

# Review of Generative Adversarial Imitation Learning

Guannan Hu

April 11, 2018

---

**Algorithm 1** Generative adversarial imitation learning

---

- 1: **Input:** Expert trajectories  $\tau_E \sim \pi_E$ , initial policy and discriminator parameters  $\theta_0, w_0$
- 2: **for**  $i = 0, 1, 2, \dots$  **do**
- 3:     Sample trajectories  $\tau_i \sim \pi_{\theta_i}$
- 4:     Update the discriminator parameters from  $w_i$  to  $w_{i+1}$  with the gradient

$$\hat{\mathbb{E}}_{\tau_i}[\nabla_w \log(D_w(s, a))] + \hat{\mathbb{E}}_{\tau_E}[\nabla_w \log(1 - D_w(s, a))] \quad (1)$$

- 5:     Take a policy step from  $\theta_i$  to  $\theta_{i+1}$ , using the TRPO rule with cost function  $\log(D_{w_{i+1}}(s, a))$ . Specifically, take a KL-constrained natural gradient step with

$$\begin{aligned} & \hat{\mathbb{E}}_{\tau_i}[\nabla_{\theta} \log \pi_{\theta}(a|s)Q(s, a)] - \lambda \nabla_{\theta} H(\pi_{\theta}) \\ & \text{where } Q(\bar{s}, \bar{a}) = \hat{\mathbb{E}}_{\tau_i}[\log(D_{w_{i+1}}(s, a)) | s_0 = \bar{s}, a_0 = \bar{a}] \end{aligned} \quad (2)$$

- 6: **end for**
-