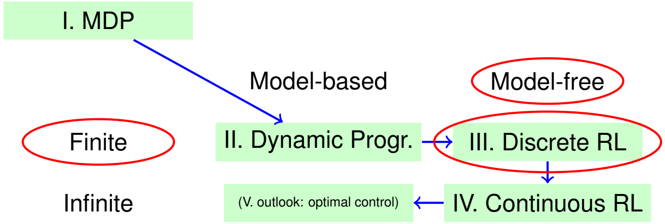
AIC/RL – Continuous Model-Free RL (Part IV)
Function Approximation Basics

Freek Stulp

Université Paris-Saclay

Where are we?



Finite MDP

S State space — *"all possible states the environment can have"*

A Action space — *"all possible actions the agent can take"*

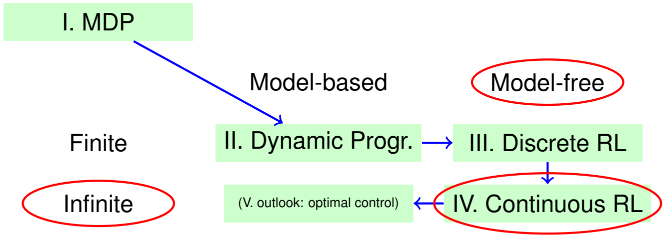
$\mathcal{P}^a_{ss'}$ Transition function — *"probability of going from s to s′ when doing a"*

$$\mathcal{P}^a_{ss'} = Pr\{s_{t+1} = s' | s_t = s, a_t = a\}$$

$\mathcal{R}^a_{ss'}$ Reward function — *"immediate reward in s / when going from s to s′"*

$$\mathcal{R}^a_{ss'} = \mathrm{E}\{r_{t+1} | s_t = s, a_t = a, s_{t+1} = s'\}$$

Where are we?



Continuous, infinite MDPs

- S State space $S \subseteq \mathbb{R}^{D_S}$ (D_S-dimensional vector)
- A Action space $A \subseteq \mathbb{R}^{D_A}$ (D_A-dimensional vector)
- f Transition rate function $f : S \times A \to \Delta S$
- r Reward function $r : S \times A \to \mathbb{R}$

- Bad news: infinite number of states and actions...
- Good news: smoothness, i.e. ΔS usually not so big

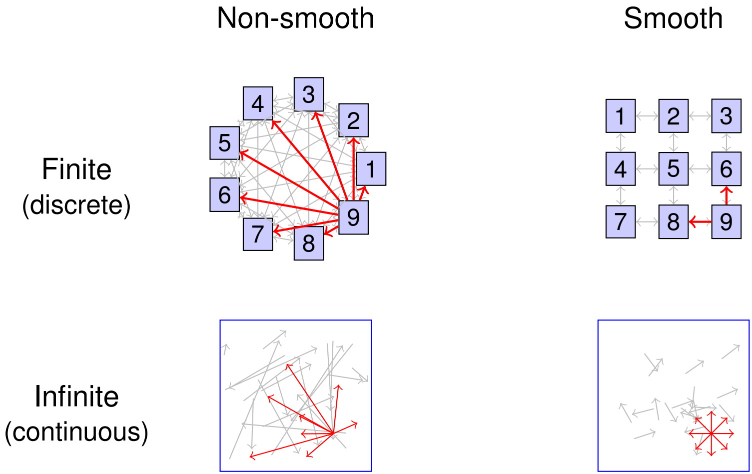
- Two solution strategies
  1. Function Approximation
  2. Direct Policy Search

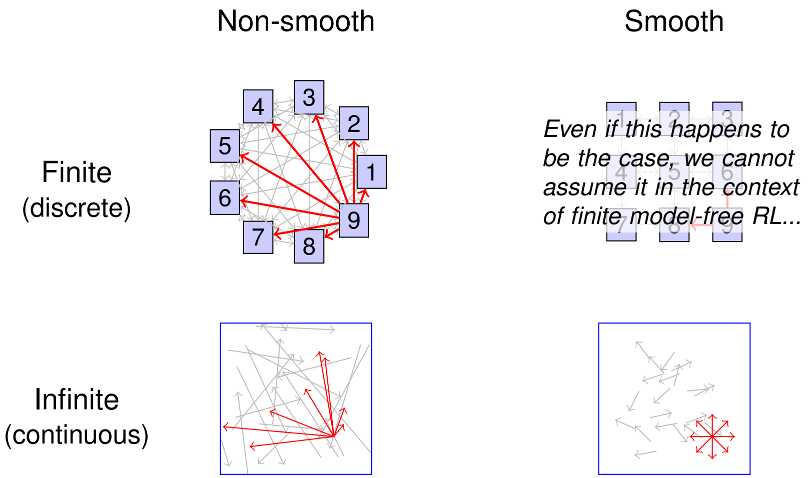
Function Approximation

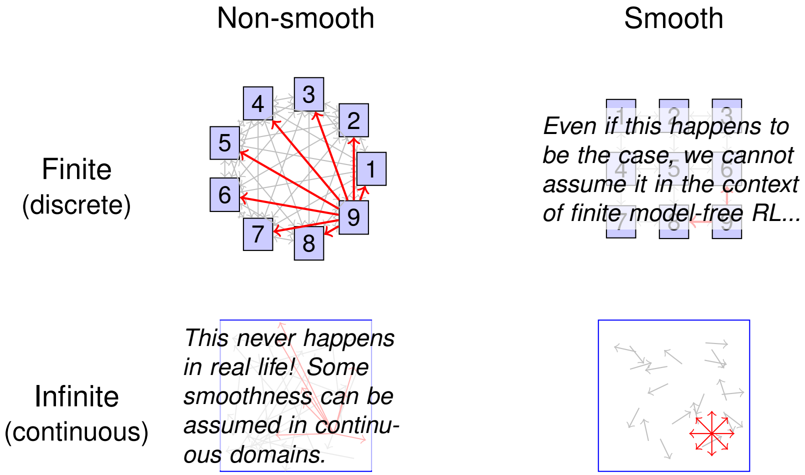
Today Only very basic introduction of the key idea

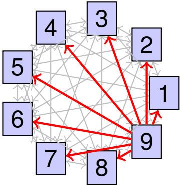
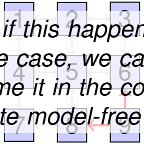
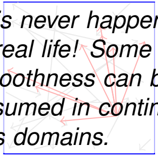
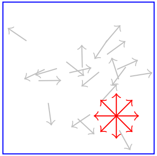
- To have more time to progress with coding

Next time Learning with function approximation

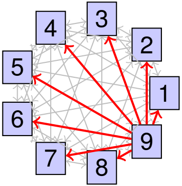
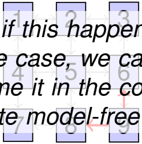
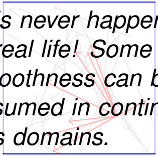
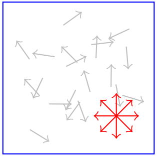

Non-smooth
Smooth
Finite
(discrete)
Infinite
(continuous)
1
2
3
4
5
6
7
8
9
1
2
3
4
5
6
7
8
9


Non-smooth
Smooth
Finite
(discrete)
Infinite
(continuous)
1 2 3 4 5 6 7 8 9
Even if this happens to be the case, we cannot assume it in the context of finite model-free RL...


Non-smooth
Smooth
Finite
(discrete)
Infinite
(continuous)
1 2 3 4 5 6 7 8 9
Even if this happens to be the case, we cannot assume it in the context of finite model-free RL...
This never happens in real life! Some smoothness can be assumed in continuous domains.

	Non-smooth	Smooth
Finite (discrete)		 *Even if this happens to be the case, we cannot assume it in the context of finite model-free RL...*
Infinite (continuous)	 *This never happens in real life! Some smoothness can be assumed in continuous domains.*	

- Assumptions in inifite, continuous problems
  1. transitions in state space will be (mostly) smooth ($f : S \times A \rightarrow \Delta S$)
  2. similar actions have similar effects

	Non-smooth	Smooth
Finite (discrete)		*Even if this happens to be the case, we cannot assume it in the context of finite model-free RL...*
Infinite (continuous)	*This never happens in real life! Some smoothness can be assumed in continuous domains.*	

- Assumptions in inifite, continuous problems
  1. transitions in state space will be (mostly) smooth ($f : S \times A \rightarrow \Delta S$)
  2. similar actions have similar effects
- Apply *function approximation*

Value Prediction with Function Approximation (SuBa8.1)

With lookup tables

- $V^{\pi}(s)$ or $Q^{\pi}(s,a)$: stored in lookup tables, i.e. and array (V) or matrix (Q)

With function approximation

- $V^{\pi}_{\theta}(s)$ or $Q^{\pi}_{\theta}(s,a)$ represented as a function approximator
  - with parameters θ
- Example: $V^{\pi}_{\theta}(s)$ represented as a (deep) neural network
  - each input neuron corresponds to one dimension of s
  - output neuron is the estimated value V
  - θ contains weights of the neural network

Example: Radial Basis Functions

$$\phi^i(x) = \exp\left(-\frac{\|x - c_i\|^2}{2\sigma_i^2}\right) \tag{1}$$

$$f(x) = \sum_{i=1}^{n} \theta^i \phi^i(x) \tag{2}$$

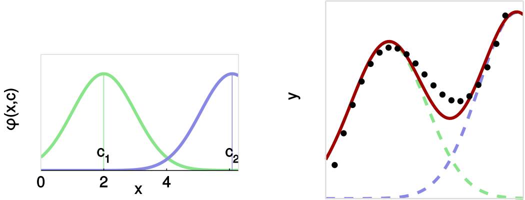


Figure : Radial Basis Function Network

Example: Radial Basis Functions

$$\phi^i(x) = \exp\left(-\frac{\|x - c_i\|^2}{2\sigma_i^2}\right) \tag{1}$$

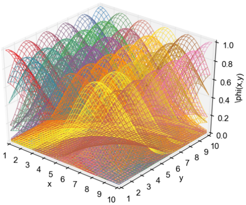
$$f(x) = \sum_{i=1}^{n} \theta^i \phi^i(x) \tag{2}$$

$y =$	1	2	3	4
$x = 1$	T	100	99	98
$x = 2$	99	98	97	96
$x = 3$	98	97	96	95
$x = 4$	97	96	95	94
$x = 5$	96	95	94	93
$x = 6$	95	94	93	92
$x = 7$	94	93	92	91
$x = 8$	93	92	91	90
$x = 9$	92	91	90	89
$x = 10$	91	90	89	88

Example: Radial Basis Functions

$$\phi^i(s) = \exp\left(-\frac{\|s - c_i\|^2}{2\sigma_i^2}\right) \tag{1}$$

$$V_\theta(s) = f(s) = \sum_{i=1}^{n} \theta^i \phi^i(s) \tag{2}$$

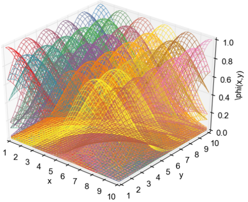


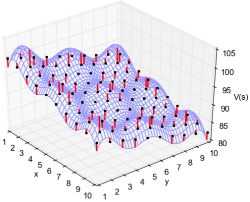
$y =$	1	2	3	4
$x = 1$	T	100	99	98
$x = 2$	99	98	97	96
$x = 3$	98	97	96	95
$x = 4$	97	96	95	94
$x = 5$	96	95	94	93
$x = 6$	95	94	93	92
$x = 7$	94	93	92	91
$x = 8$	93	92	91	90
$x = 9$	92	91	90	89
$x = 10$	91	90	89	88

Example: Radial Basis Functions

$$\phi^i(s) = \exp\left(-\frac{\|s - c_i\|^2}{2\sigma_i^2}\right) \tag{1}$$

$$V_\theta(s) = f(s) = \sum_{i=1}^{n} \theta^i \phi^i(s) \tag{2}$$





$y =$	1	2	3	4
$x = 1$	T	100	99	98
$x = 2$	99	98	97	96
$x = 3$	98	97	96	95
$x = 4$	97	96	95	94
$x = 5$	96	95	94	93
$x = 6$	95	94	93	92
$x = 7$	94	93	92	91
$x = 8$	93	92	91	90
$x = 9$	92	91	90	89
$x = 10$	91	90	89	88

Up next in the course

- More about function approximation
- Direct policy search
- Case study
    - learning to play atari games

Today's exercise

- Continue with discrete RL
  - Monte Carlo to learn V (like policy evaluation)
  - Monte Carlo to learn Q (like policy evaluation)
  - Implement ϵ-greedy exploration (like value iteration)
  - Temporal Differencing (use immediate reward r_t instead of return R)