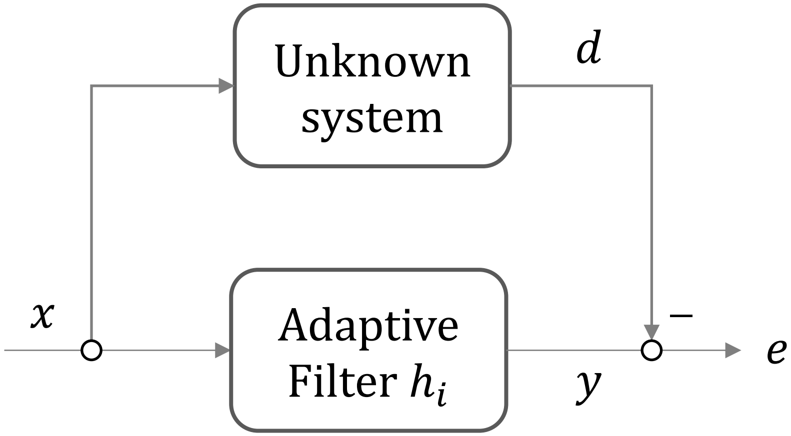
Weekly Report

M2 Ibuki Takeuchi

- Correction of colloquium
  - Recursive Least Square



$$\begin{cases} y_t = \sum_{i=0}^{k-1} h_i x_{t-i} = \boldsymbol{h}^T \boldsymbol{x}_t \\ e_t = d_t - y_t \end{cases}$$

$$\boxed{\boldsymbol{h}_{opt} = \underset{\boldsymbol{h}}{argmin} \sum_{i=1}^{t} {e_i}^2}$$

Weekly Report

M2 Ibuki Takeuchi

- Update coefficient recursively in every timesteps as following:

RLS update

$$k_t = \frac{P_t \boldsymbol{x}_{t+1}}{1 + \boldsymbol{x}_{t+1}{}^T P_t \boldsymbol{x}_{t+1}}$$

$$P_{t+1} = [I - k_t \boldsymbol{x}_t{}^T] P_t$$

$$\boldsymbol{h}_{t+1} = \boldsymbol{h}_t + k_t (d_t - y_t)$$

$$\boldsymbol{h}_t = (\boldsymbol{x}_t \boldsymbol{x}_t{}^T)^{-1} \boldsymbol{x}_t d_t$$: LS solution

⇩ correction

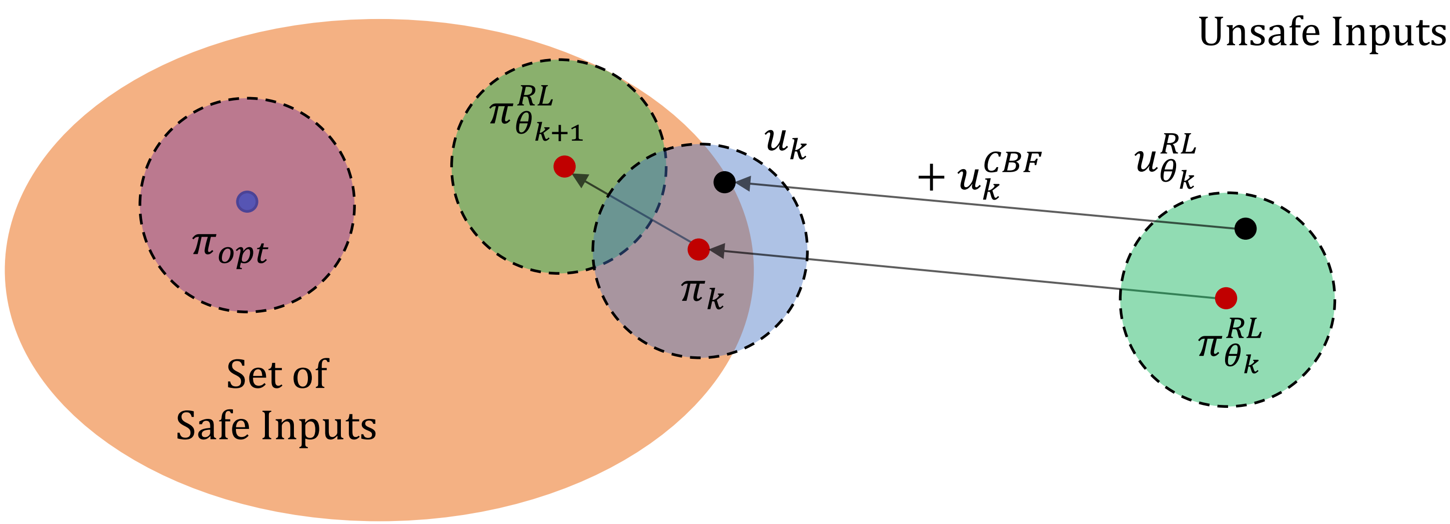
$$\boldsymbol{h}_t = (\boldsymbol{X}_t \boldsymbol{X}_t{}^T)^{-1} \boldsymbol{X}_t \boldsymbol{d}_t$$

$$\boldsymbol{X}_t = [\boldsymbol{x}_1, \cdots, \boldsymbol{x}_t]$$
$$\boldsymbol{d}_t = [d_1, \cdots, d_t]^T$$

Weekly Report

M2 Ibuki Takeuchi

- Study theme
  - Recall : safe reinforcement learning



Weekly Report

M2 Ibuki Takeuchi

- Problem formulation
  - Nominal control affine model

$$s_{t+1} = f(s_t) + g(s_t)a_t + d(s_t)$$

where f, g are known and d is uncertain

  - We want to search optimal policy maintaining safety

$$\max \mathbb{E}(\sum_{t=0}^{\infty} \gamma^t r(s_t, a_t))$$

Weekly Report

M2 Ibuki Takeuchi

- Control Barrier Function (CBF)
  - Safe set C

$$C : \{s \in \mathbb{R}^n : h(s) \geq 0\}$$

  - To maintain safety during learning process, the set above must be forward invariant.

  - Condition for function $h(s)$ to be CBF

$$\exists \eta \in [0,1], \forall s_t \in C$$

$$\sup_{a_t \in A} [(h(f(s_t) + g(s_t)a_t + d(s_t)) + (\eta - 1)h(s_t))] \geq 0$$

  - If $h(s)$ is a CBF, it is certified that safe inputs exist[1]

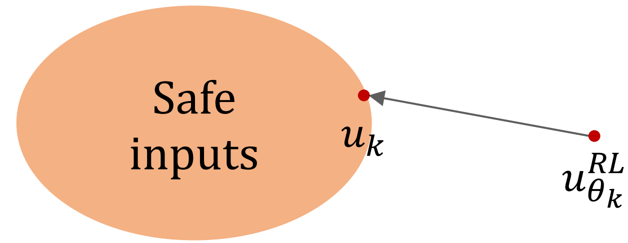
[1] : A. D. Ames, X. Xu, J. W. Grizzle and P. Tabuada. "Control Barrier Function Based Quadratic Programs for Safety Critical Systems." in IEEE Transactions on Automatic Control, vol. 62, no. 8, pp/ 3861-3876, 2017.

Weekly Report

M2 Ibuki Takeuchi

- Construct safe input using CBF[2]

$$u_k(s) = u_{\theta_k}^{RL}(s) + u_k^{CBF}(s, u_{\theta_k}^{RL})$$



- u_k^{CBF} is a solution of

$$h(s) = p^T s + q$$

$$\begin{aligned}
(a_t, \varepsilon) = \underset{a_t, \varepsilon}{\operatorname{argmin}} \quad & \|a_t\|_2 + K_\varepsilon \varepsilon \\
\text{s.t.} \quad & p^\top f(s_t) + p^\top g(s_t)\left(u_{\theta_k}^{RL}(s) + a_t\right) + p^\top \mu_d(s_t) \\
& - k_\delta |p|^\top \sigma_d(s_t) + q \geq (1 - \eta) h(s_t) - \varepsilon \\
& a_{low}^i \leq u_{\theta_k}^{RL(i)}(s) + a_t \leq a_{high}^i \text{ for } i = 1, \ldots, M
\end{aligned}$$

[2] : R. Cheng, G. Orosz, R. M. Murray, and J. W. Burdick. " End-to-end safe reinforcement learning through barrier functions for safety-critical continuous control tasks. " *Thirty-Third AAAI Conference on Artificial Intelligence (AAAI-19),* 2019.

Weekly Report

M2 Ibuki Takeuchi

- Related research
  - J. Achiam, D. Held, A. Tamar and P. Abbeel. "Constrained Policy Optimization." *Proceedings of the 32nd International Conference on Machine Learning,* vol. 70, pp. 22-31, 2017.
  - Chow et.al, "A Lyapunov-based Approach to Safe Reinforcement Learning. " *32nd Conference on Neural Information Processing Systems*, pp. 8103-8112, 2018.