

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

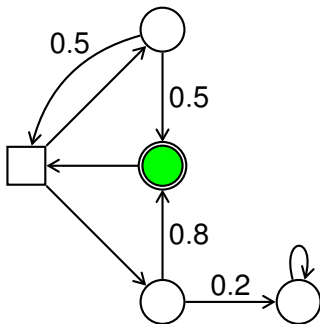


Patrick Totzke

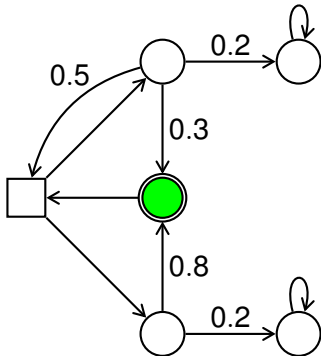


Dominik Wojtczak

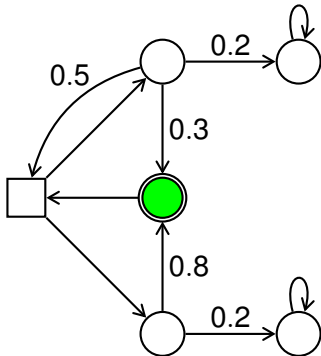
A Finite MDP



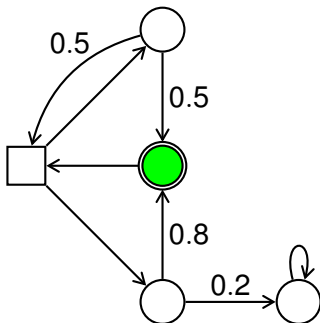
A Finite MDP



A Finite MDP

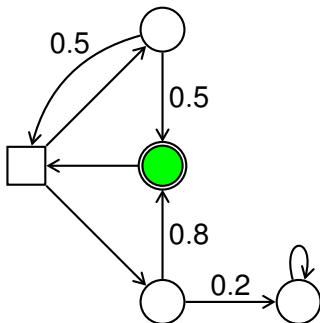


In finite MDPs there exists an optimal memoryless strategy.



In finite MDPs there exists an optimal memoryless strategy.

Even for parity objectives.



In finite MDPs there exists an optimal memoryless strategy.

Even for parity objectives.

Even in stochastic 2-player games.

MDPs are Everywhere

scholar.google.com/scholar?as_ylo=2016&q="Markov+decision+process"&hl=en&as_sdt=0,5

Google Scholar

"Markov decision process"

Articles

About 22.600 results (0.03 sec)

My profile

Any time

Since 2020

Since 2019

Since 2016

Custom range...

Reinforcement learning to rank with Markov decision process

[PDF] acm.org

Z Wei, J Xu, Y Lan, J Guo, X Cheng - ... of the 40th International ACM SIGIR ..., 2017 - dl.acm.org

One of the central issues in learning to rank for information retrieval is to develop algorithms that construct ranking models by directly optimizing evaluation measures such as normalized discounted cumulative gain~(ND CG). Existing methods usually focus on ...

☆ Cited by 39 Related articles All 5 versions

Dynamic service migration in mobile edge computing based on markov decision process

[PDF] ieee.org

S Wang, R Uргаonkar, M Zafer, T He... - IEEE/ACM ..., 2019 - ieeexplore.ieee.org

In mobile edge computing, local edge servers can host cloud-based services, which reduces network overhead and latency but requires service migrations as users move to new locations. It is challenging to make migration decisions optimally because of the uncertainty ...

☆ Cited by 10 Related articles All 10 versions

Distributed autonomous virtual resource management in datacenters using finite-markov decision process

[PDF] ieee.org

H Shen, L Chen - IEEE/ACM Transactions on Networking, 2017 - ieeexplore.ieee.org

To provide robust infrastructure as a service, clouds currently perform load balancing by migrating virtual machines (VMs) from heavily loaded physical machines (PMs) to lightly

Sort by relevance

Sort by date

include patents

include citations

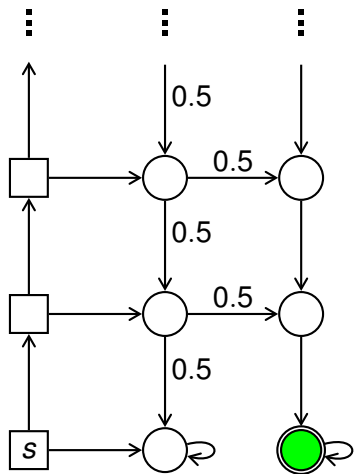
Create alert

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

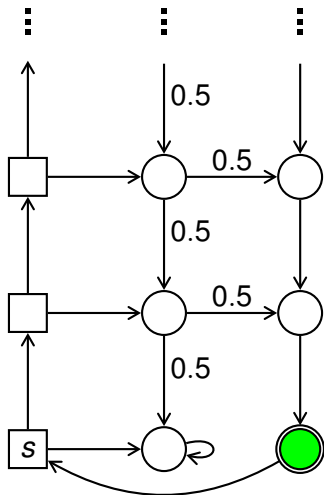
An Infinite MDP



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

$$val_s(\text{Reach } \textcircled{\textcircled{\color{green}\bullet}}) = 1$$

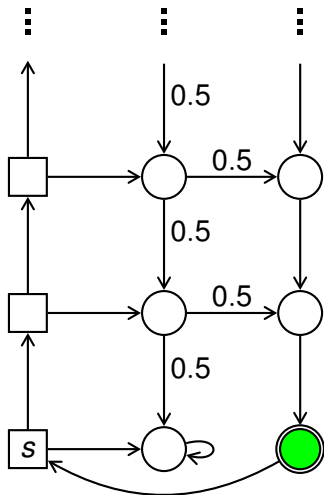
An Infinite MDP



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

$$val_s(\text{Reach } \textcircled{\textcircled{\color{green}\bullet}}) = 1$$

An Infinite MDP

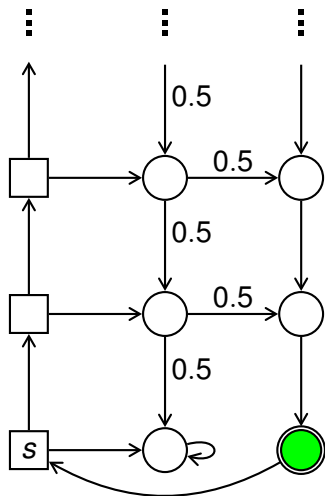


$$val_s(E) \stackrel{\text{def}}{=} \sup_{\sigma} \mathcal{P}_s^{\sigma}(E)$$

$$val_s(\text{Reach } \textcircled{\textcircled{\color{green}\bullet}}) = 1$$

$$val_s(\text{Büchi } \textcircled{\textcircled{\color{green}\bullet}}) = 1$$

An Infinite MDP



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

$$val_s(\text{Reach } \textcircled{\textcircled{\color{red}\bullet}}) = 1$$

$$val_s(\text{Büchi } \textcircled{\textcircled{\color{red}\bullet}}) = 1$$

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

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 much memory does a good strategy need?

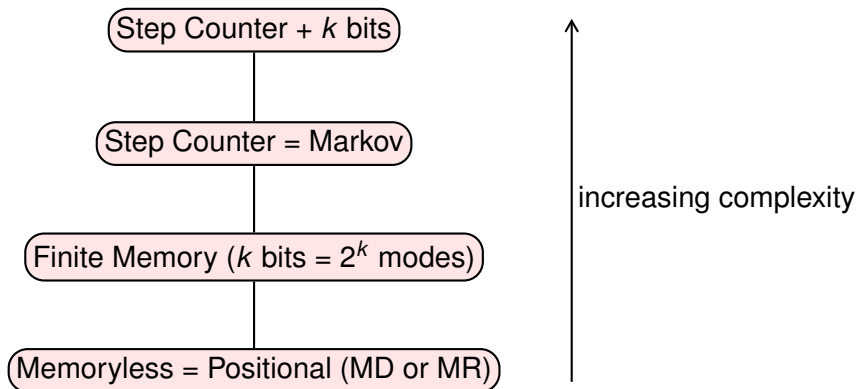
Answer depends on

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

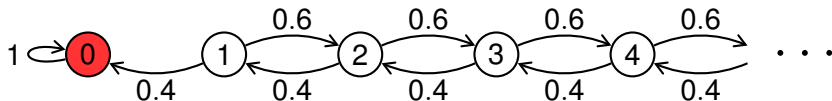
Theorem (ICALP'19)

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.



Gambler's Ruin

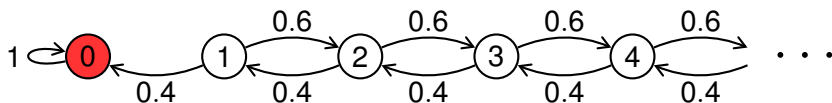


The probability of reaching **0** is

- 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 $\textcircled{0}$ is

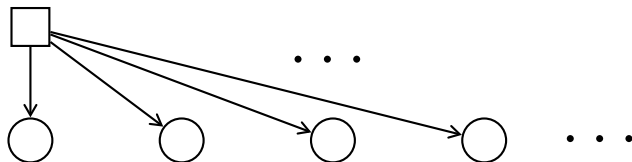
- positive everywhere
- less than 1 everywhere except in $\textcircled{0}$

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

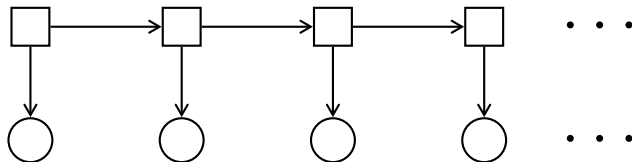
Algorithmics of finitely presented MDPs is a different topic.

From Infinite to Finite Branching

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



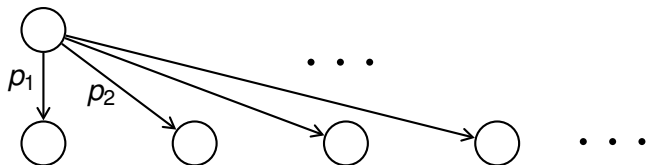
is replaced by



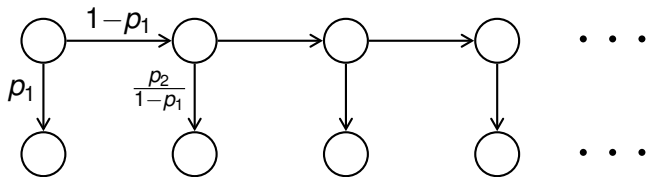
Then a good strategy in the new MDP can be translated back.

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.

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, $\mathcal{P}_s^\sigma(\text{Reach } T) = \text{val}_s(\text{Reach } T)$ 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., $\mathcal{P}_s^\sigma(\text{Reach } T) \geq \text{val}_s(\text{Reach } T) - \varepsilon$.

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):

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

Then there is n such that

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

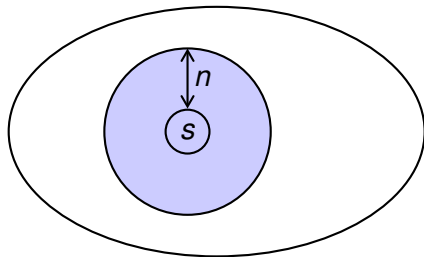
Reachability in Countable MDPs

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., $\mathcal{P}_s^\sigma(\text{Reach } T) \geq \text{val}_s(\text{Reach } T) - \varepsilon$.

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, $\mathcal{P}_s^\sigma(\text{Reach } T) \geq \text{val}_s(\text{Reach } T) - \varepsilon$ for every state s .

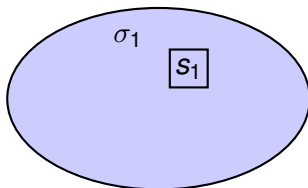
Proof idea: “plaster” the state space

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, $\mathcal{P}_s^\sigma(\text{Reach } T) \geq \text{val}_s(\text{Reach } T) - \varepsilon$ for every state s .

Proof idea: “plaster” the state space



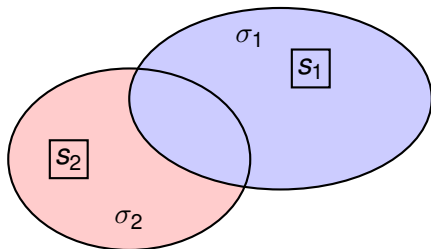
Enumerate all states s_1, s_2, s_3, \dots

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, $\mathcal{P}_s^\sigma(\text{Reach } T) \geq \text{val}_s(\text{Reach } T) - \varepsilon$ for every state s .

Proof idea: “plaster” the state space



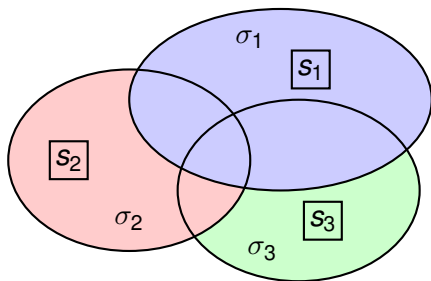
Enumerate all states s_1, s_2, s_3, \dots

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, $\mathcal{P}_s^\sigma(\text{Reach } T) \geq \text{val}_s(\text{Reach } T) - \varepsilon$ for every state s .

Proof idea: “plaster” the state space



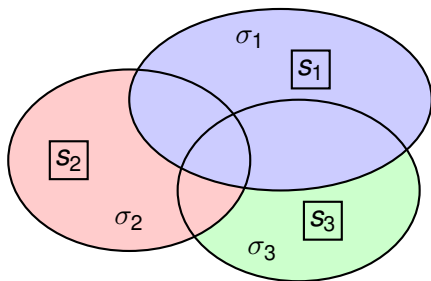
Enumerate all states s_1, s_2, s_3, \dots

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, $\mathcal{P}_s^\sigma(\text{Reach } T) \geq \text{val}_s(\text{Reach } T) - \varepsilon$ for every state s .

Proof idea: “plaster” the state space

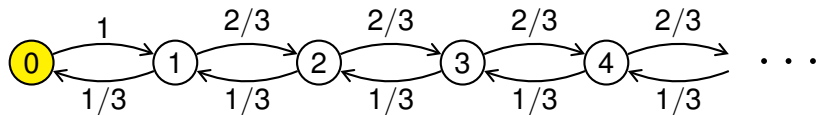


Enumerate all states s_1, s_2, s_3, \dots

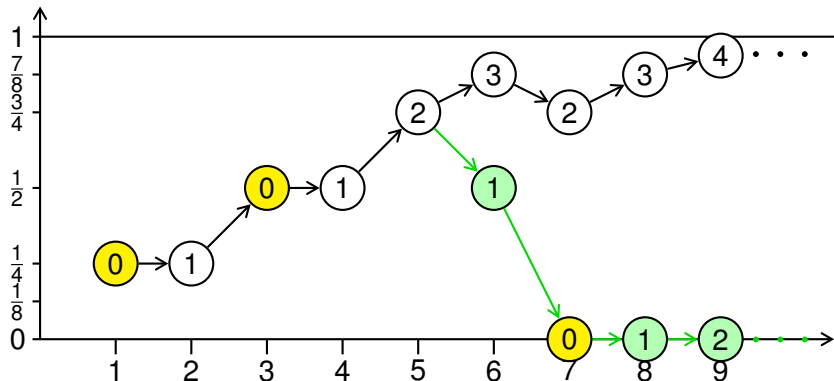
Fix σ_i in a region that is

- (A) large enough for s_i
- (B) not too damaging for s_{i+1}, \dots

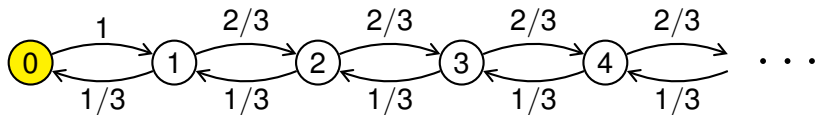
Lévy's Zero-One Law



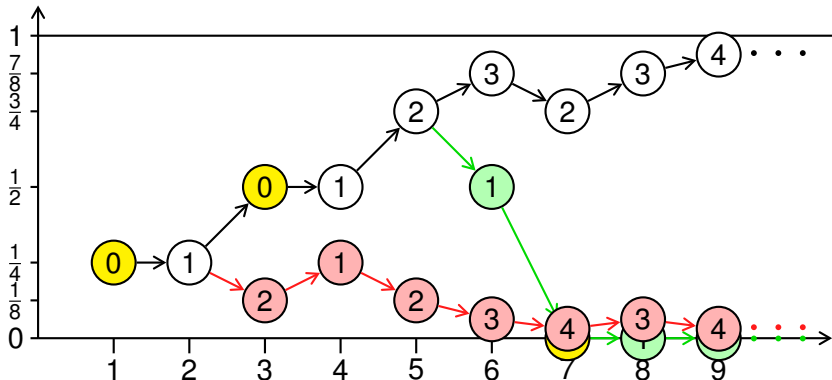
$E \stackrel{\text{def}}{=} \text{Starting from } \textcircled{0} \text{ re-visit } \textcircled{0} \text{ exactly once}$



Lévy's Zero-One Law



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



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 and $\left\{ s_1 s_2 \cdots \mid \lim_{i \rightarrow \infty} \mathcal{P}_{s_i}(E) = 1 \right\}$ 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 and $\left\{ s_1 s_2 \cdots \mid \lim_{i \rightarrow \infty} \mathcal{P}_{s_i}(E) = 1 \right\}$ 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 $\text{Reach } T$ 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. □

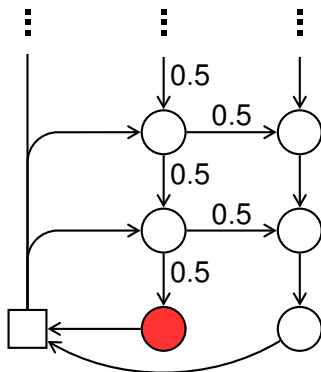
Strategy Complexity of Co-Büchi

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.

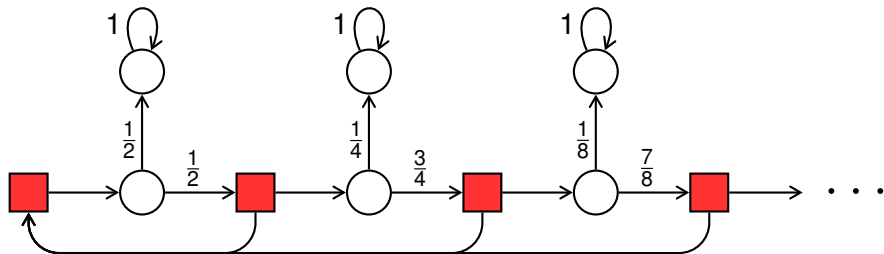
false for infinite branching:



Strategy Complexity of Co-Büchi: Safety-First Strategies

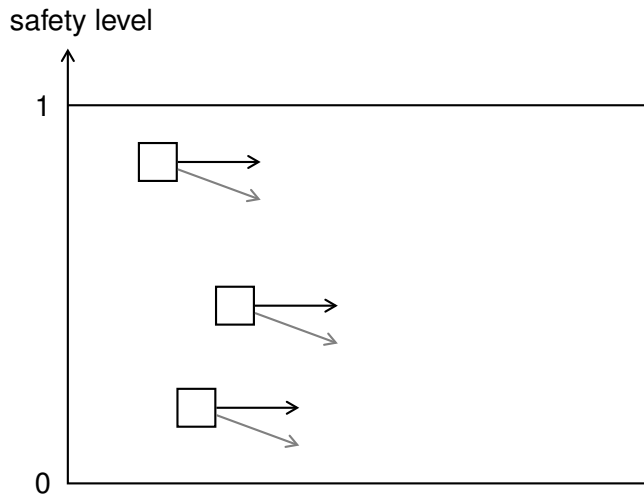
safety-first = in each state minimize prob to ever visit  again

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

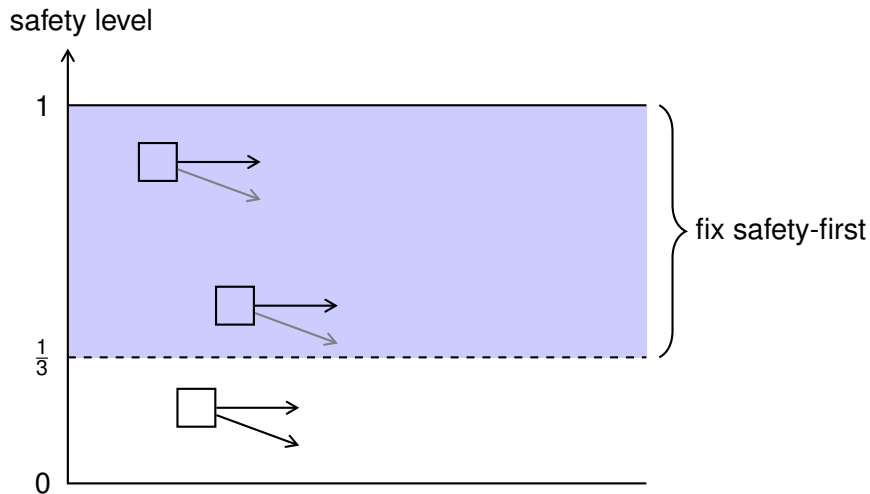


We will **combine** MD strategies for safety and reachability.

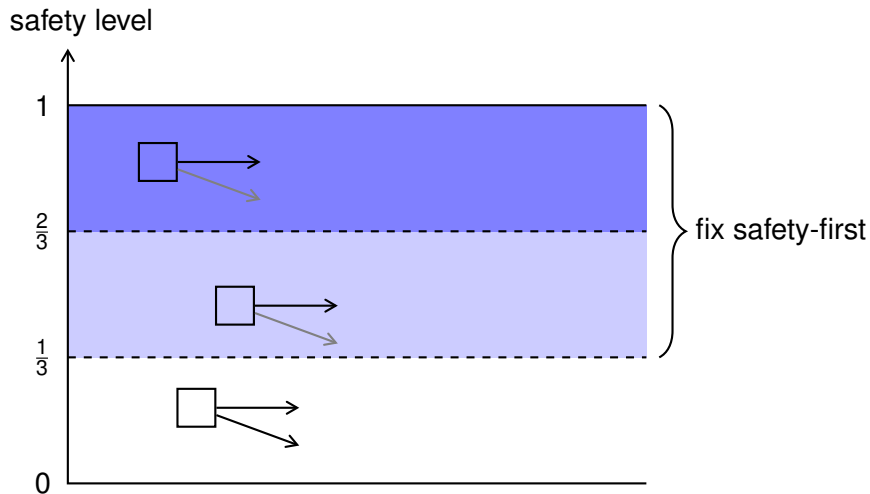
Strategy Complexity of Co-Büchi: Flag Construction



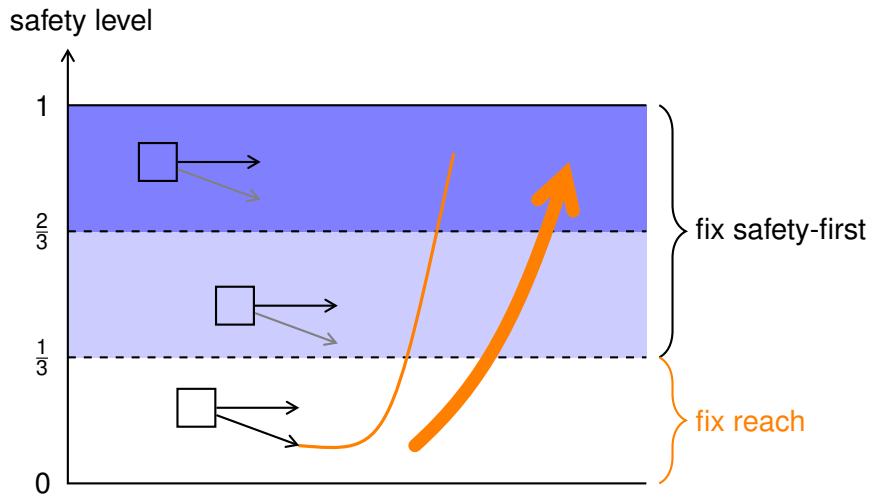
Strategy Complexity of Co-Büchi: Flag Construction



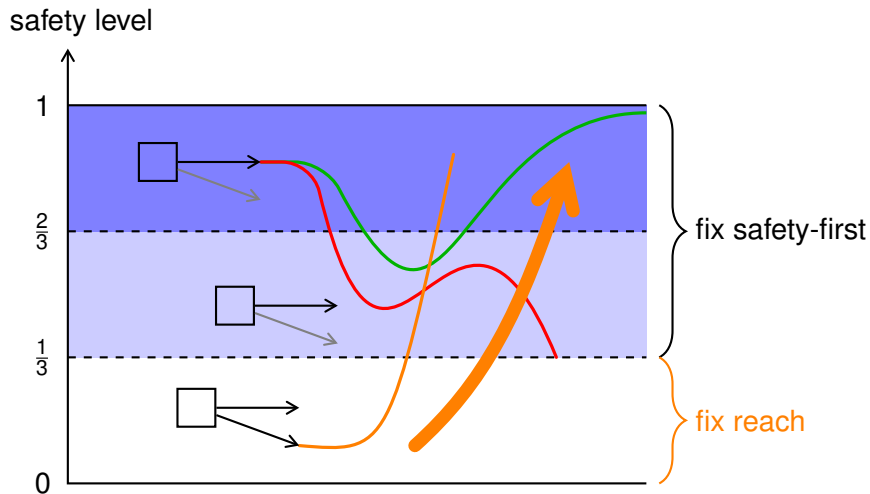
Strategy Complexity of Co-Büchi: Flag Construction



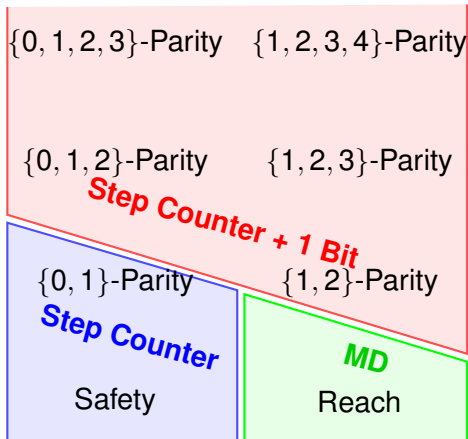
Strategy Complexity of Co-Büchi: Flag Construction



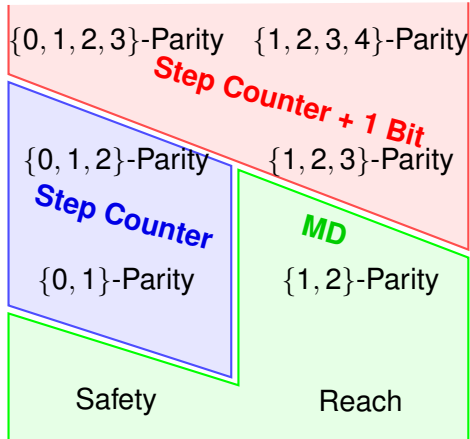
Strategy Complexity of Co-Büchi: Flag Construction



Strategy Complexity of Parity Objectives (Concur'20)



ϵ -optimal (infinite branching)



optimal (infinite branching)

With finite branching,

Step Counter

becomes

MD