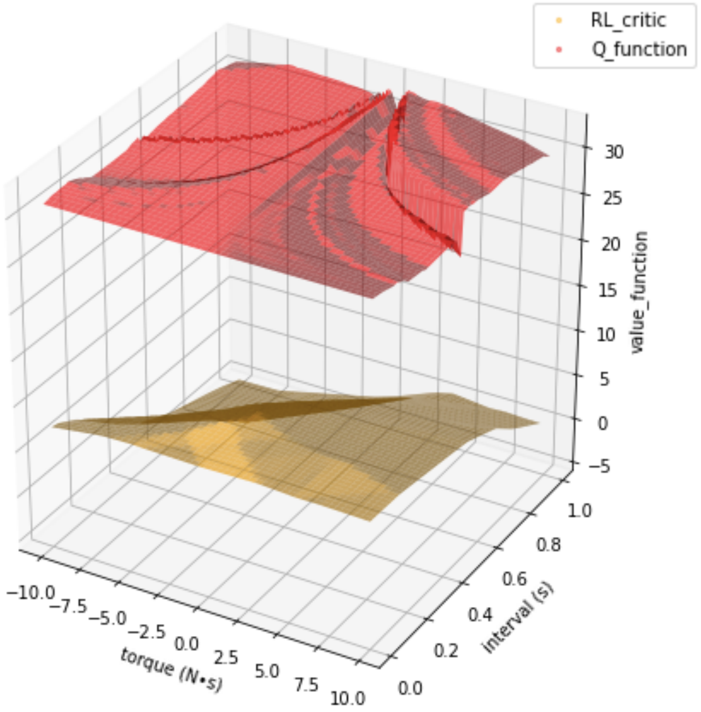
Weekly Report

M2 Ibuki Takeuchi

- Low accuracy of Q-function approximation

Graph of $Q(s, a)$ for fixed s

(function of a)



- Q-function are learned with supervised learning (Least Square)

$$\min_{\omega} \frac{1}{N} \sum_{(s,a)\in E} \{Q(s,a|\omega) - \underbrace{(r(s,a) + \gamma Q(s,\pi(s)|\omega))}_{\text{teacher data}}\}^2$$

- Low accuracy may come from..
  - Optimization algorithm
  - Data bias in E

(This may be the reason)

Weekly Report

M2 Ibuki Takeuchi

- Check as if optimization is well conducted
  - Define ω_{RL} be the Q-function parameter learned with RL
  - Loss function for data set E should be minimized by ω_{RL}

$$Loss(\omega) = \frac{1}{N} \sum_{(s,a)\in E} \{Q(s,a|\omega) - (r(s,a) + \gamma Q(s,\pi(s)|\omega))\}^2$$

  - By comparing the loss function with some ω

$$Loss(\omega) = 17, Loss(\omega_{RL}) = 0.02$$

  - I want to check again on what should be compared, with Kashima Sensei

Weekly Report

M2 Ibuki Takeuchi

- Overcome data bias
  - Various experiences of (s, a) are needed
  - Teacher‘s data are collected by agent‘s experience

  - How does the agent collect data?
    - Data exploration:

$$a = \pi_\theta(s) + e$$

    - Store data of (s, a) to set E

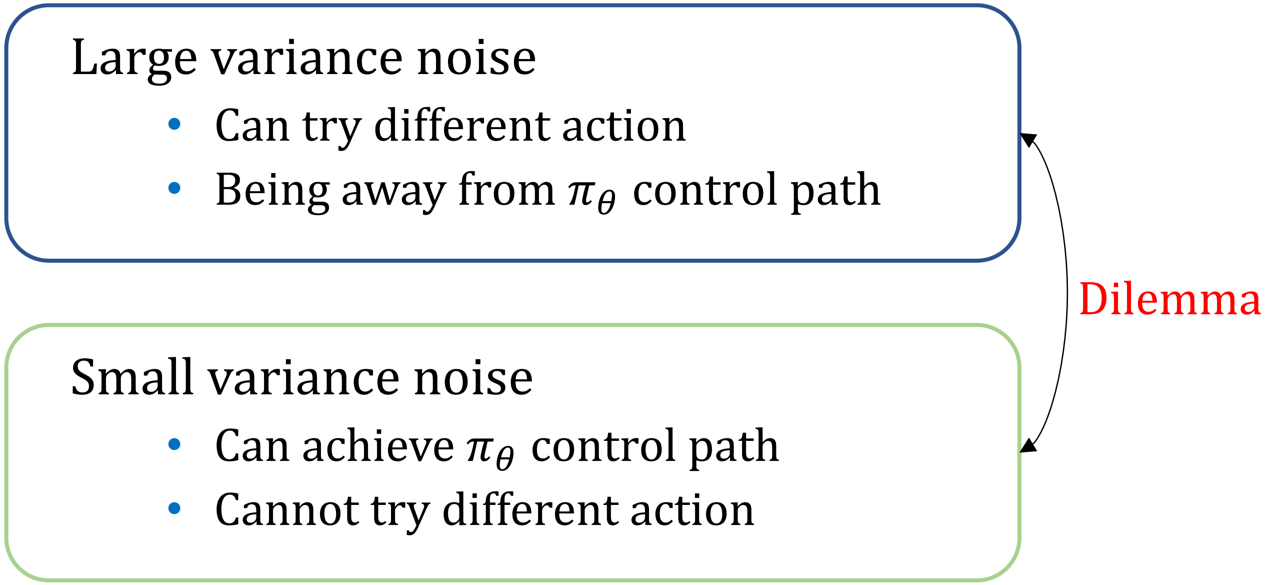
    - If the variance of e is large, that of (s, a) become large

Weekly Report

M2 Ibuki Takeuchi

- DDPG requires agent to experience states on control path w.r.t. current policy π_θ

- Exploration-Exploitation Dilemma



Large variance noise

- Can try different action
- Being away from π_θ control path

Small variance noise

- Can achieve π_θ control path
- Cannot try different action

Dilemma

Weekly Report

M2 Ibuki Takeuchi

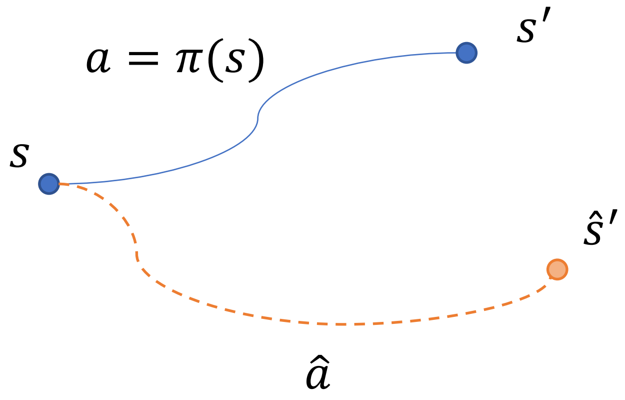
- Adaptive noise scaling
  - By changing action…

if the change of next state is large ⇔ $\frac{\partial s'}{\partial a}$ is Large

→ small noise

if the change of next state is small ⇔ $\frac{\partial s'}{\partial a}$ is Small

→ large noise



Weekly Report

M2 Ibuki Takeuchi

- Try following noise scaling

$$\frac{c}{\|g\| + c} \times \mathcal{N}(0,1) \quad : c \text{ is hyper parameter}$$

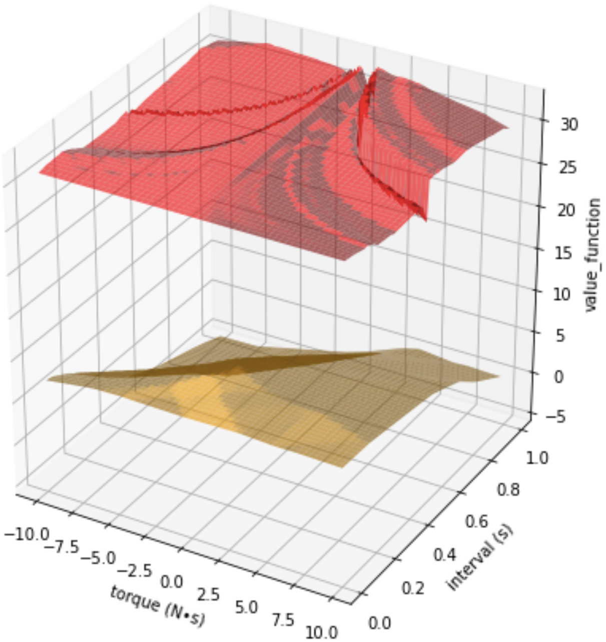
- Because I have not summarized my consideration, I want to report the result on colloquium next week

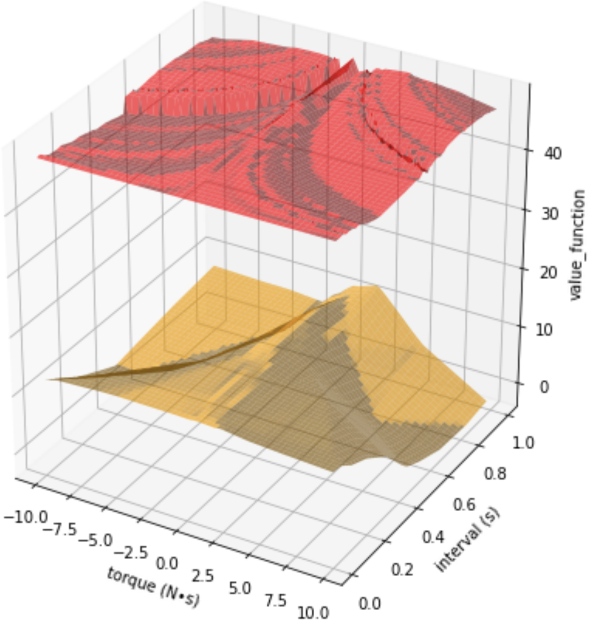
Weekly Report

M2 Ibuki Takeuchi

- Try noise

$$\frac{c}{\|g\| + c} \times \mathcal{N}(0,1) \quad : c \text{ is hyper parameter}$$


value_function
torque (N•s)
interval (s)


value_function
torque (N•s)
interval (s)