



Disturber Policy Gradient

$$C_t = r_{max}^{task} - r_t^{task}$$

$$d = \operatorname{argmax}_d L_{disturber}(d)$$

H_∞ Policy Gradient

$$\pi = \operatorname{argmax}_\pi L_{actor}^{PPO}(\pi) + \lambda \cdot L^{Hinf}(\pi)$$

$$\lambda = \lambda - \alpha \cdot L^{Hinf}(\pi)$$