Simple optimal strategies in limsup stochastic games

János Flesch*, Arkadi Predtetchinski*, William Sudderth**

** Maastricht University, The Netherlands*
*** University of Minnesota, US*

Mons, Belgium
Theory and Algorithms in Graph and Stochastic Games
March 14-15, 2019

We consider zero–sum stochastic games:

- dynamic games: time periods 1,2,...
- played by two players having opposite interests (antagonistic)
- simultaneous moves are possible (concurrent)

We focus on the limsup payoff.

Player 1: maximizes it (he wants high payoffs infinitely often).

Player 2: minimizes it.

Importance: Several classes of games can be seen as limsup games.

In a limsup game, the players may have no optimal strategies.

Main question: Provided that a player has an optimal strategy in the game, does he have a "simple" optimal strategy?

Answer: Yes. So, no need to use complicated optimal strategies.

(Literature: "conditional" strategic complexity of optimal strategies.)

Agenda

1. Zero–sum stochastic games with the limsup payoff
2. For player 1: simple optimal strategies
3. For player 2: simple optimal strategies
4. Extensions
5. The total payoff

An example:

	L_1	R_1
T_1	0 $\circlearrowleft$	2 $\rightarrow$
B_1	1 0.5 0.5	0 $\circlearrowleft$

state 1*

T_2	1 0.5 0.5
B_2	2 0.7 0.3

state 2

- States: matrices – finitely many
- Actions: – finitely many
  - player 1: rows
  - player 2: columns
- Cells of the form:

payoff transition

Play: at periods $1, 2, \ldots$

state 1*

	L_1	R_1
T_1	0 ↻	2 →
B_1	1 (0.5 0.5)	0 ↻

state 2

T_2	1 (0.5 0.5)
B_2	2 (0.7 0.3)

1. Period 1: in state 1
   - players choose actions (simul. and indep.), say T_1 and L_1
   - player 1 receives payoff 0 from player 2
   - play remains in state 1
2. Period 2: in state 1
   - say actions B_1 and L_1
   - player 1 receives payoff 1 from player 2
   - play moves to each state with prob 0.5, and so on ...

Play: at periods $1, 2, \ldots$

	L_1	R_1
T_1	0 $\circlearrowleft$	2 $\rightarrow$
B_1	1 0.5 0.5	0 $\circlearrowleft$

state 1^*

T_2	1 0.5 0.5
B_2	2 0.7 0.3

state 2

1. Period 1: in state 1
   - players choose actions (simul. and indep.), say T_1 and L_1
   - player 1 receives payoff 0 from player 2
   - play remains in state 1
2. Period 2: in state 1
   - say actions B_1 and L_1
   - player 1 receives payoff 1 from player 2
   - play moves to each state with prob 0.5, and so on ...

Play: at periods $1, 2, \ldots$

	L_1	R_1
T_1	0 ↻	2 →
B_1	1 0.5 0.5	0 ↻

state 1*

T_2	1 0.5 0.5
B_2	2 0.7 0.3

state 2

1. Period 1: in state 1
   - players choose actions (simul. and indep.), say T_1 and L_1
   - player 1 receives payoff 0 from player 2
   - play remains in state 1
2. Period 2: in state 1
   - say actions B_1 and L_1
   - player 1 receives payoff 1 from player 2
   - play moves to each state with prob 0.5, and so on ...

When choosing an action, the objective of each player:

- to get a good payoff (for today)
- and to get a good transition (for the future).

These can be conflicting !

Strategies:

	L_1	R_1
T_1	0 ↻	2 →
B_1	1 0.5 0.5	0 ↻

state 1*

T_2	1 0.5 0.5
B_2	2 0.7 0.3

state 2

Stationary strategy: state dependent. For player 1:

$$x = \begin{cases} (p, 1-p) & \text{for state 1} \\ (q, 1-q) & \text{for state 2} \end{cases}$$

"When you are in state 1: play T_1 with prob p and B_1 with prob 1-p;
When you are in state 2: play T_2 with prob q and B_2 with prob 1-q."

History-dependent strategy: action depends on past play as well.

Formally:

A zero-sum stochastic game is given by:

- finite set of states S
- finite sets of actions $A(s)$ and $B(s)$ – for each state $s \in S$
- payoff $r(s,a,b) \in \mathbb{R}$ and transition $p(s,a,b) \in \Delta(S)$
  – for each state $s \in S$ and actions $a \in A(s)$ and $b \in B(s)$.

History at period t: The sequence

$$h^t = (s_1, a_1, b_1, \ldots, s_{t-1}, a_{t-1}, b_{t-1}).$$

The set of all possible histories is H.

Strategies: A strategy σ for player 1 is a mapping that assigns to each $(h, s) \in H \times S$ a mixed action $\sigma(h, s) \in \Delta(A(s))$. A strategy τ for player 2 is similar.

A strategy is stationary if it only considers the current state: $x = x(s)_{s \in S}$ for player 1 and $y = y(s)_{s \in S}$ for player 2.

Limsup payoff: It measures the long–term best performance:

$$u(r_1, r_2, \ldots) \;=\; \limsup_{t\to\infty}\; r_t.$$

It is the highest payoff received infinitely often, e.g.:

$$u(1, 0, 1, 0, \ldots) \;=\; 1.$$

Expected limsup payoff: For initial state s, strategies σ and τ:

$$u(s, \sigma, \tau) \;=\; \mathbb{E}_{(s,\sigma,\tau)}\Big[\limsup_{t\to\infty}\; r_t\Big]\;.$$

- player 1: maximizes
- player 2: minimizes, which is the same as maximizing

$$\mathbb{E}_{(s,\sigma,\tau)}\Big[\liminf_{t\to\infty}\;(-r_t)\Big].$$

Value: Maitra and Sudderth [1992, 1998], Martin [1998]

The game has a value for every initial state s:

$$v(s) := \sup_\sigma \inf_\tau u(s,\sigma,\tau) = \inf_\tau \sup_\sigma u(s,\sigma,\tau).$$

Optimality: For $\varepsilon \geq 0$, a strategy σ^* is called ε–optimal if

$$u(s,\sigma^*,\tau) \geq v(s) - \varepsilon \qquad \forall\tau\ \forall s.$$

Similarly, τ^* is called ε–optimal if

$$u(s,\sigma,\tau^*) \leq v(s) + \varepsilon \qquad \forall\sigma\ \forall s.$$

A 0-optimal strategy is called optimal.

Several classes of games can be seen as limsup games:

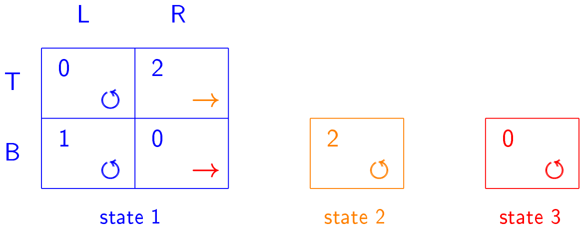
- e.g. recursive games, games with reachability/safety conditions or with Büchi/co-Büchi conditions ...

Related literature: decision/game theory and computer science

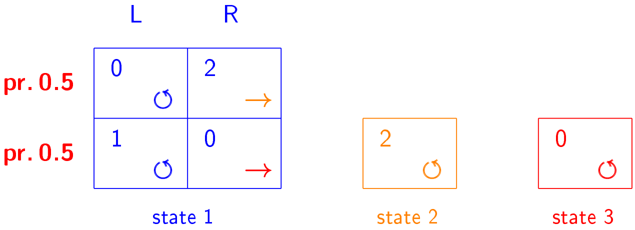
- Limsup (and liminf) payoff: e.g. Dubins, Savage [1965], Maitra, Sudderth [1992, 1993, 1996], Secchi [1998], Chatterjee, Henzinger [2007], Chatterjee, Doyen, Henzinger [2009], Hunter, Raskin [2014], Sudderth [2016], Gimbert, Renault, Sorin, Venel, Zielonka [2016], Bruyère [2017] ...

- Other “simplification” results: e.g. Dubins, Savage [1965], Strauch [1966], Blackwell [1971], Orkin [1974], Flesch, Thuijsman, Vrieze [1998] ...

Example 1 – Leading example.



Example 1 – Leading example.

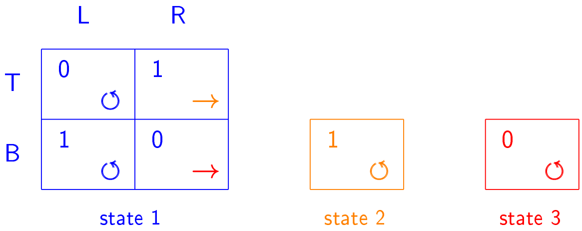


Consider the stationary strategy $x = (0.5, 0.5)$.

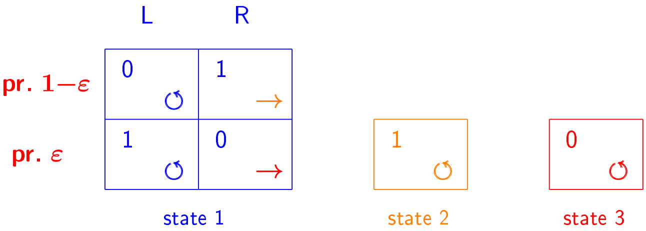
- If always action L: the limsup payoff is 1. **!!**
- If ever action R: also 1.

So: The value is $v(1) = 1$, and x is optimal.

Example 2. Now (T, R) gives payoff 1.



Example 2. Now (T, R) gives payoff 1.

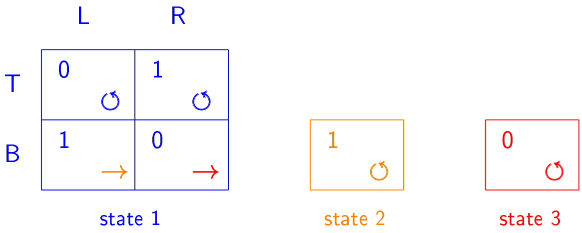


Consider the stationary strategy $(1-\varepsilon, \varepsilon)$, where $\varepsilon \in (0, 1)$.

- If always action L: the limsup payoff is 1.
- If ever action R: the limsup payoff is $1 - \varepsilon$.

So: The value is $v(1) = 1$. But player 1 has **no optimal strategy**.

Example 3. The Big Match, Gillette [1957] – later Kristoffer's talk
But now with the limsup payoff !



- The value is $v(1) = 1$.
- Player 1 has no optimal strategy.
- For small $\varepsilon > 0$, player 1 has only history-dependent ε-optimal strategies.

Agenda

1. Zero–sum stochastic games with the limsup payoff
2. For player 1: simple optimal strategies
3. For player 2: simple optimal strategies
4. Extensions
5. The total payoff

Theorem. (Flesch, Predtetchinski and Sudderth)

Suppose that player 1 has an optimal strategy. Then, player 1 has a stationary optimal strategy.

Theorem. (Flesch, Predtetchinski and Sudderth)

Suppose that player 1 has an optimal strategy. Then, player 1 has a stationary optimal strategy.

Outline of the proof:

1. We define a stationary strategy $x^* = (x^*(s))_{s \in S}$ for player 1.

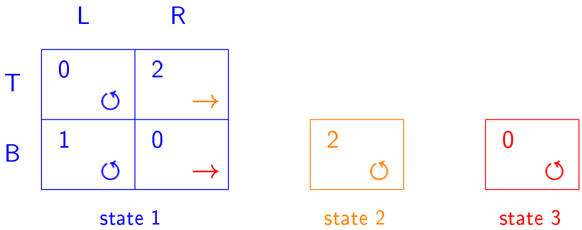
   In each state s:

   1.1: We examine an auxiliary one-shot game $M(s)$.

   1.2: Based on $M(s)$, we select a mixed action $x^*(s)$.

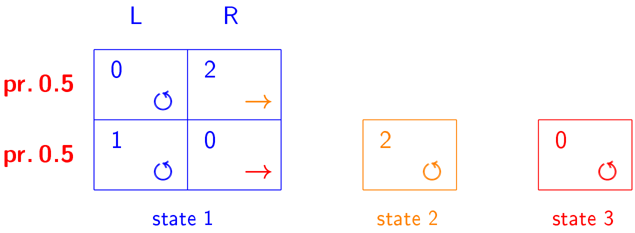
2. We show that x^* is optimal, provided that player 1 has an optimal strategy (we do not need to know this strategy).

We illustrate these steps with the leading example:



Recall: The value is $v(1) = 1$, and the stationary strategy $x = (0.5, 0.5)$ is optimal.

We illustrate these steps with the leading example:



Recall: The value is $v(1) = 1$, and the stationary strategy $x = (0.5, 0.5)$ is optimal.

	L	R
T	0 ↺	2 →
B	1 ↺	0 →

state 1
$v(1) = 1$

2 ↺

state 2
$v(2) = 2$

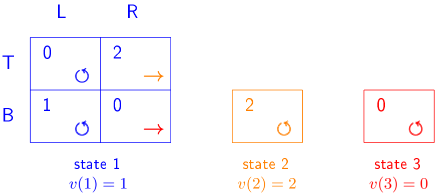
0 ↺

state 3
$v(3) = 0$

1.1 For state 1, the one-shot game $M(1)$:

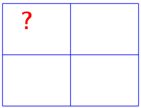
Each cell of $M(1)$ contains "the value of the next state".

M(1) for state 1

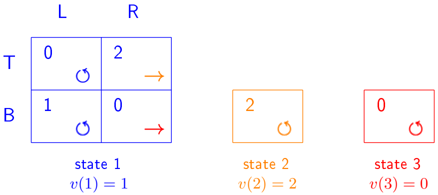


(1.1) For state 1, the one-shot game $M(1)$:

Each cell of $M(1)$ contains "the value of the next state".



M(1) for state 1



1.1 For state 1, the one-shot game $M(1)$:

Each cell of $M(1)$ contains "the value of the next state".

1	

M(1) for state 1

	L	R
T	0 ↺	2 →
B	1 ↺	0 →

state 1
$v(1) = 1$

2 ↺

state 2
$v(2) = 2$

0 ↺

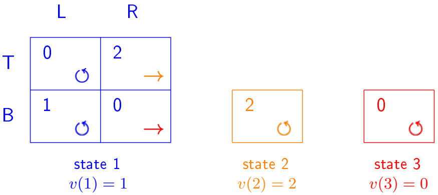
state 3
$v(3) = 0$

1.1 For state 1, the one-shot game $M(1)$:

Each cell of $M(1)$ contains "the value of the next state".

1	?

M(1) for state 1



1.1 For state 1, the one-shot game $M(1)$:

Each cell of $M(1)$ contains “the value of the next state”.

1	2

M(1) for state 1

	L	R
T	0 ↺	2 →
B	1 ↺	0 →

state 1
$v(1) = 1$

2 ↺

state 2
$v(2) = 2$

0 ↺

state 3
$v(3) = 0$

1.1 For state 1, the one-shot game $M(1)$:

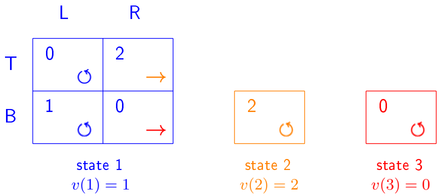
Each cell of $M(1)$ contains "the value of the next state".

1	2
1	0

M(1) for state 1

Val $M(1)$ equals $v(1) = 1$, and the optimal mixed actions are:

$$O(1) \; = \; \{(q, 1-q) : q \geq 0.5\}.$$



1.1 For state 1, the one-shot game $M(1)$:

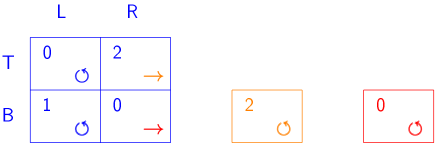
Each cell of $M(1)$ contains "the value of the next state".

pr. 1	1	2
pr. 0	1	0

M(1) for state 1

Val $M(1)$ equals $v(1) = 1$, and the optimal mixed actions are:

$$O(1) \ = \ \{(q, 1-q) : q \geq 0.5\}.$$



state 1 $v(1) = 1$ | state 2 $v(2) = 2$ | state 3 $v(3) = 0$

1.1 For state 1, the one-shot game $M(1)$:

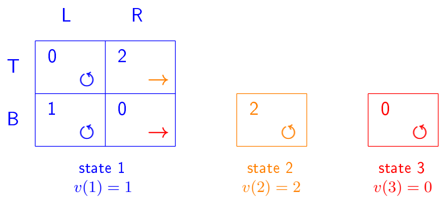
Each cell of $M(1)$ contains "the value of the next state".

pr. 0.5	1	2
pr. 0.5	1	0

M(1) for state 1

Val $M(1)$ equals $v(1) = 1$, and the optimal mixed actions are:

$$O(1) \;=\; \{(q, 1-q) : q \geq 0.5\}.$$



1.2 For state 1, selecting $x^*(1)$:

Let $x^*(1)$ be a mixed action that:

- **i.** belongs to $O(1) \; = \; \{(q, 1-q) : q \geq 0.5\}$ $\rightarrow$ good for transitions
- **ii.** has the largest support, subject to i. $\rightarrow$ good for payoffs

For example: $x^*(1) = (0.5, 0.5)$.

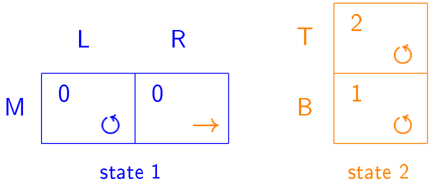
But not: $(1, 0)$ – bad payoff !

Agenda

1. Zero–sum stochastic games with the limsup payoff
2. For player 1: simple optimal strategies
3. For player 2: simple optimal strategies
4. Extensions
5. The total payoff

A strategy is called **subgame-optimal** if it induces an optimal strategy in each subgame.

It is stronger than optimality:



The value for initial state 1 is $v(1) = 0$.

The stationary strategy (M,B) is optimal for initial state 1, but **not subgame-optimal**.

Theorem. (Flesch, Predtetchinski and Sudderth)

Assume that player 2 has a subgame-optimal strategy.
Then, player 2 has a subgame-optimal strategy that only uses the the current state and the previous state and actions.

Finite memory is sufficient: at time t, we need $s_{t-1}, a_{t-1}, b_{t-1}, s_t$.

Open question.

Assume that player 2 has an optimal strategy.
Then, player 2 has a stationary optimal strategy.

Let $U(s)$ denote the mixed actions for player 2 that are optimal in the one-shot game $M(s)$.

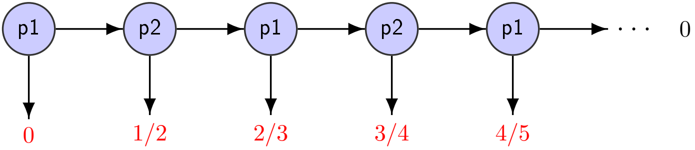
- A natural idea: take an element of $U(s)$ with a smallest support. But it does not work.

- Instead: We apply a difficult and technical result by Secchi [1998] (that for any $\varepsilon > 0$, player 2 has a stationary ε-optimal strategy).

Agenda

1. Zero–sum stochastic games with the limsup payoff
2. For player 1: simple optimal strategies
3. For player 2: simple optimal strategies
4. Extensions
5. The total payoff

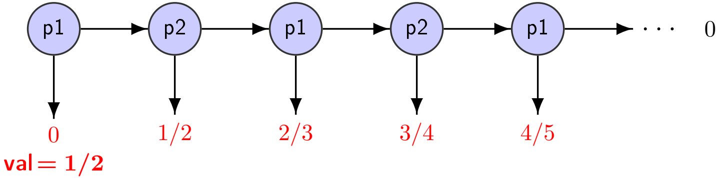
Infinite state space: An example



Player 1 has an optimal strategy: to continue (necessary !) in the initial state and to quit at a later state.

It is stationary, but depends on the initial state.

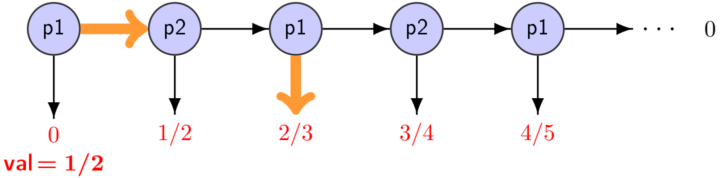
Infinite state space: An example



Player 1 has an optimal strategy: to continue (necessary !) in the initial state and to quit at a later state.

It is stationary, but depends on the initial state.

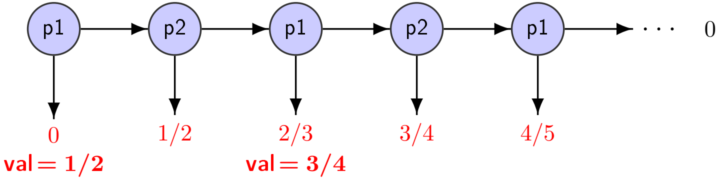
Infinite state space: An example



Player 1 has an optimal strategy: to continue (necessary !) in the initial state and to quit at a later state.

It is stationary, but depends on the initial state.

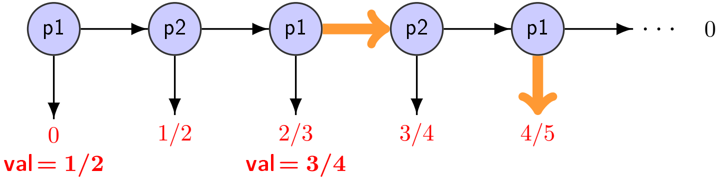
Infinite state space: An example



Player 1 has an optimal strategy: to continue (necessary !) in the initial state and to quit at a later state.

It is stationary, but depends on the initial state.

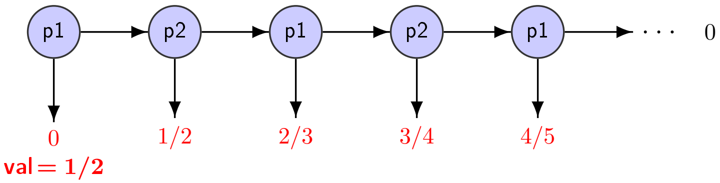
Infinite state space: An example



Player 1 has an optimal strategy: to continue (necessary !) in the initial state and to quit at a later state.

It is stationary, but depends on the initial state.

Infinite state space: An example



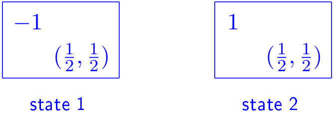
Player 1 has an optimal strategy: to continue (necessary !) in the initial state and to quit at a later state.

It is stationary, but depends on the initial state.

Even worse: Add a new state s^* such that from s^* we jump to each state with a positive probability. Then, for initial state s^*, there is no stationary optimal strategy !

Recall the payoff: $\mathbb{E}_{(s,\sigma,\tau)}\left[\limsup_{t\to\infty} r_t\right]$.

A related evaluation: $\limsup_{t\to\infty} \mathbb{E}_{(s,\sigma,\tau)}\left[r_t\right]$.



- $\mathbb{E}\left[\limsup_{t\to\infty} r_t\right] =$ **?**
  with prob 1 we visit state 2 infinitely often, so **it is equal to 1**

- $\limsup_{t\to\infty} \mathbb{E}\left[r_t\right] =$ **?**
  at each period $t \geq 2$ we have $\mathbb{E}\left[r_t\right] = 0$, so **it is equal to 0**

simple optimal strategies?

Agenda

1. Zero–sum stochastic games with the limsup payoff
2. For player 1: simple optimal strategies
3. For player 2: simple optimal strategies
4. Extensions
5. The total payoff

Total payoff: "considers the sum of the payoffs". So, for initial state s and strategy pair (σ, τ):

$$u^{+}(s,\sigma,\tau) \;=\; \mathbb{E}_{(s,\sigma,\tau)}\Big[\sum_{t=1}^{\infty} r_t\Big].$$

Assumptions:

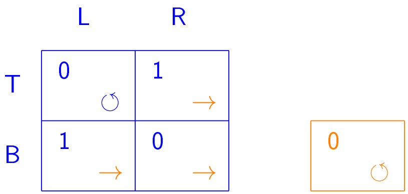
1) non-negative payoffs:

$$r(s,a,b) \geq 0 \qquad \forall s \in S,\ \forall a \in A(s),\ \forall b \in B(s).$$

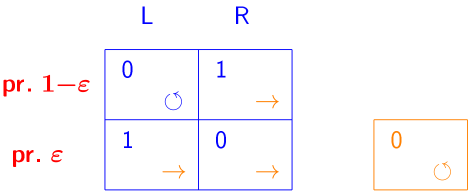
2) finite expected total payoffs:

$$u^{+}(s,\sigma,\tau) < \infty \qquad \forall s \in S,\ \forall \sigma,\ \forall \tau.$$

Example. Kumar and Shiau [1981], Maitra and Sudderth [1996]



Example. Kumar and Shiau [1981], Maitra and Sudderth [1996]



Consider the stationary strategy $(1-\varepsilon, \varepsilon)$, where $\varepsilon \in (0,1)$.

- If always action L: the total payoff is 1.
- If ever action R: the total payoff is $1-\varepsilon$.

So: The value is $v^{+}(1) = 1$. But player 1 has **no optimal strategy**.

Theorem. (Flesch, Predtetchinski and Sudderth)

Suppose that player 1 has an optimal strategy. Then, player 1 has a stationary optimal strategy.

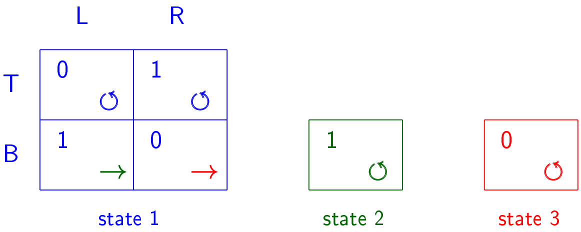
The proof is based on similar ideas.

Remark: Player 2 always has a stationary optimal strategy.
Parthasarathy [1973] and many other sources.

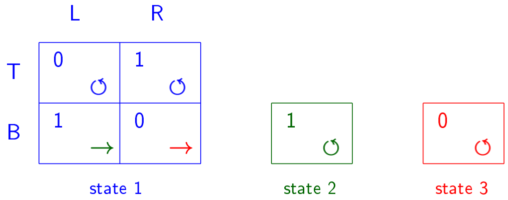
Final remark: probably simplification results are possible in other classes of games too.

THANK YOU

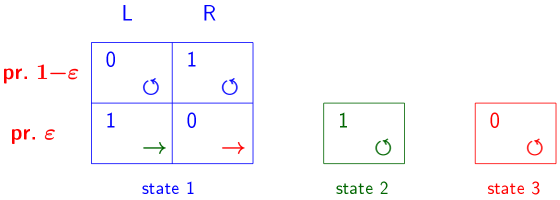
Example 3. The Big Match, Gillette [1957] – later Kristoffer's talk
But now with the limsup payoff !



- The value is $v(1) = 1$.
- Player 1 has no optimal strategy.
- For small $\varepsilon > 0$, player 1 has only history-dependent ε-optimal strategies.

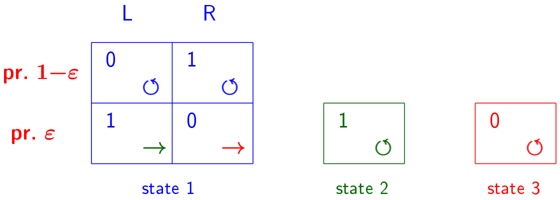


Fix $\varepsilon > 0$. Play in phases according to #R.



Fix $\varepsilon > 0$. Play in phases according to #R.

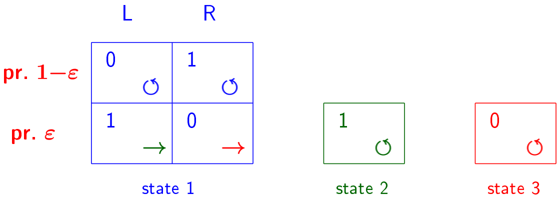
0 Play $(1-\varepsilon, \varepsilon)$ as long as only L



Fix $\varepsilon > 0$. Play in phases according to #R.

0 Play $(1-\varepsilon, \varepsilon)$ as long as only L

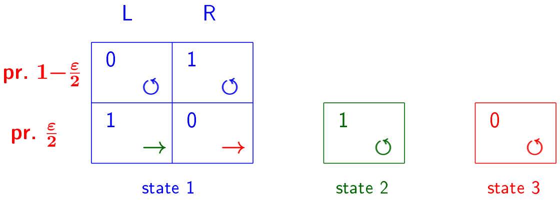
- If always L: we reach state 2



Fix $\varepsilon > 0$. Play in phases according to #R.

0 Play $(1-\varepsilon, \varepsilon)$ as long as only L

- If always L: we reach state 2
- If ever R:
  - with prob $1-\varepsilon$: payoff 1 and Next Phase
  - with prob ε: we reach state 3



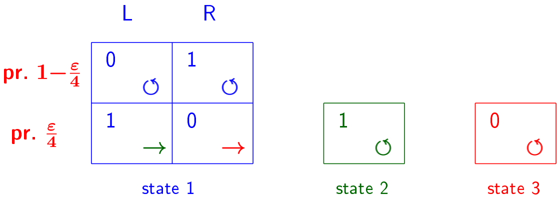
Fix $\varepsilon > 0$. Play in phases according to #R.

0 Play $(1-\varepsilon, \varepsilon)$ as long as only L

- If always L: we reach state 2
- If ever R:
  - with prob $1-\varepsilon$: payoff 1 and Next Phase
  - with prob ε: we reach state 3

1 Play $(1-\frac{\varepsilon}{2}, \frac{\varepsilon}{2})$ as long as only L

similar, except with prob $\frac{\varepsilon}{2}$: we reach state 3



Fix $\varepsilon > 0$. Play in phases according to #R.

0 Play $(1-\varepsilon, \varepsilon)$ as long as only L

- If always L: we reach state 2
- If ever R:
  - with prob $1-\varepsilon$: payoff 1 and Next Phase
  - with prob ε: we reach state 3

1 Play $(1-\frac{\varepsilon}{2}, \frac{\varepsilon}{2})$ as long as only L

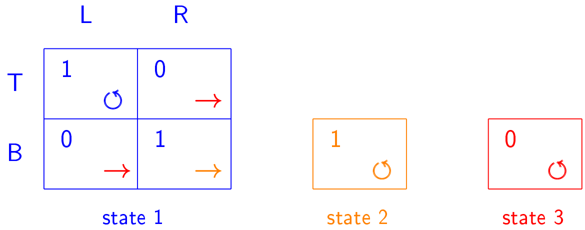
similar, except with prob $\frac{\varepsilon}{2}$: we reach state 3

2 Play $(1-\frac{\varepsilon}{4}, \frac{\varepsilon}{4})$ as long as only L

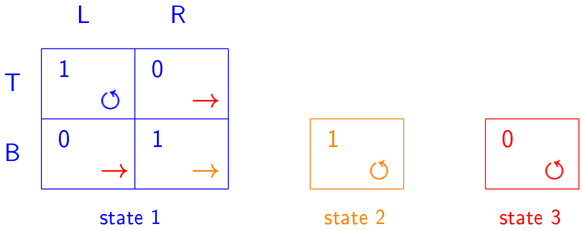
similar, except with prob $\frac{\varepsilon}{4}$: we reach state 3

3 and so on ...

Example: player 2 has no optimal strategy.



Example: player 2 has no optimal strategy.



For player 2, consider the stationary strategy $y = (1-\varepsilon, \varepsilon)$, where $\varepsilon \in (0, 1)$.

- If always action T: the limsup payoff is 0.
- If ever action B: the limsup payoff is ε.

So: The value is $v(1) = 0$. But player 2 has **no optimal strategy**.

Theorem. (Secchi 1998)

For any $\varepsilon > 0$, player 2 has a stationary ε-optimal strategy y such that for any initial state s and strategy σ for player 1

$$\mathbb{P}_{(s,\sigma,y)}\left[r_t \leq v(s_t) \text{ for all but finitely many } t\right] = 1.$$