Average Cost Optimization in Markov Control Processes with Unbounded Cost: Ergodicity and Finite Horizon Approximation*

Evgueni Gordienko[†], J. Adolfo Minjárez-Sosa[‡], Raúl Montes-de-Oca[§]

Abstract

Using the value iteration procedure for discrete-time Markov control processes on general Borel spaces we study a scheme of approximation of average cost optimal policies by solving a sequence of finite horizon optimization problems. In order to work with unbonded costs and to provide the geometric rate of convergence we propose the generalization of a well-known ergodicity condition and of use the technique of weighted norms in spaces of functions and signed measures. Applications of the approximation found could be construction of adaptive policies for Markov control processes with unbounded cost.

Key words. Markov control process, average cost optimal policy, value iteration, finite horizon approximation, geometric convergence.

1 INTRODUCTION

In the theory of discrete-time average cost Markov control processes (MCPs for short) with bounded cost one of a current ergodicity assumption is the following (see Arapostathis, et al (1993)):

*This research was supported in part by the Consejo Nacional de Ciencia y Tecnología (CONACyT) under grant 0635P-E9506, in part by the Fondo del Sistema de Investigación del Mar de Cortés under grant SIMAC/94/ CT-005.

[†]Departamento de Matemáticas, Universidad Autónoma Metropolitana Iztapalapa. A. Postal 55-534, C.P. 09340, México, D.F. MEXICO.

[‡]Departamento de Matemáticas, Universidad de Sonora. Rosales s/n, Col. Centro, C.P. 83000, Hermosillo, Son. MEXICO.

[§]Departamento de Matemáticas, Universidad Autónoma Metropolitana Iztapalapa. A. Postal 55-534, C.P. 09340, México, D.F. MEXICO.

$$\|p(\cdot\,|x,a) - p(\cdot|\,x',a')\|_{\tau} \leq 2\beta, \tag{1}$$

for all states $x, x' \in X$, and actions $a \in A(x)$, $a' \in A(x')$, where $\beta < 1$, $\|\cdot\|_{\tau}$ denotes the total variation norm, and p is the transition kernel of the MCPs considered. In this paper we generalize (1) to MCPs with unbounded costs to prove the existence of infinite horizon average cost optimal policies, and to show that these policies can be approximated by solving a sequence of n-stage optimization problems. Our main goal is to establish the geometric rate of convergence for such approximation and to obtain the similar rate of convergence in the value iteration procedure. Specifically, the Average Cost optimality Equation (ACOE) together with the value iteration procedure is used to approximate the solution of ACOE by means of solutions of the optimality equations for n-stage optimization problems.

The same problem for MCPs with bounded costs was studied in Hernández-Lerma (1989), and the geometric convergence in the uniform norm was obtained, both for value iteration and for approximation of optimal policies. For MCPs with finite state and action spaces the geometric convergence in the value iteration procedure was shown in Federgruen and Schweitzer (1980), Schweitzer and Federgruen (1979), and White (1963).

The value iteration (VI) scheme for MCPs has been studied intensively for the last twenty years. Most of contributions were made for processes with bounded costs. For unbounded cost functions the convergence of VI was investigated, for example, in Cavazos-Cadena (1996), Gordienko and Hernández-Lerma (1995b), Hernández-Lerma (1995), Hordijk, Schweitzer and Tijms (1975), Montes-de-Oca and Hernández-Lerma (1996), Sennott (1991), Spieksma (1990).

To obtain a finite horizon approximation of average cost optimal policies we derive the exponential estimation of the rate of convergence of VI with respect to the weighted norm in a suitable space of unbounded functions. The convergence in VI closely relates to the geometric convergence of distributions of a process with respect to the total variation norm in the space of signed measures. For discrete-time Markov processes (non-controlled) this type of convergence was studied, for example, in Kartashov (1985) and Meyn and Tweedie (1993) using Lyapunov-like ergodicity conditions. Bounds of rate of convergence of VI allows us to prove the geometric convergence in a optimal policy approximation procedure. Constants in the bounds found are calculated in terms of quantities involved in assumption 3 in Section 3.

In Section 2 we present the class of Markov Control Processes we are interested in. In Section 3 we list the assumptions which we use to obtain desired results. Preliminaries are formulated and proved in Section 4. Main results are given in Section 5. Remarks and an example of a control system that satisfies all our assumptions are given in Section 6.

2 CONTROL MODEL

A discrete-time Markov control model $(X, A, A(x), p, c)$ consists of a state space X, a control (or action) space A, sets $A(x)$ of admissible actions in the state $x \in X$, transition law p, and one-stage cost c, satisfying the following. Both X and A are Borel spaces (i.e. some measurable subsets of complete and separable metric spaces). Here and in what follows measurability refer to measurability with respect to a corresponding *Borel σ-algebra*, denoted by $\mathcal{B}$. For each $x \in X$ the set $A(x)$ is supposed to be nonempty and compact, and the set

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

of admissible state-action pairs is assumed to be a measurable subset of $X \times A$. Transition law $p(B|x, a)$, where $B \in \mathcal{B}(X)$ and $(x, a) \in \mathcal{K}$, is a stochastic kernel on X given $\mathcal{K}$. Finally, the one-stage cost $c(x,a)$ is a nonnegative measurable function on $\mathcal{K}$ (possibly unbounded).

Denote by $x_t \in X$ and $a_t \in A(X)$, respectively, the states of the process and the actions chosen at the moments $t = 0, 1, 2, ...,$ and define the spaces of admissible histories up to time $t \geq 1$ by setting $H_t : \mathcal{K}^{t-1} \times X$. An element of H_t is a vector, or history, $h_t = (x_0, a_0, ..., a_{t-1}, x_t)$ where $(x_s, a_s) \in \mathcal{K}$ for $s = 0, 1, ..., t - 1$.

A control policy is a sequence $\pi = \{\pi_t\}$ such that for each $t = 0, 1, ...,$ π_t is a stochastic kernel on A given H_t, and which satisfies the constraint $\pi_t(A(x_t)|h_t) = 1$ for all $h_t \in H_t$. The set of all control policies is denoted by Π

A control policy $\pi = \{\pi_t\}$is said to be stationary policy if there exists a measurable function $f : X \to A$ with $graph(f) \subset \mathcal{K}$ such that the measure $\pi_t(.|h_t)$ is concentrated at the point $f(x_t)$ for every $t = 0, 1, ...$ We will identify a stationary policy with corresponding function f, and use the notation: $f \in \Pi_s$, where $\Pi_s \subset \Pi$ is the class of all stationary policies. The stationary policy f uses the action $a_t = f(x_t)$ if the process is in the state x_t at stage t.

For each policy $\pi \in \Pi$ and initial state $x \in X$ a probability measure P_x^π is defined on the space $\Omega := (X \times A)^\infty$ in a canonical way. (See, e.g. Dynkin and Yushkevich (1979) or Hinderer (1970)). We will denote by E_x^π the corresponding expectation operator.

For $\pi \in \Pi$, $x \in X$, and $n = 1, 2...$, set:

$$J_n(x, \pi) := E_x^\pi \sum_{t=0}^{n-1} c(x_t, a_t), \tag{2}$$

$$J(x, \pi) := \limsup_{n \longrightarrow \infty} J_n(x, \pi)/n. \tag{3}$$

Then $J_n(x, \pi)$ and $J(x, \pi)$ are, respectively, the expected $n-$stage cost and the average expected cost (over infinite horizon) when the policy π is used given the initial state x.

The stationary policy $f_* \in \Pi_s$ is said to be average cost optimal, if

$$J(x, f_*) = \inf_{\Pi} J(x, \pi) \text{ for all } x \in X. \tag{4}$$

In the rest of the paper we will be concerned with the approximation of the policy f_* as in (4) by solving optimization problems involving n-stage costs $J_n(x, \pi),\ n = 1, 2, ...$

3 ASSUMPTIONS

For a given measurable function $v : X \to [\bar{v}, \infty)$ $(\bar{v} > 0)$ let L_v^∞ denote the normed linear space of all measurable functions $u : X \to \Re$ with

$$\|u\|_v := \sup_{x \in X} |u(x)| / v(x) < \infty.$$

We define the weighted total variation norm of a signed measure μ on $\mathcal{B}(X)$ as follows (see Kartashov (1985):

$$\|\mu\|_v := \int_X v(x) |\mu| (dx), \tag{5}$$

where $|\mu|$ denotes the variation of the measure μ . The space of all signed measures on $\mathcal{B}(X)$ with $\|\mu\|_v < \infty$ is denoted by M_v.

Assumption 1. (a) The one-stage cost c is a measurable nonnegative real-valued function on $\mathcal{K}$ with the property that $a \to c(x,a)$ is l.s.c. (lower semicontinuous) on $A(x)$ for every $x \in X$

(b) $\sup_{A(x)} c(x,a) \leq v(x), x \in X$.

(c) For each $u \in L_v^\infty$ the set

$$\left\{(x,a) \in \mathcal{K} | \int_X u(y)p(dy|x,a) \leq r\right\}$$

is Borel in $\mathcal{K}$ for every $r \in \Re$; and the function

$$a \to \int_X u(y)p(dy|x,a),$$

is l.s.c for every $x \in X$. (This function takes finite values due to Assumption 3(b) below).

Assumption 2. For every stationary policy f the (state) Markov process with the transition probability $p(\cdot|x, f(x))$ possesses an unique invariant probability μ_f.

Assumption 3. (a) There is a number $\beta < 1$ such that

$$\|p(\cdot|x,a) - p(\cdot|x^{,}, a^{,})\|_v \leq \beta \left[v(x) + v(x^{,})\right], \tag{6}$$

for each $x, x^{,} \in X, a \in A(x), a^{,} \in A(x^{,})$.

(b) There are $x^* \in X, a^* \in A(x^*)$ such that

$$\|p(\cdot|x^*, a^*)\|_v < \infty. \tag{7}$$

Remark 1 *For non-controlled Markov processes the hypothesis of type (6) was introduced in Kartashov (1985).*

An example of controlled autoregression process will be given in Section 6 for which all above assumptions hold.

4 PRELIMINARIES

The convergence of the approximation procedure of average optimal policies proved in Section 5 depends essentially on behavior of value functions v_n for finite horizon costs and on ergodicity properties of processes when stationary policies are applied.

For each $n = 1, 2...$ and initial state $x \in X$ we define the value function $v_n(x)$ for n-stage optimization problem as follows:

$$v_n(x) := \inf_{\pi \in \Pi} J_n(x, \pi), x \in X, \tag{8}$$

where the expected n-stage cost $J_n(x, \pi)$ was given in (2).

The following simple lemma is used to specify the properties of finite horizon value functions v_n. The proof is a combination of the inequalities (6) and (7).

In what follows we will write $\int$ instead of $\int_X$.

Lemma 2 *Assumption 3 implies the following inequality:*

$$\sup_{f \in \Pi_s} \int v(y) p(dy|x, f(x)) \leq \beta[v(x) + v(x^*)] + \|p(\cdot|x^*, a^*)\|_v \, . \tag{9}$$

Corollary 3 *Under Assumptions 1 and 3 for each $x \in X$ we have:*

$$\sup_{a \in A(x)} \int v(y) p(dy|x, a) \leq \beta[v(x) + v(x^*)] + \|p(\cdot|x^*, a^*)\|_v \, . \tag{10}$$

The last inequality is due to the fact that for each $x \in X, a \in A(x)$ there is stationary policy f with $f(x) = a$ which, in turn, is a consequence of Example 2.6 in Rieder (1978).

Lemma 4 below shows that the functions $v_n, n = 1, 2, ...$ are well-defined, belong to the space L_v^∞, and furthermore, they could be calculated recursively.

Lemma 4 *Suppose that Assumptions 1,2, and 3 hold. Then for each $n \geq 1, v_n \in L_v^\infty$, and*

$$v_n(\chi) = \min_{A(x)} \left[c(x, a) + \int v_{n-1}(y) p(dy|x, a) \right], \quad x \in X. \tag{11}$$

with $v_0 := 0$. *Moreover, there exists a measurable function* $f_n : X \to A$ *such that* $f_n(x) \in A(x)$ *for each* x, *and for every* $x \in X$

$$\min_{A(x)} \left[c(x,a) + \int v_{n-1}(y) p(dy|x,a) \right] = c(x, f_n(x) + \int v_{n-1}(y) p(dy|x, f_n(x)). \tag{12}$$

The finite horizon Dynamic Programming Equations (11) for v_n are well-known, provided that Assumption 1 holds (see, for instance, Bertsekas and Shreeve (1978) for universally measurable solutions of (11)). To ensure the functions v_n are Borel measurable we exploit recurrently the equations (11), Assumption 1 (a), (c) and Corollary 4.3 in Rieder (1978). The fact that $v_n \in L_v^\infty$ is a simple consequence of (11), Assumption 1(b) and (10). At last, the existence of a measurable minimizers f_n in (12) follows from Assumption 1(a), (c) and the mentioned result in Rieder (1978). The lower semicontinuity of the functions $a \to c(x,a) + \int v_{n-1}(y) p(dy|x,a)$, needed in order to use Corollary 4.3 in Rieder (1978), can be verified similarly to the proof of Lemma 4.2 in Gordienko and Hernández-Lerma (1995a).

The proof of the following Lemma is given in Gordienko and Herrnández-Lerma (1995b).

Lemma 5 *Suppose that Assumption 1, 2 and 3 hold. Then:*
(i) for every stationary policy f *the average cost* $J(x,f)$ *is finite;*
(ii) $J(x,f) \equiv J(f)$ *does not depend on initial states* $x \in X$*; and moreover,*
(iii) $J(f) = \int c(y, f(y)) \mu_f(dy)$.

Now we study the ergodicity properties of the processes under consideration which will be used in the proofs in Section 5. The point is to establish that the process with the transition probability $p(\cdot|x, f(x))$ is geometrically ergodic (uniformly in stationary policies $f \in \Pi_s$) with respect to the weighted total variation norm $\|\cdot\|_v$ defined in (5).

Given any stationary policy $f \in \Pi_s$ and initial state $x \in X$, let $\mu_{x,f}^{(t)}$ denote the distribution of x_t.

Lemma 6 *Suppose that Assumptions 2 and 3 hold. Then for each stationary policy* $f \in \Pi_s$ *and every* $x \in X$,

$$\left\| \mu_{x,f}^{(t)} - \mu_f \right\|_v \leq \bar{\nu}^{-1} \|\mu_f\|_v \, v(x) \beta^t, \quad t = 0, 1, 2, ..., \tag{13}$$

where $\bar{v} = \inf_{x \in X} v(x)$, *and the constant* β *is from (6).*

Proof. Let $f \in \Pi_s$ be an arbitrary stationary policy. Consider a Markov process with the transition probability $p(\cdot|x, f(x)), x \in X$. Under Assumption 3 the corresponding transition operator T_f defined by the formula:

$$T_f\mu(\cdot) := \int p(\cdot|x, f(x))\mu(dx)$$

is a bounded operator on M_v. Indeed, as it was shown in Kartashov (1985),

$$\|T_f\| := \sup_{\|\mu\|_v \le 1} \|T_f\mu\|_v = \sup_{x\in X}[v(x)]^{-1} \int v(y)p(dy|x, f(x)), \tag{14}$$

where $\|\cdot\|$ stands for the operator norm corresponding to the norm $\|\cdot\|_v$ in M_v.

From (14) and Assumption 3 follows that

$$\begin{aligned}
\|T_f\| &\le \sup_{x\in X}[v(x)]^{-1} \left| \int v(y)p(dy|x, f(x)) - \int v(y)p(dy|x^*, a^*) \right| \\
&+ \sup_{x\in X}[v(x)]^{-1} \int v(y)p(dy|x^*, a^*) \\
&\le \sup_{x\in X}[v(x)]^{-1} \int v(y) \left|p(dy|x, f(x)) - p(dy|x^*, a^*)\right| + \bar{v}^{-1} \int v(y)p(dy|x^*, a^*) \\
&\le \sup_{x\in X}[v(x)]^{-1}\beta[v(x) + v(x^*)] + \bar{v}^{-1} \int v(y)p(dy|x^*, a^*) < \infty.
\end{aligned}$$

Boundedness of the operator T_f and Assumption 3(a) provide the validity of the hypotheses of Theorem D in Kartashov(1985). This theorem yields that the Markov process with transition probability $p(\cdot|x, f(x))$ is uniformly ergodic with respect to the norm $\|\cdot\|_v$. In particular, the stationary projector P_f of the kernel $p(\cdot|x, f(x))$ is a bounded operator on M_v. Therefore,

$$\|\mu_f\|_v < \infty. \tag{15}$$

Moreover, from Theorem 4 in Kartashov (1985) we get

$$\left\|T_f^t - P_f\right\| \le \bar{v}^{-1} \|\mu_f\|_v \beta^t, \qquad t = 1, 2, ...$$

Now, denoting by δ_x the Dirac measure concentrated at the point $x \in X$, we can write

$$\begin{aligned}
\left\|\mu_{x,f}^{(n)} - \mu_f\right\|_v &= \left\|T_f^n\delta_x - P_f\delta_x\right\|_v \le \left\|T_f^n - P_f\right\| \|\delta_x\|_v \\
&\le \bar{v}^{-1} \|\mu_f\|_v v(x)\beta^n, \qquad n = 1, 2, ...,
\end{aligned}$$

because of $\|\delta_x\|_v = \int v(y)\delta_x(dy) = v(x)$. ■

Lemma 7 *Suppose that Assumption 2 and 3 hold. Then*

$$\sup_{f\in\Pi_s} \|\mu_f\|_v \le B, \tag{16}$$

where $B := (1-\beta)^{-1}[\beta v(x^*) + \|p(\cdot|x^*, a^*)\|_v] < \infty$.

Proof. Let $f \in \Pi_s$ be an arbitrary stationary policy. By invariance of the measure μ_f we have

$$\|\mu_f\|_v = \int v(x)\mu_f(dx) = \int v(x) \int p(dx|y, f(y))\mu_f(dy). \tag{17}$$

In view of (15)we can apply the Fubini Theorem to obtain from (17) the following:

$$\|\mu_f\|_v = \int \mu_f(dy) \int v(x)p(dx|y, f(y))$$

$$= \int \mu_f(dy) \int v(x)\lambda_y(dx) + \int \mu_f(dy) \int v(x)p(dx|x^*, a^*),$$

where $\lambda_y(\cdot) := p(\cdot|y, f(y)) - p(\cdot|x^*, a^*)$.

Therefore

$$\begin{aligned}\|\mu_f\|_v &\le \int \mu_f(dy) \int v(x)|\lambda_y|(dx) + \int v(x)p(dx|x^*, a^*)\\ &\le \int \mu_f(dy)\beta[v(y) + v(x^*)] + \int v(x)p(dx|\ x^*, a^*),\end{aligned}$$

by Assumption 3. Hence

$$(1-\beta)\|\mu_f\|_v \le \beta v(x^*) + \int v(x)p(dx|x^*, a^*),$$

where the right-hand side of the last inequality is finite, and it does not depend on $f \in \Pi_s$. ■

5 MAIN RESULTS

The following theorem states that Assumptions 1, 2 and 3 provide the existence of the solution of ACOE and thus, the existence of an average optimal stationary policy.

Theorem 8 *Under the assumptions 1, 2 and 3 there exist a constant* $\rho*$, *a function* ϕ *in* L_v^∞ *and a stationary policy* $f_* \in \Pi_s$ *such that*

$$\rho_* + \phi(x) = \min_{A(x)} \left\{ c(x,a) + \int \phi(y) p(dy|x,a) \right\} \tag{18}$$

$$= c(x, f_*(x)) + \int \phi(y) p(dy|x, f_*(x)), \quad x \in X;$$

$$\rho_* = J(x, f_*) = \inf_{\Pi} J(x, \pi), \;\; x \in X. \tag{19}$$

Moreover, ϕ *is unique (up to adding a constant) function in* L_v^∞ *satisfying ACOE (18), and the policy* $f*$ *is average cost optimal due to (19).*

Theorem 1 was proved in Hernández-Lerma (1995) under Lyapunov-Like conditions that are a little different from used in this paper. Nevertheless, we can use this proof since it is based on the following:

(i) Certain continuity properties of transition of transition law of MCP under consideration (to ensure the existence of measurable selectors).

(ii) Geometrical ergodicity of a process with respect to the norm $\|\cdot\|_v$ when using stationary policies.

Assumption 1 yields continuity properties required in (i). On the other hand, geometrical ergodicity was proved in Lemma 6.

Now we are ready to estimate a rate of convergence in the following value iteration procedure (see for instance, Hernández-Lerma (1989)).

Let $z \in X$ be an arbitrary, but fixed state. Define a sequence of real-valued functions ϕ_n as $\phi_n(x) := v_n(x) - v_n(z), x \in X$, where the functions $v_n(x)$ are from (8). The solution ϕ to (18) can be taken in such a way that $\phi(z) = 0$. If $\lim_{n\to\infty} \phi_n(x) = \phi(x)$ for each $x \in X$ it is said that one has the convergence of the value iteration procedure.

Theorem 9 *Suppose that Assumptions 1, 2 and 3 hold. Then*

$$|\phi_n(x) - \phi(x)| \leq B \left\|\phi\right\|_v [\bar{v}^{-1} + v(z)] \beta^n v(x), \quad x \in X \tag{20}$$

where $B = (1-\beta)^{-1}[\beta v(x^*) + \|p(\cdot|x^*, a^*)\|_v]$, *and the constant* $\beta < 1$ *is from Assumption 3.*

Remark 10 *As it is shown in Gordienko and Montes-de-Oca (1994)*

$$\|\phi\|_v \leq (1-\beta)^{-1} \max[1, \bar{v}^{-1}(1-\beta)^{-1}\{\beta v(x^*) + \|p(\cdot|x^* a^*)\|_v\}][1 + \bar{v}^{-1} v(z)].$$

Proof. Let $A_*(x) \subset A(x), x \in X$, denote the set of actions for which the minimum of the right-hand side of the first equality in (18) is attained. Consider the MCP $(X, A, A_*(x), p, -\phi)$ with the sets of admissible controls $A_*(x), x \in X$ and the cost function $c_1(x,a) := -\phi(x), (x,a) \in \mathcal{K}$, where ϕ is the solution to ACOE (18). Theorem 8 implies $\phi \in L_v^\infty$, and consequently, Assumption 1 holds true for this process. Since the maps $a \to c(x,a)$ and $a \to \int \phi(y)p(dy|x,a)$ are l.s.c., and $A(x)$ is compact, the set $A_*(x)$ is compact for every x$\in X$. Taking in consideration Assumption 1(c), and Corollary 4.3 in Rieder (1978) about measurable minimizers, we can apply Theorem 8 to the process $(X, A, A_*(x), p, -\phi)$ to get a stationary policy $f_1$that satisfies the ACOE:

$$\begin{aligned} \rho_1 + \varphi(x) &= -\phi(x) + \min_{A_*(x)} \int \varphi(y) p dy|x,a) \\ &= -\phi(x) + \int \varphi(y) p(dy|x, f_1(x)), \ x \in X. \end{aligned} \tag{21}$$

Let $\Pi_1$be the class of all stationary policies for which $f(x) \in A_*(x), x \in X$. Using Lemma 5, ergodicity of the process in $\|\cdot\|_v$ and the fact that $\phi \in L_v^\infty$ we easily verify that $\int [-\phi] d\mu_{f_1} = \inf_{f \in \Pi_1} \int [-\phi] d\mu_f$. In Hernández-Lerma (1995) is proved more:

$$\int \phi d\mu_{f_1} = \sup_{f \in \Pi_s} \int d\mu_f. \tag{22}$$

As it was shown in Montes-de-Oca and Hernández-Lerma (1996) the equation (21) yields the policy f_1 to be a canonical, i.e. if

$$J_n(x, \pi; \phi) := J_n(x, \pi) + E_x^\pi \phi(x_n), \ \pi \in \Pi, \ n = 1, 2, ... \tag{23}$$

then

$$n\rho_* + \phi(x) = J_n(x, f_1; \phi) = J_n^*(x, \phi) := \inf_{\pi \in \Pi} J_n(x, \pi; \phi) \tag{24}$$

for each $x \in X, \ n = 1, 2, ...$ Comparing (23) and (24) we conclude that

$$n\rho_* + \phi(x) = \inf_{f \in \Pi_s} \{J_n(x, \pi) + E_x^\pi \phi(x_n)\} \leq v_n(x) + \sup_{f \in \Pi_s} E_x^\pi \phi(x_n). \tag{25}$$

In view of (22), (13) and (16) the last inequality implies the following:

$$\begin{aligned}
n\rho_* + \phi(x) - v_n(x) &\le \sup_{f\in\Pi} E_x^\pi \phi(x_n) - \sup_{f\in\Pi_s}\int \phi d\mu_f + \int \phi d\mu_{f_1} \\
&\le \sup_{f\in\Pi_s}\left|\int \phi d\mu_{x,f}^{(n)} - \int \phi d\mu_f\right| + \int \phi d\mu_{f_1} \\
&\le \sup_{f\in\Pi_s}\int \sup_X [|\phi|/v| \quad v \left|d\mu_{x,f}^{(n)} - \mu_f\right| + \int \phi d\mu_{f_1} \\
&\le \|\phi\|_v \bar{v}^{-1} B v(x)\beta^n + \int \phi d\mu_{f_1} \\
&\equiv B_1 v(x)\beta^n + \int \phi d\mu_{f_1}, \qquad (26)
\end{aligned}$$

where $B_1 := \|\phi\|_v \bar{v}^{-1} B$.

Since $v_n(z) \le J_n(z, f_1)$ the inequality:

$$n\rho_* + \phi(z) - v_n(z) = J_n(z, f_1) + E_z^{f_1}\phi(x_n) - v_n(z) \ge E_z^{f_1}\varphi(x_n). \qquad (27)$$

follows from (24).

Remembering that $\phi(z) = 0$ and $\phi_n(x) = v_n(x) - v_n(z)$, we get from the inequalities (26) and (27) that

$$\begin{aligned}
&- B_1\beta^n v(x) + E_z^{f_1}\phi(x_n) - \int \phi d\mu_{f_1} \le \phi_n(x) - \phi(x) \qquad (28) \\
&\le B_1\beta^n v(x) + \int \phi d\mu_{f_1} - E_x^{f_1}\phi(x_n).
\end{aligned}$$

Now

$$-B_1 v(z) + \beta^n \le E_z^{f_1}\phi(x_n) - \int \phi d\mu_{f_1},$$

and

$$\int \phi d\mu_{f_1} - E_z^{f_1}\phi(x_n) \le B_1 v(x)\beta^n$$

by virtue of Lemma 6.

The last inequalities together with (28) provide the following inequality:

$$-B_1\beta^n v(x) - B_1\beta^n v(z) \le \phi_n(x) - \phi(x) \le B_1\beta^n v(x) + B_1\beta^n v(z),$$

that, finally proves the desired bound (20):

$$|\phi_n(x) - \phi(x)| \leq B_1\beta^n[v(x) + v(z)v(x)/\inf_X v(x)] = \beta^n v(x)B_1[1 + v(z)/\bar{v}].$$

■

Now following Hernández-Lerma (1989), Gordienko and Hernández-Lerma (1995b) we introduce the stationary policies $f_n, n = 1, 2, ...$ determined by the finite horizon value functions v_n in (8) and (11). We use these policies to approximate the average cost optimal policies in the sense that for each $x \in X$,

$$\lim_{n\to\infty} J(x, f_n) = \rho_* = J(x, f_*) = \inf_\Pi J(x, \pi),$$

with the notations of Theorem 8. In view of Lemma 5 we can rewrite last equalities as $\lim_{n\to\infty} J(f_n) = \rho_*$.

For each n, the stationary policy f_n is defined by the function $f_n(x)$ from Lemma 4, *i.e.* $f_n(x)$ is a measurable minimizer of $c(x, a)+\int v_{n-1}(y)p(dy|x, a)$ over $A(x)$. For the calculation of v_n by means of (11) some numerical procedures could be offered (at least when the state space X is a compact), while to find a solution ϕ to the equation (18) in order to get the average cost optimal policy f_* is a very difficult problem. Also, we give the upper bound for a rate of convergence which allows to inspect an accuracy of approximation due to the fact that all constants in this bound are calculable.

Theorem 11 *Suppose that Assumptions 1, 2 and 3 hold. Then there exists constant $d < \infty$ such that*

$$0 \leq J(f_n) - \rho_* \leq d\beta^n, \text{ for } n = 1, 2, ... \tag{29}$$

Remark 12 *The constant d can be easily calculated explicitly in terms of β, $\bar{v}$, $v(x^*)$, $v(z)$ and $\|p(\cdot|x^*, a^*)\|_v$.(See the proof below.)*

Theorem 13 ***Remark 14*** ***Proof.*** *First we estimate the difference $J(f_n) - \rho_*$ in terms of discrepancy function*

$$D(x, a) := c(x, a) + \int \phi(y)p(dy|x, a) - \phi(x) - \rho_*, \tag{30}$$

used often to prove average cost optimality. (See e.g. Arapostathis, et al (1993), Hernández-Lerma (1989)). By virtue of invariance of the measure μ_f *and Lemma 5 we have*

$$\int D(x, f(x))d\mu_f = \int c(x, f(x))d\mu_f + \int \phi(y) \int p(dy|x, f(x))d\mu_f(x) - \int \phi d\mu_f - \rho_*$$

$$= J(f) + \int \phi d\mu_f - \int \phi d\mu_f - \rho_* = J(f) - \rho_*.$$

The definition of $\|\cdot\|_v$ *and Lemma 5 provide the inequality*

$$\begin{aligned} 0 \le J(f) - \rho_* &= \int D(x, f(x))v(x)/v(x)d\mu_f \\ &\le \|D(\cdot, f(\cdot))\|_v \|\mu_f\|_v \le B \|D(\cdot, f(\cdot))\|_v , \end{aligned} \tag{31}$$

that holds for each stationary policy $f \in \Pi_s$.

Applying (31) to the policy f_n *we estimate* $\|D(\cdot, f_n(\cdot))\|_v$.

The equations (12) are equivalent to the following equalities

$$j_n + \phi_n(x) = \min_{A(x)} \left\{ c(x, a) + \int \phi_{n-1}(y)p(dy|x, a) \right\} \tag{32}$$

$$= c(x, f_n) + \int \phi_{n-1}(y)p(dy|x, f_n(x)), \quad x \in X, n = 1, 2, ...,$$

where $v_v(z) - v_{n-1}(z)$ *is denoted by* j_n. *Then we substitute*

$$c(x, f_n) = j_n + \phi_n(x) - \int \phi_{n-1}(y)p(dy|x, f_n(x))$$

to the definition of D *in (30) to obtain:*

$$\begin{aligned} |D(x, f_n)| &= |(j_n - \rho_*) + [\phi_n(x) - \phi(x)] - \int [\phi_{n-1}(y) - \phi(y)]p(dy|x, f_n(x)| \\ &\le |j_n - \rho_*| + |\phi_n(x) - \phi(x)| \\ &+ \|\phi_{n-1} - \phi\|_v \int v(y)p(dy|x, f_n(x)). \end{aligned} \tag{33}$$

Lemma 2 implies the following inequality for the integral in (33)

$$\begin{aligned} \int v(y)p(dy|x, f_n(x)) &\le \beta[v(x) + v(x^*)] + \| \ p(\cdot|x^*, a^*)\|_v . \\ &\le v(x)[\beta + \{\beta v(x^*) + \|p(\cdot|x^*, a^*)\|_v\}/\bar{v}]. \end{aligned}$$

The geometric upper bounds for $|\phi_n(x) - \phi(x)|$ and $\|\phi_{n-1} - \phi\|_v$ in (33) are supplied by Theorem 9. To complete the proof we estimate the term $|j_n - \rho_*|$ in (33), making use of the equality (32), ACOE (18) and the inequality (10). We have

$$\begin{aligned}
|j_n - \rho_*| &\le |\phi_n(x) - \phi(x)| \\
&\quad + \left| \min_{a\in A(x)} \left\{ c(x,a) - \int \phi_{n-1}(y)p(dy|x,a) \right\} - \min_{a\in A(x)} \left\{ c(x,a) - \int \phi(y)p(dy|x,a) \right\} \right| \\
&\le |\phi_n(x) - \phi(x)| + \sup_{a\in A(x)} \int |\phi_{n-1}(y) - \phi(y)| p(dy|x,a) \\
&\le |\phi_n(x) - \phi(x)| + |\, \|\phi_{n-1} - \phi\|_v \sup_{a\in A(x)} \int v(y)p(dy|x,a) \\
&\le |\phi_n(x) - \phi(x)| + \|\phi_{n-1} - \phi\|_v \, v(x)[\beta + \{\beta v(x^*) + \|p(\cdot|x^*, a^*\|_v]/\bar{v}]
\end{aligned}$$

Again exploiting Theorem 2 we, finally, obtain (29). ■

6 Remarks and an example

Remark 15 *In chapter 3 in Hernández-Lerma (1989) inequalities similar to (20) and (26) were proved for MCPs with bounded one-stage costs c. It was done under the assumption (1)*

Remark 16 *The definition of stationary policies f_n as minimizers of $c(x,a) + \int v_{n-1}(y)p(dy|x,a)$ in (11) presupposes precise calculation of the value functions v_n. This is not realistic condition from the point of view of constructing numerical algorithms . Analysis of the proof of Theorem 11 shows that it can be extended in the following direction. Suppose we get a sequence of measurable functions $\{\bar{v}_n\}$ for which*

$$|v_n(x) - \bar{v}_n(x)| \le \varepsilon_n v(x),\ \varepsilon_n \ge 0,\ x \in X.$$

Let us define for $n = 1, 2, ...$ stationary policies $\bar{f}_n$ as minimizers on $A(x)$ of the functions

$$c(x,a) + \int \bar{v}_{n-1}(y)p(dy|x,a).$$

The upper bounds for "errors of approximation" $J(x, \bar{f}_n) - \rho_*$ can be obtained similarly to the proof of Theorem 11. Besides the term $d\beta^n$ as in (26) these *bounds contain a summand depending on values of ε_n. Such bounds could be also useful to construct adaptive control policies for MCPs with unknown transition laws $p(\cdot|x,a)$ which need to be estimated recurrently in the course of realization of a process. (For more information on this type policies see, for example, Hernández-Lerma (1989) or Gordienko (1985)).*

Another possible application of the extension of Theorem 11 mentioned above is a use of it to obtain upper bounds of robustness of MCPs of type considered here (see Gordienko (1992)).

Remark 17 *It seems to be truth that under assumption made the rate in (26) could not be improved. On the order hand, it is interesting to compare (29) with the result in Puterman and Brumelle (1979) which proves the rate of convergence of the policy iteration procedure to be faster than geometric one. In Puterman and Brumelle (1979) this fact was shown for some finite state MCPs with a discounted reward. The numerical experiments presented in White and Scherer (1994) show that a rate of convergence of value iteration can be faster than exponential in the discounted problem for finite state-action MCPs.*

Now we give an example of MCP that satisfies Assumptions 1,2, and 3 in Section 3.

Example 18 *A controlled autoregression process.*

Consider the process of the form:

$$x_{t+1} = \rho(a_t)x_t + \xi_t,\ t = 0, 1, ..., \tag{34}$$

where : $\xi_0, \xi_1, \xi_2, ...,$ are independent uniformly distributed on $[0, 1]$ random variables. The process (34) is MCP when we choose the state space $X := [0, \infty)$, the sets of admissible actions $A(x) \equiv A,\quad x \in X$, with $A \subset \Re$ being a compact set, and define some nonnegative measurable, and lower semicontinuous in a one-stage cost function $c(x, a)$. We suppose that $\rho : A \to (0, \alpha]$ is a given measurable function and $\alpha < \frac{1}{2}$.

We will verify Assumptions 1,2 and 3 taking $v(x) := x+\delta,\ x \in X,$ where $\delta = (1 - 2\alpha)/2$. Also we suppose that $\sup_{a\in A} |c(x, a)| \leq x + \delta,\ x \in X$. It is easily to check fulfillment of Assumption 1(c) for this example. Straightforward calculations of $\|p(\cdot|x, a) - p(\cdot|x', a')\|_v$ show that the inequality (6) in

Assumption 3 holds with $\beta = \alpha + \frac{1}{2}$. On the other hand, the condition (7) is satisfied because $E(\xi_0)$ is finite. Finally, Assumption 2 follows from next proposition.

Proposition 19 *Consider the Markov process*

$$x_{t+1} = L(x_t)x_t + \xi_t, \quad t = 0, 1, ..., \tag{35}$$

with independent uniformly distributed on $[0, 1]$ *random variables* $\xi_0, \xi_1, \xi_2, ...,$ $x_0 \in [0, \infty)$, *and* $L : X \to \Re$ *a measurable function. If* $L(x_t) \leq \gamma < 1$, *and* m *is the Lebesgue measure on* $[0, 1]$, *then the process (32) is* m*-irreducible, aperiodic and satisfies the Doeblin´s condition for the measure* m.

The Doeblin´s condition is valid evidently. The irreducibility can be derived from the facts that

$$x_n = x_0 \prod_{t=0}^{n-1} L(x_t) + \zeta_n + \xi_n,$$
$$\zeta_n = \sum_{t=0}^{n-1} \prod_{s=1}^{t} L(x_s)\xi_{n-s},$$

ζ_n and ξ_n are independent, and for any $\varepsilon > 0$,

$$x_0 \prod_{t=0}^{n-1} L(x_t) < \varepsilon, \;\; P(\zeta_n < \varepsilon) > 0 \;\text{ for some }\; n \geq 1.$$

If we suppose that the process (35) has a period $k > 1$, then there are disjoint sets $C_1, ..., C_k \subset [0, \infty)$ such that $p(\cup_{i=1}^{k} C_i|x) = 1,\; x \in \cup_{i=1}^{k} C_i$, and $p(C_i|x) = 0$ if $x \in C_i, i = 1, ..., k$. Writing the transition probability $p(\cdot|x)$ of (35) trough the Lebesgue measure we can find $z \in [0, 1] \cap C_j$, for some $j \leq k$ such tha $p(C_j|x) > 0$.

Remark 20 *The above example is like to be degenerate if we are thinking of MCPs with unbounded costs. The reason is that the support of the invariant probability* μ_f *is in the interval* $[0, 2]$ *for each* $f \epsilon \Pi_s$. *Nevertheless, to avoid this effect it is possible to choose a distribution of the random variable* ξ_0 *with an unbounded support, but to be close with respect to the norm* $\|\cdot\|_v$ *to the uniform distribution on* $[0, 1]$. *Then the inequality (6) in Assumption 3 holds true with some* $\beta > \alpha + \frac{1}{2}$.

It seems to be a difficult problem to satisfy Assumption 3, when one deals with particular MCPs in applied fields. Indeed, it is not clear how to look for a suitable "excessive function" $v(x)$ in (6). Sometimes it is easier to verify the following set of conditions which could be used instead of Assumption 3 to prove Theorems 1,2 and 3.

a) $\left\| \mu_{x,f}^{(t)} - \mu_f \right\|_v \leq bv(x)\beta^t, \;\; t = 0, 1, 2, ...$ for some $\beta < 1$;

b) $\sup_{f \in \Pi_s} \|\mu_f\|_v < \infty$;

c) $\left\| \sup_{f \in \Pi_s} \int v(y) p(dy|x, f(x)) \right\|_v < \infty$.

In the case of bounded cost function c we need only the condition (a) with $v(x) =$ constant and $\|\cdot\|_v$ to be reduced to the usual total variation norm.

References

[1] Araposthatis, A., B. Borkar, E. Fernández-Gaucherand, M. K. Ghosh, S.I. Marcus (1993). Discrete-time controlled Markov processes with average cost criterion: a survey. SIAM J. Control Opt., 31, 283-344.

[2] Bertsekas, D.P., S.E. Shreve (1978). Stochastic Optimal Control: The Discrete Time Case. Academic Press, New York.

[3] Cavazos-Cadena, R. (1996). Value iteration in a class of communicating Markov decision chains with the average cost criterion. To appear in SIAM J. Control Opt.

[4] Dynkin, E.B., A.A. Yushkevich (1979). Controlled Markov Processes. Springer-Verlag, New York.

[5] Federgruen, A., P. J. Schweitzer (1980). A survey of asymptotic value-iteration for undisconted markovian decision processes. In "Recent Developments in Markov Decision Processes" R. Hartley, L.C. Thomas, and D.J. White ed. Academic Press.

[6] Gordienko, E.I. (1985). Adaptive strategies for certain class of controlled Markov processes. Theory Prob. Appl., 29, 504-518.

[7] Gordienko, E.I. (1992). An estimate of the stability of optimal control of certain stochastic and deterministic systems. J. Soviet Math., 59, 891-899.

[8] Gordienko, E.I:, O. Hernández-Lerma (1995a). Average cost Markov control processes with weighted norms: existence of canonical policies. Applicationes Math. 23 199-218.

[9] Gordienko, E.I., O. Hernández-Lerma (1995b). Average cost Markov control processes with weighted norms: value iteration. Applicationes Math, 23, 219-237.

[10] Gordienko, E.I., R. Montes-de-Oca (1994). The existence of average optimal policies for Markov control processes with unbounded cost under ergodicity conditions. Report 4.0405.101.011.94. UAM-I, México, D.F.

[11] Hernández-Lerma, O. (1989). Adaptive Markov Control Processes. Springer-Verlag, New York.

[12] Hernández-Lerma, O. (1995). Infinite-horizon Markov control processes with undiscounted cost criteria: from average to overtaking optimality. Reporte Interno Num. 165. Departamento de Matemáticas, CINVESTAV-IPN, Apartado Postal 14-740, 07000 México, D.F., MEXICO.

[13] Hinderer, K. (1970). Foundations of Non-Stationary Dynamic Programming with Discrete Time Parameter. Lecture Notes Oper. Res. 33. Springer-Verlag, New York.

[14] Hordijk, A., P.J. Schweitzer, H.C. Tijms (1975). The asymptotic behavior of the minimal total expected cost for the denumerable state Markov decision model J. Appl. Prob., 12, 298-305.

[15] Kartashov, N.V. (1985). Inequalities in theorems of ergodicity and stability for Markov chains with common phase space. J. Theory Probab. Appl., 30, 507-515.

[16] Meyn, S.P., R.L. Tweedie (1993). Markov Chains and Stochastic Stability. Springer-Verlag, New York.

[17] Montes-de-Oca, R., O. Hernández-Lerma (1996). Value iteration in average cost Markov control processes on Borel spaces. Acta Applicandae Mathematicae, 42, 203-222.

[18] Puterman, M.L., S.L. Brumelle (1979). On the convergence of policy iteration in stationary dynamic programming. Math . Oper. Res. 4 60-69.

[19] Rieder, U. (1978). Measurable selection theorems for optimization problems. Manuscripta Math., 24, 115-131.

[20] Schweitzer, P.J., A. Federgruen (1979). Geometric convergence of value iteration in multichain Markov decission problems. Adv. Appl. Prob., 11, 188-217.

[21] Sennott, L.I. (1991). Value iteration in countable state average cost Markov decision processes with unbounded costs. Ann. Oper. Res. 29, 261-271.

[22] Spieksma, F.M. (1990). Geometrically ergodic Markov chains and the optimal control of queues. Ph. D. Thesis Univ. of Leiden , Netherlands.

[23] White, D. (1963). Dynamic programming, Markov chains, and the method of successive approximations. J. Math. Anal. Appl. 6, 373-376.

[24] White, D., W.T. Scherer (1994). The convergence of value iteration in discounted Markov decision processes. J. Math. Anal. Appl. 182, 348-360.