SIAM J. CONTROL OPTIM.
Vol. 48, No. 3, pp. 1377–1404

OPTIMAL PERFORMANCE OF NETWORKED CONTROL SYSTEMS WITH NONCLASSICAL INFORMATION STRUCTURES*

ADITYA MAHAJAN† AND DEMOSTHENIS TENEKETZIS‡

Abstract. A discrete time stochastic feedback control system consisting of a nonlinear plant, a sensor, a controller, and a noisy communication channel between the sensor and the controller is considered. The sensor has limited memory and, at each time, it transmits an encoded symbol over the channel and updates its memory. The controller receives a noise-corrupted copy of the transmitted symbol. It generates a control action based on all its past observations and all its past actions. This control action is fed back to the plant. At each time instant the system incurs an instantaneous cost depending on the state of the plant and the control action. The objective is to choose encoding, memory update, and control strategies to minimize an expected total cost over a finite horizon, or an expected discounted cost over an infinite horizon, or an average cost per unit time over an infinite horizon. A solution methodology for obtaining a sequential decomposition of the global optimization problem is developed. This solution methodology is extended to the case when the sensor makes an imperfect observation of the state of the plant.

Key words. optimal control over noisy communication, sequential stochastic control, decentralized optimal control, nonclassical information structures, dynamic teams, common knowledge

AMS subject classifications. 93E03, 93E20, 93A14, 62B05, 49N30

DOI. 10.1137/060678130

1. Introduction.

1.1. Preliminaries and literature overview. Recent advances in network and communication technologies have led to an increasing interest in NCS (networked control systems), in particular in understanding the interaction between control and communication components of the system (see [1, 2, 9]). Many researchers have looked at traditional control systems with a communication component and tried to understand the limitations imposed by the communication component in the feedback loop.

Stability analysis of a plant with finite data rate feedback has been investigated under different modeling assumptions in [4, 5, 7, 10, 12, 15, 17, 23, 25, 26, 27, 28, 30, 45]. A unified overview of stabilization with finite data rate feedback was provided in [29]. Linear quadratic Gaussian (LQG) stability under different plant models (linear, deterministic, stochastic, stable, unstable) and different channel models (rate limited noiseless or noisy additive white Gaussian noise (AWGN)) was investigated in [6, 35, 36, 37, 38]. Mean square stability of a linear plant over noisy forward and reverse channels was considered in [46]. Asymptotic limitations of arbitrary time-invariant feedback for a linear time-invariant plant were investigated in [21, 22]. Most of the results for stability of NCS have found some kind of relationship between the unstable plant dynamics and some notion of capacity of the communication channel. Plant

*Received by the editors December 19, 2006; accepted for publication (in revised form) January 15, 2009; published electronically April 15, 2009. A preliminary version of this paper appeared in the Proceedings of the 45th IEEE Conference on Decision and Control, San Diego, December 2006. This research was supported in part by ONR grant N00014-03-1-0232, NSF grant CCR-0325571, and NASA grant NNX06AD47G.
http://www.siam.org/journals/sicon/48-3/67813.html

†Department of Electrical Engineering, Yale University, New Haven, CT 06520 (aditya.mahajan@yale.edu).
‡Department of Electrical Engineering and Computer Science, University of Michigan, Ann Arbor, MI 48109–2212 (teneket@eecs.umich.edu).

stability and channel capacity are both asymptotic concepts so, in retrospect, the connection between them is not surprising.

Performance analysis of NCS has received less attention that stability analysis in the literature. Optimal performance of a linear plant with a rate-limited noiseless communication channel was considered in [24, 33]: in [24] the plant disturbance is Gaussian and the controller is memoryless; in [33] the plant is undisturbed and the controller has perfect recall. Optimal performance of a linear plant with Gaussian disturbance, either a rate-limited noiseless channel or a Gaussian memoryless channel, and various information structures at the encoder was considered in [38]. Optimal performance of a nonlinear plant and a noisy channel with noiseless feedback from the output of the channel to the encoder was considered in [39].

The nature of results on performance analysis of NCS depends on the information structure of the problem, in particular, on whether or not the encoder/sensor knows the information at the receiver/controller. When the encoder knows the information available at the receiver, as is the case in the models of [24, 33, 39] and the instances in [38] with noiseless channel or information structure A (defined in [38, p. 1550]), optimal encoding and control strategies have been determined. When the encoder does not know the information at the receiver, as is the case in the model of [38] with information pattern B (defined in [38, p. 1550]), only suboptimal encoding and control strategies have been proposed.

The former case (when the encoder knows the receiver's information) has a partially nested information structure, while the latter case has a strictly nonclassical information structure. Problems with strictly nonclassical information structures are considerably harder to analyze than problems with partially nested information structures. This is reflected in the nature of results in performance analysis of NCS.

In this paper we consider a model of a simple NCS where the encoder does not know the receiver's information (and hence, the model has a strictly nonclassical information structure). We model the performance analysis of NCS as a stochastic control problem and determine a method to sequentially identify optimal encoding and control strategies. For the finite horizon problem, optimal encoding and control strategies are determined by the solution of nested optimality equations; for the infinite horizon problems, optimal strategies are determined by the fixed point of functional equations.

1.2. Features of the problem. We consider a discrete time feedback control system with a communication channel between the sensor and the controller, as shown in Figure 2.1. Such problems arise when the plant and the controller are geographically separated. We assume that there is a noisy discrete memoryless channel between the sensor and the controller. (The rate-limited communication channel is the degenerate case where the channel is noiseless.) We model problems in which the sensor has limited resources in terms of the power at which it can transmit and the data it can store and process. The encoder connected with the sensor is assumed to have a finite memory, and thus it cannot remember all its past observations and actions, and at each stage must selectively shed some information. At each time, the sensor generates a symbol, using its current observation and the contents of its memory, and transmits it over the noisy channel to the controller. We assume that there is no resource constraint at the controller. It has infinite memory and infinite power. Thus we assume that the controller has perfect recall—it remembers everything that it has seen and done in the past—and the communication channel between the controller and the plant is noiseless.[1] At each stage t the system incurs an instantaneous cost depending on the state of the plant at t and the control action at t. The objective is to choose globally optimal encoding, memory update, and control strategies to minimize the expected total cost over a finite horizon, or the expected discounted cost over an infinite horizon, or the expected average cost per unit time over an infinite horizon.

The problem has two decision-makers, the sensor and the controller. Due to the noise in the communication channel, the sensor and the controller observe different information about what is happening in the system. Due to the finite memory at the sensor, the sensor forgets information and at any given time instant the sensor may not know what actions it took in the past and why it took those actions. These two considerations, the noise in the channel and the finite memory at the sensor, result in a *decentralized* control problem. There is no known solution methodology for solving infinite horizon decentralized stochastic control problems.

Markov decision theory [16] provides a solution methodology for *centralized* stochastic control problems. For centralized problems with imperfect observations, Markov decision theory shows that there is no loss of optimality in taking a control action based on the controller's belief about the state of the plant, which is obtained using all the data available at the controller. Centralization of information and perfect recall at the controller are crucial for this idea to work. Consequently this idea does not extend to decentralized control problems: decentralization of information implies that one decision-maker cannot infer the data available with the other decision-makers and therefore cannot infer their beliefs. So if all decision-makers act according to their beliefs about the state of the plant, they will act in an inconsistent manner, and the system will not achieve globally optimal performance. Hence, Markov decision theory is not appropriate for this problem.

Orthogonal search techniques [32] provide a solution methodology for decentralized stochastic control problems. There are different variations of the orthogonal search algorithm, but the key idea is the following. Initialize by arbitrarily choosing the decision strategies of all agents; then pick an agent, say i, and determine the *best response* of agent i to the strategies of the other agents. Fix this best response as agent i's strategy. Next pick another agent j, $j \neq i$, and update agent j's strategy by its best response to the other agents' strategies. Continue in this way. If this procedure converges, the resultant strategies are member by member optimal; i.e., unilateral deviations by a single agent do not improve the system's performance. Fictitious play techniques [8, 31, 34] are philosophically similar to orthogonal search and result in member by member optimal solutions. Since decentralized stochastic control problems are, in general, nonconvex in strategy space, the above procedure may not converge to globally optimal strategies; that is, it does not guarantee that there do not exist any other tuple of strategies for all agents that outperform the member by member optimal strategies found by the above procedure. Thus, orthogonal search cannot be used to obtain globally optimal strategies for the problem under consideration.

Witsenhausen's standard form [42] is the only known solution methodology for general sequential decentralized stochastic control problems. It proceeds by first converting the problem into a *standard form* and then obtaining a sequential decomposition for the standard form. The standard form is a finite horizon stochastic control problem whose state evolution satisfies some properties, the cost is a stopping cost incurred at the last time step, and the cost has certain measurability properties. Since

[1] In what follows we show that assuming a noiseless feedback channel does not entail any loss of generality.

all the cost is incurred at the last time step in the standard form, infinite horizon problems cannot be converted into the standard form. Hence the solution methodology of [42] is not appropriate for the problem under consideration.

1.3. Contributions of this paper. The main contribution of this paper is identifying information states sufficient for performance analysis (or sufficient statistics for control) for a simple model of NCS with a nonclassical information structure. We show that an optimal choice of information states converts the design of NCS into a centralized problem and enables us to use analytic and computational tools from Markov decision theory.

In contrast to conditional probability measures, which are used as information states in centralized stochastic control problems, we use unconditional probability measures as information states. This is similar to the choice of information states in Witsenhausen's standard form [44]. However, in the standard form the domain of the information state increases with time, which restricts the results of the standard form to finite horizon problems. In contrast, the domain of the information states used in this paper does not increase with time, which allows us to extend our results to infinite horizon problems.

We would like to emphasize that identifying information states is nontrivial. Decentralized control problems have been investigated since the 1970s, but even now there is no general method for obtaining information states that work for both finite and infinite horizon problems. This paper identifies information states for a particular decentralized control problem and explains why these information states work. This explanation may provide some insights for choosing information states for other classes of decentralized stochastic control problems.

1.4. Organization of the paper. The remainder of this paper is organized as follows. We formulate the performance analysis of feedback control systems with limited communication over a noisy channel as a decentralized stochastic optimization problem. To illustrate the key concepts associated with our solution methodology we first consider the finite horizon problem. In section 2, we establish structural results of an optimal controller and obtain a methodology for sequential global optimization of the encoding, memory update, and control strategies for the finite horizon problem. We provide an explanation of the methodology in section 3. In section 4 we extend the methodology to infinite horizon problems. In section 5 we consider the case of uncountable state space. In section 6 we consider the feedback control problem when the encoder has imperfect observation of the state of the plant, and we extend the results of sections 2 and 4 to this problem. We conclude in section 7.

1.5. Notation. Throughout this paper we use the following notation. Uppercase letters (X, Y, Z) denote random variables, lowercase letters (x, y, z) denote their realizations, and calligraphic letters $(\mathcal{X}, \mathcal{Y}, \mathcal{Z})$ denote their alphabets. For random variables and functions, x^t is shorthand for $x_1, \dots, x_t$. $\mathbb{E}\{\,\cdot\,\}$ denotes the expectation of a random variable, $\Pr(\cdot)$ denotes the probability of an event, and $\mathbb{1}[\cdot]$ denotes the indicator function of a statement. To denote the expectation or probability of a random variable or an event that depends on a function φ, we use $\mathbb{E}\{\,\cdot \mid \varphi\,\}$ and $\Pr(\cdot \mid \varphi)$, respectively. We have chosen this slightly unusual notation because we want to keep track of all the functional dependencies and because the conventional notation of $\mathbb{E}^{\varphi}\{\,\cdot\,\}$ and $\Pr^{\varphi}(\cdot)$ is too cumbersome.

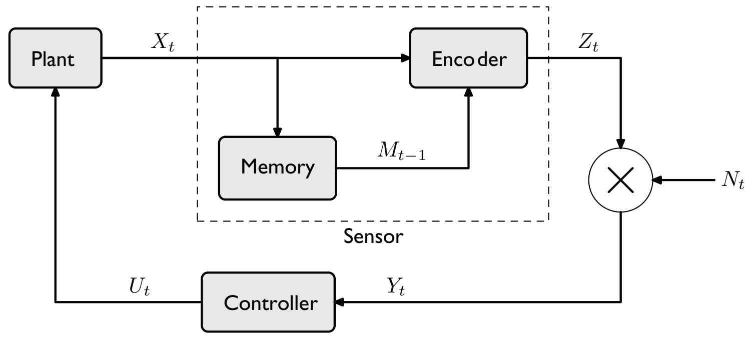

FIG. 2.1. *Feedback control system with noisy communication.*

2. The finite horizon problem.

2.1. Problem formulation. Consider the discrete time feedback control system of Figure 2.1 which operates for a horizon T. The state evolution is given by

$$X_{t+1} = f(X_t, U_t, W_t), \tag{2.1}$$

where f is the *plant evolution function* and the variables X_t, U_t, W_t denote the state of the plant, the control action, and the plant disturbance respectively, at time t. We assume that all variables are finite valued. For all t, X_t takes values in a finite set $\mathcal{X}$, U_t takes values in a finite set $\mathcal{U}$, and W_t takes values in a finite set $\mathcal{W}$. The initial state X_1 is a random variable with the probability mass function (PMF) P_{X_1}. The random variables $W_1, \dots, W_T$ are i.i.d. (independent and identically distributed) with PMF P_W and are also independent of X_1.

The sensor, consisting of an encoder and a memory, makes perfect observations of the state of the plant. At each time instant t the encoder generates an encoded symbol Z_t, taking values in a finite set $\mathcal{Z}$, as follows:

$$Z_t = c_t(X_t, M_{t-1}), \tag{2.2}$$

where c_t is the *encoding function* at time t and M_{t-1} denotes the content of the sensor's memory at $t-1$. M_t takes values in a finite set $\mathcal{M}$ and is updated according to

$$M_t = l_t(X_t, M_{t-1}), \tag{2.3}$$

where l_t is the *memory update function* at time t. Observe that the sensor has a finite size memory, and, although it makes perfect observations of the state of the plant, it cannot store all the past observations. Thus, it does not have perfect recall, and at each stage it must selectively shed information.

The encoded symbol Z_t is transmitted over a noisy communication channel, and a channel output Y_t is generated according to

$$Y_t = h(Z_t, N_t), \tag{2.4}$$

where h is the *channel function* and N_t denotes the channel noise. Y_t takes values in a finite set $\mathcal{Y}$ and N_t takes values in a finite set $\mathcal{N}$. The sequence of random variables $N_1, \dots, N_T$ is i.i.d. with given PMF P_N and is also independent of $X_1, W_1, \dots, W_T$.

The controller observes the channel outputs and generates a control action U_t as follows:

$$U_t = g_t(Y^t, U^{t-1}), \tag{2.5}$$

where g_t is the *control law* at time t. U_t takes values in a finite set $\mathcal{U}$. A uniformly bounded cost function $\rho : \mathcal{X} \times \mathcal{U} \to [0, K]$, where $K < \infty$, is given. At each t, an instantaneous cost $\rho(X_t, U_t)$ is incurred.

The collection $(\mathcal{X}, \mathcal{W}, \mathcal{M}, \mathcal{Z}, \mathcal{N}, \mathcal{Y}, \mathcal{U}, P_{X_1}, P_W, P_N, f, h, \rho, T)$ is called a *perfect observation system*. The choice of (C, L, G), $C := (c_1, \dots, c_T)$, $L := (l_1, \dots, l_T)$, $G := (g_1, \dots, g_T)$, is called a *design*.

The performance of a design is quantified by the expected total cost under that design and is given by

$$\mathcal{J}_T(C, L, G) := \mathbb{E}\left\{ \sum_{t=1}^{T} \rho(X_t, U_t) \,\middle|\, C, L, G \right\}, \tag{2.6}$$

where the expectation in (2.6) is with respect to a joint measure on $(X_1, \dots, X_T, U_1, \dots, U_T)$ generated by P_W, P_N, f, h and the choice of design (C, L, G). We are interested in the following optimization problem.

PROBLEM 2.1. *Given a perfect observation system $(\mathcal{X}, \mathcal{W}, \mathcal{M}, \mathcal{Z}, \mathcal{N}, \mathcal{Y}, \mathcal{U}, P_{X_1}, P_W, P_N, f, h, \rho, T)$, choose a design (C^*, L^*, G^*) such that*

$$\mathcal{J}_T(C^*, L^*, G^*) = \mathcal{J}_T^* := \min_{C, L, G \in \mathscr{C}^T \times \mathscr{L}^T \times \mathscr{G}^T} \mathcal{J}_T(C, L, G), \tag{2.7}$$

where $\mathscr{C}^T := \mathscr{C} \times \cdots \times \mathscr{C}$ (T times), $\mathscr{C}$ is the space of functions from $\mathcal{X} \times \mathcal{M}$ to $\mathcal{Z}$, $\mathscr{L}^T := \mathscr{L} \times \cdots \times \mathscr{L}$ (T times), $\mathscr{L}$ is the space of functions from $\mathcal{X} \times \mathcal{M}$ to $\mathcal{M}$, $\mathscr{G}^T := \mathscr{G}_1 \times \cdots \times \mathscr{G}_T$, and $\mathscr{G}_t$ is the space of functions from $\mathcal{Y}^t \times \mathcal{U}^{t-1}$ to $\mathcal{U}$.

Remarks.

1. There is no loss of generality in assuming a noiseless channel between the controller and the plant. Suppose that the channel between the controller and the plant is noisy. Let the input $\hat{U}_t$ to the plant be a noise-corrupted version of U_t given by

   $$\hat{U}_t = \hat{h}(U_t, \hat{N}_t), \tag{2.8}$$

   where $\hat{h}$ is the feedback channel and $\hat{N}_t$ denotes the noise in the feedback channel. $\hat{N}_1, \dots, \hat{N}_T$ is a sequence of independent variables that is also independent of $X_1, W_1, \dots, W_T$ and $N_1, \dots, N_T$.[2] Then this model can be transformed into one equivalent to (2.1)–(2.5) by setting

   $$\hat{W}_t = (W_t, \hat{N}_t), \tag{2.9}$$

   $$X_{t+1} = f\left(X_t, \hat{h}(U_t, \hat{N}_t), W_t\right) := \hat{f}(X_t, U_t, \hat{W}_t). \tag{2.10}$$

   Thus, without any loss of generality we can assume a noiseless feedback channel.

[2] We only require $\hat{W}_1, \dots, \hat{W}_T$, where $\hat{W}_t = (W_t, \hat{N}_t)$, to be an independent process. So, $\hat{N}_t$ need not be independent of W_t.

2. A globally optimal design for Problem 2.1 always exists because there are finitely many designs, and we can always choose one with the best performance.

2.2. Salient features of the problem. Problem 2.1 is a decentralized multi-agent stochastic optimization problem. There are two agents, the sensor and the controller, having different information about the system but a common objective of minimizing an expected total cost over a finite horizon. Multi-agent problems in which the agents have a common objective are called team problems [20]. Team problems are further classified as static teams or dynamic teams on the basis of their information structure. See [41] for a definition of information structure (also called information pattern). In static teams the actions taken by one agent do not affect the information structure of the other agents; in dynamic teams they do. In Problem 2.1, the actions taken by the sensor affect the observations of the controller and the actions taken by the controller affect the observations of the sensor; furthermore, the sensor and the controller have different information about the system. Moreover, due to the finite memory at the sensor and the noise in the channel, Problem 2.1 has a strictly nonclassical information structure; thus Problem 2.1 is a dynamic team.

Determining globally optimal strategies for dynamic teams is difficult because they are, in general, nonconvex functional optimization problems having a complex interdependence among their decision rules [14]. As pointed out in the introduction, Markov decision theory, orthogonal search, and standard form are not appropriate for solving infinite horizon dynamic team problems.

The solution concept that we are looking for is to decompose the global optimization problem into a sequence of nested optimization subproblems, where each subproblem is easier to solve than the original problem. This is called *sequential decomposition*, and it exponentially reduces the search complexity of finding an optimal strategy. A crucial step in obtaining a sequential decomposition of the global optimization problem is to identify information states that are sufficient for performance evaluation. Properties that such states must satisfy are explained in [19, 18]. All the known techniques of identifying appropriate information states, viz., Markov decision theory, orthogonal search, and standard form, are not appropriate for infinite horizon dynamic team problems. The information states in Markov decision theory—the conditional probability densities of the state given all the past observations and all the past control actions—works only when there is a single controller with perfect recall, so they are inappropriate for dynamic teams. The information states in orthogonal search are obtained under the assumption that the strategies of other agents are fixed. These information states determine only member by member optimal strategies, so they are not appropriate for determining globally optimal strategies for dynamic teams. The information states in Witsenhausen's standard form belong to a space that increases with time; hence, it is not appropriate for infinite-horizon problems. Thus a new methodology for identifying information states is needed for the problem under consideration. We provide one such methodology in this paper.

The sequential order in which the system variables are generated is the key to understanding the solution methodology that we present in this paper. For this purpose we need to refine the notion of time. We call each step of the system a *stage*. At any stage t, we consider three time instants[3] t^+, $(t+1/2)$, and $(t+1)^-$. For ease of notation, we will denote these time instants by $\underline{t}$, t, and $\overline{t}$, respectively. From now on,

[3] The actual values of these time instants are not important; we just need three values in increasing order.

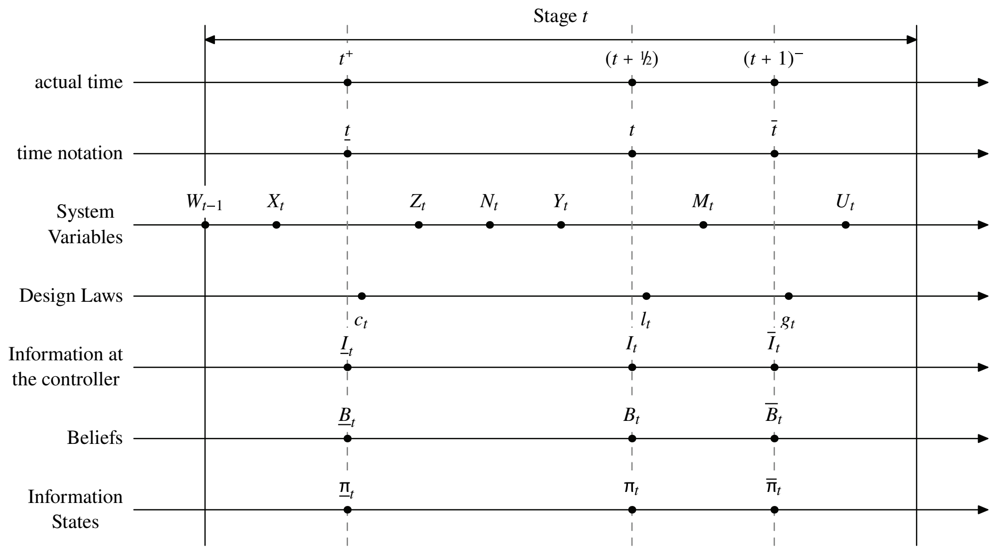

FIG. 2.2. *Problem* 2.1 *as a sequential stochastic optimization problem. This figure shows the ordering relation between the system variables, design rules, and information states.*

we will assume that the system has three agents—the encoder, the memory update, and the controller—even though the encoder and the memory update are located in the same device and have the same information. We assume that the sensor encodes just after $\underline{t}$, the sensor's memory is updated just after t, and the controller takes a control action just after $\overline{t}$. The order in which the variables are generated in the system is shown in Figure 2.2. Since the ordering of the decision-makers can be done independently of the realization of the system variables, the problem is a *sequential* stochastic optimization problem [43].

To obtain a sequential decomposition of Problem 2.1, we proceed in two steps. In step one, we derive structural properties of optimal controllers. In step two, we use the structural results of step one to identify an information state sufficient for performance evaluation, transform Problem 2.1 into an equivalent deterministic optimization problem, and obtain a sequential decomposition for this equivalent problem. This sequential decomposition gives an algorithm for obtaining an optimal design for Problem 2.1.

As pointed out in the introduction, step two is the crucial step. The key difficulty in step two is to identify an information state appropriate for performance evaluation. Even when the structural results of step one are available, identifying such an information state is a highly nontrivial task. Once an appropriate information state is identified, the transformation to a deterministic problem and the sequential decomposition follow.

2.3. Structure of optimal controllers. In this section we present structural properties of optimal controllers. We first define random variables that capture the information available just before the decision rules c_t, l_t, and g_t act on the system.

DEFINITION 2.1. *Let* $\underline{I}_t$, I_t, *and* $\overline{I}_t$ *denote the information available at the controller at time* $\underline{t}$, t, *and* $\overline{t}$, *respectively. Specifically*

1. $\underline{I}_t := (Y^{t-1}, U^{t-1}, c^{t-1}, l^{t-1}, g^{t-1})$.
2. $I_t := (Y^t, U^{t-1}, c^t, l^{t-1}, g^{t-1})$.
3. $\overline{I}_t := (Y^t, U^{t-1}, c^t, l^t, g^{t-1})$.

We have included the past decision rules in the definition of information because the distribution of the random variables depends on the choice of the past decision rules. Observe that

$$\underline{I}_t = (\overline{I}_{t-1}, U_{t-1}, g_{t-1}), \quad I_t = (\underline{I}_t, Y_t, c_t), \quad \text{and} \quad \overline{I}_t = (I_t, l_t). \tag{2.11}$$

Next we define the belief of the controller about the state of the plant and the memory contents of the sensor at times $\underline{t}^-$, t^-, and $\overline{t}^-$.

DEFINITION 2.2. *Let* $\underline{B}_t$, B_t, *and* $\overline{B}_t$ *be random vectors defined as follows:*

1. $\underline{B}_t(x, m) \coloneqq \Pr(X_t = x, M_{t-1} = m \,|\, \underline{I}_t)$.
2. $B_t(x, m) \coloneqq \Pr(X_t = x, M_{t-1} = m \,|\, I_t)$.
3. $\overline{B}_t(x, m) \coloneqq \Pr(X_t = x, M_t = m \,|\, \overline{I}_t)$.

For any particular realization $\underline{i}_t$ of $\underline{I}_t$, that is, for any particular realization y^{t-1}, u^{t-1} of Y^{t-1}, U^{t-1} and arbitrary (but fixed) choice of c^{t-1}, l^{t-1}, and g^{t-1}, the realization $\underline{b}_t$ of $\underline{B}_t$ is a PMF on $\mathcal{X} \times \mathcal{M}$. If $\underline{I}_t$ is a random vector, then $\underline{B}_t$ is a random vector belonging to $\mathcal{P}^{\mathcal{X}\times\mathcal{M}}$, the space of PMFs on $\mathcal{X} \times \mathcal{M}$. Similar interpretations hold for B_t and $\overline{B}_t$.

The random vectors $\underline{B}_t$, B_t, and $\overline{B}_t$ represent the belief of the controller about the state of the plant and the encoder's memory content at $\underline{t}$, t, and $\overline{t}$, respectively. The sequential ordering of these beliefs with respect to the other variables in the system is shown in Figure 2.2. The time evolution of these beliefs is coupled as follows.

LEMMA 2.3. *For each stage* t, *there exist deterministic functions* $\underline{F}$, F, *and* $\overline{F}$ *such that*

1. $\underline{B}_t = \underline{F}(\overline{B}_{t-1}, U_{t-1})$.
2. $B_t = F(\underline{B}_t, Y_t, c_t)$.
3. $\overline{B}_t = \overline{F}(B_t, l_t)$.

Proof.

1. Consider a component of $\underline{b}_t$,

$$\begin{aligned} \underline{b}_t(x_t, m_{t-1}) &= \Pr(X_t = x_t, M_{t-1} = m_{t-1} \,|\, \underline{i}_t) \\ &= \Pr(X_t = x_t, M_{t-1} = m_{t-1} \,|\, \overline{i}_{t-1}, u_{t-1}, g_{t-1}) \end{aligned} \tag{2.12}$$
$$= \frac{\Pr(X_t = x_t, M_{t-1} = m_{t-1}, U_{t-1} = u_{t-1} \,|\, \overline{i}_{t-1}, g_{t-1})}{\sum_{(x'_t, m'_{t-1}) \in \mathcal{X}\times\mathcal{M}} \Pr\left(X_t = x'_t, M_{t-1} = m'_{t-1}, U_{t-1} = u_{t-1} \,\middle|\, \overline{i}_{t-1}, g_{t-1}\right)}.$$

Now consider

$$\begin{aligned} &\Pr(X_t = x_t, M_{t-1} = m_{t-1}, U_{t-1} = u_{t-1} \,|\, \overline{i}_{t-1}, g_{t-1}) \\ &\quad = \Pr(x_t, m_{t-1}, u_{t-1} \,|\, \overline{i}_{t-1}, g_{t-1}) \\ &\quad = \sum_{x_{t-1}\in\mathcal{X}} \Pr(x_{t-1}, m_{t-1} \,|\, \overline{i}_{t-1}, g_{t-1}) \\ &\qquad\quad \times \Pr(u_{t-1} \,|\, x_{t-1}, m_{t-1}, \overline{i}_{t-1}, g_{t-1}) \\ &\qquad\quad \times \Pr(x_t \,|\, x_{t-1}, m_{t-1}, u_{t-1}, \overline{i}_{t-1}, g_{t-1}) \\ &\quad \overset{(a)}{=} \sum_{x_{t-1}\in\mathcal{X}} \Pr(x_{t-1}, m_{t-1} \,|\, \overline{i}_{t-1}) \mathbb{1}\left[u_{t-1} = g_{t-1}(y^{t-1}, u^{t-2})\right] \\ &\qquad\quad \times \Pr(x_t \,|\, x_{t-1}, u_{t-1}) \end{aligned} \tag{2.13}$$
$$\begin{aligned} &\quad = \mathbb{1}\left[u_{t-1} = g_{t-1}(y^{t-1}, u^{t-2})\right] \\ &\qquad\quad \times \sum_{x_{t-1}\in\mathcal{X}} \overline{b}_{t-1}(x_{t-1}, m_{t-1}) \Pr(x_t \,|\, x_{t-1}, u_{t-1}), \end{aligned} \tag{2.14}$$

where equality (a) follows from (2.1) and (2.5) and $\mathbb{1}[\cdot]$ is the indicator function. Substitute (2.14) into (2.12) and cancel $\mathbb{1}[u_{t-1} = g_{t-1}(y^{t-1}, u^{t-2})]$ from the numerator and the denominator, giving

$$\underline{b}_t(x_t, m_{t-1}) = \frac{\sum_{x_{t-1}\in\mathcal{X}} \overline{b}_{t-1}(x_{t-1}, m_{t-1}) \Pr(x_t \mid x_{t-1}, m_{t-1})}{\sum_{(x'_t, x'_{t-1}, m'_{t-1})\in\mathcal{X}\times\mathcal{X}\times\mathcal{M}} \overline{b}_{t-1}(x'_{t-1}, m'_{t-1}) \Pr\left(x'_t \,\middle|\, x'_{t-1}, m'_{t-1}\right)}. \tag{2.15}$$

Hence,

$$\underline{b}_t = \underline{F}(\overline{b}_{t-1}, u_{t-1}), \tag{2.16}$$

where $\underline{F}$ is determined by (2.15).

2. Consider a component of b_t,

$$\begin{aligned} b_t(x_t, m_{t-1}) &= \Pr(X_t = x_t, M_{t-1} = m_{t-1} \mid i_t) \\ &= \Pr(X_t = x_t, M_{t-1} = m_{t-1} \mid \underline{i}_t, y_t, c_t) \\ &= \frac{\Pr(X_t = x_t, M_{t-1} = m_{t-1}, Y_t = y_t \mid \underline{i}_t, c_t)}{\sum_{(x'_t, m'_{t-1})\in\mathcal{X}\times\mathcal{M}} \Pr\left(X_t = x'_t, M_{t-1} = m'_{t-1}, Y_t = y_t \,\middle|\, \underline{i}_t, c_t\right)}. \end{aligned} \tag{2.17}$$

Now consider

$$\begin{aligned} &\Pr(X_t = x_t, M_{t-1} = m_{t-1}, Y_t = y_t \mid \underline{i}_t, c_t) \\ &= \Pr(x_t, m_{t-1}, y_t \mid \underline{i}_t, c_t) \\ &= \Pr(x_t, m_{t-1} \mid \underline{i}_t, c_t) \Pr(y_t \mid x_t, m_{t-1}, \underline{i}_t, c_t) \\ &\overset{(b)}{=} \Pr(x_t, m_{t-1} \mid \underline{i}_t) \Pr(y_t \mid x_t, m_{t-1}, c_t) \\ &= \underline{b}_t(x_t, m_{t-1}) \Pr(y_t \mid x_t, m_{t-1}, c_t), \end{aligned} \tag{2.18}$$

where equality (b) follows from (2.1)–(2.4). Combining (2.17) and (2.18), we have

$$b_t(x_t, m_{t-1}) = \frac{\underline{b}_t(x_t, m_{t-1}) \Pr(y_t \mid x_t, m_{t-1}, c_t)}{\sum_{(x'_t, m'_{t-1})\in\mathcal{X}\times\mathcal{M}} \underline{b}_t(x'_t, m'_{t-1}) \Pr\left(y_t \,\middle|\, x'_t, m'_{t-1}, c_t\right)}. \tag{2.19}$$

Hence,

$$b_t = F(\underline{b}_t, y_t, c_t), \tag{2.20}$$

where F is given by (2.19).

3. Consider a component of $\overline{b}_t$,

$$\begin{aligned}\overline{b}_t(x_t, m_t) &= \Pr(x_t, m_t \mid \overline{i}_t) = \Pr(x_t, m_t \mid i_t, l_t) \\ &= \sum_{m_{t-1}\in\mathcal{M}} \Pr(x_t, m_t, m_{t-1} \mid i_t, l_t) \\ &= \sum_{m_{t-1}\in\mathcal{M}} \Pr(x_t, m_{t-1} \mid i_t, l_t) \Pr(m_t \mid x_t, m_{t-1}, i_t, l_t) \\ &\overset{(c)}{=} \sum_{m_{t-1}\in\mathcal{M}} \Pr(x_t, m_{t-1} \mid i_t) \Pr(m_t \mid x_t, m_{t-1}, l_t) \\ &= \sum_{m_{t-1}\in\mathcal{M}} b_t(x_t, m_{t-1}) \mathbb{1}[m_t = l_t(x_t, m_{t-1})],\end{aligned} \tag{2.21}$$

where equality (c) follows from (2.1) and (2.3), and $\mathbb{1}[\cdot]$ is the indicator function. Hence,

$$\overline{b}_t = \overline{F}(b_t, l_t), \tag{2.22}$$

where $\overline{F}$ is given by (2.21). ∎

The above relationships between the controller's beliefs lead to the structural results of the optimal controllers.

THEOREM 2.4. *Consider Problem* 2.1 *for any arbitrary (but fixed) encoding and memory update strategies* $C := (c_1, \dots, c_T)$ *and* $L := (l_1, \dots, l_T)$*, respectively. Then, without loss of optimality, we can restrict our attention to control laws of the form*

$$U_t = g_t(\overline{B}_t). \tag{2.23}$$

Proof. The plant dynamics (2.1) and sensor dynamics (2.3) imply that for any fixed encoding and memory update strategies the process $\{(X_t, M_{t-1}),\ t = 1, \dots, T\}$ is a controlled Markov process with control actions U_t. The observations Y_t of the controller can be written as

$$Y_t = h_t(c_t(X_t, M_{t-1}), N_t) =: \hat{h}_t((X_t, M_{t-1}), N_t).$$

So, the controller partially observes the state (X_t, M_{t-1}). The instantaneous cost is also a function of the state (X_t, M_{t-1}) and the control action U_t (this function does not depend on M_{t-1}). Moreover, the controller has perfect recall. Thus, for any fixed encoding and memory update strategies, the design of an optimal controlled is a partially observed centralized stochastic control problem. From Markov decision theory [16] we know that there is no loss of optimality in restricting our attention to control laws of the form (2.23). ∎

2.3.1. Implication of the structural results. Theorem 2.4 implies that at each stage t, without loss of optimality, we can restrict our attention to controllers belonging to the family $\hat{\mathscr{G}}$ of functions from $\mathcal{P}^{\mathcal{X}\times\mathcal{M}}$ to $\mathcal{U}$. With this modification Problem 2.1 is equivalent to the following problem.

PROBLEM 2.2. *Given a perfect observation system* $(\mathcal{X}, \mathcal{W}, \mathcal{M}, \mathcal{Z}, \mathcal{N}, \mathcal{Y}, \mathcal{U}, P_{X_1}, P_W, P_N, f, h, \rho, T)$*, choose a design* (C^*, L^*, G^*) *that is optimal with respect to the performance criterion of* (2.6)*, i.e.,*

$$\mathcal{J}_T(C^*, L^*, G^*) = \mathcal{J}_T^* := \min_{C,L,G\in\mathscr{C}^T\times\mathscr{L}^T\times\hat{\mathscr{G}}^T} \mathcal{J}_T(C, L, G), \tag{2.24}$$

where $\hat{\mathscr{G}}^T := \hat{\mathscr{G}} \times \cdots \times \hat{\mathscr{G}}$ *(T times).*

Using the structural results results of Theorem 2.4, we can transform Problem 2.1 into an equivalent problem, Problem 2.2, in which the domain of all the decision rules—the encoding rules, the memory update rules, and the control rules—is not changing with time. This is in contrast to Problem 2.1, where the domain of the control rules was increasing with time. This reduction to a time-invariant domain is necessary for extending the solution methodology for the finite horizon problems to infinite horizon.

In the next section we provide a sequential decomposition of Problem 2.2.

2.4. Global optimization. As explained in section 2.2, Problems 2.1 and 2.2 are dynamic teams with a strictly nonclassical information structure. To obtain a sequential decomposition we need to identify information states sufficient for performance evaluation, or equivalently, find sufficient statistics for performance evaluation. The sequential nature of the problem suggests choosing an information state for each decision rule. Suppose $\underline{\pi}_t$, π_t, and $\overline{\pi}_t$ are information states at time t for the encoder, memory update, and the controller, respectively. Due to the decentralization of information, these information states should depend only on the decision rules (which are common knowledge) and not on the observation of any agent. For $\underline{\pi}_t$, π_t, and $\overline{\pi}_t$ to be information states in the sense of [16], at each instant of time, π_t must be determined from $\underline{\pi}_t$ and c_t; $\overline{\pi}_t$ must be determined from π_t and l_t; and $\underline{\pi}_{t+1}$ must be determined from $\overline{\pi}_t$ and g_t. However, a system can have more than one information state, and not all of them are sufficient for performance evaluation (see [44]). To be sufficient for performance evaluation, the information states must *absorb/summarize* the effect of past decision rules on the expected future cost;[4] that is, they should satisfy

$$\begin{aligned} \mathbb{E}\left\{ \sum_{s=t}^{T} \rho(X_s, U_s) \;\middle|\; C, L, G \right\} &= \mathbb{E}\left\{ \sum_{s=t}^{T} \rho(X_s, U_s) \;\middle|\; \underline{\pi}_t, c_t^T, l_t^T, g_t^T \right\} \\ &= \mathbb{E}\left\{ \sum_{s=1}^{T} \rho(X_s, U_s) \;\middle|\; \pi_t, c_{t+1}^T, l_t^T, g_t^T \right\} \\ &= \mathbb{E}\left\{ \sum_{s=t}^{T} \rho(X_s, U_s) \;\middle|\; \overline{\pi}_t, c_{t+1}^T, l_{t+1}^T, g_t^T \right\}, \end{aligned} \tag{2.25}$$

or equivalently,

$$\mathbb{E}\left\{ \rho(X_t, U_t) \mid C, L, G \right\} = \mathbb{E}\left\{ \rho(X_t, U_t) \mid \overline{\pi}_t, g_t \right\}. \tag{2.26}$$

These properties, which must be satisfied by information states that are sufficient for performance evaluation, are explained in more detail in [19].

For sequential problems, one way to obtain information states satisfying the above properties is by converting the model to Witsenhausen's standard form [42]. However, in the standard form the space in which information states belong increases with time, so such a transformation to the standard form does not lead to a formulation that can

[4]Note that in problems with classical information structure, we can find an information state that is independent of the control law [16]. For problems with strictly nonclassical information structures it is not always possible to find information states that are independent of the control law. However, as long as the expected future cost conditioned on the information state is conditionally independent of the past control laws, a sequential decomposition can be obtained using that information state. See [42] for a proof.

be extended to infinite horizon problems. We want information states that will be appropriate for both finite and infinite horizon problems. This is possible only when the space in which the information states belong is time-invariant.

Thus information states sufficient for performance evaluation should satisfy the following properties:

(P1) *They must be states*, that is, at each instant of time, π_t should be a function of $\underline{\pi}_t$ and c_t; $\overline{\pi}_t$ should be a function of π_t and l_t; and $\underline{\pi}_{t+1}$ should be a function of $\overline{\pi}_t$ and g_t.

(P2) *They must be sufficient for performance evaluation*, that is, they should satisfy (2.25) or (2.26).

(P3) They should take values in a time-invariant space.

Next we present information states that have the above properties and show how these information states lead to a sequential decomposition of Problem 2.2. *We want to re-emphasize that the hardest part in our solution methodology is to identify the appropriate information states; there are no known solution methodologies for identifying information states for decentralized stochastic control problems like Problem* 2.2.

The information states defined below have all the above-mentioned desired features.

DEFINITION 2.5. *Let* Π *be the space of probability measure on* $\mathcal{X} \times \mathcal{M} \times \mathcal{P}^{\mathcal{X}\times\mathcal{M}}$, *and let* $\mathscr{B}(\cdot)$ *denote the Borel* σ*-algebra. For any* $x \in \mathcal{X}$, $m \in \mathcal{M}$, *and* $A_{\underline{B}}$, A_B, $A_{\overline{B}} \in \mathscr{B}(\mathcal{P}^{\mathcal{X}\times\mathcal{M}})$, *define* $\underline{\pi}_t$, π_t, $\overline{\pi}_t$, $t = 1, \dots, T$, *as follows.*

1. $\underline{\pi}_t(x, m, A_{\underline{B}}) := \Pr\left(X_t = x, M_{t-1} = m, \underline{B}_t \in A_{\underline{B}} \,\middle|\, c^{t-1}, l^{t-1}, g^{t-1}\right)$.
2. $\pi_t(x, m, A_B) := \Pr\left(X_t = x, M_{t-1} = m, B_t \in A_B \,\middle|\, c^t, l^{t-1}, g^{t-1}\right)$.
3. $\overline{\pi}_t(x, m, A_{\overline{B}}) := \Pr\left(X_t = x, M_t = m, \overline{B}_t \in A_{\overline{B}} \,\middle|\, c^t, l^t, g^{t-1}\right)$.

Here $\underline{\pi}_t$, π_t, and $\overline{\pi}_t$ are probability measures (or probability laws) on the probability space $\left(\mathcal{X} \times \mathcal{M} \times \mathcal{P}^{\mathcal{X}\times\mathcal{M}},\ 2^{\mathcal{X}\times\mathcal{M}}\mathscr{B}(\mathcal{P}^{\mathcal{X}\times\mathcal{M}})\right)$, where $\mathscr{B}(\mathcal{P}^{\mathcal{X}\times\mathcal{M}})$ is the Borel σ-algebra on $\mathcal{P}^{\mathcal{X}\times\mathcal{M}}$. These probability measures are information states sufficient for the performance evaluation of Problem 2.2. Specifically, they satisfy the following properties.

LEMMA 2.6. $\underline{\pi}_t, \pi_t, \overline{\pi}_t$ *are information states for the encoder, the memory update, and the controller respectively, i.e.,*

1. *there is a linear transformation* $\underline{Q}(c_t)$ *such that*

$$\pi_t = \underline{Q}(c_t)\underline{\pi}_t. \tag{2.27}$$

2. *there is a linear transformation* $Q(l_t)$ *such that*

$$\overline{\pi}_t = Q(l_t)\pi_t. \tag{2.28}$$

3. *there is a linear transformation* $\overline{Q}(g_t)$ *such that*

$$\underline{\pi}_{t+1} = \overline{Q}(g_t)\overline{\pi}_t. \tag{2.29}$$

4. *the conditional expected instantaneous cost can be expressed as*

$$\mathbb{E}\rho(X_t, U_t)|c^t, l^t, g^t = \tilde{\rho}(\overline{\pi}_t, g_t), \tag{2.30}$$

where $\tilde{\rho}$ *is a deterministic function.*

Proof.

1. Consider a component of π_t,

$$\begin{aligned}\pi_t(x,m,db) &= \sum_{y\in\mathcal{Y}}\int_{\underline{A}(b,y,c_t)} \underline{\pi}_t(x,m,\underline{b})P_N\big(n\in\mathcal{N} : y = h(c_t(x,m),n)\big)d\underline{b} \\ &=: \underline{Q}_t(c_t)\underline{\pi}_t,\end{aligned} \tag{2.31}$$

where $\underline{A}(b,y,c) = \{\underline{b}\in\mathcal{P}^{\mathcal{X}\times\mathcal{M}} : b = F(\underline{b},y,c)\}$.

2. Consider a component of $\overline{\pi}_t$,

$$\begin{aligned}\overline{\pi}_t(x,m,d\overline{b}) &= \sum_{\{m'\in\mathcal{M}:m'=l_t(x,m)\}}\int_{A(\overline{b},l_t)} \pi_t(x,m',b)db \\ &=: Q(l_t)\pi_t,\end{aligned} \tag{2.32}$$

where $A(\overline{b},l) = \{b\in\mathcal{P}^{\mathcal{X}\times\mathcal{M}} : \overline{b} = \overline{F}(b,l)\}$.

3. Consider a component of $\underline{\pi}_{t+1}$,

$$\begin{aligned}\underline{\pi}_{t+1}(x,m,d\underline{b}) &= \sum_{x_t\in\mathcal{X}}\int_{\overline{A}(\underline{b},g_t)} \overline{\pi}_t(x_t,m,\overline{b}) \\ &\qquad\times P_W\big(w\in\mathcal{W} : x = f(x_t,g_t(\overline{b}_t,w)\big) \\ &=: \overline{Q}(g_t)\overline{\pi}_t,\end{aligned} \tag{2.33}$$

where $\overline{A}(\underline{b},g) = \{\overline{b}\in\mathcal{P}^{\mathcal{X}\times\mathcal{M}} : \underline{b} = \underline{F}\big(\overline{b},g(\overline{b})\big)\}$.

4. Consider $\mathbb{E}\{\,\rho(X_t,U_t)\mid c^t,l^t,g^t\,\}$. By the problem formulation, $\underline{\pi}_1$ is known to all agents. For specified c^t,l^t, and g^{t-1}, the information state $\overline{\pi}_t$ can be evaluated using the transformations of previous steps of this lemma. Thus,

$$\begin{aligned}\mathbb{E}\{\,\rho(X_t,U_t)\mid c^t,l^t,g^t\,\} &= \mathbb{E}\{\,\rho(X_t,U_t)\mid c^t,l^t,g^t,\overline{\pi}_t\,\} \\ &= \sum_{x_t\in\mathcal{X}}\int_{\mathcal{P}^{\mathcal{X}\times\mathcal{M}}} \overline{\pi}_t(x_t,\overline{b}_t)\rho\big(x_t,g_t(\overline{b}_t)\big)d\overline{b}_t := \tilde{\rho}(\overline{\pi}_t,g_t),\end{aligned} \tag{2.34}$$

where $\overline{\pi}_t(x_t,\overline{b}_t)$ is the marginal of $\overline{\pi}_t(x_t,m_t,\overline{b}_t)$. ∎

Points 1, 2, and 3 of Lemma 2.6 shows that the information states $\underline{\pi}_t$, π_t, and $\overline{\pi}_t$ satisfy property (P1); point 4 shows that these information states satisfy property (P2). Property (P3) is satisfied by definition. Thus, $\underline{\pi}_t$, π_t, and $\overline{\pi}_t$ are information states sufficient for performance evaluation. In order to obtain a sequential decomposition, first reconsider the performance criterion of (2.6), which can be rewritten as

$$\mathbb{E}\left\{\sum_{t=1}^{T}\rho(X_t,U_t)\,\middle|\,C,L,G\right\} = \sum_{t=1}^{T}\mathbb{E}\{\,\rho(X_t,U_t)\mid c^t,l^t,g^t\,\} =: \sum_{t=1}^{T}\tilde{\rho}(\overline{\pi}_t,g_t), \tag{2.35}$$

where the sequence $\{\overline{\pi}_1,\dots,\overline{\pi}_T\}$ depends on the choice of (C,L,G). Hence, Problem 2.2 is equivalent to the following deterministic problem.

PROBLEM 2.3. *Consider a deterministic system with states $\underline{\pi}_t,\pi_t,\overline{\pi}_t$. The initial state $\underline{\pi}_1$ is known, and, for $t\geq 1$, the system evolves as follows.*

$$\pi_t = \underline{Q}(c_t)\underline{\pi}_t, \qquad \overline{\pi}_t = Q(l_t)\pi_t, \qquad \text{and} \qquad \underline{\pi}_{t+1} = \overline{Q}(g_t)\overline{\pi}_t, \tag{2.36}$$

where c_t, l_t, g_t *belong to* $\mathscr{C}, \mathscr{L}, \hat{\mathscr{G}}$*, respectively, and* $\underline{Q}$*,* Q*,* $\overline{Q}$ *are known linear transformations given by Lemma* 2.6*. At time* t*, an instantaneous cost* $\tilde{\rho}(\overline{\pi}_t, g_t)$ *is incurred.*

The optimization problem is to determine design (C, L, G)*, where* $C := (c_1, \ldots, c_T)$*,* $L := (l_1, \ldots, l_T)$*, and* $G := (g_1, \ldots, g_T)$*, to minimize the total cost over horizon* T*, i.e.,*

$$\min_{(C,L,G)\in \mathscr{C}^T \times \mathscr{L}^T \times \hat{\mathscr{G}}^T} \sum_{t=1}^{T} \tilde{\rho}(\overline{\pi}_t, g_t). \tag{2.37}$$

This is a classical deterministic optimal control problem in function space; optimal functions (C^*, L^*, G^*) can be determined as follows.

THEOREM 2.7. *An optimal design* (C^*, L^*, G^*) *for Problem* 2.3 *(and consequently for Problem* 2.2 *and thereby for Problem* 2.1*) is given the following nested optimality equations:*

$$\overline{V}_T(\overline{\pi}) = \inf_{g_T \in \hat{\mathscr{G}}} \tilde{\rho}(\overline{\pi}, g_T), \tag{2.38}$$

and for $t = 1, \ldots, T$*,*

$$\underline{V}_t(\underline{\pi}) = \min_{c_t \in \mathscr{C}} V_t\big(\underline{Q}(c_t)\underline{\pi}\big), \tag{2.39}$$

$$V_t(\pi) = \min_{l_t \in \mathscr{L}} \overline{V}_t\big(Q(l_t)\pi\big), \tag{2.40}$$

$$\overline{V}_t(\overline{\pi}) = \inf_{g_t \in \hat{\mathscr{G}}} \big\{\tilde{\rho}(\overline{\pi}, g_t) + \underline{V}_{t+1}\big(\overline{Q}(g_t)\overline{\pi}\big)\big\}. \tag{2.41}$$

The arg min *(or* arg inf*) at each step determines the corresponding optimal design for that stage. Furthermore, the optimal performance is given by*

$$\mathcal{J}_T^* = \underline{V}_1(\underline{\pi}_1). \tag{2.42}$$

Proof. This is a standard result; see [16, Chapter 2]. ∎

2.5. Discussion of Problem 2.3. We present an alternative look at Problem 2.3 which will be useful when we study the infinite-horizon version of Problem 2.1. As pointed out in section 2.4, Problem 2.3 is a deterministic control problem with state space Π and action space alternating between $\mathscr{C}$, $\mathscr{L}$, and $\hat{\mathscr{G}}$. We now introduce a sequence of *metafunctions* $\underline{\Delta}_t$, Δ_t, and $\overline{\Delta}_t$, $t = 1, \ldots, T$, where $\underline{\Delta}_t$ is a function from Π to $\mathscr{C}$, Δ_t is a function from Π to $\mathscr{L}$, and $\overline{\Delta}_t$ is a function from Π to $\hat{\mathscr{G}}$. These metafunctions describe the rationale used to select the "action" (i.e., the design c^t, l^t, g^t) at time t. The choice of all metafunctions for horizon T is called *metadesign*. Problem 2.3 is equivalent to the following feedback control problem.

PROBLEM 2.4. *Consider a deterministic system with states* $\underline{\pi}_t$*,* π_t*,* $\overline{\pi}_t \in \Pi$*, and "control actions"* $c_t \in \mathscr{C}$*,* $l_t \in \mathscr{L}$*, and* $g_t \in \hat{\mathscr{G}}$*. The initial state* $\underline{\pi}_1$ *is known, and, for* $t \geq 1$*, the system evolves as follows:*

$$\pi_t = \underline{Q}(c_t)\underline{\pi}_t, \qquad \overline{\pi}_t = Q(l_t)\pi_t, \qquad \textit{and} \qquad \underline{\pi}_{t+1} = \overline{Q}(g_t)\overline{\pi}_t, \tag{2.43}$$

where $\underline{Q}$*,* Q*, and* $\overline{Q}$ *are known transformations given by Lemma* 2.6*. The "control actions"* c_t*,* l_t*, and* g_t *are chosen according to the metafunctions* $\underline{\Delta}_t$*,* Δ_t*, and* $\overline{\Delta}_t$ *as follows:*

$$c_t = \underline{\Delta}_t(\underline{\pi}_t), \qquad l_t = \Delta_t(\pi_t), \qquad \textit{and} \qquad g_t = \overline{\Delta}_t(\overline{\pi}_t). \tag{2.44}$$

At each time, an instantaneous cost $\tilde{\rho}(\overline{\pi}_t, g_t)$ *is incurred. The optimization problem is to determine the metadesign* $\tilde{\Delta}^T := (\underline{\Delta}_1,\, \Delta_1,\, \overline{\Delta}_1,\, \dots, \underline{\Delta}_T,\, \Delta_T,\, \overline{\Delta}_T)$ *to minimize the total cost over horizon* T*, i.e.,*

$$\min \sum_{t=1}^{T} \tilde{\rho}(\underline{\pi}_t, g_t) \tag{2.45}$$

where the minimization is over the choice of $\tilde{\Delta}^T$*.*

The nested optimality equations of Theorem 2.7 determine the globally optimal metafunctions $\underline{\Delta}_t$, Δ_t, $\overline{\Delta}_t$ for $t = 1, \dots, T$, i.e., the optimal feedback laws for Problem 2.4. Since Problem 2.4 is a deterministic control problem with a known initial state, we need only specify the control "actions" c_t, l_t, g_t for $t = 1, \dots, T$. This is why we have considered Problem 2.3 instead of Problem 2.4. Nevertheless, Problem 2.4 will be useful in clarifying the nature of the solution of the infinite horizon problem corresponding to Problem 2.1.

3. Explanation of the solution methodology. The sequential decomposition obtained above can be interpreted as follows. Suppose that before the system is started the sensor and the controller get together to determine an optimal design that they will use. Instead of testing the performance of each design one by one, they decide to choose the designs sequentially. So, they need to agree on a mechanism (or an algorithm) that will, at each time instant and for any choice of past design rules,[5] determine the future design rules optimally. To do so, for any choice of past design rules, the sensor and the controller must be able to consistently evaluate the optimal future performance. To be consistent in their evaluation, each agent must "know" what the other agent is "thinking." Suppose the design rules until time t, denoted by $\gamma^t := (c_1, c_2, \dots, c_t, l_1, l_2, \dots, l_{t-1}, g_1, g_2, \dots, g_{t-1})$, have been agreed upon (by some mechanism) and the sensor and the controller want to determine the next design rule l_t. If they allow the system to run until time t, the sensor will know the values of X_t and M_{t-1}, while the controller will know the values of Y^t and U^{t-1}. However, they do not know the other agent's observations. They can form a belief on the other agent's observations, but then they do not know the other agent's belief on their observations. If they form a belief on the other agent's belief on their observation, they will not know the former belief on the latter belief. This process of forming a belief on what the other agent is "thinking" will continue until the sensor and the controller agree upon what they are thinking.

In [3] Aumann showed that such an agreement will occur in the "common knowledge" between the two agents. Formally, suppose $(\Omega, \mathcal{F}, P)$ is the probability space of the primitive random variables of the system. For any fixed γ^t, (X_t, M_{t-1}) and (Y^t, U^{t-1}) are random vectors on $(\mathcal{X} \times \mathcal{M}, 2^{\mathcal{X}\times\mathcal{M}})$ and $(\mathcal{Y}^t \times \mathcal{U}^{t-1}, 2^{\mathcal{Y}\times\mathcal{U}^{t-1}})$, respectively. Let $\sigma(X_t, M_{t-1})$ and $\sigma(Y^t, U^{t-1})$ denote the smallest subfields of $\mathcal{F}$ with respect to which (X_t, M_{t-1}) and (Y^t, M^{t-1}) are, respectively, measurable. Then the common knowledge between (X_t, M_{t-1}) and (Y^t, U^{t-1}) is $\sigma(X_t, M_{t-1}) \cap \sigma(Y^t, U^{t-1}) =: K_t(\gamma^t)$. Thus, to do a sequential decomposition, the agents should decide what to do for all $K_t(\gamma^t)$ obtained by varying γ^t over all possible values. However, it is difficult to identify the space of all possible realizations of $K_t(\gamma^t)$. So instead of using $K_t(\gamma^t)$ as an (information) state, the agents can use $\sigma(X_t, M_{t-1}, B_t) =: \hat{K}_t(\gamma^t)$, which is a

[5] In this description, we use *design rule* to refer to either the encoding rule, the memory update, rule, or the control law.

superfield of $K_t(\gamma_t)$ (see Appendix A for proof). $\hat{K}_t(\gamma^t)$ also depends on γ^t, and it is difficult to evaluate the space of realization of $\hat{K}_t(\gamma^t)$ obtained by varying γ^t over all possible values. However, if we go to the image space of the random variables, we can obtain an "overapproximation" of $\hat{K}_t(\gamma^t)$. Consider the image space of the random vectors $(X_t, M_{t-1}, \underline{B}_t) : (\Omega, \mathcal{F}, P) \to \big(\mathcal{X} \times \mathcal{M} \times \mathscr{B}(\mathbb{R}^2), \mathscr{B}(2^{\mathcal{X}\times\mathcal{M}} \times \mathscr{B}(\mathbb{R}^2)), \hat{P}_t(\gamma^t)\big) =: \Lambda_t(\gamma^t)$, where $\mathscr{B}(\cdot)$ denotes the Borel set. In $\Lambda_t(\gamma^t)$ only the measure $\hat{P}_t(\gamma^t)$ depends on the choice of past design rules. Although it is difficult to evaluate all reachable realizations of $\hat{P}(\gamma^t)$ obtained by varying γ^t over all possible values, the space of all realizations of $\hat{P}_t(\gamma^t)$ is known and is equal to all probability measures on $\big(\mathcal{X} \times \mathcal{M} \times \mathscr{B}(\mathbb{R}^2), \mathscr{B}(2^{\mathcal{X}\times\mathcal{M}} \times \mathscr{B}(\mathbb{R}^2))\big)$. So the sensor and the controller can decide on what action to take for each probability measure $\hat{P}_t$, that is, for any probability space $\Lambda_t := \big(\mathcal{X} \times \mathcal{M} \times \mathscr{B}(\mathbb{R}^2), \mathscr{B}(2^{\mathcal{X}\times\mathcal{M}} \times \mathscr{B}(\mathbb{R}^2)), \hat{P}_t\big)$, and not worry whether the space is reachable or not. Notice that the information state π_t is equivalent to Λ_t defined here. In the definition of π_t the sample space and the σ-algebra are implicitly specified. Similar interpretations hold for $\underline{\pi}_t$ and $\overline{\pi}_t$.

If the rules for breaking ties are made common knowledge, the nested optimality equations of Theorem 2.7 allow the sensor and the controller (or anyone who knows the model and rules for breaking ties, henceforth referred to as the *designer*) to sequentially and consistently determine optimal design rules in two stages. In the first stage, for each time instant and for each realization of the information state determine an optimal design rule to be used if that information state is actually realized. In the second stage, sequentially determine, for every t, the optimal design rules c_t, l_t, g_t, to be implemented as follows. For the first time instant using the information state $\underline{\pi}_1$, which is part of the model, the sensor and the controller (and the designer) can determine an optimal c_1^*. This choice of c_1^* is common knowledge between the sensor and the controller since the model and the rule for breaking ties are common knowledge. For these values of $\underline{\pi}_1$ and c_1^*, Lemma 2.6 gives the value of the realization of π_1. This value is common knowledge between the sensor and the controller (and the designer). Now, using the result of the first stage, an optimal l_1^* can be determined, which in turn gives the realization of $\overline{\pi}_1$. This realization is common knowledge between the sensor and controller (and the designer). This processes can be continued until all the design rules c_1^*, l_1^*, g_1^*,...,c_T^*, l_T^*, g_T^* are determined. This design is optimal and common knowledge between the sensor and the controller (and the designer).

In view of the discussion in section 2.5 the first stage corresponds to determining optimal metafunctions $\underline{\Delta}_t$, Δ_t, and $\overline{\Delta}_t$, $t = 1, \dots, T$, while the second stage corresponds to determining optimal design rules c_t, l_t, g_t, $t = 1, \dots, T$, that are implemented. These design rules correspond to the control actions in Problem 2.4; since the problem is deterministic, they can be specified before the system starts running.

The nested optimality equations of Theorem 2.7 are functional optimization problems: for each realization of the information state we need to determine an optimal design rule (a function) to be used if that state is actually realized. Contrast this with the centralized stochastic optimization problems where the dynamic programming equations result in parameter optimization problems: for each realization of the information state we need to determine an optimal control action (a parameter) to be taken if that state is actually realized. Functional optimization problems are an order of magnitude harder to solve than parameter optimization problems. The cardinality of the function space (e.g., $\mathscr{C}$) increases exponentially with a linear increase in the cardinality of the "action" space ($\mathcal{Z}$). Moreover the function space (e.g., $\hat{\mathscr{G}}$) can be uncountable even when the action space ($\mathcal{U}$) is finite because the size of the function space is determined by both its domain and range. This increase in complexity makes decentralized stochastic control problems harder to solve than centralized stochastic optimization problems.

We believe that this is a fundamental feature of decentralized optimization problems and not something specific to our solution. Since an agent does not know another agent's observations, in order to consistently interpret the other agent's action, it should know the design rule of the other agent. So, in any sequential decomposition, at each instant of time an agent needs to determine its design rule and not just its control action. So any sequential decomposition will result in functional optimization problems.

Our solution is "simpler" than the only other known methodology for sequentially solving dynamic teams—Witsenhausen's standard form [42]. In [42] Witsenhausen showed how to convert *any* sequential optimization problem into "standard form" and showed how to obtain a sequential decomposition for the standard form. Similar to our result, the information states in the standard form are unconditional probability measures which evolve in a linear manner. If Problem 2.1 is converted into standard form, the information state at time t will be $\sigma(X_t, M_{t-1}, Y^t, U^{t-1}) =: \tilde{K}_t(\gamma^t)$. Observe that $\hat{K}_t(\gamma^t) \subset \tilde{K}_t(\gamma^t)$. So, our information state is a subfield of the information state in the standard form, and a sufficient statistic for the decomposition presented in standard form. However, the image space $\Lambda_t(\gamma^t)$ in our decomposition is bigger than the corresponding image space in the standard form. But the image space in the standard form increases with time, so the standard form cannot be used to solve the infinite horizon problem. The image space in our decomposition does not change with time, which enables us to tackle infinite horizon problems, as shown in the next section.

4. The infinite horizon problem. In this section we extend the model of section 2.1 to an infinite horizon ($T \to \infty$) using two performance criteria: the expected discounted cost and the average cost per unit time. Let (C, L, G), $C \coloneqq (c_1, c_2, \dots)$, $L \coloneqq (l_1, l_2, \dots)$, $G \coloneqq (g_1, g_2, \dots)$ denote an infinite horizon policy. The two performance criteria that we consider are as follows:

1. The expected discounted cost, where the performance of a design is determined by

$$\mathcal{J}^\beta(C, L, G) = \mathbb{E}\left\{ \sum_{t=1}^{\infty} \beta^{t-1} \rho(X_t, U_t) \,\middle|\, C, L, G \right\}, \tag{4.1}$$

where $0 < \beta < 1$ is called the discount factor.

2. The average cost per unit time, where the performance of a design is determined by

$$\overline{\mathcal{J}}(C, L, G) = \limsup_{T \to \infty} \frac{1}{T} \mathbb{E}\left\{ \sum_{t=1}^{T} \rho(X_t, U_t) \,\middle|\, C, L, G \right\}. \tag{4.2}$$

We take the $\limsup$, rather than $\lim$, as for some designs (C, L, G) the limit may not exist.

Ideally, while implementing a solution for infinite horizon problems, we would like to use time-invariant designs. This motivates the following definitions.

Definition 4.1. *A design (C, L, G), $C \coloneqq (c_1, c_2, \dots)$, $L \coloneqq (l_1, l_2, \dots)$, $G \coloneqq (g_1, g_2, \dots)$ is called stationary (or time-invariant) if $c_1 = c_2 = \cdots = c$, $l_1 = l_2 = \cdots = l$, $g_1 = g_2 = \cdots = g$.*

Definition 4.2. *Let* $\tilde{\Delta}_t := (\underline{\Delta}_t, \Delta_t, \overline{\Delta}_t)$*. A metadesign* $\tilde{\Delta}^\infty := (\tilde{\Delta}_1, \tilde{\Delta}_2, \dots)$ *is called stationary (or time-invariant) if* $\tilde{\Delta}_1 = \tilde{\Delta}_2 = \cdots = \tilde{\Delta}$.

In centralized stochastic control problems with time-homogenous evolution and time-homogenous cost function, one can restrict attention to stationary designs without any loss of optimality. This greatly simplifies the search for an optimal design. It is natural to wonder if such a result also holds for dynamic teams. It is not known whether, in general, restricting attention to stationary designs is optimal or not in dynamic teams. In this section we show that for the problem under consideration, stationary designs may not be optimal. However, there exist stationary metadesigns that are optimal: for the expected discounted cost problem one can restrict attention to stationary metadesigns without any loss of optimality; for the average cost per unit time problem, under a technical condition, one can restrict attention to stationary metadesigns. The optimal design corresponding to an optimal stationary metadesign is, in general, time-varying.

4.1. The expected discounted cost problem. Consider the infinite horizon problem with expected discounted cost criterion given by (4.1). For this problem the relations of Lemma 2.3 hold, hence the structural result of Theorem 2.4 is valid, and we can restrict our attention to controllers belonging to $\hat{\mathscr{G}}$. Consider $\underline{\pi}_t, \pi_t, \overline{\pi}_t$ as in Definition 2.5. Lemma 2.6 can be proved as before. The transformations $\underline{Q}, Q, \overline{Q}$ and the expected instantaneous cost $\tilde{\rho}$ are the same as in the finite horizon case. Let $\gamma_t := (c_t, l_t, g_t)$ denote the design at time t, and let Γ denote the function space $\mathscr{C} \times \mathscr{L} \times \hat{\mathscr{G}}$. We can combine (2.43) and (2.44) as

$$\underline{\pi}_{t+1} = \tilde{Q}(\gamma_t)\underline{\pi}_t, \qquad \gamma_t = \tilde{\Delta}_t(\underline{\pi}_t), \tag{4.3}$$

where $\tilde{Q}(\gamma_t) := \overline{Q}(g_t) \circ Q(l_t) \circ \underline{Q}(c_t)$ and $\tilde{\Delta}_t(\underline{\pi}_t) = (\underline{\Delta}(\underline{\pi}_t), \Delta_t(\pi_t), \overline{\Delta}_t(\overline{\pi}_t))$. The instantaneous cost at time t can be rewritten as

$$\bar{\rho}(\underline{\pi}_t, \gamma_t) := \hat{\rho}\big(\big(Q(l_t) \circ \underline{Q}(c_t)\big)\underline{\pi}_t, g_t\big). \tag{4.4}$$

Hence, the infinite horizon problem with the expected discounted cost criterion given by (4.1) is equivalent to the following deterministic optimization problem.

Problem 4.1. *Consider a deterministic system with state space* Π *and action space* Γ*. The system dynamics are given by*

$$\underline{\pi}_{t+1} = \tilde{Q}(\gamma_t)\underline{\pi}_t, \qquad \gamma_t = \tilde{\Delta}_t(\underline{\pi}_t), \tag{4.5}$$

where $\tilde{Q}$ *is a known transformation and* $\tilde{\Delta} : \Pi \to \Gamma$ *for all* t*. At each instant of time an instantaneous cost* $\bar{\rho}(\underline{\pi}_t, \gamma_t)$ *is incurred. The objective is to choose metadesign* $\tilde{\Delta}^\infty := (\tilde{\Delta}_1, \tilde{\Delta}_2, \dots)$ *so as to minimize the infinite horizon cost given by*

$$\mathcal{J}^\beta(\tilde{\Delta}^\infty) := \sum_{t=1}^{\infty} \beta^{t-1} \bar{\rho}(\underline{\pi}_t, \gamma_t). \tag{4.6}$$

Problem 4.1 is a standard infinite horizon discounted cost feedback control problem. Since we have assumed that $0 \le \rho < K$, where $K < \infty$, which in turn implies $0 \le \bar{\rho} < K$, an optimal metadesign is guaranteed to exist, and we have the following result.

Theorem 4.3. *For Problem* 4.1*, and consequently for the infinite horizon expected discounted cost problem with the performance criterion given by* (4.1)*, one can* *restrict attention to stationary metadesigns without any loss of optimality. Specifically there exists a stationary metadesign* $\tilde{\Delta}^{*,\infty} := (\tilde{\Delta}^*, \tilde{\Delta}^*, \dots)$, *and a corresponding infinite horizon design* (C^*, L^*, G^*), $C^* := (c_1^*, c_2^*, \dots)$, $L := (l_1^*, l_2^*, \dots)$, $G := (g_1, g_2, \dots)$ *such that*

$$\mathcal{J}^\beta(\tilde{\Delta}^{*,\infty}) = V(\underline{\pi}_1), \tag{4.7}$$

where V *is the unique uniformly bounded fixed point of*

$$V(\underline{\pi}) = \min_{\gamma \in \Gamma} \big\{ \bar{\rho}(\underline{\pi}, \gamma) + \beta V\big(\tilde{Q}(\gamma)(\underline{\pi})\big) \big\}, \tag{4.8}$$

and $\tilde{\Delta}^*$ *satisfies*

$$V(\underline{\pi}) = \bar{\rho}\big(\underline{\pi}, \tilde{\Delta}^*(\underline{\pi})\big) + \beta V\big(\tilde{Q}\big(\tilde{\Delta}^*(\underline{\pi})\big)(\underline{\pi})\big). \tag{4.9}$$

An optimal design (c_t^*, l_t^*, g_t^*) *to be implemented at time* t *is given by*

$$(c_t^*, l_t^*, g_t^*) =: \gamma_t^* = \tilde{\Delta}^*(\underline{\pi}_t). \tag{4.10}$$

Proof. This is a standard result; see [11, Chapter 6]. ∎

4.2. The average cost per unit time problem. Consider the infinite horizon problem with average cost per unit time criterion given by (4.2). Using the argument of the first paragraph of section 4.1, this problem is equivalent to the following deterministic problem.

PROBLEM 4.2. *Consider a deterministic system with state space* Π *and action space* Γ. *The system dynamics are given by*

$$\underline{\pi}_{t+1} = \tilde{Q}(\gamma_t)\underline{\pi}_t, \qquad \gamma_t = \tilde{\Delta}_t(\underline{\pi}_t), \tag{4.11}$$

where $\tilde{Q}$ *is a known transformation and* $\tilde{\Delta}_t : \Pi \to \Gamma$ *for all* t. *At each instant of time an instantaneous cost* $\bar{\rho}(\underline{\pi}_t, \gamma_t)$ *is incurred. The objective is to choose metadesign* $\tilde{\Delta}^\infty := (\tilde{\Delta}_1, \tilde{\Delta}_2, \dots)$ *so as to minimize the infinite horizon cost given by*

$$\overline{\mathcal{J}}(\tilde{\Delta}^\infty) := \limsup_{T \to \infty} \frac{1}{T} \sum_{t=1}^{T} \bar{\rho}(\underline{\pi}_t, \gamma_t). \tag{4.12}$$

For this problem an optimal metadesign may not exist. However, under suitable conditions, we can guarantee the existence of ε-optimal metadesigns. Specifically, we have the following result.

THEOREM 4.4. *For Problem* 4.2, *and consequently for the infinite horizon average cost per unit time problem with the performance criterion given by* (4.2), *assume the following:*

(A1) *For any* $\varepsilon > 0$ *there exist bounded measurable functions* $v(\cdot)$ *and* $r(\cdot)$ *and meta-function* $\tilde{\Delta}^* : \Pi \to \Gamma$ *such that for all* $\underline{\pi}$,

$$v(\underline{\pi}) = \min_{\gamma \in \Gamma} v\left(\tilde{Q}(\gamma)\underline{\pi}\right) = v\left(\tilde{Q}\big(\tilde{\Delta}^*(\underline{\pi})\big)\underline{\pi}\right) \tag{4.13}$$

and

$$\min_{\gamma \in \Gamma} \left\{ \bar{\rho}(\underline{\pi}, \gamma) + r\big(\tilde{Q}(\gamma)\underline{\pi}\big) \right\} \le v(\underline{\pi}) + r(\underline{\pi}) \le \bar{\rho}\big(\underline{\pi}, \tilde{\Delta}^*(\underline{\pi})\big) + r\big(\tilde{Q}\big(\tilde{\Delta}^*(\underline{\pi})\big)\underline{\pi}\big) + \varepsilon. \tag{4.14}$$

Then for any horizon T *and any metadesign* $\tilde{\Delta}^T \coloneqq (\tilde{\Delta}_1, \dots, \tilde{\Delta}_T)$*, the stationary metadesign* $\tilde{\Delta}^{*,T} \coloneqq (\tilde{\Delta}^*, \dots, \tilde{\Delta}^*)$ *(*T*-times) satisfies*

$$\mathcal{J}_T(\tilde{\Delta}^{*,T}) = r(\underline{\pi}_1) + Tv(\underline{\pi}_1) \leq \mathcal{J}_T(\tilde{\Delta}^T) + T\varepsilon. \tag{4.15}$$

Further, the stationary metadesign $\tilde{\Delta}^{*,\infty} \coloneqq (\tilde{\Delta}^*, \tilde{\Delta}^*, \dots)$ *is* ε*-optimal. That is, for any infinite horizon metadesign* $\tilde{\Delta}^\infty \coloneqq (\tilde{\Delta}_1, \tilde{\Delta}_2, \dots)$ *we have*

$$\overline{\mathcal{J}}(\tilde{\Delta}^{*,\infty}) = v(\underline{\pi}_1) \leq \underline{\mathcal{J}}(\tilde{\Delta}^\infty) + \varepsilon, \tag{4.16}$$

where

$$\overline{\mathcal{J}}(\tilde{\Delta}^{*,\infty}) \coloneqq \limsup_{T\to\infty} \frac{1}{T} \sum_{t=1}^{T} \bar{\rho}\big(\underline{\pi}_t, \tilde{\Delta}^*(\underline{\pi}_t)\big) \tag{4.17}$$

with $\underline{\pi}_{t+1} = \tilde{Q}\big(\tilde{\Delta}^*(\underline{\pi}_t)\underline{\pi}_t\big)$ *and*

$$\underline{\mathcal{J}}(\tilde{\Delta}^\infty) \coloneqq \liminf_{T\to\infty} \frac{1}{T} \sum_{t=1}^{T} \bar{\rho}\big(\underline{\pi}_t, \tilde{\Delta}_t(\underline{\pi}_t)\big) \tag{4.18}$$

with $\underline{\pi}_{t+1} = \tilde{Q}\big(\tilde{\Delta}_t(\underline{\pi}_t)\underline{\pi}_t\big)$.

Proof. This is a standard result; see [11, Chapter 7]. □

4.3. Discussion of the results. The results of this section show that for infinite horizon problems stationary designs are not necessarily optimal (or ε-optimal). In view of the discussion in section 2.5, this result is not surprising. The design rules c_t, l_t, g_t of the problems under consideration correspond to the control actions and the metafunctions correspond to the control law in classical deterministic optimization problems. In classical infinite horizon deterministic optimization problems, restricting attention to stationary *control laws* does not entail any loss of optimality; however, even for a stationary control law, control actions change with time. By analogy, in the infinite horizon problems considered in this section, restricting attention to stationary metadesigns does not entail any loss of optimality; however, even for a stationary metadesign, optimal design rules change with time. In the absence of a systematic framework, the task of finding and implementing an optimal infinite horizon design is intractable. Conveniently, the methodology and results presented in this section suggest a method for obtaining and implementing time-varying optimal designs, i.e., obtaining and implementing optimal stationary metadesigns. The offline problem simplifies to obtaining the fixed point of a functional equation, which also gives an optimal stationary metadesign. This metadesign can be implemented at the sensor and the controller. When the system is running, the sensor and the controller need to keep track of the information state of the system and to use the metadesign and the current information state to determine the current optimal design rules. This greatly simplifies the online implementation of a time-varying optimal design.

4.4. Some additional remarks.

1. In Theorem 4.3 the fixed point equation (4.8) can be simplified as

$$\underline{V}(\underline{\pi}) = \min_{c\in\mathscr{C}} V'\big(\underline{Q}(c)\underline{\pi}\big), \tag{4.19}$$

$$V'(\pi) = \min_{l\in\mathscr{L}} \overline{V}\big(Q(l)\pi\big), \tag{4.20}$$

$$\overline{V}(\overline{\pi}) = \inf_{g\in\hat{\mathscr{G}}} \big\{\tilde{\rho}(\overline{\pi}, g) + \underline{V}_{t+1}\big(\overline{Q}(g)\overline{\pi}\big)\big\}, \tag{4.21}$$

with $\underline{V}$ being equivalent to V of (4.8). Here we are further decomposing the problem into its "natural" sequential form. This system of equations (4.19)–(4.21) is the infinite horizon analogue of the optimality equations (2.39)–(2.41) of Theorem 2.7. The system (4.19)–(4.21) may be easier to solve (numerically) than (4.8).

2. If the ergodicity conditions of [13, section 3.3] are satisfied, then assumption (A1) of Theorem 4.4 is satisfied for all ε, and an optimal average cost per unit time exists.

5. Uncountable state space. Consider the model of section 2.1 with the following differences: the state of the plant X_t, the plant disturbance W_t, and the control action U_t belong to uncountable spaces, i.e., $\mathcal{X} = \mathbb{R}^{d_X}$, $\mathcal{W} = \mathbb{R}^{d_W}$, and $\mathcal{U} = \mathbb{R}^{d_U}$, where d_X, d_W, and d_U are positive integers. The initial state X_1 is a random variable belonging to $(\mathbb{R}^{d_X}, \mathscr{B}(\mathbb{R}^{d_X}), \mu_{X_1})$, where $\mathscr{B}(\mathbb{R}^{d_X})$ is the Borel σ-algebra on $\mathbb{R}^{d_X}$ and the probability law μ_{X_1} is given. The plant disturbances $W_1, \dots, W_T$ are i.i.d. random variables belonging to $(\mathbb{R}^{d_W}, \mathscr{B}(\mathbb{R}^{d_W}), \mu_W)$ where the probability law μ_W is given. The rest of the model is the same as that of section 2.1. The sensor has finite memory $\mathcal{M}$, and the channel is a discrete memoryless channel with input $\mathcal{Z}$ and output $\mathcal{Y}$. The plant function f, the design (C, L, G), $C \coloneqq (c_1, \dots, c_T)$, $L \coloneqq (l_1, \dots, l_T)$, $G \coloneqq (g_1, \dots, g_T)$ and the cost ρ are Borel measurable with respect to appropriate σ-algebras. The objective is to choose a design (C, L, G) that minimizes the total expected cost under that design.

The fact that the state of the plant, the plant disturbance, and the control action belong to uncountable spaces does not change the problem fundamentally. The methodology of section 2 applies here—the technical details are a bit more involved. Notice that the existence of an optimal design is not guaranteed for this problem. However, since the function spaces are compact, there exist ε-optimal designs.

5.1. Solution methodology. For a fixed encoder, the design of an optimal controller is a centralized stochastic control problem as in the case of Problem 2.1. We need to modify the definition of beliefs, given by Definition 2.2, to take the uncountable state space into account.

Definition 5.1. *For any $A_X \in \mathscr{B}(\mathbb{R}^{d_X})$, $A_M \in 2^{\mathcal{M}}$, where $2^{\mathcal{M}}$ denotes the power set of $\mathcal{M}$, define the measurable transforms $\underline{B}_t$, B_t, and $\overline{B}_t$ as follows:*

1. $\underline{B}_t(A_X, A_M) \coloneqq \Pr(X_t \in A_X, M_{t-1} \in A_M \,|\, \underline{I}_t)$.
2. $B_t(A_X, A_M) \coloneqq \Pr(X_t \in A_X, M_{t-1} \in A_M \,|\, I_t)$.
3. $\overline{B}_t(A_X, A_M) \coloneqq \Pr(X_t \in A_X, M_t \in A_M \,|\, \overline{I}_t)$.

Lemma 2.3 can be proved as before by using Bayes rule for continuous valued random variables. Lemma 2.3 implies that Theorem 2.4 also holds in this case. Thus without loss of optimality, we can restrict our attention to controllers of the form

$$U_t = g_t(\overline{B}_t), \tag{5.1}$$

that is, the controller belonging to $\mathscr{G}_S$, the family of $\mathscr{B}(\mathcal{P}^{\mathcal{X}\times\mathcal{M}})/\mathscr{B}(\mathbb{R}^{d_U})$ measurable functions from $\mathcal{P}^{\mathcal{X}\times\mathcal{M}}$ to $\mathbb{R}^{d_U}$. Thus, at each stage we can optimize over a fixed (rather than a time-varying) domain.

With this reduction, we can define information states $\underline{\pi}_t$, π_t, $\overline{\pi}_t$ as in Definition 2.5, with the beliefs given by Definition 5.1. It is easy to show that these information states satisfy Lemma 2.6. Thus they are sufficient for performance evaluation and lead to a sequential decomposition of the problem. An ε-optimal design can be obtained by the nested optimality equations (2.38)–(2.41). Similar results extend to infinite horizon problems using the ideas of section 4.

5.2. Computational issues. Numerically, problems where the state space is uncountable are much harder than problems with finite state space. This increase in complexity does not arise from the increase in dimensionality of the information state; as a matter of fact, the information states for finite and uncountable state space problems belong to isomorphic spaces. The uncountable state space problems are harder to solve due to the increase in the complexity of the action space. Let us first consider some results from probability theory [11, Appendices 1–5] to show that these information states belong to isomorphic spaces.

DEFINITION 5.2 (Borel space). *A measurable space B is called* Borelian *or a* Borel space *if it is isomorphic to a measurable subset of a Polish (i.e., a complete separable metric) space E.*

Consider the following Borel spaces:

1. A finite or countable space D, with the σ-algebra of all subsets.
2. The unit interval J with the σ-algebra of all open subintervals.

THEOREM 5.3. *Every Borel space is isomorphic to either D or J.*

THEOREM 5.4. *Suppose that $\mathcal{P}^E$ is the set of all probability measures on the space E. If E is a Borel space, then $\mathcal{P}^E$ is also a Borel space.*

For the finite state space problem, let E denote the space $\mathcal{X} \times \mathcal{M}$ with σ-algebra $2^{\mathcal{X}\times\mathcal{M}}$; for the uncountable state space problem, let E denote the space $\mathbb{R}^{d_X} \times \mathcal{M}$ with σ-algebra $\mathscr{B}(\mathbb{R}^{d_X}) \times 2^{\mathcal{M}}$. Then E is Borelian, and, by Theorem 5.4, the space $\mathcal{P}^E$ of probability measures on E is a Borel space. By the same argument, the space Π of probability measures on $\big(E \times \mathcal{P}^E, \mathscr{B}(E \times \mathcal{P}^E)\big)$ is a Borel space. Thus the information state for the finite state space problem and the information state for the infinite state space problem are isomorphic; each is isomorphic to J, the unit interval with σ-algebra of all open subintervals.

The dimensionality of the information state is only one component that determines the complexity of the numerical solution; the dimensionality of the action space is another. In our problem the action spaces alternate between $\mathscr{C}$, $\mathscr{L}$, and $\hat{\mathscr{G}}$. For the finite state space problem, $\mathscr{C}$, $\mathscr{L}$, and $\hat{\mathscr{G}}$ are the family of functions from $\mathcal{X} \times \mathcal{M}$ to $\mathcal{Z}$, $\mathcal{X} \times \mathcal{M}$ to $\mathcal{M}$, and Π to $\mathcal{U}$, respectively. For the uncountable state space problem, $\mathscr{C}$, $\mathscr{L}$, and $\hat{\mathscr{G}}$ are the family of functions from $\mathbb{R}^{d_X} \times \mathcal{M}$ to $\mathcal{Z}$, $\mathbb{R}^{d_X} \times \mathcal{M}$ to $\mathcal{M}$, and Π to $\mathbb{R}^{d_U}$, respectively. Thus the complexity of all three function spaces increases when we go to the uncountable state space problems; this increase in complexity makes it harder to obtain numerical solutions in the case of uncountable state space problems.

5.3. Unstable systems. Consider a system with uncountable (and unbounded) state and action spaces with average cost per unit time as the performance criterion. Suppose the instantaneous cost equals $\rho_{\max}$ whenever the L_p norm of the state is greater than some constant; further, the plant dynamics and the channel are such that the system is unstable under any communication and control strategy (see the papers on stability of NCS mentioned in the introduction for various conditions under which this can happen). Then, all policies would asymptotically incur the maximum cost, and the average cost per unit time would be equal to $\rho_{\max}$. Thus, if a system with uncountable state and action spaces cannot be stabilized under any policy and the cost satisfies the above property, then all policies are optimal or ε-optimal for the average cost per unit time criterion.

6. Imperfect observations. So far we have assumed that the sensor perfectly observes the state of the plant. However, in many practical systems, the sensor observations are noisy due to external disturbances and the intrinsic noise in the measurement hardware. In this section we model this scenario and show that noisy

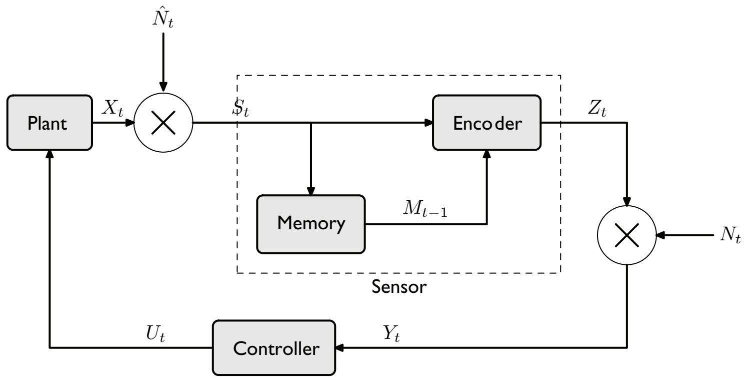

FIG. 6.1. *Feedback control system with noisy communication and imperfect observations.*

observations by the sensor do not alter the nature of the problem. We first consider the finite horizon case.

6.1. Problem formulation. Consider a discrete time imperfect observation system, as shown in Figure 6.1, which operates for T time steps. The state of the plant X_t evolves according to (2.1). The observations S_t made by the observer at time t are a noise-corrupted version of the state of the plant and are given by

$$S_t = \hat{h}(X_t, \hat{N}_t), \tag{6.1}$$

where $\hat{N}_t$ denotes the observation noise and $\hat{h}$ is the observation channel. S_t takes values in $\mathcal{S} := \{1, \dots, |\mathcal{S}|\}$ and $\hat{N}_t$ takes values in $\hat{N} := \{1, \dots, |\hat{N}|\}$. The sequence of random variables $\hat{N}_1, \dots, \hat{N}_T$ is i.i.d. with PMF $P_{\hat{N}}$. The sequence $\hat{N}_1, \dots, \hat{N}_T$ is also independent of $X_1, W_1, \dots, W_T, N_1, \dots, N_T$.

The sensor is modeled as in section 2.1 and operates as follows:

$$Z_t = c_t(S_t, M_{t-1}), \tag{6.2}$$

$$M_t = l_t(S_t, M_{t-1}). \tag{6.3}$$

All other components of the system (the channel, the controller, and the performance metric) are modeled as in section 2.1. The collection of $(\mathcal{X}, \mathcal{W}, \hat{\mathcal{N}}, \mathcal{S}, \mathcal{M}, \mathcal{Z}, \mathcal{N}, \mathcal{Y}, \mathcal{U}, P_{X_1}, P_W, P_{\hat{N}}, P_N, f, h, \hat{h}, \rho, T)$ is called an *imperfect observation system*. The choice of (C, L, G), $C := (c_1, \dots, c_T)$, $L := (l_1, \dots, l_T)$, $G := (g_1, \dots, g_T)$ is called a *design*. The performance of a design, quantified by the expected total cost under that design, is given by (2.6). We are interested in the following optimization problem.

PROBLEM 6.1. *Given an imperfect observation system* $(\mathcal{X}, \mathcal{W}, \hat{\mathcal{N}}, \mathcal{S}, \mathcal{M}, \mathcal{Z}, \mathcal{N}, \mathcal{Y}, \mathcal{U}, P_{X_1}, P_W, P_{\hat{N}}, P_N, f, h, \hat{h}, \rho, T)$*, choose a design* (C^*, L^*, G^*) *such that*

$$\mathcal{J}_T(C^*, L^*, G^*) = \mathcal{J}_T^* := \min_{C,L,G \in \mathscr{C}^T \times \mathscr{L}^T \times \mathscr{G}^T} \mathcal{J}_T(C, L, G), \tag{6.4}$$

where $\mathscr{C}^T := \mathscr{C} \times \cdots \times \mathscr{C}$ *(*T *times),* $\mathscr{C}$ *is the space of functions from* $\mathcal{S} \times \mathcal{M}$ *to* $\mathcal{Z}$, $\mathscr{L}^T := \mathscr{L} \times \cdots \times \mathscr{L}$ *(*T *times),* $\mathscr{L}$ *is the space of functions from* $\mathcal{S} \times \mathcal{M}$ *to* $\mathcal{M}$, $\mathscr{G}^T := \mathscr{G}_1 \times \cdots \times \mathscr{G}_T$*, and* $\mathscr{G}_t$ *is the space of functions from* $\mathcal{Y}^t \times \mathcal{U}^{t-1}$ *to* $\mathcal{U}$.

Although in Problem 6.1 the encoder does not know the state of the plant, the problem is conceptually the same as Problem 2.1, and the solution methodology of Problem 2.1 works for Problem 6.1 with very minor changes.

6.2. Solution methodology. First, we present structural properties of optimal controllers. Then we use these structural properties to obtain a sequential decomposition of Problem 6.1. For this purpose, define the following.

DEFINITION 6.1. *Let* $\underline{B}_t$ *and* $\overline{B}_t$ *be defined as in Definition* 2.5. *Define* B_t *as follows:*

$$B_t(x, s, m) := \Pr(X_t = x, S_t = s, M_{t-1} = m \mid I_t).$$

These beliefs are related, as in Lemma 2.3, which implies that the structural results of Theorem 2.4 also hold for Problem 6.1. Thus, without loss of optimality, we can restrict our attention to controllers of the form (2.23). These structural results imply that we can formulate a problem similar to Problem 6.1 with a time-invariant action space.

Now define $\underline{\pi}_t$, π_t, and $\overline{\pi}_t$ as in Definition 2.5, with B_t defined as in Definition 6.1. These information states $\underline{\pi}_t$, π_t, and $\overline{\pi}_t$ satisfy Lemma 2.6. Hence, Problem 6.1 is equivalent to a deterministic problem similar to that of Problem 2.3 with the transformations $\underline{Q}$, Q, and $\overline{Q}$ appropriately defined. The solution of this deterministic problem is given by nested optimality equations similar to those of Theorem 2.7. Hence, we obtain a sequential decomposition of Problem 6.1. Similar results extend to infinite horizon problems using the ideas of section 4.

7. Conclusion. We have presented a methodology for determining globally optimal (or globally ε-optimal) encoding and control strategies for networked control systems (NCS) with nonclassical information structure. The methodology is applicable to finite horizon problems with an expected total cost criterion, to infinite horizon problems with an expected discounted cost criterion, and to infinite horizon problems with an average cost per unit time criterion. We have extended this methodology to problems where the encoder/sensor makes imperfect observations about the state of the system. The resulting optimality equations can be viewed as partially observed Markov decision problems (POMDPs) where the state space is a real-valued vector and the action space is uncountable. There are very few results on efficient computational techniques for this class of POMDPs. We hope that the problem of optimal control over a noisy communication channel will motivate researchers to investigate numerical methods for optimization problems that are of the type in Problem 2.3.

For the problems considered in this paper, the action space is uncountable because of the assumption of perfect recall at the controller's site. In light of the sequential decomposition for decentralized team problems presented in this paper, this assumption of perfect recall needs to be reconsidered. For most applications, the assumption of perfect recall, that is, the assumption that an agent remembers everything that it has seen and everything that it has done in the past, is impractical. Nevertheless, in centralized stochastic control problems perfect recall is assumed since it implies a classical information structure, and it simplifies the solution methodology. In decentralized problems, the information structure is nonclassical, and it remains nonclassical even with the unrealistic assumption of perfect recall at each agent's site. Further, the assumption of perfect recall makes it harder to obtain a numerical solution of the resultant nested optimality equations. In the problems considered in this paper, the sensor/encoder has finite memory, while the controller has perfect recall. In the nested optimality equations of Theorem 2.7, to obtain an optimal encoder and memory update rule in (2.39) and (2.40) we need to choose c_t and l_t belonging to $\mathscr{C}$ and $\mathscr{L}$, respectively; both $\mathscr{C}$ and $\mathscr{L}$ are finite spaces. On the other hand, to obtain an optimal controller in (2.41) we need to choose g_t belonging to $\hat{\mathscr{G}}$, which is an uncountable space; even though the action space $\mathcal{U}$ is finite, to choose an optimal g_t we have to search over an uncountable space. If we had assumed a finite memory at the controller, we would have obtained equations in which we need to choose a control law and a memory update rule at the controller from a finite set, and this problem is similar to a POMDP with finite action space. Thus, the unrealistic assumption of perfect recall at any agent's site does not simplify the analysis but rather makes the problem numerically more difficult to solve, while the realistic assumption of a finite memory at all agents' sites results in a solution algorithm that is easier to solve.

It is important to identify special cases in which the information states $\underline{\pi}_t$, π_t, and $\overline{\pi}_t$ can be restricted to a parametric family of distributions. In centralized stochastic control problems, linear quadratic Gaussian (LQG) systems possess such a property—the information state can be restricted to Gaussian distributions. This is because in LQG systems with a classical information pattern, without any loss of optimality we can restrict our attention to affine control laws, which implies that the state of the plant is always Gaussian. Thus the information state—which is the conditional probability of the state of the plant, conditioned on all the past observations and all the past control actions of the controller—is also Gaussian. This simplifies the search for an optimal design. Unfortunately, in decentralized systems (more precisely, in systems with a nonclassical information structure) nonlinear control laws can outperform affine control laws even in linear systems where all primitive random variables are Gaussian, as illustrated by the Witsenhausen counterexample [40]. So, the state of the plant may not be Gaussian, and hence the information state need not be Gaussian. However, there may be other special cases for which information states in a decentralized system belong to a parametric family of distributions. Finding such special cases remains a challenging open problem.

The results of section 4 show that for infinite horizon problems stationary designs are not optimal, and in order to implement a time-varying optimal design, we need to implement an optimal stationary metadesign. Thus, implementing optimal designs for decentralized systems is an order of magnitude more complicated as compared to centralized systems. Traditionally, for infinite horizon decentralized control problems, performance limitations of only stationary designs is considered. It will be worthwhile to characterize the performance difference between an optimal time-varying design and the best stationary design. It will also be important to obtain performance limitations of time-varying optimal designs.

Appendix. Nested σ-algebras. We first present a general lemma and then use its result to justify the statement made in the discussion in section 3.

LEMMA A.1. *Consider a probability space* $(\Omega, \mathcal{F}, P)$. *Let* X *and* Y *be real-valued random variables defined on* $(\Omega, \mathcal{F}, P)$, *and let* $g : (\mathbb{R}, \mathscr{B}(\mathbb{R})) \to (\mathbb{R}, \mathscr{B}(\mathbb{R}))$ *be a measurable real-valued function. Then*

$$\sigma(X) \cap \sigma(Y) \subseteq \sigma(X, g(Y)). \tag{A.1}$$

Proof. Consider any set A belonging to $\sigma(X) \cap \sigma(Y)$. Then, there exist sets B_1 and B_2 belonging to $\mathscr{B}(\mathbb{R})$ such that $A = X^{-1}(B_1)$ and $A = Y^{-1}(B_2)$. Define a real-valued random variable Z on $(\Omega, \mathcal{F}, P)$ by $Z(\omega) = g\big(Y(\omega)\big)$. Let $B_3 \coloneqq g(B_2)$. Now, $g^{-1}(B_3) \supseteq B_2$, so $Z^{-1}(B_3) \coloneqq Y^{-1}\big(g^{-1}(B_3)\big) \supseteq Y^{-1}(B_2) = A$. Thus,

$$X^{-1}(B_1) \cap Z^{-1}(B_3) = A. \tag{A.2}$$

Hence, $A \in \sigma(X, Z)$, and thus

$$\sigma(X) \cap \sigma(Y) \subseteq \sigma(X, Z) = \sigma\big(X, g(Y)\big). \qquad \square \tag{A.3}$$

Now, in the discussion in section 3, we claimed that

$$K_t(\gamma^t) \coloneqq \sigma(X_t, M_{t-1}) \cap \sigma(Y^t, U^{t-1}) \subseteq \sigma(X_t, M_{t-1}, \underline{B}_t) =: \hat{K}_t(\gamma^t). \tag{A.4}$$

This follows by taking

$$X = (X_t, M_{t-1}), Y = (Y^t, U^{t-1}) \text{ and } g(Y^t, U^{t-1}) = \Pr\left(X_t, M_{t-1} \,\middle|\, Y^t, U^{t-1}, \gamma^t\right)$$

in Lemma A.1.

Acknowledgments. We are grateful to the referees and the associate editor for their comments which helped us to significantly improve the presentation of the paper.

REFERENCES

[1] P. J. Antsaklis and J. Baillieul, eds., *Special issue on networked control systems*, IEEE Trans. Automat. Control, 49 (2004), pp. 1424–1603.

[2] P. J. Antsaklis and J. Baillieul, eds., *Special issue on technology of networked control systems*, Proc. IEEE, 95 (2007) pp. 1–317.

[3] R. J. Aumann, *Agreeing to disagree*, Ann. Statist., (1976), pp. 1236–39.

[4] J. Baillieul, *Feedback design in information based control*, in Proceedings of the Workshop in Stochastic Theory and Control, B. Pasik-Duncan, ed., Springer-Verlag, New York, 2001, pp. 35–27.

[5] J. Baillieul, *Feedback coding for information based control—operating near the data-rate limit*, in Proceedings of the 41st IEEE Conference on Decision and Control, IEEE, Piscataway, NJ, 2002, pp. 3229–3236.

[6] J. H. Braslavsky, R. H. Middleton, and J. S. Freudenberg, *Feedback stabilization over signal-to-noise ratio constrained channels*, IEEE Trans. Automat. Control, 52 (2007), pp. 1391–1403.

[7] R. W. Brockett and D. Liberzon, *Quantized feedback stabilization of linear systems*, IEEE Trans. Automat. Control, 45 (2000), pp. 1279–1289.

[8] G. W. Brown, *Iterative solutions of games by fictitious play*, in Activity Analysis of Production and Allocation, T. C. Koopmans, ed., Wiley, New York, 1951, pp. 374–376.

[9] L. G. Bushnell, ed., *Special section on networks and control*, IEEE Control Systems Mag., 21 (2001), pp. 22-99.

[10] D. F. Delchamps, *Stabilizing a linear system with quantized state feedback*, IEEE Trans. Automat. Control, 35 (1990), pp. 916–924.

[11] E. B. Dynkin and A. A. Yushkevich, *Controlled Markov Processes*, Grundlehren. Math. Wiss. 23, Springer-Verlag, Berlin, 1979.

[12] N. Elia and S. K. Mitter, *Stabilization of linear systems with limited information*, IEEE Trans. Automat. Control, 46 (2001), pp. 1384–1400.

[13] O. Hernández-Lerma, *Adaptive Markov Control Processes*, Springer-Verlag, New York, 1989.

[14] Y.-C. Ho, *Team decision theory and information structures*, Proc. IEEE, 68 (1980), pp. 644–654.

[15] H. Ishii and B. A. Francis, *Quadratic stabilization of sampled-data systems with quantization*, Automatica, 39 (2003), pp. 1793–1800.

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

[17] D. Liberzon, *On stabilization of linear systems with limited information*, IEEE Trans. Automat. Control, 48 (2003), pp. 304–307.

[18] A. Mahajan, *Sequential Decomposition of Sequential Dynamic Teams: Applications to Real-Time Communication and Networked Control Systems*, Ph.D. thesis, University of Michigan, Ann Arbor, MI, 2008.

[19] A. Mahajan and D. Teneketzis, *Optimal design of sequential real-time communication systems*, submitted. Available as Control Group Report CGR-06-03, Department of EECS, University of Michigan, Ann Arbor, MI 48109-2122.

[20] J. Marschak and R. Radner, *Economic Theory of Teams*, Yale University Press, New Haven, CT, 1972.

[21] N. C. Martins and M. A. Dahleh, *Feedback control in the presence of noisy channels: "Bode-like" fundamental limitations of performance*, IEEE Trans. Automat. Control, 53 (2008), pp. 1604–1615.

[22] N. C. Martins, M. A. Dahleh, and J. C. Doyle, *Fundamental limitations of disturbance attenuation in the presence of side information*, IEEE Trans. Automat. Control, 52 (2007), pp. 56–66.

[23] N. C. Martins, M. A. Dahleh, and N. Elia, *Feedback stabilization of uncertain systems in the presence of a direct link*, IEEE Trans. Automat. Control, 51 (2006), pp. 438–447.

[24] A. S. Matveev and A. V. Savkin, *Problem of LQG optimal control via a limited capacity communication channel*, System Control Lett., (2004), pp. 51–64.

[25] G. N. Nair and R. J. Evans, *Stabilization with data-rate limited feedback: Tightest attainable bounds*, Systems Control Lett., 41 (2000), pp. 304–307.

[26] G. N. Nair and R. J. Evans, *Exponential stabilisability of finite-dimensional linear systems with limited data rates*, Automatica, 39 (2003), pp. 585–593.

[27] G. N. Nair and R. J. Evans, *Stabilizability of stochastic linear systems with finite feedback data rates*, SIAM J. Control Optim., 43 (2004), pp. 413–436.

[28] G. N. Nair, R. J. Evans, I. M. Y. Mareels, and W. Moran, *Topological feedback entropy and nonlinear stabilization*, IEEE Trans. Automat. Control, 49 (2004), pp. 1585–1597.

[29] G. N. Nair, F. Fagnani, S. Zampieri, and R. J. Evans, *Feedback control under data rate constraints: An overview*, Proc. IEEE, 95 (2007), pp. 108–137.

[30] I. R. Peterson and A. V. Savkin, *Multi-rate stabilization of multivariable discrete-time linear systems via a limited capacity communication channel*, in Proceedings of the 40th IEEE Conference on Decision and Control, 2001, IEEE, Piscataway, NJ, pp. 304–309.

[31] J. Robinson, *An iterative method for solving a game*, Ann. Math., 54 (1951), pp. 296–301.

[32] N. R. Sandell, Jr., *Control of Finite-State, Finite-Memory Stochastic Systems*, Ph.D. thesis, Massachusetts Institute of Technology, Cambridge, MA, 1974.

[33] A. V. Savkin, *Analysis and synthesis of networked control systems: Topological entropy, observability, robustness and optimal control*, Automatica, 42 (2006), pp. 51–62.

[34] L. S. Shapley, *Some topics in two-person games*, in Advances in Game Theory, M. Dresher, L. S. Shapley, and A. W. Tucker, eds., Ann. Math. Stud. 52, Princeton University Press, Princeton, NJ, 1964, pp. 1–28.

[35] S. Tatikonda, *Control under Communication Constraints*, Ph.D. thesis, Department of EECS, Massachusetts Institute of Technology, Cambridge, MA, 2000.

[36] S. Tatikonda and S. K. Mitter, *Control over noisy channels*, IEEE Trans. Automat. Control, 49 (2004), pp. 1196–1201.

[37] S. Tatikonda and S. K. Mitter, *Control under communication constraints*, IEEE Trans. Automat. Control, 49 (2004), pp. 1056–1068.

[38] S. Tatikonda, A. Sahai, and S. K. Mitter, *Stochastic linear control over a communication channel*, IEEE Trans. Automat. Control, 49 (2004), pp. 1549–1561.

[39] J. C. Walrand and P. Varaiya, *Causal coding and control of Markov chains*, System Control Lett., 3 (1983), pp. 189–192.

[40] H. S. Witsenhausen, *A counterexample in stochastic optimum control*, SIAM J. Control, 6 (1968), pp. 131–147.

[41] H. S. Witsenhausen, *Separation of estimation and control for discrete time systems*, Proc. IEEE, 59 (1971), pp. 1557–1566.

[42] H. S. Witsenhausen, *A standard form for sequential stochastic control*, Math. Systems Theory, 7 (1973), pp. 5–11.

[43] H. S. Witsenhausen, *The intrinsic model for discrete stochastic control: Some open problems*, in Control Theory, Numerical Methods and Computer System Modelling, A. Bensoussan and J. L. Lions, eds., Lecture Notes in Econom. and Math. Systems 107, Springer, Berlin, 1975, pp. 322–335.

[44] H. S. Witsenhausen, *Some remarks on the concept of state*, in Directions in Large-Scale Systems, Y. C. Ho and S. K. Mitter, eds., Plenum Press, New York, 1976, pp. 69–75.

[45] W. S. Wong and R. W. Brockett, *Systems with finite communcation bandwidth constraints* II*: Stabilization with limited information feedback*, IEEE Trans. Automat. Control, 44 (1999), pp. 1049–1053.

[46] S. Yüksel and T. Başar, *Control over noisy forward and reverse channels*, IEEE Trans. Automat. Control, to appear.