KYBERNETIKA — VOLUME 36 (2000), NUMBER 2, PAGES 195–210

ESTIMATES OF STABILITY OF MARKOV CONTROL PROCESSES WITH UNBOUNDED COSTS[1]

EVGUENI I. GORDIENKO AND FRANCISCO SALEM

For a discrete-time Markov control process with the transition probability p, we compare the total discounted costs $V_\beta\ (\pi_\beta)$ and $V_\beta(\tilde{\pi}_\beta)$, when applying the optimal control policy π_β and its approximation $\tilde{\pi}_\beta$. The policy $\tilde{\pi}_\beta$ is optimal for an approximating process with the transition probability $\tilde{p}$.

A cost per stage for considered processes can be unbounded. Under certain ergodicity assumptions we establish the upper bound for the relative stability index $[V_\beta(\tilde{\pi}_\beta) - V_\beta(\pi_\beta)]/V_\beta(\pi_\beta)$. This bound does not depend on a discount factor $\beta \in (0,1)$ and this is given in terms of the total variation distance between p and $\tilde{p}$.

1. INTRODUCTION

The problem of stability (or robustness) of policy optimization in control processes, naturally arises when a controller has no complete information on a law governing a dynamics of a process. In most cases of interest a controller needs to rely on some approximation of a law obtained from theoretical models or/and from statistical data. Such an uncertainty is typical in real applications of optimal control theory. It appears also in adaptive control models.

In setting of a problem of stability estimation for general discrete-time Markov control processes (MCP's) we will follow the approach proposed in [2, 3]. Let $\mathcal{P}$ and $\tilde{\mathcal{P}}$ be two discrete-time MCP's defined on the same Borel state space X and action space A equipped with the same nonnegative one-stage cost (possibly unbounded). The only difference between processes $\mathcal{P}$ and $\tilde{\mathcal{P}}$ is their transition probabilities denoted, respectively, by p and $\tilde{p}$. We interpret $\tilde{p}$ as a known approximation to an unknown "true" transition probability p of a "real" process $\mathcal{P}$.

The original goal of control optimization is to look for a policy for $\mathcal{P}$ that provides performance as close to the optimal value $V_\beta\ (x, \pi_\beta)$ as possible. Here π_β is *the optimal policy* for $\mathcal{P}$ (supposing the existence of it). As a performance criterion we use the *expected total discounted cost* $V_\beta\ (x, \pi)$ which is a function of an initial state x of a process and of a policy π applied; $\beta \in (0,1)$ is a given *discount factor*.

[1]Research supported by Consejo Nacional de Ciencia y Tecnología CONACYT under grant 4002000-5-25159-E.

Not having an opportunity to find the policy π_β without knowing p, one can try the policy $\tilde{\pi}_\beta$ optimal for the completely known process $\tilde{\mathcal{P}}$ as an approximation to π_β. The answer to the question:

"In what extent is it a good approximation?" depends on the value of $V_\beta(x, \tilde{\pi}_\beta) - V_\beta(x, \pi_\beta) \geq 0$ of the additional cost paid for replacing π_β by $\tilde{\pi}_\beta$. Sometimes it is more reasonable to be interested in a value of a relative increase of cost:

$$\Delta_\beta(x) := [V_\beta(x, \tilde{\pi}_\beta) - V_\beta(x, \pi_\beta)]/V_\beta(x, \pi_\beta), \tag{1.1}$$

that we call *"relative stability index"*.

The rate of vanishing of $\Delta_\beta(x)$ as $\tilde{p}$ approaches p depends heavily on properties of a class of processes under consideration, and even more, on type of convergence $\tilde{p}$ to p. In general, the relative stability index might not vanish as $\tilde{p} \to p$. (In the example 1 of the last section $\Delta_\beta(x)$ does not go to zero and $\sup_{\beta \in (0,1)} \Delta_\beta(x) = \infty$ in spite of $\tilde{p}$ converges to p in the weak topology.)

The aim of the present paper is finding upper bounds for $\Delta_\beta(x)$ which do not depend on a discount factor β. In the papers [2, 3, 17, 18, 20] some results have been obtained on the similar problem with the average cost per unit of time as a criterion of optimization of control. In particular, in [3] we used Zolotarev's metric approach developed in [21] for some uncontrollable stochastic processes. In the case of the discounted total cost optimization some known bounds of $V_\beta(\tilde{\pi}_\beta) - V_\beta(\pi_\beta)$ (see [7, 6, 17, 18, 20]) have the following structure:

$$V_\beta(\tilde{\pi}_\beta) - V_\beta(\pi_\beta) \leq M(1-\beta)\, \psi(p, \tilde{p}), \tag{1.2}$$

where ψ is some "measure of difference" between p and $\tilde{p}$, and $M(y)$ $y \in (0,1)$ is some given function.

It is well-known that under broad conditions $0 < \limsup_{\beta \to 1}(1-\beta)\, V_\beta(\pi_\beta) < \infty$ (see, for instance, [8]). Thus, having in the mind finding bounds for $\Delta_\beta(x)$ independent of β we need something as (1.2) with $M(1-\beta) = O((1-\beta)^{-1})$ as $\beta \to 1$.

In the paper [7] optimality equation and the technique of contractive operators were used to establish bounds as in (1.2). Unfortunately, this approach provides $M(1-\beta) = \mathcal{O}((1-\beta)^{-2})$ as $\beta \to 1$. The papers [17, 18, 20] apply other methods than [7], and deal with Markov control processes on denumerable state spaces. The results of [17, 18], for example, allow to obtain simple and tight bounds for many queueing control systems. The order of the constant $M(1-\beta)$ (as $\beta \to 1$) depends in [17, 18, 20] on properties of some quantities involving an optimal discounted cost V_β, and, finally, on assumptions about processes considered. Under some mild assumptions, again the constant is $M(1-\beta) = \mathcal{O}((1-\beta)^{-2})$ as $\beta \to 1$. To extract from the results of [17, 18, 20] the bounds as in (1.2) with M of order of $\mathcal{O}((1-\beta)^{-1})$, one needs to exploit some sort of ergodicity assumptions.

In this paper we work out the approach different from those used in [7, 17, 18, 20] which allows to get the explicit bounds of the form:

$$\sup_{\beta \in (0,1)} \Delta_\beta(x) \leq \psi(\|p - \tilde{p}\|), \tag{1.3}$$

where ψ is the power function of the total variation distance between $p(\cdot|x, a)$ and $\tilde{p}(\cdot|x, a)$ (uniform over states x and actions a). To achieve the goal as (1.3) we are forced to impose rather restrictive *ergodicity* hypotheses on a class of processes dealt with. As it was shown in the paper [4], such hypotheses lead to uniform (over stationary policies) geometric ergodicity of a control process. To show this we have applied in [4] the results from [11] on geometric ergodicity of uncontrolled discrete-time Markov processes. The uniform ergodicity (together with some additional technique) permits to reduce the problem to the study of stability of ergodic uncontrolled processes, and so to apply Zolotarev's metric technique [21]. This technique proposes using some estimates of the proximity of the distributions of a couple of processes on finite time-intervals, and the known rate of convergence these distributions to stationary ones.

2. CONTROL MODEL AND ASSUMPTIONS

We consider a pair of standard discrete-time ($t = 0, 1, 2, \ldots$) *Markov control models* (see, for instance, [1]) $\mathcal{P} = (X, A, A(x), p, c)$ and $\tilde{\mathcal{P}} = (X, A, A(x), \tilde{p}, c)$ with Borel *state space* X and *action space* A. The sets $A(x) \subset A$, $x \in X$ are sets of *feasible control actions* when a process is in state x. We assume $A(x)$ to be nonempty and *compact* for each $x \in X$, and we suppose the set $\mathbb{K} := \{(x, a) : x \in X, a \in A(x)\}$ of admissible state-action pairs to be a Borel subset of the Cartesian product $X \times A$. Saying about measurability we will mean in what follows measurability with respect to a corresponding Borel σ-algebra $\mathcal{B}$.

The *one-stage cost* is a nonnegative measurable function $c : \mathbb{K} \to \mathbb{R}$, possibly unbounded. The only different components of the above models are the *transition probabilities* $p(B|x, a)$ and $\tilde{p}(B|x, a)$, $(x, a) \in \mathbb{K}$, $B \in \mathcal{B}_X$, those are stochastic kernels on X given $\mathbb{K}$.

We use the standard definition [1] of a *control policy* π, and of a *stationary (deterministic) policy* as a given measurable function $f : X \to A$ with graph $(f) \subset K$ such that at the state x_t the control $a_t = f(x_t)$ is used. We denote by Π the class of all policies, and by $S \subset \Pi$ the subclass of all stationary policies, using the notation $\mathrm{f} = (f, f, \ldots)$ for policies from S. By E_x^π we will denote the expectation in the space of trajectories of a process when using the policy π with the initial state x of a process.

As a performance criterion we use the *total expected β-discounted cost* ($\beta \in (0, 1)$) defined for the process $\mathcal{P}$ as:

$$V_\beta(x, \pi) := \sum_{t=0}^{\infty} \beta^t E_x^\pi c(x_t, a_t), \tag{2.1}$$

when using the policy π with the initial state of the process $x_0 = x$.

Similarly, to (2.1) the total expected β-discounted cost $\tilde{V}_\beta(x, \pi)$ is defined for the process $\tilde{\mathcal{P}}$.

Policies π_β, $\tilde{\pi}_\beta$ are called *optimal*, respectively, for the process $\mathcal{P}$ and $\tilde{\mathcal{P}}$ if:

$$V_\beta(x, \pi_\beta) = V_\beta^*(x) := \inf_{\pi \in \Pi} V_\beta(x, \pi); \tag{2.2}$$

$$\tilde{V}_\beta(x, \tilde{\pi}_\beta) = \tilde{V}^*_\beta(x) := \inf_{\pi\in\Pi} \tilde{V}_\beta(x, \pi), \tag{2.3}$$

for all $x \in X$.

Also we will appeal to the optimal value of *long-run expected average cost* for $\mathcal{P}$:

$$\mathcal{J}_*(x) := \inf_{\pi\in\Pi} \limsup_{n\to\infty} n^{-1} E^\pi_x \sum_{t=0}^{n-1} c(x_t, a_t). \tag{2.4}$$

Let us fix throughout a measurable function ("test function") $W : X \to [1, \infty)$ and introduce the assumptions which we use to prove our results.

Assumption 1. (*Continuity and bounding conditions*)

(a) $\inf_{(x,a)\in I\!K} c(x, a) \geq \kappa > 0$;

$$\sup_{a\in A(x)} c(x, a) \leq [W(x)]^{1/s}, \quad x \in X;$$

where κ and $s > 1$ are given constants;

(b) for each $x \in X$ the map $a \to c(x, a)$ is lower semicontinuous on $A(x)$;

(c) both kernels p and $\tilde{p}$ are strongly continuous on A in the sense: for every measurable and bounded function $u : X \to I\!R$ and $x \in X$ the following maps are continuous:

$$a \to \int_X u(y)\, p(\mathrm{d}y|x, a), \quad a \to \int_X u(y)\tilde{p}(\mathrm{d}y|x, a),$$

$$a \to \int_X W(y)\, p(\mathrm{d}y|x, a), \quad a \to \int_X W(y)\tilde{p}(\mathrm{d}y|x, a).$$

Assumption 2. (*Recurrence condition*) For each stationary policy $\mathbf{f} \in S$ the Markov processes with the transition probabilities $p(\cdot|x, f(x))$ and $\tilde{p}(\cdot|x, f(x))$ are positive Harris-recurrent.

Remark that the Markov process $x_0, x_1, x_2, \ldots$ in the state space X is said to be *Harris-recurrent* if there exists a nontrivial σ-finite measure λ on X such that

$$P(x_t \in B \quad \text{for some } t|x_0 = x) = 1 \quad \text{for all } \ x \in X,$$

whenever $\lambda(B) > 0$ [13].

Assumption 3. (*Ergodicity conditions*) There exist a probability ν on $(X, \mathcal{B}_X)$ and a number $\alpha \in [0, 1)$ for which the following holds:

For each stationary policy $\mathbf{f} \in S$ there is a nonnegative measurable function h_f on X such that for every $x \in X$ and $B \in \mathcal{B}_X$:

(a) $p(B|x, f(x)) \geq h_f(x)\, \nu(B)$;

(b) $\int_X W(y)\, p(\mathrm{d}y|x, f(x)) \leq \alpha W(x) + h_f(x) \int_X W(y)\, \nu(\mathrm{d}y) < \infty$;

(c) $\inf_{\mathbf{f}\in S} \int_X h_f(x)\, \nu(\mathrm{d}x) =: \gamma > 0$.

Assumption 3*. Assumption 3 holds for the transition probability $\tilde{p}$ with the same W, ν, α, γ and some $\tilde{h}_f$ ($\mathbf{f} \in S$).

Comments on the assumptions.

(1) In the last section of the paper we give an example of $GI|GI|1|\infty$ queueing system with controlled service rates. For this example all above assumptions are satisfied.

(2) In view of Proposition 1 below the above assumptions guarantee the existence of optimal stationary policies for both processes $\mathcal{P}$ and $\tilde{\mathcal{P}}$. This is useful for our purposes, but it is not necessary (see the comments (4) below). Moreover, optimal stationary policies exist under less restrictive conditions. What is important for us is that under Assumptions 2 and 3 for each stationary policy the corresponding Markov process is Harris-recurrent with a unique invariant distribution, and with the geometric rate of convergence to this distribution. It is stated in Proposition 2 below. Moreover, the cumbersome constants in (26), (27) allow to give *explicit bounds* for the constants involved in the estimation of the rate of convergence. This is important for us to obtain explicitly calculated bounds in the stability inequality. The proof of Proposition 2 given in [4] relies on corresponding results by Kartashov [11], where the estimates in the geometric rate of convergence were given for some uncontrolled Markov processes.

To get the inequality similar to (3.1) in Theorem of the next section one can use any known estimates of geometric convergence, for example, the new rather tight estimates in [16] for some particular classes of Markov processes.

(3) The parts (b) and (c) of Assumption 3, can be checked for the process $\tilde{\mathcal{P}}$ with some $\tilde{\gamma} > 0$, $\tilde{\alpha} < 1$; then to satisfy Assumption 3* one can take $\max(\alpha, \tilde{\alpha})$ and $\min(\gamma, \tilde{\gamma})$.

(4) In view of given below Proposition 1 Assumptions 1, 2, 3 and 3* ensure the existence of optimal stationary policies $\mathbf{f}_\beta$, $\tilde{\mathbf{f}}_\beta$ for the processes $\mathcal{P}$ and $\tilde{\mathcal{P}}$. These policies are used to define the relative stability index in (1.1). On the other hand, Assumptions 2, 3 and 3* which, as it will be seen guarantee ergodicity of processes, can fail to hold for some (and even for many) stationary policies. To see this, one can consider MCP's given by linear recurrent equations (see, for example [12]). Examination of the proof of Theorem in the next section shows that we can significantly relax Assumptions 2, 3 and 3* postulating the existence of optimal policies $\mathbf{f}_\beta$ and $\tilde{\mathbf{f}}_\beta$ and some subset $S_0 \subset S$ of stationary policies such that $\mathbf{f}_\beta$, $\tilde{\mathbf{f}}_\beta \in S_0$ and Assumptions 2, 3 and 3* are satisfied for each $\mathbf{f} \in S_0$ (but probably not for each $\mathbf{f} \in S$). The bound (3.1) holds true under such modification of hypotheses, and, moreover, in this situation we do not need to use parts (b) and (c) of Assumption 1, and the supposition about compactness of sets $A(x)$ ($x \in X$). Also, it is not difficult to modify slightly the definition of $\Delta_\beta(x)$, the inequality (3.1) and its proof in order to make it sufficient to suppose only the existence of ϵ-optimal policies $\mathbf{f}_{\epsilon,\beta}$, $\tilde{\mathbf{f}}_{\epsilon,\beta} \in S_0$ ($\epsilon > 0$).

Proposition 1. ([4]) Suppose Assumptions 1, 2, 3 and 3* hold and let $\beta \in (0,1)$ be an arbitrary, but fixed discount factor. Then there exist stationary optimal policies $\mathbf{f}_\beta, \tilde{\mathbf{f}}_\beta \in S$, correspondingly, for the processes $\mathcal{P}$ and $\tilde{\mathcal{P}}$.

Also, the optimal average cost $\mathcal{J}_*(x)$ in (2.4) is independent of $x \in X$.

We will use the notation $\mathcal{J}_*$ for $\mathcal{J}_*(x)$. Also, in view of Proposition 1 we can rewrite the value functions defined in (2.2), (2.3) as follows $V_\beta^*(x) = V_\beta(x, \mathbf{f}_\beta)$ and $\tilde{V}_\beta^* = \tilde{V}_\beta(x, \tilde{\mathbf{f}}_\beta)$.

To write down the next proposition we introduce below certain constants which have arisen in the estimate of the rate of convergence in the ergodic Lemma 3.4 in [4]. The appearance of these constants is a little bit intricate, but it is important that they are calculated precisely in terms of quantities involved in Assumptions 1, 2, 3 and 3*. Also, all inequalities given below for constants follow from the proof of Lemma 3.4 in [4].

Fix any positive number $\bar{\gamma}$ such that $\bar{\gamma} \le \gamma \le 1$ (γ is from Assumption 3, part (c)), and using the convention: $(\log 1)/(1-1) := -1$, we define:

$$\omega = 2\exp\{[1 - \|\nu\|_W/(1-\alpha)](\log\bar{\gamma})/(1-\bar{\gamma})\} - 1,$$

where $\|\nu\|_W := \int_X W(y)\,\nu(\mathrm{d}y)$, and then we set:

$$\begin{aligned} d &= 1 - (1-\alpha^2)/[(1-\alpha) + \alpha\omega\|\nu\|_W] < 1; \\ \rho &:= d + q < 1, \end{aligned} \tag{2.6}$$

where q is an arbitrary positive number such that $q < 1-d$. Now let $t_* = [d/(1-d)]$ ($[\cdot]$ is the integer part), $\tau = \max\{0, 1/\log(1+q/d) - 2\}$, and

$$\begin{aligned} b_1 &= d^\tau(\tau+2)/(d+q)^\tau \quad &\text{if} \quad \tau > t_*, \\ b_1 &= d^{t_*}(t_*+2)/(d+q)^{t_*} \quad &\text{if} \quad \tau \le t_*. \end{aligned}$$

Finally, we define:

$$\begin{aligned} \bar{B} = \max\big\{&[1 + b_1 de/\alpha][1 + \|\nu\|_W/(1-\alpha)], \big[\max\{1, (\|\nu\|_W + \alpha/\rho]^{t_*} \\ &+ \|\nu\|_W/[(1-\alpha)\rho^{t_*}]\big\}. \end{aligned} \tag{2.7}$$

Let us fix an arbitrary stationary policy $\mathbf{f} \in S$, and let $\{x_0^{(\mathbf{f})}, x_1^{(\mathbf{f})}, \ldots\}$, $\{\tilde{x}_0^{(\mathbf{f})}, \tilde{x}_1^{(\mathbf{f})}, \ldots\}$ be Markov processes with the transition probabilities, correspondingly, $p(\cdot|x, f(x))$ and $\tilde{p}(\cdot|x, f(x))$. Assumption 2 ensures the existence and uniqueness of the invariant probabilities q_f and $\tilde{q}_f$ for these processes (see for instance, [14]). The following result has been proven in [4] provides the estimates of the rate of convergence of distributions of the processes to invariant distributions with respect to the total variation metric σ.

Proposition 2. ([4]) Under Assumptions 2, 3 and 3*

$$\sup_{\mathbf{f} \in S} \sigma(x_t^{(\mathbf{f})}, x_\infty^{(\mathbf{f})}) \le \bar{B}W(x)\rho^t; \tag{2.8}$$

$$\sup_{\mathbf{f}\in S} \sigma(\tilde{x}_t^{(\mathbf{f})}, \tilde{x}_\infty^{(\mathbf{f})}) \le \bar{B} W(x)\rho^t; \tag{2.9}$$

$t = 1, 2, \ldots$; where the random elements $x_\infty^{(\mathbf{f})}, \tilde{x}_\infty^{(\mathbf{f})}$ (with values in X) have, respectively, the distributions $q_f, \tilde{q}_f$, and the constants $\bar{B} < \infty$, $\rho < 1$ were introduced in (2.6), (2.7).

The total variation distance σ, used in (2.8), (2.9) and in the theorem of the next section, is defined as follows (see, for instance, [15]):

$$\sigma(\xi, \zeta) \equiv \sigma(\mu_\xi, \mu_\zeta) := 2\sup\{|P(\xi \in B) - P(\zeta \in B)| : B \text{ are Borel}\}. \tag{2.10}$$

Here μ_ξ and μ_ζ are the distributions of the random elements ξ and ζ taking values in Borel space X.

If ξ and ζ are random vectors in $\mathbb{R}^k$, having the densities, respectively, g_ξ and g_ζ then,

$$\sigma(\xi, \zeta) = \int_{\mathbb{R}^k} |g_\xi(y) - g_\zeta(y)| \, \mathrm{d}y. \tag{2.11}$$

Also, for random variables ξ and ζ taking values in the same countable set $\{y_1, y_2, \ldots\}$ we have:

$$\sigma(\xi, \zeta) = \sum_{j=1}^{\infty} |P(\xi = y_j) - P(\zeta = y_j)|.$$

3. STABILITY INEQUALITIES

Throughout of the rest of the paper we fix an arbitrary $x \in X$ as an initial state of both processes $\mathcal{P}$ and $\tilde{\mathcal{P}}$, i.e. $x_0 = x$, $\tilde{x}_0 = x$. By virtue of the above Proposition 1 we rewrite the relative stability index in (1.1) as

$$\Delta_\beta(x) = [V_\beta(x, \tilde{\mathbf{f}}_\beta) - V_\beta(x, \mathbf{f}_\beta)]/V_\beta^*(x).$$

By Assumption 1, (a) this quantity is well-defined, i.e. $V_\beta^*(x) > 0$.

Our main results looks as follows.

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

$$\Delta_\beta(x) \le M(\beta, x)\, \delta^{(s-1)/s} \max\{1, \log_\rho \delta\}, \tag{3.1}$$

where:

$$\delta = \sup_{(x,a)\in\mathbb{K}} \sigma(p(\cdot|x,a), \tilde{p}(\cdot|x,a)); \tag{3.2}$$

$$M(\beta, x) = B(x)/[(1-\beta)\, V_\beta^*(x)] \le B(x)/\kappa; \tag{3.3}$$

$$\lim_{\beta\to 1}(1-\beta)\, V_\beta^*(x) = \mathcal{J}_*, \tag{3.4}$$

and, finally,

$$B(x) = 2\{2W(x)\,[1 + 2\bar{B}\rho^{-1}] + 2(1-\alpha)^{-1}\|\nu\|_W + 1\}.$$

Corollary 1. (The inequality for the value functions)

$$|V_\beta^*(x) - \tilde{V}_\beta^*(x)| \leq [2(1-\beta)]^{-1} B(x)\, \delta^{(s-1)/s} \max\{1, \log_\rho \delta\}. \tag{3.5}$$

Proof. In view of Proposition 1 $V_\beta^*(x) = \sup_{\mathbf{f} \in S} V_\beta(x, \mathbf{f})$ and $\tilde{V}_\beta^*(x) = \sup_{\mathbf{f} \in S} \tilde{V}_\beta(x, \mathbf{f})$.

Thus,

$$|V_\beta^*(x) - \tilde{V}_\beta^*(x)| \leq \sup_{\mathbf{f} \in S} |V_\beta(x, \mathbf{f}) - \tilde{V}_\beta(x, \mathbf{f})|. \tag{3.6}$$

On the other hand, the proof of Theorem given below is, in fact, finding upper bound for the right-hand side of the inequality (3.6). This bound appears in (3.1). Therefore, (3.6) is a consequence of (3.1) and (3.3). □

Corollary 2. Let models $\mathcal{P}$ and $\tilde{\mathcal{P}}$ be given by recurrent equations:

$$x_{t+1} = F(x_t, a_t, \xi_t), \tag{3.7}$$

$$\tilde{x}_{t+1} = F(\tilde{x}_t, a_t, \tilde{\xi}_t), \quad t = 0, 1, \ldots \tag{3.8}$$

where $\{\xi_t\}$, $\{\tilde{\xi}_t\}$ are sequences of independent and identically distributed (i.i.d. for short) elements in some Borel space (Y, B_Y) with the common distributions μ_ξ and, respectively, $\mu_{\tilde{\xi}}$.

Then under hypotheses of Theorem

$$\Delta_\beta(x) \leq M(\beta, x)\, [\sigma(\mu_\xi, \mu_{\tilde{\xi}})]^{(s-1)/s}, \tag{3.9}$$

provided that $\sigma(\mu_\xi, \mu_{\tilde{\xi}}) \leq e^{-s/(s-1)}$.

Remark. When the second inequality in Assumption 1, (a) holds for large s the power of δ in (3.1) is closed to the best possible value 1. In the example of the next section $X = [0, \infty)$ and $W(x) = \bar{b}e^{hx}$, $h > 0$, and $\bar{b} > 0$ is arbitrary. Therefore, if $\sup_{a \in A(x)} c(x, a)$ is bounded by some polynomial then, for each $\epsilon > 0$ one can choose $\bar{b} = \bar{b}(\epsilon)$ is such a way that (3.1) holds with $\delta^{1-\epsilon}$.

Proof of Theorem. In view of Proposition 1 of the previous section we get (see (2.1)–(2.3))

$$\begin{aligned}
&V_\beta(x, \tilde{\mathbf{f}}_\beta) - V_\beta(x, \mathbf{f}_\beta) \\
&\leq |V_\beta(x, \tilde{\mathbf{f}}_\beta) - \tilde{V}_\beta(x, \tilde{\mathbf{f}}_\beta| + \left| \inf_{\mathbf{f} \in S} \tilde{V}_\beta(x, \mathbf{f}) - \inf_{\mathbf{f} \in S} V_\beta(x, \mathbf{f}) \right| \\
&\leq 2 \sup_{\mathbf{f} \in S} |\tilde{V}_\beta(x, \mathbf{f}) - V_\beta(x, \mathbf{f})| \\
&\leq 2 \sup_{\mathbf{f} \in S} \sum_{t=0}^{\infty} \beta^t |E_x^{\mathbf{f}} c(x_t, f(x_t)) - E_x^{\mathbf{f}} c(\tilde{x}_t, f(\tilde{x}_t))| \\
&\leq \frac{2}{1-\beta} \sup_{\mathbf{f} \in S} \sup_{t \geq 1} \epsilon_t(\mathbf{f}),
\end{aligned} \tag{3.10}$$

where

$$\epsilon_t(\mathbf{f}) := |E_x^{\mathbf{f}} c(x_t, f(x_t)) - E_x^{\mathbf{f}} c(\tilde{x}_t, f(\tilde{x}_t))|. \tag{3.11}$$

Thus, to prove (3.1) it suffices to show that for all $\mathbf{f}$ and t

$$\epsilon_t(\mathbf{f}) \leq (B(x)/2)\, \delta^{(s-1)/s} \max\{1, \log_\rho \delta\}. \tag{3.12}$$

To reduce the last inequality to estimation of quantities similar to those as in (3.11), but with bounded cost functions we first ensure that for every policy $\mathbf{f} \in S$

$$\sup_{t \geq 0} E_x^{\mathbf{f}} W(x_t) \leq W(x) + \frac{\|\nu\|_W}{1-\alpha} =: \boldsymbol{C}, \tag{3.13}$$

$$\sup_{t \geq 0} E_x^{\mathbf{f}} W(\tilde{x}_t) \leq \boldsymbol{C}. \tag{3.14}$$

Indeed, from Assumption 3(a) and 3(b) we have $\int_X W(y)\, p(\mathrm{d}y | x_{t-1}, f(x_{t-1})) \leq \alpha W(x_{t-1}) + |\nu|_W$, and by Markov property of $\{x_t\}$ we get for any fixed history h_{t-1} $(t \geq 1)$:

$$E_x^{\mathbf{f}}[W(x_t)|h_{t-1}] = \int_X W(y)\, p(\mathrm{d}y)|x_{t-1}, f(x_{t-1})) \leq \alpha W(x_{t-1}) + |\nu|_W,$$

or

$$E_x^{\mathbf{f}} W(x_t) \leq \alpha E_x^{\mathbf{f}} W(x_{t-1}) + |\nu|_W.$$

Iterating the last inequality we obtain

$$\begin{aligned} E_x^{\mathbf{f}} W(x_t) &\leq \alpha^t W(x) + \|\nu\|_W (1 + \alpha + \cdots + \alpha^{t-1}) \\ &\leq W(x) + \|\nu\|_W/(1-\alpha) \text{ that implies (3.13).} \end{aligned}$$

The proof of (3.14) is the same.

Now, for arbitrary, but fixed $b > 0$ we define

$$c_b(x,a) := \begin{cases} c(x,a) & \text{if } \; c(x,a) \leq b \\ 0 & \text{otherwise.} \end{cases}$$

Applying Assumption 1, (a), the Hölder and the Chebyshev inequalities we get for every $t \geq 0$, $\mathbf{f} \in S$ (below: $1/s + 1/\ell = 1$):

$$\begin{aligned} &|E_x^{\mathbf{f}} c(x_t, f(x_t)) - E_x^{\mathbf{f}} c_b(x_t, f(x_t))| \;=\; E_x^{\mathbf{f}} \{c(x_t, f(x_t)); c(x_t, f(x_t)) > b\} \\ &\leq\; E_x^{\mathbf{f}}[W^{1/s}(x_t) I_{\{W(x_t) > b^s)\}}] \;\leq\; \{E_x^{\mathbf{f}} W(x_t)\}^{1/s} \{P(W(x_t) > b^s)\}^{1/\ell} \\ &\leq\; \{E_x^{\mathbf{f}} W(x_t)\}^{1/s} \{E_x^{\mathbf{f}} W(x_t)\}^{1/\ell} b^{-s/\ell} \;\leq\; \boldsymbol{C} b^{1-s}. \end{aligned}$$

Similarly we obtain:

$$|E_x^{\mathbf{f}} c(\tilde{x}_t, f(\tilde{x}_t)) - E c_b(\tilde{x}_t, f(\tilde{x}_t))| \leq \boldsymbol{C} b^{1-s},$$

$\mathbf{f} \in S$, $t = 0, 1, \ldots$.

Therefore (see (3.11)).

$$\begin{aligned} \epsilon_t(\mathbf{f}) \ \le\ & |E_x^{\mathbf{f}} c(x_t, f(x_t)) - E_x^{\mathbf{f}} c_b(x_t, f(x_t))| \\ & + |E_x^{\mathbf{f}} c_b(x_t, f(x_t)) - E_x^{\mathbf{f}} c_b(\tilde{x}_t, f(\tilde{x}_t))| \\ & + |E_x^{\mathbf{f}} c_b(\tilde{x}_t, f(\tilde{x}_t)) - E_x^{\mathbf{f}} c(\tilde{x}_t))| \\ & \le 2\boldsymbol{C} b^{1-s} + \epsilon_{b,t}(\mathbf{f}), \end{aligned} \tag{3.15}$$

where $\epsilon_{b,t}(\mathbf{f})$ stands for next to the last term in the first inequality in (3.15).

Fix an arbitrary $\mathbf{f} \in S$, and let $q_{\mathbf{f}}$, $\tilde{q}_{\mathbf{f}}$ be invariant probabilities for Markov processes with transition probabilities, correspondingly, $p(\cdot|x, f(x))$ and $\tilde{p}(\cdot|x, f(x))$. To apply Zolotarev's approach (see [21]) for estimation of $\sup_{\mathbf{f}\in S} \epsilon_{b,t}(\mathbf{f})$ we need the uniform over $\mathbf{f} \in S$ convergence of distributions of x_t and $\tilde{x}_t$ to invariant probabilities. Such convergence is provided by Proposition 2.

Let $n \ge 1$ be some fixed integer.

Then

$$\epsilon_{b,t}(\mathbf{f}) \le b\sigma(x_t, \tilde{x}_t) \le b \max_{t\le n} \sigma(x_t, \tilde{x}_t) \quad \text{if} \quad t \le n, \tag{3.16}$$

and for $t > n$ by (2.8), (2.9) following [21] we get

$$\begin{aligned} & \epsilon_{b,t}(\mathbf{f}) \le b[\sigma(x_t, x_\infty^{(\mathbf{f})}) + \sigma(x_\infty^{(\mathbf{f})}, \tilde{x}_\infty^{(\mathbf{f})}) + \sigma(\tilde{x}_t, \tilde{x}_\infty^{(\mathbf{f})})] \\ & \le b[2\bar{B}W(x)\rho^n + \sigma(x_\infty^{(\mathbf{f})}, \tilde{x}_\infty^{(\mathbf{f})})]. \end{aligned} \tag{3.17}$$

On the other hand,

$$\begin{aligned} & \sigma(x_\infty^{(\mathbf{f})}, \tilde{x}_\infty^{(\mathbf{f})}) \le \sigma(x_\infty^{(\mathbf{f})}, x_n) + \sigma(x_n, \tilde{x}_n) + \sigma(\tilde{x}_n, \tilde{x}_\infty^{(\mathbf{f})}) \\ & \le 2\bar{B}W(x)\rho^n + \max_{t\le n} \sigma(x_t, \tilde{x}_t). \end{aligned} \tag{3.18}$$

Combining (3.16) – (3.18) we see

$$\epsilon_{b,t}(\mathbf{f}) \le b[4\bar{B}W(x)\rho^n + \max_{t\le n} \sigma(x_t, \tilde{x}_t)], \quad t = 1, 2, \ldots. \tag{3.19}$$

Exploiting the following dual representation

$$\sigma(\mu_\xi, \mu_\zeta) = \sup_{\varphi: \|\varphi\|_\infty \le 1} \left| \int_X \varphi(x)\, \mathrm{d}\mu_\xi - \int_X \varphi(x)\, \mathrm{d}\mu_\zeta \right|$$

of the total variation metric defined in (2.10) (see, for instance, [15]), the Markov property of $x_t, \tilde{x}_t$, the assumption $x_0 = \tilde{x}_0$, and induction arguments one can easily show that for every $\mathbf{f} \in S$

$$\max_{t\le n} \sigma(x_t, \tilde{x}_t) \le n \sup_{x\in X} \sigma(p(\cdot|x, f(x)), \tilde{p}(\cdot|x, f(x)) \le n\delta. \tag{3.20}$$

From (3.15), (3.19) and (3.20) for any $n \geq 1$, $b > 0$

$$\epsilon_t(\mathbf{f}) \leq 2\boldsymbol{C}b^{1-s} + b(4\bar{B}W(x)\rho^n + n\delta) \tag{3.21}$$

with the right-hand side being independent of a policy $\mathbf{f}$. Choose in (3.21) $n = \max\{1, [\log_\rho \delta]\}$, $b = \delta^{-1/s}$; here $[z]$ means the greatest integer $\leq z$. Then, by elementary calculations

$$\begin{aligned}\epsilon_t(f) &\leq 2\boldsymbol{C}\delta^{(s-1)/s} + \delta^{-1/s}(4\bar{B}W(x)\rho^{-1}\delta + \max\{1, \log_\rho \delta\}\delta) \\ &\leq \{2\boldsymbol{C} + 4\bar{B}W(x)\rho^{-1} + 1\}\delta^{(s-1)/s}\max\{1, \log_\rho \delta\}.\end{aligned} \tag{3.22}$$

The last inequality implies (3.12), and hence in view of (3.10) we come to the required inequality (3.1).

The inequality (3.3) is an evident consequence of the definition (2.2) and Assumption 1, (a).

Finally, the existence and the value of the limit in (3.4) can be readily established by insignificant changes in the proof of the Theorem 4.2 in [8]; (see also the proof of theorem 2.6 in [4]).

To get the inequality (3.9) in Corollary 2 it is enough to observe than the function $z^{(s-1)/s}\max\{1, \log_\rho z\}$ is increasing for $z \leq e^{-s/(s-1)}$ and the following inequality for the processes (3.7), (3.8):

$$\sigma(p(\cdot|x,a), \tilde{p}(\cdot|x,a)) = 2 \sup_{B\in\mathcal{B}_X} |P(F(x,a,\xi_0) \in B) - P(F(x,a,\hat{\xi}_0) \in B)|$$

$$= 2 \sup_{B\in\mathcal{B}_X} |P(\xi_0 \in G_{x,a}^{-1}(B) - P(\tilde{\xi}_0 \in G_{x,a}^{-1}(B))| \leq \sigma(\xi_0\tilde{\xi}_0), \text{ where } G_{x,a}(\cdot) := F(x,a,\cdot).$$

□

4. EXAMPLES

Example 1. We start with an example of *unstable* model of the discounted cost optimization. This example illustrates also a point that the convergence of the value functions $\tilde{V}_\beta \to V_\beta$ (as in Corollary 1) can not be regarded as stability of a problem if we are interested in magnitude of Δ_β.

Consider the pair of MCP's given by equations:

$$x_{t+1} = x_t a_t \xi_t; \tag{4.1}$$

$$\tilde{x}_{t+1} = \epsilon + \tilde{x}_t a_t \tilde{\xi}_t; \quad t = 0, 1, \ldots \tag{4.2}$$

on the spaces $X = [0, \infty)$ $A(x) = A := \{0, 1\}$, with the same initial states $x_0 = \tilde{x}_0 = x = 1$. In (4.1), (4.2) ϵ is some number from $(0,1)$; $\{\xi_t\}$, $\{\tilde{\xi}_t\}$ are sequences of i.i.d. random variables (r.v's, for short) such that ξ_0 has the uniform distribution on $[1/2, 1]$, and

$$P(\tilde{\xi}_0 \in (y, y + \mathrm{d}y)) = 2(1-\epsilon)\,\mathrm{d}y, \quad y \in [1/2, 1],$$

$P(\tilde{\xi}_0 = b_\epsilon) = \epsilon$ with some $b_\epsilon \geq 1$ which will be chosen later on.

Let $\beta \in (0,1)$ be a discount factor and the one-stage cost be defined as follows:

$$c(x,0) = \begin{cases} 0 & \text{if } x = 0; \\ 2-x & \text{if } x \in (0,1]; \\ 1 & \text{otherwise;} \end{cases} \tag{4.3}$$

$$c(x,1) = 2 - x, \quad x \in [0,\infty). \tag{4.4}$$

It is clear for every policy π $x_t \in [0,1]$, $t = 0,1,\ldots$, hence the stationary policy with $f_*(x) = 0$, $x \in X$ is optimal for the process (4.1), and $V^*_\beta(1) = V_\beta(1,\mathbf{f}_*) = 1$. On the other hand, applying any policy π to the process (4.2) we get

$$\tilde{x}_t \in [\epsilon,\infty), \tilde{x}^*_t \geq \tilde{x}_t, \quad t = 0,1,\ldots, \tag{4.5}$$

where $\{\tilde{x}^*_t\}$ is the trajectory corresponding to use of the stationary policy with $\tilde{f}_*(x) = 1$, $x \in X$. From (4.3)–(4.5) we have the policy $\tilde{\mathbf{f}}_*$ to be optimal for the process (4.2), and taking b_ϵ to make $\mu = \mu(b_\epsilon) := E\tilde{\xi}_0 \in (1, 1/\beta)$ we obtain

$$\begin{aligned} \tilde{V}^*_\beta(1) = \tilde{V}_\beta(1,\tilde{\mathbf{f}}_*) &= \sum_{t=0}^{\infty} \beta^t E_1^{\tilde{\mathbf{f}}_*} C(\tilde{x}^*_t, 1) \\ &= \sum_{t=0}^{\infty} \beta^t E(2 - \tilde{x}^*_t) = \frac{1}{1-\beta}\left(2 - \frac{\epsilon}{1-\mu}\right) + \frac{1}{1-\beta\mu}\left(\frac{\epsilon}{1-\mu} - 1\right). \end{aligned} \tag{4.6}$$

The last inequality in (4.6) is due to the fact that

$$\tilde{x}^*_t = \tilde{\xi}_0\tilde{\xi}_2\ldots\tilde{\xi}_{t-1} + \epsilon(\tilde{\xi}_1\ldots\tilde{\xi}_{t-1} + \tilde{\xi}_2\ldots\tilde{\xi}_{t-1} + \ldots + \tilde{\xi}_{t-1} + 1),$$

and, thus,

$$E(2 - \tilde{x}^*_t) = 2 - \frac{\epsilon}{1-\mu} + \mu^t\left(\frac{\epsilon}{1-\mu} - 1\right).$$

It is easily to calculate that the limit of the right-hand side of (4.6) as $\mu \to 1$ is $(1+\epsilon)(1-\beta)^{-1} > 1$, and this is equal to $-\infty$ as $\mu \to 1/\beta$. Therefore, by continuity, there exists $b_\epsilon = b_\epsilon(\beta)$ such that

$$\tilde{V}^*_\beta(1) = 1 = V^*_\beta(1) \tag{4.7}$$

for each ϵ. At the same time, the relative stability index in this situation is

$$\begin{aligned} \Delta_\beta(1) &= \left[\sum_{t=0}^{\infty} \beta^t E_1^{\tilde{\mathbf{f}}_*} C(x_t, a_t) - 1\right] / 1 \\ &= \sum_{t=1}^{\infty} \beta^t E(2 - \xi_0\xi_1\ldots\xi_{t-1}) = \frac{2\beta}{1-\beta} - \frac{3\beta}{4-3\beta} > 0. \end{aligned} \tag{4.8}$$

Now, let $\bar{\nu}$ be the Fortét–Mourier probability metric (corresponding to the weak convergence of random variables, see [15]). From the well-known properties of $\bar{\nu}$ and (4.1), (4.2) we readily find that in this example

$$\sup_{x,a} \bar{\nu}(p(\cdot|x,a), \tilde{p}(\cdot|x,a)) \to 0 \quad \text{as} \quad \epsilon \to 0$$

independently of a choice of b_ϵ.

On the other hand, for a fixed discount factor β using b_ϵ to guarantee (4.7) we have equality of the value functions and $\inf_{\epsilon \in (0,1)} \Delta_\beta(1) > 0$. Moreover, from (4.8) $\sup_{\beta \in (0,1)} \Delta_\beta(1) = \infty$ for each $\epsilon \in (0,1)$.

Example 2. The next example presents a stable Markov control model for which all Assumptions of Section 2 are satisfied, and what is important all these assumptions and constants involved are expressed here in rather simple terms of existence of exponential moments. We appealed to this example in [5] for other purposes.

Let $X = [0, \infty)$, $A(x) = A$ for all $x \in X$ with A being a compact subset of some interval $(0, \theta]$ (with $\theta \in A$). Define

$$x_{t+1} = (x_t + a_t \eta_t - \xi_t)^+, \tag{4.9}$$

$$\tilde{x}_{t+1} = (\tilde{x}_t + a_t \tilde{\eta}_t - \tilde{\xi}_t)^+, \quad t = 0, 1, \ldots, \tag{4.10}$$

$x_0 = \tilde{x}_0 = x$ given, where $z^+ = \max(0, z)$; $\{\eta_t\}$, $\{\xi_t\}$, $\{\tilde{\eta}_t\}$ and $\{\tilde{\xi}_t\}$ are sequences of nonnegative i.i.d. r.v.'s such that $\{\eta_t\}$ is independent of $\{\xi_t\}$ and $\{\tilde{\eta}_t\}$ is independent of $\{\tilde{\xi}_t\}$. The equations (4.9) and (4.10) represent controlled versions of random walk on a half-line, which arises in several applied models, for example, in inventory-production or water resources models (see [9]). Other important application of the process given by (4.9) is a model of control of service rates a_t in a single server queueing system of type $GI|GI|1|\infty$. In this case x_t is interpreted as the waiting time of the tth customer, while ξ_t is the interarrival time between the tth and $(t+1)$th customers. The r.v.'s η_t $(t = 0, 1, \ldots)$ describe deviations of real services times from designed values a_t. The variables in the approximating process (4.10) are interpreted in the same way. Comments on applications of such control model can be found, for instance, in [19].

We are going to check the hypotheses of the theorem of the previous section supposing the following Assumption E1 and E2. We write η, ξ, $\tilde{\eta}$, $\tilde{\xi}$ for generic r.v.'s distributed, respectively, as η_0, ξ_0, $\tilde{\eta}_0$, $\tilde{\xi}_0$, and denote

$$\zeta = \theta\eta - \xi, \quad \tilde{\zeta} = \theta\tilde{\eta} - \tilde{\xi}.$$

Assumption E1.

(a) R.v.'s $\eta, \xi, \tilde{\eta}, \tilde{\xi}$ have bounded densities continuous on $[0, \infty)$;

(b)

$$E\zeta < 0, \quad E\tilde{\zeta} < 0; \tag{4.11}$$

and there are positive constants h', $\tilde{h}'$ such that

$$E\exp(h'\zeta) < \infty, \quad E\exp(\tilde{h}'\tilde{\zeta}) < \infty. \tag{4.12}$$

As it was observed in [5] (4.11) and (4.12) yield the existence of $h > 0$ such that

$$\alpha := \max\{E\exp(h\zeta), E\exp(h\tilde{\zeta})\} < 1. \tag{4.13}$$

Assumption E2. The one-stage cost c is a strictly positive measurable function such that, for every $x \in [0, \infty)$, $c(x, \cdot)$ is lower semicontinuous on A, and

$$\sup_{a \in A} c(x, a) \leq (\bar{b})^{1/s} \exp(xh/s),$$

where $s > 1$, and $\bar{b}$ is an arbitrary positive constant.

Remark. For many particular distributions of $\eta, \xi, \tilde{\eta}$ and $\tilde{\xi}$ it is not too hard to evaluate h and α in (4.13). (See [5] for the explicit formulas in the case of exponential distributions.)

Under above Assumptions E1, E2 the work to verify Assumptions $1, 2, 3, 3^*$ in Section 2 was done in [5] provided we choose: $W(x) = \bar{b}\exp(hx)$,

$$\begin{aligned} h_f(x) &= P(x + f(x)\eta - \xi \leq 0), \\ \tilde{h}_f(x) &= P(x + f(x)\tilde{\eta} - \tilde{\xi} \leq 0), \quad x \in [0, \infty); \\ \nu &= \delta_0 \quad \text{(the Dirac distribution).} \end{aligned}$$

Thus the bounds (3.1), (3.9) and Corollary 1 hold for the processes (4.9) and (4.10). Moreover, using (2.11) the total variation distance in (3.9) is easy to calculate in terms of densities of η, ξ, $\tilde{\eta}$, $\tilde{\xi}$. The power $(s-1)/s$ can be chosen as close to 1 as desired if $\sup_{a \in A} c(x, a)$ is dominated by some polynomial. Unfortunately, we are not able in this example to simplify expressions for the constants involved in $B(x)$ in (3.3) (compared with (2.6), (2.7)). The only observation is that $\|\nu\|_W = \bar{b}$. For this reason it seems better using the same arguments as in the theorem of Section 3 to prove for this example the bound as (3.1) replacing in (3.3) $B(x)$ by another constant $D(x)$. The point is to replace the inequalities of Proposition 2 by other estimates of the rate of convergence found in [16] due to specific properties of a particular class of Markov processes on $[0, \infty)$ called "stochastically ordered". Surely, the bounds on the rate of convergence in [16] are much simpler and more tight compared with $\bar{B}$ in (2.8), (2.9). (See Example 4.1 in [16]). The use of them allows in the example considered to get a more tight bound for $\Delta_\beta(x)$ (comparing with (3.1)).

ACKNOWLEDGEMENT

The authors are grateful to the referees for valuable suggestions to improve the text of the paper.

(Received January 23, 1998.)

REFERENCES

[1] E. B. Dynkin and A. A. Yushkevich: Controlled Markov Processes. Springer–Verlag, New York 1979.

[2] E. I. Gordienko: Stability estimates for controlled Markov chains with a minorant. J. Soviet Math. *40* (1988), 481–486. (Translated from the Russian publication of 1985.)

[3] E. I. Gordienko: An estimate of the stability of optimal control of certain stochastic and deterministic systems. J. Soviet Math. *50* (1992), 891–899. (Translated from the Russian publication of 1989.)

[4] E. Gordienko and O. Hernández–Lerma: Average cost Markov control processes with weighted norms: exitence of canonical policies. Appl. Math. *23* (1995), 199–218.

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

[6] E. I. Gordienko, M. E. Isauro-Martínez and R. M. Marcos Carrillo: Estimation of stability in controlled storage systems. Research Report No. 04.0405.I.01.001.97, Dep. de Matemáticas, Universidad Autónoma Metropolitana, México 1997.

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

[8] O. Hernández-Lerma and J. B. Lasserre: Average cost optimal policies for Markov control processes with Borel state space and unbounded costs. Systems Control Lett. *15* (1990), 349–356.

[9] O. Hernández-Lerma and J. B. Lassere: Discrete–time Markov Control Processes. Springer–Verlag, New York 1995.

[10] H. Hinderer: Foundations of Non–Stationary Dynamic Programming with Discrete Time Parameter. (Lecture Notes in Operations Research 33.) Springer–Verlag, New York 1970.

[11] N. V. Kartashov: Inequalities in theorems of ergodicity and stability for Markov chains with common phase space. II. Theory Probab. Appl. *30* (1985), 507–515.

[12] P. R. Kumar and P. Varaiya: Stochastic Systems: Estimation, Identification and Adaptive Control. Prentice–Hall, Englewood Cliffs, N. J. 1986.

[13] S. P. Meyn and R. L. Tweedie: Markov Chains and Stochastic Stability. Springer–Verlag, Berlin 1993.

[14] E. Nummelin: General Irreducible Markov Chains and Non–Negative Operators. Cambridge University Press, Cambridge 1984.

[15] S. T. Rachev: Probability Metrics and the Stability of Stochastic Models. Wiley, New York 1991.

[16] D. J. Scott and R. L. Tweedie: Explicit rates of convergence of stochastically ordered Markov chains. In: Proc. Athens Conference of Applied Probability and Time Series Analysis: Papers in Honour of J. M. Gani and E. J. Hannan (C. C. Heyde, Yu. V. Prohorov, R. Pyke and S. T. Rachev, eds.). Springer–Verlag, New York 1995, pp. 176–191.

[17] N. M. Van Dijk: Perturbation theory for unbounded Markov reward processes with applications to queueing. Adv. in Appl. Probab. *20* (1988), 99–111.

[18] N. M. Van Dijk and M. L. Puterman: Perturbation theory for Markov reward processes with applications to queueing systems. Adv. in Appl. Probab. *20* (1988), 79–98.

[19] R. R. Weber and S. Stidham jr.: Optimal control of service rates in networks of queues. Adv. in Appl. Probab. *19* (1987), 202–218.

[20] W. Whitt: Approximations of dynamic programs I. Math. Oper. Res. *3* (1978), 231–243.

[21] V. M. Zolotarev: On stochastic continuity of queueing systems of type $G|G|1$. Theory Probab. Appl. *21* (1976), 250–269.

Prof. Dr. Evgueni I. Gordienko, Departamento de Matemáticas, Universidad Autónoma Metropolitana, Unidad Iztapalapa, A. P. 55–534, C. P. 09340 México, D. F. México.
e-mail: gord@xanum.uam.mx

Dr. Francisco Salem, Facultad de Ciencias Físico Matemáticas, Benemérita Universidad Autónoma de Puebla, A. P. 1152, C. P. 72000, Puebla, Pue. México.
e-mail: fsalem@fcfm.buap.mx