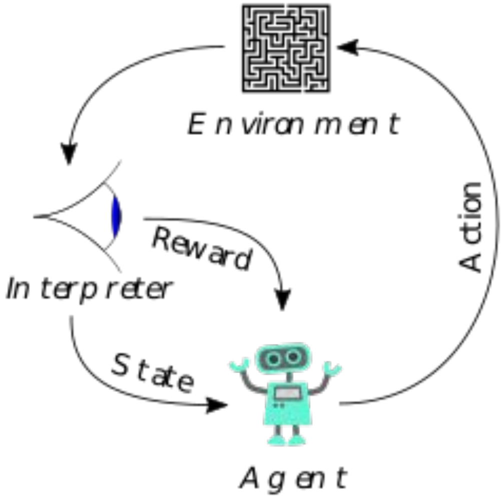
Reinforcement learning from the basics: a tale of learning and control

Eloïse Berthier
Inria Paris – Junior Seminar
June 21, 2022

1. What is reinforcement learning?

What is reinforcement learning?

It defines ways for an agent to **learn to behave** in an **unknown environment**, in order to **maximize** an expected **future reward**.



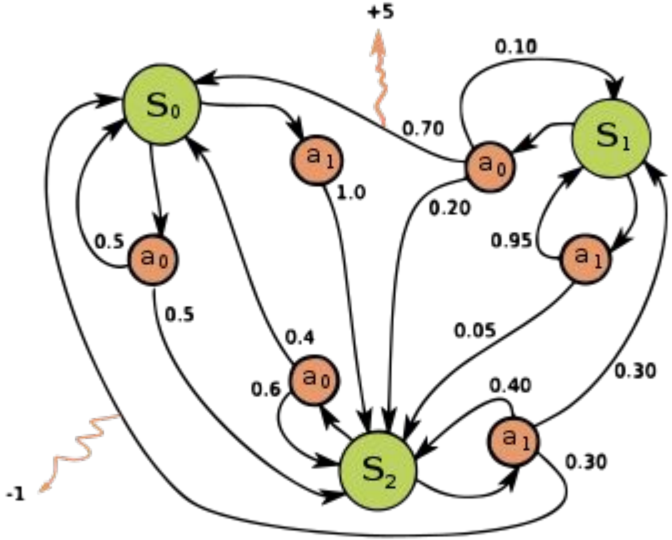
What is reinforcement learning?

The **environment** is modelled by a Markov Decision Process (MPD):

- a set of **states** S
- a set of **actions** A
- the **transition** probabilities

$$P_a(s, s') = \mathbb{P}(s_{t+1} = s' | s_t = s, a_t = a)$$

- a **reward** function $R_a(s, s')$
- a **discount factor** $\gamma \in [0, 1)$



What is reinforcement learning?

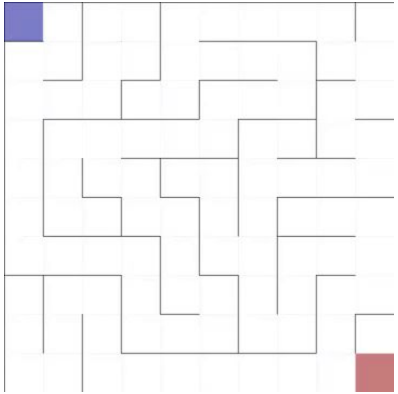
Interacting with the MDP, the aim is to **find a policy** π that **maximizes**:

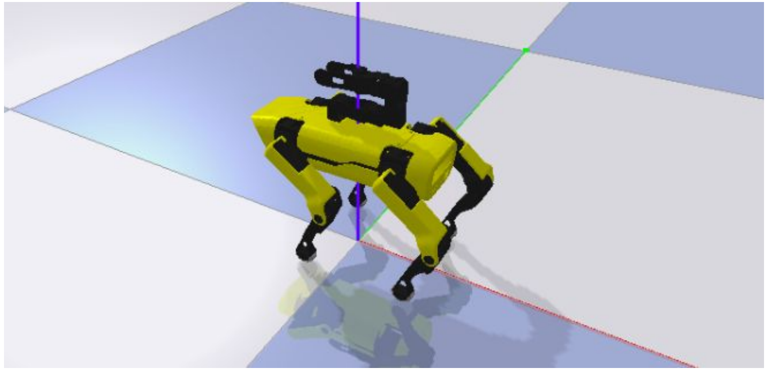
$$J(\pi) = \mathbb{E}\Big[\sum_{t=0}^{\infty} \gamma^t R_{\pi(s_t)}(s_t, s_{t+1})\Big]$$

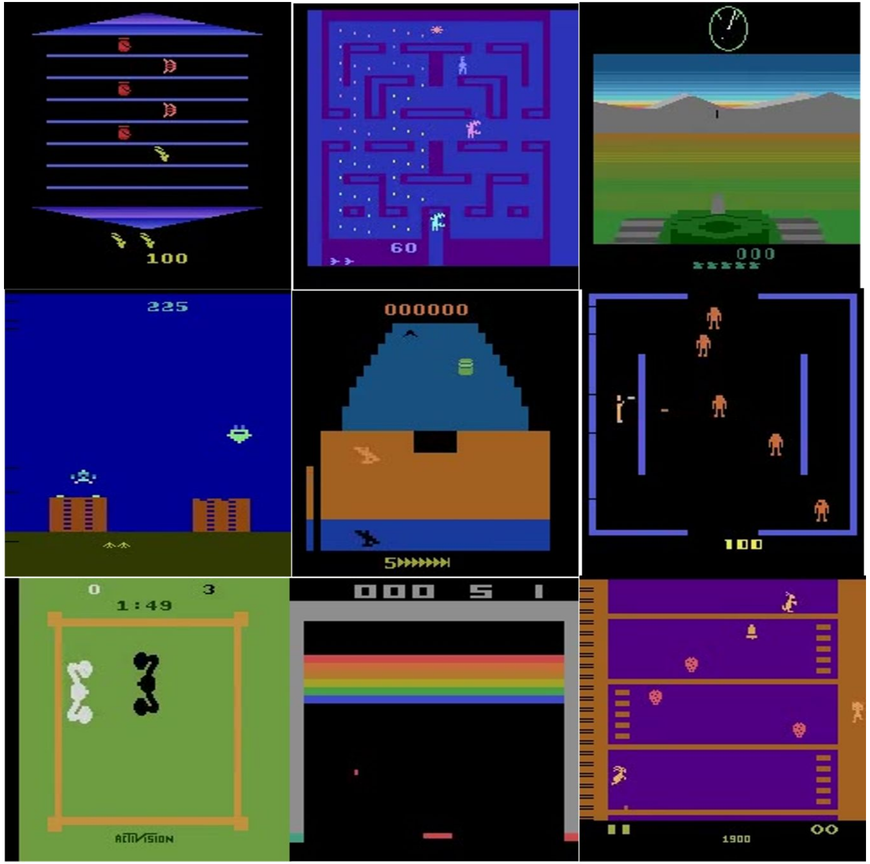
This is an **optimization** problem… but rather hard to solve:

- the MDP is unknown,
- ideally the agent must optimize and interact at the same time
- what if the state or action is a continuous variable?

Examples of environments







Recent successes

Dynamic Programming

One of the simplest RL algorithms is **value iteration**:

The value function $V(s) = \sup_\pi \mathbb{E}\left[\sum_{t=0}^{\infty} \gamma^t R_{\pi(t)}(s_t, s_{t+1}) \middle| s_0 = s\right]$ is a solution of the fixed-point equation:

$$\forall s, V(s) = \sup_a \mathbb{E}\left[R_a(s, s') + \gamma V(s')\right] = (TV)(s)$$

Value iteration algorithm:

$$V_{k+1} = TV_k$$

2. A tale of learning and control

A tale of learning and control

RL = learning to act in a **dynamic** environment which is **unknown**.

Let us look at **simpler problems**:

1) Assume the environment is known: **optimal control**

2) Assume the environment is reduced to one state: **online learning**

Optimal control *vs.* Reinforcement learning

$$\min_{u(.)} \int_0^T \ell(x(t), u(t))\mathrm{d}t$$
$$\dot{x}(t) = f(x(t), u(t)).$$

$$\max_{a_0,\ldots,a_{n-1}} \sum_{t=1}^{n} r_t$$
$$s_{t+1} \sim P(s_t, a_t)$$
$$r_{t+1} \sim R(s_t, a_t).$$

Online learning: the multi-armed bandit

→ Trade-off between **Exploration** *vs.* **exploitation**

Link with supervised learning

- Reinforcement learning deals with **large dimensional spaces** (*e.g.*, number of pixels of an image).
- All decisions are based on **observations**:
  - to learn a model of the environment and then control (model-based RL),
  - or to directly learn a policy (model-free RL).

➜ Like for supervised learning: use **function approximation**

- parametric: linear models, neural networks...
- non-parametric: *e.g.*, kernel methods

3. (Selected) Challenges for modern RL

Challenges for modern RL

Challenge 1: Scalability and computational burden

*“ OpenAI’s major game-mastery project for Dota 2 employed **10 real-time months (about 800 petaflop/days) of training time** to defeat world champion, human players.*

*[…] estimates fall in the ballpark of a **12 to 18 million USD cloud compute bill to train champion Alphastar and OpenAI Five**, respectively. “*

https://www.kdnuggets.com/2020/04/2-things-reinforcement-learning.html

Challenges for modern RL

Challenge 2: Theoretical guarantees

- Behavior of algorithms with **function approximation** (*e.g.*, Q-learning)

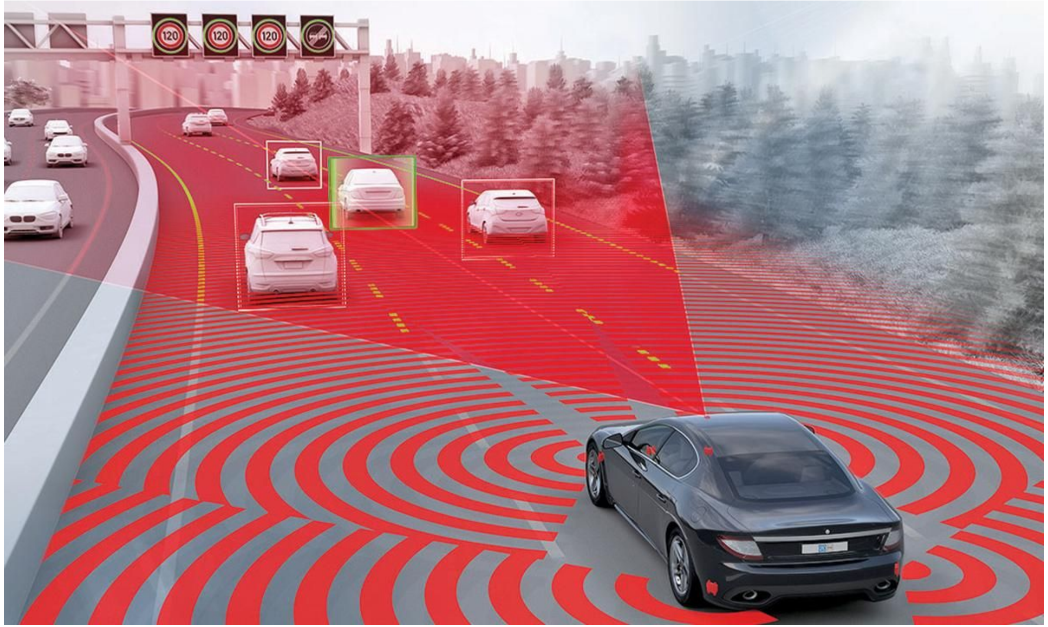
- What **performance measure** should be optimized?

  Fujimoto, Scott, et al. "Why Should I Trust You, Bellman? The Bellman Error is a Poor Replacement for Value Error." arXiv preprint arXiv:2201.12417 (2022).

- Sample **complexities**: which problems are intrinsically easy or hard?

Challenges for modern RL

Challenge 3: Safety for real-life critical systems



References

- Sutton, R. S., & Barto, A. G. (2018). *Reinforcement Learning: An Introduction*. MIT press.
- Liberzon, D. (2012). *Calculus of Variations and Optimal Control Theory – A Concise Introduction.* Princeton University Press.
- Bertsekas, D. P. (2019). *Reinforcement Learning and Optimal Control*. Athena Scientific.
- Meyn, S. (2022). *Control Systems and Reinforcement Learning*. Cambridge University Press