Math Meth Oper Res (2015) 81:269–298
DOI 10.1007/s00186-015-0496-y

ORIGINAL ARTICLE

Continuity of the optimal average cost in Markov decision chains with small risk-sensitivity

Selene Chávez-Rodríguez · Rolando Cavazos-Cadena · Hugo Cruz-Suárez

Received: 5 October 2014 / Accepted: 10 February 2015 / Published online: 22 February 2015

Abstract This note concerns discrete-time controlled Markov chains driven by a decision maker with constant risk-sensitivity λ. Assuming that the system evolves on a denumerable state space and is endowed with a bounded cost function, the paper analyzes the continuity of the optimal average cost with respect to the risk-sensitivity parameter, a property that is promptly seen to be valid at each no-null value of λ. Under standard continuity-compactness conditions, it is shown that a general form of the simultaneous Doeblin condition allows to establish the continuity of the optimal average cost at $\lambda = 0$, and explicit examples are given to show that, even if every state is positive recurrent under the action of any stationary policy, the above continuity conclusion can not be ensured under weaker recurrence requirements, as the Lyapunov function condition.

Keywords Risk-sensitive optimality equation · Discounted approach · Uniformly bounded solutions to the optimality equation · Jensen's inequality · Stopping times · Transient states · Continuity at risk-neutrality

This work was supported by PSFO under Grant No. 14-300-01, and by PRODEP under Grant No. 17332-UAAAN-CA-23.

S. Chávez-Rodríguez · H. Cruz-Suárez
Facultad de Ciencias Físico-Matemáticas, Ave. San Claudio y Río Verde, Col. San Manuel CU,
Benemérita Universidad Autónoma de Puebla, 72570 Puebla, PUE, Mexico
e-mail: selenechavez@alumnos.fcfm.buap.mx

H. Cruz-Suárez
e-mail: hcs@fcfm.buap.mx

R. Cavazos-Cadena (✉)
Departamento de Estadística y Cálculo, Universidad Autónoma Agraria Antonio Narro,
Boulevard Antonio Narro 1923, Buenavista, 25315 Saltillo, COAH, Mexico
e-mail: rcavazos@uaaan.mx

Mathematics Subject Classification 93E20 · 60J05 · 93C55

1 Introduction

This work is concerned with discrete-time Markov decision processes (MDPs) evolving on a denumerable state space. Assuming that the cost function is bounded and that the controller has a constant risk-sensitivity coefficient, denoted by λ, the performance of a control policy is measured by the corresponding λ-sensitive average cost criterion. In this context, the continuous dependence of the optimal value function $J^*(\lambda, \cdot)$ on the risk-sensitivity parameter is analyzed. As it is shown in Sect. 2, under the sole condition that the cost function is bounded, the mapping $\lambda \mapsto J^*(\lambda, \cdot)$ is continuous at each no-null value of λ but, for a general transition structure, the continuity at $\lambda = 0$ is not guaranteed, a fact that leads to consider *the main problem* studied in this work:

- To determine conditions on the transition law ensuring that, for every state x, the optimal λ-sensitive average cost $J^*(\lambda, x)$ is a continuous function of λ at zero.

This problem is interesting by itself, but an additional and *strong* motivation for its analysis stems from a fact recently presented in Bäuerle and Rieder (2013). In that paper it was proved that, for strictly positive costs, the average index with respect to the power utility $U(x) = x^p$ coincides with the classical risk-neutral average criterion, which corresponds to the (identity) utility function $V(x) = x$; in this direction, it was shown in Cavazos-Cadena and Hernández-Hernández (2015) that the same occurs for the logarithmic utility, so that *very different utilities render the same optimal average cost*. An explanation for this phenomenon was given in the aforementioned paper, where the main result, established for finite models with positive costs, can be summarized as follows: For a general utility function $U(x)$ assume that $U''(x)/U'(x) \to \lambda_U \in \mathbb{R}$ as $x \to \infty$. In this case, if the mapping $\lambda \mapsto J^*(\lambda, \cdot)$ is continuous at λ_U, then the optimal average cost with respect to the utility function U coincides with the optimal average index $J^*(\lambda_U, \cdot)$, which is associated to a controller with constant risk-sensitivity λ_U. This last result provides a strong motivation to study the problem posed above, since the case $\lambda_U = 0$ is common in applications (Stokey and Lucas 1989). As it will be mentioned later in this section, the analysis of the main problem naturally leads to consider other interesting questions, which are related with the recurrence–communication properties of the transition law.

The study of Markov models endowed with a risk-sensitive average criterion can be traced back, at least, to the seminal paper by Howard and Matheson (1972). Assuming that the controller has constant risk-sensitivity, in that paper finite and communicating models were studied via the Perron–Frobenius theory of positive matrices, and the optimal average cost, as well as an optimal stationary policy, were obtained form a solution to the risk-sensitive average cost optimality equation; this technique has been recently applied in Sladký (2008) to study finite models with general communication structure. A different perspective of analysis uses 'the discounted approach', which is based on contractive operators whose fixed points are used to generate convergent approximations to a solution of the optimality equation. For denumerable models, the discounted technique was combined with game theoretical ideas in Hernández-Hernández and

Marcus (1996), and with the total cost criterion in Cavazos-Cadena and Fernández-Gaucherand (2002). For models with Borel state space, the discounted method is the main instrument in Masi and Stettner (1999, 2000, 2007), or Jaśkiewicz (2007). In all of these papers it is assumed that the controller has constant risk-sensitivity; for a general utility function, risk-sensitive criteria were recently analyzed in Bäuerle and Rieder (2013), and applications of MDPs to financial problems are presented in Bäuerle and Rieder (2011).

Besides standard continuity-compactness conditions, *the main structural requirement* used in the paper is a general form of the simultaneous Doeblin condition, under which a given stationary policy may have several recurrence classes, but in such a case it is possible to travel form one class to another under the action of a different policy; see Assumption 3.1. In this context, *the main result* of the paper, which is stated as Theorem 3.1, establishes that, for every state x, the optimal λ-sensitive average cost $J^*(\lambda, x)$ is a continuous function of the risk-sensitivity parameter at $\lambda = 0$; also, examples are given to show that (i) the continuity result is not valid for a general transition law, and (ii) if Assumption 3.1 is replaced by the *Lyapunov function condition*, which is a a weaker communication-recurrence requirement, then the continuity of the mapping $\lambda \mapsto J^*(\lambda, \cdot)$ can not be ensured. Essentially, the examples illustrating these facts make use of an important difference between the risk-neutral and risk-sensitive average criteria, namely, the class of transient states plays an important role in the determination of the risk-sensitive average index, but does not have any influence in the risk-neutral case (Cavazos-Cadena and Fernández-Gaucherand 1999). Thus, the following is a most interesting question:

- Assume that an MDP satisfies the Lyapunov function condition, and that each state is positive recurrent under the action of every stationary policy. In this context, is it true that the mapping $\lambda \mapsto J^*(\lambda, x)$ is continuous at $\lambda = 0$ for each state x?

This question will be analyzed in Sect. 8, where an explicit example will be constructed to show that the answer is negative.

The approach of the paper is based on the discounted technique, which is used to establish the main technical tool of this work stated as Theorem 4.1. That result ensures that, under Assumption 3.1, there exists a neighborhood of 0, say $\mathcal{V}$, such that if $0 \neq \lambda \in \mathcal{V}$, then the optimality equation characterizing $J^*(\lambda, \cdot)$ has a bounded solution and, moreover, the bound is *uniform* for all no-null values of λ in the neighborhood $\mathcal{V}$.

The organization of the subsequent material is as follows: In Sect. 2 the decision model is briefly described, the λ-sensitive average index is introduced, and it is shown that the optimal value function $J^*(\lambda, \cdot)$ depends continuously on λ at any no-null value of the parameter, but not necessarily at $\lambda = 0$. In Sect. 3 the simultaneous Doeblin assumption used in the paper is formulated, and the main result, establishing the continuity of the optimal average cost function at $\lambda = 0$, is stated as Theorem 3.1; also, an explicit example is given to show that such a conclusion can not be obtained under the Lyapunov function condition. Next, in Sect. 4 the main technical tool of the paper, concerning the existence of uniformly bounded solutions of the λ-sensitive optimality equation, is stated as Theorem 4.1, a result that, via the preliminaries on the discounted approach presented in Sect. 5, is proved in Sect. 6. Finally, Theorem 3.1 is established in Sect. 7, and the exposition concludes in Sect. 8, where a brief discussion of the previous results is presented, and an example is given to show that, even when every state is positive recurrent, the continuity conclusion in Theorem 3.1 can not be guaranteed under the Lyapunov function condition.

Notation The set of all nonnegative integers is denoted by $\mathbb{N}$ and, for a given topological space K, the space $\mathcal{B}(K)$ consists of all functions $C : K \to \mathbb{R}$ which are bounded, that is, satisfy that

$$\|C\| := \sup_{x\in K} |C(x)| < \infty.$$

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. Finally, for $a, b \in \mathbb{R}$, $a \wedge b := \min\{a, b\}$.

2 Decision model

Let $\mathcal{M} = (S, A, \{A(x)\}_{x\in S}, C, P)$ be an MDP, where the state space S is a denumerable set endowed with the discrete topology, the action set A is a metric space and, for each state $x \in S$, $A(x) \subset A$ is the nonempty set of admissible actions at x; the space $\mathbb{K} := \{(x, a) \,|\, a \in A(x), x \in S\}$ is the class of admissible pairs. On the other hand,

$$C \in \mathcal{B}(\mathbb{K})$$

is the cost function and $P = [p_{xy}(\cdot)]$ is the controlled transition law on S given $\mathbb{K}$, that is, for each $(x, a) \in \mathbb{K}$ and $z \in S$, $p_{xz}(a) \geq 0$ and $\sum_{y\in S} p_{xy}(a) = 1$. This model is interpreted as follows: At each time $t \in \mathbb{N}$ the decision maker observes the state of a dynamical system, say $X_t = x \in S$, and chooses an action (control) $A_t = a \in A(x)$. Then, a cost $C(x, a)$ is incurred and, regardless of the previous states and controls, the state of the system at time $t + 1$ will be $X_{t+1} = y \in S$ with probability $p_{xy}(a)$; this is the Markov property of the decision process.

Assumption 2.1 (i) For each $x \in S$, $A(x)$ is a compact subset of A.
(ii) For every $x, y \in S$, the mappings $a \mapsto C(x, a)$ and $a \mapsto p_{xy}(a)$ are continuous in $a \in A(x)$.

Policies The space $\mathbb{H}_t$ of possible histories up to time $t \in \mathbb{N}$ is defined by $\mathbb{H}_0 := S$ and $\mathbb{H}_t := \mathbb{K}^t \times S, t \geq 1$. 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 $t \in \mathbb{N}$ and $\mathbf{h}_t \in \mathbb{H}_t$, $\pi_t(\cdot|\mathbf{h}_t)$ is a probability measure on A concentrated on $A(x_t)$, and for each Borel subset $B \subset A$, the mapping $\mathbf{h}_t \mapsto \pi_t(B|\mathbf{h}_t)$, $\mathbf{h}_t \in \mathbb{H}_t$, is Borel measurable; under a policy π, the control A_t applied at time t belongs to $B \subset A$ with probability $\pi_t(B|\mathbf{h}_t)$. The class of all policies is denoted by $\mathcal{P}$. Given the policy π being used for choosing actions and the initial state $X_0 = x$, the distribution of the state-action process $\{(X_t, A_t)\}$ is uniquely determined (Arapostathis et al. 1993; Hernández-Lerma 1989; Puterman 1994); such a distribution and the corresponding expectation operator are denoted by P_x^π and E_x^π, respectively. Next, define $\mathbb{F} := \prod_{x\in S} A(x)$ and note that $\mathbb{F}$ is a compact metric space, which consists of all functions $f : S \to A$ such that $f(x) \in A(x)$ for each $x \in S$. A policy π is *stationary* if there exists $f \in \mathbb{F}$ such that the probability measure $\pi_t(\cdot|\mathbf{h}_t)$ is always concentrated at $f(x_t)$, and in this case π and f are naturally identified; with this convention, $\mathbb{F} \subset \mathcal{P}$.

Performance criteria Throughout the remainder it is supposed that the decision maker has a constant risk-sensitivity coefficient $\lambda \in \mathbb{R}$, that is, the controller assesses a (bounded) random cost Y using the expectation of $U_\lambda(Y)$, where the utility function U_λ is given as follows: For each $y \in \mathbb{R}$,

$$U_\lambda(y) = \operatorname{sign}(\lambda)e^{\lambda y} \text{ when } \lambda \neq 0, \quad \text{and} \quad U_0(y) = y; \tag{2.1}$$

note that for every $c, x \in \mathbb{R}$,

$$U_\lambda(x + c) = e^{\lambda c}U_\lambda(x), \quad \lambda \neq 0. \tag{2.2}$$

The certainty equivalent of Y corresponding to U_λ is the real number $\mathcal{E}_\lambda[Y]$ satisfying

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

so that the controller is indifferent between paying the certainty equivalent $\mathcal{E}_\lambda[Y]$ for sure, or incurring the random cost Y. It follows from the above relation and (2.1) that

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

and from this expression it is not difficult to see that $\mathcal{E}_\lambda[\cdot]$ is monotone and homogeneous, that is, for (bounded) random variables Y and W,

$$Y \leq W \text{ a.s.} \implies \mathcal{E}_\lambda[Y] \leq \mathcal{E}_\lambda[W] \tag{2.4}$$

and

$$\mathcal{E}_\lambda[Y + a] = \mathcal{E}_\lambda[Y] + a, \quad a \in \mathbb{R}. \tag{2.5}$$

Suppose now that the controller is driving the system using policy $\pi \in \mathcal{P}$ starting at $x \in S$, and let $J_n(\lambda, \pi, x)$ be the certainty equivalent of the total cost $\sum_{t=0}^{n-1} C(X_t, A_t)$ incurred before time n, that is,

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

With this notation, the (superior limit) λ-sensitive average cost at $x \in S$ under policy π is given by

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

whereas

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

is the optimal λ-sensitive average cost function; a policy $\pi^* \in \mathcal{P}$ is λ-optimal if $J(\lambda, \pi^*, x) = J^*(\lambda, x)$ for each $x \in S$. The criterion (2.7) measures the behavior of the policy π at $x \in S$ in terms of the largest limit point of the sequence of average certainty equivalents $\{J_n(\lambda, \pi, x)/n\}$, and a more 'optimistic' point of view uses the smallest limit point of that sequence: When $X_0 = x$, the inferior limit λ-sensitive average criterion $J_-(\lambda, \pi, x)$ corresponding to $\pi \in \mathcal{P}$ is defined by

$$J_-(\lambda, \pi, x) := \liminf_{n\to\infty} \frac{1}{n} J_n(\lambda, \pi, x), \tag{2.9}$$

and the corresponding inferior limit λ-sensitive average optimal value function is given by

$$J_*(\lambda, x) := \inf_{\pi\in\mathcal{P}} J_-(\lambda, \pi, x), \quad x \in S, \tag{2.10}$$

so that

$$J_*(\lambda, \cdot) \le J^*(\lambda, \cdot). \tag{2.11}$$

Under the stability condition in the following section, it will be shown that the optimal value functions $J_*(\lambda, \cdot)$ and $J^*(\lambda, \cdot)$ coincide.

The problem As already mentioned, the objective of this work is to study the continuity of the optimal value function $J^*(\lambda, \cdot)$ with respect to the risk-sensitivity coefficient λ, a property that can be immediately verified at every $\lambda \neq 0$.

Proposition 2.1 *For each $x \in S$, the mapping $\lambda \mapsto J^*(\lambda, x)$ is continuous on $\mathbb{R} \setminus \{0\}$.*

Proof Let $x \in S$, the positive integer n, and the no-null real numbers λ and ν be arbitrary. Now, observe that $\lambda \sum_{t=0}^{n-1} C(X_t, A_t) = \nu \sum_{t=0}^{n-1} C(X_t, A_t) + (\lambda - \nu) \sum_{t=0}^{n-1} C(X_t, A_t) \le \nu \sum_{t=0}^{n-1} C(X_t, A_t) + n|\lambda - \nu| \|C\|$, an inequality that via (2.6) implies that

$$\begin{aligned} \lambda J_n(\lambda, \pi, x) &= \log \left(E_x^\pi \left[e^{\lambda \sum_{t=0}^{n-1} C(X_t, A_t)} \right] \right) \\ &\le \log \left(E_x^\pi \left[e^{\nu \sum_{t=0}^{n-1} C(X_t, A_t)} \right] \right) + n|\lambda - \nu| \, \|C\| \\ &= \nu J_n(\nu, \pi, x) + n|\lambda - \nu| \|C\|, \end{aligned}$$

and then

$$\lambda \frac{J_n(\lambda, \pi, x)}{n} \le \nu \frac{J_n(\nu, \pi, x)}{n} + |\lambda - \nu| \, \|C\|. \tag{2.12}$$

Next, suppose that λ and ν are both positive. Taking the superior limit as n goes to ∞ in both sides of the above inequality, via (2.7) it follows that $\lambda J(\lambda, \pi, x) \le$

$\nu J(\nu, \pi, x) + |\lambda - \nu| \|C\|$, and then, taking the infimum over $\pi \in \mathcal{P}$ in both side of this relation, (2.8) leads to $\lambda J^*(\lambda, x) \le \nu J^*(\nu, x) + |\lambda - \nu| \|C\|$; consequently, since $\lambda, \nu \in (0, \infty)$ are arbitrary,

$$\lambda J^*(\lambda, x) - \nu J^*(\nu, x)| \le |\lambda - \nu| \|C\|. \tag{2.13}$$

Assume now that λ and ν are both negative. In this context, taking the inferior limit as n goes to ∞ in both sides of (2.12), from (2.7) it follows that $\lambda J(\lambda, \pi, x) \le \nu J(\nu, \pi, x) + |\lambda - \nu| \|C\|$, a relation that, after taking the supremum over $\pi \in \mathcal{P}$, leads to $\lambda J^*(\lambda, x) \le \nu J^*(\nu, x) + |\lambda - \nu| \|C\|$, and interchanging the roles of λ and ν it follows that (2.13) is also valid in this case. In short, it has been shown that $\lambda J^*(\lambda, x)$ is a Lipschitz function of λ on each one of the intervals $(0, \infty)$ and $(-\infty, 0)$, so that $\lambda \mapsto J^*(\lambda, x)$ is continuous at each point in $\mathbb{R} \setminus \{0\}$. □

The conclusion of the previous proposition depends only on the boundedness of the cost function; in contrast, as the following example shows, the continuity of $J^*(\cdot, x)$ at $\lambda = 0$ can not be generally ensured under the sole assumption of bounded costs.

Example 2.1 Let the state space and the action set be given by $S = \{0, 1, -1\}$ and the singleton $A = \{a\}$, respectively, so that $A(x) = \{a\}$ for every $x \in S$. Define the cost function by

$$C(x, a) \equiv C(x) = x, \quad x \in S,$$

and let the transition law $[p_{xy}(a)] \equiv [p_{xy}]$ be determined by

$$p_{0,x} = \frac{1}{2}, \quad p_{xx} = 1, \quad x = 1, -1.$$

For this model there is a unique (stationary) policy, say f, and it will not be indicated in the diverse quantities appearing below. Note that the states 1 and -1 are absorbing, and then the specification of the cost function yields that

$$J(\lambda, x) = x, \quad x = -1, 1, \quad \lambda \in \mathbb{R}.$$

Now suppose that the system starts at $X_0 = 0$. In this case at time 1 the system will arrive to state 1 or -1 with probability 1/2 and, for each integer $n > 1$, the Markov property and the definition of the cost function together yield that

$$E_0\left[\sum_{t=0}^{n-1} C(X_t)\right] = C(0) + \frac{1}{2}\left(E_{-1}\left[\sum_{t=0}^{n-2} C(X_t)\right] + E_1\left[\sum_{t=0}^{n-2} C(X_t)\right]\right) = 0,$$

and then

$$J(0, 0) = 0. \tag{2.14}$$

On the other hand, for $\lambda \neq 0$,

$$E_0\left[e^{\lambda \sum_{t=0}^{n-1} C(X_t)}\right] = \frac{1}{2}\left(e^{\lambda(n-2)} + e^{-\lambda(n-2)}\right) = \frac{1}{2}\left(e^{|\lambda|(n-2)} + e^{-|\lambda|(n-2)}\right),$$

so that

$$e^{|\lambda|(n-2)}/2 \leq E_0\left[e^{\lambda \sum_{t=0}^{n-1} C(X_t)}\right] \leq e^{|\lambda|(n-2)},$$

a relation that via (2.6)–(2.8) leads to

$$J(\lambda, 0) = \lim_{n\to\infty} \frac{1}{n\lambda} \log\left(E_0\left[e^{\lambda \sum_{t=0}^{n-1} C(X_t)}\right]\right) = \frac{|\lambda|}{\lambda} = \operatorname{sign}(\lambda), \quad \lambda \neq 0.$$

Combining this last display with (2.14) it follows that $J(\cdot, 0)$ is not continuous at $\lambda = 0$.

In the above example, the discontinuity of the mapping $\lambda \mapsto J^*(\lambda, 0)$ can be traced back to the fact that the recurrent states -1 and 1 do not communicate, and the main problem considered in this note can be now stated as follows:

- To determine *conditions on the transition law* ensuring that, for every continuous and bounded cost function, the mapping $\lambda \mapsto J^*(\lambda, x)$ is continuous at $\lambda = 0$ for each $x \in S$.

The result in this direction is presented in the following section.

3 Main result

The continuity of the optimal value function $J^*(\cdot, x)$ at zero will be studied under the (variant of the) simultaneous Doeblin condition stated below. The following notation will be used: For each nonempty set $F \subset S$, the return time T_F is defined by

$$T_F := \min\{n \in \mathbb{N} \setminus \{0\} \mid X_n \in F\} \tag{3.1}$$

where, according to the usual convention, the minimum of the empty set is ∞; when $F = \{z\}$ is a singleton, instead of $T_{\{z\}}$ the simpler notation T_z is employed.

Assumption 3.1 There exists a finite set $F \subset S$ and a constant $M > 0$ satisfying the following properties (i) and (ii):

(i) $E_x^f\,[T_F] \leq M, \quad x \in S \setminus F, \quad f \in \mathbb{F};$

(ii) For each $y \in F$ there exists a policy $f^y \in \mathbb{F}$ such that

$$E_x^{f^y}\left[T_y\right] \leq M, \quad x \in F \setminus \{y\}. \tag{3.2}$$

Frequently, this simultaneous Doeblin condition is stated using a singleton instead of F; in this case the second condition is vacuously satisfied and each stationary policy has a unique positive recurrent class, a property that is not necessarily valid under Assumption 3.1. On the other hand, observe that in Example 2.1 the first part of Assumption 3.1 holds with $F = \{-1, 1\}$ and $M = 1$, but the second part fails. The main objective of this note is to establish the following result.

Theorem 3.1 *Under Assumptions 2.1 and 3.1, for each state $x \in S$ the mapping*

$$\lambda \mapsto J^*(\lambda, x) \quad \textit{is continuous at } \lambda = 0.$$

The proof of this theorem will be presented after establishing the necessary technical tools in the following two sections. At this point it is interesting to observe that, under Assumption 2.1, the simultaneous Doeblin condition in Assumption 3.1 has an important consequence in the analysis of the *risk-neutral* average cost criterion, namely, the corresponding optimality equation has a *bounded* solution, a fact that has two substantial implications: (a) the optimal average cost is constant, as well as (b) an optimal stationary policy exists (Arapostathis et al. 1993; Hernández-Lerma 1989; Puterman 1994). These two properties can be also ensured by imposing conditions under which the optimality equation has a *possibly unbounded solution*, like the *Lyapunov function condition* (LFC) introduced in Hordijk (1974) which, under Assumption 2.1, can be formulated as follows in the present context of bounded costs (Cavazos-Cadena and Hernández-Lerma 1992):

$$\text{There exists a state } z \text{ such that}$$
$$\lim_{n\to\infty} \sup_{f\in\mathbb{F}} E_x^f[T_z I[T_z > n]] = 0, \quad x \in S. \tag{3.3}$$

It is natural to ask if the continuity result in Theorem 3.1 still holds when the LFC replaces Assumption 3.1. As it is shown in the following example (concerning an uncontrolled model), the answer to this question is negative.

Example 3.1 Suppose that $S = \mathbb{N}$ and that $A = \{a\} = A(x)$ for every $x \in S$. Let the cost function be given by

$$C(0, a) \equiv C(0) = 0, \quad \text{and } C(x, a) \equiv C(x) = 1 \quad \text{for } x \in S \setminus \{0\}, \tag{3.4}$$

and consider the transition law $[p_{xy}(a)] \equiv [p_{xy}]$ determined by

$$p_{0,0} = 1, \quad p_{x\,x+1} = \frac{x^2}{(x+1)^2} = 1 - p_{x0}, \quad x = 1, 2, 3, \ldots$$

There is only one policy f in this model, and it will not be explicitly indicated in the diverse expressions appearing below.

- It will be shown that *the Lyapunov function condition holds* for this model with $z = 0$. To achieve this goal observe that, since there is only one policy, (3.3) is equivalent to
$$E_x[T_0] < \infty, \quad x \in S. \tag{3.5}$$

To verify this property note that $E_0[T_0] = 1$, because 0 is an absorbing state. Now, consider a no-null state x and observe that the specification of the transition law yields that for every positive integer n

$$\begin{aligned} P_x[T_0 > n] &= P_x[X_r = x + r,\ 0 < r \le n] \\ &= \frac{x^2}{(x+1)^2}\frac{(x+1)^2}{(x+2)^2}\cdots\frac{(x+n-1)^2}{(x+n)^2} \\ &= \frac{x^2}{(x+n)^2} \end{aligned} \tag{3.6}$$

and then

$$E_x[T_0] = 1 + \sum_{n=1}^{\infty} P_x[T_0 > n] = 1 + \sum_{n=1}^{\infty} \frac{x^2}{(x+n)^2} \le 1 + x, \tag{3.7}$$

completing the verification of (3.5).

- Next, it will be proved that, for each $x \neq 0$, the mapping $\lambda \mapsto J(\lambda, x)$ is not continuous at zero. First, note that

$$J(\lambda, 0) = 0, \quad \lambda \in \mathbb{R},$$

because the state 0 is absorbing and $C(0) = 0$; since the Lyapunov function condition implies that the risk-neutral average cost $J(0, x)$ does not depend on the state x (Hordijk 1974), it follows that

$$J(0, x) = 0, \quad x \in S. \tag{3.8}$$

Now, let the state $x \neq 0$ and $\lambda > 0$ be arbitrary, and observe that $X_t \neq 0$ occurs P_x-almost surely on the event $[t < T_0]$, so that (3.4) and (3.6) lead to

$$e^{\lambda n} \ge E_x\left[e^{\lambda \sum_{t=0}^{n-1} C(X_t)}\right] \ge E_x\left[e^{\lambda \sum_{t=0}^{n-1} C(X_t)} I[T_0 > n]\right] = e^{\lambda n} x^2/(x+n)^2.$$

Thus, using (2.6) and (2.7), $1 \ge \frac{1}{n} J_n(\lambda, x) > 1 + \log(x^2/(x+n)^2)/(\lambda n)$, and it follows that

$$1 = \lim_{n\to\infty} \frac{1}{n} J_n(\lambda, x) = J(\lambda, x), \quad x \neq 0, \quad \lambda > 0,$$

and then $J(\cdot, x)$ is not continuous at $\lambda = 0$, by (3.8). In short, the (uncontrolled) model in this example satisfies the Lyapunov function condition, but the mapping $\lambda \mapsto J^*(\lambda, x)$ is not continuous at zero when the state x is no-null.

4 Optimality equation

The proof of Theorem 3.1 relies on the following optimality equation for the average criterion in (2.7):

$$U_\lambda(g_\lambda + h_\lambda(x)) = \inf_{a \in A(x)} \sum_{y \in S} p_{xy}(a) U_\lambda(C(x, a) + h_\lambda(y)), \quad x \in S, \tag{4.1}$$

where $g_\lambda \in \mathbb{R}$ and $h_\lambda(\cdot)$ is a real valued function defined on S. The important role of this equation in the study of the λ-sensitive average index is signaled by the following verification result.

Lemma 4.1 *Given $\lambda \in \mathbb{R}$, let $g_\lambda \in \mathbb{R}$ and $h_\lambda \in \mathcal{B}(S)$ be such that the optimality equation (4.1) holds. In this case,*

(i) For each $x \in S$ the summation in the right-hand side of (4.1) is a continuous function of $a \in A(x)$, and then has a minimizer $f_\lambda(x) \in A(x)$.
(ii) For every state $x \in S$,

$$g_\lambda = \lim_{n\to\infty} \frac{1}{n} J_n(\lambda, f_\lambda, x) = J^*(\lambda, x) = J_*(\lambda, x).$$

For the risk-averse case $\lambda > 0$ this result has been established, for instance, in Di Masi and Stettner (1999). By completeness, a short proof is given below in the present context in which the sign of λ is unrestricted; the argument shows that the conclusion relies heavily on the fact that the exponential utility $U_\lambda(\cdot)$ is always (strictly) increasing.

Proof (i) Using that the cost function and h_λ are bounded, the conclusion follows combining the bounded convergence theorem with Assumption 2.1.
(ii) Note that (4.1) implies that

$$E_x^\pi\left[U_\lambda(C(X_0, A_0) + h_\lambda(X_1))\right] \geq U_\lambda(g_\lambda + h_\lambda(x))$$

for every $x \in S$ and $\pi \in \mathcal{P}$; from this point an induction argument using the Markov property yields that for every positive integer n

$$\begin{aligned} E_x^\pi\left[U_\lambda\left(\sum_{t=0}^{n-1} C(X_t, A_t) + \|h_\lambda\|\right)\right] &\geq E_x^\pi\left[U_\lambda\left(\sum_{t=0}^{n-1} C(X_t, A_t) + h_\lambda(X_n)\right)\right] \\ &\geq U_\lambda(ng_\lambda + h_\lambda(x)) \end{aligned}$$

where the first inequality used that U_λ is increasing. Together with (2.3)–(2.5) this last display leads to $\|h_\lambda\| + J_n(\lambda, \pi, x) \geq ng_\lambda + h_\lambda(x)$ and, dividing both sides of this relation by n and taking the inferior limit as n goes to ∞ in both sides of the resulting inequality, it follows that $J_-(\lambda, \pi, x) \geq g_\lambda$; since the policy π and the state x are arbitrary, via (2.9) and (2.10), this yields that

$$\liminf_{n\to\infty} \frac{1}{n} J_n(\lambda, f_\lambda, x) \geq J_*(\lambda, x) \geq g_\lambda, \quad x \in S. \tag{4.2}$$

On the other hand, the definition of the policy f_λ implies that

$$E_x^{f_\lambda}\left[U_\lambda(C(X_0, A_0) + h_\lambda(X_1))\right] = U_\lambda(g_\lambda + h_\lambda(x)), \quad x \in S,$$

and an induction argument yields that, for every $n = 1, 2, 3, \ldots,$

$$U_\lambda(ng_\lambda + h_\lambda(x)) = E_x^{f_\lambda}\left[U_\lambda\left(\sum_{t=0}^{n-1} C(X_t, A_t) + h_\lambda(X_n)\right)\right]$$
$$\geq E_x^{f_\lambda}\left[U_\lambda\left(\sum_{t=0}^{n-1} C(X_t, A_t) - \|h_\lambda\|\right)\right],$$

where the inequality is due to the fact that U_λ is increasing. Via (2.3)–(2.5) this relation yields that $ng_\lambda + h(x) \geq J_n(\lambda, f_\lambda, x) - \|h_\lambda\|$ and then

$$g_\lambda \geq \limsup_{n\to\infty} \frac{1}{n} J_n(\lambda, f_\lambda, x) \geq J^*(\lambda, x), \quad x \in S;$$

see (2.8) for the second inequality. Now, the conclusion follows combining this last display with (4.2) and (2.11). □

The main tool that will be used to establish Theorem 3.1 is the following result, which establishes that, for some $\delta > 0$, the optimality equations corresponding to a risk-sensitivity parameter $\lambda \in (-\delta, \delta) \setminus \{0\}$ has a solution $(g_\lambda, h_\lambda(\cdot))$, in such a way that the family $\{h_\lambda\}_{0<|\lambda|<\delta}$ is bounded in $\mathcal{B}(S)$.

Theorem 4.1 *Under Assumptions 2.1 and 3.1, there exist positive numbers δ and B such that*

for every $\lambda \in (-\delta, \delta)\setminus\{0\}$ the optimality equation (4.1) has a solution $(g_\lambda, h_\lambda(\cdot))$ satisfying $\|h_\lambda\| \leq B$.

This theorem extends the main result in Cavazos-Cadena (2003), where assuming that (i) the state and action sets are finite, (ii) that λ is positive, and (iii) that Assumption 3.1 holds for a singleton $F \subset S$, it was establish that, *for a given* $\lambda \in (-\delta, \delta)$, (4.1) has a bounded solution. The proof of Theorem 4.1 uses the discounted approach, that will be presented in the following section, and relies heavily on the next consequence of Assumption 3.1, whose proof uses classical ideas in the analysis of the simultaneous Doeblin condition in Hordijk (1974).

Lemma 4.2 *Suppose that Assumption 3.1 holds. In this case, there exist $b \in (1, \infty)$ and $\rho \in (0, 1)$ such that the following assertions (i) and (ii) hold for every $n \in \mathbb{N}$:*

(i) $P_x^f[T_F \geq n] \leq b\rho^n$ for every $x \in S \setminus F$ and $f \in \mathbb{F}$.
(ii) For each $y \in F$, the stationary policy f^y in the second part of Assumption 3.1 satisfies that

$$P_x^{f^y}[T_y \geq n] \leq b\rho^n, \quad x \in S \setminus \{y\},$$

Proof Using the notation in Assumption 3.1 let the positive integer N be such that $N \geq 2M$. In this case Markov's inequality implies that

$$P_x^f[T_F \geq N] \leq \frac{E_x^f[T_F]}{N} \leq \frac{M}{N} \leq \frac{1}{2}, \quad x \in S \setminus F, \quad f \in \mathbb{F}.$$

Starting from this relation, an induction argument using the Markov property yields that, for every state $x \in S \setminus F$ and $f \in \mathbb{F}$, the inequality $P_x^f[T_F \geq kN] \leq (1/2)^k$ holds for every nonnegative integer k. Next, given $n \in \mathbb{N}$ write $n = kN + r$ where $r \in \{0, 1, 2, \ldots, N-1\}$, and note that $P_x^f[T_F \geq n] \leq P_x^f[T_F \geq kN] \leq (1/2)^k$, so that

$$P_x^f[T_F \geq n] \leq \tilde{b}\tilde{\rho}^n, \quad x \in S \setminus F, \quad f \in \mathbb{F}, \quad n \in \mathbb{N}, \tag{4.3}$$

where $\tilde{\rho} = (1/2)^{1/N} \in (0, 1)$ and $\tilde{b} = \tilde{\rho}^{-N} \in (1, \infty)$. Now, for each $y \in F$ consider the policy $f^y \in \mathbb{F}$ in Assumption 3.1. It will be shown that

$$E_x^{f^y}[T_y] \leq 2M, \quad x \in S \setminus \{y\}. \tag{4.4}$$

To achieve this goal note that the inequality holds for $x \in F \setminus \{y\}$, by (3.2). To complete the argument note that the inclusion $y \in F$ implies that $T_F \leq T_y$, so that

$$\begin{aligned} E_x^{f^y}[T_y] &= E_x^{f^y}[T_F] + E_x^{f^y}[(T_y - T_F)I[T_y > T_F]] \\ &\leq M + E_x^{f^y}[(T_y - T_F)I[T_y > T_F]], \quad x \in S \setminus F, \end{aligned}$$

where the inequality stems form Assumption 3.1(i). Next, observing that $X_{T_F} \in F \setminus \{y\}$ on the event $T_y > T_F$ and using (3.1), the strong Markov property applied to the Markov chain induced by f^y yields, via (3.2), that $E_x^{f^y}[(T_y - T_F)I[T_y > T_F] | T_y > T_F, X_{T_F}] = E_{X_{T_F}}[T_y] \leq M$, and then

$$E_x^{f^y}[(T_y - T_F)I[T_y > T_F]] \leq M,$$

a relation that together with the previous display yields that the inequality in (4.4) is also valid when $x \in S \setminus F$. Consider now the 'reduced' model $\mathcal{M}_{f^y}$ obtained from $\mathcal{M}$ by restricting the available actions at each $x \in S$ to the singleton $\{f^y(x)\}$. For this model $\mathcal{M}_{f^y}$, relation (4.4) establishes that Assumption 3.1 holds with $F = \{y\}$, and then the first part of the proof yields that there exist $b_y \in (1, \infty)$ and $\rho_y \in (0, 1)$ such that

$$P_x^{f^y}[T_y \geq n] \leq b_y \rho_y^n, \quad n \in \mathbb{N}, \quad x \in S \setminus \{y\}.$$

Setting $b = \max\{\tilde{b}, b_y, y \in F\} \in (1, \infty)$ and $\rho := \max\{\tilde{\rho}, \rho_y, y \in F\} \in (0, 1)$, the above display and (4.3) together yield that the desired conclusions (i) and (ii) hold. □

5 Discounted approach

This section presents the auxiliary results that will be used to establish the uniform boundedness result in Theorem 4.1, which will play a central role in the proof of Theorem 3.1. The approach relies on following discounted operators introduced in Di Masi and Stettner (1999).

Definition 5.1 For $\alpha \in (0, 1)$ and $\lambda \in \mathbb{R} \setminus \{0\}$, the operator $T_{\alpha,\lambda} : \mathcal{B}(S) \to \mathcal{B}(S)$ is defined as follows: For each $V \in \mathcal{B}(S)$, the function $T_{\alpha,\lambda}[V]$ is determined by

$$U_\lambda(T_{\alpha,\lambda}[V](x)) = \inf_{a\in A(x)} \left[\sum_{y\in S} p_{xy}(a) U_\lambda(C(x, a) + \alpha V(y)) \right], \quad x \in S. \tag{5.1}$$

Combining (2.2) with the fact that U_λ is increasing, it is not difficult to see that $T_{\alpha\,\lambda}$ is a monotone and α-homogeneous operator, that is, given $V, W \in \mathcal{B}(S)$, $T_{\alpha,\lambda}[V] \geq T_{\alpha,\lambda}[W]$ when $V \geq W$, and $T_{\alpha,\lambda}[V + r] = T_{\alpha,\lambda}[V] + \alpha r$ for every $r \in \mathbb{R}$. These properties yield that $T_{\alpha,\lambda}$ is a contractive operator on the space $\mathcal{B}(S)$ endowed with the supremum norm, and that its contraction module is α, that is Di Masi and Stettner (1999),

$$\|T_{\alpha,\lambda}[V] - T_{\alpha,\lambda}[W]\| \leq \alpha\|V - W\|, \quad V, W \in \mathcal{B}(S). \tag{5.2}$$

Consequently, by Banach's fixed point theorem, there exists a unique function $V_{\alpha,\lambda} \in \mathcal{B}(S)$ satisfying $T_{\alpha,\lambda}[V_{\alpha,\lambda}] = V_{\alpha,\lambda}$; more explicitly,

$$U_\lambda(V_{\alpha,\lambda}(x)) = \inf_{a\in A(x)} \left[\sum_{y\in S} p_{xy}(a) U_\lambda(C(x, a) + \alpha V_{\alpha,\lambda}(y)) \right], \quad x \in S. \tag{5.3}$$

Observe now that (5.1) yields that $T_{\alpha,\lambda}[0](x) = \inf_{a\in A(x)} C(x, a)$, so that $\|T_{\alpha,\lambda}[0]\| \leq \|C\|$. Using (5.2) with $V_{\alpha,\lambda}$ and 0 instead of V and W, respectively, it follows that

$$(1 - \alpha)\|V_{\alpha,\lambda}\| \leq \|C\|. \tag{5.4}$$

Next, for each $x \in S$ define

$$g_{\alpha,\lambda}(x) = (1 - \alpha)V_{\alpha,\lambda}(x) \in [-\|C\|, \|C\|], \quad \alpha \in (0, 1), \quad \lambda \in \mathbb{R} \setminus \{0\}, \tag{5.5}$$

and note that the boundedness of $V_{\alpha,\lambda}$ and $C(\cdot, \cdot)$ together with Assumption 2.1 imply that, for every $\lambda \neq 0$ and $\alpha \in (0, 1)$, there exists a stationary policy $f_{\alpha,\lambda}$ satisfying

$$\begin{aligned} U_\lambda(V_{\alpha,\lambda}(x)) &= \sum_{y\in S} p_{xy}(f_{\alpha,\lambda}(x)) U_\lambda(C(x, f_{\alpha,\lambda}(x)) + \alpha V_{\alpha,\lambda}(y)) \\ &= E_x^{f_{\alpha,\lambda}} \left[U_\lambda(C(X_0, A_0,) + \alpha V_{\alpha,\lambda}(X_1)) \right], \quad x \in S. \end{aligned} \tag{5.6}$$

The main result of this section is the following theorem.

Theorem 5.1 *Under Assumptions 2.1 and 3.1, there exist $\delta > 0$ and $B^* > 0$ such that, for every $\lambda \in (-\delta, \delta) \setminus \{0\}$ and $\alpha \in (0, 1)$,*

$$\alpha|V_{\alpha,\lambda}(x) - V_{\alpha,\lambda}(z)| \leq B^*, \quad x, z \in F, \tag{5.7}$$

and

$$\alpha|V_{\alpha,\lambda}(x) - V_{\alpha,\lambda}(z)| \leq 2B^* \quad x \in S, \quad z \in F. \tag{5.8}$$

The proof of this theorem is based on the following lemma.

Lemma 5.1 *Suppose that Assumptions 2.1 and 3.1 hold. Let $b > 1$ and $\rho \in (0, 1)$ be as in Lemma 4.2 and set*

$$\delta := \frac{1}{2\|C\|}\log\left(\frac{1+\rho}{2\rho}\right) \quad \text{and} \quad B = \log\left(\frac{2b}{1-\rho}\right). \tag{5.9}$$

With this notation, for each $\alpha \in (0, 1)$ the assertions (i)–(iv) below hold, where F is the finite set in Assumption 3.1, and the return time T_F is as in (3.1).

(i) For each $x \in S \setminus F$ and $f \in \mathbb{F}$,

$$e^{-|\lambda|B/\delta} \le E_x^f[e^{2\lambda\|C\|T_F}] < e^{|\lambda|B/\delta}, \quad 0 < |\lambda| \le \delta.$$

(ii) For each $z \in F$ the stationary policy f^z in Assumption 3.1(ii) satisfies that

$$e^{-|\lambda|B/\delta} \le E_x^{f^z}[e^{2\lambda\|C\|T_z}] < e^{|\lambda|B/\delta}, \quad x \in S \setminus \{z\}, \quad 0 < |\lambda| \le \delta.$$

(iii) For each $x \in S \setminus F$, $f \in \mathbb{F}$ and $\lambda \in (-\delta, \delta) \setminus \{0\}$,

$$\begin{aligned}\lim_{n\to\infty} E_x^f&\left[U_\lambda\left(\sum_{t=0}^{T_F\wedge n-1}[C(X_t, A_t) - g_{\alpha,\lambda}(X_t)] + \alpha V_{\alpha,\lambda}(X_{T_F\wedge n})\right)\right]\\ &= E_x^f\left[U_\lambda\left(\sum_{t=0}^{T_F-1}[C(X_t, A_t) - g_{\alpha,\lambda}(X_t)] + \alpha V_{\alpha,\lambda}(X_{T_F})\right)\right];\end{aligned} \tag{5.10}$$

see (5.5) for the definition of $g_{\alpha,\lambda}(\cdot)$.

(iv) For each $\lambda \in (-\delta, \delta) \setminus \{0\}$, and $z \in F$, the policy $f^z \in \mathbb{F}$ in Assumption 3.1(ii) satisfies that, for every $x \in S \setminus \{z\}$,

$$\begin{aligned}\lim_{n\to\infty} E_x^{f^z}&\left[U_\lambda\left(\sum_{t=0}^{T_z\wedge n-1}[C(X_t, A_t) - g_{\alpha,\lambda}(X_t)] + \alpha V_{\alpha,\lambda}(X_{T_z\wedge n})\right)\right]\\ &= E_x^{f^z}\left[U_\lambda\left(\sum_{t=0}^{T_z-1}[C(X_t, A_t) - g_{\alpha,\lambda}(X_t)] + \alpha V_{\alpha,\lambda}(X_{T_z})\right)\right].\end{aligned} \tag{5.11}$$

Proof (i) Let $x \in S \setminus F$ and $f \in \mathbb{F}$ be arbitrary. By Lemma 4.2(i), the inequality $P_x^f[T_F \ge n] \le b\rho^n$ holds for every n. Thus, since (5.9) yields that $e^{2\delta\|C\|}\rho = (1+\rho)/2 < 1$, it follows that

$$\begin{aligned}E_x^f[e^{2\delta\|C\|T_F}] &= \sum_{k=1}^{\infty} e^{2\delta\|C\|k} P_x^f[T_F = k]\\ &\le b\sum_{k=1}^{\infty} e^{2\delta\|C\|k}\rho^k \le \frac{b}{1-\rho e^{2\|C\|\delta}} = \frac{b}{1-(1+\rho)/2} = e^B,\end{aligned}$$

where (5.9) was used to set the last equality. Now let $\lambda \in (-\delta, \delta) \setminus \{0\}$ be arbitrary but fixed, and note that $E_x^f[e^{2\lambda\|C\|T_F}] \le E_x^f[e^{(|\lambda|/\delta)2\delta\|C\|T_F}]$; from this point, Jensen's inequality applied to the concave function $\varphi(x) = x^{|\lambda|/\delta}$ on $(0, \infty)$ yields that $E_x^f[e^{2\lambda\|C\|T_F}] \le E_x^f[e^{2\delta\|C\|T_F}]^{(|\lambda|/\delta)}$, and together with the above display this leads to

$$E_x^f\left[e^{2\lambda\|C\|T_F}\right] \le e^{|\lambda|B/\delta}.$$

Observe now that, if W is a positive random variable, then $E[W^{-1}] \ge 1/E[W]$, by Jensen's inequality. Combining this fact with the above display it follows that

$$e^{|\lambda|B/\delta} \ge E_x^f\left[e^{2\lambda\|C\|T_F}\right] \ge E_x^f\left[e^{-2|\lambda|\|C\|T_F}\right] \ge 1/E_x^f\left[e^{2|\lambda|\|C\|T_F}\right] \ge e^{-|\lambda|B/\delta}.$$

(ii) Using Lemma 4.2(ii), the conclusion follows paralleling the argument used to establish the previous part.

(iii) Let $\lambda \in (-\delta, \delta) \setminus \{0\}$, $x \in S \setminus F$ and $f \in \mathbb{F}$ be arbitrary, and observe the following properties (a) and (b):

(a) For each positive integer n, the equality

$$\begin{aligned}\sum_{t=0}^{T_F \wedge n-1} \left[C(X_t, A_t) - g_{\alpha,\lambda}(X_t)\right] + \alpha V_{\alpha,\lambda}(X_{T_F \wedge n})\\ = \sum_{t=0}^{T_F} \left[C(X_t, A_t) - g_{\alpha,\lambda}(X_t)\right] + \alpha V_{\alpha,\lambda}(X_{T_F})\end{aligned}$$

holds on the event $[T_F \le n]$. Since T_F is finite P_x^f-a. s., it follows that

$$\begin{aligned}\lim_{n\to\infty} U_\lambda\left(\sum_{t=0}^{T_F \wedge n-1} [C(X_t, A_t) - g_{\alpha,\lambda}(X_t)] + \alpha V_{\alpha,\lambda}(X_{T_F \wedge n})\right)\\ = U_\lambda\left(\sum_{t=0}^{T_F-1} [C(X_t, A_t) - g_{\alpha,\lambda}(X_t)] + \alpha V_{\alpha,\lambda}(X_{T_F})\right) \quad P_x^f\text{-a. s.}\end{aligned}$$

(b) By (2.1) and (5.5),

$$\begin{aligned}&\left|U_\lambda\left(\sum_{t=0}^{T_F \wedge n-1} [C(X_t, A_t) - g_{\alpha,\lambda}(X_t)] + \alpha V_{\alpha,\lambda}(X_{T_F \wedge n})\right)\right|\\ &\le e^{|\lambda| \sum_{t=0}^{T_F \wedge n-1} |C(X_t, A_t) - g_{\alpha,\lambda}(X_t)| + \alpha|\lambda||V_{\alpha,\lambda}(X_{T_F \wedge n})|}\\ &\le e^{2|\lambda|(T_F \wedge n)\|C\| + \alpha|\lambda|\|C\|/(1-\alpha)}\\ &\le e^{2\delta\|C\|T_F} e^{\alpha|\lambda|\|C\|/(1-\alpha)}.\end{aligned}$$

Since $E_x^f\left[e^{2\delta\|C\|T_F}\right]$ is finite, by part (i), the property (5.10) follows combining (a) and (b) with the bounded convergence theorem. A similar argument using part (ii) can be used to establish the convergence (5.11). □

Proof of Theorem 5.1. Given $z \in F$, let f^z be the stationary policy in Assumption 3.1(ii). From (5.3) it follows that for every $x \in S$

$$U_\lambda(V_{\alpha,\lambda}(x)) \le \sum_{y\in S} p_{xy}(f^z(x))U_\lambda(C(x, f^z(x)) + \alpha V_{\alpha,\lambda}(y)),$$

an inequality that via (2.1) and (5.5) leads to

$$\begin{aligned} &U_\lambda(\alpha V_{\alpha,\lambda}(x)) \\ &\quad\le \sum_{y\in S} p_{xy}(f^z(x))U_\lambda(C(x, f^z(x)) - g_{\alpha,\lambda}(x) + \alpha V_{\alpha,\lambda}(y)) \\ &\quad= E_x^{f^z}\left[U_\lambda(C(X_0, A_0) - g_{\alpha,\lambda}(X_0) + \alpha V_{\alpha,\lambda}(X_1))\right], \quad x \in S. \end{aligned} \tag{5.12}$$

It will be proved by induction that, for every positive integer n, and $x \in S$,

$$\begin{aligned} &U_\lambda(\alpha V_{\alpha,\lambda}(x)) \\ &\quad\le E_x^{f^z}\left[U_\lambda\left(\sum_{t=0}^{T_z\wedge n-1}[C(X_t, A_t) - g_{\alpha,\lambda}(X_t)] + \alpha V_{\alpha,\lambda}(X_{T_z\wedge n})\right)\right]. \end{aligned} \tag{5.13}$$

To establish this claim note that for $n = 1$ the above relation is equivalent to (5.12), because $T_z \ge 1$. Suppose now that (5.13) holds for some positive integer n. Let $x \in S$ be arbitrary and observe that

$$\begin{aligned} &E_x^{f^z}\left[U_\lambda\left(\sum_{t=0}^{T_z\wedge n-1}[C(X_t, A_t) - g_{\alpha,\lambda}(X_t)] + \alpha V_{\alpha,\lambda}(X_{T_z\wedge n})\right)\right] \\ &= E_x^{f^z}\left[U_\lambda\left(\sum_{t=0}^{T_z-1}[C(X_t, A_t) - g_{\alpha,\lambda}(X_t)] + \alpha V_{\alpha,\lambda}(X_{T_z})\right)I[T_z \le n]\right] \\ &\quad+ E_x^{f^z}\left[U_\lambda\left(\sum_{t=0}^{n-1}[C(X_t, A_t) - g_{\alpha,\lambda}(X_t)] + \alpha V_{\alpha,\lambda}(X_n)\right)I[T_z > n]\right]. \end{aligned} \tag{5.14}$$

Next, using (2.2) note that

$$\begin{aligned} &U_\lambda\left(\sum_{t=0}^{n-1}[C(X_t, A_t) - g_{\alpha,\lambda}(X_t)] + \alpha V_{\alpha,\lambda}(X_n)\right)I[T_z > n] \\ &\quad= e^{\lambda\sum_{t=0}^{n-1}[C(X_t,A_t)-g_{\alpha,\lambda}(X_t)]} I[T_z > n]U_\lambda\left(\alpha V_{\alpha,\lambda}(X_n)\right), \end{aligned}$$

and observe that (5.12) implies that

$$
\begin{aligned}
&U_\lambda\left(\alpha V_{\alpha,\lambda}(X_n)\right)\\
&\quad\leq \sum_{y\in S} p_{X_n\,y}(f^z(X_n))U_\lambda(C(X_n, f^z(X_n)) - g_{\alpha,\lambda}(X_n) + \alpha V_{\alpha,\lambda}(y))\\
&\quad= E_x^{f^z}\left[U_\lambda(C(X_n, A_n) - g_{\alpha,\lambda}(X_n) + \alpha V_{\alpha,\lambda}(X_{n+1}))\,|X_k, k \leq n\right],
\end{aligned}
$$

where the equality is due to the Markov property. Using (2.2) the two last displays yield that

$$
\begin{aligned}
&U_\lambda\left(\sum_{t=0}^{T_z\wedge n-1}[C(X_t, A_t) - g_{\alpha,\lambda}(X_t)] + \alpha V_{\alpha,\lambda}(X_{T_z\wedge n})\right) I[T_z > n]\\
&\leq E_x^{f^z}\left[U_\lambda\left(\sum_{t=0}^{n}[C(X_t, A_t) - g_{\alpha,\lambda}(X_t)] + \alpha V_{\alpha,\lambda}(X_{n+1})\right) I[T_z > n]\,\middle|\, X_k, k \leq n\right]
\end{aligned}
$$

and then

$$
\begin{aligned}
&E_x^{f^z}\left[U_\lambda\left(\sum_{t=0}^{T_z\wedge n-1}[C(X_t, A_t) - g_{\alpha,\lambda}(X_t)] + \alpha V_{\alpha,\lambda}(X_{T_z\wedge n})\right) I[T_z > n]\right]\\
&\quad\leq E_x^{f^z}\left[U_\lambda\left(\sum_{t=0}^{n}[C(X_t, A_t) - g_{\alpha,\lambda}(X_t)] + \alpha V_{\alpha,\lambda}(X_{n+1})\right) I[T_z > n]\right]\\
&\quad= E_x^{f^z}\left[U_\lambda\left(\sum_{t=0}^{T_z-1}[C(X_t, A_t) - g_{\alpha,\lambda}(X_t)] + \alpha V_{\alpha,\lambda}(X_{T_z})\right) I[T_z = n+1]\right]\\
&\qquad+ E_x^{f^z}\left[U_\lambda\left(\sum_{t=0}^{n}[C(X_t, A_t) - g_{\alpha,\lambda}(X_t)] + \alpha V_{\alpha,\lambda}(X_{n+1})\right) I[T_z > n+1]\right];
\end{aligned}
$$

combining this relation with (5.14) it follows that

$$
\begin{aligned}
&E_x^{f^z}\left[U_\lambda\left(\sum_{t=0}^{T_z\wedge n-1}[C(X_t, A_t) - g_{\alpha,\lambda}(X_t)] + \alpha V_{\alpha,\lambda}(X_{T_z\wedge n})\right)\right]\\
&\quad\leq E_x^{f^z}\left[U_\lambda\left(\sum_{t=0}^{T_z-1}[C(X_t, A_t) - g_{\alpha,\lambda}(X_t)] + \alpha V_{\alpha,\lambda}(X_{T_z})\right) I[T_z \leq n+1]\right]
\end{aligned}
$$

$$+ E_x^{f^z}\left[U_\lambda\left(\sum_{t=0}^{n}[C(X_t, A_t) - g_{\alpha,\lambda}(X_t)] + \alpha V_{\alpha,\lambda}(X_{n+1})\right) I[T_z > n+1]\right]$$
$$= E_x^{f^z}\left[U_\lambda\left(\sum_{t=0}^{T_z\wedge(n+1)-1}[C(X_t, A_t) - g_{\alpha,\lambda}(X_t)] + \alpha V_{\alpha,\lambda}(X_{T_z\wedge(n+1)})\right)\right],$$

a relation that together with the induction hypothesis yields that (5.13) holds with $n+1$ instead of n, completing the induction argument. Taking the limit as n goes to ∞ in (5.12), via the convergence (5.11) in Lemma 5.1(ii) it follows that, for every $x \in S$

$$U_\lambda(\alpha V_{\alpha,\lambda}(x)) \leq E_x^{f^z}\left[U_\lambda\left(\sum_{t=0}^{T_z-1}[C(X_t, A_t) - g_{\alpha,\lambda}(X_t)] + \alpha V_{\alpha,\lambda}(X_{T_z})\right)\right];$$

using that T_z is finite $P_x^{f^z}$-almost surely and that $X_{T_z} = z$ when $T_z < \infty$, this inequality and (2.2) yield that

$$U_\lambda(\alpha[V_{\alpha,\lambda}(x) - V_{\alpha,\lambda}(z)]) \leq E_x^{f^z}\left[U_\lambda\left(\sum_{t=0}^{T_z-1}[C(X_t, A_t) - g_{\alpha,\lambda}(X_t)]\right)\right]$$
$$\leq E_x^{f^z}\left[U_\lambda\left(2\|C\|T_z\right)\right], \quad x \in S, \tag{5.15}$$

where the fact that U_λ is increasing was used in the last step. Consider now the following exahustive cases about λ:

Case 1 $\lambda \in (0, \delta)$. In this context $U_\lambda(w) = e^{\lambda w}$ for every $w \in \mathbb{R}$, so that (5.15) yields that $e^{\lambda\alpha[V_{\alpha,\lambda}(x) - V_{\alpha,\lambda}(z)]} \leq E_x^{f^z}\left[e^{2\lambda\|C\|T_z}\right]$, and then $e^{\lambda\alpha[V_{\alpha,\lambda}(x) - V_{\alpha,\lambda}(z)]} \leq e^{\lambda B/\delta}$ for $x \neq z$, by Lemma 5.1(ii); since this last inequality also holds for $x = z$ it follows that

$$\alpha[V_{\alpha,\lambda}(x) - V_{\alpha,\lambda}(z)] \leq \frac{B}{\delta}, \quad x \in S. \tag{5.16}$$

Case 2 $\lambda \in (-\delta, 0)$. In this framework $U_\lambda(w) = -e^{\lambda w}$ for every $w \in \mathbb{R}$, and (5.15) leads to $e^{\lambda\alpha[V_{\alpha,\lambda}(x) - V_{\alpha,\lambda}(z)]} \geq E_x^{f^z}\left[e^{2\lambda\|C\|T_z}\right]$, and via Lemma 5.1(ii) this implies that, for $x \neq z$, $e^{\lambda\alpha[V_{\alpha,\lambda}(x) - V_{\alpha,\lambda}(z)]} \geq e^{-|\lambda|B/\delta} = e^{\lambda B/\delta}$, an inequality that is also valid for $x = z$. From this point, recalling that λ is negative, it follows that (5.16) also occurs in the present case.

Since $z \in F$ was arbitrary in this argument, (5.16) implies that $\alpha[V_{\alpha,\lambda}(x) - V_{\alpha,\lambda}(z)] \leq B/\delta$ for every $x, z \in F$, so that (5.7) holds with

$$B^* = \frac{B}{\delta}. \tag{5.17}$$

Next, it will be shown that this constant B^* also satisfies (5.8). To begin with, note that by (5.7) it is sufficient to show that the inequality in (5.8) holds when $x \in S \setminus F$.

Now let $f_{\alpha,\lambda}$ be the stationary policy in (5.6), and multiply both sides of that equality by $e^{-\lambda(1-\alpha)V_{\alpha,\lambda}(x)}$ to obtain, via (2.2) and (5.5), that

$$U_\lambda(\alpha V_{\alpha,\lambda}(x)) = E_x^{f_{\alpha,\lambda}}\left[U_\lambda\left([C(X_0, A_0) - g_{\alpha,\lambda}(X_0)] + \alpha V_{\alpha,\lambda}(X_1)\right)\right], \quad x \in S.$$

Starting from this equality, an induction argument along the lines used in the above proof of (5.7) yields that, for every positive integer n and $x \in S \setminus F$,

$$U_\lambda(\alpha V_{\alpha,\lambda}(x)) = E_x^{f_{\alpha,\lambda}}\left[U_\lambda\left(\sum_{t=0}^{T_F\wedge n-1}[C(X_t, A_t) - g_{\alpha,\lambda}(X_t)] + \alpha V_{\alpha,\lambda}(X_{T_F\wedge n})\right)\right],$$

so that

$$U_\lambda(\alpha V_{\alpha,\lambda}(x)) = E_x^{f_{\alpha,\lambda}}\left[U_\lambda\left(\sum_{t=0}^{T_F-1}[C(X_t, A_t) - g_{\alpha,\lambda}(X_t)] + \alpha V_{\alpha,\lambda}(X_{T_F})\right)\right],$$

by (5.10). Multiplying both sides of this equality by $e^{-\lambda\alpha V_{\alpha,\lambda}(z)}$ where $z \in F$ is arbitrary but fixed, via (2.2) it follows that for every state x in $S \setminus F$,

$$\begin{aligned} &U_\lambda(\alpha[V_{\alpha,\lambda}(x) - V_{\alpha,\lambda}(z)]) \\ &\quad = E_x^{f_{\alpha,\lambda}}\left[U_\lambda\left(\sum_{t=0}^{T_F-1}[C(X_t, A_t) - g_{\alpha,\lambda}(X_t)] + \alpha[V_{\alpha,\lambda}(X_{T_F}) - V_{\alpha,\lambda}(z)]\right)\right], \end{aligned}$$

that is,

$$\begin{aligned} &e^{\lambda\alpha[V_{\alpha,\lambda}(x)-V_{\alpha,\lambda}(z)]} \\ &\quad = E_x^{f_{\alpha,\lambda}}\left[e^{\lambda\sum_{t=0}^{T_F-1}[C(X_t,A_t)-g_{\alpha,\lambda}(X_t)]+\lambda\alpha[V_{\alpha,\lambda}(X_{T_F})-V_{\alpha,\lambda}(z)]}\right]. \end{aligned} \tag{5.18}$$

Combining this equality with (5.5) and (5.7) it follows that for every $x \in S \setminus F$

$$E_x^{f^z}\left[e^{-2|\lambda|\|C\|T_F-|\lambda|B^*}\right] \le e^{\lambda\alpha[V_{\alpha,\lambda}(x)-V_{\alpha,\lambda}(z)]} \le E_x^{f^z}\left[e^{2|\lambda|\|C\|T_F+|\lambda|B^*}\right],$$

and via Lemma 5.1(i) this leads to

$$e^{-|\lambda|B/\delta}e^{-|\lambda|B^*} \le e^{\lambda\alpha[V_{\alpha,\lambda}(x)-V_{\alpha,\lambda}(z)]} \le e^{|\lambda|B/\delta}e^{|\lambda|B^*}.$$

Finally, using (5.17), this relation implies that $\alpha|V_{\alpha,\lambda}(x)-V_{\alpha,\lambda}(z)| \le B^*+B/\delta = 2B^*$ for every $x \in S \setminus F$, completing the verification of the inequality (5.8). □

6 Proof of the uniform boundedness theorem

In this section a proof of Theorem 4.1 will be presented. Throughout the remainder

$$z \in F \quad \text{and} \quad \lambda \in (-\delta, \delta) \setminus \{0\}$$

are *arbitrary but fixed*, and for each $\alpha \in (0, 1)$ the function $h_{\alpha,\lambda} \in \mathcal{B}(S)$ is given by

$$h_{\alpha,\lambda}(x) = \alpha[V_{\alpha,\lambda}(x) - V_{\alpha,\lambda}(z)], \quad x \in S. \tag{6.1}$$

Observe the following relation between $h_{\alpha,\lambda}$ and $g_{\alpha,\lambda}$ in (5.5):

$$g_{\alpha,\lambda}(x) - g_{\alpha,\lambda}(z) = \frac{1-\alpha}{\alpha} h_{\alpha,\lambda}(x), \tag{6.2}$$

and note that (5.4), (5.5) and Theorem 5.1 imply that

$$|g_{\alpha,\lambda}(z)| \le \|C\| \quad \text{and} \quad |h_{\alpha,\lambda}(x)| \le B, \quad x \in S, \quad \alpha \in (0, 1), \tag{6.3}$$

where

$$B := 2B^*.$$

From this point, Cantor's diagonal method yields that there exists a sequence $(\alpha_n) \subset (0, 1)$ such that

$$\alpha_n \nearrow 1, \tag{6.4}$$

and the following limits exists:

$$\begin{aligned} &\lim_{n\to\infty} g_{\alpha_n,\lambda}(z) = g^* \in [-\|C\|, \|C\|], \\ &\lim_{n\to\infty} h_{\alpha_n,\lambda}(x) = h^*_\lambda(x) \in [-B, B], \quad x \in S; \end{aligned} \tag{6.5}$$

combining these two last displays with (6.2) it follows that

$$\lim_{n\to\infty} g_{\alpha_n,\lambda}(x) = g^*, \quad x \in S. \tag{6.6}$$

Now, let $f_{\alpha,\lambda} \in \mathbb{F}$ be the stationary policy in (5.6). Since $\mathbb{F}$ is a compact metric space, taking a subsequence of (α_n), if necessary, without loss of generality it can be assumed $(f_{\alpha_n,\lambda})$ converges in $\mathbb{F}$:

$$\lim_{n\to\infty} f_{\alpha_n,\lambda}(x) = f^*_\lambda(x) \in A(x), \quad x \in S. \tag{6.7}$$

Proof of Theorem 4.1. It will be shown that the pair $(g^*, h^*_\lambda(\cdot))$ in (6.5) is a solution of the λ-optimality Eq. (4.1). To begin with, multiply both sides of (5.3) by $e^{-\lambda[\alpha V_{\alpha,\lambda}(z)+(1-\alpha)V_{\alpha,\lambda}(x)]}$ to obtain, via (2.2), (5.5) and (6.1), that for every $x \in S$

$$U_\lambda(h_{\alpha,\lambda}(x)) = \inf_{a\in A(x)} \left[\sum_{y\in S} p_{xy}(a) U_\lambda(C(x,a) - g_{\alpha,\lambda}(x) + h_{\alpha,\lambda}(y))\right], \tag{6.8}$$

and

$$U_\lambda(h_{\alpha,\lambda}(x)) = \sum_{y\in S} p_{xy}(f_{\alpha,\lambda}(x)) U_\lambda(C(x, f_{\alpha,\lambda}(x)) - g_{\alpha,\lambda}(x) + h_{\alpha,\lambda}(y)), \tag{6.9}$$

where $f_{\alpha,\lambda}$ is the stationary policy in (5.6). Also observe that (2.1), (5.5) and (6.3) together yield the following bound:

$$\begin{aligned}&|U_\lambda(C(x,a) - g_{\alpha,\lambda}(x) + h_{\alpha,\lambda}(y))| \\ &\quad \le e^{|\lambda|(|C(x,a)|+|g_{\alpha,\lambda}(x)|+|h_{\alpha,\lambda}(y)|)} \le e^{|\lambda|(2\|C\|+B)}, \quad (x,a) \in \mathbb{K}.\end{aligned} \tag{6.10}$$

Now let $(x,a) \in \mathbb{K}$ be fixed. Replacing α by α_n in (6.8) it follows that

$$U_\lambda(h_{\alpha_n,\lambda}(x)) \le \sum_{y\in S} p_{xy}(a) U_\lambda(C(x,a) - g_{\alpha_n,\lambda}(x) + h_{\alpha_n,\lambda}(y)),$$

and then, taking the limit as n goes to ∞ in both sides of this inequality, the bounded convergence theorem, (6.5) and (6.10) together imply that

$$U_\lambda(h^*_\lambda(x)) \le \sum_{y\in S} p_{xy}(a) U_\lambda(C(x,a) - g^*_\lambda + h^*_\lambda(y)), \quad (x,a) \in \mathbb{K}. \tag{6.11}$$

Select now a finite set $G \subset S$ and note that (6.9) and (6.10) yield that

$$\begin{aligned}U_\lambda(h_{\alpha_n,\lambda}(x)) &\ge \sum_{y\in G} p_{xy}(f_{\alpha_n,\lambda}(x)) U_\lambda(C(x, f_{\alpha_n,\lambda}(x)) - g_{\alpha_n,\lambda}(x) + h_{\alpha_n,\lambda}(y)) \\ &\quad - \sum_{y\in S\setminus G} p_{xy}(f_{\alpha_n,\lambda}(x)) e^{|\lambda|(2\|C\|+B)} \\ &= \sum_{y\in G} p_{xy}(f_{\alpha_n,\lambda}(x)) U_\lambda(C(x, f_{\alpha_n,\lambda}(x)) - g_{\alpha_n,\lambda}(x) + h_{\alpha_n,\lambda}(y)) \\ &\quad - \left(1 - \sum_{y\in G} p_{xy}(f_{\alpha_n,\lambda}(x))\right) e^{|\lambda|(2\|C\|+B)},\end{aligned}$$

and then, letting n go to ∞, (6.5)–(6.7) and Assumption 2.1 together imply that

$$\begin{aligned} U_\lambda(h_\lambda^*(x)) \geq & \sum_{y\in G} p_{xy}(f_\lambda^*(x)) U_\lambda(C(x, f_\lambda^*(x)) - g_\lambda^*(x) + h_\lambda^*(y)) \\ & - \left(1 - \sum_{y\in G} p_{xy}(f_\lambda^*(x))\right) e^{|\lambda|(2\|C\|+B)}; \end{aligned}$$

letting G increase to S this relation leads to

$$U_\lambda(h_\lambda^*(x)) \geq \sum_{y\in S} p_{xy}(f_\lambda^*(x)) U_\lambda(C(x, f_\lambda^*(x)) - g_\lambda^*(x) + h_\lambda^*(y)).$$

Combining this inequality with (6.11) it follows that the pair $(g_\lambda^*, h_\lambda^*(\cdot))$ satisfies the λ-optimality equation (4.1). Since $\|h_\lambda^*\| \leq B = 2B^*$ and $\lambda \in (-\delta, \delta) \setminus \{0\}$ is arbitrary, this establishes the conclusion of Theorem 4.1. □

7 Proof of the continuity result

After the previous preliminaries, in this section Theorem 3.1 will be established. Throughout the remainder δ and B are the positive numbers in Theorem 4.1, and for each $\lambda \in (-\delta, \delta) \setminus \{0\}$ the pair $(g_\lambda, h_\lambda(\cdot)) \in \mathbb{R} \times \mathcal{B}(S)$ is a solution of the optimality equation (4.1) satisfying that

$$\|h_\lambda\| \leq B; \tag{7.1}$$

note that Lemma 4.1 yields that g_λ is the optimal average cost at every initial state x, so that

$$|g_\lambda| \leq \|C\|. \tag{7.2}$$

Next, recalling the $U_\lambda(x) = \operatorname{sign}(\lambda) e^{\lambda x}$ for every $x \in \mathbb{R}$, divide both sides of (4.1) by $\operatorname{sign}(\lambda)\lambda = |\lambda| > 0$ and substract $1/\lambda$ from both sides of the resulting equality to obtain

$$\begin{aligned} & \frac{e^{\lambda(g_\lambda + h_\lambda(x))} - 1}{\lambda} \\ & = \inf_{a\in A(x)} \sum_{y\in S} p_{xy}(a) \frac{e^{\lambda(C(x,a)+h_\lambda(y))} - 1}{\lambda}, \quad x \in S, \quad 0 < |\lambda| < \delta, \end{aligned} \tag{7.3}$$

and observe that (7.1) and Assumption 2.1 together yield that there exists

$$f_\lambda \in \mathbb{F} \tag{7.4}$$

such that, for each $x \in S$, $f_\lambda(x) \in A(x)$ minimizes the right hand side of (7.3), that is,

$$\frac{e^{\lambda(g_\lambda + h_\lambda(x))} - 1}{\lambda} \\ = \sum_{y \in S} p_{xy}(f_\lambda(x)) \frac{e^{\lambda(C(x, f_\lambda(x)) + h_\lambda(y))} - 1}{\lambda}, \quad x \in S, \quad 0 < |\lambda| < \delta. \tag{7.5}$$

Finally, it is convenient to point out the following fact, which follows from the second order Taylor's expansion of the exponential function:

$$\lim_{\lambda \to 0} \Delta(\lambda) = 0, \quad \text{where} \quad \Delta(\lambda) := \sup_{x:\, |x| \le \|C\| + B} \left| \frac{e^{\lambda x} - 1}{\lambda} - x \right|. \tag{7.6}$$

With this notation, (7.1) yields that

$$\left| \frac{e^{\lambda(C(x,a) + h_\lambda(y))} - 1}{\lambda} - [C(x, a) + h_\lambda(y)] \right| \le \Delta(\lambda).$$

Proof of Theorem 3.1. It will be proved that

$$\lim_{\lambda \to 0} g_\lambda = J^*(0, x), \quad x \in S. \tag{7.7}$$

To achieve this goal, let $\{\lambda_n\}_{n \in \mathbb{N}}$ be an arbitrary sequence such that

$$\lambda_n \in (-\delta, \delta) \setminus \{0\}, \quad n = 0, 1, 2, \ldots, \quad \text{and} \quad \lim_{n \to \infty} \lambda_n = 0. \tag{7.8}$$

Recalling that the set $\mathbb{F}$ of stationary policies is a compact metric space, (7.4), (7.2) and (7.1) allow to use Cantor's diagonal method to construct a subsequence $\{\lambda_{n_k}\}_{k \in \mathbb{N}}$ such that the following limits exist:

$$\lim_{k \to \infty} g_{\lambda_{n_k}} =: g_0^*, \quad \lim_{k \to \infty} h_{\lambda_{n_k}}(x) =: h_0^*(x), \\ \lim_{k \to \infty} f_{\lambda_{n_k}}(x) =: f_0^*(x), \quad x \in S. \tag{7.9}$$

Now, let $(x, a) \in \mathbb{K}$ be arbitrary and note that (7.3) yields that

$$\frac{e^{\lambda_{n_k}(g_{\lambda_{n_k}} + h_{\lambda_{n_k}}(x))} - 1}{\lambda_{n_k}} \le \sum_{y \in S} p_{xy}(a) \frac{e^{\lambda_{n_k}(C(x,a) + h_{\lambda_{n_k}}(y))} - 1}{\lambda_{n_k}},$$

an inequality that via (7.1) and (7.6) leads to

$$g_{\lambda_{n_k}} + h_{\lambda_{n_k}}(x) - \Delta(\lambda_{n_k}) \le \sum_{y \in S} p_{xy}(a)(C(x, a) + h_{\lambda_{n_k}}(y)) + \Delta(\lambda_{n_k}).$$

Thus, since $\Delta(\lambda_{n_k}) \to 0$ as $k \to \infty$, taking the limit as k goes to ∞ in the above inequality, (7.9) and the bounded convergence theorem together imply that

$$g_0^* + h_0^*(x) \leq \sum_{y \in S} p_{xy}(a)(C(x,a) + h_0^*(y)), \quad (x,a) \in \mathbb{K}. \tag{7.10}$$

Next, let $x \in S$ be arbitrary and observe that (7.5) and (7.6) lead to

$$\begin{aligned} &g_{\lambda_{n_k}} + h_{\lambda_{n_k}}(x) + \Delta(\lambda_{n_k}) \\ &\quad \geq \frac{e^{\lambda_{n_k}(g_{\lambda_{n_k}} + h_{\lambda_{n_k}}(x))} - 1}{\lambda_{n_k}} \\ &\quad \geq \sum_{y \in S} p_{xy}(f_{\lambda_{n_k}}(x)) \frac{e^{\lambda_{n_k}(C(x, f_{\lambda_{n_k}}(x)) + h_{\lambda_{n_k}}(y))} - 1}{\lambda_{n_k}} \\ &\quad \geq \sum_{y \in S} p_{xy}(f_{\lambda_{n_k}}(x))(C(x, f_{\lambda_{n_k}}(x)) + h_{\lambda_{n_k}}(y)) - \Delta(\lambda_{n_k}), \end{aligned}$$

and then, using (7.1), it follows that for any nonempty and finite set G of the state space S,

$$\begin{aligned} g_{\lambda_{n_k}} + h_{\lambda_{n_k}}(x) \geq &\sum_{y \in G} p_{xy}(f_{\lambda_{n_k}}(x))(C(x, f_{\lambda_{n_k}}(x)) + h_{\lambda_{n_k}}(y)) \\ &-(\|C\| + B)\left[1 - \sum_{y \in G} p_{xy}(f_{\lambda_{n_k}}(x))\right] - 2\Delta(\lambda_{n_k}); \end{aligned}$$

letting k increase to ∞, Assumption 2.1 and (7.9) together yield that

$$\begin{aligned} g_0^* + h_0^*(x) \geq &\sum_{y \in G} p_{xy}(f_0^*(x))(C(x, f_0^*(x)) + h_0^*(y)) \\ &-(\|C\| + B)\left[1 - \sum_{y \in G} p_{xy}(f_0^*(x))\right], \end{aligned}$$

an inequality that letting G increase to S leads to

$$g_0^* + h_0^*(x) \geq \sum_{y \in S} p_{xy}(f_0^*(x))(C(x, f_0^*(x)) + h_0^*(y)), \quad x \in S. \tag{7.11}$$

Combining this relation with (7.10) it follows that

$$g_0^* + h_0^*(x) = \inf_{a \in A(x)} \left[C(x,a) + \sum_{y \in S} p_{xy}(a) h_0^*(y)\right], \quad x \in S, \tag{7.12}$$

showing that the pair $(g_0^*, h_0^*(\cdot))$ satisfies the (risk-neutral) optimality equation corresponding to the utility function $U_0(x) = x$; since h_0^* is a bounded function, it follows that

$$g_0^* = J^*(0, x), \quad x \in S.$$

Summarizing: It has been proved that every sequence $(\lambda_n)_{n\in\mathbb{N}}$ satisfying (7.8) has a subsequence $(\lambda_{n_k})_{k\in\mathbb{N}}$ such that $\lim_{k\to\infty} g_{\lambda_{n_k}} = J^*(0, \cdot)$, a property that is equivalent to (7.7); since $g_\lambda = J^*(\lambda, \cdot)$ when $0 < |\lambda| < \delta$, by Lemma 4.1 and Theorem 4.1, it follows that $\lim_{\lambda\to 0} J^*(\lambda, x) = J^*(0, x)$ for every $x \in S$. □

Remark 7.1 By (7.11), the policy f_0^* in the above proof satisfies that, for every state x, the action $f_0^*(x)$ is a minimizer of the right-hand side of the optimality equation in (7.12), and then f^* is risk-neutral average optimal (Hernández-Lerma (1989); Puterman (1994)).

8 A communicating model under LFC

In the previous sections, the continuity of the optimal risk-sensitive average cost function $J^*(\lambda, \cdot)$ with respect to λ has been studied. As it was shown in Proposition 2.1, for bounded costs such a property always holds at $\lambda \neq 0$, but the continuity at zero may fail, as in the simple model presented in Example 2.1. Under the version of the simultaneous Doeblin condition in Assumption 3.1, it was established in Theorem 3.1 that $J^*(\lambda, \cdot)$ is also continuous at $\lambda = 0$. Note that the first part of Assumption 3.1 ensures that, under any stationary policy, the class of recurrent states is non-empty, but there may be several recurrence classes; if such is the case, the second part of Assumption 3.1 guarantees that, employing *other* stationary policy, a given recurrence class is accessible from any other one, a condition that fails in Example 2.1. On the other hand, the simultaneous Doeblin condition is a strong requirement ensuring, via the existence of a bounded solution to the risk-neutral optimality equation, that (a) the optimal (risk-neutral) average cost is constant, and that (b) an optimal stationary exists, properties that can be guaranteed under the Lyapunov function condition (3.3), under which the risk-neutral optimality equation has a (generally) unbounded solution. However, it was shown in Example 3.1 that, under LFC, the mapping $\lambda \mapsto J^*(\lambda, x)$ may have a discontinuity at $\lambda = 0$. At this point it is interesting to observe a common feature in Examples 2.1 and 3.1, namely, in those examples the discontinuity at zero of the mapping $\lambda \mapsto J^*(\lambda, x)$ occurs *when x is a transient state*. This fact highlights the prominent role of the transient states in the determination of the risk-sensitive average cost (Cavazos-Cadena and Fernández-Gaucherand 1999) and, naturally, leads to consider the following question:

Assume that the Lyapunov function condition holds, and that the state space is a communicating class under each stationary policy, so that no state is transient. In this context, is it true that the mapping $\lambda \mapsto J^*(\lambda, x)$ is continuous at zero for every state x?

As it is shown by the following example, the answer to this question is negative.

Example 8.1 Consider a denumerable set S^* whose elements are denoted by x^*, where $x = 1, 2, 3, \ldots$, and define

$$S = \mathbb{N} \cup S^* = \mathbb{N} \cup \{x^* \,|\, x = 1, 2, 3, \ldots\}.$$

Now let the transition law $[p_{s,s_1}]$ on S be determined as follows:

$$\begin{aligned}&\text{For} x = 1, 2, 3, \ldots,\\ &1 - p_{x,0} = p_{x,x+1} = \frac{x^2}{(x+1)^2} = p_{x^*,(x+1)^*} = 1 - p_{x^*,0},\\ &p_{0,x^*} = \frac{\gamma}{x^3} = p_{0,x},\end{aligned} \tag{8.1}$$

where $\gamma^{-1} = \sum_{x=1}^{\infty} 2/x^3$. Finally, define the cost function $C : S \to \mathbb{R}$ by

$$C(x) = 1, \quad C(x^*) = -1, \quad x = 1, 2, 3, \ldots, \quad \text{and} \quad C(0) = 0; \tag{8.2}$$

setting the action set A equal to a singleton, the above quantities determine an MDP with a unique (stationary) policy, which will not be explicitly indicated in the analysis below.

In the next proposition it is shown that the Lyapunov function condition is satisfied in this example, and the risk-neutral average cost is determined.

Proposition 8.1 *In Example 8.1 the following assertions (i)–(iii) hold:*

(i) The Lyapunov function condition (3.3) is satisfied with $z = 0$.
(ii) The state process $(X_t)_{t\in\mathbb{N}}$ is irreducible, that is, $P_w[T_y < \infty] > 0$ for every $w, y \in S$.
(iii) The risk neutral average cost function is null: $J(0, y) = 0$ for every $y \in S$.

Proof (i) Since there is only one stationary policy, it is sufficient to verify that

$$E_y[T_0] < \infty, \quad y \in S. \tag{8.3}$$

To establish this assertion note that, paralleling the argument used in Example 3.1, the definition of the transition law in (8.1) yields that for every positive integer n

$$\begin{aligned}P_x[T_0 > n] &= P_x[X_r = x + r,\ 0 \le r \le n]\\ &= \frac{x^2}{(x+n)^2}\\ &= P_{x^*}[X_r = (x+r)^*,\ 0 \le r \le n]\\ &= P_{x^*}[T_0 > n], \quad x = 1, 2, 3, \ldots,\end{aligned} \tag{8.4}$$

and then $E_{x^*}[T_0] = E_x[T_0] \le \sum_{n=0}^{\infty} x^2/(x+n)^2 \le 1 + x$ for $x = 1, 2, 3, \ldots$ whereas, by the Markov property,

$$E_0[T_0] = 1 + \sum_{x=1}^{\infty} \left(p_{0,x} E_x[T_0] + p_{0,x^*} E_{x^*}[T_0]\right) \le 1 + \gamma \sum_{x=1}^{\infty} [2(1+x)/x^3] < \infty.$$

(ii) Observing that $P_0[X_1 = y] > 0$ for every $y \in S \setminus \{0\}$, the irreducibility of the state process follows from (8.3).

(iii) Combining the previous part and (8.3) it follows that the transition law in (8.1) has a unique invariant distribution $(\nu_y)_{y \in S}$ and then the ergodic theorem yields that, for every $y \in S$,

$$J(0, y) = \lim_{n\to\infty} \frac{1}{n} E_y \left[\sum_{t=0}^{n-1} C(X_t)\right] = \sum_{x\in S} \nu_x C(x).$$

On the other hand, using (8.1) it is not difficult to see that $\nu_x = \nu_{x^*}$ for every $x = 1, 2, 3, \ldots$, a property that together with (8.2) leads to $J(0, \cdot) = 0$. □

Next, the average cost function corresponding to a non-null risk sensitivity parameter will be determined.

Proposition 8.2 *For the model in Example 8.1 the following properties (i)–(iii) are valid.*

(i) For each $\lambda \neq 0$, the λ-sensitive average cost $J(\lambda, \cdot)$ is constant
(ii) If $\lambda > 0$ then $J(\lambda, \cdot) = 1$, and
(iii) $J(\lambda, \cdot) = -1$ for each $\lambda < 0$.

Proof (i) Let $\lambda \in \mathbb{R} \setminus \{0\}$ and $w, y \in S$ be arbitrary. Using Proposition 8.1(ii), select $k \in \mathbb{N}$ such that $P_w[X_k = y] > 0$ and note that, for $n > k$ the Markov property and (8.2) yield that

$$\begin{aligned} e^{\lambda J_n(\lambda, w)} &= E_w\left[e^{\lambda \sum_{t=0}^{n-1} C(X_t)}\right] \\ &\ge E_w\left[e^{\lambda \sum_{t=0}^{k-1} C(X_t)} I[X_k = y] e^{\lambda \sum_{t=k}^{n-1} C(X_t)}\right] \\ &\ge e^{-k|\lambda|} P_w[X_k = y] E_y\left[e^{\lambda \sum_{t=0}^{n-k-1} C(X_t)}\right] \\ &= e^{-k|\lambda|} P_w[X_k = y] e^{\lambda J_{n-k}(\lambda, y)} \end{aligned}$$

and then

$$\lambda \frac{J_n(\lambda, w)}{n} \ge \frac{\log(e^{-k|\lambda|} P_w[X_k = y])}{n} + \lambda \frac{J_{n-k}(\lambda, y)}{n} \tag{8.5}$$

If $\lambda > 0$, taking the superior limit as n goes to ∞ in both side of this inequality, it follows that

$$\lambda J(\lambda, w) \ge \lambda J(\lambda, y); \tag{8.6}$$

when $\lambda < 0$, this last relation also follows form (8.5) after taking the inferior limit as n increases to ∞. Thus, since the states w and y are arbitrary, (8.6) yields that $J(\lambda, \cdot)$ is constant for every no-null risk-sensitivity coefficient λ.

(ii) Suppose that $\lambda > 0$. Using that $C(x) = 1$ for every $x = 1, 2, 3, \ldots$, it follows that for every positive integer n

$$\begin{aligned} e^{\lambda J_n(\lambda,1)} &= E_1\left[e^{\lambda \sum_{t=0}^{n-1} C(X_t)}\right] \\ &\geq E_1\left[e^{\lambda \sum_{t=0}^{n-1} C(X_t)} I[X_r = 1 + r, 0 \leq r < n]\right] \\ &= e^{\lambda n} P_1\left[X_r = 1 + r,\ 0 \leq r < n\right] = e^{\lambda n}\frac{1}{n^2}, \end{aligned}$$

where (8.4) was used to set the last equality. Thus,

$$J_n(\lambda, 1)/n \geq 1 - 2\log(n)/(\lambda n),$$

and then $J(\lambda, 1) = \limsup_{n\to\infty} J_n(\lambda, 1)/n \geq 1$; recalling the $C(\cdot) \leq 1$ it follows that $J(\lambda, 1) = 1$ so that $J(\lambda, \cdot) = 1$, by part (i).

(iii) Suppose that $\lambda < 0$. Recall that $C(x^*) = -1$ for every $x = 1, 2, 3, \ldots$ and observe that for each positive integer n

$$\begin{aligned} e^{\lambda J_n(\lambda,1^*)} &= E_{1^*}\left[e^{\lambda \sum_{t=0}^{n-1} C(X_t)}\right] \\ &\geq E_{1^*}\left[e^{\lambda \sum_{t=0}^{n-1} C(X_t)} I[X_r = (x + r)^*, 0 \leq r < n]\right] \\ &= e^{-\lambda n} P_{1^*}\left[X_r = (x + r)^*, 0 \leq r < n\right] = e^{-\lambda n}\frac{1}{n^2}; \end{aligned}$$

since λ is negative, this relation yields that $J_n(\lambda, 1^*)/n \leq -1 + 2\log(n)/(n\lambda)$, a relation that combined with the inequality $C(\cdot) \geq -1$ leads to $J(\lambda, x) = \limsup_{n\to\infty} J_n(\lambda, x)/n = -1$; from this point, part (i) implies that $J(\lambda, \cdot) = -1$. □

For the model in Example 8.1, the Lyapunov function condition holds and every state is positive recurrent, by parts (i) and (ii) of Proposition 8.1. However, assertions (ii) and (iii) of Proposition 8.2 and the third part of Proposition 8.1 together yield that, for every state y, the mapping $\lambda \mapsto J(\lambda, y)$ is not continuous at $\lambda = 0$, neither form the left nor from the right.

Acknowledgments The authors are grateful to the reviewers for their careful reading of the original manuscript and their helpful suggestions to improve the paper.

References

Arapostathis 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

Bäuerle N, Rieder U (2011) Markov decision processes with applications to finance. Springer, New York

Bäuerle N, Rieder U (2013) More risk-sensitive Markov decision processes. Math Oper Res 39:105–120

Cavazos-Cadena R (2003) Solution to the risk-sensitive average cost optimality equation in a class of Markov decision processes with finite state space. Math Method 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 Method 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-Lerma O (1992) Equivalence of Lyapunov stability criteria in a class of Markov decision processes. Appl Math Optim 26:113–137

Cavazos-Cadena R, Hernández-Hernández D (2015) A Characterization of the Optimal Certainty Equivalent of the Average Cost via the Arrow-Pratt Sensitivity Function. Math Oper Res (to appear)

Di Masi GB, Stettner L (1999) Risk-sensitive control of discrete time Markov processes with infinite horizon. SIAM J Control Optim 38:61–78

Di Masi GB, Stettner L (2000) Infinite horizon risk sensitive control of discrete time Markov processes with small risk. Syst Control Lett 40:15–20

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

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

Hernández-Lerma O (1989) Adaptive Markov control processes. Springer, New York

Hordijk A (1974) Dynamic programming and Markov potential theory, Mathematical Centre Tracts 51, Mathematisch Centrum, Amsterdam

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

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

Stokey NL, Lucas RE (1989) Recursive methods in economic dynamics. Harvard University Press, Cambridge

Sladký K (2008) Growth rates and average optimality in risk-sensitive Markov decision chains. Kybernetika 44:205–226