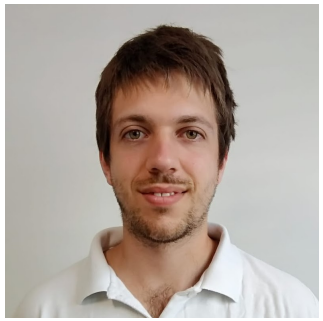
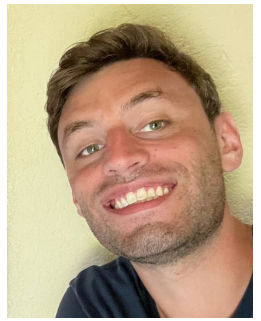
Ergodic Blind MDPs: Decidability of Approximation

K. Chatterjee[1]



R. Saona[1]

B. Ziliotto[2,3]



D. Lurie[2]

[1]ISTA

[2]CEREMADE, Université Paris Dauphine,

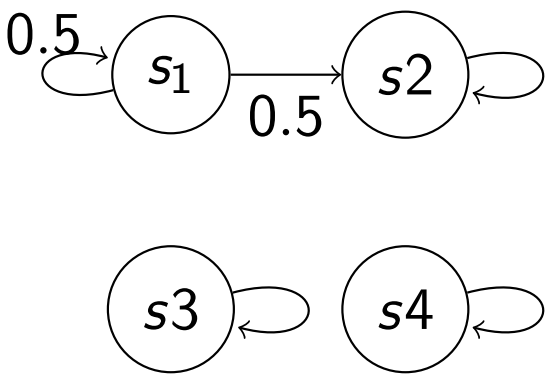
[3]CNRS, PSL Research Institute

Results

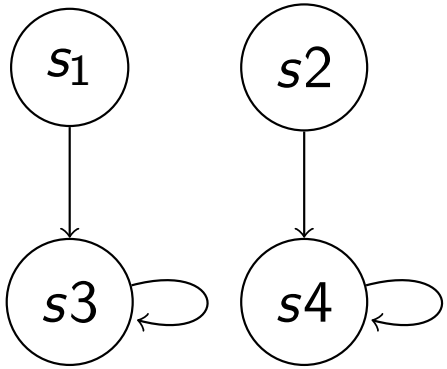
For ergodic blind MDPs with liminf average objective,

- **COMPUTING** the value is **UNDECIDABLE**.
- **APPROXIMATING** the value is **DECIDABLE**.

Blind MDPs

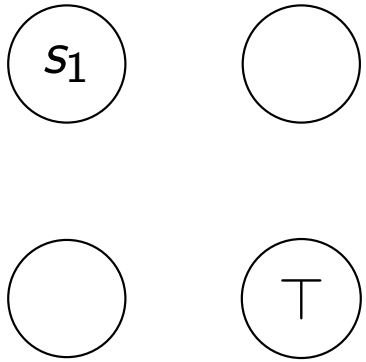


Action *wait*



Action *go*

Reachability objective



We foucs in the maximum probability of reaching the target

$$\sup_{\sigma} \mathbb{P}^{\sigma}(\exists n \quad S_n = \top)$$

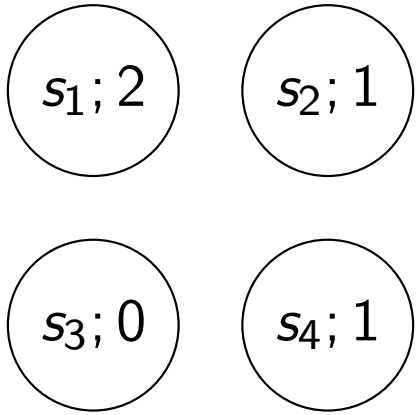
Previous results

Theorem (Madani et. al. 2003)

Approximating *the value of a blind MDP with reachability objective is* **undecidable**.

Via a reduction of the acceptance of the empty word in a two-counter Turing Machine.

Liminf average objective



We focus in the maximum expected longrun average reward

$$\sup_{\sigma} \mathbb{E}^{\sigma} \left(\liminf_{n \to \infty} \frac{1}{n} \sum_{m \geq n} r(S_m) \right)$$

Previous results

Corollary (Madani et. al. 2003)

Approximating *the value of a blind MDP with liminf average objective is* **undecidable**.

Theorem (Saona et. al. 2021)

Consider blind MDPs with liminf average objective. For all $\varepsilon > 0$, *there exists a* **finite-recall** ε**-optimal strategy**.

Via an ergodic theorem and block strategies.

Main difficulty

By the **undeciability** of approximation, there is **no explicit bound** on the amount of recall needed to achieve ε-optimality.

What **subclasses** of blind MDPs **allow a bound** on the recall?

An explicit bound leads to decidability of approximation.

Ergodicity

An **ergodic** blind MDP approximately **forgets the initial belief** after long time.
Formally, for all $\varepsilon > 0$, there exists $n_0 \geq 1$ such that, for all $n \geq n_0$,

$$\mathbb{P}^{\sigma}_{p_1}(S_n = \cdot) \approx_{\varepsilon} \mathbb{P}^{\sigma}_{q_1}(S_n = \cdot)$$

for all σ, p_1, q_1.

Lemma

Checking if a given blind MDP is ergodic is in EXPSPACE.

Approximating is decidable

Theorem

For ergodic blind MDPs with liminf average objective, **approximating** *the value is* **decidable**.

- Consider $\varepsilon > 0$.
- After n the belief is ε-"independent" of the initial belief.
- After n step approximate the belief.
- Since approximations are "independent", the error does not accumulate over time.

Computing is undecidable

Theorem

For ergodic blind MDPs with liminf average objective, **computing** *the value is* **undecidable**.

- Consider a PFA.
- Construct a ergodic blind MDP by leaking to a state and allowing restart.
  - value(blind MDP) $> 1/2 \quad \Leftrightarrow \quad$ value(PFA) $> 1/2$

Back to reachability

Proposition

All ergodic blind MDPs with reachability objective have value one.

Starting from the target state or the initial state leads approximately to the same belief, i.e., having reached the target.

Thank you!