

Offline POMDP Algorithms

Infinite-Horizon POMDP Value Iteration

$\Gamma \leftarrow$ blind lower bound

loop

$\Gamma \leftarrow \Gamma \cup \text{backup}(\Gamma)$

$\Gamma \leftarrow \text{prune}(\Gamma)$

Point-Based Value Iteration (PBVI)

backup(Γ, b)

 for $a \in A$

 for $o \in O$

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

$$\alpha_{a,o} \leftarrow \operatorname{argmax}_{\alpha \in \Gamma} \alpha^\top b'$$

 for $s \in S$

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

return $\operatorname{argmax}_{\alpha_a} \alpha_a^\top b$

Original PBDVI

$B \leftarrow b_0$

loop

 for $b \in B$

$\Gamma \leftarrow \Gamma \cup \{\text{point_backup}(\Gamma, b)\}$

 for $b \in B$

$\tilde{B} \leftarrow \{\tau(b, a, o) : a \in A, o \in O\}$

$B' \leftarrow B' \cup \left\{ \operatorname{argmax}_{b' \in \tilde{B}} \|B, b'\| \right\}$

$B \leftarrow B \cup B'$

PERSEUS: Randomly Selected Beliefs

Two Phases:

1. Random Exploration
2. Value Backup

Random Exploration:

$$B \leftarrow \emptyset$$

$$b \leftarrow b_0$$

loop until $|B| = n$

$$a \leftarrow \text{rand}(A)$$

$$o \leftarrow \text{rand}(P(o \mid b, a))$$

$$b \leftarrow \tau(b, a, o)$$

$$B = B \cup \{b\}$$

Heuristic Search Value Iteration (HSVI)

while $\bar{V}(b_0) - \underline{V}(b_0) > \epsilon$

 explore($b_0, 0$)

function explore(b, t)

 if $\bar{V}(b) - \underline{V}(b) > \epsilon\gamma^t$

$a^* = \operatorname{argmax}_a \bar{Q}(b, a)$

$o^* = \operatorname{argmax}_o P(o \mid b, a) (\bar{V}(\tau(b, a^*, o)) - \underline{V}(\tau(b, a^*, o)) - \epsilon\gamma^t)$

 explore($\tau(b, a^*, o^*), t + 1$)

$\underline{\Gamma} \leftarrow \underline{\Gamma} \cup \text{point_backup}(\underline{\Gamma}, b)$

$\bar{V}(b) = B_b [\bar{V}(b)]$

Sawtooth Upper Bounds

SARSOP

Successive Approximation of Reachable Space under Optimal Policies

Policy Graphs

Monte Carlo Value Iteration (MCVI)