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 self-optimizing 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 \( P^{\alpha_0}(i,j) \), where \( i, j \in E = \{1, \ldots, s\} \), the control \( v \) lies in a compact metric space \( V \) and the parameter \( \alpha_0 \) belongs to a compact metric space \( A \). The initial state \( x_0 \) is assumed to be distributed according to a given initial law \( \mu_0 \) and \( \alpha_0 \) stands for an unknown parameter. The process \( x_k \) is partially observed via the \( s \)-dimensional process

\[
y_k = h(x_k) + w_k
\]

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

\[
h^i(j) = \begin{cases} 
0 & \text{for } j \neq i, \\
h_i & \text{for } j = i.
\end{cases}
\]

Therefore the information available at time \( k \) is provided by the \( \sigma \)-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 \( i \)th 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 \( Y_k \)-measurable \( V \)-valued random variable. The purpose of the control is to minimize the pathwise long run average cost

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

or the long run expected average cost

\[
J^{\alpha_0}(\{v_k\}) = \limsup_{n \to \infty} n^{-1} \mathbb{E}^\mu_0 \sum_{k=0}^{n-1} c(x_k, v_k)
\]

where \( c(x, \cdot) \) is a continuous function and \( \mathbb{E}^\mu_0 \) denotes expectation with respect to the probability \( P^\mu_0 \) induced on the space of trajectories by the process \( x_k \) with parameter \( \alpha_0 \), 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

\[
\pi_k(i) = P^\mu_0 \{ x_k = i \mid Y_k \}.
\]

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

\[
\pi_k(i) = \frac{\sigma_k(i)}{\sum_{j=1}^{\infty} \sigma_k(j)}
\]

(see [12]) where

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

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

\[
q(i; \alpha_0, v_{k-1}, \pi_{k-1}) = P^\mu_0 \{ x_k = i \mid Y_{k-1} \} = \sum_{j=1}^{s} \mathbb{P}^{v_{k-1}}(j, i) \pi_{k-1}(j)
\]

and initial condition given by

\[
\pi_0(i) = \mu_0(i).
\]

The recursive formula for the filter is concisely written as

\[
\pi_{k+1} = G^{\alpha_0}(\pi_k, y_{k+1}, v_k),
\]
which provides the evolution of the “completely observed” state $\pi_k$. The control functional $J^{\alpha_0}$ can be written in terms of $\pi_k$ as

$$ J^{\alpha_0}(\{v_k\}) = \lim sup_{n \to \infty} n^{-1} \sum_{k=0}^{n-1} \sum_{i=1}^{s} c(i, v_k) \pi_k(i) $$

$$ =: \lim sup_{n \to \infty} n^{-1} \sum_{k=0}^{n-1} C(\pi_k, v_k). $$

It is shown in Corollary 1 that the optimal controls for the cost function (10) are functions of the “completely observed” state $\pi_k$; more precisely, denoting by $S$ the $(s-1)$-dimensional simplex and by $\mathcal{B}(S)$ the Borel $\sigma$-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. $P^u(i, j)$ is used for $P^u(i, j)_{\alpha_0}$. In what follows we require the following assumptions:

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

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

The aim of the paper is to find a control procedure which guarantees the $\varepsilon$-optimal value of the cost functional for the state process $x_k$ corresponding to the unknown parameter $\alpha_0$. In the next Section 2, using uniform ergodicity arguments, we show that for each $\varepsilon$, 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 $\varepsilon$-optimal for $\alpha$. 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 $\alpha_0$ 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 $\alpha$, 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 Rung-
galdi 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 $\pi_k$ corresponding to a generic value of the parameter $\alpha$. For this purpose we need to investigate some properties of the transition kernel for $\pi_k$, which in turn are derived from the corresponding properties of the unnormalized filter $\sigma_k$.

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

$$ P^\alpha\{\sigma_k(i) \leq z_i, \; i = 1, \ldots, s \mid Y_{k-1}\} = P^\alpha\{w_k^i \leq -h^i(x_k) + h^{-1}_i[\ln z_i - \ln q(i; \alpha, v_{k-1}, \pi_{k-1}) + \frac{1}{2}h_i], \; i = 1, \ldots, s \mid Y_{k-1}\} $$

$$ = \sum_{\mu=1}^{s} P^\alpha\{w_k^i \leq -h^i(r) + h^{-1}_i[\ln z_i - \ln q(i; \alpha, v_{k-1}, \pi_{k-1}) + \frac{1}{2}h_i], \; i = 1, \ldots, s \mid Y_{k-1}, x_k = \mu\} P^\alpha_{\mu}\{x_k = r \mid Y_{k-1}\} $$

$$ = \sum_{\mu=1}^{s} P^\alpha\{w_k^i \leq -h^i(r) + h^{-1}_i[\ln z_i - \ln q(i; \alpha, v_{k-1}, \pi_{k-1}) + \frac{1}{2}h_i], \; i = 1, \ldots, s \mid Y_{k-1}\} q(r; \alpha, v_{k-1}, \pi_{k-1}). $$

The conditional distribution function given by (11) has density

$$ g(z_1, \ldots, z_s; \alpha, v_{k-1}, \pi_{k-1}) := \frac{d^s}{dz_1 \ldots dz_s} P^\alpha\{\sigma_k(i) \leq z_i, \; i = 1, \ldots, s \mid Y_{k-1}\} $$

$$ = \sum_{\mu=1}^{s} \frac{1}{(2\pi)^{s/2}} \prod_{i=1}^{s} \exp\left[-\frac{1}{2}\left[-h^i(r) + h^{-1}_i(\ln z_i - \ln q(i; \alpha, v_{k-1}, \pi_{k-1}) + \frac{1}{2}h_i)\right] \right] q(r; \alpha, v_{k-1}, \pi_{k-1}) $$

$$ = \sum_{\mu=1}^{s} g(r(z_1, \ldots, z_s; \alpha, v_{k-1}, \pi_{k-1}) q(r; \alpha, v_{k-1}, \pi_{k-1}).$$
We have the following

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

\[
0 < g(z_1, \ldots, z_s) \leq g_r(z_1, \ldots, z_s; \alpha, v, \nu) \leq \bar{g}(z_1, \ldots, z_s).
\]

**Proof.** Define

\[
d_k(i, r) = \ln \frac{q(i; \alpha, v_{k-1}, \pi_{k-1})}{\sqrt{2 \pi \sigma_i^2}} - \frac{1}{2} h_i^2(r) h_i.
\]

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

\[
d(i, r) = \ln \beta - \frac{1}{2} h_i^2(r) h_i, \quad \bar{d}(i, r) = -\frac{1}{2} h_i^2(r) h_i.
\]

Consider now the set of functions

\[
D(i, r) = \{ [\ln z_i - d(i, r)][\ln z_i - \bar{d}(i, r)]; [\ln z_i - d(i, r)][\ln z_i - \bar{d}(i, r)]; [\ln z_i - \bar{d}(i, r)][\ln z_i - \bar{d}(i, r)] \}.
\]

Then

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

and as a consequence

\[
g(z_1, \ldots, z_s) \leq g_r(z_1, \ldots, z_s; \alpha, v_{k-1}, \pi_{k-1}) \leq \bar{g}(z_1, \ldots, z_s)
\]

where

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

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

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

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

\[
B^c = \left\{ (\xi_1, \ldots, \xi_{s-1}) \in \mathbb{R}_+^{s-1} : (\xi_1, \ldots, \xi_{s-1}, 1 - \sum_i \xi_i) \in B \right\}
\]

and

\[
B^n = B^c \times \mathbb{R}_+.
\]

From (6) we deduce that the conditional probabilities \( \pi_k(i), i = 1, \ldots, s-1 \), can be obtained from the unnormalized probabilities \( \sigma \) using the
transformation \( H : \mathbb{R}^s_+ \rightarrow S^o := S^s \times \mathbb{R}_+ \) given by
\[
\begin{bmatrix}
\xi_1 \\
\vdots \\
\xi_{s-1} \\
\xi_s
\end{bmatrix} = H
\begin{bmatrix}
z_1 \\
\vdots \\
z_{s-1} \\
z_s
\end{bmatrix} = \begin{bmatrix}
z_1/\sum_i z_i \\
\vdots \\
z_{s-1}/\sum_i z_i \\
\sum_i z_i
\end{bmatrix}.
\]

Clearly
\[
\begin{bmatrix}
z_1 \\
\vdots \\
z_{s-1} \\
z_s
\end{bmatrix} = H^{-1}
\begin{bmatrix}
\xi_1 \\
\vdots \\
\xi_{s-1} \\
\xi_s
\end{bmatrix} = \begin{bmatrix}
\xi_1 \xi_s \\
\vdots \\
\xi_{s-1} \xi_s \\
(1 - \sum_{i=1}^{s-1} \xi_i) \xi_s
\end{bmatrix}
\]
and its associated Jacobian is given by
\[
\frac{\partial H^{-1}}{\partial \xi} = \xi_s^{-1}.
\]

We have
\[
P_{\mu_0}^o \{ \pi_k \in B \mid \mathcal{Y}_{k-1} \} = P\{ H(\sigma_k) \in B^o \mid \mathcal{Y}_{k-1} \}
= P_{\mu_0}^o \{ \sigma_k \in H^{-1}(B^o) \mid \mathcal{Y}_{k-1} \}
= \int_{H^{-1}(B^o)} g(z_1, \ldots, z_s; \alpha, u_{k-1}, \pi_{k-1}) dz_1 \ldots dz_s
= \int_{B^o} \int_0^\infty \int_0^s \xi_s^{-1}
\times g(\xi_1, \xi_s, \ldots, \xi_{s-1} \xi_s, \left(1 - \sum_{i=1}^{s-1} \xi_i\right) \xi_s; \alpha, u_{k-1}, \pi_{k-1})
d\xi_s \, d\xi_1 \ldots d\xi_{s-1}.
\]

**Proposition 1.** The filtering process \( \pi_k \) corresponding to an admissible control function \( u : S \rightarrow V \) is a Markov process with respect to the \( \sigma \)-field \( \mathcal{Y}_k \) with transition kernel
\[
\Pi^{u\alpha}(\nu, B) = \int_B f(\xi_1, \ldots, \xi_{s-1}; \alpha, u(\nu), \nu) d\xi_1 \ldots d\xi_{s-1}
\]
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
\begin{equation}
\begin{aligned}
f_r(\xi_1, \ldots, \xi_{s-1}; \alpha, u(\nu), \nu) := \\
\int_0^{\xi_s} g_r(\xi_1, \xi_2, \ldots, \xi_{s-1} + \xi_s - \sum_{i=1}^{s-1} \xi_i; \alpha, u(\nu), \nu) \, d\xi_s
\end{aligned}
\end{equation}

with \( g_r \) as in (12).

Furthermore, there exist integrable functions \( f(\xi_1, \ldots, \xi_{s-1}) \) and \( f_r(\xi_1, \ldots, \xi_{s-1}; \alpha, u(\nu), \nu) \) on \( \mathbb{S}^r \) such that for all \( \alpha \in \mathbb{A} \), \( u \in \mathcal{U} \) and \( \nu \in \mathbb{S} \),
\begin{equation}
0 < f(\xi_1, \ldots, \xi_{s-1}) \leq f_r(\xi_1, \ldots, \xi_{s-1}; \alpha, u(\nu), \nu) \leq f(\xi_1, \ldots, \xi_{s-1}).
\end{equation}

\textbf{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.

\textbf{Lemma 2.} For any continuous function \( F \) on \( \mathbb{S}^r \) the map
\begin{equation}
V \times \mathbb{S} \ni (v, \nu) \\
\rightarrow \int_{\mathbb{S}^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}
\end{equation}
is continuous.

\textbf{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 \( (v, \nu) \). The result then follows using the last inequality in (22) and the Lebesgue dominated convergence theorem.

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

\textbf{Corollary 1.} For each \( \alpha \in \mathbb{A} \) there exists a continuous function \( w^\alpha: \mathbb{S} \rightarrow \mathbb{R} \), a constant \( \lambda^\alpha \) and a control function \( u^\alpha \in \mathcal{U} \) such that for each \( \nu \in \mathbb{S} \),
\begin{equation}
w^\alpha(\nu) + \lambda^\alpha = \inf_{v \in V} \{ \Pi^{\nu \alpha}w^\alpha(\nu) + C(\nu, v) \} = \Pi^{\nu \alpha}w^\alpha(\nu) + C(\nu, u^\alpha(\nu))
\end{equation}
with \( C(\cdot, \cdot) \) as in (10). The constant \( \lambda^\alpha \) is the optimal value for the functional \( J^\alpha \) in (10) and the control \( u^\alpha \) is an optimal control rule, namely
\begin{equation}
\lambda^\alpha = \inf_{\{v_k\}} J^\alpha(\{v_k\}) = J^\alpha(\{u^\alpha(\pi_k)\})
\end{equation}
where \( \pi_k \) is the filtering process corresponding to the true parameter \( \alpha \) and stationary control law \( u^\alpha \in \mathcal{U} \). Furthermore, if for all \( \nu \in \mathbb{S} \) the control law \( u \in \mathcal{U} \) satisfies
\begin{equation}
\Pi^{\nu \alpha}w^\alpha(\nu) + C(\nu, u(\nu)) \leq w^\alpha(\nu) + \lambda^\alpha + \varepsilon
\end{equation}
then $u$ is $\varepsilon$-optimal for $J^\alpha$, namely
\begin{equation}
J^\alpha(\{u(\pi_k)\}) \leq \lambda^\alpha + \varepsilon.
\end{equation}

Finally, for all $u \in U$,
\begin{equation}
\limsup_{n \to \infty} \frac{1}{n-1} \sum_{k=0}^{n-1} C(\pi_k, u^\alpha(\pi_k)) \geq \lambda^\alpha \text{ P-a.s.}
\end{equation}
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
\begin{enumerate}[(i)]
\item $c(x, \cdot)$ in (4) is a continuous function on the compact set $V$;
\item the mapping in (23) is continuous;
\item by Proposition 1, there exists a nontrivial measure $\phi$ on $\mathcal{B}(S)$ such that for each $u \in U$, $\alpha \in A$ and $B \in \mathcal{B}(S)$, we have
\begin{equation}
\phi(B) \leq \Pi^u\alpha(\nu, B).
\end{equation}
\end{enumerate}

Finally, inequality (28) can be easily derived from (24) and the law of large numbers for the martingale [5, Theorem VII.9.3]
\begin{equation}
\sum_{k=0}^{n-1} \left[\Pi^u\alpha wo(\pi_k) - wo(\pi_{k+1})\right] < \gamma^n.
\end{equation}

In what follows we say that a control is optimal [$\varepsilon$-optimal] for $\alpha$ if it is optimal [$\varepsilon$-optimal] when the value of the actual parameter is $\alpha$.

**Corollary 2** [uniform ergodicity]. There exists a constant $0 < \gamma < 1$ and measures $\Phi^u\alpha$ on $\mathcal{B}(S)$, for all admissible $u \in U$ and $\alpha \in A$, such that
\begin{equation}
\sup_{u \in U} \sup_{\alpha \in A} \sup_{\nu \in S} \sup_{B \in \mathcal{B}(S)} \left|\Pi^u\alpha(\nu, B) - \Phi^u\alpha(B)\right| < \gamma^n.
\end{equation}
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 $\varrho_A$ the metric in $A$ and letting $B(\alpha_1, \delta) := \{\alpha \in A : \varrho_A(\alpha, \alpha_1) < \delta\}$, we recall that a set $\{\alpha_1, \ldots, \alpha_r\}$ is called a $\delta$-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 $\varrho_A(\alpha, \alpha') < \delta$
\begin{equation}
\sup_{u \in U} \sup_{B \in \mathcal{B}(S)} \left|\Phi^u\alpha(B) - \Phi^u\alpha'(B)\right| < \varepsilon.
\end{equation}

**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 $\alpha$ of $\Pi^{u\alpha}$. We then have to show that for all $\varepsilon > 0$ there exists a $\delta > 0$ such that for all $\alpha, \alpha'$ with $g_A(\alpha, \alpha') < \delta$,

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

We have

$$
|\Pi^{u\alpha}(\nu, B) - \Pi^{u\alpha'}(\nu, B)| \leq \int_{\mathcal{B}'} |f(\xi_1, \ldots, \xi_{s-1}; \alpha, u(\nu)) - f(\xi_1, \ldots, \xi_{s-1}; \alpha', u(\nu))| d\xi_1 \ldots d\xi_{s-1}.
$$

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 $\alpha, v$ and $\nu$, (32) follows immediately.

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

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

Then, using (30), we have

$$
J^\alpha(\{u_i(\pi_k)\}) = \int_{\mathcal{S}} C(\nu, u_i(\nu)) \Phi^{u_i, \alpha}(d\nu)
$$

$$
\leq \int_{\mathcal{S}} C(\nu, u_i(\nu))|\Phi^{u_i, \alpha}(d\nu) - \Phi^{u_i, \alpha}(d\nu)| + \int_{\mathcal{S}} C(\nu, u_i(\nu)) \Phi^{u_i, \alpha}(d\nu)
$$

$$
\leq \varepsilon/4 + \int_{\mathcal{S}} C(\nu, u^\alpha(\nu)) \Phi^{u^\alpha, \alpha}(d\nu) + \varepsilon/2
$$

$$
\leq 3\varepsilon/4 + \int_{\mathcal{S}} C(\nu, u^\alpha(\nu)) \Phi^{u^\alpha, \alpha}(d\nu)
$$

$$
\leq 3\varepsilon/4 + \int_{\mathcal{S}} C(\nu, u^\alpha(\nu))\Phi^{u^\alpha, \alpha}(d\nu) - \Phi^{u^\alpha, \alpha}(d\nu)
$$

$$
+ \int_{\mathcal{S}} C(\nu, u^\alpha(\nu)) \Phi^{u^\alpha, \alpha}(d\nu)
$$

$$
\leq \varepsilon + \lambda^\alpha.
$$

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

As mentioned in the introduction, in applications we are not able to calculate the exact value of the filter \( \pi_k \) corresponding to the true value \( \alpha_0 \), 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

\[
\pi_{k+1}^\alpha = G^\alpha(\pi_k^\alpha, y_{k+1}, v_k)
\]

in which we assume for the moment that \( \alpha \) is close to \( \alpha_0 \). We then use a control sequence of the form \( v_k = u(\pi_k^\alpha) \). If \( \alpha = \alpha_0 \) then the approximate filter coincides with the optimal filter \( \pi_k \). On the other hand, if \( \alpha \neq \alpha_0 \) it is clear from the derivation of (11) that \( \pi_k^\alpha \) 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 \( x_k^\alpha \), namely the original signal process driven by controls based on the values of the approximate filter.

It is also conceivable that the initial measure \( \mu_0 \) is known only approximately so that we assume that the initial condition of the approximate filter is given by a measure \( \mu_0^\alpha \) possibly different from \( \mu_0 \); however, this is not explicitly indicated in the notation. Also notice that, although not explicitly indicated, \( \pi_k^\alpha \) and \( x_k^\alpha \) depend on \( \alpha_0 \) since \( x_k^\alpha \) also evolves according to \( P^{v_0} \).

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

**Lemma 3.** For all \( u \in U \) the pair \([x_k^\alpha, \pi_k^\alpha]\) is a Markov process with transition kernel given by

\[
\Gamma^{u,\alpha}(i, \nu, F, B) = \sum_{j \in F} P^{u,\alpha}(i, j) \int f_j(\xi_1, \ldots, \xi_{s-1}; \alpha, u(\nu), \nu) d\xi_1 \ldots d\xi_{s-1}
\]

for \( i \in E \), \( \nu \in S \) and \( F \in P(E) \), \( B \in B(S) \).

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 \( P(E) \times B(S) \) for all \( u \in U \) and \( \alpha \in A \) such that

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

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

\[
(36) \quad \sup_{u \in U} \sup_{F \in P(S)} \sup_{B \in B(S)} |\Psi^{u,\alpha}(F, B) - \Psi^{u,\alpha'}(F, B)| < \varepsilon.
\]
Proof. 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 $\alpha$ of $\Gamma^{\alpha 0}$. 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

$$J^{\alpha 0}(\{u(\pi^0_k)\}) \leq J^{\alpha 0}(\{u(\pi_k)\}) + \varepsilon.$$  

In particular, if $u \in U$ is $\varepsilon$-optimal for $\alpha_0$, then the control $u(\pi^0_k)$ is $2\varepsilon$-optimal.

Proof. 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

$$J^{\alpha 0}(\{u(\pi^0_k)\}) = \sum_{i=1}^{s} \int_{S} c(i, u(\nu)) \psi^{\alpha 0}(i, d\nu)$$

$$\leq \sum_{i=1}^{s} \int_{S} c(i, u(\nu)) [\psi^{\alpha 0}(i, d\nu) - \psi^{\alpha 0}(i, d\nu)]$$

$$+ \sum_{i=1}^{s} \int_{S} c(i, u(\nu)) \psi^{\alpha 0}(i, d\nu)$$

$$\leq \varepsilon + \int_{S} C(\nu, u(\nu)) \psi^{\alpha 0}(d\nu) = \varepsilon + J^{\alpha 0}(\{u(\pi_k)\}).$$

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 $u_i : S \rightarrow V$, $i = 1, \ldots, r$, such that if $\alpha \in A_i$ then the control $u_i(\pi^0_k)$ is $\varepsilon/4$-optimal for $\alpha' \in A_i$. In particular, if $\alpha_0$ belongs to $A_i$ then the control is $\varepsilon/4$-optimal.

The problem of computation of $\varepsilon$-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
\[
\begin{align*}
&\sup_{\alpha \in \mathcal{A}} \sup_{x_0 \in \mathcal{E}} \sup_{\pi_0 \in \mathcal{S}} \sup_{u \in \mathcal{U}} \left| (MT)^{-1} \mathbb{E}_{x_0 \pi_0} \sum_{k=0}^{MT-1} c(x_k^\alpha, u(\pi_k^0)) - \sum_{j=1}^S \int_S c(j, u(z)) \Psi^u(j, dz) \right| \\
&\quad \leq \varepsilon / 4.
\end{align*}
\]
Then choose a sequence of integers \( k_1, k_2, \ldots \) with \( k_i \geq r \), and \( r \) as in Corollary 5, such that
\[
\sum_{i=1}^n k_i \to 0 \quad \text{as } n \to \infty,
\]
and let the sequence \( \{a_i : i = 1, 2, \ldots \} \) be defined by \( a_0 = 0, a_{i+1} - a_i = k_iMT \).

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

Starting from \( k = a_0 = 0 \) we use controls \( u_i(\pi_k^0) \) 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 \( u_i(\pi_k^0) \), \( i = 1, \ldots, r \), 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 \( u_i(\pi_k^0) \), \( i = 1, \ldots, r \), 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 \( u_i(\pi_k^0) \) 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” \([a_i, a_i + r MT]\). 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 \(\Pr\{N\} = 0\) such that for \(\omega \in \Omega \setminus N\) we have

\[
J^{\alpha_0}(\{v^*_k\}) \leq \inf_{\{v_k\}} J^{\alpha_0}(\{v_k\}) + \varepsilon.
\]

**Proof.** Let \(\beta_k\) denote the index of control used at time \(k\). Define recursively

\[
\sigma^1(i) = \inf\{k \geq 0 : \beta_k MT = i\}, \quad \sigma_{n+1}(i) = \inf\{k > \sigma_n(i) : \beta_k MT = i\}.
\]

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

\[
m^{-1}\sum_{j=1}^{m} \left[ \sum_{k = \sigma_j(i)MT}^{(\sigma_j(i)+1)MT-1} c(x_{k}^{\alpha_i}, u_i(\pi_{k}^{\alpha_i})) \right.

\[
- \mathbb{E}_{x_{\sigma_j(i)MT}^{\alpha_i}, \pi_{\sigma_j(i)MT}^{\alpha_i}} \left\{ \sum_{k=0}^{MT-1} c(x_{k}^{\alpha_i}, u_i(\pi_{k}^{\alpha_i})) \right\} \to 0 \quad \text{as} \quad m \to \infty.
\]

Since by (37),

\[
\left| \mathbb{E}_{x_{\sigma_j(i)MT}^{\alpha_i}, \pi_{\sigma_j(i)MT}^{\alpha_i}} \left\{ \sum_{k=0}^{MT-1} c(x_{k}^{\alpha_i}, u_i(\pi_{k}^{\alpha_i})) \right\} \right| - MT \sum_{j=1}^{s} \int_{S} c(j, u(z)) \psi_{u_i(\alpha_i)}(j, dz) \leq \varepsilon MT/4,
\]

from (39) we have

\[
\limsup_{m \to \infty} m^{-1} \left[ \sum_{j=1}^{m} \left[ \sum_{k = \sigma_j(i)MT}^{(\sigma_j(i)+1)MT-1} c(x_{k}^{\alpha_i}, u_i(\pi_{k}^{\alpha_i})) \right.

\[
- \mathbb{E}_{x_{\sigma_j(i)MT}^{\alpha_i}, \pi_{\sigma_j(i)MT}^{\alpha_i}} \left\{ \sum_{k=0}^{MT-1} c(x_{k}^{\alpha_i}, u_i(\pi_{k}^{\alpha_i})) \right\} \right| \right] \leq \varepsilon MT/4.
\]

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

$$\limsup_{k \to \infty} m_k^{-1} \sum_{j=1}^{m_k} (\sigma_j(i) + 1)MT - 1 \sum_{k=\sigma_j(i)MT}^{\sigma_j(i)MT - 1} c(x_i^{\alpha_i}, u_i(\pi_i^{\alpha_i}))$$

$$\leq \limsup_{k \to \infty} m'_k^{-1} \sum_{j=1}^{m'_k} (\sigma_j(i') + 1)MT - 1 \sum_{k=\sigma_j(i')MT}^{\sigma_j(i')MT - 1} c(x_i'^{\alpha_i'}, u_i'(\pi_i'^{\alpha_i'}))$$

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

$$MT \sum_{j=1}^{s} \int_{S} c(j, u_i(z)) \psi^{u_i, \alpha_i}(j, dz) - \varepsilon MT/4$$

$$\leq \limsup_{n \to \infty} n^{(i+1)MT - 1} \sum_{k=iMT}^{(i+1)MT - 1} c(x_i^*, v_i^*)$$

$$- \mathbf{E}_{x_i^*, \pi_i^*} \left\{ \sum_{k=0}^{MT-1} c(x_i^*, v_i^*) \right\} \to 0 \quad \text{as } n \to \infty .$$

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

Denoting by $x_i^*$ and $\pi_i^*$ 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 $P\{N_2\} = 0$,

$$\sum_{j=1}^{s} \int_{S} c(j, u_i(z)) \psi^{u_i, \alpha_i}(j, dz) \leq \lambda^{\alpha_i} + 3\varepsilon/4 .$$

From this, using (37) we have

$$\limsup_{n \to \infty} n(MT)^{-1} \sum_{i=1}^{n} \sum_{k=iMT}^{(i+1)MT - 1} c(x_i^*, v_i^*)$$

$$\leq \limsup_{n \to \infty} n^{-1} \sum_{i=1}^{n} \sum_{j=1}^{r} \sum_{h=1}^{s} c(h, u_j(z)) \psi^{u_j, \alpha_j}(j, dz) + \varepsilon/4 .$$
Denoting by $C$ the set of Cesàro frequent indices of control, we have

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

and consequently

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

Then from (42) and (41) we conclude that the adaptive procedure is $\varepsilon$-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

$$J^{\alpha_0}(\{v^*_k\}) \leq \inf_{\{v_k\}} J^{\alpha_0}(\{v_k\}) + \varepsilon. \quad ■$$

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

**References**


