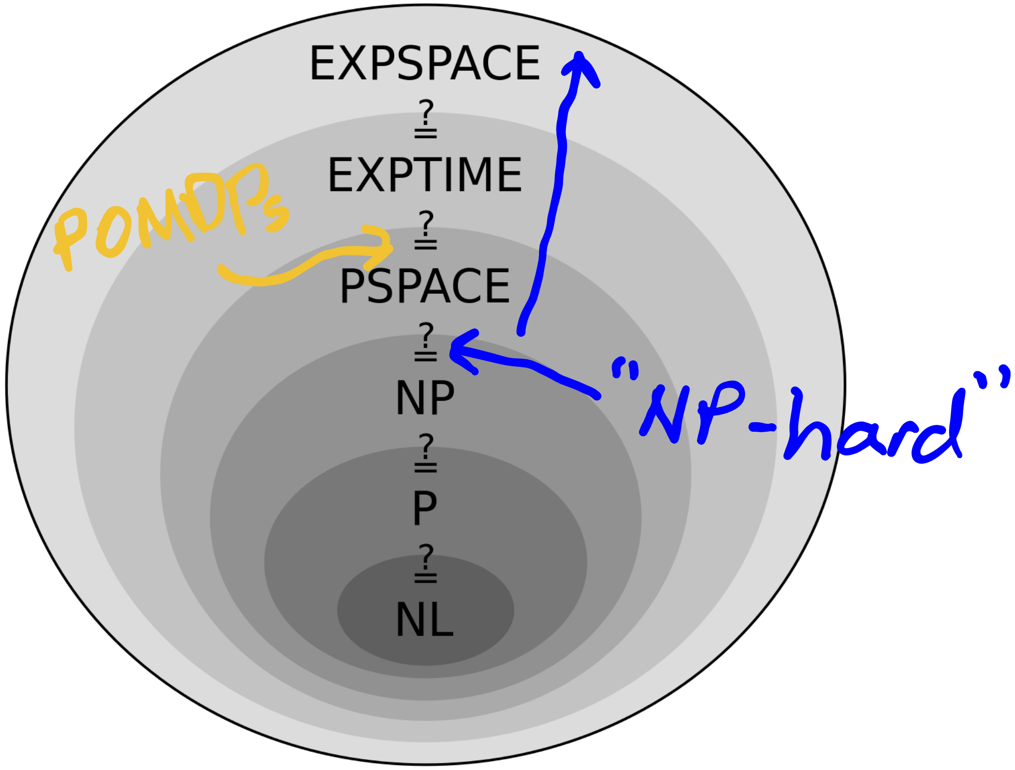
POMDP Formulation Approximations

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

Numerical Approximations

(approximately solve original problem)

Offline

Last week

Online

Thursday

Formulation Approximations

(solve a slightly different problem)

Today!

POMDP Objective

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

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

Certainty Equivalent

POMDP Objective

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

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

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

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

Certainty Equivalent

Optimal for LQG

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

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

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

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

QMDP

POMDP Objective

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

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

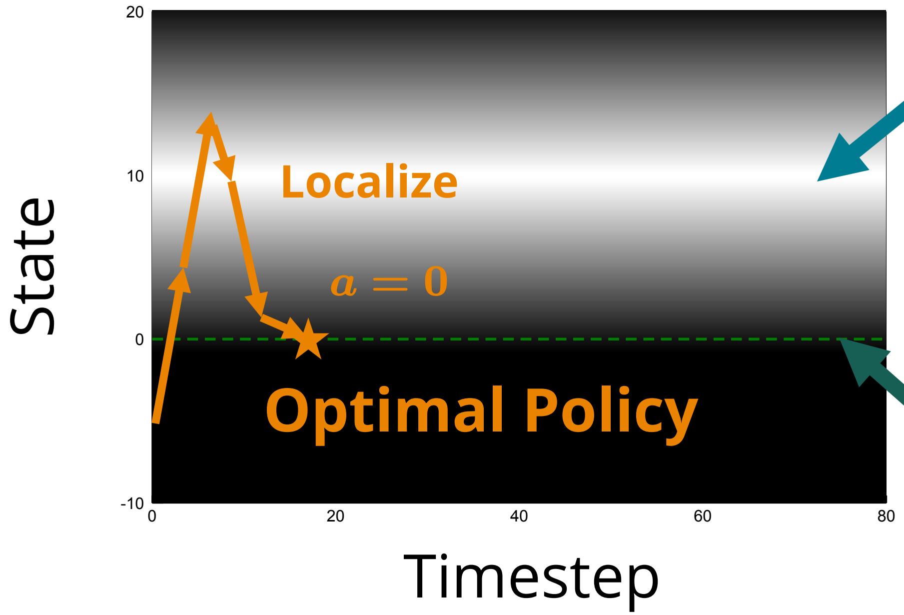
$$\pi_{\mathrm{QMDP}}(b) = \underset{a\in A}{\operatorname{argmax}} \; \underset{s\sim b}{\mathrm{E}}\left[Q_{\mathrm{MDP}}(s, a)\right]$$

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

Example: Tiger POMDP with Waiting

POMDP Example: Light-Dark

Accurate Observations



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

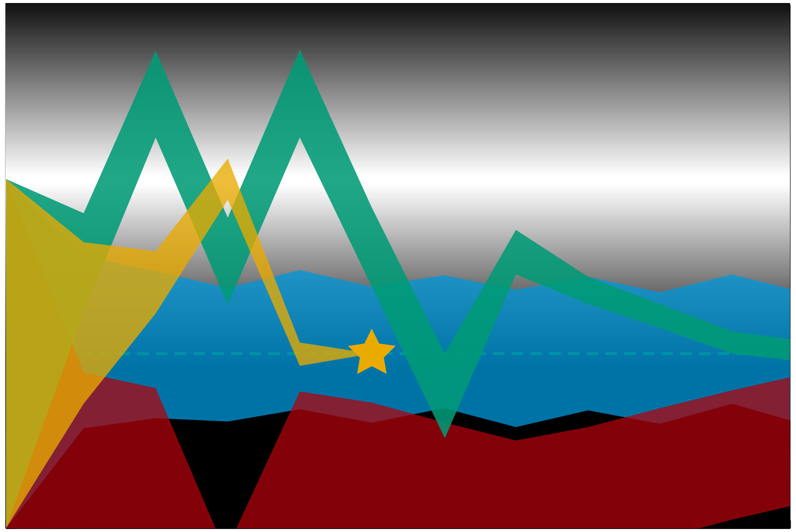
$$s' = s + a \qquad o \sim \mathcal{N}(s, s - 10)$$

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

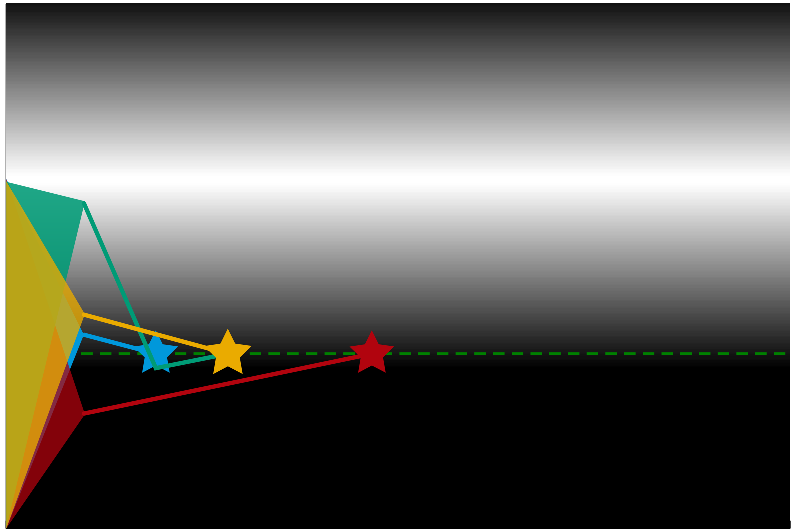
$$R(s,a) = \begin{cases} 100 & \text{if } a = 0, s = 0 \\ -100 & \text{if } a = 0, s \neq 0 \\ -1 & \text{otherwise} \end{cases}$$

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

POMDP Solution



QMDP

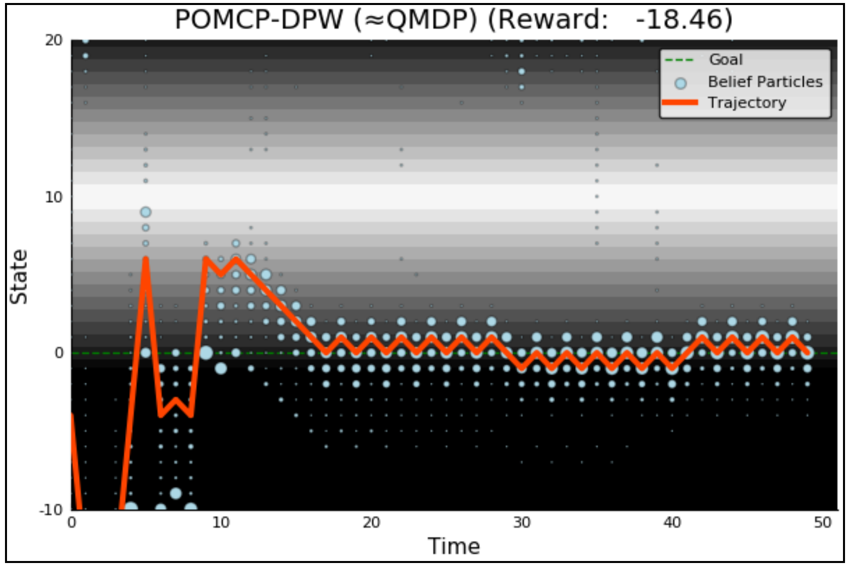


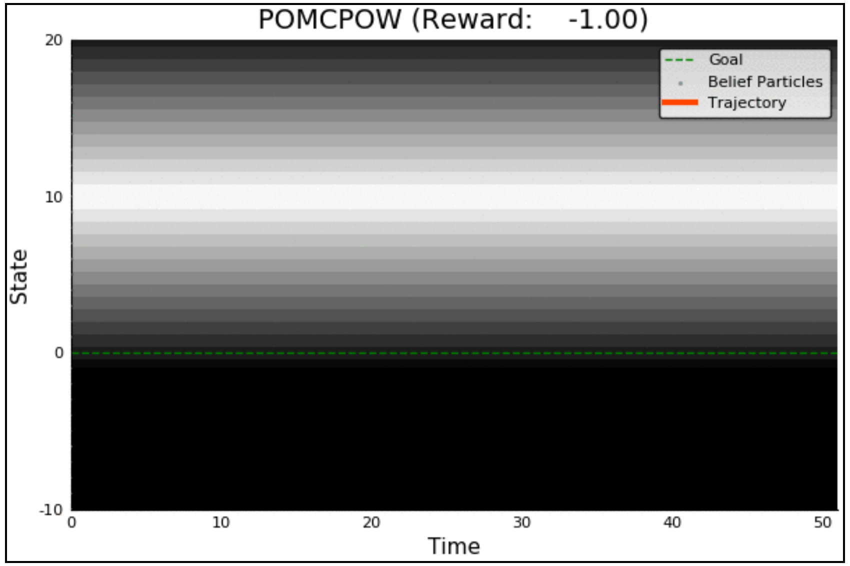
Same as **full observability**
on the next step

Information Gathering

QMDP

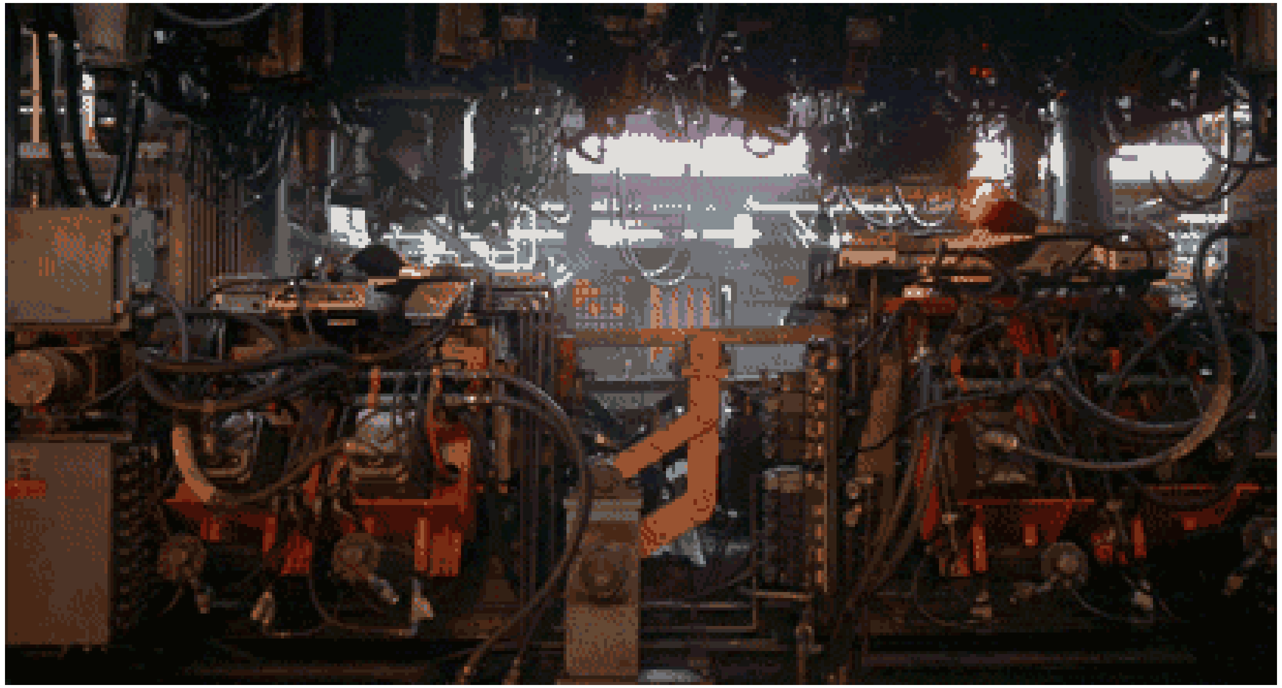
Full POMDP





QMDP

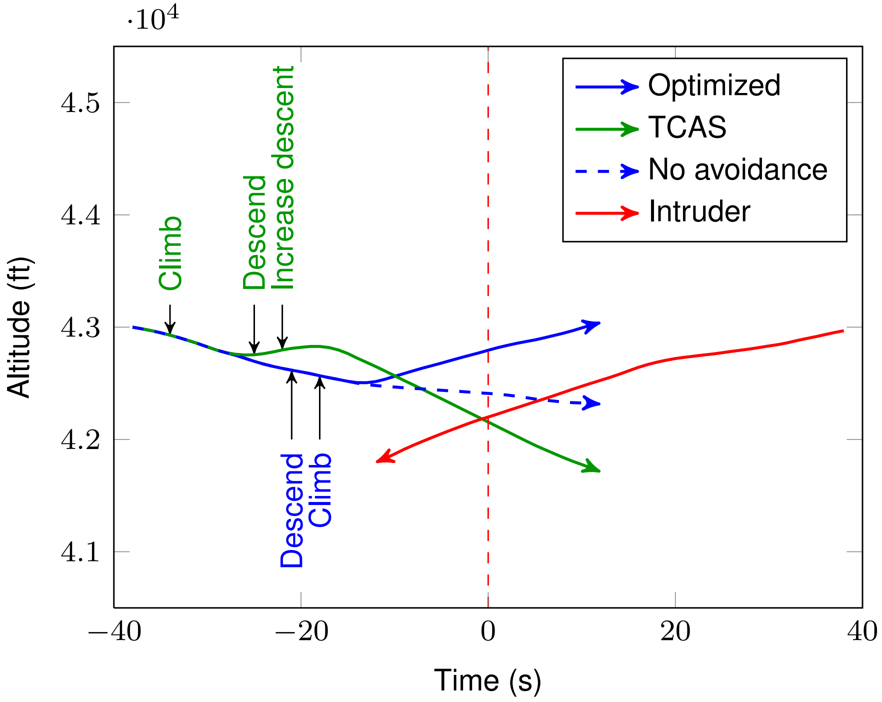
INDUSTRIAL GRADE



QMDP

ACAS X

[Kochenderfer, 2011]



Hindsight Optimization

POMDP Objective

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

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

FIB

POMDP Objective

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

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

k-Markov

POMDP Objective

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

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

Open Loop

POMDP Objective

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

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