Chapter 3
Sample-Path Optimality in Average Markov Decision Chains Under a Double Lyapunov Function Condition

Rolando Cavazos-Cadena and Raúl Montes-de-Oca

3.1 Introduction

This note concerns discrete-time Markov decision processes (MDPs) evolving on a denumerable state space. The performance index of a control policy is an (long run) average criterion, and besides standard continuity compactness conditions, the main structural assumption on the model is that (a) the (possibly unbounded) cost function has a Lyapunov function $\ell(\cdot)$ and (b) a power of order larger than 2 of ℓ also admits a Lyapunov function [14]. Within this context, the main purpose of the paper is to analyze the *sample-path* average optimality of some policies whose *expected* optimality is well known. More specifically, the main results in this direction are as follows:

(i) The stationary policy f obtained by optimizing the right-hand side of the optimality equation is sample-path optimal in the strong sense, that is, under the action of f, the observed average costs in finite times converge almost surely to the optimal expected average cost g, whereas if the system is driven by any other policy, then with probability 1, the inferior limit of those averages is at least g.

(ii) The Markovian policies obtained from procedures frequently used to approximate a solution of the optimality equation, like the vanishing discount or the successive approximations methods, are sample-path average optimal in the strong sense.

R. Cavazos-Cadena (✉)
Departamento de Estadística y Cálculo, Universidad Autónoma Agraria Antonio Narro, Buenavista, Saltillo, Coahuila 25315, México
e-mail: rcavazos@uaaan.mx

R. Montes-de-Oca
Departamento de Matemáticas, Universidad Autónoma Metropolitana–Iztapalapa, México D.F. 09340, México
e-mail: momr@xanum.uam.mx

D. Hernández-Hernández and A. Minjárez-Sosa (eds.), *Optimization, Control, and Applications of Stochastic Systems*, Systems & Control: Foundations & Applications,
DOI 10.1007/978-0-8176-8337-5_3,

The expected average criteria have been intensively studied, and a fairly complete account of the theory can be found in [9, 12, 13]; see also [1]. In this last paper, it was shown that, for a general MDP, if the optimality equation has a bounded solution, then the stationary policy f referred to in the point (i) above is optimal in the sample-path sense. In [3, 4], a similar conclusion was obtained for models with denumerable state space if the cost function has an almost monotone (or penalized) structure, in the sense that the costs are sufficiently large outside a compact set; such a conclusion was extended to models on Borel spaces by [10, 16, 20]. More recently, for models with denumerable state space and finite actions sets, the sample-path average criterion was studied in [15] under the uniform ergodicity assumption. On the other hand, the first result described above is an extension of Theorem 4.1 in [6], where the sample-path optimality of the policy f mentioned above was established in a weaker sense than the one used in the present work.

The approach of this note relies on basic probabilistic ideas, like Kolmogorov's inequality and the first Borel-Cantelli lemma, and was motivated by the elementary analysis of the strong law of large numbers as presented in [2].

The organization of the subsequent material is as follows: In Sect. 3.2, the decision model is presented and the conditions to obtain an (expected) average optimal stationary policy from a solution of the optimality equation are briefly described. Next, in Sect. 3.3, the idea of Lyapunov function is introduced and some of its elementary properties are established, whereas in Sect. 3.4, the basic structural restriction on the model, namely, the double Lyapunov function condition, is formulated as Assumption 3.4.1, and the main result of the chapter, solving problem (i) above, is stated as Theorem 3.4.1. The argument to establish this result relies on some properties of the sequence of innovations associated with the sequence of optimal relative costs, which are presented in Sect. 3.5, and then, the main theorem is proved in Sect. 3.6. Next, the result on the sample-path optimality of Markovian policies is stated as Theorem 3.7.1 in Sect. 3.7, and the necessary technical tools to prove that result, concerning tightness of the sequence of empirical measures and uniform integrability of the cost function, are established in Sect. 3.8. Finally, the exposition concludes with the proof of Theorem 3.7.1 in Sect. 3.9.

Notation. Throughout the remainder, $\mathbb{N}$ stands for the set of all nonnegative integers and the indicator function of a set A is denoted by I_A, so that $I_A(x) = 1$ if $x \in A$ and $I_A(x) = 0$ when $x \notin A$. On the other hand, for a topological space $\mathbb{K}$, the class of all continuous functions defined on $\mathbb{K}$ and the Borel σ-field of $\mathbb{K}$ are denoted by $\mathscr{C}(\mathbb{K})$ and $\mathscr{B}(\mathbb{K})$, respectively, whereas $\mathbb{P}(\mathbb{K})$ stands for the class of all probability measures defined in $\mathscr{B}(\mathbb{K})$. Finally, for an event G, the corresponding indicator function is denoted by $I[G]$.

3.2 Decision Model

Let $\mathscr{M} = (S, A, \{A(x)\}_{x\in S}, C, P)$ be an MDP, where the state space S is a denumerable set endowed with the discrete topology and the action set A is a metric space. For each $x \in S$, $A(x) \subset A$ is the nonempty subset of admissible actions at x and, defining the class of admissible pairs by $\mathbb{K} := \{(x,a) \,|\, a \in A(x), x \in S\}$, the mapping $C\colon \mathbb{K} \to \mathbb{R}$ is the cost function, whereas $P = [p_{xy}(\cdot)]$ is the controlled transition law on S given $\mathbb{K}$, that is, for all $(x,a) \in \mathbb{K}$ and $y \in S$, the relations $p_{xy}(a) \geq 0$ and $\sum_{y\in S} p_{xy}(a) = 1$ are satisfied. This model $\mathscr{M}$ is interpreted as follows: At each time $t \in \mathbb{N}$, the decision maker observes the state of a dynamical system, say $X_t = x \in S$, and selects an action (control) $A_t = a \in A(x)$ incurring a cost $C(x,a)$. Then, regardless of the previous states and actions, the state at time $t+1$ will be $X_{t+1} = y \in S$ with probability $p_{xy}(a)$; this is the Markov property of the decision process.

Assumption 3.2.1

(i) For each $x \in S$, $A(x)$ is a compact subset of A.
(ii) For every $x, y \in S$, the mappings $a \mapsto C(x,a)$ and $a \mapsto p_{xy}(a)$ are continuous in $a \in A(x)$.

Policies. The space $\mathbb{H}_t$ of possible histories up to time $t \in \mathbb{N}$ is defined by $\mathbb{H}_0 := S$ and $\mathbb{H}_t := \mathbb{K}^t \times S$ for $t \geq 1$, whereas a generic element of $\mathbb{H}_t$ is denoted by $\mathbf{h}_t = (x_0, a_0, \ldots, x_t, a_t, \ldots, x_t)$, where $a_i \in A(x_i)$. A policy $\pi = \{\pi_t\}$ is a special sequence of stochastic kernels: For each $t \in \mathbb{N}$ and $\mathbf{h}_t \in \mathbb{H}_t$, $\pi_t(\cdot|\mathbf{h}_t)$ is a probability measure on $\mathscr{B}(A)$ concentrated on $A(x_t)$, and for each Borel subset $B \subset A$, the mapping $\mathbf{h}_t \mapsto \pi_t(B|\mathbf{h}_t)$, $\mathbf{h}_t \in \mathbb{H}_t$, is Borel measurable. The class of all policies is denoted by $\mathscr{P}$ and when the controller chooses actions according to π, the control A_t applied at time t belongs to $B \subset A$ with probability $\pi_t(B|\mathbf{h}_t)$, where $\mathbf{h}_t$ is the observed history of the process up to time t. Given the policy π being used for choosing actions and the initial state $X_0 = x$, the distribution of the state-action process $\{(X_t, A_t)\}$ is uniquely determined [9], and such a distribution and the corresponding expectation operator are denoted by P_x^π and E_x^π, respectively. Next, define $\mathscr{F} := \prod_{x\in S} A(x)$ and notice that $\mathscr{F}$ is a compact metric space, which consists of all functions $f\colon S \to A$ such that $f(x) \in A(x)$ for each $x \in S$. A policy π is *Markovian* if there exists a sequence $\{f_t\} \subset \mathscr{F}$ such that the probability measure $\pi_t(\cdot|\mathbf{h}_t)$ is always concentrated at $f_t(x_t)$, and if $f_t \equiv f$ for every t, the Markovian policy π is is referred to as *stationary*. The classes of stationary and Markovian policies are naturally identified with $\mathscr{F}$ and $\mathscr{M} := \prod_{t=0}^{\infty} \mathscr{F}$, respectively, and with these conventions $\mathscr{F} \subset \mathscr{M} \subset \mathscr{P}$.

Performance Criteria. Suppose that the cost function $C(\cdot,\cdot)$ is such that

$$E_x^\pi [|C(X_t, A_t)|] < \infty, \quad x \in S, \quad \pi \in \mathscr{P}, \quad t \in \mathbb{N}. \tag{3.1}$$

In this case, the (long-run superior limit) average cost corresponding to $\pi \in \mathscr{P}$ at state $x \in S$ is defined by

$$J(x,\pi) := \limsup_{k\to\infty} \frac{1}{k} E_x^\pi \left[\sum_{t=0}^{k-1} C(X_t, A_t) \right], \tag{3.2}$$

and the corresponding optimal value function is specified by

$$J^*(x) := \inf_{\pi\in\mathscr{P}} J(x,\pi), \quad x \in S; \tag{3.3}$$

a policy $\pi^* \in \mathscr{P}$ is (superior limit) average optimal if $J(x,\pi^*) = J^*(x)$ for every $x \in S$. The criterion (3.2) evaluates the performance of a policy in terms of the largest among the limit points of the expected average costs in finite times. In contrast, the following index assesses a policy in terms of the smallest of such limit points:

$$J_-(x,\pi) := \liminf_{k\to\infty} \frac{1}{k} E_x^\pi \left[\sum_{t=0}^{k-1} C(X_t, A_t) \right] \tag{3.4}$$

is the (long run) inferior limit average criterion associated with $\pi \in \mathscr{P}$ at a state x, whereas the optimal value function associated with this criterion is given by

$$J_-^*(x) := \inf_{\pi\in\mathscr{P}} J_-(x,\pi), \quad x \in S. \tag{3.5}$$

From these specifications, it follows that

$$J_-^*(\cdot) \le J^*(\cdot), \tag{3.6}$$

and within the context described below, it will be shown that the equality holds in this last relation.

Optimality Equation. A basic instrument to analyze the above average criteria is the following *optimality equation*:

$$g + h(x) = \inf_{a\in A(x)} \left[C(x,a) + \sum_{y\in S} p_{xy}(a) h(y) \right], \quad x \in S, \tag{3.7}$$

where $g \in \mathbb{R}$ and $h \in \mathscr{C}(S)$ are given functions, and it is supposed that

$$E_x^\pi[|h(X_n)|] < \infty, \quad x \in S, \quad \pi \in \mathscr{P}, \quad n \in \mathbb{N}.$$

Under this condition and (3.1), a standard induction argument combining (3.7) and the Markov property yields that, for every nonnegative integer n,

$$(n+1)g+h(x) \le E_x^{\pi}\left[\sum_{t=0}^{n} C(X_t,A_t)+h(X_{n+1})\right], \quad x \in S, \quad \pi \in \mathscr{P}.$$

Moreover, if $f \in \mathscr{F}$ satisfies that

$$g+h(x) = C(x,f(x))+\sum_{y\in S} p_{xy}(f(x))h(y), \quad x \in S, \tag{3.8}$$

then

$$(n+1)g+h(x) = E_x^{f}\left[\sum_{t=0}^{n} C(X_t,A_t)+h(X_{n+1})\right], \quad x \in S.$$

Therefore, assuming that the condition

$$\lim_{n\to\infty} \frac{E_x^{\pi}[h(X_{n+1})]}{n+1} = 0 \tag{3.9}$$

holds for every $x \in S$ and $\pi \in \mathscr{P}$, it follows that the relation

$$\lim_{n\to\infty} \frac{1}{n+1} E_x^{f}\left[\sum_{t=0}^{n} C(X_t,A_t)\right] = g \le \liminf_{n\to\infty} \frac{1}{n+1} E_x^{\pi}\left[\sum_{t=0}^{n} C(X_t,A_t)\right] \tag{3.10}$$

is always valid, and then, (3.2)–(3.6) immediately yield that

$$J_*^*(x) = J^*(x) = g = \lim_{n\to\infty} \frac{1}{n+1} E_x^{f}\left[\sum_{t=0}^{n} C(X_t,A_t)\right], \quad x \in S, \tag{3.11}$$

so that:

(i) The superior and inferior limit average criteria render the same optimal value function,
(ii) A stationary policy f satisfying (3.8) is average optimal, and
(iii) The optimal average cost is constant and is equal to g.

3.3 Lyapunov Functions

In this section, a structural condition on the model $\mathscr{M}$ will be introduced under which (a) the basic condition (3.1) holds, (b) the optimality equation (3.7) has a solution $(g,h(\cdot))$ such that the convergence (3.9) occurs, and (c) a policy $f \in \mathscr{F}$

satisfying (3.8) exists, so that the conclusions (i)–(iii) stated above hold. Throughout the rest of this chapter

$$z \in S \quad \text{is a fixed state}$$

and T stands for the first return time to state z, that is,

$$T := \min\{n > 0 \,|\, X_n = z\}, \tag{3.12}$$

where, as usual, the minimum of the empty set is ∞. The following idea was introduced in [14] and was analyzed in [5]:

Definition 3.3.1. Let $D \in \mathscr{C}(\mathbb{K})$ and $\ell : S \to [1,\infty)$ be given functions. The function ℓ is a Lyapunov function for D, or "D has the Lyapunov function ℓ", if the following conditions (i)–(iii) hold:

(i) $1 + |D(x,a)| + \sum_{y \neq z} p_{xy}(a)\ell(y) \leq \ell(x)$ for all $(x,a) \in \mathbb{K}$.

(ii) For each $x \in S$, the mapping $f \mapsto \sum_y p_{xy}(f(x))\ell(y) = E_x^f[\ell(X_1)]$ is continuous in $f \in \mathscr{F}$.

(iii) For each $f \in \mathscr{F}$ and $x \in S$, $E_x^f[\ell(X_n)I[T > n]] \to 0$ as $n \to \infty$.

The sentence "D admits a Lyapunov function" means that there exists a function $\ell : S \to [1,\infty)$ such that conditions (i)–(iii) above hold.

The following simple lemma will be useful.

Lemma 3.3.1. *Suppose that C has the Lyapunov function $\ell(\cdot)$. In this case the assertions (i) and (ii) below are valid.*

(i) For every $n \in \mathbb{N}$ and $\pi \in \mathscr{P}$,

$$\frac{1}{n+1} E_x^\pi \left[\sum_{t=0}^{n} (1 + |C(X_t, A_t)|) + \ell(X_{n+1}) \right] \leq B(x) := \ell(x) + \ell(z), \quad x \in S;$$

in particular, the basic condition (3.1) holds.

(ii) $\lim_{n\to\infty} \frac{1}{n} E_x^\pi[\ell(X_n)] \to 0.$

Proof. Notice that the inequality $1 + |C(x,a)| + \sum_{y \in S} p_{xy}(a)\ell(y) \leq \ell(x) + \ell(z)$ is always valid, by Definition 3.3.1(i), and then, an induction argument using the Markov property yields that for arbitrary $x \in S$ and $\pi \in \mathscr{P}$,

$$E_x^\pi \left[\sum_{t=0}^{n} (1 + |C(X_t, A_t)|) + \ell(X_{n+1}) \right] \leq \ell(x) + (n+1)\ell(z), \quad n \in \mathbb{N},$$

a relation that immediately yields part (i); a proof of the second assertion can be found in Lemma 3.2 of [6]. □

The following lemma, originally established by [14], shows that the existence of a Lyapunov function has important implications for the analysis of the average criteria in (3.2) and (3.4).

Lemma 3.3.2. *Suppose that the cost function C has a Lyapunov function ℓ. In this case, there exists a unique pair $(g,h(\cdot)) \in \mathbb{R} \times \mathscr{C}(S)$ such that the following assertions (i)–(v) hold:*

(i) $g = J^*(x)$ *for each* $x \in S$.
(ii) $h(z) = 0$ *and* $|h(x)| \le (1+\ell(z)) \cdot \ell(x)$ *for all* $x \in S$. *Therefore, by Lemma 3.3.1 (ii), the convergence (3.9) holds.*
(iii) The pair $(g,h(\cdot))$ satisfies the optimality equation (3.7).
(iv) For each $x \in S$, the mapping $a \mapsto \sum_{y \in S} p_{xy}(a)h(y)$ is continuous in $a \in A(x)$.
(v) An optimal stationary policy exists: For each $x \in S$, the term within brackets in the right-hand side of (3.7) has a minimizer $f(x) \in A(x)$, and the corresponding policy $f \in \mathscr{F}$ is optimal. Moreover, (3.11) holds.

A proof of this result can be essentially found in Chapter 5 of [14]; see also Lemma 3.1 in [6] for a proof of the inequality in part (ii).

Remark 3.3.1. Notice that g in Lemma 3.3.2 is uniquely determined, since it is *the* optimal (expected) average cost at every state. The function $h(\cdot)$ in the above lemma is also unique, as established in Lemma A.2(iv) in [7]. Indeed, defining the relative cost function as $C(\cdot,\cdot) - g$, the function $h(\cdot)$ is the optimal total relative cost incurred before the first return time to state z; more explicitly, $h(x) = \inf_{\pi \in \mathscr{P}} E_x^{\pi}\left[\sum_{t=0}^{T-1}[C(X_t,A_t) - g]\right]$ for all $x \in S$.

This section concludes with some simple but useful properties of Lyapunov functions stated in Lemma 3.3.3 below, whose statement involves the following notation.

Definition 3.3.2. The class $\mathscr{L}(\ell)$ consists of all functions $D \in \mathscr{C}(\mathbb{K})$ such that a positive multiple of ℓ is a Lyapunov function for D, that is, $D \in \mathscr{L}(\ell)$ if and only if

$$\text{for some } c > 0, \quad 1 + |D(x,a)| + \sum_{y \neq z} p_{xy}(a)[c\,\ell(y)] \le c\,\ell(x), \quad (x,a) \in \mathbb{K}.$$

Notice that the function $c\,\ell(\cdot)$ inherits the properties (ii) and (iii) of the function $\ell(\cdot)$ in Definition 3.3.1.

Lemma 3.3.3. *Suppose that $\ell\colon S \to [1,\infty)$ is a Lyapunov function for a function $\tilde{D} \in \mathscr{C}(\mathbb{K})$:*

(i) If $D_0 \in \mathscr{C}(\mathbb{K})$ is such that $|D_0| \le |\tilde{D}|$, then ℓ is also a Lyapunov function for D_0.
(ii) With the notation in Definition 3.3.2, the following properties (a) and (b) hold:

(a) $\mathscr{L}(\ell)$ is a vector space that contains the constant functions.
(b) If $D_1, D_2 \in \mathscr{L}(\ell)$, then $\max\{D_1,D_2\}$ and $\min\{D_1,D_2\}$ also belong to $\mathscr{L}(\ell)$.

Proof. The first part follows directly from Definition 3.3.1. To establish part (ii), first notice that $\mathscr{L}(\ell)$ is nonempty since $\tilde{D} \in \mathscr{L}(\ell)$. Next, suppose that $D_1, D_2 \in \mathscr{L}(\ell)$ and observe that

$$1 + |D_i(x,a)| + \sum_{y \neq z} p_{xy}(a)[c_i\,\ell(y)] \leq c_i\,\ell(x), \quad (x,a) \in \mathbb{K}, \quad i = 1,2,$$

where c_1 and c_2 are positive constants. If d_1 and d_2 are real numbers, multiplying both sides of the above equality by $1 + |d_i|$, it follows that, for every $(x,a) \in \mathbb{K}$,

$$([1 + |d_i|) + (1 + |d_i|)|D_i(x,a)| + \sum_{y \neq z} p_{xy}(a)[(1 + |d_i|)c_i\,\ell(y)] \leq (1 + |d_i|)c_i\,\ell(x),$$

and then,

$$1 + |d_i D_i(x,a)| + \sum_{y \neq z} p_{xy}(a)[(1 + |d_i|)c_i\,\ell(y)] \leq (1 + |d_i|)c_i\,\ell(x), \quad i = 1,2.$$

these inequalities, it follows that

$$\begin{aligned} 1 + [1 + |d_1 D_1(x,a)| + |d_2 D_2(x,a)|] + \sum_{y \neq z} p_{xy}(a)[(1 + |d_1|)c_1 + (1 + |d_2|)c_2]\,\ell(y) \\ \leq [(1 + |d_1|)c_1 + (1 + |d_2|)c_2]\,\ell(x), \end{aligned}$$

showing that

$$1 + |d_1 D_1| + |d_2 D_2| \in \mathscr{L}(\ell),$$

and because the function in the left-hand side of this inclusion dominates $|d_1 D_1 + d_2 D_2|$ and 1, part (i) yields that (a) $d_1 D_1 + d_2 D_2 \in \mathscr{L}(\ell)$ and $1 \in \mathscr{L}(\ell)$, so that $\mathscr{L}(\ell)$ is a vector space that contains the constant functions. Finally, since $|\max\{D_1, D_2\}|, |\min\{D_1, D_2\}| \leq |D_1| + |D_2|$, the above displayed inclusion and part (i) together imply that (b) $\max\{D_1, D_2\}, \min\{D_1, D_2\} \in \mathscr{L}(\ell)$. □

3.4 A Double Lyapunov Function Condition and Sample-Path Optimality

In this section, a notion of sample-path average optimality is introduced, and using the idea of Lyapunov function, a structural condition on the decision model $\mathscr{M}$ is formulated. Under such an assumption, it is stated in Theorem 3.4.1 below that, in addition to being expected average optimal, a stationary policy f satisfying the (3.8) is also average optimal in the sample-path sense.

Definition 3.4.1. A policy $\pi^* \in \mathscr{P}$ is sample-path average optimal with optimal value $g^* \in \mathbb{R}$ if the following conditions (i) and (ii) are valid:

(i) For each state $x \in S$, $\lim_{n\to\infty} \frac{1}{n} \sum_{t=0}^{n-1} C(X_t, A_t) = g^* \quad P_x^{\pi^*}$-a.s. ;

(ii) For every $\pi \in \mathscr{P}$ and $x \in S$, $\liminf_{n\to\infty} \frac{1}{n} \sum_{t=0}^{n-1} C(X_t, A_t) \geq g^* \quad P_x^{\pi}$-a.s..

The existence and construction of sample-path optimal policies will be studied under the following structural condition on the model $\mathscr{M}$.

Assumption 3.4.1 [Double Lyapunov function condition]

(i) The cost function $C(\cdot,\cdot)$ has a Lyapunov function ℓ.
(ii) For some $\beta > 2$, the mapping ℓ^β admits a Lyapunov function.

A simple class of models satisfying this assumption is presented below.

Example 3.4.1. For each $t \in \mathbb{N}$, let $X_t \in S = \mathbb{N}$ be the number of customers waiting for a service at a time t in a single-server station. To describe the evolution of $\{X_t\}$, let the action set A be a compact metric space and set $A(x) = A$ for every x. Next, let $\{\Delta_t(a) \,|\, t \in \mathbb{N}, a \in A\}$ and $\{\xi_t(a) \,|\, t \in \mathbb{N}, a \in A\}$ be two families of independent and identically distributed random variables taking values in the set $\mathbb{N}$, and suppose that the following conditions hold:

(i) For each $k \in \mathbb{N}$, the mappings $a \mapsto P[\Delta_t(a) = k]$, $a \mapsto P[\xi_t(a) = k]$, and $a \mapsto E[\xi_t(a)^r] \in (0,\infty)$, $1 \leq r \leq 2m+3$, are continuous, where $m \in \mathbb{N} \setminus \{0\}$ is fixed.
(ii) $P[\Delta_t(a) = 1] = \mu(a) = 1 - P[\Delta_t(a) = 0]$ and $E[\xi_t(a)] - \mu(a) \leq -\rho < 0$ for all $a \in A$.

When the action $a \in A$ is chosen, (the Bernoulli variable) $\Delta_t(a)$ and $\xi_t(a)$ represent the number of service completions and arrivals in $[t, t+1)$, respectively, and the evolution of the state process is determined by

$$X_{t+1} = X_t + \xi_t(a) - \Delta_t(a) I_{\mathbb{N}\setminus\{0\}}(X_t) \quad \text{if } A_t = a, \quad t \in \mathbb{N}, \tag{3.13}$$

an equation that allows to obtain the transition law and to show that $p_{xy}(a)$ is a continuous function of $a \in A$, by the first of the conditions presented above. In the third part of the following proposition, a class of cost functions satisfying Assumption 3.4.1 is identified. □

Proposition 3.4.1. *In the context of Example 3.4.1, the following assertions hold when $z = 0$ and T is the first return time in (3.12):*

(i) For each $r = 1, 2, \ldots, 2m+2$, there exist positive constants b_r and c_r such that the function $\ell_{r+1} \in \mathscr{C}(S)$ given by

$$\ell_{r+1}(x) = x^{r+1} + b_r x + c_r, \quad x \in S, \tag{3.14}$$

satisfies

$$\rho x^r + 1 + E[\ell_{r+1}(X_{t+1}) \,|\, X_t = x, A_t = a] \le \ell_{r+1}(x), \quad x \in S, \tag{3.15}$$

where $\rho > 0$ is the number in condition (ii) stated in Example 3.4.1. Moreover, for each $x \in S$,

$$a \mapsto E[\ell_{r+1}(X_{t+1}) \,|\, X_t = x, A_t = a], \quad a \in A, \quad \textit{is continuous}, \tag{3.16}$$

and for every $\pi \in \mathscr{P}$,

$$\lim_{n\to\infty} E_x^{\pi}[(1+\rho X_n^r) I[T > n]] = 0. \tag{3.17}$$

Consequently,

(ii) For $j = 1, 2, \ldots, 2m+1$, the above mapping $\ell_{j+1}(\cdot)$ is a Lyapunov function for the cost function

$$C_j(x) :\, = \rho x^j, \quad x \in S.$$

(iii) Suppose that, for some integer $j = 1, 2, \ldots, m-1$, the cost function $C \in \mathscr{C}(\mathbb{K})$ is such that

$$\max_{a\in A} |C(x,a)| \le b_1 x^j + b_0, \quad x \in S,$$

where b_0 and b_1 are positive constants. In this case, the function C satisfies Assumption 3.4.1.

Proof. (i) For $X_t = x \neq 0$ and $1 \le r \le 2m+2$, the evolution equation (3.13) yields that

$$\begin{aligned} E[(X_{t+1})^{r+1} | X_t = x, A_t = a] &= E[(x + \xi_t(a) - \Delta_t(a))^{r+1}] \\ &\le x^{r+1} - (r+1)\rho x^r + R(x,a), \end{aligned} \tag{3.18}$$

where $\sup_{a\in A} |R(x,a)| = O(x^{r-1})$; thus, there exists a constant $b > 0$ such that $R(x,a) \le \rho x^r + b$ for every $x \neq 0$, and it follows that

$$\rho x^r - b + E[(X_{t+1})^{r+1} | X_t = x, A_t = a] \le x^{r+1}.$$

When $r = 0$, the term $R(x,a)$ in (3.18) is null, so that $\rho + E[X_{t+1} | X_t = x, A_t = a] \le x$; multiplying both sides of this relation by a sufficiently large constant b_r such that $\rho b_r > b + 1$ and combining the resulting inequality with the one displayed above, it follows that

$$\rho x^r + 1 + E[(X_{t+1})^{r+1} + b_r X_{t+1} | X_t = x, A_t = a] \le x^{r+1} + b_r x, \quad x \in S \setminus \{0\}.$$

Defining $c_r :\, = 1 + \max_{a\in A} E[\xi_t(a)^{r+1} + b_r \xi_t(a)]$, it follows that the function ℓ_{r+1} in (3.14) satisfies the inequality (3.15), whereas (3.16) follows from the

continuity properties of the distributions of the departure and arrival streams in Example 3.4.1. On the other hand, (3.15) yields that the inequality $E_x^\pi[\rho X_0^r + 1 + \ell_{r+1}(X_1)I[T > 1]] \le \ell_{r+1}(x)$ is always valid, and an induction argument using the Markov property leads to

$$E_x^\pi\left[\sum_{t=0}^{n-1}(\rho X_t^r + 1)I[T > t] + \ell_{r+1}(X_n)I[T > n]\right] \le \ell_{r+1}(x), \quad n \in \mathbb{N};$$

taking the limit as n goes to $+\infty$, this implies that $E_x^\pi[\sum_{t=0}^{\infty}(\rho X_t^r + 1)I[T > t]] \le \ell_{r+1}(x)$, and (3.17) follows.

(ii) The relations (3.15) and (3.16) immediately show that ℓ_{j+1} satisfies the requirements (i) and (ii) in Definition 3.3.1 of a Lyapunov function for C_j. Next, observe that $\ell_{j+1} \le c_0 x^{j+1} + c_1$ for some constants c_0 and c_1, and then, (3.17) with $j+1 (\le m+2)$ instead of r implies that ℓ_{j+1} also satisfies the third property in Definition 3.3.1.

(iii) Observe that the condition on the function $C(\cdot,\cdot)$ can be written as $|C| \le b_1 C_j + b_0$, and then, it is sufficient to show that C_j satisfies Assumption 3.4.1, since in this case the corresponding conclusion for C follows from Lemma 3.3.3. Let the integer j between 1 and $m-1$ be arbitrary and notice that part (ii) yields that ℓ_{j+1} is a Lyapunov function for C_j. Next, set $\beta = (2j+3)/(j+1) > 2$ and observe that (3.14) implies that there exist positive constants c_0 and c_1 such that $\ell_{j+1}^\beta(x) \le c_1 x^{2j+3} + c_0 = c_1 C_{2j+3}(x) + c_0$; since $2j+3 \le 2m+1$, part (ii) shows that C_{2j+3} has a Lyapunov function, and then, ℓ_{j+1}^β also admits a Lyapunov function, by Lemma 3.3.3.

□

The following result establishes the existence of sample-path average optimal stationary polices.

Theorem 3.4.1. *Suppose that Assumptions 3.2.1 and 3.4.1 hold, and let $(g, h(\cdot))$ be the solution of the optimality equation guaranteed by Lemma 3.3.2. In this case, if the stationary policy f satisfies (3.8), then f is sample-path average optimal with the optimal value g. More explicitly, for each $x \in S$ and $\pi \in \mathscr{P}$,*

$$\lim_{n\to\infty}\frac{1}{n}\sum_{t=0}^{n-1} C(X_t, A_t) = g \quad P_x^f\text{-a.s.} \tag{3.19}$$

and

$$\liminf_{n\to\infty}\frac{1}{n}\sum_{t=0}^{n-1} C(X_t, A_t) \ge g \quad P_x^\pi\text{-a.s.} \tag{3.20}$$

Remark 3.4.1. This theorem is related to Theorem 4.1 in [6] where MDPs with average reward criteria were considered. In the context of the present work, Theorem 4.1 in that paper establishes that, under Assumption 3.2.1, if the cost function has the Lyapunov function ℓ, then $\limsup_{n\to\infty} n^{-1}\sum_{t=0}^{n-1} C(X_t, A_t) \ge g$ P_x^π-a.s. for each

$\pi \in \mathscr{P}$ and $x \in S$ and that the equality holds if $\pi = f$ satisfies (3.8). In the present Theorem 3.4.1, the additional condition in Assumption 3.4.1(ii) is incorporated, and in this context, the stronger conclusions (3.19) and (3.20) are obtained.

A proof of Theorem 3.4.1 will be given after presenting the necessary preliminaries in the following section:

3.5 Innovations of the Sequence of Optimal Relative Costs

This section contains the main technical tool that will be used to establish Theorem 3.4.1. The necessary result concerns properties of the sequence of innovations associated to $\{h(X_t)\}$ which is introduced below. Throughout the remainder of this chapter Assumptions 3.2.1 and 3.4.1 are supposed to be valid even without explicit reference, and $(g, h(\cdot)) \in \mathbb{R} \times \mathscr{C}(S)$ stands for the pair satisfying the optimality equation (3.7) as described in Lemma 3.3.2. Next, for each positive integer n, let $\mathscr{F}_n$ be the σ-field generated by the states observed and actions applied up to time n:

$$\mathscr{F}_n := \sigma(X_t, A_t, 0 \leq t \leq n), \quad n = 1, 2, 3, \ldots, \tag{3.21}$$

and observe that for each initial state x and $\pi \in \mathscr{P}$, the Markov property of the decision process yields that

$$E_x^{\pi}[h(X_n) \,|\, \mathscr{F}_{n-1}] = \sum_{y \in S} p_{X_{n-1}y}(A_{n-1}) h(y). \tag{3.22}$$

Definition 3.5.1. The process of $\{Y_k, k \geq 1\}$ of *innovations* associated to the sequence of observed optimal relative costs $\{h(X_k), k \geq 1\}$ is given by

$$Y_n = h(X_n) - \sum_{y \in S} p_{X_{n-1},y}(A_{n-1}) h(y), \quad n = 1, 2, 3, \ldots .$$

Now, let $x \in S$ and $\pi \in \mathscr{P}$ be arbitrary but fixed, and notice that combining the definition above with (3.21) and (3.22), it follows that (i) Y_n is $\mathscr{F}_n$ measurable, and (ii) the innovations Y_n can be written as

$$Y_n = h(X_n) - E_x^{\pi}[h(X_n) \,|\, \mathscr{F}_{n-1}], \tag{3.23}$$

and then, Y_n is uncorrelated with the σ-field $\mathscr{F}_{n-1}$ with respect to P_x^{π}, that is,

$$E_x^{\pi}[Y_n W] = 0 \text{ if } W \text{ is } \mathscr{F}_{n-1} \text{ measurable and } Y_n W \text{ is } P_x^{\pi} \text{ integrable} \tag{3.24}$$

([2]). The following is the main result of this section.

Theorem 3.5.1. *Suppose that Assumptions 3.2.1 and 3.4.1 hold and let the pair $(g, h(\cdot)) \in \mathbb{R} \times \mathscr{C}(S)$ be as in Lemma 3.3.2. In this context, for each initial state $x \in S$ and $\pi \in \mathscr{P}$, the following convergences hold:*

$$\lim_{n\to\infty} \frac{h(X_n)}{n} = 0 \quad P_x^\pi\text{-a.s.} \tag{3.25}$$

and

$$\lim_{n\to\infty} \frac{1}{n} \sum_{k=1}^{n} Y_k = 0 \quad P_x^\pi\text{-a.s.} \tag{3.26}$$

This theorem will be established using two elementary facts stated in the following lemmas. The first one is a criterion for almost sure convergence, which is a consequence of the first Borel-Cantelli lemma [2].

Lemma 3.5.1. *Let $\{W_n\}$ be a sequence of random variables defined on a probability space $(\Omega, \mathscr{F}, P)$. In this case, if $\sum_{n=1}^{\infty} P[|W_n| > \varepsilon] < \infty$ for each $\varepsilon > 0$, then $\lim_{n\to\infty} W_n = 0$ P-a.s..*

The second result involved in the proof of Theorem 3.5.1 is the following inequality by Kolmogorov.

Lemma 3.5.2. *If n and k are two positive integers such that $n > k$, then for every $\alpha > 0$*

$$P_x^\pi \left[\max_{r:k\le r\le n} \left| \sum_{t=k}^{r} Y_t \right| \ge \alpha \right] \le \frac{1}{\alpha^2} \sum_{t=k}^{n} E_x^\pi[Y_t^2].$$

This classical result is established as Theorem 22.4 in [2] for the case in which the Y_n's are independent. In the present context, from the relations (3.27)–(3.29) below, it follows that $E_x^\pi[Y_n^2]$ is always finite, and then, (3.24) yields that, if $n > k$, then $E_x^\pi[Y_n Y_k I[G]] = 0$ for every $G \in \mathscr{F}_k$; from this last observation, the same arguments in the aforementioned book allow to establish Lemma 3.5.2. Now, let ℓ be a Lyapunov function for the cost function C such that ℓ^β admits a Lyapunov function for some $\beta > 2$, as ensured by Assumption 3.4.1. Applying Lemma 3.3.1 to the cost function ℓ^β, it follows that there exists a function $b\colon S \to (0,\infty)$ such that

$$\frac{1}{n+1} E_x^\pi \left[\sum_{t=0}^{n} \ell^2(X_t) \right] \le \frac{1}{n+1} E_x^\pi \left[\sum_{t=0}^{n} \ell^\beta(X_t) \right] \le b(x), \quad x \in S, \tag{3.27}$$

where the first inequality is due to the fact that $\ell(\cdot) \ge 1$.

Proof of Theorem 3.5.1. Let $\varepsilon > 0$ be arbitrary and notice that, by Lemma 3.3.2(ii),

$$|h(\cdot)| \le c\ell(\cdot), \tag{3.28}$$

where $c = 1 + \ell(z)$. Combining this relation with Markov's inequality, it follows that, for each positive integer n,

$$P_x^\pi[|h(X_n)/n| > \varepsilon] \le \frac{E_x^\pi[|h(X_n)|^\beta]}{n^\beta \varepsilon^\beta} \le \frac{c^\beta E_x^\pi[\ell(X_n)^\beta]}{n^\beta \varepsilon^\beta}$$

and then, (3.27) yields that

$$P_x^\pi[|h(X_n)/n| > \varepsilon] \le \frac{c^\beta (n+1)b(x)}{n^\beta \varepsilon^\beta} \le 2\frac{c^\beta b(x)}{n^{\beta-1}\varepsilon^\beta};$$

since $\beta > 2$, it follows that $\sum_{n=1}^\infty P_x^\pi[|h(X_n)/n| > \varepsilon] < \infty$, and the convergence (3.25) follows from Lemma 3.5.1. Next, using that the (unconditional) variance of a random variable is an upper bound for the expectation of its conditional variance [19], from (3.23), it follows that

$$\begin{aligned} E_x^\pi[Y_t^2] &= E_x^\pi[(h(X_t) - E_x^\pi[h(X_t)|\mathscr{F}_{t-1}])^2] \\ &\le E_x^\pi[(h(X_t) - E_x^\pi[h(X_t)])^2] \\ &\le E_x^\pi[h(X_t)^2] \\ &\le c^2 E_x^\pi[\ell(X_t)^2]; \end{aligned} \tag{3.29}$$

see (3.28) for the last inequality. This fact and Lemma 3.5.2 together lead to

$$P_x^\pi\left[\max_{r:\,k\le r\le n}\left|\sum_{t=k}^r Y_t\right| > \alpha\right] \le \frac{c^2}{\alpha^2}\sum_{t=k}^n E_x^\pi[\ell(X_t)^2],$$

and then, (3.27) yields that

$$P_x^\pi\left[\max_{r:\,k\le r\le n}\left|\sum_{t=k}^r Y_t\right| > \alpha\right] \le \frac{c^2(n+1)b(x)}{\alpha^2}, \quad \alpha > 0, \quad n > k \ge 1. \tag{3.30}$$

Using this relation with $k = 1$, $n = m^2$, and $\alpha = \varepsilon m^2$, it follows that

$$\begin{aligned} q_m := P_x^\pi\left[m^{-2}\left|\sum_{t=1}^{m^2} Y_t\right| > \varepsilon\right] &\le P_x^\pi\left[\max_{r:\,1\le r\le m^2}\left|\sum_{t=1}^r Y_t\right| > m^2\varepsilon\right] \\ &\le \frac{c^2(m^2+1)b(x)}{\varepsilon^2 m^4}. \end{aligned}$$

Therefore, $\sum_{m=1}^\infty q_m < \infty$, and recalling that $\varepsilon > 0$ is arbitrary, an application of Lemma 3.5.1 implies that

$$\lim_{m\to\infty}\frac{1}{m^2}\sum_{t=1}^{m^2} Y_t = 0 \quad P_x^\pi\text{-a.s.} \tag{3.31}$$

On the other hand, given a positive integer m, from the inclusion

$$\left[\max_{j:0\le j\le 2m}\left|(m^2+j)^{-1}\sum_{t=m^2}^{m^2+j}Y_t\right|\ge\varepsilon\right]\subset\left[\max_{j:0\le j\le 2m}\left|\sum_{t=m^2}^{m^2+j}Y_t\right|\ge m^2\varepsilon\right],$$

it follows that

$$\begin{aligned}p_m:&=P_x^\pi\left[\max_{j:0\le j\le 2m}\left|(m^2+j)^{-1}\sum_{t=m^2}^{m^2+j}Y_t\right|\ge\varepsilon\right]\\&\le P_x^\pi\left[\max_{j:0\le j\le 2m}\left|\sum_{t=m^2}^{m^2+j}Y_t\right|\ge m^2\varepsilon\right]\\&=P_x^\pi\left[\max_{r:m^2\le r\le(m+1)^2-1}\left|\sum_{t=m^2}^{r}Y_t\right|\ge m^2\varepsilon\right]\\&\le\frac{c^2(m+1)^2b(x)}{\varepsilon^2m^4,}\end{aligned}$$

where the last inequality was obtained from (3.30) with $n=(m+1)^2-1$, $k=m^2$, and $\alpha=m^2\varepsilon$. Thus, $\sum_{m=1}^{\infty}p_m<\infty$, and then, Lemma 3.5.1 implies that

$$\lim_{m\to\infty}\left\{\max_{j:0\le j\le 2m}\left|(m^2+j)^{-1}\sum_{t=m^2}^{m^2+j}Y_t\right|\right\}=0\quad P_x^\pi\text{-a.s.}\tag{3.32}$$

To conclude, let n be a positive integer and let m be the integral part of $\sqrt{n}$, so that

$$n=m^2+i,\quad 0\le i\le 2m.$$

Assume that i is positive and notice that in this case

$$\left|\frac{1}{n}\sum_{t=1}^{n}Y_t\right|\le\frac{m^2}{n}\left|\frac{1}{m^2}\sum_{t=1}^{m^2}Y_t\right|+\frac{1}{m^2+i}\left|\sum_{t=m^2+1}^{m^2+i}Y_t\right|,$$

and then,

$$\left|\frac{1}{n}\sum_{t=1}^{n}Y_t\right|\le\left|\frac{1}{m^2}\sum_{t=1}^{m^2}Y_t\right|+\max_{j:0\le j\le 2m}\left\{\frac{1}{m^2+j}\left|\sum_{t=m^2+1}^{m^2+j}Y_t\right|\right\},$$

a relation that is also valid when $i=0$, that is, if $n=m^2$. Taking the limit when m goes to $+\infty$ in both sides of this last inequality, the convergences in (3.31) and (3.32) together imply that (3.26) holds. □

3.6 Proof of Theorem 3.4.1

In this section, a criterion for the sample-path average optimality of a policy will be derived from Theorem 3.5.1 and that result will be used to establish Theorem 3.4.1. The arguments use the following notation:

Definition 3.6.1. The discrepancy function $\Phi\colon \mathbb{K} \to \mathbb{R}$ associated to the pair $(g, h(\cdot))$ in Lemma 3.3.2 is defined by

$$\Phi(x,a) := C(x,a) + \sum_{y\in S} p_{xy}(a)h(y) - h(x) - g.$$

Notice that Φ is a continuous mapping, by Assumption 3.2.1 and Lemma 3.3.2(iv). Also, observe that the optimality equation (3.7) yields that

$$\Phi(x,a) \geq 0, \quad (x,a) \in \mathbb{K}.$$

Lemma 3.6.1. *Suppose that Assumptions 3.2.1 and 3.4.1 hold. In this context, a policy $\pi^* \in \mathscr{P}$ is sample-path average optimal if and only if*

$$\lim_{n\to\infty} \frac{1}{n} \sum_{t=0}^{n-1} \Phi(X_t, A_t) = 0 \quad P_x^{\pi^*}\text{-a.s.} \tag{3.33}$$

Proof. It will be verified that for all $x \in S$ and $\pi \in \mathscr{P}$,

$$\lim_{n\to\infty} \frac{1}{n} \sum_{t=0}^{n-1} [C(X_t, A_t) - g - \Phi(X_t, A_t)] = 0 \quad P_x^{\pi}\text{-a.s.} \tag{3.34}$$

Assuming that this relation holds, the desired conclusion can be established as follows: Observing that

$$\frac{1}{n} \sum_{t=0}^{n-1} C(X_t, A_t) = \frac{1}{n} \sum_{t=0}^{n-1} [C(X_t, A_t) - g - \Phi(X_t, A_t)] + g + \frac{1}{n} \sum_{t=0}^{n-1} \Phi(X_t, A_t),$$

and taking the inferior limit as n goes to ∞ in both sides of this equality, the nonnegativity of the discrepancy function and (3.34) together imply that the relation

$$\liminf_{n\to\infty} \frac{1}{n} \sum_{t=0}^{n-1} C(X_t, A_t) \geq g, \quad P_x^{\pi}\text{-a.s.}$$

is always valid. Thus, by Definition 3.4.1, $\pi^* \in \mathscr{P}$ is sample-path average optimal if and only if

$$\lim_{n\to\infty} \frac{1}{n} \sum_{t=0}^{n-1} C(X_t, A_t) = g \quad P_x^{\pi^*}\text{-a.s.}, \quad x \in S,$$

a property that, by (3.34) with π^* instead of π, is equivalent to the criterion (3.33).

Thus, to conclude the argument, it is sufficient to verify the statement (3.34). To achieve this goal, notice that the definition of the discrepancy function yields that following equality is always valid for $t \geq 1$:

$$C(X_{t-1}) - g - \Phi(X_{t-1}, A_{t-1}) = h(X_{t-1}) - \sum_{y \in S} p_{X_{t-1},y}(A_{t-1}) h(y),$$

a relation that, *via* the specification of the innovation Y_t in Definition 3.5.1, leads to

$$C(X_{t-1}) - g - \Phi(X_{t-1}, A_{t-1}) = h(X_{t-1}) - h(X_t) + Y_t.$$

Therefore,

$$\sum_{t=1}^{n} [C(X_{t-1}) - g - \Phi(X_{t-1}, A_{t-1})] = h(X_0) - h(X_n) + \sum_{t=1}^{n} Y_t,$$

and then, for every initial state $X_0 = x$ and $\pi \in \mathscr{P}$,

$$\frac{1}{n} \sum_{t=1}^{n} [C(X_{t-1}) - g - \Phi(X_{t-1}, A_{t-1})] = \frac{h(x)}{n} - \frac{h(X_n)}{n} + \frac{1}{n} \sum_{t=1}^{n} Y_t, \quad P_x^{\pi}\text{-a.s.},$$

where the equality $P_x^{\pi}[X_0 = x] = 1$ was used; from this point, (3.34) follows directly from Theorem 3.5.1. □

Proof of Theorem 3.4.1. From Definitions 3.6.1 and (3.8), it follows that $\Phi(x, f(x)) = 0$ for every state x. Thus, using that $A_t = f(X_t)$ when the system is running under the policy f, it follows that, for every initial state x and $t \in \mathbb{N}$, the equality $\Phi(X_t, A_t) = \Phi(X_t, f(X_t)) = 0$ holds with probability 1 with respect to P_x^f. Therefore, the criterion (3.33) is satisfied by f, and then, Lemma 3.6.1 yields that the policy f is sample-path average optimal. □

3.7 Approximations Schemes and Sample-Path Optimality

In the remainder of this chapter the sample-path optimality of Markovian policies is analyzed. The interest in this problem stems from the fact that an explicit solution $(g, h(\cdot))$ of the optimality equation (3.7) is seldom available, and in this case, the sample-path average optimal policy f in (3.8) cannot be determined. When a solution of the optimality equation is not at hand, an iterative approximation procedure is implemented and (i) approximations $\{(g_n, h_n(\cdot))\}_{n \in \mathbb{N}}$ for $(g, h(\cdot))$ are generated, and (ii) a stationary policy f_n is obtained from $(g_n, h_n(\cdot))$. Such a policy f_n is "'nearly optimal"' in the sense that, for each fixed x, the convergence $\Phi(x, f_n(x)) \to 0$ occurs as $n \to \infty$, and the next objective is to establish the sample-path average optimality of the Markovian policy $\{f_n\}$.

Remark 3.7.1. Two procedures that can be used to approximate the solution of the optimality equation and to generate a Markovian policy $\{f_n\}$ such that the f_n's are nearly optimal are briefly described below; for details see, for instance, [1,9,11–13], or [17].

(i) The discounted method. For each $\alpha \in (0,1)$, the total expected α-discounted cost at the state x under $\pi \in \mathscr{P}$ is given by $V_\alpha(x,\pi) := E_x^\pi\left[\sum_{t=0}^\infty \alpha^t C(X_t,A_t)\right]$, whereas $V_\alpha^*(x) := \inf_{\pi\in\mathscr{P}} V_\alpha(x,\pi)$, $x \in S$, is the α-optimal value function, which satisfies the optimality equation

$$V_\alpha(x) = \inf_{a\in A(x)}\left[C(x,a) + \alpha\sum_{y\in S} p_{xy}(a)V_\alpha(y)\right], \quad x \in S.$$

Now let $\{\alpha_n\} \subset (0,1)$ be a sequence increasing to 1, and define

$$(g_n,h_n) := ((1-\alpha_n)V_{\alpha_n}(z), V_{\alpha_n}(\cdot) - V_{\alpha_n}(z))$$

and let the policy f_n be such that

$$V_{\alpha_n}(x) = C(x,f_n(x)) + \alpha_n\sum_{y\in S} p_{xy}(f_n(x))V_{\alpha_n}(y), \quad x \in S. \tag{3.35}$$

(ii) Value iteration. This procedure approximates the solution $(g,h(\cdot))$ of the optimality equation (3.7) using the total cost criterion over a finite horizon. For each $n \in \mathbb{N}\setminus\{0\}$ let $J_n(x,\pi)$ be the total cost incurred when the system runs during n steps under policy π starting at x, that is, $J_n(x,\pi) := E_x^\pi\left[\sum_{t=0}^{n-1} C(X_t,A_t)\right]$, and let $J_n^*(x) := \inf_{\pi\in\mathscr{P}} J_n(x,\pi)$ be the corresponding optimal value function; the sequence $\{J_n^*(\cdot)\}$ satisfies the relation

$$J_n^*(x) = \inf_{a\in A(x)}\left[C(x,a) + \sum_{y\in S} p_{xy}(a)J_{n-1}^*(y)\right], \quad x\in S, \quad n = 1,2,3,\ldots, \tag{3.36}$$

where $J_0^*(\cdot) = 0$, so that the functions $J_n^*(\cdot)$ are determined recursively, which is an important feature of the method. The approximations to $(g,h(\cdot))$ are given by

$$(g_n,h_n(\cdot)) := (J_n^*(z) - J_{n-1}^*(z),\ J_n^*(\cdot) - J_n^*(z)),$$

whereas the policy f_n is such that $f_n(x)$ is a minimizer of the term within brackets in (3.36):

$$J_n^*(x) = C(x,f_n(x)) + \sum_{y\in S} p_{xy}(f_n(x))J_{n-1}^*(y), \quad x\in S, \quad n = 1,2,3,\ldots$$

Under Assumptions 3.2.1 and 3.4.1(i), the approximations $(g_n,h_n(\cdot))$ generated by the discounted method converge pointwise to $(g,h(\cdot))$, and the policies f_n in (3.35) are nearly optimal in the sense that $\Phi(x,f_n(x)) \to 0$ as $n\to\infty$. Similar

conclusions hold for the value iteration scheme if, additionally, the transition law satisfies that

$$p_{xx}(a) > 0, \quad a \in A(x), \quad x \in S;$$

this requirement can be avoided if the transformation by Schewitzer (1971) is applied to the transition law and the value iteration method is applied to the transformed model; see, for instance, [7] or [8].

The following theorem establishes a sufficient condition for the sample-path optimality of a Markovian policy.

Theorem 3.7.1. *Suppose that Assumptions 3.2.1 and 3.4.1 hold, and let* $\mathbf{f} = \{f_t\}$ *be a Markov policy such that*

$$\lim_{n\to\infty} \Phi(x, f_n(x)) = 0, \quad x \in S, \tag{3.37}$$

where Φ *is the discrepancy function introduced in Definition 3.6.1. In this case, the policy* $\mathbf{f}$ *is sample-path average optimal; see Definition 3.4.1.*

The proof of this result relies on some consequences of Theorem 3.4.1 which will be analyzed below.

3.8 Tightness and Uniform Integrability

This section contains the technical preliminaries that will be used to establish Theorem 3.7.1. The necessary results are concerned with properties of the sequence of empirical measures, which is now introduced.

Definition 3.8.1. The random sequence $\{\nu_n\}$ of empirical measures associated with the state-action process $\{(X_t, A_t)\}$ is defined by

$$\nu_n(B) := \frac{1}{n} \sum_{t=0}^{n-1} \delta_{(X_t, A_t)}(B), \quad B \in \mathscr{B}(\mathbb{K}), \quad n = 1, 2, 3, \ldots,$$

where $\delta_{\mathbf{k}}$ stands for the Dirac measure concentrated at $\mathbf{k}$, that is, $\delta_{\mathbf{k}}(B) = 1$ if $\mathbf{k} \in B$ and $\delta_{\mathbf{k}}(B) = 0$ when $\mathbf{k} \notin B$.

Notice that this specification yields that, for each positive integer n and $D \in \mathscr{C}(\mathbb{K})$,

$$\nu_n(D) := \int_{\mathbf{K}} D(\mathbf{k}) \nu_n(d\mathbf{k}) = \frac{1}{n} \sum_{t=1}^{n} D(X_t, A_t) \,.$$

The main result of this section concerns the asymptotic behavior of $\{\nu_n\}$ and involves the following notation: Given a set $\tilde{S} \subset S$, for each $D \in \mathscr{C}(\mathbb{K})$, defines the new function $D_{\tilde{S}} \in \mathscr{C}(\mathbb{K})$ as follows:

$$D_{\tilde{S}}(x,a) := \max\{|D(x,a)|,1\}I_{\tilde{S}}(x), \quad (x,a) \in \mathbb{K}. \tag{3.38}$$

Theorem 3.8.1. *Suppose that Assumptions 3.2.1 and 3.4.1 hold and let $x \in S$ and $\pi \in \mathscr{P}$ be arbitrary but fixed. In this context, for each $\varepsilon > 0$, there exists a finite set $F_\varepsilon \subset S$ such that*

$$\limsup_{n\to\infty} \nu_n(C_{S\setminus F_\varepsilon}) \leq \varepsilon \quad P_x^\pi\text{-a.s.;} \tag{3.39}$$

see (3.38).

Remark 3.8.1. For a positive integer r, let $F_{1/r}$ be the set corresponding to $\varepsilon = 1/r$ in the above theorem and define the event Ω^* by

$$\Omega^* := \bigcap_{r=1}^{\infty} \left[\limsup_{n\to\infty} \nu_n(C_{S\setminus F_{1/r}}) \leq 1/r\right].$$

(i) Let the set $\mathbb{K}_r \subset \mathbb{K}$ be given by

$$\mathbb{K}_r = \{(x,a) \in \mathbb{K} \,|\, x \in F_{1/r}\}. \tag{3.40}$$

With this notation, $\mathbb{K}_r$ is a compact set, since $F_{1/r}$ is finite, and (3.38) yields that $C_{S\setminus F_{1/r}}(x,a) \geq 1$ for $(x,a) \in \mathbb{K} \setminus \mathbb{K}_r$, so that

$$\nu_n(\mathbb{K} \setminus \mathbb{K}_r) \leq \int_{\mathbb{K}} C_{S\setminus F_{1/r}}(\mathbf{k})\nu_n(d\mathbf{k}) = \nu_n(C_{S\setminus F_{1/r}}),$$

and then,

$$\limsup_{n\to\infty} \nu_n(\mathbb{K} \setminus \mathbb{K}_r) \leq \limsup_{n\to\infty} \nu_n(C_{S\setminus F_{1/r}}).$$

Consequently, along a sample trajectory $\{(X_t,A_t)\}$ in Ω^* the corresponding sequence $\{\nu_n\}$ satisfies $\limsup_{n\to\infty} \nu_n(\mathbb{K} \setminus \mathbb{K}_r) \leq 1/r$ for every $r > 0$ so that $\{\nu_n\}$ is tight.

(ii) Given a probability measure μ defined in $\mathscr{B}(\mathbb{K})$, a function $D \in \mathscr{C}(\mathbb{K})$ is integrable with respect to μ if, and only if, for each positive integer r, there exists a compact set $\tilde{\mathbb{K}}_r$ such that

$$\int_{\mathbb{K}\setminus\tilde{\mathbb{K}}_r} |D(\mathbf{k})|\mu(d\mathbf{k}) \leq 1/r.$$

Notice now that (3.38) and (3.40) yield that $C_{S\setminus F_{1/r}}(x,a) \geq |C(x,a)|$ for $(x,a) \in \mathbb{K} \setminus \mathbb{K}_r$, so that

$$\int_{\mathbf{K}\setminus\mathbf{K}_r} |C(\mathbf{k})|\nu_n(d\mathbf{k}) \leq \int_{\mathbf{K}\setminus\mathbf{K}_r} C_{S\setminus F_{1/r}}(\mathbf{k})\nu_n(d\mathbf{k}) = \nu_n(C_{S\setminus F_{1/r}}).$$

Thus,

$$\limsup_{n\to\infty} \int_{\mathbf{K}\setminus\mathbf{K}_r} |C(\mathbf{k})|\nu_n(d\mathbf{k}) \leq \limsup_{n\to\infty} \nu_n(C_{S\setminus F_{1/r}}),$$

and then, along a sample trajectory in Ω^*,

$$\limsup_{n\to\infty} \int_{\mathbf{K}\setminus\mathbf{K}_r} |C(\mathbf{k})|\nu_n(d\mathbf{k}) \leq 1/r, \quad r = 1,2,3,\ldots,$$

showing that the cost function C is uniformly integrable with respect to the family $\{\nu_n\}$ of empirical measures.

(iii) Since Theorem 3.8.1 yields that $P_x^\pi[\Omega^*] = 1$ for every $x \in S$ and $\pi \in \mathscr{P}$, the previous discussion can be summarized as follows: Regardless of the initial state and the policy used to drive the system, the following assertions hold with probability 1: (a) The sequence $\{\nu_n\}$ is tight and (b) the cost function is uniformly integrable with respect to $\{\nu_n\}$.

The proof of Theorem 3.8.1 relies on the following lemma.

Lemma 3.8.1. *Let $\varepsilon > 0$ be arbitrary, and suppose that Assumptions 3.2.1 and 3.4.1 hold, and let $g_{S\setminus F}$ be the optimal expected average cost associated with the cost function $-C_{S\setminus F}$, that is,*

$$g_{S\setminus F} := \inf_{\pi\in\mathscr{P}} J(x,\pi,-C_{S\setminus F}), \tag{3.41}$$

where $J(x,\pi,-C_{S\setminus F})$ is given by the right-hand side of (3.2) with the function $-C_{S\setminus F}$ instead of C. With this notation, there exists a finite set $F \subset S$ such that

$$g_{S\setminus F} \geq -\varepsilon\,. \tag{3.42}$$

Proof. Let $\{F_k\}$ be a sequence of finite subsets of S such that

$$F_k \subset F_{k+1}, \quad k = 1,2,3,\ldots, \quad \text{and} \quad \bigcup_{k=1}^{\infty} F_k = S; \tag{3.43}$$

from (3.38), it follows that $-C_{S\setminus F_k} \nearrow 0$ as $k \nearrow \infty$, a property that *via* (3.41) immediately yields that

$$g_{S\setminus F_k} \leq g_{S\setminus F_{k+1}} \leq 0, \quad k = 1,2,3,\ldots,$$

so that $\{g_{S\setminus F_k}\}$ is a convergent sequence; set

$$\bar{g}\colon = \lim_{k\to\infty} g_{S\setminus F_k}. \tag{3.44}$$

To establish the desired conclusion, it is sufficient to show that

$$\bar{g} = 0,$$

since in this case (3.42) occurs when F is replaced by F_k with k large enough. To verify the above equality, let ℓ be a Lyapunov function for the function C and notice that Lemma 3.3.3 yields that, for some constant $c > 0$, the mapping $c\ell$ is a Lyapunov function for $-C_{S\setminus F_k}$, and then, this last function also satisfies Assumption 3.4.1. Thus, from Lemma 3.3.2 applied to the cost function $C_{S\setminus F_k}$, it follows that there exists a function $h_k\colon S \to \mathbb{R}$ as well as a policy $f_k \in \mathscr{F}$ such that

$$g_{S\setminus F_k} + h_k(x) = -C_{S\setminus F_k}(x, f_k(x)) + \sum_{y\in S} p_{xy}(f_k(x))h_k(y), \quad x \in S, \tag{3.45}$$

where $h_k(\cdot) \le c\,\ell(\cdot)$, that is,

$$h_k(\cdot) \in \prod_{x\in S}[c\,\ell(x), c\,\ell(x)]. \tag{3.46}$$

Using the fact that the right-hand side of this inclusion as well as $\mathscr{F}$ are compact metric spaces, it follows that there exists a sequence $\{k_r\}$ of positive integers increasing to ∞ such that the following limits exist:

$$\bar{f}(x)\colon = \lim_{r\to\infty} f_{k_r}(x), \quad \bar{h}(x)\colon = \lim_{r\to\infty} h_{k_r}(x), \quad x \in S.$$

Next, observe that (3.38) and (3.43) together yield that for each state x,

$$C_{S\setminus F_k}(x, f_k(x)) = 0 \ \text{ when } k \text{ is large enough},$$

whereas, *via* Proposition 2.18 in p. 232 of [18], the continuity property in Definition 3.3.1(ii) and (3.46) leads to

$$\lim_{r\to\infty} \sum_{y\in S} p_{xy}(f_{k_r}(x))h_{k_r}(y) = \sum_{y\in S} p_{xy}(\bar{f}(x))\bar{h}(y).$$

Replacing k by k_r in (3.45) and taking the limit as r goes to ∞ in both sides of the resulting equation, (3.44) and the three last displays allow to write that

$$\bar{g} + \bar{h}(x) = \sum_{y\in S} p_{xy}(\bar{f}(x))\bar{h}(y), \quad x \in S.$$

Starting from this relation, an induction argument yields that $(n+1)\bar{g}+\bar{h}(x) = E_x^{\bar{f}}[\bar{h}(X_{n+1})]$ for every $x \in S$ and $n \in \mathbb{N}$, that is,

$$\bar{g} = \frac{1}{n+1} E_x^{\bar{f}}[\bar{h}(X_{n+1})] - \frac{\bar{h}(x)}{n+1},$$

and taking the limit as n goes to ∞, the inclusion (3.46) and Lemma 3.3.1(ii) together imply that $\bar{g} = 0$; as already mentioned, this completes the proof of the lemma. □

Proof of Theorem 3.8.1. Recalling that Assumption 3.4.1 is in force, let ℓ be a Lyapunov function for the cost function C such that ℓ^β admits a Lyapunov function for some $\beta > 2$. As already noted, for some constant $c > 0$, the function $c\ell$ is a Lyapunov function for $-C_{S\setminus F}$, a fact that immediately implies that this last function also satisfies Assumption 3.4.1. Therefore, applying Theorem 3.4.1 with the cost function $-C_{S\setminus F}$ instead of C, it follows that for every $x \in S$ and $\pi \in \mathscr{P}$,

$$\liminf_{n\to\infty} \nu_n(-C_{S\setminus F}) \geq g_{S\setminus F}, \quad P_x^\pi\text{-a. s.},$$

and selecting F as the finite set in Lemma 3.8.1, it follows that

$$\liminf_{n\to\infty} \nu_n(-C_{S\setminus F}) \geq -\varepsilon, \quad P_x^\pi\text{-a. s.},$$

a statement that is equivalent to (3.39). □

3.9 Proof of Theorem 3.7.1

In this section a proof of the sample-path average optimality of a Markovian policy satisfying condition (3.37) will be presented. The argument combines Theorem 3.8.1 with the following lemma.

Lemma 3.9.1. *Let the Markovian policy $\mathbf{f} = \{f_t\}$ be such that (3.37) holds, and consider a fixed sample trajectory $\{(X_t, A_t)\}$ along which properties (i) and (ii) below hold:*

(i) The sequence $\{\nu_n\}$ of empirical measures is tight.
(ii) $A_t = f(X_t)$ for all $t \in N$.

In this context, if ν^* is a limit point of the sequence $\{\nu_n\}$ in the weak convergence topology, then ν^* is supported on

$$\mathbb{K}^* = \{(x,a) \in \mathbb{K} \,|\, \Phi(x,a) = 0\}, \tag{3.47}$$

that is,

$$\nu^*(\mathbb{K}^*) = 1.$$

Proof. For each $x \in S$ and $\varepsilon > 0$ defines the set

$$\mathbb{K}(x,\varepsilon) = \{(x,a) \,|\, a \in A(x) \text{ and } \Phi(x,a) > \varepsilon\},$$

which is an open subset of $\mathbb{K}$, since the function Φ is continuous and S is endowed with the discrete topology. In this case,

$$\delta_{(X_t,A_t)}(\mathbb{K}(x,\varepsilon)) = \delta_{(X_t,f_t(X_t))}(\mathbb{K}(x,\varepsilon)) = 1 \iff X_t = x \text{ and } \Phi(x,f_t(x)) > \varepsilon.$$

Therefore, using condition (3.37), it follows that there exists an integer $N > 0$ such that

$$\delta_{(X_t,A_t)}(\mathbb{K}(x,\varepsilon)) = 0, \quad t > N,$$

so that

$$\nu_n(\mathbb{K}(x,\varepsilon)) = \frac{1}{n}\sum_{t=0}^{n-1} \delta_{(X_t,A_t)}(\mathbb{K}(x,\varepsilon)) = \frac{1}{n}\sum_{t=0}^{N} \delta_{(X_t,f_t(X_t))}(\mathbb{K}(x,\varepsilon)), \quad n > N,$$

by Definition 3.8.1, and it follows that $\nu_n(\mathbb{K}(x,\varepsilon)) \to 0$ as $n \to \infty$. Therefore, recalling that $\mathbb{K}(x,\varepsilon)$ is an open subset of $\mathbb{K}$, the fact that ν^* is a limit point of the sequence $\{\nu_n\}$ implies that

$$\nu^*(\mathbb{K}(x,\varepsilon)) \le \limsup_{n\to\infty} \nu_n(\mathbb{K}(x,\varepsilon)) = 0, \quad x \in S, \quad \varepsilon > 0$$

[2]. Finally, using that $\mathbb{K} \setminus \mathbb{K}^* = \bigcup_{x\in S,\, r\in\mathbf{N}} \mathbb{K}(x,1/r)$ (because of the nonnegativity of Φ), the above inequality leads to $\nu^*(\mathbb{K} \setminus \mathbb{K}^*) = 0$. □

Proof of Theorem 3.7.1. It will be proved that

$$\text{the discrepancy function } \Phi \text{ satisfies Assumption 3.4.1.} \tag{3.48}$$

Assuming that this assertion holds, the conclusion of Theorem 3.7.1 can be obtained as follows: Let $\mathbf{f} = \{f_t\}$ be a Markovian policy satisfying the property (3.37) and, for $\tilde{S} \subset S$, define

$$\Phi_{(\tilde{S})}(x,a) := \Phi(x,a) I_{\tilde{S}}(x), \quad (x,a) \in \mathbb{K}, \tag{3.49}$$

so that the equality

$$\Phi = \Phi_{(\tilde{S})} + \Phi_{(S\setminus\tilde{S})} \tag{3.50}$$

is always valid. Next, given $\varepsilon > 0$, observe the following facts (a) and (b):

(a) The property (3.48) allows to apply Theorem 3.8.1 with the cost function C replaced by Φ to conclude that there exists a finite set $F_\varepsilon \subset S$ such that for each $x \in S$, the relation $\limsup_{n\to\infty} \nu_n(\Phi_{S\setminus F_\varepsilon}) \le \varepsilon$ occurs almost surely with respect

to $P_x^{\mathbf{f}}$; since (3.49) and (3.38) together imply that $\Phi_{(S\setminus F_\varepsilon)} \leq \Phi_{S\setminus F_\varepsilon}$, it follows that

$$\limsup_{n\to\infty} \nu_n(\Phi_{(S\setminus F_\varepsilon)}) \leq \varepsilon \quad P_x^{\mathbf{f}}\text{-a.s.} \tag{3.51}$$

(b) Let $\{(X_t, A_t)\}$ be a fixed sample trajectory along which

(i) $\{\nu_n\}$ is tight, and
(ii) $A_t = f_t(X_t)$.

Now select a sequence $\{n_k\}$ of positive integers such that $\lim_{k\to\infty} n_k = \infty$ and

$$\lim_{k\to\infty} \nu_{n_k}(\Phi_{(F_\varepsilon)}) = \limsup_{n\to\infty} \nu_n(\Phi_{(F_\varepsilon)}).$$

Because of the tightness of $\{\nu_n\}$, taking a subsequence—if necessary—it can be assumed that $\{\nu_{n_k}\}$ converges weakly to some $\nu^* \in \mathbb{P}(\mathbb{K})$, and in this case, observing that $\Phi_{(F_\varepsilon)}$ is continuous and has compact support (since $F_\varepsilon \subset S$ is finite), it follows that

$$\lim_{k\to\infty} \nu_{n_k}(\Phi_{(F_\varepsilon)}) = \nu^*(\Phi_{(F_\varepsilon)});$$

on the other hand, using that ν^* is supported in the set $\mathbb{K}^*$ specified in (3.47), by Lemma 3.9.1, and that Φ_{F_ε} is null on that set (see (3.47) and (3.49)), it follows that $\nu^*(\Phi_{F_\varepsilon}) = 0$. Combining this equality with the two last displays, it follows that $\limsup_{n\to\infty} \nu_n(\Phi_{(F_\varepsilon)}) = 0$ along a sample-path satisfying conditions (i) and (ii) above. Observing that condition (i) holds almost surely with respect to $P_x^{\mathbf{f}}$, by Remark 3.8.1, and that $P_x^{\mathbf{f}}[A_t = f_t(X_t)] = 1$ for all t, it follows that

$$\limsup_{n\to\infty} \nu_n(\Phi_{(F_\varepsilon)}) = 0 \quad P_x^{\mathbf{f}}\text{-a.s.},$$

a relation that, combined with (3.50), (3.51) and the nonnegativity of Φ, yields that the convergence $\lim_{n\to\infty} \nu_n(\Phi) = 0$ occurs with probability 1 with respect to $P_x^{\mathbf{f}}$, and then, the policy $\mathbf{f}$ is sample-path average optimal, by Lemma 3.6.1. Thus, to conclude the argument, it is sufficient to verify (3.48). To achieve this goal, let ℓ be a Lyapunov function for the cost function C such that ℓ^β admits a Lyapunov function for some $\beta > 2$, and recall that the solution $(g, h(\cdot))$ of the optimality Equation (3.7) satisfies

$$|h(\cdot)| \leq c\ell(\cdot) \tag{3.52}$$

for some $c > 0$, as well as $h(z) = 0$. Combining this last equality with the specification of the discrepancy function, it follows that, for all $(x, a) \in \mathbb{K}$,

$$h(x) = C(x,a) - g - \Phi(x,a) + \sum_{y\neq z} p_{xy}(a)h(y). \tag{3.53}$$

On the other hand, Lemma 3.3.3 yields that $|C(\cdot) - g|$ admits a Lyapunov function of the form $c_1\,\ell$ where $c_1 > 0$ so that

$$c_1\,\ell(x) \geq |C(x,a) - g| + 1 + \sum_{y\in S,\, y\neq z} p_{xy}(a) c_1\,\ell(y);$$

multiplying both sides of this relation by a constant c_2 satisfying

$$c_2\,c_1 > c \text{ and } c_2 > 1, \tag{3.54}$$

it follows that

$$c_2 c_1\,\ell(x) \geq c_2|C(x,a) - g| + c_2 + \sum_{y\in S,\, y\neq z} p_{xy}(a) c_2 c_1\,\ell(y),$$

and then,

$$c_2 c_1\,\ell(x) \geq |C(x,a) - g| + 1 + \sum_{y\in S,\, y\neq z} p_{xy}(a) c_2 c_1\,\ell(y).$$

Combining this inequality with (3.53), it is not difficult to obtain that

$$\tilde{\ell}(x) \geq 1 + \boldsymbol{\Phi}(x,a) + \sum_{y\in S,\, y\neq z} p_{xy}(a)\tilde{\ell}(y), \quad (x,a) \in \mathbb{K}, \tag{3.55}$$

where $\tilde{\ell}(\cdot) := c_1 c_2 \ell(\cdot) - h(\cdot) \geq 0$ and the inequality follows from (3.52) and (3.54). Recalling that $\boldsymbol{\Phi}$ is nonnegative, the above display yields that that $\tilde{\ell}$ takes values in $[1,\infty)$ and that $\tilde{\ell}$ satisfies the first requirement for being a Lyapunov function for $\boldsymbol{\Phi}$; setting $\tilde{c} := c_1 c_2 + c > 0$, (3.53) implies that $\tilde{\ell}(\cdot) \leq \tilde{c}\,\ell$, and then, $\tilde{\ell}$ inherits the second and third properties in Definition 3.3.1 from the corresponding ones of ℓ. Thus, $\tilde{\ell}$ is a Lyapunov function for $\boldsymbol{\Phi}$, and using that $\tilde{\ell}^{\beta} \leq \tilde{c}^{\beta}\ell^{\beta}$, it follows that $\tilde{\ell}^{\beta}$ also admits a Lyapunov function, by Lemma 3.3.3. Thus, the statement (3.48) holds, and as already mentioned, this concludes the proof of Theorem 3.7.1. □

Acknowledgment With sincere gratitude and appreciation, the authors dedicate this work to Professor Onésimo Hernández-Lerma on the occasion of his 65th anniversary, for his friendly and generous support and clever guidance.

References

1. Arapostathis A., Borkar V.K., Fernández-Gaucherand E., Ghosh M.K., Marcus S.I.: Discrete time controlled Markov processes with average cost criterion: a survey. SIAM J. Control Optim. **31**, 282–344 (1993)
2. Billingsley P. : Probability and Measure, 3rd edn. Wiley, New York (1995)

3. Borkar V.K: On minimum cost per unit of time control of Markov chains, SIAM J. Control Optim. **21**, 652–666 (1984)
4. Borkar V.K.: Topics in Controlled Markov Chains, Longman, Harlow (1991)
5. Cavazos-Cadena R., Hernández-Lerma O.: Equivalence of Lyapunov stability criteria in a class of Markov decision processes, Appl. Math. Optim. **26**, 113–137 (1992)
6. Cavazos-Cadena R., Fernández-Gaucherand E.: Denumerable controlled Markov chains with average reward criterion: sample path optimality, Math. Method. Oper. Res. **41**, 89–108 (1995)
7. Cavazos-Cadena R., Fernández-Gaucherand E.: Value iteration in a class of average controlled Markov chains with unbounded costs: Necessary and sufficient conditions for pointwise convergence, J. App. Prob. **33**, 986–1002 (1996)
8. Cavazos-Cadena R.: Adaptive control of average Markov decision chains under the Lyapunov stability condition, Math. Method. Oper. Res. **54**, 63–99 (2001)
9. Hernández-Lerma O.: Adaptive Markov Control Processes, Springer, New York (1989)
10. Hernández-Lerma O.: Existence of average optimal policies in Markov control processes with strictly unbounded costs, Kybernetika, **29**, 1–17 (1993)
11. Hernández-Lerma O., Lasserre J.B.: Value iteration and rolling horizon plans for Markov control processes with unbounded rewards, J. Math. Anal. Appl. **177**, 38–55 (1993)
12. Hernández-Lerma O., Lasserre J.B.: Discrete-time Markov control processes: Basic optimality criteria, Springer, New York (1996)
13. Hernández-Lerma O., Lasserre J.B.: Further Topics on Discrete-time Markov Control Processes, Springer, New York (1999)
14. Hordijk A.: Dynamic Programming and Potential Theory (Mathematical Centre Tract 51.) Mathematisch Centrum, Amsterdam (1974)
15. Hunt F.Y.: Sample path optimality for a Markov optimization problem, Stoch. Proc. Appl. **115**, 769–779 (2005)
16. Lasserre J.B:: Sample-Path average optimality for Markov control processes, IEEE T. Automat. Contr. **44**, 1966–1971 (1999)
17. Montes-de-Oca R., Hernández-Lerma O.: Value iteration in average cost Markov control processes on Borel spaces, Acta App. Math. **42**, 203–221 (1994)
18. Royden H.L.: Real Analysis, 2nd edn. MacMillan, New York (1968)
19. Shao J.: Mathematical Statistics, Springer, New York (1999)
20. Vega-Amaya O.: Sample path average optimality of Markov control processes with strictly unbounded costs, Applicationes Mathematicae, **26**, 363–381 (1999)