JOURNAL OF OPTIMIZATION THEORY AND APPLICATIONS: Vol. 76, No. 3, MARCH 1993

Denumerable State Stochastic Games with Limiting Average Payoff[1]

V. S. Borkar[2] and M. K. Ghosh[3]

Communicated by P. Varaiya

Abstract. We study stochastic games with countable state space, compact action spaces, and limiting average payoff. For N-person games, the existence of an equilibrium in stationary strategies is established under a certain Liapunov stability condition. For two-person zero-sum games, the existence of a value and optimal strategies for both players are established under the same stability condition.

Key Words. Ergodic occupation measure, stationary strategies, Shapley equation, equilibrium.

1. Introduction

We study noncooperative stochastic games with countable state space, compact action spaces, and with ergodic or limiting average payoff. The existing literature in stochastic games with limiting average payoff seems to be very limited. To the best of our knowledge, notable contributions are due to Gillette (Ref. 1), Sobel (Ref. 2), Bewley and Kohlberg (Ref. 3), Federgruen (Ref. 4), and Mertens and Neyman (Ref. 5). As in Markov decision processes (MDP), the standard approach to stochastic games with limiting average payoff is to treat it as a limiting case of β-discount payoff as the discount factor $\beta \to 1$. For β-discount payoff criterion, much more is known. For two-person zero-sum games with β-discount payoff, the

[1]The authors wish to thank Prof. T. Parthasarathy for pointing out an error in an earlier version of this paper. M. K. Ghosh wishes to thank Prof. A. Arapostathis and Prof. S. I. Marcus for their hospitality and support.
[2]Associate Professor, Department of Electrical Engineering, Indian Institute of Science, Bangalore, India.
[3]Assistant Professor, Department of Mathematics, Indian Institute of Science, Bangalore, India.

existence of stationary β-discount optimal strategies for both players is established in Ref. 6 for Borel state space. For N-person game, the existence of β-discount Nash equilibrium in stationary strategies is established in Ref. 4 for countable state space. For more general state space the problem is much more involved. Parthasarathy and Sinha (Ref. 7) have established the existence of β-discount equilibrium in stationary strategies for Borel state space and finite action spaces, but with a transition law which is independent of the state. Mertens and Parthasarathy (Ref. 8) have proved the existence of subgame perfect equilibrium for β-discount criterion for general state and action spaces.

The limiting average payoff case is drastically different from other cases, because here the finite-time evolution of the processes is irrelevant in some sense; it is only the asymptotic behavior of the time-averaged processes that matters. For the big match, Blackwell and Ferguson (Ref. 9) have established the nonexistence of an optimal strategy for the maximizer. Therefore, it is clear that some conditions on the transition law are necessary to establish the existence of equilibrium for this case. Federgruen (Ref. 4) has established the existence of equilibrium in stationary strategy for this case under a geometric ergodicity condition. He has also given some other recurrence conditions which imply geometric ergodicity. For two-person zero-sum games with this payoff criterion, the existence of a value has been established by Mertens and Neyman (Ref. 5) for finite state and action spaces under no assumptions and for countable state and action spaces under certain assumptions.

Our main objectives in this paper are: (i) for two person zero-sum games, to establish the existence of value and stationary optimal strategies for both players and characterize the same via the Shapley equation for the average payoff criterion; (ii) for N-person games, to establish the existence of equilibrium in stationary strategies.

We achieve our goal by imposing a Liapunov-type stability assumption. We also introduce the notion of almost-sure optimality (for the two-person zero-sum case) and almost-sure equilibrium (for N-person games) by a pathwise treatment of the problems. Pathwise solutions, apart from yielding a mathematically stronger result, are very useful in many practical applications, since we often deal with only one realization; in this case, the expectation may not be appropriate in the payoff function. We first treat the N-person game. Following Borkar (Refs. 10, 11), we introduce certain measure-valued processes and use a technique involving disintegration of measures to show that, if all but one (say the kth) player employ stationary strategies, then player k cannot improve his payoff by going beyond the set of stationary strategies. When all but one player (say k) choose fixed stationary strategies, then the stochastic game problem reduces to a Markov decision problem for player k. Using the results in MDP, several results in stochastic games are often derived using appropriate fixed-point and minimax theorems. The MDP with ergodic payoff criterion has been studied extensively in Refs. 10, 11, 12. We use the results in these references to establish the existence of equilibrium in stationary strategies for N-person games. As a corollary of this, we get the existence of a value and stationary optimal strategies for both players for two-person zero-sum games. Using these existence results, we derive the corresponding Shapley equations and characterize the optimal strategies via these equations. Our previous papers (Refs. 10, 11, 12) thus form the basis for this work. We have tried to make this paper self-contained as far as the essential ideas in stochastic games are concerned, but for technical details we often refer to Ref. 11 and Ref. 12. To give complete details in every step would enormously increase the length of this paper. We structure our paper as follows. Section 2 introduces the notation and describes the problems. We follow the notation of Ref. 13. The notation is unconventional, but has the advantage of economy as far as the present approach is concerned. We introduce the concept of ergodic occupation measures and study some of these properties in Section 3. Section 4 deals with N-person games. Two-person zero-sum games are treated in Section 5.

2. Problem Description

2.1. Two-Person Zero-Sum Stochastic Games. A two-person zero-sum stochastic game is determined by five objects $\langle S, U_1, U_2, p, r\rangle$. Here, $S=\{1, 2, \dots\}$ is the state space. At each stage (time), players I and II observe the current state $i \in S$ of the system and then players I and II independently choose actions $u^1 \in U_1$, $u^2 \in U_2$, respectively, U_1 and U_2 being prescribed compact metric spaces. As a result of this, two things happen:

(i) player I receives an immediate payoff $r(i, u^1, u^2)$ from player II, where $r: S \times U_1 \times U_2 \to \mathbb{R}$ is a given bounded and continuous function;

(ii) the system moves to a new state j with probability $p(i, u^1, u^2, j)$.

The map $p: S \times U_1 \times U_2 \times S \to [0, 1]$ is assumed to be continuous. More generally, U_1 and U_2 can be allowed to depend on i. Then, for each $i \in S$, U_1^i, U_2^i are prescribed compact metric spaces. However, replacing U_1^i by $\prod_1^k U_1^k$, U_2^i by $\prod_k U_2^k$ and $p(i, \dots, j)$ by its composition with the projections $(\prod_k U_1^k, \prod_k U_2^k) \to (U_1^i, U_2^i)$, one may assume that U_1^i, U_2^i are replicas of fixed compact metric spaces U_1, U_2, respectively.

Payoff accumulates throughout the course of the game. Player I wants to maximize his accumulated income, while player II wants to minimize the same. The problem is to choose a strategy for player I that will maximize his total expected income and a strategy for player II that will minimize the same.

For any set Y, let Y^n, $n=1, 2, \dots, \infty$, denote the n-fold Cartesian product of Y with itself. For any Polish space X, $\mathscr{P}(X)$ will denote the spaces of probability measures endowed with the topology of weak convergence.

A pair of strategies (ξ^1, ξ^2) for players I and II is a pair of sequences $(\{\xi_n^1\}, \{\xi_n^2\})$, $\xi_n^l = [\xi_n^l(1), \xi_n^l(2), \dots]$ of U_l^∞-valued random variables, $l=1, 2$, such that, for each $i \in S$, $n \geq 0$,

$$P_{\xi^1,\xi^2}(X_{n+1}=i \mid X_m, \xi_m^1, \xi_m^2, m \leq n) = p(X_n, \xi_n^1(X_n), \xi_n^2(X_n), i),$$

and ξ_n^1, ξ_n^2 are conditionally independent given X_m, $m \leq n$, ξ_m^1, ξ_m^2, $m<n$. We say that $\{X_n\}$ is governed by (ξ^1, ξ^2) whenever the above hold. If in addition ξ_n^l is independent of X_m, $m \leq n$, ξ_m^l, $l=1, 2, m<n$, call $\xi^l = \{\xi_n^l\}$ a Markov strategy for player l. If furthermore ξ_n^l are identically distributed, call $\{\xi_n^l\}$ a stationary strategy for player l. A stationary strategy is called a pure strategy if the law of ξ_n^l is a Dirac measure. Let A_l, M_l, S_l, D_l denote the set of arbitrary, Markov, stationary, and pure deterministic strategies for player l. We shall be interested only in the laws of the sequences

$$\{(X_n, \xi_n^1(X_n), \xi_n^2(X_n)), n \geq 0\}.$$

Therefore, for a stationary strategy $\xi^l = \{\xi_n^l\}$ for player l, the common law of ξ_n^l can be taken to be of the form

$$\Phi^l = \textstyle\prod_k \hat{\Phi}_k^l \in \mathscr{P}_0(U_l^\infty)$$

$$:= \text{the space of product probability measures on } U_l^\infty,$$

$$\hat{\Phi}_k^l \in \mathscr{P}(U_l).$$

Such a stationary strategy of player l will be denoted by $\gamma[\Phi^l]$ and we will often use the notation $\{\xi_n^l\} \sim \gamma[\Phi^l]$. For a pair of stationary strategies $(\gamma[\Phi^1], \gamma[\Phi^2]) \in S_1 \times S_2$, the corresponding process $\{X_n\}$ is a stationary Markov chain with transition probabilities $P[\Phi^1, \Phi^2]$ (i, j) given by

$$P[\Phi^1, \Phi^2](i,j) = \int_{U_2}\int_{U_1} p(i, u_1, u_2, j)\hat{\Phi}_i^1(du_1)\hat{\Phi}_i^2(du_2). \tag{1}$$

Let $P[\Phi^1, \Phi^2]$ denote the infinite stochastic matrix $[[P[\Phi^1, \Phi^2](i,j)]]$. A pair $(\gamma[\Phi^1], \gamma[\Phi^2]) \in S_1 \times S_2$ is called stable if the corresponding chain is positive recurrent and thus has a unique invariant probability measure denoted by $\pi[\Phi^1, \Phi^2] \in \mathscr{P}(S)$. Writing $\pi[\Phi^1, \Phi^2]$ as a row vector, it satisfies

$$\pi[\Phi^1, \Phi^2]P[\Phi^1, \Phi^2] = \pi[\Phi^1, \Phi^2]. \tag{2}$$

Ergodic or Limiting Average Payoff. Let $(\xi^1, \xi^2) \in A_1 \times A_2$ and $\{X_n\}$ be the corresponding process with initial law π. The ergodic payoff is defined as

$$L[\xi^1, \xi^2](\pi) = \liminf_{N\to\infty} (1/N) E_{\xi^1,\xi^2}\left[\sum_{n=0}^{N-1} r(X_n, \xi_n^1(X_n), \xi_n^2(X_n))\right]. \quad (3)$$

We call a strategy $\xi^{*1} \in A_1$ optimal for player I for initial law π if

$$L[\xi^{*1}, \bar{\xi}^2](\pi) \geq \inf_{\xi^2 \in A_2} \sup_{\xi^1 \in A_1} L[\xi^1, \xi^2](\pi), \quad (4)$$

for any $\bar{\xi}^2 \in A_2$; ξ^{*1} is called optimal if it is optimal for any initial law. A strategy $\xi^{*2} \in A_2$ is called optimal for player II for initial law π if

$$L[\bar{\xi}^1, \xi^{*2}](\pi) \leq \sup_{\xi^1 \in A_1} \inf_{\xi^2 \in A_2} L[\xi^1, \xi^2](\pi), \quad (5)$$

for any $\bar{\xi}^1 \in A_1$; ξ^{*2} is called optimal if it is optimal for any initial law. The stochastic game with ergodic payoff has a value if

$$\inf_{\xi^2 \in A_2} \sup_{\xi^1 \in A_1} L[\xi^1, \xi^2](\pi) = \sup_{\xi^1 \in A_1} \inf_{\xi^2 \in A_2} L[\xi^1, \xi^2](\pi), \quad (6)$$

for any initial law π. We will establish the existence of a value and stationary optimal strategies for both players. We also consider pathwise ergodic payoff, i.e., the right-hand side of (3) with E_{ξ^1,ξ^2} deleted. Player I a.s. wants to maximize

$$\liminf_{N\to\infty} \frac{1}{N} \sum_{n=0}^{N-1} r(X_n, \xi_n^1(X_n), \xi_n^2(X_n)), \quad (7)$$

whereas player II a.s. wants to minimize the same. For a stable $(\gamma[\Phi^1], \gamma[\Phi^2]) \in S_1 \times S_2$, (3) and (7) a.s. equal

$$\rho[\Phi^1, \Phi^2] := \sum_{i\in S} \int_{U_2} \int_{U_1} r(i, u_1, u_2) \hat{\Phi}_i^1(du_1) \hat{\Phi}_i^2(du_2) \pi[\Phi^1, \Phi^2](i). \quad (8)$$

In the next section, we will impose some stability conditions [(A1)–(A3)] under which all $(\gamma[\Phi^1], \gamma[\Phi^2]) \in S_1 \times S_2$ are stable. Various other sufficient conditions for stability can be found in Ref. 10. Let

$$\bar{\rho} = \inf_{\gamma[\Phi^2] \in S_2} \sup_{\gamma[\Phi^1] \in S_1} \rho[\Phi^1, \Phi^2], \quad (9)$$

$$\underline{\rho} = \sup_{\gamma[\Phi^1] \in S_1} \inf_{\gamma[\Phi^2] \in S_2} \rho[\Phi^1, \Phi^2]. \quad (10)$$

Under our blanket stability assumption, we will show that $\bar{\rho} = \underline{\rho} = \rho^*$ (say), the stochastic game with ergodic payoff has a value equal to ρ^*, and each player has an optimal stationary strategy for both (3) and (7).

2.2. Noncooperative N-Person Stochastic Game. A noncooperative N-person stochastic game is determined by $2N+2$ objects $\langle S, U_l, r_l, p, l=1,\ldots,N\rangle$. Here, $S=\{1,2,\ldots\}$ is the state space. For each l, U_l is the action space of the lth player, assumed to be a compact metric space. $r_l: S\times U_1\times\cdots\times U_N\to\mathbb{R}$ is the payoff function of the lth player, which is assumed to be bounded and continuous. $p(\cdot,\ldots,\cdot): S\times U_1\times\cdots\times U_n\times S\to[0,1]$ is the transition function which is assumed to be continuous. At each stage, all the players observe the present state of the system $i\in S$ and choose actions independently. If u^l is the action chosen by the lth player and $u=(u^1,\ldots,u^N)\in U_1\times\cdots\times U_N$, then the lth player, $l=1,\ldots,N$, receives a payoff $r_l(i,u)$ and the system moves to a new state j with probability $p(i,u,j)$. An N-tuple of strategies $\xi=(\xi^1,\ldots,\xi^N)$, where for each $l=1,\ldots,N$, $\xi^l=\{\xi^l_n\}$, $\xi^l_n=[\xi^l_n(1),\xi^l_n(2),\ldots]$, $n\geq 0$, is a $\prod_k U_k^\infty$-valued random variable such that

$$P_\xi(X_{n+1}=i\,|\,X_m,\xi^l_m, m\leq n, l=1,\ldots,N)$$
$$=p(X_n,\xi^1_n(X_n),\ldots,\xi^N_n(X_n),i),$$

and ξ^l_n are conditionally independent given X_m, $m\leq n$, ξ^l_m, $m<n$, $l=1,\ldots,N$. Markov, stationary, pure strategies are defined the same way as before. Let A_l, M_l, S_l, D_l denote the set of arbitrary, Markov, stationary, and pure strategies for player l, $l=1,2,\ldots,N$. We will always assume that each S_l is endowed with the pointwise convergence topology and $S_1\times\cdots\times S_N$ with the product topology.

We will use the following notation. For $\xi=[\xi^1,\ldots,\xi^N]\in A_1\times\cdots\times A_N$, $\lambda^l\in A_l$, $\xi^{\hat{l}}=(\xi_1,\ldots,\xi^{l-1},\xi^{l+1},\xi^N)$,

$$(\xi^{\hat{l}},\lambda^l)=(\xi^1,\ldots,\xi^{l-1},\lambda^k,\xi^{l+1},\ldots,\xi^N).$$

We will use similar notation for any other N-tuple.

As before, we will assume that all

$$\gamma[\Phi]=(\gamma[\Phi^1],\ldots,\gamma[\Phi^N])\in S_1\times\cdots\times S_N$$

are stable. For such $\gamma[\Phi]$, let $\pi[\Phi]$ denote the corresponding invariant probability measure. Then, writing $\pi[\Phi]$ as a row vector,

$$\pi[\Phi]P[\Phi]=\pi[\Phi]. \tag{11}$$

For $\xi\in A_1\times\cdots\times A_N$ and initial law $\pi\in\mathscr{P}(S)$, the ergodic payoff to the lth player is

$$L_l[\xi](\pi)=\liminf_{N\to\infty}(1/N)E_\xi\left[\sum_{n=0}^{N-1} r_l(X_n,\xi_n(X_n))\right], \tag{12}$$

where $\{X_n\}$ is governed by ξ with initial law π. An N-tuple $\xi^* = (\xi^{*1}, \ldots, \xi^{*N}) \in A_1 \times \cdots \times A_N$ is called an equilibrium in the sense of Nash if, for each $l = 1, \ldots, N$,

$$L_l[\xi^*](\pi) \geq L_l[\xi^{*\hat{k}}, \xi^k](\pi), \tag{13}$$

for any $\xi^l \in A_l$ and any initial law π. As before, we also consider pathwise ergodic payoff, i.e., the right-hand side of (12) with E_ξ deleted. Under our stability hypothesis for $\gamma[\Phi] \in S_1 \times \cdots \times S_N$, (12) and its pathwise counterpart a.s. equal

$$\rho_l[\Phi] := \sum_{i \in S} \pi[\Phi](i) \int_{U_N} \cdots \int_{U_1} r_l(i, u)\hat{\Phi}_i^1(du_1) \cdots \hat{\Phi}_i^N(du_N), \tag{14}$$

for $u = (u_1, \ldots, u_N)$. We will show that there exists $\gamma[\Phi^*] \in S_1 \times \cdots \times S_N$ such that, for each l,

$$\rho_l[\Phi^*] \geq \rho_l[\Phi^{*\hat{l}}, \Phi^l], \tag{15}$$

for any $\gamma[\Phi^l] \in S_l$.

We note at this point that the existence of a value and optimal strategies for both player in two-person zero-sum game can be deduced from the existence of equilibrium in N-person games. Indeed, take $N = 2$, and $r_1 = -r_2 = r$. Then, if $(\xi^{*1}, \xi^{*2}) \in A_1 \times A_2$ is an equilibrium, it is easily seen that ξ^{*1}, ξ^{*2} are optimal strategies for players I and II, respectively. In this case, (ξ^{*1}, ξ^{*2}) is often referred to as a saddle-point equilibrium. Thus, if the N-person game admits equilibrium, then the desirable results in two person zero-sum games automatically follows. However, in general, to show the existence of equilibrium is more difficult.

3. Ergodic Occupation Measures

Here, we will introduce the concept of ergodic occupation measures and study some related properties. We will carry out our study (both for two person zero-sum and N-person games) under the following stability assumptions.

(A1) For any $(\gamma[\Phi^1], \ldots, \gamma[\Phi^N]) \in S_1 \times \cdots \times S_N$, the corresponding process $\{X_n\}$ has S as a single communicating class.

(A2) For each $i \in S$, there is a finite set $R_i \subset S$ such that $p(i, \cdot, j) = 0$ for $j \notin R_i$.

(A3) Liapunov-Type Stability Assumption. There exists a function $w: S \to \mathbb{R}_+$ such that

(i) $\lim_{i \to \infty} w(i) = \infty$.

(ii) There exist $a>0$, $\epsilon>0$ such that

$$E_\xi[(w(X_{n+1})-w(X_n)+\epsilon)I\{w(X_n)>a\}\,|\,\mathscr{F}_n]\leq 0,$$

for any $\xi\in A_1\times\cdots\times A_N$, $\{X_n\}$ governed by ξ and

$$\mathscr{F}_n=\sigma(X_m, \xi_m^l, m\leq n, l=1,\ldots,N).$$

(iii) There exist a random variable Z and a scalar $\lambda>0$ such that

$$E[\exp(\lambda Z)]<\infty,$$

and for any $c\in\mathbb{R}$,

$$P_\xi(|w(X_{n+1})-w(X_n)|>c)\leq P(Z>c), \qquad n\geq 0,$$

under any $\xi\in A_1\times\cdots\times A_N$.

We will now summarize some consequences of the above assumptions which will be needed later. For proofs and details, see Section 3 in Ref. 7. By (A3)(i), the set $\{i\in S: w(i)\leq a\}$ is finite. Without loss of generality, we may assume it to be of the form $S_M=\{1,\ldots,M\}$. Let $\xi=(\xi^1,\ldots,\xi^N)\in A_1\times\cdots\times A_N$, and let $\{X_n\}$ be governed by ξ. Let

$$\tau_a=\min\{n\geq 0\,|\,X_n\in S_M\}, \qquad \tau'=\min\{n\geq 0\,|\,X_n\notin S_M\},$$

$$\tau(1)=\min\{n\geq 0\,|\,X_n=1\}.$$

Lemma 3.1. Under (A1)–(A3), the following results hold:

(i) $\sup_\xi E_\xi[\tau_a^m\,|\,X_0=i]<\infty$, $i\in S$, $m\geq 1$.

(ii) $\sup_{i\in S_M}\sup_\xi E_\xi[(\tau')^m\,|\,X_0=i]<\infty$, $m\geq 1$.

(iii) All $\gamma[\Phi]=(\gamma[\Phi^1],\ldots,\gamma[\Phi^N])\in S_1\times\cdots\times S_N$ are stable.

(iv) Let $i>M$ and $\gamma[\Phi]$ as above; then,

$$E_\Phi[\tau_a\,|\,X_0=i]\leq w(i)/\epsilon.$$

(v) $\sum_i w(i)\pi[\Phi](i)<\infty$.

(vi) $\lim_{n\to\infty}(1/n)E_\Phi[w(X_n)]=0$.

(vii) $\sup_{\gamma[\Phi]}E_\Phi[(\tau(1))^2\,|\,X_0=i]<\infty$, $\gamma[\Phi]\in S_1\times\cdots\times S_N$, $i\in S$.

(viii) The set $\{\pi[\Phi]\in\mathscr{P}(S)\,|\,\gamma[\Phi]\in S_1\times\cdots\times S_N\}$ is compact.

Let $\gamma[\Phi]=(\gamma[\Phi^1],\ldots,\gamma[\Phi^N])\in S_1\times\cdots\times S_N$, and let $\pi[\Phi]\in\mathscr{P}(S)$ be the invariant measure of the corresponding process $\{X_n\}$. Define the ergodic occupation measure $\nu_E[\Phi]\in\mathscr{P}(S\times U_1\times\cdots\times U_N)$ as follows:

$$\int f\,d\nu_E[\Phi]=\sum_{i\in S}\pi[\Phi](i)\int_{U_1\times\cdots\times U_N} f(i,u_1,\ldots,u_N)\prod_{l=1}^{N}\hat{\Phi}_i^l(du_k), \tag{16}$$

for $f\in C_b(S\times U_1\times\cdots\times U_N)$. In terms of $\nu_E[\Phi]$, (14) becomes

$$\rho_l[\Phi]=\int r_l\,d\nu_E[\Phi]. \tag{17}$$

Let

$$\nu_E[S_1\times\cdots\times S_N]=\{\nu_E[\Phi]\,|\,\gamma[\Phi]\in S_1\times\cdots\times S_N\}. \tag{18}$$

Lemma 3.2. $\nu_E[S_1\times\cdots\times S_N]$ is compact.

Proof. By Lemma 3.1(viii), the set $\{\pi[\Phi]\,|\,\gamma[\Phi]\in S_1\times\cdots\times S_N\}$ is compact. Since $U_1,\ldots,U_N$ are compact metric spaces, it follows that $\nu_E[S_1\times\cdots\times S_N]$ is tight. We will now show that it is closed. Let

$$\gamma[\Phi_n]=(\gamma[\Phi_n^1],\ldots,\gamma[\Phi_n^N])\in S_1\times\cdots\times S_N.$$

For each l, let $\Phi_n^l\to\Phi_\infty^l$ in $\mathscr{P}_0(U_l^\infty)$, i.e., $\Phi_{ni}^l\to\hat{\Phi}_{\infty i}^l$ in $\mathscr{P}(U_l)$ for each $i\in S$. Then, $P[\Phi_n^k]\to P[\Phi_\infty]$ by the continuity of $p(\cdot,\cdot,\cdot)$. Let π be a limit point of $\{\pi[\Phi_n]\}$ in $\mathscr{P}(S)$. Therefore, $\pi[\Phi_n]\to\pi$ along a subsequence of $\{n\}$ in $\mathscr{P}(S)$. Denoting this subsequence by $\{n\}$ again by abuse of notation, $\pi[\Phi_n]\to\pi$ in total variation by Scheffe's theorem (Ref. 14, p. 224). It is easily checked that

$$\pi[\Phi_n]P[\Phi_n]\to\pi P[\Phi_\infty].$$

Since

$$\pi[\Phi_n]P[\Phi_n]=\pi[\Phi_n],\qquad\text{for all } n,$$

we have

$$\pi P[\Phi_\infty]=\pi,$$

i.e., $\pi=\pi[\Phi_\infty]$. Using once again the fact that $\pi[\Phi_n]\to\pi$ in total variation and the fact that $\hat{\Phi}_{ni}^l\to\hat{\Phi}_{\infty i}^l$ in $\mathscr{P}(U_l)$ for each i, we have, for $f\in C_b(S\times U_1\times\cdots\times U_N)$,

$$\int f(i,u_1,\ldots,u_N)\prod_{l=1}^{N}\hat{\Phi}_{ni}^l(du_l)\to\int f(i,u_1,\ldots,u_N)\prod_{l=1}^{N}\hat{\Phi}_{\infty i}^l(du_l),$$

and thus

$$\int f\,d\nu_E[\Phi_n] \to \int f\,d\nu_E[\Phi_\infty].$$

The claim follows. □

From the proof of this lemma, the following result is immediate.

Corollary 3.1. The map $S_1 \times \cdots \times S_N \ni \gamma[\Phi] = (\gamma[\Phi^1], \ldots, \gamma[\Phi^N]) \to \nu_E[\Phi] \in \nu_E[S_1 \times \cdots \times S_N]$ is continuous.

Let $\gamma[\Phi^l] \in S_l$, $l \neq k$, and $\xi^k = \{\xi_n^k\} \in A_k$. Define the $\mathscr{P}(S \times U_1 \times \cdots \times U_N)$-valued empirical process ν_n as follows: for $i \in S$, $B_i \subset U_i$ Borel, $i = 1, \ldots, N$,

$$\nu_n(\{i\} \times B_1 \times \cdots \times B_N) = \frac{1}{n} \sum_{m=0}^{n-1} I\{X_m = i\} \prod_{\substack{l=1 \\ l \neq k}}^{N} \hat{\Phi}_i^l(B_l) \hat{\Phi}_{in}^k(B_k), \tag{19}$$

where $\hat{\Phi}_{in}^k \in \mathscr{P}(U_k)$ is the law of $\xi_n^k(i)$.

Lemma 3.3. $\{\nu_n\}$ is tight and any limit point ν of $\{\nu_n\}$ as $n \to \infty$ a.s. belongs to $\nu_E[S_1 \times \cdots \times S_N]$.

Proof. Define $p' : S \times U_k \times S \to [0, 1]$ as

$$p'(i, u_k, j) = \int_{U_N} \cdots \int_{U_{k-1}} \int_{U_{k+1}} \cdots \int_{U_1} p(i, u_1, \ldots, u_n, j) \prod_{\substack{l=1 \\ l \neq k}}^{N} \hat{\Phi}_i^l(du_l).$$

Let $\{X_n\}$ be governed by $(\gamma[\Phi^{\hat{k}}], \xi^k) \in S_1 \times \cdots \times S_{k-1} \times A_k \times S_{k+1} \times \cdots \times S_N$. Let

$$\tau = \tau(1) = \min\{n > 0 \mid X_n = 1\}.$$

Define the stopping times

$$\tau_0 = 0,$$

$$\tau_{n+1} = \min\{m > \tau_n \mid X_n = 1\}, \qquad n \geq 0.$$

Using Lemma 3.1(vii) and an argument analogous to the one leading to Lemma 5.2 in Ref. 13, it can be shown that

$$\sup_{\xi^k \in A_k} E_{\Phi^{\hat{k}}, \xi^k}[\tau^2 \mid X_0 = 1] < \infty. \tag{20}$$

By (20) and the standard concatenation properties of controlled Markov chains, it follows that

$$\sup_n E_{\Phi^{\hat{k}},\xi^k}[(\tau_{n+1}-\tau_n)^2 | X_0=1] < \infty.$$

In particular, $\tau_n < \infty$ a.s. for all n. Let $\bar{S} = S \cup \{\infty\}$ denote the one-point compactification of S and $\bar{\nu}_n$ the extension of ν_n to $\mathscr{P}(\bar{S} \times U_1 \times \cdots \times U_N)$, obtained by setting $\bar{\nu}_n(\{\infty\} \times U_1 \times \cdots \times U_N) = 0$. Each $\nu \in \mathscr{P}(\bar{S} \times U_1 \times \cdots \times U_N)$ has a decomposition

$$\nu(B) = \delta(\nu)\nu'(B \cap (S \times U_1 \times \cdots \times U_N)) + (1-\delta(\nu))\nu''(B \cap (\{\infty\} \times U_1 \times \cdots \times U_N)),$$

for B Borel in $\bar{S} \times U_1 \times \cdots \times U_N$, where

$$\nu' \in \mathscr{P}(S \times U_1 \times \cdots \times U_N), \qquad \nu'' \in \mathscr{P}(\{\infty\} \times U_1 \times \cdots \times U_N),$$
$$0 \le \delta(\nu) \le 1.$$

This decomposition can be rendered unique by picking a prescribed ν' [resp. ν''] when $\delta(\nu)=0$ [resp. 1]. Since $\mathscr{P}(\bar{S} \times U_1 \times \cdots \times U_N)$ is compact, $\{\bar{\nu}_n\}$ converges to a sample path-dependent compact set in $\mathscr{P}(\bar{S} \times U_1 \times \cdots \times U_N)$. The desired tightness will follow if we show that, outside a set of zero probability, every limit point ν of $\{\bar{\nu}_n\}$ in $\mathscr{P}(\bar{S} \times U_1 \times \cdots \times U_n)$ satisfies $\delta(\nu)=1$. By (20) and the martingale stability theorem (Ref. 15, p. 53), it follows that, for any $M \ge 1$,

$$\lim_{n\to\infty} (1/n) \sum_{i=0}^{n-1} \left[\left(\sum_{m=\tau_i}^{\tau_{i+1}-1} I\{X_m \ge M\} - E_{\Phi^{\hat{k}},\xi^k}\left[\sum_{m=\tau_i}^{\tau_{i+1}} I\{X_m \ge M\} \middle| \mathscr{F}_{\tau_i} \right] \right) \right] = 0, \quad \text{a.s.} \tag{21}$$

Thus,

$$\limsup_{n\to\infty} \nu_n(\{M, M+1, \ldots\} \times U_1 \times \cdots \times U_N)$$
$$= \limsup_{n\to\infty} \frac{1}{n} \sum_{m=0}^{n-1} I\{X_m \ge M\} \le \limsup_{n\to\infty} \frac{1}{n} \sum_{i=0}^{n-1} \left[\sum_{m=\tau_i}^{\tau_{i+1}-1} I\{X_m \ge M\} \right]$$
$$= \limsup_{n\to\infty} \frac{1}{n} \sum_{i=0}^{n-1} E_{\Phi^{\hat{k}},\xi^k}\left[\sum_{m=\tau_i}^{\tau_{i+1}-1} I\{X_m \ge M\} \middle| \mathscr{F}_{\tau_i} \right]$$
$$\le \sup_{\xi^k \in A_k} E_{\Phi^{\hat{k}},\xi^k}\left[\sum_{m=0}^{\tau-1} I\{X_m \ge M\} \middle| X_0 = 1 \right]; \tag{22}$$

the last inequality follows by the concatenation properties of controlled Markov chains. By an argument similar to the one leading to Lemma 5.2 in Ref. 13, the supremum over all A_k in (22) may be replaced by the supremum over S_k. This in turn equals

$$\sup_{\gamma[\Phi^k]\in S_k}\left[E_\Phi[\tau\,|X_0=1]\left(\sum_{i\geq M}\pi[\Phi](i)\right)\right]. \tag{23}$$

In view of Lemma 3.1, the right-hand side of (23) can be made smaller than any prescribed $\epsilon>0$ by choosing M sufficiently large. Let $\{\epsilon_n\}$ be a sequence in (0, 1) decreasing to zero. Consider a sample point for which the above holds for all $\epsilon=\epsilon_n$, $n\geq 1$. Such sample points have probability one, and for them $\{\nu_n\}$ is a tight sequence. This establishes the first claim. Again by the martingale stability theorem, we have

$$(1/n)\sum_{m=0}^{n-1}\left[I\{X_{m+1}=j\}-\sum_{i\in S}I\{X_n=i\}\int_{U_k}p'(i,u_k,j)\hat{\Phi}^k_{in}(du_k)\right]\to 0, \qquad \text{a.s.} \tag{24}$$

Let ν be a limit point of $\{\nu_n\}$ as $n\to\infty$. Disintegrate ν as

$$\begin{aligned}\nu(\{i\}\times A_1\times\cdots\times A_N)&=\tilde{\nu}(\{i\}\times A_k)\prod_{\substack{l=1\\ l\neq k}}^{N}\hat{\Phi}^l_i(A_l)\\ &=\pi(i)\hat{\Phi}^k_i(A_k)\prod_{\substack{l=1\\ l\neq k}}^{N}\hat{\Phi}^l_i(A_l),\end{aligned}$$

where $\tilde{\nu}\in\mathscr{P}(S\times U_k)$ is the image of ν under the projection $S\times U_1\times\cdots\times U_N\to S\times U_k$, $\pi\in\mathscr{P}(S)$ is the image of $\tilde{\nu}$ under the projection $S\times U_k\to S$, and $\hat{\Phi}^k_i\in\mathscr{P}(U_k)$ is any version of the appropriate regular conditional law. Let $\gamma[\Phi^k]\in S_k$ be defined as

$$\Phi^k=\textstyle\prod_{i=1}^{\infty}\hat{\Phi}^k_i\in\mathscr{P}_0(U_k^\infty).$$

Then, by (24),

$$\begin{aligned}\pi(j)&=\sum_{i\in S}\pi(i)\int_{U_k}p'(i,u_k,j)\hat{\Phi}^k_i(du_k)\\ &=\sum_{i\in S}\pi(i)\int_{U_N}\cdots\int_{U_1}p(i,u_1,\ldots,u_N,j)\prod_{l=1}^{N}\hat{\Phi}^l_i(du_l).\end{aligned}$$

Denoting

$$\gamma[\Phi] = (\gamma[\Phi^1], \ldots, \nu[\Phi^N]) \in S_1 \times \cdots \times S_N,$$

we have from the above

$$\pi = \pi[\Phi].$$

Thus, $\nu \in \nu_E[S_1 \times \cdots \times S_N]$. □

4. N-Person Game

Let

$$\gamma[\Phi] = (\gamma[\Phi^1], \ldots, \gamma[\Phi^N]) \in S_1 \times \cdots \times S_N.$$

Let

$$\rho_k^*[\Phi] := \max_{\gamma[\Phi^k] \in S_k} \rho_k[\Phi] = \rho_k[\Phi^{\hat{k}}, \Phi^{*k}]. \tag{25}$$

If players $l, l \neq k$, choose fixed stationary strategies $\gamma[\Phi^l]$, $l \neq k$, then for player k, the game problem reduces to a Markov decision problem with ergodic payoff. Under assumptions (A1)–(A3), $\rho_k^*[\Phi]$ is the optimal ergodic payoff and $\gamma[\Phi^{*k}]$ an optimal strategy for this problem (Refs. 10, 11, 12). $\gamma[\Phi^{*k}]$ will be called an optimal response for player k given $\gamma[\Phi]$. For $\gamma[\Phi]$ as above, define $V_k[\Phi]: S \to \mathbb{R}$ as follows. Let $\gamma[\Phi^{*k}] \in S_k$ be as in (25). Let $\{X_n\}$ be governed by $(\gamma[\Phi^{\hat{k}}], \gamma[\Phi^{*k}])$. Let

$$\tau(1) = \min\{n \geq 1 \mid X_n = 1\}, \tag{26}$$

$$V_k[\Phi](i) = E_{\Phi^{\hat{k}}, \xi^{*k}}\left[\sum_{n=0}^{\tau(1)-1} (r_k(X_n, \xi_n^{\hat{k}}(X_n), \xi_n^{*k}(X_n)) - \rho_k^*[\Phi]) \middle| X_0 = i\right], \tag{27}$$

where

$$\{\xi_n^{*k}\} \sim \gamma[\Phi^{*k}], \qquad \{\xi_n^l\} \sim \gamma[\Phi^l], \qquad l \neq k.$$

Then by Lemma 3.1, $V_k[\Phi]$ is well defined. We say that a function $f: S \to \mathbb{R}$ is $0(w(\cdot)) \times \mathbb{R}$ if

$$\limsup_{i \to \infty} |f(i)|/w(i) < \infty,$$

where $w(\cdot)$ is the Liapunov function described in Section 3.

In view of Lemma 3.3, the following result follows as in Theorem 3.11 and the analogue of Theorem 3.2 stated immediately after the end of the proof of Theorem 3.11 in Ref. 12.

Lemma 4.1. Given $\gamma[\Phi]\in S_1\times\cdots\times S_N$, for each $k\in\{1,\ldots,N\}$, $(V_k[\Phi], \rho_k^*[\Phi])$ is the unique solution in $0(w(\cdot))\times\mathbb{R}$, with $V_k[\Phi](1)=0$, of

$$\rho+V(i)$$

$$=\max_{\phi^k\in\mathscr{P}(U_k)}\left[\int_{U_N}\cdots\int_{U_1}\left[r_k(i,u_1,\ldots,u_N)+\sum_{j\in S}V(j)\times p(i,u_1,\ldots,u_N,j)\right]\right.$$
$$\left.\times\prod_{\substack{l=1\\l\neq k}}^{N}\hat{\Phi}_i^l(du_l)\phi^k(du_k)\right]. \tag{28}$$

A $\gamma[\Phi^{*k}]\in S_k$ is an optimal response for player k, given $\gamma[\Phi]$, if and only if the maximum in the above $[(V_k[\Phi], \rho_k^*[\Phi])$ replacing $(V,\rho)]$ is attained at $\hat{\Phi}_i^{*k}$ for each $i\in S$.

Fix $k\in\{1,\ldots,N\}$ and $\gamma[\Phi]\in S_1\times\cdots\times S_N$. For $\lambda^k\in\mathscr{P}(U_k)$, set

$$L_k(i,\Phi^{\hat{k}},\lambda^k)=\int_{U_N}\cdots\int_{U_1}\left[r_k(i,u_1,\ldots,u_N)+\sum_{j\in S}p(i,u_1,\ldots,u_N,j)\right.$$
$$\left.\times V_k[\Phi](j)\prod_{\substack{l=1\\l\neq k}}^{N}\hat{\Phi}_i^l(du_l)\lambda^k(du_k)\right].$$

For each $i\in S$, define a point-to-set mapping

$$H^k[\Phi]=\left\{\gamma[\Phi^k]\in S_k\,\middle|\,\max_{\xi^k\in\mathscr{P}(U_k)}L_k(i,\Phi^{\hat{k}},\xi^k)=L_k(i,\Phi^{\hat{k}},\hat{\Phi}_i^k),\text{ for each } i\in S\right\}.$$

Then, $H^k[\Phi]$ is easily seen to be nonempty compact and convex. Here, we view S_k as a subset of the topological vector space of all measurable functions $\mu: S\to$ the space of finite signed measures on U_k. Let

$$H[\Phi]=\prod_{k=1}^{N}H^k[\Phi].$$

Then, $H[\Phi]$ is a nonempty compact convex subset of $S_1\times\cdots\times S_N$.

Theorem 4.1. There exists an equilibrium $\gamma[\Phi^*]=(\gamma[\Phi^{*1}],\ldots,\gamma[\Phi^{*N}])\in S_1\times\cdots\times S_N$.

Proof. The map $\gamma[\Phi]=(\gamma[\Phi^1],\ldots,\gamma[\Phi^N])$ to $H[\Phi]$ defines a point-to-set map from $S_1\times\cdots\times S_N$ to $2^{S_1\times\cdots\times S_N}$. We show that this map is upper semicontinuous. Let

$$\begin{aligned}\gamma[\Phi_n]&=(\gamma[\Phi_n^1],\ldots,\gamma[\Phi_n^N])\rightarrow\gamma[\Phi_\infty]\\&=(\gamma[\Phi_\infty^1],\ldots,\gamma[\Phi_\infty^1]),\qquad \text{in } S_1\times\cdots\times S_N.\end{aligned}$$

Fix k and pick

$$\gamma[\tilde{\Phi}_n^k]\in H^k[\Phi_n],\qquad m=1,2,\ldots.$$

By dropping to a subsequence if necessary, we may assume that

$$\gamma[\tilde{\Phi}_n^k]\rightarrow\gamma[\tilde{\Phi}_\infty^k],\qquad \text{for some } \gamma[\tilde{\Phi}_\infty^k]\in S_k.$$

Define

$$\gamma[\tilde{\Phi}_n]=(\gamma[\tilde{\Phi}_n^1],\ldots\,\gamma[\tilde{\Phi}_n^N]),\qquad n=1,\ldots,\infty,$$

by

$$\begin{aligned}\gamma[\tilde{\Phi}_n^l]&=\gamma[\Phi_n^l],\qquad \text{for } l\neq k,\\ \gamma[\tilde{\Phi}_n^l]&=\gamma[\tilde{\Phi}_n^k],\qquad \text{for } l=k.\end{aligned}$$

Then, $\gamma[\tilde{\Phi}_n]\rightarrow\gamma[\tilde{\Phi}_\infty]$. We next show that, for each $i\in S$,

$$V_k[\tilde{\Phi}_n](i)\rightarrow V_k[\tilde{\Phi}_\infty](i).$$

By Lemma 3.1, $V_k[\tilde{\Phi}_n](i)$ is bounded uniformly in n for each $i\in S$. Let $V: S\rightarrow\mathbb{R}$ be such that $V_k[\tilde{\Phi}_n](i)\rightarrow V(i)$ for each i along a subsequence of $\{n\}$. We claim that $V(\cdot)$ is $0(w(\cdot))$. We have that

$$\begin{aligned}V_k[\tilde{\Phi}_n]&=E_{\tilde{\Phi}_n}\left[\sum_{m=0}^{\tau(1)-1}(r_k(X_m,\tilde{\xi}_m(X_m))-\rho_k^*[\tilde{\Phi}_n])\,\middle|\,X_0=i\right],\ \text{where } \{\tilde{\xi}_m^n\}\sim\gamma[\tilde{\Phi}_n],\\&=E_{\tilde{\Phi}_n}\left[\sum_{m=0}^{\tau_a-1}(r_k(X_m,\tilde{\xi}_m^n(X_m))-\rho_k^*[\tilde{\Phi}_n])\,\middle|\,X_0=i\right]\\&\quad+E_{\tilde{\Phi}_n}\left[\sum_{m=\tau a}^{\tau(1)-1}(r_k(X_m,\tilde{\xi}_m(X_m))-\rho_k^*[\tilde{\Phi}_n])\,\middle|\,X_0=i\right]\\&\leq 2KE_{\tilde{\Phi}_n}[\tau_a\,|\,X_0=i]+\sup_{j\in S_M}V_k[\tilde{\Phi}_n](j),\end{aligned}$$

where K is a bound on r_k. Hence, in view of Lemma 3.1, there exist constants C_1, C_2 independent of n, such that

$$|V_k[\tilde{\Phi}_n](i)| \leq C_1 + C_2 w(i).$$

Hence,

$$|V(i)| \leq C_1 + C_2 w(i).$$

Thus, $V(\cdot)$ is $0(w(\cdot))$. Clearly, $V(0)=0$. For each n, $(V_k[\tilde{\Phi}_n], \rho_k^*[\tilde{\Phi}_n])$ satisfies (28) with $\rho = \rho_k^*[\tilde{\Phi}_n]$. Dropping to a further subsequence of $\{n\}$ if necessary, let $\rho_k^*[\tilde{\Phi}_n] \to \tilde{\rho}$ for some $\tilde{\rho} \in \mathbb{R}$. Consider (28) with $(V_k[\tilde{\Phi}_n], \rho_k^*[\tilde{\Phi}_n])$ replacing (V, ρ), and let $n \to \infty$. Passing to the limit [justified by (A2)], we observe that $(V, \tilde{\rho})$ satisfies (28). Hence by Lemma 4.1,

$$V = V_k[\tilde{\Phi}_\infty], \qquad \tilde{\rho} = \rho_k^*[\tilde{\Phi}_\infty].$$

Using this and (A2), it follows that, for each $i \in S$,

$$L_k(i, \tilde{\Phi}_n) \to L_k(i, \tilde{\Phi}_\infty).$$

Now, fix $\gamma[\Phi^k] \in S_k$. Define

$$\gamma[\bar{\Phi}_n] = (\gamma[\bar{\Phi}_n^1], \ldots, \gamma[\bar{\Phi}_n^N]), \qquad n = 1, \ldots, \infty,$$

by

$$\gamma[\bar{\Phi}_n^l] = \gamma[\Phi_n^l], \qquad \text{for } l \neq k,$$
$$\gamma[\bar{\Phi}_m^l] = \gamma[\Phi^k], \qquad \text{for } l = k.$$

Repeat the above arguments for $\gamma[\bar{\Phi}_n]$ in place of $\gamma[\tilde{\Phi}_n]$ for each $i \in S$. Then,

$$L_k(i, \bar{\Phi}_n) \to L_k(i, \bar{\Phi}_\infty).$$

Now, for each $i \in S$,

$$L_k(i, \Phi_n) \leq L_k(i, \tilde{\Phi}_n), \qquad n = 1, 2, \ldots.$$

Thus, for each $i \in S$,

$$L_k(i, \bar{\Phi}_\infty) \leq L_k(i, \tilde{\Phi}_\infty).$$

Hence, $\tilde{\Phi}_\infty^k \in H^k[\Phi_\infty]$. The upper semicontinuity thus follows. By Fan's fixed-point theorem (Ref. 16), there exists

$$\gamma[\Phi^*] = (\gamma[\Phi^{*1}], \ldots, \gamma[\Phi^{*N}]) \in S_1 \times \cdots \times S_N,$$

such that $\gamma[\Phi^*] \in H[\Phi^*]$. This $\gamma[\Phi^*]$ is clearly an equilibrium. □

5. Two-Person Zero-Sum Games

Theorem 5.1. The stochastic game with ergodic payoff criterion admits a value and both players have stationary optimal strategies.

Proof. Take $N=2$, and $r_1=-r_2=r$. By Theorem 4.2, there exists an equilibrium $(\gamma[\Phi^{*1}], \gamma[\Phi^{*2}]) \in S_1 \times S_2$. Clearly,

$$\begin{aligned}\min_{\gamma[\Phi^2]\in S_2} \rho[\Phi^{*1}, \Phi^2] &= \max_{\gamma[\Phi^1]\in S_1} \min_{\gamma[\Phi^2]\in S_2} \rho[\Phi^1, \Phi^2] = \underline{\rho} \\ &= \min_{\gamma[\Phi^2]\in S_2} \max_{\gamma[\Phi^1]\in S_1} \rho[\Phi^1, \Phi^2] = \bar{\rho} \\ &= \max_{\gamma[\Phi^1]\in S_1} \rho[\Phi^1, \Phi^{*2}] := \rho^*. \end{aligned} \tag{29}$$

Obviously, $\gamma[\Phi^{*1}]$, $\gamma[\Phi^{*2}]$ are optimal strategies for players I, II, respectively. □

We now study the Shapley equation for the ergodic payoff criterion. The equation is

$$\begin{aligned}\rho + V(i) &= \min_{\phi^2\in P(U_2)} \max_{\phi^1\in P(U_1)} \left[\int_{U_2}\int_{U_1} \left\{r(i, u_1, u_2) + \sum_{j\in S} p(i, u_1, u_2, j) V(j)\right\} \right. \\ &\qquad \left. \times\, \phi^1(du_2)\phi^2(du_2)\right] \\ &= \max_{\phi^1\in P(U_1)} \min_{\phi^2\in P(U_2)} \left[\int_{U_1}\int_{U_2} \left\{r(i, u_1, u_2) + \sum_{j\in S} p(i, u_1, u_2, j) V(j)\right\} \right. \\ &\qquad \left. \times\, \phi^2(du_2)\phi^1(du_2)\right], \end{aligned} \tag{30}$$

where ρ is a scalar and $V: S \to \mathbb{R}$. A solution (30) is a pair (ρ, V) satisfying it. Let $(\gamma[\Phi^{*1}], \gamma[\Phi^{*2}]) \in S_1 \times S_2$ be a pair of optimal strategies. Let $\{X_n\}$ be the corresponding process. Let

$$\tau(1) = \min\{n \geq 1 \mid X_n = 1\}.$$

Define $V: S \to \mathbb{R}$ as follows:

$$V(i) = E_{\Phi^{*1},\Phi^{*2}}\left[\sum_{n=0}^{\tau(1)-1} (r(X_n, \xi_n^{*1}(X_n), \xi_n^{*2}(X_n)) - \rho^*) \middle| X_0 = i\right], \tag{31}$$

where

$$\{\xi_n^{*k}\} \sim \gamma[\Phi^{*k}], \qquad k=1, 2.$$

Then, by Lemma 3.1, $V(i)$ is well defined.

Theorem 5.2. (V, ρ^*) is the unique solution of (31) in the class $0(w(\cdot)) \times \mathbb{R}$ with $V(1)=0$.

Proof. By the results of Refs. 11 and 12, as noted in the previous section, $V(1)=0$ and V satisfies

$$\rho^* + V(i) = \int_{U_2}\int_{U_1}\left[r(i, u_1, u_2) + \sum_{j\in S} p(i, u_1, u_2, j)V(j)\right] \hat{\Phi}_i^{*1}(du_1)\hat{\Phi}_i^{*2}(du_2). \tag{32}$$

Also from Lemma 3.1 and the strong Markov property, it follows that V belongs to $0(w(\cdot))$. Fix $\gamma[\Phi^{*2}]$. By Lemma 4.1,

$$\rho^* + V(i) = \max_{\phi^1\in\mathscr{P}(U_1)}\left[\int_{U_1}\int_{U_2}[r(i, u_1, u_2) + \sum_{j\in S} p(i, u_1, u_2, j)V(j)]\phi^1(du_1)\hat{\Phi}_i^{*2}(du_2)\right].$$

Hence,

$$\rho^* + V(i) \geq \min_{\phi^2\in\mathscr{P}(U_2)}\max_{\phi^1\in\mathscr{P}(U_1)}\left[\int_{U_1}\int_{U_2}\left[r(i, u_1, u_2) + \sum_{j\in S} p(i, u_1, u_2, j)V(j)\right] \phi^1(du_1)\phi^2(du_2)\right].$$

Similarly, it can be shown that

$$\rho^* + V(i) \leq \max_{\phi^2\in\mathscr{P}(U_2)}\min_{\phi^1\in\mathscr{P}(U_1)}\left[\int_{U_1}\int_{U_2}\left[r(i, u_1, u_2) + \sum_{j\in S} p(i, u_1, u_2, j)V(j)\right] \phi^1(du_1)\phi^2(du_2)\right].$$

Since min max $\geq$ max min, (C, ρ^*) satisfies (30). Let (V', ρ') be another solution of (30) in the class $0(w(\cdot))$ satisfying $V'(1)=0$. Let $(\gamma[\tilde{\Phi}^1], \gamma[\tilde{\Phi}^2]) \in S_1 \times S_2$ be such that

$$\rho' + V'(i) = \int_{U_2}\int_{U_1}\left[r(i, u_1, u_2) + \sum_{j} p(i, u_1, u_2, j)V'(j)\right] \hat{\tilde{\Phi}}_i^1(du_1)\hat{\tilde{\Phi}}_i^1(du_2).$$

Let $\{X_n\}$ be governed by $(\gamma[\tilde{\Phi}^1], \gamma[\tilde{\Phi}^2])$ with $X_0=1$. Then, using (A2) to justify conditional expectations/expectations,

$$E_{\tilde{\Phi}^1,\tilde{\Phi}^2}\left[\sum_{m=1}^{n} (V'(X_{m+1})-E_{\tilde{\Phi}^1,\tilde{\Phi}^2}[V'(X_{m+1})\,|\,\mathscr{F}_m])\right]=0.$$

But

$$\begin{aligned}
&E_{\tilde{\Phi}^1,\tilde{\Phi}^2}[V'(X_{m+1})\,|\,\mathscr{F}_m]\\
&=\sum_j \int_{U_2}\int_{U_1} p(X_m, u_1, u_2, j)V'(j)\hat{\tilde{\Phi}}^1_{X_m}(du_1)\hat{\tilde{\Phi}}^2_{X_m}(du_2)\\
&=\rho'+V'(X_m)-\int_{U_2}\int_{U_1} r(X_m, u_1, u_2)\hat{\tilde{\Phi}}^1_{X_m}(du_1)\hat{\tilde{\Phi}}^2_{X_m}(du_2).
\end{aligned}$$

Summing this over $m=1, \ldots, n$, taking expectation, and dividing by n, we have

$$\begin{aligned}
0=\rho'&-(1/n)E_{\tilde{\Phi}^1,\tilde{\Phi}^2}[V'(X_n)]\\
&-(1/n)E_{\tilde{\Phi}^1,\tilde{\Phi}^2}\left[\sum_{m=1}^{n}\int_{U_2}\int_{U_1} r(X_m, u_1, u_2)\hat{\tilde{\Phi}}^1_{X_m}(du_1)\hat{\tilde{\Phi}}^2_{X_m}(du_2)\right].
\end{aligned}$$

Letting $n\to\infty$ and using Lemma 3.1(vi), we get

$$\rho'=\rho[\tilde{\Phi}^1, \tilde{\Phi}^2].$$

On the other hand, if $(\gamma[\Phi^1], \gamma[\Phi^2])\in S_1\times S_2$, then using similar arguments with $(\gamma[\Phi^1], \gamma[\tilde{\Phi}^2])$ and $(\gamma[\tilde{\Phi}^1], \gamma[\Phi^2])$ would lead to

$$\begin{aligned}
\min_{\gamma[\Phi^2]}\max_{\gamma[\Phi^1]} \rho[\Phi^1, \Phi^2]\le\rho[\Phi^1, \tilde{\Phi}^2]\le\rho'\le\rho[\tilde{\Phi}^1, \Phi^2]&\\
\le\max_{\gamma[\Phi^1]}\min_{\gamma[\Phi^2]} \rho[\Phi^1, \Phi^2].&
\end{aligned}$$

Since min max $\geq$ max min, we have

$$\rho'=\bar{\rho}=\underline{\rho}=\rho^*.$$

Let $(\gamma[\bar{\Phi}^1], \gamma[\bar{\Phi}^2])\in S_1\times S_2$ be such that, for each $i\in S$,

$$\begin{aligned}
&\rho^*+V(i)\\
&=\min_{\phi^2\in\mathscr{P}(U_2)}\left[\int_{U_1}\int_{U_2}\left[r(i, u_1, u_2)+\sum_{j\in S} p(i, u_1, u_2, j)V(j)\right]\phi^2(du_2)\hat{\bar{\Phi}}^1_i(du_1)\right],
\end{aligned}$$

$$\rho^* + V'(i)$$

$$= \max_{\phi^1 \in \mathscr{P}(U_1)} \left[\int_{U_2} \int_{U_1} \left[r(i, u_1, u_2) + \sum_{j \in S} p(i, u_1, u_2, j) V'(j) \right] \hat{\bar{\Phi}}_i^2(du_2) \phi^1(du_1) \right].$$

Let $\{X_n\}$ be governed by $(\gamma[\bar{\Phi}^1], \gamma[\bar{\Phi}^2])$ with the law of $X_0 = \pi[\bar{\Phi}^1, \bar{\Phi}^2]$. Then,

$$\rho^* + V(X_n) \le \int_{U_2} \int_{U_1} \left[r(X_n, u_1, u_2) + \sum_{j \in S} p(X_n, u_1, u_2, j) V'(j) \right]$$

$$\times \bar{\Phi}^1_{X_n}(du_1) \bar{\Phi}^2_{X_n}(du_2)$$

$$= \int_{U_2} \int_{U_1} r(X_n, u_1, u_2) \bar{\Phi}^1_{X_n}(du_1) \bar{\Phi}^2_{X_n}(du_2)$$

$$+ E_{\bar{\Phi}^1, \bar{\Phi}^2}[V(X_{n+1}) \mid \mathscr{F}_n].$$

Similarly,

$$\rho^* + V'(X_n) \ge \int_{U_2} \int_{U_1} r(X_n, u_1, u_2) \bar{\Phi}^1_{X_n}(du_1) \bar{\Phi}^2_{X_n}(du_2)$$

$$+ E_{\bar{\Phi}^1, \bar{\Phi}^2}[V'(X_{n+1}) \mid \mathscr{F}_n].$$

Thus, $V(X_n) - V'(X_n)$, $n \ge 0$, is seen to be a submartingale satisfying [cf. Lemma 3.1(v)]

$$\sup_n E_{\bar{\Phi}^1, \bar{\Phi}^2}[|V(X_n) - V'(X_n)|] \le C \left[\sum_{i \in S} w(i) \pi[\bar{\Phi}^1, \bar{\Phi}^2](i) + 1 \right] < \infty,$$

where C is some constant. By the submartingale convergence theorem, it must converge a.s. Since $\{X_n\}$ is positive recurrent, this is possible only if $V(i) - V'(i)$ is a constant independent of i. Considering $i = 1$, we have $V'(1) - V(1) = 0$. Hence, $V \equiv V'$. □

Theorem 5.3. A $\gamma[\Phi^{*1}] \in S_1$ [resp. $\gamma[\Phi^{*2}] \in S_2$] is optimal for player I [resp. player II] if and only if the outer maximum [resp. outer minimum] in (30), with (V, ρ^*) the unique solution of (30) in the class $0(w(\cdot)) \times \mathbb{R}$, $V(0) = 0$, as described in Theorem 5.2, is attained at Φ_i^{*1} [resp. Φ_i^{*2}] for each $i \in S$.

Proof. The necessity has already been proved. Let $\Phi_i^{*1} \in \mathscr{P}(U_1)$ attain the outer maximum in (30). Let

$$\Phi^{*1} = \prod_i \Phi_i^{*1} \in \mathscr{P}_0(U_1^\infty).$$

Then, $\gamma[\Phi^{*1}]\in S_1$. Pick any $\gamma[\Phi^2]\in S_2$. Let $\{X_n\}$ be governed by $(\gamma[\Phi^{*1}], \gamma[\Phi^2])$. Then, using the same argument as in the proof of the previous theorem, we can show that

$$\rho[\Phi^{*1}, \Phi^2] \geq \rho^*,$$

proving the optimality of $\gamma[\Phi^{*1}]$. The claim for player II can be proved similarly. □

References

1. GILLETTE, D., *Stochastic Games with Zero Stop Probabilities*, Contributions to the Theory of Games, Edited by M. Dresher et al., Princeton University Press, Princeton, New Jersey, pp. 179–188, 1957.
2. SOBEL, M. J., *Noncooperative Stochastic Games*, Annals of Mathematical Statistics, Vol. 42, pp. 1930–1935, 1971.
3. BEWLEY, T., and KOHLBERG, E., *The Asymptotic Theory of Stochastic Games*, Mathematics of Operations Research, Vol. 1, pp. 197–208, 1976.
4. FEDERGRUEN, A., *On N-Person Stochastic Games with Denumerable State Space*, Advances in Applied Probability, Vol. 10, pp. 452–471, 1978.
5. MERTENS, J. F., and NEYMAN, A., *Stochastic Games*, International Journal of Game Theory, Vol. 10, pp. 53–66, 1981.
6. KUMAR, P. R., and SHIAU, T. H., *Existence of Value and Randomized Strategies in Zero-Sum Discrete-Time Stochastic Dynamic Games*, SIAM Journal on Control and Optimization, Vol. 19, pp. 617–634, 1981.
7. PARTHASARATHY, T., and SINHA, S., *Existence of Stationary Equilibrium Strategies in Nonzero-Sum Discounted Games with Uncountable State Space and State Independent Transitions*, International Journal of Game Theory, Vol. 18, pp. 189–194, 1989.
8. MERTENS, J. F., and PARTHASARATHY, T., *Equilibria for Discounted Stochastic Games*, Discussion Paper 8750, Center for Operations Research and Econometrics, Université Catholique de Louvain, 1989.
9. BLACKWELL, D., and FERGUSON, T. S., *The Big Match*, Annals of Mathematical Statistics, Vol. 39, pp. 159–163, 1968.
10. BORKAR, V. S., *Control of Markov Chains with Long-Run Average Cost Criterion*, Proceedings of the IMA Workshop on Stochastic Differential Systems, with Applications to Electrical/Computer Engineering, Control Theory, and Operations Research, Edited by W. Fleming and P. L. Lions, Springer, New York, New York, pp. 57–77, 1987.
11. BORKAR, V. S., *Control of Markov Chains with Long-Run Average Cost Criterion: The Dynamic Programming Equations*, SIAM Journal on Control and Optimization, Vol. 27, pp. 642–657, 1989.
12. BORKAR, V. S., and GHOSH, M. K., *Ergodic and Adaptive Control of Nearest Neighbor Motions*, Mathematics of Control, Systems, and Signals, Vol. 4, pp. 81–98, 1990.

13. BORKAR, V. S., *A Convex Analytic Approach to Markov Decision Processes*, Probability Theory and Related Fields, Vol. 78, pp. 583–602, 1988.
14. BILLINGSLEY, P., *Convergence of Probability Measures*, Wiley, New York, New York, 1968.
15. LOEVE, M., *Probability Theory, Part 1*, 4th Edition, Springer, New York, New York, 1977.
16. FAN, K., *Fixed-Point and Minimax Theorems in Locally Convex Topological Linear Spaces*, Proceedings of the National Academy of Sciences, USA, Vol. 38, pp. 121–126, 1952.