RAP Kickoff

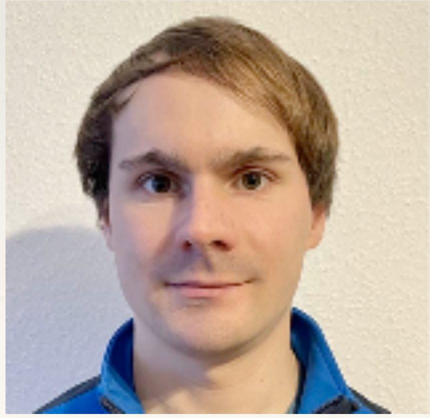
Padova, March 9, 2024

Approximated Fixpoints of Approximated Functions

Paolo Baldan

Barbara Koenig

Sebastian Gurke



Tommaso Padoan

Florian Wittbold

The setting

Given a function

$$f : X \to X$$

X endowed with

- an **order** $\leq$
- a **distance** d

(Least) fixpoint μf:

(smallest) element $x \in X$ such that $f(x) = x$

Classical of results ensuring existence of fixpoints …

In a complete lattice

$(X, \leq)$ complete lattice

$f : X \to X$ **monotone**

μf **exists** by Knaster-Tarski

if $f : X \to X$ **(ω)-continuous**

$$\begin{cases} x_0 = \bot \\ x_{n+1} = f(x_n) \end{cases}$$

then

$$\mu f = \bigsqcup_{n \in \omega} x_n$$

μf

$\vdots$

$f(f(\bot))$

$f(\bot)$

$\bot$

In a complete metric space

(X, d) **complete metric space**

$f : X \to X$ **contraction**

μf **unique fixpoint** by Banach

$$\begin{cases} x_0 & \text{arbitrary} \\ x_{n+1} = f(x_n) \end{cases}$$

then

$x_n \to \mu f$

x_0

$f(x_0)$

$f(f(x_0))$

$\dots$

μf

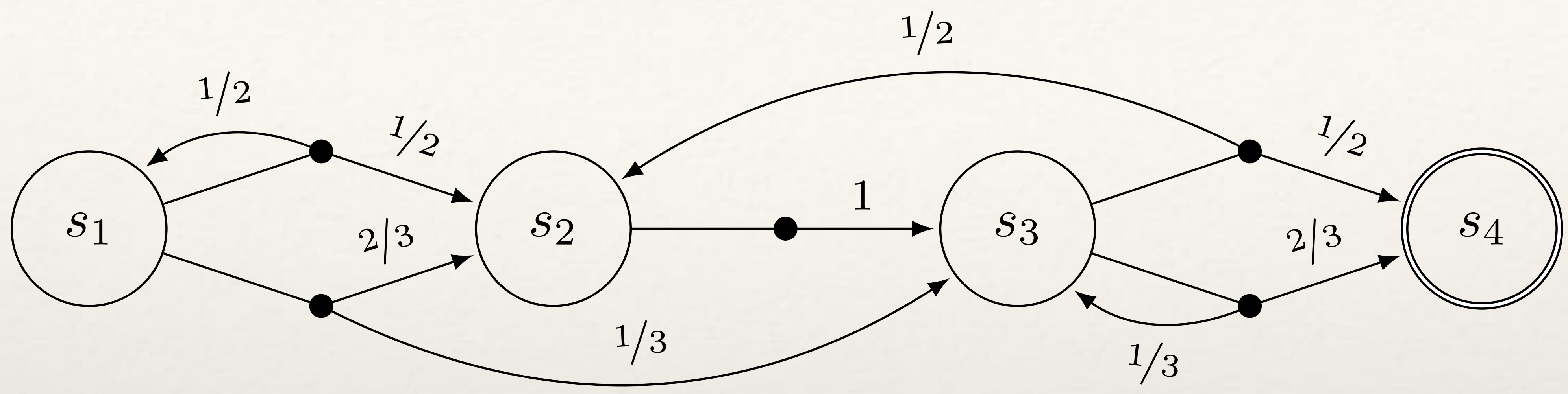
What if f can be only approximated?

Function f **not known** exactly …

… but we can construct $f_n : L \to L$ such that

$$f_n \to f$$

Markov Decision Processes

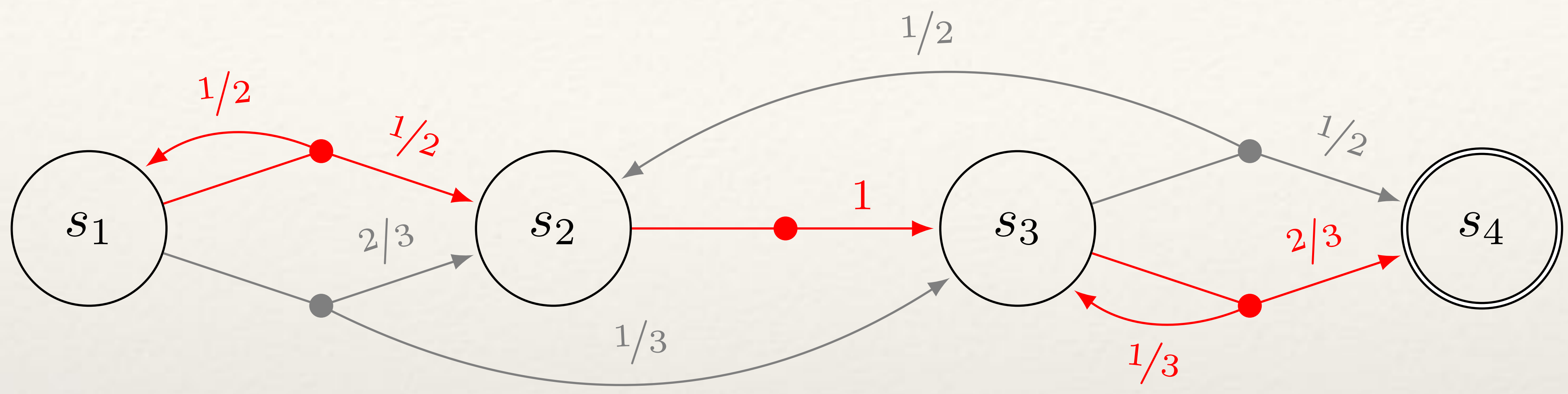


$M = (S, A, T)$

- S **states**
- $A(s) \subseteq A$ set of **actions** in state s
- $T(s' \mid s, a)$ **probability** of $s \to s'$ when choosing action a

Final states $F = \{s \mid A(s) = \varnothing\}$

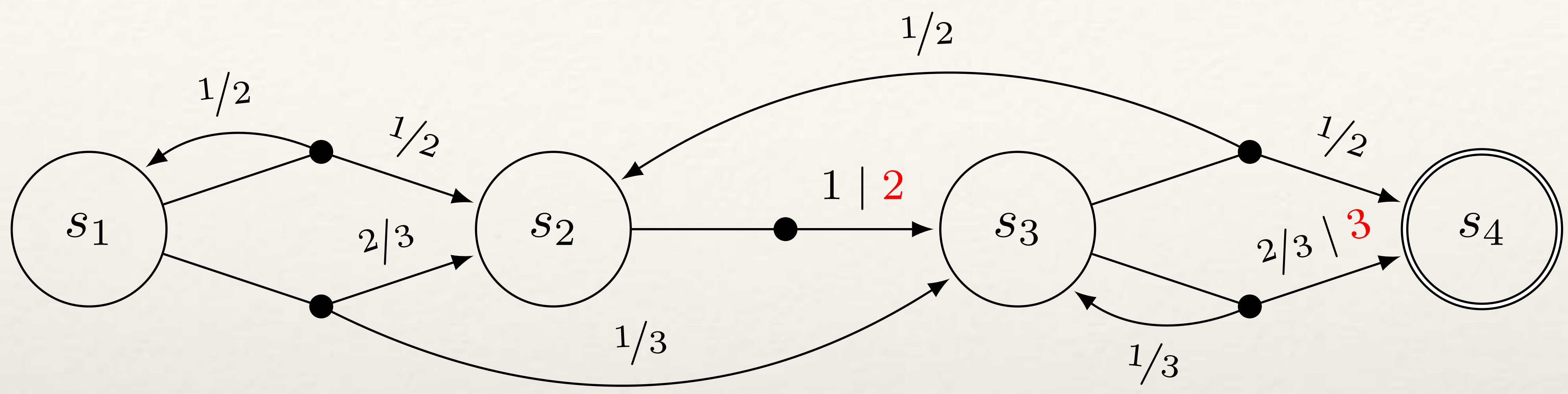
Markov Decision Processes



Policy: $\pi : S \setminus F \to A$ such that $\pi(s) \in A(s)$

M^π **Markov chain** induced by policy π

Markov Decision Processes



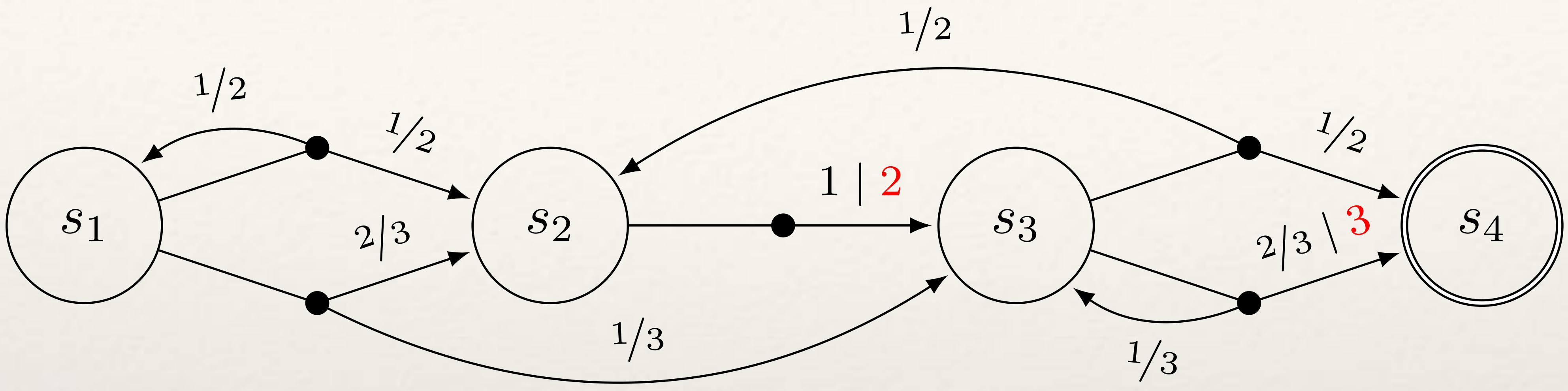
Possible **objectives** for the agent:

- **Reachability** / **avoidance** objectives (e.g., termination)
- Collect **reward** given by (step-wise) reward function

$$R : S \times A \times S \to \mathbb{R}$$

- …

Markov Decision Processes



Bellman optimality operator

Assume $S = \{1, \ldots, d\}$ and define $f : \mathbb{R}^d_{\geq 0} \to \mathbb{R}^d_{\geq 0}$ as

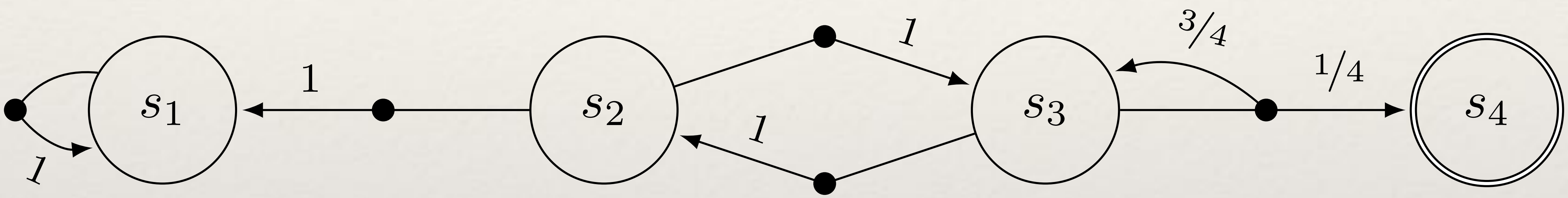
$$f(v)(s) = \max_{a \in A(s)} \sum_{s' \in S} T(s' \mid s, a) \cdot \big(R(s, a, s') + \gamma\, v(s')\big)$$

where $0 < \gamma \leq 1$ **discount factor** (**contractive** if $\gamma < 1$)

Then the **expected reward** is the **(least) fixpoint** μf

Markov Decision Processes

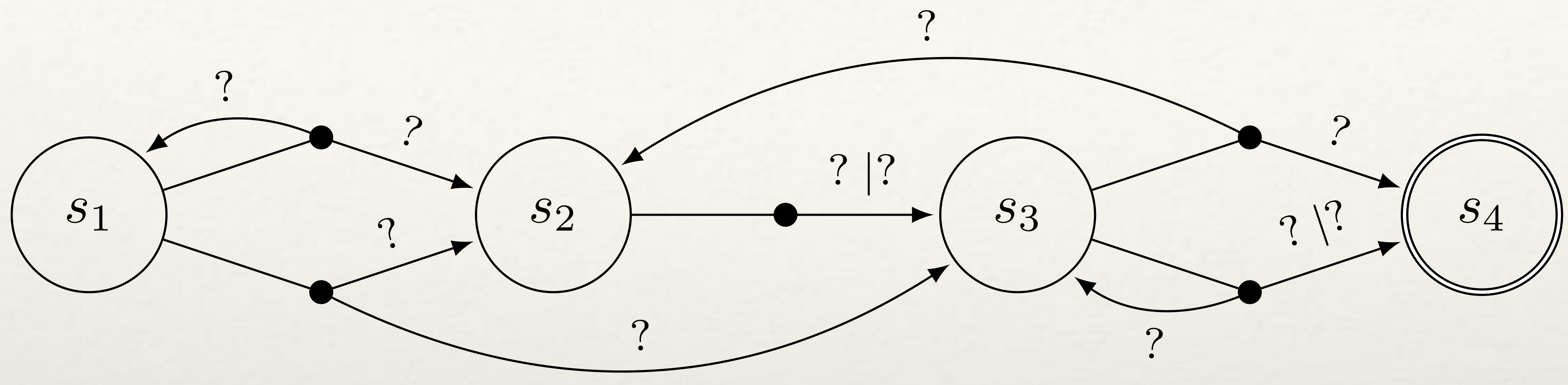
When $\gamma = 1$ … multiple fixpoints



0	1/2	1/2	1
0	1	1	1
1	1	1	1

We want the **least fixpoint** μf

Markov Decision Processes

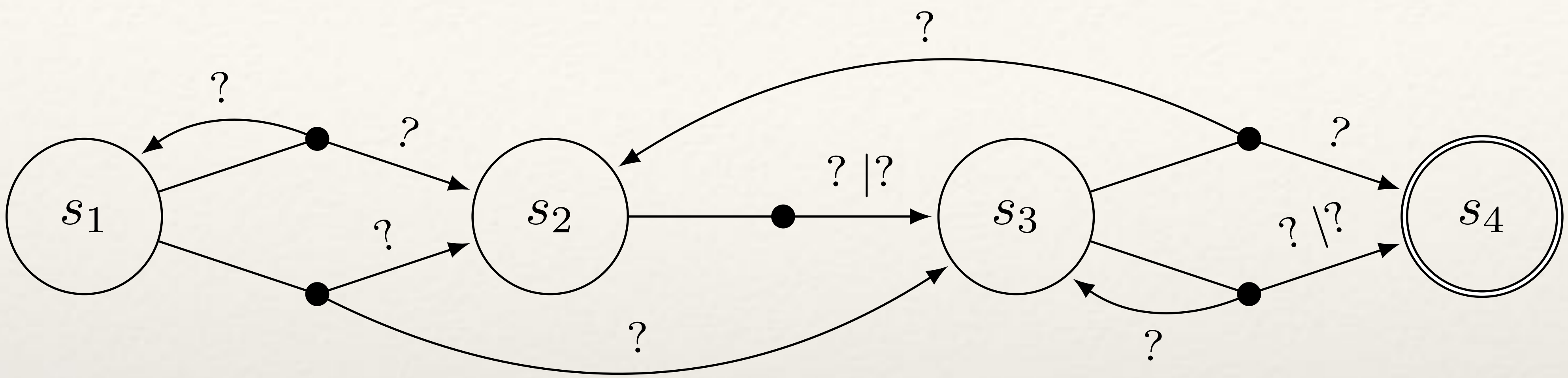


What if **probabilities** (and **rewards**) are **unknown** and can only be approximated **sampling** the MDP?

~ **Reinforcement learning**

Synthesise strategies for an MDP with unknown parameters while exploring it (Q-Learning, SARSA, . . .).

Markov Decision Processes



Typically, **contractive** case ($\gamma < 1$) is considered
$\rightsquigarrow$ fixpoint unique, errors removed by iterations, …

We focus on $\gamma = 1$ (**non-expansive**) and **model-based** (construct a model of the MDP while exploring it).

Quantitative fixpoint logics

Quantitative μ**-calculus** [Huth,Kwiatkowska], [Mio,Simpson]

$$\varphi \quad ::= \quad \mathbf{1} \mid \mathbf{0} \mid x \mid p \mid r \cdot \varphi \mid \max\{\varphi, \varphi'\} \mid \min\{\varphi, \varphi'\} \mid \Diamond\varphi \mid \Box\varphi \mid \mu x.\phi \mid \nu x.\phi$$

Formulae evaluated on an MDP as $[\![\varphi]\!]_\rho : S \to [0,1]$

$$[\![\Diamond\varphi]\!]_\rho(s) = \max_{a \in A(s)} \sum_{s' \in S} T(s' \mid s, a) \cdot [\![\varphi]\!]_\rho(s')$$

$$[\![\mu x.\varphi]\!]_\rho = \mu(\lambda v.[\![\varphi]\!]_{\rho \cup [x \mapsto v]})$$

non-expansive

Quantitative fixpoint logics

How to model check if the **MDP** can be only **approximated**?

Fixpoints can be nested
With arbitrary (non-expansive) operators outer fixpoints have to deal with approximated functions.

The setting, again

$X = [0, +\infty)^d$, with

- **pointwise order** $\leq$ and
- **sup-distance** $d(x, y) = \|x - y\|_\infty = \max_{1 \leq i \leq n} |x_i - y_i|$

Monotone and **non-expansive** function

$$f : X \to X$$

(When f has fixpoints, the least fixpoint μf exists and given by Kleene's iteration)

The aim

Given a sequence of **approximations** $f_n : X \to X$ with

$$f_n \to f \qquad \text{(uniformly)}$$

Construct a sequence approximating μf

$$x_n \to \mu f$$

Non-expansiveness wrt. sup distance covers many cases: termination probabilities in Markov chains, MDPs, stochastic games, behavioural metrics, …

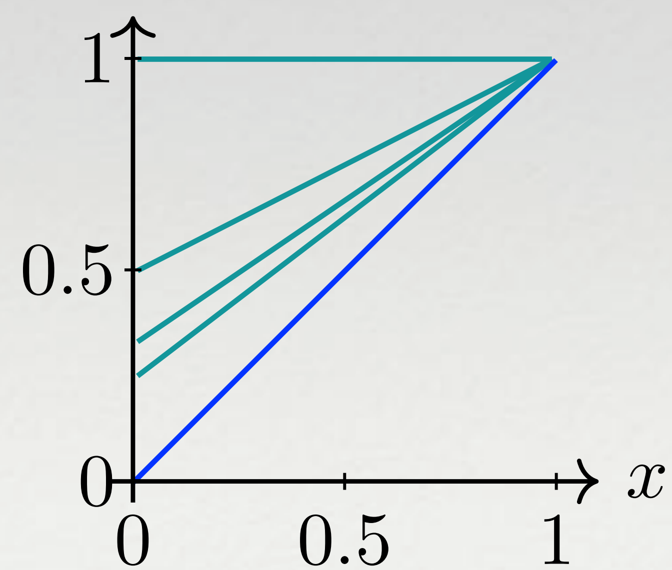
Problem: Discontinuity of μf

If f_n **converges** to f in the **order** (f_n chain and $f = \bigsqcup f_n$)

$$\mu f_n \to \mu f$$

For metric convergence the above fails

Example: $X = [0,1]$ $\quad f(x) = x \quad f_n(x) = \frac{1}{n} + (1 - \frac{1}{n}) \cdot x$



$$\mu f = 0$$

$$\mu f_n = 1$$

Problem: Unrecoverable errors

When $\mu f_n \to \mu f$ in principle one can

- take a sufficiently good approximation f_n
- perform Kleene iteration for f_n

But, when considering more precise approximations the previous computation is not reusable !

Mixing approximation/Kleene iteration does not work

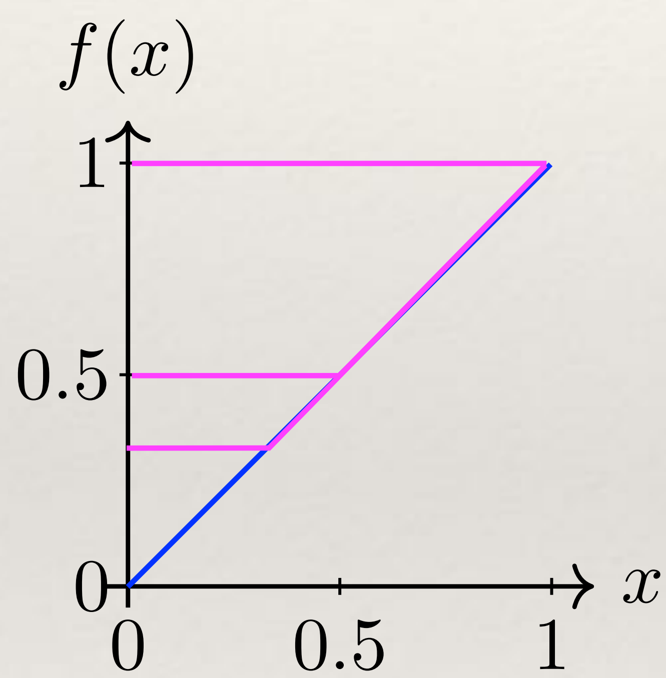
$$\begin{cases} x_0 = \perp \\ x_{n+1} = f_n(x_n) \end{cases}$$

… in general $x_n \not\to \mu f$

Problem: Unrecoverable errors

Example: $X = [0,1]$ $\quad f(x) = x$

$$f_n(x) = \begin{cases} 1/n & \text{if } x \leq 1/n \\ x & \text{otherwise} \end{cases}$$



Independently of how good the first approximation is, it overapproximates μf and next approximations will not decrease it

Dampened Mann Iteration

$$x_{n+1} = \boxed{(1-\beta_n)} \cdot \boxed{\big(\alpha_n \cdot x_n + (1-\alpha_n) \cdot f_n(x_n)\big)}$$

dampening factor

Mann iteration

- $\lim_{n\to\infty} \alpha_n < 1$,
- $\lim_{n\to\infty} \beta_n = 0$ and $\sum_{n=1}^{\infty} \beta_n = \infty$
- $\sum_{n=1}^{\infty} |\alpha_n - \alpha_{n+1}| < \infty$ and $\sum_{n=1}^{\infty} |\beta_n - \beta_{n+1}| < \infty$.

Canonical choices: $\beta_n = 1/n$ and $\alpha_n = 1/n$ (or $\alpha_n = 0$)

Exact case ("sanity check")

Iterating the **exact function**, i.e., with $f_n = f$ then

$\boxed{x_0 \quad \text{arbitrary}}$

$$x_{n+1} = (1 - \beta_n) \cdot \left(\alpha_n \cdot x_n + (1 - \alpha_n) \cdot f_n(x_n)\right)$$

It holds

$$x_n \to \mu f$$

Dealing with Approximations

Assume $f_n \to f$ and consider Dampened Mann iteration

$$x_0 \quad \text{arbitrary}$$

$$x_{n+1} = (1-\beta_n)\cdot\big(\alpha_n \cdot x_n + (1-\alpha_n)\cdot f_n(x_n)\big)$$

It holds $x_n \to \mu f$ when

1. f is a **(power) contraction** (f^k contraction for some k)
2. $\mu f_n \to \mu f$ and $f_n \to f$ **monotonically**
3. $f_n \to f$ **normally**

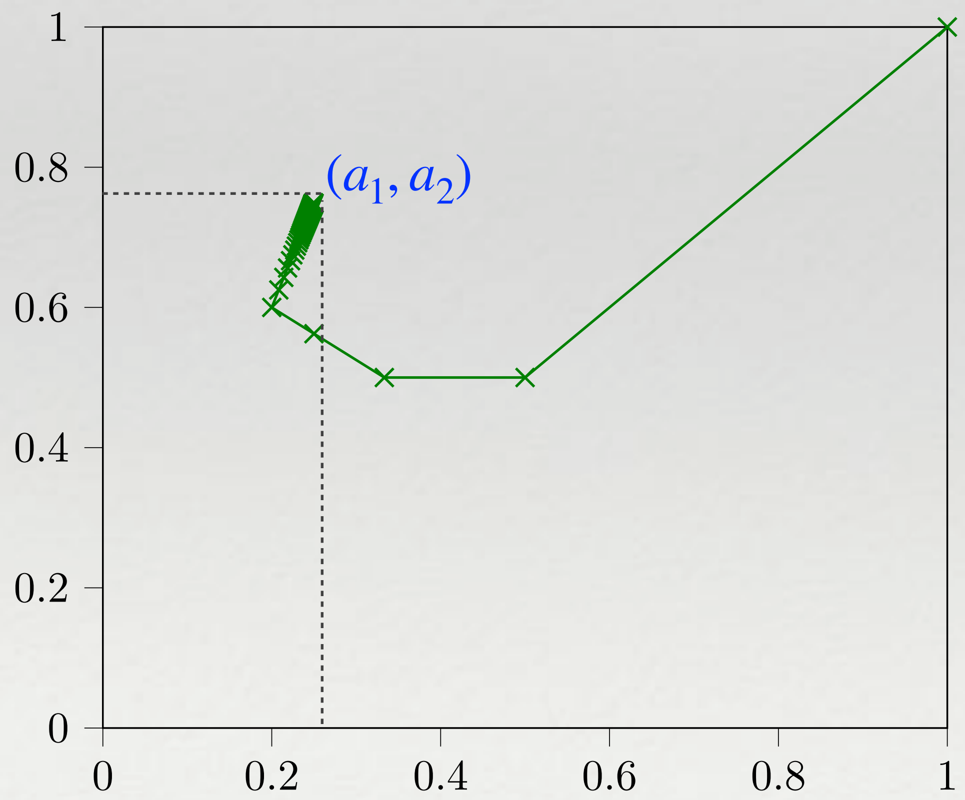
$$\sum \|f_n - f\|_\infty < \infty$$

Example

$f : [0,1]^2 \to [0,1]^2 \qquad f(x_1, x_2) = (\max\{x_1, a_1\}, \max\{x_2, a_2\})$

$(a_1, a_2) = (1/4, 3/4)$

$f_n(x_1, x_2) = (\max\{x_1, (1 - a_1)x_1^n + a_1\}, \max\{x_2, (1 - a_2)x_2^n + a_2)$



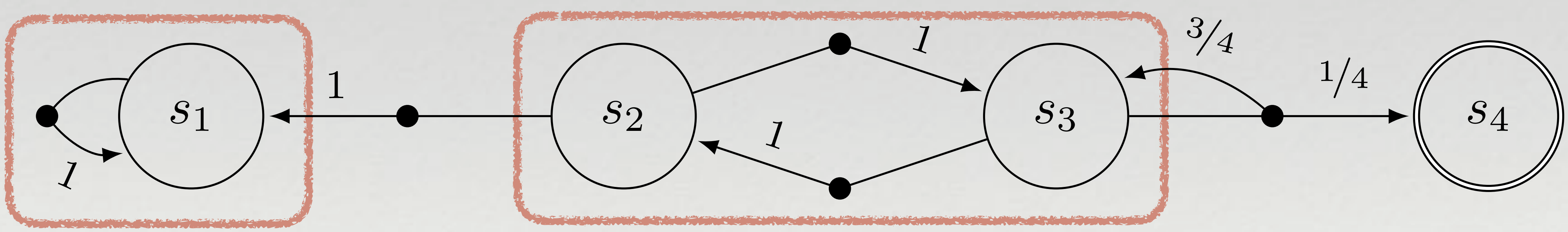
Dampened Mann Iteration

$(\alpha_n = 0)$

Dealing with MDPs

For MDPs the previous condition do not hold in general, still we can have positive results

End Component in an MDP
~ Set of states E inducing a strongly connected sub-MDP
$\Longrightarrow$ there is a policy that once in E, stays in E



Powercontractions?

For an MDP M, the following are equivalent:

- M has **no end components** (**terminating** MDP)
- the Bellmann function f_M is a **power-contraction**.

Dampened Mann “works” for terminating MDPs

General MDPs?

With end components function f_M is just **non-expansive**

Assuming that **rewards** are given **only outside** end components, the **expected reward** is

- **bounded**
- **identical** for all states in an end-component
- computed **quotienting** maximal end components, obtaining a terminating MDP

Using the above, one can show that Dampened Mann works for MDPs with end components.

With end components?

Ideas

- When sampling, once non-zero approximated probabilities will never converge to 0
- End components will eventually stabilise
- The error due to approximations when leaving a end component behaves similarly to rewards and vanishes with better approximations

What to do in general?

What to do in cases that escapes all sufficient conditions (e.g., stochastic games)?

Consider a subsequence of functions $f_{n_1}, f_{n_2}, f_{n_3}, \ldots$ that converges **normally** to f

Intuition
Perform enough samplings before the next iteration.

Hoeffding's inequality
$$\mathbb{P}\left[\,|T^n(s' \mid s, a) - T(s' \mid s, a)| > \varepsilon\right] \leq 2e^{-2n\varepsilon^2}$$

What to do in general?

Choose n_i such that

$$\mathbb{P}\left[\|f_{n_i} - f\| > \gamma_i\right] \leq \delta_i$$

where $\Sigma_i \gamma_i < \infty$ and $\Sigma_i \delta_i < \infty$ (e.g., $\gamma_i = \delta_i = 1/i^2$)

By **Borel-Cantelli** Lemma

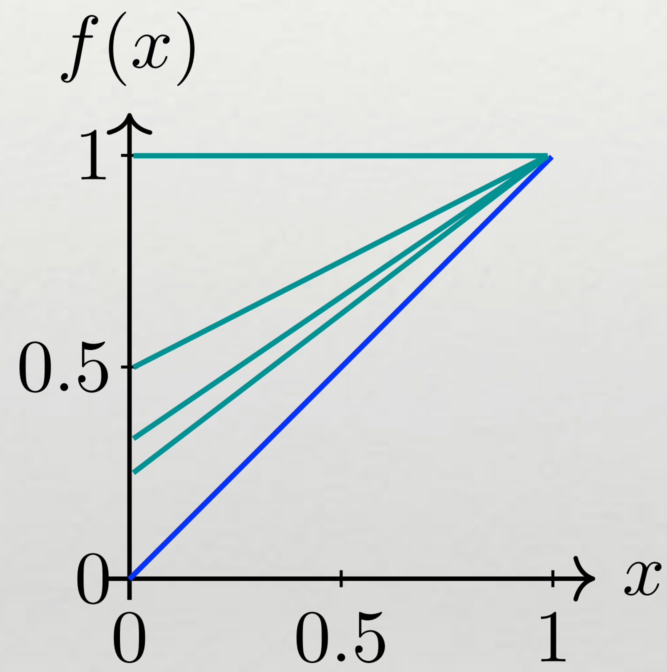
$$\mathbb{P}\left[\|f_{n_i} - f\| > \gamma_i \text{ for infinitely many } i\right] = 0$$

$\|f_{n_i} - f\| \leq \gamma_i$ (almost sure normal convergence)

Almost surely $x_n \to \mu f$

Example

Example: $X = [0,1]$ $\quad f(x) = x$ $\quad f_n(x) = \frac{1}{n} + (1 - \frac{1}{n}) \cdot x$



$\mu f = 0$

$\mu f_n = 1$

Dampened Mann works with the sequence $(f_{n^2})_n$ works (despite $\mu f_n \not\to \mu f$)

Conclusions

Runtime and accuracy after n steps ~ to computing μf_n by Kleene iteration.

Quantitative mu-calculi?

Chaotic iteration?

Negative rewards?

Conclusions

What if $\alpha_n \to 1$ (Mann iteration converging to identity)

$$x_{n+1} = (1 - \beta_n) \cdot \left(\alpha_n \cdot x_n + (1 - \alpha_n) \cdot f_n(x_n)\right)$$

~ Model-free learning

Idea: $f_n \to f$ in the limit-average $(\frac{1}{n}\sum_{i=1}^{n} f_i \to f)$

model-free RL algorithms as special cases?