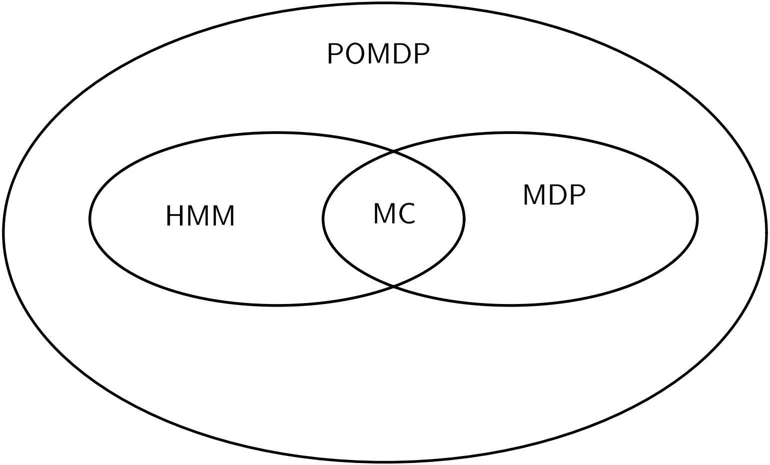
Easy strategies in complex games

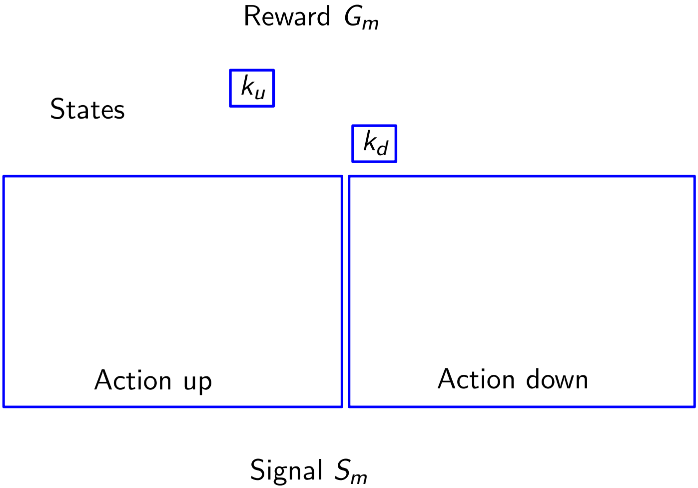
Finite memory strategies in POMDPs with long-run average objective

K. Chatterjee[1] **R. Saona**[1] B. Ziliotto[2]

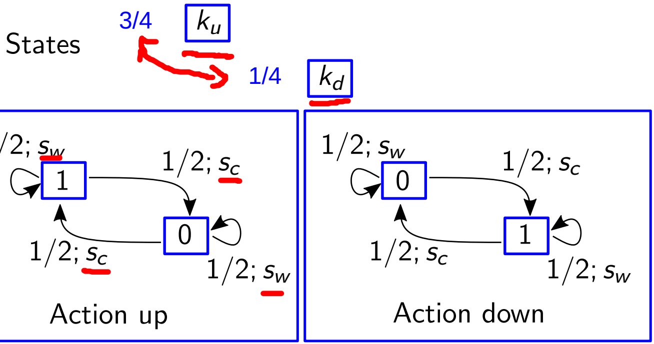
[1]IST Austria

[2]CEREMADE, CNRS, Université Paris Dauphine, PSL Research Institute


POMDP
HMM
MC
MDP

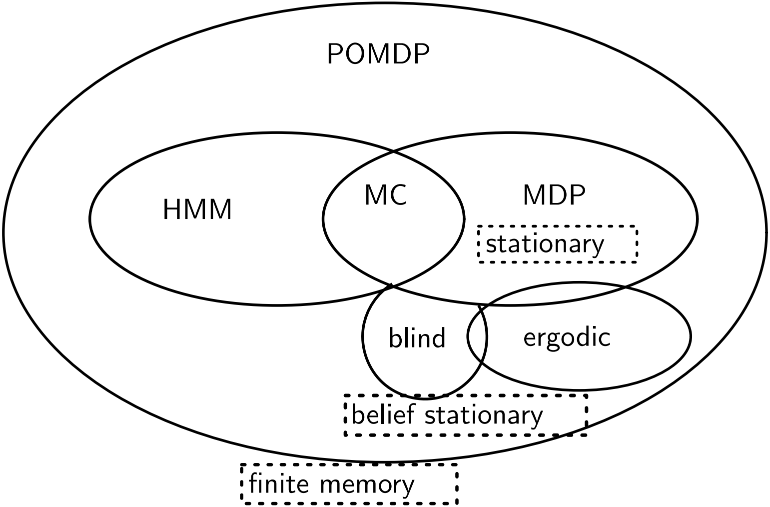

Reward G_m
States
k_u
k_d
Action up
Action down
Signal S_m

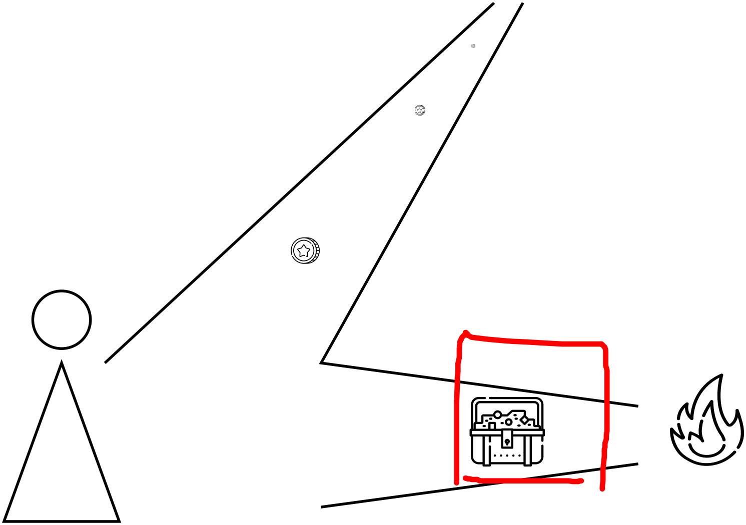
Reward G_m



Signal S_m

$$
\begin{aligned}
v_\infty(p_1) &:= \sup_{\sigma\in\Sigma} \mathbb{E}_\sigma^{p_1} && \left(\liminf_{n\to\infty} \frac{1}{n} \sum_{m=1}^{n} G_m \right) \\
&= \lim_{n\to\infty} v_n && = \lim_{n\to\infty} \sup_{\sigma\in\Sigma} \mathbb{E}_\sigma^{p_1} \left(\frac{1}{n} \sum_{m=1}^{n} G_m \right) \\
&= \lim_{\lambda\to 0^+} v_\lambda && = \lim_{\lambda\to 0^+} \sup_{\sigma\in\Sigma} \mathbb{E}_\sigma^{p_1} \left(\sum_{m=1}^{\infty} \lambda(1-\lambda)^{m-1} G_m \right)
\end{aligned}
$$


POMDP
HMM
MC
MDP
stationary
blind
ergodic
belief stationary
finite memory



Icons made by Freepik from www.flaticon.com

Approximation.

$$\left| v - v_\infty(p_1) \right| \leq \varepsilon \,.$$

This is impossible.

Lower bound.

$$(v_n) \nearrow v_\infty(p_1).$$

Our result.

Upper bound.

$$(v_n) \searrow v_\infty(p_1).$$

This is impossible.

Continuity(?).

$$v_\infty(p_1) = F(\text{rewards}, \text{transitions}) \,.$$

Continuous with respect to rewards and lower semi-continuous with respect to transitions.

Most recent history.

Are last actions and signals enough to approximate the value?

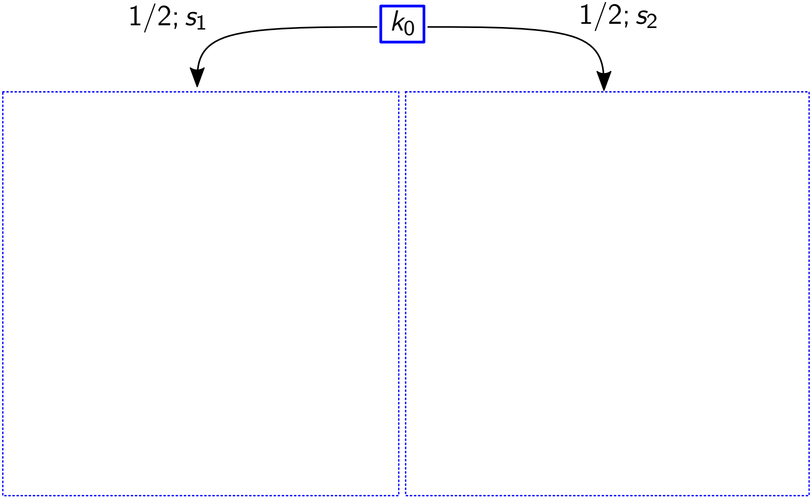
No in general, but it is enough in blind MDP

Other objectives.

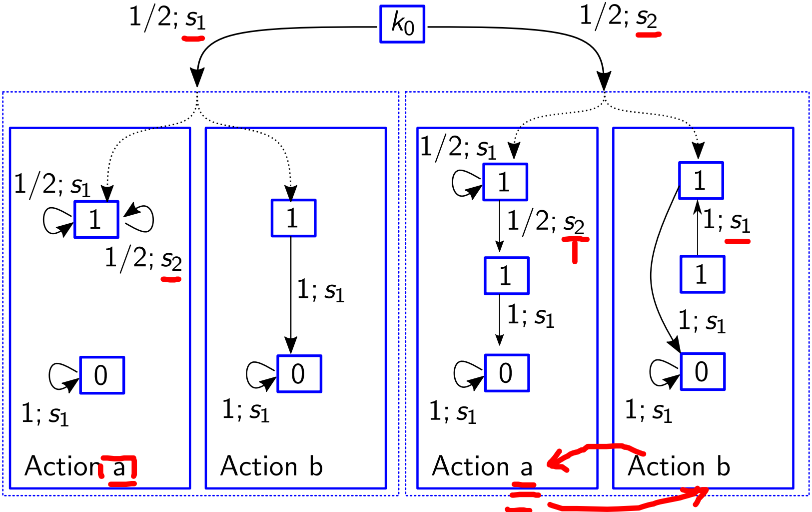
Consider the lim sup:

$$w_\infty := \sup_{\sigma \in \Sigma} \mathbb{E}_\sigma^{p_1} \left(\limsup_{n \to \infty} \frac{1}{n} \sum_{m=1}^{n} G_m \right)$$

Finite memory is not enough

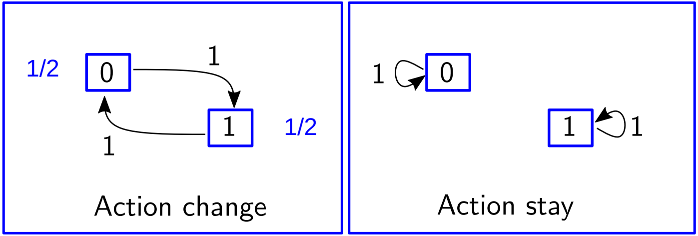


Property. We need to recall the first signal to play ε-optimally.



Property. We need to recall the first signal to play ε-optimally.

Blind MDP



$$\mathbb{E}_{\sigma}^{p_1}\left(\limsup_{n\to\infty}\frac{1}{n}\sum_{m=1}^{n} G_m\right)$$

$$\sigma = (wait)^{2^{0^2}}(change)(wait)^{2^{1^2}}\cdots(change)(wait)^{2^{N^2}}\cdots$$

Due to Venel and Ziliotto

Continuity.

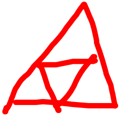
$$v_\infty(p_1) = F(\text{rewards}, \text{transitions}) \, .$$

Is v_∞ continuous with respect to transitions?

Belief partition.

$$v_\infty(p_1) = \sup_{\sigma \in ??} \mathbb{E}_\sigma^{p_1} \left(\liminf_{n \to \infty} \frac{1}{n} \sum_{m=1}^{n} G_m \right) \, .$$

Do belief partition strategies have this property?



Decidability.

Is there a class of POMDPs which is decidable?

Probability objective.

$$w_\infty(p_1;\gamma) = \sup_{\sigma \in \Sigma} \mathbb{P}_\sigma^{p_1}\left(\liminf_{n\to\infty} \frac{1}{n}\sum_{m=1}^{n} G_m > \gamma\right).$$

Do finite-memory strategies approximate the value?

Reference.

This presentation is based in the following paper:

K. Chatterjee, R. Saona and B. Ziliotto.
The Complexity of POMDPs with Long-run Average Objectives.
arXiv prepint, abs/1904.13360, 2020.
https://arxiv.org/abs/1904.13360