On Markov games with the expected average reward criterion (II)

By

Kensuke TANAKA and Kazuyoshi WAKUTA

(Received September 15, 1975)

1. Introduction

This paper is a continuation of our paper with the same title [2] in which we have showed, mainly, that under some assumptions the Markov games with the criterion of long-run average reward has a value and both players have optimal strategies.

Here, we shall show that the solution $(d, u(\cdot))$ of the functional equation (3. 1) of assumption 4 in [2] can be solved by the method of successive approximations under certain conditions. We can obtain this method as an application of the method introduced by D. J. White in Dynamic Programming [3].

2. Preliminaries

In order to state the method of successive approximations, we assume the same conditions as those in [2], that is, (1) S, A and B are compact metric spaces, (2) $r=r(s, a, b)$ is a continuous function on $S\times A\times B$, (3) 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 $q(\cdot\,|\,s_0, a_0, b_0)$, (4) there exist a continuous function $u(s)$ on S and a constant d such that for each $s\in S$,

$$d+u(s)=\sup_{\lambda\in P_A}\inf_{\mu\in P_B}\{r(s, \mu, \lambda)+\int u(s')\,dq(s'\,|\,s, \mu, \lambda)\}, \tag{2. 1}$$

where P_A and P_B are the sets of all probability measures on $(A, \mathfrak{B}(A))$ and $(B, \mathfrak{B}(B))$, respectively,

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

and

$$q(E\,|\,s, \mu, \lambda)=\int\int q(E\,|\,s, a, b)\,d\mu(a)\,d\lambda(b).$$

Moreover, we need the same lemmas as those in [2], that is,

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

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

LEMMA 3. *Let $u(x, y)$ be a bounded, continuous function on $X\times Y$, where X is a Borel subset of a Polish space and Y is a compact metric space. Then, there exist Borel measurable maps f and g from X to Y such that $u(x, f(x))=\max_{y\in Y} u(x, y)$, $x\in X$ and $u(x, g(x))=\min_{y\in Y} u(x, y)$, $x\in X$.*

The reader is referred to [1] for the proofs,

Then, from Lemma 1 and Lemma 2, the equation (1, 1) can be replaced by the following:

$$d+u(s)=\max_{\mu\in P_A}\min_{\lambda\in P_B}\{r(s, \mu, \lambda)+\int u(s')dq(s'|s, \mu, \lambda)\} \tag{2. 2}$$

$$=\min_{\lambda\in P_B}\max_{\mu\in P_A}\{r(s, \mu\ \lambda)+\int u(s')dq(s'|s, \mu, \lambda)\}.$$

We want to show that the solution $(u(\cdot), d)$ of (2. 2) can be solved by the method of successive approximations.

3. The method of successive approximations

Let Π^n be the set of all Markov policies for player I until the nth time and Γ^n the set of all Markov policies for player II until the nth time. That is, for $\pi\in\Pi^n$ and $\sigma\in\Gamma^n$, $\pi=(f_1, f_2,\cdots, f_n)$ and $\sigma=(g_1,\cdots, g_n)$, where f_k and g_k, $k=1, 2,\cdots, n$, are Borel measurable maps from S to P_A and P_B, respectively. Then, for each $\pi=(f_1, f_2,\cdots, f_n)\in\Pi^n$, $\sigma=(g_1, g_2,\cdots, g_n)\in\Gamma^n$, $E\in\mathcal{B}(S)$ and $s_1\in S$, we define the following:

$$q^{(n)}(E|s_1, \pi, \sigma)=\int\cdots\int q(E|s_n, f_n(s_n), g_n(s_n))$$

$$\prod_{i=1}^{n-1} dq(s_{i+1}|s_i, f_i(s_i), g_i(s_i)). \tag{3. 1}$$

And, moreover, we assume the following assumption.

ASSUMPTION. There exist an integer $u\geqq 0$, a quantity $\alpha(0<\alpha\leqq 1)$ and a state $s_0\in S$ such that, for each $\pi=(f_1, f_2,\cdots, f_{u+1})\in\Pi^{u+1}$, $\sigma=(g_1, g_2,\cdots, g_{u+1})\in\Gamma^{u+1}$ and $s_1\in S$,

$$q^{(u+1)}(s_0|s_1, \pi, \sigma)\geqq\alpha>0. \tag{3. 2}$$

Then, it should be noted that assumption 5 in [2] implies this assumption.

THEOREM. *Under the above assumption, the sequence* $\{d_n, v_n(\cdot), n\geqq 0\}$ *defined by,* $s\in S$

$$V_n(s)=\max_{\mu\in P_A}\min_{\lambda\in P_B}\{r(s,\mu,\lambda)+\int_S v_{n-1}(s')dq(s'|s,\mu,\lambda)\},$$

$$d_n=V_n(s_0), \tag{3. 3}$$

$$v_n(s)=V_n(s)-d_n,$$

converges uniformly to the solution $\{d, u(\cdot)\}$ *of the functional equation* (2. 1), *where*

$$v_0(s)=\max_{\mu\in P_A}\min_{\lambda\in P_B} r(s,\mu,\lambda).$$

PROOF. By Lemma 1 and Lemma 2, it follows that $V_n(\cdot)$, $v_n(\cdot)$ are well defined and continuous. We need only prove the uniformity of convergence, since this implies that the limiting form is a solution of (2. 1). For any sequence of continuous functions $\{Z_n(s)\}$ define

$$\nabla_n(Z)=\inf_{s\in S}[Z_n(s)-Z_{n-1}(s)] \tag{3. 4}$$

and

$$\Delta_n(Z)=\sup_{s\in S}[Z_n(s)-Z_{n-1}(s)]. \tag{3. 5}$$

When $n\geqq u+4$, we can show that

$$V_n(s)-V_{n-1}(s) \tag{3. 6}$$

$$\geqq\min_{\mu}[\min_{\lambda}\{r(s,\mu,\lambda)+\int v_{n-1}(s')dq(s'|s,\mu,\lambda)\}$$

$$-\min_{\lambda}\{r(s,\mu,\lambda)+\int v_{n-2}(s')dq(s'|s,\mu,\lambda)\}]$$

$$\geqq\min_{\mu}[\min_{\lambda}\int(v_{n-1}(s')-v_{n-2}(s'))dq(s'|s,\mu,\lambda)]$$

$$=\min_{\mu,\lambda}[\int(v_{n-1}(s')-v_{n-2}(s'))dq(s'|s,\mu,\lambda)]$$

$$=\min_{\mu,\lambda}[\int(V_{n-1}(s')-V_{n-2}(s'))dq(s'|s,\mu,\lambda)]$$

$$-(d_{n-1}-d_{n-2}).$$

Hence,

$$V_n(s)-V_{n-1}(s)+(d_{n-1}-d_{n-2})$$

$$\geqq\min_{\mu,\lambda}[\int(V_{n-1}(s')-V_{n-2}(s'))dq(s'|,\mu,\lambda)]. \tag{3. 7}$$

Then, by Lemma 3, we can show that

$$V_n(s)-V_{n-1}(s)+(d_{n-1}-d_{n-2})+(d_{n-2}-d_{n-3}) \tag{3. 8}$$

$$\geqq \min_{\mu,\lambda}\Big[\int\{\min_{\mu',\lambda'}\int(V_{n-2}(s'')-V_{n-3}(s''))dq(s''|s',\mu',\lambda')\}$$

$$dq(s'|s,\mu,\lambda)\Big].$$

$$\geqq \min_{\pi\in\Pi^2,\sigma\in\Gamma^2}\Big[\int(V_{n-2}(s')-V_{n-3}(s'))dq^{(2)}(s'|s,\pi,\sigma)\Big].$$

Repeating this process we derive

$$V_n(s)-V_{n-1}(s)+\sum_{k=1}^{u}(d_{n-k}-d_{n-k-1}) \tag{3. 9}$$

$$\geqq \min_{\pi\in\Pi^u,\sigma\in\Gamma^u}\Big[\int(V_{n-u}(s')-V_{n-u-1}(s'))dq^{(u)}(s'|s,\pi,\sigma)\Big],$$

namely,

$$V_n(s)-V_{n-1}(s)+(d_{n-1}-d_{n-u}) \tag{3. 10}$$

$$\geqq \min_{\pi\in\Pi^{u+1},\sigma\in\Gamma^{u+1}}\Big[\int(v_{n-u-1}(s')-v_{n-u-2}(s'))dq^{(u+1)}(s'|s,\pi,\sigma)\Big].$$

Now by hypothesis (3. 3)

$$v_{n-u-1}(s_0)=v_{n-u-2}(s_0)=0$$

and

$$q^{(u+1)}(s_0|s,\pi,\sigma)\geqq\alpha>0 \quad \text{for all} \quad \pi\in\Pi^{u+1},\ \sigma\in\Gamma^{u+1} \quad \text{and} \quad s\in S.$$

Thus, we have

$$\nabla_n(V)+(d_{n-1}-d_{n-u})\geqq(1-\alpha)\nabla_{n-u-1}(v). \tag{3. 11}$$

Similarly, for $n\geqq u+4$,

$$V_n(s)-V_{n-1}(s)+(d_{n-1}-d_{n-u})$$

$$\leqq \max_{\pi\in\Pi^{u+1},\sigma\in\Gamma^{u+1}}\Big[\int(v_{n-u-1}(s')-v_{n-u-2}(s'))$$

$$dq^{(u+1)}(s'|s,\pi,\sigma)\Big]. \tag{3. 12}$$

Thus, by hypothesis (3. 3),

$$\Delta_n(V)+(d_{n-1}-d_{n-u})\leqq(1-\alpha)\Delta_{n-u-1}(v). \tag{3. 13}$$

(3. 11) and (3. 13) imply that

$$\Delta_n(V)-\nabla_n(V)\leqq(1-\alpha)(\Delta_{n-u-1}(v)-\nabla_{n-u-1}(v)). \tag{3. 14}$$

At the same time,

$$\nabla_n(v)=\nabla_n(V)-(d_n-d_{n-1})$$

and

$$\Delta_n(v)=\Delta_n(V)-(d_n-d_{n-1})$$

yield

$$\Delta_n(v)-\nabla_n(v)=\Delta_n(V)-\nabla_n(V). \tag{3. 15}$$

Thus, by (3. 15), (3. 14) becomes

$$\Delta_n(v)-\nabla_n(v)\leqq(1-\alpha)(\Delta_{n-u-1}(v)-\nabla_{n-u-1}(v)).$$

Let $D_n(v)=\Delta_n(v)-\nabla_n(v)$. Then, for $n=N(u+1)+r$ $(1\leqq r\leqq u+1)$,

$$D_n(v)\leqq(1-\alpha)^N D_r(v)\leqq(1-\alpha)^N A,$$

where $A=\max\{D_1(v), D_2(v),\ldots, D_{u+1}(v)\}$.

Since $v_n(s_0)-v_{n-1}(s_0)=0$, it follows that

$$\nabla_n(v)\leqq 0\leqq\Delta_n(v).$$

Let

$$U_n(v)=\sup_{s\in S}|v_n(s)-v_{n-1}(s)|$$

and

$$U_n(V)=\sup_{s\in S}|V_n(s)-V_{n-1}(s)|,$$

then

$$U_n(v)\leqq D_n(v)\leqq(1-\alpha)^N A. \tag{3. 16}$$

This is sufficient to prove that the sequence $\{v_n(s)\}$ converges uniformly to the function $v(s)$.

On the other hand, we easily see that

$$U_n(V)\leqq U_{n-1}(v). \tag{3. 17}$$

From (3. 17), the sequence $\{V_n(s)\}$ also converges uniformly to a function $V(s)$. Since $\{V_n(s)\}$ converges uniformly, so does the sequence $\{d_n\}=\{V_n(s_0)\}$ converge. Thus, the proof is complete.

References

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

[2] K. TANAKA, S. IWASE and K. WAKUTA, *On Markov games with the expected average reward criterion*, Sci. Rep. Niigata Univ., Ser. A, No. 13 (1976), 31–41.

[3] D. J. WHITE, *Dynamic Programming, Markov chains, and the method of successive approximations*, Journ. Math. Anal. and Appli., 6 (1963), 373–376.