Recap

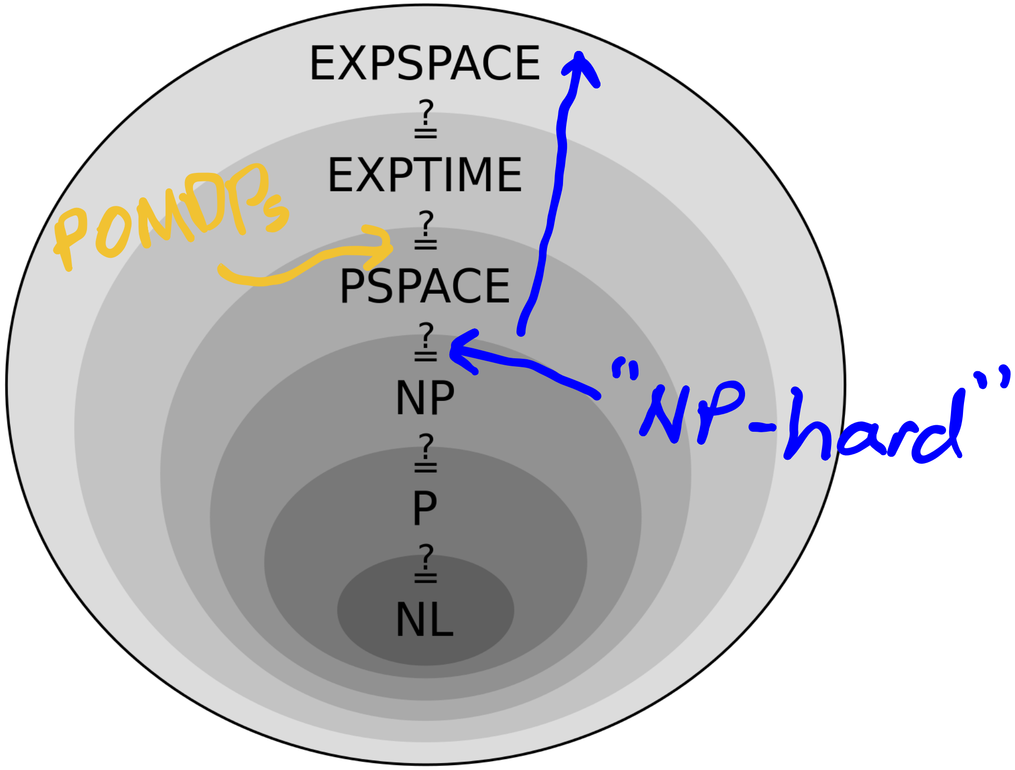
- Alpha Vectors
- Best solver for discrete POMDPs:

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

Tuesday

Online

After spring break

Formulation Approximations

(solve a slightly different problem)

Today!

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

$\pi(s) = -Ks$

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

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

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

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

$s_{t+1} = A s_t + B a_t + w_t \qquad w_t \sim N(0, \bar{W})$

$R = -s_t^T R_s s + -a_t^T R_a a_t$

$o_t = C s + v_t \qquad v_t \sim N(0, \bar{V})$

Certainty Equivalent

Optimal for LQG

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

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

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

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

QMDP

POMDP Objective

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

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

$$= \operatorname*{argmax}_{a} \alpha_a^\top b$$

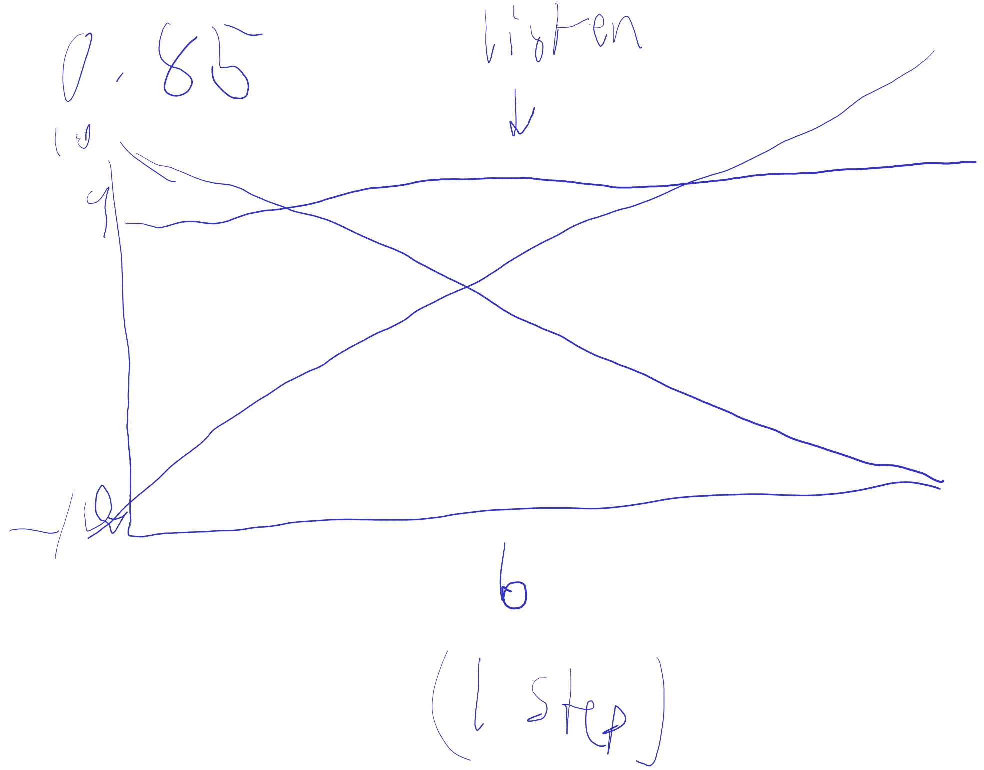
$$\pi_{\mathrm{QMDP}}(b) = \operatorname*{argmax}_{a\in A} \mathop{\mathrm{E}}_{s\sim b}\left[Q_{\mathrm{MDP}}(s, a)\right]$$

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

Example: Tiger POMDP with Waiting

$$P(o = TL \mid a = \text{listen}, s = TL) = 0.85$$

$$R = \begin{cases} 10 & a = TL \quad s = TR \\ -1 & a = \text{listen} \\ -100 & a = TL \quad s = TL \\ 0 & a = \text{wait} \end{cases}$$

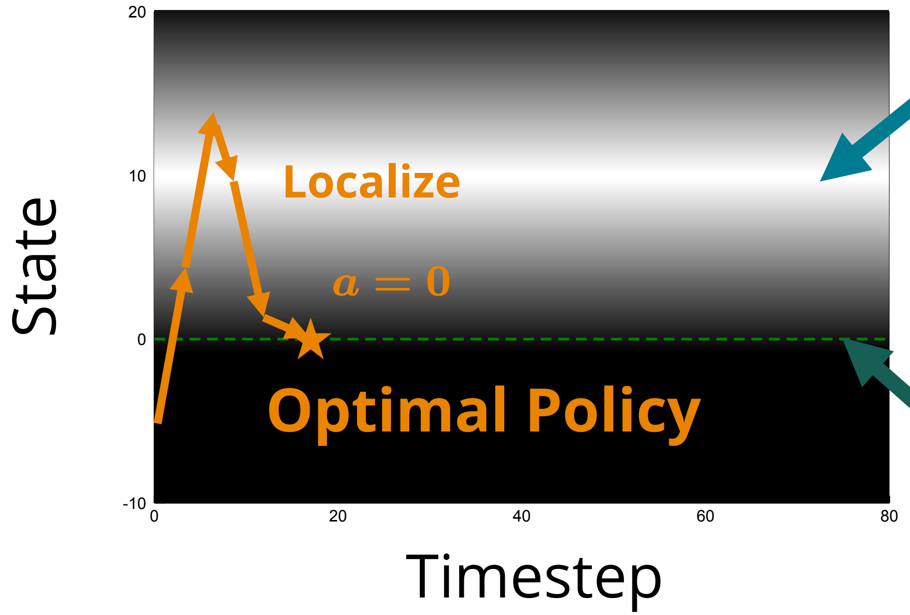


$$\gamma = 0.9$$

QMDP fails when information gathering is costly

POMDP Example: Light-Dark

Accurate Observations



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

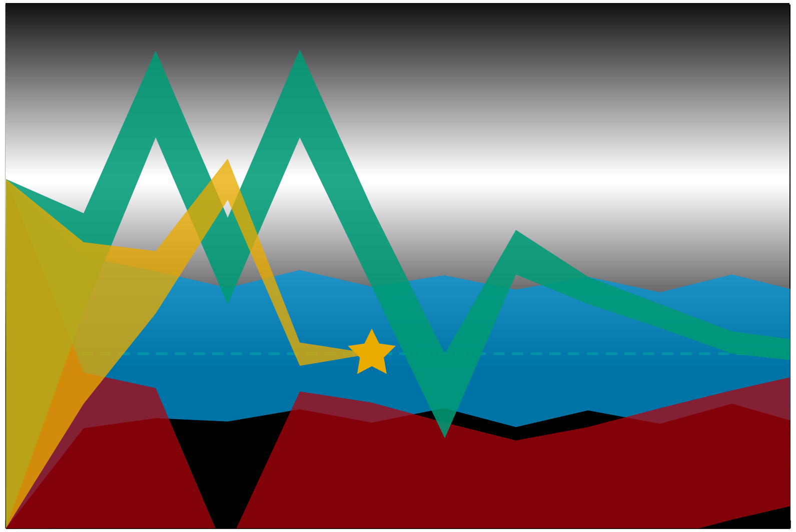
$$s' = s + a \qquad o \sim \mathcal{N}(s, s - 10)$$

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

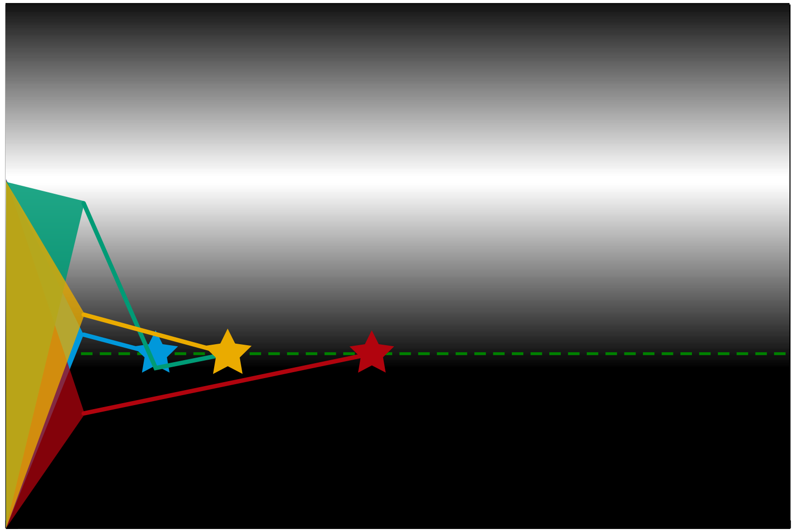
$$R(s,a) = \begin{cases} 100 & \text{if } a = 0, s = 0 \\ -100 & \text{if } a = 0, s \neq 0 \\ -1 & \text{otherwise} \end{cases}$$

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

POMDP Solution



QMDP

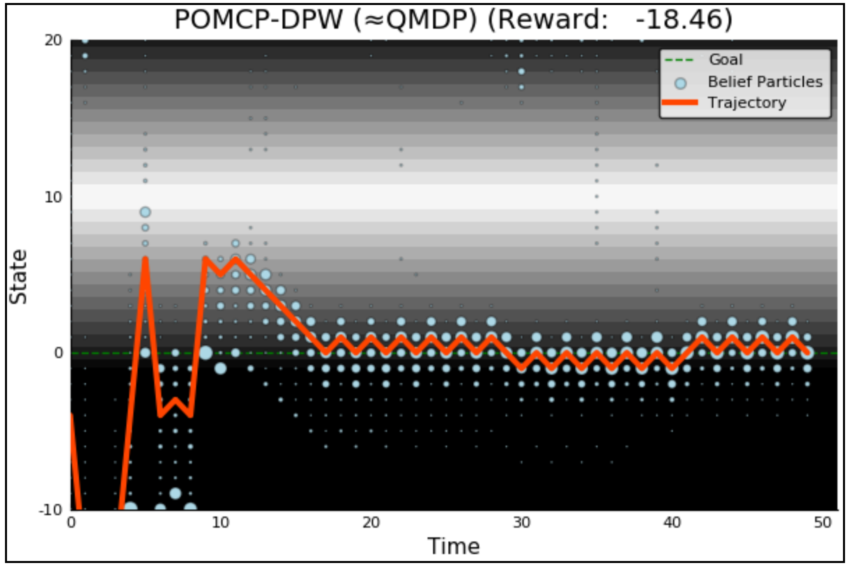


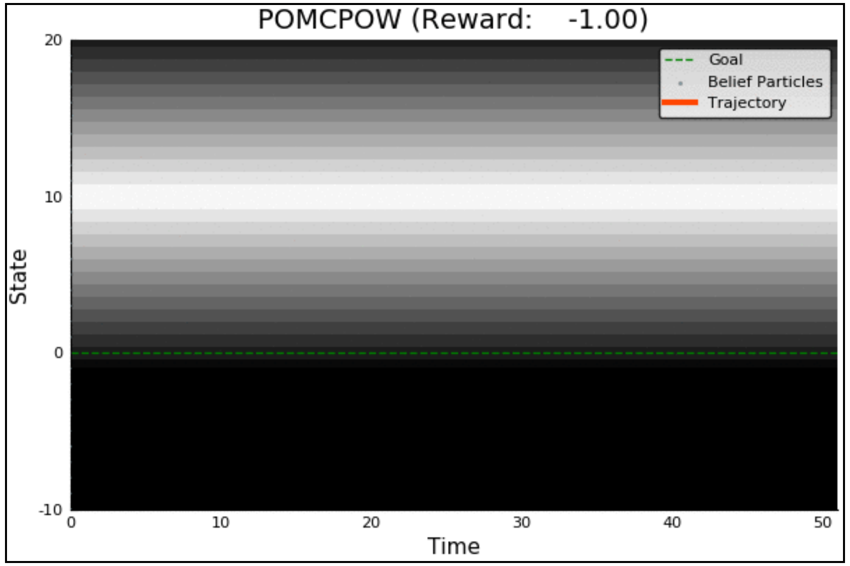
Same as **full observability** on the next step

Information Gathering

QMDP

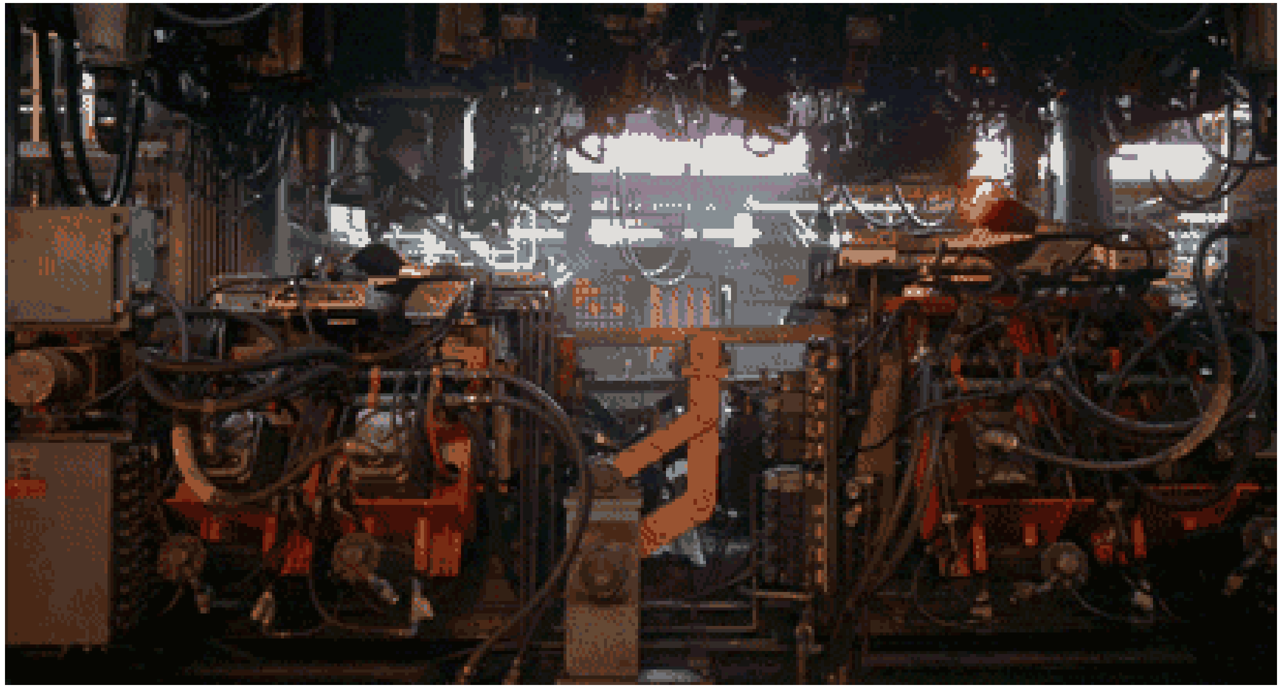
Full POMDP





QMDP

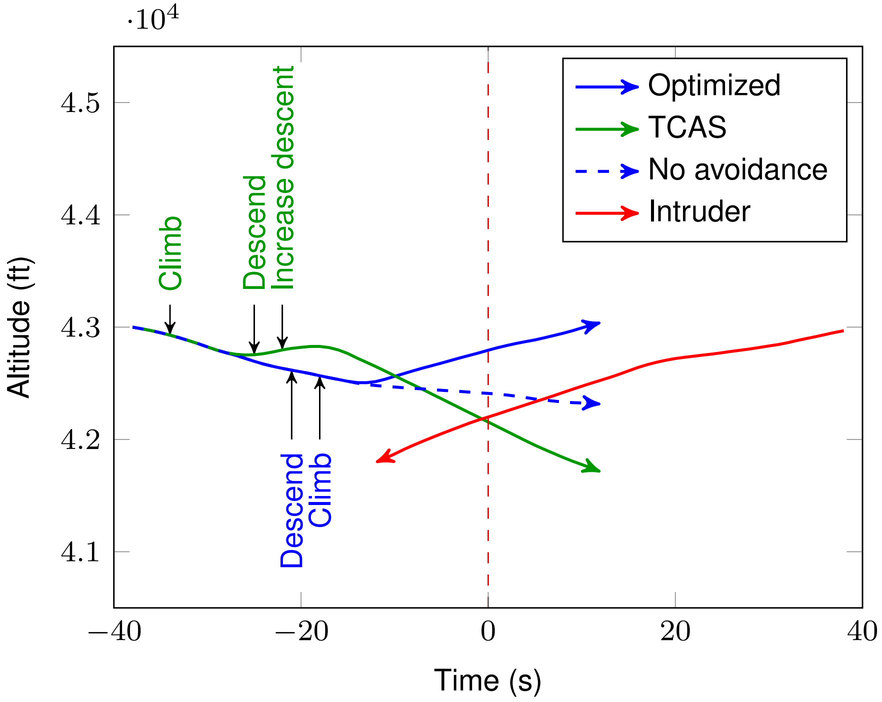
INDUSTRIAL GRADE



QMDP

ACAS X

[Kochenderfer, 2011]



Fast Informed Bound

FIB

POMDP Objective

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

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

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

QMDP (VI)

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

Hindsight Optimization

POMDP Objective

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

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

$s' = G(s, \underline{a}, w_t)$

Choose w_t

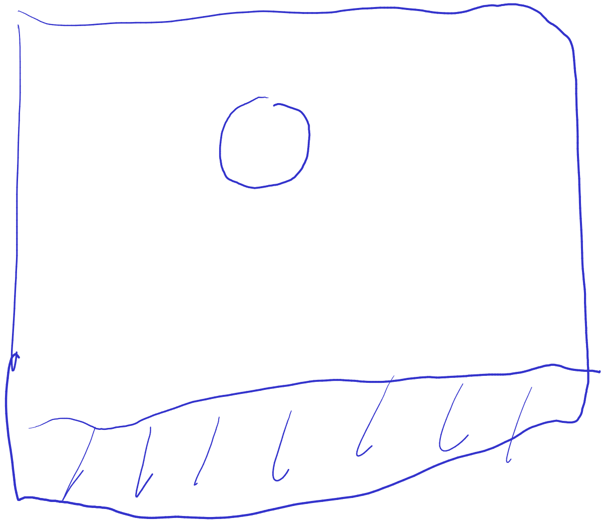
$a = \underset{a_1, a_2, \ldots}{\operatorname{argmax}} \; \mathbb{E}\left[R(s, a) + \ldots \right]$

k-Markov

POMDP Objective

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

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



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

Open Loop

POMDP Objective

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

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

$$(a_1, a_2, \ldots) = \underset{(a_1, a_2, \ldots)}{\operatorname{argmax}} \; \underline{\mathbb{E}}\left[\sum_{t=0}^{T} \gamma^T R(s_t, a_t)\right]$$

Comparison

Name	Description	Properties	Usefulness
Certainty Equivalence	Control as if the true state is mean of belief	Optimal for LQG	5 star
QMDP	Full observability after 1 time step	QMDP produces an upper bound for the true V	5 star
Hindsight Optimization	Hindsight knowledge of state and outcome uncertainty	Looser upper bound than QMDP	4 star
FIB	Takes 1 observation into account	Tighter upper bound than QMDP	2 star
k-Markov	Pretend that last k observations make up the state and solve that MDP	Great for Atari!	4 star
Open Loop	Choose sequence of actions	Good if alleatory is low, and epistemic is hard to reduce	3 star

Comparison

Name	**Description**	**Properties**	**Usefulness**
Certainty Equivalence	Control as if the true state is mean of belief	Optimal for LQG	5 star
QMDP	Full observability after 1 time step	QMDP produces an upper bound for the true V	5 star
Hindsight Optimization	Hindsight knowledge of state and outcome uncertainty	Looser upper bound than QMDP	4 star
FIB	Takes 1 observation into account	Tighter upper bound than QMDP	2 star
k-Markov	Pretend that last k observations make up the state and solve that MDP	Great for Atari!	4 star
Open Loop	Choose sequence of actions	Good if alleatory is low, and epistemic is hard to reduce	3 star