An anytime algorithm for reachability in uncountable MDPs

Kush Grover

Technical University of Munich

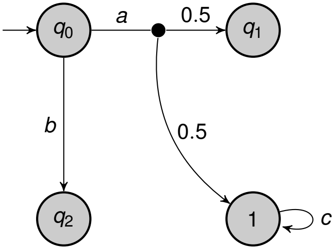
Joint work with **J. Křetínský**, **T. Meggendorfer** and **M. Weininger**

Schedule

1. Uncountable MDPs
2. Reachability Problem
3. Assumptions
4. Algorithm

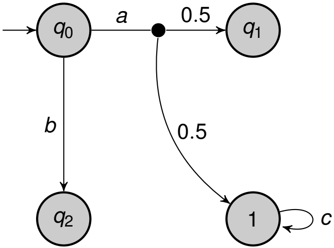
Uncountable MDPs

MDPs



(S, Act, Av, Δ)

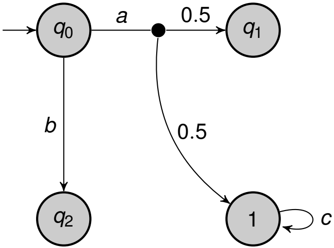
MDPs



$$(S, Act, Av, \Delta)$$

$$S = \{q_0, q_1 \dots\}$$

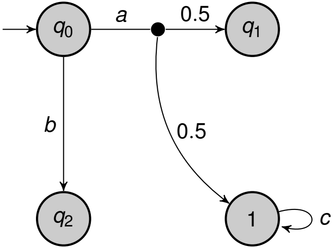
MDPs



$$(S, Act, Av, \Delta)$$

$$Act = \{a, b, c \dots\}$$

MDPs

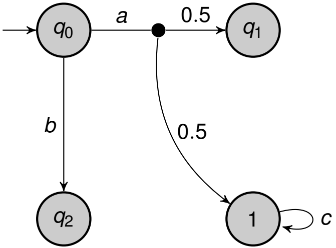


(S, Act, Av, Δ)

$Av : S \to Act$

$Av(q_0) = \{a, b\}$
$Av(1) = \{c\}$

MDPs

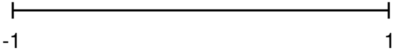


$$(S, Act, Av, \Delta)$$

$$\Delta : S \times Act \to Dist(S)$$

Uncountable MDPs

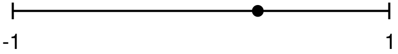
$\mathcal{M} = (S$



S: Compact metric space

Uncountable MDPs

$\mathcal{M} = (S, Act$

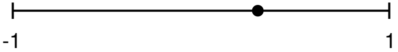


Act: Compact metric space

For e.g. $Act = [-1, 1]$

Uncountable MDPs

$$\mathcal{M} = (S, Act, Av$$

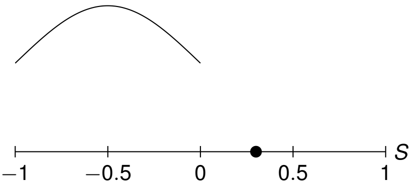


$$Av\colon S \to \Sigma_{Act} \setminus \{\phi\}$$

$$Av(s) = [-1, 1]$$

Uncountable MDPs

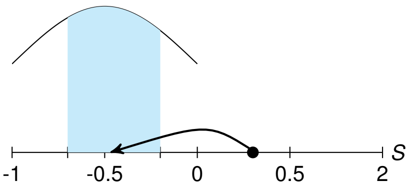
$$\mathcal{M} = (S, Act, Av, \Delta)$$



$$\Delta \colon S \times Act \to \Pi(S)$$

Uncountable MDPs

$$\mathcal{M} = (S, Act, Av, \Delta)$$



$$\Delta\colon S \times Act \to \Pi(S)$$

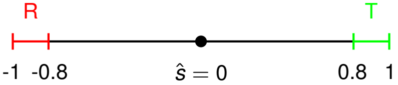
Reachability Problem

Reachability in Uncountable MDPs

Find the probability of reaching a target set T from an initial state $\hat{s}$ ($V(\hat{s})$).

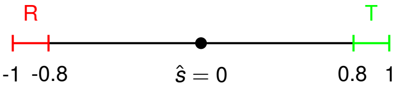
Reachability in Uncountable MDPs

Find the probability of reaching a target set T from an initial state $\hat{s}$ ($V(\hat{s})$).



Reachability in Uncountable MDPs

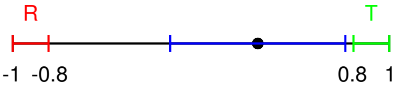
Find the probability of reaching a target set T from an initial state $\hat{s}$ ($V(\hat{s})$).



$$\Delta(s, a) = \Delta(s) = unif(s - a_c \frac{0.8 - s}{0.8}, s + a_c \frac{0.8 - s}{0.8})$$

Reachability in Uncountable MDPs

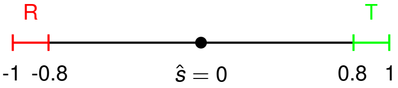
Find the probability of reaching a target set T from an initial state $\hat{s}$ ($V(\hat{s})$).



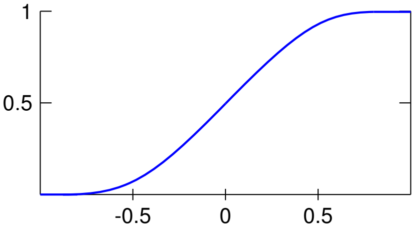
$$\Delta(s, a) = \Delta(s) = unif(s - a_c \frac{0.8 - s}{0.8}, s + a_c \frac{0.8 - s}{0.8})$$

Reachability in Uncountable MDPs

Find the probability of reaching a target set T from an initial state $\hat{s}$ ($V(\hat{s})$).

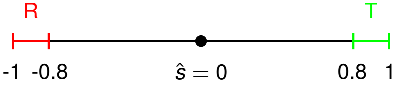


$$\Delta(s,a) = \Delta(s) = unif(s - a_c\frac{0.8-s}{0.8}, s + a_c\frac{0.8-s}{0.8})$$



Reachability in Uncountable MDPs

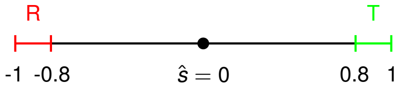
Find the probability of reaching a target set T from an initial state $\hat{s}$ ($V(\hat{s})$).



Computing the exact value $V(\hat{s})$ is undecidable.

Reachability in Uncountable MDPs

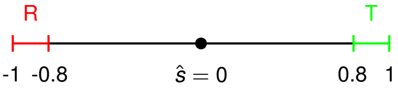
Find the probability of reaching a target set T from an initial state $\hat{s}$ ($V(\hat{s})$).



Next best: Compute approximate values with a converging bound on the error.

Reachability in Uncountable MDPs

Find the probability of reaching a target set T from an initial state $\hat{s}$ ($V(\hat{s})$).

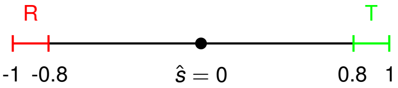


Next best: Compute approximate values with a converging bound on the error.

I

Reachability in Uncountable MDPs

Find the probability of reaching a target set T from an initial state $\hat{s}$ ($V(\hat{s})$).

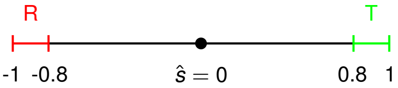


Next best: Compute approximate values with a converging bound on the error.

$$V(\hat{s}) \in \ I$$

Reachability in Uncountable MDPs

Find the probability of reaching a target set T from an initial state $\hat{s}$ ($V(\hat{s})$).



Next best: Compute approximate values with a converging bound on the error.

$$V(\hat{s}) \in I, |I| \leq \epsilon$$

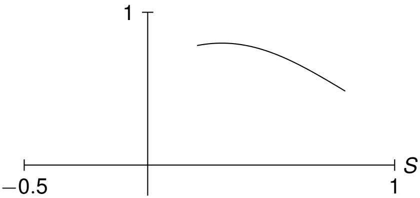
Reachability in Uncountable MDPs

Find the probability of reaching a target set T from an initial state $\hat{s}$ ($V(\hat{s})$).

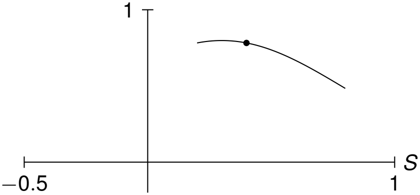
Solution: Extend BRTDP (Křetínský et. al. '14) to the uncountable setting.

Assumptions

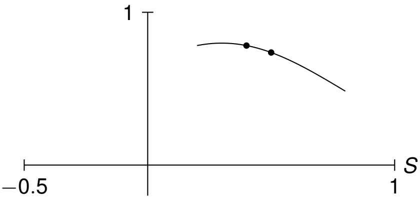
Lipschitz Continuity



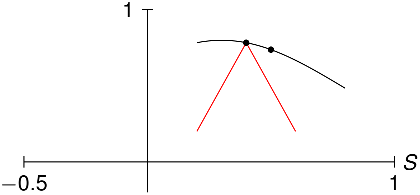
Lipschitz Continuity



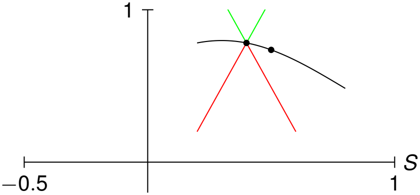
Lipschitz Continuity



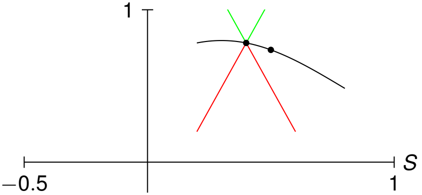
Lipschitz Continuity



Lipschitz Continuity

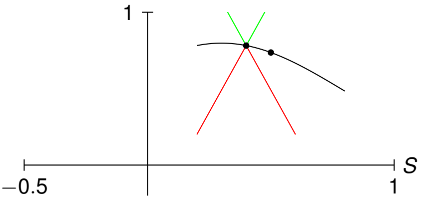


Lipschitz Continuity



The value functions are **Lipschitz continuous**.

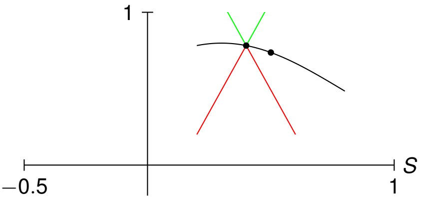
Lipschitz Continuity



The value functions are **Lipschitz continuous**.

$$|V(s) - V(s')| \leq K \cdot d(s, s')$$

Lipschitz Continuity



The value functions are **Lipschitz continuous**.

$$|V(s) - V(s')| \leq K \cdot d(s, s')$$

$$|V(s, a) - V(s', a')| \leq K_{\times} \cdot d_{\times}((s, a), (s', a'))$$

State-Action Maximum Approximation

s

State-Action Maximum Approximation

$$f\colon Av(s) \to [0,1]$$

State-Action Maximum Approximation

Lipschitz

$f \colon Av(s) \to [0, 1]$

State-Action Maximum Approximation

Lipschitz Computable

$$f\colon Av(s) \to [0,1]$$

State-Action Maximum Approximation

Lipschitz   Computable

$$f\colon Av(s) \rightarrow [0,1]$$

We can under (and over) approximate the value $\max_{a \in Av(s)} f(a)$ arbitrarily close.

Transition Approximation

$$g \colon S \to [0, 1]$$

Transition Approximation

Lipschitz

$$g\colon S \to [0,1]$$

Transition Approximation

Lipschitz Computable

$$g\colon S \to [0,1]$$

Transition Approximation

Lipschitz Computable

$$g\colon S \to [0,1]$$

We can under (and over) approximate the value $\Delta(s,a)\langle g\rangle$.

State-Action Sampling

Sampling fairness

State-Action Sampling

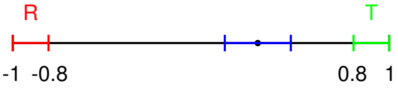
Sampling fairness

Always eventually sample “near" all the reachable state-action pair.

State-Action Sampling

Sampling fairness

Always eventually sample “near" all the reachable state-action pair.



Sink Computability and Attractor

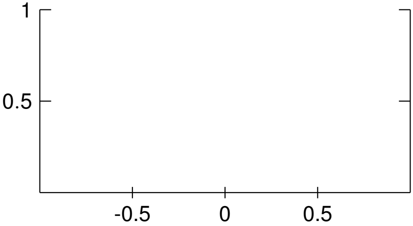
Sets T and R are **decidable** and **measurable**.

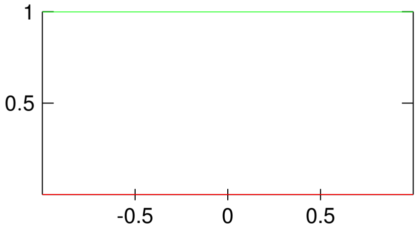
For any state s and strategy π we have $Pr^{\pi}_{\mathcal{M},s}[\diamond(T \cup R)] = 1$.

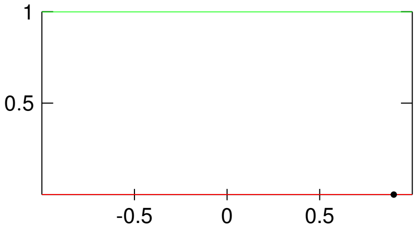
Summary of assumptions

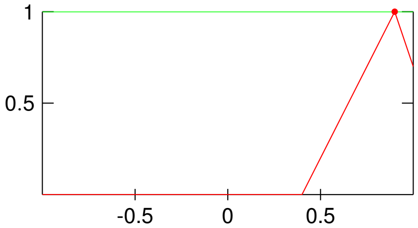
- Lipschitz continuity.
- State-Action Maximum approximation.
- Transition approximation.
- State-Action sampling.
- Sink Computability and Attractor.

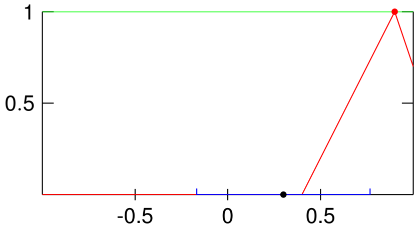
Algorithm

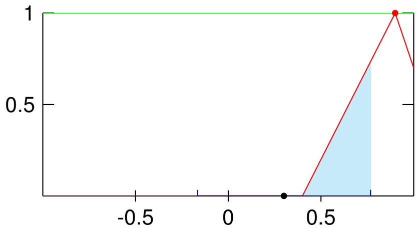

1
0.5
-0.5
0
0.5


1
0.5
-0.5
0
0.5


1
0.5
-0.5
0
0.5

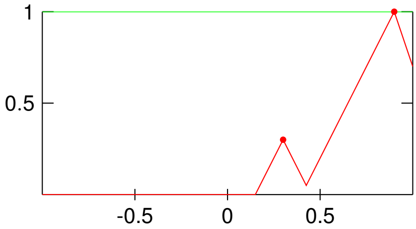

1
0.5
-0.5
0
0.5


1
0.5
-0.5
0
0.5

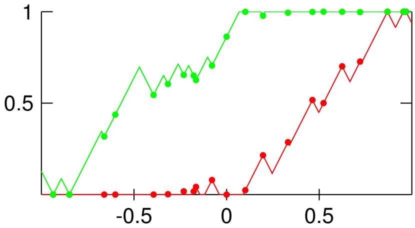


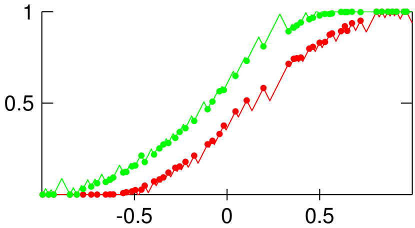
Compute expected value of L under $\Delta(s, a)$ i.e.

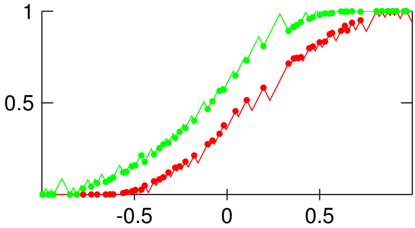
$$L_{new}(s, a) = \Delta(s, a)\langle L \rangle.$$



1
0.5
-0.5
0
0.5


1
0.5
-0.5
0
0.5

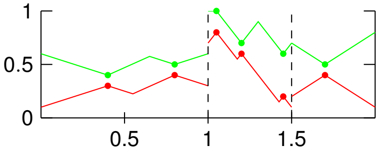

1
0.5
-0.5
0
0.5



Anytime

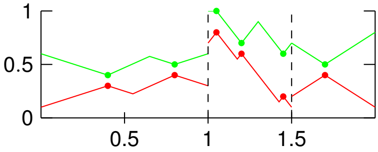
Possible extensions

- **Discontinuities:**



Possible extensions

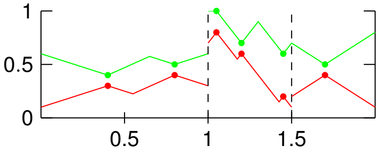
- **Discontinuities:**



- **LTL:** Can handle “reach-avoid" properties directly.

Possible extensions

- **Discontinuities:**



- **LTL:** Can handle “reach-avoid" properties directly.
- **Apply learning:** Use learning heuristics to guide the algorithm.

Implementation

- Prototype implementation in python.
- Evaluated it on the example showed earlier.

Conclusion and Future work

Conclusion:

- We gave an anytime algorithm for reachability under mild assumptions.
- Guaranteed converging bound on the error.

Future work:

- Extend implementation which can handle uncountable action spaces and some discontinuities of the value function.