CHARACTERIZATION OF THE OPTIMAL AVERAGE COST IN MARKOV DECISION CHAINS DRIVEN BY A RISK-SEEKING CONTROLLER

ROLANDO CAVAZOS-CADENA,* *Universidad Autónoma Agraria Antonio Narro*
HUGO CRUZ-SUÁREZ (iD),** *Benemérita Universidad Autónoma de Puebla*
RAÚL MONTES-DE-OCA,*** *Universidad Autónoma Metropolitana-Iztapalapa*

Abstract

This work concerns Markov decision chains on a denumerable state space endowed with a bounded cost function. The performance of a control policy is assessed by a long-run average criterion as measured by a *risk-seeking* decision maker with constant risk-sensitivity. Besides standard continuity–compactness conditions, the framework of the paper is determined by the following conditions: (i) the state process is communicating under each stationary policy, and (ii) the simultaneous Doeblin condition holds. Within this framework it is shown that (i) the optimal superior and inferior limit average value functions coincide and are constant, and (ii) the optimal average cost is characterized via an extended version of the Collatz–Wielandt formula in the theory of positive matrices.

Keywords: Risk-lover controller; exponential utility; optimality inequality; extended Collatz–Wielandt formula; optional sampling theorem; truncated cost function

2020 Mathematics Subject Classification: Primary 93E20
Secondary 93C55

1. Introduction

This work is concerned with Markov decision processes (MDPs) on a denumerable state space endowed with a bounded cost function. It is assumed that *the decision maker is risk-seeking* with a constant risk-sensitivity coefficient, and the performance of a control policy is measured by a long-run average cost index. In addition to standard continuity–compactness conditions, the framework of the paper is mainly determined by two assumptions on the transition structure: (a) the state process is communicating under each stationary policy, and (b) the simultaneous Doeblin condition holds. In this context, the main objectives of this work can be described as follows:

(i) to prove that the optimal inferior and superior limit average cost functions coincide and are constant;

(ii) to characterize the optimal average cost.

The results for these problems extend to the present framework conclusions established in [7] under the condition that the decision maker is risk-averse and, as in that paper, the characterization of the optimal average cost presented in Theorem 1(ii) is a generalized version of the Collatz–Wielandt formula for the largest eigenvalue of a positive matrix. On the other hand, an additional objective of the paper is the following:

(iii) to establish the existence of a stationary policy whose average index differs from the optimal one by less than a given tolerance.

The analysis of discrete-time Markov models endowed with a risk-sensitive average criterion can be traced back, at least, to the seminal paper [14]. In that paper the Perron–Frobenius theory of positive matrices [17] was employed to study MDPs with finite state and action spaces, and the optimal average cost was characterized via an optimality equation. On the other hand, motivated by important connections with the theory of large deviations and mathematical finance, risk-sensitive average criteria have recently been studied. Models with a countable state space are considered in [5, 8, 9, 20], whereas MDPs on a general state space are analyzed in [10, 11, 12, 15]. Connections of risk-sensitive average criteria with the theory of large deviations are presented in [2, 16], whereas applications to mathematical finance are given in [3, 18, 21].

On the other hand, there are important differences between the risk-neutral and risk-sensitive average criteria. For instance, in the risk-neutral case the simultaneous Doeblin condition ensures that the optimal average cost is constant and is characterized via an optimality equation [13, 19], a result that is not generally valid in the risk-sensitive case [6]. For this reason, in this work the simultaneous Doeblin condition will be supplemented with the requirement that the state space is communicating under the action of any stationary policy.

The remainder of the paper is organized as follows: In Section 2 the decision model is formally introduced, the basic assumptions of the paper are formulated, and the risk-sensitive average criteria are defined. Next, in Section 3 the idea of a subsolution of the optimality equation is formulated and the main conclusions of the paper are stated in Theorem 1, which is followed by a brief outline of the strategy used to prove that result. Section 4 contains the basic properties of the family of subsolutions which allow us to establish the fundamental auxiliary result in Theorem 2 of Section 5, establishing that if the space of stationary policies is finite and the cost function is nonpositive with compact support, then the optimality equation has a bounded solution. After those preliminaries, Theorem 1 is finally proved in Section 6, and the paper concludes with some brief comments in Section 7.

1.1. Notation

In the following, the set of nonnegative integers is denoted by $\mathbb{N}$, and, for a real-valued function h, $\|h\| := \sup\{|h(x)| \colon$ belongs to the domain of $h\}$ is the corresponding supremum norm. Given a nonempty set S, the space of all bounded real-valued functions defined on S is denoted by $\mathcal{B}(S)$, i.e. $h \colon S \to \mathbb{R}$ belongs to $\mathcal{B}(S)$ if and only if $\|h\| < \infty$. On the other hand, for real numbers a and b, $a \wedge b = \min\{a, b\}$, and, if $F \subset S$, the indicator function of the subset F is denoted by $\mathbf{1}_F$, that is, $\mathbf{1}_F(y) := 1$, $y \in F$, and $\mathbf{1}_F(y) = 0$, $y \in S \setminus F$. For an event W, the corresponding

indicator function is denoted by $\mathbf{1}[W]$, and every relation between random variables holds almost surely with respect to the underlying probability measure.

2. Decision model

Throughout, $\mathcal{M} = (S, A, \{A(x)\}_{x \in S}, C, P)$ represents an MDP, a model for a dynamical system whose components are as follows:

(i) The state space S is a *denumerable* set endowed with the discrete topology.

(ii) The control set A is a *metric* space.

(iii) For each state $x \in S$, the nonempty set $A(x) \subset A$ is the class of possible actions at x.

(iv) $C \colon \mathbb{K} \to \mathbb{R}$ is the cost function, where $\mathbb{K} := \{(x, a) \mid a \in A(x), x \in S\}$ is the collection of admissible pairs.

(v) $P = [p_{x,y}(\cdot)]$ 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 a decision maker (controller) chooses and applies an action $A_t = a \in A(x)$. As a consequence of such an intervention, a cost $C(X_t, A_t) = C(x, a)$ is incurred and, regardless of the previous states and actions, the state of the system at time $t + 1$ will be $X_{t+1} = y \in S$ with probability $p_{X_t, X_{t+1}}(A_t) = p_{x,y}(a)$, where $\sum_{y \in S} p_{x,y}(a) = 1$; this is the Markov property of the decision model.

Assumption 1.

(i) *For every $x \in S$, $A(x)$ is a compact subset of A.*

(ii) *For each $x, y \in S$, the mappings $a \mapsto p_{x,y}(a)$ and $a \mapsto C(x, a)$ are continuous in $a \in A(x)$.*

(iii) *The cost function C is bounded, i.e. $C \in \mathcal{B}(\mathbb{K})$.*

The observed history of the process up to time n is denoted by

$$H_n := (X_0, A_0, X_1, A_1, \ldots, X_{n-1}, A_{n-1}, X_n), \qquad n = 0, 1, 2, 3, \ldots, \tag{1}$$

whereas

$$\mathcal{F}_n := \sigma(H_n) \tag{2}$$

is the σ-field generated by H_n.

2.1. Policies

A policy is a measurable rule for choosing actions. Formally, a control policy $\pi = \{\pi_n\}_{n \in \mathbb{N}}$ is a (special) sequence of stochastic kernels π_n on the action space A given $\mathbb{H}_n$, where the space $\mathbb{H}_n$ of possible histories up to time n is given by $\mathbb{H}_0 = S$, and $\mathbb{H}_n = \mathbb{K}^n \times S$ for $n \geq 1$; the vector $\mathbf{h}_n = (x_0, a_0, \ldots, x_{n-1}, a_{n-1}, x_n)$ represents a generic element of $\mathbb{H}_n$, so that $a_i \in A(x_i)$ for $i < n$ and $x_i \in S$ for $i \leq n$. When the controller drives the system using policy $\pi = \{\pi_n\}$, given that the history $\mathbb{H}_n$ attains the value $\mathbf{h}_n \in \mathbb{H}_n$, the probability of choosing action A_n inside a Borel subset B of $A(x_n)$ is $\pi_n(B \mid \mathbf{h}_n)$, where $\pi_n(A(x_n) \mid \mathbf{h}_n) = 1$. The family of all policies is denoted by $\mathcal{P}$ and, given the initial state $X_0 = x$ and the policy $\pi \in \mathcal{P}$ used to direct the system, the distribution $\mathbb{P}_x^\pi$ of the state–action process $\{(X_k, A_k)\}_{k \in \mathbb{N}}$ is uniquely determined [13, 19] and $\mathbb{E}_x^\pi[\,\cdot\,]$ represents the corresponding expectation operator. Next, set $\mathbb{F} := \prod_{x \in S} A(x)$, so that $\mathbb{F}$

is a compact metric space consisting of all functions $f \colon S \to A$ such that $f(x) \in A(x)$ for every $x \in S$. A policy π is stationary if there exists $f \in \mathbb{F}$ such that, under π, at each time $t \in \mathbb{N}$ the action $A_t = f(X_t)$ is applied. The class of stationary policies and $\mathbb{F}$ are naturally identified, a convention allowing us to write $\mathbb{F} \subset \mathcal{P}$.

2.2. Utility function and average criteria

In this paper we suppose that the controller has a constant risk-sensitivity coefficient $\lambda \neq 0$, so that a random cost Y is assessed via the expected value of $U_\lambda(Y)$, where the (dis-)utility function U_λ is the strictly increasing mapping given by

$$U_\lambda(x) = \operatorname{sign}(\lambda)e^{\lambda x}, \qquad x \in \mathbb{R}; \tag{3}$$

notice that

$$U_\lambda(c + h) = e^{\lambda c}U_\lambda(h), \qquad c, h \in \mathbb{R}. \tag{4}$$

When the decision maker has to choose between two random costs Y_0 and Y_1, paying Y_0 will be preferred if $\mathbb{E}[U_\lambda(Y_1)] > \mathbb{E}[U_\lambda(Y_0)]$, whereas the controller will be indifferent between both costs if $\mathbb{E}[U_\lambda(Y_1)] = \mathbb{E}[U_\lambda(Y_0)]$. Given a random cost Y such that $\mathbb{E}[U_\lambda(Y)]$ is finite, the *certainty equivalent* of Y with respect to U_λ is the unique real number $\mathcal{E}_\lambda[Y]$ determined by $U_\lambda(\mathcal{E}_\lambda[Y]) = \mathbb{E}[U_\lambda(Y)]$, so that the controller will be willing to pay the amount $\mathcal{E}_\lambda[Y]$ to avoid facing the uncertainty associated with Y; note that

$$\mathcal{E}_\lambda[Y] = U_\lambda^{-1}(\mathbb{E}[U_\lambda(Y)]) = \frac{1}{\lambda} \log (\mathbb{E}[e^{\lambda Y}]), \tag{5}$$

a relation that immediately leads to

$$\mathbb{P}[|Y| \leq b] = 1 \implies |\mathcal{E}_\lambda(Y)| \leq b. \tag{6}$$

Observe that $U_\lambda(\cdot)$ is strictly convex (resp. concave) if $\lambda > 0$ (resp. $\lambda < 0$), and in that case Jensen's inequality yields that $\mathcal{E}_\lambda(Y) > \mathbb{E}[Y]$ (resp. $\mathcal{E}_\lambda(Y) < \mathbb{E}[Y]$) if Y is a non-constant random variable. When $\lambda > 0$ (resp. $\lambda < 0$) the controller is referred to as *risk-averse* (resp. *risk-seeking*). Using the idea of the certainty equivalent, risk-sensitive average criteria can be defined as follows: Let $\pi \in \mathcal{P}$ be the policy used to operate the system starting at $X_0 = x$. The application of the first n actions $A_0, A_1, \ldots, A_{n-1}$ generates the cost $\sum_{k=0}^{n-1} C(X_k, A_k)$ and, using (5), the associated certainty equivalent is

$$J_n(\pi, x) := \frac{1}{\lambda} \log \left(\mathbb{E}_x^\pi \left[\exp \left\{ \lambda \sum_{t=0}^{n-1} C(X_t, A_t) \right\} \right] \right), \qquad n = 1, 2, 3, \ldots, \tag{7}$$

which represents an average of $J_n(\pi, x)/n$ per step. The superior limit (λ-sensitive) average performance index of policy π at state x is given by

$$J(\pi, x) := \limsup_{n \to \infty} \frac{1}{n} J_n(\pi, x), \tag{8}$$

and the corresponding optimal value function is

$$J^*(x) := \inf_{\pi \in \mathcal{P}} J(\pi, x), \qquad x \in S; \tag{9}$$

a policy $\pi^* \in \mathcal{P}$ is lim sup (λ-)average optimal if $J(\pi^*, x) = J^*(x)$ for every $x \in S$. The criterion (8) represents the worst-case point of view about the behavior of the sequence $\{J_n(\pi, x)/n\}$ of averages over a finite horizon. The optimistic perspective assesses the performance of π via the smallest limit point of that sequence:

$$J_-(\pi, x) := \liminf_{n \to \infty} \frac{1}{n} J_n(\pi, x), \tag{10}$$

with the corresponding optimal value function given by

$$J_*(x) := \inf_{\pi \in \mathcal{P}} J_-(\pi, x), \qquad x \in S, \tag{11}$$

and $\pi_* \in \mathcal{P}$ is lim inf (λ-)average optimal if $J_-(\pi_*, \cdot) = J_*(\cdot)$. It follows that

$$J_-(\pi, \cdot) \leq J(\pi, \cdot), \quad \pi \in \mathcal{P}, \qquad \text{and then } J_*(\cdot) \leq J^*(\cdot). \tag{12}$$

From the definitions in (7)–(11), using (6), it is not difficult to verify that

$$-\|C\| \leq J_-(\pi, \cdot) \leq J(\pi, \cdot) \leq \|C\|, \quad \pi \in \mathcal{P}, \qquad -\|C\| \leq J_*(\cdot) \leq J^*(\cdot) \leq \|C\|.$$

Remark 1. In Section 6, several cost functions are considered simultaneously. In that case, the cost function is explicitly indicated in the above notation for the average criteria and optimal value functions. For instance, $J_*(C, \cdot)$ and $J^*(C, \cdot)$ are used instead of $J_*(\cdot)$ and $J^*(\cdot)$; notice that

$$J_*(C + \beta, \cdot) = J_*(C, \cdot) + \beta, \quad J^*(C + \beta, \cdot) = J^*(C, \cdot) + \beta, \qquad \beta \in \mathbb{R}. \tag{13}$$

2.3. Optimality equation

Under appropriate requirements, the characterization of the optimal value functions in (9) and (11) can be based on the following optimality equation:

$$U_\lambda(g + h(x)) = \inf_{a \in A(x)} \left[\sum_{y \in S} p_{x,y}(a) U_\lambda(C(x, a) + h(y)) \right], \quad x \in S, \tag{14}$$

where $g \in \mathbb{R}$ and $h(\cdot)$ is a function defined on S. Suppose that this equality holds, and set

$$Y_0 = U_\lambda(h(X_0)), \qquad Y_n = U_\lambda \left(\sum_{t=0}^{n-1} (C(X_t, A_t) - g) + h(X_n) \right), \quad n = 1, 2, 3, \ldots \tag{15}$$

Remark 2. Direct consequences of (15) are the following facts:

(i) Let $\pi = \{\pi_n\} \in \mathcal{P}$ be a fixed policy. From (14), it follows that, for every $n \in \mathbb{N}$ and $a \in A(X_n)$, $U_\lambda(g + h(X_n)) \leq \sum_{y \in S} p_{X_n, y}(a) U_\lambda(C(X_n, a) + h(y))$; integrating both sides of this relation with respect to $\pi_n(\mathrm{d}a \mid H_n)$ and using the Markov property, it follows that

$$U_\lambda(h(X_n)) \leq \int_{A(X_n)} \sum_{y \in S} p_{X_n, y}(a) U_\lambda(C(X_n, a) + h(y)) \pi_n(\mathrm{d}a \mid H_n)$$

$$= \mathbb{E}_x^\pi [U_\lambda(C(X_n, A_n) - g + h(X_{n+1})) \mid \mathcal{F}_n], \quad x \in S, \ n \in \mathbb{N}. \tag{16}$$

Next, using that $\exp\left\{\lambda \sum_{t=0}^{n-1}(C(X_t, A_t) - g)\right\}$ is $\mathcal{F}_n$-measurable, (4) and the two previous displays together yield

$$
\begin{aligned}
Y_n &= U_\lambda\left(\lambda \sum_{t=0}^{n-1}(C(X_t, A_t) - g) + h(X_n)\right) \\
&= \exp\left\{\lambda \sum_{t=0}^{n-1}(C(X_t, A_t) - g)\right\} U_\lambda(h(X_n)) \\
&\leq \exp\left\{\lambda \sum_{t=0}^{n-1}(C(X_t, A_t) - g)\right\} \mathbb{E}_x^\pi[U_\lambda(C(X_n, A_n) - g + h(X_{n+1})) \mid \mathcal{F}_n] \\
&= \mathbb{E}_x^\pi\left[\exp\left\{\lambda \sum_{t=0}^{n-1}(C(X_t, A_t) - g)\right\} U_\lambda(C(X_n, A_n) - g + h(X_{n+1})) \mid \mathcal{F}_n\right] \\
&= \mathbb{E}_x^\pi\left[U_\lambda\left(\sum_{t=0}^{n}(C(X_t, A_t) - g) + h(X_{n+1})\right) \mid \mathcal{F}_n\right] = \mathbb{E}_x^\pi[Y_{n+1} \mid \mathcal{F}_n], \qquad (17)
\end{aligned}
$$

that is, $\{(Y_n, \mathcal{F}_n)\}_{n \in \mathbb{N}}$ is a *submartingale* with respect to P_x^π for every $x \in S$ and $\pi \in \mathcal{P}$.

(ii) Suppose that function $h(\cdot)$ in (14) is *bounded*. In this context, using Assumption 1, it is not difficult to see that, for each $x \in S$, the term within brackets in (14) is a continuos function of $a \in A(x)$, and then it has a maximizer $f^*(x) \in A(x)$. Thus, $U_\lambda(g + h(x)) = \mathbb{E}_x^{f^*}[U_\lambda(C(x, f^*(x)) + h(x))]$ for every $x \in S$ and, paralleling the argument in part (i), it follows that the equality always holds in (16) and (17) when π is replaced by f^*, so that $\{(Y_n, \mathcal{F}_n)\}_{n \in \mathbb{N}}$ is a *martingale* with respect to $\mathbb{P}_x^{f^*}$ for every $x \in S$.

(iii) Maintaining the assumption that $h(\cdot)$ is bounded, observe that, for every $n \in \mathbb{N} \setminus \{0\}$, $x \in S$, and $\pi \in \mathcal{P}$, part (i) implies that $\mathbb{E}_x^\pi[Y_0] \leq \mathbb{E}_x^\pi[Y_n]$, an inequality that, using (15), is equivalent to $U_\lambda(h(x)) \leq \mathbb{E}_x^\pi[U_\lambda(\sum_{t=0}^{n-1}(C(X_t, A_t) - g) + h(X_{n+1}))]$; recalling that $U_\lambda(\cdot)$ is increasing, via (4) and (7) it follows that

$$
\begin{aligned}
U_\lambda(h(x)) &\leq \mathbb{E}_x^\pi\left[U_\lambda\left(\sum_{t=0}^{n-1}(C(X_t, A_t) - g) + \|h\|\right)\right] \\
&= e^{\lambda(\|h\| - ng)} \mathbb{E}_x^\pi\left[U_\lambda\left(\sum_{t=0}^{n-1} C(X_t, A_t)\right)\right] \\
&= e^{\lambda(\|h\| - ng)} U_\lambda(J_n(\pi, x)) = U_\lambda(J_n(\pi, x) - ng + \|h\|).
\end{aligned}
$$

Thus, $h(x) \leq J_n(\pi, x) - ng + \|h\|$, and then $g \leq (1/n)(J_n(\pi, x) + \|h\| - h(x))$ for every positive integer n. Therefore,

$$
g \leq J_-(\pi, \cdot), \quad \pi \in \mathcal{P}, \qquad \text{so that } g \leq J_*(\cdot), \qquad (18)
$$

by (10) and (11). On the other hand, part (ii) above yields $\mathbb{E}_x^{f^*}[Y_0] = \mathbb{E}_x^{f^*}[Y_n]$, which is equivalent to $U_\lambda(h(x)) = \mathbb{E}_x^{f^*}[U_\lambda(\sum_{t=0}^{n-1}(C(X_t, A_t) - g) + h(X_{n+1}))]$; paralleling the

above argument, it follows that

$$U_\lambda(h(x)) \geq \mathbb{E}_x^{f^*}\left[U_\lambda\left(\sum_{t=0}^{n-1}(C(X_t, A_t) - g) - \|h\|\right)\right]$$

$$= e^{-\lambda(ng+\|h\|)}\mathbb{E}_x^{f^*}\left[U_\lambda\left(\sum_{t=0}^{n-1}C(X_t, A_t)\right)\right]$$

$$= e^{-\lambda(ng+\|h\|)}U_\lambda(J_n(f^*, x)) = U_\lambda(-(ng + \|h\|) + J_n(f^*, x)).$$

Therefore, $h(x) \geq J_n(f^*, x) - (ng + \|h\|)$, and then $g \geq (1/n)(J_n(f^*, x) - (\|h\| + h(x)))$; via (8) and (9), it follows that $g \geq J(f^*, x) \geq J^*(x)$, $x \in S$. Combining this relation with (12) and (18), it follows that $g \leq J_* \leq J^*(\cdot) \leq g$, so that $J_*(\cdot) = g = J^*(\cdot)$ and $J_-(f^*, \cdot) = g = J(f^*, \cdot)$, i.e. $g = \lim_{n\to\infty} n^{-1}J_n(f^*, \cdot)$.

2.4. Assumptions for a constant average cost

Even under strong communication-ergodicity conditions, like those in Assumption 2 below, in the present context of a denumerable state space the existence of a solution of the optimality equation cannot be generally guaranteed for an arbitrary bounded cost function; an (uncontrolled) example illustrating this phenomenon was given in Section 9 of [7]. However, as we verify later, the strong requirements imposed below ensure that the optimal value functions are constant. To continue, it is convenient to introduce the following notation.

For each set $F \subset S$, define the first return time to F by

$$T_F := \min\{n \geq 1 \mid X_n \in F\}, \tag{19}$$

where, by convention, the minimum of the empty set is ∞; if $F = \{x\}$ is a singleton, the simpler notation

$$T_x \equiv T_{\{x\}} \tag{20}$$

is used. Combining the above definition with (1) and (2), it follows that $[T_F = n] \in \mathcal{F}_n$ for every $n \in \mathbb{N}$, so that T_F is a stopping time with respect to the filtration $\{\mathcal{F}_n\}$.

Assumption 2. *There exists $z \in S$ such that the following properties hold.*

(i) *Accessibility from z:*
    Under the action of any stationary policy, every state $y \in S$ is accessible from z, i.e.
    $\mathbb{P}_z^f[T_y < \infty] > 0$, $y \in S, f \in \mathbb{F}$.

(ii) *Simultaneous Doeblin condition:*
     The first return time T_z satisfies $\sup_{x \in S, f \in \mathbb{F}} \mathbb{E}_x^f[T_z] < \infty$.

The proof of the following result can be found in [7].

Lemma 1. *Suppose that Assumptions 1 and 2 hold. Then:*

(i) *For every $y \in S$, there exists a finite constant M_y such that*

$$\mathbb{E}_x^\pi[T_y] \leq M_y, \qquad x \in S, \ \pi \in \mathcal{P}. \tag{21}$$

(ii) *For arbitrary different states $x, y \in S$, $\mathbb{P}_x^\pi[T_y < T_x] > 0$, $\pi \in \mathcal{P}$.*

2.5. The problems

As already observed, under Assumptions 1 and 2, the optimal average cost cannot be generally characterized via the optimality equation (14), since it does not necessarily admit a solution. Thus, looking for a characterization of the optimal average cost is an interesting problem. The main objectives of this paper are:

- to show that the optimal value functions $J^*(\cdot)$ and $J_*(\cdot)$ coincide and are constant;

- to characterize the optimal average cost.

These problems were addressed in [7] for the case $\lambda > 0$ of a risk-averse controller, so, from this point onwards, *our analysis focuses on the risk-seeking case $\lambda < 0$.*

3. Main result

In this section the main conclusions of this work are stated in Theorem 1, which extends to the risk-seeking context results established in [7] for the risk-averse case. The ideas in the following definition will be useful.

Definition 1.

(i) The set $\tilde{\mathcal{G}}$ of subsolutions of the optimality equation (14) for model $\mathcal{M}$ consists of all pairs $(g, h(\cdot))$, where $g \in \mathbb{R}$ and the function $h \colon S \to \mathbb{R}$ satisfy

$$U_\lambda(g + h(x)) \leq \inf_{a \in A(x)} \left[\sum_{y \in S} p_{x,y}(a) U_\lambda(C(x, a) + h(y)) \right], \qquad x \in S. \qquad (22)$$

(ii) The set $\mathcal{G}$ is the projection of $\tilde{\mathcal{G}}$ on its first coordinate, i.e. $\mathcal{G} = \{g \in \mathbb{R} \mid (g, h) \in \tilde{\mathcal{G}} \text{ for some } h \colon \mathbb{R} \to \mathbb{R}\}$.

Remark 3. When two cost functions are simultaneously being examined, the sets $\tilde{\mathcal{G}}$ and $\mathcal{G}$ in the above definition are denoted by $\tilde{\mathcal{G}}(C)$ and $\mathcal{G}(C)$, respectively, making it clear what cost function is being considered. With this notation, it follows that, for each $\beta \in \mathbb{R}$,

$$(g, h(\cdot)) \in \tilde{\mathcal{G}}(C) \iff (g + \beta, h(\cdot)) \in \tilde{\mathcal{G}}(C + \beta),$$

whereas

$$g \in \mathcal{G}(C) \iff g + \beta \in \mathcal{G}(C + \beta). \qquad (23)$$

Via (3), it is not difficult to see that if $\lambda < 0$, then inequality (22) is equivalent to

$$e^{\lambda g + \lambda h(x)} \geq \sup_{a \in A(x)} \left[e^{\lambda C(x,a)} \sum_{y \in S} p_{x,y}(a) e^{\lambda h(y)} \right], \qquad x \in S. \qquad (24)$$

Using the above notation, the main result of the paper can be stated as follows.

Theorem 1. *Suppose that $\lambda < 0$, and that Assumptions 1 and 2 hold. In this context, the following assertions are valid:*

(i) *The inclusion $g \in \mathcal{G}$ is equivalent to $g \leq J_*(\cdot)$.*

(ii) *For each $x \in S$, $J_*(x) = \sup\{g \mid g \in \mathcal{G}\} = J^*(x)$.*

(iii) *Set* $g^* = \sup\{g \mid g \in \mathcal{G}\}$. *With this notation,* $g^* \in \mathcal{G}$. *More explicitly, there exists* $h^* : S \to \mathbb{R}$ *such that* $(g^*, h^*(\cdot)) \in \tilde{\mathcal{G}}$, *that is,*

$$e^{\lambda g^* + \lambda h^*(x)} \geq \sup_{a \in A(x)} \left[e^{\lambda C(x,a)} \sum_{y \in S} p_{x,y}(a) e^{\lambda h^*(y)} \right], \qquad x \in S. \tag{25}$$

(iv) *For each* $\varepsilon > 0$, *there exists a stationary policy* $f \in \mathbb{F}$ *which is* ε-*optimal in the sense that* $g^* \leq J_-(f, x) \leq J(f, x) \leq g^* + \varepsilon$, $x \in S$.

The equality in Theorem 1(ii) is an extension of the Collatz–Wielandt formula for the largest eigenvalue of a positive (finite-dimensional) matrix ([17]); see, for instance, the discussion on this point presented in Remark 3.2 of [7]. The rather technical proof of the above result will be presented in Section 6. Roughly, the backbone of the argument consists in 'approximating' the optimal average cost of the original model $\mathcal{M}$, via MDPs with finite spaces of stationary policies and cost functions with finite support. This strategy is developed in the following two sections, and can be briefly outlined as follows: Section 4 concerns basic properties of the class of subsolutions. First, it is shown that if $(g, h(\cdot)) \in \tilde{\mathcal{G}}$ then $h(\cdot)$ is bounded from above, and g is a lower bound for the optimal inferior limit average index; moreover, given that the system is driven by $\pi \in \mathcal{P}$ starting at $x \in S$, it is shown that $h(x)$ is bounded from above by the certainty equivalent of the total relative cost $\sum_{t=0}^{T_w-1} (C(X_t, A_t) - g)$ up to the first return time to a given state w. Additionally, such a certainty equivalent is studied for the case in which the system evolves under a stationary policy, and the results in this direction are used in Section 5 to show that, if the space of stationary policies is finite and C has finite support, then the optimality equation (14) admits a solution $(g, h(\cdot))$ where the mapping $h(\cdot)$ is bounded. After these preliminaries, the proof of Theorem 1 is finally presented in Section 6 before the concluding remarks.

4. Fundamental tools

This section presents auxiliary results that will be used to establish Theorem 1. The starting point is the following lemma, which concerns two basic properties of the pairs in $\tilde{\mathcal{G}}$, namely, if $(g, h(\cdot)) \in \tilde{\mathcal{G}}$ then (i) the functional part h is bounded from above, and (ii) g is a lower bound of the optimal inferior limit average index.

Lemma 2. *Given that* $\lambda < 0$, *under Assumptions 1 and 2 the following properties hold for each* $w \in S$ *and* $(g, h(\cdot)) \in \tilde{\mathcal{G}}$:

(i) *For each* $x \in S$ *and* $\pi \in \mathcal{P}$,

$$e^{\lambda h(x) - \lambda h(w)} \geq \mathbb{E}_x^\pi \left[\exp \left\{ \lambda \sum_{t=0}^{T_w-1} (C(X_t, A_t) - g) \right\} \right], \tag{26}$$

and then

$$1 \geq \mathbb{E}_w^\pi \left[\exp \left\{ \lambda \sum_{t=0}^{T_w-1} (C(X_t, A_t) - g) \right\} \right]. \tag{27}$$

(ii) $h(\cdot) - h(w) \leq \|C - g\| M_w$, *where* M_w *is as in* (21).

(iii) $g \leq J_*(x)$ *for each* $x \in S$.

Proof. For (i), let $\pi \in \mathcal{P}$ and $x \in S$ be arbitrary. Given $n \in \mathbb{N}$, observe that, combining Definition 1 with (24), it follows that $e^{\lambda h(X_n)} \geq \sum_{y \in S} e^{\lambda(C(X_n, a) - g)} p_{X_n, y}(a) e^{\lambda h(y)}$ for every $a \in A(X_n)$, and then

$$e^{\lambda h(X_n)} \geq \sum_{y \in S} \int_{A(X_n)} e^{\lambda(C(X_n, a) - g)} p_{x,y}(a) \, \pi_n(\mathrm{d}a \mid H_n) e^{\lambda h(y)}$$
$$= \mathbb{E}_x^\pi \left[e^{\lambda(C(X_n, A_n) - g) + \lambda h(X_{n+1})} \mid \mathcal{F}_n \right],$$

where the equality is due to the Markov property. Now set $Z_0 = e^{\lambda h(X_0)}$, $Z_n = \exp\left\{ \lambda \sum_{t=0}^{n-1} (C(X_t, A_t) - g) + \lambda h(X_n) \right\}$, $n = 1, 2, 3, \ldots$ Using these and the previous display, and the fact that $\exp\left\{ \lambda \sum_{t=0}^{n-1} (C(X_t, A_t) - g) \right\}$ is $\mathcal{F}_n$-measurable, an argument along the lines of the one used in Remark 2 yields that $\{(Z_n, \mathcal{F}_n)\}$ is a supermartingale with respect to $\mathbb{P}_x^\pi$, i.e. $Z_n \geq \mathbb{E}_x^\pi[Z_{n+1} \mid \mathcal{F}_n]$, $n \in \mathbb{N}$, $\mathbb{P}_x^\pi$-almost surely (a.s.). Therefore, since T_w is a stopping time with respect to the filtration $\{\mathcal{F}_n\}$, the optional sampling theorem yields $\mathbb{E}_x^\pi[Z_0] \geq \mathbb{E}_x^\pi[Z_{n \wedge T_w}]$ (Theorem 7.7.3, p. 304 of [1]; Theorem 35.2, p. 405 of [4]), i.e.

$$e^{\lambda h(x)} \geq \mathbb{E}_x^\pi \left[\exp\left\{ \lambda \sum_{t=0}^{n \wedge T_w - 1} (C(X_t, A_t) - g) + \lambda h(X_{n \wedge T_w}) \right\} \right]. \tag{28}$$

Now observe that $\mathbb{P}_x^\pi[T_w < \infty] = 1$, by (21), and that $X_{T_w} = w$ on the event $[T_w < \infty]$, by (19). Therefore, with probability 1 with respect to $\mathbb{P}_x^\pi$,

$$\lim_{n \to \infty} \exp\left\{ \lambda \sum_{t=0}^{n \wedge T_w - 1} (C(X_t, A_t) - g) + \lambda h(X_{n \wedge T_w}) \right\}$$
$$= \exp\left\{ \lambda \sum_{t=0}^{T_w - 1} (C(X_t, A_t) - g) + \lambda h(X_{T_w}) \right\} = \exp\left\{ \lambda \sum_{t=0}^{T_w - 1} (C(X_t, A_t) - g) + \lambda h(w) \right\}.$$

Thus, via Fatou's lemma, (28) yields

$$e^{\lambda h(x)} \geq \liminf_{n \to \infty} \mathbb{E}_x^\pi \left[\exp\left\{ \lambda \sum_{t=0}^{n \wedge T_w - 1} (C(X_t, A_t) - g) + \lambda h(X_{n \wedge T_w}) \right\} \right]$$
$$\geq \mathbb{E}_x^\pi \left[\exp\left\{ \lambda \sum_{t=0}^{T_w - 1} (C(X_t, A_t) - g) + \lambda h(w) \right\} \right]$$
$$= \mathbb{E}_x^\pi \left[\exp\left\{ \lambda \sum_{t=0}^{T_w - 1} (C(X_t, A_t) - g) \right\} \right] e^{\lambda h(w)},$$

and (26) follows; setting x equal to w leads to (27).

For (ii), from Jensen's inequality it follows that

$$\mathbb{E}_x^\pi\left[\exp\left\{\lambda\sum_{t=0}^{T_w-1}(C(X_t,A_t)-g)\right\}\right] \geq \exp\left\{\mathbb{E}_x^\pi\left[\lambda\sum_{t=0}^{T_w-1}(C(X_t,A_t)-g)\right]\right\}$$

$$\geq \mathrm{e}^{-\|\lambda(C-g)\|\mathbb{E}_x^\pi[T_w]}$$

$$\geq \mathrm{e}^{\lambda\|C-g\|M_w}, \qquad x\in S,\ \pi\in\mathcal{P}, \tag{29}$$

where the negativity of λ and (21) were used in the last step. The above relation and (26) together imply that $\lambda h(\cdot) - \lambda h(w) \geq \lambda\|C-g\|M_w$, and the conclusion follows.

For (iii), let $x\in S$ and $\pi\in\mathcal{P}$ be arbitrary. Recalling that $\{(Z_n,\mathcal{F}_n)\}$ is a supermartingale with respect to $\mathbb{P}_x^\pi$, for every positive integer n,

$$\mathrm{e}^{\lambda h(x)} = \mathbb{E}_x^\pi[Z_0] \geq \mathbb{E}_x^\pi[Z_n]$$

$$= \mathbb{E}_x^\pi\left[\exp\left\{\lambda\sum_{t=0}^{n-1}(C(X_t,A_t)-g)+\lambda h(X_n)\right\}\right]$$

$$\geq \mathbb{E}_x^\pi\left[\exp\left\{\lambda\sum_{t=0}^{n-1}(C(X_t,A_t)-g)\right\}\right]\mathrm{e}^{\lambda(h(w)+\|C-g\|M_w)}$$

$$= \mathbb{E}_x^\pi\left[\exp\left\{\lambda\sum_{t=0}^{n-1}C(X_t,A_t)\right\}\right]\mathrm{e}^{-n\lambda g+\lambda(h(w)+\|C-g\|M_w)}$$

$$= \exp\{\lambda J_n(\pi,x)-n\lambda g+\lambda(h(w)+\|C-g\|M_w)\},$$

where the second inequality is due to part (ii), and (7) was used in the last step. It follows that $\lambda h(x) \geq \lambda J_n(\pi,x) - n\lambda g + \lambda(h(w)+\|C-g\|M_w)$, and then

$$g \leq \frac{1}{n}J_n(\pi,x)+\frac{1}{n}(h(w)-h(x)+\|C-g\|M_w).$$

Taking the inferior limit as n goes to ∞ in both sides of this relation, it follows that $g \leq J_-(\pi,x)$, and then $g \leq J_*(\cdot)$, since $x\in S$ and $\pi\in\mathcal{P}$ are arbitrary in this argument. $\quad\square$

The remainder of the section is dedicated to studying inequality (27) for a stationary policy f, and it is convenient to introduce additional notation.

Definition 2. Let $f\in\mathbb{F}$, $g\in\mathbb{R}$, and $w\in S$ be arbitrary but fixed, and define the total relative cost function up to the first return time to state w under f by

$$h_{f,g,w}(x) := \frac{1}{\lambda}\log\left(\mathbb{E}_x^f\left[\exp\left\{\lambda\sum_{t=0}^{T_w-1}(C(X_t,A_t)-g)\right\}\right]\right), \qquad x\in S.$$

Notice that this definition and (29) with f instead of π together yield $\lambda h_{f,g,w}(x) \geq \lambda\|C-g\|M_w$, and the negativity of λ leads to

$$h_{f,g,w}(\cdot) \leq \|C-g\|M_w. \tag{30}$$

On the other hand, a conditional argument combining Definition 2 with the Markov property yields

$$
\mathbb{E}_x^f \left[\exp\left\{ \lambda \sum_{t=0}^{T_w-1} (C(X_t, A_t) - g) \right\} \mathbf{1}[X_1 \neq w] \mid H_1 \right]
$$

$$
= \mathrm{e}^{\lambda(C(X_0, f(X_0)) - g)} \mathbf{1}[X_1 \neq w] \mathbb{E}_x^f \left[\exp\left\{ \lambda \sum_{t=1}^{T_w-1} (C(X_t, A_t) - g) \right\} \mathbf{1}[X_1 \neq w] \mid H_1 \right]
$$

$$
= \mathrm{e}^{\lambda(C(X_0, f(X_0)) - g)} \mathbf{1}[X_1 \neq w] \mathbb{E}_{X_1} \left[\exp\left\{ \lambda \sum_{t=0}^{T_w-1} (C(X_t, A_t) - g) \right\} \right]
$$

$$
= \mathrm{e}^{\lambda(C(X_0, f(X_0)) - g)} \mathbf{1}[X_1 \neq w] \mathrm{e}^{\lambda h_{f,g,w}(X_1)},
$$

and then

$$
\mathrm{e}^{\lambda h_{f,g,w}(x)} = \mathbb{E}_x^f \left[\exp\left\{ \lambda \sum_{t=0}^{T_w-1} (C(X_t, A_t) - g) \right\} \mathbf{1}[X_1 = w] \right]
$$

$$
+ \mathbb{E}_x^f \left[\exp\left\{ \lambda \sum_{t=0}^{T_w-1} (C(X_t, A_t) - g) \right\} \mathbf{1}[X_1 \neq w] \right]
$$

$$
= \mathrm{e}^{\lambda(C(x, f(x)) - g)} \left(p_{x,w}(f(x)) + \sum_{y \neq w} p_{x,y}(f(x)) \mathrm{e}^{\lambda h_{f,g,w}(y)} \right). \tag{31}
$$

Next, the following lemma shows that if $h_{f,g,w}(w)$ is finite at some point, then $h_{f,g,w}(\cdot)$ is finite on S, and a fundmental property of the mapping $y \mapsto h_{f,g,y}(y)$ will be established, namely, if that function attains a non-negative value at some point, then it is *always* non-negative.

Lemma 3. *Assume that $\lambda < 0$ and that Assumptions 1 and 2 are valid. In this context, the following assertions hold for arbitrary $f \in \mathbb{F}$, $g \in \mathbb{R}$, and $w \in S$.*

  (i) *If $h_{f,g,w}(w) > -\infty$, then $h_{f,g,w}(x)$ is finite for every $x \in S$.*

 (ii) *The following assertions are equivalent:*

   (a) $h_{f,g,w}(w) \geq 0$.

   (b) $h_{f,g,w}(\cdot)$ *is finite and the following Poisson inequality holds:*

$$
\mathrm{e}^{\lambda h_{f,g,w}(x)} \geq \mathrm{e}^{\lambda(C(x, f(x)) - g)} \sum_{y \in S} p_{x,y}(f(x)) \mathrm{e}^{\lambda h_{f,g,w}(y)}, \qquad x \in S. \tag{32}
$$

   (c) $h_{f,g,y}(y) \geq 0$ *for every $y \in S$.*

Proof. For (i), given $x \in S \setminus \{w\}$, using Lemma 1(ii) pick a positive integer k satisfying

$$
\mathbb{P}_w[T_x = k < T_w] > 0, \tag{33}
$$

and notice that $[T_x = k < T_w] = [X_t \notin \{x, w\}, 1 \leq t < k, X_k = x]$ belongs to $\mathcal{F}_k$ (see (2)), so that

$$
\mathbb{E}_w^f \left[\exp \left\{ \lambda \sum_{t=0}^{T_w-1} (C(X_t, A_t) - g) \right\} \mid \mathcal{F}_k \right]
$$

$$
\geq \mathbb{E}_w^f \left[\mathbf{1}[T_x = k < T_w] \exp \left\{ \lambda \sum_{t=0}^{T_w-1} (C(X_t, A_t) - g) \right\} \mid \mathcal{F}_k \right]
$$

$$
= \exp \left\{ \lambda \sum_{t=0}^{k-1} (C(X_t, A_t) - g) \right\} \mathbf{1}[T_x = k < T_w] \mathbb{E}_w^f \left[\exp \left\{ \lambda \sum_{t=k}^{T_w-1} (C(X_t, A_t) - g) \right\} \mid \mathcal{F}_k \right]
$$

$$
\geq e^{\lambda k \|C - g\|} \mathbf{1}[T_x = k < T_w] \mathbb{E}_w^f \left[\exp \left\{ \lambda \sum_{t=k}^{T_w-1} (C(X_t, A_t) - g) \right\} \mid \mathcal{F}_k \right]
$$

$$
= e^{\lambda k \|C - g\|} \mathbf{1}[T_x = k < T_w] \mathbb{E}_{X_k}^f \left[\exp \left\{ \lambda \sum_{t=0}^{T_w-1} (C(X_t, A_t) - g) \right\} \right],
$$

where the last equality is due to the Markov property. Since $X_k = x$ on the event $[T_x = k]$, it follows from Definition 2 that

$$
\mathbb{E}_w^f \left[\exp \left\{ \lambda \sum_{t=0}^{T_w-1} (C(X_t, A_t) - g) \right\} \mid \mathcal{F}_k \right]
$$

$$
\geq e^{\lambda k \|C - g\|} \mathbf{1}[T_x = k < T_w] \mathbb{E}_x^f \left[\exp \left\{ \lambda \sum_{t=0}^{T_w-1} (C(X_t, A_t) - g) \right\} \right]
$$

$$
= e^{\lambda k \|C - g\|} \mathbf{1}[T_x = k < T_w] e^{\lambda h_{f,g,w}(x)},
$$

so that

$$
e^{\lambda h_{f,g,w}(w)} = \mathbb{E}_w^f \left[\exp \left\{ \lambda \sum_{t=0}^{T_w-1} (C(X_t, A_t) - g) \right\} \right] \geq e^{\lambda k \|C - g\|} e^{\lambda h_{f,g,w}(x)} \mathbb{P}_w^f[T_x = k < T_w],
$$

and then (33) yields

$$
h_{f,g,w}(w) > -\infty \implies \lambda h_{f,g,w}(w) < \infty \implies \lambda h_{f,g,w}(x) < \infty \implies h_{f,g,w}(x) > -\infty.
$$

Since $x \in S \setminus \{w\}$ is arbitrary in this argument, this last display and (30) together yield that if $h_{f,g,w}(w)$ is finite then $h_{f,g,w}(x)$ is finite for every $x \in S$.

For (ii) (c) $\implies$ (a), the implication is clear. For (a) $\implies$ (b), suppose that $h_{f,g,w}(w) \geq 0$, so that $h_{f,g,w}(\cdot)$ is a finite function, by part (i), and $1 \geq e^{\lambda h_{f,g,w}(w)}$, since λ is negative. Thus, from (31),

$$
e^{\lambda h_{f,g,w}(x)} \geq e^{\lambda(C(x,f(x))-g)} \left(p_{x,w}(f(x)) e^{\lambda h_{f,g,w}(w)} + \sum_{y \neq w} p_{x,y}(f(x)) e^{\lambda h_{f,g,w}(y)} \right), \quad x \in S,
$$

and (32) follows. For (b) $\implies$ (c), let $y \in S$ be arbitrary but fixed. Consider the new MDP $\mathcal{M}_f$ obtained from $\mathcal{M}$ by setting $A(x) = \{f(x)\}$. In this case f is the unique policy in this new

model, and (32) establishes that the pair $(g, h_{f,g,w}(\cdot))$ belongs to the set $\tilde{\mathcal{G}}(\mathcal{M}_f)$ of subsolutions corresponding to $\mathcal{M}_f$, so that the conclusions of Lemma 2(i) applied to model $\mathcal{M}_f$ with y instead of w hold. In particular, inequality (27) with f instead of π is valid, and then $1 \geq \mathbb{E}_y^f\big[\exp\big\{\lambda \sum_{t=0}^{T_y-1}(C(X_t, A_t) - g)\big\}\big] = e^{\lambda h_{f,g,y}(y)}$, so that $h_{f,g,y}(y) \geq 0$. $\qquad\square$

The next result is a natural complement of Lemma 3(ii).

Lemma 4. *Let $f \in \mathbb{F}$, $g \in \mathbb{R}$, and $w \in S$ be arbitrary but fixed. Under the conditions in Lemma 3, the following assertions hold.*

(i) $h_{f,g,w}(w) > 0 \iff h_{f,g,y}(y) > 0$ *for every $y \in S$.*

(ii) $h_{f,g,w}(w) = 0 \iff h_{f,g,y}(y) = 0$ *for every $y \in S$.*

Proof. For (i), the $\Longleftarrow$ part is clear. To prove the $\Longrightarrow$, suppose that $h_{f,g,w}(w) > 0$. In this case $h_{f,g,w}(\cdot)$ is finite by Lemma 3, and, recalling that λ is negative, $1 > e^{\lambda h_{f,g,w}(w)} = \mathbb{E}_w^f\big[\exp\big\{\lambda \sum_{t=0}^{T_w-1}(C(X_t, A_t) - g)\big\}\big]$. Thus,

$$1 = e^{\lambda \rho} \mathbb{E}_w^f\left[\exp\left\{\lambda \sum_{t=0}^{T_w-1}(C(X_t, A_t) - g)\right\}\right] \quad \text{for some } \rho < 0. \qquad (34)$$

Next, define the cost function $\hat{C}$ by $\hat{C}(x, \cdot) = C(x, \cdot)$, $x \neq w$, $\hat{C}(w, \cdot) = C(w, \cdot) + \rho$, and consider the new model $\hat{\mathcal{M}}$ obtained from $\mathcal{M}$ by replacing C by $\hat{C}$. In this case, if $\hat{h}_{f,g,w}$ is the total relative cost in Definition 2 associated with model $\hat{\mathcal{M}}$, (34) yields that $e^{\lambda \hat{h}_{f,g,w}(w)} = 1$, i.e. $\hat{h}_{f,g,w}(w) = 0$, and then the equivalence of assertions (a) and (c) in Lemma 3(ii) applied to model $\hat{\mathcal{M}}$ implies that

$$\hat{h}_{f,g,y}(y) \geq 0 \quad \text{for every state } y. \qquad (35)$$

Next, combining this fact with the inequality $h_{f,g,w}(w) > 0$, we show that $h_{f,g,y}(y) > 0$ for $y \in S$. Since $h_{f,g,w}(w) > 0$, to achieve this goal it is sufficient to establish the above inequality for $y \neq w$. In this context, observe that (35) yields

$$\begin{aligned}
1 \geq e^{\lambda \hat{h}_{f,g,y}(y)} &= \mathbb{E}_y^f\left[\exp\left\{\lambda \sum_{t=0}^{T_y-1}(\hat{C}(X_t, A_t) - g)\right\}\right] \\
&= \mathbb{E}_y^f\left[\exp\left\{\lambda \sum_{t=0}^{T_y-1}(C(X_t, A_t) - g)\right\} \exp\left\{\lambda \rho \sum_{t=0}^{T_y-1} \mathbf{1}_{\{w\}}(X_t)\right\}\right] \\
&> \mathbb{E}_y^f\left[\exp\left\{\lambda \sum_{t=0}^{T_y-1}(C(X_t, A_t) - g)\right\}\right] = e^{\lambda h_{f,g,y}(y)},
\end{aligned}$$

where the strict inequality follows by observing that (a) $\exp\big\{\lambda \rho \sum_{t=0}^{T_y-1} \mathbf{1}_{\{w\}}(X_t)\big\} \geq 1$, since λ and ρ are negative, and that (b) $\exp\big\{\lambda \rho \sum_{t=0}^{T_y-1} \mathbf{1}_{\{w\}}(X_t)\big\} > 1$ on the event $[T_w < T_y]$, which has positive probability with respect to $\mathbb{P}_y^f$ by Lemma 1(ii), whereas the last equality is due to Definition 2. Thus, $1 > e^{\lambda h_{f,g,y}(y)}$, and then $h_{f,g,y}(y) > 0$, since λ is negative. This completes the proof of part (i).

For (ii), the assertion follows by combining part (i) with the equivalence of properties (a) and (c) in Lemma 3(ii). $\qquad\square$

5. Optimality equation in a special case

The objective of this section is to provide sufficient conditions to ensure the existence of a *bounded* solution to the optimality equation (14). The result in this direction requires special conditions on the cost function and the action sets, and is stated in Theorem 2. That result will be established using the following lemma.

Lemma 5. *Let $\lambda < 0$ and $f \in \mathbb{F}$ be arbitrary but fixed. Suppose that the cost function $C(\cdot, f(\cdot))$ under f satisfies the following requirements:*

(a) $C(x, f(x)) \le 0$ *for every $x \in S$;*

(b) $C(\cdot, f(\cdot))$ *has finite support, i.e.*

$$F := \{x \in S \mid C(x, f(x)) < 0\} \text{ is finite.} \tag{36}$$

In this case there exists a real number $g(f)$ such that the following properties are valid:

(i) $g(f) \le 0$, *and the functions $\{h_{f,g(f),w}\}_{w \in S}$ satisfy*

$$\mathrm{e}^{\lambda h_{f,g(f),w}(w)} = \mathbb{E}_w^f\left[\exp\left\{ \lambda \sum_{t=0}^{T_w-1} (C(X_t, A_t) - g(f)) \right\} \right] = 1, \qquad w \in S; \tag{37}$$

see Definition 2.

(ii) *For each $w \in S$, the function $h_{f,g(f),w}(\cdot)$ is finite and the following Poisson equation holds:*

$$\mathrm{e}^{\lambda h_{f,g(f),w}(x)} = \mathrm{e}^{\lambda(C(x,f(x))-g(f))} \sum_y p_{x,y}(f(x)) \mathrm{e}^{\lambda h_{f,g(f),w}(y)}, \qquad x \in S. \tag{38}$$

(iii) *The function $h_{f,g(f),w}(\cdot)$ is bounded for each $w \in S$. More explicitly,*

$$-\infty < \inf_{x \in S} h_{f,g(f),w}(x) \le h_{f,g(f),w}(\cdot) \le M_w \|C - g(f)\|.$$

(iv) *Given $w \in S$, for $n \in \mathbb{N}$ set*

$$W_n(f) := \exp\left\{ \lambda \sum_{t=0}^{n \wedge T_w-1} (C(X_t, A_t) - g(f)) + \lambda h_{f,g,w}(X_{n \wedge T_w}) \right\}. \tag{39}$$

With this notation, $W_n(f) = \mathbb{E}_x^f\left[\exp\left\{ \lambda \sum_{t=0}^{T_w-1} (C(X_t, A_t) - g(f)) \right\} \mid \mathcal{F}_n \right]$, $\mathbb{P}_x^f$-a.s., $x \in S$, $n \in \mathbb{N}$; see (2), (19), and (20).

Proof. Recall that the equality $A_t = f(X_t)$ is always valid when the system is driven using *f*.

The argument to prove (i) is by induction on $|F|$, the number of elements of the set *F* in (36). If $|F| = 0$ then $C(\cdot, f(\cdot)) = 0$, so that, setting $g(f) = 0$, it follows that (37) holds for each $x \in S$. Next, assume that, for some nonnegative integer *n*, the desired conclusion is valid when the support of $C(\cdot, f(\cdot))$ has *n* elements, and suppose that $|F| = n + 1$. Pick $\tilde{x} \in F$, so that

$$C(\tilde{x}, f(\tilde{x})) < 0, \tag{40}$$

and set $\tilde{F} = F \setminus \{\tilde{x}\}$, $\tilde{C}(x, a) = C(x, a)\mathbf{1}_{\tilde{F}}(x)$, $(x, a) \in \mathbb{K}$, so that

$$F = \tilde{F} \cup \{\tilde{x}\}, \quad C(x, f(x)) = C(\tilde{x}, f(\tilde{x}))\mathbf{1}_{\{\tilde{x}\}}(x) + \tilde{C}(x, f(x)), \qquad x \in S. \tag{41}$$

It follows that the support of the function $\tilde{C}(\cdot, f(\cdot))$ is the set $\tilde{F}$, which has n elements. Thus, by the induction hypothesis there exists $\tilde{g}(f) \leq 0$ such that (37) holds with $\tilde{C}$ and $\tilde{g}(f)$ instead of C and $g(f)$, respectively; in particular,

$$\mathbb{E}^f_{\tilde{x}}\left[\exp\left\{ \lambda \sum_{t=0}^{T_{\tilde{x}}-1} (\tilde{C}(X_t, A_t) - \tilde{g}(f)) \right\} \right] = 1.$$

Next, observe that (41) yields

$$\sum_{t=0}^{T_{\tilde{x}}-1} (C(X_t, f(X_t)) - \tilde{g}(f)) = C(\tilde{x}, f(\tilde{x})) \sum_{t=0}^{T_{\tilde{x}}-1} \mathbf{1}_{\{\tilde{x}\}}(X_t) + \sum_{t=0}^{T_{\tilde{x}}-1} (\tilde{C}(X_t, f(X_t)) - \tilde{g}(f)))$$

$$= C(\tilde{x}, f(\tilde{x})) \mathbf{1}_{\{\tilde{x}\}}(X_0) + \sum_{t=0}^{T_{\tilde{x}}-1} (\tilde{C}(X_t, f(X_t)) - \tilde{g}(f)))$$

$$= C(\tilde{x}, f(\tilde{x})) + \sum_{t=0}^{T_{\tilde{x}}-1} (\tilde{C}(X_t, f(X_t)) - \tilde{g}(f))) \qquad \mathbb{P}^f_{\tilde{x}}\text{-a.s.},$$

where the second equality was obtained using that $X_t \neq \tilde{x}$ for $1 \leq t < T_{\tilde{x}}$, by (19) and (20), whereas the relation $\mathbb{P}^f_{\tilde{x}}[X_0 = \tilde{x}] = 1$ was used in the last step. Thus, (37) leads to

$$\mathbb{E}^f_{\tilde{x}}\left[\exp\left\{ \lambda \sum_{t=0}^{T_{\tilde{x}}-1} (C(X_t, A_t) - \tilde{g}(f)) \right\} \right] = e^{\lambda C(\tilde{x}, f(\tilde{x}))} > 1, \tag{42}$$

where, recalling that $\lambda < 0$, the inequailty is due to (40). Now observe that

$$\exp\left\{ \lambda \sum_{t=0}^{T_{\tilde{x}}-1} (C(X_t, A_t) - \gamma) \right\} \searrow 0 \text{ as } \gamma \searrow -\infty,$$

whereas, since the mapping $\gamma \mapsto \exp\left\{ \lambda \sum_{t=0}^{T_{\tilde{x}}-1} (C(X_t, A_t) - \gamma) \right\}$ is increasing,

$$\exp\left\{ \lambda \sum_{t=0}^{T_{\tilde{x}}-1} (C(X_t, A_t) - \gamma) \right\} \leq \exp\left\{ \lambda \sum_{t=0}^{T_{\tilde{x}}-1} (C(X_t, A_t) - \tilde{g}(f)) \right\}, \qquad \gamma \leq \tilde{g}(f).$$

Combining (42) with the two previous displays, the dominated convergence theorem implies that $v(\gamma) := \mathbb{E}^f_{\tilde{x}}\left[\exp\left\{ \lambda \sum_{t=0}^{T_{\tilde{x}}-1} (C(X_t, A_t) - \gamma) \right\} \right]$ is continuous in $\gamma \in (-\infty, \tilde{g}(f)]$, and that $v(\gamma) \searrow 0$ as $\gamma \searrow -\infty$; since $v(\tilde{g}(f)) = e^{\lambda C(\tilde{x}, f(\tilde{x}))} > 1$, by (42), the intermediate value property implies that there exists $g(f) \in (-\infty, \tilde{g}(f)) \subset (-\infty, 0)$ such that

$$1 = v(g(f)) = \mathbb{E}^f_{\tilde{x}}\left[\exp\left\{ \lambda \sum_{t=0}^{T_{\tilde{x}}-1} (C(X_t, A_t) - g(f)) \right\} \right] = e^{\lambda h_{f, g(f), \tilde{x}}(\tilde{x})}.$$

From this point, Lemma 4(ii) yields that (37) holds for every $x \in S$. This completes the induction argument.

For (ii), since $e^{\lambda h_{f,g(f),w}(w)} = 1$ for every $w \in S$, by part (i), Lemma 3(i) yields that $h_{f,g(f),w}(\cdot)$ is finite, and (38) follows from (31).

For (iii), recall that the equality $\mathbb{P}_x^\pi[T_w < \infty] = 1$ is always valid, by Lemma 1(i). Now, define the finite set $G = \{w\} \cup F$ such that $T_G \leq T_w$ and $T_G \leq T_F$, and then

$$C(X_t, f(X_t)) = 0, \qquad 1 \leq t < T_G. \tag{43}$$

Next, select an initial state

$$X_0 = x \in (S \setminus G) \subset (S \setminus F), \tag{44}$$

so that

$$C(X_0, f(X_0)) = 0, \tag{45}$$

and observe that

$$\exp\left\{\lambda \sum_{t=0}^{T_w - 1} (C(X_t, f(X_t)) - g(f))\right\}$$

$$= \mathbf{1}[T_G = T_w] \exp\left\{\lambda \sum_{t=0}^{T_w - 1} (C(X_t, f(X_t)) - g(f))\right\}$$

$$\quad + \mathbf{1}[T_G < T_w] \exp\left\{\lambda \sum_{t=0}^{T_w - 1} (C(X_t, f(X_t)) - g(f))\right\}$$

$$= \mathbf{1}[T_G = T_w] \exp\left\{\lambda \sum_{t=0}^{T_G - 1} (C(X_t, f(X_t)) - g(f))\right\}$$

$$\quad + \mathbf{1}[T_G < T_w] \exp\left\{\lambda\left(\sum_{t=0}^{T_G - 1} (C(X_t, f(X_t)) - g(f)) + \sum_{t=T_G}^{T_w - 1} (C(X_t, f(X_t)) - g(f))\right)\right\}$$

$$= \mathbf{1}[T_G = T_w]e^{-\lambda g(f)T_G} + \mathbf{1}[T_G < T_w] \exp\left\{-\lambda g(f)T_G + \lambda \sum_{t=T_G}^{T_w - 1} (C(X_t, f(X_t)) - g(f))\right\}$$

$$\leq \mathbf{1}[T_G = T_w] + \mathbf{1}[T_G < T_w] \exp\left\{\lambda \sum_{t=T_G}^{T_w - 1} (C(X_t, f(X_t)) - g(f))\right\},$$

where the third equality is due to (43)–(45), and the inequality stems from the negativity of λ and $g(f)$. It follows that, for every $x \in S \setminus G$,

$$e^{\lambda h_{f,g(f),w}(x)} = \mathbb{E}_x^f\left[\exp\left\{\lambda \sum_{t=0}^{T_w - 1} (C(X_t, f(X_t)) - g(f))\right\}\right]$$

$$\leq \mathbb{P}_x^f[T_G = T_w] + \mathbb{E}_x^f\left[\mathbf{1}[T_G < T_w] \exp\left\{\lambda \sum_{t=T_G}^{T_w - 1} (C(X_t, f(X_t)) - g(f))\right\}\right]$$

$$= \mathbb{P}_x^f[T_G < T_w] + \mathbb{E}_x^f\left[\mathbf{1}[T_G < T_w]e^{\lambda h_{f,g(f),w}(X_{T_G})}\right],$$

where, via a conditional argument, the second equality follows from the Markov property. Thus, since $X_{T_G} \in G$ on the event $[T_G < \infty]$ and $\mathbb{P}_x^f[T_G < \infty] \geq \mathbb{P}_x^f[T_w < \infty] = 1$, it follows that

$$
e^{\lambda h_{f,g(f),w}(x)} \leq 1 + \exp\left\{ \max_{y \in G} |\lambda h_{f,g(f),w}(y)| \right\} \leq 2 \exp\left\{ \max_{y \in G} |\lambda h_{f,g(f),w}(y)| \right\}, \quad x \in S \setminus G,
$$

and then $h_{f,g(f),w}(x) \geq \lambda^{-1}[\log(2) + \max_{y \in G} |\lambda h_{f,g(f),w}(y)|]$ when $x \in S \setminus G$; since $h_{f,g(f),w}(\cdot)$ is a finite function and the set G is finite, it follows that $h_{f,g(f),w}(\cdot)$ is *bounded from below*.

For (iv), looking at (2), (19), and (20) we observe that, for each $x, w \in S$ and $n, k \in \mathbb{N} \setminus \{0\}$ with $k \leq n$, the random variables $\mathbf{1}[T_w = k]$, $\sum_{t=0}^{k-1}(C(X_t, A_t) - g(f))$, and $\mathbf{1}[T_w > n]$ are $\mathcal{F}_n$-measurable, and then

$$
\mathbb{E}_x^f\left[\exp\left\{ \lambda \sum_{t=0}^{T_w-1}(C(X_t, A_t) - g(f)) \right\} \mid \mathcal{F}_n \right]
$$

$$
= \mathbb{E}_x^f\left[\sum_{k=1}^{n} \mathbf{1}[T_w = k] \exp\left\{ \lambda \sum_{t=0}^{k-1}(C(X_t, A_t) - g(f)) \right\} \mid \mathcal{F}_n \right]
$$

$$
+ \mathbb{E}_x^f\left[\mathbf{1}[T_w > n] \exp\left\{ \lambda \sum_{t=0}^{n-1}(C(X_t, A_t) - g(f)) \right\} \exp\left\{ \lambda \sum_{t=n}^{T_w-1}(C(X_t, A_t) - g(f)) \right\} \mid \mathcal{F}_n \right]
$$

$$
= \sum_{k=1}^{n} \mathbf{1}[T_w = k] \exp\left\{ \lambda \sum_{t=0}^{k-1}(C(X_t, A_t) - g(f)) \right\}
$$

$$
+ \mathbf{1}[T_w > n] \exp\left\{ \lambda \sum_{t=0}^{n-1}(C(X_t, A_t) - g(f)) \right\} \mathbb{E}_x^f\left[\exp\left\{ \lambda \sum_{t=n}^{T_w-1}(C(X_t, A_t) - g(f)) \right\} \mid \mathcal{F}_n \right]
$$

$$
= \mathbf{1}[T_w \leq n] \exp\left\{ \lambda \sum_{t=0}^{n \wedge T_w-1}(C(X_t, A_t) - g(f)) \right\}
$$

$$
+ \mathbf{1}[T_w > n] \exp\left\{ \lambda \sum_{t=0}^{n-1}(C(X_t, A_t) - g(f)) \right\} e^{\lambda h_{f,g(f),w}(X_n)}
$$

$$
= \mathbf{1}[T_w \leq n] \exp\left\{ \lambda \sum_{t=0}^{n \wedge T_w-1}(C(X_t, A_t) - g(f)) \right\}
$$

$$
+ \mathbf{1}[T_w > n] \exp\left\{ \lambda \sum_{t=0}^{n \wedge T_w-1}(C(X_t, A_t) - g(f)) \right\} e^{\lambda h_{f,g(f),w}(X_{n \wedge T_w})},
$$

where the Markov property was used to set the third equality. Next, notice that on the event $[T_w \leq n]$ the equality $X_{n \wedge T_w} = w$ holds, and in this case $e^{\lambda h_{f,g(f),w}(X_{n \wedge T_w})} = e^{\lambda h_{f,g(f),w}(w)} = 1$, by part (i). Combining this fact with the previous display, it follows that

$$
\mathbb{E}_x^f\left[\exp\left\{ \lambda \sum_{t=0}^{T_w-1}(C(X_t, A_t) - g(f)) \right\} \mid \mathcal{F}_n \right]
$$

$$
= \mathbf{1}[T_w \leq n] \exp\left\{ \lambda \sum_{t=0}^{n \wedge T_w-1}(C(X_t, A_t) - g(f)) \right\} e^{\lambda h_{f,g(f),w}(X_{n \wedge T_w})}
$$

$$+ \mathbf{1}[T_w > n] \exp \left\{ \lambda \sum_{t=0}^{n \wedge T_w - 1} (C(X_t, A_t) - g(f)) \right\} e^{\lambda h_{f,g(f),w}(X_{n \wedge T_w})}$$

$$= \exp \left\{ \lambda \sum_{t=0}^{n \wedge T_w - 1} (C(X_t, A_t) - g(f)) \right\} e^{\lambda h_{f,g(f),w}(X_{n \wedge T_w})} = W_n(f),$$

completing the proof. □

Next, the previous lemma will be used to establish the main result of the section, which will be derived under the following conditions.

Assumption 3. *There exists a finite set $F \subset S$ such that the cost function and the action sets satisfy the following properties.*

(i) *For each $x \in S \setminus F$, the action set $A(x)$ is a singleton, whereas $A(x)$ is finite for $x \in F$.*

(ii) *If $x \in S \setminus F$, $C(x, \cdot) = 0$ and $C(x, \cdot) \leq 0$ if $x \in F$.*

Theorem 2. *Suppose that $\lambda < 0$, and that Assumptions 2 and 3 are valid. In this context, for each $w \in S$, the following assertions hold.*

(i) *For each stationary policy $\tilde{f}$, the cost function $C(\cdot, \tilde{f}(\cdot))$ satisfies conditions (a) and (b) in Lemma 5, and there exists $f \in \mathbb{F}$ such that the pair $(g(f), h_{f,g(f),w}(\cdot)) \in (-\infty, 0] \times \mathcal{B}(S)$ satisfies*

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

$$= e^{\lambda C(x,f(x))} \sum_{y \in S} p_{x,y}(f(x)) e^{\lambda h_{f,g(f),w}(y)}, \qquad x \in S.$$

(ii) *$J_*(\cdot) = g(f) = J^*(\cdot)$, and $g(f) = \lim_{n \to \infty} (1/n) J_n(f, x)$, $x \in S$.*

Proof. Starting with part (i), notice that Assumption 3 implies that the space $\mathbb{F}$ of stationary policies is *finite* and that, for each $\tilde{f} \in \mathbb{F}$, the nonpositive cost function $C(\cdot, \tilde{f}(\cdot))$ has finite support, so that conditions (a) and (b) in Lemma 5 hold for every $\tilde{f} \in \mathbb{F}$; see (36). It follows that, for each $\tilde{f} \in \mathbb{F}$, there exists $g(\tilde{f})$ such that, for each state w, the pair $(g(\tilde{f}), h_{\tilde{f},g(\tilde{f}),w}(\cdot)) \in (-\infty, 0] \times \mathcal{B}(S)$ satisfies the conclusions in Lemma 5. Pick $f \in \mathbb{F}$ such that

$$g(f) = \min_{\tilde{f} \in \mathbb{F}} g(\tilde{f}). \tag{46}$$

Now let $w \in S$ be arbitrary but fixed, and note that (38) holds by Lemma 5(ii), so that

$$e^{\lambda h_{f,g(f),w}(x)} \leq \sup_{a \in A(x)} \left[e^{\lambda(C(x,a) - g(f))} \sum_y p_{x,y}(a) e^{\lambda h_{f,g(f),w}(y)} \right], \qquad x \in S. \tag{47}$$

Thus, to establish the first assertion it is sufficient to show that the equality holds in the above relation. To achieve this goal, recall that the space $A(y)$ is finite for every $y \in S$, which allows

us to pick $f^* \in \mathbb{F}$ such that

$$e^{\lambda(C(x,f^*(x))-g(f))} \sum_y p_{x,y}(f^*(x))e^{\lambda h_{f,g(f),w}(y)}$$

$$= \sup_{a \in A(x)} \left[e^{\lambda(C(x,a)-g(f))} \sum_y p_{x,y}(a)e^{\lambda h_{f,g(f),w}(y)} \right], \qquad x \in S. \quad (48)$$

Combining these two last displays, it follows that, for each state $x \in S$,

$$e^{\lambda h_{f,g(f),w}(x)} \leq e^{\lambda(C(x,f^*(x))-g(f))} \sum_y p_{x,y}(f^*(x))e^{\lambda h_{f,g(f),w}(y)},$$

and then there exists a function $\Delta \colon S \to \mathbb{R}$ such that

$$\Delta(\cdot) \geq 0 \qquad (49)$$

and

$$e^{\lambda h_{f,g(f),w}(x)} = e^{-\Delta(x)+\lambda(C(x,f^*(x))-g(f))} \sum_y p_{x,y}(f^*(x))e^{\lambda h_{f,g(f),w}(y)}, \qquad x \in S; \quad (50)$$

notice that $g(f) - g(f^*) \leq 0$, by (46), and then

$$\lambda(g(f) - g(f^*)) \geq 0, \qquad (51)$$

since $\lambda < 0$. To complete the proof of part (i) it is sufficient to show that

$$\Delta(\cdot) = 0, \qquad (52)$$

since this equality combined with (48) and (50) yields that the equality occurs in (47). To establish (52), recall that the equality $A_n = f^*(X_n)$ always holds with probability 1 under f^*, so that, via (50) and the Markov property, it follows that, for each $x \in S$ and $n \in \mathbb{N}$,

$$e^{\lambda h_{f,g(f),w}(X_n)} = e^{-\Delta(X_n)+\lambda(C(X_n,f^*(X_n))-g(f))} \sum_y p_{X_n,y}(f^*(X_n))e^{\lambda h_{f,g(f),w}(y)}$$

$$= e^{-\Delta(X_n)+\lambda(C(X_n,A_n)-g(f))} \mathbb{E}_x^{f^*}\left[e^{\lambda h_{f,g(f),w}(X_{n+1})} \mid \mathcal{F}_n \right], \quad \mathbb{P}_x^{f^*}\text{-a.s.} \quad (53)$$

Now define the positive random variables

$$V_0 = e^{\lambda h_{f,g(f),w}(X_0)},$$

$$V_n = \exp\left\{ \sum_{t=0}^{n-1} [-\Delta(X_t) + \lambda(C(X_t, A_t) - g(f))] + \lambda h_{f,g(f),w}(X_n) \right\}, \quad n \in \mathbb{N}. \quad (54)$$

Observing that under f^* the random variable $\sum_{t=0}^{n} [-\Delta(X_t) + \lambda(C(X_t, A_t) - g(f))]$ is $\mathcal{F}_n$-measurable, and using the Markov property, it follows that, for every $x \in S$ and $n \in \mathbb{N}$,

$$\mathbb{E}_x^{f^*} \left[\exp \left\{ \sum_{t=0}^{n} [-\Delta(X_t) + \lambda(C(X_t, A_t) - g(f))] + \lambda h_{f,g(f),w}(X_{n+1}) \right\} \mid \mathcal{F}_n \right]$$

$$= \exp \left\{ \sum_{t=0}^{n} [-\Delta(X_t) + \lambda(C(X_t, A_t) - g(f))] \right\} \mathbb{E}_x^{f^*} [e^{\lambda h_{f,g(f),w}(X_{n+1})} \mid \mathcal{F}_n]$$

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

where the last equality is due to (53). This last relation and (54) lead to $\mathbb{E}_x^{f^*}[V_{n+1} \mid \mathcal{F}_n] = V_n$, i.e. $\{(V_n, \mathcal{F}_n)\}_{n \in \mathbb{N}}$ is a martingale with respect to $\mathbb{P}_x^{f^*}$. Via the optional sampling theorem it follows that, for every $x \in S$ and $n \in \mathbb{N}$,

$$e^{\lambda h_{f,g(f),w}(x)} = \mathbb{E}_x^{f^*}[V_0] = \mathbb{E}_x^{f^*}[V_{n \wedge T_w}]$$

$$= \mathbb{E}_x^{f^*} \left[\exp \left\{ \sum_{t=0}^{n \wedge T_w - 1} [\lambda(C(X_t, A_t) - g(f)) - \Delta(X_t)] + \lambda h_{f,g(f),w}(X_{n \wedge T_w}) \right\} \right].$$

Recalling that $\mathbb{P}_x^{f^*}[T_w < \infty] = 1$, by Lemma 1(i), and using that $X_{T_w} = w$ on the event $[T_w < \infty]$, it follows that

$$\lim_{n \to \infty} V_{n \wedge T_w} = \lim_{n \to \infty} \exp \left\{ \sum_{t=0}^{n \wedge T_w - 1} [\lambda(C(X_t, A_t) - g(f)) - \Delta(X_t)] + \lambda h_{f,g(f),w}(X_{n \wedge T_w}) \right\}$$

$$= \exp \left\{ \sum_{t=0}^{T_w - 1} [\lambda(C(X_t, A_t) - g(f)) - \Delta(X_t)] + \lambda h_{f,g(f),w}(w) \right\}, \quad \mathbb{P}_x^{f^*}\text{-a.s.} \quad (55)$$

Next, we compare V_n with the random variable $W_n(f^*)$ introduced in Lemma 5; see (39). Observe that

$$\sum_{t=0}^{n \wedge T_w - 1} [-\Delta(X_t) + \lambda(C(X_t, A_t) - g(f))] = \sum_{t=0}^{n \wedge T_w - 1} \lambda(C(X_t, A_t) - g(f^*)) - Q_n,$$

where

$$Q_n := (n \wedge T_w)\lambda(g(f) - g(f^*)) + \sum_{t=0}^{n \wedge T_w - 1} \Delta(X_t) \geq 0,$$

and the inequality is due to (49) and (51). Thus,

$$0 \le V_{n \wedge T_w} = \exp\left\{ \sum_{t=0}^{n \wedge T_w - 1} [-\Delta(X_t) + \lambda(C(X_t, A_t) - g(f))] + \lambda h_{f,g(f),w}(X_{n \wedge T_w}) \right\}$$

$$= \exp\left\{ \sum_{t=0}^{n \wedge T_w - 1} [\lambda(C(X_t, A_t) - g(f^*))] + \lambda h_{f,g(f),w}(X_{n \wedge T_w}) \right\} e^{-Q_n}$$

$$\le \exp\left\{ \sum_{t=0}^{n \wedge T_w - 1} [\lambda(C(X_t, A_t) - g(f^*))] + \lambda h_{f,g(f),w}(X_{n \wedge T_w}) \right\}$$

$$\le \exp\left\{ \sum_{t=0}^{n \wedge T_w - 1} [\lambda(C(X_t, A_t) - g(f^*))] + \lambda h_{f^*,g(f^*),w}(X_{n \wedge T_w}) \right\} e^{|\lambda| \, \|h_{f,g(f),w} - h_{f^*,g(f^*),w}\|}$$

$$= W_n(f^*) e^{|\lambda| \, \|h_{f,g(f),w} - h_{f^*,g(f^*),w}\|},$$

where $W_n(f^*)$ is as in (39). Since the equality

$$W_n(f^*) = \mathbb{E}_x^{f^*}\left[\exp\left\{ \lambda \sum_{t=0}^{T_w - 1} (C(X_t, A_t) - g(f^*)) \right\} \mid \mathcal{F}_n \right]$$

holds $\mathbb{P}_x^{f^*}$-a.s. for every $x \in S$ and $n \in \mathbb{N}$, by Lemma 5(iv), it follows that $\{W_n(f^*)\}_{n \in \mathbb{N}}$ is *uniformly integrable* with respect to $\mathbb{P}_x^{f^*}$, i.e. $\mathbb{E}_x^{f^*}[W_n(f^*)\mathbf{1}[W_n(f^*) > c]] \to 0$ as $c \to \infty$ uniformly in n; note that $W_n(f^*) \ge 0$, and see, for instance, Theorem 7.6.6, p. 300 of [1], or Theorem 16.14, p. 230 of [4]. Since $\|h_{f,g(f),w} - h_{f^*,g(f^*),w}\|$ is finite, by Lemma 5(iii), the relation $0 \le V_{n \wedge T_w} \le W_n(f^*) e^{|\lambda| \, \|h_{f,g(f),w} - h_{f^*,g(f^*),w}\|}$ obtained from the above display immediately implies that $\{V_{n \wedge T_w}\}_{n \in \mathbb{N}}$ is uniformly integrable with respect to $\mathbb{P}_x^{f^*}$. Combining this property with (55), an application of Theorem 7.5.2, p. 295 in [1] allows us to interchange the limit and the expectation to obtain that, for every $x \in S$,

$$e^{\lambda h_{f,g,w}(x)} = \lim_{n \to \infty} \mathbb{E}_x^{f^*}[V_{n \wedge T_w}] = \mathbb{E}_x^{f^*}\left[\lim_{n \to \infty} V_{n \wedge T_w} \right]$$

$$= \mathbb{E}_x^{f^*}\left[\exp\left\{ \sum_{t=0}^{T_w - 1} [\lambda(C(X_t, A_t) - g(f)) - \Delta(X_t)] + \lambda h_{f,g(f),w}(w) \right\} \right].$$

Setting $x = w$, it follows that

$$e^{\lambda h_{f,g(f),w}(w)} = \mathbb{E}_w^{f^*}\left[\exp\left\{ \sum_{t=0}^{T_w - 1} [\lambda(C(X_t, A_t) - g(f)) - \Delta(X_t)] + \lambda h_{f,g(f),w}(w) \right\} \right]$$

and then

$$1 = \mathbb{E}_w^{f^*}\left[\exp\left\{ \sum_{t=0}^{T_w - 1} [\lambda(C(X_t, A_t) - g(f)) - \Delta(X_t)] \right\} \right]$$

$$= \mathbb{E}_w^{f^*}\left[\exp\left\{ \lambda \sum_{t=0}^{T_w - 1} (C(X_t, A_t) - g(f^*)) \right\} e^{-Q} \right],$$

where $Q := T_w \lambda(g(f) - g(f^*)) + \sum_{t=0}^{T_w-1} \Delta(X_t) \geq 0$, and the inequality is due to (49) and (51). Observe now that $1 = \mathbb{E}_w^{f^*}\big[\exp\big\{\lambda \sum_{t=0}^{T_w-1}(C(X_t, A_t) - g(f^*))\big\}\big]$ by Lemma 5(i), and combining this equality with the two previous expressions it follows that $1 = \mathbb{P}_w^{f^*}[Q = 0]$, an equality that, combined with (49) and (51), implies that $\mathbb{P}_w^{f^*}\big[\sum_{t=0}^{T_w-1} \Delta(X_t) = 0\big] = 1$.

To conclude, observe that $\sum_{t=0}^{T_w-1} \Delta(X_t) \geq \Delta(X_0) \geq 0$ and $\sum_{t=0}^{T_w-1} \Delta(X_t) \geq \Delta(x)\mathbf{1}[T_x < T_w] \geq 0$, $x \in S \setminus \{w\}$, and recall that $\mathbb{P}_w^{f^*}[T_x < T_w] > 0$ by Lemma 1(ii), whereas $1 = \mathbb{P}_w[X_0 = w]$. Combining these relations with the two previous expressions, it follows that $\Delta(\cdot) = 0$. As already mentioned, this completes the proof of part (i). From this point, part (ii) follows via Remark 2(iii). $\qquad\square$

6. Proof of Theorem 1

In this section the main result of the paper is established. Essentially, the approach used to prove Theorem 1 consists in 'approximating' the original model $\mathcal{M}$ by a sequence $\{\mathcal{M}_n\}$ of MDPs satisfying the conditions in Theorem 2. Throughout the remainder of the section Assumptions 1 and 2, as well as the condition $\lambda < 0$, are enforced. The argument relies heavily on the preliminaries established in Section 5, as well as on Lemma 6. To begin with, *suppose* that the cost function C satisfies

$$C(x, a) \leq 0, \qquad (x, a) \in \mathbb{K}, \tag{56}$$

let $\{F_n\}_{n\in\mathbb{N}}$ be a sequence of subsets of S such that

$$F_k \text{ is finite}, \quad F_k \subset F_{k+1}, \quad k \in \mathbb{N}, \qquad \bigcup_{k\in\mathbb{N}} F_k = S, \tag{57}$$

and define the truncated cost function C_n by

$$C_n(x, a) := C(x, a)\mathbf{1}_{F_n}(x), \qquad (x, a) \in \mathbb{K}. \tag{58}$$

Now, let $f_0 \in \mathbb{F}$ be a fixed stationary policy, and recall that, for each $x \in S$, the action set $A(x)$ is a compact subspace of the metric space A, so that there exists a sequence of $\{D_n(x)\}_{n\in\mathbb{N}}$ of subsets of $A(x)$ such that

$$D_k(x) \text{ is finite}, \quad D_k(x) \subset D_{k+1}(x), \quad k \in \mathbb{N}, \qquad D(x) := \bigcup_{k\in\mathbb{N}} D_k(x) \text{ is dense in } A(x). \tag{59}$$

With this notation, for each $n \in \mathbb{N}$ define the collection $\{A_n(x)\}_{x\in S}$ of action sets as follows:

$$A_n(x) = D_n(x), \quad x \in F_n, \qquad A_n(x) = \{f_0(x)\}, \quad x \in S \setminus F_n. \tag{60}$$

Next, let

$$\mathcal{M}_n = (S, A, \{A_n(x)\}_{x\in S}, C_n, P), \qquad n \in \mathbb{N}, \tag{61}$$

be the MDP obtained from $\mathcal{M}$ after replacing the actions sets $\{A(x)\}_{x\in S}$ and the cost function C by $\{A_n(x)\}$ and C_n, respectively. It follows from (56)–(61) that, for every $n \in \mathbb{N}$, the conditions in Theorem 2 are satisfied by $\mathcal{M}_n$, and then there exists a stationary policy $f_n \in \prod_{x\in S} A_n(x)$ such that, given $w \in S$, the pair

$$(g_n, h_n(\cdot)) = (g(f_n), h_{f_n, g(f_n), w}(\cdot)) \in (-\infty, 0] \times \mathcal{B}(S) \tag{62}$$

has the following properties:

$$
e^{\lambda g_n + \lambda h_n(x)} = \sup_{a \in A_n(x)} \left[e^{\lambda C_n(x,a)} \sum_{y \in S} p_{x,y}(a) e^{\lambda h_n(y)} \right] =
$$
$$
= e^{\lambda C_n(x, f_n(x))} \sum_{y \in S} p_{x,y}(f_n(x)) e^{\lambda h_n(y)}, \qquad x \in S, \tag{63}
$$

$$
h_n(\cdot) \leq M_w \|C_n - g_n\|, \qquad h_n(w) = 0, \qquad -\|C\| \leq -\|C_n\| \leq g_n \leq 0, \tag{64}
$$

by Lemma 5 and (56). It follows that $g_n \in [-\|C\|, 0]$ and $h_n(x)b \in (-\infty, 2M_w\|C\|]$, $x \in S$, $n \in \mathbb{N}$. Via Cantor's diagonal method, it is possible to determine a subsequence of $(g_n, h_n(\cdot))$ (by notational convenience, still denoted by $(g_n, h_n(\cdot))$) which is convergent:

$$
\lim_{n \to \infty} g_n =: g^* \in [-\|C\|, 0], \qquad \lim_{n \to \infty} h_n(x) =: h^*(x) \in [-\infty, 2M_w\|C\|], \quad x \in S. \tag{65}
$$

The following lemma, using the conditions and definitions in (56)–(65), is the last step before the proof of Theorem 1.

Lemma 6. *The pair $(g^*, h^*(\cdot))$ in (65) satisfies the following properties.*

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

$$
e^{\lambda g^* + \lambda h^*(x)} \geq \sup_{a \in A(x)} \left[e^{\lambda C(x,a)} \sum_{y \in S} p_{x,y}(a) e^{\lambda h^*(y)} \right], \qquad x \in S, \tag{66}
$$

and then $(g^*, h^*(\cdot)) \in \tilde{\mathcal{G}}$; see Definition 1 and (24).

(ii) *For each $x \in S$,*
$$
J_*(x) = g^* = J^*(x). \tag{67}
$$

Proof. For (i), let $x \in S$ be arbitrary but fixed and select $n_x \in \mathbb{N}$ such that $x \in F_k$ if $k \geq n_x$. Next, let $d \in D(x)$ be arbitrary, so that there exists $n_{(d)} \in \mathbb{N}$ such that $d \in D_k(x)$ for $k \geq n_{(d)}$. It follows from (60) that $d \in A_n(x)$ for $n \geq \max\{n_x, n_{(d)}\}$, and then (63) implies

$$
e^{\lambda g_n + \lambda h_n(x)} \geq e^{\lambda C_n(x,d)} \sum_{y \in S} p_{x,y}(d) e^{\lambda h_n(y)} = e^{\lambda C(x,d)} \sum_{y \in S} p_{x,y}(d) e^{\lambda h_n(y)}, \quad n \geq \max\{n_x, n_{(d)}\},
$$

where the equality is due to the definition of C_n in (58). Taking the inferior limit as n goes to ∞ in this relation, (65) and Fatou's lemma together lead to

$$
e^{\lambda g^* + \lambda h^*(x)} \geq e^{\lambda C(x,d)} \liminf_{n \to \infty} \sum_{y \in S} p_{x,y}(d) e^{\lambda h_n(y)}
$$
$$
\geq e^{\lambda C(x,d)} \sum_{y \in S} p_{x,y}(d) \liminf_{n \to \infty} e^{\lambda h_n(y)}
$$
$$
= e^{\lambda C(x,d)} \sum_{y \in S} p_{x,y}(d) e^{\lambda h^*(y)}, \qquad d \in D(x).
$$

Next, given $a \in A(x)$, observe that (59) implies that there exists a sequence $\{d_k\}_{k \in \mathbb{N}}$ contained in $D(x)$ such that $\lim_{k \to \infty} d_k = a$; replacing d by d_k in the above display and taking the inferior

limit as k goes to ∞, from Fatou's lemma and the continuity conditions in Assumption 1, it follows that

$$
\begin{aligned}
\mathrm{e}^{\lambda g^* + \lambda h^*(x)} &\geq \liminf_{k \to \infty} \mathrm{e}^{\lambda C(x, d_k)} \sum_{y \in S} p_{x,y}(d_k) \mathrm{e}^{\lambda h^*(y)} \\
&\geq \mathrm{e}^{\lambda C(x, a)} \sum_{y \in S} \liminf_{k \to \infty} p_{x,y}(d_k) \mathrm{e}^{\lambda h^*(y)} = \mathrm{e}^{\lambda C(x, a)} \sum_{y \in S} p_{x,y}(a) \mathrm{e}^{\lambda h^*(y)}.
\end{aligned}
$$

Since $x \in S$ and $a \in A(x)$ are arbitrary, it follows that (66) holds. To complete the proof of part (i) we show that $h^*(\cdot)$ is a finite function. To achieve this goal, pick a stationary policy $f \in \mathbb{F}$ and define the sequence $\{\mathcal{S}_n\}$ of subsets of S as follows:

$$
\mathcal{S}_0 = \{w\}, \quad \mathcal{S}_k = \{y \in S \mid p_{xy}(f(x)) > 0 \text{ for some } x \in \mathcal{S}_{k-1}\}, \qquad k = 1, 2, 3, \ldots,
$$

and notice that $S = \bigcup_{k=0}^{\infty} \mathcal{S}_k$, by Lemma 1(ii). Using (65), to show that $h^*(\cdot)$ is finite it is sufficient to verify that, for each $k \in \mathbb{N}$,

$$
h^*(x) > -\infty \text{ for } x \in \mathcal{S}_k, \tag{68}
$$

a claim that will be proved by induction. Notice that $h^*(w) = 0$, by (64) and (65), and then the desired conclusion holds for $k = 0$. Suppose that (68) is valid for some $k \in \mathbb{N}$, let $\hat{y} \in \mathcal{S}_{k+1}$, and pick $\hat{x} \in \mathcal{S}_k$ such that

$$
p_{\hat{x}, \hat{y}}(f(\hat{x})) > 0. \tag{69}
$$

Now observe that (66) implies that

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

Recalling that $\lambda < 0$, the condition $h^*(\hat{x}) > -\infty$ implies that the left-most term in the above relation is finite, and then (69) allows us to conclude that $\mathrm{e}^{\lambda h^*(\hat{y})} < \infty$, so that $h^*(\hat{y}) > -\infty$; since $\hat{y} \in \mathcal{S}_{n+1}$ is arbitrary, it follows that (68) holds with $k + 1$ instead of k, concluding the induction argument. As already mentioned, this completes the proof of part (i).

For (ii), for each $n \in \mathbb{N}$ let $J^{*,n}(\cdot)$ be the superior limit optimal average cost function for model $\mathcal{M}_n$. In this case, (62) and part (ii) of Theorem 2 yield that $g_n = J^{*,n}(\cdot)$. Moreover, if f_n is the stationary policy in (63), then Theorem 2(ii) applied to model $\mathcal{M}_n$ yields

$$
g_n = J^{*,n}(x) = \lim_{k \to \infty} \frac{1}{\lambda k} \log \left(\mathbb{E}_x^{f_n} \left[\exp \left\{ \lambda \sum_{t=0}^{k-1} C_n(X_t, A_t) \right\} \right] \right), \qquad x \in S;
$$

see (7). Observe now that $C_n \geq C$, since the cost function is negative, so that the above display leads to

$$
\begin{aligned}
g_n &\geq \limsup_{k \to \infty} \frac{1}{\lambda k} \log \left(\mathbb{E}_x^{f_n} \left[\exp \left\{ \lambda \sum_{t=0}^{k-1} C(X_t, A_t) \right\} \right] \right) \\
&= \limsup_{k \to \infty} \frac{1}{k} J_k(f_n, x) = J(f_n, x) \geq J^*(x), \qquad x \in S;
\end{aligned}
$$

see (8) and (9). From this point, (65) leads to $g^* \geq J^*(\cdot)$. Recall now that $(g^*, h^*(\cdot)) \in \tilde{\mathcal{G}}$, by Lemma 6(ii), so that $g^* \leq J_*(\cdot)$, by Lemma 2(iii). Combining these two expressions with the inequality $J_*(\cdot) \leq J^*(\cdot)$ in (12), it follows that $J_*(\cdot) = g^* = J^*(\cdot)$. $\qquad\square$

The proof of the main result is finally presented below.

Proof of Theorem 1. Let $\lambda < 0$ be fixed, and suppose that Assumptions 1 and 2 hold. The proof has been split into two cases.

Case 1: $C(x, a) \leq 0$ *for every* $(x, a) \in \mathbb{K}$. Let $g^* \in \mathbb{R}$ and $h^*(\cdot)$ be as in (65). For (i), suppose that $g \in \mathcal{G}$. In this case, $(g, h(\cdot)) \in \tilde{\mathcal{G}}$ for some function $h(\cdot)$, and then $g \leq J_*(\cdot)$ by Lemma 2(iii). Assume that $g \leq J_*(\cdot)$. Since $J_*(\cdot) = g^*$, by Lemma 6(ii) it follows that $g \leq g^*$, and then $\lambda g \geq \lambda g^*$, since λ is negative. Combining this last inequality with (66), and recalling that $h^*(\cdot)$ is a finite function (by Lemma 6(ii)), it follows that $(g, h^*) \in \tilde{\mathcal{G}}$, and then $g \in \mathcal{G}$; see Definition 1 and (24).

For (ii), since $J_*(\cdot) = g^*$, by Lemma 6(ii), part (i) implies that $g \in \mathcal{G} \iff g \leq g^*$; then $g^* = \sup\{g \mid g \in \mathcal{G}\}$, and the desired conclusion follows from equality (67) established in Lemma 6(ii).

For (iii), since $h^*(\cdot)$ is finite, by Lemma 6(i), and $g^* = \sup\{g \mid g \in \mathcal{G}\}$, by part (ii), the conclusion follows from inequality (66) established in Lemma 6(i).

For (iv), consider the model $\mathcal{M}_n$ in (61), let $(g_n, h_n(\cdot)) \in [-\|C\|, 0] \times \mathcal{B}(S)$, and let $f_n \in \mathbb{F}$ be as in (62) and (63). It follows from Remark 2(iii) that $g_n = \lim_{k \to \infty} (1/k) J_k^{(n)}(f_n, x)$, $x \in S$, where $J_k^{(n)}(f_n, x)$ is the $(\lambda\text{-})$certainty equivalent of the random cost $\sum_{t=0}^{k-1} C_n(X_t, A_t)$ with respect to $\mathbb{P}_x^{f_n}$, so that, for each initial state x,

$$g_n = \lim_{k \to \infty} \frac{1}{k\lambda} \log \left(\mathbb{E}_x^{f_n} \left[\exp \left\{ \lambda \sum_{t=0}^{k-1} C_n(X_t, A_t) \right\} \right] \right).$$

We show that policy f_n is ε-optimal if n is large enough. To achieve this goal, using that $C \leq 0$, observe that $C \leq C_n$, by (58), and then the above display implies that

$$g_n \geq \limsup_{k \to \infty} \frac{1}{k\lambda} \log \left(\mathbb{E}_x^{f_n} \left[\exp \left\{ \lambda \sum_{t=0}^{k-1} C(X_t, A_t) \right\} \right] \right)$$

$$\geq \liminf_{k \to \infty} \frac{1}{k\lambda} \log \left(\mathbb{E}_x^{f_n} \left[\exp \left\{ \lambda \sum_{t=0}^{k-1} C(X_t, A_t) \right\} \right] \right),$$

i.e. $g_n \geq J(f_n, x) \geq J_-(f_n, x) \geq J_*(x)$, $x \in S$, $n \in \mathbb{N}$; see (7)–(11). Since $J_*(\cdot) = g^*$, by part (iii), using that $g_n \to g^*$ as $n \to \infty$, pick $n_* \in \mathbb{N}$ such that $g^* + \varepsilon \geq g_{n_*}$ to conclude from the above display that $g^* + \varepsilon \geq J(f_{n_*}, \cdot) \geq J_-(f_{n_*}, \cdot) \geq g^*$, so that $f_{n_*} \in \mathbb{F}$ is ε-optimal.

Case 2: $C \in \mathcal{B}(\mathbb{K})$. The notation in Lemma 1 and Remark 3 will be used. For (i), consider the cost function $C - \|C\|$, which is nonpositive. Using Remark 3, note that

$$g \in \mathcal{G}(C) \iff g - \|C\| \in \mathcal{G}(C - \|C\|)$$
$$\iff g - \|C\| \leq J_*(C - \|C\|, \cdot)$$
$$\iff g - \|C\| \leq J_*(C, \cdot) - \|C\|,$$

where the second equivalence is due to part (i) in Case 1 applied to the cost function $C - \|C\|$, and Lemma 1 was used in the last step. It follows that $g \in \mathcal{G}(C) \iff g \leq J_*(C, \cdot)$.

For (ii), an application of part (ii) of Case 1 to the cost function $C - \|C\| \leq 0$ yields

$$J_*(C - \|C\|, \cdot) = \sup\{g \mid g \in \mathcal{G}(C - \|C\|)\} = J^*(C - \|C\|, \cdot),$$

a relation that, via (13) and (23), leads to $J_*(C, \cdot) = \sup\{g \mid g \in \mathcal{G}(C)\} = J^*(C, \cdot)$.

For (iii), since $C - \|C\| \leq 0$, the third part of Case 1 yields that

$$\sup\{g \mid g \in \mathcal{G}(C - \|C\|)\} \in \mathcal{G}(C - \|C\|);$$

since $\mathcal{G}(C - \|C\|) = \mathcal{G}(C) - \|C\|$, it follows that $\sup\{g \mid g \in \mathcal{G}(C)\} \in \mathcal{G}(C)$.

For (iv), let $\varepsilon > 0$ be fixed. Applying the fourth part of Case 1 to the cost function $C - \|C\| \leq 0$, there exists $f_\varepsilon \in \mathbb{F}$ such that

$$\sup\{g \mid g \in \mathcal{G}(C - \|C\|)\} + \varepsilon \geq J(C - \|C\|, f_\varepsilon, \cdot)$$
$$\geq J_-(C - \|C\|, f_\varepsilon, \cdot) \geq \sup\{g \mid g \in \mathcal{G}(C - \|C\|)\},$$

a relation that, via (13) and (23), leads to $g^* + \varepsilon \geq J(C, f_\varepsilon, \cdot) \geq J_-(C, f_\varepsilon, \cdot) \geq g^*$, where $g^* = \sup\{g \mid g \in \mathcal{G}(C)\}$, showing that f_ε is ε-optimal for the cost function C. □

Remark 4. Finally, it is important to point out that the literature on risk-seeking is scarce. The problem addressed in this manuscript is interesting by itself. The case $\lambda < 0$ is associated with a risk-loving (risk-seeking) controller. A risk-seeking controller is willing to accept greater economic uncertainty in exchange for higher returns. Using a prospect theory approach, it has been found that people are risk-averse in the domain of gains but become risk-seeking in the domain of losses; see, for instance, [22].

7. Concluding remarks

This work has studied Markov decision chains endowed with average cost criteria. Besides standard continuity–compactness conditions, three essential assumptions determined the framework of the paper: (i) the simultaneous Doeblin condition holds, (ii) the state space is communicating under the action of each stationary policy, and (iii) the decision maker driving the system is risk-seeking. Within this context, a characterization of the optimal average cost was provided in Theorem 1, extending conclusions in [7] obtained under the condition that the controller is risk-averse. Another interesting result stated in part (iv) of the main theorem is the existence of ε-optimal stationary policies, which was obtained by truncating the cost function, instead of taking a policy that 'almost optimizes' the term within brackets of the optimality inequality (25). In this direction there is, at least, an interesting question to be analyzed in the future: Given $\varepsilon > 0$, assume that $f \in \mathbb{F}$ is derived from the optimality inequality in such a way that action $f(x)$ is an $\varepsilon|\lambda|$-maximizer of the logarithm of the term within brackets in (25), i.e. $e^{\lambda(g^*+\varepsilon)+\lambda h^*(x)} \leq \left[e^{\lambda C(x,f(x))} \sum_{y \in S} p_{x,y}(f(x))e^{\lambda h^*(y)}\right]$ for every $x \in S$. With this notation, establishing that f is ε-optimal is an interesting problem.

Funding information

Competing interests

There were no competing interests to declare which arose during the preparation or publication process of this article.

References

[1] ASH, R. B. (1972). *Real Analysis and Probability*. Academic Press, New York.

[2] BALAJI, S. AND MEYN, S. P (2000). Multiplicative ergodicity and large deviations for an irreducible Markov chain. *Stoch. Process. Appl.* **90**, 123–144.

[3] BÄUERLE, N. AND REIDER, U. (2011). *Markov Decision Processes with Applications to Finance*. Springer, New York.

[4] BILLINGSLEY, P. (2012). *Probability and Measure*. Wiley, New York.

[5] BORKAR, V. S. AND MEYN, S. P. (2002). Risk-sensitive optimal control for Markov decision process with monotone cost. *Math. Operat. Res.* **27**, 192–209.

[6] CAVAZOS-CADENA, R. (2009). Solutions of the average cost optimality equation for finite Markov decision chains: Risk-sensitive and risk-neutral criteria. *Math. Meth. Operat. Res.* **70**, 541–566.

[7] CAVAZOS-CADENA, R. (2018). Characterization of the optimal risk-sensitive average cost in denumerable Markov decision chains. *Math. Operat. Res.* **43**, 1025–1050.

[8] CAVAZOS-CADENA, R. AND FERNÁNDEZ-GAUCHERAND, E. (2002). Risk-sensitive control in communicating average Markov decision chains. In *Modelling Uncertainty: An Examination of Stochastic Theory, Methods and Applications*, eds. M. Dror, P. L'Ecuyer and F. Szidarovsky. Kluwer, Boston, MA, pp. 525–544.

[9] DENARDO, E. V. AND ROTHBLUM, U. G. (2006). A turnpike theorem for a risk-sensitive Markov decision process with stopping. *SIAM J. Control Optimization* **45**, 414–431.

[10] DI MASI, G. B. AND STETTNER, L. (2000). Infinite horizon risk sensitive control of discrete time Markov processes with small risk. *Systems Control Lett.* **40**, 305–321.

[11] DI MASI, G. B. AND STETTNER, L. (2007). Risk-sensitive control of discrete-time Markov processes with infinite horizon. *SIAM J. Control Optimization* **38**, 61–78.

[12] DI MASI, G. B. AND STETTNER, L. (2007). Infinite horizon risk sensitive control of discrete time Markov processes under minorization property. *SIAM J. Control Optimization* **46**, 231–252.

[13] HERNÁNDEZ-LERMA, O. (1989). *Adaptive Markov Control Processes*. Springer, New York.

[14] HOWARD, R. A. AND MATHESON, J. E. (1972). Risk-sensitive Markov decision processes. *Manag. Sci.* **18**, 356–369.

[15] JAŚKIEWICZ, A. (1989). Average optimality for risk sensitive control with general state space. *Ann. Appl. Prob.* **17**, 654–675.

[16] KONTOYIANNIS, I. AND MEYN, S. P. (2013). Spectral theory and limit theorems for geometrically ergodic Markov processes. *Ann. Appl. Prob.* **13**, 304–362.

[17] MEYER, C. D. (2000). *Matrix Analysis and Applied Linear Algebra*. SIAM, Philadelphia.

[18] PITERA, M. AND STETTNER, L. (2015). Long run risk sensitive portfolio with general factors. *Math. Meth. Operat. Res.* **82**, 265–293.

[19] PUTERMAN, M. (1994). *Markov Decision Processes*. Wiley, New York.

[20] SLADKÝ, K. (2008). Growth rates and average optimality in risk-sensitive Markov decision chains. *Kybernetika* **44**, 205–226.

[21] STETTNER, L. (1999). Risk sensitive portfolio optimization. *Math. Meth. Operat. Res.* **50**, 463–474.

[22] ZALESKIEWICZ, T. (2001). Beyond risk seeking and risk aversion: Personality and the dual nature of economic risk taking. *Europ. J. Pers.* **15**, S105–S122.