ON DETERMINISTIC STATIONARY STRATEGIES FOR MARKOV GAMES

Iwamoto, Seiichi
Department of Mathematics, Kyushu University

Kai, Yu
Research Institute of Fundamental Information Science, Kyushu University

https://doi.org/10.5109/13084

出版情報：統計数理研究. 16 (1/2), pp.71-82, 1974-03. Research Association of Statistical Sciences
バージョン：
権利関係：

ON DETERMINISTIC STATIONARY STRATEGIES FOR MARKOV GAMES

By

Seiichi IWAMOTO* and Yû KAI**

(Received August 31, 1973)

1. Introduction.

We are concerned with the existence of optimal deterministic stationary strategies for Markov Game. Our Markov Game is specified by six tuple (S, A, B, q, r, β): S is a nonempty Borel subset of a Polish space, the set of a system; A is a nonempty Borel subset of a Polish space, the set of actions available to player I; B is a nonempty Borel subset of a Polish space, the set of actions available to player II; q is the law of motion of the system; it associates Borel measurably with each tuple $(s, a, b) \in S \times A \times B$ a probability measure $q(\cdot \mid s, a, b)$ on the Borel measurable space $(S, \mathcal{B}(S))$, where $\mathcal{B}(X)$ is the σ-field generated by the metric on X; r, the payoff function, is a bounded Borel measurable function on $S \times A \times B$; $0 \leqq \beta < 1$ is a discount factor. When the system is in s, and players I and II choose actions a and b respectively, player II pays player I payoff $r(s, a, b)$ units of money and system moves to next state s' according to the conditional distribution $q(\cdot \mid s, a, b)$. Then, the whole process is repeated from the new state s'. Since β is the discount factor, the unit income at n-th day in future is worth β^n-times of the unit one today. Then, our optimization problem is to maximize the total expected discounted gain of player I as the game proceeds over the infinite future and to minimize the expected discounted loss of player II.

A *strategy π for player I* is a sequence of $\pi_1, \pi_2, \cdots$, where π_n specifies the action to be chosen by player I on the n-th day by associating Borel measurably with each history $h = (s_1, a_1, b_1, \cdots, s_{n-1}, a_{n-1}, b_{n-1}, s_n)$ of the system a probability distribution $\pi_n(\cdot \mid h)$ on $(A, \mathcal{B}(A))$. Π denotes the class of all strategies for player I. A strategy π for player I is called *semi-Markov* (*Markov*) if each π_n is a function of s_1 and $s_n(s_n)$ alone; a strategy π is said to be *deterministic stationary* if there is a Borel measurable map f from S to A such that $\pi_n = f$ for each $n \geqq 1$; and, in this case, π is denoted by $f^{(\infty)}$. Strategies, semi-Markov strategies and deterministic stationary strategies for player II are defined analogously. Γ denotes the class of all strategies for player II.

A pair (π, σ) of strategies for players I and II associates with each initial state

* Department of Mathematics, Kyushu University, Fukuoka.

** Research Institute of Fundamental Information Science, Kyushu University, Fukuoka.

s_1 the n-th day expected gain for player I

$$I_n(\pi, \sigma)(s_1) = \pi_1\sigma_1 q\pi_2\sigma_2 q \cdots \pi_n\sigma_n qr$$

$$= \int_{\underbrace{ABS\cdots SAB}_{(3n-1)\text{ factors}}} r(s_n, a_n, b_n)d\pi_1\sigma_1 q\pi_2\sigma_2 q \cdots \pi_n\sigma_n q(\cdot \mid s_1)$$

and a total expected discounted gain for player I

$$I(\pi, \sigma)(s_1) = \sum_{n=1}^{\infty} \beta^{n-1} I_n(\pi, \sigma)(s_1).$$

It is clear that $I_n(\pi, \sigma)$ is Borel measurable and consequently $I(\pi, \sigma)$ is Borel measurable.

A strategy π^* is *optimal for player I* if for all $\sigma' \in \Gamma$ and all $s \in S$

$$\inf_{\sigma \in \Gamma} \sup_{\pi \in \Pi} I(\pi, \sigma)(s) \leqq I(\pi^*, \sigma')(s).$$

A strategy σ^* is *optimal for player II* if for all $\pi' \in \Pi$ and all $s \in S$

$$\sup_{\pi \in \Pi} \inf_{\sigma \in \Gamma} I(\pi, \sigma)(s) \geqq I(\pi', \sigma^*)(s).$$

We shall say that the *Markov game has a value* if for all $s \in S$

$$\sup_{\pi \in \Pi} \inf_{\sigma \in \Gamma} I(\pi, \sigma)(s) = \inf_{\sigma \in \Gamma} \sup_{\pi \in \Pi} I(\pi, \sigma)(s).$$

When the game has a value, the quantity $\sup_{\pi \in \Pi} \inf_{\sigma \in \Gamma} I(\pi, \sigma)$, as a function on S, is called the *value function*. Recently Maitra and Parthasarathy [5] and Parthasarathy [6] have proved, under some conditions, that the Markov game (S, A, B, q, r, β) has a value and that both players have optimal random stationary strategies.

In this paper we shall study the Markov game (S, A, B, q, r, β), too, and prove, under somewhat restrictive conditions on the payoff function r and on the conditional distribution q, that both players have optimal "deterministic" stationary strategies as well as that the game has a value. The difference between [5] and our paper is the following: for $w \in M(S)$ let's consider the two expressions

$$r(s, \mu, \lambda) + \beta \int w(\cdot) dq(\cdot \mid s, \mu, \lambda) \tag{1.1}$$

and

$$r(s, a, b) + \beta \int w(\cdot) dq(\cdot \mid s, a, b) \tag{1.2}$$

where $M(X)$ is the class of all bounded Borel measurable functions on Borel subset X, $r(s, \mu, \lambda) = \int_B \int_A r(s, a, b) d\mu(a) d\lambda(b)$ and $q(\cdot \mid s, \mu, \lambda) = \int_B \int_A q(\cdot \mid s, a, b)\, d\mu(a) d\lambda(b)$ for $\mu \in P_A$, $\lambda \in P_B$, where P_X is the class of all probability measures on $(X, \mathcal{B}(X))$. The expression in (1.1) which was treated in [5] is always concave-convex in (μ, λ), because of its bilinearity in (μ, λ). On the other hand, the expression in (1.2) is not always concave-convex in (a, b).

This paper gives a sufficient condition on A, B, q and r that makes the expres-

sion (1.2) concave-convex in (a, b) as well as continuous in (a, b). We shall show that under this condition there exist optimal deterministic stationary strategies. Our proofs are partially owing to [5] and to the results by Blackwell [1] and Strauch [8].

2. Some preliminaries.

Let X be a topological space. Then 2^X denotes the set of all non-empty closed subsets of X. In 2^X we introduce the topology (called exponential topology or Vietoris topology) which is the coarsest one in which the sets 2^A for open A (in X) are open (in 2^X) and for closed A are closed, where 2^A (rel. to X) is the set of all $F=\bar{F}\subset A$. Further by $\mathcal{O}(X)$, $\mathcal{F}(X)$ and $\mathcal{B}(X)$ we mean the set of all open, closed and topological Borel subsets in X respectively.

In the following definitions we assume that X, Y are topological spaces.

DEFINITION 2.1. A function $F: X\to 2^Y$ is $(\mathcal{B}(X), \mathcal{O}(Y))$-, $(\mathcal{B}(X), \mathcal{F}(Y))$- or $(\mathcal{B}(X), \mathcal{B}(Y))$-*measurable* iff for each $G\in\mathcal{O}(Y)$, $K\in\mathcal{F}(Y)$ or $B\in\mathcal{B}(Y)$ $F^{-1}(G)$, $F^{-1}(K)$ or $F^{-1}(B)$ belongs to $\mathcal{B}(X)$ respectively, where $F^{-1}(A)=\{x\in X\mid F(x)\cap A\neq\emptyset\}$ for $A\subset Y$.

Note that when Y is perfectly normal space, $(\mathcal{B}(X), \mathcal{F}(Y))$-measurability of $F(\cdot)$ implies $(\mathcal{B}(X), \mathcal{O}(Y))$-measurability. From above definition it is clear that $(\mathcal{B}(X), \mathcal{B}(Y))$-measurability of $F(\cdot)$ implies both $(\mathcal{B}(X), \mathcal{O}(Y))$- and $(\mathcal{B}(X), \mathcal{F}(Y))$-measurability.

DEFINITION 2.2. A function $F: X\to 2^Y$ is $\mathcal{B}(X)/\mathcal{O}(2^Y)$- or $\mathcal{B}(X)/\mathcal{B}(2^Y)$-*measurable* iff for each $\boldsymbol{G}\in\mathcal{O}(2^Y)$ or $\boldsymbol{B}\in\mathcal{B}(2^Y)$ $F^{-1}(\boldsymbol{G})$ or $F^{-1}(\boldsymbol{B})$ belongs to $\mathcal{B}(X)$ respectively, where in this case $F^{-1}(\boldsymbol{A})=\{x\in X\mid F(x)\in\boldsymbol{A}\}$ for $\boldsymbol{A}\in\mathcal{B}(2^Y)$.

Of course these two measurabilities are equivalent.

Let (Y, d) be a compact metric space. Then $(2^Y, d_H)$ with Hausdorff metric $d_H(A, B)=\max(\sup_{a\in A}\rho(a, B), \sup_{b\in B}\rho(A, b))$ is a compact metric space, too. Here $\rho(x, C)=\inf_{y\in C} d(x, y)$ for $C\subset Y$. Furthermore, by Theorem in 42-II of [4] the identity mapping: $(2^Y, \mathcal{O}(2^Y))\to(2^Y, d_H)$ is a homeomorphism.

LEMMA 2.1. *Let X be a metric space and Y a compact metric space. A function $F: X\to 2^Y$ is $\mathcal{B}(X)/\mathcal{B}(2^Y)$-measurable if and only if F is $(\mathcal{B}(X), \mathcal{F}(Y))$-measurable.*

PROOF. It is sufficient to prove that $\mathcal{B}(X)/\mathcal{O}(2^Y)$-measurability is equivalent to $(\mathcal{B}(X), \mathcal{F}(Y))$-measurability. From the definition of Vietoris topology $\mathcal{O}(2^Y)$ is generated by the open subbase

$$\mathcal{O}^{**}(2^Y)=\{\{B\in 2^Y;\ B\subset G\},\ \{B\in 2^Y;\ B\cap H\neq\phi\};\ G,\ H\in\mathcal{O}(Y)\}.$$

We have for any $G\in\mathcal{O}(Y)$

$$\begin{aligned}\{x;\ F(x)\in\{B\in 2^Y;\ B\subset G\}\}&=\{x;\ F(x)\subset G\}\\&=\{x;\ F(x)\cap G^c=\phi\}\\&=\{x;\ F(x)\cap G^c\neq\phi\}^c\end{aligned}\tag{2.1}$$

and for any $H\in\mathcal{O}(Y)$

$$\{x;\ F(x)\in\{B\in 2^Y;\ B\cap H\neq\phi\}\}=\{x;\ F(x)\cap H\neq\phi\}.\tag{2.2}$$

Let $F(\cdot)$ be $(\mathcal{B}(X), \mathcal{F}(Y))$-measurable. Since Y is a metric space, Y is perfectly normal space. Then, remark below Definition 2.1 assures that $F(\cdot)$ is $(\mathcal{B}(X), \mathcal{O}(Y))$-measurable. Hence by (2.1) and (2.2) it holds that for any $\boldsymbol{O}^{**}$ in the open subbase $\mathcal{O}^{**}(2^Y)$ $\{x;\ F(x)\in \boldsymbol{O}^{**}\}\in\mathcal{B}(X)$. Through the proof of Theorem in 42-II of [4] we have for each open ball $\boldsymbol{R}$ with center $A\in 2^Y$, namely, $\boldsymbol{R}=\{B\in 2^Y \mid d_H(A, B)<\varepsilon\}$,

$$\boldsymbol{R}=\bigcup_{k=1}^{\infty}[\{B\in 2^Y;\ B\subset G\}\cap\{B\in 2^Y;\ B\cap G_1^k\neq\phi\}\cap\cdots\cap\{B\in 2^Y;\ B\cap G_{n_k}^k\neq\phi\}]$$

where $G_i^k=\{x\mid d(x, a_i^k)<\varepsilon-1/k\}$ and $G=\{x\mid\rho(x, A)<\varepsilon\}$ are open in Y and $\{a_i^k, \cdots, a_{n_k}^k\}_{k\geq 1}$ is a finite system of points of A such that for each $x\in A$ and k we have $d(a_i^k, x)<1/k$ for some i. (The compactness of A enables us to choose such a system). Therefore we have for any ball $\boldsymbol{R}$ $\{x;\ F(x)\in\boldsymbol{R}\}\in\mathcal{B}(X)$. Since $(2^Y, d_H)$ is compact metric space, any open set $\boldsymbol{O}$ in $(2^Y, d_H)$ is a countable union of the base $\{\boldsymbol{R}_i;\ i\geq 1\}$. Consequently we have for any open set $\boldsymbol{O}\in\mathcal{O}(2^Y)$ $\{x;\ F(x)\in\boldsymbol{O}\}\in\mathcal{B}(X)$.

Conversely if $F(\cdot)$ is $\mathcal{B}(X)/\mathcal{O}(2^Y)$-measurable, (2.1) yields that for any closed set K in Y

$$\{x;\ F(x)\cap K\neq\phi\}=\{x;\ F(x)\in\{B\in 2^Y;\ B\subset K^c\}\}^c.$$

Then in this case $F^{-1}(K)\in\mathcal{B}(X)$ for any $K\in\mathcal{F}(Y)$. This completes the proof.

LEMMA 2.2. *Let X be a metric space, Y be a compact metric space and $v: XY\to R$ be bounded, Borel measurable in x for each $y\in Y$ and continuous in y for each $x\in X$. Define $v^*: X\to 2^Y$ by*

$$v^*(x)=\{y;\ v(x, y)=\max_{y'\in Y} v(x, y')\}.$$

Then $v^*(\cdot)$ is $\mathcal{B}(X)/\mathcal{B}(2^Y)$ measurable.

PROOF. By Lemma 2.1 it sufficies to prove $(\mathcal{B}(X), \mathcal{F}(Y))$-measurability of $v^*(\cdot)$. Since any $L\in\mathcal{F}(Y)$ is compact, there exists a countable set D in L such that $\bar{D}=L$. Then we have

$$\{x;\ v^*(x)\cap L\neq\phi\}=\bigcap_{n=1}^{\infty}\Big[\bigcup_{y_i\in D}\Big\{x;\ |\max_{y'\in Y} v(x, y')-v(x, y_i)|<\frac{1}{n}\Big\}\Big]. \tag{2.3}$$

If $v^*(x)\cap L\neq\phi$, then there exists a $y^*\in L$ such that $v(x, y^*)=\max_{y'\in Y} v(x, y')$. We can choose subsequence $\{y_{i_j};\ j\geq 1\}\subset D$ such that $y_{i_j}\to y^*$ as $j\to\infty$. Since $v(x, \cdot)$ is continuous in y, for any $n\geq 1$ there exists a $y_{i_{j_n}}\in D$ such that

$$|v(x, y^*)-v(x, y_{i_{j_n}})|<\frac{1}{n}.$$

Consequently we have $x\in\bigcap_{n=1}^{\infty}\Big[\bigcup_{y_i\in D}\Big\{x;\ |\max_{y'\in Y} v(x, y')-v(x, y_i)|<\frac{1}{n}\Big\}\Big]$. Conversely if $x\in\bigcap_{n=1}^{\infty}\Big[\bigcup_{y_i\in D}\Big\{x;\ |\max_{y'\in Y} v(x, y')-v(x, y_i)|<\frac{1}{n}\Big\}\Big]$, then there exists a subsequence $\{y_{i_n};\ n\geq 1\}\subset D\subset L$ such that

$$|\max_{y'\in Y} v(x, y')-v(x, y_{i_n})|<\frac{1}{n}\qquad\text{for each } n\geq 1.$$

Since L is compact, we can choose subsequence $\{y_{i_{n_k}};\ k\geq 1\}$ such that $y_{i_{n_k}}\to y^*\in L$ as $k\to\infty$. By y-continuity of $v(x, \cdot)$, we have

$$v(x, y^*) = \max_{y' \in Y} v(x, y') .$$

Then (2.3) yields $\{x ; v^*(x) \cap L \neq \phi\} \in \mathcal{B}(X)$ because of Borel measurability in x of the functions $\max_{y' \in Y} v(\cdot, y')$ and $v(\cdot, y)$ for fixed $y \in Y$. This completes the proof of Lemma 2.2.

Note that Lemma 2.2 shows that the assumption (ii) in [6] is redundant.

LEMMA 2.3. *General Selector Theorem (Kuratowski and Ryll-Nardzewski* [3]). *Let X be a metric space, Y be a separable metric space and $V : X \to 2^Y$ be $\mathcal{B}(X)/\mathcal{B}(2^Y)$-measurable. Then there exists a Borel measurable selector v such that*

$$v(x) \in V(x) \qquad \text{for all } x \in X .$$

PROOF. This lemma is due to Kuratowski and Ryll-Nardzewski [3].

LEMMA 2.4. *Let $v : XY \to R$ be a bounded continuous function, where X is a metric space and Y a compact metric space. Then $v^* : X \to R$ defined by $v^*(x) = \max_{y \in Y} v(x, y)$ is continuous. Similarly, $v_* : X \to R$ defined by $v_*(x) = \min_{y \in Y} v(x, y)$ is also continuous.*

PROOF. This lemma is stated in Lemma 2.2 of [5].

We shall set the following assumptions:

ASSUMPTION (I). $u : SAB \to R$ is bounded on SAB, continuous in (a, b) for each $s \in S$ and Borel measurable in s for each $(a, b) \in AB$.

ASSUMPTION (II). $u : SAB \to R$ is concave in a for each $b \in B$ and convex in b for each $a \in A$ (abbreviated hereafter concave convex in (a, b)), where A, B are compact convex metric spaces.

Now we can prove the Minimax Selector Theorem for our Markov game that is fundamental in finding optimal deterministic stationary strategies over all possible strategies.

THEOREM 2.1 (*Minimax Selector Theorem*). *Under the Assumptions* (I), (II), *there exist Borel measurable $f : S \to A$ and $g : S \to B$ such that for all $s \in S$*

$$u_*(s, f(s)) = \max_{a \in A} u_*(s, a)$$

$$u^*(s, g(s)) = \min_{b \in B} u^*(s, b) ,$$

where $u_*(s, a) = \min_{b \in B} u(s, a, b)$, $u^*(s, b) = \max_{a \in A} u(s, a, b)$. Hence, for all $s \in S$

$$u(s, f(s), g(s)) = \min_{b \in B} \max_{a \in A} u(s, a, b) = \max_{a \in A} \min_{b \in B} u(s, a, b) .$$

PROOF. Lemma 2.4, together with Assumption (I), implies a-continuity of $u_*(s, \cdot)$ for fixed $s \in S$. Since B is compact, $u_*(\cdot, a)$ is measurable in s for fixed $a \in A$. By Lemma 2.2 and 2.3, there exists a Borel measurable $f : S \to A$ such that

$$u_*(s, f(s)) = \max_{a \in A} u_*(s, a) \qquad s \in S . \tag{2.4}$$

Similarly, there exists a Borel measurable $g : S \to B$ such that

$$u^*(s, g(s)) = \min_{b \in B} u^*(s, a) \qquad s \in S . \tag{2.5}$$

Namely,

$$\min_{b\in B} u(s, f(s), b) = \max_{a\in A} \min_{b\in B} u(s, a, b) \tag{2.6}$$

and

$$\max_{a\in A} u(s, a, g(s)) = \min_{b\in B} \max_{a\in A} u(s, a, b)\,. \tag{2.7}$$

By Assumption (II), appealing to Sion's minimax theorem [7], we have

$$u(s, f(s), g(s)) = \min_{b\in B} \max_{a\in A} u(s, a, b) = \max_{a\in A} \min_{b\in B} u(s, a, b)\,.$$

This completes the proof.

Let $M^+(S)$ denote the class of all bounded nonnegative Borel measurable functions on S. For $u\in M(S)$ we mean $\|u\| = \sup_{s\in S} |u(s)|$. Note that $M^+(S)$ is a closed subset in the complete metric space $(M(S), d)$, where $d(u, v)=\|u-v\|$. Hence $(M^+(S), d)$ is a complete metric space.

We further consider the Markov Game (S, A, B, q, β, r^*), where

$$r^*(s, a, b) = r(s, a, b)+\|r\| \geqq 0\,, \qquad (s, a, b)\in SAB\,.$$

That is, in the new-defined Markov Game its payoff function is an element of $M^+(SAB)$, other components being not altered. We call the game (S, A, B, q, β, r) *original Markov Game* and the game (S, A, B, q, β, r^*) *modified Markov Game*, or simply "*original M.G.*" and "*modified M.G.*", respectively. Then we have the following relationship between two Markov Games, which was pointed by Kai [2] in the case of finite state Markov Game.

LEMMA 2.5. *Any strategy (π, σ) in the original M.G. (S, A, B, q, β, r) can be regarded as a strategy (π, σ) in the modified M.G. (S, A, B, q, β, r^*) and vice versa. Furthermore it follows that*

$$I(\pi, \sigma) = I^*(\pi, \sigma) - \frac{\|r\|}{1-\beta}\,,$$

where $I^*(\pi, \sigma)$ is the total expected discounted payoff associated with modified M.G. (S, A, B, q, β, r^*).

PROOF. The proof of this lemma is straighforward.

In the following Lemmas 2.6 and 2.7 we assume that A and B are compact convex sets.

LEMMA 2.6. *Let for each $n\geqq 1$ $g_n(a, b): A\times B\to R$ be concave-convex in (a, b).*

(i) *If $\alpha_n \geqq 0$ for $n=1, 2, \cdots, l$, then $\sum_{n=1}^{l} \alpha_n g_n(a, b): A\times B\to R$ is concave-convex in (a, b).*

(ii) *If $g_n(a, b)$ converges to $g(a, b)$ as $n\to\infty$, then $g(a, b)$ is concave-convex in (a, b).*

PROOF. Easy.

LEMMA 2.7. *Let $q=q(B\,|\,s, a, b): \mathcal{B}(S)\times S\times A\times B\to R$ be concave-convex in (a, b) for each $(B, s)\in \mathcal{B}(S)\times S$. If $v=v(s): S\to R$ is a nonnegative bounded Borel measurable function, then $\int_S v(s')dq(s'\,|\,s, a, b): S\times A\times B\to R$ is concave-convex in (a, b) for each $s\in S$.*

PROOF. This is a trivial consequence of Lemma 2.6.

We further set the following assumption:

ASSUMPTION (III). The action spaces A, B are compact convex, the payoff function $r=r(s, a, b)$ is continuous and concave-convex in (a, b) for each $s\in S$ and the law of motion $q=q(B\,|\,s, a, b)$ is continuous and concave-convex in (a, b) for each $(B, s)\in\mathcal{B}(S)\times S$.

DEFINITION 2.3. Let Assumption (III) be satisfied. For any $w\in M^+(S)$, we associate $(Tw)(s)$ defined by

$$(Tw)(s)=\max_{a\in A}\min_{b\in B}[r^*(s, a, b)+\beta\int_S w(s')dq(s'\,|\,s, a, b)] \tag{2.8}$$

$$=\min_{b\in B}\max_{a\in A}[r^*(s, a, b)+\beta\int_S w(s')dq(s'\,|\,s, a, b)]\,.$$

Note that T is a well-defined operator on $M^+(S)$ under the Assumption (III).

LEMMA 2.8. *Under the Assumption* (III), *the operator T is a contraction mapping on $M^+(S)$ with contraction coefficient $\beta<1$. Hence T has a unique fixed point w^* in $M^+(S)$.*

PROOF. Easy.

Throughout the remainder of this section, we shall assume Assumption (III). For above unique fixed point $w^*\in M^+(S)$, we define $u_{w^*}: SAB\to R$ by $u_{w^*}(s, a, b)=r^*(s, a, b)+\beta\int w^*(\cdot)dq(\cdot\,|\,s, a, b)$. Minimax Selector Theorem together with Assumption (III), then, enables us to choose Borel measurable $f: S\to A$ and $g: S\to B$ such that for all $s\in S$

$$\begin{aligned}\min_{b\in B}u_{w^*}(s, f(s), b)&=\max_{a\in A}\min_{b\in B}u_{w^*}(s, a, b)\\&=\min_{b\in B}\max_{a\in A}u_{w^*}(s, a, b)\\&=\max_{a\in A}u_{w^*}(s, a, g(s))\\&=u_{w^*}(s, f(s), g(s))\,.\end{aligned} \tag{2.9}$$

Now we can introduce a dummy game $(S, A, B, u_{w^*}(\cdot))$, where u_{w^*} is the unique fixed point of the operator T on $M^+(S)$. By (2.9), this dummy game is strictly determined, its value is $w^*(s)$ and $f(s)$ and $g(s)$ are optimal strategies in the dummy game for players I and II respectively.

3. Semi-Markov strategies are enough.

Player I wishes to maximize his expected discounted gain by knowing his complete past history up to date. However, if he can maximize his gain by knowing just any partial history without knowing whole past history, he will do so. Similar question will arise to player II. By the method of Strauch [7], we shall prove that it is enough for both players only to know initial and present states, that is, to use only semi-Markov strategies.

LEMMA 3.1. *Let (π, σ) be any strategies and $p\in P_S$.*

(i) *There exist random semi-Markov strategies (π^*, σ^*) such that for any n and any $r\in M(SSABS)$*

$$e_\pi^\sigma r(s_1, s_n, a_n, b_n, s_{n+1}) = e_{\pi^*}^{\sigma^*} r(s_1, s_n, a_n, b_n, s_{n+1}),$$

where $e_\pi^\sigma = \pi_1\sigma_1 q\pi_2\sigma_2 q \cdots$.

(ii) *There exist random Markov strategies* (π^{**}, σ^{**}) *such that for any* n *and any* $r \in M(SABS)$

$$pe_\pi^\sigma r(s_n, a_n, b_n, s_{n+1}) = pe_{\pi^{**}}^{\sigma^{**}} r(s_n, a_n, b_n, s_{n+1}).$$

PROOF. Let π_n^*, σ_n^* be the conditional distributions of a_n, b_n respectively given s_n and s_1 under e_π^σ, and let π_n^{**}, σ_n^{**} be the conditional distributions of a_n, b_n respectively given s_n under pe_π^σ. We shall prove (i), the proof of (ii) is similar. The lemma is true for $n=1$, since $\pi_1^* = \pi_1$ and $\sigma_1^* = \sigma_1$, hence

$$e_\pi^\sigma r(s_1, a_1, b_1, s_2) = \pi_1\sigma_1 qr = \pi_1^*\sigma_1^* qr = e_{\pi^*}^{\sigma^*} r(s_1, a_1, b_1, s_2).$$

Now assume the lemma is true of $n < N$. All expectations are under the conditional probability e_π^σ.

$$\begin{aligned} e_\pi^\sigma r(s_1, s_N, a_N, b_N, s_{N+1}) &= E[r(s_1, s_N, a_N, b_N, s_{N+1}) \mid s_1] \\ &= E\{E(r(s_1, s_N, a_N, b_N, s_{N+1}) \mid s_1, s_N) \mid s_1\} \\ &= E\{u(s_1, s_N) \mid s_1\} \\ &= e_\pi^\sigma u(s_1, s_N) \end{aligned}$$

where

$$\begin{aligned} u(s_1, s_N) &= E(r(s_1, s_N, a_N, b_N, s_{N+1}) \mid s_1, s_N) \\ &= \pi_N^* \sigma_N^* qr(s_1, s_N, a_N, b_N, s_{N+1}) \end{aligned}$$

by the properties of conditional distribution. But $u(s_1, s_N) = v(s_1, s_{N-1}, a_{N-1}, b_{N-1}, s_N) \in M(SSABS)$, and by the induction hypothesis

$$e_\pi^\sigma u(s_1, s_N) = e_{\pi^*}^{\sigma^*} u(s_1, s_N).$$

Thus

$$\begin{aligned} e_\pi^\sigma r(s_1, s_N, a_N, b_N, s_{N+1}) &= e_{\pi^*}^{\sigma^*} u(s_1, s_N) \\ &= e_{\pi^*}^{\sigma^*} \pi_N^* \sigma_N^* qr(s_1, s_N, a_N, b_N, s_{N+1}) \\ &= e_{\pi^*}^{\sigma^*} r(s_1, s_N, a_N, b_N, s_{N+1}). \end{aligned}$$

This completes the proof.

THEOREM 3.1. *Let* (π, σ) *be any strategies and* $p \in P_S$. *Then there exist random semi-Markov strategies* (π^*, σ^*) *and random Markov strategies* (π^{**}, σ^{**}) *such that*

$$I(\pi, \sigma) = I(\pi^*, \sigma^*)$$

and

$$pI(\pi, \sigma) = pI(\pi^{**}, \sigma^{**})$$

for any payoff function $r \in M(SABS)$.

PROOF. From Lemma 3.1 it follows that for any payoff function r

$$I_n(\pi, \sigma) = I_n(\pi^*, \sigma^*)$$

and

$$pI_n(\pi, \sigma) = pI_n(\pi^{**}, \sigma^{**}).$$

Since $0 \leqq \beta < 1$ is a discount factor, $I_n(\pi, \sigma)$ converges to $I(\pi, \sigma)$ as $n \to \infty$ for any

$(\pi, \sigma) \in \Pi \times \Gamma$. Hence

$$I(\pi, \sigma) = I(\pi^*, \sigma^*) .$$

Furthermore, dominated convergence theorem yields

$$pI(\pi, \sigma) = pI(\pi^{**}, \sigma^{**})$$

This completes the proof.

COROLLARY. *Let $\Pi_{s,m}$, $\Gamma_{s,m}$ be the classes of all random semi-Markov strategies for players I and II respectively. Then*

$$\sup_{\pi \in \Pi} \inf_{\sigma \in \Gamma} I(\pi, \sigma) = \sup_{\Pi_{s,m}} \inf_{\Gamma_{s,m}} I(\pi, \sigma) ,$$

$$\inf_{\sigma \in \Gamma} \sup_{\pi \in \Pi} I(\pi, \sigma) = \inf_{\Gamma_{s,m}} \sup_{\Pi_{s,m}} I(\pi, \sigma) .$$

4. Relations to Markovian decision processes.

Throughout this section, let's assume that player II is allowed to use only the strategy $g^{(\infty)}$, where $g: S \to B$ is the Borel measurable function defined at the end Section 2, while that the choice of player I remains unrestricted.

Associated with the Markov game (S, A, B, q, r^*, β), we consider the Markovian decision process (hereafter abbreviated by M.D.P.) (S, A, q', r', β) such that

$$q'(s' \mid s, a) = q(s' \mid s, a, g(s)) ,$$

$$r'(s, a) = r^*(s, a, g(s))$$

$$= r(s, a, g(s)) + \|r\| .$$

$I'(\pi')$ denotes the total expected discounted reward from π', where π' is a policy for M.P.D. (S, A, q', r', β).

LEMMA 4.1. *If π is a semi-Markov (stationary) strategy for player I in the M.G. (S, A, B, q, r^*, β), then π is a semi-Markov (stationary) policy in the M.D.P. (S, A, q', r', β) and $I^*(\pi, g^{(\infty)}) = I'(\pi)$. Conversely, if π is a semi-Markov (stationary) policy in the M.D.P. (S, A, q', r', β), then π is a semi-Markov (stationary) strategy for player I in the M.G. (S, A, B, q, r^*, β) with $I'(\pi) = I^*(\pi, g^{(\infty)})$.*

PROOF. The proof is easily verified and is omitted.

This lemma looks like a same version as Lemma 3.1 of [5], but it actually different from that one. Because in our lemma $g: S \to B$ is Borel measurable, while in [5] $g: S \to P_B$.

LEMMA 4.2. *A policy π' is optimal for the M.D.P. (S, A, q', r', β) if and only if its return $I'(\pi')$ satisfies the optimality equation, that is,*

$$I'(\pi') = \sup_{a \in A} T_a I'(\pi') ,$$

where $T_a u(s) = r'(s, a) + \beta \int u(\cdot) dq'(\cdot \mid s, a)$ for $u \in M(S)$.

PROOF. This lemma is stated as Theorem 6(f) in [1].

THEOREM 4.1. *If $f^{(\infty)}$ is an optimal deterministic stationary policy in the M.D.P. (S, A, q', r', β), then*

$$I^*(f^{(\infty)}, g^{(\infty)}) = \sup_{\pi \in \Pi} I^*(\pi, g^{(\infty)}) .$$

PROOF. Let $\pi \in \Pi$ be an arbitrary strategy for player I in the M.G. (S, A, B, q, r, β). We apply Theorem 3.1 to $\sigma = g^{(\infty)}$, then there exists a semi-Markov strategy π^* for the M.G. (S, A, B, q, r, β) such that $I^*(\pi^*, g^{(\infty)}) = I^*(\pi, g^{(\infty)})$. Since $f^{(\infty)}$ is optimal for the M.D.P. (S, A, q', r', β), $I'(f^{(\infty)}) \geqq I'(\pi)$. By Lemma 4.1, $I^*(f^{(\infty)}, g^{(\infty)}) \geqq I^*(\pi, g^{(\infty)})$ for arbitrary strategy π for player I. This completes the proof.

5. Existence of optimal deterministic strategies.

In this section we assume that Assumption (III) remains operative. Then, by Lemma 2.8, there exists an unique fixed point w^* in $M^+(S)$ of the operator T. Furthermore, by Theorem 2.1, there exist Borel measurable $f: S \to A$ and $g: S \to B$ such that for all $s \in S$

$$\begin{aligned} & r^*(s, f(s), g(s)) + \beta \int w^*(\cdot) dq(\cdot \mid s, f(s), g(s)) \\ & = \min_{b \in B} \max_{a \in A} [r^*(s, a, b) + \beta \int w^*(\cdot) dq(\cdot \mid s, a, b)] \\ & = \max_{a \in A} \min_{b \in B} [r^*(s, a, b) + \beta \int w^*(\cdot) dq(\cdot \mid s, a, b)] . \end{aligned} \tag{5.1}$$

DEFINITION 5.1. For any Borel measurable $f: S \to A$ and $g: S \to B$, define an operator $L(f, g)$ by

$$L(f, g)w(s) = r^*(s, f(s), g(s)) + \beta \int w(\cdot) dq(\cdot \mid s, f(s), g(s)) \qquad s \in S,\ w \in M^+(S) .$$

The following lemma is trivial.

LEMMA 5.1. *The operator $L(f, g)$ is a contraction mapping on $M^+(S)$ with contraction coefficient $\beta < 1$ and $I^*(f^{(\infty)}, g^{(\infty)})$ is its unique fixed point in $M^+(S)$.*

PROOF. Easy.

Finally we have

THEOREM 5.1. *Let A, B be compact metric spaces. Uuder Assumption* (III), *the Markov Game (S, A, B, q, r, β) has a value, the value function is Borel measurable and both players have optimal deterministic stationary strategies.*

PROOF. By Lemma 2.5, it suffices to show that the same result is true for modified M.G. (S, A, B, q, r^*, β). By virtue of Assumption (III), Lemma 2.8 shows the existence of the unique fixed point w^* in $M^+(S)$ of the operator T. Furthermore, by Theorem 2.1 (Minimax Selector Theorem), there exist Borel measurable $f: S \to A$ and $g: S \to B$ such that for all $s \in S$

$$\begin{aligned} & r^*(s, f(s), g(s)) + \beta \int w^*(\cdot) dq(\cdot \mid s, f(s), g(s)) \\ & = \min_{b \in B} \max_{a \in A} [r^*(s, a, b) + \beta \int w^*(\cdot) dq(\cdot \mid s, a, b)] \\ & = \max_{a \in A} \min_{b \in B} [r^*(s, a, b) + \beta \int w^*(\cdot) dq(\cdot \mid s, a, b)] \\ & = w^*(s) . \end{aligned}$$

This is equivalent to

$$L(f, g)w^* = Tw^* = w^*.$$

Consequently, Lemma 5.1 implies that

$$w^* = I^*(f^{(\infty)}, g^{(\infty)}).$$

Hence we have

$$I^*(f^{(\infty)}, g^{(\infty)})(s) = \max_{a \in A} [r^*(s, a, g(s)) + \beta \int I^*(f^{(\infty)}, g^{(\infty)})(\cdot) dq(\cdot \mid s, a, g(s))] \quad s \in S. \tag{5.2}$$

Then by virtue of Lemma 4.1, we can write (5.2) as follows:

$$I'(f^{(\infty)})(s) = \max_{a \in A} [r'(s, a) + \beta \int I'(f^{(\infty)})(\cdot) dq'(\cdot \mid s, a)] \qquad s \in S. \tag{5.3}$$

Hence $I'(f^{(\infty)})$ satisfies the optimality equation for the M.D.P. (S, A, q', r', β). Lemma 4.2 implies $f^{(\infty)}$ is optimal for this M.D.P. Furthermore, by Theorem 4.1,

$$I^*(f^{(\infty)}, g^{(\infty)}) = \sup_{\pi \in \Pi} I^*(\pi, g^{(\infty)}). \tag{5.4}$$

Following the similar argument,

$$I^*(f^{(\infty)}, g^{(\infty)})(s) = \min_{b \in B} [r^*(s, f(s), b) + \beta \int I^*(f^{(\infty)}, g^{(\infty)})(\cdot) dq(\cdot \mid s, f(s), b)] \quad s \in S \tag{5.5}$$

leads to

$$I^*(f^{(\infty)}, g^{(\infty)}) = \inf_{\sigma \in \Gamma} I^*(f^{(\infty)}, \sigma). \tag{5.6}$$

Consequently

$$I^*(f^{(\infty)}, g^{(\infty)}) = \sup_{\pi \in \Pi} I^*(\pi, g^{(\infty)}) \geqq \inf_{\sigma \in \Gamma} \sup_{\pi \in \Pi} I^*(\pi, \sigma). \tag{5.7}$$

Similarly

$$I^*(f^{(\infty)}, g^{(\infty)}) = \inf_{\sigma \in \Gamma} I^*(f^{(\infty)}, \sigma) \leqq \sup_{\pi \in \Pi} \inf_{\sigma \in \Gamma} I^*(\pi, \sigma). \tag{5.8}$$

Hence

$$\inf_{\sigma \in \Gamma} \sup_{\pi \in \Pi} I^*(\pi, \sigma) = I^*(f^{(\infty)}, g^{(\infty)}) = \sup_{\pi \in \Pi} \inf_{\sigma \in \Gamma} I^*(\pi, \sigma).$$

From (5.7) and (5.8) we have

$$I^*(f^{(\infty)}, \sigma') \geqq \inf_{\sigma \in \Gamma} \sup_{\pi \in \Pi} I^*(\pi, \sigma) \qquad \sigma' \in \Gamma,$$

$$I^*(\pi', g^{(\infty)}) \leqq \sup_{\pi \in \Pi} \inf_{\sigma \in \Gamma} I^*(\pi, \sigma) \qquad \pi' \in \Pi.$$

This completes the proof.

Acknowledgement. The authors wish to express their hearty thanks to Professsr N. Furukawa for his advices in preparing this paper.

References

[1] D. Blackwell, *Discounted Dynamic Programming.* Ann. Math. Statist. 37 (1966), 226-235.

[2] Y. Kai, *On optimal non-random stationary policies in finite state stochastic games,* Bull. Math. Statist., 15 (1973), 93-99.

[3] K. KURATOWSKI and C. RYLL-NARDZEWSKI, *A general theorem on selectors*, Bull. Acad. Polon. Sci. Ser. Math. Astronom. Phys. **13** (1965), 397-403.
[4] K. KURATOWSKI, Topology, Vol. II, Academic Press. (1966).
[5] A. MAITRA and T. PARTHASARATHY, *On stochastic games*, J. Optimization Theory Appl. **5** (1970), 289-300.
[6] T. PARTHASARATHY, *Discounted and Positive Stochastic Games*, Bull. Amer. Math. Soc. **77** (1971), 134-136.
[7] M. SION, *On general minimax theorems*, Pacific J. Math. **8** (1958), 171-176.
[8] R. E. STRAUCH, *Negative Dynamic Programming*, Ann. Math. Statist. **37** (1966), 871-890.