# Strategy recovery for stochastic mean payoff games

Marcello Mamino
TU Dresden

GRASTA '15, October 19–23, 2015, Montreal

# Outline

- Stochastic games
- What is the solution of a game?
- Complexity of stochastic games
- Strategy recovery
- Proof

# Stochastic games

Definition (stochastic game)

- Two player 0-sum complete information game.
- Finite **directed graph** $G$, a token rests on one of the vertices.
- Each vertex $v$ has an **owner** $o(v)$ which is a player.
- Each directed edge $x \xrightarrow{A,p} y$ has an **action** $A \in \{a, b, c \ldots\}$ and a **probability** $p \in \mathbb{Q} \cap [0, 1]$.
- Each action $A$ has a **reward** $r(A) \in \mathbb{Q}$.
- Play starts at some vertex $v_0$.
- Play never ends.

# Stochastic games

A play of a stochastic game $G$ produces an infinite squence of vertices and actions

$$v_0 \xrightarrow{A_0} v_1 \xrightarrow{A_1} v_2 \xrightarrow{A_2} \quad \cdots$$

Definition
For $0 < \beta < 1$, the $\beta$-**discounted payoff** is

$$v_\beta(A_0, A_1 \dots) = (1-\beta)\sum_{i=0}^{\infty} r(A_i)\beta^i$$

The **mean payoff** is

$$v_1(A_0, A_1 \dots) = \liminf_{n\to\infty} \frac{1}{n+1}\sum_{i=0}^{n} r(A_i)$$

# Stochastic games

- Introduced by Gillette in 1957 generalizing Shapley.
- Used to model reactive systems with randomized and adversarial behaviour (*competitive Markov decision processes*).
- Pseudo-polynomial time algorithms in some cases (discounted payoff, *ergodic* mean payoff if most states are deterministic).
- No polynomial time algorithm known.

Theorem (Gillette '57, Liggett–Lippman '69)

Stochastic discounted payoff and mean payoff games are determined. Moreover, the optimal strategies are **positional**.

Corollary

Stochastic discounted payoff and mean payoff games are in NP $\cap$ co-NP

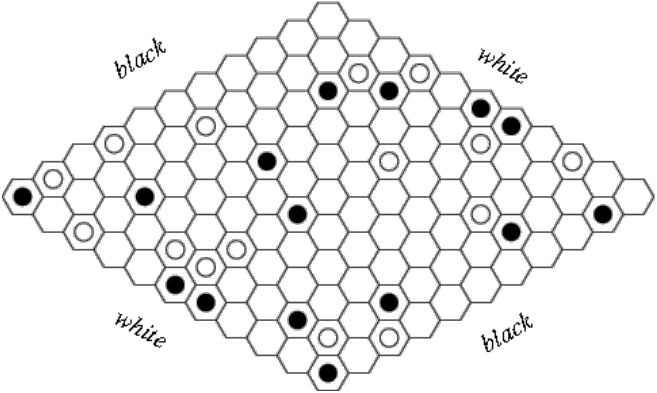
# What is the solution of a game?

Definition
We call **strategic solution** a pair of optimal strategies.

Definition
We call **quantitative solution** a method to evaluate all possible positions in a game.

Observation
If the plays of a class of games have **finite length**, then – *under reasonable hypotheses* – the problems of finding a strategic solution and a quantitative solution are **equivalent**.

# What is the solution of a game?



# What is the solution of a game?

Definition
We call **strategic solution** a pair of optimal strategies.

Definition
We call **quantitative solution** a method to evaluate all possible positions in a game.

Observation
If the plays of a class of games have **finite length**, then – *under reasonable hypotheses* – the problems of finding a strategic solution and a quantitative solution are **equivalent**.

# What is the solution of a game?

Observation

In **general**, to find a quantitative solution, given a strategic solution, is not harder than playing two strategies against each other (**quantitative** $\prec$ **strategic**).

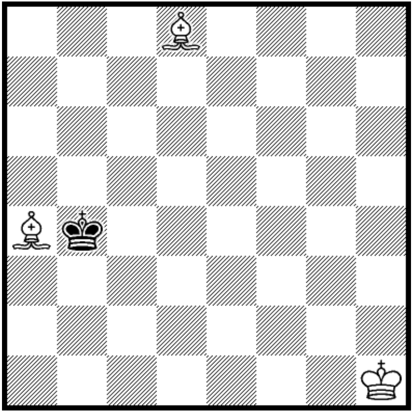
Fact

There are **inperfect information** stochastic games whose $\epsilon$-optimal strategies require **exponential space** to be represented in binary.

Question (strategy recovery)

Given the quantitative solution of a specific game, how hard is it to derive a strategic solution?

# What is the solution of a game?



# What is the solution of a game?

Observation

In **general**, to find a quantitative solution, given a strategic solution, is not harder than playing two strategies against each other (**quantitative** $\prec$ **strategic**).

Fact

There are **inperfect information** stochastic games whose $\epsilon$-optimal strategies require **exponential space** to be represented in binary.

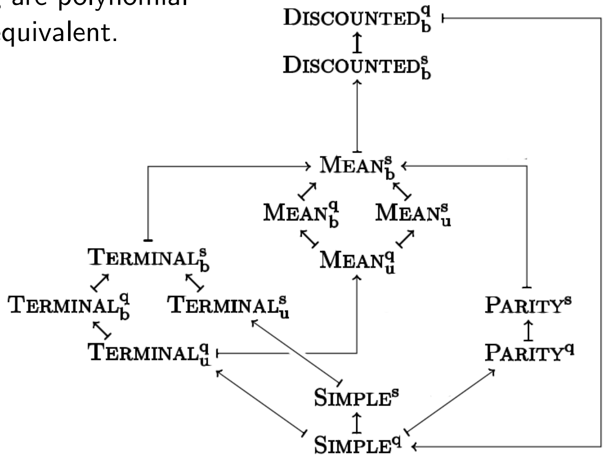
Question (strategy recovery)

Given the quantitative solution of a specific game, how hard is it to derive a strategic solution?

# Complexity of stochastic games

Theorem (Andersson–Miltersen '09)

The following are polynomial time Turing equivalent.



# Strategy recovery

Observation
For **discounted** payoff stochastic games strategy recovery can be performed in linear time.

Theorem (Andersson–Miltersen '09)
Strategy recovery for **terminal** and **simple** stochastic games can be done in linear time.

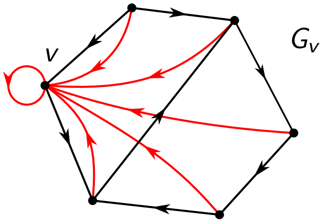
Theorem
For **mean** payoff stochastic games, strategy recovery is as hard as it possibly can, namely polynomial time Turing **equivalent to strategic solution.**

Idea of the proof: reduce all stochastic mean payoff games to a subclass of games with the property that, by a reason of symmetry, all positions have expected value zero.

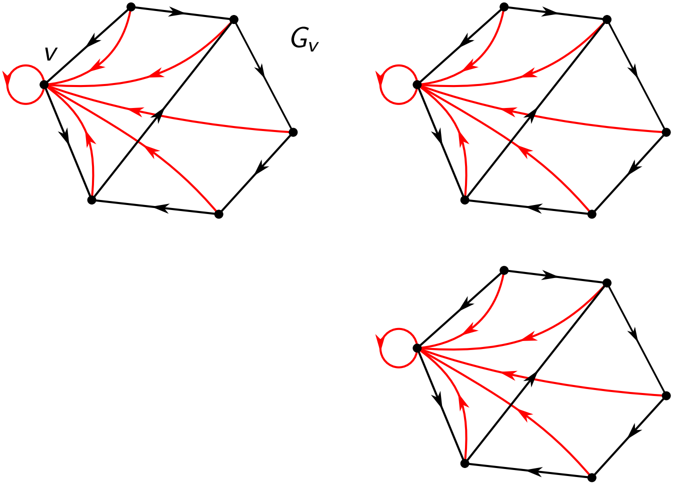
# Steps of the proof

1. The mean payoff game on $G$ is strategically equivalent to the $\beta$-discounted game on $G$ for $\beta$ close enough to 1.
2. Fix a vertex $v$ of $G$ and replace all edges $x \xrightarrow{A,p} y$ with $x \xrightarrow{A,\beta p} y$ and $x \xrightarrow{A,(1-\beta)p} v$, yielding a new game $G_v$.
3. This immediately forces the expected mean payoff of all initial positions of $G_v$ to be the same.
4. Moreover the expected mean payoff of $G_v$ coincides with the expected $\beta$-discounted value of $G$ starting at $v$.
5. Summarizing, if we can find optimal strategies for all $G_v$, then we can evaluate all $G_v$, hence we can compute the $\beta$-discounted value of all positions in $G$, and by a previous observation we can compute optimal $\beta$-discounted strategies, which coincide with optimal mean payoff strategies.

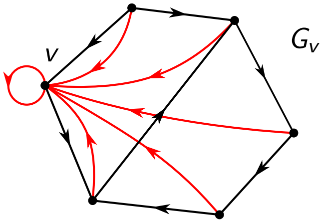
# Steps of the proof

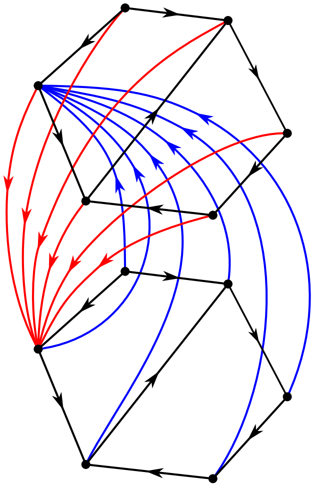


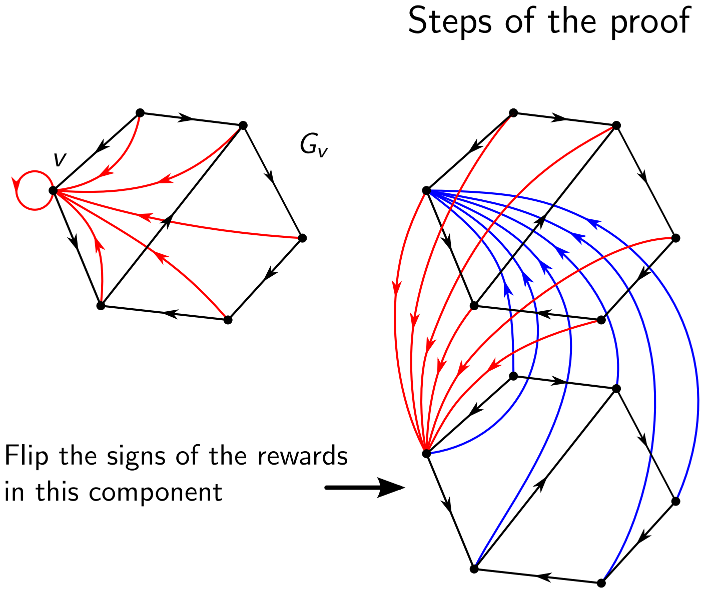
# Steps of the proof



# Steps of the proof




Steps of the proof
v
G_v
Flip the signs of the rewards in this component

Thank you!