Math Meth Oper Res (2009) 70:541–566
DOI 10.1007/s00186-008-0277-y

ORIGINAL ARTICLE

Solutions of the average cost optimality equation for finite Markov decision chains: risk-sensitive and risk-neutral criteria

Rolando Cavazos-Cadena

Received: 26 May 2008 / Revised: 8 November 2008 / Published online: 23 December 2008

Abstract This work is concerned with controlled Markov chains with finite state and action spaces. It is assumed that the decision maker has an arbitrary but constant risk sensitivity coefficient, and that the performance of a control policy is measured by the long-run average cost criterion. Within this framework, the existence of solutions of the corresponding risk-sensitive optimality equation for arbitrary cost function is characterized in terms of communication properties of the transition law.

Keywords Closed set · Arrival time · Constant average cost · Strong simultaneous Doeblin condition · Multiplicative optimality equation

Mathematics Subject Classification (2000) 93E20 · 60J05 · 93C55

1 Introduction

This note concerns Markov decision chains with finite state and action spaces. It is assumed that the controller grades a random cost through the utility function with constant risk sensitivity coefficient λ, and the performance of a control policy is measured by the corresponding (long-run) risk-sensitive average cost criterion. In his context, the *main objective* of this note is the following:

Dedicated to Professor Onésimo Hernández-Lerma, on the occasion of his sixtieth birthday.

This work was supported by the PSF Organization under Grant No. 08-04, and in part by CONACYT under Grant 25357.

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

To determine necessary and sufficient conditions on the transition law so that, for arbitrary cost function, the corresponding risk-sensitive average cost optimality equation has a solution.

The results on this problem are stated in Theorems 3.1–3.3 below and can be roughly described as follows:

(i) Given $\lambda \neq 0$, a solution of the corresponding risk-sensitive average cost optimality equation exists for arbitrary cost function if, and only if, a certain *strong form* of the simultaneous Doeblin condition holds, whose precise formulation depends on the sign of λ, and

(ii) Regardless of the sign of $\lambda \neq 0$, the conditions guaranteeing that the risk-sensitive average optimality equation has a solution for each cost function are (substantially) stronger than the corresponding requirements for the risk-neutral case $\lambda = 0$.

The interest on stochastic systems endowed with the risk-sensitive average criterion can be traced back, at least, to the seminal papers by Howard and Matheson (1972), Jacobson (1973) and Jaquette (1973, 1976). Particularly, in Howard and Matheson (1972) finite communicating models were considered and, using the Perron-Frobenius theory of positive matrices (Seneta 1980), the existence of solutions to the (risk-averse) optimality equation was established; a different approach to this problem, based on the risk-sensitive total cost criterion, was presented in Cavazos-Cadena and Fernández-Gaucherand (2002). Recently, there has been an intensive work on stochastic system endowed with the risk-sensitive average criterion; see, for instance, Fleming and McEneany (1995), Di Masi and Stettner (2000, 2007), Jaśkiewicz (2007) and the references there in. On the other hand, it is known that the risk-neutral average cost optimality equation admits a solution under (diverse variants of) the simultaneous Doeblin condition (Thomas 1980; Puterman 1994); however, in Cavazos-Cadena and Fernández-Gaucherand (1999), a simple example was given to show that such a condition is not sufficient to ensure the existence of solutions to the optimality equation in the risk-sensitive case, a fact that provides the motivation to study the problem considered in this work.

The analysis in this note relies heavily on *the discounted approach*, which involves a family $\{T_\alpha \mid \alpha \in (0, 1)\}$ of contractive (discounted) operators whose fixed points $\{V_\alpha\}$ are used to obtain approximate solutions to the average cost optimality equation, a classical idea that has been widely used; see, for instance, Hernández-Lerma (1988), Arapstathis et al. (1993), or Puterman (1994) for the risk-neutral case, and Cavazos-Cadena and Hernández-Hernández (2003), Cavazos-Cadena (2003), or Jaśkiewicz (2007) for the risk-sensitive case.

The organization of the paper is as follows: firstly, in Sect. 2 the decision model is formally described, and the average criteria analyzed in this work, as well as the corresponding optimality equations, are introduced. Next, in Sect. 3 the results on the conditions characterizing the solvability of the average cost optimality equation are stated as Theorems 3.1–3.3, which concern the risk-averse, risk-neutral and risk-seeking criteria, respectively; also, the conditions involved in each case—which are expressed in terms of accessibility properties of closed sets in the sense of Definition 3.1—are briefly discussed and compared. Then, in Sect. 4 the family $\{T_\alpha\}$ of contractive operators involved in the discounted approach is introduced, and the basic results on this method are stated as Lemmas 4.1 and 4.2, which are finally used in Sect. 5 to provide a proof of Theorems 3.1–3.3.

Notation. The set of all nonnegative integers is denoted by $\mathbb{N}$ and, for a finite set $\mathbb{K}$,

$$\#(\mathbb{K}) := \text{number of elements of } \mathbb{K},$$

whereas the space of real valued functions defined on $\mathbb{K}$ is denoted by $\mathcal{B}(\mathbb{K})$; the maximum norm of $C \in \mathcal{B}(\mathbb{K})$ is given by $\|C\| := \max_{x\in\mathbb{K}} |C(x)|$. If A is an event, the corresponding indicator function is denoted by $I[A]$ and, as usual, all relations involving conditional expectations are supposed to hold almost surely with respect to the underlying probability measure.

2 Decision model

Throughout the remainder $\mathcal{M} = (S, A, \{A(x)\}_{x\in S}, C, P)$ is a Markov decision process (MDP), where the state space S and the action set A are finite sets endowed with the discrete topology while, for each $x \in S$, $A(x) \subset A$ is the nonempty set of admissible actions at x; the class $\mathbb{K}$ of admissible pairs is given by $\mathbb{K} = \{(x, a) \mid a \in A(x), x \in S\}$. On the other hand, $C \in B(\mathbb{K})$ is the cost function and $P = [p_{x\,y}(\cdot)]$ is the controlled transition law. This model $\mathcal{M}$ is interpreted as follows: At each time $t \in \mathbb{N}$ the state of a dynamical system is observed, say $X_t = x \in S$, and an action $A_t = a \in A(x)$ is chosen. Then, a cost $C(x, a)$ is incurred and the state at time $t + 1$ will be $X_{t+1} = y \in S$ with probability $p_{x\,y}(a)$, where $\sum_{y\in S} p_{x\,y}(a) = 1$.

Policies. For each $t \in \mathbb{N}$, the space $\mathbb{H}_t$ of possible histories up to time t is given $\mathbb{H}_0 := S$ and $\mathbb{H}_t := \mathbb{K} \times \mathbb{H}_{t-1}$, $t \geq 1$. A generic element of $\mathbb{H}_t$ is denoted by $\mathbf{h}_t = (x_0, a_0, \ldots, x_i, a_i, \ldots, x_t)$, where $a_i \in A(x_i)$. A policy $\pi = \{\pi_t\}$ is a special sequence of stochastic kernels: For each $t \in \mathbb{H}_t$ and $\mathbf{h}_t \in \mathbb{H}_t$, $\pi_t(\cdot|\mathbf{h}_t)$ is a probability measure on A concentrated on $A(x_t)$. Under the action of policy 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; the class of all policies is denoted by $\mathcal{P}$. Given the policy $\pi \in \mathcal{P}$ 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 (Hernández-Lerma 1988; Arapstathis et al. 1993; Puterman 1994); such a distribution is denoted by P_x^π, while E_x^π stands for the corresponding expectation operator. Next, define $\mathbb{F} := \prod_{x\in S} A(x)$, so that $\mathbb{F}$ consists of all functions $f : S \to A$ such that $f(x) \in A(x)$ for each $x \in S$. A policy π is Markovian if there exists a sequence $(f_0, f_1, f_2, \ldots) \in \prod_{t\in\mathbb{N}} \mathbb{F}$ such that $\pi_t(\cdot|\mathbf{h}_t)$ is always concentrated at $f_t(x_t)$, and in this case π and the corresponding sequence $(f_0, f_1, f_2, \ldots)$ are naturally identified; the class of Markovian policies is denoted by $\mathcal{P}_M$. A policy $\pi = (f_0, f_2, f_2 \ldots) \in \mathcal{P}_M$ is stationary if $f = f_t$ for all t, and the class of stationary policies and $\mathbb{F}$ are identified, so that, with this convention, $\mathbb{F} \subset \mathcal{P}_M \subset \mathcal{P}$.

Utility Functions. For each $\lambda \in \mathbb{R}$, the utility function corresponding to the constant risk sensitivity λ is the function $U_\lambda : \mathbb{R} \to \mathbb{R}$ specified as follows: For each $x \in \mathbb{R}$

$$U_\lambda(x) = \begin{cases} \text{sign}(\lambda)e^{\lambda x}, & \text{if } \lambda \neq 0, \\ x, & \text{if } \lambda = 0. \end{cases} \tag{2.1}$$

Given a random cost Y, the certain equivalent of Y with respect to $U_\lambda(\cdot)$ is denoted by $\mathcal{E}[\lambda, Y]$ and is implicitly defined by

$$U_\lambda(\mathcal{E}[\lambda, Y]) = E[U_\lambda(Y)], \tag{2.2}$$

so that a decision maker with risk sensitivity λ—assessing a random cost according to the expectation of $U_\lambda(Y)$— is indifferent between paying the certain equivalent $\mathcal{E}[\lambda, Y]$ for sure, or incurring the random cost Y. Combining the above displays the following expression is obtained:

$$\mathcal{E}[\lambda, Y] = \begin{cases} \frac{1}{\lambda}\log\left(E[e^{\lambda Y}]\right), & \lambda \neq 0, \\ E[Y], & \lambda = 0, \end{cases} \tag{2.3}$$

and from Jensen's inequality it follows that $\mathcal{E}[\lambda, Y] > E[Y] = \mathcal{E}[0, Y]$ if $\lambda > 0$, whereas $\mathcal{E}[\lambda, Y] < E[Y]$ when $\lambda < 0$. A controller assessing a random cost Y via the expectation of $U_\lambda(Y)$ is referred to as risk-averse if $\lambda > 0$, and as risk-seeking when $\lambda < 0$; if $\lambda = 0$ the decision maker is risk-neutral. Notice that for $r, \lambda \in \mathbb{R}$

$$\mathcal{E}[\lambda, Y + r] = \mathcal{E}[\lambda, Y] + r. \tag{2.4}$$

Average Performance Criteria. Assume that the controller has risk sensitivity λ, and given $\pi \in \mathcal{P}$, $x \in S$ and a positive integer n, let $J_{C,n}(\lambda, \pi, x)$ be the certain equivalent of the total cost $\sum_{t=1}^{n-1} C(X_t, A_t)$ incurred before time n when the system is driven by π starting at $X_0 = x$, that is,

$$U_\lambda\left(J_{C,n}(\lambda, \pi, x)\right) = E_x^\pi\left[U_\lambda\left(\sum_{t=0}^{n-1} C(X_t, A_t)\right)\right]; \tag{2.5}$$

more explicitly,

$$J_{C,n}(\lambda, \pi, x) := \begin{cases} \frac{1}{\lambda}\log\left(E_x^\pi\left[e^{\lambda\sum_{t=0}^{n-1} C(X_t, A_t)}\right]\right), & \text{if } \lambda \neq 0, \\ E_x^\pi\left[\sum_{t=0}^{n-1} C(X_t, A_t)\right], & \text{when } \lambda = 0; \end{cases} \tag{2.6}$$

see (2.3). With this notation, the long-run λ-sensitive average cost at state x under policy π is given by

$$J_C(\lambda, \pi, x) := \limsup_{n\to\infty} \frac{1}{n} J_{C,n}(\lambda, \pi, x), \tag{2.7}$$

whereas

$$J_C^*(\lambda, x) := \inf_{\pi \in \mathcal{P}} J_C(\lambda, \pi, x) \tag{2.8}$$

is the optimal λ-sensitive average cost at x; a policy $\pi^* \in \mathcal{P}$ is λ-optimal if $J_C(\lambda, \pi^*, x) = J_C^*(\lambda, x)$ for each $x \in S$.

Optimality Equations. For each $\lambda \in \mathbb{R}$, the optimality equation corresponding to the average criterion in (2.7) and (2.8) is given by

$$U_\lambda(g + h(x)) = \min_{a \in A(x)} \left[\sum_{y \in S} p_{x\,y}(a) U_\lambda(C(x, a) + h(y)) \right], \quad x \in S, \tag{2.9}$$

where g is a real number and $h : S \to \mathbb{R}$ is a given function. Assume now that this equation is satisfied by the pair $(g, h(\cdot)) \in \mathbb{R} \times \mathcal{B}(S)$, and notice that the finiteness of the action set implies that there exists $f^* \in \mathbb{F}$ such that

$$U_\lambda(g + h(x)) = \sum_{y \in S} p_{x\,y}(f^*(x)) U_\lambda(C(x, f^*(x)) + h(y)), \quad x \in S. \tag{2.10}$$

Combining the specification of the utility function $U_\lambda(\cdot)$ with the Markov property, an induction argument using the above displays yields that the following relations hold for every positive integer n, $\pi \in \mathcal{P}$ and $x \in S$:

$$U_\lambda(ng + h(x)) \le E_x^\pi \left[U_\lambda \left(\sum_{t=0}^{n-1} C(X_t, A_t) + h(X_n) \right) \right],$$

and

$$U_\lambda(ng + h(x)) = E_x^{f^*} \left[U_\lambda \left(\sum_{t=0}^{n-1} C(X_t, A_t) + h(X_n) \right) \right].$$

Using that $U_\lambda(\cdot)$ is strictly increasing it follows that

$$U_\lambda(ng + h(x)) \le E_x^\pi \left[U_\lambda \left(\sum_{t=0}^{n-1} C(X_t, A_t) + \|h\| \right) \right] = U_\lambda(J_{C,n}(\lambda, \pi, x) + \|h\|),$$

and

$$U_\lambda(ng + h(x)) \ge E_x^{f^*} \left[U_\lambda \left(\sum_{t=0}^{n-1} C(X_t, A_t) - \|h\| \right) \right] = U_\lambda(J_{C,n}(\lambda, f^*, x) - \|h\|),$$

where (2.2), (2.4) and (2.5) were combined to set the equalities. Therefore,

$$g \leq \frac{1}{n} J_{C,n}(\lambda, \pi, x) + \frac{\|h\| - h(x)}{n} \quad \text{and} \quad g \geq \frac{1}{n} J_{C,n}(\lambda, f^*, x) - \frac{\|h\| - h(x)}{n},$$

relations that together with (2.7) and (2.8) lead to the following verification result (Hernández-Lerma 1988; Puterman 1994; Hernández-Hernández and Marcus 1996).

Lemma 2.1 *Given* $\lambda \in \mathbb{R}$, *suppose that the pair* $(g, h(\cdot)) \in \mathbb{R} \times \mathcal{B}(S)$ *satisfies the* λ*-optimality equation* (2.9). *In this case, assertions* (i)–(iii) *below hold*:

- (i) *The* λ*-optimal average cost function* $J_C^*(\lambda, \cdot)$ *is constant and equal to* g;
- (ii) *The stationary policy* f^* *in* (2.10) *is* λ*-optimal, i. e.,* $J_C(\lambda, f^*, x) = g$ *for each* $x \in S$.

  Moreover,
- (iii) *For each* $x \in S$, $g = \lim_{n\to\infty} \frac{1}{n} J_{C,n}(\lambda, f^*, x)$.

The Problem. In the risk-neutral case $\lambda = 0$, the optimality equation (2.9) becomes

$$g + h(x) = \min_{a\in A(x)} \left[C(x, a) + \sum_{y\in S} p_{x\,y}(a)h(y) \right], \quad x \in S \tag{2.11}$$

(see 2.1), and it is known that if the transition law satisfies (some variant of) the simultaneous Doeblin condition (Thomas 1980), then for each $C \in \mathcal{B}(S)$ there exists a pair $(g_C, h_C(\cdot)) \equiv (g, h(\cdot))$ satisfying the above equality. However, if $\lambda \neq 0$, an explicit example was given in Cavazos-Cadena and Fernández-Gaucherand (1999), showing that the simultaneous Doeblin condition does not ensure the existence of a solution to (2.9), a fact that provides the motivation to analyze the following problem: Given $\lambda \in \mathbb{R}$, determine necessary and sufficient conditions on the transition law so that, for each $C \in \mathcal{B}(\mathbb{K})$, there exists a pair $(g_C, h_C(\cdot)) \equiv (g, h(\cdot)) \in \mathbb{R} \times \mathcal{B}(S)$ satisfying the λ-optimality equation (2.9). The results on this problem are stated in the following section.

3 Solvability of the optimality equations

In this section necessary and sufficient conditions on the transition law will be provided to ensure that the λ-optimality equation has a solution for arbitrary $C \in \mathcal{B}(\mathbb{K})$. For each $\lambda \in \mathbb{R}$, consider the following conditions $\mathbf{C}_i^\lambda$, $i = 1, 2$.

$\mathbf{C}_1^\lambda$: For each $C \in \mathcal{B}(S)$ there exists $g_C \equiv g$ and $h_C(\cdot) \equiv h(\cdot) : S \to \mathbb{R}$ such that the λ-optimality equation (2.9) holds.

$\mathbf{C}_2^\lambda$: For each $C \in \mathcal{B}(\mathbb{K})$, the λ-optimal average cost function $J_C^*(\lambda, \cdot)$ is constant; see (2.8).

These conditions will be related to communication properties of the transition law involving the following ideas of closed sets and first arrival time.

Definition 3.1 (i) Let $f \in \mathbb{F}$ be an arbitrary stationary policy. A nonempty set $K \subset S$ is f-closed if

$$x \in K \quad \text{and} \quad p_{x\,y}(f(x)) > 0 \Rightarrow y \in K.$$

(ii) A nonempty set $W \subset S$ is $\mathcal{M}$-closed if

$$x \in W \;\; \text{and} \;\; p_{x\,y}(a)) > 0 \quad \text{for some} \;\; a \in A(x) \;\; \Rightarrow \;\; y \in W.$$

(iii) If $U \subset S$ is an arbitrary set, the first arrival time to set U is defined by

$$T_U := \min\{n \geq 1 \,|\, X_n \in U\},$$

where the minimum of the empty set is ∞.

The existence of solutions of the λ-optimality equation will be characterized in terms of the accessibility properties of closed sets. To state the result in the risk-averse case consider the following condition:

$\mathbf{C}_3^1$: If $K \neq S$ is f-closed for some $f \in \mathbb{F}$, then there exists $f_K \in \mathbb{F}$ such that

$$P_x^{f_K}[T_K \leq \#(S \backslash K)] = 1, \quad x \in S \backslash K. \tag{3.1}$$

The following result is an extension of Theorem 2.1 in Cavazos-Cadena and Hernández-Hernández (2008), where the case of an uncontrolled chain was analyzed.

Theorem 3.1 *Let $\lambda > 0$ be arbitrary but fixed. In this case conditions $\mathbf{C}_1^\lambda$, $\mathbf{C}_2^\lambda$ and $\mathbf{C}_3^1$ are equivalent.*

According to this theorem, the risk-averse λ-optimality equation has a solution for each $C \in \mathcal{B}(\mathbb{K})$ if, and only if, each proper f-closed set K is visited in at most $\#(S \backslash K)$ steps when the system starting outside K is driven by an appropriate stationary policy (possibly depending on K). This latter condition is substantially stronger than the following requirement which, as stated below, characterizes the solvability of the risk-neutral optimality equation.

$\mathbf{C}_3^0$: If $K \neq S$ is f-closed for some $f \in \mathbb{F}$, then there exists $f_K \in \mathbb{F}$ such that

$$P_x^{f_K}[T_K < \infty] = 1, \quad x \in S \backslash K. \tag{3.2}$$

Theorem 3.2 *Conditions $\mathbf{C}_1^0$, $\mathbf{C}_2^0$ and $\mathbf{C}_3^0$ are equivalent.*

The result on the solvability of the risk-seeking optimality equation is expressed in terms of the following condition involving $\mathcal{M}$-closed sets.

$\mathbf{C}_3^{-1}$: If $W \neq S$ is an $\mathcal{M}$-closed set, then

$$P_x^f[T_W \leq \#(S \backslash W)] = 1, \quad x \in S \backslash W, \quad f \in \mathbb{F}. \tag{3.3}$$

Theorem 3.3 *If $\lambda < 0$, then conditions $\mathbf{C}_1^\lambda$, $\mathbf{C}_2^\lambda$ and $\mathbf{C}_3^{-1}$ are equivalent.*

The above theorems will be proved after establishing the necessary technical tools in the following section. Each one of the conditions $\mathbf{C}_3^i$, $i = -1, 0, 1$ is a variant of the simultaneous Doeblin condition (Thomas 1980), and in the remainder of the section the relations among them are briefly analyzed. The discussion uses the following simple result.

Lemma 3.1 *Let K be a subset of the state space satisfying $\emptyset \neq K \neq S$.*

(i) *Suppose that the following property holds*:

$$\text{For each } x \in S\setminus K \text{ there exists a } \pi_x \in \mathcal{P}_M \text{ such that } P_x^{\pi_x}[T_K < \infty] > 0. \tag{3.4}$$

Under this condition there exists a policy $f \in \mathbb{F}$ such that

$$P_x^f[T_K] < \infty \quad x \in S\setminus K. \tag{3.5}$$

(ii) *Define the set W by*

$$W := \{x \in S\setminus K \mid P_x^\pi[T_K = \infty] = 1 \text{ for all } \pi \in \mathcal{P}_M\}. \tag{3.6}$$

In this case

$$x \in W \quad \text{and} \quad p_{x\,y}(a) > 0 \quad \text{for some } a \in A(x) \Rightarrow y \in W;$$

particularly, if $W \neq \emptyset$ then W is $\mathcal{M}$-closed (see Definition 3.1(ii)).

Proof (i) Suppose that (3.4) holds and, for each positive integer r, set

$$\tilde{K}_r := \{x \in S\setminus K \mid P_x^\pi[T_K = r] > 0 \text{ for some } \pi \in \mathcal{P}_M\}, \tag{3.7}$$

a specification that together with (3.4) yields that

$$S\setminus K = \bigcup_{r\geq 1} \tilde{K}_r. \tag{3.8}$$

Now, define the disjoint sets K_r as follows:

$$K_1 := \tilde{K}_1, \quad K_r := \tilde{K}_r \setminus \bigcup_{s=1}^{r-1} \tilde{K}_s, \quad r = 2, 3, \ldots \tag{3.9}$$

and notice that

$$\bigcup_{s=1}^{r} K_s = \bigcup_{s=1}^{r} \tilde{K}_s \tag{3.10}$$

for each r, so that $S\setminus K = \bigcup_{r\geq 1} K_r$, by (3.8). Since the sets K_i are disjoint and $S\setminus K$ is finite, it follows that there exists a positive integer R such that

$$S\setminus K = \bigcup_{s=1}^{R} K_s. \tag{3.11}$$

Next, define a policy $f \in \mathbb{F}$ as follows:

(a) If $x \in K$, set $f(x) := f^*(x)$, where $f^* \in \mathbb{F}$ is arbitrary but fixed.
(b) If $x \in K_r$ where $1 \leq r \leq R$, first notice that $x \in \tilde{K}_r$, by (3.9), so that there exists a policy

$$\pi_x = (f_{x\,0}, f_{x\,1}, f_{x\,2}, \ldots) \in \mathcal{P}_M \tag{3.12}$$

satisfying

$$P_x^{\pi_x}[T_K = r] > 0; \tag{3.13}$$

see (3.7). With this notation, set

$$f(x) := f_{x\,0}(x), \quad x \in K_r. \tag{3.14}$$

It will be proved that this policy f satisfies (3.5). As a first step in this direction it will be shown, by induction, that for $s = 1, 2, \ldots, R$

$$P_x^f[T_K \leq s] > 0, \quad x \in K_s. \tag{3.15}$$

To achieve this goal, let $x \in K_1$ be arbitrary and notice that (3.13) and (3.14) with $r = 1$ together yield that

$$\sum_{y\in K} p_{x\,y}(f(x)) = \sum_{y\in K} p_{x\,y}(f_{x\,0}(x)) = P_x^{\pi_x}[X_1 \in K] = P_x^{\pi_x}[T_K = 1] > 0$$

so that (3.15) holds for $s = 1$. Suppose now that (3.15) is valid for $s \leq r - 1$, where the positive integer r satisfies $1 < r \leq R$. Let $x \in K_r$ be arbitrary, and let the policy π_x be as in (3.12) and (3.13); setting $\pi = (f_{x\,1}, f_{x\,2}, \ldots) \in \mathcal{P}_M$, the Markov property and Definition 3.1(iii) together yield

$$0 < P_x^{\pi_x}[T_K = r] = \sum_{y\in S\setminus K} p_{x\,y}(f_{x\,0}(x)) P_y^{\pi}[T_K = r - 1],$$

and then there exists $y \in S\setminus K$ such that $p_{x\,y}(f_{x\,0}(x)) > 0$ and $P_y^{\pi}[T_K = r - 1] > 0$. It follows from (3.7) that

$$y \in \tilde{K}_{r-1} \subset K_1 \cup \cdots \cup K_{r-1} \quad \text{and} \quad p_{x\,y}(f(x)) = p_{x\,y}(f_{x\,0}(x)) > 0;$$

see (3.10) and (3.14). From the inclusions in this statement the induction hypothesis yields that $P_y^f[T_K \le r-1] > 0$, and then $P_x^f[T_K \le r] \ge P_x^f[X_1 = y, T_K \le r] = p_{x\,y}(f(x))P_y^f[T_K \le r-1] > 0$; since $x \in K_r$ is arbitrary, this shows that (3.15) holds for $s = r$, completing the induction argument. Notice now that, since S is finite, (3.11) and (3.15) together imply that $\rho := \min_{y\in S\setminus K} P_y^f[T_K \le R] > 0$, and then

$$P_x^f[T_K > R] \le 1-\rho < 1, \quad x \in S\setminus K.$$

On the other hand, for each integer $t \ge 2$ and $x \in S\setminus K$, Definition 3.1(iii) yields that

$$\begin{aligned}
P_x^f[T_K > tR] &= P_x^f[X_i \in S\setminus K,\ i = 1,2,\ldots,R,\ T_K > tR] \\
&= \sum_{y\in S\setminus K} P_x^f[X_i \in S\setminus K,\ 1 \le i < R,\ X_R = y,\ T_K > tR] \\
&= \sum_{y\in S\setminus K} P_x^f[X_i \in S\setminus K,\ 1 \le i < R,\ X_R = y]P_y^f[T_K > (t-1)R] \\
&\le \max_{y\in S\setminus K}\{P_y^f[T_K > (t-1)R]\} \\
&\quad \times \sum_{y\in S\setminus K} P_x^f[X_i \in S\setminus K,\ 1 \le i < R,\ X_R = y] \\
&= \max_{y\in S\setminus K}\{P_y^f[T_K > (t-1)R]\}P_x^f[T_K > R],
\end{aligned}$$

where the Markov property was used to set the third equality. Combining these two last displays it follows that

$$\max_{x\in S\setminus K} P_x^f[T_K > tR] \le (1-\rho)^t, \quad t = 1,2,\ldots,$$

a relation that immediately leads to (3.5).
(ii) Let $x \in W$ and $a \in A(x)$ be arbitrary and select $f^* \in \mathbb{F}$ satisfying $f^*(x) = a$. Given a policy $\pi = (f_0, f_2, \ldots) \in \mathcal{P}_M$, let π^* be the Markovian policy that at time $t = 0$ chooses actions according to f^*, whereas from time 1 onwards π^* selects actions using π as if the process had started again: formally, $\pi^* = (f_0^*, f_1^*, \ldots) \in \mathcal{P}_M$ is given by $f_0^* = f^*$, and $f_t^* = f_{t-1}$ for $t \ge 1$. Since $x \in W$, (3.6) and the Markov property together yield that

$$1 = P_x^{\pi^*}[T_K = \infty] = \sum_{y\in S\setminus K} p_{x\,y}(f^*(x))P_y^{\pi}[T_K = \infty] = \sum_{y\in S\setminus K} p_{x\,y}(a)P_y^{\pi}[T_K = \infty].$$

Consequently, $P_y^{\pi}[T_K = \infty] = 1$ if $p_{x\,y}(a) > 0$, and since $\pi \in \mathcal{P}_M$ is arbitrary, from (3.6) it follows that $x \in W$ and $p_{x\,y}(a) > 0 \Rightarrow y \in W$. □

The relations among the conditions $\mathbf{C}_3^i$, $i = 1, 0, -1$ are discussed in the lemma and the two examples below

Lemma 3.2 (i) $\mathbf{C}_3^1 \Rightarrow \mathbf{C}_3^0$;
(ii) $\mathbf{C}_3^{-1} \Rightarrow \mathbf{C}_3^0$.

Proof (i) As already noted, this part follows immediately from the statement of conditions $\mathbf{C}_3^1$ and $\mathbf{C}_3^0$; see (3.1) and (3.2).

(ii) Assume that $\mathbf{C}_3^{-1}$ holds. To establish $\mathbf{C}_3^0$, let $\tilde{f} \in \mathbb{F}$ and the $\tilde{f}$-closed set $K \neq S$ be arbitrary. From Definition 3.1 it follows that K is nonempty and $P_z^{\tilde{f}}[X_n \in K] = 1$ for each $z \in K$ and $n \in \mathbb{N}$, so that

$$P_z^{\tilde{f}}[T_K = 1] = 1 = P_z^{\tilde{f}}[T_{S\setminus K} = \infty], \quad z \in K. \tag{3.16}$$

Now, let W be the set in (3.6). It will be shown that

$$W = \emptyset, \tag{3.17}$$

a relation that immediately leads to the desired conclusion. In fact, (3.6) and the above display together imply that condition (3.4) holds, so that there exists a policy $f \in \mathbb{F}$ such that $P_x^f[T_K < \infty] = 1$ for every $x \in S\setminus K$, by Lemma 3.1(i), and then $\mathbf{C}_3^0$ holds. To conclude the argument (3.17) will be established by contradiction. Assume that $W \neq \emptyset$, so that W is an $\mathcal{M}$-closed set, by Lemma 3.1(ii), and notice that the first equality in (3.16) and (3.6) together imply that $K \cap W = \emptyset$, so that

$$T_{S\setminus K} \le T_W,$$

by Definition 3.1(iii). Recalling that $\mathbf{C}_3^{-1}$ is in force, it follows from condition (3.3) that $P_z^{\tilde{f}}[T_W < \infty] = 1$ for each $z \in K \subset S\setminus W$ which together with the above display leads to

$$P_z^{\tilde{f}}[T_{S\setminus K} < \infty] = 1, \quad z \in K,$$

a relation that contradicts (3.16). It follows that W is empty, completing the proof. □

It is not difficult to see that the condition $\mathbf{C}_3^0$ does not imply $\mathbf{C}_3^i$, $i = 1, -1$; see, for instance Example 3.1 in Cavazos-Cadena and Fernández-Gaucherand (1999). On the other hand, in the following examples it is shown that no general implication exists between $\mathbf{C}_3^1$ and $\mathbf{C}_3^{-1}$.

Example 3.1 Let the state space and the action set be given by $S = \{0, 1, 2\}$ and $A = \{0, 1\}$, respectively, set $A(x) = \{0, 1\}$ for $x = 0, 1$, and $A(2) = \{0\}$, and define the transition law $P = [p_{x\,y}(\cdot)]$ as follows:

$$p_{0\,0}(0) = 1 = p_{1\,1}(0) = p_{2\,1}(0)$$

and

$$p_{x\,y}(1) = 1/2, \quad x, y \in \{0, 1\}.$$

In this context it will be shown that $\mathbf{C}_3^{-1}$ is valid but $\mathbf{C}_3^1$ fails.
$\mathbf{C}_3^{-1}$ holds: From the above specifications it follows that there is exactly one proper $\mathcal{M}$-closed subset of S, namely, $W_1 = \{0, 1\}$. Since $S\setminus W_1 = \{2\}$ and $A(2) = \{0\}$, using that $p_{2\,1}(0) = 1$ it follows that (3.3) occurs with W_1 instead of W.
$\mathbf{C}_3^1$ fails: If $f^* \in \mathbb{F}$ is such that $f^*(0) = 0$ it follows that $K = \{0\}$ is f^*-closed, since $p_{0\,0}(0) = 1$. Now, if $f \in \mathbb{F}$ and n is a positive integer, using that $p_{1\,1}(0) = 1$ and $p_{1\,1}(1) = 1/2$, it follows that, for each $n = 1, 2, 3, \ldots$, $P_1^f[T_K \le n] = 0$ if $f(1) = 0$ and $P_1^f[T_K \le n] = 1 - 1/2^n < 1$ if $f(1) = 1$, so that (3.1) fails.

Example 3.2 As before, let the state space and the action set be given by $S = \{0, 1, 2\}$ and $A = \{0, 1\}$, respectively, but now set $A(x) = \{0\}$ for $x = 0, 1$ and $A(2) = A$. Next, define the transition law $P = [p_{x\,y}(\cdot)]$ as follows:

$$p_{0\,1}(0) = 1 = p_{1\,0}(0) = p_{2\,1}(0), \quad \text{and} \quad p_{2\,2}(1) = p_{2\,1}(1) = 1/2$$

As it will be shown below, in this framework $\mathbf{C}_3^1$ is valid but $\mathbf{C}_3^{-1}$ fails.
$\mathbf{C}_3^1$ holds: For each $f^* \in \mathbb{F}$ there is exactly one proper f^*-closed set, namely, $K = \{0, 1\}$. Since $p_{2\,1}(0) = 1$, if $f \in \mathbb{F}$ is such that $f(2) = 0$, it follows that (3.1) holds with $f_K = f$, so that $\mathbf{C}_3^1$ holds.
$\mathbf{C}_3^{-1}$ fails: The set $W = \{0, 1\}$ is $\mathcal{M}$-closed, whereas if $f \in \mathbb{F}$ is such that $f(2) = 1$, using that $p_{2\,2}(2) = 1/2$, it follows that $P_2[T_W \le n] = 1 - 1/2^n < 1$ for each positive integer n, so that (3.3) does not hold.

4 Technical tools

This section contains the auxiliary results that will be used to prove Theorems 3.1–3.3. The necessary preliminaries concern the (discounted) operators on $\mathcal{B}(S)$ introduced below where, hereafter, $C \in \mathcal{B}(S)$ and $\lambda \in \mathbb{R}$ are arbitrary but fixed.

Definition 4.1 Given $\alpha \in (0, 1)$ define the operator $T_\alpha : \mathcal{B}(S) \to \mathcal{B}(S)$ as follows: For each $V \in \mathcal{B}(S)$ and $x \in S$, $T_\alpha[V](x)$ is implicitly determined by

$$U_\lambda(T_\alpha[V](x)) = \min_{a \in A(x)} \sum_{y \in S} p_{x\,y}(a) U_\lambda(C(x, a) + \alpha V(y)), \quad x \in S. \tag{4.1}$$

Notice that $T_\alpha[W](x)$ is the minimum certain equivalent of the random cost $C(X_0, A_0) + \alpha V(X_1)$ that can be achieved when the initial state is $X_0 = x$; with this in mind, via (2.4) it is not difficult to see that

$$T_\alpha[V + r] = T_\alpha[V] + \alpha r, \quad r \in \mathbb{R}, \quad V \in \mathcal{B}(S).$$

Now, let $V, \tilde{V} \in \mathcal{B}(S)$ be arbitrary and notice that, using that $U_\lambda(\cdot)$ is strictly increasing, from (4.1) it follows that T_α is a monotone operator, that is,

$$T_\alpha[\tilde{V}] \ge T_\alpha[V] \quad \text{if } \tilde{V} \ge V, \tag{4.2}$$

and combining these two last displays with the relation $-\|\tilde{V} - V\| + V \le \tilde{V} \le V + \|\tilde{V} - V\|$ it follows that $-\alpha\|\tilde{V} - V\| + T_\alpha[V] \le T_\alpha[\tilde{V}] \le T_\alpha[V] + \alpha\|\tilde{V} - V\|$, that is,

$$\|T_\alpha[\tilde{V}] - T_\alpha[V]\| \le \alpha\|\tilde{V} - V\|, \tag{4.3}$$

so that T_α is a contractive operator. Consequently, since $\mathcal{B}(S)$ endowed with the maximum norm is a Banach space, there exists a unique function $V_\alpha \in \mathcal{B}(S)$ satisfying $T_\alpha[V_\alpha] = V_\alpha$, that is,

$$U_\lambda(V_\alpha(x)) = \min_{a \in A(x)} \sum_{y \in S} p_{x\,y}(a) U_\lambda(C(x, a) + \alpha V_\alpha(y)), \quad x \in S, \tag{4.4}$$

and then, recalling that the action set is finite, from this equation it follows that for each $\alpha \in (0, 1)$ there exists a policy $f_\alpha \in \mathbb{F}$ such that

$$U_\lambda(V_\alpha(x)) = \sum_{y \in S} p_{x\,y}(f_\alpha(x)) U_\lambda(C(x, f_\alpha(x)) + \alpha V_\alpha(y)), \quad x \in S. \tag{4.5}$$

Notice that (4.3) with V_α and 0 instead of $\tilde{V}$ and V, respectively, implies that $\|V_\alpha\| - \|T_\alpha[0]\| \le \|V_\alpha - T_\alpha[0]\| = \|T_\alpha[V_\alpha] - T_\alpha[0]\| \le \alpha\|V_\alpha - 0\| = \alpha\|V_\alpha\|$, and then $(1 - \alpha)\|V_\alpha\| \le \|T_\alpha[0]\|$; since $T_\alpha[0](x) = \min_{a \in A(x)} C(x, a)$ for each $x \in S$, by Definition 4.1, it follows that $\|T_\alpha[0]\| \le \|C\|$, so that

$$(1 - \alpha)\|V_\alpha\| \le \|C\|. \tag{4.6}$$

Next, the fixed point V_α will be used to construct approximate solutions to the optimality equation (2.9).

Definition 4.2 Given $\alpha \in (0, 1)$, let $x_\alpha, x_\alpha^+ \in S$ be points where the function $V_\alpha(\cdot) \in \mathcal{B}(S)$ attains its extreme values:

$$V_\alpha(x_\alpha) = \min_{x \in S} V_\alpha(x), \quad \text{and} \quad V_\alpha(x_\alpha^+) = \max_{x \in S} V_\alpha(x); \tag{4.7}$$

notice that the finiteness of S guarantees the existence of such points. With this notation, $g_\alpha, g_\alpha^+ \in \mathbb{R}$ and $h_\alpha, h_\alpha^+ \in \mathcal{B}(S)$ are specified as follows:

$$g_\alpha := (1 - \alpha) V_\alpha(x_\alpha), \quad g_\alpha^+ := (1 - \alpha) V_\alpha(x_\alpha^+), \tag{4.8}$$

while

$$h_\alpha(x) := V_\alpha(x) - V_\alpha(x_\alpha) \quad \text{and} \quad h_\alpha^+(x) := V_\alpha(x) - V_\alpha(x_\alpha^+), \quad x \in S. \tag{4.9}$$

Combining this Definition with (4.4) and (4.5) direct calculations using (2.1) yield that the following equalities hold for each state x and $\alpha \in (0, 1)$:

$$U_\lambda(g_\alpha + h_\alpha(x)) = \min_{a\in A(x)} \sum_{y\in S} p_{x\,y}(a) U_\lambda(C(x, a) + \alpha h_\alpha(y)); \tag{4.10}$$

$$U_\lambda(g_\alpha^+ + h_\alpha^+(x)) = \min_{a\in A(x)} \sum_{y\in S} p_{x\,y}(a) U_\lambda(C(x, a) + \alpha h_\alpha^+(y)), \tag{4.11}$$

as well as

$$U_\lambda(g_\alpha + h_\alpha(x)) = \sum_{y\in S} p_{x\,y}(f_\alpha(x)) U_\lambda(C(x, f_\alpha(x)) + \alpha h_\alpha(y)) \tag{4.12}$$

and

$$U_\lambda(g_\alpha^+ + h_\alpha^+(x)) = \sum_{y\in S} p_{x\,y}(f_\alpha(x)) U_\lambda(C(x, f_\alpha(x)) + \alpha h_\alpha^+(y)). \tag{4.13}$$

On the other hand, since the sate space S and the class $\mathbb{F}$ of stationary policies are finite sets, from the inclusions $x_\alpha, x_\alpha^+ \in S$ and $f_\alpha \in \mathbb{F}$ it follows that there exists a sequence $\{\alpha_k\} \subset (0, 1)$ and $(x_*, x_*^+, f^*) \in S \times S \times \mathbb{F}$ satisfying

$$\alpha_k \nearrow 1 \ \text{ as } \ k \nearrow \infty, \tag{4.14}$$

as well as

$$f_{\alpha_k} = f^*, \quad x_{\alpha_k} = x_* \quad \text{and} \quad x_{\alpha_k}^+ = x_*^+, \quad k \in \mathbb{N}. \tag{4.15}$$

Also, observe that (4.6)–(4.9) together yield that the following assertions hold for every k:

$$\begin{gathered} g_{\alpha_k}^+, \quad g_{\alpha_k} \in [-\|C\|, \ \|C\|\,], \\ h_{\alpha_k}(x) \in [0, \infty), \quad \text{and} \quad h_{\alpha_k}^+(x) \in (-\infty, \ 0], \quad x \in S. \end{gathered} \tag{4.16}$$

Since $[-\|C\|, \ \|C\|]$, $[0, \infty]$ and $[-\infty, \ 0]$ are compact metric spaces, taking a subsequence of $\{\alpha_k\}$, if necessary, it can be supposed that there exist real numbers g_* and g_*^+ as well as functions $h_*(\cdot)$ and $h_*^+(\cdot)$ defined on S such that the following statements hold:

$$\lim_{k\to\infty} g_{\alpha_k} = g_* \in [-\|C\|, \ \|C\|\,], \quad \lim_{k\to\infty} g_{\alpha_k}^+ = g_*^+ \in [-\|C\|, \ \|C\|\,], \tag{4.17}$$

and

$$\lim_{k\to\infty} h_{\alpha_k}(x) = h_*(x) \in [0, \infty], \quad \lim_{k\to\infty} h_{\alpha_k}^+(x) = h_*^+(x) \in [-\infty, \ 0], \quad x \in S, \tag{4.18}$$

where, via (4.9) and (4.15),

$$h_*(x_*) = 0 = h_*^+(x_*^+). \tag{4.19}$$

Defining the sets $\mathcal{H}_*, \mathcal{H}_*^+ \subset S$ by

$$\mathcal{H}_* := \{x \in S \mid h_*(x) < \infty\} \quad \text{and} \quad \mathcal{H}_*^+ := \{x \in S \mid h_*^+(x) > -\infty\}, \tag{4.20}$$

the basic result of the discounted approach to the average criterion can be stated as follows.

Lemma 4.1 (i) *If $\mathcal{H}_* = S$, then the pair $(g_*, h_*(\cdot))$ in (4.17) and (4.18) satisfies the λ-optimality equation (2.9).*
Similarly,
(ii) *If $\mathcal{H}_*^+ = S$, then $(g_*^+, h_*^+(\cdot))$ is a solution of the λ-optimality equation.*

Proof Suppose that $\mathcal{H}_* = S$, so that the function $h_*(\cdot)$ is finite; see (4.18) and (4.20). In this case, replacing α by α_k in (4.10) and taking the limit as k goes to ∞ in both sides of the resulting equality, the finiteness of the state and action spaces and the convergences in (4.18) together yield that $(g^*, h^*(\cdot))$ satisfies the λ-optimality equation (2.9). This establishes the first part, while the second one can be obtained along similar lines. □

The application of this lemma to the proof of Theorems 3.1–3.3 relies on the following simple result involving the ideas in Definition 3.1.

Lemma 4.2 *Let $\{\alpha_k\} \subset (0, 1)$ be such that (4.14)–(4.18) hold and let $\mathcal{H}_*$ and $\mathcal{H}_*^+$ be the sets in (4.20). In this context, statements* (i) *and* (ii) *below are valid:*

(i) *If $\lambda \ge 0$, then the set $\mathcal{H}_*$ is f^*-closed, where f^* is the stationary policy in (4.15).*
(ii) *If $\lambda < 0$ then $\mathcal{H}_*^+$ is $\mathcal{M}$-closed.*

Proof To begin with, notice that the sets $\mathcal{H}_*$ and $\mathcal{H}_*^+$ are nonempty, by (4.19) and (4.20).

(i) Let $\lambda \ge 0$ be arbitrary, so that

$$U_\lambda(\infty) := \lim_{x\to\infty} U_\lambda(x) = \infty; \tag{4.21}$$

see (2.1). Now let $x, y \in S$ be arbitrary, and notice that replacing α by α_k in (4.12) and using the first equality in (4.15) it follows that

$$U_\lambda(g_{\alpha_k} + h_{\alpha_k}(x)) = \sum_{z\in S} p_{x\,z}(f^*(x)) U_\lambda(C(x, f^*(x)) + \alpha_k h_{\alpha_k}(z)),$$

so that

$$\begin{aligned}U_\lambda(g_{\alpha_k}+h_{\alpha_k}(x)) &= \sum_{z\in S\setminus\{y\}} p_{x\,z}(f^*(x))U_\lambda(C(x,\,f^*(x))+\alpha_k h_{\alpha_k}(z))\\ &\quad +p_{x\,y}(f^*(x))U_\lambda(C(x,\,f^*(x))+\alpha_k h_{\alpha_k}(y))\\ &\geq \sum_{z\in S\setminus\{y\}} p_{x\,z}(f^*(x))U_\lambda(C(x,\,f^*(x)))\\ &\quad +p_{x\,y}(f^*(x))U_\lambda(C(x,\,f^*(x))+\alpha_k h_{\alpha_k}(y)),\end{aligned}$$

where, using that $U_\lambda(\cdot)$ is strictly increasing, the inequality stems from the fact that $h_{\alpha_k}(\cdot)\geq 0$; see (4.16). Thus,

$$\begin{aligned}U_\lambda(g_{\alpha_k}+h_{\alpha_k}(x)) \geq (1-p_{x\,y}(f^*(x))U_\lambda(C(x,\,f^*(x)))\\ +p_{x\,y}(f^*(x))U_\lambda(C(x,\,f^*(x))+\alpha_k h_{\alpha_k}(y))\end{aligned}$$

and taking the limit as k goes to ∞, the first convergences in (4.17) and (4.18) together lead to

$$\begin{aligned}U_\lambda(g_*+h_*(x)) \geq (1-p_{x\,y}(f^*(x))U_\lambda(C(x,\,f^*(x)))\\ +p_{x\,y}(f^*(x))U_\lambda(C(x,\,f^*(x))+h_*(y)).\end{aligned}$$

Suppose now that $x\in\mathcal{H}_*$. In this case $h_*(x)$ is finite and the above inequality yields that $p_{x\,y}(f^*(x))U_\lambda(C(x,\,f^*(x))+h_*(y))<\infty$, so that, via (4.21), $p_{x\,y}(f^*(x))>0$ implies that $h_*(y)<\infty$, that is, $y\in\mathcal{H}_*$, showing that $\mathcal{H}_*$ is f^*-closed; see (4.20) and Definition 3.1(i).

(ii) Assume that $\lambda<0$. In this case it follows from (2.1) that

$$U_\lambda(-\infty) := \lim_{x\to-\infty} U_\lambda(x) = -\infty. \tag{4.22}$$

Next, let $x,y\in S$ and $a\in A(x)$ be arbitrary, and notice that (4.11) with α_k instead of α yields that

$$\begin{aligned}U_\lambda(g^+_{\alpha_k}+h^+_{\alpha_k}(x)) &\leq \sum_{z\in S} p_{x\,z}(a)U_\lambda(C(x,a)+\alpha_k h^+_{\alpha_k}(z))\\ &\leq p_{x\,y}(a)U_\lambda(C(x,a)+\alpha_k h^+_{\alpha_k}(y)) + \sum_{z\in S\setminus\{y\}} p_{x\,z}(a)U_\lambda(C(x,a))\\ &= p_{x\,y}(a)U_\lambda(C(x,a)+\alpha_k h^+_{\alpha_k}(y))+(1-p_{x\,y}(a))U_\lambda(C(x,a)),\end{aligned}$$

where the relation $h^+_{\alpha_k}(\cdot)\leq 0$ was used to set the inequality. From this point, the second convergences in (4.17) and (4.18) yield that

$$U_\lambda(g^+_*+h^+_*(x)) \leq (1-p_{x\,y}(a))U_\lambda(C(x,a)) + p_{x\,y}(a)U_\lambda(C(x,a)+h^+_*(y))$$

so that

$$\text{If}\quad p_{x\,y}(a) > 0,\ \text{then}$$
$$U_\lambda(C(x,a) + h_*^+(y)) = -\infty \quad\Rightarrow\quad U_\lambda(g_*^+ + h_*^+(x)) = -\infty;$$

observing that if c is a finite number then $U_\lambda(c + w) = -\infty$ if and only if $w = -\infty$ (see 4.22), from the above display it follows that

$$h_*^+(x)) > -\infty \quad \text{and} \quad p_{x\,y}(a) > 0 \;\Rightarrow\; h_*^+(y)) > -\infty,$$

establishing that $\mathcal{H}_*^+$ is $\mathcal{M}$-closed; see (4.20) and Definition 3.1(ii). □

5 Proof of Theorems 3.1–3.3

In this section the above preliminaries will be used to establish the theorems stated in Sect. 3, which can be summarized in a single statement as follows: For each $\lambda \in \mathbb{R}$

$$\mathbf{C}_1^\lambda \Rightarrow \mathbf{C}_2^\lambda \Rightarrow \mathbf{C}_3^{\mathrm{sign}(\lambda)} \Rightarrow \mathbf{C}_1^\lambda,$$

where $\mathrm{sign}(0) = 0$, $\mathrm{sign}(\lambda) = 1$ if $\lambda > 0$ and $\mathrm{sign}(\lambda) = -1$ if $\lambda < 0$. Since the first implication has been already proved in Lemma 2.1(i), without any additional comment each one of the proofs below consists in establishing that $\mathbf{C}_2^\lambda \Rightarrow \mathbf{C}_3^{\mathrm{sign}(\lambda)}$ and $\mathbf{C}_3^{\mathrm{sign}(\lambda)} \Rightarrow \mathbf{C}_1^\lambda$. However, although the arguments have the same structure, due to the differences among the conditions $\mathbf{C}_3^k$, $k = -1, 0, 1$, the specific details depend heavily on $\mathrm{sign}(\lambda)$.

Proof of Theorem 3.1 Let $\lambda > 0$ be arbitrary but fixed.
$\mathbf{C}_2^\lambda \Rightarrow \mathbf{C}_3^1$: Assume that for each $C \in \mathcal{B}(S)$ the λ-optimal average cost function $J_C^*(\lambda, \cdot)$ is constant. To establish $\mathbf{C}_3^1$ let $K \neq S$ be an f-closed set for some $f \in \mathbb{F}$, so that K is nonempty, by Definition 3.1(i). Define the sequence $\{K_t\}$ of subsets of S as follows:

$$K_0 := K,$$
$$K_{t+1} := \left\{ x \in S \setminus \bigcup_{0 \le m \le t} K_m \;\middle|\; \sum_{y \in K_0 \cup \cdots \cup K_t} p_{x\,y}(a) = 1 \text{ for some } a \in A(x) \right\}, \quad t \in \mathbb{N}. \tag{5.1}$$

Thus, the sets K_t are disjoint, and the finiteness of S yields that there exists $r \in \mathbb{N}$ such that

$$K_{r+1} = \emptyset, \quad \text{and} \quad K_t \neq \emptyset, \quad 0 \le t \le r. \tag{5.2}$$

Setting

$$\tilde{K} = K_0 \cup \cdots \cup K_r, \tag{5.3}$$

it will be shown, by contradiction, that

$$S = \tilde{K}. \tag{5.4}$$

To achieve this goal, *assume* that $S \setminus \tilde{K} \neq \emptyset$. Since K_{r+1} is empty, (5.1) and (5.3) together yield that

$$\sum_{y \in S \setminus \tilde{K}} p_{x\,y}(a) > 0, \quad x \in S \setminus \tilde{K}, \quad a \in A(x),$$

and then the finiteness of S and A implies that

$$\min_{x \in S \setminus \tilde{K},\, a \in A(x)} \sum_{y \in S \setminus \tilde{K}} p_{x\,y}(a) =: \rho > 0. \tag{5.5}$$

Next, it will be shown by induction that, for each positive integer n

$$P_x^\pi[T_{\tilde{K}} > n] \geq \rho^n, \quad x \in S \setminus \tilde{K}, \quad \pi \in \mathcal{P}, \tag{5.6}$$

an assertion that for $n = 1$ is equivalent to (5.5). Assume that this statement holds for $n = m$ and let $x, y \in S \setminus \tilde{K}$, $a \in A(x)$ and $\pi \in \mathcal{P}$ be arbitrary but fixed. Define the new policy $\tilde{\pi} = \{\tilde{\pi}_t\}$ by $\tilde{\pi}_t(\cdot|\mathbf{h}_t) = \pi_{t+1}(\cdot|x, a, \mathbf{h}_t)$ for each t and $\mathbf{h}_t \in \mathbb{H}_t$, and notice that combining Definition 3.1(iii) and (5.5) with the Markov property it follows that

$$\begin{aligned}
&P_x^\pi\left[T_{\tilde{K}} > m+1 \middle| A_0 = a, X_1 = y\right] \\
&\quad = E_x^\pi\left[I[X_t \notin \tilde{K},\ 1 \leq t \leq m+1] \middle| A_0 = a, X_1 = y\right] \\
&\quad = I[X_1 = y] E_y^{\tilde{\pi}}\left[I[X_t \notin \tilde{K},\ 1 \leq t \leq m]\right] \\
&\quad = I[X_1 = y] P_y^{\tilde{\pi}}\left[T_{\tilde{K}} \geq m\right] \\
&\quad \geq I[X_1 = y]\rho^m,
\end{aligned}$$

where the inequality is due to the induction hypothesis. Therefore, recalling that x, y belong to $S \setminus \tilde{K}$, via (5.5) it follows that

$$P_x^\pi\left[T_{\tilde{K}} > m+1\right] \geq \rho^m \sum_{y \in S \setminus \tilde{K}} p_{x\,y}(a) \geq \rho^{m+1},$$

so that (5.6) also holds for $n = m + 1$, completing the induction argument. Next, define $\tilde{C} \in \mathcal{B}(S) (\subset \mathcal{B}(\mathbb{K}))$ as follows:

$$\tilde{C}(x) := 1 - \log(\rho)/\lambda, \quad x \in S \setminus \tilde{K}, \quad \tilde{C}(x) := 0, \quad x \in \tilde{K}, \tag{5.7}$$

and observe that $\tilde{C} \geq 0$, so that $J^*_{\tilde{C}}(\lambda, \cdot) \geq 0$. Since $K = K_0 \subset \tilde{K}$ and K is f-closed, from (2.6) and (2.7) the above specification of $\tilde{C}$ yields that $J_{\tilde{C}}(\lambda, f, x) = 0$ if x

belongs to K, so that

$$J^*_{\tilde{C}}(\lambda, x) = 0, \quad x \in K, \tag{5.8}$$

while if $x \in S \setminus \tilde{K}$ and $\pi \in \mathcal{P}$ are arbitrary, (2.6), (5.6) and (5.7) together yield that for each positive integer n

$$\begin{aligned} J_{\tilde{C},n}(\lambda, \pi, x) &= \frac{1}{\lambda} \log\left(E^\pi_x \left[e^{\lambda \sum_{t=0}^{n-1} \tilde{C}(X_t, A_t)}\right]\right) \\ &\geq \frac{1}{\lambda} \log\left(E^\pi_x \left[e^{\lambda \sum_{t=0}^{n-1} \tilde{C}(X_t, A_t)} I[T_{\tilde{K}} > n]\right]\right) \\ &= \frac{1}{\lambda} \log\left(E^\pi_x \left[e^{\lambda n(1-\log(\rho)/\lambda)} I[T_{\tilde{K}} > n]\right]\right) \\ &\geq \frac{1}{\lambda} \log\left(e^{\lambda n(1-\log(\rho)/\lambda)} \rho^n\right) = n, \end{aligned}$$

so that

$$J^*_{\tilde{C}}(\lambda, x) \geq 1, \quad x \in S \setminus \tilde{K};$$

see (2.7) and (2.8). Combining this fact with (5.8) it follows that $J^*_{\tilde{C}}(\lambda, \cdot)$ is not constant, which is a contradiction since $\mathbf{C}^\lambda_2$ is in force. Thus, (5.4) holds and it follows that

$$S = K_0 \cup \cdots \cup K_r, \tag{5.9}$$

where $r \in \mathbb{N}$ is as in (5.2); since $K_0 = K$ is a proper subset of S it follows that r is positive, and recalling that K_t is nonempty for $1 \leq t \leq r$ (see 5.2),

$$r \leq \#(S \setminus K_0) = \#(S \setminus K). \tag{5.10}$$

Now, for each $x \in S$ define the set $A_K(x)$ as follows:

$$\begin{aligned} &A_K(x) := A(x), \quad x \in K_0 = K, \\ &A_K(x) := \left\{ a \in A(x) \mid \sum_{y \in K_0 \cup \cdots \cup K_{t-1}} p_{x\,y}(a) = 1 \right\}, \quad x \in K_t, \quad t = 1, 2, \ldots, r, \end{aligned}$$

and notice that the sets $A_K(x)$ are nonempty, by (5.1) and (5.2). If $f_K \in \prod_{x \in S} A_K(x) \subset \mathbb{F}$ is arbitrary, it follows that

$$P^{f_K}_x[X_1 \in K] = 1, \quad x \in K_1, \qquad P^{f_K}_x[X_1 \in K \cup K_1 \cup \cdots \cup K_{t-1}] = 1, \\ x \in K_t, \quad 1 < t \leq r,$$

relations that combining Definition 3.1(iii) with a simple conditioning argument yield that

$$P_x^{f_K}[T_K \le t] = 1, \quad x \in K_t, \quad t = 1, 2, \ldots, r$$

and, via (5.9) and (5.10), this implies that $P_x^{f_K}[T_K \le \#(S\setminus K)] = 1$ for $x \in S\setminus K$, establishing $\mathbf{C}_3^1$.

$\mathbf{C}_3^1 \Rightarrow \mathbf{C}_1^\lambda$. Suppose that $\mathbf{C}_3^1$ holds and let $C \in \mathcal{B}(\mathbb{K})$ be arbitrary but fixed. If $K \ne S$ is an f-closed set for some $f \in \mathbb{F}$, with the notation in Sect. 4 it will be shown that for each $\alpha \in (0, 1)$

$$h_\alpha(x) \le 2\|C\|N + \max_{y\in K} h_\alpha(y), \quad x \in S, \quad \text{where } N = \#(S\setminus K). \tag{5.11}$$

To verify this assertion let $f_K \in \mathbb{F}$ be as in (3.1) and define the new stationary policy f_K^* as follows: $f_K^*(x) = f_K(x)$ if $x \in S\setminus K$, and $f_K^*(x) = f(x)$ for $x \in K$. It follows that, starting at $x \in S\setminus K$, T_K has the same distribution with respect to $P_x^{f_K}$ and $P_x^{f_K^*}$, so that $P_x^{f_K^*}[T_K \le N] = P_x^{f_K^*}\left[\bigcup_{i=1}^N [X_i \in K]\right] = 1$ and then, since f_K^* and f coincide on K and this set is f-closed, it is not difficult to see that

$$P_x^{f_K^*}[X_N \in K] = 1, \quad x \in S. \tag{5.12}$$

Notice now that (2.1) and (4.10) together yield that the inequality

$$e^{\lambda h_\alpha(x)} \le E_x^{f_K^*}\left[e^{\lambda(C(X_0,A_0)-g_\alpha)} e^{\lambda\alpha h_\alpha(X_1)}\right]$$

always holds, and then $e^{\lambda h_\alpha(x)} \le e^{2\lambda\|C\|} E_x^{f_K^*}\left[e^{\lambda h_\alpha(X_1)}\right]$ for each $x \in S$; see (4.6) and (4.8), and recall that $h_\alpha(\cdot) \ge 0$. From this point it follows that

$$e^{\lambda h_\alpha(x)} \le e^{2\lambda\|C\|N} E_x^{f_K^*}\left[e^{\lambda h_\alpha(X_N)}\right], \quad x \in S,$$

a fact that using (5.12) leads to (5.11). Replacing α by α_k in (5.11) and taking the limit as k goes to ∞, the finiteness of the state space implies that

$$h_*(x) \le 2\|C\|\#(S\setminus K) + \max_{y\in K} h_*(y), \quad x \in S;$$

see (4.18). Also, notice that the above statement is certainly true if $K = S$. Recalling that the set $\mathcal{H}_*$ in (4.20) is f^*-closed, by Lemma 4.2(i), the above relation with $\mathcal{H}_*$ instead of K yields that $h_*(x) < \infty$ for every $x \in S$, that is, $\mathcal{H}_* = S$, a fact that implies that the λ-optimality equation has a solution, by Lemma 4.1(i); since $C \in \mathcal{B}(\mathbb{K})$ was arbitrary in this argument, it follows that $\mathbf{C}_1^\lambda$ holds. □

Proof of Theorem 3.2 $\mathbf{C}_2^0 \Rightarrow \mathbf{C}_3^0$: Assume that the risk-neutral optimal average cost function $J_C^*(0, \cdot)$ is constant for each $C \in \mathcal{B}(\mathbb{K})$ and let $K \neq S$ be an $\tilde{f}$-closed set for some $\tilde{f} \in \mathbb{F}$, so that

$$1 = P_x^{\tilde{f}}[T_K = 1] = P_x^{\tilde{f}}[X_n \in K, n = 1, 2, 3, \ldots], \quad x \in K.$$

Next, let W be the set in (3.6) and notice that the above display yields that K and W are disjoint. Suppose now that W is nonempty. In this case W is $\mathcal{M}$-closed, by Lemma 3.1(ii), so that

$$P_x^{\pi}[X_n \in W, \; n = 1, 2, 3, \ldots] = 1, \quad x \in W, \quad \pi \in \mathcal{P},$$

and defining $\tilde{C} \in \mathcal{B}(S)$ by $\tilde{C}(x) = 1$ for $x \in W$ and $\tilde{C}(x) = 0$ if $x \in S \setminus W$, from (2.6) and the two displays above it follows that, for each positive integer n, the following assertions (a) and (b) hold: (a) $J_{\tilde{C},n}(0, \tilde{f}, x) = 0$ if $x \in K \subset S \setminus W$, and (b) $J_{\tilde{C},n}(0, \pi, x) = 1$ for $x \in W$ and $\pi \in \mathcal{P}$. Thus, $J_{\tilde{C}}^*(0, x) = 1$ if $x \in W$ and $J_{\tilde{C}}^*(0, x) = 0$ for $x \in K$, so that $J_{\tilde{C}}^*(0, \cdot)$ is not constant, which is a contradiction. Therefore, W in (3.6) is the empty set; this yields that condition (3.4) occurs, and then there exists a policy $f \in F$ satisfying $P_x^f[T_K < \infty] = 1$ for each $x \in S \setminus K$, by Lemma 3.1(i), establishing $\mathbf{C}_3^0$.

$\mathbf{C}_3^0 \Rightarrow \mathbf{C}_1^0$: Assume that $\mathbf{C}_3^0$ holds, let $K \neq S$ be an f-closed set and select a policy $f_K \in \mathbb{F}$ such that (3.2) holds. In this case, there exists a positive integer R_K such that $\max_{x \in S \setminus K} P_x^{f_K}[T_K \geq R_K] =: \rho_K < 1$. As in the proof of Lemma 3.1, it follows from the Markov property that $P_x^{f_K}[T_K \geq t R_K] \leq \rho_K^t$ for each $x \in S \setminus K$ and $t = 1, 2, 3, \ldots$, so that

$$E_x^{f_K}[T_K] \leq \frac{R_K}{1 - \rho_K} < \infty, \quad x \in S \setminus K. \tag{5.13}$$

With the notation in Sect. 4 it will be shown that, for each $\alpha \in (0, 1)$,

$$h_\alpha(x) \leq 2\|C\| \frac{R_K}{1 - \rho_K} + \max_{y \in K} h_\alpha(y), \quad x \in S \setminus K. \tag{5.14}$$

To achieve this goal notice that (2.1) and (4.10) with $\lambda = 0$ together imply that the inequality $g_\alpha + h_\alpha(x) \leq C(x, f_K(x)) + \alpha \sum_{y \in S} p_{x\,y}(f_K(x)) h_\alpha(y)$ always holds. Recalling that $h_\alpha(\cdot) \geq 0$ (see 4.7 and 4.9) a glance at (4.6) and (4.8) leads to

$$h_\alpha(x) \leq 2\|C\| + E_x^{f_K}[h_\alpha(X_1)], \quad x \in S. \tag{5.15}$$

It will be proved, by induction, that for each positive integer n

$$h_\alpha(x) \le 2\|C\| \sum_{t=0}^{n-1} P_x^{f_K}[T_K > t] + E_x^{f_K}\left[h_\alpha(X_{T_K}) I[T_K \le n]\right]$$
$$+E_x^{f_K}\left[h_\alpha(X_n) I[T_K > n]\right], \quad x \in S. \tag{5.16}$$

For $n = 1$ this statement is equivalent to (5.15), since T_K is always ≥ 1. Assume that this claim holds for $n = m$ and notice that (5.15) and the Markov property together yield

$$\begin{aligned}
&E_x^{f_K}\left[h_\alpha(X_m) I[T_K > m] | X_1, \ldots, X_m\right] \\
&\quad = I[T_K > m] h_\alpha(X_m) \\
&\quad \le I[T_K > m]\{2\|C\| + E_{X_m}^{f_K}[h_\alpha(X_{m+1})]\} \\
&\quad = 2\|C\| I[T_K > m] + E_x^{f_K}[h_\alpha(X_{m+1}) I[T_K > m] | X_1, \ldots, X_m]
\end{aligned}$$

so that

$$\begin{aligned}
E_x^{f_K}\left[h_\alpha(X_m) I[T_K > m]\right] &\le 2\|C\| P_x^{f_K}][T_K > m] + E_x^{f_K}[h_\alpha(X_{m+1}) I[T_K > m]] \\
&= 2\|C\| P_x^{f_K}][T_K > m] + E_x^{f_K}[h_\alpha(X_{m+1}) I[T_K = m+1]] \\
&\quad + E_x^{f_K}[h_\alpha(X_{m+1}) I[T_K > m+1]]
\end{aligned}$$

and together with the induction hypothesis, this yields that (5.16) is also valid for $n = m + 1$. Taking the limit as n goes to ∞ in (5.16), it follows from (5.13) that for each $x \in S \setminus K$,

$$\begin{aligned}
h_\alpha(x) &\le 2\|C\| \sum_{t=0}^{\infty} P_x^{f_K}[T_K > t] + E_x^{f_K}\left[h_\alpha(X_{T_K})\right] \\
&= 2\|C\| E_x^{f_K}[T_K] + E_x^{f_K}\left[h_\alpha(X_{T_K})\right] \le 2\|C\| \frac{R_K}{1-\rho_K} + E_x^{f_K}\left[h_\alpha(X_{T_K})\right]
\end{aligned}$$

and (5.14) follows, since $X_{T_K} \in K$ on the event $[T_K < \infty]$ and $P_x^{f_K}[T_K < \infty] = 1$. Replacing α by α_k in (5.14) and taking the limit as $k \to \infty$ in both sides of the resulting inequality, it follows that

$$h_*(x) \le 2\|C\| \frac{R_K}{1-\rho_K} + \max_{y \in K} h_*(y), \quad x \in S \setminus K; \tag{5.17}$$

see (4.18). This relation implies that the f^*-closed set $\mathcal{H}_*$ in (4.20) is equal to S. Indeed, if $\mathcal{H}_*$ is a proper subset of S, then setting $K = \mathcal{H}_*$ the right-hand side of the above inequality is finite, whereas the left-hand side is ∞ if $x \in S \setminus \mathcal{H}_*$, so that the condition $S \ne \mathcal{H}_*$ leads to a contradiction. Therefore, $\mathcal{H}_* = S$ and then the 0-optimality equation has a solution, by Lemma 4.1, establishing $\mathbf{C}_1^0$. □

Proof of Theorem 3.3 Let $\lambda < 0$ and $C \in \mathcal{B}(\mathbb{K})$ be arbitrary but fixed.
$\mathbf{C}_2^\lambda \Rightarrow \mathbf{C}_3^{-1}$: To establish $\mathbf{C}_3^{-1}$ let $K \neq S$ be an $\mathcal{M}$-closed set and define the sequence $\{K_t\}$ of disjoint subsets of S as follows:

$$K_0 := K,$$
$$K_{t+1} := \left\{ x \in S \setminus \bigcup_{0 \le m \le t} K_m \;\middle|\; \sum_{y \in K_0 \cup \cdots \cup K_t} p_{x\,y}(a) = 1 \text{ for all } a \in A(x) \right\}, \quad t \in \mathbb{N}; \tag{5.18}$$

since S is finite it follows that there exists $r \in \mathbb{N}$ such that

$$K_{r+1} = \emptyset, \quad \text{and} \quad K_t \neq \emptyset, \quad 0 \le t \le r, \tag{5.19}$$

and it will be shown, by contradiction, that

$$S = \tilde{K}, \tag{5.20}$$

where

$$\tilde{K} = K_0 \cup \cdots \cup K_r; \tag{5.21}$$

using that $K_0 = K$ is $\mathcal{M}$-closed, from (5.18) it is not difficult to see that $\tilde{K}$ is also $\mathcal{M}$-closed. Moreover, the above definition of the sets K_i yields that $P_x^\pi[X_1 \in K_0 \cup \cdots \cup K_{i-1}] = 1$ for every $\pi \in \mathcal{P}$, $x \in K_i$ and $i > 1$, whereas $P_x^\pi[X_1 \in K_0] = 1$ if $x \in K_1$; since $K_0 = K$, these relations immediately yield that, for every policy π,

$$P_x^\pi[T_K \le i] = 1, \quad x \in K_1 \cup \cdots \cup K_i, \quad i = 1, 2 \ldots, r. \tag{5.22}$$

Assume that $S \setminus \tilde{K} \neq \emptyset$ and notice that, since K_{r+1} is empty, (5.18) and (5.21) together yield that for each $x \in S \setminus \tilde{K}$, there exists $a_x \in A(x)$ such that $\sum_{y \in S \setminus \tilde{K}} p_{x\,y}(a_x) > 0$. Let f be a stationary policy such that $f(x) = a_x$ for each $x \in S \setminus \tilde{K}$, and notice that the previous inequality yields that

$$\min_{x \in S \setminus \tilde{K}} P_x^f[T_{\tilde{K}} > 1] = \min_{x \in S \setminus \tilde{K}} \sum_{y \in S \setminus \tilde{K}} p_{x\,y}(a_x) =: \rho > 0, \tag{5.23}$$

and an induction argument leads to

$$P_x^f[T_{\tilde{K}} > n] \ge \rho^n, \quad x \in S \setminus \tilde{K}, \quad n = 1, 2, 3, \ldots; \tag{5.24}$$

see the proof of Theorem 3.1. Next, define $\tilde{C} \in \mathcal{B}(S)$ as follows:

$$\tilde{C}(x) := -1 - \log(\rho)/\lambda, \quad x \in S \setminus \tilde{K}, \quad \tilde{C}(x) := 0, \quad x \in \tilde{K}, \tag{5.25}$$

and observe that $\tilde{C} \leq 0$, so that $J^*_{\tilde{C}}(\lambda, \cdot) \leq 0$. Since $\tilde{K}$ is $\mathcal{M}$-closed, from (2.6) and (2.7) the above specification of $\tilde{C}$ yields that $J_{\tilde{C}}(\lambda, \pi, x) = 0$ for $x \in \tilde{K}$ and $\pi \in \mathcal{P}$, so that

$$J^*_{\tilde{C}}(\lambda, x) = 0, \quad x \in \tilde{K}, \tag{5.26}$$

while if $x \in S\setminus\tilde{K}$ and f is as in (5.24),

$$\begin{aligned} E_x^\pi\left[e^{\lambda\sum_{t=0}^{n-1}\tilde{C}(X_t,A_t)}\right] &\geq E_x^\pi\left[e^{\lambda\sum_{t=0}^{n-1}(-1-\rho/\lambda)}I[T_{\tilde{K}} > n]\right] \\ &= e^{-n\lambda}\rho^{-n}P_x^\pi\left[T_{\tilde{K}} > n\right] \geq e^{-n\lambda} \end{aligned}$$

and then, recalling that λ is negative,

$$J_{\tilde{C},n}(\lambda, f, x) = \frac{1}{\lambda}\log\left(E_x^\pi\left[e^{\lambda\sum_{t=0}^{n-1}\tilde{C}(X_t,A_t)}\right]\right) \leq -n,$$

and it follows that

$$J^*_{\tilde{C}}(\lambda, x) \leq -1, \quad x \in S\setminus\tilde{K};$$

see (2.7) and (2.8). Combining this fact with (5.26) it follows that $J^*_{\tilde{C}}(\lambda, \cdot)$ is not constant, which is a contradiction since $\mathbf{C}_3^\lambda$ is in force. Thus, (5.20) holds and it follows that

$$S = K_0 \cup \cdots \cup K_r,$$

where $r \in \mathbb{N}$ is as in (5.19); since $K_0 = K$ is a proper subset of S it follows that r is positive, and recalling that K_t is nonempty for $1 \leq t \leq r$ (see 5.19),

$$r \leq \#(S\setminus K_0) = \#(S\setminus K).$$

Combining these two last displays with (5.22), it follows that $P_x^\pi[T_K \leq \#(S\setminus K)] = 1$ for each $x \in S\setminus K$ and $\pi \in \mathcal{P}$, so that $\mathbf{C}_3^{-1}$ certainly holds.

$\mathbf{C}_3^{-1} \Rightarrow \mathbf{C}_1^\lambda$: Suppose that $\mathbf{C}_3^{-1}$ is valid and let $C \in \mathcal{B}(\mathbb{K})$ be arbitrary but fixed. Now let K be an $\mathcal{M}$-closed set with $K \neq S$. Given $\alpha \in (0, 1)$, let $f_\alpha \in \mathbb{F}$ be as in (4.13), so that

$$e^{\lambda g_\alpha^+ + \lambda h_\alpha^+(x)} = e^{\lambda C(x,a)}\sum_{y\in S}p_{x\,y}(f_\alpha(x))e^{\lambda\alpha h_\alpha^+(y)}$$

for all $x \in S$; see (2.1). Since $h_*^+(\cdot) \leq 0$ (see 4.7 and 4.9) and λ is negative, using (4.6) and (4.8) the above relation yields that for each $x \in S$,

$$e^{\lambda h_\alpha^+(x)} \leq e^{-2\lambda\|C\|}\sum_{y\in S}p_{x\,y}(f_\alpha(x))e^{\lambda h_\alpha^+(y)} = E_x^{f_\alpha}\left[e^{-2\lambda\|C\| + \lambda h_\alpha^+(X_1)}\right]$$

a fact that immediately leads to

$$e^{\lambda h_\alpha^+(x)} \le e^{-2n\lambda\|C\|} E_x^{f_\alpha}\left[e^{\lambda h_\alpha^+(X_n)}\right], \quad x \in S, \quad n = 1, 2, 3, \ldots. \tag{5.27}$$

On the other hand, since $K \neq S$ is $\mathcal{M}$-closed and $\mathbf{C}_3^{-1}$ is in force, $P_x^{f_\alpha}[T_K \le N] = 1$ if $x \in S\backslash K$, where $N = \#(S\backslash K)$; consequently, $P_x^{f_\alpha}[X_N \in K] = 1$, since once the set K is reached, the system does not leave K. Therefore, the above display with $n = N$ yields that $e^{\lambda h_\alpha^+(x)} \le e^{-2N\lambda\|C\|} e^{\lambda \min_{y\in K} h_\alpha^+(y)}$ for $x \in S\backslash K$, and then the negativity of λ leads to $h_\alpha^+(x) \ge -2\|C\|N + \min_{y\in K} h_\alpha^+(y)$. Replacing α by α_k and taking the limit as k goes to ∞, it follows that

$$h_*^+(x) \ge -2\|C\|\#(S\backslash K) + \min_{y\in K} h_*^+(y), \quad x \in S\backslash K,$$

a relation that implies that the $\mathcal{M}$-closed set $\mathcal{H}_*^+$ in (4.20) coincides with S. Indeed, if $\mathcal{H}_*^+ \neq S$ then, setting $K = \mathcal{H}_*^+$ in the above display, the right-hand side of the inequality is finite, whereas the left-hand side is $-\infty$ when $x \in S\backslash\mathcal{H}_*^*$. This contradiction shows that $\mathcal{H}_*^+ = S$, and then the λ-optimality equation has solution, by Lemma 4.1(ii), showing that $\mathbf{C}_1^\lambda$ holds. □

References

Arapstathis A, Borkar VK, Fernández-Gaucherand E, Gosh MK, Marcus SI (1993) Discrete-time controlled Markov processes with average cost criteria: a survey. SIAM J Control Optim 31:282–334

Cavazos-Cadena R (2003) Solution to the risk-sesnitive average cost optimality equation in a class of markov decision processes with finite state space. Math Methods Oper Res 57:263–285

Cavazos-Cadena R, Fernández-Gaucherand E (1999) Controlled Markov chains with risk-sensitive criteria: average cost, optimality equations and optimal solutions. Math Methods Oper Res 43:121–139

Cavazos-Cadena R, Fernández-Gaucherand E (2002) Risk-sensitive control in communicating average Markov decision chains. In: Dror M, L'Ecuyer P, Szidarovsky F (eds) Modelling uncertainty: an examination of stochastic theory, methods and applications. Kluwer, Boston, pp 525–544

Cavazos-Cadena R, Hernández-Hernández D (2003) Solution to the risk-sensitive average cost optimality equation in communicating Markov decision chains with finite state space: An alternative approach. Math Methods Oper Res 56:473–479

Cavazos-Cadena R, Hernández-Hernández D (2008) Necessary and sufficient conditions for a solution to the risk-sensitive Poisson equation on a finite state space. Syst Control Lett (to appear)

Di Masi GB, Stettner L (2000) Infinite horizon risk sensitive control of discrete time Markov processes with small risk. Syst Control Lett 40:305–321

Di Masi GB, Stettner L (2007) Infinite horizon risk sensitive control of discrete time Markov processes under minorization property. SIAM J Control Optim 46:231–252

Fleming WH, McEneany WM (1995) Risk-sensitive control on an infinite horizon. SIAM J Control Optim 33:1881–1915

Hernández-Hernández D, Marcus SI (1996) Risk-sensitive control of Markov processes in countable state space. Syst Control Lett 29:147–155

Hernández-Lerma O (1988) Adaptive Markov control processes. Springer, New York

Howard AR, Matheson JED (1972) Risk-sensitive Markov decision processes. Manage Sci 18:356–369

Jacobson DH (1973) Optimal stochastic linear systems with exponential performance criteria and their relation to stochastic differential games. IEEE Trans Automat Control 18:124–131

Jaquette SC (1973) Markov decison processes with a new optimality criterion: discrete time. Ann Stat 1:496–505

Jaquette SC (1976) A utility criterion for Markov decision processes. Manage Sci 23:43–49

Jaśkiewicz A (2007) Average optimality for risk sensitive control with general state space. Ann Appl Probab 17:654–675

Puterman ML (1994) Markov decision processes. Wiley, New York

Seneta E (1980) Nonnegative matrices. Springer, New York

Thomas LC (1980) Conectedness conditions for denumerable state Markov decision processes. In: Hartley R, Thomas LC, White DJ (eds) Recent advances in Markov decision processes. Academic Press, New York