Mathematical Programming 59 (1993) 249-259
North-Holland

A finite step algorithm via a bimatrix game to a single controller non-zero sum stochastic game

A.S. Nowak
Institute of Mathematics, Wrocław Technical University, Wrocław, Poland

T.E.S. Raghavan*
Department of Mathematics, Statistics, and Computer Science, University of Illinois at Chicago, USA

Received 6 March 1989
Revised manuscript received 24 June 1991

Given a non-zero sum discounted stochastic game with finitely many states and actions one can form a bimatrix game whose pure strategies are the pure stationary strategies of the players and whose penalty payoffs consist of the total discounted costs over all states at any pure stationary pair. It is shown that any Nash equilibrium point of this bimatrix game can be used to find a Nash equilibrium point of the stochastic game whenever the law of motion is controlled by one player. The theorem is extended to undiscounted stochastic games with irreducible transitions when the law of motion is controlled by one player. Examples are worked out to illustrate the algorithm proposed.

Key words: Stochastic game theory.

1. Introduction

Stochastic games were formulated by Shapley [12] in 1953. In this seminal paper, he proved that zero-sum discounted stochastic games have a value in the class of stationary strategies of the players. Gillette [6] introduced zero-sum stochastic games with the long run average cost per play for a player.

Shapley's stochastic games were extended to non-zero sum games by many researchers [5, 14, 13]. For discounted games they proved the existence of Nash equilibrium points in stationary strategies. From an algorithmic point of view, it is desirable to look for subclasses of stochastic games that admit stationary optimal strategies or stationary Nash equilibria. If the algorithm has to terminate in a finite number of arithmetic steps, it is desirable to look for stochastic games which have the added property that the smallest field containing the data entries defining the stochastic game also contains the entries of some solution to the stochastic game.

Correspondence to: Dr. Andrzej Nowak, Instytut Matematyki, Politechniki Wrocławskieg, Wybrzeze Wyspianskiego 27, 50-370 Wrocław, Poland.
* The work of this author was supported in part by the NSF grants DMS-9024408 and DMS 8802260.

Only some special classes of stochastic games have this additional property. For further details on this topic see [2, 9, 10].

Among many such subclasses of stochastic games, single controller stochastic games have many desirable properties. They possess stationary optimal strategies in both discounted and undiscounted zero-sum case. They possess the above mentioned *ordered field property* for both zero-sum and non-zero sum case if we take as solutions Nash equilibria in stationary strategies [9, 3].

Single player control zero-sum games can be solved by a single linear program (see [9] for discounted games, and [7, 15] for undiscounted games).

In this paper, we prove two theorems (see Theorems 3.1 and 3.2) that can be used to construct *finite step algorithms* for Nash equilibrium points in both discounted and irreducible undiscounted non-zero sum stochastic games. The theorems sharpen an earlier result of Filar and Raghavan [4] proved in the context of zero-sum undiscounted single-controller games. Incidentally, *these are the first algorithms* (to the best of our knowledge) which directly rely on the Lemke–Howson algorithm to solve for stationary equilibria in stochastic games.

2. Definitions and preliminaries

Let $S=\{1,\dots,N\}$ be the set of states of a multistage game. Let $I=\{1,\dots,k\}$ and $J=\{1,\dots,p\}$ be finite sets of actions available to players 1 and 2, respectively. The game initiates at a state, say s_1, on the first day. Knowing the state s_1, players 1 and 2 secretly select actions $i_1\in I$ and $j_1\in J$, respectively. This results in an immediate cost, given by $r_1(s_1, i_1, j_1)$ for player 1 and $r_2(s_1, i_1, j_1)$ for player 2. The game moves to a new state s_2, the next day. The players know that this movement is Markovian and it depends only on the state just passed and the action just selected. The players, knowing the new state s_2, choose actions $i_2\in I$, $j_2\in J$ and incur immediate costs $r_1(s_2, i_2, j_2)$ and $r_2(s_2, i_2, j_2)$ and so on.

Let $\hat{q}(s_n | s_1, i_1, j_1, \dots, s_{n-1}, i_{n-1}, j_{n-1}) = q(s_n | s_{n-1}, i_{n-1}, j_{n-1})$, $n=2,3,\dots$. If $h_n=(s_1, i_1, j_1, \dots, s_{n-1}, i_{n-1}, j_{n-1})$, the history on the nth day, in general a player, say player 1, can select an $i_n\in I$ by a random mechanism which could depend on the partial history h_n. Such a strategy is called a behavior strategy for player 1. In case a strategy depends only on the current state, then such a strategy can be denoted by $f(s)=(f_1(s),\dots,f_k(s))$, $s\in S$, (where $f_i(s)$ is the probability that action i is selected in state s by player 1) and is called a *stationary strategy* for player 1. The idea is that player 1 ignores all past information such as the states and actions and dates, but retains only the information about the state he is currently in. A stationary strategy $g(s)=(g_1(s),\dots,g_p(s))$, $s\in S$, is similarly defined for player 2. The simplest among stationary strategies are the *pure stationary strategies* where, say player 1, chooses the same action i_s whenever state $s\in S$ is reached. We can identify any such pure stationary strategy f with an N-tuple $f=(i(1),\dots,i(N))$, where $i(s)\in I$ for each $s\in S$. We can similarly define pure stationary strategy for player 2.

Once we specify the initial state and any strategy pair (f, g) for the players, we have a probability distribution over all sequences of states and actions. Let $r_k^{(n)}(f, g)(s)$ be the expected cost to player k on the nth day, when the game starts at state s and f, g are the strategy choices of the players.

A *β-discounted cost (penalty)* for the infinite horizon game is given by

$$\Phi_k^\beta(f, g)(s) = \sum_{n=1}^{\infty} \beta^{n-1} r_k^{(n)}(f, g)(s),$$

where $k = 1, 2$, and $s = s_1$.

An *undiscounted cost* for the infinite horizon game is given by

$$\Phi_k(f, g)(s) = \limsup_T \frac{1}{T} \sum_{n=1}^{T} r_k^{(n)}(f, g)(s),$$

where $k = 1, 2$, and $s = s_1$.

We call a stationary strategy pair a *Nash equilibrium point* to the β-discounted stochastic game iff for each $s \in S$,

$$\Phi_1^\beta(f^*, g^*)(s) \leq \Phi_1^\beta(\pi, g^*)(s),$$

for any behavior strategy π of player 1 and

$$\Phi_2^\beta(f^*, g^*)(s) \leq \Phi_2^\beta(f^*, \sigma)(s),$$

for any behavior strategy σ of player 2.

Nash equilibria for undiscounted games are similarly defined.

We call a stochastic game a *player* 2 *control stochastic game* iff the transition probabilities depend on the current state and current actions of player 2 only, that is,

$$q(t \mid s, i, j) = q(t \mid s, j) \quad \text{for all } s, t, i, j.$$

3. Main results

We quote the following fact which we need in the sequel (see [9, Lemma 5.1]):

Lemma. *Any pair (f^*, g^*) of stationary strategies is a Nash equilibrium to a discounted player* 2 *control stochastic game iff it is a Nash equilibrium to a stochastic game where player* 1 *has no immediate cost beyond the first day.* □

Now we are ready to prove the first main theorem which gives a recipe for finding Nash equilibrium points for player 2 control discounted stochastic games in a finite number of arithmetic steps.

Theorem 3.1. *Let $0<\beta<1$ be a discount factor. Let N be the number of states of the game. Let $f_1, f_2, \ldots, f_m$ be an enumeration of all pure stationary strategies for player 1. Let $g_1, g_2, \ldots, g_n$ be an enumeration of all pure stationary strategies for player 2. Let (A, B) be an $m \times n$ bimatrix game with entries*

$$A=\left[\sum_s r_1(s, f_i(s), g_j(s))\right],$$

$$B=\left[\sum_s \Phi_2^\beta(f_i, g_j)(s)\right],$$

where $i=1, \ldots, m$, $j=1, \ldots, n$. We will view A, B as cost matrices to players 1 and 2. Let (ξ^*, η^*) be any mixed strategy Nash equilibrium point to the bimatrix game (A, B). Then

(a) $$f^*=\sum_i \xi_i^* f_i \quad \text{and} \quad g^*=\sum_j \eta_j^* g_j$$

constitute a Nash equilibrium point to the β-discounted stochastic game.

(b) *For each state s,*

$$\Phi_2^\beta(f^*, g^*)(s)=\sum_i \sum_j \Phi_2^\beta(f_i, g_j)(s)\xi_i^* \eta_j^*.$$

Proof. Since (ξ^*, η^*) is a Nash equilibrium point to (A, B), we have

$$\sum_s r_1(s, f^*(s), g^*(s)) \leq \sum_s r_1(s, f(s), g^*(s)), \tag{1}$$

for every pure stationary f. Clearly,

$$\min_f r_1(s, f(s), g^*(s)) \leq r_1(s, f^*(s), g^*(s)), \tag{2}$$

for every $s \in S$. Hence,

$$\sum_s \min_f r_1(s, f(s), g^*(s)) \leq \sum_s r_1(s, f^*(s), g^*(s)). \tag{3}$$

At the same time

$$\sum_s \min_f r_1(s, f(s), g^*(s)) = \min_{\{f(1), \ldots, f(N)\}} \sum_{s=1}^N r_1(s, f(s), g^*(s))$$
$$= \min_f \sum_s r_1(s, f(s), g^*(s)). \tag{4}$$

From (4), (3) and (1), we get

$$\sum_s \min_f r_1(s, f(s), g^*(s)) = \sum_s r_1(s, f^*(s), g^*(s)),$$

which, together with (2), gives

$$r_1(s, f^*(s), g^*(s)) = \min_f r_1(s, f(s), g^*(s)),$$

for every $s \in S$. This and the above lemma imply that

$$\Phi_1^\beta(f^*, g^*)(s) \leq \Phi_1^\beta(f, g^*)(s), \tag{5}$$

for every $s \in S$ and for any pure stationary strategy f of player 1. Thus, (5) holds for any strategy f of player 1 (see [1] or [12]).

Consider the equilibrium inequalities for player 2 in (A, B). We have

$$\sum_s \sum_i \sum_j \Phi_2^\beta(f_i, g_j)(s) \xi_i^* \eta_j^* \leq \sum_s \sum_i \Phi_2^\beta(f_i, g)(s) \xi_i^*,$$

for every pure stationary strategy g of player 2. Thus,

$$\sum_s \sum_j \Phi_2^\beta(f^*, g_j)(s) \eta_j^* \leq \sum_s \Phi_2^\beta(f^*, g)(s), \tag{6}$$

for every pure stationary strategy g of player 2. (Here we use the fact that the transition probability q is independent of player 1's actions.)

We claim that the stronger inequality

$$\sum_j \Phi_2^\beta(f^*, g_j)(s) \eta_j^* \leq \Phi_2^\beta(f^*, g)(s)$$

holds, for every pure stationary strategy g of player 2 and for all states $s \in S$. Suppose not. Then, for some state s' and some pure stationary strategy g', we have

$$\sum_j \Phi_2^\beta(f^*, g_j)(s') \eta_j^* > \Phi_2^\beta(f^*, g')(s'). \tag{7}$$

Let g_0 be a pure stationary strategy for player 2, which is optimal against f^* [1]. That is,

$$\Phi_2^\beta(f^*, g_0)(s) = \min_g \Phi_2^\beta(f^*, g)(s), \tag{8}$$

for every $s \in S$. From (7) and (8), we conclude that

$$\sum_s \sum_j \Phi_2^\beta(f^*, g_j)(s) \eta_j^* > \sum_s \Phi_2^\beta(f^*, g_0)(s), \quad s \in S,$$

which contradicts (6). It is therefore obvious that

$$v_2(s) := \min_g \Phi_2^\beta(f^*, g)(s) = \sum_j \Phi_2^\beta(f^*, g_j)(s) \eta_j^*,$$

for all $s \in S$. Moreover, if $\eta_k^* > 0$, then

$$v_2(s) = \Phi_2^\beta(f^*, g_k)(s) \quad \text{for all } s \in S.$$

From the optimality equation for discounted dynamic programming [1], we have

$$v_2(s) = r_2(s, f^*(s), g_k(s)) + \beta \sum_t v_2(t) q(t \mid s, g_k(s)),$$

for all s and for each k such that $\eta_k^* > 0$. Hence, if we put $g^* = \sum_j \eta_j^* g_j$, we get

$$v_2(s) = r_2(s, f^*(s), g^*(s)) + \beta \sum_t v_2(t) q(t \mid s, g^*(s)),$$

for all $s \in S$. This implies that

$$v_2(s) = \Phi_2^{\beta}(f^*, g^*)(s) \quad \text{for all } s \in S,$$

and the proof is completed. □

Our second main result concerns *irreducible undiscounted stochastic games*, where we assume that the transition matrix induced by any pure stationary strategy for player 2 is irreducible. It is well known that the expected costs $\Phi_k(f, g)(s)$ to the players in an irreducible undiscounted stochastic game, under any stationary strategy pair (f, g), are independent of the initial state s, and thus, they will be denoted by $\Phi_k(f, g)$.

Theorem 3.2. *Let $f_1, f_2, \ldots, f_m$ be an enumeration of all pure stationary strategies for player 1. Let $g_1, g_2, \ldots, g_n$ be an enumeration of all pure stationary strategies for player 2. Let (A, B) be an $m \times n$ bimatrix game with entries*

$$A = \left[\sum_s r_1(s, f_i(s), g_j(s)) \right],$$

$$B = [\Phi_2(f_i, g_j)],$$

where $i = 1, \ldots, m$, $j = 1, \ldots, n$. Let (ξ^*, η^*) be any mixed strategy Nash equilibrium point to the bimatrix game (A, B). Then

$$f^* = \sum_i \xi_i^* f_i \quad \text{and} \quad g^* = \sum_j \eta_j^* g_j$$

constitute a Nash equilibrium point to the irreducible undiscounted stochastic game.

Proof. Repeating verbatim the arguments of the discounted case we have

$$r_1(s, f^*(s), g^*(s)) \leq r_1(s, f_i(s), g^*(s)), \tag{9}$$

for all s, i. Hence, we have

$$\sum_t q(t \mid s, g^*(s)) r_1(t, f^*(t), g^*(t)) \leq \sum_t q(t \mid s, g^*(s)) r_1(t, f_i(t), g^*(t)), \tag{10}$$

for all s and i. Now using (9) and (10), since transitions are controlled only by player 2, we conclude that

$$\Phi_1(f^*, g^*) \leq \Phi_1(f_i, g^*) \quad \text{for all } i,$$

which together with [1] implies that (f^*, g^*) satisfies the equilibrium condition for player 1.

For player 2, we obtain

$$\sum_i \sum_j \Phi_2(f_i, g_j) \xi_i^* \eta_j^* = \sum_j \Phi_2(f^*, g_j) \eta_j^* \leq \Phi_2(f^*, g), \tag{11}$$

for each pure stationary strategy g of player 2. From (11), we conclude that every g_j such that $\eta_j^* > 0$ is optimal for player 2 against f^*, that is,

$$\Phi_2(f^*, g_j) = \min_g \Phi_2(f^*, g).$$

Let $v = \min_g \Phi_2(f^*, g)$. From the dynamic programming literature (see, e.g., [11]), we infer that there exists a function $w: S \to \mathbb{R}$ such that

$$\begin{aligned} v + w(s) &= \min_{j \in J} \left[r_2(s, f^*(s), j) + \sum_t w(t) q(t \mid s, j) \right] \\ &= r_2(s, f^*(s), g_k(s)) + \sum_t w(t) q(t \mid s, g_k(s)) \end{aligned} \tag{12}$$

for every k such that $\eta_k^* > 0$ and all $s \in S$. It follows that

$$v + w(s) = r_2(s, f^*(s), g^*(s)) + \sum_t w(t) q(t \mid s, g^*(s)),$$

where $s \in S$ and $g^* = \sum_k \eta_k^* g_k$. This and (12) imply (see [11]) that

$$\Phi_2(f^*, g^*) = \min_g \Phi_2(f^*, g),$$

which completes the proof. □

4. Examples

We will use the theorems proved above to constructively solve for Nash equilibria for discounted and also irreducible undiscounted player 2 control games.

Example 4.1. Consider the following stochastic game with 3 states and two actions for each player at each state. The transitions which consist of an exact law of motion are indicated by an arrow entry underneath each column that corresponds to action of player 2. Player 1 chooses rows. Let the discount factor $\beta = 0.8$.

$$\begin{array}{ccc} s=1 & s=2 & s=3 \\ \begin{bmatrix} (6,3) & (0,8) \\ (0,5) & (7,1) \end{bmatrix}, & \begin{bmatrix} (0,10) & (9,2) \\ (7,5) & (0,8) \end{bmatrix}, & \begin{bmatrix} (3,0) & (0,5) \\ (0,4) & (4,0) \end{bmatrix}. \\ \begin{matrix} \downarrow & \quad \downarrow \\ 1 & \quad 2 \end{matrix} & \begin{matrix} \downarrow & \quad \downarrow \\ 2 & \quad 3 \end{matrix} & \begin{matrix} \downarrow & \quad \downarrow \\ 3 & \quad 1 \end{matrix} \end{array}$$

We can enumerate pure stationary strategies by $f_1 = (111)$, $f_2 = (112), \ldots, f_8 = (222)$. For example $f_6 = (212)$ for player 1 selects row 2 in state 1, row 1 in state 2 and row 2 in state 3. The enumeration of pure stationary g_j's for player 2 is done similarly. The bimatrix game (A, B) is given as

$$A=\begin{bmatrix} 9 & 6 & 18 & 15 & 3 & 0 & 12 & 9 \\ 6 & 10 & 15 & 19 & 0 & 4 & 9 & 13 \\ 16 & 13 & 9 & 6 & 10 & 7 & 3 & 0 \\ 13 & 17 & 6 & 10 & 7 & 11 & 0 & 4 \\ 3 & 0 & 12 & 9 & 10 & 7 & 19 & 16 \\ 0 & 4 & 9 & 13 & 7 & 11 & 16 & 20 \\ 10 & 7 & 3 & 0 & 17 & 14 & 10 & 7 \\ 7 & 11 & 0 & 4 & 14 & 18 & 7 & 11 \end{bmatrix},$$

$$B=\begin{bmatrix} 65 & 82 & 17 & 47.6 & 98 & 141.4 & 11.6 & 75 \\ 85 & 77 & 53 & 38.6 & 118 & 136.4 & 60.4 & 50 \\ 40 & 57 & 23 & 53.6 & 53 & 80.4 & 22.4 & 105 \\ 60 & 52 & 59 & 44.6 & 73 & 75.4 & 71.2 & 80 \\ 75 & 100 & 27 & 72.0 & 91 & 128.8 & 4.6 & 40 \\ 95 & 95 & 63 & 63.0 & 111 & 123.8 & 53.4 & 15 \\ 50 & 75 & 33 & 78.0 & 46 & 67.8 & 15.4 & 70 \\ 70 & 70 & 69 & 69.0 & 66 & 62.8 & 64.2 & 45 \end{bmatrix}.$$

Using the Lemke–Howson algorithm [8], the following Nash equilibrium point was found

$$\xi^* = (0, 0, \tfrac{192}{1613}, \tfrac{408}{1613}, 0, 0, 0, \tfrac{1013}{1613}), \qquad \eta^* = (0, 0, \tfrac{10}{91}, \tfrac{39}{91}, 0, 0, \tfrac{42}{91}, 0).$$

We get $f^* = \sum_i \xi_i^* f_i$ and $g^* = \sum_j \eta_j^* g_j$ given by

$$f^*(s) = \begin{cases} (\tfrac{600}{1613}, \tfrac{1013}{1613}) & \text{for } s=1, \\ (0, 1) & \text{for } s=2, \\ (\tfrac{192}{1613}, \tfrac{1421}{1613}) & \text{for } s=3, \end{cases} \qquad g^*(s) = \begin{cases} (\tfrac{7}{13}, \tfrac{6}{13}) & \text{for } s=1, \\ (0, 1) & \text{for } s=2, \\ (\tfrac{4}{7}, \tfrac{3}{7}) & \text{for } s=3. \end{cases}$$

The equilibrium expected costs $w(1)$, $w(2)$, $w(3)$ for player 2 at states 1, 2, and 3, respectively, are given by

$$w(1) = \tfrac{34325}{1613}, \quad w(2) = \tfrac{35640}{1613}, \quad w(3) = \tfrac{28420}{1613}.$$

These are the equilibrium expected costs for player 2 corresponding to the Nash equilibrium pair (f^*, g^*). To find the equilibrium expected costs for player 1 we have to compute $v(s) = \Phi_1^\beta(f^*, g^*)(s)$, which are given by the unique solution to the equations

$$v(s) = r_1(s, f^*(s), g^*(s)) + \beta \sum_t v(t) q(t \mid s, g^*(s)), \quad s = 1, 2, 3.$$

We get

$$v(1)=\tfrac{25710}{2363},\quad v(2)=\tfrac{18960}{2363},\quad v(3)=\tfrac{23700}{2363}.$$

Example 4.2. We consider here a stochastic game with undiscounted costs and with irreducible transition. This game has 2 states. In each state player 1 has 2 actions and player 2 has 3 actions. The arrow underneath each column indicates the chances of moving to states 1 and 2.

$$\overset{s=1}{\begin{bmatrix}(9,4) & (6,7) & (4,8)\\ (5,7) & (7,5) & (9,3)\end{bmatrix}},\qquad \overset{s=2}{\begin{bmatrix}(2,9) & (9,1) & (8,2)\\ (9,3) & (0,9) & (8,1)\end{bmatrix}}.$$
$$\begin{matrix}\downarrow & \downarrow & \downarrow & & \downarrow & \downarrow & \downarrow\\ (0.5,0.5) & (0.8,0.2) & (0,1) & & (0.5,0.5) & (0.2,0.8) & (1,0)\end{matrix}$$

For example if $s=2$ and column 2 is chosen by player 2, the game moves to state 1 with chance 0.2 and stays at state 2 with chance 0.8. It is easy to see that the game is irreducible.

The bimatrix game (A, B) has the cost entries as follows:

$$A=\begin{bmatrix}11 & 18 & 17 & 8 & 15 & 14 & 6 & 13 & 12\\ 18 & 9 & 17 & 15 & 6 & 14 & 13 & 4 & 12\\ 7 & 14 & 13 & 9 & 16 & 15 & 11 & 18 & 17\\ 14 & 5 & 13 & 16 & 7 & 15 & 18 & 9 & 17\end{bmatrix},$$

$$B=\begin{bmatrix}6.5 & \frac{13}{7} & \frac{10}{3} & \frac{53}{7} & 4 & \frac{37}{6} & \frac{26}{3} & \frac{13}{6} & 5\\ 3.5 & \frac{53}{7} & 3 & \frac{41}{7} & 8 & 6 & \frac{14}{3} & \frac{53}{6} & 4.5\\ 8 & \frac{19}{7} & \frac{16}{3} & \frac{43}{7} & 3 & \frac{27}{6} & 7 & \frac{4}{3} & 2.5\\ 5 & \frac{59}{7} & 5 & \frac{31}{7} & 7 & \frac{26}{6} & 3 & 8 & 2\end{bmatrix}.$$

By the Lemke-Howson algorithm we get

$$\xi^*=(0,\tfrac{2}{3},0,\tfrac{1}{3})\quad\text{and}\quad \eta^*=(0,0,\tfrac{5}{9},0,0,0,0,0,\tfrac{4}{9})$$

as a Nash pair.

Since the transitions are irreducible the expected Nash equilibrium cost $w(s)=w=\frac{11}{3}$ for player 2. Further the induced stationary Nash equilibrium point (f^*, g^*) is given by

$$f^*(s)=\begin{cases}(\frac{2}{3},\frac{1}{3}), & s=1,\\ (0,1), & s=2,\end{cases}\qquad g^*(s)=\begin{cases}(\frac{5}{9},0,\frac{4}{9}), & s=1,\\ (0,0,1), & s=2.\end{cases}$$

The Nash equilibrium cost for player 1 is $v(s)=v=\frac{226}{31}$, $s=1,2$.

Remark. In general any Nash equilibrium point (ξ^*, η^*) of (A, B) may not necessarily induce a Nash equilibrium point for the player 2 control undiscounted stochastic game. The following is a counter example.

Counter example. We consider an undiscounted stochastic game with 2 states, where player 1 has always a null cost. Transitions and costs of player 2 are as follows:

$$\begin{array}{cc} s=1 & s=2 \\ [0,1], & [0,1]. \\ \downarrow\ \downarrow & \downarrow\ \downarrow \\ 1\ \ 2 & 2\ \ 1 \end{array}$$

Let $g_1=(11)$, $g_2=(12)$, $g_3=(21)$, and $g_4=(22)$. For example, g_3 selects column 2 in state 1 and column 1 in state 2. Player 1 has exactly one pure strategy, say f^*. Clearly,

$$A=[0 \quad 0 \quad 0 \quad 0]$$

and

$$B=[\Phi_2(f^*, g_j)(1)+\Phi_2(f^*, g_j)(2)]=[0 \quad 0 \quad 0 \quad 2].$$

For example, $\xi^*=1$, $\eta^*=(0, 0.5, 0.5, 0)$ is a Nash equilibrium point to (A, B). However, $g^*=0.5g_2+0.5g_3$ selects each action with chance 0.5, in every state, and $r_2(1, f^*(1), g^*(1))=r_2(2, f^*(2), g^*(2))=0.5$. Thus $\Phi_2(f^*, g^*)=0.5>\Phi_2(f^*, g_3)=0$. Hence (f^*, g^*) is not a Nash equilibrium point for the stochastic game.

References

[1] D. Blackwell, "Discrete dynamic programming," *Annals of Mathematical Statistics* 33 (1962) 719–726.

[2] J.A. Filar, "Ordered field property for stochastic games when the player who controls transition changes from state to state," *Journal of Optimization Theory and Applications* 34 (1981) 503–513.

[3] J.A. Filar, "On stationary equilibria of a single-controller stochastic game," *Mathematical Programming* 30 (1984) 313–325.

[4] J.A. Filar and T.E.S. Raghavan, "A matrix game solution of a single controller stochastic game," *Mathematics of Operations Research* 9 (1984) 356–362.

[5] A.M. Fink, "Equilibrium in a stochastic *n*-person game," *Journal of Science of Hiroshima University, Series, A-I* 28 (1964) 89–93.

[6] D. Gillette, "Stochastic games with zero stop probabilities," in: *Contributions to the Theory of Games III. Annals of Mathematical Studies No. 39* (Princeton University Press, Princeton, NJ, 1957) pp. 179–187.

[7] A. Hordijk and L.C.M. Kallenberg, "Linear programming and Markov games I, II," in: O. Moeschlin and D. Pallaschke, eds., *Game Theory and Mathematical Economics* (North-Holland, Amsterdam, 1981).

[8] T. Parthasarathy and T.E.S. Raghavan, *Some Topics in Two-Person Games* (American Elsevier Publishing Corporation, New York, 1971).

[9] T. Parthasarathy and T.E.S. Raghavan, "An orderfield property for stochastic games when one player controls transition probabilities," *Journal of Optimization Theory and Applications* 33 (1981) 375–392.

[10] T.E.S. Raghavan and J.A. Filar, "Algorithms for stochastic games — a survey," to appear in: *Zeitschrift fur Operations Research.*

[11] S.M. Ross, "Non-discounted denumerable Markovian decision models," *Annals of Mathematical Statistics* 39 (1968) 412–423.

[12] L.S. Shapley, "Stochastic games," *Proceedings of the National Academy of Sciences of the U.S.A.* 39 (1953) 1095–1100.

[13] M.J. Sobel, "Noncooperative stochastic games," *Annals of Mathematical Statistics* 42 (1971) 1930-1935.

[14] M. Takahashi, "Equilibrium points of stochastic noncooperative *n*-person games," *Journal of Science of Hiroshima University, Series A-I* 28 (1964) 95-99.

[15] O.J. Vrieze, "Linear programming and undiscounted stochastic games in which one player controls transitions," *OR Spektrum* 3 (1981) 29-35.