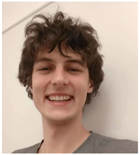
Concurrent Stochastic Games with Stateful-discounted and Parity Objectives: Complexity and Algorithms


A. Asadi[1]


K. Chatterjee[1]


R. Saona[1]

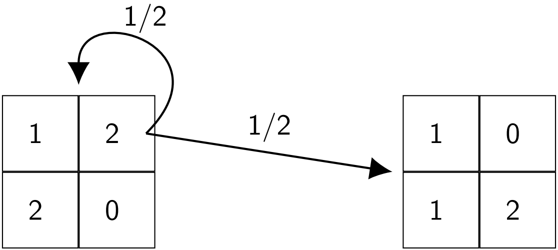

J. Svoboda[1]

[1]Institute of Science and Technology Austria (ISTA)

FSTTCS 2024

Institute of Science and Technology Austria

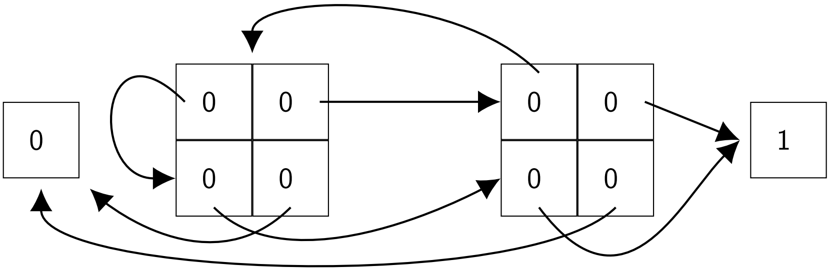
Game



Concurrent zero-sum stochastic game

States: s_1, s_2
2 opposite players
Actions: a, b
Rewards per state and action profile: $r(s, a, b)$
Stochastic transitions: $\delta(s, a, b)$

Deterministic game



Concurrent zero-sum deterministic game

Objectives and values

- **Reachability**: Probability of reaching a state
- **Discounted**: Discounted sum of rewards
- **Parity**: ω-regular objective
- **Stateful discounted**:
  State-dependent discounted sum of rewards

$$\mathrm{Disc}_\Lambda((s_i, a_i, b_i)_{i \geq 0}) := \sum_{i \geq 0} \left(r(s_i, a_i, b_i) \Lambda(s_i) \prod_{j < i} 1 - \Lambda(s_j) \right)$$

$$\mathrm{val}_\Lambda(s) := \sup_{\sigma \in \Sigma^S} \inf_{\tau \in \Gamma^S} \mathbb{E}_s^{\sigma,\tau} [\mathrm{Disc}_\Lambda] \,.$$

- **Limit (Stateful discounted) value**:
  Vanishing state-dependent discounted sum of rewards

$$\mathrm{val}_\chi(s) := \lim_{\lambda_1 \to 0^+} \cdots \lim_{\lambda_d \to 0^+} \mathrm{val}_\Lambda(s) \,.$$

Limit Value approximation.

How hard is to approximate the limit value?

Hint: Parity is a special case of limit value.

Our contribution

Previous.

- **Limit value**: EXPSPACE upper bound and double exponential time algorithm
- **Parity**: PSPACE upper bound and exponential time algorithm

Our contribution.

For both values,

- TFNP[NP] upper bound
- Exponential time algorithm, which is polynomial for fixed number of states

Hardness

Why approximation and not exact value computation?
Even for reachability objectives,

- **Irrational value**: exact value is irrational.
- **SQRT-SUM hardness**: at least as hard as SQRT-SUM, which is not known to be in NP.

What is the problem with guessing ε-optimal strategies?
Even for reachability objectives in deterministic games,

- **Double exponential patience**: ε-optimal strategies require very small numbers.

Previous work

Theorem (Kristoffer et. al., 2013)

Approximating the value of concurrent stochastic games with reachability objectives is in TFNP[NP].

Theorem (Attia and Oliu-Barton, 2019)

Approximating the undiscounted value (vanishing discounted value) of concurrent stochastic games can be done in exponential space and time.

Simplification

Theorem (Alfaro et. al. 2003; Gimbert and Zielonka, 2005)

There is a linear reduction from the computational problems of approximating the parity value to the approximation of the limit value of stateful-discounted objectives.

Technical tool

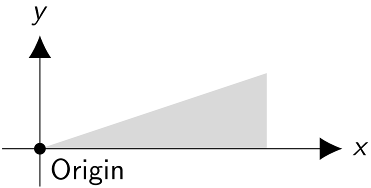
Our technical contribution

Lemma

Consider a nonzero polynomial P in $x_1, \cdots, x_\ell$ of degrees $D_1, \cdots, D_\ell$ with integer coefficients of bit-size B. Let $D := \max(D_1, \cdots, D_\ell)$ and $\mathrm{B}_1 := 4\ell\,\mathrm{bit}(D) + \mathrm{B} + 1$. Then,

$$\forall\, x_1 \in (0, \exp(-\mathrm{B}_1)] \quad \cdots \quad \forall\, x_\ell \in \left(0, (x_{\ell-1})^{D+1}\right]$$

$$|P(x_1, \cdots, x_\ell)| \geq \exp(\mathrm{B}_1 - \ell) \cdot (x_\ell)^{D+1}\,.$$



Root free zone in 2 dimensions

Ideas

Simple model: discounted reachability objective

Consider the reachability objective as the limit of the following discounted reachability objective.

Discounted reachability.

$$\mathrm{Disc}_\lambda((s_i)_{i\geq 0}) \coloneqq \sum_{i\geq 0} \mathbb{1}[s_i = \top]\, \lambda(1-\lambda)^{(i-1)_+}$$

$$\mathrm{val}_\lambda(s) \coloneqq \sup_{\sigma\in\Sigma^S} \inf_{\tau\in\Gamma^S} \mathbb{E}_s^{\sigma,\tau}[\mathrm{Disc}_\Lambda]\,.$$

Idea: Algorithm

Algorithm

The discounted value is characterized as the unique parameter that makes a parameterized matrix game have value zero.

Bellman fixpoint equation. Fixing stationary strategies σ, τ we obtain a Markov chain with

- **payoff**: $\nu^{\sigma,\tau}(s) := \mathbb{E}_s^{\sigma,\tau}[\mathrm{Disc}_\lambda]$.
- **transition**: $\delta^{\sigma,\tau}(s,s') := \sum_{\substack{a\in\mathcal{A}\\ b\in\mathcal{B}}} \sigma(s)(a)\cdot\tau(s)(b)\cdot\delta(s,a,b)(s')$
- **reward**: $r^{\sigma,\tau}(s) := \sum_{\substack{a\in\mathcal{A}\\ b\in\mathcal{B}}} \sigma(s)(a)\cdot\tau(s)(b)\cdot r(s,a,b)$

In matrix form, the Bellman operator defined by Shapley can be written as a recursive expression:

$$\nu^{\sigma,\tau} = \lambda Id \odot r^{\sigma,\tau} + (1-\lambda) Id \odot \left(\delta^{\sigma,\tau}\nu^{\sigma,\tau}\right) .$$

Algorithm

Bellman fixpoint equation.

$$\nu^{\sigma,\tau} = \lambda Id \odot r^{\sigma,\tau} + (1-\lambda) Id \odot \left(\delta^{\sigma,\tau} \nu^{\sigma,\tau}\right) .$$

By Cramer's rule, we have

$$\nu^{\sigma,\tau}(s) = \frac{\nabla^s_\lambda(\sigma,\tau)}{\nabla_\lambda(\sigma,\tau)} ,$$

where $\nabla_\lambda(\sigma,\tau)$ and $\nabla^s_\lambda(\sigma,\tau)$ are determinants of $n \times n$ matrices. Linearizing the equation, we get

$$0 = \nabla^s_\lambda(\sigma,\tau) - \nu^{\sigma,\tau}(s) \nabla_\lambda(\sigma,\tau) .$$

Algorithm

Linear equation

$$0 = \nabla_\lambda^s(\sigma, \tau) - \nu^{\sigma,\tau}(s)\nabla_\lambda(\sigma, \tau)\,.$$

Define the parameterized matrix game on pure stationary strategies

$$M_\lambda[z](\hat{\sigma}, \hat{\tau}) := \nabla_\lambda^s(\hat{\sigma}, \hat{\tau}) - z\,\nabla_\lambda(\hat{\sigma}, \hat{\tau})\,.$$

Lemma (Attia and Oliu-Barton, 2019)

The discounted value $\mathrm{val}_\lambda(s)$ *is the unique parameter such that*

$$\mathrm{val}\, M_\lambda[z] = 0\,.$$

Moreover, $z \mapsto \mathrm{val}\, M_\lambda[z]$ *is strictly decreasing.*

Algorithm

Lemma (Attia and Oliu-Barton, 2019)

The discounted value $\mathrm{val}_\lambda(s)$ *is the unique parameter such that* $\mathrm{val}\, M_\lambda[z] = 0\,.$ *Moreover,* $z \mapsto \mathrm{val}\, M_\lambda[z]$ *is strictly decreasing.*

As a consequence

$$\lambda \mapsto \mathrm{val}_\lambda(s)$$

is a rational function by parts and there is an explicit bound on the degree and coefficients of the polynomials involved.

Theorem (Attia and Oliu-Barton, 2019)

Consider $\varepsilon > 0$. *There exists an explicit doubly exponentially small discount factor* $\lambda > 0$ *such that*

$$|\,\mathrm{val}_\lambda(s) - \lim_{\lambda \to 0} \mathrm{val}_\lambda(s)| \leq \varepsilon\,.$$

Algorithm

Given a concurrent stochastic game with reachability objective,

1. Consider the discounted reachability objective that **approximates the limit value**.
2. Construct an (exponentially large) **parametric matrix game** whose only parameter that makes it zero value is the discounted value.
3. Use **binary search** to approximate the discounted value.

Idea: Complexity

TFNP[NP] complexity

Given a concurrent stochastic game,

1. Guess ε-**optimal strategies** for each player.

2. Verify the ε-optimality of the guessed strategy using an **optimal counter strategy**.

3. Given "ε-optimal" and optimal counter strategies, **approximate the value** of the corresponding Markov chain.

TFNP[NP] complexity

Theorem (Hansen et. al., 2009)

Even deterministic concurrent games with reachability objectives require playing actions with **doubly exponentially small** *probabilities to get approximately optimal strategies.*

Guessing ε-optimal strategies is impossible... using classic representations.

Theorem (Hansen et. al., 2009)

For all stochastic concurrent games with reachability objectives, there exists ε-optimal strategies that all positive probabilities are at least **doubly exponentially big** *of the form $\varepsilon^{2^{|A|}}$.*

Guessing ε-optimal strategies is possible using floating point numbers.

TFNP[NP] complexity

Theorem (Frederiksen and Miltersen, 2013)

Approximating the value of a Markov chain with reachability objective and floating point probabilities can be done in polynomial time.

Summary

TFNP[NP] complexity

Given a concurrent stochastic game with limit objective,

1. Consider the discounted objective that **approximates the limit value**.
2. Construct a **reachability objective** game with the same value.
3. Guess ε-**optimal strategies** for each player.
4. Verify the ε-optimality of the guessed strategy using an **optimal counter strategy**.
5. Given "ε-optimal" and optimal counter strategies, **approximate the value** of the corresponding Markov chain.

Algorithm

Given a concurrent stochastic game with limit objective,

1. Consider the discounted objective that **approximates the limit value**.
2. Construct an (exponentially large) **parametric matrix game** whose only parameter that makes it zero value is the discounted value.
3. Use **binary search** to approximate the discounted value.

Thank you!