Appl Math Optim 38:283–301 (1998)

Applied Mathematics & Optimization

Stochastic Games with Average Payoff Criterion*

M. K. Ghosh[1] and A. Bagchi[2]

[1]Department of Mathematics, Indian Institute of Science, Bangalore 560012, India

[2]Department of Applied Mathematics, University of Twente, P.O. Box 217, 7500 AE Enschede, The Netherlands

Communicated by A. Bensoussan

Abstract. We study two-person stochastic games on a Polish state and compact action spaces and with average payoff criterion under a certain ergodicity condition. For the zero-sum game we establish the existence of a value and stationary optimal strategies for both players. For the nonzero-sum case the existence of Nash equilibrium in stationary strategies is established under certain separability conditions.

Key Words. Stochastic game, Stationary strategy, Value, Nash equilibrium, Ergodicity

AMS Classification. 90D15, 93E05, 90D25.

1. Introduction

We study noncooperative stochastic games on uncountable state and action spaces, and with ergodic or limiting average payoff. Although there is a vast literature on general state Markov decision processes (MDP) with average payoff criterion, the corresponding results on stochastic games seem to be very limited. For finite or countable state space there are several papers, e.g., [10], [22], [4], [15], [8], and [5]. Uncountable state and action spaces arise quite often in practical problems. When the planning horizon is infinite, two usual payoff criteria that are treated are discounted payoff and limiting

* Part of this research was performed when the first author was visiting the University of Twente, The Netherlands.

average (ergodic) payoff. In many applications, the ergodic payoff criterion is more realistic than the discounted one because the former represents a limiting time average while the latter involves a discount factor which may be difficult to evaluate. From a mathematical point of view, the discounted case is comparatively easier to analyse since the discount factor takes care of the asymptotic behavior of the state process. On the other hand, without some stability property of the state process it is difficult to handle the ergodic case and therefore this case is much more involved. There are several papers on the uncountable state stochastic game with discounted payoff. For example, the two-person zero-sum case has been studied by Maitra and Parthasarathy [14], and by Kumar and Shiau [13]. For the nonzero-sum case, the existence of Nash equilibrium in stationary strategies is a challenging problem and is still open. For finite action spaces Himmelberg et al. [12] and Parthasarathy [20] have established the existence of Nash equilibrium in stationary strategies under a certain separability condition on the reward functions and transition kernel. Parthasarathy and Sinha [21] have obtained the same existence results under the assumption that the transition law is independent of the initial state. Mertens and Parthasarathy [16] have proved the existence of subgame perfect equilibrium for general state and action spaces. Amir [1] has established the existence of equilibrium in Markov strategies for finite action spaces. Recently Nowak and Raghaven [18] have proved the existence of correlated equilibrium in stationary strategies for general state and action spaces under very general conditions In [9] we have studied the stochastic game problem on a Polish state space and compact action space and with average payoff criterion. We have studied both the zero-sum and nonzero-sum cases.

Under certain conditions we have established the existence of saddle point equilibrium for the zero-sum case and Nash equilibrium for the nonzero-sum case in stationary strategies. Nowak [19] has established the existence of correlated equilibrium in stationary strategies for the average payoff criterion for Borel state space and compact action space. He has used a certain geometric ergodicity condition to obtain the equilibrium. Under further separability conditions on the transition law and payoff functions he has also established the existence of Nash equilibrium for the average payoff case. Thus the results for the nonzero-sum case in [9] and [19] are similar, although the methodology and proofs are quite different. In [19] Nowak has employed the vanishing discount method to obtain the equilibrium, i.e., first he has used the existence of equilibrium for the discounted case from [18] and then he has let the discount factor go to one to obtain the corresponding result for the average payoff case. On the contrary, in [9] the optimality equation for the average case is directly used to obtain the Nash equilibrium. To the best of our knowledge the zero-sum case for the average payoff criterion for the uncountable state space has not been treated in the literature before [9].

In this paper which is a revised version of [9], we first study the zero-sum game. Under a certain ergodicity assumption we establish the existence of a value and stationary optimal strategies for both players for the ergodic payoff criterion. We then study the value iteration scheme and develop algorithms for finding optimal strategies for both players. For the nonzero-sum case, we assume the same ergodicity conditions together with separability on reward functions and transition kernel. Under these conditions, we establish the existence of Nash equilibrium in stationary strategies.

Our paper is organized as follows. Section 2 introduces the notation and contains some preliminaries. Section 3 deals with the zero-sum game. Nash equilibrium for the

nonzero-sum game is treated in Section 4. Section 5 concludes the paper with some remarks.

2. Notation and Preliminaries

A two-person stochastic game is determined by six objects $\langle X, U, V, r_1, r_2, q\rangle$, where X is the state space, assumed to be a Polish space. U and V are action spaces of players 1 and 2, respectively, assumed to be compact metric spaces. $r_1\colon X \times U \times V \to \mathbb{R}$, $i = 1, 2$, is the one-stage payoff function for player i, assumed to be bounded and continuous. $q\colon X \times U \times V \to \mathcal{P}(X)$ (the space of probability measures on X endowed with the topology of weak convergence) is the transition law, assumed to be continuous (in the topology of weak convergence). The game is played as follows. At each stage (time) players observe the current state $x \in X$ of the system and then players 1 and 2 independently choose actions $u \in U$, $v \in V$, respectively. As a result of this two things happen:

(i) player i, $i = 1, 2$, receives an immediate payoff $r_i(x, u, v)$,
(ii) the system moves to a new state x' with the distribution $q(\cdot \mid x, u, v)$.

The whole process then repeats from the new state x'. Payoff accumulates throughout the course of the game. The planning horizon or total number of stages is infinite, and each player wants to maximize his time average payoff.

At each stage the players choose their actions independently on the basis of past information. The available information for decision making at time $t \in \mathbb{N} := \{0, 1, 2, \ldots\}$ is given by the history of the process up to that time

$$h_t := (x_0, u_0, v_0, x_1, u_1, v_1, \ldots, u_{t-1}, v_{t-1}, x_t) \in H_t,$$

where $H_0 = X$, $H_t = H_{t-1} \times (U \times V \times X), \ldots, H_\infty = (X \times U \times V)^\infty$ are the history spaces. A strategy for player 1 is a sequence $\pi^1 = \{\pi_t^1\}_{t\in\mathbb{N}}$ of stochastic kernels $\pi_t^1\colon H_t \to \mathcal{P}(U)$. The set of all strategies for player 1 is denoted by Π_1. A strategy $\pi^1 \in \Pi_1$ is called a Markov strategy if

$$\pi_t^1(h_{t-1}, u, v, x)(\cdot) = \pi_t^1(h'_{t-1}, u', v', x)(\cdot)$$

for all $h_{t-1}, h'_{t-1} \in H_{t-1}$, $u, u' \in U$, $v'v \in V$, $x \in X$, $t \in \mathbb{N}$. Thus a Markov strategy for player 1 can be identified with a sequence of measurable maps $\{\Phi_t^1\}$, $\Phi_t^1\colon X \to \mathcal{P}(U)$. A Markov strategy $\{\Phi_t^1\}$ is called a stationary strategy if $\Phi_t^1 = \Phi\colon X \to \mathcal{P}(U)$ for all t. A stationary strategy is called deterministic or pure of $\Phi\colon X \to U$. Let M_1, S_1,and D_1 denote the set of Markov, stationary, and deterministic strategies for player 1, respectively. The strategies for player 2 are defined similarly. Let Π_2, M_2, S_2, and D_2 denote the set of arbitrary, Markov, stationary, and deterministic strategies for player 2, respectively.

Given an initial distribution $\mu \in \mathcal{P}(X)$ and a pair of strategies $(\pi^1, \pi^2) \in \Pi_1 \times \Pi_2$, the corresponding state and action processes $\{X_t\}$, $\{U_t\}$, $\{V_t\}$ are stochastic processes defined on the canonical space $(H_\infty, \mathcal{B}(H_\infty), P_\mu^{\pi_1,\pi_2})$ (where $\mathcal{B}(H_\infty)$ = Borel σ-field on H_∞) via the projections $X_t(h_\infty) = x_t$, $U_t(h_\infty) = u_t$, $V_t(h_\infty) = v_t$, where $P_\mu^{\pi_1,\pi_2}$ is uniquely determined by π^1, π^2, and μ by Ionescu Tulcea's theorem [3]. The corresponding expectation is denoted by $E_\mu^{\pi_1,\pi_2}$. When $\mu = \delta_x$, $x \in X$, we simply write $E_x^{\pi_1,\pi_2}$.

Note that the processes X_t, U_t, V_t will satisfy: for $A \in \mathcal{B}(X)$, $B \in \mathcal{B}(U)$, $C \in \mathcal{B}(V)$, $h_{t-1} \in H_{t-1}$, $h_t \in H_t$, $x \in X$, $u \in U$, $v \in V$

$$P_\mu^{\pi_1,\pi_2}(X_0 \in A) = \mu(A), \tag{1}$$

$$P_\mu^{\pi_1,\pi_2}(U_t \in B | h_t) = \pi_t^1(h_t)(B), \tag{2}$$

$$P_\mu^{\pi_1,\pi_2}(V_t \in C | h_t) = \pi_t^2(h_t)(C), \tag{3}$$

$$P_\mu^{\pi_1,\pi_2}(U_t \in B, V_t \in C | h_t) = \pi_t^1(h_t)(B)\pi_t^2(h_t)(C), \tag{4}$$

and

$$P_\mu^{\pi_1,\pi_2}(X_{t+1} \in A | h_{t-1}, X_t = x, U_t = u, V_t = v) = q(A|x,u,v). \tag{5}$$

Notice that (4) reflects the fact that at each stage the actions are chosen independently and hence the processes U_t and V_t are conditionally independent given the past history.

For a pair of stationary strategies $(\Phi, \Psi) \in S_1 \times S_2$, the corresponding state process $\{X_t\}$ is a Markov process with stationary transition probabilities $P[\Phi, \Psi]$ given by

$$P[\Phi,\Psi](x, dy) = \int_V \int_U q(dy \mid x,u,v)\Phi(x)(du)\Psi(x)(dv). \tag{6}$$

For $\varphi \in \mathcal{P}(U)$, $\psi \in \mathcal{P}(V)$, we use the notation

$$\hat{r}_i(x,\varphi,\psi) = \int_V \int_U r_i(x,u,v)\varphi(du)\psi(dv), \qquad i = 1,2, \tag{7}$$

and

$$\hat{q}(dy \mid x,\varphi,\psi) = \int_V \int_U q(dy \mid x,u,v)\varphi(du)\psi(dv). \tag{8}$$

A pair of stationary strategies $(\Phi, \Psi) \in S_1 \times S_2$ is called stable if the corresponding state process $\{X_i\}$ is ergodic, i.e., it has a unique invariant measure denoted as $\eta[\Phi, \Psi] \in \mathcal{P}(X)$ and

$$\frac{1}{T}\sum_{t=0}^{T-1} \hat{q}^{\,t}(\cdot \mid x, \Phi(x), \Psi(x)) \to \eta[\Phi,\Psi] \tag{9}$$

in $\mathcal{P}(X)$ as $T \to \infty$ for any $x \in X$, where $\hat{q}^{\,t}(\cdot \mid x, \Phi(x), \Psi(x))$ denotes the t-step transition function under (Φ, Ψ).

Let $(\pi^1, \pi^2) \in \Pi_1 \times \Pi_2$ and let (X_t, U_t, V_t) be the corresponding process with $X_0 = x \in X$. The ergodic payoff for player i, $i = 1, 2$, is defined as

$$L_i(\pi^1,\pi^2)(x) = \liminf_{T\to\infty} \frac{1}{T} E_x^{\pi^1,\pi^2}\left[\sum_{t=0}^{T-1} r_i(X_t,U_t,V_t)\right]. \tag{10}$$

A pair of strategies (π^{*1}, π^{*2}) is called a Nash equilibrium (for the ergodic payoff criterion) if

$$\left.\begin{aligned} &L_1(\pi^{*1},\pi^{*2})(x) \geq L_1(\pi^1,\pi^{*2})(x) \qquad \text{for any} \quad \pi^1 \in \Pi_1, \quad x \in X \\ &\text{and} \\ &L_2(\pi^{*1},\pi^{*2})(x) \geq L_2(\pi^{*1},\pi^2)(x) \qquad \text{for any} \quad \pi^2 \in \Pi_2, \quad x \in X. \end{aligned}\right\} \tag{11}$$

Our aim is to establish the existence of a Nash equilibrium $(\Phi^*, \Psi^*) \in S_1 \times S_2$.

The game is called zero-sum iff

$$r_1(x, u, v) + r_2(x, u, v) = 0$$

for any $x \in X$, $U \in U$, $v \in V$. In this case a Nash equilibrium is often referred to as a saddle-point equilibrium. More generally, let $r = r_1 = -r_2$, and $L(\pi^{*1}, \pi^{*2})(x) = L_1(\pi^{*1}, \pi^{*2})(x)$. A strategy $\pi^{*1} \in \Pi_1$ is called optimal for player 1 if, for any $x \in X$,

$$L(\pi^{*1}, \tilde{\pi}^2)(x) \geq \inf_{\pi^2 \in \Pi_2} \sup_{\pi^1 \in \Pi_1} L(\pi^1, \pi^2)(x) \tag{12}$$

for any $\tilde{\pi}^2 \in \Pi_2$. A strategy $\pi^{*2} \in \Pi_2$ is called optimal for player 2 if, for any $x \in X$,

$$L(\tilde{\pi}^1, \pi^{*2})(x) \geq \inf_{\pi^1 \in \Pi_1} \sup_{\pi^2 \in \Pi_2} L(\pi^1, \pi^2)(x) \tag{13}$$

for any $\tilde{\pi}^1 \in \Pi_1$. The game has a value if

$$\inf_{\pi^2 \in \Pi_2} \sup_{\pi^1 \in \Pi_1} L(\pi^1, \pi^2)(x) = \sup_{\pi^1 \in \Pi_1} \inf_{\pi^2 \in \Pi_2} L(\pi^1, \pi^2)(x) \tag{14}$$

for any $x \in X$. Note that if (π^{*1}, π^{*2}) is a saddle-point equilibrium, then π^{*i} is an optimal strategy for player i, $i = 1, 2$, and the game has a value. Our aim is to establish the existence of a value and stationary optimal strategies for each player.

We subsequently assume an ergodic condition under which all $(\Phi, \Psi) \in S_1 \times S_2$ will be stable. We then denote

$$\rho_i(\Phi, \Psi) = \int_X \widehat{r}_i(x, \Phi(x), \Psi(x)) \eta[\Phi, \Psi](dx), \qquad i = 1, 2. \tag{15}$$

Note that under such a condition

$$L_i(\Phi, \Psi)(x) = \rho_i(\Phi, \Psi), \qquad i = 1, 2,$$

for any $x \in X$.

We can also consider a "pathwise" ergodic payoff, i.e., the right-hand side of (10) with $E_x^{\pi_1, \pi_2}$ deleted. Player i wants to a.s. maximize

$$\liminf_{T \to \infty} \frac{1}{T} \sum_{t=0}^{T-1} r_i(X_t, U_t, V_t). \tag{16}$$

The notation of pathwise equilibrium and optimal strategies are analogous. Pathwise solutions, apart from yielding mathematically stronger results, are 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. Under our ergodicity assumption, to be made later, a pair $(\Phi, \Psi) \in S_1 \times S_2$, (16) will be a.s. equal to $\rho_i[\Phi, \Psi]$ and the usual solutions and pathwise solutions will coincide. Therefore we do not distinguish the two payoff concepts.

We denote by $B(X)$ and $C_b(X)$ the Banach spaces of bounded measurable and bounded continuous functions on X respectively with sup norm.

3. Zero-Sum Game

In this section we study the zero-sum case. We carry out our study under the following ergodicity assumption:

(A1)

(i) There exists a number $\alpha < 1$ such that

$$\sup \|\widehat{q}(\cdot \mid x, \varphi, \psi) - \widehat{q}(\cdot \mid x', \varphi', \psi')\|_{\mathrm{TV}} \leq 2\alpha, \tag{17}$$

where the supremum is over all $x, x' \in X$, $\varphi, \varphi' \in \mathcal{P}(U)$, $\psi, \psi' \in \mathcal{P}(V)$ and $\|\cdot\|_{\mathrm{TV}}$ denotes the total variation norm.

(ii) For $x \in X$, $u \in U$, $v \in V$, $q(A \mid x, u, v) > 0$ for any open set $A \subset X$.

Lemma 3.1. *Under* **(A1)**(i), *for any pair of stationary strategies* $(\Phi, \Psi) \in S_1 \times S_2$, *the corresponding state process* $\{X_t\}$ *is ergodic and its invariant measure* $\eta[\Phi, \Psi]$ *satisfies*

$$\|\widehat{q}^{\,t}(\cdot \mid x, \Phi(x), \Psi(x)) - \eta[\Phi, \Psi](\cdot)\|_{\mathrm{TV}} \leq 2\alpha^t. \tag{18}$$

Also,

$$\lim_{t\to\infty} \int f(y)\widehat{q}^{\,t}(dy \mid x, \Phi(x), \Psi(x)) = \int f(y)\eta[\Phi, \Psi](dy) \tag{19}$$

for any $f \in B(X)$.

Proof. We can closely mimic the proof of Lemma 3.3 on p. 57 of [11] to draw the desired conclusions. We omit the details. □

In the next lemma we present a set of sufficient conditions which imply **(A1)**. The proof closely follows that of Lemma 3.3 on p. 57 of [11] and is therefore omitted.

Lemma 3.2. *The following two conditions separately imply* **(A1)**(i):

(i) *There exists a measure* μ *on* X *such that*

$$\mu(X) > 0 \quad \textit{and} \quad q(\cdot \mid x, u, v) \geq \mu(\cdot)$$

for all $x \in X$, $u \in U$, $v \in V$.

(ii) *There exists a measure* ν *on* X *such that*

$$\nu(X) < 2 \quad \textit{and} \quad q(\cdot \mid x, u, v) \leq \nu(\cdot)$$

for all $x \in X$, $u \in U$, $v \in V$.

Example 3.3. Let $(\Omega, \mathcal{F}, P)$ be a probability space and let $\{W_t\}$ be a sequence of independent $N(0, 1)$ ($N(a, b)$ denotes the Gaussian distribution with mean a and variance b) random variables. Let $U, V \subset \mathbb{R}$ be compact sets. Let $f\colon \mathbb{R} \times U \times V \to \mathbb{R}$, $g\colon \mathbb{R} \to \mathbb{R}$ be bounded continuous functions with $g(\cdot) > 0$. Let the state process be given by the following difference equation:

$$X_{t+1} = f(X_t, U_t, V_t) + g(X_t)W_t \tag{20}$$

where U_t, V_t are action processes. Then

$$q(\cdot \mid x, u, v) = N(f(x, u, v), g^2(x)).$$

Since f is bounded we can easily find a nontrivial measure μ (in fact we can choose μ to be equivalent to the Lebesque measure on $\mathbb{R}$) such that

$$q(\cdot \mid x, u, v) \geq \mu(\cdot).$$

In view of Lemma 3.2, condition **(A1)** is satisfied. This example can easily be extended to multidimensional state space. Many pursuitevasion games in discrete time can be modeled after (20).

The existence of a value and optimal strategies are usually derived from the solution of appropriate dynamic programming (or Shapley) equations. For the ergodic payoff criterion the Shapley equations are

$$\begin{aligned}\rho + v(x) &= \min_{\psi \in \mathcal{P}(V)} \max_{\varphi \in \mathcal{P}(U)} \left[\widehat{r}(x, \varphi, \psi) + \int_X v(y) \widehat{q}(dy \mid x, \varphi, \psi) \right] \\ &= \max_{\varphi \in \mathcal{P}(U)} \min_{\psi \in \mathcal{P}(V)} \left[\widehat{r}(x, \varphi, \psi) + \int_X v(y) \widehat{q}(dy \mid x, \varphi, \psi) \right]. \end{aligned} \tag{21}$$

A solution to (21) is a pair (ρ, v) satisfying (21) where ρ is a scalar and $v \in C_b(X)$ (one can choose $v \in B(X)$ also). We have the following standard result.

Theorem 3.4. *Let $(\rho^*, V^*) \in \mathbb{R} \times C_b(X)$ be a solution of* (21). *Then*:

(i) *ρ^* is the value of the game.*

(ii) *Let $(\Phi^*, \Psi^*) \in S_1 \times S_2$ be such that for each $x \in X$*

$$\begin{aligned}\rho^* + v^*(x) &= \min_{\psi \in \mathcal{P}(V)} \left[\widehat{r}(x, \Phi^*(x), \psi) + \int_X v^*(y) \widehat{q}(dy \mid x, \Phi^*(x), \psi) \right] \\ &= \max_{\varphi \in \mathcal{P}(U)} \left[\widehat{r}(x, \varphi, \Psi^*(x)) + \int_X v^*(y) \widehat{q}(dy \mid x, \varphi, \Psi^*(x)) \right], \end{aligned} \tag{22}$$

then Φ^* is an optimal strategy for player 1 *and Ψ^* is an optimal strategy for player* 2. *Under our assumption such Φ^*, Ψ^* always exist.*

(iii) *Let $D(x, \varphi, \psi)$ be defined as*

$$D(x, \varphi, \psi) = \widehat{r}(x, \varphi, \psi) + \int_X v^*(y) \widehat{q}(dy \mid x, \varphi, \psi) - v^*(x) - \rho^*. \tag{23}$$

For any $\pi^1, \pi^2 \in \Pi_1 \times \Pi_2$ and $x \in X$

$$\lim_{T \to \infty} \frac{1}{T} \sum_{t=0}^{T-1} \widehat{r}(X_t, \pi_t^1(h_t), \pi_t^2(h_t)) = \rho^* \qquad P_x^{\pi^1, \pi^2} \text{ a.s.} \tag{24}$$

if and only if

$$\lim_{T \to \infty} \frac{1}{T} \sum_{t=0}^{T-1} D(X_t, \pi_t^1(h_t), \pi_t^2(h_t)) = 0 \qquad P_x^{\pi^1, \pi^2} \text{ a.s.} \tag{25}$$

(iv) *If (π^1, π^2) satisfies* (24) *then it is a pair of optimal strategies for both players.*

Proof. Let $\pi^2 \in \Pi_2$ be arbitrary and let $\Phi^* \in S_1$ be as in (22). The existence of such a Φ^* is guaranteed by a standard measurable selection theorem, e.g. Lemma 1 in [2]. Let $\{X_t\}$ be the corresponding state process with initial condition $X_0 = x$. By (22) we have

$$\rho^* + v^*(X_t) \le \widehat{r}(X_t, \Phi^*(X_t), \pi_t^2(h_t)) + \int_X v^*(y)\widehat{q}(dy \mid X_t, \Phi^*(X_t), \pi_t^2(h_t)).$$

Now,

$$\begin{aligned} E_x^{\Phi^*,\pi^2}[v^*(X_{t+1}) \mid h_t] &= \int_X v^*(y)\widehat{q}(dy \mid X_t, \Phi^*(X_t), \pi^2(h_t)) \\ &\ge \rho^* + v^*(X_t) - \widehat{r}(X_t, \Phi^*(X_t), \pi_t^2(h_t)). \end{aligned}$$

Therefore,

$$\begin{aligned} &\sum_{t=0}^{T-1}[v^*(X_t) - E_x^{\Phi^*,\pi^2}[v^*(X_{t+1}) \mid h_t]] \\ &\qquad \le v^*(X_T) - v^*(x) + \sum_{t=0}^{T-1}\widehat{r}(X_t, \Phi^*(X_t), \pi_t^2(h_t)) - T\rho^*. \end{aligned}$$

Taking expectation, dividing by T, and letting $T \to \infty$, we obtain

$$\rho^* \le L(\Phi^*, \pi^2)(x).$$

Hence

$$\rho^* \le \sup_{\pi^1 \in \Pi_1} \inf_{\pi^2 \in \Pi_2} L(\pi^1, \pi^2)(x)$$

for any $x \in X$. Similarly, it can be shown that

$$\rho^* \ge L(\pi^1, \Psi^*)(x) \qquad \text{for any} \qquad \pi^1 \in \Pi_1, x \in X.$$

Thus

$$\rho^* \ge \inf_{\pi^2 \in \Pi_2} \sup_{\pi^1 \in \Pi_1} L(\pi^1, \pi^2)(x)$$

for any $x \in X$. Therefore it follows that ρ^* is the value of the game, Φ^* is an optimal strategy for player 1, and Ψ^* is an optimal strategy for player 2. This proves (i) and (ii). (iii) can be proved by using similar steps and the martingale stability theorem, see Chapter 3 of [11]. (iv) follows from (iii). □

In view of the above theorem, we look forward to establishing the existence of a solution to (21) in $\mathbb{R} \times C_b(X)$. Indeed, under **(A1)** such a result will be proved. We use the span-contraction method to achieve this. The idea dates back to Tijms [23] and

Federgruen and Tijms [7] in the context of MDP; see Chapter 3 of [11] for an excellent presentation.

Definition 3.5. Let $v \in B(X)$. The span seminorm of v is defined as

$$sp(v) = \sup_{x \in X} v(x) - \inf_{x \in X} v(x).$$

Let $T\colon B(X) \to B(X)$. We say that T is a span-contraction if, for some $\beta \in [0, 1)$,

$$sp(Tu - Tv) \leq \beta sp(u - v)$$

for all $u, v \in B(X)$.

Clearly, $sp(v) = 0$ if and only if $v =$ constant. We introduce the following equivalence relation $\sim$ in $B(X)$. We say that $v_1 \sim v_2$ if and only if $v_1 - v_2 =$ constant. Let $\widetilde{B}(X) = B(X)/\sim$, the quotient space endowed with the quotient norm $\| \ \|_\sim$. If $T\colon B(X) \to B(X)$ is a span-contraction, then it is easily seen that the canonically induced map $\widetilde{T}\colon \widetilde{B}(X) \to \widetilde{B}(X)$ is a contraction and thus has a unique fixed point. It then follows that T itself has a span-fixed point, i.e., there exists a function $v \in B(X)$ such that $sp(Tv - v) = 0$ or equivalently $Tv - v =$ constant, and any two such fixed points must differ by some constant.

Let $v \in C_b(X)$. Define $Tv(x)$ as

$$\begin{aligned} Tv(x) &= \min_{\psi \in \mathcal{P}(V)} \max_{\varphi \in \mathcal{P}(U)} [\widehat{r}(x, \varphi, \psi) + \int_X v(y) \widehat{q}(dy \mid x, \varphi, \psi)] \\ &= \max_{\varphi \in \mathcal{P}(U)} \min_{\psi \in \mathcal{P}(V)} [\widehat{r}(x, \varphi, \psi) + \int_X v(y) \widehat{q}(dy \mid x, \varphi, \psi)]. \end{aligned} \tag{26}$$

Using the (weak) continuity of $g(\cdot \mid \cdot, \cdot, \cdot)$ and the fact that U and V and hence $\mathcal{P}(U)$ and $\mathcal{P}(V)$ are compact metric spaces, it is not difficult to see that $Tv \in C_b(X)$. Thus $T\colon C_b(X) \to C_b(X)$.

Lemma 3.6. *Under* **(A1)**, *T is a span-contraction on $C_b(X)$.*

Proof. Let $v_1, v_2 \in C_b(X)$. Let $\Phi_1^*, \Phi_2^* \in S_1$ and $\Psi_1^*, \Psi_2^* \in S_2$ be such that, for each $x \in X$,

$$\begin{aligned} Tv_1 &= \min_{\psi \in \mathcal{P}(V)} \left[\widehat{r}(x, \Phi_1^*(x), \psi) + \int_X v_1(y) \widehat{q}(dy \mid x, \Phi_1^*(x), \psi) \right] \\ &= \max_{\varphi \in \mathcal{P}(U)} \left[\widehat{r}(x, \varphi, \Psi_1^*(x)) + \int_X v_1(y) \widehat{q}(dy \mid x, \varphi, \Psi_1^*(x)) \right] \end{aligned}$$

and

$$\begin{aligned} Tv_2 &= \min_{\psi \in \mathcal{P}(V)} \left[\widehat{r}(x, \Phi_2^*(x), \psi) + \int_X v_2(y) \widehat{q}(dy \mid x, \Phi_2^*(x), \psi) \right] \\ &= \max_{\varphi \in \mathcal{P}(U)} \left[\widehat{r}(x, \varphi, \Psi_2^*(x)) + \int_X v_2(y) \widehat{q}(dy \mid x, \varphi, \Psi_2^*(x)) \right]. \end{aligned}$$

Then, for any $x, x' \in X$,

$$(Tv_1 - Tv_2)(x) \leq \int_X (v_1 - v_2)(y)\widehat{q}(dy \mid x, \Phi_1^*(x), \Psi_2^*(x)) \tag{27}$$

and

$$(Tv_1 - Tv_2)(x') \geq \int_X (v_1 - v_2)(y)\widehat{q}(dy \mid x', \Phi_2^*(x'), \Psi_1^*(x')). \tag{28}$$

From (27) and (28), it follows that, for any $x, x' \in X$,

$$(Tv_1 - Tv_2)(x) - (Tv_1 - Tv_2)(x') \leq \int_X (Tv_1 - Tv_2)(y)\lambda(dy), \tag{29}$$

where $\lambda(\cdot) = \widehat{q}(\cdot \mid x, \Phi_1^*(x), \Psi_2^*(x)) - \widehat{q}(\cdot \mid x', \Phi_2^*(x'), \Psi_1^*(x'))$.

It can be shown as in Lemma 3.5 p. 59 of [11] that, for any $v \in C_b(X)$ and for any finite signed measure μ,

$$\left| \int v(y)\mu(dy) \right| \leq \tfrac{1}{2} sp(v) \|\mu\|_{\mathrm{TV}}. \tag{30}$$

Applying (30) to (29) and using **(A1)**, it follows that, for any $x, x' \in X$,

$$(Tv_1 - Tv_2)(x) - (Tv_1 - Tv_2)(x') \leq \alpha sp(v_1 - v_2), \tag{31}$$

where $\alpha < 1$ is as in **(A1)**.

Since (31) is true, for all $x, x' \in X$, it follows that

$$sp(Tv_1 - Tv_2) \leq \alpha sp(v_1 - v_2). \qquad \square$$

Theorem 3.7. *Let $x_0 \in X$ be arbitrary. Then under* **(A1)** *there exists a unique solution $(\rho^*, v^*) \in \mathbb{R} \times C_b(X)$ to* (21) *satisfying $v^*(x_0) = 0$.*

Proof. By Lemma 3.6, $T\colon C_b(X) \to C_b(X)$ as defined in (26) is a span-contraction. Therefore there exists a $v \in C_b(X)$ and a constant $\rho^* \in \mathbb{R}$ such that

$$\rho^* + v(x) = Tv(x).$$

Let $v^*(x) = v(x) - v(x_0)$. Then $v^*(x_0) = 0$ and it satisfies

$$\rho^* + v^*(x) = Tv^*(x).$$

Let (ρ', v') be another solution of (21) satisfying $v'(x_0) = 0$. Then clearly v' is also a span-fixed point of T. Hence $v^*(x) - v'(x) =$ constant. Since $v^*(x_0) - v'(x_0) = 0$, it follows that $v^* \equiv v'$. It then easily follows that $\rho' = \rho^*$. $\square$

Based on the above existence results, we now develop the value iteration scheme to obtain uniform approximations to ρ^*. This extends the result of Federgruen and Tijms

[7] to the present problem. Our presentation closely follows Chapter 3 of [11], therefore we omit proofs in several places. Throughout we assume **(A1)**.

We define the value iteration functions $v_t \in C_b(X)$ as follows: $v_0 \in C_b(X)$ is arbitrary, and

$$v_t := T v_{t-1} = T^t v_0, \qquad \text{for} \quad t \geq 1, \tag{32}$$

where T is as defined in (26). It can be easily seen that v_t is the value function for the game with the length of horizon being t and v_0 is the terminal reward. Let $\{\Phi_0\} \in M_1$ and $\{\Psi_t\} \in M_2$ be such that for each $x \in X$

$$\begin{aligned} v_t(x) &= \min_{\psi \in \mathcal{P}(V)} \left[\widehat{r}(x, \Phi_t(x), \psi) + \int_X v_{t-1}(y) \widehat{q}(dy \mid x, \Phi_t(x), \psi) \right] \\ &= \max_{\varphi \in \mathcal{P}(U)} \left[\widehat{r}(x, \varphi, \Psi_t(x)) + \int_X v_{t-1}(dy) \widehat{q}(dy \mid x, \varphi, \Psi_t(x)) \right]. \end{aligned} \tag{33}$$

We say that $\{\Phi_t\}$ and $\{\Psi_t\}$ are value iterations for players 1 and 2, respectively. We will show that these strategies are optimal for the players. We define a sequence of functions ℓ_t in $C_b(X)$ by

$$\begin{aligned} \ell_t(x) &= T^t v_0(x) - T^t v^*(x) \\ &= v_t(x) - v^*(x) - t\rho^*, \end{aligned} \tag{34}$$

where (v^*, ρ^*) is as Theorem 3.7. Then, for each $x \in X$,

$$\begin{aligned} \ell_{t+1}(x) &= \min_{\psi \in \mathcal{P}(V)} \max_{\varphi \in \mathcal{P}(U)} \left[D(x, \varphi, \psi) + \int_X \ell_t(y) \widehat{q}(dy \mid x, \varphi, \psi) \right] \\ &= \max_{\varphi \in \mathcal{P}(U)} \min_{\psi \in \mathcal{P}(V)} \left[D(x, \varphi, \psi) + \int_X \ell_t(y) \widehat{q}(dy \mid x, \varphi, \psi) \right], \end{aligned} \tag{35}$$

where $D(x, \varphi, \psi)$ is as in (23).

Lemma 3.8. *Let $\alpha < 1$ be as in* **(A1)**. *Then*:

(i) $sp(\ell_t) \leq \alpha sp(\ell_{t-1}) \leq \alpha^t sp(\ell_0)$ *for all* $t \geq 0$.
(ii) *Let* $\ell_t^+ = \sup_x \ell_t(x)$, $\ell_t^- = \inf_x \ell_t(x)$. *Then* $\{\ell_t^+\}$ *is nonincreasing and* $\{\ell_t^-\}$ *is nondecreasing.*
(iii) $\sup |\ell_t(x) - c| \leq \alpha^t sp(\ell_0)$ *for all* t *where* $c = \lim_{t\to\infty} \ell_t^+ = \lim_{t\to\infty} \ell_t^-$.
(iv) $\|\ell_t\| \leq \|\ell_{t-1}\| \leq \|\ell_0\|$ *for all* t.

Proof. (i) $sp(\ell_t) = sp(T(T^{t-1}v_0) - T(T^{t-1}v^*)) \leq \alpha sp(T^{t-1}v_0 - T^{t-1}v^*) \leq \alpha^t sp(\ell_0)$.
(ii) Let $(\Phi, \Psi) \in S_1 \times S_2$ such that, for each x,

$$\begin{aligned} \ell_{t+1}(x) &= \max_{\varphi \in \mathcal{P}(U)} \left[D(x, \varphi, \Psi(x)) + \int_X \ell_t(y) \widehat{q}(dy \mid x, \varphi, \Psi(x)) \right] \\ &= \min_{\psi \in \mathcal{P}(V)} \left[D(x, \Phi(x), \psi) + \int_X \ell_t(y) \widehat{q}(dy \mid x, \Phi(x), \psi) \right]. \end{aligned} \tag{36}$$

Let $(\Phi^*, \Psi^*) \in S_1 \times S_2$ be as in (22). Then $D(x, \varphi, \Psi^*(x)) \leq 0$ for any $\varphi \in \mathcal{P}(U)$ and $D(x, \Phi^*(x), \psi) \geq 0$ for any $\psi \in \mathcal{P}(V)$. Thus from (36) it follows that, for each $x \in X$,

$$\ell_{t+1}(x) \leq \int_X \ell_t(y)\widehat{q}(dy \mid x, \Phi^*(x), \Psi^*(x)). \tag{37}$$

Therefore, $\ell^+_{t+1} \leq \ell^+_t$. Similarly, it can be shown that $\ell^-_{t+1} \geq \ell^-_t$.

(iii) This follows from (i) and (ii).

(iv) From (37) it follows that

$$\ell_{t+1}(x) \leq \|\ell_t\|.$$

Similarly it can be shown that $\ell_{t+1}(x) \geq -\|\ell_t\|$. Hence,

$$\|\ell_{t+1}\| \leq \|\ell_t\| \leq \|\ell_0\|. \qquad \square$$

Theorem 3.9. *Let V_t^+ and V_t^- be defined by*

$$V_t^+ = \sup_x w_t(x), \qquad V_t^- = \inf_x w_t(x),$$

where $w_t(x) = v_t(x) - v_{t-1}(x)$ for $t \geq 1$. Then:

(i) *The sequence V_t^+ is nonincreasing and V_t^- is nondecreasing and both converge exponentially fast to ρ^*. More precisely, for all $t \geq 1$*

$$-\alpha^{t-1} sp(\ell_0) \leq V_t^- - \rho^* \leq V_t^+ - \rho^* \leq \alpha^{t-1} sp(\ell_0).$$

(ii) *$V_t^- \leq \rho^* \leq V_t^+$ for each t, and*

$$V_t^- \leq \rho(\Phi', \Psi') \leq V_t^+$$

when $(\Phi', \Psi') \in S_1 \times S_2$ is such that $\Phi' = \Phi_t$, $\Psi' = \Psi_t$ for the fixed but arbitrary t, Φ_t, Ψ_t are as in (33). *Moreover,*

$$|\rho(\Phi', \Psi') - \rho^*| \leq \alpha^{t-1} sp(\ell_0).$$

(iii) $|\rho(\Phi', \Psi') - \rho^*| \leq \sup_x |w_t(x) - \rho^*| \leq \alpha^{t-1} sp(\ell_0)$.

(iv) *For every fixed $z \in X$,*

$$\sup |(v_t(x) - v_t(z)) - (v^*(x) - v^*(z))| \leq 2\alpha^t sp(\ell_0).$$

(v) $\sup_x |D(x, \Phi_t(x), \Psi_t(x))| \leq 2\alpha^{t-1} sp(\ell_0)$.
Thus by Theorem 3.4(*iv*), *$\{\Phi_t\} \in M_1$ is an optimal strategy for player* 1 *and $\{\Psi_t\} \in M_2$ is an optimal strategy for player* 2.

Proof. In view of Lemma 3.8, the proof of Theorem 4.8 on p. 64 of [11] can be closely mimicked to draw the desired conclusions. $\square$

Remark 3.10. From (iii) above it follows that, for large t, $\Phi_t \in S_1$, $\Psi_t \in S_2$ give nearly optimal strategies for both players. This theorem can readily provide algorithms

for finding optimal strategies for both players. This extends the well-known results in MDP to the stochastic game.

Remark 3.11. We can also extend the successive averaging to our case. For $t \geq 1$, set $u_t(x) = v_t(x)/t$. Let $T_t\colon C_b(x) \to C_b(x)$ be defined by

$$T_t v(x) = \min_{\psi \in \mathcal{P}(V)} \max_{\varphi \in \mathcal{P}(U)} \left[t^{-1} \widehat{r}(x, \varphi, \psi) + t^{-1}(t-1) \int_X v(y) \widehat{q}(dy \mid x, \varphi, \psi) \right]$$
$$= \max_{\varphi \in \mathcal{P}(U)} \min_{\psi \in \mathcal{P}(V)} \left[t^{-1} \widehat{r}(x, \varphi, \psi) + t^{-1}(t-1) \int_X v(y) \widehat{q}(dy \mid x, \varphi, \psi) \right]. \quad (38)$$

As in the proof of Lemma 3.6, one can show that $T_t\colon C_b(x) \to C_b(x)$ is a contraction. Thus there exists a unique $u_t^* \in C_b(x)$ such that

$$u_t^*(x) = T_t u_t^*(x). \quad (39)$$

Then the following result can easily be proved.

Corollary 3.12.

(i) $\sup_x |u_t(x) - \rho^*| \to 0$ *as* $t \to \infty$.
(ii) $\|u_t^* - u_t\| \to 0$ *as* $t \to \infty$.
(iii) $\sup_x \|u_t^*(x) - \rho^*\| \to 0$ *as* $t \to \infty$.

4. Nonzero Sum Game

In this section we establish the existence of a pair of stationary equilibrium strategies for a nonzero-sum game. To this end we first strengthen our earlier assumptions.

(A2) $q\colon X \times U \times V \to \mathcal{P}(X)$ is strongly continuous, i.e., continuous in the total variation norm (here $\mathcal{P}(X)$ is regarded as a subset of the space of finite signed measures on X).

(A3) There exists two substochastic kernels

$$q_1\colon X \times U \to \mathcal{P}(X), \qquad q_2\colon X \times V \to \mathcal{P}(X)$$

such that

$$q(\cdot \mid x, u, v) = q_1(\cdot \mid x, u) + q_2(\cdot \mid x, v), \qquad x \in X, \quad u \in U, \quad v \in V.$$

(A4) The reward functions r_i, $i = 1, 2$, are separable in action variables, i.e., there exist bounded continuous functions

$$r_{i1}\colon X \times U \to \mathbb{R}, \qquad r_{i2}\colon X \times V \to \mathbb{R}, \qquad i = 1, 2,$$

such that

$$r_i(x, u, v) = r_{i1}(x, u) + r_{i2}(x, v), \qquad x \in X, \quad u \in U, \quad v \in V.$$

Throughout this section we assume **(A1)** as well. First we give an example where our assumptions (**(A1)**–**(A3)**) are satisfied.

Example 4.1. Let $U, V \subset \mathbb{R}$ be compact. Let

$$\begin{aligned} f_1&: \mathbb{R} \times U \to \mathbb{R} && \text{be bounded continuous,}\\ f_2&: \mathbb{R} \times V \to \mathbb{R} && \text{be bounded continuous,}\\ g_i&: \mathbb{R} \to \mathbb{R}, \quad i = 1, 2, && \text{be bounded continuous,} \end{aligned}$$

and $g_i > 0$. Let $\{W_t^1\}$, $\{W_t^2\}$ be independent $N(0, 1)$ random variables. Assume that, for $i = 1, 2$, Ω_i: $\mathbb{N} \to \mathcal{F}$ ($(\Omega, \mathcal{F}, P)$ is the basic probability space) such that, for each $t \in \mathbb{N}$,

$$\Omega_1(t) \cap \Omega_2(t) = \varphi \quad \text{and} \quad \Omega_1(t) \cup \Omega_2(t) = \Omega.$$

Let $\{X_t\}$ be a real-valued process given by

$$\begin{aligned} X_{t+1} = &(f_1(X_t, u_t) + g_1(X_t)W_t^1)I\{\omega \in \Omega_1(t)\}\\ &+ (f_2(X_t, v_t) + g_2(X_t)W_t^2)I\{\omega \in \Omega_2(t)\}, \end{aligned} \tag{40}$$

$X_0 = x \in \mathbb{R}$, where $u_t \in U$, $v_t \in V$ are actions chosen by the players. We analyse (40) in two cases:

(i) Assume that, for each $t \in \mathbb{N}$, $P(\Omega_1(t)) = \gamma \in [0, 1]$ and $\Omega_i(t)$, $i = 1, 2$, is independent of X_t, u_t, v_t. Then it is easily seen that

$$q(\cdot \mid x, u, v) = \gamma N(f_1(x, u), g_1^2(x)) + (1 - \gamma)N(f_2(x, v), g_2^2(x)).$$

Since f_i, g_i are bounded and continuous, it is easy to see that q satisfies all our assumptions.

(ii) Let $\mathbb{R}_+ = \{x \in \mathbb{R} \mid x \geq 0\}$ and $\mathbb{R}_- = \{x \in \mathbb{R} \mid x < 0\}$. Let $\Omega_1(t) = \{\omega\colon X_t \in \mathbb{R}_+\}$, $\Omega_2(t) = \{\omega\colon X_t \in \mathbb{R}_-\}$. Assume further that $f_1(0, \cdot) = f_2(0, \cdot) =$ constant, and $g_1(0) = g_2(0)$. Then

$$q(\cdot \mid x, u, v) = I\{x \in \mathbb{R}_+\}N(f_1(x, u), g_1^2(x)) + I\{x \in \mathbb{R}_-\}N(f_2(x, v), g_2^2(x)).$$

Clearly, q satisfies all our assumptions. These two cases can be extended to higher dimensions. Also, since f_i, g_i are quite general, many problems of practical interest can be modeled after these. This is particularly true in economic systems. In an economic organization with two decision makers, suppose that only one decision maker is allowed by a higher authority to take action depending on the current economic state of the organization. The evolution of the state will then follow (40) so that assumption **(A3)** is satisfied

The following result, which will be repeatedly used in what follows, is proved in [1].

Lemma 4.2. *Under* **(A2)** *there exists a* $\mu \in \mathcal{P}(X)$ *such that* $q(\cdot \mid x, u, v) \ll \mu$ *for all* $x \in X$, $u \in U$, $v \in V$. *This* μ *will be used throughout. Let* h: $X \times U \times V \times X \to \mathbb{R}_+$ *be the Radon–Nikodym derivative of* $q(\cdot \mid x, u, v)$ *with respect to* μ. *We assume that*

(A5) *For each* $x \in X$, *if* $(u_n, v_n) \to (u, v)$ *in* $U \times V$, *then*

$$\|h(x, u^n, v^n, \cdot) - h(x, u, v, \cdot)\|_{L^1(\mu)} \to 0 \qquad \text{as} \quad n \to \infty.$$

Remark 4.3. If $h(x, \cdot, \cdot, y)$ is continuous for each $x, y \in X$, then **(A5)** is satisfied. In particular, this is satisfied in our example.

Following Himmelberg et al. [12] and Parthasarathy [20] we topologize the spaces S_1 and S_2 with the topology of relaxed controls introduced by Warga [24]. We identify two elements $\Phi_1, \Phi_2 \in S$ if $\Phi_1 = \Phi_2$ a.e. μ (where μ is as in 4.2). Let

$$Y_1 = \{ f\colon X \times U \to \mathbb{R} \mid \begin{array}{l} f \text{ is measurable in the first argument and} \\ \text{continuous in the second, and there exists} \\ g \in L^1(\mu) \text{ such that } |f(x,u)| \le g(x) \text{ for} \\ \text{every } u \in U \}. \end{array}$$

Then Y_1 is a Banach space with norm [24]

$$\| f \| = \int_X \sup_u |f(x,u)| \mu(dx).$$

Every $\Phi \in S_1$ (with the μ-a.e. equivalence relation) can be identified with the element $\Lambda_\Phi \in Y_1^*$ defined as

$$\Lambda_\Phi(f) = \int_X \int_U f(x,u) \Phi(x)(du) \mu(dx).$$

Thus S_1 can be identified with a subset of Y_1^*. Equip S_1 with the weak* topology. Then it can be shown as in [20] that S_1 is compact and metrizable. S_2 is topologized analogously. The following result is immediate.

Lemma 4.4. *Let $\{\Phi_n\} \in S_1$, $\{\Psi_n\} \in S_2$ and $\Phi_n \to \Phi$ in S_1 and $\Psi_n \to \Psi$ in S_2. Let $\ell\colon X \times U \times V \to \mathbb{R}$ be such that $\ell(x,u,v) = \ell_1(x,u) + \ell_2(x,v)$ for $\ell_1\colon X \times U \to \mathbb{R}$ and $\ell_2\colon X \times V \to \mathbb{R}$, each is measurable in the first argument and continuous in the second. Then, for any $h \in L^1(\mu)$,*

$$\begin{aligned} &\int_X h(x) \int_U \int_V \ell(x,u,v) \Phi_n(x)(du) \Psi_n(x)(dv) \mu(dx) \\ &\quad \longrightarrow \int_X h(x) \int_U \int_V \ell(x,u,v) \Phi(x)(du) \Psi(x)(dv) \mu(dx). \end{aligned}$$

Lemma 4.5. *Let $\{v_n\}$ be a sequence of uniformly bounded functions $v_n\colon X \to \mathbb{R}$. Let v be a weak* limit point of v. Let*

$$f_n(x,u,v) = \int_X v_n(y) q(dy \mid x,u,v)$$

and

$$f(x,u,v) = \int_X v(y) q(dy \mid x,u,v), \qquad x \in X, \quad u \in U, \quad v \in V.$$

Then under **(A2)** *and* **(A5)**, *for each* x,

$$\sup_{(u,v)\in U\times V} |f_n(x,u,v) - f(x,u,v)| \to 0 \qquad as \quad n\to\infty.$$

Proof. The pointwise convergence follows from **(A2)** in view of Lemma 4.2. The uniformity of the convergence in u, v can be proved using **(A5)**, see Lemma 3.1 of [17]. □

We now proceed to prove the existence of Nash equilibrium in stationary strategies. Let $(\Phi_1, \Phi_2) \in S_1 \times S_2$. The under **(A1)**, by the results of Chapter 3 of [11] for each x,

$$\sup_{\pi^1\in\Pi_1} L_1(\pi^1, \Phi_2)(x) = \sup_{\pi^1\in S_1} L_1(\pi^1, \Phi_2)(x)$$

and

$$\sup_{\pi^2\in\Pi_2} L_2(\Phi_1, \pi^2)(x) = \sup_{\pi^2\in S_2} L_2(\Phi_1, \pi^2)(x).$$

In fact under **(A1)** the above suprema can be replaced by maxima. Thus there exists $(\Phi_1^*, \Phi_2^*) \in S_1 \in S_2$ such that, for each $x \in X$,

$$\sup_{\pi^1\in\Pi_1} L_1(\pi^1, \Phi_2)(x) = L_1(\Phi_1^*, \Phi_2)(x) = \rho_1(\Phi_1^*, \Phi_2),$$
$$\sup_{\pi^2\in\Pi_2} L_2(\Phi_1, \pi^2)(x) = L_2(\Phi_1, \Phi_2^*)(x) = \rho_2(\Phi_1, \Phi_2^*).$$

(See [3].) Let

$$\rho_1^*(\Phi_2) = \max_{\Phi_1\in S_1} \rho_1(\Phi_1, \Phi_2) \tag{41}$$

and

$$\rho_2^*(\Phi_1) = \max_{\Phi_2\in S_2} \rho_2(\Phi_1, \Phi_2). \tag{42}$$

Let $(\Phi_1^*, \Phi_2^*) \in S_1 \times S_2$ realize the maxima in (41) and (42). Then Φ_1^* (resp. Φ_2^*) is an optimal response of player 1 (resp. player 2) given player 2 (resp. player 1) employs Φ_2 (resp. Φ_1). Fix an $x_0 \in X$. Let $v_1[\Phi_2]$: $X \to \mathbb{R}$ be defined as

$$v_1[\Phi_2](x) = E_x^{\Phi_1^*,\Phi_2}\left[\sum_{t=0}^{\infty}(\widehat{r}_1(X_t, \Phi_1^*(X_t), \Phi_2(X_t)) - \rho_1^*(\Phi_2))\right]. \tag{43}$$

Using **(A1)** it is not difficult to see that $v_1[\Phi_2]$ is well defined and $|v_1[\Phi_2](x)| \le 2k_1/(1-\alpha)$, where k_1 is an upper bound on r_1. Set

$$v_1^*[\Phi_2](x) = v_1[\Phi_2](x) - v_1[\Phi_2](x_0). \tag{44}$$

Then $v_1^*[\Phi_2] \in B(X)$ and $v_1^*[\Phi_2](x_0) = 0$. The following result can be proved using the well-known result in MDP (the techniques of Section 3 can be used to supply the details).

Lemma 4.6. *Given* $\Phi_2 \in S_2$, $(v_1^*[\Phi_2], \rho_1^*(\Phi_2))$ *is the unique solution in* $(B(X) \times \mathbb{R})$ *to*

$$\rho + v(x) = \max_{\varphi \in \mathcal{P}(U)} \left[\widehat{r}_1(x, \varphi, \Phi_2(x)) + \int_X v(y) \widehat{q}(dy \mid x, \varphi, \Phi_2(x)) \right] \tag{45}$$

satisfying $v(x_0) = 0$. *A strategy* $\Phi_1^* \in S_1$ *is an optimal response of player* 1 *given player* 2 *employs* Φ_2 *if and only if* Φ_1^* *realize the pointwise maximum in* (45) (*with* (ρ, v) *replaced by* $(\rho_1^*(\Phi_2), v_1^*[\Phi_2])$.

Remark 4.7. We define $v_2^*[\Phi_1]$ similarly and a result analogous to the above lemma holds for the optimal response of player 2.

Theorem 4.8. *Under* **(A1)–(A5)** *there exists a Nash equilibrium in stationary strategies*.

Proof. Let $(\Phi_1, \Phi_2) \in S_1 \times S_2$. Let

$$\tau \colon S_1 \times S_2 \to 2^{S_1 \times S_2}$$

be defined as

$$\tau(\Phi_1, \Phi_2) = \{(\Phi_1^*, \Phi_2^*) | \rho_1(\Phi_1^*, \Phi_2) = \rho_1^*(\Phi_2), \rho_2(\Phi_1, \Phi_2^*) = \rho_2^*(\Phi_1)\}.$$

We first show that τ is upper semicontinuous. Let $(\Phi_{1n}, \Phi_{2n}) \to (\Phi_{1\infty}, \Phi_{2\infty})$ in $S_1 \times S_2$. Let $(\Phi_{1n}^*, \Phi_{2n}^*) \in \tau(\Phi_{1n}, \Phi_{2n})$ be such that $\Phi_{1n}^* \to \Phi_{1\infty}^*$ and $\Phi_{2n}^* \to \Phi_{2\infty}^*$ in S_2. We need to show that $(\Phi_{1\infty}^*, \Phi_{2\infty}^*) \in \tau(\Phi_{1\infty}, \Phi_{2\infty})$. Let $x_0 \in X$ be fixed. Let

$$\overline{v}_n^1(x) = E_x^{\Phi_{1n}^*, \Phi_{2n}} \left[\sum_{t=0}^{\infty} (\widehat{r}_1(X_t, \Phi_{1n}^*(X_t), \Phi_{2n}(X_t)) - \rho_1^*(\Phi_{2n})) \right]$$

and

$$\overline{v}_n^2(x) = E_x^{\Phi_{1n}, \Phi_{2n}^*} \left[\sum_{t=0}^{\infty} (\widehat{r}_2(X_t, \Phi_{1n}(X_t), \Phi_{2n}^*(X_t)) - \rho_2^*(\Phi_{1n})) \right].$$

Let $v_n^1(x) = \overline{v}_n^1(x) - \overline{v}_n^1(x_0)$, $v_n^2(x) = \overline{v}_n^2(x) - \overline{v}_n^2(x_0)$. Using **(A1)** it can easily be shown that

$$|v_n^i(x)| \leq 4k_i/(1-\alpha),$$

where k_i is an upper bound on r_i and α is as in **(A1)**. Let $v_n^1 \to v_1$ and $v_n^2 \to v_2$ in the weak* sense along a suitable subsequence. Let $\rho_1^*[\Phi_{2n}] \to \rho_1^*$ and $\rho_2^*[\Phi_{1n}] \to \rho_2^*$ along

a subsequence. Then since

$$\rho_1^*(\Phi_{2n}) + v_n^1(x) = \widehat{r}_1(x, \Phi_{1n}^*(x), \Phi_{2n}(x)) + \int_X v_n^1(y)\widehat{q}(dy \mid x, \Phi_{1n}^*(x), \Phi_{2n}(x))$$

and

$$\rho_2^*(\Phi_{1n}) + v_n^2(x) = \widehat{r}_2(x, \Phi_{1n}(x), \Phi_{2n}^*(x)) + \int_X v_n^2(y)\widehat{q}(dy \mid x, \Phi_{1n}(x), \Phi_{2n}^*(x)).$$

Using Lemmas 4.4 and 4.5 it follows that

$$\begin{aligned}
\rho_1^* + v_1(x) &= \widehat{r}_1(x, \Phi_{1\infty}^*(x), \Phi_{2\infty}(x)) \\
&\quad + \int_X v_1(y)\widehat{q}(dy \mid x, \Phi_{1\infty}^*(x), \Phi_{2\infty}(x)) \quad \mu\text{-a.e.}, \\
\rho_2^* + v_2(x) &= \widehat{r}_2(x, \Phi_{1\infty}(x), \Phi_{2\infty}^*(x)) \\
&\quad + \int_X v_2(y)\widehat{q}(dy \mid x, \Phi_{1\infty}(x), \Phi_{2\infty}^*(x)) \quad \mu\text{-a.e.}
\end{aligned}$$

Let $\overline{v}_1(x) = v_1(x) - v_1(x_0)$ and $\overline{v}_2(x) = v_2(x) - v_2(x_0)$. Then using Lemma 4.6 it is seen that

$$\begin{aligned}
&\rho_1^*(\Phi_{2\infty}^*) + \overline{v}_1(x) \\
&\quad = \max_{\varphi \in \mathcal{P}(U)} \left[\widehat{r}_1(x, \varphi, \Phi_{2\infty}^*) + \int_X \overline{v}_1(y)\widehat{q}(dy \mid x, \varphi, \Phi_{2\infty}^*(x)) \right] \quad \mu\text{-a.e.} \\
&\rho_2^*(\Phi_{1\infty}^*) + \overline{v}_2(x) \\
&\quad = \max_{\psi \in \mathcal{P}(V)} \left[\widehat{r}_2(x, \Phi_{1\infty}^*, \psi) + \int_X \overline{v}_2(y)\widehat{q}(dy \mid x, \Phi_{1\infty}^*(x), \psi) \right] \quad \mu\text{-a.e.}
\end{aligned}$$

From this the upper semicontinuity follows. Hence by Fan's fixed point theorem [6] the map τ has a fixed point $(\Phi_1^*, \Phi_2^*) \in S_1 \times S_2$. The pair $(\Phi_1^*, \Phi_2^*) \in S_1 \times S_2$ obviously forms a μ-equilibrium one (i.e., (11) holding in a set of μ-measure 1). Then by a construction analogous to Theorem 1 of [20] the existence of the desired Nash equilibrium follows. □

5. Conclusion

We have established the existence of saddle-point equilibrium in stationary strategies under a geometric ergodicity condition. However for Nash equilibrium we have imposed further separability conditions on rewards and transition kernels. An interesting open problem is to establish a similar existence result without such separability conditions. For Nash equilibrium we have treated the two-person game for notational convenience. The result can easily be extended to an N-person game. Also, our result can be extended to the Borel state and action spaces under some additional assumptions.

References

1. Amir, R. (1991), On stochastic games with uncountable state and action spaces, in Stochastic Games and Related Topics, edited by T.E.S. Raghavan et al., Kluwer, Dordsecht, pp. 149-159.
2. Beneš, V.E. (1970), Existence of optimal strategies based on specified information for a class of stochastic decision problems, SIAM Journal of Control, 8:179–188.
3. Bertsekas, D.P., and Shreve, S.E. (1978), Stochastic Optimal Control: The Discrete Time Case, Academic Press, New York.
4. Bewley, T., and Kohlberg, E. (1976), The asymptotic theory of stochastic games, Mathematics of Operations Research, 1:197–208.
5. Borkar, V.S., and Ghosh, M.K. (1993), Denumerable state stochastic games with limiting average payoff, Journal of Optimization Theory and Applications, 67:539–560.
6. Fan, K. (1952)Fixed point and minimax theorems in locally convex topological linear spaces, Proceedings of the National Academy of Science, U.S.A., 38:121–126.
7. Federgruen, A., and Tijms, H.C. (1978), The optimality equation in average cost denumerable state semi-Markov decision problems, recurrency conditions and algorithms, Journal of Applied Probability, 5:356–373.
8. Federgruen, A. (1978), On N-person stochastic games with denumerable state space, Advances in Applied Probability, 10:452–471.
9. Ghosh, M.K., and Bagchi, A. (1991), Stochastic Games with Average Payoff Criterion, Technical Report No. 985, Department of Applied Mathematics, University of Twente, Enschede.
10. Gilette, D. (1957), Stochastic games with zero stop probabilities, in Contributions to the Theory of Games, edited by M. Dresher et al., Princeton University Press, Princeton, NJ, pp. 179–188.
11. Hernández-Lerma, O. (1989), Adaptive Markov Control Processes, Springer-Verlag, New York.
12. Himmelberg, C., Parthasarathy, T., Raghavan, T.E.S., and Van Vleck, F. (1976), Existence of ρ-equilibrium and optimal stationary strategies in stochastic games, Proceedings of the American Mathematical Society, 60: 245–251.
13. Kumar, P.R., and Shiau, T.H. (1981), Existence of value and randomized strategies in zero-sum discrete-time stochastic dynamic games, SIAM Journal of Control and Optimization, 19:617–634.
14. Maitra, A., and Parthasarathy, T. (1970), On stochastic games, Journal of Optimization Theory and Applications, 5:289–300.
15. Mertens, J.F., and Neyman, A. (1981), Stochastic games, International Journal of Game Theory, 10:53–66.
16. Mertens, J.F., and Parthasarathy, T., Equilibria for discounted stochastic game, CORE Discussion Paper # 8750.
17. Nowak, A.S. (1987), Nonrandomized strategy equilibria in noncooperative stochastic games with additive transition and reward structure, Journal of Optimization Theory and Applications, 52:429–441.
18. Nowak, A.S., and Raghavan, T.E.S. (1992), Existence of stationary correlated equilibria with symmetric information to discounted stochastic games, Mathematics of Operations Research, 17:519–526.
19. Nowak, A.S. (1993), Stationary equilibria for nonzero-sum average payoff ergodic stochastic games with general state space, Annals of the International Society of Dynamic Games, 1:232–246.
20. Parthasarathy, T. (1982), Existence of equilibrium stationary strategies in discounted stochastic games, Shankhya Series A, 44:114–127.
21. Parthasarathy, T., and Sinha, S. (1989), Existence of stationary equilibrium strategies in nonzero sum discounted games with uncountable state space and state independent transitions, International Journal of Game Theory, 18: 189–194.
22. Sobel, M.J. (1971), Noncooperative stochastic games, Annals of Mathematical Statistics 42:1930–1935.
23. Tijms, H.C. (1975), On dynamic programming with arbitrary state space, Compact Action Space and the Average Reward as Criterion, Report BW 55/75, Mathematisch Centrum, Amsterdam.
24. Warga, J. (1967), Functions of relaxed controls, SIAM Journal on Control, 5:628–641.

Accepted 9 *January* 1997