11}$$

In the above the cost summands have been moved to the left outside the minimizations that are not relevant for them.

Rewriting (3.11) in (backward) recursive form yields the Bellman equation

$$\begin{aligned}
J^*(k,I^k)=\min_{u(k)} E\{c\left[k,x(k),u(k)\right]\\
+J^*(k+1,I^{k+1})|I^k\} \qquad k=N-1,\cdots,0
\end{aligned} \tag{3.12}$$

where $J^*(k,I^k)$ is the optimal cost-to-go from time k to the end and its dependence on the available information set at k is explicitly pointed out. The terminal condition for (3.12) is

$$J^*(N,I^N)=E\{c\left[N,x(N)\right]|I^N\} \tag{3.13}$$

where the last measurement is irrelevant since it is averaged out immediately.

The stochastic dynamic programming functional equation (3.12) resulted from the use of the principle of optimality embodied in (3.7) for the additive cost (3.8). The recursion was obtained by moving to the left in (3.11) the cost summands.

An equivalent approach, based on the "basic lemma of stochastic control" [2] is as follows. This basic lemma states that

$$\min_u E[c(x,u)] = \min_u E\{E[c(x,u)|y]\} \geq E \min_{u(y)} E[c(x,u)|y], \quad (3.14)$$

i.e., if a measurement y related to x is available then the minimization of the conditional expectation [the right-hand side (RHS) of (3.14)] yields the absolute minimum. This is equivalent to the statement that to minimize an integral [the outside expectation in (3.14)] is best done by minimizing the integrand at each point via the function $u(y)$, i.e., "feedback," instead of a single value for the entire integral, i.e., "open loop." In other words, moving a minimization inside a sequence of expectations, to be in front of a conditional expectation (conditioned on all the available information) is what is needed for the global minimum. Thus, based upon (3.14) the expected cost is minimized as follows:

$$\begin{aligned}&\min_{u(0),\cdots,u(N-2),u(N-1)} E\{C|I^0\}\\ &= \min_{u(0),\cdots,u(N-2),u(N-1)} E\{\cdots E[E(C|I^N)|I^{N-1}]\cdots|I^0\}\\ &= \min_{u(0)} E\left\{\cdots \min_{u(N-2)} E\left[\min_{u(N-1)} E(C|I^{N-1})|I^{N-2}\right]\cdots|I^0\right\},\end{aligned} \quad (3.15)$$

i.e., exactly (3.7). Note that the nestedness property (2.3) of the sequence I^k was used above.

IV. Dual Effect: Caution and Probing

The solution of multistage stochastic decision processes, either in the general form (3.7) or in the stochastic dynamic programming form (3.12) for an additive cost is a formidable problem. Unless an explicit form is found for the optimal cost-to-go in (3.12) one cannot solve this functional equation except numerically. The curse of dimensionality [6] afflicted upon the deterministic dynamic programming is further compounded by the expectation operators in the stochastic case making it unsolvable with a few exceptions (in addition to numerical minimization, numerical calculation of the conditional expectations also has to be carried out, which is practically impossible).

The few exceptions are the linear-quadratic problem [1], [2], [7], the linear-exponential-quadratic-Gaussian problem [24] and a linear system with a special form cost (even powers of the state up to sixth) [25].

Since one cannot obtain the optimal stochastic controller it is of interest to find suitable approximations for the stochastic dynamic programming. Such an approximation should preserve the preposterior analysis property of the principle of optimality mentioned in the previous section and allow an assessment of the effect of uncertainties (imperfect information: present and future) on the controller and its performance.

The approximations of the stochastic dynamic programming fall in the following two classes.

1) Feedback Type Algorithms: In this case the control depends only on the current information

$$u(k) = u(k, I^k) \quad (4.1)$$

but does not use the prior statistical description of the future posterior information

$$p[y(j+1)|I^j], \qquad j \geq k. \quad (4.2)$$

2) Closed-Loop Type Algorithms: Such a controller utilizes feedback (4.1) *and* anticipates future feedback via (4.2), i.e., that the loop will stay closed.

Feldbaum [14] introduced the concept of dual effect in the control of stochastic dynamic systems. In a stochastic problem the control has, in general, two effects.

1) It affects the state (control action).

2) It affects the uncertainty of the state (augmented by the possibly unknown parameters).

A rather general mathematical definition of this has been given in [7] in terms of conditional central moments of the state vector. To illustrate it, let the conditional covariance of the state at k be

$$\Sigma(k|k) = E\{[x(k) - \hat{x}(k|k)][x(k) - \hat{x}(k|k)]'|I^k\} \quad (4.3)$$

where $\hat{x}(k|k)$ denotes the conditional mean. Then if $\Sigma(k|k)$ does not depend on the past controls U^{k-1}, the control has no dual effect (of second order), i.e., it is neutral. This is the case in linear dynamic systems with additive but not necessarily Gaussian noise [7], [32]. In nonlinear systems the state estimation accuracy is in general control dependent—the control has a dual effect.

If the system has unknown parameters, modeled as a realization of a vector valued random variable, the control values will affect, in general, the information about them derived from the measurements. Since having more accurate estimates of the system parameters is intuitively beneficial for the controller, the idea that the controller should enhance their identification is appealing. The initial control should account for the fact that it is applied to a system with parameters drawn from the prior distribution *and* for the fact that their value can be further identified during the process. This is the adaptive or learning feature of the controller. A simple example that illustrates the dual effect of the control is given in the Appendix.

Therefore, the controller can be used for "active information storage" (estimation enhancement or uncertainty reduction) via what has been called probing [14]. Note that only a "closed-loop" algorithm can do this active information gathering. On the other hand, the existence of uncertainty in the system, might have another effect. Since, in general, uncertainty in the system will increase the expected cost, the controller should be "cautious" not to increase further the effect of the existing uncertainties on the cost. A simple example to illustrate this "caution" effect is also given in the Appendix.

The open-loop feedback (OLF) control [1], which belongs to the feedback class, works well in some problems. Nevertheless, it can suffer from the "turn-off" phenomenon which can be avoided only by a closed-loop controller [15], [36]. As pointed out in [7] the optimal solution of the linear-quadratic control problem belongs to the feedback class because in this problem the control has no dual effect. Among the algorithms that belong to the feedback class are the heuristic certainty equivalence ("enforced separation") [10], [28], the self-tuning regulator [3], the cautious control [36], and the multiple model partitioned control [4], [13]. Algorithms of the closed-loop type are the wide-sense adaptive [8], [29], [30], the dual controllers of [27], [36], the innovations dual controller of [20], and the model adaptive dual controller for multiple models [37].

V. Caution and Probing Effects from the Stochastic Dynamic Programming

The previous discussion pointed out qualitatively that a controller

1) has a direct control effect on the state;

2) can perform active information gathering (probing) to improve the accuracy of subsequent control actions; and

3) has to be cautious because of the existing uncertainties in the system.

While there is no universal agreement on the notions of caution and probing this author believes these concepts are valuable in the derivation of suboptimal algorithms. In this section a quantification of the above properties is presented. This is obtained by an approximation of the optimal cost from the stochastic dynamic programming that results in a decomposition of the cost into three terms, each associated with one of the above items.

The stochastic dynamic programming equation (3.12) is approximated as follows [8], [29], [30]. First, instead of the exact information state, the following approximate "wide-sense" information state is used:

$$\mathcal{P}^k=\{\hat{x}(k|k),\Sigma(k|k)\}, \qquad (5.1)$$

i.e., the (approximate) conditional mean and covariance of $x(k)$ obtained, e.g., via an extended Kalman filter. The use of this "quasi-sufficient statistic" is needed for an algorithm that is implementable. Assume now that the system is at time k and a closed-loop control (in the sense defined earlier) is to be computed using $\mathcal{P}^k$ *and* the present knowledge (statistical) about the future observations.

The principle of optimality with the information state (5.1) yields the following stochastic dynamic programming equation for the closed-loop-optimal expected cost-to-go at time k

$$J^*(k,\mathcal{P}^k)=\min_{u(k)}E\{c[k,x(k),u(k)] + J^*(k+1,\mathcal{P}^{k+1})|\mathcal{P}^k\}. \qquad (5.2)$$

The main problem is to obtain an approximate expression for $E\{J^*(k+1,\mathcal{P}^{k+1})|\mathcal{P}^K\}$ preserving its closed-loop feature, i.e., this expression should incorporate the "value" of the future observations. In order to find an explicit solution, the cost-to-go $C(k+1)$ defined in (3.8) is expanded about a nominal trajectory (designated by subscript 0) generated by the recursion

$$x_0(j+1)=f[j,x_0(j),u_0(j),\bar{v}(j)], \qquad j=k+1,\cdots,N-1 \qquad (5.3)$$

where $u_0(j)$, $j=k+1,\cdots,N-1$ is a sequence of nominal controls and $\bar{v}(j)$ is the mean of the process noise. The initial condition $x_0(k+1)$ is taken as the predicted value of the state at $k+1$ given $\mathcal{P}^k$ and the control (yet to be found) $u(k)$. The expansion of the cost-to-go from time $k+1$ is

$$C(k+1)=C_0(k+1)+\Delta C_0(k+1) \qquad (5.4)$$

where $C_0(k+1)$ is the cost along the nominal (ignoring all the uncertainties) and $\Delta C_0(k+1)$ is the variation of the cost about the nominal with terms up to second order obtained from a Taylor expansion, which will capture the stochastic effects. The approximation of the closed-loop-optimal expected cost-to-go from time $k+1$ is done now as follows:

$$J^*(k+1)=C_0(k+1)+\Delta J_0^*(k+1) \qquad (5.5)$$

where the optimal "closed-loop" perturbation cost is

$$\Delta J_0^*(k+1) = \min_{\delta u(k+1)} E\Big\{\cdots \min_{\delta u(N-1)} E\big[\Delta C_0(k+1)|\mathcal{P}^{N-1}\big]\cdots|\mathcal{P}^{k+1}\Big\} \qquad (5.6)$$

and $\delta u(k)=u(k)-u_0(k)$. This minimization problem is quadratic since, by construction, $\Delta C_0(k+1)$ is quadratic in $\delta u(j)$, $k+1\leq j\leq N-1$ as well as in the variations about the nominal trajectory, $\Delta x(j)=x(j)-x_0(j)$, $k+1\leq j\leq N$. Using a Taylor series expansion of (2.1) and including second-order terms results in a set of perturbation state equations in $\delta x(j)$ with $\delta x(k+1)=x(k+1)-x_0(k+1)$ as an initial condition. Thus, the problem posed in (5.6) consists of minimizing a quadratic cost given a quadratic system of state equations, and is somewhat similar to the linear-quadratic control problem. Then, by assuming a solution quadratic in the perturbed state (i.e., neglecting higher order terms) and evaluating the expectations permits the optimal closed-loop (CL) cost-to-go to be obtained explicitly. See [8] for the development of the details. This result, obviously, depends on the approximations used in the derivation.

The Cost Decomposition

The explicit expression of the (approximate) cost obtained can be decomposed as follows:

$$J^{\text{CL}}(k)\triangleq J_D(k)+J_C(k)+J_P(k) \qquad (5.7)$$

where the subscript D stands for deterministic, C stands for caution, and P stands for probing components.

It will be assumed, for simplicity, that

$$c[k,x(k),u(k)]=c_1[k,x(k)]+c_2[k,u(k)] \quad (5.8)$$

and that the process noise, whose covariance is V, enters additively in (2.1). Then the deterministic component of the cost-to-go is, excluding c_1 (which does not depend on the control) is given by

$$J_D(k) \triangleq c_2[k,u(k)]+C_0(k+1)+\gamma_0(k+1) \quad (5.9)$$

and the stochastic terms obtained via the perturbation problem are

$$J_C(k) \triangleq 1/2 \operatorname{tr}[K_0(k+1)\Sigma(k+1|k)] + 1/2 \sum_{j=k+1}^{N-1} \operatorname{tr}[K_0(j+1)V(j)] \quad (5.10)$$

$$J_P(k) \triangleq 1/2 \sum_{j=k+1}^{N-1} \operatorname{tr}[\mathcal{C}_{0,xx}(j)\Sigma_0(j|j)]. \quad (5.11)$$

Σ is the covariance of the augmented state and γ, K, and $\mathcal{C}$ are given by appropriate recursions detailed in [8].

The stochastic term (5.10) reflects the effect of the uncertainty at time k summarized by $\Sigma(k|k)$ and subsequent process noises on the cost. These uncertainties cannot be affected by $u(k)$ but their weightings do depend on it, e.g., $\Sigma(k+1|k)$ depends on $\Sigma(k|k)$ and $u(k)$. 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 stochastic term (5.11) accounts for the effect of uncertainties when subsequent decisions (corrective actions) will be made. The weighting of these future uncertainties is nonnegative ($\mathcal{C}_{0,xx}$ is positive semidefinite). If the control can reduce by probing (experimentation) the future updated covariance, it can thus reduce the cost. The weighting matrix $\mathcal{C}_{0,xx}$ yields approximately the *value of future information* for the problem under consideration. Therefore, this is called the *probing* term. Note that even if the control has no dual effect, i.e., it does not affect the future covariance Σ of the augmented state (which includes the random parameters), the weighting of these covariances might still be affected by the control. Therefore, this (admittedly approximate) procedure accounts not only for the dual effect but all the stochastic effects in the performance index.

Thus, starting from the stochastic dynamic programming one can see the following: 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.

The closed-loop control $u(k)$ is found from the minimization of (5.7) using a search procedure. At every k to each control $u(k)$ for which (5.7) is evaluated during the search there corresponds a predicted state and to this predicted state a sequence of deterministic controls is attached that defines the nominal trajectory. 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 to be obtained.

The "quality" of the approximations used in the derivations outlined above, in particular, the second-order expansions, is an open question. Only extensive Monte Carlo simulations with rigorous comparison with other algorithms (see, e.g., [37]) can answer these questions. For some problems [29], [30] significant performance improvements have been found. In other cases where probing is not significant the CL algorithm performed close to the OLF [8].

The cost decomposition is believed to provide the only insight we now have towards the understanding of complex stochastic control problems for which the optimal solution is unknown. Furthermore, the classification of various stochastic control problems presented in the next section, which is based on this decomposition, can be used as a tool to assess for which nonlinear problems stochastic control algorithms can provide significant performance improvements.

VI. Implications of the Cost Decomposition and Examples

The decomposition of J^{CL} presented above yields an explicit evaluation of the tradeoffs between direct control, active probing, and a cautious action on the part of the controller. Thus, the ability of the control to affect learning as well as steer the system to its targets can be numerically evaluated using this decomposition. This is a particularly attractive feature for it captures both the need (and desire) of the controller to extract more information from the system as well as the aversion for drastic actions which may result in undesirable outcomes (risk aversion [12]). Furthermore, this also gives indication whether the uncertainty dominates the problem when the stochastic part of the cost (J_C+J_P) exceeds significantly the deterministic part (J_D).

If the uncertainty dominates the problem, then one can distinguish two cases.

1) The caution component J_C dominates. Then, since this is "uncontrollable" uncertainty, one has a highly uncertain model which cannot be improved in the course of the control period.

2) The probing component J_P dominates. Then, with the dual effect of the control, one can reduce the uncertainty of the model—thus, the model, while uncertain at the beginning, might prove to be ultimately adequate for the control problem under consideration.

A third case occurs when we have the following.

3) The deterministic component of the cost J_D dominates; then the parameter uncertainties are of no significant consequence.

The last case is the most desirable because then the controller can be of the certainty equivalence type [7], i.e., it can ignore the uncertainties by replacing all the random variables by their (conditional) means. This is the least expensive algorithm because it is essentially deterministic and will yield near optimum performance. However, the stochastic control approach outlined above has to be used to reach this conclusion.

Wonham [33] stated, about ten years ago, the following. In the case of (stochastic) feedback controls the general conclusion is that only marginal improvement can be obtained (over a controller ignoring the stochastic features), unless the disturbance level is very high; in this case the fractional improvement may be large but the system is useless anyway.

This statement implies that with high-level disturbances (in which one can include large parameter uncertainties) one has a "hopeless" situation. The other extreme is the situation with low level disturbances. These two situations seem to match, respectively, cases 1) and 3) from above. It was also pointed out in [33] that Feldbaum's dual control which probes the system might hold the promise of useful applications of stochastic control. However, at that time it was not clear whether there are such problems and, if yes, then how to obtain a (dual) controller that can effectively probe the system to reduce uncertainties. The wide-sense dual (or stochastic closed-loop) control algorithm [8], [29], presented in Section V, can then be used to obtain significant performance improvement.

As will be shown in the sequel, the cost decomposition presented above can answer affirmatively the question whether there are probing-dominated stochastic control problems, i.e., problems falling in case 2) from above.

In the following a number of examples are discussed to illustrate the usefulness of the cost decomposition and its implications. Some of these examples have appeared earlier in the literature and they are reexamined in light of the recently gained quantitative understanding of the caution and probing effects from the cost decomposition.

A. A Probing-Dominated Problem (Terminal Guidance)

The first example is the interception problem from [30]. In this case a third-order linear system with six unknown (random) parameters and both process and measurement noises was considered. The augmented nine-dimensional state (for which the dynamic equation is obviously nonlinear) had an initial estimate and an associated covariance. The elements of this covariance matrix corresponding to the parameters reflected the fact the initial estimates of the parameters were poor. The goal was to steer one of the (proper) state components to a target value by the terminal time, which was $N=20$. This was expressed by a quadratic term for the terminal state. There was no cost associated with the state prior to the terminal time and the cost weighting of the control, also entering quadratically, was low.

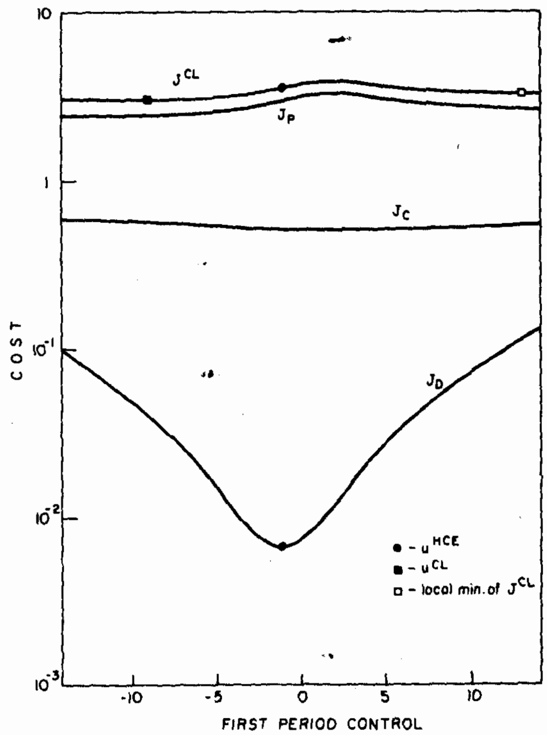


Fig. 1. Cost decomposition for a probing-dominated stochastic control problem (terminal guidance for a third-order system with six unknown parameters).

Fig. 1 presents the plot of the cost decomposition for the first period control. It can be seen that this is a probing-dominated stochastic control problem: the probing component of the cost is approximately 80 percent of the total cost.

The performance of the wide-sense dual [or closed-loop (CL)] control described in Section V was compared in [30] via Monte Carlo runs to the HCE (heuristic certainty equivalence) where the parameters' estimates were used as if they were the true values. The observed improvement of the CL algorithm versus HCE was, from (the modest number of) 20 Monte Carlo runs, around 85 percent [30]. This fractional improvement is quite close to the share of the probing cost from the total as indicated above. The CL controller, via its dual effect helped identify the system, i.e., it was actively adaptive and this was the key factor in its better performance. This decomposition, which was not known at the time of the original work [30], can now be used to provide the explanation for the observed performance improvement.

An important observation is that the probing component of the cost is not convex—the parameter identification is enhanced by large magnitude first period control values, both negative and positive. This lack of convexity of the probing component leads to local minima, as can be seem from Fig. 1. This phenomenon was pointed out in [27], [36]. The behavior of the multiple minima is discussed later in more detail.

The example discussed above, which is of the terminal state penalty type, belongs to the second class of problems, i.e., probing dominated.

B. A Caution-Dominated and an Essentially Deterministic Problem (Econometric Models)

Two additional problems, derived from econometrics are discussed next. Both are macroeconometric models of the U.S., derived from the same data but under different assumptions. For a concise description of the models see [9], [10]. The first econometric model has three states (gross national product, investment, and consumption), is driven by the government expenditures input, and has five unknown parameters characterized by an initial estimate and covariance matrix. The second econometric model has 11 states (as above plus increments of these variables and some lagged values), same input, and three unknown parameters.

The first model was obtained by Kendrick using ordinary least squares [17] while the second, more elaborate model, was obtained by Wall using the full information maximum likelihood method [34], [35]. The cost was quadratic in the deviations of the three economic variables and the input from target values along the entire trajectory consisting of seven periods (economic quarters).

The analysis of the cost $J^{\mathrm{CL}}(0)$ for the first econometric model, shown in Fig. 2, points to the fact that this problem is dominated by the caution term. This is due to the relatively large uncertainties in the initial parameter estimates. The probing component is negligible—this problem is completely dominated by the initial uncertainty—it belongs to the first class defined at the beginning of the section. Note that both the caution as well as the probing term tend to reduce the value of u^{CL} versus u^{HCE}, i.e., they are not conflicting in this case.

Fig. 3 shows the cost for the second econometric model. The deterministic component dominates here and $u^{\mathrm{CL}}(0)$ is very close to $u^{\mathrm{HCE}}(0)$. The probing component is again negligible. This problem belongs to the third class—it is essentially deterministic.

C. A Scalar Problem: Parametric Study of the Cost Shape

Another example of the application of the cost decomposition deals with a scalar linear system over $N=2$ time periods discussed in [19],

$$x(k+1)=ax(k)+bu(k)+v(k) \qquad k=0,1 \quad (6.1)$$

with $a=0.7$ known, the unknown input gain b with initial estimate $\hat{b}(0)=0.6$, and variance $\sigma_b^2(0)$. The process noise $v(k)$ is zero mean, white with variance V. The goal is to keep the state x, which is perfectly observed, around zero. This is expressed by the quadratic cost

$$C=1/2Q(2)x^2(2)+1/2r\left[u^2(0)+u^2(1)\right] \quad (6.2)$$

with terminal state weighting $Q(2)$ and control weighting $r=0.1$. The initial state is $x(0)=1$.

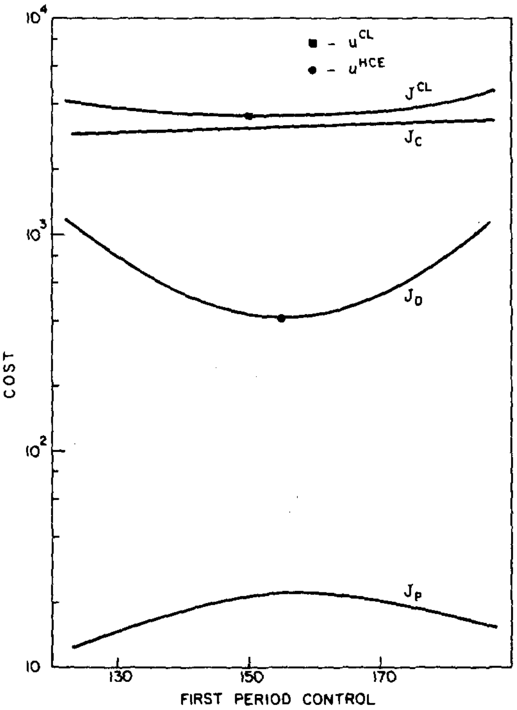


Fig. 2. Cost decomposition for a caution-dominated stochastic control problem (third-order econometric model with five unknown parameters).

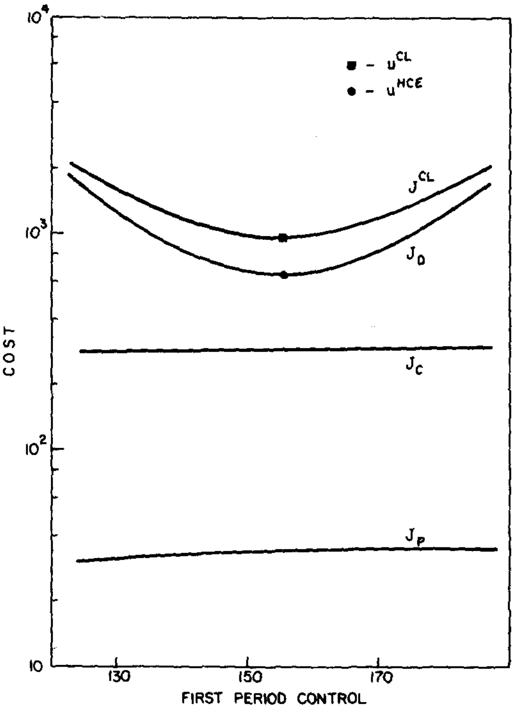


Fig. 3. Cost decomposition for an essentially deterministic stochastic control problem (11th-order econometric model with three unknown parameters).

Fig. 4 presents the cost decomposition at $k=0$ (first period) for the initial gain uncertainty $\sigma_b^2(0)=0.52$ and process noise variance $V=0.2$ and terminal state weighting $Q(2)=10$. The probing component of the cost, which varies drastically with the control, yields two minima for the total cost. It is of interest to see how these minima behave as the terminal state weighting changes. This is illustrated in Fig. 5. For even larger terminal weighting the two minima get further apart while for a lower weighting, $Q(2)=1$, there is

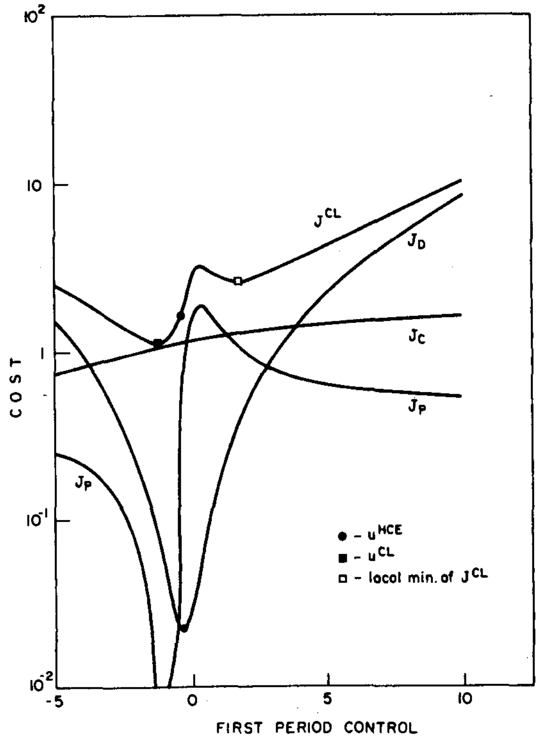


Fig. 4. Cost decomposition for a two-stage problem with unknown input gain (scalar system).

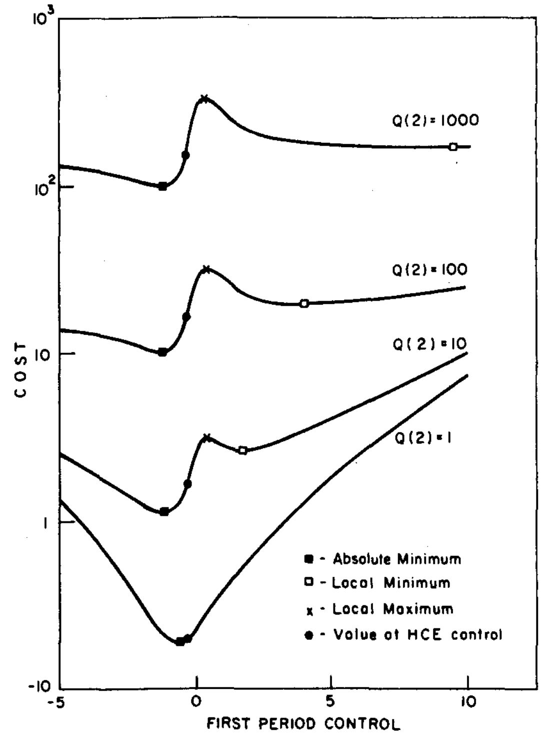


Fig. 5. The closed-loop cost for various terminal state weightings (scalar system).

only one minimum left. In this latter case the lighter terminal penalty does not justify a major control effort to identify accurately the parameter b and u^{CL} is quite close to u^{HCE}.

Another aspect of interest is how the anticipated future learning changes the present behavior of the CL controller. To this purpose the variance of the process noise was varied. Fig. 6 shows the cost $J^{\mathrm{CL}}(0)$ for $Q(2)=1000$, $\sigma_b^2=2$, and various values of V. For large process noise variance, less learning is anticipated and the cost curve is relatively flat, even though it has two minima, wide apart. For low process noise variance the cost curve has a very high maximum at $u(0)=0$ (when no learning of b occurs) and then two sharp minima around this point.

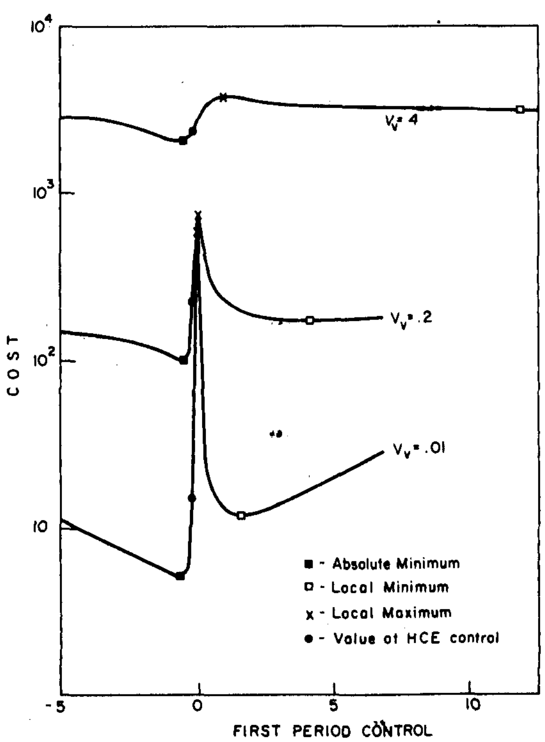


Fig. 6. Effect of the anticipated future learning on the control (scalar system).

VII. Conclusions

While still very few stochastic control problems have been solved optimally, insight into such problems can be gained by using the decomposition of the expected cost. This decomposition, based on the stochastic dynamic programming, yields three cost components: one deterministic and two stochastic ones. The stochastic terms quantify the effect of the various uncertainties on the performance index. The effects these stochastic terms have been associated with Feldbaum's concepts of caution and probing. Furthermore, this decomposition revealed three classes of stochastic control problems: caution dominated, probing dominated, and essentially deterministic. This, admittedly fuzzy, classification pointed out that there are stochastic control problems where significant improvements can be expected when using an appropriate sophisticated control algorithm. The examples show that one can assess, before extensive simulations, whether significant performance improvement can be expected in a stochastic control problem. It has also been shown that the various cost components can vary drastically with changes in the performance index weightings. The probing component of the cost can be nonconvex thus leading to local minima in the total cost.

Appendix
Simple Examples of Probing and Caution

Consider the scalar system

$$x(k+1)=ax(k)+bu(k)+v(k) \qquad (A.1)$$

with a known, b an unknown parameter with prior mean $\hat{b}(0)$ and variance $\sigma_b^2(0)$, and $v(k)$ a zero-mean white noise sequence with variance σ_v^2. Letting

$$I^k=\{X^k,U^{k-1}\}, \qquad (A.2)$$

i.e., perfect state observations, it follows that the uncertainty about parameter b at time $k+1$ is, from a standard least-squares argument, dependent on the control at k as follows:

$$\sigma_b^2(k+1)=\frac{\sigma_b^2(k)\sigma_v^2}{\sigma_b^2(k)u^2(k)+\sigma_v^2}. \qquad (A.3)$$

This clearly illustrates the control's dual effect, in addition to its effect on the state the control also affects the future information accuracy.

Consider next the same system with the (one-step horizon or myopic) cost

$$C(k)=x^2(k+1)+\lambda u^2(k). \qquad (A.4)$$

The control that minimizes

$$J(k)=E\{C(k)|I^k\} \qquad (A.5)$$

can be obtained easily as

$$u^*(k)=-\frac{ax(k)\hat{b}(k)}{\hat{b}^2(k)+\sigma_b^2(k)+\lambda}. \qquad (A.6)$$

Note that, because of the myopicity of the cost (A.4), this controller ignores any possibility of learning. On the other hand, because of the uncertainty in b, this control can be very cautious—a large variance $\sigma_b^2(k)$ can decrease significantly the value of the control in (A.6) compared to the case where there is no uncertainty in b or when this uncertainty is ignored as an HCE controller would do

$$u^{\mathrm{HCE}}(k)=-\frac{ax(k)\hat{b}(k)}{\hat{b}^2(k)+\lambda}. \qquad (A.7)$$

The optimal myopic controller (A.6) can then exhibit the turn-off phenomenon [15], [36]—it can be small because of large uncertainty in b and this will then prevent, according to (A.3), the reduction of this uncertainty.

Acknowledgment

The constructive comments of the anonymous reviewers are gratefully acknowledged.

References

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

[2] K. J. Åström, *Introduction to Stochastic Control Theory*. New York: Academic, 1976.

[3] K. J. Åström, U. Borisson, L. Ljung, and B. Wittenmark, "Theory and application of self-tuning regulators," *Automatica*, vol. 13, pp. 457–476, Sept. 1977.

[4] M. Athans *et al.*, "The stochastic control of the F-8C aircraft using a multiple model adaptive control (MMAC) method—Part I: Equilibrium flight," *IEEE Trans. Automat. Contr.*, vol. AC-22, pp. 768–780, Oct. 1977.

[5] R. Bellman, *Dynamic Programming*. Princeton, NJ: Princeton Univ. Press, 1957.

[6] —, *Adaptive Control Processes: A Guided Tour*. Princeton, NJ: Princeton Univ. Press, 1961.

[7] Y. Bar-Shalom and E. Tse, "Dual effect, certainty equivalence, and separation in stochastic control," *IEEE Trans. Automat. Contr.*, vol. AC-19, pp. 494–500, Oct. 1974.

[8] —, "Caution, probing, and the value of information in the control of uncertain systems," *Ann. Econ. Social Measurement*, vol. 5, pp. 323–337, Summer 1976.

[9] Y. Bar-Shalom, "Effects of uncertainties on the control performance of linear systems with unknown parameters and trajectory confidence tubes," *Ann. Econ. Social Measurement*, vol. 6, pp. 599–611, 1978.

[10] Y. Bar-Shalom and K. D. Wall, "Dual adaptive control and uncertainty effects in macroeconomic system optimization," *Automatica*, vol. 16, pp. 147–156, 1980.

[11] D. P. Bertsekas, *Dynamic Programming and Stochastic Control*. New York: Academic, 1976.

[12] M. DeGroot, *Optimal Statistical Decisions*. New York: McGraw-Hill, 1970.

[13] J. G. Deshpande, T. N. Upadhyay, and D. G. Lainiotis, "Adaptive control of linear stochastic systems," *Automatica*, vol. 9, pp. 107–115, 1973.

[14] A. A. Feldbaum, *Optimal Control Systems*. New York: Academic, 1965.

[15] D. J. Hughes and O. L. R. Jacobs, "Turn-off, escape and probing in nonlinear stochastic control," in *Proc. IFAC Symp. Adaptive Contr.*, Budapest, Hungary, Sept. 1974.

[16] O. L. R. Jacobs and J. W. Patchell, "Caution and probing in stochastic control," *Int. J. Contr.*, vol. 15, pp. 189–199, 1972.

[17] D. Kendrick, "Adaptive control of macroeconomic models with measurement error," in *Optimal Control for Econometric Models*, S. Holly, B. Rustem, and M. Zarrop, Eds. London: Macmillan, 1979.

[18] H. Kushner, *Introduction to Stochastic Control*. New York: Holt, 1971.

[19] E. C. MacRae, "Linear decision with experimentation," *Ann. Econ. Social Measurement*, vol. 1, pp. 437–447, 1972.

[20] R. Milito, C. Padilla, and R. Padilla, "An innovations approach to dual control," *IEEE Trans. Automat. Contr.*, vol. AC-26, 1981.

[21] G. L. Nemhauser, *Introduction to Dynamic Programming*. New York: Wiley, 1967.

[22] H. Raiffa and R. Schlaifer, *Applied Statistical Decision Theory*. Cambridge, MA: M.I.T. Press, 1972.

[23] M. H. Richardson, G. W. Keller, and R. E. Larson, "Principle of optimality and dynamic programming—Forward & backward," unpublished manuscript, 1972.

[24] J. L. Speyer, J. J. Deyst, and D. H. Jacobson, "Optimization of stochastic linear systems with additive measurement and process noise using exponential performance criteria," *IEEE Trans. Automat. Contr.*, vol. AC-19, pp. 358–366, Aug. 1974.

[25] J. L. Speyer, "A nonlinear control law for a stochastic infinite time problem," *IEEE Trans. Automat. Contr.*, vol. AC-21, pp. 560–564, Aug. 1976.

[26] C. Streibel, "Sufficient statistics in the optimum control of stochastic systems," *J. Math. Anal. Appl.*, vol., 12, pp. 576–592, Dec. 1965.

[27] J. Sternby, "A regulator for time-varying stochastic systems," in *Proc. 7th IFAC World Congr.*, Helsinki, Finland, June 1978.

[28] G. Stein and G. N. Saridis, "A parameter adaptive control technique," *Automatica*, vol. 5, pp. 731–739, 1969.

[29] E. Tse, Y. Bar-Shalom, and L. Meier, "Wide-sense adaptive dual control of stochastic nonlinear systems," *IEEE Trans. Automat. Contr.*, vol. AC-18, pp. 98–108, Apr. 1973.

[30] E. Tse and Y. Bar-Shalom, "An actively adaptive control for discrete-time systems with random parameters," *IEEE Trans. Automat. Contr.*, vol. AC-18, pp. 109–117, Apr. 1973.

[31] —, "Generalized certainty equivalence and dual effect in stochastic control," *IEEE Trans. Automat. Contr.*, vol. AC-20, pp. 817–819, Dec. 1975.

[32] W. M. Wonham, "On the separation theorem of stochastic control," *SIAM J. Contr.*, vol. 6, no. 2, pp. 312–326, 1968.

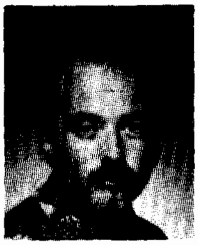
[33] —, "Optimal stochastic control," *Automatica*, vol. 5, pp. 113–118, Jan. 1969.

[34] K. D. Wall, "FIML estimation of rational distributed lag structural form models," *Ann. Econ. Social Measurement*, vol. 5, Winter 1976.

[35] K. D. Wall, Y. Bar-Shalom, and D. Kendrick, "Adaptive control of macroeconometric models with correlated measurement errors," presented at the 5th NBER Conf. Stochastic Contr. Econ., New Haven, CT, May 1977.

[36] B. Wittenmark, "An active suboptimal dual controller for systems with stochastic parameters," *Automat. Contr. Theory Appl.*, Canada, vol. 3, pp. 13–19, 1975.

[37] C. J. Wenk and Y. Bar-Shalom, "A multiple model adaptive dual control algorithm for stochastic systems with unknown parameters," *IEEE Trans. Automat. Contr.*, vol. AC-25, pp. 703–710, Aug. 1980.



Yaakov Bar-Shalom (S'63–M'66–SM'80) was born on May 11, 1941. He received the B.S. cum laude 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, NJ, in February 1970, all in electrical engineering.

From 1963 to 1966 he served in the Israel Defense Army as an Electronics Officer. From 1966 to 1967 he was a Teaching Assistant in the Department of Electrical Engineering of the Technion. Between 1967 and 1970 he was at Princeton University, first as a Research Assistant, then as a Research Associate. During the academic year 1968–1969 he was appointed Charles Grosvenor Osgood Fellow in recognition for outstanding performance. From 1970 to 1976 he was a Senior Research Engineer with Systems Control, Inc., Palo Alto, CA. At SCI he has developed and applied advanced stochastic control and estimation procedures to problems in the areas of defense, environmental, and economic systems. Between 1973 and 1976 he taught graduate courses in system theory at the University of Santa Clara. Currently, he is Professor of Electrical Engineering and Computer Science, University of Connecticut, Storrs. He has been consulting to Systems Control, Inc., ORI, Inc., System Development Corporation, C.S. Draper Laboratory, Bolt, Beranek and Newman, General Electric and CTEC, Inc. His research interests are in stochastic adaptive control, resource allocation under uncertainty, estimation theory, decision theory, system identification, and modeling, analysis and control of environmental and economic systems. During 1976 and 1977 he served as Chairman of the Technical Committee on Stochastic Control and Associate Editor of the IEEE TRANSACTIONS ON AUTOMATIC CONTROL and currently is Associate Editor of *Automatica*. He was Chairman of the 18th IEEE Symposium on Adaptive Process in 1979 and is Program Chairman of the 1982 American Control Conference. Also, he is Guest Associate Editor of this Special Issue.

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

A Multiobjective Dynamic Programming Method for Capacity Expansion

VIRA CHANKONG, YACOV Y. HAIMES, FELLOW, IEEE, AND DAVID M. GEMPERLINE

***Abstract*—This paper integrates two existing methodologies—a single-objective dynamic programming method for capacity expansion and the surrogate worth tradeoff (SWT) method for optimizing multiple objectives—into a unified schema. In particular it shows 1) how a multiobjective mixed integer programming formulation representing the multiobjective capacity expansion problem can be translated into a multiobjective dynamic programming formulation, 2) how such DP formulation can be used to generate noninferior solutions, and 3) how tradeoff information can be obtained from solutions in 2). The necessary theoretical machinery for 3) is developed. To demonstrate the computational viability of the proposed schema, an example problem is formulated and solved.**

I. INTRODUCTION

THE PROBLEM of optimal scheduling, sequencing, construction, and capacity expansion of water resource projects, energy generation, and storage units, etc., has been on the planners' agenda for a long time. For simplicity, the above problem will be referred to here as the "capacity expansion" problem. Numerous mathematical models addressing this problem have been suggested in the literature.

This paper extends a well-documented dynamic programming schema for optimal scheduling, sequencing, and capacity expansion of water resource projects from a single objective function to multiple objective functions. The extended schema—multiobjective dynamic programming—incorporates fixed as well as variable cost elements associated with the candidate projects, as well as other relevant planning objectives.

To develop the multiobjective dynamic programming (MODP) schema for the capacity expansion problem in a pedagogical way, it is imperative that two other methodolo-

Manuscript received December 1, 1980; revised March 20, 1981 and May 27, 1981. This work was supported in part by the National Science Foundation under Grant ENG79-03605 and in part by the U.S. Department of Energy for the project entitled "Modeling and Synthesis of Energy Storage Devices in Power Systems," and Contract DEACO-180-RA-50256 for the project entitled "Industrial Functional Modeling."

V. Chankong was on leave at the Department of Systems Engineering, Case Institute of Technology, Case Western Reserve University, Cleveland, OH 44106. He is with the Department of Electrical Engineering, Khon Kaen University, Khon Kaen, Thailand.

Y. Y. Haimes is with the Department of Systems Engineering, Case Institute of Technology, Case Western Reserve University, Cleveland, OH 44106.

D. M. Gemperline was with the Department of Systems Engineering, Case Institute of Technology, Case Western Reserve University, Cleveland, OH 44106. He is now with the Water Resources Planning Section, Harza Engineering Company, Chicago, IL 60004.

0018-9286/81/1000-1195$00.75 ©1981 IEEE