KYBERNETIKA — VOLUME 41 (2005), NUMBER 6, PAGES 757 - 772

ESTIMATES FOR PERTURBATIONS OF AVERAGE MARKOV DECISION PROCESSES WITH A MINIMAL STATE AND UPPER BOUNDED BY STOCHASTICALLY ORDERED MARKOV CHAINS

RAÚL MONTES-DE-OCA AND FRANCISCO SALEM-SILVA*

This paper deals with Markov decision processes (MDPs) with real state space for which its minimum is attained, and that are upper bounded by (uncontrolled) stochastically ordered (SO) Markov chains. We consider MDPs with (possibly) unbounded costs, and to evaluate the quality of each policy, we use the objective function known as the *average cost*. For this objective function we consider two Markov control models $\mathbb{P}$ and $\mathbb{P}_1$. $\mathbb{P}$ and $\mathbb{P}_1$ have the same components except for the transition laws. The transition q of $\mathbb{P}$ is taken as unknown, and the transition q_1 of $\mathbb{P}_1$, as a known approximation of q. Under certain irreducibility, recurrence and ergodic conditions imposed on the bounding SO Markov chain (these conditions give the rate of convergence of the transition probability in t-steps, $t = 1, 2, \ldots$ to the invariant measure), the difference between the optimal cost to drive $\mathbb{P}$ and the cost obtained to drive $\mathbb{P}$ using the optimal policy of $\mathbb{P}_1$ is estimated. That difference is defined as *the index of perturbations*, and in this work upper bounds of it are provided. An example to illustrate the theory developed here is added.

Keywords: stochastically ordered Markov chains, Lyapunov condition, invariant probability, average Markov decision processes

AMS Subject Classification: 90C40, 93E20

1. INTRODUCTION

This paper concerns with Markov Decision Processes (MDPs) with real state space for which its minimum is attained, and that are upper bounded by (uncontrolled) stochastically ordered Markov Chains. The MDPs, considered (possibly) have an unbounded one-step cost function. The quality of each policy will be evaluated by the objective function (or the performance index) known as the average cost. Denote it by $J(\Pi, x)$, where Π is the policy that drives the system, and x is the initial state. Now consider the following:

There are two Markov control models (see [7] and [8]): $\mathbb{P}$ and $\mathbb{P}_1$, and we suppose that they have the same state and action spaces, and the same one-step cost function,

*Francisco Salem-Silva supported by grant VIEP-BUAP II 33.

but different transition probability laws. They are denoted by q and q_1, respectively. It is supposed that q_1 is known but near, in the sense of the total variation metric, to the "unknown" transition probability law q. Since q is unknown, q_1 will be used as an approximation of q to find an optimal control for $\mathbb{P}_1$ and then it will be used to control $\mathbb{P}$. Hence, assuming the existence of stationary optimal policies f^* and f_1^* for $\mathbb{P}$ and $\mathbb{P}_1$, respectively, the additional cost can be evaluated when using f_1^* to control $\mathbb{P}$, instead of f^*, by means of the so-called *Index of Perturbations* (see, for instance, [1, 2, 3, 5, 6] and [14]). For the average case, this Index is defined as:

$$\widehat{\Delta}(x) := J(f_1^*, x) - J(f^*, x), \tag{1.1}$$

where x is the initial state.

The main goal of the present paper is to find a measure for the perturbation of the MDP generated for $\mathbb{P}$. That means an inequality with the following structure is wanted to be found for an upper bound for Index (1.1):

$$\widehat{\Delta}(x) \leq M\Gamma_x(\|q - q_1\|) \tag{1.2}$$

where M is a constant, and $\Gamma_x(\cdot)$ is a function such that $\Gamma_x(y) \to 0$ if $y \to 0$, and x is the initial state.

For the discounted case, i. e. when the objective function is the total discounted expected cost $V(\Pi, x)$, where Π is the policy that drives the system, and x is the initial state; upper bounds have been obtained as in (1.2), for the corresponding Index of Perturbations i. e.,

$$\Delta_1(x) := V(f_1^*, x) - V(f^*, x), \tag{1.3}$$

where f^* and f_1^* are optimal policies for $\mathbb{P}$ and $\mathbb{P}_1$, respectively (see, e. g. [2, 3, 5] and [6]). (Also look at [1] for the case of total cost with finite horizon.)

In this paper, for MDPs with real state space for which its minimum is attained; the Zolotarev's Method will be used (see [17]) to reduce the problem to one of the perturbation of uncontrolled processes. Here, the rate of convergence provided by Stochastically Ordered (SO) Markov chains that satisfy certain irreducibility, recurrence and ergodic conditions will be applied (see [12]). With this new rate of convergence, it will be discovered that the term $\Gamma_x(\|q - q_1\|)$ can be calculated explicitly in a simple way and more precise bounds are expected.

In order to use the rate of convergence of the SO it will be supposed that the MDPs are upper bounded for SO Markov chains.

The paper is organized as follows. Firstly, in Section 2 we present the basics on stochastically ordered Markov chains including the main assumption (Assumption 2.1) that assures the rate of convergence. Secondly, in Section 3 we give the preliminaries about average MDPs. Section 4 provides the main result of the paper (Theorem 4.1). Sections 5 and 6 complete the proof of Theorem 4.1. Finally, in the last section an example is presented.

2. STOCHASTICALLY ORDERED MARKOV CHAINS

Notation and terminology

Let $Y = \{y_t\}$ be a homogeneous Markov chain with values in the state space X with discrete time $t = 0, 1, 2 \ldots$, and with transition kernel $p(B|x)$, $B \in \mathbb{B}(X)$, $x \in X$ where $\mathbb{B}(X)$ denotes the sigma-algebra of Borel of X.

Let P_x and E_x be respectively the probability law and the expectation of the chain under the initial condition $y_0 = x \in X$.

The transition probabilities in t-steps of the chain are denoted by $p^t(B|x)$, $x \in X$ $B \in \mathbb{B}(X)$, i.e. $p^t(B|x) = P_x[y_t \in B]$, $t = 0, 1, 2, \ldots$

For $x \in X$, $B, D \in \mathbb{B}(X)$, and, $t = 0, 1, 2, \ldots$, it is written:

$$_B p^t(D|x) := P_x[y_t \in D \text{ and } y_j \notin B \text{ for } 1 \leq j \leq t-1]. \tag{2.1}$$

Let $\mathbb{M} := \{\mu | \mu \text{ be a probability on } \mathbb{B}(X)\}$, and let $\mathbb{B}_M := \{g : X \to \mathbb{R} : g \text{ is measurable and bounded}\}$.

Denote by $\|\cdot\|$ the total variation metric defined on $\mathbb{M}$, i.e. for μ_1, $\mu_2 \in \mathbb{M}$,

$$\|\mu_1 - \mu_2\| := 2 \sup_{D \in \mathbb{B}(X)} \{|\mu_1(D) - \mu_2(D)|\}, \tag{2.2}$$

or equivalently,

$$\|\mu_1 - \mu_2\| := \sup \left\{ \left| \int g \mathrm{d}\mu_1 - \int g \mathrm{d}\mu_2 \right| : g \in \mathbb{B}_M \text{ and } |g| \leq 1 \right\}. \tag{2.3}$$

Remark 2.1. For random elements χ and κ taking values in X, we write

$$\|\chi - \kappa\| \equiv \|\mu_\chi - \mu_\kappa\|,$$

where μ_χ and μ_κ are the distributions of χ and κ, respectively.

$\mu \in \mathbb{M}$ is supposed to be *invariant* (with respect to the Markov chain $Y = \{y_t\}$) if it has the property that

$$\mu(D) = \int_D p(\mathrm{d}y|x)\mu(\mathrm{d}y), \tag{2.4}$$

where $D \in \mathbb{B}(X)$, $x \in X$.

The Markov chain $Y = \{y_t\}$ is said to be *Harris-recurrent* if there exists a non-trivial σ-finite measure γ such that

$$P_x[y_t \in B \text{ for some } t] = 1, \tag{2.5}$$

for all $x \in X$ whenever $B \in \mathbb{B}(X)$, satisfies $\gamma(B) > 0$.

It is said that a Harris-recurrent Markov chain $Y = \{y_t\}$ is *positive* if it has an invariant probability measure m_Y, i.e. m_Y is a probability measure and satisfies (2.4).

Stochastically ordered Markov chains

This paper specifically deals with a Markov chain $Y = \{y_t\}$ having state space X of the form: $[d, \infty), d \in \mathbb{R}$, or more concretely of the form $X = [0, \infty)$, for simplicity.

Remark 2.2. The important clue is that the state space has a minimal element (see, [12]). Due to this fact let us obtain explicit bounds on the rate of convergence to the invariant measure.

Let $x, y \in X$. Consider the canonical probability spaces (Ω, F, P_x) and (Ω, F, P_y) induced by the kernel p and the initial distributions δ_x and δ_y (here δ_x and δ_y denote the probabilities concentrated in x, and y, respectively), in which it is possible to define two copies of the chain $Y_1 = \{y_t^1\}$, $Y_2 = \{y_t^2\}$, respectively, whenever $y_0^1 = x$ and $y_0^2 = y$. Hence, taking the product space $(\Omega \times \Omega, F \times F, P_x \times P_y) = (\Omega^*, F^*, P^*)$ the chains Y^1 and Y^2 can be described jointly (see [10]).

Let W and Z be nonnegative random variables defined on the probability space (Ω', F', P'). W is considered to be *stochastically larger than* Z if $P'[W \leq x] \leq P'[Z \leq x]$ for all $x \in \mathbb{R}$.

The chain $Y = \{y_t\}$ is *stochastically ordered* (or stochastically ordered in its initial state) if for two copies of the chain $Y^1 = \{y_t^1\}$, $Y^2 = \{y_t^2\}$, whenever $y_0^1 = x$ and $y_0^2 = y$ and $y < x$, then y_t^1 is stochastically larger than y_t^2 for all $t \geq 1$, i.e. $P^*[y_t^1 \leq z] \leq P^*[y_t^2 \leq z]$ for all $z \in \mathbb{R}$ and $t \geq 1$, where $P^* := P_x \times P_y$.

Besides it is supposed that the chain $Y = \{y_t\}$ is *pathwise ordered* if for two copies of the chain $Y^1 = \{y_t^1\}$, $Y^2 = \{y_t^2\}$, whenever $y_0^1 = x$ and $y_0^2 = y$ and $y < x$, then $y_t^2(\omega) \leq y_t^1(\omega)$ for all $\omega \in \Omega^* = \Omega \times \Omega$.

Remark 2.3. As it was mentioned in [12] (see also, [10]), if the chain is stochastically ordered but not pathwise ordered, then it is possible to change the underlying probability space and construct a new chain that is pathwise ordered and distributionally equivalent to original chain. Hence, it is possible to assume that a ordered chain is pathwise ordered.

Assumption 2.1. Let $Y = \{y_t\}$ be a Markov chain. Suppose that

(a) For each $x \in X$, there exists a positive integer t^* such that ${}_{\{0\}}p^{t^*}([x, \infty)|0) > 0$;

(b) Y is stochastically ordered;

(c) Let $\tau_0 = \inf\{t > 0 : y_t = 0\}$. For each $x \in X$ we assume that $E_x(\tau_0) < \infty$;

(d) There exists $L : [0, \infty) \to [1, \infty)$ with $L(0) = 1$ and constants λ and b with $0 < \lambda < 1$, $0 \leq b < \infty$, such that,

$$\int L(y)p(\mathrm{d}y|x) \leq \lambda L(x) + bI_{\{0\}}(x), \tag{2.6}$$

where $x \in X$ and $I_{\{0\}}$ denotes the indicator function of the set $\{0\}$.

Remark 2.4. The conditions in Assumption 2.1 are the same that appear in [12], except for Assumption 2.1(c). In [12] is assumed the existence of the invariant measure. Notice that latter is implied by Assumption 2.1(c).

Now, the result that provides the rate of convergence of the transition probabilities $p^t(\cdot|x)$, $x \in X$ to the invariant measure m_Y (when there are) in the sense of the total variation metric (see Assumption 2.1(c)), is presented without proof. This proof can be found in Lund and Tweedie [12].

Lemma 2.1. Let $Y = \{y_t\}$ be a Markov chain. Suppose that Assumption 2.1 holds. Then, for each $t = 0, 1, 2 \ldots$,

$$\|p^t(\cdot|x) - m_Y(\cdot)\| \leq r^{-t} h_x(r), \tag{2.7}$$

for all $r < \lambda^{-1}$ and $x \in X$, where $h_x(r) = E_\nu r^{\tau_0}$, $\nu = \max\{Z, x\}$, and Z is a random variable with distribution m_Y, and

$$h_x(r) \leq E_x[r^{\tau_0}] + b/(1-\lambda) < \infty, \tag{2.8}$$

$\tau_0 = \inf\{t > 0 : y_t = 0\}$, and b and λ are the constants in Assumption 2.1.

Remark 2.5. a) In Section 7 an example that satisfies the assumptions of Lemma 2.1, is presented.

b) Notice that from (2.7) and (2.8), for each $x \in X$,

$$\left\|p^t(\cdot|x) - m_Y(\cdot)\right\| \to 0, \quad t \to \infty.$$

3. AVERAGE MARKOV DECISION PROCESSES

In this section a special kind of MDPs, is shown; i. e. MDPs satisfying the condition that both the state of space and the action space are subsets of $\mathbb{R}$, are dealt with.

Specifically, let $\mathbb{P} = (X, A, \{A(x) : x \in X\}, q, c)$ be a standard Markov control model (see, [7]) which consists of the state space X, the action space A. Both X and A are assumed to be measurable subsets of $\mathbb{R}$ endowed with the usual metric, and in fact, it is supposed that $X = [0, \infty)$ (see Remark 2.1). The sets $A(x)$, $x \in X$ are nonempty measurable subsets of A, and represent the constrained action sets. Let $\mathbb{K} = \{(x, a) : x \in X, a \in A(x)\}$, which is considered to be measurable in the product $X \times A$. The transition law q is a stochastic kernel of X given $\mathbb{K}$ (i. e. $q(\cdot|x, a)$ is a probability measure on X, for each $(x, a) \in \mathbb{K}$, and $q(B|\cdot)$ is a measurable function on $\mathbb{K}$, for each measurable set $B \subset X$), and the one-step cost $c : \mathbb{K} \to \mathbb{R}$ is a measurable function.

A *policy* is defined as a sequence $\Pi = \{\pi_t\}$ satisfying that, for each $t = 0, 1, 2, \ldots$, π_t is a stochastic kernel of A given H_t, where H_t denotes the set of all admissible histories $h_t = (x_0, a_0, x_1, a_1, \ldots, x_{t-1}, a_{t-1}, x_t)$, with $(x_i, a_i) \in \mathbb{K}$, $i = 0, 1, \ldots, t-1$, $x \in X$, and π_t concentrated on $A(x_t)$.

Let Δ be the set of all policies, and let $\mathbb{F}$ denote the set of all measurable functions $f : X \to A$ such that $f(x) \in A(x)$, for all $x \in X$.

A policy $\Pi = \{\pi_t\}$ is called *stationary* if there exists $f \in \mathbb{F}$ such that for each $t = 0, 1, \ldots, \pi_t$ is concentrated on $f(x)$ if $x_t = x$. In this case, we identify Π with f, and the set of all stationary policies with $\mathbb{F}$.

Remark 3.1. It is well-known (see, [7] and [8]) that a MDP in which a stationary policy $g \in \mathbb{F}$ is used to drive the system gives that the sequence of states $\{x_t\}$ is a homogeneous Markov chain with stationary transition kernel given by $p(\cdot|x) := q(\cdot|x, g(x))$, $x \in X$. This is the connection with the previous section. On the other hand, taking in account a stationary policy $g \in \mathbb{F}$, the corresponding state process is denoted by $\{x_t^g\}$.

Given a policy $\Pi \in \Delta$ and $x \in X$, P_x^Π stays for the probability measure induced in canonical way by the model $\mathbb{P}$ (see, [9] for the construction of P_x^Π), and E_x^Π stays for the expectation corresponding to P_x^Π.

Let $\mathbb{P} = (X, A, \{A(x) : x \in X\}, q, c)$ be a Markov control model.

The *long-run* expected average cost (AC) while using a policy Π, given the initial state $x_0 = x$, is defined as:

$$J(\Pi, x) := \limsup_{n \to \infty} \frac{E_x \left[\sum_{t=0}^{n-1} c(x_t, a_t)\right]}{n}. \tag{3.1}$$

A policy $\Pi^* \in \Delta$ is *AC-optimal* if,

$$J(\Pi^*, x) = \inf_{\Pi \in \Delta} J(\Pi, x), \quad x \in X, \tag{3.2}$$

and the *optimal* AC-function is designated as:

$$J^*(x) := \inf_{\Pi \in \Delta} J(\Pi, x), \quad x \in X. \tag{3.3}$$

The following Assumption is supposed to be valid throughout the paper:

Assumption 3.1.

(a) The existence of a stationary policy $\widehat{f}$ which is AC-optimal is assumed.

(b) It is also supposed that for every stationary policy $f \in \mathbb{F}$, the average cost $J(f, \cdot)$ is a constant $J(f)$ given by

$$J(f, x) = J(f) = \int c(y, f(y)) m_f(\mathrm{d}y), \tag{3.4}$$

where m_f is invariant probability corresponding to the stochastic kernel induced by f and $x \in X$.

Remark 3.2.

(a) For sufficient conditions for Assumption 3.1(a), see Assumptions 2.1, 2.2 and 2.3 in [4].

(b) Sufficient conditions for Assumption 3.1(b) are the following: for bounded cost, see Section (3.3) in [7], and for unbounded cost, see Assumptions 2.1, 2.2 and 2.3 in [4].

4. BOUNDS FOR THE INDEX OF PERTURBATIONS

Let $\mathbb{P} = (X, A, \{A(x) : x \in X\}, q, c)$ and $\mathbb{P}_1 = (X, A, \{A(x) : x \in X\}, q_1, c)$ be two average Markov control models. Both of them satisfy the definitions and the Assumption 3.1 of the previous section.

Remark 4.1.

(i) Notice that $\mathbb{P}$ and $\mathbb{P}_1$ defer only in the transition probability, but q is supposed to be unknown and q_1 is an approximation known of q.

(ii) Let $\mathbb{F}$ and $\mathbb{F}_1$ be the corresponding sets of stationary policies for the models $\mathbb{P}$ and $\mathbb{P}_1$, respectively. Observe that $\mathbb{F} = \mathbb{F}_1$, since $\mathbb{P}$ and $\mathbb{P}_1$ have the same state and action spaces.

(iii) Given an initial state x and stationary policies f and g, there exist canonical spaces (Ω', F', P_x^f) and (Ω', F', P_x^g) to describe, in particular, the processes $\{x_t^f\}$ and $\{x_t^g\}$, respectively (see Section 3). Notice that they have the same measurable space (Ω', F') (see [9]).

Assumption 4.1. There is a stationary policy g for which the following points hold, considering the Markov chain $\{x^g\}$ (see, Remark 3.1):

(a) The Assumption 2.1, for some constants λ and b with $0 < \lambda < 1$, $0 \leq b < \infty$, and function $L : [0, \infty) \to [1, \infty)$, it is also supposed that the function L is increasing;

(b) Let $x \in X$. If $x_0^f = x_0^g = x$, then $x_t^f(\omega) \leq x_t^g(\omega)$, for all $f \in \mathbb{F}$, $\omega \in \Omega'$ and $t = 1, 2, \ldots$. (Here Ω' is the set defined in the canonical space – see Remark 4.1(iii).)

Moreover, we assume:

(c) There exists a constant $s \geq 1$ such that

$$\sup_{a \in A(x)} |c(x, a)| \leq [L(x)]^{\frac{1}{s}}, \quad x \in X. \tag{4.1}$$

Assumption 4.2. For the model $\mathbb{P}_1$ there exists a stationary policy g_1 such that Assumption 4.1 holds for the transition kernel q_1 with the same λ, b, s and L.

Remark 4.2.

(a) Notice that as L is increasing, then for both models $\mathbb{P}$ and $\mathbb{P}_1$ it can be applied that

$$\int L(y)q(\mathrm{d}y|x, f(x)) \leq \lambda L(x) + bI_{\{0\}}(x), \tag{4.2}$$

and

$$\int L(y)q_1(\mathrm{d}y|x, f_1(x)) \leq \lambda L(x) + bI_{\{0\}}(x), \tag{4.3}$$

where λ, b are the constants in Assumptions 4.1 and 4.2, $f, f_1 \in \mathbb{F}$, and $x \in X$.

(b) Under Assumption 3.1(a) there exist f^* and f_1^* stationary average optimal policies for $\mathbb{P}$ and $\mathbb{P}_1$, respectively.

Remember that the Index of Perturbations has already been defined as:

$$\widehat{\Delta}(x) := J(f_1^*, x) - J(f^*, x), x \in X. \tag{4.4}$$

Notice that $\widehat{\Delta}(x) \geq 0$, for all $x \in X$.

Theorem 4.1. Consider the models $\mathbb{P}$ and $\mathbb{P}_1$. Suppose that Assumption 3.1 holds for both of these models, and let f^*, $f_1^* \in \mathbb{F}$ be average optimal policies for $\mathbb{P}$ and $\mathbb{P}_1$, respectively. Also, suppose that Assumption 4.1 and 4.2 hold. Then

$$\widehat{\Delta}(x) \leq 2\left(\frac{2b}{1-\lambda} + 2rh_x(r) + 1\right)\delta^{\frac{s-1}{s}} \max\left\{1, \log_\rho \delta\right\}, \tag{4.5}$$

where $\delta = \sup_{x\in X} \sup_{a\in A(x)} \|q_1(\cdot|x,a) - q(\cdot|x,a)\|$ and $\rho = \frac{1}{r}$.

Remark 4.3. If the models $\mathbb{P}$ and $\mathbb{P}_1$ are obtained by the recurrent equations

$$x_{t+1} = F(x_t, a_t, \xi_t), \tag{4.6}$$

and

$$\widetilde{x}_{t+1} = F(\widetilde{x}_t, a_t, \widetilde{\xi}_t), \tag{4.7}$$

$t = 0, 1, 2, \ldots$, respectively, it can be proved (see [6]) that

$$\delta^{\frac{s-1}{s}} \max\left\{1, \log_\rho \delta\right\} \leq \left\|\mu_\xi - \mu_{\widetilde{\xi}}\right\|^{\frac{s-1}{s}}, \tag{4.8}$$

provided that $\|\mu_\xi - \mu_{\widetilde{\xi}}\| \leq e^{\frac{-s}{s-1}}$, where μ_ξ and $\mu_{\widetilde{\xi}}$ are the distributions of ξ and $\widetilde{\xi}$, respectively.

5. TECHNICAL PRELIMINARIES

Lemma 5.1. Under Assumption 4.1, for each $f \in \mathbb{F}$, there exists an invariant (actually the limit, in the sense of the total variation metric – see Remark 2.4) probability m_f corresponding to the kernel q.

Proof. Fix $f \in \mathbb{F}$ and let g be the distinguished policy in Assumption 4.1. Denote by τ^f and τ^g the time of the first return of $\{x_t^f\}$ and $\{x_t^g\}$ to $x_0 = 0$, given $x_0^f = x_0^g = 0$, respectively. By Assumption 4.1b), for $t = 1, 2, \ldots$, we get

$$E[\tau^f] \leq E[\tau^g] < \infty.$$

Therefore, by Corollary 5.3 of [15], $\{x_t^f\}$ is positive Harris-recurrent. The existence of m_f follows. Now from Theorem 4.1 in [12], it is clear that m_f is the limit of $\{x_t^f\}$, in the sense of the total variation metric. □

Let $\vartheta > 0$ be a fixed number. Define $c_\vartheta(x,a) = c(x,a)$ if $c(x,a) \leq \vartheta$ and $c_\vartheta(x,a) = 0$ if $c(x,a) > \vartheta$.

Lemma 5.2. Under Assumptions 3.1 and 4.1, for every stationary policy $f \in \mathbb{F}$,

$$\left| \int c(y, f(y)) m_f(\mathrm{d}y) - \int c_\vartheta(y, f(y)) m_f(\mathrm{d}y) \right| \leq \left[\int L(y) m_f(\mathrm{d}y) \right] \vartheta^{1-s}, \qquad (5.1)$$

where m_f is the invariant probability corresponding to the stochastic kernel induced by f; b and s are the constants in Assumption 4.1, and $\vartheta > 0$.

Proof. First, the definition of c_ϑ, (4.1), and $\{c(y, f(y) > \vartheta\} \subseteq \{L(y) > \vartheta^s\}$ yield:

$$\begin{aligned} & \left| \int c(y, f(y)) m_f(\mathrm{d}y) - \int c_\vartheta(y, f(y)) m_f(\mathrm{d}y) \right| \\ \leq & \int c(y, f(y)) I_{\{c(y,f(y)) > \vartheta\}}(y) m_f(\mathrm{d}y) \\ \leq & \int [L(y)]^{\frac{1}{s}} I_{\{L(y) > \vartheta^s\}}(y) m_f(\mathrm{d}y). \end{aligned} \qquad (5.2)$$

Now, using the Hölder and Chebyshev inequalities, where $1/\ell = 1 - 1/s$, it follows that:

$$\begin{aligned} \int [L(y)]^{\frac{1}{s}} I_{\{L(y) > \vartheta^s\}}(y) m_f(\mathrm{d}y) & \leq \left[\int [L(y) m_f(\mathrm{d}y) \right]^{\frac{1}{s}} [P(L(y) > \vartheta^s)]^{\frac{1}{\ell}} \\ & \leq \left[\int [L(y) m_f(\mathrm{d}y) \right]^{\frac{1}{s}} \left[\int [L(y) m_f(\mathrm{d}y)]^{\frac{1}{\ell}} \vartheta^{\frac{-s}{\ell}} \right] \\ & = \int [L(y) m_f(\mathrm{d}y)] \quad \vartheta^{1-s} \end{aligned}$$

hence (5.1) is obtained from (5.2). □

Remark 5.1. Notice that Lemmas 5.1 and 5.2 also hold under Assumption 4.2 for $\mathbb{P}_1$.

Lemma 5.3. Suppose that Assumptions 3.1, 4.1 and 4.2 hold. Consider for $f \in \mathbb{F}$, the processes $\{x_t^f\}$ and $\{\widetilde{x}_t^f\}$ which correspond to the models $\mathbb{P}$ and $\mathbb{P}_1$, respectively. Then, for each $t = 0, 1, \ldots, x \in X$, we get

$$\left\|\widetilde{x}_{t+1}^f - x_{t+1}^f\right\| \leq \left\|\widetilde{x}_t^f - x_t^f\right\| + \sup_{x\in X}\sup_{a\in A(x)} \|q_1(\cdot|x,a) - q(\cdot|x,a)\|\,. \tag{5.3}$$

Proof. Let $\mathbb{H} = \{h \in \mathbb{B}_M : |h| \leq 1\}$ and applying the Chapman–Kolmogorov equation we get for $x \in X$:

$$\begin{aligned} I_{t+1} &\doteq \left\|\widetilde{x}_{t+1}^f - x_{t+1}^f\right\| = \sup_{h\in\mathbb{H}}\left|\int h(y)\left\{q_1^{t+1}(\mathrm{d}y|x,f(x)) - q^{t+1}(\mathrm{d}y|x,f(x))\right\}\right| \\ &= \sup_{h\in\mathbb{H}}\left|\int h(y)\int q_1^t(\mathrm{d}z|x,f(x))\,q_1(\mathrm{d}y|z,f(z))\right. \\ &\quad \left. -\int h(y)\int q^t(\mathrm{d}z|x,f(x))\,q(\mathrm{d}y|z,f(z))\right|. \end{aligned} \tag{5.4}$$

Now, applying Fubini

$$\begin{aligned} I_{t+1} &= \sup_{h\in\mathbb{H}}\left|\int q_1^t(\mathrm{d}z|x,f(x))\int h(y)q_1(\mathrm{d}y|z,f(z))\right. \\ &\quad \left. -\int q^t(\mathrm{d}z|x,f(x))\int h(y)q(\mathrm{d}y|z,f(z))\right| \\ &\leq \sup_{h\in\mathbb{H}}\left|\int q_1^t(\mathrm{d}z|x,f(x))\int h(y)q_1(\mathrm{d}y|z,f(z))\right. \\ &\quad -\int q^t(\mathrm{d}z|x,f(x))\int h(y)q(\mathrm{d}y|z,f(z)) \\ &\quad +\int q_1^t(\mathrm{d}z|x,f(x))\int h(y)q(\mathrm{d}y|z,f(z)) \\ &\quad \left. -\int q_1^t(\mathrm{d}z|x,f(x))\int h(y)q(\mathrm{d}y|z,f(z))\right| \\ &\leq \sup_{h\in\mathbb{H}}\left|\int q_1^t(\mathrm{d}z|x,f(x))\int h(y)q_1(\mathrm{d}y|z,f(z))\right. \\ &\quad \left. -\int q_1^t(\mathrm{d}z|x,f(x))\int h(y)q(\mathrm{d}y|z,f(z))\right| \\ &\quad +\sup_{h\in\mathbb{H}}\left|\int q^t(\mathrm{d}z|x,f(x))\int h(y)q(\mathrm{d}y|z,f(z))\right. \\ &\quad \left. -\int q_1^t(\mathrm{d}z|x,f(x))\int h(y)q(\mathrm{d}y|z,f(z))\right| \end{aligned}$$

$$\leq \sup_{h\in\mathbb{H}}\sup_{z\in X}\left|\int h(y)q_1(\mathrm{d}y|z,f(z)) - \int h(y)q(\mathrm{d}y|z,f(z))\right| \tag{5.5}$$
$$+\sup_{h\in\mathbb{H}}\left|\int q^t(\mathrm{d}z|x,f(x))\,\widetilde{h}(z) - \int q_1^t(\mathrm{d}z|x,f(x))\,\widetilde{h}(z)\right|$$

where $\widetilde{h}(z) = \int h(y)q(\mathrm{d}z|x,f(z)) \in \mathbb{H}$ since,

$$\left|\widetilde{h}(z)\right| \leq \int |h(y)|\,q(\mathrm{d}z|x,f(z)) \leq \int q(\mathrm{d}y|z,f(x)) = 1.$$

Then we can observe that the last member on the right side of (5.5) is less than

$$\sup_{\widetilde{h}\in\mathbb{H}}\left|\int q^t(\mathrm{d}z|x,f(x))\,\widetilde{h}(z) - \int q_1^t(\mathrm{d}z|x,f(x))\,\widetilde{h}(z)\right| = \left\|\widetilde{x}_t^f - x_t^f\right\|.$$

Also interchanging sups in the first right of (5.5) we obtain

$$\begin{aligned}
&\sup_{h\in\mathbb{H}}\sup_{z\in X}\left|\int h(y)q_1(\mathrm{d}y|z,f(z)) - \int h(y)q(\mathrm{d}y|z,f(z))\right| \\
&= \sup_{z\in X}\|q_1(\cdot|z,f(z)) - q(\cdot|z,f(z))\| \\
&\leq \sup_{z\in X}\sup_{a\in A(z)}\|q_1(\cdot|z,a) - q(\cdot|z,a)\| \qquad (5.6)
\end{aligned}$$

since $f(z) \in A(z)$ for each $f \in \mathbb{F}$. Hence combining (5.5) and (5.6) we get (5.3). □

Lemma 5.4. Let $Y = \{y_t\}$ be a Markov chain with state space $[0,\infty)$. Let $\tau_0 = \inf\{t > 0 : y_t = 0\}$ and denote by $N_x(r) = E_x(r^{\tau_0})$, $r \in \mathbb{R}$. Assume that Y is pathwise ordered and that $N_0(r) < \infty$ for some $r > 1$. Then the function L defined by $L(0) = 1$ and $L(x) = N_x(r)$ for $x > 0$, and the constants $\lambda = r^{-1}$ and $b = r^{-1}(L_0 - 1)$ satisfy (2.6) the equality.

Proof. This is Theorem 5.1 in [12]. □

6. PROOF OF THE THEOREM 1

Let $x \in X$ and consider $\vartheta > 0$.

Then

$$\begin{aligned}
\widehat{\Delta}(x) &= |J(f_1^*,x) - J(f^*,x)| \\
&\leq |J(f_1^*,x) - J_1(f_1^*,x)| + \left|\inf_{f\in\mathbb{F}} J_1(f) - \inf_{f\in\mathbb{F}} J(f)\right| \\
&\leq 2\sup_{f\in\mathbb{F}}|J(f) - J_1(f)|, \qquad (6.1)
\end{aligned}$$

where $J_1(\cdot)$ is the average cost for the model $\mathbb{P}_1$. Let $D = |J(f) - J_1(f)|$, so

$$\begin{aligned} D &= \left| \int c(y, f(y)) \widetilde{m}_f(\mathrm{d}y) - \int c(y, f(y)) m_f(\mathrm{d}y) \right| \\ &\leq D_1 + D_2 + D_3, \end{aligned} \tag{6.2}$$

where

$$\begin{aligned} D_1 &= \left| \int c(y, f(y)) \widetilde{m}_f(\mathrm{d}y) - \int c_\vartheta(y, f(y)) \widetilde{m}_f(\mathrm{d}y) \right| \\ D_2 &= \left| \int c(y, f(y)) m_f(\mathrm{d}y) - \int c_\vartheta(y, f(y)) m_f(\mathrm{d}y) \right|, \end{aligned}$$

and

$$D_3 = \left| \int c_\vartheta(y, f(y)) \widetilde{m}_f(\mathrm{d}y) - \int c_\vartheta(y, f(y)) m_f(\mathrm{d}y) \right|,$$

and, m_f and $\widetilde{m}_f$ are the invariant measures for $\{x_t^f\}$ and $\{\tilde{x}_t^f\}$ respectively.

Observe that from inequality (5.1) it is obtained, for $i = 1, 2$

$$D_i \leq \left[\int L(y) \widetilde{m}_f(\mathrm{d}y) \right] \vartheta^{1-s} \tag{6.3}$$

where s appears in Assumption 4.1c).

In [12] (see also [13]) it has been proved that

$$\int L(y) \widetilde{m}_f(\mathrm{d}y) \leq \frac{b}{1-\lambda}, \tag{6.4}$$

where b and λ are the same as in assumption (2.1).

Hence, from (6.3) and (6.4), it is concluded

$$D_1 + D_2 \leq 2 \frac{b}{1-\lambda} \vartheta^{1-s}. \tag{6.5}$$

On the other hand, provided that $|c_\vartheta(\cdot, \cdot)| \leq \vartheta$ and the definition of the total variation metric, it is obtained that

$$D_3 \leq \vartheta \left\| \widetilde{m}_f - m_f \right\|. \tag{6.6}$$

Now, an estimation of the right side of (6.6) is going to be giving:

Let x_∞^f and $\tilde{x}_\infty^f$ be random variables with distribution m_f and $\widetilde{m}_f$, respectively. Then, for each positive integer n, we have:

$$\begin{aligned} \left\| \widetilde{m}_f - m_f \right\| &= \left\| \tilde{x}_\infty^f - x_\infty^f \right\| \\ &\leq \left\| \tilde{x}_\infty^f - \tilde{x}_n^f \right\| + \left\| \tilde{x}_n^f - x_n^f \right\| + \left\| \tilde{x}_\infty^f - x_n^f \right\|. \end{aligned} \tag{6.7}$$

The first and the last terms in (6.7) are less than $r^{-n}h_x(r)$ for each $r \geq \lambda^{-1}$ and each $x \in X$ (see (2.7) in Lemma 2.1). Then we have:

$$\left\|\widetilde{x}_\infty^f - x_\infty^f\right\| \leq 2r^{-n}h_x(r) + \max_{0\leq t\leq n}\left\|\widetilde{x}_t^f - x_t^f\right\|. \tag{6.8}$$

Applying inductively Lemma 5.3, it can be shown that

$$\max_{0\leq t\leq n}\left\|\widetilde{x}_\infty^f - x_\infty^f\right\| \leq n \sup_{x\in X}\sup_{a\in A(x)}\left\|p_1(\cdot|x,a) - p(\cdot|x,a)\right\|. \tag{6.9}$$

Hence, if $\delta = \sup_{x\in X}\sup_{a\in A(x)}\left\|p_1(\cdot|x,a) - p(\cdot|x,a)\right\|$ results in (6.8):

$$\left\|\widetilde{x}_\infty^f - x_\infty^f\right\| \leq 2r^{-n}h_x(r) + n\delta. \tag{6.10}$$

Taking $n = \max\left\{1, [\log_\rho \delta]\right\}$, where $[z]$ means the greatest integer $\leq z$, $\rho = \frac{1}{r}$ and $\vartheta = \delta^{\frac{s-1}{s}}$ in (6.10) we get

$$\begin{aligned} D_3 &\leq \delta^{-\frac{1}{s}}\left(2\rho h_x(r)\delta\right) + \max\left\{1, \log_\rho \delta\right\}\delta \\ &\leq \left(2\rho^{-1}h_x(r) + 1\right)\delta^{\frac{s-1}{s}}\max\left\{1, \log_\rho \delta\right\}. \end{aligned} \tag{6.11}$$

Notice that the right side of (6.11) is independent of $f \in \mathbb{F}$.

Now combining (6.1), (6.2), (6.5) and (6.11) it is gotten:

$$\widehat{\Delta}(x) \leq 2\left(2\frac{b}{1-\lambda} + 2rh_x(r) + 1\right)\delta^{\frac{s-1}{s}}\max\left\{1, \log_\rho \delta\right\}.$$

7. AN EXAMPLE

The following example has been studied in [4] in order to show the existence of AC-optimal policies and the convergence of the value iteration method. Here assumptions on the example which allow to illustrate the main results in this paper are provided, and conclusions about the average criterion are obtained.

Let $X = [0, \infty)$ and $A(x) = A$, for all $x \in X$, where A is a compact subset of the interval $(0, \Theta]$ (with $\Theta \in A$). Define the models:

$$x_{t+1} = (x_t + a_t\eta_t - \varepsilon_t)^+, \tag{7.1}$$

and

$$\widetilde{x}_{t+1} = (\widetilde{x}_t + a_t\widetilde{\eta}_t - \widetilde{\varepsilon}_t)^+, \tag{7.2}$$

where $t = 0, 1, 2, \ldots$, $x_0 = \widetilde{x}_0 \in X$ is given, $z^+ = \max\{0, z\}$, and $\{\eta_t\}, \{\widetilde{\eta}_t\}$, $\{\varepsilon_t\}$ and $\{\widetilde{\varepsilon}_t\}$ are sequences of independent and identically distributed random variables that satisfy the following assumptions:

Let $\eta, \widetilde{\eta}, \varepsilon$ and $\widetilde{\varepsilon}$ be generic random variables distributed as $\eta_0, \widetilde{\eta}_0, \varepsilon_0$ and $\widetilde{\varepsilon}_0$, respectively.

Let g and $g_1 \in \mathbb{F}$ be defined as:

$$g(x) = \Theta, \quad \text{for all } x \in X, \tag{7.3a}$$

and

$$g_1(x) = \Theta, \quad \text{for all } x \in X. \tag{7.3b}$$

Assumption 7.1.

a) Let $\eta, \widetilde{\eta}, \varepsilon$ and $\widetilde{\varepsilon}$ have continuous and bounded densities, concentrated on $[0, \infty)$;

b) For each $t = 0, 1, 2, \ldots$ η_t is independent of ε_t and $\widetilde{\eta}_t$ is independent of $\widetilde{\varepsilon}_t$;

Let $\xi := \Theta\eta - \varepsilon$ and $\widetilde{\xi} := \Theta\widetilde{\eta} - \widetilde{\varepsilon}$. Also, let $\phi(r) := E(r^{\xi})$ and $\widetilde{\phi}(r) := E(r^{\widetilde{\xi}})$, $r \in \mathbb{R}$.

c) $E(\xi) < 0$ and $E\left(\widetilde{\xi}\right) < 0$;

d) It is supposed that there exist $r_0 > 1$ and $\widetilde{r}_0 > 1$ such that:

$$\phi(r_0) < \infty, \quad \widetilde{\phi}(r_0) < \infty \quad \text{and} \quad \phi'(r_0) = \widetilde{\phi}'(r_0) = 0.$$

e) The function c satisfies Assumption 4.1c) with

$$L(x) = \max\left\{E_x(r^{\tau_0}), E_x\left(r^{\widetilde{\tau}_0}\right)\right\}, \quad x > 0, \text{ and} \tag{7.4}$$

$$L(x) = 1, \quad x = 0, \tag{7.5}$$

where $\tau_0 = \min\{t > 0 : x_t^g = 0\}$, $\widetilde{\tau}_0 = \min\{t > 0 : \widetilde{x}_t^{g_1} = 0\}$ and r will be defined later (see (7.8) below).

It will be seen that this example satisfies the hypotheses of Theorem 4.1. First, in [4] it has been proved that Assumptions 7.1a), 7.1b) and 7.1c) imply that the processes (7.1) and (7.2) satisfy the Assumption 3.1.

It is known that the random walks that are obtained when substituting the policies defined in (7.3a) and (7.3b) in the models in (7.1) and (7.2) i. e.

$$x_{t+1} = (x_t + \Theta\eta_t - \xi_t)^+, \tag{7.6}$$

and

$$\widetilde{x}_{t+1} = \left(\widetilde{x}_t + \Theta\widetilde{\eta}_t - \widetilde{\xi}_t\right)^+, \tag{7.7}$$

are ordered Markov chains (see [12]).

Also, for every policy $f \in \mathbb{F}$ it can be gotten that:

$$x_{t+1} = (x_t + f(x_t)\eta_t - \xi_t)^+ \leq (x_t + \Theta\eta_t - \xi_t)^+,$$

and

$$\widetilde{x}_{t+1} = \left(\widetilde{x}_t + f(x_t)\widetilde{\eta}_t - \widetilde{\xi}_t\right)^+ \leq \left(\widetilde{x}_t + \Theta\widetilde{\eta}_t - \widetilde{\xi}_t\right)^+.$$

hence Assumption 4.1b) and 4.2 hold.

Taking in consideration Assumptions 7.1a), 7.1b) and 7.c), it can be concluded that (7.6) and (7.7) are irreducible and recurrent (see [15]) so Assumptions 2.1a), 2.1b) and 2.1 c) hold for (7.6) and (7.7).

Now using Assumption 7.1d) it is obtained that

$$E_0(r^{\tau_0}) < \infty \quad \text{for} \quad 1 < r < \phi^{-1}(r_0)$$

and

$$E_0\left(r^{\tilde{\tau}_0}\right) < \infty \quad \text{for} \quad 1 < r < \phi^{-1}(\tilde{r}_0),$$

(see [11]).

Then taking $\widetilde{M} = \min\left\{\phi^{-1}(r_0), \phi^{-1}(\tilde{r}_0)\right\}$, we have that $E_0(r^{\tau_0}) < \infty$ and $E_0\left(r^{\tilde{\tau}_0}\right) < \infty$ for some r such that

$$1 < r < \widetilde{M}. \tag{7.8}$$

Then (7.6) and (7.7) satisfy Assumption 2.1d) with

$$\begin{aligned} v(x) &= E_0(r^{\tau_0}), x > 0 \quad \text{and} \\ v(x) &= 1, x = 0, \end{aligned}$$

and $\lambda = r^{-1}$, $b_1 = r^{-1}\left[E_0(r^{\tau_0}) - 1\right]$ for (7.6), and

$$\begin{aligned} \tilde{v}(x) &= E_0\left(r^{\tilde{\tau}_0}\right), x > 0 \quad \text{and} \\ \tilde{v}(x) &= 1, x = 0, \end{aligned}$$

and $\lambda = r^{-1}$, $b = r^{-1}\left[E_0\left(r^{\tilde{\tau}_0}\right) - 1\right]$, for (7.7) (see Lemma 5.4). Then the function L defined in Assumption 7.1e) with r defined in (7.8) satisfies Assumption 4.1a) and 4.2a).

Assumptions 4.1c) and 4.2c) are part of Assumption 7.1e). Then the Theorem 4.1 and the Remark 4.3 can be applied to obtain the following bound for the Index of Perturbations for $x \in X$:

$$\widehat{\Delta}(x) \leq 2\left(2\frac{b}{1-\lambda} + 2rh_x(r) + 1\right)\left\|\mu_\xi - \mu_{\tilde{\xi}}\right\|^{\frac{s-1}{s}},$$

provided that $\left\|\mu_\xi - \mu_{\tilde{\xi}}\right\| \leq e^{\frac{-s}{s-1}}$.

Remember that $\lambda = r^{-1}$, $b = \max\{r^{-1}[E_0(r^{\tau_0}) - 1], r^{-1}[E_0(r^{\tilde{\tau}_0}) - 1]\}$, and $h_x(r) \leq E_x(r^{\tau_0}) + \frac{b}{1-\lambda}$. Observe that even h_x can be estimated for some distribution of ξ.

(Received April 22, 2004.)

REFERENCES

[1] F. Favero and W. J. Runglandier: A robustness result for stochastic control. Systems Control Lett. *46* (2002), 91–97.

[2] E. I. Gordienko: An estimate of the stability of optimal control of certain stochastic and deterministic systems. J. Soviet Math. *50* (1992), 891–899.

[3] E. I. Gordienko: Lecture Notes on Stability Estimation in Markov Decision Processes. Universidad Autónoma Metropolitana, México D.F., 1994.

[4] E. I. Gordienko and O. Hernández-Lerma: Average cost Markov control processes with weighted norms: value iteration. Appl. Math. *23* (1995), 219–237.

[5] E. I. Gordienko and F. S. Salem-Silva: Robustness inequality for Markov control processes with unbounded costs. Systems Control Lett. *33* (1998), 125–130.

[6] E. I. Gordienko and F. S. Salem-Silva: Estimates of stability of Markov control processes with unbounded costs. Kybernetika *36* (2000), 2, 195–210.
[7] O. Hernández-Lerma: Adaptive Markov Control Processes. Springer–Verlag, New York 1989.
[8] O. Hernández-Lerma and J. B. Lasserre: Further Topics on Discrete-Time Markov Control Processes. Springer–Verlag, New York 1999.
[9] K. Hinderer: Foundations of Non-stationary Dynamic Programming with Discrete Time Parameter. (Lectures Notes in Operations Research and Mathematical Systems 33.) Springer–Verlag, Berlin – Heidelberg – New York 1970.
[10] T. Lindvall: Lectures on the Coupling Method. (Wiley Series in Probability and Mathematical Statistics.) Wiley, New York 1992.
[11] R. Lund: The geometric convergence rates of a Lindley random walk. J. Appl. Probab. *34* (1997), 806–811.
[12] R. Lund and R. Tweedie: Geometric convergence rates for stochastically ordered Markov chains. Math. Oper. Res. *20* (1996), 182–194.
[13] S. Meyn and R. Tweedie: Markov Chains and Stochastic Stability. Springer–Verlag, New York 1993.
[14] R. Montes-de-Oca, A. Sakhanenko, and F. Salem-Silva: Estimates for perturbations of general discounted Markov control chains. Appl. Math. *30* (2003), 3, 287–304.
[15] E. Nummelin: General Irreducible Markov Chains and Non-negative Operators. Cambrigde University Press, Cambridge 1984.
[16] S. T. Rachev: Probability Metrics and the Stability of Stochastic Models. Wiley, New York 1991.
[17] V. M. Zolotarev: On stochastic continuity of queueing systems of type G/G/1. Theory Probab. Appl. *21* (1976), 250–269.

Raúl Montes-de-Oca, Departamento de Matemáticas AM-Iztapalapa, Ave. San Rafael Atlixco #186. Col. Vicentina, 09340 México D.F. México.
e-mail: momr@xanum.uam.mx

Francisco Salem-Silva, Programa de investigación en Matemáticas aplicadas y computación IMP eje central Lazaro Cárdenas #152, col. C.P. 07730, México D. F. México.
e-mail: fsalem@fcfm.buap.mx