POMDP Value Iteration $|S| \sim 10\text{-}20$

SARSOP $|S| \sim 10{,}000$

POMDP Formulation Approximations

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

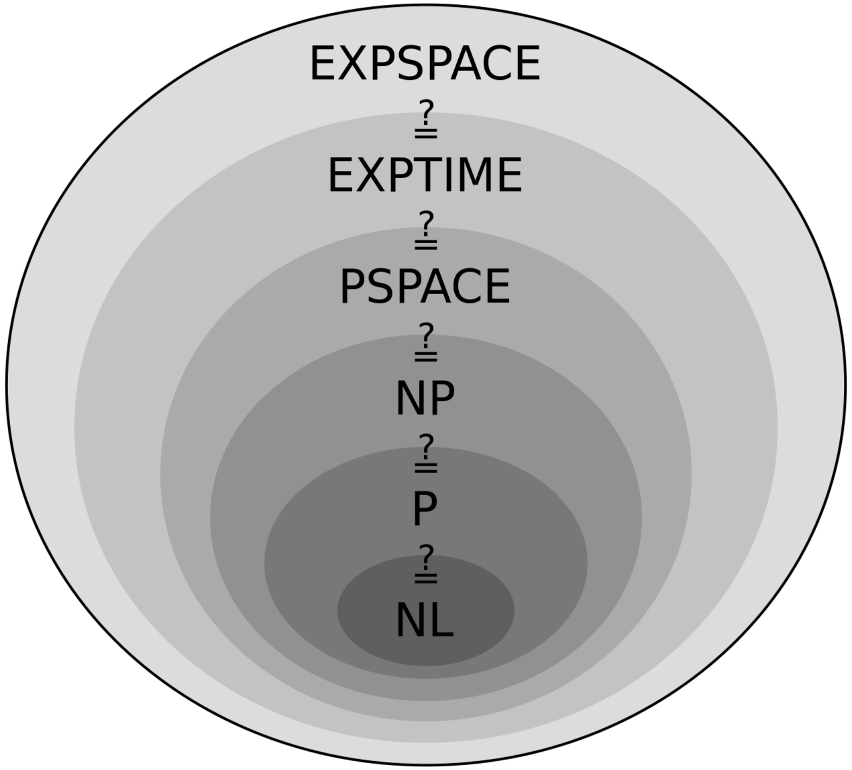
Sad facts ●

- 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

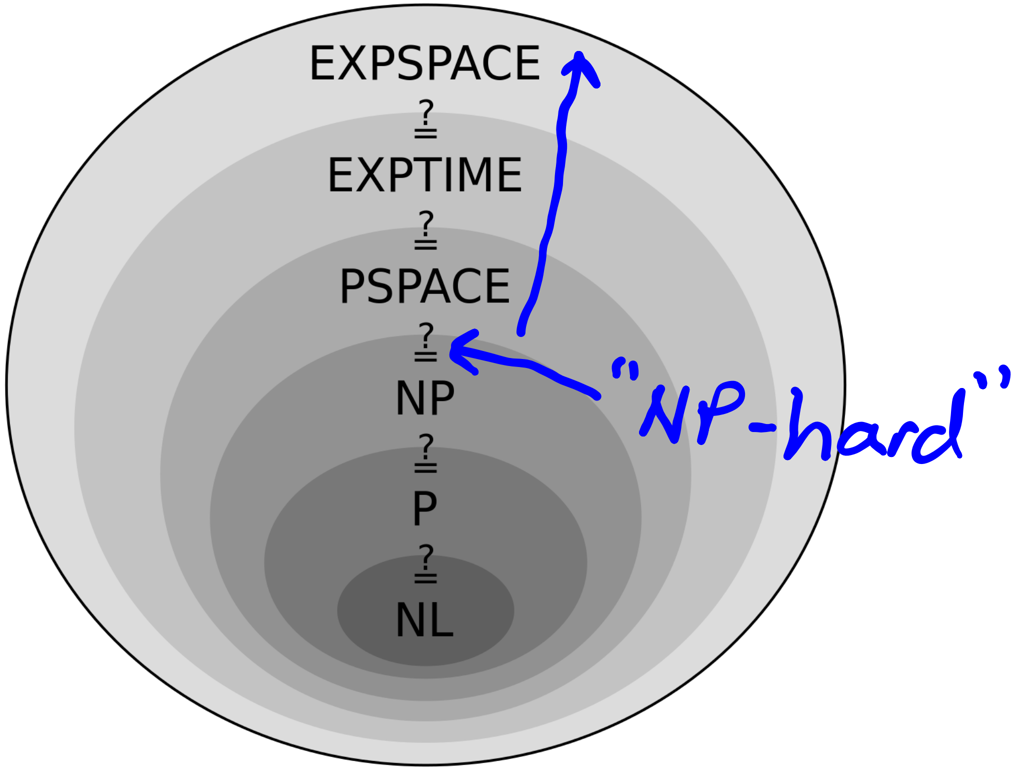
- Infinite horizon POMDPs are *undecidable*
- Finite horizon POMDPs are *PSPACE Complete*
    - Among the hardest problems that can be solved using a polynomial amount of space



POMDP Computational Complexity

Sad facts ●

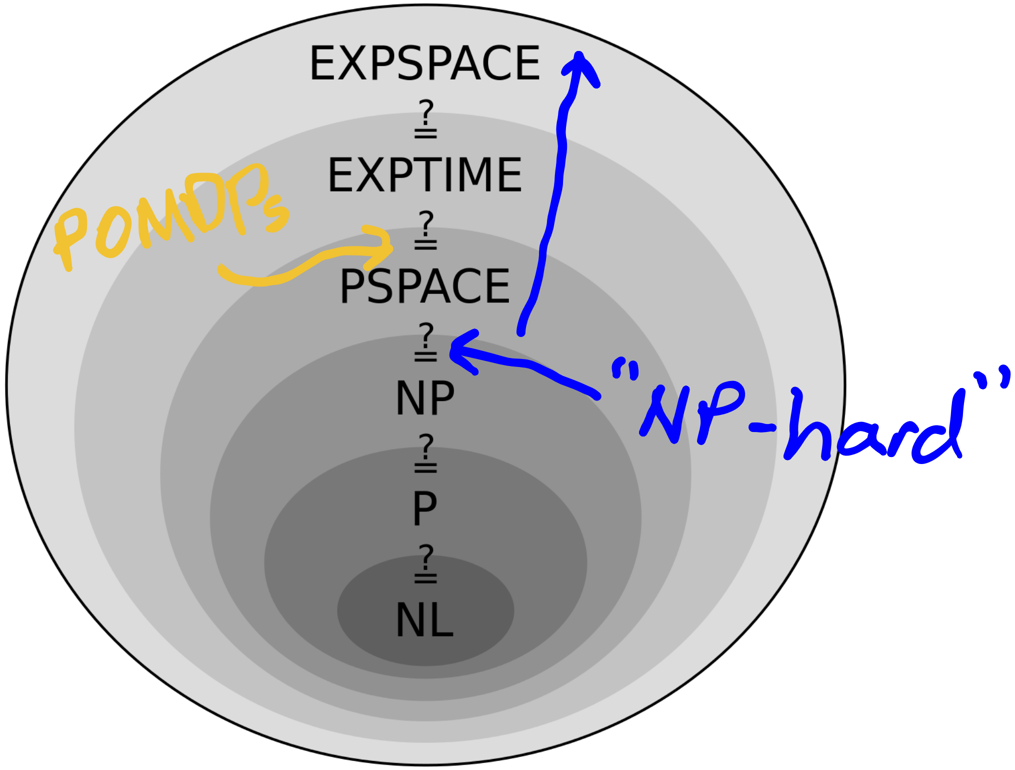
- Infinite horizon POMDPs are *undecidable*
- Finite horizon POMDPs are *PSPACE Complete*
  - Among the hardest problems that can be solved using a polynomial amount of space



POMDP Computational Complexity

Sad facts ●

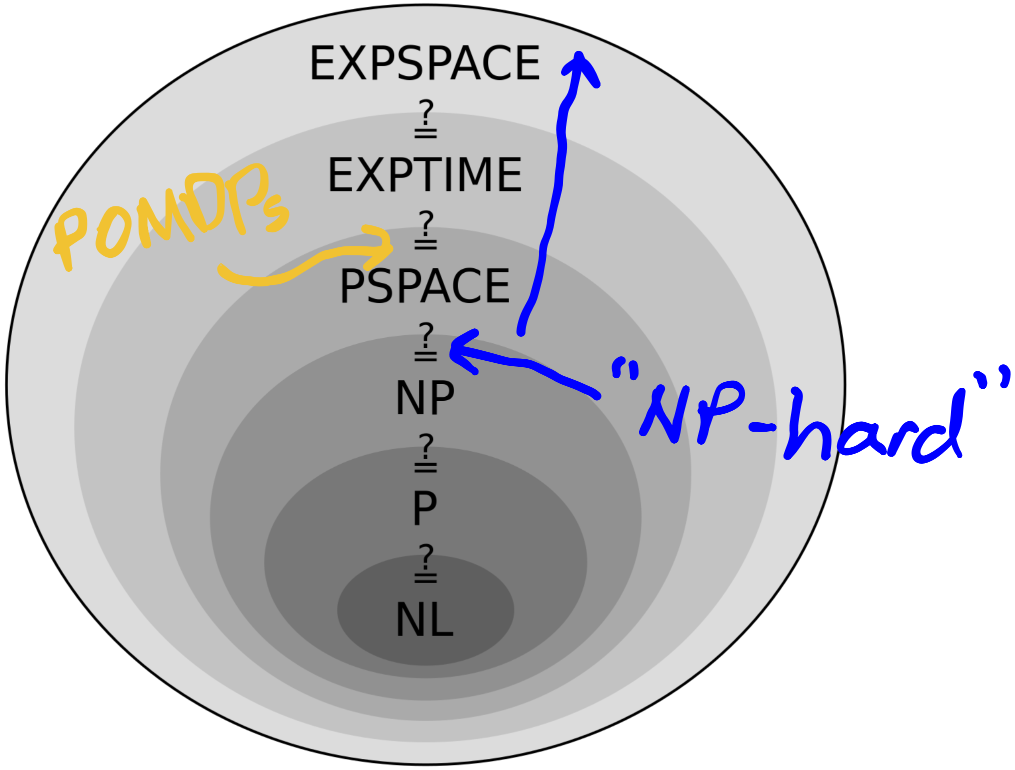
- Infinite horizon POMDPs are *undecidable*
- Finite horizon POMDPs are *PSPACE Complete*
  - Among the hardest problems that can be solved using a polynomial amount of space



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

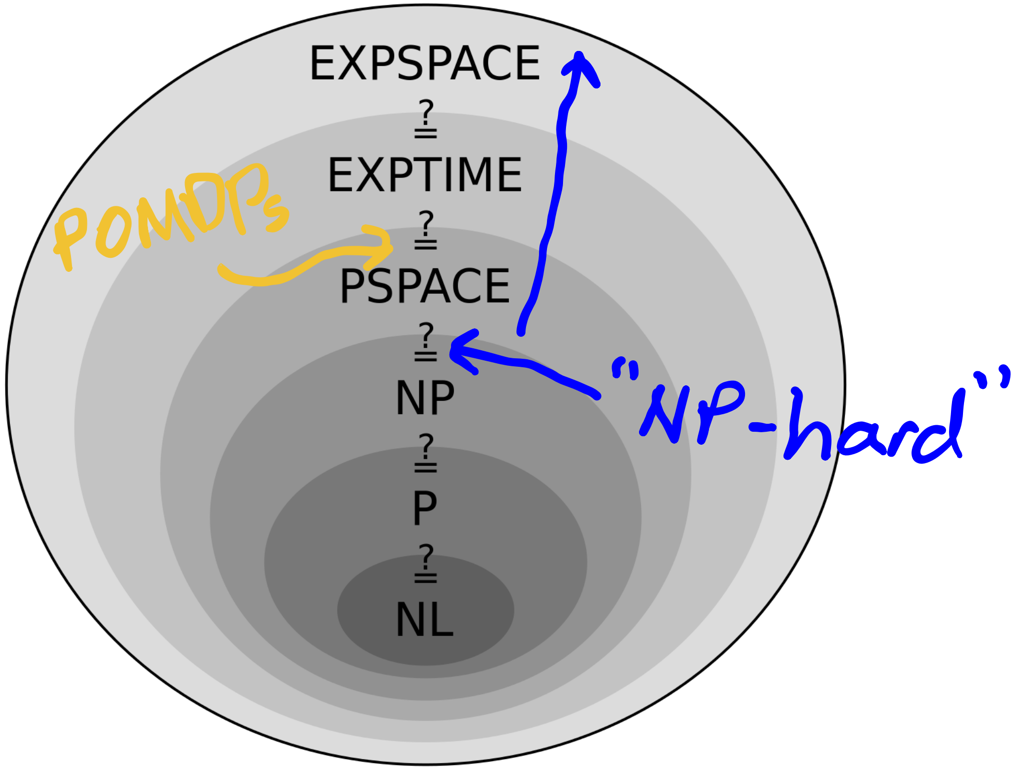


POMDP Computational Complexity

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

$\overline{V} - \underline{V} = \varepsilon$

Numerical Approximations

(approximately solve original problem)

Offline

~~Last week~~

Today

scalable to $|S| \sim 10{,}000$

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!

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

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

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

.

Certainty Equivalent

POMDP Objective

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

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

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

Certainty Equivalent

POMDP Objective

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

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

or mode or median

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

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

Certainty Equivalent

MDP LQR

Optimal for LQG

LQG POMDP

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

Linear Dynamics

Gaussian Process Noise

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

Gaussian Observation Noise

mean covariance

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

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

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

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

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

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

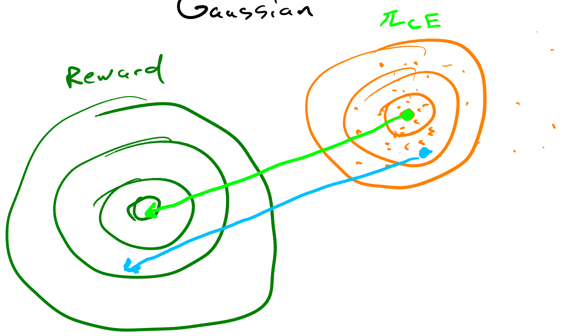
Bayesian Update

Kalman Filter

$$\pi^*(b) = -K_{LQR}\,\mu_b$$

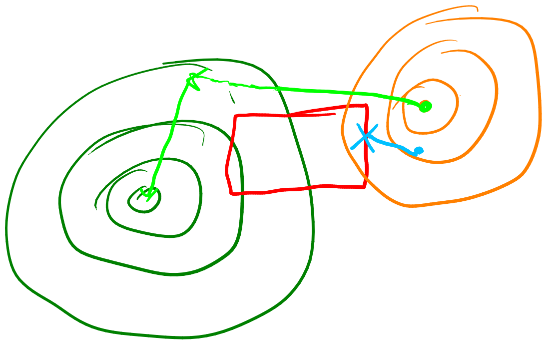
$S = \mathbb{R}^2$ Linear Dynamics + Observations Gaussian   Reward Quadratic

Case 1:



C.E. is optimal

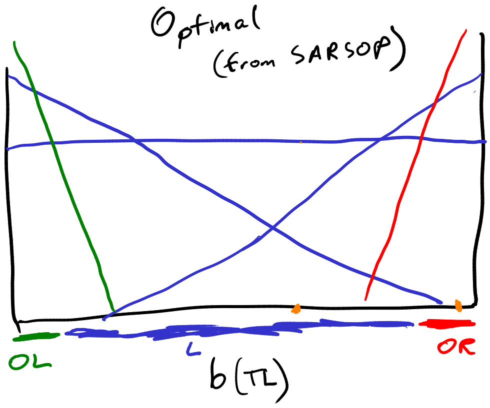
Case 2:



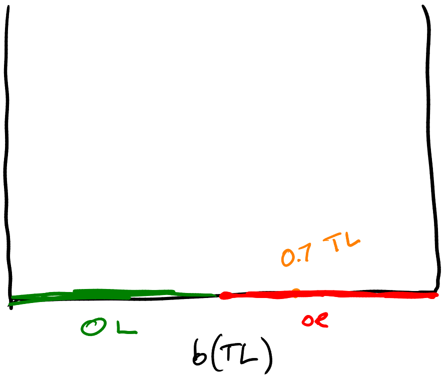
C.E. is bad

Tiger POMDP

Optimal (from SARSOP)



C.E. using mode



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

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

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

solve underlying MDP to get QMDP

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

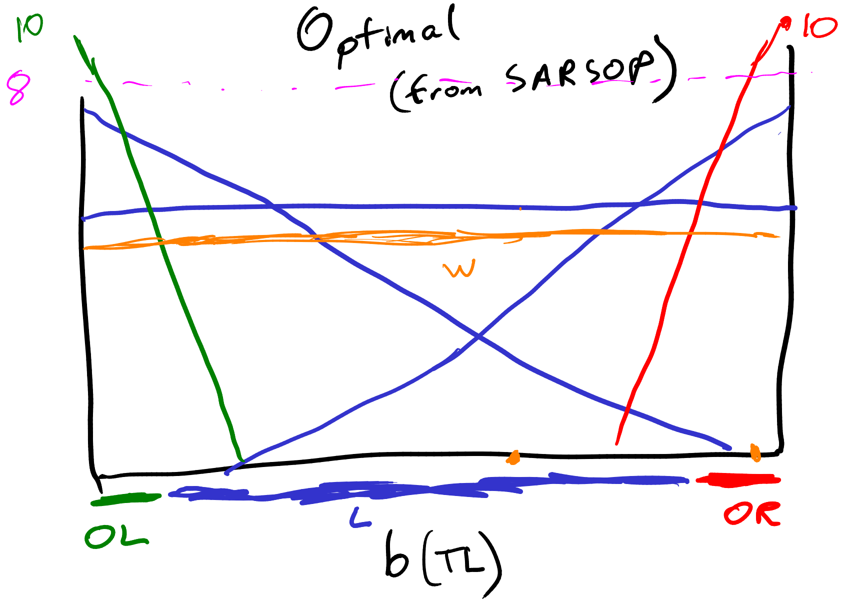
$$b' = \tau(b, a, o)$$

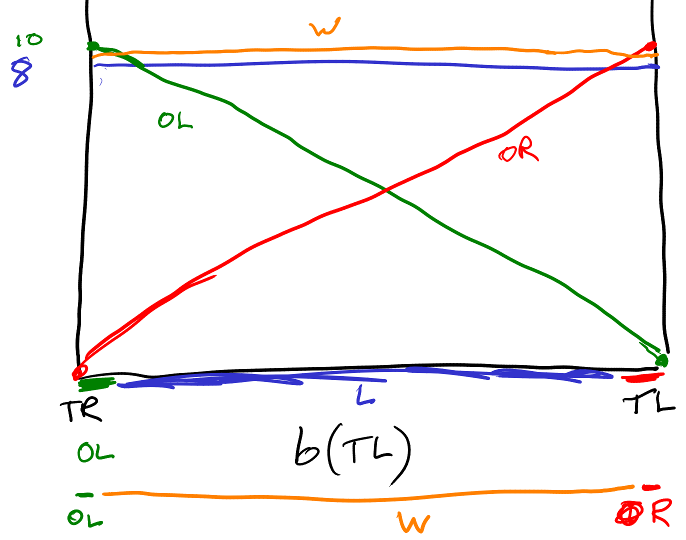
Example: Tiger POMDP with Waiting

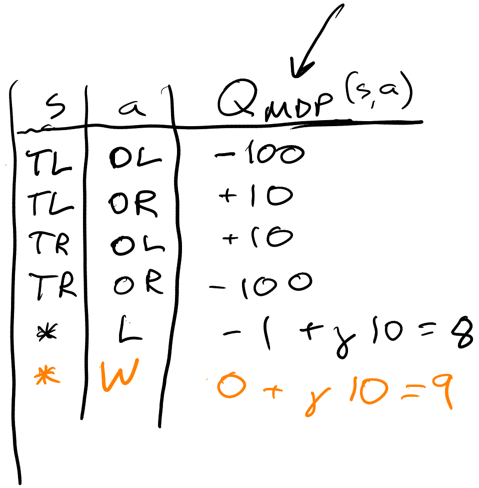
Terminates when door is open

Always an upper bound on true POMDP Value function $\rightarrow$ $\underset{s \sim b}{E}[Q_{MDP}(s,a)] = b(TL)\, Q_{MDP}(TL,a) + (1 - b(TL))\, Q_{MDP}(TR,a)$

$\gamma = 0.9$







s	a	$Q_{MDP}(s,a)$
TL	OL	−100
TL	OR	+10
TR	OL	+10
TR	OR	−100
*	L	$-1 + \gamma 10 = 8$
*	W	$0 + \gamma 10 = 9$

add new action Wait

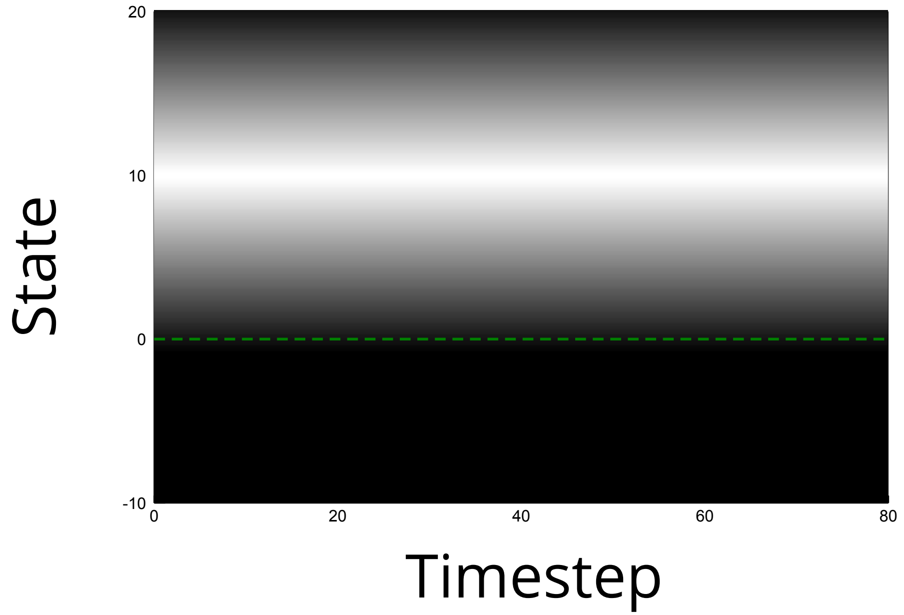
50-50 observation

0 immediate reward

QMDP is bad at <u>costly</u> information gathering + long-lasting uncertainty

o.w. it's pretty good

POMDP Example: Light-Dark



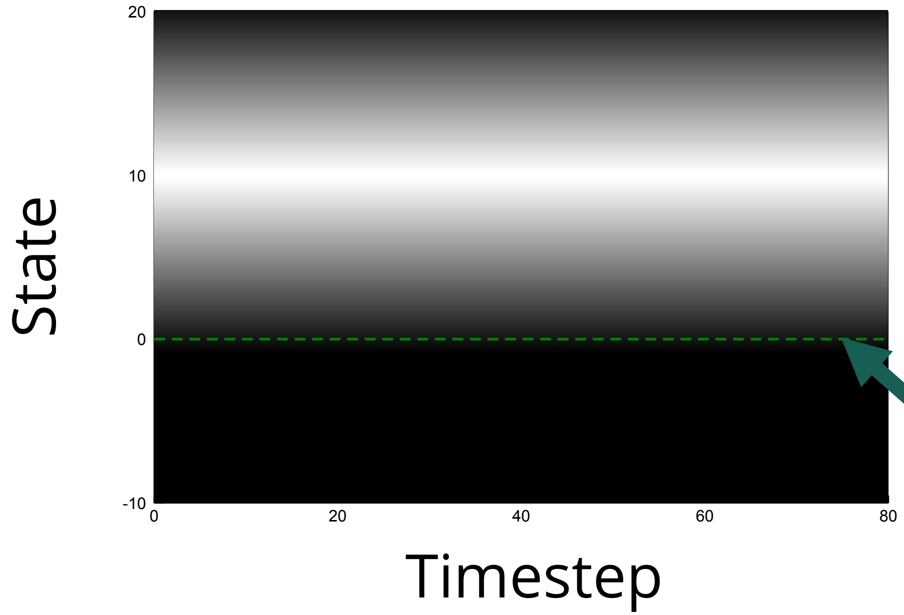
$$\mathcal{S} = \mathbb{Z} \qquad\qquad \mathcal{O} = \mathbb{R}$$

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

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

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

POMDP Example: Light-Dark



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

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

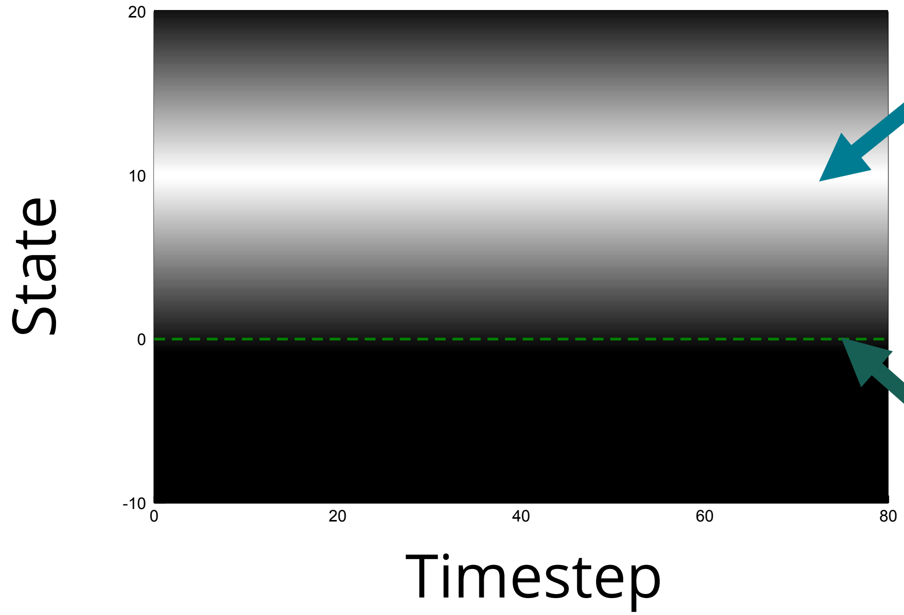
$$\mathcal{A} = \{-10, -1, 0, 1, 10\}$$

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

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

POMDP Example: Light-Dark

Accurate Observations



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

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

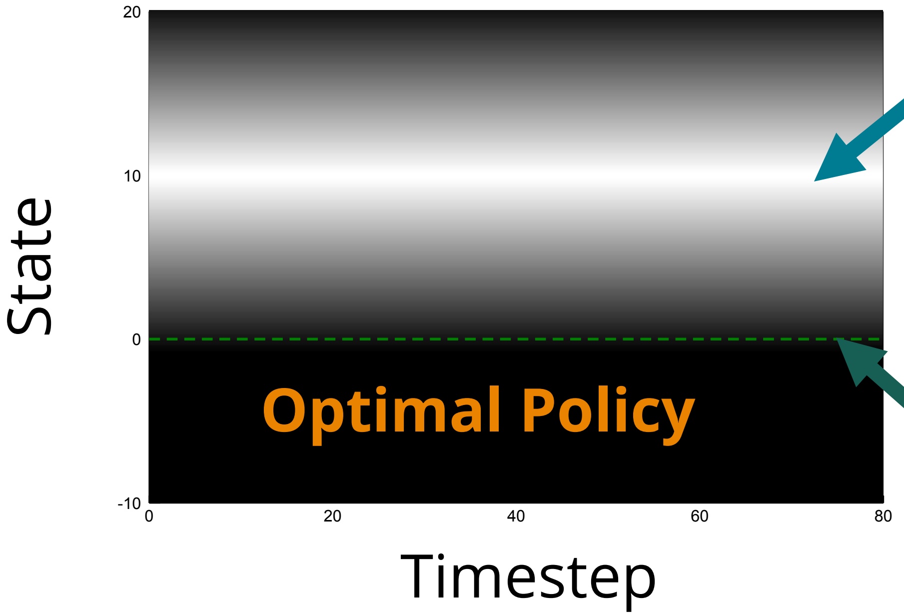
$$\mathcal{A} = \{-10, -1, 0, 1, 10\}$$

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

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

POMDP Example: Light-Dark

Accurate Observations



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

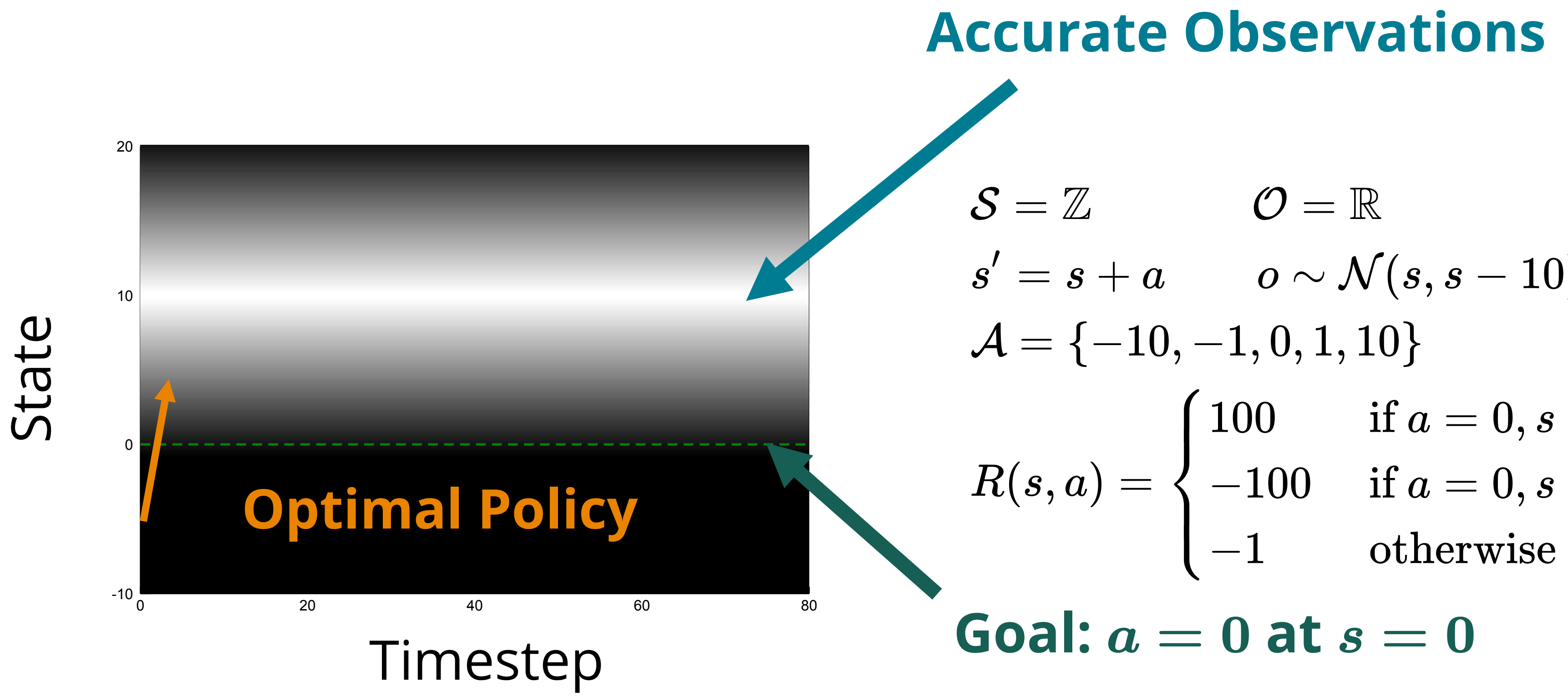
$$s' = s + a \qquad o \sim \mathcal{N}(s, s - 10)$$

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

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

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

POMDP Example: Light-Dark



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

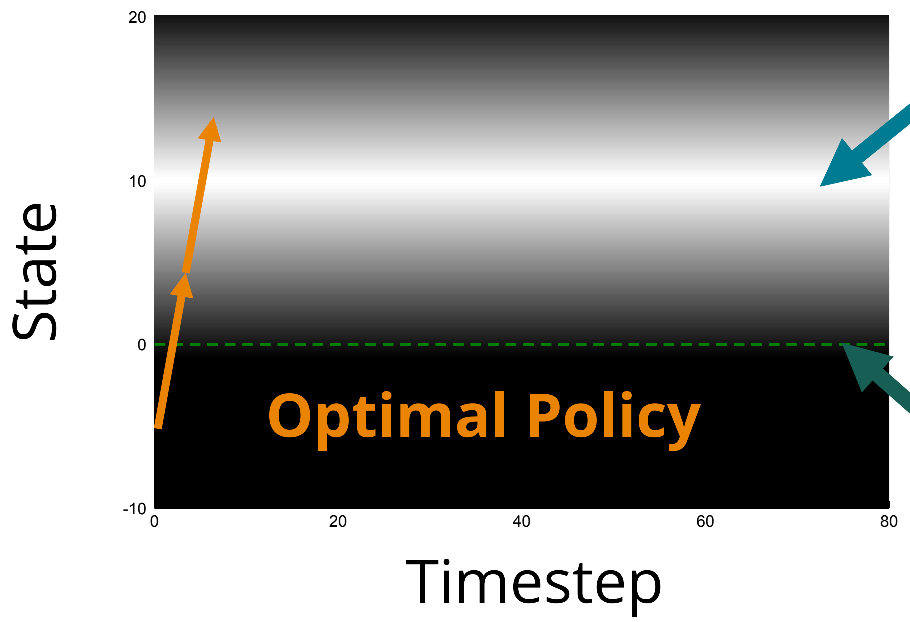
$$s' = s + a \qquad o \sim \mathcal{N}(s, s - 10)$$

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

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

POMDP Example: Light-Dark

Accurate Observations



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

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

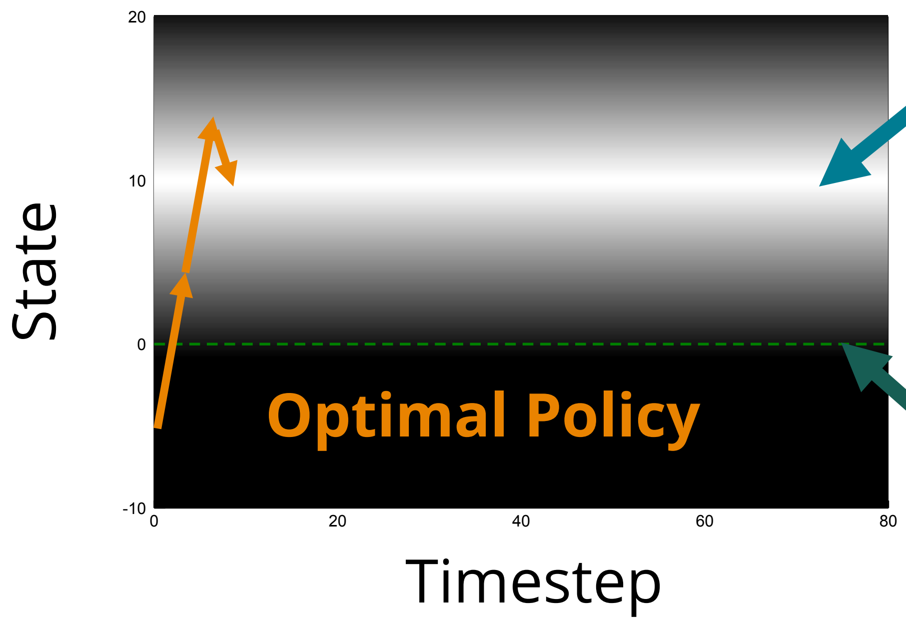
$$\mathcal{A} = \{-10, -1, 0, 1, 10\}$$

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

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

POMDP Example: Light-Dark

Accurate Observations



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

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

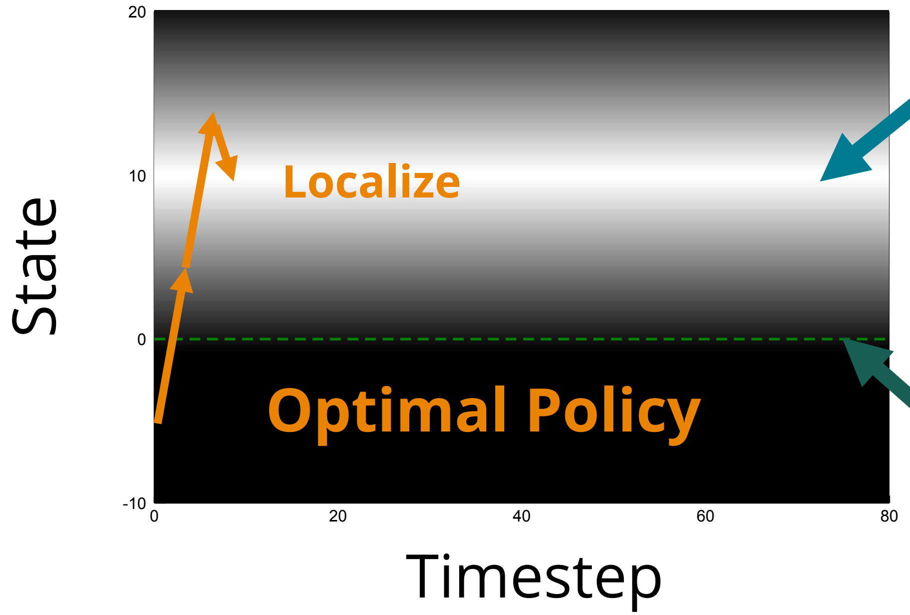
$$\mathcal{A} = \{-10, -1, 0, 1, 10\}$$

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

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

POMDP Example: Light-Dark

Accurate Observations



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

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

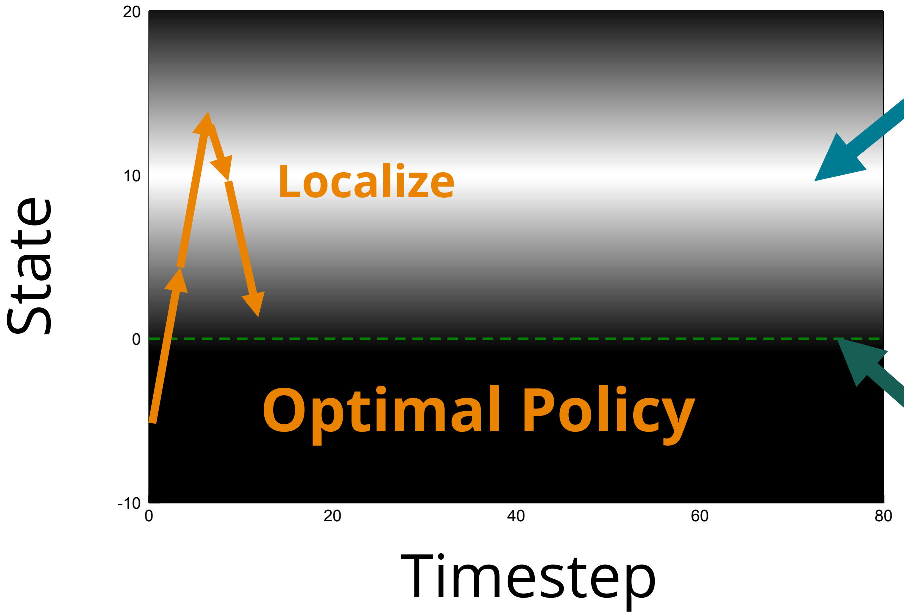
$$\mathcal{A} = \{-10, -1, 0, 1, 10\}$$

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

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

POMDP Example: Light-Dark

Accurate Observations



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

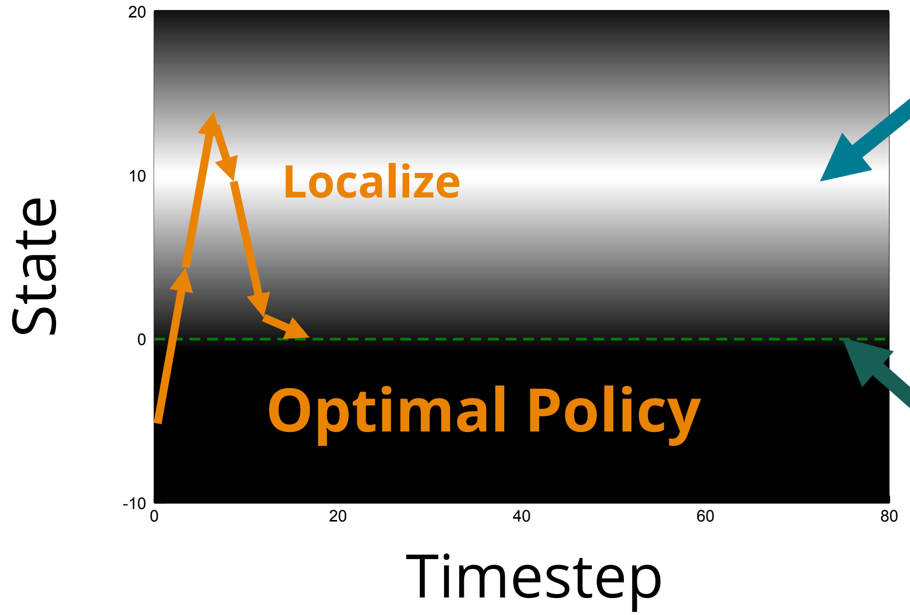
$$s' = s + a \qquad o \sim \mathcal{N}(s, s - 10)$$

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

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

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

POMDP Example: Light-Dark



Accurate Observations

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

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

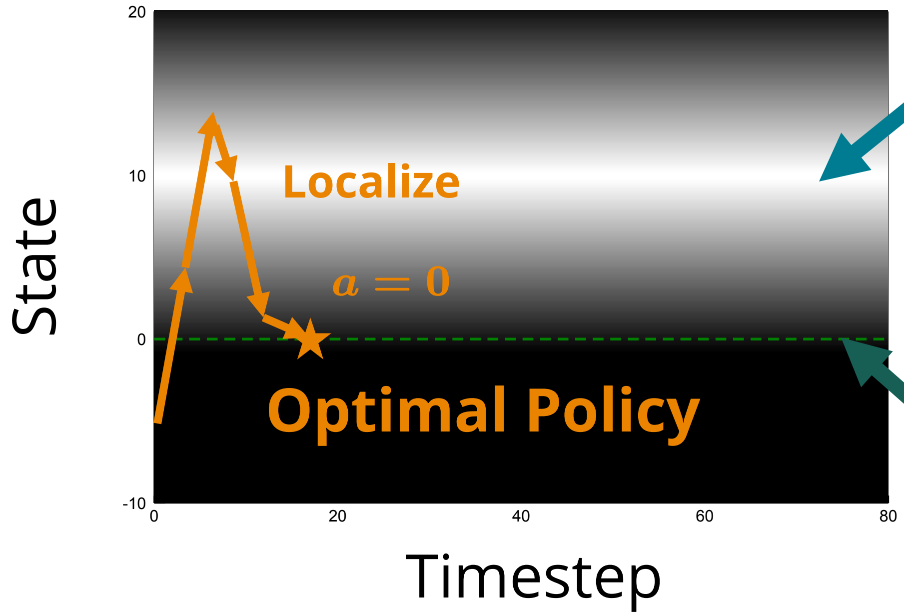
$$\mathcal{A} = \{-10, -1, 0, 1, 10\}$$

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

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

POMDP Example: Light-Dark

Accurate Observations



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

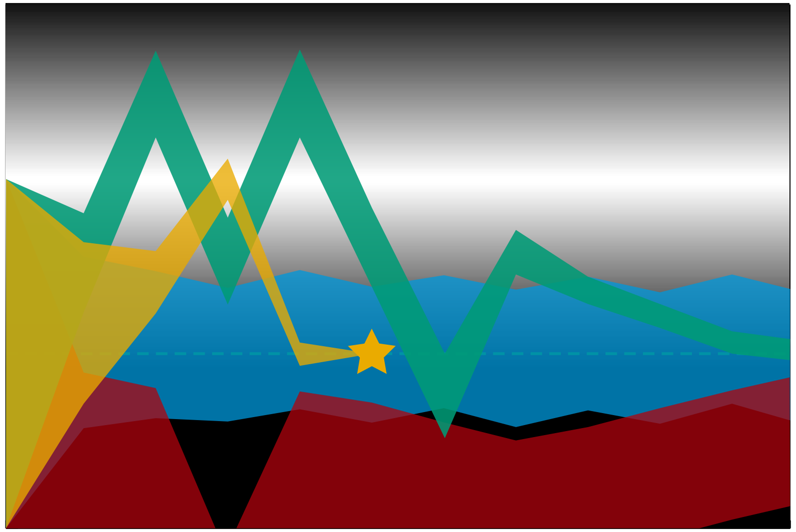
$$s' = s + a \qquad o \sim \mathcal{N}(s, s - 10)$$

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

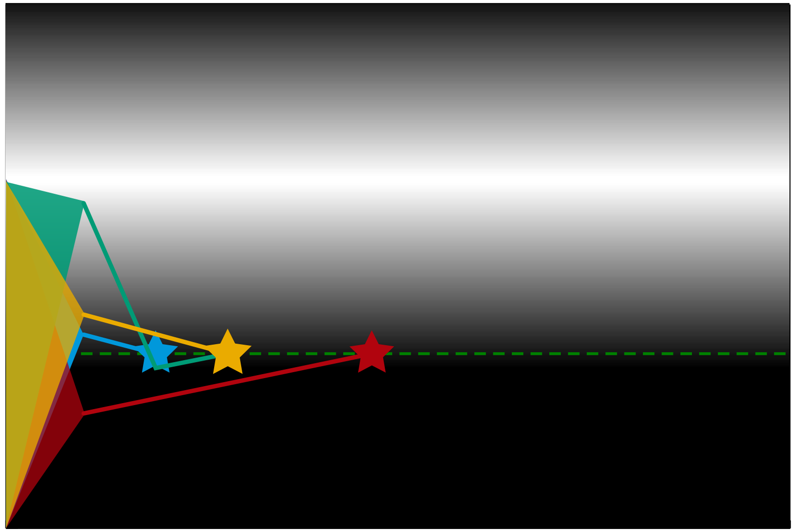
$$R(s,a) = \begin{cases} 100 & \text{if } a = 0, s = 0 \\ -100 & \text{if } a = 0, s \neq 0 \\ -1 & \text{otherwise} \end{cases}$$

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

POMDP Solution



QMDP

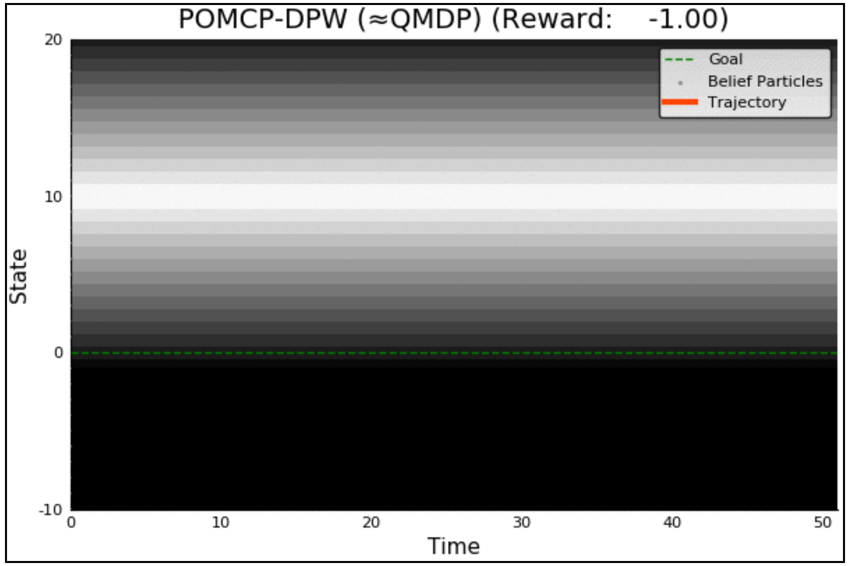


Same as **full observability** on the next step

Information Gathering

QMDP

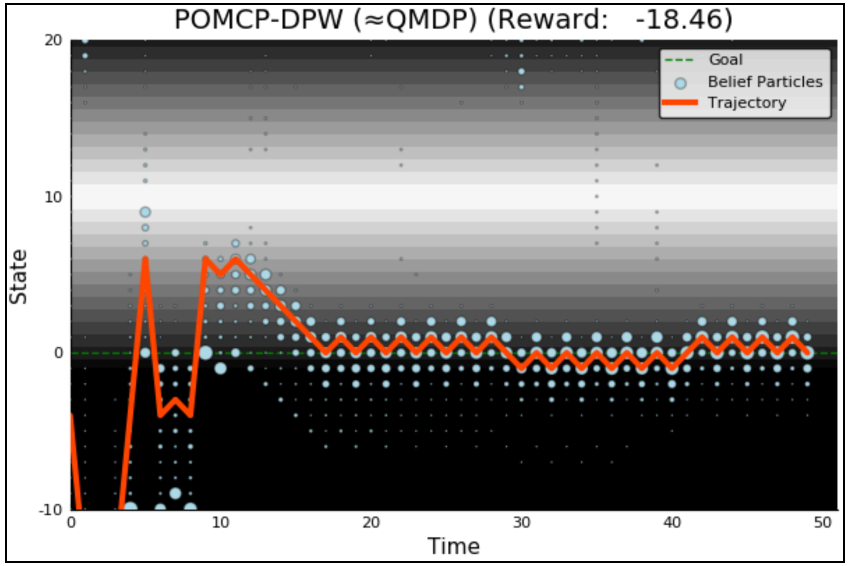
Full POMDP



Information Gathering

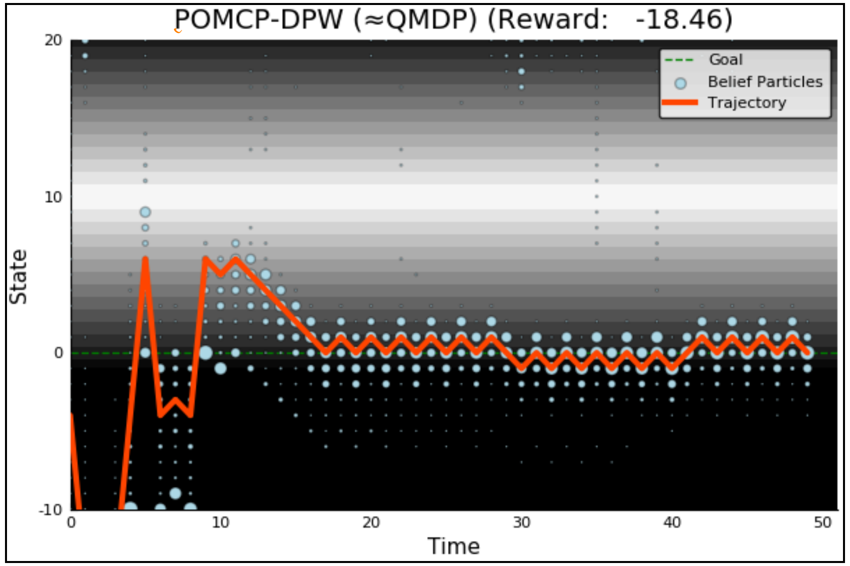
QMDP

Full POMDP

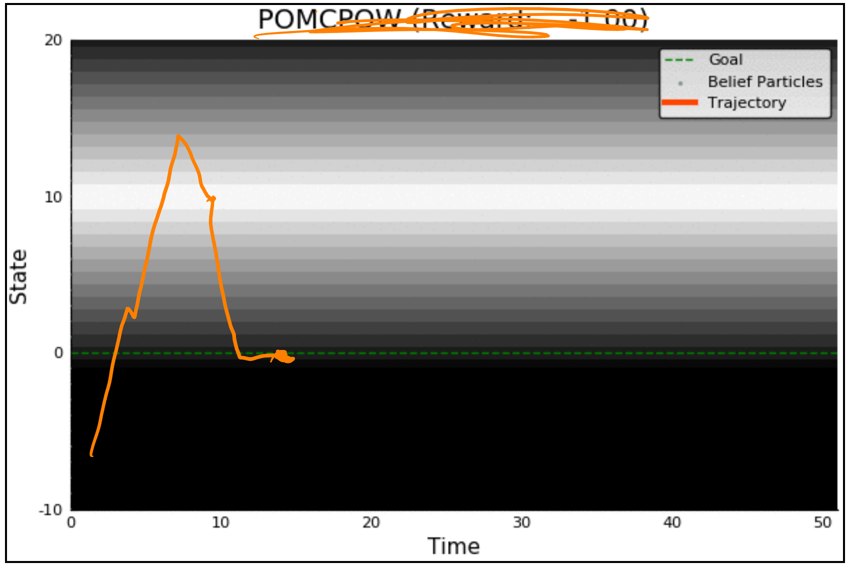


Information Gathering

QMDP

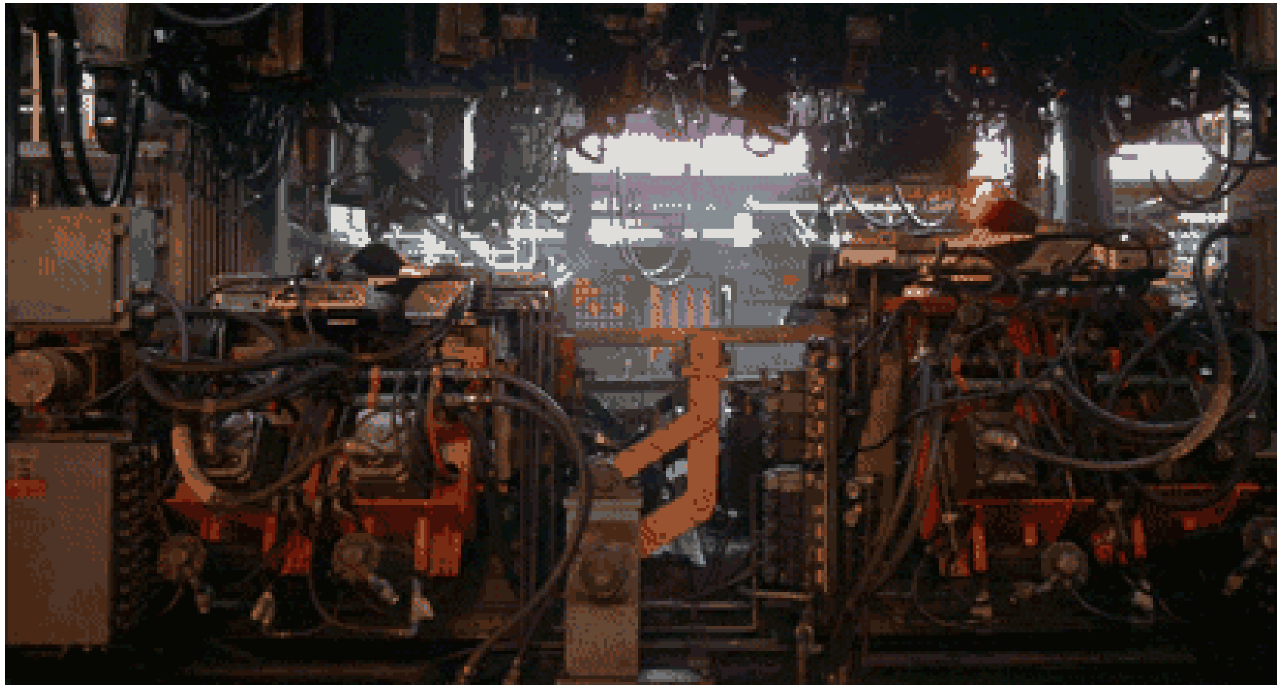


Full POMDP



QMDP

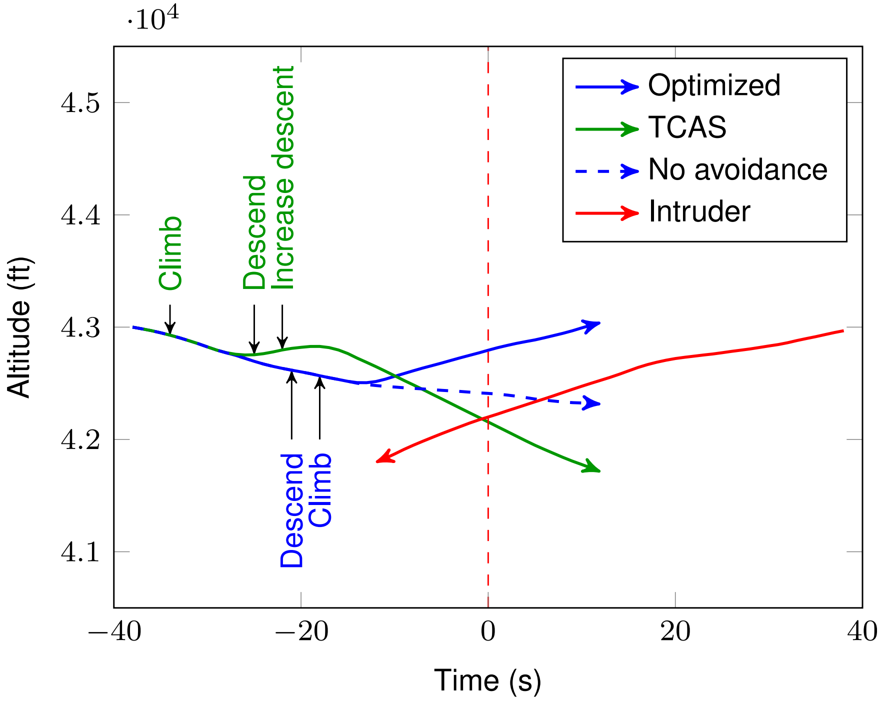
INDUSTRIAL GRADE



QMDP

ACAS X

[Kochenderfer, 2011]



Hindsight Optimization

POMDP Objective

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

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

pre-sample w_t^k

$$Q_{HS}(s,a) = \frac{1}{K}\sum_{k=1}^{K}\max_{a_{1:T}}\sum_{t=0}^{T} R(s_t, a_t)$$

s.t. $s_{t+1} = G(s_t, a_t, w_t^k)$

$a_0 = a$

$$\pi_{HS} = \underset{a}{\text{argmax}}\ \underset{s\sim b}{\mathbb{E}}\left[Q_{HS}(s,a)\right]$$

FIB

POMDP Objective

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

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

FIB

loop

$$\alpha_a[s] \leftarrow R(s,a) + \gamma \sum_o \max_{a'} \sum_{s'} T(s' | s,a) Z(o | a, s') \alpha_{a'}[s']$$

Compared to QMDP

loop

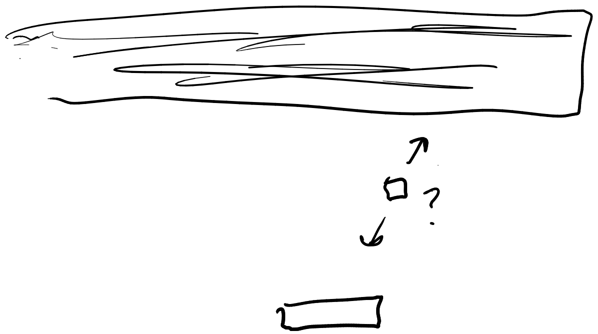
$$\alpha_a[s] \leftarrow R(s,a) + \gamma \sum_{s'} T(s' | s,a) \max_{a'} \alpha_{a'}[s']$$

k-Markov

POMDP Objective

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

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



solve and MDP where

$s_t = [o_t, o_{t-1}, \ldots o_{t-(k-1)}]$

$k=4$ -markov approx

s_t = last 4 observations

Open Loop

POMDP Objective

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

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