Solving Concurrent Games in PSPACE

Patrick Totzke

Concurrent Games

á la Shapley [S'53], Everett [E'57], Gillette [G'57]:

- Played on a finite directed graph;
- time is discrete and not bounded;
- finite action sets;

At any time, players simultaneously chose an action each which, together with the current vertex, determines where to move next and a payoff vector.

$\rightsquigarrow$ probability space over infinite paths and expected payoffs.

Concurrent Games

á la Shapley [S'53], Everett [E'57], Gillette [G'57]:

- Played on a finite directed graph;
- time is discrete and not bounded;
- finite action sets;

At any time, players simultaneously chose an action each which, together with the current vertex, determines where to move next and a payoff vector.

$\rightsquigarrow$ probability space over infinite paths and expected payoffs.

Objectives: discounted sum; reachability; parity; mean-payoff

Concurrent Games

á la Shapley [S'53], Everett [E'57], Gillette [G'57]:

- Played on a finite directed graph;
- time is discrete and not bounded;
- finite action sets;

At any time, players simultaneously chose an action each which, together with the current vertex, determines where to move next and a payoff vector.

$\rightsquigarrow$ probability space over infinite paths and expected payoffs.

Objectives: discounted sum; reachability; parity; mean-payoff

Concurrent Games

á la Shapley [S'53], Everett [E'57], Gillette [G'57]:

- Played on a finite directed graph;
- time is discrete and not bounded;
- finite action sets;

At any time, players simultaneously chose an action each which, together with the current vertex, determines where to move next and a payoff vector.

$\rightsquigarrow$ probability space over infinite paths and expected payoffs.

Objectives: discounted sum; reachability; parity; mean-payoff

Concurrent Games

á la Shapley [S'53], Everett [E'57], Gillette [G'57]:

- Played on a finite directed graph;
- time is discrete and not bounded;
- finite action sets;

At any time, players simultaneously chose an action each which, together with the current vertex, determines where to move next and a payoff vector.

$\rightsquigarrow$ probability space over infinite paths and expected payoffs.

Objectives: discounted sum; reachability; parity; mean-payoff

Concurrent Games

á la Shapley [S'53], Everett [E'57], Gillette [G'57]:

- Played on a finite directed graph;
- time is discrete and not bounded;
- finite action sets;

At any time, players simultaneously chose an action each which, together with the current vertex, determines where to move next and a payoff vector.

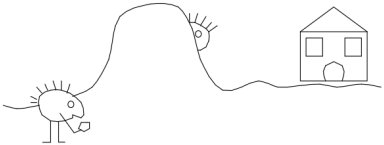
$\rightsquigarrow$ probability space over infinite paths and expected payoffs.

Objectives: discounted sum; reachability; parity; mean-payoff $= \liminf_{n\to\infty} \sum_{i=1}^{n} r_n/n$

Concurrent Reachability Games

Example: Snowball game [dAHK'98]

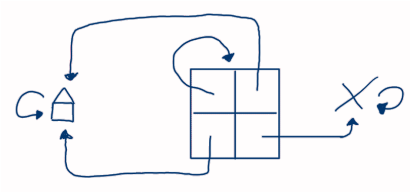
- two players simultaneously chose actions (T/W, R/H);
- P1 wins on T+H or W+R, P2 wins on T+R, else repeat



Concurrent Reachability Games

Example: Snowball game [dAHK'98]

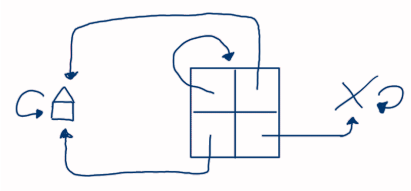
- two players simultaneously chose actions (T/W, R/H);
- P1 wins on T+H or W+R, P2 wins on T+R, else repeat



Concurrent Reachability Games

Example: Snowball game [dAHK'98]

- two players simultaneously chose actions (T/W, R/H);
- P1 wins on T+H or W+R, P2 wins on T+R, else repeat



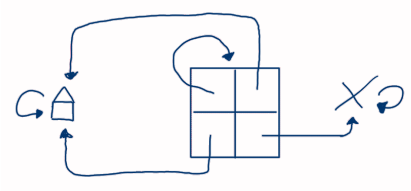
In general,

- zero-sum values exist but optimal strategies may not
- values may not be rational, so can only be approximated [S'53]
- ε-optimal stationary and rational-valued strategies exist [E'57] but require doubly-exponentially small probabilities ($\varepsilon^{2^{\mathcal{O}(N)}}$)[HKM'09]
- games can be solved by reduction to $\mathrm{FO}(\mathbb{R})$, value or strategy iteration, in EXP.

Concurrent Reachability Games

Example: Snowball game [dAHK'98]

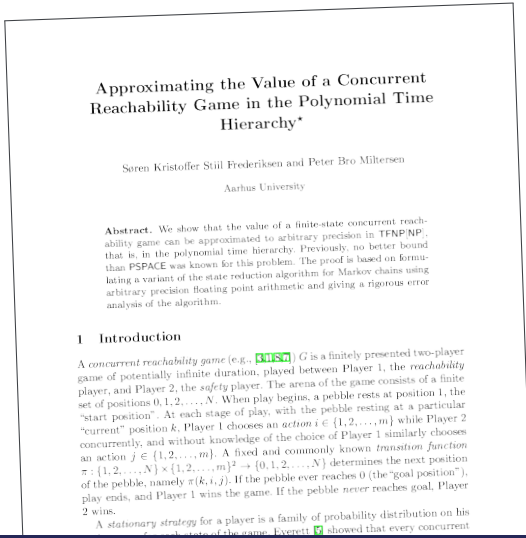
- two players simultaneously chose actions (T/W, R/H);
- P1 wins on T+H or W+R, P2 wins on T+R, else repeat



In general,

- zero-sum values exist but optimal strategies may not
- values may not be rational, so can only be approximated [S'53]
- ε-optimal stationary and rational-valued strategies exist [E'57] but require doubly-exponentially small probabilities ($\varepsilon^{2^{\mathcal{O}(N)}}$)[HKM'09]
- games can be solved by reduction to $\mathrm{FO}(\mathbb{R})$, value or strategy iteration, in EXP.

Solving Concurrent Reachability Games in $\mathsf{FNP}^{\mathsf{NP}}$



Approximating the Value of a Concurrent Reachability Game in the Polynomial Time Hierarchy*

Søren Kristoffer Stiil Frederiksen and Peter Bro Miltersen

Aarhus University

Abstract. We show that the value of a finite-state concurrent reachability game can be approximated to arbitrary precision in $\mathsf{TFNP}^{\mathsf{NP}}$, that is, in the polynomial time hierarchy. Previously, no better bound than PSPACE was known for this problem. The proof is based on formulating a variant of the state reduction algorithm for Markov chains using arbitrary precision floating point arithmetic and giving a rigorous error analysis of the algorithm.

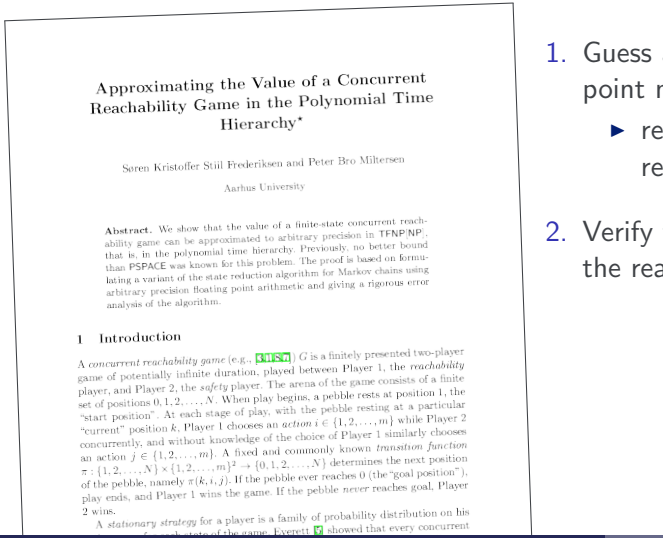
1 Introduction

A *concurrent reachability game* (e.g., [3,4,18,7]) G is a finitely presented two-player game of potentially infinite duration, played between Player 1, the *reachability* player, and Player 2, the *safety* player. The arena of the game consists of a finite set of positions $0, 1, 2, \ldots, N$. When play begins, a pebble rests at position 1, the "start position". At each stage of play, with the pebble resting at a particular "current" position k, Player 1 chooses an *action* $i \in \{1, 2, \ldots, m\}$ while Player 2 concurrently, and without knowledge of the choice of Player 1 similarly chooses an action $j \in \{1, 2, \ldots, m\}$. A fixed and commonly known *transition function* $\pi : \{1, 2, \ldots, N\} \times \{1, 2, \ldots, m\}^2 \to \{0, 1, 2, \ldots, N\}$ determines the next position of the pebble, namely $\pi(k, i, j)$. If the pebble ever reaches 0 (the "goal position"), play ends, and Player 1 wins the game. If the pebble *never* reaches goal, Player 2 wins.

A *stationary strategy* for a player is a family of probability distribution on his

1. Guess a strategy profile (and value) in floating point notations $x \cdot 10^y$ with $x, y \leq 2^{\mathcal{O}(n)}$
2. Verify that the resulting Markov chain satisfies the reachability probability

Solving Concurrent Reachability Games in $\mathsf{FNP}^{\mathsf{NP}}$



Approximating the Value of a Concurrent Reachability Game in the Polynomial Time Hierarchy*

Søren Kristoffer Stiil Frederiksen and Peter Bro Miltersen

Aarhus University

Abstract. We show that the value of a finite-state concurrent reachability game can be approximated to arbitrary precision in $\mathsf{TFNP}^{\mathsf{NP}}$, that is, in the polynomial time hierarchy. Previously, no better bound than PSPACE was known for this problem. The proof is based on formulating a variant of the state reduction algorithm for Markov chains using arbitrary precision floating point arithmetic and giving a rigorous error analysis of the algorithm.

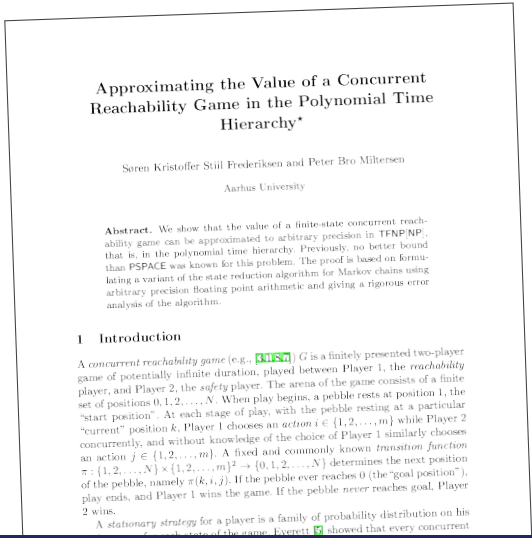
1 Introduction

A *concurrent reachability game* (e.g., [3,4,18,7]) G is a finitely presented two-player game of potentially infinite duration, played between Player 1, the *reachability* player, and Player 2, the *safety* player. The arena of the game consists of a finite set of positions $0, 1, 2, \ldots, N$. When play begins, a pebble rests at position 1, the "start position". At each stage of play, with the pebble resting at a particular "current" position k, Player 1 chooses an *action* $i \in \{1, 2, \ldots, m\}$ while Player 2 concurrently, and without knowledge of the choice of Player 1 similarly chooses an action $j \in \{1, 2, \ldots, m\}$. A fixed and commonly known *transition function* $\pi : \{1, 2, \ldots, N\} \times \{1, 2, \ldots, m\}^2 \to \{0, 1, 2, \ldots, N\}$ determines the next position of the pebble, namely $\pi(k, i, j)$. If the pebble ever reaches 0 (the "goal position"), play ends, and Player 1 wins the game. If the pebble *never* reaches goal, Player 2 wins.

A *stationary strategy* for a player is a family of probability distribution on his

1. Guess a strategy profile (and value) in floating point notations $x \cdot 10^y$ with $x, y \leq 2^{\mathcal{O}(n)}$
   - requires "approximately normalised" representations of distributions
2. Verify that the resulting Markov chain satisfies the reachability probability

Solving Concurrent Reachability Games in $\mathsf{FNP}^{\mathsf{NP}}$



Approximating the Value of a Concurrent Reachability Game in the Polynomial Time Hierarchy*

Søren Kristoffer Stiil Frederiksen and Peter Bro Miltersen

Aarhus University

Abstract. We show that the value of a finite-state concurrent reachability game can be approximated to arbitrary precision in $\mathsf{TFNP}^{\mathsf{NP}}$, that is, in the polynomial time hierarchy. Previously, no better bound than PSPACE was known for this problem. The proof is based on formulating a variant of the state reduction algorithm for Markov chains using arbitrary precision floating point arithmetic and giving a rigorous error analysis of the algorithm.

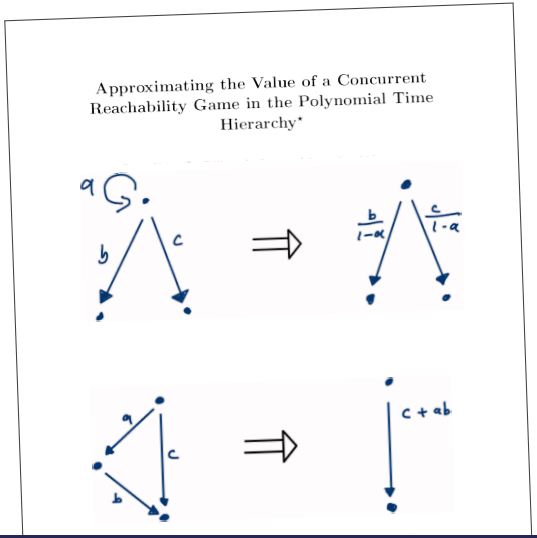
1 Introduction

A *concurrent reachability game* (e.g., [3,4,18,7]) G is a finitely presented two-player game of potentially infinite duration, played between Player 1, the *reachability* player, and Player 2, the *safety* player. The arena of the game consists of a finite set of positions $0, 1, 2, \ldots, N$. When play begins, a pebble rests at position 1, the "start position". At each stage of play, with the pebble resting at a particular "current" position k, Player 1 chooses an *action* $i \in \{1, 2, \ldots, m\}$ while Player 2 concurrently, and without knowledge of the choice of Player 1 similarly chooses an action $j \in \{1, 2, \ldots, m\}$. A fixed and commonly known *transition function* $\pi : \{1, 2, \ldots, N\} \times \{1, 2, \ldots, m\}^2 \to \{0, 1, 2, \ldots, N\}$ determines the next position of the pebble, namely $\pi(k, i, j)$. If the pebble ever reaches 0 (the "goal position"), play ends, and Player 1 wins the game. If the pebble *never* reaches goal, Player 2 wins.

A *stationary strategy* for a player is a family of probability distribution on his

1. Guess a strategy profile (and value) in floating point notations $x \cdot 10^y$ with $x, y \leq 2^{\mathcal{O}(n)}$
   - requires "approximately normalised" representations of distributions
2. Verify that the resulting Markov chain satisfies the reachability probability
   - state-space reduction uses few operations, each introduces a bounded error.

Solving Concurrent Reachability Games in $\mathsf{FNP}^{\mathsf{NP}}$

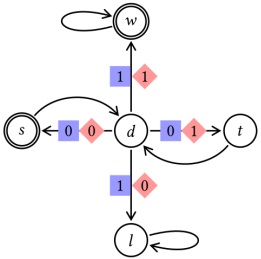

Approximating the Value of a Concurrent Reachability Game in the Polynomial Time Hierarchy*

1. Guess a strategy profile (and value) in floating point notations $x \cdot 10^y$ with $x, y \leq 2^{\mathcal{O}(n)}$
   - requires "approximately normalised" representations of distributions
2. Verify that the resulting Markov chain satisfies the reachability probability
   - state-space reduction uses few operations, each introduces a bounded error.

Concurrent Parity Games

Example: The Bad Match [T'92, MS'96]

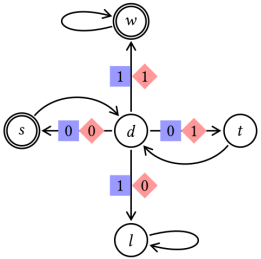
- two players simultaneously chose 0 or 1;
- if P1 picks 1 both choices are fixed forever
- (only) matches are accepting;



Concurrent Parity Games

Example: The Bad Match [T'92, MS'96]

- two players simultaneously chose 0 or 1;
- if P1 picks 1 both choices are fixed forever
- (only) matches are accepting;

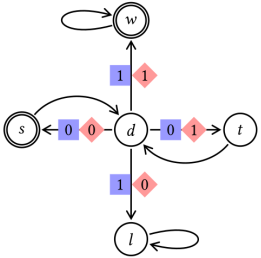


Value(d)=1. No optimal strategy exists and no FM or Markov strategy is ε-optimal.

Concurrent Parity Games

Example: The Bad Match [T'92, MS'96]

- two players simultaneously chose 0 or 1;
- if P1 picks 1 both choices are fixed forever
- (only) matches are accepting;



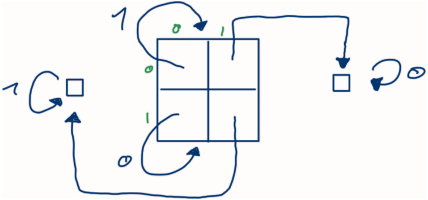
Value(d)=1. No optimal strategy exists and no FM or Markov strategy is ε-optimal.

- Markov+1bit ε-optimal strategies always exists for Büchi Games [KMST'24]
- Parity games reduce to stateful-discounted games [dAHM'03, GZ'06], which can be solved in $\mathrm{FNP}^{\mathrm{NP}}$ [ACSS'24].

Concurrent Mean-Payoff Games

Example: The Big Match [BF'68]

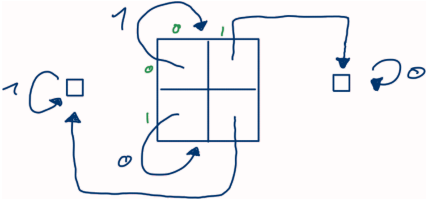
- two players simultaneously chose 0 or 1;
- if P2 picks 1 both choices are fixed forever;
- P1 gains 1 on a match, 0 otherwise.



Concurrent Mean-Payoff Games

Example: The Big Match [BF'68]

- two players simultaneously chose 0 or 1;
- if P2 picks 1 both choices are fixed forever;
- P1 gains 1 on a match, 0 otherwise.



In general,

- zero-sum values exist but may be irrational [S'53]
- optimal strategies and NE need not exist! [E'57]
- stationary ε-optimal need not exist! [BF'68]

Status Quo: Solving Concurrent MPG

- Mertens & Neymann '89: Value iteration $\varepsilon^{-m^{\mathcal{O}(n)}}$
- Hansen, Koucjy, Lauritzen, Miltersen, Tsigaridas'11: Recursive binary search $\log(\varepsilon^{-1})m)^{2^{\mathcal{O}(n)}}$
- Oliu-Barton '20: Binary search $\log(\varepsilon^{-1})m^{\mathcal{O}(n)}$
- Hansen, Koucky, Lauritzen, Miltersen'11: For any discount factor $\varepsilon^{m^{\mathcal{O}(n)}}$ the discounted game approximates the value of the mean-payoff game within ε. Etessami, Yannakakis '05 lets us find the value of a discounted game in $\exists\mathrm{FO}(\mathbb{R})$

Finding good Bounded Strategies

Definition

A strategy ε-*achieves* value $v \in \mathbb{R}$ with respect to $\mathcal{S}$ if it guarantees expected outcome of at least $v - \varepsilon$ against all opponent strategies in $\mathcal{S}$.

Let's consider bounded-memory strategies.

Theorem

The following is in $\mathsf{FNP}^{\mathsf{NP}}$*. Given a game, bounds* $\mathbf{b} \in \mathbb{N}^k$*,* $\varepsilon \in (0, 1/2)$ *and* $v \in \mathbb{R}$*, find a strategy that* ε*-achieves* v *with respect to* $\mathcal{B}(\mathbf{b})$ *if for some* $\varepsilon' \in [0, \varepsilon/3)$*, some strategy* ε'*-achieves* v *with respect to* $\mathcal{B}(\mathbf{b})$*.*

We assume $\mathbf{b}$ is unary, ε is binary, and transition matrices are in ANR floating points.

Proof Ideas

We follow the approach of [FM'13] as much as possible.

Part 1 (guess profiles):
Extra Lemma (**): If good profiles exist, then also one in ANR.

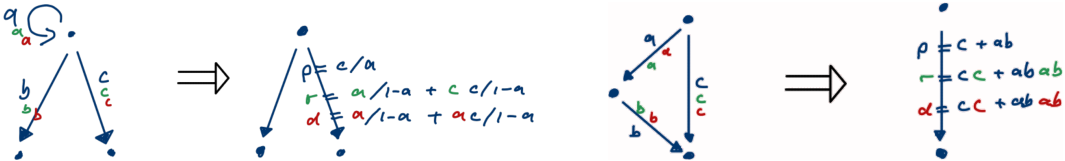
Part 2 (approximate values in MCs): via state-reductions

Proof Ideas

We follow the approach of [FM'13] as much as possible.

Part 1 (guess profiles):
Extra Lemma (**): If good profiles exist, then also one in ANR.

Part 2 (approximate values in MCs): via state-reductions that preserve probabilities, rewards and durations.

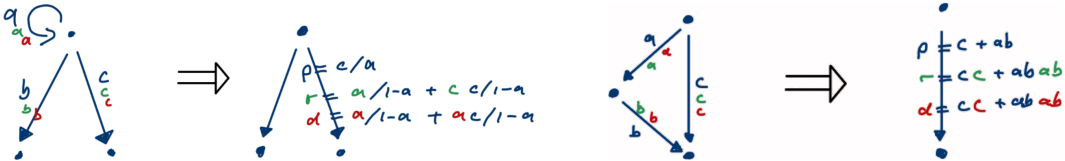


Proof Ideas

We follow the approach of [FM'13] as much as possible.

Part 1 (guess profiles):
Extra Lemma (**): If good profiles exist, then also one in ANR.

Part 2 (approximate values in MCs): via state-reductions that preserve probabilities, rewards and durations.



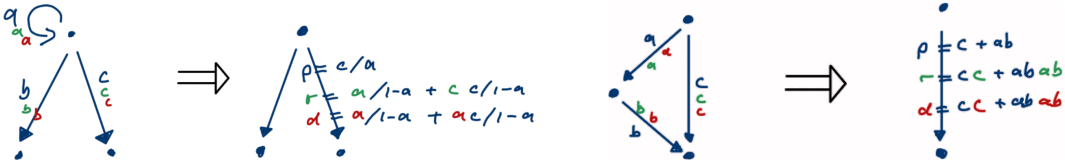
and therefore the "mpo-ratio" $\liminf_{n\to\infty} \sum_{i=1}^{n} r_i / \sum_{i=1}^{n} d_i$

Proof Ideas

We follow the approach of [FM'13] as much as possible.

Part 1 (guess profiles):
Extra Lemma (**): If good profiles exist, then also one in ANR.

Part 2 (approximate values in MCs): via state-reductions that preserve probabilities, rewards and durations.



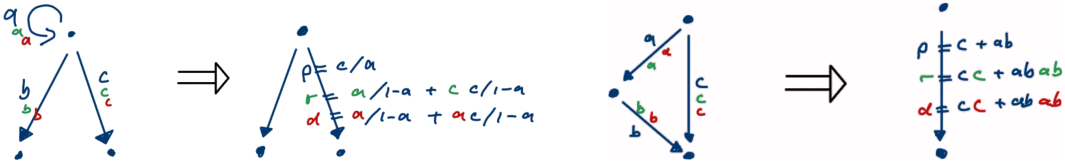
and therefore the "mpo-ratio" $\liminf_{n\to\infty} \sum_{i=1}^{n} r_i / \sum_{i=1}^{n} d_i = mpv$.

Proof Ideas

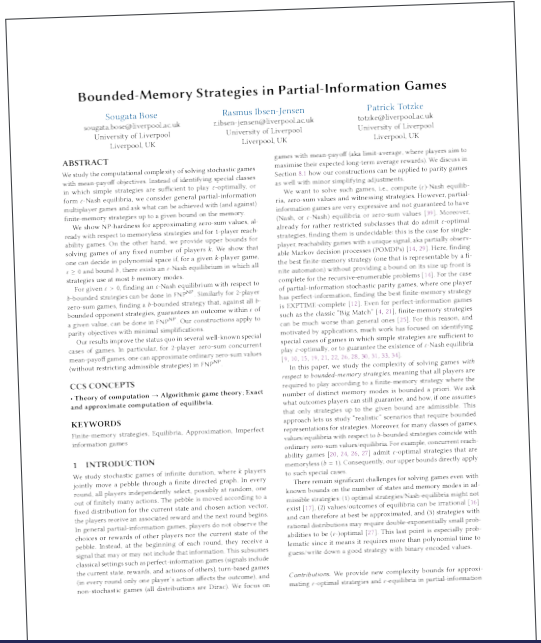
We follow the approach of [FM'13] as much as possible.

Part 1 (guess profiles):
Extra Lemma (**): If good profiles exist, then also one in ANR.

Part 2 (approximate values in MCs): via state-reductions that preserve probabilities, rewards and durations.



and therefore the "mpo-ratio" $\liminf_{n\to\infty} \sum_{i=1}^{n} r_i / \sum_{i=1}^{n} d_i = mpv$.
This can be encoded in $\mathrm{FO}(\mathbb{R})$, + perturbation lemma gives (**).



Bounded-Memory Strategies in Partial-Information Games

Sougata Bose
sougata.bose@liverpool.ac.uk
University of Liverpool
Liverpool, UK

Rasmus Ibsen-Jensen
ribsen-jensen@liverpool.ac.uk
University of Liverpool
Liverpool, UK

Patrick Totzke
totzke@liverpool.ac.uk
University of Liverpool
Liverpool, UK

ABSTRACT

We study the computational complexity of solving stochastic games with mean-payoff objectives. Instead of identifying special classes in which simple strategies are sufficient to play ε-optimally, or form ε-Nash equilibria, we consider general partial-information multiplayer games and ask what can be achieved with (and against) finite-memory strategies up to a given bound on the memory.

We show NP-hardness for approximating zero-sum values, already with respect to memoryless strategies and for 1-player reachability games. On the other hand, we provide upper bounds for solving games of any fixed number of players k. We show that one can decide in polynomial space if, for a given 2-player game, $\varepsilon \geq 0$ and bound b there exists an ε-Nash equilibrium in which all strategies use at most b memory modes.

For given $\varepsilon > 0$, finding an ε-Nash equilibrium with respect to b-bounded strategies can be done in $\mathsf{FNP}^{\mathsf{NP}}$. Similarly for 2-player zero-sum games, finding a b-bounded strategy that, against all b-bounded opponent strategies, guarantees an outcome within ε of a given value, can be done in $\mathsf{FNP}^{\mathsf{NP}}$. Our constructions apply to parity objectives with minimal simplifications.

Our results improve the status quo in several well-known special cases of games. In particular, for 2-player zero-sum concurrent mean-payoff games one can approximate ordinary zero-sum values (without restricting admissible strategies) in $\mathsf{FNP}^{\mathsf{NP}}$.

CCS CONCEPTS

• **Theory of computation → Algorithmic game theory; Exact and approximate computation of equilibria.**

KEYWORDS

Finite-memory strategies, Equilibria, Approximation, Imperfect information games

1 INTRODUCTION

We study stochastic games of infinite duration, where k players jointly move a pebble through a finite directed graph. In every round, all players independently select, possibly at random, one out of finitely many actions. The pebble is moved according to a fixed distribution for the current state and chosen action vector. The players receive an associated reward and the next round begins. In general partial-information games, players do not observe the choices or rewards of other players nor the current state of the pebble. Instead, at the beginning of each round, they receive a signal that may or may not include that information. This subsumes classical settings such as perfect-information games (signals include the current state, rewards, and actions of others), turn-based games (in every round only one player's action affects the outcome), and non-stochastic games (all distributions are Dirac). We focus on games with mean-payoff (aka limit-average, where players aim to maximise their expected long-term average rewards). We discuss in Section 3.1 how our constructions can be applied to parity games as well with minor simplifying adjustments.

We want to solve such games i.e., compute (ε)-Nash equilibria, zero-sum values and witnessing strategies. However, partial-information games are very expensive and not guaranteed to have (Nash, or ε-Nash) equilibria or zero-sum values [39]. Moreover, already for rather restricted subclasses that do admit ε-optimal strategies, finding them is undecidable: this is the case for single-player reachability games with a unique signal aka partially observable Markov decision processes (POMDPs) [14, 29]. Here, finding the best finite-memory strategy (one that is representable by a finite automaton) without providing a bound on its size up front is complete for the recursive-enumerable problems [14]. For the case of partial-information mean-payoff games, where one player has perfect information, finding the best finite-memory strategy is EXPTIME-complete [12]. Even for perfect-information games such as the classic "Big Match" [1, 21], finite-memory strategies can be much worse than general ones [25]. For this reason, and motivated by applications, much work has focused on identifying special cases of games in which simple strategies are sufficient to play ε-optimally, or to guarantee the existence of ε-Nash equilibria [3, 10, 13, 18, 21, 22, 26, 28, 30, 33, 34].

In this paper, we study the complexity of solving games with respect to *bounded-memory strategies* meaning that all players are required to play according to a finite-memory strategy where the number of distinct memory modes is bounded a priori. We ask what outcomes players can still guarantee, and how, if one restricts that only strategies up to the given bound are admissible. This approach lets us study "realistic" scenarios that require bounded representations for strategies. Moreover, for many classes of games values/equilibria with respect to b-bounded strategies coincide with ordinary zero-sum values/equilibria. For example, concurrent reachability games [19, 24, 26, 27] admit ε-optimal strategies that are memoryless ($b = 1$). Consequently, our upper bounds directly apply to such special cases.

There remain significant challenges for solving games even with known bounds on the number of states and memory modes in admissible strategies: (1) optimal strategies/Nash equilibria might not exist [17], (2) values/outcomes of equilibria can be irrational [36] and can therefore at best be approximated, and (3) strategies with rational distributions may require double-exponentially small probabilities to be (ε-)optimal [27]. This last point is especially problematic since it means it requires more than polynomial time to guess-write down a good strategy with binary encoded values.

Contributions. We provide new complexity bounds for approximating ε-optimal strategies and ε-equilibria in partial-information

This works for any fixed number k of players with individual MP objectives and partial information!

Implications

- zero-sum values in (perfect-info) MPG can be approximated, existence of memoryless (ε-)NE can be checked in FNP[NP]. (via DSGs with 2-exp discounts)
- results apply to parity with minimal adjustments in SSRs
- Stay-in-a-set games
- Quitting games