Jaime Martínez Sánchez; Elena Zaitseva

Note on stability estimation in average Markov control processes

Kybernetika, Vol. 51 (2015), No. 4, 629–638

Persistent URL: http://dml.cz/dmlcz/144471

KYBERNETIKA — VOLUME *51* (2015), NUMBER 4, PAGES 629–638

NOTE ON STABILITY ESTIMATION IN AVERAGE MARKOV CONTROL PROCESSES

Jaime Martínez Sánchez and Elena Zaitseva

We study the stability of average optimal control of general discrete-time Markov processes. Under certain ergodicity and Lipschitz conditions the stability index is bounded by a constant times the Prokhorov distance between distributions of random vectors determinating the "original and the perturbated" control processes.

Keywords: discrete-time Markov control processes, average criterion, stability index, Prokhorov metric

Classification: 90C40, 93E20

1. SETTING OF THE PROBLEM

Let on a Borel space $(X, \mathcal{B}_X)$ two following Markov control processes be given:

$$x_t = F(x_{t-1},\ a_t,\ \xi_t), \qquad t = 1, 2, \dots, \tag{1.1}$$

$$\widetilde{x}_t = F(\widetilde{x}_{t-1},\ \widetilde{a}_t,\ \widetilde{\xi}_t), \qquad t = 1, 2, \dots, \tag{1.2}$$

where $a_t \in A(x_{t-1}) \subset A$, $\widetilde{a}_t \in A(\widetilde{x}_{t-1}) \subset A$ are the controls (actions) forming the *control policies* $\pi = (a_1, a_2, \dots)$, $\widetilde{\pi} = (\widetilde{a}_1, \widetilde{a}_2, \dots)$ (see, e. g., [4, 8] for definitions); $\{\xi_t\}$ and $\{\widetilde{\xi}_t\}$ are sequences of i.i.d. random vectors in a separable metric space (S, r). In what follows the distributions of ξ_1 and $\widetilde{\xi}_1$ are denoted by μ and $\widetilde{\mu}$ respectively.

We will suppose that A is a Borel space with a metric d and that $A(x)$ is compact for every $x \in X$. Denoting

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

(equipped with the metric $\nu := \max\{\rho, d\}$, where ρ is a metric in X), let $c : \mathbb{K} \to \mathbb{R}$ be a given *bounded measurable one-step cost function.*

For any *initial state* $x \in X$ *and control policy* $\pi \in \Pi$ *(*Π *is the set of all control*

DOI: 10.14736/kyb-2015-4-0629

policies, see [4]) the average per unit of time costs are as follows:

$$J(x,\pi) \;:=\; \limsup_{n\to\infty} \frac{1}{n} \sum_{t=1}^{n} E_x^{\pi}\, c(x_{t-1},a_t), \quad x\in X; \tag{1.3}$$

$$\widetilde{J}(x,\pi) \;:=\; \limsup_{n\to\infty} \frac{1}{n} \sum_{t=1}^{n} E_x^{\pi}\, c(\widetilde{x}_{t-1},\widetilde{a}_t), \quad x\in X. \tag{1.4}$$

The following assertion is well-known (see e. g. [8]).

Proposition 1. Under Assumptions 2.1 and 2.2 given in the next section, there exist *stationary optimal policies* f_* and $\widetilde{f}_*$ such that $J(x,f_*)$, $\widetilde{J}(x,\widetilde{f}_*)$ do not depend on $x\in X$; and

$$J_* := J(f_*) = \inf_{\pi\in\Pi} J(x,\pi); \qquad \widetilde{J}_* := \widetilde{J}(\widetilde{f}_*) = \inf_{\pi\in\Pi} \widetilde{J}(x,\pi), \quad x\in X. \tag{1.5}$$

To set the stability estimation problem, first, suppose that process (1.2) is interpreted as an "available approximation" to process (1.1) (i. e. $\widetilde{\mu}$ is an approximation to μ). Second, the policy $\widetilde{f}_*$ (optimal with respect to (1.4)) is applied to control the "original" process (1.1) (instead of "unavailable" optimal policy f_*).

Following the definitions given in [5, 6, 7, 9], we introduce the *stability index*:

$$\Delta := \; J(\widetilde{f}_*) - J(f_*) \;\ge\; 0, \tag{1.6}$$

where J is the average cost defined in (1.3). This definition means that Δ represents an extra cost paid for using $\widetilde{f}_*$ instead of the optimal policy f_*.

Under Lyapunov-like ergodicity hypotheses and certain Lipschitz conditions in [5] it was proved (for the processes with unbounded costs c) that

$$\Delta \;\le\; \overline{K}\,\kappa(\mu,\widetilde{\mu}), \tag{1.7}$$

where $\overline{K}$ is an explicitly calculated constant, and κ is the *Kantorovich metric.* The convergence in κ is equivalent to the weak convergence plus the convergence of first absolute moments (see, e. g. [10]).

Unfortunately, the Lyapunov-like conditions (or "drift conditions"; Assumption 1 in [5]) used there, in the *particular case of a bounded cost c* lead to too strong ergodicity hypotheses (known as "minorization conditions", see e. g. [4, 6]). Particularly, in [4] it is shown that under the minorization conditions the average cost optimization problem can be reduced to optimization of the expected total discounted cost. When the one-stage cost c is bounded, the problem of estimation of the stability index for the discounted total costs is indeed very simple.

The aim of the present paper is making advantage of boundedness of c and using the well-known *ergodicity condition* given in Assumption 2.1 below (see, e. g. [1, 8]) to prove the "stability inequality" as in (1.7), but with the *Lévy–Prokhorov distance* $\ell\pi(\mu,\widetilde{\mu})$ on its right-hand side.

The Lévy–Prokhorov metric on the space of probability distributions on $(S, \mathcal{B}_S)$ ($\mathcal{B}_S$ is the Borel $\sigma-$algebra) is defined as follows (see [2]):

$$\ell\pi(\mu, \eta) := \inf\left\{\epsilon > 0 : \ \mu(B) \le \eta\left(B^{\epsilon}\right) + \epsilon, \ \eta(B) \le \mu\left(B^{\epsilon}\right) + \epsilon; \text{ for all } B \in \mathcal{B}_S\right\},$$

where $B^{\epsilon} := \{s \in S : \ \ r(s, B) < \epsilon\}$.

It is well-known ((see, e. g. [2]) that $\ell\pi$ metrizes the weak convergence of distributions.

2. ASSUMPTIONS AND THE RESULT

We will denote: $k := (x, a) \in \mathbb{K}$. Let for $k = (x, a) \in \mathbb{K}$, $\quad B \in \mathcal{B}_X$ (where $\mathcal{B}_X$ denotes the Borel $\sigma-$algebra),

$$p(B|k) := P(F(x, a, \xi_1) \in B), \ \ \widetilde{p}(B|k) := P(F(x, a, \widetilde{\xi}_1) \in B)$$

be *the transition probabilities* of process (1.1) and (1.2), respectively.

We recall that for probability measures p and p' on $(X, \mathcal{B}_X)$ the total variation norm $\|p - p'\|$ is

$$\sup\left\{\left|\int_X \varphi \, \mathrm{d}p - \int_X \varphi \, \mathrm{d}p'\right| : \ \varphi : \ X \to \mathbb{R} \ \text{ with } \ \|\varphi\|_{\infty} \le 1\right\}.$$

Also in the below assumptions we use the *Hausdorff distance* h between compact subsets of the metric space (A, d):

$$h(C, D) := \max\left\{\sup_{x_1 \in C} \ \inf_{x_2 \in D} \ d(x_1, x_2), \ \sup_{x_2 \in D} \ \inf_{x_1 \in C} \ d(x_2, x_1)\right\}.$$

The ergodicity conditions given in the Assumption 2.1 below were intensively exploited in the literature on Markov control processes with the average cost (for example, they can be found in the paper [1] and the book [8]).

Assumption 2.1. (*Ergodicity conditions*) There exists a number $\lambda < 1$ such that,

$$\sup_{k, k' \in \mathbb{K}} \|p(\cdot|k) - p(\cdot|k')\| \ \le \ 2\lambda;$$

$$\sup_{k, k' \in \mathbb{K}} \|\widetilde{p}(\cdot|k) - \widetilde{p}(\cdot|k')\| \ \le \ 2\lambda;$$

where $\|\cdot\|$ is the total variation norm.

Assumption 2.2. *(Lipschitz conditions; the same as in the paper [5])* There exist finite constants L_0, L, L_1 and L_* such that for all $x, x' \in X$, $k, k' \in \mathbb{K}$, $s, s' \in S$,

(a) $h(A(x), \ A(x')) \le L_0 \rho(x, x')$, ($h$ is the Hausdorff metric);

(b) $|c(k) - c(k')| \ \le \ L_1 \nu(k, k')$;

(c) $\|F(k, \xi_1) - F(k', \xi_1)\| \ \le \ L\nu(k, k')$;

(d) $\rho(F(k,s), F(k,s')) \le L_* r(s,s')$;

(e) for each $x \in X$ and a bounded measurable function $u: X \to \mathbb{R}$, the map $a \to Eu[F(x,a,\widetilde{\xi})]$ is continuous on $A(x)$.

Now we are in position to formulate the main result of the paper.

Theorem 1. Under Assumptions 2.1 and 2.2,

$$\Delta \le K\, \ell\pi(\mu, \widetilde{\mu}), \tag{2.1}$$

where

$$K = 8\left(1 + \frac{2}{1-\lambda}\right)\left[(1+L_0)L_*\left(L_1 + \frac{2bL}{1-\lambda}\right) + \frac{2b}{1-\lambda}\right], \tag{2.2}$$

and in (2.2) the constant K is calculated through the constants involved in Assumptions 2.1 and 2.2. Also $b := \sup_{(x,a)\in\mathbb{K}} |c(x,a)|$.

Remark 1. If "a contractive parameter" $\lambda \uparrow 1$ then, K in (2.2) is of order $M(1-\lambda)^{-2}$. This is better comparing with the constant $\overline{K}$ in (1.7), which is of order $M'(1-\lambda)^{-3}$ (see [5]).

Let us consider the important case when $S = \mathbb{R}^m$ and the "approximating distribution" $\widetilde{\mu}$ (defining process (1.2)) is the empirical distribution

$$\widetilde{\mu} \equiv \widehat{\mu}_n(\xi_1, \dots, \xi_n) := \frac{1}{n}\sum_{i=1}^{n} \delta_{\xi_i} \quad (\delta \text{ is the Dirac measure})$$

obtained from i.i.d. realizations of the random vectors ξ_t in process (1.1). The following result is a direct consequence of inequality (2.1) and the bound on the rate of convergence of $E\ell\pi(\mu, \widehat{\mu}_n)$ obtained in [3].

Corollary 1. Let γ be any fixed number such that $\gamma > \max(2, m)$, and

$$\alpha := \frac{m\gamma}{(\gamma - m)(\gamma - 2)}.$$

Suppose that $E\|\xi_1\|^\alpha < \infty$ and in (2.1) $\widetilde{\mu} = \widehat{\mu}_n$. Then there is $M < \infty$ such that

$$E\Delta \le Mn^{-1/\gamma}, \qquad n = 1, 2, \dots.$$

Remark 2. The following natural question arises: Is Assumption 2.1 essential in order to the stability bound (2.1) be true? In the paper [5] an example of MCP's as in (1.1) and (1.2) with a bounded cost c was constructed in which $\ell\pi(\mu, \widetilde{\mu}) \to 0$, while the stability index Δ in (1.6) keeps to be greater than 1. It is easy to check that for this example (Example 1 in [5]) Assumption 2.1 is not satisfied.

3. SOME EXAMPLES

Example 1. (Borrowed from [1]) Let $X = S = \mathbb{R}$, $A(x) = A, x \in X$, with A being a compact set in $\mathbb{R}$. We set:

$$x_t = H(x_{t-1}, a_t) + G(x_{t-1})\xi_t, \quad t \geq 1;$$

$$\widetilde{x}_t = H(\widetilde{x}_{t-1}, \widetilde{a}_t) + G(\widetilde{x}_{t-1})\widetilde{\xi}_t, \quad t \geq 1;$$

where $\xi_t \sim \mathit{Norm}\ (a, \sigma)$, $\widetilde{\xi}_t \sim \mathit{Norm}\ (\widetilde{a}, \widetilde{\sigma})$; $H : \mathbb{R} \times A \to \mathbb{R}$, $G : \mathbb{R} \to \mathbb{R}$ are bounded Lipschitzian functions and $G(\cdot) > 0$. In [1] it is stated that Assumption 2.1 is satisfaced for these MCP's. It is not difficult to verify Assumption 2.2. Thus, inequality (2.1) can be applied for this example, provided that the cost c is bounded and Lipschitzian. (And if needed the constant K in (2.2) could be evaluated.) In the particular case when $\sigma = \tilde{\sigma} = 1$ and $|a - \widetilde{a}| \leq 2/3$, then in (2.1), $\ell\pi(\mu, \widetilde{\mu}) \leq \sqrt{\frac{3}{2}}\ |a - \widetilde{a}|^{1/2}$.

Example 2. (Water release control model) In this simplest model (see, e. g. [8]) the processes of water stocks are described as follows:

$$x_t = \min\{x_{t-1} - a_t + \xi_t,\ V\}, \quad t \geq 1;$$

$$\widetilde{x}_t = \min\{\widetilde{x}_{t-1} - \widetilde{a}_t + \widetilde{\xi}_t,\ V\}, \quad t \geq 1;$$

where $V < \infty$ is the capacity of a reservoir, $a_t \in A(x_t) := [0, x_t]$, $t \geq 1$ are water consumptions, and $\{\xi_t\}, \{\widetilde{\xi}_t\}$ are i.i.d. nonnegative random variables representing water inflow. Correspondingly, $X = [0, V]$, $S = [0, \infty)$ and the one-step cost is supposed to be Lipschitzian (for instance, $c(x, a) := c_0 a$ where c_0 is the cost of unit of water).

Suppose that random variables ξ_1 and $\widetilde{\xi}_1$ have bounded, Lipschitzian on $[0, V]$ densities g and $\widetilde{g}$, which are strictly positive in some open interval $(0, \Gamma) \supset (0, V]$. Then a positive constant β can found such that $p(B|x, a) \geq \beta\delta_V$; $\widetilde{p}(B|x, a) \geq \beta\delta_V$ for all $(x, a) \in \mathbb{K}$, $B \in \mathcal{B}_{[0,V]}$ (δ_V is the Dirac measure). Therefore (see [8]) Assumption 2.1 holds for this example. It is also easy to verify the fulfillment of Assumption 2.2. For all these reasons inequality (2.1) is applicable. As it could be seen from the proof of the above theorem, using the properties of the Dudley metric [2], the following rough, but simpler inequality holds:

$$\Delta \ \leq \ \frac{K}{2} \int_0^\infty |g(s) - \widetilde{g}(s)|\, \mathrm{d}s.$$

Example 3. (Controlled random walk on a circle) Let $X = S = A = [0, 1)$; $A(x) \equiv A$;

$$x_t = a_t x_{t-1} + \xi_t (\mathrm{mod} 1), \ \ \widetilde{x}_t = \widetilde{a}_t \widetilde{\xi}_{t-1} + \widetilde{\xi}_t (\mathrm{mod} 1), \quad t \geq 1.$$

Then supposing the existence of smooth enough, positive densities of ξ_1 and $\widetilde{\xi}_1$ one can easily check Assumptions 2.1 and 2.2 (provided that c is Lipschitzian).

4. THE PROOF OF THE THEOREM

To prove inequality (2.1) we take advantage of method proposed in the paper [5]. Nevertheless we need to modify this technique. In [5] the combination of certain Lyapunov-like conditions and the results of the paper [11] allows to use contractive properties of the operators related to the average cost optimality equations. It appeared that these operators are contractive with respect the uniform weighted norm in the space of real-valued functions on X. Under Assumption 2.1 we have to use the span seminorm in the space of *bounded functions* and the well-known fact (see, e. g. [8]) about the contractibility with respect of such seminorm.

Let $\mathbb{B}$ denote the space of all measurable bounded functions $u : X \to \mathbb{R}$. For $u \in \mathbb{B}$ we will use the supremum norm $\|u\|_\infty := \sup_{x\in X} |u(x)|$, and the span seminorm $\|u\|_{sp} := \sup_{x\in X} u(x) - \inf_{x\in X} u(x)$. It is clear that $\|u + \beta\|_{sp} = \|u\|_{sp}$ for any $\beta \in \mathbb{R}$, and $\|u\|_{sp} \le 2\|u\|_\infty$. In what follows the random vectors ξ_1 and $\widetilde{\xi}_1$ in (1.1), (1.2) are denoted by ξ and $\widetilde{\xi}$.

Using the notation from (1.5) we define two following operators $T,\ \widetilde{T}:\ \mathbb{B} \to \mathbb{B}:$

$$Tu(x) := \inf_{a\in A(x)} \left\{c(x, a) - J_* \;+\; Eu[F(x, a, \xi)]\right\}, \quad x \in X; \tag{4.1}$$

$$\widetilde{T}u(x) := \inf_{a\in A(x)} \left\{c(x, a) - \widetilde{J}_* \;+\; Eu[F(x, a, \widetilde{\xi})]\right\}, \quad x \in X. \tag{4.2}$$

It is well-known (see, e. g. [8]) that under Assumption 2.1, for every $u, v \in \mathbb{B}$,

$$\|Tu - Tv\|_{sp} \;\le\; \lambda\|u - v\|_{sp}, \qquad \|\widetilde{T}u - \widetilde{T}v\|_{sp} \;\le\; \lambda\|u - v\|_{sp}, \tag{4.3}$$

and therefore there exist functions $h, \widetilde{h} \in \mathbb{B}$ that are solutions of the corresponding optimality equations:

$$h = Th, \qquad \widetilde{h} = \widetilde{T}\,\widetilde{h}. \tag{4.4}$$

Since for any $\beta, \beta' \in \mathbb{R}$, $h + \beta$ and $\widetilde{h} + \beta'$ are also solutions of (4.4) we can choose (in what follows) h and $\widetilde{h}$ in (4.4) in such way that,

$$\|h\|_{sp} = \|h\|_\infty, \qquad \|h - \widetilde{h}\|_{sp} = \|h - \widetilde{h}\|_\infty. \tag{4.5}$$

Then from (4.3) it follows that

$$\|h - \widetilde{h}\|_{sp} \;\le\; 2\|Th - \widetilde{T}h\|_\infty \;+\; \lambda\|h - \widetilde{h}\|_\infty,$$

or in view of (4.1) and (4.2),

$$\|h - \widetilde{h}\|_\infty \;\le\; \frac{2}{1-\lambda}\;\left[|J_* - \widetilde{J}_*| \;+\; Q\right], \tag{4.6}$$

where

$$Q := \sup_{k\in\mathbb{K}} \left|Eh[F(k, \xi)] - Eh[F(k, \widetilde{\xi})]\right|. \tag{4.7}$$

Let $f := \widetilde{f}_*$ denote the stationary policy optimal for process (1.2), and $\{x_t, t \geq 0\}$ be the Markov process induced by the application of the policy f to the control process (1.1).

Using the Markov property of $\{x_t\}$ and the optimality equations (4.4) by simple calculations we obtain that for each $t \geq 1, x \in X$

$$\begin{aligned} E_x^f\, h(x_t) \;&=\; E_x^f\, h(x_{t-1}) \;-\; E_x^\pi\, c\big(x_{t-1}, f(x_{t-1})\big) \;+\; J_* \\ &\quad + E_x^f\, \left[H\big(x_{t-1}, f(x_{t-1})\big) \;-\; \inf_{a \in A(x_{t-1})} H(x_{t-1}, a)\right], \end{aligned} \tag{4.8}$$

where

$$H(x,a) := \; c(x,a) \;+\; Eh[F(x,a,\xi)] - J_*, \quad (x,a) \in \mathbb{K}. \tag{4.9}$$

Summing equalities (4.8) over $t = 1, 2, \ldots, n$, dividing by n and taking lim sup (when $n \to \infty$) we get that

$$J(x,f) \;=\; J_* + \limsup_{n\to\infty} \frac{1}{n} \sum_{t=1}^{n} E_x^f\, \big[H\big(x_{t-1}, f(x_{t-1})\big) \;-\; \inf_{a \in A(x_{t-1})} H(x_{t-1}, a)\big]. \tag{4.10}$$

Similarly to (4.9) let

$$\widetilde{H}(x,a) := c(x,a) + E\widetilde{h}\,[F(x,a,\widetilde{\xi})] \;-\; \widetilde{J}_*, \quad (x,a) \in \mathbb{K}. \tag{4.11}$$

Since for the optimal for process (1.2) policy $\widetilde{f}_*$ the action $a_t = \widetilde{f}_*(x_{t-1})$ yields the infimum of $\widetilde{H}(x_{t-1}, a)$ over $a \in A(x_{t-1})$ we get that in (4.10)

$$I_t := H\big(x_{t-1}, f(x_{t-1})\big) - \inf_{a \in A(x_{t-1})} H(x_{t-1}, a)$$

$$= H\big(x_{t-1}, f(x_{t-1})\big) - \widetilde{H}\big(x_{t-1}, f(x_{t-1})\big) + \inf_{a \in A(x_{t-1})} \widetilde{H}(x_{t-1}, a) - \inf_{a \in A(x_{t-1})} H(x_{t-1}, a).$$

Using this equality and (4.9), (4.11) we find that

$$|I_t| \leq 2|J_* - \tilde{J}_*| + 2 \sup_{a \in A(x_{t-1})} \big|Eh[F(x_{t-1}, a, \xi)] - E\widetilde{h}[F(x_{t-1}, a, \widetilde{\xi})]\big|, \tag{4.12}$$

and the last term on the right-hand side of (4.12) is less (see (4.7)) than $2Q + 2\|h - \widetilde{h}\|_\infty$.

Comparing (1.6), (4.10), (4.6) and the last inequalities we obtain that

$$\Delta \;\leq\; 2\left(1 + \frac{2}{1-\lambda}\right)\; \Big[|J_* - \tilde{J}_*| + Q\Big]. \tag{4.13}$$

Next step is to find a bound for $|J_* - \tilde{J}_*|$ expressed in terms of Q. We choose some sequence $\{\alpha_n\} \subset (0,1)$ such that $\alpha_n \uparrow 1$, and for each n we define the total discounted cost

$$V_{\alpha_n}(x,\pi) := \; E_x^\pi \sum_{t=1}^{\infty} \alpha_n^{t-1}\, c(x_{t-1}, a_t), \quad x \in X,\ \pi \in \Pi$$

for the MCP (1.1). Let also $V^*_{\alpha_n}$ be the corresponding value function, and $h_n(x) := V^*_{\alpha_n}(x) - V^*_{\alpha_n}(z)$, where $z \in X$ is an arbitrary but fixed state.

Applying the standard vanishing discount approach and using arguments quite similar to those given in the last part of the proof in Section 4 in [5], we find that

$$|J_* - \widetilde{J}_*| \le \limsup_{n\to\infty} \sup_{k\in\mathbb{K}} \left|Eh_n[F(k,\xi)] - Eh_n[F(k,\widetilde{\xi})]\right|. \tag{4.14}$$

It is well-known (see, e. g., [1]) that Assumption 2.1 implies that for every stationary policy f' and each $x \in X$, for $t = 1, 2, \dots$

$$\left|E^{f'}_x c(x_{t-1}, f'(x_{t-1})) - \int_X c(x, f'(x)) q_{f'}(\mathrm{d}x)\right| \le 2b\lambda^{t-1}, \tag{4.15}$$

where $q_{f'}$ is the corresponding invariant probability.

From (4.15) it is easy to see that for any $n \ge 1$, $\|h_n\| \le \frac{4b}{1-\lambda}$.

It is well-known (see, e. g. [8]) that for every $x \in X$, $h_n(x) \to h'(x)$, where h' is a solution of the optimality equation $h' = Th'$ (see (4.4)). Then using the bounded convergence theorem and (4.14) we easily obtain that

$$|J_* - \widetilde{J}_*| \le \sup_{k\in\mathbb{K}} |Eh'[F(k,\xi)] - Eh'[F(k,\widetilde{\xi})]|. \tag{4.16}$$

Because the solution of the optimality equation is unique up to adding of an arbitrary constant, we conclude (comparing (4.7), (4.13) and (4.16)) that

$$\Delta \le 4\left(1 + \frac{2}{1-\lambda}\right) Q. \tag{4.17}$$

Now, using (4.4), (4.5) and (4.6) we get that

$$\|h\|_\infty = \|h\|_{sp} \le \|Th - T0\|_{sp} + \|T0\|_{sp}, \quad \text{or}$$

$$\|h\|_\infty \le \frac{1}{1-\lambda} \left\|\inf_{a\in A(\cdot)} c(\cdot, a) - J_*\right\|_{sp}$$

$$= \frac{1}{1-\lambda} \left\|\inf_{a\in A(\cdot)} c(\cdot, a)\right\|_{sp} \le \frac{2b}{1-\lambda}. \tag{4.18}$$

Thus in the definition of Q in (4.7) the function h is bounded by the constant $\frac{2b}{1-\lambda}$.

Next goal is to show that in (4.7), and therefore in inequality (4.17) the functions $h[F(k,\cdot)] : S \to \mathbb{R}$ satisfy the Lipschitz condition with a Lipschitz constant independent of $k \in \mathbb{K}$.

Let us define (see (4.1)):

$$g(k) := c(k) + Eh[F(k,\xi)], \quad k \in \mathbb{K}.$$

Using the definition of the total variation norm $\|p - p'\|$ (see Section 2), Assumption 2.2, (b) and (c), we find that for every $k, k' \in \mathbb{K}$,

$$|g(k) - g(k')| \le \left[L_1 + \frac{2bL}{1-\lambda}\right] \nu(k, k'). \tag{4.19}$$

Now definition (4.1) and the first equation in (4.4) suggest that for each $x \in X$

$$h(x) = \inf_{a \in A(x)} \{g(x, a) - J_*\}. \tag{4.20}$$

In the course of the proof of the main result in [5] the following lemma was proved (which in fact was not formulated as a separate assertion):

Under condition (4.19) and Assumption 2.2, (a) for the function h in (4.20) we have:

$$|h(x) - h(x')| \le (1 + L_0) \left[L_1 + \frac{2bL}{1-\lambda}\right] \rho(x, x') \ \text{ for every } x, x' \in X. \tag{4.21}$$

Now fixing arbitrary $s, s' \in S$ by (4.21) and Assumption 2.2, (e) we get (see (4.7)):

$$\big|h[F(k, s)] - h[F(k, s')]\big| \le \overline{L}\, L_*\, r(s, s'), \tag{4.22}$$

where $\overline{L}$ denotes the constant on the right-hand side of (4.21). Therefore for each $k \in \mathbb{K}$ the function $h[F(k, \cdot)] : S \to \mathbb{R}$ satisfies the Lipschitz condition with the constant $\overline{L}L_*$ and is bounded by $\frac{2b}{1-\lambda}$ (see (4.18)).

Let

$$\partial(\mu, \eta) := \sup\left\{\int_S \varphi \,\mathrm{d}\mu - \int_S \varphi \,\mathrm{d}\eta : \quad \|\varphi\|_\infty + \|\varphi\|_L \le 1\right\} \tag{4.23}$$

be the Dudley metric in the space of probability distributions on $(S, \mathcal{B}_S)$ (see [2] for definition and properties of ∂). Applying (4.18), (4.22) and definition (4.23) to (4.7) and using inequality (4.17) we get that

$$\Delta \le 4\left(1 + \frac{2}{1-\lambda}\right)\left(\overline{L}L_* + \frac{2b}{1-\lambda}\right) \partial(\mu, \widetilde{\mu}).$$

To obtain the desired inequality (2.1) we should use the last inequality and the well-known relation $\partial \le 2\ell\pi$ between the Dudley and Lévy–Prokhorov metrics.

ACKNOWLEDGEMENT

The authors wish to thank referees for valuable suggestions on improvement of the previous version of the paper.

(Received June 11, 2014)

REFERENCES

[1] A. Arapostathis, V. S. Borkar, E. Fernández-Gaucherand, M. K. Ghosh, and S. I. Marcus: Discrete-time controlled Markov processes with average cost criterion: A survey. SIAM J. Control and Optimization *31* (1993), 282–344. DOI:10.1137/0331018

[2] R. M. Dudley: Real Analysis and Probability. Wadsworth and Brooks/Cole, Pacific Grove 1989. DOI:10.1017/s0013091500018265

[3] R. M. Dudley: The speed of mean Glivenko–Cantelly convergence. Ann. Math. Stat. *40* (1969), 40–50. DOI:10.1214/aoms/1177697802

[4] E. B. Dynkin and A. A. Yushkevich: Controlled Markov Processes. Springer-Verlag, New York 1979. DOI:10.1002/zamm.19810610317

[5] E. Gordienko, E. Lemus-Rodríguez, and R. Montes-de-Oca: Average cost Markov control processes: stability with respect to the Kantorovich metric. Math. Methods Oper. Res. *70* (2009), 13–33. DOI:10.1007/s00186-008-0229-6

[6] E. I. Gordienko: Stability estimates for controlled Markov chains with a minorant. J. Soviet. Math. *40* (1988), 481–486. DOI:10.1007/bf01083641

[7] E. I. Gordienko and A. A. Yushkevich: Stability estimates in the problem of average optimal switching of a Markov Chain. Math. Methods Oper. Res. *57* (2003), 345–365.

[8] O. Hernández-Lerma: Adaptive Markov Control Processes. Springer-Verlag, New York 1989. DOI:10.1007/978-1-4419-8714-3

[9] R. Montes-de-Oca and F. Salem-Silva: Estimates for perturbations of average Markov decision processes with a minimal state and upper bounded by stochastically ordered Markov chains. Kybernetika *41* (2005), 757–772.

[10] S. T. Rachev and L. Rüschendorf: Mass Transportation Problem, Vol. II: Applications. Springer, New York 1998.

[11] O. Vega-Amaya: The average cost optimality equation: a fixed point approach. Bol. Soc. Math. Mexicana *9* (2003), 185–195.

Jaime Martínez Sánchez, Department of Mathematics, Universidad Autónoma Metropolitana-Iztapalapa. San Rafael Atlixco 186, Col. Vicentina C.P. 09340, Mexico City. México.

e-mail: j_edum@yahoo.com

Elena Zaitseva, Instituto Tecnológico Autónomo de México. Rio Hondo 1, Col. Progreso Tizapan. Del. Alvaro Obregón, C.P. 01080, Mexico City. México.

e-mail: lenagordi@hotmail.com