Math Meth Oper Res (2010) 71:47–84
DOI 10.1007/s00186-009-0285-6

ORIGINAL ARTICLE

Optimality equations and inequalities in a class of risk-sensitive average cost Markov decision chains

Rolando Cavazos-Cadena

Received: 24 July 2008 / Accepted: 26 January 2009 / Published online: 18 February 2009

Abstract This note concerns controlled Markov chains on a denumerable sate space. The performance of a control policy is measured by the risk-sensitive average criterion, and it is assumed that (a) the simultaneous Doeblin condition holds, and (b) the system is communicating under the action of each stationary policy. If the cost function is bounded below, it is established that the optimal average cost is characterized by an optimality inequality, and it is to shown that, even for bounded costs, such an inequality may be strict at every state. Also, for a nonnegative cost function with compact support, the existence an uniqueness of bounded solutions of the optimality equation is proved, and an example is provided to show that such a conclusion generally fails when the cost is negative at some state.

Keywords First arrival time · Stopping problem with total cost index · Relative value function · Constant average cost · Stochastic matrix associated with a multiplicative Poisson equation

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

Dedicated to Professor Onésimo Hernández-Lerma, on the occasion of his 60th birthday.

This work was supported by the PSF Organisation under Grant No. 08-05(450), 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

1 Introduction

This work concerns Markov decision chains with denumerable state space and compact action sets. The performance of a control policy is measured by the risk-sensitive average cost criterion associated with a constant risk-sensitivity coefficient $\lambda > 0$, and the main objective of the paper is to provide conditions under which

(a) the (λ-sensitive) optimal average cost function is constant, and
(b) there exists a solution of an optimality equation or *inequality*, from which an optimal stationary policy can be obtained.

Besides standard continuity assumptions, the framework under which these problems are analyzed is determined by the following two requirements:

(i) The simultaneous Doeblin condition holds (Thomas 1980; Hernández-Lerma 1988), and
(ii) The whole state space is a communicating class under the action of each stationary policy.

In this context, it will be shown that the optimal average cost function is constant, say g^*, and the main results of the paper are expressed in terms of the corresponding relative value function $h_{g^*}(\cdot)$, which will be formally introduced in Sect. 3, and whose intuitive meaning can be described as follows: first, let $C(\cdot)$ be the one-step cost structure, and define the relative cost function as $C(\cdot) - g^*$. With this notation, assuming that x is the initial state of the system, $h_{g^*}(x)$ is the λ-sensitive measure of the minimum total relative cost incurred before the first visit to a fixed state z. In terms of this idea, the main results of the paper, stated as Theorems 3.1 and 3.2 in Sect. 3, can be described as follows:

(1) If the cost function is nonnegative and has compact support, then the pair $(g^*, h_{g^*}(\cdot))$ is a solution of the λ-sensitive optimality equation, and
(2) If the cost function is bounded below—or, without loss of generality, a nonnegative function—then the pair $(g^*, h_{g^*}(\cdot))$ is a solution of a λ-sensitive optimality *inequality*.

In both cases, λ-optimal stationary policies can be obtained from the equation or inequality in a standard way. On the other hand, explicit examples are given to show that

(E1) If a cost function has compact support but takes a negative value, then the pair $(g^*.h_{g^*}(\cdot))$ may not satisfy the optimality equation at any state, and
(E2) If $C(\cdot)$ is a general nonnegative function the optimality inequality may be strict at every state, even if the cost function is bounded.

The analysis of 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 Markov decision chains were studied under the communication assumption (ii) described above and, using the Perron-Frobenius theory of positive matrices (Seneta 1980), the existence of solutions to the λ-sensitive optimality equation was established. 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), Borkar and Meyn (2002), Jaśkiewicz (2007) and the references there in.

The motivation to study the solvability of the λ-sensitive average optimality equation (or inequality) under the assumptions (i) and (ii) above stems from the following remarks. First, the risk neutral-average index has been successfully analyzed under (diverse variants of) the simultaneous Doeblin condition (SDC) ensuring the existence of an appropriate solution of the corresponding optimality equation, which yields that the optimal average cost function is constant, as well as an optimal stationary policy (Thomas 1980; Hernández-Lerma 1988; Arapostathis et al. 1993; Puterman 1994); moreover, under mild requirements, the SDC is also necessary to have that, for each bounded cost function, a bounded solution of the risk-neutral optimality equation exists (Cavazos-Cadena 1988). Thus, at least at an initial stage, it seems natural to include the simultaneous Doeblin condition as a working assumption to analyze the risk-sensitive average index. However, in Cavazos-Cadena and Fernández-Gaucherand (1999), an example was given to show that, even in a finite model, such a condition does not guarantee that the λ-optimal average cost is constant, and in this case the optimality equation can not be solved; moreover, even when the risk-sensitive average cost function is constant, the solvability of the optimality equation is not guaranteed under the SDC (Cavazos-Cadena and Hernández-Hernández 2004). This behavior does not occur when the state space is communicating with respect to each stationary policy (Howard and Matheson 1972) so that it is also natural to include assumption (ii) above within the basic framework.

On the other hand, the result (1) described above is an extension of the main theorem in Howard and Matheson (1972), in that the existence of bounded solutions of the λ-optimality equation is established in a model with denumerable state space, while the result (2) is related to more recent work by Hernández-Hernández and Marcus (1999) and Jaśkiewicz (2007), where the existence of a solution to an optimality inequality was established. The results in these two papers were obtained implementing the discounted approach in a similar way to that used by Sennot (1986, 1995) in the risk-neutral context. Thus, a family of optimal value functions corresponding to a risk-sensitive discounted criterion was considered, and conditions were imposed on the behavior of the family as the discount factor increases to 1. The required conditions can be ensured if the cost function has a 'penalized' structure, in the sense that it takes 'large values' outside appropriate compact sets (Borkar and Meyn 2002). In contrast, although the present work assumes that the state space is denumerable, no condition is imposed on any derived quantity, and the result (2) does not require any special structure on the cost function beyond the nonnegativity or, more generally, the existence of a lower bound.

The approach used in the paper relies on the analysis of stopping time problems endowed with the risk-sensitive total cost criterion, and extends ideas in Cavazos-Cadena and Fernández-Gaucherand (2002), were finite models were studied; indeed, each relative value function considered below is the optimal index associated with one stopping time problem. On the other hand, *the key technical instrument* in this note is an auxiliary probability matrix associated with a solution to the risk-sensitive (multiplicative) Poisson equation corresponding to a stationary policy. Such a matrix

is introduced in Sect. 5 and its properties allow to obtain the results concerning cost functions with compact support and, from that point, the conclusions for the general case are established via an approximation process.

The organization of the paper is as follows: first, in Sect. 2 the decision model is formally described and the λ-sensitive average criterion, as well as the corresponding optimality equation, are briefly discussed. Next, in Sect. 3 the idea of relative value function is introduced and, after proving some of its basic properties, the main results of the paper are stated as Theorems 3.1 and 3.2, which establish the results (1) and (2) described above, respectively. The proofs of these theorems rely on the properties of the relative value functions presented in the following two sections. Thus, Sect. 4 concerns general properties, which are valid when the cost function $C(\cdot)$ is nonnegative, while the results in Sect. 5 hold under the additional condition that $C(\cdot)$ has compact support. After these preliminaries, the main results are finally proved in Sect. 6, and the exposition concludes in Sect. 7 with an explicit example illustrating the facts (E1) and (E2) discussed above.

Notation Throughout the remainder $\mathbb{N}$ stands for the set of nonnegative integers whereas, for a topological space $\mathbb{K}$, $\mathcal{B}(\mathbb{K})$ is the space of all real-valued and bounded functions defined on $\mathbb{K}$, that is, $C\colon \mathbb{K} \to \mathbb{R}$ belongs to $\mathcal{B}(\mathbb{K})$ if and only its supremum norm $\|C\|$ is finite, where $\|C\| := \sup_{x\in\mathbb{K}} |C(x)|$. On the other hand, for an event A 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

Let $\mathcal{M} = (S, A, \{A(x)\}_{x\in S}, C, P)$ be a Markov decision process (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 $x \in S$, $A(x) \subset A$ is the nonempty and compact set of admissible actions at x; the class $\mathbb{K}$ of admissible pairs is given by $\mathbb{K} = \{(x, a) \,|\, a \in A(x), x \in S\} \subset S \times A$. On the other hand, $C\colon S \to \mathbb{R}$ 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, regardless of the states observed and actions applied before time t, the state of the system at time $t + 1$ will be $X_{t+1} = y \in S$ with probability $p_{x\,y}(a)$, where $\sum_{y\in S} p_{x\,y}(a) = 1$; this is the Markov property of the process.

Assumption 2.1 (i) For each $(x, a) \in \mathbb{K}$, $C(x, a) \geq 0$;
(ii) For each $x, y \in S$, the mappings $a \mapsto C(x, a)$ and $a \mapsto p_{x\,y}(a)$ are continuous in $a \in A(x)$.

Policies. For each $t \in \mathbb{N}$ the space $\mathbb{H}_t$ of histories up to time t is recursively determined by $\mathbb{H}_0 := S$ and $\mathbb{H}_t = \mathbb{K} \times \mathbb{H}_{t-1}$, $t = 1, 2, 3, \ldots$; a generic element of $\mathbb{H}_t$ is denoted by $\mathbf{h}_t = (x_0, a_0, \ldots, x_{t-1}, a_{t-1}, x_t)$ where $x_i \in S$ and $a_i \in A(x_i)$. A policy is a sequence $\pi = \{\pi_t\}$, where each π_t is a special stochastic kernel on A given $\mathbb{H}_t$, that is,

for each $\mathbf{h}_t \in \mathbb{H}_t$, $\pi_t(\cdot|\mathbf{h}_t)$ is a probability measure on the Borel class of A satisfying $\pi_t(A(x_t)|\mathbf{h}_t) = 1$, and for each Borel subset $B \subset A$, the mapping $\mathbf{h}_t \mapsto \pi_t(B|\mathbf{h}_t)$ is measurable; the number $\pi_t(B|\mathbf{h}_t)$ is the probability of choosing action A_t within the set B when the system is driven by π and $\mathbf{h}_t$ is the observed history of the process up to time t; the class of all policies is denoted by $\mathcal{P}$. Given the initial state $x \in S$ and the policy $\pi \in \mathcal{P}$ being used for choosing actions, the distribution of the state-action process $\{(X_t, A_t)\}$ is uniquely determined (Hernández-Lerma 1988; Arapostathis et al. 1993; Puterman 1994) and 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 stationary if there exists $f \in \mathbb{F}$ such that, under π, the action selected at each time t is given by $A_t = f(X_t)$. The class of stationary policies is naturally identified with $\mathbb{F}$ and, under the action of each stationary policy, the state process $\{X_t\}$ is a Markov chain with stationary transition mechanism.

Average performance index. As already mentioned, it is supposed that the controller has constant risk sensitivity $\lambda > 0$, which means the the decision maker assesses a random cost Y using the expectation of $e^{\lambda Y}$; the number $\mathcal{E}(Y) := \log(E[e^{\lambda Y}]/\lambda$, which satisfies $e^{\lambda \mathcal{E}(Y)} = E[e^{\lambda Y}]$, is referred to as the certain equivalent of Y, and the decision maker is indifferent between incurring the random cost Y or paying the certain equivalent $\mathcal{E}(Y)$ for sure. Next, given $\pi \in \mathcal{P}$, $x \in S$ and a positive integer n, let $J_{C,n}(\pi, x)$ be the certain equivalent of the total cost $\sum_{t=0}^{n-1} C(X_t, A_t)$ incurred before time n when the system is driven by π starting at $X_0 = x$, that is,

$$J_{C,n}(\pi, x) := \frac{1}{\lambda} \log \left(E_x^\pi \left[e^{\lambda \sum_{t=0}^{n-1} C(X_t, A_t)} \right] \right), \tag{2.1}$$

and define the (limit superior) λ-sensitive average cost at state x under policy π by

$$J_C(\pi, x) := \limsup_{n \to \infty} \frac{1}{n} J_{C,n}(\pi, x); \tag{2.2}$$

the λ-optimal average cost function $J_C^*(\cdot)$ is specified by

$$J_C^*(x) := \inf_{\pi \in \mathcal{P}} J_C(\pi, x), \quad x \in S, \tag{2.3}$$

and a policy $\pi^* \in \mathcal{P}$ is λ-optimal if $J_C(\pi^*, x) = J_C^*(x)$ for each $x \in S$. If the cost function is bounded, it follows from these definitions that $\|J_C^*(\cdot)\| \leq \|C\| < \infty$. To analyze the case of general nonnegative costs, the following condition is enforced.

Assumption 2.2 For some state $z \in S$, $J_C^*(z) < \infty$.

Remark 2.1 Replacing the limit superior by limit inferior in (2.1) the criterion $\tilde{J}_C(\pi, x) := \liminf_{n \to \infty} J_{C,n}(\pi, x)/n$ is obtained. The corresponding optimal value function is $\tilde{J}_C^*(x) := \inf_{\pi \in \mathcal{P}} \tilde{J}_C(\pi, x)$, so that $\tilde{J}_C^*(\cdot) \leq J_C^*(\cdot)$. Under the full set of conditions imposed in this work, the limit superior and limit inferior λ-sensitive average criteria render the same optimal value function.

Optimality Equation. The optimality equation associated with the λ-sensitive average cost criterion in (2.1)–(2.3) is

$$e^{\lambda g+\lambda h(x)} = \inf_{a\in A(x)} \left[e^{\lambda C(x,a)} \sum_{y\in S} p_{x\,y}(a) e^{\lambda h(y)} \right], \quad x \in S, \tag{2.4}$$

where g is a real number and $h\colon S \to \mathbb{R}$ is a given function.

The following verification criterion was established in Hernández-Hernández and Marcus (1996).

Lemma 2.1 *Given $\lambda \in [0,\infty)$ suppose that the pair $(g, h(\cdot)) \in \mathbb{R} \times \mathcal{B}(S)$ satisfies the λ-optimality Eq. (2.4). In this case, the assertions (i) and (ii) below hold:*

(i) The λ-optimal average cost function $J_C^*(\cdot)$ is constant and equal to g;

(ii) If the stationary policy f^* satisfies

$$e^{\lambda g+\lambda h(x)} = e^{\lambda C(x,f^*(x))} \sum_{y\in S} p_{x\,y}(f^*(x)) e^{\lambda h(y)}, \quad x \in S, \tag{2.5}$$

then f^* is λ-optimal and $g = \lim_{n\to\infty} \frac{1}{n} J_{C,n}(f^*, x)$, $x \in S$.

Remark 2.2 (i) Suppose that Assumption 2.1(ii) holds, let $x \in S$ be arbitrary but fixed, and let $h\colon S \to [-\infty,\infty]$ be a given function. In this context, Fatou's lemma yields that

$$a \mapsto e^{\lambda C(x,a)} \sum_{y\in S} p_{x\,y}(a) e^{\lambda h(y)}, \quad a \in A(x),$$

is a lower semi-continuous function taking values in $[0,\infty]$ and, since $A(x)$ is a compact space, this mapping has a minimizer $f^*(x) \in A(x)$. Therefore, the corresponding policy $f^* \in \mathbb{F}$ satisfies

$$\inf_{a\in A(x)} \left[e^{\lambda C(x,a)} \sum_{y\in S} p_{x\,y}(a) e^{\lambda h(y)} \right] = e^{\lambda C(x,f^*(x))} \sum_{y\in S} p_{x\,y}(f^*(x)) e^{\lambda h(y)}, \quad x \in S,$$

so that the infimum can be replaced by minimum. Also, in the context of Lemma 2.1, Assumption 2.1(ii) implies that a policy f^* satisfying (2.5) exists.

(ii) Suppose that (2.4) is satisfied, with $h(\cdot) \in \mathcal{B}(S)$. In this case an induction argument yields that the relation

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

always holds, and then $g \leq \liminf_{n\to\infty} J_{C,n}(\pi,\cdot)/n = \tilde{J}_C(\pi,\cdot)$; see Remark 2.1. Thus, $g \leq \tilde{J}_C^*(\cdot)$ and then, since $\tilde{J}_C^*(\cdot) \leq J_C^*(\cdot)$ and $J_C^*(\cdot) = g$—by Lemma 2.1—it follows that $\tilde{J}_C^*(\cdot) = J_C^*(\cdot) = g$, that is, the limit superior and the limit inferior average criterion render the same optimal value function.

(iii) Equation (2.5), to be solved for the pair $(g, h(\cdot))$, will be referred to as the *Poisson equation* associated with policy $f^* \in \mathbb{F}$.

Conditions on the transition law. As already mentioned, the main objective of the paper is to establish the existence of a solution $(g, h(\cdot))$ of the optimality Eq. (2.4), or a similar inequality, implying that (a) the optimal value function $J_C^*(\cdot)$ is constant and equal to g, and (b) a λ-optimal stationary policy f^* exists. These problems will be analyzed under the ergodicity and communication conditions stated below, whose formulations involve the following terminology.

Definition 2.1 (i) The hitting time corresponding to $F \subset S$ is given by

$$T_F := \min\{n \geq 1 \mid X_n \in F\}, \tag{2.6}$$

where the minimum of the empty set is ∞; if $F = \{x\}$ is a singleton, $T_x \equiv T_{\{x\}}$.

(ii) The simultaneous Doeblin condition at state x—briefly, SDC(x)—is specified as follows:

$$\begin{aligned}&\text{SDC}(x)\text{: there exist } \widehat{N}(x) \in \mathbb{N} \setminus \{0\} \text{ and } \widehat{\rho}(x) \in (0,1)\\ &\text{such that } P_y^f[T_x > \widehat{N}(x)] \leq \widehat{\rho}(x), \quad y \in S, \quad f \in \mathbb{F}.\end{aligned}$$

Assumption 2.3 For some $z_0 \in S$, SDC(z_0) holds.

Remark 2.3 Under Assumption 2.3, an induction argument using Definition 2.1(ii) yields that every $x \in S$, $f \in \mathbb{F}$ and $k \in \mathbb{N}$, $P_x^f[T_{z_0} > k\widehat{N}(z_0)] \leq \widehat{\rho}(z_0)^k$, so that $P_x^f[T_{z_0} > k] \leq \widetilde{M}(z_0)\widetilde{\rho}(z_0)^k$, where $\widetilde{M}(z_0) := 1/\widehat{\rho}(z_0)$ and $\widetilde{\rho}(z_0) := \widehat{\rho}(z_0)^{1/\widehat{N}(z_0)}$. Therefore, $E_{z_0}^f[T_{z_0}] = \sum_{k=0}^{\infty} P_x^f[T_{z_0} > k] \leq \widetilde{M}(z_0)/(1-\widetilde{\rho}(z_0)) < \infty$, so that z_0 is positive recurrent with respect to the Markov chain induced by each $f \in \mathbb{F}$.

Assumption 2.4 The state space S is communicating under the action of each stationary policy. More precisely, for each $f \in \mathbb{F}$ and $x, y \in S$, $P_x^f[T_y < \infty] > 0$.

This section concludes with the following consequence of the two last assumptions which will be useful later.

Lemma 2.2 *Under Assumptions* 2.1(*ii*), 2.3 *and* 2.4, *the following assertions* (*i*) *and* (*ii*) *below hold.*

(i) For each $y \in S$ there exists a positive integer $N(y)$ and $\rho(y) \in (0,1)$ such that

$$P_x^\pi[T_y > N(y)] \leq \rho(y), \quad x \in S, \quad \pi \in \mathcal{P}, \tag{2.7}$$

so that, for each $y \in S$, SDC(y) *holds; see Definition* 2.1(*ii*).

(ii) For each $y \in S$, $\sup_{x\in S,\,\pi\in\mathcal{P}} E_x^\pi[T_y] \le \rho(y)^{-1}/(1-\rho(y)^{1/N(y)}) < \infty$.

Proof First, using Assumption 2.3 select $z_0 \in S$ such that SDC(z_0) holds, so that, by Remark 2.3, for each $f \in \mathbb{F}$ the state z_0 is positive recurrent with respect to the Markov chain induced by f, which is communicating, by Assumption 2.4. It follows that for each $f \in \mathbb{F}$ there exists a unique probability distribution μ_f on the state space S such that

$$\mu_f(x) > 0 \quad \text{and} \quad \mu_f(x) = \sum_{v\in S} \mu_f(v)p_{v,x}(f(v)), \quad x \in S; \tag{2.8}$$

see, for instance, Loève (1980). Now, let $y \in S$ be arbitrary but fixed, and let the reward function $R_y \in \mathcal{B}(S)$ be given by

$$R_y(x) = 1, \quad x \in S \setminus \{y\}, \quad R_y(y) = 0. \tag{2.9}$$

From the theory of risk-neutral average criterion (Thomas 1980; Hernández-Lerma 1988; Puterman 1994), Assumption 2.3 yields that there exists $g_y \in \mathbb{R}$ and $h_y \in \mathcal{B}(S)$ such that the following optimality equation is satisfied:

$$g_y + h_y(x) = \sup_{a\in A(x)} \left[R_y(x) + \sum_{w\in S} p_{x\,w}(a)h_y(w) \right]. \tag{2.10}$$

Using that $h_y(\cdot)$ is bounded, Assumption 2.1(ii) and the bounded convergence theorem together yield that for each $x \in S$ the term within brackets in (2.10) is a continuous function of $a \in A(x)$; since the action sets are compact, it follows that there exists $f_y \in \mathbb{F}$ such that $g_y + h_y(x) = R_y(x) + \sum_{w\in S} p_{x\,w}(f_y(x))h_y(w)$ for each $x \in S$, a fact that via (2.8) and (2.9) leads to

$$g_y = \sum_{x\in S} \mu_{f_y}(x)R_y(x) = \sum_{x\in S,\,x\neq y} \mu_{f_y}(x) = 1 - \mu_{f_y}(y) < 1. \tag{2.11}$$

Now let $x \in S$, $\pi \in \mathcal{P}$ and $n \in \mathbb{N}$ be arbitrary, and notice that (2.10) and the Markov property together yield that the following relation holds P_x^π-almost surely:

$$h(X_n) \ge R_y(X_n) - g + \sum_y p_{X_n\,y}(A_n)h(y) = R_y(X_n) - g + E_x^\pi[h(X_{n+1})|X_s, A_s, s \le n],$$

so that $E_x^\pi[h(X_n)] \ge E_x^\pi[R_y(X_n) - g + h(X_{n+1})]$, an inequality that via an induction argument leads to

$$\begin{aligned} \|h_y\| \ge h_y(x) &\ge E_x^\pi \left[\sum_{t=0}^{n} (R_y(X_t) - g_y) + h_y(X_{n+1}) \right] \\ &\ge -g_y + E_x^\pi \left[\sum_{t=1}^{n} (R_y(X_t) - g_y) \right] - \|h_y\|, \end{aligned}$$

and then, if n is positive,

$$\frac{2\|h_y\| + g_y}{n} + g_y \geq \frac{1}{n} E_x^\pi \left[\sum_{t=1}^{n} R_y(X_t) \right] \geq \frac{1}{n} E_x^\pi \left[I[T_y > n] \sum_{t=1}^{n} R_y(X_t) \right].$$

Observing that $X_t \neq y$ if $1 \leq t \leq n < T_y$, by Definition 2.1(i), it follows that $\sum_{t=1}^{n} R_y(X_t) = n$ on $[T_y > n]$ [see 2.9)], and the above display yields that

$$\frac{2\|h_y\| + g_y}{n} + g_y \geq P_x^\pi [T_y > n], \quad x \in S, \quad \pi \in \mathcal{P}, \quad n = 1, 2, 3, \ldots$$

Selecting the positive integer $N(y)$ in such a way that $(2\|h_y\| + g_y)/N(y) < (1 - g_y)/2$ it follows that

$$P_x^\pi [T_y > N(y)] \leq \rho(y) := \frac{2\|h_y\| + g_y}{N(y)} + g_y < \frac{1 + g_y}{2} < 1, \quad x \in S, \quad \pi \in \mathcal{P},$$

where the last inequality is due to (2.11). This establishes part (i), and the second part can be proved paralleling the argument outlined in Remark 2.3. □

3 Main results

In this section the main results of this note are sated as Theorems 3.1 and 3.2 below. Throughout the remainder Assumptions 2.1–2.4 are supposed to be valid even without explicit reference and, to begin with, the idea of relative value function is introduced and some of its basic properties are established.

Definition 3.1 Let $z \in S$ be such that $J_C^*(z)$ is finite; such a state will be fixed throughout the remainder of the paper—see Assumption 2.2.

(i) Given $g \in \mathbb{R}$ the corresponding relative value function $h_g : S \to [-\infty, \infty]$ is defined by

$$h_g(x) := \frac{1}{\lambda} \inf_{\pi \in \mathcal{P}} \log \left(E_x^\pi \left[e^{\lambda \sum_{t=0}^{T_z - 1} (C(X_t, A_t) - g)} \right] \right), \quad x \in S, \tag{3.1}$$

where T_z is the hitting time in Definition 2.1(i).

(ii) The set G is given by

$$G := \{ g \in \mathbb{R} \mid h_g(z) \leq 0 \}. \tag{3.2}$$

Lemma 3.1 *The relative value function $h_g(\cdot)$ satisfies the following properties* (i)–(iii):

(i) *If $g > J_C^*(z)$ then $h_g(z) < \infty$, and*
(ii) *For each $x \in S$, the mapping $g \mapsto h_g(x)$ is decreasing and $h_g(z) \to -\infty$ as $g \to \infty$.*

Consequently,

(iii) *G is nonempty and $G \subset [0, \infty)$.*

Proof (i) Let $g > J_C^*(z)$ be arbitrary but fixed. Next, select $g_0 \in (J_C^*(z), g)$ and notice that (2.1)–(2.3) together yield that, for some policy $\pi^0 \in \mathcal{P}$, the following inequality holds:

$$\limsup_{n\to\infty} \frac{1}{n\lambda} \log\left(E_z^{\pi^0}\left[e^{\lambda\sum_{t=0}^{n-1} C(X_t,A_t)}\right]\right) < g_0.$$

Therefore, there exists a positive integer N_0 such that $E_z^{\pi^0}[e^{\lambda\sum_{t=0}^{n-1} C(X_t,A_t)}] < e^{n\lambda g_0}$ for $n \geq N_0$, a relation that, since $C(\cdot,\cdot)$ is nonnegative, leads to

$$E_z^{\pi^0}\left[e^{\lambda\sum_{t=0}^{n-1} C(X_t,A_t)}\right] \leq Me^{n\lambda g_0}, \quad n \in \mathbb{N}\setminus\{0\},$$

where $M := E_z^{\pi^0}[e^{\lambda\sum_{t=0}^{N_0-1} C(X_t,A_t)}] < \infty$. Next, using that T_z is finite $P_z^{\pi^0}$-almost surely, by Lemma 2.2(ii), notice that

$$\begin{aligned} E_z^{\pi^0}\left[e^{\lambda\sum_{t=0}^{T_z-1}(C(X_t,A_t)-g)}\right] &= \sum_{n=1}^{\infty} E_z^{\pi^0}\left[e^{\lambda\sum_{t=0}^{n-1}(C(X_t,A_t)-g)} I[T_z = n]\right] \\ &\leq \sum_{n=1}^{\infty} E_z^{\pi^0}\left[e^{\lambda\sum_{t=0}^{n-1}(C(X_t,A_t)-g)}\right] \\ &= \sum_{n=1}^{\infty} E_z^{\pi^0}\left[e^{\lambda\sum_{t=0}^{n-1} C(X_t,A_t)}\right] e^{-n\lambda g}, \end{aligned}$$

a fact that via the previous display leads to

$$E_z^{\pi^0}\left[e^{\lambda\sum_{t=0}^{T_z-1}(C(X_t,A_t)-g)}\right] \leq \sum_{n=1}^{\infty} Me^{n\lambda g_0}e^{-n\lambda g} = \sum_{n=1}^{\infty} Me^{-n\lambda(g-g_0)} < \infty;$$

recall that $g > g_0$ for the last inequality. From this point, Definition 3.1(i) yields that $h_g(z) < \infty$.

(ii) Let $x \in S$ and $g \in \mathbb{R}$ be arbitrary but fixed. Since T_z is always larger than or equal to 1, it follows that for each $\delta > 0$ the inequality $\sum_{t=0}^{T_z-1}(C(X_t, A_t) - [g+\delta]) \leq \sum_{t=0}^{T_z-1}(C(X_t, A_t) - g) - \delta$ always holds, so that

$$\begin{aligned} &E_x^{\pi}\left[e^{\lambda\sum_{t=0}^{T_z-1}(C(X_t,A_t)-[g+\delta])}\right] \\ &\leq E_x^{\pi}\left[e^{\lambda\sum_{t=0}^{T_z-1}(C(X_t,A_t)-g)}\right] e^{-\lambda\delta}, \quad \pi \in \mathcal{P}, \quad \delta > 0 \end{aligned}$$

and the specification of $h_g(\cdot)$ yields that

$$h_{g+\delta}(x) \le h_g(x) - \delta, \quad \delta > 0,$$

establishing that the mapping $g \mapsto h_g(x)$ is decreasing. Now, let $g_0 > J_C^*(z)$ be arbitrary, so that $h_{g_0}(z) < \infty$, by part (i). In this case, using the above display with z instead of x it follows that $\lim_{g\to\infty} h_g(z) = \lim_{\delta\to\infty} h_{g_0+\delta}(z) \le \lim_{\delta\to\infty}[h_{g_0}(z) - \delta] = -\infty$.

(iii) Let $g < 0$ be arbitrary. Since $T_z \ge 1$ and the cost function is nonnegative, it follows that the inequality

$$E_x^\pi \left[e^{\lambda \sum_{t=0}^{T_z-1}(C(X_t,A_t)-g)} \right] \ge e^{-\lambda g}$$

always holds, so that $h_g(x) \ge -g > 0$ for every $x \in S$, and then $g \notin G$; see Definition 3.1. Finally, G is nonempty, since the part (ii) and (3.2) together yield that $g \in G$ when g is large enough.

□

The main results of the paper are stated in the following two theorems. Set

$$g^* := \inf G \in [0, \infty), \tag{3.3}$$

where the inclusion is due to Lemma 3.1(iii), and notice that (3.2) and the previous lemma together imply that

$$G = (g^*, \infty) \quad \text{or} \quad G = [g^*, \infty). \tag{3.4}$$

In the following theorem the existence and uniqueness of a solution of the optimality equation (2.4) is established when the cost function has *compact support*, and such a solution is used to characterize the λ-optimality of a stationary policy.

Theorem 3.1 *Suppose that Assumptions 2.1, 2.3 and 2.4 hold, and that the cost function has compact support, that is, there exists a finite set $F \subset S$ such that*

$$C(x, a) = 0, \quad x \in S \setminus F, \qquad a \in A(x), \tag{3.5}$$

In this case, the following assertions (i)–(v) *are valid.*

(i) $g^* \in G$, $h_{g^*}(z) = 0$, *and* $h_{g^*} \in \mathcal{B}(S)$; *see Definition 3.1 and* (3.3).

(ii) *The pair* $(g^*, h_{g^*}(\cdot))$ *satisfies the optimality equation* (2.4). *Consequently,*

(iii) $J_C^*(\cdot) = g^*$, *and*

(iv) *A policy $f \in \mathbb{F}$ is λ-optimal if, and only if,*

$$e^{\lambda g^* + \lambda h_{g^*}(x)} = e^{\lambda C(x, f(x))} \sum_{y\in S} p_{xy}(f(x)) e^{\lambda h_{g^*}(y)}, \quad x \in S, \tag{3.6}$$

so that a λ-optimal policy certainly exists; see Remark 2.2.

(v) *If $g \in \mathbb{R}$ and $h \in \mathcal{B}(S)$ are such that the pair $(g, h(\cdot))$ satisfies (2.4), then $g = g^*$ and $h(\cdot) = h_{g^*}(\cdot) + h(z)$.*

Concerning the part (ii) of this theorem, an explicit example will be presented in Sect. 7 showing that, if a cost function has compact support but takes a negative value, then the pair $(g^*, h_{g^*}(\cdot))$ does not necessarily satisfy the optimality equation (2.4). On the other hand, the part (v) of this result establishes the uniqueness of a pair $(g, h(\cdot))$ satisfying the optimality equation (2.4) as well as the conditions $h(z) = 0$ and $h \in \mathcal{B}(S)$. It is interesting to note that, as illustrated in the following example, if the condition $h \in \mathcal{B}(S)$ is not required, then (2.4) may admit solutions $(g, h(\cdot))$ where g is not equal to the optimal average cost.

Example 3.1 Suppose that $S = \mathbb{N}$, $A = \{0\}$ and let $p \in (0, 1)$ be fixed. Next, let the transition law $[p_{x\,y}(0)] \equiv [p_{x\,y}]$ be determined as follows:

$$p_{x\,0} := p, \quad p_{x\,x+1} := 1 - p, \qquad x \in \mathbb{N}. \tag{3.7}$$

Since the action space is a singleton, it is clear that Assumption 2.1 holds, and observing that $P_x[T_0 > 1] = P_x[X_1 \neq 0] = 1 - p \in (0, 1)$, it follows that the simultaneous Doeblin condition holds at $z = 0$. On the other hand, observing that $P_x[X_1 = 0] = p$ and $P_0[X_x = x] = (1 - p)^x > 0$ for every $x \in S$, it is not difficult to see that $P_x[T_y < \infty] \geq P_x[T_y \leq y + 1] > 0$ for every $x, y \in S$, so that Assumption 2.4 is also valid. Thus, setting $C(\cdot) = 0$, all the conditions in Theorem 3.1 are satisfied and the optimal average cost is null. Observe now that the optimality equation (2.4) associated to $C(\cdot) = 0$ becomes

$$e^{\lambda g + \lambda h(x)} = p e^{\lambda h(0)} + (1 - p) e^{\lambda h(x+1)}, \quad x \in S. \tag{3.8}$$

Next, given $g \geq 0$, a function $H_g \colon S \to [0, \infty]$ will be recursively constructed such that the pair $(g, H_g(\cdot))$ satisfies the above Poisson equation: Set

$$H_g(0) := 0$$

and, assuming that $H_g(n) \geq 0$ has been specified for some $n \in \mathbb{N}$, observe that $e^{\lambda g + \lambda H_g(n)} - p \geq 1 - p$, and define

$$H_g(n + 1) := \frac{1}{\lambda} \log\left(\frac{e^{\lambda g + \lambda H_g(n)} - p}{1 - p} \right)$$

which is certainly a nonnegative number. From these two last displays it follows immediately that the pair $(g, H_g(\cdot))$ satisfies the Poisson equation (3.8). By Theorem 3.1(v), H_g is unbounded for each $g > 0$.

In the following result the assumption of a compact support for the cost function is dropped, and the main qualitative difference with respect to Theorem 3.1 is that, instead of the optimality equation (2.4), it is asserted that the pair $(g^*, h_{g^*}(\cdot))$ satisfies

an inequality which, as it will be shown by an explicit example in Sect. 7, may be always *strict*.

Theorem 3.2 *Suppose that Assumptions 2.1–2.4 hold. In this context, assertions* (i)–(iv) *below are valid.*

(i) *The inclusion* $g^* \in G$ *holds and*

$$J_C^*(\cdot) = g^*;$$

moreover,

$$\liminf_{n\to\infty} \frac{1}{n} J_{C,n}(\pi, x) \geq g^*, \qquad x \in S, \quad \pi \in \mathcal{P}. \tag{3.9}$$

(ii) *The relative value function* $h_{g^*}(\cdot)$ *is bounded below and the pair* $(g^*, h_{g^*}(\cdot))$ *satisfies the following optimality inequality:*

$$e^{\lambda g^* + \lambda h_{g^*}(x)} \geq \min_{a\in A(x)} \left[e^{\lambda C(x,a)} \sum_{y\in S} p_{x\,y}(a) e^{\lambda h_{g^*}(y)} \right], \quad x \in S, \tag{3.10}$$

Consequently,

(iii) *There exists a policy* $f \in \mathbb{F}$ *satisfying*

$$e^{\lambda g^* + \lambda h_{g^*}(x)} \geq e^{\lambda C(x, f(x))} \sum_{y\in S} p_{x\,y}(f(x)) e^{\lambda h_{g^*}(y)}, \quad x \in S, \tag{3.11}$$

and each stationary policy satisfying this condition is λ*-optimal.*

(iv) *If* $g \in \mathbb{R}$ *and* $h\colon S \to \mathbb{R}$ *are such that*

$$e^{\lambda g + \lambda h(x)} \geq \min_{a\in A(x)} \left[e^{\lambda C(x,a)} \sum_{y\in S} p_{x\,y}(a) e^{\lambda h(y)} \right], \quad x \in S, \tag{3.12}$$

then one of the following assertions (a) or (b) hold:

(a) $g > g^*$;

(b) $g = g^*$ *and* $h(\cdot) - h(z) \geq h_{g^*}(\cdot)$.

Remark 3.1 (i) Using that the average criterion $J_C(\pi, \cdot)$ is additively homogeneous, that is,

$$J_{C+\beta}(\pi, \cdot) = J_C(\pi, \cdot) + \beta, \qquad \pi \in \mathcal{P}, \quad \beta \in \mathbb{R},$$

it is not difficult to verify that the conclusions in Theorem 3.2 remain valid if, instead of the nonnegativity requirement in Assumption 2.1(i), it is supposed that $C(\cdot, \cdot)$ is just bounded below.

(ii) Relation (3.9) establishes that the limit inferior average cost index in Remark 2.1 satisfies $\tilde{J}_C(\pi, x) \geq g^*$, for each $x \in S$ and $\pi \in \mathcal{P}$; thus $g^* \leq \tilde{J}_C^*(\cdot) \leq J_C^*(\cdot) = g^*$, and it follows that the limit superior an limit inferior average criteria have the same optimal value function.

(iii) If a policy π is λ-optimal, that is $J_C(\pi, \cdot) = J_C^*(\cdot)$ [see (2.1)–(2.3)], then (3.9) implies that $\lim_{n\to\infty} J_{C,n}(\pi, \cdot)/n = g^*$, extending the conclusion in the second part of Lemma 2.1. In particular, $J_{C,n}(f, \cdot)/n \to g^*$ when $f \in \mathbb{F}$ is as in (3.11).

The proofs of Theorems 3.1 and 3.2 are rather technical and will be presented after the auxiliary results on the relative value functions presented in the following two sections. Essentially, Sect. 4 concerns general properties, which are valid for arbitrary nonnegative costs, while Sect. 5 is dedicated to analyze the case of a a cost function $C(\cdot, \cdot) \geq 0$ with compact support.

4 Basic properties of the relative value functions

In this section some general properties of the relative value functions are established in the following three lemmas, whose conclusions can be roughly described as follows:

(a) A relative value function h_g is bounded below, it can be realized by using a stationary policy, and the dynamic programming equation satisfied by h_g is determined;

(b) A sufficient criterion on a function $H(\cdot)$ is given so that it dominates a relative value function $h_g(\cdot)$, the finiteness of $h_g(\cdot)$ for $g \in G$ is established, and it is shown that g^* in (3.3) is an upper bound for the optimal average cost function $J_C^*(\cdot)$; finally,

(c) The inclusion $g^* \in G$ is proved.

Lemma 4.1 *Let $g \in \mathbb{R}$ be arbitrary but fixed and suppose that Assumptions 2.1–2.4 hold. In this context, the following properties* (i)–(iii) *are satisfied by the relative value function $h_g(\cdot)$ in Definition 3.1.*

(i) *For each $x \in S$*

$$h_g(x) \geq -N(z)g + \log(1 - \rho(z))/\lambda =: M_g, \tag{4.1}$$

where $N(z) \in \mathbb{N} \setminus \{0\}$ and $\rho(z) \in (0, 1)$ are as in Lemma 2.2.

(ii) *The function $h_g(\cdot)$ satisfies the following dynamic programming equation:*

$$e^{\lambda h_g(x)} = \inf_{a\in A(x)} \left[e^{\lambda(C(x,a)-g)} \left(p_{xz}(a) + \sum_{y\in S\setminus\{z\}} p_{xy}(a) e^{\lambda h_g(y)} \right) \right], \quad x \in S. \tag{4.2}$$

(iii) *There exists a policy* $f_g \in \mathbb{F}$ *such that*

$$e^{\lambda h_g(x)} = e^{\lambda(C(x,f_g(x))-g)}\left(p_{x\,z}(f_g(x)) + \sum_{y\in S\setminus\{z\}} p_{x\,y}(f_g(x))e^{\lambda h_g(y)}\right), \quad x \in S. \tag{4.3}$$

Proof (i) Let $x \in S$ and $\pi \in \mathcal{P}$ be arbitrary and notice that

$$\begin{aligned} E_z^\pi\left[e^{\lambda\sum_{t=0}^{T_z-1}(C(X_t,A_t)-g)}\right] &\geq E_z^\pi\left[e^{\lambda\sum_{t=0}^{T_z-1}(C(X_t,A_t)-g)}I[T_z \leq N(z)]\right] \\ &\geq e^{-\lambda N(z)g}P_z^\pi[T_z \leq N(z)] \\ &\geq e^{-\lambda N(z)g}(1-\rho(z)), \end{aligned}$$

where the condition $C(\cdot,\cdot) \geq 0$ in Assumption 2.1(i) was used to set the second inequality and Lemma 2.2 was used in the last step. Now, the conclusion follows combining the above display with Definition 3.1.

(ii) Let $\pi \in \mathcal{P}$ and $x \in S$ be arbitrary. Using that $[T_z = 1] = [X_1 = z]$, the Markov property yields that, for each $a \in A(x)$,

$$\begin{aligned} &E_x^\pi\left[e^{\lambda\sum_{t=0}^{T_z-1}(C(X_t,A_t)-g)}I[T_z > 1]\,\middle|\, A_0 = a, X_1\right] \\ &= e^{\lambda(C(x,a)-g)}I[X_1 \neq z]E_{X_1}^{\pi^{(x,a)}}\left[e^{\lambda\sum_{t=0}^{T_z-1}(C(X_t,A_t)-g)}\right] \\ &\geq e^{\lambda(C(x,a)-g)}I[X_1 \neq z]e^{\lambda h_g(X_1)}, \end{aligned}$$

where the shifted p olicy $\pi^{(x,a)}$ is determined by $\pi_t^{(x,a)}(\cdot|\mathbf{h}_t) = \pi_{t+1}(\cdot|x,a,\mathbf{h}_t)$, and (3.1) was used to set the inequality. Thus,

$$\begin{aligned} &E_x^\pi\left[e^{\lambda\sum_{t=0}^{T_z-1}(C(X_t,A_t)-g)}I[T_z > 1]\,\middle|\, A_0 = a\right] \\ &\geq e^{\lambda(C(x,a)-g)}\sum_{y\in S\setminus\{z\}} p_{x\,y}(a)e^{\lambda h_g(y)}, \end{aligned}$$

and combining this relation with

$$E_x^\pi\left[e^{\lambda\sum_{t=0}^{T_z-1}(C(X_t,A_t)-g)}I[T_z = 1]\,\middle|\, A_0 = a\right] = e^{\lambda(C(x,a)-g)}p_{x\,z}(a)$$

it follows that

$$\begin{aligned} &E_x^\pi\left[e^{\lambda\sum_{t=0}^{T_z-1}(C(X_t,A_t)-g)}\,\middle|\, A_0 = a\right] \\ &\geq e^{\lambda(C(x,a)-g)}\left(p_{x\,z}(a) + \sum_{y\in S\setminus\{z\}} p_{x\,y}(a)e^{\lambda h_g(y)}\right) \\ &\geq \inf_{a\in A(x)}\left[e^{\lambda(C(x,a)-g)}\left(p_{x\,z}(a) + \sum_{y\in S\setminus\{z\}} p_{x\,y}(a)e^{\lambda h_g(y)}\right)\right]; \end{aligned}$$

integrating with respect to the distribution of A_0, it follows that $E_x^{\pi}[e^{\lambda\sum_{t=0}^{T_z-1}(C(X_t,A_t)-g)}]$ is larger than or equal to the infimum in the above display, a fact that yields

$$e^{\lambda h_g(x)} \geq \inf_{a\in A(x)}\left[e^{\lambda(C(x,a)-g)}\left(p_{x\,z}(a)+\sum_{y\in S\setminus\{z\}}p_{x\,y}(a)e^{\lambda h(y)}\right)\right], \quad x\in S, \tag{4.4}$$

by Definition 3.1(i). To establish the reverse inequality, let $\varepsilon > 0$ be arbitrary and notice that, for each $y \in S$, (3.1) yields that there exists a policy $\pi^y \in \mathcal{P}$ such that

$$E_y^{\pi^y}\left[e^{\lambda\sum_{t=0}^{T_z-1}(C(X_t,A_t)-g)}\right] \leq e^{\lambda(h_g(y)+\varepsilon)}.$$

Next, for each $f \in \mathbb{F}$, let the new policy $\gamma^f = \{\gamma_t^f\} \in \mathcal{P}$ be specified as follows: $\gamma_0^f(\{f(x)\}|x) = 1$ for each $x \in S$, while $\gamma_{t+1}^f(\cdot|\mathbf{h}_{t+1}) = \pi_t^{x_1}(\cdot|x_1,\ldots,a_t,x_{t+1})$ for each $t \in \mathbb{N}$ and $\mathbf{h}_{t+1} \in \mathbb{H}_{t+1}$. A controller choosing actions according to γ^f operates as follows: At time $t = 0$ the action applied is selected using f and, after observing $X_1 = y$, from time 1 onwards the actions are selected using π^y as if the precess had started again. Using the Markov property and Definition 3.1(i) it is not difficult to see that

$$\begin{aligned} e^{\lambda h_g(x)} &\leq E_y^{\gamma^f}\left[e^{\lambda\sum_{t=0}^{T_z-1}(C(X_t,A_t)-g)}\right] \\ &= e^{\lambda(C(x,f(x))-g)} \\ &\quad\times\left(p_{x\,z}(f(x))+\sum_{y\in S\setminus\{z\}}p_{x\,y}(f(x))E_y^{\pi^y}\left[e^{\lambda\sum_{t=0}^{T_z-1}(C(X_t,A_t)-g)}\right]\right), \end{aligned}$$

and combining the two last displays it follows that, for each $x \in S$,

$$e^{\lambda h_g(x)} \leq e^{\varepsilon}e^{\lambda(C(x,f(x))-g)}\left(p_{x\,z}(f(x))+\sum_{y\in S\setminus\{z\}}p_{x\,y}(f(x))e^{\lambda h_g(y)}\right);$$

since $\varepsilon > 0$ and $f \in \mathbb{F}$ are arbitrary, this relation implies that

$$e^{\lambda h_g(x)} \leq \inf_{a\in A(x)}\left[e^{\lambda(C(x,a)-g)}\left(p_{x\,z}(a)+\sum_{y\in S\setminus\{z\}}p_{x\,y}(a)e^{\lambda h_g(y)}\right)\right], \quad x\in S,$$

a fact that, via (4.4), leads to (4.2).

(iii) From Assumption 2.1 the term within brackets in (4.2) has a a minimizer $f_g(x) \in A(x)$, and (4.3) follows via (4.2); see Remark 2.2. □

Lemma 4.2 *Let $H : S \to (-\infty, \infty]$, $f \in \mathbb{F}$ and $g \in \mathbb{R}$ be such that the following conditions* (a) *and* (b) *are satisfied:*

(a) $H(z) < \infty$, *and*

(b) *For each* $x \in S$, $e^{\lambda H(x)} \geq e^{\lambda(C(x,f(x))-g)} \Big(p_{x\,z}(f(x)) + \sum_{y \in S\setminus\{z\}} p_{x\,y}(f(x))\, e^{\lambda H(y)} \Big)$.

In this framework the following assertions (i)–(iv) *hold.*

(i) $H(\cdot)$ *is a finite function , and*

(ii) $e^{\lambda H(x)} \geq E_x^f \left[e^{\lambda \sum_{t=0}^{T_z-1} (C(X_t,A_t)-g)} \right] \geq e^{\lambda h_g(x)}$ *for each state* x. *Consequently,*

(iii) *If* $g \in G$, *then the relative value function* $h_g(\cdot)$ *is finite and*

$$e^{\lambda h_g(x)} = E_x^{f_g} \left[e^{\lambda \sum_{t=0}^{T_z-1} (C(X_t,A_t)-g)} \right], \quad x \in S,$$

where the policy $f_g \in \mathbb{F}$ *is as in (4.3), and*

(iv) $g^* \geq J_C^*(\cdot)$.

Proof (i) Since $H(\cdot) > -\infty$, it is sufficient to show that $H(\cdot) < \infty$. To achieve this goal observe that, since $H(z)$ is finite, the inequality in condition (b) yields that

$$H(x) < \infty \quad \text{and} \quad p_{x\,y}(f(x)) > 0 \Rightarrow H(y) < \infty.$$

Now, let $y \in S$ be arbitrary and notice that Assumption 2.4 implies that there exists a positive integer k as well as states $x_0, x_1, \ldots, x_k$ such that $x_0 = z$, $x_k = y$ and

$$p_{x_{i-1}\,x_i}(f(x_{i-1})) > 0, \quad i = 1, 2, \ldots, k.$$

Using that $H(x_0) = H(z) < \infty$, these two last displays together lead to $H(y) = H(x_k) < \infty$, and then $H(\cdot) < \infty$, since $y \in S$ is arbitrary.

(ii) Recalling that $A_t = f(X_t)$ under the action of policy f, condition (b) and Definition 2.1(i) together yield that

$$\begin{aligned} e^{\lambda H(x)} &\geq E_x^f \Big[e^{\lambda(C(X_0,A_0)-g)} I[X_1 = z] \\ &\qquad + e^{\lambda(C(X_0,A_0)-g)} e^{\lambda H(X_1)} I[X_1 \neq z] \Big] \\ &= E_x^f \Big[e^{\lambda(C(X_0,A_0)-g)} I[T_Z = 1] \\ &\qquad + e^{\lambda(C(X_0,A_0)-g)} e^{\lambda H(X_1)} I[T_z > 1] \Big], \quad x \in S. \end{aligned} \tag{4.5}$$

Next, observe that for every positive integer n and $x \in S$, the Markov property and condition (b) together imply that

$$
\begin{aligned}
&E_x^f\left[I[T_z > n]e^{\lambda\sum_{t=0}^{n-1}(C(X_t,A_t)-g)}e^{\lambda H(X_n)}\Big|\, X_s, s \le n\right]\\
&= I[T_z > n]e^{\lambda\sum_{t=0}^{n-1}(C(X_t,A_t)-g)}e^{\lambda H(X_n)}\\
&\ge I[T_z > n]e^{\lambda\sum_{t=0}^{n-1}(C(X_t,A_t)-g)}e^{\lambda(C(X_n,A_n)-g)}\\
&\quad\left(p_{X_n z}(f(X_n)) + \sum_{y\in S\setminus\{z\}} p_{X_n y}(f(X_n))e^{\lambda H(y)}\right)\\
&= I[T_z > n]e^{\lambda\sum_{t=0}^{n}(C(X_t,A_t)-g)}E_x^f\left[I[X_{n+1}=z]+I[X_{n+1}\ne z]e^{\lambda H(X_{n+1})}\Big|\, X_s, s\le n\right]\\
&= E_x^f\left[e^{\lambda\sum_{t=0}^{n}(C(X_t,A_t)-g)}I[T_z > n]I[X_{n+1}=z]\Big|\, X_s, s\le n\right]\\
&\quad + E_x^f\left[I[T_z > n]I[X_{n+1}\ne z]e^{\lambda\sum_{t=0}^{n}(C(X_t,A_t)-g)}e^{\lambda H(X_{n+1})}\Big|\, X_s, s\le n\right]\\
&= E_x^f\left[e^{\lambda\sum_{t=0}^{n}(C(X_t,A_t)-g)}I[T_z = n+1]\Big|\, X_s, s\le n\right]\\
&\quad + E_x^f\left[I[T_z > n+1]e^{\lambda\sum_{t=0}^{n}(C(X_t,A_t)-g)}e^{\lambda H(X_{n+1})}\Big|\, X_s, s\le n\right],
\end{aligned}
$$

and then

$$
\begin{aligned}
&E_x^f\left[I[T_z > n]e^{\lambda\sum_{t=0}^{n-1}(C(X_t,A_t)-g)}e^{\lambda H(X_n)}\right]\\
&\ge E_x^f\left[e^{\lambda\sum_{t=0}^{n}(C(X_t,A_t)-g)}I[T_z = n+1]\right]\\
&\quad + E_x^f\left[I[T_z > n+1]e^{\lambda\sum_{t=0}^{n}(C(X_t,A_t)-g)}e^{\lambda H(X_{n+1})}\right],
\end{aligned}
$$

an inequality that, via an induction argument using (4.5), yields that for every $x \in S$ and $n = 1, 2, 3, \ldots$

$$
\begin{aligned}
e^{\lambda H(x)} &\ge \sum_{k=1}^{n} E_x^f\left[I[T_z = k]e^{\lambda\sum_{t=0}^{k-1}(C(X_t,A_t)-g)}\right]\\
&\quad + E_x^f\left[I[T_z > n]e^{\lambda\sum_{t=0}^{n-1}(C(X_t,A_t)-g)}e^{\lambda H(X_n)}\right],
\end{aligned}
$$

and then

$$
\begin{aligned}
e^{\lambda H(x)} &\ge \lim_{n\to\infty}\sum_{k=1}^{n} E_x^f\left[I[T_z = k]e^{\lambda\sum_{t=0}^{k-1}(C(X_t,A_t)-g)}\right]\\
&= \sum_{k=1}^{\infty} E_x^f\left[I[T_z = k]e^{\lambda\sum_{t=0}^{k-1}(C(X_t,A_t)-g)}\right]\\
&= E_x^f\left[e^{\lambda\sum_{t=0}^{T_z-1}(C(X_t,A_t)-g)}\right] \ge e^{\lambda h_g(x)}.
\end{aligned}
$$

where it was used that $P_x^f[T_z < \infty] = 1$ to set the second equality, by Lemma 2.2(ii), and the second inequality follows from (3.1).

(iii) Let $g \in G$ be arbitrary so that $h_g(z) \leq 0$, and let $f_g \in \mathbb{F}$ be as in (4.3). In this case conditions (a) and (b) hold with h_g and f_g instead of H and f, respectively, and the conclusion follows from parts (i) and (ii).

(iv) Given $g \in G$, let policy $f_g \in \mathbb{F}$ be such that (4.3) holds, and notice that $1 \geq e^{\lambda h_g(z)}$, since $h_g(z) \leq 0$, so that (4.3) immediately implies that $e^{\lambda h_g(x)} \geq e^{C(x,f_g(x))-g} \sum_{y\in S} p_{xy} e^{\lambda h_g(y)}$ for every state x. From this point, an induction argument using the Markov property yields that, for each positive integer n and $x \in S$,

$$e^{\lambda n g + \lambda h_g(x)} \geq E_x^{f_g}\left[e^{\lambda \sum_{t=0}^{n-1} C(X_t,A_t)+h_g(X_n)}\right] \geq e^{\lambda J_{C,n}(f_g,x)+\lambda M_g}$$

where (2.1) and the lower bound in (4.1) were used to set the second inequality. Thus,

$$g + \frac{h_g(x) - M_g}{n} \geq \frac{1}{n} J_{C,n}(f_g, x)$$

and, since $h_g(\cdot)$ is finite, via (2.2) and (2.3) this leads to $g \geq J_C(f_g, \cdot) \geq J_C^*(\cdot)$; since $g \in G$ is arbitrary, it follows that $\inf G = g^* \geq J_C^*(\cdot)$.

□

Lemma 4.3 *With the notation in* (3.3), $g^* \in G$.

Proof Let $\{g_k\} \subset G$ be such that

$$g_k \searrow g^* \quad \text{as} \quad k \nearrow \infty \tag{4.6}$$

and, applying Lemma 4.1 (iii), select a policy $f_{g_k} \in \mathbb{F}$ such that

$$e^{\lambda h_{g_k}(x)} = e^{\lambda(C(x,f_{g_k}(x))-g_k)}\left(p_{xz}(f_{g_k}(x)) + \sum_{y\in S\setminus\{z\}} p_{xy}(f_{g_k}(x)) e^{\lambda h_{g_k}(y)}\right), \quad x \in S; \tag{4.7}$$

using the fact that $\mathbb{F}$ is a compact metric space, after taking a subsequence (if necessary) it can be assumed that $\{f_{g_k}\}$ converges to a policy $f^* \in \mathbb{F}$:

$$\lim_{k\to\infty} f_{g_k}(x) = f^*(x), \quad x \in S. \tag{4.8}$$

Next, observe that the monotonicity property in Lemma 3.1 and (4.6) together yield that there exists a function $H^*(\cdot)$ defined on S such that

$$h_{g_1}(x) \leq \lim_{k\to\infty} h_{g_k}(x) = H^*(x) \leq h_{g^*}(x), \quad x \in S. \tag{4.9}$$

Since $h_{g_1}(\cdot)$ is bounded below, by Lemma 4.1, the first inequality implies that $H^*(\cdot) > -\infty$ while using that $g_k \in G$, so that $h_{g_k}(z) \leq 0$, the above convergence yields that

(a) $H^*(z) \leq 0$.
On the other hand, taking limit as k goes to ∞ in both sides of (4.7), Assumption 2.1(ii), (4.6), Fatou's lemma and the two last displays together imply that

(b) $$e^{\lambda H^*(x)} \geq e^{\lambda(C(x,f^*(x))-g^*)} \times \left(p_{x\,z}(f^*(x)) + \sum_{y\in S\setminus\{z\}} p_{x\,y}(f^*(x)) e^{\lambda H^*(y)} \right), \quad x \in S.$$

These two facts (a) and (b) show that the conditions in Lemma 4.2 are satisfied with H^* and f^* instead of H and f, respectively, and it follows that $H^*(\cdot)$ is a finite function as well as

$$e^{\lambda H^*(x)} \geq E_x^{f^*}\left[e^{\lambda \sum_{t=0}^{T_z-1}(C(X_t,A_t)-g^*)} \right] \geq e^{\lambda h_{g^*}(x)}, \quad x \in S.$$

This relation and (4.9) together lead to $H(\cdot) = h_{g^*}(\cdot)$, and from this point the fact (a) above yields that $h_{g^*}(z) \leq 0$, so that $g^* \in G$; see (3.2). □

5 Nonnegative costs with compact support

In this section additional properties of the relative value functions are established under the assumption that the cost function $C(\cdot,\cdot) \geq 0$ has compact support. In this context, the main goals to be established can be described as follows: (a) The relative value function $h_{g^*}(\cdot)$ is bounded, and (b) the equality $h_{g^*}(z) = 0$ holds; this latter property is the backbone of the argument used in Sect. 6 to establish Theorem 3.1. Finally, the exposition concludes with a result about (c) the uniqueness of solutions of the Poisson equation associated with a stationary policy f; see Remark 2.2(iii).

Lemma 5.1 *Suppose that Assumptions 2.1, 2.3 and 2.4 hold, and that the cost function C has compact support, that is, the condition (3.5) holds for some finite set F. In this case the relative value function $h_{g^*}(\cdot)$ is bounded.*

Proof Let the finite set F be as in (3.5) and without loss of generality assume that $z \in F$, so that

$$T_z \geq T_F; \tag{5.1}$$

see Definition 2.1(i). Now, set

$$M^* := 1 + \max_{y\in F} |h_{g^*}(y)| \tag{5.2}$$

and notice that M^* is finite, since F is a finite set and $h_{g^*}(\cdot)$ is a finite function; see Lemmas 4.2(iii) and 4.3. Next, it will be shown that

$$h_{g^*}(x) \le M^*, \quad x \in S \setminus F. \tag{5.3}$$

To achieve this goal, let $f_{g^*} \in \mathbb{F}$ be as in Lemma 4.2(iii), so that

$$e^{\lambda h_{g^*}(y)} = E_y^{f_{g^*}}\left[e^{\lambda \sum_{t=0}^{T_z-1}(C(X_t,A_t)-g^*)}\right], \quad y \in S, \tag{5.4}$$

and let the initial state $X_0 = x \in S \setminus F$ be arbitrary but fixed. Since $X_t \notin F$ for $1 \le t < T_F$ [see Definition 2.1(i)] and $X_0 = x \notin F$, condition (3.5) yields that the following equalities hold $P_x^{f_{g^*}}$-almost surely:

$$\sum_{t=0}^{T_z-1}(C(X_t, A_t) - g^*) = -g^* T_F \quad \text{on the event } [T_z = T_F]$$

and

$$\sum_{t=0}^{T_z-1}(C(X_t, A_t) - g^*) = -g^* T_F + \sum_{t=T_F}^{T_z-1}(C(X_t, A_t) - g^*) \quad \text{on the event } [T_z > T_F].$$

Thus, since $g^* \ge 0$ [see (3.3)],

$$\begin{aligned} E_x^{f_{g^*}}\left[e^{\sum_{t=0}^{T_z-1}(C(X_t,A_t)-g^*)} I[T_z = T_F]\right] &= E_x^{f_{g^*}}\left[e^{-g^* T_F} I[T_z = T_F]\right] \\ &\le P_x^{f_{g^*}}[T_z = T_F], \end{aligned} \tag{5.5}$$

while, using that $I[T_z > T_F = k] = I[X_s \notin F, 1 \le s < k, X_k \in F \setminus \{z\}]$, the Markov property yields that for every positive integer k

$$\begin{aligned} &E_x^{f_{g^*}}\left[e^{\sum_{t=0}^{T_z-1}(C(X_t,A_t)-g^*)} I[T_z > T_F = k] | X_1, \ldots, X_k\right] \\ &= E_x^{f_{g^*}}\left[e^{-g^* T_F} e^{\sum_{t=T_F}^{T_z-1}(C(X_t,A_t)-g^*)} I[T_z > T_F = k] | X_1, \ldots, X_k\right] \\ &= e^{-g^* T_F} I[T_z > T_F = k] E_{X_k}^{f_{g^*}}\left[e^{\sum_{t=0}^{T_z-1}(C(X_t,A_t)-g^*)}\right] \\ &= e^{-g^* T_F} I[T_z > T_F = k] e^{\lambda h_{g^*}(X_k)} \\ &\le I[T_z > T_F = k] e^{\lambda M^*} \end{aligned}$$

where (5.4) was used to set the third equality and, since $X_k \in F$ on the event $[T_F = k]$, the inequality follows from (5.2). Therefore, $E_x^{f_{g^*}}\left[e^{\sum_{t=0}^{T_z-1}(C(X_t,A_t)-g^*)} I[T_z > T_F]\right] \le$

$P_x^{f_{g^*}}[T_z > T_F]e^{\lambda M^*}$, and combining this with (5.1) and (5.5) it follows that

$$e^{\lambda h_{g^*}(x)} = E_x^{f_{g^*}}\left[e^{\sum_{t=0}^{T_z-1}(C(X_t,A_t)-g^*)}\right] \le P_x^{f_{g^*}}[T_z = T_F] + P_x^{f_{g^*}}[T_z > T_F]e^{\lambda M^*} \le e^{\lambda M^*},$$

where (5.4) was used to set the equality; this implies (5.3), since x is an arbitrary state in $S \setminus F$. Combining (5.2) with (5.3) it follows that $h_{g^*}(\cdot) \le M^* < \infty$, and the desired conclusion is obtained recalling that $h_{g^*}(\cdot) \ge M_{g^*} > -\infty$, by Lemma 4.1(i). □

The main result of this section is the following.

Theorem 5.1 *Under the conditions in Lemma 5.1,* $h_{g^*}(z) = 0$.

The proof of this theorem relies on properties of the following matrix Q, which is derived from a solution of the Poisson equation associated with a stationary policy; see Remark 2.2(iii).

Definition 5.1 Let $g \in \mathbb{R}$, $f \in \mathbb{F}$ and $D, H : S \to \mathbb{R}$ be such that the following Poisson equation holds:

$$e^{\lambda H(x)} = e^{\lambda D(x)-\lambda g}\sum_{y\in S} p_{x\,y}(f(x))e^{\lambda H(y)}, \quad x \in S. \tag{5.6}$$

In this context, on the space S define the matrix $Q(f, g, D, H) \equiv Q = [q_{x\,y}]_{x,y\in S}$ as follows:

$$q_{x\,y} := \frac{e^{\lambda D(x)-\lambda g}\, p_{x\,y}(f(x))e^{\lambda H(y)}}{e^{\lambda H(x)}}, \quad x, y \in S. \tag{5.7}$$

Remark 5.1 Notice that (5.6) and the specification (5.7) immediately yield that the matrix Q is stochastic. The distribution of the state process $\{X_n\}$ when Q is the one-step transition matrix and $X_0 = x$ is the initial state is denoted by P_x^Q, with E_x^Q standing for the corresponding expectation operator.

Lemma 5.2 *In the context of Definition 5.1, suppose that* $H(\cdot) \in \mathcal{B}(S)$ *and that Assumption 2.3 holds. In this situation, the assertions* (i)–(iii) *below hold*:

(i) $D \in \mathcal{B}(S)$.

(ii) *For each* $y \in S$, *the matrix* Q *in* (5.7) *satisfies the simultaneous Doeblin condition at* y. *More precisely, if* $N(y) \in \mathbb{N} \setminus \{0\}$ *and* $\rho(y) \in (0, 1)$ *are as in* (2.7), *then*

$$P_x^Q[T_y \le N(y)] \ge 1 - \tilde{\rho}(y), \quad x \in S,$$

where $\tilde{\rho}(y) \in (0, 1)$ *is determined by*

$$1 - \tilde{\rho}(y) = (1 - \rho(y))e^{-\lambda N(y)\|D(\cdot)-g\|-2\lambda\|H\|} > 0, \quad y \in S.$$

(iii) *For each* $\varepsilon \in (-\infty, -\log(\tilde{\rho}(y))/[\lambda N(y)])$ *the function* $H_\varepsilon : S \to \mathbb{R}$ *defined by*

$$H_\varepsilon(y) = 0, \quad \text{and} \quad H_\varepsilon(x) = \frac{1}{\lambda} \log\left(E_x^Q[e^{\lambda \varepsilon T_y}]\right), \quad x \in S \setminus \{y\}, \tag{5.8}$$

is bounded and satisfies the following Poisson equation:

$$e^{\lambda H_\varepsilon(x)} = e^{\lambda\varepsilon - \lambda\alpha(\varepsilon) D_y(x)} \sum_{w \in S} q_{x\,w} e^{\lambda H_\varepsilon(w)}, \quad x \in S, \tag{5.9}$$

where

$$\alpha(\varepsilon) = \frac{1}{\lambda} \log\left(E_y^Q[e^{\lambda \varepsilon T_y}]\right) \tag{5.10}$$

and D_y *is the indicator function of point* y*, that is*

$$D_y(x) = 0, \quad \text{if } x \in S \setminus \{y\}, \quad \text{and} \quad D_y(y) = 1. \tag{5.11}$$

Proof (i) From (5.6) it follows that $e^{-\lambda\|H\|} \le e^{\lambda H(x)} \le e^{\lambda D(x) - \lambda g} e^{\lambda \|H\|}$ and $e^{\lambda\|H\|} \ge e^{\lambda H(x)} \ge e^{\lambda D(x) - \lambda g} e^{-\lambda\|H\|}$ for every $x \in S$, so that $\|D\| \le 2\|H\| + |g|$.

(ii) From (5.7), for each integer $k > 1$ the following relations for the probability of the event $[T_y = k]$ with respect to P_x^Q hold, where $x = x_0$ and $x_k = y$:

$$\begin{aligned}
P_x^Q[T_y = k] &= \sum_{\substack{x_i \neq y \\ i=1,2,\ldots,k-1}} \prod_{i=1}^{k} q_{x_{i-1}\,x_i} \\
&= \sum_{\substack{x_i \neq y \\ i=1,2,\ldots,k-1}} \prod_{i=1}^{k} \frac{e^{\lambda D(x_{i-1}) - \lambda g} p_{x_{i-1}\,x_i}(f(x_{i-1})) e^{\lambda H(x_i)}}{e^{\lambda H(x_{i-1})}} \\
&= \sum_{\substack{x_i \neq y \\ i=1,2,\ldots,k-1}} \left(\prod_{i=1}^{k} e^{\lambda D(x_{i-1}) - \lambda g}\right) \\
&\quad \times \left(\prod_{i=1}^{k} p_{x_{i-1}\,x_i}(f(x_{i-1}))\right) \prod_{i=1}^{k} \frac{e^{\lambda H(x_i)}}{e^{\lambda H(x_{i-1})}} \\
&\ge \sum_{\substack{x_i \neq y \\ i=1,2,\ldots,k-1}} e^{-k\lambda\|D(\cdot) - g\|} \left(\prod_{i=1}^{k} p_{x_{i-1}\,x_i}(f(x_{i-1}))\right) \frac{e^{\lambda H(x_k)}}{e^{\lambda H(x_0)}} \\
&= e^{-k\lambda\|D(\cdot) - g\|} \frac{e^{\lambda H(y)}}{e^{\lambda H(x)}} \sum_{\substack{x_i \neq y \\ i=1,2,\ldots,k-1}} \prod_{i=1}^{k} p_{x_{i-1}\,x_i}(f(x_{i-1}))
\end{aligned}$$

and then

$$P_x^Q[T_y = k] \geq e^{-k\lambda\|D(\cdot)-g\|-2\lambda\|H\|} P_x^f[T_y = k]).$$

Also, it is not difficult to see that this inequality is also valid for $k = 1$, so that

$$\begin{aligned} P_x^Q[T_y \leq N(y)] &= \sum_{k=1}^{N(y)} P_x^Q[T_y = k] \\ &\geq \sum_{k=1}^{N(y)} e^{-k\lambda\|D(\cdot)-g\|-2\lambda\|H\|} P_x^f[T_y = k] \\ &\geq e^{-N(y)\lambda\|D(\cdot)-g\|-2\lambda\|H\|} \sum_{k=1}^{N(y)} P_x^f[T_y = k] \\ &= e^{-N(y)\lambda\|D(\cdot)-g\|-2\lambda\|H\|} P_x^f[T_y \leq N(y)] \end{aligned}$$

and the conclusion follows from (2.7).

(iii) An induction argument using the previous part yields that $P_x^Q[T_y > kN(y)] \leq \tilde{\rho}(y)^k$ for each positive integer k and $x \in S$, and this leads to

$$P_x^Q[T_y \geq n] \leq \tilde{\rho}(y)^{n/N(y)-1}, \qquad n = 1, 2, 3, \ldots, \quad x \in S;$$

thus, if $\varepsilon < -\log(\tilde{\rho}(y))/[\lambda N(y)]$, so that $e^{\lambda\varepsilon}\tilde{\rho}(y)^{1/N(y)} < 1$, then

$$\begin{aligned} E_x^Q[e^{\lambda\varepsilon T_y}] &= \sum_{n=1}^{\infty} e^{\lambda\varepsilon n} P_x[T_y = n] \\ &\leq \sum_{n=1}^{\infty} e^{\lambda\varepsilon n} P_x[T_y \geq n] \leq \sum_{n=1}^{\infty} e^{\lambda\varepsilon n} \rho(y)^{n/N(y)-1} < \infty, \quad x \in S, \end{aligned}$$

and observing that the inequality $E_x^Q[e^{\lambda\varepsilon T_y}] \geq E_x^Q[e^{\lambda\varepsilon T_y} I[T_y \leq N(y)]] \geq e^{-\lambda|\varepsilon|N(y)}(1 - \tilde{\rho}(y))$ is always valid, it follows that $H_\varepsilon(\cdot)$ in (5.8) is bounded. On the other hand, via the Markov property, a conditioning argument yields that for each $x \in S$

$$E_x^Q[e^{\lambda\varepsilon T_y}] = e^{\lambda\varepsilon}\left(q_{x\,y} + \sum_{w \neq y} q_{x\,w} E_w^Q[e^{\lambda\varepsilon T_y}]\right),$$

an equality that together with (5.8) and (5.11) shows that the Poisson equation (5.9) holds if $x \neq y$. On the other hand, setting $x = y$ the above display and

(5.10) allow to write

$$e^{\lambda\alpha(\varepsilon)} = e^{\lambda\varepsilon}\left(q_{y\,y} + \sum_{w\neq y} q_{y\,w} E_w^Q[e^{\lambda\varepsilon T_y}]\right),$$

that is

$$1 = e^{\lambda\varepsilon-\lambda\alpha(\varepsilon)}\left(q_{y\,y} + \sum_{w\neq y} q_{y\,w} E_w^Q[e^{\lambda\varepsilon T_y}]\right),$$

and a glance at (5.11) and (5.8) shows that the equality in (5.9) also holds for $x = y$.

□

Proof of Theorem 5.1 Set

$$\Delta := -h_{g^*}(z) \geq 0, \tag{5.12}$$

where, using that $g^* \in G$—by Lemma 4.3—the inequality follows from (3.2). It will be shown, by contradiction, that Δ is null.
Assume that $\Delta > 0$. Recall that $h_{g^*}(\cdot)$ is bounded, by Lemma 5.1, while Lemma 4.1(iii) yields that there exists a policy $f_{g^*} \in \mathbb{F}$ such that

$$e^{\lambda h_{g^*}(x)} = e^{\lambda(C(x,f_{g^*}(x))-g^*)}\left(p_{x\,z}(f_{g^*}(x)) + \sum_{y\in S\setminus\{z\}} p_{x\,y}(f_{g^*}(x))e^{\lambda h_{g^*}(y)}\right), \quad x \in S. \tag{5.13}$$

Now define the function $H \in \mathcal{B}(S)$ by

$$H(x) := h_{g^*}(x) + \Delta D_z(x), \quad x \in S, \tag{5.14}$$

where D_z is the indicator function of point z; see (5.11); since

$$H(z) = 0, \tag{5.15}$$

[see (5.12)] and $H(x) = h_{g^*}(x)$ for $x \neq z$, it follows form (5.13) that

$$e^{\lambda H(x)} = e^{\lambda(C(x,f_{g^*}(x))-g^*)} \sum_{y\in S} p_{x\,y}(f_{g^*}(x))e^{\lambda H(y)}, \quad x \in S \setminus \{z\}$$

whereas using the the notation in (5.12), the equality in (5.13) with $x = z$ is equivalent to

$$e^{\lambda H(z)} = 1 = e^{\lambda(C(z,f_{g^*}(z))+\Delta-g^*)}\left(p_{z\,z}(f_{g^*}(z)) + \sum_{y\in S\setminus\{z\}} p_{z\,y}(f_{g^*}(z))e^{\lambda H(y)}\right)$$
$$= e^{\lambda(C(z,f_{g^*}(z))+\Delta-g^*)}\left(p_{z\,z}(f_{g^*}(z))e^{\lambda H(z)} + \sum_{y\in S\setminus\{z\}} p_{z\,y}(f_{g^*}(z))e^{\lambda H(y)}\right).$$

These two last displays lead to

$$e^{\lambda H(x)} = e^{\lambda(D(x)-g^*)}\sum_{y\in S} p_{x\,y}(f_{g^*}(x))e^{\lambda H(y)}, \quad x \in S, \tag{5.16}$$

where

$$D(x) = C(x, f_{g^*}(x)) + \Delta D_z(x), \quad x \in S. \tag{5.17}$$

Therefore, defining the matrix $[q_{x\,y}]$ by

$$q_{x\,y} := \frac{e^{\lambda D(x)-\lambda g^*}\, p_{x\,y}(f_{g^*}(x))e^{\lambda H(y)}}{e^{\lambda H(x)}} \tag{5.18}$$

it follows from Lemma 5.2(iii) that for each $\varepsilon > 0$ small enough, there exists a bounded function H_ε such that

$$H_\varepsilon(z) = 0 \tag{5.19}$$

and

$$e^{\lambda H_\varepsilon(x)} = e^{\lambda\varepsilon-\lambda\alpha(\varepsilon)D_z(x)}\sum_{y\in S} q_{x\,y}e^{\lambda H_\varepsilon(y)}, \quad x \in S, \tag{5.20}$$

where $\alpha(\varepsilon)$ is given by (5.10) with $y = z$, so that $\alpha(\varepsilon) \searrow 0$ as $\varepsilon \searrow 0$. Thus, without loss of generality, it can be assumed that $\varepsilon > 0$ is chosen in such a way that

$$\alpha(\varepsilon) < \Delta. \tag{5.21}$$

Combining (5.18) and (5.20) it follows that for each $x \in S$

$$\begin{aligned} e^{\lambda H_\varepsilon(x)} &= e^{\lambda\varepsilon-\lambda\alpha(\varepsilon)D_z(x)} \sum_{y\in S} q_{x\,y} e^{\lambda H_\varepsilon(y)} \\ &= e^{\lambda\varepsilon-\lambda\alpha(\varepsilon)D_z(x)} \sum_{y\in S} \frac{e^{\lambda D(x)-\lambda g^*} p_{x\,y}(f_{g^*}(x)) e^{\lambda H(y)}}{e^{\lambda H(x)}} e^{\lambda H_\varepsilon(y)} \\ &= \frac{e^{\lambda[D(x)-\alpha(\varepsilon)D_z(x)]-\lambda(g^*-\varepsilon)}}{e^{\lambda H(x)}} \sum_{y\in S} p_{x\,y}(f_{g^*}(x)) e^{\lambda[H(y)+H_\varepsilon(y)]} \end{aligned}$$

and then

$$e^{\lambda[H_\varepsilon(x)+H(x)]} = e^{\lambda[D(x)-\alpha(\varepsilon)D_z(x)]-\lambda(g^*-\varepsilon)} \sum_{y\in S} p_{x\,y}(f_{g^*}(x)) e^{\lambda[H(y)+H_\varepsilon(y)]};$$

observing that (5.17) and (5.21) together yield that $D(x)-\alpha(\varepsilon)D_z(x) = C(x, f_{g^*}(x)) + \Delta D_z(x) - \alpha(\varepsilon)D_z(x) \geq C(x, f_{g^*}(x))$, it follows that

$$e^{\lambda[H_\varepsilon(x)+H(x)]} \geq e^{\lambda C(x,f_{g^*}(x))-\lambda(g^*-\varepsilon)} \sum_{y\in S} p_{x\,y}(f_{g^*}(x)) e^{\lambda[H(y)+H_\varepsilon(y)]}, \quad x \in S.$$

Since $H_\varepsilon(z) + H(z) = 0$, by (5.15) and (5.19), conditions (a) and (b) in Lemma 4.2 are valid with $H_\varepsilon + H$, f_{g^*} and $g^* - \varepsilon$ instead of H, f and g, respectively, so that conclusion (ii) of Lemma 4.2 yields that $H_\varepsilon(\cdot) + H(\cdot) \geq h_{g^*-\varepsilon}(\cdot)$; in particular, $0 = H_\varepsilon(z) + H(z) \geq h_{g^*-\varepsilon}(z)$, so that $g^* - \varepsilon \in G$, an inclusion that contradicts (3.3), since $\varepsilon > 0$. Consequently, $\Delta = 0$. □

The following result will be useful to establish the uniqueness of bounded solutions of the λ-sensitive optimality equation (2.4).

Lemma 5.3 *Suppose that Assumption 2.3 holds, and let $f \in \mathbb{F}$, $g \in \mathbb{R}$ and H, $H_1 \in \mathcal{B}(S)$ be such that*

$$H(z) = H_1(z) = 0.$$

Additionally, assume that for some functions D, $D_1 : S \to \mathbb{R}$, the following Poisson equations equations hold:

$$\begin{aligned} e^{\lambda g+\lambda H(x)} &= e^{\lambda D(x)} \sum_{y\in S} p_{x\,y}(f(x)) e^{\lambda H(y)} \quad x \in S, \\ e^{\lambda g+\lambda H_1(x)} &= e^{\lambda D_1(x)} \sum_{y\in S} p_{x\,y}(f(x)) e^{\lambda H_1(y)} \quad x \in S. \end{aligned}$$

In this context,

$$\text{if } D_1(\cdot) \geq D(\cdot), \text{ then } D(\cdot) = D_1(\cdot) \text{ and } H(\cdot) = H_1(\cdot).$$

Proof Assume that $D_1(\cdot) \geq D(\cdot)$ and let the stochastic matrix $Q = [q_{x\,y}]$ be as in (5.7). Next, observe that for each $x \in S$

$$\begin{aligned} e^{\lambda H_1(x)} &= e^{\lambda D_1(x)-\lambda g} \sum_{y\in S} p_{x\,y}(f(x)) e^{\lambda H_1(y)} \\ &= e^{\lambda H(x)} e^{\lambda D_1(x)-\lambda D(x)} \sum_{y\in S} \frac{e^{\lambda D(x)-\lambda g} p_{x\,y}(f(x)) e^{\lambda H(y)}}{e^{\lambda H(x)}} e^{\lambda H_1(y)-H(y)} \\ &= e^{\lambda H(x)} e^{\lambda D_1(x)-\lambda D(x)} \sum_{y\in S} q_{x\,y} e^{\lambda H_1(y)-H(y)} \end{aligned}$$

and then

$$e^{\lambda H_1(x)-\lambda H(x)} = e^{\lambda D_1(x)-\lambda D(x)} \sum_{y\in S} q_{x\,y} e^{\lambda H_1(y)-H(y)} \geq \sum_{y\in S} q_{x\,y} e^{\lambda H_1(y)-H(y)}, \quad x \in S; \tag{5.22}$$

recall the $D_1(\cdot) \geq D(\cdot)$ for the inequality. From this point an induction argument yields that that the inequality $e^{\lambda H_1(x)-\lambda H(x)} \geq E_x^Q[e^{\lambda H_1(X_n)-\lambda H(x_n)}]$ always holds, so that

$$\begin{aligned} e^{\lambda H_1(x)-\lambda H(x)} &\geq \frac{1}{n} \sum_{t=1}^{n} E_x^Q[e^{\lambda H_1(X_t)-\lambda H(x_t)}] \\ &\geq \frac{1}{n} \sum_{t=1}^{n} E_x^Q[e^{\lambda H_1(X_t)-\lambda H(x_t)} I[X_t \in \tilde{F}]], \qquad x \in S, \quad n = 1, 2, 3, \end{aligned}$$

where $\tilde{F}$ is a finite subset of S. Since the transition matrix $Q = [q_{x\,y}]$ satisfies the simultaneous Doeblin condition at each state y, by Lemma 5.2(ii), it follows that Q that it is communicating and positive recurrent (see Remark 2.3), and then there exists a unique probability distribution $\mu(\cdot)$ on S such that $\mu(x) > 0$ and $\sum_{t=1}^{n} E_x^Q[R(X_t)]/n \to \sum_{y\in S} \mu(y) R(y)$ for every $x \in S$ and $R \in \mathcal{B}(S)$. Therefore, the above display yields that $e^{\lambda H_1(x)-\lambda H(x)} \geq \sum_{y\in\tilde{F}} \mu(y) e^{\lambda H_1(y)-\lambda H(y)}$ and, since $x \in S$ and the finite set $\tilde{F} \subset S$ are arbitrary, it follows that the relations

$$e^{\lambda H_1(x)-\lambda H(x)} \geq \sum_{y\in S} \mu(y) e^{\lambda H_1(y)-\lambda H(y)} \geq \inf_{y\in S} e^{\lambda H_1(y)-\lambda H(y)}$$

hold for every $x \in S$. After taking the infimum with respect to x, this implies that

$$\sum_{y\in S} \mu(y) e^{\lambda H_1(y)-\lambda H(y)} = \inf_{y\in S} e^{\lambda H_1(y)-\lambda H(y)},$$

and then $e^{\lambda H_1(\cdot)-\lambda H(\cdot)}$ is constant, since $\mu(\cdot) > 0$. Thus, $H_1(\cdot) - H(\cdot) = H_1(z) - H(z) = 0$ which, via (5.22), yields that $D_1(\cdot) - D(\cdot) = 0$. □

6 Proof of Theorems 3.1 and 3.2

In this section the main results of the paper stated in Sect. 3 will be finally established.

Proof of Theorem 3.1 Suppose that Assumptions 2.1, 2.3 and 2.4, as well as condition (3.5) for some finite set $F \subset S$ hold.

(i) The inclusion $g^* \in G$ was proved in Lemma 4.3, while the boundedness of $h_{g^*}(\cdot)$ and the equality $h_{g^*}(z) = 0$ were established in Lemma 5.1 and Theorem 5.1, respectively.

(ii) Since $h_{g^*}(z) = 0$, the dynamic programming equation (4.2) applied to the case $g = g^*$ is equivalent to

$$e^{\lambda g^* + \lambda h_{g^*}(x)} = \inf_{a \in A(x)} \left[e^{\lambda C(x,a)} \sum_{y \in S} p_{x\,y}(a) e^{\lambda h_{g^*}(y)} \right] \quad x \in S,$$

so that the pair $(g^*, h_{g^*}(\cdot))$ satisfies the optimality equation (2.4).

(iii) Since $h_{g^*}(\cdot)$ is bounded, the previous part and Lemma 2.1(i) together imply that $J_C^*(\cdot) = g^*$.

(iv) It will be shown that $f \in \mathbb{F}$ is λ-optimal if and only if (3.6) holds.
Assume that f is λ-optimal, that is $g^* = J_C(f, \cdot)$. In this case, applying parts (i) and (ii) above to the reduced MDP $\mathcal{M}_f$ obtained by restricting the set of admissible actions at each state x to $\{f(x)\}$, it follows that there exists a bounded function $h_f(x)$ such that $h_f(z) = 0$ and

$$e^{\lambda g^* + \lambda h_f(x)} = e^{\lambda C(x, f(x))} \sum_{y \in S} p_{x\,y}(f(x)) e^{\lambda h_f(y)} \quad x \in S.$$

On the other hand, from the optimality equation in part (ii) it follows that there exists $\delta \colon S \to [0, \infty)$ such that

$$e^{\lambda g^* + \lambda h_{g^*}(x)} = e^{\lambda C(x, f(x)) - \lambda \delta(x)} \sum_{y \in S} p_{x\,y}(f(x)) e^{\lambda h_{g^*}(y)} \quad x \in S. \tag{6.1}$$

From this point, an application of Lemma 5.3 yields that $\delta(\cdot) = 0$, so that (3.6) holds.
Assume that (3.6) is valid. In this situation the λ-optimality of f follows from the second part of Lemma 2.1.

(v) Let $g \in \mathbb{R}$ and $h \in \mathcal{B}(S)$ be such that (2.4) holds. In this case $g = g^*$, by Lemma 2.1(i), and there exists a policy $f \in \mathbb{F}$ such that

$$e^{\lambda g^* + \lambda \tilde{h}(x)} = e^{\lambda C(x, f(x))} \sum_{y \in S} p_{x\,y}(f(x)) e^{\lambda \tilde{h}(y)} \quad x \in S,$$

where $\tilde{h}(\cdot) = h(\cdot) - h(z)$; see Remark 2.2(i). For this policy f the optimality equation in part (ii) implies that (6.1) holds for some $\delta \geq 0$ and, since $h_{g^*}(z) = \tilde{h}(z) = 0$, via Lemma 5.3 the above display and (6.1) together imply that $h^*(\cdot) = \tilde{h}(\cdot) = h(\cdot) - h(z)$. □

The results in Theorem 3.2 will be derived from Theorem 3.1 by approximating an arbitrary nonnegative cost function by functions with compact support.

Proof of Theorem 3.2 Suppose that Assumptions 2.1–2.4 hold.

(i) Let $\{F_k\}$ be a sequence of finite subsets of S such that

$$F_k \nearrow S \quad \text{as} \quad k \nearrow \infty, \tag{6.2}$$

and define the sequence of cost functions $\{C_k : S \times \mathbb{K} \to [0, \infty)\}$ as follows: for each $(x, a) \in \mathbb{K}$,

$$C_k(x, a) = C(x, a), \quad x \in F_k, \quad \text{and} \quad C_k(x, a) = 0, \quad x \in S \setminus F_k. \tag{6.3}$$

so that optimal average cost associated with C_k is constant, say

$$J^*_{C_k}(\cdot) = g_k; \tag{6.4}$$

see Theorem 3.1. Observe now that from (6.2) and (6.3) it follows that $0 \leq C_k \nearrow C$, and then $\{J^*_{C_k}(\cdot)\}$ is an increasing sequence which is bounded above by $J^*_C(\cdot)$; see (2.1)–(2.3). Since $J^*_C(z_0)$ is finite for some $z_0 \in S$, by Assumption 2.2, from (6.4) it follows that there exists $\tilde{g} \in [0, \infty)$ such that

$$\lim_{k \to \infty} g_k = \tilde{g} \leq J^*_C(\cdot). \tag{6.5}$$

Now, it is claimed that

$$\tilde{g} \in G. \tag{6.6}$$

Assuming that this assertion holds, the conclusions in part (i) can be obtained as follows: The specification of g^* in (3.3) and the above display together yield that $\tilde{g} \geq g^*$; since $g^* \geq J^*_C(\cdot)$, by Lemma 4.2(v), it follows that

$$\tilde{g} \geq g^* \geq J^*_C(\cdot),$$

relations that together with (6.5) lead to

$$J^*_C(\cdot) = \lim_{k \to \infty} g_k = g^* \in G.$$

On the other hand, for each positive integers n and k, form the inequality $C \geq C_k$ it follows that $J_{C,n}(\pi, x) \geq J_{C_k,n}(\pi, x)$ for every $x \in S$ and $\pi \in \mathcal{P}$, so that

$$\liminf_{n\to\infty} \frac{1}{n} J_{C,n}(\pi, x) \geq \liminf_{n\to\infty} \frac{1}{n} J_{C_k,n}(\pi, x) \geq g_k,$$

where, using that the λ-optimality equation associated with C_k has a bounded solution—by Theorem 3.1—the second inequality is due to Remark 2.2(ii); letting k increase to ∞ it follows that

$$\liminf_{n\to\infty} \frac{1}{n} J_{C,n}(\pi, x) \geq g^*,$$

as stipulated in part (i). To conclude, the inclusion in (6.6) will be established. Since C_k has compact support, from Theorem 3.1 it follows that there exist a function $h_k \in \mathcal{B}(S)$ and a policy $f_k \in \mathbb{F}$ such that

$$h_k(z) = 0 \quad \text{and} \quad e^{\lambda g_k + \lambda h_k(x)} = e^{\lambda C(x, f_k(x))} \sum_{y\in S} p_{x\,y}(f_k(x)) e^{\lambda h_k(y)}, \quad x \in S. \tag{6.7}$$

On the other hand, since $\mathbb{F}$ and $[-\infty, \infty]$ are compact metric spaces, taking a subsequence, if necessary, it can be assumed that there exists a policy $\tilde{f} \in \mathbb{F}$ and $\tilde{h}\colon S \to [-\infty, \infty]$ such that, for each $x \in S$,

$$\lim_{k\to\infty} f_k(x) = \tilde{f}(x) \in A(x) \quad \text{and} \quad \lim_{k\to\infty} h_k(x) = \tilde{h}(x) \in [-\infty, \infty],$$
$$\text{where } \tilde{h}(z) = 0.$$

Taking the limit inferior as k goes to ∞ in both sides of the equality in (6.7), the above display and the convergence in (6.5) together imply, via Assumption 2.1(ii) and Fatou's lemma, that

$$e^{\lambda \tilde{g} + \lambda \tilde{h}(x)} \geq e^{\lambda C(x, \tilde{f}(x))} \sum_{y\in S} p_{x\,y}(\tilde{f}(x)) e^{\lambda \tilde{h}(y)}, \quad x \in S;$$

since $\tilde{h}(z) = 0$, it follows from Lemma 4.2(ii) that $1 = e^{\lambda \tilde{h}(z)} \geq e^{\lambda h_{\tilde{g}}(z)}$, and then $h_{\tilde{g}}(z) \leq 0$, that is, $\tilde{g} \in G$; see (3.2).

(ii) By Lemma 4.1(i),

$$h_{g^*}(\cdot) \geq M_{g^*} = -N(z)g^* + \log(1 - \rho(z))/\lambda, \tag{6.8}$$

where the notation is as in Lemma 2.2. On the other hand, since $1 \geq e^{h_{g^*}(z)}$—by the inclusion $g^* \in G$ in the part (i)—the inequality (3.10) follows from the dynamic programming equation established in Lemma 4.1(ii).

(iii) By Remark 2.2(ii), there exists a policy $f \in \mathbb{F}$ such that, for each $x \in S$, $f(x)$ minimizes the mapping

$$a \mapsto e^{\lambda C(x,a)} \sum_{y \in S} p_{x\,y}(a) e^{h_{g^*}(y)}, \quad a \in A(x),$$

and then

$$e^{\lambda g^* + \lambda h_{g^*}(x)} \geq e^{\lambda C(x,f(x))} \sum_{y \in S} p_{x\,y}(f(x)) e^{h_{g^*}(y)}, \quad x \in S.$$

by the previous part. Now, let f be an arbitrary stationary policy satisfying this relation. Using an induction argument, it follows that the inequality

$$e^{n\lambda g^* + \lambda h_{g^*}(x)} \geq E_x^f \left[e^{\lambda \sum_{t=0}^{n-1} C(X_t, A_t) + \lambda h_{g^*}(X_n)} \right]$$

always holds, and then (6.8) yields that

$$e^{n\lambda g^* + \lambda h_{g^*}(x)} \geq E_x^f \left[e^{\lambda \sum_{t=0}^{n-1} C(X_t, A_t)} \right] e^{\lambda M_{g^*}} = e^{\lambda J_{C,n}(f,x) + \lambda M_{g^*}}, \quad x \in S;$$

see (2.1). Thus, $g^* + (h_{g^*}(x) - M_{g^*})/n \geq J_{C,n}(f,x)/n$ and, since h_{g^*} is finite, (2.1), (2.3) and the part (i) together yield that $J_C^*(\cdot) = g^* \geq J_C(f,\cdot) \geq J_C^*(\cdot)$, so that f is λ-optimal.

(iv) Let $g \in \mathbb{R}$ and $h : S \to \mathbb{R}$ be such that (3.12) holds and, using Remark 2.2(i), select a policy $f \in \mathbb{F}$ such that

$$e^{\lambda g + \lambda h(x)} \geq e^{\lambda C(x,f(x))} \sum_{y \in S} p_{x\,y}(f(x)) e^{\lambda h(y)}, \quad x \in S.$$

After multiplying both sides of this inequality by $e^{-\lambda g - \lambda h(z)}$ and setting

$$\tilde{h}(\cdot) = h(\cdot) - h(z),$$

it follows that

$$\begin{aligned} e^{\lambda \tilde{h}(x)} &\geq e^{\lambda(C(x,f(x))-g)} \sum_{y \in S} p_{x\,y}(f(x)) e^{\lambda \tilde{h}(y)} \\ &= e^{\lambda(C(x,f(x))-g)} \left(p_{x\,z} e^{\lambda \tilde{h}(z)} + \sum_{y \in S \setminus \{z\}} p_{x\,y}(f(x)) e^{\lambda \tilde{h}(y)} \right) \\ &= e^{\lambda(C(x,f(x))-g)} \left(p_{x\,z} + \sum_{y \in S \setminus \{z\}} p_{x\,y}(f(x)) e^{\lambda \tilde{h}(y)} \right), \quad x \in S. \end{aligned}$$

Since $\tilde{h}(\cdot)$ is a finite function, an application of Lemma 4.2(ii) yields that

$$\tilde{h}(x) \geq h_g(x), \quad x \in S. \tag{6.9}$$

In particular, $0 = \tilde{h}(z) \geq h_g(z)$, so that $g \in G$ and then

$$g \geq g^*;$$

see (3.2) and (3.3). These two last displays show that one of the following alternatives hold: (a) $g > g^*$, or (b) $g = g^*$ and $h(\cdot) - h(z) = \tilde{h}(\cdot) \geq h_{g^*}(\cdot)$, completing the proof. □

7 An example on the optimality inequality and the optimality equation

As already mentioned, the main qualitative difference between Theorems 3.1 and 3.2 is that Theorem 3.1 ensures that the pair $(g^*, h_{g^*}(\cdot))$ satisfies the the optimality *equation* (2.4) when $C(\cdot, \cdot) \geq 0$ has compact support, while for a nonnegative cost function satisfying Assumption 2.2, Theorem 3.2 asserts that the *inequality* (3.10) holds. The objective of the section is to provide an example to show that,

(a) If the cost function is nonnegative but does not have compact support, then the strict inequality in (3.10) may occur for each state x, even if $C(\cdot, \cdot)$ is bounded, and
(b) If the cost function has compact support but assumes negative values, then the λ-optimality equation (2.4) may fail at each state x, so that, in general, the nonnegativity of the cost function is necessary to ensure that the conclusion of Theorem 3.1(iv) holds.

Example 7.1 On the state space $S = \mathbb{N}$, consider the transition matrix $P = [p_{x\,y}]$ determined by

$$p_{x\,x+1} := \left(\frac{x+1}{x+2}\right)^2 p, \qquad p_{x\,0} := 1 - p_{x\,x+1}, \quad x = 0, 1, 2, 3, \ldots \tag{7.1}$$

where

$$p \in (0, 1)$$

is fixed. If $C : S \to \mathbb{R}$ is bounded function, then the triple (S, P, C) is naturally identified with an MDP in which the action set A is a singleton, and Assumptions 2.1 and 2.2 automatically hold. Since

$$P_x[T_0 = 1] = 1 - \left(\frac{x+1}{x+2}\right)^2 \quad p > 1 - p > 0,$$

the simultaneous Doeblin condition at

$$z = 0$$

is valid, and combining this fact with

$$\begin{aligned} P_0[X_1 = 0] &= 1 - P_0[X_1 = 1] = 1 - p/4 \\ P_0[X_n = n] &= P_0[X_k = k,\ k = 1, 2, \ldots n] \\ &= \prod_{k=1}^{n} p_{k-1\,k} = \prod_{k=1}^{n} \left[\left(\frac{k}{k+1} \right)^2 p \right] = \frac{p^n}{(n+1)^2}, \quad n = 1, 2, 3, \ldots, \end{aligned} \tag{7.2}$$

it is not difficult to see that Assumption 2.4 is also verified.

For the model in Example 7.1, (3.10) becomes

$$e^{\lambda g^* + \lambda h_{g^*}(x)} \geq e^{\lambda C(x)} \left[p_{x\,0} e^{\lambda h_{g^*}(x)} + p_{x\,x+1} e^{\lambda h_{g^*}(x+1)} \right], \quad x \in S, \tag{7.3}$$

and in the following proposition a bounded cost function will be specified so that the strict inequality always occurs in this relation.

Proposition 7.1 *In the context of Example 7.1*

(i) $E_0 \left[e^{-\log(p)(T_0 - 1)} \right] \in (1, \infty)$.

(ii) *Set*

$$\alpha := \log(E_0 \left[e^{-\log(p)(T_0 - 1)} \right] \in (0, \infty), \tag{7.4}$$

where the inclusion follows from part(i), and define $C \in \mathcal{B}(S)$ *by*

$$C(z) = C(0) := 0, \qquad C(x) := \frac{-\log(p) + 2\alpha}{\lambda}, \quad x \in S \setminus \{0\}. \tag{7.5}$$

In this framework,

$$g^* = \frac{2\alpha}{\lambda} \quad \textit{and} \quad h_{g^*}(z) = h_{g^*}(0) = -\frac{\alpha}{\lambda} < 0. \tag{7.6}$$

Consequently,

(iii) *For the cost function in (7.5), the strict inequality holds in (7.3) for each* $x \in S$.

Proof (i) For each $n \in \mathbb{N}$, Definition 2.1(i) and the specification of the transition law together yield that $P_0[T_0 > n] = P_0[X_n = n]$. Using that

$$P_0[T = n] = P_0[T_0 > n-1] - P_0[T_0 > n] = P_0[X_{n-1} = n-1] - P_0[X_n = n]$$

for a positive integer n, it follows that

$$P_0[T_0 = n] = \frac{p^{n-1}}{n^2}\left[1 - \left(\frac{n}{n+1}\right)^2 p\right], \quad n = 1, 2, 3, \ldots; \tag{7.7}$$

see (7.2). Observe now that $E_0\left[e^{-\log(p)(T_0-1)}\right] > 1$, since $P_0[T_0 = 1] < 1$ and $-\log(p) > 0$. To show that the expectation is finite, notice that $P_0[T_0 < \infty] = 1$, by Remark 2.3, so that

$$\begin{aligned} E_0\left[e^{-\log(p)(T_0-1)}\right] &= \sum_{n=1}^{\infty} e^{-\log(p)(n-1)} P_0[T_0 = n] \\ &= \sum_{n=1}^{\infty} \frac{1}{p^{n-1}} P_0[T_0 = n] \\ &= \sum_{n=1}^{\infty} \frac{1}{n^2}\left[1 - \left(\frac{n}{n+1}\right)^2 p\right] < \infty. \end{aligned} \tag{7.8}$$

(ii) Given $g \in \mathbb{R}$ write

$$g = \frac{2\alpha + t}{\lambda}. \tag{7.9}$$

Since $X_n \neq 0$ for $1 \leq n < T_0$, using (7.5) it follows that

$$\lambda \sum_{n=1}^{T_0-1} (C(X_n) - g) = \lambda \sum_{n=1}^{T_0-1} \left(\frac{-\log(p) + 2\alpha}{\lambda} - \frac{2\alpha + t}{\lambda}\right) = -(T_0 - 1)[\log(p) + t].$$

Thus,

$$X_0 = 0 \Rightarrow \lambda \sum_{n=0}^{T_0-1} (C(X_n) - g) = -\lambda g - (T_0 - 1)[\log(p) + t]$$

and then

$$\begin{aligned} e^{\lambda h_g(0)} &= E_0\left[e^{\lambda \sum_{n=0}^{T_0-1} (C(X_n)-g)}\right] \\ &= E_0\left[e^{-\lambda g - (T_0-1)[\log(p)+t]}\right] = e^{-\lambda g} E_0\left[e^{-(T_0-1)[\log(p)+t]}\right] \end{aligned}$$

$$= e^{-\lambda g}\sum_{n=1}^{\infty} e^{-(n-1)[\log(p)+t]} P_0[T_0 = n]$$
$$= e^{-\lambda g}\sum_{n=1}^{\infty} e^{-(n-1)[\log(p)+t]} \frac{p^{n-1}}{n^2}\left[1 - \left(\frac{n}{n+1}\right)^2 p\right]$$
$$= e^{-\lambda g}\sum_{n=1}^{\infty} \frac{e^{-(n-1)t}}{n^2}\left[1 - \left(\frac{n}{n+1}\right)^2 p\right]. \tag{7.10}$$

This relation and (7.9) show that

$$h_g(0) < \infty \iff t \geq 0 \iff g \geq \frac{2\alpha}{\lambda},$$

and, since $t = 0$ corresponds to $g = 2\alpha/\lambda$, via (7.4) and (7.8), from (7.10) it follows that

$$g = \frac{2\alpha}{\lambda} \Rightarrow e^{\lambda h_g(0)} = e^{-2\alpha}\sum_{n=1}^{\infty} \frac{1}{n^2}\left[1 - \left(\frac{n}{n+1}\right)^2 p\right]$$
$$= e^{-2\alpha} E_0\left[e^{-\log(p)(T_0-1)}\right] = e^{-2\alpha} e^{\alpha} = e^{-\alpha} < 1,$$

and (7.6) follows from these two last displays.
(iii) For the present example the relative value function is determined by

$$e^{\lambda h_{g^*}(x)} = E_0\left[e^{\lambda \sum_{t=0}^{T_0-1}(C(X_t)-g^*)}\right], \tag{7.11}$$

and Lemma 4.1(ii) yields that

$$e^{\lambda h_{g^*}(x)} = e^{\lambda(C(x)-g^*)}\left[p_{x\,0} + p_{x\,x+1} e^{\lambda h_{g^*}(x+1)}\right], \quad x \in S;$$

since $h_{g^*}(0) < 0$ and $p_{x\,0} > 0$ for each $x \in S$, by the previous part and (7.1), respectively, it follows that

$$e^{\lambda h_{g^*}(x)} > e^{\lambda(C(x)-g^*)}\left[p_{x\,0} e^{\lambda h_{g^*}(0)} + p_{x\,x+1} e^{\lambda h_{g^*}(x+1)}\right], \quad x \in S, \tag{7.12}$$

completing the proof. □

Now, using the notation in (7.4)–(7.6) define the cost function $C_1 : S \to \mathbb{R}$ by

$$C_1(x) := 0, \quad x \in S \setminus \{0\}, \quad \text{and} \quad C_1(z) = C_1(0) := \frac{\log(p) - 2\alpha}{\lambda} \equiv \beta < 0, \tag{7.13}$$

so that

$$C_1(\cdot) = C(\cdot) + \beta, \tag{7.14}$$

and then

$$g_1^* = J_{C_1}^*(\cdot) = J_C^*(\cdot) + \beta = g^* + \beta. \tag{7.15}$$

Therefore, $C_1(\cdot) - g_1^* = C(\cdot) - g^*$ and the relative value function $h_{g_1^*}^1$ associated with C_1 and g_1^* satisfies $e^{\lambda h_{g_1^*}^1(x)} = E_x[e^{\lambda \sum_{t=0}^{T_0-1}(C_1(X_t)-g_1^*)}] = E_x[e^{\lambda \sum_{t=0}^{T_0-1}(C(X_t)-g^*)}] = e^{\lambda h_{g^*}(x)}$; see (7.11). Thus,

$$h_{g_1^*}^1(\cdot) = h_{g^*}(\cdot). \tag{7.16}$$

Using the cost function C_1 above, via Proposition 7.1 it is shown below that the λ-optimality Equation (2.4) does not necessarily hold if the cost function takes a negative value, and all the other conditions in Theorem 3.1 hold.

Proposition 7.2 *In the context of Example 7.1, let the function $C_1(\cdot)$ be as in (7.13), so that C_1 has compact support and takes a negative value. In this case, the optimal average cost $g_1^* = J_{C_1}^*(\cdot)$, and the relative value function $h_{g_1^*}^1(\cdot)$ associated with $C_1(\cdot)$ and g_1^*, satisfy the following relation:*

$$e^{\lambda g_1^* + \lambda h_{g_1^*}^1(x)} > e^{\lambda C_1(x)} \left[p_{x\,0} e^{\lambda h_{g_1^*}^1(0)} + p_{x\,x+1} e^{\lambda h_{g_1^*}^1(x+1)} \right], \quad x \in S. \tag{7.17}$$

Proof Using (7.14)–(7.16), it follows that the above relation is equivalent to the inequality (7.12) established in Proposition 7.1. □

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

Borkar VS, Meyn SP (2002) Risk-sensitive optimal control for Markov decison process with monotone cost. Math Oper Res 27:192–209

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

Cavazos-Cadena R, Fernández-Gaucherand E (1999) Controlled Markov chains with risk-sensitive criteria: average cost, optimality equations and optimal solutions. Math Meth 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 (2004) A characterization of exponential functionals in finite Markov chains. Math Methods Oper Res 60:399–414

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 properrty. 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-Hernández D, Marcus SI (1999) Existence of risk-sensitive optimal stationary policies for controlled Markov processes. Appl Math Optim 40:273–285
Hernández-Lerma O (1988) Adaptive Markov control processes. Springer, New York
Howard AR, Matheson JE (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 Contr 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
Loève M (1980) Probability theory I. Springer, New York
Puterman ML (1994) Markov decision processes. Wiley, New York
Seneta E (1980) Nonnegative matrices. Springer, New York
Sennot L (1986) A new condition for the existence of optimum stationary policies in average cost Maarkov decision processes. Oper Res Lett 5:17–23
Sennot L (1995) Another set of conditions for average optimality in Maarkov control processes. Syst Control Lett 24:147–151
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