Chapter 10:
Dimensions of Reinforcement Learning

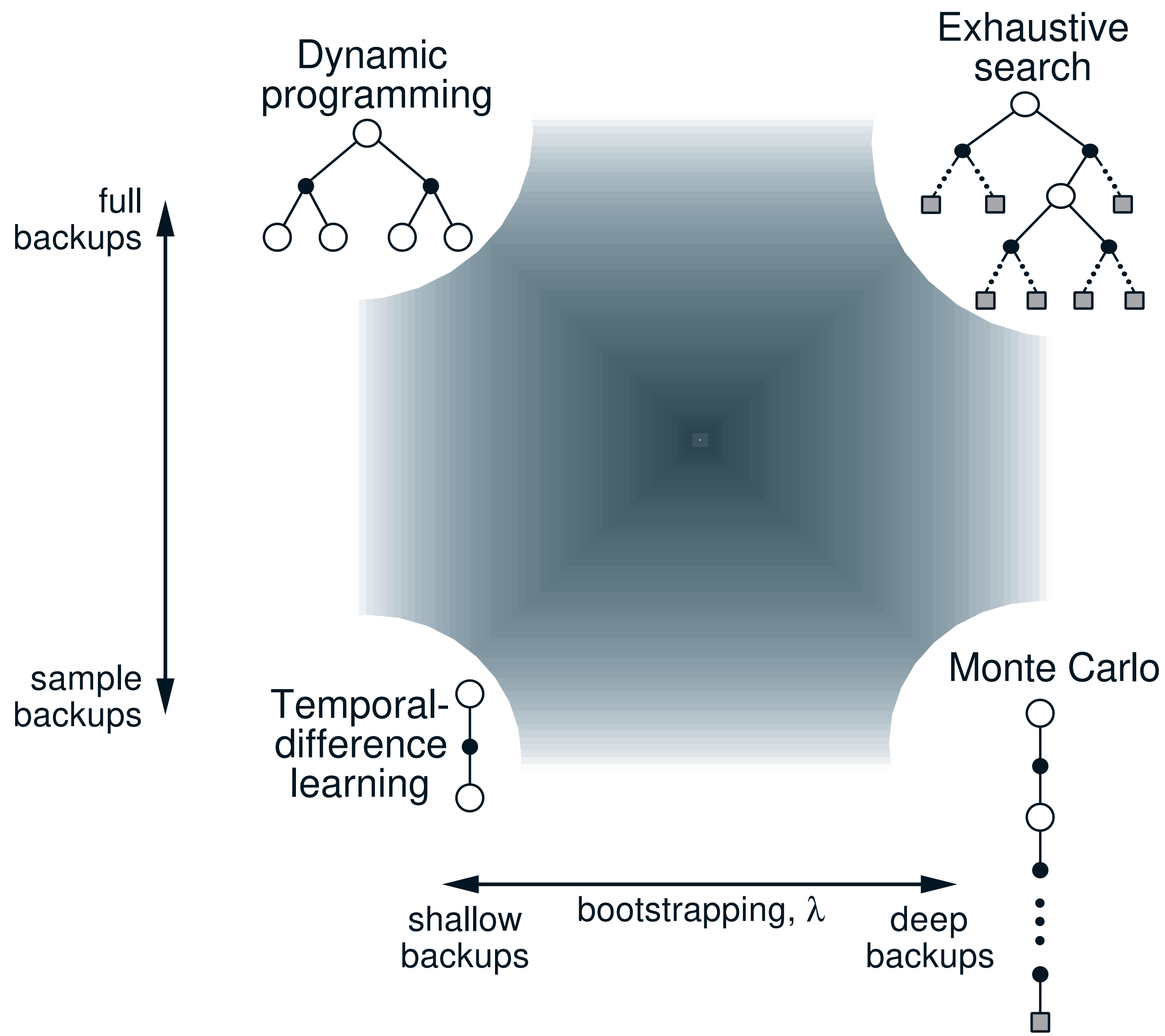
Objectives of this chapter:

- Review the treatment of RL taken in this course
- What have left out?
- What are the hot research areas?

Three Common Ideas

- Estimation of value functions
- Backing up values along real or simulated trajectories
- Generalized Policy Iteration: maintain an approximate optimal value function and approximate optimal policy, use each to improve the other

Backup Dimensions



Other Dimensions

- Function approximation
  - tables
  - aggregation
  - other linear methods
  - many nonlinear methods
- On-policy/Off-policy
  - On-policy: learn the value function of the policy being followed
  - Off-policy: try learn the value function for the best policy, irrespective of what policy is being followed

Still More Dimensions

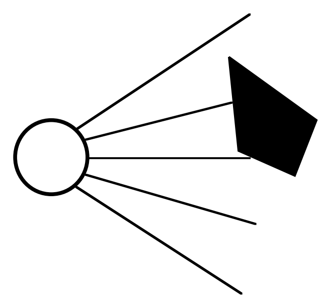
- Definition of return: episodic, continuing, discounted, etc.
- Action values vs. state values vs. afterstate values
- Action selection/exploration: ε-greed, softmax, more sophisticated methods
- Synchronous vs. asynchronous
- Replacing vs. accumulating traces
- Real vs. simulated experience
- Location of backups (search control)
- Timing of backups: part of selecting actions or only afterward?
- Memory for backups: how long should backed up values be retained?

Frontier Dimensions

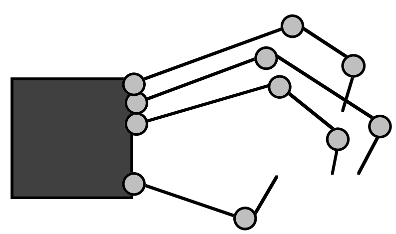
- Prove convergence for bootstrapping control methods.
- Trajectory sampling
- Non-Markov case:
  - Partially Observable MDPs (POMDPs)
    - Bayesian approach: belief states
    - construct state from sequence of observations
  - Try to do the best you can with non-Markov states
- Modularity and hierarchies
  - Learning and planning at several different levels
    - Theory of options

More Frontier Dimensions

- Using more structure
  - factored state spaces: dynamic Bayes nets



  - factored action spaces



Still More Frontier Dimensions

- Incorporating prior knowledge
  - advice and hints
  - trainers and teachers
  - shaping
  - Lyapunov functions
  - etc.