Simultaneous Input and State Set-Valued Observers with Applications to Attack-Resilient Estimation

Sze Zheng Yong

Abstract— **In this paper, we present a fixed-order set-valued observer for linear discrete-time bounded-error systems that simultaneously finds bounded sets of compatible states and unknown inputs that are optimal in the minimum $\mathcal{H}_\infty$-norm sense, i.e., with minimum average power amplification. We also analyze the necessary and sufficient conditions for the stability of the observer and its connection to a system property known as strong detectability. Next, we show that the proposed set-valued observer can be used for attack-resilient estimation of state and attack signals when cyber-physical systems are subject to false data injection attacks on both actuator and sensor signals. Moreover, we discuss the implication of strong detectability on resilient state estimation and attack identification. Finally, the effectiveness of our set-valued observer is demonstrated in simulation, including on an IEEE 14-bus electric power system.**

I. Introduction

Cyber-physical systems (CPS) are systems in which computational and communication elements collaborate to control physical entities. Such systems include the power grid, autonomous vehicles, medical devices, etc. Most of these systems are *safety-critical* and if compromised or malfunctioning, can cause serious harm to the controlled physical entities and the people operating or utilizing them. Recent incidents of attacks on CPS, e.g., the Ukrainian power grid, the Maroochy water service and an Iranian nuclear plant [1]–[3] highlight a need for CPS security and for new designs of resilient estimation and control.

In particular, false data injection attack is one of the most serious forms of attacks on CPS, where malicious and strategic attackers intrude and inject fake data signals into the sensor measurements and actuator signals with the goal of causing harm, energy theft etc. Given the strategic nature of these false data injection signals, they are not well-modeled by a zero-mean, Gaussian white noise nor by signals with known bounds. Hence, traditional Kalman filtering and unknown input observers do not apply. Nevertheless, reliable estimates of states and unknown inputs are valuable and needed for purposes of resilient control, attack identification, etc. Similar state and input estimation problems can be found across a wide range of disciplines, from the estimation of mean areal precipitation [4] to fault detection and diagnosis [5] to input estimation in physiological systems [6].

Literature review. Much of the research focus on simultaneous input and state estimation has been on obtaining *point* estimates for deterministic systems with unknown inputs via asymptotic and sliding mode observers (e.g., [7]–[9]) or for stochastic systems with unknown inputs via unbiased minimum-variance estimation (e.g., [10]–[14]). These methods do not directly apply to bounded-error models, i.e., uncertain dynamic systems with set-valued uncertainties, where instead, the sets of states and unknown inputs that are compatible/consistent with sensor observations are desired. Similarly, while $\mathcal{H}_2, \mathcal{H}_\infty, \mathcal{L}_1$ filters (e.g., [15]–[17]) can deal with bounded modeling errors, only point estimates of states are obtained in addition to the fact that it is unsuited to handle large unknown inputs.

In contrast, set-membership or set-valued state observers are capable of estimating the set of compatible states and are preferable to stochastic estimation when hard accuracy bounds are important [18], e.g., to guarantee safety. Since its conception, it was apparent that characterizing the set of states that are compatible with measurements is in general computationally intensive. The complexity of optimal observers [19] grows with time, and also for methods based on an ℓ_1 model matching problem [20] and polyhedral set computation using Fourier-Motzkin elimination [21]. Thus, fixed-order recursive filters were designed with *equalized performance* (i.e., with invariant estimation errors) for *superstable* systems in [18], [22]. However, all these set-membership approaches can only compute the set of compatible states and do not apply when the unknown input signals have unknown bounds, as is often required in attack-resilient estimation where the attack signals are malicious and strategic.

In the context of attack-resilient estimation against false data injection attacks, numerous approaches were proposed for deterministic systems (e.g., [23]–[26]), stochastic systems (e.g., [27]–[29]) and bounded-error systems [30]–[32], but they share the common theme of only obtaining *point* estimates. In particular, error bounds were computed in [30] for only the initial state and in [31] with the assumption of zero initial state and without optimality considerations. More importantly, only sensor attacks are considered and set-valued estimates of state and attack signals are not computed.

Contributions. The goal of this paper is to bridge the gap between set-valued state estimation without unknown inputs and point-valued state and unknown input estimation. We propose a fixed-order set-valued observer for linear discrete-time bounded-error systems that simultaneously finds bounded sets of states and unknown inputs that contain the true state and unknown input, are compatible/consistent with measurement outputs and are optimal in the minimum $\mathcal{H}_\infty$-norm sense, i.e., with minimum average power amplification. In addition, we provide the necessary and sufficient

S.Z. Yong is with the School for Engineering of Matter, Transport and Energy, Arizona State University, Tempe, AZ, USA (e-mail: szyong@asu.edu).

conditions for observer stability and boundedness of the set-valued estimates, which we show is closely related to a system property known as strong detectability.

We further show that the proposed set-valued observer is applicable for achieving attack-resiliency in cyber-physical systems against false data injection attacks on both actuator and sensor signals. Specifically, the set-valued observer can compute the sets of states and attack signals that are compatible with measurements, where the latter enables not only attack detection but also identification. Moreover, we discuss the implication of strong detectability on resilient state estimation and attack identification. Finally, the effectiveness of our set-valued observer is demonstrated using a benchmark system and an IEEE 14-bus electric power system.

Notation. $\mathbb{R}^n$ denotes the n-dimensional Euclidean space, $\mathbb{C}$ the field of complex numbers and $\mathbb{N}$ nonnegative integers. For a vector $v \in \mathbb{R}^n$ and a matrix $M \in \mathbb{R}^{p \times q}$, $\|v\| \triangleq \sqrt{v^\top v}$ and $\|M\|$ denote their (induced) 2-norm. Moreover, the transpose, inverse, Moore-Penrose pseudoinverse and rank of M are given by $M^\top$, M^{-1}, $M^\dagger$ and $\mathrm{rk}(M)$. For a symmetric matrix S, $S \succ 0$ ($S \succeq 0$) is positive (semi-) definite.

II. Problem Statement

System Assumptions. Consider the linear time-invariant discrete-time bounded-error system

$$\begin{aligned} x_{k+1} &= Ax_k + Bu_k + Gd_k + Ww_k, \\ y_k &= Cx_k + Du_k + Hd_k + v_k, \end{aligned} \tag{1}$$

where $x_k \in \mathbb{R}^n$ is the state vector at time $k \in \mathbb{N}$, $u_k \in \mathbb{R}^m$ is a known input vector, $d_k \in \mathbb{R}^p$ is an unknown input vector, and $y_k \in \mathbb{R}^l$ is the measurement vector. The process noise $w_k \in \mathbb{R}^n$ and the measurement noise $v_k \in \mathbb{R}^l$ are assumed to be bounded, with $\|w_k\| \leq \eta_w$ and $\|v_k\| \leq \eta_v$ (thus, they are ℓ_∞ sequences). We also assume an estimate $\hat{x}_0$ of the initial state x_0 is available, where $\|\hat{x}_0 - x_0\| \leq \delta^x_0$. The matrices A, B, C, D, G, H and W are known and of appropriate dimensions, where G and H are matrices that encode the *locations* through which the unknown input or attack signal can affect the system dynamics and measurements. Note that no assumption is made on H to be either the zero matrix (no direct feedthrough), or to have full column rank when there is direct feedthrough. Without loss of generality, we assume that $\mathrm{rk}[G^\top \ H^\top] = p$, $n \geq l \geq 1$, $l \geq p \geq 0$ and $m \geq 0$.

Unknown Input (or Attack) Signal Assumptions. The unknown inputs d_k are not constrained to be a signal of any type (random or strategic) nor to follow any model, thus no prior 'useful' knowledge of the dynamics of d_k is available (independent of $\{d_\ell\}$ $\forall k \neq \ell$, $\{w_\ell\}$ and $\{v_\ell\}$ $\forall \ell$). We also do not assume that d_k is bounded or has known bounds and thus, d_k is suitable for representing adversarial attack signals.

The simultaneous input and state set-valued observer design problem is twofold and can be stated as follows:

1) *Given a linear discrete-time bounded-error system with unknown inputs* (1)*, design an optimal and stable filter that simultaneously finds bounded sets of compatible states and unknown inputs in the minimum $\mathcal{H}_\infty$-norm sense, i.e., with minimum average power amplification.*
2) *Develop an attack-resilient set-valued observer for system* (1) *that computes a bounded set of state estimates that contains the true state and identifies the set of compatible attack signals irrespective of the magnitude of false data injection attacks on its actuators and sensors. In addition, recommend preventative attack mitigation strategies based on detectability conditions.*

III. Preliminary Material

A. System Transformation

We first carry out a transformation of the system to decouple the output equation into two components, one with a full rank direct feedthrough matrix and the other without direct feedthrough. Note, however, that this similarity transformation is different from the one in [14], which is no longer applicable as it was based on the noise error covariance.

Let $p_H \triangleq \mathrm{rk}(H)$. Using singular value decomposition, we rewrite the direct feedthrough matrix H as $H = \begin{bmatrix} U_1 & U_2 \end{bmatrix} \begin{bmatrix} \Sigma & 0 \\ 0 & 0 \end{bmatrix} \begin{bmatrix} V_1{}^\top \\ V_2{}^\top \end{bmatrix}$, where $\Sigma \in \mathbb{R}^{p_H \times p_H}$ is a diagonal matrix of full rank, $U_1 \in \mathbb{R}^{l \times p_H}$, $U_2 \in \mathbb{R}^{l \times (l-p_H)}$, $V_1 \in \mathbb{R}^{p \times p_H}$ and $V_2 \in \mathbb{R}^{p \times (p-p_H)}$, while $U \triangleq \begin{bmatrix} U_1 & U_2 \end{bmatrix}$ and $V \triangleq \begin{bmatrix} V_1 & V_2 \end{bmatrix}$ are unitary matrices. When there is no direct feedthrough, Σ, U_1 and V_1 are empty matrices[a], and U_2 and V_2 are arbitrary unitary matrices.

Then, we define two orthogonal components of the unknown input given by

$$d_{1,k} = V_1^\top d_k, \quad d_{2,k} = V_2^\top d_k. \tag{2}$$

Since V is unitary, $d_k = V_1 d_{1,k} + V_2 d_{2,k}$ and the system (1) can be rewritten as

$$\begin{aligned} x_{k+1} &= Ax_k + Bu_k + GV_1 d_{1,k} + GV_2 d_{2,k} + Ww_k \\ &= Ax_k + Bu_k + G_1 d_{1,k} + G_2 d_{2,k} + Ww_k, \end{aligned} \tag{3}$$

$$\begin{aligned} y_k &= Cx_k + Du_k + HV_1 d_{1,k} + HV_2 d_{2,k} + v_k \\ &= Cx_k + Du_k + H_1 d_{1,k} + v_k, \end{aligned} \tag{4}$$

where $G_1 \triangleq GV_1$, $G_2 \triangleq GV_2$ and $H_1 \triangleq HV_1 = U_1 \Sigma$. Next, we decouple the output y_k using a nonsingular transformation $T = \begin{bmatrix} T_1^\top & T_2^\top \end{bmatrix}^\top \triangleq U^\top = \begin{bmatrix} U_1 & U_2 \end{bmatrix}^\top$ to obtain $z_{1,k} \in \mathbb{R}^{p_H}$ and $z_{2,k} \in \mathbb{R}^{l-p_H}$ given by

$$\begin{aligned} z_{1,k} &\triangleq T_1 y_k = U_1^\top y_k = C_1 x_k + D_1 u_k + \Sigma d_{1,k} + v_{1,k} \\ z_{2,k} &\triangleq T_2 y_k = U_2^\top y_k = C_2 x_k + D_2 u_k + v_{2,k} \end{aligned} \tag{5}$$

where $C_1 \triangleq U_1^\top C$, $C_2 \triangleq U_2^\top C$, $D_1 \triangleq U_1^\top D$, $D_2 \triangleq U_2^\top D$, $v_{1,k} \triangleq U_1^\top v_k$ and $v_{2,k} \triangleq U_2^\top v_k$. This transform is also chosen such that $\| \begin{bmatrix} v_{1,k}^\top & v_{2,k}^\top \end{bmatrix}^\top \| = \|U^\top v_k\| = \|v_k\|$.

B. Input and State Detectability (a.k.a. Strong Detectability)

Similar to the stability of the deterministic and stochastic input and state observers/filters, we will show in Section IV-B that the stability of the set-valued observer is directly related to the notion of strong detectability. Without loss of generality, we assume that $B = 0$ and $D = 0$ in this section, since u_k is known.

[a] We adopt the convention that the inverse of an empty matrix is also an empty matrix and assume that operations with empty matrices are possible.

Definition 1 (Strong detectability)**.** *The linear system* (1) *is* strongly detectable *if*

$$y_k = 0 \;\forall\, k \geq 0 \quad \textit{implies} \quad x_k \to 0 \textit{ as } k \to \infty$$

for all initial states and input sequences $\{d_i\}_{i \in \mathbb{N}}$.

Definition 2 (Invariant Zeros)**.** *The invariant zeros* z *of the Rosenbrock system matrix* $\mathcal{R}_S(z) := \begin{bmatrix} zI - A & -G \\ C & H \end{bmatrix}$ *of system* (1) *are the finite values of* z *for which* $\mathcal{R}_S(z)$ *drops rank, i.e.,* $\mathrm{rk}(\mathcal{R}_S(z)) < \mathrm{nrank}(\mathcal{R}_S)$*, where* $\mathrm{nrank}(\mathcal{R}_S)$ *is the normal rank (maximum rank over* $z \in \mathbb{C}$*) of* $\mathcal{R}_S(z)$*.*

Theorem 1 (Strong detectability)**.** *A linear time-invariant discrete-time system is strongly detectable if and only if either of the following holds for all* $z \in \mathbb{C}, |z| \geq 1$*:*

(i) $\mathrm{rk}\,\mathcal{R}_S(z) \triangleq \mathrm{rk} \begin{bmatrix} zI - A & -G \\ C & H \end{bmatrix} = n + p$,

(ii) $\mathrm{rk}\,\hat{\mathcal{R}}_S(z) \triangleq \mathrm{rk} \begin{bmatrix} zI - \hat{A} & -G_2 \\ C_2 & 0 \end{bmatrix} = n + p - p_H$,

(iii) $\mathrm{rk}\,\overline{\mathcal{R}}_S(z) \triangleq \mathrm{rk} \begin{bmatrix} zI - \overline{A} & -G_2 \\ C_2 & 0 \end{bmatrix} = n + p - p_H$,

(iv) $\mathrm{rk}\,\overline{\mathcal{R}}^{\star}_S(z) \triangleq \mathrm{rk} \begin{bmatrix} zI - \overline{A} & -G_2 \\ C_2\overline{A} & C_2 G_2 \end{bmatrix} = n + p - p_H$,

where $\hat{A} \triangleq A - G_1 \Sigma^{-1} C_1$ *and* $\overline{A} \triangleq (I - G_2 \tilde{M}_2 C_2)\hat{A}$ *for any* $\tilde{M}_2 \in \mathbb{R}^{(p-p_H)\times(p-p_H)}$*. The above conditions are equivalent to the system being minimum-phase (i.e., the invariant zeros of* $\mathcal{R}_S(z)$ *in Condition (i) are stable).*

Moreover, strong detectability implies that the pairs (A, C), $(\hat{A}, C_2)$, $(\overline{A}, C_2)$ *and* $(\overline{A}, C_2\overline{A})$ *are detectable; and if* $l = p$*, then strong detectability implies that the pairs* (A, G), $(\hat{A}, G_2)$ *and* $(\overline{A}, G_2)$ *are stabilizable.*

Proof. The equivalence of Conditions (i) and (ii) with strong detectability in Definition 1 can be found in [14]. Thus, it is sufficient to show the equivalence of Conditions (ii), (iii) and (iv) using the following identity for all $z \in \mathbb{C}, z \neq 0$:

$$\begin{aligned} \mathrm{rk} \begin{bmatrix} zI - \hat{A} & -G_2 \\ C_2 & 0 \end{bmatrix} &= \mathrm{rk} \begin{bmatrix} zI - \hat{A} & -G_2 \\ C_2 & 0 \end{bmatrix} \begin{bmatrix} I & 0 \\ -\tilde{M}_2 C_2 \hat{A} & I \end{bmatrix} \\ = \mathrm{rk} \begin{bmatrix} zI - \overline{A} & -G_2 \\ C_2 & 0 \end{bmatrix} &= \mathrm{rk} \begin{bmatrix} I & 0 \\ -C_2 & zI \end{bmatrix} \begin{bmatrix} zI - \overline{A} & -G_2 \\ C_2 & 0 \end{bmatrix} \\ = \mathrm{rk} \begin{bmatrix} zI - \overline{A} & -G_2 \\ C_2\overline{A} & C_2 G_2 \end{bmatrix}. \end{aligned}$$

Finally, comparing Conditions (i)–(iv) to the PBH rank test for detectability (and stabilizability), we see that strong detectability implies that (A, C), $(\hat{A}, C_2)$, $(\overline{A}, C_2)$ and $(\overline{A}, C_2\overline{A})$ are detectable (and that (A, G), $(\hat{A}, G_2)$, $(\overline{A}, G_2)$ are stabilizable if $l = p$). ∎

IV. Fixed-Order Simultaneous Input and State Set-Valued Observers

A. Set-Valued Observer Design

We consider a recursive three-step set-valued observer design (similar to [12], [14]), composed of an *unknown input estimation* step that uses the current measurement and the set of compatible states to estimate the set of compatible unknown inputs, a *time update* step that propagates the compatible set of states based on the system dynamics, and a *measurement update* step that updates the set of compatible states using the current measurement. In brief, our goal is to design a recursive three-step set-valued observer of the form:

$$\begin{aligned} \textit{Unknown Input Estimation: } \hat{D}_{k-1} &= \mathcal{F}_d(\hat{X}_{k-1}, u_k), \\ \textit{Time Update: } \quad \hat{X}^{\star}_k &= \mathcal{F}^{\star}_x(\hat{X}_{k-1}, \hat{D}_{k-1}, u_k), \\ \textit{Measurement Update: } \quad \hat{X}_k &= \mathcal{F}_x(\hat{X}^{\star}_k, u_k, y_k), \end{aligned}$$

where $\mathcal{F}_d$, $\mathcal{F}^{\star}_x$ and $\mathcal{F}_x$ are the to-be-designed set mappings while $\hat{D}_{k-1}$, $\hat{X}^{\star}_k$ and $\hat{X}_k$ are the sets of compatible unknown inputs at time $k-1$, propagated and updated states at time k, respectively. Note that we have a (one-step) delayed estimate of $\hat{D}_{k-1}$ because it is the only estimate we can obtain in light of (5) since $d_{2,k-1}$ does not affect $z_{1,k}$ and $z_{2,k}$, and hence, cannot be estimated from y_k. The reader is referred to a previous work [13] for a detailed discussion on when a delay is absent or when further delays are expected.

Since the complexity of optimal observers grows with time [19]–[21], we will only consider fixed-order recursive filters as in [18], [22], where set-valued estimates are of the form:

$$\hat{D}_{k-1} = \{d \in \mathbb{R}^p : \|d_{k-1} - \hat{d}_{k-1}\| \leq \delta^d_{k-1}\}, \qquad (6)$$

$$\hat{X}^{\star}_k = \{x \in \mathbb{R}^n : \|x_k - \hat{x}^{\star}_{k|k}\| \leq \delta^{x,\star}_k\}, \qquad (7)$$

$$\hat{X}_k = \{x \in \mathbb{R}^n : \|x_k - \hat{x}_{k|k}\| \leq \delta^x_k\}, \qquad (8)$$

i.e., we confine the estimation errors to balls of norm δ. In this case, the observer design problem is reduced to finding the centroids $\hat{d}_{k-1}$, $\hat{x}^{\star}_{k|k}$ and $\hat{x}_{k|k}$ as well as the radii δ^d_{k-1}, $\delta^{x,\star}_k$ and δ^x_k of the sets $\hat{D}_{k-1}$, $\hat{X}^{\star}_k$ and $\hat{X}_k$, respectively.

We further limit our attention to observers for the centroids $\hat{d}_{k-1}$, $\hat{x}^{\star}_{k|k}$ and $\hat{x}_{k|k}$ that belong to the class of three-step recursive filters given in [12], [14], defined as follows for each time k (with $\hat{x}_{0|0} = \hat{x}_0$):

Unknown Input Estimation:

$$\hat{d}_{1,k} = M_1(z_{1,k} - C_1\hat{x}_{k|k} - D_1 u_k), \qquad (9)$$

$$\hat{d}_{2,k-1} = M_2(z_{2,k} - C_2\hat{x}_{k|k-1} - D_2 u_k), \qquad (10)$$

$$\hat{d}_{k-1} = V_1\hat{d}_{1,k-1} + V_2\hat{d}_{2,k-1}, \qquad (11)$$

Time Update:

$$\hat{x}_{k|k-1} = A\hat{x}_{k-1|k-1} + Bu_{k-1} + G_1\hat{d}_{1,k-1}, \qquad (12)$$

$$\hat{x}^{\star}_{k|k} = \hat{x}_{k|k-1} + G_2\hat{d}_{2,k-1}, \qquad (13)$$

Measurement Update:

$$\begin{aligned} \hat{x}_{k|k} &= \hat{x}^{\star}_{k|k} + L(y_k - C\hat{x}^{\star}_{k|k} - Du_k) \\ &= \hat{x}^{\star}_{k|k} + \tilde{L}(z_{2,k} - C_2\hat{x}^{\star}_{k|k} - D_2 u_k), \end{aligned} \qquad (14)$$

where $L \in \mathbb{R}^{n\times l}$, $\tilde{L} \triangleq LU_2 \in \mathbb{R}^{n\times(l-p_H)}$, $M_1 \in \mathbb{R}^{p_H\times p_H}$ and $M_2 \in \mathbb{R}^{(p-p_H)\times(l-p_H)}$ are observer gain matrices that are chosen in the following lemma and theorem to minimize the "volume" of the set of compatible states and unknown inputs, quantified by the radii δ^d_{k-1}, $\delta^{x,\star}_k$ and δ^x_k. Their proofs will be provided in the appendix. Note also that we applied $L = LU_2U_2^\top = \tilde{L}U_2^\top$ from the following lemma into (14).

Lemma 1 (Necessary Conditions for Boundedness of Set–Valued Estimates)**.** *The input and state estimation errors,* $\tilde{d}_{k-1} \triangleq d_{k-1} - \hat{d}_{k-1}$ *and* $\tilde{x}_{k|k} \triangleq x_k - \hat{x}_{k|k}$*, are bounded for* *all k (i.e., the set-valued estimates are bounded with radii $\delta^d_{k-1}, \delta^{x,\star}_k, \delta^x_k < \infty$), only if $M_1\Sigma = I$, $M_2C_2G_2 = I$ and $LU_1 = 0$. Consequently, $\mathrm{rk}(C_2G_2) = p - p_H$, $M_1 = \Sigma^{-1}$, $M_2 = (C_2G_2)^\dagger$ and $L = LU_2U_2^\top = \tilde{L}U_2^\top$.*

Theorem 2. *Suppose Lemma 1 holds, and let $T_{\tilde{x},w,v}$ denote the transfer function matrix that maps the noise signals $\vec{w}_k \triangleq \begin{bmatrix} w_k^\top & v_k^\top \end{bmatrix}^\top$ to the updated state estimation error $\tilde{x}_{k|k} \triangleq x_k - \hat{x}_{k|k}$. Moreover, assume that the following hold:*

- *(A.1)* (A_∞, C_∞) *is detectable,*
- *(A.2)* $D_\infty D_\infty^\top \succ 0$ *and*
- *(A.3)* $\mathrm{rk}\begin{bmatrix} A_\infty - e^{j\omega}I & B_\infty \\ C_\infty & D_\infty \end{bmatrix} = n + l - p_H,\ \forall \omega \in [0, 2\pi]$,

with $\hat{A} \triangleq A - G_1M_1C_1$, $\Phi \triangleq I - G_2M_2C_2$, $\overline{A} \triangleq \Phi\hat{A}$, $A_\infty \triangleq \overline{A}$, $B_\infty \triangleq \begin{bmatrix} \Phi W & -\Phi G_1M_1 & -\sqrt{2}G_2M_2 & 0 \end{bmatrix}$, $C_\infty \triangleq C_2\overline{A}$ and $D_\infty \triangleq \begin{bmatrix} C_2\Phi W & -C_2\Phi G_1M_1 & -\sqrt{2}C_2G_2M_2 & \sqrt{2}I \end{bmatrix}$. Then, there exists an $\mathcal{H}_\infty$-observer that satisfy $\|T_{\tilde{x},w,v}\|_\infty \leq \gamma$, i.e., the maximum average signal power amplification is upper-bounded by γ^2:

$$\frac{\lim_{k\to\infty} \frac{1}{k+1} \sum_{i=0}^{k} \tilde{x}_{i|i}^\top \tilde{x}_{i|i}}{\lim_{k\to\infty} \frac{1}{k+1} \sum_{i=0}^{k} \vec{w}_i^\top \vec{w}_i} = \|T_{\tilde{x},w,v}\|_\infty^2 \leq \gamma^2, \tag{15}$$

if and only if there exists $P = P^\top \succeq 0$ that satisfies the following discrete-time algebraic Riccati equation (DARE)[b]:

$$\begin{aligned} P = &-(A_\infty P + C_l^\top + B_\infty)(C_lPC_l^\top + R_l)^{-1} \\ &(A_\infty P + C_l^\top + B_\infty)^\top + B_\infty B_\infty^\top + A_\infty P A_\infty^\top, \end{aligned} \tag{16}$$

with $C_l \triangleq \begin{bmatrix} C_\infty \\ \gamma^{-1}I \end{bmatrix}$, $D_l \triangleq \begin{bmatrix} D_\infty \\ 0 \end{bmatrix}$ and $R_l \triangleq \begin{bmatrix} D_\infty D_\infty^\top & 0 \\ 0 & -I \end{bmatrix}$ such that $U_\infty \triangleq I - \gamma^{-2}P \succ 0$ and $\breve{A} \triangleq A_\infty - (A_\infty PC_\infty^\top + B_\infty D_l^\top)(C_lPC_l^\top + R_l)^{-1}C_l$ is asymptotically stable, i.e., $|\lambda_i(\breve{A})| < 1$ for all eigenvalues $\lambda_i(\breve{A})$ of $\breve{A}$. When such a P matrix exists, the filter gain $\tilde{L}$ is given by

$$\tilde{L} = (B_\infty D_\infty^\top + A_\infty V_\infty C_\infty^\top)(C_\infty V_\infty C_\infty^\top + D_\infty D_\infty^\top)^{-1} \tag{17}$$

with $V_\infty = P + \gamma^{-2}PU_\infty^{-1}P$. Moreover, $A_e \triangleq (I - \tilde{L}C_2)\overline{A}$ and $A_e^\star \triangleq \overline{A}(I - \tilde{L}C_2)$ are asymptotically stable.

Thus, we can search over γ (e.g., via bisection) to find the smallest γ and the corresponding optimal observer gain $\tilde{L}$ in the minimum $\mathcal{H}_\infty$-norm sense. Further, by upper-bounding the estimation errors, we find the radii of the sets of compatible inputs and states to be (cf. proof in the appendix):

$$\begin{aligned} \delta^d_{k-1} = &\ \delta^x_0\|V_eA_e^{k-1}\| + \eta_w(\|V_2M_2C_2\| \\ &+ \sum_{i=0}^{k-2}\|V_eA_e^iB_{e,w}\|) + \eta_v(\sum_{i=0}^{k-3}\|V_eA_e^i(B_{e,v1} + A_eB_{e,v2})\| \\ &+\|V_2M_2T_2\| + \|V_eB_{e,v2} + (V_1 - V_2M_2C_2G_1)M_1T_1\| \\ &+\|V_eA_e^{k-2}B_{e,v1}\|), \end{aligned} \tag{18}$$

$$\begin{aligned} \delta^{x,\star}_k = &\ \delta^x_0\|\overline{A}A_e^{k-1}\| + \eta_w(\sum_{i=0}^{k-2}\|\overline{A}A_e^iB_{e,w}\| + \|B^\star_{e,w}\|) \\ &+\eta_v(\sum_{i=0}^{k-3}\|\overline{A}A_e^i(B_{e,v1} + A_eB_{e,v2})\| + \|B^\star_{e,v2}\| \\ &+\|\overline{A}A_e^{k-2}B_{e,v1}\| + \|B^\star_{e,v1} + \overline{A}B_{e,v2}\|), \end{aligned} \tag{19}$$

$$\begin{aligned} \delta^x_k = &\ \delta^x_0\|A_e^k\| + \eta_v(\|B_{e,v2}\| + \|A_e^{k-1}B_{e,v1}\| \\ &+ \sum_{i=0}^{k-2}\|A_e^i(B_{e,v1} + A_eB_{e,v2})\|) + \eta_w\sum_{i=0}^{k-1}\|A_e^iB_{e,w}\|, \end{aligned} \tag{20}$$

where $\hat{A} \triangleq A - G_1M_1C_1$, $\Phi \triangleq I - G_2M_2C_2$, $\overline{A} \triangleq \Phi\hat{A}$, $V_e \triangleq V_1M_1C_1 + V_2M_2C_2\hat{A}$, $A_e \triangleq (I - \tilde{L}C_2)\overline{A}$, $B^\star_{e,w} \triangleq \Phi W$, $B^\star_{e,v1} \triangleq -\Phi G_1M_1T_1$, $B^\star_{e,v2} \triangleq -G_2M_2T_2$, $B_{e,w} \triangleq (I - \tilde{L}C_2)B^\star_{e,w}$, $B_{e,v1} \triangleq (I - \tilde{L}C_2)B^\star_{e,v1}$ and $B_{e,v2} \triangleq (I - \tilde{L}C_2)B^\star_{e,v2} - \tilde{L}T_2$. Moreover, since A_e is stable as a consequence of Theorem 2 (hence, $\lim_{k\to\infty} A_e^k = 0$), it is straightforward to see that the radii converge to finite steady-state values given by $\lim_{k\to\infty}\delta^d_{k-1} = \delta^d$, $\lim_{k\to\infty}\delta^{x,\star}_k = \delta^{x,\star}$ and $\lim_{k\to\infty}\delta^x_k = \delta^x$, where

$$\begin{aligned} \delta^d \triangleq &\ \eta_w(\|V_2M_2C_2\| + \lim_{k\to\infty}\textstyle\sum_{i=0}^{k-2}\|V_eA_e^iB_{e,w}\|) \\ &+\eta_v(\lim_{k\to\infty}\textstyle\sum_{i=0}^{k-3}\|V_eA_e^i(B_{e,v1} + A_eB_{e,v2})\| + \|V_2M_2T_2\| \\ &+\|V_eB_{e,v2} + (V_1 - V_2M_2C_2G_1)M_1T_1\|) < \infty, \\ \delta^{x,\star} \triangleq &\ \eta_w(\lim_{k\to\infty}\textstyle\sum_{i=0}^{k-2}\|\overline{A}A_e^iB_{e,w}\| + \|B^\star_{e,w}\|) \\ &+\eta_v(\lim_{k\to\infty}\textstyle\sum_{i=0}^{k-3}\|\overline{A}A_e^i(B_{e,v1} + A_eB_{e,v2})\| + \|B^\star_{e,v2}\| \\ &+\|\overline{A}A_e^{k-2}B_{e,v1}\| + \|B^\star_{e,v1} + \overline{A}B_{e,v2}\|) < \infty, \\ \delta^x \triangleq &\ \eta_w\lim_{k\to\infty}\textstyle\sum_{i=0}^{k-1}\|A_e^iB_{e,w}\| + \eta_v(\|B_{e,v2}\| \\ &+ \lim_{k\to\infty}\textstyle\sum_{i=0}^{k-2}\|A_e^i(B_{e,v1} + A_eB_{e,v2})\|) < \infty. \end{aligned}$$

Algorithm 1 summarizes the three steps of the fixed-order input and state set-valued observer, in which $d_{2,k-1}$ is estimated before the time update, followed by the measurement update and finally, the estimation of $d_{1,k}$. Note that we did not include $\delta^{x,\star}$ and $\hat{X}^\star_k$ in the algorithm for conciseness.

B. Observer Stability and Strong Detectability

Next, we provide the relationship between observer stability and strong detectability (cf. Definition 1), whose proof will be provided in the appendix. Note that $\mathrm{rk}(C_2G_2) = p - p_H$ is a necessary condition for the boundedness of the set-valued estimates by Lemma 1 and this is assumed in the following results. It is also noteworthy that C_2G_2 is the first *invertibility matrix* in [13] (similar to a Markov parameter).

Lemma 2. *All non-zero invariant zeros of the system* (1) *are eigenvalues/poles of the state matrices $A_e^\star \triangleq \overline{A}(I - \tilde{L}C_2)$ and $A_e \triangleq (I - \tilde{L}C_2)\overline{A}$ of the propagated and updated state estimation error dynamics $\tilde{x}^\star_{k|k}$ and $\tilde{x}_{k|k}$, respectively, for any observer gain $\tilde{L}$, where $\overline{A} \triangleq \Phi\hat{A}$ and $\Phi \triangleq I - G_2M_2C_2$.*

Theorem 3 (Strong Detectability ⇔ Observer Stability)**.** *Suppose Lemma 1 holds. Then, strong detectability is necessary and sufficient for asymptotic stability of the observer dynamics with $A_e^\star$ and A_e, and for the boundedness of the set-valued input and state estimates for any non-zero δ^x_0.*

Theorem 4 (Strong Detectability and Existence of an $\mathcal{H}_\infty$-Observer)**.** *Suppose Lemma 1 holds. Then, strong detectability guarantees that the Assumptions (A.1) and (A.2) in Theorem 2 hold. If $p = l$, then strong detectability also satisfies Assumption (A.3). Otherwise, that $(\overline{A}, G_2)$ or (A_∞, B_∞) is stabilizable on the unit circle also satisfies*

[b] The DARE equation in (16) can be solved with control system software. For example, in MATLAB's Control System Toolbox, we can use the command $\mathtt{DARE}(A_\infty^\top, C_l^\top, \Phi(I + G_1M_1M_1^\top G_1^\top)\Phi^\top, (C_2^\top C_2 - \frac{1}{\gamma^2}I)^{-1})$.

Algorithm 1 Fixed-Order Input & State Set-Valued Observer

```
Initialize: $M_1 = \Sigma^{-1}$; $M_2 = (C_2 G_2)^\dagger$; $\hat{A} = A - G_1 M_1 C_1$;
            $\Phi = I - G_2 M_2 C_2$; $\overline{A} = \Phi \hat{A}$;
            $V_e = V_1 M_1 C_1 + V_2 M_2 C_2 \hat{A}$;
            Compute $\tilde{L}$ via Theorem 2 and perform bisection to minimize $\gamma$.
            $A_e = (I - \tilde{L} C_2)\overline{A}$; $B_{e,w} = (I - \tilde{L} C_2)\Phi W$;
            $B_{e,v1} = -(I - \tilde{L} C_2)\Phi G_1 M_1 T_1$;
            $B_{e,v2} = -((I - \tilde{L} C_2) G_2 M_2 + \tilde{L}) T_2$;
            $\hat{x}_{0|0} = \hat{x}_0 = \text{centroid}(\hat{X}_0)$;
            $\delta_0^x = \min_{\delta} \{\|x - \hat{x}_{0|0}\| \leq \delta, \forall x \in \hat{X}_0\}$;
            $\hat{d}_{1,0} = M_1(z_{1,0} - C_1 \hat{x}_{0|0} - D_1 u_0)$;
for $k = 1$ to $N$ do
  ▷ Estimation of $d_{2,k-1}$ and $d_{k-1}$
    $\hat{x}_{k|k-1} = A\hat{x}_{k-1|k-1} + Bu_{k-1} + G_1 \hat{d}_{1,k-1}$;
    $\hat{d}_{2,k-1} = M_2(z_{2,k} - C_2 \hat{x}_{k|k-1} - D_2 u_k)$;
    $\hat{d}_{k-1} = V_1 \hat{d}_{1,k-1} + V_2 \hat{d}_{2,k-1}$;
    $\delta_{k-1}^d = \delta_0^x \|V_e A_e^{k-1}\| + \eta_w (\sum_{i=0}^{k-2} \|V_e A_e^{k-2-i} B_{e,w}\|$
        $+ \|V_2 M_2 C_2\|) + \eta_v (\|V_2 M_2 T_2\| + \|V_e A_e^{k-2} B_{e,v1}\|$
        $+ \|V_e B_{e,v2} + (V_1 - V_2 M_2 C_2 G_1) M_1 T_1\|$
        $+ \sum_{i=1}^{k-2} \|V_e A_e^{k-2-i} (B_{e,v1} + A_e B_{e,v2})\|)$;
    $\hat{D}_{k-1} = \{d \in \mathbb{R}^l : \|d - \hat{d}_{k-1}\| \leq \delta_{k-1}^d\}$;
  ▷ Time update
    $\hat{x}_{k|k}^\star = \hat{x}_{k|k-1} + G_2 \hat{d}_{2,k-1}$;
  ▷ Measurement update
    $\hat{x}_{k|k} = \hat{x}_{k|k}^\star + \tilde{L}(z_{2,k} - C_2 \hat{x}_{k|k}^\star - D_2 u_k)$;
    $\delta_k^x = \delta_0^x \|A_e^k\| + \eta_w \sum_{i=0}^{k-1} \|A_e^i B_{e,w}\| + \eta_v (\|B_{e,v2}\|$
        $+ \|A_e^{k-1} B_{e,v1}\| + \sum_{i=0}^{k-2} \|A_e^i (B_{e,v1} + A_e B_{e,v2})\|)$;
    $\hat{X}_k = \{x \in \mathbb{R}^n : \|x - \hat{x}_{k|k}\| \leq \delta_k^x\}$;
  ▷ Estimation of $d_{1,k}$
    $\hat{d}_{1,k} = M_1(z_{1,k} - C_1 \hat{x}_{k|k} - D_1 u_k)$;
end for
```

Assumption (A.3) in Theorem 2. Then, an $\mathcal{H}_\infty$-observer with cost γ exist if $U_\infty \succ 0$ and $\breve{A}$ is asymptotically stable (with U_∞ and $\breve{A}$ as defined in Theorem 2).

V. Application to Attack-Resilient Estimation

In this section, we show that the proposed set-valued observer can be applied for attack-resilient estimation of state and attack signals when a system is subject to false data injection attacks on both the actuator and sensor signals. Note that we are not considering *sparse* false data injection attacks, which is a subject of ongoing research. But rather, we assume that the attack locations are known (i.e., the G and H matrices that encode the attack locations are given) while the attack magnitudes d_k at any time k are unknown.

As previously discussed, the false data injection attack magnitudes d_k on the actuator and sensor signals are adversarial and strategic. Hence, we ought not make any assumptions on the attack model (random or deterministic) because a strategic adversary could otherwise simply select a different attack model than is assumed. This 'non-assumption' aligns perfectly with the unknown input signal model in Section II.

Therefore, for any linear time-invariant bounded-error models of a cyber-physical system, the system description in (1) can capture false data injection attacks on the actuator and sensor signals on the system without any limitations. Moreover, the fixed-order simultaneous input and state set-valued observer proposed in Section IV is capable not only of reliably estimating the set of all compatible states, $\hat{X}_k$, but also of identifying the attack signals via the estimation of the set of all compatible unknown inputs, $\hat{D}_{k-1}$.

A. Implications on Attack-Resilient Estimation

Having established that the proposed set-valued observer is applicable to attack-resilient estimation, we now discuss the implication of the relationship between observer stability and strong detectability on resilient state estimation and attack identification. First, we introduce the following definitions:

Definition 3 (Resilient Set of Compatible States)**.** *We say that the estimated set of compatible states is* resilient*, if for any initial state $x_0 \in \mathbb{R}^n$ and signal attack sequence $\{d_j\}_{j \in \mathbb{N}}$ in $\mathbb{R}^p$, the true state x_k is contained in the set estimates $\hat{X}_k^\star$ and $\hat{X}_k$, and these sets remain bounded for all k.*

Definition 4 (Data Injection Attack Identification)**.** *A false data injection attack is* identified*, if for any initial state $x_0 \in \mathbb{R}^n$ and signal attack sequence $\{d_j\}_{j \in \mathbb{N}}$ in $\mathbb{R}^p$, the true attack signal d_{k-1} is contained in the set estimate $\hat{D}_{k-1}$ and this set remains bounded for all k.*

Definition 5 (Correctable false data injection attacks[c])**.** *We say that false data injection attacks on p actuators and sensors are correctable, if for any initial state $x_0 \in \mathbb{R}^n$ and signal attack sequence $\{d_j\}_{j \in \mathbb{N}}$ in $\mathbb{R}^p$, we have a set-valued observer to compute the resilient set of compatible states and to identify the false data injection attack signals.*

Based on these definitions as well as the observer design in Section IV, we have the following conclusions:

Proposition 1 (Resilient Set-Valued State Estimation)**.** *A resilient set of compatible sets can be obtained if and only if the system* (1) *is strongly detectable and Lemma 1 holds.*

Proposition 2 (Attack Identification)**.** *A false data injection attack is* identified *if and only if the system* (1) *is strongly detectable and Lemma 1 holds.*

Note that strong detectability is a system property that is independent of the observer design. Hence, the necessity of strong detectability can serve as a guide to determine and recommend which actuators or sensors need to be safeguarded to guarantee resilient estimation as a preventative attack mitigation method (cf. Section VI-B for an example).

Moreover, we can derive an upper bound on the maximum number of false data injection attacks that can be asymptotically corrected based on strong detectability.

Theorem 5. *The maximum number of correctable actuators and sensors signal attacks, p^*, for system* (1) *is equal to the number of sensors, l, i.e., $p^* \leq l$ (upper bound is achievable)[d].*

Proof. The theorem follows immediately as an implication of Theorem 3 and [27, Theorem 1], [28, Theorem 4.3]. ■

[c] This definition is distinct from [25, Definition 1] that is defined for *exact finite-time point estimation* (after n steps) and requires strong observability [25]. Instead, it is related to *boundedness* in *infinite time*, similar to *infinite-time point estimation* that only requires strong detectability [27], [28].

[d] By contrast, the *stronger* requirement of strong observability in [25] (implies strong detectability [14]) leads to a maximum of $p^* \leq \lceil \frac{l}{2} - 1 \rceil$.

VI. Simulation Examples

A. Benchmark System

In this example, we consider a system that has been used as a benchmark for many state and input filters (e.g., [14]):

$$A = \begin{bmatrix} 0.5 & 2 & 0 & 0 & 0 \\ 0 & 0.2 & 1 & 0 & 1 \\ 0 & 0 & 0.3 & 0 & 1 \\ 0 & 0 & 0 & 0.7 & 1 \\ 0 & 0 & 0 & 0 & 0.1 \end{bmatrix}; G = \begin{bmatrix} 1 & 0 & -0.3 \\ 1 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix}; H = \begin{bmatrix} 0 & 0 & 1 \\ 0 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix};$$

$$B = 0_{5\times 1}; C = I_5; D = 0_{5\times 1}.$$

The unknown inputs used in this example are as given in Fig. 1, while the initial state estimate and noise signals (drawn uniformly) have bounds $\delta_x = 0.5$, $\eta_w = 0.02$ and $\eta_v = 10^{-4}$. The invariant zeros of the system matrix $\mathcal{R}_S(z)$ are $\{0.3, 0.8\}$. Thus, this system is strongly detectable.

We observe from Fig. 1 that the proposed algorithm is able to find set-valued estimates of the states and unknown inputs. The actual estimation errors are also within the predicted upper bounds (cf. Fig. 2), which converges to a steady-state as established in Section IV-A. Moreover, with the gain $\tilde{L}$ of the $\mathcal{H}_\infty$ observer, the eigenvalues of $A_e^\star = \overline{A}(I - \tilde{L}C_2)$ and $A_e = (I - \tilde{L}C_2)\overline{A}$ are $\{0, 0.107, 0.3, 0.407, 0.8\}$. Hence, as is predicted in Lemma 2, all invariant zeros of the system are eigenvalues of the set-valued observer.

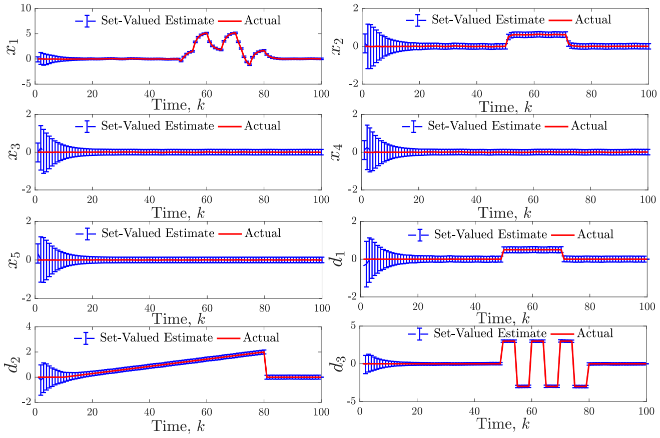

Fig. 1: Actual states x_1, x_2, x_3, x_4, x_5 and their estimates, as well as unknown inputs d_1, d_2 and d_3 and their estimates.

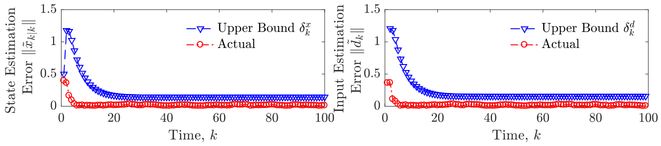

Fig. 2: Actual estimation errors and radii of set-valued estimates of states, $\|\tilde{x}_{k|k}\|$, δ_k^x, and unknown inputs, $\|\tilde{d}_k\|$, δ_k^d.

B. Attack-Resilient Estimation

We now demonstrate the effectiveness of our attack-resilient $\mathcal{H}_\infty$ observer using an IEEE 14-bus system [33] that is subject to data injection attacks. The system consists of 5 synchronous generators and 14 buses, with secure phasor measurement units (PMUs) being installed on the buses depicted in Fig. 3. It is represented by $n = 10$ states comprising the rotor angles and frequencies of each generator. The dynamics of the system can be represented by an uncertain LTI model [23], [25] that is discretized with a sampling interval of $dT = 0.05s$ to obtain the model in (1), where $p = 35$ sensors are deployed to measure the real power injections at every bus, the real power flows along every branch and the rotor angle of generator 1. The initial state estimate and noise signals (drawn uniformly) have bounds $\delta_x = 1$, $\eta_w = 0.03$ and $\eta_v = 0.03$.

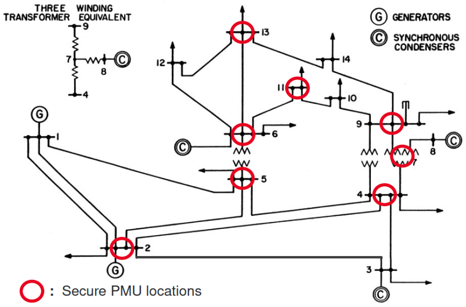

Fig. 3: IEEE 14-bus electric power system [33].

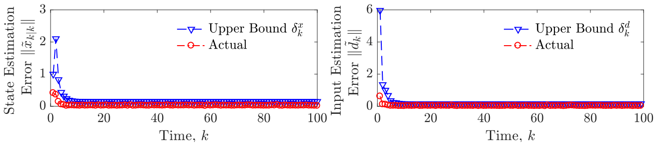

Fig. 4: Resilient estimation errors and radii of set-valued estimates of states, $\|\tilde{x}_{k|k}\|$, δ_k^x, and attack signals, $\|\tilde{d}_k\|$, δ_k^d.

In this example, we assume that all unsecured PMU measurements are attacked (sensor attacks with known locations). Nevertheless, we observe from Fig. 4 that the proposed algorithm is able to find set-valued state estimates that contain the true state, as well as to identify a bounded set that contains the actual attack signals. Individual state and attack signals as in Fig. 1 can also be estimated but are omitted for brevity. Moreover, performing strong detectability tests (necessary condition by Propositions 1 and 2) for various attack locations, i.e., G and H, we found that false data injection attacks on the sensors are correctable if at least one sensor from among sensors 10, 14 and 15 is protected.

VII. Conclusion

We presented an optimal fixed-order set-valued observer that simultaneously computes the set of compatible states and unknown inputs for linear discrete-time bounded-error systems in the minimum $\mathcal{H}_\infty$-norm sense. Necessary and sufficient conditions for the stability of the observer and its connection to strong detectability were also derived. Then, we showed that the proposed set-valued observer is applicable for attack-resilient estimation of state and attack signals when cyber-physical systems are subject to malicious false data injection attacks on both actuator and sensor signals, as well as discussed the implication of strong detectability on resilient state estimation and attack identification.

References

[1] A.A. Cárdenas, S. Amin, and S. Sastry. Research challenges for the security of control systems. In *Conference on Hot Topics in Security*, pages 6:1–6:6, 2008.

[2] J.P. Farwell and R. Rohozinski. Stuxnet and the future of cyber war. *Survival*, 53(1):23–40, 2011.

[3] K. Zetter. Inside the cunning, unprecedented hack of Ukraine's power grid. Wired Magazine, 2016.

[4] P.K. Kitanidis. Unbiased minimum-variance linear state estimation. *Automatica*, 23(6):775–778, November 1987.

[5] R. Patton, R. Clark, and P.M. Frank. *Fault diagnosis in dynamic systems: theory and applications*. Prentice Hall, 1989.

[6] G. De Nicolao, G. Sparacino, and C. Cobelli. Nonparametric input estimation in physiological systems: Problems, methods, and case studies. *Automatica*, 33(5):851–870, 1997.

[7] M. Corless and J. Tu. State and input estimation for a class of uncertain systems. *Automatica*, 34(6):757–764, 1998.

[8] C.P. Tan and C. Edwards. Sliding mode observers for detection and reconstruction of sensor faults. *Automatica*, 38(10):1815–1821, 2002.

[9] K. Kalsi, J. Lian, S. Hui, and S.H. Zak. Sliding-mode observers for systems with unknown inputs: A high-gain approach. *Automatica*, 46(2):347–353, 2010.

[10] H. Fang, Y. Shi, and J. Yi. On stable simultaneous input and state estimation for discrete-time linear systems. *International Journal of Adaptive Control and Signal Processing*, 25(8):671–686, 2011.

[11] S. Gillijns and B. De Moor. Unbiased minimum-variance input and state estimation for linear discrete-time systems. *Automatica*, 43(1):111–116, January 2007.

[12] S. Gillijns and B. De Moor. Unbiased minimum-variance input and state estimation for linear discrete-time systems with direct feedthrough. *Automatica*, 43(5):934–937, 2007.

[13] S.Z. Yong, M. Zhu, and E. Frazzoli. On strong detectability and simultaneous input and state estimation with a delay. In *IEEE Conference on Decision and Control (CDC)*, pages 468–475, 2015.

[14] S.Z. Yong, M. Zhu, and E. Frazzoli. A unified filter for simultaneous input and state estimation of linear discrete-time stochastic systems. *Automatica*, 63:321–329, 2016.

[15] I. Yaesh and U. Shaked. A transfer function approach to the problems of discrete-time systems: $\mathcal{H}_\infty$-optimal linear control and filtering. *IEEE Transactions on Automatic Control*, 36(11):1264–1271, 1991.

[16] P.P. Khargonekar and M.A. Rotea. Mixed $\mathcal{H}_2/\mathcal{H}_\infty$ filtering. In *IEEE Conference on Decision and Control*, pages 2299–2304, 1992.

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

[18] F. Blanchini and M. Sznaier. A convex optimization approach to synthesizing bounded complexity ℓ^∞ filters. *IEEE Transactions on Automatic Control*, 57(1):216–221, 2012.

[19] M. Milanese and A. Vicino. Optimal estimation theory for dynamic systems with set membership uncertainty: An overview. *Automatica*, 27(6):997–1009, 1991.

[20] M.A. Dahleh and I.J. Diaz-Bobillo. *Control of uncertain systems: a linear programming approach*. Prentice-Hall, Inc., 1994.

[21] J.S. Shamma and K. Tu. Set-valued observers and optimal disturbance rejection. *IEEE Trans. on Automatic Control*, 44(2):253–264, 1999.

[22] J. Chen and C.M. Lagoa. Observer design for a class of switched systems. In *IEEE Conference on Decision and Control European Control Conference*, pages 2945–2950, 2005.

[23] F. Pasqualetti, F. Dörfler, and F. Bullo. Attack detection and identification in cyber-physical systems. *IEEE Transactions on Automatic Control*, 58(11):2715–2729, November 2013.

[24] M.S. Chong, M. Wakaiki, and J.P. Hespanha. Observability of linear systems under adversarial attacks. In *IEEE American Control Conference (ACC)*, pages 2439–2444, 2015.

[25] H. Fawzi, P. Tabuada, and S. Diggavi. Secure estimation and control for cyber-physical systems under adversarial attacks. *IEEE Transactions on Automatic Control*, 59(6):1454–1467, June 2014.

[26] Y. Shoukry, P. Nuzzo, A. Puggelli, A.L. Sangiovanni-Vincentelli, S.A. Seshia, M. Srivastava, and P. Tabuada. Imhotep-SMT: A satisfiability modulo theory solver for secure state estimation. In *13th International Workshop on Satisfiability Modulo Theories (SMT)*, 2015.

[27] S.Z. Yong, M. Zhu, and E. Frazzoli. Resilient state estimation against switching attacks on stochastic cyber-physical systems. In *IEEE Conference on Decision and Control*, pages 5162–5169, 2015.

[28] S.Z. Yong, M. Zhu, and E. Frazzoli. Switching and data injection attacks on stochastic cyber-physical systems: Modeling, resilient estimation and attack mitigation. *arXiv:1707.07112*, 2017.

[29] H. Kim, P. Guo, M. Zhu, and P. Liu. Attack-resilient estimation of switched nonlinear cyber-physical systems. *arXiv:1609.03600*, 2016.

[30] M. Pajic, P. Tabuada, I. Lee, and G.J. Pappas. Attack-resilient state estimation in the presence of noise. In *IEEE Conference on Decision and Control (CDC)*, pages 5827–5832, 2015.

[31] Y. Nakahira and Y. Mo. Dynamic state estimation in the presence of compromised sensory data. In *IEEE Conference on Decision and Control (CDC)*, pages 5808–5813, 2015.

[32] S.Z. Yong, M.Q. Foo, and E. Frazzoli. Robust and resilient estimation for cyber-physical systems under adversarial attacks. In *IEEE American Control Conference (ACC)*, pages 308–315, 2016.

[33] J. Kim and L. Tong. On phasor measurement unit placement against state and topology attacks. In *IEEE International Conference on Smart Grid Communications*, pages 396–401, 2013.

[34] L. Xie, C. E. de Souza, and M. Fu. $\mathcal{H}_\infty$ estimation for discrete-time linear uncertain systems. *International Journal of Robust and Nonlinear Control*, 1(2):111–123, 1991.

[35] C. De Souza, M. Gevers, and G. Goodwin. Riccati equations in optimal filtering of nonstabilizable systems having singular state transition matrices. *IEEE Trans. on Automatic Control*, 31(9):831–838, 1986.

[36] R.A. Horn and C.R. Johnson. *Matrix analysis*. Cambridge University Press, 2012.

Appendix: Proofs

A. *Proof of Lemma 1*

We observe from (5), (9) and (10) that

$$\hat{d}_{1,k} = M_1(C_1\tilde{x}_{k|k} + \Sigma d_{1,k} + v_{1,k}), \tag{21}$$

$$\begin{aligned}\hat{d}_{2,k-1} &= M_2(C_2(A_{k-1}\tilde{x}_{k-1|k-1} + G_1\tilde{d}_{1,k-1} + w_{k-1}) \\ &\quad + v_{2,k} + C_2G_2d_{2,k-1}).\end{aligned} \tag{22}$$

Then, from (12) and (13), as well as (4) and (14), the propagated and updated state estimate errors are found as

$$\tilde{x}^\star_{k|k} = \ A\tilde{x}_{k-1|k-1} + G_1\tilde{d}_{1,k-1} + G_2\tilde{d}_{2,k-1} + w_{k-1}, \tag{23}$$

$$\tilde{x}_{k|k} = (I - LC)\tilde{x}^\star_{k|k} - LU_1\Sigma d_{1,k} - Lv_k. \tag{24}$$

We show by induction that the estimates $\hat{d}_k$, $\hat{x}_{k|k}$ and $\hat{x}^\star_{k|k}$ are bounded, assuming at the moment that their dynamics are stable (which we will show to hold in Theorem 3). For the base case, since $\hat{x}_{0|0}$, $\hat{x}^\star_{0|0}$ and the noise signals are bounded by assumption, from (21) and (22), we find that $\hat{d}_{1,0}$ and $\hat{d}_{2,0}$ are bounded, only if $M_1\Sigma = I$ and $M_2C_2G_2 = I$, since we assumed that $d_{1,0}$ and $d_{2,0}$ can be unbounded. Hence, $\hat{d}_0$ is bounded. In the inductive step, we assume that $\tilde{x}_{k-1|k-1}$ and $\tilde{x}^\star_{k-1|k-1}$ are bounded. Then, the input estimates $\tilde{d}_{1,k-1}$ and $\tilde{d}_{2,k-1}$ are bounded, only if $M_1\Sigma = I$ and $M_2C_2G_2 = I$, since the unknown inputs $d_{1,k}$ and $d_{2,k-1}$ can be unbounded although the noise signals are bounded. Similarly, from (24) with a bounded measurement noise, we need the constraint $LU_1 = 0$ such that $\tilde{x}_{k|k}$ remains bounded. Thus, by induction, $\tilde{x}^\star_{k|k}$ and $\tilde{x}_{k|k}$ are bounded for all k. Since we require $M_2C_2G_2 = I$ for the existence of bounded input estimates, it follows that $\text{rk}(C_2G_2) = p - p_H$ is a necessary condition. Furthermore, $L = LUU^\top = LU_2U_2^\top = \tilde{L}U_2^\top$ since $LU_1 = 0$. ■

B. *Proof of Theorem 2*

First, we reduce the system with unknown inputs to one without unknown inputs. From (14) and (5), we have $\tilde{x}_{k|k} = \tilde{x}^\star_{k|k} - \tilde{L}(C_2\tilde{x}^\star_{k|k} + v_{2,k})$. Then, substituting (21) with $M_1 = \Sigma^{-1}$ into (23) and the above, and rearranging, we obtain

$$\begin{aligned}\tilde{x}_{k|k} &= \overline{A}\tilde{x}_{k-1|k-1} + \overline{w}_{k-1} - \tilde{L}_k(C_2\overline{A}\tilde{x}_{k-1|k-1} \\ &\quad + C_2\overline{w}_{k-1} + v_{2,k}),\end{aligned} \tag{25}$$

with the state matrix $\overline{A} \triangleq (I - G_2M_2C_2)\hat{A}$ and noise signal $\overline{w}_{k-1} \triangleq -(I - G_2M_2C_2)(G_1M_1v_{1,k-1} - w_{k-1}) -$

$G_2 M_2 v_{2,k}$. As it turns out, the *updated* state estimate error dynamics above is the same for an *a posteriori* $\mathcal{H}_\infty$ filter [15, Eq. (5.2)] for a linear system without unknown inputs

$$x^{e,\star}_{k+1} = \overline{A} x^{e,\star}_k + \overline{w}_k, \quad y^{e,\star}_k = C_2 x^{e,\star}_k + v_{2,k}.$$

Since the objective for both systems is the same, i.e., to obtain the observer gain $\tilde{L}$ with an *a posteriori* $\mathcal{H}_\infty$ filter, they are equivalent systems from the perspective of estimation. Furthermore, from the analysis in [15, Eq. (5.3)], it can be seen that an equivalent observer gain $\tilde{L}$ can be obtained with the *a priori* $\mathcal{H}_\infty$ filter for the following system

$$x^e_{k+1} = A_\infty x^e_k + B_\infty w^e_k, \quad y^e_k = C_\infty x^e_k + D_\infty w^e_k,$$

with $A_\infty \triangleq \overline{A}$, $B_\infty \triangleq \begin{bmatrix}\Phi W & -\Phi G_1 M_1 & -\sqrt{2} G_2 M_2 & 0\end{bmatrix}$, $C_\infty \triangleq C_2 \overline{A}$, $D_\infty \triangleq \begin{bmatrix}0 & 0 & 0 & \sqrt{2} I\end{bmatrix} + C_2 B_\infty$ and $w^e_k \triangleq \begin{bmatrix}w_k^\top & v_{1,k}^\top & \frac{1}{\sqrt{2}} v_{2,k+1}^\top & \frac{1}{\sqrt{2}} v_{2,k}^\top\end{bmatrix}^\top$, where the factors $\frac{1}{\sqrt{2}}$ are included such that $\lim_{k\to\infty} \sum_{i=0}^k \|w^e_i\| = \lim_{k\to\infty} \sum_{i=0}^k \| \begin{bmatrix} w_i^\top & v_i^\top \end{bmatrix}^\top \|$. The design of the observer gain $\tilde{L}$ is then a direct application of the *a priori* $\mathcal{H}_\infty$ filter (e.g., [34, Theorem 2.2]). Next, we consider the following identity $\forall \omega \in [0, 2\pi]$ (equivalently, $\forall z \in \mathcal{C}, |z| = 1$)

$$\begin{aligned}
&\mathrm{rk} \begin{bmatrix} A_\infty - e^{j\omega} I & B_\infty \\ C_\infty & D_\infty \end{bmatrix} = \mathrm{rk} \begin{bmatrix} A_\infty - zI & B_\infty \\ C_\infty & D_\infty \end{bmatrix} \\
&= \mathrm{rk} \begin{bmatrix} \overline{A} - zI & \Phi W & -\Phi G_1 M_1 & -\sqrt{2} G_2 M_2 & 0 \\ C_2 \overline{A} & 0 & 0 & 0 & \sqrt{2} I \end{bmatrix} \\
&= \mathrm{rk} \begin{bmatrix} \overline{A} - zI & \Phi W & -\Phi G_1 M_1 & -\sqrt{2} G_2 M_2 & 0 \\ 0 & 0 & 0 & 0 & \sqrt{2} I \end{bmatrix} \\
&= \mathrm{rk} \begin{bmatrix} \overline{A} - zI & \Phi W & -\Phi G_1 M_1 & -\sqrt{2} G_2 M_2 \end{bmatrix} + l - p_H,
\end{aligned} \tag{26}$$

where the third equality is obtained by subtracting C_2 times the first row from the second row, and then subtracting $\frac{\sqrt{2}}{z}$ times the first column to the final column. Hence, Assumption (A.3) implies that (A_∞, B_∞) is stabilizable on the unit circle, which along with the Assumption (A.1) are are necessary and sufficient conditions for the stability of A_e [35] and by extension, $A^\star_e$, since A_e and $A^\star_e$ have the same eigenvalues [36, Theorem 1.3.22]. ■

C. Proof of Error Bounds/Set Radii (18), (19) *and* (20)

From (21)–(24), we find the state estimation errors as

$$\begin{aligned}
\tilde{x}^\star_{k|k} &= \overline{A} \tilde{x}_{k-1|k-1} + B^\star_{e,w} w_{k-1} + B^\star_{e,v1} v_{k-1} + B^\star_{e,v2} v_k, \\
\tilde{x}_{k|k} &= (I - \tilde{L} C_2) \tilde{x}^\star_{k|k} - \tilde{L} v_{2,k},
\end{aligned} \tag{27}$$

where $\hat{A} \triangleq A - G_1 M_1 C_1$, $\Phi \triangleq I - G_2 M_2 C_2$, $\overline{A} \triangleq \Phi \hat{A}$, $B^\star_{e,w} \triangleq \Phi W$, $B^\star_{e,v1} \triangleq -\Phi G_1 M_1 T_1$ and $B^\star_{e,v2} \triangleq -G_2 M_2 T_2$. Then, via substitution, we obtain the estimation errors in terms of the initial state error $\tilde{x}_{0|0}$ and noise signals:

$$\begin{aligned}
\tilde{x}_{k|k} &= A_e^k \tilde{x}_{0|0} + A_e^{k-1} \begin{bmatrix} B_{e,w} & B_{e,v1} \end{bmatrix} \vec{w}_0 + B_{e,v2} v_k \\
&\quad + \textstyle\sum_{i=1}^{k-1} A_e^{k-1-i} \begin{bmatrix} B_{e,w} & B_{e,v1} + A_e B_{e,v2} \end{bmatrix} \vec{w}_i, \\
\tilde{x}^\star_{k|k} &= \overline{A} A_e^{k-1} \tilde{x}_{0|0} + \overline{A} A_e^{k-2} \begin{bmatrix} B_{e,w} & B_{e,v1} \end{bmatrix} \vec{w}_0 \\
&\quad + B^\star_{e,w} w_{k-1} + (B^\star_{e,v1} + \overline{A} B_{e,v2}) v_{k-1} + B^\star_{e,v2} v_k \\
&\quad + \textstyle\sum_{i=1}^{k-2} \overline{A} A_e^{k-1-i} \begin{bmatrix} B_{e,w} & B_{e,v1} + A_e B_{e,v2} \end{bmatrix} \vec{w}_i,
\end{aligned}$$

with $\vec{w}_i = \begin{bmatrix} w_i^\top & v_i^\top \end{bmatrix}^\top$, $A_e \triangleq (I - \tilde{L} C_2) \overline{A}$, $B_{e,w} \triangleq (I - \tilde{L} C_2) B^\star_{e,w}$, $B_{e,v1} \triangleq (I - \tilde{L} C_2) B^\star_{e,v1}$ and $B_{e,v2} \triangleq (I - \tilde{L} C_2) B^\star_{e,v2} - \tilde{L} T_2$. Moreover, we find the unknown input estimation error as

$$\begin{aligned}
\tilde{d}_{k-1} &= V_1 \tilde{d}_{1,k-1} + V_2 \tilde{d}_{2,k-1} \\
&= -V_e A_e^{k-1} \tilde{x}_0 - V_e A_e^{k-2} \begin{bmatrix} B_{e,w} & B_{e,v1} \end{bmatrix} \vec{w}_0 \\
&\quad - V_2 M_2 C_2 w_{k-1} - V_2 M_2 T_2 v_k \\
&\quad + (V_e B_{e,v1} + (V_1 - V_2 M_2 C_2 G_1) M_1 T_1) v_{k-1} \\
&\quad + \textstyle\sum_{i=1}^{k-2} V_e A_e^{k-1-i} \begin{bmatrix} B_{e,w} & B_{e,v1} + A_e B_{e,v2} \end{bmatrix} \vec{w}_i,
\end{aligned}$$

with $V_e \triangleq V_1 M_1 C_1 + V_2 M_2 C_2 \hat{A}$. Finally, the error bounds (18), (19), (20) can be found using triangle inequalities. ■

D. Proof of Lemma 2

Let z be any invariant zero of system (1), i.e., when $\overline{\mathcal{R}}_S(z)$ drops rank (cf. Theorem 1). Then, there exists $\begin{bmatrix} \nu^\top & \mu^\top \end{bmatrix}^\top \neq 0$ such that

$$(zI - \overline{A}) \nu - G_2 \mu = 0, \quad C_2 \nu = 0.$$

Premultiplying the former with $(I - G_2 M_2 C_2)$ and applying the latter as well as the fact that $M_2 C_2 G_2 = I$, we have

$$\begin{aligned}
&(I - G_2 M_2 C_2)(zI - \hat{A}) \nu + (I - G_2 M_2 C_2) G_2 \mu = 0 \\
&= (zI - \overline{A}) \nu = (zI - \overline{A}) \nu + \overline{A} \tilde{L} C_2 \nu = (zI - \overline{A}(I - \tilde{L} C_2)) \nu.
\end{aligned}$$

If $\nu = 0$, then $G_2 \mu = 0$ and, in turn, $\mu = 0$, which is a contradiction. Hence, $\nu \neq 0$ and the determinant of $zI - (\overline{A} - \overline{A} \tilde{L} C_2)$ is zero, i.e., any invariant zero of $\mathcal{R}_S(z)$ is also an eigenvalue of the propagated state estimation error dynamics of $\tilde{x}^\star_{k|k}$, which can be found from (27) to be

$$\begin{aligned}
\tilde{x}^\star_{k|k} &= \overline{A}(I - \tilde{L} C_2) \tilde{x}^\star_{k-1|k-1} + B^\star_{e,w} w_{k-1} \\
&\quad + (B^\star_{e,v1} - \overline{A} \tilde{L} T_2) v_{k-1} + B^\star_{e,v2} v_k,
\end{aligned} \tag{28}$$

and by extension, the state matrix A_e in (25), since A_e and $A^\star_e$ have the same eigenvalues [36, Theorem 1.3.22]. ■

E. Proof of Theorem 3

Let Lemma 1 hold. We will prove that strong detectability implies that state estimation errors with $A^\star_e = \overline{A}(I - \tilde{L} C_2)$ and $A_e = (I - \tilde{L} C_2) \overline{A}$ (they have same eigenvalues [36]) are asymptotically stable and bounded, and vice versa.
($\Rightarrow$): By Theorem 1, strong detectability implies that $(\overline{A}, C_2 \overline{A})$ is detectable. Hence, there exists $\tilde{L}$ such that A_e, and by extension $A^\star_e$, are asymptotically stable, which in turn implies bounded-input, bounded-state stability of (25), (28).
($\Leftarrow$): We will show this by contraposition. By Lemma 2, we know that (1) being not strongly detectable implies that $A^\star_e$, and by extension A_e, are unstable. This implies that $\lim_{k\to\infty} \|A_e^k\| = \lim_{k\to\infty} \|(A^\star_e)^k\| = \infty$. Thus, from (19) and (20), the estimation errors are unbounded for any $\delta^x_0 \neq 0$.

F. Proof of Theorem 4

By Theorem 1, strong detectability implies that Assumption (A.1) holds. Moreover, $D_\infty D_\infty^\top = C_2(G_2 M_2 M_2^\top G_2^\top + \Phi(W W^\top + G_1 M_1 M_1^\top G_1^\top) \Phi^\top) C_2^\top + 2I \succ 0$ readily satisfies Assumption (A.2). Finally, for Assumption (A.3) to hold, we require that $\mathrm{rk} \begin{bmatrix} \overline{A} - zI & \Phi W & -\Phi G_1 M_1 & -\sqrt{2} G_2 M_2 \end{bmatrix} = n$ (cf. (26)), which is satisfied if $(\overline{A}, G_2)$ or (A_∞, B_∞) is stabilizable on the unit circle. Moreover, if $p = l$, from Theorem 1, strong detectability satisfies the rank condition above and thus Assumption (A.3). ■