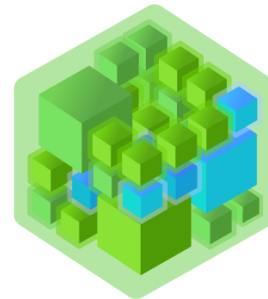


Navegação - Conceitos gerais

Real-time hierarchical POMDPs for autonomous robot navigation

Michael André Gonçalves



lcad

LABORATÓRIO DE COMPUTAÇÃO
DE ALTO DESEMPENHO

- Muitos métodos de navegação de robôs contemporâneos não levam em conta a incerteza do movimento.
- Métodos baseados em *Partially Observable Markov Decision Process* (POMDP) poder ser utilizados para a navegação probabilística.
- Nesse trabalho é estudado uma implementação hierárquica do POMDP aplicado à navegação robótica (*Robotic Navigation Hierarchical POMDP*).

- Introdução
- Partially Observable Markov Decision Processes (POMDP).
- Formulação do POMDP em navegação robótica.
- Hierarchical POMDP aplicado à navegação robótica.
- Algoritmo.
- Resultado.

- O POMDP pode modelar o estado oculto de um robô que não é completamente observável.
- Ele possui um alto custo computacional quando aplicados em ambientes reais o tornando intratável de ser resolvido por métodos exatos.
- Métodos de aproximação podem ser aplicados, porem são limitados a tratar apenas alguns milhares de estados.
- Portanto, o POMDP é aplicado principalmente no planejamento de navegação à alto nível.
- Nesse trabalho é estudado uma proposta de um POMDP hierárquico.

- O POMDP recebe como entrada o estado do ambiente e gera ações como saída.
- Ele considera que a ação do sistema sobre o ambiente não é garantida.
- Conseqüentemente, não se pode garantir o estado futuro.
- Formado por dois módulos essenciais:
 - *Belief Update*
 - *policy component*

- O POMDP é formalmente representado por:

$$\mathcal{M} = \langle \mathcal{S}, \mathcal{A}, \mathcal{T}, \mathcal{R}, \mathcal{Z}, \mathcal{O} \rangle$$

- \mathcal{S} : Conjunto de estados;
- \mathcal{A} : Conjunto das ações;
- \mathcal{Z} : Conjunto de observações;
- $T(s,a,s')$: Função de transição de estado $\rightarrow P(s'|s,a)$;
- $R(s,a)$: Função de recompensa;
- $O(s',a,z)$: Probabilidade sob a observação $\rightarrow P(z|s',a)$
- $b_t(s)$: Probabilidade de se estar no estado s no tempo t .
- B : é o conjunto de b_t possíveis.



- *Belief Update*

$$\begin{aligned} b_{t+1}(s') &= P(s'|z, a, b_t) \\ &= \frac{P(z|s', a, b_t)P(s'|a, b_t)}{P(z|a, b_t)} \\ &= \frac{P(z|s', a, b_t) \sum_{s \in S} P(s'|a, b_t, s)P(s|a, b_t)}{P(z|a, b_t)} \\ &= \frac{P(z|s', a) \sum_{s \in S} P(s'|s, a)P(s|b_t)}{P(z|a, b_t)} \\ &= \frac{\mathcal{O}(s', a, z) \sum_{s \in S} \mathcal{T}(s, a, s')b_t(s)}{P(z|a, b_t)} \end{aligned}$$

- *Belief Update*

$$\begin{aligned} b_{t+1}(s') &= P(s'|z, a, b_t) \\ &= \frac{P(z|s', a, b_t)P(s'|a, b_t)}{P(z|a, b_t)} \\ &= \frac{P(z|s', a, b_t) \sum_{s \in S} P(s'|a, b_t, s)P(s|a, b_t)}{P(z|a, b_t)} \\ &= \frac{P(z|s', a) \sum_{s \in S} P(s'|s, a)P(s|b_t)}{P(z|a, b_t)} \\ &= \frac{\mathcal{O}(s', a, z) \sum_{s \in S} \mathcal{T}(s, a, s')b_t(s)}{P(z|a, b_t)} \end{aligned}$$

O denominador $P(z|a, b_t)$, é um fator de normalização e é equivalente à probabilidade total da percepção da observação z .

$$\begin{aligned} P(z|a, b_t) &= \sum_{s' \in S} P(z|s', a)P(s'|s, a)b_t(s) \\ &= \sum_{s' \in S} \mathcal{O}(z, s', a)\mathcal{T}(s, a, s')b_t(s) \end{aligned}$$



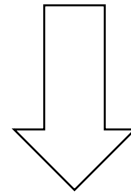
- *policy component*

$$V_t^*(s) = \max_{a \in \mathcal{A}} \left[\mathcal{R}(s, a) + \gamma \sum_{s' \in \mathcal{S}} \mathcal{T}(s, a, s') V_{t-1}(s') \right]$$



- *policy component*

$$V_t^*(s) = \max_{a \in \mathcal{A}} \left[\mathcal{R}(s, a) + \gamma \sum_{s' \in \mathcal{S}} \mathcal{T}(s, a, s') V_{t-1}(s') \right]$$



$$V_t^*(b) \triangleq \max_{a \in \mathcal{A}} \left[\rho(b, a) + \gamma \sum_{b' \in \mathcal{B}} \tau(b, a, b') V_{t-1}(b') \right]$$



- *policy component*

$$V_t^*(s) = \max_{a \in \mathcal{A}} \left[\mathcal{R}(s, a) + \gamma \sum_{s' \in \mathcal{S}} \mathcal{T}(s, a, s') V_{t-1}(s') \right]$$

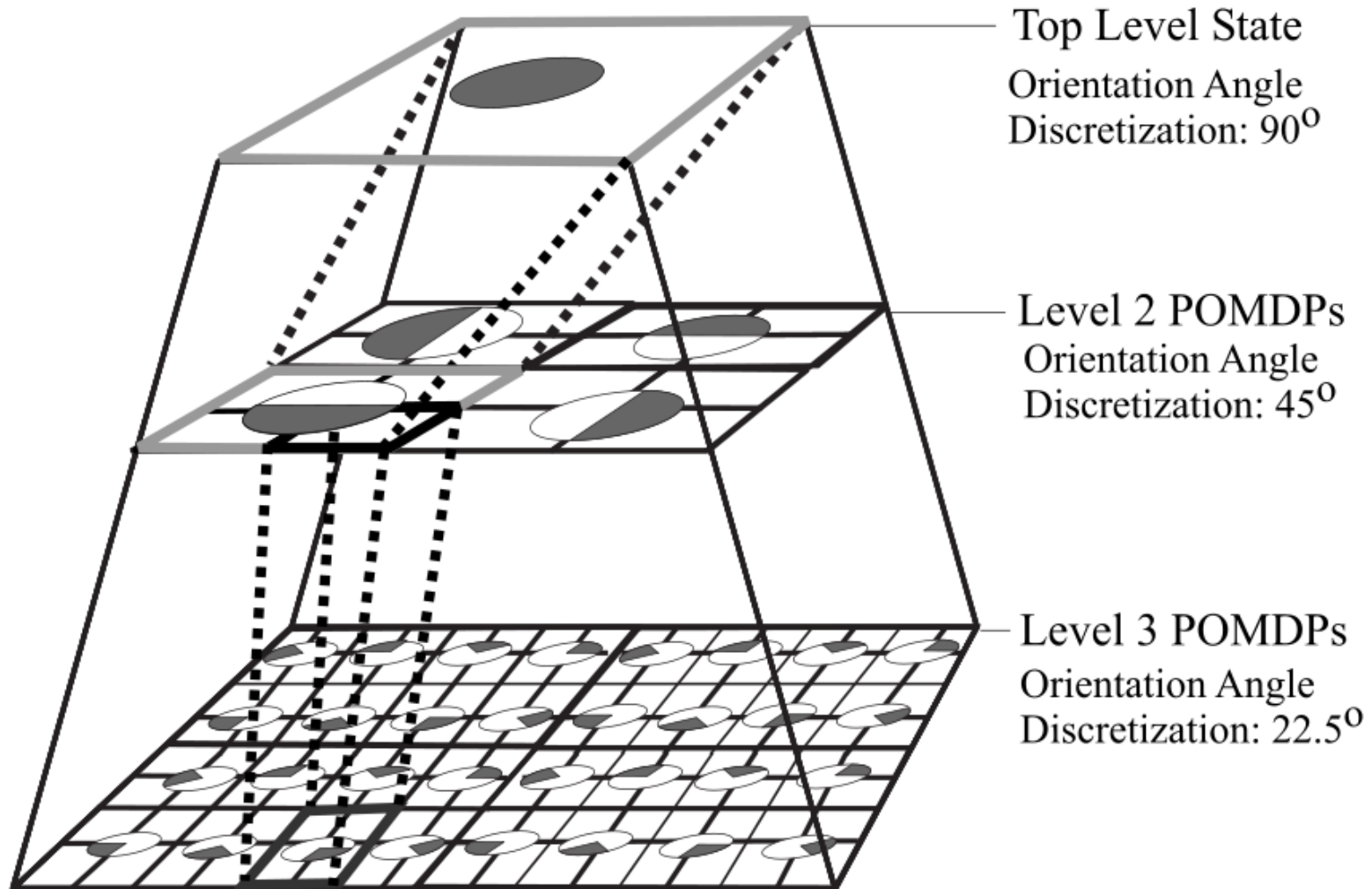
$$V_t^*(b) \triangleq \max_{a \in \mathcal{A}} \left[\rho(b, a) + \gamma \sum_{b' \in \mathcal{B}} \tau(b, a, b') V_{t-1}(b') \right]$$

$$\rho(b, a) = \sum_{s \in \mathcal{S}} b(s) R(s, a) \quad \tau(b, a, b') = P(b' | a, b)$$

- S: Localização a partir de um *grid map* e um ângulo de orientação.
- A: Ângulo de rotação;
- Z: Sensores Auxiliares para a localização;
- R: Recompensa atribuída por uma ação sobre uma posição do *grid map*;

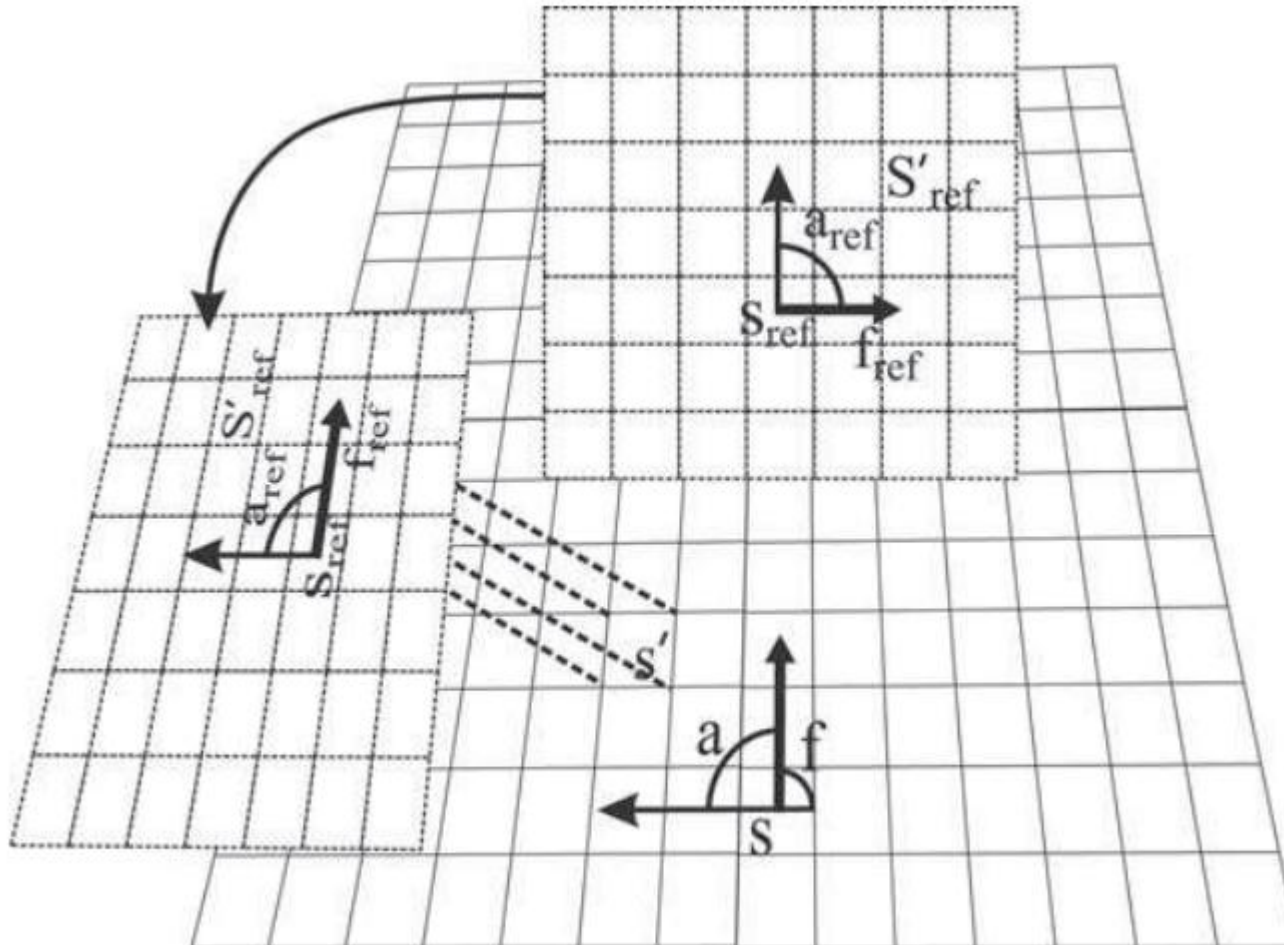
- Divisão do POMDP de forma hierárquica
 - A instância do nível do topo trabalha com ações/estados em ângulos de 90° com $[0^\circ, 360^\circ]$ de atuação.
 - Cada nível descido dobra a discretização, por exemplo: o nível seguinte do topo trabalha com 45° .

Estrutura do HPOMDP



	Top level	Intermediate level l	Bottom level
No of POMDPs	1	$ \mathcal{A}^{l-1} \times \mathcal{S}^{l-1} $	$ \mathcal{A}^{L-1} \times \mathcal{S}^{L-1} $
Size of \mathcal{S}	$ \mathcal{S}^0 /2^{2(L-1)}$	20	$5 \times (2 + r)^2$
Range of orientation angles	$[0^\circ, 360^\circ]$	$[\theta_p - \frac{90^\circ}{2^{l-1}}, \theta_p + \frac{90^\circ}{2^{l-1}}]$	$[\theta_p - \frac{90^\circ}{2^{L-1}}, \theta_p + \frac{90^\circ}{2^{L-1}}]$
Resolution of orientation angles	90°	$90^\circ/2^{l-1}$	$90^\circ/2^{L-1}$
Size of \mathcal{A}	4	5	5
Range of action angles	$[0^\circ, 360^\circ]$	$[a_p - \frac{90^\circ}{2^{l-2}}, a_p + \frac{90^\circ}{2^{l-2}}]$	$[a_p - \frac{90^\circ}{2^{L-2}}, a_p + \frac{90^\circ}{2^{L-2}}]$
Resolution of action angles	90°	$90^\circ/2^{l-1}$	$90^\circ/2^{L-1}$

Estrutura do HPOMDP

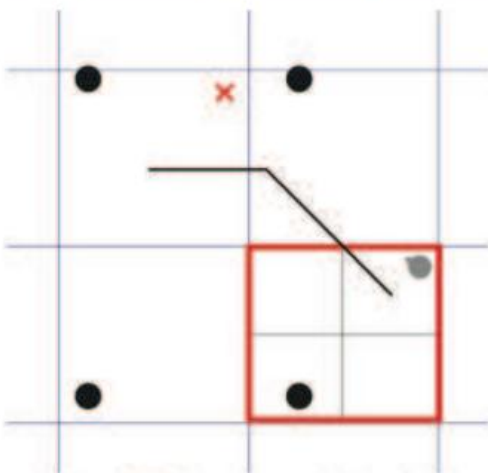
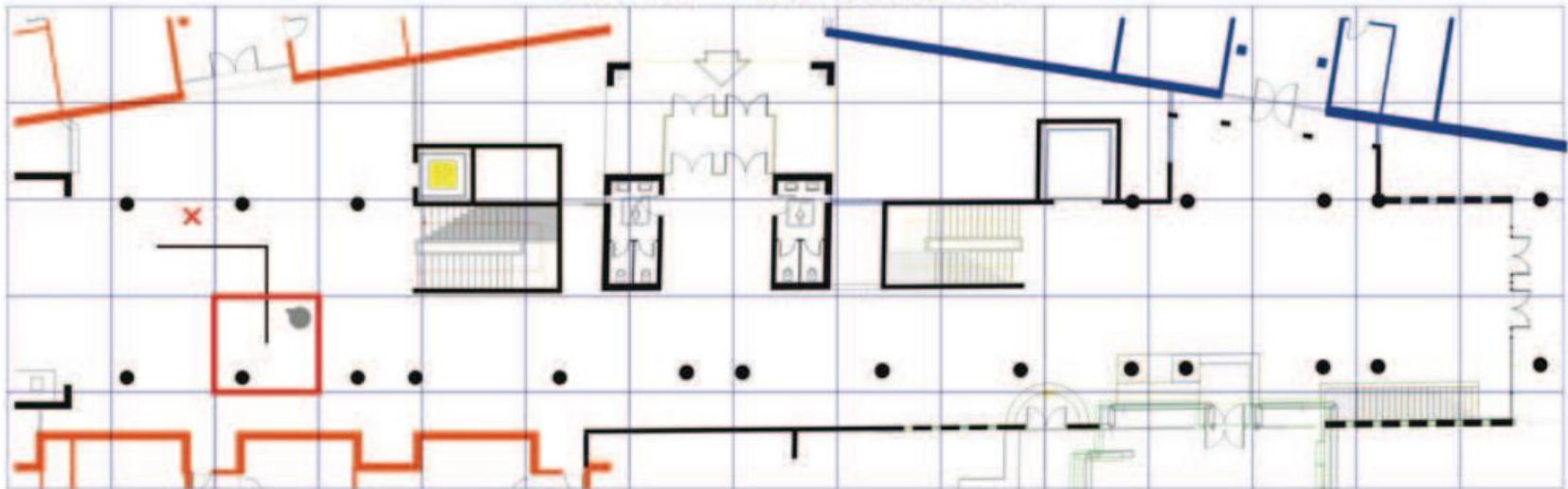


RN-HPOMDP planning

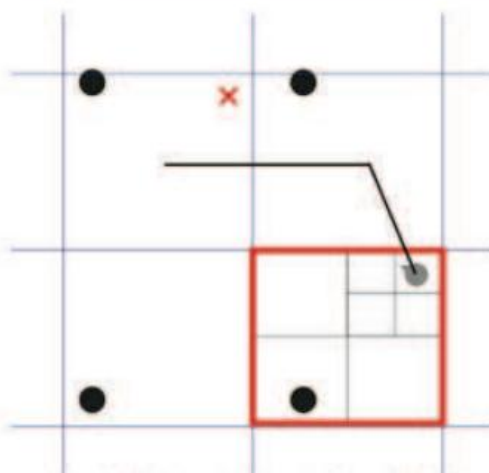
```
while not reached the goal state
  compressTopBelief(top level)
   $a_p = \text{solveTopLevel}(\textit{top level})$ 
  for  $l = 2$  to  $L$ 
     $\textit{whichPOMDP} = \text{selectPOMDP}(l, a_p)$ 
    compressBelief( $l, \textit{whichPOMDP}$ )
     $a_p = \text{solveLevel}(l, \textit{whichPOMDP})$ 
  end
  executeAction( $a_p$ )
   $z = \text{getObservation}()$ 
   $\textit{belief}_L = \text{updateBelief}(\textit{whichPOMDP}, a_p, z)$ 
   $\textit{full belief} = \text{updateFullBelief}(\textit{belief}_L, \textit{whichPOMDP})$ 
end
```

Exemplo

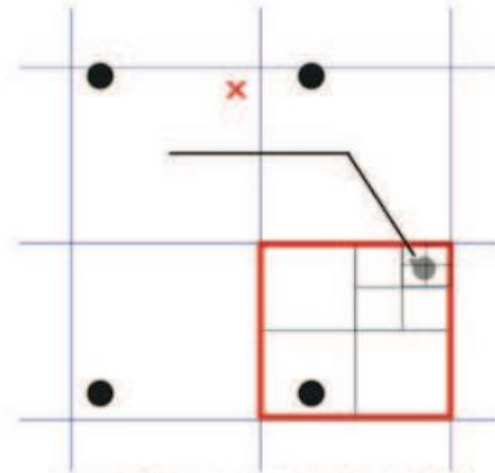
Top Level : Angle resolution 90°



Level 1 : Angle resolution 45°
Action : 135°



Level 2 : Angle resolution 22.5°
Action : 112.5°



Level 3 : Angle resolution 11.25°
Action : 123.75°

Computation time required to solve the RN-HPOMDP with varying grid size and five levels

Grid size	POMDP size		CPU time (s)
5 cm × 5 cm	$ S = 18,411,520$	$ A = 64$	18.520
10 cm × 10 cm	$ S = 4,602,880$	$ A = 64$	0.911
15 cm × 15 cm	$ S = 2,038,080$	$ A = 64$	0.426
20 cm × 20 cm	$ S = 1,150,720$	$ A = 64$	0.257
25 cm × 25 cm	$ S = 734,976$	$ A = 64$	0.262
30 cm × 30 cm	$ S = 503,808$	$ A = 64$	0.251

Computation time required to solve the RN-HPOMDP with varying number of levels and grid size of 10 cm × 10 cm

No. of levels	POMDP size		CPU time (s)
3	$ S = 1,150,720$	$ A = 16$	201.210
4	$ S = 2,301,440$	$ A = 32$	16.986
5	$ S = 4,602,880$	$ A = 64$	0.911
6	$ S = 9,205,760$	$ A = 128$	0.460
7	$ S = 18,411,520$	$ A = 256$	0.411

Computation time required to solve a HPOMDP with the compared approaches

	POMDP size		CPU time (s)
Theocharous [25]	$ S = 575$	$ A = 4$	2.11–5.7
	$ S = 1385$	$ A = 4$	5.05–26.12
Pineau et al. [21]	$ S = 11$	$ A = 6$	2.84
	$ S = 20$	$ A = 30$	77.99