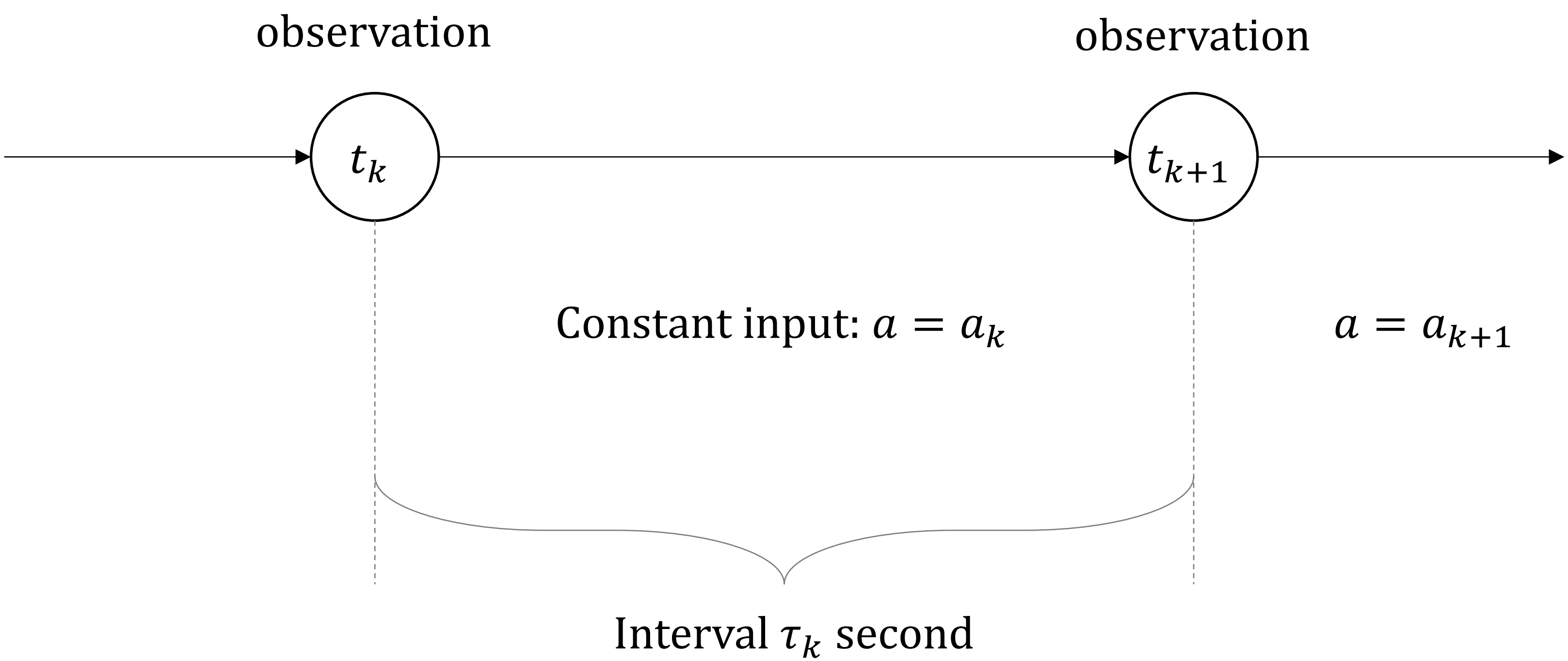
Weekly Report

M2 Ibuki Takeuchi

- Self-triggered control (cf. : Event-triggered control)
  - Controller decides action signal and next observation time
  - (State feedback) control low function: $\pi(s) = [a(s) \quad \tau(s)]$



Weekly Report

M2 Ibuki Takeuchi

- Reinforcement learning for optimal self-triggered control π^*

$$\pi^*(s) = \mathrm{argmax}_\pi\, J(\pi)$$

$$J(\pi) = \mathbb{E}_{s_0}[V^\pi(s_0)]$$

$$V^\pi(s) = \sum_{i=0}^{\infty} \gamma^i r_i^\pi$$

$$r_i^\pi = -\int_{t_k}^{t_{k+1}} s(t)^T Q s(t) dt - \tau_i a_i^T R a_i + \lambda \tau_i$$

- Assume every *i-th* step's access to data tuple (s_t, r_i^π)
- Assume the system to be control affine

$$\dot{s} = f(s) + g(s)a$$

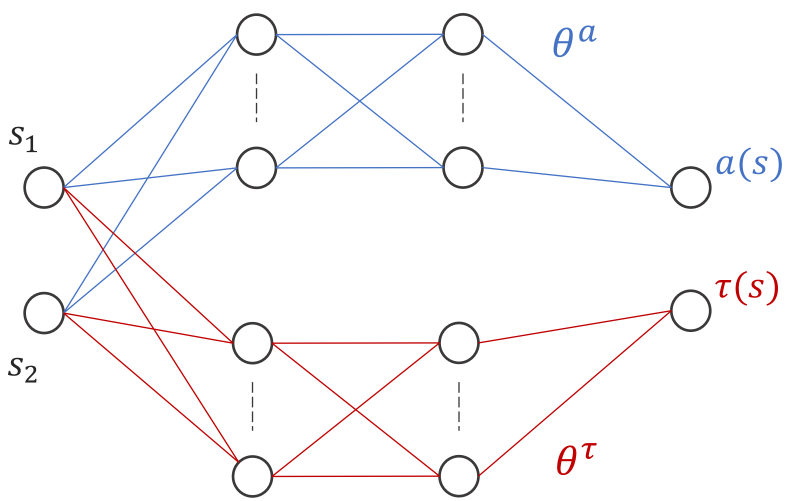
Weekly Report

M2 Ibuki Takeuchi

- How to get π^* ?
  - Parametrize the function as π_{θ^π} (express with NN)
  - Use gradient method to update NN, with policy gradient $\nabla_{\theta^\pi} J(\theta^\pi)$

$\nabla_{\theta^\pi} J(\pi_{\theta^\pi})$

- Policy function has 2 elements : $\pi(s) = [a(s) \quad \tau(s)]$
  - Divide policy network into 2 parametrized function



- θ^π expresses the combined network

$$\theta^\pi = [\theta^a \quad \theta^\tau]$$

Weekly Report

M2 Ibuki Takeuchi

- Theme for master thesis
  - Research the relation between $\nabla_{\theta^a} J(\theta^\pi), \nabla_{\theta^\tau} J(\theta^\pi)$ limited to this problem (RL for self-triggered control)

- Analytical calculation
  - By the definition of $J(\theta^\pi)$, focus on $\nabla_\theta V^{\theta^\pi}(s)$
  - $\nabla_\theta V^{\theta^\pi}(s) = \nabla_\theta [r\big(s, \pi(s|\theta^\pi)\big) + \gamma V^{\theta^\pi}\big(s'(\theta^\pi)\big)]$
    $= \nabla_\theta r\big(s, \pi(s|\theta^\pi)\big) + \gamma \nabla_\theta \{V^{\theta^\pi}\big(s'(\theta^\pi)\big)\}$

  - 1st element is gradient for step reward

$$\nabla_{\theta^a} r\big(s, \pi(s|[\theta^a, \theta^\tau])\big) = \nabla_{\theta^a} a(s|\theta^a) \nabla_a r(s, [a\ \tau])|_{a=a(s|\theta^a),\, \tau=\tau(s|\theta^\tau)}$$
$$\nabla_{\theta^\tau} r\big(s, \pi(s|[\theta^a, \theta^\tau])\big) = \nabla_{\theta^\tau} \tau(s|\theta^\tau) \nabla_\tau r(s, [a\ \tau])|_{a=a(s|\theta^a),\, \tau=\tau(s|\theta^\tau)}$$

  - Calculate $\nabla_a r(s,u)|_{u=\pi(s|\theta^\pi)}, \nabla_\tau r(s,u)|_{u=\pi(s|\theta^\pi)}$

Weekly Report

M2 Ibuki Takeuchi

- Analytical calculation
  - $\nabla_\theta V^{\pi_\theta}(s) = \nabla_\theta r\big(s, \pi(s|\theta^\pi)\big) + \gamma \nabla_\theta \{V^{\theta^\pi}\big(s'(\theta^\pi)\big)\}$
  - 2nd element is gradient for V^{θ^π} at next state
  - When parameter θ changes
    - $V^{\theta^\pi}(\cdot)$ changes
    - $s'(\theta^\pi)$ changes
  - How to calculate this gradient …?

$$\gamma \nabla_\theta \{f^\theta\big(g(\theta)\big)\} = \lim_{h \to 0} \frac{f^{\theta+h}\big(g(\theta+h)\big) - f^\theta\big(g(\theta+h)\big)}{h}$$

$$+ \lim_{h \to 0} \frac{f^\theta\big(g(\theta+h)\big) - f^\theta\big(g(\theta)\big)}{h}$$