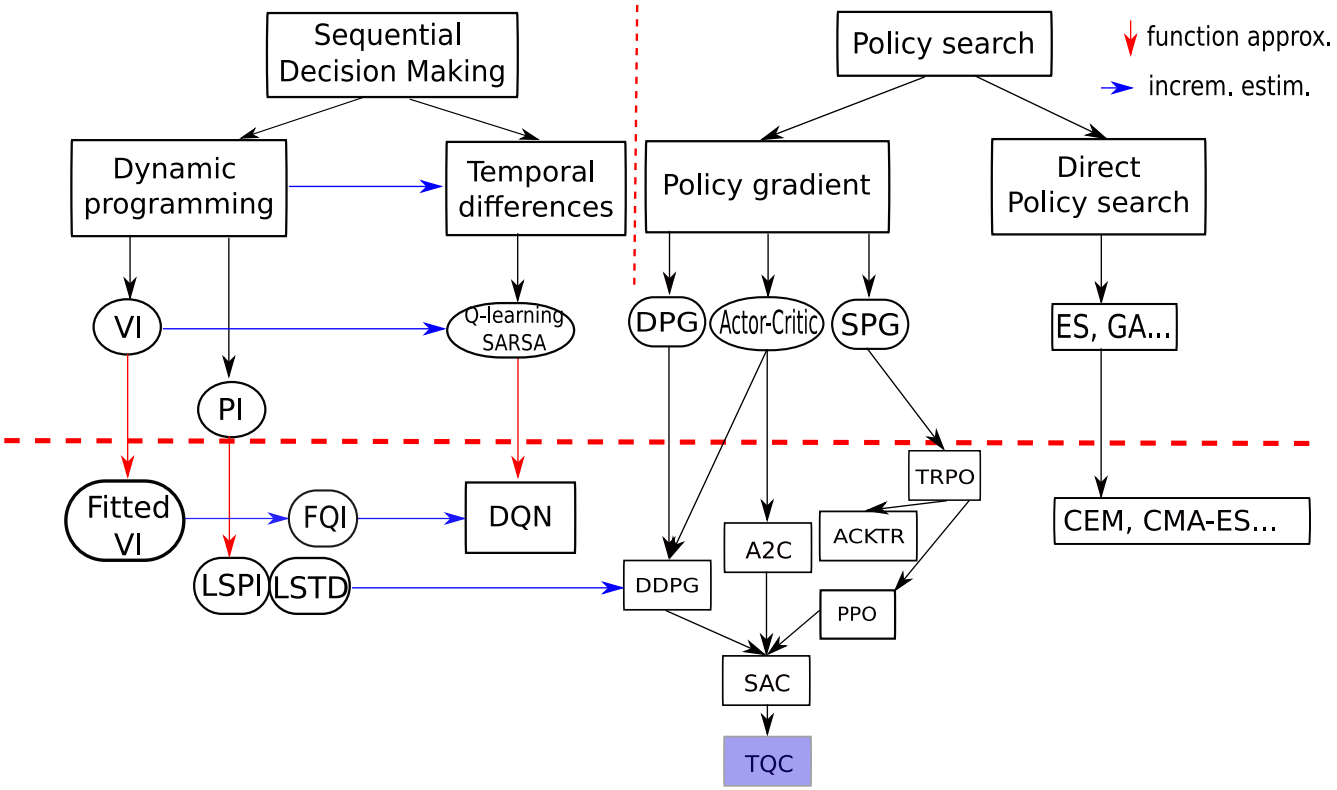
Teaching Reinforcement Learning

The 4 routes to deep RL

Olivier Sigaud

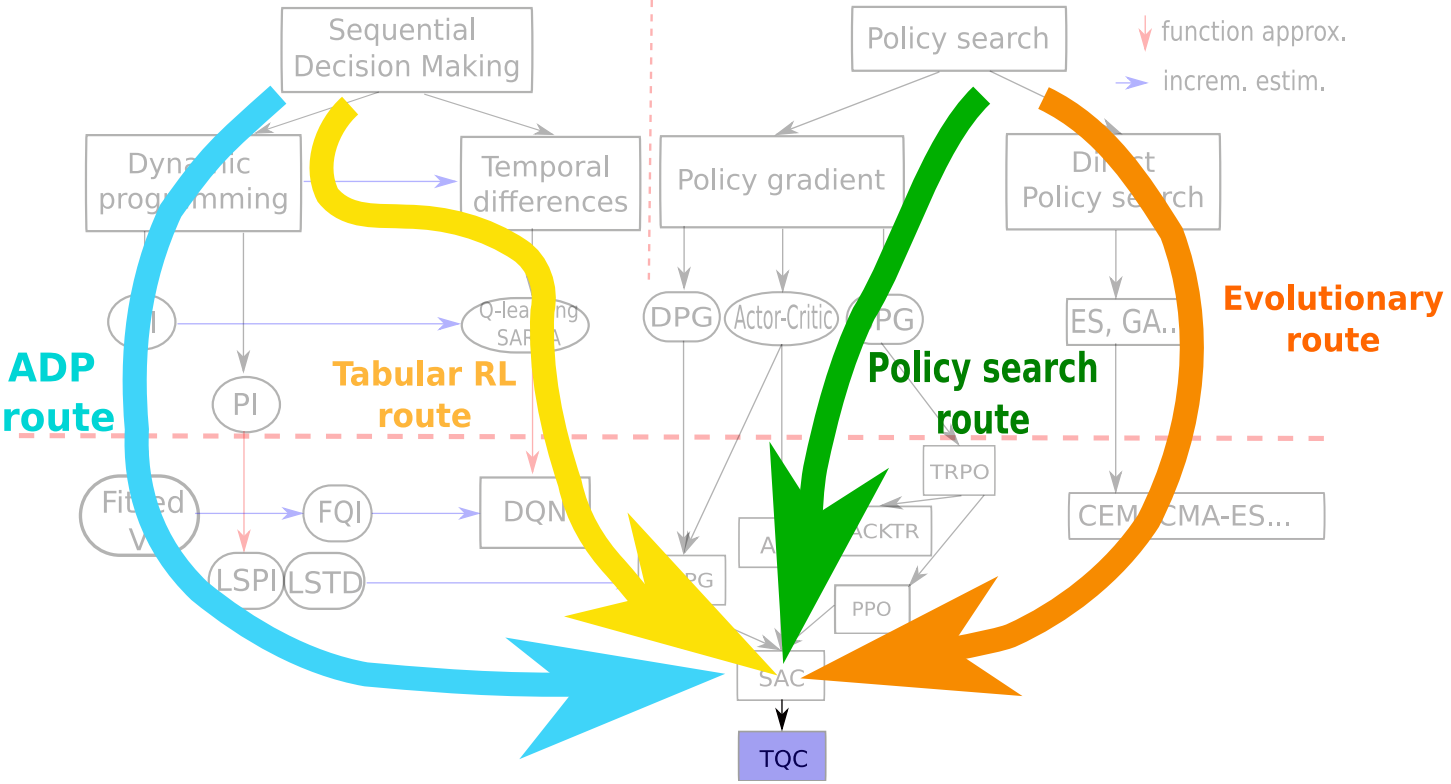
Sorbonne Université
http://people.isir.upmc.fr/sigaud

The Big Picture



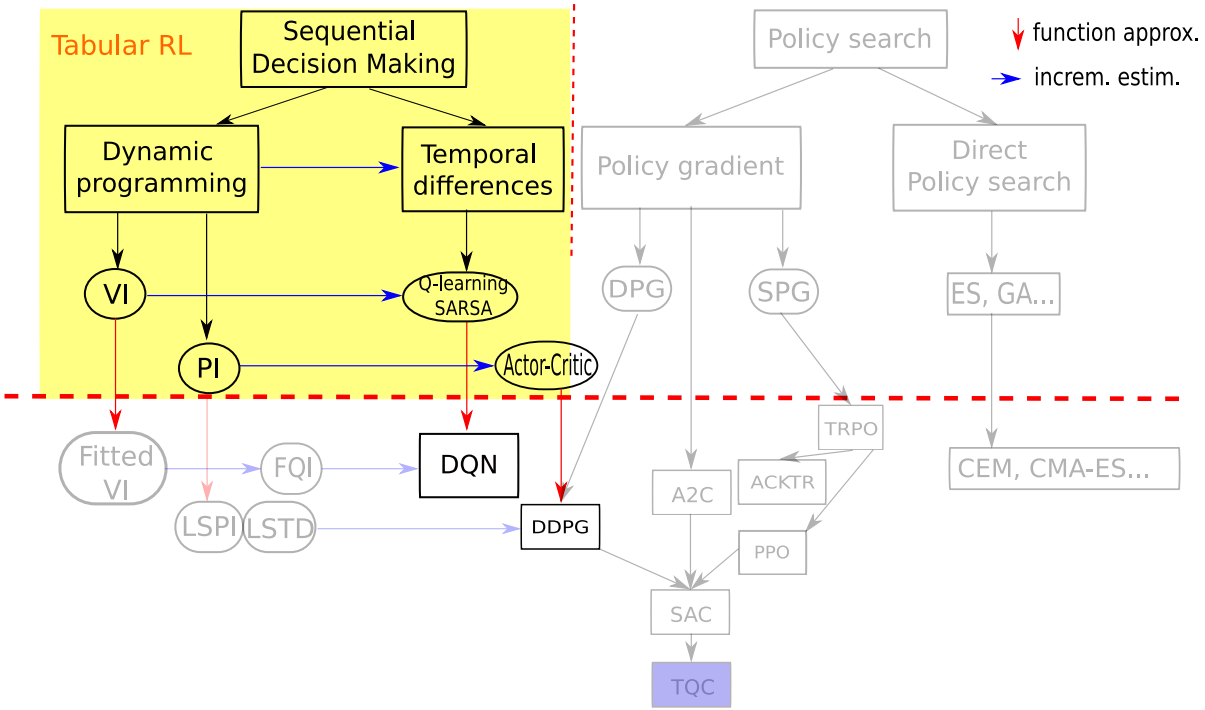
- A very partial view of the whole RL literature

The four routes to deep RL



- Four different ways to come to Deep RL

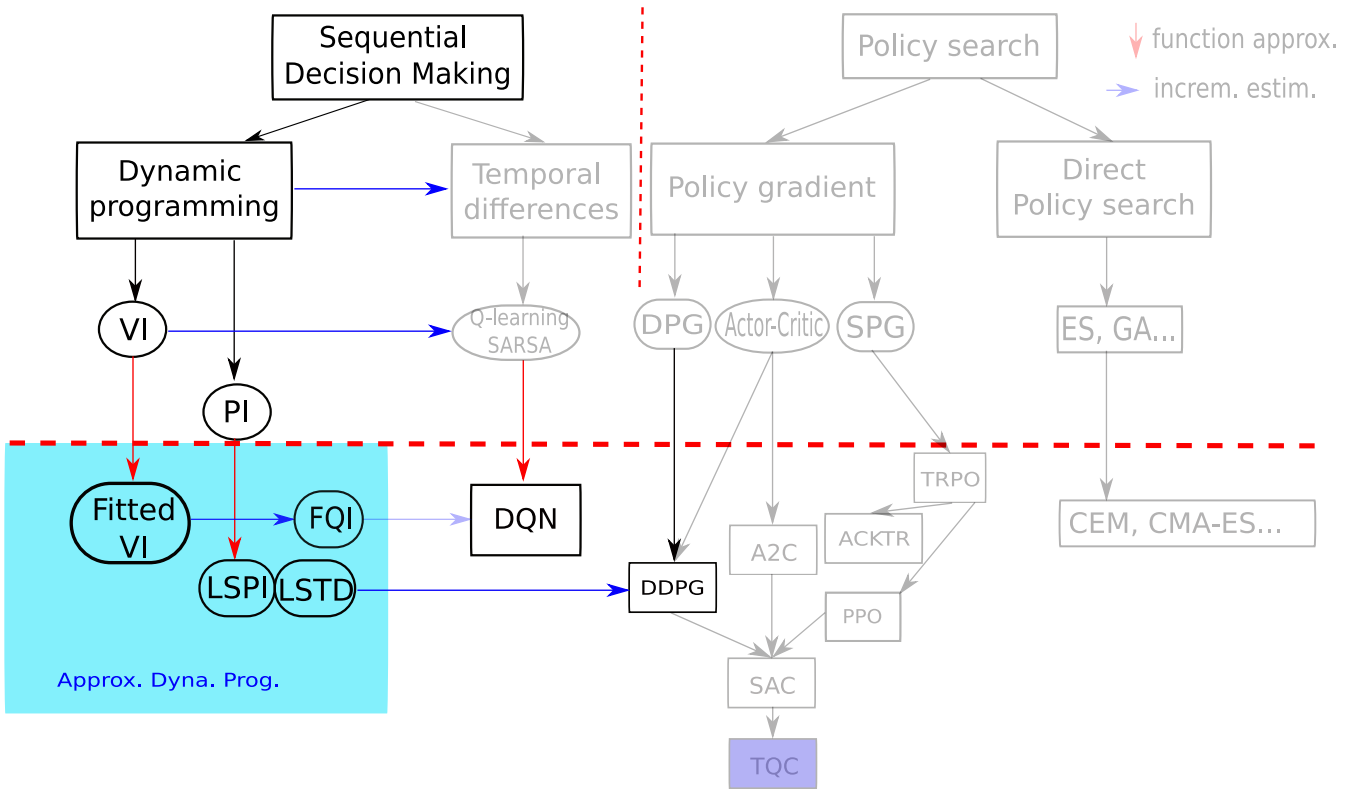
The Tabular RL route



- 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
- The route of the first two days

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

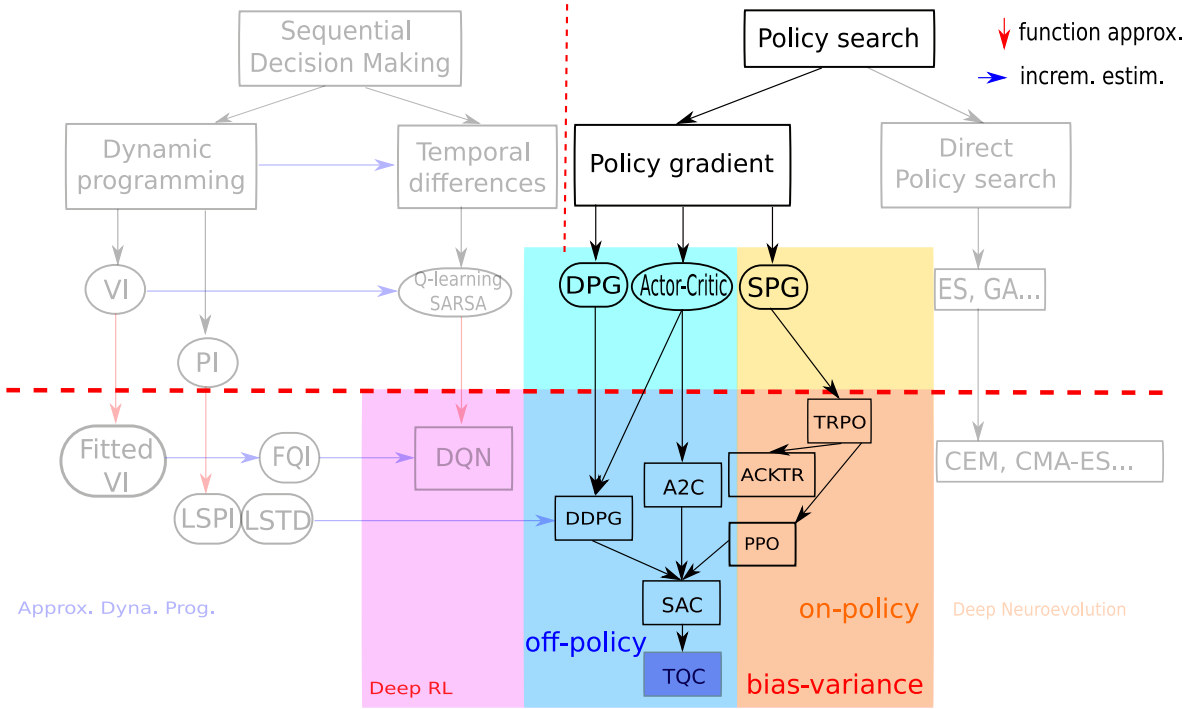
The Approximate Dynamic Programming route



- 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

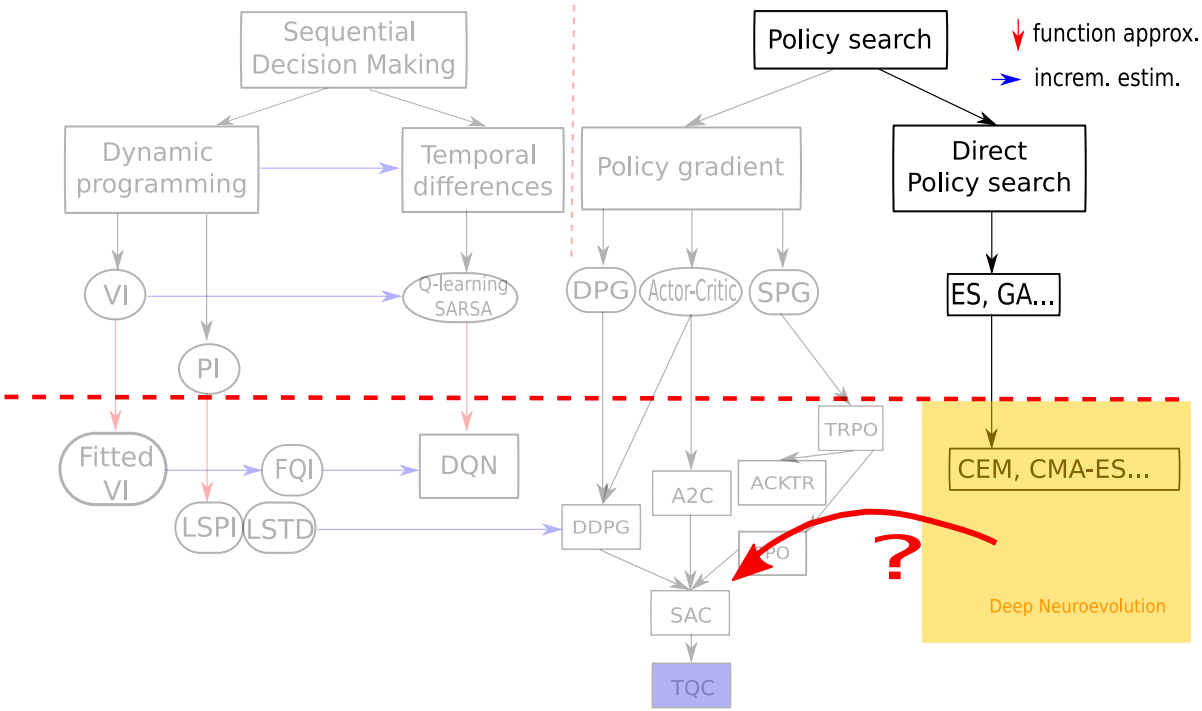
The Policy Search route



- The favorite route of roboticists
- The route of the third day

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

The Evolutionary route



- The favorite route of black-box optimisation people
- Much more efficient than RL people think
- Proposed for the last day

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

Organization

- Morning: classes about algorithms and concepts
- After lesson: rehearsal questions about the concepts
- Afternoon: coding, practicing, tuning hyper-parameters
- Practicing based on the BBRL library
- Some mini-projects to train on your own
- Lessons are available in video
- Everything available in advance, supplementary material...
- Look at the Moodle page for more details

Evaluation

- Answer to lessons related questions: 1 point (0 - 0.5 - 1)
- 8 mini-projects: 16 points (2 per project, 0 - 0.5 - 1 - 1.5 - 2)
- Submit your best policies on LunarLander-v2 and LunarLanderContinous-v2: 3 points per policy (based on ranking)
- Bonus: submit an always landing policy on RocketLander-v0: 3 points
- All mini-projects and policies are to be given by december 15, midnight
- Delay is penalized (1 point per day)
- Technical details on the Moodle page
- Advices
  - Listen to lessons in advance
  - Do mini-projects in advance or each week

Any question?

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