Point Based Value Iteration with Optimal Belief Compression for Dec-POMDPs

Georgia Tech

Liam MacDermed

Charles L. Isbell

Problem: **An agents' belief set has doubly exponential growth with the horizon due to nested beliefs.**

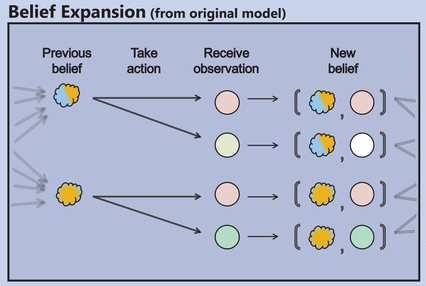
Key Idea: **Allow agent's to decide how to best compress their own beliefs and fold this belief choice into the model itself thereby allowing our model solver to compute an optional compression scheme.**

Formal Solution:

Given a DecPOMDP $\langle N, A, S, O, P, R, s^{(0)} \rangle$ (with states $\omega \in S$ and observations $\sigma \in O$) we define the BB-DecPOMDP approximation model $\langle N', A', O', S', P', R', s'^{(0)} \rangle$ (which optimally compresses beliefs) with belief set size parameters $t_1, \cdots, t_n$ as:

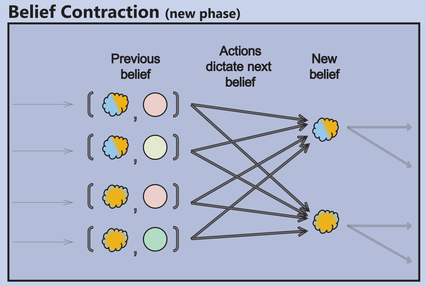
- $N' = N$ and $A'_i = \{a_1, \cdots, a_{\max(|A_i|,\ t_i)}\}$
- $O'_i = \{O_i, \emptyset\} \bigtimes \{1, 2, \cdots, t_i\}$ with factored observation $o_i = \langle \sigma_i, \theta_i \rangle \in O'$
- $S' = S \bigtimes O'$ with factored state $s = \langle \omega, \sigma_1, \theta_1, \cdots, \sigma_n, \theta_n \rangle \in S'$.
- $$P'(s'|s,a) = \begin{cases} P(\omega', \langle \sigma'_1, \cdots, \sigma'_n \rangle | \omega, a) & \text{if } \forall i : \sigma_i = \emptyset, \sigma'_i \neq \emptyset \text{ and } \theta'_i = \theta_i \\ 1 & \text{if } \forall i : \sigma_i \neq \emptyset, \sigma'_i = \emptyset \text{ and } \theta'_i = a_i, \omega' = \omega \\ 0 & \text{otherwise} \end{cases}$$
- $$R'(s,a) = \begin{cases} R(\omega, a) & \text{if } \forall i : \sigma_i = \emptyset \\ 0 & \text{otherwise} \end{cases}$$
- $s'^{(0)} = \langle s^{(0)}, \emptyset_1, 1, \cdots, \emptyset_n, 1 \rangle$ is the initial state distribution

Belief Expansion (from original model)



Model alternates between these two phases

Belief Contraction (new phase)



Results on Benchmark Problems

				Previous Best	1-Belief		2-Beliefs	
	$\lvert S\rvert$	$\lvert A_i\rvert$	$\lvert O_i\rvert$	Utility	Utility	$\lvert\Gamma\rvert$	Utility	$\lvert\Gamma\rvert$
Dec-Tiger	2	2	2	13.4486	-20.000	2	4.6161	187
Broadcast	4	2	2	9.1	9.2710	36	9.2710	44
Recycling	4	3	2	31.92865	26.3158	8	31.9291	13
Grid small	16	5	2	6.89	5.2716	168	6.8423	206
Box Pushing	100	4	5	149.854	127.1572	258	223.8674	357
Wireless	64	2	6	-175.40	-208.0437	99	-167.1025	374

	3-Beliefs		4-Beliefs		5-Beliefs	
	Utility	$\lvert\Gamma\rvert$	Utility	$\lvert\Gamma\rvert$	Utility	$\lvert\Gamma\rvert$
Dec-Tiger	13.4486	231	13.4486	801	13.4486	809
Broadcast	9.2710	75	9.2710	33	9.2710	123
Recycling	31.9291	37	31.9291	498	31.9291	850
Grid small	6.9826	276	6.9896	358	6.9958	693
Box Pushing	224.1387	305	-	-	-	-

Utility achieved by our PBVI-BB-DecPOMDP algorithm compared to the previously best known policies on a series of standard benchmarks. Higher is better. Our algorithm beats all previous results except on Dec-Tiger where we believe an optimal policy has already been found.

Approach: Solve a DecPOMDP by transforming it into a POMDP

DecPOMDP → *Optimal Belief Compression (lossy)* → **Bounded Belief DecPOMDP** → *Centralize the BB-DecPOMDP (lossless)* → **Belief POMDP** → *Modified Point Based Value Iteration* → **Policy Over Joint Strategies** → *Direct Mapping* → **Joint Policy**

Problem: **How can we centralize the model in order to make the problem tractable?**

Key Ideas: **Factor the state so that it contains each agent's current belief.**
Beliefs are only labels and can be decoupled from history.
We can convert a BB-DecPOMDP into a POMDP with exponential actions corresponding to strategies of a CBG by utilizing the fact that the common knowledge distribution over joint-beliefs is a sufficient statistic for planning (previously proved)

Formal Solution:

Belief-POMDP $\langle A', S', O', P', R', s'^{(0)} \rangle$ converted from BB-DecPOMDP $\langle N, A, S, O, P, R, s^{(0)} \rangle$ (with belief labels Θ_i for each agent).

Factored states: $S' = S \times \prod_{i=1}^{n} \Theta$ with $\langle \omega, \theta_1, \cdots, \theta_n \rangle \in S'$
where $\omega \in S$ is the underlying state
and $\theta_i \in \Theta_i$ is agent i's belief.

No observations: $O' = \{\}$

Actions are strategies: $A' = \prod_{i=1}^{n} \prod_{j=1}^{|\Theta_i|} A_i$
(one action for each agent for each belief)

Transition function: $P'(s'|s,a), = \sum_{[\theta,o]=\theta'} P(\omega', o|\omega, \langle a_{\theta_1}, \cdots, a_{\theta_n} \rangle)$
where a_{θ_i} is the action agent i would take in belief θ_i.

Reward function: $R'(s,a) = R(\omega, \langle a_{\theta_1}, \cdots, a_{\theta_n} \rangle)$

Initial state: $s'^{(0)}_{\omega} = s^{(0)}$

Problem: **The Belief POMDP has an exponential action space corresponding to joint-strategies.**

Key Idea: **Modify PBVI so that instead of finding the best action (e.g. strategy) exhaustively, solve the corresponding CBG.**

Formal Solution:

Inputs: DecPOMDP M, discount γ
belief bounds $|\Theta_i|$, stopping criterion ϵ_Γ
Output: value function Γ

$\langle N, A, O, S, P, R, s^{(0)} \rangle \Leftarrow$ BB-DecPOMDP approximation of M
$B^{\forall} \Leftarrow$ sampling of states using a random walk from $s^{(0)}$
$\Gamma' \Leftarrow \{\langle R_{\min}/\gamma, \cdots, R_{\min}/\gamma \rangle\}$
repeat
  $B \Leftarrow B^{\forall}$; $\Gamma \Leftarrow \Gamma'$; $\Gamma' \Leftarrow \emptyset$
  while $B \neq \emptyset$ **do**
    $b \Leftarrow$ Rand$(b \in B)$
    $\alpha \Leftarrow \Gamma(b)$
    $\alpha' \Leftarrow$ optimal point of integer program solving CBG at b
    if $\alpha'(b) > \alpha(b)$ **then**
      $\alpha \Leftarrow \alpha'$
    $\Gamma' \Leftarrow \Gamma' \bigcup \alpha$
    for all $b \in B$ **do**
      if $\alpha(b) > \Gamma(b)$ **then**
        $B \Leftarrow B/b$
until $\Gamma' - \Gamma < \epsilon_\Gamma$
return Γ

Problem: **Problem: How do we solve a Cooperative Bayesian Game (CBG) efficiently given that it's NP complete?**

Key Idea: **Solve for a deterministic agent normal form correlated equilibrium. This is an ILP. The linear relaxation is very often integral.**

Formal Solution:

Maximize: $$V_{\vec{a},\alpha}(b) = \sum_{s \in S} b(s) \sum_{\vec{a} \in A} \sum_{s' \in S} P(s'|s, \vec{a}) \alpha(s') x_{\vec{a},\theta \in s}$$

Over: $$x \in \{0,1\}^{|S||A|}, \quad \alpha \in \Gamma$$

Subject to: For each agent i, joint-type θ, and partial joint-actions of other agents $\vec{a}_{-i}$

$$\sum_{a_i \in A_i} x_{\vec{a},\theta} = x_{\langle \vec{a}_{-i}, \theta_{-i} \rangle}$$

$$\text{for each } \theta \in \Theta\text{: } \sum_{\vec{a} \in A} x_{\vec{a},\theta} = 1$$

$$\text{for each } \theta \in \Theta,\ \vec{a} \in A\text{: } \quad x_{\vec{a},\theta} \geq 0$$