J. Math. Anal. Appl. 396 (2012) 58–69

Contents lists available at SciVerse ScienceDirect

Journal of Mathematical Analysis and Applications

journal homepage: www.elsevier.com/locate/jmaa

Average control of Markov decision processes with Feller transition probabilities and general action spaces☆

O.L.V. Costa [a], F. Dufour [b,*]

[a] Departamento de Engenharia de Telecomunicações e Controle, Escola Politécnica da Universidade de São Paulo, CEP: 05508 900-São Paulo, Brazil

[b] Université Bordeaux, IMB, Institut Mathématiques de Bordeaux, INRIA Bordeaux Sud Ouest, Team: CQFD, 351 cours de la Liberation, 33405 Talence Cedex, France

ARTICLE INFO

Article history:
Received 2 December 2011
Available online 9 June 2012
Submitted by Hans Zwart

Keywords:
Markov Decision Processes
Average cost
General Borel spaces
Feller transition probabilities
Non-compact action set
Policy iteration

ABSTRACT

This paper studies the average control problem of discrete-time Markov Decision Processes (MDPs for short) with general state space, Feller transition probabilities, and possibly non-compact control constraint sets $A(x)$. Two hypotheses are considered: either the cost function c is strictly unbounded or the multifunctions $A_r(x) = \{a \in A(x) : c(x,a) \leq r\}$ are upper-semicontinuous and compact-valued for each real r. For these two cases we provide new results for the existence of a solution to the average-cost optimality equality and inequality using the vanishing discount approach. We also study the convergence of the policy iteration approach under these conditions. It should be pointed out that we do not make any assumptions regarding the convergence and the continuity of the limit function generated by the sequence of relative difference of the α-discounted value functions and the Poisson equations as often encountered in the literature.

1. Introduction

There exists nowadays an extensive literature on discrete-time Markov Decision Processes (MDPs for short). Without attempting to present an exhaustive panorama of this vast field of research, the interested reader may consult a significant list of references on MDPs in the survey paper [1], the papers [2–9] and the books [10–16]. For the general state X and control Borel spaces $A(x)$ most of the works on the average control of MDPs assume that the cost function c satisfies some kind of restricted-growth unbounded condition, (that is, for some function $w(x) \geq 1$, $\sup_{a\in A(x)} |c(x,a)| \leq w(x)$, and that $Qw(x,a) \leq \beta w(x) + b$ for some $0 < \beta < 1$ and constant b), the control action set $A(x)$ is compact valued, the transition probabilities $Q(\cdot|x,a)$ are strongly continuous, and that the stationary policies give rise to geometric ergodic transition kernels. For instance, recently these kinds of assumptions have been considered in [2] to derive new conditions based on two inequalities which yield the existence of an average optimal deterministic stationary policy. However, as pointed out in [13] there are several important examples where some of these conditions are not satisfied. For instance, the quadratic cost associated to the LQG control problem does not satisfy the restricted-growth unbounded condition as pointed out in [17]. Due to that different combinations of assumptions have been considered in the literature.

The objective of this work is to study the optimality equations (equality and inequality) and the policy iteration algorithm for the average control problem of MDPs in general Borel spaces under a set of assumptions that, as far as the authors are aware of, had not been considered previously. We consider Feller transition probabilities and possibly non-compact control constraint sets $A(x)$. Two hypotheses are considered: either the cost function c is strictly unbounded or the multifunctions

☆ This work was partially supported by USP project MaCLinC.

* Corresponding author.
E-mail addresses: oswaldo@lac.usp.br (O.L.V. Costa), dufour@math.u-bordeaux1.fr (F. Dufour).

0022-247X/$ – see front matter © 2012 Elsevier Inc. All rights reserved.
doi:10.1016/j.jmaa.2012.05.073

$A_r(x) = \{a \in A(x) : c(x,a) \leq r\}$ are upper-semicontinuous and compact-valued for each real r. Our approach is partially based on the notion of *generalized liminf* of some sequence of functions, so that the limit function always exists and is lower semi-continuous, which yields the existence of an optimal selector. In this way there is no need to assume that the sequence of functions converges. For the inequality and equality optimality equations a similar approach was considered in [6] but, unlike the present paper, under the framework of compact valued control sets $A(x)$, the cost functions satisfying a restricted-growth unbounded condition, and that stationary policies give rise to geometric ergodic transition kernels.

The first main result of this paper is to provide new results for the existence of a solution to the average-cost optimality inequality (ACOI for short) and equality (ACOE for short) using the vanishing discount approach, under the hypotheses previously described. For the ACOE it is supposed that stationary policies give rise to positive Harris recurrent transition kernels (instead of geometric ergodic as in most of the previous papers). To the best knowledge of the authors, the existence of solutions to the ACOE in the context of not necessarily compact action spaces, weak Feller transition kernels and strictly unbounded cost functions has been discussed only in [9]. In [9], the author assumes that the state space is locally compact and that there exists a sequence of relative difference of the α-discounted value functions having the property to be equicontinuous. Here we work with a general Borel state space and we do not assume such *continuity* property but, on the other hand, we suppose a positive Harris recurrent property for the transition kernels. Moreover, it must be pointed out that our existence result for the ACOE does not guarantee any regularity property for the solution. To this end, we provide a technical condition implying that there exists a solution to the ACOE which is lower semi-continuous. Compared to the conditions imposed in [9], our result appears to be different and complementary.

The second main result of our work gives conditions for the convergence of the so-called policy iteration algorithm (PIA for short). The PIA has been considered for the strictly unbounded case in the Refs. [7,13]. In [7], the author works under some stability assumptions such as a uniform accessibility hypothesis to ensure that the sequence given by the solutions of the Poisson equations converge to a continuous limit function leading to the convergence of the PIA. In [13], the authors follow another approach based on the existence of a continuous limit function of a suitable subsequence of the Poisson equations. Here we follow a similar approach as in [13], but we assume a weaker hypothesis, see Eq. (46) in Assumption F. We prove that the PIA converges to the minimum average cost as defined in item (a) of Definition 11.1.1 in [14]. Again, compared to the conditions imposed in [7,13], our result appears to be different and complementary.

Notice that our results are derived without making any assumptions regarding the convergence and the continuity of the limit function generated by the sequence of relative difference of the α-discounted value functions and the Poisson equations as often encountered in the literature.

The paper is organized as follows. In Section 2 we introduce the notation, definitions, problem formulation, some of the main assumptions and some auxiliary results. Section 3 presents the main results regarding the existence of a solution to ACOI and ACOE. Finally the PIA is dealt with in Section 4.

2. Problem formulation and auxiliary results

The main goal of this section is to introduce the notation, definitions, problem formulation, and some of the main assumptions that will be used throughout the paper. We also present some auxiliary results that will be useful along the paper.

We follow closely the notation in [13]. We recall that X is a Borel space if it is a Borel subset of a complete and separable metric space, and its Borel σ-algebra is denoted by $\mathcal{B}(X)$. For X, Y Borel spaces, the family of all stochastic kernels on X given Y is denoted by $\mathcal{P}(X|Y)$. $\mathcal{P}(X)$ denotes the set of all probability measures on $(X, \mathcal{B}(X))$. Moreover, $\mathcal{P}(X)$ is considered as a topological space equipped with the weak topology. The Dirac measure centered on a fixed point $x \in X$ is denoted by δ_x.

We will denote by $\mathbb{M}(X)$ the set of measurable functions from X to $\mathbb{R}$. For $w \in \mathbb{M}(X)$ with $w : X \to [1, \infty)$, referred to as weight function, and $u \in \mathbb{M}(X)$, we define the w-norm of u as $\|u\|_w = \sup_{x \in X} \frac{|u(x)|}{w(x)}$. A function u is said to be w-bounded if $\|u\|_w < \infty$ (bounded if $\|u\| < \infty$ where $\|\cdot\|$ is the sup-norm). The set of w-bounded (bounded respectively) measurable functions defined on X is denoted by $\mathbb{B}_w(X)$ ($\mathbb{B}(X)$ respectively). We denote by $\mathbb{C}(X)$ the set of bounded continuous functions from X to $\mathbb{R}$, and by $\mathbb{L}_+(X)$ the set of non-negative lower semi-continuous functions from X to $\mathbb{R}$. For Q a stochastic kernel on X given Y, a probability measure $\mu \in \mathcal{P}(X)$ and $v \in \mathbb{M}(X)$ we define $Qv : Y \to \mathbb{R}$ as

$$Qv(y) := \int_X v(z) Q(dz|y), \tag{1}$$

and $\mu(v)$ as

$$\mu(v) := \int_X v(z) \mu(dz), \tag{2}$$

provided that the corresponding integrals are well defined and finite.

As in Definition 2.2.1 of [13] we consider a five-tuple for a Markov control model

$$\left(X, A, \{A(x)|x \in X\}, Q, c\right) \tag{3}$$

consisting of

(a) a Borel space X, representing the state space;
(b) a Borel space A, representing the control or action set;
(c) a family $\{A(x)|x \in X\}$ of non-empty measurable subsets $A(x)$ of A, where $A(x)$ denotes the set of feasible controls or actions when the system is in state $x \in X$, and with the property that

$$\mathbb{K} := \{(x, a)|x \in X, a \in A(x)\} \tag{4}$$

is a measurable subset of $X \times A$. Moreover $\mathbb{K}$ contains the graph of a measurable function from X to A;
(d) a stochastic kernel Q on X given $\mathbb{K}$;
(e) a measurable function $c : \mathbb{K} \to \mathbb{R}$.

We need the following definitions.

Definition 2.1 (*See Definition 2.3.1 of [13]*)**.** Φ denotes the set of all stochastic kernels φ in $\mathcal{P}(A|X)$ such that $\varphi(A(x)|x) = 1$ for all $x \in X$, and $\mathbb{F}$ stands for the set of all measurable functions $f : X \to A$, satisfying that $f(x) \in A(x)$ for all $x \in X$.

Definition 2.2. For $\varphi \in \Phi$ and $f \in \mathbb{F}$ we define the Markov kernels $Q_\varphi(C|x) := Q(C|x, \varphi(x)) := \int_{A(x)} Q(C|x, a)\varphi(x, da)$ and $Q_f(C|x) := Q(C|x, f(x))$ for any $C \in \mathcal{B}(A)$ and $x \in X$. Similarly we set $c_\varphi(x) := c(x, \varphi(x)) := \int_{A(x)} c(x, a)\varphi(x, da)$ and $c_f(x) := c(x, f(x))$ for any $x \in X$.

To introduce the optimal control problem we are concerned with, it is necessary to introduce different classes of control policy.

Definition 2.3. Define $H_0 = X$ and $H_n = \mathbb{K} \times H_{n-1}$ for $n \geq 1$. A control policy is a sequence $\pi = \{\pi_n\}$ of stochastic kernels π_n on A given H_n satisfying the following constraint: for all $h_n \in H_n$ and $n \geq 1$, $\pi_n(A(x_n)|h_n) = 1$, where $h_n = (x_0, a_0, \ldots, x_{n-1}, a_{n-1}, x_n)$. Let Π be the class of all policies. A policy $\pi = \{\pi_n\}$ is said to be a randomized stationary policy if there exists $\phi \in \Phi$ such that $\pi_n(\cdot|h_n) = \phi(\cdot|x_n)$, and it is denoted by ϕ^∞. A policy $\pi = \{\pi_n\}$ is said to be a stationary policy if there exists $f \in \mathbb{F}$ such that $\pi_n(\cdot|h_n) = \delta_{f(x_n)}(\cdot)$, and it is denoted by f^∞.

According to a standard convention, we identify $\mathbb{F}$ (respectively, Φ) with the class of all stationary (respectively, randomized stationary) policies. Therefore, we have $\mathbb{F} \subset \Phi \subset \Pi$. Let $(\Omega, \mathcal{F})$ be the canonical space consisting of the sample path $\Omega = (X \times A)^\infty$ and the associated σ-algebra $\mathcal{F}$. For any policy $\pi \in \Pi$ and any initial distribution ν on X, it can be defined a probability, labeled P_ν^π, and a stochastic process $\{(x_t, a_t)\}_{t\in\mathbb{N}}$ where $\{x_t\}_{t\in\mathbb{N}}$ is the state process and $\{a_t\}_{t\in\mathbb{N}}$ is the control process satisfying for any $B \in \mathcal{B}(X)$, $C \in \mathcal{B}(A)$ and $h_t \in H_t$ with $t \in \mathbb{N}$, $P_\nu^\pi(x_0 \in B) = \nu(B)$, $P_\nu^\pi(a_t \in C|h_t) = \pi_t(C|h_t)$, and $P_\nu^\pi(x_{t+1} \in B|h_t, a_t) = Q(B|x_t, a_t)$, see for example [13, Chapter 2] for such a construction. The expectation with respect to P_ν^π is denoted by E_ν^π. If $\nu = \delta_x$ for $x \in X$, we write P_x^π for P_ν^π and E_x^π for E_ν^π.

We consider the long run average cost problem, defined for any initial distribution ν on X, as

$$V(\nu) = \inf_{\pi\in\Pi} V(\nu, \pi), \tag{5}$$

where

$$V(\nu, \pi) = \overline{\lim_{n\to\infty}} \frac{1}{n} E_\nu^\pi \left(\sum_{t=0}^{n-1} c(x_t, a_t) \right) \tag{6}$$

and for notational simplicity we set $V(x) = V(\delta_x)$, $V(x, \pi) = V(\delta_x, \pi)$.

In what follows we define, for each $r > 0$, the multifunction $A_r(\cdot) : X \to A$ as

$$A_r(x) = \{a \in A(x) : c(x, a) \leq r\}. \tag{7}$$

We consider the following conditions.

Assumption A.

A.1 Q is weakly continuous on $\mathbb{K}$.
A.2 $A(\cdot)$ is u.s.c.

and the following assumptions on the cost function.

Assumption B.

B.1 $c(\cdot, \cdot)$ is l.s.c. on $\mathbb{K}$ and non-negative.
B.2 One of the two following conditions hold:
  (a) $c(\cdot, \cdot)$ is strictly unbounded on $\mathbb{K}$, that is, there exists an increasing sequence of compact sets $\mathbb{K}_n \uparrow \mathbb{K}$ such that $\lim_{n\to\infty} \inf_{(x,a)\in\mathbb{K}_n^c} c(x, a) = \infty$.
  (b) for each $r > 0$ the multifunction $A_r(\cdot) : X \to A$ defined in (7) is u.s.c. and compact-valued.

We have the following 2 auxiliary lemmas.

Lemma 2.4. *Suppose Assumptions* B.1 *and* B.2(a) *hold. Then the mapping c is inf-compact, that is, $A_r(x)$ is compact for any $x \in X$ and $r \in \mathbb{R}$.*

Proof. Consider $x \in X$. Since c is lower semi-continuous on $\mathbb{K}$, $A_r(x)$ is closed in $A(x)$ for any $r \in \mathbb{R}$. Consequently, $\{x\} \times A_r(x)$ is closed in $\mathbb{K}$. Denote $K_r = \{(x, a) \in \mathbb{K} : c(x, a) \leq r\}$. From Assumption B.2(a), there exists $n \in \mathbb{N}$ such that $K_r \subset \mathbb{K}_n$. Therefore, it follows that the closure of K_r is compact. However, since $\{x\} \times A_r(x) \subset K_r$, we get that $\{x\} \times A_r(x)$ is compact. Let us denote by pr_X, the projection of $X \times A$ into A. By definition of the product topology [18, Section 2.14], this mapping is continuous and so $pr_X\big(\{x\} \times A_r(x)\big) = A_r(x)$ is compact. □

Lemma 2.5. *Suppose Assumptions* A.1 *and* B *hold. Let v be a lower semi-continuous function on $\mathbb{K}$ bounded below by a constant K. Then the mapping $c + Qv$ is inf-compact, that is, for any $x \in X$ and $r \in \mathbb{R}$, $\{a \in A(x) : c(x, a) + Qv(x, a) \leq r\}$ is compact.*

Proof. Consider $x \in X$ and $r \in \mathbb{R}$. The set $\{a \in A(x) : c(x, a) + Qv(x, a) \leq r\}$ is closed in $A(x)$ since $c + Qv$ is lower semi-continuous on $\mathbb{K}$. However, $v \geq K$ and so

$$\{a \in A(x) : c(x, a) + Qv(x, a) \leq r\} \subset \{a \in A(x) : c(x, a) \leq r - K\}.$$

If Assumption B.2(a) holds then from Lemma 2.4 we have that $\{a \in A(x) : c(x, a) \leq r - K\}$ is compact and the result follows. If Assumption B.2(b) holds then by definition we have that $\{a \in A(x) : c(x, a) \leq r - K\}$ is compact, completing the proof. □

As in [19] the following definition of the generalized inferior limit will be used along the paper.

Definition 2.6. Let S be a Borel space and let $\{w_n\}$ be a sequence of functions in $\mathbb{M}(S)$. The generalized inferior limit of the sequence $\{w_n\}$ denoted by $\underline{\lim}^g{}_{n\to\infty} w_n$ is defined as

$$\underline{\lim_{n\to\infty}}^g w_n(s) = \sup_{k\geq 1}\sup_{\epsilon>0}\Big(\inf_{m\geq k}\inf_{\{y:d(y,s)<\epsilon\}} w_m(y)\Big) \tag{8}$$

where $d(\cdot, \cdot)$ is the metric in S. For notational convenience, $\underline{\lim}^g{}_{n\to\infty} w_n$ will be also denoted by w_*.

The following properties from the generalized inferior limit will be used in the sequel.

Proposition 2.7. *Let $\{w_n\}$ be a sequence of nonnegative functions in $\mathbb{M}(S)$ and consider an arbitrary $s \in S$. In this case, $w_*(s)$ as defined in (8) satisfies the following properties:*

(i) *For any sequence $\{s_n\}$ such that $s_n \to s$, it follows that $\underline{\lim}_{n\to\infty} w_n(s_n) \geq w_*(s)$, and there exists a sequence $\{s_n^*\}$ such that $s_n^* \to s$ and $\underline{\lim}_{n\to\infty} w_n(s_n^*) = w_*(s)$.*
(ii) *$w_* \in \mathbb{L}_+(S)$.*
(iii) *(Generalized Fatou's Lemma) Suppose that $\{\mu_n\}$ is a sequence of probability measures in $\mathcal{P}(S)$ and that $\{\mu_n\}$ converges weakly to a $\mu \in \mathcal{P}(S)$. Then*

$$\underline{\lim_{n\to\infty}} \int_S w_n(s)\mu_n(ds) \geq \int_S w_*(s)\mu(ds). \tag{9}$$

Proof. For the proof of (i) see Lemma 4.1 in [20]. For (ii) see Lemma 3.1 in [21]. For (iii), see [22]. □

As a consequence of the previous result, we obtain the following technical proposition.

Proposition 2.8. *Suppose Assumption* A.1 *holds. Let $\{w_n\}$ be a sequence of nonnegative functions in $\mathbb{M}(X)$ and $\{Qw_n\}$ the associated sequence of nonnegative functions in $\mathbb{M}(\mathbb{K})$ defined as in (1). Then*

$$\underline{\lim_{n\to\infty}}^g Qw_n \geq Qw_*. \tag{10}$$

Proof. Let $\{(y_n, a_n)\}$ be a sequence in $\mathbb{K}$ such that $(y_n, a_n) \to (y, a)$. From Assumptions A.1 we have that $Q(\cdot|y_n, a_n)$ converges weakly to $Q(\cdot|y, a)$. From (9) we have that

$$\begin{aligned}\underline{\lim_{n\to\infty}} Qw_n(y_n, a_n) &= \underline{\lim_{n\to\infty}} \int_X w_n(z)Q(dz|y_n, a_n)\\ &\geq \int_X w_*(z)Q(dz|y, a) = Qw_*(y, a)\end{aligned} \tag{11}$$

and the result follows from (i) in Proposition 2.7 since (11) holds for any convergent sequence $\{(y_n, a_n)\}$ to (y, a). □

3. The average cost optimality inequality and equality

The goal of this section is to derive new results for the existence of a solution to the ACOI and ACOE using the so-called vanishing discount approach. Some auxiliary results and the main assumptions related to the vanishing discount approach are presented in Section 3.1. Section 3.2 presents the main result regarding the existence of a solution for the ACOI, while Section 3.3 deals with the ACOE.

3.1. The vanishing discount approach

The main goal of this subsection is to present some auxiliary results and the main assumptions related to the so-called vanishing discount approach. For this we consider the following expected discounted Markov control problems for $0 < \alpha < 1$:

$$V_\alpha(x) = \inf_{\pi\in\Pi} V_\alpha(x,\pi), \tag{12}$$

where

$$V_\alpha(x,\pi) = E_x^\pi\left(\sum_{t=0}^{\infty}\alpha^t c(x_t,a_t)\right). \tag{13}$$

The following result has been shown in Theorem 4.2 of [23].

Proposition 3.1. *Suppose Assumptions A and B.1 hold. Consider $0 < \alpha < 1$ (arbitrarily fixed). If $V_\alpha(x) < \infty$ for each $x \in X$ then V_α is the (pointwise) minimal function in $\mathbb{L}_+(X)$ that satisfies*

$$V_\alpha(x) = \min_{a\in A(x)}\Big(c(x,a) + \alpha QV_\alpha(x,a)\Big). \tag{14}$$

Moreover there exists $f_\alpha \in \mathbb{F}$ which minimizes the right hand side of (14), that is, for each $x \in X$,

$$\min_{a\in A(x)}\Big(c(x,a) + \alpha QV_\alpha(x,a)\Big) = c(x,f_\alpha(x)) + \alpha QV_\alpha(x,f_\alpha(x)). \tag{15}$$

We consider the following conditions.

Assumption C. There exists positive constants N, M, a state $\bar{x} \in X$, an upper semi-continuous function $b : X \to [1,\infty)$ and $0 < \alpha_0 < 1$ such that for any $\alpha \in [\alpha_0, 1)$,

(a) $(1-\alpha)V_\alpha(\bar{x}) \le M$, and
(b) $-N \le V_\alpha(x) - V_\alpha(\bar{x}) \le b(x)$ for all $x \in X$.

For notational convenience, define

$$H_\alpha(x) = V_\alpha(x) - V_\alpha(\bar{x}), \qquad \rho_\alpha = (1-\alpha)V_\alpha(\bar{x}). \tag{16}$$

From Assumption C and Lemma 3.2 in [8] there exists ρ_* and a sequence of discount factors $\alpha_n \uparrow 1$ such that $\lim_{n\to\infty}(1-\alpha_n)V_{\alpha_n}(x) = \rho_*$ for all $x \in X$. Set $\rho_n = (1-\alpha_n)V_{\alpha_n}(\bar{x})$, $h_n = H_{\alpha_n}$. From (14) and (15) we have that there exists $f_{\alpha_n} \in \mathbb{F}$ such that for any $x \in X$,

$$\begin{aligned}\rho_n + h_n(x) &= \min_{a\in A(x)}\big(c(x,a) + \alpha_n Qh_n(x,a)\big)\\ &= c(x,f_{\alpha_n}(x)) + \alpha_n Qh_n(x,f_{\alpha_n}(x)).\end{aligned} \tag{17}$$

Moreover, we have $-N \le h_n(x) \le b(x)$, for any $n \in \mathbb{N}$, and $x \in X$. Consequently, for any $x \in X$,

$$-N \le h_*(x) \le b(x), \tag{18}$$

since b is upper semi-continuous.

3.2. The average cost optimality inequality

We have the following theorem ensuring the existence of a solution to the ACOI. Notice that this solution is lower semi-continuous and bounded below.

Theorem 3.2. *Suppose that Assumptions A–C hold. Then for some $f_* \in \mathbb{F}$, we have that*

$$\begin{aligned}\rho_* + h_*(x) &\geq \min_{a\in A(x)}\Big(c(x,a) + Qh_*(x,a)\Big)\\ &= c(x, f_*(x)) + Qh_*(x, f_*(x)).\end{aligned} \tag{19}$$

Moreover f_*^∞ is AC-optimal and $\rho_* = V(x) = V(x, f_*^\infty)$ for all $x \in X$.

Proof. First let us show the result by assuming B.2(a). Consider $\hat{x} \in X$. From Proposition 2.7(i) we can find a sequence $\{x_n\}$ in X such that $x_n \to \hat{x}$ and $\underline{\lim}_{n\to\infty} h_n(x_n) = h_*(\hat{x})$. In what follows set $a_n = f_{\alpha_n}(x_n)$ where f_α is as defined in (15). We have that there exists a further subsequence such that $\lim_{p\to\infty} h_{n_p}(x_{n_p}) = h_*(\hat{x})$. Given an arbitrary $\epsilon > 0$ consider a positive integer p_ϵ such that $h_*(\hat{x}) + \frac{\epsilon}{2} \geq h_{n_p}(x_{n_p})$ and $\rho_* + \frac{\epsilon}{2} \geq \rho_{n_p}$ whenever $p \geq p_\epsilon$. From (17) we get that for all $p \geq p_\epsilon$,

$$\epsilon + \rho_* + h_*(\hat{x}) \geq \rho_{n_p} + h_{n_p}(x_{n_p}) = c(x_{n_p}, a_{n_p}) + \alpha_{n_p} Qh_{n_p}(x_{n_p}, a_{n_p}). \tag{20}$$

Noticing that, from item (b) of Assumption C, $h_{n_p}(y) + N \geq 0$ for all $y \in X$, we get from (20) that for all $p \geq p_\epsilon$,

$$\epsilon + N + \rho_* + h_*(\hat{x}) \geq c(x_{n_p}, a_{n_p}). \tag{21}$$

Set now the measure $\mu_p \in \mathcal{P}(\mathbb{K})$ as follows: $\mu_p(dx, da) = \delta_{x_{n_p}}(dx)\delta_{a_{n_p}}(da)$. From (21) we get that for all $p \geq p_\epsilon$, $\sup_{p\geq p_\epsilon} \int_{\mathbb{K}} c(y,a)\mu_p(dy,da) = \sup_{p\geq p_\epsilon} c(x_{n_p}, a_{n_p}) < \infty$. From Assumption B.2, we get that the sequence $\{\mu_p\}_{p\geq p_\epsilon}$ is tight in $\mathcal{P}(\mathbb{K})$ (see, for instance, Proposition 1.4.15 in [24]). It follows that we can find a further subsequence such that μ_{p_i} converges weakly to a probability measure $\mu \in \mathcal{P}(\mathbb{K})$. Define now μ_X the projection of μ on X, that is, $\mu_X(B) = \mu(B \times A)$, $\forall B \in \mathcal{B}(X)$. From Proposition D.8 in [13], there exists a stochastic kernel $\varphi \in \Phi$ such that $\mu(B \times C) = \int_B \varphi(C|x)\mu_X(dx)$, $\forall B \in \mathcal{B}(X)$, $\forall C \in \mathcal{B}(A)$.

Consider any function $\hat{v} \in \mathbb{C}(X)$ and set the function $v \in \mathbb{C}(\mathbb{K})$ as $v(x,a) = \hat{v}(x)$ for all $(x,a) \in \mathbb{K}$. From continuity of $\hat{v}$ we get that $\hat{v}(x_{n_{p_i}}) \to \hat{v}(\hat{x})$ and from the weak convergence of μ_{p_i} to μ we get that $\hat{v}(x_{n_{p_i}}) = \mu_{p_i}(v) \to \mu(v) = \mu_X(\hat{v})$. Thus we have that for all $\hat{v} \in \mathbb{C}(X)$, $\delta_{\hat{x}}(\hat{v}) = \mu_X(\hat{v})$. Since $\mathbb{C}(X)$ is a separating family for $\mathcal{P}(X)$ (see for example Theorem 13.11 in [25]), we get that $\mu_X(dx) = \delta_{\hat{x}}(dx)$.

From Assumption B.1, $c(\cdot,\cdot)$ is l.s.c. and therefore, from Proposition E.2 in [13], it follows that

$$\underline{\lim}_{i\to\infty} \int_{\mathbb{K}} c(y,a)\mu_{p_i}(dy,da) \geq \int_{\mathbb{K}} c(y,a)\mu(dy,da) = \int_{A(\hat{x})} c(\hat{x},a)\varphi(da|\hat{x}). \tag{22}$$

Since $\underline{\lim}^g{}_{i\to\infty} h_{n_{p_i}} \geq \underline{\lim}^g{}_{n\to\infty} h_n = h_*$, we obtain from Proposition 2.8 (adding N to get non negative functions, for simplicity we omit this) that for any $(x,a) \in \mathbb{K}$

$$\underline{\lim}^g_{i\to\infty} Qh_{n_{p_i}}(x,a) \geq Qh_*(x,a),$$

and using Proposition 2.7(iii) we get that

$$\underline{\lim}_{i\to\infty} \int_{\mathbb{K}} Qh_{n_{p_i}}(y,a)\mu_{p_i}(dy,da) \geq \int_{\mathbb{K}} Qh_*(y,a)\mu(dy,da) = \int_{A(\hat{x})} Qh_*(\hat{x},a)\varphi(da|\hat{x}). \tag{23}$$

Noticing that $c(x_{n_{p_i}}, a_{n_{p_i}}) = \int_{\mathbb{K}} c(y,a)\mu_{p_i}(dy,da)$ and $Qh_{n_{p_i}}(x_{n_{p_i}}, a_{n_{p_i}}) = \int_{\mathbb{K}} Qh_{n_{p_i}}(y,a)\mu_{p_i}(dy,da)$ we get, by combining (20) with (22) and (23), that

$$\begin{aligned}\epsilon + \rho_* + h_*(\hat{x}) &\geq \underline{\lim}_{i\to\infty}\Big(c(x_{n_{p_i}}, a_{n_{p_i}}) + \alpha_{n_{p_i}} Qh_{n_{p_i}}(x_{n_{p_i}}, a_{n_{p_i}})\Big)\\ &\geq \underline{\lim}_{i\to\infty} \int_{\mathbb{K}} c(y,a)\mu_{p_i}(dy,da) + \underline{\lim}_{i\to\infty} \int_{\mathbb{K}} Qh_{n_{p_i}}(y,a)\mu_{p_i}(dy,da)\\ &\geq \int_{A(\hat{x})}\Big(c(\hat{x},a) + Qh_*(\hat{x},a)\Big)\varphi(da|\hat{x}).\end{aligned} \tag{24}$$

From Proposition D.8 in [13] we get that there exists $\hat{a} \in A(\hat{x})$ such that

$$\int_{A(\hat{x})}\Big(c(\hat{x},a) + Qh_*(\hat{x},a)\Big)\varphi(da|\hat{x}) \geq c(\hat{x},\hat{a}) + Qh_*(\hat{x},\hat{a}) \geq \min_{a\in A(\hat{x})}\Big(c(\hat{x},a) + Qh_*(\hat{x},a)\Big) \tag{25}$$

and from (24) and (25) we get that for every $\epsilon > 0$, $\epsilon + \rho_* + h_*(\hat{x}) \geq \min_{a\in A(\hat{x})}\Big(c(\hat{x},a) + Qh_*(\hat{x},a)\Big)$. Taking the limit as $\epsilon \downarrow 0$ we get the first inequality in (19). From Proposition 2.7(ii) we get that $h_* + N$ is in $\mathbb{L}_+(X)$ and therefore h_* is l.s.c. and bounded below. Combining Proposition D.6 in [13] and Lemma 2.5, it follows that there exists $f_* \in \mathbb{F}$ satisfying (19). Finally,

the claim that f_*^∞ is AC-optimal and $\rho_* = V(x) = V(x, f_*^\infty)$ for all $x \in X$ follows as in Theorem 3.3 of [8]. This shows the result with Assumption B.2(a).

For Assumption B.2(b), the proof follows the same steps up to Eq. (21). Considering $r > \epsilon + N + \rho_* + h_*(\hat{x})$ we have that for all $p \geq p_\epsilon$, $a_{n_p} \in A_r(x_{n_p})$ (recall the definition of $A_r(\cdot)$ from (7)). From Assumption B.2(b), $A_r(\cdot) : X \to A$ is u.s.c. and compact-valued. From Berge's theorem (see [26] or [27, Theorem 7.4.2]), $G := \{\hat{x} \times A_r(\hat{x})\} \cup \left(\cup_{\{p \geq p_\epsilon\}}(\{x_{n_p}\} \times A_r(x_{n_p}))\right)$ is a compact subset of $\mathbb{K}$. Therefore we can find a convergent subsequence $(x_{n_{p_i}}, a_{n_{p_i}}) \to (\hat{x}, \hat{a}) \in G$, and thus $\hat{a} \in A_r(\hat{x})$. The proof follows then the same steps as from Eq. (22), setting $\mu_{p_i}(dx, da) = \delta_{x_{n_{p_i}}}(dx)\delta_{a_{n_{p_i}}}(da)$ and $\mu(dx, da) = \delta_{\hat{x}}(dx)\delta_{\hat{a}}(da)$. □

3.3. The average cost optimality equality

We want to derive now a sufficient condition for the existence of a solution to the ACOE. We introduce the following assumptions.

Assumption D.

D.1 $Q_f b(x) < \infty$ for any $x \in X$, and $f \in \mathbb{F}$,
D.2 for each $f \in \mathbb{F}$ there exists $\mu_f \in \mathbb{P}(X)$ such that
(a) $\mu_f(b) < \infty$,
(b) for any $x \in X$ and $v \in \mathbb{B}_b(X)$, $|Q_f^n v(x) - \mu_f(v)| \to 0$ as $n \to \infty$.

We have the following result.

Theorem 3.3. *Suppose that Assumptions A–D hold. Consider $f_* \in \mathbb{F}$ and ρ_* as in Theorem 3.2 and μ_{f_*} as in Assumption D.2. Then there exists $w \in \mathbb{B}_b(X)$ such that*

$$\rho_* + w(x) = \inf_{a \in A(x)} \Big(c(x, a) + Q\, w(x, a) \Big) \tag{26}$$

and moreover

$$\inf_{a \in A(x)} \Big(c(x, a) + Q\, w(x, a) \Big) = c(x, f_*(x)) + Q\, w(x, f_*(x)) \quad \mu_{f_*}\text{-a.s.} \tag{27}$$

Proof. Define $h^*(x) = \overline{\lim}_{n \to \infty} h_n(x)$. Remark that $h^* \in \mathbb{B}_b(X)$ by using Assumption C. From (17) we have that for any $a \in A(x)$,

$$\rho_n + h_n(x) \leq c(x, a) + \alpha_n Q h_n(x, a). \tag{28}$$

According to Assumption D.1, $Qb(x, a) < \infty$ for any $a \in A(x)$ and since, from Assumption C, $h_n \leq b$, we have from the Fatou's Lemma that $\overline{\lim}_{n \to \infty} Q h_n(x, a) \leq Q h^*(x, a)$. Therefore from (28) we get that

$$\rho_* + h^*(x) \leq c(x, a) + Q h^*(x, a). \tag{29}$$

From (29) we obtain that

$$\rho_* + h^*(x) \leq \inf_{a \in A(x)} \big(c(x, a) + Q h^*(x, a) \big) \leq c_{f_*}(x) + Q_{f_*} h^*(x). \tag{30}$$

Set $u = h_* - h^*$. Combining the fact that $h^* \in \mathbb{B}_b(X)$ and Eq. (18), it follows that $u \in \mathbb{B}_b(X)$. From Eqs. (19) and (30), we get that for all $x \in X$,

$$u(x) \geq Q_{f_*} u(x). \tag{31}$$

From Assumption D.2 and that $u \in \mathbb{B}_b(X)$ we get that $\mu_{f^*}(|u|) < \infty$. Now, from Assumption D.2 and (31) we have that $u(x) \geq Q_{f_*}^n u(x) \to \mu_{f^*}(u)$ as $n \to \infty$, showing that $\inf_{y \in X}(h_*(y) - h^*(y))$ is finite. Moreover we get from Assumption D.2 and the same arguments as in Lemma 7.5.12 in [14] that on a set $\mathcal{N} \in \mathcal{B}(X)$ such that $\mu_{f_*}(\mathcal{N}) = 1$ we have that $h_*(x) = h^*(x) + d$ for all $x \in \mathcal{N}$, where $d = \inf_{y \in X}(h_*(y) - h^*(y))$. In any case we have that for all $x \in X$, $h_*(x) \geq h^*(x) + d$. Define now $w_0 = h_*$ and

$$w_{n+1}(x) = \inf_{a \in A(x)} \big(c(x, a) + Q\, w_n(x, a) \big) - \rho_*. \tag{32}$$

We have that for any $n \in \mathbb{N}$,

$$h^*(x) + d \leq w_{n+1}(x) \leq w_n(x) \leq h_*(x). \tag{33}$$

Indeed for $n = 1$ it follows from (19) that

$$w_1(x) = \min_{a\in A(x)}\big(c(x,a) + Qh_*(x,a)\big) - \rho_* \le h_*(x) = w_0(x),$$

and from (30) that

$$\begin{aligned} w_1(x) &= \min_{a\in A(x)}\big(c(x,a) + Qh_*(x,a)\big) - \rho_* \\ &\ge \inf_{a\in A(x)}\big(c(x,a) + Qh^*(x,a)\big) + d - \rho_* \\ &\ge h^*(x) + d \ge h_*(x) + d. \end{aligned}$$

Suppose it holds for n. Then

$$w_{n+1}(x) = \inf_{a\in A(x)}\big(c(x,a) + Q\,w_n(x,a)\big) - \rho_* \le \inf_{a\in A(x)}\big(c(x,a) + Q\,w_{n-1}(x,a)\big) - \rho_* = w_n(x) \tag{34}$$

and from (30) again that

$$\begin{aligned} w_{n+1}(x) &= \inf_{a\in A(x)}\big(c(x,a) + Q\,w_n(x,a)\big) - \rho_* \\ &\ge \inf_{a\in A(x)}\big(c(x,a) + Qh^*(x,a)\big) + d - \rho_* \ge h^*(x) + d. \end{aligned} \tag{35}$$

Consequently, combining Eqs. (34) and (35), we get (33).

Since $\{w_n\}$ is monotone decreasing, there exists $w \in \mathbb{B}_b(X)$ such that $w_n \downarrow w$. Clearly from (32) we have that for any $a \in A(x)$,

$$w_{n+1}(x) \le c(x,a) + Q\,w_n(x,a) - \rho_*. \tag{36}$$

Since $w_n \in \mathbb{B}_b(X)$ we have, from Assumption D.1 and Eq. (36) and the monotone convergence theorem, that

$$\begin{aligned} w(x) + \rho_* &= \lim_{n\to\infty} w_{n+1}(x) + \rho_* \le c(x,a) + \lim_{n\to\infty}\int_X w_n(y)Q(dy|x,a) \\ &= c(x,a) + \int_X \lim_{n\to\infty} w_n(y)Q(dy|x,a) = c(x,a) + Q\,w(x,a). \end{aligned} \tag{37}$$

From (37) it follows that

$$w(x) + \rho_* \le \inf_{a\in A(x)}\,(c(x,a) + Q\,w(x,a)). \tag{38}$$

On the other hand, Eq. (32) and the fact that $w_n \ge w$ yields that

$$w_{n+1}(x) \ge \inf_{a\in A(x)}\big(c(x,a) + Q\,w(x,a)\big) - \rho_*$$

and taking the limit as $n \to \infty$ we get that

$$w(x) + \rho_* \ge \inf_{a\in A(x)}\big(c(x,a) + Q\,w(x,a)\big). \tag{39}$$

Combining (38) and (39) we get (26).

Finally, it follows from (33) that for any $x \in \mathcal{N}$, $w(x) = h_*(x)$. Consequently, we get from (19) that for any $x \in \mathcal{N}$, $\inf_{a\in A(x)}\big(c(x,a) + Q\,w(x,a)\big) = w(x) + \rho_* = h_*(x) + \rho_* \ge c(x, f_*(x)) + Qh_*(x, f_*(x))$. However, $w \le h_*$ and so $\inf_{a\in A(x)}\big(c(x,a) + Q\,w(x,a)\big) = w(x) + \rho_* \ge c(x, f_*(x)) + Q\,w(x, f_*(x))$ for any $x \in \mathcal{N}$, showing the result. □

We conclude this subsection presenting Theorem 3.5 which provides conditions to guarantee that the limit function w as in Theorem 3.3 is l.s.c. and bounded below, so that the infimum in (26) can be replaced by minimum. We need first the following proposition.

Proposition 3.4. *Suppose that Assumptions A, B.1, B.2(b), C, D hold. Consider w_n as in Theorem 3.3. Then w_n is l.s.c. and bounded below by $-N + d$.*

Proof. First we notice that $w_n(x) \ge h^*(x) + d \ge -N + d$. Let us show now that w_n are l.s.c. by induction on n. For $n = 0$ we have that $w_0 = h_*$ which is l.s.c. and the result follows. Suppose now that w_n is l.s.c. We want to show that w_{n+1} is l.s.c. We have that $c + Q\,w_n$ is l.s.c. on $\mathbb{K}$ and bounded below, so that from Lemma 2.5 we have that the mapping $c + Q\,w_n$ is inf-compact. From Proposition D.6 in [13] we have that there exists $f_n \in \mathbb{F}$ such that

$$w_{n+1}(x) = \min_{a\in A(x)}\Big(c(x,a) + Q\,w_n(x,a)\Big) = c_{f_n}(x) + Q_{f_n}w_n(x).$$

Fix $\hat{x} \in X$ and consider a sequence $\{x_k\}$ such that $x_k \to \hat{x}$ and

$$\lim_{k\to\infty} w_{n+1}(x_k) = \underline{\lim}_{y\to\hat{x}} w_{n+1}(y).$$

Write $a_k = f_n(x_k)$. Since b is upper semi-continuous (see Assumption C), given $\epsilon > 0$ we can find k_ϵ such that whenever $k \geq k_\epsilon$, we have that

$$\begin{aligned}\epsilon + \rho_* + b(\hat{x}) &\geq \rho_* + b(x_k) \geq \rho_* + h_*(x_k) \geq \rho_* + w_{n+1}(x_k)\\ &= c(x_k, a_k) + Q w_n(x_k, a_k) \geq c(x_k, a_k) - N + d.\end{aligned}$$

So setting $r = \epsilon + \rho_* + b(\hat{x}) + N - d$ we get that $c(x_k, a_k) \leq r$ and thus $a_k \in A_r(x_k)$ for all $k \geq k_\epsilon$. From Assumption B.2(b), $A_r(\cdot) : X \to A$ is u.s.c. and compact-valued. From Berge's theorem (see [26] or [27, Theorem 7.4.2]), $G := \{\hat{x} \times A_r(\hat{x})\} \cup \left(\cup_{\{k\geq k_\epsilon\}}\{x_k \times A_r(x_k)\}\right)$ is a compact subset of $\mathbb{K}$. Therefore we can find a convergent subsequence $(x_{k_i}, a_{k_i}) \to (\hat{x}, \hat{a}) \in G$, and thus $\hat{a} \in A_r(\hat{x})$. Recalling that $c + Q w_n$ is l.s.c. we get that

$$\begin{aligned}\underline{\lim}_{y\to\hat{x}} w_{n+1}(y) &= \lim_{k\to\infty} w_{n+1}(x_k) = \underline{\lim}_{k\to\infty} \left(c(x_k, a_k) + Q w_n(x_k, a_k)\right)\\ &\geq \left(c(\hat{x}, \hat{a}) + Q w_n(\hat{x}, \hat{a})\right) \geq \min_{a\in A(\hat{x})} \left(c(\hat{x}, a) + Q w_n(\hat{x}, a)\right) = w_{n+1}(\hat{x})\end{aligned}$$

showing the desired result. □

Theorem 3.5. *Suppose that Assumptions A, B.1, B.2(b), C, D hold. Moreover, assume that for any set $\Gamma \in \mathcal{B}(X)$ such that $\mu_f(\Gamma) = 0$ for some $f \in \mathbb{F}$ we have that*

$$\sup_{x\in\Gamma} \sup_{a\in A(x)} Q(\Gamma; (x, a)) < 1. \tag{40}$$

Consider w as in Theorem 3.3. Then w is l.s.c. and bounded below by $-N + d$.

Proof. We use the same notation as in Theorem 3.3. We have that for any $z \in \mathcal{N}$, $h_*(z) = w(z) \leq w_n(z) \leq h_*(z)$, that is, $w_n(z) = w(z)$ for all n. Define now $\ell_n = \sup_{z\in\mathcal{N}^c}\left[w_n(z) - w(z)\right] \geq 0$. From (33) and recalling that $h_*(x) \leq h^*(x)$ we have that

$$h_*(x) + d \leq h^*(x) + d \leq w(x),$$

and therefore $0 \leq \ell_0 \leq -d$. Since d is finite according to the proof of Theorem 3.3, we get that ℓ_0 is finite.

From Eq. (32), it follows that for any $x \in X$

$$\begin{aligned}w_n(x) &= \min_{a\in A(x)}\left[c(x, a) + Q w(x, a) + Q(w_{n-1} - w)(x, a)\right] - \rho_*\\ &\leq \inf_{a\in A(x)}\left[c(x, a) + Q w(x, a)\right] - \rho_* + \sup_{a\in A(x)} Q(w_{n-1} - w)(x, a)\\ &\leq w(x) + \ell_{n-1} \sup_{a\in A(x)} Q(\mathcal{N}^c|(x, a))\end{aligned} \tag{41}$$

since $w_{n-1}(z) - w(z) = 0$ for $z \in \mathcal{N}$. From (40) and (41), and recalling that $\mu_{f_*}(\mathcal{N}) = 1$, we conclude that for some $\beta < 1$, $\ell_n \leq \beta\ell_{n-1}$, and thus $\ell_n \leq \beta^n\ell_0$. This shows that $0 \leq w_n(x) - w(x) \leq \beta^n\ell_0$ since $w_n(z) = w(z)$ for $z \in \mathcal{N}$, leading to the uniform convergence of w_n to w. The result follows after noticing, from Proposition 3.4, that w_n is l.s.c. for each n. Indeed given arbitrary $\epsilon > 0$ we can find n_ϵ such that $0 \leq w_n(y) - w(y) \leq \epsilon$ for all $y \in X$ and all $n \geq n_\epsilon$. Therefore for all $n \geq n_\epsilon$,

$$\epsilon + \underline{\lim}_{y\to x} w(y) \geq \underline{\lim}_{y\to x} w_n(y) \geq w_n(x) \geq w(x)$$

and since $\epsilon > 0$ is arbitrary, the result follows. □

4. Polity iteration algorithm

In this section we consider the so-called policy iteration algorithm for the average cost Markov control problem. After introducing the main assumptions and some auxiliary results we present Theorem 4.3, which shows the existence of a limit function satisfying an average optimality equation. From this optimality equation it is shown in Theorem 4.5 the convergence of the PIA to the minimum average cost criteria (see Definition 11.1.1 in [14]).

We start by considering the following assumptions.

Assumption E.

E.1 For each $f \in \mathbb{F}$, there exists an invariant probability measure μ_f for $Q(\cdot|\cdot, f(\cdot))$ such that $J(f) := \int_X c(x, f(x))\mu_f(dx) < \infty$.

E.2 For each $f \in \mathbb{F}$ there exists $v_f : X \to \mathbb{R}$ solution of the Poisson equation

$$\gamma_f + v_f(x) = c(x, f(x)) + Q v_f(x, f(x)) \tag{42}$$

such that v_f is lower semi-continuous.

E.3 There exists $M > 0$ and an upper semi-continuous function $d : X \to [1, \infty)$ such that, for any $f \in \mathbb{F}$, $-M \le v_f(x) \le d(x)$ and $\mu_f(d) < \infty$.

The PIA works as follows:

(1) For $f_n \in \mathbb{F}$, obtain (γ_n, v_n) solution of the Poisson equation (42) as in E.2, so that $\mu_{f_n}(|v_n|) < \infty$, v_n is l.s.c. and

$$\gamma_n + v_n(x) = c(x, f_n(x)) + Q v_n(x, f_n(x)). \tag{43}$$

(2) Obtain $f_{n+1} \in \mathbb{F}$ such that

$$\min_{a \in A(x)} \Big(c(x, a) + Q v_n(x, a) \Big) = c(x, f_{n+1}(x)) + Q v_n(x, f_{n+1}(x)). \tag{44}$$

We have the following proposition.

Proposition 4.1. *Suppose that Assumptions A, B.1, E hold. Then there exists $f_{n+1} \in \mathbb{F}$ such that (44) holds and $\gamma_n = J(f_n) \ge J(f_{n+1}) = \gamma_{n+1}$.*

Proof. Consider $f_n \in \mathbb{F}$ and v_n a solution of the P.E. as stated in E.2 associated to f_n. It means that (43) holds. Integrating (43) with respect to μ_{f_n} we get from E.1–E.3 that $\gamma_n = J(f_n)$. From B.1–A.2 and noticing from E.2–E.3 that v_n is l.s.c. and uniformly bounded below, we get from Proposition D.6 in [13] and Lemma 2.5 that there exists $f_{n+1} \in \mathbb{F}$ such that (44) holds. Since $\min_{a \in A(x)} \Big(c(x, a) + Q v_n(x, a) \Big) \le c(x, f_n(x)) + Q v_n(x, f_n(x)) = \gamma_n + v_n(x)$ we have from (44) that $\gamma_n + v_n(x) \ge c(x, f_{n+1}(x)) + Q v_n(x, f_{n+1}(x))$. Integrating with respect to $\mu_{f_{n+1}}$ we conclude that $\gamma_n = J(f_n) \ge J(f_{n+1}) = \gamma_{n+1}$. □

From Proposition 4.1 we have that $\gamma_n \downarrow \gamma_* \ge 0$. Define $v_* = \underline{\lim}^g_{n \to \infty} v_n$. We have the following proposition.

Proposition 4.2. *Suppose that Assumptions A, B and E hold. Then for some $g_* \in \mathbb{F}$, we have that*

$$\begin{aligned} \gamma_* + v_*(x) &\ge \min_{a \in A(x)} \Big(c(x, a) + Q v_*(x, a) \Big) \\ &= c(x, g_*(x)) + Q v_*(x, g_*(x)). \end{aligned} \tag{45}$$

Proof. It follows from (43) and the same arguments as in the proof of Theorems 3.2. □

We consider now the following assumptions.

Assumption F.

F.1 For each $f \in \mathbb{F}$, we have
(a) $Q_f d(x) < \infty$ for any $x \in X$,
(b) for any $x \in X$ and $v \in \mathbb{B}_d(X)$, $\lim_{n \to \infty} |Q_f^n v(x) - \mu_f(v)| \to 0$.

F.2 For (γ_n, v_n) solution of the Poisson equation (43), we have that for every $x \in X$,

$$\overline{\lim_{n \to \infty}} Q_{f_n}(v_n - v_{n-1})(x) \le 0. \tag{46}$$

We have the following theorem, showing the existence of a limit function satisfying an average optimality equation.

Theorem 4.3. *Suppose that Assumptions A, B, E and F hold. Consider $g_* \in \mathbb{F}$ and γ_* as in Proposition 4.2 and μ_{g_*} as in Assumption F.1. Then there exists $s \in \mathbb{B}_d(X)$ such that*

$$\gamma_* + s(x) = \inf_{a \in A(x)} \Big(c(x, a) + Qs(x, a) \Big) \tag{47}$$

and moreover

$$\inf_{a \in A(x)} \Big(c(x, a) + Qs(x, a) \Big) = c(x, g_*(x)) + Qs(x, g_*(x)) \quad \mu_{g_*}\text{-a.s.} \tag{48}$$

Proof. This proof follows the same steps as the proof of Theorem 3.3. Let $v^*(x) = \overline{\lim}_{n\to\infty} v_n(x)$. From (43) and (44) we have that for any $a \in A(x)$,

$$\gamma_n + v_n(x) \le c(x,a) + Q v_{n-1}(x,a) + Q_{f_n}(v_n - v_{n-1})(x). \tag{49}$$

From Assumption E.3, we have for any $n \in \mathbb{N}$, $-M \le v_n(x) \le d(x)$. Consequently, combining Assumption F.2 and Fatou's Lemma (see Assumption F.1), we get from (49) that

$$\gamma_* + v^*(x) \le c(x,a) + Q v^*(x,a), \tag{50}$$

and therefore from (50) we obtain that

$$\gamma_* + v^*(x) \le \inf_{a\in A(x)} \big(c(x,a) + Q v^*(x,a)\big) \le c_{g_*}(x) + Q_{g_*} v^*(x). \tag{51}$$

Define $r = v_* - v^*$. Notice that from Assumption E.3, it follows that v_* and v^* are in $\mathbb{B}_d(X)$, implying that $r \in \mathbb{B}_d(X)$. Moreover, we get from (51) and (45) that for all $x \in X$, $r(x) \ge Q_{g_*} r(x)$. As in the proof of Theorem 3.3, $\inf_{y\in X}(v_*(y) - v^*(y))$ is finite. By using Assumption F.2 and the same arguments as in Lemma 7.5.12 in [14] we have that there exists a set $\mathcal{N}_P \in \mathcal{B}(X)$ such that $\mu_{g_*}(\mathcal{N}_P) = 1$ and $v_*(x) = v^*(x) + d_P$ for all $x \in \mathcal{N}_P$, where $d_P = \inf_{y\in X}(v_*(y) - v^*(y))$. In any case we have that for all $x \in X$, $v_*(x) \ge v^*(x) + d_P$.

Define now $s_0 = v_*$ and $s_{n+1}(x) = \inf_{a\in A(x)}\big(c(x,a) + Q s_n(x,a)\big) - \gamma_*$. By repeating the same arguments as in the proof of Theorem 3.3 we get that $v^*(x) + d \le s_{n+1}(x) \le s_n(x) \le v_*(x)$. Since $\{s_n\}$ is monotone decreasing, there exists $s \in \mathbb{B}_d(X)$ such that $s_n \downarrow s$. Clearly we have that for any $a \in A(x)$, $s_{n+1}(x) \le c(x,a) + Q s_n(x,a) - \gamma_*$. Since $s_n \in \mathbb{B}_d(X)$ we have from the monotone convergence theorem, that

$$\begin{aligned} s(x) + \gamma_* &= \lim_{n\to\infty} s_{n+1}(x) + \gamma_* \le c(x,a) + \lim_{n\to\infty} \int_X s_n(y) Q(dy|x,a) \\ &= c(x,a) + \int_X \lim_{n\to\infty} s_n(y) Q(dy|x,a) = c(x,a) + Qs(x,a), \end{aligned} \tag{52}$$

showing that $s(x) + \gamma_* \le \inf_{a\in A(x)}(c(x,a) + Qs(x,a))$.

On the other hand we have that $s_{n+1}(x) \ge \inf_{a\in A(x)}\big(c(x,a) + Qs(x,a)\big) - \gamma_*$ and taking the limit as $n \to \infty$ we get that $s(x) + \gamma_* \ge \inf_{a\in A(x)}\big(c(x,a) + Qs(x,a)\big)$, which yields (47).

Since for any $x \in \mathcal{N}_P$, $s(x) = v_*(x)$ we get that for any $x \in \mathcal{N}_P$,

$$\inf_{a\in A(x)} \big(c(x,a) + Qs(x,a)\big) = s(x) + \gamma_* = v_*(x) + \gamma_* \ge c(x, g_*(x)) + Q v_*(x, g_*(x)).$$

However, $s \le v_*$ and so $\inf_{a\in A(x)}\big(c(x,a) + Qs(x,a)\big) = s(x) + \gamma_* \ge c(x, g_*(x)) + Qs(x, g_*(x))$ for any $x \in \mathcal{N}_P$, completing the proof. □

We recall next the definition of a stable policy (see Definition 5.7.7 in [13]).

Definition 4.4. A randomized stationary policy $\varphi^\infty \in \Phi$ is said to be a stable policy if there exists an invariant probability $p_\varphi \in \mathcal{P}(X)$ for $Q_\varphi(\cdot|\cdot)$ such that $V(p_\varphi, \varphi^\infty) < \infty$, so that $V(p_\varphi, \varphi^\infty) = \int_X c_\varphi(x) p_\varphi(dx)$.

Now from Assumptions A, B.1, B.2(a) and Theorem 5.7.9 in [13] it follows that there exists a stable policy φ_*^∞, with invariant probability μ_{φ_*}, satisfying

$$\inf_{\nu\in\mathcal{P}(X)} \inf_{\pi\in\Pi} V(\nu,\pi) = V(\mu_{\varphi_*}, \varphi_*^\infty). \tag{53}$$

We have the following theorem showing the convergence of the PIA to the minimum average cost criteria (see Definition 11.1.1 in [14]).

Theorem 4.5. *Suppose that Assumptions A, B.1, B.2(a), E and F hold. If $\mu_{\varphi_*}(d) < \infty$ then the PIA converges to the minimum average cost criteria, that is,*

$$\inf_{\nu\in\mathcal{P}(X)} \inf_{\pi\in\Pi} V(\nu,\pi) = \gamma_*. \tag{54}$$

Proof. From Assumption E.3, we have that $\mu_{g_*}(|s|) \le \mu_{g_*}(d) < \infty$. Combining Eqs. (47) and (48), we get

$$\gamma_* + s(x) = c(x, g_*(x)) + Qs(x, g_*(x)) \quad \mu_{g_*}\text{-a.s.}$$

Integrating both sides of the previous equation with respect to μ_{g_*} gives

$$\int_X c(x, g_*(x)) \mu_{g_*}(dx) = \gamma_*,$$

implying that

$$V(\mu_{g_*}, g_*^\infty) = \gamma_*. \tag{55}$$

Now, we have from Eq. (47) that

$$\gamma_* + s(x) \le c(x, \varphi_*(x)) + Qs(x, \varphi_*(x)),$$

for any $x \in X$. Since $\mu_{\varphi_*}(d) < \infty$, we have that $\mu_{\varphi_*}(|s|) < \infty$ and so

$$\gamma_* \le \int_X c(x, \varphi_*(x))\mu_{\varphi_*}(dx) = V(\mu_{\varphi_*}, \varphi_*^\infty) = \inf_{\nu\in\mathcal{P}(X)} \inf_{\pi\in\Pi} V(\nu, \pi). \tag{56}$$

Finally, combining Eqs. (55) and (56), the result follows. □

Acknowledgments

The first author received financial support from CNPq (Brazilian National Research Council), grant 301067/09-0. The second author was supported by ARPEGE program of the French National Agency of Research (ANR), project "FAUTOCOES", number ANR-09-SEGI-004.

References

[1] Aristotle Arapostathis, Vivek S. Borkar, Emmanuel Fernández-Gaucherand, Mrinal K. Ghosh, Steven I. Marcus, Discrete-time controlled Markov processes with average cost criterion: a survey, SIAM J. Control Optim. 31 (2) (1993) 282–344.
[2] X. Guo, Q. Zhu, Average optimality for Markov decision processes in Borel spaces: a new condition and approach, J. Appl. Probab. 43 (2006) 318–334.
[3] Onésimo Hernández-Lerma, Existence of average optimal policies in Markov control processes with strictly unbounded costs, Kybernetika (Prague) 29 (1) (1993) 1–17.
[4] Onésimo Hernández-Lerma, Jean B. Lasserre, Average cost optimal policies for Markov control processes with Borel state space and unbounded costs, Systems Control Lett. 15 (4) (1990) 349–356.
[5] Onésimo Hernández-Lerma, Raúl Montes-de Oca, Rolando Cavazos-Cadena, Recurrence conditions for Markov decision processes with Borel state space: a survey, Ann. Oper. Res. 28 (1–4) (1991) 29–46.
[6] A. Jaśkiewicz, A.S. Nowak, On the optimality equation for average cost Markov control processes with Feller transition probabilities, J. Math. Anal. Appl. 316 (2006) 495–509.
[7] S.P. Meyn, The policy iteration algorithm for average reward Markov decision processes with general state space, IEEE Trans. Automat. Control 42 (12) (1997) 1663–1680.
[8] Raúl Montes-de Oca, Onésimo Hernández-Lerma, Conditions for average optimality in Markov control processes with unbounded costs and controls, J. Math. Systems Estim. Control 4 (1) (1994) 19 (electronic).
[9] Raúl Montes-de Oca, The average cost optimality equation for Markov control processes on Borel spaces, Systems Control Lett. 22 (5) (1994) 351–357.
[10] Eitan Altman, Constrained Markov Decision Processes, in: Stochastic Modeling, Chapman & Hall, CRC, Boca Raton, FL, 1999.
[11] D.P. Bertsekas, S.E. Shreve, Stochastic Optimal Control: The Discrete Time Case, in: Mathematics in Science and Engineering, vol. 139, Academic Press Inc., New York, 1978.
[12] Jerzy Filar, Koos Vrieze, Competitive Markov Decision Processes, Springer-Verlag, New York, 1997.
[13] O. Hernández-Lerma, J.B. Lasserre, Discrete-Time Markov Control Processes: Basic Optimality Criteria, in: Applications of Mathematics, vol. 30, Springer-Verlag, New York, 1996.
[14] O. Hernández-Lerma, J.B. Lasserre, Further Topics On Discrete-Time Markov Control Processes, in: Applications of Mathematics, vol. 42, Springer-Verlag, New York, 1999.
[15] Martin L. Puterman, Markov Decision Processes: Discrete Stochastic Dynamic Programming, in: Wiley Series in Probability and Mathematical Statistics: Applied Probability and Statistics, John Wiley & Sons Inc., New York, 1994, A Wiley-Interscience Publication.
[16] Linn I. Sennott, Stochastic Dynamic Programming and The Control of Queueing Systems, in: Wiley Series in Probability and Statistics: Applied Probability and Statistics, John Wiley & Sons Inc., New York, 1999, A Wiley-Interscience Publication.
[17] Onésimo Hernández-Lerma, Jean B. Lasserre, Policy iteration for average cost Markov control processes on Borel spaces, Acta Appl. Math. 47 (2) (1997) 125–154.
[18] Charalambos D. Aliprantis, Kim C. Border, Infinite Dimensional Analysis: A Hitchhiker's Guide, third ed., Springer, Berlin, 2006.
[19] M. Schäl, Averge optimality in dynamic programming with general state space, Math. Oper. Res. 18 (1993) 163–172.
[20] R. Cavazos-Cadena, F. Salem-Silva, The discounted method and equivalence of average criteria for risk-sensitive Markov decision processes on Borel spaces, Appl. Math. Optim. 61 (2010) 167–190.
[21] A. Jaśkiewicz, A.S. Nowak, Zero-sum ergodic stochastic games with feller transition probabilities, SIAM J. Control Optim. 45 (2006) 773–789.
[22] R. Serfozo, Convergence of Lebesgue integrals with varying measures, Sankhyã, Ser. A 44 (1982) 380–402.
[23] O. Hernández-Lerma, M. Muñoz de Ozak, Discrete-time Markov control processes with discounted costs: optimality criteria, Kybernetika 28 (1992) 191–212.
[24] Onésimo Hernández-Lerma, Jean Bernard Lasserre, Markov Chains and Invariant Probabilities, in: Progress in Mathematics, vol. 211, Birkhäuser Verlag, Basel, 2003.
[25] Achim Klenke, Probability Theory. A comprehensive Course, in: Universitext, Springer, London, 2008.
[26] E. Berge, Topological Spaces, Macmillan, New York, 1963.
[27] E. Klein, A.C. Thompson, Theory of Correspondences, Wiley, New York, 1984.