On stochastic games with additive reward and transition structure

Raghavan, T.E.S.; Tijs, S.H.; Vrieze, O.J.

Published in:
Journal of Optimization Theory and Applications

Publication date:
1985

Document Version
Peer reviewed version

On Stochastic Games with Additive Reward and Transition Structure

T.E.S. RAGHAVAN[1], S.H. TIJS[2] and O.J. VRIEZE[2]

Abstract. In this paper we introduce a new class of two-person stochastic games with nice properties. For games in this class the payoffs as well as the transitions in each state consist of a part which depends only on the action of the first player and a part dependent only on the action of the second player.

For the zero-sum games in this class we prove that the orderfield property holds in the infinite horizon case, and that there exist optimal pure stationary strategies for the discounted as well as the undiscounted payoff criterion. For both criteria also finite algorithms are given to solve the game. An example shows that for non-zero sum games in this class, there are not necessarily pure stationary equilibria. But if such a game possesses a stationary equilibrium point, then there also exists a stationary equilibrium point which uses in each state at most two pure actions for each player.

Key Words. Game Theory, Stochastic Game, Pure Stationary Optimal Strategy, Additive Stochastic Game.

---

[1] Department of Mathematics, University of Illinois, Chicago, U.S.A.

[2] Department of Mathematics, Catholic University, Nijmegen, The Netherlands

1. Introduction

We consider *stochastic games* of the form $\Gamma = \langle S, \{A_s; s \in S\}, \{B_s; s \in S\}, r, p \rangle$. Here $S := \{1,2,\ldots,z\}$ is the *state space* and the finite sets A_s and B_s are the *action spaces*, available to the players I and II respectively, in state s. Further $r = (r_1, r_2)$ is a vector valued function with domain $T := \{(s,i,j);\ s \in S,\ i \in A_s,\ j \in B_s\}$ and range $\mathbb{R}^2$, where r_1 and r_2 are the *reward functions* of players I and II, respectively. Finally, $p = \{p(t|s,i,j);\ t \in S,\ (s,i,j) \in T\}$ prescribes the *law of motion*, where $p(t|s,i,j)$ denotes the probability that the state moves from s to t when i and j are the actions of the players in state s. Of course, for the *transition probabilities* we have $p(t|s,i,j) \geq 0$ and $\sum_t p(t|s,i,j) = 1$ for all $(s,i,j) \in T$.

We say that the game Γ possesses *additive rewards* if for all $(s,i,j) \in T$:

$$r_1(s,i,j) = r_{11}(s,i) + r_{12}(s,j)$$

$$r_2(s,i,j) = r_{21}(s,i) + r_{22}(s,j)$$

for some functions r_{11}, r_{12}, r_{21} and r_{22} on the appropriate domain. The game Γ is said to be *controlled by one player*, say player II, if $p(t|s,i,j) = p(t|s,i',j)$ for all $i,i' \in A_s$ and all $s,t \in S$ and $j \in B_s$. We write $p(t|s,i,j) = p(t|s,j)$ if no confusion is possible. Thus in such games the transition probabilities are not influenced by player I.

The game Γ is said to be a *switching control game* if the states can be partitioned into two sets S_1 and S_2, such that for $s \in S_1$, $p(t|s,i,j) = p(t|s,i)$ and for $s \in S_2$, $p(t|s,i,j) = p(t|s,j)$. That is, the law of motion from states of S_1 is independent of the action of player II and similarly the law of motion from states of S_2 is independent of the actions of player I.

The game Γ is said to possess *additive transitions* if, for all $(s,i,j) \in T$, $p(t|s,i,j) = p_1(t|s,i)+p_2(t|s,j)$, where p_1 is a function of the state and the action of player I and p_2 a function of the state and the action of player II.

The game is called a *zero-sum game* if $r_1+r_2 = 0$ on T. Otherwise the game is called a *nonzero-sum game*. In the following we suppose that the players have an infinite horizon. A play proceeds as usual in stochastic games (cf. Ref.1). We will be concerned with both discounted and undiscounted payoffs. The state and actions on the τ-th day will be denoted by s_τ, i_τ and j_τ.

A *stationary strategy* f for player I consists of a z-tuple $f = (f_1,f_2,\ldots,f_z)$, where f_s is a probability distribution on A_s. Intuitively this means that when the game is in s, player I, when adopting f as strategy, chooses an action according to f_s. A *behavioural strategy* μ is a sequence $\mu = (\mu_0,\mu_1,\ldots)$ where on the τ-th day $\mu_\tau(s_0,i_0,j_0,s_1,i_1,j_1,\ldots, i_{\tau-1},j_{\tau-1},s_\tau)$ is a probability distribution on A_{s_τ}, which depends on the history $h_\tau = (s_0,i_0,j_0,s_1,\ldots,j_{\tau-1},s_\tau)$ up to the τ-th day. A stationary strategy f for player I is called *pure* if f_s is degenerate for each $s \in S$, i.e. f_s selects a particular action with probability 1. Let g and ν be similarly defined as stationary and behavioural strategy, respectively, for player II.

Let $V_\beta(\mu,\nu)(s)$ denote the pair of *expected* β-*discounted rewards* when μ and ν are the strategies of players I and II, s is the starting state and the discount factor equals $\beta \in [o,1)$. Thus

$$V_\beta(\mu,\nu)(s) = (E_{\mu\nu s}(\sum_{\tau=0}^{\infty} \beta^\tau r_1(s_\tau,i_\tau,j_\tau)),E_{\mu\nu s}(\sum_{\tau=0}^{\infty} \beta^\tau r_2(s_\tau,i_\tau,j_\tau))).$$

Here $E_{\mu\nu s}$ denotes the expectation with respect to μ,ν and initial state s.

The second important evaluation rule in vogue is the so called *undiscounted* (or *average*) *payoff* defined by

$$V_1(\mu,\nu)(s) = (\liminf_{T\to\infty} E_{\mu\nu s}(\frac{1}{T+1}\sum_{\tau=0}^{T} r_1(s_\tau,i_\tau,j_\tau)), \limsup_{T\to\infty} E_{\mu\nu s}(\frac{1}{T+1}\sum_{\tau=0}^{T} r_2(s_\tau,i_\tau,j_\tau))$$

Here again $s_0 = s$ is the starting state. When $r_1+r_2 = 0$ we denote by V_β, V_1, etc., the payoff corresponding to player I.

In this paper special attention is paid to the class of stochastic games with additive rewards and additive transitions (ARAT-*games*). Zero-sum ARAT-games turn out to have nice optimal strategies and there are simple algorithms to solve such games as we will see in section 2. In the last section 3 some results for non-zero sum ARAT-games are derived.

2. The zero-sum case

A zero-sum stochastic game is said to possess a *value* if for each $s \in S$:

$$\inf_\nu \sup_\mu V_\beta(\mu,\nu)(s) = \sup_\mu \inf_\nu V_\beta(\mu,\nu)(s) =: V_\beta(s) \qquad (1)$$

Here equation (1) corresponds to the undiscounted case if $\beta = 1$. Strategies μ^* and ν^* for players I and II, respectively, are called *optimal strategies* if for each $s \in S$:

$$\inf_\nu V_\beta(\mu^*,\nu)(s) = V_\beta(s), \quad \sup_\mu V_\beta(\mu,\nu^*)(s) = V_\beta(s) \qquad (2)$$

Shapley introduced stochastic games and showed in his fundamental paper (Ref.1) that β-discounted stochastic games have a value and that both players possess stationary strategies which are optimal for each starting state. That is inf and sup can be replaced by min and max in the above equations (1) for $\beta \in [0,1)$. However, for undiscounted stochastic games the existence of a value was unknown till recently (cf. Mertens & Neyman, Ref.2). In general, however optimal strategies even in the class of behavioural strategies may not exist for this evaluation rule. Thus, without further restrictions on the rewards or the law of motion one cannot hope for stationary optimals.

If stochastic games have to be solved in finite steps one has to hope for the orderfield property. A zero-sum stochastic game is said to have the *orderfield property* if the coordinates $V_\beta(s)$ of the value of the game and the coordinates of suitable optimal strategies lie in the same ordered subfield of the reals as the data of the stochastic game.

Stern (Ref.3) in his PhD thesis first proved the existence of a value in stationary strategies for undiscounted stochastic games controlled by one player. Parthasarathy and Raghavan (Ref.4) showed, that for this class for both discounted and undiscounted payoffs the orderfield property holds. Also they gave a linear programming algorithm for solving these games, when the payoffs are discounted. Vrieze (Ref.5) and independently Hordijk and Kallenberg (Ref.6) gave a linear programming algorithm for solving these games with undiscounted payoffs. Filar (Ref.7) proved the existence of a value in stationary strategies for switching control stochastic games and he proved that also for this class the orderfield property holds. Vrieze et al (Ref.8) have given a finite step algorithm to solve these switching control games.

In looking for stochastic games with the orderfield property and optimal stationary strategies one needs conditions on the immediate payoffs or on the transition probabilities or on both. We note that one player control games and switching control games can be considered as subclasses of games with additive transition functions. For such games: $p(t|s,i,j) = p_1(t|s,i)+p_2(t|s,j)$, and if $p_1 = 0$, then such a game reduces to a player II controlling stochastic game , and if for each $s \in S$, $p_1(t|s,i) = 0$ for all $(t,i) \in S\times A_s$ or $p_2(t|s,j) = 0$ for all $(t,j) \in S\times B_s$, then the game corresponds to a switching control stochastic game . Thus a natural question to ask is whether games with additive transition functions admit stationary optimals and whether the orderfield property holds for this more general class.

Shapley's theorem implies the existence of stationary optimals in the discounted case. For the undiscounted case we have good indications that optimal stationary strategies exist. However we have not yet been able to prove this. The following example shows that the orderfield property does not hold.

Example 2.1.

$$\begin{bmatrix} 0 & -1 \\ -1 & 2 \end{bmatrix} \begin{matrix} \to 1 \\ \to 2 \end{matrix} \qquad \begin{bmatrix} 0 & -1 \\ -1 & 1 \end{bmatrix} \begin{matrix} \to 1 \\ \to 2 \end{matrix}$$

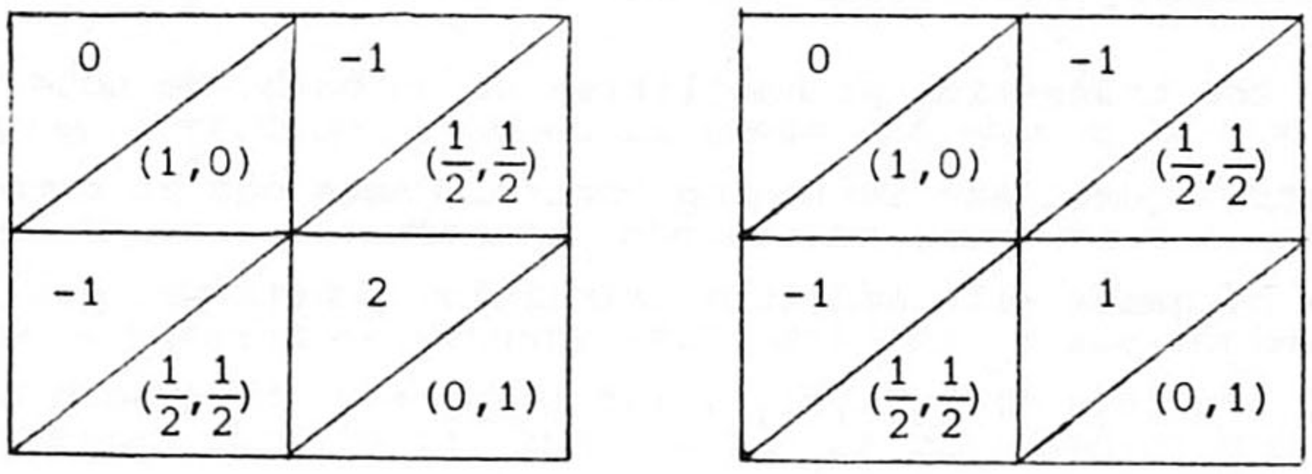
$\downarrow \quad \downarrow$ $\qquad$ $\downarrow \quad \downarrow$

$1 \quad 2$ $\qquad$ $1 \quad 2$

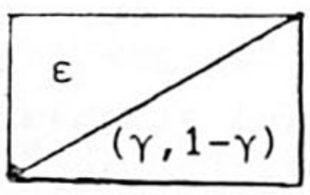
state 1 $\qquad$ state 2

In both states the player controlling the transitions is determined by the toss of an unbiased coin, after the players have chosen their action. The arrows associated with the pure actions indicate the state to which the game moves when that player controls the transitions. Trivially, the transition function is additive.

This game can also be written as



state 1 $\qquad$ state 2

where a box  corresponds to an immediate payoff ε and a jump with probability γ to state 1 and probability $1-\gamma$ to state 2.

For $\beta \in [0,1)$ the β-discounted value $(V_\beta(1), V_\beta(2))$ is given by the unique solution of

$$v_1 = \mathrm{Val}\begin{bmatrix} 0+\beta v_1 & -1+\frac{1}{2}\beta v_1+\frac{1}{2}\beta v_2 \\ -1+\frac{1}{2}\beta v_1+\frac{1}{2}\beta v_2 & 2+\beta v_2 \end{bmatrix} \text{ and } v_2 = \mathrm{Val}\begin{bmatrix} 0+\beta v_1 & -1+\frac{1}{2}\beta v_1+\frac{1}{2}\beta v_2 \\ -1+\frac{1}{2}\beta v_1+\frac{1}{2}\beta v_2 & 1+\beta v_2 \end{bmatrix}$$

Both matrix games are completely mixed, resulting in

$$v_1 = \frac{1}{4}(\beta v_1(2+\beta v_2)-(-1+\frac{1}{2}\beta v_1+\frac{1}{2}\beta v_2)^2) \qquad (3a)$$

$$v_2 = \frac{1}{3}(\beta v_1(1+\beta v_2)-(-1+\frac{1}{2}\beta v_1+\frac{1}{2}\beta v_2)^2) \qquad (3b)$$

Combining (3a) and (3b) yields

$$4v_1-\beta v_1(2+\beta v_2) = 3v_2-\beta v_1(1+\beta v_2) \text{ or}$$
$$v_2 = \frac{1}{3}(4-\beta)v_1 \qquad (4)$$

Substitution of (4) into (3a) results into

$$4v_1 = \beta v_1(2+\frac{1}{3}\beta(4-\beta)v_1)-(-1+\frac{1}{2}\beta v_1+\frac{1}{6}\beta(4-\beta)v_1)^2,$$

which leads to

$$V_\beta(1) = v_1 = \frac{12(12-\beta)-12\sqrt{144-24\beta}}{-2\beta^2(1-\beta)} = \frac{-6}{(1-\beta)(12-\beta+\sqrt{144-24\beta})} \qquad (5)$$

Since the value of the undiscounted game with initial state s = 1 equals $\lim_{\beta\uparrow 1}(1-\beta)V_\beta(1)$ it follows by (5), that neither for the discounted case nor for the undiscounted case the orderfield property holds, by noting that $V_1(1)$ is irrational, while the game parameters all are in the rational field, and that for the rational discount factor $\beta = \frac{1}{2}$, $V_\beta(1)$ is also irrational. This gives an indication that in order to obtain a nice solution of the game, one has to look for a further constraint on the game components. Such a constraint is additivity of the rewards. When the rewards and the transitions are both additive the problem is manageable as we will show below. Related work on additive games can be found in Parthasarathy & Raghavan (Ref.9) and Himmelberg, et al. (Ref.10). A main result for ARAT-zerosum games is given in the following theorem.

Theorem 2.1. Let the rewards and transitions be additive in a zero-sum stochastic game Γ. Then there are pure optimal stationary strategies for both players for discounted as well as undiscounted payoffs. Furthermore, the orderfield property holds for both criteria. Also there are pure stationary strategies for both players which are uniformly optimal for all discount factors sufficiently near to one.

Proof. The following is well-known (cf. Bewley & Kohlberg Ref.11). For a zero-sum stochastic game with finite state and action spaces, there exists a series

$$W(\alpha) = \Sigma_{k=-\infty}^{K} w_k (\alpha(1-\alpha)^{-1})^{k/K}$$

in fractional powers of $\alpha(1-\alpha)^{-1}$ and vectors $w_k \in \mathbb{R}^Z$ as coefficients, such that for each $\beta \in (0,1)$, sufficiently near to one, $W(\beta)$ equals the value V_β of the β-discounted game. Moreover $W(\alpha)$ satisfies for each $s \in S$ the so called limit discount equation:

$$W_s(\alpha) = \underset{A_s \times B_s}{\mathrm{Val}} [G_s(W(\alpha))], \tag{6}$$

where the (i,j)-th cell of the matrix game $[G_s(W(\alpha))]$ has content

$$g(s,i,j) = r(s,i,j) + \alpha \Sigma_t p(t|s,i,j) W_t(\alpha). \tag{7}$$

Here the matrix game $[G_s(W(\alpha))]$ is a game in the field of real Puiseux series (cf. Bewley & Kohlberg, Ref.11).

Furthermore it is known that the value of the undiscounted stochastic game equals (cf. Mertens & Neyman, Ref.2):

$$w_K = \lim_{\alpha \uparrow 1} (1-\alpha) W(\alpha) \tag{8}$$

In general optimal actions for the matrix game in (6) are quite complex and belong to the same Puiseux field as to which $W(\alpha)$ belongs. However for an additive game, $G_s(W(\alpha))$ can be decomposed as follows:

$G_s(W(\alpha)) = G_{1s}(W(\alpha))+G_{2s}(W(\alpha))$, where

$g_1(s,i,j) = r_1(s,i)+\alpha\Sigma_t p_1(t|s,i)W_t(\alpha)$ and

$g_2(s,i,j) = r_2(s,j)+\alpha\Sigma_t p_2(t|s,j)W_t(\alpha)$.

So $G_{1s}(W(\alpha))$ has identical columns and $G_{2s}(W(\alpha))$ has identical rows.

But then, when solving $G_s(W(\alpha))$, player I only needs to consider $G_{1s}(W(\alpha))$ and player II only needs to look at $G_{2s}(W(\alpha))$. This observation results in the fact that both players have optimal real pure actions in the limit discount equations. Let f^* be a pure stationary strategy for player I such that f^*_s is an optimal action in $[G_s(W(\alpha))]$ for each $s \in S$ and let g^* be similar for player II. Then by the theorems 6.1 and 6.2 of Bewley & Kohlberg, Ref.11, it follows that f^* and g^* are uniformly discount optimal and optimal for the undiscounted case.

That for each $\beta \in [0,1)$ both players have optimal pure stationary strategies can be shown in a similar way. Namely when in (6) we replace α by a fixed $\beta \in [0,1)$ and $W(\alpha)$ by V_β, we obtain Shapley's equation for the β-discounted game. Again the matrix game $[G_s(V_\beta)]$ can be decomposed into a part independent of player I and a part independent of player II. Application of Shapley's theorem does the rest. The orderfield property for the discounted case follows from the fact that for a pair of stationary strategies the associated discounted payoff is a rational function of β, and for the undiscounted case then the orderfield property follows from (8). [Cf. Ref.4]. □

Knowing that the orderfield property holds for ARAT-games, we now have an indication that there exists a finite algorithm. We do not know whether for ARAT-games there exists a one-step solution method, like solving one linear program. But for the discounted and also for the undiscounted criterion we will indicate now a finite step solution method. For the discounted ARAT-game the method of Hoffman & Karp (Ref.12) can be used, which proceeds as follows.

(i) Choose $v_0 = M(1-\beta)^{-1}1_z$ with $M := \min_{(s,i,j)\in T} r(s,i,j)$ and $1_z = (1,1,\ldots,1) \in \mathbb{R}^z$. Put $\tau := 0$.

(ii) Determine for player I a pure stationary strategy $f^\tau = (f_1^\tau, f_2^\tau, \ldots, f_z^\tau)$, such that f_s^τ is an optimal action for player I in the matrix game $[G_s(v_\tau)]$ for each $s \in S$ (cf. (6) and (7)).

(iii) Solve for player II the discounted Markov decision problem which results when player I fixes f^τ. This can be done for example by solving one linear programming problem. Let $v_{\tau+1}$ be the optimal value of this problem.

(iv) If $v_{\tau+1} \neq v_\tau$ put $\tau := \tau+1$ and return to (ii); else stop.

It is straightforward to show that $v_{\tau+1} \geq v_\tau$ componentwise and when f^τ is not optimal then $v_{\tau+1} \neq v_\tau$. That in step (ii) of the algorithm player I possesses optimal pure actions in $[G_s(v_\tau)]$ follows again from the fact that $G_s(v_\tau)$ can be decomposed into a part only depending on player I and a part only depending on player II. Since in each iteration player I strictly improves his strategy and since there are a finite number of pure stationary strategies it is clear that the algorithm stops after a finite number of iterations.

For the undiscounted additive game a finite algorithm can be developed which resembles the algorithm of Vrieze, et al. (Ref.8). Like the algorithm above of Hoffman & Karp also this algorithm can be described by the term "value oriented policy iteration". We will not give this algorithm in detail here, but indicate how the algorithm of Vrieze, et al (Ref.8) should be adapted. (The notations in their paper are used.) Throughout the algorithm we have $S_1 = S$ and $S_2 = \emptyset$. Further $\sigma^c(\tau) = \{\sigma_k^c(\tau); k\in S_1\}$, $\tau = 0,1,2,\ldots$, is a pure stationary strategy now, with as consequence that $\tilde{\Gamma}(\sigma^c(\tau+1))$ is a Markov decision problem. This Markov decision problem can be solved by the same LP1. More changes are not needed. The proof that also this modified algorithm stops after a finite number of iterations proceeds in the same way as in Vrieze, et al,

using the fact that there are a finite number of pure stationary strategies.

3. The nonzero-sum case

For nonzero-sum games the concept of equilibrium points is relevant.

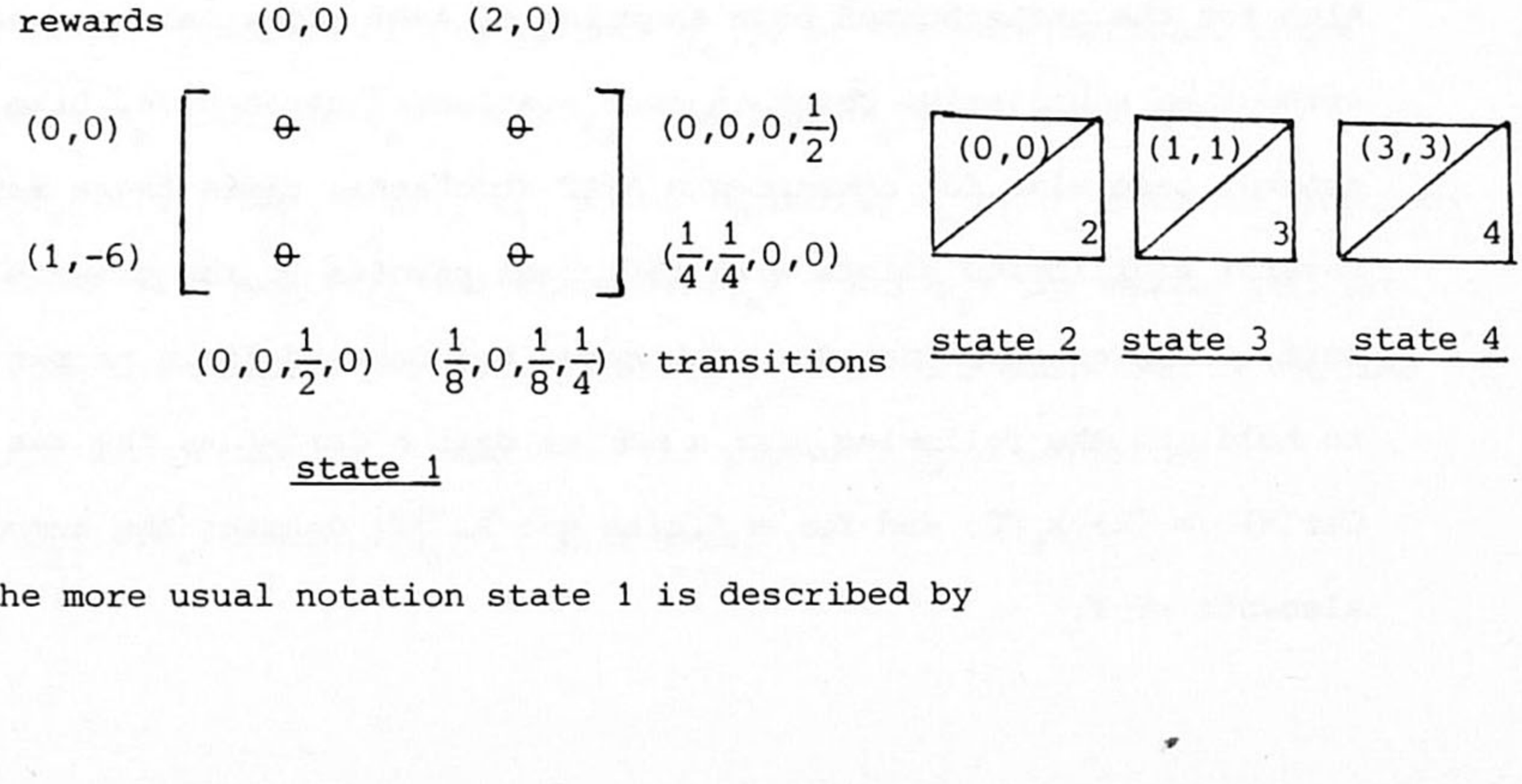
A pair of strategies (μ^*,ν^*) forms an *equilibrium point* if for all strategies μ and ν:

$$V_{\beta 1}(\mu,\nu^*) \leq V_{\beta 1}(\mu^*,\nu^*) \text{ and } V_{\beta 2}(\mu^*,\nu) \leq V_{\beta 2}(\mu^*,\nu^*) \qquad (9)$$

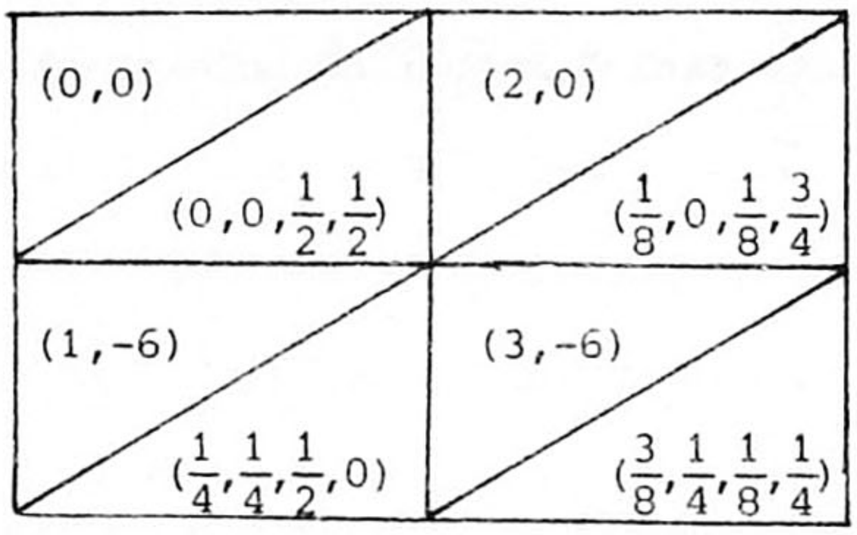
Again, in (9) to $\beta = 1$ we associate the undiscounted case.

It is well-known for the discounted case that there exist equilibrium points of stationary strategy pairs. (Refs. 13 and 14). For the undiscounted version in general the existence of equilibria is unknown. For different subclasses of stochastic games this problem is settled by Rogers (Ref.14), Federgruen (Ref.15), Parthasarathy & Raghavan (Ref. 4) and Parthasarathy, Tijs & Vrieze (Ref.16). In view of the results of the zero-zum case, the question arises whether for nonzero-sum ARAT-games, there exist equilibrium points of pure stationary strategy pairs. The following example answers this question in the negative.

Example 3.1. Consider the stochastic game with 4 states given by:



In the more usual notation state 1 is described by



The states 2,3 and 4 are absorbing. Both players have two pure stationary strategies corresponding to choosing their first and second action respectively in state 1. Let us denote these strategies by f^1, f^2, g^1 and g^2 respectively.

Take $\beta = \frac{1}{2}$. When we compute $V_{\frac{1}{2}^n}(f^k, g^\ell)$ for $n = 1,2$, $k = 1,2$ and $\ell = 1,2$, then we obtain

$$\begin{array}{c} \\ f^1 \\ f^2 \end{array} \begin{array}{c} \begin{array}{cc} g^1 & \quad g^2 \end{array} \\ \begin{bmatrix} (2,2) & (4\frac{2}{3}, 2\frac{8}{15}) \\ (1\frac{5}{5}, -6\frac{2}{7}) & (4\frac{10}{13}, -6\frac{4}{13}) \end{bmatrix} \end{array} \tag{10}$$

For example $V_{\frac{1}{2}^1}(f^1, g^2)$ can be computed as the unique solution v of

$$v = 2 + \frac{1}{2}\cdot\frac{1}{8}\cdot v + \frac{1}{2}\cdot\frac{1}{8}\cdot(1-\frac{1}{2})^{-1} + \frac{1}{2}\cdot\frac{3}{4}(1-\frac{1}{2})^{-1}\cdot 3, \text{ resulting in } v = 4\frac{2}{3}.$$

From (10) it can be seen that there exists no equilibrium point in pure stationary strategies and that for this example the unique equilibrium point in stationary strategies is completely mixed.

Also for the undiscounted case examples of ARAT games can be constructed without an equilibrium point in pure stationary strategies. Like in the general case also for nonzero-sum ARAT stochastic games there may be several equilibrium points with different payoffs to the players. Furthermore, examples show that for such games the ordered field property fails to hold. In the following, for $x \in \mathbb{R}$ we define $\mathrm{Car}(x)$ as the set $\mathrm{Car}(x) := \{k;\ x_k \neq 0\}$ and for a finite set T, $|T|$ denotes the number of elements of T.

For discounted nonzero-sum additive stochastic games we have the next remarkable theorem.

Theorem 3.1. If for a discounted nonzero-sum ARAT stochastic game the pair (f^*,g^*) forms an equilibrium point of stationary strategies, then there exists an equilibrium point $(\tilde{f},\tilde{g})$, such that $V_{\beta 1}(f^*,g^*) = V_{\beta 1}(\tilde{f},\tilde{g})$, $V_{\beta 2}(f^*,g^*) = V_{\beta 2}(\tilde{f},\tilde{g})$ and such that $|Car(\tilde{f}_s)| \leq 2$ and $|Car(\tilde{g}_s)| \leq 2$ for each state $s \in S$.

Proof. Let (f^*,g^*) be a stationary equilibrium point and let $V_1^* = V_{\beta 1}(f^*,g^*)$ and $V_2^* = V_{\beta 2}(f^*,g^*)$. This is equivalent to

$$\max_i (r_1(s,i,g_s^*)+\beta\Sigma_t p(t|s,i,g_s^*)V_1^*) = V_1^* \qquad (11)$$

and

$$\max_j (r_2(s,f_s^*,j)+\beta\Sigma_t p(t|s,f_s^*,j)V_2^*) = V_2^* \qquad (12)$$

where the maximum in (11) is reached at least for each $i \in Car(f_s^*)$ and in (12) the maximum is attained at least for each $j \in Car(g_s^*)$.

By the additivity of the game, (11) and (12) are equivalent to

$$\max_i (r_{11}(s,i)+\beta\Sigma_t p_1(t|s,i)V_1^*)+r_{12}(s,g_s^*)+\beta\Sigma_t p_2(t|s,g_s^*)V_1^* = V_1^* \qquad (13)$$

and

$$r_{21}(s,f_s^*)+\beta\Sigma_t p_1(t|s,f_s^*)V_2^* + \max_j (r_{22}(s,j)+\beta\Sigma_t p_2(t|s,j)V_2^*) = V_2^* \qquad (14)$$

Put $W_1 = r_{12}(s,g_s^*)+\beta\Sigma_t p_2(t|s,g_s^*)V_1^*$. Since $g_s \mapsto r_{12}(s,g_s)+\beta\Sigma_t p_2(t|s,g_s)V_1^*$ is a linear function of the weights $g_s(j)$ on the pure actions, there exists a $\tilde{g}_s$ with $Car(\tilde{g}_s) \subset Car(g_s^*)$ and $|Car(\tilde{g}_s)| \leq 2$ such that $W_1 = r_{12}(s,\tilde{g}_s)+\beta\Sigma_t p_2(t|s,\tilde{g}_s)V_1^*$. Hence replacing g_s^* by $\tilde{g}_s$ does not disturb equations (13) and (11) and since $Car(\tilde{g}_s) \subset Car(g_s^*)$ the maxima in (14) and (12) are reached for each $j \in Car(\tilde{g}_s)$. This procedure can be carried out for each state $s \in S$ and also for player I by considering $W_2 = r_{21}(s,f_s^*)+\beta\Sigma_t p(t|s,f_s^*)V_2^*$. This leads to

$$\max_i (r_1(s,i,\tilde{g}_s) + \beta \Sigma_t p(t|s,i,\tilde{g}_s) V_1^*) = V_1^* \qquad (15)$$

and

$$\max_j (r_2(s,\tilde{f}_s,j) + \beta \Sigma_t p(t|s,\tilde{f}_s,j) V_2^*) = V_2^* \qquad (16)$$

for each $s \in S$, where in (15) the maximum is attained at least for each $i \in Car(\tilde{f}_s)$ and in (16) the maximum is reached at least for each $j \in Car(\tilde{g}_s)$. Hence $(\tilde{f},\tilde{g})$ forms an equilibrium point and $V_{\beta 1}(\tilde{f},\tilde{g}) = V_1^* = V_{\beta 1}(f^*,g^*)$ and $V_{\beta 2}(\tilde{f},\tilde{g}) = V_2^* = V_{\beta 2}(f^*,g^*)$. □

Example 3.1 above shows that theorem 3.1 cannot be sharpened. An analogous statement like theorem 3.1 can be given for the undiscounted case. However the proof is far more complicated using Markov chain theory and Markov decision theory. Moreover we do not know whether for the undiscounted case equilibria of stationary strategies always exist, though we have good reasons to believe that this indeed is the case.