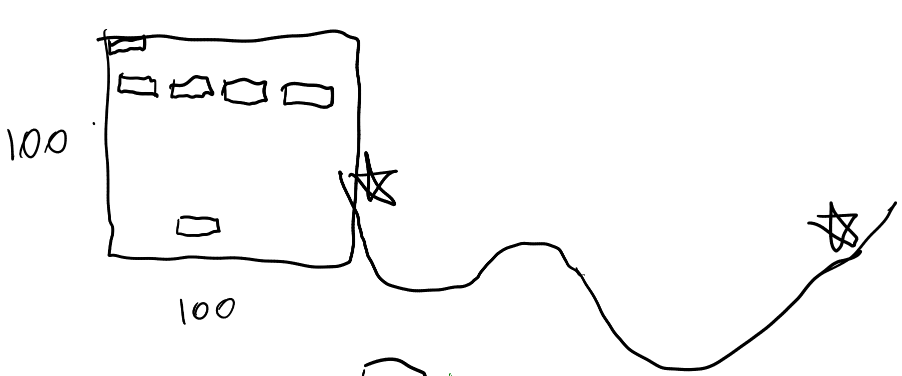
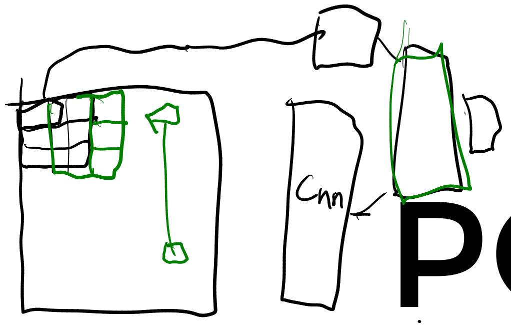


$$Q' \leftarrow Q$$

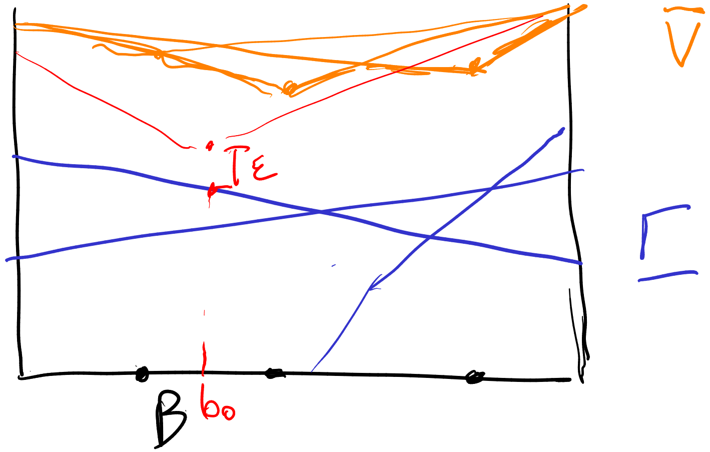
$$Q' \leftarrow (1-\tau)Q' + \tau Q$$

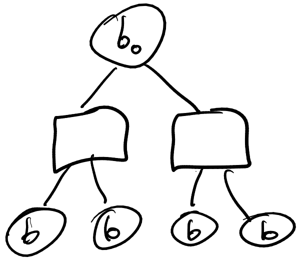


POMDP Formulation Approximations

Convolutional
Neural
Network

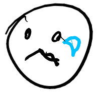
SARSOP
~100,000





POMDP Computational Complexity

POMDP Computational Complexity

Sad facts ● 

POMDP Computational Complexity

Sad facts

- Infinite horizon POMDPs are *undecidable*

POMDP Computational Complexity

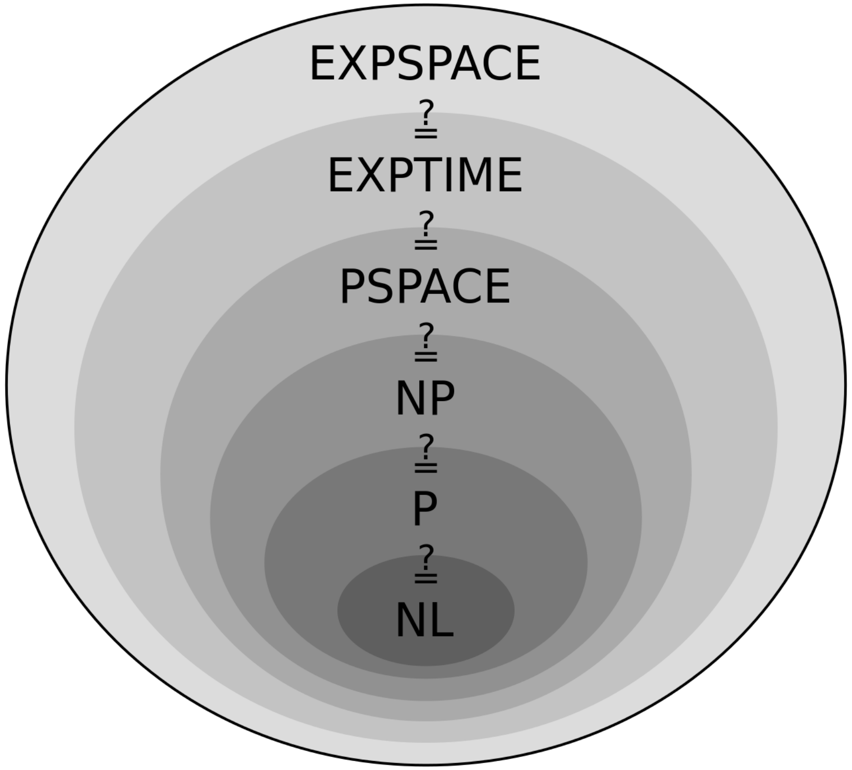
Sad facts

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

POMDP Computational Complexity

Sad facts ●

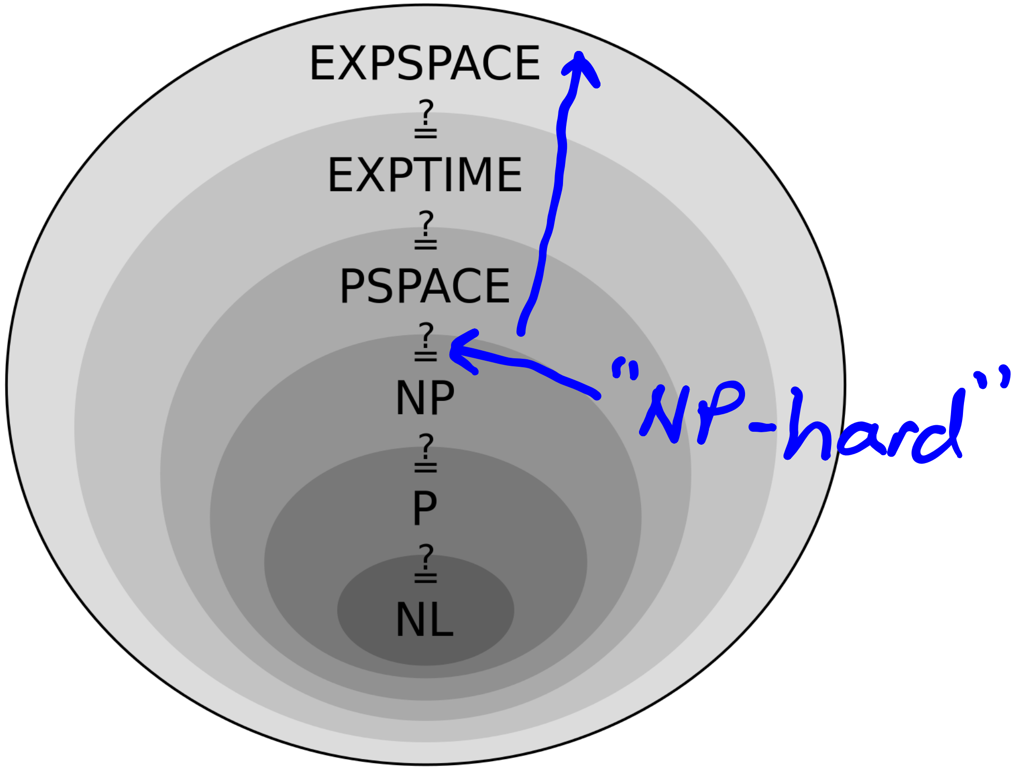
- Infinite horizon POMDPs are *undecidable*
- Finite horizon POMDPs are *PSPACE Complete*
  - Among the hardest problems that can be solved using a polynomial amount of space



POMDP Computational Complexity

Sad facts ●

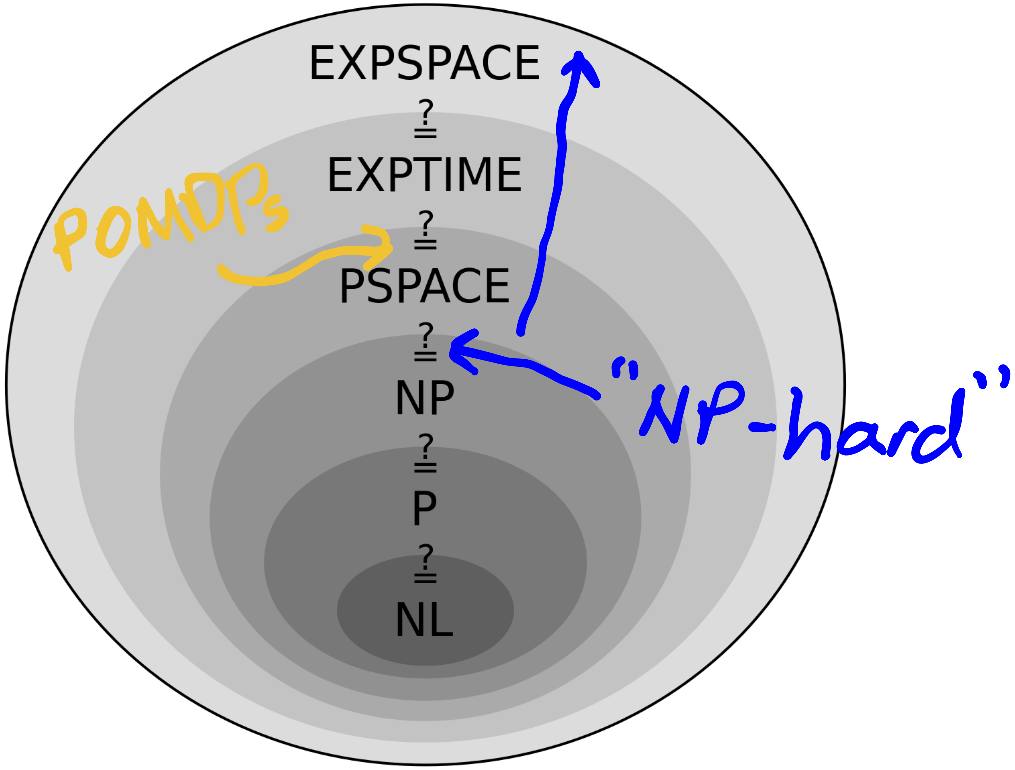
- Infinite horizon POMDPs are *undecidable*
- Finite horizon POMDPs are *PSPACE Complete*
  - Among the hardest problems that can be solved using a polynomial amount of space



POMDP Computational Complexity

Sad facts ●

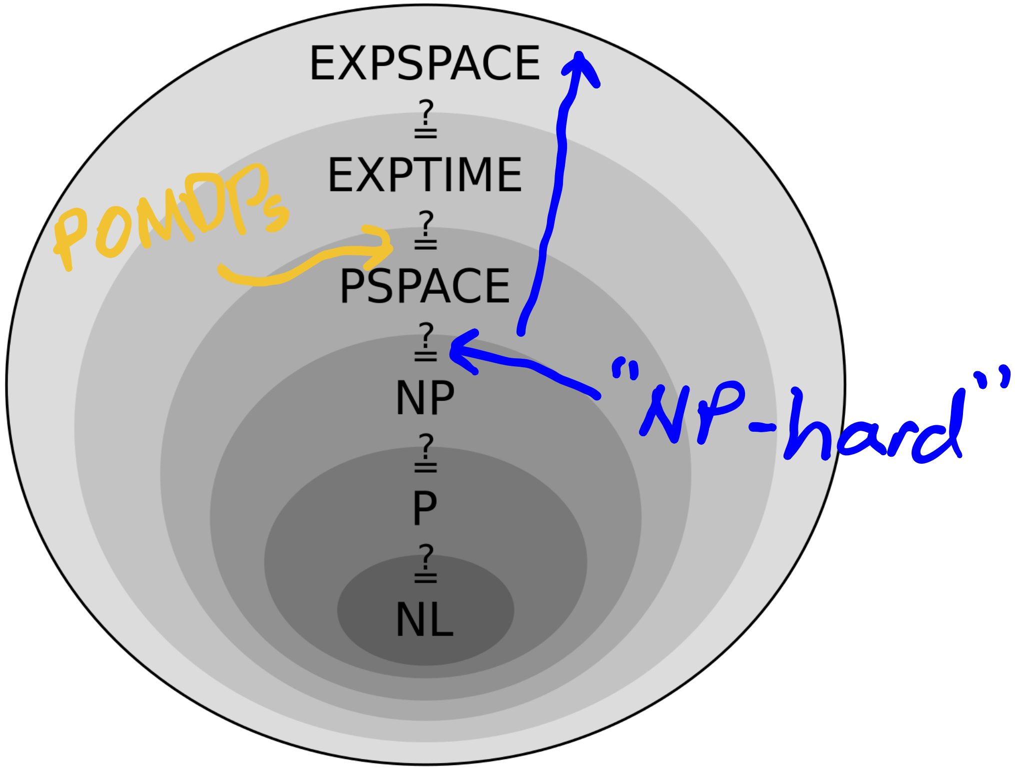
- Infinite horizon POMDPs are *undecidable*
- Finite horizon POMDPs are *PSPACE Complete*
  - Among the hardest problems that can be solved using a polynomial amount of space



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

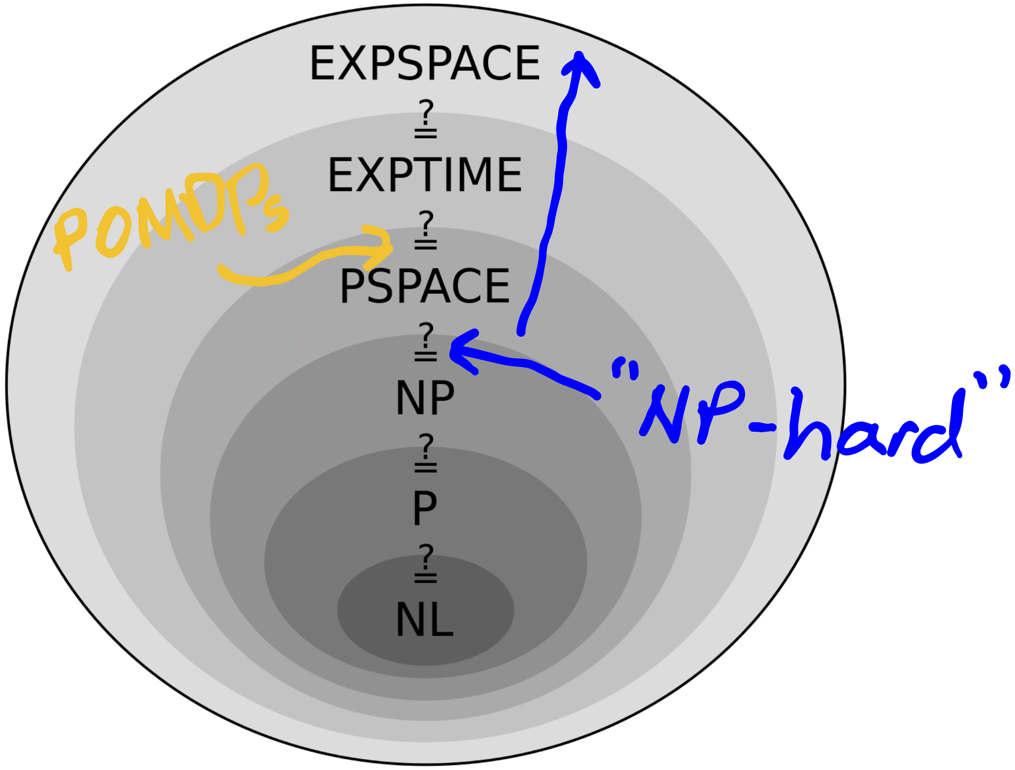


POMDP Computational Complexity

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!

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

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

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

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 $\underset{s \in S}{\operatorname{argmax}}\, b(s)$

mode

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

mean

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

Certainty Equivalent

Certainty Equivalent

MDP analogy
LQR

Optimal for LQG

Linear
Quadratic
Gaussian

Linear Quadratic Gaussian

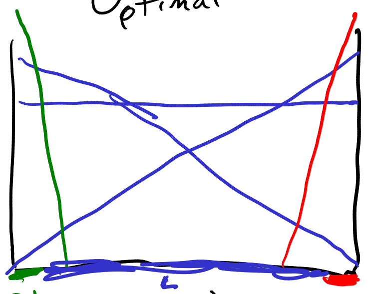
$s \in \mathbb{R}^n$
$a \in \mathbb{R}^m$

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

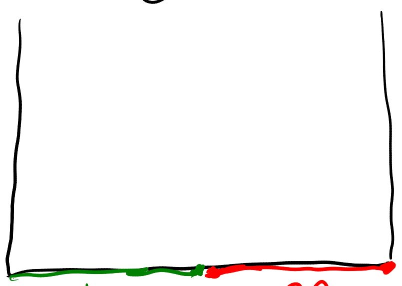
$$O(\mathbf{o} \mid \mathbf{s}') = \mathcal{N}(\mathbf{o} \mid \mathbf{O}_s \mathbf{s}', \mathbf{\Sigma}_o)$$

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

Optimal

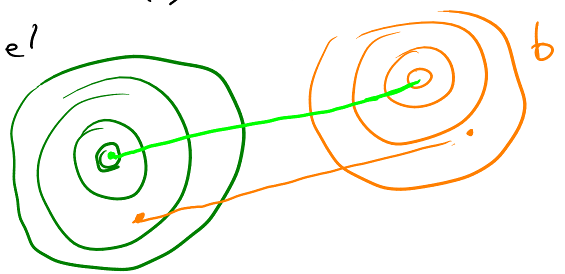


CE

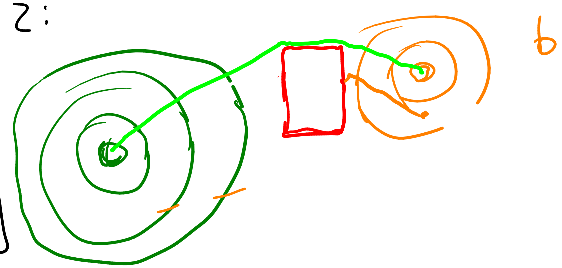


When is CE Good

Case 1



Case 2:



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

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

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

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

b'

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

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

Kalman Filter (belief update)

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

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

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

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

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

Given $(S, A, O, T, Z, R, \gamma)$

(S, A, T, R, γ)

$\downarrow$

Q^*

QMDP

POMDP Objective

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

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

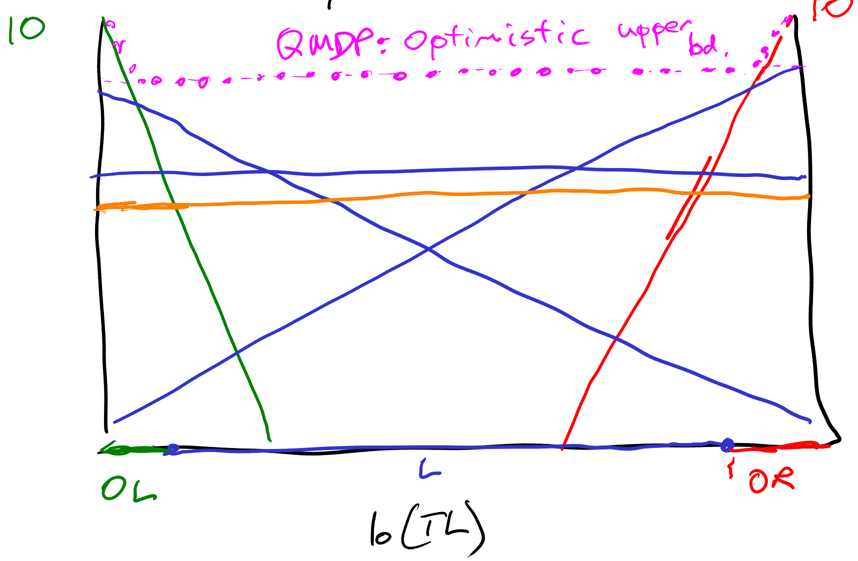
$$\pi_{\text{QMDP}}(b) = \underset{a\in A}{\operatorname{argmax}} \underset{s\sim b}{\operatorname{E}} \left[\underline{Q_{\text{MDP}}(s, a)}\right]$$

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

Example: Tiger POMDP with Waiting

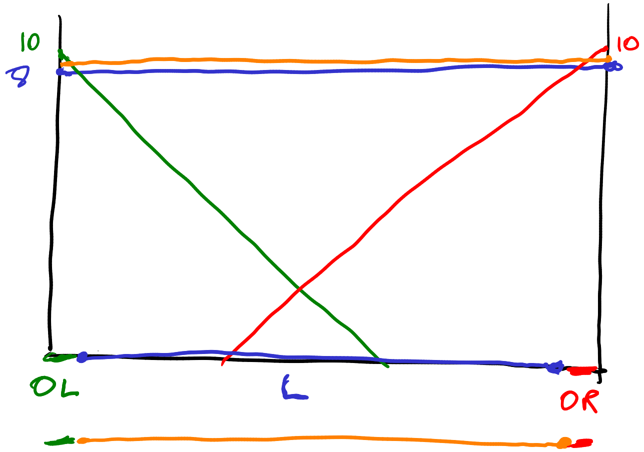
Terminates when door is open, $\gamma = 0.9$

Optimal (from SARSOP)



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

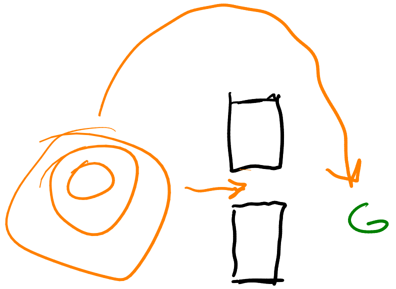
$\underset{s \sim b}{E}\left[Q_{MDP}(s,a)\right] = b(TL)\,Q_{MDP}(TL,a) + (1-b(TL))\,Q_{MDP}(TR,a)$



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

Is QMDP good for tiger?

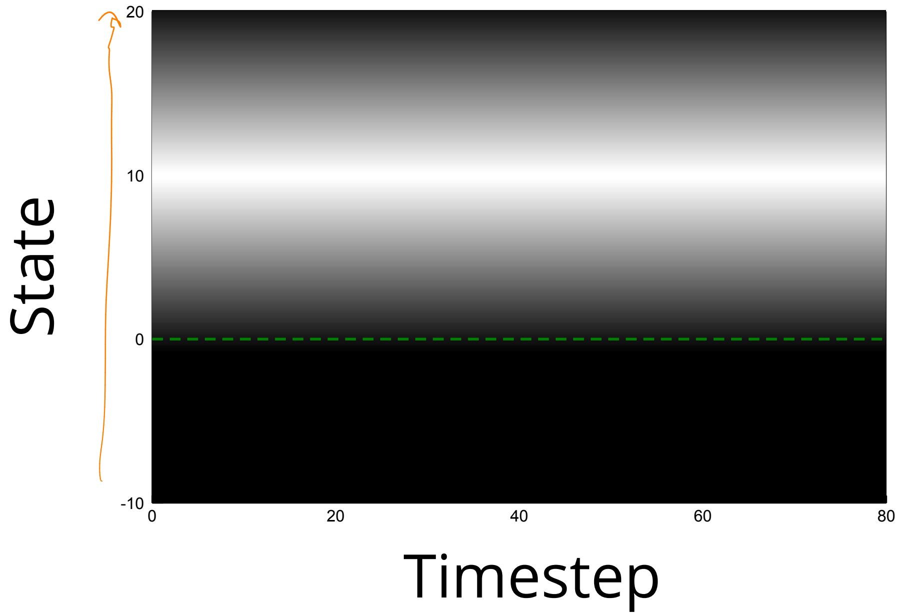
Wait
- 50-50 observation
- 0 immediate reward



QMDP is bad at costly information gathering + long-lasting uncertainty

o.w. QMDP is pretty good + Much easier to solve

POMDP Example: Light-Dark



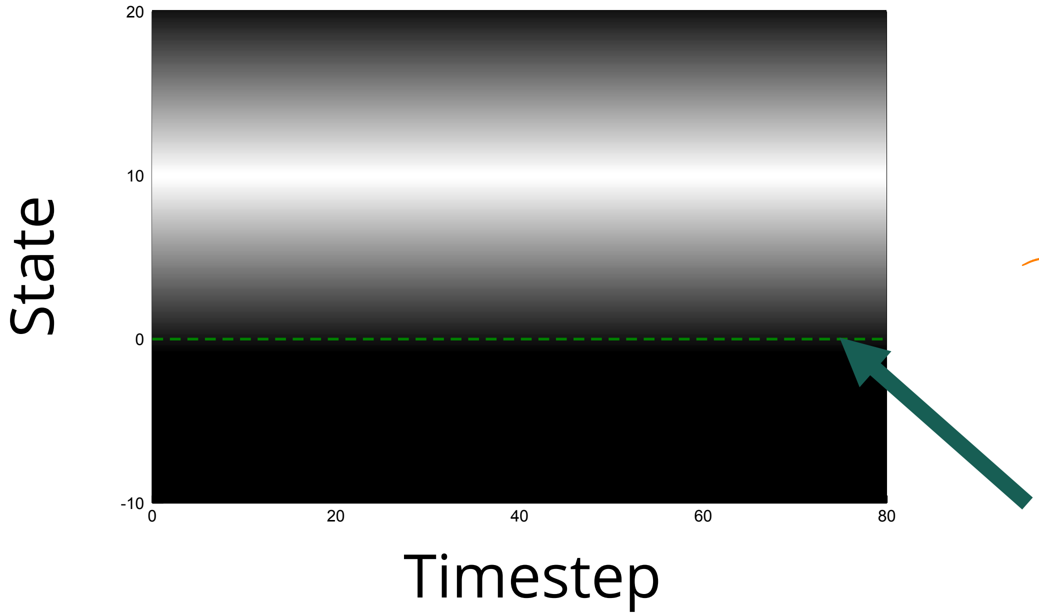
$$\mathcal{S} = \mathbb{Z} \qquad \mathcal{O} = \mathbb{R}$$

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

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

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

POMDP Example: Light-Dark



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

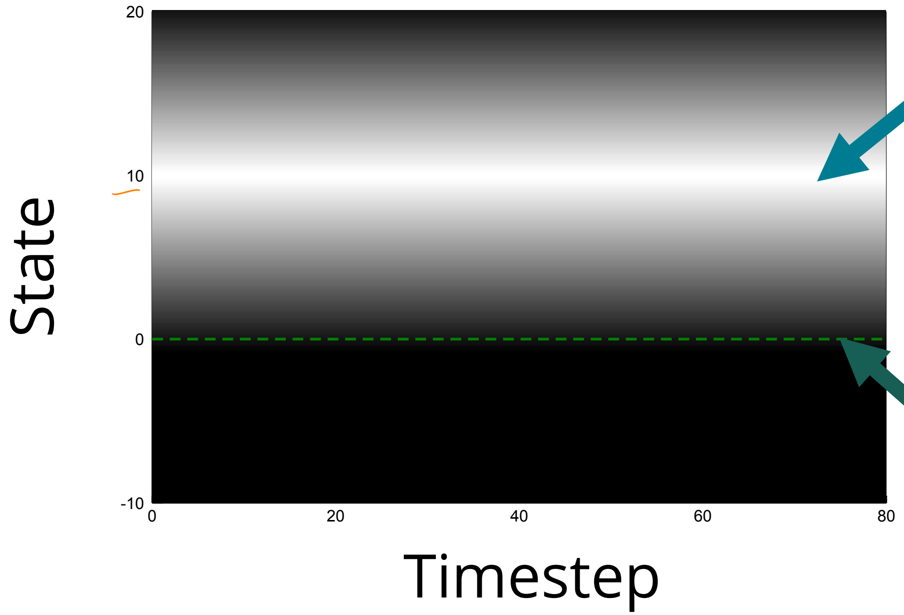
$$s' = s + a \qquad o \sim \mathcal{N}(s, s - 10)$$

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

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

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

POMDP Example: Light-Dark



Accurate Observations

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

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

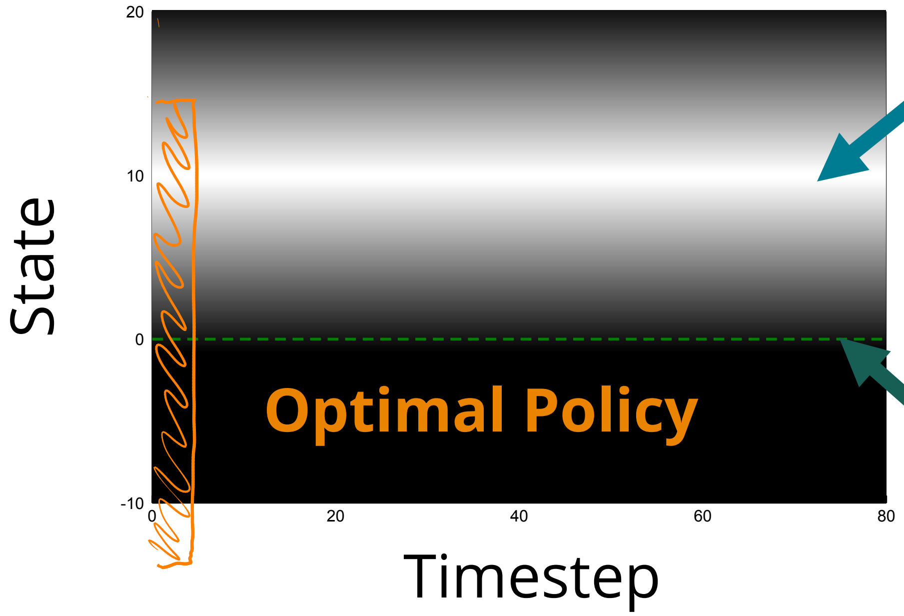
$$\mathcal{A} = \{-10, -1, 0, 1, 10\}$$

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

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

POMDP Example: Light-Dark

Accurate Observations



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

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

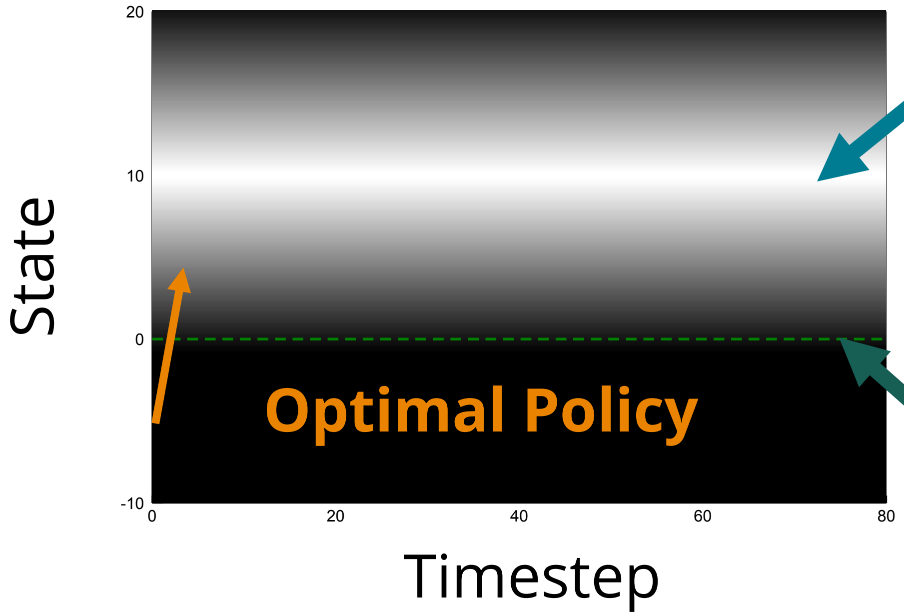
$$\mathcal{A} = \{-10, -1, 0, 1, 10\}$$

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

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

POMDP Example: Light-Dark

Accurate Observations



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

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

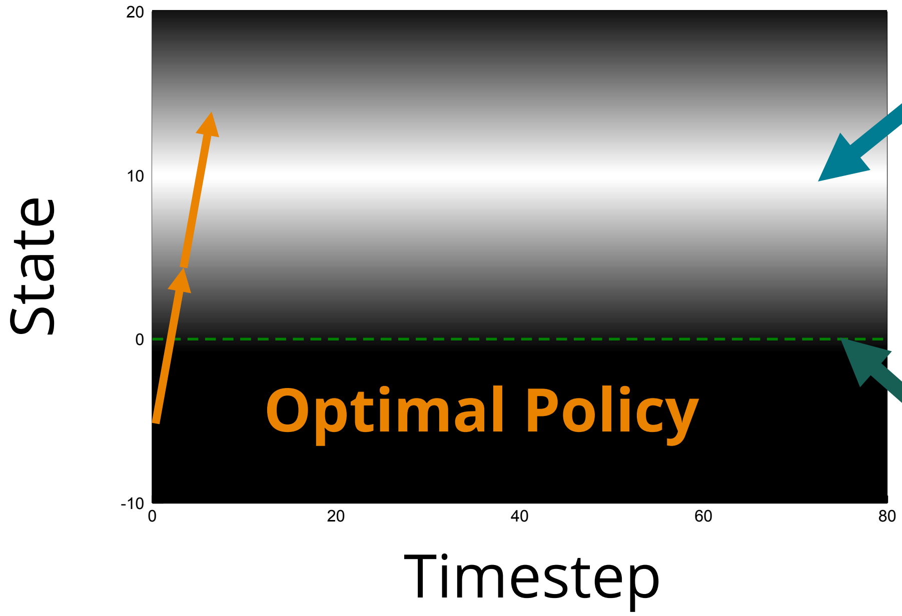
$$\mathcal{A} = \{-10, -1, 0, 1, 10\}$$

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

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

POMDP Example: Light-Dark

Accurate Observations



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

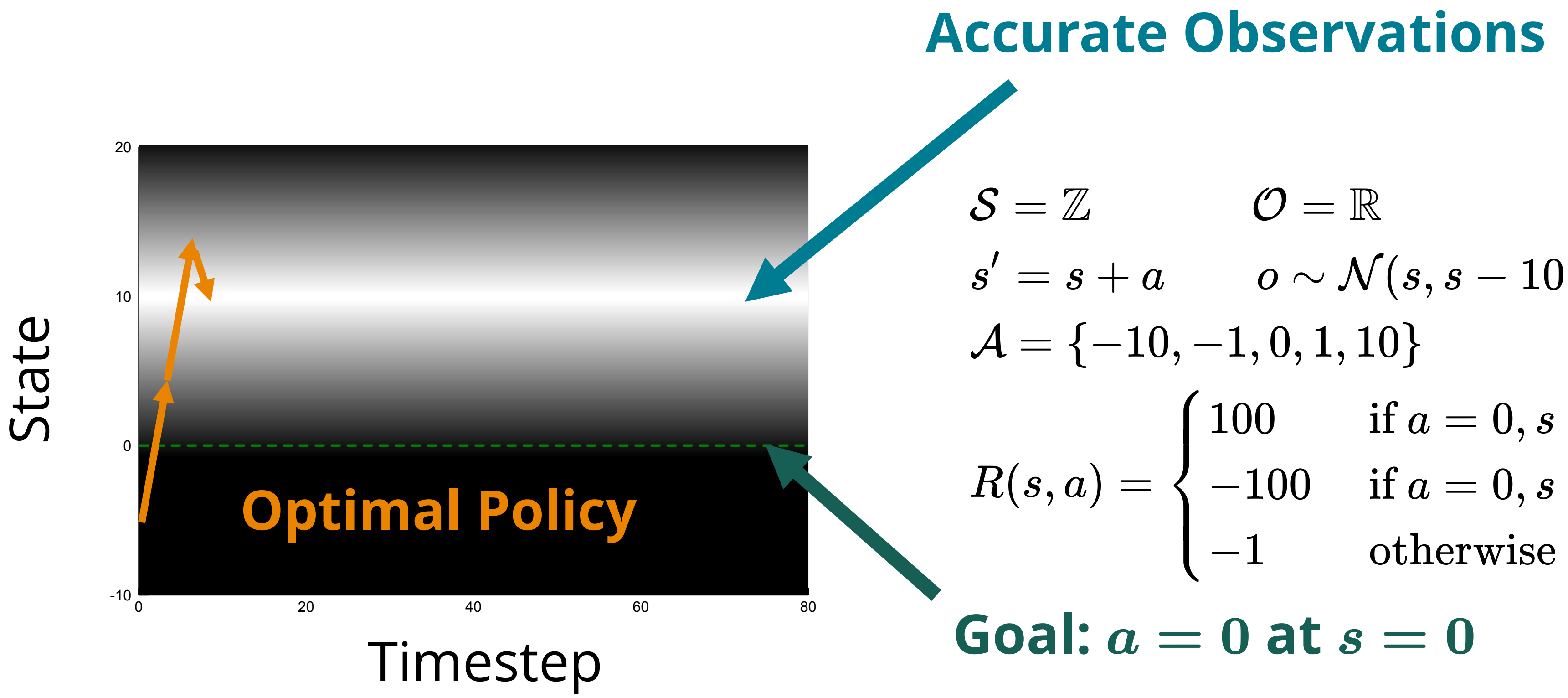
$$s' = s + a \qquad o \sim \mathcal{N}(s, s - 10)$$

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

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

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

POMDP Example: Light-Dark



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

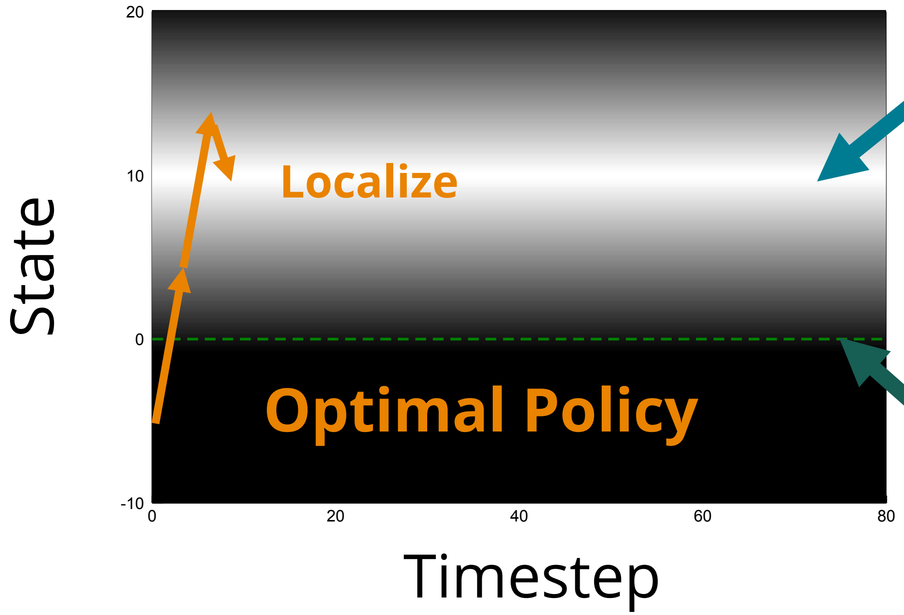
$$s' = s + a \qquad o \sim \mathcal{N}(s, s - 10)$$

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

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

POMDP Example: Light-Dark

Accurate Observations



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

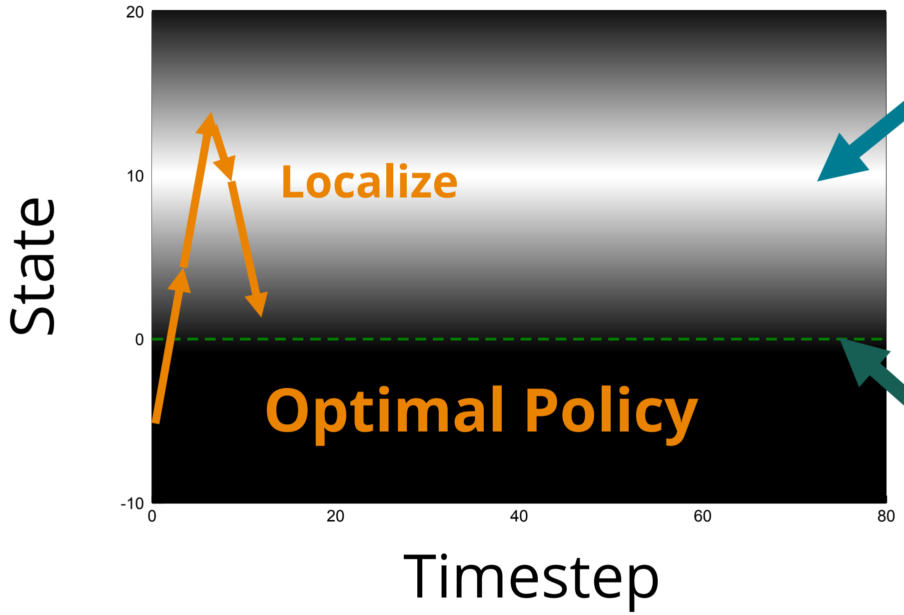
$$s' = s + a \qquad o \sim \mathcal{N}(s, s-10)$$

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

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

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

POMDP Example: Light-Dark



Accurate Observations

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

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

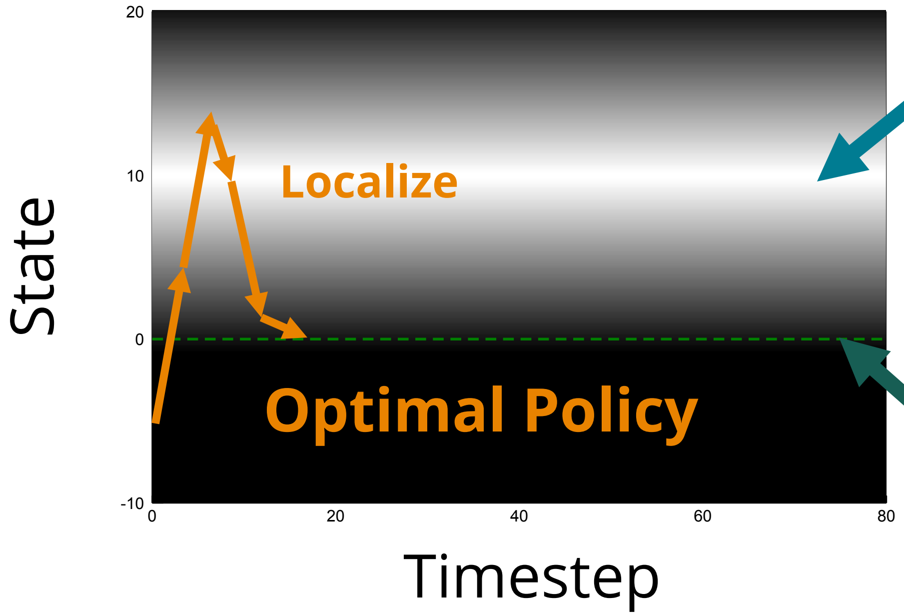
$$\mathcal{A} = \{-10, -1, 0, 1, 10\}$$

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

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

POMDP Example: Light-Dark

Accurate Observations



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

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

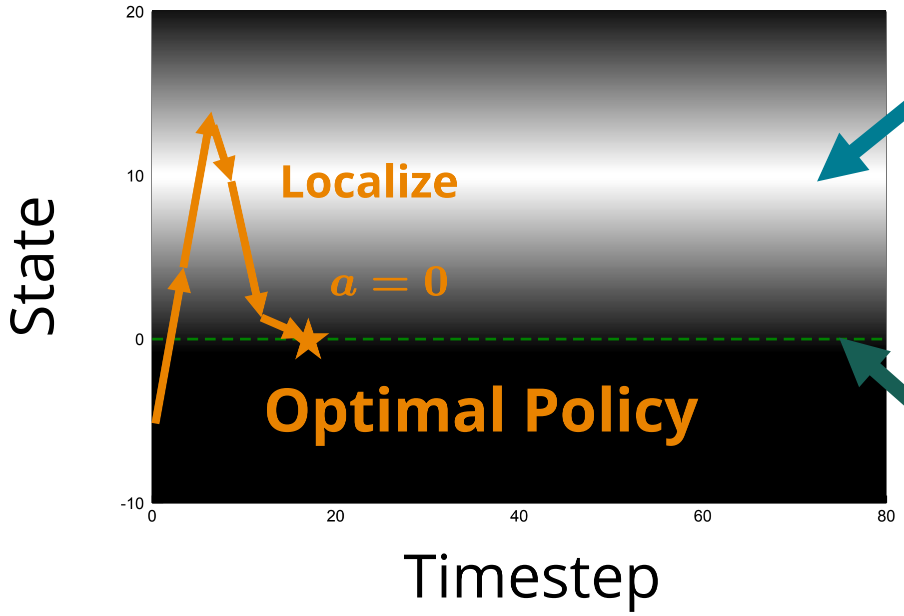
$$\mathcal{A} = \{-10, -1, 0, 1, 10\}$$

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

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

POMDP Example: Light-Dark

Accurate Observations



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

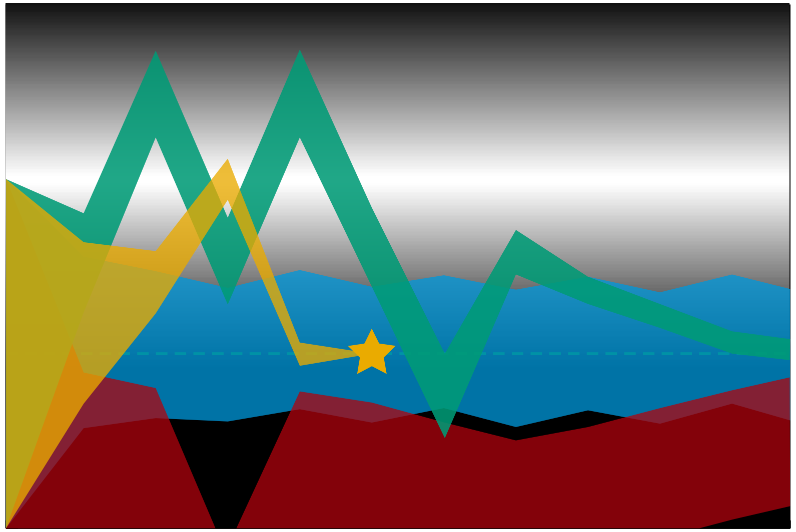
$$s' = s + a \qquad o \sim \mathcal{N}(s, s - 10)$$

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

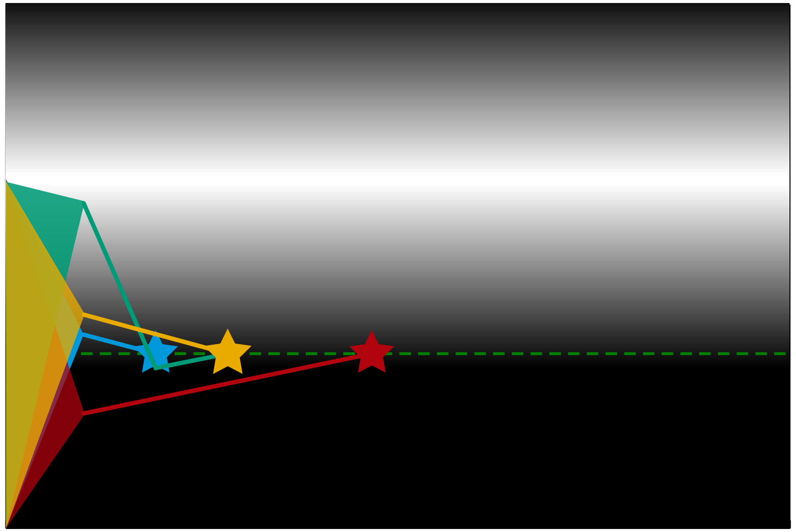
$$R(s,a) = \begin{cases} 100 & \text{if } a = 0, s = 0 \\ -100 & \text{if } a = 0, s \neq 0 \\ -1 & \text{otherwise} \end{cases}$$

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

POMDP Solution



QMDP

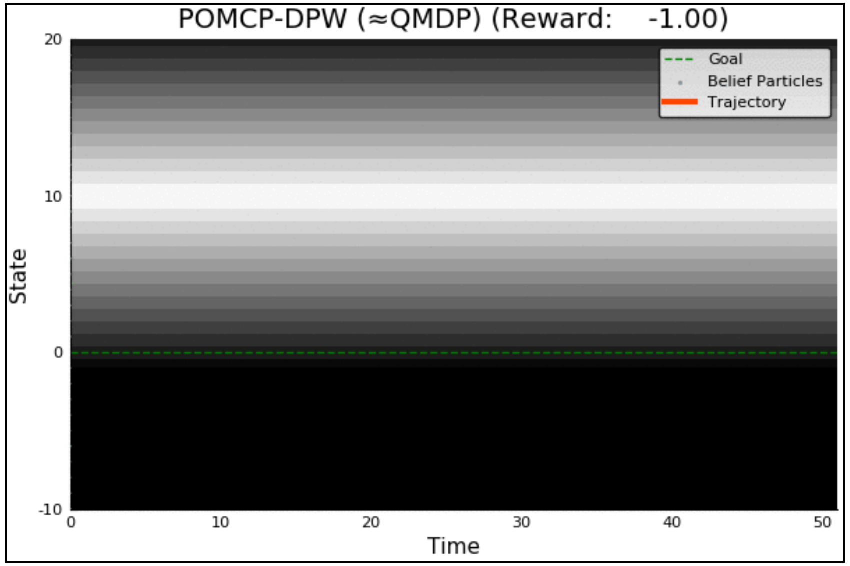


Same as **full observability** on the next step

Information Gathering

QMDP

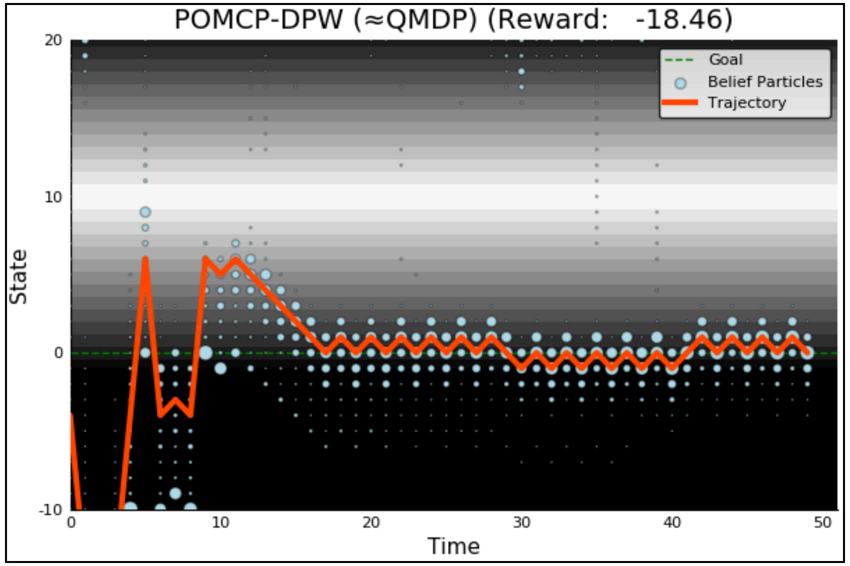
Full POMDP



Information Gathering

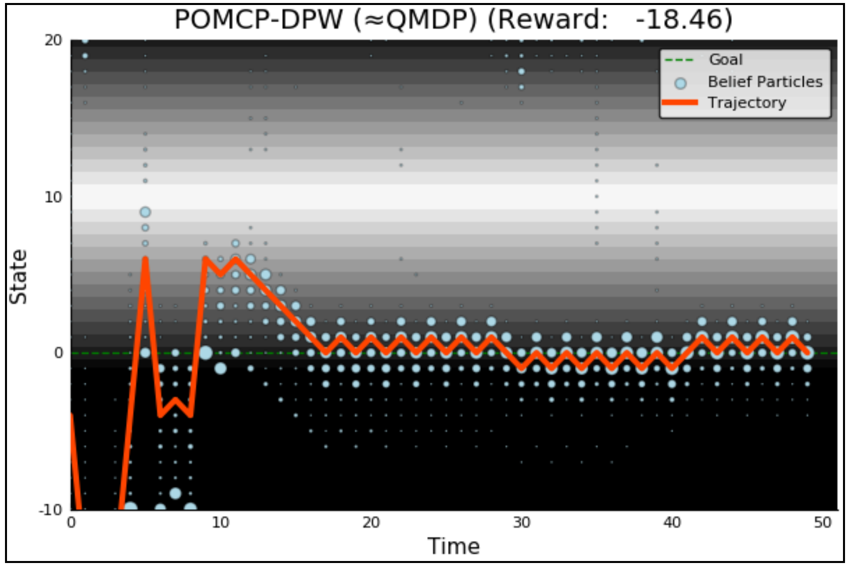
QMDP

Full POMDP

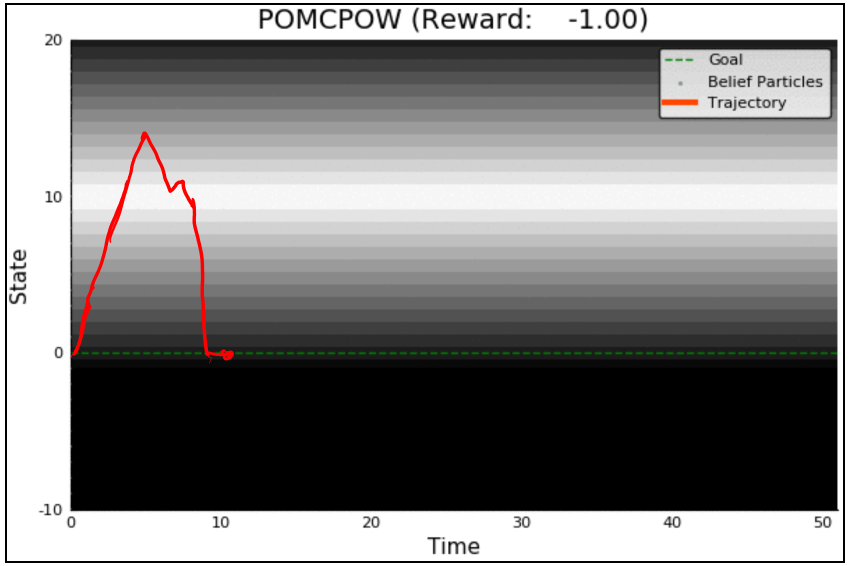


Information Gathering

QMDP

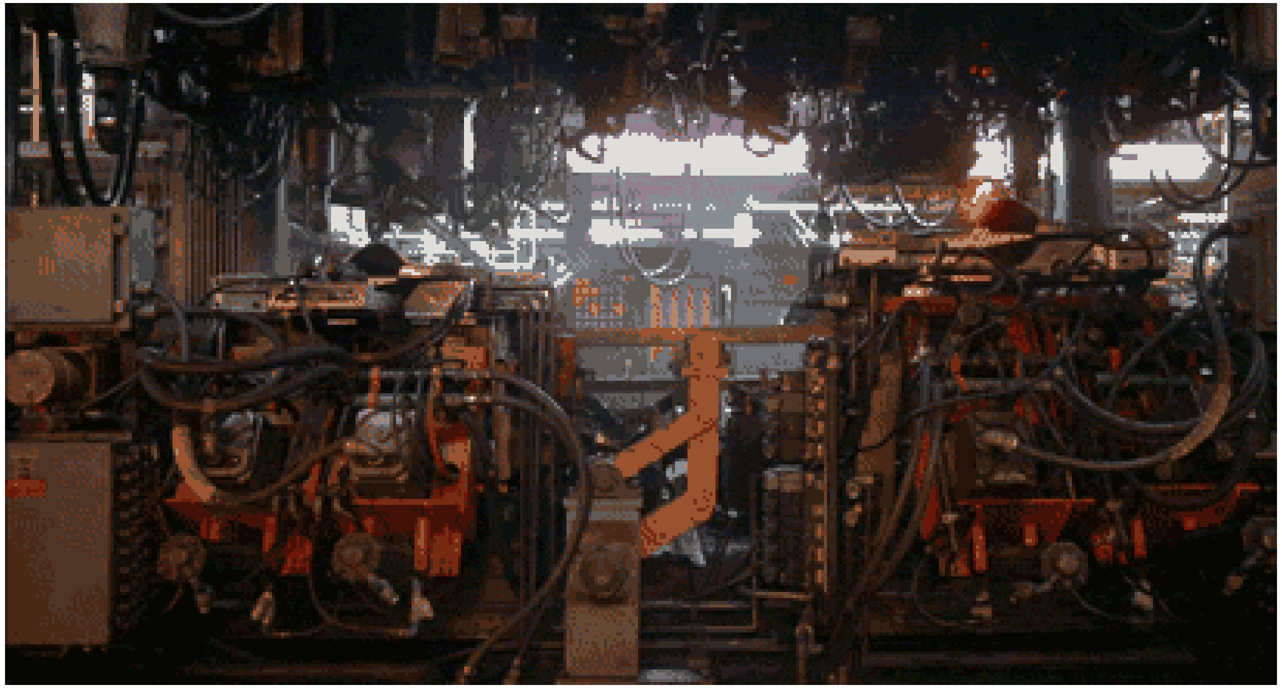


Full POMDP



QMDP

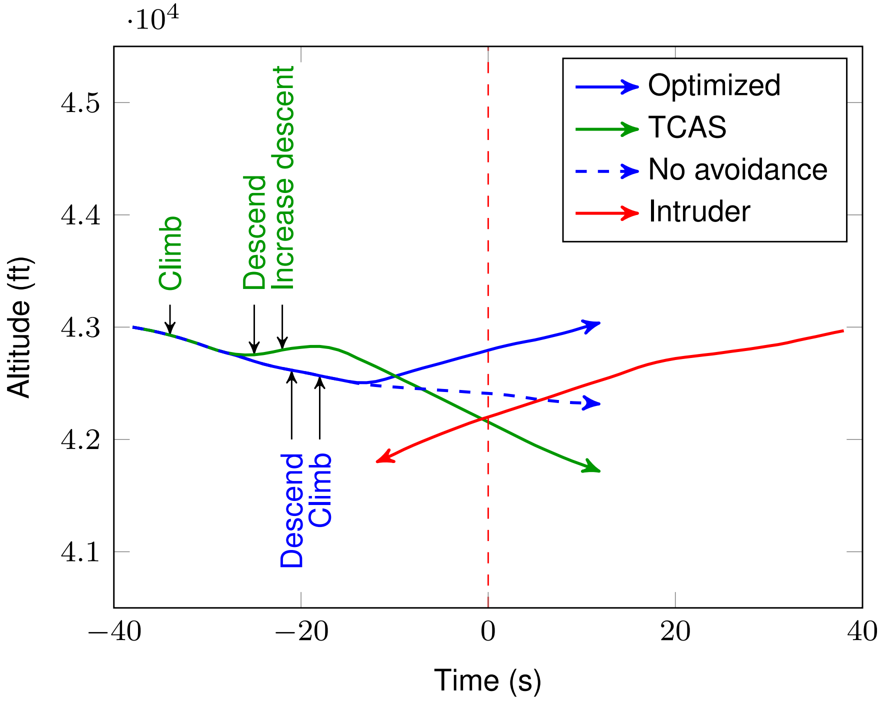
INDUSTRIAL GRADE



QMDP

ACAS X

[Kochenderfer, 2011]



Hindsight Optimization

POMDP Objective

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

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

K scenarios

pre-sample w_t^k

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

subject to $\quad s_{t+1} = G(s_t, a_t, w_t^k)$

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

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

FIB

POMDP Objective

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

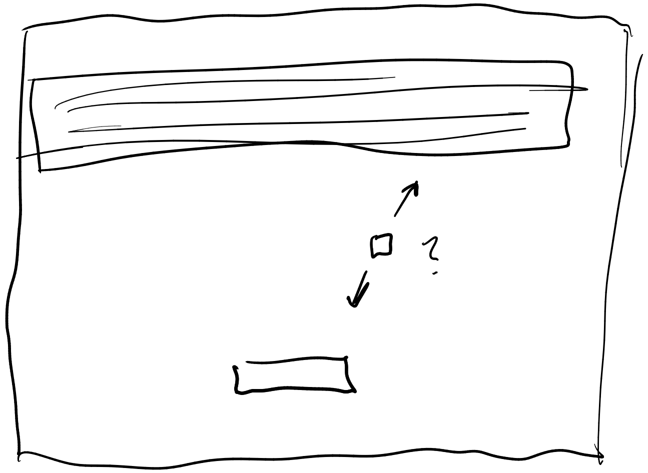
$$b' = \tau(b, a, o)$$

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 an MDP where state is

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

Open Loop

POMDP Objective

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

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