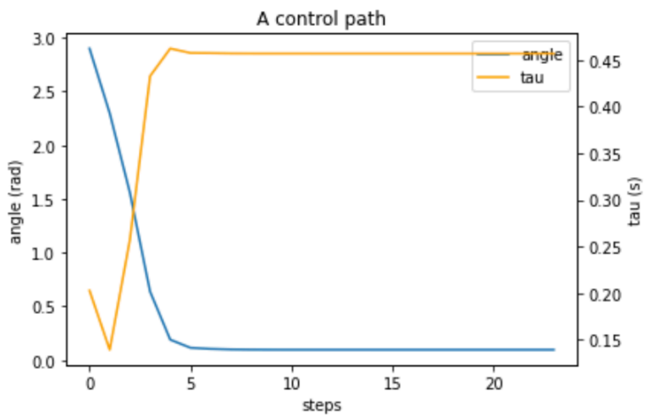
Weekly Report

M2 Ibuki Takeuchi

- Report on last week
  - I could improve policy on RL for optimal self-triggered control



There is no guarantee that this policy is the best policy …

- Wide interval around origin and frequent otherwise
- Stabilize the system

- This week
  - Discuss the next step for master thesis
  - Check that
    - evaluation function for learned policy is larger than that for initial policy
    - approximation accuracy of value function $V^{\pi}(s)$
    - learned policy's dependence for initial policy

Weekly Report

M2 Ibuki Takeuchi

- 1: Comparison of evaluation function
  - Policies

$$\pi_{init}(s) = \begin{bmatrix} lqr(s) \\ 0.2 \end{bmatrix} \quad \text{v.s.} \quad \pi_{RL}(s)\text{: learned policy}$$

  - Evaluation criteria: $J(\pi) = \mathbb{E}_{s_0}[\sum_{i=0}^{\infty} \gamma^i r(s_i, \pi(s_i))]$
  - Result

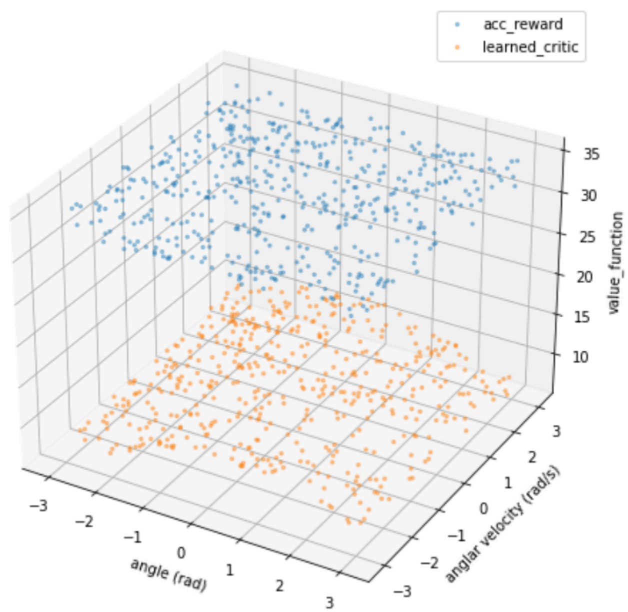
$$J(\pi_{init}) = -14.769 < J(\pi_{RL}) = 45.092$$

Weekly Report

M2 Ibuki Takeuchi

- 2: Approximation accuracy of value function $V^{\pi}(s)$
  - $V^{\pi}(s) = Q^{\pi}(s, \pi(s))$
  - Agent fits $Q(s, a|\omega)$ to approximate $Q^{\pi}(s, a)$
  - Evaluation criteria

Does $Q(s, \pi(s)|\omega)$ approximates $\underbrace{\sum_{i=0}^{\infty} \gamma^{i}\, r(s_i, \pi(s_i))}_{\text{definition of } V^{\pi}(s)}$ well?



Weekly Report

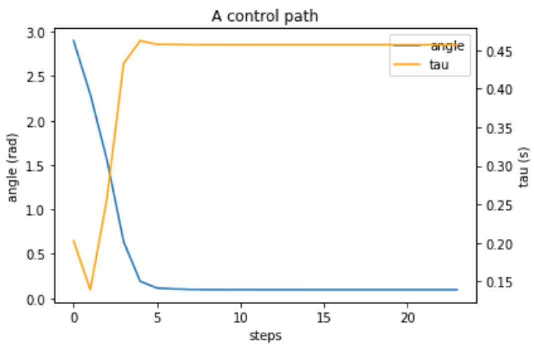
M2 Ibuki Takeuchi

- 3: Learned policy's dependence for initial policy
  - Initial policies

$$\pi_{init}(s): \underbrace{\begin{bmatrix} lqr(s) \\ 0.01 \end{bmatrix}}_{\pi_1}, \underbrace{\begin{bmatrix} lqr(s) \\ 0.1 \end{bmatrix}}_{\pi_2}, \underbrace{\begin{bmatrix} lqr(s) \\ 0.5 \end{bmatrix}}_{\pi_3}, \underbrace{\begin{bmatrix} lqr(s) \\ 1.0 \end{bmatrix}}_{\pi_4}$$

  - 3 patterns of learning
    - adaptive interval and stabilizing: π_2, π_3, π_4

      ※interval around origin point is different



    - constant interval (minimum) and stabilizing
    - constant interval (minimum) and unstabilizing

    } π_1, π_2, π_4