2020 American Control Conference
Denver, CO, USA, July 1-3, 2020

Gain Design via LMIs to Minimize the Impact of Stealthy Attacks

Navid Hashemi[1] and Justin Ruths[1]

Abstract— **The goal of this paper is to design the gain matrices for estimate-based feedback to minimize the impact that falsified sensor measurements can have on the state of a stochastic linear time invariant system. Here we consider attackers that stay stealthy, by raising no alarms, to a chi-squared anomaly detector, thereby restricting the set of attack inputs within an ellipsoidal set. We design linear matrix inequalities (LMIs) to find a tight outer ellipsoidal bound on the convex set of states reachable due to the stealthy inputs (and noise). Subsequently considering the controller and estimator gains as design variables requires further linearization to maintain the LMI structure. Without a competing performance criterion, the solution of this gain design is the trivial uncoupling of the feedback loop (setting either gain to zero). Here we consider - and convexify - an output constrained covariance (OCC) $\|H\|_2$ gain constraint on the non-attacked system. Through additional tricks to linearize the combination of these LMI constraints, we propose an iterative algorithm whose core is a combined convex optimization problem to minimize the state reachable set due to the attacker while ensuring a small enough $\|H\|_2$ gain during nominal operation.**

I. Introduction

One of the most fundamental problems in Control theory is the design problem to select gain matrices for controllers and estimators that guarantee certain properties of a feedback control system. It is fitting then that the literature on security of control systems consider this question as well. Incorporating multiple competing criteria for gain design creates the inherent trade-offs that make this class of problem interesting and challenging. In this paper, we develop semidefinite programming (SDP) problems to minimize the impact an attacker can have on a feedback control system, balanced by a ($\|H\|_2$ gain) performance criterion.

A major thrust to defend against attacks on control systems has been to leverage methods from fault detection to build model-based observers that monitor the evolution of the output and raise alarms when the output changes in ways that is unexpected. In this line of work, groups have tuned and analyzed different types of such detectors such as chi-squared, CUSUM [1], and MEMWA [2]; quantified the impact of attackers [2], [3], [4], [5]; and designed attacks that are stealthy to detectors [6], [7]. This work has helped to quantify the impact of an attack, when the adversary wishes to remain stealthy to the detector as the detector imposes a constraint on the attack. From this framework we are now able to consider the gain design problem to minimize the impact of attacks on control systems. The core security metric that has risen from this literature is the size of the states that are reachable by the system, when the system is driven by the attack.

The first versions of this work focused only on observer gain design and aimed at minimizing the reachable estimation errors instead of the states themselves [7]. One of the important realizations that came from this work was that there is a trade-off between security and performance, since it is possible to set the observer gain to zero, cutting off the attacker and hence minimizing their impact, but also cutting the feedback loop. More recently, the combined controller and observer design problem with state reachable set has been studied [5]. While the goal here is fundamentally the same, the difference in controller and observer model introduces a different set of technical challenges. In this paper, we consider estimate feedback, where the estimate is generated by an observer of Luenberger form. This creates a dependency between the controller and observer gains that requires extra effort to linearize. In [5], the authors consider a dynamic output feedback controller and the independence of the controller matrices from the observer gain (the estimator is used for detection, not for feedback) makes the problem marginally easier to linearize.

In this paper, we introduce an output covariance constrained (OCC) $\|H\|_2$ performance criterion to avoid the trivial zero-gain solution. In this context, the structure (covariance) of the noise distribution gives important information about which directions are more or less vulnerable to attack and to amplification. To our knowledge, the gain design problem for OCC $\|H\|_2$ has not been completely convexified in the way we have done in this paper (Theorem 1). This is a necessary step to yield an overall SDP problem for the combined security and performance gain design. To our knowledge, this problem has, in the past, been solved with an iterative algorithm [8], [9]. This performance choice also differentiates our work with [5], which considers a distributionally robust $\|H\|_\infty$ constraint for performance, in the sense that it does not incorporate information about the noise distributions.

II. Background

Here we consider a discrete-time linear time invariant (LTI) system of the form

$$x_{k+1} = Fx_k + Gu_k + \nu_k, \tag{1}$$

$$y_k = Cx_k + \eta_k. \tag{2}$$

in which the state $x_k \in \mathbb{R}^n$, $k \in \mathbb{N}$, evolves due to the state update provided by the state matrix $F \in \mathbb{R}^{n\times n}$, the control input $u_k \in \mathbb{R}^m$ shaped by the input matrix $G \in \mathbb{R}^{n\times m}$, and the i.i.d Gaussian system noise $\nu_k \sim \mathcal{N}(0, R_1)$, $R_1 \in \mathbb{V}^{n\times n}$

[1] The authors are with the Departments of Mechanical and Systems Engineering at the University of Texas at Dallas. Email: (nxh150030, jruths)@utdallas.edu

978-1-5386-8266-1/$31.00 ©2020 AACC

($\mathbb{V}$ is the set of positive definite matrices). The output $y_k \in \mathbb{R}^p$ aggregates a linear combination of the states, given by the observation matrix $C \in \mathbb{R}^{p\times n}$, and the i.i.d Gaussian measurement noise $\eta_k \sim \mathcal{N}(0, R_2)$, $R_2 \in \mathbb{V}^{p\times p}$. For the simplicity of the exposition, we have considered Gaussian noises, however, the approach we present is applicable for general noise distributions. In addition we assume that F is stable, the pair (F, C) is detectable, (F, G) is stabilizable and system and measurement noises are mutually independent.

In this work, we consider the scenario that the actual measurement y_k can be corrupted by an attack, $\delta_k \in \mathbb{R}^p$. The attack is injected at some point between the measurement and reception of the output by the controller,

$$\bar{y}_k = y_k + \delta_k = Cx_k + \eta_k + \delta_k. \tag{3}$$

If the attacker has access to the measurements, then it is possible for the attack δ_k to cancel some or all of the original measurement y_k - so an additive attack can achieve arbitrary control over the "effective" output of the system.

Because our system is stochastic, we require an estimator to produce a prediction of the system behavior

$$\hat{x}_{k+1} = F\hat{x}_k + Gu_k + L(\bar{y}_k - C\hat{x}_k), \tag{4}$$

where $\hat{x}_k \in \mathbb{R}^n$ is the estimated state and the observer gain L is designed to force the estimate to track the system states.

We consider observer-based output feedback controllers

$$u_k = K\hat{x}_k, \tag{5}$$

where $K \in \mathbb{R}^{m\times n}$ is the controller gain matrix. Next, we define the residual sequence

$$r_k = \bar{y}_k - C\hat{x}_k, \tag{6}$$

as the difference between what we actually receive ($\bar{y}_k$) and expect to receive ($C\hat{x}_k$), which evolves according to

$$\begin{aligned} x_{k+1} &= (F + GK)x_k - GKe_k + \nu_k \\ e_{k+1} &= \big(F - LC\big)e_k - L\eta_k + \nu_k - L\delta_k, \\ r_k &= Ce_k + \eta_k + \delta_k, \end{aligned} \tag{7}$$

where $e_k = x_k - \hat{x}_k$ is the estimation error. In the absence of attacks (i.e., $\delta_k = 0$), we can show that the steady-state distribution of r_k is Gaussian with covariance $\Sigma = C\mathbf{P}_eC^T + R_2$ [4], where the steady state covariance of the estimation error $\mathbf{P}_e = \lim_{k\to\infty} P_k = \lim_{k\to} \mathbf{E}[e_ke_k^T]$ is the solution of

$$\mathbf{P}_e = (F - LC)\mathbf{P}_e(F - LC)^T + LR_2L^T + R_1. \tag{8}$$

In this work, we consider the chi-squared detector, although similar analysis can be done with other detector choices [10], [11], [12]. The chi-squared detector constructs a quadratic distance measure z_k to be sensitive to changes in the variance of the distribution as well as the expected value,

$$z_k = r_k^T\Sigma^{-1}r_k, \tag{9}$$

and generates alarms when the distance measure exceeds a threshold $\alpha \in \mathbb{R}_{>0}$, i.e., $z_k > \alpha$ raises alarms.

A. *Definition of Attack*

Detectors are designed to identify anomalies in system behavior. If an attacker aims to remain undetected, the choice detector and its parameters limit what the attacker is able to accomplish. The type of attacks we consider here require strong knowledge of and access to system dynamics, statistics of noise, current estimate ($\hat{x}$), and the detector configuration. The goal of this powerful stealthy attack is to construct the worst case scenario to aid the design of more robust systems.

Zero-alarm attacks are attack sequences δ_k that maintain the distance measure at or below the threshold of detection, i.e., $z_k \le \alpha$, thus producing no alarms,

$$\delta_k = \phi_k - (y_k - C\hat{x}_k) = -Ce_k - \eta_k + \phi_k, \tag{10}$$

where $\phi_k \in \mathbb{R}^p$ is any vector such that $\phi_k^T\Sigma^{-1}\phi_k \le \alpha$. So,

$$\begin{aligned} z_k &= r_k^T\Sigma^{-1}r_k = (Ce_k + \eta_k + \delta_k)^T\Sigma^{-1}(Ce_k + \eta_k + \delta_k), \\ &= \phi_k^T\Sigma^{-1}\phi_k \le \alpha. \end{aligned} \tag{11}$$

Thus $z_k \le \alpha$ and no alarms are raised.

B. *Reachable Set*

Under a stealthy zero-alarm attack (10), the attacked system dynamics become, defining a stacked state $\xi_k = \big[x_k^T, \hat{x}_k^T, e_k^T\big]^T$ and input $\mu_k = \big[\nu_k^T, \phi_k^T\big]^T$,

$$\xi_{k+1} = A\xi_k + B\mu_k, \tag{12}$$

with

$$A = \begin{bmatrix} F & GK & 0 \\ LC & F + GK - LC & -LC \\ 0 & 0 & F \end{bmatrix}, \; B = \begin{bmatrix} I & 0 \\ 0 & L \\ I & -L \end{bmatrix}. \tag{13}$$

Throughout the rest of the paper we will use a selection matrix $E_x = [I_n, 0_{n\times n}, 0_{n\times n}]$ to pull out quantities relevant to the state $x_k = E_x\xi_k$.

The reachable set of states is then,

$$\mathcal{R} = \left\{ x_k = E_x\xi_k \;\middle|\; \begin{array}{l} \xi_{k+1} = A\xi_k + B\mu_k, \\ \xi_1 = 0,\ \phi_k^T\Sigma^{-1}\phi_k \le \alpha, \\ \nu_k^TR_1^{-1}\nu_k \le \bar{\nu},\ \forall k \in \mathbb{N} \end{array} \right\}, \tag{14}$$

where the ellipsoidal bound on the attack ϕ_k is imposed by the attacker's desire to remain stealthy (11), and the ellipsoidal bound on the noise is created by truncating the Gaussian system noise to a desired probability, i.e., $\Pr[\nu_k^TR_1^{-1}\nu_k \le \bar{\nu}] = p_\nu$, where p_ν is some desired (typically high) probability.

III. LMI Approach to Design K and L

In the following subsections we build the framework of linear matrix inequalities to design the observer and controller gains to minimize the impact of a stealthy attacker subject to a performance constraint. The proofs have been omitted for space but are available online [13].

A. Bounding Ellipsoid LMI (given K and L)

Before we move on to the *synthesis* problem of designing the gain matrices, we first provide a solution to the *analysis* problem of finding a tight outer ellipsoidal bound of the reachable set given K and L, when the system is driven by the system noise and attack. A similar analysis result appears in [4], however, there the problem is split into two optimizations - one to find a bound on the estimation error reachable set, the result of which is used in the second optimization to bound the state reachable set. Here, in Lemma 1, we solve these simultaneously through the stacked states ξ_k and inputs μ_k. We will use the notation $\mathcal{E}(\mathcal{Q}) = \{x \mid x^T\mathcal{Q}^{-1}x \leq 1\}$ to define an ellipsoid with shape matrix $\mathcal{Q}$.

Lemma 1: Given the stacked system matrices A and B in (13), gain matrices K, L, detector threshold α with steady state residual covariance Σ, system noise truncation threshold $\bar{\nu}$ with covariance R_1, if there exists constants $a, a_1, a_2 \in [0, 1)$, the solution of

$$\begin{cases} \min\limits_{a_1,a_2,\mathcal{Q}} \mathbf{tr}(E_x\mathcal{Q}E_x^T) \\ \quad \text{s.t.} \;\; 0 \leq a_1, a_2 < 1, \quad a_1 + a_2 \geq a, \\ \qquad \begin{bmatrix} a\mathcal{Q} & \mathcal{Q}A^T & 0 \\ A\mathcal{Q} & \mathcal{Q} & B \\ 0 & B^T & \frac{1-a}{2-a}W \end{bmatrix} \geq 0, \end{cases} \tag{15}$$

provides the shape matrix $\mathcal{Q}$ of the ellipsoidal bound on the reachable set of states, i.e., $\mathcal{R} \subseteq \mathcal{E}(E_x^T\mathcal{Q}E_x)$, where

$$W = \begin{bmatrix} \frac{1-a_1}{\bar{\nu}^2}R_1^{-1} & 0 \\ 0 & \frac{(1-a_2)}{\alpha}\Sigma^{-1} \end{bmatrix}. \tag{16}$$

Proof: This is a direct result of Lemma 1 in [5] where the stacked dynamics (12) is driven by two ellipsoidally bounded inputs with $W_1 = R_1^{-1}$ and $W_2 = \Sigma^{-1}$. We define the positive definite function to be a quadratic function of the state, $V_k = \xi_k^T\big(\frac{2-a}{1-a}\mathcal{P}\big)\xi_k \leq \frac{2-a}{1-a}$, which results in the LMI,

$$\mathcal{H} = \begin{bmatrix} a\mathcal{P} & A^T\mathcal{P} & 0 \\ \mathcal{P}A & \mathcal{P} & \mathcal{P}B \\ 0 & B^T\mathcal{P} & \frac{1-a}{2-a}W \end{bmatrix} \geq 0, \tag{17}$$

where $\mathcal{P} > 0$ is the inverse of the shape matrix of the ellipsoidal bound for the ξ reachable set ($\mathcal{P}^{-1} = \mathcal{Q}$), such that the first block $E_x\mathcal{Q}E_x^T$ is the shape matrix of the ellipsoidal bound of the reachable set of system states. To make this ellipsoidal bound tight (as small as possible), the cost is selected to minimize the trace of the shape matrix $E_x\mathcal{Q}E_x^T$. To use $\mathcal{Q}$ as the variable of the optimization instead of $\mathcal{P}$ we apply the transformation $T = \operatorname{diag}[\mathcal{Q}, \mathcal{Q}, I_n]$, to (17), i.e., $T^T\mathcal{H}T$, which results in the LMI in (15). ∎

B. Output Covariance Constrained (OCC) $\|H\|_2$ Constraint

The introduction of this section and past related work has identified that trivial solutions exist unless a performance criteria is imposed in the optimization [7], [5]. One of the distinguishing features of this work is that we consider an output covariance constrained $\|H\|_2$ constraint, which involves the covariances of the system and sensor noises. The challenge, tackled in the next subsection, is to convexify and linearlize this inherently nonlinear constraint. Most optimizations in the literature either use a distributionally robust constraint that is already convex [5] or solve the OCC $\|H\|_2$ using iterative algorithms [8], [9]. Here, for the system *without attack*, we consider the system driven by system and measurement noise and enforce an $\|H\|_2$ constraint between the output y_k and excitation $\omega_k = \big[\nu_k^T,\, \eta_k^T\big]^T$.

When there is no attack the system evolves according to

$$\zeta_{k+1} = \hat{A}\zeta_k + \hat{B}\omega_k, \tag{18}$$

using the stacked state $\zeta_k = E_{x\hat{x}}\xi_k = \big[x_k^T,\, \hat{x}_k^T\big]^T$ and with $E_{x\hat{x}} = [I_{2n},\, 0_{2n\times n}]$ making $\hat{A} = E_{x\hat{x}}AE_{x\hat{x}}^T$ and $\hat{B} = E_{x\hat{x}}BE_{x\hat{x}}^T$.

Remark 1: Note that the seemingly redundant definition of ξ_k allows the state matrix without attack $\hat{A}$ to be expressed as a sub-block of the state matrix under attack A. Establishing this parallel structure is key towards being able to integrate the $\|H\|_2$ constraint (without attack) with the reachable set calculation (under attack).

The OCC $||H||_2$ criteria specifies the gain from the noise to the output should be less than a desired value $\bar{\gamma}$,

$$\lim_{N\to\infty}\sqrt{\frac{\frac{1}{N}\sum_{k=1}^N y_k^Ty_k}{\frac{1}{N}\sum_{k=1}^N \omega_k^T\omega_k}} = \sqrt{\frac{\mathbf{E}[y_k^Ty_k]}{\mathbf{E}[\omega_k^T\omega_k]}} \leq \bar{\gamma}. \tag{19}$$

Lemma 2: Given the dynamics in (18), the OCC $\|H\|_2$ constraint in (19) is satisfied if the steady state covariance

$$\mathbf{P} = \begin{bmatrix} \mathbf{P}_x & \mathbf{P}_{x\hat{x}} \\ \mathbf{P}_{x\hat{x}}^T & \mathbf{P}_{\hat{x}} \end{bmatrix} = \lim_{k\to\infty}\mathbf{P}_k = \lim_{k\to\infty}\mathbf{E}[\zeta_k\zeta_k^T], \tag{20}$$

satisfies the Lyapunov equation

$$\mathbf{P} = \hat{A}\mathbf{P}\hat{A}^T + \hat{R}, \qquad \mathbf{P} \geq 0, \quad \hat{R} = \begin{bmatrix} R_1 & 0 \\ 0 & LR_2L^T \end{bmatrix} \tag{21}$$

and the following convex inequality holds,

$$\begin{aligned} \mathcal{C}_h = \mathbf{tr}\big(\hat{E}_x^TC^TC\hat{E}_x\mathbf{P}\big) + \mathbf{tr}(R_2) \\ - \bar{\gamma}^2\big(\mathbf{tr}(R_1) + \mathbf{tr}(R_2)\big) \leq 0, \end{aligned} \tag{22}$$

where $\hat{E}_x = [I_n,\, 0_{n\times n}]$.

Now in order to use this $\|H\|_2$ constraint in a convex optimization we need to linearize the Lyapunov equation constraint. We state this result as part of a complete convex optimization problem to design the gains K and L to achieve the optimal (smallest) $\|H\|_2$ gain.

Theorem 1: Given the dynamics in (18), the smallest output covariance constrained $||H||_2$ gain defined by (19) is

$$\gamma^* = \sqrt{\frac{\mathbf{tr}\big(C\mathbf{P}_x^*C^T\big) + \mathbf{tr}(R_2)}{\mathbf{tr}(R_2) + \mathbf{tr}(R_1)}}, \tag{23}$$

where $\mathbf{P}_x^*$ is the solution of

$$\begin{cases} \min\limits_{\mathbf{P}_x,\mathbf{Q}_1,X,Y,Z} \mathbf{tr}(C\mathbf{P}_xC^T) \\ \qquad \text{s.t.} \;\; \mathcal{C}_L \geq 0, \end{cases} \tag{24}$$

with

$$\mathcal{C}_L = \begin{bmatrix} \mathbf{Q}_1 & I_n & \mathbf{Q}_1 F + XC & Z & \mathbf{Q}_1 R_1 & XR_2 \\ * & \mathbf{P}_x & F & F\mathbf{P}_x + GY & R_1 & 0 \\ * & * & \mathbf{Q}_1 & I_n & 0 & 0 \\ * & * & * & \mathbf{P}_x & 0 & 0 \\ * & * & * & * & R_1 & 0 \\ * & * & * & * & * & R_2 \end{bmatrix}. \quad (25)$$

There exists at most $\binom{2n}{n}$ distinct real-valued, control gains $L = \mathbf{Q}_{12}^{-1} X$, $K = Y\mathbf{P}_{x\hat{x}}^{-T}$ satisfying $\gamma = \gamma^*$, where $\mathbf{P}_{x\hat{x}} = (I_n - \mathbf{P}_x \mathbf{Q}_1)\mathbf{Q}_{12}^{-T}$ and $\mathbf{Q}_{12}$ is the solution of following generalized algebraic Ricatti equation,

$$\mathbf{Q}_{12}\Gamma_1\mathbf{Q}_{12} + \mathbf{Q}_{12}\Gamma_2 + \Gamma_3\mathbf{Q}_{12} + \Gamma_4 = 0, \quad (26)$$

with known matrices

$$\begin{aligned} &\Gamma_1 = GY(I_n - \mathbf{Q}_1\mathbf{P}_x)^{-1}, \\ &\Gamma_2 = F, \\ &\Gamma_3 = (\mathbf{Q}_1 GY + XC\mathbf{P}_x + \mathbf{Q}_1 F\mathbf{P}_x - Z)(I_n - \mathbf{Q}_1\mathbf{P}_x)^{-1}, \\ &\Gamma_4 = -XC. \end{aligned}$$

Remark 2: Here in Theorem 1, we transformed the nonlinear Lyapunov equation (21) into the linear convex optimization (24) through algebraic techniques with replacement of this equality constraint with the very similar inequality $\mathbf{P} - \hat{A}\mathbf{P}\hat{A}^T - \hat{R} \geq 0$. This relaxation is justified because the objective function $\mathbf{tr}(C\mathbf{P}_x C^T)$ minimizes the decision variable $\mathbf{P}$ and drives the optimization to the bound of the inequality, which would yield equality - hence driving the relaxed form, to the equality (21).

C. Bounding ellipsoid LMI (designing K and L)

The goal of this paper is to construct an optimization to design K and L such that the impact of an attacker on the reachable states is minimized. However, when K and L are considered variables of the Lemma 1 optimization, (17), and therefore, (15) contains nonlinear terms. In the discussion that follows, we impose some structure on the solution so that we can linearize the overall design problem. Each choice will be motivated individually, but it is also the combined effect of the these structures taken together that yield the final *linear* matrix inequality.

Imposed Structure 1: We structure the inverse of the shape matrix of the stacked state ξ, $\mathcal{P}$, as

$$\mathcal{P} = \begin{bmatrix} \mathcal{P}_1 & \mathcal{P}_{12} & \\ \mathcal{P}_{12}^T & \mathcal{P}_2 & \\ & & \mathcal{P}_3 \end{bmatrix}, \quad (27)$$

which assumes the independence of the ellipsoidal bound on the estimation error e_k from the ellipsoidal bound on the combined state x_k and estimate $\hat{x}_k$ (inspired by a similar assumption made in [5]). This structure enables us to utilize the parallel dynamics with and without attack (see Remark 1) and linearize the original LMI with respect to K and L.

This structure also permits inverting each block separately, such that $\mathcal{P}_3^{-1} = \mathcal{Q}_e$ and

$$\begin{bmatrix} \mathcal{P}_1 & \mathcal{P}_{12} \\ \mathcal{P}_{12}^T & \mathcal{P}_2 \end{bmatrix}^{-1} = \begin{bmatrix} \mathcal{Q}_x & \mathcal{Q}_{x\hat{x}} \\ \mathcal{Q}_{x\hat{x}}^T & \mathcal{Q}_{\hat{x}} \end{bmatrix} \quad (28)$$

Consider the linearizing change of co-ordinates used in [5],

$$T_2 = \begin{bmatrix} T_3 & & \\ & T_3 & \\ & & I_n \end{bmatrix}, \qquad T_3 = \begin{bmatrix} \mathcal{Q}_x & I_n & 0 \\ \mathcal{Q}_{x\hat{x}}^T & 0 & 0 \\ 0 & 0 & I_n \end{bmatrix}. \quad (29)$$

Although (17) is not entirely linearized with this transformation, due to the presence of term Σ which depends on L, we will introduce an iterative approach later to avoid this nonlinearity. The LMI $\mathcal{H}$, (17), becomes

$$\mathcal{H}_L = T_2^T \mathcal{H} T_2 = \begin{bmatrix} a\mathcal{P}_L & A_L^T & 0 \\ A_L & \mathcal{P}_L & B_L \\ 0 & B_L^T & \frac{1-a}{2-a}W \end{bmatrix}, \quad (30)$$

where

$$\mathcal{P}_L = T_3^T \mathcal{P} T_3 = \begin{bmatrix} \mathcal{Q}_x & I_n & 0 \\ I_n & \mathcal{P}_1 & 0 \\ 0 & 0 & \mathcal{P}_3 \end{bmatrix}, \quad (31)$$

$$\begin{aligned} &B_L = T_3^T \mathcal{P} B = \begin{bmatrix} I_n & 0 \\ \mathcal{P}_1 & Y_1 \\ \mathcal{P}_3 & -\mathcal{P}_3 L \end{bmatrix}, \\ &A_L = T_3^T \mathcal{P} A T_3 = \begin{bmatrix} F\mathcal{Q}_x + GX_1 & F & 0 \\ Z_1 & \mathcal{P}_1 F + Y_1 C & -Y_1 C \\ 0 & 0 & \mathcal{P}_3 F \end{bmatrix}, \\ &Y_1 = \mathcal{P}_{12} L, \quad X_1 = K\mathcal{Q}_{x\hat{x}}^T, \\ &Z_1 = \mathcal{P}_1 F\mathcal{Q}_x + \mathcal{P}_{12} LC\mathcal{Q}_x + \mathcal{P}_1 GK\mathcal{Q}_{x\hat{x}}^T + \mathcal{P}_{12} F\mathcal{Q}_{x\hat{x}}^T \\ &\quad + \mathcal{P}_{12} GK\mathcal{Q}_{x\hat{x}}^T - \mathcal{P}_{12} LC\mathcal{Q}_{x\hat{x}}^T. \end{aligned}$$

One of the useful features of this transformation is that $\mathcal{Q}_x = E_x^T \mathcal{Q} E_x$, the quantify used in the objective function of Lemma 1, appears as a variable of the LMI. This section provides the linearization necessary to separate the gains K and L as variables in Lemma 1 (and could then be used as the starting point if a different performance criteria was used, as opposed to the $\|H\|_2$ constraint considered in this paper).

D. Combining Both Constraints into the Design

In this work, we design the controller and estimator gains to minimize the impact of attacks on the system state, which is measured by an outer ellipsoidal bound on the reachable states when the system is driven by the attack and system noise. There are an infinite number of potential outer bounding - and tight - ellipsoids. To combine the reachable set and $\|H\|_2$ LMI constraints, we make a specific choice about the outer ellipsoidal bound we select.

Imposed Structure 2: We select the shape matrix of the ellipsoidal bound of the states x_k and estimate $\hat{x}_k$ under attack $E_{x\hat{x}}\mathcal{Q}E_{x\hat{x}}^T$ - see (28) - to have the same orientation as the covariance of the states and estimate without attack (ζ_k),

$$\sigma\mathbf{P} = E_{x\hat{x}}\mathcal{Q}E_{x\hat{x}}^T, \quad (32)$$

where σ is a scaling factor and becomes a new variable of the method. Since $\mathbf{Q} = \mathbf{P}^{-1}$, this sets up a common set of variables to link the $\|H\|_2$ (left) and ellipsoidal bound (right)

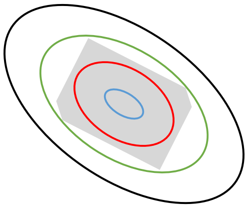

Fig. 1: The reachable set (gray) is approximated by ellipsoids with shape matrix $\sigma\mathbf{P}$. The role of σ is to identify the tight approximation (green), where $\sigma = \sigma^*$. When $\sigma < \sigma^*$ the optimization will be infeasible because the red ellipsoid cannot contain the reachable set. When $\sigma > \sigma^*$ the black ellipsoid loosely contains the reachable set.

constraints,

$$\begin{aligned}\sigma \begin{bmatrix}\mathbf{P}_x & \mathbf{P}_{x\hat{x}}\\ \mathbf{P}_{x\hat{x}}^T & \mathbf{P}_{\hat{x}}\end{bmatrix} &= \begin{bmatrix}\mathcal{Q}_x & \mathcal{Q}_{x\hat{x}}\\ \mathcal{Q}_{x\hat{x}}^T & \mathcal{Q}_{\hat{x}}\end{bmatrix},\\ \underbrace{\frac{1}{\sigma}\begin{bmatrix}\mathbf{Q}_1 & \mathbf{Q}_{12}\\ \mathbf{Q}_{12}^T & \mathbf{Q}_2\end{bmatrix}}_{\|H\|_2} &= \underbrace{\begin{bmatrix}\mathcal{P}_1 & \mathcal{P}_{12}\\ \mathcal{P}_{12}^T & \mathcal{P}_2\end{bmatrix}}_{\mathcal{E}(\mathcal{Q})}.\end{aligned} \tag{33}$$

The structure above allows us to replace variables in the ellipsoidal bound optimization $\mathcal{Q}$ and $\mathcal{P}$ with quantities from the performance criteria, $\mathbf{P}$ and $\mathbf{Q}$, respectively.

Remark 3: For small values of σ the optimization will be infeasible and for extremely large values of σ the ellipsoid trivially bounds the reachable set. The optimal value σ^*, where the ellipsoidal bound is tangent to the reachable set, will be calculated in the algorithm that iterates the optimization. If $\sigma > \sigma^*$ the optimization is feasible and if $\sigma < \sigma^*$ the problem is infeasible (see Fig. 1).

Remark 4: Due to our techniques to linearize the problem, the value of σ^* does not in practice achieve tangency to the reachable set characterized by optimized gain matrices in this paper. Nonetheless, the notion that a σ^* exists for feasibility of a solution is still valid, even if it does incorporate an extra amount of conservatism.

Based on (33) we can link the variables of bounding ellipsoid LMI with $\|H\|_2$ constraint,

$$\begin{aligned}&X = \sigma Y_1 = \mathbf{Q}_{12}L, \quad Y = \frac{X_1}{\sigma} = K\mathbf{P}_{x\hat{x}}^T,\\ &Z = Z_1 = \mathbf{Q}_1 F\mathbf{P}_x + XC\mathbf{P}_x + \mathbf{Q}_1 GY + \mathbf{Q}_{12}F\mathbf{P}_{x\hat{x}}^T\\ &\qquad + \mathbf{Q}_{12}GY - XC\mathbf{P}_{x\hat{x}}^T.\end{aligned} \tag{34}$$

Now we can re-write A_L, B_L , $\mathcal{P}_L$ based on $\mathbf{P}_x$, $\mathbf{Q}_1$, $\mathcal{P}_3$, X, Y, Z,

$$A_L = \begin{bmatrix}\sigma(F\mathbf{P}_x + GY) & F & 0\\ Z & \frac{1}{\sigma}(\mathbf{Q}_1 F + XC) & -\frac{1}{\sigma}XC\\ 0 & 0 & \mathcal{P}_3 F\end{bmatrix},$$

$$B_L = \begin{bmatrix}I_n & 0\\ \frac{1}{\sigma}\mathbf{Q}_1 & \frac{1}{\sigma}X\\ \mathcal{P}_3 & -\mathcal{P}_3 L\end{bmatrix}, \quad \mathcal{P}_L = \begin{bmatrix}\sigma\mathbf{P}_x & I_n & 0\\ I_n & \frac{1}{\sigma}\mathbf{Q}_1 & 0\\ 0 & 0 & \mathcal{P}_3\end{bmatrix}.$$

Thus the choice in (32) has facilitated integrating these optimizations.

```
Algorithm III.1: THEOREM 2($F, G, C, R_1, R_2$)
  $L \leftarrow$ Theorem 1
  $\sigma \leftarrow \infty$
  while true do
    if (35) is infeasible then break
    $\mathbf{P}_x, \mathbf{Q}_1, X, Y, Z \leftarrow$ (35)
    $\mathcal{K}, \mathcal{L} \leftarrow$ real solutions of Ricatti eqn (26)
    $(\tilde{K}, \tilde{L}) \leftarrow$ select pair $(K, L) \in (\mathcal{K}, \mathcal{L})$ by
        smallest Lemma 1 objective value
    if $\|L - \tilde{L}\| \leq \epsilon$ then $\sigma \leftarrow \sigma - \varepsilon$
    else $K, L \leftarrow \tilde{K}, \tilde{L}$
  return $(K, L)$
```

Theorem 2: Consider a LTI system (1) with desired output covariance constrained $\|H\|_2$ gain $\bar{\gamma}$ (19), chi-squared detector threshold α (11) and zero-alarm stealthy attacker (11). Algorithm III.1 returns (approximately) optimal controller K^* and observer L^* gains to minimize the reachable set of states possible by the attacker, while maintaining an OCC $\|H\|_2$ gain no bigger than $\bar{\gamma}$. Algorithm III.1 uses the Ricatti equation (26) to update L based on the solution of the combined convex optimization problem which has a solution if for some $a \in [0, 1)$,

$$\begin{cases}\displaystyle\min_{\substack{a_1, a_2, \mathbf{P}_x, \mathbf{Q}_1,\\ X, Y, Z, \mathcal{P}_3}} \ \mathbf{tr}(\mathbf{P}_x)\\ \qquad \text{s.t.} \quad 0 \leq a_1, a_2 < 1, \quad a_1 + a_2 \geq a,\\ \qquad\qquad \mathcal{H}_L \geq 0,\\ \qquad\qquad \mathcal{C}_h \leq 0, \ \mathcal{C}_L \geq 0.\end{cases} \tag{35}$$

Proof: In the past sections we have linearized the LMIs associated with the ellipsoidal outer bound on the reachable set ($\mathcal{H}_L$) and with the $\|H\|_2$ constraint ($\mathcal{C}_h$ and $\mathcal{C}_L$) and finally made a structural connection between these two optimizations (33) to use a common set of decision variables. Because the controller gain K appears within the decision variables of the optimization, we implicitly optimize K.

There are two remaining challenges to be addressed in this proof. First, L appears, as K does, implicitly in the decision variables, but also explicitly in the nonlinear term $\mathcal{P}_3 L$ in matrix B_L and in the dependence of covariance Σ in matrix W on L, both of which are in $\mathcal{H}_L$. Second, the magnification factor σ multiplies most of the decision variables of the optimization. Thus neither L nor σ can be taken as variable in the optimization. We solve this by applying an iterative algorithm over both L and σ that leverages the structure presented in Figure 1. If we select a large enough value for the magnification parameter σ, any choice of L easily satisfies $\mathcal{H}_L$ by creating a very large ellipsoidal outer bound. We satisfy the other constraints of the optimization (the $\|H\|_2$ constraints) by selecting the initial value for L as the $\|H\|_2$ optimal L (from Theorem 1), hence satisfying $\mathcal{C}_h$ and $\mathcal{C}_L$. For these fixed values of σ, L, and Σ, the optimization

(35) is solved. The solution then suggests a new value of L - solved from the Ricatti equation in (26) - that minimizes the ellipsoidal bound while satisfying all constraints. Using the same value for σ but the updated L (and hence updated Σ), optimization (35) is again solved, yielding another value for L. This iteration is repeated until L has sufficiently converged. Once convergence is achieved, σ is reduced and the process is repeated with the existing L as the initial value for the L iteration. At some point, the magnification factor σ will be too small for any choice of L to permit the ellipsoidal bound to contain the reachable set, hence the optimization (35) will become infeasible. This is the stopping condition for the algorithm.

■

IV. Case Study

We now demonstrate these tools and consider an LTI system (with matrices given below) for this study with the chi-squared detector tuned to a threshold $\alpha = 3.841$, and system noise truncated $\Pr[\nu_k^T R_1^{-1} \nu_k \leq \bar{\nu}] = p_\nu = 95\%$. In this work we solve the semi-definite programming problems with the software YALMIP, with solver SeDuMi [14].

$$F = \begin{bmatrix} 0.84 & 0.23 \\ -0.47 & 0.12 \end{bmatrix}, \; G = \begin{bmatrix} 0.07 \\ 0.23 \end{bmatrix}, \; C = \begin{bmatrix} 1 & 0 \end{bmatrix},$$
$$R_1 = \begin{bmatrix} 0.45 & -0.11 \\ -0.11 & 0.20 \end{bmatrix}, \; R_2 = 1, \; \bar{\gamma} = 1.08. \tag{36}$$

From Theorem 1, we calculate the minimum $\|H\|_2$ gain $\gamma^* = 1.06$, which is achieved by the $\binom{4}{2}$ solutions of the Ricatti equation (26). We pick the (real) solution that yields the minimum trace of the ellipsoidal bound (using Lemma 1). The gains that achieve the optimum $\|H\|_2$ gain γ^* are $L = [0.31, -0.21]^T$ and $K = [-12, -3.29]$. The empirical reachable set corresponding to the $\|H\|_2$ optimal gains is plotted in Fig. 2 in gray with the green ellipsoidal bound.

We now run the algorithm in Theorem 2 with initial magnification factor $\sigma = 160$, convergence threshold $\epsilon = 0.02$, and decrement $\varepsilon = 1$. The resulting gain matrices that minimize the reachable set while preserving an $\|H\|_2$ gain $\gamma^* \leq \gamma \leq \bar{\gamma} = 1.08$, are $L = [0.5, 0.1]^T$ and $K = [-1.58, -1.96]$ for final magnification value of $\sigma = 146$. In Fig. 2, the red ellipsoid corresponds to the shape matrix found by Theorem 2, inclusive of all linearization steps and imposed structures. In blue, we show the ellipsoidal bound corresponding to the same optimal K and L, but using Lemma 1, which does not contain the linearizations and imposed structures. In black we have empirically computed the exact reachable set, which evidences a significant reduction in the reachable set over the $\|H\|_2$ optimal gains.

V. Conclusion

This paper presents a set of tools to design the feedback controller and observer gains for observer-based feedback control to minimize the effect of an attacker that intelligently falsifies sensor measurements. This approach allows us to make a trade-off between ensuring performance and minimizing the effect of the attacker.

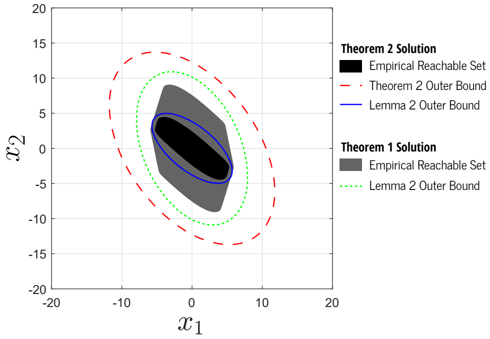

Fig. 2: The comparison of empirical reachable sets demonstrates the effectiveness of our tools (Theorem 2, black) to design gains K and L to minimize the reachable set due to the attacker compared to Theorem 1 (gray, $\|H\|_2$ optimal). The optimal gains are designed in Theorem 2 using the (red) ellipsoid with shape matrix $\sigma \mathbf{P}_x$. Lemma 1 provides a better (blue) ellipsoid given the same optimal gains. Lemma 1 with the $\|H\|_2$ optimal gains provides the (green) ellipsoid.

References

[1] C. Murguia and J. Ruths, "On model-based detectors for linear time-invariant stochastic systems under sensor attacks," *IET Control Theory & Applications*, vol. 13, no. 8, pp. 1051–1061, 2019.

[2] D. Umsonst and H. Sandberg, "Anomaly detector metrics for sensor data attacks in control systems," in *2018 Annual American Control Conference (ACC)*. IEEE, 2018, pp. 153–158.

[3] Y. Mo and B. Sinopoli, "On the performance degradation of cyber-physical systems under stealthy integrity attacks," *IEEE Transactions on Automatic Control*, vol. 61, pp. 2618–2624, 2016.

[4] N. Hashemi, C. Murguia, and J. Ruths, "A comparison of stealthy sensor attacks on control systems," in *2018 Annual American Control Conference (ACC)*. IEEE, 2018, pp. 973–979.

[5] C. Murguia, I. Shames, J. Ruths, and D. Nešić, "Security metrics and synthesis of secure control systems," *Automatica*, vol. 115, p. 108757, 2020.

[6] Z. Guo, D. Shi, K. Johansson, and L. Shi, "Worst-case stealthy innovation-based linear attack on remote state estimation," *Automatica*, vol. 89, pp. 117–124, 03 2018.

[7] C. Murguia, N. van de Wouw, and J. Ruths, "Reachable sets of hidden cps sensor attacks: Analysis and synthesis tools," in *proceedings of the IFAC World Congress*, 2016.

[8] G. Zhu, M. Rotea, and R. Skelton, "A convergent algorithm for the output covariance constraint control problem," *SIAM Journal on Control and Optimization*, vol. 35, no. 1, pp. 341–361, 1997.

[9] F. Xu, K. H. Lee, and B. Huang, "Monitoring control performance via structured closed-loop response subject to output variance/covariance upper bound," *Journal of Process Control*, vol. 16, no. 9, pp. 971–984, 2006.

[10] C. Murguia and J. Ruths, "Characterization of a cusum model-based sensor attack detector," in *proceedings of the 55th IEEE Conference on Decision and Control (CDC)*, 2016.

[11] ——, "Cusum and chi-squared attack detection of compromised sensors," in *proceedings of the IEEE Multi-Conference on Systems and Control (MSC)*, 2016.

[12] R. Tunga, C. Murguia, and J. Ruths, "Tuning windowed chi-squared detectors for sensor attacks," in *2018 Annual American Control Conference (ACC)*. IEEE, 2018, pp. 1752–1757.

[13] N. Hashemi and J. Ruths, "Co-design for performance and security: Lmi tools," 2020. [Online]. Available: arXiv:1909.12452

[14] J. Lofberg, "Yalmip : a toolbox for modeling and optimization in matlab," in *Computer Aided Control Systems Design, 2004 IEEE International Symposium on*, 2004, pp. 284–289.