ASEN 6519, Spring 2017

Algorithms for Aerospace Autonomy

Lecture 24:
Approximate POMDPS

Luke Burks
04/20/2017

University of Colorado **Boulder**
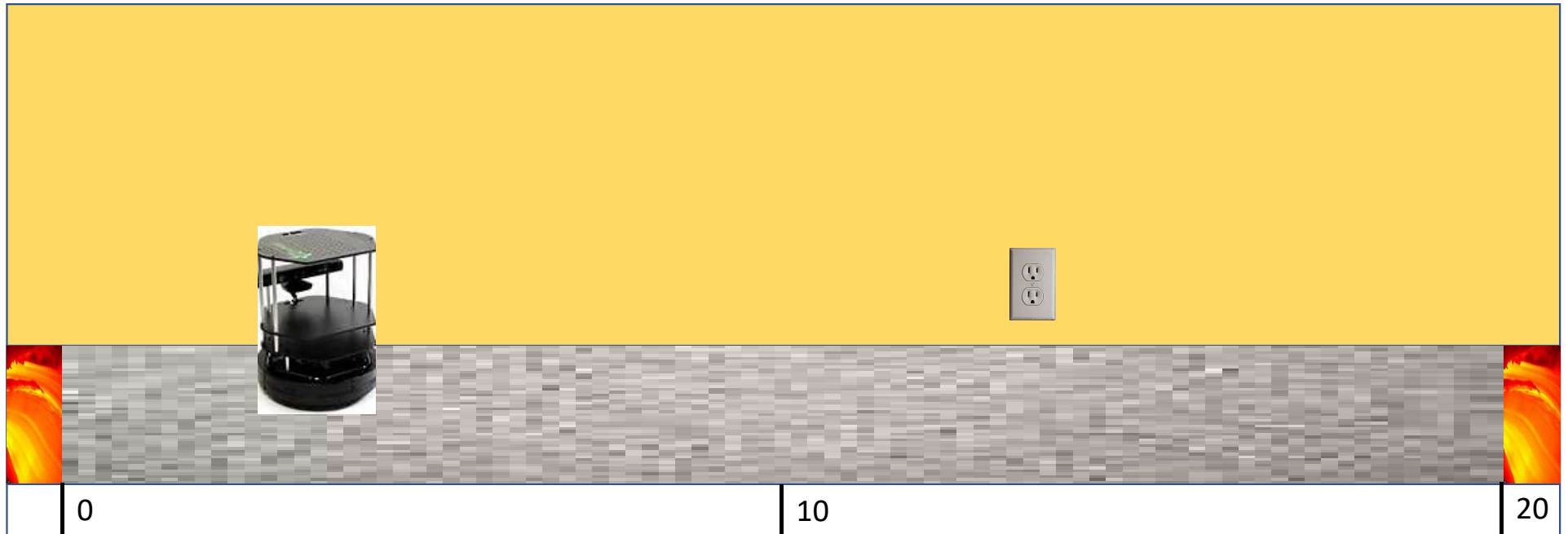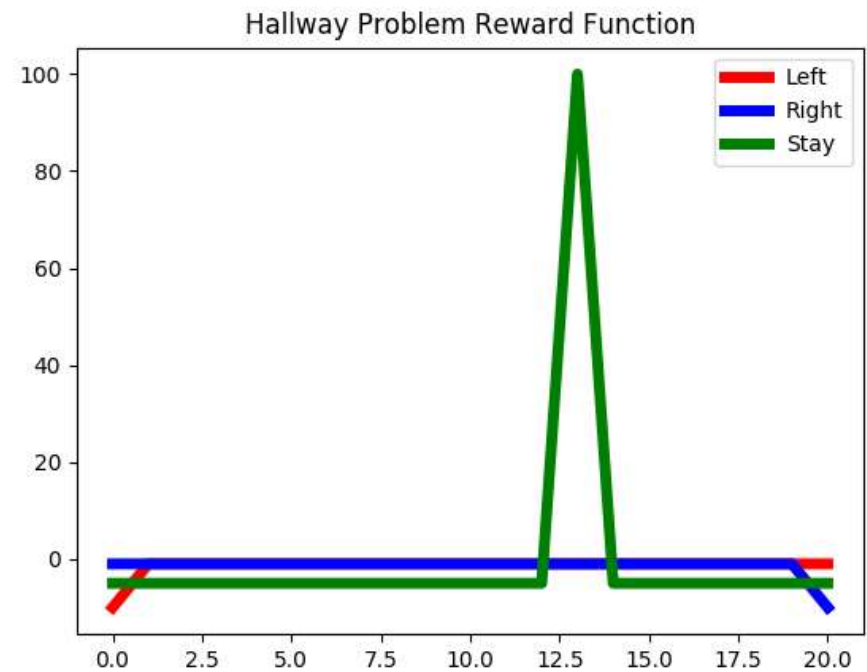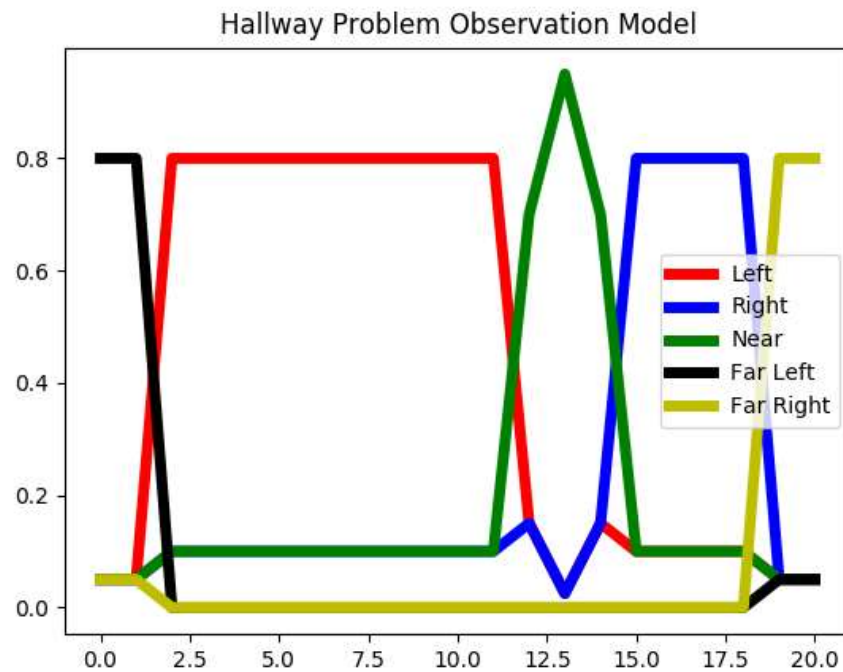
MDP: Modified Europa Hopper Problem
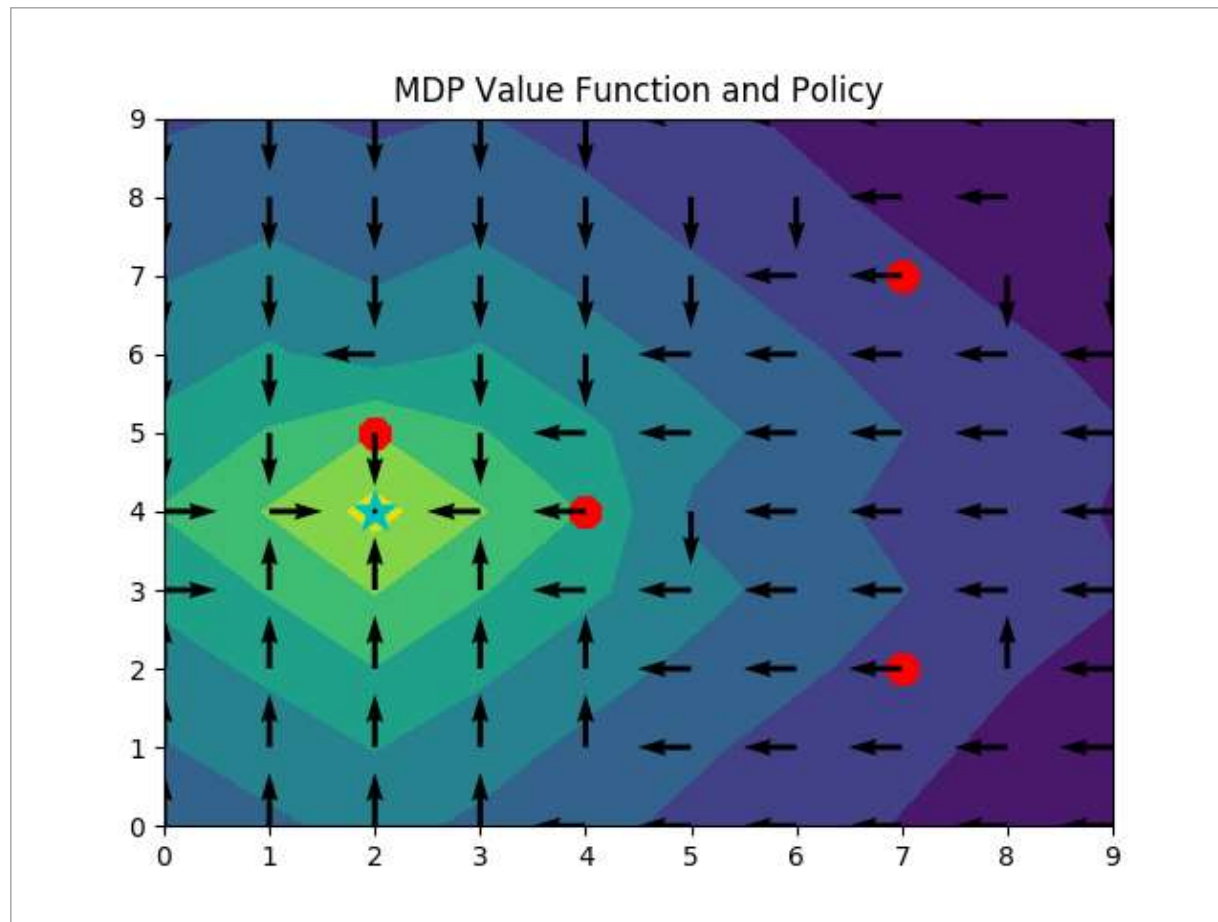
MDP: The simpler hallway problem

- A robot is in a hallway trying to find the power outlet
- Also, there's lava at each end of the hallway…
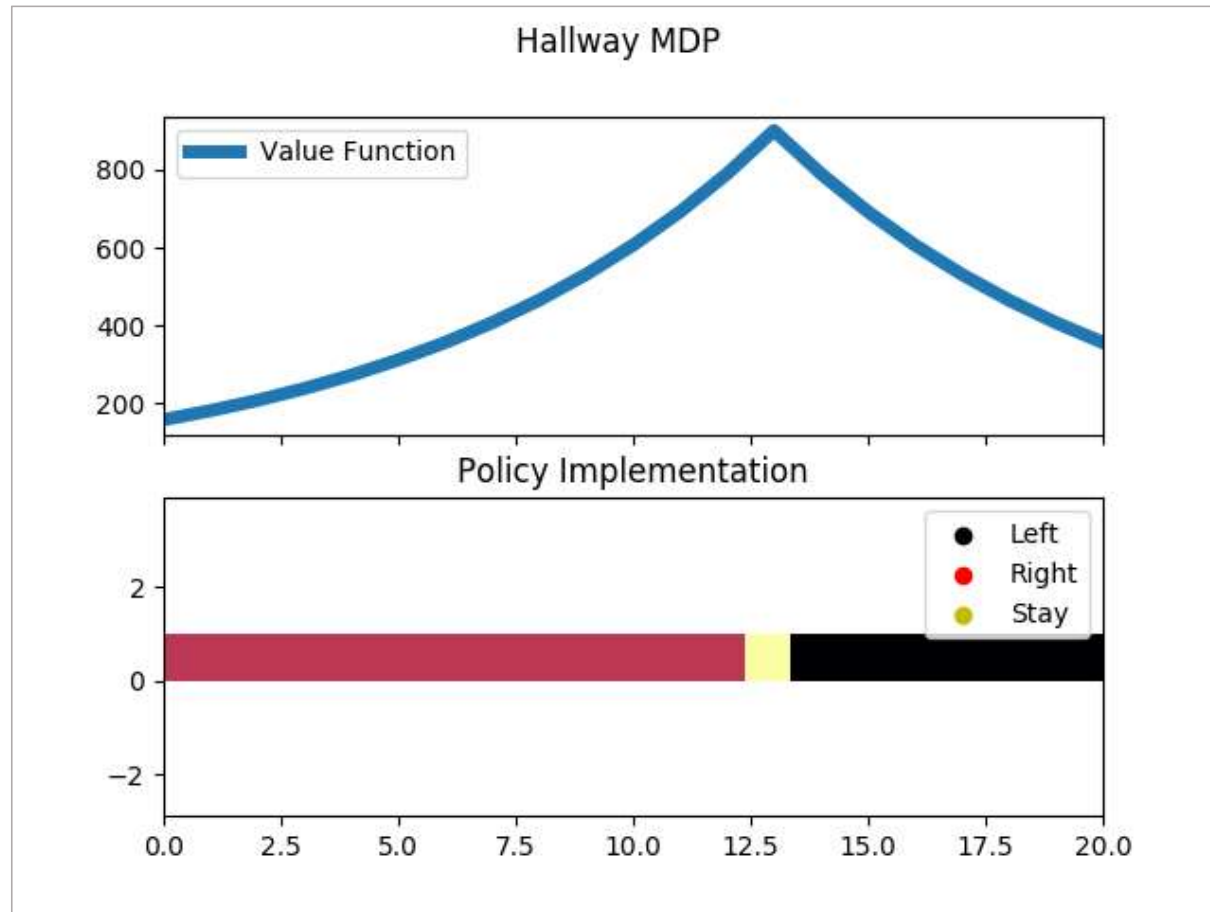

0   10   20

MDP: The simpler hallway problem

- The hallway is discretized into 21 spaces
- The robot can take 3 actions, and receive 5 observations

MDP: The simpler hallway problem

Why this is hard with a POMDP

- In the (super weird) baby problem from the book there were two states, so we could visualize the alpha vectors easily enough.

- Here the value function is defined by a 21 dimensional piece-wise set of alpha vectors...

- That awful set of calculations we did on Thursday, the additional actions and observations lead to far more of those
  - 2 step look ahead: 729 alphas
  - 4 step look ahead: 10^{74} alphas
  - 5 step look ahead: 10^{372} alphas
  - 20 step look ahead: $10^{10^{13}}$ alphas

QMDP: What if we pretend it's actually an MDP?

- QMDP [Littman et. Al 1995]
- Calculates the MDP-Value of taking each possible action

$$Q(s_i, a) = r(s_i, a) + \sum_{j}^{N} V_{MDP}(s_j)p(s_j|s_i, a)$$

- The policy for a belief is the action that maximizes the expectation of that value

$$\pi(b) = argmax_a \sum_{i}^{N} b(s_i)Q(s_i, a)$$

- Is exact for cases where the state becomes fully observed after the action
- Can be shown to be an upper bound on the value function
- Preforms poorly at pure information gathering actions

QMDP Simulation Example
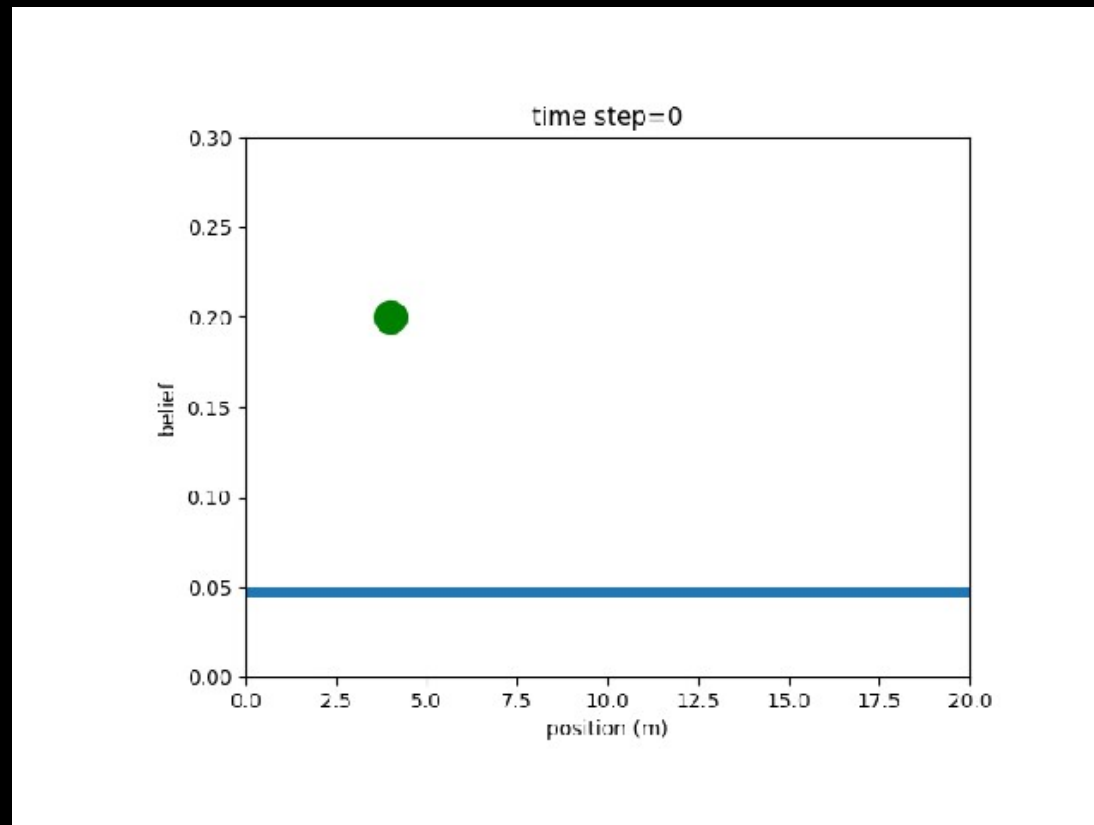
FIB: What if we only kind of pretend its actually an MDP?

- Fast Informed Bound [Hauskrecht 2000]

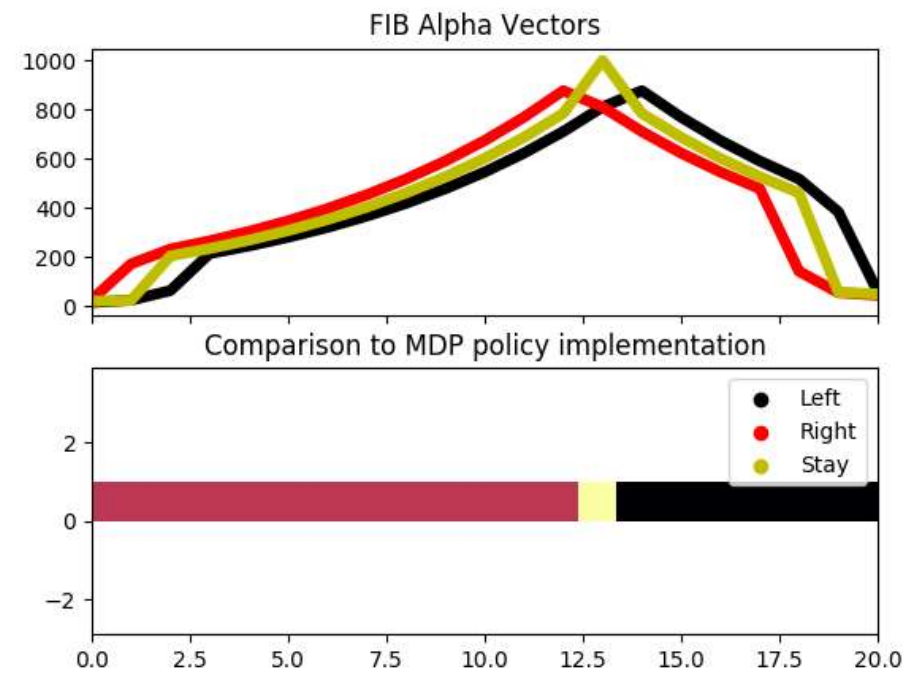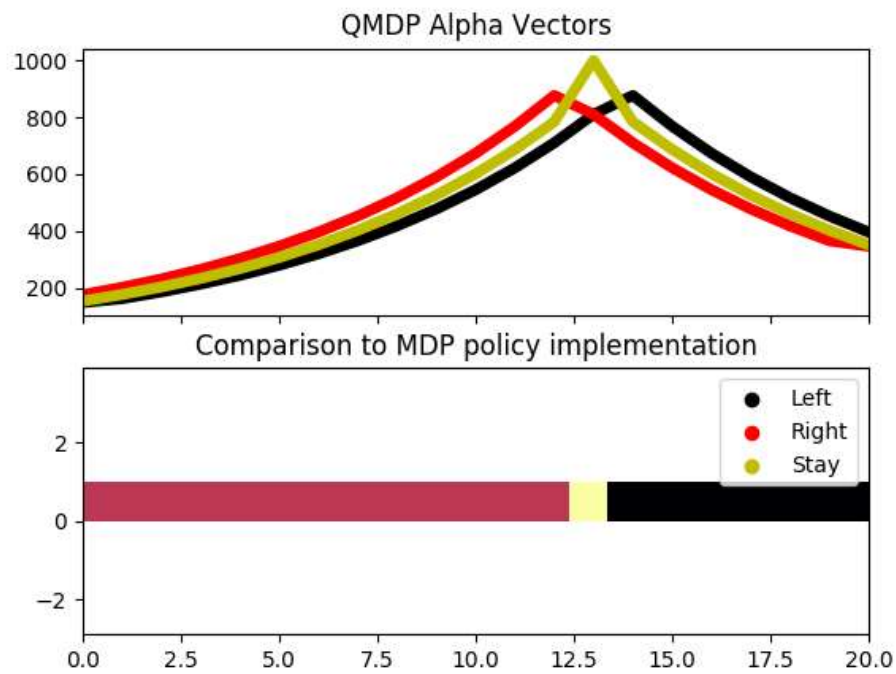- Takes all possible observations into account when calculating the value of each action

$$Q(s_i, a) = r(s_i, a) + \sum_{j}^{N} \sum_{o}^{|O|} p(o|s_j) \, p(s_j|s_i, a) V_{MDP}(s_j)$$

- The value estimate is upper bounded by QMDP and lower bounded by the true value

- Takes longer than QMDP, especially for problems with lots of observations
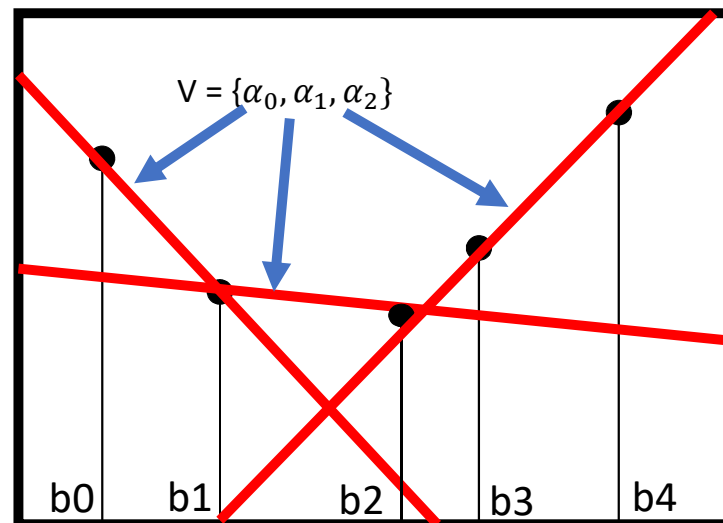
MDP: Just a reminder
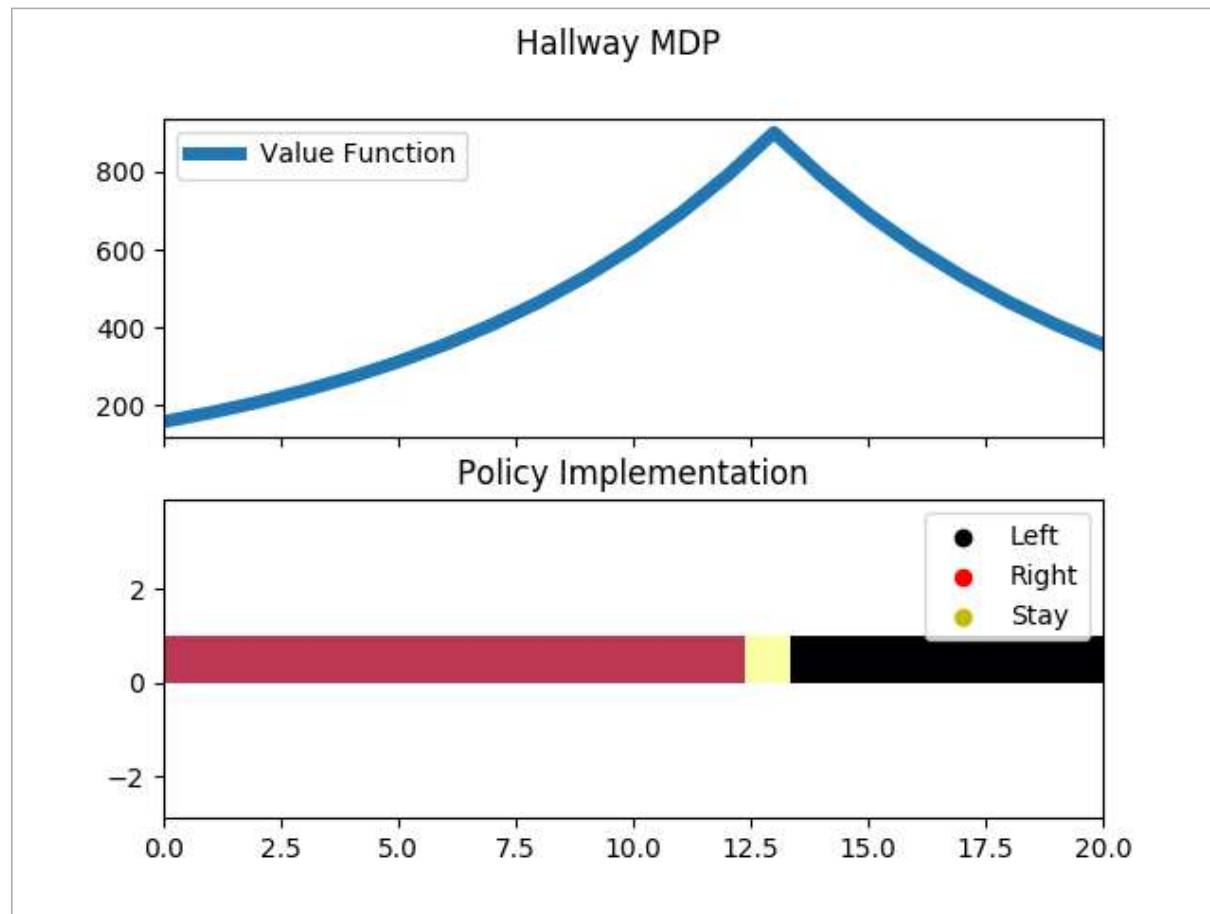
QMDP vs. FIB

PBVI: Do we have to care about ALL the beliefs?

- Point-Based Value Iteration [Pineau et. al 2003]

- Given some starting belief, there are many beliefs you'll never encounter no matter the sequence of actions and observations

- Use what you know about the beliefs you've seen to tell you what to do at new beliefs

$V = \{\alpha_0, \alpha_1, \alpha_2\}$

b0   b1   b2   b3   b4

# PBVI: Do we have to care about ALL the beliefs?
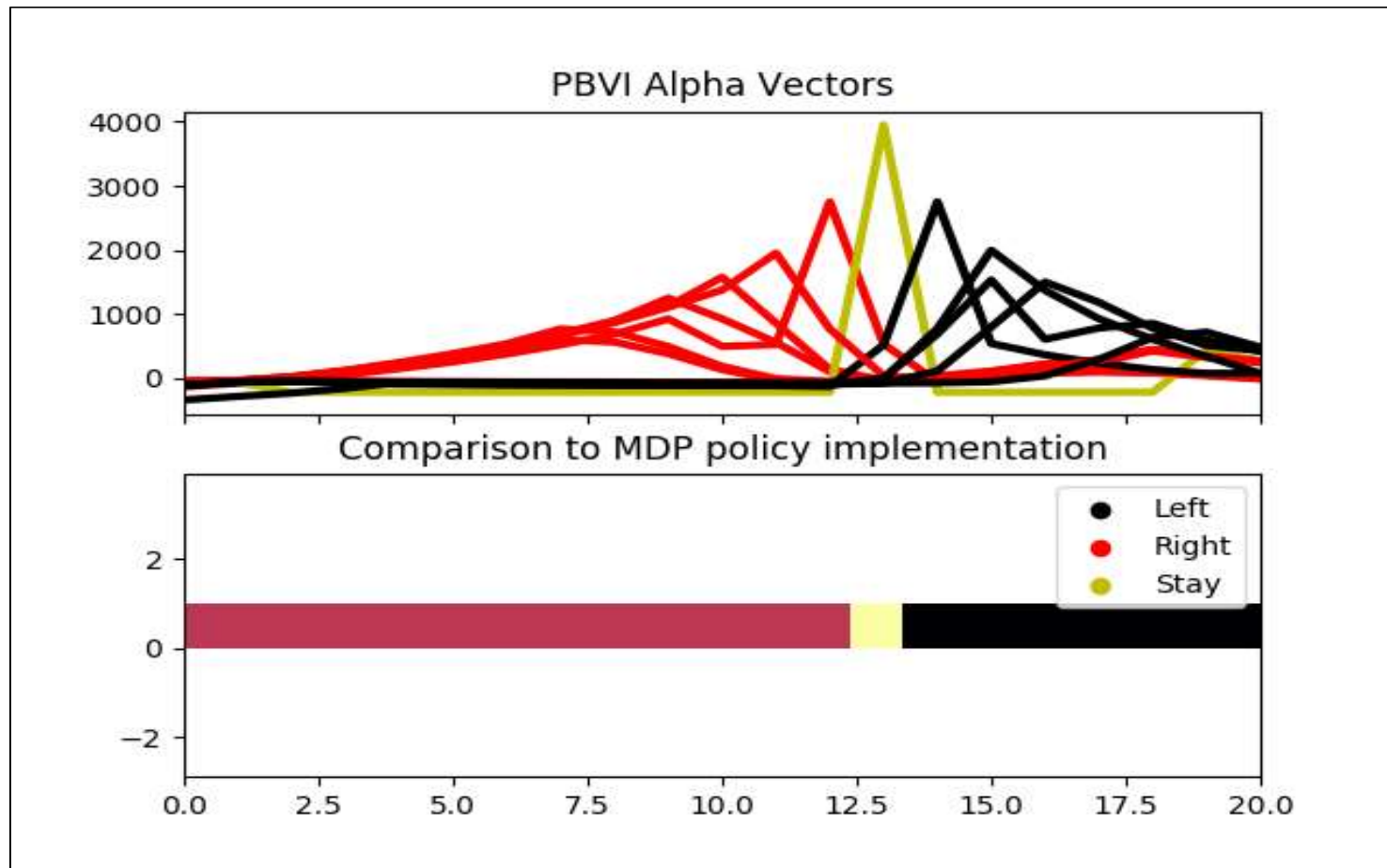
MC-POMDP: Because putting Monte Carlo in front of anything makes it possible right?

- Monte Carlo POMDP [Thrun 2003]
- Represent the belief by a set of (x) particles

$$b(s) \rightarrow \mathcal{X}, \; x \in \mathcal{X}$$

- Push those particles through your transition function
- Use importance resampling to kick out particles that end up in unlikely places
- Each particle set you find has a value calculated for it
- Can use nearest neighbor interpolation between particle sets to find the action for a new set
- Works in continuous spaces

MC-POMDP: Because putting Monte Carlo in front of anything makes it possible right?
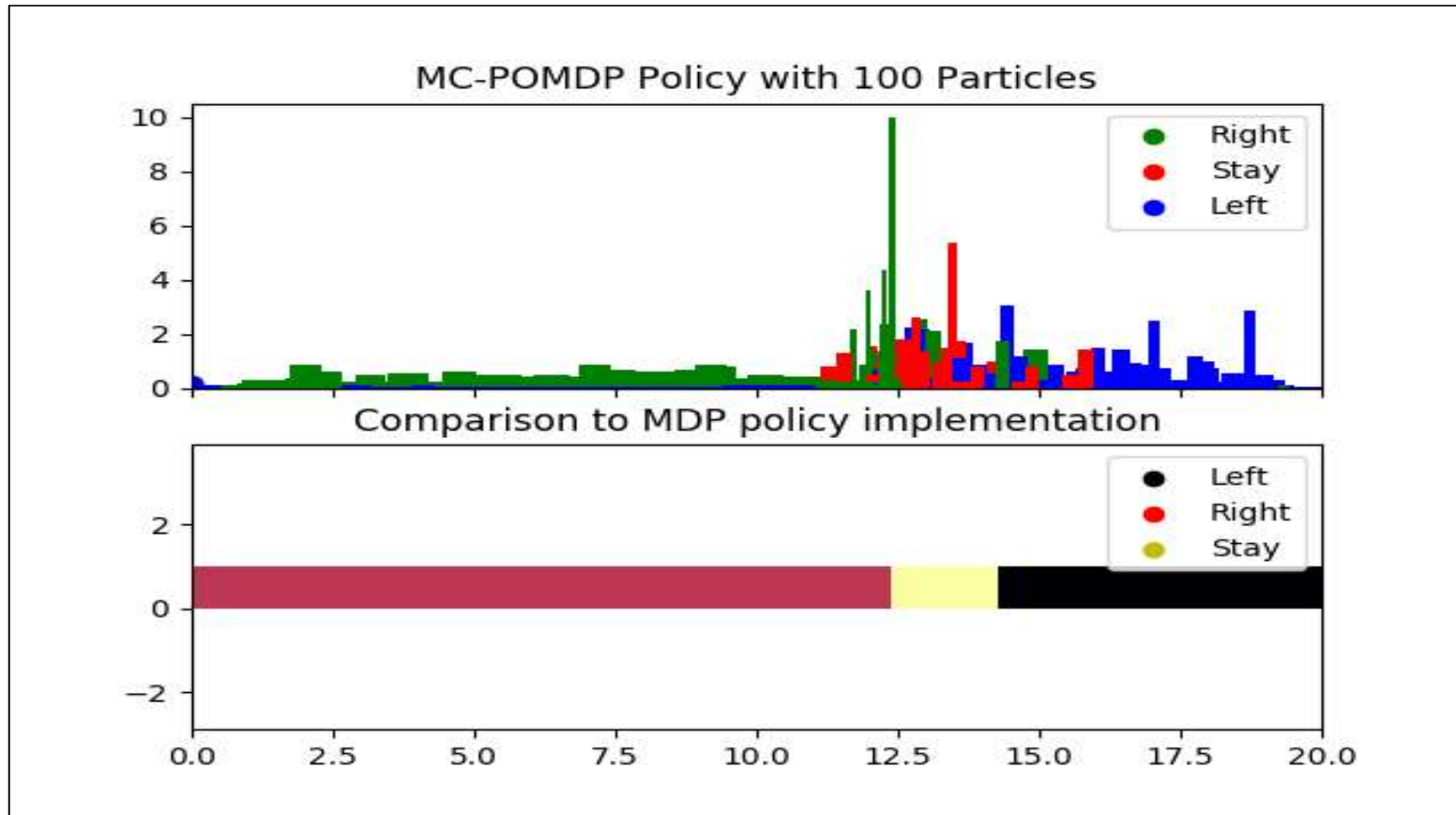
POMCP: When your problem requires a state for every atom in the solar system…

- Partially Observable Monte Carlo Planning [Silver 2010]
- Online POMDP Solver
- Plans forward from the current belief state
  - Guarantees reachable beliefs are explored
- Cares about histories, not beliefs
  - Histories are sequences of past actions and observations
  - Given a prior belief and a history, you can always construct the posterior belief
- For history h_t using K particles,

$$\hat{X}(s, h_t) = \sum_i^K \delta_{s x_t^i} \text{ , where } \delta_{ss'} \text{ is the kronecker delta}$$

POMCP: When your problem requires a state for every atom in the solar system…

- Builds a tree of possible future histories
- When an action is taken and an observation received, the root is moved down the corresponding branches while other branches are pruned
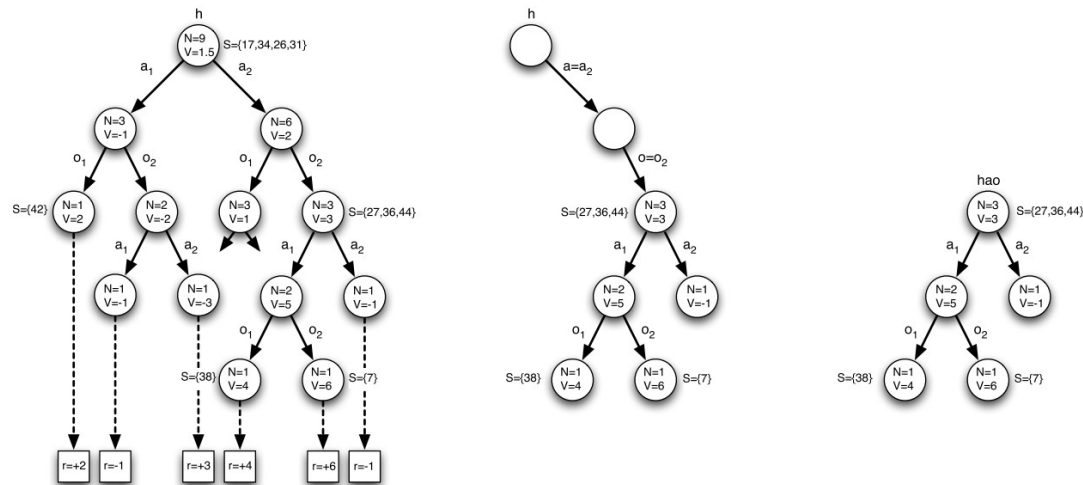
POMCP: In action

- 10^56 States

- 1024 Obs

- 4 Actions