Multi-Observer Approach for Estimation and Control Under Adversarial Attacks

Tianci Yang

Submitted in partial fulfilment of the requirements of the degree of
Doctor of Philosophy

Department of Electrical and Electronic Engineering
THE UNIVERSITY OF MELBOURNE

June 2019
Abstract

Traditional control systems composed of interconnected controllers, sensors, and actuators use point-to-point communication architectures. This is no longer suitable when new requirements – such as modularity, decentralisation of control, integrated diagnostics, quick and easy maintenance, and low cost – are necessary. To meet these requirements, 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. However, these new architectures, especially wireless NCSs, are more susceptible to adversarial attacks. For instance, one of the most well-known examples of attacks on NCSs is the StuxNet virus that targeted Siemens’ supervisory control and data acquisition systems which are used in many industrial processes. Another very recent incident is the attack on the Ukraine power grid system, where an adversarial attack caused a power outage affecting more than 80,000 people for almost 3 hours. These incidents (and many other not mentioned here) show that there is an acute need for strategic defence mechanisms to identify and deal with adversarial attacks on NCSs.

In this thesis, based on sensor and actuator redundancy, we develop a “multi-observer based estimation framework” to address the problem of state estimation for discrete-time nonlinear systems with general dynamics under sensor and actuator false data injection attacks. Although there exist results in the literature addressing similar problems, in general, they are only applicable to some specific classes of nonlinear systems. To the best of the author’s knowledge, a unifying estimation framework that works for general nonlinear systems in the presence of attacks has not been proposed. The estimation scheme provided here can be applied to a large class of nonlinear systems as long as a bank of
observers with certain stability properties exist.

Once an estimate of the system states is obtained from the multi-observer estimator, we provide detection and isolation algorithms for attack detection and for identifying attacked sensors and actuators. For nonlinear systems in the presence of sensor attacks, process disturbance and measurement noise, we detect and isolate attacked sensors by designing multiple observers and comparing their estimates. For noise-free nonlinear systems under sensor and actuator attacks, we isolate attacked sensors and actuators by reconstructing the attack signals. Furthermore, for LTI systems, we provide a simple yet effective control method to stabilize the system despite of sensor and actuator attacks by switching off the isolated actuators and closing the system loop with the proposed estimator and a switching output feedback controller.

Finally, we use a class of nonlinear systems with positive-slope nonlinearities under sensor attacks and measurement noise as a detailed case study where we provide a deeper discussion about the tools that we propose. In particular, we give sufficient conditions under which our tools are guaranteed to work; we also give sufficient conditions under which such methods cannot work. These results have been published in our previous work [74,76].
Declaration

This is to certify that

1. the thesis comprises only my original work towards the PhD,

2. due acknowledgement has been made in the text to all other material used,

3. the thesis is less than 100,000 words in length, exclusive of tables, maps, bibliographies and appendices.

Tianci Yang, June 2019
Preface

The contents of this thesis are the results of the original research unless otherwise stated and have not been submitted for a higher degree at any other university or institution. The material described in this thesis has been obtained under the supervision of Dragan Nešić, Margreta Kuijper, Carlos Murguia. The majority of work, approximately 90%, is my own.

The journal papers that follow from the material presented in the thesis and the contribution of each author are listed in the following:

   First author: problem formulation; finding suitable approaches based on literature review; tackling the problem and completing the mathematics proofs; simulations verification; writing the paper; revision.
   Second to fourth author: supervision; provide technical comments; proofreading; paper polishing.

   First author: problem formulation; finding suitable approaches based on literature review; tackling the problem and completing the mathematics proofs; simulations verification; writing the paper; revision.
   Second to fourth author: supervision; provide technical comments; proofreading;
Several conference papers that follow from the results presented in the thesis and the contribution of each author are listed in the following:

   First author: problem formulation; finding suitable approaches based on literature review; tackling the problem and completing the mathematics proofs; simulations verification; writing the paper; revision.
   Second to fourth author: supervision; provide technical comments; proofreading; paper polishing.

   First author: problem formulation; finding suitable approaches based on literature review; tackling the problem and completing the mathematics proofs; simulations verification; writing the paper; revision.
   Second to fourth author: supervision; provide technical comments; proofreading; paper polishing.

   First author: problem formulation; finding suitable approaches based on literature review; tackling the problem and completing the mathematics proofs; simulations verification; writing the paper; revision.
   Second to fourth author: supervision; provide technical comments; proofreading; paper polishing.
Acknowledgements

I would like to take this opportunity to thank my principal supervisor Prof. Dragan Nešić for his encouragement, support, and patience. I was not a very independent researcher when I started my PhD because I did not have any research experience before. Without much progress in the first two years of my PhD, my progress review meetings were extended several times, which nearly leads to my desperation. However, Dragan has never given up on me, instead, he encouraged me and gave me his best help. His high standards in research has kept me on the right track in regards to pursuing my PhD degree, and made my PhD life very fulfilling. I would not get my PhD degree without his efforts. Every year, Dragan invited me and all his PhD students to his home with enthusiasm. He cooked (he is indeed good at cooking), played guitar and sang for us, and more importantly, he inspired, encouraged us by telling us his stories when he was young.

My co-supervisors Margreta Kuijper and Carlos Murguia also helps me a lot during my PhD. Margreta often gives me valuable feedback on my papers. She is often able to point out the shortcomings of the papers that are difficult to realize. I benefit a lot from her stringent requirements on quality of paper writing. Carlos is a very kind supervisor. He is always full of patience, enthusiasm and new ideas. Although Carlos officially became my supervisor in the third year of my PhD, I have to say that I have learned a lot from him, especially in the aspect of paper writing. I would like to thank Carlos a lot for spending a lot of time discussing the papers, sharing ideas with me, and giving me suggestions on how to follow Dragan’s ideas. The quality of my paper writing has increased a lot with his great help.

I would also like to thank my mum Zhe Qiu and dad Yongli Yang for their support
and understanding, especially when I quited my PhD candidature in the University of New South Wales. Without their encouragement, I would not be so brave to change my research topic and start a new PhD life at the University of Melbourne.
To my parents
Zhe Qiu and Yongli Yang
Notation and Definitions

0.1 Notation

\( \mathbb{R}, \mathbb{N}, \mathbb{Z} \) \hspace{1cm} The set of real, natural, integer, numbers.
\( \mathbb{R}^n \) \hspace{1cm} The set of all n-tuples (vectors) of real numbers.
\( \mathbb{R}^{n \times m} \) \hspace{1cm} The set of \( n \times m \) matrices for any \( m, n \in \mathbb{N} \).
\( A \cup B \) \hspace{1cm} Union of sets \( A \) and \( B \).
\( A \cap B \) \hspace{1cm} Intersection of sets \( A \) and \( B \).
\( \text{rank}(C) \) \hspace{1cm} The rank of matrix \( C \).
\( C^\top \) \hspace{1cm} Transpose of matrix \( C \).
\( C^{-1} \) \hspace{1cm} The inverse of matrix \( C \).
\( (C)_{\text{left}}^{-1} \) \hspace{1cm} The Moore-Penrose pseudoinverse of matrix \( C \).
\( \text{supp}(v) \) \hspace{1cm} The support of vector \( v \), i.e., the set of indices of nonzero elements of vector \( v \).
\( \text{card}(A) \) \hspace{1cm} The cardinality of set \( A \).
\( A \setminus B \) \hspace{1cm} The set \( \{ x : x \in A, x \notin B \} \).
\( \forall \) \hspace{1cm} Universal quantifier.
\( \exists \) \hspace{1cm} Existential quantifier.
\( C > 0 \) \hspace{1cm} Matrix \( C \) is positive definite.
\( C \geq 0 \) \hspace{1cm} Matrix \( C \) is positive semi-definite.
\( A \subseteq B \) \hspace{1cm} \( A \) is a subset of \( B \).
\( A \subset B \) \hspace{1cm} \( A \) is a proper subset of \( B \).
\( a \in A \) \hspace{1cm} \( a \) is an element of set \( A \).
Zero matrix, identity matrix of dimension $n \times n$.

Zero matrix, identity matrix with appropriate dimensions.

The stacking of all the elements of vector $v \in \mathbb{R}^n$ in subset $J \subset \{1, \ldots, n\}$.

The stacking of all rows of matrix $C \in \mathbb{R}^{p \times n}$ in subset $J \subset \{1, \ldots, p\}$.

The binomial coefficient, where $a, b$ are nonnegative integers.

A variable uniformly distributed in the interval $(z_1, z_2)$.

A variable normally distributed with mean $\mu$ and variance $\sigma^2$.

### 0.2 Definitions and lemmas

Several definitions and lemmas that are important in this dissertation are introduced here.

**Definition 0.1.** [37] A continuous function $\alpha : [0, a) \to [0, \infty)$ is said to belong to class $K$, if it is strictly increasing and $\alpha(0) = 0$.

**Definition 0.2.** [37] A continuous function $\beta : [0, a) \times [0, \infty) \to [0, \infty)$ is said to belong to class $KL$ if, for fixed $s$, the mapping $\beta(r, s)$ belongs to class $K$ with respect to $r$ and, for fixed $r$, the mapping $\beta(r, s)$ is decreasing with respect to $s$ and $\beta(r, s) \to 0$ as $s \to \infty$.

**Definition 0.3.** [37] A continuous function $\beta : [0, a) \times [0, \infty) \to [0, \infty)$ is said to belong to class $exp-KL$ if there exist $c > 0$, $\lambda \in (0, 1)$, such that $\beta(s, k) = c \lambda^k \cdot s$.

**Definition 0.4.** [33] (Input-to-State Stability) Consider a discrete-time system:

$$ e^+ = f(e, u), $$

with input $u$. (1) is Input-to-State Stable (ISS) if there exist a KL-function: $\beta(\cdot, \cdot)$ and a K-function $\gamma(\cdot)$ such that

$$ |e(k)| \leq \beta(|e(0)|, k) + \gamma(||u||_\infty) $$

for all $e(0) \in \mathbb{R}^n$, $\{u(k)\} \in l_\infty$, and $k \geq 0$. In particular, system (1) is said to be Input-to-State...
Stable (ISS) with linear gain $\gamma$ and exp – KL function, if $\beta(|e(0)|, k) = c\lambda^k|e(0)|$ with $c > 0$, $\lambda \in (0, 1)$, and $\gamma(||u||_\infty) = \gamma||u||_\infty$ with $\gamma \geq 0$.

**Definition 0.5.** [7] A linear matrix inequality (LMI) has the form:

$$F(x) \triangleq F_0 + \sum_{i=1}^{m} x_i F_i > 0, \quad (3)$$

where $x \in \mathbb{R}^m$ is the variable and the symmetric matrices $F_i = F_i^T \in \mathbb{R}^{n \times n}, i = 0, \ldots, m,$ are given.

**Definition 0.6.** [60] (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 $q_1, q_2 \in \mathbb{R}^{n_q}$:

$$\begin{bmatrix} \Delta q \\ \Delta f \end{bmatrix}^T M \begin{bmatrix} \Delta q \\ \Delta f \end{bmatrix} \geq 0, \quad (4)$$

where $\Delta q = q_1 - q_2$ and $\Delta f = f(q_1) - f(q_2)$. 
# Abbreviations

<table>
<thead>
<tr>
<th>Abbreviation</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>LTI</td>
<td>Linear Time-Invariant</td>
</tr>
<tr>
<td>ISS</td>
<td>Input-to-State Stable</td>
</tr>
<tr>
<td>GES</td>
<td>Globally Exponentially Stable</td>
</tr>
<tr>
<td>GAS</td>
<td>Globally Asymptotically Stable</td>
</tr>
<tr>
<td>UIO</td>
<td>Unknown Input Observer</td>
</tr>
<tr>
<td>LMI</td>
<td>Linear Matrix Inequality</td>
</tr>
<tr>
<td>NCS</td>
<td>Networked Control System</td>
</tr>
<tr>
<td>CPS</td>
<td>Cyber Physical System</td>
</tr>
</tbody>
</table>
Contents

0.1 Notation ................................................................. xiii
0.2 Definitions and lemmas ................................................ xiv

1 Introduction ............................................................... 1
  1.1 Motivation ............................................................. 1
    1.1.1 Networked Control Systems ................................. 1
    1.1.2 Security in NCSs .............................................. 5
  1.2 Overview of the Literature ......................................... 10
    1.2.1 Fault diagnosis and Attack Detection .................... 11
    1.2.2 Secure Control ............................................... 12
    1.2.3 Secure Estimation .......................................... 13
    1.2.4 Multi-Observer Approach .................................. 15
  1.3 Outline of the Thesis .............................................. 17

2 Secure Estimation for Nonlinear Systems under Sensor Attacks .... 21
  2.1 Overview ............................................................. 21
  2.2 Multi-Observer Estimator (Noise-free Case) .................... 23
    2.2.1 Application Examples ....................................... 26
  2.3 Robust Multi-Observer Based Estimator .......................... 30
    2.3.1 Application Example ....................................... 34
  2.4 Isolation of Attacked Sensors .................................... 35
  2.5 Conclusion ......................................................... 40

3 Estimation for Nonlinear Systems Under Sensor and Actuator Attacks 43
  3.1 Overview ............................................................. 43
  3.2 Estimation ............................................................ 45
    3.2.1 Complete Unknown Input Observers .......................... 46
    3.2.2 Partial Unknown Input Observers ............................ 49
    3.2.3 An Application Example ....................................... 54
  3.3 Isolation of Attacks .............................................. 57
  3.4 Conclusion .......................................................... 61

4 Case Study: Systems with Positive-Slope Nonlinearities .......... 67
  4.1 Overview ............................................................. 67
  4.2 Robust circle-criterion observer .................................. 68
  4.3 Observer-based estimator under sensor attacks and sensor noise .... 74
4.4 Detection and Isolation of sensor attacks ............................ 79
  4.4.1 Detection .......................................................... 79
  4.4.2 Isolation .......................................................... 85
4.5 Conclusion ............................................................. 91

5 Estimation and Control for LTI Systems Under Sensor and Actuator Attacks 93
  5.1 Overview ............................................................. 93
  5.2 Estimation .......................................................... 94
    5.2.1 Complete Unknown Input Observers .......................... 95
    5.2.2 Partial Unknown Input Observers ............................. 100
  5.3 Attack Isolation and Reconstruction ............................... 106
  5.4 Control ............................................................. 109
    5.4.1 Sensor attacks only ........................................... 109
    5.4.2 Sensor and actuator attacks ................................. 113
  5.5 Conclusion ........................................................ 116

6 Conclusions and Future Work ............................................ 117
  6.1 Conclusions ....................................................... 117
  6.2 Future Work ....................................................... 118
    6.2.1 Enhanced Secure Estimation Techniques ..................... 118
    6.2.2 Attack Isolation for Noisy Systems ......................... 118
    6.2.3 Secure Control of Nonlinear Systems under Actuator Attacks ... 119
List of Figures

1.1 The connected community. http://www.emfexplained.info/?ID=25916  
1.2 A traditional control system.  
1.3 A typical NCS.  
1.4 Cyber attacks on critical infrastructure in the US [32].  
1.5 The cyberphysical attack space. Each point depicts the qualitative classification of a given attack scenario.[65]  
1.6 CIA in Cyber Security [4].  
1.7 A NCS under attacks.  
1.8 Multi-observer estimator for LTI systems under sensor attacks.  
2.1 Estimated states $\hat{x}$ converges to the true states $x$ when $a_2 \sim U(-10,10)$. Legend: $\hat{x}$ (blue), true states (black).  
2.2 Estimated states $\hat{x}$ converges to the true states $x$ when $a_2 \sim U(-10,10)$. Legend: $\hat{x}$ (blue), true states (black).  
2.3 Estimated states $\hat{x}$ converges to a neighbourhood of the true states $x$ when $a_3 \sim U(-10,10)$. Legend: $\hat{x}$ (blue), true states (black).  
2.4 Attack isolation, $a_3 \sim U(-1,1)$.  
3.1 Complete UIO-based estimator.  
3.2 Partial UIO-based estimator.  
3.3 Estimated states $\hat{x}$ converges to the true states $x$ when $a_{u1}, a_{u2}, a_{y2} \sim U(-1,1)$. Legend: $\hat{x}$ (grey), true states (black).  
3.4 Estimated states $\hat{x}$ converges to the true states $x$ when $a_{u1}, a_{u2}, a_{y2} \sim U(-10,10)$. Legend: $\hat{x}$ (grey), true states (black).  
3.5 Estimated states $\hat{x}$ converges to the true states $x$ when $a_{u1}, a_{y2} \sim U(-1,1)$. Legend: $\hat{x}$ (grey), true states (black).  
3.6 Estimated states $\hat{x}$ converges to the true states $x$ when $a_{u1}, a_{y2} \sim U(-10,10)$. Legend: $\hat{x}$ (grey), true states (black).  
3.7 Estimate of $a_y$ when $a_{u1}, a_{u2}, a_{y2} \sim U(-1,1)$.  
3.8 Estimate of $a_u$ when $a_{u1}, a_{u2}, a_{y2} \sim U(-1,1)$.  
3.9 Estimate of $a_y$ when $a_{u1}, a_{u2}, a_{y2} \sim U(-10,10)$.  
3.10 Estimate of $a_u$ when $a_{u1}, a_{u2}, a_{y2} \sim U(-10,10)$.  
3.11 Estimate of $a_y$ when $a_{u3}, a_{y2} \sim U(-1,1)$.  
3.12 Estimate of $a_u$ when $a_{u3}, a_{y2} \sim U(-1,1)$.  
3.13 Estimate of $a_y$ when $a_{u3}, a_{y2} \sim U(-10,10)$.  
3.14 Estimate of $a_u$ when $a_{u3}, a_{y2} \sim U(-10,10)$.
4.1 Estimated states $\hat{x}$ converges to a neighbourhood of the true states $x$. Legend: Observer obtained via Theorem 4.1 (red), Observer from [60] (grey), true states (black) .......................... 74
4.2 Estimated states $\hat{x}$ converges to a neighbourhood of the true states $x$. Legend: Observer obtained via Theorem 4.1 (red), Observer from [30] (grey), true states (black) .......................... 75
4.3 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) ................... 80
4.4 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) ................... 81
4.5 Attack detection, $a_2 \sim \mathcal{U}(-0.7,0.7)$ .................. 85
4.6 Attack detection, $a_2 \sim \mathcal{U}(-1,1)$ .......................... 86
4.7 Attack isolation, $a_3 \sim \mathcal{U}(-2,2)$ .......................... 91
4.8 Attack isolation, $a_3 \sim \mathcal{U}(-5,5)$ .......................... 92
5.1 Estimated states $\hat{x}$ converges to the true states $x$ when $a_u, a_y \sim \mathcal{U}(-10,10)$. Legend: $\hat{x}$ (blue), true states (black) .................. 100
5.2 Estimated states $\hat{x}$ converges to the true states $x$ when $a_u, a_y \sim \mathcal{U}(-10,10)$. Legend: $\hat{x}$ (blue), true states (black) .................. 106
5.3 Estimated actuator attacks $\hat{a}_u^+$ converges to $a_u$ when $a_u, a_y \sim \mathcal{U}(-10,10)$. Legend: $\hat{a}_u^+$ (blue), $a_u$ (black) .................. 108
5.4 Estimated sensor attacks $\hat{a}_y$ converges to $a_y$ when $a_u, a_y \sim \mathcal{U}(-10,10)$. Legend: $\hat{a}_y$ (blue), $a_y$ (black) .................. 109
5.5 Estimated actuator attacks $\hat{a}_u^+$ converges to $a_u$ when $a_u, a_y \sim \mathcal{U}(-10,10)$. Legend: $\hat{a}_u^+$ (blue), $a_u$ (black) .................. 110
5.6 Estimated sensor attacks $\hat{a}_y$ converges to $a_y$ when $a_u, a_y \sim \mathcal{U}(-10,10)$. $\hat{a}_y$ (blue), $a_y$ (black) .................. 111
5.7 Controlled states when $a_2 \sim \mathcal{U}(-10,10)$ .......................... 112
5.8 Estimation, isolation, and control diagram .................. 114
5.9 State trajectories when $a_u, a_y \sim \mathcal{U}(-10,10)$ .......................... 116
List of Tables

2.1 Systems/observers satisfying Definition 2.1 in the literature. . . . . . 24
2.2 Systems/observers satisfying Definition 2.2 in the literature. . . . . . 32
Chapter 1
Introduction

1.1 Motivation

1.1.1 Networked Control Systems

Control systems are rapidly evolving into highly interconnected, distributed and heterogeneous systems where control algorithms rely on distributed computation and interact with heterogeneous digital communication networks. The digitization of control systems has significantly enlarged the control systems applications and enhanced existing control applications as it offers simpler installation and maintenance, it is cheaper and it can reduce the weight and volume of the system. Moreover, wireless communication networks are necessary in some applications, such as control of autonomous vehicles operating on smart highways. X-by-wire technologies in the automotive and aerospace applications are nowadays standard, whereas Internet of Things is an area of active research. Smart cities, smart highways or smart farms are applications that are enabled by control algorithms implemented over communication networks. We elaborate in more detail on several such applications.

- **Industrial Control Systems (ICSs):** These systems are used to monitor and control plants and equipment in industry, e.g., nuclear plants and water/sewage systems. ICSs rely heavily on the network structure. For instance, the programmable logic controller (PLC) [6], which is a popular controller for ICS and designed for operating in hostile environments, is often equipped with wireless and wired communication capacity that is configured depending on the surrounding environments. It
can be connected to PC systems in a control center that monitors and controls the
operations. Moreover, the vision of Industry 4.0 [2], also known as the fourth in-
dustrial revolution, is the integration of massively deployed smart computing and
network technologies in industrial production and manufacturing settings. By in-
terconnecting equipment through networks, productivity, efficiency, and safety of
manufacturing system can be improved.

• **Internet of Things:** “Internet of Things (IoTs)” refers to millions of physical devices
around the world that are connected to the Internet, collecting and sharing data.
Any physical object, from smart cameras to aircrafts or self-driving cars that are
connected and controlled through the Internet are considered IoTs devices. Other
examples are air conditioners that can be adjusted using smartphones, or the wear-
able devices, from FitBits to smart watches or anything that we are wearing that is
connected to the Internet.

• **X-by-wire technology:** X-by-wire technology in the automotive industry uses elec-
trical or electro-mechanical systems to perform vehicle functions. For instance, in
a fly-by-wire aircraft, the pilot’s control inputs are fed and interpreted by the flight
computer which moves the control surfaces to achieve the desired flight path mod-
ification. This is in contrast with the traditional mechanical transmission, where
the pilot’s inputs are directly transmitted to the flight control actuators through
hydraulic transmission.

• **Smart grids:** Smart grids consist of two major components, power application and
supporting infrastructure. Power application provides the core functions of the grid
and the supporting infrastructure are the intelligent components that use software,
hardware, and communication networks to monitor and control the core opera-
tions of the smart grid. Smart grids are of great benefit to society since they enable
global load balancing, smart generation, and energy savings. They also allow home
customers to better control their energy usage.

• **Smart cars and smart highways:** The emergence of smart vehicles is an impor-
tant evolution of the automotive industry. They are more environmentally friendly,
fuel-efficient, and possess enhanced safety features. By connecting processors to a central computing platform through communication networks, various functions such as engine emission control, brake control, entertainment, and comfort features can be carried out. Smart highway is a type of roadway that allows for technological integration into current transport roadways, where smart cars communicate to each other to reduce traffic congestion and improve safety.

- **Wireless medical devices**: Over the last decade, there has been a radical shift from wired to wireless medical devices. By connecting patients’ monitors to wireless networks, vital signs can be continuously monitored through the hospital access point. Moreover, doctors now have real-time access to patient data throughout hospitals, allowing them to make immediate critical care decisions and perform administrative tasks.

- **Smart buildings**: Smart buildings leverage pervasive wireless connectivity, sensors, and the cloud to remotely monitor and control a range of buildings’ systems, e.g., access control, security systems, lighting, and air conditioning. They provide better efficiency, safety, and comfort, while providing cost savings.

- **Smart Community**: In Figure 1.1, the paradigm of “connected community” is depicted, where various NCSs applications, e.g., smart farms, smart hospitals, smart homes, smart utilities, are prevalently used. All these smart objects share the same common feature, i.e., they use communication networks as the channel between controllers and plants.

Advanced digital technologies in control systems have led to novel control architectures that include digital communication networks in control loops. A traditional control system, see Figure 1.2, consists of a plant and controller that are connected with dedicated communication channels. On the other hand, modern control systems often have the structure given in Figure 1.3, where the controller and plant communicate via a communication network. The network may be either wired, like CAN or Flexray, or wireless, such as WirelessHART. In Figure 1.3, the signals \( \hat{y} \) and \( \hat{u} \) are sampled, quantized and potentially delayed versions of \( y \) and \( u \), which poses a number of design challenges.
More precisely, the presence of network in control loops leads to the following undesirable phenomena that complicate the design, reduce the performance and may even cause instability in the system:

- Sampling jitter and uncertain sampling times
- Quantization and finite capacity of the communication channel
- Data scheduling via network communication protocols
- Communication and computation delays
- Data dropouts
- Vulnerability of the system to adversarial attacks and lack of privacy
As a result of the above mentioned phenomena, the modeling, analysis and design of NCSs is significantly more complex than in their traditional counterparts. The undesirable effects of the network induced phenomena can be somewhat mitigated by the use of distributed computation that gives rise to “smart sensors” and “smart actuators” but this further complicates the analysis and design of NCSs. The literature that addresses the first five of the above phenomena in NCSs is rich and there have been significant advances in the past two decades. Security in NCSs is a relatively new research area and it is the main topic of this thesis. We will address in more detail security in NCSs in the next section.

1.1.2 Security in NCSs

Wireless networks might render NCSs more vulnerable to attacks since opponents may inject signals into the closed-loop dynamics by hacking into the communication network causing poor system performance or damage. Cyber-physical attacks on control systems have caused substantial damage to a number of engineering systems. One of the most well-known examples is the attack on Maroochy Shire Council’s 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 StuxNet virus that targeted Siemens’ supervisory control and data acquisition systems which are used in many industrial processes. In 2014, the computers of a German steel mill were hacked through a minor support system for environmental control, which resulted in a massive destruction of a blast furnace in the steel mill. Besides these incidents, a report
published by the US Department of Homeland Security shows that the number of cyber attacks on critical infrastructure in the US from 2009 to 2015 has been increasing significantly (3,000% in only six years), see Figure 1.4. Hence, intelligent actions are needed for countering attacks to NCSs.

There are different classes of possible attacks to NCSs. Namely, Denial of Service (DoS), bias injection, replay, zero dynamic) and eavesdropping [22, 26, 44, 53, 65]. The control community (see [65] and references therein) has characterized all these classes of attacks by the resources needed to launch them, i.e., model knowledge, disclosure resources, and disruption resources, see Figure 1.5. 

Disclosure resources enable attackers to obtain sensitive information about the system operation, e.g., the computed control actions $u$ and real-time measurements $y$. Note that the system dynamics can not be modified by attackers with disclosure resources only. Disruption resources are related to the attack action used to affect the system dynamics by violating data integrity and availability. Model knowledge refers to the amount of a priori information the attacker has about the system dynamics. Model knowledge is a core component of the adversarial model as it enables attackers to construct intelligent attack signals that are difficult to detect. For instance, if the closed-loop system possesses an unstable zero [81], a zero dynamics attack that
1.1 Motivation

leads to unbounded trajectories could be launched without being detected by monitoring system [27, 39, 67]. To implement such an intelligent attack, the opponent needs disruption resources and knowledge of the exact system model, see Figure 1.5. Bias injection attacks also require model knowledge and disruption resources to inject constant biases without being detected [40, 41]. To implement a covert (undetectable) false data injection attack, an attacker needs detailed model knowledge and full access to sufficiently many sensor and actuator channels [14, 57]. Besides these “intelligent” attacks, some other attacks might also cause serious damage to NCSs without having model knowledge. For instance, replay attacks is an attack scenario where the opponent records control and sensor data sequences, and then replays them to actuators and monitors, respectively. Disclosure and disruption resources are only needed to launch replay attacks [36, 80, 82]. Denial-of-service attacks prevent data from reaching its receiver by jamming the communication channel. This attack only requires disruption resources [10, 19, 79]. Eavesdropping attacks only require disclosure resources to obtain sensitive information about the system. Note that attacks might also be classified based on what main properties of data and IT services (e.g., confidentiality, integrity, and availability [4]) the attackers aim at violating. As depicted in Figure 1.6 (a), attackers might only aim at violating data confidentiality using disclosure resources, e.g., eavesdropping attacks. False data injection attacks are often used to violate data integrity so that the system operation is disrupted. As shown in Figure 1.6 (b), in the false data injection attack scenario, attack signals are injected to original signals to deceive the receiver’s side. Attackers might also aim at violating data availability using Denial-of-Service (DoS) attacks as shown in Figure 1.6 (c).

Security control refers to a set of techniques used in IT systems to secure the integrity of systems. For instance, cryptography, access control, and intrusion detection. In what follows, we give a brief introduction of these techniques.

- **Cryptography** is about constructing protocols that prevent third parties from reading private messages [45, 51]. There are five primary functions of cryptography:
  1. Privacy/confidentiality: Ensuring that no one can read the message except the intended receiver.
  2. Authentication: The process of proving one’s identity.
Figure 1.5: The cyberphysical attack space. Each point depicts the qualitative classification of a given attack scenario.[65]

3. Integrity: Assuring the receiver that the received message has not been altered in any way from the original.
4. Non-repudiation: A mechanism to prove that the sender really sent this message.
5. Key exchange: The method by which crypto keys are shared between sender and receiver.

- **Access control** is a security technique that regulates who can access to certain information in computing environments [45]. At a high level, access control is a selective restriction of access to data. It consists of two main components: authentication, which is used to verify that someone is who they claim to be, and authorization, which determines whether a user should be given access to the data or resources.

- **Intrusion detection** is about detecting if someone is trying to compromise information systems through malicious activities or through policy violations [43]. A intrusion detection is done by monitoring system activity through examining vulnerabilities in the system, the integrity of files and conducting an analysis of patterns based on already known attacks. Moreover, it also monitors the Internet automatically to search for any latest threats that could cause a future attack.

In general, the IT community is protecting the cyber components of NCSs aiming at
1.1 Motivation

preventing cyber-physical attacks from happening. However, IT security tools do not deal with the dynamic physical components of NCSs. The physical layer of NCSs can be manipulated by attacks without even tampering with IT components. Therefore, to complement IT security approaches, we need the systems and control theoretic perspective of the problem to secure NCSs.

The control community mainly focuses on mitigation of attacks, i.e., reducing the damage caused by adversarial signals. The first step towards attack mitigation is to model cyber-physical attacks. In [65], the authors consider discrete-time linear time-invariant (LTI) systems subject to adversarial attacks. Several attack scenarios are modeled as additive signals, i.e.,

\[
\begin{align*}
    x^+ &= Ax + B(u + a_u) + Gd + Ff, \\
    y &= Cx + a_y + m,
\end{align*}
\]

with system state \( x \in \mathbb{R}^n \), control input \( u \in \mathbb{R}^nu \), actuator and sensor additive attacks \( a_u \in \mathbb{R}^nu \) and \( a_y \in \mathbb{R}^ny \), process disturbance \( d \in \mathbb{R}^nd \), measurement noise \( m \in \mathbb{R}^ny \),

Figure 1.6: CIA in Cyber Security [4].
Figure 1.7: A NCS under attacks.

and possible faulty signals $f \in \mathbb{R}^{n_f}$. Zero-dynamics attacks can be modeled as $a_u(k) = g\nu^k$ and $a_y(k) = 0$ for all $k \geq 0$, where $\nu$ denotes a system zero [81], and $g$ denotes the input-zero direction [81]. Bias injection attacks can be formulated by letting $a_u^+ = \beta a_u + (1 - \beta)\bar{a}$, where $a_u(0) = 0$, $\beta \in (0, 1)$, and $\bar{a}$ is the desired bias. DoS attacks can be modeled as $a_{ui} = -u_i$, for all $i \in W_u$, where $W_u \subseteq \{1, \ldots, n_u\}$ denotes the set of attacked actuators. In the replay attack scenario, $a_{ui}(k) = u_i(k - T) - u_i(k)$ for all $i \in W_u \subseteq \{1, \ldots, n_u\}$, $a_{yi}(k) = y_i(k - T) - y_i(k)$ for all $i \in W_y \subseteq \{1, \ldots, n_y\}$, where $W_y$ denotes the set of attacked sensors, and $T$ is the recording duration.

In this thesis, we consider attacks that can be modeled as additive signals to sensor and actuator data. Hence, zero dynamics, DoS, bias injection, and replay attacks all fit in our problem setting. However, we assume that attackers have limited resources and they can only attack some sensors/actuators. For instance, attackers only record and replay some of the sensor measurements in a replay attack scenario, or only prevents some of sensor measurement from reaching the controller in a DoS attack scenario. In Figure 1.7, a diagram of NCSs under additive attacks is depicted. After the injection of $a_u$ and $a_y$, the controller output $u$ now becomes $\hat{u} + a_u$ after transmission, and $y$ becomes $\hat{y} + a_y$.

### 1.2 Overview of the Literature

Here, instead of trying to be exhaustive with the literature overview, we present some representative results that consider different attack scenarios; then we would present in more detail results that are directly related to this thesis.
1.2 Overview of the Literature

1.2.1 Fault diagnosis and Attack Detection

Fault diagnosis is a well-established area of research [5,13,38]. However, given the strategic and adaptive nature of attacks, classical fault-tolerant filters that can detect faults on actuators/sensors might not be adequate when considering cyber-physical attacks. For example, in [5], filters are designed to detect faulty actuators and sensors by assuming a priori knowledge (statistical or temporal) of the fault signals. This is not possible with attacks as the signal injected by opponents are arbitrarily and possibly unbounded. In [13], unknown input observers are used for detecting faulty actuators under the assumption that the exact number of faulty actuators is known. This assumption is not reasonable in an adversarial setting as the number of attacked sensors/actuators is usually unknown. It follows that there is an acute need for efficient detection, isolation, and mitigation techniques to deal with cyber-physical attacks on NCSs.

DoS Attacks

The problem of mitigation of DoS attacks is considered in [44]. The authors present two mitigation methods using network intrusion detection and response techniques. In [66], a mechanism called Change-Point Monitoring (CPM) is presented to detect DoS attacks.

Replay Attacks

A detection strategy against replay attacks is designed in [48] using the idea of water-marking. They add random vectors (the water-mark signals) to sensor and actuator data and then check if the statistics induced by these vectors is present in real-time data. If there are replay attacks, the effect of water-marks on sensor/actuator data would be absent.

Bias Injection Attacks

A detection mechanism for bias injection attacks is proposed in [41]. It is based on the extraction of suitable robust invariant sets for networked power system. Whenever the
state vector exits these invariant sets, alarm signals are triggered indicating a potential security breach.

**Covert Attacks and Zero Dynamics Attacks**

In [26, 52], the authors provide mathematical tools for detecting covert and zero dynamics attacks on NCSs. The main idea is that they disturb the system dynamics so that attackers lose their perfect model knowledge. Then, attacks that were designed using the undisturbed system model would not be “stealthy” to the disturbed one.

**False Data Injection Attacks on Some Sensors/Actuators**

Assuming that only some sensors are corrupted by false data injection attacks, the problem of sensor attack detection for LTI systems is addressed in [64] using the residual computation method. Similarly, in our paper [74], the problem of attack detection for a class of nonlinear systems with positive-slope nonlinearities under sensor false data injection attacks is solved by using a bank of observers with linear ISS gains.

### 1.2.2 Secure Control

The problem of secure control considers how to design controllers to stabilize the system in spite of attacks. More specifically, the secure control problem refers to all the algorithms and architectures designed to cope with attacks against NCSs under well-defined adversarial models and trust assumptions [11].

**DoS Attacks**

In [19], a systematic design framework for output-based dynamic event-triggered control (ETC) systems under DoS attacks is proposed. For a class of DoS attack characterized by frequency and duration properties, the stability of system is preserved by the proposed ETC scheme. The stability analysis for NCSs under DoS attacks is given in [10], where DoS attacks are modeled as time-varying delays. Then, an attack-based delay-dependent
criterion for stability is given in terms of LMIs. An $H_\infty$ controller is designed for mitigating DoS attacks in [77].

**Replay Attacks**

An LQG controller is designed in [48] to mitigate replay attacks. Similarly, in [46], using a stochastic game approach, an optimal control policy for switching between control-cost optimal and secure controllers is proposed. In [82], a variation of the receding-horizon control scheme is proposed to deal with the replay attacks and the resulting system performance degradation is also analyzed.

**False Data Injection Attacks on Some Sensors/Actuators**

An adaptive controller is designed in [71] to guarantee ultimate boundedness of the closed-loop dynamics under additive attacks. In [23], an output-feedback controller that stabilizes the system despite of sensor attacks is proposed and a principle of separation between estimation and control is proved to hold. Assuming that a sufficiently small subset of actuators are under attack, a state feedback controller is designed in our previous work [75] to stabilize the system dynamics by switching off the attacked actuators.

### 1.2.3 Secure Estimation

Secure estimation is about providing robust state estimates in spite of attacks. It is an important step towards mitigating adversarial attacks since state estimates provide knowledge about the system dynamics which can be used to detect the presence of attacks.

**False Data Injection Attacks and DoS Attacks**

Assuming attackers can only compromise a sufficiently small subset of sensors/actuators, two major categories of secure estimation approaches have been proposed. The first category performs state estimation by analyzing sensor information collected during a finite time window. For example, in [22], compressed sensing techniques are used for
state reconstruction for discrete-time LTI systems under sensor and actuator attacks. The authors consider sensor attacks as errors in the sensors and formulate the problem of initial state reconstruction as an error correction problem. The authors use the compressed sensing techniques in [9] to convexify the optimization problem used to construct their estimators. Similarly, for a class of power systems under sensor attacks, the authors in [28] provide an estimator of the system state using compressed sensing techniques. Another example in this category is the estimation method proposed in [53], the Satisfiability Modulo Theory (SMT) approach. In [53], a Luenberger-like observer is proposed based on efficient SMT techniques. A family of filters sharing a single dynamical equation for the states, but different output equations, are used to generate estimates corresponding to different subsets of sensors being attacked. Then, they provide an SMT-based detection procedure to detect which sets of sensors are attack-free. This approach has also been used for state estimation for differentially flat systems under sensor attacks in [54]. The Gramian-based observer proposed in [15] also belongs to this category.

The second category of secure estimation methods explores redundancy in sensors and actuators by designing recursive observers/filters [15, 47, 50, 74]. In [15, 47, 63], a multi-observer approach is proposed for state estimation for LTI systems under sensor attacks. It is assumed that only a subset of sensors is under attack. If observers exist for the remaining attack-free sensors and we are able to find such observers, we can obtain robust state estimate despite sensor attacks. Therefore, the basic idea of the multi-observer approach is designing a bank of observers for different subsets of sensors, and then, using some estimation strategy to find attack-free observer(s), we can obtain true estimates of the system state. Following the idea of [15], in our previous work [74], a multi-observer estimator is provided for a class of nonlinear systems with positive-slope nonlinearities. In this thesis, we extend the results in [15] by considering different classes of systems and formulations. Next, we give a more detailed introduction of the multi-observer approach proposed in [15].
1.2 Overview of the Literature

1.2.4 Multi-Observer Approach

The core of our estimation scheme is based on the work in [15], where the problem of state estimation for continuous-time LTI systems under sensor attacks is addressed. The system considered in [15] is formulated as follows:

\[
\begin{align*}
\dot{x} &= Ax + Bu \\
y &= Cx + ay
\end{align*}
\]  

(1.2)

with state \( x \in \mathbb{R}^n \), input \( u \in \mathbb{R}^{n_u} \), \( i \)-th sensor measurement \( y_i \in \mathbb{R} \), stacked measurements \( y := (y_1, \ldots, y_{n_y})^T \in \mathbb{R}^{n_y} \), attack signal \( a_i \in \mathbb{R} \). If the \( i \)-th sensor is not attacked, \( a_i(k) = 0 \) for \( k \geq 0 \); otherwise, sensor \( i \) is under attack and \( a_i(k) \) is arbitrary and possibly unbounded. The following assumptions on the system have been made:

**Assumption 1.1.** The unknown set of attacked sensors do not change over time.

**Assumption 1.2.** \((A, C^T)\) is observable for every \( J \subset \{1, \ldots, n_y\} \) with \( \text{card}(J) = n_y - 2q > 0 \), and \( q \) is a positive integer.

**Assumption 1.3.** There are at most \( q > 0 \) sensors attacked.

Assumption 1.1 is a realistic assumption as the time it takes for an attacker to gain access to another sensor is often large compared to the estimation time scale. From Assumption 1.2 and 1.3, it can be seen that \( q < \frac{n_y}{2} \), that is the number of attacked sensors must be less than half of the total number of sensors. Therefore, having redundant sensors would increase the success rate of the estimator. For example, in order to endure 1 sensor attack, at least 3 sensors are needed for the system. The main idea of multi-observer estimator proposed in [15] is shown in Figure 1.8.

Among all the \( n_y \) sensors, the selectors choose every \( J_s \subset \{1, \ldots, n_y\} \) set of \( n_y - q \) sensors and every \( S_s \subset \{1, \ldots, n_y\} \) set of \( n_y - 2q \) sensors, and then, a Luenberger observer is designed for each \( J_s \) and \( S_s \). Their estimates are denoted by \( \hat{x}_{J_s} \) and \( \hat{x}_{S_s} \), respectively. Then, what the estimator does is to find one of \( J_s \) that is attack-free. This “special” \( J_s \) exists because in Assumption 1.3, it is assumed that at most \( q \) sensors are attacked, therefore, among all the \( n_y \) sensors, at least \( n_y - q \) of them are attack-free. Hence, at least one \( J_s \) is a...
subset of attack-free sensors. Note that if $J_s$ is a subset of attack-free sensors, then for all $S_s \subset J_s$ with $\text{card}(J_s) = n_y - q$, $\hat{x}_{J_s}(k) = 0$ for all $k \geq 0$. Therefore, for this special $J_s$, $\hat{x}_{J_s}$ and $\hat{x}_{S_s}$ for all $S_s \subset J_s$ are all exponentially stable estimates of the system states and they will be consistent for the attack-free $J_s$. Therefore for each $J_s$, the estimator computes the largest difference between $\hat{x}_{J_s}(k)$ and $\hat{x}_{S_s}(k)$ for all $k \geq 0$, and selects the observers leading to the smallest difference. Then, it can be proved that by using the selected observers for estimation, the estimator provides exponential estimate that is independent of sensor attacks.

This multi-observer approach can be extended to a large class of nonlinear systems if observers with certain stability properties exist. Moreover, it can also be extended to the case when actuators are under attack, or when sensor and actuator attacks both occur. This is the main topic of this thesis and we present an overview in the next section. It should be noted that secure estimation problem is combinatorial in nature since we never really know a priori the subset of sensors/actuators under attack, and hence, the complexity of the estimation scheme in Figure 1.8 grows exponentially with the growth

Figure 1.8: Multi-observer estimator for LTI systems under sensor attacks
of the number of sensors $n_y$. Therefore, the multi-observer estimator considered in this thesis work well for small or medium-scale systems.

1.3 Outline of the Thesis

We present below the outline of the thesis and summarise our contributions in each chapter. We assume in this thesis that attackers have limited resources that they can only compromise a sufficiently small subset of sensors/actuators, i.e., we always have redundant sensors/actuators for state observation. Moreover, we assume that the unknown sets of attacked sensors and actuators do not change over time. This assumption is valid and realistic since the time it takes for the attacker to gain control of another sensor/actuator is large compared to the time scale of the estimation algorithm.

In Chapter 2, we consider the problem of state estimation and attack isolation for noisy discrete-time nonlinear systems under sensor attacks of the form:

$$
\begin{align*}
    x^+ &= F(x,u,d), \\
    y_i &= g_i(x,u,m_i,a_i), i \in \{1, \ldots, p\},
\end{align*}
$$

(1.3)

with state $x \in \mathbb{R}^n$, input $u \in \mathbb{R}^{n_u}$, disturbance $d \in \mathbb{R}^s$, $\{d(k)\} \in l_\infty$, $i$-th sensor measurement $y_i \in \mathbb{R}$, stacked measurements $y := (y_1, \ldots, y_p)^\top \in \mathbb{R}^p$, attack signal $a_i \in \mathbb{R}$, measurement noise $m_i \in \mathbb{R}$, $\{m_i(k)\} \in l_\infty$, and nonlinear functions $F : \mathbb{R}^n \times \mathbb{R}^{n_u} \times \mathbb{R}^s \to \mathbb{R}^n$ and $g_i : \mathbb{R}^n \times \mathbb{R}^{n_u} \times \mathbb{R} \times \mathbb{R} \to \mathbb{R}$. If the $i$-th sensor is not attacked, $a_i(k) = 0$ for $k \geq 0$; otherwise, sensor $i$ is under attack and $a_i(k)$ is arbitrary and possibly unbounded. We develop a multi-observer estimator, using a bank of observers, which provides an ISS-like estimate of the system states with respect to measurement noise, process disturbance and independent of sensor attacks. To pinpoint the attacked sensors, we propose an estimator-based isolation algorithm by assuming measurement noise and system disturbance are bounded and their bounds are known. Compared with the existing estimation methods given in [28,54], which assume no system disturbances or noise occur and work for a class of nonlinear systems, our estimation framework can deal with a much larger class of nonlinear systems at the price of having to design multiple observers.
In Chapter 3, we consider noise-free discrete-time nonlinear systems subject to sensor and actuator attacks of the following form:

\[
\begin{align*}
    x^+ &= f(x) + B(u + a_u) \\
    y &= h(x) + a_y
\end{align*}
\]

with state \(x \in \mathbb{R}^n\), output \(y \in \mathbb{R}^n\), known input \(u\), actuator attack \(a_u \in \mathbb{R}^n_u\), i.e., \(a_{ui}(k) = 0, \forall k \geq 0\) if the \(i\)-th actuator is attack-free; otherwise, \(a_{ui}(k) \neq 0\) for some \(k \geq 0\) and can be arbitrarily large, sensor attack \(a_y \in \mathbb{R}^n_y\), i.e., \(a_{yi}(k) = 0, \forall k \geq 0\) if the \(i\)-th actuator is attack-free; otherwise, \(a_{yi}(k) \neq 0\) for some \(k \geq 0\) and can be arbitrarily large. Assume \(B\) has full column rank. We propose an unknown input multi-observer estimator that is able to provide robust estimate of the system states and similarly, attacked sensors and actuators are isolated by reconstructing the attack signals and checking their sparsity pattern when \(k\) is sufficiently large. Compared with the SMT-based estimator proposed in [56], where state estimation is performed by analyzing sensor information collected within a window of finite length, the unknown input observer-based estimators we consider are applicable to nonlinear systems and show a higher promise of scalability since observers incorporate real-time sensor measurements as soon as they become available. Hence, memory for storing sensor information for a period of time is not required in contrast to the SMT scheme.

In Chapter 4, we present a case study by addressing the problem of state estimation, attack detection and isolation for a class of noisy nonlinear systems with positive-slope nonlinearities. We introduce our previous work in [74,76] and explain how we use a bank of robust circle-criterion observers for reconstructing the state of the discrete-time nonlinear systems subject to sensor attacks with the following structure:

\[
\begin{align*}
    x^+ &= Ax + G f(Hx) + \rho(u, y), \\
    \tilde{y} &= \tilde{C}x + a + \tilde{m},
\end{align*}
\]

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 \(i \in \{1, \ldots, p\}\) 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. The nonlinearity $f(Hx)$ is an $r$-dimensional vector where each entry is a function of a linear combination of the states:

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

(1.6)

where $H_{ij}$ denotes the entries of the matrix $H$, and $f_i$ is globally Lipschitz for $i \in \{1, \ldots, r\}$.

We first consider the attack-free case and propose the design method of a robust circle-criterion observer that is ISS-like with respect to measurement noise with linear ISS gain minimized. Then we use our design method to construct a bank of circle-criterion observers and build a multi-observer estimator to provide robust estimate of system (1.5) in spite of sensor attacks. Moreover, we present observer-based algorithms for detection and isolating attacked sensors. Sufficient conditions under which the algorithms are guaranteed to work are provided; sufficient conditions under which our tools cannot work are also given.

In Chapter 5, the problem of state estimation, attack isolation and control is considered for discrete-time Linear Time Invariant (LTI) systems under actuator and sensor attacks of the form:

$$\begin{cases} x^+ = Ax + B(u + a_u) \\ y = Cx + a_y \end{cases}$$

(1.7)

with state $x \in \mathbb{R}^n$, output $y \in \mathbb{R}^n_y$, known input $u \in \mathbb{R}^n_u$, vector of actuator attacks $a_u \in \mathbb{R}^n_u, a_u = (a_{u1}, \ldots, a_{un_u})^\top$, i.e., $a_{ui}(k) = 0$ for all $k \geq 0$ if the $i$-th actuator is attack-free; otherwise, $a_{ui}(k) \neq 0$ for some $k_i \geq 0$ and can be arbitrarily large, and vector of sensor attacks $a_y \in \mathbb{R}^n_y, a_y = (a_{y1}, \ldots, a_{yn_y})^\top$, i.e., $a_{yi}(k) = 0$ for all $k \geq 0$ if the $i$-th sensor is attack-free; otherwise, $a_{yi}(k) \neq 0$ for some $k_i \geq 0$ and can be arbitrarily large. Matrices $A, B, C$ are of appropriate dimensions, and we assume that $(A, B)$ is stabilizable, $(A, C)$ is detectable, and $B$ has full column rank. We use a bank of Unknown Input Observers (UIOs) to reconstruct the system state and attack signals so that attacked actuators and sensors can be isolated by checking the sparsity pattern of the reconstructed attack signals when $k$ is sufficiently large. A switching output-feedback controller is designed to
stabilize the system by closing the system loop with the multi-observer estimator we provide. Compared to the adaptive controller proposed in [34], where a particular class of attacks is considered and ultimate boundedness of the closed-loop system is guaranteed only, our controller is able drive the system state asymptotically to the origin despite arbitrary and potentially unbounded attack signals.

In Chapter 6, we conclude the thesis by summarizing our main contribution. Some directions for further research are presented.
Chapter 2
Secure Estimation for Nonlinear Systems under Sensor Attacks

2.1 Overview

In this chapter, we address the problem of state estimation and attack isolation for general discrete-time nonlinear systems when sensors are corrupted by (potentially unbounded) attack signals. For a large class of nonlinear plants and observers, we provide a general estimation scheme, built around the idea of sensor redundancy and multi-observer, capable of reconstructing the system state in spite of sensor attacks and noise.

The core of our estimation scheme is based on the work in [15], where the problem of state estimation for continuous-time LTI systems is addressed. The authors propose a multi-observer estimator that we have introduced in Section 1.2.4, using a bank of Luenberger observers, which provides a robust estimate of the system state in spite of sensor attacks. In this chapter, we extend the results in [15,74,76] by considering discrete-time systems with general nonlinear dynamics. We cast the multi-observer estimation scheme in terms of the existence of a bank of (local and practical) nonlinear observers with Input-to-State-Stable (ISS) (with respect to process disturbance and measurement noise) estimator error dynamics. We consider the setting where the system has \( p \) sensors and up to \( q < p \) of them are attacked, i.e., we have redundant sensors for state observation. Following the multi-observer approach given in [15], we use a bank of observers to construct an estimator that provides a robust state estimate in the presence of false data injection attacks and noise.

The main idea behind the multi-observer estimator is the following: Each observer in...
the bank is driven by a different subset of sensors. Then, for every pair of observers in the bank, the estimator computes the difference between their estimates and selects the observers leading to the smallest difference. If there are attacks on some of the sensors, the observers driven by those sensors produce larger differences than the attack-free ones, in general, and thus they are not selected by the estimator. The main difference between this chapter and Section 1.2.4 is that here we consider nonlinear systems with general dynamics and nonlinear observers that are locally ISS with respect to process disturbance and measurement noise are used to construct our multi-observer estimator. We first consider the noise-free case and show that our estimator converges to the true state of the system in spite of sensor attacks. Next, we consider the case when process disturbances and measurement noise are present. Assuming each observer’s error is Input-to-State Stable (ISS) with respect to measurement noise and disturbances in the attack-free case, our estimator provides estimates whose errors satisfy an ISS-like property with respect to disturbances and independent of the attack signals. Compared to the estimation methods given in [28, 54], where no system disturbances and noise are considered, our estimation framework can deal with a much larger class of nonlinear systems at the price of having to design multiple observers.

Finally, we provide an algorithm for isolating attacked sensors using the proposed estimator and assuming that upper bounds on the process disturbance and measurement noise are known. The problem of attack isolation is not considered in [15]. The idea behind our isolation algorithm is: For each pair of observers, when driven by attack-free sensors, the largest difference between their estimates is proved to be bounded by a threshold that depends on the bounds on process disturbance and measurement noise. For every time-step, we select and take the union of all the subsets of sensors such that the corresponding threshold is not crossed; then, the remaining sensors are isolated as attacked ones. To improve the isolation performance, we carry out the isolation over windows of \( N \) time-steps. That is, we select the subset of sensors that are isolated most often in every time window as the attacked ones. In [64], [55], the problem of isolation of attacked sensors for LTI systems is addressed using the majority-vote method and satisfiability modulo theory, respectively. Compared to those results, our isolation algorithm can
be applied to nonlinear and noisy systems. Simulation results are presented to illustrate the performance of our tool.

### 2.2 Multi-Observer Estimator (Noise-free Case)

A multi-observer based estimator for continuous-time LTI systems has been proposed in [15]. Here, we generalize these results by considering general discrete-time nonlinear systems. Consider the nonlinear system

$$\begin{cases}
    x^+ = f(x, u), \\
y_i = h_i(x, u, a_i), \ i \in \{1, \ldots , p\}.
\end{cases} \quad (2.1)$$

with state $x \in \mathbb{R}^n$, input $u \in \mathbb{R}^{n_u}$, $i$-th sensor measurement $y_i \in \mathbb{R}$, stacked output $y := (y_1, \ldots , y_p)^\top \in \mathbb{R}^p$, attack signal $a_i \in \mathbb{R}$, stacked attack vector $a := (a_1, \ldots , a_p)^\top \in \mathbb{R}^p$, and functions $f : \mathbb{R}^n \times \mathbb{R}^{n_u} \to \mathbb{R}^n$ and $h_i : \mathbb{R}^n \times \mathbb{R}^{n_u} \times \mathbb{R} \to \mathbb{R}$. If the $i$-th sensor is not attacked, $a_i(k) = 0$ for $k \geq 0$; otherwise, sensor $i$ is under attack and $a_i(k)$ is arbitrary and possibly unbounded. The unknown set of attacked sensors is denoted as $W$, $W \subset \{1, \ldots , p\}$.

**Assumption 2.1.** The set of attacked sensors does not change over time, i.e., $W$ is constant (time-invariant) and $\text{supp}(a(k)) \subseteq W$, for all $k \geq 0$.

Consider the observer

$$\begin{cases}
    z^+_J = \Gamma_J(z_J, y^J, u), \\
    \hat{x}_J = \eta_J(z_J, y^J, u),
\end{cases} \quad (2.2)$$

where $y^J \in \mathbb{R}^{\text{Card}(J)}$ denotes the stacking of all $y_i$, $i \in J$, $J \subset \{1, \ldots , n\}$, $z_J \in \mathbb{R}^l$ is the observer state, $\hat{x}_J \in \mathbb{R}^n$ denotes the estimate of the plant state, and $\Gamma_J : \mathbb{R}^l \times \mathbb{R}^{\text{Card}(J)} \times \mathbb{R}^{n_u} \to \mathbb{R}^l$ and $\eta_J : \mathbb{R}^l \times \mathbb{R}^{\text{Card}(J)} \times \mathbb{R}^{n_u} \to \mathbb{R}^n$ are some functions.

**Definition 2.1.** (Local Asymptotic Practical Observer) *System~(2.2) is said to be a local*
Table 2.1: Systems/observers satisfying Definition 2.1 in the literature.

<table>
<thead>
<tr>
<th>Convergence</th>
<th>References</th>
</tr>
</thead>
<tbody>
<tr>
<td>Global exponential</td>
<td>[30, 42, 49, 60, 69, 73, 74]</td>
</tr>
<tr>
<td>Global asymptotic</td>
<td>[12, 70, 72, 78]</td>
</tr>
<tr>
<td>Local exponential</td>
<td>[29, 49, 61, 62]</td>
</tr>
<tr>
<td>Local asymptotic</td>
<td>[8, 16, 25, 31, 58]</td>
</tr>
<tr>
<td>Finite-time</td>
<td>[35, 49]</td>
</tr>
</tbody>
</table>

Asymptotic practical observer for system (2.1) if, for \( a^I(k) = 0, k \geq 0 \), there exists a set-valued map \( D_J(x) \subseteq \mathbb{R}^l \), such that, for any pair of initial conditions \((x(0), z_J(0)) \in \mathbb{R}^n \times D_J(x(0))\) and \( e_J(k) := \hat{x}_J(k) - x(k) \), there exist KL-function \( \beta_J(\cdot) \) and \( \nu_J \geq 0 \) satisfying: \(|e_J(k)| \leq \beta_J(|e_J(0)|, k) + \nu_J, k \geq 0\).

In this chapter, we assume that observers of form described in Definition 2.1 exist and are known for different subsets of sensors \( y^J, J \subseteq \{1, \ldots, p\} \). Any technique available in literature can be used to construct these observers as long as the corresponding convergence properties satisfy Definition 2.1. Note that all observers guaranteeing global (local) asymptotic convergence satisfy Definition 2.1 with \( \nu_J = 0 \). In Table 2.1, we present a list of publications where design methods for nonlinear observers satisfying Definition 2.1 are given. We also list the corresponding convergence properties that these observers guarantee. The results in this chapter apply to all the listed systems/observers.

**Assumption 2.2.** At most \( q \) sensors are attacked, i.e.,

\[
\text{card}(W) \leq q < \frac{p}{2}, \tag{2.3}
\]

where \( q \) denotes the largest integer such that for all \( I \subset \{1, \ldots, p\} \) with \( \text{card}(J) \geq p - 2q > 0 \), an observer of the form (2.2) exists for any \( y^I \in \mathbb{R}^{\text{card}(J)} \).

Following the ideas in [15], we use a local asymptotic practical observer for each subset \( J \subset \{1, \ldots, p\} \) of sensors with \( \text{card}(J) = p - q \) and for each subset \( S \subset \{1, \ldots, p\} \) with \( \text{card}(S) = p - 2q \). By Assumption 2.2, among the \( p \) sensors, there exists at least one subset of sensors \( \bar{I}, \bar{I} \subset \{1, \ldots, p\} \), with \( \text{card}(\bar{I}) = p - q \) satisfying \( y^\bar{I} = h^\bar{I}(x,u) \), i.e., there is
2.2 Multi-Observer Estimator (Noise-free Case)

a set $\bar{I}$ of sensors that is attack-free and thus $a^I(k) = 0$ for all $k \geq 0$. Then, in general, the difference between estimate $\hat{x}_I(k)$ and the estimate $\hat{x}_S(k)$ given by any subset $S \subset \bar{I}$ with $\text{card}(S) = p - 2q$ is smaller than the other subsets $J$ with $\text{card}(J) = p - q$ and $a^J(k) \neq 0$. This motivates the following estimation strategy.

For each subset $J \subset \{1, \ldots, p\}$ with $\text{card}(J) = p - q$, define $\pi_J(k)$ as the largest deviation between the estimates $\hat{x}_J(k)$ and $\hat{x}_S(k)$ for any $S \subset J$ with $\text{card}(S) = p - 2q$:

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

for all $k \geq 0$, and define the sequence $\sigma(k)$ as

$$\sigma(k) := \arg\min_{J \subset \{1, 2, \ldots, p\} : \text{card}(J) = p - q} \pi_J(k).$$

Then, as proven below, the estimate indexed by $\sigma(k)$:

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

is an asymptotic attack-free estimate of the system state. The following result uses the terminology presented above.

**Theorem 2.1.** Consider system (2.1), observer (2.2), estimator (2.4)-(2.6), and the estimation error $e(k) = \hat{x}_{\sigma(k)}(k) - x(k)$. Let Assumption 2.1-2.2 be satisfied; then, there exist a constant $\nu \geq 0$ and a class KL-function $\beta(\cdot)$ satisfying:

$$|e(k)| \leq \beta(e_0, k) + \nu,$$

for all $k \geq 0$.

We omit the proof of Theorem 1 since we later prove a more general result in Section 2.3.

**Remark:** The optimization problem in (2.5) amounts to searching among a number of
26 Secure Estimation for Nonlinear Systems under Sensor Attacks

\[
\begin{pmatrix}
p \\
p - q
\end{pmatrix}
\] sets \( J \) to find the one with the smallest value of \( \pi_J(k) \). The complexity of finding the smallest number among a set of \( n \) numbers is \( O(n) \) [20]. Therefore, the complexity of (2.5) is \( O\left(\begin{pmatrix} p \\
p - q \end{pmatrix}\right)\).

\[\text{2.2.1 Application Examples}\]

In this subsection, we show the performance of the proposed estimation scheme for two classes of nonlinear systems and observers.

**High Gain Observers:** Consider the nonlinear system

\[
\begin{aligned}
\dot{x}^+ &= f(x), \\
y &= h(x) + a,
\end{aligned}
\]

(2.8)

with state \( x \in \mathbb{R}^n \), output \( y \in \mathbb{R}^p \), attack vector \( a \in \mathbb{R}^p \), and functions \( f: \mathbb{R}^n \to \mathbb{R}^n \) and \( h: \mathbb{R}^n \to \mathbb{R}^p \).

**Assumption 2.3.** The origin of (2.8) is locally stable [37].

Consider the observer

\[
\dot{\hat{x}}_J = f(\hat{x}_J) + K_J(y^l - h(\hat{x}_J)),
\]

(2.9)

with state estimate \( \hat{x}_J \in \mathbb{R}^n \) and observer gain matrix \( K_J \in \mathbb{R}^{n \times \text{card}(J)} \). The observer gain \( K_J \) is designed following the results in [61].

**Proposition 2.1.** Let Assumption 2.3 be satisfied and \( q \) be the largest integer such that for all \( J \subset \{1, \ldots, p\} \) with \( \text{card}(J) \geq p - 2q \) an observer of the form (2.9) for system (2.8) exists for any \( y^l \in \mathbb{R}^{\text{card}(J)} \). Then, for \( a^l(k) = 0, k \geq 0 \), there exists a set-valued map \( \mathcal{D}_J(x) \subseteq \mathbb{R}^n \), such that, for any \( (x(0),\hat{x}_J(0)) \in \mathbb{R}^n \times \mathcal{D}_J(x(0)) \), there are \( \lambda_J \in (0,1) \) and \( c_J > 0 \) satisfying \( |e_J(k)| \leq c_J \lambda^k_J |e_J(0)| \), \( k \geq 0 \), where \( e_J = \hat{x}_J - x \).

**Proof:** Proposition 2.1 follows from [61, Theorem 3]. By Proposition 2.1, system (2.8) with observer (2.9) satisfy Definition 2.1 with \( \beta(|e_J(0)|,k) = c_J \lambda^k_J |e_J(0)|, \nu_I = 0 \), and some
set-valued map $D_J(x)$. Hence, we can write the following corollary of Theorem 2.1 and Proposition 2.1.

**Corollary 2.1.** Consider system (2.8), observer (2.9), the estimator (2.4)-(2.6), and the corresponding estimation error $e(k) = \hat{x}_v(k) - x(k)$. Let Assumptions 2.2 be satisfied; then, there exist $c > 0$ and $\lambda \in (0, 1)$ satisfying: $|e(k)| \leq c\lambda^k e_0$, $k \geq 0$, for $e_0$ as defined in (2.7).

**Example 1:** Consider the following nonlinear system subject to sensor attacks

$$
\begin{align*}
x_1^+ &= x_1 - x_1^2 + x_2 x_1^2 - x_2^2 x_1^3, \\
x_2^+ &= -x_2, \\
y_1 &= 2x_1 + x_1^2, \\
y_2 &= x_1 + x_2 + a_2, \\
y_3 &= 2x_1 + x_2.
\end{align*}
$$

(2.10)

We have three sensors, i.e., $p = 3$. Using the design method given in [61], we have found that observers of the form (2.9) exist for each subset $J \subset \{1, 2, 3\}$ with $\text{card}(J) \geq 1$. By Assumption 2.2, $q = 1$, i.e., at most one sensor is attacked. We let $W = \{2\}$ and design an observer for each $J \subset \{1, 2, 3\}$ with $\text{card}(J) = 2$ and each $S \subset \{1, 2, 3\}$ with $\text{card}(S) = 1$. Therefore, totally $\binom{3}{2} + \binom{3}{1} = 6$ observers are designed. We fix the initial condition of the observers to $\hat{x}(0) = [0, 0]^\top$, select $(x_1(0), x_2(0)) \sim \mathcal{N}(0, 1)$, and let $a_2 \sim \mathcal{U}(-10, 10)$. For $k \in [0, 49]$, we use (2.9),(2.4)-(2.6) to construct $\hat{x}(k)$. The performance of the estimator is shown in Figure 2.1.

**Reduced Order Observers:** Consider the system

$$
\begin{align*}
x^+ &= Ax + f(x, y), \\
y &= Cx + a,
\end{align*}
$$

(2.11)

with state $x \in \mathbb{R}^n$, output $y \in \mathbb{R}^p$, attack $a \in \mathbb{R}^p$, matrices $A \in \mathbb{R}^{n \times n}$ and $C \in \mathbb{R}^{p \times n}$, and nonlinear function $f : \mathbb{R}^n \times \mathbb{R}^p \rightarrow \mathbb{R}^n$.

**Assumption 2.4.** $f(x, y)$ is globally Lipschitz in $x$. 

Consider the partial output vector
\[ y^I = C^I x + a^I \]
and attack \( a^I \), with \( y^I, a^I \in \mathbb{R}^{\text{card}(I)} \), and the reduced state
\[ \zeta_I = L_I x \in \mathbb{R}^{n - \text{card}(I)} \]
where \( L_I \in \mathbb{R}^{(n - \text{card}(I)) \times n} \) is such that
\[ \begin{bmatrix} L_I^T & (C^I)^T \end{bmatrix} \]
is nonsingular. Let
\[
(N_I, M_I) := \begin{bmatrix} L_I \\ C^I \end{bmatrix}^{-1} ;
\]
then, \( x = N_I \zeta_I + M_I y^I \), and we can write the dynamics of the reduced state \( \zeta_I \) as
\[
\zeta_I^+ = A_{L_I} \zeta_I + L_I \phi_I(\zeta_I, y^I) + B_{L_I} y^I,
\]
where \( A_{L_I} := L_I AN_I \in \mathbb{R}^{(n - \text{card}(I)) \times (n - \text{card}(I))} \), \( B_{L_I} := L_I AM_I \in \mathbb{R}^{(n - \text{card}(I)) \times \text{card}(I)} \), and function \( \phi_I(z_I, y^I) := f(N_I z_I + M_I y^I, y^I) \). Consider the reduced order observer
\[
\begin{cases}
  z_I^+ = A_{L_I} z_I + \phi_I(z_I, y^I) + B_{L_I} y^I + K_I (y^I^+ - C^I \hat{x}_I^+) , \\
  \hat{x}_I = N_I z_I + M_I y^I ,
\end{cases}
\]
2.2 Multi-Observer Estimator (Noise-free Case)

Figure 2.2: Estimated states $\hat{x}$ converges to the true states $x$ when $a_2 \sim \mathcal{U}(-10, 10)$. Legend: $\hat{x}$ (blue), true states (black)

with observer state $\hat{z}_J \in \mathbb{R}^{n - \text{card}(J)}$, estimated state $\hat{x}_J \in \mathbb{R}^n$, and observer matrix $K_J \in \mathbb{R}^{(n - \text{card}(J)) \times \text{card}(J)}$. We design $K_J$ following the results in [78].

**Proposition 2.2.** Let Assumption 2.4 be satisfied and $q$ be the largest integer such that for all $J \subset \{1, \ldots, p\}$ with $\text{card}(J) \geq p - 2q$ an observer of the form (2.13) for system (2.12) exists for any $y^l \in \mathbb{R}^{\text{card}(J)}$. Then, for $a^l(k) = 0$, $k \geq 0$, and any $(x(0), z_J(0)) \in \mathbb{R}^n \times \mathbb{R}^0$, there exists a KL-function $\beta_J(\cdot)$ satisfying: $|e_J(k)| \leq \beta_J(|e_J(0)|, k)$, $k \geq 0$, where $e_J = \hat{x}_J - x$.

**Proof:** Proposition 2.2 follows from [78, Theorem 4].

By Proposition 2.2, system (2.11) with observer (2.13) satisfy Definition 2.1 for some KL-function, $\nu_J = 0$, and set-valued map $D_J(x) = \mathbb{R}^n$. Hence, we can write the following corollary of Theorem 2.1 and Proposition 2.2.

**Corollary 2.2.** Consider system (2.11), observer (2.13), the estimator (2.4)-(2.6), and the corresponding estimation error $e(k) = \hat{x}_{\sigma(k)}(k) - x(k)$. Let Assumptions 2.2 be satisfied; then, there exists a class KL-function $\beta(\cdot)$ satisfying: $|e(k)| \leq \beta(e_0, k)$, $k \geq 0$, for $e_0$ as defined in (2.7).
Example 2: Consider the following nonlinear system under sensor attacks:

\[
\begin{align*}
    x^+ &= \begin{bmatrix} 0.5 & 0 & 0 & 0 \\ 0 & 0.8 & 1 & 0 \\ 0.5 & 0.1 & 0.3 & 0 \\ 0.3 & 1 & 0 & 0.5 \end{bmatrix} x + \begin{bmatrix} 1 \\ 0 \\ 0 \\ -1.25 \tanh x_4 - 0.6 \end{bmatrix}, \\
    y &= \begin{bmatrix} 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} x + \begin{bmatrix} 0 \\ a_2 \\ 0 \end{bmatrix}.
\end{align*}
\] (2.14)

Using the design method proposed in [78], we have found that observers of the form (2.13) exist for each subset \( J \subset \{1, 2, 3\} \) with \( \text{card}(J) \geq 1 \) and \( p = 3 \). By Assumption 2.2, \( q = 1 \), i.e., at most one sensor is attacked. For randomly selected initial conditions, we attack sensor two, i.e., \( W = \{2\} \), and let \( a_2 \sim U(-10, 10) \). We use (2.13), (2.4)-(2.6) to reconstruct \( x(k) \). The performance of the estimator is shown in Figure 2.2.

2.3 Robust Multi-Observer Based Estimator

The tools given in this section, generalize the results in [15, 74] by considering systems with general nonlinear dynamics, disturbances, and noise. Consider the system

\[
\begin{align*}
    x^+ &= F(x, u, d), \\
    y_i &= g_i(x, u, m_i, a_i), i \in \{1, \ldots, p\},
\end{align*}
\] (2.15)

with state \( x \in \mathbb{R}^n \), input \( u \in \mathbb{R}^{n_u} \), disturbance \( d \in \mathbb{R}^s \), \( \{d(k)\} \in l_\infty \), \( i \)-th sensor measurement \( y_i \in \mathbb{R} \), stacked measurements \( y := (y_1, \ldots, y_p)^T \in \mathbb{R}^p \), attack signal \( a_i \in \mathbb{R} \), measurement noise \( m_i \in \mathbb{R} \), \( \{m_i(k)\} \in l_\infty \), and nonlinear functions \( F : \mathbb{R}^n \times \mathbb{R}^{n_u} \times \mathbb{R}^s \rightarrow \mathbb{R}^n \) and \( g_i : \mathbb{R}^n \times \mathbb{R}^{n_u} \times \mathbb{R} \times \mathbb{R} \rightarrow \mathbb{R} \).

Consider the observer

\[
\begin{align*}
    z_j^+ &= \Gamma_j(z_j, y_j, u), \\
    \dot{x}_j &= \eta_j(z_j, y_j, u),
\end{align*}
\] (2.16)
where $z_J \in \mathbb{R}^{l_I}$ is the observer state, $\hat{x}_J \in \mathbb{R}^n$ denotes the state estimate, and $\Gamma_J : \mathbb{R}^{l_I} \times \mathbb{R}^{|J|} \times \mathbb{R}^n \rightarrow \mathbb{R}^{l_I}$ and $\eta_J : \mathbb{R}^{l_I} \times \mathbb{R}^{|J|} \times \mathbb{R}^n \rightarrow \mathbb{R}^n$ are some functions.

**Definition 2.2.** (Local ISS Practical Observer). System (2.16) is said to be a local ISS practical observer for system (2.15) if, for a \( J(k) = 0, k \geq 0 \), there exists a set-valued map $D_J(x) \subseteq \mathbb{R}^{l_I}$, such that for any pair of initial conditions \((x(0), z_J(0)) \in \mathbb{R}^n \times D_J(x(0))\) and $e_J = \hat{x}_J - x$, there exist a KL-function $\beta_J(\cdot)$, K-functions $\gamma_{1,J}(\cdot)$ and $\gamma_{2,J}(\cdot)$, and constant $v_J \geq 0$ satisfying:

$$
|e_J(k)| \leq \beta_J(|e_J(0)|, k) + \gamma_{1,J}(|m|^\infty) + \gamma_{2,J}(|d|^\infty) + v_J, \quad k \geq 0.
$$

We assume that observers of form given in Definition 2.2 exist and are known for different subsets of sensors $y^I$, $J \subseteq \{1, \ldots, p\}$. In Table 2.2, we present a list of references where design methods for nonlinear observers satisfying Definition 2.2 can be found. All these observers can be used to construct the proposed estimator.

**Assumption 2.5.** At most $q$ sensors are attacked, i.e.,

$$
\text{card}(W) \leq q < \frac{p}{2},
$$

where $q$ denotes the largest integer such that for all $J \subset \{1, \ldots, p\}$ with $\text{card}(J) \geq p - 2q > 0$, an observer of the form (2.16) exists for any $y^I \in \mathbb{R}^{\text{card}(J)}$.

**Theorem 2.2.** Consider system (2.15), observer (2.16), estimator (2.4)-(2.6), and the estimation error $e(k) = \hat{x}_{\sigma(k)}(k) - x(k)$. Let Assumptions 2.5 be satisfied; then, there exist a class KL-function $\beta(\cdot)$, class K-functions $\gamma_{1}(\cdot)$ and $\gamma_{2}(\cdot)$, and a constant $v \geq 0$ satisfying:

$$
\left\{ \begin{array}{l}
|e(k)| \leq \beta(e_0, k) + \gamma_1(||m||^\infty) + \gamma_2(||d||^\infty) + v, \\
e_0 := \max_{J: \text{card}(J) = p - q} \{|e_J(0)|, |e_S(0)|\}.
\end{array} \right.
$$

for all $k \geq 0$ and $\{m(k)\}, \{d(k)\} \in l_\infty.$
Table 2.2: Systems/observers satisfying Definition 2.2 in the literature.

<table>
<thead>
<tr>
<th>Convergence</th>
<th>References</th>
</tr>
</thead>
<tbody>
<tr>
<td>Global exponential</td>
<td>[42, 60, 69, 74]</td>
</tr>
<tr>
<td>Global asymptotic</td>
<td>[1]</td>
</tr>
</tbody>
</table>

**Proof:** Under Assumption 2.5, there exist at least one subset $I$ with $\text{card}(I) = p - q$ and $a^I(k) = 0$ for all $k \geq 0$. Then, by Definition 2.2, there exist a KL-function $\beta_I(\cdot)$, class K-functions $\gamma_{1,I}(\cdot)$ and $\gamma_{2,I}(\cdot)$, and $\nu_I \geq 0$ such that

$$|e_I(k)| \leq \beta_I(e_0, k) + \gamma_{1,I}(||m^I||_{\infty}) + \gamma_{2,I}(||d||_{\infty}) + \nu_I, \quad (2.20)$$

for all $k \geq 0$. For all $S \subset I$ with $\text{card}(S) = p - 2q$, there exist a KL-function $\beta_S(\cdot)$, class K-functions $\gamma_{1,S}(\cdot)$ and $\gamma_{2,S}(\cdot)$, and $\nu_S \geq 0$ such that

$$|e_S(k)| \leq \beta_S(e_0, k) + \gamma_{1,S}(||m^S||_{\infty}) + \gamma_{2,S}(||d||_{\infty}) + \nu_S, \quad (2.21)$$

for all $k \geq 0$, which yields

$$\pi_I(k) = \max_{S \subset I} |\hat{x}_I(k) - \hat{x}_S(k)|$$

$$= \max_{S \subset I} |\hat{x}_I(k) - x(k) + x(k) - \hat{x}_S(k)|$$

$$\leq |e_I(k)| + \max_{S \subset I} |e_S(k)|$$

$$\leq 2(\beta'(e_0, k) + \gamma'_1(||m^I||_{\infty}) + \gamma'_2(||d||_{\infty}) + \nu'), \quad (2.22)$$

for all $k \geq 0$, where

$$\gamma'_1(||m^I||_{\infty}) = \max_{S \subset I} \left\{ \gamma_{1,I}(||m^I||_{\infty}), \gamma_{1,S}(||m^I||_{\infty}) \right\},$$

$$\gamma'_2(||d||_{\infty}) = \max_{S \subset I} \left\{ \gamma_{2,I}(||d||_{\infty}), \gamma_{2,S}(||d||_{\infty}) \right\}.$$ 

Under Assumption 2.5, for each subset $J \subset \{1, \ldots, p\}$ with $\text{card}(J) = p - q$, there exists $\bar{S} \subset J$ with $\text{card}(\bar{S}) = p - 2q$ such that $a^{\bar{S}}(k) = 0$ for all $k \geq 0$, and there exist a KL-
2.3 Robust Multi-Observer Based Estimator

function $\beta_S(\cdot)$, class K-functions $\gamma_{1,S}(\cdot)$ and $\gamma_{2,S}(\cdot)$, and $\nu_S \geq 0$ such that

$$|e_S(k)| \leq \beta_S(e_0,k) + \gamma_{1,S}(||m_S||_\infty) + \gamma_{2,S}(||d||_\infty) + \nu_S,$$  \hspace{1cm} (2.23)

for all $k \geq 0$. From (2.4), by construction

$$\pi_{\sigma}(k) = \max_{S \supset \sigma(k) : \text{card}(S) = 2q} |\hat{x}_{\sigma}(k) - \hat{x}_S(k)|$$

using the above lower bound on $\pi_{\sigma}(k)$ and the triangle inequality, we have that

$$|e_{\sigma}(k)| = |\hat{x}_{\sigma}(k) - x(k)|$$
$$= |\hat{x}_{\sigma}(k) - \hat{x}_S(k) + \hat{x}_S(k) - x(k)|$$
$$\leq |\hat{x}_{\sigma}(k) - \hat{x}_S(k)| + |e_S(k)|$$
$$\leq \pi_{\sigma}(k) + |e_S(k)|$$
$$\leq \pi_1(k) + |e_S(k)|,$$  \hspace{1cm} (2.24)

for all $k \geq 0$. Hence, from (2.22) and (2.23), we have

$$|e_{\sigma}(k)| \leq 3(\beta_1(e_0,k) + \gamma_{1,1}(||m||_\infty) + \gamma_{2,1}(||d||_\infty) + \nu_1),$$  \hspace{1cm} (2.25)

for all $k \geq 0$, where

$$\gamma_{1,1}(||m||_\infty) = \max \\{ \gamma'_1(||m||_\infty), \gamma_{1,S}(||m||_\infty) \},$$

$$\gamma_{2,1}(||d||_\infty) = \max \\{ \gamma'_2(||d||_\infty), \gamma_{1,S}(||d||_\infty) \}.$$

Inequality (2.25) is of the form (2.19) with KL-function $\beta(e_0,k) = 3\beta_1(e_0,k)$, nonnegative constant $\nu = 3\nu_1$, and K-functions $\gamma_1(||m||_\infty) = 3\gamma_{1,1}(||m||_\infty)$, and $\gamma_2(||d||_\infty) = 3\gamma_{2,1}(||d||_\infty)$.

**Remark:** If $d$ and $m$ are independent stochastic noises with normal distribution rather than bounded deterministic, e.g., $d \sim \mathcal{N}(0,\sigma_1^2)$, $m \sim \mathcal{N}(0,\sigma_2^2)$, then (2.19) holds with
\[ ||d||_\infty = 4\sigma_1, ||m||_\infty = 4\sigma_2 \] and possibility 0.9998.

### 2.3.1 Application Example

**H\(_\infty\) Observers:** Consider the nonlinear system

\[
\begin{aligned}
    x^+ &= Ax + F(x, u) + Bd, \\
    y &= Cx + a,
\end{aligned}
\]

with state \(x \in \mathbb{R}^n\), control \(u \in \mathbb{R}^nu\), output \(y \in \mathbb{R}^p\), attack vector \(a \in \mathbb{R}^p\), process disturbance \(d \in \mathbb{R}^nu\), \(\{d(k)\} \in l_\infty\), matrices \(A \in \mathbb{R}^{n \times n}\), \(C \in \mathbb{R}^{p \times n}\), and nonlinear function \(F : \mathbb{R}^n \times \mathbb{R}^nu \rightarrow \mathbb{R}^n\).

**Assumption 2.6.** \(F(x, u)\) is locally Lipschitz with respect to \(x\) in a region \(D\) uniformly in \(u\), i.e., for all \(x_1, x_2 \in D\):

\[
|F(x_1, u^*) - F(x_2, u^*)| \leq \gamma_d |x_1 - x_2|,
\]

where \(u^*\) is any admissible control sequence and \(\gamma_d > 0\) is called the Lipschitz constant.

Consider the observer

\[
\hat{x}_J^+ = A\hat{x}_J + F(\hat{x}_J, u) + L_J(y^l - C^l\hat{x}_J),
\]

with estimated state \(\hat{x}_J \in \mathbb{R}^n\) and matrix \(L_J \in \mathbb{R}^{n \times \text{card}(J)}\). The observer gain \(L_J\) is designed following the results in [1].

**Proposition 2.3.** Let Assumption 2.6 be satisfied, and \(q\) be the largest integer such that for all \(J \subset \{1, \ldots, p\}\) with \(\text{card}(J) \geq p - 2q > 0\) an ISS practical observer of the form (2.28) for system (2.26) exists for any \(y^l \in \mathbb{R}^{\text{card}(J)}\). Then, for all \(k \geq 0\), and \((x(0), \hat{x}_J(0)) \in \mathbb{R}^n \times \mathbb{R}^n\), there exists a class KL-function \(\beta_J(\cdot, \cdot), \gamma_{2J} > 0\) satisfying

\[
|e_J^l(k)| \leq \beta_J(|e_J^l(0)|, k) + \gamma_{2J} ||d||_\infty, k \geq 0,
\]

**Proof:** Proposition 2.3 follows from [1].

By Proposition 2.3, system (2.26) with observer (2.28) satisfy Definition 2.2 for some KL-function, \(m = 0\), linear function \(\gamma_{2J}, v_J = 0\), and set-valued map \(D_J(x) = \mathbb{R}^n\). Hence,
we can write the following corollary of Theorem 2.2 and Proposition 2.3.

**Corollary 2.3.** Consider system (2.26), observer (2.28), and the estimator (2.4)-(2.6) with estimation error \( e(k) = \hat{x}(k) - x(k) \). Let Assumptions 2.1, 2.2, and 2.6 be satisfied. Then, there exist a class KL-function \( \beta(\cdot, \cdot) \) and constant \( \gamma_2 > 0 \) satisfying \( |e(k)| \leq \beta(e_0, k) + \gamma_2 ||d||_{\infty} \), for all \( e_0 \geq 0, k \geq 0, \{d(k)\} \in l_{\infty} \).

**Example 3** Consider the following nonlinear system subject to process disturbance and sensor attacks:

\[
\begin{align*}
x^+ &= \begin{bmatrix} 0.7 & 0.01 \\ 0 & 0.5 \end{bmatrix} x + \begin{bmatrix} 0.05 \sin(x_1 + x_2) \\ 0.1 \sin(x_1 + x_2) \end{bmatrix} + \begin{bmatrix} d_1 \\ d_2 \end{bmatrix}, \\
y &= \begin{bmatrix} 1 & 0 \\ 0.5 & 1 \end{bmatrix} x + \begin{bmatrix} 0 \\ a_3 \end{bmatrix},
\end{align*}
\]

(2.29)

with \( d_1, d_2 \sim \mathcal{U}(-0.1, 0.1) \). Using the design method proposed in [1], we have found that observers of the form (2.28) exist for each subset \( J \subset \{1, 2, 3\} \) with \( \text{card}(J) \geq 1 \) and \( p = 3 \). It follows that, by Assumption 2.5, \( q = 1 \), i.e., our estimator is able to reconstruct the system states if at most one sensor can be attacked. We attack the sensor three, i.e., \( W = \{3\} \). We fix the initial conditions of the observers at \( \hat{x}(0) = [0, 0]^T \), and select \( (x_1(0), x_2(0)) \sim \mathcal{N}(0, 1^2) \). We let \( a_3 \sim \mathcal{U}(-10, 10) \). For \( k \in [0, 199] \), we use (2.28), (2.4)-(2.6) to construct \( \hat{x}(k) \). The performance of the estimator is shown in Figures 2.3.

### 2.4 Isolation of Attacked Sensors

Consider system (2.15) and let \( q \) be the largest integer such that an observer of the form (2.16) satisfying Definition 2.2 exists for each subset \( J \subset \{1, \ldots, p\} \) with \( \text{card}(J) \geq p - 2q \).
Assumption 2.7. Bounds on the process disturbance $d$ and the sensor noise $m$ are known, i.e.,

$$||d||_\infty = \bar{d}, \ ||m||_\infty = \bar{m},$$  \hspace{1cm} (2.30)

where $\bar{d} \geq 0$ and $\bar{m} \geq 0$ are known constants.

To perform the isolation, we construct an observer satisfying Definition 2.2 for each subset $J \subset \{1, \ldots, p\}$ of sensors with $\text{card}(J) = p - q$ and each subset $S \subset \{1, \ldots, p\}$ with $\text{card}(S) = p - 2q$. Hence, by Definition 2.2, for $a^S(k) = 0, k \geq 0$, there exist a KL-function, $\beta_S(\cdot)$, K-functions, $\gamma_{1, S}(\cdot)$ and $\gamma_{2, S}(\cdot)$, and $\nu_S \geq 0$ satisfying:

$$|e_S(k)| \leq \beta_S(|e(0)|, k) + \gamma_{1, S}(\bar{m}) + \gamma_{2, S}(\bar{d}) + \nu_S,$$  \hspace{1cm} (2.31)

for all $k \geq 0$. Note that, there always exist a $k^*_S$ such that $\beta_S(|e(0)|, k) \leq \epsilon$, for any $\epsilon > 0$ and $k \geq k^*_S$. Then,

$$|e_S(k)| \leq \epsilon + \gamma_{1, S}(\bar{m}) + \gamma_{2, S}(\bar{d}) + \nu_S,$$  \hspace{1cm} (2.32)

for all $k \geq k^*_S$. Define $\bar{k}^* := \max_{J, S} \{k^*_j, k^*_S\}$. By Assumption 2.5, there are at most
2.4 Isolation of Attacked Sensors

$q$ sensors under attack; then, we know there exists at least one $\bar{I} \subset \{1, \ldots, p\}$ with $\text{card}(\bar{I}) = p - q$ such that $a^I(k) = 0, k \geq 0$, and

$$|e_I(k)| \leq \epsilon + \gamma_{1,I}(\bar{m}) + \gamma_{2,I}(\bar{d}) + v_I,$$

(2.33)

for all $k \geq k_1^*$. Then, we have

$$\pi_I(k) := \max_{S \subset I} |\hat{x}_I(k) - \hat{x}_S(k)|$$

$$= \max_{S \subset I} |\hat{x}_I(k) - x(k) + x(k) - \hat{x}_S(k)|$$

(2.34)

$$\leq |e_I(k)| + \max_{S \subset I} |e_S(k)|.$$

From (2.32) and (2.33), we obtain

$$\pi_I(k) \leq 2(\epsilon + \gamma'_{1,I}(\bar{m}) + \gamma'_{2,I}(\bar{d}) + \nu'_I),$$

for all $k \geq k^*$, where

$$\gamma'_{1,I}(\bar{m}) := \max_{S \subset I, \text{card}(S) = p - 2q} \{\gamma_{1,I}(\bar{m}), \gamma_{1,S}(\bar{m})\},$$

and

$$\gamma'_{2,I}(\bar{d}) := \max_{S \subset I, \text{card}(S) = p - 2q} \{\gamma_{2,I}(\bar{d}), \gamma_{2,S}(\bar{d})\}.$$

However, if the subset $J$ of sensors is under attack at time $k$, i.e., $a^J(k) \neq 0$, then $\hat{x}_J(k)$ and $\hat{x}_S(k)$ in $\pi_J(k)$ are more inconsistent and produce larger $\pi_J(k)$. Define

$$\bar{\pi}_J := 2(\epsilon + \gamma'_{1,J}(\bar{m}) + \gamma'_{2,J}(\bar{d}) + \nu'_J),$$

(2.35)

for each $J \subset \{1, \ldots, p\}$ with $\text{card}(J) = p - q$, where

$$\gamma'_{1,J}(\bar{m}) := \max_{S \subset J, \text{card}(S) = p - 2q} \{\gamma_{1,J}(\bar{m}), \gamma_{1,S}(\bar{m})\},$$
and

\[
\gamma'_{2,j}(d) := \max_{S \subset \{1, \ldots, p\} : \text{card}(S) = p - 2q} \{ \gamma_{2,j}(d), \gamma_{2,S}(d) \};
\]

then, \( \bar{\pi}_j \) can be used as a threshold to isolate attacked sensors. For all \( k \geq \bar{k}^* \), we select from all the subsets \( J \subset \{1, \ldots, p\} \) with \( \text{card}(J) = p - q \), the ones that satisfy

\[
\pi_j(k) \leq \bar{\pi}_j. \quad (2.36)
\]

Denote as \( \bar{W}(k) \) the set of sensors that we regard as attack-free at time \( k \). We construct \( \bar{W}(k) \) as the union of all subsets \( J \) satisfying (2.36):

\[
\bar{W}(k) := \bigcup_{J \subset \{1, \ldots, p\} : \text{card}(J) = p - q, \pi_j(k) \leq \bar{\pi}_j} J. \quad (2.37)
\]

Thus, the set \( \{1, \ldots, p\} \setminus \bar{W}(k) \) is isolated as the set of attacked sensors at time \( k \). Note, however, that, for small persistent attacks, it is still possible that for some \( k \geq \bar{k}^* \) and some \( J \subset \{1, \ldots, p\} \) with \( \text{card}(J) = p - q \), \( a^j(k) \neq 0 \) but (2.36) still holds. This implies that \( J \subset \bar{W}(k) \) even if \( a^j(k) \neq 0 \) and would result in wrong isolation at time \( k \). To improve the isolation performance, we carry out the isolation over windows of \( N \) time-steps, \( N \in \mathbb{N} \). That is, for each \( k \in [\bar{k}^* + (i - 1)N, \bar{k}^* + iN], i \in \mathbb{N} \), we compute and collect \( \bar{W}(k) \) for every \( k \) in the window and select the subset \( J \) with \( \text{card}(J) \geq p - q \) that is equal to \( \bar{W}(k) \) most often in the \( i \)-th window. We denote this \( J \) as \( J(i) \). Then, we select \( \{1, \ldots, p\} \setminus J(i) \) as the set of sensors under attack in the \( i \)-th window. This isolation strategy is stated in Algorithm 1.
Algorithm 1 Attack Isolation.

1. Design an observer satisfying Definition 2.2 for each subset $J \subset \{1, \ldots, p\}$ with $\text{card}(J) = p - q$ and each subset $S \subset \{1, \ldots, p\}$ with $\text{card}(S) = p - 2q$.
2. Initialize the counter variable $n_j(i) = 0$ for all $J$ with $\text{card}(J) \geq p - q$ and all $i \in \mathbb{Z}_{>0}$.
3. Compute $\tilde{n}_j$ for each $J$ with $\text{card}(J) = p - q$ as (2.35).
4. For $i \in \mathbb{Z}_{>0}$ and $\forall k \in [\bar{k}^* + (i - 1)N, \bar{k}^* + iN - 1]$, compute $\pi_j(k)$, $\forall J$ with $\text{card}(J) = p - q$, as

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

5. For all $k \in [\bar{k}^* + (i - 1)N, \bar{k}^* + iN - 1]$, take the union of all the subsets $J$ such that $\pi_j(k) \leq \pi_j$:

$$\bar{W}(k) = \bigcup_{J \subset \{1, \ldots, p\}: \text{card}(J) = p - q, \pi_j \leq \pi_j} J.$$ 

6. For $k \in [\bar{k}^* + (i - 1)N, \bar{k}^* + iN - 1]$, if $\bar{W}(k) = J$ for some $J$ with $\text{card}(J) \geq p - q$, then update its corresponding counter variable as $n_j(i) = n_j(i) + 1$.
7. For all $i \in \mathbb{Z}_{>0}$, select the subset $J$ with $\text{card}(J) \geq p - q$ that is equal to $\bar{W}(k)$ most often, i.e.,

$$J(i) = \arg \max_{J \in \{1, \ldots, p\}: \text{card}(J) \geq p - q} n_j(i).$$

8. For all $i \in \mathbb{Z}_{>0}$, the set of sensors under attack is given by $\bar{A}(i) = \{1, \ldots, p\} \setminus J(i)$.
9. For all $i \in \mathbb{Z}_{>0}$, return $\bar{A}(i)$.

Example 4: Consider the following nonlinear system subject to process disturbance and sensor attacks:

$$\begin{align*}
\begin{cases}
x^+ = 
\begin{bmatrix}
0.7 & 0.01 \\
0 & 0.5
\end{bmatrix} x + 
\begin{bmatrix}
0.05 \sin(x_1 + x_2) \\
0.1 \sin(x_1 + x_2)
\end{bmatrix} +
\begin{bmatrix}
d_1 \\
d_2
\end{bmatrix}, \\
y = 
\begin{bmatrix}
1 & 0 \\
0 & 1
\end{bmatrix} x +
\begin{bmatrix}
a_1 \\
a_2
\end{bmatrix}
\end{cases}
\end{align*}$$

(2.38)

with $d_1, d_2 \sim \mathcal{U}(-0.1, 0.1)$. Using the design method proposed in [1], we have found that observers of the form (2.28) exist for each subset $J \subset \{1, 2, 3\}$ with $\text{card}(J) \geq 1$ and $p = 3$. It follows that, by Assumption 2.5, $q = 1$. We design an observer for each $J \subset \{1, 2, 3\}$ with $\text{card}(J) = 2$ and each $S \subset \{1, 2, 3, 4\}$ with $\text{card}(S) = 1$. Therefore, in total, $(3^2 + 3^1) = 6$ observers are designed. We obtain their ISS gains by Monte Carlo simulations, initialize the observers at $\hat{x}(0) = x(0)$, select $(x_1(0), x_2(0))$ from a standard normal dis-
tribution, and fix $\epsilon = 0$. We let $N = 50, 100, 200$, and follow the evolution of Algorithm 1 for 1000 time-steps. We attack sensor three, i.e., $W = \{3\}$, and let $a_3 \sim U(-1, 1)$. The isolation results are shown in Figures 2.4. In this figure, for visualization only, we depict $\bar{A}_i = \emptyset$ (no isolated sensors) by sensor 0 being isolated in the $i$-th time window.

### 2.5 Conclusion

Following the idea of sensor redundancy and multi-observer in [15], a general estimation scheme has been proposed for a large class of nonlinear plants and observers, which provides robust estimate of the system state when a sufficiently small subset of sensors are corrupted by (potentially unbounded) attack signals and system plant as well as all sensors are affected by bounded noise. We have posed the multi-observer estimation scheme in terms of the existence of a bank of (local and practical) nonlinear observers with ISS (with respect to disturbances and noise) estimation error dynamics. We have proved that the proposed estimator provides ISS-like estimates of the system state with respect to disturbances only and independent of sensor attacks. This scheme has been proposed in [15], for linear systems/observers. Here, we have proposed a unifying framework.
for a much larger class of nonlinear systems/observers and provided the corresponding
stability properties that the estimator yields in the nonlinear case. Using the proposed
estimator, we have provided an isolation algorithm to pinpoint sensor attacks during fi-
nite time windows. Simulations results have been provided to illustrate the performance
of our tools.
Chapter 3

Estimation for Nonlinear Systems
Under Sensor and Actuator Attacks

3.1 Overview

In this chapter, similar to Chapter 2, we address the problem of state estimation, and
attack isolation of discrete-time nonlinear systems, but now we consider the noise-free
case and with possibility of actuator attacks as well, i.e., sensors and actuators are both
under (potentially unbounded) false data injection attacks. Using a bank of unknown
input observers, each observer leading to an asymptotically stable estimation error (in
the attack-free case), we propose an observer-based estimator that provides asymptotic
estimates of the system state in spite of actuator and sensor attacks.

Unknown input observers are dynamical systems capable of estimating the state of
the plant without using any input signals. If such an observer exists for different subsets
of sensors, then, using a bank of observers, we can perform state estimation and attack
isolation when a sufficiently small subset of sensors are attacked (even if all inputs are
under attack). The main idea behind our multi-observer estimator is similar to the esti-
mation strategy described in Section 1.2.4. The main difference is that we use UIO instead
of Luenberger observers to construct the multi-observer estimator. See Figure 3.1 for a
schematic description of the proposed estimator.

If there does not exist a UIO capable of reconstructing the system state when all in-
puts are unknown, but there are UIOs that can estimate the state when some inputs are
known and some unknown, then using a bank of these UIOs, we can use similar ideas
to perform state estimation and attack isolation at the price of only being able to isolate
when a sufficiently small subset of actuators and sensors are under attack. If the inputs assumed to be unknown to the UIO include all the attacked ones and the outputs it uses for estimation are attack-free, this UIO produces asymptotically stable estimation errors. Then, for every pair of UIOs in the bank, we compute the largest difference between their estimates and select the pair leading to the smallest difference. We prove that these observers provide asymptotic estimates of the system state. See Figure 3.2 for a schematic description of the proposed estimator.

Once we have an estimate of the state, we provide tools for reconstructing attack signals using model matching techniques. Attacked actuators and sensors are isolated by simply checking the sparsity of the estimated attack signals. Simulation results are provided to illustrate the performance of our tools.
### 3.2 Estimation

Consider the discrete-time nonlinear system under sensor and actuator attacks:

\[
\begin{align*}
x^+ &= f(x) + B(u + a_u) \\
y &= h(x) + a_y
\end{align*}
\]  

with state \( x \in \mathbb{R}^n \), output \( y \in \mathbb{R}^n_y \), known input \( u \in \mathbb{R}^n_u \), vector of actuator attacks \( a_u \in \mathbb{R}^n_u \), \( a_u = (a_{u1}, \ldots, a_{un_u})^\top \), i.e., \( a_{ui}(k) = 0 \) for all \( k \geq 0 \) if the \( i \)-th actuator is attack-free; otherwise, \( a_{ui}(k_i) \neq 0 \) for some \( k_i \geq 0 \) and can be arbitrarily large, and vector of sensor attacks \( a_y \in \mathbb{R}^n_y \), \( a_y = (a_{y1}, \ldots, a_{yn_y})^\top \), i.e., \( a_{yi}(k) = 0 \) for all \( k \geq 0 \) if the \( i \)-th sensor is attack-free; otherwise, \( a_{yi}(k_i) \neq 0 \) for some \( k_i \geq 0 \) and can be arbitrarily large. Let \( B \) have full column rank, \( W_u \subseteq \{1, \ldots, n_u\} \) denote the unknown set of actuators under attack, and \( W_y \subseteq \{1, \ldots, n_y\} \) be the unknown set of sensors under attack. We assume the following.

**Assumption 3.1.** The sets of attacked actuators and sensors do not change over time, i.e., \( W_u \subseteq \)
\{1, \ldots, n_u\}, W_y \subset \{1, \ldots, n_y\} \text{ are constant (time-invariant) and } \text{supp}(a_u(k)) \subseteq W_u, \text{ supp}(a_y(k)) \subseteq W_y, \text{ for all } k \geq 0.

### 3.2.1 Complete Unknown Input Observers

We first treat \(a_u\) in (3.1) as an unknown input to system (3.1) and consider an observer of the form:

\[
\hat{x}_{J_s}^+ = f_{J_s}(\hat{x}_{J_s}, u, y_{J_s}, (y_{J_s})^+) \tag{3.2}
\]

where \(\hat{x}_{J_s}\) is the observer state and \(f_{J_s} : \mathbb{R}^n \times \mathbb{R}^{n_u} \times \mathbb{R}^{\text{card}(J_s)} \times \mathbb{R}^{\text{card}(J_s)} \to \mathbb{R}^n\) denotes some nonlinear function. Define \(e_{J_s} = \hat{x}_{J_s} - x\). System (3.2) is a complete unknown input observer for system (3.1) if, for all \(a_u \in \mathbb{R}^{n_u}\), and \(a_{J_s}y(k) = 0, \forall k \geq 0\), there exist a KL-function \(\beta_{J_s}(\cdot, \cdot)\), such that:

\[
|e_{J_s}(k)| \leq \beta_{J_s}(|e_{J_s}(0)|, k) \tag{3.3}
\]

for all \(e_{J_s}(0) \in \mathbb{R}^n\) and \(k \geq 0\).

Let \(q\) be the largest integer such that for each \(y_{J_s} \in \mathbb{R}^{\text{card}(J_s)}\) with \(J_s \subset \{1, \ldots, n_y\}\) and \(\text{card}(J_s) \geq n_y - 2q \geq 0\), a complete UIO of the form (3.2) satisfying (3.3) exists.

**Assumption 3.2.** There are at most \(q\) sensors attacked by an adversary, i.e.,

\[
\text{card}(W_y) \leq q < \frac{n_y}{2}, \tag{3.4}
\]

where \(q\) denotes the largest integer such that for all \(J_s \subset \{1, \ldots, n_y\}\) with \(\text{card}(J_s) \geq n_y - 2q\), a complete UIO (3.2) exists for any \(y_{J_s} \in \mathbb{R}^{\text{card}(J_s)}\).

**Lemma 3.1.** Under Assumption 3.2, among each set of \(n_y - q\) sensors, at least \(n_y - 2q > 0\) of them are attack-free.

**Proof:** Lemma 3.1 follow trivially from Assumption 3.2. \(\blacksquare\)

We construct a complete UIO for each set \(J_s \subset \{1, \ldots, n_y\}\) with \(\text{card}(J_s) = n_y - q\) and for each set \(S_s \subset \{1, \ldots, n_y\}\) with \(\text{card}(S_s) = n_y - 2q\). Under Assumption 3.2, there exist at least one set \(\overline{J}_s \subset \{1, \ldots, n_y\}\) with \(\text{card}(\overline{J}_s) = n_y - q\) such that \(a_{\overline{J}_s}y(k) = 0, \forall k \geq 0\). Then, the estimate given by the UIO for \(\overline{J}_s\) is a correct estimate, and the estimates given
3.2 Estimation

by the UIOs for any \( S_s \subset J_s \) with \( \text{card}(S_s) \) which we denote as \( \hat{x}_{S_s} \) are consistent with \( \hat{x}_{J_s} \). This motivates the following estimation strategy: for each set \( J_s \) with \( \text{card}(J_s) = n_y - q \), we define \( \pi_{J_s}(k) \) as the largest deviation between \( \hat{x}_{J_s}(k) \) and \( \hat{x}_{S_s}(k) \) that is given by any \( S_s \subset J_s \) with \( \text{card}(S_s) = n_y - 2q \), i.e.,

\[
\pi_{J_s}(k) = \max_{S_s \subset J_s : \text{card}(S_s) = n_y - 2q} |\hat{x}_{J_s}(k) - \hat{x}_{S_s}(k)|. \tag{3.5}
\]

For all \( k \geq 0 \),

\[
\sigma_s(k) = \arg\min_{J_s \subset \{1, \ldots, n_y\} : \text{card}(J_s) = n_y - q} \pi_{J_s}(k), \tag{3.6}
\]

and then we say that, for all \( k \geq 0 \), the estimate given by \( \sigma_s(k) \) is a correct estimate,

\[
\hat{x}(k) = \hat{x}_{\sigma_s(k)}(k) \tag{3.7}
\]

where \( \hat{x}_{\sigma_s(k)}(k) \) represents the estimates given by \( \sigma_s(k) \). The follow result summarizes the above discussion.

**Theorem 3.1.** Consider system (3.1), observer (3.2), and the estimator (3.5)-(3.7). Let Assumptions 3.1-3.2 be satisfied and define the estimation error \( e(k) := \hat{x}_{\sigma_s(k)}(k) - x(k) \); then, there exists a KL-function \( \bar{\beta}(-, -) \) satisfying:

\[
\begin{cases}
|e(k)| \leq \bar{\beta}(e_0, k) \\
\quad e_0 := \max_{J_s : \text{card}(J_s) = n_y - q} \{ |e_{J_s}(0)|, |e_{S_s}(0)| \}, \tag{3.8}
\end{cases}
\]

for all \( k \geq 0 \).

**Proof:** Under Assumption 3.2, there exists at least one set \( J_s \) with \( \text{card}(J_s) = n_y - q \) such that \( a_{J_s}(k) = 0, \forall k \geq 0 \); then, there exists a KL-function \( \beta_{J_s}(-, -) \) such that

\[
|e_{J_s}(k)| \leq \beta_{J_s}(e_0, k), \tag{3.9}
\]

for all \( e_0 \in \mathbb{R}^n \) and \( k \geq 0 \). Also for any set \( S_s \subset J_s \) with \( \text{card}(S_s) = n_y - 2q \), we have

...
\[ \dot{a}_y^S(k) = 0, \forall k \geq 0; \] hence, there exists a KL-function \( \beta_{S_s}(\cdot, \cdot) \) such that

\[ |e_{S_s}(k)| \leq \beta_{S_s}(e_0, k), \quad (3.10) \]

for all \( e_0 \in \mathbb{R}^n \) and \( k \geq 0 \). From the definition of \( \pi_{J_s} \) in (3.5), we can write the following

\[
\pi_{J_s}(k) = \max_{S_s \subset J_s} \left| \dot{x}_{J_s}(k) - \dot{x}_{S_s}(k) \right|
\]

\[
= \max_{S_s \subset J_s} \left| \dot{x}_{J_s}(k) - x(k) + x(k) - \dot{x}_{S_s}(k) \right|
\]

\[
\leq |e_{J_s}(k)| + \max_{S_s \subset J_s} |e_{S_s}(k)| \quad (3.11)
\]

for all \( k \geq 0 \). From (3.9) and (3.10), we obtain

\[
\pi_{J_s}(k) \leq 2\beta'_{J_s}(e_0, k), \quad (3.12)
\]

for all \( e_0 \in \mathbb{R}^n \) and \( k \geq 0 \), where

\[
\beta'_{J_s}(e_0, k) := \max_{S_s \subset J_s} \{ \beta_{J_s}(e_0, k), \beta_{S_s}(e_0, k) \}
\]

for all \( k \geq 0 \). From (3.6), \( \pi_{\sigma_s(k)}(k) \leq \pi_{J_s}(k) \). By Lemma 3.1, we know that there exists at least one set \( \tilde{S}_s \subset J_s \) with \( \text{card}(\tilde{S}_s) = n_y - 2q \) such that \( \dot{a}_y^S(k) = 0 \) for all \( k \geq 0 \), and there exists a KL-function \( \beta_{\tilde{S}_s} (\cdot, \cdot) \) such that

\[
|e_{\tilde{S}_s}(k)| \leq \beta_{\tilde{S}_s}(e_0, k), \quad (3.13)
\]

for all \( e_0 \in \mathbb{R}^n \) and \( k \geq 0 \). From (3.5), we have that

\[
\pi_{\sigma_s(k)}(k) = \max_{S_s \subset \sigma_s(k)} \left| \dot{x}_{\sigma_s(k)}(k) - \dot{x}_{S_s}(k) \right|
\]

\[
\geq \left| \dot{x}_{\sigma_s(k)}(k) - \dot{x}_{\tilde{S}_s}(k) \right|. 
\]
By the triangle inequality, we can write
\[
|e_{\sigma_s}(k)| = |\hat{x}_{\sigma_s}(k) - x(k)| \\
= |\hat{x}_{\sigma_s}(k) - \hat{x}_{\bar{S}_s}(k) + \hat{x}_{\bar{S}_s}(k) - x(k)| \\
\leq |\hat{x}_{\sigma_s}(k) - \hat{x}_{\bar{S}_s}(k)| + |e_{\bar{S}_s}(k)| \\
\leq \pi_{\sigma_s}(k) + |e_{\bar{S}_s}(k)| \\
\leq \bar{\beta}(e_0, k),
\]
(3.14)
for all \( k \geq 0 \). From (3.12) and (3.13), we have
\[
|e_{\sigma_s}(k)| \leq \bar{\beta}(e_0, k),
\]
(3.15)
for all \( e_0 \in \mathbb{R}^n \) and \( k \geq 0 \), where
\[
\bar{\beta}(e_0, k) = 3 \cdot \max \left\{ \beta_{\bar{S}_s}(e_0, k), \beta'_{\bar{J}_s}(e_0, k) \right\},
\]
for all \( k \geq 0 \). Inequality (3.15) is of the form (3.8) and the result follows. \( \blacksquare \)

Remark: The optimization problem in (3.6) amounts to searching among a number of \( \binom{n_y}{n_y - q} \) sets \( J_s \) to find the one with the smallest value of \( \pi_{J_s}(k) \). Therefore, the complexity of (3.6) is \( O \left( \binom{n_y}{n_y - q} \right) \).

3.2.2 Partial Unknown Input Observers

Let \( B \) be partitioned as \( B = [b_1, \ldots, b_i, \ldots, b_n] \) with \( b_i \in \mathbb{R}^{n_y \times 1} \). Then, system (3.1) can be written as
\[
x^+ = f(x) + Bu + b_{W_i} a^{W_i} \]
\[
y = h(x) + ay
\]
(3.16)
where the vector of attacks $a^W_u$ can be regarded as an unknown input to the system dynamics. The columns of $b^W_u$ are all $b_i$ such that $i \in W_u$. Let $(q_1, q_2)$ be the largest integers such that a partial unknown input observer of the form

$$
\dot{\hat{x}}_{J_u} = f_{J_u} (\hat{x}_{J_u}, u, y_{J_s}, (y^h)^+) 
$$

exists for each $b_{J_u}$ and $J_u \subset \{1, \ldots, n_u\}$ with $\text{card}(J_u) \leq 2q_1 < n_u$ and each $y^h$ with $\text{card}(J_s) \geq n_y - 2q_2 > 0$, where columns of $b_{J_u}$ are $b_i, i \in J_u$, i.e., an unknown input observer of the form (3.17) exists for the following system:

$$
x^+ = f(x) + Bu + b_{J_u} a^W_u \\
y^h = h^h(x) + a^l_y
$$

with known input $u$ and unknown input $a^W_u$. We refer to UIOs of the form (3.17) as partial UIOs for the pair $(J_u, J_s)$. We assume the following.

**Assumption 3.3.** At most $q_1$ actuators and $q_2$ sensors are under attack, i.e.,

$$
\text{card}(W_u) \leq q_1 < \frac{n_u}{2} \\
\text{card}(W_y) \leq q_2 < \frac{n_y}{2}
$$

where $q_1$ and $q_2$ denote the largest integer such that for any $J_u \subset \{1, \ldots, n_u\}$ with $\text{card}(J_u) \leq 2q_1$ and $J_s \subset \{1, \ldots, n_y\}$ with $\text{card}(J_s) \geq n_y - 2q_2$, a partial UIO of the form (3.17) exists for the pair $(J_u, J_s)$.

**Lemma 3.2.** Under Assumption 3.3, for each set of $q_1$ actuators, among all its supersets with $2q_1$ actuators, at least one set is a superset of $W_u$.

**Lemma 3.3.** Under Assumption 3.3, among each set of $n_y - q_2$ sensors, at least $n_y - 2q_2 > 0$ of them are attack-free.

**Proof:** Lemmas 3.2 and 3.3 follow trivially from Assumption 3.3. ■

We say that a UIO exists for each pair $(J_u, J_s)$ with $\text{card}(J_u) \leq 2q_1$ and $\text{card}(J_s) \geq n_y - 2q_2$,
if for $W_u \subseteq J_u$, $a_y^I(k) = 0$, and $k \geq 0$, there exists a $KL$-function $\beta_{f_u}(\cdot, \cdot)$ such that

$$|e_{f_u}(k)| \leq \beta_{f_u}(|e_{f_u}(0)|, k), \quad (3.21)$$

where $e_{f_u} = \hat{x}_{f_u} - x$. We construct a partial UIO for each $(J_u, I_s)$ with $\text{card}(J_u) = q_1$ and $\text{card}(I_s) = n_y - q_2$ and for each pair $(S_u, S_s)$ with $\text{card}(S_u) = 2q_1$ and $\text{card}(S_s) = n_y - 2q_2$. Then, under Assumption 3.3, there exists at least one set $\bar{J}_u$ with $\text{card}(\bar{J}_u) = q_1$ such that $W_u \subseteq \bar{J}_u$ and at least one set $\bar{I}_s$ with $\text{card}(\bar{I}_s) = n_y - q_2$ such that $a_y^I(k)(k) = 0$, for all $k \geq 0$. Thus, the estimate given by the UIO for $(J_u, I_s)$ is a correct estimate, and the estimates given by the UIOs for any $(S_u, S_s)$ where $S_u \supset \bar{J}_u$ with $\text{card}(S_u) = 2q_1$ and $S_s \subset \bar{I}_s$ with $\text{card}(I_s) = n_y - 2q_2$ (denotes as $\hat{x}_{S_u}$) are consistent with $\hat{x}_{f_u}$. This motivates the following estimation strategy: for each $(J_u, I_s)$ with $\text{card}(J_u) = q_1$ and $\text{card}(I_s) = n_y - q_2$, we define $\pi_{f_u}(k)$ as the largest deviation between $\hat{x}_{f_u}(k)$ and $\hat{x}_{S_u}(k)$ that is given by any $(S_u, S_s)$ where $S_u \supset \bar{J}_u$ with $\text{card}(S_u) = 2q_1$ and $S_s \subset \bar{I}_s$ with $\text{card}(S_s) = n_y - 2q_2$, i.e.,

$$\pi_{f_u}(k) := \max_{S_u \supset \bar{J}_u, S_s \subset \bar{I}_s} |\hat{x}_{f_u}(k) - \hat{x}_{S_u}(k)|. \quad (3.22)$$

for all $k \geq 0$, and

$$(\sigma_u(k), \sigma_s(k)) = \arg\min_{J_u, I_s} \pi_{f_u}(k); \quad (3.23)$$

then, we say that the estimate given by $(\sigma_u(k), \sigma_s(k))$ is a correct estimate, i.e.,

$$\hat{x}(k) = \hat{x}_{\sigma_u}(k)(k), \quad (3.24)$$

where $\hat{x}_{\sigma_u}(k)(k)$ denotes the estimate indexed by $(\sigma_u(k), \sigma_s(k))$. The following result summarizes the ideas above presented.

**Theorem 3.2.** Consider system (3.1), observer (3.17), and the estimator (3.22)-(3.24). Let Assumption 3.3 be satisfied and define the estimation error $e(k) = \hat{x}_{\sigma_u}(k)(k) - x(k)$; then, there
exists a KL-function \( \tilde{\beta}(\cdot, \cdot) \) satisfying:

\[
\begin{cases}
|e(k)| \leq \tilde{\beta}(e_0, k) \\
 e_0 := \max_{(J_u, J_s)} \{|e_{J_u}(0)|, |e_{J_s}(0)|\},
\end{cases}
\tag{3.25}
\]

for all \( e_0 \in \mathbb{R}^n, k \geq 0 \).

**Proof:** Under Assumption 3.3, there exist at least one set \( J_u \) with \( \text{card}(J) = q_1 \) such that \( J_u \supset W_u \) and at least one set \( J_s \) with \( \text{card}(J_s) = n_y - q_2 \) such that \( a_{J_s}^S(k) = 0, \forall k \geq 0 \), then, there exist a KL-function \( \beta_{J_us}(\cdot, \cdot) \), such that

\[
|e_{J_us}(k)| \leq \beta_{J_us}(e_0, k),
\tag{3.26}
\]

for all \( e_0 \in \mathbb{R}^n \) and \( k \geq 0 \). Also for any set \( S_u \supset J_u \) with \( \text{card}(S_u) = 2q_1 \) and \( S_s \subset J_s \) with \( \text{card}(S_s) = n_y - 2q_2 \), we have \( S_u \supset W_u \) and \( a_{J_s}^S(k) = 0, \forall k \geq 0 \), hence there exist a KL-function \( \beta_{S_us}(\cdot, \cdot) \), such that

\[
|e_{S_us}(k)| \leq \beta_{S_us}(e_0, k),
\tag{3.27}
\]

for all \( e_0 \in \mathbb{R}^n \) and \( k \geq 0 \). Recalling the definition of \( \pi_{J_us} \) from (3.22), we have that

\[
\pi_{J_us}(k) = \max_{S_u \supset J_u, S_s \subset J_s} \left| \hat{x}_{J_us}(k) - \hat{x}_{S_us}(k) \right| \\
= \max_{S_u \supset J_u, S_s \subset J_s} \left| \hat{x}_{J_us}(k) - x(k) + x(k) - \hat{x}_{S_us}(k) \right| \\
\leq |e_{J_us}(k)| + \max_{S_u \supset J_u, S_s \subset J_s} |e_{S_us}(k)|
\tag{3.28}
\]

for all \( k \geq 0 \). From (3.26) and (3.27), we obtain

\[
\pi_{J_us}(k) \leq 2 \beta_{J_us}(e_0, k),
\tag{3.29}
\]
for all \( e_0 \in \mathbb{R}^n \) and \( k \geq 0 \), where
\[
\beta_{Ju}^e(e_0, k) := \max_{S_u \supset J_u, S_s \subset J_s} \{ \beta_{Jus}^e(e_0, k), \beta_{Su}^e(e_0, k) \},
\]
for all \( k \geq 0 \). Recall from (3.23) that \( \pi_{\sigma u}(k) \leq \pi_{Jus}(k) \). From Lemmas 3.2, 3.3, we know that there exist at least one set \( \bar{S}_u \supset \sigma(k) \) with \( |\bar{S}_u| = 2q_1 \) and at least one set \( \bar{S}_s \subset \sigma(k) \) with \( |\bar{S}_s| = n_y - 2q_2 \) such that \( \bar{S}_u \supset W_u \) and \( a_y^\mathbf{\bar{S}}(k) = 0 \) for all \( k \geq 0 \), and there exist a class \( KL \)-function \( \beta_{\bar{S}_u}^\cdot(\cdot, \cdot) \), such that
\[
|e_{\bar{S}_u}(k)| \leq \beta_{\bar{S}_u}(e_0, k), \quad (3.30)
\]
for all \( e_0 \in \mathbb{R}^n \) and \( k \geq 0 \). From (3.22), there is a fact that
\[
\pi_{\sigma u}(k) = \max_{S_u \supset \sigma(k), S_s \subset \sigma(k)} |\hat{x}_{\sigma u}(k) - \hat{x}_{\bar{S}_u}(k)|
\]
\[
\geq |\hat{x}_{\sigma u}(k) - \hat{x}_{\bar{S}_u}(k)|.
\]

From the triangle inequality we have that
\[
|e_{\sigma u}(k)| = |\hat{x}_{\sigma u}(k) - x(k)|
\]
\[
= |\hat{x}_{\sigma u}(k) - \hat{x}_{\bar{S}_u}(k) + \hat{x}_{\bar{S}_u}(k) - x(k)|
\]
\[
\leq |\hat{x}_{\sigma u}(k) - \hat{x}_{\bar{S}_u}(k)| + |e_{\bar{S}_u}(k)|
\]
\[
\leq \pi_{\sigma u}(k) + |e_{\bar{S}_u}(k)| \leq \pi_{Jus}(k) + |e_{\bar{S}_u}(k)|
\]
for all \( k \geq 0 \). From (3.29) and (3.30), we have
\[
|e_{\sigma u}(k)| \leq \bar{\beta}(e_0, k), \quad (3.32)
\]
for all \( e_0 \in \mathbb{R}^n \) and \( k \geq 0 \), where
\[
\bar{\beta}(e_0, k) = 3 \cdot \max \left\{ \beta_{Su}^e(e_0, k), \beta_{Jus}^e(e_0, k) \right\}.
\]
(3.32) is of the form (3.25) and the result follows.

\[ O\left(\begin{pmatrix} n_y \\ n_y - q_2 \end{pmatrix} \times \begin{pmatrix} n_u \\ q_1 \end{pmatrix}\right). \]

**Remark:** The computation complexity of (3.23) is $O\left(\begin{pmatrix} n_y \\ n_y - q_2 \end{pmatrix} \times \begin{pmatrix} n_u \\ q_1 \end{pmatrix}\right)$.

### 3.2.3 An Application Example

Consider the nonlinear system:

\[
x^+ = Ax + f(x) + B(u + a_u)
y = Cx + a_y
\]  

(3.33)

with matrix $C \in \mathbb{R}^{n_y \times n}$ and nonlinear function $f : \mathbb{R}^n \rightarrow \mathbb{R}^n$ satisfying the following Lipschitz condition:

\[
|f(x_1) - f(x_2)| \leq \gamma|x_1 - x_2|, \forall x_1, x_2 \in \mathbb{R}^n,
\]

where $\gamma > 0$ denotes the Lipschitz constant. Consider a complete UIO of the form:

\[
\hat{x}^+_J = \bar{A}_J \hat{x}_J + \bar{B}_J u + \bar{f}_J(\hat{x}_J) + K_J(y^h - C^h \hat{x}_J) + B_J y^h
\]  

(3.35)

where $K_J \in \mathbb{R}^{n_x \times \text{card}(J)}$ is the observer gain. Let $H_J := (C^h B)^{-1} I - B H_J C^h$, $\bar{A}_J := \bar{G} J A$, $\bar{B}_J := \bar{G} J B$, and $f_J(\cdot) = \bar{G} J f(\cdot)$. If for all $J \subset \{1, \ldots, n_y\}$ with $\text{card}(J) \geq n_y - 2q$, it is satisfied that $\text{rank}(C^h B) = n_y$; then, complete UIOs can be designed using the tools given in [68] for all $y^h$ with $\text{card}(J) \geq n_y - 2q$. Using the estimation strategy (3.5)-(3.7) and Theorem 3.1, we can conclude that (3.8) is satisfied for all $e_0 \in \mathbb{R}^n$ and $k \geq 0$. If $n_y - 2 < n_u$, then, complete UIOs cannot be designed for any $y^h$ with $\text{card}(J) = n_y - 2$ using the design methods given in [68]. Then, in that case, consider partial UIOs of the form:

\[
\hat{x}^+_J = \bar{A}_{J_{us}} \hat{x}_{J_{us}} + \bar{B}_{J_{us}} u + \bar{f}_{J_{us}}(\hat{x}_{J_{us}}) + K_{J_{us}}(y^h - C^h \hat{x}_{J_{us}}) + B_{J_{us}} y^h
\]  

(3.36)
Let $H_{J_u} := (C^h b_{J_u})^{-1}$, $\bar{G}_{J_u} := I - b_{J_u} H_{J_u} C^h$, $\bar{b}_{J_u} := b_{J_u} H_{J_u}$, $A_{J_u} := \bar{G}_{J_u} A$, $B_{J_u} := \bar{G}_{J_u} B$, and $\bar{f}_{J_u} (\cdot) := \bar{G}_{J_u} f (\cdot)$. If for all $J_u \subset \{1, \ldots, n_u\}$, $\text{card}(J_u) \leq 2q_1$, and $J_s \subset \{1, \ldots, n_y\}$, $\text{card}(J_s) \geq n_y - 2q_2$, it is satisfied that $\text{rank}(C^h b_{J_u}) = \text{rank}(b_{J_u}) = \text{card}(J_u)$; then, partial UIOs can be designed using the method given in [68], for all $(J_u, J_s)$ with $\text{card}(J_u) \leq 2q_1$, $\text{card}(J_s) \geq n_y - 2q_2$. Under Assumption 3.3, using the estimation strategy (3.22)-(3.24) and Theorem 3.2, we can conclude that (3.25) is satisfied for all $e_0 \in \mathbb{R}^n$ and $k \geq 0$.

**Example 1.** Consider the nonlinear system under sensor and actuator attacks:

\[
\begin{align*}
    x^+ &= \begin{bmatrix} 0.2 & 0.5 \\ 0.2 & 0.7 \end{bmatrix} x + \begin{bmatrix} 0.5 \sin x_1 \\ 0.5 \sin x_2 \end{bmatrix} + \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} (u + a_u) \\
    y &= \begin{bmatrix} 1 & 3 \\ 1 & 1 \\ 3 & 2 \\ 4 & 1 \end{bmatrix} x + a_y 
\end{align*}
\] (3.37)

Using the method given in [68], a complete UIO can be designed for each $y^{J_s}$ with $\text{card}(J_s) \geq 2$. Therefore, we have $q = 1$. We let $W_u = \{1, 2\}$, which means both actuators are under attack, and $W_y = \{2\}$, which means the 2-nd sensor is under attack. We let $(u_1, u_2) \sim \mathcal{U}(-5, 5)$, $(a_{u1}, a_{u2}, a_{y2}) \sim \mathcal{U}(-c, c)$ with $c = 1, 10$. Then, we design a complete UIO for each $J_s \subset \{1, 2, 3, 4\}$ with $\text{card}(J_s) = 3$ and each $S_y \subset \{1, 2, 3, 4\}$ with $\text{card}(S_y) = 2$. Therefore, totally $\binom{4}{3} + \binom{4}{2} = 10$ complete UIOs are designed, which are all initialized at $[0, 0]^\top$. For all $k \geq 0$, (3.5) – (3.7) is used to construct $\hat{x}(k)$. The performance of the estimator is shown in Figures 3.3-3.4.
Figure 3.3: Estimated states \( \hat{x} \) converges to the true states \( x \) when \( a_{u1}, a_{u2}, a_{y2} \sim U(-1, 1) \). Legend: \( \hat{x} \) (grey), true states (black)

**Example 2.** Consider the nonlinear system:

\[
x^+ = \begin{bmatrix} 0.5 & 0 & 0.1 \\ 0.2 & 0.7 & 0 \\ 1 & 0 & 0.3 \end{bmatrix} x + \begin{bmatrix} 0.5 \sin x_1 \\ 0.5 \sin x_2 \\ 0.5 \sin x_3 \end{bmatrix} + \begin{bmatrix} 1 & 0 & 1 \\ 1 & 1 & 0 \\ 0 & 1 & 1 \end{bmatrix} (u + a_u),
\]

\[
y = \begin{bmatrix} 1 & 3 & 1 \\ 1 & 1 & 1 \\ 3 & 2 & 1 \\ 4 & 1 & 1 \end{bmatrix} x + a_y.
\]

(3.38)

We have \( n_y = 4 \) and \( n_u = 3 \); then, \( n_y - 2 < n_u \) and it can be verified that complete UIOs cannot be designed for any \( y^J \) with \( \text{card}(J_s) = 2 \) using the design methods given in [68]. Instead, partial UIOs can be designed for each pair \( (J_u, J_s) \) with \( \text{card}(J_u) \leq 2 \) and \( \text{card}(J_s) \geq 2 \). We let \( q_1 = q_2 = 1 \), \( (u_1, u_2, u_3) \sim U(-1, 1) \), \( W_u = \{1\}, W_y = \{2\}, (a_{u3}, a_{y2}) \sim U(-c, c) \) with \( c = 1, 10 \). We construct a partial UIO for each set pair \( (J_u, J_s) \) with \( \text{card}(J_u) = 1 \), \( \text{card}(J_s) = 3 \) and each set pair \( (S_u, S_s) \) with \( \text{card}(S_u) = 2 \), \( \text{card}(S_s) = 2 \). Therefore, totally \( \binom{3}{1} \times \binom{3}{2} + \binom{3}{2} \times \binom{4}{2} = 30 \) partial UIOs are constructed and they are
3.3 Isolation of Attacks

Once we have an estimate $\hat{x}(k)$ of $x(k)$, either using the complete multi-observer estimator in Section 3.2.1 or the partial multi-observer estimator in Section 3.2.2, we can use these estimates, the system dynamics (3.1), and the known inputs to asymptotically reconstruct the attack signals. Note that $e = \hat{x} - x \Rightarrow x = \hat{x} - e \Rightarrow x^+ = \hat{x}^+ - e^+$. Then, the system dynamics (3.1) can be written in terms of $e$ and $\hat{x}$ as follows:

$$
\begin{align*}
\hat{x}^+ &= e^+ + A(\hat{x} - e) + f(\hat{x} - e) + B(u + a_u), \\
\downarrow \\
a_u &= B_{left}^{-1}(\hat{x}^+ - A\hat{x} - f(\hat{x} - e)) - u - B_{left}^{-1}(e^+ + Ae),
\end{align*}
$$

(3.39)

\[ \text{Figure 3.4: Estimated states } \hat{x} \text{ converges to the true states } x \text{ when } a_{u1}, a_{u2}, a_{y2} \sim U(-10, 10). \text{ Legend: } \hat{x} \text{ (grey), true states (black)} \]

initialized at $\hat{x}(0) = [0,0]^T$. For all $k \geq 0$, (3.22)-(3.24) is used to construct $\hat{x}(k)$. The performance of the estimator is shown in Figures 3.5-3.6.
Estimation for Nonlinear Systems Under Sensor and Actuator Attacks

Figure 3.5: Estimated states $\hat{x}$ converges to the true states $x$ when $a_{u1}, a_{y2} \sim U(-1, 1)$. Legend: $\hat{x}$ (grey), true states (black)

where, because $B$ has full column rank, $B^{-1}_{left}$ denotes the Moore-Penrose pseudoinverse of $B$. Similarly, we have

$$\begin{align*}
y &= Cx + a_y = C\hat{x} - Ce + a_y, \\
\text{⇓} \\
a_y &= y - C\hat{x} + Ce.
\end{align*}$$

(3.40)

First, consider the complete multi-observer in Section 3.2.1. Let the estimation error dynamics characterized by (3.5)-(3.7) be given by

$$e^+ = f_1(e, x, a_y, a_y^+, a_u),$$

(3.41)

where $f_1 : \mathbb{R}^n \times \mathbb{R}^n \times \mathbb{R}^{n_y} \times \mathbb{R}^{n_y} \times \mathbb{R}^{n_u} \to \mathbb{R}^n$ denotes some nonlinear function. That is, the estimation error dynamics is given by some nonlinear function of the state and the attack signals. However, in Theorem 3.1, we have proved that $e$ converges to the origin asymptotically. Hence, the terms depending on $e$ and $e^+$ in the expression for $a_u$ and $a_y$ in (3.39) and (3.40), respectively, vanish asymptotically and therefore, the following attack
3.3 Isolation of Attacks

Figure 3.6: Estimated states $\hat{x}$ converges to the true states $x$ when $a_{u1}, a_{y2} \sim \mathcal{U}(-10, 10)$. Legend: $\hat{x}$ (grey), true states (black)

estimates:

$$\hat{a}_u(k) = B_{1e1}^{-1}(\hat{x}(k) - A\hat{x}(k - 1) - f(\hat{x}(k - 1))) - u(k - 1), \quad (3.42)$$

and

$$\hat{a}_y(k) = y(k) - C\hat{x}(k), \quad (3.43)$$

reconstruct the attack signals $a_u(k - 1)$ and $a_y(k)$, i.e.,

$$\lim_{k \to \infty} (\hat{a}_u(k) - a_u(k - 1)) = 0, \quad (3.44)$$

and

$$\lim_{k \to \infty} (\hat{a}_y(k) - a_y(k)) = 0. \quad (3.45)$$

Then, for sufficiently large $k$, the sparsity pattern of $\hat{a}_u(k)$ and $\hat{a}_y(k)$ can be used to isolate attacks at time $k$, i.e.,

$$\hat{W}_u(k) = \text{supp}(\hat{a}_u(k)), \quad (3.46)$$

and

$$\hat{W}_y(k) = \text{supp}(\hat{a}_y(k)), \quad (3.47)$$
where \( \hat{W}_u(k) \) denotes the set of isolated attacked actuators, and \( \hat{W}_y(k) \) denotes the set of isolated attacked sensors. Note that we can only estimate \( a_u \) from \( \hat{x}^+ \) and \( e^+ \), which implies that we always have, at least, one-step delay for actuator attacks isolation.

Next, consider the partial multi-observer estimator given in Section 3.2.2. In this case, the attack vector \( a_u \) and \( a_y \) can also be written as (3.39) and (3.40), and the estimation error dynamics is given by some nonlinear difference equation characterized by the estimator structure in (3.22)-(3.24). Let the estimation error dynamics be given by

\[
e^+ = f_2(e, x, a_y, a_u^+, a_u),
\]

for some nonlinear function \( f_2 : \mathbb{R}^n \times \mathbb{R}^n \times \mathbb{R}^{ny} \times \mathbb{R}^{ny} \times \mathbb{R}^{nu} \to \mathbb{R}^n \). In Theorem 3.2, we have proved that \( e \) converges to the origin asymptotically. Hence, the attack estimate in (3.42) and (3.43) asymptotically reconstructs the attack signals. Again, the sparsity pattern of \( \hat{a}_u(k) \) and \( \hat{a}_y(k) \) can be used to isolate actuator and sensor attacks using (3.46) and (3.47).

**Example 3.** We consider model (3.37) in Example 1. We let \( q = 1, W_u = \{1, 2\}, W_y = \{2\}, (u_1, u_2) \sim \mathcal{U}(-5, 5), (a_{u1}, a_{u2}, a_{y2}) \sim \mathcal{U}(-c, c), \) with \( c = 1, 10, \) and \( (x_1(0), x_2(0)) \sim \mathcal{N}(0, 1) \). We run \( \left(\frac{3}{1}\right) + \left(\frac{3}{4}\right) = 10 \) complete UIOs initialized at \( \hat{x}(0) = [0, 0]^\top \). We reconstruct \( a_y \) and \( a_u \) from (3.43) and (3.42) in 19 time-steps. The performance of the attack estimation is shown in Figures 3.7-3.10. By checking the sparsity, actuators 1 and 2 and sensor 2 can be isolated as the attacked ones.

**Example 4.** We consider model (3.38) in Example 3.2.3. We let \( q_1 = q_2 = 1, W_u = \{3\}, W_y = \{2\}, (u_1, u_2, u_3) \sim \mathcal{U}(-5, 5), (a_{u3}, a_{y2}) \sim \mathcal{U}(-c, c), \) with \( c = 1, 10, \) and \( (x_1(0), x_2(0), x_3(0)) \sim \mathcal{N}(0, 1) \). We run \( \left(\frac{3}{2}\right) \times \left(\frac{4}{2}\right) \times \left(\frac{4}{1}\right) = 30 \) partial UIOs initialized at \( \hat{x}(0) = [0, 0]^\top \). We reconstruct \( a_y \) and \( a_u \) from (3.43)-(3.42) in 19 time-steps. The performance is shown in Figures 3.11-3.14. By checking the sparsity of \( a_y \) and \( a_u \), actuator 3 and sensor 2 can be isolated as the attacked ones.
3.4 Conclusion

We have addressed the problem of state estimation and attack isolation for discrete-time nonlinear systems under (potentially unbounded) actuator and sensor false data injection attacks. Using a bank of Unknown Input Observers (UIOs), we have proposed an estimator that reconstructs the system states and the attack signals. We use these estimates to control the system and isolate attacks. Simulation results are provided to illustrate our results.
Figure 3.8: Estimate of $a_u$ when $a_{u1}, a_{u2}, a_{y2} \sim \mathcal{U}(-1, 1)$.

Figure 3.9: Estimate of $a_y$ when $a_{u1}, a_{u2}, a_{y2} \sim \mathcal{U}(-10, 10)$. 
Figure 3.10: Estimate of $a_u$ when $a_{u1}, a_{u2}, a_{y2} \sim U(-10, 10)$.

Figure 3.11: Estimate of $a_y$ when $a_{u3}, a_{y2} \sim U(-1, 1)$. 
Figure 3.12: Estimate of $a_u$ when $a_{u3}, a_{y2} \sim \mathcal{U}(-1, 1)$.

Figure 3.13: Estimate of $a_y$ when $a_{u3}, a_{y2} \sim \mathcal{U}(-10, 10)$. 
Figure 3.14: Estimate of \( a_u \) when \( a_u, a_y \sim \mathcal{U}(-10, 10) \).
Chapter 4
Case Study: Systems with Positive-Slope Nonlinearities

4.1 Overview

In this chapter, we present a case study by addressing the problem of state estimation, attack detection and isolation for a class of nonlinear systems with positive-slope nonlinearities. We introduce our previous work in [74, 76] and explain how we use a bank of robust circle-criterion observers to construct a multi-observer estimator, as well as detect and isolate sensor attacks for the system we consider. We use the same estimation strategy as we have introduced in Chapter 2, but the observer-based detection algorithm that we provide in this chapter has not been presented in previous chapters. Furthermore, we give a deeper discussion about the tools we propose by giving the sufficient conditions under which our detection and isolation algorithms are guaranteed to work; we also provide sufficient conditions under which such methods cannot work.

We consider the setting when the system has \( p \) sensors, all of which are subject to measurement noise and up to \( q < p \) of them are attacked. We assume that \( q \) is known but the exact subset of sensors being attacked is unknown. First, we follow the results in [15] for linear systems by using a bank of circle-criterion observers [3, 21, 30, 60], each observer leading to an ISS estimation error, to construct an estimator that provides robust estimates of the system state in spite of sensor attacks. The idea of the estimation strategy is the same as the one provided in [15], which has been described in Section 1.2.4, but we consider a different class of systems/observers, i.e., a class of nonlinear systems with positive-slope nonlinearities and the circle-criterion observers. In addition, we also
address the problem of attack detection and isolation for this particular class of nonlinear systems. In the case when the bound on the measurement noise is known. Using a bank of circle-criterion observers, each observer leading to an ISS estimation error, we propose two algorithms for detecting and isolating false data sensor attacks, respectively. These algorithms make use of the ISS property of the circle-criterion observers to check whether the trajectories of observers are “consistent” with the attack-free trajectories of the system. The main idea behind our algorithms is the following. Each observer in the bank is driven by a different set 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 set 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.

In order to increase the performance of the detection and isolation algorithms that we provide, we perform the detection and isolation in every finite-length time window. For example, if there is one time instant in a time window that sensor attacks are detected by our detection algorithm, then it is stated that the sensors are under attack in this time window. For attack isolation, we select the subset of sensors that are isolated most often in each time window and isolate this subset of sensors as the one under attack. To design the observers in the bank, we give an extension to the result in [30] for designing robust discrete-time circle-criterion observers. In particular, we use the incremental multiplier technique introduced in [60] 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. Simulation results are presented to show the performance of our tools.

4.2 Robust circle-criterion observer

In [30], 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 [30]
by considering measurement noise. The design method follows the ideas in [60] where
the nonlinearity is characterized by an incremental multiplier matrix. We present an
extension to the result in [60] 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:

\[ \begin{align*}
    x^{+} &= Ax + Gf(Hx) + \rho(u, y), \\
    y &= Cx + m,
\end{align*} \]

(4.1)

with state \( x \in \mathbb{R}^n \), output \( y \in \mathbb{R}^{ny} \), sensor noise \( m \in \mathbb{R}^{ny} \), \{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:

\[ f_i = f_i \left( \sum_{j=1}^{n} H_{ij}x_j \right), \quad i = 1, \ldots, r, \]

(4.2)

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

**Assumption 4.1.** For all \( i \in \{1, \ldots, r\} \), the following holds

\[ \frac{f_i(v_i) - f_i(w_i)}{v_i - w_i} \geq 0, \forall v_i, w_i \in \mathbb{R}, v_i \neq w_i. \]

(4.3)

Consider the circle criterion observer:

\[ \dot{x}^{+} = A\dot{x} + Gf(H\dot{x} + K(C\dot{x} - y)) + L(C\dot{x} - y) + \rho(u, y) \]

(4.4)

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

\[ e^{+} = (A + LC)e - Lm + G\Delta f, \]

(4.5)
where

\[ \Delta f := f(\hat{q}) - f(\tilde{q}), \] (4.6)

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

\[ \Delta \tilde{q} := \hat{q} - \tilde{q} = (H + KC)e - Km. \] (4.7)

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.

**Proposition 4.1.** Consider system (4.1), for given \( c_3 \in (0, 1) \), suppose there exist matrix \( P \in \mathbb{R}^{n \times n} \) and \( P > 0 \), \( K \in \mathbb{R}^{r \times ny} \) and \( Y \in \mathbb{R}^{ny \times ny} \), an incremental multiplier matrix \( M \) for the nonlinearity \( f \), and scalars \( \mu > 0 \) and \( \mu_1 > 0 \) that satisfy the matrix inequalities:

\[
\begin{bmatrix}
-P & * \\
\Xi_{21} & \Xi_{22}
\end{bmatrix}
+ \begin{bmatrix}
0 & 0 \\
0 & \Gamma^T M \Gamma
\end{bmatrix} \leq 0,
\] (4.8)

\[
\begin{bmatrix}
P & I \\
I & \mu I
\end{bmatrix} \geq 0,
\]

where

\[
\Xi_{21}^T = \begin{bmatrix}
PA + YC & -Y & PG
\end{bmatrix},
\]

\[
\Xi_{22} = \begin{bmatrix}
(c_3 - 1)P & 0 & 0 \\
0 & -c_3 \mu_1 I & 0 \\
0 & 0 & 0
\end{bmatrix},
\] (4.9)

and

\[
\Gamma = \begin{bmatrix}
H + KC & -K & 0 \\
0 & 0 & 1
\end{bmatrix},
\] (4.10)

then the observer (4.4) characterized by gains \( L = P^{-1}Y \) and \( K \) has ISS error dynamics with a linear gain \( \gamma = \sqrt{\mu \mu_1} \) and an \( \exp - KL \) function with respect to \( m \).

**Proof.** The proof of Proposition 4.1 can be obtained from the proof of Theorem 1 in [60].
4.2 Robust circle-criterion observer

by letting $H = I$, $B = 0$, $D = I$, $D_q = 0$ and adding a new variable $\mu_1$ in $\Xi_{22}$.

From Proposition 4.1, we see that if we could solve (4.8) while minimizing $\sqrt{\mu_1}$, then the designed observer is robust to measurement noise. We take advantage of the results in [60] by using an incremental multiplier matrix to characterize the nonlinearity $f$ in the design of a robust circle-criterion observer, but we do not fix $\mu_1 = 1$ as a constant as [60] does. Hence, our observer could provide estimates more robust to measurement noise in some circumstances.

From (4.3), we have

$$(\dot{q} - \dot{\tilde{q}})^T (f(\dot{q}) - f(\dot{\tilde{q}})) \geq 0. \quad (4.11)$$

Recalling (4.6) and (4.7), we know $\Delta q^T \Delta f \geq 0 \forall \tilde{q} \in \mathbb{R}^r$ and $\forall \hat{q} \in \mathbb{R}^r$. Hence, any matrix

$$M = \kappa \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}, \quad (4.12)$$

with $\kappa > 0$ is an incremental multiplier matrix for $f$. The following linear matrix inequality is equivalent to (4.8).

**Lemma 4.1.** [60] For some matrix $Y_2 \in \mathbb{R}^{r \times n_y}$, consider the linear matrix inequality

$$
\begin{bmatrix} -P & \ast \\ \Xi_{21} & \Xi_{22} \end{bmatrix} + \begin{bmatrix} 0 & 0 \\ 0 & \Gamma_1^T \Gamma_1 + \Gamma_1^T \Gamma_2 + \Gamma_2^T \Gamma_2 \end{bmatrix} \leq 0, \quad (4.13)
$$

where $\Xi_{21}, \Xi_{22}$ are described in (4.9), 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 with

$$L = P^{-1} Y, K = \frac{Y_2}{\kappa}, \quad (4.14)$$

(4.13) and (4.8) are equivalent.

**Proof.** Recalling (4.10), we let $\Gamma = \Gamma_1 + \tilde{\Gamma}_2$ where $\tilde{\Gamma}_2 = \begin{bmatrix} KC & -K & 0 \\ 0 & 0 & 0 \end{bmatrix}$. Note that $M \tilde{\Gamma}_2 = \ldots$
Case Study: Systems with Positive-Slope Nonlinearities

\[
\begin{bmatrix}
0 & 0 & 0 \\
\kappa KC & \kappa K & 0
\end{bmatrix},
\text{with } K \text{ given by (4.14), we see that } \kappa K = Y_2. \text{ Therefore, } M\Gamma_2 = \begin{bmatrix} 0 & 0 & 0 \\ Y_2C & Y_2 & 0 \end{bmatrix} = \Gamma_2. \text{ As we have } \Gamma_2^T M\Gamma_2 = 0, \text{ thus }
\]

\[
\begin{align*}
\Gamma_1^T M\Gamma & = (\Gamma_1 + \tilde{\Gamma}_2)^T M(\Gamma_1 + \tilde{\Gamma}_2) \\
& = \Gamma_1^T M\Gamma_1 + \tilde{\Gamma}_1^T M\Gamma_2 + \tilde{\Gamma}_2^T M\Gamma_1 + \tilde{\Gamma}_2^T M\tilde{\Gamma}_2 \\
& = \Gamma_1^T M\Gamma_1 + \Gamma_1^T \Gamma_2 + \Gamma_2^T \Gamma_1
\end{align*}
\]

which implies that (4.13) and (4.8) are equivalent. The proof is complete. \(\square\)

By replacing (4.8) with (4.13) in Proposition 4.1, we obtain the following result.

**Theorem 4.1.** Consider the system (4.1), for given \(c_3 \in (0, 1)\), suppose there exist matrix \(P \in \mathbb{R}^{n \times n}\) and \(P > 0\), matrix \(Y \in \mathbb{R}^{n \times n_y}\), matrix \(Y_2 \in \mathbb{R}^{r \times n_y}\), scalars \(\kappa > 0, \mu > 0, \mu_1 > 0\) that satisfy the linear matrix inequalities

\[
\begin{align*}
- P & \preceq 0 \\
\Xi_{21} & \preceq 0 \\
\Xi_{22} & \preceq 0 \\
\Gamma_1^T M\Gamma_1 & + \Gamma_1^T \Gamma_2 + \Gamma_2^T \Gamma_1 & \leq 0, \\
\begin{bmatrix} P & I \\ I & \mu I \end{bmatrix} & \succeq 0,
\end{align*}
\]

where \(\Xi_{21}, \Xi_{22}\) are described in (4.9), \(M\) is given by (4.12), and \(\Gamma_1 = \begin{bmatrix} H & 0 & 0 \\ 0 & 0 & I \end{bmatrix}, \Gamma_2 = \begin{bmatrix} 0 & 0 & 0 \\ Y_2C & -Y_2 & 0 \end{bmatrix}\), then the observer (4.4) characterized by gains given in (4.14) has ISS error dynamics with a linear gain \(\gamma = \sqrt{\mu \mu_1}\) and an exp – KL function, which means there exist \(c > 0, \lambda \in (0, 1)\) such that

\[
|e(k)| \leq c\lambda^k |e(0)| + \gamma \|m\|_k,
\]

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

**Corollary 4.1.** A circle-criterion observer robust to measurement noise can be obtained by solving (4.16) while minimizing \(\mu + \mu_1\).
Theorem 4.1 provides a way to design the observer for (4.1). If we solve (4.16) while minimizing $\sqrt{\mu \mu_1}$, we obtain an observer robust to measurement noise. To make the objective function convex, we consider using $\mu + \mu_1$ instead as our objective function. We know that $(\mu + \mu_1)^2 \geq 4 \cdot \mu \mu_1$, which yields $\gamma = \sqrt{\mu \mu_1} \leq \frac{1}{2} (\mu + \mu_1)$ as $\mu, \mu_1$ are positive. Therefore, we can minimize the upper bounds of $\gamma$ by minimizing $\mu + \mu_1$. By solving (4.16) while minimizing $\mu + \mu_1$, we can obtain an observer that attenuates measurement noise. Since $c_3 \in (0, 1)$ is in a bounded set, we do a grid-search over $c_3$, i.e. we make a grid in $(0, 1)$ and for each grid point we solve (4.16) while minimizing $\mu + \mu_1$ and then we choose the $c_3$ that minimizes $\sqrt{\mu \mu_1}$. In our design method, besides regarding $\mu_1$ as a variable, we also do not assume $c_3$ is a fixed constant with a given value as [60] does, which makes our LMIs less conservative than those in [60] and further improves the robustness of our observer to measurement noise in some circumstances. We use the model used in Example 1 in [30] and compare their performance by introducing measurement noise $m$. All LMIs were solved using PENLAB [24] in MATLAB.

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

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

We let $\delta = 0.1$ and $\alpha = 1$. (4.18) can be rewritten in the form of (4.1) with Assumption 4.1 holding, see [30] for more details. We solve (4.16) while minimizing $\mu + \mu_1$ and doing a grid search over $c_3$ to obtain observer matrices $K, L, c_3 = 0.900$, and $\gamma = 0.924$. We obtain $K, L$ via solving the LMIs in [30]. We solve the LMIs in [60] by letting $c_3 = 0.500$, $M$ given as (4.12), $H = I$ and minimizing $\mu$ to obtain observer with $\gamma = 22.4$. We let $m \sim U(-0.5, 0.5)$. $x(0)$ is randomly selected from a normal distribution and $\hat{x}(0) = [0 \ 0]^T$. The performance of these observers are compared respectively in Figures 4.1-4.2.
4.3 Observer-based estimator under sensor attacks and sensor noise

In this section, we introduce a circle-criterion observer-based estimator for the class of systems described in Section 4.2. We assume that a small number of sensors are subject to sensor attacks:

\[
x^+ = Ax + Gf(Hx) + \rho(u, y),
\]
\[
\hat{y} = C\hat{x} + a + \hat{m},
\]

(4.19)

where \(\hat{y} \in \mathbb{R}^p\) is the vector of sensor measurement under attacks, \(\hat{m} \in \mathbb{R}^p\), \(\{\hat{m}(k)\} \in l_\infty\) is the measurement noise, and \(a \in \mathbb{R}^p\) is the vector of attacks. If sensor \(i \in \{1, \ldots, p\}\) 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 \(W \subseteq \{1, \ldots, p\}\) 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 \(\hat{y}\left(k; x(0), a_{[0,k]}, \hat{m}_{[0,k]}\right)\) as the output of the system at time \(k\) when the initial state is \(x(0)\) and the outputs are subject to measurement noise \(\hat{m}\) and sensor attacks \(a\). The nonlinearity \(f(\cdot)\) satisfies (4.2) and (4.3). 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-like 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.

In what follows, we use the same estimation strategy as proposed in [15], which is introduced in Section 1.2.4, whereas we use the robust circle-criterion observer that we design in Section 4.2 to construct the multi-observer estimator. For (4.19), let $0 < q < \frac{p}{2}$ be the largest integer such that for each set $J \subset \{1, \ldots, p\}$ of sensors with $\text{card}(J) \geq p - 2q$, a circle-criterion observer of the form:

$$\hat{x}_J^* = A\hat{x}_J + Gf(H\hat{x}_J + K_J(\tilde{C}_J\hat{x}_J - \tilde{y}_J)) + L_J(\tilde{C}_J\hat{x}_J - \tilde{y}_J) + \rho(u, y),$$

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 $e^+_k(k) = \dot{x}_k(k) - x(k)$ with the following dynamics

$$e^+_k = (A + L\dot{C}^i) e_k - L\dot{m}^l + G\Delta f_k,$$  \hspace{1cm} (4.21)

with $\Delta f_k := f(q) - f(\bar{q})$, $\bar{q}_k := Hx$, $\bar{q}_k := H\dot{x}_k + Kf(\bar{g}^l - \bar{g})$, $\bar{g}^l := \bar{C}^i x + \Delta m^l$ is ISS with a linear gain $\gamma_f$ and an $exp - KL$ function with respect to measurement noise $\dot{m}^l$. This implies that there exist $c_\ell > 0$, $\lambda_f \in (0, 1)$, $\gamma_f \geq 0$ satisfying:

$$|e_f(k)| \leq c_\ell \lambda_f |e_f(0)| + \gamma_f ||\dot{m}^l||_k,$$ \hspace{1cm} (4.22)

for all $e_f(0) \in \mathbb{R}^n$, $k \geq 0$, and $\dot{m}^l \in \mathbb{R}^{\text{card}(\ell)}$ with $\{\dot{m}^l(k)\} \in l_\infty$.

We make the following assumption.

**Assumption 4.2.** There are at most $q$ sensors attacked,

$$\text{card}(W) \leq q < \frac{p}{2}.$$ \hspace{1cm} (4.23)

Using the design method proposed in Section 4.2, we construct a circle-criterion observer for each set $J \subset \{1, \ldots, p\}$ with $\text{card}(\ell) = p - q$ and for each set $S \subset \{1, \ldots, p\}$ with $\text{card}(\ell) = p - 2q$. For each set $J$ with $\text{card}(\ell) = p - q$, we define $\pi_J(k)$ for all $k \geq 0$ as the largest deviation between the estimate $\hat{x}_J(k)$ and the estimate $\hat{\dot{x}}_S(k)$ that is given by any set $S \subset J$ with $\text{card}(\ell) = p - 2q$. That is

$$\pi_J(k) := \max_{S \subset J : \text{card}(\ell) = p - 2q} |\hat{x}_J(k) - \hat{\dot{x}}_S(k)|,$$ \hspace{1cm} (4.24)

By assumption 4.2, among the $p$ sensors, there is at least one set $\bar{I} \subset \{1, \ldots, p\}$ of sensors with $\text{card}(\bar{I}) = p - q$ that $\bar{g}^l(k) = C^l x(k) + \bar{m}^l(k)$ as $a^l(k) = 0$ for all $k \geq 0$; then, in general, all the estimates that appear in the definition of $\pi_J(k)$ are more consistent than those corresponding to the sets $J$ with $\text{card}(\ell) = p - q$ and $\bar{g}^l(k) = \bar{C}^l x(k) + a^l(k) + \bar{m}^l(k)$ with $a^l(k) \neq 0$. This motivates the following state estimation strategy: For all $k \geq 0$,

$$\sigma(k) := \arg \min_{J \subset \{1, \ldots, p\} : \text{card}(\ell) = p - q} \pi_J(k);$$ \hspace{1cm} (4.25)
then, we say that the estimate given by the set $\sigma(k)$ is a good estimate, i.e.,

$$\hat{x}(k) = \hat{x}_{\sigma(k)}(k),$$  \hspace{1cm} (4.26)

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 4.2.** Consider system (4.19), observer (4.20), the estimator (4.24)-(4.26), and define the estimation error $e(k) := \hat{x}(k) - x(k)$ with $\hat{x}(k)$ as in (4.26). Let Assumptions 4.1 and Assumption 4.2 be satisfied; then, there exist constants $\bar{c} > 0$, $\bar{\lambda} \in (0, 1)$, and $\bar{\gamma}_y \geq 0$ satisfying:

$$|e(k)| \leq \bar{c}\bar{\lambda}ke_0 + \bar{\gamma}_y||\tilde{m}||_k,$$

$$e_0 := \max_{J:\text{card}(J) = p-2q} \max_{S:\text{card}(S) = p-2q} \{|e_J(0)|, |e_S(0)|\}.$$

(4.27)

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

**Proof.** From the result of Section 4.2, we know for each subset $J \subset \{1, \cdots, p\}$ with card$(J) \geq p - 2q$, the observation error dynamics satisfies (4.22). Since $a_i(k) = 0$ for all $i \in \{1, \cdots, p\} \setminus W$ and $\forall k \geq 0$, we conclude for $J = \bar{I} \subseteq \{1, \cdots, p\} \setminus W$ with card$(\bar{I}) = p - q$, there exist $c_{\bar{I}} > 0$, $\lambda_{\bar{I}} \in (0, 1)$ and $\gamma_{\bar{I}} \geq 0$, such that

$$|e_{\bar{I}}(k)| \leq c_{\bar{I}}\lambda_{\bar{I}}^k \cdot e_0 + \gamma_{\bar{I}}||\tilde{m}_{\bar{I}}||_k,$$

(4.28)

for all $e_0 \in \mathbb{R}^n$ and $k \geq 0$. Also for any set $S \subseteq \bar{I}$ with card$(S) = p - 2q$, we have $a^S(k) = 0$ $\forall k \geq 0$, hence there exist $c_S > 0$, $\lambda_S \in (0, 1)$ and $\gamma_S \geq 0$ such that

$$|e_S(k)| \leq c_S\lambda_S^k \cdot e_0 + \gamma_S||\tilde{m}_S||_k,$$

(4.29)
for all $e_0 \in \mathbb{R}^n$ and $k \geq 0$. Recalling the definition of $\pi_I$ from (5.26), we have that

$$
\pi_I(k) = \max_{S \subset I} |\hat{x}_I(k) - \hat{x}_S(k)| = \max_{S \subset I} |\hat{x}_I(k) - x(k) + x(k) - \hat{x}_S(k)|
$$

(4.30)

$$
\leq |e_I(k)| + \max_{S \subset I} |e_S(k)|
$$

for all $k \geq 0$. From (4.28) and (4.29), we obtain

$$
\pi_I(k) \leq 2c_I^* \lambda_I^k \cdot e_0 + 2\gamma_I^k ||\hat{m}^I||_k,
$$

(4.31)

for all $e_0 \in \mathbb{R}^n$ and $k \geq 0$, where $c_I^* := \max_{S \subset I} \{c_I, c_S\}$, $\lambda_I^* := \max_{S \subset I} \{\lambda_I, \lambda_S\}$, and $\gamma_I^* := \max_{S \subset I} \{\gamma_I, \gamma_S\}$. Observe that since $S \subset \bar{I}$ with $\text{card}(S) = p - 2q$. Recall from (4.24)-(4.26) that $\hat{x}(k) = \hat{x}_{c(k)}(k)$ where $c(k) = \arg \min_{j \in \{1, 2, \ldots, p\} : \text{card}(S) = p - q} \pi_I(k)$, hence $\pi_{c(k)}(k) \leq \pi_I(k)$. We know that there exist at least one set $\bar{S} \subset o(k)$ with $\text{card}(S) = p - 2q$ such that $a^\bar{S}(k) = 0$ $\forall k \geq 0$, and there exist $c_{\bar{S}} > 0$, $\lambda_{\bar{S}} \in (0, 1)$ and $\gamma_{\bar{S}} \geq 0$ such that

$$
|e_{\bar{S}}(k)| \leq c_{\bar{S}} \lambda_{\bar{S}}^k \cdot e_0 + \gamma_{\bar{S}} ||\hat{m}^\bar{S}||_k,
$$

(4.32)

for all $e_0 \in \mathbb{R}^n$ and $k \geq 0$. From (4.24), there is a fact that

$$
\pi_{c(k)}(k) = \max_{S \subset c(k) : \text{card}(S) = p - 2q} |\hat{x}_{c(k)}(k) - \hat{x}_S(k)| \geq |\hat{x}_{c(k)}(k) - \hat{x}_S(k)|.
$$

From the triangle inequality we have that

$$
|e_{c(k)}(k)| = |\hat{x}_{c(k)}(k) - x(k)| = |\hat{x}_{c(k)}(k) - \hat{x}_S(k) + \hat{x}_S(k) - x(k)| \leq |\hat{x}_{c(k)}(k) - \hat{x}_S(k)| + |e_S(k)|
$$

(4.33)

$$
\leq \pi_{c(k)}(k) + |e_S(k)| \leq \pi_I(k) + |e_S(k)|
$$

for all $k \geq 0$. From (4.31) and (4.32), we have

$$
|e_{c(k)}(k)| \leq c_{\bar{S}} \lambda_{\bar{S}}^k \cdot e_0 + \gamma_{\bar{S}} \max \left\{ ||\hat{m}^\bar{S}||_k, ||\hat{m}^I||_k \right\},
$$

(4.34)
4.4 Detection and Isolation of sensor attacks

for all \( e_0 \in \mathbb{R}^n \) and \( k \geq 0 \), where \( \bar{c} = 3 \cdot \max \{ c_5, c'_i \} \), \( \bar{\lambda} = \max \{ \lambda_5, \lambda'_i \} \), \( \bar{\gamma} = 3 \cdot \max \{ \gamma_5, \gamma'_i \} \). Since \( ||\tilde{m}||_k \geq \max \{ ||\tilde{m}^5||_k, ||\tilde{m}^i||_k \} \). (4.34) is of the form (4.27) and the proof is complete. \( \square \)

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

\[
\begin{align*}
\dot{x}^+ &= \begin{bmatrix} 1 & \delta \\ 0 & 1 \end{bmatrix} x + \begin{bmatrix} \frac{1}{2} \delta \alpha \sin(x_1 + x_2) \\ \delta \alpha \sin(x_1 + x_2) \end{bmatrix}, \\
\tilde{y} &= \begin{bmatrix} 3 & 0.3 \\ 3 & 0.6 \\ 6 & 0.9 \\ 1.2 & 12 \end{bmatrix} x + a + \tilde{m}.
\end{align*}
\] (4.35)

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 (4.20) 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 4.2, 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 \( k \in [0, 500] \), (4.24)-(4.26) is used to construct \( \hat{x}(k) \). The performance of the designed estimator is shown in Figures 4.3-4.4.

### 4.4 Detection and Isolation of sensor attacks

In this section, we aim at detecting and isolating attacked sensors for system (4.2) under attacks satisfying Assumption 4.2.

#### 4.4.1 Detection

Let Assumption 4.2 hold, we assume the following:
Assumption 4.3. The bound on measurement noise is known, i.e.,

\[ ||\tilde{m}\|_{\infty} = \bar{m}, \]  

(4.36)

and \( \bar{m} > 0 \) is a known constant.

We construct a circle-criterion observer (4.20) for system (4.19), i.e., considering all sensors, and for each subset \( J \subset \{1, \ldots, p\} \) of sensors with \( \text{card}(J) = p-q \). The obtained estimates are denoted as \( \hat{x}_A \) and \( \hat{x}_J \), respectively. Define \( e_A = \hat{x}_A - x \) and let \( a(k) = 0 \) for all \( k \geq 0 \); then, there exist \( c_A > 0 \), \( \lambda_A \in (0,1) \), and \( \gamma_A \geq 0 \) such that

\[ |e_A(k)| \leq c_A \lambda_A^k |e_A(0)| + \gamma_A ||\tilde{m}\|_k, \]  

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

Because \( \lambda_A \in (0,1) \), it can be easily verified that, for every \( \epsilon > 0 \), there exist \( k^* \) such that

\[ c_A \lambda_A^k |e_A(0)| \leq \epsilon, \]  

for all \( k \geq k^* \), which implies \( |e_A(k)| \leq \epsilon + \gamma_A ||\tilde{m}\|_k \leq \epsilon + \gamma_A \bar{m}, \) for \( k \geq k^* \). Also, for each subset \( J \subset \{1, \ldots, p\} \) with \( \text{card}(J) = p-q \), if \( a_j(K) = 0 \) for all \( k \geq 0 \), there exist \( c_J > 0 \), \( \lambda_J \in (0,1) \), and \( \gamma_J \geq 0 \) such that

\[ |e_J(k)| \leq c_J \lambda_J^k |e_J(0)| + \gamma_J ||\tilde{m}^J\|_k, \]  

for all \( e_J(0) \in \mathbb{R}^n, k \geq 0 \), and \( \tilde{m}^J \in \mathbb{R}^{p-q}, \{\tilde{m}^J(k)\} \in l_{\infty} \). Because \( \lambda_J \in (0,1) \), there exists \( k^*_J \) such that

\[ c_J \lambda_J^k |e_J(0)| \leq \epsilon, \]  

for all \( k \geq k^*_J \), and thus \( |e_J(k)| \leq \epsilon + \gamma_J ||m^J\|_k \leq \epsilon + \gamma_J \bar{m}, \) for
Let $k \geq k_0^*$. Let
\[
  k^* := \max_{J \subset \{1, \ldots, p\} : \text{card}(J) = p-q} \{k^*_1, k^*_2\},
\]
and define
\[
  \pi(k) := \max_{J \subset \{1, \ldots, p\} : \text{card}(J) = p-q} |\hat{x}_A(k) - \hat{x}_J(k)|.
\]
Let $J(k) = \arg\max_{J \subset \{1, \ldots, p\} : \text{card}(J) = p-q} |\hat{x}_A(k) - \hat{x}_J(k)|$, for all $k \geq k^*$. Then, if sensors are attack-free, i.e. $a(k) = 0$ for all $k \geq 0$, we have
\[
  \pi(k) = |\hat{x}_A(k) - \hat{x}_J(k)|
  = |\hat{x}_A(k) - x(k) + x(k) - \hat{x}_J(k)|
  = |e_A(k) - e_J(k)|
  \leq |e_A(k)| + |e_J(k)|
  \leq 2(e + \bar{\gamma}\bar{m}),
\]
for all $k \geq k^*$, where
\[
  \bar{\gamma} := \max_{J \subset \{1, \ldots, p\} : \text{card}(J) = p-q} \{\gamma, \gamma_J\}.
\]
However, if sensors are under attack, i.e., \( a(k) \neq 0 \) for some \( k \geq 0 \); then, the estimates \( \hat{x}_A(k) \) and \( \hat{x}_f(k) \) in \( \pi(k) \) are likely to be inconsistent and thus lead to larger \( \pi(k) \) than the attack-free case. Define

\[
\bar{z} := 2(\varepsilon + \gamma \bar{m}); \tag{4.39}
\]

then, \( \bar{z} \) can be used as a threshold to detect sensor attacks for \( k \geq k^* \). However, it is still possible that for some \( k \geq k^* \) and \( a(k) \neq 0 \), inequality (4.38) still holds, which would result in non-detection. Then, to improve the detection rate, we perform the detection over windows of \( N \in \mathbb{N} \) time-steps. That is, for each \( k \in [k^* + (i - 1)N, k^* + iN], i \in \mathbb{N} \), we compute \( \pi(k) \) and compare it with \( \bar{z} \) for every \( k \) in the window. If there exists \( k_1 \in [k^* + (i - 1)N, k^* + iN], i \in \mathbb{N} \) such that \( \pi(k_1) > \bar{z} \), then we say that sensors are under attack in the \( i \)-th window. Otherwise, we say sensors are attack-free in this window. This is formally stated in Algorithm 2.

Algorithm 2 Attack Detection.

1: Design a circle-criterion observer for system (4.19) and for each subset \( J \subset \{1, \ldots, p\} \) with \( \text{card}(J) = p - q \).
2: Fix the window size \( N \in \mathbb{N} \).
3: Calculate \( \bar{z} \) as in (4.39).
4: For \( i \in \mathbb{Z}_{>0} \), calculate \( \pi(k) \) for \( k \in [k^* + (i - 1)N, k^* + iN - 1] \).
5: For \( i \in \mathbb{Z}_{>0} \), if \( \exists k_1 \in [k^* + (i - 1)N, k^* + iN - 1] \) such that \( \pi(k_1) > \bar{z} \), then sensor attacks occurs in the \( i \)-th window, and

\[
detection(i) = 1;
\]

otherwise, sensors are attack-free in the \( i \)-th window, and

\[
detection(i) = 0.
\]
6: Return \( detection(i) \).

Because our knowledge of \( ||m||_\infty \) might be conservative, we consider the case when the actual bound on measurement noise is smaller than \( \bar{m} \), i.e., \( ||m||_\infty = \tau \bar{m} \) and \( \tau \in (0, 1) \). We give a sufficient condition under which sensor attacks cannot be detected by Algorithm 2 in the \( i \)-th time window for a given \( N > 0 \).
Proposition 4.2. Given a time window length $N > 0$, if

$$||a||_{k^* + iN - 1} \leq (1 - \tau)\bar{m};$$ \hspace{1cm} (4.40)

where $\tau \in (0, 1)$, then, $\pi(k) \leq z$ for all $k \in [k^* + (i - 1)N, k^* + iN - 1]$ and sensor attacks cannot be detected by Algorithm 2 in the $i$-th time window.

Proof. For a given time window length $N > 0$, sensor attacks cannot be detected by Algorithm 2 in the $i$-th time window for $a(k) \neq 0$ for some $k \geq 0$, if we have

$$\pi(k) \leq z,$$ \hspace{1cm} (4.41)

for all $k \in [k^* + (i - 1)N, k^* + iN - 1]$. For $a(k) \neq 0$ for some $k \geq 0$, we have

$$\pi(k) \leq |e(k)| + |e_f(k)| \leq 2\epsilon + \gamma(||\tilde{m}||_{k^* + iN - 1} + ||a||_{k^* + iN - 1}) + \gamma_f(||\tilde{m}||_{k^* + iN - 1} + ||a||_{k^* + iN - 1}) \leq 2(\epsilon + \bar{\gamma}(||\tilde{m}||_{\infty} + ||a||_{k^* + iN - 1})) = 2(\epsilon + \bar{\gamma}(\tau\bar{m} + ||a||_{k^* + iN - 1})).$$

for all $k \in [k^* + (i - 1)N, k^* + iN - 1]$. It follows that the inequality (4.41) is satisfied for $a$ satisfying (4.40) for all $k \in [k^* + (i - 1)N, k^* + iN - 1]$ and thus sensor attacks cannot be detected by Algorithm 2 in the $i$-th time window.

Next, we give a sufficient condition under which sensor attacks can always be detected by Algorithm 2 in the $i$-th time window for a given $N > 0$.

Proposition 4.3. For a given time window length $N > 0$, if there exist $k_1 \in [k^* + (i - 1)N, k^* + iN - 1]$ such that

$$|e(k_1)| > 3(\epsilon + \bar{\gamma}\bar{m});$$ \hspace{1cm} (4.42)

then, $\pi(k_1) > z$ and thus sensor attacks can be detected by Algorithm 2 in the $i$-th time window.
Proof. For a given time window length $N > 0$, sensor attacks can be detected by Algorithm 2 in the $i$-th time window for $a(k) \neq 0$ for some $k \geq 0$, if there exist $k_1 \in [k^* + (i-1)N, k^* + iN - 1]$ such that $\pi(k_1) > \bar{z}$. Since there are at most $q$ sensors under attack, we know there exist at least one $\bar{I} \subset \{1, \ldots, p\}$ with $\text{card}(\bar{I}) = p - q$ such that $a^{\bar{I}}(k) = 0, \forall k \geq 0$, and

$$|e^{\bar{I}}(k)| \leq \epsilon + \gamma_{\bar{I}}||\bar{m}^{\bar{I}}||_k,$$

(4.43) for $k \geq k^*$. From (4.37), we know $\pi(k) \geq |e(k) - e^{\bar{I}}(k)|$ for $k \geq k^*$. If (4.42) holds, then

$$\pi(k_1) \geq ||e(k_1)| - |e^{\bar{I}}(k_1)||$$

$$> 3(\epsilon + \bar{\gamma} \bar{m}) - \epsilon - \gamma_{\bar{I}}||\bar{m}^{\bar{I}}||_{k_1}$$

$$> 2(\epsilon + \bar{\gamma} \bar{m}),$$

(4.44) which implies sensor attacks can be detected by Algorithm 2 in the $i$-th time window. \qed

**Example 3** Consider the discrete-time nonlinear system subject to measurement noise and sensor attacks:

$$x^+ = \begin{bmatrix} 1 & \delta \\ 0 & 1 \end{bmatrix} x + \begin{bmatrix} \frac{1}{2} \delta \alpha \sin(x_1 + x_2) \\ \delta \alpha \sin(x_1 + x_2) \end{bmatrix},$$

(4.45)

$$\bar{y} = \begin{bmatrix} 3 & 0.3 \\ 3 & 0.6 \\ 6 & 0.9 \\ 1.2 & 12 \end{bmatrix} x + a + \bar{m}.$$  

(4.46)

with $\delta = 0.1$, $\alpha = 1$, and $\bar{m}_i \sim \mathcal{U}(-0.5, 0.5)$ for $i \in \{1, 2, 3, 4\}$. Using the design method proposed in [74], we find that circle-criterion observers of the form (4.20) exist for each subset $J \subset \{1, 2, 3, 4\}$ with $\text{card}(J) \geq 1$. Since $p = 4$, by Assumption 4.2, the maximum number of attacks is $q = 1$. We design a circle-criterion observer for the whole system and for each $J \subset \{1, 2, 3, 4\}$ with $\text{card}(J) = 3$. Therefore, in total, $\binom{4}{3} + 1 = 5$ observers are designed. We obtain their ISS gains by montecarlo simulations. Theses eleven observers are initialized at $\hat{x}(0) = x(0)$ and $x_1(0), x_2(0)$ are randomly selected from a standard
normal distribution; thus, $\epsilon = 0$. We let $N = 50, 100, 200$ and evaluate Algorithm 1 and Algorithm 2 for 1000 time-steps. We let $W = \{2\}$, which means the 2-nd sensor is under attack, and $a_2 \sim \mathcal{U}(-c, c)$ with $c$ given by $0.7$ and $1$. We run Algorithm 2 with $\binom{q}{0} + 1 = 5$ observers. The detection results are shown in Figures 4.5-4.6.

### 4.4.2 Isolation

Let Assumptions 4.2 and 4.3 hold. To perform the isolation, we construct a circle-criterion observer using the design method given in Section 4.2 for each subset $J \subset \{1, \ldots, p\}$ of sensors with $\text{card}(J) = p - q$ and each subset $S \subset \{1, \ldots, p\}$ of sensors with $\text{card}(S) = p - 2q$. Hence, for $a^S(k) = 0$ for some $k \geq 0$, there exist $c_S > 0, \lambda_S \in (0, 1)$, and $\gamma_S \geq 0$ satisfying

$$|e_S^V(k)| \leq c_S \lambda_S^k |e(0)| + \gamma_S \|\tilde{m}_S\|_k,$$

for all $e(0) \in \mathbb{R}^n$ and $k \geq 0$. Note that, because $\lambda_S \in (0, 1)$, there always exist $k_S^*$ such that $c_S \lambda_S^k |e(0)| \leq \epsilon$, for any $\epsilon > 0$ and $k \geq k_S^*$. Define $\bar{k}^* := \max_{J,S} \left\{k_J^* \right\}$. For each subset $J$
with \( \text{card}(J) = p - q \), define \( \pi_I(k) \) as

\[
\pi_I(k) := \max_{S \subset I: \text{card}(S) = p - 2q} |\hat{x}_I(k) - \hat{x}_S(k)|. \tag{4.48}
\]

Since there are at most \( q \) sensors under attack, we know there exist at least one \( \bar{I} \subset \{1, \ldots, p\} \) with \( \text{card}(\bar{I}) = p - q \) such that \( a^{\bar{I}}(k) = 0, \forall k \geq 0 \). Define

\[
\pi_{\bar{I}}(k) := \max_{S \subset \bar{I}} |\hat{x}_{\bar{I}}(k) - \hat{x}_S(k)|
\]

\[
= \max_{S \subset \bar{I}} |\hat{x}_{\bar{I}}(k) - x(k) + x(k) - \hat{x}_S(k)|
\]

\[
\leq |e_{\bar{I}}(k)| + \max_{S \subset \bar{I}} |e_S(k)|. \tag{4.49}
\]

From (4.43) and (4.47), we obtain \( \pi_{\bar{I}}(k) \leq 2(e + \gamma_I ||\bar{m}||_k) \), for all \( k \geq \bar{k}^* \), where

\[
\gamma_{\bar{I}} := \max_{S \subset \bar{I}: \text{card}(S) = p - 2q} \{\gamma_I, \gamma_S\}.
\]
4.4 Detection and Isolation of sensor attacks

However, if the subset $J$ of sensors is under attack, i.e., $a^J(k) \neq 0$ for some $k \geq 0$, then $\hat{x}_J(k)$ and $\hat{x}_S(k)$ in $\pi_J(k)$ are more inconsistent and might produce larger $\pi_J(k)$. Define

$$z_J = 2(\epsilon + \gamma_{J}' \bar{m}),$$

(4.50)

for each $J \subset \{1, \ldots, p\}$ with $\text{card}(J) = p - q$, where

$$\gamma_{J}' := \max_{S \subset J: \text{card}(S) = p - 2q} \{\gamma_J, \gamma_S\};$$

then, $z_J$ can be used as a threshold to isolate attacked sensors. For all $k \geq \bar{k}^*$, we select out all the subsets $J \subset \{1, \ldots, p\}$ with $\text{card}(J) = p - q$ that satisfy

$$\pi_J(k) \leq z_J.$$  

(4.51)

Denote as $\bar{W}(k)$ the set of sensors that we regard as attack-free at time $k$. Then, $\bar{W}(k)$ is given as the union of all subsets $J$ such that (4.51) holds:

$$\bar{W}(k) := \bigcup_{J \subset \{1, \ldots, p\}: \text{card}(J) = p - q, \pi_J(k) \leq z_J} J.$$  

(4.52)

Thus, the set $\{1, \ldots, p\} \setminus \bar{W}(k)$ is isolated as the set of attacked sensors at time $k$. However, note that it is still possible that for some $k \geq \bar{k}^*$ and some $J \subset \{1, \ldots, p\}$ with $\text{card}(J) = p - q$, $a^J(k) \neq 0$ for some $k \geq 0$ but (4.51) still holds. This implies that $J \subset \bar{W}(k)$ even if $a^J(k) \neq 0$ for some $k \geq 0$ and would result in wrong isolation. Therefore, we perform the isolation over windows of $N \in \mathbb{N}$ time-steps. That is, for each $k \in [\bar{k}^* + (i - 1)N, \bar{k}^* + iN], i \in \mathbb{N}$, we compute and collect $\bar{W}(k)$ for every $k$ in the window and select the subset $J$ with $\text{card}(J) \geq p - q$ that is equal to $\bar{W}(k)$ most often in the $i$-th window. We denote this $J$ as $J(i)$. Then, we select $\{1, \ldots, p\} \setminus J(i)$ as the set of sensors under attack in the $i$-th window. This is formally stated in Algorithm 3.

Next, we give a sufficient condition under which none of the attacked sensors can be isolated by Algorithm 3 in the $i$-th time window for a given $N > 0$ when $||m||_\infty = \tau \cdot \bar{m}$ where $\tau \in (0, 1)$.
Algorithm 3 Attack Isolation.

1: Design a circle-criterion observer for each subset $J \subset \{1, \ldots, p\}$ with $\text{card}(J) = p - q$ and each subset $S \subset \{1, \ldots, p\}$ with $\text{card}(S) = p - 2q$.
2: Initialize the counter variable $n_J(i) = 0$ for all $J$ with $\text{card}(J) \geq p - q$ and all $i \in \mathbb{Z}_{>0}$.
3: Calculate $\bar{z}_J$ for each $J$ with $\text{card}(J) = p - q$ as (4.50).
4: For $i \in \mathbb{Z}_{>0}$ and $\forall k \in [\bar{k}^* + (i - 1)N, \bar{k}^* + iN - 1]$, calculate $\pi_J(k)$, $\forall J$ with $\text{card}(J) = p - q$ as follows:
$$\pi_J(k) = \max_{S \subset J, \text{card}(S) = p - 2q} |\hat{x}_J(k) - \hat{x}_S(k)|.$$
5: For all $k \in [\bar{k}^* + (i - 1)N, \bar{k}^* + iN - 1]$, take the union of all the subsets $J$ such that $\pi_J(k) \leq \bar{z}_J$:
$$\bar{W}(k) = \bigcup_{J \subset \{1, \ldots, p\}, \text{card}(J) = p - q, \pi_J(k) \leq \bar{z}_J} J.$$
6: For all $k \in [\bar{k}^* + (i - 1)N, \bar{k}^* + iN - 1]$, if $\bar{W}(k) = J$ for some $J$ with $\text{card}(J) \geq p - q$, then update its corresponding counter variable as follows:
$$n_J(i) = n_J(i) + 1.$$
7: For all $i \in \mathbb{Z}_{>0}$, select the subset $J$ with $\text{card}(J) \geq p - q$ that is equal to $\bar{W}(k)$ most often, i.e.,
$$J(i) = \arg\max_{J \in \{1, \ldots, p\}, \text{card}(J) \geq p - q} n_J(i).$$
8: For all $i \in \mathbb{Z}_{>0}$, the set of sensors under attack is given as:
$$\tilde{A}(i) = \{1, \ldots, p\} \setminus J(i).$$
9: For all $i \in \mathbb{Z}_{>0}$, return $\tilde{A}(i)$.
4.4 Detection and Isolation of sensor attacks

Proposition 4.4. Given a time window length $N > 0$, if

$$||a||_{k^*+iN-1} \leq (1 - \tau)\tilde{m}; \quad (4.53)$$

where $\tau \in (0, 1)$, then, for all $J \subset \{1, \ldots, p\}$ with $\text{card}(J) = p - q$ and $a^J(k) \neq 0$ for some $k \geq 0$, $\pi_J(k) \leq z_J$ for all $k \in [k^* + (i - 1)N, k^* + iN - 1]$ and none of attacked sensors can be isolated by Algorithm 3 in the $i$-th time window.

Proof. For given time window length $N > 0$, none of attacked sensors can be isolated by Algorithm 3 in the $i$-th window if $\forall J \subset \{1, \ldots, p\}$ with $\text{card}(J) = p - q$ and $a^J(k) \neq 0$ for some $k \geq 0$, we have $\pi_J(k) \leq z_J$ for all $k \in [k^* + (i - 1)N, k^* + iN - 1]$. For all $k \in [k^* + (i - 1)N, k^* + iN - 1]$ and $a^J(k) \neq 0$ for some $k \geq 0$, we have

$$\pi_J(k) \leq |e_f(k)| + |e_S(k)|$$

$$\leq 2\epsilon + \gamma_J(||\tilde{m}||_{\infty} + ||a^J||_{\infty})$$

$$+ \gamma_S(||\tilde{m}^S||_{k^*+iN-1} + ||a^S||_{k^*+iN-1})$$

$$\leq 2\epsilon + 2\gamma_J(||\tilde{m}||_{k^*+iN-1} + ||a^J||_{k^*+iN-1})$$

$$\leq 2(\epsilon + \gamma_J(\tau\tilde{m} + ||a||_{k^*+iN-1})).$$

If (4.53) holds, then $\pi_J(k) \leq z_J$ for all $J \subset \{1, \ldots, p\}$ with $\text{card}(J) = p - q$ and all $k \in [k^* + (i - 1)N, k^* + iN - 1]$. Then, none of attacked sensors can be isolated by Algorithm 3 in the $i$-th time window.

Next, we give a sufficient condition under which all of attacked sensors can be isolated by Algorithm 3 in the $i$-th time window for a given time window length $N > 0$.

Proposition 4.5. Given a time window length $N > 0$, if for all $J \subset \{1, \ldots, p\}$ with $\text{card}(J) = p - q$ and $a^J(k) \neq 0$ for some $k \geq 0$, we have

$$|e_f(k)| > 3(\epsilon + \gamma'_J\tilde{m}), \quad (4.54)$$

for at least $N/2$ time-steps in the $i$-th time window, then for all $J \subset \{1, \ldots, p\}$ with $\text{card}(J) =$
\( p - q \) and \( a^l(k) \neq 0 \) for some \( k \geq 0 \), we have \( \pi_J(k) > \bar{z}_J \) for at least \( N/2 \) time-steps in the \( i \)-th time window, and all of attacked sensors can be isolated by Algorithm 3 in the \( i \)-th time window.

**Proof.** Since there are at most \( q \) sensors under attack, for each subset \( J \) with \( \text{card}(J) = p - q \) we know there exist at least one \( \bar{S} \subset J \) with \( \text{card}(\bar{S}) = p - 2q \) such that \( a^{\bar{S}}(k) = 0 \) for all \( k \geq 0 \), and \( |e_{\bar{S}}| \leq \epsilon + \gamma_{\bar{S}} \| \bar{m}^{\bar{S}} \|_k \) for all \( k \in [\bar{k}^* + (i - 1)N, \bar{k}^* + iN - 1] \). By construction of (4.48), it is satisfied that

\[
\pi_J(k) = \max_{S \subset J, \text{card}(S) = p - 2q} |\hat{x}_J(k) - \hat{x}_{\bar{S}}(k)| \\
\geq |e_J(k) - e_{\bar{S}}(k)|.
\]

for all \( k \geq \bar{k}^* \). If (4.54) holds at least \( N/2 \) time-steps in the \( i \)-th time window, then from triangle inequality, for all \( J \subset \{1, \ldots, p\} \) with \( \text{card}(J) = p - q \) and \( a^l(k) \neq 0 \) for some \( k \geq 0 \), we have

\[
\pi_J(k) \geq ||e_J(k)| - |e_{\bar{S}}(k)|| \\
> 3(\epsilon + \gamma_J^* \bar{m}) - \epsilon - \gamma_{\bar{S}} \| \bar{m}^{\bar{S}} \|_k \\
> 2(\epsilon + \gamma_J^* \bar{m}), \tag{4.55}
\]

for at least \( N/2 \) time-steps in the \( i \)-th time window, which implies all of attacked sensors can be isolated by Algorithm 3 in the \( i \)-th time window.

**Remark:** The performance of Algorithm 3 can be arbitrarily improved by increasing the length of the time window \( N \) at the price of increasing the time needed for isolation.

**Example 4** Consider the discrete-time nonlinear system subject to measurement noise and sensor attacks:

\[
x^+ = \begin{bmatrix} 1 & \delta \\ 0 & 1 \end{bmatrix} x + \begin{bmatrix} \frac{1}{2} \alpha \sin(x_1 + x_2) \\ \delta \alpha \sin(x_1 + x_2) \end{bmatrix}, \tag{4.56}
\]
4.5 Conclusion

We have provided a design method for discrete-time circle-criterion observers robust to measurement noise in terms of semidefinite programs. We have developed 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 er-

\[
\hat{y} = \begin{bmatrix} 3 & 0.3 \\ 3 & 0.6 \\ 6 & 0.9 \\ 1.2 & 12 \end{bmatrix} x + a + \tilde{m}. \quad (4.57)
\]

with \( \delta = 0.1, \alpha = 1, \) and \( \tilde{m}_i \sim \mathcal{U}(-0.5, 0.5) \) for \( i \in \{1, 2, 3, 4\} \). We let \( W = \{3\} \) and \( a_3 \sim \mathcal{U}(-c, c) \) with \( c \) given by 2 and 5. We run Algorithm 3 with \( \binom{4}{3} + \binom{4}{2} = 10 \) observers. We say sensor 0 is under attack in the \( i \)-th window when \( \bar{A}_i = \emptyset \). The isolation results are shown below in Figures 4.7-4.8.

Figure 4.7: Attack isolation, \( a_3 \sim \mathcal{U}(-2, 2) \).
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. In the case when the bound on measurement noise is known, two algorithms for detecting and isolating sensor attacks are proposed using multi-observer approach. Simulations results are provided to illustrate the performance of tools. The performance of our algorithms can be improved by increasing the length of the time window $N$ at the price of increasing the time needed for detection and isolation.

Figure 4.8: Attack isolation, $a_3 \sim U(-5, 5)$. 
5.1 Overview

In this chapter, we address the problem of state estimation, attack isolation, and control of discrete-time linear time-invariant systems under (potentially unbounded) actuator and sensor false data injection attacks. Using banks of unknown input observers, each observer leading to an exponentially stable estimation error (in the attack-free case), we develop an observer-based estimator using the same estimation strategy as we have described in Chapter 3, that provides exponential estimates of the system state in spite of actuator and sensor attacks.

Once we have an estimate of the state, we provide tools for reconstructing attack signals using model matching techniques. Attacked actuators and sensors are isolated by simply checking the sparsity of the estimated attack signals. Finally, after obtaining state estimates and isolation has been performed, we provide a control scheme for stabilizing the closed-loop dynamics. In the case with sensor attacks only (no actuators attacks), we show that the system can be stabilized by closing the loop with the multi-observer estimator and a static output feedback controller whose design is independent of the multi-observer estimator. When both sensors and actuator are attacks, we propose an effective technique to stabilize the system by switching off the isolated actuators, and closing the loop with a multi-observer based output time-varying feedback controller.
Because attack signals might be zero for some time instants, actuators isolated as attack-free might arbitrarily switch among all the supersets of the set of attack-free actuators. Therefore, we need a controller able to stabilize the closed-loop dynamics under the arbitrary switching induced by turning off the isolated actuators. To achieve this, we assume that a state feedback controller that stabilizes the switching closed-loop system exists, and use this controller together with the multi-observer estimator to stabilize the system. We use Input-to-State Stability (ISS) [59] of the closed-loop system with respect to the exponentially stable estimation error to conclude on stability of the closed-loop dynamics. Compared to the adaptive controller proposed in [71], where a particular class of attacks is considered and ultimate boundedness of the closed-loop system is guaranteed only, our controller is able drive the system state asymptotically to the origin under arbitrary and potentially unbounded attack signals. Simulation results are presented to illustrate the performance of our tool.

5.2 Estimation

In [15], the problem of state estimation for continuous-time LTI system under sensor attacks is solved using a bank of Luenberger observers. Inspired by these results, we use banks of UIOs to estimate the state of the system when sensor and actuator attacks both occur. Consider a discrete-time linear system under sensor and actuator attacks:

\[
\begin{align*}
\dot{x}^+ &= Ax + B(u + a_u) \\
y &= C x + a_y
\end{align*}
\] (5.1)

with state \(x \in \mathbb{R}^n\), output \(y \in \mathbb{R}^m\), known input \(u \in \mathbb{R}^n\), vector of actuator attacks \(a_u \in \mathbb{R}^{n_u}, a_u = (a_{u1}, \ldots, a_{un_u})^\top\), i.e., \(a_{ui}(k) = 0\) for all \(k \geq 0\) if the \(i\)-th actuator is attack-free; otherwise, \(a_{ui}(k_i) \neq 0\) for some \(k_i \geq 0\) and can be arbitrarily large, and vector of sensor attacks \(a_y \in \mathbb{R}^{n_y}, a_y = (a_{y1}, \ldots, a_{yn_y})^\top\), i.e., \(a_{yi}(k) = 0\) for all \(k \geq 0\) if the \(i\)-th sensor is attack-free; otherwise, \(a_{yi}(k_i) \neq 0\) for some \(k_i \geq 0\) and can be arbitrarily large. Matrices \(A, B, C\) are of appropriate dimensions, and we assume that \((A, B)\) is stabilizable, \((A, C)\) is detectable, and \(B\) has full column rank. Let \(W_u \subset \{1, \ldots, n_u\}\) denote the unknown set
of attacked actuators, and \( W_y \subset \{1, \ldots, n_y\} \) denote the unknown set of attacked sensors.

**Assumption 5.1.** The sets of attacked actuators and sensors do not change over time, i.e., \( W_u \subseteq \{1, \ldots, n_u\} \), \( W_y \subset \{1, \ldots, n_y\} \) are constant (time-invariant) and \( \text{supp}(a_u(k)) \subseteq W_u \), \( \text{supp}(a_y(k)) \subseteq W_y \), for all \( k \geq 0 \).

### 5.2.1 Complete Unknown Input Observers

We first treat \((u + a_u)\) as an unknown input to system (5.1) and consider a UIO with the following structure:

\[
\begin{align*}
\dot{z}_{J_s} &= N_{J_s} z_{J_s} + L_{J_s} y_{J_s}, \\
\dot{x}_{J_s} &= z_{J_s} + E_{J_s} y_{J_s},
\end{align*}
\tag{5.2}
\]

where \( z_{J_s} \in \mathbb{R}^{n} \) is the state of the observer, \( \hat{x}_{J_s} \in \mathbb{R}^{n} \) denotes the estimate of the system state, \((N_{J_s}, L_{J_s}, E_{J_s})\) are observer matrices of appropriate dimensions to be designed. It is easy to verify that if \((N_{J_s}, L_{J_s}, E_{J_s})\) satisfy the following equations:

\[
\begin{align*}
N_{J_s} (I - E_{J_s} C_{J_s}) + L_{J_s} C_{J_s} + (E_{J_s} C_{J_s} - I) A &= 0, \\
(E_{J_s} C_{J_s} - I) B &= 0;
\end{align*}
\tag{5.3}
\]

then, the estimation error \( e_{J_s} = \hat{x}_{J_s} - x \) satisfies:

\[
e_{J_s} = N_{J_s} e_{J_s}.
\tag{5.4}
\]

If \( N_{J_s} \) is Schur, system (5.2) is called a UIO for (5.1). In [18], it is proved that such observer exists if and only if the following two conditions are satisfied:

\((c_1)\) \( \text{rank}(C_J^T B) = \text{rank}(B) = n_u \).

\((c_2)\) The pair \((C_J, A - E_J C_J)\) is detectable.

Let \( q \) be the largest integer such that for all \( J_s \subset \{1, \ldots, n_y\} \) with \( \text{card}(J_s) \geq n_y - 2q > 0 \), conditions \((c_1)\) and \((c_2)\) are satisfied; then, observer (5.2) can be constructed for any \( C_J \) with \( \text{card}(J_s) \geq n_y - 2q \) by solving (5.3) for a Schur matrix \( N_{J_s} \). Hence, for such an
observer, if \( a_y^I(k) = 0 \) for all \( k \geq 0 \), there exist \( c_I > 0 \), \( \lambda_I \in (0, 1) \) satisfying:

\[
|e_I(k)| \leq c_I \lambda_I^k |e_I(0)|,
\]

for all \( e_I(0) \in \mathbb{R}^n, k \geq 0 \) [18], where \( e_I = \hat{x}_I - x \).

**Assumption 5.2.** There are at most \( q \) sensors attacked by an adversary, i.e.,

\[
\text{card}(W_y) \leq q < \frac{n_y}{2},
\]

where \( q \) is the largest positive integer satisfying conditions (c1) and (c2).

**Lemma 5.1.** Under Assumption 5.2, among each set of \( n_y - q \) sensors, at least \( n_y - 2q > 0 \) of them are attack-free.

**Proof:** Lemma 5.1 follows trivially from Assumption 5.2.

Let Assumption 5.2 be satisfied. Inspired by the ideas in [15], we use a UIO for each subset \( J_s \subset \{1, \ldots, n_y\} \) of sensors with \( \text{card}(J_s) = n_y - q \) and for each subset \( S_s \subset \{1, \ldots, n_y\} \) of sensors with \( \text{card}(S_s) = n_y - 2q \). Under Assumption 5.2, there exists at least one set \( J_s \subset \{1, \ldots, n_y\} \) with \( \text{card}(J_s) = n_y - q \) such that \( a_y^I(k) = 0 \) for all \( k \geq 0 \). Then, the estimate given by the UIO for \( J_s \) is a correct estimate, and the estimate given by the UIO for any \( S_s \subset J_s \) with \( \text{card}(S_s) = n_y - 2q \) is consistent with that given by \( J_s \). This motivates the following estimation strategy.

For each set \( J_s \) with \( \text{card}(J_s) = n_y - q \), we define \( \pi_{J_s}(k) \) as the largest deviation between \( \hat{x}_{J_s}(k) \) and \( \hat{x}_{S_s}(k) \) that is given by any \( S_s \subset J_s \) with \( \text{card}(S_s) = n_y - 2q \), i.e.,

\[
\pi_{J_s}(k) := \max_{S_s \subset J_s: \text{card}(S_s) = n_y - 2q} |\hat{x}_{J_s}(k) - \hat{x}_{S_s}(k)|,
\]

for all \( k \geq 0 \), and the sequence \( \sigma_s(k) \) as

\[
\sigma_s(k) := \arg \min_{J_s \subset \{1, \ldots, n_y\}: \text{card}(J_s) = n_y - q} \pi_{J_s}(k).
\]
Then, as proved below, the estimate indexed by $\sigma_s(k)$:

$$\hat{x}(k) := \hat{x}_{\sigma_s(k)}(k),$$

is an exponential attack-free estimate of the system state. For simplicity and without generality, for all $J_s$ and $S_s$, $z_{J_s}(0)$ and $z_{S_s}(0)$ are chosen such that $\hat{x}_{J_s}(0) = \hat{x}_{S_s}(0) = \hat{x}(0)$. The following result summarizes the ideas presented above.

**Theorem 5.1.** Consider system (5.1), observer (5.2), and the complete multi-observer estimator (5.7)-(5.9). Define the estimation error $e(k) := \hat{x}_{\sigma_s(k)}(k) - x(k)$, and let conditions (c1)-(c2) and Assumptions 5.1-5.2 be satisfied; then, there exist constants $\bar{c} > 0$, $\bar{\lambda} \in (0, 1)$ satisfying:

$$|e(k)| \leq \bar{c}\bar{\lambda}^k|e(0)|,$$

(5.10)

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

**Proof:** Under Assumption 5.2, there exists at least one set $\bar{J}_s$ with $\text{card}(\bar{J}_s) = n_y - q$ such that $a_{y_{\bar{J}_s}}(k) = 0$ for all $k \geq 0$. Then, there exist $c_{\bar{J}_s} > 0$ and $\lambda_{\bar{J}_s} \in (0, 1)$ such that

$$|e_{\bar{J}_s}(k)| \leq c_{\bar{J}_s}\lambda_{\bar{J}_s}^k|e(0)|,$$

(5.11)

for all $e(0) \in \mathbb{R}^n$ and $k \geq 0$. Moreover, for any set $S_s \subset \bar{J}_s$ with $\text{card}(S_s) = n_y - 2q$, we have $a_{y_{S_s}}(k) = 0 \forall k \geq 0$; hence, there exist $c_{S_s} > 0$ and $\lambda_{S_s} \in (0, 1)$ such that

$$|e_{S_s}(k)| \leq c_{S_s}\lambda_{S_s}^k|e(0)|,$$

(5.12)

for all $e(0) \in \mathbb{R}^n$ and $k \geq 0$. Consider $\pi_{\bar{J}_s}$ in (5.7). Combining the above inequalities, we
have

\[
\pi_{J_k}(k) = \max_{S_s \subset J_k} |\hat{x}_{J_k}(k) - \hat{x}_{S_s}(k)| \\
= \max_{S_s \subset J_k} |\hat{x}_{J_k}(k) - x(k) + x(k) - \hat{x}_{S_s}(k)| \\
\leq |e_{J_k}(k)| + \max_{S_s \subset J_k} |e_{S_s}(k)| \\
\leq 2c'_{J_k} \lambda'_{J_k} |e(0)|,
\]

for all \(e(0) \in \mathbb{R}^n\) and \(k \geq 0\), where

\[
c'_{J_k} := \max_{S_s \subset J_k} \{c_{J_s}, c_{S_s}\},
\]

\[
\lambda'_{J_k} := \max_{S_s \subset J_k} \{\lambda_{J_s}, \lambda_{S_s}\}.
\]

Note that \(S_s \subset \bar{J}_s\), \(\text{card}(S_s) = n_y - 2q\). Then, from (5.8), we have \(\pi_{\sigma_s(k)}(k) \leq \pi_{J_k}(k)\). From Lemma 5.1, we know that there exist at least one set \(\bar{S}_s \subset \sigma_s(k)\) with \(\text{card}(\bar{S}_s) = n_y - 2q\), such that \(a_{\bar{S}_s}(k) = 0\) for all \(k \geq 0\), and there exist \(c_{\bar{S}_s} > 0\) and \(\lambda_{\bar{S}_s} \in (0, 1)\) such that

\[
|e_{\bar{S}_s}(k)| \leq c_{\bar{S}_s} \lambda_{\bar{S}_s}^k |e(0)|,
\]

for all \(e(0) \in \mathbb{R}^n\) and \(k \geq 0\). From (5.7), we have

\[
\pi_{\sigma_s(k)}(k) = \max_{S_s \subset \sigma_s(k)} |\hat{x}_{\sigma_s(k)}(k) - \hat{x}_{S_s}(k)| \\
\geq |\hat{x}_{\sigma_s(k)}(k) - \hat{x}_{\bar{S}_s}(k)|.
\]

Using this lower bound on \(\pi_{\sigma_s(k)}(k)\) and the triangle inequality we have that

\[
|e_{\sigma_s(k)}(k)| = |\hat{x}_{\sigma_s(k)}(k) - x(k)| \\
= |\hat{x}_{\sigma_s(k)}(k) - \hat{x}_{S_s}(k) + \hat{x}_{S_s}(k) - x(k)| \\
\leq |\hat{x}_{\sigma_s(k)}(k) - \hat{x}_{S_s}(k)| + |e_{S_s}(k)| \\
\leq \pi_{\sigma_s(k)}(k) + |e_{S_s}(k)| \\
\leq \pi_{J_k}(k) + |e_{S_s}(k)|,
\]
for all $k \geq 0$. Hence, from (5.13) and (5.14), we have

$$|e_{e_t}(k)| \leq \bar{c} \bar{\lambda}^k |e(0)|,$$  \hspace{1cm} (5.16)

for all $e(0) \in \mathbb{R}^n$ and $k \geq 0$, where $\bar{c} = 3 \max \{ c_{\bar{S}} , c_{\bar{J}}' \}$ and $\bar{\lambda} = \max \{ \lambda_{\bar{S}}, \lambda_{\bar{J}}' \}$. Inequality (5.16) is of the form (5.10) and the result follows.

\[\square\]

Remark: The optimization problem in (5.8) amounts to searching among a number of \( \binom{n_y}{n_y-q} \) sets $J_s$ to find the one with the smallest value of $\pi_{J_s}(k)$. Therefore, the complexity of (5.8) is $O\left( \binom{n_y}{n_y-q} \right)$.

**Example 1:** Consider the following system subject to actuator and sensor attacks:

\begin{align*}
\begin{cases}
  x^+ &= \begin{bmatrix} 0.2 & 0.5 \\ 0.2 & 0.7 \end{bmatrix} x + \begin{bmatrix} 1 \\ 2 \end{bmatrix} (u + a_u), \\
  y &= \begin{bmatrix} 1 & 3 \\ 1 & 1 \\ 3 & 2 \end{bmatrix} x + a_y.
\end{cases}
\end{align*}

It can be verified that a UIO of the form (5.2) exists for each $C^h$ with $J_s \subset \{1,2,3,4\}$ and $\text{card}(J_s) \geq 2$; then, $4 - 2q = 2$, i.e., $q = 1$ and at most one sensor is attacked. We attack the actuator and let $W_y = \{3\}$, i.e., the third sensor is attacked. We let $u \sim \mathcal{U}(-1,1)$, $a_u, a_{y3} \sim \mathcal{U}(-10,10)$, and $(x_1(0), x_2(0)) \sim \mathcal{N}(0,1^2)$. We design a UIO for each $J_s$ with $\text{card}(J_s) = 3$, and for each $S_s$ with $\text{card}(S_s) = 2$. Therefore, totally $\binom{4}{3} + \binom{4}{2} = 10$ UIOs are designed and they are all initialized at $\hat{x}(0) = [0,0]^\top$. For $k \in [0,19]$, the estimator (5.2), (5.7)-(5.9) is used to construct $\hat{x}(k)$. The performance of the estimator is shown in Figure 5.1.
\section{Partial Unknown Input Observers}

Here, we are implicitly assuming that either condition (c\textsubscript{1}) or (c\textsubscript{2}) (or both) cannot be satisfied for any \( C^h \) with \( \text{card}(J_u) = n_y - 2q \) with \( q \geq 1 \). Let \( B \) be partitioned as \( B = [b_1, \ldots, b_{i_r}, \ldots, b_{n_u}] \) where \( b_i \in \mathbb{R}^{n \times 1} \) is the \( i \)-th column of \( B \). Then, the attacked system (5.1) can be written as

\[
\begin{cases}
x^+ = Ax + Bu + b_{W_u}a_{W_u}, \\
y = Cx + a_y,
\end{cases}
\tag{5.18}
\]

where the attack input \( a_{W_u} \) can be regarded as an unknown input and the columns of \( b_{W_u} \) are \( b_i, i \in W_u \). Denote by \( b_{J_u} \) the matrix whose columns are \( b_i \) for \( i \in J_u \). Let \( q_1 \) and \( q_2 \) be the largest integers such that for all \( J_u \subset \{1, \ldots, n_u\} \) with \( \text{card}(J_u) \leq 2q_1 < n_u \) and \( J_s \subset \{1, \ldots, n_y\} \) with \( \text{card}(J_s) \geq n_y - 2q_2 > 0 \), the following is satisfied:

\[ (c_3) \quad \text{rank}(C^h b_{J_u}) = \text{rank}(b_{J_u}) = \text{card}(J_u). \]
(c₄) There exists \((N_{Ju}, L_{Ju}, E_{Ju}, T_{Ju})\) satisfying:

\[
\begin{cases}
N_{Ju}(I - E_{Ju}C^h) + L_{Ju}C^h + (E_{Ju}C^h - I)A = 0, \\
(T_{Ju} + E_{Ju}C^h - I)B = 0, \\
(E_{Ju}C^h - I)b_{Ju} = 0,
\end{cases}
\] (5.19)

with detectable pair \((C^h, A - E_{Ju}C^hA)\) and Schur \(N_{Ju}\). If conditions \((c_3)\) and \((c_4)\) are satisfied, a UIO with the following structure exists for each \(b_{Ju}\) with \(J_u \subset \{1, \ldots, n_u\}\), card \((J_u) \leq 2q_1 < n_u\) and each \(C^h\) with \(J_s \subset \{1, \ldots, n_y\}\), card \((J_s) \geq n_y - 2q_2 > 0\):

\[
\begin{cases}
z_{Ju}^+ = N_{Ju}z_{Ju} + T_{Ju}Bu + L_{Ju}y^h, \\
\hat{x}_{Ju} = z_{Ju} + E_{Ju}y^h,
\end{cases}
\] (5.20)

where \(z_{Ju} \in \mathbb{R}^n\) is the observer state, \(\hat{x}_{Ju}\) denotes the state estimate, and \((N_{Ju}, L_{Ju}, T_{Ju}, E_{Ju})\) are the observer matrices satisfying (5.19), see [18] for further details. That is, system (5.20) is a UIO for the system:

\[
\begin{cases}
x^+ = Ax + Bu + b_{Ju}a_{Ju}^h, \\
y^h = C^hx + a_{Ju}^h,
\end{cases}
\] (5.21)

with unknown input \(b_{Ju}a_{Ju}^h\) and known input \(Bu\). It follows that the estimation error \(e_{Ju} = \hat{x}_{Ju} - x\) satisfies:

\[
e_{Ju}^+ = N_{Ju}e_{Ju},
\] (5.22)

for some Schur matrix \(N_{Ju}\). We refer to UIOs of the form (5.21) as partial UIOs for the pair \((J_u, J_s)\).

**Assumption 5.3.** There are at most \(q_1\) actuators and at most \(q_2\) sensors attacked by an adversary, i.e.,

\[
\text{card}(W_u) \leq q_1 < \frac{n_u}{2},
\] (5.23)

\[
\text{card}(W_y) \leq q_2 < \frac{n_y}{2},
\] (5.24)

where \(q_1\) and \(q_2\) are the largest positive integers satisfying \((c_3)\) and \((c_4)\).
Remark: Note that if conditions \((c_3)\) and \((c_4)\) are satisfied for \(b_{J_u}\) with \(\text{card}(J_u) = 2q_1 = n_u\), then conditions \((c_1)\) and \((c_2)\) are satisfied, and (5.20) is a complete UIO for (5.1) for \(T_{J_u} = 0\). Since we are considering partial UIOs, we assume \(2q_1 < n_u\) to exclude this case.

Lemma 5.2. Under Assumption 5.3, for each set of \(q_1\) actuators, among all its supersets with \(2q_1\) actuators, at least one set is a superset of \(W_u\).

Lemma 5.3. Under Assumption 5.3, among each set of \(n_y - q_2\) sensors, at least \(n_y - 2q_2 > 0\) sensors are attack-free.

Proof: Lemmas 5.2 and 5.3 follow trivially from Assumption 5.3.

Note that the existence of a UIO for each pair \((J_u, J_s)\) with \(\text{card}(J_u) \leq 2q_1\) and \(\text{card}(J_s) \geq n_y - 2q_2\) means that if \(W_u \subseteq J_u\) and \(a_y(k) = 0\) for all \(k \geq 0\), the estimation error \(e_{J_u} = \hat{x}_{J_u} - x\) satisfies

\[
|e_{J_u}(k)| \leq c_{J_u} \lambda_{J_u}^k |e_{J_u}(0)|, \tag{5.25}
\]

for some \(c_{J_u} > 0\) and \(\lambda_{J_u} \in (0,1)\), all \(e_{J_u}(0) \in \mathbb{R}^n\), and \(k \geq 0\). Let Assumption 5.3 be satisfied. We use a UIO for each pair \((J_u, J_s)\) with \(\text{card}(J_u) = q_1\) and \(\text{card}(J_s) = n_y - q_2\). Then, we use a UIO for each pair \((S_u, S_s)\) with \(S_u \subseteq \{1, \ldots, n_u\}\), \(\text{card}(S_u) = 2q_1\) and \(S_s \subseteq \{1, \ldots, n_y\}\), \(\text{card}(S_s) = n_y - 2q_2\). Under Assumption 5.3, there exists at least one set \(\bar{J}_u\) with \(\text{card}(\bar{J}_u) = q_1\) such that \(W_u \subseteq \bar{J}_u\) and at least one set \(\bar{J}_s\) with \(\text{card}(\bar{J}_s) = n_y - q_2\) such that \(a_y(k) = 0\) for all \(k \geq 0\). Then, the estimate given by the UIO for \((\bar{J}_u, \bar{J}_s)\) is a correct estimate, and the estimates given by the UIOs for any \((S_u, S_s)\) (denoted as \(\hat{x}_{S_u}\)), where \(S_u \supseteq \bar{J}_u\), \(\text{card}(S_u) = 2q_1\) and \(S_s \subseteq \bar{J}_s\), \(\text{card}(\bar{J}_s) = n_y - 2q_2\), are consistent with \(\hat{x}_{J_u}\).

This motivates the following estimation strategy.

For each pair \((J_u, J_s)\) with \(\text{card}(J_u) = q_1\) and \(\text{card}(J_s) = n_y - q_2\), define \(\pi_{J_u}(k)\) as the largest deviation between \(\hat{x}_{J_u}(k)\) and \(\hat{x}_{S_u}(k)\) that is given by any pair \((S_u, S_s)\), where \(S_u \supseteq J_u\) with \(\text{card}(S_u) = 2q_1\) and \(S_s \subseteq J_s\) with \(\text{card}(S_s) = n_y - 2q_2\). That is,

\[
\pi_{J_u}(k) := \max_{S_u \supseteq J_u, S_s \subseteq J_s} |\hat{x}_{J_u}(k) - \hat{x}_{S_u}(k)|, \tag{5.26}
\]
for all $k \geq 0$. Define the sequences $\sigma_u(k)$ and $\sigma_s(k)$ as

$$
(\sigma_u(k), \sigma_s(k)) := \arg\min_{J_u, J_s} \mathcal{P}_{fu}(k).
$$

(5.27)

Then, as proven below, the estimate indexed by $(\sigma_u(k), \sigma_s(k))$:

$$
\hat{x}(k) = \hat{x}_{\sigma_u(k)}(k),
$$

(5.28)

is an exponential attack-free estimate of the system state. For simplicity and without
generality, for all $J$ and $S$, $z_{J_u}(0)$ and $z_{S_u}(0)$ are chosen such that

$$
\hat{x}_{J_u}(0) = \hat{x}_{S_u}(0) = \hat{x}(0).
$$

The following result summarizes the ideas presented above.

**Theorem 5.2.** Consider system (5.1), observer (5.20), and the partial multi-observer estimator (5.26)-(5.28). Define the estimation error $e(k) := \hat{x}_{\sigma_u(k)}(k) - x(k)$ and let (c3)-(c4) and Assumptions 5.1, 5.3 be satisfied; then, there exist positive constants $\bar{c} > 0$ and $\bar{\lambda} \in (0, 1)$ satisfying:

$$
|e(k)| \leq \bar{c} \bar{\lambda}^k |e(0)|, \quad (5.29)
$$

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

**Proof:** Under Assumption 5.3, there exists at least one set $\bar{J}_u$ with card($\bar{J}_u$) = $q_1$ such that $\bar{J}_u \supset W_u$, and at least one set $\bar{J}_s$ with card($\bar{J}_s$) = $n_y - q_2$ such that $a_y^{\bar{J}_s}(k) = 0$ for all $k \geq 0$; then, there exist $c_{J_u} > 0$ and $\lambda_{J_u} \in (0, 1)$ satisfying

$$
|e_{J_u}(k)| \leq c_{J_u} \lambda_{J_u}^k |e(0)|, \quad (5.30)
$$

for all $e(0) \in \mathbb{R}^n$ and $k \geq 0$. Moreover, for any set $S_u \supset \bar{J}_u$ with card($S_u$) = $2q_1$ and $S_s \subset \bar{J}_s$ with card($S_s$) = $n_y - 2q_2$, we have $S_u \supset W_u$ and $a_y^{S_u}(k) = 0$ for all $k \geq 0$; hence, there exist $c_{S_u} > 0$ and $\lambda_{S_u} \in (0, 1)$ such that

$$
|e_{S_u}(k)| \leq c_{S_u} \lambda_{S_u}^k |e(0)|, \quad (5.31)
$$

for all $e(0) \in \mathbb{R}^n$ and $k \geq 0$. Consider $\pi_{J_u}$ in (5.26). Combining the above results, we
have that
\[ \pi_{\bar{J}_u}(k) = \max_{S_u \supset I_u, S_s \subset I_s} |\hat{x}_{\bar{J}_u}(k) - \hat{x}_{S_u}(k)| \]
\[ = \max_{S_u \supset I_u, S_s \subset I_s} |\hat{x}_{\bar{J}_u}(k) - x(k) + x(k) - \hat{x}_{S_u}(k)| \]
\[ \leq |e_{\bar{J}_u}(k)| + \max_{S_u \supset I_u, S_s \subset I_s} |e_{S_u}(k)|, \]
for all \( k \geq 0 \). From (5.30) and (5.31), we obtain
\[ \pi_{\bar{J}_u}(k) \leq 2c'_{\bar{J}_u} \lambda'_{\bar{J}_u} |e(0)|, \] (5.32)
for all \( e(0) \in \mathbb{R}^n \) and \( k \geq 0 \), where
\[ c'_{\bar{J}_u} := \max_{S_u \supset I_u, S_s \subset I_s} \{c_{\bar{J}_u}, c_{S_u}\}, \]
\[ \lambda'_{\bar{J}_u} := \max_{S_u \supset I_u, S_s \subset I_s} \{\lambda_{\bar{J}_u}, \lambda_{S_u}\}. \]
Note that \( S_u \supset \bar{J}_u \), \( \text{card}(S_u) = 2q_1 \), and \( S_s \subset \bar{J}_s \), \( \text{card}(S_s) = n_y - 2q_2 \). Then, from (5.27), we have \( \pi_{\bar{J}_u}(k) \leq \pi_{\bar{J}_u}(k) \). By Lemmas 5.2 and 5.3, we know that there exists at least one set \( \bar{S}_u \supset \sigma_u(k) \) with \( \text{card}(\bar{S}_u) = 2q_1 \) and at least one set \( \bar{S}_s \subset \sigma_s(k) \) with \( \text{card}(\bar{S}_s) = n_y - 2q_2 \) such that \( \bar{S}_u \supset W_{\bar{J}_u} \) and \( u_{\bar{J}_u}(k) = 0 \) for all \( k \geq 0 \). Hence, there exist \( c_{\bar{S}_u} > 0 \) and \( \lambda_{\bar{S}_u} \in (0, 1) \) satisfying
\[ |e_{\bar{S}_u}(k)| \leq c_{\bar{S}_u} \lambda_{\bar{S}_u}^k |e(0)|, \] (5.33)
for all \( e(0) \in \mathbb{R}^n \) and \( k \geq 0 \). From (5.26), by construction
\[ \pi_{\bar{J}_u}(k) = \max_{S_u \supset \sigma_u(k), S_s \subset \sigma_s(k)} |\hat{x}_{\bar{J}_u}(k) - \hat{x}_{S_u}(k)| \]
\[ \geq |\hat{x}_{\bar{S}_u}(k) - \hat{x}_{S_u}(k)|. \]
5.2 Estimation

Using the above lower bound on $\pi_{\nu}(k)$ and the triangle inequality, we have that

$$|e_{\nu}(k)| = |\hat{x}_{\nu}(k) - x(k)|$$

$$= |\hat{x}_{\nu}(k) - \hat{x}_{\nu}(k) + \hat{x}_{\nu}(k) - x(k)|$$

$$\leq |\hat{x}_{\nu}(k) - \hat{x}_{\nu}(k)| + |\nu_{\nu}(k)|$$

$$\leq \pi_{\nu}(k) + |\nu_{\nu}(k)|$$

$$\leq \pi_{\tilde{f}_{us}}(k) + |\nu_{\nu}(k)|,$$

for all $k \geq 0$. Hence, from (5.32) and (5.33), we have

$$|e_{\nu}(k)| \leq \tilde{e}\tilde{\lambda}|e(0)|,$$  (5.35)

for all $e(0) \in \mathbb{R}^n$ and $k \geq 0$, where $\tilde{e} = 3 \max\{\epsilon_{\nu}, \epsilon'_{\nu}\}$, $\tilde{\lambda} = \max\{\lambda_{\nu}, \lambda'_{\nu}\}$. Inequality (5.35) is of the form (5.29), and the result follows. ■

Remark: The computation complexity of (5.27) is $O\left(\left(\begin{array}{l} n_y \\ n_y - q_2 \end{array}\right) \times \left(\begin{array}{l} n_u \\ q_1 \end{array}\right)\right)$.

Example 2: Consider a linear system subject to actuator and sensor attacks:

$$\begin{align*}
\begin{bmatrix}
0.5 & 0 & 0.1 \\
0.2 & 0.7 & 0 \\
1 & 0 & 0.3
\end{bmatrix}
\begin{bmatrix}
x^+ \\
x + \begin{bmatrix}
0.1 & 1 & 0.1 \\
0 & 0 & 0.5
\end{bmatrix}(u + a_u),
\end{bmatrix}
\begin{bmatrix}
1 & 2 & 0 \\
0 & 1 & 1 \\
0 & 1 & 2 \\
1 & 1 & 1
\end{bmatrix}
\begin{bmatrix}
x + a_y.
\end{bmatrix}
\end{align*}$$

(5.36)

It can be verified that complete UIOs do not exist for any $C^l$ with $\text{card}(J_s) \leq 2$. However, a partial UIO exists for each pair $(J_u, J_s)$ with $\text{card}(J_u) \leq 2$ and $\text{card}(J_s) \geq 2$; then, $2q_1 = 2$ and $4 - 2q_2 = 2$, i.e., $q_1 = q_2 = 1$. We let $W_u = \{3\}$, $W_y = \{2\}$, i.e., the third actuator and the second sensor are attacked, $(u_1, u_2, u_3) \sim \mathcal{U}(-1, 1)$, and $a_{13}, a_{22} \sim \mathcal{U}(-10, 10)$, and $(x_1(0), x_2(0), x_3(0)) \sim \mathcal{N}(0, 1^2)$. We construct a partial UIO for each pair $(J_u, J_s)$
Figure 5.2: Estimated states $\hat{x}$ converges to the true states $x$ when $a_{u3}, a_{y2} \sim U(-10, 10)$. Legend: $\hat{x}$ (blue), true states (black)

with $\text{card}(J_u) = 1, \text{card}(J_s) = 3$ and each set $(S_u, S_s)$ with $\text{card}(S_u) = 2, \text{card}(S_s) = 2$. Therefore, totally $\binom{3}{1} \times \binom{4}{3} + \binom{3}{2} \times \binom{4}{2} = 30$ partial UIOs are designed. We initialize the observers at $\hat{x}(0) = [0, 0, 0]^\top$. Estimator (5.20), (5.26)-(5.28) is used to construct $\hat{x}(k)$. The performance of the estimator is shown in Figure 5.2.

### 5.3 Attack Isolation and Reconstruction

Once we have an estimate $\hat{x}(k)$ of $x(k)$, either using the complete multi-observer estimator in Section 5.2.1 or the partial multi-observer estimator in Section 5.2.2, we can use these estimates, the system model (5.1), and the known inputs to exponentially reconstruct the attack signals. Note that $e = \hat{x} - x \Rightarrow x = \hat{x} - e \Rightarrow x^+ = \hat{x}^+ - e^+$. Then, the system dynamics (5.1) can be written in terms of $e$ and $\hat{x}$ as follows:

\[
\begin{cases}
\dot{\hat{x}}^+ = e^+ + A(\hat{x} - e) + B(u + a_u), \\
\downarrow \\
a_u = B^{-1}_{\text{lef}}(\hat{x}^+ - A\hat{x}) - u - B^{-1}_{\text{lef}}(e^+ - Ae),
\end{cases}
\]

(5.37)
because $B$ has full column rank (as introduced in the system description), where $B_{\text{left}}^{-1}$ denotes the Moore-Penrose pseudoinverse of $B$. Similarly, we have

$$
\begin{align*}
\begin{cases}
y = Cx + ay = C\hat{x} - Ce + ay,
\Downarrow
\end{cases}
\end{align*}
$$

First, consider the complete multi-observer in Section 5.2.1. Let the estimation error dynamics characterized by (5.7)-(5.9) be given by

$$
e^+ = f_1(e, x, ay, au),
$$

where $f_1 : \mathbb{R}^n \times \mathbb{R}^n \times \mathbb{R}^{ny} \times \mathbb{R}^{nu} \to \mathbb{R}^n$ denotes some nonlinear function. That is, the estimation error is given by some nonlinear function of the state and the attack signals. However, in Theorem 5.1, we have proved that $e$ converges to the origin exponentially. Hence, the terms depending on $e$ and $e^+$ in the expression for $au$ and $ay$ in (5.37) and (5.38) vanishes exponentially and therefore, the following attack estimate:

$$
\hat{a}_u(k) = B^{-1}_{\text{left}}(\hat{x}(k) - A\hat{x}(k - 1)) - u(k - 1),
$$

and

$$
\hat{a}_y(k) = y(k) - C\hat{x}(k),
$$

exponentially reconstruct the attack signals $a_u(k - 1)$ and $a_y(k)$, i.e.,

$$
\lim_{k \to \infty} (\hat{a}_u(k) - a_u(k - 1)) = 0,
$$

and

$$
\lim_{k \to \infty} (\hat{a}_y(k) - a_y(k)) = 0.
$$

Then, for sufficiently large $k$, the sparsity pattern of $\hat{a}_u(k)$ and $\hat{a}_y(k)$ can be used to isolate attacks, i.e.,

$$
\hat{W}_u(k) = \text{supp}(\hat{a}_u(k)),
$$
Figure 5.3: Estimated actuator attacks $\hat{a}_u^+$ converges to $a_u$ when $a_u, a_y \sim \mathcal{U}(-10, 10)$. Legend: $\hat{a}_u^+$ (blue), $a_u$ (black).

and

$$\hat{W}_y(k) = \text{supp}(\hat{a}_y(k)),$$

(5.45)

where $\hat{W}_u(k)$ denotes the set of isolated attacked actuators, and $\hat{W}_y(k)$ denotes the set of isolated attacked sensors. Note that we can only estimate $a_u$ from $\hat{x}^+$ and $e^+$, which implies that we always have, at least, one-step delay for actuator attacks isolation.

Next, consider the partial multi-observer estimator given in Section 5.2.2. In this case, the attack vector $a_u$ and $a_y$ can also be written as (5.37) and (5.38), and the estimation error dynamics is given by some nonlinear difference equation characterized by the estimator structure in (5.26)-(5.28). Let the estimation error dynamics be given by

$$e^+ = f_2(e, x, a_y, a_u),$$

(5.46)

for some nonlinear function $f_2 : \mathbb{R}^n \times \mathbb{R}^n \times \mathbb{R}^n_y \times \mathbb{R}^n_u \rightarrow \mathbb{R}^n$. In Theorem 5.2, we have proved that $e$ converges to the origin exponentially. Hence, the attack estimate in (5.40) and (5.41) exponentially reconstructs the attack signals. Again, the sparsity pattern of $\hat{a}_u(k)$ and $\hat{a}_y(k)$ can be used to isolate actuator and sensor attacks using (5.44) and (5.45).

Example 3: Consider system (5.17) and the complete multi-observer estimator in Example 1. Let $W_u = \{1\}$, $W_y = \{3\}$, $u \sim \mathcal{U}(-1, 1)$, $a_u, a_y \sim \mathcal{U}(-10, 10)$, and $(x_1(0), x_2(0)) \sim \mathcal{N}(0, 1^2)$. We obtain $\hat{a}_u(k)$ and $\hat{a}_y(k)$ from (5.40) and (5.41). The reconstructed attack signals are depicted in Figures 5.3-5.4. By checking the sparsity of these signals, actuator and sensor 3 are isolated as attacked.

Example 4: Here we consider system (5.36) and the partial multi-observer estimator in
5.4 Control

In this section, we introduce a method to use the proposed multi-observer estimators to asymptotically stabilize the system dynamics.

5.4.1 Sensor attacks only

We first consider the case when only sensors are attacked and actuators are attack-free. Then, the system is given by

\[
\begin{align*}
    x^+ &= Ax + Bu, \\
    y &= Cx + a_y.
\end{align*}
\]  

(5.47)
Let \( u = K \hat{x} \), where \( \hat{x} \) is the estimate given by the complete multi-observer estimator in Section 5.2.1 or the partial multi-observer estimator in Section 5.2.2, and \( K \) is chosen such that \( A + BK \) is Schur. Then, the closed-loop system is given by

\[
x^+ = Ax + BK \hat{x},
\]

or in terms of the estimation error as

\[
x^+ = Ax + B(K(\hat{x} - x + x)),
\]

\[
= (A + BK)x + BKe.
\]

For the complete multi-observer estimator, let the estimation error dynamics be given by

\[
e^+ = f_1(e, x, a_y),
\]
for some nonlinear function $f_1 : \mathbb{R}^n \times \mathbb{R}^n \times \mathbb{R}^{n_y} \to \mathbb{R}^n$. For the partial multi-observer estimator, let the estimation error dynamics be given by

$$e^+ = f_2(e, x, a_y), \quad (5.51)$$

for some nonlinear function $f_2 : \mathbb{R}^n \times \mathbb{R}^n \times \mathbb{R}^{n_y} \to \mathbb{R}^n$. Since $A + BK$ is Schur, the closed-loop dynamics (5.49) is Input-to-State Stable (ISS) with respect to input $e(k)$ and some linear gain, see [33]. Moreover, in Theorems 5.1 and 5.2, we have proved that (5.50) and (5.51) are exponentially stable uniformly in $x(k)$ and $a_y(k)$. The latter and ISS of the system dynamics imply that $\lim_{k \to \infty} x(k) = 0$ [33].
Example 5: Consider the open-loop unstable system

\[
\begin{cases}
    x^+ = \begin{bmatrix} 1.2 & 0.5 \\ 0.2 & 0.7 \end{bmatrix} x + \begin{bmatrix} 1 \\ 0 \end{bmatrix} K\hat{x}, \\
y = \begin{bmatrix} 1 & 3 \\ 1 & 1 \\ 3 & 2 \end{bmatrix} x + a_y.
\end{cases}
\]

(5.52)

It can be verified that a UIO of the form (5.2) exists for each \( J_s \subset \{1, 2, 3, 4\} \) with \( \text{card}(J_s) \geq 2 \); then, \( 4 - 2q = 2 \) and \( q = 1 \). We let \( W_y = \{2\} \) and \( a_{y2} \sim \mathcal{U}(-10, 10) \). We construct \( \binom{4}{3} + \binom{4}{2} = 10 \) UIOs initialized at \( \hat{x}(0) = [0, 0]^T \) and let

\[
K = \begin{bmatrix} -1.2 & 0.7 \\ -0.2 & -0.7 \end{bmatrix}.
\]

We use the complete multi-observer in Section 5.2.1 to estimate the state. The state of the closed-loop system is shown in Figure 5.7.
5.4 Control

5.4.2 Sensor and actuator attacks

Here, we consider sensor and actuator attacks. We propose a simple yet effective technique to stabilize the system by switching off the isolated attacked actuators, i.e., by removing the columns of $B$ that correspond to the isolated actuators, and closing the loop with a multi-observer based output dynamic feedback controller, see Figure 5.8. We introduce a switching signal $ho(k) \subseteq \{1, \ldots, n_u\}$, containing the isolated attack-free actuators, i.e., $\rho(k) := \{1, \ldots, n_u\} \setminus \hat{W}_u(k)$. This $\rho(k)$ is used to denote actuators that are kept switched on. That is, $\rho(k) = J$ if the subset $J \subseteq \{1, \ldots, n_u\}$ of actuators are switched on and the remaining actuators are switched off at time $k$. Here, we are assuming that there is a secure communication channel that can be used to transmit $\rho(k)$ to the actuators of the system so that, once $\rho(k)$ has arrived, we could physically turn the isolated actuators off. Again, let $B$ be partitioned as $B = [b_1, \ldots, b_i, \ldots, b_{n_u}]$. After switching off the subset $\{1, \ldots, n_u\} \setminus \rho(k)$ of actuators, system (5.1) is written as follows

$$
\begin{align*}
    x^+ &= Ax + b_{\rho(k)}(u^{\rho(k)} + a_{\rho(k)}'), \\
    y &= Cx + a_y,
\end{align*}
$$

(5.53)

where $b_{\rho(k)}$ is the matrix whose columns are $b_i, i \in \rho(k)$, vectors $u^{\rho(k)}$ and $a_{\rho(k)}'$ are the inputs and attacks corresponding to the switched-on actuators, respectively. We first consider the case when the complete multi-observer estimator in Section 5.2.1 exists, i.e., $\hat{x}$ is generated by (5.7)-(5.9). We estimate $\hat{a}_{u}(k)$ using (5.40) and obtain $\hat{W}_u(k)$ from (5.44).

Then, we switch off the set $\hat{W}_u$ of actuators by letting $\rho(k) = \bar{J}(k) = \{1, \ldots, n_u\} \setminus \hat{W}_u(k)$. Since $a_{ui}(k) = 0, i \in \bar{J}(k)$, system (5.53) has the following form:

$$
x^+ = Ax + b_{\bar{J}(k)}u^{\bar{J}(k)}
$$

(5.54)

where $u^{\bar{J}(k)} \in \mathbb{R}^{\text{card}(\bar{J}(k))}$ is the set of isolated attack-free inputs. Let $0 < q^* < n_u$ be the largest integer such that $(A, b_J)$ is stabilizable for each set $J \subseteq \{1, \ldots, n_u\}$ with $\text{card}(J) \geq n_u - q^*$ where $b_J$ denotes a matrix whose columns are $b_i$ for $i \in J$. We assume that at most $q^*$ actuators are attacked. It follows that $n_u - q^* \leq \text{card}(\bar{J}(k)) \leq n_u$. We assume the following.
Assumption 5.4. For any subset $J$ with cardinality $\text{card}(J) = n_u - q^*$, there exists a linear switching state feedback controller $u_{\bar{J}}(k) = K_{\bar{J}(k)}x$ such that the closed-loop dynamics:

$$x^+ = (A + b_{\bar{J}(k)}K_{\bar{J}(k)})x + b_{\bar{J}(k)}K_{\bar{J}(k)}e,$$  \hspace{1cm} (5.55)

is ISS with respect to input $e$ for $b_{\bar{J}(k)}$ arbitrarily switching among all $b_{J'}$ with $J \subset J' \subset \{1, \ldots, n_u\}$ and $n_u - q^* \leq \text{card}(J') \leq n_u$.

Remark: We do not give a method for designing the linear switching state feedback controller $u_{\bar{J}(k)} = K_{\bar{J}(k)}x$. Standard results for designing switching controllers, for instance results in [17] and references therein, can be used to design controllers satisfying Assumption 5.4. By switching off the set $\hat{W}_u(k)$ of actuators at time $k$, using the controller designed for the set $\bar{J}(k)$, and letting $u_{\bar{J}(k)} = K_{\bar{J}(k)}\hat{x}$, the closed-loop system can be written as (5.55) with estimation error $e = \hat{x} - x$ generated by some nonlinear difference equation (5.39). Because in Theorem 5.1, we have proved that $e(k)$ converges to zero exponentially uniformly in $x(k)$, $a_u(k)$ and $a_y(k)$, the error $e(k)$ in (5.55) is a vanishing perturbation. Hence, under Assumption 5.4, it follows that $\lim_{k \to \infty} x(k) = 0$.

Next, assume that a complete multi-observer estimator does not exist but a partial multi-observer estimator exists (Section 5.2.2), i.e., $\hat{x}$ is generated from (5.26)-(5.28) and $q_1 \leq q^*$. We assume that at most $q_1$ actuators are attacked. We construct $\hat{x}(k)$ from (5.26)-(5.28), estimate $\hat{a}_u(k)$ using (5.40), and obtain $\hat{W}_u(k)$ from (5.44). After switching off the set $\hat{W}_u(k)$ of actuators, the system has the form (5.54) with $n_u - q_1 \leq \text{card}(\bar{J}(k)) \leq n_u$. 

We assume the following.

**Assumption 5.5.** For any subset $J$ with cardinality $\text{card}(J) = n_u - q_1$, there exists a linear switching state feedback controller $u^{(k)}_J = K^{(k)}_J x$ such that the closed-loop dynamics (5.55) is ISS with respect to input $e$ for $b^{(k)}_J$ arbitrarily switching among all $b^{(k)}_J'$ with $J \subset J' \subset \{1, \ldots, n_u\}$ and $n_u - q_1 \leq \text{card}(J') \leq n_u$.

Using the controller designed for the set $\bar{J}(k)$, and letting $u^{(k)}_{\bar{J}} = K^{(k)}_{\bar{J}} x$, the closed-loop dynamics can be written in the form (5.55). Then, in this case, $e(k)$ is generated by some nonlinear difference equation of the form (5.46). Under Assumption 5.5, the closed-loop dynamics (5.55) is ISS with input $e(k)$, see [33]. Moreover, in Theorem 5.2, we have proved that $e(k)$ converges to the origin exponentially uniformly in $x(k), a_u(k)$ and $a_y(k)$.

The latter and ISS of the system dynamics imply that $\lim_{k \to \infty} x(k) = 0$ [33].

**Example 6:** Consider the following system:

\[
\begin{aligned}
\begin{cases}
    x^+ &= \begin{bmatrix} 0.5 & 0 & 0.1 \\ 0.2 & 1.7 & 0 \\ 1 & 0 & 0.3 \end{bmatrix} x + \begin{bmatrix} 0.5 & 0 & 1 \\ 1 & 1 & 1 \\ 0 & 0 & 1 \end{bmatrix} (u + a_u), \\
    y &= \begin{bmatrix} 1 & 2 & 0 \\ 0 & 1 & 1 \\ 0 & 1 & 2 \\ 1 & 1 & 1 \end{bmatrix} x + a_y.
\end{cases}
\end{aligned}
\] (5.56)

Since $(A, b_i)$ is stabilizable for $i \in \{1, 2, 3\}$, we have $q^* = 2$. It can be verified that there does not exist a complete UIO for any $S_s \subset \{1, 2, 3, 4\}$ with $\text{card}(S_s) = 2$, but partial UIOs exists for each pair $(J_u, J_s)$ with $\text{card}(J_u) \leq 2$ and $\text{card}(J_s) \geq 2$; then, we have $q_1 = q_2 = 1$ and $q_1 < q^*$. We let $W_u = \{3\}, W_y = \{2\}$, and $a_{u3}, a_{y2} \sim U(-10, 10)$. We construct $(3^3) \times (3^1) + (3^2) \times (4^1) = 30$ UIOs and use the design method given in [17] to build controllers for actuators $\{1, 2\}, \{1, 3\}, \{2, 3\}, \{1, 2, 3\}$. Then, we use the partial multi-observer approach in Section 5.2.2 to estimate the state, reconstruct the attack signals and control the system. The state of the system is shown in Figure 5.9.
5.5 Conclusion

We have addressed the problem of state estimation, attack isolation, and control for discrete-time linear time-invariant (LTI) systems under (potentially unbounded) actuator and sensor false data injection attacks. Using a bank of Unknown Input Observers (UIOs), we have proposed an estimator that reconstructs the system states and the attack signals. We use these estimates to isolate attacks and control the system. We propose an effective technique to stabilize the system by switching off the isolated actuators. Simulation results are provided to illustrate our results.
Chapter 6
Conclusions and Future Work

6.1 Conclusions

In this thesis, we have addressed particular problems on secure estimation, attack detection and isolation, and control for discrete-time linear and nonlinear systems under sensor and actuator attacks. In what follows, we summarize the main contributions of the thesis and provide some future research directions.

As the first contribution of the thesis, we have extended the multi-observer approach for secure estimation—which was first introduced for linear continuous-time systems—to a large class of nonlinear systems under sensor attacks, disturbances, and measurement noise. We have provided a “unifying estimation framework” that covers many classes of nonlinear systems and observers. Despite of sensor attacks and noise, the proposed multi-observer estimator is capable of providing robust state estimates. Compared to existing secure estimation techniques (which mainly deal with linear systems), the multi-observer approach has a broader range of potential applications as it can deal with a large class of nonlinear systems as long as observers with certain stability property exist. We have also derived the corresponding estimation results for noise-free nonlinear systems under both, actuator and sensor attacks.

For discrete-time linear systems under sensor and actuator attacks, we have provided tools for designing multi-observer based estimators such that: 1) the system state is exponentially reconstructed; 2) actuator and sensor attacks are isolated; 3) the adversarial signals injected by attackers (by tampering with sensors and actuators) are exponentially estimated; and 4) the closed-loop dynamics is stabilized despite of sensor/actuator at-
tacks.

We have provided extensive simulation results for several classes of nonlinear (and linear) systems to illustrate the performance of our results.

6.2 Future Work

6.2.1 Enhanced Secure Estimation Techniques

The proposed multi-observer approach can solve the problem of state estimation for a large class of nonlinear systems under sensor and actuator attacks. However, when the numbers of sensors and actuators increase, the number of the observers that the estimator requires grow considerably. Then, a future research direction might be to reduce the number of observers needed for constructing the estimator without compromising the estimation performance. A possible solution could be to secure some sensors/actuators, i.e., we can select a group of sensors/actuators to allocate security equipment so that attacks to those secured ones are practically impossible. This would considerably reduce the required number of observers that our estimator needs at the price of enabling extra hardware. However, the following question arises: given a limited number of sensors/actuators that we can secure, which ones should be selected to maximize robustness of the estimation scheme against attacks? This class of problems is known in the literature as security allocation problems.

6.2.2 Attack Isolation for Noisy Systems

When sensor/actuator attacks, sensor noise, and system disturbances are present, isolation of attacks is, in general, more challenging than in the noise-free case. This is because it might be difficult to distinguish between noise/disturbances and attacks. In Chapter 2, we develop a threshold-based algorithm for isolating attacked sensors. In Chapter 3 and 5, assuming sensors and actuators are both attacked, the isolation results are derived for systems without perturbations. In this case, we can isolate attacked sensors/actuators by checking the sparsity pattern of the reconstructed attack signals. This is no longer
6.2 Future Work

possible when noise/disturbances drive the system dynamics. Efficient isolation algorithms for disturbed systems under both sensor and actuator attacks are missing in the literature. A possible solution to this isolation problem could be to use known bounds or probability distributions on the driving disturbances together with standard change detection methods to construct new efficient isolation algorithms.

6.2.3 Secure Control of Nonlinear Systems under Actuator Attacks

We have provided a control scheme for LTI systems under sensor/actuator attacks. However, we have not considered the problem of secure control for nonlinear systems when sensor and actuator attacks both occur, i.e., how to synthesize controllers that are capable of stabilizing nonlinear systems in the presence of sensors/actuator attacks. This is an interesting (and very challenging) problem that the community has not even begun to address.
Bibliography


Minerva Access is the Institutional Repository of The University of Melbourne

Author/s:
Yang, Tianci

Title:
Multi-observer approach for estimation and control under adversarial attacks

Date:
2019

Persistent Link:
http://hdl.handle.net/11343/227777

File Description:
Final thesis file

Terms and Conditions:
Terms and Conditions: Copyright in works deposited in Minerva Access is retained by the copyright owner. The work may not be altered without permission from the copyright owner. Readers may only download, print and save electronic copies of whole works for their own personal non-commercial use. Any use that exceeds these limits requires permission from the copyright owner. Attribution is essential when quoting or paraphrasing from these works.