Gain Design via LMIs to Minimize the Impact of Stealthy Attacks

Navid Hashemi and Justin Ruths

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. It is, again, not surprising then that introducing a security metric to be optimized pulls the optimal gains in a new direction. 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.

Before work on security, Control Theory’s adversaries included model uncertainties, nonlinearities, disturbances, and other extremely problematic, but non-intelligent sources. The presence of an intelligent and exploitative attacker gives rise to all-new questions for Control Theory. In response, the work has branched into reviewing all aspects of fundamental theory with the new lens of security. The branch that is most relevant to the work in this paper leverages 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 (and tuning of 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. While a dynamic output feedback controller is the more general case, 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. 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
\[
\begin{align*}
    x_{k+1} &= Fx_k + Gu_k + v_k, \\ 
    y_k &= Cx_k + \eta_k.
\end{align*}
\] (1, 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 Gaussian system noise \( v_k \sim N(0, R_1) \), \( R_1 \in \mathbb{V}^{n \times n} \) (\( \mathbb{V} \) 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 Gaussian measurement noise \( \eta_k \sim N(0, R_2) \), \( R_2 \in \mathbb{V}^{p \times p} \). In addition we assume that the pair \((F, C)\) is detectable and \((F, G)\) is stabilizable.

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,
\[
y_k = y_k + \delta_k = Cx_k + \eta_k + \delta_k.
\] (3)
If the attacker has access to the measurements, then it is possible for the attack \( \delta_k \) 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
\[
\hat{x}_{k+1} = F\hat{x}_k + Gu_k + L(y_k - C\hat{x}_k),
\] (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 out-of-band feedback controllers
\[
u_k = K\hat{x}_k,
\] (5)
where \( K \in \mathbb{R}^{m \times n} \) is the controller gain matrix. Next, we define the residual sequence
\[
r_k = y_k - C\hat{x}_k,
\] (6)
as the difference between what we actually receive (\( y_k \)) and expect to receive (\( C\hat{x}_k \)), which evolves according to
\[
\begin{align*}
    x_{k+1} &= (F + GK)\hat{x}_k - GKe_k + v_k, \\
    e_{k+1} &= (F - LC)e_k - L\eta_k + v_k - L\delta_k, \\
    r_k &= Ce_k + \eta_k + \delta_k,
\end{align*}
\] (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,
\[
\Sigma = \mathbf{E}[r_k r_k^T] = C\mathbf{E}[e_k e_k^T]C^T + \mathbf{E}[\eta_k \eta_k^T],
\] (8)
where the steady state covariance of the estimation error
\[
\mathbf{P}_e = \lim_{k \to \infty} \mathbf{P}_k = \mathbf{E}[e_k e_k^T]
\] is
\[
\mathbf{P}_e = (F - LC)\mathbf{P}_e(F - LC)^T + LR_2L^T + R_1.
\] (9)

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,
\[
z_k = r_k^T \Sigma^{-1} r_k.
\] (10)
The chi-squared detector generates alarms when the distance measure exceeds a threshold \( \alpha \in \mathbb{R}_{>0} \)
\[
\begin{align*}
    z_k &\leq \alpha \quad \rightarrow \quad \text{no alarm}, \\
    z_k &> \alpha \quad \rightarrow \quad \text{alarm: } k' = k,
\end{align*}
\] (11)
such that alarm time(s) \( k' \) are produced. The \( \Sigma^{-1} \) factor in the definition of \( z_k \) re-scales the distribution \( \mathbf{E}[z_k] = p, \mathbf{E}[z_k z_k^T] = 2p \) so that the threshold \( \alpha \) can be designed independent of the specific statistics (mean and covariance) of the noises \( v_k \) and \( \eta_k \); instead, it can be selected simply based on the number of sensors, \( p \).

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 generate attack sequences that maintain the distance measure at or below the threshold of detection, i.e., \( z_k \leq \alpha \). Hence, these attacks generate no alarms. To satisfy this condition we define the attack as
\[
\delta_k = \phi_k - (y_k - C\hat{x}_k) = -C e_k - \eta_k + \phi_k,
\] (12)
where \( \phi_k \in \mathbb{R}^p \) is any vector such that \( \phi_k^T \Sigma^{-1} \phi_k \leq \alpha \) (recall the attacker has access to the sensor, \( y_k \), and knowledge of the estimator, \( \hat{x}_k \)). Based on this attack strategy,
\[
z_k = r_k^T \Sigma^{-1} r_k,
\] (13)
Thus \( z_k \leq \alpha \) and no alarms are raised.

B. Reachable Set

Under a stealthy zero-alarm attack \([12]\), the attacked system dynamics become
\[
\begin{align*}
    x_{k+1} &= Fx_k + GK\hat{x}_k + v_k, \\
    \hat{x}_{k+1} &= LCx_k + (F + GK - LC)\hat{x}_k - LCe_k + L\phi_k, \\
    e_{k+1} &= F e_k - L\phi_k + v_k.
\end{align*}
\]
We stack these into a combined state \( \xi_k = [x_k^T, \hat{x}_k^T, e_k^T]^T \) and combined input \( \mu_k = [v_k^T, \phi_k^T]^T \),
\[
\xi_{k+1} = A\xi_k + B\mu_k,
\] (14)
with
\[
A = \begin{bmatrix} F & GK \\ LC & F + GK - LC & -LC \end{bmatrix}, \quad B = \begin{bmatrix} I \\ 0 \\ I \end{bmatrix}.
\] (15)

**Remark 1:** The choice of including \(x_k, \tilde{x}_k\), and \(e_k\) seems redundant at this point since \(e_k = x_k - \tilde{x}_k\); however, this choice is crucial as we layer additional constraints into the design optimization.

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

The reachable set of states is then,
\[
\mathcal{R} = \left\{ x_k = E_x \xi_k \left| \begin{array}{l}
\xi_{k+1} = A \xi_k + B \mu_k, \\
\xi_1 = 0, \\
\phi_k^T \Sigma^{-1} \phi_k \leq \alpha, \\
\nu_k^T R_1^{-1} \nu_k \leq \bar{\nu}, \forall k \in \mathbb{N}
\end{array} \right. \right\},
\] (16)

where the ellipsoidal bound on the attack \(\phi_k\) is imposed by the attacker’s desire to remain stealthy [13], and the ellipsoidal bound on the noise is created by truncating the Gaussian system noise to a desired probability, i.e., \(\Pr[\nu_k^T R_1^{-1} \nu_k \leq \bar{\nu}] = p_\nu\), where \(p_\nu\) is some desired (typically high) probability. In principle, the noise has unbounded support, and hence the reachable set is unbounded. To ensure bounded reachable sets, we apply this truncation at the desired confidence level.

**III. LMI APPROACH TO DESIGN K AND L**

In the first section, we reframed an existing result more concisely, which identifies a minimal outer ellipsoidal bound on the set of states reachable by a stealthy (zero-alarm) attacker. When we then move to consider minimizing this set further through the design of the feedback and estimator gains \(K\) and \(L\). As has been discussed in previous studies, a trivial solution exists to this design problem - to make either \(K = 0_{n \times n}\), or \(L = 0_{n \times p}\). Doing so cuts the feedback loop and guarantees that corrupted measurements do not impact the system state. Simultaneously, this destroys the purpose - more specifically the performance - of the feedback loop. While many performance metrics could be used, in Section III-B we impose a \(\|H\|_2\) constraint to avoid these trivial solutions. Unlike prior work where the performance criteria was distributionally robust (agnostic), this \(\|H\|_2\) constraint is specific to the covariance of the noise, thereby allowing our design optimization to leverage this important knowledge. This OCC \(\|H\|_2\) constraint is non-convex; to our knowledge, this paper offers the first convexification of the OCC \(\|H\|_2\) criteria into an LMI framework.

**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 2 we solve these simultaneously through the stacked states \(\xi_k\) and inputs \(\mu_k\). The following lemma provides a bound on a Lyapunov-inspired function given an elliptically bounded input.

**Lemma 1:** [5] Let \(V_k\) be a positive definite function with \(V_1 = 0\) and \(\mu_k^T W_i \mu_k \leq 1, i = 1, \ldots, N\), where \(W_i\) is positive definite. If there exists a constant \(a \in (0, 1)\) and \(a_i \in (0, 1)\) such that \(\sum_{i=1}^N a_i \geq a\) and
\[
V_{k+1} - a V_k - \sum_{i=1}^N (1 - a_i) \mu_k^T W_i \mu_k \leq 0,
\] (17)
then \(V_k \leq N \frac{a}{1-a}\).

When we select the positive definite \(V_k\) to be a quadratic function of the state, the result above provides an outer ellipsoidal bound on the reachable states. We will use the notation \(E(Q) = \{x \mid x^T Q^{-1} x \leq 1\}\).

**Lemma 2:** Given the stacked system matrices \(A\) and \(B\) in [15], gain matrices \(K, L\), detector threshold \(\alpha\) with steady state residual covariance \(\Sigma\), system noise truncation threshold \(\bar{\nu}\) with covariance \(R_3\), if there exists constants \(a, a_1, a_2 \in [0, 1]\), the solution of
\[
\begin{aligned}
\min_{a_1, a_2, Q} & \mathbf{tr}(E_x^T Q E_x) \\
\text{s.t.} & \quad 0 \leq a_1, a_2 < 1, \quad a_1 + a_2 \geq a,
\end{aligned}
\] (18)

provides the shape matrix \(Q\) of the ellipsoidal bound on the reachable set of states, i.e., \(\mathcal{R} \subseteq E(E_x^T Q E_x)\), where
\[
W_k = \begin{bmatrix}
\frac{1-a_1}{\alpha^2} R_1^{-1} & 0 \\
0 & \frac{1-a_2}{\alpha} \Sigma^{-1}
\end{bmatrix}.
\] (19)

**Proof:** The stacked dynamics [14] is driven by two inputs which are both ellipsoidally bounded. Letting \(W_1 = R_1^{-1}\) and \(W_2 = \Sigma^{-1}\), (17) becomes
\[
V_{k+1} - a V_k - \frac{1 - a_2}{\alpha} \phi_k^T \Sigma^{-1} \phi_k - \frac{1 - a_1}{\bar{\nu}^2} \nu_k^T R_1^{-1} \nu_k \leq 0.
\]

Substituting the choice \(V_k = E_k^T Q E_x\) [4] to \(E_k^T Q E_x \leq \frac{2-a}{\alpha} Q\), \(P > 0\), into this equation and expanding using the dynamics [14] results in the LMI,
\[
\mathcal{H} = \begin{bmatrix}
P A & P A^T & 0 \\
P B & P B^T & 0 \\
0 & 0 & \frac{1-a}{\alpha^2} W
\end{bmatrix} \geq 0,
\] (20)
where \(P\) is the inverse of the shape matrix of the ellipsoidal bound for the \(\xi\) reachable set \((P^{-1} = Q)\), such that the first block \(E_x^T Q E_x\) 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 \(E_x^T Q E_x\). To use \(Q\) as the variable of the optimization instead of \(P\) we apply the
transformation $T = \text{diag}[Q, Q, I_s]$, to (20), i.e., $T^T H T$, which results in the LMI in (18).

**Remark 2:** For any convex shape (i.e., reachable set) there are an infinite number of tight outer ellipsoidal bounds. These different ellipsoids can be visualized as being tangent to the reachable set at different points. The minimum trace objective minimizes the sum of the squared principal axes, which tends to avoid solutions with, for example, low volume but one large principal axis.

**Remark 3:** Note that the parameter $a$ is not a decision variable of the optimization in (18). It appears nonlinearly (multiplying $Q$). Since $a$ belongs to a compact interval, the conventional choice is to solve (18) across a grid search in $a$ and select the minimal, feasible solution.

### 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 linearize 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]. To specify the performance robust control, in general, studies the gain observed in the signal $h_k = H_1 x_k + H_2 y_k + H_3 y_k$. 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 $h_k = y_k$ and excitation $\omega_k = [\nu^T_k, \eta_k^T]^T$, making $H_1 = C$, $H_2 = I_{p \times p}$, and $H_3 = 0_{p \times n}$.

When there is no attack the system evolves according to

$$x_{k+1} = F x_k + G K \dot{x}_k + \nu_k,$$
$$\dot{x}_{k+1} = L C x_k + (F + G K - L C) \dot{x}_k - L \eta_k,$$
$$y_k = C x_k + \eta_k,$$

which can be combined using the stacked state $\xi_k = [x^T_{k+1}, x^T_k]^T$.

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

with $E_{\hat{x}x} = \text{diag} [I_n, I_n, 0_{n \times n}]$ making $\hat{A} = E_{\hat{x}x} A E_{\hat{x}x}$.

**Remark 4:** It is here that we can start to appreciate the value of the seemingly redundant definition of $\xi_k$ (see Remark 1). By doing so, the state matrix without attack $\hat{A}$ can 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 $\bar{\gamma}$,

$$\text{OCC } \|H\|_2 : \sqrt{\frac{1}{N} \sum_{k=1}^{N} y_k^T y_k} \leq \sqrt{\frac{1}{N} \sum_{k=1}^{N} \omega_k^T \omega_k} \leq \bar{\gamma}. \quad (25)$$

**Lemma 3:** Given the dynamics in (24), the OCC $\|H\|_2$ constraint in (25) is satisfied if the steady state covariance

$$P = \begin{bmatrix} P_x & P_{x\hat{x}} \\ P_{\hat{x}x} & P_{\hat{x}\hat{x}} \end{bmatrix} = \lim_{k \to \infty} P_k = \lim_{k \to \infty} E[\xi_k \xi_k^T], \quad (26)$$

satisfies the Lyapunov equation

$$\hat{P} = \hat{A} \hat{P} \hat{A}^T + \hat{R}, \quad \hat{P} \geq 0, \quad (27)$$

and the following convex inequality holds,

$$C_h = \text{tr}(\hat{E}_x^T C_h C \hat{E}_x P) + \text{tr}(R_2) - \overline{\gamma}^2 (\text{tr}(R_1) + \text{tr}(R_2)) \leq 0, \quad (28)$$

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

**Proof:** From (23) and the definition of $\omega_k$, we can calculate the quadratic terms in (25),

$$y_k^T y_k = x_k^T C_k x_k + 2 x_k^T C_k \eta_k + \eta_k^T \eta_k,$$
$$\omega_k \bar{\omega}_k = \nu_k^T \nu_k + \eta_k^T \eta_k. \quad (29)$$

Taking the expectation (recall $x_k$ and $\eta_k$ are independent),

$$E[y_k^T y_k] = E[x_k^T C_k x_k] + E[\eta_k^T \eta_k] = \text{tr}(C_k C \ E[x_k x_k^T]) + \text{tr}(R_2), \quad (31)$$

and similarly,

$$E[\omega_k \bar{\omega}_k] = E[\nu_k^T \nu_k] + E[\eta_k^T \eta_k] = \text{tr}(R_1) + \text{tr}(R_2). \quad (32)$$

The unknown quantity is then the covariance of the state $E[x_k x_k^T]$, which is the first block of the stacked state $\xi_k$ covariance $P_k = E[\xi_k \xi_k^T]$. This covariance follows the update, evaluating $E[\xi_k \xi_k^T]$ with (24),

$$P_{k+1} = \hat{A} P_k \hat{A}^T + \hat{R}, \quad \hat{R} = \begin{bmatrix} R_1 & 0 \\ 0 & LR_2 L^T \end{bmatrix}. \quad (33)$$

Because the matrix $\hat{A}$ is stable the covariance converges to an steady value $\lim_{k \to \infty} P_k = P$ which must satisfy the Lyapunov equation (27). Combining this $P$ with (31)-(32) the $\|H\|_2$ constraint becomes

$$\sqrt{\frac{\text{tr}(C P_x C^T) + \text{tr}(R_2)}{\text{tr}(R_2) + \text{tr}(R_1)}} \leq \bar{\gamma}. \quad (34)$$

Rearranging this leads to the condition (28). \hfill \Box

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 program to design the gains $K$ and $L$ to achieve the optimal (smallest) $\|H\|_2$ gain.

**Theorem 1:** Given the dynamics in (24), the smallest output covariance constrained $\|H\|_2$ gain defined by (25) is

$$\gamma^* = \sqrt{\frac{\text{tr}(C P^*_x C^T) + \text{tr}(R_2)}{\text{tr}(R_2) + \text{tr}(R_1)}}, \quad (35)$$

where $P^*_x$ is the solution of

$$\begin{cases}
    P_x, Q_1, X, Y, Z, \\
    \text{s.t. } C_L \geq 0,
\end{cases} \quad (36)$$
with
\[
C_L = \begin{bmatrix}
Q_1 & I & Q_1 F + XC & Z & Q_1 R_1 & X R_2 \\
* & P_x & F & F P_x + G Y & 0 & 0 \\
* & * & Q_1 & I & 0 & 0 \\
* & * & * & P_x & 0 & 0 \\
* & * & * & * & R_1 & 0 \\
* & * & * & * & R_2 & 0
\end{bmatrix}.
\] (37)

There exists at most \(2^n\) distinct real-valued, control gains \(L = Q_{12} X, \ K = Y P_{x x}^{-T}\) satisfying \(\gamma = \gamma^*\), where \(P_{x x} = (I - P_x Q_1) Q_{12}^{-T}\) and \(Q_{12}\) is the solution of following generalized algebraic Ricatti equation,
\[
Q_{12} \Gamma_1 + Q_{12} \Gamma_2 + \Gamma_3 Q_{12} + \Gamma_4 = 0,
\] (38)
with known matrices
\[
\Gamma_1 = G Y (I - Q_1 P_x)^{-1},
\]
\[
\Gamma_2 = F,
\]
\[
\Gamma_3 = (Q_1 G Y + X C P_x + Q_1 F P_x - Z)(I - Q_1 P_x)^{-1},
\]
\[
\Gamma_4 = -X C.
\]

Proof: The formula for the optimal gain \(\gamma^*\), comes naturally from the \(\|H\|_2\) bound derived in (34). Since all other terms are constant, minimizing \(\text{tr}(C P_x C^T)\) is equivalent to minimizing the gain. This covariance is constrained by the Lyapunov equation in (27). Here, which is a standard technique for incorporating Lyapunov equations into convex optimizations, we replace this equality constraint with the very similar inequality,
\[
P - \dot{A} P \dot{A}^T - \dot{R} \geq 0, \quad P \succeq 0.
\] (39)

We can now combine these two inequality constraints into one using the Schur complement [13],
\[
C = \begin{bmatrix}
P & \dot{R} \\
P \dot{A}^T & P
\end{bmatrix} \succeq 0.
\] (40)

This relaxation is justified because the objective function \(\text{tr}(C P_x C^T)\) minimizes the decision variable \(P\) and drives the optimization to the bound of the inequality, which would yield equality - hence driving the relaxed form (39) to the equality (27).

We use the following transformation to linearize \(C\),
\[
C_L = \begin{bmatrix}
T_1 & T_1^T C & T_1 & T_1
\end{bmatrix} \begin{bmatrix}
P - R_L \\
\dot{A}_L \\
\dot{P}_L
\end{bmatrix},
\] (41)
with
\[
T_1 = \begin{bmatrix}
Q_1 & I \\
Q_{12} & 0
\end{bmatrix}, \quad P^{-1} = Q = \begin{bmatrix}
Q_1 & Q_{12} \\
Q_{12} & Q_2
\end{bmatrix},
\] (42)
\[
P_L = T_1^T P T_1 = \begin{bmatrix}
Q_1 & I \\
I & P_x
\end{bmatrix},
\] (43)
\[
R_L = T_1^T R T_1 = \begin{bmatrix}
Q_1 R_1 Q_1 + Q_{12} L R_2 L^T Q_{12}^T & Q_1 R_1 \\
R_1 Q_1 & R_1
\end{bmatrix},
\]
\[
\dot{A}_L = T_1^T \dot{A} P T_1 = \begin{bmatrix}
Q_1 F + XC \\
F P_x + G Y
\end{bmatrix},
\]
where \(X, Y, \) and \(Z\) are defined as
\[
X = Q_{12} L, \quad Y = K P_{x x}^T, \quad Z = Q_1 F P_x + X C P_x + Q_1 G Y + Q_{12} F P_{x x}^T + Q_{12} G Y - X C P_{x x}^T.
\] (44)

The term \(P_L - R_L\) can be linearized by applying a Schur complement to recover \(C_L\) in (37). This transformation changes the set of decision variables from \((P_x, P_{x x}, L, K)\) to \((P_x, Q_{12}, X, Z, Y)\). The solution in these new decision variables is then used to calculate \(P_{x x}\) and \(Q_{12}\) using (46) and the identify
\[
P_x Q_1 + P_{x x} Q_{12}^T = I,
\] (47)
which comes from the first block of the definition \(P Q = I\). The definition of \(Z\) in (46) and (47) combine to form the general algebraic Ricatti equation (38), which, in general, has \(2^n\) different answers. Finally, the gain matrices can be found by, \(L = Q_{12}^{-1} X\) and \(K = Y P_{x x}^{-T}\).

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 2 optimization, (20), and therefore, (18) contains nonlinear terms. In the sections that follow, 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 \(\xi, P\), as
\[
P = \begin{bmatrix}
P_1 & P_{12} \\
P_{12}^T & P_2
\end{bmatrix},
\] (48)
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 \(\hat{x}_k\). This is inspired by a similar assumption made in [5]. This structure enables us to utilize the parallel dynamics with and without attack (see Remarks [1] and [4]) and linearize the original LMI with respect to \(K\) and \(L\).

This structure also permits inverting each block separately, such that \(P_2^{-1} = Q_x\) and
\[
\begin{bmatrix}
P_1 & P_{12} \\
P_{12}^T & P_2
\end{bmatrix}^{-1} = \begin{bmatrix}
Q_x & Q_{x x} \\
Q_{x x}^T & Q_x
\end{bmatrix}.
\] (49)

Consider the linearizing change of co-ordinates used in [5],
\[
T_2 = \begin{bmatrix}
T_3 & T_3 \\
T_3 & 0
\end{bmatrix}, \quad T_3 = \begin{bmatrix}
Q_x & I \\
0 & I
\end{bmatrix}.
\] (50)
Although (20) is not entirely linearized with this transformation, due to the presence of term \(\Sigma\) which depends on \(L\),
Fig. 1: The reachable set (gray) is approximated by ellipsoids with shape matrix \( \sigma \mathbf{P} \). Without relaxation effects, the role of \( \sigma \) 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. Due to our techniques to linearize the optimization, the reachable set approximation will have some extra conservatism and we do not expect the optimal ellipsoid to be exactly tangent.

we will introduce an iterative approach later to avoid this nonlinearity. The LMI \( \mathcal{H}_L \), \((20)\), becomes

\[
\mathcal{H}_L = T_2^T \mathcal{H}_T T_2 = \begin{bmatrix}
\sigma \mathbf{P}_L & A_L^T \mathbf{P}_L & B_L \\
A_L & \mathbf{P}_L & B_L \\
0 & B_L^T & \frac{1}{2} W
\end{bmatrix},
\]

(51)

where

\[
\mathbf{P}_L = T_3^T \mathbf{P} T_3 = \begin{bmatrix}
\mathbf{Q}_x & I \\
I & \mathbf{P}_3 \\
0 & 0
\end{bmatrix},
\]

(52)

\[
B_L = T_3^T \mathbf{P} B = \begin{bmatrix}
I \\
\mathbf{P}_1 & Y_1 \\
\mathbf{P}_3 & -\mathbf{P}_3 L_4
\end{bmatrix},
\]

\[
A_L = T_3^T \mathbf{P} A T_3 = \begin{bmatrix}
\mathbf{F} \mathbf{Q}_x + \mathbf{G}_x X_1 & \mathbf{F} \\
Z_1 & \mathbf{P}_1 F + Y_1 C -Y_1 C \\
0 & \mathbf{P}_3 F
\end{bmatrix},
\]

\[
Y_1 = \mathbf{P}_3 L_4, \quad X_1 = K \mathbf{Q}_x^*, \\
Z_1 = \mathbf{P}_1 F \mathbf{Q}_x + \mathbf{P}_3 L C \mathbf{Q}_x + \mathbf{P}_1 G K \mathbf{Q}_x^T + \mathbf{P}_3 L F \mathbf{Q}_x^T + \mathbf{P}_3 K \mathbf{Q}_x^T \\
+ \mathbf{P}_3 L C \mathbf{Q}_x^T - \mathbf{P}_3 L C \mathbf{Q}_x^T.
\]

One of the useful features of this transformation is that \( \mathbf{Q}_x = \mathbf{E}_x^T \mathbf{Q} \mathbf{E}_x \), the quantity used in the objective function of Lemma 2 appears as a variable of the LMI. This section provides the linearization necessary to separate the gains \( K \) and \( L \) as variables in Lemma 2 (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. As Remark 2 states, there are an infinite number of potential outer bounding - and tight - ellipsoids. In order to combine the LMI constraints from the reachable set and \( \| H \|_2 \) calculations, 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 \( E_{\hat{x}}^T \mathbf{Q} E_{\hat{x}} \) - see (19) - to have the same orientation as the covariance of the states and estimate without attack (\( \zeta_k \)),

\[
\sigma \mathbf{P} = E_{\hat{x}}^T \mathbf{Q} E_{\hat{x}},
\]

(53)

where \( \sigma \) 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) constraints,

\[
\sigma \begin{bmatrix}
\mathbf{P}_x \\
\mathbf{P}_{xx} \\
\mathbf{P}_{x} \\
\mathbf{P}_{xx}
\end{bmatrix} = \begin{bmatrix}
\mathbf{Q}_x \\
\mathbf{Q}_{xx} \\
\mathbf{Q}_x \\
\mathbf{Q}_{xx}
\end{bmatrix},
\]

\[
\begin{bmatrix}
\frac{1}{\sigma} \mathbf{Q}_{k \hat{x}} \\
\mathbf{Q}_{k \hat{x}}
\end{bmatrix} = \begin{bmatrix}
\mathbf{P}_1 \\
\mathbf{P}_{12}
\end{bmatrix}.
\]

(54)

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

Remark 5: It is intuitive that for small values of \( \sigma \) the optimization will be infeasible and for extremely large values of \( \sigma \) the ellipsoid trivially bounds the reachable set. The optimal value \( \sigma^* \), where the ellipsoidal bound is tangent to the reachable set, will be calculated in the algorithm that wraps the convex optimization. If \( \sigma > \sigma^* \) the optimization is feasible and if \( \sigma < \sigma^* \) the problem is infeasible. Figure 1 illustrates this intuition.

Remark 6: Our techniques to linearize the problem and the existence of overestimation that is inherent in Lemma 1 the value of \( \sigma^* \) does not in practice achieve tangency to the reachable set characterized by optimized gain matrices in this paper. Nonetheless, the notion that a \( \sigma^* \) exists for feasibility of a solution is still valid, even if it does incorporate an extra amount of conservatism.

Based on (54) we can link the variables of bounding ellipsoidal LMI with \( \| H \|_2 \) constraint,

\[
\begin{align*}
X &= \sigma Y_1 = \mathbf{Q}_{k \hat{x}} L, \\
Y &= \frac{X_1}{\sigma} = K \mathbf{P}_{x}^*,
\end{align*}
\]

\[
Z = Z_1 = \mathbf{Q}_1 F \mathbf{P}_x + X \mathbf{C} \mathbf{P}_x + \mathbf{Q}_1 G \mathbf{Y} + \mathbf{Q}_{k \hat{x}} F \mathbf{P}_{xx},
\]

\[
+ \mathbf{Q}_{k \hat{x}} G \mathbf{Y} - X \mathbf{C} \mathbf{P}_{xx}.
\]

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

\[
A_L = \begin{bmatrix}
\sigma (F \mathbf{P}_x + \mathbf{G} \mathbf{Y}) & F \\
Z & \frac{1}{\sigma} (\mathbf{Q}_1 F + X \mathbf{C}) & -\frac{1}{2} \mathbf{X} \mathbf{C}
\end{bmatrix},
\]

\[
B_L = \begin{bmatrix}
I \\
\frac{1}{\sigma} \mathbf{Q}_1 \\
\mathbf{P}_3 + \mathbf{Q}_1 L
\end{bmatrix}, \quad \mathbf{P}_L = \begin{bmatrix}
\sigma \mathbf{P}_x & I \\
\mathbf{P}_3 & 0
\end{bmatrix}.
\]

Thus the choice in (53) has facilitated integrating these optimizations.

Theorem 2: Consider a LTI system (11) with desired output covariance constrained \( \| H \|_2 \) gain \( \bar{\gamma} \) (25), chi-squared detector threshold \( \alpha \) (11) and zero-alarm stealthy attacker (13). Algorithm (III.1) returns optimal controller \( K^* \) and observer
and finally made a structural connection between these two
this proof. First, using the same value for \( \sigma \) in (38) - that minimizes the ellipsoidal bound while satisfying
\( P \) decision variables, but also explicitly in the nonlinear term

\( L \) suggests a new value of \( L \) to permit the ellipsoidal bound to contain the reachable set, hence the optimization (56) will become infeasible. This is the stopping condition for the algorithm.

Note that the convergence criteria for \( \sigma \) and the decrement amount for \( \sigma \) are selected by the user. In principle, these should be small, but making them larger will allow the algorithm to require fewer iterations. In practice the number of iterations for \( \sigma \) to converge tends to be quite small (typically 2-3).

Recall that, in general, solving the Ricatti equation (38) yields up to \( \binom{n}{m} \) solutions. We assert that there will always exist at least one real solution due to matrices \( P_L, P, C_L, \) and \( C \) being positive definite. To chose between the real solutions, we choose the one that yields the smallest ellipsoidal outer bound, as determined by Lemma 2 (which provides a tighter approximation of the reachable set, since it does not require the additional linearization steps).

Finally, the stability of the closed loop system is implicitly guaranteed if \( H_L > 0 \) and \( C_L > 0 \) (see Appendix).

**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 \( \alpha = 5.99 \), and system noise truncated \( \Pr[v_k^T R_1^{-1} v_k \leq \bar{\nu}] = p_\nu = 95\% \). In this work we solve the semi-definite programming problems with the software YALMIP, with solver SeDuMi [14].

\[
F = \begin{bmatrix}
0.84 & 0.23 \\
-0.47 & 0.12
\end{bmatrix}, \quad
G = \begin{bmatrix}
0.07 & 0.23 \\
1 & 0
\end{bmatrix}, \quad
C = \begin{bmatrix}
0.45 & -0.11 \\
-0.11 & 0.20
\end{bmatrix}, \quad
R_2 = 1, \quad \gamma = 1.08.
\]

From Theorem 1 we calculate the minimum \( \|H\|_2 \) gain \( \gamma^* = 1.06 \), which is achieved by the \( \binom{4}{3} \) solutions of the Ricatti equation (38). We pick the (real) solution that yields the minimum trace of the ellipsoidal bound (using Lemma 2). The gains that achieve the optimum \( \|H\|_2 \) gain \( \gamma^* \) are

\[
L = \begin{bmatrix}
0.31 \\
-0.21
\end{bmatrix}, \quad
K = \begin{bmatrix}
-12.00 & -3.29
\end{bmatrix}.
\]

To test the validity of the relaxation in (39), we compute \( P_x \) again by plugging these gains into the *unrelated* Lyapunov equation (27). The error between the two values of \( CP_x C^T \) is 0%, demonstrating that the objective function in Theorem 1 drives the relaxed inequality toward equality. 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
Fig. 2: The comparison of empirical reachable sets corresponding to gains $K$ and $L$ designed with Theorem 2 (black, proposed algorithm) versus Theorem 1 (gray, $\|H\|_2$ optimal) demonstrates the effectiveness of our tools to minimize the reachable set due to the attacker. The optimal gains are designed in Theorem 2 using the (red) ellipsoid with shape matrix $\sigma P_\alpha$. Lemma 2 provides a better (blue) ellipsoid given the same optimal gains. Lemma 2 with the $\|H\|_2$ optimal gains provides the (green) ellipsoid.

minimize the reachable set while preserving an $\|H\|_2$ gain $\gamma^* \leq \gamma \leq \tilde{\gamma} = 1.08$, are

$$L = \begin{bmatrix} 0.50 \\ 0.10 \end{bmatrix}, \quad K = \begin{bmatrix} -1.58 & -1.96 \end{bmatrix},$$

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 ellipsoid bound corresponding to the same optimal $K$ and $L$, but using Lemma 2 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. The figure demonstrates the margin of conservatism induced by constructing LMIs from the original nonlinear constraints. Despite the gap caused by these relaxations and imposed structures, the figure also highlights the effectiveness of our tools to optimize gains that minimize the reachable set.

We evaluate the effectiveness of the relaxation (39) now in the context of Theorem 2 and observe an error of 2% between the two values of $CP_x C^T$. The small but nonzero difference suggests the presence of the $H_L$ constraint keeps the objective from completely driving the relaxed version (39) to equality to match (27) exactly.

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 compromises sensor measurements. As past work has observed, there is a necessary trade-off between ensuring performance and minimizing the effect of the attacker. Here impact of the attack is quantified as the set of states reachable through the action of the attacker and nominal closed-loop performance is specified through a desired output covariance constrained $\|H\|_2$ gain. We frame this problem as an LMI wrapped in an iterative algorithm and a large part of the effort here is to linearize the constraints involved. Along the way, we also contribute a convex optimization to design optimal OCC $\|H\|_2$ gains. Future work looks toward devising new transformations to eliminate the iterative part of the algorithm and also to impose less structure on the solution we find.

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. Nesci, “Security metrics of networked control systems under sensor attacks (extended preprint),” ArXiv, vol. abs/1809.01108, 2018.
[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. Rota, 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] C. Murguia and J. Ruths, “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] S. Boyd, L. El Ghaoui, E. Feron, and V. Balakrishnan, Linear Matrix Inequalities in System and Control Theory, ser. Studies in Applied Mathematics. Philadelphia, PA: SIAM, 1994, vol. 15.
[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.

APPENDIX

We show that the LMI constraints $H_L > 0$ and $C_L > 0$ imply that the attacked system and nominal system are stable. For $H_L$, $H_L > 0$ implies $H > 0$ which implies $P > 0$ and $aP - A^TPA > 0$. Since $a \in [0, 1]$, $(1-a)P + aP = A^TPA > 0$. Thus, $A^T$, and hence, $A$ is stable.

Similarly, for $C_L$, $C_L > 0$ implies $C > 0$ which implies $P > 0$ and $P - A\hat{P} \hat{A}^T < 0$. Therefore, $P - A\hat{P} \hat{A}^T > 0$, which makes $\hat{A}$ stable as well.