From Policy Gradient to Actor-Critic methods
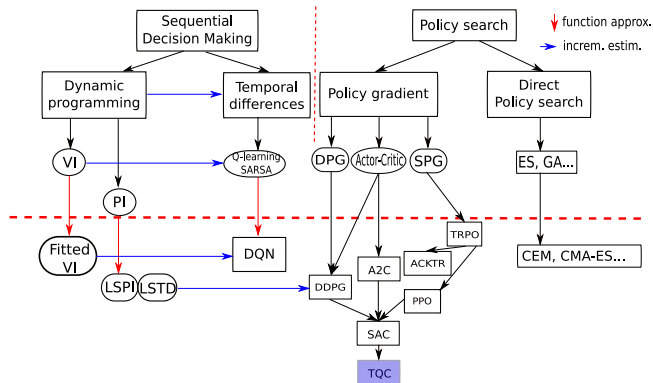## Introduction: the 4 routes to deep RL

Olivier Sigaud

Sorbonne Université
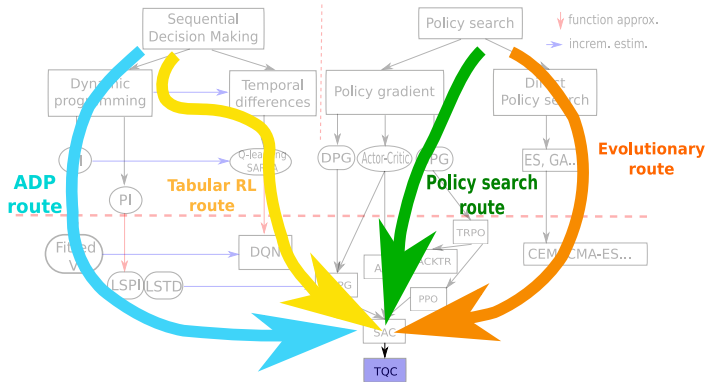http://people.isir.upmc.fr/sigaud
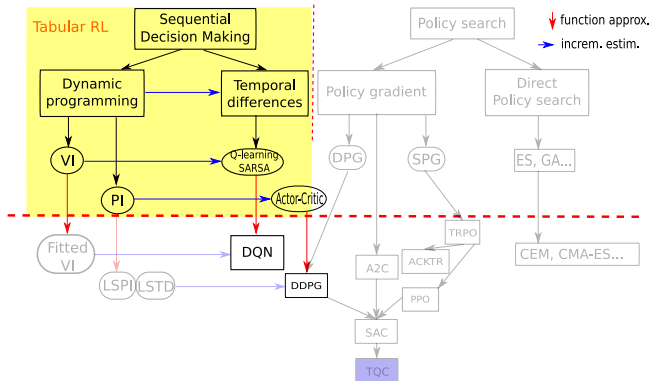
The Big Picture

▶ A very partial view of the whole RL literature
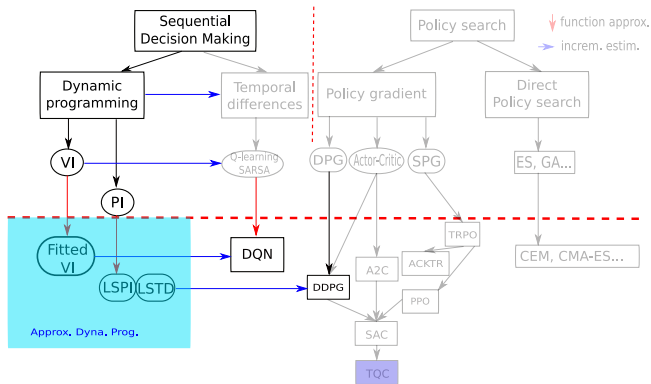
The four routes to deep RL

▶ Four different ways to come to Deep RL

The Tabular RL route

- ▶ The favorite route of beginners
- ▶ Start from Sutton&Barto, present Q-learning, SARSA and Actor-Critic
- ▶ Add function approximation with NNs, go to DQN, then DDPG

Sutton, R. S. & Barto, A. G. (1998) *Reinforcement Learning: An Introduction.* MIT Press.

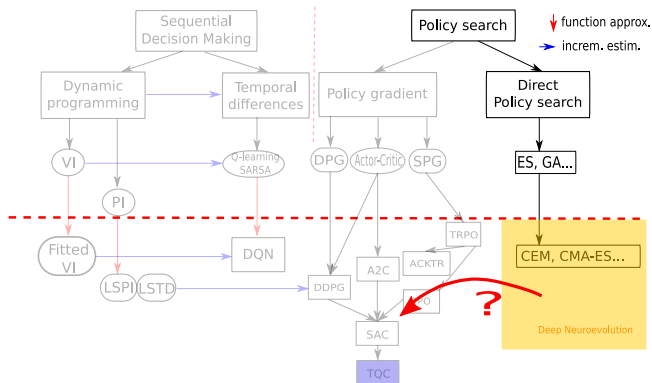# The Approximate Dynamic Programming route

- ▶ The favorite route of mathematicians
- ▶ I never travelled this route

Warren B. Powell. *Approximate Dynamic Programming: Solving the curses of dimensionality*, volume 703. John Wiley & Sons, 2007
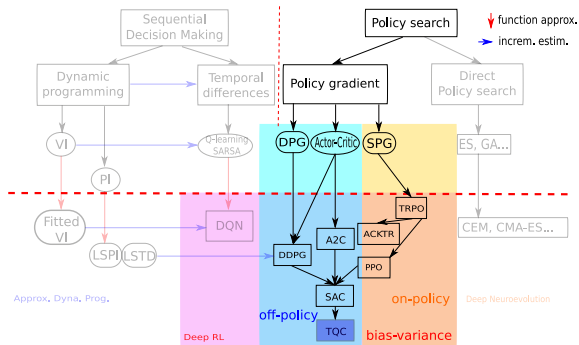
The Evolutionary route

- ▶ The favorite route of black-box optimisation people
- ▶ Much more efficient than RL people think

Tim Salimans, Jonathan Ho, Xi Chen, and Ilya Sutskever. Evolution strategies as a scalable alternative to reinforcement learning.
arXiv preprint arXiv:1703.03864, 2017

The Policy Search route

- ▶ The favorite route of roboticists
- ▶ The one I'm travelling in these lessons
- ▶ Central question: difference between PG with baseline and Actor-Critic

Marc P. Deisenroth, Gerhard Neumann, Jan Peters, et al. A survey on policy search for robotics. *Foundations and Trends® in Robotics*, 2(1–2):1–142, 2013

Outline

1. The policy search problem
2. Policy Gradient derivation (3 parts)
3. From policy gradient with baseline to actor-critic
4. Bias-variance trade-off
5. On-policy vs off-policy
6. TRPO, ACKTR
7. PPO
8. DDPG, TD3
9. SAC
10. RWR
11. Wrap-up

Any question?

Send mail to: `Olivier.Sigaud@upmc.fr`

Marc Peter Deisenroth, Gerhard Neumann, Jan Peters, et al.

A survey on policy search for robotics.
Foundations and Trends® in Robotics, 2(1–2):1–142, 2013.

Warren B. Powell.

Approximate Dynamic Programming: Solving the curses of dimensionality, volume 703.
John Wiley & Sons, 2007.

Tim Salimans, Jonathan Ho, Xi Chen, and Ilya Sutskever.

Evolution strategies as a scalable alternative to reinforcement learning.
arXiv preprint arXiv:1703.03864, 2017.

Richard S. Sutton and Andrew G. Barto.

Reinforcement Learning: An Introduction.
MIT Press, 1998.