APPLICATIONES MATHEMATICAE
24,2 (1996), pp. 169–188

**V. S. BORKAR (Bangalore)**

# RECURSIVE SELF-TUNING CONTROL OF FINITE MARKOV CHAINS

*Abstract.* A recursive self-tuning control scheme for finite Markov chains is proposed wherein the unknown parameter is estimated by a stochastic approximation scheme for maximizing the log-likelihood function and the control is obtained via a relative value iteration algorithm. The analysis uses the asymptotic o.d.e.s associated with these.

**1. Introduction.** One popular approach for adaptive control of Markov chains has been the self-tuning scheme of Mandl [18]. In this approach, a parametrized model set is postulated and the parameter is estimated "on line" by a suitable statistical method. The control used is the corresponding "certainty equivalent" control, i.e., the control that would be optimal at a given time for a given state if the current parameter estimate were the true parameter. Mandl proved the asymptotic optimality of this scheme under a strong identifiability condition which requires complete model discrimination under arbitrary control policies. It was brought out in [7] that this condition cannot in general be relaxed. To work around this difficulty, various modifications were proposed, such as randomization of the control or the parameter estimate [8], [10] and introduction of an explicit cost bias in the estimation scheme [4], [5], [15], [16], [19]. There remained, however, another problem with the basic scheme, viz., that a priori, it is not in a computationally amenable form. There are two reasons for this. One is that it requires the computation of optimal control policies (and, in the latter case, costs) as a function of the parameter. Although this computation is "off-line", so to say, the computational and memory overheads can be considerable. Secondly, the statistical schemes employed (mostly maximum likelihood) were

1991 *Mathematics Subject Classification*: Primary 93E35.
*Key words and phrases*: adaptive control, self-tuning control, controlled Markov chains, stochastic approximation, relative value iteration.
Work supported by the Homi Bhabha Fellowship.

in an idealized form where the entire likelihood function (say) is available at each step and an exact maximization is required at each step. This is not always computationally amenable. This has prompted modifications such as a finite grid approximation of the parameter space [22] or recursive computation of control assuming a consistent parameter estimation scheme in the background [14]. The only fully recursive schemes we know are those of El Fattah [11], [12] where both the control policy and the parameter estimate are obtained through stochastic approximation procedures. These works, however, use extremely strong and nontransparent conditions. We propose here an alternative scheme which, while using weaker hypotheses, retains the recursiveness and computational feasibility. Specifically, we use a stochastic approximation algorithm for maximizing the log-likelihood and a relative value iteration to obtain the control policy.

The paper is organized as follows. The next section sets up the notation and describes the adaptive control scheme. Section 3 studies the stochastic approximation scheme for parameter estimation. Almost sure consistency of the estimation scheme is established under suitable conditions. Section 4 considers the asymptotic behaviour of the relative value iteration algorithm and proves the a.s. $\varepsilon$-optimality of the adaptive control scheme. An appendix recalls two important results from [6, 13] used in the main text of the paper.

**2. Preliminaries.** We shall follow the notation of [3], since we shall be referring to it for some key results. Let $X_n, n \geq 0$, be a controlled Markov chain on the state space $S = \{1, \ldots, d\}$ with transition matrix

$$P_u^\theta = [[p(i,j,u_i,\theta)]], \qquad i,j \in S,$$

indexed by the control vector $u = [u_1, \ldots, u_d]$ and the unknown parameter $\theta$. Here $u_i \in D_i$ for some prescribed compact metric space $D_i$, $i \in S$. By replacing each $D_i$ by $\prod_i D_i := D_1 \times \ldots \times D_d$ and $p(i,j,\cdot,\theta)$ by its composition with the projection $\prod_k D_k \to D_i$ for each $i,j,\theta$, we may and do assume that all $D_i$'s are replicas of a fixed compact metric space $D$. The parameter $\theta$ takes values in a compact convex subset $A$ of $\mathbb{R}^m$, $m \geq 1$, containing a distinguished element $\theta_0$, the true parameter. The actual system is assumed to correspond to $\theta_0$, which is unknown. The functions $p(i,j,\cdot,\cdot)$ are assumed to be continuous, and continuously differentiable in the last argument uniformly with respect to the rest. Denote by $P^\theta(\cdot), E_\theta(\cdot)$ the probabilities (resp. expectations) under $\theta$, dropping the $\theta$ when $\theta = \theta_0$. Finally, for any Polish (i.e., separable and metrizable with a complete metric) space $Y$, $\mathcal{P}(Y)$ will denote the Polish space of probability measures on $Y$ with the Prokhorov topology.

A *control strategy* (CS for short) is a sequence $\{\xi_n\}$, $\xi_n = [\xi_n(1), \ldots,$

$\dots \xi_n(d)]$, of $D^d$-valued random variables such that for $i \in S$ and $n \geq 0$,

$$P^\theta(X_{n+1} = i \mid X_m, \xi_m, m \leq n) = p(X_n, i, \xi_n(X_n), \theta).$$

We then say that $\{X_n\}$ is *governed* by the CS $\{\xi_n\}$. If $\xi_n$ is independent of $X_m$, $m \leq n$, and $\xi_m$, $m < n$, for each $n$, and $\{\xi_n\}$ are identically distributed, call the CS a *stationary randomized strategy* (SRS). If the common law of each $\xi_n$ therein is $\Phi \in \mathcal{P}(D^d)$, denote the SRS by $\gamma[\Phi]$. As argued in [3], $\Phi$ may be taken to be a product measure $\prod_i \phi_i$ with $\phi_i \in \mathcal{P}(D)$ for all $i$. Conversely, each such measure can be identified with an SRS. For later reference, let $\mathcal{P}_0(D^d) \subset \mathcal{P}(D^d)$ denote the compact set of product measures. If $\Phi$ is a Dirac measure at $\xi \in D^d$ (say), call the corresponding SRS a *stationary strategy* (SS), denoted by $\gamma\{\xi\}$. Under an SRS (resp. SS), $\{X_n\}$ is a Markov chain with stationary transitions, the transition matrix being given by

$$\begin{aligned} P^\theta[\Phi] &= [[p^\theta_\Phi(i,j)]] \\ &:= \left[\left[\int p(i,j,u,\theta)\, \phi_i(du)\right]\right], \quad i,j \in S \text{ (resp., } P^\theta\{\xi\} = P^\theta_\xi). \end{aligned}$$

We assume throughout that $S$ is a single communicating class under each $\gamma[\Phi]$. The chain then has a unique invariant probability measure denoted by

$$\begin{gathered} \pi^\theta[\Phi] = [\pi^\theta[\Phi](1), \dots, \pi^\theta[\Phi](d)] \\ (\text{resp., } \pi^\theta\{\xi\} = [\pi^\theta\{\xi\}(1), \dots, \pi^\theta\{\xi\}(d)]). \end{gathered}$$

Define $\widehat{\pi}^\theta[\Phi] \in \mathcal{P}(S \times D)$ by

$$\int f\, d\widehat{\pi}^\theta[\Phi] = \sum_{i \in S} \int f(i,u)\, \phi_i(du)\, \pi^\theta[\Phi](i), \qquad f \in C(S \times D).$$

Define $\widehat{\pi}^\theta\{\xi\}$ analogously. Let $k \in C(S \times D)$. The *ergodic* or *long run average cost control problem* is to a.s. minimize over all CS the quantity

$$\limsup_{n \to \infty} \frac{1}{n} \sum_{m=0}^{n-1} k(X_m, \xi_m(X_m)).$$

Under $\gamma[\Phi]$ or $\gamma\{\xi\}$ with $\theta$ as the operative parameter, this a.s. equals $\int k\, d\widehat{\pi}^\theta[\Phi]$ (resp., $\int k\, d\widehat{\pi}^\theta\{\xi\}$). If $\theta_0$ were known, this is the classical ergodic control problem. Since it is not, one has to resort to some adaptive control scheme. We propose one below, following the statement of some additional assumptions.

For each $\theta$ and $\gamma[\Phi]$ with $\Phi = \prod_i \phi_i$, define

$$F(\Phi, \theta) = -\sum \pi^\theta[\Phi](i) \int \phi_i\,(du) \left( \sum_j p(i,j,u,\theta_0) \ln \frac{p(i,j,u,\theta)}{p(i,j,u,\theta_0)} \right).$$

This is continuously differentiable in $\theta$. To see this, recall our differentiability condition on $p(i,j,u,\cdot)$. Now $\pi^\theta[\Phi]$ is the unique solution to the linear

system $\pi^\theta[\Phi]P^\theta[\Phi] = \pi^\theta[\Phi]$, $\sum_i \pi^\theta[\Phi](i) = 1$. Dropping one (say, the first) equation from the former, we get a linearly independent set and Cramer's rule then allows us to write $\pi^\theta[\Phi]$ explicitly as ratio of polynomials in the entries of $P^\theta[\Phi]$ with a nonvanishing determinant (the latter being a consequence of our irreducibility condition on $P^\theta[\Phi]$ for all $\theta$). It follows that $\pi^\theta[\Phi]$ is continuously differentiable in $\theta$ and we are done.

A straightforward application of Jensen's inequality shows that $F(\Phi, \theta) \geq 0$, with $F(\Phi, \theta_0) = 0$. Let $\psi \in \mathcal{P}(D)$ be a prescribed probability measure with support$(\psi) = D$. We say that $\Phi \in \mathcal{P}(D^d)$ is *a-thick* for some $a > 0$ if for all $i \in S$ and Borel $B \subset D$, $\phi_i(B) \geq a\psi(B)$. Our main assumptions are:

(A1) For any $\theta \neq \theta_0$ in $A$, there exist $i, j \in S$ and $u \in D$ such that

$$p(i, j, u, \theta) \neq p(i, j, u, \theta_0).$$

(A2) For any $a > 0$ sufficiently small, there exists a $V : \mathbb{R}^m \to \mathbb{R}^+$ such that $V(\theta) = 0$ if and only if $\theta = \theta_0$ and furthermore,

(i) $\lim_{||x|| \to \infty} V(x) = \infty$,

(ii) for any $\varepsilon > 0$, $\sup\langle \nabla V(\theta), \nabla_\theta F(\Phi, \theta)\rangle < 0$, where the supremum is over all $\theta$ with $\|\theta - \theta_0\| \geq \varepsilon$ and all *a-thick* $\Phi$, and $\nabla_\theta$ is the gradient in the $\theta$ variable,

(iii) for $\theta \in \partial A$ (= the boundary of $A$), $\nabla V(\theta)$ is transversal to $\partial A$ and directed towards interior$(A)$.

Some comments regarding these assumptions are in order here. (A1) is a weaker identifiability condition than Mandl's. The latter requires that the said inequality hold for all $u$. We shall, in fact, argue later that (A1) is no restriction at all. It implies in particular that for some $i, j$, $p(i, j, u, \theta) \neq p(i, j, u, \theta_0)$ for $u$ in an open set. Using the strict convexity of $x \to x \ln x$ and Jensen's inequality, it is then easily verified that $F(\Phi, \theta) > 0$ for $\theta \neq \theta_0$ and $a$-thick $\Phi$, $a > 0$. Thus for given $a > 0, F(\Phi, \cdot)$ with $a$-thick $\Phi$ have a common unique minimum at $\theta_0$. (A2) then ensures a common Lyapunov function for the corresponding gradient flows. An example is the case when $p(i, j, u, \theta)$ are affine in $\theta$. Such parameterizations have been studied in [2], [21]. Then $F(\Phi, \cdot)$ are strictly convex for $a$-thick $\Phi, a > 0$, and $V(\theta) = \|\theta - \theta_0\|^2$ will do the job. It should be kept in mind that we only need the existence of $V$ and not its explicit knowledge in the algorithms proposed. Nevertheless, we require the following.

(A3) There exist $\overline{a} > 0$ and a known continuously differentiable function $W : \mathbb{R}^m \to \mathbb{R}^+$ such that $\nabla W$ is Lipschitz and

$$\langle \nabla W, \nabla V\rangle \geq \overline{a} \quad \text{outside } A.$$

For example, for convex $A$ with a smooth boundary, a suitable $W$ with $\nabla W$ along the outward normal on $\partial A$ will do.

Let $\{a(n)\} \subset (0,1)$ be a decreasing sequence satisfying

$$\sum_n a(n) = \infty, \quad \sum_n a(n)^2 < \infty.$$

Let $K = \max_{i,j,u,\theta} \|\nabla_\theta \ln(p(i,j,u,\theta))\|_\infty$ and $\overline{K} \geq 2K/\overline{a}$, where $\overline{a} > 0$ is as in (A3). Our parameter estimation scheme is

$$\theta(n+1) = \theta(n) + a(n)[G(X_n, X_{n+1}, \xi_n(X_n), \theta(n)) \\ -\overline{K}\nabla W(\theta(n)) I\{\theta(n) \notin A\}],$$

where $G(i,j,u,\theta)$ is any continuous extension of $\nabla_\theta \ln(p(i,j,u,\theta))$ to $S \times S \times D \times \mathbb{R}^m$ satisfying $\|G(\cdot,\cdot,\cdot,\cdot)\|_\infty \leq K$. It should be noted that we are hereby taking the penalty function approach to enforce the constraint $\theta \in A$: The estimation scheme is allowed excursions outside $A$, but is forced back towards $A$ by using the penalty term involving $W$. An alternative approach would be to drop the latter term, but project $\theta(n)$ back into $A$ in a suitable manner at each iteration. Such a scheme is followed, e.g., in [17]. The analysis to follow will have to be correspondingly different, but not in any crucial way.

We also consider a relative value iteration algorithm adapted from [1]. Let $[\theta]$ = the point in $A$ nearest to $\theta$ on the line joining $\theta$ to a prescribed $\theta^* \in A$. For convex $A$, the map $\theta \to [\theta]$ is continuous. For $i \in S$,

$$(2.1) \quad h_{n+1}(i) \\ = h_n(i) + a(n)\Big[\min_u \Big(\sum_j p(i,j,u,[\theta(n)])h_n(j) - h_n(i) + k(i,u)\Big) - h_n(1)\Big].$$

Let $\mathcal{G}_n = \sigma(X_m, \xi_m, m \leq n), \mathcal{F}_n = \sigma(X_m, m \leq n, \xi_m, m < n)$ and $a \in (0,1)$ sufficiently small. For $n \geq 0$, let

$$Z_n = \operatorname{argmin}\Big(\sum_j p(X_n, j, \cdot, [\theta(n)])h_n(j) + k(X_n, \cdot)\Big),$$

any tie being resolved according to some fixed ordering. Let $Z_n'$ be a $D$-valued random variable with law $\psi$, independent of $\mathcal{F}_n$. Pick $\xi_n(X_n)$ according to: $\xi_n(X_n) = Z_n$ with probability $1-a$ and $= Z_n'$ with probability $a$, the randomization being independent of $\mathcal{F}_n, Z_n'$. This completes the description of our adaptive control scheme.

**3. Convergence of parameter estimates.** Define

$$\overline{G}(i,u,\theta) = \sum_j p(i,j,u,\theta)G(i,j,u,\theta),$$

$$\widehat{G}(\mu,\theta) = \int G(\cdot,\cdot,\theta)\, d\mu, \quad \mu \in \mathcal{P}(S \times D),$$

$$L_n(\theta) = \overline{G}(X_n, \xi_n(X_n), \theta),$$
$$\overline{\Delta M}_n = G(X_n, X_{n+1}, \xi_n(X_n), \theta(n)) - L_n(\theta(n)),$$
$$M_n = \sum_{m=0}^{n} a(m)\overline{\Delta M}_m.$$

LEMMA 3.1. *$\{M_n\}$ converges a.s.*

Proof. $(M_n, \mathcal{G}_n)$ is a zero mean martingale with bounded increments $\{\Delta M_n\}$ satisfying $|\Delta M_n| \leq Ka(n)$. Since $\sum a(n)^2 < \infty$, its quadratic variation process converges a.s. By Prop. VII-2-3, pp. 149–150 of [20], $\{M_n\}$ converges a.s. ∎

Let $t_0 = 0$ and $t_n = \sum_{i=0}^{n-1} a(i)$. Define $\overline{\theta}(\cdot) : \mathbb{R}^+ \to \mathbb{R}^m$ by $\overline{\theta}(t_n) = \theta(n)$, $n \geq 0$, with linear interpolation. For $k \geq 0, n \geq k$, define $\widetilde{\theta}^k(\cdot) : [t_k, \infty) \to \mathbb{R}^m$ by $\widetilde{\theta}^k(t_k) = \theta(k)$ and

$$\widetilde{\theta}^k(t_{n+1}) = \widetilde{\theta}^k(t_n) + a(n)(L_n(\widetilde{\theta}^k(t_n)) - \overline{K}\nabla_\theta W(\widetilde{\theta}^k(t_n))I\{\theta(n) \notin A\}) \tag{3.1}$$

with linear interpolation.

LEMMA 3.2. *For each $T > 0$,*

$$\lim_{n\to\infty} \sup_{[t_n, t_n+T]} \|\overline{\theta}(t) - \widetilde{\theta}^n(t)\| = 0.$$

Proof. For $n \geq k$,

$$\overline{\theta}(t_{n+1}) = \overline{\theta}(t_n) + a(n)(L_n(\overline{\theta}(t_n)) - \overline{K}\nabla_\theta W(\overline{\theta}(t_n))I\{\theta(n) \notin A\}) + M_n - M_k, \tag{3.2}$$

where, by the preceding lemma,

$$\lim_{n \geq k \to \infty} (M_n - M_k) = 0 \quad \text{a.s.} \tag{3.3}$$

By subtracting (3.2) from (3.1) and using an appropriate discrete Gronwall inequality, the claim follows by standard arguments in view of (3.3). ∎

Let $U_1$ (resp. $U_2$) denote the space of $\mathcal{P}(S \times D)$-valued (resp., $\mathcal{P}(\{0,1\})$-valued) trajectories $\overline{\mu} = \{\mu_t, t \geq 0\}$ (resp., $\overline{\eta} = \{\eta_t, t \geq 0\}$) with the coarsest topology that renders continuous the maps $\overline{\mu} \to \int_0^T f(t) \int g \, d\mu_t \, dt$, $g \in C(S \times D)$ (resp. $\eta \to \int_0^T f(t)\eta_t(i)\, dt$, $i = 0, 1$), for $T \geq 0$ and $f \in L_2[0, T]$. Then $U_1$ is metrizable by the metric

$$d(\overline{\mu}, \overline{\nu}) = \sum_{k,m,n} 2^{-(k+m+n)} \left| \int_0^n e_k^n(t) \int g_m \, d\mu_t \, dt - \int_0^n e_k^n(t) \int g_m \, d\nu_t \, dt \right| \wedge 1$$

where $\{e_k^n(\cdot)\}_{k=1}^\infty$ is a CONS for $L_2[0, n]$ and $\{g_m\}$ is countable dense in the unit ball of $C(D)$. $U_1$ is also compact. To see this, note that this is equivalent to verifying for each $T > 0$ the compactness of the measures

$dt\,d\mu_t$ on $[0,T]\times D$, i.e., positive measures of total mass $T$ on $[0,T]\times D$ whose marginal on $[0,T]$ is Lebesgue, in the topology of weak convergence. This is immediate from Prokhorov's theorem. Similarly, one shows that $U_2$ is compact metrizable. Consider the o.d.e.

$$\dot\theta(t) = \widehat G(\mu_t,\theta(t)) - \eta_t(1)\overline K\nabla_\theta W(\theta(t)), \qquad \theta(0)=\theta, \tag{3.4}$$

where $\overline\mu\in U_1$ and $\overline\eta\in U_2$.

Lemma 3.3. *The map $U_1\times U_2\times\mathbb{R}^m\ni(\overline\mu,\overline\eta,\theta)\to\theta(\cdot)\in C([0,\infty);\mathbb{R}^m)$ defined by* (3.4) *is continuous.*

Proof. Let $(\overline\mu^n,\overline\eta^n,\theta^n)\to(\overline\mu^\infty,\overline\eta^\infty,\theta^\infty)$ in $U_1\times U_2\times\mathbb{R}^m$. For $n\ge1$, let $\theta^n(\cdot)$ satisfy (3.4) with $\overline\mu=\overline\mu^n,\overline\eta=\overline\eta^n,\theta=\theta^n$. Using the Gronwall lemma and the Arzelà–Ascoli theorem, one verifies that $\{\theta^n(\cdot)\}$ is relatively compact in $C([0,\infty);\mathbb{R}^m)$. By dropping to a subsequence if necessary, let $\theta^n(\cdot)\to\theta^\infty(\cdot)$. Then $\theta^\infty(0)=\theta^\infty$ and for $t\ge0$ and $n\ge1$,

$$\begin{aligned}\theta^n(t) = \theta^n &+ \int_0^t(\widehat G(\mu_s^n,\theta^n(s)) - \eta_s^n(1)\overline K\nabla_\theta W(\theta^n(s))\\ &- \widehat G(\mu_s^n,\theta^\infty(s)) + \eta_s^n(1)\overline K\nabla_\theta W(\theta^\infty(s)))\,ds\\ &+ \int_0^t(\widehat G(\mu_s^n,\theta^\infty(s)) - \eta_s^n(1)\overline K\nabla_\theta W(\theta^\infty(s))\\ &- \widehat G(\mu_s^\infty,\theta^\infty(s)) + \eta_s^\infty(1)\overline K\nabla_\theta W(\theta^\infty(s)))\,ds\\ &+ \int_0^t(\widehat G(\mu_s^\infty,\theta^\infty(s)) - \eta_s^\infty(1)\overline K\nabla_\theta W(\theta^\infty(s)))\,ds.\end{aligned}$$

As $n\to\infty$, the first integral goes to zero because $\theta^n(\cdot)\to\theta^\infty(\cdot)$ and the second does so in view of our topology on $U_1,U_2$. Thus $\theta^\infty(\cdot)$ satisfies (3.4) with $\overline\mu=\overline\mu^\infty,\overline\eta=\overline\eta^\infty$. The claim follows. ∎

Define $\overline\mu'\in U_1$ and $\overline\eta'\in U_2$ by

$$\mu_t'(i,B) = I\{X_n=i,\ \xi_n(X_n)\in B\},\ i\in S,\ B\subset D \text{ Borel},\ t_n\le t<t_{n+1},$$
$$\eta_t'(1) = I\{\overline\theta(t)\notin A\},$$

for $t\ge0$. For $n\ge0$, let $\widehat\theta^n(\cdot)$ denote the solution of (3.4) on $[t_n,\infty)$ when $\overline\mu=\overline\mu',\ \overline\eta=\overline\eta'$ and $\widehat\theta^n(t_n)=\theta(n)$.

Lemma 3.4. *For each $T>0$,*

$$\lim_{n\to\infty}\sup_{t\in[t_n,t_n+T]}\|\widehat\theta^n(t)-\widetilde\theta^n(t)\| = 0.$$

This is straightforward from the Gronwall inequality. In conjunction with the preceding lemmas, this suggests that we can study the time asymp-

totics of our algorithm by looking at limit points of $\widehat{\theta}^n(\cdot)$ in $C([0,\infty);\mathbb{R}^m)$ as $n \to \infty$. Let $(\overline{\mu} = \{\mu_t, t \geq 0\}, \overline{\nu} = \{\nu_t, t \geq 0\}, \widehat{\theta}(\cdot))$ be a limit point of $(\overline{\mu}^n, \overline{\nu}^n, \widehat{\theta}^n(t_n + \cdot))$ in $U_1 \times U_2 \times C((0,\infty);\mathbb{R}^m)$, where $\overline{\mu}^n = \{\mu'_{t_n+t},\ t \geq 0\}$, $\overline{\nu}^n = \{\nu'_{t_n+t}, t \geq 0\},\ n \geq 0$ (i.e., $\mu^n_t = \mu'_{t_n+t},\ \nu^n_t = \nu'_{t_n+t},\ t \geq 0$).

LEMMA 3.5. *Almost surely, the following holds: For any $\overline{\mu}$ as above, and $t \geq 0$, there exists a-thick $\Phi_t \in \mathcal{P}_0(D^d)$ such that $\mu_t = \widehat{\pi}[\Phi_t]$ for the SRS $\gamma[\Phi_t]$.*

Proof. For $i \in S$,

$$\widetilde{M}_n = \sum_{m=1}^{n} a(m)\Big(I\{X_m = i\} - \sum_j I\{X_{m-1} = j\}p(j,i,\xi_{m-1}(j),\theta_0)\Big)$$

is a zero mean bounded increment martingale with respect to $\{\mathcal{G}_n\}$, with a convergent quadratic variation process in view of $\sum a(n)^2 < \infty$. By Prop. VII-2-3(c), pp. 149–150 of [20], it converges a.s. For $n \geq 0$, let

$$\overline{n}(s) = \min\Big\{m > n \;\Big|\; \sum_{j=n}^{m} a(j) \geq s\Big\}, \qquad s > 0.$$

Then

$$\lim_{n\to\infty}(\widetilde{M}_{\bar{n}(s)} - \widetilde{M}_n) = 0 \text{ a.s.} \quad \text{and} \quad \sum_{m=n}^{\bar{n}(s)} a(m) \geq s$$

together imply

$$\frac{\sum_{m=n}^{\bar{n}(s)} a(m) I\{X_m = i\}}{\sum_{m=n}^{\bar{n}(s)} a(m)} - \frac{\sum_{m=n}^{\bar{n}(s)} a(m) \sum_j p(j,i,\xi_{m-1}(j),\theta_0) I\{X_{m-1} = j\}}{\sum_{m=n}^{\bar{n}(s)} a(m)} \to 0 \quad \text{a.s.} \tag{3.5}$$

Define $\varphi_{n,s} \in \mathcal{P}(S \times D)$ by

$$\varphi_{n,s}(B \times C) = \frac{\sum_{m=n}^{\bar{n}(s)} a(m) I\{X_m \in B,\ \xi_m(X_m) \in C\}}{\sum_{m=n}^{\bar{n}(s)} a(m)},$$

for $B \subset S$ and $C \subset D$ Borel. Our conditions on $\{a(m)\}$ imply $a(m+1)/a(m) \to 0$. In view of (3.5) one then has: Almost surely, any limit point $\varphi$ of $\varphi_{n,s}$ in $\mathcal{P}(S \times D)$ as $n \to \infty$ must satisfy

$$\varphi(\{i\} \times D) = \sum_j \int p(j,i,u,\theta_0)\varphi(\{j\} \times du), \qquad i \in S. \tag{3.6}$$

Then $\varphi$ must be of the form $\widehat{\pi}^{\theta_0}[\Phi]$ for some SRS $\gamma[\Phi]$. Recalling our definitions of $\{\mu^n_t\}, \{\overline{n}(s)\}$, etc., it follows that any limit point $\varphi$ in $\mathcal{P}(S \times D)$

of the measures

$$\frac{1}{s}\int_t^{t+s} d\mu'_y\,dy$$

as $t\to\infty$ must be as above. Since $\overline{\mu}$ is a limit point of $\{\overline{\mu}_n\}$, it then follows that for any $t\ge 0$ and $s>0$, there exists a $\Phi=\Phi_{t,s}$ (to make the $t,s$ dependence explicit) in $\mathcal{P}_0(D^d)$ such that

$$\frac{1}{s}\int_t^{t+s}\int f\,d\mu_y\,dy=\int f\,d\widehat{\pi}^{\theta_0}[\Phi_{t,s}],\qquad f\in C(S\times D). \tag{3.7}$$

But (3.6) completely characterizes $\varphi\in\mathcal{P}(S\times D)$ of the form $\widehat{\pi}^{\theta_0}[\Phi]$. Also, (3.6) is preserved under convergence in the compact space $\mathcal{P}(S\times D)$. Therefore one may let $s\to 0$ in (3.7) to conclude that almost surely, for a.e. $t$, there exists $\Phi_t\in\mathcal{P}_0(D^d)$ such that $\mu_t=\widehat{\pi}^{\theta_0}[\Phi_t]$. Since the dependence $\Phi\to\widehat{\pi}^{\theta_0}[\Phi]$ is continuous (see, e.g., [3], Ch. 5), a standard measurable selection argument ensures a measurable version of $t\to\Phi_t$. The qualification "a.e. $t$" may also be dropped by modifying $\overline{\mu}$ suitably on a set of zero Lebesgue measure without affecting anything.

We still need to show that $\{\Phi_t\}$ are $a$-thick. An argument analogous to that employed at the beginning of this proof shows that for any $i\in S$ and Borel $C\subset D$,

$$\frac{\sum_{m=n}^{\bar{n}(s)}a(m)I\{X_m=i,\ \xi_m(i)\in C\}}{\sum_{m=n}^{\bar{n}(s)}a(m)} - \frac{\sum_{m=n}^{\bar{n}(s)}a(m)I\{X_m=i\}\varphi_i^m(C)}{\sum_{m=n}^{\bar{n}(s)}a(m)}\to 0\quad\text{a.s.}$$

as $n\to\infty$, where $\varphi_i^m\in\mathcal{P}(D)$ is the regular conditional law of $\xi_m(i)$ given $\mathcal{F}_m$. By our choice of $\xi_m$, $\varphi_i^m(C)\ge a\psi(C)$. Thus passing to the limit in the above along an appropriate subsequence $\{n_k\}$ (with $\varphi$ as in (3.6)), we get

$$\varphi(\{i\}\times C)\ge\liminf_{k\to\infty}\frac{\sum_{m=n_k}^{n_k(s)}a(m)I\{X_m=i\}\varphi_i^m(C)}{\sum_{m=n_L}^{n_k(s)}a(m)}\ge a\varphi(\{i\}\times D)\psi(C).$$

It follows that the $\Phi_{t,s}$ and hence the $\Phi_t$ above are $a$-thick for a.e. $t$, where the "a.e. $t$" may be dropped as before. ∎

Lemma 3.6. *Almost surely,* $\widehat{\theta}(t)\in\text{interior}(A)\Rightarrow\eta_t(1)=0$ *and* $\widehat{\theta}(t)\notin\overline{A}\Rightarrow\eta_t(1)=1$, $t\ge 0$.

Proof. Let $f\in C(\mathbb{R}^m)$ be nonnegative, smooth with compact support in interior$(A)$. Then recalling that by definition, $\eta_y^n(1)=I\{\overline{\theta}(t_n+y)\notin A\}$,

we have

$$\int_t^{t+s} f(\overline{\theta}(t_n + y))\eta_y^n(1)\, dy = 0 \quad \forall t, s \geq 0.$$

Letting $n \to \infty$ along an appropriate subsequence and using Lemmas 3.2 and 3.4, we have, almost surely,

$$\int_t^{t+s} f(\widehat{\theta}(y))\eta_y(1)\, dy = 0 \quad \forall t, s \geq 0.$$

From our choice of $f$, it follows that $\widehat{\theta}(t) \in \text{interior}(A)$ implies $\eta_t(1) = 0$ for a.e. $t$, where "a.e. $t$" may be dropped by taking a suitable modification. The second claim is proved similarly. ∎

By Lemma 3.3, we have

$$\dot{\widehat{\theta}}(t) = \widehat{G}(\mu_t, \widehat{\theta}(t)) - \eta_t(1)\overline{K}\nabla_\theta W(\widehat{\theta}(t)). \tag{3.8}$$

Using Lemma 3.5, $\widehat{\theta}(t) \in A$ implies

$$\widehat{G}(\mu_t, \widehat{\theta}(t)) = -\nabla_\theta F(\Phi_t, \widehat{\theta}(t)) \tag{3.9}$$

for some $\mathcal{P}_0(D^d)$-valued process $\{\Phi_t\}, t \geq 0$.

THEOREM 3.1. *$\theta(n) \to \theta_0$ a.s.*

Proof. It suffices to prove that $\overline{\theta}(t) \to \theta_0$ a.s. By our choice of $W$ and $\overline{K}$, $\overline{\theta}(\cdot)$ does not exit a prescribed bounded neighbourhood $\widehat{A}$ of $A$. Thus the initial conditions of (3.1) remain in this set. By (A2) and (A3), our choice of $\overline{K}$, (3.8), (3.9) and Lemma 3.6, one has

$$\frac{d}{dt}V(\widehat{\theta}(t)) < 0 \quad \text{when } \widehat{\theta}(t) \neq \theta_0.$$

By the standard Lyapunov stability argument, $\widehat{\theta}(t) \to \theta_0$, uniformly with respect to $\{\Phi_t\}$ and $\widehat{\theta}(0) \in \widehat{A}$. In view of Lemmas 3.2 and 3.4, the claim follows by a standard approximation argument. (See, e.g., Theorem 1, p. 339 of [13], recalled in the appendix as Theorem A.1.) ∎

**4. $\varepsilon$-Optimality.** This section establishes the $\varepsilon$-optimality of the proposed scheme. Before doing so, recall the dynamic programming equations associated with the ergodic control problem [3]:

$$\overline{V}(i) = \min_n \Big(k(i,u) + \sum_j p(i,j,u,\theta_0)\overline{V}(j) - \beta\Big), \quad i \in S. \tag{4.1}$$

These have a solution $(\overline{V}, \beta) \in \mathbb{R}^d \times \mathbb{R}$ where $\beta$ is uniquely specified as the optimal cost

$$\beta = \min_{\gamma\{\xi\}} \int k\, d\widehat{\pi}^{\theta_0}\{\xi\}$$

and $\overline{V}$ is unique up to an additive constant. Let $(V^*, \beta)$ be the unique solution satisfying $V^*(1) = \beta$. Then, for $1_c := [1, \dots, 1]^T$, the solution set is $\{(\overline{V}, \beta) \mid \overline{V} \in J\}$ for

$$J = \{V^* + b1_c \mid b \in \mathbb{R}\}.$$

Define $F^1 : \mathbb{R}^d \to \mathbb{R}^d$ and $F^2 : \mathbb{R}^d \to \mathbb{R}^d$ by

$$F_i^1(x) = \min_u \Big(k(i,u) + \sum_j p(i,j,u,\theta_0)x_j - x_1\Big),$$

$$F_i^2(x) = \min_u \Big(k(i,u) + \sum_j p(i,j,u,\theta_0)x_j - \beta\Big),$$

for $x = [x_1, \dots, x_d]$ and $i \in S$. Then defining the norm $\|\cdot\|_\infty$ and the seminorm $|\cdot|^\sim$ by

$$\|x\|_\infty = \max |x_i|, \qquad |x|^\sim = \max_i x_i - \min_i x_i,$$

we have

$$\begin{aligned} \|F^2(x) - F^2(y)\|_\infty &\le \|x - y\|_\infty, \\ |F^i(x) - F^i(y)|^\sim &\le |x - y|^\sim, \qquad i = 1, 2. \end{aligned} \tag{4.2}$$

Note that $|x|^\sim = 0$ if and only if $x = b1_c$ for some $b \in \mathbb{R}$. Also, $J = \{x \mid F^2(x) = x\}$. Consider the o.d.e.s

$$\dot{x}(t) = F^1(x(t)) - x(t), \tag{4.3}$$

$$\dot{y}(t) = F^2(y(t)) - y(t). \tag{4.4}$$

LEMMA 4.1. *If $x(0) = y(0)$ then $|x(t) - y(t)|^\sim = 0$ for all $t \ge 0$.*

Proof. From (4.3) and (4.4), we have (noting that $F_i^1(x) = F_i^2(x) - (x_1 - \beta)$)

$$x(t) - y(t) = \int_0^t e^{-(t-s)}[(F^2(x(s)) - F^2(y(s))) - (x_1(s) - \beta)1_c]\, ds.$$

Thus

$$\max_i(x_i(t) - y_i(t)) \le \int_0^t e^{-(t-s)}(\max_i(F_i^2(x(s)) - F_i^2(y(s))) - (x_1(s) - \beta))\, ds,$$

$$\min_i(x_i(t) - y_i(t)) \ge \int_0^t e^{-(t-s)}(\min_i(F_i^2(x(s)) - F_i^2(y(s))) - (x_1(s) - \beta))\, ds.$$

Using (4.2), one then has

$$|x(t) - y(t)|^\sim \le \int_0^t e^{-(t-s)}|x(s) - y(s)|^\sim\, ds,$$

from which the claim follows by the Gronwall inequality. ∎

LEMMA 4.2. *For any $x \in J$, $\|y(t) - x\|_\infty$ is nonincreasing and $y(t) \to \overline{y} \in J$, which may depend on $y(0)$.*

This is proved in Theorem 3.1 of [6] (recalled in the appendix as Theorem A.2).

COROLLARY 4.1. *$V^*$ is the globally asymptotically stable equilibrium point of* (4.3).

Proof. By the above lemmas, $|x(t)|^\sim = |y(t)|^\sim \leq 2\|y(t)\|_\infty$ and thus $\{|x(t)|^\sim\}$ is bounded. To show that $x(\cdot)$ is, it then suffices to show that $x_1(t)$ is bounded. Now

$$\begin{aligned}|F_1^2(x(t)) - x_1(t)| &= \left|\min_u \sum_j p(x_1(t), j, u, \theta_0)(x_j(t) - x_1(t)) + k(x(t), u) - \beta\right| \\ &\leq |x(t)|^\sim + C\end{aligned}$$

for a suitable constant $C$. Thus

$$\dot{x}_1(t) = F_1^1(x(t)) - x_1(t) = F_1^2(x(t)) - (x_1(t) - \beta) - x_1(t) = b(t) - (x_1(t) - \beta)$$

for a bounded $b(\cdot)$. Explicitly integrating this linear o.d.e., one sees that $x_1(\cdot)$ is bounded. Hence $x(\cdot)$ is. Since $|x(t) - V^*|^\sim = |y(t) - V^*|^\sim \leq 2\|y(t) - V^*\|_\infty$ and $|y(t) - V^*|^\sim \to 0$ by Lemma 4.2, $x(t) \to \{x \mid |x - V^*|^\sim = 0\} = J$ in a bounded fashion. In particular, since $J = \{x \mid F^2(x) = x\}$, we have $F^2(x(t)) - x(t) \to 0$. Thus

$$\dot{x}_1(t) = b(t) - (x_1(t) - \beta)$$

with $b(t) \to 0$. Integrating explicitly gives

$$x_1(t) - \beta = e^{-t}(x(0) - \beta) + \int_0^t e^{-(t-s)} b(s)\, ds.$$

Since $b(t) \to 0$, l'Hospital's rule can be used to conclude that $x_1(t) \to \beta$. Since $x(t) \to J$ anyway, $x(t) \to V^*$. To conclude asymptotic stability, we also need to show the stability in the sense of Lyapunov. Now, since $V^*(1) = \beta$, we get

$$\begin{aligned}\|x(t) - V^*\|_\infty &\leq |x(t) - V^*|^\sim + |x_1(t) - \beta| = |y(t) - V^*|^\sim + |x_1(t) - \beta| \\ &\leq 2\|y(t) - V^*\|_\infty + |x_1(t) - \beta| \leq 2\|x(0) - V^*\|_\infty + |x_1(t) - \beta|\end{aligned}$$

by the preceding lemma and the fact that $x(0) = y(0)$. Since $V^* \in J$, we have

$$b(t) = F_1^2(x(s)) - x_1(s) - (F_1^2(V^*) - V^*(1)).$$

It is easily verified that

$$(F_1^2(x) - x_1) - (F_1^2(y) - y_1) \leq \max_i((x_i - x_1) - (y_i - y_1)) \leq |x - y|^\sim,$$

$$(F_1^2(x) - x_1) - (F_1^2(y) - y_1) \geq \min_i((x_i - x_1) - (y_i - y_1)) \geq -|x - y|^\sim.$$

Thus,

$$|b(t)| \le |x(t) - V^*|^\sim = |y(t) - V^*|^\sim \le 2\|y(t) - V^*\|_\infty \le 2\|x(0) - V^*\|_\infty.$$

Since

$$x_1(t) - \beta = e^{-t}(x_1(0) - \beta) + \int_0^t e^{-(t-s)} b(s)\, ds,$$

it follows that

$$\begin{aligned} |x_1(t) - \beta| &\le e^{-t}|x_1(0) - \beta| + 2\int_0^t e^{-(t-s)}\|x(0) - V^*\|_\infty\, ds \\ &\le 3\|x(0) - V^*\|_\infty. \end{aligned}$$

Hence $\|x(t) - V^*\|_\infty \le 5\|x(0) - V^*\|_\infty$, implying stability in the sense of Lyapunov. This completes the proof. ∎

Just as we established the convergence of $\{\theta(n)\}$ by linking its iterations with (3.8), we shall establish the convergence of $\{h_n\}$ by linking (2.1) with (4.3). To do so, we first need to establish that $\{h_n\}$ remains bounded.

LEMMA 4.3. *Sample path-wise, if the iterations* (2.1) *remain bounded for one initial condition, they do so for all initial conditions.*

P r o o f. Let $\{h'_n\}, \{h''_n\}$ be two sequences generated by (2.1) with different initial conditions, with $\{h''_n\}$ bounded. Write (2.1) for $\{h'_n\}, \{h''_n\}$, subtract and take the seminorm $|\cdot|^\sim$ on both sides of the resulting equation to obtain

$$\begin{aligned} |h'_{n+1} - h''_{n+1}|^\sim &\le (1 - a(n))|h'_n - h''_n|^\sim + a(n)|h'_n - h''_n|^\sim \\ &= |h'_n - h''_n|^\sim \le \ldots \le |h'_0 - h''_0|^\sim. \end{aligned}$$

But $|h'_n|^\sim \le |h'_n - h''_n|^\sim + |h''_n|^\sim$. Thus $|h'_n|^\sim$ remains bounded. It is then enough to show that any one component of $\{h'_n\}$ remains bounded in order to conclude that $\{h'_n\}$ itself is bounded. Consider $\{h'_n(1)\}$. We have

$$\begin{aligned} &\Big|\min_u \sum_j p(1, j, u, [\theta(n)])h'_n(j) - h'_n(1)\Big| \\ &\qquad = \Big|\min_u \sum_j p(1, j, u, [\theta(n)])(h'_n(j) - h'_n(1))\Big| \le |h'_n|^\sim, \end{aligned}$$

which is bounded. Thus the iteration for $\{h'_n(1)\}$ has the form

$$h'_{n+1}(1) = (1 - a(n))h'_n(1) + a(n)H_n$$

where $\{H_n\}$ is a uniformly bounded sequence. A simple induction argument establishes the boundedness of $\{h'_n(1)\}$ and therefore of $\{h'_n\}$. ∎

LEMMA 4.4. *The sequence* $\{h_n\}$ *generated by* (2.1) *is a.s. bounded.*

Proof. Let $\varepsilon > 0$ and $T > 0$. Define $\{T_n\}$ by $T_0 = 0$ and $T_n = t_{m(n)}$ where $m(n)$ is chosen so that

$$t_{m(n+1)-1} < t_{m(n)} + T \leq t_{m(n+1)}, \quad n \geq 0.$$

Thus $T_{i+1} - T_i \in [T, T+1]$ always. Let $B$ be a large closed ball containing $h_0$ and the $\varepsilon$-neighbourhood of $V^*$ in its interior. Consider $\{h'_n\}$ generated by a modification of (2.1) as follows: Whenever $h'_{m(n)} \in B^c$, reset it to $h_0$. Define $z(t)$, $t \geq 0$, by $z(t_n) = h'_n$ with linear interpolation on $[t_n, t_{n+1}]$, $n \geq 0$. For $n \geq 0$, let $x^n(t)$, $t \in [T_n, T_{n+1}]$, be the solutions of (4.3) satisfying $x^n(T_n) = z(T_n)$. Since $\theta(n) \to \theta_0$ a.s., a routine approximation argument shows that almost surely (i.e., whenever $\theta(n) \to \theta_0$),

$$\lim_{n\to\infty} \sup_{t\in[T_n,T_{n+1}]} \|z(t) - x^n(t)\| = 0.$$

Corollary 4.1 and the converse Lyapunov theorem (Theorem 17.5, p. 100 of [23]) imply that there exists a Lyapunov function for (4.3) that strictly decreases along the nonconstant trajectories of (4.3). Now we can invoke Theorem 1, p. 339 of [13] (Theorem A.1 of the appendix) to conclude that $z(t)$ and therefore $h'_n$ converges a.s. to the $\varepsilon$-neighbourhood of $V^*$. This implies in particular that $h'_{m(n)}$ was reset to $h_0$ at most finitely many times, i.e., $h'_n$ evolved as per (2.1) from some (random) $n$ on. Now appeal to the preceding lemma to conclude. ∎

Theorem 4.1. *$h_n \to V^*$ a.s.*

Proof. In the light of Lemma 4.4, exactly the same argument as in the proof thereof ensures that $h_n$ converges a.s. to the $\varepsilon$-neighbourhood of $V^*$ for a given $\varepsilon$. Since the $\varepsilon > 0$ was arbitrary, we are done. ∎

Theorem 4.2. *For any $\varepsilon > 0$, there exists an $a_0(\varepsilon) > 0$ such that if $a < a_0(\varepsilon)$, the proposed adaptive control policy is $\varepsilon$-optimal.*

Proof. From (4.1), we have

$$V^*(X_n) = \min_u \Big(k(X_n, u) + \sum_j p(X_n, j, u, \theta_0)V^*(j) - \beta\Big), \quad n \geq 0.$$

Thus

$$\begin{aligned}
&\beta + V^*(X_n) - E[V^*(X_{n+1}) \mid \mathcal{G}_n] - k(X_n, \xi_n(X_n)) \\
&\quad = \Big[\min_u \Big(k(X_n, u) + \sum_j p(X_n, j, u, \theta_0)V^*(j)\Big) \\
&\qquad - \min_u \Big(k(X_n, u) + \sum_j p(X_n, j, u, [\theta(n)])V^*(j)\Big)\Big] \\
&\qquad + \Big[\min_u \Big(k(X_n, u) + \sum_j p(X_n, j, u, [\theta(n)])V^*(j)\Big)
\end{aligned}$$

$$
\begin{aligned}
&- \min_n \Big(k(X_n,u) + \sum_j p(X_n,j,u,[\theta(n)])h_n(j)\Big)\Big] \\
&+ \Big[\min_n \Big(k(X_n,u) + \sum_j p(X_n,j,u,[\theta(n)])h_n(j)\Big) \\
&- \Big(k(X_n,\xi_n(x_n)) + \sum_j p(X_n,j,\xi_n(X_n),[\theta(n)])h_n(j)\Big)\Big] \\
&+ \Big[\Big(k(X_n,\xi_n(X_n)) + \sum_j p(X_n,j,\xi_n(X_n),[\theta(n)])h_n(j)\Big) \\
&- \Big(k(X_n,\xi_n(X_n)) + \sum_j p(X_n,j,\xi_n(X_n),\theta_0)V^*(j)\Big)\Big].
\end{aligned}
$$

Let $\delta > 0$. Since $[\theta(n)] \to \theta^*$ and $h_n \to V^*$ a.s., outside a zero probability set (ignored henceforth), the expressions in the first, second and fourth square brackets do not exceed $\delta/3$ for sufficiently large $n$. That in the third square bracket is bounded by $K'I\{\xi_n(X_n) = Z'_n\}$ for a suitable constant $K'$. Sum both sides over $n = 0, 1, \ldots, N-1$, divide by $N$ and let $N \to \infty$. By the strong law of large numbers for square integrable martingales ([9], p. 244), we have

$$
\frac{1}{N}\sum_{n=1}^{N}(V^*(X_n) - E[V^*(X_n) \mid \mathcal{G}_{n-1}]) \to 0 \quad \text{a.s.}
$$

Hence

$$
\begin{aligned}
&\limsup_{n\to\infty} \Big|\beta - \frac{1}{n}\sum_{m=0}^{n-1} k(X_m,\xi_m(X_m))\Big| \\
&\qquad \le \delta + K' \limsup_{n\to\infty} \frac{1}{n}\sum_{m=0}^{n-1} I\{\xi_m(X_m) = Z'_m\} \le \delta + K'a \quad \text{a.s.}
\end{aligned}
$$

Since $\delta$ was arbitrary, the claim follows for $a_0(\varepsilon) = \varepsilon/K'$. ∎

In conclusion, observe that if (A1) were relaxed, one could analogously obtain convergence of $\{\theta(n)\}$ to the set of $\theta$ for which $p(i,j,u,\theta) = p(i,j,u,\theta_0)$ for all $i,j,u$. The $\varepsilon$-optimality argument does not get affected.

**Appendix.** We recall here two key results from [6], [13] resp. used in this paper. We start with Theorem 1, p. 339 of [13], which is Theorem A.1 below.

Consider the $d$-dimensional o.d.e.

$$
\dot{x}(t) = f(x(t),t) \tag{A.1}
$$

which has a globally, uniformly asymptotically stable equilibrium point $x_0$ and an associated continuously differentiable Lyapunov function $V : \mathbb{R}^d \to$

$\mathbb{R}^+$ satisfying $\sup_{t,\|x-x_0\|\geq\varepsilon} \nabla V \cdot f(x,t) < 0$ for any $\varepsilon > 0$. Given $T > 0$ and $\delta > 0$, we call a bounded measurable function $y(\cdot) : \mathbb{R}^+ \to \mathbb{R}^d$ a $(T,\delta)$-*perturbation* of (A.1) if there exist $0 = T_0 < T_1 < T_2 < \ldots$ such that $T_{j+1} - T_j \geq T$ for all $j$ and there exist solutions $x^j(t)$, $t \in [T_j, T_{j+1}]$, of (A.1) for $j \geq 0$ such that

$$\sup_{t\in[T_j,T_{j+1}]} \|x^j(t) - y(t)\| < \delta \quad \forall j.$$

THEOREM A.1. *Given $T > 0$ and $\varepsilon > 0$, there exists a $\delta_0 > 0$ sufficiently small such that for $0 < \delta < \delta_0$, any $(T,\delta)$-perturbation $y(\cdot)$ of* (A.1) *converges to the $\varepsilon$-neighbourhood of $x_0$.*

Proof. Clearly $V(x_0) < V(x)$, $x \neq x_0$. For $\eta > 0$, define $B(\eta) = \{x \mid V(x) < V(x_0) + \eta\}$. Then $B(\eta)$ is an open neighbourhood of $x_0$. Since $y(\cdot)$ is bounded, we may suppose that $y(\cdot)$ and the trajectories $\{x^j(\cdot)\}$ of (A.1) as above which we shall consider below, a priori lie in a sufficiently large closed bounded ball $\overline{B}$. Let

$$K = \max_{x\in\overline{B}} \|\nabla V(x)\|, \quad \Delta = -\sup_{t\geq 0,\, x\notin B(\eta)} \nabla V(x) \cdot f(x,t).$$

Then $\Delta > 0$ and for $\{x^j(\cdot)\}$ as above,

$$V(x^j(T_{j+1})) \leq V(x^j(T_j)) - \Delta T$$

whenever $x^j(t)$, $t \in [T_j, T_{j+1}]$, does not intersect $B(\eta)$. If $\delta < \Delta T/(4K)$, we also have

$$V(y(T_{j+1})) \leq V(y(T_j)) - \Delta T/2. \tag{A.2}$$

Call $y(t), t \in [T_i, T_{i+1}]$, a *patch* of $y(\cdot)$. If a patch of $y(\cdot)$ does not intersect $B(\eta + \delta/K)$, the corresponding $x^j(\cdot)$ cannot intersect $B(\eta)$ and (A.2) holds. Since (A.2) can hold for at most finitely many consecutive $j$, eventually $x^j(\cdot)$ must intersect $B(\eta)$ whence the corresponding patch of $y(\cdot)$ intersects $B(\eta + \delta/K)$. Now

$$V(x^j(t)) \leq V(x^j(s)) \quad \text{for } s,t \in [T_j, T_{j+1}],\ t \geq s,$$

always and thus

$$V(y(t)) \leq V(y(s)) + 2\delta K \quad \text{for } s,t \in [T_j, T_{j+1}],\ t \geq s,$$

for all $j$. Hence the patch of $y(\cdot)$ that intersects $B(\eta + \delta/K)$ remains in $B(\eta + \delta/K + 2\delta K)$ after hitting $B(\eta + \delta/K)$. Since $2\delta K < \Delta T/2$, (A.2) ensures that the subsequent patch also hits $B(\eta + \delta/K)$. It follows that $y(\cdot)$ remains in $B(\eta + \delta/K + 2\delta K)$ once it hits $B(\eta + \delta/K)$. Pick $\eta, \delta$ sufficiently small so that $B(\eta + \delta/K + 2\delta K)$ is in the $\varepsilon$-neighbourhood of $x_0$. This completes the proof. ∎

It should be remarked that this is a slight variant of the original result of [13], where the o.d.e. is autonomous. In applying this result in Theorem 3.1, one notes that for a given $T > 0$, $\overline{\theta}(t + \cdot)$ is a $(T, \delta)$-perturbation of (3.8) for any $\delta > 0$ for sufficiently large $t$, by virtue of Lemmas 3.2 and 3.4. Thus the above applies for every $\varepsilon > 0$, implying the desired convergence.

We now turn to Theorem 3.1 of [6], which is Theorem A.2 below. The proof is very lengthy, so we shall proceed through a sequence of lemmas. Consider the $d$-dimensional o.d.e.

$$\dot{x}(t) = F(x(t)) - x(t) \tag{A.3}$$

where $F$ satisfies $\|F(x) - F(y)\|_\infty \le \|x - y\|_\infty$ and $J = \{x \mid F(x) = x\} \neq \emptyset$. Let $x^* \in J$.

LEMMA A.1. *$t \to \|x(t) - x^*\|_\infty$ is nonincreasing.*

Proof. For $x \in \mathbb{R}^d$, define $\|x\|_p = (d^{-1} \sum_{i=1}^d |x_i|^p)^{1/p}$ for $p \in (1, \infty)$. It is easily verified that $\|x\|_p \to \|x\|_\infty$ as $p \to \infty$. Direct differentiation leads to

$$\frac{d}{dt}\|x(t) - x^*\|_p = -\|x(t) - x^*\|_p + \|x(t) - x^*\|_p^{1-p} \Gamma(t)$$

where

$$\begin{aligned}\Gamma(t) &= \frac{1}{d} \sum_{i=1}^d |x_i(t) - x_i^*|^{p-1} \operatorname{sgn}(x_i(t) - x_i^*)(F_i(x(t)) - F_i(x^*)) \\ &\le \|x(t) - x^*\|_p^{p-1} \|F(x(t)) - F(x^*)\|_p \qquad \text{(by Hölder's inequality)}\end{aligned}$$

Integrating over $[s, t], t \ge s$, gives

$$\|x(t) - x^*\|_p \le \|x(s) - x^*\|_p + \int_s^t (-\|x(y) - y^*\|_p + \|F(x(y)) - F(x^*)\|_p)\, dy.$$

Let $p \to \infty$ and use $\|F(x(y)) - F(x^*)\|_\infty \le \|x(y) - x^*\|_\infty$ to conclude. ∎

Thus $\|x(t) - x^*\|_\infty \to b \ge 0$. If $b = 0$, we are done. Suppose $b > 0$. At this juncture, we need some additional terminology.

For $m \le d$, an *$m$-face* is a set of the type

$$\{x = [x_1, \ldots, x_d] \mid x_{i_k} \in [a_k, b_k],\ k \le m,\ x_{i_k} = c_k,\ k > m\}$$

where $\{i_1, \ldots, i_d\}$ is a permutation of $\{1, \ldots, d\}$ and $c_k, b_k > a_k$ are scalars. Let $B_b = \{x \in \mathbb{R}^d \mid \|x - x^*\|_\infty = b\}$, which then is the union of $(d-1)$-faces of the type

$$\{x \mid x_i - x_i^* = b \text{ or } -b, |x_j - x_j^*| \le b \text{ for } j \neq i\}.$$

Then $x(t) \to B_b$, i.e., $\Omega =$ the $\omega$-limit set of $x(\cdot)$, is contained in $B_b$. If $\Omega = \{\overline{x}\}$ then $\overline{x}$ is an equilibrium point for (A.3). Thus $F(\overline{x}) = \overline{x}$ and we

are done. If not, let $\widetilde{x}(\cdot)$ be a trajectory of (A.3) in $\Omega$. By abuse of notation, let $\{\widetilde{x}(\cdot)\} = \{\widetilde{x}(t) \mid t \in \mathbb{R}\}$.

Finally, for a $(d-1)$-face $A$, define $G_A = \{x \in A \mid F(x) \in A\}$. Then $G_A$ is closed, possibly empty.

LEMMA A.2. $\{\widetilde{x}(\cdot)\} \cap A \subset G_A$.

Proof. If both sets are empty, there is nothing to prove. Suppose $\{\widetilde{x}(\cdot)\} \cap A \neq \emptyset$. For simplicity, let $A = \{x \mid x_1^* = a, |x_i - x_i^*| \leq a, i \neq 1\}$. By suitable choice of $\widetilde{x}(0)$, suppose that $\{\widetilde{x}(t) \mid t \in [0,\overline{t}]\} \subset A$ for some $\overline{t} > 0$. Then for $t \in [0,\overline{t}]$, $\widetilde{x}_1(t) = a + x_1^*$. Hence

$$0 = \frac{d}{dt}\widetilde{x}_1(t) = F_1(\widetilde{x}(t)) - \widetilde{x}_1(t), \quad t \in [0,\overline{t}].$$

Also, $|F_i(\widetilde{x}(t)) - x_i^*| \leq \|\widetilde{x}(t) - x^*\|_\infty = b$ for $i \geq 2$, $t \in [0,\overline{t}]$. It follows that $\widetilde{x}(t) \in G_A$ for $t \in [0,\overline{t}]$. Thus all connected segments of $\{\widetilde{x}(\cdot)\} \cap A$ containing more than one point are in $G_A$. Clearly, those containing a single point must be in the relative boundary $\partial A$ of $A$, which is a union of its faces which are $(d-2)$-faces. Let $x \in \{\widetilde{x}(\cdot)\} \cap \partial A$. It suffices to show that $F(x) \in \partial A$. If not, $F(x) - x$ would be transversal to $\partial A$ at $x$, which contradicts the fact that $\{\widetilde{x}(\cdot)\}$ is a differentiable trajectory confined to $B_b$. (It cannot make "sharp turns".) This completes the proof. ∎

Fix a $(d-1)$-face $A$ for the time being.

LEMMA A.3. *If $G_A \neq \emptyset$, then $F : G_A \to A$ can be extended to a map $\widetilde{F} : A \to A$ satisfying $\|\widetilde{F}(x) - \widetilde{F}(y)\|_\infty \leq \|x - y\|_\infty$ for $x, y \in A$. Further, $\widetilde{F}$ has a fixed point $\widetilde{x}$ in $A$.*

Proof. The second claim follows from the first by the Brouwer fixed point theorem. To prove the first, suppose for simplicity that $A = \{x \mid x_1 = x_1^* + b,\ |x_j - x_j^*| \leq b$ for $j > 1\}$. Fix $i, 1 < i \leq d$. Define

$$g_i(x) = \inf_{y \in G_A} (F_i(y) + \|x - y\|_\infty), \quad x \in A.$$

Then $g_i(x) \leq F_i(x)$ for $x \in G_A$. For $x, y \in G_A$,

$$|F_i(x) - F_i(y)| \leq \|x - y\|_\infty$$

leads to

$$F_i(y) + \|x - y\|_\infty \geq F_i(x).$$

Thus $g_i(x) \geq F_i(x)$, implying $F_i = g_i$ on $G_A$. For $x, z \in A$,

$$g_i(x) \leq \inf_{y \in G_A} (F_i(y) + \|y - z\|_\infty + \|z - x\|_\infty) \leq g_i(z) + \|z - x\|_\infty.$$

Similarly, $g_i(z) \leq g_i(x) + \|z - x\|_\infty$. Hence

$$|g_i(x) - g_i(z)| \leq \|z - x\|_\infty.$$

Let $\widetilde{F}_i(x) = (g_i(x) \wedge (x_i^* + b)) \vee (x_i^* - a)$. Then

$$|\widetilde{F}_i(x) - \widetilde{F}_i(y)| \le \|x - y\|_\infty.$$

Let $\widetilde{F}_1(x) = x_1^* + b$ for $x \in A$. Then $\widetilde{F}(\cdot) = [\widetilde{F}_1(\cdot), \dots, \widetilde{F}_d(\cdot)]$ has the desired properties. ∎

The same argument can be used once again to extend $\widetilde{F}$ to a map $\widehat{F} : \mathbb{R}^d \to \mathbb{R}^d$ that restricts to $\widetilde{F}$ on $A$ and to $F$ on $\bigcup_{A'} G_{A'}$ (the union is over all $(d-1)$-faces of $B_b$) and satisfies $\|\widehat{F}(x) - \widehat{F}(y)\|_\infty \le \|x - y\|_\infty$ for $x, y \in \mathbb{R}^d$. Now repeat the earlier argument with $\widehat{F}, \widetilde{x}$ replacing $F, x^*$ to conclude that $\|\widetilde{x}(t) - \widetilde{x}\|_\infty$ is nonincreasing and thus converges to a $c \ge 0$. If $c = 0$, we are done. If not, $\widetilde{x}(t) \to B_c$. Also, it is clear that no $(d-1)$-face of $B_c$ is coplanar with $A$. This argument can be repeated for each $(d-1)$-face of $B_b$ that intersects $\{\widetilde{x}(\cdot)\}$, leading to possibly more $\|\cdot\|_\infty$-spheres $B_q, B_r, \dots$ defined analogously to $B_c$ such that $\widetilde{x}(t) \to B_c \cap B_q \cap B_r \cap \dots$ The above remarks also imply that this intersection is a union of $m$-faces with $m$ at most $d-2$. Now consider a trajectory $\overline{x}(\cdot)$ of (A.3) in the $\omega$-limit set of $\widetilde{x}(\cdot)$ and repeat the above argument to conclude that $\overline{x}(t)$ converges to a union of $m$-faces with $m$ at most $d-3$. Iterating this argument at most $d$ times, we are left with a union of finitely many points to one of which $\widetilde{x}(\cdot), \overline{x}(\cdot), \dots$ and therefore $x(\cdot)$ must converge and which then must be a fixed point of $F$. Thus we have:

THEOREM A.2. *Any solution $x(\cdot)$ of* (A.3) *converges to a point in $J$ that may depend on $x(0)$. Also, for any $x^* \in J$, $\|x(t) - x^*\|_\infty$ is nonincreasing.*

## References


[1] D. Bertsekas, *Dynamic Programming—Deterministic and Stochastic Models*, Prentice-Hall, Englewood Cliffs, N.J., 1987.

[2] V. S. Borkar, *Identification and adaptive control of Markov chains*, Ph.D. Thesis, Dept. of Electrical Engrg. and Computer Science, Univ. of California, Berkeley, 1980.

[3] —, *Topics in Controlled Markov Chains*, Pitman Res. Notes in Math. 240, Longman Scientific and Technical, Harlow, 1991.

[4] —, *The Kumar–Becker–Lin scheme revisited*, J. Optim. Theory Appl. 66 (1990), 289–309.

[5] —, *On Milito–Cruz adaptive control scheme for Markov chains*, ibid. 77 (1993), 385–393.

[6] V. S. Borkar and K. Soumyanath, *A new analog parallel scheme for fixed point computation I—theory*, submitted.

[7] V. S. Borkar and P. P. Varaiya, *Adaptive control of Markov chains I: finite parameter case*, IEEE Trans. Automat. Control AC-24 (1979), 953–957.

[8] —, —, *Identification and adaptive control of Markov chains*, SIAM J. Control Optim. 20 (1982), 470–488.

[9] Y.-S. Chow and H. Teicher, *Probability Theory*: *Independence*, *Interchangeability*, *Martingales*, Springer, New York, 1979.
[10] B. Doshi and S. Shreve, *Randomized self-tuning control of Markov chains*, J. Appl. Probab. 17 (1980), 726–734.
[11] Y. El Fattah, *Recursive algorithms for adaptive control of finite Markov chains*, IEEE Trans. Systems Man Cybernet. SMC-11 (1981), 135–144.
[12] —, *Gradient approach for recursive estimation and control in finite Markov chains*, Adv. Appl. Probab. 13 (1981), 778–803.
[13] M. Hirsch, *Convergent activation dynamics in continuous time networks*, Neural Networks 2 (1987), 331–349.
[14] A. Jalali and M. Ferguson, *Adaptive control of Markov chains with local updates*, Systems Control Lett. 14 (1990), 209–218.
[15] P. R. Kumar and A. Becker, *A new family of adaptive optimal controllers for Markov chains*, IEEE Trans. Automat. Control AC-27 (1982), 137–142.
[16] P. R. Kumar and W. Lin, *Optimal adaptive controllers for Markov chains*, ibid., 756–774.
[17] H. Kushner and D. Clark, *Stochastic Approximation for Constrained and Unconstrained Systems*, Springer, Berlin, 1978.
[18] P. Mandl, *Estimation and control in Markov chains*, Adv. Appl. Probab. 6 (1974), 40–60.
[19] R. Milito and J. B. Cruz Jr., *An optimization oriented approach to adaptive control of Markov chains*, IEEE Trans. Automat. Control AC-32 (1987), 754–762.
[20] J. Neveu, *Discrete-Parameter Martingales*, North-Holland, Amsterdam, 1975.
[21] B. Sagalovsky, *Adaptive control and parameter estimation in Markov chains*: *a linear case*, IEEE Trans. Automat. Control AC-27 (1982), 414–417.
[22] Ł. Stettner, *On nearly self-optimizing strategies for a discrete-time uniformly ergodic adaptive model*, Appl. Math. Optim. 27 (1993), 161–177.
[23] T. Yoshizawa, *Stability Theory by Liapunov's Second Method*, The Mathematical Society of Japan, 1966.

Vivek S. Borkar
Department of Computer Science and Automation
Indian Institute of Science
Bangalore-560012, India
E-mail: borkar@csa.iisc.ernet.in

*Received on 4.10.1995;*
*revised version on 2.4.1996*