Safety-Critical Online Control with Adversarial Disturbances

Bhaskar Ramasubramanian[1], Baicen Xiao[1], Linda Bushnell[1], and Radha Poovendran[1]

Abstract— **This paper studies the control of safety-critical dynamical systems in the presence of adversarial disturbances. We seek to synthesize state-feedback controllers to minimize a cost incurred due to the disturbance while respecting the safety constraint. The safety constraint is given by a bound on an $\mathcal{H}_\infty$ norm, while the cost is specified as an upper bound on an $\mathcal{H}_2$ norm of the system. We consider an online setting where costs at each time are revealed only after the controller at that time is chosen. We propose an iterative approach to the synthesis of the controller by solving a modified discrete-time Riccati equation. Solutions of this equation enforce the safety constraint. We compare the cost of this controller with that of the optimal controller when one has complete knowledge of disturbances and costs in hindsight. We show that the *regret* function, which is defined as the difference between these costs, called varies logarithmically with the time horizon. We validate our approach on a process control setup that is subject to two kinds of adversarial attacks.**

I. Introduction

The recent advances and successes of reinforcement learning (RL) [1] in robotics, games, and mobile networks [2]–[6] has spurred its use in other areas where RL algorithms interact with the physical environment over long periods of time [7]–[9]. An increasingly popular domain where RL methods are being deployed are to safety-critical systems like large-scale power systems [10], which are susceptible to attacks by an intelligent adversary [11], [12]. Since these systems have an underlying dynamic model, actions of the system are typically a function of (a history of) the system states. Rules governing these actions can be designed so that the overall system behaves in a desired way. At the same time, the actions may have to be chosen to minimize a cost.

We consider a safety-critical linear time invariant (LTI) system affected by adversarial inputs. Our goal is to design state-feedback controllers to minimize the cost incurred due to this input while satisfying a safety constraint. This is also called the '*combined $\mathcal{H}_2/\mathcal{H}_\infty$ problem*' [13]–[15]. The $\mathcal{H}_\infty$ safety constraint enforces a bound on the ratio of the magnitude of the output to that of the adversarial disturbance. The $\mathcal{H}_2$ cost is the expected mean square output value when the disturbance input is a white noise process. The $\mathcal{H}_\infty$ constraint is embedded in the optimization process by solving a modified discrete-time Riccati equation, whose solution yields an upper bound on the $\mathcal{H}_2$ cost.

In this paper, we study an online scenario of the combined $\mathcal{H}_2/\mathcal{H}_\infty$ problem. At each time, the adversary inserts a disturbance, after which the cost incurred is revealed to the system. Our aim is to iteratively design controllers to minimize this cost, while satisfying the safety constraint. We compare the cost of this controller with that of the optimal controller if all adversarial inputs and costs were known apriori. The difference between these costs is termed as the *regret* faced by the system.

Regret bounds for partially and fully adversarial disturbances for LTI dynamics and convex costs were presented in [16]–[20], where the authors considered a richer class of 'disturbance action policies'. These policies depend not only on the current state, but also on a history of disturbances, and provide stronger regret bounds (poly-logarithmic) than we present (logarithmic), since policies in this paper only depend on the current state. Also, while the analyses in [16]–[20] fix a stabilizing controller at the start of their algorithms, we adopt a different approach and iteratively solve a set of Riccati equations to update the controller at each time step.

A. Contributions

We aim to minimize an upper bound on the $\mathcal{H}_2$ cost for an LTI system with adversarial inputs with an $\mathcal{H}_\infty$ constraint on the disturbance-output map in an online setting. At each time, the adversary inserts a disturbance input. The cost functions are revealed to the system only after it has determined a controller. We make the following contributions:

- We introduce strongly stable disturbance attenuating (S2DA) policies. This generalizes strongly stable policies from [21]. S2DA policies are strongly stable and satisfy an additional condition on the $\mathcal{H}_\infty$ norm.
- We show that initializing our procedure with a stabilizing disturbance attenuating policy will yield stabilizing disturbance attenuating policies at successive time steps. If solutions of the Riccati recursion are bounded, we show that these policies will also be strongly stable.
- We establish bounds on the difference between solutions to the Riccati equation at successive time steps. We use the above results to show that the regret bound is $O(\log T)$, where T is the time horizon of interest.
- We validate our method on a model of the Tennessee Eastman control challenge [22] that is subject to arbitrary adversarial inputs, and a denial-of-service attack.

B. Outline of Paper

The rest of this paper is organized as follows: Section II summarizes related work. We state our problem and detail our solution in Sections III and IV. Section V illustrates our approach on a model of the Tennessee Eastman control challenge. Section VI concludes the paper.

[1] Network Security Lab, Department of Electrical and Computer Engineering, University of Washington, Seattle, WA 98195, USA. {bhaskarr, bcxiao, lb2, rp3}@uw.edu

This work was supported by the U.S. Army Research Office and the Office of Naval Research via Grants W911NF-16-1-0485 and N00014-17-S-B001 respectively.

II. Related Work

Simultaneous $\mathcal{H}_2/\mathcal{H}_\infty$ policy synthesis for discrete-time linear systems is a well-studied problem. The structure of an upper bound for the LQG cost, minimizing which would solve the mixed problem was first proposed in [13]. The authors of this paper also presented a closed-form controller that would minimize this upper bound. Two other upper bounds for the LQG cost were proposed in [14], and it was shown that the same controller would be optimal in each case when restricted to static, full-state feedback. In this case, a static, time-invariant state-feedback is sufficient for optimal performance [15]. In discrete-time, this does not hold for the full-information feedback (states and disturbances available), or partial information cases. This problem was studied for nonlinear discrete-time systems in [23]. An orthogonal approach to solve disturbance attenuation and rejection problems using geometric control theory was presented in [24]–[29]. However, these works considered the $\mathcal{H}_2$ and $\mathcal{H}_\infty$ cases separately. These problems have also been studied in robust and model predictive control [30]–[32].

There has been a renewed interest in the use of RL techniques in learning to control linear dynamical systems. Recent developments in this field are surveyed in [33]. An online version of LQ control with Gaussian disturbances was presented in [21], where the authors presented a regret bound for known LTI dynamics and adversarial quadratic costs. In [16], the authors considered strongly convex costs, and presented stronger regret bounds for a larger class of policies they termed disturbance action policies. This was generalized to semi-adversarial disturbance inputs and convex costs in [17]. The authors of [34] adopted a different approach to determine regret bounds for the LQR. They iteratively solved a sequence of Riccati equations to generate a sequence of stabilizing controllers, and compared the cost of this sequence of controllers with that of the optimal static controller if all costs were known in hindsight.

Minimizing the regret over sequentially revealed adversarial convex costs against the class of linear policies when a model of the dynamics was unknown was studied in [18]. This was generalized to partially observed systems with semi-adversarial disturbances in [19], and for the Kalman filter in [35]. More recently, [20] presented regret bounds for the case of fully adversarial disturbances.

Sample complexity bounds for the LQR for an LTI system with unknown dynamics were given in [36], and for the Kalman filter in [37]. The convergence of policy-gradient methods for the LQR was studied in [38]–[40]. Convergence guarantees for policy gradient methods for the mixed $\mathcal{H}_2/\mathcal{H}_\infty$ problem was reported in [41]. The authors of [42] studied a trade-off between exploration for learning and safety for the LQR under bounded disturbances and constraints on the state and input sets.

III. Preliminaries and Problem Formulation

For a matrix M, we write $M \geq 0$ when M is positive semi-definite. We write $M_1 \geq M_2$ when $M_1 - M_2 \geq 0$. $Tr(M)$ and $\lambda_{max}(M)$ denote the trace and maximum eigenvalue of M. Consider a discrete-time linear system:

$$x_{t+1} = Ax_t + Bu_t + Dw_t \quad (1)$$

$$z_t = Cx_t + Eu_t, \quad (2)$$

where x_t, u_t, w_t, z_t denote the state, control, disturbance, and controlled output. In this case, the optimal stabilizing control will be a static state feedback $u_t = -Kx_t$ [15].

Let T_{zw} denote the input-output map from the disturbance to the measured output. We want the $\mathcal{H}_\infty$ norm of the closed-loop system $||T_{zw}||_\infty := \sup\{\frac{||z||_2}{||w||_2} : 0 < ||w||_2 < \infty\}$ to remain below a desired threshold, γ. When $||T_{zw}||_\infty < \gamma$, the controller is deemed to have *attenuated* the disturbance[1].

The $\mathcal{H}_2$ norm of a linear system is the expected root mean square value of z_t when w_t is a white noise process [43]. In this case, the $\mathcal{H}_2$-cost will be given by $\lim_{t\to\infty} \mathbb{E}[z_t^T z_t]$. Since w_t in this paper can be a more general adversarial input, we will choose to minimize an upper bound on the (squared) $\mathcal{H}_2$ norm of the closed-loop system [14].

Assumption 1. *Assume the following:*

1) (A, B) *is stabilizable. This will ensure the existence of a controller* K *such that* $(A - BK)$ *is stable.*
2) $C^TC = Q \geq 0$, $E^TC = 0$ *and* $E^TE = R \geq 0$. *This will ensure elimination of cross-weighting terms between state and control variables [14].*

Since we are interested in stabilizing controllers that additionally attenuate the adversarial input, we define the *valid* set of controllers as:

$$\mathcal{K} := \{K : |\lambda_{max}(A - BK)| < 1, ||T_{zw}^K||_\infty < \gamma\}, \quad (3)$$

where T_{zw}^K is the input-output map from w to z under the controller K. An upper bound on the infinite-horizon $\mathcal{H}_2$-cost that we seek to minimize in this paper is [13], [14]:

$$J(K) := Tr(PDD^T) \quad \text{where } P \text{ solves} \quad (4)$$

$$(A - BK)^T\tilde{P}(A - BK) + Q + K^TRK - P = 0 \quad (5)$$

$$\tilde{P} := P + PD(\gamma^2 I - D^TPD)^{-1}D^TP \quad (6)$$

A typical objective to achieve mixed $\mathcal{H}_2$/ $\mathcal{H}_\infty$ goals can then be stated as: '*For the system in Equations (1) - (2), determine a sequence of controls* $\{u_t\}_{t>0}$ *so that:*

1) *the cost in Equation (4) is minimized, subject to*
2) $u_t = -Kx_t$, $K \in \mathcal{K}$, *where* $\mathcal{K}$ *is as in Equation (3).* '

If $P \geq 0$ is a solution to Eqn. (5), then $(A - BK)$ is stable if and only if $(A, (Q + K^TRK)^{1/2})$ is detectable [13]. If $P^* \geq 0$ is a solution and $P^* \leq P$ for all other solutions P, then P^* is a *minimal* solution. The controller that minimizes the cost while achieving $||T_{zw}^{K^*}||_\infty < \gamma$ is [13], [41]:

$$K^* = (R + B^T\tilde{P}^*B)^{-1}B^T\tilde{P}^*A \quad (7)$$

[1] We will say that the disturbance has been attenuated if $||T_{zw}||_\infty < \gamma$ is true even when $\gamma > 1$. In the time-invariant case, $||T_{zw}||_\infty$ can be computed as the maximum singular value of the transfer matrix from w to z, restricted to the boundary of the unit-circle .

We focus on an online setting of Equations (1) - (2). At each time t, the adversary chooses w_t. The learner chooses $u_t = -K_t x_t$, and suffers a loss determined as a function of the matrices C_t and E_t. We assume that the sequence of matrices $\{C_t, E_t\}$ is determined before the start of the learning process. However, they are revealed to the learner only after it chooses u_t. Therefore, the learner faces a *regret*, defined as the difference between the cost when using the aforementioned controller and the optimal controller from the set $\mathcal{K}$. We aim to minimize this regret, and ensure that it grows sub-linearly with the time horizon T. Formally,

Problem 1. *At each time t, the learner observes state x_t, and commits to a controller $u_t = K_t x_t$. After this, cost matrices $Q_t := C_t^T C_t, R_t := E_t^T E_t$ such that $C_t^T E_t = 0$ are revealed to the learner. This cost incurred to the learner is upper-bounded by $J_t(K_t) = Tr(P_t D D^T)$, P_t being the solution of Equation (5). With $J_0(\cdot) := 0$, determine a sequence of policies $\{K_t\}$ such that for some large enough time T, a regret term, defined as $R(T) := J_T(K_T) - \min_{K \in \mathcal{K}} J_T(K)$ grows sub-linearly with T.*

IV. Solution Method

We briefly summarize our solution approach. First, we introduce strongly stable disturbance attenuating policies. This is motivated by strongly stable policies introduced in [21] to quantify the stability of a stabilizing policy. Then, we show that if we initialize our procedure with a stable disturbance attenuating policy, successive iterates will continue to yield policies that are stable and disturbance attenuating. When solutions of the Riccati recursion are uniformly bounded, we show that the sequence of stabilizing and disturbance attenuating policies are also strongly stable for an appropriate choice of parameters. In order to establish our regret bounds, we determine upper bounds on the difference between solutions to the Riccati recursion at successive time steps. We put these together to get the final regret bound. The regret bound comprises a *burn-in cost*, a cost that is incurred before we start to obtain meaningful bounds, while a second term gives the bound for a large enough time horizon T.

A. Strong Stability

We leverage the notion of a strongly stable controller first proposed in [21] for the LQR problem. This was subsequently used in [17] for the more general case.

Definition 1 (Strongly Stable Policies [21])**.** *A policy K is stable if $|\lambda_{max}(A - BK)| < 1$. It is $(\kappa, \epsilon)-$strongly stable for $\kappa > 0$, $\epsilon \in (0, 1]$ if $||K|| \leq \kappa$, and there exist matrices L, H such that $A - BK = HLH^{-1}$, with $||L|| \leq 1 - \epsilon$ and $||H|| ||H^{-1}|| \leq \kappa$.*

Sequentially strongly stable controllers were used in [21], [34] to reason about a sequence of strongly stable policies.

Definition 2 (Sequentially Strongly Stable Policies [21])**.** *A sequence of policies $\{K_t\}_{t \geq 1}$ is sequentially $(\kappa, \epsilon)-$strongly stable for $\kappa > 0$, $\epsilon \in (0, 1]$ if there exist sequences of matrices $\{L_t\}_{t \geq 1}, \{H_t\}_{t \geq 1}$ such that for all $t \geq 1$, $A - BK_t = H_t L_t H_t^{-1}$, and:*

1) *$||L_t|| \leq 1 - \epsilon$, $||K_t|| \leq \kappa$,*
2) *$||H_t|| \leq \beta$, $||H_t^{-1}|| \leq 1/\alpha$, where $\kappa = \beta/\alpha$, $\beta > 0$,*
3) *$||H_{t+1}^{-1} H_t|| \leq 1 + \epsilon$.*

In the above, observe that $|\lambda_{max}(A - BK_t)| = |\lambda_{max}(L_t)| \leq ||L_t|| \leq 1 - \epsilon$. Since we are interested in stable policies that will also achieve disturbance attenuation, we introduce the notion of strongly stable and sequentially strongly stable disturbance attenuating policies.

Definition 3 (Strongly Stable Disturbance Attenuating (S2DA) Policies)**.** *A policy K is $(\kappa, \epsilon, \gamma)-$S2DA if it is $(\kappa, \epsilon)-$strongly stable and $||T_{zw}^K||_\infty < \gamma$.*

Definition 4 (Sequentially Strongly Stable Disturbance Attenuating (S3DA) Policies)**.** *A sequence of policies $\{K_t\}_{t \geq 1}$ is sequentially $(\kappa, \epsilon, \gamma)-$S3DA if $\{K_t\}_{t \geq 1}$ is sequentially $(\kappa, \epsilon)-$strongly stable and $||T_{zw}^{K_t}||_\infty < \gamma$ for all $t \geq 1$.*

B. Set $\mathcal{K}$ is Invariant

In this part, we will show that if $K_1 \in \mathcal{K}$, then $K_t \in \mathcal{K}$ for all $t > 1$. That is, if we start with a stabilizing and disturbance attenuating controller, then successive updates of the controller will retain this property. The sequence of controllers is then said to be *regularized* [41]. We adapt the Riccati recursion update procedure in [44] to the setting of disturbance attenuation. We use representations of solutions to Lyapunov and Riccati equations to establish stability and disturbance attenuation for the updated controllers. Further, since matrices C_t and E_t are fixed at time t, we can use results specific to the time-invariant case. Before proving our result (Theorem 1), we state a useful result from robust control [30] that transforms the constraints in Equation (3) to a the solution of a Riccati inequality.

Lemma 1. *[30] For a discrete-time linear time-invariant system, the following conditions are equivalent:*

1) *The controller gain $K \in \mathcal{K}$.*
2) *There exists $P > 0$ such that:* i)*: $I - \gamma^{-2} D^T P D > 0$, and* ii)*: $Q + K^T R K - P + (A - BK)^T (P + PD(\gamma^2 I - D^T P D)^{-1} D^T P)(A - BK) < 0$.*
3) *The Riccati equation (5) admits a unique stabilizing solution $P \geq 0$ such that:* i)*: $I - \gamma^{-2} D^T P D > 0$, and* ii)*: $(I - \gamma^{-2} D^T P D)^{-T} (A - BK)$ is stable.*

In the sequel, for $t \geq 1$, define:

$$\bar{R}_t := \frac{t-1}{t} \bar{R}_{t-1} + \frac{1}{t} R_t \tag{8}$$

$$\bar{Q}_t := \frac{t-1}{t} \bar{Q}_{t-1} + \frac{1}{t} Q_t \tag{9}$$

We perform the update in this manner in order to obtain useful bounds on differences between successive updates as a function of the time index t.

Theorem 1. *Let Assumption 1 hold, $K_1 \in \mathcal{K}$, and there is *a solution $P_1 \geq 0$ to Equation (5). Suppose at time t,*

$$\begin{aligned}(A - BK_t)^T \tilde{P}_t (A - BK_t) + \bar{Q}_t + K_t^T \bar{R}_t K_t = P_t,\\ \tilde{P}_t := P_t + P_t D(\gamma^2 I - D^T P_t D)^{-1} D^T P_t\end{aligned} \tag{10}$$

and K_t is updated as:

$$K_{t+1} = (\bar{R}_t + B^T \tilde{P}_t B)^{-1} B^T \tilde{P}_t A. \tag{11}$$

Then $K_t \in \mathcal{K}$ for all $t > 1$.

Proof. We will begin by showing that $K_t \in \mathcal{K}$ will ensure that the solution to the Equation (10) is bounded. To do this, we use the fact that for a stabilizing K_t, the solution to an associated Lyapunov equation will be bounded. Then, we will show that K_{t+1} will be stabilizing, and finally show that K_{t+1} will also be disturbance attenuating. We use induction.

A. *Base Case*:

Since (A, B) is stabilizable, there exists a stable controller K_1. Since there exists a solution P_1 to Equation (5), K_1 is also disturbance attenuating (from Lemma 2.1 of [13]), which establishes the base case of our induction.

B. *P_t is Bounded*:

Let $K_t \in \mathcal{K}$ for some $t > 1$. Since K_t is stabilizing, there is a unique solution $\bar{P}_t \geq 0$ to the Lyapunov equation (12), with $\bar{P}_t$ given by [45]:

$$(A - BK_t)^T \bar{P}_t (A - BK_t) - \bar{P}_t = -(\bar{Q}_t + K_t^T \bar{R}_t K_t), \tag{12}$$

$$\bar{P}_t = \sum_{i=0}^{\infty} ((A - BK_t)^T)^i (\bar{Q}_t + K_t^T \bar{R}_t K_t)(A - BK_t)^i.$$

Now consider P_t given by Equation (10). Subtracting Equation (12) from Equation (10), we get:

$$\begin{aligned}P_t - \bar{P}_t = (A - BK_t)^T (P_t - \bar{P}_t)(A - BK_t)\\ +(A - BK_t)^T (P_t D(\gamma^2 I - D^T P_t D)^{-1} D^T P_t)((A - BK_t)\end{aligned} \tag{13}$$

Since $K_t \in \mathcal{K}$, from Lemma 1, $(\gamma^2 I - D^T P_t D) > 0$. Therefore, the second term of Equation (13) is positive definite, which means that (13) is a Lyapunov equation for $P_t - \bar{P}_t$. The (unique) solution to this equation is given by:

$$\begin{aligned}P_t - \bar{P}_t = \sum_{i=0}^{\infty} ((A - BK_t)^T)^i ((A - BK_t)^T \times\\ (P_t D(\gamma^2 I - D^T P_t D)^{-1} D^T P_t)((A - BK_t))(A - BK_t)^i.\end{aligned} \tag{14}$$

Since $(A - BK_t)$ is stable, both $\bar{P}_t$ and $P_t - \bar{P}_t$ are bounded. Therefore, $P_t = \bar{P}_t + (P_t - \bar{P}_t)$ is bounded.

C. *K_{t+1} is Stabilizing*:

Expanding $(A - BK_t)^T \tilde{P}_t (A - BK_t) + K_t^T \bar{R}_t K_t + \bar{Q}_t$ and using Equation (11) to write $B^T \tilde{P}_t A = (\bar{R}_t + B^T \tilde{P}_t B) K_{t+1}$:

$$\begin{aligned}
&A^T \tilde{P}_t A - K_t^T B^T \tilde{P}_t A - A^T \tilde{P}_t B K_t\\
&\quad + K_t^T (B^T \tilde{P}_t B + \bar{R}_t) K_t + \bar{Q}_t\\
=&A^T \tilde{P}_t A + (K_{t+1} - K_t)^T (\bar{R}_t + B^T \tilde{P}_t B)(K_{t+1} - K_t)\\
&\quad - K_{t+1}^T B^T \tilde{P}_t A - A^T \tilde{P}_t B K_{t+1}\\
&\quad + K_{t+1}^T (\bar{R}_t + B^T \tilde{P}_t B) K_{t+1} + \bar{Q}_t\\
=&(A - BK_{t+1})^T \tilde{P}_t (A - BK_{t+1}) + K_{t+1}^T \bar{R}_t K_{t+1}\\
&\quad + \bar{Q}_t + (K_{t+1} - K_t)^T (\bar{R}_t + B^T \tilde{P}_t B)(K_{t+1} - K_t)\\
\Rightarrow &P_t = (A - BK_{t+1})^T P_t (A - BK_{t+1}) + M
\end{aligned} \tag{15}$$

In Equation (15), M is a positive definite matrix defined as:

$$\begin{aligned}
M :=& K_{t+1}^T \bar{R}_t K_{t+1} + \bar{Q}_t\\
&+(A - BK_{t+1})^T (P_t D(\gamma^2 I - D^T P_t D)^{-1} D^T P_t)\\
&\qquad \times (A - BK_{t+1})\\
&+ (K_{t+1} - K_t)^T (\bar{R}_t + B^T \tilde{P}_t B)(K_{t+1} - K_t),
\end{aligned}$$

The terms in the first and third lines in the above equation are positive definite by assumption, and $(\gamma^2 I - D^T P_t D) > 0$ from Lemma 1. Since P_t is bounded, and we can write $P_t = \sum_{i=0}^{\infty} ((A - BK_{t+1})^T)^i M (A - BK_{t+1})^i$, $(A - BK_{t+1})$ must be stable so that the sum on the right hand side does not diverge. Therefore, K_{t+1} is stabilizing.

D. *K_{t+1} is Disturbance Attenuating*:

Since $(A - BK_{t+1})$ is stable, there exists $P \geq 0$ that solves $(A - BK_{t+1})^T P (A - BK_{t+1}) - P = -V$, where $V > 0$. Choose V to be:

$$\begin{aligned}
V :=& K_{t+1}^T \bar{R}_t K_{t+1} + \bar{Q}_t + \rho I\\
&+ (A - BK_{t+1})^T (PD(\gamma^2 I - D^T PD)^{-1} D^T P)\\
&\qquad \times (A - BK_{t+1}),
\end{aligned}$$

where $\rho > 0$ is chosen so that V is positive definite, and

$$\begin{aligned}
&\rho I + (A - BK_{t+1})^T (PD(\gamma^2 I - D^T PD)^{-1} D^T P)\\
&\qquad \times (A - BK_{t+1})\\
&\leq (A - BK_{t+1})^T (P_t D(\gamma^2 I - D^T P_t D)^{-1} D^T P_t)\\
&\qquad \times (A - BK_{t+1})\\
&\quad + (K_{t+1} - K_t)^T (\bar{R}_t + B^T \tilde{P}_t B)(K_{t+1} - K_t).
\end{aligned} \tag{16}$$

Rearranging these equations gives us $(A - BK_{t+1})^T \tilde{P} (A - BK_{t+1}) - P + K_{t+1}^T \bar{R}_t K_{t+1} + \bar{Q}_t = -\rho I < 0$, where $\tilde{P}$ is according to Equation (5). This satisfies the second part of the second condition in Lemma 1.

When $K_t \in \mathcal{K}$, $\gamma^2 I - D^T P_t D > 0$ from Lemma 1. We can write $\gamma^2 I - D^T PD = \gamma^2 I - D^T P_t D + D^T (P_t - P) D$. Now, $P_t - P = (A - BK_{t+1})^T (P_t - P)(A - BK_{t+1}) + N$, where N is got by subtracting the term on the left of the inequality in (16) from the term on the right. This is a Lyapunov equation in $P_t - P$. Since K_{t+1} is stabilizing and $N > 0$, there is a positive semi-definite solution, which gives us $P_t - P \geq 0$. Therefore, $0 < \gamma^2 I - D^T P_t D \leq \gamma^2 I - D^T PD$, which satisfies the first part of the second condition in Lemma 1. Then, from Lemma 1, K_{t+1} is also such that $||T_{zw}^{K_{t+1}}||_\infty < \gamma$, and therefore, $K_{t+1} \in \mathcal{K}$, which completes the proof. □

C. S2DA and S3DA Policies

In this part, we present results quantifying the stability and disturbance attenuation of a sequence of valid policies. We begin by showing that there exist values of parameters κ, ϵ such that any stable and disturbance attenuating policy is S2DA. The proofs are omitted due to space constraints.

Proposition 1. *Assume that $K \in \mathcal{K}$. Then, there exist values κ, ϵ such that K is $(\kappa, \epsilon, \gamma)-$S2DA.*

Suppose that a sequence of positive definite matrices P_t is generated according to Equation (10), where K_{t+1} is given by Equation (11), and K_1 is an initial stable and disturbance attenuating policy. Then, we have the following result, assuming that the updates P_t are uniformly bounded.

Proposition 2. *Let $Q_t, R_t \geq \mu I$, $P_t \leq \nu I$, and $K_t \in \mathcal{K}$. Then, $\{K_t\}_{t\geq 1}$ is $(\bar{\kappa}, \frac{1}{2\bar{\kappa}^2}, \gamma)-$S2DA, where $\bar{\kappa} := \sqrt{\nu/\mu}$.*

Additionally, if $||P_t - P_{t+1}|| \leq p \leq \mu^2/\nu$, then, $\{K_t\}_{t\geq 1}$ is $(\bar{\kappa}, \frac{1}{2\bar{\kappa}^2}, \gamma)-$S3DA, where $\bar{\kappa} := \sqrt{\nu/\mu}$.

In the sequel, we will use $\mathcal{K}_{\kappa,\epsilon}$ to denote the set of $(\kappa, \epsilon, \gamma)-$S2DA or $(\kappa, \epsilon, \gamma)-$S3DA policies.

D. Bound on Riccati Recursion Updates

Our next result yields a bound on the difference between successive updates of the Riccati recursion (10). We achieve this by reducing our framework to the form of the recursive updates for the traditional LQR that was shown in [34], and assuming that parameter values are chosen so that an inequality in the proof will not depend on a constant term.

Theorem 2. *Let $Q_t, R_t \geq \mu I$, $Tr(Q_t), Tr(R_t) < \sigma, P_t \leq \nu I$, and $\{K_t\}_{t\geq 1}$ be $(\kappa, \epsilon, \gamma)-$S2DA. Then, there exist constants p^* and t^* such that $||P_{t+1} - P_t|| \leq p^*/t$ for all $t > t^*$.*

Proof. From Equations (10), (15), and (11),

$$\begin{aligned} P_{t+1} - P_t &= (A - BK_{t+1})^T(\tilde{P}_{t+1} - \tilde{P}_t)(A - BK_{t+1}) \\ &\quad + (\bar{Q}_{t+1} - \bar{Q}_t) + K_{t+1}^T(\bar{R}_{t+1} - \bar{R}_t)K_{t+1} \\ &\quad - (K_{t+1} - K_t)^T(\bar{R}_t + B^T\tilde{P}_tB)(K_{t+1} - K_t) \qquad (17) \\ K_{t+1} - K_t &= (B^T\tilde{P}_tB + \bar{R}_t)^{-1} \qquad (18) \\ &\quad \times (B^T(\tilde{P}_t - \tilde{P}_{t-1})(A - BK_t) + (\bar{R}_{t-1} - \bar{R}_t)K_t), \end{aligned}$$

where the last term in the last equation uses the fact that $\bar{R}_{t-1}K_t + B^T\tilde{P}_{t-1}BK_t - B^T\tilde{P}_{t-1}A = 0$. Therefore,

$$\begin{aligned} P_{t+1} - P_t = (A - BK_{t+1})^T(P_{t+1} - P_t)(A - BK_{t+1}) \\ + M_t, \qquad (19) \end{aligned}$$

where $M_t := M_{t_1} + M_{t_2} + M_{t_3}$, and

$$\begin{aligned} M_{t_1} &:= (A - BK_{t+1})^T \\ &\quad \times (P_{t+1}D(\gamma^2 I - D^TP_{t+1}D)^{-1}D^TP_{t+1} \\ &\quad - P_tD(\gamma^2 I - D^TP_tD)^{-1}D^TP_t) \\ &\quad \times (A - BK_{t+1}) \qquad (20) \\ M_{t_2} &:= (\bar{Q}_{t+1} - \bar{Q}_t) + K_{t+1}^T(\bar{R}_{t+1} - \bar{R}_t)K_{t+1} \qquad (21) \\ -M_{t_3} &:= (B^T(\tilde{P}_t - \tilde{P}_{t-1})(A - BK_t) + (\bar{R}_{t-1} - \bar{R}_t)K_t)^T \\ &\quad \times (B^T\tilde{P}_tB + \bar{R}_t)^{-1} \qquad (22) \\ &\quad \times (B^T(\tilde{P}_t - \tilde{P}_{t-1})(A - BK_t) + (\bar{R}_{t-1} - \bar{R}_t)K_t) \end{aligned}$$

Equation (19) is a Lyapunov equation. Therefore,

$$\begin{aligned} P_{t+1} - P_t &= \sum_{i=0}^{\infty}((A - BK_{t+1})^T)^i M_t (A - BK_{t+1})^i \\ &\leq ||M_t|| \sum_{i=0}^{\infty}((A - BK_{t+1})^T)^i (A - BK_{t+1})^i \end{aligned}$$

Now, $||M_t|| \leq \sum_{i=1}^{3}||M_{t_1}||$. Since K_t is $(\kappa, \epsilon, \gamma)-$S2DA, $||K_t|| \leq \kappa$, $(A - BK_{t+1}) = H_{t+1}L_{t+1}H_{t+1}^{-1}$, and we have:

$$\begin{aligned} &||\sum_{i=0}^{\infty}((A - BK_{t+1})^T)^i(A - BK_{t+1})^i|| \\ &\leq \sum_{i=0}^{\infty}\kappa^2(1-\epsilon)^{2i} = \frac{\kappa^2}{\epsilon(2-\epsilon)} \leq \frac{\kappa^2}{\epsilon} \qquad (23) \end{aligned}$$

From $Tr(Q_t), Tr(R_t) \leq \sigma$, we can write:

$$\begin{aligned} ||\bar{Q}_{t+1} - \bar{Q}_t|| &= \frac{1}{t+1}||Q_{t+1} - \bar{Q}_t|| \leq \frac{2\sigma}{t+1} \\ ||\bar{R}_{t+1} - \bar{R}_t|| &= \frac{1}{t+1}||R_{t+1} - \bar{R}_t|| \leq \frac{2\sigma}{t+1} \\ \Rightarrow ||M_{t_2}|| &\leq \frac{2\sigma(1+\kappa^2)}{t+1} \qquad (24) \end{aligned}$$

Now, consider M_{t_1}. We can write $||A - BK_{t+1}|| \leq \kappa(1 - \epsilon) \leq \kappa$, since $\epsilon \in (0, 1]$[2]. Since $0 < P_t \leq \nu I$, we can write $(\gamma^2 I - D^TP_tD)^{-1} \leq (\gamma^2 I - \nu D^TD)^{-1}$. Therefore,

$$\begin{aligned} &||(\gamma^2 I - D^TP_tD)^{-1}|| \leq ||(\gamma^2 I - \nu D^TD)^{-1}|| \\ &= ||\gamma^{-2}(I - \frac{\nu}{\gamma^2}D^TD)^{-1}|| \leq \frac{1}{\gamma^2} + \frac{\nu}{\gamma^4}||D^TD|| \end{aligned}$$

A lower bound on the norm of the middle term of M_{t_1} is

$$\begin{aligned} ||M_{t_1}|| &\leq \kappa^2 m_D, \text{ where} \qquad (25) \\ m_D &:= \frac{2\nu^2}{\gamma^2}(1 + \frac{\nu}{\gamma^2}||D^TD||)||D^TD|| \end{aligned}$$

Since $R_t \geq \mu I$, $||(B^T\tilde{P}_tB + \bar{R}_t)^{-1}|| \leq \frac{1}{\mu}$. Then, we have:

$$||M_{t_3}|| \leq \frac{1}{\mu}(\frac{2\sigma\kappa}{t} + \kappa||B||(||P_t - P_{t-1}|| + m_D))^2 \qquad (26)$$

Using the bounds in Equations (23)-(26), we have:

$$\begin{aligned} &||P_{t+1} - P_t|| \leq \frac{\kappa^2}{\epsilon}(\kappa^2 m_D + \frac{2\sigma(1+\kappa^2)}{t+1}) \\ &+ \frac{\kappa^2}{\epsilon\mu}(\frac{2\sigma\kappa}{t} + \kappa||B||(||P_t - P_{t-1}|| + m_D))^2 \\ &= \frac{2\kappa^2\sigma(1+\kappa^2)}{\epsilon(t+1)} + \frac{\kappa^2}{\epsilon\mu}(\frac{2\sigma\kappa}{t} + \kappa||B||(||P_t - P_{t-1}||))^2 \\ &\qquad (27) \\ &+ \frac{\kappa^4 m_D}{\epsilon}(\frac{2||B||}{\mu}(\frac{2\sigma}{t} + ||B||||P_t - P_{t-1}||) + ||B||^2 m_D + 1) \end{aligned}$$

To complete the proof, we make the following assumption.

[2] Note that $|\lambda_{max}(A - BK)| < ||A - BK||$, where the (two-)norm of a matrix is given by its maximum singular value.

Assumption 2. *μ, ν, γ are chosen so that the inequality (27) will be true independent of the last term of (27) for all $t > t^*$.*

Future work will examine the relaxation of this assumption in greater detail. This setting is now similar to that in Lemma A.6 in [34]. Therefore, if there is some p^* and t^* such that for all $t > t^*$, $(||P_t - P_{t-1}||) \leq p^*/t$, then $(||P_{t+1} - P_t||) \leq p^*/(t+1)$. Specifically, this will be true for[3]:

$$\begin{aligned} t > t^* &= \frac{8\sigma\kappa^4||B||}{\epsilon\mu}(1 + \frac{\kappa^2||B||(1+\kappa^2)}{\epsilon}) \\ p^* &\leq \frac{2\sigma}{||B||} + \frac{4\kappa^2\sigma(1+\kappa^2)}{\epsilon} \end{aligned}$$

The base case of the induction can be shown as in [34]. □

E. Online Algorithm

Algorithm 1 Safety-Critical Online Controller Synthesis

```
 1: procedure GENERATE $\{K_t\}_{t>1}$
 2:     Input: System: $x_{t+1} = Ax_t + Bu_t + Dw_t$, initial state, parameters $\mu, \nu, \kappa := \sqrt{\nu/\mu}, \epsilon = 1/(2\kappa^2), \gamma, \sigma, K_1 \in \mathcal{K}_{\kappa,\epsilon}$, time horizon $T$
 3:     Output: $\{K_t\}_{t>1}$, such that $K_t \in \mathcal{K}_{\kappa,\epsilon}$
 4:     for $t = 1, 2, \ldots, T$ do
 5:         obtain current state $x_t$
 6:         generate $u_t = -K_t x_t$
 7:         adversary plays $w_t$
 8:         adversary generates $C_t, E_t$ (Assumption 1)
 9:         $Q_t := C_t^T C_t$; $R_t := E_t^T E_t$
10:         update $\bar{R}_t$, $\bar{Q}_t$ acc. to Eqns. (8)-(9)
11:         update $P_t$ according to Eqn. (10)
12:         if $t = \lceil \frac{8\sigma\kappa^4||B||}{\epsilon\mu}(1 + \frac{\kappa^2||B||(1+\kappa^2)}{\epsilon}) \rceil$ then
13:             $d := 0$, $P_0 := P_{t^*}$, $K_0 := K_{t^*}$
14:             $d \leftarrow d + 1$
15:             successively solve Eqn. (10) as long as $||P_d - P_{d-1}|| > p^*/t^*$; update $K_d$ according to Eqn. (11)
16:         end if
17:         return $K_{t+1}$ according to Eqn. (11)
18:     end for
19: end procedure
```

From Assumption 1 and Theorem 1, if we start at $t = 1$ from a stabilizing policy that attenuates the disturbance, then our update procedure will continue to yield stabilizing, disturbance attenuating policies for all $t > 1$. At each step, we compute $u_t = -K_t x_t$, and the output and cost are revealed in terms of the matrices C_t, and E_t, where C_t and E_t satisfy Assumption 1. The update is carried out according to Equations (8)-(9) by averaging over previous values of $Q_t := C_t^T C_t$ and $R_t := E_t^T E_t$. From Theorem 2, $||P_{t+1} - P_t|| < p^*/t$ for $t > t^*$ (*Lines 12-16*). Algorithm 1 formally presents this procedure.

F. Regret Bounds

Iterative solutions to the Riccati equation in the LTI case exhibit quadratic convergence to an optimal solution P^* [44]. In [41], this convergence rate was also shown to hold for the variant of the Riccati equation that we use in this paper. Specifically, for some $c > 0$, $||P_t - P^*|| \leq c||P_{t-1} - P^*||^2$, and $||P_{t+1} - P_t|| \leq c||P_t - P_{t-1}||^2$. Further, observe that $R(T)$ in Problem 1 can be written as $R(T) = Tr(P_T DD^T) - Tr(P_T^* DD^T)$, where P_T^* corresponds to the solution of the Riccati equation that yields the optimal controller from the set $\mathcal{K}$. We use these results to establish a bound on the growth of the regret for sufficiently large T.

Theorem 3. *Let the conditions of Assumption 1 hold, and let $Q_t, R_t \geq \mu I$, $Tr(Q_t), Tr(R_t) < \sigma$, $P_t \leq \nu I$, $\kappa = \sqrt{\nu/\mu}$, $\epsilon = 1/2\kappa^2$. Let the controllers $\{K_t\}_{t\geq 1}$ be $(\kappa, \epsilon, \gamma)-$S2DA, and $DD^T > 0$. Then, for $T \geq t^* = \frac{8\sigma\kappa^4||B||}{\epsilon\mu}(1 + \frac{\kappa^2||B||(1+\kappa^2)}{\epsilon})$, $p^* \leq \frac{2\sigma}{||B||} + \frac{4\kappa^2\sigma(1+\kappa^2)}{\epsilon}$, and some constant $m > 0$, $R(T) \leq Tr(DD^T)(\log(T) + \frac{2mp^*}{t^*+1} - \log(t^*))$.*

Proof. With $J_0(\cdot) = 0$, we can express $R(T)$ as:

$$\begin{aligned} R(T) &= \sum_{t=1}^{T}(Tr(P_t DD^T) - Tr(P_{t-1}DD^T)) - Tr(P_T^* DD^T) \\ &= \sum_{t=1}^{t^*}(Tr(P_t DD^T) - Tr(P_{t-1}DD^T)) - Tr(P_T^* DD^T) \\ &+ \sum_{t=t^*}^{T}(Tr(P_t DD^T) - Tr(P_{t-1}DD^T)) \\ &= Tr(P_{t^*}DD^T) - Tr(P^* DD^T) + Tr(P^* DD^T) \\ -Tr(P_T^* DD^T) &+ \sum_{t=t^*}^{T}(Tr(P_t DD^T) - Tr(P_{t-1}DD^T)) \\ &\leq 2Tr(DD^T)||P_{t^*} - P^*|| + Tr(DD^T)\sum_{t=t^*}^{T}||P_t - P_{t-1}||, \end{aligned}$$

where P^* is the optimal solution to the time-invariant, infinite-horizon Riccati equation. In the above, the first term can be interpreted as a *burn-in cost*, that is, the cost incurred before the procedure starts to yield meaningful regret bounds, while the second term gives the bound for large enough T.

From Theorem 2, $\sum_{t=t^*}^{T}||P_t - P_{t-1}|| \leq \sum_{t=t^*}^{T} p^*/t \leq \log(\frac{T}{t^*})$, while for the first term, we use the quadratic convergence to P^* to obtain $||P_{t^*} - P^*|| \leq \frac{mp^*}{t^*+1}$. Here, m is a constant associated with $\lim\limits_{t\to\infty}\sum\limits_{i=0}^{t}||P_{t^*+i} - P_{t^*+i+1}||$. Therefore, $R(T) \leq Tr(DD^T)(\log(T) + \frac{2mp^*}{t^*+1} - \log(t^*))$. □

The regret bound in our case differs from those shown in related work (e.g. [17], [19], [34]) due to the nature of the cost function that we seek to optimize in this paper. Since we are interested in the minimization of an (upper bound on

[3] These thresholds can be obtained by expanding the quadratic term on the right-hand side of Equation (27) and using Assumption 2 to get a quadratic inequality in p^*. That is, we get a quadratic $a(p^*)^2 + bp^* + c \leq 0$, where a, b, c are terms involving t and the constants in Equation (27). The bound on t is obtained by recognizing that $(t+1)/t^2 \approx 1/t$, and requiring that the roots of this quadratic inequality be real, that is, $\sqrt{b^2 - 4ac} > 0$. The bound on p^* is then got by requiring that $p^* \in [p_1, p_2]$, where p_1 and p_2 are roots of the quadratic equation $a(p^*)^2 + bp^* + c = 0$. Specifically, we set $p^* \leq -b/2a$ so that the quadratic inequality will be satisfied.

the) $\mathcal{H}_2$ cost, given by $\lim_{t\to\infty} \mathbb{E}[z_t^T z_t]$, when w_t is white noise, our regret term of the form in Problem 1 can be recast in the form on the first line of the above proof.

V. Experimental Evaluation

We validate our method on a well-studied problem from process control called the Tennessee Eastman control challenge [22]. The irreversible and exothermic process (Figure 1) produces two products (G, H) from four reactants (A, C, D, E); component F represents other products formed from side reactions in the process. The open-loop process is unstable, which necessitates the use of feedback control. This model has been adapted to demonstrate the use of machine learning methods to study resilience to attacks [46], fault detection [47], and impacts of advanced persistent threats [48]. A continuous-time LTI model of the plant presented in [49] consisted of eight states, four inputs, and ten outputs. We use values of the A, B, C matrices from [49], and discretize the model, assuming a zero order hold. We additionally assume $w_t \in \mathbb{R}^8$, $D = I_{8\times 8}$, and E chosen to satisfy Assumption 1. We use these values of C and E to determine the (optimal) counterfactual static controller $K \in \mathcal{K}$.

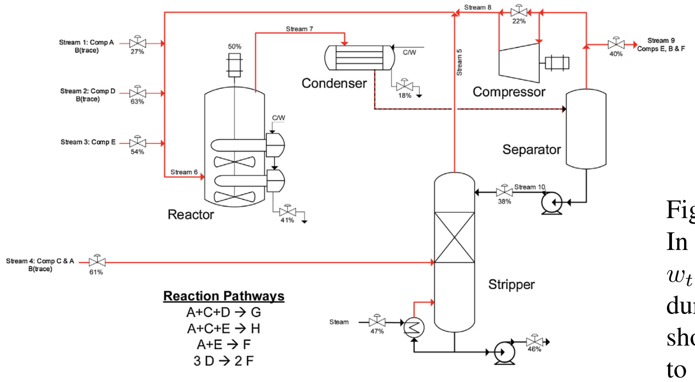
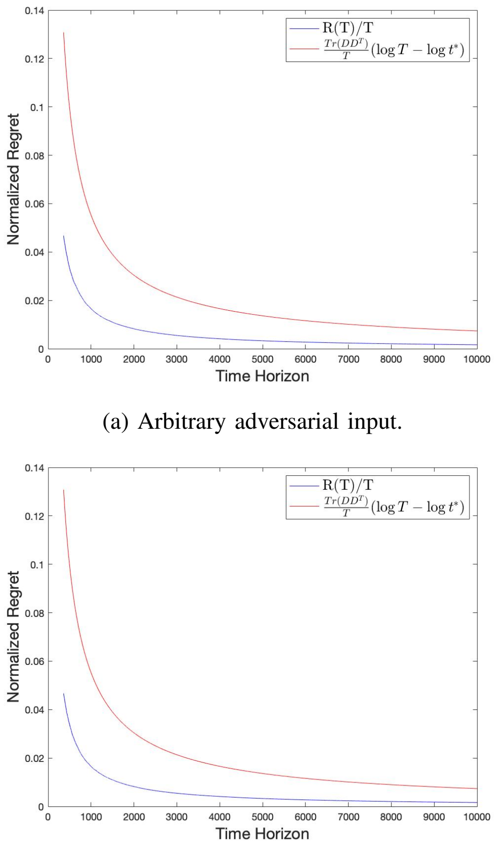


Fig. 1: Flow diagram of the Tennessee Eastman Process [50]

We consider two attack scenarios. In the first, at each time step, the adversary injects an arbitrary input w_t. We make no assumptions on the nature of this input (except that it is bounded, for the purpose of simulation). In the second, we simulate a denial-of-service attack, by setting $w_t = -Bu_t$ for $t \in [t_a, t'_a]$, and arbitrary at other times. $[t_a, t'_a]$ is the attack duration such that $1 < t_a \leq t'_a$. In this case, the impact of the controller on the evolution of the state is canceled during the attack, but the learner still incurs a cost associated to the control (as a result of the E_t term). In each case, the matrices C_t and E_t are perturbed versions of C and E.

The normalized regret for the two attacks are shown in Figure 2. In particular, we observe that the regret of a sequence of controllers computed according to Algorithm 1 with respect to the optimal, counter-factual, time-invariant static controller that is obtained by solving the Riccati equation for the time-invariant case satisfies the bounds determined in Theorem 3. For the denial-of-service attack, although the effect of the controller is canceled for the duration of the attack, as long as this attack starts at $t_a > 1$, Algorithm 1 will continue to produce stabilizing, disturbance attenuating controllers if we start from an initial controller that is stabilizing and disturbance attenuating (Theorem 1).



(a) Arbitrary adversarial input.

(b) Denial-of-service attack.

Fig. 2: Normalized regret for two types of adversarial input. In Fig. 2a, the adversary input, w_t is arbitrary. In Fig. 2b, w_t cancels the effect of the control u_t on the state evolution during the attack, denoting denial-of-service. The blue curves show the normalized regret of a controller chosen according to Algorithm 1 with respect to the optimal, counter-factual, time-invariant static controller. The red curves denote the (normalized) right-hand side of the regret bound of Thm. 3.

VI. Conclusion

This paper presented an iterative solution to an online control problem in the presence of bounded adversarial disturbances. In this setting, costs incurred by the system at each time due to an adversarial disturbance input were revealed only after the input was given. We synthesized controllers to minimize (an upper bound of) a quadratic cost while simultaneously satisfying a safety constraint. This was achieved by solving a Riccati equation in an iterative manner. Solutions to the Riccati equation enforced the safety constraint. We showed that initializing the procedure with a stabilizing and disturbance attenuating controller ensured that controllers at successive time steps retained this property. We showed that the regret of this controller, compared to the optimal controller when all costs and disturbances were known in hindsight, varied logarithmically with the time horizon. We validated our approach on a model of the Tennessee Eastman chemical process that was subject to arbitrary adversarial inputs and a denial of service attack.

Future work will study the partial information setting, where one will have to synthesize dynamic output feedback controllers, and the more generalized problem of minimizing the $\mathcal{H}_\infty$ norm of the output to disturbance map. We will also extend our analysis to the case of unknown system dynamics.

References

[1] R. S. Sutton and A. G. Barto, *Reinforcement Learning: An Introduction*. MIT press, 2018.

[2] R. Hafner and M. Riedmiller, "Reinforcement learning in feedback control," *Machine Learning*, vol. 84, pp. 137–169, 2011.

[3] V. Mnih *et al.*, "Human-level control through deep reinforcement learning," *Nature*, vol. 518, no. 7540, 2015.

[4] T. P. Lillicrap *et al.*, "Continuous control with deep reinforcement learning," in *International Conference on Learning and Representations*, 2016.

[5] D. Silver *et al.*, "Mastering the game of Go with deep neural networks and tree search," *Nature*, vol. 529, no. 7587, 2016.

[6] C. Zhang, P. Patras, and H. Haddadi, "Deep learning in mobile and wireless networking: A survey," *IEEE Communications Surveys & Tutorials*, vol. 21, no. 3, pp. 2224–2287, 2019.

[7] D. Sadigh, S. Sastry, S. A. Seshia, and A. D. Dragan, "Planning for autonomous cars that leverage effects on human actions." in *Robotics: Science and Systems*, 2016.

[8] Z. Yan and Y. Xu, "Data-driven load frequency control for stochastic power systems: A deep reinforcement learning method with continuous action search," *IEEE Transactions on Power Systems, 34(2)*, 2018.

[9] C. You, J. Lu, D. Filev, and P. Tsiotras, "Advanced planning for autonomous vehicles using reinforcement learning and deep inverse RL," *Robotics and Autonomous Systems*, vol. 114, pp. 1–18, 2019.

[10] F. L. Lewis, D. Vrabie, and K. G. Vamvoudakis, "Reinforcement learning and feedback control: Using natural decision methods to design optimal adaptive controllers," *IEEE Control Systems Magazine*, vol. 32, no. 6, pp. 76–105, 2012.

[11] A. Banerjee, K. K. Venkatasubramanian, T. Mukherjee, and S. K. Gupta, "Ensuring safety, security, and sustainability of mission-critical cyber-physical systems," *Proceedings of the IEEE, 100(1)*, 2012.

[12] J. E. Sullivan and D. Kamensky, "How cyber-attacks in Ukraine show the vulnerability of the US power grid," *The Electricity Journal*, vol. 30, no. 3, pp. 30–35, 2017.

[13] W. M. Haddad, D. S. Bernstein, and D. Mustafa, "Mixed-norm H_2/H_∞ regulation and estimation: The discrete-time case," *Systems & Control Letters*, vol. 16, no. 4, pp. 235–247, 1991.

[14] D. Mustafa and D. S. Bernstein, "LQG cost bounds in discrete-time H_2/H_∞ control," *Transactions of the Institute of Measurement and Control*, vol. 13, no. 5, pp. 269–275, 1991.

[15] I. Kaminer, P. P. Khargonekar, and M. A. Rotea, "Mixed H_2/H_∞ control for discrete-time systems via convex optimization," *Automatica*, vol. 29, no. 1, pp. 57–70, 1993.

[16] N. Agarwal, E. Hazan, and K. Singh, "Logarithmic regret for online control," in *Advances in Neural Information Processing Systems*, 2019.

[17] N. Agarwal, B. Bullins, E. Hazan, S. Kakade, and K. Singh, "Online control with adversarial disturbances," in *International Conference on Machine Learning*, 2019, pp. 111–119.

[18] E. Hazan, S. M. Kakade, and K. Singh, "The nonstochastic control problem," in *Algorithmic Learning Theory*, 2020, pp. 408–421.

[19] M. Simchowitz, K. Singh, and E. Hazan, "Improper learning for non-stochastic control," in *Conference on Learning Theory*, 2020.

[20] D. J. Foster and M. Simchowitz, "Logarithmic regret for adversarial online control," in *International Conference on Machine Learning*, 2020.

[21] A. Cohen, A. Hasidim, T. Koren, N. Lazic, Y. Mansour, and K. Talwar, "Online linear quadratic control," in *International Conference on Machine Learning*, 2018, pp. 1029–1038.

[22] J. J. Downs and E. F. Vogel, "A plant-wide industrial process control problem," *Computers & Chemical Engineering, 17(3)*, 1993.

[23] M. Aliyu and E. Boukas, "Discrete-time mixed H_2/H_∞ nonlinear filtering," in *Proc. American Control Conference*, 2008.

[24] W. M. Wonham, *Linear multivariable control: A Geometric Approach*. Springer, 1974.

[25] J. Willems, "Almost invariant subspaces: An approach to high gain feedback design–part I: Almost controlled invariant subspaces," *IEEE Transactions on Automatic Control*, vol. 26, no. 1, pp. 235–252, 1981.

[26] G. Basile and G. Marro, *Controlled and conditioned invariants in linear system theory*. Prentice Hall Englewood Cliffs, NJ, 1992.

[27] K. Furuta and M. Wongsaisuwan, "Closed-form solutions to discrete-time LQ optimal control and disturbance attenuation," *Systems & Control Letters*, vol. 20, no. 6, pp. 427–437, 1993.

[28] A. Saberi, Z. Lin, and A. A. Stoorvogel, "H_2 and H_∞ almost disturbance decoupling problem with internal stability," *International Journal of Robust and Nonlinear Control*, vol. 6(8), 1996.

[29] Z. Lin and B. M. Chen, "Solutions to general H_∞ almost disturbance decoupling problem with measurement feedback and internal stability for discrete-time systems," *Automatica*, vol. 36(8), 2000.

[30] K. Zhou, J. C. Doyle, and K. Glover, *Robust and optimal control*. Prentice Hall New Jersey, 1996, vol. 40.

[31] A. Bemporad and M. Morari, "Robust model predictive control: A survey," in *Robustness in identification and control*. Springer, 1999.

[32] S. V. Raković and W. S. Levine, *Handbook of model predictive control*. Springer, 2018.

[33] B. Recht, "A tour of reinforcement learning: The view from continuous control," *Annual Review of Control, Robotics, and Autonomous Systems*, vol. 2, pp. 253–279, 2019.

[34] M. Akbari, B. Gharesifard, and T. Linder, "An iterative Riccati algorithm for online linear quadratic control," *arXiv preprint arXiv:1912.09451*, 2019.

[35] A. Tsiamis and G. Pappas, "Online learning of the Kalman filter with logarithmic regret," *arXiv preprint arXiv:2002.05141*, 2020.

[36] S. Dean, H. Mania, N. Matni, B. Recht, and S. Tu, "On the sample complexity of the linear quadratic regulator," *Foundations of Computational Mathematics*, pp. 1–47, 2019.

[37] A. Tsiamis, N. Matni, and G. J. Pappas, "Sample complexity of Kalman filtering for unknown systems," in *Learning for Dynamics and Control*, 2020, pp. 435–444.

[38] M. Fazel, R. Ge, S. Kakade, and M. Mesbahi, "Global convergence of policy gradient methods for the linear quadratic regulator," in *International Conference on Machine Learning*, 2018, pp. 1467–1476.

[39] S. Tu and B. Recht, "The gap between model-based and model-free methods on the linear quadratic regulator: An asymptotic viewpoint," in *Conference on Learning Theory*, 2019, pp. 3036–3083.

[40] B. Gravell, P. M. Esfahani, and T. Summers, "Learning robust control for linear quadratic systems with multiplicative noise via policy gradient," *arXiv preprint arXiv:1905.13547*, 2019.

[41] K. Zhang, B. Hu, and T. Başar, "Policy optimization for H_2 linear control with H_∞ robustness guarantee: Implicit regularization and global convergence," in *Learning for Dynamics and Control*, 2020, pp. 179–190.

[42] S. Dean, S. Tu, N. Matni, and B. Recht, "Safely learning to control the constrained linear quadratic regulator," in *Proc. American Control Conference*, 2019, pp. 5582–5588.

[43] M. Green and D. J. Limebeer, *Linear robust control*. Dover, 2012.

[44] G. Hewer, "An iterative technique for the computation of the steady state gains for the discrete optimal regulator," *IEEE Transactions on Automatic Control*, vol. 16, no. 4, pp. 382–384, 1971.

[45] W. J. Rugh, *Linear System Theory*. Prentice Hall, 1996.

[46] A. Keliris, H. Salehghaffari, B. Cairl, P. Krishnamurthy, M. Maniatakos, and F. Khorrami, "Machine learning-based defense against process-aware attacks on industrial control systems," in *IEEE International Test Conference*, 2016, pp. 1–10.

[47] W. Zou, Y. Xia, and H. Li, "Fault diagnosis of Tennessee-Eastman process using orthogonal incremental extreme learning machine based on driving amount," *IEEE Transactions on Cybernetics, 48(12)*, 2018.

[48] L. Huang and Q. Zhu, "A dynamic games approach to proactive defense strategies against advanced persistent threats in cyber-physical systems," *Computers & Security*, vol. 89, p. 101660, 2020.

[49] N. L. Ricker, "Model predictive control of a continuous, nonlinear, two-phase reactor," *Journal of Process Control*, vol. 3(2), 1993.

[50] I. A. Udugama, K. V. Gernaey, M. A. Taube, and C. Bayer, "A novel use for an old problem: The Tennessee Eastman challenge process as an activating teaching tool," *Education for Chemical Engineers*, vol. 30, pp. 20–31, 2020.