On Markov games with the expected average reward criterion

By

Kensuke TANAKA, Seiichi IWASE and Kazuyoshi WAKUTA

(Received May 20, 1975)

1. Introduction

In recent years considerable attention has been given to Markov games with a specified discount factor β, $0 \leqq \beta < 1$, so that the unit reward on the nth day is worth only β^{n-1}. The optimization problem, then, is to maximize the total expected discounted gain of player I as the game proceeds over the infinite future and to minimize the total expected loss of player II.

But, if $\beta = 1$, the total expected gain of player I or total expected loss of player II may diverge. Then, we need to consider the expected average reward as a criterion of optimality. For this reason, in this paper, we shall treat the problem, that is, to maximize the expected average gain per unit time of player I as the game continues on the infinite future and to minimize the expected average loss per unit time of player II. Then, under this criterion and some assumptions, we shall show that Markov game has a value and, moreover, we shall give an algorithm for finding the value of the game and for finding ε-optimal strategies. Furthermore, we shall show that, under some assumptions, Markov game with the expected average reward criterion is reduced to one with some specified discount factor.

This paper consists of four sections. In Section 2, we shall give the formulation of the problem treated by us in this paper. In Section 3, we shall show the existence of optimal stationary strategies and give an algorithm for finding ε-optimal strategies.

2. The formulation of the problem

In this paper, our Markov game is determined by a tupel (S, A, B, q, r). Here, S is a non-empty Borel subset of a Polish space, the set of states of a system; A is a non-empty Borel subset of a Polish space, the set of actions available to player I; B is a non-empty 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 triple $(s, a, b) \in S \times A \times B$ a probability measure $q(\cdot \mid s, a, b)$ on the Borel measurable space $(S, \mathfrak{B}(S))$, where $\mathfrak{B}(S)$ is

the σ-field generated by the metric on S; r, the reward function, is a bounded Borel measurable function on $S\times A\times B$.

Periodically, player I and player II observe the current state s and choose actions a and b according to the current state s and the full knowledge of the history of the system. As a consequence of the actions chosen by the players, player II pays player I reward $r(s, a, b)$ units of money and the system moves to a new state s' according to the conditional distribution $q(\cdot|s, a, b)$. Then, the whole process is repeated from the new state s'. Here, our optimization problem is to maximize the expected average gain per unit time of player I and to minimize the expected average loss per unit time of player II.

A *strategy* π for player I is a sequence of $\pi_1, \pi_2,\ldots$, where π_n specifies the action to be chosen by player I on the nth time by associating Borel measurably with each history $h_n=(s_1, a_1, b_1, s_2,\ldots, a_{n-1}, b_{n-1}, s_n)$ of the system a probability distribution $\pi_n(\cdot|h_n)$ on $(A, \mathfrak{B}(A))$. A strategy π is, particularly, said to be *stationary* if there is a Borel measurable map f from S into P_A such that $\pi_n=f$, for all n, where P_A is the set of all probability measures on $(A, \mathfrak{B}(A))$. Π denotes the class of all strategies for player I. Strategies and stationary strategies for player II are defined analogously. Γ denotes the class of all strategies for player II.

For each n, a pair (π, σ) of the strategies for player I and player II associates with each initial state s the total expected reward $I_n(\pi, \sigma)(s)$ of player I up to the nth time and the expected average gain per unit time of player I up to the nth time:

$$\frac{I_n(\pi, \sigma)(s)}{n} \tag{2. 1}$$

Then, player I wants to maximize

$$\overline{\lim_{n\to\infty}} \frac{I_n(\pi, \sigma)(s)}{n} \tag{2. 2}$$

and player II wants to minimize

$$\underline{\lim_{n\to\infty}} \frac{I_n(\pi, \sigma)(s)}{n} \tag{2. 3}$$

For each n the function $I_n(\pi, \sigma)$ is, plainly, Borel measurable and, consequently,

$$\overline{\lim_{n\to\infty}} \frac{I_n(\pi, \sigma)(s)}{n} \quad \text{and} \quad \underline{\lim_{n\to\infty}} \frac{I_n(\pi, \sigma)(s)}{n}$$

are Borel measurable.

A strategy π^* is *optimal* for player I if for all strategies σ' for player II and all $s\in S$,

$$\inf_{\sigma\in\Gamma} \sup_{\pi\in\Pi} \overline{\lim_{n\to\infty}} \frac{I_n(\pi, \sigma)(s)}{n} \leqq \underline{\lim_{n\to\infty}} \frac{I_n(\pi^*, \sigma')(s)}{n} \tag{2. 4}$$

A strategy σ^* is *optimal* for player II if for all strategies π' for player I and all $s\in S$,

$$\sup_{\pi\in\Pi}\inf_{\sigma\in\Gamma}\varliminf_{n\to\infty}\frac{I_n(\pi,\sigma)(s)}{n}\geqq\varlimsup_{n\to\infty}\frac{I_n(\pi',\sigma^*)(s)}{n}. \tag{2. 5}$$

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

$$\inf_{\sigma\in\Gamma}\sup_{\pi\in\Pi}\varlimsup_{n\to\infty}\frac{I_n(\pi,\sigma)(s)}{n}=\sup_{\pi\in\Pi}\inf_{\sigma\in\Gamma}\varliminf_{n\to\infty}\frac{I_n(\pi,\sigma)(s)}{n}. \tag{2. 6}$$

In case the Markov game has a value, the quantity

$$\inf_{\sigma\in\Gamma}\sup_{\pi\in\Pi}\varlimsup_{n\to\infty}\frac{I_n(\pi,\sigma)(s)}{n},$$

as a function on S, is called the value function.

3. Existence of optimal stationary strategies

In this section, we shall study the existence of optimal stationary strategies for our Markov game. First, we shall state several assumptions and lemmas necessary for the proofs of main results. We shall assume the following assumptions: (A1) S, A and B are compact metric spaces, (A2) $r=r(s,a,b)$ is a continuous function on $S\times A\times B$, (A3) whenever $s_n\longrightarrow s_0$, $a_n\longrightarrow a_0$ and $b_n\longrightarrow b_0$, $q(\cdot|s_n,a_n,b_n)$ converges weakly to $q(\cdot|s_0,a_0,b_0)$, (A4) there exist a continuous function $u(s)$ on S and a constant d such that for each $s\in S$,

$$d+u(s)=\sup_{\mu\in P_A}\inf_{\lambda\in P_B}\left\{r(s,\mu,\lambda)+\int_S u(s')dq(s'|s,\mu,\lambda)\right\}, \tag{3. 1}$$

where, for each $\mu\in P_A$ and $\lambda\in P_B$,

$$r(s,\mu,\lambda)=\iint r(s,a,b)\,d\mu(a)\,d\lambda(b) \tag{3. 2}$$

and, for each $E\in\mathcal{B}(S)$, $\mu\in P_A$ and $\lambda\in P_B$,

$$q(E|s,\mu,\lambda)=\iint q(E|s,a,b)\,d\mu(a)\,d\lambda(b). \tag{3. 3}$$

Then, by (A1), P_A and P_B, endowed with weak topology, are compact metric spaces. Moreover, from (A1) and (A3), we can show the following lemmas:

LEMMA 3. 1 *Let $f(s,a,b)$ be a continuous, real-valued function on $S\times A\times B$. Then, $f(s,\mu,\lambda)=\iint f(s,a,b)\,d\mu(a)\,d\lambda(b)$, $s\in S$, $\mu\in P_A$, $\lambda\in P_B$ is a continuous function on $S\times P_A\times P_B$.*

LEMMA 3. 2 *Let u be a bounded, continuous, real-valued function on $X\times Y$, where X is a Borel subset of a Polish space and Y is a compact metric space. Then, $u^*(x)=\max_{y\in Y}u(x,y)$ is continuous. Moreover, $u_*(x)=\min_{y\in Y}u(x,y)$ is also continuous.*

These lemmas are given in [1].

Let $C(S)$ denote the family of all bounded, continuous functions on S. For $u\in C(S)$ we define $\|u\|=\sup_{s\in S}|u(s)|$. Then $(C(S), d)$ is a complete metric space, where $d(u, v)=\|u-v\|$ for each $u, v\in C(S)$. For each $\mu\in P_A$ and $\lambda\in P_B$, we define an operator $L(\mu, \lambda)$ on $C(S)$ as follows: for each $u\in C(S)$ and $s\in S$,

$$L(\mu, \lambda)u(s)=r(s, \mu, \lambda)+\int_S u(s')dq(s'|s, \mu, \lambda) \tag{3. 4}$$

Then, by virtue of Lemma 3. 1, $L(\mu, \lambda)u$ is a continuous function on $S\times P_A\times P_B$. Since $L(\mu, \lambda)u$, P_A and P_B satisfy the conditions of Sion's minimax theorem (Theorem 3. 4 of [4]), we can show that for each $u\in C(S)$,

$$\sup_{\mu\in P_A}\inf_{\lambda\in P_B} L(\mu, \lambda)u(s)=\inf_{\lambda\in P_B}\sup_{\mu\in P_A} L(\mu, \lambda)u(s). \tag{3. 5}$$

Moreover, since $L(\mu, \lambda)u(s)$ is continuous on $S\times P_A\times P_B$ and P_A and P_B are compact, sup and inf can be replaced by max and min, respectively, and we can prove the following lemma under (A1), (A2) and (A3).

LEMMA 3. 3 *For each $u\in C(S)$, there exist Borel measurable maps μ_* and λ_* from S into P_A and P_B, such that*

$$\begin{aligned}\min_{\lambda\in P_B} L(\mu_*, \lambda)u(s)&=\max_{\mu\in P_A}\min_{\lambda\in P_B} L(\mu, \lambda)u(s) \\ &=\min_{\lambda\in P_B}\max_{\mu\in P_A} L(\mu, \lambda)u(s) \\ &=\max_{\mu\in P_A} L(\mu, \lambda_*)u(s) \\ &=L(\mu_*, \lambda_*)u(s).\end{aligned} \tag{3. 6}$$

The proof of this lemma is stated in Lemma 2. 4 of [1].

Then, under (A1), (A2), (A3) and (A4), we can prove Theorem 3. 1, Theorem 3. 2 and Theorem 3. 3 by using the lemmas.

THEOREM 3. 1 *Our Markov game has a value in a sense of* (2. 6), *i.e., for all $s\in S$,*

$$\inf_{\sigma\in\Gamma}\sup_{\pi\in\Pi}\overline{\lim_{n\to\infty}}\frac{I_n(\pi, \sigma)(s)}{n}=\sup_{\pi\in\Pi}\inf_{\sigma\in\Gamma}\underline{\lim_{n\to\infty}}\frac{I_n(\pi, \sigma)(s)}{n}, \tag{3. 7}$$

and player I and player II have optimal stationary strategies.

PROOF. By (A4) and Lemma 3. 3, there exists a Borel measurable map μ_* from S into P_A such that

$$\begin{aligned}d+u(s)&=\sup_{\mu\in P_A}\inf_{\lambda\in P_B}\left\{r(s, \mu, \lambda)+\int_S u(s')dq(s'|s, \mu, \lambda)\right\} \\ &=\sup_{\mu\in P_A}\inf_{\lambda\in P_B} L(\mu, \lambda)u(s)\end{aligned}$$

$$= \inf_{\lambda \in P_B} L(\mu_*, \lambda) u(s).$$

For a pair (π, σ) of the strategies of player I and player II, let $E_{\pi,\sigma}$ be an integral oparator associated with a probability measure on S generated by strategies π and σ. Then, for a stationary strategy $\mu_*{}^\infty = (\mu_*, \mu_*, \ldots)$ for player I and any strategy σ for player II,

$$E_{\mu_*{}^\infty, \sigma}\left\{\sum_{t=2}^{n+1}[u(s_t) - E_{\mu_*{}^\infty, \sigma}[u(s_t)|h_{t-1}]]\right\} = 0, \tag{3. 8}$$

where s_t is the state of the system on tth time.
But, for each t, it holds that

$$\begin{aligned}
&E_{\mu_*{}^\infty, \sigma}[u(s_t)|h_{t-1}] \qquad (3.\ 9)\\
&= \int_S u(s')\, dq(s'|s_{t-1}, \mu_*(s_{t-1}), \lambda_{t-1})\\
&= r(s_{t-1}, \mu_*(s_{t-1}), \lambda_{t-1}) + \int_S u(s')\, dq(s'|s_{t-1}, \mu_*(s_{t-1}), \lambda_{t-1})\\
&\quad - r(s_{t-1}, \mu_*(s_{t-1}), \lambda_{t-1})\\
&\geqq \min_{\lambda \in P_B}\left\{ r(s_{t-1}, \mu_*(s_{t-1}), \lambda) + \int_S u(s')\, dq(s'|s_{t-1}, \mu_*(s_{t-1}), \lambda)\right\}\\
&\quad - r(s_{t-1}, \mu_*(s_{t-1}), \lambda_{t-1})\\
&= \min_{\lambda \in P_B} L(\mu_*, \lambda) u(s_{t-1}) - r(s_{t-1}, \mu_*(s_{t-1}), \lambda_{t-1})\\
&= d + u(s_{t-1}) - r(s_{t-1}, \mu_*(s_{t-1}), \lambda_{t-1}),
\end{aligned}$$

where λ_{t-1} denotes a probability measure on B determined by $\sigma_{t-1}(\cdot|h_{t-1})$.
Hence, by (3. 8) and (3. 9), we have

$$0 \leqq E_{\mu_*{}^\infty, \sigma}\left\{\sum_{t=2}^{n+1}[u(s_t) - (d + u(s_{t-1}) - r(s_{t-1}, a_{t-1}, b_{t-1}))]\right\} \tag{3. 10}$$

or

$$d \leqq n^{-1} E_{\mu_*{}^\infty, \sigma}[u(s_{n+1})] - n^{-1} E_{\mu_*{}^\infty, \sigma}[u(s_1)] + n^{-1} I_n(\mu_*{}^\infty, \sigma)(s_1), \tag{3. 11}$$

where

$$I_n(\mu_*, \sigma)(s_1) = \sum_{t=1}^{n} E_{\mu_*{}^\infty, \sigma}[r(s_t, a_t, b_t)].$$

Using the fact that $\|u\| \leqq M$, we get, for any strategy σ for player II and all $s \in S$,

$$d \leqq \varliminf_{n\to\infty} \frac{I_n(\mu_*^\infty, \sigma)(s)}{n}, \tag{3. 12}$$

Thus, from (3. 12), it holds that for all $s \in S$,

$$d \leqq \inf_{\sigma\in\Gamma} \varliminf_{n\to\infty} \frac{I_n(\mu_*^\infty, \sigma)(s)}{n} \leqq \sup_{\pi\in\Pi} \inf_{\sigma\in\Gamma} \varliminf_{n\to\infty} \frac{I_n(\pi, \sigma)(s)}{n}. \tag{3. 13}$$

Similarly, we get, for all $s \in S$,

$$d \geqq \sup_{\pi\in\Pi} \varlimsup_{n\to\infty} \frac{I_n(\pi, \lambda_*^\infty)(s)}{n} \geqq \inf_{\sigma\in\Gamma} \sup_{\pi\in\Pi} \varlimsup_{n\to\infty} \frac{I_n(\pi, \sigma)(s)}{n}. \tag{3. 14}$$

where λ_* is a Borel measurable map from S into P_B satisfying the equation in Lemma 3. 3. On the other hand, it is generally true that, for all $s \in S$,

$$\sup_{\pi\in\Pi} \inf_{\sigma\in\Gamma} \varliminf_{n\to\infty} \frac{I_n(\mu, \sigma)(s)}{n} \leqq \inf_{\sigma\in\Gamma} \sup_{\pi\in\Pi} \varliminf_{n\to\infty} \frac{I_n(\pi, \sigma)(s)}{n} \tag{3. 15}$$

$$\leqq \inf_{\sigma\in\Gamma} \sup_{\pi\in\Pi} \varlimsup_{n\to\infty} \frac{I_n(\pi, \sigma)(s)}{n}.$$

By (3. 13), (3. 14) and (3. 15), we have a constant d as the value of the game and μ_*^∞ and λ_*^∞ are optimal stationary strategies for player I and player II, respectivery. Thus the proof is complete.

Now, we define an operator T on $C(S)$ as follows: for each $u \in C(S)$,

$$Tu(s) = \sup_{\mu\in P_A} \inf_{\lambda\in P_B} L(\mu, \lambda)u(s). \tag{3. 16}$$

Then, by Lemma 3. 2, we can define a sequence of bounded, continuous functions $d_n \in C(S)$, $n=0, 1, 2, \ldots$, suth that, for each $s \in S$,

$$d_0(s) = \sup_{\mu\in P_A} \inf_{\lambda\in P_B} r(s, \mu, \lambda) \tag{3. 17}$$

and

$$d_{n+1}(s) = Td_n(s), \quad n=0, 1, 2, \ldots. \tag{3. 18}$$

THEOREM 3. 2 *There exists a constant M such that, for all $s \in S$ and n,*

$$|d_n(s) - nd| \leqq 2M. \tag{3. 19}$$

PROOF. Since $u(s)$ and $r(s, a, b)$ are bounded, there exists a constant M such that, for all $s \in S$, $|u(s)| \leqq M$ and $u(s) - M \leqq d_0(s) \leqq u(s) + M$.
By (3. 18) and (A4), we have, for all $s \in S$,

$$d_1(s) = Td_0(s)$$

$$= \max_{\mu\in P_A} \min_{\lambda\in P_B} L(\mu, \lambda)d_0(s)$$

$$\leqq \max_{\mu \in P_A} \min_{\lambda \in P_B} L(\mu, \lambda)(u(s)+M)$$

$$= d+u(s)+M.$$

By the same way, we get, for all $s \in S$,

$$d_1(s) \geqq (d+u(s))-M.$$

Repeating the above calculation, we can obtain, for all $s \in S$ and n,

$$nd+u(s)-M \leqq d_n(s) \leqq nd+u(s)+M. \tag{3. 20}$$

By (3. 20), we get (3. 19). Thus, the proof is complete.

In order to find ε-optimal strategies, for any strategies $\pi=(\mu_1, \mu_2, \cdots)$ and $\sigma=(\lambda_1, \lambda_2, \cdots)$, we define ${}^n\pi$ and ${}^n\sigma$ as follows: ${}^n\pi=(\mu_1, \mu_2, \cdots, \mu_n)$ and ${}^n\sigma=(\lambda_1, \lambda_2, \cdots, \lambda_n)$.

THEOREM 3. 3 *For any fixed $\varepsilon>0$, then there exist a number N and strategies π, σ for player I, II, respectively, such that, for all $s \in S$ and all $n \geqq N$,*

$$d \leqq \frac{I_n({}^n\pi^\varepsilon, \sigma)(s)}{n} + \varepsilon \qquad \textit{for any Markov strategy } \sigma \tag{3. 21}$$

and

$$d \geqq \frac{I_n(\pi, {}^n\sigma^\varepsilon)(s)}{n} - \varepsilon \qquad \textit{for any Markov strategy } \pi, \tag{3. 22}$$

where a strategy π is said to be Markov if, for all n, π_n is a Borel measurable map from S into P_A.

PROOF. If we take n_0 such that $n_0>4M/\varepsilon$, then for all $n \geqq n_0$, it holds that

$$\left| d - \frac{d_n(s)}{n} \right| \leqq \frac{2M}{n} \leqq \frac{2M}{n_0} < \frac{\varepsilon}{2}. \tag{3. 23}$$

On the other hand, by the definition of d_n and Lemma 3. 3, there exists a Borel measurable map $\mu_n{}^\varepsilon$ from S into P_A such that

$$d_n(s) = Td_{n-1}(s)$$

$$= \sup_{\mu \in P_A} \inf_{\lambda \in P_B} \left\{ r(s, \mu, \lambda) + \int_S d_{n-1}(s')\, dq(s' \mid s, \mu, \lambda) \right\}$$

$$= \inf_{\lambda \in P_B} \left\{ r(s, \mu_n{}^\varepsilon, \lambda) + \int_S d_{n-1}(s')\, dq(s' \mid s, \mu_n{}^\varepsilon, \lambda) \right\}$$

$$\leqq L(\mu_n{}^\varepsilon, \lambda_n) d_{n-1}(s)$$

for any Borel measurable map λ_n from S into P_B.
Thus, repeating the above calculation, we get

$$d_n(s) \leqq L(\mu_n{}^\varepsilon, \lambda_n) d_{n-1}(s)$$

$$\leqq L(\mu_n{}^\varepsilon, \lambda_n)L(\mu_{n-1}{}^\varepsilon, \lambda_{n-1})d_{n-2}(s)$$

$$\vdots$$

$$\leqq \prod_{i=0}^{n-1} L(\mu_{n-1}{}^\varepsilon, \lambda_{n-1})d_0(s)$$

$$= I_n({}^n\pi^\varepsilon, {}^n\sigma)(s)+\int_S d_0(s')dq(s'|s, {}^n\pi^\varepsilon, {}^n\sigma),$$

where ${}^n\pi^\varepsilon=(\mu_n{}^\varepsilon, \mu_{n-1}{}^\varepsilon, \ldots, \mu_1{}^\varepsilon)$ and ${}^n\sigma=(\lambda_n, \lambda_{n-1}, \ldots, \lambda_1)$.
Since $d_n(s)\geqq nd-n\varepsilon/2$ and $\|d_0\|\leqq M$, by (3. 24),

$$nd\leqq I_n({}^n\pi^\varepsilon, {}^n\sigma)(s)+M+n\varepsilon/2. \tag{3. 25}$$

Hence, we get, for all sufficiently large n and all $s\in S$,

$$d\leqq \frac{I_n({}^n\pi^\varepsilon, \sigma)(s)}{n}+\varepsilon \qquad \text{for any Markov strategy } \sigma. \tag{3. 26}$$

Similarly, for all sufficiently large n and all $s\in S$,

$$d\geqq \frac{I_n(\pi, {}^n\sigma^\varepsilon)(s)}{n}-\varepsilon \qquad \text{for any Markov strategy } \pi. \tag{3. 27}$$

Thus, the proof is complete.

Next, we shall show that, under (A1), (A2), (A3) and some conditions, our Markov game with the expected average reward criterion is reduced to one with some specified discount factor, In order to show the fact, we need impose on q the following additional assumption: (A5) there exists a state s_0 and $1>\alpha>0$ such that, for all $s\in S$, $a\in A$ and $b\in B$,

$$q(\{s_0\}|s, a, b)\geqq\alpha. \tag{3. 28}$$

Then, from (3. 28), it holds that, for all $\mu\in P_A$ and $\lambda\in P_B$,

$$q(\{s_0\}|s, \mu, \lambda)\geqq\alpha. \tag{3. 29}$$

For Markov game with the law of motion of the system satisfying the above assumption, consider a new Markov game with identical state and action space, with identical rewards, but with the law of motion of the system given for $E\in\mathfrak{B}(S)$ by

$$q'(E|s, a, b)=\begin{cases}\dfrac{q(E|s, a, b)}{1-\alpha} & \text{for } s_0\notin E\\[2ex] \dfrac{q(E|s, a, b)-\alpha}{1-\alpha} & \text{for } s_0\in E.\end{cases}$$

We now call the game (S, A, B, q, r) original Markov game and the game (S, A, B, q', r) modified Markov game, or simply "original M. G." and "modified M. G.", respectively.

It should be noted that $q'(\cdot|s, a, b)$ also satisfies the assumption (A3) since for all $u\in C(S)$, $a\in A$ and $b\in B$,

$$\int_S u(s')dq'(s'|s, a, b)=\frac{1}{1-\alpha}\int_S u(s')dq(s'|s, a, b)-\frac{\alpha}{1-\alpha}u(s_0) \tag{3. 30}$$

Since modified M. G. (S, A, B, q', r) satisfies the assumptions (A1), (A2) and (A3), we have the following lemma by the results of [1].

LEMMA 3. 4 *The modified M. G. (S, A, B, q', r) with a discount factor β, $0\leqq\beta<1$, has a value, the value function is continuous, and player I and player II have optimal stationary strategies.*

Let $f_\beta^{*(\infty)}$ and $g_\beta^{*(\infty)}$ be optimal stationary strategies for player I and player II in the modified M. G. with a discount factor $0\leqq\beta<1$, respectively, and $w_\beta^*(s)$ be the value of the game. Then, for f_β^*, g_β^* and $w_\beta^*(s)$, it holds that

$$w_\beta^*(s)=\min_{\lambda\in P_B}\max_{\mu\in P_A}\left\{r(s, \mu, \lambda)+\beta\int_S w_\beta^*(s')dq'(s'|s, \mu, \lambda)\right\} \tag{3. 31}$$

$$=\max_{\mu\in P_A}\min_{\lambda\in P_B}\left\{r(s, \mu, \lambda)+\beta\int_S w^*(s')dq'(s'|s, \mu, \lambda)\right\}$$

$$=r(s, f_\beta^*(s), g_\beta^*(s))+\beta\int_S w^*(s')\,dq'(s'|s, f_\beta^*(s), g_\beta^*(s))$$

Then we can prove the following theorem.

THEOREM 3. 4 *Under the assumptions (A1), (A2), (A3) and (A5), there exist optimal stationary strategies for player I and player II for the original M. G. with the expected average reward criterion and the value of the game is $\alpha I'_{1-\alpha}(f_{1-\alpha}^{*(\infty)}, g_{1-\alpha}^{*(\infty)})(s_0)$. Further, the optimal stationary strategies for player I and player II are the sequences of Borel measurable maps from S into P_A and P_B, respectively, satisfying the following equation:*

$$\alpha w^*_{1-\alpha}(s_0)+u'(s)=\max_{\mu\in P_A}\min_{\lambda\in P_B}\left\{r(s, \mu, \lambda)+\int_S u'(s')\,dq(s'|s, \mu, \lambda)\right\} \tag{3. 32}$$

where

$$u'(s)=w^*_{1-\alpha}(s)-w^*_{1-\alpha}(s_0) \tag{3. 33}$$

PROOF. A. Maitra and T. Parthasarathy [1] have shown that, since $w^*_{1-\alpha}$ is the value of the modified M. G. with a discount factor $1-\alpha$, $w^*_{1-\alpha}$ satisfies the min max equation, namely, for all $s\in S$,

$$w^*_{1-\alpha}(s)=\max_{\mu\in P_A}\min_{\lambda\in P_B}\left\{r(s, \mu, \lambda)+(1-\alpha)\int_S w^*_{1-\alpha}(s')\,dq'(s'|s, \mu, \lambda)\right\}.$$

Using (3. 33), we get

$$u'(s)+w^*_{1-\alpha}(s_0)=(1-\alpha)w^*_{1-\alpha}(s_0)+\max_{\mu\in P_A}\min_{\lambda\in P_B}\Big\{r(s,\mu,\lambda)+$$

$$+(1-\alpha)\int_S u'(s')\,dq'(s'|s,\mu,\lambda)\Big\}. \tag{3. 34}$$

From $u'(s_0)=0$, (3. 34) yields that

$$\alpha w^*_{1-\alpha}(s_0)+u'(s)=\max_{\mu\in P_A}\min_{\lambda\in P_B}\Big\{r(s,\mu,\lambda)+\int_S u'(s')\,dq(s'|s,\mu,\lambda)\Big\}. \tag{3. 35}$$

This shows that the original M. G. satisfies, also, the assumption (A4). Hence, by Theorem 3. 1, a value of the game is $\alpha w^*_{1-\alpha}(s_0)$ and there exist optimal stationary strategies for player I and player II. Thus, the proof is complete.

In order to give sufficient conditions for the existence of the assumption (A4), fix some state s_0 and let

$$u_\beta(s)=w^*_\beta(s)-w^*_\beta(s_0). \tag{3. 36}$$

Then, from (3. 31) and (3. 36), it holds that

$$d_\beta+u_\beta(s)=\max_{\mu\in P_A}\min_{\lambda\in P_B}\Big\{r(s,\mu,\lambda)+\beta\int_S u_\beta(s')\,dq(s'|s,\mu,\lambda)\Big\}, \tag{3. 37}$$

where $d_\beta=(1-\beta)w^*_\beta(s_0)$.

We can prove the following theorem under the assumptions (A1), (A2) and (A3).

THEOREM 3. 5 *If $\{u_\beta\}$ is a uniformly bounded equicontinuous family of functions, then it follows that (A4) holds and $(1-\beta)w^*_\beta(s)$ converges to d as $\beta\longrightarrow 1^-$ for all $s\in S$.*

PROOF. By the Ascoli-Arzela's theorem there exist a sequence $\beta_\nu\longrightarrow 1$ and a continuous function $u(s)$ such that $u_{\beta_\nu}(s)$ converges uniformly to $u(s)$ on S. Now, since u_β is bounded, we can also require that d_{β_ν} converges to d. Hence, from (3. 37), we get

$$d+u(s)=\max_{\mu\in P_A}\min_{\lambda\in P_B}\Big\{r(s,\mu,\lambda)+\int_S u(s')\,dq(s'|s,\mu,\lambda)\Big\}. \tag{3. 38}$$

For any sequence $\beta_\nu\longrightarrow 1$, there is a subsequence β'_ν such that $\lim d_{\beta'_\nu}$ exists. By the above this limit is d. Thus, $d=\lim_{\beta\to 1} d_\beta$. The result follows since s_0 is any arbitrary state. Thus, the proof is complete.

4. Acknowledgment

The authors are deeply indebted to Professor N. Furukawa of Kyushu University for his valuable advices.

References

[1] A. MAITRA and T. PARTHASARATHY, *On stochastic games*, Journ. Opti. Theory and its Appl., 5 (1970), 289–300.

[2] T. PARTHASARATHY and T. E. RAGHAVEN, Some topics in two-pesron games, American Elsevier, New York, 1971.

[3] M. Ross, *Arbitrary state Markovian decision processes*, Ann. Math. Statist., 39 (1968), 2118–2122.

[4] M. SION, *On general minimax theorems*, Pacific J. Math., 8 (1958), 171–176.