On continuous time Markov games with the expected average reward criterion

By

Kensuke TANAKA* and Kazuyoshi WAKUTA**

(Received November 8, 1976)

1. Introduction

This paper is concerned with continuous time Markov game in which the state space is countable and the action spaces of player I and player II are compact metric spaces. In the game, the players continuously observe the state of the system and then the players choose independently actions. As a result, the reward is paid to player I from player II and the system moves to a new state which is governed by the known transition probability rates. Then, the optimization problem is to maximize the long-run expected average gain of player I as the game proceeds over the infinite future and to minimize the long-run expected average loss of player II.

However, so far as we know, such the game has not been tried up to the present. Hence, at first, we shall give the formulation of a continuous time Markov game with the criterion of long-run expected average reward. Then, we shall show that the game has a value and there exist the optimal stationary strategies for both players under this criterion and some assumptions. Moreover, we shall give the sufficient conditions for some important assumption.

2. The formulation of the problem

In this paper, we determine "*continuous time Markov game*" by five objects (S, A, B, q, r). Here, S is a countable state space labeled $\{1, 2, 3, \cdots\cdots\}$, 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 transition probability rates which govern the law of motion of the system and is a bounded function $q(\cdot \mid i, a, b)$ on S for each triple $(s, a, b) \in S \times A \times B$; r, the reward function, is a bounded Borel measurable function on $S \times A \times B$.

In this game, player I and player II continuously observe the state of the system and

* Niigata University.
** Nagaoka Technical College.

classify it into one of the possible state $i \in S$, then player I and player II choose actions $a \in A$ and $b \in B$, respectively. As a consequence of the present state i and the actions a and b chosen by the players, player II pays player I reward $r(i, a, b)$ unit of money and the system moves to a new state $j \in S$, which is governed by the transition probability rate $q(j \mid i, a, b)$. Then, our optimization problem is to maximize the expected average gain of player I as the game proceeds over the infinite future and to minimize the expected average loss of player II.

We assume that the strategies for player I and playre II are independent of the past history of the system and depend only on the present state of the system. Such a strategy $\pi = \pi(t)$ for player I is specified by a family $\{\mu_t\}$, where μ_t is a probability measure $\mu_t(\cdot \mid i)$ on the Borel measurable space $(A, \mathcal{B}(A))$ in which $\mathcal{B}(A)$ is the σ-field generated by the metric on A for each $i \in S$ and $t \in [0, \infty)$ and μ_t is a Lebesgue measurable function $\mu_\cdot(M \mid i)$ on $[0, \infty)$ for each $M \in \mathcal{B}(A)$ and $i \in S$. Then, we call such a strategy a Markov strategy. Moreover, such a Markov strategy $\pi = \pi(t)$ is said to be stationary if $\pi(t)$ is independent of t, that is, there exists a map μ from S into P_A such that $\mu_t = \mu$ for all $t \in [0, \infty)$, where P_A is the set of all probability measures on $(A, \mathcal{B}(A))$. Π denotes the class of all Markov strategies for player I. Markov strategies and stationary strategies for player II are defined analogously. Γ denotes the class of all Markov strategies for player II.

Throughout the paper, we shall assume, for the transition rate matrix $Q(a, b) = \{q(j \mid i, a, b); i, j \in S\}$ corresponding to $a \in A$ and $b \in B$, the following

ASSUMPTION 1. For each $i, j \in S$, $q(j \mid i, a, b)$ is a continuous function on $A \times B$ and for all $a \in A$, $b \in B$, $q(j \mid i, a, b) \geqq 0$, $i \neq j$, $\sum_j q(j \mid i, a, b) = 0$ and $|q(i \mid i, a, b)| \leqq M$ for all $i \in S$ and a positive number $M < \infty$.

When a pair of the Markov strategies (π, σ) for player I and player II is used, the transition probability rates are defined as follows: for each $t \in [0, \infty)$

$$q(j \mid i, t, \pi, \sigma) = \iint q(j \mid i, a, b) d\mu_t(a \mid i) d\lambda_t(b \mid i),$$

where the strategies π and σ are specified by the families $\{\mu_t\}$ and $\{\lambda_t\}$, respectively. Then, for all $i, j \in S$ and $t \in [0, \infty)$, $q(j \mid i, t, \pi, \sigma)$ satisfies also the following conditions

$$q(j \mid i, t, \pi, \sigma) \geqq 0, \ j \neq i \quad \sum_j q(j \mid i, t, \pi, \sigma) = 0 \tag{2. 1}$$

and

$$|q(i \mid i, t, \pi, \sigma)| \leqq M. \tag{2. 1}$$

We write the transition rate matrix corresponding to π and σ as $Q(t, \pi, \sigma) = \{q(j \mid i, t, \pi, \sigma); i, j \in S\}$ and, if π and σ are stationary strategies, we write $Q(\pi, \sigma)$ instead of $Q(t, \pi, \sigma)$. Under the conditions (2. 1) and (2. 2), it has already been showed in [1] to exist a unique stochastic transition probability matrix $F(s, t, \pi, \sigma) = \{f_{ij}(s, t, \pi, \sigma); i, j \in S\}$ corresponding to $Q(t, \pi, \sigma)$. Moreover, $F(s, t, \pi, \sigma)$ satisfies the kolmogorov forward differential

equations

$$\frac{\partial}{\partial t} F(s, t, \pi, \sigma) = F(s, t, \pi, \sigma)\ Q(t, \pi, \sigma) \tag{2. 3}$$

with $F(s, s, \pi, \sigma) = I$, for almost all $t \in [s, \infty)$, where I is the infinite unit matrix.

Also, a measurable Markov process $\{X(t, \pi, \sigma);\ t \geqq 0\}$ corresponding to the stochastic transition probability matrix $F(s, t, \pi, \sigma)$ exists and is well-behaved. Throughout our discussion the game starts from the origin. In view of this we write $F(t, \pi, \sigma)$ instead of $F(0, t, \pi, \sigma)$.

Now, we define the expected average criterion function. When a pair of the Markov strategies (π, σ) is chosen by player I and player II, at any time t the expected gain rate out of state $i \in S$ is given by

$$r(i, t, \pi, \sigma) = \iint r(i, a, b)\, d\mu_t(a \mid i)\, d\lambda_t(b \mid i). \tag{2. 4}$$

It is clear that $r(i, t, \pi, \sigma)$ is a Lebesgue measurable function of t. Thus, when the system starts from a state $i \in S$ and a pair of the Markov strategies (π, σ) is used, the total expected gain of player I up to the time T is defined to be

$$\phi(i, T, \pi, \sigma) = \int_0^T \sum_j f_{ij}(t, \pi, \sigma)\ r(j, t, \pi, \sigma)\, dt. \tag{2. 5}$$

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

$$\inf_{\sigma \in \Gamma} \sup_{\pi \in \Pi} \overline{\lim_{T \to \infty}} \frac{\phi(i, T, \pi, \sigma)}{T} \leqq \lim_{T \to \infty} \frac{\phi(i, T, \pi^*, \sigma')}{T}.$$

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

$$\sup_{\pi \in \Pi} \inf_{\sigma \in \Gamma} \lim_{T \to \infty} \frac{\phi(i, T, \pi, \sigma)}{T} \geqq \overline{\lim_{T \to \infty}} \frac{\phi(i, T, \pi', \sigma^*)}{T}.$$

We shall say that a continuous time Markov game has a value if for all $i \in S$

$$\sup_{\pi \in \Pi} \inf_{\sigma \in \Gamma} \lim_{T \to \infty} \frac{\phi(i, T, \pi, \sigma)}{T} = \inf_{\sigma \in \Gamma} \sup_{\pi \in \Pi} \overline{\lim_{T \to \infty}} \frac{\phi(i, T, \pi, \sigma)}{T}.$$

When the game has a value, the quantity

$$\sup_{\pi \in \Pi} \inf_{\sigma \in \Gamma} \lim_{T \to \infty} \frac{\phi(i, T, \pi, \sigma)}{T}$$

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

3. On the existence of optimal stationary strategies

Throughout our discussion in this section, we shall assume the following:

ASSUMPTION 2. A and B are compact metric spaces and, for each $i \in S$, $r(i, a, b)$ is a continuous function on $A \times B$.

ASSUMPTION 3. There exist a bounded function ν on S and a constant g such that, for all $i \in S$,

$$g = \sup_{\mu \in P_A} \inf_{\lambda \in P_B} \{r(i, \mu, \lambda) + \sum_j q(j \mid i, \mu, \lambda)\nu(j)\}, \tag{3, 1}$$

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

$$r(i, \mu, \lambda) = \int\int r(i, a, b)\, d\mu(a)\, d\lambda(b)$$

and

$$q(j \mid i, \mu, \lambda) = \int\int q(j \mid i, a, b)\, d\mu(a)\, d\lambda(b).$$

Hence, by Assumption 2, P_A and P_B endowed with weak topology are compact metric spaces and, for each $i \in S$ and $j \in S$, by Assumption 1 and Assumption 2, $r(i, \mu, \lambda)$ and $q(j \mid i, \mu, \lambda)$ are continuous bounded functions on $P_A \times P_B$.

Then, under Assumption 1 and Assumption 2, we can prove the following theorem.

THEOREM 3. 1. *For a pair of the stationary strategies (μ, λ), if there exist a bounded function ν on S and a constant g such that, for each $i \in S$.*

$$g = r(i, \mu, \lambda) + \sum_j q(j \mid i, \mu, \lambda)\nu(j), \tag{3. 2}$$

then, it holds that, for each $i \in S$,

$$g = \lim_{T \to \infty} \frac{\phi(i, T, \mu, \lambda)}{T}.$$

PROOF. Let $\{f_{li}(t, \mu, \lambda); l, i \in S\}$ be the transition probability matrix corresponding to $Q(\mu, \lambda)$. Multiplying both sides of (3. 2) by $f_{li}(t, \mu, \lambda)$ and summing over all $i \in S$, we have, for each $l \in S$,

$$g = \sum_i f_{li}(t, \mu, \lambda) r(i, \mu, \lambda) + \sum_i f_{li}(t, \mu, \lambda) \sum_j q(j \mid i, \mu, \lambda)\nu(j). \tag{3. 3}$$

Since $\sum_i \sum_j | f_{li}(t, \mu, \lambda) q(j \mid i, \mu, \lambda) | < \infty$, the summation signs in the seconds term of (3. 3) can be interchanged.

Using the Kolmogorov forward differential equation (2. 3), we get for each $l \in S$,

$$g = \sum_i f_{li}(t, \mu, \lambda)\, r(i, \mu, \lambda) + \sum_j \frac{\partial f_{lj}(t, \mu, \lambda)}{\partial t} \nu(j). \tag{3. 4}$$

By integrating on both sides of (3. 4) with respect to t from 0 to $T < \infty$, we have

$$Tg = \int_0^T \sum_i f_{li}(t, \mu, \lambda)\, r(i, \mu, \lambda)\, dt + \int_0^T \sum_j \frac{\partial f_{lj}(t, \mu, \lambda)}{\partial t} \nu(j)\, dt$$

$$= \int_0^T \sum_i f_{li}(t, \mu, \lambda)\, r(i, \mu, \lambda)\, dt + \sum_j f_{lj}(T, \mu, \lambda)\nu(j) - \nu(l).$$

Dividing by T and taking the limit as $T \to \infty$, we get

$$g=\lim_{T\to\infty}\frac{\phi(l, T, \mu, \lambda)}{T} \qquad \text{for each} \quad l\in S.$$

Thus, the proof is complete.

Moreover, the following lemma is important.

LEMMA. *If ν is a bounded function on S, for each $i\in S$, $\sum_j q(j|i, a, b)\nu(j)$ converges uniformly in a and b. As this result, $\sum_j q(j|i, a, b)\nu(j)$ is a bounded continuous function on $A\times B$.*

The proof of the lemma is given in our paper [9].

Next, for ν in Assumption 3, we define

$$K(i, \mu, \lambda)\equiv r(i, \mu, \lambda)+\sum_j q(j|i, \mu, \lambda)\,\nu(j).$$

Then, by the lemma, $K(i, \mu, \lambda)$ is a continuous function on $P_A\times P_B$ for each $i\in S$. $K(i, \mu, \lambda)$, P_A and P_B satisfy the conditions of Sion's minimax theorem (Theorem 3. 4 of [6]) because of its bilinearity in (μ, λ) and, consequently, for each $i\in S$.

$$\sup_{\mu\in P_A}\inf_{\lambda\in P_B} K(i, \mu, \lambda)=\inf_{\lambda\in P_B}\sup_{\mu\in P_A} K(i, \mu, \lambda).$$

Moreover, since $K(i, \mu, \lambda)$ is continuous on $P_A\times P_B$ for each $i\in S$ and P_A, P_B are compact, sup and inf can be replaced by max and min, respectively. Thus, (3. 1) is written as follows: for each $i\in S$

$$g=\max_{\mu\in P_A}\min_{\lambda\in P_B} K(i, \mu, \lambda)=\min_{\lambda\in P_B}\max_{\mu\in P_A} K(i, \mu, \lambda).$$

From [5], there exist maps μ^* and λ^* from S into P_A and P_B, respectively, such that, for each $i\in S$,

$$\begin{aligned}\min_{\lambda\in P_B} K(i, \mu^*, \lambda) &= \max_{\mu\in P_A}\min_{\lambda\in P_B} K(i, \mu, \lambda) \qquad (3.\ 5)\\ &= \min_{\lambda\in P_B}\max_{\mu\in P_A} K(i, \mu, \lambda)\\ &= \max_{\mu\in P_A} K(i, \mu, \lambda^*).\end{aligned}$$

Then, under Assumption 1, Assumption 2 and Assumption 3, we can prove the following theorem.

THEOREM 3. 2. *The game has a value and both players have optimal stationary strategies.*

PROOF. From (3. 5), there exists a map μ^* from S into P_A such that, for each $i\in S$,

$$\begin{aligned}g &= \max_{\mu\in P_A}\min_{\lambda\in P_B}\{r(i, \mu, \lambda)+\sum_j q(j|i, \mu, \lambda)\,\nu(j)\} \qquad (3.\ 6)\\ &= \min_{\lambda\in P_B}\max_{\mu\in P_A} K(i, \mu, \lambda)\end{aligned}$$

$$=\min_{\lambda\in P_B} K(i, \mu^*, \lambda)$$

and, moreover, (3. 6) is written as follows: for each $i\in S$ and all $\lambda\in P_B$

$$g\leqq r(i, \mu^*, \lambda)+\sum_j q(j\,|\,i, \mu^*, \lambda)\nu(j).$$

Hence, for a stationary strategy μ^* for player I and any Markov strategy σ for player II, we have for each $t\in[0, \infty)$

$$g\leqq r(i, t, \mu^*, \sigma)+\sum_j q(j\,|\,i, t, \mu^*, \sigma)\nu(j). \tag{3. 7}$$

Now, let the Kolmogorov forward differential equation corresponding to the strategies (μ^*, σ) be

$$\frac{\partial f_{ij}(t, \mu^*, \sigma)}{\partial t}=\sum_l f_{il}(t, \mu^*, \sigma)q(j\,|\,l, t, \mu^*, \sigma) \tag{3. 8}$$

$$\text{for each}\quad i, j\in S \text{ and } t\geqq 0.$$

Multiplying both sides of (3. 8) by $\nu(j)$ and summing over all $j\in S$, we obtain, for each $i\in S$ and $t\geqq 0$,

$$\sum_j \frac{\partial f_{ij}(t, \mu^*, \sigma)}{\partial t}\nu(j)=\sum_j\sum_l f_{il}(t, \mu^*, \sigma)\, q(j\,|\,l, t, \mu^*, \sigma)\,\nu(j). \tag{3. 9}$$

Since $\sum_j\sum_l |f_{il}(t, \mu^*, \sigma)q(j\,|\,l, t, \mu^*, \sigma)\,\nu(j)\,|<\infty$, the summation signs in the right-hand side can be interchanged. Using (3. 7), we obtain, for each $i\in S$,

$$\begin{aligned}\sum_j \frac{\partial f_{ij}(t, \mu^*, \sigma)}{\partial t}\nu(j)&=\sum_l f_{il}(t, \mu^*, \sigma)\sum_j q(j\,|\,l, t, \mu^*, \sigma)\nu(j) \qquad (3.10)\\ &\geqq\sum_l f_{il}(t, \mu^*, \sigma)[g-r(l, t, \mu^*, \sigma)]\\ &=g-\sum_l f_{il}(t, \mu^*, \sigma)r(l, t, \mu^*, \sigma).\end{aligned}$$

Also, (3. 10) is written as follows

$$g\leqq\sum_l f_{il}(t, \mu^*, \sigma)\, r(l, t, \mu^*, \sigma)+\sum_j \frac{\partial f_{ij}(t, \mu^*, \sigma)}{\partial t}\nu(j). \tag{3.11}$$

By integrating of both sides of (3. 11) with respect to t from 0 to $T<\infty$, we have

$$Tg\leqq\int_0^T\sum_l f_{il}(t, \mu^*, \sigma)\, r(l, t, \mu^*, \sigma)\, dt+\sum_j f_{ij}(T, \mu^*, \sigma)\nu(j)-\nu(i).$$

Dividing by T and taking the inferior limit as $T\to\infty$, we get, for each $i\in S$,

$$g\leqq\varliminf_{T\to\infty}\frac{1}{T}\int_0^T\sum_l f_{il}(t, \mu^*, \sigma)\, r(l, t, \mu^*, \sigma)\, dt. \tag{3.12}$$

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

$$g \leqq \inf_{\sigma \in \Gamma} \lim_{T \to \infty} \frac{\phi(i, T, \mu^*, \sigma)}{T} \leqq \sup_{\pi \in \Pi} \inf_{\sigma \in \Gamma} \lim_{T \to \infty} \frac{\phi(i, T, \pi, \sigma)}{T}. \tag{3.13}$$

Similarly, it holds that, for each $i \in S$.

$$g \geqq \sup_{\pi \in \Pi} \overline{\lim_{T \to \infty}} \frac{\phi(i, T, \pi, \lambda^*)}{T} \geqq \inf_{\sigma \in \Gamma} \sup_{\pi \in \Pi} \overline{\lim_{T \to \infty}} \frac{\phi(i, T, \pi, \sigma)}{T}, \tag{3.14}$$

where λ^* is the stationary strategy defined from (3. 5) for player II.

On the other hand, it is generally true that, for each $i \in S$,

$$\sup_{\pi \in \Pi} \inf_{\sigma \in \Gamma} \lim_{T \to \infty} \frac{\phi(i, T, \pi, \sigma)}{T} \leqq \inf_{\sigma \in \Gamma} \sup_{\pi \in \Pi} \overline{\lim_{T \to \infty}} \frac{\phi(i, T, \pi, \sigma)}{T}. \tag{3.15}$$

By (3. 13), (3. 14) and (3. 15), we have g as the value function of the game and μ^* and λ^* are optimal stationary strategies for player I and player II, respectively. Thus, the proof is complete.

4. The sufficient conditions of Assumption 3

In this section, we need the following results which we have given in our paper [9]. If μ and λ are the stationary strategies of player I and player II, respectively, $\Psi(i, \alpha, \mu, \lambda)$ is the unique bounded solution to

$$\alpha \Psi(i, \alpha, \mu, \lambda) = r(i, \mu, \lambda) + \sum_j q(j \mid i, \mu, \lambda) \Psi(j, \alpha, \mu, \lambda), \tag{4. 1}$$

where

$$\Psi(i, \alpha, \mu, \lambda) = \int_0^\infty e^{-\alpha t} \sum_j f_{ij}(t, \mu, \lambda)\, r(j, \mu, \lambda)\, dt.$$

Further, if μ^* and λ^* are α-discounted optimal stationary strategies of both players, then $\Psi(i, \alpha, \mu^*, \lambda^*)$ is the unique bounded solution to

$$\alpha \Psi(i, \alpha, \mu^*, \lambda^*) = \max_{\mu \in P_A} \min_{\lambda \in P_B} \{ r(i, \mu, \lambda) + \sum_j q(j \mid i, \mu, \lambda) \Psi(j, \alpha, \mu^*, \lambda^*) \}. \tag{4. 2}$$

Now, for a pair of α-discounted optimal stationary strategies (μ^*, λ^*) of both players, we define the notation as follows

$$\nu_{ij}(\alpha, \mu^*, \lambda^*) = \Psi(i, \alpha, \mu^*, \lambda^*) - \Psi(j, \alpha, \mu^*, \lambda^*).$$

We assume the following assumption regarding ν_{ij}.

ASSUMPTION 4. For some sequence $\alpha_m \to 0$ as $m \to \infty$ and for some state $k \in S$, there exists a positive number $M < \infty$ such that, for all $m = 1, 2, \ldots\ldots$ and $i \in S$,

$$|\nu_{ik}(\alpha_m, \mu_m, \lambda_m)| \leqq M.$$

Then, under Assumption 1 and Assumption 2, we can prove the following theorem.

THEOREM 4.1 *Under Assumption 4, there exist a bounded function ν on S and a constant g such that, for all $i \in S$,*

$$g = \max_{\mu \in P_A} \min_{\lambda \in P_B} \{ r(i, \mu, \lambda) + \sum_j q(j|i, \mu, \lambda)\nu(j) \}. \tag{4. 3}$$

PROOF. In Assumption 4, (μ_m, λ_m) is a pair of α_m-discounted optimal stationary strategies of both players. From (2. 1) and (4. 2), $\Psi(i, \alpha_m, \mu_m, \lambda_m)$ satisfies the following equation: for a fixed state $k \in S$, all $m=1, 2, 3, \ldots\ldots$ and $i \in S$,

$$\alpha_m \Psi(i, \alpha_m, \mu_m, \lambda_m) = \max_{\mu \in P_A} \min_{\lambda \in P_B} \{ r(i, \mu, \lambda) + \sum_j q(j|i, \mu, \lambda)\nu_{jk}(\alpha_m, \mu_m, \lambda_m) \}. \tag{4. 4}$$

Since S is a countable space and $|\alpha_m \Psi(i, \alpha_m, \mu_m, \lambda_m)| \leqq \|r\|$, by Assumption 4, there exist the subsequences $\{\mu_{m'}\}$ and $\{\lambda_{m'}\}$ of $\{\mu_m\}$ and $\{\lambda_m\}$, respectively, such that, for all $i \in S$,

$$\lim_{m' \to \infty} \alpha_{m'} \Psi(i, \alpha_{m'}, \mu_{m'}, \lambda_{m'}) = g$$

and for all $j \in S$ and a fixed $k \in S$,

$$\lim_{m' \to \infty} \nu_{jk}(\alpha_{m'}, \mu_{m'}, \lambda_{m'}) = \nu(j)$$

where $\alpha_{m'} \longrightarrow 0$ as $m' \longrightarrow \infty$.

Clearly, ν is a bounded function on S. Since $\sum_j q(j|i, \mu, \lambda)$ converges uniformly in μ and λ by the lemma, taking the limits as $m' \to \infty$ on both sides of (4. 4), we get the equation (4. 3). Thus, the proof is complete.

Moreover, since P_A and P_B are compact from Assumption 2, there exist the subsequences $\{\mu_{m'}\}$ and $\{\lambda_{m'}\}$ of the sequences $\{\mu_m\}$ and $\{\lambda_m\}$ of α_m-discounted optimal stationary strategies of both players such that

$$\lim_{m' \to \infty} \alpha_{m'} \Psi(i, \alpha_{m'}, \mu_{m'}, \lambda_{m'}) = g, \qquad \text{for all } i \in S$$

$$\lim_{m' \to \infty} \nu_{jk}(\alpha_{m'}, \mu_{m'}, \lambda_{m'}) = \nu(j), \quad \text{for a fixed } k \text{ and all } j \in S$$

and

$$\mu_{m'} \Longrightarrow \mu^* \in P_A, \quad \lambda_{m'} \Longrightarrow \lambda^* \in P_B,$$

where the notation $\Longrightarrow$ denotes weak convergence.

Then, we can prove the following theorem.

THEOREM 4. 2 *μ^* and λ^* are the optimal stationary strategies for player* I *and player* II,

respectively.

PROOF. $(\mu_{m'}, \lambda_{m'})$ is a pair of $\alpha_{m'}$-discounted optimal stationary strategies of both players. From (2. 1), (4. 1) and (4. 2), we get for all $i \in S$

$$\alpha_{m'}\Psi(i, \alpha_{m'}, \mu_{m'}, \lambda_{m'}) = r(i, \mu_{m'}, \lambda_{m'}) + \\ + \sum_j q(j|i, \mu_{m'}, \lambda_{m'})\nu_{jk}(\alpha_{m'}, \mu_{m'}, \lambda_{m'}). \qquad (4.\ 5)$$

Since $\sum_j q(j|i, \mu, \lambda)$ converges uniformly in μ and λ by the lemma, taking the limits on both sides of (4. 5) as $m' \to \infty$, we obtain for all $i \in S$

$$g = r(i, \mu^*, \lambda^*) + \sum_j q(j|i, \mu^*, \lambda^*)\nu(j).$$

Thus, the proof is complete.

Next, we impose on q the following assumption.

ASSUMPTION 5. There exist some state $k \in S$ and a positive number δ such that, for all $i \neq k$, $a \in A$ and $b \in B$,

$$q(k|i, a, b) \geqq \delta > 0.$$

Now, under this assumption, we define new transition rates as follows: for each i and $j \in S$,

$$\bar{q}(j|i, a, b) = q(j|i, a, b) + \delta_{ij}\delta \qquad j \neq k$$

$$\bar{q}(k|i, a, b) = q(k|i, a, b) + \delta_{ik}\delta - \delta, \qquad (4.\ 6)$$

where δ_{ij} is the Kronecker delta.

Clearly, new transition probability rate matrix $\bar{Q}(a, b)$ satisfies Assumption 1. Hence, there exists a unique stochastic transition probability matrix $\bar{F}(s, t, \pi, \sigma)$ for any given pair of Markov strategies (π, σ).

Now, we consider a new Markov game $(S, A, B, \bar{q}, r, \delta)$ with identical state space, identical action spaces, identical reward and a discount factor $\delta > 0$. Since this new game satisfies Assumption 1 and Assumption 2, from the result of [9], there exists a pair of optimal stationary strategies (μ^*, λ^*) of both players.

Then, we can prove the following theorem.

THEOREM 4. 3 *Under Assumption 5, there exist a bounded function ν on S and a constant g such that, for all $i \in S$,*

$$g = \max_{\mu \in P_A} \min_{\lambda \in P_B} \{r(i, \mu, \lambda) + \sum_j q(j|i, \mu, \lambda)\nu(j)\}.$$

PROOF. Since μ^* and λ^* are the optimal stationary strategies of both players in new Markov game, from (4. 2), we have for all $i \in S$,

$$\partial\bar{\Psi}(i, \partial, \mu^*, \lambda^*) = \max_{\mu \in P_A} \min_{\lambda \in P_B} \{r(i, \mu, \lambda) + \sum_j \bar{q}(j \mid i, \mu, \lambda)\bar{\Psi}(j, \partial, \mu^*, \lambda^*)\}. \tag{4. 7}$$

Substituting $q(j \mid i, \mu, \lambda)$ for $\bar{q}(j \mid i, \mu, \lambda)$ in (4. 7), we obtain, after some simplification, for all $i \in S$,

$$\partial\bar{\Psi}(k, \partial, \mu^*, \lambda^*) = \max_{\mu \in P_A} \min_{\lambda \in P_B} \{r(i, \mu, \lambda) + \sum_j q(j \mid i, \mu, \lambda)\bar{\Psi}(j, \partial, \mu^*, \lambda^*)\}.$$

Thus, the proof is complete.

REMARK. Under Assumption 5, the pair of ∂-discounted optimal stationary strategies in the new Markov game is the same as the pair of optimal stationary strategies in the original Markov game.

References

[1] P. KAKUMANU, *Continuous time Markov decision models with applications to optimization problems*, Tech. Rep. 63, Dept. O. R., Cornell University.

[2] P. KAKUMANU, *Continuously discounted Markov decision model with countable state and action space*, Ann. Math. Statist., 42 (1971), 919–926.

[3] P. KAKUMANU, *Nondiscounted continuous time Markovian decision process with countable state space*, SIAM J. control, 10 (1972), 210–220.

[4] P. KAKUMANU, *Continuous time Markovian decision processes average return criterion*, J. Math. Anal. and Appli., 52 (1975), 173–188.

[5] A. MAITRA and T. PARTHASARATHY, *On stochastic games*, J. Opti. Theory and Appli., 5 (1970), 289–300.

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

[7] K. TANAKA, S. IWASE and K. WAKUTA, *On Markov games with the expected average reward criterion*, Sci. Rep. Niigata Univ., Ser. A, 13 (1976), 31–41.

[8] K. TANAKA and K. WAKUTA, *On semi-Markov games*, Sci. Rep. Niigata Univ., Ser. A, 13 (1976), 55–64.

[9] K. TANAKA and K. WAKUTA, *On continuous time Markov games with countable state space*, to appear.