POMDP Model Approx. "Formulation" Approx

Last two Lectures: Numerically Approximate Solutions

- Offline PBVI / α vectors
- Online tree search

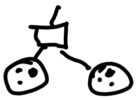
DESPOT α

Model Approximation

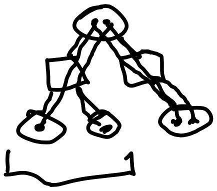
- Solve a slightly different opt. prob. "model"
- Toolbox
- when to use

DESPOT-α

POMCPOW : MCTS + Prog. Wid. + Weighted Particle Beliefs



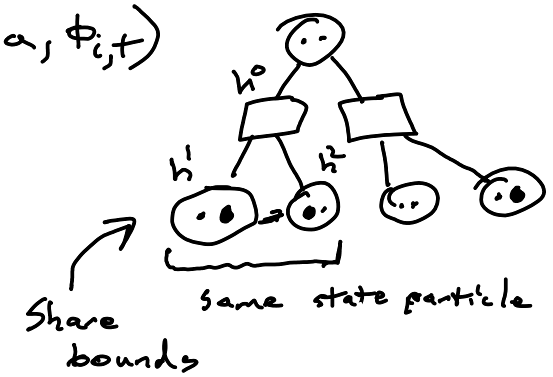
DESPOT: Heuristic Search + Scenarios

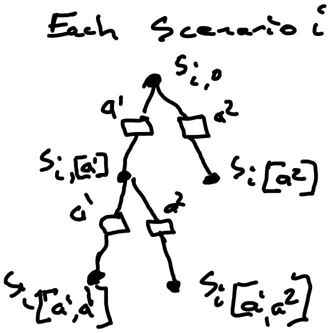


DESPOT-α

All scenarios in *all* belief node weghted by Z

$s',o,r = G(s,a,\phi_{i,t})$





$$U(h^1) = \sum_{i=1}^{K} w_i^{h^1} \underline{U(s_{i,h^0})}$$

$$L(h^1) = \sum_{i=1}^{K} w_i^{h^1} \cdot \underline{\alpha^p[s_{i,h^0}]}$$

any suboptimal plan

Model Approximations

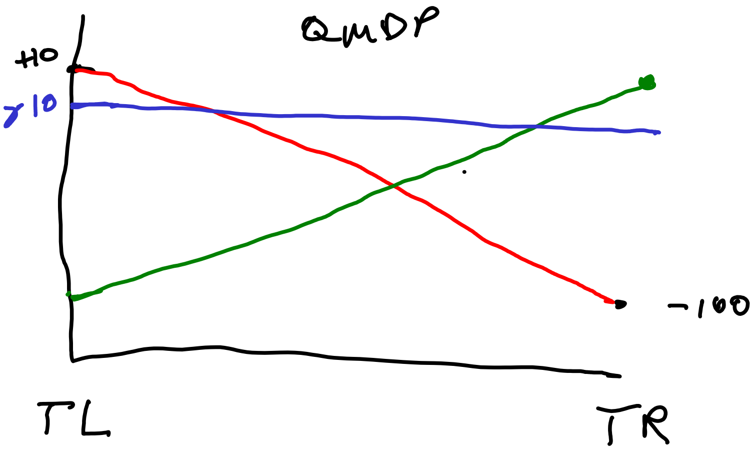
QMDP

α-vectors ~ Q values $\alpha^p[s] = Q^p(s, p[\cdot])$

$$\alpha^a[s] = Q_{MDP}(s,a)$$

↖ solution to underlying MDP

One shot Tiger POMDP



POMDP objective

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

QMDP

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

- belief at current time
- F.O. state in future

~ Optimistic

ACAS X

$$s = (\underbrace{\overbrace{\dot{h}, h_o, \dot{h}_i, adv}^{\text{Fully Observable}}}_{\text{millions}}, \overset{\text{partially}}{ps})$$

↖ pilot state ∈ {responsive, ignoring}

QMDP

$$\underset{s\sim b}{E}\left[QMDP(s,a)\right]$$

Name	Desc.	Properties	Usefullness
QMDP	Full obs. after 1 step hindsight kn. of epistemic	Upper Bound on true Value Function	☆☆☆☆☆
FIB	takes 1 step observation into account	Tighter upper Bound than QMDP	☆
Hindsight Optimization	Hindsight knowledge of state + outcome (alleatory + epistemic) uncertainty	Looser upper bound than QMDP	☆☆☆☆
Certainty Equivalence	Control as if mean (or median or mode) is true state	Optimal for LQR	☆☆☆☆☆
Open Loop action sequence - - - - - - no observations	Choose action sequence that optimizes objective in expectation	Good when hard to reduce epistemic	☆☆☆
Last k observations "k-markov"	Pretend last k observations are state and solve MDP	Great for Atari!	☆☆☆☆
Most Likely Observation	Plan assuming $b' = \tau(b, a, \hat{o}(b))$	No observation branching Good when Z unimodal	☆☆☆
Epistemic → Alleatory	Assume that partially observable part of s takes a random value at each time step	Conservative	☆☆

FIB

$$\pi_{FIB}(b) = \underset{a \in A}{\arg\max}\, {\alpha_{FIB}^{a}}^{T} b$$

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

Hindsight

POMDP objective

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

QMDP

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

HOP

$$V_{hs}(b) = E\left[\underset{(a_1 \ldots a_T)}{\overset{\downarrow}{\max}} \sum_{t=0}^{T} \gamma^t R(s_t, a_t) \right]$$

$$\pi_{hs}(b) = \underset{a \in A}{\arg\max} \underset{s_0 \sim b}{E}\left[R(s_0, a) + \max_{\substack{a_t \\ t \in 1..T}} \sum_{t=1}^{T} \gamma^t R(s_t, a_t) \right]$$

in VI
E

CE

$$\pi_s : S \to A$$

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

O.L.

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

no observation

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