

ON ε -OPTIMAL STRATEGIES IN DISCOUNTED MARKOV GAMES

HEINZ-UWE KÜENLE

Sektion Mathematik, Karl-Marx-Universität, Leipzig, DDR

1. Introduction

In this paper we consider the following situation: The state of a dynamic system is periodically controlled. In every control point the first player or decisionmaker chooses an action from an action space \mathfrak{A} , and then the second player or opponent chooses an action from an action space \mathfrak{B} . Then the system moves to a new state from the state space \mathfrak{X} according to the transition probability q . In the costs $k(x, a, b, x')$ the present state is x , actions a and b are chosen by the decisionmaker and by the opponent, respectively, and the following state is x' . The costs are discounted by means of the discount factor α and the system is considered infinitely many times. The aim of the decisionmaker is to minimize the expected total discounted costs, but the purpose of the opponent is to maximize these costs.

We consider the so-called case of perfect information, when the opponent knows the action of the decisionmaker in the present period, and also the usual case of simultaneous action choice, when the opponent does not know this action.

Markov games in which the state space and the action spaces may be some Borel sets are considered by Dietzsch [3] and Küenle [6], [7] in the perfect information case and by Maitra and Parthasarathy [9], Idzik [5], Couwenbergh [2] and Küenle [7] in the other case.

Semi-continuity and compactness conditions are given, which are sufficient for the existence of ε -optimal or optimal strategies and of the value of the game, too.

Since in opposition to the above-mentioned papers universally measurable strategies are allowed in our treatment, our assumptions are partially weaker than the assumptions in these papers. Semi-continuity and compactness assumptions similar to these in our paper are in [7], but in [7] the existence of the value of the game is not considered. On the other hand, we use here some results which were shown in [7].

In the next section we give a mathematical model for the situation which is described above.

Some general optimality conditions are given in the third section. In Sections 4 and 5 we use the results of Section 3 to ensure the existence of optimal or ε -optimal strategies under semi-continuity and compactness conditions and of the value in the case of perfect information and in the case of simultaneous action choice, respectively.

2. The mathematical model

Let M be a tuple $M = ((\mathfrak{X}, X), (\mathfrak{U}, A), (\mathfrak{B}, B), q, \mathfrak{E}, \mathfrak{F}, k, a)$, where

(a) (\mathfrak{X}, X) , (\mathfrak{U}, A) and (\mathfrak{B}, B) are measure spaces, \mathfrak{X} is called *state space*, \mathfrak{U} *action space*, and \mathfrak{B} *action space of the opponent*. It is assumed that \mathfrak{X} , \mathfrak{U} , and \mathfrak{B} are non-empty;

(b) q is a transition probability from $(\mathfrak{X} \times \mathfrak{U} \times \mathfrak{B}, X \otimes A \otimes B)$ to (\mathfrak{X}, X) , the transition law;

(c) $h \in \mathfrak{H}_n := \mathfrak{X} \times \mathfrak{U} \times \mathfrak{B} \times \dots \times \mathfrak{U} \times \mathfrak{B} \times \mathfrak{X}$ ($3n+1$ factors) is called *history at time n* ($n \in N := \{0, 1, 2, \dots\}$). Write

$$H_n := X \otimes A \otimes B \otimes \dots \otimes B \otimes X \quad (3n+1 \text{ factors}).$$

A transition probability π_n from $(\mathfrak{H}_n, \overline{H_n})$ to (\mathfrak{U}, A) is called *decision rule (of the decisionmaker) at time n* and a transition probability ϱ_n from $(\mathfrak{H}_n \times \mathfrak{U}, \overline{H_n} \otimes A)$ to (\mathfrak{B}, B) is called *decision rule (of the opponent) at time n* ($n \in N$). (If Σ is a σ -algebra, then $\overline{\Sigma}$ denotes the σ -algebra of Σ -universally measurable sets.) We denote by E_n and F_n the sets of all admissible decision rules at time n of the decisionmaker and of the opponent, respectively ($E_n \neq \emptyset$, $F_n \neq \emptyset$ for all $n \in N$);

(d) $\mathfrak{E} := \bigtimes_{n \in N} E_n$ and $\mathfrak{F} := \bigtimes_{n \in N} F_n$ are called *admissible strategy sets* and $\Pi \in \mathfrak{E}$ and $P \in \mathfrak{F}$ (*admissible*) *strategies of the opponent* and *of the decisionmaker*, respectively;

(e) k is a real bounded $X \otimes A \otimes B \otimes X$ -measurable function on $\mathfrak{X} \times \mathfrak{U} \times \mathfrak{B} \times \mathfrak{X}$, which is called *cost function*;

(f) $a \in [0, 1)$ is called *discount factor*.

We call the tuple M a *Markov game*. First of all we shall give suitable probability spaces by means of the Ionescu-Tulcea Theorem (see [4] for example).

If $\Pi = (\pi_n)_{n \in N}$ and $P = (p_n)_{n \in N}$ are strategies of the decisionmaker and of opponent, resp., and if $x \in \mathfrak{X}$ is an initial state, then we have the probability space $(\Omega, \Sigma, P_{\Pi P}^x)$, where

$$\begin{aligned}\Omega &:= \mathfrak{X} \times \mathfrak{A} \times \mathfrak{B} \times \mathfrak{X} \times \mathfrak{A} \times \mathfrak{B} \times \dots, \\ \Sigma &:= X \otimes A \otimes B \otimes X \otimes A \otimes B \otimes \dots\end{aligned}$$

and $P_{\Pi P}^x$ is the unique probability measure on Σ with

$$\begin{aligned}& \int_{\Omega} u(x_0, a_0, b_0, \dots, x_{n-1}, a_{n-1}, b_{n-1}, x_n) P_{\Pi P}^x(d\omega) \\ &= \int_{\mathfrak{A}} \pi_0(da_0/x) \int_{\mathfrak{B}} p_0(db_0/x_0, a_0) \int_{\mathfrak{X}} q(dx_1/x_0, a_0, b_0) \dots \\ & \dots \int_{\mathfrak{B}} p_n(db_{n-1}/x_0, a_0, b_0, \dots, x_{n-1}, a_{n-1}) \int_{\mathfrak{X}} q(dx_n/x_{n-1}, a_{n-1}, b_{n-1}) \\ & \quad u(x, a_0, b_0, \dots, x_{n-1}, a_{n-1}, b_{n-1}, x_n) \quad (\omega = (x_0, a_0, b_0, x_1, \dots))\end{aligned}$$

for every real bounded H_n -measurable function u on \mathfrak{S}_n ($n \in N$). Let

$$K(x_0, a_0, b_0, x_1, a_1, b_1, \dots) := \sum_{n=0}^{\infty} \alpha^n k(x_n, a_n, b_n, x_{n+1})$$

and

$$V_{\Pi P}(x) := \int_{\Omega} P_{\Pi P}^x(d\omega) K(\omega).$$

$V_{\Pi P}(x)$ are the *expected total discounted costs* if the strategies Π and P are chosen and the initial state is x .

Let

$$\bar{V}(x) := \inf_{\Pi \in \mathfrak{E}} \sup_{P \in \mathfrak{F}} V_{\Pi P}(x)$$

and

$$\underline{V}(x) := \sup_{P \in \mathfrak{F}} \inf_{\Pi \in \mathfrak{E}} V_{\Pi P}(x)$$

for all $x \in \mathfrak{X}$. \underline{V} is called the *lower value* and \bar{V} the *upper value of the game*. If $\underline{V} = \bar{V} =: V$, then V is called the *value of the game*.

$V_{\Pi}(x) := \sup_{P \in \mathfrak{F}} V_{\Pi P}(x)$ is called *maximal expected total discounted costs*.

In this paper we consider the game from the point of view of the first player.

DEFINITION 2.1. Let $\varepsilon \geq 0$. A strategy $\Pi^* \in \mathfrak{E}$ is called ε -optimal, if $V_{\Pi^*} \leq \bar{V} + \varepsilon$. It is called *optimal* if it is 0-optimal. ■

This definition of optimality is used in [6], [7] and partially in [3].

The concept of optimality, which is used in [2], [5], [9] and in many other papers, is the following one (see also [12]).

DEFINITION 2.2. A strategy $\Pi^* \in \mathfrak{B}$ is called *strong ε -optimal* or *strong optimal* if it is ε -optimal or optimal, resp., and the game has a value. ■

We are interested in finding optimal or ε -optimal strategies in some classes of simple strategies as Markov or deterministic strategies. Let us put $\mathfrak{G}_n = \mathfrak{X} \times \mathfrak{U} \times \mathfrak{B} \times \dots \times \mathfrak{X} \times \mathfrak{U} \times \mathfrak{B}$ ($3n$ factors) for $n \in \mathbb{N} \setminus \{0\}$ to simplify our notation. " $g \in \mathfrak{G}_0$ " means that g must be omitted in the formula.

DEFINITION 2.3. A decision rule $\pi_n \in E_n$ is called *Markov* if there exists a transition probability π from $(\mathfrak{X}, \bar{\mathfrak{X}})$ to $(\mathfrak{U}, \mathfrak{A})$ such that $\pi_n(\cdot/g, x) = \pi(\cdot/x)$ for all $g \in \mathfrak{G}_n$, $x \in \mathfrak{X}$, $n \in \mathbb{N}$.

It is called *deterministic* if a $(\bar{\mathfrak{H}}_n - \mathfrak{A}$ -measurable) function e from \mathfrak{H}_n to \mathfrak{U} exists with $\pi_n(\{e(h)\}/h) = 1$ for all $h \in \mathfrak{H}_n$. Such a deterministic decision rule we denote by δ_e , too.

A decision rule $\varrho_n \in F_n$ is called *Markov* if a transition probability ϱ from $(\mathfrak{X} \times \mathfrak{U}, \overline{\mathfrak{X} \otimes \mathfrak{A}})$ to $(\mathfrak{B}, \mathfrak{B})$ exists with $\varrho_n(\cdot/g, a, b) = \varrho(\cdot/x, a)$ for all $g \in \mathfrak{G}_n$, $x \in \mathfrak{X}$, $a \in \mathfrak{U}$.

It is called *deterministic* if a $(\overline{\mathfrak{H}_n \otimes \mathfrak{A}} - \mathfrak{B}$ -measurable) function f from $\mathfrak{H}_n \times \mathfrak{U}$ to \mathfrak{B} exists with $\varrho_n(\{f(h, a)\}/h, a) = 1$ for all $(h, a) \in \mathfrak{H}_n \times \mathfrak{U}$. In the last case we denote ϱ_n by δ_f , too.

A strategy $\Pi = (\pi_n) \in \mathfrak{E}$ or $P = (\varrho_n) \in \mathfrak{F}$ is called *Markov* or *deterministic* if all decision rules π_n or ϱ_n , $n \in \mathbb{N}$, are Markov or deterministic, respectively. ■

Remark 2.4. In this paper there is often no difference between a function v on a product space $Y \times Z$ and a function v' on Z if $v(y, z) = v'(z)$ for all $(y, z) \in Y \times Z$. ■

That means, for example, that we put $\pi_n = \pi$ if π_n is a Markov decision rule and the connection between π and π_n is as in Definition 2.3.

π^∞ or ϱ^∞ denotes a *stationary strategy*, that is, a Markov strategy (π_n) or (ϱ_n) with $\pi_n = \pi$ or $\varrho_n = \varrho$ for all $n \in \mathbb{N}$, resp.

We now consider some operators. Let \underline{q} , \underline{T} , $\underline{\pi}_n$ and $\underline{\varrho}_n$ be given by

$$\begin{aligned} \underline{q}u(x_0, a_0, b_0, \dots, x_n, a, b) &:= \int_{\bar{\mathfrak{X}}} \underline{q}(dx/x_n, a, b) u(x_0, a_0, b_0, \dots, \\ &\quad \dots, x_n, a, b, x), \end{aligned}$$

$$\underline{T}u := \underline{q}(k + a \cdot u),$$

$$\pi_n u'(h) = \int_{\mathfrak{A}} \pi_n(da/h) u'(h, a),$$

$$\varrho_n u''(h, a) = \int_{\mathfrak{B}} \varrho_n(db/h, a) u''(h, a, b)$$

for all $n \in N$, $h = (x_0, a_0, b_0, \dots, x_n) \in \mathfrak{H}_n$, $\pi_n \in E_n$, $\varrho_n \in F_n$, $a \in \mathfrak{A}$, $b \in \mathfrak{B}$, $x \in \mathfrak{X}$ and all real bounded measurable functions u, u', u'' on \mathfrak{H}_{n+1} , $\mathfrak{H}_n \times \mathfrak{A}$, $\mathfrak{H}_n \times \mathfrak{A} \times \mathfrak{B}$, resp.

The proof of the following lemma is easy and here omitted. We denote by \mathfrak{U} the set of all real bounded \mathbf{X} -measurable functions on \mathfrak{X} .

LEMMA 2.5. *Let $\Pi = (\pi_n) \in \mathfrak{E}$ and $P = (\varrho_n) \in \mathfrak{F}$ and $u \in \mathfrak{U}$. Then*

$$V_{\Pi P}(x) = \lim_{n \rightarrow \infty} \pi_0 \varrho_0 T \dots \pi_n \varrho_n T u(x)$$

holds for every $x \in \mathfrak{X}$ and the convergence is uniform. ■

3. Optimality conditions

The next definition is helpful in our treatment.

DEFINITION 3.1. Let $u \in \mathfrak{U}$ and $\varepsilon \geq 0$. A decision rule $\pi_n \in E_n$ is called (ε, u) -optimal if $\pi_n \varrho_n T u \leq u + \varepsilon$ for all $\varrho_n \in F_n$ (that means, in conformity with Remark 2.4,

$$\begin{aligned} \int_{\mathfrak{A}} \pi_n(da/x_0, a_0, b_0, \dots, x_n) \int_{\mathfrak{B}} \varrho_n(db/x_0, a_0, b_0, \dots, x_n, a) \\ + \int_{\mathfrak{X}} q(dx/x_n, a, b) (k(x_n, a, b, x) + \alpha \cdot u(x)) \leq u(x_n) + \varepsilon \end{aligned}$$

for all $(x_0, a_0, b_0, \dots, x_n) \in \mathfrak{H}_n$).

A decision rule $\bar{\varrho}_n \in F_n$ is called (ε, u) -optimal if

$$\pi_n \bar{\varrho}_n T u \geq u - \varepsilon$$

for all $\pi_n \in E_n$. ■

We get the following lemma.

LEMMA 3.2. *If $\bar{\Pi} = (\bar{\pi}_n)$ is an admissible strategy such that for each $n \in N$ $\bar{\pi}_n$ is (ε, u) -optimal for some $u \in \mathfrak{U}$, then*

$$V_{\bar{\Pi}} \leq u + \frac{\varepsilon}{1 - \alpha}.$$

Proof. From the monotonicity of the operators $\underline{\pi}_n$, \underline{q}_n and \underline{T} it follows successively

$$\underline{\pi}_k \underline{q}_k \underline{T} \dots \underline{\pi}_n \underline{q}_n \underline{T} u \leq u + \varepsilon \cdot \sum_{i=0}^{n-k} \alpha^i$$

and from Lemma 2.6, $V_{\bar{\Pi}P} \leq u + \frac{\varepsilon}{1-\alpha}$ for all $P \in \mathfrak{P}$. ■

The proof of the next lemma is similar to the previous one.

LEMMA 3.3. *If $\bar{P} = (\bar{q}_n)$ is an admissible strategy of the opponent and if every \bar{q}_n is (ε, u) -optimal for some $u \in \mathfrak{U}$, then we have*

$$u \leq V_{\bar{\Pi}\bar{P}} + \frac{\varepsilon}{1-\alpha} \quad \text{for every } \Pi \in \mathfrak{E}.$$

We get now the following theorem, which gives sufficient conditions for existence of ε -optimal strategies and of the value of a Markov game M .

THEOREM 3.4. (a) *If for some $\bar{u} \in \mathfrak{U}$ for every $n \in N$ an (ε, \bar{u}) -optimal decision rule $\bar{\pi}_n \in \bar{E}_n$ and an (ε', \bar{u}) -optimal decision rule $\bar{q}_n \in F_n$ exist, then $\bar{\Pi} := (\bar{\pi}_n)$ is $\frac{\varepsilon + \varepsilon'}{1-\alpha}$ -optimal.*

(b) *If the assumptions in (a) are fulfilled for every $\varepsilon > 0$ and $\varepsilon' > 0$, then the Markov game M has the value \bar{u} .*

Proof. (a): From Lemmas 3.2 and 3.3 it follows that

$$V_{\bar{\Pi}} \leq \bar{u} + \frac{\varepsilon + \varepsilon'}{1-\alpha} \leq V_{\Pi} + \frac{\varepsilon + \varepsilon'}{1-\alpha} \quad \text{for each } \Pi \in \mathfrak{E}.$$

(b): We get from Lemmas 3.2 and 3.3

$$\begin{aligned} \inf_{\Pi \in \mathfrak{E}} \sup_{P \in \mathfrak{P}} V_{\Pi P} - \frac{\varepsilon}{1-\alpha} &\leq V_{\bar{\Pi}} - \frac{\varepsilon}{1-\alpha} \leq \bar{u} \leq \inf_{\Pi \in \mathfrak{E}} V_{\Pi \bar{P}} + \frac{\varepsilon'}{1-\alpha} \\ &\leq \sup_{P \in \mathfrak{P}} \inf_{\Pi \in \mathfrak{E}} V_{\Pi P} + \frac{\varepsilon'}{1-\alpha} \quad \text{for all } \varepsilon, \varepsilon' > 0. \quad \blacksquare \end{aligned}$$

4. Markov games with perfect information

We assume in the last two sections that a map \mathcal{A} from \mathfrak{X} to A and a map \mathcal{B} from $\mathfrak{X} \times \mathfrak{U}$ to B are given with the properties

$$\hat{A} := \{(x, a) : a \in \mathcal{A}(x)\} \in X \otimes A$$

and

$$\hat{B} := \{(x, a, b) : b \in \mathcal{B}(x, a)\} \in X \otimes A \otimes B.$$

We call decision rules π_n or ϱ_n *\mathcal{A} -admissible* or *\mathcal{B} -admissible* if for all $g \in \mathfrak{G}_n$, $x \in \mathfrak{X}$ and $a \in \mathfrak{A}$, $\pi_n(\mathcal{A}(x)/g, x) = 1$ or $\varrho_n(\mathcal{B}(x, a)/g, x, a) = 1$, resp.

We say that M is a *Markov game with perfect information* if E_n is a subset of the set of all \mathcal{A} -admissible decision rules and contains all \mathcal{A} -admissible deterministic decision rules at time n and if F_n is the set of all \mathcal{B} -admissible decision rules of the opponent at time n . This can be interpreted in the following way: At time n both players know all previous states and actions. First the decisionmaker chooses an action from \mathfrak{A} and then the opponent from \mathfrak{B} , knowing decisionmaker's last action, too.

First of all in this section we give some definitions and lemmas which we use later. Then we will consider some semi-continuity and compactness conditions, which are sufficient for the existence of strong ϵ -optimal strategies.

Let \underline{U} and \underline{L} be operators which are defined by

$$\underline{L}u'(g, x) = \inf_{\pi \in E_n} \pi u'(g, x)$$

for all $g \in \mathfrak{G}_n$, $x \in \mathfrak{X}$ and all real bounded $\overline{H_n \otimes A}$ -measurable functions u' ($n \in N$) and

$$\underline{U}u''(g, x, a) = \sup_{\varrho \in F_n} \varrho u''(g, x, a)$$

for all $g \in \mathfrak{G}_n$, $x \in \mathfrak{X}$, $a \in \mathfrak{A}$ and all real bounded $\overline{H_n \otimes A \otimes B}$ -measurable functions u'' ($n \in N$).

In the case of perfect information we have then

$$\underline{L}u'(g, x) = \inf_{a \in \mathcal{A}(x)} u'(g, x, a)$$

and

$$\underline{U}u''(g, x, a) = \sup_{b \in \mathcal{B}(x, a)} u'(g, x, a, b) \quad \text{for all } g \in \mathfrak{G}_n, x \in \mathfrak{X}, a \in \mathfrak{A}, \\ n \in N.$$

We can use these properties to extend the domain of \underline{L} and \underline{U} to all real functions on $\mathfrak{H}_n \times \mathfrak{A}$ and $\mathfrak{H}_n \times \mathfrak{A} \times \mathfrak{B}$, respectively. Let \mathfrak{Y} be a Borel set in a Polish space. The σ -algebra of all Borel subsets of \mathfrak{Y} is denoted by $\sigma_{\mathfrak{Y}}$ and the σ -algebra of all $\sigma_{\mathfrak{Y}}$ -universally-measurable subsets of \mathfrak{Y} by $\bar{\sigma}_{\mathfrak{Y}}$. For $Y \in \sigma_{\mathfrak{Y}}$ we denote by \mathfrak{M}_Y the set of all probability measures on $\sigma_{\mathfrak{Y}}$ which are concentrated on Y . Since for every probability measure on $\sigma_{\mathfrak{Y}}$ there exists a unique extension to a probability measure on $\bar{\sigma}_{\mathfrak{Y}}$, we can identify \mathfrak{M}_Y with the subset of all probability measures on $\bar{\sigma}_{\mathfrak{Y}}$ which are concentrated on Y . We consider on \mathfrak{M} the weak topology. Then $\mathfrak{M}_{\mathfrak{Y}}$ is a Borel set in a Polish space, the corresponding Borel- σ -algebra is de-

noted by $\sigma_{\mathfrak{Y}}^*$, the σ -algebra of all $\sigma_{\mathfrak{Y}}^*$ -universally-measurable sets is denoted by $\bar{\sigma}_{\mathfrak{Y}}^*$.

DEFINITION 4.1. Let \mathfrak{Y} and \mathfrak{Z} be Borel sets in Polish spaces. A *transition probability* p from \mathfrak{Y} to $\sigma_{\mathfrak{Z}}$ is called *Borel* if

$$\{y: p(\cdot/y) \in Q\} \in \sigma_{\mathfrak{Y}} \quad \text{for all } Q \in \sigma_{\mathfrak{Z}}^*.$$

A strategy $\Pi = (\pi_n)_{n \in N}$ or $P = (p_n)_{n \in N}$ is called *Borel* if all π_n , $n \in N$, or p_n , $n \in N$, are Borel, respectively. ■

DEFINITION 4.2. A Markov game M is called *Borel* if the following assumptions hold: \mathfrak{X} , \mathfrak{A} and \mathfrak{B} are Borel sets in Polish spaces, $X = \bar{\sigma}_{\mathfrak{X}}$, $A = \bar{\sigma}_{\mathfrak{A}}$, $B = \bar{\sigma}_{\mathfrak{B}}$, q is a Borel transition probability, k is a Borel function. ■

In this section it is assumed that M is Borel and $\hat{A} \in \sigma_{\mathfrak{X} \times \mathfrak{A}}$, $\hat{B} \in \sigma_{\mathfrak{X} \times \mathfrak{A} \times \mathfrak{B}}$. A proof of the following selection theorem can be found in [1].

LEMMA 4.3. Let \mathfrak{Y} and \mathfrak{Z} be Borel sets in Polish spaces, \mathcal{Z} is a map from \mathfrak{Y} to $\sigma_{\mathfrak{Z}}$ with the property $\{(y, z): z \in \mathcal{Z}(y)\} \in \sigma_{\mathfrak{Y} \times \mathfrak{Z}}$ and r is a real bounded $\sigma_{\mathfrak{Y} \times \mathfrak{Z}}$ -measurable function. Then

(a) For every $\varepsilon > 0$ there exists a $\bar{\sigma}_{\mathfrak{Y}} - \bar{\sigma}_{\mathfrak{Z}}$ -measurable function s from \mathfrak{Y} to \mathfrak{Z} such that $s(y) \in \mathcal{Z}(y)$ and

$$r(y, s(y)) \leq \inf_{z \in \mathcal{Z}(y)} r(y, z) + \varepsilon \quad \text{for all } y \in \mathfrak{Y};$$

(b) If $r(y, \cdot)$ is lower semi-continuous on $\mathcal{Z}(y)$ and $\mathcal{Z}(y)$ is compact for every $y \in \mathfrak{Y}$, then there exists a $\sigma_{\mathfrak{Y}} - \sigma_{\mathfrak{Z}}$ -measurable function s from \mathfrak{Y} to \mathfrak{Z} such that $s(y) \in \mathcal{Z}(y)$ and

$$r(y, s(y)) = \inf_{z \in \mathcal{Z}(y)} r(y, z). \quad \blacksquare$$

Since the function s in part (a) of this theorem need not to be a Borel one, the use of universally measurable strategies in this paper is very favourable.

In this paper we use the following definition of semi-continuity of set valued functions, which can be found in [6].

DEFINITION 4.4. Let \mathfrak{Y} and \mathfrak{Z} be Borel sets in Polish spaces and \mathcal{Z} a function from \mathfrak{Y} to $\sigma_{\mathfrak{Z}}$. Let d be a metric on \mathfrak{Z} which determines the topology on \mathfrak{Z} . We put

$$\mathcal{Z}^*(y) = \{z: z \in \mathfrak{Z}, d(z, z') \leq \varepsilon \text{ for some } z' \in \mathcal{Z}(y)\}.$$

\mathcal{Z} is called *upper semi-continuous* (u.s.c.) or *lower semi-continuous* (l.s.c.) if for every y and every sequence (y_n) , $y, y_n \in \mathfrak{Y}$ and $\lim_{n \rightarrow \infty} y_n = y$,

$$\bigcap_{\varepsilon > 0} \bigcap_{n=0}^{\infty} \bigcup_{k=n}^{\infty} \mathcal{Z}^*(y_n) \leq \mathcal{Z}(y)$$

or

$$\bigcap_{\varepsilon > 0} \bigcup_{n=0}^{\infty} \bigcap_{k=n}^{\infty} \mathcal{X}^{\varepsilon}(y_n) \geq \mathcal{X}(y),$$

respectively. ■

We remark without proof that these definitions of semi-continuity are the same as in [8] if \mathcal{Z} and $\mathcal{X}(y)$ for all $y \in \mathcal{Y}$ are compact.

Now it is possible to formulate the following existence theorems.

THEOREM 4.5. *If*

- (a) $\mathcal{A}(\cdot)$ is u.s.c., $\mathcal{A}(x)$, $x \in \mathcal{X}$, and \mathcal{A} are compact,
- (b) $\mathcal{B}(\cdot)$ is l.s.c. on \hat{A} ,
- (c) $\underline{q}k$ is l.s.c. on \hat{B} ,
- (d) $\int_{\mathcal{X}} u(x) \underline{q}(dx|\cdot)$ is continuous for every real bounded continuous u on \mathcal{X} ,

then there exists an optimal deterministic stationary Borel strategy and the value of the game is the unique bounded l.s.c. function v on \mathcal{X} satisfying the identity $v = \underline{LUT}v$.

Proof. In [7] it is shown that \underline{LUT} has a unique fixpoint v in the set of all bounded l.s.c. functions on \mathcal{X} and that $\underline{UT}v$ is l.s.c. and hence Borel on A . We get by Lemma 4.3 (a) for every given $\varepsilon > 0$ the existence of a $\sigma_{\mathcal{X} \times \mathcal{A}} - \sigma_{\mathcal{B}}$ -measurable function f with $\underline{\delta}_f T v \geq \underline{UT}v - \varepsilon$ and the existence of a $\sigma_{\mathcal{X}} - \sigma_{\mathcal{A}}$ -measurable function e with $\underline{\delta}_e \underline{UT}v = \underline{LUT}v$.

By Theorem 3.4, it follows that δ_e^∞ is an optimal strategy and that v is the value of the game M . ■

The assumptions in the following theorem are obtained essentially by changing "l.s.c." into "u.s.c.".

THEOREM 4.6. *If*

- (a) $\mathcal{A}(\cdot)$ is l.s.c.,
- (b) $\mathcal{B}(\cdot)$ is u.s.c. on \hat{A} , $\mathcal{B}(x, a)$, $(x, a) \in \mathcal{X} \times \mathcal{A}$, and \mathcal{B} are compact,
- (c) $\underline{q}k$ is u.s.c. on \hat{B} ,
- (d) $\int_{\mathcal{X}} u(x) \underline{q}(dx|\cdot)$ is continuous for every real bounded continuous u on \mathcal{X} ,

then for every $\varepsilon > 0$ there exists an ε -optimal deterministic stationary strategy and the value of the game is the unique u.s.c. function v on \mathcal{X} satisfying the identity $v = \underline{LUT}v$.

Proof. It is shown also in [7] that \underline{LUT} has a unique fixpoint v in the set of all bounded u.s.c. functions and that $\underline{UT}v$ is u.s.c. on \hat{A} . The rest of the proof is similar to the proof of Theorem 4.5. ■

At the end of this section we give the following existence theorem.

THEOREM 4.7. *If*

- (a) $\mathcal{A}(x)$ is compact for every $x \in \mathfrak{X}$,
- (b) $\mathcal{B}(x, \cdot)$ is l.s.c. on $\mathcal{A}(x)$ for every $x \in \mathfrak{X}$, $\mathcal{B}(x, a)$ is compact for every $(x, a) \in \hat{A}$,
- (c) $\underline{q}k(x, \cdot, \cdot)$ is l.s.c. on $\{(a, b) : a \in \mathcal{A}(x), b \in \mathcal{B}(x, a)\}$ for every $x \in \mathfrak{X}$, $\underline{q}k(x, a, \cdot)$ is continuous on $\mathcal{B}(x, a)$ for all $(x, a) \in \hat{A}$,
- (d) $\int_{\mathfrak{X}} u(x) \underline{q}(dx/\cdot)$ is continuous for every real bounded $\sigma_{\mathfrak{X}}$ -measurable function on \mathfrak{X} ,

then there exists an optimal deterministic stationary Borel strategy and the value of the game is the unique bounded $\sigma_{\mathfrak{X}}$ -measurable function v on \mathfrak{X} with $v = \underline{LUT}v$.

Proof. The existence of a bounded Borel fixpoint v of \underline{LUT} is shown again in [7] as well as the Borel measurability of $\underline{UT}v$ and the fact that $\underline{UT}v(x, \cdot)$ is l.s.c. for all $x \in \mathfrak{X}$. The rest follows from the proof of Theorem 4.5. ■

5. Markov games with simultaneous action choice

A Markov game M is called a *Markov game with simultaneous action choice* if $\mathcal{B}(x, a)$ is independent of a , and if E_n is the set of all \mathcal{A} -admissible decision rules π at time n and F_n is the set of all \mathcal{B} -admissible decision rules ϱ of the opponent at time n where $\varrho(\cdot/g, x, a)$ is independent of a for all $g \in \mathfrak{G}_n$, $x \in \mathfrak{X}$, $n \in \mathbb{N}$. (We write here $\varrho(\cdot/g, x)$ for $\varrho(\cdot/g, x, a)$ and $\mathcal{B}(x)$ for $\mathcal{B}(x, a)$). In opposition to a Markov game with perfect information, in this case the opponent chooses his action without knowledge of the last action choice of the decisionmaker.

In this section we always assume that M is a Markov game with simultaneous action choice.

Let us consider two other Markov games, which are connected with a Markov game with simultaneous action choice. The first one is got by change of the positions of both players. Formally this is given in the following definition.

DEFINITION 5.1. A Markov game M^* is called *dual to* M if

$$M^* = ((\mathfrak{X}, X), (\mathcal{B}, B), (\mathfrak{A}, A), q^*, \mathfrak{E}^*, \mathfrak{F}^*, k^*, a)$$

with

$$q^*(\cdot/x, b, a) := q(\cdot/x, a, b), \quad k^*(x, b, a, x') := -k(x, a, b, x')$$

for all $(x, a, b, x') \in \mathfrak{X} \times \mathfrak{A} \times \mathfrak{B} \times \mathfrak{X}$ and $\mathfrak{E}^* := \bigcap_{n=0}^{\infty} E_n^*$, $\mathfrak{F}^* := \bigcap_{n=0}^{\infty} F_n^*$, where E_n^* is the set of all decision rules ϱ^* of the first player in M^* with

$$\varrho^*(\cdot/x_0, b_0, a_0, \dots, x_{n-1}, b_{n-1}, a_{n-1}, x_n) := \varrho(\cdot/x_0, a_0, b_0, \dots, x_{n-1}, a_{n-1}, b_{n-1}, x_n)$$

for some $\varrho \in F_n$ and F_n^* is the set of all decision rules π^* of the second player in M^* with

$$\pi^*(\cdot/x_0, b_0, a_0, \dots, x_{n-1}, b_{n-1}, a_{n-1}, x_n) := \pi(\cdot/x_0, a_0, b_0, \dots, x_{n-1}, a_{n-1}, b_{n-1}, x_n)$$

for some $\pi \in E_n$ and for all $(x_0, a_0, b_0, \dots, x_{n-1}, a_{n-1}, b_{n-1}, x_n) \in \mathfrak{H}_n$, $n \in \mathbb{N}$. ■

Obviously, M^* is a Markov game with simultaneous action choice, too. We say that the second player in M has an ϵ -optimal strategy if the first player in M^* has such a strategy. We assume in this section that M is a Borel Markov game.

Now we define a Markov game with perfect information, connected with one with simultaneous action choice.

DEFINITION 5.2. A tuple M' is called *derivated game of the Markov game M* if

$$M' = ((\mathfrak{X}, \sigma_{\mathfrak{X}}), (\mathfrak{W}_{\mathfrak{A}}, \sigma_{\mathfrak{A}}^*), (\mathfrak{B}, \sigma_{\mathfrak{B}}), q', \mathfrak{E}', \mathfrak{F}', k', a)$$

with

$$q'(\cdot/x, w, b) := \int_{\mathfrak{A}} q(\cdot/x, a, b) w(da)$$

and

$$k'(x, w, b) := \int_{\mathfrak{A}} \int_{\mathfrak{X}} k(x, a, b, x') q(dx'/x, a, b) w(da) \quad \text{for all } x \in \mathfrak{X}.$$

$$w \in \mathfrak{W}_{\mathfrak{A}}, \quad b \in \mathfrak{B}, \quad \mathfrak{E}' := \bigcap_{n=0}^{\infty} E'_n, \quad \mathfrak{F}' := \bigcap_{n=0}^{\infty} F'_n,$$

where F'_n is the set of all decision rules ϱ' of the opponent at time n with

$$\varrho'(\mathcal{B}(x)/g, x) = 1 \quad \text{for all } g \in \mathfrak{G}'_n := (\mathfrak{X} \times \mathfrak{W}_{\mathfrak{A}} \times \mathfrak{B} \times \dots \times \mathfrak{X} \times \mathfrak{W}_{\mathfrak{A}} \times \mathfrak{B})$$

($3n$ factors) and E'_n is the set of all decision rules δ_n of the decisionmaker in M' with $\pi \in E'_n$. ■

It is well known that q' and k' have the corresponding measurability properties such that M' is a Borel Markov game.

M^{**} means the derivated game of the Markov game M^* . \underline{L}^i , \underline{U}^i and \underline{T}^i mean the operators \underline{L} , \underline{U} and \underline{T} , connected with the Markov game M^i ($i \in \{*, ', **\}$). The following lemma gives a relation between M' and M^{**} .

LEMMA 5.3. *If u is a real bounded $\bar{\sigma}_x$ -measurable function, then we have $\underline{L}' \underline{U}' \underline{T}' u = -\underline{L}^{**} \underline{U}^{**} \underline{T}^{**}(-u)$.*

Proof. We have for every $x \in \mathfrak{X}$,

$$\begin{aligned} \underline{L}' \underline{U}' \underline{T}' u(x) &= \inf_{\mu \in \mathfrak{M}_{\mathcal{A}}(x)} \sup_{b \in \mathcal{B}(x)} \int_{\mathfrak{A}} \underline{T}u(x, a, b) \mu(da) \\ &= \inf_{\mu \in \mathfrak{M}_{\mathcal{A}}(x)} \sup_{v \in \mathfrak{M}_{\mathcal{B}}(x)} \int_{\mathfrak{B}} \int_{\mathfrak{A}} \underline{T}u(x, a, b) \mu(da) v(db). \end{aligned}$$

Similarly we get

$$\begin{aligned} -\underline{L}^{**} \underline{U}^{**} \underline{T}^{**}(-u)(x) &= - \inf_{v \in \mathfrak{M}_{\mathcal{B}}(x)} \sup_{a \in \mathcal{A}(x)} \int_{\mathfrak{B}} \int_{\mathfrak{A}} \underline{T}^*(-u)(x, b, a) v(db) \\ &= \sup_{v \in \mathfrak{M}_{\mathcal{B}}(x)} \inf_{\mu \in \mathfrak{M}_{\mathcal{A}}(x)} \int_{\mathfrak{B}} \int_{\mathfrak{A}} \underline{T}u(x, a, b) v(db) \mu(da). \end{aligned}$$

By Fubini's Theorem and by Sion's Minimax Theorem (see [12]) we get the statement. ■

The next lemma gives us the possibility to use the games M' and M^{**} for getting ε -optimal strategies for the game M .

LEMMA 5.4. *Let u be a real bounded $\bar{\sigma}_x$ -measurable function.*

(a) *If $\delta_{\pi_n} \in E'_n$ is (ε, u) -optimal in M' , then we have $\pi_n \in E_n$ and π_n is (ε, u) -optimal in M .*

(b) *If $\delta_{\varrho_n^*} \in E'_n$ is (ε, u) -optimal in M^{**} , then there exists $\varrho_n \in F_n$ such that ϱ_n is an $(\varepsilon, -u)$ -optimal decision rule of the opponent in M , where the correspondence between ϱ_n and ϱ_n^* is the same as in Definition 5.1.*

Proof. (a): Since δ_{π_n} is (ε, u) -optimal in M' , we have

$$\delta_{\pi_n} \underline{q} \underline{T}' u \leq u + \varepsilon \quad \text{for all } \underline{q} \in F'_n.$$

Then for every Markov $\underline{q} \in F_n$ we get $\pi_n \underline{q} \underline{T}u \leq u + \varepsilon$ since we can consider the set of all Markov decision rules from F_n as a subset of F'_n . But then this inequality holds for all $\underline{q} \in F_n$.

(b): We get, as in the proof of (a), $\underline{q}_n^* \pi^* \underline{T}u \leq u + \varepsilon$ for all $\pi^* \in F_n^*$. By Fubini's Theorem follows $\pi \underline{q}_n \underline{T}(-u) \geq -u - \varepsilon$ for each $\pi \in E_n$. ■ We can now give the following two existence theorems.

THEOREM 5.5. *If a Markov game M fulfills the assumptions of Theorem 4.5 and if $\mathcal{B}(x, a)$, $(x, a) \in \hat{A}$, and \mathfrak{B} are compact, then the first player has an optimal stationary Borel strategy and the second player has for every*

$\varepsilon > 0$ an ε -optimal stationary strategy. The value of the game is the unique real bounded u.s.c. function v on \mathfrak{X} with $v = \underline{L}' \underline{U}' \underline{T}' v$.

Proof. In [7] it is shown that the assumptions of Theorem 4.5 are fulfilled by M' and that the assumptions of Theorem 4.6 are fulfilled by M^{**} .

For the u.s.c. fixpoint v of $\underline{L}' \underline{U}' \underline{T}'$ we have $-v = \underline{L}^{**} \underline{U}^{**} \underline{T}^{**}(-v)$ by Lemma 5.3. As in the proof of Theorem 4.5 and of Theorem 4.6 we get a Borel Markov $\delta_\pi \in E'_n$, which is $(0, v)$ -optimal in M' and for every $\varepsilon > 0$ a Markov $\delta_\varepsilon \in E_n^{**}$, which is $(\varepsilon, -v)$ -optimal in M^{**} . The statement follows by Lemma 5.4 and Theorem 3.4. ■

THEOREM 5.6. *If the Markov game M fulfills the assumptions of Theorem 4.7, then the first player has an optimal stationary Borel strategy and the second player has for every $\varepsilon > 0$ an ε -optimal stationary strategy. The value of the game is the unique real bounded $\sigma_{\mathfrak{X}}$ -measurable function v with $v = \underline{L}' \underline{U}' \underline{T}' v$.*

Proof. As it is shown in [7], the assumptions of Theorem 4.7 are fulfilled by M' . For the Borel fixpoint v of $\underline{L}' \underline{U}' \underline{T}'$ we again have $-v = \underline{L}^{**} \underline{U}^{**} \underline{T}^{**}(-v)$. We get as in the proof of Theorem 4.7 the existence of a Borel Markov $\delta_\pi \in E'_n$, $(0, v)$ -optimal in M' . Further,

$$\underline{T}^{**}(-v)(x, w, a) = - \int_{\mathfrak{B}} \left(k(x, a, b) + \alpha \int_{\mathfrak{X}} v(x') q(dx'/x, a, b) \right) w(db)$$

is u.s.c. in (w, a) for every $x \in \mathfrak{X}$ since $k(x, \cdot)$ is l.s.c. and $\int_{\mathfrak{X}} v(x') q(dx'/x, \cdot)$ is continuous (see [10] for example).

By Lemma 4.3 (b), $\underline{U}^{**} \underline{T}^{**}(-v)$ is a Borel function. Then by Lemma 4.3 (a), for every $\varepsilon > 0$ there exists a Markov $\delta_\varepsilon \in E_n^{**}$, which is $(\varepsilon, -v)$ -optimal in M^{**} . The statement follows again from Lemma 5.4 and Theorem 3.4. ■

At the end we remark that for a Markov game with simultaneous action choice, $\mathfrak{B}(x, \cdot)$ is always l.s.c. since $\mathcal{B}(x, a)$ is independent of a (see assumption (b) of Theorem 4.7)

References

- [1] L. D. Brown and R. Purves, *Measurable selections of extrema*, Ann. Statist. 1 (1973), 902–912.
- [2] H. A. M. Couwenbergh, *Stochastic games with metric state space*, Internat. J. Game Theory 9 (1980), 25–36.
- [3] V. Dietzsch, *Dynamische Minimax-Entscheidungsmodelle*, Dissertation A, Karl-Marx-Universität, Sektion Mathematik, Leipzig 1977.
- [4] K. Hinderer, *Foundations of Non-stationary Dynamic Programming with Discrete Time Parameter*, Lecture Notes in OR, vol. 33, Springer-Verlag, Berlin 1970.

- [5] A. Idzik, *Remarks on discounted stochastic games*, Trans. of the 8th Prague Conference, vol. C, Academia, Prague 1979.
 - [6] H.-U. Küenle, *Über die Optimalität von Strategien in stochastischen-dynamischen Minimax-Entscheidungsmodellen I*, Math. Operationsforsch. Statist. Ser. Optimization 12 (1981), 421–435.
 - [7] —, *Über die Optimalität von Strategien in stochastischen dynamischen Minimax-Entscheidungsmodellen II*, *ibid.* 14 (1983), 301–313.
 - [8] K. Kuratowski and A. Mostowski, *Set Theory*, PWN—Polish Scientific Publishers, Warszawa 1976.
 - [9] A. Maitra and T. Parthasarathy, *On stochastic games*, J. Optimization Theory Appl. 5 (1970), 289–300.
 - [10] M. Schäl, *Dynamische Optimierung unter Stetigkeits- und Kompaktheitsbedingungen*, Habilitationsschrift, Universität Hamburg, 1972.
 - [11] F. H. Simons and J. G. F. Thiemann, *A note on the Ionescu Tulcea theorem*, Technological University Eindhoven, Depart. of Math., 1978.
 - [12] M. Sion, *On general minimax theorems*, Pacific J. Math. 8 (1958), 171–176.
 - [13] J. van der Wal, *Markov games: An annotated bibliography*, Technological University Eindhoven, Depart. of Math., 1975.
-