International Game Theory Review, Vol. 3, No. 4 (2001) 283–290

STATIONARY STRATEGIES IN ZERO-SUM STOCHASTIC GAMES

J. FLESCH, F. THUIJSMAN and O. J. VRIEZE

Department of Mathematics, Maastricht University,
P.O. Box 616, 6200 MD Maastricht, The Netherlands

We deal with zero-sum stochastic games. We demonstrate the importance of stationary strategies by showing that stationary strategies are better (in terms of the rewards they guarantee for a player, against any strategy of his opponent) than (1) pure strategies (even history-dependent ones), (2) strategies which may use only a finite number of different mixed actions in any state, and (3) strategies with finite recall. Examples are given to clarify the issues.

1. Introduction

A zero-sum stochastic game Γ can be described by a state space $S := \{1, \ldots, z\}$, and a corresponding collection $\{M_1, \ldots, M_z\}$ of matrices, where matrix M_s has size $m_s^1 \times m_s^2$ and, for $a \in A_s := \{1, \ldots, m_s^1\}$ and $b \in B_s := \{1, \ldots, m_s^2\}$, entry (a, b) of M_s consists of a payoff $r_s(a, b) \in \mathbb{R}$ and a probability vector $p_s(a, b) = (p_s(t|a, b))_{t \in S}$. The elements of S are called states and for each state $s \in S$ the elements of A_s and B_s are called (pure) actions of players 1 and 2 in state s. The game is to be played at stages in $\mathbb{N}$ in the following way. The play starts at stage 1 in an initial state, say in state $s^1 \in S$, where, simultaneously and independently, both players are to choose an action: player 1 chooses a row $a^1 \in A_{s^1}$, while player 2 chooses a column $b^1 \in B_{s^1}$. These choices induce an immediate payoff $r_{s^1}(a^1, b^1)$ from player 2 to player 1. Next, the play moves to a new state according to the probability vector $p_{s^1}(a^1, b^1)$, say to state s^2. At stage 2, new actions $a^2 \in A_{s^2}$ and $b^2 \in B_{s^2}$ are to be chosen by the players in state s^2. Then player 1 receives payoff $r_{s^2}(a^2, b^2)$ from player 2 and the play moves to some state s^3 according to the probability vector $p_{s^2}(a^2, b^2)$, and so on.

The sequence $h^n = (s^1, a^1, b^1; \ldots; s^n, a^n, b^n)$ is called the history up to stage n. The players are assumed to have complete information and perfect recall.

A mixed action for a player in state s is a probability distribution on the set of his actions in state s. Mixed actions in state s will be denoted by x_s for player 1 and by y_s for player 2, and the sets of mixed actions in state s by X_s and Y_s, respectively.

Naturally, $A_s \subset X_s$ and $B_s \subset Y_s$ for all states $s \in S$. A (history-dependent) strategy π for player 1 is a decision rule that prescribes a mixed action $\pi_s(h)$

in the present state s for any past history h of the play. For player 2, (history-dependent) strategies σ are defined similarly. We use the notations Π and Σ for the respective (history-dependent) strategy spaces of the players. If the mixed actions prescribed by a strategy only depend on the current state, then the strategy is called stationary. Thus, the stationary strategy spaces are $X := \times_{s \in S} X_s$ for player 1 and $Y := \times_{s \in S} Y_s$ for player 2. We will use the respective notations x and y for stationary strategies for players 1 and 2.

A strategy π is called pure if it always prescribes a pure action with probability 1, namely $\pi_s(h) \in A_s$ for all states s and histories h. Pure strategies are defined in a similar way for player 2. A pair of strategies (π, σ) together with an initial state $s \in S$ determine a stochastic process on the payoffs. The sequences of payoffs are evaluated by the average reward, given by

$$\gamma_s(\pi, \sigma) := \liminf_{N \to \infty} \mathbb{E}_{s\pi\sigma} \left(\frac{1}{N} \sum_{n=1}^{N} r_n \right),$$

where r_n denotes the random variable for the payoff at stage n.

For any initial state $s \in S$, it is in the spirit of the game to evaluate a strategy π of player 1 by the reward $\phi_s(\pi)$ that π guarantees when starting in s; so let

$$\phi_s(\pi) := \inf_{\sigma \in \Sigma} \gamma_s(\pi, \sigma).$$

Similarly, for strategies σ of player 2 define

$$\psi_s(\sigma) := \sup_{\pi \in \Pi} \gamma_s(\pi, \sigma).$$

Mertens and Neyman (1981) showed that

$$\sup_{\pi \in \Pi} \phi_s(\pi) = \inf_{\sigma \in \Sigma} \psi_s(\sigma) =: v_s \qquad \forall\ s \in S.$$

Here, $v := (v_s)_{s \in S}$ is called the average value of the game. A strategy π of player 1 is called ε-optimal, $\varepsilon \geq 0$, if

$$\phi_s(\pi) \geq v_s - \varepsilon \qquad \forall\ s \in S.$$

Similarly, a strategy σ of player 2 is called ε-optimal, $\varepsilon \geq 0$, if

$$\psi_s(\sigma) \leq v_s + \varepsilon \qquad \forall\ s \in S.$$

Because of the definition of the value v, both players have ε-optimal strategies for all $\varepsilon > 0$. However, 0-optimal strategies do not generally exist and stationary strategies are not always sufficient to achieve ε-optimality for all $\varepsilon > 0$ [see the famous game, the Big Match in Gillette (1957) and Blackwell and Ferguson (1968)].

2. The Results

We start with the following theorem, which states that, for each player, there exists a pure stationary strategy which is at least as good as any other pure (possibly history-dependent) strategy.

Theorem 2.1. *In any zero-sum stochastic game G, there exists a pure stationary strategy x for player* 1 *such that for any initial state $s \in S$ and any pure strategy π*

$$\phi_s(x) \geq \phi_s(\pi)\,.$$

A similar statement holds for player 2.

Proof. We only prove the statement for player 1; for player 2 a similar proof can be given. When player 1 uses a pure strategy (possibly a history-dependent one), then at any stage of the play, player 2 knows in advance which action player 1 is going to choose. Therefore, we examine a related zero-sum game $\bar{G}$ in which player 1 chooses an action first and then player 2 has to make his move. This can be done by replacing each state s by a state $(s,0)$ and, for each $a \in A_s$, a state (s,a). State $(s,0)$ has actions sets A_s for player 1 and $\{1\}$ for player 2; all payoffs are 0 and action $a \in A_s$ leads to state (s,a) with probablility 1. State (s,a) has actions sets $\{1\}$ for player 1 and B_s for player 2; action $b \in B_s$ gives a payoff $2 \cdot r_s(a,b)$ to player 1 and leads to state $(t,0)$ with probability $p_s(t|a,b)$. The game $\bar{G}$ is a so-called perfect information game, i.e. a stochastic game where in each state at most one player has a non-trivial set of actions. For such games, it is well known [cf. Liggett and Lippman (1969)] that player 1 has pure stationary optimal strategies (and so has player 2). Any such optimal strategy is better than any pure strategy for player 1 in the game $\bar{G}$ and, therefore, it corresponds in an obvious one-to-one fashion to a pure stationary strategy in the original game with the required property. □

Example 2.1. Consider the following game G:

5 1	4 2
0 1	6 1

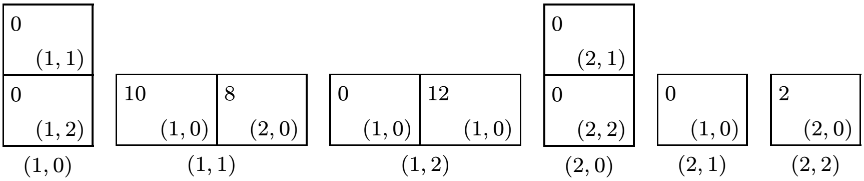
state 1

0 1
1 2

state 2

Player 1's actions are the rows, while players 2's actions are the columns. In each entry, the payoff is placed in the upper left corner, while the number in the bottom right corner denotes the state to which transition occurs with probability 1.

Following the construction in the proof of Theorem 2.1, the related zero-sum game $\bar{G}$ is as follows:



One can easily verify that the pure stationary strategies

$$\bar{x} = ((1,0),(1),(1),(1,0),(1),(1))$$

$$\bar{y} = ((1),(0,1),(1,0),(1),(1),(1))$$

are optimal for the respective players in $\bar{G}$, guaranteeing the value $\bar{v} = 2$. So, the corresponding pure stationary strategy for player 1 in the original game G is

$$x = ((1,0),(1,0))\,.$$

Clearly, x guarantees $\phi_1(x) = \phi_2(x) = 2$ in G, and no pure strategy (not even a history-dependent one) is able to guarantee more than 2.

Theorem 2.1 has the following consequence, which says that if a player is restricted to using only a finite number of *mixed* actions, in all states, then he cannot do better than to use a stationary strategy.

Corollary 2.1. *Let $\tilde{X}_s$ be a non-empty and finite subset of X_s, for all $s \in S$. Then there exists a stationary strategy $\tilde{x}$ for player 1 with the following properties*:

(1) *$\tilde{x}_s \in \tilde{X}_s$ for all $s \in S$.*
(2) *Suppose that π is a strategy for player 1 such that $\pi_s(h) \in \tilde{X}_s$ for any present state $s \in S$ and past history h. Then, for any initial state $s \in S$*

$$\phi_s(\tilde{x}) \geq \phi_s(\pi)\,. \tag{1}$$

A similar statement holds for player 2 as well.

Proof. The proof is straightforward. One only has to define a zero-sum game $\tilde{G}$ with the same state space $\tilde{S} = S$, such that in any state $s \in S$, player 1's actions are exactly the elements of $\tilde{X}_s$, player 2's action space remains B_s, while the payoffs and transitions are given by taking expectations. Then by applying Theorem 2.1 for $\tilde{G}$, we obtain a pure stationary strategy $\tilde{x}$ in $\tilde{G}$, which is also a stationary strategy of the original game G (not necessarily pure though). Both required properties of $\tilde{x}$ follow immediately. □

Next, we deal with strategies which, for some $m \in \mathbb{N} \cup \{0\}$, use only the present state and the past history of the last m stages (strategies with finite recall of m

stages). It turns out that, perhaps surprisingly, these strategies are not any better than stationary strategies (which have no recall at all).

Corollary 2.2. *Suppose a strategy* π *of player* 1 *satisfies for some* $m \in \mathbb{N} \cup \{0\}$ *that for any present state* $s \in S$ *and for all histories* $(s_1, a_1, b_1; \ldots; s_{n+m}, a_{n+m}, b_{n+m})$, *for all* $n \in \mathbb{N}$, *the prescribed mixed action*

$$\pi_s(s_1, a_1, b_1; \ldots; s_{n+m}, a_{n+m}, b_{n+m})$$

is independent of $(s_1, a_1, b_1; \ldots; s_n, a_n, b_n)$.

Then, there exists a stationary strategy x *for player* 1 *such that for any initial state* $s \in S$

$$\phi_s(x) \geq \phi_s(\pi) .$$

Proof. Since m is fixed, strategy π can only prescribe finitely many different mixed actions, in all states of the game. Therefore, the result immediately follows from Corollary 2.1. □

Finally, we wish to make some remarks regarding Corollary 2.2. (Similar remarks could be made on Corollary 2.1 as well.)

Remark 2.1. In Corollary 2.2, the condition that π has finite recall is crucial, as shown by a famous example, called the Big Match, which was introduced by Gillette (1957) and solved by Blackwell and Ferguson (1968). In that example, if player 1 is able to recall all the past actions of player 2 then he can guarantee higher rewards than by stationary strategies.

What is more, in another example by Flesch *et al.* (1999), it is sufficient to know only the present stage of the play (the "length" of the history) in order to beat stationary strategies. The example examined in Flesch *et al.* (1999) is the following one.

	L_1	R_1
T_1	1 1	0 1
B_1	1 2	1 4

state 1

	L_2	R_2
T_2	1 1	0 2
B_2	0 3	1 4

state 2

0 3

state 3

1 4

state 4

The notation is similar to that of Example 2.1. States 3 and 4 are so called "absorbing states". By entries (B_1, L_1) and (T_2, L_2) play can move between the two non-absorbing states.

This game has the following properties for initial states 1 and 2:

(a) The value is $v_1 = v_2 = 1$.

(b) For all $\varepsilon > 0$, player 1 has ε-optimal strategies that depend only on the stage number. Indeed, define strategy f^K for player 1, where $K \in \mathbb{N}$, as follows:

$$\text{at stage } n \text{ play } T_1 \text{ or } T_2 \text{ with probability } \sqrt[K]{\frac{n}{n+1}} \qquad \text{for all } n \in \mathbb{N}\,.$$

Observe that this strategy f^K is symmetric in the sense that the prescribed mixed actions in states 1 and 2 are the same for any stage n. Note that these Top probabilities converge to 1 as n tends to infinity, so f^K assigns less and less probabilities to actions B_1 and B_2.

For initial states 1 and 2, for all $\varepsilon > 0$, if $K \in \mathbb{N}$ is large, then player 1 can guarantee a reward at least $1 - \varepsilon$ by playing the strategy f^K, namely

$$\phi_1(f^K) \geq 1 - \varepsilon, \qquad \phi_2(f^K) \geq 1 - \varepsilon\,.$$

(c) Player 1 has no stationary ε-optimal strategy for initial states 1 and 2, if $\varepsilon \in [0, 1)$. In fact, player 1 can get at most 0 for initial states 1 and 2 by playing stationary strategies, namely

$$\phi_1(x) = \phi_2(x) = 0 \qquad \forall\, x\,.$$

Note that (b) implies (a), because the highest payoff in the game is 1.

Now we briefly explain (b). The question here is how player 2 can reply to the strategy f^K. Intuitively, player 2 has two hopes to decrease player 1's reward. The first one is achieving absorption in entry (B_2, L_2) with payoff 0. Player 2's best candidate would be playing actions L_1 and L_2 whenever the play is in state 1 or in state 2. But then whenever the play is in state 2, a transition occurs to state 1 with a large probability, and it takes a long time until the play comes back to state 2 again. Because the strategy f^K assigns decreasing probabilities to action B_1, the lengths of stay in state 1 will increase fast during the play and the frequency of visits to state 2 will tend to zero. As a consequence, the frequency of stages when absorption could occur is zero (in the limit) and the probabilities on action B_2 at those stages will decrease "rapidly". Therefore, the overall probability of absorption in entry (B_2, L_2) will be small. In conclusion, playing L_1 and L_2 gives player 2 little hope.

On the other hand, since the payoffs in entries (T_1, R_1) and (T_2, R_2) equal 0, player 2 could try to play actions R_1 and R_2 with a "positive" frequency and hope that the play will not absorb. But in that case, the frequency of stages when absorption could occur is positive and the probabilities on B_1 and B_2 at those stages decrease "slowly". Hence, it will appear that the play must eventually absorb with probability 1, and then the zero payoffs in entries (T_1, R_1) and (T_2, R_2) will have no influence on the reward.

Finally, we discuss (c). Note that strategies are completely determined by the choices for states 1 and 2. For each stationary strategy $x = (x_1, x_2)$ of player 1, we define a strategy $y^x = (y_1^x, y_2^x)$ for player 2: let

$$y_1^x := \begin{cases} (0,1) & \text{if } x_1 = (1,0) \\ (1,0) & \text{otherwise} \end{cases}, \qquad y_2^x := \begin{cases} (0,1) & \text{if } x_2 = (1,0) \\ (1,0) & \text{otherwise.} \end{cases}$$

Notice that, for $s = 1, 2$, we have $\gamma_s(x, y^x) = 0$ for all x, so the proof of (c) is complete.

Remark 2.2. Note that, in Corollary 2.2, the stationary strategy x depends on π. Consider the following game:

1 1	0 1
0 1	1 2

state 1

1 2

state 2

Let π_n be the stationary strategy for player 1, for any $n \in \mathbb{N}$, given by

$$\pi_n = \left(\left(\frac{n-1}{n}, \frac{1}{n} \right), (1) \right).$$

Clearly,

$$\phi_1(\pi_n) = \frac{n-1}{n} \quad \forall\, n \in \mathbb{N}.$$

However, there is no stationary strategy which would guarantee 1 for initial state 1, hence there is no x with the property that

$$\phi(x) \geq \phi(\pi_n) \quad \forall\, n \in \mathbb{N}.$$

Nevertheless, in any zero-sum stochastic game, for any $\varepsilon > 0$ one can show the existence of a stationary strategy x_ε such that for all initial states $s \in S$

$$\phi_s(x_\varepsilon) \geq \sup_x \phi_s(x) - \varepsilon$$

(in the example above, take π_n with a large $n \in \mathbb{N}$). Then, by Corollary 2.2, we obtain for all initial states $s \in S$ that

$$\phi_s(x_\varepsilon) \geq \phi_s(\pi) - \varepsilon$$

whenever π has finite recall. (So x_ε is independent of π).

References

Blackwell, D. and T. S. Ferguson (1968). "The Big Match". *Annals of Mathematical Statistics*, Vol. 39, 159–163.

Flesch, J., F. Thuijsman and O. J. Vrieze (1999). "Markov Strategies are Better than Stationary Strategies". *International Game Theory Review*, Vol. 1, 9–31.

Gillette, D. (1957). "Stochastic Games with Zero Stop Probabilities". In M. Dresher, A. W. Tucker and P. Wolfe (eds.), *Contributions to the Theory of Games III, Annals of Mathematical Studies*, Vol. 39. Princeton University Press, 179–187.

Liggett, T. M. and S. A. Lippman (1969). "Stochastic Games with Perfect Information and Time Average Payoff". *SIAM Review*, Vol. 11, 604–607.

Mertens, J. F. and A. Neyman (1981). "Stochastic Games". *International Journal of Game Theory*, Vol. 10, 53–66.