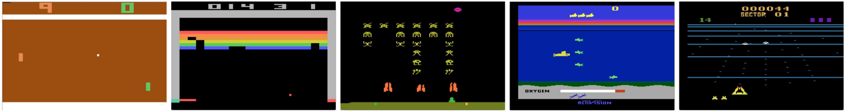
A Game Theoretic Analysis of LQG Control under Adversarial Attack

Zuxing Li, György Dán, and Dong Liu

Motivation

- Deep reinforcement learning[1]
  - DNN function approximator for complex control tasks
  - Wide-range of promising applications
  - Inherits vulnerability of DNN[2,3]
- Need for adversarial reinforcement learning



[1] Mnih et al., 2015. Human-level control through deep reinforcement learning.

[2] Szegedy et al., 2013. Intriguing properties of neural networks.

[3] Huang et al., 2016. Adversarial attacks on neural network policies.

State of the art

- Attack techniques: Generate adversarial examples[3-5]
- Defense techniques: Use perturbations in the training[6]
- Game formulations: Capture the strategic interaction
  - Variants of stochastic game[6-8]
  - Stackelberg game + POMDP or LQG[9,10]
  - Cheap talk game + Linear dynamic system[11]

---

[3] Huang et al., 2016. Adversarial attacks on neural network policies.

[4] Lin et al., 2017. Tactics of adversarial attack on deep reinforcement learning agents.

[5] Behzadan and Munir, 2017. Vulnerability of deep reinforcement learning to policy induction attacks.

[6] Pinto et al., 2017. Robust adversarial reinforcement learning.

[7] Horák et al., 2017. Manipulating adversary's belief: A dynamic game approach to deception by design for proactive network security.

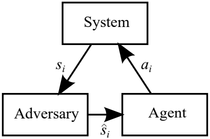
[8] Gleave et al., 2020. Adversarial policies: Attacking deep reinforcement learning.

[9] Osogami, 2015. Robust partially observable Markov decision process.

[10] Sayin et al., 2019. Hierarchical multistage Gaussian signaling games in noncooperative communication and control systems.

[11] Saritas et al., 2017. Nash and Stackelberg equilibria for dynamic cheap talk and signaling games.

Adversarial LQG control



- N-stage LQG:

$$s_{i+1} = \alpha_i s_i + \beta_i a_i + z_i, \text{ given } \alpha_i \neq 0, \ \beta_i \neq 0$$
$$\hat{s}_i = \pi_i s_i + c_i$$
$$a_i = \kappa_i \hat{s}_i + \rho_i$$
$$r_i = R_i(s_i, a_i) = -\theta_i s_i^2 - \phi_i a_i^2, \text{ given } \theta_i > 0, \ \phi_i > 0$$
$$S_1 \sim b_1 \triangleq \mathcal{N}(\mu_1, \sigma_1^2), \text{ given } \mu_1, \ \sigma_1^2 > 0$$
$$Z_i \sim \mathcal{N}(0, \omega_i^2), \text{ given } \omega_i^2 > 0$$
$$C_i \sim \mathcal{N}(0, \delta_i^2)$$

Adversarial LQG control

- Belief of the agent in the beginning of i-th stage $b_i \triangleq \mathcal{N}(\mu_i, \sigma_i^2)$: Posterior distribution of S_i after observing $\{(\hat{s}_k, a_k)\}_{k=1}^{i-1}$
- Belief update $b_i \to b_{i+1}$:

$$\mu_{i+1} = \Lambda_\mu(b_i, \pi_i, \delta_i^2, \hat{s}_i, a_i) = \alpha_i \frac{\pi_i \sigma_i^2 \hat{s}_i + \mu_i \delta_i^2}{\pi_i^2 \sigma_i^2 + \delta_i^2} + \beta_i a_i$$

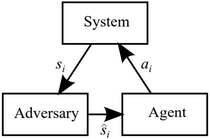
$$\sigma_{i+1}^2 = \Lambda_\nu(b_i, \pi_i, \delta_i^2) = \frac{\alpha_i^2 \sigma_i^2 \delta_i^2}{\pi_i^2 \sigma_i^2 + \delta_i^2} + \omega_i^2$$

- Adversarial manipulation constraints:

$$-\infty < \varepsilon' \leq \pi_i \leq \varepsilon < \infty, \text{ given } \varepsilon', \ \varepsilon$$

$$I(\hat{S}_i; S_i) = \frac{1}{2} \log \frac{\pi_i^2 \sigma_i^2 + \delta_i^2}{\delta_i^2} \geq \frac{1}{2} \log \lambda > 0, \text{ given } \lambda > 1$$

Adversarial LQG control

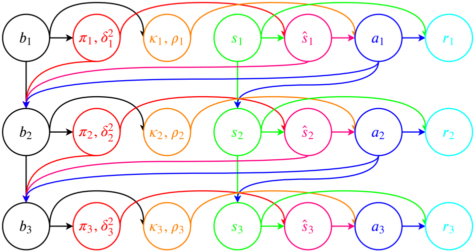


- Asymmetric information
  - The adversary manipulates the system states.
  - The agent chooses actions based on the manipulated observations.
- Conflicting objectives
  - The agent aims at improving the control reward.
  - The adversary aims at degrading the control reward.

Question

How to formulate the interaction of players with asymmetric information in an LQG control?

Adversarial LQG game



- In the beginning of the i-th stage
  - Adversarial strategy: $g_i(\pi_i, \delta_i^2 | b_i)$ or $(\pi_i, \delta_i^2) = g_i(b_i)$
  - Agent strategy: $f_i(\kappa_i, \rho_i | b_i)$ or $(\kappa_i, \rho_i) = f_i(b_i)$
- Running follows the LQG equations
- In the end of the i-th stage
  - Adversary reveals the chosen parameters (π_i, δ_i^2) to the agent
  - Both players update the next belief b_{i+1}

Subgame perfect equilibrium

- Strategies (g^{N*}, f^{N*}) form an SPE.
- Value function of a subgame starting from the i-th stage:

$$V_i^N(b_i) = E_{b_i, g_i^{N*}, f_i^{N*}} \left(\sum_{j=i}^{N} R_j(S_j, A_j) \right)$$

- Backward dynamic programming:

$$\begin{aligned} V_i^N(b_i) &= \min_{g_i} E_{b_i, g_i, f_i^*} \{ R_i(S_i, A_i) \\ &\quad + V_{i+1}^N(\mathcal{N}(\Lambda_\mu(b_i, \Pi_i, \Delta_i^2, \hat{S}_i, A_i), \Lambda_\nu(b_i, \Pi_i, \Delta_i^2))) \} \\ &= \max_{f_i} E_{b_i, g_i^*, f_i} \{ R_i(S_i, A_i) \\ &\quad + V_{i+1}^N(\mathcal{N}(\Lambda_\mu(b_i, \Pi_i, \Delta_i^2, \hat{S}_i, A_i), \Lambda_\nu(b_i, \Pi_i, \Delta_i^2))) \} \end{aligned}$$

Single-stage ALQG game

Proposition 1

Let $N = 1$. An SPE *always exists* and consists of (f_1^*, g_1^*), where $(\kappa_1^*, \rho_1^*) = f_1^*(b_1) = (0, 0)$ for any belief b_1; and g_1^* can be any adversarial strategy subject to constraints.

Pure strategy equilibria

Theorem 1

Let $N \geq 2$. If $\varepsilon' \neq \varepsilon$ or if $\varepsilon' = \varepsilon = 0$, then there is *no* pure strategy SPE for the ALQG game. If $\varepsilon' = \varepsilon \neq 0$, then there is a *unique* pure strategy SPE. The SPE strategies for $1 \leq i \leq N$ are given by

$$\tilde{\theta}_{N+1} = \hat{\theta}_{N+1} = 0;$$

$$\tilde{\theta}_i = \theta_i + \tilde{\theta}_{i+1}\alpha_i^2 - \frac{\tilde{\theta}_{i+1}^2 \alpha_i^2 \beta_i^2}{\phi_i + \tilde{\theta}_{i+1}\beta_i^2};$$

$$\hat{\theta}_i = \theta_i + \hat{\theta}_{i+1}\alpha_i^2 - \left(\frac{\tilde{\theta}_{i+1}^2 \alpha_i^2 \beta_i^2}{\phi_i + \tilde{\theta}_{i+1}\beta_i^2} + (\hat{\theta}_{i+1} - \tilde{\theta}_{i+1})\alpha_i^2 \right) \frac{\lambda - 1}{\lambda};$$

$$(\pi_i^*, \delta_i^{2*}) = g_i^*(b_i) = \left(\varepsilon, \frac{\varepsilon^2 \sigma_i^2}{\lambda - 1} \right);$$

$$(\kappa_i^*, \rho_i^*) = f_i^*(b_i) = \left(-\frac{\tilde{\theta}_{i+1}\alpha_i\beta_i(\lambda - 1)}{(\phi_i + \tilde{\theta}_{i+1}\beta_i^2)\lambda\varepsilon}, -\frac{\tilde{\theta}_{i+1}\alpha_i\beta_i\mu_i}{(\phi_i + \tilde{\theta}_{i+1}\beta_i^2)\lambda} \right).$$

Pure strategy equilibria

Corollary 1

If $\varepsilon' = \varepsilon \neq 0$, the value function induced by the unique pure strategy SPE is

$$V_i^N(b_i) = -\tilde{\theta}_i \mu_i^2 - \hat{\theta}_i \sigma_i^2 - \sum_{j=i+1}^{N} \hat{\theta}_j \omega_{j-1}^2.$$

Observations

- A rational adversary will always apply a manipulation with the largest variance.
- The value function V_i^N consists of two separable terms of μ_i and σ_i^2.

Time-invariant model

- Time-invariant parameters: $\alpha_i = \alpha \neq 0$, $\beta_i = \beta \neq 0$, $\omega_i^2 = \omega^2 > 0$, $\theta_i = \theta > 0$, and $\phi_i = \phi > 0$ for $i \geq 1$
- Define the mapping $L : \mathbb{R}^2_{\geq 0} \to \mathbb{R}^2_{\geq 0}$ as

$$L\left(x, y\right) = \left(\theta + \frac{\phi\alpha^2 x}{\phi + \beta^2 x}, \theta + \frac{\phi\alpha^2 x}{\phi + \beta^2 x}\frac{\lambda - 1}{\lambda} + \alpha^2 y \frac{1}{\lambda}\right).$$

Proposition 2

Let $\lambda > \alpha^2$. Then the mapping L admits a least fixed point $(\tilde{\theta}, \hat{\theta}) \in \mathbb{R}^2_{\geq 0}$, for which

$$\lim_{n\to\infty} L^n(0, 0) = L(\tilde{\theta}, \hat{\theta}) = (\tilde{\theta}, \hat{\theta}).$$

Time-invariant model

Theorem 2

Let $\lambda > \alpha^2$, $\varepsilon' = \varepsilon \neq 0$, and $N \to \infty$. Then the ALQG game of the time-invariant model has a stationary SPE in pure strategies as: For $i \geq 1$,

$$\left(\pi_i^*, \delta_i^{2*}\right) = g_i^*(b_i) = \left(\varepsilon, \frac{\varepsilon^2 \sigma_i^2}{\lambda - 1}\right);$$

$$(\kappa_i^*, \rho_i^*) = f_i^*(b_i) = \left(-\frac{\tilde{\theta}\alpha\beta(\lambda-1)}{(\phi+\tilde{\theta}\beta^2)\lambda\varepsilon}, -\frac{\tilde{\theta}\alpha\beta\mu_i}{(\phi+\tilde{\theta}\beta^2)\lambda}\right).$$

Corollary 2

Let $b_1 \triangleq \mathcal{N}(\mu_1, \sigma_1^2)$ with *bounded* mean and variance. For the stationary SPE in pure strategies, the expected average reward per stage in steady state is *independent* of the initial belief:

$$\lim_{N\to\infty} \frac{V_1^N(b_1)}{N} = -\hat{\theta}\omega^2.$$

Behavioral strategy equilibria

Theorem 3

Let $N \geq 2$, $\varepsilon' < 0 < \varepsilon$, and $\check{\theta}_{N+1} = 0$. Then for $1 \leq i \leq N$,

$$\check{\theta}_i = \theta_i + \check{\theta}_{i+1}\alpha_i^2 - (\check{\theta}_{i+1} - \tilde{\theta}_{i+1})\alpha_i^2 \frac{\lambda - 1}{\lambda}.$$

There is a continuum of SPEs in behavioral strategies. Each SPE in the i-th stage consists of a behavioral strategy g_i^* and a pure strategy f_i^* satisfying

$$\mathbb{S}(g_i^*|b_i) \triangleq \left\{(\pi_i, \delta_i^2) : \pi_i \neq 0, \varepsilon' \leq \pi_i \leq \varepsilon, \delta_i^2 = \frac{\pi_i^2\sigma_i^2}{\lambda - 1}\right\};$$

$$||\mathbb{S}(g_i^*|b_i)|| \geq 2;$$

$$E_{g_i^*}(\Pi_i) = 0;$$

$$(\kappa_i^*, \rho_i^*) = f_i^*(b_i) = \left(0, -\frac{\tilde{\theta}_{i+1}\alpha_i\beta_i\mu_i}{\phi_i + \tilde{\theta}_{i+1}\beta_i^2}\right).$$

Behavioral strategy equilibria

Corollary 3

Let $\varepsilon' < 0 < \varepsilon$. For any SPE in behavioral strategies, we have

$$V_i^N(b_i) = -\tilde{\theta}_i \mu_i^2 - \check{\theta}_i \sigma_i^2 - \sum_{j=i+1}^{N} \check{\theta}_j \omega_{j-1}^2.$$

Observations

- It is sufficient for the agent to use a pure strategy.
- Although the adversary cannot use $\pi_i = 0$, the behavioral strategy g_i^* needs to achieve zero-mean of the random coefficient Π_i.
- A rational adversary will always use a manipulation with the largest variance.
- The value function V_i^N consists of two separable terms of μ_i and σ_i^2.
- Stronger adversary $\Rightarrow$ The value function of an SPE in behavioral strategies $\leq$ The value function of a pure strategy SPE.

Time-invariant model

- Define the mapping $J : \mathbb{R}^2_{\geq 0} \to \mathbb{R}^2_{\geq 0}$ as

$$J(x, y) = \left(\theta + \frac{\phi \alpha^2 x}{\phi + \beta^2 x}, \theta + \alpha^2 x \frac{\lambda - 1}{\lambda} + \alpha^2 y \frac{1}{\lambda} \right).$$

Proposition 3

Let $\lambda > \alpha^2$. Then the mapping J admits a least fixed point $(\tilde{\theta}, \check{\theta}) \in \mathbb{R}^2_{\geq 0}$, for which

$$\lim_{n \to \infty} J^n(0, 0) = J(\tilde{\theta}, \check{\theta}) = (\tilde{\theta}, \check{\theta}).$$

Time-invariant model

Theorem 4

Let $\lambda > \alpha^2$, $\varepsilon' < 0 < \varepsilon$, and $N \to \infty$. Then the ALQG game of the time-invariant model has a stationary SPE in behavioral strategies as: For $i \geq 1$,

$$g_i^* \left(\pi_i = \varepsilon', \, \delta_i^2 = \frac{\varepsilon'^2 \sigma_i^2}{\lambda - 1} \middle| \, b_i \right) = \frac{\varepsilon}{\varepsilon - \varepsilon'};$$

$$g_i^* \left(\pi_i = \varepsilon, \, \delta_i^2 = \frac{\varepsilon^2 \sigma_i^2}{\lambda - 1} \middle| \, b_i \right) = -\frac{\varepsilon'}{\varepsilon - \varepsilon'};$$

$$(\kappa_i^*, \rho_i^*) = f_i^*(b_i) = \left(0, -\frac{\tilde{\theta}\alpha\beta\mu_i}{\phi + \tilde{\theta}\beta^2} \right).$$

Corollary 4

Let $b_1 \triangleq \mathcal{N}(\mu_1, \sigma_1^2)$ with *bounded* mean and variance. For the stationary SPE in behavioral strategies, the expected average reward per stage in steady state is

$$\lim_{N \to \infty} \frac{V_1^N(b_1)}{N} = -\check{\theta}\omega^2.$$

Other cases

- Theorem 5: Let $N \geq 2$. If $0 = \varepsilon' < \varepsilon$ or if $\varepsilon' < \varepsilon = 0$, there is *no* SPE for the ALQG game.
- Theorem 6: Let $N = 2$. If $0 < \varepsilon' < \varepsilon$ or if $\varepsilon' < \varepsilon < 0$, there is a *unique* SPE in behavioral strategies for the ALQG game: For any belief $b_1 \triangleq \mathcal{N}(\mu_1, \sigma_1^2)$,

$$g_1^* \left(\pi_1 = \varepsilon', \, \delta_1^2 = \frac{\varepsilon'^2 \sigma_1^2}{\lambda - 1} \middle| \, b_1 \right) = \frac{\varepsilon}{\varepsilon' + \varepsilon};$$

$$g_1^* \left(\pi_1 = \varepsilon, \, \delta_1^2 = \frac{\varepsilon^2 \sigma_1^2}{\lambda - 1} \middle| \, b_1 \right) = \frac{\varepsilon'}{\varepsilon' + \varepsilon};$$

$$\kappa_1^* = f_1^*(b_1) = \frac{-\theta_2 \alpha_1 \beta_1 E_{g_1^*}(\Pi_1) \sigma_1^2}{(\phi_1 + \theta_2 \beta_1^2) \left(E_{g_1^*}(\Pi_1^2) \left(\mu_1^2 + \frac{\lambda}{\lambda - 1} \sigma_1^2 \right) - E_{g_1^*}^2(\Pi_1) \mu_1^2 \right)};$$

$$\rho_1^* = f_1^*(b_1) = -E_{g_1^*}(\Pi_1) \mu_1 \kappa_1^* - \frac{\theta_2 \alpha_1 \beta_1 \mu_1}{\phi_1 + \theta_2 \beta_1^2};$$

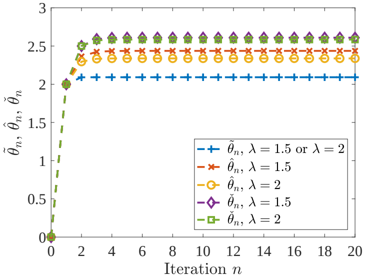
Other cases

$$V_1^2(b_1) = -\left(\theta_1 + \theta_2\alpha_1^2 - \frac{\theta_2^2\alpha_1^2\beta_1^2}{\phi_1 + \theta_2\beta_1^2}\right)\mu_1^2 - \theta_2\omega_1^2 - (\theta_1 + \theta_2\alpha_1^2)\sigma_1^2 + \frac{\theta_2^2\alpha_1^2\beta_1^2 E_{g_1^*}^2(\Pi_1)\sigma_1^4}{(\phi_1 + \theta_2\beta_1^2)\left(E_{g_1^*}(\Pi_1^2)\left(\mu_1^2 + \frac{\lambda}{\lambda-1}\sigma_1^2\right) - E_{g_1^*}^2(\Pi_1)\mu_1^2\right)}.$$

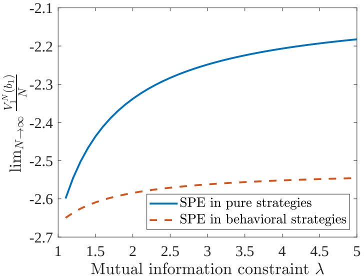
Numerical results

Time-invariant LQG model parameters

Parameter	μ_1	σ_1^2	α	β	ω^2	θ	ϕ
Value	0	1	-0.5	-1.5	1	2	1



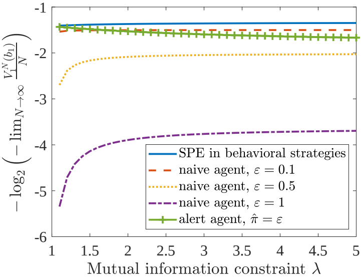
$\tilde{\theta}_n$, $\hat{\theta}_n$, $\check{\theta}_n$ computed as $L^n(0,0)$ and $J^n(0,0)$ v.s. the number of iterations n, for $\lambda = 1.5$ and $\lambda = 2$ ($\lambda > \alpha^2$).



Expected average reward per stage v.s. mutual information constraint λ, for stationary SPEs in pure strategies and behavioral strategies.

Numerical results

- Naive agent: Unaware of the adversary and take the optimal LQG strategy
- Alert agent: Assume an adversarial strategy and take the best response



Expected average reward per stage for stationary SPE in behavioral strategies, that for a naive agent, and that for an alert agent v.s. mutual information constraint λ.

Summary

- Adversarial LQG game
  - Strategic interaction
  - Asymmetric information
  - System dynamics
- Subgame perfect equilibria
  - Pure strategy SPE
  - Behavioral strategy SPE
- Improvement by considering strategic interactions
- Future work
  - Non-scalar state dynamic system
  - Relax the assumption that the adversarial strategy is revealed to the agent after each stage

Summary

- Adversarial LQG game
  - Strategic interaction
  - Asymmetric information
  - System dynamics
- Subgame perfect equilibria
  - Pure strategy SPE
  - Behavioral strategy SPE
- Improvement by considering strategic interactions
- Future work
  - Non-scalar state dynamic system
  - Relax the assumption that the adversarial strategy is revealed to the agent after each stage

Thank you for your attention!