Recap

- Alpha Vectors
- Best solver for discrete POMDPs:

SARSOP

POMDP Computational Complexity

POMDP Computational Complexity

Sad facts ● 

- Infinite horizon POMDPs are *undecidable*

POMDP Computational Complexity

Sad facts

- Infinite horizon POMDPs are *undecidable*
- Finite horizon POMDPs are *PSPACE Complete*

POMDP Computational Complexity

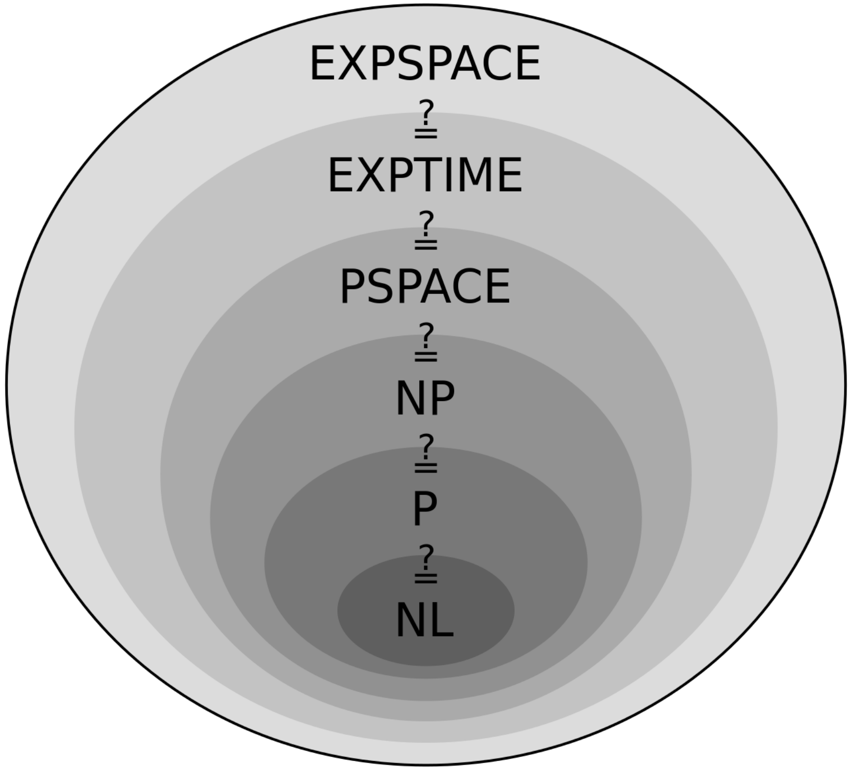
Sad facts ●

- Infinite horizon POMDPs are *undecidable*
- Finite horizon POMDPs are *PSPACE Complete*
  - Among the hardest problems that can be solved using a polynomial amount of space

POMDP Computational Complexity

Sad facts ●

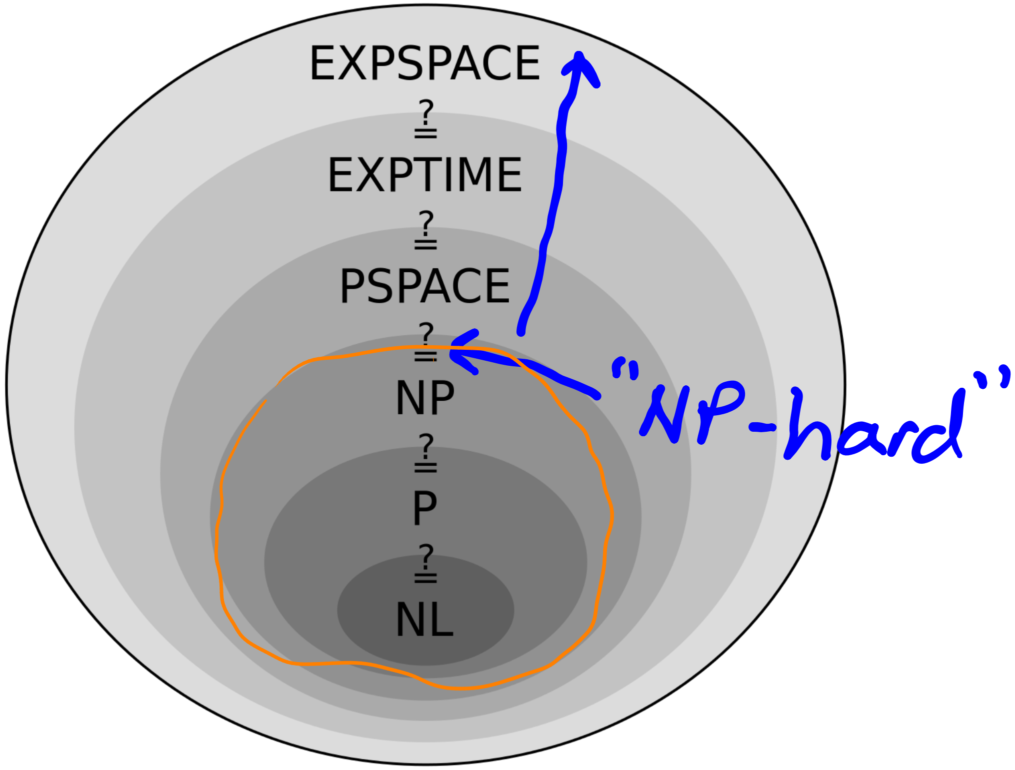
- Infinite horizon POMDPs are *undecidable*
- Finite horizon POMDPs are *PSPACE Complete*
  - Among the hardest problems that can be solved using a polynomial amount of space



POMDP Computational Complexity

Sad facts ●

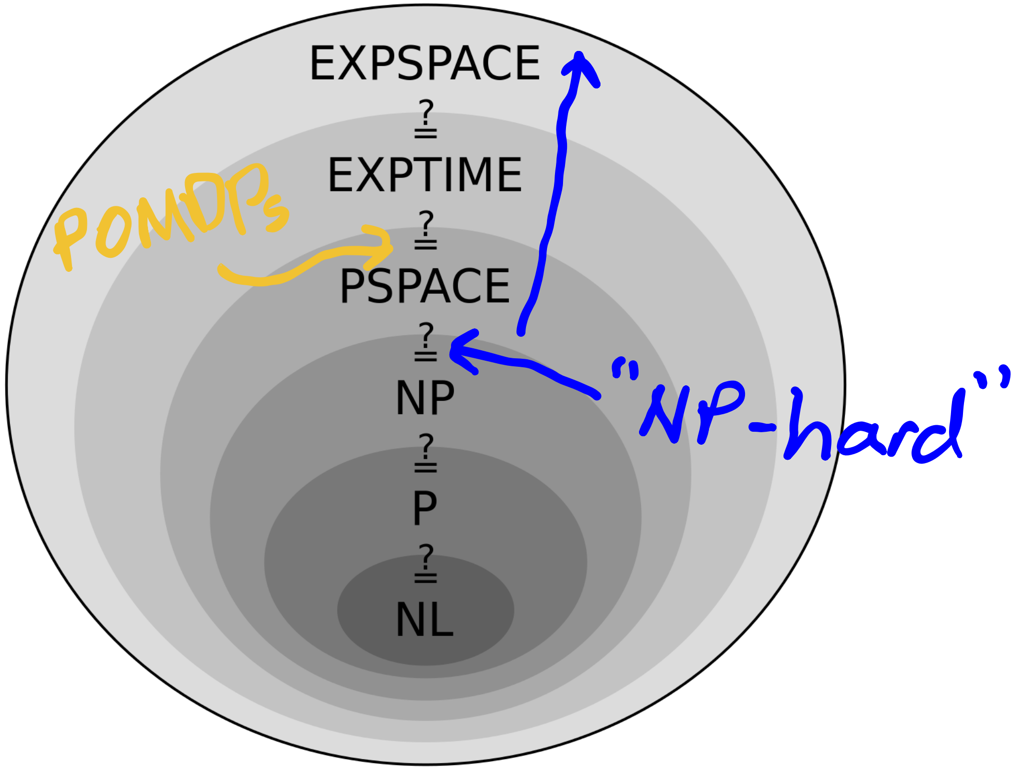
- Infinite horizon POMDPs are *undecidable*
- Finite horizon POMDPs are *PSPACE Complete*
  - Among the hardest problems that can be solved using a polynomial amount of space



POMDP Computational Complexity

Sad facts ●

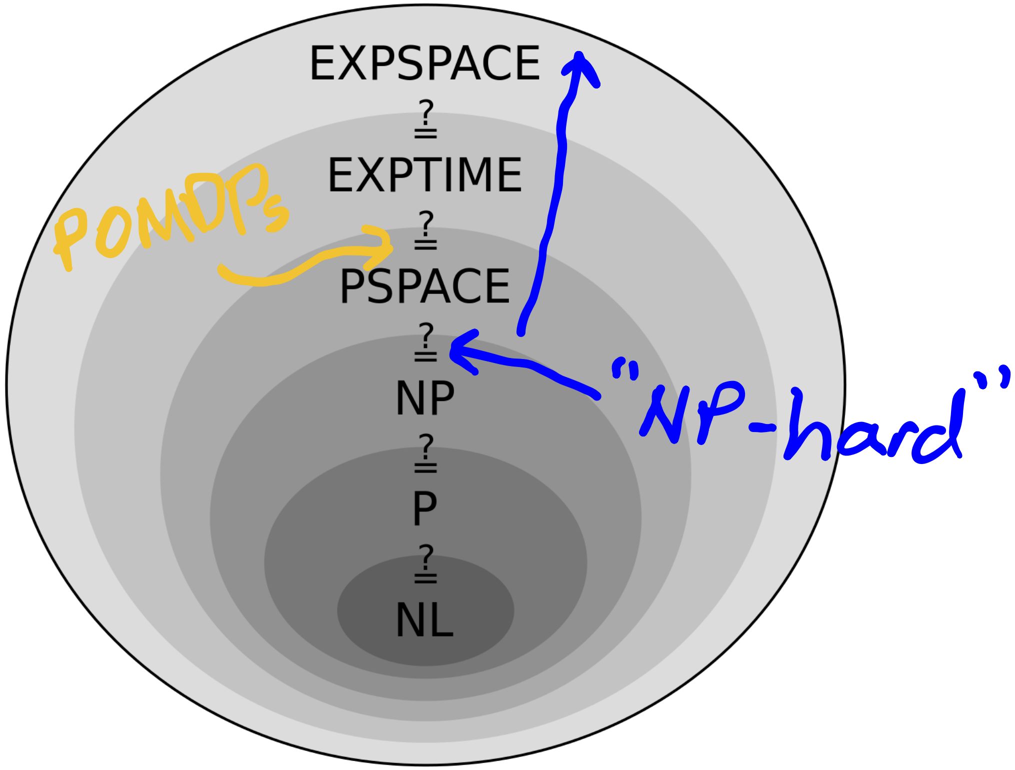
- Infinite horizon POMDPs are *undecidable*
- Finite horizon POMDPs are *PSPACE Complete*
  - Among the hardest problems that can be solved using a polynomial amount of space



POMDP Computational Complexity

Sad facts ●

- Infinite horizon POMDPs are *undecidable*
- Finite horizon POMDPs are *PSPACE Complete*
  - Among the hardest problems that can be solved using a polynomial amount of space
  - Any algorithm that can solve a general POMDP will have exponential complexity

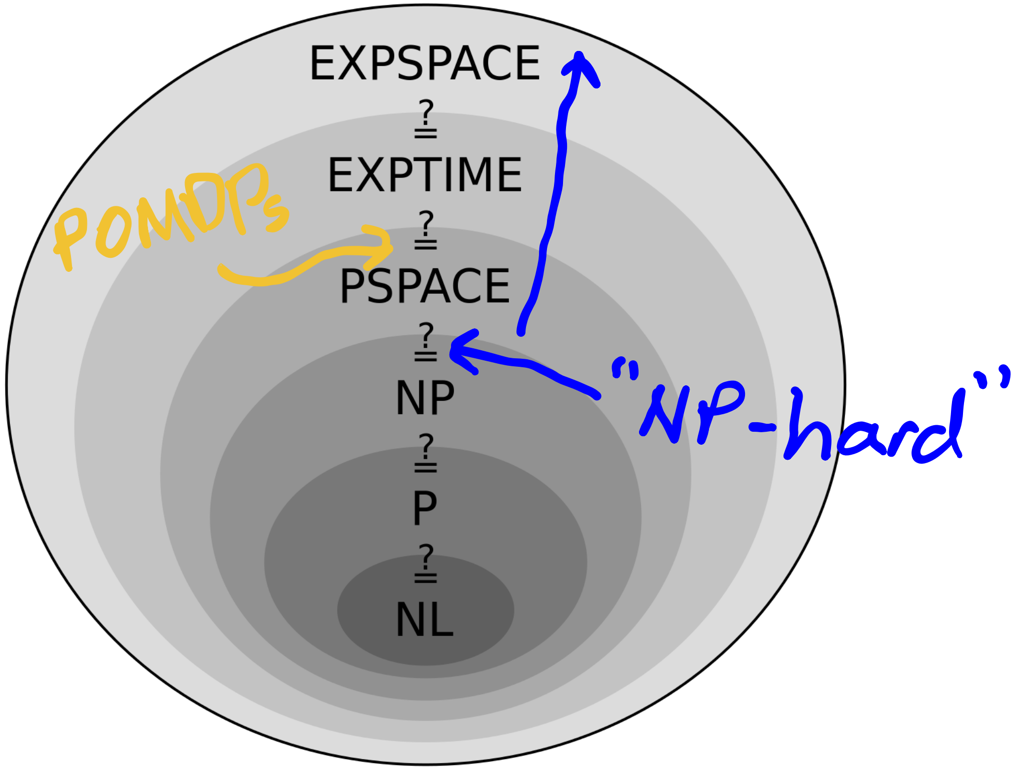


POMDP Computational Complexity

Sad facts ☹

- Infinite horizon POMDPs are *undecidable*
- Finite horizon POMDPs are *PSPACE Complete*
  - Among the hardest problems that can be solved using a polynomial amount of space
  - Any algorithm that can solve a general POMDP will have exponential complexity

  (we think)



Approximate POMDP Solutions

Approximate POMDP Solutions

Numerical Approximations

(approximately solve original problem)

Approximate POMDP Solutions

Numerical Approximations

(approximately solve original problem)

Offline

Approximate POMDP Solutions

Numerical Approximations

(approximately solve original problem)

Offline

Last week

Approximate POMDP Solutions

Numerical Approximations

(approximately solve original problem)

Offline

Last week

Online

Approximate POMDP Solutions

Numerical Approximations

(approximately solve original problem)

Offline

Last week

Online

Thursday

Approximate POMDP Solutions

Numerical Approximations

(approximately solve original problem)

Offline

Last week

Online

Thursday

Formulation Approximations

(solve a slightly different problem)

Approximate POMDP Solutions

Numerical Approximations

(approximately solve original problem)

Offline

Last week

Online

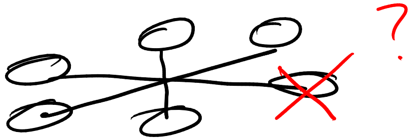
Thursday

Formulation Approximations

(solve a slightly different problem)

Today!

Rotor Failure Example



$$s = (\underbrace{x, y, z, \dot{x}, \dot{y}, \dot{z}, \phi, \theta, \psi, p, q, r}_{\text{known}}, \underbrace{m_1, m_2 \cdots m_6}_{\text{unkown}})$$

$\cancel{m \in \{\text{nominal}, \text{failed}\}}$

Certainty - Equivalence

POMDP Objective

POMDP Objective

$$\pi^* = \underset{\pi : B \to A}{\operatorname{argmax}} \; \mathrm{E}\left[\sum_{t=0}^{\infty} \gamma^t R(s_t, \pi(b_t))\right]$$

POMDP Objective

$$\pi^* = \underset{\pi: B \to A}{\operatorname{argmax}} \; \mathrm{E}\left[\sum_{t=0}^{\infty} \gamma^t R(s_t, \pi(b_t))\right]$$

$$b' = \tau(b, a, o)$$

Certainty Equivalent

POMDP Objective

$$\pi^* = \underset{\pi : B \to A}{\mathrm{argmax}} \; \mathrm{E}\left[\sum_{t=0}^{\infty} \gamma^t R(s_t, \pi(b_t))\right]$$

$$b' = \tau(b, a, o)$$

Certainty Equivalent

POMDP Objective

$$\pi^* = \underset{\pi:B\to A}{\operatorname{argmax}} \; \mathrm{E}\left[\sum_{t=0}^{\infty} \gamma^t R(s_t, \pi(b_t))\right]$$

$$b' = \tau(b, a, o)$$

$$b' = \tau(b, a, o)$$

Certainty Equivalent

POMDP Objective

$$\pi^* = \underset{\pi:B\to A}{\operatorname{argmax}} \; \mathrm{E}\left[\sum_{t=0}^{\infty} \gamma^t R(s_t, \pi(b_t))\right]$$

$$b' = \tau(b, a, o)$$

$$\pi_{\mathrm{CE}}(b) = \pi_s(\underset{s\sim b}{\mathrm{E}}[s])$$

π^*_{MDP} $\pi_s(\hat{s})$ $\hat{s} = \underset{s\sim b}{\mathrm{E}}[s]$

$$b' = \tau(b, a, o)$$

Certainty Equivalent

Optimal for LQG ← Gaussian

Linear Quadratic

$S = \mathbb{R}^n$

$A = \mathbb{R}^m$

$O = \mathbb{R}^p$

$$T(\mathbf{s}' \mid \mathbf{s}, \mathbf{a}) = \mathcal{N}(\mathbf{s}' \mid \mathbf{T}_s\mathbf{s} + \mathbf{T}_a\mathbf{a}, \mathbf{\Sigma}_s)$$

$$O(\mathbf{o} \mid \mathbf{s}') = \mathcal{N}(\mathbf{o} \mid \mathbf{O}_s\mathbf{s}', \mathbf{\Sigma}_o)$$

$$R(s,a) = -s^\top R_s s - a^\top R_a a$$

Kalman Filter

$$b(\mathbf{s}) = \mathcal{N}(\mathbf{s} \mid \boldsymbol{\mu}_b, \mathbf{\Sigma}_b)$$

Prediction

$$\boldsymbol{\mu}_p \leftarrow \mathbf{T}_s\boldsymbol{\mu}_b + \mathbf{T}_a\mathbf{a}$$

$$\mathbf{\Sigma}_p \leftarrow \mathbf{T}_s\mathbf{\Sigma}_b\mathbf{T}_s^\top + \mathbf{\Sigma}_s$$

Update

$$\mathbf{K} \leftarrow \mathbf{\Sigma}_p\mathbf{O}_s^\top\left(\mathbf{O}_s\mathbf{\Sigma}_p\mathbf{O}_s^\top + \mathbf{\Sigma}_o\right)^{-1}$$

$$\boldsymbol{\mu}_b \leftarrow \boldsymbol{\mu}_p + \mathbf{K}\left(\mathbf{o} - \mathbf{O}_s\boldsymbol{\mu}_p\right)$$

$$\mathbf{\Sigma}_b \leftarrow (\mathbf{I} - \mathbf{K}\mathbf{O}_s)\mathbf{\Sigma}_p$$

$$\pi^*(b) = K_a \operatorname{mean}(b)$$

QMDP

POMDP Objective

$$\pi^* \;= \underset{\pi : B \to A}{\operatorname{argmax}} \; \mathrm{E}\left[\sum_{t=0}^{\infty} \gamma^t R(s_t, \pi(b_t))\right]$$

$$b' = \tau(b, a, o)$$

QMDP

POMDP Objective

$$\pi^* = \underset{\pi: B \to A}{\operatorname{argmax}} \; \mathrm{E}\left[\sum_{t=0}^{\infty} \gamma^t R(s_t, \pi(b_t))\right]$$

$$b' = \tau(b, a, o)$$

$$b' = \tau(b, a, o)$$

QMDP

POMDP Objective

$$\pi^* = \underset{\pi:B\rightarrow A}{\text{argmax}}\ \mathrm{E}\left[\sum_{t=0}^{\infty} \gamma^t R(s_t, \pi(b_t))\right]$$

$$b' = \tau(b, a, o)$$

$$\underbrace{(S, A, T, R, O, Z, \gamma)}_{\text{solve}}$$

optimal Q-value for the fully observable MDP

$$\pi_{\text{QMDP}}(b) = \underset{a\in A}{\text{argmax}}\ \underset{s\sim b}{\mathrm{E}}\left[Q_{\text{MDP}}(s,a)\right]$$

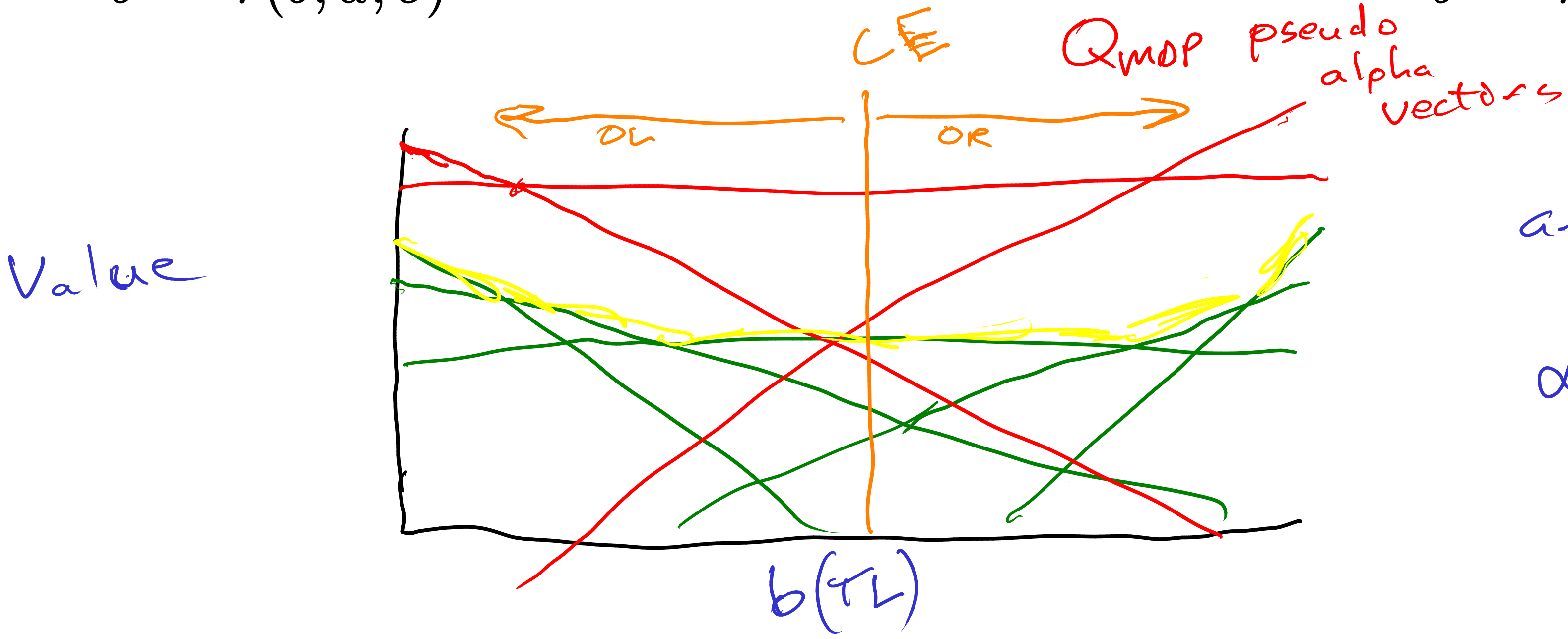
$$\underset{a\in A}{\text{argmax}}\ \pi_s\left(\underset{s\sim b}{\mathrm{E}}[s]\right)$$

CE

$$\underset{a\in A}{\text{argmax}}\ \alpha_a \cdot b$$

$$\alpha_a[s] = Q_{\text{MDP}}(s,a)$$

$$b' = \tau(b, a, o)$$



Example: Tiger POMDP with Waiting

$A = \{OL, OR, \text{Listen}, \text{Wait}\}$

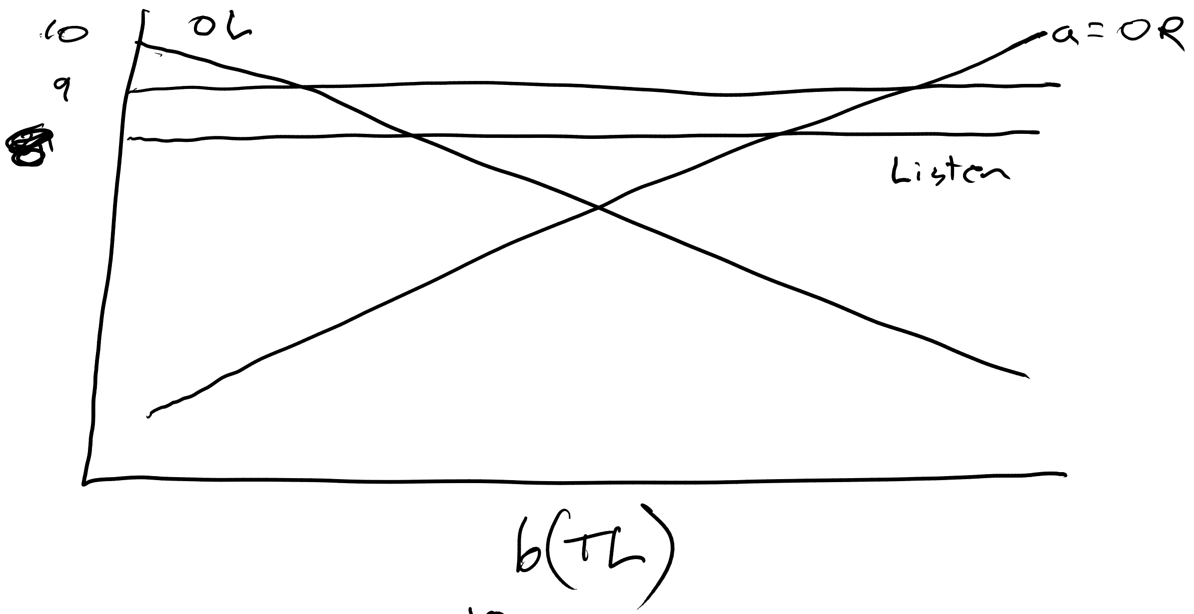
+10 if good

−100 if bad

−1 reward, 85% correct observation

0 reward, random observation

pseudo alpha vectors



$b(TL) = 50\%$

optimal = L

QMDP solution?

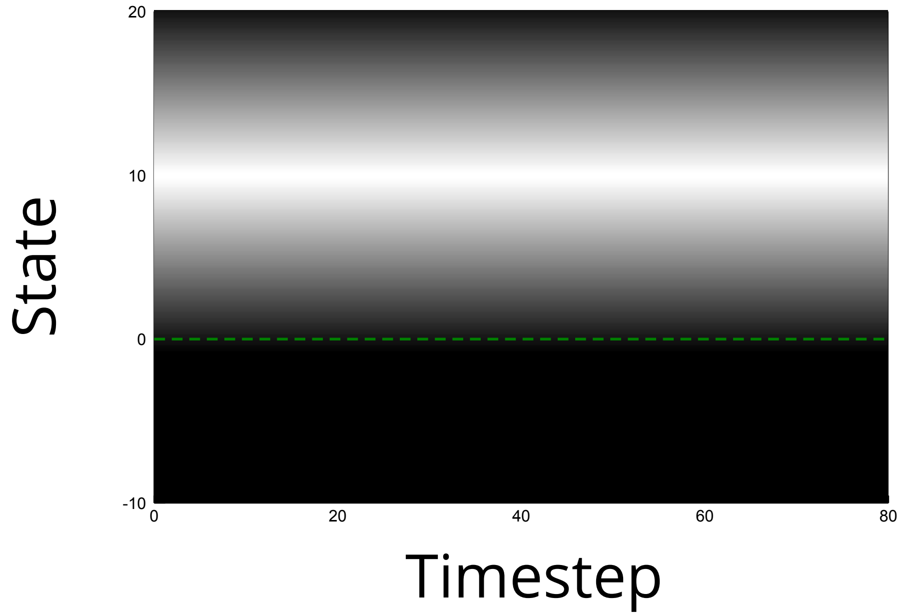
$Q_{MDP}(\text{any}, \text{open}) = 10$

$\gamma = 0.9$

+10

$-1 + \gamma \max_a Q_{MDP}(s,a)$

POMDP Example: Light-Dark



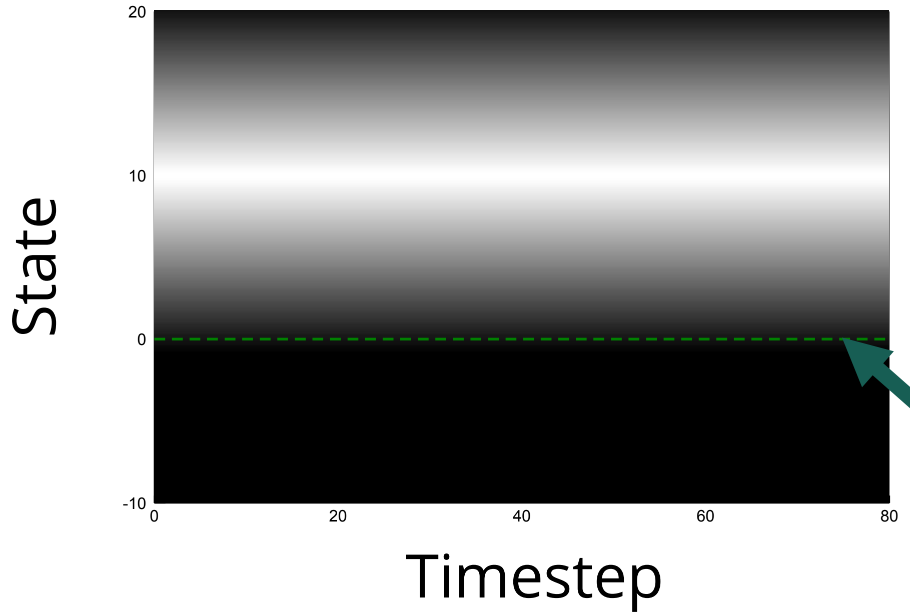
$$\mathcal{S} = \mathbb{Z} \qquad \mathcal{O} = \mathbb{R}$$

$$s' = s + a \qquad o \sim \mathcal{N}(s, s - 10)$$

$$\mathcal{A} = \{-10, -1, 0, 1, 10\}$$

$$R(s,a) = \begin{cases} 100 & \text{if } a = 0, s = 0 \\ -100 & \text{if } a = 0, s \neq 0 \\ -1 & \text{otherwise} \end{cases}$$

POMDP Example: Light-Dark



$$\mathcal{S} = \mathbb{Z} \qquad \mathcal{O} = \mathbb{R}$$

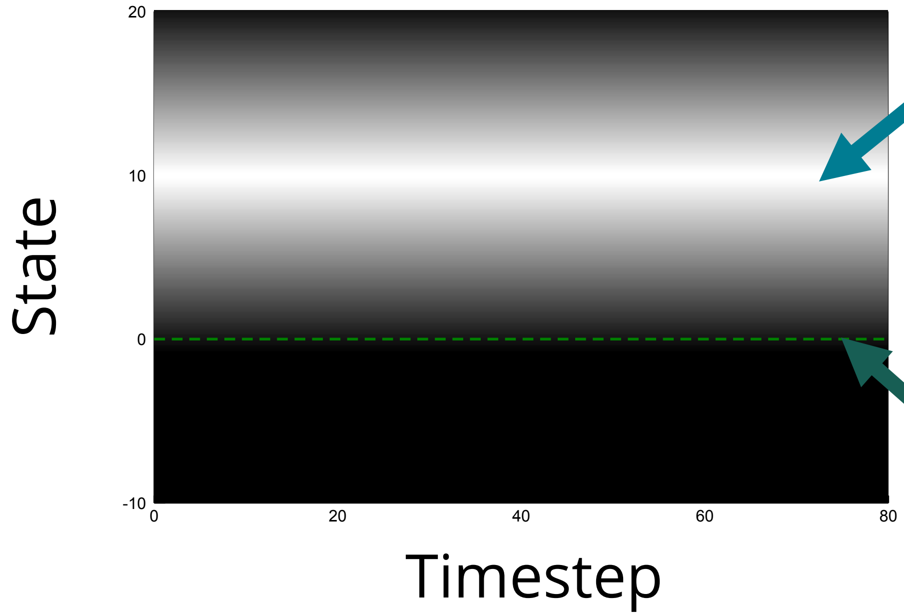
$$s' = s + a \qquad o \sim \mathcal{N}(s, s - 10)$$

$$\mathcal{A} = \{-10, -1, 0, 1, 10\}$$

$$R(s,a) = \begin{cases} 100 & \text{if } a = 0, s = 0 \\ -100 & \text{if } a = 0, s \neq 0 \\ -1 & \text{otherwise} \end{cases}$$

Goal: $a = 0$ at $s = 0$

POMDP Example: Light-Dark



Accurate Observations

$\mathcal{S} = \mathbb{Z} \qquad \mathcal{O} = \mathbb{R}$

$s' = s + a \qquad o \sim \mathcal{N}(s, s - 10)$

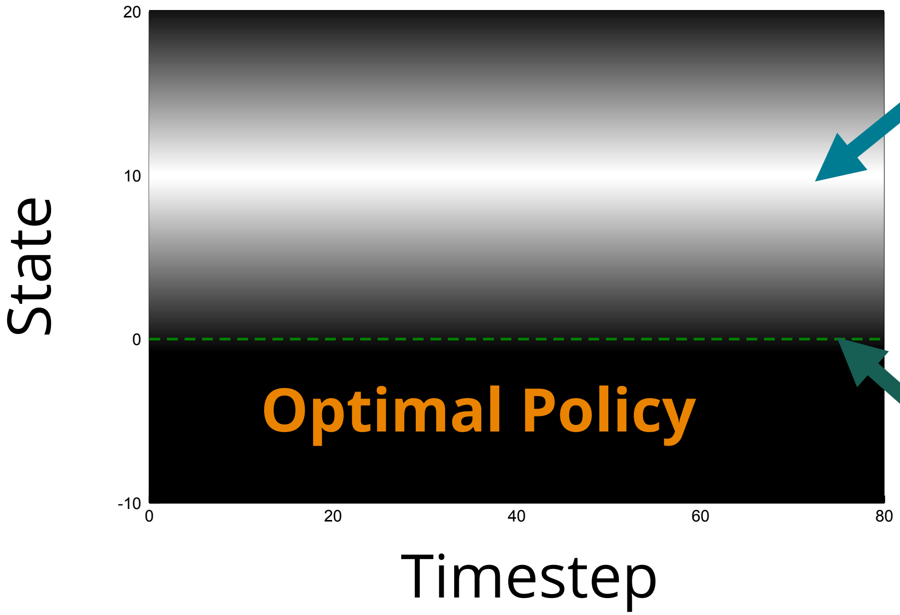
$\mathcal{A} = \{-10, -1, 0, 1, 10\}$

$$R(s,a) = \begin{cases} 100 & \text{if } a = 0, s = 0 \\ -100 & \text{if } a = 0, s \neq 0 \\ -1 & \text{otherwise} \end{cases}$$

Goal: $a = 0$ at $s = 0$

POMDP Example: Light-Dark

Accurate Observations



$$\mathcal{S} = \mathbb{Z} \qquad \mathcal{O} = \mathbb{R}$$

$$s' = s + a \qquad o \sim \mathcal{N}(s, s - 10)$$

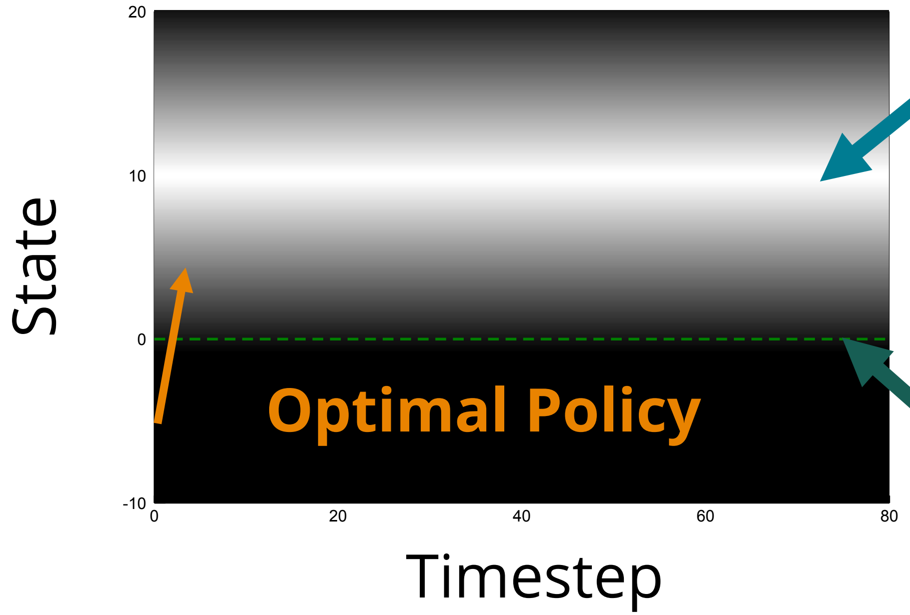
$$\mathcal{A} = \{-10, -1, 0, 1, 10\}$$

$$R(s,a) = \begin{cases} 100 & \text{if } a = 0, s = 0 \\ -100 & \text{if } a = 0, s \neq 0 \\ -1 & \text{otherwise} \end{cases}$$

Goal: $a = 0$ at $s = 0$

POMDP Example: Light-Dark

Accurate Observations



$$\mathcal{S} = \mathbb{Z} \qquad \mathcal{O} = \mathbb{R}$$

$$s' = s + a \qquad o \sim \mathcal{N}(s, s - 10)$$

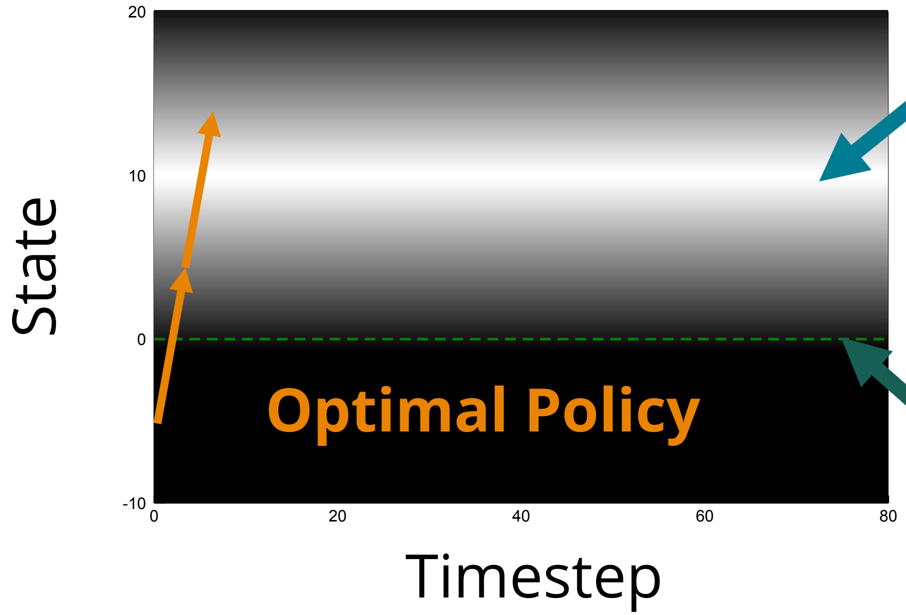
$$\mathcal{A} = \{-10, -1, 0, 1, 10\}$$

$$R(s,a) = \begin{cases} 100 & \text{if } a = 0, s = 0 \\ -100 & \text{if } a = 0, s \neq 0 \\ -1 & \text{otherwise} \end{cases}$$

Goal: $a = 0$ at $s = 0$

POMDP Example: Light-Dark

Accurate Observations



$$\mathcal{S} = \mathbb{Z} \qquad \mathcal{O} = \mathbb{R}$$

$$s' = s + a \qquad o \sim \mathcal{N}(s, s - 10)$$

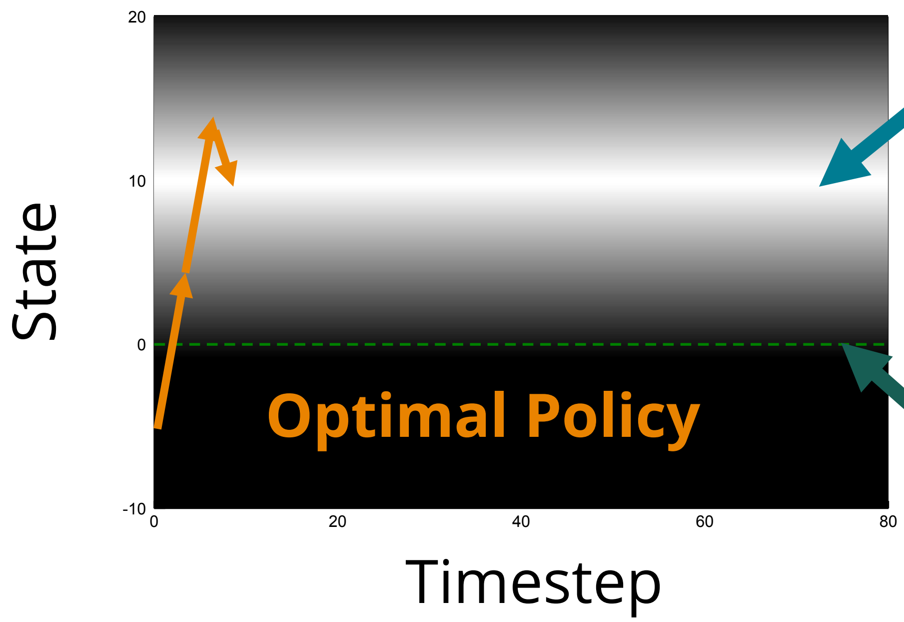
$$\mathcal{A} = \{-10, -1, 0, 1, 10\}$$

$$R(s,a) = \begin{cases} 100 & \text{if } a = 0, s = 0 \\ -100 & \text{if } a = 0, s \neq 0 \\ -1 & \text{otherwise} \end{cases}$$

Goal: $a = 0$ at $s = 0$

POMDP Example: Light-Dark

Accurate Observations



$$\mathcal{S} = \mathbb{Z} \qquad \mathcal{O} = \mathbb{R}$$

$$s' = s + a \qquad o \sim \mathcal{N}(s, s - 10)$$

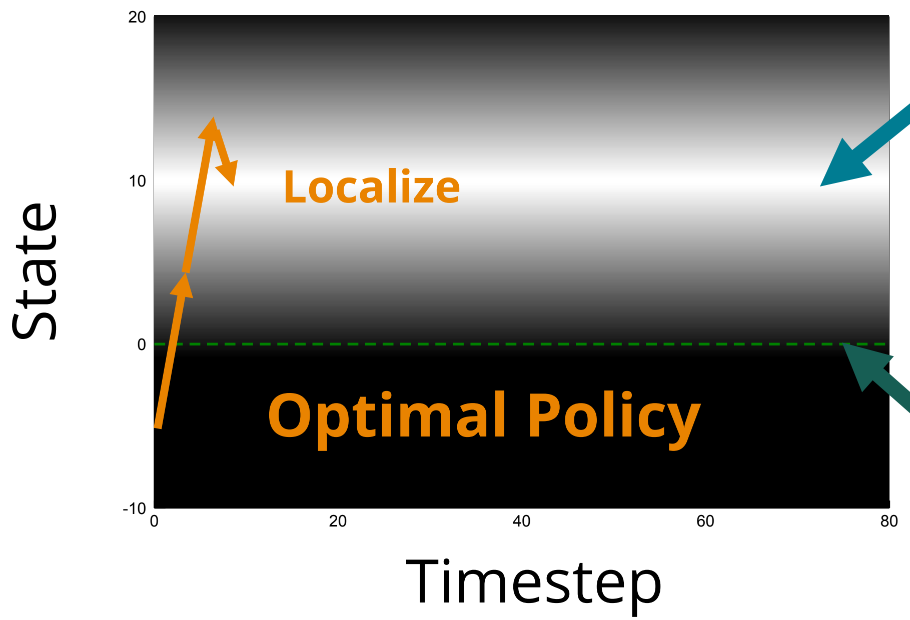
$$\mathcal{A} = \{-10, -1, 0, 1, 10\}$$

$$R(s,a) = \begin{cases} 100 & \text{if } a = 0, s = 0 \\ -100 & \text{if } a = 0, s \neq 0 \\ -1 & \text{otherwise} \end{cases}$$

Goal: $a = 0$ at $s = 0$

POMDP Example: Light-Dark

Accurate Observations



$$\mathcal{S} = \mathbb{Z} \qquad \mathcal{O} = \mathbb{R}$$

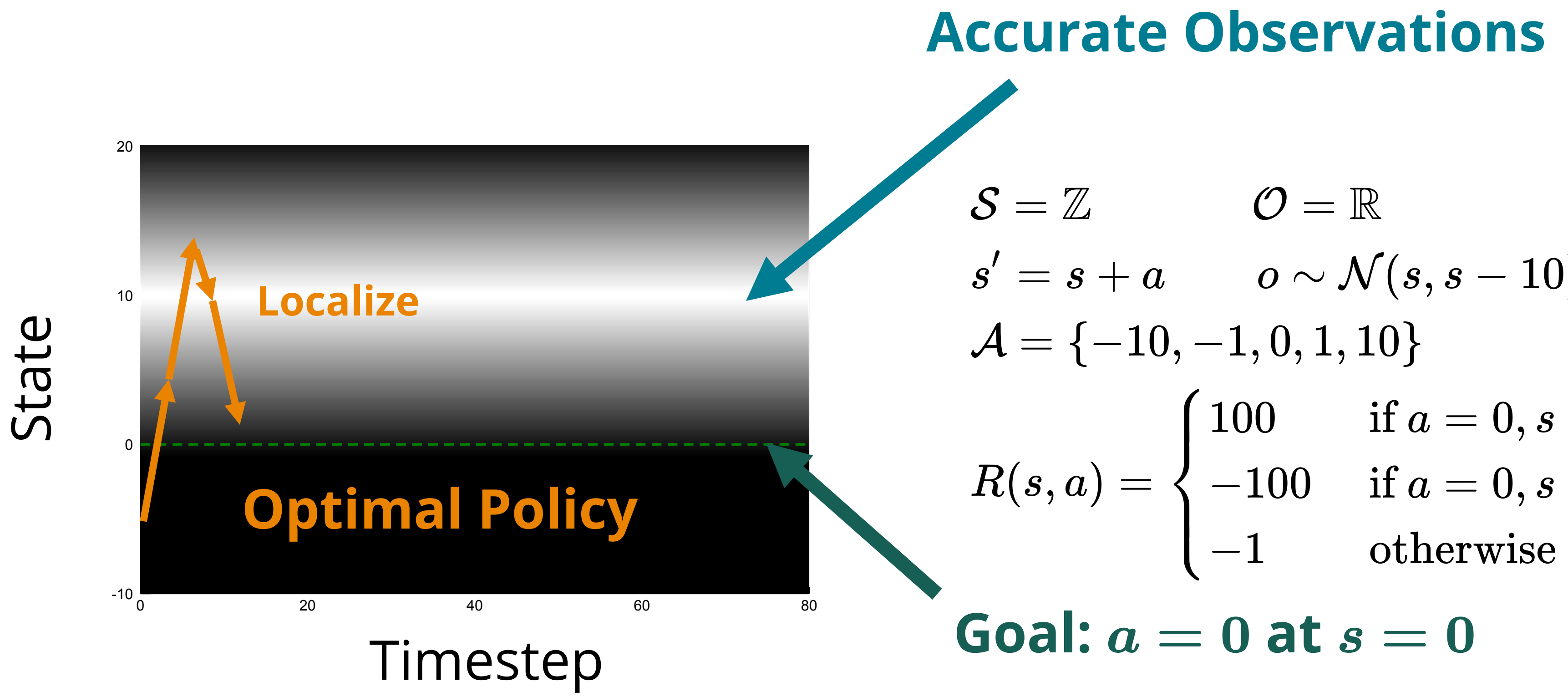
$$s' = s + a \qquad o \sim \mathcal{N}(s, s - 10)$$

$$\mathcal{A} = \{-10, -1, 0, 1, 10\}$$

$$R(s,a) = \begin{cases} 100 & \text{if } a = 0, s = 0 \\ -100 & \text{if } a = 0, s \neq 0 \\ -1 & \text{otherwise} \end{cases}$$

Goal: $a = 0$ at $s = 0$

POMDP Example: Light-Dark



$$\mathcal{S} = \mathbb{Z} \qquad \mathcal{O} = \mathbb{R}$$

$$s' = s + a \qquad o \sim \mathcal{N}(s, s - 10)$$

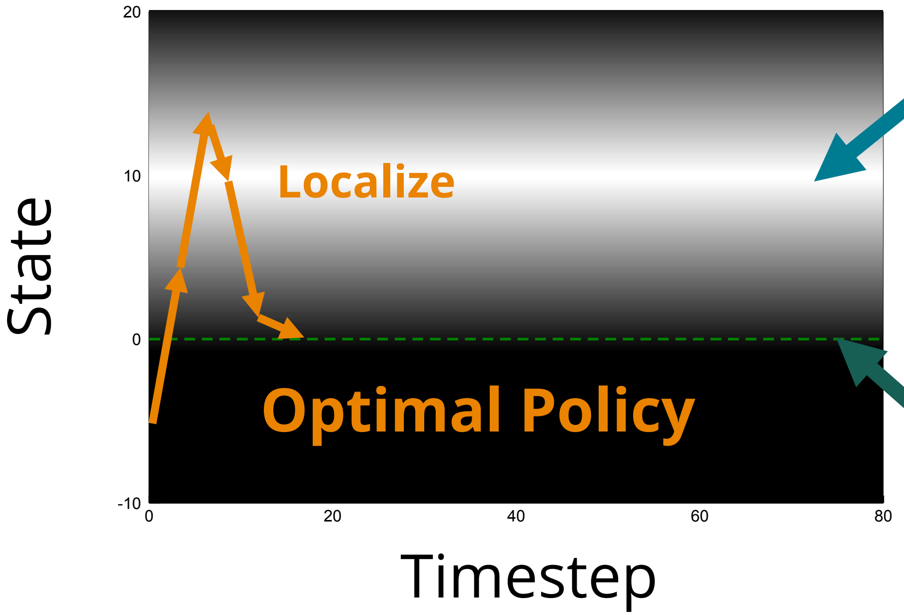
$$\mathcal{A} = \{-10, -1, 0, 1, 10\}$$

$$R(s,a) = \begin{cases} 100 & \text{if } a = 0, s = 0 \\ -100 & \text{if } a = 0, s \neq 0 \\ -1 & \text{otherwise} \end{cases}$$

Goal: $a = 0$ at $s = 0$

POMDP Example: Light-Dark

Accurate Observations



$$\mathcal{S} = \mathbb{Z} \qquad \mathcal{O} = \mathbb{R}$$

$$s' = s + a \qquad o \sim \mathcal{N}(s, s - 10)$$

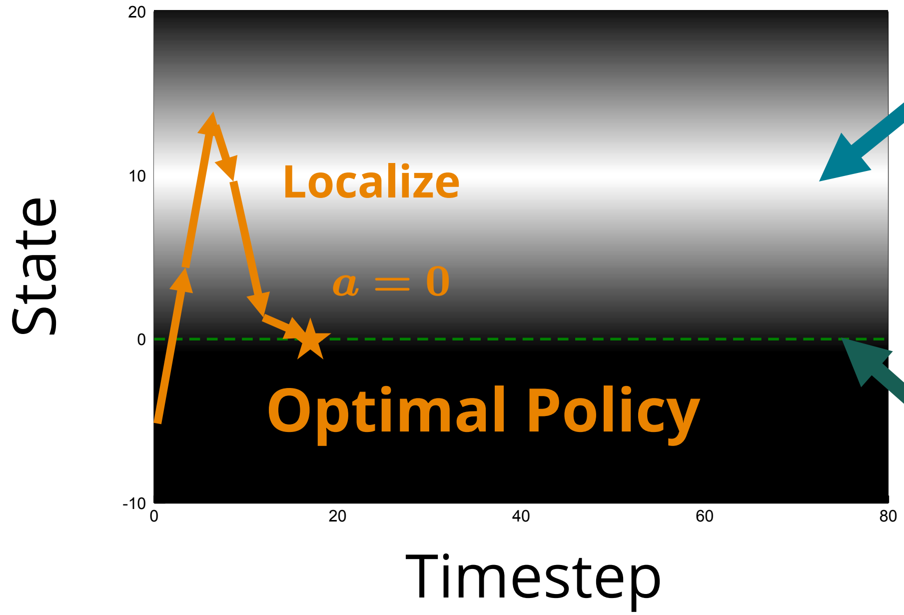
$$\mathcal{A} = \{-10, -1, 0, 1, 10\}$$

$$R(s,a) = \begin{cases} 100 & \text{if } a = 0, s = 0 \\ -100 & \text{if } a = 0, s \neq 0 \\ -1 & \text{otherwise} \end{cases}$$

Goal: $a = 0$ at $s = 0$

POMDP Example: Light-Dark

Accurate Observations



$$\mathcal{S} = \mathbb{Z} \qquad \mathcal{O} = \mathbb{R}$$

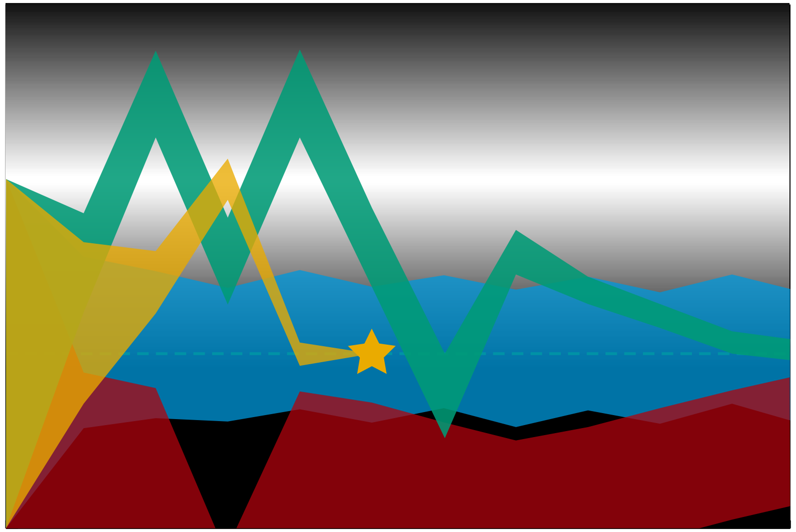
$$s' = s + a \qquad o \sim \mathcal{N}(s, s - 10)$$

$$\mathcal{A} = \{-10, -1, 0, 1, 10\}$$

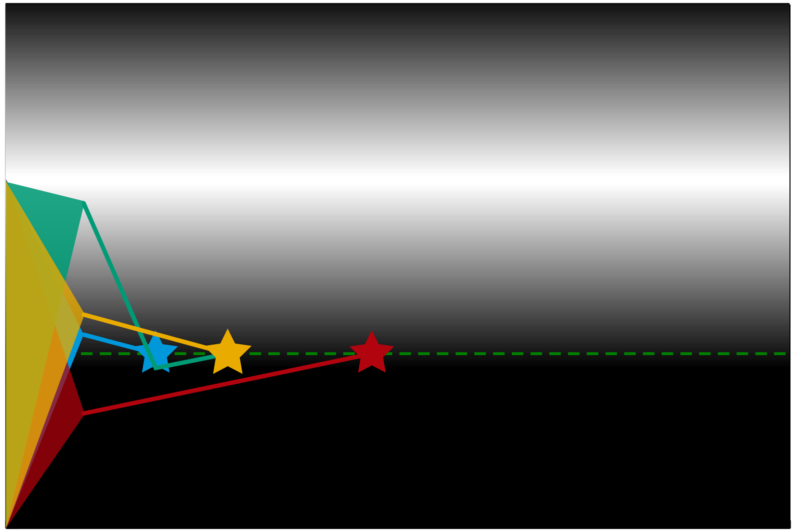
$$R(s,a) = \begin{cases} 100 & \text{if } a = 0, s = 0 \\ -100 & \text{if } a = 0, s \neq 0 \\ -1 & \text{otherwise} \end{cases}$$

Goal: $a = 0$ at $s = 0$

POMDP Solution



QMDP

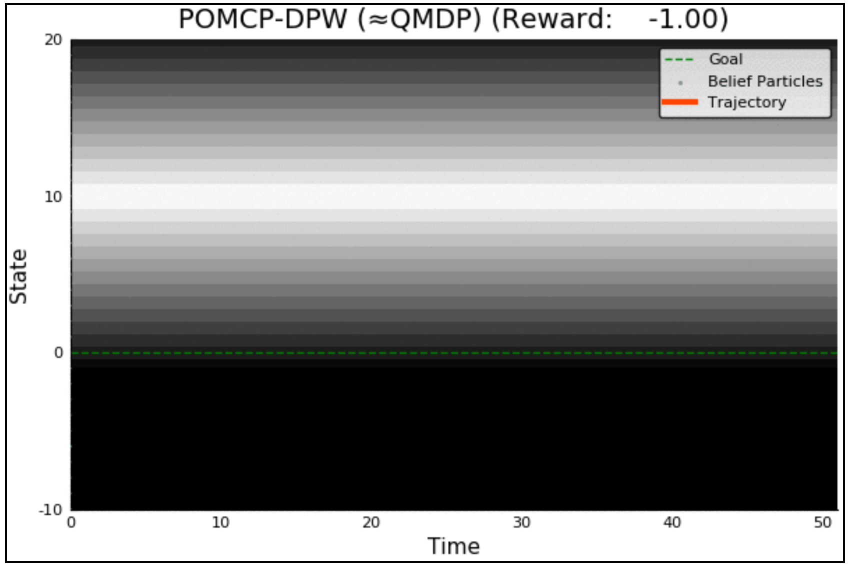


Same as **full observability** on the next step

Information Gathering

QMDP

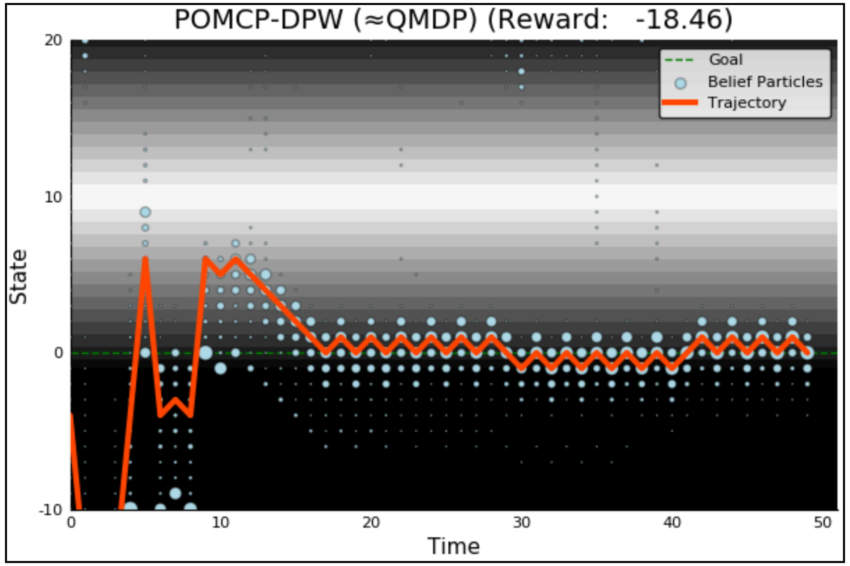
Full POMDP



Information Gathering

QMDP

Full POMDP

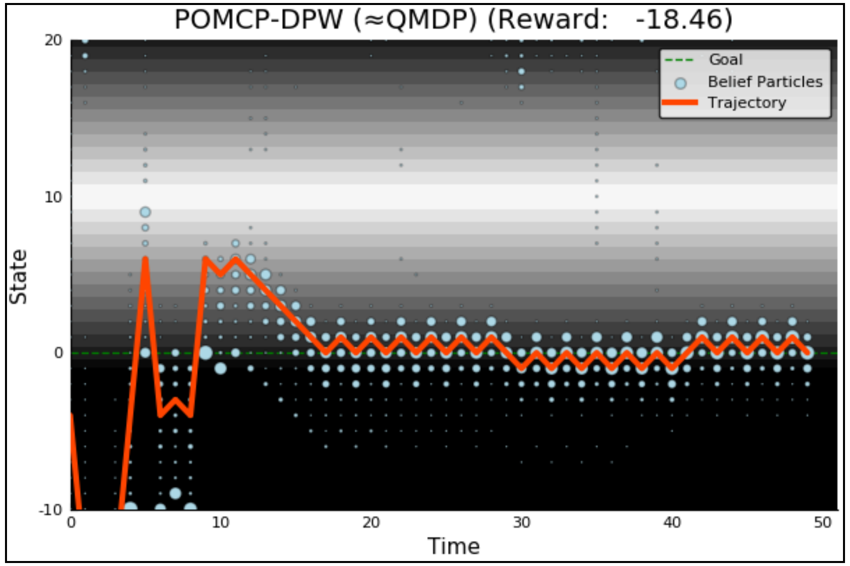


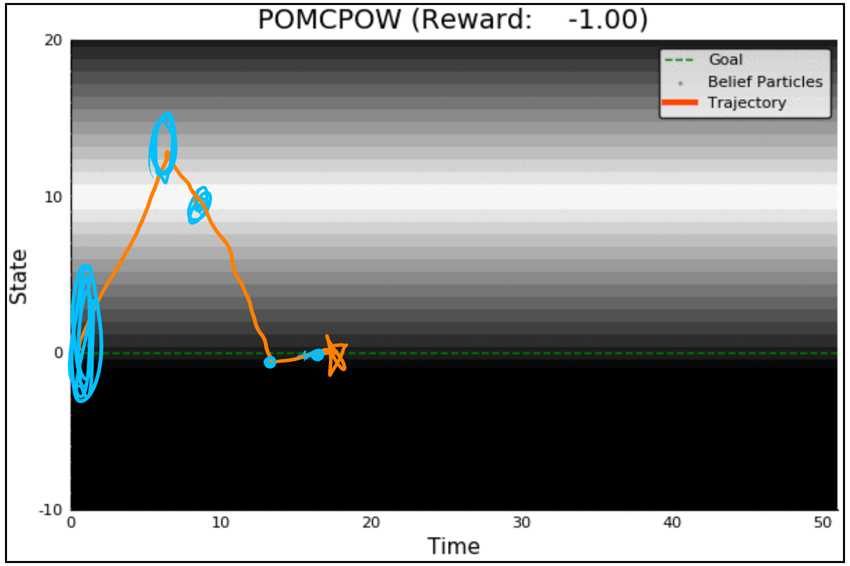
Information Gathering

QMDP

Full POMDP

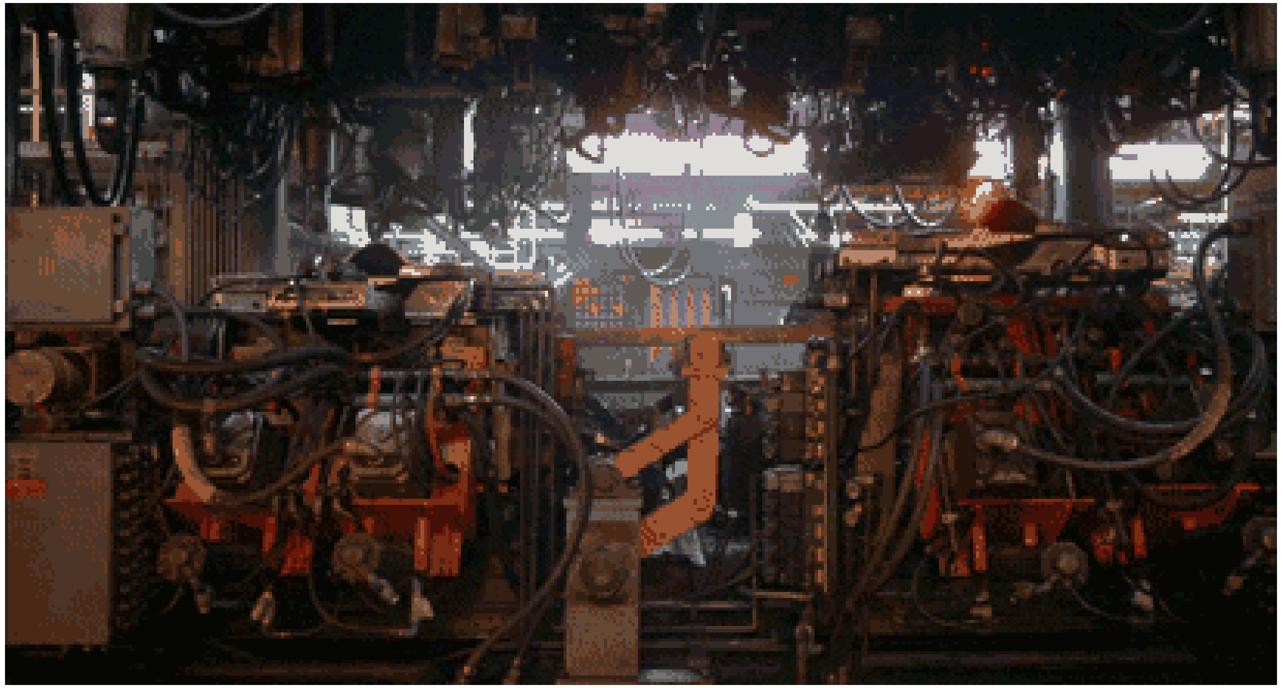
50





QMDP

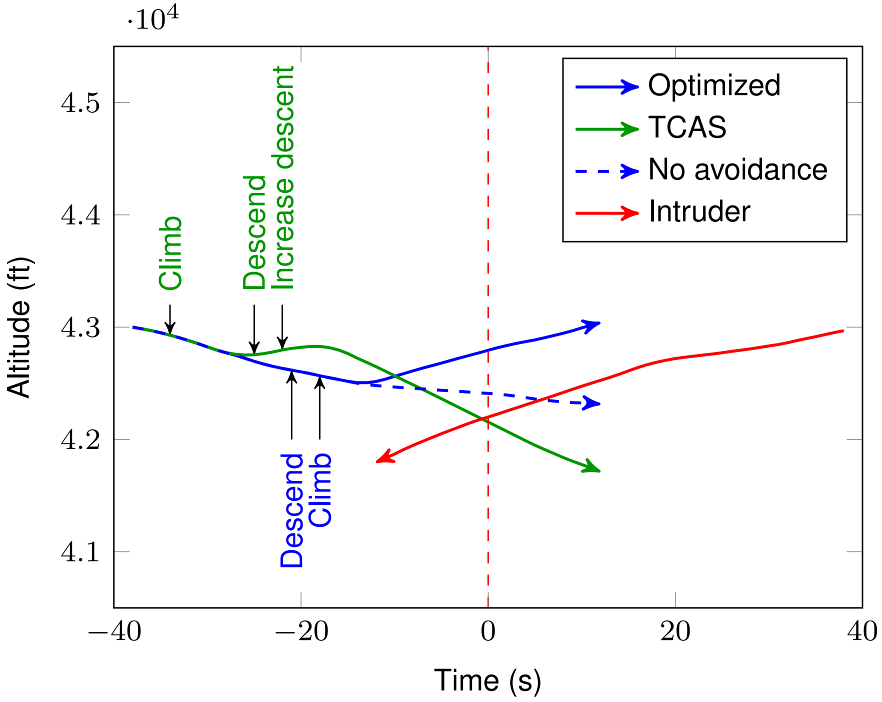
INDUSTRIAL GRADE



QMDP

ACAS X

[Kochenderfer, 2011]



Hindsight Optimization

POMDP Objective

$$\pi^* = \underset{\pi : B \to A}{\operatorname{argmax}} \, \mathrm{E}\left[\sum_{t=0}^{\infty} \gamma^t R(s_t, \pi(b_t))\right]$$

$$b' = \tau(b, a, o)$$

$$\pi_{HOP}(b) = \underset{a^i_{0:\infty}}{\operatorname{argmax}} \frac{1}{m} \sum_{i=1}^{m} \gamma^t R(s^i_t, a^i_t)$$

Subject to $\quad s^i_{t+1} = G(s^i_t, a^i_t, \phi^i_t)$

fixed outcome

$a^i_1 = a^j_1 \quad \forall i, j$

m scenarios

FIB

POMDP Objective

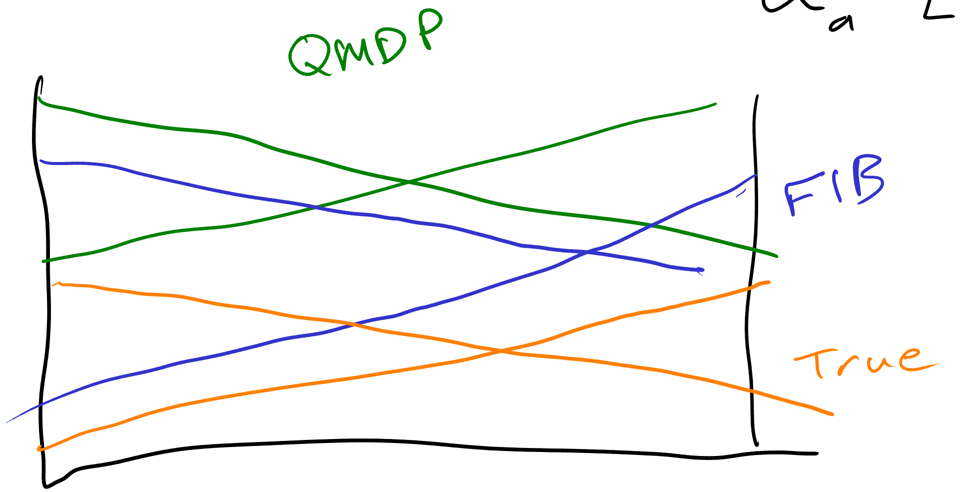
$$\pi^* = \underset{\pi:B\to A}{\operatorname{argmax}} \; \mathrm{E}\left[\sum_{t=0}^{\infty} \gamma^t R(s_t, \pi(b_t))\right]$$

$$b' = \tau(b, a, o)$$

$$\pi_{FIB}(b) = \underset{a \in A}{\operatorname{argmax}} \; \alpha_a^T b$$

iterate

$$\alpha_a^{(k+1)}[s] = R(s,a) + \gamma \sum_o \max_{a'} \sum_{s'} Z(o \mid a, s') \, T(s' \mid s,a) \, \alpha_{a'}^{(k)}[s']$$



k-Markov

POMDP Objective

$$\pi^* = \underset{\pi:B\to A}{\operatorname{argmax}}\ \mathrm{E}\left[\sum_{t=0}^{\infty}\gamma^t R(s_t, \pi(b_t))\right]$$

$$b' = \tau(b, a, o)$$

k previous observations

Open Loop

POMDP Objective

$$\pi^* = \underset{\pi : B \to A}{\operatorname{argmax}} \, \mathrm{E}\left[\sum_{t=0}^{\infty} \gamma^t R(s_t, \pi(b_t))\right]$$

$$\max_{a_1, a_2, \ldots a_T} E\left[\sum_{t=0}^{T} \gamma^t R(s_t, a_t)\right]$$

$$b' = \tau(b, a, o)$$