Semi-infinite Stochastic Games[1,2]

Sagnik Sinha, Frank Thuijsman and Stef H. Tijs

Abstract: We investigate two-person zero-sum stopping stochastic games with a finite number of states, for which the action sets of player I are finite and those for player II are countably infinite. Concerning the payoffs no restrictions are made. We show that for such games the value, possibly $-\infty$ in some coordinates, exists; player I possesses optimal stationary strategies and player II possesses near-optimal stationary strategies with finite support. Furthermore we relate the existence of value and of (near-)optimal stationary strategies with a maximal solution to the Shapley-equation.

1. Introduction: Let $M = [a_{ij}]_{i=1,j=1}^{m\ \ n}$ be an $m \times n$-matrix with $a_{ij} \in \mathbb{R}$ for all i and j. Then this matrix represents the following non-cooperative two-person game, called the matrix game M. There are two players, player I and player II, and a play in the game runs as follows. Simultaneously and independently of each other, player I chooses a row i and player II chooses a column j; then these choices are announced and player II has to pay the amount a_{ij} to player I.

Each of the players is allowed to randomize over pure action choices, which gives rise to the notion of mixed actions $p \in \Delta^m := \left\{x \in \mathbb{R}^m : x \geq 0, \sum_{i=1}^m x_i = 1\right\}$ for player I and mixed actions $q \in \Delta^n$ for player II. So, if player I uses the mixed action p and player II uses q, then with probability $p_i q_j$ entry (i, j) will be chosen, hence the expected payoff, by player II to player I, is pAq.

---

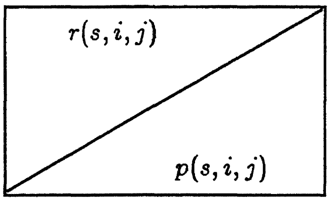
[1]This research started when Tijs was visiting the Indian Statistical Institute in New Delhi. We thank Professor T. Parthasarathy for his hospitality.

[2]Support by The Netherlands Organization for Scientific Research NWO, project 10-64-10, is gratefully acknowledged.

T. E. S. Raghavan et al. (eds.), Stochastic Games and Related Topics, 71–83.

Now assuming that player I wants to maximize the expected payoff and player II wants to minimize the same, it is interesting to know if there exists something like the 'worth' of the matrix game M and 'best' mixed actions for the players. This problem was solved in 1928 by von Neumann [1] who showed that for each matrix M, as defined above, there exists a unique $v \in \mathbf{R}$ and further $p^* \in \Delta^m$ and $q^* \in \Delta^n$ such that for all $p \in \Delta^m$, $q \in \Delta^n$ it holds that $pAq^* \leq v \leq p^*Aq$. This v is called the value of the matrix game M and is denoted by $\text{val}(M)$ or Let $M = \text{val}[a_{ij}]_{i=1,j=1}^{m\ \ n}$ Further p^* is called an optimal mixed action for player I and q^* an optimal mixed action for player II. So using p^*, when repeating this matrix game M over and over again, guarantees player I's average income to be at least v; a similar interpretation can be made for player II.

In 1953 Shapley [2] introduced the idea of not playing one and the same matrix game every day, but playing one of a finite collection of matrix games every day (stage), with a motion along the matrix games from stage to stage governed by the current matrix game and the actions chosen there. To be more precise, there are z matrices $M^1, M^2, \ldots, M^z$, matrix M^s has size $m_s \times n_s$ and entry (i,j) of matrix M^s is given as:



where $r(s,i,j) \in \mathbf{R}$ is a payoff to player I and

$$p(s,i,j) = (p(1|s,i,j), \ldots, p(z|s,i,j)) \in \Delta_o^z := \{x \in \mathbf{R}^z : x \geq 0, \sum_{s=1}^{z} x_s < 1\}.$$

The interpretation is: if in state s player I chooses $i \in A_s := \{1, 2, \ldots, m_s\}$ and player II chooses $j \in B_s := \{1, 2, \ldots, n_s\}$ then player II pays $r(s,i,j)$ to player I and next the play moves from state(matrix) s to state t with probability $p(t|s,i,j)$, and the play stops with probability $1 - \sum_{t=1}^{z} p(t|s,i,j) > 0$.

Thus, these 'stopping stochastic games' were naturally descended from matrix games. In stochastic games the players consider strategies; infinite plans that tell a player at each stage, at each state and for every history of the play, what mixed action to choose at that stage in that state. Strategies for which the prescribed mixed actions do not depend on the histories are called Markov strategies; Markov strategies for which the prescribed mixed actions do not depend on the stages are called stationary strategies. Stationary strategies are simply z-tuples of mixed actions, one mixed action for each state. It is obvious that stationary strategies are most easy to handle.

An initial state $s \in S$, together with a pair of strategies (π_1, π_2), for players I and II respectively, determine a stochastic process over the set of states $S = \{1, 2, \ldots, z\}$, and hence for all stages $\tau \in \mathbf{N}$ an expected direct payoff $R_{s\pi_1\pi_2}(\tau)$ by player II to player I is determined. The players evaluate this stream of expected payoffs $(R_{s\pi_1\pi_2}(1), R_{s\pi_1\pi_2}(2), \ldots)$ to be worth the reward $v(s, \pi_1, \pi_2) := \sum_{\tau=1}^{\infty} R_{s\pi_1\pi_2}(\tau)$.

By Shapley [2] it is known that there exists a unique $v \in \mathbf{R}^z$, a stationary strategy ρ^* for player I and a stationary strategy σ^* for player II such that for all π_1 and π_2 the following holds:

$$v(s, \pi_1, \sigma^*) \leq v(s) \leq v(s, \rho^*, \pi_2).$$

Here v is called the value of the stochastic game and $\rho^*(\sigma^*)$ is called an optimal stationary strategy for player I(II).

Furthermore, Shapley [2] shows that: v is the value of the stochastic game, ρ^* is an optimal stationary strategy for player I and σ^* an optimal stationary strategy for player II if and only if for all $s \in S$ one has that $v(s) = \text{val}M_s(v)$ and that $\rho^*(s)$ is an optimal mixed action for player I in $M_s(v)$ and that $\sigma^*(s)$ is an optimal mixed action for player II in $M_s(v)$, where $M_s(v) = [r(s,i,j) + \sum_{t=1}^{z} p(t|s,i,j)v(t)]_{i=1,j=1}^{m_s\ n_s}$. Hence the value and optimal stationary strategies can be found by solving the set of equations: $x_s = \text{val}M_s(x)$ for $s \in S, x \in \mathbf{R}^z$. This set of equations is known as the Shapley equation and can be shortened to $x = T(x)$ $(x \in \mathbf{R}^z)$, where T is the function from $\mathbf{R}^z$ to $\mathbf{R}^z$ defined by $(T(x))_s = \text{val}M_s(x)$. The equation $x = Tx$ $(x \in \mathbf{R}^z)$ has a unique solution in $\mathbf{R}^z$ since T is a continuous contraction map.

Since 1953 the theory of stochastic games has been extended in many ways. Non-stopping stochastic games, i.e. stochastic games for which $p(s,i,j) \in \Delta^z$ for all s, i, j have been examined under the β-discounted as well as the limiting average criterion, two different ways of evaluating the infinite streams of expected payoffs; nonzero-sum stochastic games have been studied; stochastic games have been examined with infinite state and/or action spaces. This paper belongs to the last category. In this last category one mostly needs some boundedness conditions on the payoffs to be sure that the rewards are well-defined for all pairs of strategies. In this paper we investigate stopping stochastic games with a finite state space where for each state player I has a finite action space and player II has a countably infinite action space. We put no restrictions on the payoffs. Player II's strategy space however is restricted to the set of strategies having finite support, i.e. strategies for which there exists an $N \in \mathbf{N}$ such that for all states, stages and histories, the probability that player II will choose a column in $\{1, 2, \ldots, N\}$ equals 1. This is done to assure that all direct expected payoffs are well defined.

In section 2 we show that these semi-infinite stochastic games have a value in $\mathbf{R}_-^z$, where $\mathbf{R}_- = \mathbf{R} \cup \{-\infty\}$, that player I possesses optimal stationary strategies and that player II possesses near-optimal stationary strategies. In section 3 we relate solutions for these stochastic games with solutions of the Shapley equation in $\mathbf{R}_-^z$.

We finish this section with some examples to illustrate that these semi-infinite (stopping) stochastic games are the only reasonable countably infinite extension of the finite stochastic games as introduced in [2], if one wants to avoid putting some boundedness conditions on the payoffs.

Example 1.1: For a stochastic game with countable state space and finite action spaces one needs bounded payoffs to get well-defined rewards. Take for example the stochastic game with state space $\mathbf{N}$ and action set $\{1\}$ for both players in all states. The transition is given by $p(s+1|s,1,1) = \frac{1}{2}$ for all $s \in \mathbf{N}$ and $p(t|s,1,1) = 0$ otherwise; the payoffs to player I are given by $r(s,1,1) = (-2)^{s-1}$. Consider the game starting in state 1. In that case the reward $v(s,\pi_1,\pi_2)$ is not defined if one wants to take $\sum_{\tau=1}^{\infty} R_{s\pi_1\pi_2}(\tau)$.

Example 1.2: For a stochastic game with finite state space and both players having countably many actions in the same state, the value need not exist. Take for example the stochastic game with $S = \{1\}$, $A_1 = B_1 = \mathbf{N}$, the transition is given by $p(1|1,i,j) = \frac{1}{2}$ for all i, j, and the payoffs are given by $r(1,i,j) = 1$ if $i > j$, $r(1,i,j) = 0$ if $i = j$, $r(1,i,j) = -1$ if $i < j$. It is clear that $sup_{\pi_1} inf_{\pi_2} v(1,\pi_1,\pi_2) = -2$, whereas $inf_{\pi_2} sup_{\pi_1} v(1,\pi_1,\pi_2) = 2$.

Example 1.3: For a stochastic game with finite state space for which for each state one of the players has a finite action set, the value need not exist either. Consider the stochastic game with $s = \{1,2\}$; $A_1 = B_2 = \mathbf{N}$, $A_2 = B_1 = \{1\}$; $p(2|1,i,j) = \frac{1}{2} = p(1|2,i,j)$ for all i and j and $p(t|s,i,j) = 0$ else; $r(1,i,1) = i$, $r(2,1,j) = -j$ for all i and j.

If the players are not restricted to strategies with finite support, then $\sum_{\tau=1}^{\infty} R_{s\pi_1\pi_2}(R(\tau))$ need not exist. If the players are restricted to strategies with finite support, then
$sup_{\pi_1} inf_{\pi_2} v(s,\pi_1,\pi_2) = -\infty$ whereas $inf_{\pi_2} sup_{\pi_1} v(s,\pi_1,\pi_2) = +\infty$ for both $s = 1$ and $s = 2$.

2. Semi-infinite games: Tijs [3] examined semi-infinite matrix games and showed that for such games the value exists, though it may be $-\infty$ if player II is the player with action set $\mathbf{N}$. For player II one cannot allow all probability distributions over $\mathbf{N}$ as mixed actions without encountering difficulties concerning the expected direct payoffs. Tijs [3] considers several restrictions on player II's set of mixed actions. One of them is restricting player II to use only mixed actions with finite support, i.e. mixed actions in $\bigcup_{n\in\mathbf{N}} \Delta^n$; where we identify Δ^n with the set $\{x \in \mathbf{R}^\infty : x \geq 0, \sum_{i=1}^n x_i = 1,\ x_i = 0 \text{ for } i > n\}$.

In this section we extend the work of Tijs [3] on semi-infinite matrix games, with the above restriction for player II, to semi-infinite stochastic games.

Preliminaries: A semi-infinite stochastic game Γ_∞ is given by a finite set of states S, where for each $s \in S$ there is a $m_s \times \infty$ matrix M^s of which entry (i,j) contains $r(s,i,j) \in \mathbf{R}$, the payoff to player I, and $p(s,i,j) \in \Delta_o^z$, a transition vector. However, we assume that $sup_{s,i,j}\{\sum_{t\in S} p(t|s,i,j)\} < 1$. Play proceeds in stages as explained in section 1. Player II is restricted to strategies π_2 with finite support, i.e. for each π_2 there exists an $N \in \mathbf{N}$ such that, before stopping, with probability 1, player II will choose one of the first N columns at all stages, in all states, for all histories.

Related with Γ_∞, for all $n \in \mathbf{N}$ one can look at the n-truncated stochastic game Γ_n, which we get by deleting all columns j with $j > n$, for all states. By completion with zeros, strategies for player II in Γ_n can be identified with finite support strategies for player II in Γ_∞; similarly, by deleting zeros, finite support strategies for player II in Γ_∞ can be identified with strategies for player II in Γ_n for n sufficiently large. We also identify Markov strategies for player I in Γ_∞ with those in Γ_n, $n \in \mathbf{N}$.

If there exists a $v_\infty \in \mathbf{R}_-^z$ such that for all $\epsilon > 0$ and all $\delta < 0$ there are strategies $\pi_{\epsilon\delta}^1$ and $\pi_{\epsilon\delta}^2$ for the respective players for which for all $s \in S$, all π^1, π^2 (with finite support)

$$v_\infty(s, \pi_{\epsilon\delta}^1, \pi^2) \geq v_\infty(s) - \epsilon$$

$$v_\infty(s, \pi^1, \pi_{\epsilon\delta}^2) \leq \begin{cases} v_\infty(s) + \epsilon & \text{if } v_\infty(s) \in \mathbf{R} \\ \delta & \text{if } v_\infty(s) = -\infty \end{cases}$$

then v_∞ is called the value of Γ_∞, and also $\pi_{\epsilon\delta}^1$ is called an $\epsilon\delta$-optimal strategy for player I and similarly $\pi_{\epsilon\delta}^2$is called an $\epsilon\delta-$ optimal strategy for player II. If one can take $\epsilon = 0$ and $\delta = -\infty$, then such $\epsilon\delta$-optimal strategies are called optimal strategies.

As we know the value v_n of Γ_n exists for all n.

Theorem 2.1:

(i) $lim_{n\to\infty} v_n$ exists in $\mathbf{R}_-^z$ and equals v_∞, the value of Γ_∞.

(ii) Player I has optimal stationary strategies in Γ_∞.

(iii) Player II has near-optimal stationary strategies in Γ_∞.

Proof: For all $n \in \mathbf{N}$ it holds that $v_n \geq v_{n+1}$, coordinatewise, since in Γ_{n+1} player II's action set is larger than in Γ_n, and player I's action set remains the same. As $v_1, v_2, \ldots$ is non-increasing it converges to some $w \in \mathbf{R}^z_-$, coordinatewise. We prove that w is the value of Γ_∞.

For every $n \in \mathbf{N}$ player I has an optimal stationary strategy ρ^n in Γ_n. All ρ^n, $n \in \mathbf{N}$, can be seen as elements of the compact set $\Delta^{m_1} \times \Delta^{m_2} \times \ldots \times \Delta^{m_z}$. Hence, without loss of generality we can assume that $\rho^1, \rho^2, \ldots$ converges to some stationary strategy ρ^*. Now let π^2 be any strategy with finite support for player II. Then π^2 can be seen as a strategy for player II in Γ_n for n sufficiently large. Fix n_0 large enough. It is well-known that playing against a fixed stationary strategy in a stopping stochastic game with finite state and action spaces, Γ_{n_0} for instance, an optimal reply can be found among the stationary strategies. Blackwell [4] proves this for the special class of β-discounted stochastic games, but his proof can be applied to stopping games as well. Hence there is a σ for player II in Γ_{n_0} such that $v(\rho^*, \sigma) \leq v(\rho^*, \pi^2)$, coordinatewise. We also know $v(\rho^n, \sigma) \geq v_n$ for all $n \geq n_0$. Now, using the fact that $\rho \mapsto v(\rho, \sigma)$ is a continuous function on the set of stationary strategies of player I, and using the fact that $v_1, v_2, \ldots$ converges to w, we conclude $v(\rho^*, \sigma) \geq w$, and finally $v(\rho^*, \pi^2) \geq w$. So the lower value of Γ_∞ is at least w.

Take $\epsilon > 0$, $\delta < 0$. Since $v_1, v_2, \ldots$ converges to w there exists $n_1 \in \mathbf{N}$ such that $v_{n_1}(s) \leq w(s) + \epsilon$ if $w(s)$ is finite, and $v_{n_1}(s) \leq \delta$ if $w(s) = -\infty$, for all $s \in S$. Let σ^* be an optimal stationary strategy for player II in Γ_{n_1}. Then again applying the argument that against σ^* player I has a stationary strategy as best answer ([4]), we have that for all strategies π^1 for player I and all $s \in S$ it holds that: $v(s, \pi^1, \sigma^*) \leq v_{n_1}(s) \leq w(s) + \epsilon$ if $w(s)$ is finite; $v(s, \pi^1, \sigma^*) \leq v_{n_1}(s) \leq \delta$ if $w(s) = -\infty$. So the upper value of Γ_∞ is at most w. Combination of the above arguments proves the theorem. ◇

It is clear that player II needs not possess an optimal stationary strategy (with finite support). A simple example for this is the stochastic game consisting of one $1 \times \infty$ state matrix for which $r(1,1,j)$ equals $-j$ and $p(1,1,j) = 0$ for all $j \in \mathbf{N}$.

In the following theorem we give a necessary and sufficient condition for player II to possess optimal stationary strategies. First we introduce the concept 'critical number' for a semi-infinite stochastic game Γ_∞. The critical number c is defined by $c = min\{n \in \mathbf{N} : v_n = v_\infty\}$, where $min\,\emptyset = \infty$. If c is finite, then Γ_∞ is called an essentially finite game.

Theorem 2.2: A semi-infinite stochastic game Γ_∞ is essentially finite if and only if player II has optimal stationary strategies.

Proof: Suppose $c < \infty$. Let σ^* be an optimal stationary strategy for player II in Γ_c. Playing against σ^* in Γ_∞ player I has a stationary best answer ([4]). For all stationary ρ we have $v(\rho, \sigma^*) \leq v_c = v_\infty$, so for all strategies π^1 for player I in $\Gamma_\infty : v(\pi^1, \sigma^*) \leq v_\infty$. Hence σ^* is an optimal stationary strategy for player II in Γ_∞.

Starting with a stationary optimal strategy $\hat{\sigma}$ for player II in Γ_∞ which only uses the first n_1 columns, it is clear that for all stationary strategies ρ for player I we have $v(\rho, \hat{\sigma}) \leq v_\infty$. Hence $v_{n_1} \leq v_\infty$. But for all $n \in \mathbf{N}$ we also had $v_n \geq v_\infty$. So $v_{n_1} = v_\infty$ and therefore $c \leq n_1 < \infty$. ◇

For essentially finite stochastic games Γ_∞ we look at the relation between sets of optimal stationary strategies in Γ_∞, O^1_∞ and O^2_∞, and the sets of optimal stationary strategies in Γ_n, O^1_n and O^2_n, $n \in \mathbf{N}$. Note that for an essentially finite stochastic game both O^1_n and O^2_n are non-empty for all $n \in \mathbf{N} \cup \{\infty\}$ by Theorems 2.1 and 2.2.

Theorem 2.3: Let Γ_∞ be an essentially finite stochastic game. Then $O^1_\infty = \bigcap_{n \geq c} O^1_n$ and $O^2_\infty = \bigcup_{n \geq c} O^2_n$.

Proof: By the proof of Theorem 2.1 we know that, if some sequence of stationary strategies $\rho^1, \rho^2, \ldots$, with $\rho^n \in O^1_n$ for all $n \in \mathbf{N}$, converges to a stationary strategy ρ^*, then $\rho^* \in O^1_\infty$. So we have $O^1_\infty \supset \bigcap_{n \geq c} O^1_n$. Conversely, if $\rho^* \in O^1_\infty$ then for all stationary strategies σ for player II in Γ_n, $n \geq c$, it holds that $v(\rho^*, \sigma) \geq v_\infty = v_n$. Hence $O^1_\infty \subset \bigcap_{n \geq c} O^1_n$. Combining we have proved the player I part of the theorem.

The player II part follows directly from the proof of Theorem 2.2. ◇

3. The Shapley Equation: As already mentioned in the introduction, for finite two-person stopping stochastic games the value and optimal stationary strategies can be found by solving the Shapley equation $x = Tx$ $(x \in \mathbf{R}^z)$, where

$$(Tx)_s = \text{val} M_s(x) = \text{val}[r(s,i,j) + \sum_{t=1}^{z} p(t|s,i,j) x_t]_{i=1, j=1}^{m_s, n_s}$$

which, in the finite case, has a unique solution.

One can also consider the Shapley equation for semi-infinite stochastic games. The difference is that for these stochastic games the $M_s(x)$ will be semi-infinite matrices and the system $x = Tx$ should be solved over $\mathbf{R}^z_-$. Examining $x = Tx$ $(x \in \mathbf{R}^z_-)$ one quickly notes that $(-\infty, -\infty, \ldots, -\infty)$ will often be a solution, hence unlike in the finite case we may no longer have a unique solution. However, some interesting results can be derived.

Theorem 3.1: The value of the semi-infinite stochastic game, v_∞, is a solution of the Shapley equation.

Proof: For all $x \in \mathbf{R}_-^z$ and all $s \in S$ let $M_s(x)$ denote the semi-infinite matrix game with entries $(r(s,i,j) + \sum_{t=1}^{z} p(t|s,i,j)x_t)$ in $\mathbf{R}_-$, and let $M_s^n(x)$ denote the corresponding n-truncated matrix game for $n \in \mathbf{N}$. We know by the result of Shapley [2] that $v_n(s) = \text{val}M_s^n(v_n)$ for all $s \in S$.

First, suppose $v_\infty \in \mathbf{R}^z$ and take $\epsilon > 0$. There exists $n_0 \in \mathbf{N}$ such that for all $n \geq n_0$:

(i) $|v_\infty(s) - v_n(s)| < \epsilon$ for all $s \in S$, because $lim_{n\to\infty} v_n(s) = v_\infty(s) \in \mathbf{R}$ for all $s \in S$

(ii) $|\text{val}M_s^n(v_n) - \text{val}M_s^n(v_\infty)| < \epsilon$ for all $s \in S$, because payoffs in $M_s^n(v_n)$ differ less than ϵ from corresponding payoffs in $M_s^n(v_\infty)$ for all $s \in S$.

(iii) $|\text{val}M_s^n(v_\infty) - \text{val}M_s(v_\infty)| < \epsilon$ for all $s \in S$, because$lim_{k\to\infty}$ $\text{val}M_s^k(v_\infty) = \text{val}M_s(v_\infty)$.

Combining (i), (ii) and (iii) yields $|v_\infty(s) - \text{val}M_s(v_\infty)| < 3\epsilon$ for all $s \in S$. Since ϵ was arbitrary we have $v_\infty(s) = \text{val}M_s(v_\infty)$ for all $s \in S$.

Second, if $v_\infty \notin \mathbf{R}^z$ then, without loss of generality, there is $k \in \{0, 1, \ldots, z-1\}$ such that $v_\infty(s) = -\infty$ if and only if $s \in \{k+1, k+2, \ldots, z\}$. For all $s \in \{1, 2, \ldots, k\}$ player I can prevent the play to move to any state outside $\{1, 2, \ldots, k\}$, otherwise $v_\infty(s)$ would be $-\infty$ as well. So for $s \in \{1, 2, \ldots, k\}$ player I can restrict to rows i in $M_s(v_\infty)$ for which $p(t|s,i,j) = 0$ for all $t \in \{k+1, k+2, \ldots, z\}$ and all $j \in \mathbf{N}$. Hence, restricting player I's actions this way yields a stochastic game $\widetilde{\Gamma}$ for which play remains in $\{1, 2, \ldots, k\}$ forever. It is obvious that the value of this new stochastic game with state space $\{1, 2, \ldots, k\}$ equals the value of the original stochastic game on those states. So the value of the new stochastic game is finite and as we have just seen we can conclude $v_\infty(s) = \text{val}M_s(v_\infty)$ for $s \in \{1, 2, \ldots, k\}$. If $v_\infty(s) = -\infty$ for some $s \in S$, then this implies that player I cannot force a play to remain within the set of states $\{1, 2, \ldots, k\}$ with probability 1. From this one can conclude that in each row of $M_s(v_\infty)$ some entries equal $-\infty$. Hence $v_\infty(s) = \text{val}M_s(v_\infty)$ for all states (with value $-\infty$ or otherwise). ◇

Theorem 3.2: If $x \in \mathbf{R}^z$ is a solution to the Shapley equation, then $x = v_\infty$.

Proof: Suppose $x \in \mathbf{R}^z$ is a solution to the Shapley equation. Then for all $s \in S : x_s = \text{val}M_s(x) = \text{val}[r(s,i,j) + \sum_{t=1}^{z} p(t|s,i,j)x_t]_{i=1,j=1}^{m_s,\infty}$.

Let ρ_s^* be an optimal mixed action for player I in $M_s(x)$ for all $s \in S$. Then for all stationary strategies σ for player II it holds that $x_s \leq r(s,\rho_s^*,\sigma_s) + \sum_{t=1}^{z} p(t|s,\rho_s^*,\sigma_s)x_t$, where $r(s,\rho_s^*,\sigma_s) = \sum_{i=1}^{m_s}\sum_{j=1}^{\infty} \rho_s^*(i) r(s,i,j)\sigma_s(j)$ and $p(t|s,\rho_s^*,\sigma_s) = \sum_{i=1}^{m_s}\sum_{j=1}^{\infty} \rho_s^*(i) p(t|s,i,j)\sigma_s(j)$. In vector-notation, letting $r(\rho^*,\sigma) = (r(1,\rho_1^*,\sigma_1), \ldots r(z,\rho_z^*,\sigma_z))$ and $P(\rho^*,\sigma)$ the $z \times z$-matrix with $p(t|s,\rho_s^*,\sigma_s)$ in entry (s,t), we get $x \leq r(\rho^*,\sigma) + P(\rho^*,\sigma)x$. This implies $x \leq \sum_{\tau=1}^{n} P^{\tau-1}(\rho^*,\sigma) r(\rho^*,\sigma) + P^n(\rho^*,\sigma)x$ for all $n \in \mathbf{N}$, where $P^\tau(\rho^*,\sigma)$ is the τ-fold product of $P(\rho^*, \sigma)$ and $P^0(\rho^*, \sigma)$ is the identity matrix. Since by the stopping play assumption, $lim_{n\to\infty} P^n(\rho^*, \sigma)x = 0$ we have $x \leq \sum_{\tau=1}^{\infty} P^{\tau-1}(\rho^*, \sigma)r(\rho^*, \sigma) = v(\rho^*, \sigma)$. Hence $x \leq v_\infty$.

Conversely, let σ_s^* be an ϵ-optimal mixed action for player II in $M_s(x)$, for all $s \in S$. Then we have that for all stationary strategies ρ for player I: $x \geq r(\rho, \sigma^*) + P(\rho, \sigma^*)x - \epsilon 1_z$. From this one can derive $x \geq v(\rho, \sigma^*) - \epsilon(1-\alpha)^{-1}1_z$ where $\alpha = max_{s,i} \sum_{t=1}^{z} p(t|s, i, \sigma_s^*) < 1$. Since this can be done for all $\epsilon > 0$, we have $x \geq v_\infty$. ◇

Observe that Theorem 3.2 says, that there is at most one real solution to the Shapley equation, and if there is one, it necessarily equals v_∞.

The following theorem tells us how to find v_∞ among all solutions.

Theorem 3.3: If $x \in \mathbf{R}_-^z$ is a solution to the Shapley equation, then $x \leq v_\infty$.

Proof: Let $x \in \mathbf{R}_-^z$ be a solution. For $s \in S$ with $x_s = -\infty$, it is clear that $x_s \leq v_\infty(s)$. For $s \in S$ with $x_s \in \mathbf{R}$ we have $x_s = \text{val}[r(s,i,j) + \sum_{t=1}^{z} p(t|s,i,j)x_t]_{i=1,j=1}^{m_s\ \infty}$. This implies that for an optimal mixed action ρ_s^* of player I in $M_s(x)$ we have $\sum_{t=1}^{z} p(t|s, \rho_s^*, j)x_t \in \mathbf{R}$ for all $j \in \mathbf{N}$, and hence $p(t|s, \rho_s^*, j) = 0$ for all t with $x_t = -\infty$ and all $j \in \mathbf{N}$. Without loss of generality let $k \in S$ be such that $x_t = -\infty$ for $t > k$ and $x_t \in \mathbf{R}$ for $t \leq k$. Then, for $s \in \tilde{S} := \{1, 2, \ldots, k\}$, we have that for all stationary strategies σ for player II: $x_s \leq r(s, \rho_s^*, \sigma_s) + \sum_{t=1}^{k} p(t|s, \rho_s^*, \sigma_s)x_t$. In vector notation over $\tilde{S} : x \leq r(\rho^*, \sigma) + P(\rho^*, \sigma)x$. Now, using $x_s \in \mathbf{R}$ for all $s \in \tilde{S}$, iteration of the above inequality gives $x \leq v(\rho^*, \sigma)$. Hence we have shown that $x_s \leq v_\infty(s)$ for all $s \in \tilde{S}$. ◇

Theorem 3.4: A stationary strategy ρ^* for player I is optimal if and only if for each $s \in S$ the mixed action ρ_s^* is optimal in the semi-infinite matrix game $M_s(v_\infty)$.

Proof: If for each $s \in S$ the mixed action ρ_s^* is optimal in $M_s(v_\infty)$, then, by the proof of Theorem 3.3, it follows that ρ^* is an optimal stationary strategy in the stochastic game.

Conversely, let ρ^* be an optimal stationary strategy for player I. Suppose, for some $s \in S$, that ρ_s^* is not optimal in $M_s(v_\infty)$, then, for some $j \in \mathbf{N}$ and some $\epsilon > 0$, it holds that $r(s, \rho_s^*, j) + \sum_{t=1}^{z} p(t|s, \rho_s^*, j)v_\infty(t) < \text{val}M_s(v_\infty) - \epsilon = v_\infty(s) - \epsilon$. Hence, if for the stochastic game starting in s, player II initially chooses column j and from then on uses an $\epsilon\delta$-optimal stationary strategy $\pi_{\epsilon\delta}^2$ against ρ^*, then for this strategy $(j, \pi_{\epsilon\delta}^2)$ of player II we have $v(s, \rho^*, (j, \pi_{\epsilon\delta}^2)) \leq r(s, \rho_s^*, j) + \sum_{t=1}^{z} p(t|s, \rho_s^*, j)(v_\infty(t) + \epsilon) < v_\infty(s)$. This contradicts the optimality of ρ^* in the stochastic game. ◇

For player II near-optimal stationary strategies cannot directly be found as extensions of mixed actions in the matrix games $M_s(v_\infty)$. Think for instance of the stochastic game consisting of one $1 \times \infty$ state with $r(1,1,j) = -j$ and $p(1|1,1,j) = \frac{1}{2}$ for all $j \in \mathbf{N}$. It is obvious that $v_\infty = -\infty$. Hence $M_1(v_\infty)$ is the $1 \times \infty$ matrix game with payoff $-\infty$ in all entries. Clearly, the mixed action: "choose column j" is an optimal mixed action for player II in the matrix game $M_1(v_\infty)$, but does not give any information about near-optimal strategies in the stochastic game. This is due to the fact that a state s, with $v_\infty(s) = -\infty$, is either 'directly good' or 'indirectly good' for player II. The following two lemmas illustrate this phenomenon.

Lemma 3.5: If $v_\infty \notin \mathbf{R}^z$ and $\delta < 0$, then there is at least one state $s \in S$ in which player II has a mixed action q_s, such that the expected direct payoff in state s is at most δ, if player II uses the mixed action q_s.

Proof: If not, then player I has a stationary strategy ρ such that all expected direct payoffs are at least δ, and hence $v \in \mathbf{R}^z$. Contradiction. $\diamond$

Let $D \subset S$, be the set of states for which player II has, for all $\delta < 0$, a mixed action to keep the expected direct payoff below δ. If $v_\infty \notin \mathbf{R}^z$, then D is non-empty by Lemma 3.5. D is called the set of states that are directly good for player II.

Lemma 3.6: If $v_\infty \notin \mathbf{R}^z$, then let $ID = \{s \in S \backslash D : v_\infty(s) = -\infty\}$ and suppose $ID \neq \emptyset$. Then player II has mixed actions q_s, $s \in ID$, such that for all stationary strategies σ for player II with $\sigma_s = q_s$ for $s \in ID$, and all stationary strategies for player I, any play started in some state in ID will reach the set of states D with positive probability.

Proof: If not, then $v_\infty(s) \in \mathbf{R}$ for some $s \in ID$. Contradiction. $\diamond$

ID may be empty and is called the set of indirectly good states for player II. Player II can construct a near-optimal strategy in the following way. In directly good states (belonging to D), player II ensures that the expected directed payoff is low enough; in indirectly good states (belonging to ID) player II ensures that the transitions will lead to D; in the other states, with finite value, player II has to consider direct payoffs as well as transitions.

Theorem 3.7: A near-optimal stationary strategy for player II can be constructed by taking mixed actions q_s, $s \in S$, which are near-optimal in the matrix game $[r(s,i,j)]_{i=1,j=1}^{m_s,\infty}$ for $s \in D$, in the matrix game $[r(s,i,j) + \sum_{t=1}^{z} p(t|s,i,j)v_\infty(t)]_{i=1,j=1}^{m_s,\infty}$ for $s \in S\backslash(D \cup ID)$, and for $s \in ID(k)$ to be taken near-optimal in the matrix game $\left[\sum_{t=1}^{z} p(t|s,i,j)w^k(t)\right]_{i=1,j=1}^{m_s,\infty}$, where $ID(k) = \{s \in ID$; player I cannot avoid the play to move from s to D within k stages, with positive probability}, and where $w^k(t) = -\infty$ for $t \in D \cup \left(\bigcup_{\ell=1}^{k} ID(\ell)\right)$ and $w^k(t) = 0$ for other $t, k \in \{0, 1, 2, \ldots, z-1\}$.

Proof: Follows straightforward from Lemma 3.5, Lemma 3.6 and part of the proof of Theorem 3.2. ◇

For stochastic games with finite state and action spaces and any $x \in \mathbb{R}^z$ it holds that $lim_{n\to\infty} T^n(x) = v_\infty$, by the facts that T is a contraction operator and v_∞ is its unique fixed point. Since for semi-infinite stochastic games the value may be $-\infty$ in some coordinates, it is not directly clear whether the above method can be used to find the value. The following theorem answers this problem.

Theorem 3.8: For all $x \in \mathbb{R}^z$, $lim_{n\to\infty} T^n(x) = v_\infty$.

Proof: $D = \{s \in S : (T(x))_s = -\infty\}$, $ID = \{s \in S \backslash D : (T^z(x))_s = -\infty\}$. If $S = D \cup ID$, we are finished, else we continue in the following way. For the remaining states $(T^z(x))_S \in \mathbb{R}$ and hence $(T^n(x))_s \in \mathbb{R}$ for all these remaining states. Let $R := S\backslash(D \cup ID)$. Without loss of generality suppose $R = \{1, 2, \ldots, k\}$. For $y = (y_1, y_2, \ldots, y_k, -\infty, \ldots, -\infty) \in \mathbb{R}_-^z$, with $\hat{y} = (y_1, y_2, \ldots, y_k) \in \mathbb{R}^k$, and $s \in R$ it holds that $(T(y))_s = (\widehat{T}(\hat{y}))_s$ where

$$(\widehat{T}(\hat{y}))_s = \text{val}\left[r(s,i,j) + \sum_{t=1}^{k} p(t|s,i,j)\hat{y}_t\right]_{i=1,j=1}^{m_s(R)\infty} =: \text{val}\widehat{M}_s(\hat{y}),$$

where, without loss of generality, $\{1, 2, \ldots, m_s(R)\}$ is the set of rows of $M_s(y)$ for which all entries are reals, and where $\widehat{M}_s(\hat{y})$ is the matrix consisting of those rows.

For $a, b \in \mathbb{R}^k$, $a \neq b$, it holds that $\|\widehat{T}(a) - \widehat{T}(b)\| < \|a - b\|$, where $\|x\| = max_s |x_s|$. Hence $\widehat{T}$ is a contraction map on $\mathbb{R}^k$ and has a unique fixed point which necessarily equals $lim_{n\to\infty} \widehat{T}^n(\hat{x})$ for all $\hat{x} \in \mathbb{R}^k$. For $s \in R$, player I's optimal stationary strategies never lead the play to $D \cup ID$ and hence, starting in R, the stochastic game can be seen as a stochastic game with R as set of states: $\widehat{\Gamma}$. Simply restrict player I's action sets as is done above. Applying Theorems 3.1 and 3.2 to $\widehat{\Gamma}$ we have that $v_\infty(s) = lim_{n\to\infty}(\widehat{T}^n(\hat{x}))_s = lim_{n\to\infty}(T^n(x))_s$ for $s \in R$. For $s \in D \cup ID$ and $n \geq z : (T^n(x))_s = -\infty = v_\infty(s)$. Hence $lim_{n\to\infty} T^n(x) = v_\infty$. ◇

Observe that $lim_{n\to\infty} T^n(x)$ need not equal v_∞ if we start with $x \in \mathbb{R}_-^z \backslash \mathbb{R}^z$. This is illustrated by the following example:

Take $S = \{1,2,3\}$; $A_1 = \{1,2\}$, $A_2 = A_3 = \{1\}$, $B_1 = B_2 = B_3 = \mathbf{N}$; $r(1,1,j) = r(2,1,j) = r(3,1,j) = 0$ for all $j \in \mathbf{N}$, $r(1,2,j) = 2$ for all $j \in \mathbf{N}$; $p(1|1,1,j) = p(3|1,2,j) = p(1|2,1,j) = p(3|3,1,j) = \frac{1}{2}$, $p(t|s,i,j) = 0$ else. Then, starting with $x = (0,-\infty,-\infty)$, it follows that $T^n(x) = (0,0,-\infty)$ for all $n \in \mathbf{N}$, where as $v = (2,1,0)$.

Closing Remarks: In the literature β-discounted stochastic games have been studied extensively. Those are stochastic games that continue infinitely, since there one takes in the definition of the game that $\sum_{t=1}^{z} p(t|s,i,j) = 1$ for all s,i,j. The players however, discount future payoffs by some factor $\beta \in (0,1)$. So for a pair of strategies (π_1,π_2) and an initial state s the β-discounted reward to player I is given by $\sum_{\tau=1}^{\infty} \beta^{\tau-1} R_{s\pi_1\pi_2}(\tau)$.

Notice that a β-discounted stochastic game can be seen as a stopping stochastic game if we relate transition probabilities $p(t|s,i,j)$ in the β-discounted non-stopping game with transition probabilities $\beta p(t|s,i,j)$ in the stopping game.

Hence all the results derived in this paper for stopping stochastic games, also hold for β-discounted stochastic games.

In section 2 we made the assumption $sup_{s,i,j}\{\sum_{t\in S} p(t|s,i,j)\} < 1$. This condition played no role until Lemma 3.5.

References

[1] *Von Neumann, J.* (1928). Zur Theorie der Gesellschaftsspiele, *Mathematische Annalen*, **100**, 295-320.

[2] *Shapley, L.S.* (1953). Stochastic Games. *Proceedings of the National Academy of Sciences, U.S.A.*, **39**, 1095-1100.

[3] *Tijs, S.H.* (1975). Semi-Infinite and Infinite Matrix Games and Bimatrix Games. Ph.D. Thesis, University of Nijmegen.

[4] *Blackwell, D.* (1962). Discrete Dynamic Programming. *Annals of Mathematical Statistics*, **33**, 719-726.

Semi-infinite Stochastic Games

S. Sinha
Indian Statistical Institute
7 S.J. Sansanwal Marg
New Delhi 110016
India.

F. Thuijsman
Department of Mathematics
Faculty of General Sciences
University of Limburg
P.O. Box 616
6200 MD Maastricht
The Netherlands.

S.H.Tijs
Department of Mathematics
Catholic University
Toernooiveld,
6525 ED Nijmegen
The Netherlands