On an Optimal Multistrategy and a Weak Optimal Multistrategy of a Markov Game

By
LIU Zhaohua and Kensuke TANAKA

(Received September 16, 1986)

1. Introduction

In the dynamic game theory, the multiperson game model with a discount factor on infinite horizon has been studied by many authors. The literature in this area is mostly concerned with the noncooperative equilibrium point. Such an equilibrium point gives the individual stability to each player, but it does not guarantee the collective stability. Actually, in many cases, the players may find a multistrategy which will yield a smaller total expected discounted loss if they cooperate. So, in [**9**], we proposed to find the D-solution which is analogous to the domination structure for a multiobjective decision problem.

In the paper, we introduce the distance from the total expected discounted loss constructed by all players to some given point as their collective loss function. All players cooperate in choosing a multistrategy to minimize this distance. But, in general, it would seem difficult to us to find directly such an optimal multistrategy. So, using some weighting factor vector, we modify our game system to a new one with the loss function weighting by this vector. Then, we develop the theory to find an optimal multistrategy, which is called a weak optimal multistrategy in the paper, in the modified game system. Moreover, we discuss the relation between an optimal multistrategy and a weak optimal multistrategy in the case which there exists an optimal one in the original game system. Finally, we show that a weak optimal stationary multistrategy in Theorem 1 is a D-solution under a domination structure determined by some convex cone D.

This paper is organized in the following way. In Section 2, we give a standard formulation for a cooperative m-person discounted Markov game. In Section 3, we give the necessary lemmas and definitions. In Section 4, we show the existence of a weak optimal multistrategy and discuss the relation between an optimal multistrategy and a weak optimal multistrategy. Finally, we show that a weak optimal multistrategy is a D-solution.

2. Formulation of a cooperative m-person Markov game

A cooperative m-person Markov game with a discount factor is given by a set of $2m+3$ objects:

$$(S, A^1, A^2, \ldots, A^m, q, r^1, r^2, \ldots, r^m, \beta), \tag{2.1}$$

where

(i) $S=\{1, 2, \ldots, s, \ldots\}$ is a countable set of states in the game system, namely, the state space.

(ii) A^i is the action space of the ith player, $i=1, 2, \ldots, m$. We assume that each A^i is a compact metric space.

(iii) q is a transition probability measure which governs the law of motion in the game process, in other words, for any state s and any multiaction $\bar{a}$ such that

$$(s, \bar{a}) = (s, a^1, a^2, \ldots, a^m) \in S \times \prod_{i=1}^{m} A^i = S \times A, \tag{2.2}$$

there corresponds a probability measure $q(\cdot \mid s, \bar{a})$ defined on S which decides the transition from s to a new state.

(iv) r^i is the loss function of the ith player, $i=1, 2, \ldots, m$, it is a real valued function defined at $(s, \bar{a}) \in S \times A$.

(v) β is a given discount factor, $0<\beta<1$.

Throughout this paper, we assume that all multistrategies chosen by the players are Markov multistrategies, that is, each multistrategy is independent of the past history in the game process, it depends only on the present state. We denote the Markov multistrategy by $\bar{\pi}=(\pi^1, \pi^2, \ldots, \pi^m)$, where π^i is a Markov strategy of the ith player and each strategy π^i is specified by a sequence of probability measures $\pi_t^i(\cdot \mid s_t)$ on $(A^i, B(A^i))$ for a given state s_t at the time t, where $B(A^i)$ is the Borel field of A^i. If each π_t^i is independent of the time t, that is, $\pi_t^i=\mu^i$ is stationary, we write

$$\pi^i=(\mu^i, \mu^i, \ldots\ldots)=\mu^i, \tag{2.3}$$

namely, stationary strategy. This is $\mu^i \in [P(A^i)]^S$. For simplicity, we write $P(A^i)$ instead of $[P(A^i)]^S$ as the stationary strategy space of the ith player. $P(A^i)$ is the set of all probability measures on $(A^i, B(A^i))$. We denote by Π^i the class of all Markov strategies of the ith player, $i=1, 2, \ldots, m$.

Then the Markov game process is interpreted as follows: If a Markov multistrategy $\bar{\pi}=(\pi^1, \pi^2, \ldots, \pi^m)$ is chosen, at the successive discrete time t, $t=1, 2, \ldots$, all players observe the state of the game system and classify the state s_t for the game process. Then, at the present state s_t, each player i chooses an action $a_t^i \in A^i$ by the probability measure $\pi_t^i(\cdot \mid s_t)$. As a sequence of the chosen $\bar{a}_t=(a_t^1, a_t^2, \ldots, a_t^m) \in A$ at the state s_t, the ith player will have loss $r^i(s_t, \bar{a}_t)$ and then the game process moves to a new state s_{t+1} according to the transition probability measure $q(\cdot \mid s_t, \bar{a}_t)$. After that the whole process of the game is restarted from the state s_{t+1}. So, for any initial distribution

$p_1(\cdot)$, the sequence of the states $\{s_t\}$, $t=1, 2, \ldots$, in the game process makes a Markov chain with the distribution $p_t(\cdot|\bar{\pi})$, $t=1, 2, \ldots\ldots$, which is repeatedly given by

$$p_t(s_t|\bar{\pi})=\sum_{s'\in S} q_{t-1}(s_t|s', \bar{\pi})p_{t-1}(s'|\bar{\pi}) \tag{2.4}$$

and

$$p_1(s_1|\bar{\pi})=p_1(s_1) \qquad \text{for all } s_1\in S,$$

where, for all $\bar{\pi}=(\pi^1, \pi^2, \ldots, \pi^m)$, $\pi^i=(\pi_1^i, \pi_2^i, \ldots\ldots)$ and all $s', s\in S$,

$$q_t(s'|s, \bar{\pi})=\int_A q(s'|s, \bar{a})d\bar{\pi}_t(\bar{a}|s) \tag{2.5}$$

and

$$d\bar{\pi}_t(\bar{a}|s)=\prod_{i=1}^m d\pi_t^i(a^i|s), \tag{2.6}$$

$$\bar{a}=(a^1, a^2, \ldots, a^m)\in A=\prod_{i=1}^m A^i$$

(the probability that the chain moves from the state s at time t to a new state s' at the next time).

And, the expected loss of the ith player at each time t, $t=1, 2, \ldots$, is given by

$$E_{\bar{\pi}}[r^i(s_t, t, \bar{\pi})]=\sum_{s_t} r^i(s_t, t, \bar{\pi})p_t(s_t|\bar{\pi}) \tag{2.7}$$

where,

$$r^i(s, t, \bar{\pi})=\int_A r^i(s, \bar{a})\,d\bar{\pi}_t(\bar{a}|s)$$

(the loss of the ith player at the state $s_t=s$).

So, if a multistrategy $\bar{\pi}=(\pi^1, \pi^2, \ldots, \pi^m)$ is chosen, under the discount factor β, the total expected discounted loss of the ith player is defined by

$$\begin{aligned} I^i(\bar{\pi}) &= \sum_{t=1}^{\infty} \beta^{t-1}E_{\bar{\pi}}[r^i(s_t, t, \bar{\pi})] \\ &= \sum_{t=1}^{\infty} \beta^{t-1}\sum_{s} r^i(s, t, \bar{\pi})p_t(s|\bar{\pi}) \\ &= E_{\bar{\pi}}[\sum_{t=1}^{\infty} \beta^{t-1}r^i(s_t, t, \bar{\pi})]. \end{aligned} \tag{2.8}$$

The vector expresion for $I^i(\bar{\pi})$ is given by

$$\begin{aligned} I(\bar{\pi}) &= (I^1(\bar{\pi}), I^2(\bar{\pi}), \ldots I^m(\bar{\pi})) \\ &= \sum_{t=1}^{\infty} \beta^{t-1}E_{\bar{\pi}}[r(s_t, t, \bar{\pi})], \end{aligned} \tag{2.9}$$

where

$$r(s_t, t, \bar{\pi})=(\ldots, r^i(s_t, t, \bar{\pi}), \ldots)_{i=1}^m$$

In our game system, all players cooperate in choosing a multistrategy $\bar{\pi}$ to minimize the distance from $I(\bar{\pi})$ to some given point $z\in R^m$. This means that no other multistrategy yields a smaller total expected discounted loss in the sense of the distance from the point z. In other words, for the given point $z=(z^1, z^2, \dots, z^m)\in R^m$, the players make a collective loss function

$$\| I(\bar{\pi})-z\|=\sqrt{\sum_{i=1}^{m} |I^i(\bar{\pi})-z^i|^2}, \tag{2.10}$$

in which they wish to find an optimal multistrategy to minimize (2. 10) over $\bar{\pi}\in \Pi = \prod_{i=1}^{m} \Pi^i$. But, in general, it is difficult to find an optimal multistrategy. So, we would develop the discussion to find a weighting factor d_* and a multistrategy $\bar{\pi}^*$ such that

$$\begin{aligned} \inf_{\bar{\pi}} \|I(\bar{\pi})-z\| &= \max_{\|d\|\leqq 1} \min_{\bar{\pi}} [\langle d, I(\bar{\pi})\rangle - \langle d, z\rangle] \\ &= \langle d_*, I(\bar{\pi}^*)\rangle - \langle d_*, z\rangle, \end{aligned} \tag{2.11}$$

where, in (2. 11), $\langle \cdot, \cdot \rangle$ denotes the inner product.

3. Preliminary lemmas in the game system

In order to prove the necessary lemmas in the game system, we need an assumption on the convexity of the set of all total expected discounted multiloss. So, we introduce a notation as follows

$$K=\{ I(\bar{\pi}) \text{ for all } \bar{\pi}\in \prod_{i=1}^{m} \Pi^i=\Pi \}. \tag{3.1}$$

and impose the following assumption on the set K

(A1) K is a convex subset in R^m, that is, for all $\bar{\pi}_1, \bar{\pi}_2\in \Pi$ and all α, $0<\alpha<1$

$$\alpha I(\bar{\pi}_1)+(1-\alpha)I(\bar{\pi}_2)\in K.$$

For the set K, the function $\delta(d|K)=\inf_{\bar{\pi}} \langle d, I(\bar{\pi})\rangle$ defined on R^m is said to be the support function of K.

Lemma 1. *If the distance ρ_0 from the origin to K is positive under* (A1), *then*

$$\rho_0=\inf_{\bar{\pi}} \|I(\bar{\pi})\|=\max_{\|d\|\leqq 1} \delta(d|K), \tag{3.2}$$

where the maximum on the right side of (3. 2) *is attained by some $d_*\in R^m$, $\|d_*\|=1$, that is,*

$$\rho_0=\delta(d_*|K).$$

Proof. In order to show that $\rho_0\geqq\delta(d|K)$ for all $d\in R^m$, $\|d\| \leqq 1$, we may limit to those *d's* which render the support function $\delta(d|K)$ positive. If $\delta(d|K)>0$, the half-space H_+

$$H_+=\{ x\in R^m \mid \langle d, x\rangle \geqq \delta(d|K) \}\supset K.$$

And $0\notin H_+$ because of $\langle d, 0\rangle=0$. So, since K is convex, the hyperplane H

$$H=\{ x\in R^m \mid \langle d, x\rangle=\delta(d|K) \}$$

separates K and the origin 0. Now, let $S(\varepsilon)$ be an open sphere with radius $\varepsilon>0$ centered at 0. For any $d \in R^m$, $\|d\| \leqq 1$, having $\delta(d|K) > 0$, let ε^* be the supremum of those ε's for which the hyperplane H separates K and $S(\varepsilon)$. Then, $0 \leqq \varepsilon^* \leqq \rho_0$, and $\delta(d|K) = \sup_{\|x\| < \varepsilon^*} < d, x > \leqq \varepsilon^*$. Thus, for every $d \in R^m$, $\|d\| \leqq 1$, we have $\delta(d|K) \leqq \rho_0$.

On the other hand, since $K \cap S(\rho_0) = \phi$, there is a hyperplane separating $S(\rho_0)$ and K. Therefore, there is a $d_* \in R^m$, $\|d_*\| = 1$, such that $\delta(d_*|K) = \rho_0$, that is,

$$\max_{\|d\| \leqq 1} \delta(d|K) = \delta(d_*|K) = \rho_0.$$

Whence the proof is completed.

LEMMA 2. *If the distance ρ_1 from the given point $z \in R^m$, $z \neq 0$, to K is positive under* (A1), *then*

$$\rho_1 = \inf_{\bar{\pi}} \|I(\bar{\pi}) - z\| = \max_{\|d\| \leqq 1} [\delta(d|K) - < d, z >], \tag{3.3}$$

where the maximum on the right side of (3.3) *is attained by some $d_* \in R^m$, $\|d_*\| = 1$.*

PROOF. Putting $K' = K - z$, K' is a convex set in R^m according to (A1). So, from Lemma 1, it follows that

$$\begin{aligned} \rho_1 = \inf_{x \in K'} \|x\| &= \max_{\|d\| \leqq 1} \delta(d|K') \\ &= \delta(d_*|K'),\ \|d_*\| = 1. \end{aligned} \tag{3.4}$$

Then, (3.4) can be rewritten as follows

$$\begin{aligned} \rho_1 &= \inf_{\bar{\pi} \in \Pi} \|I(\bar{\pi}) - z\| = \max_{\|d\| \leqq 1} [\inf_{\bar{\pi} \in \Pi} < d, I(\bar{\pi}) - z >] \\ &= \max_{\|d\| \leqq 1} [\delta(d|K) - < d, z >] \\ &= \delta(d_*|K) - < d_*, z >. \end{aligned}$$

Thus, this completes the proof.

Now, it is necessary to introduce the definitions of an optimal multistrategy and a weak optimal multistrategy.

DEFINITION 1. A multistrategy $\bar{\pi}^*$ is called an optimal multistrategy for the point z in the game system if $\bar{\pi}^*$ satisfies

$$\|I(\bar{\pi}^*) - z\| \leqq \|I(\bar{\pi}) - z\| \qquad \text{for all } \bar{\pi} \in \Pi.$$

DEFINITION 2. A multistrategy $\bar{\pi}^*$ is called a weak optimal multistrategy for the point z (with respect to a weighting factor $\bar{d}$) if $\bar{\pi}^*$ satisfies

$$\inf_{\bar{\pi}} \|I(\bar{\pi}) - z\| = < \bar{d}, I(\bar{\pi}^*) > - < \bar{d}, z >.$$

4. The existence of a weak optimal multistrategy in the game system

In this game system, we assume that each player uses the stationary multistrategy so that the stationary multistrategy is specified as a multiprobability measure:

$$\bar{\mu} = (\mu^1, \mu^2, \ldots, \mu^m) \in \prod_{i=1}^{m} P(A^i) = P(A)$$

which is depending on the state. For this stationary mulistrategy $\bar{\mu}$ and the state s, we may write the loss function of the ith player and the transition probability measure as follows

$$r^i(s, \bar{\mu}) = \int_A r^i(s, \bar{a}) d\bar{\mu}(\bar{a} \mid s) \tag{4.1}$$

and

$$q(\cdot \mid s, \bar{\mu}) = \int_A q(\cdot \mid s, \bar{a}) d\bar{\mu}(\bar{a} \mid s), \tag{4.2}$$

where

$$\bar{a} \in \prod_{i=1}^{m} A^i = A \quad \text{and} \quad d\bar{\mu}(\bar{a} \mid s) = \prod_{i=1}^{m} d\mu^i(a^i \mid s).$$

Since we have assumed that each A^i is a compact metric space, it is separable and so $C(A^i)$, the space of all continuous functions on A^i, is a separable Banach space with the supnorm. The dual space $C(A^i)^* = M(A^i, B(A^i))$ of $C(A^i)$ is a bounded regular measure space, so that the probability measure space $P(A^i)$ in $C(A^i)^*$ is weak* compact. Since $C(A^i)^*$ is separable, the unit sphere $P(A^i)$ in $C(A^i)^*$ is weak* compact metrizable subspace, it follows that $P(A) = \prod_{i=1}^{m} P(A^i)$ is weak* compact metrizable subspace in $C(A)^*$, where $C(A) = \prod_{i=1}^{m} C(A^i)$.

In order to prove main results in the game system (2.1), we need some additional assumptions on q and r^i, $i = 1, 2, \ldots, m$.

(A2) Let $q(s' \mid s, \bar{a})$ be continuous function on $\bar{a} \in A$ for every $(s', s) \in S \times S$.

(A3) The loss function $r^i(s, \bar{a})$ of the ith player is bounded on $S \times A$ and is continuous on A for every $s \in S$.

Then, we would show that there exists a weak optimal multistrategy which minimizes the distance ρ_0 from the origin to K. From Lemma 1, it is sufficient to find an optimal multistrategy of the total expected discounted numerical loss function $\langle d_*, I(\pi) \rangle$ weighting by the factor d_* on Π. And, from (2.9), this numerical loss function can be rewritten as

$$\begin{aligned} \langle d_*, I(\pi) \rangle &= \sum_{i=1}^{m} d_*^i I^i(\pi), \quad d_* = (d_*^1, d_*^2, \ldots, d_*^m) \qquad (4.3) \\ &= \sum_{i=1}^{m} d_*^i \sum_{t=1}^{\infty} \beta^{t-1} \sum_{s} r^i(s, t, \pi) p_t(s \mid \pi) \\ &= \sum_{t=1}^{\infty} \beta^{t-1} E_\pi [\langle d_*, r(s_t, t, \pi) \rangle]. \end{aligned}$$

So we modify our game system (2.1) to one with the numerical loss function

$$(S, A, q, \langle d_*, r \rangle, \beta), \tag{4.4}$$

where

$$A=\prod_{i=1}^{m} A^i \quad \text{and} \quad <d_*, r> = \sum_{i=1}^{m} d_*^i r^i.$$

In the new game system, the form of the collective function is different from the original game system (2. 1) and, from (A2) and (A3), $<d_*, r>$ is a bounded and continuous loss function on $S\times A$.

Let $C(S)$ be the set of all bounded (continuous) real valued function on the countable state space S. Then, for a weighting factor d_* in Lemma 1, we define an operator T_0 on $C(S)$ by

$$T_0u(s)=\min_{\bar{\mu}\in P(A)}[<d_*, r(s, \bar{\mu})>+\beta\sum_{s'} u(s')q(s'\,|\,s, \bar{\mu})], \tag{4. 5}$$

where

$$P(A)=\prod_{i=1}^{m} P(A^i)$$

and

$$<d_*, r(s, \bar{\mu})> = \sum_{i=1}^{m} d_*^i r^i (s, \bar{\mu}) = \int_A \sum_{i=1}^{m} d_*^i r^i(s, \bar{a})d\bar{\mu}(\bar{a}\,|\,s).$$

Evidently, $T_0u(s)\in C(S)$ whenever $u\in C(S)$. For simplicity, we let

$$L(\bar{\mu})u(s)=<d_*, r(s, \bar{\mu})>+\beta\sum_{s'} u(s')q(s'\,|\,s, \bar{\mu}). \tag{4. 6}$$

Thus, the expression of (4. 5) can be rewritten by

$$T_0u(s)=\min_{\bar{\mu}\in P(A)} L(\bar{\mu})u(s).$$

In our discussion, it is important that T_0 is a contraction operator on $C(S)$.

THEOREM 1. *Suppose that the game system* (2. 1) *satisfies Assumptions* (A1), (A2) *and* (A3). *Then, if the distance* ρ_0 *from the origin to* K *is positive, there exists a weak optimal stationary multistrategy* $\bar{\mu}^*\in \Pi$ *with respect to a weighting factor* d_* *such that*

$$\begin{aligned}\rho_0=\inf_{\bar{\pi}} \|I(\bar{\pi})\| &= \max_{\|d\|\leq 1} \delta(d\,|\,K) \\ &= <d_*, I(\bar{\mu}^*)>,\end{aligned} \tag{4. 7}$$

where $\delta(d\,|\,K)$ *is the support function of* K.

If the infimum in the left side in (4. 7) *is attained by some multistrategy* $\bar{\pi}_0$, *that is, an optimal multistrategy, then,*

$$\rho_0=\|I(\bar{\pi}_0)\| =<d_*, I(\bar{\mu}^*)>=<d_*, I(\bar{\pi}_0)>. \tag{4. 8}$$

PROOF. By Lemma 1, (4. 7) can be rewritten as

$$\begin{aligned}\rho_0=\inf_{\bar{\pi}} \|I(\bar{\pi})\| &= \max_{\|d\|\leq 1} \delta(d\,|\,K) \\ &=\delta(d_*\,|\,K) \\ &=\inf_{\bar{\pi}} <d_*, I(\bar{\pi})>.\end{aligned}$$

So, we would find a weak optimal multistrategy to minimize $<d_*, I(\bar{\pi})>$ on Π. The operator T_0 defined by (4. 5) is a contraction operator on $C(S)$ because of the discount factor β, $0<\beta<1$. Since $C(S)$ is a Banach space with supnorm, it follows that T_0 has a unique fixed point in $C(S)$, say u^*, then for each $s\in S$ we have

$$u^*(s)=T_0u^*(s)=\min_{\bar{\mu}} L(\bar{\mu})u^*(s). \tag{4.9}$$

Moreover, since $L(\bar{\mu})u(s)$ is continuous on the compact set $P(A)$ by (A2) and (A3), the minimum of (4.9) is attained by a stationary multistrategy $\bar{\mu}^*$, whence for each $s\in S$

$$\begin{aligned} u^*(s)&=L(\bar{\mu}^*)u^*(s) \\ &\leqq L(\bar{\mu})u^*(s) \qquad \text{for all } \bar{\mu}\in P(A). \end{aligned} \tag{4.10}$$

Consequently, from the iterative substitution for u^* in the first equation of (4.10), we obtain

$$u^*(s)=\langle d_*, I(\bar{\mu}^*)(s)\rangle \qquad \text{for all initial state } s\in S.$$

On the other hand, using a similar argument to the inequality of (4.10), we obtain

$$u^*(s)\leqq \langle d_*, I(\bar{\pi})(s)\rangle \qquad \text{for all initial state } s\in S.$$

These iterative methods are given, in detail, in the proof of Theorem 4.1 in [**9**]. Since, for any probability measure $p\in P(S)$,

$$\sum_s \langle d_*, I(\bar{\pi})(s)\rangle p(s)=\langle d_*, I(\bar{\pi})\rangle,$$

we have

$$\langle d_*, I(\bar{\mu}^*)\rangle \leqq \langle d_*, I(\bar{\pi})\rangle \qquad \text{for all } \bar{\pi}\in \Pi.$$

To prove (4.8), suppose that $\bar{\pi}_0$ is an optimal multistrategy such that

$$\rho_0=\|I(\bar{\pi}_0)\|=\inf_{\bar{\pi}} \|I(\bar{\pi})\|.$$

Since $I(\bar{\pi}_0)\in K$, we have

$$\begin{aligned} \langle d_*, I(\bar{\pi}_0)\rangle &\geqq \inf_{\bar{\pi}} \langle d_*, I(\bar{\pi})\rangle \\ &= \langle d_*, I(\bar{\mu}^*)\rangle=\inf_{\bar{\pi}}\|I(\bar{\pi})\|=\rho_0. \end{aligned} \tag{4.11}$$

However, since $\|d_*\|\leqq 1$, we have

$$\begin{aligned} \langle d_*, I(\bar{\pi}_0)\rangle &\leqq \|d_*\|\,\|I(\bar{\pi}_0)\| \\ &\leqq \inf_{\bar{\pi}} \|I(\bar{\pi})\| = \langle d_*, I(\bar{\mu}^*)\rangle =\rho_0. \end{aligned} \tag{4.12}$$

So, from (4.11) and (4.12), we obtain (4.8). Whence, the proof is completed.

Now, in order to find a weak optimal multistrategy which minimizes the distance ρ_1 from the given point $z\neq 0$ to K. From Lemma 2, we would find an optimal multistrategy $\bar{\pi}^*$ of the total expected discounted numerical loss function $\langle d_*, I(\bar{\pi})-z\rangle$ on Π. And, this numerical loss function can be rewritten as

$$\langle d_*,I(\bar{\pi})-z\rangle = \langle d_*, I(\bar{\pi})\rangle - \langle d_*, z\rangle \tag{4.13}$$

$$=\sum_{t=1}^{\infty}\beta^{t-1}\sum_{s}p_t(s\,|\,\bar{\pi})\,[\langle d_*, r(s,t,\bar{\pi})-(1-\beta)z\rangle]$$

$$=\sum_{t=1}^{\infty}\beta^{t-1}E_{\bar{\pi}}[\langle d_*, r(s,t,\bar{\pi})-(1-\beta)z\rangle].$$

So, we modify game system (2.1) to one with the collective loss function

$$(S, A, q, \langle d_*, r-(1-\beta)z\rangle, \beta). \tag{4.14}$$

Then, for a weighting factor d_* in Lemma 2, we define an operator T_1 on $C(S)$ by

$$T_1u(s)=\min_{\bar{\mu}\in P(A)}[\langle d_*, r-(1-\beta)z\rangle+\beta\sum_{s'}u(s')q(s'\,|\,s,\bar{\mu})].$$

Using a similar argument to the proof of Theorem 1 with the operator T_1 instead of T_0, we can prove the following theorem.

THEOREM 2. *Suppose that the game system* (2.1) *satisfies Assumptions* (A1), (A2) *and* (A3). *Then, if the distance ρ_1 from a given point $z\neq 0$ to K is positive, there exists a weak optimal stationary multistrategy $\bar{\mu}^*$ with respect to a weighting factor d_* such that*

$$\rho_1=\inf_{\bar{\pi}}\|I(\bar{\pi})-z\|=\max_{\|d\|\leq 1}[\delta(d\,|\,K)-\langle d, z\rangle]$$
$$=\delta(d_*\,|\,K)-\langle d_*, z\rangle.$$

If there exists an optimal multistrategy $\bar{\pi}_1^*$ in the game system (4.14), *then*

$$\rho_1=\|I(\bar{\pi}_1^*)-z\|=\langle d_*, I(\bar{\mu}^*)-z\rangle$$
$$=\langle d_*, I(\bar{\pi}_1^*)-z\rangle.$$

5. The relation between a D-solution and a weak optimal multistrategy in the game system

In order to show that a weak optimal stationary multistrategy in Theorem 1 is a D-solution under a domination structure determined by some convex cone D, we need the concept of convex cone.

A subset F in R^m is said to be a cone with vertex at the origin 0 if $x\in F$ implies that $\lambda x\in F$ for all $\lambda\geqq 0$. A convex cone is, of course, defined as a set which is both convex and cone.

Now, in this section, consider a subset $L\subset R^m$ such that

(i) $L\not\ni 0$

(ii) $L^+=\{y\in R^m \mid \langle x, y\rangle>0 \quad \text{for all } x\in L\}\neq\phi$

(iii) $L\cup\{0\}=D$ is a convex cone with vertex at the origin 0, where $\langle\cdot,\cdot\rangle$ denotes the inner product.

Note that this L is a convex cone without the vertex 0 in R^m.

DEFINITION 3. A multistrategy $\bar{\pi}^*$ is a D-solution if there is no other multistrategy $\bar{\pi}\in\Pi$ such that

$$I(\bar{\pi}^*)\in I(\bar{\pi})+L. \tag{5.1}$$

REMARK. For a given closed convex cone F in R^m, if $L=\text{int } F$ (resp. $L=F-\{0\}$), the multistrategy $\bar{\pi}^*$ in (5.1) is usually said to be a F-weak solution (resp. F-strong solution), where int$F$ denotes the interior of the set F. See Part II in Aubin [1] for the concepts of a weak Pareto solution and a strong Pareto solution in a game system.

Now, in Theorem 1, under Assumptions (A1), (A2), (A3) and $\rho_0>0$, we show that there exists a weak optimal stationary multistrategy $\bar{\mu}^*\in \Pi$ with respect to a weighting factor d_* such that

$$\rho_0=\inf_{\bar{\pi}} \|I(\bar{\pi})\|=\delta(d_*|K)=\langle d_*,I(\bar{\mu}^*)\rangle, \tag{5.2}$$

where $\delta(d_*|K)$ is the support function of K.

Then, for a weighting factor d_* and a weak optimal multistrategy $\bar{\mu}^*$ in Theorem 1, we can prove the following theorem.

THEOREM 3. For a weighting factor d_* and a weak optimal stationary multistrategy $\bar{\mu}^*$ in Theorem 1, if $d_*\in L^+$, the multistrategy $\bar{\mu}^*$ is a D-solution.

PROOF. Suppose that $\bar{\mu}^*$ is not D-solution. Then, from (5.1), there is a $\bar{\pi}\in \Pi$ such that

$$I(\bar{\pi}^*)\in I(\bar{\pi})+L,$$

that is, there exists $\bar{d}\in L$ such that

$$I(\bar{\pi}^*)=I(\bar{\pi})+\bar{d}. \tag{5.3}$$

Taking an inner product on both sides of (5.3) with the weighting factor $d_*\in L^+$, we obtain an inequality

$$\langle d_*,I(\bar{\pi})\rangle < \langle d_*, I(\bar{\mu}^*)\rangle.$$

This contradicts to the fact that

$$\delta(d_*|K)=\langle d_*, I(\bar{\mu}^*)\rangle$$

and the proof is completed.

References

[1] J. P. Aubin, Mathematical Methods of Game and Economic Theory, North-Holland, Amsterdam, 1979.

[2] J. P. Aubin, Applied Functional Analysis, Wiley-Interscience, New York, 1979.

[3] M. S. Bazaraa and C. M. Shetty, Nonlinear Programming, John Wiley and Sons, New York, 1979.

[4] P. Billingsley, Convergence of Probability Measures, Wiley-Interscience, New York, 1968.

[5] K. Fan, *Fixed point and minimax theorem in locally convex topological linear spaces*, Proc. Nat. Acad. Sci. U. S. A., 38 (1958), 121–126.

[6] H. C. Lai and K. Tanaka, *Noncooperative n-person game with a stopped set*, J. Math. Anal. Appl., 88 (1982), 153–171.

[7] H. C. Lai and K. Tanaka, *A noncooperative n-person semi-Markov game with a separable metric space*, Appl. Math. Optim., 11 (1984), 23–42.

[8] H. C. Lai and K. Tanaka, *On an N-person noncooperative Markov game with a metric state space*, J. Math. Anal. Appl., 101 (1984), 78–96.

[9] H. C. Lai and K. Tanaka, *On a D-solution of a cooperative m-person discounted Markov game*, J Math. Anal. Appl., 115 (1986), 578–591.

[10] H. C. Lai and K. Tanaka, *An N-person noncooperative discounted vector-valued dynamic game with a metric space*, to appear in Appl. Math. Optim.

[11] H. C. Lai and K. Tanaka, *An N-person noncooperative discounted vector-valued dynamic game with a stopped set*, to appear in J. Computer and Math. with Applications (Special Issue, Pursuit-Evasion Games).

[12] H. C. Lai and K. Tanaka, *A vector-minimization problem in a stochastic continuous-time n-person game*, to appear in Proceeding of the Conference on Functional Analysis in honor of Professor Ky Fan, Lecture Notes in Pure and Applied Mathematics, Marcel Dekker.

[13] H. C. Lai and K. Tanaka, *On vector reward n-person cooperative game, preprint.*

[14] D. G. Luenberger, Optimization by Vector Space Methods, Wiley-Interscience, New York, 1969.

[15] K. Tanaka, *On the learning algorithm of 2-person zero-sum Markov game with expected average reward criterion*, Bull. Inform. Cyber., 21 (1985), 1–17.

[16] K. Tanaka, *On some vector valued Markov game*, Japan, J. Appl. Math., 2 (1985), 293–308.

[17] K. Tanaka and H. C. Lai, *A two-person zero-sum Markov game with a stopped set*, J. Math. Anal. Appl., 86 (1982), 54–68.

[18] K. Tanaka and Y. Maruyama, *The multiobjective optimization problem of set function*, J. Information and Optimization Sciences, 5 (1984), 293–306.

Liu Zhaohua
Department of Mathematics
Harbin Teacher's University
Harbin,
The People's Republic of China

Kensuke Tanaka
Department of Mathematics
Faculty of Science
Niigata University
Niigata, 950–21, Japan