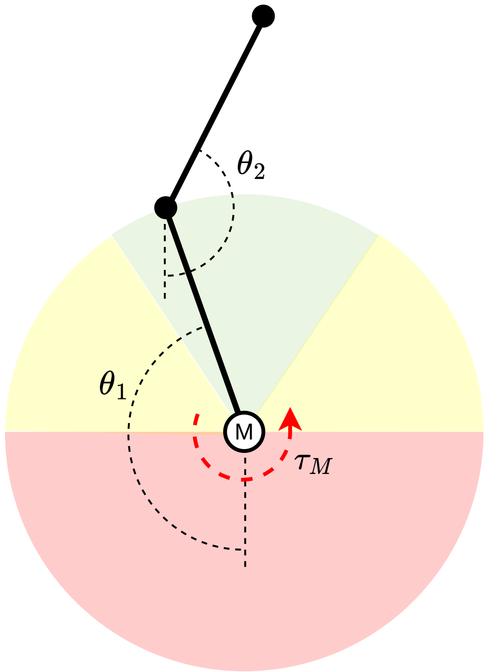

State:

$$s = (\theta_1, \theta_1', \theta_1'', \theta_2, \theta_2', \theta_2'', \tau_M)$$

$$\theta_1, \theta_2 \in [0, 180]$$

Goal state:

$$s_g = (180, 0, 0, 180, 0, 0, 0)$$

Normalized state:

$$s_n = (\theta_{1n}, \theta_{1n}', \theta_{1n}'', \theta_{2n}, \theta_{2n}', \theta_{2n}'', \tau_{Mn})$$

$$\forall 1 \leq i \leq 7 : s_n^{(i)} \in [-1, 1]$$

Normalized goal state:

$$s_{gn} = (1, 0, 0, 1, 0, 0, 0)$$

Reward strategy:

a) Green area (=balancing)

State boundaries: $\theta_{1n} \in]0.8, 1]$

Strategy: eliminate all remaining deltas...

Reward:

$$r_a(t) = |s_{gn} - s_n(t-1)| - |s_{gn} - s_n(t)|$$

b) Yellow area (=swinging up outer pole)

State boundaries: $\theta_{1n} \in]0.5, 0.8]$

Strategy: mainly the state of outer pole is considered (75%), but the angle of inner pole as well (25%)...

State: $s_b = (\theta_{1n}, \theta_{2n}, \theta_{2n}', \theta_{2n}'')$

Goal state: $s_{gb} = (1, 1, 0, 0)$

Reward:

$$r_b(t) = |s_{gb} - s_b(t-1)| - |s_{gb} - s_b(t)|$$

c) Red area (=swinging up)

State boundaries: $\theta_{1n} \in [0, 0.5]$

Strategy: motion of inner pole in one direction is maximized...

Reward:

$$r_c(t) = (\theta_{1n}(t) - \theta_{1n}(t-1)) + (|\theta_{1n}'(t) + \theta_{1n}''(t)| - |\theta_{1n}'(t-1) + \theta_{1n}''(t-1)|)$$