MULTILAYER CONTROL OF LARGE MARKOV CHAINS[1]

Jean-Pierre Forestier[2,3] and Pravin Varaiya[3]

ABSTRACT

The computational burden associated with controlling a plant modelled as a Markov chain with a large number of states is addressed by proposing a two-layer feedback control structure. At the lower layer a regulator continuously monitors the plant. When the state of the plant reaches an extreme value, the supervisor at the higher layer intervenes to reset the regulator. It is shown that the plant dynamics and cost originally defined at the lower layer can be "lifted" to the supervisor layer and that the supervisor's control task can be defined in a way that permits wide flexibility in the design of the regulator.

---

[1]Research sponsored in part by a grant from IRIA (Institut de Recherche en Informatique et Automatique) and by the National Science Foundation Grant ENG76-16816. The results given here were announced at the IFAC Workshop on Control and Management of Integrated Industrial Complexes, Toulouse, 1977.

[2]C.E.R.T./D.E.R.A., 2 Av. E. Belin, 31400 Toulouse, France.

[3]Department of Electrical Engineering and Computer Sciences and the Electronics Research Laboratory, University of California, Berkeley, California 94720.

I. INTRODUCTION

We have in mind the situation of a plant being continuously controlled by a local regulator. Once in a while the "parameters" of the regulator are "reset" by a "supervisor". This can happen in at least two different ways: perhaps some components internal to the plant are malfunctioning and so the supervisor has to carry out some repairs (in this case the regulator description includes that of the plant), or there is a change in the external environment and the supervisor intervenes and resets the regulator to alter the plant's operating condition. We represent the state of the plant as well as of the relevant environment by s_t, $t = 0,1,\ldots$. s_t takes values in a finite set $S = \{1,\ldots,s\}$.

A control structure of this kind, symbolically drawn in Figure 1, is called a two-layer structure [1] in contrast to a two-level hierarchical structure. In the former the determination of control is split into algorithms which operate at different time scales whereas in the latter there is a "spatial" division into algorithms operating at the same time scale.

While there is a voluminous literature dealing with multilevel structures [2], little effort has been devoted to the study of multilayer control structures although casual observation suggests that in practice it is widely adopted in the control of large processes. The reason for this lack of attention seems in part due to the difficulty in satisfactorily exhibiting within a single problem formulation three essential features of a multilayer structure:

(i) the supervisor must intervene less frequently than the regulator (in terms of Fig. 1 $\tau >> 1$);

(ii) the supervisor must use less information than the regulator; and

(iii) the supervisor must solve a "higher" level problem.

A satisfactory formulation should have as a consequence that

(iv) the system performance improves as the supervisor intervenes more frequently or receives more information.

Chong and Athans [3] consider an interconnected linear system

$$\dot{x}_i = A_{ii}x_i + B_{ii}u_i + \sum_{j\neq i} [A_{ij}x_j + B_{ij}u_j] + \xi_i,$$

$$y_i = H_i x_i + \eta_i, \qquad i = 1,\ldots,k. \tag{1}$$

Here the subscript i refers to the ith subsystem and the notation is standard. The cost is a quadratic function. They propose a two-layer structure in which at the lower layer the ith local regulator chooses a linear feedback law assuming the model

$$\dot{x}_i = A_{ii}x_i + B_{ii}u_i + v_i + \xi_i,$$

$$y_i = H_i x_i + \eta_i,$$

where v_i is the "prediction" of the neglected interaction terms received from the upper layer supervisor. At time 0 the supervisor chooses $v(t) = (v_1(t),\ldots,v_k(t))$ for $0 \leq t \leq T$, at time T it chooses v(t) for $T \leq t \leq 2T$ based upon all the information available up to T, and so on. Thus feature (ii) mentioned above is absent. A variety of "reasonable" ways for making the supervisor choice is possible each of which can be intuitively rationalized [4], but there is no theoretically attractive formulation of the supervisor's higher level problem. Finally, it has not been possible to prove that the performance improves if the time T between successive supervisor interventions is reduced.

Earlier Donoghue and Lefkowitz [5] had considered a static optimization problem with a similar two-layer structure in which the supervisor intervened periodically and had full information. One of the variables under the supervisor's control was the frequency with which its intervention was carried out.

Periodic intervention is appropriate if the lower layer represents a production cycle of fixed duration and the supervisor intervenes at a fixed stage of each cycle as in [5]. It is inappropriate if the lower layer is a continuous or a periodic process. In the model presented here the supervisor intervenes only when the state reaches some "extreme" or "boundary" value. Precisely, we assume given a fixed subset B of S such that the supervisor intervenes only at those instances τ when s_τ is in B. Furthermore the supervisor observes the state only at these instances. Thus the moments of intervention are randomly spaced and are determined "intrinsically" by the plant and environment rather than being arbitrarily preselected. (Of course periodic intervention is a special case.[1]) Finally, we pay little attention to the way in which the lower layer regulation is carried out, and concentrate mainly on the supervisor's actions. This permits very different design procedures to be employed for the two layers.

We formulate the supervisor's problem in the next section. This requires "lifting" the system dynamics and the cost function defined originally at the lower layer to the supervisor's layer. This lifting is

---

[1] To see this suppose the period is N. Consider the process $\hat{s}_t = (s_t, t \bmod N) \in S \times \{0,1,\ldots,N-1\}$ and take $B = S \times \{0\}$. Then $\hat{s}_\tau^t \in B$ for $\tau = 0,N,2N,\ldots$.

examined more abstractly in Section III. We return to the supervisor's problem in Section IV. In Section V we give optimality conditions for the supervisor's problem and propose an algorithm for finding the optimal supervisor strategy.

II. THE SUPERVISOR'S PROBLEM

The lower layer process is denoted s_t, $t = 0,1,\ldots$ with values in $S = \{1,\ldots s\}$. If $s_t = i$ the regulator can choose any control u_t from $U(i)$, a prespecified set. A (stationary) regulator strategy is any element $u = (u(1),\ldots,u(s)) \in U = U(1)x,\ldots,U(s)$. For each u the process s_t is a Markov chain with (stationary) transition probability matrix $P(u) = \{P_{ij}(u)\}$ where

$$P_{ij}(u) = P_{ij}(u(i)) = \text{Prob}\{s_{t+1} = j | s_t = i,\ u_t = u(i)\}. \qquad (2)$$

Note that the ith row $(P_{i1},\ldots,P_{is})$ depends only on $u(i)$.

The cost associated with the regulator strategy u is

$$J(u) = \lim_{T\to\infty} \frac{1}{T+1} E \sum_{t=0}^{T} k(s_t,u(s_t)) \qquad (3)$$

where $k(i,u(i))$ is the prespecified "instantaneous" cost defined for i in S and $u(i)$ in $U(i)$.

To make (3) meaningful we impose the following assumption.

Strong ergodicity assumption. For each u the chain s_t has a single ergodic class consisting of all the states.

An equivalent assumption is that for each u there is a unique row vector, called the steady state probability distribution, $\pi(u) = (\pi_1(u),\ldots,\pi_s(u))$ with $\pi_i(u) > 0$ for all i, such that

$$\pi(u) = \pi(u)P(u) \text{ and } \pi(u)\underset{\sim}{1} = 1, \qquad (4)$$

where $\underset{\sim}{1} = (1,\ldots,1)'$. (This assumption can be considerably weakened as in [10].) Under this assumption (3) can be rewritten more simply as

$$J(u) = \pi(u)k(u) \tag{5}$$

where $k(u) = (k(1,u(1)),\ldots,k(s,u(s)))'$.

We now formulate the supervisor's observation and decision processes. A distinguished subset $B = \{1,\ldots,b\} \subset S$ is preselected; states in B are called <u>boundary</u> states. We assume that $s_0 = b_0 \in B$ and let $0 \equiv T_0 < T_1 < \ldots < T_n$ be the random times at which s_t is in B i.e.

$$T_{n+1}(\omega) = \min\{t > T_n(\omega) \,|\, s_t(\omega) \in B\}. \tag{6}$$

Here ω denotes the sample path. The supervisor's observation process is $b_0,b_1,\ldots,b_n,\ldots$ where

$$b_n(\omega) = s_{T_n(\omega)}(\omega). \tag{7}$$

Note that the $\{b_n\}$ process operates at a <u>different "time scale"</u> slower than $\{s_t\}$ as seen in Figure 2 adapted from [6]. $\{b_n\}$ is obtained from $\{s_t\}$ by "erasing" the time that $\{s_t\}$ does not spend in the boundary states.

Consider the following proposition whose proof can be found in Revuz [7,p.25].

<u>Theorem 1</u> Let s_t $t = 0,1.,,,$ be any Markov chain with stationary transition probabilities and state space S. Let $B \subset S$, $s_0 = b_0 \in B$, and define b_n, $n = 0,1,\ldots$ by (6), (7). Then $\{b_n\}$ is also a Markov chain with stationary transition probabilities. Furthermore, if $\{s_t\}$ is strongly ergodic, so is $\{b_n\}$.

It follows that for every regulator strategy u the supervisor's observation process is a strongly ergodic Markov chain thus inheriting the properties of the lower layer process.

The next task is to propose a reasonable class of strategies for the supervisor. These must of course be based on the $\{b_n\}$ process, and

they must reflect the idea that the supervisor resets the regulator each time a boundary state is reached. This is formulated as follows. Each time a boundary state say β in B is reached the supervisor selects a regulator strategy v^β from a prespecified subset $V^\beta \subset U$. Thus a supervisor strategy is a b-tuple $v = (v^1,\ldots,v^b) \in V = V^1 x \ldots x V^b$.

Now suppose $v = (v^1,\ldots,v^b)$ is chosen. Then during the random time interval $[T_0,T_1]$ the evolution of s_t is governed by the transition probability matrix $P(v^{b_0})$, during $(T_1,T_2]$ by the matrix $P(v^{b_1}),\ldots$, during $(T_n,T_{n+1}]$ by the matrix $P(v^{b_n})$ etc., where $b_0,b_1,\ldots$ is the supervisor's observation process, and $0 \equiv T_0,T_1,\ldots$ are the reset times given by (6).

Several remarks should be made before proceeding with the analysis. Firstly, the process $\{s_t\}$ will <u>not</u> generally be Markovian any longer, although it is Markovian within each interval $(T_n,T_{n+1}]$. Secondly, if we take $V^\beta = U$ for all β in B, then essentially the supervisor takes over the task of the regulator and the two layers "collapse". The subset V^β must be restricted, presumably on the basis of simplicity of implementation of the regulator. One way of doing this is to identify a different regulator with each regulator strategy u. V^β is then the set of regulators available to the supervisor when the boundary state β is reached. Thus the supervisor's task is of a higher level: it consists of choosing a regulator. Alternatively we may think of a more versatile regulator with some variable parameters which are adjusted by the supervisor.

The supervisor's task is to select an optimal v. To formulate this precisely we must "lift" the dynamics (2) and the cost (3) or (5) defined at the regulator layer to the supervisor layer. This is accomplished by the next result whose proof is given in Section IV.

Theorem 2 Let $v = (v^1,\ldots,v^b) \in V$, and let $v^\alpha = (v^\alpha(1),\ldots,v^\alpha(s))$.

(i) The supervisor process b_n, $n = 0,1\ldots$ is a strongly ergodic Markov chain with a unique steady state probability distribution $p(v) = (p_1(v),\ldots,p_b(v))$. $p(v)$ is the solution of

$$p(v)P(v) = p(v),\ p(v)\underset{\sim}{1} = 1 \tag{8}$$

where $P(v) = \{P_{\alpha\beta}(v);\ \alpha,\beta \text{ in } B\}$ is such that its αth row $P_\alpha(v) = (P_{\alpha 1},\ldots,P_{\alpha b})$ depends only on v^α.

(ii) The (non-Markovian) regulator process s_t, $t = 0,1,\ldots$ is such that the limit

$$\mathcal{J}(v) = \lim_T \frac{1}{T+1} E \sum_0^T k(s_t, v^{\beta_t}(s_t))$$

exists, where $\beta_t = b_n$ for t in $[T_n, T_{n+1})$.

(iii) Moreover there exist functions $K(\beta,v^\beta)$ and $T(\beta,v^\beta)$ defined for β in B and v^β in V^β such that

$$\mathcal{J}(v) = \frac{\sum_1^b p_\beta(v)K(\beta,v^\beta)}{\sum_1^b p_\beta(v)T(\beta,v^\beta)} \tag{9}$$

Thus the supervisor's problem is defined by the dynamics (8) and cost function (9). Observe that (8) has the same form as (2) whereas (9) differs from (5). The denominator in (9) arises as will be seen from the fact that the time scale of the supervisor process is slower than the plant process.

III. INTERMEDIATE RESULTS

Let x_t, $t = 0,1,\ldots$ be a strongly ergodic Markov chain with state space $X = \{1,\ldots,x\}$ and transition probability matrix P^X. Let $p^X = (p_1^X,\ldots,p_x^X)$ be its steady state probability distribution. Let $k^X = (k_1^X,\ldots,k_x^X)'$ be a given instantaneous cost and let

$$J = p^X k^X \tag{10}$$

Let $Y = \{1,\ldots,y\} \subset X$. Suppose $x_0 = y_0 \in Y$. Define $0 \equiv T_0 < T_1 < \ldots$ by

$$T_{n+1}(\omega) = \min\{t > T_n(\omega) \,|\, x_t(\omega) \in Y\}$$

and the process y_n, $n = 0,1,\ldots$ by $y_n(\omega) = x_{T_n(\omega)}(\omega)$. By Theorem 1 $\{y_n\}$ is a strongly ergodic Markov chain. Denote its transition probability matrix by P^Y and its steady state probability distribution by $p^Y = (p_1^Y,\ldots,p_y^Y)$. We first relate these to P^X, p^X.

Let $Z = X-Y$ and partition P^X as

$$P^X = \left[\begin{array}{c|c} P^{YY} & P^{YZ} \\ \hline P^{ZY} & P^{ZZ} \end{array}\right]. \tag{11}$$

Then according to [8,p.134] we have

$$P^Y = P^{YY} + P^{YZ}[I-P^{ZZ}]^{-1}P^{ZY}. \tag{12}$$

Using the fact that p^Y is determined by $p^Y P^Y = p^Y$, $p^Y \underset{\sim}{1} = 1$, it follows from (12) that p^Y is just the restriction of p^X to Y, i.e.,

$$p_i^Y = p_i^X[\sum_{j \in Y} p_j^X]^{-1}, \quad i \in Y.$$

We proceed to lift the cost (10) to the chain $\{y_n\}$. The next result is known (see, e.g. [9,p.151] or [10,Lemma 3.1]).

Lemma 1 Consider the x linear equations in the 1+x variables $\gamma \in R$, $c \in R^X$,

$$\gamma \underset{\sim}{1} = [P^X - I]c + k^X. \tag{13}$$

(i) If (γ,c) is a solution, then $\gamma=J$. (ii) If (γ,c) is a solution, then so is $(\gamma,c+\delta\underset{\sim}{1})$ for every δ. (iii) A solution always exists.

Let (γ,c) be a solution to (13). Partition c, k, $\underset{\sim}{1}$ as $c = (c^Y,c^Z)'$, $k = (k^Y,k^Z)'$, $\underset{\sim}{1} = (\underset{\sim}{1}^Y,\underset{\sim}{1}^Z)'$, so that (13) can be partitioned as

$$\gamma \underset{\sim}{1}^Y = [P^{YY} - I]c^Y + P^{YZ}c^Z + k^Y, \tag{14}$$

$$\gamma \underset{\sim}{1}^Z = P^{ZY}c^Y + [P^{ZZ} - I]c^Z + k^Z. \tag{15}$$

Premultiplying (15) by $P^{YZ}[I-P^{ZZ}]^{-1}$ and adding to (14) we get

$$\gamma(\underset{\sim}{1}^Y + P^{YZ}[I-P^{ZZ}]^{-1}\underset{\sim}{1}^Z) = [P^{YY} - I]c^Y + P^{YZ}[I-P^{ZZ}]^{-1}P^{ZY}c^Y + k^Y + P^{YZ}[I-P^{ZZ}]^{-1}k^Z. \tag{16}$$

Let $K = (K_1,\ldots,K_y)'$, $T = (T_1,\ldots,T_y)'$ be defined by

$$K = k^Y + P^{YZ}[I-P^{ZZ}]^{-1}k^Z, \quad T = \underset{\sim}{1}^Y + P^{YZ}[I-P^{ZZ}]^{-1}\underset{\sim}{1}^Z \tag{17}$$

From (12), (16), (17) we see that

$$\gamma T = [P^Y - I]c^Y + K \tag{18}$$

which can be compared with (13). Also since $\gamma=J$ by Lemma 1, we get

$$\gamma = J = \frac{p^Y K}{p^Y T} \tag{19}$$

We can interpret T, K. Since $[I-P^{ZZ}]^{-1} = \sum_{t=0}^{\infty} [P^{ZZ}]^t = \{N_{ij}\}$, i,j in Z say, therefore N_{ij} is the expected time that the process $\{x_t\}$, starting in state i, spends in state j before entering Y. It follows from (17) that for $i \in Y$, T_i is the expected time that the process $\{x_t\}$, starting in state i, spends in $X \cup \{i\}$ before entering Y, whereas K_i is the expected cost incurred during this time.

IV. PROOF OF THEOREM 2

Let $v = (v^1,\ldots,v^b)$ be a fixed supervisor strategy with $v^\beta = (v^\beta(1),\ldots,v^\beta(s))$. Let b_n, n = 0,1,... be the resulting supervisor process and s_t, t = 0,1,... the regulator process. The latter is not necessarily Markovian since knowing s_t does not tell us which regulator strategy v^β is in place at time t. That information is given by the last boundary state visited by $\{s_t\}$ before t, that is by the process β_t, t = 0,1,... where $\beta_t = b_n$ for t in $[T_n,T_{n+1})$. Hence the augmented process $x_t = (\beta_t,s_t)$, t = 0,1,... is Markov. Since we must have $\beta_t = s_t$ whenever $s_t \in B$ therefore x_t takes values in

$$X = \{(\beta,\beta) \mid \beta \in B\} \cup \{(\beta,i) \mid \beta \in B,\ i \notin B\} = Y \cup Z \text{ say.} \qquad (20)$$

The transition probabilities of $\{x_t\}$ can be easily evaluated, using the notation of (2), as

$$P^X_{(\alpha,i)(\beta,j)} = \text{Prob}\{x_{t+1} = (\beta,j) \mid x_t = (\alpha,i)\}$$

$$= \begin{cases} P_{ij}(v^\alpha(i)) & \text{if } \beta=j \text{ or if } \beta=\alpha \text{ and } j \notin B \\ 0 & \text{if } \beta\neq\alpha \text{ and } \beta\neq j \end{cases} \qquad (21)$$

(Here (α,i), (β,j) are of course restricted to X.) It is easy to check using the strong ergodicity assumption that $\{x_t\}$ is strongly ergodic also. Hence it has a unique steady state probability vector $p^X = \{p^X_{(\alpha,i)};\ (\alpha,i) \in X\}$.

We proceed with the proof of Theorem 2. Consider first the subset Y of X specified in (20) and define the chain y_n, n = 0,1,... as in Section III. That is $y_n(\omega) = x_{T_n(\omega)}(\omega)$ where $0 \equiv T_0$ and

$$T_{n+1}(\omega) = \min\{t > T_n(\omega) \,|\, x_t(\omega) \in Y\}$$

$$= \min\{t > T_n(\omega) \,|\, s_t(\omega) \in B\}$$

by (20). Hence

$$y_n(\omega) \equiv (b_n(\omega),\ b_n(\omega))$$

so that $\{y_n\}$ is essentially identical to the supervisor process $\{b_n\}$. By Theorem 1 $\{y_n\}$, hence $\{b_n\}$, is a strongly ergodic Markov chain and so it has a unique steady state distribution $p(v) = (p_1(v),\ldots,p_b(v))$.

Secondly, observe that if we define a cost function q for $\{x_t\}$ by

$$q((\alpha,i),v(\alpha,i)) = k(i,v^{\alpha}(i)),\ (\alpha,i) \in X$$

then the cost can be rewritten as

$$\mathcal{J}(v) = \lim_T \frac{1}{T+1} E \sum_0^T k(s_t, v^{\beta_t}(s_t)) = \lim_T \frac{1}{T+1} E \sum_0^T q(x_t, v(x_t))$$

and the second limit exists since $\{x_t\}$ is strongly ergodic. In fact it can be rewritten as

$$\mathcal{J}(v) = \sum_{(\alpha,i) \in X} p^X_{(\alpha,i)}\ k(i,v^{\alpha}(i)).$$

It remains to obtain the representations (8), (9). We do this by using the representations (12), (17). So partition p^X as in (11) with Y,Z given by (20). Denote the components of the transition probability matrix P^Y of the process $\{y_n\}$, hence also of $\{b_n\}$, by

$$P_{\alpha\beta} = P_{\alpha\beta}(v) = \text{Prob}\{y_{n+1} = (\beta,\beta) \,|\, y_n = (\alpha,\alpha)\}$$

$$= \text{Prob}\{b_{n+1} = \beta \,|\, b_n = \alpha\},\ \alpha,\beta \in B,$$

and let its αth row be denoted

$$P_\alpha(v) = (P_{\alpha 1},\ldots,P_{\alpha b}).$$

Next let $A = S-B = \{b+1,\ldots,s\}$ and consider the following row vectors and matrices:

$$P_{\alpha B}(v^\alpha) = (P_{\alpha\beta}(v^\alpha(\alpha));\ \beta \in B),$$

$$P_{\alpha A}(v^\alpha) = (P_{\alpha i}(v^\alpha(\alpha));\ i \in A),$$

$$P_{AA}(v^\alpha) = \{P_{ij}(v^\alpha(i));\ i,j \in A\},$$

$$P_{AB}(v^\alpha) = \{P_{i\beta}(v^\alpha(i));\ i \in A,\ \beta \in B\}.$$

Then we can substitute for P^X from (21) into (12) and verify that the αth row of P^Y is given by

$$P_\alpha(v) = P_{\alpha B}(v^\alpha) + P_{\alpha A}(v^\alpha)[I-P_{AA}(v^\alpha)]^{-1}P_{AB}(v^\alpha) \tag{22}$$

which depends only on v^α, thereby proving (8). Now from (19) we know that there exist vector $K = (K_1,\ldots,K_b)'$ and $T = (T_1,\ldots,T_b)'$ such that

$$\mathcal{J}(v) = \frac{p(v)K}{p(v)T}.$$

From (17) we can verify that

$$K_\beta = k(\beta,v^\beta(\beta)) + P_{\beta A}(v^\beta)[I-P_{AA}(v^\beta)]^{-1}k_A(v^\beta) = K(\beta,v^\beta) \tag{23}$$

where the column vector $k_A(v^\beta)$ has components $(k(i,v^\beta(i));\ i \in A)$. From (17) again, we verify that

$$T_\beta = 1 + P_{\beta A}(v^\beta)[I-P_{AA}(v^\beta)]^{-1}\underset{\sim}{1}^A = T(\beta,v^\beta). \tag{24}$$

Theorem 2 is proved.

$T(\beta,v^{\beta})$ is the expected time that the process $\{s_t\}$ takes starting at β and before reaching the boundary states B and $K(\beta,v^{\beta})$ is the expected cost incurred during this time. Thus the $T(\beta,v^{\beta})$ give the expected lower layer or "real" time between transitions of the supervisor process and the $K(\beta,v^{\beta})$ give the expected cost between these transitions.

The interpretation of K,T suggest the following practical consideration. Suppose the supervisor selects a strategy v. Then to reckon its effect on the dynamics and the cost it is necessary to evaluate (22), (23), (24) which requires complete knowledge of the lower layer transition probabilities and costs. To a certain extent this is self-defeating since the higher layer should have reduced a priori as well as reduced "on line" information. Suppose that this prior knowledge is not available but that the supervisor observes the transition times $T_0,T_1,\ldots$, the process b_n, $n = 0,1,\ldots$ and the cummulative cost

$$k_n = \sum_0^{T_n} k(s_t,v^{\beta_t}(s_t)).$$

Then the supervisor can use these observations to estimate $P_{\alpha}(v)$, $K(\beta,v^{\beta})$ and $T(\beta,v^{\beta})$ by means of obvious empirical averages. Since the augmented process $\{x_t\}$ is ergodic these estimates will be consistent. It may therefore be possible to combine such estimates with the algorithm proposed in the next section to obtain an adaptive control scheme for the supervisor.

V. OPTIMALITY CONDITIONS

From Theorem 2 we see that an optimal supervisor strategy is the solution to the following problem,

$$\text{Min } \mathcal{J}(v) = \sum_1^b p_\beta(v)K(\beta,v^\beta)[\sum_1^b p_\beta(v)T(\beta,v^\beta)]^{-1}$$

$$\text{s.t. } p(v)P(v) = p(v),\ p(v)\underset{\sim}{1} = 1,$$

$$v = (v^1,\ldots,v^b) \in V = V^1x\ldots xV^b.$$

As before let $P_\beta(v^\beta)$ be the βth row of $P(v)$. For any $c = (c_1,\ldots,c_b)'$ and v in V define $H(c,v) = (H_1(c,v^1),\ldots,H_b(c,v^b))'$ and $G(c,v) = (G_1(c,v^1),\ldots G_b(c,v^b))'$ by

$$H_\beta(c,v^\beta) = P_\beta(v^\beta)c - c_\beta + K(\beta,v^\beta);\ G_\beta(c,v^\beta) = H_\beta(c,v^\beta)[T(\beta,v^\beta)]^{-1}. \tag{25}$$

Let $g(c)$ be given by

$$g_\beta(c) = \inf\{G_\beta(c,v^\beta)\,|\,v^\beta \in V^\beta\},\ \beta = 1,\ldots,b. \tag{26}$$

In obtaining the optimality conditions we need the next result which can be compared with Lemma 1.

Lemma 2 Let v be fixed. Consider the b linear equations in the $1+b$ variables γ,c.

$$\gamma\underset{\sim}{1} = G(c,v). \tag{27}$$

(i) If (γ,c) is a solution, then $\gamma = \mathcal{J}(v)$. (ii) If (γ,c) is a solution, then so is $(\gamma,c+\delta\underset{\sim}{1})$ for every δ. (iii) A solution always exists.

Proof Rewrite (27) as

$$\gamma T(v) = [P(v)-I]c + K(v) \tag{28}$$

where $T(v) = T(1,v^1),\ldots,T(b,v^b))'$, $K(v) = (K(1,v^1),\ldots,K(b,v^b))'$. Premultiplying (28) by $p(v)$ gives

$$\gamma p(v)T(v) = p(v)[P(v)-I]c + p(v)K(v) = 0 + p(v)K(v)$$

so that $\gamma = \mathcal{J}(v)$. (ii) follows from the fact that $[P(v)-I]\underset{\sim}{1} = 0$. Finally note that

$$p(v)[\mathcal{J}(v)T(v)-K(v)] = 0$$

so that $\mathcal{J}(v)T(v)-K(v)$ is orthogonal to the null space of $[P(v)-I]'$, hence it is in the range of $P(v)-I$.

¤

Theorem 3 (Minimum principle) v is optimal if and only if there exist γ, c such that

$$\gamma\underset{\sim}{1} = G(c,v) = g(c). \tag{29}$$

Proof Sufficiency: By Lemma 2 $\gamma = \mathcal{J}(v)$. Let $w \in V$. By definition of $g(c)$

$$\mathcal{J}(v)\underset{\sim}{1} \leq G(c,w)$$

and since $T(w) > 0$ therefore

$$\mathcal{J}(v)T(w) \leq [P(w)-I]c + K(w).$$

Premultiplying both sides by $p(w)$ gives $\mathcal{J}(v)p(w)T(w) \leq p(w)K(w)$ and so $\mathcal{J}(v) \leq \mathcal{J}(w)$.

Necessity: Suppose v is optimal. Let c be such that

$$\mathcal{J}(v)\underset{\sim}{1} = G(c,v).$$

Let $w \in V$ be such that $\mathcal{J}(v)\underset{\sim}{1} \geq G(c,w)$, hence

$$\mathcal{J}(v)T(w) \geq [P(w)-I]c + K(w).$$

Premultiplying both sides by $p(w)$ yields

$$\mathcal{J}(v) \geq \mathcal{J}(w),$$

and since v is optimal we must have equality. Since $p_\beta(w) > 0$ for each β this implies that $\mathcal{J}(v)\underset{\sim}{1} = G(c,w)$. Hence $\mathcal{J}(v)\underset{\sim}{1} = g(c)$.

□

We can also obtain bounds similar to [10, Theorem 4.1]. Define

$$\underline{g}(c) = \min_\beta g_\beta(c) = \min_\beta \inf_{v^\beta} G_\beta(c,v^\beta)$$

$$\bar{g}(c) = \max_\beta g_\beta(c) = \max_\beta \inf_{v^\beta} G_\beta(c,v^\beta)$$

Theorem 4 (Bounds) Let c be arbitrary and let v be a minimizer in (26) i.e., $G(c,v) = g(c)$. Then

$$\underline{g}(c) \leq \mathcal{J}(v) \leq \bar{g}(c)$$

$$\underline{g}(c) \leq \mathcal{J}^* \leq \bar{g}(c)$$

where $\mathcal{J}^* = \inf_v \mathcal{J}(v)$. Also v is optimal if and only if $\underline{g}(c) = \bar{g}(c)$.

Proof $\underline{g}(c)\underset{\sim}{1} \leq G(c,v) \leq \bar{g}(c)\underset{\sim}{1}$, and so

$$\underline{g}(c)T(v) \leq [P(v)-I]c + K(v) \leq \bar{g}(c)T(v).$$

Premultiplication by p(v) gives

$$\underline{g}(c)p(v)T(v) \leq p(v)K(v) \leq \bar{g}(c)p(v)T(v)$$

from which

$$\underline{g}(c) \leq \mathcal{J}(v) \leq \bar{g}(c).$$

Next let $w \in V$. Then

$$\underline{g}(c)\underset{\sim}{1} \leq G(c,w)$$

from which, as above, we can conlcude $\underline{g}(c) \leq \mathcal{J}(w)$. Hence $\underline{g}(c) \leq \inf \mathcal{J}(w) = \mathcal{J}^*$. The final assertion follows from Theorem 3.

□

With Theorems 3,4 in hand we propose the following algorithm for finding an optimal supervisor strategy.

Step 0 Let v_0 be any initial strategy and solve (27) to obtain c_0 such that $\mathcal{J}(v_0)\underset{\sim}{1} = G(c_0,v_0)$. If v_0 is unavailable, start with any c_0. Go to Step 1.

Step 1 Let c_n be given. Let v_{n+1} be a minimizer of (26) i.e. $g(c_n) = G(c_n,v_{n+1})$. If $\bar{g}(c_n) - \underline{g}(c_n) \leq \varepsilon$ stop, because by Theorem 4 $\mathcal{J}(v_{n+1}) - \mathcal{J}^* \leq \varepsilon$, otherwise go to Step 2.

Step 2 Let $c_{n+1} = c_n + \Delta\theta(c_n)$ where $\Delta > 0$ is a "small" number and for any c,

$$\theta(c) = g(c) - \frac{1}{b}[\underset{\sim}{1}'g(c)]\underset{\sim}{1}$$

Set n+1 = n and return to Step 1.

The behavior of the algorithm is partially analyzed in the next result which deals with its time-continuous version. The proof is omitted since it is virtually identical with that of [10,Theorem 5.1].

Theorem 5 Consider the differential equation

$$\frac{dc}{dt} = \theta(c),\ c(0) = c_0 \tag{30}$$

(i) There is a unique solution c(t) of (30) defined for all $t \geq 0$.

(ii) c(t) converges to a unique limit c* which is the solution of

$$\underline{g}(c) = \bar{g}(c),\ \underset{\sim}{1}'c = \underset{\sim}{1}'c_0.$$

(iii) $\underline{g}(c(t))$ increases, and $\bar{g}(c(t))$ decreases, strictly monotonically to $\mathcal{J}^*$.

As our final result we will show that the minimum cost decreases if the supervisor's information increases. More precisely, let

$B = \{1,\ldots,b\}$, $V = V^1 x \ldots x V^b$, $\hat{B} = \{1,\ldots,b+1\}$, $\hat{V} = V^1 x \ldots x V^{b+1}$ where

$V^{b+1} \supset V^1 \cup \ldots \cup V^b$.

We claim that

$$\inf_{V} \mathcal{J}(v) \geq \inf_{\hat{V}} \mathcal{J}(\hat{v}). \tag{31}$$

This assertion is not immediate because a (stationary) supervisor strategy $v \in V$ cannot be implemented as a strategy $\hat{v} \in \hat{V}$. However it can be implemented as a feedback control which depends on previous states. To see this let $v = (v^1,\ldots,v^b) \in V$ and let b_n, $n = 0,1,\ldots$ be the supervisor process with values in B, and the lower layer process be s_t, $t = 0,1,\ldots$. For the more informed supervisor denote the corresponding processes by $\hat{b}_n$, $n = 0,1,\ldots$ and $\hat{s}_t$, $t = 0,1\ldots$. Consider the following feedback control

$$\hat{v}_n = v(b_0,\ldots,b_n) = v^{b_m} \text{ where } m = \max\{\ell \leq n \,|\, \hat{b}_\ell \in B\},\ n = 0,1\ldots$$

It is evident then that $\hat{v}_n \equiv v^{b_n}$, $n = 0,1\ldots$ and so $\hat{s}_t \equiv s_t$. Hence $\mathcal{J}(\{\hat{v}_n\}) = \mathcal{J}(v)$. But by a slight modification of a standard argument (see e.g. [9,p.159] or [10, Corollary 3.17]) it can be shown that

$$\mathcal{J}(\{\hat{v}_n\}) \geq \inf_{\hat{V}} \mathcal{J}(\hat{v})$$

from which (31) follows.

V. CONCLUSIONS

The two layer control structure presented here posseses to some degree the desirable features mentioned in the Introduction. Its most attractive aspect is that the dynamics and cost can be lifted to the higher layer in a form which inherits the most important properties of the lower layer problem. The least attractive aspect of the structure is that to obtain

this higher layer problem the supervisor needs to know all of the a priori information. This is mitigated somewhat by the fact that the problem parameters can be estimated by the supervisor in a consistent way. Nevertheless the practical usefulness of the scheme will become clear only if the estimation of the supervisor's problem and its control can be combined in an "adaptive" control scheme. We hope to report such a scheme in the future.

REFERENCES

[1] W. Findeisen, "A survey of problems in hierarchical control," Proc. Workshop on Multilevel Control, Inst. of Automatic Control, Technical University, Warsaw, 1975.

[2] M. S. Malmoud, "Multilevel systems control and applications: a survey," IEEE Trans. Systems, Man, Cybernetics SMC-7, 125-143, 1977.

[3] C. Y. Chong and M. Athans, "On the periodic coordination of linar stochastic systems," Proc. 6th IFAC World Congress, Boston, Mass, 1975.

[4] A. Benveniste, P. Bernhard and A. Cohen, "On the decomposition of stochastic control problems," Rapport de Recherche No. 187, IRIA, Rocquencourt, France.

[5] J. F. Donoghue and I. Lefkowitz, "Economic tradeoffs associated with a multilayer control strategy for a class of static systems," IEEE Trans. Automatic Control, AC-17, 7-15, 1972.

[6] D. Freedman, Approximating Countable Markov Chains, Holden-Day, San Francisco, 1971.

[7] D. Revuz, Markov Chains, North Holland, New York, 1975.

[8] J. G. Kemeny, J. L. Snell and A. W. Knapp, Denumerable Markov Chains, Springer-Verlag, New York, 1976.

[9] H. Kushner, Introduction to Stochastic Control, Holt, Rinehart and Winston, New York, 1971.

[10] P. Varaiya, "Optimal and suboptimal stationary controls for Markov chains," Elect. Res. Lab. Memo. No. UCB/ERL 77/17, Univ. of Calif., Berkeley, 1977.

FIGURE CAPTIONS

Figure 1. A two-layer control structure

Figure 2. The lower layer process $\{s_t\}$ and supervisor process $\{b_n\}$.

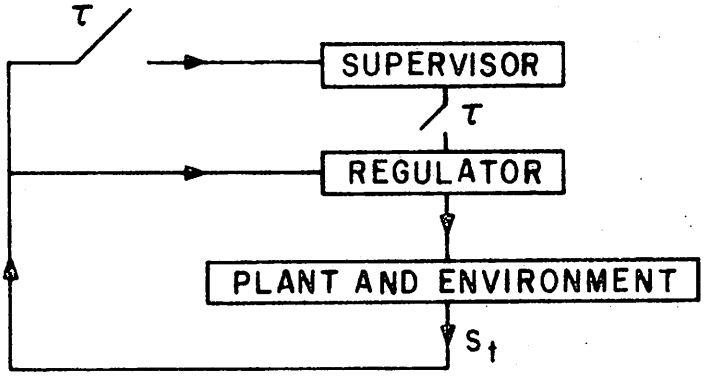

Figure 1

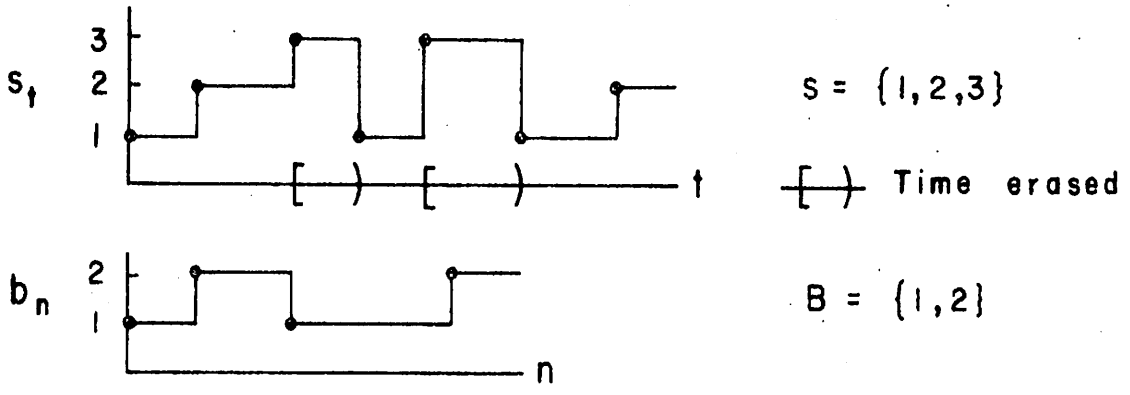

Figure 2