Solutions to a Class of Nonstandard Stochastic Control Problems with Active Learning

TAMER BAŞAR, FELLOW, IEEE

Abstract—**We formulate and solve a dynamic stochastic optimization problem of a nonstandard type, whose optimal solution features active learning. The proof of optimality and the derivation of the corresponding control policies is an indirect one, which relates the original single-person optimization problem to a sequence of nested zero-sum stochastic games. Existence of saddle points for these games implies the existence of optimal policies for the original stochastic control problem, which, in turn, can be obtained from the solution of a nonlinear deterministic optimal control problem. The paper also studies the problem of existence of stationary optimal policies when the time horizon is infinite and the objective function is discounted.**

I. Introduction

One of the major challenges of optimum stochastic control theory has been to deal effectively with problems which do not satisfy the condition of the *separation theorem* [8], known more casually as the *separation principle*, which refers to situations where the control and estimation (filtering) functions can be separated out and dealt with individually, either independently or sequentially. The simplest form in which the principle manifests itself is the so-called *certainty-equivalence* property which is responsible for the complete separation of the control and filtering functions in the linear-quadratic-Gaussian (LQG) optimal control problem. Here, the control does not affect the quality of information to be carried along the line, and the LQG problem therefore constitutes a prime example of a control system that is of the neutral type; moreover, the control design could be carried out independently of the noise corrupted measurements available at the stations.

If a stochastic control problem is not of the neutral type, there is the possibility that the control input will affect the information content of the measurements to be made at future stages, in which case, the estimation part cannot be decoupled from the control actions. In addition to the control actions impacting the filtering, the estimators developed will in turn impact the control actions and thereby their performance, thus bringing in an intricate interplay between the two. The presence of such multiple roles of control design (to improve control performance and simultaneously the quality of estimation—which are at times conflicting objectives) makes stochastic control problems with active learning quite intractable, both analytically and numerically, unless one resorts to some approximation schemes and is content with suboptimal laws [1], [6]. There is no general theory available for such problems, and one cannot even identify a subclass whose complete solution can be obtained.

The main objective of this paper is to define a class of discrete-time stochastic control problems of the nonneutral type, and to solve it for both finite and infinite horizons, by developing an indirect method for the derivation of optimal policies and for the proof of existence and optimality. The problem involves active learning, and is motivated by a macro-economics model of credibility and monetary policy developed recently (but not solved) in [5]. This is a model of monetary policy and inflation that incorporates asymmetric information between the private sector and the monetary authority. The former is a *passive* player who simply forms conditional (rational) expectations of the current inflation rate, which constitutes the surprise component of the policy maker's (the monetary authority) objective function. The policy maker tries to maximize the objective function by choosing a control policy which also affects the information carried to the passive player whose rational expectations in turn influence the performance of that policy (see also, [3] for a full discussion on and the economics aspects of this model).

The problem is formulated in precise mathematical terms in the next section as a nonstandard stochastic control problem. The complete solution to the two-stage version is provided in Section III, which clearly displays the active learning role played by the optimal control. An interesting aspect of the derivation and proof of optimality is that, even though the original problem is a single person stochastic optimization problem, one has to introduce a seemingly related stochastic zero-sum game and study its saddle-point solution. In Section IV we study the solution to the general finite-horizon problem with affine policies, and in Section V we discuss the more general case as well as the infinite-horizon case. The concluding remarks of Section VI end the paper. We should note at the outset that even though the original problem was motivated by a model arising in macro-economics, no background in economics is needed in order to follow the analysis and the methodology developed in this paper. We also feel that the solution technique introduced here for the first time should be of independent interest to researchers in stochastic control.

II. The General Mathematical Model

The rational expectations model alluded to in Section I (see the Appendix for an economic interpretation) leads to the (nonstandard) stochastic control problem where the objective is to maximize over $\gamma^N := (\gamma_0, \cdots, \gamma_N)$ the functional

$$J_N(\gamma^N) = E\sum_{i=0}^{N}(\beta)^i\left[x_i(u_i - E[u_i|I_i]) - \frac{1}{2}(u_i)^2\right] \tag{2.1}$$

subject to

$$x_{i+1} = \rho x_i + c_i + v_i \qquad i = 0, 1, \cdots \tag{2.2}$$

$$x_0 \sim N(\bar{x}_0, \sigma_0);\ v_i \sim N(0, \sigma_v),\ \rho \in (0, 1), \tag{2.3}$$

$$u_i = \gamma_i(I_i, x_i) \tag{2.4}$$

$$I_i = y^{i-1} := (y_0, \cdots, y_{i-1}) \tag{2.5}$$

$$y_i = u_i + w_i,\ w_i \sim N(0, \sigma_w) \tag{2.6}$$

Manuscript received August 21, 1987; revised April 12, 1988. Paper recommended by Associate Editor, D. A. Castanon. This work was supported in part by the U.S. Air Force Office of Scientific Research under Grant AFOSR 84-0056.

The author is with the Decision and Control Laboratory, Coordinated Science Laboratory and Department of Electrical and Computer Engineering, University of Illinois, Urbana, IL 61801.

IEEE Log Number 8823519.

0018-9286/88/1200-1122$01.00

where $\beta \in (0, 1)$ is a discount factor, γ_i is a general Borel measurable mapping, c_i is a scalar, and x_0, v_i, w_i are independent Gaussian random variables, for $i = 0, 1, \cdots, N$. Furthermore, $E[u_i | I_i]$ denotes the conditional expectation of the random variable u_i with respect to the sigma field generated by the information vector I_i, and $E\{\cdot\}$ denotes the unconditional expectation. When N is finite, we call this a *finite horizon problem* (with $N + 1$ stages), and if $N = \infty$ we will refer to it as an *infinite horizon problem*, in which case we replace γ^N in (2.1) formally by γ^∞.

We note that the control u_i enters the problem not through the state equation (2.2), but through the message process (2.6) and the cost function (2.1) to be minimized. The presence of the conditional expectation term in (2.1) makes this a nonstandard stochastic dynamic optimization problem not treated heretofore; furthermore, the problem is nonneutral since the choice of u_i has a direct effect on the content of the information carried by the measurement y_i regarding the state x_i.

Our first observation here is that since

$$E\{x_i E[u_i | I_i]\} = E\{E[x_i | I_i] u_i\} \tag{2.7}$$

the objective function (2.1) can equivalently be written as

$$J_N(\gamma^N) = E\left\{\sum_{i=0}^{N} (\beta)^i \left[u_i(x_i - E[x_i | I_i]) - \frac{1}{2}(u_i)^2\right]\right\} \tag{2.8}$$

which is a more convenient form to work with.

Our second observation is a reiteration of the earlier remark that even though the stochastic control problem formulated above is one with perfect information, it is not of the standard type because of the presence of the conditional expectation term $E[x_i | I_i]$ in the cost functional, which depends on the past control values $u_{i-1}, \cdots, u_0$.

Because of the nonstandard nature of the problem, it will be illuminating first to study the two-stage version, which clearly displays the dual role control plays in these stochastic optimization problems. The complete solution to this problem, presented in the next section, and the method of derivation and verification of optimality should be of independent interest.

III. The Two-Stage Version: A Complete Solution

In view of (2.8), the problem here is to maximize

$$J_1(\gamma^1) = E\left\{\beta\left[u_1(x_1 - E[x_1 | I_1]) - \frac{1}{2}(u_1)^2\right] + u_0(x_0 - \bar{x}_0) - \frac{1}{2}(u_0)^2\right\} \tag{3.1a}$$

over $\gamma^1 = (\gamma_0, \gamma_1)$, where

$$u_0 = \gamma_0(x_0), \quad u_1 = \gamma_1(x_1, y_0)$$
$$I_1 = \{y_0\}, \quad y_0 = u_0 + w_0$$
$$x_1 = \rho x_0 + c_o + v_0. \tag{3.1b}$$

Since J_1 is quadratic and strictly concave in γ_1, it has a unique maximum over γ_1 for each γ_0, given by

$$\gamma_1^*(x_1, y_0) = x_1 - E[x_1 | y_0] \tag{3.2}$$

yielding

$$\max_{\gamma_1} J_1(\gamma_0, \gamma_1) = E\left\{\frac{1}{2}\beta(x_1 - E[x_1 | y_0])^2 + u_0(x_0 - \bar{x}_0) - \frac{1}{2}(u_0)^2\right\} := F(\gamma_0). \tag{3.3}$$

Since

$$x_1 - E[x_1 | y_0] = \rho(x_0 - E[x_0 | y_0]) + v_0$$

expression (3.3) can equivalently be written as

$$F(\gamma_0) = \frac{1}{2}\beta\sigma_v + E\left\{\frac{1}{2}\beta\rho^2(x_0 - E[x_0 | y_0])^2 + u_0(x_0 - \bar{x}_0) - \frac{1}{2}(u_0)^2\right\} \tag{3.4}$$

$$\equiv \frac{1}{2}\beta(\sigma_v + \rho^2\sigma_0) - E\left\{\frac{1}{2}\beta\rho^2(E[\tilde{x}_0 | y_0])^2 + \frac{1}{2}(u_0)^2 - u_0\tilde{x}_0\right\} \tag{3.5}$$

where $\tilde{x}_0 := x_0 - \bar{x}_0$, and the last line follows since the error $x_0 - E[x_0 | y_0]$ is orthogonal to the estimate $E[x_0 | y_0]$.

Note that the control u_0 enters F both *directly* (through the last two terms, which are quadratic) and also *indirectly* through the conditional estimate $E[x_0 | y_0]$, which is *nonquadratic* in u_0. This is a clear illustration of the *dual* role of control, leading to an optimization problem of the type not treated heretofore.

Now, if F had not included the term $(E[\tilde{x}_0 | y_0])^2$, its maximum would be attained uniquely by a linear function of $\tilde{x}_0$.[1] Motivated by this appealing structure, we now restrict γ_0 to lie in the class $\gamma_0(x_0) = L(x_0 - \bar{x}_0)$, L being an arbitrary scalar, and maximize F over this class. Since

$$E[\tilde{x}_0 | y_0 = L\tilde{x}_0 + w_0] = \frac{L\sigma_0}{L^2\sigma_0 + \sigma_w} y_0$$

$F(\gamma_0)$ can be evaluated on this class to be

$$\tilde{F}(L) = \frac{1}{2}\beta(\sigma_v + \rho^2\sigma_0) - \frac{1}{2}\beta\rho^2\frac{L^2\sigma_0^2}{L^2\sigma_0 + \sigma_w} - \frac{1}{2}L^2\sigma_0 + L\sigma_0 \tag{3.6}$$

which is continuous and bounded above, and therefore admits a maximum. By differentiating $\tilde{F}(L)$ with respect to L and setting the resulting expression equal to zero, we obtain the equation

$$1 - L = \frac{L\sigma_0\sigma_w}{(L^2\sigma_0 + \sigma_w)^2}\beta\rho^2$$

which can be rewritten as

$$1 - L = \frac{Lp_0}{(L^2p_0 + 1)^2}\beta\rho^2 \tag{3.7}$$

where

$$p_0 := \sigma_0/\sigma_w.$$

The polynomial equation (3.7) will admit one or three or five real solutions, but every such solution will lie in the open interval (0, 1). Let L_0 be that real solution of (3.7) that provides the largest value for $\tilde{F}(L)$ defined by (3.6). If there is more than one maximizing solution, then L_0 could be taken to be any one of them, which we will henceforth refer to as an $\tilde{F}$-maximizing solution of (3.7). We are now in a position to state our first result.

Proposition 3.1: When γ_0 is restricted to the class of affine policies, the stochastic control problem (3.1) admits the globally

[1] The minimum of $(E[\tilde{x}_0 | y_0])^2$, on the other hand, is zero and is attained by choosing $\gamma_0(x_0) = 0$, again showing the conflicting roles of control in this optimization problem.

optimal solution

$$\gamma_1^*(x_1, y_0) = x_1 - E[x_1 | y_0] \equiv x_1 - \hat{x}_{1|0}, \tag{3.8a}$$

$$\gamma_0^*(x_0) = L_0(x_0 - \bar{x}_0) \tag{3.8b}$$

where L_0 is an $\tilde{F}$-maximizing solution of (3.7), $0 < L_0 < 1$, and

$$\hat{x}_{1|0} = \rho \hat{x}_{0|0} + c_0$$

$$\hat{x}_{0|0} = \bar{x}_0 + \frac{L_0 p_0}{L_0^2 p_0 + 1} y_0.$$

□

The question now is whether it would be possible to improve upon the performance attained under (γ_0^*, γ_1^*) above by going outside the class of affine policies for γ_0^*. The following theorem provides a set of (sufficient) conditions under which this is not possible, and thus the pair (3.8) is overall optimal.

Theorem 3.1: The policy pair (γ_0^*, γ_1^*) given in Proposition 3.1 is the *unique globally optimal* solution of (3.1) if there exists an $\tilde{F}$-maximizing solution L_0 of (3.7), satisfying the strict inequality

$$L_0(1 - L_0)p_0 < 1. \tag{3.9}$$

Proof: Before proceeding with the proof, it is important to note that implicit in the statement of the theorem is the property that there can be at most one $\tilde{F}$-maximizing solution of (3.7) that satisfies (3.9), which is a result we are also going to prove. Now, the proof of the theorem is an indirect one, relating the maximizing solution of (3.4) to the saddle point of an auxiliary zero-sum game. Towards this end, introduce a two-person zero-sum game in normal form [2], defined by the kernel

$$G(\delta, \gamma) = E\left\{\frac{1}{2}\beta\rho^2(\delta(y_0) - x_0)^2 + u_0(x_0 - \bar{x}_0) - \frac{1}{2}(u_0)^2\right\} + \frac{1}{2}\beta\sigma_v \tag{3.10}$$

which is to be minimized by a suitable choice of $\delta = \delta(y_0)$ and maximized by an appropriate $u_0 = \gamma(x_0)$, where $y_0 = u_0 + w_o$ is as defined before. We now claim that this game admits the unique saddle-point solution

$$\delta^*(y_0) = \bar{x}_0 + \frac{L_0 p_0}{L_0^2 p_0 + 1} y_0; \quad \gamma^*(x_0) = L_0(x_0 - \bar{x}_0) \tag{3.11}$$

under the condition (3.9), where L_0 is the unique $\tilde{F}$-maximizing solution of (3.7). To prove this claim, it will be sufficient to verify the validity of the two saddle-point inequalities

$$G(\delta^*, \gamma^*) \le G(\delta, \gamma^*), \quad \text{for all } \delta$$

$$G(\delta^*, \gamma^*) \ge G(\delta^*, \gamma), \quad \text{for all } \gamma$$

and show that the optimum solution is unique in each case. Initially, we do not know whether L_0 is unique, therefore, we take it as any $\tilde{F}$-maximizing solution of (3.7) that satisfies (3.9). Then, the first inequality above follows since $G(\delta, \gamma)$ is minimized over δ for any γ uniquely by the conditional mean of x_0, and when $\gamma = \gamma^*$, this conditional mean is affine in y_0 as given. For the second inequality, note that $G(\delta^*, \gamma)$ is a quadratic function of $u_0 = \gamma(x_0)$, with the coefficient of $(u_0)^2$ being

$$-\frac{1}{2}\left(1 - \frac{L_0^2 p_0^2}{(L_0^2 p_0 + 1)^2}\beta\rho^2\right) =: \alpha.$$

The condition $\alpha < 0$ directly implies that $G(\delta^*, \gamma)$ is a strictly concave function of γ, and being quadratic, it admits a unique solution, which is

$$\gamma(x_0) = \left\{\left(1 - \frac{L_0 p_0}{L_0^2 p_0 + 1}\beta\rho^2\right) \Big/ \left[1 - \frac{L_0^2 p_0^2}{(L_0^2 p_0 + 1)^2}\beta\rho^2\right]\right\}(x_0 - \bar{x}_0).$$

Using the fact that L_0 satisfies (3.7), the gain term above can be simplified to give

$$\gamma(x_0) = L_0(x_0 - \bar{x}_0)$$

thus verifying the validity of the second saddle-point inequality, under the condition $\alpha < 0$. Again using the fact that L_0 satisfies (3.7), α can be simplified to

$$\alpha = \frac{1}{2}[L_0(1 - L_0)p_0 - 1]$$

and, hence the concavity condition is indeed equivalent to (3.9). Note that under this condition, $G(\delta^*, \gamma)$ admits a *unique* maximum. Now, using the ordered interchangeability property of multiple saddle-point equilibria [2, p. 24], it readily follows that (3.11) is indeed the *unique* saddle-point solution of G under (3.9), and hence that there can be at most one $\tilde{F}$-maximizing solution of (3.7) under (3.9). Otherwise, there will be at least two different γ^*'s maximizing $G(\delta^*, \gamma)$ for the same δ^*, which is impossible since the kernel is strictly concave under (3.9).

Thus, completing verification of our claim on the saddle point of G, we now proceed with the proof of the theorem. A crucial observation now is the equality

$$\max_{\gamma} F(\gamma) = \max_{\gamma} \min_{\delta} G(\delta, \gamma)$$

that is, the unrestricted maximum value of our function F is equal to the lower value of the game with kernel G. Since the upper and lower values of the game are equal under (3.9), and the saddle point is unique and linear, it follows that F admits a unique maximizing policy which is linear in $\bar{x}_0$. □

Remark 3.1: The condition of Theorem 3.1 is sufficient for the affine solution to be overall maximizing, but there is no indication that it is also necessary; in fact, it is quite plausible that the result is valid for all values of the parameters defining the problem. Nonsatisfaction of (3.9) simply means that the auxiliary game defined in the proof of the theorem does not admit a saddle point; that is, the upper value is strictly larger than the lower value; however, this does not outrule the possibility that the maximizing solution for $F(\gamma)$ is still affine. □

Condition (3.9) was given in terms of the solution of (3.7). It is possible, however, to derive another condition, directly in terms of the parameters of the problem, which would guarantee satisfaction of (3.9). The following corollary (to Theorem 3.1) does precisely that.

Corollary 3.1: The pair of policies given in Proposition 3.1 is the unique globally optimal solution if

$$p_0\rho^2\beta < 4. \tag{3.12}$$

Proof: The result follows from Theorem 3.1 if we can show that (3.12) implies (3.9). Towards this end rewrite (3.9), in view of (3.7), as

$$1 - \frac{L_0^2 p_0^2}{(L_0^2 p_0 + 1)^2}\rho^2\beta$$

which is bounded below by

$$\ge 1 - \frac{1}{4}p_0\rho^2\beta$$

which further is positive by (3.12). □

Remark 3.1: Condition (3.12) above is in fact a very reasonable condition which is satisfied in all practical cases. Note that since $\rho^2\beta < 1$, it is satisifed whenever $\sigma_0 \le 4\sigma_w$, that is, the variance of the initial state x_0 should not be much larger than the variance of the noise w_0. □

IV. Solution to the Finite Horizon Problem

Guided by the results of the previous section, we now seek a solution to the general finite horizon problem formulated in Section II, first in the class of policies

$$u_i=\gamma_i(x_i, I_i)=\mu_i(x_i-E[x_i|I_i]), \qquad i=0, 1, \cdots \tag{4.1a}$$

where μ_i is a general linear mapping, say

$$\mu_i(x)=L_ix, \qquad \text{for all } x \in \mathbb{R}. \tag{4.1b}$$

Later we will show (in Proposition 4.2) that the optimum solution obtained in this class is in fact optimum in the larger class of policies where γ_i above is allowed to be any affine mapping, and furthermore (in Section V) that under certain conditions it is optimal even in the class of general nonlinear policies.

The first step in our derivation is to obtain a recursive algorithm for $E[x_i|I_i]$ when γ^N is restricted to the class given above. Note that x_i is generated by

$$x_{i+1}=\rho x_i+c_i+v_i, \qquad i=0, 1, \cdots \tag{4.2a}$$

and the measurement equation is

$$y_i=L_i(x_i-E[x_i|I_i])+w_i; \qquad i=1, \cdots. \tag{4.2b}$$

Since the underlying statistics are Gaussian, $E[x_i|I_i]$ can be computed recursively using the Kalman filter, with the corresponding expressions given in the following lemma.

Lemma 4.1: With $u_i = L_i(x_i - E[x_i|I_i])$, $i = 0, \cdots, N$, the expressions for the conditional mean $\hat{x}_{i|i-1} := E[x_i|I_i]$ can be obtained recursively by

$$\hat{x}_{i|i-1}=\rho\hat{x}_{i-1|i-1}+c_i, \; \hat{x}_{0|-1}=\bar{x}_0 \tag{4.3a}$$

$$\hat{x}_{i|i}=\hat{x}_{i|i-1}+\frac{L_i\sigma_{i|i-1}}{L_i^2\sigma_{i|i-1}+\sigma_w}y_i$$

$$\sigma_{i+1|i}=\rho^2\sigma_{i|i}+\sigma_V$$

$$\sigma_{i|i}=\sigma_{i|i-1}\sigma_w/(L_i^2\sigma_{i|i-1}+\sigma_w); \; \sigma_{0|-1}=\text{var}\,(x_0)=\sigma_0 \tag{4.3b}$$

where

$$\sigma_{i+1|i} := E[(x_{i+1}-\hat{x}_{i+1|i})^2].$$ □

Now define

$$p_i := \sigma_{i|i-1}/\sigma_w$$
$$r := \sigma_v/\sigma_w$$

and note that p_i is generated by the recursive equation

$$p_{i+1}=r+\rho^2p_i/(L_i^2p_i+1), \qquad p_0=\sigma_0/\sigma_w. \tag{4.4}$$

Hence, when the policies are restricted to the form (4.1), the stochastic control problem becomes equivalent to the deterministic optimal control problem

$$\underset{\{L_o,\cdots,L_N\}}{\text{maximize}}\; \tilde{J}_N(L^N) := \sum_{i=0}^{N}\beta^i\left[L_ip_i-\frac{L_i^2}{2}p_i\right]\sigma_w \tag{4.5}$$

under the dynamic (state equation) constraint (4.4). We now show that this optimal control problem admits a unique solution, and we characterize the solution in terms of a recursive equation.

A. Preliminary Notation

For each positive scalar p, let $\{W_k(p)\}_{k=0}^N$ be defined recursively by

$$W_{N+1}(p) \equiv 0 \tag{4.6a}$$

$$W_k(p)=\sup_L\left\{\left(L-\frac{L^2}{2}\right)p+\beta W_{k+1}\left(\frac{\rho^2p}{L^2p+1}+r\right)\right\}, \quad k=N, N-1, \cdots, 0. \tag{4.6b}$$

Let $L_k = L_k(p)$ be a maximizing solution for the RHS (right-hand side) of (4.6b), whenever it exists, and let $\{p_k^*\}_{k=0}^N$ be a trajectory sequence defined recursively by

$$p_{k+1}^*=\frac{\rho^2p_k^*}{L_k^2(p_k^*)p_k^*+1}+r; \qquad p_0^*=p_0. \tag{4.7}$$

Finally, let

$$L_k^* := L_k(p_k^*), \qquad k=0, 1, \cdots, N. \tag{4.8}$$

Proposition 4.1:

i) The maximization problem (4.6b) admits a solution for each positive p.

ii) The control problem (4.4), (4.5) admits a solution $\{L_k^*\}_{k=0}^N$ which is given by (4.8), and the corresponding optimum trajectory is generated by (4.7).

iii) The optimum solution satisfies

$$0\le L_k^*\le 1,\; k<N, \text{ and } L_N^*=1.$$

iv) The maximum value of (4.5) is $W_0(p_0)\sigma_w$.

To set the stage for the proof of this result, we will state and prove a number of auxiliary lemmata.

Lemma 4.2: The value of the optimal control problem (4.4), (4.5) is $\tilde{J}_N^* = W_0(p_0)\sigma_w$, where $W_0(\cdot)$ is obtained through the recursive equation

$$W_{N+1} \equiv 0$$

$$W_k(p)=\sup_L\left\{\left(L-\frac{L^2}{2}\right)p+\beta W_{k+1}\left(\frac{\rho^2p}{L^2p+1}+r\right)\right\}$$

$$:= \sup_L G_k(p, L), \qquad k\le N. \tag{4.9}$$

Furthermore, if the RHS above admits a solution $L_k(p)$, $k \le N$, then $L_k(p_k^*)$, $k \le N$ is the optimal solution where p_k^* is generated by (4.7).

Proof: This follows from a standard dynamic programming argument (see, e.g., [4] and [7]). □

Lemma 4.3: For each $p > 0$, we have the bounds

$$0< W_k(p)\le A_kp+a_k, \qquad k\le N \tag{4.10}$$

where $\{A_k\}$, $\{a_k\}$ are generated recursively by

$$A_N=\frac{1}{2}, \; A_k=\frac{1}{2}+\rho^2\beta A_{k+1}, \qquad k\le N-1$$

$$a_N=0, \; a_k=r\beta A_{k+1}+\beta a_{k+1}, \qquad k\le N-1.$$

Proof: The proof is by induction on k. Firstly, for $k = N$,

$$W_N(p)=\sup_L\left(L-\frac{L^2}{2}\right)p=\frac{1}{2}p \equiv A_Np$$

hence, the bound is exact for $k = N$. Now, we show that if the bound is valid for $k+1$, then it is also valid for k. Toward this end, consider the following sequence of inequalities:

$$\begin{aligned} W_k(p) &\le \sup_L \left(L-\frac{L^2}{2}\right)p+\beta \sup_L W_{k+1}\left(\frac{\rho^2 p}{L^2p+1}+r\right)\\ &\le \frac{p}{2}+\beta \sup_L \left(\frac{\rho^2 p A_{k+1}}{L^2p+1}+a_{k+1}+rA_{k+1}\right)\\ &< \frac{p}{2}+\rho^2 p\beta A_{k+1}+\beta a_{k+1}+\beta r A_{k+1}\\ &\equiv A_k p+a_k \end{aligned}$$

where the first inequality is a property of the supremum, the second inequality follows from the stipulated bound on W_{k+1}, and the third strict inequality follows since $p > 0$ and $\{A_k\}$ is a positive sequence. Since the upper bound was valid for $k = N$, by induction it is valid for all $k < N$.

To prove the strict lower bound, it is sufficient to see that $W_k(p)$ is nonnegative for all $k \le N$, and hence from (4.9)

$$W_k(p) \ge \sup_L \left(L-\frac{L^2}{2}\right)p > 0. \qquad \blacksquare$$

Lemma 4.4: For every $p > 0$, there exists a solution to the RHS of (4.9).

Proof: Let $G_k(p, L) := (L - L^2/2)p + \beta W_{k+1}([\rho^2 p/(L^2p+1)] + r)$ and note that if $W_{k+1}(p)$ is continuous in its argument, then $G_k(p, L)$ is jointly continuous in (p, L), and consequently $W_k(p) = \sup_L G_k(p, L)$ is continuous in p, provided that G_k is bounded above in L. Since $W_{N+1}(p)$ is zero (and thereby continuous in p), it follows by induction (on k) that $G_k(p, L)$ is jointly continuous in (p, L) for all $k \le N$, under the proviso that it is bounded above in L. However, using the bound (4.10) on W_k, we can readily see that $G_k(p, L) \downarrow -\infty$ as $L \to \pm\infty$. Hence, it is indeed bounded above in L, and furthermore there exists a finite positive constant K_p such that $\sup_L G_k(p, L) = \sup_{|L|\le K_p} G_k(p, L)$. It then follows from the Weierstrass theorem that a maximum exists, for all $k \le N$. □

Lemma 4.5:

i) $W_k(p)$ is an increasing function of p, $(W_k(p) \uparrow p)$, for all $k \le N$.

ii) If $L_k(p)$ is any maximizing solution, then for all $p > 0$, $0 \le L_k(p) \le 1$, $k \le N$.

Proof: The proof of i) is by induction on k. The result is clearly true for $k = N$, since $W_N(p) = (1/2)p$. Now, if $W_{k+1}(p) \uparrow p$, for each L, $W_{k+1}([\rho^2p/(L^2p+1)] + r) \uparrow p$ since $([\rho^2p/(L^2p+1)] + r) \uparrow p$. Also, $W_{k+1}([\rho^2p/(L^2p+1)] + r) \downarrow L^2$, and hence $\max_L G_k(p, L) = \max_{0\le L\le 1} G_k(p, L)$, since otherwise (i.e., for L outside the interval [0, 1]) both terms in G_k are decreasing in L (as $L \downarrow$ for $L < 0$ and as $L \uparrow$ for $L > 1$). Now, since $(L - L^2/2)p \uparrow p$ for $L \in (0, 1)$, and the second term of G_k was increasing in p for all L, as shown above, it follows that $W_k(p) = \max_L G_k(p, L) = \max_{0\le L\le 1} G_k(p, L) \uparrow p$. Therefore, the result is true for k if it is true for $k + 1 \le N + 1$, thus completing the induction. As a byproduct, we also obtain ii). □

Lemma 4.6:

$$W_k(p) > W_{k+1}(p) \qquad \forall k \le N,\ p > 0.$$

Proof: For $k = N$,

$$W_N(p) = \frac{1}{2} > W_{N+1}(p) = 0 \qquad \forall p > 0$$

which starts the induction process. We now show that $W_k(p) > W_{k+1}(p)$ if $W_{k+1}(p) > W_{k+2}(p)$, thus completing the induction process. This is accomplished by noting the following sequence of inequality and equalities:

$$\begin{aligned} &W_k(p) - W_{k+1}(p)\\ &= \max_L \left\{\left(L-\frac{L^2}{2}\right)p+\beta W_{k+1}\left(\frac{\rho^2p}{L^2p+1}+r\right)\right\}\\ &\quad -\max_L \left\{\left(L-\frac{L^2}{2}\right)p+\beta W_{k+2}\left(\frac{\rho^2p}{L^2p+1}+r\right)\right\}\\ &\ge \left(\hat{L}-\frac{\hat{L}^2}{2}\right)p+\beta W_{k+1}\left(\frac{\rho^2p}{\hat{L}^2p+1}+r\right)-\left(\hat{L}-\frac{\hat{L}^2}{2}\right)p\\ &\quad -\beta W_{k+2}\left(\frac{\rho^2p}{\hat{L}^2p+1}+r\right)\\ &= \beta\left[W_{k+1}\left(\frac{\rho^2p}{\hat{L}^2p+1}+r\right)-W_{k+2}\left(\frac{\rho^2p}{\hat{L}^2p+1}+r\right)\right] > 0 \end{aligned}$$

where $\hat{L} = L_{k+1}(p)$ is the argument of the second maximization, as introduced earlier. (If there is more than one such maximizing solution, any one of them could be chosen as $\hat{L}$.) □

Proof of Proposition 4.1: The proof follows from the sequence of Lemmas 4.2–4.6. □

The next result says that the structural restriction (4.1a) does not bring in any loss of generality in the class of affine policies.

Proposition 4.2: In the general formulation of Section II, let $u_i = \gamma_i(y^{i-1}, x_i) = L_i(y^{i-1}, x_i)$ where L_i is a general affine mapping for each $i = 0, 1, \cdots, N$. In this affine class, the maximum of $J_N(\gamma^N)$ is attained by

$$u_i = L_i^*(x_i - \hat{x}_{i|i-1}), \qquad i = 0, \cdots, N \tag{4.11}$$

where the sequence $\{L_i^*\}$ is defined by (4.8), and $\{\hat{x}_{i|i-1}\}$ is determined as in Lemma 4.1 with L_i replaced by L_i^*. Furthermore, the statement of Proposition 4.1 is valid over this larger class.

Proof: We first note that over the affine class, γ_i can be written as

$$u_i = L_i(x_i - E[x_i | y^{i-1}]) + l_i(y^{i-1}), \qquad i = 0, \cdots, N \tag{4.12}$$

where L_i is an arbitrary scalar and l_i is some arbitrary affine mapping. To save from notation let us denote the first term in (4.12) by $\tilde{u}_i$, and the second term by m_i, which are both random variables. Now, substituting this form into (2.1), and recognizing that $\tilde{u}_i$ and m_i are uncorrelated, we obtain the functional

$$\begin{aligned} J_N(\gamma^N) &= E\sum_{i=0}^{N}(\beta)^i\left[x_i(\tilde{u}_i - E[\tilde{u}_i | y^{i-1}]) - \frac{1}{2}(\tilde{u}_i)^2 - \frac{1}{2}(m_i)^2\right]\\ &\le E\sum_{i=0}^{N}(\beta)^i\left[\tilde{u}_i(x_i - E[x_i | y^{i-1}]) - \frac{1}{2}(\tilde{u}_i)^2\right] \end{aligned}$$

where in arriving at the last step we have used the fact that $(m_i)^2$ is a nonnegative random variable (hence, the inequality), and have also utilized the interchange of conditional expectations as in (2.7). Now, the final step of the proof is to recognize that the sigma field generated by y^i is the same as the sigma field generated by $\tilde{y}^i$, where

$$\tilde{y}_k = \tilde{u}_k + v_k \equiv y_k - m_k.$$

(This is true since m_k is y^{k-1}-measurable.) Hence,

$$J_N(\gamma^N) \le E\sum_{i=0}^{N}(\beta)^i\left[\tilde{u}_i(x_i - E[x_i | \tilde{y}^{i-1}]) - \frac{1}{2}(\tilde{u}_i)^2\right]$$

where the RHS is precisely the function that was maximized earlier in this section under the structural restriction (4.1a). Since this upper bound is attained (as shown in Proposition 4.1), the proof is completed. □

V. Optimality Over the Nonlinear Class, and the Infinite Horizon Case

In this section we provide two extensions to the analysis and results of Section IV: i) a study of the optimality of linear policies in the general class of nonlinear policies; and ii) the infinite-horizon problem.

A. Optimality Over the Nonlinear Class

In Section IV we have shown that the stochastic dynamic optimization problem formulated in Section II admits an optimal solution in the class of affine policies (cf. Proposition 4.2), whereas in Section III we had shown that for the two-stage version of the problem the linear solution is optimal even in the larger class of nonlinear policies. The proof given in Section III was an indirect one, relating the solution of the original problem to the saddle-point solution of a stochastic zero-sum game. The question now is whether that line of proof carries over to the general N-stage problem so as to establish overall optimality of the policies (4.11).

It turns out that a nontrivial extension of the game-theoretic approach is possible, where we now define, instead of a single game, a sequence of nested zero-sum games, each one imposing saddle-point existence conditions on the problem similar to that of (3.9). We provide below the essentials of our line of approach, and the construction of the nested games which are used in the proof of overall optimality. The procedure also leads recursively to precise conditions under which the linear solution of Proposition 4.2 would be overall optimal.

Now, starting with the objective functional (2.8), we first note the equality

$$\max_{\gamma^N} J_N(\gamma^N) = \max_{\gamma^{N-1}} J_{N-1}(\gamma^{N-1}) + \frac{1}{2} E\{(\beta)^N (x_N - E[x_N | I_N])^2\}$$

since $u_N = 1/2(x_N - E[x_N | I_N])$ is overall optimal. Denote the maximand on the RHS by $F_{N-1}(\gamma^{N-1})$, and introduce the zero-sum game ($N-1$th game) with kernel

$$G_{N-1}(\delta_N, \gamma^{N-1}) = J_{N-1}(\gamma^{N-1}) + \frac{1}{2} (\beta)^N E\{(x_N - \delta_N(I_N))^2\}.$$

Note that

$$\max_{\gamma^{N-1}} F_{N-1}(\gamma^{N-1}) = \max_{\gamma^{N-1}} \min_{\delta_N} G_{N-1}(\delta_N, \gamma^{N-1})$$

and hence a saddle-point solution for G_{N-1} would provide a maximizing solution for F_{N-1}. We now claim that (under certain conditions) the pair $\{\delta_N^* = \hat{x}_{N|N-1};\ \gamma_i^*(x_i, y^{i-1}) = L_i^*(x_i - \hat{x}_{i|i-1}),\ i = 0, \cdots, N-1\}$ provides a saddle-point solution for G_{N-1}, where the relevant terms were defined in Proposition 4.2. To prove this, we first observe that, with $\gamma^{N-1} = \gamma^{N-1*}$ as given, the optimality of δ_N^* follows readily from the Kalman filter theory; verification of the other side of the saddle-point inequality, however, is quite subtle, and is discussed next.

Now, the problem is to maximize $G_{N-1}(\delta_N^*, \gamma^{N-1})$ over γ^{N-1}, which is another nonstandard stochastic control problem and is as difficult to solve as the original one. Towards its solution we first note that, since

$$\hat{x}_{N|N-1} = \rho E[x_{N-1} | I_{N-1}] + c_N + \frac{pL_{N-1}\sigma_{N-1|N-2}}{L_{N-1}^2 \sigma_{N-1|N-2} + \sigma_w} y_{N-1}$$

$$\equiv \rho E[x_{N-1} | I_{N-1}] + c_N + K_{N-1} y_{N-1}$$

and

$$E(\hat{x}_{N|N-1} - x_N)^2 = \rho^2 E(x_{N-1} - E[x_{N-1} | I_{N-1}])^2 + K_{N-1}^2 E(u_{N-1})^2 - 2K_{N-1}\rho u_{N-1}(x_{N-1} - E[x_{N-1} | I_{N-1}]) + K_{N-1}^2 \sigma_w + \sigma_v$$

we have

$$\begin{aligned} G_{N-1}(\delta_N^*, \gamma^{N-1}) = {} & J_{N-2}(\gamma^{N-2}) + (\beta)^{N-1}(1 - \beta\rho K_{N-1}) \\ & \cdot E\{u_{N-1}(x_{N-1} - E[x_{N-1} | I_{N-1}])\} \\ & - \frac{1}{2} (\beta)^{N-1}(1 - \beta K_{N-1}^2) E(u_{N-1})^2 \\ & + \frac{1}{2} (\beta)^N \rho^2 E(x_{N-1} - E[x_{N-1} | I_{N-1}])^2 + k_{N-1} \end{aligned}$$

where

$$k_{N-1} := \frac{1}{2} (\beta)^N (\sigma_v + k_{N-1}^2 \sigma_w).$$

Now, the expression above can be maximized over γ_{N-1}, provided that $1 - \beta K_{N-1}^2 > 0$, in which case

$$\max_{\gamma^{N-1}} G_{N-1}(\delta_N^*, \gamma^{N-1}) = \max_{\gamma^{N-2}} \left\{ J_{N-2}(\gamma^{N-2}) + k_{N-1} + \frac{1}{2} (\beta)^{N-1} \tilde{p}_{N-1} E\{(x_{N-1} - E[x_{N-1} | I_{N-1}])^2\} \right\} \quad (*)$$

where the maximizing γ_{N-1} is

$$\gamma_{N-1}^*(x_{N-1}, y^{N-2}) = \frac{1 - \beta\rho K_{N-1}}{1 - \beta K_{N-1}^2} (x_{N-1} - E[x_{N-1} | I_{N-1}])$$

which is in the same form as the asserted optimum policy (4.11). In (*) above, p_{N-1} is defined by

$$p_{N-1} := [(1 - \beta\rho K_{N-1})^2 / (1 - \beta K_{N-1}^2)] + \beta\rho^2.$$

Furthermore, define the maximand in (*), without the k_{N-1} term, as $F_{N-2}(\gamma^{N-2})$, i.e.,

$$\max_{\gamma^{N-1}} G_{N-1}(\delta_N^*, \gamma^{N-1}) = \max_{\gamma^{N-2}} F_{N-2}(\gamma^{N-2}) + k_{N-1}.$$

Hence, to complete the solution of the $N-1$th game (with kernel G_{N-1}) we have to maximize $F_{N-2}(\gamma^{N-2})$ over γ^{N-2}. This is again a nonstandard stochastic control problem, for which we now introduce a new game (the $N-2$th game) with kernel

$$G_{N-2}(\delta_{N-1}, \gamma^{N-2}) = J_{N-2}(\gamma^{N-2}) + \frac{1}{2} (\beta)^{N-1} \tilde{p}_{N-1} E\{(x_{N-1} - \delta_{N-1}(I_{N-1}))^2\}$$

whose relationship with the latest maximization problem is

$$\max_{\gamma^{N-2}} F_{N-2}(\gamma^{N-2}) = \max_{\gamma^{N-2}} \min_{\delta_{N-1}} G_{N-2}(\delta_{N-1}, \gamma^{N-2}).$$

This new game is structurally similar to the $N-1$th one, and hence following the earlier procedure we can obtain only part of the saddle point, comprising the policies δ_{N-1}^* and γ_{N-2}^*, with the derivation of γ_{N-3}^* left to the next game in the sequence. Hence, this way, we define, recursively, a sequence of nested games

$\{G_n(\delta_{n+1}, \gamma^n)\}_{n=0}^{N-1}$, where

$$G_n(\delta_{n+1}; \gamma^n)=J_n(\gamma^n)+\frac{1}{2}(\beta)^{n+1}\tilde{p}_{n+1}E\{(x_{n+1}-\delta_{n+1}(I_{n+1}))^2\}$$

$$\tilde{p}_n := [(1-\beta\rho\tilde{p}_{n+1}K_n)^2/(1-\beta\tilde{p}_{n+1}K_n^2)]+\beta\rho^2\tilde{p}_{n+1}; \ \tilde{p}_N=1 \quad (5.1)$$

$$K_n := \rho L_n\sigma_{n|n-1}/(L_n^2\sigma_{n|n-1}+\sigma_w). \quad (5.2)$$

Under the condition that

$$1-\beta K_n^2\tilde{p}_{n+1}>0, \qquad n=N-1, N-2, \cdots, 0 \quad (5.3)$$

each game yields only part of the saddle-point solution, with the nth game yielding

$$\gamma_n^*(x_n, y^{n-1})=\frac{1-\beta\rho\tilde{p}_{n+1}K_n}{1-\beta\tilde{p}_{n+1}K_n^2}(x_n-E[x_n|I_n]).$$

The backward iteration halts when n reaches 0, in which case the saddle-point solution is obtained completely for the 0th game defined by the kernel $G_0(\delta_1, \gamma_0)$, which yields

$$\gamma_0^*(x_0)=\frac{1-\beta\rho\tilde{p}_1K_0}{1-\beta\tilde{p}_1K_0^2}(x_0-\bar{x}_0).$$

Now, moving in the opposite direction (forward in time) we complete the saddle-point solution of every game in the sequence, thus verifying that the very first game introduced, $G_{N-1}(\delta_N, \gamma^{N-1})$ admits a saddle point *only* in the structural form (4.11). Finally, consistency of the adopted procedure requires that

$$L_n=\frac{1-\beta\rho\tilde{p}_{n+1}K_n}{1-\beta\tilde{p}_{n+1}K_n^2}, \qquad n=0, \cdots, N. \quad (5.4)$$

Theorem 5.1: Let (5.2) and (5.4) admit a solution sequence $\{\hat{L}_n\}_{n=0}^{N-1}$, $\{\hat{K}_n\}_{n=0}^{N-1}$, satisfying (5.3), where $\tilde{p}_n$ depends on K_n through (5.1) and $\sigma_{n+1|n}$ depends on L_n as in Lemma 4.1. Then:

i) $\hat{L}_n = L_n^*$, $n = 0, \cdots, N-1$, where the latter was defined in Proposition 4.1;

ii) the linear control law (4.11) is optimal also in the class of nonlinear policies.

Proof: The proof follows from the construction of nested games prior to the statement of the theorem, and the two facts that: i) the only saddle-point solution of the game with kernel G_{N-1} is linear with the structure (4.11); and ii) the solution to the original stochastic control problem exists in the class of linear policies (cf. Proposition 4.2). □

Remark 5.1: It can be seen through some routine manipulations that when $N = 1$ and $n = 0$, the condition (5.3), and hence that of Theorem 5.1, is equivalent to that of Theorem 3.1. □

B. The Infinite Horizon Case

We treat the infinite horizon case as the limit of the finite horizon problem as $N \to \infty$, provided that the discounted payoff (2.1) remains bounded and the optimum policy sequence (4.11) converges to a well-defined limit. Note that since the optimal policy (cf. Theorem 5.1 and Proposition 4.2) is linear, the stationary limiting policy will be given by

$$\gamma_n^*(x_n, y^{n-1})=L^*(x_n-E[x_n|I_n])$$

where L^* is the stationary solution of the optimal control problem (4.4), (4.5) as $N \to \infty$. For each N, denoting the solution given in Proposition 4.1 iii) by L_k^{N*}, $k < N$, we would expect $L^* = \lim_{N\to\infty} L_k^{N*}$, for every finite k.

To study the existence of such a limit, we first recall that the value function $W_k(p)$ defined in (4.6) is strictly increasing for decreasing $k < N$, for every $p > 0$ (cf. Lemma 4.6), and further that it is bounded above by an affine function (cf. Lemma 4.3). This last property follows since in (4.10) both A_k and a_k are bounded in retrograde time. Hence, $\lim_{k\to\infty} W_{-k}(p) = W(p)$ where the limiting function satisfies

$$W(p)=\max_L\left\{\left(L-\frac{L^2}{2}\right)p+\beta W\left(\frac{\rho^2p}{L^2p+1}+r\right)\right\}. \quad (5.5)$$

Denoting the maximizing solution here by $L(p)$, we already know from Lemma 4.5 that $0 \le L(p) \le 1$ for all $p > 0$. In view of this, and the fact that $\rho^2 < 1$, (4.7) describes a stable system with L_k replaced by L, and hence $p_k \to p^*$, where p^* solves

$$p^*=\frac{\rho^2p^*}{L(p^*)p^*+1}+r. \quad (5.6a)$$

Let

$$L^* := L(p^*). \quad (5.6b)$$

Then, we have the following solution for the infinite horizon problem.

Theorem 5.2: With $N \to \infty$ in (2.1) and under the conditions of Theorem 5.1, the stochastic control problem of Section II admits the optimal stationary policies

$$\gamma_n^*(x_n, y^{n-1})=L^*(x_n-E[x_n|I_n])$$

for n sufficiently large, where L^* is defined by (5.6). For smaller values of n, the optimal policy is

$$\gamma_n^*(x_n, y^{n-1})=L(p_n)(x_n-E[x_n|I_n])$$

where p_n is obtained from

$$p_{k+1}=\frac{\rho^2p_k}{L(p_k)p_k+1}+r, \qquad p_0=\sigma_0/\sigma_w$$

and $L(\cdot)$ is a maximizing solution of (5.5). □

VI. Conclusion

One of the main messages of this paper has been that there do exist stochastic control problems of the nonneutral type which admit analytic solutions. However, even for the seemingly simple scalar problem of this paper, the derivation of optimal policies and proof of their existence is quite a nontrivial task, requiring an indirect approach. It would be interesting (and quite rewarding) to explore the possibility of devising a more direct approach towards the solution of this problem; although, by the experience of the author, this seems to be quite unlikely. We should also note that even though the solution has been obtained in closed form, it still involves (off-line) the solution of a nonlinear deterministic optimal control problem which, even numerically, is not easy.

The general approach of this paper, which relates the original single-person optimization problem to a sequence of nested zero-sum games, is original and seems to appear in the literature for the first time. One of its unique aspects is the demonstration of the utility of the powerful machinery of saddle-point equilibria even in problems which are neither formulated as, nor can directly be converted to, zero-sum games. Other applications of this approach could be seen, for example, in extending the results of this paper to more general models where (2.2) is replaced by higher order ARMA processes. Such an extension, although not immediate, seems to be possible.

Appendix

Here, we provide interpretations for the different terms used in the mathematical model of Section II, in the context of the rational expectations model alluded to in Section I. Further insight into the model can be gained from [3] and [5].

As mentioned in the Introduction, there are actually two players in this decision problem, the policy maker (say, the government) and the private (or the public) sector. The former is an active decision maker, who controls the rate of monetary growth, which we denote by u_i at time step i. This is, in fact, the desired (planned) rate, which will in general be different from the actual rate of monetary growth y_i because of independent shocks impending on the economy. Now, the other player is a passive one, who simply attempts to predict the future (one-step ahead) value of y_i, using the information available at time i, which we denote by I_i; hence, his input to the decision problem is $E[y_i|I_i]$. The difference between the actual and predicted values of y_i, $y_i - E[y_i|I_i]$ is called the monetary surprise which, according to common belief, has a positive impact on the economy, because it leads to stimulation. The benefit that accrues from stimulation will have to be traded off against the negative (inflationary) effect caused by large values of u_i. This tradeoff is captured in the preference parameter x_i, which is only known to the policy maker. Then, the problem is to find the best mix between stimulation and inflation, which is formulated as a dynamic optimization problem. The objective function adopted, J_N, reflects the tradeoff between benefit derived from stimulation $(x_i(y_i - E[y_i|I_i]))$ and loss incurred from increased inflation $((1/2)u_i^2)$. Note that since $E\{x_i(y_i - E[y_i|I_i])\} \equiv E\{x_i(u_i - E[u_i|I_i])\}$, this objective function is equivalent to the one given by (2.1).

REFERENCES

[1] Y. Bar-Shalom and E. Tse, "Concepts and methods in stochastic control," *Control Dynam. Syst.: Adv. Theory Appl.*, vol. 23, pp. 99-172, 1976.

[2] T. Başar and G. J. Olsder, *Dynamic Noncooperative Game Theory*. New York: Academic, 1982.

[3] T. Başar and M. Salmon, "Credibility and the value of information transmission in a model of monetary policy and inflation," submitted for publication, 1988.

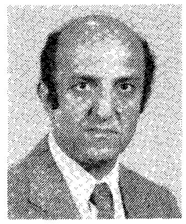
[4] D. Bertsekas, *Dynamic Programming and Stochastic Control*. New York: Academic, 1976.

[5] A. Cuckierman and A. Meltzer, "A theory of ambiguity, credibility, and inflation under discretion and asymmetric information," *Econometrica*, vol. 54, pp. 1009-1128, Sept. 1986.

[6] D. Kendrick, *Stochastic Control for Economic Models*. New York: McGraw-Hill, 1981.

[7] P. R. Kumar and P. Varaiya, *Stochastic Systems: Estimation, Identification and Adaptive Control*. Englewood Cliffs, NJ: Prentice-Hall, 1986.

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



Tamer Başar (S'71-M'73-SM'79-F'83) was born in Istanbul, Turkey. He received the B.S.E.E. degree from Robert College, Istanbul, and the M.S., M.Phil., and Ph.D. degrees from Yale University, New Haven, CT.

After being at Harvard University, Marmara Research Institute, and Boğaziçi University, he joined the University of Illinois at Urbana-Champaign in 1981, where he is currently a Professor of Electrical and Computer Engineering. He has spent two sabbatical years (1978-1979 and 1987-1988) at Twente University of Technology, Twente, The Netherlands and INRIA, Centre de Sophia Antipolis, France, respectively. He has authored or coauthored numerous papers in the general areas of optimal control, dynamic games, stochastic control, estimation theory, stochastic processes, information theory, and mathematical economics. He is the coauthor of the text *Dynamic Noncooperative Game Theory* (New York: Academic, 1982), Editor of the volume *Dynamic Games and Applications in Economics* (New York: Springer-Verlag, 1986), and Coeditor of the volume *Differential Games and Applications* (New York: Springer-Verlag, 1988).

Dr. Başar is a member of Sigma Xi. He has been active in the IEEE Control Systems Society in various capacities, and has served as Technical Committee Chairman, Associate Editor, and member of the BOG. He is currently an Associate Editor of three journals, and is the Program Chairman for the 1989 CDC.