ADVERSARIAL LINEAR QUADRATIC REGULATOR UNDER FALSIFIED ACTIONS

Chenglong Sun, Zuxing Li, and Chao Wang

School of Electronics and Information Engineering
Tongji University, Shanghai 201804, China

ABSTRACT

Reinforcement learning (RL) has been widely employed in communications, in the areas of interference management, resource allocation, signal detection, and power control, etc. Nevertheless, RL is vulnerable under various malicious attacks, such as adversarial examples and privacy intrusions. In this paper, a falsification attack on the agent actions in a scalar linear quadratic regulator (LQR) system is studied. This adversarial problem is formulated as a novel dynamic game by introducing an adversarial belief, and subgame perfect equilibria (SPEs) are characterized under different adversarial constraints. Numerical experiments show the impact of strategic interactions and justify the theoretic results.

***Index Terms*—** Adversarial analysis, dynamic game, Fisher information, subgame perfect equilibrium

1. INTRODUCTION

Recent breakthrough of deep reinforcement learning (DRL) provides an efficient solution for highly-complicated decision strategy optimization [1] and has a wide range of intelligent applications. For instance, DRL has been employed in communication systems for interference management, resource allocation, and power control [2, 3, 4, 5].

Nevertheless, DRL is vulnerable under various attacks. As shown in [6], DRL can be misled by adversarial examples. The design of adversarial examples has been studied in [6, 7, 8, 9, 10]. As a counter measure, DRL was trained by using the adversarial examples to enhance the robustness [11]. In these works, the control agent (or adversary) was optimized by fixing the adversary (or control agent), i.e., the strategic interactions between the players were not fully considered.

Stochastic game (SG) [12] and partially observable SG (POSG) model the strategic interactions between multiple players in a dynamic system, and have been employed in adversarial problems [10, 13]. However, players in SG and POSG do not interact with each other directly but through the impact of their actions on the system state transitions. Therefore, SG and POSG cannot be applied to model direct interactions in dynamic systems, e.g., adversarial examples in DRL. Cheap talk game [14] models a direct interaction between players, where a sender sends a message to a receiver based on private information; and the receiver decides an action based on the received message and a belief on the private information. In [15, 16, 17, 18], dynamic cheap talk games were formulated to study the adversarial examples in dynamic systems.

Besides adversarial examples, agent actions in a dynamic system can also be maliciously manipulated by the adversary to degrade the agent performance. In [19], this adversarial problem was studied in the context of linear quadratic Gaussian (LQG), although the strategic interactions were not fully considered. In [20], a Stackelberg game was formulated for a similar adversarial problem, where the dynamic system plays as the leader and the adversarial agent plays as the follower.

In this paper, we consider an LQR control under falsification attacks on the agent actions. We propose a novel zero-sum dynamic game to model the strategic interactions in the adversarial problem, where the agent and the adversary make decisions of their actions simultaneously in the beginning of each stage. We impose an information-theoretic constraint on the adversarial falsification. We study the conditions for the existence of pure strategy SPEs, the strategies and the expected accumulated agent reward in SPEs.

Notation: Unless otherwise specified, we denote a random variable by a capital letter, e.g., X, a realization by the corresponding lower-case letter, e.g., x, a Gaussian distribution with mean μ and variance σ^2 by $\mathcal{N}(\mu,\sigma^2)$, and the expectation operation by $\mathbb{E}(\cdot)$.

2. ADVERSARIAL LQR

We consider an N-stage adversarial LQR problem as shown in Fig. 1. In the i-th stage, the agent observes a system state $s_i \in \mathbb{R}$ and then determines an action $a_i \in \mathbb{R}$ with an objective to maximize the expected accumulated reward. We consider an attack on the agent action, i.e., an adversary manipulates the action a_i and feeds a falsified action $\hat{a}_i \in \mathbb{R}$ back to the dynamic system with an objective to minimize the expected accumulated reward. The instantaneous reward $r_i \in \mathbb{R}$ is determined by the state s_i and the falsified action $\hat{a}_i$. The current

This work was supported in part by the National Natural Science Foundation of China (62006173, 62171322), the National Key Research and Development Program of China (2018YFE0125400), the 2021-2023 China-Serbia Inter-Governmental S&T Cooperation Project (No. 6), and the Sino-German Center of Intelligent Systems at Tongji University.

Corresponding author: Zuxing Li (zuxing@tongji.edu.cn)

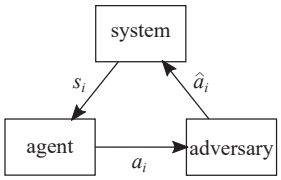

Fig. 1. The considered adversarial LQR problem.

state s_i, the next state s_{i+1}, the action a_i, the falsified action $\hat{a}_i$, and the reward r_i are described by

$$s_{i+1} = \alpha_i s_i + \beta_i \hat{a}_i + z_i, \text{ given } \alpha_i \neq 0,\ \beta_i \neq 0, \tag{1}$$

$$a_i = \kappa_i s_i, \tag{2}$$

$$\hat{a}_i = \pi_i a_i + c_i, \tag{3}$$

$$r_i = R_i(s_i, \hat{a}_i) = -\theta_i s_i^2 - \phi_i \hat{a}_i^2, \text{ given } \theta_i > 0,\ \phi_i > 0, \tag{4}$$

$$Z_i \sim \mathcal{N}(0, \omega_i^2), \text{ given } \omega_i^2 > 0, \tag{5}$$

$$C_i \sim \mathcal{N}(0, \delta_i^2). \tag{6}$$

As proved in the follows, it is optimal to consider the linear mapping (2) for the agent, which is consistent with standard LQR control [21]. The adversarial falsification is assumed to be a linear Gaussian random mapping, which is determined by the linear falsification coefficient π_i and the variance of the Gaussian random falsification δ_i^2. If $\pi_i = 1$ and $\delta_i^2 = 0$, then the problem reduces to a standard LQR control. We further assume that the adversarial falsification is "small", since a significant falsification of action can be easily detected [19] and involves a high cost. Therefore, the following constraints are imposed on the adversarial model (3):

$$-\infty < \varepsilon' \leq \pi_i \leq \varepsilon < \infty, \tag{7}$$

$$0 < \lambda \leq \frac{\pi_i^2}{\delta_i^2}. \tag{8}$$

The ratio π_i^2/δ_i^2 in (8) is the Fisher information [22], which measures the information available about any a_i given $\hat{A}_i$. Therefore, the constraint (8) means that the falsified action can convey at least a certain amount of information about the agent action. We denote by $\mathscr{A}(\varepsilon', \varepsilon, \lambda)$ the set of feasible adversarial actions (π_i, δ_i^2) satisfying (7)-(8).

For the adversarial LQR problem, we propose a novel dynamic game to capture the direct interactions of the agent and the adversary, the asymmetric information about the state available for the agent and the adversary, and the system dynamics.

We denote by b_i the belief of the adversary about S_i in the beginning of the i-th stage, i.e., the posterior distribution of S_i for the adversary after observing $\{a_j\}_{j=1}^{i-1}$ and $\{\hat{a}_j\}_{j=1}^{i-1}$. We assume that the adversarial belief b_i is also known by the agent. Although this assumption is strong, we will show it does not take effect in pure strategy SPEs.

In the i-th stage, the proposed dynamic game is played as follows. The agent uses a pure strategy $f_i(b_i)$ for choosing the parameter κ_i based on the belief b_i, and then decides an action $a_i = \kappa_i s_i$ when the agent observes a system state s_i. The adversary employs a pure strategy $g_i(b_i)$ for choosing the parameters $(\pi_i, \delta_i^2) \in \mathscr{A}(\varepsilon', \varepsilon, \lambda)$ based on the belief b_i, and then generates a random falsified action $\hat{A}_i \sim \mathcal{N}(\pi_i a_i, \delta_i^2)$ on observing the action a_i. In the end of this stage, both the agent and the adversary update the next belief b_{i+1}.

Given the agent strategies $f^N = (f_1, \ldots, f_N)$ and the adversarial strategies $g^N = (g_1, \ldots, g_N)$ over N stages, the expected accumulated reward is

$$V\left(b_1, f^N, g^N\right) = \mathbb{E}_{b_1, f^N, g^N}\left(\sum_{j=1}^{N} R_j(S_j, \hat{A}_j)\right). \tag{9}$$

We refer to the proposed dynamic game as adversarial LQR game, where the objective of the adversary is to minimize (9), while the agent aims at maximizing it.

3. SPE ANALYSIS

In this section, we first formulate SPE for the adversarial LQR game, and then characterize pure strategy SPEs.

3.1. SPE Formulation

Given an SPE with strategies (f^{N*}, g^{N*}), the value function of a subgame starting from the i-th stage with a belief b_i is

$$V_i^N(b_i) = V(b_i, f_i^{N*}, g_i^{N*}) = \mathbb{E}_{b_i, f_i^{N*}, g_i^{N*}}\left(\sum_{j=i}^{N} R_j(S_j, \hat{A}_j)\right). \tag{10}$$

If (non-equilibrium) strategies in the i-th stage are used, the Q-function of the subgame is

$$Q_i^N(b_i, f_i, g_i) = \mathbb{E}_{b_i, f_i, g_i}\left(R_i(S_i, \hat{A}_i) + V_{i+1}^N(b_{i+1})\right). \tag{11}$$

From the SPE definitions, the value functions $\{V_i^N\}_{i=1}^{N-1}$ have to satisfy the backward dynamic programming equations:

$$V_i^N(b_i) = \max_{f_i} Q_i^N(b_i, f_i, g_i^*) = \min_{g_i} Q_i^N(b_i, f_i^*, g_i). \tag{12}$$

3.2. Pure Strategy SPE

We first consider the single-stage game, i.e., $N = 1$.

Proposition 1. *Let $N = 1$. When $\mathscr{A}(\varepsilon', \varepsilon, \lambda) \neq \emptyset$, an SPE exists. For any belief b_1, the SPE strategies are*

$$\kappa_1^* = f_1^*(b_1) = 0, \tag{13}$$

$$(\pi_1^*, \delta_1^{2*}) = g_1^*(b_1) = \begin{cases} \left(\varepsilon, \dfrac{\varepsilon^2}{\lambda}\right), \text{ if } |\varepsilon| \geq |\varepsilon'| \\ \left(\varepsilon', \dfrac{\varepsilon'^2}{\lambda}\right), \text{ otherwise} \end{cases}, \tag{14}$$

i.e., the agent always chooses the action $a_1^* = 0$ and the adversary chooses the parameters leading to the largest-variance noise.

The proof of Proposition 1 is presented in the full version [23, Appendix A].

Depending on the constraint (7), the existence of pure strategy SPE unfortunately does not always hold for the multiple-stage dynamic game, i.e., $N \geq 2$.

Define parameters $\{\tilde{\theta}_i\}_{i=1}^{N+1}$ and $\{\tilde{\phi}_i\}_{i=1}^{N+1}$ by the backward iterations as

$$\tilde{\theta}_{N+1} = \tilde{\phi}_{N+1} = 0, \tag{15}$$

$$\tilde{\theta}_i = \theta_i + \frac{\tilde{\theta}_{i+1}\alpha_i^2\phi_i}{\phi_i + \tilde{\theta}_{i+1}\beta_i^2}, \tag{16}$$

$$\tilde{\phi}_i = \phi_i + \tilde{\theta}_{i+1}\beta_i^2. \tag{17}$$

We first consider the case $|\varepsilon'| = |\varepsilon| \neq 0$.

Theorem 1. *Let $N \geq 2$. When $\varepsilon' = \varepsilon \neq 0$, there is a unique pure strategy SPE. For $1 \leq i \leq N$ and any belief b_i, the SPE strategies are given by*

$$\kappa_i^* = f_i^*(b_i) = -\frac{\tilde{\theta}_{i+1}\alpha_i\beta_i}{(\phi_i + \tilde{\theta}_{i+1}\beta_i^2)\varepsilon} = -\frac{\tilde{\theta}_{i+1}\alpha_i\beta_i}{(\phi_i + \tilde{\theta}_{i+1}\beta_i^2)\varepsilon'}, \tag{18}$$

$$(\pi_i^*, \delta_i^{2*}) = g_i^*(b_i) = \left(\varepsilon, \frac{\varepsilon^2}{\lambda}\right) = \left(\varepsilon', \frac{\varepsilon'^2}{\lambda}\right). \tag{19}$$

Corollary 1. *When $\varepsilon' = \varepsilon \neq 0$, the value function induced by the unique pure strategy SPE is*

$$V_i^N(b_i) = -\tilde{\theta}_i\mathbb{E}(S_i^2) - \sum_{j=i}^{N}\tilde{\phi}_j\frac{\varepsilon^2}{\lambda} - \sum_{j=i+1}^{N}\tilde{\theta}_j\omega_{j-1}^2, \tag{20}$$

$$= -\tilde{\theta}_i\mathbb{E}(S_i^2) - \sum_{j=i}^{N}\tilde{\phi}_j\frac{\varepsilon'^2}{\lambda} - \sum_{j=i+1}^{N}\tilde{\theta}_j\omega_{j-1}^2. \tag{21}$$

Proposition 2. *Let $N \geq 2$. When $-\varepsilon' = \varepsilon > 0$, there is no pure strategy SPE for the adversarial LQR game.*

When $|\varepsilon| \neq |\varepsilon'|$, additional conditions on the beliefs are needed to guarantee the existence of pure strategy SPEs.

Theorem 2. *Let $N \geq 2$. When $|\varepsilon'| < |\varepsilon|$, a pure strategy SPE exists iff $\frac{(\varepsilon^2-\varepsilon'^2)\varepsilon^2}{(\varepsilon-\varepsilon')^2} \geq \frac{\lambda\tilde{\theta}_{i+1}^2\alpha_i^2\beta_i^2\mathbb{E}(S_i^2)}{(\phi_i+\tilde{\theta}_{i+1}\beta_i^2)^2}$ for all $1 \leq i \leq N$. For $1 \leq i \leq N$ and any b_i satisfying $\frac{(\varepsilon^2-\varepsilon'^2)\varepsilon^2}{(\varepsilon-\varepsilon')^2} \geq \frac{\lambda\tilde{\theta}_{i+1}^2\alpha_i^2\beta_i^2\mathbb{E}(S_i^2)}{(\phi_i+\tilde{\theta}_{i+1}\beta_i^2)^2}$, the SPE strategies are given by*

$$\kappa_i^* = f_i^*(b_i) = -\frac{\tilde{\theta}_{i+1}\alpha_i\beta_i}{(\phi_i + \tilde{\theta}_{i+1}\beta_i^2)\varepsilon}, \tag{22}$$

$$(\pi_i^*, \delta_i^{2*}) = g_i^*(b_i) = \left(\varepsilon, \frac{\varepsilon^2}{\lambda}\right). \tag{23}$$

Corollary 2. *When $|\varepsilon'| < |\varepsilon|$ and $\frac{(\varepsilon^2-\varepsilon'^2)\varepsilon^2}{(\varepsilon-\varepsilon')^2} \geq \frac{\lambda\tilde{\theta}_{i+1}^2\alpha_i^2\beta_i^2\mathbb{E}(S_i^2)}{(\phi_i+\tilde{\theta}_{i+1}\beta_i^2)^2}$ for all $1 \leq i \leq N$, the value function induced by the unique pure strategy SPE can be obtained as (20).*

Theorem 3. *Let $N \geq 2$. When $|\varepsilon'| > |\varepsilon|$, a pure strategy SPE exists iff $\frac{(\varepsilon'^2-\varepsilon^2)\varepsilon'^2}{(\varepsilon-\varepsilon')^2} \geq \frac{\lambda\tilde{\theta}_{i+1}^2\alpha_i^2\beta_i^2\mathbb{E}(S_i^2)}{(\phi_i+\tilde{\theta}_{i+1}\beta_i^2)^2}$ for all $1 \leq i \leq N$. For $1 \leq i \leq N$ and any b_i satisfying $\frac{(\varepsilon'^2-\varepsilon^2)\varepsilon'^2}{(\varepsilon-\varepsilon')^2} \geq \frac{\lambda\tilde{\theta}_{i+1}^2\alpha_i^2\beta_i^2\mathbb{E}(S_i^2)}{(\phi_i+\tilde{\theta}_{i+1}\beta_i^2)^2}$, the SPE strategies are given by*

$$\kappa_i^* = f_i^*(b_i) = -\frac{\tilde{\theta}_{i+1}\alpha_i\beta_i}{(\phi_i + \tilde{\theta}_{i+1}\beta_i^2)\varepsilon'}, \tag{24}$$

$$(\pi_i^*, \delta_i^{2*}) = g_i^*(b_i) = \left(\varepsilon', \frac{\varepsilon'^2}{\lambda}\right). \tag{25}$$

Corollary 3. *When $|\varepsilon'| > |\varepsilon|$ and $\frac{(\varepsilon'^2-\varepsilon^2)\varepsilon'^2}{(\varepsilon-\varepsilon')^2} \geq \frac{\lambda\tilde{\theta}_{i+1}^2\alpha_i^2\beta_i^2\mathbb{E}(S_i^2)}{(\phi_i+\tilde{\theta}_{i+1}\beta_i^2)^2}$ for all $1 \leq i \leq N$, the value function induced by the unique pure strategy SPE can be obtained as* (21)*.*

The proofs of Theorems 1, 2, 3, Corollaries 1, 2, 3, and Proposition 2 are given in the full version [23, Appendix B].

Remark 1. *The SPE strategies of both players are independent of the adversarial belief. Therefore, the assumption that both players always have the same belief does not take effect in SPEs.*

3.3. Asymptotic Analysis on Time-Invariant System

We now focus on the SPEs in the asymptotic regime as $N \to \infty$ for a time-invariant system, where $\alpha_i = \alpha \neq 0$, $\beta_i = \beta \neq 0$, $\omega_i^2 = \omega^2 > 0$, $\theta_i = \theta > 0$, and $\phi_i = \phi > 0$ for all $i \geq 1$.

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

$$L(x,y) = \left(\theta + \frac{\alpha^2\phi x}{\phi + \beta^2 x}, \phi + \beta^2 x\right), \tag{26}$$

which is effectively the parameter update rule in (16)-(17) for the time-invariant system. It follows from [18, Proposition 2] that the mapping L admits a least fixed point $(\tilde{\theta}, \tilde{\phi}) \in \mathbb{R}_{\geq 0}^2$ and satisfies

$$\lim_{n\to\infty} L^{(n)}(0,0) = L(\tilde{\theta}, \tilde{\phi}) = (\tilde{\theta}, \tilde{\phi}), \tag{27}$$

where $L^{(n)}(0,0) = \underbrace{L(L(\cdots(L}_{n \text{ iterations}}(0,0)))\cdots))$. The least fixed point can be used to characterize the pure strategy SPEs for the time-invariant system in the asymptotic regime.

Corollary 4. *Let $N \to \infty$. When $\varepsilon' = \varepsilon \neq 0$, the unique SPE for the time-invariant system consists of stationary pure strategies. For $i \geq 1$ and any belief b_i, the SPE strategies are:*

$$\kappa_i^* = f_i^*(b_i) = -\frac{\tilde{\theta}\alpha\beta}{(\phi + \tilde{\theta}\beta^2)\varepsilon} = -\frac{\tilde{\theta}\alpha\beta}{(\phi + \tilde{\theta}\beta^2)\varepsilon'}, \tag{28}$$

$$(\pi_i^*, \delta_i^{2*}) = g_i^*(b_i) = \left(\varepsilon, \frac{\varepsilon^2}{\lambda}\right) = \left(\varepsilon', \frac{\varepsilon'^2}{\lambda}\right). \tag{29}$$

If the initial belief b_1 has bounded mean and variance, the asymptotic expected reward rate is

$$\lim_{N\to\infty}\frac{V_1^N(b_1)}{N} = -\tilde{\phi}\frac{\varepsilon^2}{\lambda} - \tilde{\theta}\omega^2, \tag{30}$$

$$= -\tilde{\phi}\frac{\varepsilon'^2}{\lambda} - \tilde{\theta}\omega^2. \tag{31}$$

The proof of Corollary 4 is by substituting $(\tilde{\theta}_i, \tilde{\phi}_i)$ for all i with $(\tilde{\theta}, \tilde{\phi})$ in Theorem 1 and Corollary 1.

When $|\varepsilon'| \neq |\varepsilon|$, it follows from Theorems 2 and 3 that beliefs need to satisfy an upper bound for the existence of pure strategy SPEs. For the time-invariant system, we give in the following the sufficient and necessary conditions for the existence of pure strategy SPEs in the asymptotic regime.

Corollary 5. *Let $N \to \infty$. When $|\varepsilon'| < |\varepsilon|$, a pure strategy SPE for the time-invariant system exists iff*

$$0 < \left|\frac{\alpha\phi}{\phi + \tilde{\theta}\beta^2}\right| < 1, \tag{32}$$

$$\frac{(\varepsilon^2 - \varepsilon'^2)\varepsilon^2}{(\varepsilon - \varepsilon')^2} \geq \frac{\lambda\tilde{\theta}^2\alpha^2\beta^2\mathbb{E}(S_1^2)}{(\phi + \tilde{\theta}\beta^2)^2}, \tag{33}$$

$$\frac{(\varepsilon^2 - \varepsilon'^2)\varepsilon^2}{(\varepsilon - \varepsilon')^2} \geq \frac{\tilde{\theta}^2\alpha^2\beta^2(\beta^2\varepsilon^2 + \lambda\omega^2)}{(\phi + \tilde{\theta}\beta^2)^2 - \alpha^2\phi^2}. \tag{34}$$

For $i \geq 1$, the SPE strategies are stationary and are given by

$$\kappa_i^* = f_i^*(b_i) = -\frac{\tilde{\theta}\alpha\beta}{(\phi + \tilde{\theta}\beta^2)\varepsilon}, \tag{35}$$

$$(\pi_i^*, \delta_i^{2*}) = g_i^*(b_i) = \left(\varepsilon, \frac{\varepsilon^2}{\lambda}\right). \tag{36}$$

In the SPE, the asymptotic expected reward rate can be obtained as (30).

Corollary 6. *Let $N \to \infty$. When $|\varepsilon'| > |\varepsilon|$, a pure strategy SPE for the time-invariant system exists iff*

$$0 < \left|\frac{\alpha\phi}{\phi + \tilde{\theta}\beta^2}\right| < 1, \tag{37}$$

$$\frac{(\varepsilon'^2 - \varepsilon^2)\varepsilon'^2}{(\varepsilon - \varepsilon')^2} \geq \frac{\lambda\tilde{\theta}^2\alpha^2\beta^2\mathbb{E}(S_1^2)}{(\phi + \tilde{\theta}\beta^2)^2}, \tag{38}$$

$$\frac{(\varepsilon'^2 - \varepsilon^2)\varepsilon'^2}{(\varepsilon - \varepsilon')^2} \geq \frac{\tilde{\theta}^2\alpha^2\beta^2(\beta^2\varepsilon'^2 + \lambda\omega^2)}{(\phi + \tilde{\theta}\beta^2)^2 - \alpha^2\phi^2}. \tag{39}$$

For $i \geq 1$, the SPE strategies are stationary and are given by

$$\kappa_i^* = f_i^*(b_i) = -\frac{\tilde{\theta}\alpha\beta}{(\phi + \tilde{\theta}\beta^2)\varepsilon'}, \tag{40}$$

$$(\pi_i^*, \delta_i^{2*}) = g_i^*(b_i) = \left(\varepsilon', \frac{\varepsilon'^2}{\lambda}\right). \tag{41}$$

In the SPE, the asymptotic expected reward rate can be obtained as (31).

The proofs of Corollaries 5 and 6 are presented in the full version [23, Appendix C].

Table 1. Time-invariant system parameters.

Parameter	$\mathbb{E}(S_1^2)$	α	β	ω^2	θ	ϕ
Value	2	-1	-1.5	1	1	2

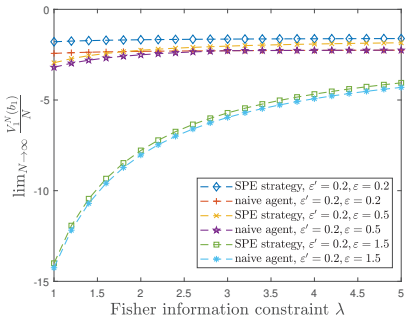

Fig. 2. Asymptotic expected reward rates achieved by SPE and a naive agent.

4. NUMERICAL EXPERIMENT

We show the impact of strategic interactions on a time-invariant LQR system. The system parameters are listed in Table 1.

The asymptotic expected reward rates achieved in SPEs for the considered time-invariant system are shown in Fig. 2. As the Fisher information lower bound λ increases, the adversary becomes weaker and the asymptotic expected reward rate increases. We fix the lower bound $\varepsilon' = 0.2$. As the upper bound ε increases, the adversary has a larger feasible set for the coefficient π_i and becomes stronger, and thus the asymptotic expected reward rate decreases.

We also consider a non-equilibrium scenario, where a naive agent uses the standard LQR strategy while an adversary uses the best response strategy subject to the constraints. As shown in Fig. 2, the asymptotic expected reward rate in SPE is always greater than that achieved by the naive agent, i.e., the agent becomes more resilient to the attacks by fully considering the strategic interactions.

5. CONCLUSION

We proposed a novel dynamic game to capture the direct interactions, the asymmetric system information, and the inherent system dynamics for an adversarial LQR problem. We focused on the pure strategy SPEs. Our study showed that: The existence of SPE depends on the adversarial constraints and the initial adversarial belief; the SPE strategies for both players are independent of the adversarial belief. Our numerical experiment justified the theoretic results and showed that an agent that is aware of the falsification attacks can be designed more resilient.

6. REFERENCES

[1] V. Mnih et al., "Human-level control through deep reinforcement learning," *Nature*, vol. 518, pp. 529–533, 2015.

[2] N. C. Luong et al., "Applications of deep reinforcement learning in communications and networking: A survey," *IEEE Communications Surveys and Tutorials*, vol. 21, pp. 3133–3174, 2019.

[3] H. Zhang, M. Min, L. Xiao, S. Liu, P. Cheng, and M. Peng, "Reinforcement learning-based interference control for ultra-dense small cells," in *Proc. of IEEE GLOBECOM*, 2018, pp. 1–6.

[4] Y. Wei, F. R. Yu, M. Song, and Z. Han, "Joint optimization of caching, computing, and radio resources for fog-enabled IoT using natural actor–critic deep reinforcement learning," *IEEE Internet of Things Journal*, vol. 6, pp. 2061–2073, 2019.

[5] Y. Zhang, D. Lan, C. Wang, P. Wang, and F. Liu, "Deep reinforcement learning-aided transmission design for multi-user V2V networks," in *Proc. of IEEE WCNC*, 2021, pp. 1–6.

[6] S. Huang, N. Papernot, I. Goodfellow, Y. Duan, and P. Abbeel, "Adversarial attacks on neural network policies," 2016.

[7] Y. C. Lin, Z. W. Hong, Y. H. Liao, M. L. Shih, M. Y. Liu, and S. Min, "Tactics of adversarial attack on deep reinforcement learning agents," in *Proc. of IJCAI*, 2017, pp. 3756–3762.

[8] V. Behzadan and A. Munir, "Vulnerability of deep reinforcement learning to policy induction attacks," in *Proc. of MLDM*, 2017, pp. 262–275.

[9] A. Russo and A. Proutiere, "Optimal attacks on reinforcement learning policies," 2019.

[10] A. Gleave, M. Dennis, C. Wild, N. Kant, S. Levine, and S. Russell, "Adversarial policies: Attacking deep reinforcement learning," in *Proc. of ICLR*, 2020.

[11] L. Pinto, J. Davidson, R. Sukthankar, and A. Gupta, "Robust adversarial reinforcement learning," in *Proc. of ICML*, 2017, pp. 2817–2826.

[12] L. Shapley, "Stochastic games," *Proc. of the National Academy of Sciences*, vol. 39, pp. 1095–1100, 1953.

[13] K. Horak, Q. Zhu, and B. Bosansky, "Manipulating adversary's belief: A dynamic game approach to deception by design for proactive network security," in *Proc. of GameSec*, 2017, pp. 273–294.

[14] V. P. Crawford and J. Sobel, "Strategic information transmission," *Econometrica*, vol. 50, pp. 1431–1451, 1982.

[15] S. Saritas, S. Yuksel, and S. Gezici, "Nash and Stackelberg equilibria for dynamic cheap talk and signaling games," in *Proc. of ACC*, 2017, pp. 3644–3649.

[16] S. Saritas, E. Shereen, H. Sandberg, and G. Dán, "Adversarial attacks on continuous authentication security: A dynamic game approach," in *Proc. of GameSec*, 2019, pp. 439–458.

[17] Z. Li and G. Dán, "Dynamic cheap talk for robust adversarial learning," in *Proc. of GameSec*, 2019, pp. 297–309.

[18] Z. Li, G. Dán, and D. Liu, "A game theoretic analysis of LQG control under adversarial attack," in *Proc. of CDC*, 2020, pp. 1632–1639.

[19] R. Zhang and P. Venkitasubramaniam, "Stealthy control signal attacks in linear quadratic Gaussian control systems: Detectability reward tradeoff," *IEEE Transactions on Information Forensics and Security*, vol. 12, pp. 1555–1570, 2017.

[20] M. O. Sayin and T. Basar, "Secure sensor design for cyber-physical systems against advanced persistent threats," in *Proc. of GameSec*, 2017, pp. 91–111.

[21] T. Soderstrom, *Discrete-Time Stochastic Systems*, Springer, 2002.

[22] R. A. Fisher, "On the mathematical foundations of theoretical statistics," *Philosophical Transactions of the Royal Society of London. Series A*, vol. 222, pp. 309–368, 1922.

[23] C. Sun, Z. Li, and C. Wang, "Adversarial linear quadratic regulator under falsified actions," 2021, `https://www.kth.se/files/view/zuxing/615ec98f842ed30013362694/icassp2022fullpaper.pdf`.