APPLICATIONES MATHEMATICAE
22,2 (1994), pp. 165–180

G. B. DI MASI (Padova) and Ł. STETTNER (Warszawa)

ON ADAPTIVE CONTROL OF A PARTIALLY OBSERVED MARKOV CHAIN

Abstract. A control problem for a partially observable Markov chain depending on a parameter with long run average cost is studied. Using uniform ergodicity arguments it is shown that, for values of the parameter varying in a compact set, it is possible to consider only a finite number of nearly optimal controls based on the values of actually computable approximate filters. This leads to an algorithm that guarantees nearly selfoptimizing properties without identifiability conditions. The algorithm is based on probing control, whose cost is additionally assumed to be periodically observable.

1. Introduction. On a given probability space $\{\Omega, \mathcal{F}, P\}$ consider a discrete-time Markov chain x_k $(k = 0, 1, \ldots)$ with controlled transition matrix $P^{v\alpha_0}(i,j)$, where $i, j \in E = \{1, \ldots, s\}$, the control v lies in a compact metric space V and the parameter α_0 belongs to a compact metric space A. The initial state x_0 is assumed to be distributed according to a given initial law μ_0 and α_0 stands for an unknown parameter. The process x_k is partially observed via the s-dimensional process

$$y_k = h(x_k) + w_k \tag{1}$$

where w_k is a sequence of s-dimensional i.i.d. random vectors with standard normal distribution and $h : E \to \mathbb{R}^s$ has components

$$h^i(j) = \begin{cases} 0 & \text{for } j \neq i, \\ h_i > 0 & \text{for } j = i. \end{cases} \tag{2}$$

Therefore the information available at time k is provided by the σ-field $\mathcal{Y}_k = \sigma\{y_1, \ldots, y_k\}$.

1991 *Mathematics Subject Classification*: Primary 93E20, 93E11; Secondary 93E10, 93E15.

Key words and phrases: adaptive control, partially observed systems, filtering process, uniform ergodicity, approximate filter, long run average cost.

Looking at the form (1)–(2) adopted for the observations, it is clear that it reflects the situation in which there are "indicators" h^i that detect that the ith state has been reached; the values provided by the indicators are, however, corrupted by white noise, but still the observation structure is such that each state can be monitored by the controller. Many of the problems discussed e.g. in [16] can be described (possibly more realistically) in terms of our model; in particular, we mention economic models (e.g. cost control), analysis of diagnostic data, computer networks problems.

The control v_k used at time k is a $\mathcal{Y}_k$-measurable V-valued random variable. The purpose of the control is to minimize the pathwise long run average cost

$$J^{\alpha_0}(\{v_k\}) = \limsup_{n\to\infty} n^{-1} \sum_{k=0}^{n-1} c(x_k, v_k) \tag{3}$$

or the long run expected average cost

$$\bar{J}^{\alpha_0}(\{v_k\}) = \limsup_{n\to\infty} n^{-1} \mathbf{E}_{\mu_0}^{\alpha_0} \sum_{k=0}^{n-1} c(x_k, v_k) \tag{4}$$

where $c(x,\cdot)$ is a continuous function and $\mathbf{E}_{\mu_0}^{\alpha_0}$ denotes expectation with respect to the probability $P_{\mu_0}^{\alpha_0}$ induced on the space of trajectories by the process x_k with parameter α_0, for fixed initial condition $x_0 \sim \mu_0$ and control sequence $\{v_k\}$.

In order to transform the partially observable control problem into a completely observable one [3], [14], we introduce the filtering process

$$\pi_k(i) = P_{\mu_0}^{\alpha_0}\{x_k = i \mid \mathcal{Y}_k\}\,. \tag{5}$$

This process can be recursively obtained for $k = 1, 2, \ldots$ by

$$\pi_k(i) = \frac{\sigma_k(i)}{\sum_{j=1}^{s} \sigma_k(j)} \tag{6}$$

(see [12]) where

$$\sigma_k(i) = \exp[-\tfrac{1}{2}\langle h(i), h(i)\rangle + \langle y_k, h(i)\rangle] q(i; \alpha_0, v_{k-1}, \pi_{k-1}) \tag{7}$$

with $\langle\,,\rangle$ denoting inner product in $\mathbb{R}^s$,

$$q(i; \alpha_0, v_{k-1}, \pi_{k-1}) = P_{\mu_0}^{\alpha_0}\{x_k = i \mid \mathcal{Y}_{k-1}\} = \sum_{j=1}^{s} P^{v_{k-1}\alpha_0}(j, i)\pi_{k-1}(j) \tag{8}$$

and initial condition given by

$$\pi_0(i) = \mu_0(i)\,.$$

The recursive formula for the filter is concisely written as

$$\pi_{k+1} = G^{\alpha_0}(\pi_k, y_{k+1}, v_k)\,, \tag{9}$$

which provides the evolution of the "completely observed" state π_k. The control functional $\bar{J}^{\alpha_0}$ can be written in terms of π_k as

$$\begin{aligned}(10)\qquad \bar{J}^{\alpha_0}(\{v_k\}) &= \limsup_{n\to\infty} n^{-1}\mathbf{E}^{\alpha_0}_{\mu_0}\sum_{k=0}^{n-1}\sum_{i=1}^{s} c(i,v_k)\pi_k(i)\\ &=: \limsup_{n\to\infty} n^{-1}\mathbf{E}^{\alpha_0}_{\mu_0}\sum_{k=0}^{n-1} C(\pi_k,v_k)\,.\end{aligned}$$

It is shown in Corollary 1 that the optimal controls for the cost function (10) are functions of the "completely observed" state π_k; more precisely, denoting by S the $(s-1)$-dimensional simplex and by $\mathcal{B}(S)$ the Borel σ-algebra on S, we have $v_k = u_k(\pi_k)$, where the control law $u_k : S \to V$ is $\mathcal{B}(S)$-measurable. In what follows it is of particular interest to consider the case of time-invariant laws, i.e. $v_k = u(\pi_k)$; the class of such laws is denoted by $\mathcal{U}$. Furthermore, several quantities are parametrized interchangeably by $u \in \mathcal{U}$ or by $v \in V$, with obvious meaning of the symbols, e.g. $P^{u\alpha_0}(i,j)$ is used for $P^{u(i)\alpha_0}(i,j)$.

In what follows we require the following assumptions:

(A.1) $\quad \inf_{v\in V}\inf_{\alpha\in A}\inf_{i,j\in E} P^{v\alpha}(i,j) \ge \beta > 0\,,$

(A.2) $\quad$ for each $i,j \in E$, $P^{v\alpha}(i,j)$ is a continuous function on $A\times V$.

The aim of the paper is to find a control procedure which guarantees the ε-optimal value of the cost functional for the state process x_k corresponding to the unknown parameter α_0. In the next Section 2, using uniform ergodicity arguments, we show that for each ε, there exists a finite set $\{u_j : S \to V : j = 1,\ldots,r\}$ of control functions such that for each $\alpha \in A$ one of such functions is ε-optimal for α. This allows us to limit the choice of control functions only to a finite set. Loosely speaking, there are only a finite number of relevant control functions. Then, in Section 3, we show that the same holds if approximate filtering is used; more precisely, if an incorrect value of α and of the initial condition is used in the filtering formula. This result is based on the joint uniform ergodicity of approximate filter and state driven by controls based on the approximate filter. Finally, in the last Section 4, we provide a direct adaptive control procedure that guarantees nearly optimal behaviour without explicit identification of the parameter α, provided that we are able to observe periodically the cost. This adaptive procedure can be applied to any uniformly ergodic stochastic system and possesses nearly selfoptimizing properties.

The notion of approximate filter appeared in a paper of Kushner and Huang [15] and was studied later in [1], [6], [7]. Optimal ergodic control with partial observations was studied for a more general model by Runggaldier and Stettner [17]. Adaptive control of partially observable Markov processes has been studied in [10], [11] for discounted cost criterion and for a particular Quality Control/Replacement model in [6], [7]. However, the general problem of control of irreducible Markov chains with partial observations and long run average cost seems to be open. Particular approaches are based either on the special form of white-noise corrupted observations [17] or on the simple structure of Quality Control/Replacement models [6], [7]. In the present paper we present an alternative approach to the problem based on a particularly rich observation structure.

2. Uniform ergodicity of the controlled filtering process. In this section we study the uniform ergodicity of the filtering process π_k corresponding to *a generic value of the parameter* α. For this purpose we need to investigate some properties of the transition kernel for π_k, which in turn are derived from the corresponding properties of the unnormalized filter σ_k.

Using (7), (1) and (2), we have

$$
\begin{aligned}
(11)\quad & P^\alpha\{\sigma_k(i)\le z_i,\ i=1,\ldots,s\mid \mathcal{Y}_{k-1}\}\\
&= P^\alpha\{w_k^i\le -h^i(x_k)+h_i^{-1}[\ln z_i-\ln q(i;\alpha,v_{k-1},\pi_{k-1})+\tfrac12 h_i^2],\\
&\qquad\qquad i=1,\ldots,s\mid \mathcal{Y}_{k-1}\}\\
&=\sum_{r=1}^s P^\alpha\{w_k^i\le -h^i(r)+h_i^{-1}[\ln z_i-\ln q(i;\alpha,v_{k-1},\pi_{k-1})+\tfrac12 h_i^2],\\
&\qquad\qquad i=1,\ldots,s\mid \mathcal{Y}_{k-1},x_k=r\}P^\alpha_{\mu_0}\{x_k=r\mid \mathcal{Y}_{k-1}\}\\
&=\sum_{r=1}^s P^\alpha\{w_k^i\le -h^i(r)+h_i^{-1}[\ln z_i-\ln q(i;\alpha,v_{k-1},\pi_{k-1})+\tfrac12 h_i^2],\\
&\qquad\qquad i=1,\ldots,s\mid \mathcal{Y}_{k-1}\}q(r;\alpha,v_{k-1},\pi_{k-1})\,.
\end{aligned}
$$

The conditional distribution function given by (11) has density

$$
\begin{aligned}
(12)\quad & g(z_1,\ldots,z_s;\alpha,v_{k-1},\pi_{k-1})\\
&:=\frac{d^s}{dz_1\ldots dz_s}P^\alpha\{\sigma_k(i)\le z_i,\ i=1,\ldots,s\mid \mathcal{Y}_{k-1}\}\\
&=\sum_{r=1}^s\frac{1}{(2\pi)^{s/2}}\Big\{\prod_{i=1}^s \exp[-\tfrac12[-h^i(r)+h_i^{-1}(\ln z_i-\ln q(i;\alpha,v_{k-1},\pi_{k-1})\\
&\qquad\qquad +\tfrac12 h_i^2)]^2](h_iz_i)^{-1}\Big\}q(r;\alpha,v_{k-1},\pi_{k-1})\\
&=:\sum_{r=1}^s g_r(z_1,\ldots,z_s;\alpha,v_{k-1},\pi_{k-1})q(r;\alpha,v_{k-1},\pi_{k-1})\,.
\end{aligned}
$$

We have the following

LEMMA 1. *There exist positive, integrable functions* $\underline{g}, \overline{g} : \mathbb{R}_+^s \to \mathbb{R}$ *such that for all* $\alpha \in A$, $v \in V$ *and* $\nu \in S$,

$$0 < \underline{g}(z_1, \ldots, z_s) \le g_r(z_1, \ldots, z_s; \alpha, v, \nu) \le \overline{g}(z_1, \ldots, z_s)\,. \tag{13}$$

Proof. Define

$$d_k(i,r) = \ln q(i; \alpha, v_{k-1}, \pi_{k-1}) - \tfrac{1}{2}h_i^2 + h^i(r)h_i\,.$$

Then using the fact that $\beta \le q(i; \alpha, v_{k-1}, \pi_{k-1}) \le 1$, with β as in (A.1), we have

$$\underline{d}(i,r) \le d_k(i,r) \le \overline{d}(i,r)$$

where

$$\underline{d}(i,r) = \ln\beta - \tfrac{1}{2}h_i^2 + h^i(r)h_i, \qquad \overline{d}(i,r) = -\tfrac{1}{2}h_i^2 + h^i(r)h_i\,.$$

Consider now the set of functions

$$\begin{aligned} D(i,r) = \{&[\ln z_i - \underline{d}(i,r)][\ln z_i - \underline{d}(i,r)];\\ &[\ln z_i - \underline{d}(i,r)][\ln z_i - \overline{d}(i,r)];\ [\ln z_i - \overline{d}(i,r)][\ln z_i - \overline{d}(i,r)]\}\,. \end{aligned}$$

Then

$$\min D(i,r) \le [\ln z_i - d_k(i,r)]^2 \le \max D(i,r)$$

and as a consequence

$$\underline{g}(z_1, \ldots, z_s) \le g_r(z_1, \ldots, z_s; \alpha, v_{k-1}, \pi_{k-1}) \le \overline{g}(z_1, \ldots, z_s)$$

where

$$\underline{g}(z_1, \ldots, z_s) = \sum_{r=1}^{s} \frac{1}{(2\pi)^{s/2}} \left\{ \prod_{i=1}^{s} \exp\left[-\frac{1}{2h_i^2} \max D(i,r)\right] (h_i z_i)^{-1} \right\} > 0, \tag{14}$$

$$\overline{g}(z_1, \ldots, z_s) = \sum_{r=1}^{s} \frac{1}{(2\pi)^{s/2}} \left\{ \prod_{i=1}^{s} \exp\left[-\frac{1}{2h_i^2} \min D(i,r)\right] (h_i z_i)^{-1} \right\}. \tag{15}$$

Due to the form of the functions in $D(i,r)$, it is also clear that $\underline{g}(z_1, \ldots, z_s)$ and $\overline{g}(z_1, \ldots, z_s)$ are integrable functions. ∎

For $B \in \mathcal{B}(S)$ define

$$B^{\mathrm{r}} = \Big\{(\xi_1, \ldots, \xi_{s-1}) \in \mathbb{R}_+^{s-1} : \Big(\xi_1, \ldots, \xi_{s-1}, 1 - \sum_i \xi_i\Big) \in B\Big\}$$

and

$$B^{\mathrm{n}} = B^{\mathrm{r}} \times \mathbb{R}_+\,.$$

From (6) we deduce that the conditional probabilities $\pi_k(i)$, $i = 1, \ldots, s-1$, can be obtained from the unnormalized probabilities σ using the transformation $H:\mathbb{R}_+^s \to S^{\mathrm{n}} := S^{\mathrm{r}} \times \mathbb{R}_+$ given by

$$\begin{bmatrix} \xi_1 \\ \vdots \\ \xi_{s-1} \\ \xi_s \end{bmatrix} = H \begin{bmatrix} z_1 \\ \vdots \\ z_{s-1} \\ z_s \end{bmatrix} = \begin{bmatrix} z_1/\sum_i z_i \\ \vdots \\ z_{s-1}/\sum_i z_i \\ \sum_i z_i \end{bmatrix}. \tag{16}$$

Clearly

$$\begin{bmatrix} z_1 \\ \vdots \\ z_{s-1} \\ z_s \end{bmatrix} = H^{-1} \begin{bmatrix} \xi_1 \\ \vdots \\ \xi_{s-1} \\ \xi_s \end{bmatrix} = \begin{bmatrix} \xi_1\xi_s \\ \vdots \\ \xi_{s-1}\xi_s \\ (1-\sum_{i=1}^{s-1}\xi_i)\xi_s \end{bmatrix} \tag{17}$$

and its associated Jacobian is given by

$$\left|\frac{\partial H^{-1}}{\partial \xi}\right| = \xi_s^{s-1}. \tag{18}$$

We have

$$\begin{aligned} (19)\quad & P_{\mu_0}^{\alpha}\{\pi_k \in B \mid \mathcal{Y}_{k-1}\} = P\{H(\sigma_k) \in B^{\mathrm{n}} \mid \mathcal{Y}_{k-1}\} \\ &= P_{\mu_0}^{\alpha}\{\sigma_k \in H^{-1}(B^{\mathrm{n}}) \mid \mathcal{Y}_{k-1}\} \\ &= \int\limits_{H^{-1}(B^{\mathrm{n}})} g(z_1,\ldots,z_s;\alpha,v_{k-1},\pi_{k-1})dz_1\ldots dz_s \\ &= \int\limits_{B^{\mathrm{r}}} \int\limits_0^{\infty} \xi_s^{s-1} \\ &\quad \times g\Big(\xi_1\xi_s,\ldots,\xi_{s-1}\xi_s,\Big(1-\sum_{i=1}^{s-1}\xi_i\Big)\xi_s;\alpha,v_{k-1},\pi_{k-1}\Big)\,d\xi_s\,d\xi_1\ldots d\xi_{s-1}. \end{aligned}$$

Proposition 1. *The filtering process π_k corresponding to an admissible control function $u:S\to V$ is a Markov process with respect to the σ-field $\mathcal{Y}_k$ with transition kernel*

$$\Pi^{u\alpha}(\nu,B) = \int\limits_{B^{\mathrm{r}}} f(\xi_1,\ldots,\xi_{s-1};\alpha,u(\nu),\nu)\,d\xi_1\ldots d\xi_{s-1} \tag{20}$$

with

$$f(\xi_1,\ldots,\xi_{s-1};\alpha,u(\nu),\nu) = \sum_{r=1}^{s} f_r(\xi_1,\ldots,\xi_{s-1};\alpha,u(\nu),\nu)q(r;\alpha,u(\nu),\nu)$$

where

$$f_r(\xi_1,\ldots,\xi_{s-1};\alpha,u(\nu),\nu) := \int\limits_0^\infty \xi_s^{s-1} g_r\Big(\xi_1\xi_s,\ldots,\xi_{s-1}\xi_s,\Big(1-\sum_{i=1}^{s-1}\xi_i\Big)\xi_s;\alpha,u(\nu),\nu\Big)\,d\xi_s \tag{21}$$

with g_r as in (12).

Furthermore, there exist integrable functions $\underline{f}(\xi_1,\ldots,\xi_{s-1})$ and $\bar{f}(\xi_1,\ldots,\xi_{s-1})$ on S^{r} such that for all $\alpha\in A$, $u\in\mathcal{U}$ and $\nu\in S$,

$$0<\underline{f}(\xi_1,\ldots,\xi_{s-1})\le f_r(\xi_1,\ldots,\xi_{s-1};\alpha,u(\nu),\nu)\le \bar{f}(\xi_1,\ldots,\xi_{s-1}). \tag{22}$$

Proof. Equations (20) and (21) are immediate consequences of (19) and (12), while inequality (22) can be easily derived from (13) using the change of variables (16) to (18) to prove integrability. ∎

Lemma 2. *For any continuous function F on S^{r} the map*

$$V\times S\ni(v,\nu) \to \int\limits_{S^{\mathrm{r}}} F(\xi_1,\ldots,\xi_{s-1})f(\xi_1,\ldots,\xi_{s-1};\alpha,v,\nu)\,d\xi_1\ldots d\xi_{s-1} \tag{23}$$

is continuous.

Proof. Using (21), (12), (8) and assumption (A.2), it is easily seen that $f(\xi_1,\ldots,\xi_{s-1};\alpha,v,\nu)$ is a continuous function of (v,ν). The result then follows using the last inequality in (22) and the Lebesgue dominated convergence theorem. ∎

In what follows the transition kernel $\Pi^{v\alpha}$ will also play the role of an operator and we use the notation $\Pi^{v\alpha}w(\nu)=\int w(\eta)\Pi^{v\alpha}(\nu,d\eta)$ for any integrable function $w:S\to\mathbb{R}$.

Corollary 1. *For each $\alpha\in A$ there exists a continuous function $w^\alpha: S\to\mathbb{R}$, a constant λ^α and a control function $u^\alpha\in\mathcal{U}$ such that for each $\nu\in S$,*

$$w^\alpha(\nu)+\lambda^\alpha=\inf_{v\in V}\{\Pi^{v\alpha}w^\alpha(\nu)+C(\nu,v)\}=\Pi^{u^\alpha\alpha}w^\alpha(\nu)+C(\nu,u^\alpha(\nu)) \tag{24}$$

with $C(\cdot,\cdot)$ as in (10). *The constant λ^α is the optimal value for the functional $\bar{J}^\alpha$ in* (10) *and the control u^α is an optimal control rule, namely*

$$\lambda^\alpha=\inf_{\{v_k\}}\bar{J}^\alpha(\{v_k\})=\bar{J}^\alpha(\{u^\alpha(\pi_k)\}) \tag{25}$$

where π_k is the filtering process corresponding to the true parameter α and stationary control law $u^\alpha\in\mathcal{U}$. Furthermore, if for all $\nu\in S$ the control law $u\in\mathcal{U}$ satisfies

$$\Pi^{u\alpha}w^\alpha(\nu)+C(\nu,u(\nu))\le w^\alpha(\nu)+\lambda^\alpha+\varepsilon \tag{26}$$

then u is ε-optimal for $\bar{J}^\alpha$, namely

$$\bar{J}^\alpha(\{u(\pi_k)\}) \le \lambda^\alpha + \varepsilon \,. \tag{27}$$

Finally, for all $u \in \mathcal{U}$,

$$\limsup_{n\to\infty} n^{-1} \sum_{k=0}^{n-1} C(\pi_k, u^\alpha(\pi_k)) \ge \lambda^\alpha \qquad P\text{-a.s.} \tag{28}$$

with equality holding for $u = u^\alpha$.

Proof. Equations (24) to (27) follow from [8, Theorem 3]. In fact the assumptions in [8, Remark 4] or [9, Ch. 3] hold since

(i) $c(x,\cdot)$ in (4) is a continuous function on the compact set V;

(ii) the mapping in (23) is continuous;

(iii) by Proposition 1, there exists a nontrivial measure ϕ on $\mathcal{B}(S)$ such that for each $u \in \mathcal{U}$, $\alpha \in A$ and $B \in \mathcal{B}(S)$, we have

$$\phi(B) \le \Pi^{u\alpha}(\nu, B) \,. \tag{29}$$

Finally, inequality (28) can be easily derived from (24) and the law of large numbers for the martingale [5, Theorem VII.9.3]

$$\sum_{k=0}^{n-1} [\Pi^{u\alpha} w^\alpha(\pi_k) - w^\alpha(\pi_{k+1})] \,. \ \blacksquare$$

In what follows we say that a control is *optimal* [*ε-optimal*] for α if it is optimal [ε-optimal] when the value of the actual parameter is α.

Corollary 2 [uniform ergodicity]. *There exists a constant $0 < \gamma < 1$ and measures $\Phi^{u\alpha}$ on $\mathcal{B}(S)$, for all admissible $u \in \mathcal{U}$ and $\alpha \in A$, such that*

$$\sup_{u\in\mathcal{U}} \sup_{\alpha\in A} \sup_{\nu\in S} \sup_{B\in\mathcal{B}(S)} |(\Pi^{u\alpha})^n(\nu, B) - \Phi^{u\alpha}(B)| < \gamma^n \tag{30}$$

where $(\Pi^{u\alpha})^n$ is the n-th iterate of the operator $\Pi^{u\alpha}$.

Proof. The result follows directly from equation (5.6) of Chapter V in [4]. In fact, using the minorization property (29), it is possible to see that condition (D′), required in Case b) there, is satisfied. ■

Denoting by ϱ_A the metric in A and letting $B(\alpha_i, \delta) := \{\alpha \in A : \varrho_A(\alpha, \alpha_i) < \delta\}$, we recall that a set $\{\alpha_1, \ldots, \alpha_r\}$ is called a *δ-net* of A if $\bigcup_i B(\alpha_i, \delta) \supset A$. We have the following

Theorem 1. *For each $\varepsilon > 0$ there exists a $\delta > 0$ such that for $\varrho_A(\alpha, \alpha') < \delta$,*

$$\sup_{u\in\mathcal{U}} \sup_{B\in\mathcal{B}(S)} |\Phi^{u\alpha}(B) - \Phi^{u\alpha'}(B)| < \varepsilon \,. \tag{31}$$

Proof. It is possible to proceed analogously to what has been done in [18, Proposition 1], where arguments from [13] are suitably adapted. The crucial point in order to exploit these results is the proof of the uniform continuity in α of $\Pi^{u\alpha}$. We then have to show that for all $\varepsilon > 0$ there exists a $\delta > 0$ such that for all α, α' with $\varrho_A(\alpha, \alpha') < \delta$,

$$\sup_{u\in\mathcal{U}} \sup_{\nu\in S} \sup_{B\in\mathcal{B}(S)} |\Pi^{u\alpha}(\nu, B) - \Pi^{u\alpha'}(\nu, B)| < \varepsilon\,. \tag{32}$$

We have

$$\begin{aligned}&|\Pi^{u\alpha}(\nu, B) - \Pi^{u\alpha'}(\nu, B)|\\ &\le \int_{B^{\mathrm{r}}} |f(\xi_1, \ldots, \xi_{s-1}; \alpha, u(\nu), \nu) - f(\xi_1, \ldots, \xi_{s-1}; \alpha', u(\nu), \nu)|\, d\xi_1 \ldots d\xi_{s-1}\,.\end{aligned}$$

Taking into account that the majorization property given by the last inequality in (22) implies uniform integrability of $f(\xi_1, \ldots, \xi_{s-1}; \alpha, v, \nu)$ and that by (A.2), (21) and (12) it is uniformly continuous in α, v and ν, (32) follows immediately. ■

COROLLARY 3. *For each $\varepsilon > 0$ there exists $\delta > 0$ and a δ-net $\{\alpha_1, \ldots, \alpha_r\} \subset A$ such that if the control function $u_i \in \mathcal{U}$ is $\varepsilon/2$-optimal for α_i and $\varrho_A(\alpha, \alpha_i) < \delta$, then u_i is ε-optimal for α.*

Proof. As in Corollary 1, denote by u^α the optimal control for α. Let $\|c\| = \sup_{i\in E} \sup_{v\in V} c(i, v)$ and take $\delta > 0$ such that (31) holds for $\varepsilon/(4\|c\|)$. Then, using (30), we have

$$\begin{aligned}\bar{J}^\alpha(\{u_i(\pi_k)\}) &= \int_S C(\nu, u_i(\nu))\, \Phi^{u_i\alpha}(d\nu)\\ &\le \int_S C(\nu, u_i(\nu))|\Phi^{u_i\alpha}(d\nu) - \Phi^{u_i\alpha_i}(d\nu)| + \int_S C(\nu, u_i(\nu))\, \Phi^{u_i\alpha_i}(d\nu)\\ &\le \varepsilon/4 + \int_S C(\nu, u^{\alpha_i}(\nu))\, \Phi^{u^{\alpha_i}\alpha_i}(d\nu) + \varepsilon/2\\ &\le 3\varepsilon/4 + \int_S C(\nu, u^{\alpha}(\nu))\, \Phi^{u^{\alpha}\alpha_i}(d\nu)\\ &\le 3\varepsilon/4 + \int_S C(\nu, u^{\alpha}(\nu))|\Phi^{u^{\alpha}\alpha_i}(d\nu) - \Phi^{u^{\alpha}\alpha}(d\nu)|\\ &\quad + \int_S C(\nu, u^{\alpha}(\nu))\, \Phi^{u^{\alpha}\alpha}(d\nu)\\ &\le \varepsilon + \lambda^\alpha\,.\end{aligned}$$

Due to the compactness of A the existence of $\{\alpha_1, \ldots, \alpha_r\}$ follows.

3. Robustness of controls. Ergodicity of the "state-filter" pair. As mentioned in the introduction, in applications we are not able to calculate the exact value of the filter π_k corresponding to the true value α_0, since this value is unknown. Therefore we cannot use the filter given in (9), but we have to resort to an approximate filter described by the recursive equation

$$\pi^{\alpha}_{k+1} = G^{\alpha}(\pi^{\alpha}_k, y_{k+1}, v_k) \tag{33}$$

in which we assume for the moment that α is close to α_0. We then use a control sequence of the form $v_k = u(\pi^{\alpha}_k)$. If $\alpha = \alpha_0$ then the approximate filter coincides with the optimal filter π_k. On the other hand, if $\alpha \neq \alpha_0$ it is clear from the derivation of (11) that π^{α}_k is no longer a Markov process and in order to exploit ergodicity results it is necessary to augment the state vector by including the process x^{α}_k, namely the original signal process driven by controls based on the values of the approximate filter.

It is also conceivable that the initial measure μ_0 is known only approximately so that we assume that the initial condition of the approximate filter is given by a measure μ possibly different from μ_0; however, this is not explicitly indicated in the notation. Also notice that, although not explicitly indicated, π^{α}_k and x^{α}_k depend on α_0 since x^{α}_k also evolves according to $P^{v\alpha_0}$.

By arguments analogous to those used for the derivation of (11), (12), (20) and (21), denoting by $\mathcal{P}(E)$ the class of all subsets of E we have the following

Lemma 3. *For all $u \in \mathcal{U}$ the pair $[x^{\alpha}_k,\ \pi^{\alpha}_k]$ is a Markov process with transition kernel given by*

$$\begin{aligned} &\Gamma^{u\alpha}(i,\nu,F,B) \\ &\quad = \sum_{j\in F} P^{u(\nu)\alpha_0}(i,j) \int\limits_{B^{\mathrm{r}}} f_j(\xi_1,\ldots,\xi_{s-1};\alpha,u(\nu),\nu)\,d\xi_1\ldots d\xi_{s-1} \end{aligned} \tag{34}$$

for $i \in E$, $\nu \in S$ and $F \in \mathcal{P}(E)$, $B \in \mathcal{B}(S)$.

The following proposition is the analogue of Corollary 2 and Theorem 1 in the case when the approximate filter (33) is used in the control procedure.

Proposition 2. *There exists a constant $0 < \gamma < 1$ and measures $\Psi^{u\alpha}$ on $\mathcal{P}(E)\times\mathcal{B}(S)$ for all $u \in \mathcal{U}$ and $\alpha \in A$ such that*

$$\sup_{u\in\mathcal{U}}\sup_{\alpha\in A}\sup_{i\in E}\sup_{\nu\in S}\sup_{F\in\mathcal{P}(S)}\sup_{B\in\mathcal{B}(S)} |(\Gamma^{u\alpha})^n(i,\nu,F,B) - \Psi^{u\alpha}(F,B)| < \gamma^n\,. \tag{35}$$

Furthermore, for all $\varepsilon > 0$ there exist $\delta > 0$ such that for $\varrho_A(\alpha,\alpha') < \delta$,

$$\sup_{u\in\mathcal{U}}\sup_{F\in\mathcal{P}(S)}\sup_{B\in\mathcal{B}(S)} |\Psi^{u\alpha}(F,B) - \Psi^{u\alpha'}(F,B)| < \varepsilon\,. \tag{36}$$

P r o o f. Analogously to (30), inequality (35) follows immediately from equation (5.6) of Chapter 5 in [4]. Similarly to (31) the uniform continuity (36) can be obtained analogously to what has been done in [18, Proposition 1]. Again the crucial point is the proof of the uniform continuity in α of $\Gamma^{u\alpha}$. This is perfectly analogous to the proof of (32). ∎

COROLLARY 4. *For each $\varepsilon > 0$, there exists a $\delta > 0$ such that for $\varrho_A(\alpha, \alpha_0) < \delta$ we have*

$$\bar{J}^{\alpha_0}(\{u(\pi_k^\alpha)\}) \le \bar{J}^{\alpha_0}(\{u(\pi_k)\}) + \varepsilon\,.$$

In particular, if $u \in \mathcal{U}$ is ε-optimal for α_0, then the control $u(\pi_k^\alpha)$ is 2ε-optimal.

P r o o f. Define $\|c\| = \sup_{i\in E}\sup_{v\in V} c(i,v)$ and take $\delta > 0$ such that (36) holds for $\varepsilon/\|c\|$. Using (35) we have

$$\begin{aligned}
\bar{J}^{\alpha_0}(\{u(\pi_k^\alpha)\}) &= \sum_{i=1}^{s} \int_S c(i,u(\nu))\,\Psi^{u\alpha}(i,d\nu)\\
&\le \sum_{i=1}^{s} \int_S c(i,u(\nu))|\Psi^{u\alpha}(i,d\nu) - \Psi^{u\alpha_0}(i,d\nu)|\\
&\quad + \sum_{i=1}^{s} \int_S c(i,u(\nu))\,\Psi^{u\alpha_0}(i,d\nu)\\
&\le \varepsilon + \int_S C(\nu,u(\nu))\,\Phi^{u\alpha_0}(d\nu) = \varepsilon + \bar{J}^{\alpha_0}(\{u(\pi_k)\})\,. \ \blacksquare
\end{aligned}$$

Considering in Corollary 4 a $\delta' > 0$ corresponding to $\varepsilon' = \varepsilon/8$ and in Corollary 3 a $\delta'' > 0$ corresponding to $\varepsilon'' = \varepsilon'$ and then taking $\delta = \min\{\delta', \delta''\}/2$, we have immediately

COROLLARY 5. *For each $\varepsilon > 0$, one can choose a partition $\{A_i : i = 1,\ldots,r\}$ of A, representative elements $\alpha_i \in A_i$ and control functions $u_i : S \to V$, $i = 1,\ldots,r$, such that if $\alpha \in A_i$ then the control $u_i(\pi_k^\alpha)$ is $\varepsilon/4$-optimal for $\alpha' \in A_i$. In particular, if α_0 belongs to A_i then the control is $\varepsilon/4$-optimal.*

The problem of computation of ε-optimal controls can be studied by the use of techniques developed in [2], [17]; for details and more references see [9].

4. Adaptive control algorithm. In this section we describe an adaptive control procedure that proves nearly optimal for the model considered.

For the purpose of this section we assume that periodically, i.e. for $k = T-1, 2T-1, \ldots$ we obtain the summarized cost incurred up to that time, i.e.

$$\sum_{k=0}^{nT-1} c(x_k, v_k)$$

is given at time nT for all $n = 1, 2, \ldots$

First notice that by (35), for any $\varepsilon > 0$ there is an integer $MT > 0$ such that

$$\sup_{\alpha\in A} \sup_{x_0\in E} \sup_{\pi_0\in S} \sup_{u\in\mathcal{U}} \Big| (MT)^{-1}\mathbf{E}_{x_0\pi_0} \sum_{k=0}^{MT-1} c(x_k^\alpha, u(\pi_k^\alpha)) - \sum_{j=1}^{s} \int_S c(j, u(z))\, \Psi^{u\alpha}(j, dz)\Big| \le \varepsilon/4\,. \tag{37}$$

Then choose a sequence of integers $k_1, k_2, \ldots$ with $k_i \ge r$, and r as in Corollary 5, such that

$$n\Big/\sum_{i=1}^{n} k_i \to 0 \qquad \text{as } n \to \infty\,, \tag{38}$$

and let the sequence $\{a_i : i = 1, 2, \ldots\}$ be defined by $a_0 = 0$, $a_{i+1} - a_i = k_i MT$.

The adaptive procedure is based on the conclusion of Corollary 5 and is the following.

Starting from $k = a_0 = 0$ we use controls $u_i(\pi_k^\alpha)$ for $k \in [a_0 + (i-1)MT,\ a_0 + iMT)$, $i = 1, \ldots, r$, respectively. Then, if $rMT < a_1$, we compare the average costs incurred in the intervals $[a_0 + (i-1)MT,\ a_0 + iMT)$ using $u_i(\pi_k^{\alpha_i})$, $i = 1, \ldots, r$, and in $[rMT, (r+1)MT)$ we use the control corresponding to the minimal cost. If $(r+1)MT < a_1$, we analogously compare the average costs incurred in *all* the past intervals by the controls u_i, and again for $k \in [(r+1)MT,\ (r+2)MT)$ we use the control corresponding to the minimum average cost. We proceed in this way until a_1 is reached.

Then we use again all controls $u_i(\pi_k^{\alpha_i})$, $i = 1, \ldots, r$, in the intervals $[a_1 + (i-1)MT,\ a_1 + iMT)$ respectively. Afterwards, if $a_1 + rMT < a_2$, in time intervals of length MT we apply the controls for which the average costs incurred in *all* the previous intervals were minimal.

In general, when a point a_j is reached, we first "test" all the controls $u_i(\pi_k^{\alpha_i})$ in the intervals $[a_j + (i-1)MT,\ a_j + iMT)$ and then we use the controls corresponding to the minimal average past cost, and this procedure is continued until a_{j+1} is reached.

Notice that the algorithm is based on the idea of comparing the average past costs incurred by the various controls and following the "leader". In order to be able to compare all possible controls we force the use of every control in the "testing intervals" $[a_i, a_i + rMT)$. Because of (38) these intervals are sparse and their influence on the final cost is going to become negligible in the long run.

Denoting by v_k^* the controls resulting from the described procedure, we have the following

THEOREM 2. *There exist $N \subset \Omega$ with $P\{N\} = 0$ such that for $\omega \in \Omega \setminus N$ we have*

$$J^{\alpha_0}(\{v_k^*\}) \le \inf_{\{v_k\}} \bar{J}^{\alpha_0}(\{v_k\}) + \varepsilon\,.$$

Proof. Let β_k denote the index of control used at time k. Define recursively

$$\sigma_1(i) = \inf\{k \ge 0 : \beta_{kMT} = i\}, \qquad \sigma_{n+1}(i) = \inf\{k > \sigma_n(i) : \beta_{kMT} = i\}\,.$$

By the law of large numbers for martingales, we find that for $i = 1, \ldots, r$ and $\omega \in \Omega \setminus N_1$ with $P\{N_1\} = 0$,

$$\tag{39} m^{-1} \sum_{j=1}^{m} \Big[\sum_{k=\sigma_j(i)MT}^{(\sigma_j(i)+1)MT-1} c(x_k^{\alpha_i}, u_i(\pi_k^{\alpha_i})) - \mathbf{E}_{x^{\alpha_i}_{\sigma_j(i)MT}\pi^{\alpha_i}_{\sigma_j(i)MT}} \Big\{ \sum_{k=0}^{MT-1} c(x_k^{\alpha_i}, u_i(\pi_k^{\alpha_i}))\Big\}\Big] \to 0 \qquad \text{as } m \to \infty\,.$$

Since by (37),

$$\Big| \mathbf{E}_{x^{\alpha_i}_{\sigma_j(i)MT}\pi^{\alpha_i}_{\sigma_j(i)MT}} \Big\{ \sum_{k=0}^{MT-1} c(x_k^{\alpha_i}, u_i(\pi_k^{\alpha_i}))\Big\} - MT \sum_{j=1}^{s} \int_S c(j, u(z))\, \Psi^{u_i\alpha_i}(j, dz)\Big| \le \varepsilon MT/4\,,$$

from (39) we have

$$\tag{40} \limsup_{m\to\infty} m^{-1} \Big| \sum_{j=1}^{m} \Big[\sum_{k=\sigma_j(i)MT}^{(\sigma_j(i)+1)MT-1} c(x_k^{\alpha_i}, u_i(\pi_k^{\alpha_i})) - MT \sum_{j=1}^{s} \int_S c(j, u_i(z))\, \Psi^{u_i\alpha_i}(j, dz)\Big]\Big| \le \varepsilon MT/4\,.$$

Assume now that i is a Cesàro frequent index of control, i.e.

$$\limsup_{n\to\infty} \frac{1}{n} \sum_{k=0}^{n-1} I_{\{\beta_k = i\}} > 0\,.$$

In this case, by (38), the control u_i is also used outside the "testing intervals" $[a_j, a_j + rMT)$, because it corresponds to a minimum average past cost. Therefore, given any other control index i' there exist sequences m_k and m'_k such that

$$\limsup_{k\to\infty} m_k^{-1} \sum_{j=1}^{m_k} \sum_{k=\sigma_j(i)MT}^{(\sigma_j(i)+1)MT-1} c(x_k^{\alpha_i}, u_i(\pi_k^{\alpha_i})) \le \limsup_{k\to\infty} m_k'^{-1} \sum_{j=1}^{m'_k} \sum_{k=\sigma_j(i')MT}^{(\sigma_j(i')+1)MT-1} c(x_k^{\alpha_{i'}}, u_{i'}(\pi_k^{\alpha_{i'}}))$$

so that, using (40) we have for all $i' = 1, \ldots, r$,

$$MT \sum_{j=1}^{s} \int_S c(j, u_i(z))\, \Psi^{u_i\alpha_i}(j, dz) - \varepsilon MT/4 \le MT \sum_{j=1}^{s} \int_S c(j, u_{i'}(z))\, \Psi^{u_{i'}\alpha_{i'}}(j, dz) + \varepsilon MT/4$$

and therefore using Corollary 5 we have

$$\sum_{j=1}^{s} \int_S c(j, u_i(z))\, \Psi^{u_i\alpha_i}(j, dz) \le \lambda^{\alpha_0} + 3\varepsilon/4\,. \tag{41}$$

Consequently, the control $u_i(\pi_k^{\alpha_i})$, corresponding to the Cesàro frequent index i, is $3\varepsilon/4$-optimal.

Denoting by x_k^* and π_k^* the state and filter at time k resulting from the adaptive procedure and using again the law of large numbers for martingales, we have for $\omega \in \Omega \setminus N_2$ with $P\{N_2\} = 0$,

$$(nMT)^{-1} \sum_{i=1}^{n} \Big[\sum_{k=iMT}^{(i+1)MT-1} c(x_k^*, v_k^*) - \mathbf{E}_{x^*_{iMT}\pi^*_{iMT}} \Big\{ \sum_{k=0}^{MT-1} c(x_k^*, v_k^*) \Big\} \Big] \to 0 \qquad \text{as } n \to \infty\,.$$

From this, using (37) we have

$$\limsup_{n\to\infty} n(MT)^{-1} \sum_{i=1}^{n} \sum_{k=iMT}^{(i+1)MT-1} c(x_k^*, v_k^*) \le \limsup_{n\to\infty} n^{-1} \sum_{i=1}^{n} \sum_{j=1}^{r} I_{\{\beta_{iMT}=j\}} \sum_{h=1}^{s} \int_S c(h, u_j(z))\, \Psi^{u_j\alpha_j}(j, dz) + \varepsilon/4\,. \tag{42}$$

Denoting by C the set of Cesàro frequent indices of control, we have

$$\limsup_{n\to\infty} n^{-1}\sum_{i=1}^{n}\sum_{j\notin C} I_{\{\beta_{iMT}=j\}} = 0$$

and consequently

$$\limsup_{n\to\infty} n^{-1}\sum_{i=1}^{n}\sum_{j\in C} I_{\{\beta_{iMT}=j\}} = 1\,.$$

Then from (42) and (41) we conclude that the adaptive procedure is ε-optimal for $\omega \in \Omega/N$ with $N = N_1 \cup N_2$. ■

From Theorem 1, a direct application of Fatou's lemma provides the following

COROLLARY 6. *The controls v_k^* are such that*

$$\bar{J}^{\alpha_0}(\{v_k^*\}) \le \inf_{\{v_k\}} \bar{J}^{\alpha_0}(\{v_k\}) + \varepsilon\,. \quad ■ \tag{43}$$

This result shows that the controls resulting from the adaptive procedure are ε-optimal for the functional $\bar{J}^{\alpha_0}$. By (28) they are also ε-optimal for the functional J^{α_0}.

References

[1] A. Arapostathis and S. I. Marcus, *Analysis of an identification algorithm arising in the adaptive estimation of Markov chains*, Math. Control Signals Systems 3 (1990), 1–29.

[2] V. V. Baranov, *A recursive algorithm in Markovian decision processes*, Cybernetics 18 (1982), 499–506.

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

[4] J. L. Doob, *Stochastic Processes*, Wiley, New York, 1953.

[5] W. Feller, *An Introduction to Probability Theory and Its Applications II*, Wiley, New York, 1971.

[6] E. Fernández-Gaucherand, A. Arapostathis and S. I. Marcus, *On the adaptive control of a partially observable Markov decision process*, in: Proc. 27th IEEE Conf. on Decision and Control, 1988, 1204–1210.

[7] —, —, —, *On the adaptive control of a partially observable binary Markov decision process*, in: Advances in Computing and Control, W. A. Porter, S. C. Kak and J. L. Aravena (eds.), Lecture Notes in Control and Inform. Sci. 130, Springer, New York, 1989, 217–228.

[8] L. G. Gubenko and E. S. Shtatland, *On discrete-time Markov decision processes*, Theory Probab. Math. Statist. 7 (1975), 47–61.

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

[10] O. Hernández-Lerma and S. I. Marcus, *Adaptive control of Markov processes with incomplete state information and unknown parameters*, J. Optim. Theory Appl. 52 (1987), 227–241.
[11] —, —, *Nonparametric adaptive control of discrete-time partially observable stochastic systems*, J. Math. Anal. Appl. 137 (1989), 312–334.
[12] A. H. Jazwinski, *Stochastic Processes and Filtering Theory*, Academic Press, New York, 1970.
[13] N. W. Kartashov, *Criteria for uniform ergodicity and strong stability of Markov chains in general state space*, Theory Probab. Math. Statist. 30 (1985), 71–89.
[14] P. R. Kumar and P. Varaiya, *Stochastic Systems*: *Estimation*, *Identification and Adaptive Control*, Prentice-Hall, Englewood Cliffs, 1986.
[15] H. J. Kushner and H. Huang, *Approximation and limit results for nonlinear filters with wide bandwidth observation noise*, Stochastics 16 (1986), 65–96.
[16] G. E. Monahan, *A survey of partially observable Markov decision processes: theory, models and algorithms*, Management Sci. 28 (1982), 1–16.
[17] W. J. Runggaldier and Ł. Stettner, *Nearly optimal controls for stochastic ergodic problems with partial observation*, SIAM J. Control Optim. 31 (1993), 180–218.
[18] Ł. Stettner, *On nearly self-optimizing strategies for a discrete-time uniformly ergodic adaptive model*, J. Appl. Math. Optim. 27 (1993), 161–177.

GIOVANNI B. DI MASI
DIPARTIMENTO DI MATEMATICA
PURA ED APPLICATA AND CNR-LADSEB
UNIVERSITÀ DI PADOVA
I-35100 PADOVA, ITALY

ŁUKASZ STETTNER
INSTITUTE OF MATHEMATICS
POLISH ACADEMY OF SCIENCES
ŚNIADECKICH 8
00-950 WARSZAWA, POLAND

Received on 2.11.1992