MARKOV CONTROL PROBLEMS UNDER COMMUNICATION CONSTRAINTS*

V. S. BORKAR[†], SANJOY MITTER[‡], AND SEKHAR TATIKONDA[‡]

Abstract. In traditional control systems theory one assumes that the controller receives the observations from the sensors instantaneously and exactly, and the control action it initiates is immediately effective. In reality, however, the controller and the sensors may be geographically separated from the plant and the observations may be transmitted from the sensors to the controller across a communication channel. In addition to adding noise, the channel can also put a constraint on the rate of information transmission and introduce delays. In a similar way the path from the controller back to the plant can be modeled as a communication channel with noise and delays. Hence the need for a control theory that explicitly takes into account these aspects. The aim of the present work is to view the problem in its totality and come up with a set of paradigms that can form a basis for analysis and synthesis of such control systems.

Keywords: Markov control problem, Communication constraints

1. Introduction. In traditional control systems theory one assumes that the controller receives the observations from the sensors instantaneously and exactly, and the control action it initiates is immediately effective. In reality, however, the controller and the sensors may be geographically separated from the plant and the observations may be transmitted from the sensors to the controller across a communication channel. In addition to adding noise, the channel can also put a constraint on the rate of information transmission and introduce delays. In a similar way the path from the controller back to the plant can be modeled as a communication channel with noise and delays. Hence the need for a control theory that explicitly takes into account these aspects. Recent papers [8], [9], and [25] address various aspects of this problem. For related work see [11], [13], and [14]. Also see Nair and Evans work on state estimation under a bit-rate constraint [19]. The aim of the present work is to view the problem in its totality and come up with a set of paradigms that can form a basis for analysis and synthesis of such control systems.

Here we analyze controlled Markov chains with communication constraints, wherein we treat the encoding scheme as well as the control scheme as a control choice. This idea has been discussed previously by Walrand and Varaiya[26], Borkar, Mitter, and Tatikonda [9], and Feng and Loparo [13]. In general for this control problem it is a very difficult to compute optimal distributed controllers. See Papadimitriou and Tsitsiklis [20]. We first consider some "solvable" cases and for the more general cases,

we motivate and propose some approximation schemes.

A broad class of such Markov control problems can be solved under the assumption of "equimemory." Specifically we assume that the encoder has access to the decoder's state. In this setting we provide solutions to the average cost Markov control problem with continuous state space subject to an average entropy rate restriction on the channel transmissions. We show that the optimal control policy is a randomized stationary Markov policy. (Recall that for controlled Markov processes with additional cost constraints, one can in general claim only optimal *randomized* stationary policies [1].) Previous work on the continuous state case can be found in [11], [13], [14], and [8].

"Equimemory" is not always a valid assumption. Relaxing this assumption, though, makes the problem become much more difficult. Here we study the discounted cost Markov control problem with finite state space subject to a fixed rate constraint on the channel transmissions. In this setting we concern ourselves with different regimes of possible actions. These regimes are determined by the complexity of the controllers, what the controllers and encoder know about each other, and the relative ordering of the cardinality of the relevant spaces: state space, observation space, code space and action space.

The fixed rate constraint implies fixed-length coding. In general one can achieve better performance through variable-length coding. But as expected, the analysis becomes more difficult. In particular, issues of appropriate prefix coding and variable delays in transmission come to the fore. See Borkar and Mitter [8] for a discussion in the context of innovation coding and Wong and Brockett [28], [29]. in the context of state estimation. There is an inherent tradeoff between coarseness of the quantization and the delays involved in transmitting the data over feedback channels.

In the most general case the sensors are geographically separate from the plant. Herein we assume that the sensors and the plant are located at the same place. Furthermore we assume that the sensors have access to a noisy observation of the state of the plant as well as a noiseless access to the past control signals to the plant. This is important as it allows the encoder access to delayed outputs of the controller. The controller is allowed noisy delayed observations of the encoder. Thus the two control actions, controller and encoder, can observe each other. We discuss later how to relax these assumptions.

The paper is organized as follows: it broadly splits into two parts. The first, consisting of sections 2 and 3, takes up some "solvable" cases, involving (not surprisingly) certain strong hypotheses. These involve various technical conditions on the control process and more crucially, the previously stated "equimemory" hypothesis which ensures that both the encoder and the controller have access to the same sufficient statistics. The importance of this condition has already been underscored in the literature on sequential source coding [15], where, in fact, the term was first coined. Under this and other strong conditions, section 2 takes up the problem of vector quantizer-encoder design and section 3 embeds it into the overall control loop.

Section 4 onwards we take up the more general problem, involving partial observation and nonclassical information structures (read "no equimemory hypothesis"). For simplicity of analysis, we take the state space of the control process to be finite

and memoryless. As stated before there are two separate decision variables involved: the decision of the encoder situated at the plant to use a particular encoding scheme and the decision of the controller to use a particular control action. We consider four distinct paradigms, depending on whether the decision is made in a "static" manner or a "dynamic" manner (i.e., using all available information at the time) at either end. Section 4 describes these paradigms, termed (in increasing order of complexity) static-static, dynamic-static, static-dynamic, and dynamic-dynamic. The first adjective refers to the encoder and the second to the controller. Section 5 takes up the analysis of each of these in order. Section 6 concludes with a list of further problems that need to be addressed.

2. The Vector Quantizer. Our controlled process will be a controlled Markov chain $X(t), t = 0, 1, 2, ...$, taking values in a closed, bounded subset $S \subset \mathbb{R}^d, d \geq 1$, satisfying: $S = \overline{\text{int}(s)}, \text{int}(S) \neq 0$, controlled by a control process $Z(t), t = 0, 1, 2, ...$, taking values in a compact metric (hence Polish) control space U. Note that because $X(t)$ is restricted to the bounded set S the system is "stable" (i.e. the state cannot run off to infinity.) The evolution of $X(\cdot)$ is described by

$$(2.1) \qquad P(X(t+1) \in A | X(s), Z(s), s \leq t) \triangleq p(X(t), Z(t), A)$$

for a Borel set $A \subset S$, where $p(x, z, dy)$ is the (controlled) transition kernel satisfying: the map $(x, z) \in S \times U \to p(x, z, dy) \in \mathcal{P}(S)$ is continuous. (Here and later on $\mathcal{P}(\cdots)$ denotes the Polish space of probability measures on a Polish space "$\cdots$" with Prohorov topology – see e.g. [7], Ch. 2.) The class $Z(\cdot)$ for which (2.1) holds will be called the "admissible controls" and will be the most general class of control that we shall admit. (Intuitively, these are the control processes satisfying the natural minimal requirement that they do not anticipate the future.) Later on, we shall specialize this class further.

Let $\Sigma = \{\alpha_1, \alpha_2, ...\}$ be an ordered set that will serve as the alphabet for the encoder. Let $\{q(t)\}$ denote the Σ-valued process that stands for the encoded, "vector quantized," version of $\{X(t)\}$. We describe next the passage from $X(\cdot)$ to $q(\cdot)$, wherein we closely mimic [9].

Let D denote the set of finite nonempty subsets of S satisfying the following condition:

(†) There exists a fixed $\Delta > 0$ such that for any $A \in D$ and any distinct $[x_1, ..., x_d], [y_1, ..., y_d] \in A, |x_i - y_i| \geq \Delta \forall i$.

We endow D with the Hausdorff metric which renders it a compact Polish space. For $A \in D$, let $l_A : S \to A$ denote the map that maps $x \in S$ to the element of A that is nearest to it with respect to the Euclidean norm $\| \cdot \|$, any tie being resolved according to some a priori fixed priority rule. Let $i_A : A \to \Sigma$ denote the one-to-one map that first orders the elements $\{a_1, ..., a_m\}$ of A lexico-graphically and then maps them to $\{\alpha_1, ..., \alpha_m\}$ preserving the order.

Note that we have chosen to use a nearest-neighbor, with respect to the Euclidean norm, type of quantizer. More generally one should use a quantizer that is somehow induced from the cost. We choose the Euclidean nearest-neighbor quantizer because the analysis is easier and it is practical in applications.

At each time t, a measurable map $\eta_t : \Sigma^\infty \to D$ is chosen. With $Q_t \triangleq \eta_t(q^t)$ (where $q^t \triangleq [q(t), q(t-1), ...]$), on sets $q(t+1) = i_{Q_t} \circ l_{Q_t}(X(t+1))$, which defines $\{q(t)\}$ recursively, with $q(0) \triangleq$ the law of $X(0)$. This is the process that is suitably encoded and transmitted across a channel.

The interpretation of this scheme is as follows: If one were to use a fixed vector quantizer, the finite subset of $\mathbb{R}^d$ onto which the signal gets mapped can itself be identified with the finite alphabet Σ. In our case, however, this set is dynamically selected and therefore time-varying. Hence it must be mapped to a fixed alphabet Σ in a uniquely invertible manner. This is achieved through the map i_A. Assuming that the receiver (decoder) knows ahead of time the maps $\{\eta_t(\cdot)\}$ (we shall comment more on this requirement later, simplifying it considerably), he can reconstruct Q_t as $\eta_t(q^t)$, having received q^t be time t. (The latter is an <u>assumption</u>. More on this later.) The contribution of the condition (†) is to render the map $A = \{a_1, ..., a_m\} \in D \to \{i_A(a_1), .., i_A(a_m)\} \in \Sigma^*$ continuous. Not only does this make sense from the point of view of robust decoding, but it also makes the control problems we formulate later well-posed.

We should comment, though, that if the standard deviation of the state conditioned on the past codewords and controls is smaller than Δ then, roughly speaking, the vector quantizer cannot adjust its points to match the source and will yield errors on the order of Δ. In other words Δ is a measure of the finest resolution our quantizers can achieve.

Recall our assumption that the decoder has at his end the complete knowledge of q^t at time t. This is because we assume noiseless transmission. This is certainly an idealization, but is one way of ensuring "equimemory" for the encoder and the decoder, necessary for making the control problem tractable. (We shall consider another variation of the "equimemory" condition later on.)

We picture the encoder as being at the plant end and the decoder as being at the controller end. (See figure 4.1.) The controller thus has $\{q(t)\}$ as his observation process. The celebrated certainty equivalence principle of stochastic control theory [2] then suggest that the correct "state" process at the controller end is $\{\pi_t\}$, the process of conditional laws of $X(t)$ given $q^t, t \geq 0$. This process is given recursively by the nonlinear filter (the derivation follows by a routine application of the Bayes rule.)

$$(2.2) \qquad \pi_{t+1}(dy) = \frac{\int I_{\{i_{Q_t} \circ l_{Q_t}(y) = q(t+1)\}} p(x, Z(t), dy) \pi_t(dx)}{\int \int I_{\{i_{Q_t} \circ l_{Q_t}(z) = q(t+1)\}} p(x, Z(t), dz) \pi_t(dx)}.$$

Of course, this makes sense only if the denominator is nonzero. [1]

---

[1] One condition that ensures this is the following. Assume transition kernel $(x, z) \to p(x, z, dy)$: $\forall x, z, \ p(x, z, dy)$ is absolutely continuous with the Lebesgue measure $\lambda(dy)$ on S, and there exist constants $\sigma_2 > \sigma_1 > 0$ and a function $\psi \in C(S)$ with $\psi(\cdot) > 0$ such that

$$(2.3) \qquad \sigma_1 \psi(y) \leq \frac{dp(x, z, dy)}{d\lambda(dy)}(y) \leq \sigma_2 \psi(y) \quad \forall \, y, x, z.$$

In particular, $p(x, z, dy)$ has a density with respect to $\lambda(dy)$ that is strictly positive almost everywhere.

Now observe that by (†), $l_A^{-1} \circ i_A^{-1}(a)$ contains a relatively open set of S for all a, A. Then (2.3) ensures that for all $x, z, \ p(x, z, B) > 0$ for all relatively open B, which takes care of potential pathologies.

The controller then chooses the control signal Z_t based on $\pi^t \triangleq \{\pi_s(dy), s \leq t\}$, and communicates it to the plant, in principle, across yet another communication channel. We shall not model the latter channel explicitly. This is because, as long as it is memoryless, it can easily be accounted for by suitably modifying the map $z \to p(x, z, dy)$.

This completes our description of the control loop under the first variation of "equimemory" hypothesis. Note that both the encoder and the decoder can compute $\{\pi_t\}$.

In the second variation, we suppose that $\{q_t\}$ is being transmitted across a noisy memoryless channel with input alphabet Σ and an output alphabet $\Lambda = \{\beta_1, \beta_2, ...\}$, with transition probabilities $r(i, j), i \in \Sigma, j \in \Lambda$. (Thus $r(i, j) \geq 0, \sum_j r(i, j) = 1, \forall i, j$). Let $\{\hat{q}(t)\}$ denote the output process of this channel. We impose "equimemory" condition by assuming that $\hat{q}(t)$ is also available to the encoder immediately. Conceptually, this can happen either if it is instantly fed back across a feedback channel, or if it can somehow be inferred from the observed control sequence $Z(s), s \leq t$. (In either case, this is an idealization.) The difference with the previous formulation is that the "state" process for the controller now is the process $\{\hat{\pi}_t\}$ of conditional laws of $X(t)$ given $\hat{q}^t \triangleq \{\hat{q}_s, s \leq t\}$, given recursively by the nonlinear filter

$$(2.4) \qquad \hat{\pi}_{t+1}(dy) = \frac{\sum_a r(a, \hat{q}(t+1)) \int I_{\{i_{Q_t} \circ l_{Q_t}(y) = a\}} p(x, Z(t), dy) \pi_t(dx)}{\sum_{\tilde{a}} r(\tilde{a}, \hat{q}(t+1)) \int \int I_{\{i_{Q_t} \circ l_{Q_t}(z) = \tilde{a}\}} p(x, Z(t), dz) \pi_t(dx)}$$

Once again, observe that both the encoder and the decoder can compute $\{\hat{\pi}_t\}$ under our hypotheses.

3. The Control Problem. Suppose our aim is to minimize a long run average (or "ergodic") cost

$$\limsup_{N \to \infty} \frac{1}{N} \sum_{t=0}^{N-1} E[\bar{g}(X(t), Z(t))]$$

for a prescribed running cost function $\bar{g} \in C(S \times U)$. Furthermore, we wish to achieves this with an average rate constraint

$$\limsup_{N \to \infty} \frac{H(q^N)}{N} \leq R,$$

where $H(q^N)$ is the Shannon entropy of the string q^N. The entropy is defined by $H(q^N) = -\sum_s \Pr(s) \log \Pr(s)$, where the sum is over all $s \in \Sigma^N$. We shall cast this as an appropriate "certainty equivalent" control problem wherein one seeks to optimally control the nonlinear filter (2.2) (or (2.4), as the case may be) subject to an appropriate criterion. In order to be specific, we shall consider only the first scenario ($\{q(n)\}$ transmitted noiselessly) leading to (2.2). The other scenario that led to (2.4) can be handled in a completely analogous manner.

Let $(\pi, z, A) \in \mathcal{P}(S) \times U \times D \to \Psi(\pi, z, A, d\pi') \in \mathcal{P}(\mathcal{P}(S))$ denote the transition kernel of the $\mathcal{P}(S)$-valued controlled Markov process $\{\pi_t\}$ described by (2.2), where both $\{Z(t)\}$ and $\{Q(t)\}$ are being viewed as respectively U-valued and D-valued control sequences. Recall the following results from [9]:

LEMMA 3.1. *The map $\Psi(\cdot, \cdot, \cdot, d\pi')$ is continuous.*

This is essentially Lemma 4.1 of [9], with a minor modification to account for the control variable $z \in U$.

LEMMA 3.2. $H(q^N) = \sum_{t=0}^{N-1} E\left[k(\pi_t, Z(t), Q_t)\right]$ *where*

$$k(\pi, z, A) \triangleq -\sum_{a \in A} h_a(\pi, z, A) \log h_a(\pi, z, A),$$

with

$$h_a(\pi, z, A) \triangleq \int \pi(dx) \int p(x, z, dy) I_{\{i_A \circ l_A(y) = a\}}.$$

Clearly,

$$E[\bar{g}(X(t), Z(t))] = E[g(\pi_t, Z(t))]$$

where $g(\pi, z) \triangleq \int \pi(dx)\bar{g}(x, z)$. Thus the above control problem can be recast as the problem of $\{\pi_t\}$ governed by (2.2) so as to minimize

$$\limsup_{N \to \infty} \frac{1}{N} \sum_{t=0}^{N-1} E[g(\pi_t, Z(t))]$$

subject to

$$\limsup_{N \to \infty} \frac{1}{N} \sum_{t=0}^{N-1} E\left[k(\pi_t, Z(t), Q_t)\right] \leq R.$$

This is the "ergodic control under constraints," extensively studied in the discrete state space set-up [6], [1]. As in [6], we have

THEOREM 3.3. *The above control problem has an optimal randomized stationary Markov policy.*

This helps little in absence of a scheme to compute the optimal policy, at least approximately. Unfortunately, the constrained control problem does not in general admit a simple dynamic programming formulation [6], [1]. Nevertheless, one can cast it as an infinite dimensional linear program as in [6], [1], and look for appropriate finite dimensional approximation. A better alternative is to change the problem to a single cost problem, by seeking to minimize

$$\limsup_{N \to \infty} \frac{1}{N} \sum_{t=0}^{N-1} E\left[g(\pi_t, Z(t)) + \lambda k(\pi_t, Z(t), Q_t)\right]$$

for a suitable $\lambda > 0$. This is amenable to traditional dynamic programming analysis. Under the assumption of asymptotic flatness [9] we have [2]

---

[2]A stochastic dynamical system is said to be *asymptotically flat* if, when $\{X(x, t)\}$ denotes its trajectory starting at initial state x,

$$E\left[\|X(x, t) - X(y, t)\|\right] \leq K\alpha^t \}x - y\|, \ t \geq 0,$$

for some $K > 0$, $0 < \alpha < 1$.

THEOREM 3.4. *There exists a measurable function* $V : \mathcal{P}(S) \to \mathbb{R}$ *and a constant* $\tilde{V}$ *such that* $\forall \pi \in \mathcal{P}(S)$,

$$V(\pi) + \tilde{V} = \min_{z,A} \left[(g(\pi, z) + \lambda k(\pi, z, A)) + \int \tilde{\psi}(\pi, z, A, d\pi') V(\pi') \right].$$

Furthermore, there exist measurable maps $v : \mathcal{P}(S) \to U, u : \mathcal{P}(S) \to D$ *such that* $(v(\pi), u(\pi))$ *attains the minimum on the right hand side and* $Z(t) = v(\pi_t), Q_t = u(\pi_t), t \geq 0$, *are the optimal control and quantization policies leading to the optimal cost, which equals* $\tilde{V}$.

This follows from the results of [23], which also appear as Theorem 4.7, p. 157 of [21]. [21] also considers related computational issues.

As already mentioned, a completely analogous statement is possible for the second problem, of controlling $\{\hat{\pi}_t\}$ governed by (2.4).

4. The General Control Problem. We now take up the more general control problem. As mentioned before, our "plant" will now be a finite state controlled Markov chain X_n, $n \geq 0$, taking values in a finite state space $S = \{s_1, s_2, ..., s_{|S|}\}$ with an associated "observation process" Y_n, $n \geq 0$ taking values in a finite "observation space" $B = \{b_1, ..., b_{|B|}\}$. It is controlled by a "control process" Z_n, $n \geq 0$, taking values in an action space A. This is selected by a "controller" based upon the noisy, delayed, encoded observation he receives from the "encoder."

The encoder maps the observation Y_n and other relevant information, (i.e. past observations and controls), to a fixed codebook $C = \{c_1, ..., c_{|C|}\}$. The encoded word at time n is $E_n \in C$. We fix the codebook for convenience. Furthermore we assume a fixed-length encoding scheme so that the transmission delay is the same for all codewords. In a more general setting the choice of codebook would also be part of the optimization problem. The delay induced on the channel is due to the length of the codeword. Longer codewords allow for more messages to be sent but at the cost of more delay. Delays lead to information stagnation. See Borkar and Mitter for discussion of this issue [8].

The encoding is transmitted across a finite memoryless channel with finite output codebook $D = \{d_1, ..., d_{|D|}\}$. The input codebook and the output codebook are related by the following channel transition probability $q(\cdot, \cdot) : C \times D \to [0, 1]$ with the interpretation that $c \in C$ gets mapped to $d \in D$ by the channel with a probability of $q(c, d)$. Thus $\sum_{d \in D} q(c, d) = 1$ for all $c \in C$. We assume that $|C| \leq |D|$. This is normally the case. (Take for example the binary symmetric channel.)

Note that Shannon theory states that one cannot transmit more than $C(q(\cdot, \cdot))$ bits while achieving a vanishingly small probability of decoding error where C is the capacity of the channel [10]. But this result holds in the limit of large code blocklength. In our case we do not have the luxury of large blocklengths. Also as we are treating our problem as a partially observed control problem we do not always have to be concerned with error-less transmission. There is no clear separation between control and encoding. See Mitter and Borkar for discussion of this issue [8].

The controller receives the channel output with a fixed delay of $\Delta_1 \geq 0$ (Δ_1 an integer) time units. Having received a channel output $W_n \in D$ at time n, the controller selects an action Z_n in A and transmits it to the plant, where it is implemented with
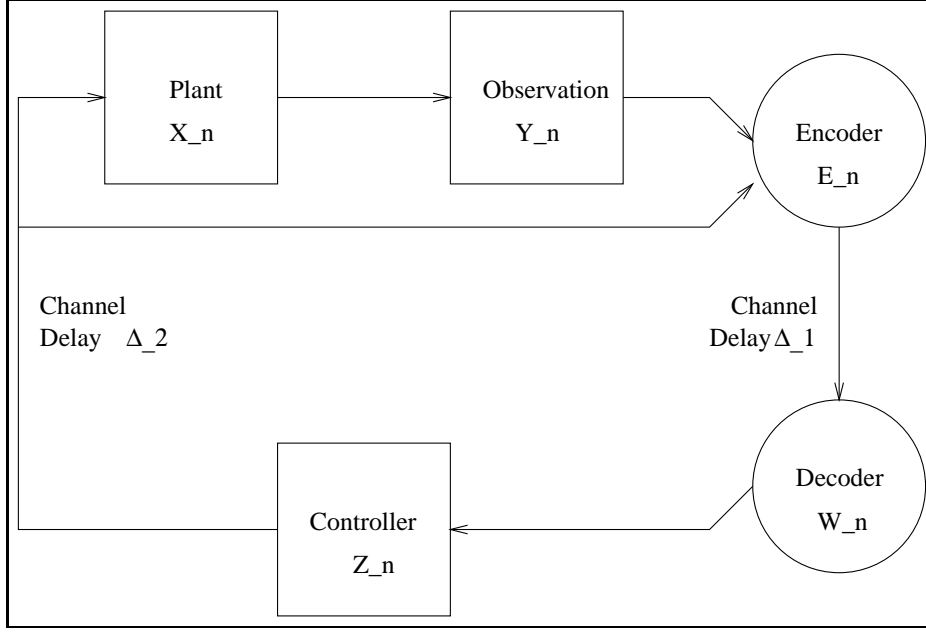
FIG. 4.1. *Closed Loop System*

a further delay of Δ_2 units, $\Delta_2 \geq 0$. Let $\Delta = \Delta_1 + \Delta_2$. For simplicity of analysis we assume this channel is noiseless. This assumption holds in the case when the controller has access to a high fidelity encoder. For example a base station can send messages to a satellite with relatively little distortion. Figure 4.1 gives a schematic drawing for the foregoing.

Let $\tilde{Z}_n \triangleq Z_{n-\Delta_2}$, the process $\{X_n, Y_n\}$ evolves by the Markov property as per

$$P(X_{n+1} = i, Y_{n+1} = j | X_m, Y_m, \tilde{Z}_m, \ m \leq n) \triangleq p(X_n, \tilde{Z}_n, i, j)$$

for $n \geq 0$, $i \in S$, $j \in B$. (Z_m for $-\Delta_2 \leq m < 0$ may be fixed according to some convention.) The controlled transition probability function $p : S \times A \times S \times B \rightarrow [0, 1]$ must satisfy $\sum_{i,j} p(k, u, i, j) = 1 \ \forall k, u$. For future reference, define $\bar{p}(i, u, j) = \sum_k p(i, u, j, k)$ for $i, j \in S$, $u \in A$.

Our control objective will be to minimize the total discounted cost

$$(4.1) \qquad\qquad E\left[\sum_{m=0}^{\infty} \beta^m \ g(X_m, Z_m)\right]$$

for a prescribed "running cost" function $g : S \times A \rightarrow \Re$ and a "discount factor" $\beta \in (0, 1)$. This is simpler to handle than the "average cost" of sections 2 and 3.

We shall consider four scenarios of increasing complexity:

  (i) <u>Static-Static</u>: Here the controller chooses Z_n at time n depending only on $\hat{Z}_n \triangleq [Z_{n-\Delta}, ..., Z_{n-1}]$ and W_n. The encoder chooses the encoding map $\Psi_n(\cdot)$ depending at most on $\tilde{Z}_m$, $m \leq n$ and Y_n.

  (ii) <u>Dynamic-Static</u>: Here the controller is as in (i), but the encoder solves an exact dynamic optimization problem given the controller behavior.

(iii) <u>Static-Dynamic</u>: Here the encoder is as in (i) but the controller solves an exact dynamic optimization problem given the encoder.

(iv) <u>Dynamic-Dynamic</u>: Here both the encoder and the controller solve exact dynamic optimization problems from their individual perspectives.

The next section elaborates on each of these four cases one by one. It should be underscored, however, that all methods involve an element of ad hocism and approximation. What one buys is a relatively simple structure of the encoder and/or the controller. This has to be justified on a case-by-case basis.

5. Schemes for Simultaneous Coding and Control.

5.1. Static-Static Case. In general the encoder and the controller have different information about the other's objectives. Because of this, it becomes necessary to state what information one has about the other.

Before taking up the general case, consider the special situation wherein $W_n = Y_{n-\Delta_1} = X_{n-\Delta_1}$ $\forall n$. That is, the controller receives the exact state information, albeit delayed. (This is the case where the encoder/decoder transmit the observation noiselessly). We solve this "delayed"-control problem by the usual technique of state augmentation. Let $\hat{X}_n \triangleq X_{n-\Delta_1}$ and recall that $\hat{Z}_n \triangleq [Z_{n-\Delta}, ..., Z_{n-1}]$. We argue below that the new state process $\{(\hat{X}_n, \hat{Z}_n)\}$ is an $S \times A^\Delta$-valued controlled Markov chain controlled by the A-valued control process $\{Z_n\}$. This, in fact, is what qualifies it as the "state process."

To see this, define for $i, j \in S, [u_1, ..., u_\Delta] \in A^\Delta$,

$$p_\Delta(i, u_1, .., u_\Delta, j) \triangleq \sum_{i_1,...,i_{\Delta-1}} \prod_{0 \le k \le \Delta-1} \bar{p}(i_k, u_{k+1}, i_{k+1}),$$

where $i_0 = i, i_\Delta = j$, to be the Δ-step transition probability under controls $u_1, .., u_\Delta$.

Then for $i, j \in S, \bar{u} = [u_1, .., u_\Delta]$, and $\bar{v} = [v_1, ..., v_\Delta] \in A^\Delta$, keeping in mind that $\Delta = \Delta_1 + \Delta_2$, we have

$$P(\hat{X}_{n+1} = j, \hat{Z}_{n+1} = \bar{v}|\hat{X}_m, \hat{Z}_m, Z_m, m < n, \hat{X}_n = i, \hat{Z}_n = \bar{u}, Z_n = u)$$
$$= P(\hat{X}_{n+1} = j, \hat{Z}_{n+1} = \bar{v}|\hat{X}_n = i, \hat{Z}_n = \bar{u}, Z_n = u)$$
$$= \begin{cases} 0 & \text{if } \bar{v} \ne [u_2, .., u_\Delta, u] \\ \bar{p}(i, u_1, j) & \text{otherwise.} \end{cases}$$

The "running cost" for this controlled Markov chain is defined as

$$\hat{g}(i, \bar{u}; u) = \beta^{\Delta_2} \sum_j p_\Delta(i, u_1, .., u_\Delta, j) g(j, u)$$

for $i \in S, \bar{u} = [u_1, ..., u_\Delta] \in A^\Delta, u \in A$. Note that the control action Z_n at time n depends directly on $X_{n-\Delta_1}$, so we must estimate the state X_n given that we know $X_{n-\Delta_1}$ and the controls. Also Z_n directly affects $X_{n+\Delta_2}$, hence the β^{Δ_2} term. Together, these lead to:

$$E[\hat{g}(\hat{X}_n, \hat{Z}_n; Z_n)] = \beta^{\Delta_2} E[E[g(X_{n+\Delta_2}, Z_n) \mid \hat{X}_n]]$$
$$= \beta^{\Delta_2} E[E[g(X_{n+\Delta_2}, Z_n) \mid \hat{X}_m, \hat{Z}_m, m \le n, Z_n]]$$

Therefore

$$E[\sum_{m=n}^{\infty} \beta^m g(X_m, \tilde{Z}_m)] = E[\sum_{m=n-\Delta_2}^{\infty} \beta^m \beta^{\Delta_2} g(X_{m+\Delta_2}, Z_m)]$$

$$= E[\sum_{m=n-\Delta_2}^{\infty} \beta^m \hat{g}(\hat{X}_m, \hat{Z}_m; Z_m)]$$

for $n \geq \Delta_2$. This establishes the equivalence of our original control problem with the problem for $(\hat{X}_n, \hat{Z}_n)$, $n \geq 0$, with running cost $\hat{g}(\cdot)$. We shall be using analogous arguments in subsequent sections to derive the corresponding "equivalent problems."

The "value function" $V : S \times A^{\Delta} \to \Re$ is defined by: $V(i, \bar{u})$ is the minimum cost when $\hat{X}_0 = i, \hat{Z}_0 = \bar{u}$. The standard dynamic programming arguments then show that it is a unique solution to the dynamic programming equations

$$(5.1) \qquad V(i, \bar{u}) = \min_u \left(\hat{g}(i, \bar{u}; u) + \beta \sum_j p(i, u_1, j) V(j, [u_2, ..., u_{\Delta}, u]) \right)$$

for $(i, \bar{u}) \in S \times A^{\Delta}$. Let $\eta(i, \bar{u})$ attain the minimum on the right hand side, any ties being resolved according to some fixed priority rule. Then $Z_n = \eta(\hat{X}_n, \hat{Z}_n)$ is the optimal control process. These claims follow by routine dynamic programming arguments.

In general, $W_n \neq \hat{X}_n$. A simple scheme then would be to use some conventional encoding-decoding scheme to arrive at an estimate $e(\hat{X}_n)$ of $\hat{X}_n$ and set $Z_n = \eta(e(\hat{X}_n), \hat{Z}_n)$ for each n. This can be expected to perform well if, for example, the following conditions hold: Let $|S| = |B| = |C| = |D|$, so that one can identify S, B, C, D via a suitable one-to-one correspondence. Suppose furthermore that this correspondence can be chosen such that $Y_n \approx X_n, W_n \approx Y_{n-\Delta_1}$ with high probability. That is, the serial channel $X_n \to Y_n \to \Psi_n(Y_n) \to W_{n+\Delta_1}$ maintains reasonable fidelity (where $E_n = \Psi_n(Y_n)$.) Under such circumstances, a traditional estimation scheme such as maximum likelihood should lead to an estimate $e(\hat{X}_n)$ "close" to $\hat{X}_n$ with high probability.

In general, however, $|B| < |S|$ and $|C| < |B|$. The former is because Y_n is usually a noisy measurement of some function of the state and not of the state itself. The latter implies an explicit data compression that is usually present. If the drop in cardinality is not too high in either inequality, the above "certainty equivalence" scheme may still be reasonable, but not otherwise.

If $Y_n \approx X_n$ with high probability (in other words, X_n is well-estimated given Y_n), but $|C|, |D| << |B|$ (i.e. a significant extent of data compression is present), one may modify the certainty equivalence schemes as follows: For each $i \in S$, let $v(i)$ denote the vector $[V(i, \bar{u})]$ as $\bar{u}$ varies over A^{Δ}, ordered (say) lexicographically. Use a good clustering algorithm to group $v(i)'s$ into disjoint clusters $C_1, ..., C_K$ where $K \approx |C|$. Given Y_n estimate X_n as (say) X'_n according to a suitable estimation scheme, such as maximum likelihood. If $v(X'_n) \in C_{k_n}$ (say), encode and transmit k_n. The controller then estimates $k_{n-\Delta_1}$ as (say) $\tilde{k}_n$ using the observation W_n, once again using some conventional estimation scheme such as maximum likelihood. He then picks a representative $i_n \in C_{\tilde{k}_n}$ and sets $Z_n = \eta(i_n, \hat{Z}_n)$.

There is yet another possibility that might be more appealing in some situations. Note that the function $\eta(\cdot, \cdot)$ can be precomputed and made available to the encoder. Suppose for the time being that $|C|, |D| \geq |A|$. Let $H(a) = \{(i, \bar{u})|\eta(i, \bar{u}) = a\}, a \in A$. Then $\{H(a)\}$ partitions $S \times A^{\Delta}$ into $|A|$ disjoint sets, tagged by the action to be taken when the "state" falls in the set in question. If the encoder knows $(\hat{X}_n, \hat{Z}_n)$ at time $n - \Delta_1$, all we have to do is to encode and transmit the tag $\hat{a}_n$ defined by: $(\hat{X}_n, \hat{Z}_n) \in H(\hat{a}_n)$ (i.e., $\hat{a}_n = \eta(\hat{X}_n, \hat{Z}_n)$). The controller receives at time n a noise-corrupted version of W_n of this from which he can estimate $\hat{a}_n$ by the maximum likelihood scheme (say) and set Z_n equal to this estimate. The catch of course is that at time $n - \Delta_1$ the encoder knows not $\hat{X}_n$, but $Y_{n-\Delta}$ and the first Δ_2 components of $\hat{Z}_n$, but not the last Δ_1 components thereof. Nevertheless, he can, in principle, estimate $X_{n-\Delta_1}$ and the last Δ_1 components of $\hat{Z}_n$ given the transition mechanism and the known behavior of the controller, and use these estimates in place of the actual values in the spirit of "certainty equivalence."

The problem is a little messier if, say, $|C|$ or $|D|$ is less than $|A|$. One then needs further aggregation in the control space so that $\hat{a}_n$ is picked from a proper subset of A. One possible schemes would be to plot for each $a \in A$ the vector $[\hat{g}(i, \bar{u}; a) + \beta \sum_j p(i, u_1, j)V(j, [u_2, ..., u_{\Delta}, a])]$ where $(i, \bar{u})$ varies over $S \times A^{\Delta}$, use a clustering algorithm to cluster them into a small number ($\approx \min(|C|, |D|)$) of clusters and let $\hat{a}_n$ be an encoding of the cluster index. The controller, having estimated $\hat{a}_n$, picks up a prespecified representative element of A corresponding to that cluster, as his choice of Z_n. See [12] for a summary of clustering algorithms.

While all these proposals remain a little ad hoc, they have a clear-cut common message: in control with communication constraints, it makes sense to replace the traditional information theoretic encoding criteria by criteria that explicitly take into account the control aspects.

5.2. Dynamic-Static Case. In this scenario, we assume that the controller uses a control of the form $Z_n = \psi(W_n, \hat{Z}_n)$ where $\psi : D \times A^{\Delta} \to A$ is a prescribed map. This map might be arrived at by one of the several methods described in the preceding subsection or in some other way, that is not our focus here. We shall be concerned with corresponding exact dynamic programming optimization problem faced by the encoder.

The encoder observes at time n, Y_n and $\tilde{Z}_n$. He then faces a partially observed control problem. Let E_n be the codeword transmitted at time n. Then

$$\tilde{Z}_{n+1} = Z_{n+1-\Delta_2} = \psi(W_{n+1-\Delta_2}, \hat{Z}_{n+1-\Delta_2}).$$

Then the conditional law of $\tilde{Z}_{n+1}$ given $\mathcal{F}_n \triangleq \sigma(Y_m, E_m, \tilde{Z}_m, \; m \leq n)$ is $r_n^o(\cdot)$, given by: for $a \in A$, $r_n^o(a) = \sum q(E_{n+1-\Delta}, j)$, where the sum is over all j such that $\psi(j, \hat{Z}_{n+1-\Delta_2}) = a$. More generally, for $k < \Delta$ the conditional law of $\tilde{Z}_{n+k}$ given $\mathcal{F}_n$ and $\tilde{Z}_{n+1} = a_1, ..., \tilde{Z}_{n+k-1} = a_{k-1}$, is $r_n^{a_1,...,a_{k-1}}(\cdot)$, given by

$$r_n^{a_1,...,a_k}(a) = \sum q(E_{n+k-\Delta}, j)$$

where the sum is over all j such that $\psi(j, \hat{Z}_{n+k-\Delta_2}) = a$, with

$$\hat{Z}_{n+k-\Delta_2} = [Z_{n+k-\Delta_2-\Delta}, ..., Z_{n+k-\Delta_2-1}]$$
$$= [\tilde{Z}_{n+k-\Delta}, ..., \tilde{Z}_{n+k-1}]$$
$$= [\tilde{Z}_{n+k-\Delta}, ..., \tilde{Z}_n, a_1, .., a_{k-1}].$$

Thus the joint conditional law of $[\tilde{Z}_{n+1}, ..., \tilde{Z}_{n+\Delta-1}]$ given $\mathcal{F}_n$ is $r_n(\cdot, ..., \cdot)$ given by:

$$r_n(a_1, ..., a_{\Delta-1}) = r_n^o(a_1) r_n^{a_1}(a_2) r_n^{a_1, a_2}(a_3) \cdots r_n^{a_1, ..., a_{\Delta-2}}(a_{\Delta-1}).$$

The derivation shows that $r_n(\cdot)$ is a fixed function of

$$E_n^* = [E_{n-1}, ..., E_{n-\Delta}] \text{ and } Z_n^* = [\tilde{Z}_n, ..., \tilde{Z}_{n-\Delta+1}].$$

Therefore we may write $r_n(\cdot) \triangleq R(E_n^*, Z_n^*)$ for a map $R : C^\Delta \times A^{\Delta-1} \to \mathcal{P}(A^\Delta)$ ($\triangleq$ probability vectors on A^Δ) defined via the foregoing. Analogously, we set $r_n^o \triangleq R^o(E_n^*, Z_n^*)$ for a suitable map $R^o : C^\Delta \times A^\Delta \to \mathcal{P}(A)$ defined via the foregoing. (Then $R^o(\bar{e}, \bar{z})$ is the image of $R(\bar{e}, \bar{z})$ under the projection $A^\Delta = A \times A^{\Delta-1} \to A$.)

Let π_n denote the conditional law of X_n given $\mathcal{F}_n, n \geq 0$. Write these as row vectors. Then $\pi_n \in \mathcal{P}(S) \triangleq$ the space of probability (row) vectors on S. Then standard arguments using Bayes rule (see, e.g., [18]) show that $\{\pi_n\}$ satisfies the recursion given by the nonlinear filter

$$(5.2) \qquad \pi_{n+1} = \pi_n P(Y_{n+1}, \tilde{Z}_n)/(\pi_n P(Y_{n+1}, \tilde{Z}_n) 1_c), n \geq 0$$

where $P(y, z)$ is the $|S| \times |S|$ matrix whose (i, j)-th element is $p(i, z, j, y)$ and 1_c is the $|S|$-dimensional column vector of all 1's.

Let $A^\Delta = A^{\Delta_1} \times A^{\Delta_2}$ and let $\Theta_1 : A^\Delta \to A^{\Delta_1}$, $\Theta_2 : A^\Delta \to A^{\Delta_2}$, denote the projection maps.

Now "control" E_n effects $X_{n+\Delta}$ and $\tilde{Z}_{n+\Delta}$. Then

$$E[g(X_{n+\Delta}, \tilde{Z}_{n+\Delta})|\mathcal{F}_n]$$
$$= E[g(X_{n+\Delta}, Z_{n+\Delta_1})|\mathcal{F}_n]$$
$$= E[g(X_{n+\Delta}, \psi(W_{n+\Delta_1}, \hat{Z}_{n+\Delta_1}))|\mathcal{F}_n]$$
$$= E[g(X_{n+\Delta}, \psi(W_{n+\Delta_1}, [\tilde{Z}_{n-\Delta_2}, ..., \tilde{Z}_{n+\Delta_1-1}]))|\mathcal{F}_n]$$
$$(5.3) \quad = \sum_l \sum_i \sum_{\bar{a} \in A^\Delta} R(E_n^*, Z_n^*)(\bar{a}) \pi_n(i) p_\Delta(i, \bar{a}, j) q(E_n, l) g(i, \psi(l, [\Theta_2(Z_n^*) : \Theta_1(\bar{a})]))$$

In view of this, we may define

$$g^*((\pi_n, E_n^*, Z_n^*), E_n) \triangleq \beta^\Delta E[g(X_{n+\Delta}, \tilde{Z}_{n+\Delta})|\mathcal{F}_n]$$

for a suitable continuous map $g^* : (\mathcal{P}(S) \times C^\Delta \times A^\Delta) \times C \to \Re$ defined via (5.3). Then

$$E[\sum_{m=n}^{\infty} \beta^m g(X_m, \tilde{Z}_m)] = E[\sum_{m=n-\Delta}^{\infty} \beta^m E[\beta^\Delta g(X_{m+\Delta}, \tilde{Z}_{m+\Delta}) \mid \mathcal{F}_m]]$$

$$= E[\sum_{m=n-\Delta}^{\infty} \beta^m g^*((\pi_m, E_m^*, Z_m^*), E_m)]$$

for $n \geq 0$, which establishes the equivalence of our original problem with that of controlling $\{(\pi_n, E_n^*, Z_n^*), n \geq 0\}$ with running cost function $g^*(\cdot)$.

The partially observed control problem for the encoder can now be stated as follows: the "state" process is $\{(\pi_n, E_n^*, Z_n^*), n \geq 0\}$, taking values in the state space $\mathcal{P}(S) \times C^\Delta \times A^\Delta$. The "control process" is $\{E_n\}$ taking values in C. The cost to be minimized is

$$(5.4) \qquad E[\sum_{n=0}^{\infty} \beta^n g^*((\pi_n, E_n^*, Z_n^*), U_n)].$$

One can write down the messy probabilistic evolution law for this stochastic control system, but we shall skip it since one can easily write down the dynamic programming equations directly. Let $V(\pi, \bar{e}, \bar{z})$ denote the "value function" evaluated at $(\pi, \bar{e}, \bar{z}) \in \mathcal{P}(S) \times C^\Delta \times A^\Delta$, i.e., it is the infimum of (5.4) over all admissible $\{E_n\}$ when $\pi_0 = \pi$, $E_0^* = \bar{e} = [e_1, ..., e_\Delta]$, $Z_0^* = \bar{z} = [z_1, ..., z_\Delta]$. Standard dynamic programming arguments lead to the dynamic programming equations

$$V(\pi, \bar{e}, \bar{z}) = \min_c (g^*((\pi, \bar{e}, \bar{z}), c) + \beta \sum_{i,j,y,a} \pi(i) R^o(\bar{e}, \bar{z})(a) p(i, z_1, j, y)$$
$$(5.5) \qquad \times V(\pi P(y, z_1)/(\pi P(y, z_1) 1_c), [c, e_1, ..., e_{\Delta-1}], [a, z_1, ..., z_{\Delta-1}]))$$

Suppose the minimum on the right hand side is attained at $\mathcal{V}(\pi, \bar{e}, \bar{z})$ for a suitable defined $\mathcal{V} : \mathcal{P}(S) \times C^\Delta \times A^\Delta \to A$. Then $E_n = \mathcal{V}(\pi_n, E_n^*, Z_n^*)$ is the optimal coding policy for the encoder, by standard dynamic programming arguments.

In fact, dynamic programming arguments show that this is a complete characterization of optimal coding policies that are "stationary", i.e., depend only on the current "state" (π_n, E_n^*, Z_n^*). The problem may thus seem to have been "solved", but only to the extent that we ignore the complexity issues altogether. Solving (5.5) can be an ominous task, warranting a good approximation scheme. We shall address these issues later.

5.3. Static-Dynamic Case. Here we hypothesize a static encoding rule given by $E_n = \Psi(Y_n, Z_n^*), n \geq 0$ for a prescribed map $\Psi : B \times A^\Delta \to C$ known to the controller. Let $\mathcal{G}_n = \sigma(W_m, Z_m | m \leq n)$ and $\tilde{\pi}_n$ the conditional law of $\hat{X}_n$ given $\mathcal{G}_n$ viewed as a (random) element of $\mathcal{P}(S)$ and written as a row vector. Using Bayes rule, one can write down the recursion for $\{\tilde{\pi}_n\}$ as

$$(5.6) \qquad \tilde{\pi}_{n+1} = \tilde{\pi}_n \tilde{P}(W_n, Z_n', Z_{n-\Delta})/(\tilde{\pi}_n \tilde{P}(W_n, Z_n', Z_{n-\Delta}) 1_c)$$

for $n \geq 0$, where $\tilde{P}(w, \bar{z}, z)$ is an $|S| \times |S|$ matrix whose (i, j)-th element is

$$\sum_y p(i, z, j, y) q(\Psi(y, \bar{z}), w),$$

and $Z_n' \triangleq Z_{n-\Delta_1}^*$. Note also that

$$(5.7) \qquad \beta^{\Delta_2} E[g(X_{n+\Delta_2}, Z_n) | \mathcal{G}_n] = \beta^{\Delta_2} \sum_{i,j} \tilde{\pi}_n(i) p_\Delta(i, Z_{n-\Delta}, ..., Z_{n-1}, j) g(j, Z_n)$$

Define $\bar{Z}_n = [Z_{n-\Delta-\Delta_1}, ..., Z_{n-1}] \in A^{\Delta+\Delta_1}$. Let Λ_1, Λ_2 be the projections from $A^{\Delta+\Delta_1}$ to its first (respectively last) Δ components. The state space for the partially observed control problem of the controller is now $\mathcal{P}(S) \times A^{\Delta+\Delta_1}$, the state process being $(\tilde{\pi}_n, \bar{Z}_n), n \geq 0$. The cost to be minimized over the control process $\{Z_n\}$ is

$$(5.8) \qquad E\left[\sum_{n=0}^{\infty} \beta^n \bar{g}((\tilde{\pi}_n, \Lambda_2(\bar{Z}_n)); Z_n)\right]$$

where $\bar{g}((\tilde{\pi}_n, \Lambda_2(\bar{Z}_n)); Z_n)$ is the right hand side of (5.7). Note that the "control" Z_n affects $X_{n+\Delta_2+1}$ and Z_n is affected by E_n which is a function of $Z_{n-\Delta-\Delta_2+1}$, ..., $Z_{n-\Delta_2}$. Now,

$$E[\sum_{m=n}^{\infty} \beta^m g(X_m, \tilde{Z}_m)] = E[\sum_{m=n-\Delta_2}^{\infty} \beta^m \beta^{\Delta_2} E[g(X_{m+\Delta_2}, Z_m) \mid \mathcal{G}_n]]$$

$$= E[\sum_{m=n-\Delta_2}^{\infty} \beta^m \bar{g}((\tilde{\pi}_m, \Lambda_2(\bar{Z}_m)); Z_m)]$$

for $n \geq 0$. Thus our original problem is equivalent to that of controlling $(\tilde{\pi}_n, \bar{Z}_n), n \geq 0$, with (5.8) as the cost.

Let $V(\pi, \bar{z})$ denote the value function evaluated at $\pi \in \mathcal{P}(S), \bar{z} = [z_1, ..., z_{\Delta+\Delta_1}] \in A^{\Delta+\Delta_1}$, i.e., the infimum of (5.8) over all admissible $\{Z_n\}$ when $\tilde{\pi}_0 = \pi, \bar{Z}_0 = \bar{z}$. Then standard dynamic programming arguments lead to:

$$V(\pi, \bar{z}) = \min_z [\bar{g}(\pi, \Lambda_2(\bar{z}); z) + \beta \sum_{i,j,y,w} \pi(i) p(i, z_{\Delta_1}, j, y) q(\Psi(y, \Lambda_1(\bar{z})), w)$$

$$(5.9) \qquad \times V(\pi \tilde{P}(w, \Lambda_1(\bar{z}), z_{\Delta_1})/(\pi \tilde{P}(w, \Lambda_1(\bar{z}), z_{\Delta_1}) 1_c), [z_2, ..., z_{\Delta+\Delta_1}, z])]$$

Once again, an "optimal stationary policy," $Z_n = \Phi(\pi_n, \bar{Z}_n), n \geq 0$, can be derived by (and characterized by): $\Phi(\pi, \bar{z}) \in$ Argmin (right hand side of (5.9)).

Once again we have a potentially "computationally" hard dynamic program. But even that seems a luxury when we face squarely the problem in its full generality in the next subsection.

We conclude this subsection with a remark that applies both to this and the previous subsection. Both present complex dynamic programs for which one needs good approximation schemes. There is a considerable literature on approximations of dynamic programs, notably the recent developments in simulation-based approximations. (See [3] and the references therein.) These possibilities need to be explored in this context, using to advantage all prior knowledge about the plant structure.

5.4. Dynamic-Dynamic Case. Now we allow both the encoder and the controller to use at each instant all information available to them. A precise formulation follows.

Let $\mathcal{B}_n = \sigma(Y_m, E_m, Z_m, W_m, m \leq n)$, and redefine π_n as the conditional law of X_n given $\mathcal{B}_n$, written as a row vector. Then $\{\pi_n\}$ is given recursively by

$$(5.10) \qquad \pi_{n+1} = \pi_n P^*(Y_n, E_n, W_n, \tilde{Z}_n)/(\pi_n P^*(Y_n, E_n, W_n, \tilde{Z}_n) 1_c),$$

where $P^*(y, e, w, z)$ is the $|S| \times |S|$ matrix whose (i, j)-th element is $p(i, z, j, y)q(e, w)$. A little thought shows that the term $q(E_n, W_n)$ cancels between the numerator and the denominator on the right hand side of (5.10) and thus (5.10) reduces to (5.2). The cost is now given by

$$E \left[\sum_{n=0}^{\infty} \beta^n g^{'} \left(\pi_n, \tilde{Z}_n \right) \right]$$

where $g^{'}(\pi, e) = \sum_i \pi(i)g(i, e) \ \forall \pi, u$. This is to be minimized overall $\{E_n\}, \{Z_n\}$ satisfying:

(i) $\forall n$, E_n is conditionally independent of $X_m, m \le n$, given $Y_m, \tilde{Z}_m, m \le n$,

(ii) $\forall n$, Z_n is conditionally independent of $X_m, Y_m, E_m, W_m, m \le n + \Delta_2$, given $W_m, m \le n$.

This can be converted to an apparently (but not really) more amenable problem by changing the choice of "state" as well as the underlying probability measure. Define a process $\{\nu_n\}$ of "unnormalized conditional laws" taking values in $\mathcal{M}(S) =$ the space of nonnegative measures on S (written as row vectors in $\Re^{|S|}$ indexed by S) by: $\nu_0 = \pi_0$ and

(5.11) $$\nu_{n+1} = |B||D|\nu_n P^*(Y_n, E_n, W_n, \tilde{Z}_n)$$

Let $(\Omega, \mathcal{F}, \mathcal{P})$ be the underlying probability space. Define on $(\Omega, \mathcal{F})$ a new probability measure P_0 according to $\frac{dP}{dP_0}|_{\mathcal{G}_n} = \nu_n(S), \ \forall n$. (For what follows, there is no loss of generality in supposing that $\mathcal{F} = \bigvee_n \mathcal{G}_n$, though this obviously need not be true.) Under P_0, standard arguments ([5]) shows that $\{(Y_n, W_n)\}$ are independent and $\{Y_n\}, \{W_n\}$ I.I.D., uniformly distributed on B, D respectively. Clearly, $\pi_n = \nu_n/\nu_n(S)$, whence the term "unnormalized conditional law." The attractive feature of (5.10) is that is a "linear" control systems with I.I.D. "noise" $\{(W_n, Y_n)\}$. (The control and noise do not, however, enter "linearly".) The control process $\{E_n\}, \{Z_n\}$ have to satisfy (i) and (ii) above.

This is an instance of a "decentralized control problem" discussed variously in [27] and [22] among other places, with the encoder and the controller operating as two distinct controllers acting on their own information. Such problems are known to be computationally hard [20]. Another important feature of these problems, as emphasized in [22], is the fact that each control here has to play two additional roles: as a "probe" to learn the other controllers' control policy and as a "signaling device" to share information with the other controller, both within the constraints of the computational limitations. Quantifying these aspect has proved elusive. But at any rate, one thing is clear – because some online "learning" by the controllers will be necessarily involved, one can expect optimal policies, if any, to be non-stationary, i.e., exhibiting explicit dependence on the time variable. Within such policies, it is easy to show the existence of optimal ones by using the techniques of [4]. ([4] considers a finite horizon control problem, but the proofs adapt easily to the infinite horizon discounted cost problem.) But this is of little help, as it leaves open the issue of computing these policies. Since we are now allowing policies that depend on the entire past observations of the encoder/controller respectively, the formalism of the

preceding subsections can no longer be used to freeze the policy of one and find the corresponding best policy of the other.

With this in mind, we propose below an ad hoc approximation scheme, with some heuristic justification. We postulate that $\{E_n\}$, $\{Z_n\}$ are generated by mechanisms of the following form:

$$
\begin{aligned}
e_{n+1} = F(e_n, e_{n-1}, ..., e_{n-n_1}; E_n, E_{n-1}, ..., E_{n-n_2}; Y_n, Y_{n-1}, ..., Y_{n-n_3}, \\
\tilde{Z}_n, \tilde{Z}_{n-1}, ..., \tilde{Z}_{n-n_4}),
\end{aligned}
$$
(5.12)

$$
E_{n+1} = f(e_{n+1}),
$$
(5.13)

$$
z_{n+1} = G(z_n, z_{n-1}, ..., z_{n-n_1}; Z_n, Z_{n-1}, ..., Z_{n-n_2}; W_n, W_{n-1}, ..., W_{n-n_3}),
$$
(5.14)

$$
Z_{n+1} = g(z_{n+1})
$$
(5.15)

where $\{e_n\}$, $\{z_n\}$ take values in appropriate finite alphabets, not necessarily the same.

To motivate this, consider the classical LQG problem. The feedback loop in LQG control is traditionally viewed as consisting of three units: the "plant" whose outputs (observations) is fed to the "estimator" (the Kalman filter) whose output (the state estimate) is fed to the "controller" whose output (the "certainty equivalence" control) is fed back to the plant. But one may also view the estimator and the controller as a single unit, a "dynamic controller" quite similar in structure to the plant itself. This dynamic controller has the state estimate as its own internal "state", the certainty equivalent control is its "output", and the linear feedback law that maps the former to the latter its "state-to-output" map.

In general, in a partially observed control system, one can expect a "dynamic controller", the dynamic aspect incorporating the state estimation, parameter tracking/estimation, signaling, and learning the controller's decisions in a distributed environment, etc. What we propose above can then be considered a finite state approximation to the unfathomable "ideal" dynamic controller.

This still leaves open the issue of how to choose (5.12)-(5.15). We propose that n_3, n_4 equal Δ while n_1, n_2 equal $\Delta + \Delta_1$. This is purely heuristic, based on the analysis of the preceding subsections. The maps F, f, G, g can be taken from suitable parameterized families, such as feedforward neural nets. The "choice" then reduces to a pure parametric (as opposed to dynamic) optimization. This can be done using a simulation-based algorithm such as Infinitesimal Perturbation Analysis (IPA) [17].

6. Conclusions. There are many more issues in control under communication constraints that we did not touch. We conclude by outlining some of them.

We have assumed that the encoder and the controller operate synchronously, i.e., on a common clock, and "act" at each tick of the clock. If they do no have a common clock and operate asynchronously, the analysis becomes very hard. Even in the synchronous case, they may not act at each tick of the clock. This could be due to block coding by the encoder, reset times for the controller, or a conscious decision

to act at fixed epochs or a combination of these. Once again, if these epochs are either random or fixed but incommensurate (i.e., not rationally related), the analysis becomes messy.

Also, we have viewed transmission of a codeword as a one shot affair. In principle, when an encoding is received in real time, each new bit conveys some information for which one does not have to wait till the end of transmission. This is particularly so for tree-based codes. Taking this into account, though better in principle, will add to the complexity of the control schemes. The choice of "epochs" can also be an additional decision that forms a part of the control loop.

By virtue of confining ourselves to a finite state set-up, we also avoided stability issues. It is of interest to see if traditional stability/stabilizablity analysis are robust to communication constraints in general.

Finally, in the dynamic-dynamic case, we have implicitly assumed "decentralized planning" along with distinct information structures at the encoder and the controller ends, as is customary. In other words these two decision makers not only have different information σ-fields, but also arrive at their policies separately. An important intermediate possibility overlooked in the literature is to have different information structures, but "centralized planning," which often makes the problems tractable in principle. See [24], chapter two, for a formal statement of this principle and related issues.

REFERENCES

[1] E. ALTMAN, *Constrained Markov Decision Processes,* Chapman and Hall, Bocca Raton, 1999.

[2] D. BERTSEKAS, *Dynamic Programming and Optimal Control,* Vol. 1 and 2, Athena Scientific, Belmont, MA, 1995.

[3] D. BERTSEKAS AND J. TSITSIKLIS, *Neurodynamic Programming,* Athena Scientific, Belmont, MA, 1996.

[4] V. BORKAR, *N-person Noncooperative Stochastic Games with Partial Information,* Proceedings of International Conference on Game Theory, New Delhi, 1990, Springer Lecture Notes in Economics, No. 369, 1992.

[5] V. BORKAR, *Topics in Controlled Markov Chains,* Pitman Research Notes in Mathematics 240, Longman Scientific and Technical, Harlow, England, 1991.

[6] V. BORKAR, *Ergodic Control of Markov Chains with Constraints - the General Case,* SIAM Journal of Control and Optimization, 32:1(1994) pp. 176–186.

[7] V. BORKAR, *Probability theory: An Advanced Course,* Springer Verlag, New York, 1995.

[8] V. BORKAR AND S. MITTER, *LQG Control with Communication Constraints,* in Communications, Computation, Control, and Signal Processing: A Tribute to Thomas Kailath, A. Paulraj, V. Roychowdhury, C. Schaper (eds.), Kluwer Academic Publishers, Norwell, Massachusetts, pp. 365–373, 1997.

[9] V. BORKAR, S. MITTER, AND S. TATIKONDA, *Optimal Sequential Vector Quantization of Markov Sources.* Submitted to SIAM Journal on Control and Optimization.

[10] T. COVER AND J. THOMAS, *Elements of Information Theory,* John Wiley, New York, 1991.

[11] D. DELCHAMPS, *Stabilizing a Linear System with Quantized State Feedback.* IEEE-AC, 35:8(1900), pp. 916–924.

[12] L. DEVROYE, L. GYORFI, AND G. LUGOSI, *A Probabilistic Theory of Pattern Recognition,* Springer-Verlag, New York, 1996.

[13] X. FENG AND K. LOPARO, *Active Probing for Information in Control Systems with Quantized State Measurements: A Minimum Entropy Approach.* IEEE-AC, 42:2(1997), pp. 216-238.

[14] X. FENG, K. LOPARO, AND Y. FANG, *Optimal State Estimation for Stochastic Systems: An Information Theoretic Approach.* IEEE-AC, 42:6(1997), pp. 771–785.

[15] G. GABOR AND Z. GYORFI, *Recursive Source Coding*, Springer Verlag, New York, 1986.

[16] A. GERSHO AND R. M. GRAY, *Vector Quantization and Signal Compression*. Boston, MA: Kluwer, 1992.

[17] Y-C HO AND X. CAO, *Perturbation Analysis of Discrete Event Systems*, Birkhauser, Boston, MA, 1991.

[18] P. R. KUMAR AND P. VARAIYA, *Stochastic Systems: Estimation, Identification, and Adaptive Control*, Prentice Hall, Inc., Englewood Cliffs, NJ, 1986.

[19] G. NAIR AND R. EVANS, *State Estimation Under Bit-Rate Constraints*, Proceedings of the 37th IEEE Conference on Decision and Control, Vol. 1, pp. 251 - 256, 1998.

[20] C. H. PAPADIMITRIOU AND J. N. TSITSIKLIS, *Intractable Problems in Control Theory*, SIAM Journal on Control and Optimization, 24:4(1986), pp. 639–654.

[21] W. RUNGGALDIER AND L. STETTNER, *Approximations of Discrete Time Partially Observed Control Problems*, Applied Mathematics Monographs, No. 6, Giardini Editori E Stampatori in Pisa, October 1994.

[22] N. R. SANDELL, P. VARAIYA, M. ATHANS, AND M. SAFANOV, *Survey of Decentralized Control Methods for Large Scale Systems*, IEEE-AC, 23:2(1978), pp. 108–128.

[23] L. STETTNER, *Ergodic Control of Partially observed Markov Processes with Equivalent Transition Probabilities*, Applicationes Mathematicae, 22:1(1993), pp. 25–38.

[24] S. TATIKONDA, *Control Under Communication Constraints*, MIT, LIDS PH.D. Thesis, September, 2000.

[25] S. TATIKONDA, A. SAHAI, AND S. MITTER, *Control of LQG Systems Under Communication Constraints*, LIDS working paper, M.I.T., May, 1998.

[26] J. WALRAND AND P. VARAIYA, *Optimal causal coding-decoding problems*, IEEE-IT, 29:6(1983), pp. 814–820.

[27] H. S. WITSENHAUSEN, *Separation of Estimation and Control for Discrete Time Systems*, Proceedings of the IEEE, 59:11(1971), pp. 1557–1566.

[28] W. WONG AND R. BROCKETT, SYSTEMS WITH FINITE COMMUNICATION BANDWIDTH CONSTRAINTS-PART I: STATE ESTIMATION PROBLEMS, IEEE-AC, 42:9(1997).

[29] W. WONG AND R. BROCKETT, *Systems with Finite Communication Bandwidth Constraints–II: Stabilization with Limited Information Feedback*, IEEE-AC, 44:5(1999), pp. 1049–1052.