Offline POMDP Algorithms

Last time: POMDP Value Iteration (horizon d)

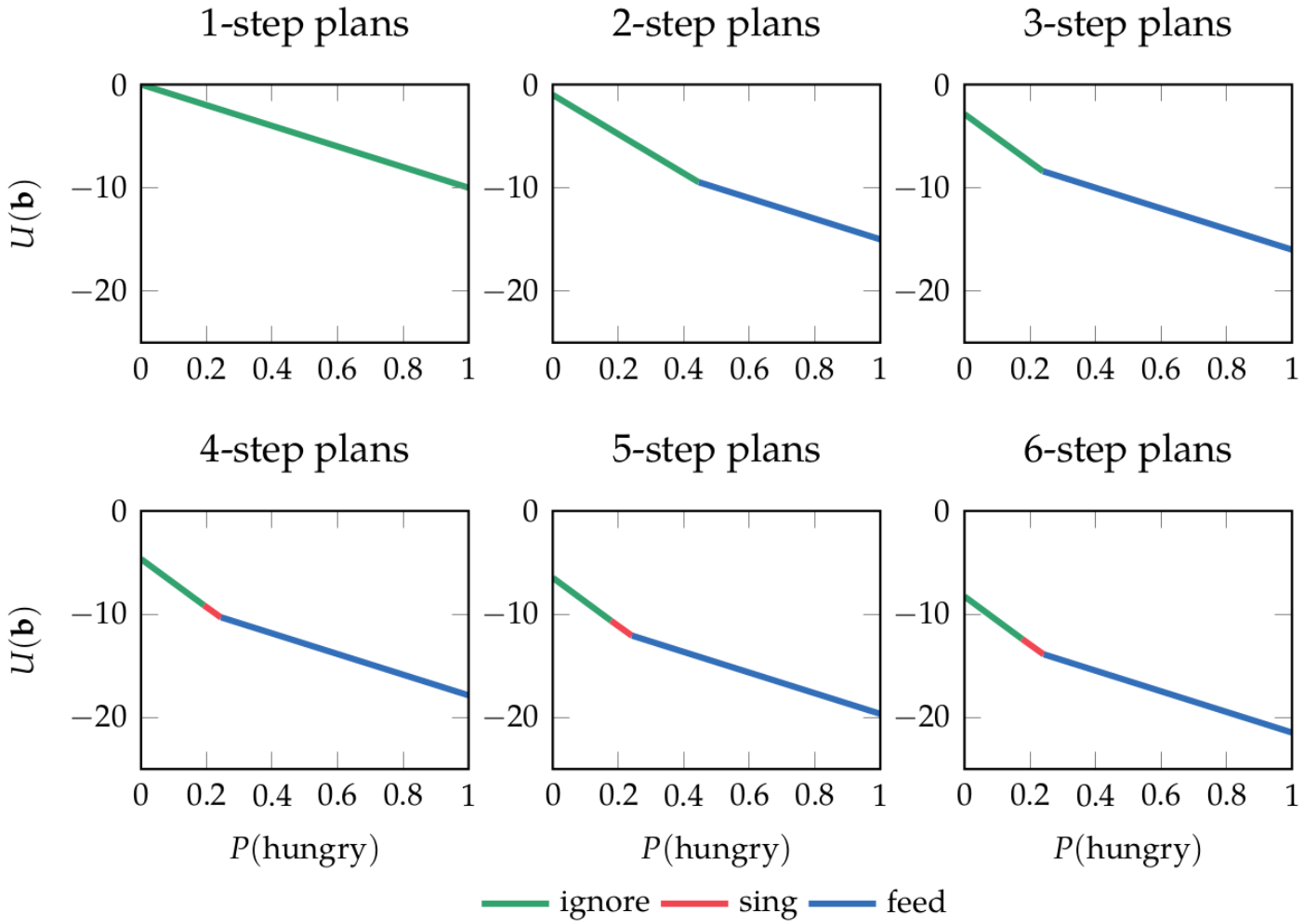
$\Gamma^0 \leftarrow \emptyset$

for $n \in 1 \dots d$

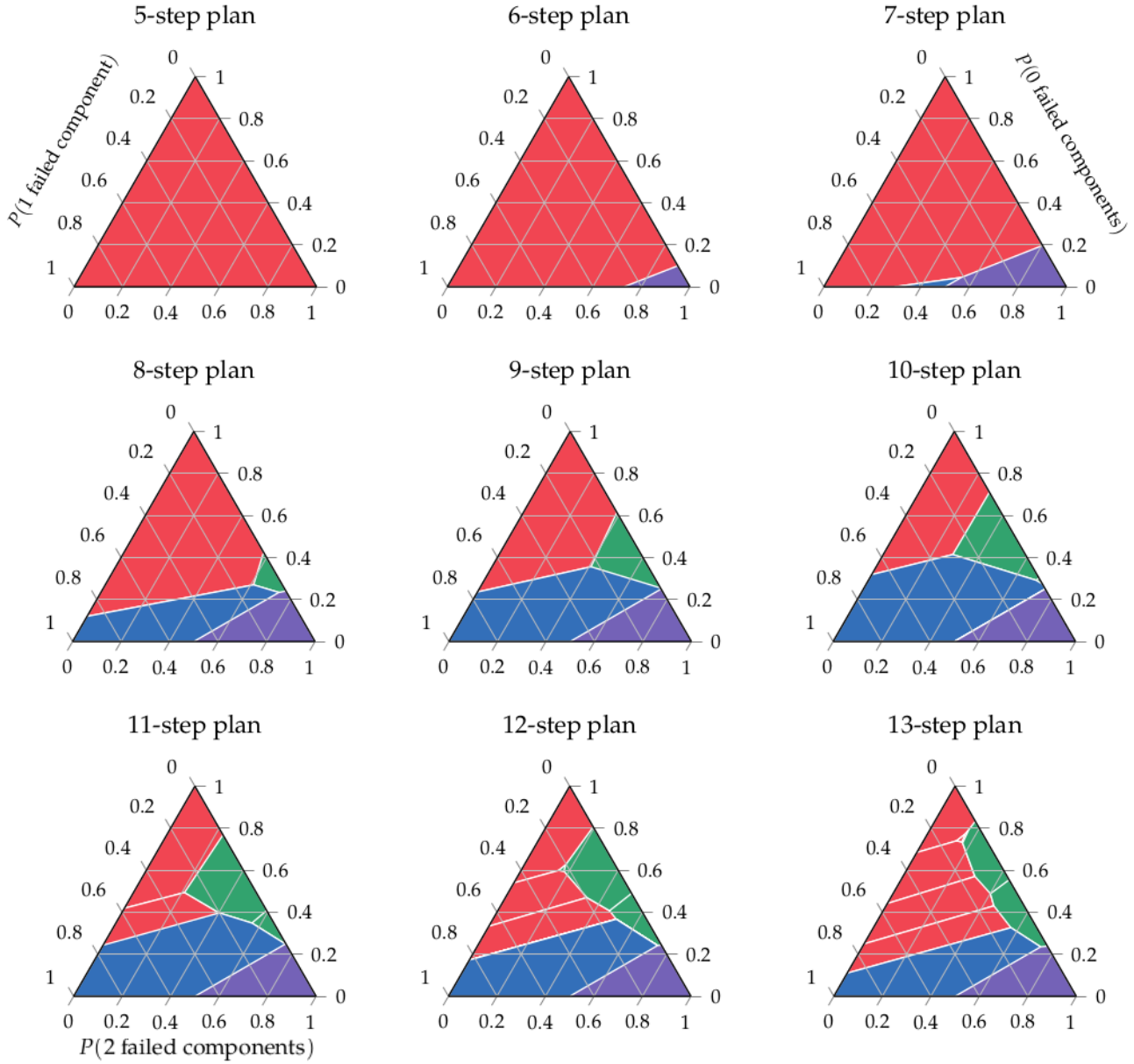
Construct Γ^n by expanding with Γ^{n-1}

Prune Γ^n

Finite Horizon POMDP Value Iteration



Finite Horizon POMDP Value Iteration



Infinite-Horizon POMDP Lower Bound Improvement

$\Gamma \leftarrow$ blind lower bound

loop

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

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

Point-Based Value Iteration (PBVI)

point_backup(Γ, b)

for $a \in A$

for $o \in O$

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

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

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 $\underset{\alpha_a}{\operatorname{argmax}}\ \alpha_a^\top b$

Original PBVI

$B \leftarrow b_0$

loop

  for $b \in B$

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

  $B' \leftarrow \emptyset$

  for $b \in B$

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

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

  $B \leftarrow B \cup B'$

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

  explore$(b_0, 0)$

explore(b, t)

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

    $a^* = \underset{a}{\text{argmax}}\ \overline{Q}(b, a)$

    $o^* = \underset{o}{\text{argmax}}\ P(o \mid b, a)\left(\overline{V}(\tau(b, a^*, o)) - \underline{V}(\tau(b, a^*, o)) - \epsilon\gamma^t\right)$

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

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

    $\overline{V}(b) = B_b\left[\overline{V}(b)\right]$

Sawtooth Upper Bounds

SARSOP

Successive Approximation of Reachable Space under Optimal Policies

Offline POMDP Algorithms

Policy Graphs

Monte Carlo Value Iteration (MCVI)

MC-Backup (G, b, N)

$R_a = 0 \quad V_{a,o,v} = 0$

for $a \in A$

for i in $1:N$

$s_i \leftarrow$ sample (b)

$s_i', o_i, r_i \leftarrow G(s_i, a)$

$R_a + r_i$

for $v \in G$

$V_{a,o_i,v} = V_{a,o_i,v} +$ Simulate(G, v, s', L)

for o in O

$V_{a,o} = \max_{v \in G} V_{a,o,v}$

$v_{a,o} = \operatorname{argmax}_{v \in G} V_{a,o,v}$

$V_a = R_a + \gamma \sum_o V_{ao}/N$

$V^* = \max_a V_a$

$a^* = \operatorname{argmax}_a V_a$

add new node to G labeled with a^*