N. V. KRYLOV

1964

SovietRxiv

Abstract

Full Text

N. V. KRYLOV

ON THE EXISTENCE OF ε-OPTIMAL HOMOGENEOUS MARKOV STRATEGIES FOR A CONTROLLED CHAIN

(Presented by Academician A. N. Kolmogorov, 7 XII 1963)

I. Definitions. The basic property of a controlled chain.

Let X be an at most countable set, and suppose that with each point $x \in X$ there is associated a collection of probability distributions $\{P_d(x, y)\}$ on $y \in X$, where d takes values in some set $D(x)$, and, in general, for individual $d \in D(x)$,

$$\sum_{y \in X} P_d(x, y) < 1.$$

Definition 1. Any sequence δ of functions $d_n(x_0, \dots, x_n)$, $n \geq 0$, such that $d_n(x_0, \dots, x_n) \in D(x_n)$, will be called a (pure) strategy. A strategy δ will be called a **homogeneous*** **Markov** strategy if, for every $n \geq 0$, $d_n = d(x_n)$. The set of Markov strategies is denoted by Δ_M. The function d_n will be called the **control at time** n.

Each strategy δ naturally induces, in the space X^∞ of sequences $(x_0, x_1, \dots)$, a random process with discrete time $\{x_n\}$, whose probability distribution P_x^δ is such that

$$P_x^\delta\{x_n = a \mid x_0, \dots, x_{n-1}\} = P_{d_{n-1}(x_0, \dots, x_{n-1})}(x_{n-1}, a)\, \delta_x^{x_0},$$

where $\delta_x^{x_0} = 1$ if $x_0 = x$, and $\delta_x^{x_0} = 0$ if $x_0 \neq x$, and $P_n^\delta\{x_0 = x\} = 1$.

Definition 2. The random process $\{x_n\}$ with probability distribution P_x^δ is a **controlled** (by means of δ) **chain** (c.c.).

We note the following basic property of c.c.' s, often used below, analogous to the well-known property for Markov chains.

Let $\xi = f(x_0, x_1, \dots)$ be a random variable, let $\delta = \{d_n, n \geq 0\}$ be some strategy, and let

$$\theta_n(a_0, \dots, a_{n-1})\xi = f(a_0, \dots, a_{n-1}, x_0, x_1, \dots),$$

$$\theta_n(a_0, \dots, a_{n-1})\delta = \delta_n(a_0, \dots, a_{n-1}) = \{d_{n+1+i}(a_0, \dots, a_{n-1}, x_0, \dots, x_i);\ i \geq 0\}.$$

If $a_0, \dots, a_{n-1}$ are regarded as parameters, then $\theta_n(a_0, \dots, a_{n-1})\xi$ will be a random variable, and $\delta_n(a_0, \dots, a_{n-1})$ a certain strategy.

Theorem 1 (basic property of a c.c.). *If ζ is a stopping time of a c.c., τ is a random variable independent of the future, and at least one of the quantities*

$$M_x^\delta \chi_{\tau<\zeta}\xi \quad \text{or} \quad M_x^\delta \chi_{\tau<\zeta}(M_{x_n}^{\theta_n(a_0,\dots,a_{n-1})\delta}\theta_n(a_0, \dots, a_{n-1})\xi)\big|_{a_i=x_i,\ i=0,\dots,n-1,\ n=\tau}$$

exists, then the other also exists and they are equal.

This fact will be written more briefly as

$$M_x^\delta \chi_{\tau<\zeta}\xi = M_x^\delta \chi_{\tau<\zeta} M_{x_\tau}^{\delta_\tau}\theta_\tau\xi.$$

The proof is carried out almost in the same way as the proof of the analogous property for Markov chains (see, for example, ([1]), p. 133).

II. Formulation of the problem and obtained results.

1. One of the main problems in the theory of c.c.' s may be formulated as follows. Let a numerical function $\varphi(x, y, d)$ be given, where $x, y \in X$ and $d \in D(x)$. Put

$$\Delta(x) = \left\{ \delta : \ M_x^\delta \sum_{i=0}^{\infty} \varphi(x_i, x_{i+1}, d_i) \text{ is defined} \right\}.$$

It is required, for $\varepsilon \geq 0$, to find a strategy δ for which, in a fixed

---

* In what follows, homogeneous Markov strategies will simply be called Markov strategies.

at the point x

$$M_x^\delta \sum_{i=0}^{\infty} \varphi(x_i, x_{i+1}, d_i) \geq v(x) - \varepsilon,$$

where

$$v(x) = \sup_{\delta \in \Delta(x)} M_x^\delta \sum_{i=0}^{\infty} \varphi(x_i, x_{i+1}, d_i).$$

If such a strategy exists, then it is called ε-optimal at the point x; 0-optimal strategies are called simply optimal. The present work is devoted mainly to the study of cases in which, among the ε-optimal strategies (for $\varepsilon > 0$ they obviously exist), there exist Markov strategies, i.e., strategies of the class Δ_M. In particular, the following is true.

Theorem 2. *If the space X is finite and the function $0 < v(x) < \infty$, then for every $\varepsilon > 0$ there exists a strategy δ such that $\delta \in \Delta_M$ and for all $x \in X$*

$$M_x^\delta \sum_{i=0}^{\infty} \varphi(x_i, x_{i+1}, d_i) \geq v(x) - \varepsilon.$$

Moreover, if at the point x there exists an optimal strategy, then it can also be chosen from Δ_M. If, in addition, the $D(x)$ are finite for every x, then there exists a strategy from Δ_M that is optimal for all points.

2. Introduce the following notation: τ_x^i is the time of the i-th visit to the point $x \in X$, not counting zero, $\tau_x^0 = 0$;

$$\xi = \sum_{i=0}^{\infty} \varphi(x_i, x_{i+1}, d_i), \qquad \xi_i(x) = \sum_{k=0}^{\tau_x^i - 1} \varphi(x_k, x_{k+1}, d_k), \qquad \xi_j^i(x) = \xi_j(x) - \xi_i(x)$$

for $0 < i < j$; if $\widetilde{\Delta} = \{\delta\}$ is some set of strategies, then

$$|\widetilde{\Delta}(x)| = \widetilde{\Delta} \cap \{\delta : M_x^\delta |\xi| < \infty\}.$$

In what follows an important role is played by

Theorem 3. *If at the point x, $|v(x)| < \infty$ and a) $v(x) > 0$ or b) $P_x^\delta\{\tau_x^1 = \infty\} < 1 - q$, where $q > 0$, for all ε-optimal strategies at the point x for sufficiently small ε, then*

$$v(x) = \sup_{\delta \in |\Delta(x)|} M_x^\delta \xi_1(x) \,/\, P_x^\delta\{\tau_x^1 = \infty\}*; \tag{1}$$

moreover, the upper bound (i.e.) $v(x) = \sup\limits_{\delta \in \Delta(x)} M_x^\delta \xi$ is attained for some strategy if and only if the upper bound in (1) is attained.

Remark 1. In formula (1), $M_x^\delta \xi_1(x)$ and $P_x^\delta\{\tau_x^1 = \infty\}$ do not depend on the values of the controls applied by the strategy δ at the point x at times different from zero.

Lemma 1. *If $\delta \in |\Delta(y)|$, $v(y) < \infty$, then: a) for any n, $M_y^\delta|\xi_n(x)| < \infty$; b)*

$$M_y^\delta \xi \leq \lim_{n\to\infty} M_y^\delta \xi_n(x).$$

Proof of the lemma. a) From Theorem 1 we obtain

$$M_y^\delta \xi = M_y^\delta \left[\xi_n(x) + \chi_{\tau_x^n<\infty} M_x^{\delta_{\tau_x^n}} \xi\right] \leq v(y), \tag{2}$$

i.e.,

$$\eta_n = \xi_n(x) + \chi_{\tau_x^n<\infty} M_x^{\delta_{\tau_x^n}} \xi$$

is absolutely integrable with respect to the measure P_y^δ; it is easy to see, however, that

$$M_x^{\delta_{\tau_x^n}} \xi \leq v(x) \quad (\text{a.s. } P_y^\delta),$$

whence $M_y^\delta \xi_n^-(x) < \infty$. Further, from (2) it follows

$$v(y) = \sup_{\delta\in|\Delta(y)|} M_y^\delta \left[\xi_n(x) + \chi_{\tau_x^n<\infty} v(x)\right] \tag{3}$$

(cf. the Bellman functional equation from dynamic programming theory), and hence $M_y^\delta \xi_n(x) < \infty$, i.e. also $M_y^\delta \xi_n^+(x) < \infty$.

* Here an indeterminacy of the form $\dfrac{0}{0}$ is taken to be $-\infty$.

b) Let F_n be the σ-algebra of events determined by the values of the process up to time τ_x^n. Then $\eta_n = M_y^\delta(\xi \mid F_n)$, and as $n \to \infty$ $\eta_n \to M_y^\delta(\xi \mid F_\infty)$ (a.s. P_y^δ), where $F_\infty = \bigcup_{n=1}^\infty F_n$. However, the random variable ξ is F_∞-measurable and hence $\eta_n \to \xi$ (a.s. P_y^δ). Consequently, $\chi_{\tau_x^n<\infty} M_x^{\delta_{\tau_x^n}} \xi \to 0$ (a.s. P_y^δ), which, together with the inequality $M_x^{\delta_{\tau_x^n}} \xi \leq v(x)$ (a.s. P_y^δ), proves part b).

Proof of Theorem 3. Put in (3) $n = 1$, $x = y$, $\delta \in |\Delta(x)|$. Then

$$v(x) \geq M_x^\delta[\xi_1(x) + \chi_{\tau_x^1<\infty} v(x)],$$

i.e.

$$M_x^\delta \xi_1(x) \le v(x) P_x^\delta\{\tau_x^1 = \infty\},$$

whence

$$v(x) \ge \sup_{\delta\in|\Delta(x)|} M_x^\delta \xi_1(x)/P_x^\delta\{\tau_x^1 = \infty\}.$$

To prove (1) in case a), suppose that in the last relation a strict inequality actually holds. Then there exists $\varepsilon > 0$ such that $v(x) > \varepsilon$ and

$$v(x) - \varepsilon \ge \sup_{\delta\in|\Delta(x)|} M_x^\delta \xi_1(x)/P_x^\delta\{\tau_x^1 = \infty\}. \tag{4}$$

According to Theorem 1, for $\delta \in |\Delta(x)|$ we obtain

$$M_x^\delta \xi_n(x) = \sum_{i=0}^{n-1} M_x^\delta \chi_{\tau_x^i<\infty} \xi_{i+1}^i(x) = \sum_{i=0}^{n-1} M_x^\delta \chi_{\tau_x^i<\infty} M_x^{\delta_{\tau_x^i}} \xi_1(x).$$

By the lemma just proved, $\left|M_x^{\delta_{\tau_x^i}} \xi\right| < \infty$ (a.s. P_x^δ) and, therefore, $\delta_{\tau_x^i} \in |\Delta(x)|$ (a.s. P_x^δ). Then from (4) we obtain

$$M_x^{\delta_{\tau_x^i}} \xi_1(x) \le (v(x) - \varepsilon) P_x^{\delta_{\tau_x^i}}\{\tau_x^1 = \infty\} \quad (\text{a.s. } P_x^\delta).$$

Finally, applying Theorem 1 once more, we find

$$M_x^\delta \xi_n(x) \le (v(x) - \varepsilon) \sum_{i=0}^{n-1} M_x^\delta \chi_{\tau_x^i<\infty} P_x^{\delta_{\tau_x^i}}\{\tau_x^1 = \infty\} \le v(x) - \varepsilon,$$

which, together with the second assertion of the lemma, leads to a contradiction: $v(x) \le v(x) - \varepsilon$.

In case b), relation (1) is derived as follows. Let $\delta(\varepsilon)$ be an ε-optimal strategy for x; then, analogously to (2):

$$v(x) \le M_x^{\delta(\varepsilon)}\left[\xi_1(x) + \chi_{\tau_x^1<\infty} M_x^{\delta(\varepsilon)_{\tau_x^1}} \xi\right] + \varepsilon \le M_x^{\delta(\varepsilon)}\left[\xi_1(x) + \chi_{\tau_x^1<\infty} v(x)\right] + \varepsilon,$$

whence it follows that

$$v(x) \le M_x^{\delta(\varepsilon)} \xi_1(x) + \varepsilon/P_x^{\delta(\varepsilon)}\{\tau_x^1 = \infty\}.$$

Pass in the last expression to the upper limit as $\varepsilon \downarrow 0$:

$$v(x) \le \varlimsup_{\varepsilon\downarrow 0} M_x^{\delta(\varepsilon)} \xi_1(x)/P_x^{\delta(\varepsilon)}\{\tau_x^1 = \infty\} \le \sup_{\delta\in|\Delta(x)|} M_x^\delta \xi_1(x)/P_x^\delta\{\tau_x^1 = \infty\}.$$

Relation (1) is proved.

Suppose now that the l.u.b. $\sup_{\delta\in|\Delta(x)|} M_x^\delta \xi$ is attained at the strategy δ. Then the l.u.b. in (3) is also attained at it, and consequently the l.u.b. in (1) is attained at the strategy δ. Conversely, suppose that the l.u.b. in (1) is attained at the strategy $\delta = \{d_i\}$; then for the strategy $\delta' = \{d_i'\}$, defined by the relations $d_i'(x_0, \dots, x_i) = d_{i-n}(x_n, \dots, x_i)$, where $n = \max\{i,\ k \le i:\ x_k = x\}$, it is easy to obtain

$$M_x^{\delta'}\xi = M_x^\delta \xi_1(x)/P_x^\delta\{\tau_x^1 = \infty\} = v(x).$$

Theorem 3 is proved.

3. The use of Theorem 3 makes it possible to prove a number of useful assertions. We first give some definitions. Let $C = \{x_1, \dots, x_n\}$, $x_i \in X$, $x_i \neq x_j$ for $i \neq j$, and let $A(x)$ be a function on C such that for each x $A(x) \in D(x)$.

Definition 3. We shall call a problem with an $R = (A, C)$-restriction the problem of finding ε-optimal strategies under the assumption that at a point $x \in C$ the set of possible values of the controls is $A(x)$. The $\Delta(x)$ and $v(x)$ corresponding to the problem with an R-restriction will be denoted by $\Delta^R(x)$, $v^R(x)$. Let $x \in C$; put $A(x) = d$. Denote $R_1 = (A, C \cup \{x\})$ and

$$a_x^R(d) = \sup_{\delta\in|\Delta^{R_1}(x)|} M_x^\delta \xi_1(x)/P_x^\delta\{\tau_x^1 = \infty\}.$$

Then formula (1) takes the form

$$v^R(x) = \sup_{d\in D(x)} a_x^R(d). \tag{5}$$

Theorem 4. If, for $R = (A, C)$, at the point $x \notin C$ the conditions of Theorem 2 are satisfied for the problem with an R-restriction, then for $\varepsilon > 0$ there is a control $d \in D(x)$ such that, for $A_1(y) = A(y)$, $y \in C$; $A_1(y) = d$, $y = x$; $R_1 = (A_1, C \cup \{x\})$: a) $v^{R_1}(y) \ge v^R(y) - \varepsilon$ for all $y \in X$; b) if there exists an optimal strategy for the point y in the problem with an R-restriction, then there exists an optimal strategy for y also in the problem with an R_1-restriction; c) if d is such that $v^{R_1}(y) \equiv v^R(y)$, then $v^R(x) = a_x^R(d)$, and conversely.

Proof. We first establish c). From (3) it follows that if $v^R(x) \le v^{R_1}(x) + \varepsilon$, then $v^R(y) \le v^{R_1}(y) + \varepsilon$ everywhere. Now at the point x,

$$v^{R_1}(x) = a_x^R(d),$$

and hence c) is proved.

Assertion a) follows from the following. Suppose that, for d, we have

$$a_x^R(d) \ge v^R(x) - \varepsilon,$$

then everywhere also

$$v^R(y) \le v^{R_1}(y) + \varepsilon.$$

Let us pass to b). Suppose that $M_y^\delta \xi = v^R(y)$ and $\delta \in \Delta^R(y)$. Then, if

$$P_y^\delta\{\tau_x^1 = \infty\} = 1,$$

one may regard $\delta \in \Delta^{R_1}(y)$, and b) follows from a). Let

$$P_y^\delta\{\tau_x^1 = \infty\} < 1.$$

From equality (2) for δ and (4) we obtain that, when $v^R(y) < \infty$, there exists an optimal strategy for x in the problem with an R-restriction. Therefore (see Theorem 3 and Remark 1), for some d in (5) the upper bound is attained, i.e., there exists d such that

$$v^R(x) = v^{R_1}(x)$$

(and hence also $v^R(y) \equiv v^{R_1}(y)$). Thus, for the problem with an R_1-restriction, in (1) the upper bound is attained, which, by Theorem 3, entails the existence of

$$\delta' = \{d_i';\, i \geq 0\} \in \Delta^{R_1}(x)$$

such that

$$M_x^{\delta'} \xi = v^R(x) = v^{R_1}(x).$$

Now let $\delta = \{d_i\}$; construct

$$\bar\delta = \{\bar d_i(x_0, \dots, x_i) = d_i(x_0, \dots, x_i) \quad \text{for } \tau_x^1 > i;$$

$$d'_{i-\tau_x^1}(x_{\tau_x^1}, \dots, x_i) \quad \text{for } \tau_x^1 \leq i;\ i \geq 0\}.$$

Then, by Theorem 1, it is easy to obtain

$$M_y^{\bar\delta} \xi = M_y^\delta[\xi_1(x) + \chi_{\tau_x^1 < \infty} v(x)] = M_y^\delta \xi = v^{R_1}(y) = v^R(y),$$

moreover,

$$\delta \in \Delta^{R_1}(y).$$

If $v^R(y) = \infty$, then in the construction of $\bar\delta$, instead of δ' we take any ε-optimal strategy from $\Delta^{R_1}(x)$, where R_1 is such that

$$v^R(x) \leq v^{R_1}(x) + \varepsilon.$$

For such $\bar\delta$ we have

$$M_y^{\bar\delta} \xi + 2\varepsilon \geq M_y^\delta \xi = \infty,$$

which was required to prove.

Theorem 2 follows easily from Theorem 4 by induction on the number of elements of C.

We point out one more theorem, whose derivation is based on Theorem 4.

Theorem 5. Suppose that $|\varphi(x, y, d)| \leq M < \infty$ and, for all x and $d \in D(x)$,

$$\sum_{y \in x} P_d(x, y) \leq 1 - q, \qquad \text{where } 0 < q \leq 1.$$

Then, for every $\varepsilon > 0$, there exists

$$\delta \in \Delta_M$$

such that

$$M_x^\delta \xi \geq v(x) - \varepsilon$$

for all x. If, moreover, for the point x there exists an optimal strategy, then it too can be chosen in the class Δ_M.

I express my deep gratitude to A. N. Shiryaev for the great assistance he rendered to the author in preparing the work for publication.

Moscow State University
named after M. V. Lomonosov

Received
6 XII 1963

References

1. E. B. Dynkin, *Foundations of the Theory of Markov Processes*, Moscow, 1959.

Note: Figure translations are in progress. See original paper for figures.