

THE COMPLETELY MIXED SINGLE-CONTROLLER STOCHASTIC GAME¹

JERZY A. FILAR

ABSTRACT. We consider a zero-sum stochastic game with finitely many states and actions. Further we assume that the transition probabilities depend on the actions of only one player (player II, in our case), and that the game is *completely mixed*. That is, every optimal stationary strategy for either player assigns a positive probability to every action in every state. For these games, properties analogous to those derived by Kaplansky [4] for the completely mixed matrix games, are established in this paper. These properties lead to the counterintuitive conclusion that the controller need not know the law of motion in order to play optimally, but his opponent does not have this luxury.

1. Introduction. Any $m \times n$ matrix $M = (m_{ij})_{i,j=1}^{m,n}$ can be regarded; as a two-person, zero-sum (matrix) game, with m_{ij} denoting the amount player II will pay player I if II chooses an action $j \in \{1, 2, \dots, n\}$ and I chooses an action $i \in \{1, 2, \dots, m\}$. A *mixed (or randomized) strategy* for player I (II) in such a game is an $m(n)$ -component probability vector $\mathbf{x}(\mathbf{y})$ whose i th (j th) entry x_i (y_j) denotes the probability that player I (II) will choose action i (j). It was von Neumann (e.g., see [9]) who proved the well-known minimax theorem for matrix games. It is a consequence of this theorem that there always exists a strategy pair $(\mathbf{x}^0, \mathbf{y}^0)$ satisfying

$$(1.1) \quad \mathbf{x}^T M \mathbf{y}^0 \leq (\mathbf{x}^0)^T M \mathbf{y}^0 \leq (\mathbf{x}^0)^T M \mathbf{y},$$

for all mixed strategies $\mathbf{x}(\mathbf{y})$ of player I (II). The strategies $\mathbf{x}^0, \mathbf{y}^0$ are called *optimal* and the number $v(M) = (\mathbf{x}^0)^T M \mathbf{y}^0$ is called *the value* of the game M .

In the classical paper Kaplansky [4] considered the class of *completely mixed* matrix games, namely, those games in which all optimal strategies are strictly positive in every component. It follows from Kaplansky's results that these games possess a number of desirable properties which general matrix games lack. In particular, if M is completely mixed, with every entry strictly positive (there is no

Received by the editors July 2, 1984 and, in revised form, January 17, 1985.

1980 *Mathematics Subject Classification*. Primary 90D20; Secondary 90D05.

Key words and phrases. Stochastic games, single-controller, completely mixed property, stationary strategies.

¹This work was supported in part by NSF Grant #ECS-820-4677. I am grateful to Lodewijk Kallenberg for reading an earlier version of this paper and for his helpful comments. Any errors in the present manuscript are the author's alone.

loss of generality in this assumption since a constant can always be added to every entry without changing the game strategically) we have that

- (a) the matrix M is square and nonsingular,
- (b) there exists a unique optimal strategy pair $(\mathbf{x}^0, \mathbf{y}^0)$,
- (c) against any pure action i (j) for player I (II) the optimal strategy \mathbf{y}^0 (\mathbf{x}^0) yields precisely the value, that is,

$$(1.2) \quad (\mathbf{x}^0)^T M = v(M) \mathbf{1}^T \quad \text{and} \quad M \mathbf{y}^0 = v(M) \mathbf{1},$$

where $\mathbf{1}$ is a vector with unity in every component. It follows from (1.2) that

$$(1.3) \quad v(M) = |M| / \left(\sum_{i=1}^n \sum_{j=1}^n M_{ij} \right),$$

where $|M|$ is the determinant, and M_{ij} is the (i, j) th cofactor of the matrix M .

In this paper we show that nearly all of the above results can be generalized to the class of *completely mixed single-controller stochastic games*. These are discrete time, dynamic games in which at each stage the players play one out of a finite set of matrix games. Further, the identity of the matrix game to be played at the next stage is determined by a stochastic “law of motion” which depends on the game which is being played presently and on the current action of only one of the two players I or II. The class of single-controller stochastic games in a sense lies halfway between the Markovian decision process and the general stochastic game. Consequently, these games possess many desirable properties that general stochastic games lack (e.g., see [8, 5, 10, 3]). In Filar and Raghavan [1] some properties of these games which resemble properties of matrix games were established. These results suggest that in a single-controller stochastic game the relationship between stationary and pure stationary strategies is in many ways analogous to the relationship between mixed and pure strategies in matrix games. In this paper we shall show that for the class of *completely mixed (c.m.)*, single-controller stochastic games, the above analogy is nearly perfect. In the process we derive simple formulae for computing the unique pair of optimal stationary strategies which also lead to the following counterintuitive conclusion: The controller can play optimally without the knowledge of the “law of motion”, while his opponent cannot!

2. Definitions and preliminaries. A *stochastic game* as formulated by Shapley [7] is played in stages. At each stage, the game is in one of finitely many states, $s = 1, 2, \dots, S$, in which players I and II are obliged to play a matrix game $A^s = (a(i, j, s))_{i,j=1}^{m,n}$ once. The “law of motion” is defined by the probabilities $q(s'|s, i, j)$, where the event $\{s'|s, i, j\}$ is the event that the game will enter state s' at the next stage given that at the current stage the state of the game is s , and that I and II choose the i th row and the j th column of A^s , respectively. In general players' strategies will depend on complete past histories. In this paper, however, we shall only be concerned with *stationary strategies*. We may represent a typical stationary strategy f for player I by a “composite” vector, $f = (f(1), f(2), \dots, f(S))$, where

each $f(s)$ is a probability vector² given by $f(s) = (f_1(s), f_2(s), \dots, f_{m_s}(s))$. Here, $f_i(s)$ is the probability that I chooses the i th row of A^s whenever the game is in state s . Player II's stationary strategies are similarly defined.

Once we specify the initial state and a strategy pair (f, g) for players I and II, we implicitly define a probability distribution over all sequences of states and actions which can occur during the game and consequently over all sequences of payoffs to player I. Let $\pi_n(f, g)(s)$ denote the expected income to player I at the n th stage when players I and II use the strategy pair (f, g) and the game begins in s . The two types of stochastic games which we shall consider are determined by the manner in which players evaluate a stream of payoffs (π_1, π_2, \dots) . They are

(a) *The β -discounted games* $\Gamma_\beta = \{\Gamma_\beta(1), \Gamma_\beta(2), \dots, \Gamma_\beta(S)\}$. Here, $\beta \in [0, 1)$ and $\Gamma_\beta(s)$ refers to the game beginning in state s . In such games, $\Phi_\beta(f, g)(s)$, the expected income to player I in $\Gamma_\beta(s)$ when the strategy pair (f, g) is used, is defined by

$$(2.1) \quad \Phi_\beta(f, g)(s) = \sum_{n=1}^{\infty} \beta^{n-1} \pi_n(f, g)(s).$$

(b) *The undiscounted or average-reward games* $\Gamma = \{\Gamma(1), \Gamma(2), \dots, \Gamma(S)\}$ are defined by

$$(2.2) \quad \Phi(f, g)(s) = \liminf_{N \rightarrow \infty} \frac{1}{N} \sum_{n=1}^N \pi_n(f, g)(s)$$

($\Phi(f, g)(s)$ has analogous meaning to $\Phi_\beta(f, g)(s)$). To show that a number $v_\beta(s)$ is the *value* of $\Gamma_\beta(s)$ for $s = 1, 2, \dots, S$, it is sufficient to show that there exists a stationary pair of strategies (f^β, g^β) (*optimal strategy pair*) such that, for each s ,

$$(2.3) \quad \Phi_\beta(f, g^\beta)(s) \leq v_\beta(s) = \Phi_\beta(f^\beta, g^\beta)(s) \leq \Phi_\beta(f^\beta, g)(s)$$

for any stationary f for player I and g for player II. For the undiscounted game $\Gamma(s)$, $v(s)$; $s = 1, 2, \dots, S$ and an optimal stationary pair (f^0, g^0) is similarly defined.

All the stochastic games considered below will be constrained by the hypothesis H1: *Player II controls the law of motion*, that is, $q(s'|s, i, j) \equiv q(s'|s, j)$ for all s', s, i and j .

It should be clear that as a result of H1, a stationary strategy g for player II determines an $S \times S$ Markov matrix $Q(g) = (q(s'|s, g))_{s, s'=1}^S$, where

$$q(s'|s, g) = \sum_{j=1}^{n_s} q(s'|s, j) g_j(s).$$

Now if (f, g) is any pair of stationary strategies we define a *current* payoff vector associated with this pair by $r(f, g) = (r(f, g, 1), \dots, r(f, g, S))$ where

$$r(f, g, s) = f(s) A^s g(s) = \sum_{i=1}^{m_s} \sum_{j=1}^{n_s} a(i, j, s) f_i(s) g_j(s).$$

²We shall not always differentiate between n -component row and column vectors. This is intended to simplify our already complicated notation and should not confuse the readers.

Further, since $Q(g)$ is a Markov matrix it is known that there exists a Markov matrix $Q^*(g)$ such that

$$(2.4) \quad Q^*(g) = \lim_{N \rightarrow \infty} \frac{1}{N+1} \sum_{n=0}^N Q^n(g)$$

where $Q^0(g) = I$ is the identity matrix.

It can be verified from the above definitions and the properties of $Q^*(g)$ that, for every s and every pair of stationary strategies (f, g) ,

$$(2.5) \quad \Phi_\beta(f, g)(s) = \{ [I - \beta Q(g)]^{-1} r(f, g) \}_s \quad \text{and} \\ \Phi(f, g)(s) = \{ Q^*(g) r(f, g) \}_s,$$

where $\{ \}_s$ denotes the s th component of a vector.

In the β -discounted game Γ_β with each state payoff matrix A^s we can associate a ‘‘Shapley matrix’’

$$A^s(\beta) = \left[a(i, j, s) + \beta \sum_{s'=1}^S q(s'|s, j) v_\beta(s') \right]_{i,j=1}^{m_s, n_s}.$$

Shapley [7] proved that if $(f^\beta(s), g^\beta(s))$ is an optimal strategy pair in the matrix game $A^s(\beta)$ for each s , then the stationary strategy pair (f^β, g^β) so created is optimal in Γ_β . Conversely, it is easy to check that if (f^β, g^β) is an optimal stationary strategy pair in Γ_β then $(f^\beta(s), g^\beta(s))$ is an optimal pair in $A^s(\beta)$ for each s .

We shall say that a stationary strategy g^0 is *uniformly discount optimal* for player II if there exists $\beta^0 \in [0, 1)$ such that g^0 is optimal for II in Γ_β for all $\beta \in (\beta^0, 1)$. For games constrained by H1, Parthasarathy and Raghavan [5] have obtained the following results: For each s there exists a square nonsingular submatrix $\dot{A}^s(\beta)$ of $A^s(\beta)$ and a corresponding submatrix \dot{A}^s of A^s such that if we define for all $\beta > \beta^0$

$$(2.6) \quad f^\beta(s) = v_\beta(s) \mathbf{1} [\dot{A}^s(\beta)]^{-1} \quad \text{and} \quad g^\beta(s) = v_\beta(s) [\dot{A}^s(\beta)]^{-1} \mathbf{1},$$

where $\mathbf{1}$ is a vector with 1 in every component, then the set of pairs $(f^\beta(s), g^\beta(s))$ obtained from (2.6) by adding 0's in the places corresponding to the rows/columns of A^s which are not in \dot{A}^s forms an optimal stationary strategy pair (f^β, g^β) in Γ_β for all $\beta > \beta^0$. Further, it is shown in [5] that, for each s , \dot{A}^s is nonsingular and there is a constant $\theta(s)$ such that

$$(2.7) \quad \dot{g}^0(s) = \theta(s) [\dot{A}^s]^{-1} \mathbf{1} = g^\beta(s) \quad \text{for all } \beta > \beta^0.$$

That is, $g^0 = (g^0(1), \dots, g^0(S))$ where each $g^0(s)$ is obtained by completing $\dot{g}^0(s)$ with 0's is a uniformly discount optimal strategy for player II. Also, it turns out that (f^0, g^0) , where $f^0(s) = \lim_{\beta \rightarrow 1^-} f^\beta(s)$, is an optimal stationary strategy pair in the undiscounted game Γ .³

A *pure stationary strategy* σ for player I is an S -tuple $(\sigma(1), \sigma(2), \dots, \sigma(S))$ where $\sigma(s)$ is an m_s -component, degenerate probability vector. There are $t = \prod_{s=1}^S m_s$ such strategies and they will be labelled $\sigma^1, \sigma^2, \dots, \sigma^t$. Similarly, let $\nu^1, \nu^2, \dots, \nu^p$

³ In [5] it was assumed without loss of generality that $a(i, j, s) > 0$ for all i, j, s . We shall continue to make the same assumption throughout this paper.

denote the pure stationary strategies for player II, where $p = \prod_{s=1}^S n_s$. Now, for each s we may define a matrix game $A(s) = (\Phi(\sigma^i, \nu^j)(s))_{i,j=1}^{l,p}$. In Filar and Raghavan [1] it was shown that for the single controller (player II) undiscounted stochastic game the value of the matrix game $v(A(s)) = v(s)$ for each s . Further, if $X(s)$ (F) is the set of mixed (stationary) strategies for I in $A(s)$ (Γ) respectively, then equations (2.5) and (2.4) in [1] define transformations T and \hat{T} such that $T: F \rightarrow X(s)$ and $\hat{T}: X(s) \rightarrow F$ for each s . It turns out that if $\xi = T(f)$ or if $f = \hat{T}(\xi)$, then

$$(2.8) \quad \Phi(f, g)(s) = \sum_{i=1}^l \xi_i \Phi(\sigma^i, g)(s), \quad s = 1, 2, \dots, S,$$

where $g \in G$ is held fixed. Analogous results can be established for the discounted stochastic games constrained by H1.

It should be noted that an undiscounted stochastic game with player I's stationary strategy held fixed can be regarded as an average reward Markovian decision process ((AMD)-process) as far as player II is concerned. Such a process can be solved with the help of a single pair of primal-dual linear programs as a consequence of results proved by Hordijk and Kallenberg [2]. In particular, if we fix player I's optimal strategy f^0 , then it follows from [2, Theorem 6] that there exists an S -vector γ such that, for each $j = 1, \dots, n_s, s = 1, \dots, S$,

$$(2.9) \quad v(s) \leq \sum_{s'=1}^S q(s'|s, j)v(s')$$

and

$$(2.10) \quad v(s) + \gamma(s) \leq \sum_{i=1}^{m_s} a(i, j, s)f_i^0(s) + \sum_{s'=1}^S q(s'|s, j)\gamma(s').$$

For the discounted Markovian decision process linear programming techniques analogous, but simpler, than those for the AMD-process have been known for some time.

3. Main results. Let Γ be an undiscounted single-controller stochastic game with positive payoffs, and let F^0 and G^0 denote the sets of optimal stationary strategies for players I and II respectively. We shall say that Γ is *completely mixed* (c.m. for short) if $(f^0, g^0) \in F^0 \times G^0$ implies that $f_i^0(s)$ and $g_j^0(s)$ are strictly positive for all i, j and s .

We shall establish analogs of (a)–(c) of §1 for the case of undiscounted games only. However, the corresponding results for the discounted case can be verified by considerably simpler arguments, since in this latter case the results of Shapley [7] can be invoked.

LEMMA 3.1. *Let Γ be as above. Then A^s is square and nonsingular for each $s = 1, 2, \dots, S$.*

PROOF. The comments in §2 imply that one pair of optimal stationary strategies (f^0, g^0) is obtained as the limit of (f^β, g^0) as $\beta \rightarrow 1^-$, where each $(f^\beta(s), g^0(s))$ is an extension (by insertion of 0's) of $(\hat{f}^\beta(s), \hat{g}^0(s))$ satisfying (2.6) for all $\beta > \beta^0$ and

with the same set of submatrices $A^s(\beta)$ of $A^s(\beta)$. But since $f^0(s)$ and $g^0(s)$ are strictly positive, $A^s(\beta)$ must equal $A^s(\beta)$ for each s . Thus A^s is square for each s , and nonsingular by Lemma 4.1 of [5]. \square

REMARK 3.2. Since the game is completely mixed and since player II controls the transitions, all strictly positive stationary strategies for II partition the states of the undiscounted game Γ into the same sets of ergodic chains C_1, C_2, \dots, C_k and the same set of transient states H . It follows from §5 of Filar and Raghavan [1] that if $s \in H$ then $n_s = 1$, and that the game can be subdivided into k subgames $\Gamma_1, \dots, \Gamma_k$ corresponding to the above ergodic chains. Of course, $v(s) = v_c$ for all $s \in C_c$ where $c = 1, 2, \dots, k$.

PROPOSITION 3.3. Let $(f^0, g^0) \in F^0 \times G^0$. Then for all pure stationary strategies σ^i and ν^j for players I and II we have that, for every s ,

$$(3.1) \quad v(s) = \Phi(\sigma^i, g^0)(s) = \Phi(f^0, \nu^j)(s).$$

PROOF. We know that $\Phi(\sigma^i, g^0)(s) \leq v(s)$ for all $i = 1, 2, \dots, t$ and $s = 1, 2, \dots, S$. Suppose that $\Phi(\sigma^m, g^0)(\bar{s}) < v(\bar{s})$ for some \bar{s} and m . Let $\xi^0 = T(f^0)$ be defined by the transformation (2.5) in [1]. Since $f^0(s) > \mathbf{0}$ for each s , $\xi^0 > \mathbf{0}$ also. Now by (2.8) we have that

$$v(\bar{s}) = \Phi(f^0, g^0)(\bar{s}) = \sum_{i=1}^t \xi_i^0 \Phi(\sigma^i, g^0)(\bar{s}) < \sum_{i=1}^t \xi_i^0 v(\bar{s}) = v(\bar{s}),$$

which is contradictory. Thus $\Phi(\sigma^i, g^0)(s) = v(s)$ for all i and s . The second equality in (3.1) is proved by first considering each of the subgames $\Gamma_1, \dots, \Gamma_k$ from Remark 3.2 separately. The strategy pair $(f^0, g^0)_c$ consisting of $(f^0(s), g^0(s))$ for $s \in C_c$ is clearly optimal in Γ_c . With player I's strategy held fixed at $f^0(s)$ for $s \in C_c$ the game becomes an AMD-process with the rewards to player II denoted by

$$r_c(f^0, j, s) = \sum_{i=1}^{m_s} a(i, j, s) f_i^0(s)$$

(when II chooses action j in state s , that is). However, $g^0(s)$ for $s \in C_c$ is still optimal for II in this process, and because of its positivity we can conclude from [2, Proposition 4] that there exists vector γ^0 which together with $v_c \mathbf{1}$ satisfy (2.9) and (2.10) with equalities. Thus we have

$$v_c + \gamma^0(s) = r_c(f^0, j, s) + \sum_{s'=1}^S q(s'|s, j) \gamma^0(s')$$

for all $j = 1, 2, \dots, n_s$ and $s \in C_c$. Hence for any pure stationary $\nu^j(s)$, $s \in C_c$,

$$(3.2) \quad v_c \mathbf{1} + \gamma^0 = r_c(f^0, \nu^j) + Q(\nu^j) \gamma^0.$$

Note that every vector in (3.2) has S_c components, where S_c is the cardinality of C_c . Now, multiplying (3.2) by the $S_c \times S_c$ matrix $Q^*(\nu^j)$ (see (2.4)) and using the fact that $Q^*(\nu^j)Q(\nu^j) = Q^*(\nu^j)$ we obtain $v_c \mathbf{1} = \Phi_c(f^0, \nu^j)$. We have now proved that, for any pure stationary strategy ν^j and any $s \in H^c$ (the complement of H), $\Phi(f^0, \nu^j)(s) = v(s)$. Suppose now that $s \in H$; that is, s is transient whenever II's stationary strategy is strictly positive everywhere. By Remark 3.2, $n_s = 1$ for all

$s \in H$ and hence in these states $g^0(s) = v^j(s)$ for all $j = 1, 2, \dots, p$. We shall define $p_g(s, s')$ to be the probability that s' is the first state of H^c encountered if the initial state is $s \in H$ and player II uses the stationary strategy g . From the above it follows that $p_{g^0}(s, s') = p_{v^j}(s, s')$ for all $j = 1, \dots, p, s \in H$ and $s' \in H^c$. The next equation is analogous to (5.1) in [1]:

$$(3.3) \quad \Phi(f^0, g)(s) = \sum_{s' \in H^c} p_g(s, s') \Phi(f^0, g)(s'),$$

where $s \in H$, and g is any stationary strategy for II (note that every g ensures exit from H in finite time with probability 1). Substituting first g^0 and then v^j for g in (3.3) we obtain identical right-hand sides (since (3.1) is already proved for $s' \in H^c$), thus even when $s \in H, \Phi(f^0, v^j)(s) = \Phi(f^0, g^0)(s) = v(s)$. This completes the proof. \square

PROPOSITION 3.4. *The set G^0 is a singleton. That is, player II possesses a unique stationary optimal strategy.*

PROOF. We already know that there exists a uniformly discount optimal strategy $g^0 \in G^0$ such that $g^0(s)$ satisfies (2.7) for each s . Suppose that g^* is also in G^0 . By Remark 3.2 the sets C_1, C_2, \dots, C_k and H are the same under g^0 and g^* (since the game is c.m.). Also it is clear that $g^0(s) = g^*(s)$ for all $s \in H$. We shall prove that $g^0(s) = g^*(s) = g^*(s)$ for all $s \in C_c$ for each $c = 1, \dots, k$, by considering the subgames Γ_c separately. Without loss of generality we consider only Γ_1 and assume that C_1 has S_1 states. Of course, $v_1 = v(s)$ is a constant for all $s \in C_1$. The stationary matrices $Q^*(g^0)$ and $Q^*(g^*)$ each have identical rows $u^0 = (u^0(1), \dots, u^0(S_1))$ and $u^* = (u^*(1), \dots, u^*(S_1))$ with $u^0(s)$ and $u^*(s) > 0$ for all $s \in C_1$. By Proposition 3.3, for every pure stationary strategy σ for player I in Γ_1 we have, for every $s \in C_1$,

$$(3.4) \quad \begin{aligned} v_1 &= \Phi(\sigma, g^0)(s) = [Q^*(g^0)r(\sigma, g)]_s \\ &= \sum_{s'=1}^{S_1} u^0(s') [\sigma(s') A^{s'} g^0(s')] \\ &= \sum_{s'=1}^{S_1} \sum_{j=1}^{n_s} r(\sigma, j, s') u_j^0(s'), \end{aligned}$$

where $u_j^0(s') = u^0(s') g_j^0(s')$ for all $j = 1, \dots, n_s$, and $s' \in C_1$. Equation (3.4) holds with g^* in place of g^0 and with $u_j^*(s') = u^*(s') g_j^*(s')$ in place of $u_j^0(s')$. Now let $z_j(s) = u_j^0(s) - u_j^*(s)$ for all $j = 1, \dots, n_s$ and $s \in C_1$. Then from (3.4) we obtain

$$(3.5) \quad \sum_{s'=1}^{S_1} \sum_{j=1}^{n_s} r(\sigma, j, s') z_j(s') = 0,$$

for every pure stationary strategy σ for I in Γ_1 . Let

$$Z = (Z_1; Z_2; \dots; Z_{S_1})^T$$

be a column vector such that $Z_s = (z_1(s), z_2(s), \dots, z_{n_s}(s))$ for each $s \in C_1$, and let t_1 be the number of pure stationary strategies for I in Γ_1 . Fix $\sigma(s)$ for each $s \geq 2$ and consider the n_1 equations extracted from (3.5) by letting $\sigma(1)$ range over the n_1 -dimensional unit basis vectors (recall that A^1 is square). These are equivalent to

$$(3.6) \quad A^1 Z_1 = \alpha \mathbf{1}.$$

Now, if $\alpha = 0$ by nonsingularity of A^1 (Lemma 3.1) we have that $Z_1 = u^0(1)g^0(1) - u^*(1)g^*(1) = 0$, and since $g^0(1)$ and $g^*(1)$ are probability vectors we obtain, from the above, that $u^0(1) = u^*(1)$ and hence $g^0(1) = g^*(1)$. If, however, $\alpha \neq 0$ then $Z_1 = \alpha[A^1]^{-1}\mathbf{1}$ which, by comparison with (2.7), implies that either $Z_1 > \mathbf{0}$ or $Z_1 < \mathbf{0}$. So, if $\hat{Z}_1 = (1/\zeta)Z_1$ where $\zeta = Z_1 \cdot \mathbf{1}$ and $\hat{\alpha} = \alpha/\zeta$, then from (3.6) and (2.7) we have

$$(3.7) \quad A^1 \hat{Z}_1 = \hat{\alpha} \mathbf{1} \quad \text{and} \quad A^1 g^0(1) = \theta(1)\mathbf{1},$$

where both \hat{Z}_1 and $g^0(1)$ are now strictly positive probability vectors. It follows from (3.7) that $\hat{\alpha} = \theta(1)$ and $\hat{Z}_1 = g^0(1)$, which from definitions of Z_1 and ζ implies that $u^0(1)g^0(1) - u^*(1)g^*(1) = (u^0(1) - u^*(1))g^0(1)$, and hence $g^*(1) = g^0(1)$ (since $u^0(1), u^*(1) > 0$). Thus we have proved that $g^*(1) = g^0(1)$ in every case. Similarly, by selecting suitable σ 's in (3.5) we can show that $g^*(s) = g^0(s)$ for every $s \in C_1$. Hence in the subgame Γ_1 player II has a unique optimal strategy; similarly, for every other subgame Γ_c where $c = 1, 2, \dots, k$. Hence in the original game Γ , $G^0 = \{g^0\}$. \square

PROPOSITION 3.5. *The set F^0 is a singleton; That is, player I possesses a unique optimal stationary strategy.*

PROOF. Once again we shall consider subgame Γ_1 corresponding to an ergodic class C_1 induced by g^0 , the optimal stationary strategy for player II. Let F_1^0 be the set of stationary optimals for player I in Γ_1 . For every $f \in F_1^0$, we have a corresponding AMD-process in which $g^0(s)$ for $s \in C_1$ is optimal for player II. That is, by an argument such as that used to derive (3.2) we can claim that there exists an S_1 -vector γ such that

$$(3.8) \quad r(f, j, s) + \sum_{s'=1}^{S_1} q(s'|s, j)\gamma(s') - \gamma(s) = v_1,$$

for $j = 1, 2, \dots, n_s$ and $s \in C_1$. Set

$$W = (f(1); \dots; f(S_1); \gamma)^T,$$

and define a set of $n_x \times S_1$ matrices P^s and E^s the j th rows of which are vectors $q(s, j) = (q(1|s, j), q(2|s, j), \dots, q(S_1|s, j))$ and e_s the s th vector of the unit basis of S_1 -dimensional space, respectively. Since we know that player I possesses at least one optimal strategy f^0 , there must be a vector γ^0 such that $W^0 = (f^0; \gamma^0)$ satisfies (3.8). Suppose f^* is another optimal policy. Then there exists γ^* such that $W^* = (f^*; \gamma^*)$ is also a solution of (3.8). Let

$$Z = W^0 - W^* = (Z(1); \dots; Z(S_1); \omega)^T.$$

Then $Z(s)$ and ω satisfy, for each $s \in C_1$ (recall that $r(f, j, s) = [f(s)A^s]j$), the equation (since $E^s\omega = \omega_s\mathbf{1}$)

$$(3.9) \quad (A^s)^T Z(s) + P^s\omega - \omega_s\mathbf{1} = \mathbf{0}.$$

Since A^s is nonsingular by Lemma 3.1, we now have

$$(3.10) \quad Z(s) = \omega_s [(A^s)^T]^{-1} \mathbf{1} - [(A^s)^T]^{-1} P^s\omega$$

for every $s \in C_1$. But each $Z(s) = f^0(s) - f^*(s)$ so $\mathbf{1}^T Z(s) = 0$. Thus for every $s \in C_1$,

$$(3.11) \quad \omega_s [(A^s)^{-1} \mathbf{1}]^T \mathbf{1} - [(A^s)^{-1} \mathbf{1}]^T P^s\omega = 0.$$

However, from (2.7), $(A^s)^{-1} \mathbf{1} = (1/\theta(s))g^0(s)$ where $g^0(s)$ for $s \in C_1$ form an optimal strategy for II (note that $\theta(s) > 0$ as a result of the assumption that $\alpha(i, j, s) > 0$, always). Substituting this in (3.11) we obtain, for each $s \in C_1$,

$$\omega_s - (g^0(s))^T P^s\omega = 0,$$

or, equivalently (using the fact that $[g^0(s)^T P^s]_{s'} = q(s'|s, g^0)$),

$$(3.12) \quad \omega = Q(g^0)\omega.$$

Hence ω is an eigenvector of $Q(g^0)$ corresponding to eigenvalue $\lambda = 1$. But since $Q(g^0)$ is an irreducible stochastic matrix with 1 as its maximal eigenvalue we have from the Perron–Frobenius Theorem that the corresponding eigenspace has dimension 1. Thus $\omega = \rho\mathbf{1}$ (since $\mathbf{1}$ satisfies (3.12)). Substituting this in (3.9) yields that, for every s ,

$$(3.13) \quad (A^s)^T Z(s) + \rho[P^s\mathbf{1} - \mathbf{1}] = (A^s)^T Z(s) = \mathbf{0}.$$

Therefore, $Z(s) = f^0(s) - f^*(s) = \mathbf{0}$ for all $s \in C_1$ and so $F_1^0 = \{f^0\}$. Since a similar argument applies to every Γ_c , $c = 1, \dots, k$, and since, for $s \in H$, A^s is a 1×1 matrix, we have proved that player I possesses a unique optimal stationary policy. \square

REMARK 3.5. Note that we have derived formulae for computing the unique pair of optimal stationary strategies in the undiscounted game. For player II, (2.7) and Lemma 3.1 imply that

$$(3.14) \quad g_j^0(s) = \sum_{i=1}^{n_s} A_{ij}^s / \sum_{i=1}^{n_s} \sum_{j=1}^{n_s} A_{ij}^s,$$

for $j = 1, \dots, n_s$; $s = 1, 2, \dots, S$, where A_{ij}^s is the (i, j) th cofactor of matrix A^s . This, of course, corresponds exactly to one of Kaplansky’s formulae (see (1.2)). On the other hand, for player I in a state $s \in C_c$ the optimal strategy $f^0(s)$ is obtained from (3.8)

$$(3.15) \quad f^0(s) = v(s) [(A^s)^T]^{-1} \mathbf{1} + [(A^s)^T]^{-1} (E^s - P^s)\gamma,$$

where γ is any solution of the system of equations (since $\mathbf{1}^T f^0(s) = 1$)

$$(3.16) \quad (I - Q(g^0))\gamma = \theta - v_c\mathbf{1}.$$

Note that $Q(g^0)$ is an $S_c \times S_c$ irreducible stochastic matrix and

$$\theta(s) = \text{Det}(A^s) / \sum_{i=1}^{n_s} \sum_{j=1}^{n_s} A_{ij}^s$$

for each $s \in C_c$ by (2.7). Hence, the value v_c in any state of Γ_c is the convex combination of $\theta(s)$'s obtained upon multiplying (3.16) on the left by $Q^*(g^0)$. Surprisingly, perhaps, the expression (3.14) implies that the controller (player II) can calculate his optimal strategy without the knowledge of the law of motion. Indeed, this optimal strategy remains invariant under all perturbations of the law of motion which do not destroy the completely mixed property! Nonetheless, his opponent must use the law of motion to compute his optimal strategy via (3.15), the first term of which corresponds to Kaplansky's formula (1.2). Of course, if $s \in H$, both players have only one action to choose from, and $v(s)$ is an appropriate convex combination of v_c 's (e.g., see (3.3) with g replaced by g^0).

REMARK 3.6. We conclude with the following question suggested by the referee: Can analogous results be established for the case of completely mixed single-controller stochastic games? Generalizations of Kaplansky's results to bimatrix games are well known (e.g., see Raghavan [6]), however, the lines of argument used in this section depend heavily on a number of results from [1 and 5] which were derived specifically for the zero-sum case.

REFERENCES

1. J. A. Filar and T. E. S. Raghavan, *A matrix game solution to a single-controller stochastic game*, Math. Oper. Res. **9** (1984), 356–362.
2. A. Hordijk and L. C. M. Kallenberg, *Linear programming and Markov decision chains*, Management Sci. **25** (1979), 352–362.
3. ———, *Linear programming and Markov games*. II, Game Theory and Mathematical Economics (O. Moeschlin and D. Pallaschke, eds.), North-Holland, Amsterdam, 1981, pp. 307–319.
4. I. Kaplansky, *A contribution to von Neumann's theory of games*, Ann. of Math. (2) **46** (1945), 474–479.
5. T. Parthasarathy and T. E. S. Raghavan, *An orderfield property for stochastic games, when one player controls the transition probabilities*, J. Optim. Theory Appl. **33** (1981), 375–392.
6. T. E. S. Raghavan, *Completely mixed strategies in bimatrix games*, J. London Math. Soc. (2) **2** (1970), 709–712.
7. L. S. Shapley, *Stochastic games*, Proc. Nat. Acad. Sci. U.S.A. **39** (1953), 1095–1100.
8. M. A. Stern, *On stochastic games with limiting average payoff*, Ph.D. Thesis, Univ. of Illinois at Chicago Circle, Chicago, Ill. 1975.
9. J. von Neumann and O. Morgenstern, *Theory of games and economic behaviour*, Princeton Univ. Press, Princeton, N. J., 1944.
10. O. J. Vrieze, *Linear programming and undiscounted stochastic games in which one player controls transitions*, OR Spektrum **3** (1981), 29–35.

DEPARTMENT OF MATHEMATICAL SCIENCES, JOHNS HOPKINS UNIVERSITY, BALTIMORE, MARYLAND 21218