Author/s:
Yang, T;Murguia, C;Kuijper, M;Nesic, D

Title:
A Robust Circle-criterion Observer-based Estimator for Discrete-time Nonlinear Systems in the Presence of Sensor Attacks

Date:
2018

Citation:
Yang, T., Murguia, C., Kuijper, M. & Nesic, D. (2018). A Robust Circle-criterion Observer-based Estimator for Discrete-time Nonlinear Systems in the Presence of Sensor Attacks. Proceedings of the 57th IEEE Conference on Decision and Control, 2018-December, pp.571-576. IEEE. https://doi.org/10.1109/CDC.2018.8619724.

Persistent Link:
https://hdl.handle.net/11343/251856

A Robust Circle-criterion Observer-based Estimator for Discrete-time Nonlinear Systems in the Presence of Sensor Attacks

Tianci Yang, Carlos Murguia, Margreta Kuijper, Dragan Nešić

***Abstract*— We address the problem of robust state estimation and attack isolation for a class of discrete-time nonlinear systems with positive-slope nonlinearities under (potentially unbounded) sensor attacks and measurement noise. We consider the case when a subset of sensors is subject to additive false data injection attacks. Using a bank of circle-criterion observers, each observer leading to an Input-to-State Stable (ISS) estimation error, we propose a estimator that provides robust estimates of the system state in spite of sensor attacks and measurement noise; and an algorithm for detecting and isolating sensor attacks. Our results make use of the ISS property of the observers to check whether the trajectories of observers are consistent with the attack-free trajectories of the system. Simulations results are presented to illustrate the performance of the results.**

I. Introduction

Networked Control Systems (NCSs) have emerged as a technology that combines control, communication, and computation and offers the necessary flexibility to meet new demands in distributed and large scale systems. Recently, security of NCSs has become an important issue as wireless communication networks might serve as new access points for attackers to adversely affect the operation of the system dynamics. Cyber-physical attacks on NCSs have caused substantial damage to a number of physical processes. One of the most well-known examples is the attack on Maroochy Shire Councils sewage control system in Queensland, Australia that happened in January 2000. The attacker hacked into the controllers that activate and deactivate valves and caused flooding of the grounds of a hotel, a park, and a river with a million liters of sewage. Another incident is the very recent SuxNet virus that targeted Siemens supervisory control and data acquisition systems which are used in many industrial processes. It follows that strategic mechanisms to identify and deal with attacks on NCSs are strongly needed.

In [1]-[22], a range of topics related to security of control systems have been discussed. In general, they provide analysis tools for quantifying the performance degradation induced by different classes of attacks; and propose reaction strategies to identify and counter their effect on the system dynamics. Most of the existing work, however, has considered control systems with linear dynamics, although in most engineering applications the dynamics of the plants being monitored and controlled is highly nonlinear. There are some results addressing the nonlinear case though. In [23], exploiting sensor redundancy, the authors address the problem of sensor attack detection and state estimation for uniformly observable continuous-time nonlinear systems. Similarly, in [24], the authors provide an algorithm for isolating sensor attacks for a class of discrete-time nonlinear systems with bounded measurement noise.

In this manuscript, we consider the case when the system has p sensors, all of which are subject to measurement noise and up to $q < p/2$ of them are attacked. Following the results in [25] for linear systems, using a bank of circle criterion observers [26]-[29], each observer leading to an ISS estimation error, we construct an estimator that provides robust estimates of the system state in spite of sensor attacks. In particular, the proposed estimator leads to estimation errors satisfying an ISS property with respect to measurement noise but independent of attack signals. Next, we propose an algorithm for detecting and isolating false data injection sensor attacks. Our results make use of the ISS property of the observers to check whether the trajectories of observers are consistent with the attack-free trajectories of the system. The main idea behind our results is the following. Each observer in the bank is driven by a different subset of sensors. Thus, without attacks, the observers produce ISS estimation errors with respect to measurement noise only. For every pair of observers in the bank, we compute the largest difference between their estimates. If a pair of observers is driven by a subset of attack-free sensors, then the largest difference between their estimates is also ISS with respect to measurement noise only. However, if there are attacks on some of the sensors, the observers driven by those sensors might produce larger differences than the attack-free ones. These ideas work well under the assumption that less than $p/2$ sensors are attacked, i.e, $q < p/2$. To design the observers in the bank, we give an extension to the result in [28] for designing robust discrete-time circle-criterion observers. In particular, we use the incremental multiplier technique introduced in [29] to cast the observer design as the solution of a semidefinite program. We minimize the ISS-gain from the measurement noise to the estimation error.

The paper is organized as follows. In Section II, we present preliminary results needed for the subsequent sections. In Section III, we provide tools for designing optimal robust circle criterion observers in the attack-free case. In Section IV, assuming that a sufficiently small number of sensors are subject to attacks, we propose an estimation scheme using a bank of robust circle criterion observers. In Section V, an algorithm for isolating sensor attacks is given. Finally, in Section VI, we give concluding remarks.

This work was supported by the Australian Research Council under the Discovery Project DP170104099.

The authors are with the Department of Electrical and Electronics Engineering, the University of Melbourne, Australia. tianciy@student.unimelb.edu.au

II. Preliminaries

A. Notation

We denote the set of real numbers by $\mathbb{R}$, the set of natural numbers by $\mathbb{N}$, the set of integers by $\mathbb{Z}$, and $\mathbb{R}^{n\times m}$ the set of $n\times m$ matrices for any $m,n\in\mathbb{N}$. For any vector $v\in\mathbb{R}^n$, we denote v_J the stacking of all v_i, $i\in J$, $J\subset\{1,\dots,n\}$, $|v|=\sqrt{v^\top v}$, and $\text{supp}(v)=\{i\in\{1,\dots,n\}\,|v_i\neq 0\}$. For a sequence of vectors $\{v(k)\}_{k=0}^{\infty}$, we denote $v_{[0,k]}$ a sequence of vectors $v(i)$, $i=0,\dots,k$, $||v||_\infty \triangleq \sup_{k\geq 0}|v(k)|$ and $||v||_T \triangleq \sup_{0\leq k\leq T}|v(k)|$. We say a sequence $\{v(k)\}\in l_\infty$ if $||v||_\infty<\infty$. We denote the cardinality of a set S as $\text{card}(S)$. We denote matrix P to be positive definite as $P>0$. The identity matrix is denoted by I. A function $\beta:\mathbb{R}_{\geq 0}\times\mathbb{R}_{\geq 0}\to\mathbb{R}_{\geq 0}$ is said to be of class $exp-KL$ if there exist $c>0$ and $\lambda\in(0,1)$, such that $\beta(s,k)=c\lambda^k s$. The binomial coefficient is denoted as $\binom{a}{b}$, where a,b are nonnegative integers. We denote a variable m uniformly distributed in the interval (a,b) as $m\sim\mathcal{U}(a,b)$.

B. Preliminary Definitions

Definition 1 [29] *(Incremental Multiplier Matrix). Let $f:\mathbb{R}^{n_q}\to\mathbb{R}^{n_f}$. A symmetric matrix $M\in\mathbb{R}^{(n_q+n_f)\times(n_q+n_f)}$ is an incremental multiplier matrix for function $f(\cdot)$ if the following incremental quadratic constraint is satisfied for all $q_1,q_2\in\mathbb{R}^{n_q}$:*

$$\begin{bmatrix}\triangle q\\ \triangle f\end{bmatrix}^T M\begin{bmatrix}\triangle q\\ \triangle f\end{bmatrix}\geq 0, \tag{1}$$

where $\triangle q=q_1-q_2$ and $\triangle f=f(q_1)-f(q_2)$.

Definition 2 *(Input-to-State Stability). Consider the discrete time system*

$$e^+=F(e,m), \tag{2}$$

with state $e\in\mathbb{R}^n$ and input $m\in\mathbb{R}^p$, $\{m(k)\}\in l_\infty$. System (2) is said to be Input-to-State Stable (ISS) with linear gain γ and $exp-KL$ function, if there exist $c>0$, $\lambda\in(0,1)$, and $\gamma\geq 0$ such that the following inequality is satisfied:

$$|e(k)|\leq c\lambda^k|e(0)|+\gamma||m||_k, \tag{3}$$

for all $e(0)\in\mathbb{R}^n$, $k\geq 0$, and $\{m(k)\}\in l_\infty$.

III. Robust circle-criterion observer

In [28], using the circle criterion, the authors design observers for discrete-time nonlinear systems with no disturbances. In this section, we give an extension to the result in [28] by considering measurement noise. The design method follows the ideas in [29] where the nonlinearity is characterized by an incremental multiplier matrix. We present an extension to the result in [29] by casting the observer design as a semidefinite program with more degrees of freedom, which might lead to a less conservative ISS gains. We consider discrete-time nonlinear systems of the form:

$$\begin{aligned}x^+ =& Ax+Gf(Hx)+\rho(u,y),\\ y=& Cx+m,\end{aligned} \tag{4}$$

with state $x\in\mathbb{R}^n$, output $y\in\mathbb{R}^{n_y}$, sensor noise $m\in\mathbb{R}^{n_y}$, $\{m(k)\}\in l_\infty$, and matrices $G\in\mathbb{R}^{n\times r}$ and $H\in\mathbb{R}^{r\times n}$. The term $\rho(u,y)$ is a known real-valued vector that depends on the system inputs and outputs. The nonlinearity $f(Hx)$ is an r-dimensional vector where each entry is a function of a linear combination of the states:

$$f_i=f_i\left(\sum_{j=1}^{n}H_{ij}x_j\right),\quad i=1,\dots,r, \tag{5}$$

where H_{ij} denotes the entries of the matrix H.

Assumption 1 *For all $i\in\{1,\dots,r\}$, the following holds*

$$\frac{f_i(v_i)-f_i(w_i)}{v_i-w_i}\geq 0,\ \forall\, v_i,w_i\in\mathbb{R},\ v_i\neq w_i. \tag{6}$$

Consider the circle criterion observer:

$$\hat{x}^+=A\hat{x}+Gf(H\hat{x}+K(C\hat{x}-y))+L(C\hat{x}-y)+\rho(u,y) \tag{7}$$

with observer state $\hat{x}\in\mathbb{R}^n$ (the estimate of x), and observer gain matrices $K\in\mathbb{R}^{r\times n_y}$ and $L\in\mathbb{R}^{n\times n_y}$ to be designed. Define the estimation error $e:=\hat{x}-x$. It follows that the estimation error dynamics is given by the following difference equation:

$$e^+=(A+LC)e-Lm+G\triangle f, \tag{8}$$

where

$$\triangle f:=f(\hat{q})-f(\tilde{q}), \tag{9}$$

$\tilde{q}:=Hx$, $\hat{q}:=H\hat{x}+K(\hat{y}-y)$, $\hat{y}:=C\hat{x}$, and

$$\triangle q:=\hat{q}-\tilde{q}=(H+KC)e-Km. \tag{10}$$

We aim at designing the observer matrices K and L such that the estimation error dynamics is ISS with a linear gain and an $exp-KL$ function with respect to the measurement noise.

Theorem 1 *Consider system (4), for given $c_3\in(0,1)$, if there exist positive definite matrix $P\in\mathbb{R}^{n\times n}$, matrices $Y\in\mathbb{R}^{n\times n_y}$ and matrix $Y_2\in\mathbb{R}^{r\times n_y}$, and scalars $\kappa>0$, $\mu>0$, and $\mu_1>0$ satisfying:*

$$\begin{bmatrix}-P & \star\\ \Xi_{21} & \Xi_{22}\end{bmatrix}+\begin{bmatrix}0 & 0\\ 0 & \Gamma_1^TM\Gamma_1+\Gamma_1^T\Gamma_2+\Gamma_2^T\Gamma_1\end{bmatrix}\leq 0,\qquad \begin{bmatrix}P & I\\ I & \mu I\end{bmatrix}\geq 0, \tag{11}$$

with Ξ_{21} and Ξ_{22} defined as

$$\begin{aligned}\Xi_{21}^T&:=\begin{bmatrix}PA+YC & -Y & PG\end{bmatrix},\\ \Xi_{22}&:=\begin{bmatrix}(c_3-1)P & 0 & 0\\ 0 & -c_3\mu_1 I & 0\\ 0 & 0 & 0\end{bmatrix},\end{aligned} \tag{12}$$

M defined as

$$M:=\kappa\begin{bmatrix}0 & 1\\ 1 & 0\end{bmatrix}, \tag{13}$$

with $\kappa > 0$, which is an incremental multiplier matrix for $f(\cdot)$, and

$$\Gamma_1 := \begin{bmatrix} H & 0 & 0 \\ 0 & 0 & I \end{bmatrix}, \Gamma_2 := \begin{bmatrix} 0 & 0 & 0 \\ Y_2 C & -Y_2 & 0 \end{bmatrix};$$

then, the observer (7) with $L = P^{-1}Y, K = \frac{Y_2}{\kappa}$ has ISS error dynamics with a linear gain $\gamma = \sqrt{\mu\mu_1}$ and an $exp - KL$ function with respect to m.

Proof: See Proposition 1 and Lemma 2 in [30]. ■

Theorem 1 provides a tool for designing the observer (4). If we minimize $\sqrt{\mu\mu_1}$ subject to (11), the obtained observer minimizes the effect of noise on the estimation error. To have a convex objective function, we use $\mu+\mu_1$ instead of $\sqrt{\mu\mu_1}$. Because $(\mu+\mu_1)^2 \geq 4\mu\mu_1$; then, $\gamma = \sqrt{\mu\mu_1} \leq \frac{1}{2}(\mu+\mu_1)$ since μ and μ_1 are positive. Therefore, by minimizing $\mu+\mu_1$, we minimize an upper bound on γ. Since $c_3 \in (0,1)$ (a bounded set), we could perform a grid-search over c_3, i.e., we make a grid in $(0,1)$, for each grid point we minimize $\mu+\mu_1$ subject to (11), and we choose the c_3 that minimizes $\sqrt{\mu\mu_1}$ to construct the observer. In our design method, besides regarding μ_1 as a variable, we also do not assume c_3 is a fixed constant as it is done in [29]. We use the model used in Example 1 in [28] and compare the performance of both observers. All the optimizations are solved using PENLAB [31] in MATLAB.

Example 1 Consider the discrete-time nonlinear system subject to measurement noise:

$$\begin{aligned} x^+ &= \begin{bmatrix} 1 & \delta \\ 0 & 1 \end{bmatrix} x + \begin{bmatrix} \frac{1}{2}\delta\alpha \sin(x_1+x_2) \\ \delta\alpha \sin(x_1+x_2) \end{bmatrix} + \begin{bmatrix} \delta u \\ \delta u \end{bmatrix}, \\ y &= \begin{bmatrix} 3 & 3 & 6 & 1.2 \\ 0.3 & 0.6 & 0.9 & 12 \end{bmatrix}^\top x + m. \end{aligned} \tag{14}$$

We let $\delta = 0.1$ and $\alpha = 1$. Matrices (14) can be written in the form (4) with Assumption 1 being satisfied, see [28] for details. We minimizing $\mu+\mu_1$ subject to (11) and perform a grid search over c_3 to obtain observer matrices K and L. The smallest ISS gain is $\gamma = 0.924$ for $c_3 = 0.9$. Next, we minimize μ subject to the LMIs in [29] with $c_3 = 0.5$, M as in (13), and $H = I$. The corresponding observer leads to $\gamma = 22.4$. We let $m_i \sim \mathcal{U}(-0.5, 0.5)$ for $i \in \{1,\dots,4\}$. The initial conditions are randomly selected from a normal distribution and $\hat{x}(0) = [0,0]^\top$. We depict the performance of these observers in Figures 1-2.

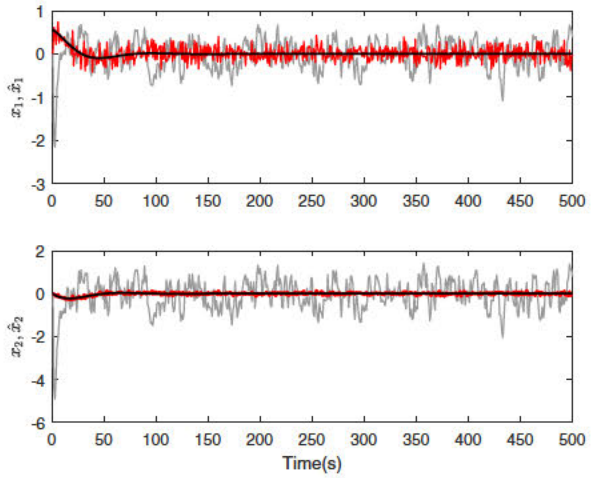

Fig. 1. Estimated states $\hat{x}$ converges to a neighbourhood of the true states x. Legend: Observer obtained via Theorem 1 (red), Observer from [29] (grey), true states (black)

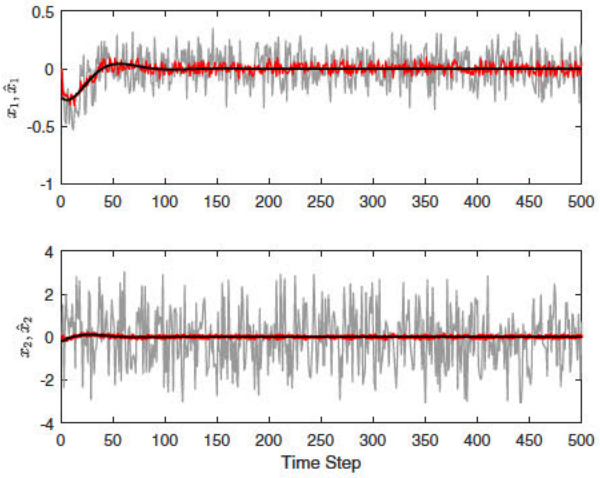

Fig. 2. Estimated states $\hat{x}$ converges to a neighbourhood of the true states x. Legend: Observer obtained via Theorem 1 (red), Observer from [28] (grey), true states (black)

IV. Observer-based estimator under sensor attacks and sensor noise

In this section, we introduce a circle-criterion observer-based estimator for the class systems described in Section III. We assume that a small number of sensors are subject to sensor attacks:

$$\begin{aligned} x^+ &= Ax + Gf(Hx) + \rho(u,y), \\ \tilde{y} &= \tilde{C}x + a + \tilde{m}, \end{aligned} \tag{15}$$

where $\tilde{y} \in \mathbb{R}^p$ is the vector of sensor measurement under attacks, $\tilde{m} \in \mathbb{R}^p$, $\{\tilde{m}(k)\} \in l_\infty$ is the measurement noise, and $a \in \mathbb{R}^p$ is the vector of attacks. If sensor $i \in \{1,\dots,p\}$ is not attacked, then the i-th component of the vector $a(k)$, $a_i(k) = 0, \forall k \geq 0$; otherwise, sensor i is attacked and $a_i(k)$ is arbitrary and possibly unbounded. We denote $W \subseteq \{1,\dots,p\}$ the set of attacked sensors, then we have $\text{supp}(a(k)) = W$ for all $k \geq 0$. We assume the set W is unknown to us. We denote $\tilde{y}\left(k; x(0), a_{[0,k]}, \tilde{m}_{[0,k]}\right)$ as the output of the system at time k when the initial state is $x(0)$ and the outputs are subject to measurement noise $\tilde{m}$ and sensor attacks a. The nonlinearity $f(\cdot)$ satisfies (5) and (6). We aim at obtaining an observer-based estimator that provides exponential convergence of the estimates $\hat{x}(k)$ to a neighborhood of the true states $x(k)$ and an ISS estimation error $e(k) = \hat{x}(k) - x(k)$ with a linear gain and $exp - KL$ function with respect to the measurement noise and independent of sensor attacks. That is, the estimation error satisfies $|e(k)| \leq \bar{c}\bar{\lambda}^k|e(0)| + \bar{\gamma}_y||\tilde{m}||_k$, for some $\bar{c} > 0$,

$\bar{\lambda} \in (0,1)$, and $\bar{\gamma}_y \geq 0$, and all $e(0) \in \mathbb{R}^n$, $k \geq 0$.

In what follows, we introduce our estimation strategy which is inspired by the method in [25] for the linear case. For (15), let $0 < q < \frac{p}{2}$ be the largest integer such that for each set $J \subset \{1,\dots,p\}$ of sensors with $\text{card}(J) \geq p-2q$, a circle-criterion observer of the form:

$$\begin{aligned}\hat{x}_J^+ =& A\hat{x}_J + Gf(H\hat{x}_J + K_J(\tilde{C}_J\hat{x}_J - \tilde{y}_J)) \\ &+ L_J(\tilde{C}_J\hat{x}_J - \tilde{y}_J) + \rho(u,y),\end{aligned} \tag{16}$$

exists for $\tilde{y}_J$. Here, $\hat{x}_J$ denotes the estimate of the state x from $\tilde{y}_J$, and K_J, L_J are the corresponding observer gains. Matrix $\tilde{C}_J$ is the stacking of all $\tilde{C}_i$, $i \in J$, where $\tilde{C}_i$ is the i-th row of $\tilde{C}$. When we say that the observer exists for $\tilde{y}_J$, we mean that, if $a_J(k) = 0$ for all $k \geq 0$, the error of each observer $e_J(k) = \hat{x}_J(k) - x(k)$ with the following dynamics

$$e_J^+ = (A + L_J\tilde{C}_J)e_J - L_J\tilde{m}_J + G\triangle f_J, \tag{17}$$

with $\triangle f_J := f(\hat{q}) - f(\tilde{q}_J)$, $\tilde{q}_J := Hx$, $\hat{q}_J := H\hat{x}_J + K_J(\hat{y}_J - \tilde{y}_J)$, $\hat{y}_J := \tilde{C}_J\hat{x}_J$, and $\tilde{y}_J := \tilde{C}_Jx + \tilde{m}_J$ is ISS with a linear gain γ_J and an $exp-KL$ function with respect to measurement noise $\tilde{m}_J$. This implies that there exist $c_J > 0$, $\lambda_J \in (0,1)$, $\gamma_J \geq 0$ satisfying:

$$|e_J(k)| \leq c_J\lambda_J^k|e_J(0)| + \gamma_J||\tilde{m}_J||_k, \tag{18}$$

for all $e_J(0) \in \mathbb{R}^n$, $k \geq 0$, and $\tilde{m}_J \in \mathbb{R}^{\text{card}(J)}$ with $\{\tilde{m}_J(k)\} \in l_\infty$.

To construct the estimator, we need that among the p sensors, there are at least $p-q$ attack-free sensors, and among each $p-q$ sensors, there are at least $p-2q$ attack-free sensors. To satisfy this, we make the following assumption.

Assumption 2 *There are at most q sensors attacked,*

$$\text{card}(W) \leq q < \frac{p}{2}. \tag{19}$$

Using the design method proposed in Section III, we construct an observer for each set $J \subset \{1,\dots,p\}$ with $\text{card}(J) = p-q$ and for each set $S \subset \{1,\dots,p\}$ with $\text{card}(S) = p-2q$. For each set J with $\text{card}(J) = p-q$, we define $\pi_J(k)$ for all $k \geq 0$ as the largest deviation between the estimate $\hat{x}_J(k)$ and the estimate $\hat{x}_S(k)$ that is given by any set $S \subset J$ with $\text{card}(S) = p-2q$. That is

$$\pi_J(k) := \max_{S \subset J : \text{card}(S) = p-2q} |\hat{x}_J(k) - \hat{x}_S(k)|. \tag{20}$$

By assumption, among the p sensors, there is at least one set $\bar{I} \subset \{1,\dots,p\}$ of sensors with $\text{card}(\bar{I}) = p-q$ that $\tilde{y}_{\bar{I}} = C_{\bar{I}}x + \tilde{m}_{\bar{I}}$ as $a_{\bar{I}} = 0$; then, in general, all the estimates that appear in the definition of $\pi_{\bar{I}}(k)$ are more consistent than those corresponding to the sets J with $\text{card}(J) = p-q$ and $\tilde{y}_J = \tilde{C}_Jx + a_J + \tilde{m}_J$ with $a_J \neq 0$. This motivates the following state estimation strategy: For all $k \geq 0$,

$$\sigma(k) := \underset{J \subset \{1,\dots,p\} : \text{card}(J) = p-q}{\arg\min} \pi_J(k); \tag{21}$$

then, we say that the estimate given by the set $\sigma(k)$ is a good estimate, i.e.,

$$\hat{x}(k) = \hat{x}_{\sigma(k)}(k), \tag{22}$$

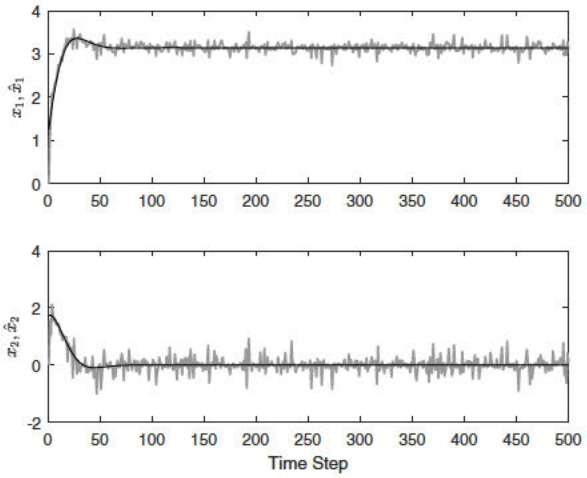

Fig. 3. Estimated states $\hat{x}$ converges to a neighbourhood of the true states x when $a_3 \sim \mathcal{U}(-1,1)$. Legend: $\hat{x}$ (grey), true states (black)

where $\hat{x}_{\sigma(k)}(k)$ denotes the estimate given by the set $\sigma(k)$. The following result states that the proposed estimator is robust with respect to sensor attacks and measurement noise.

Theorem 2 *Consider system* (15)*, the estimator* (20)-(22)*, and define the estimation error $e(k) := \hat{x}(k) - x(k)$ with $\hat{x}(k)$ as in* (22)*. Let Assumptions 1 and Assumption 2 be satisfied; then, there exist constants $\bar{c} > 0$, $\bar{\lambda} \in (0,1)$, and $\bar{\gamma}_y \geq 0$ satisfying:*

$$|e(k)| \leq \bar{c}\bar{\lambda}^k|e(0)| + \bar{\gamma}_y||\tilde{m}||_k, \tag{23}$$

for all $e(0) \in \mathbb{R}^n$, $k \geq 0$, and $\tilde{m} \in \mathbb{R}^p$, $\{\tilde{m}(k)\} \in l_\infty$.

Proof: The proof of Theorem 2 can be found in [30]. ■

Example 2 Consider the following system subject to noise and sensor attacks:

$$\begin{aligned}x^+ &= \begin{bmatrix}1 & \delta \\ 0 & 1\end{bmatrix}x + \begin{bmatrix}\frac{1}{2}\delta\alpha\sin(x_1+x_2) \\ \delta\alpha\sin(x_1+x_2)\end{bmatrix} + \begin{bmatrix}\delta u \\ \delta u\end{bmatrix}, \\ \tilde{y} &= \begin{bmatrix}3 & 3 & 6 & 1.2 \\ 0.3 & 0.6 & 0.9 & 12\end{bmatrix}^\top x + a + \tilde{m}.\end{aligned} \tag{24}$$

Again, we let $\delta = 0.1$, $\alpha = 1$, and $\tilde{m}_i \sim \mathcal{U}(-0.5, 0.5)$ for $i \in \{1,2,3,4\}$. We find that the circle-criterion observer of the form (16) exists for each set of $J \subset \{1,2,3,4\}$ with $\text{card}(J) \geq 1$ and $p = 4$. Hence, we have $q = 1$. We let $W = \{3\}$, which means the 3-rd sensor is under attack. Using the design method proposed in Section III, we design an observer for each $J \subset \{1,2,3,4\}$ with $\text{card}(J) = 3$ and each $S \subset \{1,2,3,4\}$ with $\text{card}(S) = 2$. Therefore, $\binom{4}{3} + \binom{4}{2} = 10$ observers are designed. We initialize the observer at $\hat{x}(0) = [0,0]^\top$. The initial conditions of the system are randomly selected from a standard normal distribution. We let $a_3 \sim \mathcal{U}(-b,b)$ with b given by $1, 10$. For all $k \in [0,500]$, (20)-(22) is used to construct $\hat{x}(k)$. The performance of the designed estimator is shown in Figures 3-4.

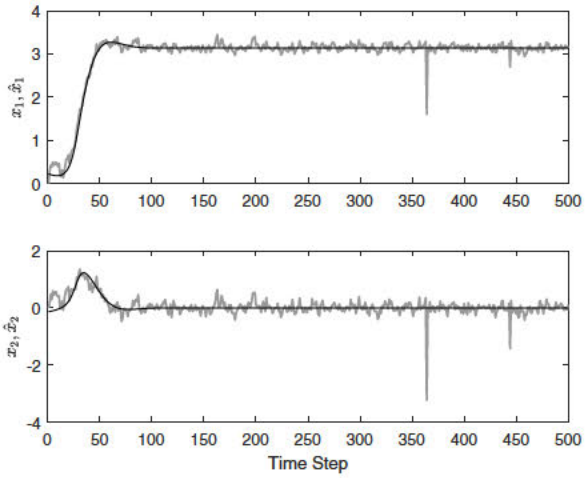

Fig. 4. Estimated states $\hat{x}$ converges to a neighbourhood of the true states x when $a_3 \sim \mathcal{U}(-10, 10)$. Legend: $\hat{x}$ (grey), true states (black)

V. Isolation of sensor attacks

Let q be the largest integer such that a circle-criterion observer exists for each set $J \subset \{1, \ldots, p\}$ with $\text{card}(J) \geq p-2q$. We propose an algorithm for isolating attacked sensors assuming that we know how many sensors are attacked, which is denoted as $q^\star$ ($q^\star \leq q$).

Assumption 3 *There are $q^\star \leq q$ attacked sensors, i.e.,*

$$\text{card}(W) = q^\star, \tag{25}$$

and $q^\star \leq q$ is a known positive integer.

We construct a circle-criterion observer for each set $J \subset \{1, \ldots, p)$ with $\text{card}(J) = p - q^\star$ and for each set $S \subset \{1, \ldots, p\}$ with $\text{card}(S) = p - 2q^\star$. For each set J with $\text{card}(J) = p - q^\star$ and for all $k \geq 0$, we define $\pi^\star_J(k)$ as

$$\pi^\star_J(k) := \max_{S \subset J : \text{card}(S) = p - 2q^\star} |\hat{x}_J(k) - \hat{x}_S(k)|. \tag{26}$$

Since there are $q^\star$ sensors under attack, we know there is one set $\bar{I} \subset \{1, \ldots, p\}$ of sensors with $\text{card}(\bar{I}) = p - q^\star$ with $\tilde{y}_{\bar{I}} = \tilde{C}_{\bar{I}} x + \tilde{m}_{\bar{I}}$ because $a_{\bar{I}} = 0$. Then, in general, all of the estimates that appear in the definition of $\pi^\star_{\bar{I}}(k)$ are more consistent than the estimates corresponding to all the sets J with $\text{card}(J) = p - q^\star$ and $\tilde{y}_J = \tilde{C}_J x + a_J + \tilde{m}_J$ with $a_J \neq 0$. For all $k > 0$, denote $\bar{J}(k)$ as the set of attack-free sensors at time k, then $\bar{J}(k)$ is given as

$$\bar{J}(k) = \operatorname*{arg\,min}_{J \subset \{1,2,\ldots,p\} : \text{card}(J) = p - q^\star} \pi^\star_J(k). \tag{27}$$

Then, the set $\{1, \ldots, p\} \setminus \bar{J}(k)$ is isolated as the set of attacked sensors at time k.

Note, however, that it is still possible that for some $k > 0$ and some $J \subset \{1, \ldots, p\}$ with $\text{card}(J) = p - q^\star$, $a_J(k) \neq 0$ but $J = \bar{J}(k)$, which would result in wrong isolation. To improve the isolation performance, we perform the isolation over windows of $N \in \mathbb{N}$ time-steps. That is, in every N time-steps, where $N \in \mathbb{Z}_{>0}$ is the window size we choose, we keep obtaining $\bar{J}(k)$ from (27) for each k, and we choose the set $J(i)$ that is equal to $\bar{J}(k)$ most often in the i-th window. Then we say that $\{1, \ldots, p\} \setminus J(i)$ is the set of sensors potentially under attack in the i-th time window.

Algorithm 1 Attack Isolation.

1. Design a circle-criterion observer for each set $J \subset \{1, \ldots, p\}$ with $\text{card}(J) = p - q^\star$ and for each set $S \subset \{1, \ldots, p\}$ with $\text{card}(S) = p - 2q^\star$.
2. Initialize the counter variable $n_J(i) = 0$ for each J with $\text{card}(J) = p - q^\star$ and for all $i \in \mathbb{Z}_{>0}$.
3. For $i \in \mathbb{Z}_{>0}$ and for $k \in [1 + (i-1)N, iN]$,calculate $\pi^\star_J(k)$ for all J with $\text{card}(J) = p - q^\star$ as (26)
4. For $i \in \mathbb{Z}_{>0}$ and for $k \in [1 + (i-1)N, iN]$, select the set $\bar{J}(k)$ as (27)
5. For all $k \in [1 + (i-1)N, iN]$, if $\bar{J}(k) = J$ for some J with $\text{card}(J) = p - q^\star$, then update its corresponding counter variable as follows:

$$n_J(i) = n_J(i) + 1.$$

6. For each $i \in \mathbb{Z}_{>0}$, select the set J that is equal to $\bar{J}(k)$ most often

$$J(i) = \operatorname*{arg\,max}_{J \in \{1,\ldots,p\} : \text{card}(J) = p - q^\star} n_J(i).$$

7. For each $i \in \mathbb{Z}_{>0}$, the set of sensors under attack is given as:

$$\tilde{A}(i) = \{1, \ldots, p\} \setminus J(i).$$

8. For each $i \in \mathbb{Z}_{>0}$, return $\tilde{A}(i)$.

Example 3 Consider system (24) with $\delta = 0.1$ and $\alpha = 1$. In each case, we let $\tilde{m}_i \sim \mathcal{U}(-0.5, 0.5)$ for $i \in \{1, \ldots, 4\}$, $q^\star = 1$,$W = \{3\}$, $a_3 \sim \mathcal{U}(-b, b)$, and $b = 1, 2.5$. We choose the window size N to be $50, 100, 200$. We apply Algorithm 1 by running $\binom{4}{2} + \binom{4}{3} = 10$ circle-criterion observers initialized at $\hat{x}(0) = [0, 0]^\top$. The system initial conditions are randomly selected from a standard normal distribution. We follow the steps in Algorithm 1 and check in 1000 time-steps which sensor is isolated in each time window. The obtained results are shown in Figures 5-6.

VI. Conclusion

We have provided a design method for discrete-time circle-criterion observers robust to measurement noise in terms of semidefinite programs. We have proposed a circle-criterion observer-based estimation strategy in the presence of measurement noise and sensor attacks. We have proved that the designed estimator provides ISS estimation errors with a linear gain and an $exp - KL$ function with respect to measurement noise when a sufficiently small subset of sensors are corrupted by (potentially unbounded) attack signals. Finally, we have provided an algorithm for isolating attacked sensors using the proposed estimator.

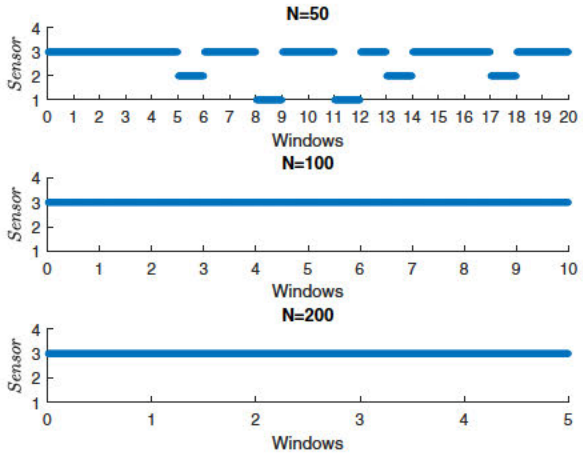

Fig. 5. The sensor isolated by Algorithm 1, $\alpha = 1$, $a_3 \sim \mathcal{U}(-1, 1)$.

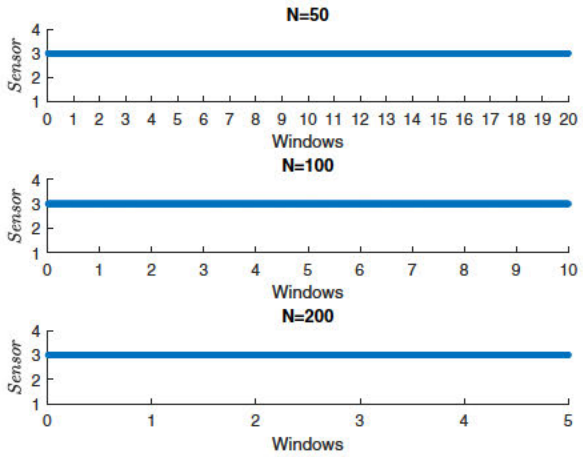

Fig. 6. The sensor isolated by Algorithm 1, $\alpha = 1$, $a_3 \sim \mathcal{U}(-2.5, 2.5)$.

REFERENCES

[1] H. Fawzi, P. Tabuada, and S. Diggavi, "Security for control systems under sensor and actuator attacks," in *IEEE 51st Conference on Desision and Control (CDC)*, 2012, pp. 3412–3417.

[2] M. Massoumnia, G. C. Verghese, and A. S. Willsky, "Failure detection and identification in linear time-invariant systems," *Technology*, no. July, 1986.

[3] M. Pajic, J. Weimer, N. Bezzo, and P. Tabuada, "Robustness of Attack-Resilient State Estimators Robustness of Attack-Resilient State Estimators," no. April, pp. 163–174, 2014.

[4] Y. Mo and B. Sinopoli, "Resilient detection in the presence of integrity attacks," *IEEE Transactions on Signal Processing*, vol. 62, no. 1, pp. 31–43, 2014.

[5] K. G. Vamvoudakis, J. P. Hespanha, B. Sinopoli, and Y. Mo, "Detection in adversarial environments," *IEEE Transactions on Automatic Control*, vol. 59, no. 12, pp. 3209–3223, 2015.

[6] M. S. Chong and M. Kuijper, "Characterising the vulnerability of linear control systems under sensor attacks using a system's security index," in *IEEE 55th Conference on Decision and Control (CDC)*, 2016, pp. 5906–5911.

[7] K. G. Vamvoudakis, J. P. Hespanha, B. Sinopoli, and Y. Mo, "Adversarial detection as a zero-sum game," in *IEEE 51st Conference on Decision and Control (CDC)*, 2012, pp. 7133–7138.

[8] Y. Shoukry, P. Nuzzo, A. Puggelli, A. L. Sangiovanni-Vincentelli, S. Seshia, and P. Tabuada, "Secure state estimation for cyber physical systems under sensor attacks: a satisfiability modulo theory approach," *IEEE Transactions on Automatic Control*, vol. 62, pp. 4917 – 4932.

[9] S. Z. Yong, M. Zhu, and E. Frazzoli, "Resilient state estimation against switching attacks on stochastic cyber-physical systems," in *IEEE 54th Conference on Decision and Control (CDC)*, 2015, pp. 5162–5169.

[10] J. Park, J. Weimer, and I. Lee, "Sensor attack detection in the presence of transient faults," *6th International Conference on Cyber-Physical Systems*, no. April, pp. 1–10, 2015.

[11] Y. Liu, P. Ning, and M. K. Reiter, "False data injection attacks against state estimation in electric power grids," *ACM Transactions on Information and System Security*, vol. 14, no. 1, pp. 21–32, 2009.

[12] A. Teixeira, I. Shames, H. Sandberg, and K. H. Johansson, "Revealing stealthy attacks in control systems," *2012 50th Annual Allerton Conference on Communication, Control, and Computing, Allerton 2012*, pp. 1806–1813, 2012.

[13] C. Murguia and J. Ruths, "Characterization of a CUSUM model-based sensor attack detector," in *2016 IEEE 55th Conference on Decision and Control, CDC 2016*, 2016, pp. 1303–1309.

[14] V. S. Dolk, P. Tesi, C. D. Persis, and W. P. M. H. Heemels, "Event-triggered control systems under denial-of-service attacks," *IEEE Transactions on Control of Network Systems*, vol. 4, pp. 93–105, 2017.

[15] N. Hashemil, C. Murguia, and J. Ruths, "A comparison of stealthy sensor attacks on control systems," in *proceedings of the American Control Conference (ACC)*, 2017.

[16] F. Pasqualetti, F. Dorfler, and F. Bullo, "Attack detection and identification in cyber-physical systems," *IEEE Transactions on Automatic Control*, vol. 58, pp. 2715–2729, 2013.

[17] C. Murguia and J. Ruths, "On reachable sets of hidden cps sensor attacks," in *proceedings of the American Control Conference (ACC)*, 2017.

[18] J. Giraldo, A. Cardenas, and N. Quijano, "Integrity attacks on real-time pricing in smart grids: Impact and countermeasures," *IEEE Transactions on Smart Grid*, 2016.

[19] 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.

[20] S. H. Kafash, J. Giraldo, C. Murguia, A. A. Cardenas, and J. Ruths, "Constraining attacker capabilities through actuator saturation," in *proceedings of the American Control Conference (ACC)*, 2017.

[21] 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.

[22] C. Murguia, I. Shames, F. Farokhi, and D. Nešić, "On privacy of quantized sensor measurements through additive noise," in *proceedings of the 57th IEEE Conference on Decision and Control (CDC)*, 2018.

[23] J. Kim, C. Lee, H. Shim, Y. Eun, and J. H. Seo, "Detection of Sensor Attack and Resilient State Estimation for Uniformly Observable Nonlinear Systems," no. Cdc, pp. 1297–1302, 2016.

[24] T. Yang, C. Murguia, M. Kuijper, and D. Nešić, "Attack detection and isolation for discrete-time nonlinear systems," in *proceedings of the Australian and New Zealand Control Conference (ANZCC)*, 2018.

[25] M. S. Chong, M. Wakaiki, and P. Hespanha, "Observability of linear systems under adversarial attacks *," *Proc. American Control Conf. (ACC)*, pp. 2439–2444, 2015.

[26] M. Arcak, "Nonlinear observers a circle criterion design and robustness analysis.pdf," *Automatica*, vol. 37, no. 12, pp. 1923–1930, 2001.

[27] X. Fan and M. Arcak, "Observer design for systems with multivariable monotone nonlinearities," *Systems and Control Letters*, vol. 50, no. 4, pp. 319–330, 2003.

[28] S. Ibrir, "Circle-criterion approach to discrete-time nonlinear observer design," *Automatica*, vol. 43, no. 8, pp. 1432–1441, 2007.

[29] S. Sundaram, "State and unknown input observers for discrete-time nonlinear systems," in *IEEE 55th Conference on Decision and Control (CDC)*, 2016, pp. 7111–7116.

[30] Y. Tianci, M. Carlos, M. Kuijper, and D. Nešić, "A robust circle-criterion observer-based estimator for discrete-time nonlinear systems in the presence of sensor attacks," *arXiv preprint arXiv:1805.04242*, 2018.

[31] J. Fiala, M. Kočvara, and M. Stingl, "PENLAB: A MATLAB solver for nonlinear semidefinite optimization," *arXiv preprint arXiv:1311.5240*, 2013.