J. Appl. Prob. **42,** 905–918 (2005)
Printed in Israel

NONSTATIONARY VALUE ITERATION IN CONTROLLED MARKOV CHAINS WITH RISK-SENSITIVE AVERAGE CRITERION

ROLANDO CAVAZOS-CADENA,* *Universidad Autónoma Agraria Antonio Narro*
RAÚL MONTES-DE-OCA,** *Universidad Autónoma Metropolitana*

Abstract

This work concerns Markov decision chains with finite state spaces and compact action sets. The performance index is the long-run risk-sensitive average cost criterion, and it is assumed that, under each stationary policy, the state space is a communicating class and that the cost function and the transition law depend continuously on the action. These latter data are not directly available to the decision-maker, but convergent approximations are known or are more easily computed. In this context, the nonstationary value iteration algorithm is used to approximate the solution of the optimality equation, and to obtain a nearly optimal stationary policy.

Keywords: Nonstationary successive approximations algorithm; approximate solution to the optimality equation; Schweitzer's transformation; Birkhoff's contraction coefficient

2000 Mathematics Subject Classification: Primary 90C40; 93E20

1. Introduction

This paper deals with Markov decision processes (MDPs) with finite state spaces and compact action sets. The controller has (arbitrary but constant) risk sensitivity $\lambda > 0$, and the performance index of a control policy is the corresponding risk-sensitive average cost criterion. Besides standard continuity requirements, it is assumed that the whole state space is a communicating class under each stationary policy. In these circumstances, it has been shown that the optimal value function is constant and is characterized by the λ-sensitive optimality equation (λ-OE) (see Cavazos-Cadena and Fernández-Gaucherand (2002)). A similar result was obtained for MDPs with denumerable state spaces in Borkar and Meyn (2002) – where a penalized cost structure was assumed – and, under appropriate mixing conditions, for MDPs on Borel spaces in Di Masi and Stettner (1999), (2000). In the finite state space context described above, our main objective is to approximate the solution to the λ-OE when the exact transition law and cost function are not immediately available to the controller, but convergent approximations to these data are known or are more easily computed. Following Federgruen and Schweitzer (1981), this problem will be approached via the nonstationary value iteration algorithm, which has been widely used to construct adaptive optimal policies in the risk-neutral case (see Hernández-Lerma (1989)). The result, stated as Theorem 3.1, is based on (a) the application of the extended Schweitzer transformation introduced in Cavazos-Cadena

Received 21 January 2004; revision received 13 June 2005.

* Postal address: Departamento de Estadística y Cálculo, Universidad Autónoma Agraria Antonio Narro, Buenavista, Saltillo COAH, 25315, México. Email address: rcavazos@narro.uaaan.mx

** Postal address: Departamento de Matemáticas, Universidad Autónoma Metropolitana, Campus Iztapalapa, Avenida San Rafael Atlixco #186, Colonia Vicentina, México 09340, D.F. México. Email address: momr@xanum.uam.mx

This work was supported by the PSF Organization under grant no. 008/100/03-3.

https://doi.org/10.1239/jap/1134587805 Published online by Cambridge University Press

and Montes-de-Oca (2003) (see also Schweitzer (1971)), which generates an equivalent MDP having a positive lower bound for the probability of observing the coincidence of two successive states; and (b) the use of the Birkhoff contraction coefficient, which was first employed in Bielecki *et al.* (1999) to analyze the stationary version of the value iteration scheme in a risk-sensitive context. A similar problem was recently studied for MDPs with locally compact state spaces in Duncan *et al.* (2001), where, under strong mixing conditions, nearly optimal adaptive policies were constructed using large deviations techniques and discrete maximum likelihood estimates.

The paper is organized as follows. In Section 2 the decision model is briefly described, and in Section 3 the original MDP is transformed and the main result stated as Theorem 3.1. After some technical preliminaries in Section 4, Theorem 3.1 is proved in Section 5.

Throughout, $\mathbb{R}$ and $\mathbb{N}$ as usual denote the sets of real numbers and nonnegative integers, respectively, and, for a topological space $\mathbb{K}$, $\mathcal{B}(\mathbb{K})$ denotes the space of all continuous and bounded real-valued functions defined on $\mathbb{K}$: for each $V \in \mathcal{B}(\mathbb{K})$, $\|V\| := \max_{x\in\mathbb{K}} |V(x)|$. If $G = [G_{x,y}]$ is a real matrix, set $\|G\| := \max_{x,y} |G_{x,y}|$.

2. Decision model

Let $M = \langle S, A, C, P\rangle$ be an MDP, where the state space S is a finite set endowed with the discrete topology; the control set A is a compact metric space and, for each $x \in S$, $A(x) \subset A$ is the measurable and nonempty subset of admissible actions at state x; $C\colon \mathbb{K} \to \mathbb{R}$ is the cost function, with $\mathbb{K} := \{(x, a),\ a \in A(x),\ x \in S\}$; and $P = [p_{x,y}(\cdot)]$ is the controlled transition law. The interpretation of M is as follows. At each time $t \in \mathbb{N}$, the state $X_t = x \in S$ of a dynamical system is observed and an action $A_t = a \in A(x)$ is chosen. A cost $C(x, a)$ is then incurred and, regardless of the previous states and actions, the state of the system at time $t + 1$ will be $X_{t+1} = y \in S$ with probability $p_{x,y}(a)$; this is the Markov property of the decision model.

Assumption 2.1. *Assume that $C \in \mathcal{B}(\mathbb{K})$ and, for each $x, y \in S$, that $p_{x,y}(\cdot) \in \mathcal{B}(A)$.*

Definition 2.1. (*Policy.*) A control policy is a rule for choosing actions, and may depend on both the current state and the record of previous states and actions; the class of all policies is denoted by $\mathcal{P}$. Given the policy π used to drive the system and the initial state $X_0 = x \in S$, the distribution P^π_x of the state–action process $\{(X_t, A_t)\}$ is uniquely determined (see Hernández-Lerma (1989) and Puterman (1994)), and E^π_x stands for the corresponding expectation operator. Now define $\mathbb{F} := \prod_{x\in S} A(x)$, the compact set consisting of all functions $f\colon S \to A$. Policy π is stationary if there exists an $f \in \mathbb{F}$ such that, under π, the action $A_t = f(X_t)$ is applied at each time $t \in \mathbb{N}$; the class of stationary policies is naturally identified with $\mathbb{F}$. Furthermore, each sequence $\{f_t\} \subset \mathbb{F}$ corresponds to a policy π that prescribes action $A_t = f_t(X_t)$ at each time $t \in \mathbb{N}$.

Definition 2.2. (*Performance index.*) When the system evolves under $\pi \in \mathcal{P}$ and $x \in S$ is the initial state, the λ-sensitive total cost up to time $n \in \mathbb{N}$ is given by

$$J_n(\pi, x) := \frac{1}{\lambda} \log\biggl(\mathrm{E}^\pi_x\biggl[\exp\biggl(\lambda \sum_{t=0}^{n} C(X_t, A_t)\biggr)\biggr]\biggr),$$

whereas the (long-run) expected λ-sensitive average cost under π, starting at x, is defined by

$$J(\pi, x) := \limsup_{n\to\infty}\biggl(\frac{1}{n+1} J_n(\pi, x)\biggr).$$

https://doi.org/10.1239/jap/1134587805 Published online by Cambridge University Press

This criterion is the performance index of the policy π, starting at x. The optimal λ-sensitive average cost at state x is given by

$$J^*(x) := \inf_{\pi \in \mathcal{P}} J(\pi, x),$$

and a policy $\pi^* \in \mathcal{P}$ is λ-optimal if $J(\pi^*, x) = J^*(x)$ for every $x \in S$.

Assumption 2.2. *For each $x, y \in S$ and $f \in \mathbb{F}$, there exists an $n \equiv n(x, y, f) \in \mathbb{N}$ such that* $\mathrm{P}_x^f[X_n = y] > 0$.

Lemma 2.1. *Under Assumptions 2.1 and 2.2, the following assertions hold.*

(i) *There exist a $g \in \mathbb{R}$ and an $h \colon S \to \mathbb{R}$ satisfying the λ-OE*

$$\mathrm{e}^{\lambda g + \lambda h(x)} = \min_{a \in A}\Bigg[\mathrm{e}^{\lambda C(x,a)} \sum_{y \in S} p_{x,y}(a)\mathrm{e}^{\lambda h(y)}\Bigg], \qquad x \in S. \tag{2.1}$$

(ii) *The real number g in (2.1) is the optimal λ-sensitive average cost at each state, i.e. $J^*(\cdot) = g$, and the function $h(\cdot)$ is uniquely determined modulo an additive constant.*

(iii) *If $f \in \mathbb{F}$ is such that action $f(x)$ is a minimizer of the right-hand side of (2.1) for each $x \in S$, then the stationary policy f is λ-optimal.*

A proof of this result can be found in Cavazos-Cadena and Fernández-Gaucherand (2002), or Borkar and Meyn (2002).

2.1. The problem

Throughout the remainder of the paper, we suppose that the cost function and the transition law are not immediately available to the controller, but, rather, that the decision-maker knows approximations $C_n \colon \mathbb{K} \to \mathbb{R}$ and $P_n = [p^n_{x,y}(\cdot)]$ satisfying the following conditions.

Assumption 2.3. (i) Assume that, for each $n \in \mathbb{N}$, $C_n \in \mathcal{B}(\mathbb{K})$ and $p^n_{x,y}(\cdot) \in \mathcal{B}(A)$, $x, y \in S$.

(ii) Assume that $\|C_n - C\| + \max_{x,y \in S} \|p^n_{x,y}(\cdot) - p_{x,y}(\cdot)\| \to 0$ as $n \to \infty$.

Under Assumptions 2.1–2.3, the main problem we consider in this note is that of how to use, for each $n \in \mathbb{N}$, the data $\{(C_k, P_k),\ k \leq n\}$ to build convergent approximations $(g_n, H_n(\cdot))$ to the unique solution $(g, h(\cdot))$ of the λ-OE satisfying $h(z) = 0$, for a fixed reference point $z \in S$, and to determine stationary policies $\{\psi_n\}$ whose performance indices converge to the optimal one (see Theorem 3.1, below). This problem is approached via an appropriate formulation of the nonstationary value iteration scheme; the results, stated in the following section, extend those obtained in Federgruen and Schweitzer (1981) for MDPs with risk-neutral criteria. The following lemma, established as Theorem A in Cavazos-Cadena and Montes-de-Oca (2003), will be useful.

Lemma 2.2. *Let $\gamma \in \mathbb{R}$ and $H \in \mathcal{B}(S)$ be fixed. Under Assumption 2.1, the following assertions hold.*

(i) *If*

$$\mathrm{e}^{\lambda\gamma + \lambda H(x)} \leq \min_{a \in A}\Bigg[\mathrm{e}^{\lambda C(x,a)} \sum_{y \in S} p_{x,y}(a)\mathrm{e}^{\lambda H(y)}\Bigg], \qquad x \in S,$$

then $\gamma \leq J^*(\cdot)$.

https://doi.org/10.1239/jap/1134587805 Published online by Cambridge University Press

(ii) *Similarly,*

$$e^{\lambda\gamma+\lambda H(x)} \geq \min_{a\in A}\left[e^{\lambda C(x,a)} \sum_{y\in S} p_{x,y}(a) e^{\lambda H(y)}\right], \qquad x \in S,$$

implies that $\gamma \geq J^*(\cdot)$.

(iii) *If* $f \in \mathbb{F}$ *satisfies*

$$e^{\lambda\gamma+\lambda H(x)} \geq e^{\lambda C(x,f(x))} \sum_{y\in S} p_{x,y}(f(x)) e^{\lambda H(y)}, \qquad x \in S,$$

then $\gamma \geq J(f,\cdot)$.

3. Transformed model and main result

Henceforth, Assumptions 2.1–2.3 are implicit. The solution to the problem posed above involves the following extension of Schweitzer's transformation, introduced in Cavazos-Cadena and Montes-de-Oca (2003).

Definition 3.1. Let $M = \langle S, A, C, P\rangle$ be as in Section 2 and let $\alpha \in (0, 1)$ be fixed.

(i) We define $D\colon \mathbb{K} \to \mathbb{R}$ and $Q = [q_{x,y}(\cdot)]$ as follows. For each $x, y \in S$ and $a \in A$,

$$\begin{aligned} D(x,a) &:= \frac{1}{\lambda}\log((1-\alpha)e^{\lambda C(x,a)} + \alpha), \\ q_{x,y}(a) &:= \frac{(1-\alpha)e^{\lambda C(x,a)} p_{x,y}(a) + \alpha\delta_{x,y}}{(1-\alpha)e^{\lambda C(x,a)} + \alpha}, \end{aligned} \tag{3.1}$$

where $\delta_{x,y} = 0$ if $x \neq y$ and $\delta_{x,x} = 1$.

(ii) Let $\{C_n\}$ and $\{P_n\}$ be as in Assumption 2.3. For every $n \in \mathbb{N}$, $a \in A$, and $x, y \in S$, set

$$\begin{aligned} D_n(x,a) &:= \frac{1}{\lambda}\log((1-\alpha)e^{\lambda C_n(x,a)} + \alpha), \\ q^n_{x,y}(a) &:= \frac{(1-\alpha)e^{\lambda C_n(x,a)} p^n_{x,y}(a) + \alpha\delta_{x,y}}{(1-\alpha)e^{\lambda C_n(x,a)} + \alpha}. \end{aligned}$$

(iii) The transformed MDP $\tilde{M}$ is given by $\tilde{M} := \langle S, A, D, Q\rangle$, and $\tilde{M}^n := \langle S, A, D_n, Q^n\rangle$ for each $n \in \mathbb{N}$, where $Q^n := [q^n_{x,y}(\cdot)]$.

For each $(x, a) \in \mathbb{K}$, (3.1) implies that $\{q_{x,y}(a),\ y \in S\}$ is a probability distribution on S and

$$q_{x,x}(a) \geq \frac{\alpha}{(1-\alpha)e^{|\lambda|\,\|C\|} + \alpha} =: \beta > 0; \tag{3.2}$$

thus, in model $\tilde{M}$ the probability of observing the equality of two successive states is bounded away from 0. Also from (3.1), it is not difficult to see that $\tilde{M}$ satisfies Assumptions 2.1 and 2.2, and an application of Lemma 2.1 to $\tilde{M}$ shows that there exists a pair $(\tilde{g}, \tilde{h}(\cdot))$ satisfying

$$e^{\lambda\tilde{g}+\lambda\tilde{h}(x)} = \min_{a\in A}\left[e^{\lambda D(x,a)} \sum_{y\in S} q_{x,y}(a) e^{\lambda\tilde{h}(y)}\right], \qquad x \in S, \tag{3.3}$$

which is the λ-OE associated with $\tilde{M}$. The number $\tilde{g}$ is the optimal λ-sensitive average cost associated with $\tilde{M}$, and $\tilde{h}(\cdot)$ is uniquely determined up to an additive constant. The following

https://doi.org/10.1239/jap/1134587805 Published online by Cambridge University Press

lemma shows that the solutions to the λ-OEs associated with M and $\tilde{M}$ are related in a simple way; for a proof see Cavazos-Cadena and Montes-de-Oca (2003).

Lemma 3.1. (i) Suppose that the pair $(g, h(\cdot))$ satisfies the optimality equation (2.1), and define $\tilde{g}$ by $\tilde{g} := (1/\lambda)\log((1-\alpha)\mathrm{e}^{\lambda g} + \alpha)$. In this case, $(\tilde{g}, h(\cdot))$ is a solution to (3.3).

(ii) If the pair $(\tilde{g}, \tilde{h}(\cdot))$ satisfies (3.3) then $\mathrm{e}^{\lambda\tilde{g}} > \alpha$ and, with $g := (1/\lambda)\log((\mathrm{e}^{\lambda\tilde{g}} - \alpha)/(1-\alpha))$, the pair $(g, \tilde{h}(\cdot))$ satisfies (2.1).

Remark 3.1. (i) Throughout the remainder of the paper, $z \in S$ is a fixed reference point and $(g, h(\cdot))$ stands for the unique solution to (2.1) satisfying $h(z) = 0$. Consequently, with $\tilde{g}$ defined as in Lemma 3.1(i), $(\tilde{g}, h(\cdot))$ is the unique solution to (3.3) for which the functional part vanishes at z (i.e. for which $h(z) = 0$).

(ii) In model $\tilde{M}$, $\tilde{\mathrm{P}}_x^\pi$ denotes the distribution of the state–action process $\{(X_t, A_t)\}$ under the action of policy π when the initial state is $X_0 = x$. From (3.2), it is not difficult to see that

$$\tilde{\mathrm{P}}_y^\pi[X_r = y] \geq \beta^r, \qquad y \in S,\ \pi \in \mathcal{P},\ r \in \mathbb{N}. \tag{3.4}$$

(iii) Notice that

$$\|D_n - D\| + \max_{x,y \in S} \|q_{x,y}^n - q_{x,y}\| \to 0 \quad \text{as } n \to \infty,$$

by Assumption 2.3 and Definition 3.1. In particular, since $D \in \mathcal{B}(\mathbb{K})$ and $\mathbb{K}$ is a compact space, $\sup_{n\in\mathbb{N}} \|D_n\| =: \Delta < \infty$.

The nonstationary value iteration algorithm is now introduced in terms of the models $\tilde{M}^n$.

Definition 3.2. (i) The sequence $\{V_n : S \to \mathbb{R},\ n = -1, 0, 1, 2, \ldots\}$ is recursively determined as follows: $V_{-1} := W$, where the seed $W \in \mathcal{B}(S)$ is arbitrary but fixed, and, for $n \in \mathbb{N}$,

$$V_n(x) := \min_{a\in A}\left[\frac{1}{\lambda}\log\left(\mathrm{e}^{\lambda D_n(x,a)} \sum_{y\in S} q_{x,y}^n(a)\mathrm{e}^{\lambda V_{n-1}(y)}\right)\right], \qquad x \in S. \tag{3.5}$$

(ii) For each $n \in \mathbb{N}$, the nth differential cost function $\tilde{g}_n : S \to \mathbb{R}$ is defined by

$$\tilde{g}_n(x) = V_n(x) - V_{n-1}(x), \qquad x \in S. \tag{3.6}$$

(iii) The nth relative value function $H_n : S \to \mathbb{R}$ is given as follows:

$$H_n(x) = V_n(x) - V_n(z), \qquad x \in S,\ n = -1, 0, 1, 2, 3, \ldots. \tag{3.7}$$

Notice that (3.5) is equivalent to

$$\mathrm{e}^{\lambda V_n(x)} = \min_{a\in A}\left[\mathrm{e}^{\lambda D_n(x,a)} \sum_{y\in S} q_{x,y}^n(a)\mathrm{e}^{\lambda V_{n-1}(y)}\right], \qquad x \in S,\ n \in \mathbb{N}, \tag{3.8}$$

and, by Assumption 2.1 and Definition 3.1, the term within brackets in this equation depends continuously on $a \in A$; since the action space A is compact, for each $n \in \mathbb{N}$ there exists a $\psi_n \in \mathbb{F}$ such that

$$\mathrm{e}^{\lambda V_n(x)} = \mathrm{e}^{\lambda D_n(x,\psi_n(x))} \sum_{y\in S} q_{x,y}^n(\psi_n(x))\mathrm{e}^{\lambda V_{n-1}(y)}, \qquad x \in S. \tag{3.9}$$

https://doi.org/10.1239/jap/1134587805 Published online by Cambridge University Press

Our main result can be now stated as follows (recall Remark 3.1(i)).

Theorem 3.1. *Under Assumptions 2.1–2.3, the following assertions hold.*

(i) $(\tilde{g}_n(z), H_n(\cdot)) \to (\tilde{g}, h(\cdot))$ *as* $n \to \infty$.

(ii) *For each* $n \in \mathbb{N}$, *we have* $\tilde{g}_n(z) > \alpha$. *With*

$$g_n := \frac{1}{\lambda} \log\left(\frac{\mathrm{e}^{\lambda \tilde{g}_n(z)} - \alpha}{1 - \alpha}\right), \tag{3.10}$$

it follows that $\lim_{n\to\infty} g_n = g$.

(iii) *For a given* $n \in \mathbb{N}$, *let* $\psi_n \in \mathbb{F}$ *satisfy (3.9). Then* $J(\psi_n, \cdot) \to g$ *as* $n \to \infty$.

The proof of this theorem will be presented in Section 5.

4. Technical preliminaries

The technical tools that will be used to prove Theorem 3.1 are established in the two lemmas below. The first one concerns a communication property of the transformed model $\tilde{M}$ with respect to arbitrary policies.

Lemma 4.1. *There exist a positive integer* N_0 *and a* $\beta^* > 0$ *such that, for each* $x, y \in S$ *and* $\pi \in \mathcal{P}$,

$$\tilde{\mathrm{P}}_x^{\pi}[X_{N_0} = y] \geq \beta^*.$$

Proof. First, for each $y \in S$ define the hitting time T_y by

$$T_y := \min\{n > 0 \colon X_n = y\}. \tag{4.1}$$

Next, let $x, y \in S$ be arbitrary but fixed, and recall that the transition kernel $Q = [q_{x,y}(\cdot)]$ of model $\tilde{M}$ satisfies the conditions in Assumptions 2.1 and 2.2. By combining the Markov property with Proposition 18 of Royden (1968, p. 232), it is not difficult to see that the mappings $f \mapsto \tilde{\mathrm{P}}_x^f[X_n = y]$ and $f \mapsto \tilde{\mathrm{P}}_x^f[T_y \leq n]$ are continuous in $f \in \mathbb{F}$. Also, given an $f' \in \mathbb{F}$, there exists an $n(x, y, f')$ satisfying

$$\tilde{\mathrm{P}}_x^{f'}[X_{n(x,y,f')} = y] > 0$$

and, by continuity, there exists a neighborhood $\mathcal{N}(f')$ of f' such that

$$f \in \mathcal{N}(f') \Rightarrow \tilde{\mathrm{P}}_x^f[X_{n(x,y,f')} = y] > 0. \tag{4.2}$$

Since $\mathbb{F}$ is compact, there exist policies $f_i' \in \mathbb{F}$, $i = 1, 2, \dots, N_1$, such that $\mathbb{F} = \bigcup_{i=1}^{N_1} \mathcal{N}(f_i')$. Set $n_0(x, y) := \max\{n(x, y, f_i'), i = 1, 2, \dots, N_1\}$ and let $f \in \mathbb{F}$ be arbitrary. In this case, $f \in \mathcal{N}(f_i')$ for some i between 1 and N_1, meaning that (4.1) and (4.2) yield

$$\tilde{\mathrm{P}}_x^f[T_y \leq n_0(x, y)] \geq \tilde{\mathrm{P}}_x^f[X_{n(x,y,f_i')} = y] > 0;$$

since $\mathbb{F}$ is compact and $f \mapsto \tilde{\mathrm{P}}_x^f[T_y \leq n_0(x, y)]$ is a continuous mapping, there exists a $\rho(x, y) > 0$ such that

$$\tilde{\mathrm{P}}_x^f[T_y \leq n_0(x, y)] \geq \rho(x, y) \quad \text{for every } f \in \mathbb{F}.$$

https://doi.org/10.1239/jap/1134587805 Published online by Cambridge University Press

With $n_0 := \max_{x,y\in S} n_0(x,y)$ and $\rho := \min_{x,y\in S} \rho(x,y)$, it follows that $n_0 < \infty$ and $\rho > 0$, by the finiteness of S, and that

$$\tilde{\mathrm{P}}_x^f[T_y \le n_0] \ge \tilde{\mathrm{P}}_x^f[T_y \le n_0(x,y)] \ge \rho(x,y) \ge \rho, \qquad x, y \in S,\ f \in \mathbb{F}.$$

Consequently, for each $f \in \mathbb{F}$, the inequality $\tilde{\mathrm{P}}_x^f[T_y > n_0] \le 1 - \rho$ holds for all states x and y, and an induction argument using the Markov property yields $\tilde{\mathrm{P}}_x^f[T_y > rn_0] \le (1-\rho)^r$ for each $r \in \mathbb{N}$, $f \in \mathbb{F}$, and $x, y \in S$, meaning that

$$\tilde{\mathrm{E}}_x^f[T_y] \le n_0 \sum_{r=0}^{\infty} (1-\rho)^r = \frac{n_0}{\rho}.$$

From this, it follows that there exists a constant B such that $\tilde{\mathrm{E}}_x^\pi[T_y] \le B$ for every policy $\pi \in \mathcal{P}$ and all $x, y \in S$ (see Thomas (1980) and Cavazos-Cadena (1988)). Therefore, if the integer N_0 is larger than $2B$, we have

$$\tilde{\mathrm{P}}_x^\pi[T_y < N_0] \ge \tfrac{1}{2}, \qquad x, y, \in S,\ \pi \in \mathcal{P},$$

by Markov's inequality. Observing that

$$\tilde{\mathrm{P}}_x^\pi[X_{N_0} = y] \ge \sum_{r=1}^{N_0-1} \tilde{\mathrm{P}}_x^\pi[T_y = r,\ X_{N_0} = y],$$

from the above display and (3.4) we find, via the Markov property, that

$$\begin{aligned}\tilde{\mathrm{P}}_x^\pi[X_{N_0} = y] &\ge \sum_{r=1}^{N_0-1} \tilde{\mathrm{P}}_x^\pi[T_y = r]\beta^{N_0-r} \\ &\ge \beta^{N_0} \sum_{r=1}^{N_0-1} \tilde{\mathrm{P}}_x^\pi[T_y = r] \\ &= \beta^{N_0}\tilde{\mathrm{P}}_x^\pi[T_y < N_0] \\ &\ge \tfrac{1}{2}\beta^{N_0} \\ &=: \beta^* > 0.\end{aligned}$$

This completes the proof.

The second preliminary result involves the following contraction coefficient.

Definition 4.1. (i) Given an $m \in \mathbb{N} \setminus \{0\}$, let $\mathcal{P}_m$ consist of all m-dimensional vectors with positive components, and, for each $\mathfrak{x}, \mathfrak{y} \in \mathcal{P}_m$, define the Birkhoff (pseudo)distance between $\mathfrak{x}$ and $\mathfrak{y}$ by

$$d(\mathfrak{x}, \mathfrak{y}) := \log\left(\frac{\max_r[x_r/y_r]}{\min_r[x_r/y_r]}\right),$$

where x_r and y_r, $r = 1, \dots, m$, are the respective components of $\mathfrak{x}$ and $\mathfrak{y}$.

https://doi.org/10.1239/jap/1134587805 Published online by Cambridge University Press

(ii) Let G be a matrix of order $n \times m$, and suppose that all the components of G are positive. The Birkhoff contraction coefficient of G is defined by

$$\tau(G) := \min\{\tau > 0 \colon d(G\mathfrak{x}, G\mathfrak{y}) \le \tau d(\mathfrak{x}, \mathfrak{y}) \text{ for all } \mathfrak{x}, \mathfrak{y} \in \mathcal{P}_m\}. \tag{4.3}$$

Lemma 4.2. *If the matrix $G = [G_{i,j}]$ has positive components, the following* Birkhoff formula *holds:*

$$\tau(G) = \frac{1 - \phi(G)^{1/2}}{1 + \phi(G)^{1/2}}, \quad \textit{where } \phi(G) := \min_{r,i,j,k} \frac{G_{r,j} G_{i,k}}{G_{i,j} G_{r,k}}. \tag{4.4}$$

A proof of this lemma can be found in Seneta (1981, pp. 100–110) or Cavazos-Cadena (2003).

5. Proof of Theorem 3.1

The foregoing preliminaries will be now used to establish Theorem 3.1. The argument relies on the three lemmas stated below, particularly on the assertion (of Lemma 5.2) that $\{\mathrm{sp}(\tilde{g}_n)\}$, where $\mathrm{sp}(\tilde{g}_n)$ is the span seminorm of $\tilde{g}_n$, converges to 0. The span seminorm is given by

$$\mathrm{sp}(W) := \max_{x \in S} W(x) - \min_{x \in S} W(x), \qquad W \in \mathcal{B}(S). \tag{5.1}$$

Before going any further it is convenient to introduce the following notation.

Definition 5.1. For each policy $f \in \mathbb{F}$ and each $n \in \mathbb{N}$, the matrices $B^f = [B^f_{x,y},\ x, y \in S]$ and $B^{n,f} = [B^{n,f}_{x,y},\ x, y \in S]$ are determined as follows:

$$B^f_{x,y} := \mathrm{e}^{\lambda D(x, f(x))} q_{x,y}(f(x)), \quad B^{n,f}_{x,y} := \mathrm{e}^{\lambda D_n(x, f(x))} q^n_{x,y}(f(x)), \qquad x, y \in S.$$

From Assumption 2.3 and Definition 3.1, it follows that

$$\lim_{n \to \infty} \Big(\max_{f \in \mathbb{F}} \| B^f - B^{n,f} \| \Big) = 0,$$

which yields the following convergence, with N_0 as in Lemma 4.1:

$$\max_{f_0, f_1, \dots, f_{N_0 - 1} \in \mathbb{F}} \left\| \prod_{i=0}^{N_0 - 1} B^{n-i, f_i} - \prod_{i=0}^{N_0 - 1} B^{f_i} \right\| \to 0 \quad \text{as } n \to \infty. \tag{5.2}$$

Next observe that, by Definition 5.1 and Remark 3.1(iii), the inequality $B^f_{x,y} \ge \mathrm{e}^{-\lambda\Delta} q_{x,y}(f(x))$ always holds. Via Lemma 4.1, it follows that, for each $f_0, f_1, \dots, f_{N_0-1} \in \mathbb{F}$,

$$\mathrm{e}^{\lambda N_0 \Delta} \ge \left[\prod_{i=0}^{N_0 - 1} B^{f_i} \right]_{x,y} \ge \mathrm{e}^{-\lambda N_0 \Delta} \tilde{\mathrm{P}}^{\pi}_x [X_{N_0} = y] \ge \mathrm{e}^{-\lambda N_0 \Delta} \beta^*, \qquad x, y \in S, \tag{5.3}$$

where policy π is given by $\pi := (f_0, f_1, \dots, f_{N_0-1}, f_{N_0-1}, f_{N_0-1}, \dots)$.

Lemma 5.1. *Let $f_0, f_1, f_2, \dots, f_{N_0} \in \mathbb{F}$ be arbitrary but fixed, and define the matrix G by*

$$G := \begin{bmatrix} \prod_{i=0}^{N_0 - 1} B^{f_i} \\ \prod_{i=1}^{N_0} B^{f_i} \end{bmatrix}. \tag{5.4}$$

https://doi.org/10.1239/jap/1134587805 Published online by Cambridge University Press

Then

$$\tau(G) \le \frac{1-\beta^* \mathrm{e}^{-2\lambda N_0 \Delta}}{1+\beta^* \mathrm{e}^{-2\lambda N_0 \Delta}} =: \tau^* < 1$$

(see (4.3) and (4.4)).

Proof. By (5.3), all the components of G lie between $\beta^* \mathrm{e}^{-\lambda N_0 \Delta}$ and $\mathrm{e}^{\lambda N_0 \Delta}$; hence, the inequalities $G_{r,j}/G_{i,j} \ge \beta^* \mathrm{e}^{-2\lambda N_0 \Delta}$ and $G_{i,k}/G_{r,k} \ge \beta^* \mathrm{e}^{-2\lambda N_0 \Delta}$ always hold, meaning that $\phi(G) \ge (\beta^* \mathrm{e}^{-2\lambda N_0 \Delta})^2$, by (4.4). Using the formula for $\tau(G)$ in (4.4), the result follows from the observation that the mapping $w \mapsto (1-\sqrt{w})/(1+\sqrt{w})$ is decreasing in $w \in [0, 1]$.

Now note that, by (5.2) and (5.3), there exists an integer $N_1 \ge N_0$ such that the components of $\prod_{i=0}^{N_0-1} B^{n-i, f_i}$ are always positive when $n \ge N_1$; set

$$\varepsilon_n := \max_{\substack{x,y \in S \\ f_i \in \mathbb{F},\, i=0,1,\dots,N_0-1}} \left| \log \frac{[\prod_{i=0}^{N_0-1} B^{n-i,f_i}]_{x,y}}{[\prod_{i=0}^{N_0-1} B^{f_i}]_{x,y}} \right| \in (0, \infty), \qquad n \ge N_1, \tag{5.5}$$

and observe that (5.2) and (5.3) together yield

$$\lim_{n \to \infty} \varepsilon_n = 0. \tag{5.6}$$

The following lemma provides the central step in the proof of Theorem 3.1.

Lemma 5.2. *The sequence* $\{\tilde{g}_n\}$ *of differential costs in Definition 3.2(ii) satisfies the following assertions.*

(i) *With the integer* N_1 *and* $\tau^* \in [0, 1)$ *as in (5.5) and Lemma 5.1, respectively, we have*

$$\mathrm{sp}(\tilde{g}_n) \le 2(\varepsilon_n + \varepsilon_{n-1})/\lambda + \tau^* \,\mathrm{sp}(\tilde{g}_{n-N_0}), \qquad n > N_1 \tag{5.7}$$

(see (5.1)).

(ii) *It follows from (i) that* $\lim_{n\to\infty} \mathrm{sp}(\tilde{g}_n) = 0$.

Proof. (i) Identify the set of all mappings $x \mapsto \mathrm{e}^{\lambda V_k(x)}$, $x \in S$, with the components of a column vector $\mathfrak{V}_k$, say (so $\mathfrak{V}_k = [\mathrm{e}^{\lambda V_k(x)}]_{x \in S}$); let $n > N_1$ be arbitrary; and note that (3.8), (3.9), and Definition 5.1 together yield the relations

$$\mathfrak{V}_n = B^{n,\psi_n} \mathfrak{V}_{n-1}, \qquad \mathfrak{V}_{n-1} \le B^{n-1,\psi_n} \mathfrak{V}_{n-2},$$
$$\mathfrak{V}_{n-1} = B^{n-1,\psi_{n-1}} \mathfrak{V}_{n-2}, \qquad \mathfrak{V}_n \le B^{n,\psi_{n-1}} \mathfrak{V}_{n-1},$$

implying that

$$\mathfrak{V}_n = \prod_{i=0}^{N_0-1} B^{n-i,\psi_{n-i}} \mathfrak{V}_{n-N_0}, \qquad \mathfrak{V}_{n-1} \le \prod_{i=0}^{N_0-1} B^{n-1-i,\psi_{n-i}} \mathfrak{V}_{n-N_0-1},$$
$$\mathfrak{V}_{n-1} = \prod_{i=0}^{N_0-1} B^{n-1-i,\psi_{n-1-i}} \mathfrak{V}_{n-1-N_0}, \qquad \mathfrak{V}_n \le \prod_{i=0}^{N_0-1} B^{n-i,\psi_{n-1-i}} \mathfrak{V}_{n-N_0}.$$

https://doi.org/10.1239/jap/1134587805 Published online by Cambridge University Press

Together with (5.5), these lead to

$$\mathfrak{V}_n \geq \mathrm{e}^{-\varepsilon_n} \prod_{i=0}^{N_0-1} B^{\psi_{n-i}} \mathfrak{V}_{n-N_0}, \qquad \mathfrak{V}_{n-1} \leq \mathrm{e}^{\varepsilon_{n-1}} \prod_{i=0}^{N_0-1} B^{\psi_{n-i}} \mathfrak{V}_{n-N_0-1}, \tag{5.8}$$

$$\mathfrak{V}_{n-1} \geq \mathrm{e}^{-\varepsilon_{n-1}} \prod_{i=0}^{N_0-1} B^{\psi_{n-1-i}} \mathfrak{V}_{n-1-N_0}, \qquad \mathfrak{V}_n \leq \mathrm{e}^{\varepsilon_n} \prod_{i=0}^{N_0-1} B^{\psi_{n-1-i}} \mathfrak{V}_{n-N_0}. \tag{5.9}$$

Now set $(f_0, f_1, \dots, f_{N_0}) := (\psi_n, \psi_{n-1}, \dots, \psi_{n-N_0})$ and let G be the matrix in (5.4). In this case $\prod_{i=0}^{N_0-1} B^{\psi_{n-i}} = \prod_{i=0}^{N_0-1} B^{f_i}$ is a submatrix of G and (5.8) implies that, for each $x \in S$,

$$\begin{aligned} \mathrm{e}^{\lambda \tilde{g}_n(x)} &= \mathrm{e}^{\lambda[V_n(x) - V_{n-1}(x)]} \\ &\geq \mathrm{e}^{-\varepsilon_n - \varepsilon_{n-1}} \frac{[\prod_{i=0}^{N_0-1} B^{\psi_{n-i}} \mathfrak{V}_{n-N_0}]_x}{[\prod_{i=0}^{N_0-1} B^{\psi_{n-i}} \mathfrak{V}_{n-N_0-1}]_x} \\ &\geq \mathrm{e}^{-\varepsilon_n - \varepsilon_{n-1}} \min_r \frac{[G\mathfrak{V}_{n-N_0}]_r}{[G\mathfrak{V}_{n-N_0-1}]_r}, \end{aligned} \tag{5.10}$$

where r is the row index of the vector to which it is affixed. Similarly, using the fact that $\prod_{i=0}^{N_0-1} B^{\psi_{n-1-i}} = \prod_{i=1}^{N_0} B^{f_i}$ is a submatrix of G, from (5.9) we find that, for every $y \in S$,

$$\mathrm{e}^{\lambda \tilde{g}_n(y)} \leq \mathrm{e}^{\varepsilon_n + \varepsilon_{n-1}} \frac{[\prod_{i=0}^{N_0-1} B^{\psi_{n-1-i}} \mathfrak{V}_{n-N_0}]_y}{[\prod_{i=0}^{N_0-1} B^{\psi_{n-1-i}} \mathfrak{V}_{n-N_0-1}]_y} \leq \mathrm{e}^{\varepsilon_n + \varepsilon_{n-1}} \max_r \frac{[G\mathfrak{V}_{n-N_0}]_r}{[G\mathfrak{V}_{n-N_0-1}]_r}.$$

Together with (5.10) and Definition 4.1, this leads to

$$\begin{aligned} \lambda[\tilde{g}_n(y) - \tilde{g}_n(x)] &\leq 2(\varepsilon_n + \varepsilon_{n-1}) + \log\biggl(\max_r \frac{[G\mathfrak{V}_{n-N_0}]_r}{[G\mathfrak{V}_{n-N_0-1}]_r}\biggr) - \log\biggl(\min_r \frac{[G\mathfrak{V}_{n-N_0}]_r}{[G\mathfrak{V}_{n-N_0-1}]_r}\biggr) \\ &= 2(\varepsilon_n + \varepsilon_{n-1}) + d(G\mathfrak{V}_{n-N_0}, G\mathfrak{V}_{n-N_0-1}) \\ &\leq 2(\varepsilon_n + \varepsilon_{n-1}) + \tau(G) d(\mathfrak{V}_{n-N_0}, \mathfrak{V}_{n-1-N_0}) \\ &= 2(\varepsilon_n + \varepsilon_{n-1}) + \lambda \tau(G) \operatorname{sp}(V_{n-N_0} - V_{n-1-N_0}), \end{aligned}$$

where the last equality follows from $d(\mathfrak{V}_{n-N_0}, \mathfrak{V}_{n-1-N_0}) = \lambda \operatorname{sp}(V_{n-N_0} - V_{n-1-N_0})$, which in turn follows from Definition 4.1(i) and (5.1). Since $x, y \in S$ and $n > N_1$ are arbitrary, the observation that $\tilde{g}_{n-N_0} = V_{n-N_0} - V_{n-1-N_0}$ now completes the proof of part (i).

(ii) By (5.5) and (5.6), there exists a $b_0 \in (0, \infty)$ such that $2(\varepsilon_n + \varepsilon_{n-1})/\lambda \leq b_0$ for each $n > N_1$. Part (i) then yields $\operatorname{sp}(\tilde{g}_n) \leq b_0 + \tau^* \operatorname{sp}(\tilde{g}_{n-N_0})$, meaning that

$$\operatorname{sp}(\tilde{g}_n) \leq b_0 \sum_{i=0}^{r-1} (\tau^*)^i + (\tau^*)^r \operatorname{sp}(\tilde{g}_{n-rN_0})$$

if $r, n \in \mathbb{N}$ satisfy $n - (r-1)N_0 > N_1$. Given an integer $n > N_1$, let k be the smallest integer such that $n - kN_0 \leq N_1$. Then

$$\operatorname{sp}(\tilde{g}_n) \leq b_0 \sum_{i=0}^{k-1} (\tau^*)^i + (\tau^*)^k \operatorname{sp}(\tilde{g}_{n-kN_0}) \leq \frac{b_0}{1-\tau^*} + b_1 < \infty,$$

https://doi.org/10.1239/jap/1134587805 Published online by Cambridge University Press

where $b_1 := \max_{t\le N_1} \operatorname{sp}(\tilde g_t)$, and, consequently, $\limsup_{n\to\infty} \operatorname{sp}(\tilde g_n)$ is finite. To conclude, note that (5.6) and (5.7) together imply that

$$\limsup_{n\to\infty} \operatorname{sp}(\tilde g_n) \le \tau^* \limsup_{n\to\infty} \operatorname{sp}(\tilde g_n)$$

and, thus, $\limsup_{n\to\infty} \operatorname{sp}(\tilde g_n) = 0$, since $\tau^* < 1$.

Lemma 5.3. *The sequences* $\{\|\tilde g_n(\cdot)\|\}$ *and* $\{\operatorname{sp}(V_n)\}$ *are bounded.*

Proof. By combining (3.6)–(3.8), we find that

$$\exp(\lambda \tilde g_n(x) + V_{n-1}(x)) = \min_{a\in A}\Bigg[\mathrm{e}^{\lambda D_n(x,a)} \sum_{y\in S} q^n_{x,y}(a) \mathrm{e}^{\lambda V_{n-1}(y)}\Bigg], \qquad x \in S,\ n \in \mathbb{N}. \quad (5.11)$$

Let $z \in S$ be the fixed reference point introduced in Definition 3.2, note that

$$\tilde g_n(z) - \operatorname{sp}(\tilde g_n) \le \tilde g_n(x) \le \tilde g_n(z) + \operatorname{sp}(\tilde g_n) \quad \text{for every } x \in S,$$

and observe that (5.11) yields

$$\begin{aligned}\exp(\lambda[\tilde g_n(z) - \operatorname{sp}(\tilde g_n)] + V_{n-1}(x)) &\le \min_{a\in A}\Bigg[\mathrm{e}^{\lambda D_n(x,a)} \sum_{y\in S} q^n_{x,y}(a) \mathrm{e}^{\lambda V_{n-1}(y)}\Bigg] \\ &\le \exp(\lambda[\tilde g_n(z) + \operatorname{sp}(\tilde g_n)] + V_{n-1}(x)).\end{aligned}$$

By applying Lemma 2.2 to model $\tilde M^n$ in Definition 3.1(iii), we find that

$$\tilde g_n(z) - \operatorname{sp}(\tilde g_n) \le J^{n*}(\cdot) \le \tilde g_n(z) + \operatorname{sp}(\tilde g_n),$$

where $J^{n*}(\cdot)$ is the optimal λ-sensitive average cost for model $\tilde M^n$; since $|J^{n*}(\cdot)| \le \|D_n\| \le \Delta < \infty$ (see Remark 3.1(iii)), it follows that $|\tilde g_n(z)| \le \operatorname{sp}(\tilde g_n) + \Delta$, meaning that

$$|\tilde g_n(x)| \le \operatorname{sp}(\tilde g_n) + |\tilde g_n(z)| \le 2\operatorname{sp}(\tilde g_n) + \Delta,$$

which, via Lemma 5.2, yields

$$\sup_{n\in\mathbb{N}} \|\tilde g_n\| =: \tilde b < \infty. \quad (5.12)$$

Observe now that, since S is finite, for each $k \in \mathbb{N}$ the function $V_k(\cdot)$ has a minimizer $x_k^* \in S$:

$$V_k(x_k^*) = \min_{x\in S} V_k(x).$$

Next, let $n \ge N_1$ be arbitrary. From (5.3) and the first inequality in (5.8) (with $n + N_0$ instead of n), it follows that

$$\mathrm{e}^{\lambda V_{n+N_0}(x_n^*)} \ge \mathrm{e}^{-\varepsilon_{n+N_0}} \mathrm{e}^{-N_0\lambda\Delta} \sum_{y\in S} \tilde{\mathrm{P}}^{\pi}_{x_n^*}[X_{N_0} = y] \mathrm{e}^{\lambda V_n(y)},$$

where $\pi = (\psi_n, \psi_{n-1}, \dots, \psi_{N_0}, \psi_{N_0}, \dots)$ and each policy ψ_n is as in (3.9). Lemma 4.1 then yields

$$\mathrm{e}^{\lambda V_{n+N_0}(x_n^*)} \ge \mathrm{e}^{-\varepsilon_{n+N_0}} \mathrm{e}^{-N_0\lambda\Delta} \beta^* \mathrm{e}^{\lambda V_n(w)} \quad \text{for every } w \in S.$$

This implies that

$$V_{n+N_0}(x_n^*) \geq V_n(w) - N_0\Delta - \frac{\varepsilon_{n+N_0}}{\lambda} + \log(\beta^*)$$

and, thus,

$$N_0\Delta + \frac{\varepsilon_{n+N_0}}{\lambda} - \log(\beta^*) + V_{n+N_0}(x_n^*) - V_n(x_n^*) \geq V_n(w) - V_n(x_n^*), \qquad w \in S,\ n \geq N_1. \tag{5.13}$$

To conclude, observe that

$$V_{n+N_0}(x_n^*) - V_n(x_n^*) = \sum_{i=0}^{N_0-1} \tilde{g}_{n+N_0-i}(x_n^*) \leq N_0\tilde{b},$$

by (3.6) and (5.12). Equation 5.13 then yields

$$N_0\Delta + \frac{\varepsilon_{n+N_0}}{\lambda} - \log(\beta^*) + N_0\tilde{b} \geq \operatorname{sp}(V_n) \quad \text{for } n \geq N_1,$$

and from (5.6) it follows that $\{\operatorname{sp}(V_n)\}$ is a bounded sequence.

Proof of Theorem 3.1. Notice that, by (3.6) and (3.7), (3.8) can be equivalently written as

$$\exp(\lambda\tilde{g}_n(z) + \lambda H_n(x)) = \min_{a\in A}\left[e^{\lambda D_n(x,a)} \sum_{y\in S} q_{x,y}^n(a) e^{\lambda H_{n-1}(y)}\right], \qquad x \in S,\ n \in \mathbb{N}. \tag{5.14}$$

Next, observe that $|H_n(x)| = |V_n(x) - V_n(z)| \leq \operatorname{sp}(V_n)$; the sequences $\{\tilde{g}_n(z)\}$ and $\{\|H_n\|\}$ are therefore bounded, by Lemma 5.3. Now let $(\gamma, H(\cdot))$ be an arbitrary limit point of $\{(\tilde{g}_n(z), H_n(\cdot))\}$, so that there exists a sequence $\{n_k\}$ of positive integers satisfying

$$\lim_{k\to\infty} \tilde{g}_{n_k}(z) = \gamma, \qquad \lim_{k\to\infty} H_{n_k}(x) = H(x), \qquad x \in S. \tag{5.15}$$

Observing that $|H_{n_k-1}(x) - H_{n_k}(x)| = |\tilde{g}_{n_k}(x) - \tilde{g}_{n_k}(z)| \leq \operatorname{sp}(\tilde{g}_n)$ by (3.6) and (3.7), we find that

$$\lim_{k\to\infty} H_{n_k-1}(x) = H(x), \qquad x \in S, \tag{5.16}$$

by Lemma 5.2. If we replace n by n_k in (5.14) and take the limit as $k \to \infty$ on both sides of the resulting equation, the finiteness of S, Assumption 2.3, (5.15), and (5.16) together imply that

$$e^{\lambda\gamma + \lambda H(x)} = \min_{a\in A}\left[e^{\lambda D(x,a)} \sum_{y\in S} q_{x,y}(a) e^{\lambda H(y)}\right], \qquad x \in S. \tag{5.17}$$

Since, by (3.7) and (5.15), $H(z) = 0$, it follows that $(\gamma, H(\cdot)) = (\tilde{g}, h(\cdot))$; see Remark 3.1(i). Thus, we have shown that an arbitrary limit point of $(\tilde{g}_n(z), H_n(\cdot))$ coincides with $(\tilde{g}, h(\cdot))$, meaning that $\lim_{n\to\infty}(\tilde{g}_n(z), H_n(\cdot)) = (\tilde{g}, h(\cdot))$, which establishes part (i). Next, using Definition 3.1(ii), (5.14) can be equivalently written as

$$\exp(\lambda\tilde{g}_n(z) + \lambda H_n(x)) = \min_{a\in A}\left[(1-\alpha)e^{\lambda C_n(x,a)} \sum_{y\in S} p_{x,y}^n(a) e^{\lambda H_{n-1}(y)} + \alpha e^{\lambda H_{n-1}(x)}\right], \qquad x \in S, \tag{5.18}$$

whence

$$\exp(\lambda\tilde{g}_n(z) + \lambda H_n(x)) > \alpha e^{\lambda H_{n-1}(x)};$$

https://doi.org/10.1239/jap/1134587805 Published online by Cambridge University Press

by setting $x = z$ in this inequality and recalling that $H_n(z) = H_{n-1}(z) = 0$, we find that $e^{\lambda \tilde{g}_n(z)} > \alpha$, meaning that g_n in (3.10) is well defined and, by part (i) and Lemma 3.1, $\lim_{n\to\infty} g_n = g$. This proves part (ii). To conclude, observe that via Definition 3.1, for each $x \in S$ and $n \in \mathbb{N}$, (3.9) yields

$$e^{\lambda V_n(x)} = (1-\alpha)e^{\lambda C_n(x,\psi_n(x))} \sum_{y\in S} p^n_{x,y}(\psi_n(x))e^{\lambda V_{n-1}(y)} + \alpha e^{\lambda V_{n-1}(x)},$$

which, using (3.6) and (3.7), is equivalent to

$$\frac{\exp(\lambda\tilde{g}_n(z) + \lambda[H_n(x) - H_{n-1}(x)]) - \alpha}{1-\alpha} e^{H_{n-1}(x)} = e^{\lambda C_n(x,\psi_n(x))} \sum_{y\in S} p^n_{x,y}(\psi_n(x))e^{\lambda H_{n-1}(y)}. \tag{5.19}$$

Next, let $x \in S$ be arbitrary but fixed, and notice that part (i) yields

$$\lim_{n\to\infty} \frac{\exp(\lambda\tilde{g}_n(z) + \lambda[H_n(x) - H_{n-1}(x)]) - \alpha}{1-\alpha} = \frac{e^{\lambda\tilde{g}} - \alpha}{1-\alpha} = e^{\lambda g},$$

whereas Assumption 2.3 and the convergence $H_k(\cdot) \to h(\cdot)$ established in part (i) together imply that

$$\lim_{n\to\infty} \frac{e^{\lambda C_n(x,\psi_n(x))} \sum_{y\in S} p^n_{x,y}(\psi_n(x))e^{\lambda H_{n-1}(y)}}{e^{\lambda C(x,\psi_n(x))} \sum_{y\in S} p_{x,y}(\psi_n(x))e^{\lambda H_{n-1}(y)}} = 1.$$

Therefore, given an $\varepsilon > 0$, there exists a positive integer N such that, for each $n > N$,

$$\frac{\exp(\lambda\tilde{g}_n(z) + \lambda[H_n(x) - H_{n-1}(x)]) - \alpha}{1-\alpha} \le e^{\lambda(g+\varepsilon)}$$

and

$$e^{\lambda C_n(x,\psi_n(x))} \sum_{y\in S} p^n_{x,y}(\psi_n(x))e^{\lambda H_{n-1}(y)} \ge e^{-\varepsilon} e^{\lambda C(x,\psi_n(x))} \sum_{y\in S} p_{x,y}(\psi_n(x))e^{\lambda H_{n-1}(y)}.$$

These inequalities and (5.19) together imply that, for each $x \in S$ and $n > N$,

$$\exp(\lambda(g + 2\varepsilon) + \lambda H_{n-1}(x)) \ge e^{\lambda C(x,\psi_n(x))} \sum_{y\in S} p_{x,y}(\psi_n(x))e^{\lambda H_{n-1}(y)}.$$

Then $J(\psi_n, \cdot) \le g + 2\varepsilon$ for $n > N$, by Lemma 2.2(iii). Since $J(\psi_n, \cdot) \ge J^*(\cdot) = g$ always holds, it follows that $g \le J(\psi_n, \cdot) \le g + 2\varepsilon$ if n is large enough, meaning that $J(\psi_n, \cdot) \to g$ as $n \to \infty$.

As mentioned in Section 1, characterizations of the optimal λ-sensitive average cost for MDPs with denumerable or Borel state spaces via the λ-OE have recently been given in Borkar and Meyn (2002) and Di Masi and Stettner (1999), (2000), respectively. Extending Theorem 3.1 to the cases considered in those papers is an interesting problem.

Acknowledgement

The authors are grateful to the reviewer for his careful reading of the original manuscript and constructive criticism.

https://doi.org/10.1239/jap/1134587805 Published online by Cambridge University Press

References

Bielecki, T., Hernández-Hernández, D. and Pliska, S. R. (1999). Risk sensitive control of finite state Markov chains in discrete time, with applications to portfolio management. *Math. Meth. Operat. Res.* **50,** 167–188.

Borkar, V. S. and Meyn, S. P. (2002). Risk-sensitive optimal control for Markov decision processes with monotone cost. *Math. Operat. Res.* **27,** 192–209.

Cavazos-Cadena, R. (1988). Necessary and sufficient conditions for a bounded solution to the optimality equation in average reward Markov decision chains. *Systems Control Lett.* **10,** 71–78.

Cavazos-Cadena, R. (2003). An alternative derivation of Birkhoff's formula for the contraction coefficient of a positive matrix. *Linear Algebra Appl.* **375,** 291–297.

Cavazos-Cadena, R. and Fernández-Gaucherand, E. (2002). Risk-sensitive optimal control in communicating average Markov decision chains. In *Modeling Uncertainty*, eds M. Dror, P. L'Ecuyer and F. Szydarovszky, Kluwer, Boston, MA, pp. 515–553.

Cavazos-Cadena, R. and Montes-de-Oca, R. (2003). The value iteration algorithm in risk-sensitive average Markov decision chains with finite state space. *Math. Operat. Res.* **28,** 752–776.

Di Masi, G. B. and Stettner, L. (1999). Risk-sensitive control of discrete-time Markov processes with infinite horizon. *SIAM J. Control Optimization* **38,** 61–78.

Di Masi, G. B. and Stettner, L. (2000). Infinite horizon risk sensitive control of discrete time Markov processes with small risk. *Systems Control Lett.* **40,** 15–20.

Duncan, T. E., Pasik-Duncan, B. and Stettner, L. (2001). Risk sensitive adaptive control of discrete time Markov processes. *Prob. Math. Statist.* **21,** 493–512.

Federgruen, A. and Schweitzer, P. J. (1981). Nonstationary Markov decision problems with converging parameters. *J. Optimization Theory Appl.* **34,** 207–241.

Hernández-Lerma, O. (1989). *Adaptive Markov Control Processes*. Springer, New York.

Puterman, M. L. (1994). *Markov Decision Processes: Discrete Stochastic Dynamic Programming*. John Wiley, New York.

Royden, H. L. (1968). *Real Analysis*. MacMillan, London.

Schweitzer, P. J. (1971). Iterative solution of the functional equations of undiscounted Markov renewal programming. *J. Math. Anal. Appl.* **34,** 495–501.

Seneta, E. (1981). *Non-Negative Matrices and Markov Chains*, 2nd edn. Springer, New York.

Thomas, L. C. (1980). Connectedness conditions for denumerable state Markov decision processes. In *Recent Advances in Markov Decision Processes*, eds R. Hartley, L. C. Thomas and D. J. White, Academic Press, New York, pp. 181–204.

https://doi.org/10.1239/jap/1134587805 Published online by Cambridge University Press