Positive Zero-Sum Stochastic Games with Countable State and Action Spaces

János Flesch¹ · Arkadi Predtetchinski² · William Sudderth³

Published online: 1 November 2018
© The Author(s) 2018

Abstract

A positive zero-sum stochastic game with countable state and action spaces is shown to have a value if, at every state, at least one player has a finite action space. The proof uses transfinite algorithms to calculate the upper and lower values of the game. We also investigate the existence of (ϵ) -optimal strategies in the classes of stationary and Markov strategies.

Keywords Zero-sum stochastic game · Value of the game · Optimal strategy · Markov strategy · Fixed point · Tarski fixed point theorem · Transfinite algorithm

Mathematics Subject Classification 91A15 · 91A05 · 60J10 · 60G42

We would like to thank Andrés Perea for pointing out a connection between the auxiliary one-shot games in our paper and non-standard analysis. The support for the visit of William Sudderth to Maastricht provided by a travel Grant Number 040.11.495 of the Netherlands Organisation for Scientific Research (NWO) is acknowledged.

✉ Arkadi Predtetchinski
a.predtetchinski@maastrichtuniversity.nl

János Flesch
j.flesch@maastrichtuniversity.nl

William Sudderth
bill@stat.umn.edu

¹ Department of Quantitative Economics, Maastricht University, P.O.Box 616, 6200 MD Maastricht, The Netherlands

² Department of Economics, Maastricht University, P.O.Box 616, 6200 MD Maastricht, The Netherlands

³ School of Statistics, University of Minnesota, Minneapolis, MN 55455, USA

1 Introduction

A positive zero-sum stochastic game is a two-person dynamic game played in stages $t = 0, 1, \dots$. At each stage t of the game the players simultaneously select actions from their action sets at the current state. These actions together with the current state determine a nonnegative reward and the distribution of the next state. The payoff from player 2 to player 1 is the infinite sum of the rewards taken over all the stages. We assume that the state space and the action sets of the two players at each state are countable.

As is well-known, such a game does not always admit a value. A typical example is the following: There is a state with action set $\mathbb{N} = \{0, 1, \dots\}$ for each player in which the reward is 1 if player 1's action is greater than player 2's action and is 0 otherwise. From this state, regardless of the chosen actions, the transition is to an absorbing state where the reward is zero. Thus player 1 wins in this game if he chooses the larger action and loses if not.

To find conditions for the existence of the value, we construct transfinite algorithms that output the upper value and the lower value of the game (cf. Theorems 6 and 9). If at every state at least one player has a finite action set, then the algorithms for the upper and lower values give the same answer. So, in this case, the game has a value. Also, in this case, player 2 has arbitrarily good Markov strategies (cf. Theorem 11).

If the action set for player 2 is finite at every state, then the algorithm for the value simplifies and becomes the limit of the sequence of values of the n -stage games. In this simpler case, player 2 has an optimal stationary strategy (cf. Theorem 12).

Related Literature For an overview on the literature of positive stochastic games, we refer to the recent survey by Jaśkiewicz and Nowak [7]. In particular, we refer to Maitra and Parthasarathy [9], Parthasarathy [15,16], Frid [6], and Nowak [13]. Transfinite algorithms were used previously for example by Blackwell [1] for G_δ games and by Maitra and Sudderth [10,11] for stochastic games with limsup payoff.

Organization of the Paper The next section presents the basic definitions needed for our model. Section 3 has some preliminary results for one-shot games with unbounded payoffs and infinite action sets. Section 4 introduces the one-shot games associated with a stochastic game, defines operators corresponding to the upper and lower values of the one-shot games, and shows that, by iterating these operators up to some countable ordinal, the upper and lower values of the stochastic game are attained. The results of Sect. 4 are applied in Sect. 5 to obtain our main theorems on the value and (ϵ) -optimal strategies for the stochastic game.

2 The Model

Positive Zero-Sum Stochastic Games We consider positive zero-sum stochastic games with countable state and action spaces. Such a game is played by two players, and is given by (1) a nonempty and countable state space S with a distinguished element $s_0 \in S$, called the initial state, (2) for each state $s \in S$, nonempty and countable action spaces $A(s)$ and $B(s)$ for player 1 and respectively player 2, (3) for each state $s \in S$ and actions $a \in A(s)$, $b \in B(s)$, a probability mea-

sure $p(s, a, b) = p(s'|s, a, b)_{s' \in S}$ on S , and (4) a non-negative reward function $r : Z \rightarrow [0, \infty)$, where $Z = \{(s, a, b) | s \in S, a \in A(s), b \in B(s)\}$. A positive zero-sum stochastic game is denoted by Γ . Whenever we need to emphasize the initial state s_0 of the game, we write $\Gamma(s_0)$.

The game is played at stages in $\mathbb{N} = \{0, 1, \dots\}$ and begins in the initial state $s_0 \in S$. At every stage $t \in \mathbb{N}$, the play is in a state $s_t \in S$. In this state, player 1 chooses an action $a_t \in A(s_t)$ and simultaneously player 2 chooses an action $b_t \in B(s_t)$, yielding a triple $z_t = (s_t, a_t, b_t) \in Z$. Then, player 1 receives reward $r(z_t)$ from player 2, and state s_{t+1} is drawn in accordance with the probability measure $p(z_t)$. Thus, play of the game induces an infinite sequence $(z_0, z_1, \dots)$ in Z . The payoff is

$$u(z_0, z_1, \dots) = \sum_{t=0}^{\infty} r(z_t),$$

which is an element of $[0, \infty]$. The payoff is paid by player 2 to player 1. Player 1's objective is to maximize the expected payoff given by u , and player 2's objective is to minimize it.

Strategies The set of histories at stage t is denoted by H_t . Thus, $H_0 = S$ and $H_t = Z^t \times S$ for every stage $t \geq 1$. Let $H = \cup_{t \in \mathbb{N}} H_t$ denote the set of all histories. For each history h , let s_h denote the final state in h .

A mixed action for player 1 in state $s \in S$ is a probability measure $x(s)$ on $A(s)$. Similarly, a mixed action for player 2 in state $s \in S$ is a probability measure $y(s)$ on $B(s)$. The respective sets of mixed actions in state s are denoted by $\Delta(A(s))$ and $\Delta(B(s))$.

A strategy for player 1 is a map π that to each history $h \in H$ assigns a mixed action $\pi(h) \in \Delta(A(s_h))$. Similarly, a strategy for player 2 is a map σ that to each history $h \in H$ assigns a mixed action $\sigma(h) \in \Delta(B(s_h))$. The set of strategies is denoted by Π for player 1 and by Σ for player 2. A strategy is called pure if it places probability 1 on one action after each history.

A strategy is called stationary if the assigned mixed actions only depend on the history through its final state. Thus, a stationary strategy for player 1 can be seen as an element x of $\times_{s \in S} \Delta(A(s))$. Similarly, a stationary strategy for player 2 can be seen as an element y of $\times_{s \in S} \Delta(B(s))$. A strategy is called Markov if the assigned mixed actions only depend on the history through its final state and the current stage.

An initial state $s \in S$ and a pair of strategies $(\pi, \sigma) \in \Pi \times \Sigma$ determine a distribution $\mathbb{P}_{s, \pi, \sigma}$ for the stochastic process $z_0, z_1, \dots$. The corresponding expectation operator is written $\mathbb{E}_{s, \pi, \sigma}$ and the expected payoff is denoted by $u(s, \pi, \sigma) = \mathbb{E}_{s, \pi, \sigma}[\sum_{t=0}^{\infty} r(z_t)]$.

Value and Optimality The lower value of the game $\Gamma(s)$ with the initial state $s \in S$, denoted by $\alpha(s)$, is defined as

$$\alpha(s) = \sup_{\pi \in \Pi} \inf_{\sigma \in \Sigma} u(s, \pi, \sigma).$$

Similarly, the upper value of the game $\Gamma(s)$ with the initial state $s \in S$, denoted by $\beta(s)$, is defined as

$$\beta(s) = \inf_{\sigma \in \Sigma} \sup_{\pi \in \Pi} u(s, \pi, \sigma).$$

The inequality $\alpha(s) \leq \beta(s)$ always holds. If $\alpha(s) = \beta(s)$, then this quantity is called the value of the game for initial state s and it is denoted by $v(s)$. Then, for $\epsilon \geq 0$, a strategy $\pi \in \Pi$ for player 1 is called ϵ -optimal for initial state s if $u(s, \pi, \sigma) \geq v(s) - \epsilon$ for every strategy $\sigma \in \Sigma$ for player 2. Similarly, a strategy $\sigma \in \Sigma$ for player 2 is called ϵ -optimal for initial state s if $u(s, \pi, \sigma) \leq v(s) + \epsilon$ for every strategy $\pi \in \Pi$ for player 1.

If the value exists for every initial state, then for every $\epsilon > 0$, each player has a strategy that is ϵ -optimal for every initial state. We call these strategies ϵ -optimal. A 0-optimal strategy is simply called optimal.

As we observed in the introduction, the value in general does not exist and a major objective here is to find interesting conditions for its existence.

Example 1 It is known that in a positive zero-sum stochastic game, even if the state and action spaces are finite, player 1 may have no optimal strategy, see Kumar and Shiau [8], Maitra and Sudderth [12], and Jaśkiewicz and Nowak [7]. Consider the following game.

	<i>L</i>	<i>R</i>
<i>T</i>	0	1
<i>B</i>	1	0

In this game, there is only one non-trivial state, shown in the figure. In this state, player 1's actions are T and B , and player 2's actions are L and R . The rewards for the corresponding action combinations are given in the matrix. The transitions are as follows: if action combination (T, L) is chosen then the state remains the same, but after any other action combination transition occurs to an absorbing state where the reward is equal to 0. Clearly, player 1 can guarantee an expected payoff of $1 - \epsilon$, for any $\epsilon \in (0, 1)$, by playing the stationary strategy $(1 - \epsilon, \epsilon)$. Thus the value is equal to 1 for the non-trivial state. Yet, player 1 has no optimal strategy. $\square$

3 One-Shot Games

In this section we present some results for one-shot games that are crucial for our study of positive zero-sum stochastic games.

We consider positive zero-sum one-shot games $G = (A, B, f)$ where A is a nonempty and countable action set for player 1, B is a nonempty and countable action set for player 2, and $f : A \times B \mapsto [0, \infty]$ is the payoff function.

The notation x and y will be used for mixed actions for players 1 and 2 respectively. Also we set $u(x, y) = \sum_a \sum_b f(a, b)x(a)y(b)$ for the expected payoff under a pair (x, y) of mixed actions. When calculating expected payoffs, we make use of the usual convention in measure theory that $\infty \cdot 0 = 0$. So, if an action combination with payoff ∞ is chosen with probability 0, then it makes no contribution to the expected payoff.

Here is a slight generalization of the usual von Neumann theorem (cf. Tijs [20], or exercise 18.3 on p. 201 in Maitra and Sudderth [12]).

Theorem 1 Consider a positive zero-sum one-shot game $G = (A, B, f)$. If f is bounded and either A or B is finite, then G has a value v . Moreover, if A , respectively B , is finite, then player 1, respectively player 2, has an optimal strategy.

The following example shows that player 1 may have no optimal strategy, even if both action sets are finite.

Example 2

	L	R
T	2	∞
B	3	1

This game has value 3. Indeed, for every $\epsilon \in (0, 1)$, player 1 can guarantee an expected payoff of $3 - \epsilon$ by playing action T with probability ϵ and action B with probability $1 - \epsilon$, and player 2 can also guarantee that the expected payoff is not more than 3 by playing action L . Clearly, player 1 has no optimal strategy. Andrés Perea pointed out that this game could also be seen as

	L	R
T	2δ	1
B	3δ	δ

where δ is an infinitesimal as in nonstandard analysis. $\square$

If both action sets are infinite, the value does not always exist. Probably the most well known example is the following game, which is essentially the game mentioned in the introduction.

Example 3 Let the action sets be $A = B = \mathbb{N}$ and the payoff for actions $a \in A$ and $b \in B$ be equal to 1 if $a > b$ and 0 otherwise. That is, in this game player 1 wins if his action is greater than the action chosen by player 2. This game has no value. The lower value of this game is 0 and the upper value is 1. $\square$

Perhaps the next two theorems are also known, but we do not have references for them.

Theorem 2 Consider a positive zero-sum one-shot game $G = (A, B, f)$. If B is finite, then G has a value $v \in [0, \infty]$ and player 2 has an optimal strategy.

Proof Let v_n be the value of the game $G_n = (A, B, f_n)$, where $f_n = \min(f, n)$ for $n = 1, 2, \dots$. Clearly, $0 \leq v_n \leq v_{n+1}$ for all n . Set $v = \lim v_n$, and write $u_n(x, y)$ for the expected payoff from x, y in G_n . Let α and β be the lower value and respectively the upper value of the game G .

We prove the following two claims:

Claim 1: $\alpha \geq v$.

Claim 2: player 2 has a strategy $y^* \in \Delta(B)$ such that for every $x \in \Delta(A)$ we have $u(x, y^*) \leq v$.

Once these two claims are proven, the statement of the theorem follows. Indeed, claim 2 implies that $\beta \leq v$, and hence by claim 1 we have $v = \alpha = \beta$. So, the value of the game is v , and y^* is an optimal strategy for player 2.

First we prove claim 1. Notice that

$$\alpha = \sup_x \inf_y u(x, y) \geq \sup_x \inf_y u_n(x, y) = v_n.$$

By taking the limit when n tends to infinity, we obtain $\alpha \geq v$.

Now we prove claim 2. There is no harm in assuming that $v < \infty$. Then, for every n , $v_n \leq v < \infty$.

By Theorem 1, player 2 has an optimal strategy $y_n \in \Delta(B)$ in the game G_n for every n . Since $\Delta(B)$ is a compact subset of a finite dimensional Euclidean space, the sequence y_n has a subsequence that converges to some $y^* \in \Delta(B)$. With a slight abuse of notation we write y_n for this convergent subsequence.

Fix a strategy $x \in \Delta(A)$ for player 1. For every $b \in B$ and $n \in \mathbb{N}$, let $g(b) = \sum_a f(a, b)x(a)$ and $g_n(b) = \sum_a f_n(a, b)x(a)$. Notice that, by the monotone convergence theorem, for every b we have $g(b) = \lim_n g_n(b)$.

Let B' denote the set of those actions $b \in B$ for which $g(b) = \infty$. Since for every n it holds that

$$u_n(x, y_n) = \sum_b g_n(b)y_n(b) \leq v_n \leq v < \infty,$$

for every $b \in B'$ we must have $y^*(b) = \lim_n y_n(b) = 0$. Thus, in accordance with our convention, for every $b \in B'$

$$g(b)y^*(b) = \infty \cdot 0 = 0.$$

Therefore,

$$\begin{aligned} u(x, y^*) &= \sum_b g(b)y^*(b) \\ &= \sum_{b \in B-B'} g(b)y^*(b) \\ &= \lim_n \sum_{b \in B-B'} g_n(b)y_n(b) \\ &\leq \lim_n \sum_{b \in B-B'} g_n(b)y_n(b) + \limsup_n \sum_{b \in B'} g_n(b)y_n(b) \\ &= \limsup_n u_n(x, y_n) \\ &\leq \lim_n v_n \\ &= v. \end{aligned}$$

For the third equality, we used the assumption that B is finite. This completes the proof of claim 2. $\square$

Theorem 2 is related to Theorem 5.1 of Nowak [13]. Nowak's result is for real-valued payoff functions in a Borel measurable setting.

Theorem 3 *Consider a positive zero-sum one-shot game $G = (A, B, f)$. If A is finite, then the game G has a value $v \in [0, \infty]$.*

Proof We can assume without loss of generality that $B = \mathbb{N}$.

Let G_n denote the game where player 1's action set is A , player 2's action set is $B_n = \{0, \dots, n\}$, and the payoff function is the restriction of f to $A \times B_n$. By Theorem 2 the game G_n has a value, say v_n . Of course, we have $v_{n+1} \leq v_n$. Let $v = \lim v_n$. We argue that v is the value of G .

Let α and β be the lower value and respectively the upper value of the game G . It is obvious that $\beta \leq v_n$ for every $n \in \mathbb{N}$, since player 2 can always ignore all but finitely many of his actions. Hence $\beta \leq v$. We show that $v \leq \alpha$.

Take an $R < v$. As $R < v_n$, player 1 has a strategy x_n that guarantees him the payoff of R in the game G_n , that is $R \leq u(x_n, y)$ for each $y \in \Delta(B_n)$. Since x_n lies in the compact set $\Delta(A)$, it has a subsequence converging to a point in $\Delta(A)$. By passing to the subsequence, we can assume that x_n converges to some $x_* \in \Delta(A)$. Fix $\rho \in (0, 1)$ and define player 1's strategy x as follows: with probability ρ , randomize uniformly on A ; with probability $(1 - \rho)$, randomize according to the distribution x_* . More precisely, we let $x(a) = \rho \cdot |A|^{-1} + (1 - \rho) \cdot x_*(a)$ for each $a \in A$.

We argue that $(1 - \rho) \cdot R \leq u(x, y)$ for each $y \in \Delta(B)$.

If $u(x, y) = \infty$, there is nothing to prove, so will assume that $u(x, y) < \infty$. Since x places positive probability on each action in A , this implies that $u(a, y) < \infty$ for each $a \in A$.

Suppose first that the distribution y is finitely supported, i.e. that y is an element of $\Delta(B_n)$ for some natural n . Since $u(a, y) < \infty$ for each $a \in A$, the sum $u(x_m, y) = \sum_{a \in A} x_m(a)u(a, y)$ converges to $u(x_*, y)$. For each $m \geq n$, since y is a feasible strategy for player 2 in the game G_m , we have $R \leq u(x_m, y)$. Taking the limit we obtain $R \leq u(x_*, y)$. And hence $(1 - \rho) \cdot R \leq (1 - \rho) \cdot u(x_*, y) \leq u(x, y)$, as claimed.

Now take an arbitrary $y \in \Delta(B)$. Let $N \in B$ be any action such that $y(N) > 0$. For each $n \geq N$ define $y_n \in \Delta(B_n)$ by letting

$$y_n(i) = \begin{cases} \frac{y(i)}{y(0) + \dots + y(n)}, & \text{if } i \leq n \\ 0, & \text{otherwise.} \end{cases}$$

Since y_n is finitely supported, it follows by the previous paragraph that $(1 - \rho) \cdot R \leq u(x, y_n)$. On the other hand the payoff

$$u(x, y_n) = \frac{1}{y(0) + \dots + y(n)} \sum_{i \leq n} y(i)u(x, i)$$

converges to $u(x, y)$ as $n \rightarrow \infty$. This is because the fraction in the above expression converges to 1, while the sum converges to $u(x, y)$ by the monotone convergence theorem. We conclude that $(1 - \rho) \cdot R \leq u(x, y)$, as desired.

This shows that $(1 - \rho) \cdot R \leq \alpha$. Since $\rho \in (0, 1)$ and $R < v$ are arbitrary, we obtain $v \leq \alpha$. $\square$

4 One-Shot Operators and Their Fixed Points

We now return to our study of positive zero-sum stochastic games. Let Γ be a game as in Sect. 2. Recall that $\alpha(s)$ and $\beta(s)$ are the lower and the upper values of the game $\Gamma(s)$ with the initial state $s \in S$. Let $\mathcal{L}^+$ be the space of all functions from S to $[0, \infty]$. We write α and β to denote the functions $s \mapsto \alpha(s)$ and $s \mapsto \beta(s)$, respectively. Both of these are elements of $\mathcal{L}^+$. The functions α and β , referred to simply as the lower and the upper values of Γ , will be at the center of our attention in this section.

For a function $f \in \mathcal{L}^+$ and a state $s \in S$, we define the one-shot game $M_f(s)$ in which the action sets are $A(s)$ for player 1 and $B(s)$ for player 2, and the payoff for actions $a \in A(s)$ and $b \in B(s)$ is

$$r(s, a, b) + \sum_{t \in S} f(t) p(t|s, a, b).$$

The game $M_f(s)$ does not have a value in general, but if either $A(s)$ or $B(s)$ is finite, then Theorems 2 and 3 guarantee that $M_f(s)$ admits a value.

We define the one-shot operator $\mathcal{A} : \mathcal{L}^+ \rightarrow \mathcal{L}^+$ by letting $\mathcal{A}f(s)$ be the lower value of the one-shot game $M_f(s)$, for a function $f \in \mathcal{L}^+$ and a state $s \in S$. That is,

$$\begin{aligned} \mathcal{A}f(s) = & \sup_{x \in \Delta(A(s))} \inf_{y \in \Delta(B(s))} \sum_{a \in A(s), b \in B(s)} \\ & \left[r(s, a, b) + \sum_{s' \in S} f(s') p(s'|s, a, b) \right] x(a) y(b). \end{aligned}$$

Similarly, we define the one-shot operator $\mathcal{B} : \mathcal{L}^+ \rightarrow \mathcal{L}^+$ by letting $\mathcal{B}f(s)$ be the upper value of the one-shot game $M_f(s)$, for each function $f \in \mathcal{L}^+$ and state $s \in S$. That is,

$$\begin{aligned} \mathcal{B}f(s) = & \inf_{y \in \Delta(B(s))} \sup_{x \in \Delta(A(s))} \sum_{a \in A(s), b \in B(s)} \\ & \left[r(s, a, b) + \sum_{s' \in S} f(s') p(s'|s, a, b) \right] x(a) y(b). \end{aligned}$$

It is clear that the operators $\mathcal{A}$ and $\mathcal{B}$ are monotone. That is, if for some $f, g \in \mathcal{L}^+$ we have $f \leq g$ then we also have $\mathcal{A}f \leq \mathcal{A}g$ and $\mathcal{B}f \leq \mathcal{B}g$.

We call a function $f \in \mathcal{L}^+$ a fixed point of the operator $\mathcal{A}$ if $\mathcal{A}f = f$. We define fixed points similarly for $\mathcal{B}$. Note that if f is a fixed point of either operator, then $f + c$ is also a fixed point for every $c \in \mathbb{R}$ as long as $f + c$ is nonnegative (so that it belongs to $\mathcal{L}^+$).

4.1 The Operator $\mathcal{A}$ for the Lower Value

Lemma 4 *The lower value α of Γ is a fixed point of the operator $\mathcal{A}$.*

Proof The proof that $\mathcal{A}\alpha = \alpha$ is presented in two steps. Fix $s \in S$.

Step 1 We prove $\mathcal{A}\alpha(s) \leq \alpha(s)$.

Let $R < \mathcal{A}\alpha(s)$ and $\epsilon > 0$. Since $\mathcal{A}\alpha(s)$ is the lower value of the one-shot game $M_\alpha(s)$, there exists $x^* \in \Delta(A(s))$ such that for all $y \in \Delta(B(s))$ we have

$$\sum_{a,b} \left[r(s, a, b) + \sum_{s'} \alpha(s') p(s'|s, a, b) \right] x^*(a) y(b) > R.$$

It could happen that some of the terms $\alpha(s')$ are infinite. By the monotone convergence theorem, there exists $n > 0$ so that

$$\sum_{a,b} \left[r(s, a, b) + \sum_{s'} \min\{\alpha(s'), n\} p(s'|s, a, b) \right] x^*(a) y(b) > R - \epsilon.$$

Use the definition of α to choose a strategy π for player 1 such that, for all states s' and all strategies σ for player 2, $u(s', \pi, \sigma) \geq \min\{\alpha(s'), n\} - \epsilon$. Let π^* be a strategy for player 1 that begins at state s with action x^* and continues from the next state with the strategy π . It is easy to check that $u(s, \pi^*, \sigma) \geq R - 2\epsilon$ for every strategy σ of player 2. This implies that $\alpha(s) \geq R - 2\epsilon$. Since this holds for all $R < \mathcal{A}\alpha(s)$ and $\epsilon > 0$, we have shown that $\alpha(s) \geq \mathcal{A}\alpha(s)$.

Step 2 We prove $\mathcal{A}\alpha(s) \geq \alpha(s)$.

Let $R < \alpha(s)$. Since $\alpha(s)$ is the lower value of the stochastic game $\Gamma(s)$, there exists a strategy π^* for player 1 such that $u(s, \pi^*, \sigma) \geq R$ for all strategies σ for player 2. Let $x^* = \pi^*(s)$, that is, the initial mixed action prescribed by π^* when starting in state s .

Let $y \in \Delta(B(s))$ and $\epsilon > 0$ be arbitrary. By the definition of α , for each state s' there is a strategy $\sigma_{s'}$ such that $u(s', \pi^*, \sigma_{s'}) \leq \alpha(s') + \epsilon$. Let σ' be the strategy that plays y in stage 0, and as of stage 1 coincides with the strategy σ_{s_1} . Then

$$\begin{aligned} R &\leq u(s, \pi^*, \sigma') \\ &\leq \sum_{a,b} \left[r(s, a, b) + \sum_{s'} (\alpha(s') + \epsilon) p(s'|s, a, b) \right] x^*(a) y(b) \\ &= \sum_{a,b} \left[r(s, a, b) + \sum_{s'} \alpha(s') p(s'|s, a, b) \right] x^*(a) y(b) + \epsilon. \end{aligned}$$

As $\epsilon > 0$ was arbitrary, we conclude

$$R \leq \sum_{a,b} \left[r(s, a, b) + \sum_{s'} \alpha(s') p(s'|s, a, b) \right] x^*(a) y(b). \quad (4.1)$$

Since $y \in \Delta(B(s))$ was arbitrary too, $R \leq \mathcal{A}\alpha(s)$. Because this holds for all $R < \alpha(s)$, we have shown that $\alpha(s) \leq \mathcal{A}\alpha(s)$.

The lemma now follows from Steps 1 and 2. $\square$

Next we define an algorithm that iterates the operator $\mathcal{A}$. First let $\alpha_0(s) = 0$ for all s . Then, for each ordinal $\xi > 0$, define

$$\alpha_\xi = \mathcal{A}\alpha_{\xi-1}$$

if ξ is a successor ordinal, and

$$\alpha_\xi = \sup_{\nu < \xi} \alpha_\nu$$

if ξ is a limit ordinal.

Note that $\alpha_1 = \mathcal{A}0 \geq 0 = \alpha_0$. As $\mathcal{A}$ is a monotone operator, it follows by induction that the sequence $(\alpha_\xi)_\xi$ is nondecreasing in ξ .

The next result is strongly related to Tarski's [19] fixed point theorem. Note the following: (1) The set $\mathcal{L}^+$ of functions from S to $[0, \infty]$ is a complete lattice with the usual $\leq$ relation between functions. (2) The operator $\mathcal{A}$ is monotone, as we mentioned already. (3) The operator $\mathcal{A}$ has a fixed point: the function f such that $f(s) = \infty$ for every state $s \in S$. In view of these three properties, Tarski's fixed point theorem implies that $\mathcal{A}$ has a least fixed point too. The next lemma identifies the least fixed point of $\mathcal{A}$.

Lemma 5 *There exists a countable ordinal ξ_* such that $\sup_\xi \alpha_\xi = \alpha_{\xi_*}$. The function α_{ξ_*} is the least fixed point of $\mathcal{A}$. Moreover, α_{ξ_*} is the least $\mathcal{A}$ -excessive function, that is, the least function $f \in \mathcal{L}^+$ such that $\mathcal{A}f \leq f$.*

Proof The existence of ξ_* follows from a cardinality argument. Indeed, for every state $s \in S$, the sequence $(\alpha_\xi(s))_\xi$ is a nondecreasing sequence of real numbers. Hence, the sequence $(\alpha_\xi(s))_\xi$ can only have countably many different elements.¹ Thus, for state s , there exists a countable ordinal ξ_s such that $\alpha_\xi(s) = \alpha_{\xi_s}$ for every $\xi \geq \xi_s$. Now take $\xi_* = \sup_{s \in S} \xi_s$. Because S is countable, ξ_* is a countable ordinal.

It also follows that $\alpha_{\xi_*} = \alpha_{\xi_*+1} = \mathcal{A}\alpha_{\xi_*}$, so α_{ξ_*} is a fixed point of $\mathcal{A}$.

Now we show that α_{ξ_*} is the least fixed point of $\mathcal{A}$. Let $f \in \mathcal{L}^+$ be any fixed point of $\mathcal{A}$. As $f \in \mathcal{L}^+$, we have $0 \leq f$. As $\mathcal{A}$ is monotone, it follows by induction that $\alpha_\xi \leq f$ for every ξ . Hence, $\alpha_{\xi_*} \leq f$.

¹ Otherwise, whenever $\alpha_\xi(s) < \alpha_{\xi+1}(s)$, we can choose a rational number $q_\xi \in [\alpha_\xi(s), \alpha_{\xi+1}(s)]$, and the cardinality of these rational numbers would be uncountable.

As α_{ξ_*} is a fixed point of $\mathcal{A}$, clearly α_{ξ_*} is $\mathcal{A}$ -excessive. Suppose that $f \in \mathcal{L}^+$ is $\mathcal{A}$ -excessive too. Then, $\alpha_0 = 0 \leq f$ implies $\alpha_1 = \mathcal{A}\alpha_0 \leq \mathcal{A}f \leq f$. Continuing by induction, we find that $\alpha_{\xi_*} \leq f$. Thus, α_{ξ_*} is the least $\mathcal{A}$ -excessive function. $\square$

Theorem 6 *The lower value α of Γ is equal to α_{ξ_*} .*

Proof It follows from Lemmas 4 and 5 that $\alpha \geq \alpha_{\xi_*}$.

To complete the proof of the theorem, take any strategy π for player 1 and let $\epsilon > 0$. We show that there exists a strategy σ for player 2 such that $u(s, \pi, \sigma) \leq \alpha_{\xi_*}(s) + 2\epsilon$ for every state $s \in S$. This will imply that $\alpha \leq \alpha_{\xi_*}$.

If at any stage t , the current state is s and the history is $h \in H_t$, then let $\sigma(h)$ be a mixed action in $\Delta(B(s))$ such that

$$\sum_{a,b} \left[r(s, a, b) + \sum_{s'} \alpha_{\xi_*}(s') p(s'|s, a, b) \right] \cdot \pi(h)(a) \cdot \sigma(h)(b) \leq \alpha_{\xi_*}(s) + \epsilon \cdot 2^{-t}.$$

A mixed action $\sigma(h)$ with the required property does exist, because the lower value of the matrix game $M_{\alpha_{\xi_*}}(s)$ is $\mathcal{A}\alpha_{\xi_*}(s)$ by the definition of the operator $\mathcal{A}$, and because $\mathcal{A}\alpha_{\xi_*}(s) = \alpha_{\xi_*}(s)$ by Lemma 5.

Consider the process

$$Q_0 = \alpha_{\xi_*}(s_0), \quad \text{and} \quad Q_n = \sum_{t=0}^{n-1} r_t + \alpha_{\xi_*}(s_n), \quad n \geq 1.$$

Then

$$\begin{aligned} \mathbb{E}_{s,\pi,\sigma}[Q_{n+1}|h_n] &= \mathbb{E}_{s,\pi,\sigma} \left[\sum_{t=0}^n r_t + \alpha_{\xi_*}(s_{n+1}) | h_n \right] \\ &\leq \sum_{t=0}^{n-1} r_t + \alpha_{\xi_*}(s_n) + \epsilon \cdot 2^{-n} = Q_n + \epsilon \cdot 2^{-n}, \end{aligned}$$

where the inequality follows by the definition of $\sigma(h_n)$. Hence we have $\mathbb{E}_{s,\pi,\sigma}[Q_{n+1}] \leq \mathbb{E}_{s,\pi,\sigma}[Q_n] + \epsilon \cdot 2^{-n}$. Combining the inequalities we obtain

$$\mathbb{E}_{s,\pi,\sigma}[Q_{n+1}] \leq \mathbb{E}_{s,\pi,\sigma}[Q_0] + \sum_{t=0}^n \epsilon \cdot 2^{-t} \leq \alpha_{\xi_*}(s) + 2\epsilon.$$

Thus, for all $n \in \mathbb{N}$,

$$\mathbb{E}_{s,\pi,\sigma} \left[\sum_{t=0}^n r_t \right] \leq \mathbb{E}_{s,\pi,\sigma}[Q_{n+1}] \leq \alpha_{\xi_*}(s) + 2\epsilon.$$

Using the monotone convergence theorem yields

$$u(s, \pi, \sigma) = \mathbb{E}_{s, \pi, \sigma} \left[\sum_{t=0}^{\infty} r_t \right] \leq \alpha_{\xi_*}(s) + 2\epsilon,$$

as desired. $\square$

4.2 The Operator $\mathcal{B}$ for the Upper Value

Lemma 7 *The upper value β of Γ is a fixed point of the operator $\mathcal{B}$.*

Proof The proof that $\mathcal{B}\beta = \beta$ is in two steps. Fix $s \in S$.

Step 1 We prove $\mathcal{B}\beta(s) \leq \beta(s)$.

We can assume that $\beta(s) < \infty$, otherwise the statement is obvious. Let $R > \beta(s)$. It can be proven similarly to (4.1) in the proof of Step 2 of Lemma 4 that there is a mixed action $y^* \in \Delta(B(s))$ such that for all $x \in \Delta(A(s))$ we have

$$\sum_{a,b} [r(s, a, b) + \sum_{s'} \beta(s') p(s'|s, a, b)] x(a) y^*(b) \leq R.$$

Hence, $\mathcal{B}\beta(s) \leq R$. Since this holds for all $R > \beta(s)$, we have shown that $\mathcal{B}\beta(s) \leq \beta(s)$.

Step 2 We prove $\mathcal{B}\beta(s) \geq \beta(s)$.

We can assume that $\mathcal{B}\beta(s) < \infty$, otherwise the statement is obvious. Let $R > \mathcal{B}\beta(s)$ and $\epsilon > 0$. It can be proven similarly to the proof of Step 1 of Lemma 4 that player 2 has a strategy σ^* such that $u(s, \pi, \sigma^*) \leq R + 2\epsilon$ for every π . This implies that $\beta(s) \leq R + 2\epsilon$. Since this holds for all $R > \mathcal{B}\beta(s)$ and $\epsilon > 0$, we have shown that $\beta(s) \leq \mathcal{B}\beta(s)$. $\square$

We also define an algorithm that iterates the operator $\mathcal{B}$. First let $\beta_0(s) = 0$ for all s . Then, for each ordinal $\xi > 0$, define

$$\beta_\xi = \mathcal{B}\beta_{\xi-1}$$

if ξ is a successor ordinal, and

$$\beta_\xi = \sup_{\nu < \xi} \beta_\nu$$

if ξ is a limit ordinal.

Note that $\beta_1 = \mathcal{B}0 \geq 0 = \beta_0$. As $\mathcal{B}$ is a monotone operator, it follows by induction that the sequence $(\beta_\xi)_\xi$ is nondecreasing in ξ .

Similarly to Lemma 5, we have the following.

Lemma 8 *There exists a countable ordinal ξ^* such that $\sup_\xi \beta_\xi = \beta_{\xi^*}$. The function β_{ξ^*} is the least fixed point of $\mathcal{B}$. Moreover, β_{ξ^*} is the least $\mathcal{B}$ -excessive function, that is, the least function $f \in \mathcal{L}^+$ such that $\mathcal{B}f \leq f$.*

Similarly to Theorem 6, we have the following.

Theorem 9 *The upper value β of Γ is equal to β_{ξ^*} .*

5 The Value and Optimal Strategies

The next theorem is a consequence of Theorems 6 and 9.

Theorem 10 *Consider a positive stochastic game. The game has a value if and only if $\alpha_{\xi_*} = \beta_{\xi^*}$. In that case, the value is equal to $\alpha_{\xi_*} = \beta_{\xi^*}$.*

The following example illustrates that a game can admit a value even when the one-shot games fail to have values at some stages of the recursive algorithms. This illustrates the advantage of our approach that uses separate algorithms to find the upper value and the lower value, rather than having only one algorithm that iterates the value of the one-shot games.

Example 4 Consider the following game with state space $S = \{1, 2, 3, 4\}$, with state 1 being the initial state. In state 1, the action spaces are $\mathbb{N}$ for both players, the reward is 0, and if actions a and b are chosen, then the next state is state 3 if $a > b$ and it is state 2 if $a \leq b$. (The transitions depend on which player chooses the higher number.) In state 2, the reward is 0, and the next state is state 3. In state 3, the reward is 1, and the next state is state 4. State 4 is absorbing with reward 0.

Clearly, whatever happens, the sum of the rewards is 1. So the game has value equal to 1 from initial state 1. In the recursion, we get $\alpha_0 = \beta_0 = (0, 0, 0, 0)$ at ordinal 0 and $\alpha_1 = \beta_1 = (0, 0, 1, 0)$ at ordinal 1. But in the next step, at ordinal 2, the one-shot game $M_{\alpha_1}(1) = M_{\beta_1}(1)$ admits no value: indeed, we have $\alpha_2(1) = 0$ and $\beta_2(1) = 1$. More precisely, $\alpha_2 = (0, 1, 1, 0)$ and $\beta_2 = (1, 1, 1, 0)$. The algorithms reach the value of the game at ordinal 3: $\alpha_3 = \beta_3 = (1, 1, 1, 0)$. $\square$

If in each state $s \in S$, either $A(s)$ or $B(s)$ is finite, then Theorems 2 and 3 guarantee that the one-shot game $M_f(s)$, for every $f \in \mathcal{L}^+$ and every state $s \in S$, admits a value. Hence, in that case, the operators $\mathcal{A}$ and $\mathcal{B}$ are equal, $\alpha_\xi = \beta_\xi$ for all ordinals ξ , and $\xi_* = \xi^*$. To emphasize this, in that case, we will use the notations $T := \mathcal{A} = \mathcal{B}$ and $v_\xi := \alpha_\xi = \beta_\xi$ for all ordinals ξ and $v^* = v_{\xi_*} = v_{\xi^*}$.

Theorem 11 *Assume that in every state s , either $A(s)$ or $B(s)$ is finite. Then the following hold:*

1. *The value of the stochastic game exists and is equal to v^* .*
2. *The value function v^* is the least T -excessive function, that is, the least function $f \in \mathcal{L}^+$ such that $Tf \leq f$.*
3. *For every $\epsilon > 0$, player 2 has an ϵ -optimal Markov strategy.*
4. *If expected payoffs are uniformly bounded in the sense that*

$$\sup_{s, \pi, \sigma} u(s, \pi, \sigma) < \infty,$$

then, for every positive integer n and $\epsilon > 0$, there is a stationary strategy π for player 1 such that, for all s and all strategies σ for player 2, $u(s, \pi, \sigma) \geq v_n(s) - \epsilon$.

Proof The first assertion follows from Theorem 10; the second is by Lemma 5 or by Lemma 8.

To prove the third assertion, let $\epsilon > 0$. Recall that $Tv^*(s) = v^*(s)$ is the value of the matrix game $M_{v^*}(s)$. Now for each state $s \in S$ and every $n \in \mathbb{N}$, choose a mixed action $y_n(s)$ in $\Delta(B(s))$ to be $\epsilon \cdot 2^{-n}$ -optimal in the one-shot game $M_{v^*}(s)$; that is, for all $x \in \Delta(A(s))$,

$$\sum_{a,b} [r(s, a, b) + \sum_{s'} v^*(s') p(s'|s, a, b)] \cdot x(a) \cdot y_n(s)(b) \leq v^*(s) + \epsilon \cdot 2^{-n}.$$

Let σ be the Markov strategy for player 2 that uses the action $y_n(s)$ when in state s at stage n , for all s and n . One can show that for any strategy π for player 1 and any state s we have $u(s, \pi, \sigma) \leq v^*(s) + 2\epsilon$, i.e. that σ is a 2ϵ -optimal strategy. The argument that σ is a 2ϵ -optimal strategy is similar to the proof of Theorem 6, and is omitted.

We also omit the proof of the final assertion, because it is only a slight variation on the proof of Theorem 3.2 in Secchi [17]. $\square$

The characterization of the value function as the least T -excessive function mirrors a similar characterization of the optimal reward function in the (one-person) gambling theory of Dubins and Savage [3]. It also appears as the réduite of a function in the potential theoretic treatment of gambling theory by Dellacherie and Meyer [2]. See also Lemma 7.10, p. 187 in Maitra and Sudderth [12] for a similar characterization of the value of a leavable game.

The following simple example shows that if player 2's action space $B(s)$ is infinite in some state s , then player 2 may not have a 0-optimal strategy or an ϵ -optimal stationary strategy for small $\epsilon > 0$.

Example 5 Consider a game in which there is only one state, player 1 has only one action and player 2's action space is $\mathbb{N}$. If player 2 chooses action n then the reward is $\frac{1}{2^n}$.

Clearly, the value of the game is 0, and player 2 has no 0-optimal strategy. Moreover, any stationary strategy of player 2 induces an expected payoff of infinity, thus player 2 has no ϵ -optimal stationary either, for any $\epsilon > 0$.

However, for any $\epsilon > 0$, player 2 has an ϵ -optimal Markov strategy, as stated in Theorem 11. Indeed, let $m \in \mathbb{N}$ be such that $m \geq 1$ and $\frac{1}{2^{m-1}} \leq \epsilon$. Consider the Markov strategy that chooses action $m + n$ at stage n . This induces a payoff of

$$\frac{1}{2^m} + \frac{1}{2^{m+1}} + \frac{1}{2^{m+2}} + \cdots = \frac{1}{2^{m-1}}.$$

By the choice of m , this Markov strategy is ϵ -optimal. $\square$

The next theorem states that, when the action space of player 2 is finite in each state, it suffices to execute the algorithm on the natural numbers to find the value of the stochastic game. As usual, ω denotes the first limit ordinal.

Theorem 12 Assume that $B(s)$ is finite for all states $s \in S$. Then:

1. The value of the positive stochastic game is equal to $v_\omega = \lim_{n \rightarrow \infty} v_n$.
2. Player 2 has an optimal stationary strategy.
3. If expected payoffs are uniformly bounded in the sense that

$$\sup_{s, \pi, \sigma} u(s, \pi, \sigma) < \infty,$$

then, for every $s \in S$ and $\epsilon > 0$, player 1 has a stationary strategy that is ϵ -optimal at state s .

Proof The Proof of Part 1 It suffices to show that $v^* = v_\omega$. It will then follow from Theorem 11 that the value of the positive stochastic game is equal to v_ω .

We already know that $v_\omega \leq v^*$. So, in view of Lemma 5 and Theorem 6 (or by Lemma 8 and Theorem 9), it suffices to show that v_ω is T -excessive: $Tv_\omega \leq v_\omega$. Fix a state $s \in S$.

Let $R < Tv_\omega(s)$. As $Tv_\omega(s)$ is the value of the one-shot game $M_{v_\omega}(s)$, there exists $x \in \Delta(A(s))$ such that for all $b \in B(s)$

$$\sum_{a \in A(s)} \left[r(s, a, b) + \sum_{z \in S} v_\omega(z) p(z|s, a, b) \right] x(a) > R.$$

By the monotone convergence theorem, for all $b \in B(s)$

$$\lim_{n \rightarrow \infty} \left[\sum_{a \in A(s)} [r(s, a, b) + \sum_{z \in S} v_n(z) p(z|s, a, b)] x(a) \right] > R.$$

As $B(s)$ is assumed to be finite, for n sufficiently large, we have for all $b \in B(s)$

$$\sum_{a \in A(s)} \left[r(s, a, b) + \sum_{z \in S} v_n(z) p(z|s, a, b) \right] x(a) > R.$$

Hence, the value of the one-shot game $M_{v_n}(s)$, namely $Tv_n(s)$, is at least R for n sufficiently large. Thus,

$$v_\omega(s) = \lim_{n \rightarrow \infty} v_n(s) = \lim_{n \rightarrow \infty} Tv_n(s) \geq R.$$

As this holds for all $R < Tv_\omega(s)$, we have shown that $v_\omega(s) \geq Tv_\omega(s)$.

The Proof of Part 2 The proof is based on using Theorem 2, which allows us to choose, for each state $s \in S$, a mixed action $y(s) \in \Delta(B(s))$ that is optimal for player 2 in the one-shot game $M_{v^*}(s)$. Let σ be the stationary strategy for player 2 that uses $y(s)$ at each state $s \in S$. The proof that σ is optimal is similar to the proof of part 3 of Theorem 11.

The Proof of Part 3 This follows from part 1 together with part 4 of Theorem 11. $\square$

Theorem 12 is almost a special case of Theorem 5.4 in Nowak [13]. Nowak's result is for a much more general setting with Borel state and action spaces, but uses a condition on n -stage games called FA that is not needed for our simpler result.

It need not be true that v_ω is the value of the stochastic game if player 2 has an infinite action set. Here is an example in which player 1 is a dummy.

Example 6 Let $S = \{0, 1, \dots\}$. Suppose play begins at state 0 and player 2 can choose any action in $B(0) = \{2, 3, \dots\}$. If he chooses action $n \in B(0)$ the next state is n . Motion is deterministic from state n to $n - 1$ for $n \geq 2$ and state 1 is absorbing. The reward r depends only on the state and is equal to 0 except at state 2 where it equals 1.

By choosing an initial action $b > n$, player 2 can guarantee a reward of 0 for the first n stages. Hence, $v_n(0) = T^n 0(0) = 0$. However, the value of the stochastic game with initial state 0 is 1. In this example, the value function is equal to $v_{\omega+1} = T v_\omega$. $\square$

The next example shows that player 1 may have no ϵ -optimal stationary strategy, even if the state and action spaces are finite, and the reward function has only finite values.

Example 7 Consider the following game.

	L	R
T	0	1
B	1	0^*

In this game, there is only one non-trivial state, shown in the figure. In this state, player 1's actions are T and B , and player 2's actions are L and R . The rewards for the corresponding action combinations are given in the matrix. The transitions are as follows: if entry (B, R) is chosen, then transition occurs to an absorbing state where the reward is equal to 0, while after any other action combination the game remains in the non-trivial state.

The value for the non-trivial state is infinity, and player 1 has an optimal strategy. To show this we construct a strategy for player 1 that guarantees that with positive probability either entry (B, L) is played infinitely often or entry (T, R) is played infinitely often. Let $\{q_n\}_{n \in \mathbb{N}}$ be a sequence in $(0, 1)$ such that $q^* := \inf_{n=0}^\infty q_n > 0$. Let H' denote the set of histories in which the final state is still the non-trivial one (so in which no absorption has taken place). For each $h \in H'$ let R_h denote the number of times that player 2 has played action R in h . Consider the following strategy for player 1: at any $h \in H'$, play the mixed action $(q_{R_h}, 1 - q_{R_h})$. It is not difficult to verify that if player 1 uses this strategy then regardless the strategy of player 2, with probability at least q^* , either entry (B, L) is played infinitely often or entry (T, R) is played infinitely often. To see this, notice that (a) the probability of absorption in entry (B, R) is at most $1 - q^*$, and (b) the probability that there is an $N \in \mathbb{N}$ such that at all stages $n > N$ entry (T, L) is played is zero.

Any stationary strategy of player 1 only guarantees a finite expected payoff. Indeed, it is clear if the stationary strategy is pure. So assume that the stationary strategy places probability $z \in (0, 1)$ on action T . Then, if player 2 plays action R at every stage, the expected payoff is

$$(1 - z) \cdot 0 + z(1 - z) \cdot 1 + z^2(1 - z) \cdot 2 + \cdots = \frac{z}{1 - z},$$

which is finite, as claimed. Hence, player 1 has no ϵ -optimal stationary strategy, for any $\epsilon > 0$.

Yet, there is a Markov strategy for player 1 that guarantees that the expected payoff is infinite. Let the Markov strategy π place probability $z_n = \frac{n+1}{n+2}$ on action T at stage n . So, the sequence of probabilities on action T is $\frac{1}{2}, \frac{2}{3}, \frac{3}{4}, \dots$.

Consider a pure Markov strategy σ for player 2. Since π is Markov, it is sufficient to consider such responses from player 2. We distinguish two cases.

Case 1: Suppose first that σ chooses action R at infinitely many stages, say at stages in $M = \{n_1, n_2, \dots\}$.

Case 1.1: Suppose that $M = \mathbb{N}$, i.e., σ chooses action R at all stages. Then, the expected payoff from the non-trivial state s is

$$\begin{aligned} u(s, \pi, \sigma) &= (1 - z_0) \cdot 0 + z_0(1 - z_1) \cdot 1 + z_0z_1(1 - z_2) \cdot 2 + \cdots \\ &= z_0 + z_0z_1 + z_0z_1z_2 + \cdots \\ &= \frac{1}{2} + \frac{1}{3} + \frac{1}{4} + \cdots \\ &= \infty. \end{aligned}$$

Case 1.2: Now, more generally, suppose that $M \subseteq \mathbb{N}$. By ignoring the non-negative rewards at stages outside M (when σ chooses action L) and only looking at the expected sum of the rewards at stages in M , we get an infinite sum similar to the one above and find that

$$u(s, \pi, \sigma) \geq \frac{1}{2} + \frac{1}{3} + \frac{1}{4} + \cdots = \infty.$$

Case 2: Suppose that σ chooses action L at all stages from stage N onwards. Note that no absorption occurs before stage N with probability at least

$$\frac{1}{2} \cdot \frac{2}{3} \cdots \frac{N}{N+1} = \frac{1}{N+1},$$

which is positive. Because

$$\sum_{n=N}^{\infty} (1 - z_n) = \frac{1}{N+2} + \frac{1}{N+3} + \frac{1}{N+4} + \cdots = \infty,$$

the second Borel–Cantelli lemma implies that, if no absorption occurs before stage N , then with probability 1, σ will choose action B infinitely often from stage N onwards. Therefore,

$$u(s, \pi, \sigma) \geq \frac{1}{N+1} \cdot \infty = \infty.$$

□

Open Access This article is distributed under the terms of the Creative Commons Attribution 4.0 International License (<http://creativecommons.org/licenses/by/4.0/>), which permits unrestricted use, distribution, and reproduction in any medium, provided you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made.

References

1. Blackwell, D.: Operator solution of infinite G_δ games of imperfect information. In: Anderson, T.W., Athreya, K., Iglehart, D.L. (eds.) *Probability, Statistics, and Mathematics: Papers in Honor of Samuel Karlin*. Academic Press, New York (1989)
2. Dellacherie, C., Meyer, P.-A.: *Probabilités et potentiel, Chapitres IX à XI*. Hermann, Paris (1983)
3. Dubins, L.E., Savage, L.J.: *How to Gamble If You Must: Inequalities for Stochastic Processes* (Dover Editions in 1976 and 2014). McGraw-Hill, New York (1965)
4. Flesch, J., Predtetchinski, A., Sudderth, W.: Simplifying optimal strategies in limsup and liminf stochastic games. *Discrete Appl. Math.* (Forthcoming)
5. Flesch, J., Predtetchinski, A., Sudderth, W.: Characterization and simplification of optimal strategies in positive stochastic games. *J. Appl. Probab.* **55**, 1–14 (2018)
6. Frid, E.: On stochastic games. *Theory Probab. Appl.* **18**, 389–393 (1973)
7. Jaśkiewicz, Nowak, : *Zero-Sum Stochastic Games*. *Handbook of Dynamic Game Theory*, pp. 1–65. Springer, Berlin (2016)
8. Kumar, P.R., Shiau, T.H.: Existence of value and randomized strategies in zero-sum discrete-time stochastic dynamic games. *SIAM J. Control Optim.* **19**, 617–634 (1981)
9. Maitra, A., Parthasarathy, T.: On stochastic games II. *J. Optim. Theory Appl.* **8**, 154–160 (1971)
10. Maitra, A., Sudderth, W.: An operator solution of stochastic games. *Isr. J. Math.* **78**, 33–49 (1992)
11. Maitra, A., Sudderth, W.: Borel stochastic games with limsup payoff. *Ann. Probab.* **21**, 861–885 (1993)
12. Maitra, A., Sudderth, W.: *Discrete Gambling and Stochastic Games*. Springer, Berlin (1996)
13. Nowak, A.S.: Universally measurable strategies in zero-sum stochastic games. *Ann. Probab.* **13**, 269–287 (1985)
14. Nowak, A.S., Raghavan, T.E.S.: Positive stochastic games and a theorem of Ornstein. In: *Stochastic Games and Related Topics, Shapley Honor Volume*. Kluwer, Dordrecht, pp. 127–134 (1991)
15. Parthasarathy, T.: Discounted and positive stochastic games. *Bull. Am. Math. Soc.* **77**, 134–136 (1971)
16. Parthasarathy, T.: Discounted, positive, and noncooperative stochastic games. *Int. J. Game Theory* **2**, 25–37 (1973)
17. Secchi, P.: Stationary strategies for recursive games. *Math. Oper. Res.* **22**, 494–512 (1997)
18. Strauch, R.: Negative dynamic programming. *Ann. Math. Stat.* **37**, 871–890 (1966)
19. Tarski, A.: A lattice-theoretical fixed point theorem. *Pac. J. Math.* **5**, 285–309 (1955)
20. Tijs, S.H.: *Semi-infinite and infinite matrix games and bimatrix games*. PhD thesis, Tilburg University, The Netherlands (1975)