How to Play in Infinite MDPs

Stefan Kiefer

University of Oxford, UK

Current Trends in Graph and Stochastic Games
Maastricht, 7 April 2022

Collaborators

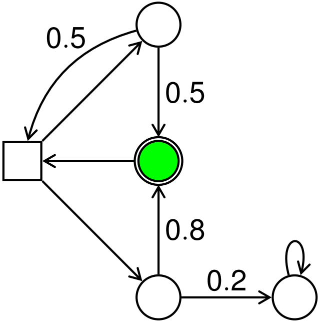
Richard Mayr

Mahsa Shirmohammadi

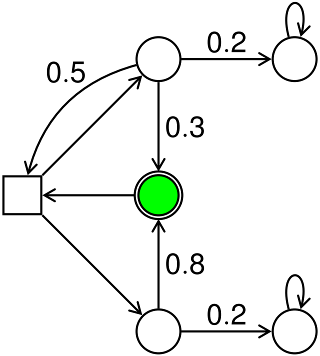
Patrick Totzke

Dominik Wojtczak

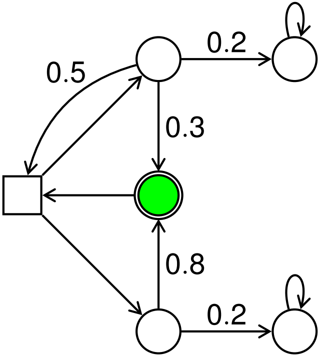
A Finite MDP



A Finite MDP

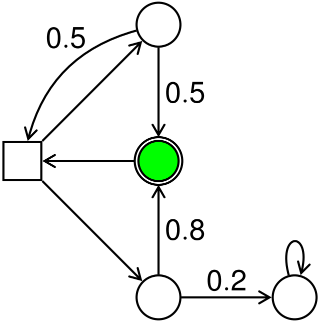


A Finite MDP



In finite MDPs there exists an optimal memoryless strategy.

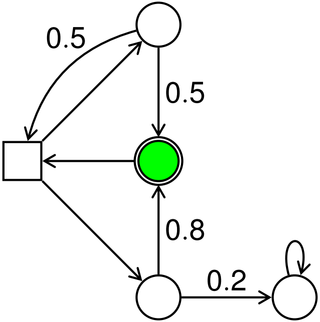
A Finite MDP



In finite MDPs there exists an optimal memoryless strategy.

Even for parity objectives.

A Finite MDP

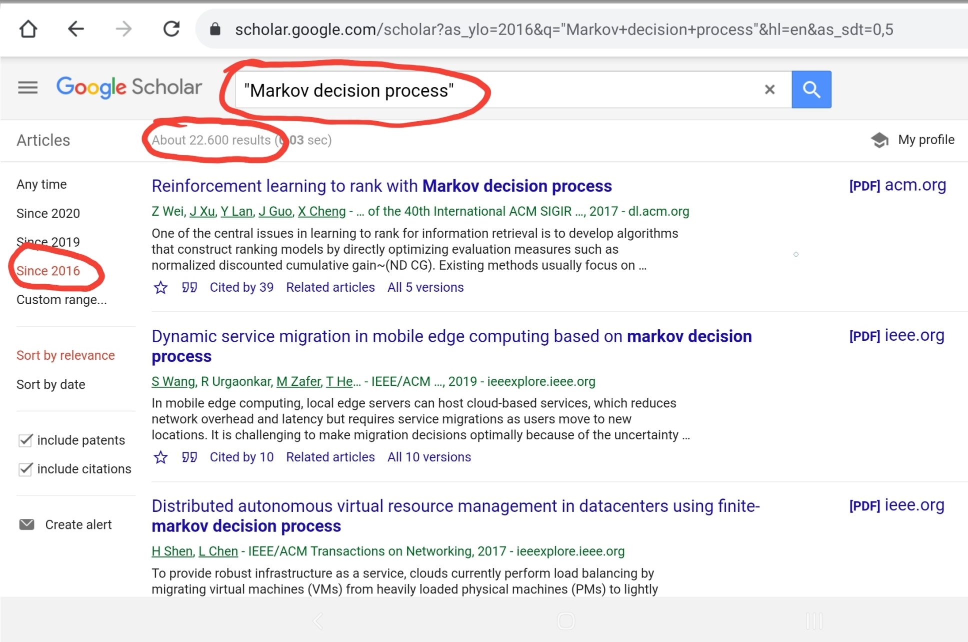


In finite MDPs there exists an optimal memoryless strategy.

Even for parity objectives.

Even in stochastic 2-player games.

MDPs are Everywhere

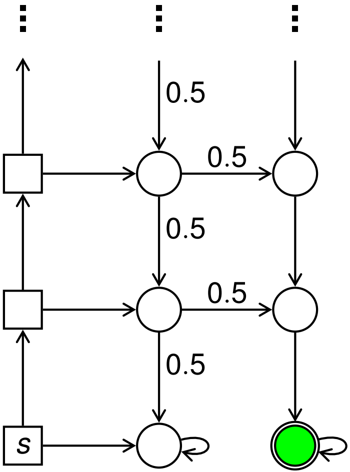


MDPs are Everywhere

The standard model for dynamic systems with both stochastic and nondeterministic behaviour

- artificial intelligence and machine learning
- control theory
- operations research and finance
- formal verification

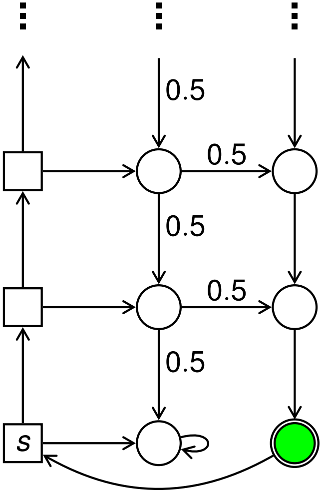
An Infinite MDP



$$val_s(E) \stackrel{\text{def}}{=} \sup_\sigma \mathcal{P}_s^\sigma(E)$$

$$val_s(\text{Reach } \bigcirc) = 1$$

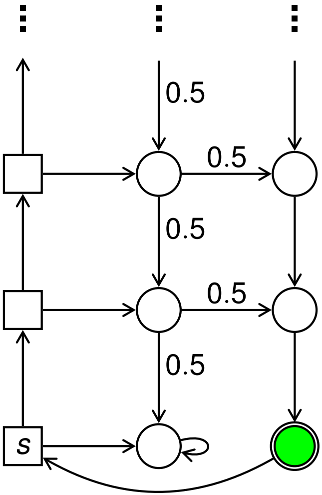
An Infinite MDP



$val_s(E) \stackrel{\text{def}}{=} \sup_\sigma \mathcal{P}_s^\sigma(E)$

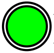
$val_s(\text{Reach}$  $) = 1$

An Infinite MDP

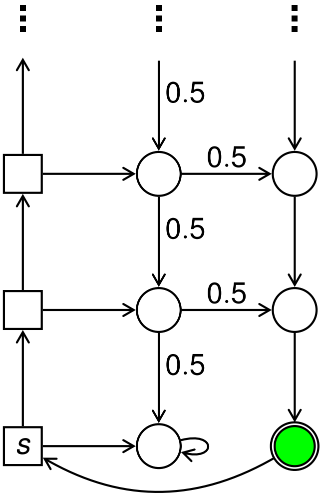


$val_s(E) \stackrel{\text{def}}{=} \sup_\sigma \mathcal{P}_s^\sigma(E)$

$val_s(\text{Reach}$  $) = 1$

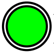
$val_s(\text{Büchi}$  $) = 1$

An Infinite MDP



$val_s(E) \stackrel{\text{def}}{=} \sup_\sigma \mathcal{P}_s^\sigma(E)$

val_s(Reach ) $= 1$

val_s(Büchi ) $= 1$

In infinite MDPs optimal strategies may not exist.
Optimal and ε-optimal strategies may require (infinite) memory.

How to Play in an MDP = Strategy Complexity

How much memory does a good strategy need?

Answer depends on

- objective: reachability, safety, Büchi, parity
- ε-optimal strategies or (where they exist) optimal strategies
- type of MDP: finite, countably infinite, uncountably infinite

How much memory does a good strategy need?

Answer depends on

- objective: reachability, safety, Büchi, parity
- ε-optimal strategies or (where they exist) optimal strategies
- type of MDP: finite, countably infinite, uncountably infinite

Theorem (ICALP'19)

In countably infinite MDPs with Büchi objective, for ε-optimal strategies, a step counter plus 1 bit of memory is necessary and sufficient.

How to Play in an MDP = Strategy Complexity

How much memory does a good strategy need?

Answer depends on

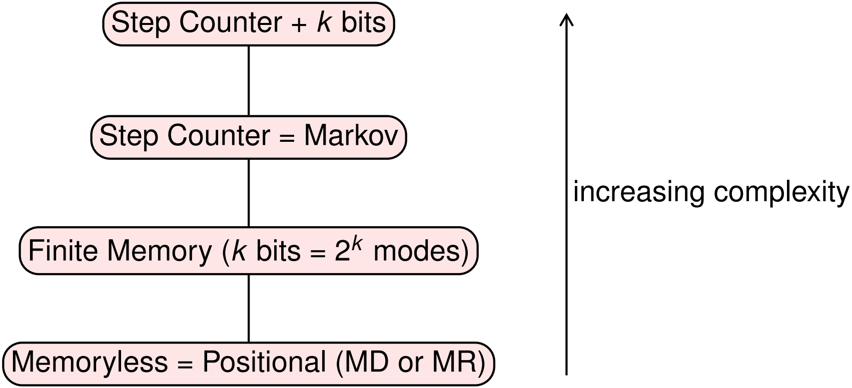
- objective: reachability, safety, Büchi, parity
- ε-optimal strategies or (where they exist) optimal strategies
- type of MDP: finite, countably infinite, uncountably infinite

Theorem (ICALP'19)

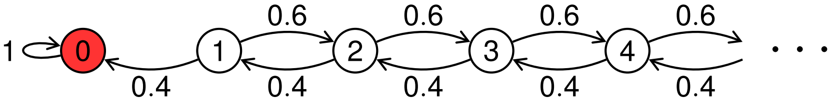
In countably infinite MDPs with Büchi objective, for ε-optimal strategies, a step counter plus 1 bit of memory is necessary and sufficient.

Randomized vs deterministic makes little difference here.

Memory


Step Counter + k bits
Step Counter = Markov
Finite Memory (k bits = 2^k modes)
Memoryless = Positional (MD or MR)
increasing complexity

Gambler's Ruin

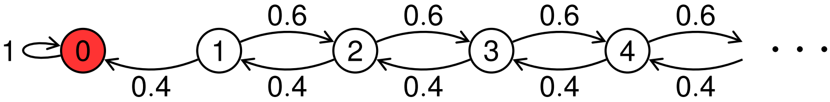


The probability of reaching 0 is

- positive everywhere
- less than 1 everywhere except in 0

Such situations do not exist in finite Markov chains or MDPs.

Gambler's Ruin



The probability of reaching (0) is

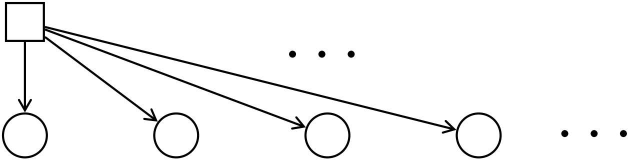
- positive everywhere
- less than 1 everywhere except in (0)

Such situations do not exist in finite Markov chains or MDPs.

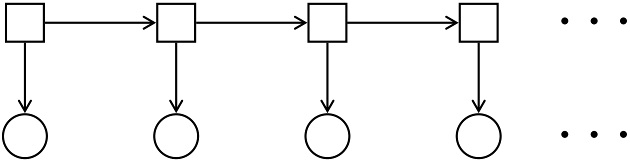
Algorithmics of finitely presented MDPs is a different topic.

From Infinite to Finite Branching

For many objectives, there is a reduction to finite branching:



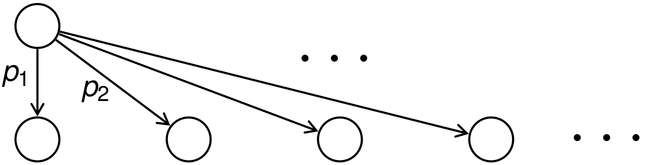
is replaced by



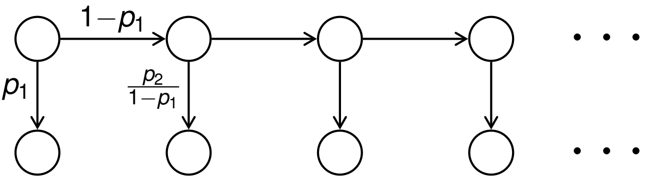
Then a good strategy in the new MDP can be translated back.

From Infinite to Finite Branching

For many objectives, there is a reduction to finite branching:



is replaced by



Then a good strategy in the new MDP can be translated back.

Reachability in Finite MDPs

Plan for next part of the talk:

- construct good strategies with little memory
- use reachability as example (not our own work)

Lemma (optimal strategies in finite MDPs)

Consider a finite MDP with reachability target T.
There exists a single MD strategy σ that is optimal everywhere.
Formally, $\mathcal{P}_s^\sigma(\mathit{Reach}\ T) = \mathit{val}_s(\mathit{Reach}\ T)$ for every state s.

Reachability in Countable MDPs

Lemma (ε-optimal strategies need no memory)

Consider a countable MDP with reachability target T.
For every $\varepsilon > 0$ and every state s there exists an MD strategy σ that is ε-optimal for s, i.e., $\mathcal{P}_s^\sigma(\text{Reach } T) \geq val_s(\text{Reach } T) - \varepsilon$.

Proof idea: reduction to the finite case

Fix state s and $\varepsilon > 0$.
Let τ be an $\frac{\varepsilon}{2}$-optimal strategy (potentially infinite memory):

$$\mathcal{P}_s^\tau(\text{Reach } T) \;\geq\; val_s(\text{Reach } T) - \frac{\varepsilon}{2}$$

Then there is n such that

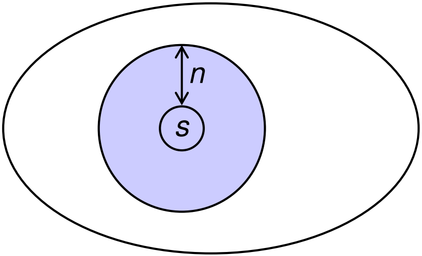
$$\mathcal{P}_s^\tau(\text{Reach } T \text{ within at most } n \text{ steps}) \;\geq\; val_s(\text{Reach } T) - \varepsilon$$

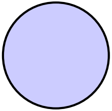
Reachability in Countable MDPs

Lemma (ε-optimal strategies need no memory)

Consider a countable MDP with reachability target T.
For every $\varepsilon > 0$ and every state s there exists an MD strategy σ that is ε-optimal for s, i.e., $\mathcal{P}_s^\sigma(Reach\ T) \geq val_s(Reach\ T) - \varepsilon$.

Proof idea: reduction to the finite case



τ does well in  alone.

That sub-MDP is finite.

It has an optimal MD strategy!

Reachability in Countable MDPs

Theorem (Ornstein'69: uniform ε-optimal strategies)

Consider a countable MDP with reachability target T. Let $\varepsilon > 0$. There is a single MD strategy σ that is ε-optimal everywhere. Formally, $\mathcal{P}_s^\sigma(Reach\ T) \geq val_s(Reach\ T) - \varepsilon$ for every state s.

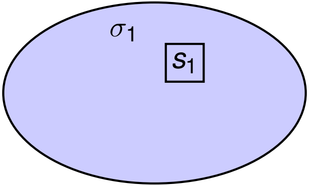
Proof idea: "plaster" the state space

Reachability in Countable MDPs

Theorem (Ornstein'69: uniform ε-optimal strategies)

Consider a countable MDP with reachability target T. Let $\varepsilon > 0$. There is a single MD strategy σ that is ε-optimal everywhere. Formally, $\mathcal{P}_s^\sigma(\mathit{Reach}\ T) \geq \mathit{val}_s(\mathit{Reach}\ T) - \varepsilon$ for every state s.

Proof idea: "plaster" the state space



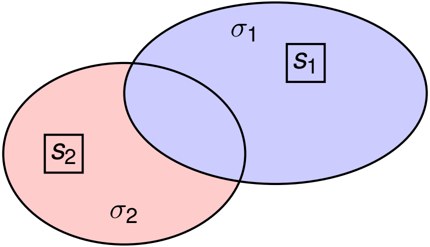
Enumerate all states $s_1, s_2, s_3, \ldots$

Reachability in Countable MDPs

Theorem (Ornstein'69: uniform ε-optimal strategies)

Consider a countable MDP with reachability target T. Let $\varepsilon > 0$. There is a single MD strategy σ that is ε-optimal everywhere. Formally, $\mathcal{P}_s^\sigma(\mathit{Reach}\ T) \geq \mathit{val}_s(\mathit{Reach}\ T) - \varepsilon$ for every state s.

Proof idea: "plaster" the state space



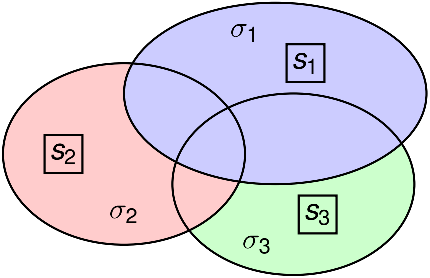
Enumerate all states $s_1, s_2, s_3, \ldots$

Reachability in Countable MDPs

Theorem (Ornstein'69: uniform ε-optimal strategies)

Consider a countable MDP with reachability target T. Let $\varepsilon > 0$. There is a single MD strategy σ that is ε-optimal everywhere. Formally, $\mathcal{P}_s^\sigma(\mathit{Reach}\ T) \geq \mathit{val}_s(\mathit{Reach}\ T) - \varepsilon$ for every state s.

Proof idea: "plaster" the state space



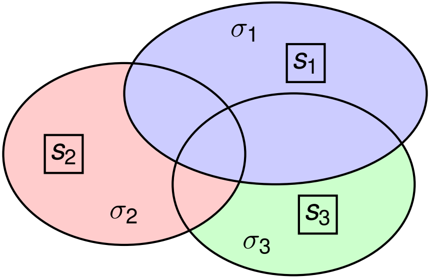
Enumerate all states $s_1, s_2, s_3, \ldots$

Reachability in Countable MDPs

Theorem (Ornstein'69: uniform ε-optimal strategies)

Consider a countable MDP with reachability target T. Let $\varepsilon > 0$. There is a single MD strategy σ that is ε-optimal everywhere. Formally, $\mathcal{P}_s^\sigma(\mathit{Reach}\ T) \geq \mathit{val}_s(\mathit{Reach}\ T) - \varepsilon$ for every state s.

Proof idea: "plaster" the state space



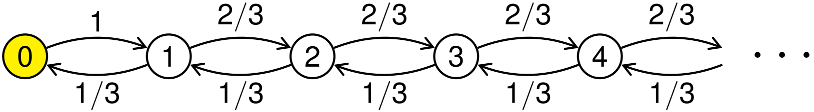
Enumerate all states $s_1, s_2, s_3, \ldots$

Fix σ_i in a region that is

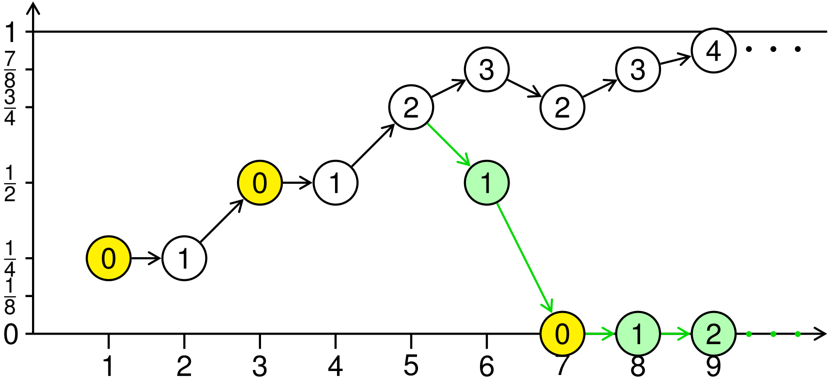
(A) large enough for s_i

(B) not too damaging for $s_{i+1}, \ldots$

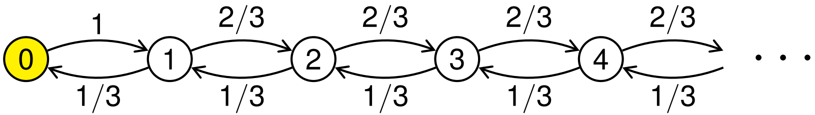
Lévy's Zero-One Law



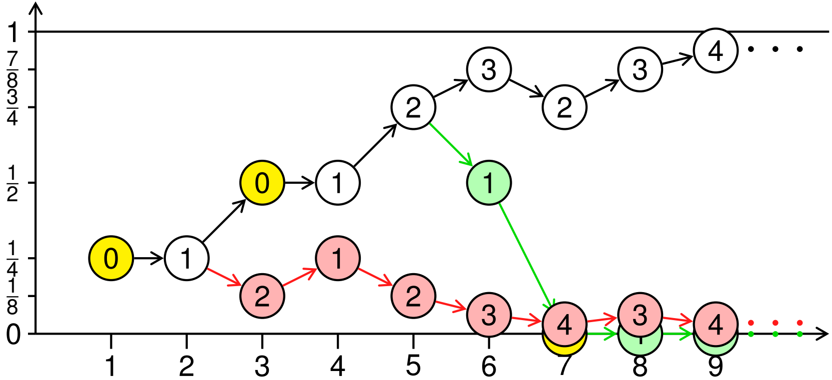
$E \stackrel{\text{def}}{=}$ Starting from 0 re-visit 0 exactly once



Lévy's Zero-One Law



$E \stackrel{\text{def}}{=}$ Starting from 0 re-visit 0 exactly once



Lévy's Zero-One Law

tail event: independent of any finite prefix (unlike E from before)

For tail events E, by Lévy's zero-one law:

$$E \text{ and } \left\{ s_1 s_2 \cdots \;\middle|\; \lim_{i \to \infty} \mathcal{P}_{s_i}(E) = 1 \right\} \text{ are equal up to a null set.}$$

Lévy's Zero-One Law

tail event: independent of any finite prefix (unlike E from before)

For tail events E, by Lévy's zero-one law:

$$E \text{ and } \left\{ s_1 s_2 \cdots \;\middle|\; \lim_{i \to \infty} \mathcal{P}_{s_i}(E) = 1 \right\} \text{ are equal up to a null set.}$$

Theorem (Ornstein'69: uniform a.s. winning strategies)

Consider a countable MDP with reachability target T. There is a single MD strategy that is almost-surely winning for all states that have an almost-surely winning strategy.

Proof.

Remove all states that do not have an a.s. winning strategy.

Make T a sink so that Reach T becomes a tail event.

Fix a uniform $\frac{1}{2}$-optimal MD strategy σ (exists as shown before).

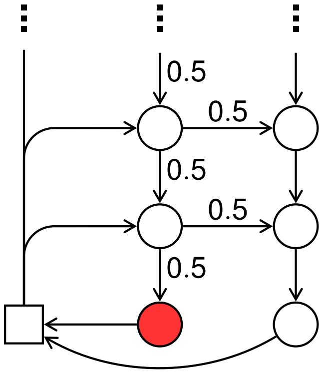
"Optimism" to reach T must converge to 1. □

Strategy Complexity of Co-Büchi

Theorem (LICS'17)

Consider a countable finitely branching MDP with co-Büchi objective.
There is a single MD strategy that is almost-surely winning for all states that have an almost-surely winning strategy.

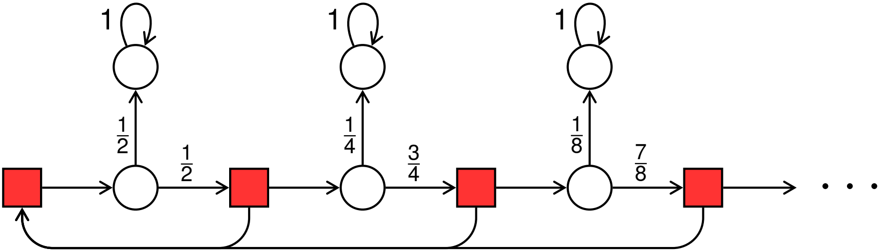
false for infinite branching:



Strategy Complexity of Co-Büchi: Safety-First Strategies

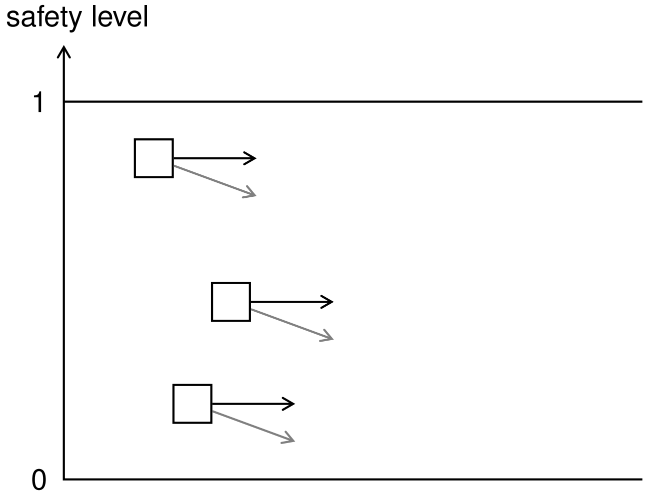
safety-first = in each state minimize prob to ever visit 🔴 again

- If it succeeds, it satisfies co-Büchi.
- There is an MD safety-first strategy that is optimal for safety in every state.

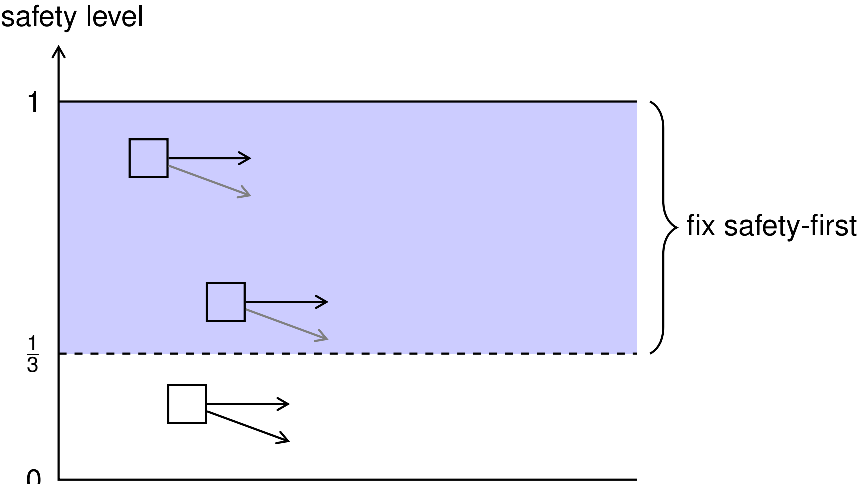


We will combine MD strategies for safety and reachability.

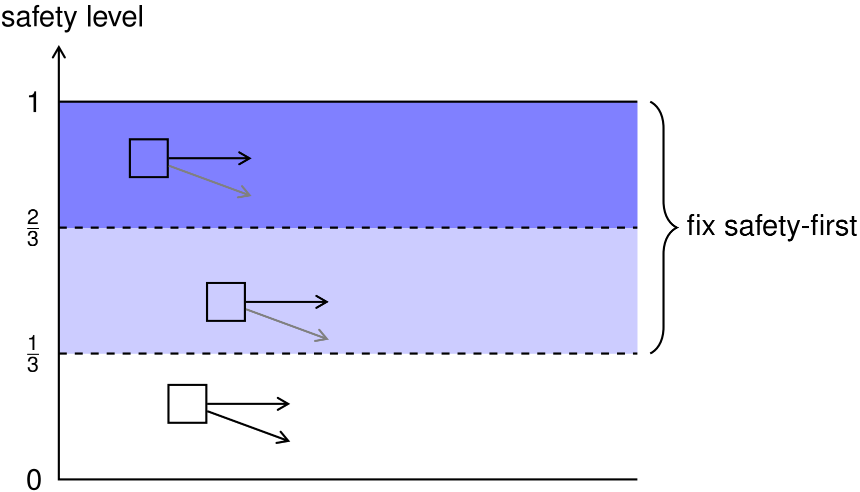
Strategy Complexity of Co-Büchi: Flag Construction



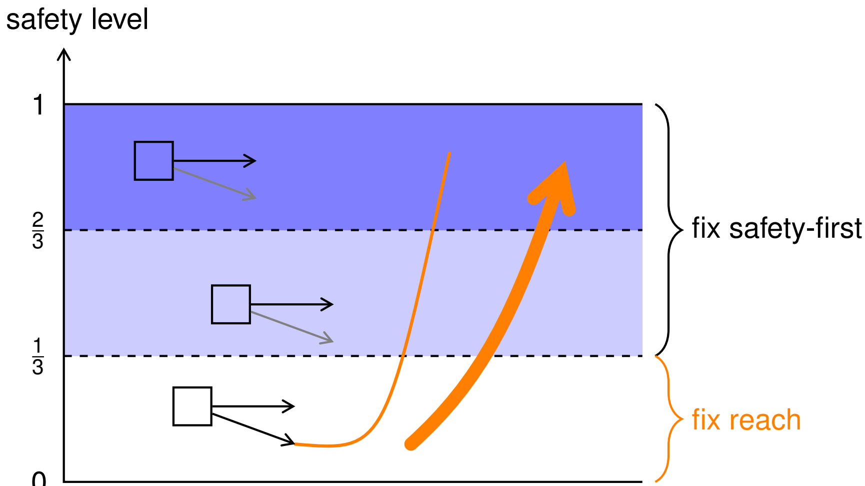
Strategy Complexity of Co-Büchi: Flag Construction



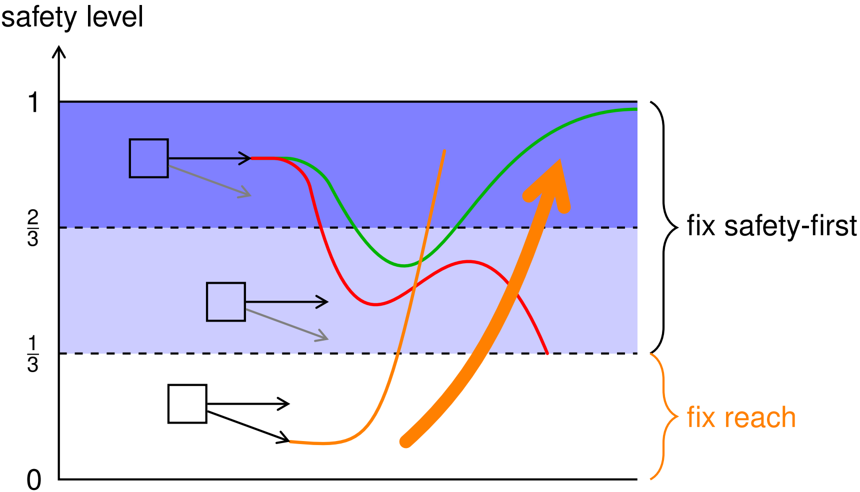
Strategy Complexity of Co-Büchi: Flag Construction



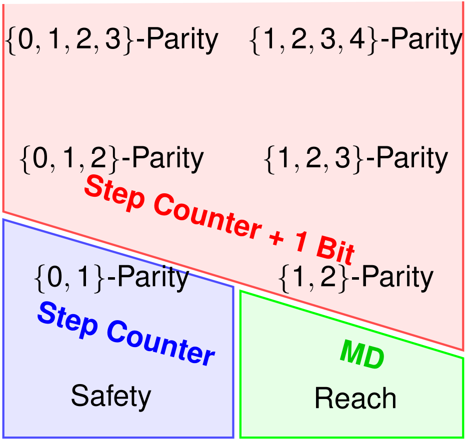
Strategy Complexity of Co-Büchi: Flag Construction



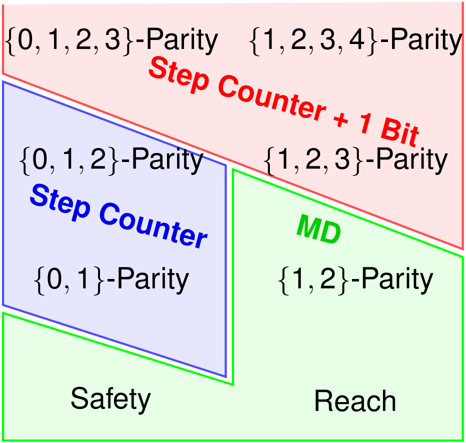
Strategy Complexity of Co-Büchi: Flag Construction



Strategy Complexity of Parity Objectives (Concur'20)



ε-optimal (infinite branching)



optimal (infinite branching)

With finite branching, **Step Counter** becomes **MD**