九州大学学術情報リポジトリ
Kyushu University Institutional Repository

CONTINUOUS TIME NON-COOPERATIVE $ N $-PERSON MARKOV GAMES

Tanaka, Kensuke
Department of Mathematics, Niigata University

Homma, Hisafumi
Department of Mathematics, Niigata University

https://doi.org/10.5109/13126

出版情報：統計数理研究. 18 (1/2), pp.93-105, 1978-03. 統計科学研究会
バージョン：
権利関係：

KYUSHU UNIVERSITY

CONTINUOUS TIME NON-COOPERATIVE n-PERSON MARKOV GAMES

By

Kensuke TANAKA*

and

Hisafumi HOMMA*

(Received October 10, 1977)

1. Introduction

This paper is a continuation of our papers [12] and [13] and is concerned with a continuous time non-cooperative n-person Markov game in which the state space is a finite set and the action spaces of all players are compact metric spaces. In the game, all players continuously observe the state of the system and then choose actions independently without collaboration with any of the others. As this result, each player gains a reward and the system moves to a new state which is governed by the known transition rates. Then, we consider the optimization problems to maximize the total expected discounted gain and the long-run expected average gain of each player. And, we show that the game has an equilibrium point and all players have the equilibrium stationary strategies under these criterions and some conditions.

2. The formulations of the games

The concept of a discrete time Markov game was first formulated by Shapley in [9]. And, by introducing a discount factor and using the results in dynamic programming, Maitra and Parthasarathy investigated a Markov game with infinite horizon in [5] and [8]. Further, in our paper [13], a continuous time Markov game with a discount factor was investigated by using the results of continuous time Markov decision process given in, mainly, [2]. At the same time, we investigated a continuous time Markov game with the expected average reward criterion in [12] by using the results in [3] and [4]. In this paper, we consider a continuous time non-cooperative n-person Markov game with a discount factor and a game with expected average reward criterion, respectively, by introducing the notion of an equilibrium point in [7].

* Department of Mathematics, Niigata University, Niigata 950-21, Japan.

So, at first, we define "continuous time non-cooperative n-person Markov game with a discount factor" by a set of $2(n+1)+1$ objects

$$(S, A_1, A_2, \cdots, A_n, q, r^{(1)}, r^{(2)}, \cdots, r^{(n)}, \alpha).$$

Here, S is a finite set labeled $1, 2, \cdots, m$, the set of states of a system; each A_i is a non-empty Borel subset of a Polish space, the set of actions available to player i, $i=1, 2, \cdots, n$,; q is a transition rate which governs the law of motion of the system and is a finite valued function $q_{s,\cdot}(a_1, a_2, \cdots, a_n)$ on S for each

$$(s, a_1, a_2, \cdots, a_n) \in S\times \prod_{k=1}^{n} A_k,;$$

$r_s^{(i)}$, a reward function of player i, is a bounded Borel measurable function on

$$\prod_{k=1}^{n} A_k,;$$

α, a discount factor, is a positive real number. And, "continuous time non-cooperative n-person Markov game with the expected average reward criterion" is defined by a set of $2(n+1)$ objects

$$(S, A_1, A_2, \cdots, A_n, q, r^{(1)}, r^{(2)}, \cdots, r^{(n)}).$$

In these games, all players continuously observe the state of the system and classify it into one of the possible states $s\in S$. Then, according to the present state s, each player i chooses independently an action $a_i \in A_i$, $i=1, 2, \cdots, n$, without collaboration with any of the others. As this result, each player i gains a reward $r_s^{(i)}(a_1, a_2, \cdots, a_n)$ unit of money and the system moves to a new state $s' \in S$, which is governed by the transition rate $q_{s,s'}(a_1, a_2, \cdots, a_n)$. Then, our optimization problems are to maximize the total expected discounted gain and the expected average gain of each player, respectively, as the games proceed to the infinite future.

We assume that strategies for each player are independent of the past history of the system and depend only on the present state of the system. Such a strategy $\pi^{(i)}=\pi^{(i)}(\mathrm{t})$ for each player i is specified by a family $\{\mu_t^{(i)}\}$, where $\mu_t^{(i)}$ is a probability measure $\mu_t^{(i)}(\cdot\,|\,s)$ on a Borel measurable space $(A_i, B(A_i))$ in which $B(A_i)$ is the σ-field generated by the metric on A_i, for each $s\in S$ and $t\in[0, \infty)$, and $\mu_t^{(i)}$ is a Lebesgue measurable function $\mu_t^{(i)}(M|s)$ on $[0, \infty)$ for each $M\in B(A_i)$ and $s\in S$. Then, we call such a strategy a Markov strategy. Moreover, such a Markov strategy $\pi^{(i)}=\pi^{(i)}(t)$ is said to be stationary if $\pi^{(i)}(t)$ is independent of t, that is, there exists a mapping μ from S into $P(A_i)$ such that $\mu_t^{(i)}=\mu$ for all $t\in[0, \infty)$, where $P(A_i)$ is a set of all probability measures on $(A_i, B(A_i))$. $\Pi^{(i)}$ denotes the class of all Markov strategies for each player i. Markov strategies and stationary strategies for other players are defined analogously.

Throughout the paper, we assume, for the transition rate matrix

$$Q(a_1, a_2, \cdots, a_n)=\{q_{s,s'}(a_1, a_2, \cdots, a_n)\,;\ s, s'\in S\}$$

corresponding to

$$(a_1, a_2, \cdots, a_n) \in \prod_{k=1}^{n} A_k,$$

the following:

ASSUMPTION 1. For each $s, s' \in S$, $q_{s,s'}(a_1, a_2, \cdots, a_n)$ is a continuous function on

$$\prod_{k=1}^{n} A_k$$

and, for all

$$(a_1, a_2, \cdots, a_n) \in \prod_{k=1}^{n} A_k,$$

$$q_{s,s'}(a_1, a_2, \cdots, a_n) \geqq 0, \quad s \neq s',$$

$$\sum_{s'} q_{s,s'}(a_1, a_2, \cdots, a_n) = 0$$

and

$$|q_{s,s}(a_1, a_2, \cdots, a_n)| \leqq M$$

for all s and some positive number $M < \infty$.

For a set of the Markov strategies $\{\pi^{(1)}, \pi^{(2)}, \cdots, \pi^{(n)}\}$ in which each $\pi^{(i)}$ is used by player i, the transition rates are defined as follows: for each $t \geqq 0$,

$$q_{s,s'}(t, \pi^{(1)}, \pi^{(2)}, \cdots, \pi^{(n)}) = \int \cdots \int q_{s,s'}(a_1, a_2, \cdots, a_n) \prod_{k=1}^{n} d\mu_t^{(k)}(a_k | s),$$

where each strategy $\pi^{(i)}$ is specified by the family $\{\mu_t^{(i)}\}$. It is clear that each $q_{s,s'}(t, \pi^{(1)}, \cdots, \pi^{(n)})$ is measurable in t. And, from Assumption 1, $q_{s,s'}(t, \pi^{(1)}, \cdots, \pi^{(n)})$ satisfies also the following conditions: for all $s, s' \in S$ and each $t \geqq 0$,

$$q_{s,s'}(t, \pi^{(1)}, \cdots, \pi^{(n)}) \geqq 0, \quad s \neq s', \tag{1}$$

$$\sum_{s'} q_{s,s'}(t, \pi^{(1)}, \cdots, \pi^{(n)}) = 0, \tag{2}$$

and

$$|q_{s,s}(t, \pi^{(1)}, \cdots, \pi^{(n)})| \leqq M. \tag{3}$$

We write the transition rate matrix corresponding to $\pi^{(1)}, \cdots, \pi^{(n-1)}$ and $\pi^{(n)}$ as $Q(t, \pi^{(1)}, \cdots, \pi^{(n)}) = \{q_{s,s'}(t, \pi^{(1)}, \cdots, \pi^{(n)}) ; s, s' \in S\}$ and, if all $\pi^{(i)}$ are stationary strategies, we write $Q(\pi^{(1)}, \pi^{(2)}, \cdots, \pi^{(n)})$ instead of $Q(t, \pi^{(1)}, \pi^{(2)}, \cdots, \pi^{(n)})$. Under (1), (2), (3), Miller showed in [6] that there exists a unique stochastic matrix $F(t, t', \pi^{(1)}, \cdots, \pi^{(n)}) = \{f_{s,s'}(t, t', \pi^{(1)}, \cdots, \pi^{(n)} ; s, s' \in S\}$ corresponding to $Q(t, \pi^{(1)}, \cdots, \pi^{(n)})$ which satisfies the Kolmogorov forward differential equations

$$\frac{\partial}{\partial t'} F(t, t', \pi^{(1)}, \cdots, \pi^{(n)}) = F(t, t', \pi^{(1)}, \cdots, \pi^{(n)}) Q(t', \pi^{(1)}, \cdots, \pi^{(n)}) \tag{4}$$

with $F(t, t, \pi^{(1)}, \cdots, \pi^{(n)})=I$ for almost all $t' \in [t, \infty)$, where I is the identity matrix. Then, a measurable Markov process $\{X(t', \pi^{(1)}, \cdots, \pi^{(n)}), t' \geqq t\}$ corresponding to the stochastic matrix $F(t, t', \pi^{(1)}, \cdots, \pi^{(n)})$ exists and is well-behaved. Throughout our discussion, the game starts from $t=0$. In this view, we write $F(t, \pi^{(1)}, \pi^{(2)}, \cdots, \pi^{(n)})$ instead of $F(0, t, \pi^{(1)}, \pi^{(2)}, \cdots, \pi^{(n)})$.

Now, we define the total expected discounted gain of each player. When a set of the Markov strategies $\{\pi^{(1)}, \pi^{(2)}, \cdots, \pi^{(n)}\}$ is used, at any time t, the expected gain rate of each player i out of state $s \in S$ is given by

$$r_s^{(i)}(t, \pi^{(1)}, \cdots, \pi^{(n)}) = \int \cdots \int r_s^{(i)}(a_1, \cdots, a_n) \prod_{k=1}^{n} d\mu_t^{(k)}(a_k | s).$$

Clearly, $r_s^{(i)}(t, \pi^{(1)}, \cdots, \pi^{(n)})$ is measurable in t. Thus, when the system starts in a state $s \in S$ and a set of Markov strategies $\{\pi^{(1)}, \pi^{(2)}, \cdots, \pi^{(n)}\}$ is used, the total expected discounted gain for each player i is defined to be

$$\psi_s^{(i)}(\alpha, \pi^{(1)}, \cdots, \pi^{(n)}) = \int_0^\infty e^{-\alpha t} \sum_{s'} f_{s,s'}(t, \pi^{(1)}, \cdots, \pi^{(n)}) r_{s'}^{(i)}(t, \pi^{(1)}, \cdots, \pi^{(n)}) \, dt.$$

We say that a set of Markov strategies $\{\pi^{(1)}, \cdots, \pi^{(n)}\}$ is an equilibrium point if, for all i and $s \in S$,

$$\psi_s^{(i)}(\alpha, \pi^{(1)}, \cdots, \pi^{(n)}) = \sup_{\sigma^{(i)} \in \Pi^{(i)}} \psi_s^{(i)}(\alpha, \pi^{(1)}, \cdots, \sigma^{(i)}, \cdots, \pi^{(n)}).$$

When the game has an equilibrium point $\{\pi^{(1)}, \pi^{(2)}, \cdots, \pi^{(n)}\}$, each $\pi^{(i)}$, $i=1, 2, \cdots, n$, is called an equilibrium strategy for player i respectively.

Next, we define the expected average gain and an equilibrium strategy of each player. When the system starts in a state $s \in S$ and a set of Markov strategies $\{\pi^{(1)}, \pi^{(2)}, \cdots, \pi^{(n)}\}$ is used, the total expected gain for each player i up to the time T

$$\phi_s^{(i)}(T, \pi^{(1)}, \cdots, \pi^{(n)}) = \int_0^T \sum_{s'} f_{s,s'}(t, \pi^{(1)}, \cdots \pi^{(n)}) r_{s'}^{(i)}(t, \pi^{(1)}, \cdots, \pi^{(n)}) \, dt$$

and the expected average gain for each player i is defined by

$$\phi_s^{(i)}(\pi^{(1)}, \cdots, \pi^{(n)}) = \overline{\lim_{T \to \infty}} \frac{\phi_s^{(i)}(T, \pi^{(1)}, \cdots, \pi^{(n)})}{T}.$$

Then, a set of Markov strategies $\{\pi^{(1)}, \pi^{(2)}, \cdots, \pi^{(n)}\}$ is called an equilibrium point if, for all i and $s \in S$,

$$\phi_s^{(i)}(\pi^{(1)}, \pi^{(2)}, \cdots, \pi^{(n)}) = \sup_{\sigma^{(i)} \in \Pi^{(i)}} \phi_s^{(i)}(\pi^{(1)}, \cdots, \sigma^{(i)}, \cdots, \pi^{(n)})$$

and each $\pi^{(i)}$, $i=1, 2, \cdots, n$, is called an equilibrium strategy for player i, respectively.

3. Existence of equilibrium stationary strategies in the Markov game with a discount factor

In this section, we show the existence of equilibrium stationary strategies for the game with a discount factor. Firstly, for a set of the stationary strategies $\{\mu^{(1)}, \mu^{(2)}, \cdots, \mu^{(n)}\}$, we assume the following notations:

$$\bar{\mu} \equiv \{\mu^{(1)}, \mu^{(2)}, \cdots, \mu^{(n)}\} \in \left(\prod_{k=1}^{n} P(A_k)\right)^S,$$

$$\hat{\mu}^{(i)} \equiv \{\mu^{(1)}, \cdots, \mu^{(i-1)}, \mu^{(i+1)}, \cdots, \mu^{(n)}\} \in \left(\prod_{\substack{k=1 \\ k\neq i}}^{n} P(A_k)\right)^S,$$

$$(\bar{\mu}\,;\sigma^{(i)}) \equiv \{\mu^{(1)}, \cdots, \mu^{(i-1)}, \sigma^{(i)}, \mu^{(i+1)}, \cdots, \mu^{(n)}\}$$

for

$$\sigma^{(i)} \in (P(A_i))^S,$$

$$r_s^{(i)}(\bar{\mu}) \equiv \int \cdots \int r_s^{(i)}(a_1, a_2, \cdots, a_n) \prod_{k=1}^{n} d\mu^{(k)}(a_k \mid s),$$

and

$$q_{s,s'}(\bar{\mu}) \equiv \int \cdots \int q_{s,s'}(a_1, a_2, \cdots, a_n) \prod_{k=1}^{n} d\mu^{(k)}(a_k \mid s),$$

where

$$\left(\prod_{k=1}^{n} P(A_k)\right)^S$$

denotes a set of all mappings from S into

$$\prod_{k=1}^{n} P(A_k).$$

Secondly, throughout the paper, we assume the following:

ASSUMPTION 2. Each A_i, $i=1, 2, \cdots, n$, is a compact metric space, respectively, and, for each i and $s \in S$, $r_s^{(i)}(a_1, a_2, \cdots, a_n)$ is a continuous function on

$$\prod_{k=1}^{n} A_k.$$

Then, since by Assumption 2 each $P(A_i)$ endowed with weak topology is a compact metric space, $P(A_i)$ is a compact metric space and, for each $s, s' \in S$ and i, $r_s^{(i)}(\bar{\mu})$ and $q_{s,s'}(\bar{\mu})$ are bounded continuous functions on

$$\prod_{i=1}^{n} P(A_i).$$

Throughout the paper, we assume that each $P(A_i)$ is endowed with weak topology.

LEMMA 3. 1. *For each i, s and*

$$\bar{\mu}\in\left(\prod_{k=1}^{n} P(A_k)\right)^S,$$

$$v_s^{(i)}(\bar{\mu})=\phi_s^{(i)}(\alpha, \bar{\mu})$$

is the unique solution to

$$\alpha v_s^{(i)}(\bar{\mu})=r_s^{(i)}(\bar{\mu})+\sum_{s'} q_{s,s'}(\bar{\mu})v_{s'}^{(i)}(\bar{\mu}), \tag{5}$$

where α is the discount factor.

The proof of this lemma is given in our paper [13].

Let X^m be a m-dimensional vector space. For $v=(v_1, v_2, \cdots, v_m)\in X^m$, we define

$$\|v\|=\max_i |v_i|\cdot$$

(X^m, d) is a complete metric space, where $d(v, u)=\|v-u\|$ for each v and $u\in X^m$.

Now, for a set of stationary strategies $\bar{\mu}=\{\mu^{(1)}, \mu^{(2)}, \cdots, \mu^{(n)}\}$, we define a new one-step stochastic matrix $P(\bar{\mu})$ with relation to $Q(\bar{\mu})$ as follows:

$$P(\bar{\mu})=I+M^{-1}Q(\bar{\mu}),$$

whose (s, s')th element is given by

$$p_{s,s'}(\bar{\mu})=\delta_{s,s'}+M^{-1}q_{s,s'}(\bar{\mu}),$$

where M is the positive number in Assumption 1.

Then, since

$$\prod_{k=1}^{n} P(A_k)$$

is a compact metric space and $r_s^{(i)}(\bar{\mu})$ and $p_{s,s'}(\bar{\mu})$ are continuous functions on

$$\prod_{k=1}^{n} P(A_k),$$

for each i and s, we can define a mapping $T^{(i)}: X^m\rightarrow X^m$ as follows: for each i and $v\in X^m$,

$$(T^{(i)}v)(s)=\max_{\sigma^{(i)}\in P(A_i)}\{(\alpha+M)^{-1}r_s^{(i)}(\bar{\mu}\,;\sigma^{(i)})$$
$$+M(\alpha+M)^{-1}\sum_{s'} p_{s,s'}(\bar{\mu}\,;\sigma^{(i)})v_{s'}\}.$$

This mapping is a contraction mapping on X^m because $0<M(\alpha+M)^{-1}<1$. Hence, $T^{(i)}$ has a unique fixed point in X^m by the Banach's fixed point theorem.

Let $v^{(i)}(\hat{\mu}^{(i)})$ be the unique fixed point of $T^{(i)}$. For $v^{(i)}(\hat{\mu}^{(i)})$, the following lemma holds.

LEMMA 3. 2. *For each i, there exists a $v^{(i)}(\hat{\mu}^{(i)})\in X^m$ such that, for each s and $\bar{\mu}$,*

$$\alpha v_s^{(i)}(\hat{\mu}^{(i)})=\max_{\sigma^{(i)}\in P(A_i)}\{r_s^{(i)}(\bar{\mu}\,;\sigma^{(i)})+\sum_{s'} q_{s,s'}(\bar{\mu}\,;\sigma^{(i)})v_{s'}^{(i)}(\hat{\mu}^{(i)})\}, \tag{6}$$

where $v_s^{(i)}(\hat{\mu}^{(i)})$ is the s-th element of $v^{(i)}(\hat{\mu}^{(i)})$.

The proof of this lemma is given in our paper [13].
Moreover, we can prove that the fixed point $v^{(i)}(\hat{\mu}^{i)})$ is continuous in $\bar{\mu}$.

LEMMA 3. 3. *If*

$$\bar{\mu}_l(s) \in \prod_{k=1}^{n} P(A_k)$$

for all l *and*

$$\bar{\mu}_l(s) \Rightarrow \bar{\mu}_0(s) \in \prod_{k=1}^{n} P(A_k) \qquad as \quad l \to \infty,$$

it holds that, for each i,

$$\| v^{(i)}(\hat{\mu}_l^{(i)}) - v^{(i)}(\hat{\mu}_0^{(i)}) \| \to 0 \quad as\ l \to \infty,$$

where the notation $\Rightarrow$ *denotes weak convergence in*

$$\prod_{k=1}^{n} P(A_k).$$

PROOF. Since $v^{(i)}(\hat{\mu}^{(i)})$ is the fixed point of $T^{(i)}$, we have, for all i and s,

$$\begin{aligned} | v_s^{(i)}(\hat{\mu}_l^{(i)}) - v_s^{(i)}(\hat{\mu}_0^{(i)}) | \leqq (\alpha+M)^{-1} \max_{\sigma^{(i)} \in P(A_i)} & | r_s^{(i)}(\bar{\mu}_l ; \sigma^{(i)}) \\ - r_s^{(i)}(\bar{\mu}_0 ; \sigma^{(i)}) | + M(\alpha+M)^{-1} \max_{\sigma^{(i)} \in P(A_i)} & | \sum_{s'} p_{s,s'}(\bar{\mu}_l ; \sigma^{(i)}) \\ \times v_{s'}^{(i)}(\hat{\mu}_l^{(i)}) - \sum_{s'} p_{s,s'}(\bar{\mu}_0 ; \sigma^{(i)}) v_{s'}^{(i)}(\hat{\mu}_0^{(i)}) | . & \end{aligned} \tag{7}$$

Further, we can rewrite (7) as follows:

$$\begin{aligned} | v_s^{(i)}(\hat{\mu}_l^{(i)}) - v_s^{(i)}(\hat{\mu}_0^{(i)}) | \leqq & (\alpha+M)^{-1} \max_{\sigma^{(i)} \in P(A_i)} | r_s^{(i)}(\bar{\mu}_l ; \sigma^{(i)}) - r_s^{(i)}(\bar{\mu}_0 ; \sigma^{(i)}) | \\ & + M(\alpha+M)^{-1} \max_{\sigma^{(i)} \in P(A_i)} \sum_{s'} | p_{s,s'}(\bar{\mu}_l ; \sigma^{(i)}) - p_{s,s'}(\bar{\mu}_0 ; \sigma^{(i)}) | \\ & \times | v_{s'}^{(i)}(\hat{\mu}_0^{(i)}) | + M(\alpha+M)^{-1} \max_{\sigma^{(i)} \in P(A_i)} \sum_{s'} p_{s,s'}(\bar{\mu}_0 ; \sigma^{(i)}) \\ & \times | v_{s'}^{(i)}(\hat{\mu}_l^{(i)}) - v_{s'}^{(i)}(\hat{\mu}_0^{(i)}) | . \end{aligned} \tag{8}$$

Since $r_s^{(i)}(\bar{\mu}_l ; \sigma^{(i)})$ and $p_{s,s'}(\bar{\mu}_l ; \sigma^{(i)})$ converge to $r_s^{(i)}(\bar{\mu}_0 ; \sigma^{(i)})$ and $p_{s,s'}(\bar{\mu}_0 ; \sigma^{(i)})$ uniformly in $\sigma^{(i)}$ respectively, the first term and second term in right-hand side of (8) become zere as $l \to \infty$. And, by using $\sum_{s'} p_{s,s'}(\bar{\lambda}) = 1$ for all $\bar{\lambda}(s) \in P(A_k)$, we obtain

$$\begin{aligned} & \overline{\lim_{l\to\infty}} \| v^{(i)}(\hat{\mu}_l^{(i)}) - v^{(i)}(\hat{\mu}_0^{(i)}) \| \\ & \qquad \leqq M(\alpha+M)^{-1} \overline{\lim_{l\to\infty}} \| v^{(i)}(\hat{\mu}_l^{(i)}) - v^{(i)}(\hat{\mu}_0^{(i)}) \| . \end{aligned} \tag{9}$$

From (9), it holds that, for each i,

$$\|v^{(i)}(\hat{\mu}_l^{(i)})-v^{(i)}(\hat{\mu}_0^{(i)})\|\to 0 \quad \text{as} \quad l\to\infty,$$

because $0<M(\alpha+M)^{-1}<1$. Thus, the lemma is proved.

Next, for $\bar{\mu}$ and $v^{(i)}(\hat{\mu}^{(i)})$ in lemma 3. 2, we introduce the following notation: for each i and $s\in S$,

$$K_s^{(i)}(\bar{\mu}\,;\sigma^{(i)})\equiv r_s^{(i)}(\bar{\mu}\,;\sigma^{(i)})+\sum_{s'} q_{s,s'}(\bar{\mu}\,;\sigma^{(i)})v_{s'}^{(i)}(\hat{\mu}^{(i)})$$

for

$$\sigma^{(i)}(s)\in P(A_i),$$

and

$$G^{(i)}(\hat{\mu}^{(i)})\equiv\{\lambda^{(i)}\,;K_s^{(i)}(\bar{\mu}\,;\lambda^{(i)})=\max_{\sigma^{(i)}\in P(A_i)} K_s^{(i)}(\bar{\mu}\,;\sigma^{(i)}) \text{ for all } s\in S\}.$$

Then,

$$\prod_{k=1}^{n} P(A_k)$$

is a compact convex set of locally convex space and $G^{(i)}(\hat{\mu}^{(i)})$ is a non-empty, closed and convex subset. So we define a mapping G:

$$\left(\prod_{k=1}^{n} P(A_k)\right)^S\to\left(\prod_{k=1}^{n} P(A_k)\right)^S$$

as follows: for each $\bar{\mu}$,

$$G(\bar{\mu})\equiv\{(\lambda^{(1)},\lambda^{(2)},\cdots,\lambda^{(n)})\,;\lambda^{(i)}\in G^{(i)}(\hat{\mu}^{(i)}) \text{ for all } i\}.$$

Hence, in order to apply the fixed point theorem in [1], the following lemma is important.

LEMMA 3. 4. *The mapping G is upper semi-continuous.*

PROOF. It will be sufficient to show that, if $\bar{\lambda}_l\Rightarrow\bar{\lambda}_0$, $\bar{\mu}_l\Rightarrow\bar{\mu}_0$ as $l\to\infty$ and $\bar{\lambda}_l\in G(\bar{\mu}_l)$ for all l, then $\bar{\lambda}_0\in G(\bar{\mu}_0)$. In fact, from Lemma 3. 2, we have for each s and $\bar{\lambda}_l=(\lambda_l^{(1)},\lambda_l^{(2)},\cdots,\lambda_l^{(n)})\in G(\bar{\mu}_l)$

$$\alpha v_s^{(i)}(\hat{\mu}_l^{(i)})=\max_{\sigma^{(i)}\in P(A_i)} K_s^{(i)}(\bar{\mu}_l\,;\sigma^{(i)})=K_s^{(i)}(\bar{\mu}_l\,;\lambda_l^{(i)}), \tag{10}$$

and, for any $\sigma^{(i)}(s)\in P(A_i)$,

$$\alpha v_s^{(i)}(\hat{\mu}_l^{(i)})\geqq K_s^{(i)}(\bar{\mu}_l\,;\sigma^{(i)}). \tag{11}$$

So passing to the limit and using Lemma 3. 3, (10) and (11) can be written as

$$\alpha v_s^{(i)}(\hat{\mu}_0^{(i)})=K_s^{(i)}(\bar{\mu}_0\,;\lambda_0^{(i)})$$

and for any $\sigma^{(i)}(s)\in P(A_i)$

$$\alpha v_s^{(i)}(\hat{\mu}_0^{(i)})\geqq K_s^{(i)}(\bar{\mu}_0\,;\sigma^{(i)}),$$

respectively. Thus, the lemma is proved.

Then, we can conclude from Fan's theorem that there exists $\bar{\mu}_*=(\mu_*^{(1)}, \mu_*^{(2)}, \cdots, \mu_*^{(n)})$ such that $\bar{\mu}_*\in G(\bar{\mu}_*)$, that is, for each i and $s\in S$,

$$\alpha v_s^{(i)}(\hat{\mu}_*^{(i)})=r_s^{(i)}(\bar{\mu}_*)+\sum_{s'} q_{s,s'}(\bar{\mu}_*)v_{s'}^{(i)}(\hat{\mu}_*^{(i)}), \tag{12}$$

and, for any $\sigma^{(i)}(s)\in P(A_i)$,

$$\alpha v_s^{(i)}(\hat{\mu}_*^{(i)})\geqq r_s^{(i)}(\bar{\mu}_*\,;\sigma^{(i)})+\sum_{s'} q_{s,s'}(\bar{\mu}_*\,;\sigma^{(i)})v_{s'}^{(i)}(\hat{\mu}_*^{(i)}). \tag{13}$$

Then, we can prove the following theorem.

THEOREM 3. 1. *The game has an equilibrium point and all players have the equilibrium stationary strategies.*

PROOF. By using Lemma 3. 1 to (12), we have for each i and s,

$$v_s^{(i)}(\hat{\mu}_*^{(i)})=\phi_s^{(i)}(\alpha, \bar{\mu}_*). \tag{14}$$

From (13), for a set of the stationary strategies $\hat{\mu}_*^{(i)}$ and any Markov strategy $\pi^{(i)}$ for player i, we have for each $t\geqq 0$

$$\alpha v_s^{(i)}(\hat{\mu}_*^{(i)})\geqq r_s^{(i)}(t, \bar{\mu}_*\,;\pi^{(i)})+\sum_{s'} q_{s,s'}(t, \bar{\mu}_*\,;\pi^{(i)})v_{s'}^{(i)}(\hat{\mu}_*^{(i)}). \tag{15}$$

Multiplying both sides of (15) by $e^{-\alpha t}f_{l,s}(t, \bar{\mu}_*\,;\pi^{(i)})$ and summing over all $s\in S$, we get for each $l\in S$

$$\begin{aligned}\alpha e^{-\alpha t}&\sum_s f_{l,s}(t, \bar{\mu}_*\,;\pi^{(i)})v_s^{(i)}(\hat{\mu}_*^{(i)})\\ &\geqq e^{-\alpha t}\sum_s f_{l,s}(t, \bar{\mu}_*\,;\pi^{(i)})v_s^{(i)}(t, \bar{\mu}_*\,;\pi^{(i)})\\ &+e^{-\alpha t}\sum_s\sum_{s'} f_{l,s}(t, \bar{\mu}_*\,;\pi^{(i)})q_{s,s'}(t, \bar{\mu}_*\,;\pi^{(i)})v_{s'}^{(i)}(\hat{\mu}_*^{(i)}).\end{aligned} \tag{16}$$

We interchange the summation signs in the second term of the right-hand side of (16). And, using the Kolmogorov forward differential equation, we obtain for each $l\in S$

$$\begin{aligned}\alpha e^{-\alpha t}&\sum_s f_{l,s}(t, \bar{\mu}_*\,;\pi^{(i)})v_s^{(i)}(\hat{\mu}_*^{(i)})\\ &\geqq\alpha e^{-\alpha t}\sum_s f_{l,s}(t, \bar{\mu}_*\,;\pi^{(i)})r_s^{(i)}(t, \bar{\mu}_*\,;\pi^{(i)})\\ &+e^{-\alpha t}\sum_s \frac{\partial}{\partial t}f_{l,s}(t, \bar{\mu}_*\,;\pi^{(i)})v_s^{(i)}(\hat{\mu}_*^{(i)}).\end{aligned} \tag{17}$$

By integrating on both sides of (17) with respect to $t\in[0, \infty)$, we have for each $l\in S$

$$\begin{aligned}v_s^{(i)}(\hat{\mu}_*^{(i)})&\geqq\int_0^\infty e^{-\alpha t}\sum_s f_{l,s}(t, \bar{\mu}_*\,;\pi^{(i)})r_s^{(i)}(t, \bar{\mu}_*\,;\pi^{(i)})\,dt\\ &=\phi_s^{(i)}(\alpha, \bar{\mu}_*\,;\pi^{(i)}).\end{aligned} \tag{18}$$

From (14) and (18), a set of stationary strategies $\bar{\mu}_* = (\mu_*^{(1)}, \mu_*^{(2)}, \cdots, \mu_*^{(n)})$ is an equilibrium point and each i-th element of $\bar{\mu}_*$ is an equilibrium stationary strategy for each player i. Thus, the theorem is proved.

4. Existence of equilibrium stationary strategies in the Markov game with the expected average reward criterion

In this section, we show the existence of equilibrium stationary strategies for the game with the expected average reward criterion. First, in addition to Assumption 1 and Assumption 2, we impose, in particular, on q the following assumption in this section.

ASSUMPTION 3. There exist some state $k \in S$ and a positive number β such that, for all $s \neq k$ and

$$(a_1, a_2, \cdots, a_n) \in \prod_{i=1}^{n} A_i,$$

$$q_{s,k}(a_1, a_2, \cdots, a_n) \geqq \beta > 0.$$

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

$$\bar{q}_{s,s'}(a_1, a_2, \cdots, a_n) = q_{s,s'}(a_1, a_2, \cdots, a_n) + \delta_{s,s'}\beta, \quad s' \neq k$$

and

$$\bar{q}_{s,k}(a_1, a_2, \cdots, a_n) = q_{s,k}(a_1, a_2, \cdots, a_n) + \delta_{s,k}\beta - \beta,$$

where $\delta_{s,k}$ denotes the Kronecker delta.

Clearly, the new transition rate matrix $\bar{Q}(a_1, a_2, \cdots, a_n)$ satisfies Assumption 1. Hence, there exists a unique stochastic matrix $\bar{F}(t, t', \pi^{(1)}, \pi^{(2)}, \cdots, \pi^{(n)})$ for any set of Markov strategies $\{\pi^{(1)}, \pi^{(2)}, \cdots, \pi^{(n)}\}$. By using this transition rate matrix, we consider a new continuous time Markov game $(S, A_1, \cdots, A_n, \bar{q}, r^{(1)}, \cdots, r^{(n)}, \beta)$ with the same state space, the same action spaces and the same rewards as the original game and with a discount factor $\beta > 0$. Since this new game satisfies Assumption 1 and Assumption 2, from Theorem 3. 1, there exists an equilibrium point $\bar{\mu}_* = (\mu_*^{(1)}, \cdots, \mu_*^{(n)})$.

Then, by the theory of a Markov game with a discount factor, we can prove the following lemma.

LEMMA 4. 1. *There exists a number $g^{(i)}(\hat{\mu}_*^{(i)})$, a m-dimensional vector $v^{(i)}(\hat{\mu}_*^{(i)}) \in X^m$ and the set of stationary strategies*

$$\bar{\mu}_* \in \Big(\prod_{k=1}^{n} P(A_k)\Big)^S$$

such that, for each i and $s \in S$,

$$g^{(i)}(\hat{\mu}_*^{(i)}) = \max_{\sigma^{(i)} \in P(A_i)} \{r_s^{(i)}(\bar{\mu}_*; \sigma^{(i)}) + \sum_{s'} q_{s,s'}(\bar{\mu}_*; \sigma^{(i)}) v_{s'}^{(i)}(\hat{\mu}_*^{(i)})\}$$

$$= r_s^{(i)}(\bar{\mu}_*) + \sum_{s'} q_{s,s'}(\bar{\mu}_*) v_{s'}^{(i)}(\hat{\mu}_*^{(i)}).$$

PROOF. Since $\bar{\mu}_*$ is an equilibrium point in the new Markov game, from (12), we have for each i and $s \in S$,

$$\beta v_s^{(i)}(\hat{\mu}_*^{(i)}) = \max_{\sigma^{(i)} \in P(A_i)} \{r_s^{(i)}(\bar{\mu}_* ; \sigma^{(i)}) + \sum_{s'} \bar{q}_{s,s'}(\bar{\mu}_* ; \sigma^{(i)}) v_{s'}^{(i)}(\hat{\mu}_*^{(i)})\}$$

$$= r_s^{(i)}(\bar{\mu}_*) + \sum_{s'} \bar{q}_{s,s'}(\bar{\mu}_*) v_{s'}^{(i)}(\hat{\mu}_*^{(i)}). \tag{19}$$

Substituting $q_{s,s'}(\bar{\mu}_* ; \sigma^{(i)})$ and $q_{s,s'}(\bar{\mu}_*)$ for $\bar{q}_{s,s'}(\bar{\mu}_* ; \sigma^{(i)})$ and $\bar{q}_{s,s'}(\bar{\mu}_*)$ in (19) respectively, we obtain, after some simplification, for each i and $s \in S$,

$$\beta v_k^{(i)}(\hat{\mu}_*^{(i)}) = \max_{\sigma^{(i)} \in P(A_i)} \{r_s^{(i)}(\bar{\mu}_* ; \sigma^{(i)}) + \sum_{s'} q_{s,s'}(\bar{\mu}_* ; \sigma^{(i)}) v_{s'}^{(i)}(\hat{\mu}_*^{(i)})\}$$

$$= r_s^{(i)}(\bar{\mu}_*) + \sum_{s'} q_{s,s'}(\bar{\mu}_*) v_{s'}^{(i)}(\hat{\mu}_*^{(i)}).$$

Thus, the lemma is proved.

In order to main theorem in this section, the following lemma is important.

LEMMA 4. 2. *For a set of the stationary strategies* $\bar{\mu} = \{\mu^{(1)}, \mu^{(2)}, \cdots, \mu^{(n)}\}$, *if there exists a number* $g^{(i)}(\hat{\mu}^{(i)})$ *and a m-dimensional vector* $v^{(i)}(\hat{\mu}^{(i)}) \in X^m$ *such that, for each* i *and* $s \in S$,

$$g^{(i)}(\hat{\mu}^{(i)}) = r_s^{(i)}(\bar{\mu}) + \sum_{s'} q_{s,s'}(\bar{\mu}) v_{s'}^{(i)}(\hat{\mu}^{(i)}), \tag{20}$$

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

$$g^{(i)}(\hat{\mu}^{(i)}) = \lim_{T \to \infty} \frac{\phi_s^{(i)}(T, \bar{\mu})}{T} = \phi_s^{(i)}(\bar{\mu}).$$

PROOF. Let $\{f_{l,s}(t, \bar{\mu}) ; l, s \in S\}$ be the stochastic matrix corresponding to $Q(\bar{\mu}) = \{q_{s,s'}(\bar{\mu}) ; s, s' \in S\}$. Multiplying both sides of (20) by $f_{l,s}(t, \bar{\mu})$ and summing over all $s \in S$, we have, for each $l \in S$,

$$g^{(i)}(\hat{\mu}^{(i)}) = \sum_s f_{l,s}(t, \bar{\mu}) r_s^{(i)}(\bar{\mu}) + \sum_s \sum_{s'} f_{l,s}(t, \bar{\mu}) q_{s,s'}(\bar{\mu}) v_{s'}^{(i)}(\hat{\mu}^{(i)}). \tag{21}$$

By interchanging the order of the double sum in the second term of the right-hand side of (21) and, then, by using the Kolmogorov forward differential equation, we get for each $l \in S$

$$g^{(i)}(\hat{\mu}^{(i)}) = \sum_s f_{l,s}(t, \bar{\mu}) r_s^{(i)}(\bar{\mu}) + \sum_s \frac{\partial}{\partial t} f_{l,s}(t, \bar{\mu}) v_s^{(i)}(\hat{\mu}^{(i)}). \tag{22}$$

By integrating on both sides of (22) with respect to t from 0 to $T < \infty$, we have

$$T g^{(i)}(\hat{\mu}^{(i)}) = \int_0^T \sum_s f_{l,s}(t, \bar{\mu}) r_s^{(i)}(\bar{\mu}) \, dt$$

$$+ \sum_s f_{l,s}(T, \bar{\mu}) v_s^{(i)}(\hat{\mu}^{(i)}) - v_l^{(i)}(\hat{\mu}^{(i)}).$$

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

$$g^{(i)}(\hat{\mu}^{(i)}) = \lim_{T\to\infty} \frac{\phi_s^{(i)}(T, \bar{\mu})}{T}.$$

Thus, the lemma is proved.

Then, we can prove the following theorem.

THEOREM 4. 1. *The game has an equilibrium point and all players have the equilibrium stationary strategies.*

PROOF. From Lemma 4. 1, for a set of the stationary strategies $\hat{\mu}_*^{(i)}$ and any Markov strategy $\sigma^{(i)}$ for player i, we have for each $t \geqq 0$

$$g^{(i)}(\hat{\mu}_*^{(i)}) \geqq r_s^{(i)}(t, \bar{\mu}_* ; \sigma^{(i)}) + \sum_{s'} q_{s,s'}(t, \bar{\mu}_* ; \sigma^{(i)}) v_{s'}^{(i)}(\hat{\mu}_*^{(i)}). \tag{23}$$

Multiplying both sides of (23) by $f_{l,s}(t, \bar{\mu}_* ; \sigma^{(i)})$ and summing over all $s \in S$, and then, using the Kolmogorov forward differential equation, we get for each $l \in S$

$$g^{(i)}(\hat{\mu}_*^{(i)}) \geqq \sum_s f_{l,s}(t, \bar{\mu}_* ; \sigma^{(i)}) r_s^{(i)}(t, \bar{\mu}_* ; \sigma^{(i)})$$

$$+ \sum_s \frac{\partial}{\partial t} f_{l,s}(t, \bar{\mu}_* ; \sigma^{(i)}) v_s^{(i)}(\hat{\mu}_*^{(i)}). \tag{24}$$

Next, integrating on both sides of (24) with respect to t from 0 to $T < \infty$, and then, dividing by T and taking the superior limit as $T \to \infty$, we get

$$g^{(i)}(\hat{\mu}_*^{(i)}) \geqq \overline{\lim_{T\to\infty}} \frac{\phi_s^{(i)}(T, \bar{\mu}_* ; \sigma^{(i)})}{T} = \phi_s^{(i)}(\bar{\mu}_* ; \sigma^{(i)}). \tag{25}$$

On the other hand, from Lemma 4. 1 and Lemma 4. 2, we get, for each i and $s \in S$,

$$g^{(i)}(\hat{\mu}_*^{(i)}) = \lim_{T\to\infty} \frac{\phi_s^{(i)}(T, \bar{\mu}_*)}{T} = \phi_*^{(i)}(\bar{\mu}_*). \tag{26}$$

From (25) and (26), it holds that, for each i and $s \in S$,

$$\phi_s^{(i)}(\bar{\mu}_*) = \sup_{\sigma^{(i)} \in P(A_i)} \phi_s^{(i)}(\bar{\mu}_* ; \sigma^{(i)}).$$

Thus, the theorem is proved.

REMARK. The equilibrium point $\bar{\mu}_*$ in the new Markov game is the equilibrium point in the original Markov game.

References

[1] FAN, K., *Fixed-point and minimax theorems in locally convex topological linear space*, Proc. Nat. Acad. Sci. USA., **88** (1952), 121-126.

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

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

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

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

[6] MILLER, B.L., *Finite state continuous time Markov decision process with applications to a class of optimization problems in queueing theory*, Tech. Rep. 15, Dept., O.R., Stanford University.

[7] NASH, J.F., *Equilibrium points in n-person games*, Proc. Nat. Acad. Sci. USA., **36** (1950), 48-49.

[8] PARTHASARATHY, T., *Discounted, positive and non-cooperative stochastic games*, Inter. J. Game Theory, **2** (1973), 25-37.

[9] SHAPLEY, L.S., *Stochastic games*, Proc. Nat. Acad. Sci. USA., **39** (1953), 1095-1100.

[10] TAKAHASHI, M., *Equilibrium points of stochastic non-cooperative n-person games*, J. Sci. Hiroshima Univ., **28** (1964), 95-99.

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

[12] TANAKA, K. and WAKUTA, K., *On continuous time Markov games with the expected average reward criterion*, Sci. Rep. Niigata Univ., **14** (1977), 15-24.

[13] TANAKA, K. and WAKUTA, K., *Continuous time Markov games with countable state space*, to appear.