Summer 2024

Zero Dynamics Attacks on Unknown Bilinear Systems: Vulnerability and Detection

Mohammad Aminul Haq
Old Dominion University

ZERO DYNAMICS ATTACKS ON UNKNOWN BILINEAR SYSTEMS: VULNERABILITY AND DETECTION

by

Mohammad Aminul Haq
B.Sc. in Electrical and Electronic Engineering, 2009
Rajshahi University of Engineering and Technology, Bangladesh
M.Sc. in Control and Automation Engineering, 2016
Yildiz Technical University, Turkey

A Dissertation Submitted to the Faculty of
Old Dominion University in Partial Fulfillment of the
Requirements for the Degree of

DOCTOR OF PHILOSOPHY

ELECTRICAL AND COMPUTER ENGINEERING

OLD DOMINION UNIVERSITY
August 2024

Approved by:

W. Steven Gray (Director)

Luis A. Duffaut Espinosa (Member)

Oscar R. González (Member)

ABSTRACT

ZERO DYNAMICS ATTACKS ON UNKNOWN BILINEAR SYSTEMS:
VULNERABILITY AND DETECTION

Mohammad Aminul Haq
Old Dominion University, 2024
Director: Dr. W. Steven Gray

Critical infrastructure requires a safe and secure operating environment because of its significant impact on society. Its large-scale size and distributed sensors and actuators make it vulnerable to cyber-physical attacks. A zero dynamics attack is a type of cyber-physical attack where an adversary keeps the output of the target constant (classically zero), while forcing some of the internal states to deviate from their nominal values. Most of the existing work in the literature assumes the system dynamics are linear and available to an adversary. The first goal of this dissertation is to show that an adversary can successfully execute a malicious zero dynamics attack on an unknown bilinear system. The main motivation is to understand the nature of the vulnerability so that appropriate defensive measures can be taken. The second goal is to develop an algorithm for attack detection so that the system can be secured against such attacks. To demonstrate the methodology, a bilinear model of a petro-chemical plant was chosen. Two types of zero dynamics attacks are considered, an observer-based approach and an analytical approach. Both approaches are simulated numerically and found to be effective. Then an observer-based attack detection method is developed. The proposed state observer for a bilinear system is designed using Lyapunov theory and convex optimization concepts. The observer monitors all the unmeasurable states so that any deviation in any state from it's nominal value can be immediately detected in the event of an attack. The designed observer is implemented for the petro-chemical plant and found to be effective in detecting the onset of zero dynamics attacks.

ACKNOWLEDGMENTS

I am profoundly grateful to my supervisor, Dr. W. Steven Gray, for his unwavering support, patience, insightful guidance, and encouragement throughout this journey. His expertise and mentorship have been invaluable in shaping this dissertation and my academic growth.

I extend my deepest gratitude to the members of my doctoral committee, Dr. Luis A. Duffaut Espinosa (Department of Electrical and Biomedical Engineering, The University of Vermont) and Dr. Oscar R. González for their constructive feedback and scholarly insights that have significantly enriched this work.

I especially want to thank the faculty and staff of the Department of Electrical and Computer Engineering at Old Dominion University for all their assistance and support. I also want to acknowledge the National Science Foundation for its support under grant CMMI-1839378, which enabled me to begin my research journey. I am indebted to my colleagues and friends whose emotional and intellectual support helped me substantially throughout this endeavor.

My heartfelt appreciation goes to my wife and children for their unwavering belief in me and their unconditional love and support. Their patience, understanding, and encouragement have sustained me through the challenges of pursuing a Ph.D. This journey would not be possible without my parents' continuous support and encouragement. Their love, support, and trust in me have strengthened me to reach this milestone.

Lastly, I am grateful to the creator of the universe, who created some beautiful minds whose biographies and discoveries always encourage me to investigate the unknown and work for the advancement of humanity.

TABLE OF CONTENTS

Page

Page

LIST OF TABLES

LIST OF FIGURES

CHAPTER 1

INTRODUCTION

1.1 CYBER-PHYSICAL SYSTEMS

Cyber-physical systems (CPS) are smart systems that link physical systems with the computing power of cyber systems. Helen Gill of the United States National Science Foundation (NSF) first coined the term cyber-physical system [49]. Integrating computation, communication, and control with physical systems laid the foundation of this highly interdisciplinary field [13]. Modern analytical tools and concepts from system theory, such as state space analysis, system identification, robust control, estimation, and optimization, are put together with the latest technology in communication and computer networking like 5G, WiFi, and multi-core computation to build efficient and smart physical systems. Internet of Things (IoT), like drone delivery and autonomous vehicle control, are making CPS more versatile and part of our daily life. Similarly, introducing the industrial internet, remote monitoring and controllability features into industry makes systems more efficient and productive. These developments in cyber-physical systems improve our quality of life, help us interact with systems seamlessly, and simplify system monitoring. Ensuring the secure operation of these systems becomes a growing concern.

Cyber-physical system security deals with any threat, not a customarily known process error, initiated from the outside world, which is not legitimate in order to induce or interfere with the system's operation in some way. Critical infrastructure, such as petrochemical plants, water plants, and power grids, are examples of cyber-physical systems that require a safe and secure operating environment because of their significant impact on society. Their large-scale size and distributed sensors and actuators make them vulnerable to cyber-physical

attacks. With the advent of the internet and modern communication systems, much of this infrastructure is monitored and controlled over complex communication networks. This adds one more inlet node/surface for cyber-physical attacks. The recent Stuxnet malware attack, the US-Canada 2003 blackout, and the Maroochy Shire Council Sewage control event are examples of such incidents [8]. In general, this kind of public infrastructure is more attractive prey to adversaries for cyber-physical attacks because of their ability to create a multi-dimensional impact on a larger population group. For example, a cyber attack on the Colonial pipeline on May 7, 2021 affected a large population and had great economical consequences. This oil pipeline originates in Houston, Texas, and supplies gas and jet fuel mainly to the southeastern states of Alabama, Florida, North Carolina, South Carolina, and as far away as New York. The East Coast of the United States gets 45% of its total gas via this pipeline. As a consequence of the attack, Charlotte Douglas International Airport and Hartsfield-Jackson Atlanta International Airport experienced a disruption in fuel supply, and panic buying created fuel shortages at filling stations in those states. Moreover, Colonial had to pay the hackers $4.4 million as ransom in Bitcoin though the Justice Department recovered $2.3 million in cryptocurrency later. In 2021, more than $45 million was paid as ransom in 292 cyber attacks on a variety of organizations. Cybercrime will cost companies worldwide an estimated $10.5 trillion annually by 2025, up from $3 trillion in 2015. This is a growth rate of 15 percent.

A typical layered architecture of a cyber-physical system is shown in Figure 1. Cyber-physical systems have three major components or layers: the physical plant layer, the communication network layer, and the computational and control system layer [7]. Each layer is modeled as a distinct system and interacts with other layers to run the whole cyber-physical system. Cyber physical attacks can happen in any of these layers. These layers are modeled as either a linear or nonlinear system. The system in each layer further can be classified as

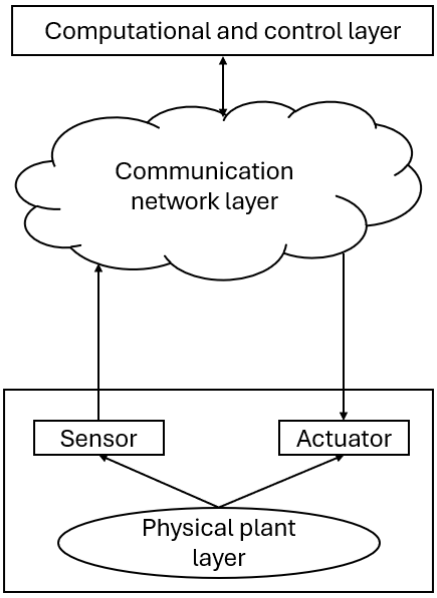

Figure 1: Cyber-physical systems layer architecture

continuous-time or discrete-time and analog or digital. In general, the communication network layers and computational and control layers are digital systems, whereas the physical plant layer is mostly a continuous-time analog system. The focus of this dissertation is on plants that can be represented in terms of a bilinear model.

1.2 BILINEAR SYSTEMS

In control theory, a bilinear system refers to a class of state space systems that exhibits a certain kind of nonlinearity in its state dynamics [52]. Specifically, bilinearity arises when the state dynamics depend on the multiplicative interaction between states and inputs. A continuous-time bilinear system is commonly expressed as

$$\dot{z} = N_0 z + \sum_{i=1}^{m} N_i z u_i \tag{1a}$$

$$y = Cz, \tag{1b}$$

where $N_0 \in \mathbb{R}^{n\times n}$ is the state matrix, and $N_i \in \mathbb{R}^{n\times n}$, $i = 1, 2, \ldots, m$ are the coupling matrices between the states and the input vector $u \in \mathbb{R}^m$. The output equation has a linear structure, where $C \in \mathbb{R}^{p\times n}$ is the output matrix.

Bilinear systems appear in a number different fields such as in engineering, biology, ecology, and medicine. In control engineering many real systems or sub-systems are modeled as bilinear systems. For example, the braking system for an automobile, fluid transfer in process control, and nuclear reactor control are all represented using bilinear models [44]. Bilinear systems are also known to be universal approximators for a large class of control affine nonlinear systems [53].

1.3 ZERO DYNAMICS ATTACKS

The zero dynamics of a system refer to the internal dynamics of the system when the output is exactly zero for all positive time. These dynamics are uniquely defined about any point in the state space where the system has relative degree and the zero output is in the range of the input-output map [26]. These internal dynamics can be related to the location of the zeros of the linearized model at an equilibrium point. A zero dynamics attack (ZDA) is a type of cyber-physical attack where an adversary keeps the output of the target constant (classically zero), while forcing some of the internal states to deviate from their nominal values. The design of the attack signal requires exact knowledge of the plant's zero dynamics [27]. This is rarely available in practice since plant uncertainty is always present. However, most of the existing work in the literature assumes the system dynamics are both linear and available to an adversary (see, for example, [8, 54, 56]). One exception to this is given in [12], where a known bilinear network with attacks modeled as additive and multiplicative disturbances can be made more or less vulnerable to such attacks.

Another example is given in [18, 17], where it is shown that estimating the Chen-Fliess series representation of the nonlinear input-output map was enough to synthesize a *universal* zero dynamics attack, i.e., no state space model of the targeted system was necessary in this instance. But it is not clear how effective this approach would be in general as nonlinear system identification is a difficult problem [50]. On the other hand, the system identification problem for bilinear systems was largely solved in [29, 30, 48]. The assertion is that this creates a ZDA vulnerability for systems with bilinear dynamics.

1.4 SYSTEM IDENTIFICATION

System identification is an active research area in the control system community. This field has a class of well documented/developed research resources which serve as a basis for research in many other fields like machine learning, optimal control, artificial intelligence, cyber-security, etc. For zero dynamics attacks, system identification plays a major role in building a model of the targeted system prior to the synthesis of an attack signal. The specific identification algorithm used is completely determined by the dynamical properties of the targeted system. Because of the diverse nature of system dynamics, researchers have developed a wide variety of identification algorithms. Linear time invariant (LTI) systems are a well understood branch of control theory. The classical approaches to identifying an LTI system are prediction error methods and subspace identification methods [47]. Researchers also have extended and/or modified these linear identification methods in several different research directions to adapt them for nonlinear system identification. However, nonlinear system identification is significantly more complex, and mainly *ad hoc* methods or partial solutions are found in the literature. Most of the available nonlinear system identification algorithm utilize features from numerical, probabilistic, and statistical analysis along with

control theory. Since this dissertation is focused on continuous-time bilinear systems, the scope of the literature review for system identification is therefore limited to this class of systems only.

In [4], the authors used Hartley modulating function (HMF) method to identify a bilinear continuous-time system. The Hartley modulating function method replaces the input-output differential equation to represent the system's behavior. It utilizes the known derivatives of the Hartley function instead of the derivatives of inputs and outputs by applying an integral transformation to the signals. In [23], the authors used shifted Legendre polynomials to identify the parameters of a bilinear system. It treats the product of two time functions as a single function. This enables one to represent the state equations in a computationally convenient matrix-algebraic form. This form is then used to determine the unknown parameters of a bilinear system via the operational properties of the Chevyshev polynomials from the input-output data [38]. In [28], an online recursive algorithm using Walsh functions for estimating the parameters of a bilinear system is presented. In [10], the authors proposed a method that uses block-pulse functions to obtain a robust estimate of a bilinear realization. They implement an approach that minimizes a robust performance criterion to reduce the effect of noise and significant errors on the expansion coefficients. These coefficients are then used to obtain the parameters of the bilinear system. The authors in [30] used a novel idea of applying a train of pulses to the system to transform the bilinear system into a linear system for a brief period of time. Then they utilize identification tools developed for linear system identification to identify a state space representation of an unknown bilinear system. As discussed later, this method was found to be very suitable for ZDA design, modulo some adjustments, so this will provide the primary identification framework for this work.

1.5 ATTACK DETECTION

The main objective of the cyber physical systems security field is to ensure secure operation of a system against any cyber attack. Different methods for attack detection have been devised depending on the nature of the attack, i.e., which layer of the CPS is compromised. A general security approach is to incorporate an anomaly detection system which uses input and output measurements to detect any anomaly in the system operation. Such methods are already widely used in industry to keep the operation of a system smooth and uninterrupted in event of a fault occurrence [15]. While fault detection theory laid the foundation of many detection algorithms, the disadvantage of this approach is that most classical detection algorithms were designed to detect and respond to machine failure, random faults and accidents, not deliberate attacks [15].

The standard fault detection approach is to use a prediction model to predict the system behavior and compare it against the currents measurements to identify any deviation from normal behavior. The governing equations of the physical system such as Newton's laws, energy preservation laws, and fluid dynamics can be utilized to develop a model of the system, or a model can be build from past input-output observations of the systems. Auto-Regressive (AR) models or Auto-Regressive Moving Average (ARMA) with exogenous inputs, in the case where there is noise, can be used to develop a prediction model of the system. For a linear dynamical system, a state space model is often used, and in case of nonlinear systems, a linearized state space model can be determined to predict the probable output given that the nonlinear system is stable in a neighborhood of an equilibrium point. The error between the model outputs and the measured outputs are checked against a threshold value to raise an alarm. But this approach for zero dynamics attack detection is not applicable since zero dynamics attacks do not make significant changes in the system's output. There are other kinds of cyber attacks described in the literature such as denial-of-service attacks, replay

attacks, and deception attacks [59]. But zero dynamics attacks are known to be among the most lethal as they can cause a significant amount of damage to the physical system in a short period of time. In this dissertation, the zero dynamics attack detection problem will be addressed in detail, thus, the literature review in this section is limited to this class of attacks only.

In [54], the authors address zero dynamics attacks on a linear time-invariant system where the attacker conducts the attack by a stealthy data-injection to the control system. They first present a method to quantify the degree of stealthiness of an attack and then describe a detection method based on modifying the system's structure. The authors also assumed that the attacker had a priori complete knowledge of the model of the state space system and use the model to synthesize the attack signal. The proposed method of zero dynamics attack detection is to modify the input, output, or system matrices such that the resultant state vectors are no longer in the kernel space of the output matrix. Therefore, the output is no longer zero during the attack.

In [56], the authors consider the problem of zero dynamic attacks on linear distributed control systems (DCSs). In general, DCSs have a diverse set of sensors and controllers which are often managed by independent agents. In this paper, the authors develop a method to prevent zero dynamics attacks when as many as p agents and sensors are corrupted. The presence of a zero dynamics attack is detected in terms of the structural interaction between agents and sensors. Graph theory is used to obtain the necessary and sufficient conditions for the presence of zero dynamics attacks in terms of the structural interactions between agents and sensors. The problem is framed as a optimization problem to minimize the cost of communication/sensing while ensuring the desired system robustness against attacks.

In [36], the authors propose a generalized sampler instead of a simple sampler to shift the zeros of a sampled-data linear system inside the unit circle, rendering zero dynamics attacks

ineffective to the system. The generalized sampler is designed by taking a weighted average of multiple samples obtained over a single sampling period. An optimal procedure is designed to select the locations of the zeros such that the error between the general sampler and the simple sampler is minimized. A similar kind of approach is presented in [35]. A generalized version of a zero-order hold is used to counter zero dynamics attacks. The effectiveness of the method is demonstrated using a DC-DC converter. The authors of [19] show that in some cases, the required amplitude of the generalized hold becomes unrealistically large, which demands the inclusion of an input system capable of supplying comparatively large input signals. Replacing an intrinsic zero (removing one and adding a new one) requires an excessively large amplitude of the hold functions, whereas just adding a new one may lead to a relatively smaller amplitude compared to the replacement case.

In [2], an auxiliary system and detection filter is introduced to detect zero dynamic attacks in a linear time-invariant system. The key feature of this approach is that attackers cannot design an undetectable attack that significantly affects the system performance and stability, even though attackers might have full knowledge of the plant, the auxiliary system, and filters. Their proposed method utilizes a plant side auxiliary (PSA) and two command and control (C&C) side filters to detect zero dynamics attacks.

The authors in [3] formally address the zero dynamics attack on a linear single-input, single-output (SISO) time-delay system. They introduce a novel zero dynamics attack input class for infinite dimensional systems. They also provide sufficient conditions for SISO time-delay systems to be resilient against this class of zero dynamics attack inputs.

In [55], the authors consider a zero dynamics attack against a wind power system. The mathematical model of the wind power system is obtained by transforming a semi-direct drive permanent magnet synchronous generator (D-PMSG) into an equivalent circuit. The zero dynamics attack is analyzed by separating the stable and unstable states of the system

with the help of relative degree. A protection scheme is developed by combining a multiple linear regression (MRL) predictive control with the Byrnes-Isidori normal form.

In [41], the authors considered second-order multi-agent systems under zero dynamics attack and devised a topology switching method to detect such attacks. They characterize the detectability of ZDAs to derive sufficient and necessary conditions for a Luenberger observer under the switching topologies to detect the attack. Moreover, the observer serves as a state estimator in the absence of attacks. In [43], the authors tackled the ZDA problem where the attacker is aware of the topology-switching strategy and employs the "pause (update and resume) attack" technique to avoid detection. The detectability of the proposed method is developed in terms of the network topology, the set of monitored agents, and the set of measurements of the monitored agents. A similar approach is used to detect cooperative zero-dynamics attacks in the context of coupled harmonic oscillators [42].

The authors in [24] used a modulation matrix in the path of the control variables to detect covert and zero dynamic attacks in cyber-physical systems. By inserting a modulation matrix in the control variables, the input behavior of the process is altered. Therefore, the adversary loses perfect knowledge of the system, which assists in revealing the attack. Though the modulation matrix is designed by focusing on covert attacks, guaranteeing the modulation matrix changes the input directions, zero dynamics attacks also can be detected because multivariable zeros depend on the frequency and direction.

In this dissertation, an estimator-based attack detection method is investigated. The idea behind this estimator-based approach is that an estimator estimates all the states with sufficient accuracy such that any deviation from its nominal trajectory will be readily detectable even though the output shows no deviation. The proposed state estimator for a bilinear system is designed using the Lyapunov stability theory and convex optimization concepts. Though convex optimization methods have been used to solve certain types of nonlinear

problems, most of them are very problem-specific, and their adaptation to other nonlinear problem-solving techniques is quite complicated. For example, three different convex optimization methods to design a general nonlinear observer are presented in [25]. A multivariable sector condition must be satisfied to incorporate the nonlinearity into the design. However, this class of systems satisfying the multivariable sector condition is very limited. Alternative design approaches in [1] and [51] mainly depend on determining maximum singular values of certain positive semi-definite matrices so that the observer's linear and bilinear error dynamics become stable for the given input class. These methods ensure the observer's stability, but the procedures to obtain the observer gains are relatively complex. Instead of employing an existing bilinear observer to detect the attack, this dissertation proposes a new bilinear observer that is simpler to design. Specifically, the error dynamics of the observer are used to form a Lyapunov function. The Lyapunov function is optimized using a convex optimization method which finds observer gains that null the contribution from the bilinear part of the system. The design method is much simpler than existing methods while yielding good performance.

1.6 PROBLEM STATEMENT

The main objectives of this dissertation are to:

1. Show how in general an adversary can successfully execute a malicious zero dynamics attack on an unknown bilinear system.

2. Demonstrate by simulation a zero dynamics attack using a bilinear model of a petrochemical plant.

3. Develop methods to detect a zero dynamics attack on a bilinear system.

4. Demonstrate by simulation the effectiveness of these attack detection methods on the petro-chemical plant.

1.7 MAIN CONTRIBUTIONS

1. Characterized the vulnerability of bilinear systems to zero dynamics attacks using system analysis techniques.
2. Showed that no a prior knowledge of the bilinear system is required to conduct a successful zero dynamics attack.
3. Demonstrated two different attack strategies, namely, the observer-based and analytical methods.
4. Designed an observer-based method to detect a zero dynamics attack on a general bilinear system.
5. Demonstrated the efficacy of the attack detection system using a model of a physical plant.

1.8 THESIS OUTLINE

The dissertation is organized as follows. The dynamics of a general bilinear system and its properties pertinent to zero dynamics attacks are described in Chapter 2. The system identification algorithm and attack design procedures are presented in Chapter 3. The attack detection methods for bilinear systems against zero dynamics attacks are described in Chapter 4. Two numerical examples demonstrating zero dynamics attack on a bilinear plant and its detection are given in Chapter 5. The conclusions are presented in Chapter 6.

CHAPTER 2

BILINEAR SYSTEMS

The main goal of this chapter is to present those elements of bilinear system theory that are used through this dissertation. One reason for choosing bilinear systems is that a wide range of nonlinear systems can be approximated to arbitrary accuracy using bilinear models [53]. This chapter begins by defining a bilinear state space representation. Then the notion of relative degree, zero dynamics, and non-minimum phase are presented in general and then specialized to the bilinear case. Finally, a detailed description of a bilinear model for a petro-chemical plant, which serves as a model for the case study in the subsequent chapters, is given. The plant is linearized around an equilibrium point, and the linearized plant's controllability, observability, and non-minimum phase properties are presented.

2.1 GENERAL BILINEAR SYSTEMS

A bilinear system is a nonlinear system where states and input interact with each other directly. A bilinear system uses a multiplicative operation between state vector and the inputs to capture the nonlinearity in the system. In the state space setting, a continuous-time bilinear system has the form

$$\dot{z} = N_0 z + \sum_{i=1}^{m} N_i z u_i \tag{2a}$$

$$y = Cz, \tag{2b}$$

where $N_0 \in \mathbb{R}^{n\times n}$ is the state matrix, $N_i \in \mathbb{R}^{n\times n}$, $i = 1, 2, \ldots, m$ are the coupling matrices between the states and the input vector, and $u \in \mathbb{R}^m$. The output equation has a linear

structure, where $C \in \mathbb{R}^{p \times n}$ is the output matrix.

2.2 RELATIVE DEGREE OF SYSTEM

For dynamical systems, the concept of relative degree is used to describe a certain relationship between the input and output, specifically the number of times the output must be differentiated before the input appears explicitly. Relative degree is fundamental in controller design, system analysis, and observer design [26]. It also plays an important role in determining whether a system is vulnerable to zero dynamics attacks. In order to develop the concept of relative degree for nonlinear systems, the relative degree of a linear time-invariant system is first described strictly in the time-domain setting.

2.2.1 Relative degree of linear system

Consider an LTI system

$$\begin{aligned} \dot{z} &= Az + Bu \\ y &= Cz + Du, \end{aligned} \tag{3}$$

where $z(t) \in \mathbb{R}^n$ is the state vector, $u(t) \in \mathbb{R}$ is the input, $y(t) \in \mathbb{R}$ is the output, and A, B, C, and D are matrices of appropriate dimensions. The relative degree r of the system is the smallest integer such that the r-th derivative of the output y depends explicitly on the input u. If $D \neq 0$, then the relative degree $r = 0$ because the output y directly depends on the input u. Taking the derivative of the output equation of (3) with $D = 0$ gives

$$\begin{aligned} \dot{y} &= C\dot{z} \\ &= C(Az + Bu) \\ &= CAz + CBu. \end{aligned}$$

The process of taking derivative of the output is continued until the input u appears explicitly or the coefficient of the input u is nonzero in the derivative, that is

$$\begin{aligned}
\ddot{y} &= C\ddot{z} \\
&= C(A\dot{z} + B\dot{u}) \\
&= C(A(Az + Bu) + B\dot{u}) \\
&= CA^2z + CABu + CB\dot{u} \\
&\vdots \\
y^{(r)} &= CA^rz + CA^{r-1}Bu + \cdots + CABu^{r-2} + CBu^{r-1}.
\end{aligned}$$

The LTI system having a relative degree r means $CA^{r-1}B \neq 0$ and $CA^kB = 0$ for $k = 0, 1, \ldots, r-2$. Every linear system has relative degree $0 \leq r \leq n$.

Example 2.1: Consider the LTI system with the following state space representation:

$$\begin{aligned}
\dot{z} &= \begin{bmatrix} 0 & 1 \\ -2 & -3 \end{bmatrix} z + \begin{bmatrix} 0 \\ 1 \end{bmatrix} u \\
y &= \begin{bmatrix} 1 & 0 \end{bmatrix} z.
\end{aligned}$$

Taking the first derivative of the output y:

$$\begin{aligned}
\dot{y} &= C\dot{z} \\
&= CAz + CBu \\
&= \begin{bmatrix} 1 & 0 \end{bmatrix} \begin{bmatrix} 0 & 1 \\ -2 & -3 \end{bmatrix} z + \begin{bmatrix} 1 & 0 \end{bmatrix} \begin{bmatrix} 0 \\ 1 \end{bmatrix} u \\
&= z_2
\end{aligned}$$

Here, input u does not appear explicitly. Compute the second derivative of the output y:

$$\ddot{y} = C\ddot{z}$$

$$
\begin{aligned}
&= CA^2z + CABu + CB\dot{u} \\
&= \begin{bmatrix} 1 & 0 \end{bmatrix} \left(\begin{bmatrix} 0 & 1 \\ -2 & -3 \end{bmatrix} \right)^2 z + \begin{bmatrix} 1 & 0 \end{bmatrix} \begin{bmatrix} 0 & 1 \\ -2 & -3 \end{bmatrix} \begin{bmatrix} 0 \\ 1 \end{bmatrix} u + \begin{bmatrix} 1 & 0 \end{bmatrix} \begin{bmatrix} 0 \\ 1 \end{bmatrix} \dot{u} \\
&= -2z_1 - 3z_2 + u.
\end{aligned}
$$

Here, input u appears explicitly. Therefore the relative degree r of the system is 2.

2.2.2 Relative degree of nonlinear system

The relative degree of a nonlinear system is a generalization of the notion of the relative degree from linear systems theory. Perfectly analogous to the linear case, the relative degree for nonlinear system is an integer that denotes the minimum number of times one needs to differentiate the output until the input explicitly appears in the resulting expression. Unlike the linear case, the relative of a nonlinear system is tied to a specific point in the state space, and the relative degree at some points many not be defined.

Consider the following single-input, single-output nonlinear system described by the state-space representation:

$$
\begin{aligned}
\dot{z} &= f(z) + g(z)u \\
y &= h(z),
\end{aligned}
\tag{4}
$$

where $z(t) \in \mathbb{R}^n$ is the state vector, $u(t) \in \mathbb{R}$ is the input, $y(t) \in \mathbb{R}$ is the output, f and g are smooth vector fields, and $h(z)$ is a smooth scalar-valued function. System (4) is said to have relative degree r at a point z_0 if:

1. $L_g L_f^k h(z) = 0$ for all z in a neighborhood U of z_0 for $k = 0, 1, \ldots, r-2$,

2. $L_g L_f^{r-1} h(z) \neq 0|_{z=z_0}$,

where $L_f h$ denotes the Lie derivative of h along f, $L_g L_f h$ denotes the Lie derivative of h first along vector field f and then along vector field g. The notation $L_f^k h$ is used when h is

differentiated k times along f. Note that the relative degree is defined at the point z_0. There might be points where the relative degree cannot be defined. Consider the following example.

Example 2.2: A controlled Van der Pol oscillator in state space form is

$$\dot{z} = f(z) + g(z)u = \begin{bmatrix} z_2 \\ 2\omega\xi(1-\mu z_1^2)z_2 - \omega^2 z_1 \end{bmatrix} + \begin{bmatrix} 0 \\ 1 \end{bmatrix} u$$

$$y = h(z) = z_1.$$

Here, μ is a scalar parameter indicating the nonlinearity and strength of the damping, ω is the angular velocity [26]. To determine its relative degree, first compute

$$\begin{aligned} L_g h(z) &= \frac{\partial h}{\partial z} g(z) \\ &= \begin{bmatrix} 1 & 0 \end{bmatrix} \begin{bmatrix} 0 \\ 1 \end{bmatrix} \\ &= 0. \end{aligned}$$

Thus, the relative degree, if well defined, must satisfy $r > 1$. Next compute:

$$\begin{aligned} L_f h(z) &= \frac{\partial h}{\partial z} f(z) \\ &= \begin{bmatrix} 1 & 0 \end{bmatrix} \begin{bmatrix} z_2 \\ 2\omega\xi(1-\mu z_1^2)z_2 - \omega^2 z_1 \end{bmatrix} \\ &= z_2 \\ L_g L_f h(z) &= \frac{\partial L_f h}{\partial z} g(z) \\ &= \begin{bmatrix} 0 & 1 \end{bmatrix} \begin{bmatrix} 0 \\ 1 \end{bmatrix} \\ &= 1. \end{aligned}$$

Since $L_g h(z) = 0$ and $L_g L_f h(z) = 1$, the system has relative degree at every point z_0. However, if the output function is, for instance

$$y = h(z) = \sin(z_2)$$

then $L_g h(z) = \cos(z_2)$. The system has relative degree 1 at any z_0 except when $z_{02} = (2k+1)\frac{\pi}{2}$, where k is any integer. In this case, the system has no relative degree.

Example 2.3: Consider the bilinear system

$$\dot{z} = \begin{bmatrix} 1 & 0 \\ -2 & 3 \end{bmatrix} z + \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} zu$$
$$y = \begin{bmatrix} 1 & 0 \end{bmatrix} z.$$

In terms of a general control-affine nonlinear system (4), observe

$$f(z) = \begin{bmatrix} 1 & 0 \\ -2 & 3 \end{bmatrix} z$$
$$g(z) = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} z$$
$$h(z) = \begin{bmatrix} 1 & 0 \end{bmatrix} z.$$

To determine relative degree, compute the following Lie derivatives:

$$\begin{aligned} L_g h(z) &= \frac{\partial h}{\partial z} g(z) \\ &= \begin{bmatrix} 1 & 0 \end{bmatrix} \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} z \\ &= z_1, \end{aligned}$$

$$L_f h(z) = \frac{\partial h}{\partial z} f(z)$$

$$= \begin{bmatrix} 1 & 0 \end{bmatrix} \begin{bmatrix} 1 & 0 \\ -2 & 3 \end{bmatrix} z$$

$$= z_1,$$

$$L_g L_f h(z) = \frac{\partial L_f h}{\partial z} g(z)$$

$$= \begin{bmatrix} 1 & 0 \end{bmatrix} \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$$

$$= z_1.$$

It is evident from the above that $r = 1$ at any point z such that $z_1 \neq 0$.

2.3 ZERO DYNAMICS OF A SYSTEM

In general, the zero dynamics of a system corresponds to the problem of zeroing the output, that is, finding an initial state z_0 and an input function $u(t)$ such that the output $y(t)$ is identically zero for all $t \geq 0$ [26]. Understanding the precise nature of the zero dynamics of a system helps determine how vulnerable the system is to zero dynamics attacks. This section first describes an LTI system's zero dynamics using the Byrnes-Isidori normal form. Based on the LTI system's zero dynamics, the zero dynamics of a control-affine general nonlinear system is then developed. Finally, the zero dynamics of a single-input, single-output (SISO) bilinear system around an equilibrium are presented.

2.3.1 Zero dynamics of linear time-invariant systems

Consider the transfer function of a linear time-invariant system with relative degree r

$$H(s) = K \frac{b_0 + b_1 s + \cdots + b_{n-r+1} s^{n-r+1} + s^{n-r}}{a_0 + a_1 s + \cdots + a_{n-1} s^{n-1} + s^n}.$$

Suppose the numerator and denominator polynomials are relatively coprime and the initial state is $z(0)$. Therefore, it has the following minimal realization

$$\begin{aligned} \dot{z} &= \underbrace{Az}_{f(z)} + \underbrace{B}_{g(z)} u \\ y &= \underbrace{Cz}_{h(z)}, \end{aligned} \tag{5}$$

where

$$A = \begin{bmatrix} 0 & 1 & 0 & \cdots & 0 \\ 0 & 0 & 1 & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \cdots & 1 \\ -a_0 & -a_1 & -a_2 & \cdots & -a_{n-1} \end{bmatrix}$$

$$B = \begin{bmatrix} 0 \\ 0 \\ \vdots \\ 0 \\ K \end{bmatrix}$$

$$C = \begin{bmatrix} b_0 & b_1 & \cdots & b_{n-r-1} & 1 & 0 & \cdots & 0 \end{bmatrix},$$

and $K \neq 0$. The goal is to obtain a normal form for the LTI system (5) so that the zero dynamics become explicit. Since the system (5) has a relative degree r, one can write

$$\begin{aligned} x_1 &= h(z) = Cz = b_0 z_1 + b_1 z_2 + \cdots + b_{n-r-1} z_{n-r} + z_{n-r+1} \\ x_2 &= L_f h(z) = CAz = b_0 z_2 + b_1 z_3 + \cdots + b_{n-r-1} z_{n-r+1} + z_{n-r+2} \\ &\vdots \\ x_r &= L_f^r h(z) = CA^{r-1} z = b_0 z_r + b_1 z_{r+1} + \cdots + b_{n-r-1} z_{n-1} + z_n. \end{aligned}$$

This defines the first r components of a coordinate transformation $x = \Phi(z)$. In order to write the equations in a more compact manner, introduce a vector notation ξ to group the first r state variables

$$\xi = \begin{bmatrix} x_1 \\ \vdots \\ x_r \end{bmatrix}.$$

The remaining $n - r$ coordinates can then be chosen independently provided that of $\frac{\partial \Phi}{\partial z}$ is nonsingular on a neighborhood of $z(0)$. The following choices are made

$$\begin{aligned} x_{r+1} &= z_1 \\ x_{r+2} &= z_2 \\ &\vdots \\ x_n &= z_{n-r}. \end{aligned}$$

Introduce another vector notation η to group the remaining $n - r$ state variable together

$$\eta = \begin{bmatrix} x_{r+1} \\ \vdots \\ x_n \end{bmatrix}. \tag{6}$$

Since the system remains linear after applying the linear coordinate transformation Φ, the following linear structure is found

$$\begin{aligned} \dot{x}_1 &= x_2 \\ \dot{x}_2 &= x_3 \\ &\vdots \\ \dot{x}_{r-1} &= x_r \\ \dot{x}_r &= R\xi + S\eta + Ku \\ \dot{\eta} &= P\xi + Q\eta, \end{aligned} \tag{7}$$

where R, S are row vectors, and P and Q are matrices of appropriate dimensions. Now, if the output is to be identically the zero function, then clearly it is necessary that

$$y(0) = y^{(1)}(0) = \cdots = y^{(r-1)}(0) = 0,$$

or equivalently, $\xi(0) = 0$. To ensure that $\xi(t) = 0$ for $t > 0$, it is necessary to apply the input

$$u^* = \frac{-S\eta}{K}$$

so that $y^{(r)} = 0$. This homogeneous linear ODE has only the trivial solution $y = 0$ since all the initial conditions are zero by assumption. Therefore, the dynamics of $\eta(t)$ reduces to

$$\dot{\eta} = Q\eta, \quad \eta(0) = \eta_0. \tag{8}$$

Using the particular choice of η in (6), one can find the structure of the Q matrix

$$\begin{aligned}
\dot{x}_{r+1} &= \dot{z}_1 = z_2 = x_{r+2} \\
\dot{x}_{r+2} &= \dot{z}_2 = z_3 = x_{r+3} \\
&\vdots \\
\dot{x}_n &= \dot{z}_{n-r} = -b_0 z_1 - \cdots - b_{n-r-1} z_{n-r} + x_1 \\
&= -b_0 x_{r+1} - \cdots - b_{n-r-1} x_n + x_1,
\end{aligned}$$

or equivalently,

$$\dot{\eta} = \begin{bmatrix} \dot{x}_{r+1} \\ \dot{x}_{r+2} \\ \vdots \\ \dot{x}_n \end{bmatrix} = \underbrace{\begin{bmatrix} 0 & 1 & 0 & \cdots & 0 \\ 0 & 0 & 1 & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & 1 \\ -b_0 & -b_1 & -b_2 & \cdots & -b_{n-r-1} \end{bmatrix}}_{Q} \eta, \tag{9}$$

since x_1 is zero by design. It is obvious from the companion structure of Q matrix in (9) that the eigenvalues of Q are exactly the zeros of the transfer function $H(s)$.

Example 2.4: Consider the LTI system

$$H(s) = \frac{s-2}{s^2+7s+12}$$

with an initial value $z(0) = [2 \;\; 1]^T$. The LTI system is minimal and has the following state space representation

$$\dot{z} = \underbrace{\begin{bmatrix} -7 & -12 \\ 1 & 0 \end{bmatrix}}_{A} z + \underbrace{\begin{bmatrix} 1 \\ 0 \end{bmatrix}}_{B} u, \quad z(0) = \begin{bmatrix} 2 \\ 1 \end{bmatrix}$$

$$y = \underbrace{\begin{bmatrix} 1 & -2 \end{bmatrix}}_{C} z.$$

It is evident from the transfer function $H(s)$ that the system's relative degree is 1 and the dimension of the realization is 2. Therefore, $\xi = [x_1]$, $\eta = [x_2]$, and

$$\xi = x_1 = Cz = \begin{bmatrix} 1 & -2 \end{bmatrix} z.$$

The normal form is

$$\dot{x}_1 = b(x) + a(x)u$$

$$\dot{x}_2 = q_2(x).$$

In the z-coordinate frame,

$$\hat{b}(z) = L_f h(z) = CAz = \begin{bmatrix} 1 & -2 \end{bmatrix} \begin{bmatrix} -7 & -12 \\ 1 & 0 \end{bmatrix} z = -9z_1 - 12z_2$$

$$\hat{a}(z) = L_g h(z) = CB = \begin{bmatrix} 1 & -2 \end{bmatrix} \begin{bmatrix} 1 \\ 0 \end{bmatrix} = 1.$$

Now, compute the required coordinate transformation

$$\Phi(z) = \begin{bmatrix} h(z) \\ \Phi_2(z) \end{bmatrix} = \underbrace{\begin{bmatrix} 1 & -2 \\ 0 & 1 \end{bmatrix}}_{T} z.$$

Select $\Phi_2(z)$ so that $\frac{\delta\Phi}{\delta z}$ is nonsingular. $x = \Phi(z) = Tz$ gives

Applying the coordinate transformation,

$$b(x) = \hat{b} \circ \Phi^{-1}(x) = CAT^{-1}x = \begin{bmatrix} 1 & -2 \end{bmatrix} \begin{bmatrix} -7 & -12 \\ 1 & 0 \end{bmatrix} \left(\begin{bmatrix} 1 & -2 \\ 0 & 1 \end{bmatrix} \right)^{-1} x = -9x_1 - 30x_2$$

$$a(x) = \hat{a} \circ \Phi^{-1}(x) = CB = \begin{bmatrix} 1 & -2 \end{bmatrix} \begin{bmatrix} 1 \\ 0 \end{bmatrix} = 1$$

$$q_2(x) = L_f\Phi_2 \circ \Phi^{-1}(x) = \begin{bmatrix} 0 & 1 \end{bmatrix} \begin{bmatrix} -7 & -12 \\ 1 & 0 \end{bmatrix} \left(\begin{bmatrix} 1 & -2 \\ 0 & 1 \end{bmatrix} \right)^{-1} x = x_1 + 2x_2.$$

Therefore, the LTI system has the normal form

$$\dot{x}_1 = \begin{bmatrix} -9 & -30 \\ 1 & 2 \end{bmatrix} x + \begin{bmatrix} 1 \\ 0 \end{bmatrix} u$$

$$y = x_1,$$

or equivalently,

$$\begin{bmatrix} \dot{\xi} \\ \dot{\eta} \end{bmatrix} = \begin{bmatrix} -9 & -30 \\ 1 & 2 \end{bmatrix} \begin{bmatrix} \xi \\ \eta \end{bmatrix} + \begin{bmatrix} 1 \\ 0 \end{bmatrix} u$$

$$y = \xi.$$

Comparing the realization above with (7), one finds that $R = -9$, $S = -30$, $P = 1$, $Q = 2$, and $K = 1$. To ensure that $\xi(t) = 0$ for $t > 0$, it is necessary to apply the input

$$\begin{aligned} u^* &= \frac{-S\eta}{K} \\ &= \frac{30}{1} = 30. \end{aligned}$$

so that $y^{(r)} = 0$. Therefore, the dynamics of the state $\eta(t)$ become

$$\dot{\eta} = 2\eta, \quad \eta(0) = 2,$$

where 2 is the zero of the transfer function $H(s)$.

2.3.2 Zero dynamics of control-affine nonlinear systems

For a nonlinear system, there are no simple notion of a zero as in the linear time-invariant case. However, one can develop an analogous concept for a control-affine nonlinear system of zero dynamics [26]. Consider the following single-input, single-output control-affine nonlinear system with relative degree r strictly less than its state dimension n, namely,

$$\begin{aligned} \dot{z} &= f(z) + g(z)u, \quad z(0) = z_0 \\ y &= h(z). \end{aligned} \tag{10}$$

The nonlinear system (10) can be represented in the normal form at some points of interest z_0, which is often an equilibrium point. For the existence of the normal form of a nonlinear system, there must exist a nonlinear coordinate transformation

$$\Phi(z) = \begin{bmatrix} \phi_1(z) \\ \vdots \\ \phi_r(z) \\ \phi_{r+1}(z) \\ \vdots \\ \phi_n(z) \end{bmatrix}, \tag{11}$$

where the first r coordinates are set to be

$$\begin{aligned} \phi_1(z) &= h(z) \\ \phi_2(z) &= L_f h(z) \\ &\vdots \\ \phi_r(z) &= L_f^{r-1} h(z) \end{aligned}$$

and the remaining coordinates $\phi_{r+1}, \ldots, \phi_n$ are selected such that $L_g\phi_{r+1} = \cdots = L_g\phi_n = 0$ and the jacobian matrix of the transformation

$$\frac{\partial \Phi}{\partial z} = \begin{bmatrix} \frac{\partial \phi_1(z)}{\partial z_1} & \cdots & \frac{\partial \phi_1(z)}{\partial z_n} \\ \vdots & \ddots & \vdots \\ \frac{\partial \phi_n(z)}{\partial z_1} & \cdots & \frac{\partial \phi_n(z)}{\partial z_n} \end{bmatrix}$$

is nonsingular on a neighborhood of z_0. To write the normal form of (10) in a compact form, a notation similar to that for an LTI system is used. That is, $\xi = [x_1, \cdots, x_r]^T$ and $\eta = [x_{r+1} \cdots, x_n]^T$ denote two group of states. Therefore,

$$\begin{aligned}
\dot{x}_1 &= x_2 \\
\dot{x}_2 &= x_3 \\
&\vdots \\
\dot{x}_{r-1} &= x_r \\
\dot{x}_r &= b(x_1, \ldots x_r, x_{r+1}, \ldots, x_n) + a(x_1, \ldots x_r, x_{r+1}, \ldots, x_n)u \\
&= b(\xi, \eta) + a(\xi, \eta)u \\
\dot{\eta} &= q(x_1, \ldots x_r, x_{r+1}, \ldots, x_n) \\
&= q(\xi, \eta).
\end{aligned}$$

It is assumed that the output $y(t)$ is zero at $t = 0$, that is, $h(z_0) = 0$. Because z_0 is an equilibrium point $f(z_0) = 0$. As in the LTI case, the relative degree r implies

$$\begin{aligned}
x_1(0) &= h(z_0) = 0 \\
x_2(0) &= L_f h(z_0) = 0 \\
&\vdots \\
x_r(0) &= L_f^{r-1} h(z_0) = 0.
\end{aligned}$$

The values of the remaining $n-r$ components of the new coordinates can be chosen arbitrarily z_0. In particular, one can chose them so that they are zero. If the output is to be identically the zero function, then clearly it is necessary that

$$y(0) = y^{(1)}(0) = \cdots = y^{(r-1)}(0) = 0,$$

or equivalently, $\xi(0) = 0$. To ensure that $\xi(t) = 0$ for $t > 0$, it is necessary to apply the input

$$u^* = \frac{b(0, \eta(t))}{a(0, \eta(t))}$$

so that $y^{(r)} = 0$. From the relative degree assumption, $a(0, \eta(t)) \neq 0$ at least for some finite interval of time. Since $\xi(t)$ is identically zero on this interval, the internal dynamics of the system are governed by the differential equation

$$\dot{\eta}(t) = q(0, \eta(t)), \quad \eta(0) = \eta_0.$$

Example 2.5: Consider the system

$$\begin{aligned} \dot{z_1} &= z_3 \\ \dot{z_2} &= (z_1 - z_2)^3 \\ \dot{z_3} &= z_3^2 + u \\ y &= z_1 \end{aligned}$$

on a neighborhood of origin. This is a nonlinear control-affine system with

$$f(z) = \begin{bmatrix} z_3 \\ (z_1 - z_2)^3 \\ z_3^2 \end{bmatrix},$$

$$g(z) = \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix},$$

$$h(z) = z_1,$$
$$z(0) = \begin{bmatrix} 0 & 0 & 0 \end{bmatrix}^T.$$

The relative degree at $z(0) = 0$ is computed as follows:

$$L_g h(z) = \begin{bmatrix} 1 & 0 & 0 \end{bmatrix} \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix} = 0 \quad \forall z \in \mathbb{R}^3 \quad \rightarrow \quad r > 1,$$
$$L_f h(z) = \begin{bmatrix} 1 & 0 & 0 \end{bmatrix} \begin{bmatrix} z_3 \\ (z_1 - z_2)^3 \\ z_3^2 \end{bmatrix} = z_3$$
$$L_g L_f h(z) = \begin{bmatrix} 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix} = 1 \quad \rightarrow \quad r = 2.$$

The relative degree about the origin is $r = 2$, and the dimension of the system is $n = 3$. The output $y(0) = 0$ at the origin. The vector $\xi = [x_1 \quad x_2]^T$ and $\eta = [x_3]$. Therefore, the normal form has the following structure

$$\dot{x}_1 = x_2$$
$$\dot{x}_2 = b(x) + a(x)u$$
$$\dot{x}_3 = q_3(x).$$

In the z-coordinate frame,

$$\hat{b}(z) = L_f^2 h(z) = \begin{bmatrix} 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} z_3 \\ (z_1 - z_2)^3 \\ z_3^2 \end{bmatrix} z = z_3^2$$

$$\hat{a}(z) = L_g L_f h(z) = \begin{bmatrix} 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix} = 1.$$

Now compute the coordinate transformation:

$$\Phi(z) = \begin{bmatrix} h(z) \\ L_f h(z) \\ \Phi_3(z) \end{bmatrix} = \begin{bmatrix} z_1 \\ z_3 \\ \Phi_3(z) \end{bmatrix},$$

where

$$\frac{\partial \Phi}{\partial z} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 0 & 1 \\ \frac{\partial \Phi_3}{\partial z_1} & \frac{\partial \Phi_3}{\partial z_2} & \frac{\partial \Phi_3}{\partial z_3} \end{bmatrix}.$$

Setting $\Phi_3(z) = z_2$ makes $\det(\frac{\partial \Phi}{\partial z})|_{z(0)} \neq 0$ and $L_g \Phi_3(z) = 0$. Therefore,

$$x = \Phi(z) = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 0 & 1 \\ 0 & 1 & 0 \end{bmatrix} z.$$

That is,

$$x_1 = z_1$$

$$x_2 = z_3$$

$$x_3 = z_2.$$

Applying the coordinate transformation,

$$b(x) = \hat{b} \circ \Phi^{-1}(x) = x_2^2$$

$$a(x) = \hat{a} \circ \Phi^{-1}(x) = 1$$

$$q_3(x) = L_f\Phi_3 \circ \Phi^{-1}(x) = \begin{bmatrix} 0 & 1 & 0 \end{bmatrix} \begin{bmatrix} z_3 \\ (z_1 - z_2)^3 \\ z_3^2 \end{bmatrix}_{z=\Phi^{-1}(z)}$$

$$= (x_1 - x_3)^3.$$

Therefore, the system has the following normal form

$$\dot{x}_1 = x_2$$

$$\dot{x}_2 = x_2^2 + u$$

$$\dot{x}_3 = (x_1 - x_3)^3$$

$$y = x_1.$$

Setting $\xi = 0$ gives the zero dynamics

$$\dot{x}_3 = -x_3^3, \quad x_3(0) = 0.$$

That is, the system has the trivial zero dynamics $\eta(t) = 0$, $t \geq 0$.

Example 2.6: Consider the bilinear system

$$\dot{z} = \begin{bmatrix} -1 & 1 & -1 \\ 3 & 0 & 6 \\ 1 & 0 & 2 \end{bmatrix} z + \begin{bmatrix} 1 & 0 & 1 \\ 0 & 1 & 1 \\ 1 & 0 & 2 \end{bmatrix} zu, \quad z(0) = \begin{bmatrix} 2 \\ 1 \\ -1 \end{bmatrix}$$

$$y = \begin{bmatrix} 0 & 1 & 1 \end{bmatrix} z.$$

Comparing the above bilinear system with the general structure of a control-affine nonlinear system (10), it follows that

$$f(z) = \begin{bmatrix} -1 & 1 & -1 \\ 3 & 0 & 6 \\ 1 & 0 & 2 \end{bmatrix} z$$

$$g(z) = \begin{bmatrix} 1 & 0 & 1 \\ 0 & 1 & 1 \\ 1 & 0 & 2 \end{bmatrix} z$$
$$h(z) = \begin{bmatrix} 0 & 1 & 1 \end{bmatrix} z.$$

The relative degree r at $z(0)$ is calculated as follows:

$$L_g h(z) = \begin{bmatrix} 0 & 1 & 1 \end{bmatrix} \begin{bmatrix} 1 & 0 & 1 \\ 0 & 1 & 1 \\ 1 & 0 & 2 \end{bmatrix} \begin{bmatrix} 2 \\ 1 \\ -1 \end{bmatrix} = 0 \quad \forall z \in \mathbb{R}^3 \quad \rightarrow \quad r > 1,$$
$$L_f h(z) = \begin{bmatrix} 0 & 1 & 1 \end{bmatrix} \begin{bmatrix} -1 & 1 & -1 \\ 3 & 0 & 6 \\ 1 & 0 & 2 \end{bmatrix} = \begin{bmatrix} 4 & 0 & 8 \end{bmatrix} z,$$
$$L_f h(z(0)) = \begin{bmatrix} 4 & 0 & 8 \end{bmatrix} \begin{bmatrix} 2 \\ 1 \\ -1 \end{bmatrix} = 0,$$
$$L_g L_f h(z(0)) = \begin{bmatrix} 4 & 0 & 8 \end{bmatrix} \begin{bmatrix} 1 & & 1 \\ 0 & 1 & 1 \\ 1 & 0 & 2 \end{bmatrix} z(0) = \begin{bmatrix} 12 & 0 & 20 \end{bmatrix} \begin{bmatrix} 2 \\ 1 \\ -1 \end{bmatrix} = 4 \neq 0 \quad \rightarrow \quad r = 2.$$

Note that $h(z_0) = 0$, so the zero function is in the range of the input-output map. The normal form has the structure,

$$\dot{x}_1 = x_2$$
$$\dot{x}_2 = b(x) + a(x)u$$
$$\dot{x}_3 = q_3(x).$$

In the z-coordinate frame,

$$\hat{b}(z) = L_f^2 h(z) = \begin{bmatrix} 4 & 0 & 8 \end{bmatrix} \begin{bmatrix} -1 & 0 & -1 \\ 3 & 0 & 6 \\ 1 & 0 & 2 \end{bmatrix} z = \begin{bmatrix} 4 & 4 & 12 \end{bmatrix} z$$

$$\hat{a}(z) = L_g L_f h(z) = \begin{bmatrix} 4 & 0 & 8 \end{bmatrix} \begin{bmatrix} 1 & 0 & 1 \\ 0 & 1 & 1 \\ 1 & 0 & 2 \end{bmatrix} z = \begin{bmatrix} 12 & 0 & 20 \end{bmatrix} z$$

The required coordinate transformation is

$$\Phi(z) = \begin{bmatrix} h(z) \\ L_f h(z) \\ \Phi_3(z) \end{bmatrix} = \begin{bmatrix} 0 & 1 & 1 \\ 4 & 0 & 8 \\ 0 & 0 & 1 \end{bmatrix} z$$

where

$$\frac{\partial \Phi}{\partial z} = \begin{bmatrix} 0 & 1 & 1 \\ 4 & 0 & 8 \\ \frac{\partial \Phi_3}{\partial z_1} & \frac{\partial \Phi_3}{\partial z_2} & \frac{\partial \Phi_3}{\partial z_3} \end{bmatrix}.$$

Setting $\Phi_3(z) = z_2$ makes $\det(\frac{\partial \Phi}{\partial z})|_{z(0)} \neq 0$ and $L_g \Phi_3(z) = 0$. Therefore,

$$x = \Phi(z) = \begin{bmatrix} 0 & 1 & 1 \\ 4 & 0 & 8 \\ 0 & 0 & 1 \end{bmatrix} z.$$

That is,

$$z_3 = x_3$$

$$x_1 = z_2 + z_3 \quad \Rightarrow \quad z_2 = x_1 - x_3$$

$$x_2 = 4z_1 + 8z_3 \quad \Rightarrow \quad z_1 = \frac{1}{4}x_2 - 2x_3$$

and

$$z = \Phi^{-1}(x) = \begin{bmatrix} 0 & \frac{1}{4} & -2 \\ 1 & 0 & -1 \\ 0 & 0 & 1 \end{bmatrix} x.$$

Applying the coordinate transformation gives,

$$b(x) = \hat{b} \circ \Phi^{-1}(x) = \begin{bmatrix} 4 & 4 & 12 \end{bmatrix} \begin{bmatrix} 0 & \frac{1}{4} & -2 \\ 1 & 0 & -1 \\ 0 & 0 & 1 \end{bmatrix} x = 4x_1 + x_2$$

$$a(x) = \hat{a} \circ \Phi^{-1}(x) = \begin{bmatrix} 12 & 0 & 20 \end{bmatrix} \begin{bmatrix} 0 & \frac{1}{4} & -2 \\ 1 & 0 & -1 \\ 0 & 0 & 1 \end{bmatrix} x = 3x_2 - 4x_3$$

$$\begin{aligned} q_3(x) = L_f\Phi_3 \circ \Phi^{-1}(x) &= \begin{bmatrix} 0 & 1 & 0 \end{bmatrix} \begin{bmatrix} -1 & 0 & -1 \\ 3 & 0 & 6 \\ 1 & 0 & 2 \end{bmatrix} \begin{bmatrix} 0 & \frac{1}{4} & -2 \\ 1 & 0 & -1 \\ 0 & 0 & 1 \end{bmatrix} x \\ &= \frac{1}{4}x_2. \end{aligned}$$

Therefore, the system has the following normal form

$$\begin{aligned} \dot{x}_1 &= x_2 \\ \dot{x}_2 &= (4x_1 + x_2) + (3x_2 - 4x_3)u \\ \dot{x}_3 &= \frac{1}{4}x_2 \\ y &= x_1. \end{aligned}$$

To ensure that $\xi(t) = 0$ for $t > 0$, it is necessary to apply the input

$$u^* = -\frac{4x_1 + x_2}{3x_2 - 4x_3}.$$

so that $y^{(r)} = 0$. Therefore, the dynamics of the state $\eta(t)$ become

$$\dot{x_3} = \frac{1}{4}x_2, \quad x(0) = \begin{bmatrix} 0 \\ 0 \\ -1 \end{bmatrix}.$$

2.4 NON-MINIMUM PHASE SYSTEM

First the non-minimum phase property of an LTI system is defined and then the definition is extended to a control-affine nonlinear system.

2.4.1 Non-minimum phase LTI system

A linear time-invariant system is said to be minimum phase if the system and its inverse are both causal and stable. Otherwise, the system is non-minimum phase. In this dissertation, the definition is narrowed so that an LTI system is non-minimum phase if it has at least one zero in the right-half complex plane. Such zeros affect the system's transient response and can lead to undesirable behaviors such as overshoot and oscillations in response to a step input.

Example 2.7: Consider the simple transfer function

$$H(s) = \frac{s-2}{s+1}.$$

The system has a zero at $s = 2$ (right-half plane) and pole at $s = -1$ (left-half plane). Therefore, the system is a non-minimum phase system. The presence of a right half-plane zero typically causes an initial response in the opposite direction of the steady-state response when a step input $u(t)$ is applied. This phenomenon is known as an undershoot. Observe

that the unit step response is

$$y(t) = (1 - 3e^{-t})u(t).$$

At $t = 0$, $y(0) = -2$, the response initially drops to -2, which is in the opposite direction of the steady-state value of 1. This behavior is typical of non-minimum phase systems, where the initial response can be counterintuitive and lead to a transient undershoot or overshoot.

2.4.2 Non-minimum phase nonlinear system

A control-affine nonlinear system is said to be *minimum phase* if its zero dynamics are stable in some sense [26]. If they are stable on some open subset of the zero dynamics manifold, then the system is called *locally minimum phase*. If they are stable everywhere on this manifold, then the system is *globally minimum phase*. Standard stability analysis methods like Lyapunov's first and second method can be used to make this determination.

Example 2.8: Consider the bilinear system

$$\dot{z} = f(z) + g(z)u = \underbrace{\begin{bmatrix} -6 & 5 \\ 1 & -1 \end{bmatrix}}_{A_0} z + \underbrace{\begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix}}_{A_1} zu$$

$$y = h(z) = \underbrace{\begin{bmatrix} 1 & -2 \end{bmatrix}}_{C}$$

at the equilibrium $z_e = [1 \quad 1]^T$, $u_e = 1$.

Linearizing the system around this point yields,

$$\begin{aligned} A &= \frac{\partial(f(z) + g(z)u)}{\partial z}\Big|_{(z_e, u_e)} \\ &= (A_0 + A_1 u)|_{(z_e, u_e)} \end{aligned}$$

$$= \begin{bmatrix} -5 & 5 \\ 1 & -1 \end{bmatrix}$$

$$B = \frac{\partial(f(z) + g(z)u)}{\partial u}|_{(z_e,u_e)}$$

$$= A_1 z|_{(z_e,u_e)}$$

$$= \begin{bmatrix} 1 \\ 0 \end{bmatrix}.$$

The linearized system around the point (z_e, u_e) is therefore

$$\dot{z} = Az + Bu = \begin{bmatrix} -5 & 5 \\ 1 & -1 \end{bmatrix} z + \begin{bmatrix} 1 \\ 0 \end{bmatrix} u$$

$$y = Cz = \begin{bmatrix} 1 & -2 \end{bmatrix} z.$$

The corresponding transfer function of the linearized system is

$$H(s) = \frac{s-1}{s^2 + 6s}.$$

The linearized system has a zero at $s = 1$, and therefore the bilinear system is not locally minimum phase about the given equilibrium point.

2.5 BILINEAR MODELLING OF A PETRO-CHEMICAL PLANT

The goal of this section is to describe the feed flow system in [45], which is used to model a petro-chemical processing plant. The system as shown in Figure 2 consists of two tanks of fluid, each with a concentration $z_i \geq 0$, $i = 1, 2$ of a dissolved solid measured in kg/m^3. The concentrations are controlled by two valves $u_1 \leq 0$ and $u_2 \leq 0$. Valve u_1 is used to transfer fluid between the two tanks, while valve u_2 is used as a drain valve for tank 1. Parameter $\alpha > 0$ represents an assumed constant exogenous fluid transfer rate into tank

2. The parameters and variables for the system are summarized in Table 1 and Table 2, respectively.

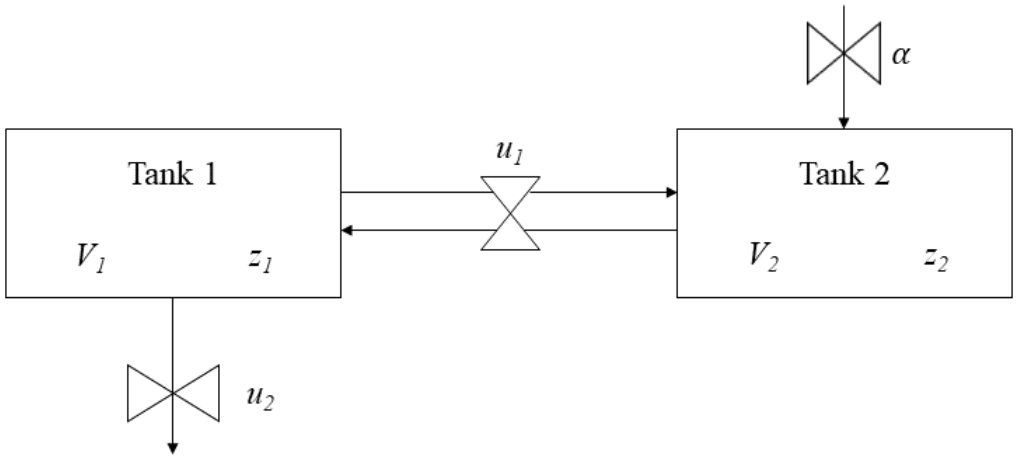

Figure 2: Two tank feed flow system

Under normal operating conditions, the valves are used simultaneously to maintain prescribed concentrations in the tanks. The larger the difference in concentrations, the faster the rate of change in a tank's concentration. Based on the conservation of matter principle, the feed flow system has the dynamics

$$V_1\dot{z}_1 = u_1(z_1 - z_2) + z_1u_2, \;\; z_1(0) = z_{10}$$

$$V_2\dot{z}_2 = -u_1(z_1 - z_2) + \alpha z_2, \;\; z_2(0) = z_{20}.$$

These dynamics with an output y can be rewritten as the bilinear system

$$\dot{z} = N_0z + N_1zu_1 + N_2zu_2, \;\; z(0) = z_0 \tag{12a}$$

$$y = Cz, \tag{12b}$$

where

$$N_0 = \begin{bmatrix} 0 & 0 \\ 0 & \frac{\alpha}{V_2} \end{bmatrix}, \; N_1 = \begin{bmatrix} \frac{1}{V_1} & -\frac{1}{V_1} \\ -\frac{1}{V_2} & \frac{1}{V_2} \end{bmatrix}, \; N_2 = \begin{bmatrix} \frac{1}{V_1} & 0 \\ 0 & 0 \end{bmatrix},$$

TABLE 1: Parameters for the feed flow system

Parameter	**Value**	**Description**
V_1	100	volume of tank 1 (m^3)
V_2	200	volume of tank 2 (m^3)
α	6	constant exogenous fluid transfer rate into tank 2 (m^3/s)
C	$[1\ \ 0]$	output vector
$\{z_{1e}, z_{2e}\}$	$\{4, 10\}$	equilibrium states (kg/m^3)
$\{u_{1e}, u_{2e}\}$	$\{-10, -15\}$	equilibrium inputs (m^3/s)

$$C = \begin{bmatrix} C_1 & C_2 \end{bmatrix}, \quad z_0 = \begin{bmatrix} z_{10} \\ z_{20} \end{bmatrix}.$$

An equilibrium point (z_e, u_e) of the system must satisfy the following condition:

$$U_e z_e = 0, \tag{13}$$

where

$$U_e = \begin{bmatrix} u_{1e} + u_{2e} & -u_{1e} \\ -u_{1e} & u_{1e} + \alpha \end{bmatrix}, \quad z_e = \begin{bmatrix} z_{1e} \\ z_{2e} \end{bmatrix}.$$

Condition (13) can be rewritten as

$$Z_e u_e + W = 0,$$

TABLE 2: Variables for the feed flow system

Variable	**Range**	**Description**
z_i	$\{0, \mathbb{R}_+\}$	concentration of dissolved solid in i-th tank (kg/m^3)
u_1	$\{\mathbb{R}_-, 0\}$	fluid transfer rate between tank 1 and 2 (m^3/s)
u_2	$\{\mathbb{R}_-, 0\}$	fluid transfer rate out of tank 1 (m^3/s)

where

$$Z_e = \begin{bmatrix} z_{1e} - z_{2e} & z_{1e} \\ -(z_{1e} - z_{2e}) & 0 \end{bmatrix}, \; W = \begin{bmatrix} 0 \\ \alpha z_{2e} \end{bmatrix}, \; u_e = \begin{bmatrix} u_{1e} \\ u_{2e} \end{bmatrix}.$$

2.6 LINEARIZED PETRO-CHEMICAL PLANT

Linearizing (12a)-(12b) around an equilibrium point (z_e, u_e) renders the linear time-invariant realization

$$A = [N_0 + N_1 u_{1e} + N_2 u_{2e}] = V^{-1} U_e, \tag{14a}$$

$$B = [N_1 z_e \;\; N_2 z_e] = V^{-1} Z_e, \tag{14b}$$

$$C = [C_1 \;\; C_2] \tag{14c}$$

with $V = \text{diag}(\text{V}_1, \text{V}_2)$. The corresponding transfer functions are

$$H_1(s) = K_1 \frac{s + b_1}{s^2 + a_1 s + a_2} \tag{15}$$

$$H_2(s) = K_2 \frac{s + b_2}{s^2 + a_1 s + a_2}, \tag{16}$$

TABLE 3: Coefficients of the transfer functions of the linearized feed flow system

Parameter	**Formula**
K_1	$\frac{(C_1V_2-C_2V_1)}{V_1V_2}$
K_2	$\frac{C_1}{V_1}$
b_1	$\frac{(C_2u_{2e}-\alpha C_1)}{(C_1V_2-C_2V_1)}$
b_2	$-\frac{1}{C_1V_2}(\alpha C_1 + C_1u_{1e} + C_2u_{1e})$
a_1	$-\frac{1}{V_1V_2}(V_1(\alpha + u_{1e}) + V_2(u_{1e} + u_{2e}))$
a_2	$\frac{1}{V_1V_2}(\alpha(u_{1e} + u_{2e}) + u_{1e}u_{2e})$

where, in general the K_i's are nonzero. The transfer functions are non-minimum phase if and only if $b_i < 0$, $i = 1, 2$, specifically, if

$$u_{2e} < \frac{\alpha C_1}{C_2}, \quad C_1V_2 - C_2V_1 > 0$$

or

$$u_{2e} > \frac{\alpha C_1}{C_2}, \quad C_1V_2 - C_2V_1 < 0$$

for $H_1(s)$, and

$$u_{1e} > -\frac{\alpha C_1}{C_1 + C_2} = d_1, \quad C_1V_2 > 0$$

or

$$u_{1e} < \frac{\alpha C_1}{C_1 + C_2} = d_1, \quad C_1V_2 < 0$$

for $H_2(s)$. The pole locations are in left-half plane if and only if $a_i > 0$, $i = 1, 2$. If $z_e \neq 0$, then from (13) it follows that

$$z_{2e} = \frac{u_{1e}}{\alpha + u_{1e}} z_{1e} \tag{17}$$

$$\det(U_e) = \alpha(u_{1e} + u_{2e}) + u_{1e}u_{2e} = 0. \tag{18}$$

It is evident from Table 4 and (18) that $a_2 = 0$. Therefore, each H_i has a pole at zero. Since $\alpha > 0$ and $u_{1e} < 0$, it is obvious from (17) that

$$\alpha + u_{1e} < 0. \tag{19}$$

This ensures that $a_1 > 0$ so that each H_i has its remaining pole in the strict left-half plane.

The system is controllable by each input acting individually if the controllability matrices

$$\mathcal{C}_1 = -\frac{(z_{1e} - z_{2e})^2}{V_1 V_2} \begin{bmatrix} 1 & \frac{u_{1e}-u_{2e}}{V_1} + \frac{u_{1e}}{V_2} \\ 1 & \frac{-u_{1e}+\alpha}{V_1} + \frac{u_{1e}}{V_2} \end{bmatrix} \tag{20}$$

and

$$\mathcal{C}_2 = -\frac{z_{1e}^2}{V_1^2} \begin{bmatrix} 1 & \frac{u_{1e}-u_{2e}}{V_1} \\ 0 & \frac{-u_{1e}+\alpha}{V_2} \end{bmatrix} \tag{21}$$

have full rank. Input u_1 loses controllability when $z_{1e} = z_{2e}$. Input u_2 loses controllability when $z_{1e} = 0$ and/or $u_{1e} = \alpha$. A dual analysis can be made using the system's observability matrix

$$\mathcal{O} = \begin{bmatrix} C_1 & C_2 \\ C_1 \frac{u_{1e}-u_{2e}}{V_1} + C_2 \frac{-u_{1e}+\alpha}{V_2} & -\frac{C_1 u_{1e}}{V_1} + \frac{C_2 u_{1e}}{V_2} \end{bmatrix}.$$

The system loses it's observability for some specific combinations of parameters and inputs. For example, if $C_1 = 0$ and $u_{1e} = \alpha$ the observability matrix $\mathcal{O}$ loses the full rank.

The relative degree of each transfer function $H_i(s)$ is $r_i = 1$. To determine when the bilinear system (53) also has this relative degree, observe that

$$\dot{y} = C\dot{z} = CN_0 z + CN_1 z u_1 + CN_2 z u_2.$$

Therefore, (53) has relative degree 1 at z in u_1 if

$$CN_1z = \left(\frac{C_1}{V_1} - \frac{C_2}{V_2}\right)(z_1 - z_2) \neq 0$$

$$\Leftrightarrow \qquad \frac{C_1}{V_1} \neq \frac{C_2}{V_2}, \;\; z_1 \neq z_2,$$

and relative degree 1 at z in u_2 if

$$CN_2z = \frac{C_1z_1}{V_1} \neq 0$$

$$\Leftrightarrow \qquad C_1 \neq 0, \;\; z_1 \neq 0.$$

If the parameters are chosen as shown in Table 1, then the bilinear system is

$$\dot{z} = \begin{bmatrix} 0 & 0 \\ 0 & \frac{6}{200} \end{bmatrix} z + \begin{bmatrix} \frac{1}{100} & -\frac{1}{100} \\ -\frac{1}{200} & \frac{1}{200} \end{bmatrix} zu_1 + \begin{bmatrix} \frac{1}{100} & 0 \\ 0 & 0 \end{bmatrix} zu_2, \qquad z_0 = \begin{bmatrix} z_{1e} \\ z_{2e} \end{bmatrix} \tag{22a}$$

$$y = \begin{bmatrix} 1 & 0 \end{bmatrix} z \tag{22b}$$

and has relative degree 1 in both inputs. When the system is linearized around the equilibrium given in Table 1, the corresponding transfer functions are

$$H_1(s) = \frac{-0.06(s - 0.03)}{s(s + 0.27)}, \;\; H_2(s) = \frac{0.04(s + 0.02)}{s(s + 0.27)}.$$

Both transfer functions are marginally stable, irreducible, and have relative degree 1. $H_1(s)$ is non-minimum phase system, while $H_2(s)$ is a minimum phase. Therefore, the mapping $u_1 \mapsto y$ for the bilinear system will be locally non-minimum phase at this equilibrium. This will be the linearized system of interest for the remainder of the dissertation.

CHAPTER 3

ZERO DYNAMICS ATTACK DESIGN

This chapter aims to develop methods for designing zero dynamic attack signals. Designing a zero dynamics attack utilizes the zero dynamics property of the targeted system. The properties of a bilinear system that make it susceptible to zero dynamics attacks are presented. The bilinear system identification algorithm adopted to synthesize an attack signal is described. The bilinear systems considered here are continuous-time systems. However, the system identification algorithm requires discrete-time input-output data and thus identifies a discrete-time bilinear system. Therefore, a rule for discretizing a continuous-time bilinear system and the relationship between a discrete-time and continuous-time bilinear realization is established. This chapter ends by presenting two attack design methods, namely, an observer-based method and an analytical method.

3.1 ZERO DYNAMICS ATTACK VULNERABILITY

The design of a zero dynamics attack signal has two steps: the system identification step and the attack signal synthesis step. For the system identification step, it is assumed that an adversary has the full access to the input and output of the target system without noise. To initiate the identification process, the adversary need to perform an input-output experiment that is benign enough to avoid detection. For example, a sequences of constant input pulses with free-decay in between is injected to identify a continuous-time bilinear system using the single experiment multiple pulses (SEMP) method [30]. Alternatively, a white uniformly distributed pseudo-random signal could be applied to identify the generating series of the system using Chen-Fliess series method [18]. After identifying the system, the next step is

to design a zero dynamics attack signal u^*. A SISO bilinear system is susceptible to a zero dynamics attack if:

i) it has a dimension $n > 1$;

ii) it has a relative degree $1 \leq r < n$;

iii) it has non-minimum phase zero dynamics.

3.2 BILINEAR SYSTEM IDENTIFICATION

3.2.1 Discretized Bilinear System

Consider the following SISO bilinear system with n states,

$$\dot{z} = N_0 z + N_1 z u_1 \tag{23a}$$

$$y = Cz, \tag{23b}$$

where N_0 is a Hurwitz matrix. If $u_1(t) = v_1 \in \mathbb{R}$ over $k\Delta t \leq t < (k+1)\Delta t$, where Δt is the sample time, then (23a) is equivalent to

$$z(k+1) = M_1 z(k),$$

where $0 \leq k \leq l_0$ and

$$M_1 = \exp\left(N_0 + N_1 v_1\right)\Delta t. \tag{24}$$

If $v_1 = 0$ then

$$z(k+1) = M_0 z(k),$$

where

$$M_0 = \exp(N_0 \Delta t). \tag{25}$$

3.2.2 Identification Algorithm

The identification algorithm used here is adapted from [29]. This algorithm by design will estimate a minimal realization of the input-output system up to a coordinate transformation. To identify (23), the pulse sequence shown in Figure 3 is first applied to the input channel. The first pulse starts at the time index $k \in \mathbb{N}$. The delay $l_0 \in \mathbb{N}$ is defined as the number

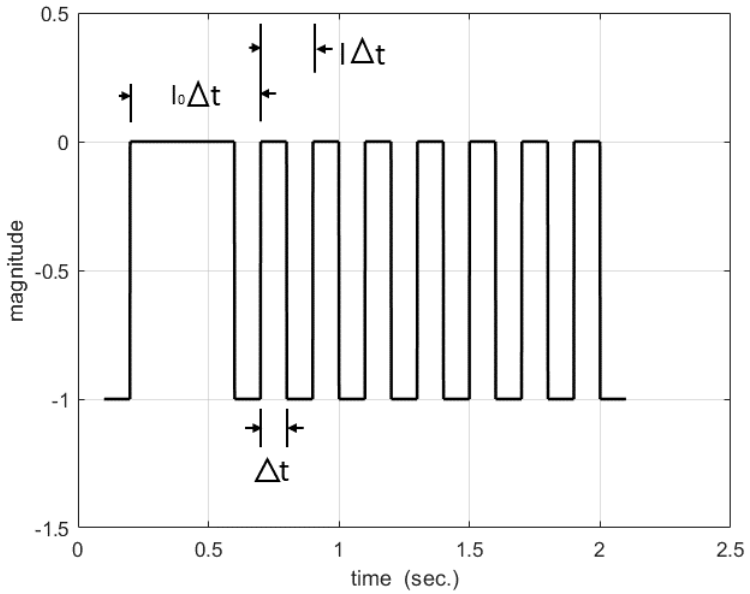

Figure 3: Input for the identification algorithm

of sampling times, Δt, between the first and the second pulse's rising edge. Therefore, the first pulse is of width $l_0 \Delta t$. The delay l_0 allows one to collect data to identify the matrix M_0. From the second pulse onward, the delay between any two consecutive rising edges is $l\Delta t$, where $l \in \mathbb{N}$. The steps of the identification algorithm are briefly described below.

Step 1. From the output samples starting at time $k+1$, build a $\beta \times \gamma$ Hankel matrix

$$H_{k+1} = \begin{bmatrix} y(k+1) & y(k+2) & \cdots & y(k+\gamma) \\ y(k+2) & y(k+3) & \cdots & y(k+\beta+1) \\ \vdots & \vdots & \ddots & \vdots \\ y(k+\beta-1) & y(k+\beta) & \cdots & y(k+\beta+\gamma-1) \end{bmatrix}.$$

The size of the Hankel matrix is assumed to be large enough so that its rank is greater than or equal to the minimal dimension of the unknown system [39, 40]. The numerical rank n of H_{k+1} will be the dimension of the identified system.

Step 2. Compute the singular value decomposition of H_{k+1} to determine an observability matrix $U_{k+1} \in \mathbb{R}^{\beta\times\beta}$ and a controllability-like matrix $\sum_{k+1} V_{k+1}^T \in \mathbb{R}^{\beta\times\gamma}$:

$$\begin{aligned} H_{k+1} &= U_{k+1} \sum_{k+1} V_{k+1}^T \\ &= \begin{bmatrix} \hat{C} \\ \hat{C}\hat{A} \\ \vdots \\ \hat{C}\hat{A}^{\beta-1} \end{bmatrix} \begin{bmatrix} \hat{z}(k+1) & \hat{M}_0\hat{z}(k+1) & \cdots & \hat{M}_0^{\gamma-1}\hat{z}(k+1) \end{bmatrix}. \end{aligned}$$

Step 3. An estimate of C is found by taking the first n columns of the first row of the matrix U_{k+1}, namely,

$$\hat{C} := U_{k+1}(1, 1:n).$$

Two matrices $U_{upper}, U_{lower} \in \mathbb{R}^{(\beta-1)\times n}$ with rank n are then formed by taking the first $\beta-1$ rows and first n columns of U_{k+1}, and the second to β-th rows and first n columns of U_{k+1}, respectively. Estimates of the M_0 matrix and state vector z at the $(k+l_0)$-th sample are found as follows:

$$\hat{M}_0 = (U_{upper})^{\dagger} U_{lower}$$

$$\hat{z}(k+l_0) = \hat{M}_0^{l_0-1}\hat{z}(k+1).$$

Here the symbol † represents the pseudo-inverse operation. The vector $\hat{z}(k+1)$ is the first column of controllability-like matrix estimated in the previous step.

Step 4. The measured sampled outputs are next grouped into two different classes with an ordered indexing. The ordered members of the first group are found by the formula

$$y(:, j_i) = (U_{upper})^{\dagger} H_{k+1}(:, j_i),$$

where $j_i = k + l_0 + (i-1)l + 1$, $i = 1, 2, \ldots, p$. The ordered members of the second class are determined by the following formulas

$$\begin{aligned} x(:, j_1) &= \hat{M}_0^{l_0-1}\hat{z}(k+1) \\ x(:, j_i) &= \hat{M}_0^{l-1} y(:, j_{i-1}), \quad i = 2, 3, \ldots, p. \end{aligned}$$

Step 5. Form two matrices using the members of the two classes in the previous step:

$$\begin{aligned} \mathcal{Y} &= \begin{bmatrix} y(j_1) & y(j_2) & \cdots & y(j_p) \end{bmatrix} \\ \mathcal{X} &= \begin{bmatrix} x(j_1) & x(j_2) & \cdots & x(j_p) \end{bmatrix}. \end{aligned}$$

Estimate the coupling matrix M_1 as

$$\hat{M}_1 = \mathcal{Y}\mathcal{X}^{\dagger}.$$

Step 6. Estimate the continuous-time state transition matrix N_0 and coupling matrix N_1 using (25) and (24), respectively,

$$\begin{aligned} \hat{N}_0 &= \frac{1}{\Delta t}\log(\hat{M}_0) \\ \hat{N}_1 &= \frac{1}{v_1}\left[\frac{1}{\Delta t}\log(\hat{M}_1) - \hat{M}_0\right]. \end{aligned}$$

Regarding the final step, the matrix logarithm is generally a multi-valued function and can take on complex values. For example, the logarithm of the rotation matrix

$$A = \begin{bmatrix} \cos(\alpha) & -\sin(\alpha) \\ \sin(\alpha) & \cos(\alpha) \end{bmatrix}$$

is the set of skew-symmetric matrices

$$B_n = (\alpha \pm 2\pi n) \begin{bmatrix} 0 & -1 \\ 1 & 0 \end{bmatrix}, \quad n \in \mathbb{N}.$$

The *necessary* conditions for the existence of only real-valued estimates in (24) and (25) are: 1) the matrices M_0 and M_1 are nonsingular, and 2) the corresponding Jordan block of each negative eigenvalue of each matrix occurs an even number of times [9]. The *unique* real-valued logarithms of matrices M_0 and M_1 exist if all the eigenvalues of M_0 and M_1 are positive real and each Jordan block belonging to any eigenvalue in each matrix does not appear more than once [9]. From the Taylor series point of view, recall the power series representation

$$\log(M_p) = \sum_{k=1}^{\infty} (-1)^{k+1} \frac{(M_p - I)^k}{k}$$

converges if $||M_p - I|| < 1$ for $p = 0, 1$ [14]. This provides a simple *sufficient* condition for the existence of real-valued estimates. Reducing the sample time Δt will often reduce $||M_p - I||$ to ensure convergence, but this can also lead to poor numerical conditioning.

3.3 ATTACK SIGNAL SYNTHESIS

This section will present two different approaches to designing an attack input: the observer-based approach and the analytical approach. The following simplifying assumptions are made regarding the execution in both approaches:

A1) The attack is initiated (including the plant identification step) when the system is in an equilibrium state (z_e, u_e).

A2) The attack is implemented/designed using the identified initial condition, which is assumed to be close to the true initial condition when the attack signal is applied.

A3) Any feedback controller present (for example to stabilize the plant) is disabled during the attack.

A4) The attack is designed assuming no type of measurement or process noise is present.

The first assumption is reasonable because many physical plants, including petro-chemical plants, operate at equilibrium under normal circumstances. The noise-free case is considered in this dissertation primarily because the robustness of the bilinear identification algorithm used throughout appears to be an unresolved issue in the literature. Though an adversary's ability to disable the feedback controller can be deemed as a type of attack, this in general would not be a zero dynamics attack since it would lack any form of stealthiness and would normally be be detected immediately.

Consider the general bilinear system,

$$\begin{aligned} \dot{z} &= N_0 z + \sum_{i=1}^{m} N_i z u_i \\ y &= Cz. \end{aligned} \tag{26}$$

Assume Q_{eq} is the set of all equilibria $(z_e, u_e) \in \mathbb{R}^{(n+m)}$ of the bilinear system (26), and Q_v is the set of equilibria where the bilinear system is a non-minimum phase system, that is, there are zeros in the right-half complex plane if the bilinear system is linearized at any equilibrium. Let $\mathcal{I}$ be the set of m inputs, $\mathcal{I}_c$ denotes the set of constant inputs, and $\mathcal{I}_a$ denotes the set of attack inputs. Assume $\mathcal{I}$ can be partitioned into $\mathcal{I}_c$ and $\mathcal{I}_a$ such that $\mathcal{I}_c \bigcap \mathcal{I}_a = \emptyset$, $\mathcal{I}_c \bigcup \mathcal{I}_a = \mathcal{I}$ with $\mathcal{I}_a \neq \emptyset$. Select any u_p from $\mathcal{I}_a$ and reorganize (26) into the

form

$$\dot{z} = \underbrace{(N_0 + \sum_{u_i \in \mathcal{I}_c} N_i u_i)}_{A_0} z + \underbrace{N_p u_p}_{A_1 u^*} z, \;\; z(0) \in Q_v, \tag{27a}$$

$$= A_0 z + A_1 z u^*, \tag{27b}$$

$$y = Cz. \tag{27c}$$

Assume the SISO bilinear system (27) has relative degree $r < n$ at an equilibrium $z(0) = z_0$. The bilinear system (27) can be represented in the normal form at z_0.

$$\begin{aligned}
\dot{x}_1 &= x_2 \\
\dot{x}_2 &= x_3 \\
&\vdots \\
\dot{x}_{r-1} &= x_r \\
\dot{x}_r &= b(x_1, \cdots x_r, x_{r+1}, \cdots, x_n) + a(x_1, \cdots x_r, x_{r+1}, \cdots, x_n)u \\
&= b(\xi, \eta) + a(\xi, \eta)u \\
\dot{\eta} &= q(x_1, \cdots x_r, x_{r+1}, \cdots, x_n) \\
&= q(\xi, \eta).
\end{aligned}$$

First, two attack signal synthesis approaches, the observer-based approach and the analytical approach, for a general bilinear system are described. Then the attack signal synthesis for the petro-chemical plant (12) using these approaches are presented in the respective subsections. In standard zero dynamics of an LTI system, the output is kept identically at zero, and the zeros of the system play roles in determining input for zeroing the output [26]. However, the objective for the petro-chemical plant is to keep the output identically at a nonzero constant level. The zero dynamics attack signal is designed in the sense that the zeros of the linearized system play roles in determining input to keep output identically constant while

attacking some internal dynamics as well. This implies if relative degree $r > 0$ then $y^{(r)} = 0$.

3.3.1 Observer-based approach

The observer-based approach assumes that state estimates are always available. Consider the general bilinear system (27). The output $y(t)$ is assumed to be zero at $t = 0$, that is, $h(z_0) = Cz_0 = 0$. Because z_0 is an equilibrium point $f(z_0) = 0$. In z coordinates, the relative degree r at z_0 implies

$$\begin{aligned} y^{(k)}(0) &= L_f^k z + L_g L_f^{k-1} z|_{z=z_0}, \qquad for\ \ k = 0, 1, \ldots, r-1 \\ &= (CA_0^k + CA_1 A_0^{k-1}) z|_{z=z_0} \\ &= 0. \end{aligned}$$

The zero dynamics attack input u^* is determined by letting $y^{(r)} = 0$,

$$\begin{aligned} y^r(t) &= L_f^r z + L_g L_f^{r-1} z u^* \\ 0 &= CA_0^r z + CA_1 A_0^{r-1} z u^* \\ u^* &= -\frac{CA_0^r z}{CA_1 A_0^{r-1} z}. \end{aligned} \tag{28}$$

In the x coordinates the attack input is,

$$u^* = -\frac{CA_0^r \Phi^{-1}(x)}{CA_1 A_0^{r-1} \Phi^{-1}(x)}.$$

Setting $\xi(0) = 0$ for $t > 0$, therefore the attack input can be written as

$$u^* = -\frac{CA_0^r \Phi^{-1}(0, \eta)}{CA_1 A_0^{r-1} \Phi^{-1}(0, \eta)}. \tag{29}$$

From the relative degree assumption, $CA_1 A_0^{r-1} \Phi^{-1}(0, \eta) \neq 0$. Since $\xi(t)$ is identically zero for all t, the internal dynamics of the system are governed by the differential equation

$$\dot{\eta}(t) = q(0, \eta(t)), \quad \eta(0) = \eta_0.$$

It is obvious from (28) or (29) that the synthesis of the attack input u^* requires the knowledge of the current state z.

Now, a zero dynamics attack signal for the petro-chemical plant (12) is synthesized using the observer-based approach. Selecting u_1 as the attack input and setting u_2 to the constant value k_2 in (12) yields

$$\begin{aligned}\dot{z} &= (N_0 + N_2 k_2)z + N_1 u_1 z \\ &= \underbrace{\begin{bmatrix} -\frac{1}{V_1} & 0 \\ 0 & \frac{\alpha k_2}{V_2} \end{bmatrix}}_{G_0} z + \underbrace{\begin{bmatrix} \frac{1}{V_1} & \frac{1}{V_1} \\ -\frac{1}{V_2} & \frac{1}{V_2} \end{bmatrix}}_{G_1} u_1 z \\ y &= \underbrace{\begin{bmatrix} 1 & 0 \end{bmatrix}}_{C} z.\end{aligned}$$

The corresponding Hankel matrix has dimension $n = 1$. Therefore, the system is not susceptible to a zero dynamics attack. On the other hand, if the input u_2 is chosen as the attack input while setting u_1 to a constant value $k_1 \in \mathbb{R}$, the bilinear system in (12) becomes

$$\begin{aligned}\dot{z} &= (N_0 + N_1 k_1)z + N_2 u_2 z \\ &= \underbrace{\begin{bmatrix} \frac{k_1}{V_1} & -\frac{k_1}{V_1} \\ -\frac{k_1}{V_2} & \frac{\alpha + k_1}{V_2} \end{bmatrix}}_{G_0} z + \underbrace{\begin{bmatrix} \frac{1}{V_1} & 0 \\ 0 & 0 \end{bmatrix}}_{G_2} u_1 z \qquad (30a) \\ y &= \underbrace{\begin{bmatrix} 1 & 0 \end{bmatrix}}_{C} z. \qquad (30b)\end{aligned}$$

The above equations can be rewritten to calculate the attack signal $u_2^*(t)$ as follows:

$$\dot{z} = \left[G_0 + G_2 \left(-\frac{CG_0 z}{CG_1 z} \right) \right] z \qquad (31a)$$

$$u_2^* = -\frac{CG_0 z}{CG_1 z}. \qquad (31b)$$

Furthermore, during the attack, by design $z_1 = z_{1e} \in \mathbb{R} - \{0\}$, and thus, the system reduces

to the linear time-invariant (LTI) system

$$\dot{z} = \underbrace{\begin{bmatrix} \frac{k_1}{V_1} & -\frac{k_1}{V_1} \\ -\frac{k_1}{V_2} & \frac{\alpha+k_1}{V_2} \end{bmatrix}}_{\bar{A}} z + \underbrace{\begin{bmatrix} \frac{z_{1e}}{V_1} \\ 0 \end{bmatrix}}_{\bar{B}} u_2 \tag{32a}$$

$$y = \underbrace{\begin{bmatrix} 1 & 0 \end{bmatrix}}_{\bar{C}} z. \tag{32b}$$

The corresponding transfer function is

$$H(s) = \frac{(s - \frac{\alpha+k_1}{V_2})\frac{z_{1e}}{V_1}}{s^2 - (\frac{k_1}{V_1} + \frac{\alpha+k_1}{V_2})s + \frac{\alpha k_1}{V_1 V_2}}.$$

The system is non-minimum phase if $\alpha + k_1 > 0$. The system has one pole in strict right-half plane and another pole in the strict left-half plane. It is also obvious that the system has relative degree one. Note that in each case the linear systems above are distinct from the *linearized* models derived in the previous section. In particular, they are exact models and are not restricted to operate near the plant's equilibrium. In the present context, a zero dynamics attack corresponds to yielding a constant, possibly nonzero, output. Therefore,

$$\begin{aligned} \dot{y}(t) &= \bar{C}\dot{z}(t) \\ 0 &= \bar{C}(\bar{A}z(t) + \bar{B}u_2^*(t)) \\ u_2^*(t) &= -\frac{\bar{C}\bar{A}z(t)}{\bar{C}\bar{B}} \\ &= -\frac{k_1}{z_{1e}}(z_1(t) - z_2(t)). \end{aligned} \tag{33}$$

The stealthiness of the attack will depend on the accuracy of both the estimated model and the state estimates.

3.3.2 Analytical approach

The analytical approach computes the attack signal directly from the identified SISO model (27) of the bilinear system and the initial value. Since the relative degree is r, the

r-th derivative of the output is zero, that is,

$$0 = C(A_0^r z(t) + A_1 A_0^{(r-1)} z(t) u^*(t))$$
$$u^*(t) = -\frac{CA_0 z(t)}{CA_1 A_0^{(r-1)} z(t)}.$$

Substituting the formula for $u^*(t)$ in (27b), the zero dynamics attack input is found in terms of the following formula

$$\dot{z}(t) = \left(A_0 - A_1 \frac{CA_0^r z(t)}{CA_1 A_0^{r-1} z(t)} \right) z(t) \tag{34a}$$
$$u_1^*(t) = -\frac{CA_0^r z(t)}{CA_1 A_0^{r-1} z(t)} \tag{34b}$$

where $CA_1 A_0^{r-1} z(t) \neq 0$ [16].

Now the analytical approach to design an attack input for the petro-chemical plant (12) is described. Substituting (33) into (32) yields an explicit formula for $u_2^*(t)$:

$$\dot{z}(t) = \left(\bar{A} + \frac{\bar{B}\bar{C}\bar{A}}{\bar{C}\bar{B}} \right) z(t) \tag{35a}$$
$$\begin{aligned} u_2^*(t) &= -\left(\frac{\bar{C}\bar{A}}{\bar{C}\bar{B}} \right) z(t) \\ &= -\left(\frac{\bar{C}\bar{A}}{\bar{C}\bar{B}} \right) \exp\left[\left(\bar{A} + \frac{\bar{B}\bar{C}\bar{A}}{\bar{C}\bar{B}} \right) t \right] z_0 \\ &= \frac{k_1}{z_{1e}} [-1 \;\; 1] \exp\left(\begin{bmatrix} \frac{2k_1}{V_1} & -\frac{2k_1}{V_1} \\ -\frac{k_1}{V_2} & \frac{\alpha + k_1}{V_2} \end{bmatrix} t \right) \begin{bmatrix} z_{1e} \\ z_{2e} \end{bmatrix}. \end{aligned} \tag{35b}$$

Further simplifying (35b) yields an attack input of the form

$$u_2^*(t) = k_4 + k_5 \exp\left(\frac{\alpha + k_1}{V_2} \right) t, \tag{36}$$

where $k_4, k_5 \in \mathbb{R}$. As expected, the critical frequency of u_2^* is at the zero location.

A second way to design a zero dynamic attack signal is the Taylor series approximation of signal u_2 as follows:

$$u_2^*(t) = u_2(t_0) + u_2'(t_0)(t - t_0) + u_2''(t_0)\frac{(t - t_0)^2}{2!} + \dots \tag{37}$$

To facilitate the calculation one may consider truncating the infinite series at N-th term with sufficient accuracy. It is assumed that the zero dynamic attack signal is a continuous and infinitely differentiable function of time. The n-th coefficient of the series is determined by taking the n-th derivative of the equation (33) at time t_0. In general, for a bilinear or a non-linear system, the taylor series converges for small t in the neighborhood of t_0, and for a linear system one might expect a longer t. Since the bilinear system (30) gains a linear time-invariant property due to the zero dynamic condition on the output one might expect an attack input with longer duration t.

CHAPTER 4

ATTACK DETECTION METHODS

This chapter addresses the security of cyber-physical systems of bilinear type. The main goal is to develop an observer-based zero dynamics attack detection technique. The preliminaries for the observer development are first summarized: linear matrix inequalities (LMI), convex optimization, semidefinite programming, and Lyapunov stability theory. Then, a bilinear observer design method is described. A theorem for the convergence of such an observer is given. The observer is then employed as an attack detector.

4.1 LINEAR MATRIX INEQUALITIES

Linear matrix inequalities are a set of inequalities involving a linear function of symmetric matrices. They may include equality constraints along with the inequality constraints. LMIs are widely used in control theory, convex optimization, and system analysis. They have their roots in control theory and mathematical optimization, particularly in the latter half of the 20th century. The concept of matrix inequalities dates back to Aleksandr Lyapunov's work on stability theory in 1892 [5]. He introduced conditions involving positive definite matrices to analyze the stability of differential equations. Yakubovich [57, 58] and Kalman [32, 33] discovered a relationship between the existence of a positive definite matrix satisfying certain matrix inequalities and the Popov criterion. Later, this becomes known as the Kalman-Yakubovich-Popov (KYP) lemma providing an important link between control theory and LMIs, thus establishing conditions for the solvability of certain control problems in terms of matrix inequalities. During the 1970s and 1980s, the development of state space methods in optimal control theory, in particular, the algebraic Riccati equation, brought about the

use of LMIs to describe system properties and design controllers. In this dissertation, LMIs play a central role in determining a suitable gain for a bilinear state observer.

An LMI is an inequality in the form

$$F(x) = F_0 + x_1F_1 + x_2F_2 + \cdots + x_nF_n \succeq 0,$$

where

- $F_0, F_1, \ldots, F_n$ are given symmetric matrices of the same size;
- $x = [x_1, x_2, \ldots, x_n]$ is a vector of variables;
- The notation $F(x) \succeq 0$ means that all eigenvalues of $F(x)$ are non-negative.

The following example illustrates one application of an LMI.

Example 4.1: Consider the problem of finding a vector $[x_1 \quad x_2]^T$ such that the following matrix is positive definite:

$$A(x) = \begin{bmatrix} 1 & x_1 \\ x_1 & x_2 \end{bmatrix}.$$

First, write the matrix $A(x)$ as a linear combination of symmetric matrices:

$$\begin{aligned} A(x) &= A_0 + x_1A_1 + x_2A_2 \succeq 0 \\ &= \begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix} + x_1 \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix} + x_2 \begin{bmatrix} 0 & 0 \\ 0 & 1 \end{bmatrix} \succeq 0 \\ &= \begin{bmatrix} 1 & x_1 \\ x_1 & x_2 \end{bmatrix} \succeq 0. \end{aligned}$$

The matrix $\begin{bmatrix} 1 & x_1 \\ x_1 & x_2 \end{bmatrix} \succeq 0$ implies:

- The leading principle minor must be non-negative: $1 > 0$.
- The determinant of the matrix must be non-negative: $\det\left(\begin{bmatrix} 1 & x_1 \\ x_1 & x_2 \end{bmatrix}\right) = x_2 - x_1^2 \geq 0$, or $x_2 \geq x_1^2$.

LMIs provide a powerful framework due to their convex nature, making them amenable to efficient numerical solutions using interior-point methods and other optimization techniques. Their integration with convex optimization techniques has expanded the range of solvable problems, contributing to advancements in both theoretical research and practical applications.

4.2 CONVEX OPTIMIZATION

The development of convex optimization is a rich and multidisciplinary story, spanning several centuries and involving contributions from mathematics, economics, engineering, and computer science [6]. Its origin can be traced back to the era of Euclid, where his *Elements* laid down early geometric principles, some of which pertain to convex shapes. The development of calculus by Newton and Leibniz during the 17th century provided tools for optimization, particularly through differentiation. Other remarkable mathematical developments during this period are the methods developed by Euler for finding extrema of functionals, namely the method of Lagrange multipliers, which is crucial for constrained optimization problems. During the mid 19th century, Dantzig introduced the simplex method, a pivotal algorithm for solving linear programming problems efficiently [11]. In 1984, Karmarkar introduced an interior-point method for linear programming which was more efficient than the simplex method for large problems and reinvigorated interest in convex optimization [34]. Tools like CVX (a MATLAB-based software for convex optimization) and various Python

libraries (such as CVXPY) have made it easier to apply convex optimization techniques in practical applications.

Convex optimization deals with optimizing convex functions defined over convex sets. Therefore, it is worth briefly reviewing what convex sets and convex functions are before providing the general form of a convex optimization problem.

Definition 4.2.1 (**Convex sets** [6])**.** *A set $\mathcal{C} \subseteq \mathbb{R}^n$ is convex if for any two points $x, y \in \mathcal{C}$ the line segment connecting them lies entirely within C. Mathematically, $\mathcal{C}$ is convex if for all $x, y \in \mathcal{C}$ and $t \in [0,\ 1]$:*

$$tx + (1-t)y \in \mathcal{C}. \tag{38}$$

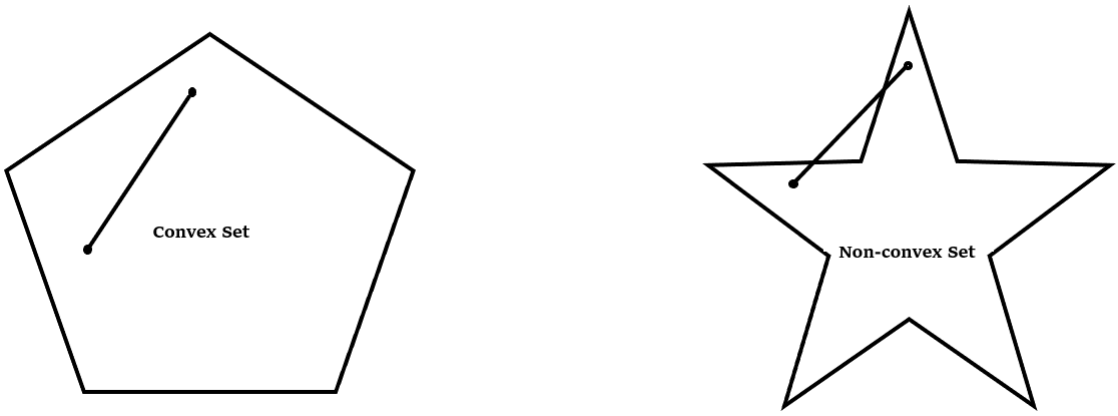

Figure 4: An example of convex and non-convex sets

Definition 4.2.2 (**Convex functions** [6])**.** *A function $f : \mathbb{R}^n \to \mathbb{R}$ is convex if its domain is a convex set and for all x, y in its domain and $t \in [0,\ 1]$:*

$$f(tx + (1-t)y) \leq tf(x) + (1-t)f(y). \tag{39}$$

This definition implies that the line segment between any two points on the graph of the function lies above or on the graph. Given a convex set $\mathcal{C}$ and a convex function f defined

on $\mathcal{C}$, a convex optimization problem has the form [6]

$$\begin{aligned} \min_{x \in \mathbb{R}^n} \quad & f(x) \\ subject\ \ to \quad & g_i(x) \geq 0, \quad i = 1, \ldots, m \\ & h_i(x) = 0, \quad i = 1, \ldots, p. \end{aligned} \tag{40}$$

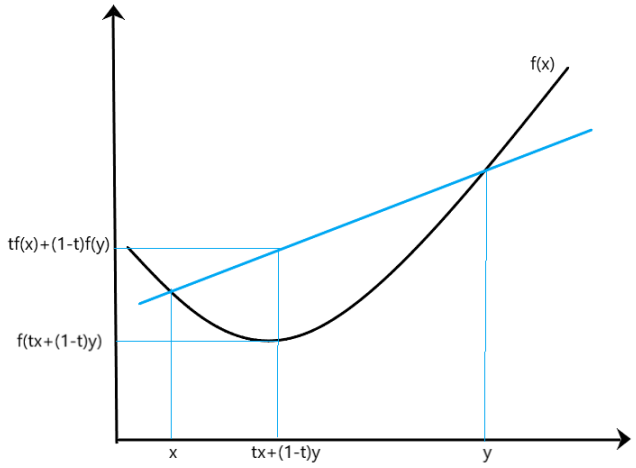

Figure 5: An example graph of a convex function

Figures 4 and 5 show typical examples of a convex and non-convex sets, and a graph of a convex function. One of the reasons for building and transforming a given mathematical or engineering problem into a convex optimization problem is that a convex optimization problem has a global solution, and there are well-developed algorithms (e.g., interior point method) to solve a convex optimization problem. There are several different types of convex optimization problems. Some widely used common types are linear programming (LP), quadratic programming (QP), second-order cone programming (SOCP), and semidefinite programming (SDP). This dissertation will use the semidefinite programming to cast the bilinear observer design problem into a convex optimization problem in order to obtain a set of feasible observer gains. In the next section, a more complete description of the semidefinite

programming is presented.

4.3 SEMIDEFINITE PROGRAMMING

The goal of semidefinite programming is to optimize a linear objective function subject to a constraint that a symmetric matrix variable is semidefinite. Such a constraint can be nonlinear but convex [6]. So semidefinite problems are convex optimization problems and can be solved using optimization softwares (YALMIP, MOSEK) [21],[22]. Consider the following optimization problem

$$\begin{aligned} &minimize \quad c^T x \\ &subject\ \ to \quad F(x) \geq 0, \end{aligned} \tag{41}$$

where

$$F(x) \triangleq F_0 + \sum_{i=1}^{m} x_i F_i.$$

Here, $c \in \mathbb{R}^m$ is the objective function data vector, $x \in \mathbb{R}^m$ is a vector comprised of the optimizing variables, and $F_0, \ldots, F_m \in \mathbb{R}^{n \times n}$ are the $m+1$ symmetric matrices. The constraint ensures that $F(x)$ is positive semidefinite, i.e., $z^T F(x) z \geq 0$ for all $z \in \mathbb{R}^n$. Since semidefinite program is a convex optimization problem, it satisfies the convexity condition, that is, for $x, y \in \mathbb{R}$ and for all $0 \leq t \leq 1$,

$$F(tx + (1-ty)) = tF(x) + (1-t)F(y) \geq 0.$$

One of the most common applications in control theory is the minimizing maximum eigenvalue problem. Suppose $A(x) = A_0 + A_1 x_1 + \cdots + A_m x_m$, where $A_i = A_i^T \in \mathbb{R}^{n \times n}$. The problem of minimizing the maximum eigenvalue of the symmetric matrix $A(x)$ can be cast as the semidefinite program [46]

$$minimize\ \ t$$

$$subject\ to\ \ tI - A(x) \geq 0$$

with variables $x \in \mathbb{R}^n$ and $t \in \mathbb{R}$. Another widely used example is casting a nonlinear (convex) optimization problem as a semidefinite problem [46]

$$\begin{aligned} minimize\ \ & \frac{(c^T x)^2}{d^T x} \\ subject\ to\ \ & Ax + b \geq 0, \end{aligned}$$

where $d^T x > 0$ if $Ax + b \geq 0$. By introducing an auxiliary variable t, the above nonlinear problem can be rewritten as

$$\begin{aligned} minimize\ \ & t \\ subject\ to\ \ & Ax + b \geq 0, \\ & \frac{(c^T x)^2}{d^T x} \leq 0. \end{aligned}$$

In this formulation, t serves as an upper bound on the objective function, and it is a linear function. These constraints, in turn, can be expressed as a semidefinite programming problem [46],

$$\begin{aligned} & minimize\ \ t \\ & subject\ to \begin{bmatrix} \mathbf{diag}(Ax + b) & 0 & 0 \\ 0 & t & c^T x \\ 0 & c^T x & d^T x \end{bmatrix} \geq 0, \end{aligned}$$

where $x \in \mathbb{R}^n$ and $t \in \mathbb{R}$ are semidefinite programming variables.

Example 4.2: Consider the following linear dynamic system example to illustrate how a simple semidefinite problem is formulated and solved

$$\dot{z} = Az. \tag{42}$$

Here the goal is to prove the stability of (42) by finding a symmetric P satisfying

$$A^T P + PA < 0 \tag{43a}$$

$$P > 0. \tag{43b}$$

The inequalities (43) denote a semidefinite programming which can be solved using any numerical software tool like CVX or CVXPY.

4.4 STABILITY ANALYSIS USING LYAPUNOV THEORY:

Lyapunov stability methods involve constructing a Lyapunov function to assess the stability of a dynamical system's equilibrium point. The key is to find a positive definite function whose time derivative is negative semidefinite or negative definite. This method provides a powerful tool for stability analysis without requiring the explicit solution of the system's differential equations. The key definitions are summarized below.

Definition 4.4.1 (**Equilibrium point**). *An equilibrium point x^* of a system described by $\dot{x} = f(x,t)$, $x(t_0) = x_0$ is a point $f(x^*,t) = 0$ for all t [37].*

From a stability point of view, an equilibrium point of a dynamical system can be classified as stable, asymptotically stable, exponentially stable, or unstable point. Without loss of generality, the origin can be regarded as an equilibrium state. However, a nonlinear system may have more than one equilibrium state.

Definition 4.4.2 (**Stable:**). *The origin of a system $\dot{x} = f(x,t)$, $x(t_0) = x_0$ is said to be stable in the sense of Lyapunov (SISL), if for every $\epsilon > 0$, there exists a $\delta = \delta(\epsilon, t_0) > 0$ such that if $\|x(t_0)\| < \delta$, then $\|x(t)\| < \epsilon$ for $t > t_0$ [37].*

Definition 4.4.3 (**Asymptotically stable**). *The origin of a system $\dot{x} = f(x,t)$, $x(t_0) = x_0$*

is said to be asymptotically stable (AS), if it is Lyapunov stable, and if there exists a $\delta(t_0) > 0$ *such that if* $\|x(t_0)\| < \delta$*, then* $\lim_{t\to\infty} \|x(t)\| = 0$ *[37].*

Definition 4.4.4 (**Exponentially stable**). *The origin of a system* $\dot{x} = f(x,t)$*,* $x(t_0) = x_0$ *is said to be exponentially stable (ES), if it is asymptotically stable and if there exists a* $\alpha, \beta, \delta > 0$ *such that if* $\|x(t_0)\| < \delta$*, then* $\|x(t)\| = \alpha\|x(t_0)\| \exp(-\beta t)$ *for* $t \geq 0$ *[37].*

Definition 4.4.5 (**Unstable**). *The origin of a system* $\dot{x} = f(x,t)$*,* $x(t_0) = x_0$ *is said to be unstable if it is not SISL [37].*

Theorem 4.4.1 (**Lyapunov's stability theorem**). *[37] Consider a system described by*

$$\dot{x} = f(x,t) \tag{44}$$

where $f(0,t) = 0$ *for all* $t > t_0$*. If there exist a scalar-valued, radially unbounded function* $V(x)$ *having continuous partial derivatives and satisfying the conditions:*

1. $V(x)$ *is positive definite, or* $V(x) > 0 \ \forall\, x \in \mathbb{R}^n \setminus \{0\}$*;*
2. $V(x) = 0$ *if and only if* $x = 0$*;*
3. $\dot{V}(x)$ *is negative definite, i.e.,* $\dot{V}(x) < 0 \ \forall\, x$*,*

then the equilibrium point is globally asymptotically stable.

The Lyapunov stability theorem 4.4.1 describes a specific kind of stability: global asymptotically stability. This dissertation will consider designing a globally asymptotically stable bilinear observer to detect a zero dynamics attack. Other notions of stability, such as global exponential stability, will be pursued in future research for observer design. For nonlinear systems, there is no general procedure for finding Lyapunov functions. Finding an appropriate candidate Lyapunov function is the key step in determining whether the corresponding nonlinear system is stable at the point of interest.

Example 4.3: Consider the simple nonlinear system

$$\dot{x} = -x^3.$$

The equilibrium point is

$$\begin{aligned} \dot{x} &= 0 \\ \Rightarrow -x^3 &= 0 \\ \Rightarrow x &= 0. \end{aligned}$$

A good candidate for a Lyapunov function is $V(x) = \frac{1}{2}x^2$. Observe:

1. $V(x)$ is positive definite since $V(x) > 0$ for all $x \neq 0$.

2. $V(0) = 0$ if and only if $x = 0$.

3. Taking the time derivative of $V(x)$ gives

$$\begin{aligned} \dot{V}(x) &= \frac{dV}{dt} \\ &= \frac{d}{dt}\Big(\frac{1}{2}x^2\Big) \\ &= x\dot{x}. \end{aligned}$$

   Substituting $x\dot{x} = -x^3$ gives

$$V(x) = x(-x^3) = -x^4.$$

   $\dot{V}(x) = -x^4$ is negative definite because $\dot{V}(x) < 0$ for all $x \neq 0$ and $\dot{V}(0) = 0$.

Since $V(x)$ is positive definite, and $\dot{V}(x)$ is negative definite, by Lyapunov's Theorem 4.4.1, the equilibrium at $x^* = 0$ is asymptotically stable.

An observer design method is proposed in the next section using the theories presented in the previous subsections. A bilinear observer is formulated to track the states of the plant.

The difference between the observer states and plant states defines an error system. An appropriate candidate Lyapunov function is chosen to ensure the global asymptotic stability of these dynamics so that the state estimates of the observer asymptotically approach the true state values of the plant. The Lyapunov function is used to formulate a convex optimization problem in the form of semidefinite programming and linear matrix inequalities. The solution of the convex optimization problem gives a feasible set for the observer gains.

4.5 OBSERVER DESIGN METHOD

Consider the following observer for a bilinear plant of the form (2)

$$\dot{\hat{z}} = N_0\hat{z} + \sum_{i=1}^{m} N_i\hat{z}u_i + L_0(y-\hat{y}) + \sum_{i=1}^{m} L_i(y-\hat{y})u_i \tag{45a}$$

$$\hat{y} = C\hat{z}, \tag{45b}$$

where $\hat{z}$, $\hat{y}$, and L_i, $i = 0, 1, \ldots, m$ are the observer states, outputs, and gains, respectively. Define the error between the plant states and observer states as $e = z - \hat{z}$. Therefore, the error dynamics are

$$\dot{e} = (N_0 - L_0C)e + \sum_{i=1}^{m}(N_i - L_iC)eu_i. \tag{46}$$

Theorem 4.5.1. *Assume there exist a symmetric positive definite matrix $P \in \mathbb{R}^{n\times n}$ and matrices $S_i \in \mathbb{R}^{n\times l}$, $i = 0, 1, \ldots, m$ for the n dimensional system (45) such that*

$$\begin{aligned} F_0 &< 0 \\ F_i &= 0, \quad i = 1, 2, \ldots, m, \end{aligned} \tag{47a}$$

where

$$F_i = N_i^TP - C^TS_i^T + PN_i - S_iC, \quad i = 0, 1, \ldots, m \tag{47b}$$

Then the observer (45) has globally asymptotically stable error dynamics (46) when the observer gains are

$$L_i = P^{-1}S_i, \quad i = 0, 1, \ldots, m. \tag{48}$$

Proof. For a given positive definite matrix P, define the quadratic function

$$V(e,t) = e^T(t)Pe(t) > 0, \tag{49}$$

where $e(t)$ satisfies (46). The positive definitiveness of P ensures that V is radially unbounded. Taking the derivative gives

$$\begin{aligned}
\dot{V} &= \dot{e}^T Pe + e^T P\dot{e} \\
&= e^T(N_0 - L_0C)^T Pe + e^T\left(\sum_{i=1}^{m}(N_i - L_iC)^T\right)Peu_i \\
&\quad + e^T P(N_0 - L_0C)e + e^T P\left(\sum_{i=1}^{m}(N_i - L_iC)\right)eu_i \\
&= e^T(N_0^T P - C^T L_0^T P + PN_0 - PL_0C)e \\
&\quad + e^T(N_1^T P - C^T L_1^T P + PN_1 - PL_1C)eu_1 \\
&\vdots \\
&\quad + e^T(N_m^T P - C^T L_m^T P + PN_m - PL_mC)eu_m.
\end{aligned}$$

Substituting $PL_i = S_i$ for $i = 0, 1, \ldots, m$ in the above equation, the following linear matrix function is obtained

$$\begin{aligned}
\dot{V} &= e^T(N_0^T P - C^T S_0^T + PN_0 - S_0C)e \\
&\quad + e^T(N_1^T P - C^T S_1^T + PN_1 - S_1C)eu_1 \\
&\vdots \\
&\quad + e^T(N_m^T P - C^T S_m^T + PN_m - S_mC)eu_m
\end{aligned}$$

$$= e^T(F_0 + F_1u_1 + \ldots + F_mu_m)e < 0, \tag{50}$$

where the F_i are defined in (47b). By assumption, there exists S_i such that (47a) is satisfied. Therefore, $\dot{V}(e) = e^T F_0 e < 0, \quad \forall e \neq 0$. Thus, by Lyapunov's Theorem 4.4.1, the error dynamics are globally asymptotically stable. The observer gains are as given in (48). □□□

The inequality (50) along with the constraints in (47a) constitute a linear matrix inequality feasibility problem. It can be efficiently solved using any convex optimization algorithm (e.g., interior-point method) and suitable software tools. Assuming $F_i = 0$ for $i = 1, 2, \ldots, m$, it is easy to show that $F_0 < 0$ if $(N_0 - L_0C)$ in equation (46) is stable. That is, a necessary condition for a feasible solution to the LMI problem is the detectability of the linear part (N_0, C) of the bilinear system.

Next, a qualitative discussion is presented comparing the proposed bilinear observer design with a linear observer design. First, the key steps for a linear observer design and the proposed bilinear observer design are listed and compared. The objectives are: to show that, in general, the bilinear observer design is a nonconvex problem, whereas linear observer design is a convex problem; to point out the source of the nonconvexity in the proposed bilinear observer design; and to explain how the nonconvex problem is turned into a convex problem.

Consider the linear system:

$$\dot{z} = N_0z + Bu, \quad z(0) = z_0 \tag{51a}$$

$$y = Cz. \tag{51b}$$

1. Observer: $\dot{\hat{z}} = N_0\hat{z} + LCe$ with gain L and error $e = z - \hat{z}$, $\hat{z}(0) = 0$.

2. Error dynamics: $\dot{e} = (N_0 - LC)e$.

3. Lyapunov function: $V(e) = e^T P e$, where $P = P^T > 0$.

4. $\dot{V} = e^T (\underbrace{N_0^T P - C^T L^T P + P N_0 - P L C}_{convex,\ F_0<0}) e < 0.$

Consider the general SISO bilinear system:

$$\dot{z} = N_0 z + N_1 z u, \quad z(0) = z_0 \tag{52a}$$

$$y = Cz. \tag{52b}$$

1. Observer: $\dot{\hat{z}} = N_0 \hat{z} + N_1 \hat{z} u + (L_0 + L_1 u) C e$, with gains L_0, L_1 and error $e = z - \hat{z}$, $\hat{z}(0) = 0$.

2. Error dynamics: $\dot{e} = (N_0 - L_0 C) e + (N_1 - L_1 C) e u$.

3. Lyapunov function: $V(e) = e^T P e$ and $P = P^T > 0$.

4. $\dot{V} = e^T [(\underbrace{N_0^T P - C^T L_0^T P + P N_0 - P L_0 C}_{convex,\ F_0<0}) + (\underbrace{N_0^T P - C^T L_1^T P + P N_0 - P L_1 C}_{nonconvex,\ F_1=0}) u] e < 0.$

Steps 1 and 2 represent the corresponding observer dynamics and the error dynamics for the linear system (51) and general bilinear system (52) with their corresponding gains. In step 3, the same Lyapunov candidate function is chosen to facilitate the comparison in both linear and bilinear cases. In step 4, the derivatives of the Lyapunov functions $\dot{V}$ are shown. For the linear observer, $\dot{V} = e^T F_0 e$ is a convex function, where F_0 can be expressed as a linear combination of matrices. It can be solved for L_0 such that the third criterion of the Lyapunov stability theory is satisfied, or $\dot{V} < 0$. This can be done by enforcing $F_0 < 0$ in the convex optimization algorithm. However, the $\dot{V}$ is a nonconvex function for the bilinear observer. It has two terms: the first term or the convex term $e^T F_0 e$, and the second term or the nonconvex term $e^T F_1 e u$, where F_0, F_1 are linear combinations of matrices. The presence of the input u in the second term is the source of nonconvexity. Putting the constraints $F_0 < 0$ and $F_1 = 0$ in the optimization algorithm turns the nonconvex problem into a

convex problem. The constraint $F_1 = 0$ gives a value of the observer gain L_1 that makes the nonconvex term zero, whereas the constraint $F_0 < 0$ gives a value of the observer gain L_0 which makes the $\dot{V} < 0$. The convex optimization problem is solved in a single step, where F_0 and F_1 serve as the equality and inequality constraint of the optimization problem, respectively. The nonconvexity is associated with the F_1 coefficient matrix, and the F_1 is enforced to be zero, thus neutralizing the effect of the nonconvexity in the function. For a multi-input general bilinear system, there will be m nonconvex terms F_i for $i = 1, 2, \ldots, m$. By putting each $F_i = 0$ one can get the corresponding observer gain L_i which neutralizes the corresponding nonconvex term stated in the Theorem 4.5.1. The fact that F_0 is a negative definite matrix makes the observer globally asymptotically stable.

CHAPTER 5

NUMERICAL EXAMPLES

Two examples are presented in this chapter. The first example shows how a zero dynamics attack is executed on a petro-chemical plant. The second example illustrates how the deviation in the states during attack are detected and tracked using a bilinear observer. This provides the enabling theory for a zero dynamics attack detector.

5.1 ATTACK SYNTHESIS

Consider the bilinear model of a petro-chemical plant shown in (12)

$$\dot{z} = \begin{bmatrix} 0 & 0 \\ 0 & \frac{\alpha}{V_2} \end{bmatrix} z + \begin{bmatrix} \frac{1}{V_1} & -\frac{1}{V_1} \\ -\frac{1}{V_2} & \frac{1}{V_2} \end{bmatrix} z u_1 + \begin{bmatrix} \frac{1}{V_1} & 0 \\ 0 & 0 \end{bmatrix} z u_2, \qquad z(0) = z_0 \tag{53a}$$

$$\dot{z} = \underbrace{\left(\begin{bmatrix} 0 & 0 \\ 0 & \frac{\alpha}{V_2} \end{bmatrix} + \begin{bmatrix} \frac{k_1}{V_1} & -\frac{k_1}{V_1} \\ -\frac{k_1}{V_2} & \frac{k_1}{V_2} \end{bmatrix} \right)}_{A_0} z + \begin{bmatrix} \frac{1}{V_1} & 0 \\ 0 & 0 \end{bmatrix} z u_2, \qquad z(0) = z_0 \tag{53b}$$

$$y = \begin{bmatrix} 1 & 0 \end{bmatrix} z. \tag{53c}$$

Choosing parameters values as in the Table 1, one gets the bilinear system

$$\dot{z} = \underbrace{\begin{bmatrix} -0.01 & 0.01 \\ 0.005 & 0.025 \end{bmatrix}}_{M_0} z + \underbrace{\begin{bmatrix} 0.01 & 0 \\ 0 & 0 \end{bmatrix}}_{M_2} z u_2, \qquad z_0 = \begin{bmatrix} 4 \\ 10 \end{bmatrix} \tag{54a}$$

$$y = \underbrace{\begin{bmatrix} 1 & 0 \end{bmatrix}}_{C} z. \tag{54b}$$

The identification algorithm was run on a set of six sampled input-output data sets. No measurement noise was included since the robustness of the identification algorithm appears

to be an open problem in the literature. The sampling time Δt was selected to be 0.1 seconds. After applying the similarity transformation T, the identified bilinear system was found to be:

$$\hat{M}_0 = \begin{bmatrix} -0.010000000000000 & 0.010000000000000 \\ 0.004999998030822 & 0.025000000784774 \end{bmatrix}$$

$$\hat{M}_2 = \begin{bmatrix} 0.010000000001673 & -0.000000000000658 \\ -0.000000002424625 & 0.000000000967701 \end{bmatrix}$$

$$\hat{C} = \begin{bmatrix} 1.000000000000000 & 0.000000000000000 \end{bmatrix}$$

$$\hat{z}_0 = \begin{bmatrix} 4.000000000000016 \\ 10.000000000000695 \end{bmatrix}$$

$$T = \begin{bmatrix} -0.5764769645900 & 0.7076425044075 \\ 1.4468084807027 & -705.3391138105984 \end{bmatrix}.$$

The zero dynamic attacks signal u_2^* was generated from the identified system using both

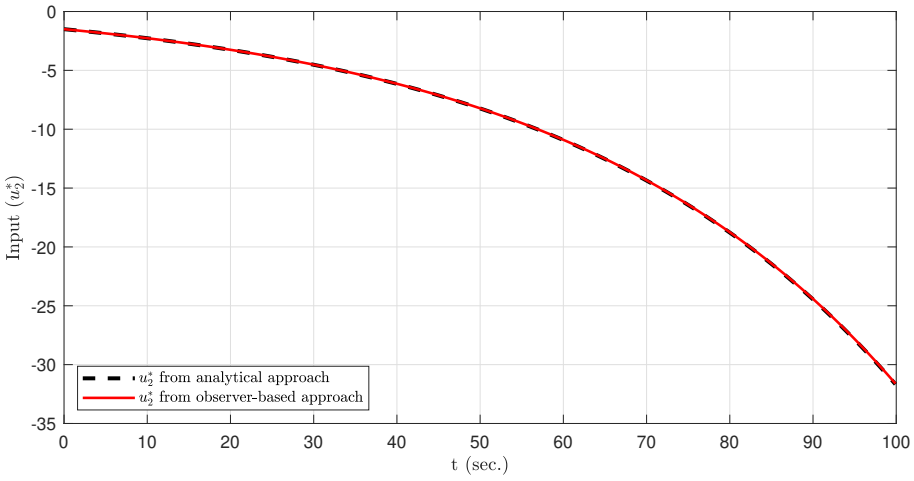

Figure 6: Attack inputs u_2^* applied to the plant

the observer-based and analytical approaches. An open-loop observer computed the state estimate using the identified model and initial condition. This approach worked reliably, even

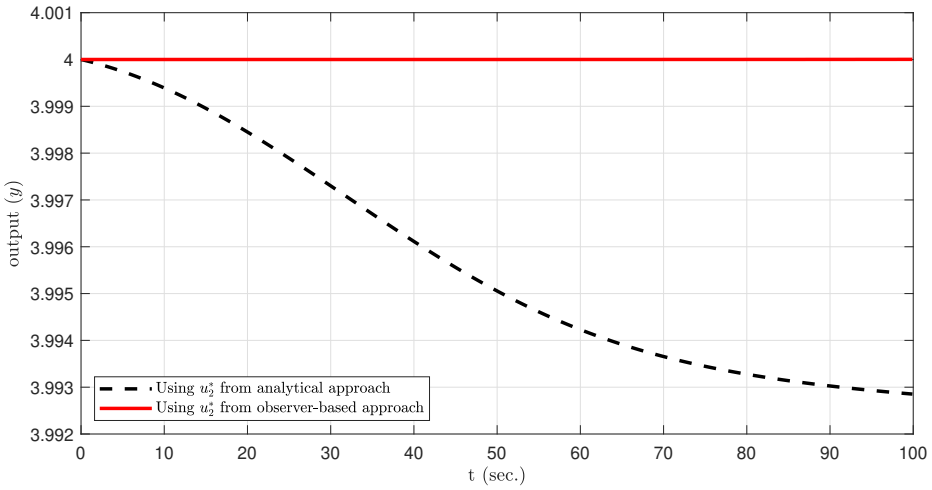

Figure 7: Outputs y when the plant is under a zero dynamics attack

though the linearized plant is marginally stable. Since there was no measurement noise, and the identified initial condition was extremely accurate. The analytical attack signal was found to be

$$\hat{u}_2^*(t) = 1.21 - 2.71 \exp(0.025t).$$

The attack signals using both approaches are shown in Figure 6 and are nearly identical. Figure 7 shows the output response to each attack signal. Both attacks are clearly very stealthy for the first 100 seconds. The observer-based approach is more robust but at the expense of more computations. Figure 8 shows the system states during each attack. As expected z_2 grows without bounds during the attack. The bilinear system (54) represents a special case, where during the attack, by design $z_1 = z_{1e} = 4$, and thus, the system reduces to the linear time-invariant (LTI) system as shown in (32). The position of the non-minimum phase zero $g := (\alpha + k_1)/V_2$ determines the severity of the attack. The larger k_1 in magnitude, the larger the growth constant of the attack signal, hence the more destructive the attack. Table 4 gives g and $z_2(100)$ for various values of k_1.

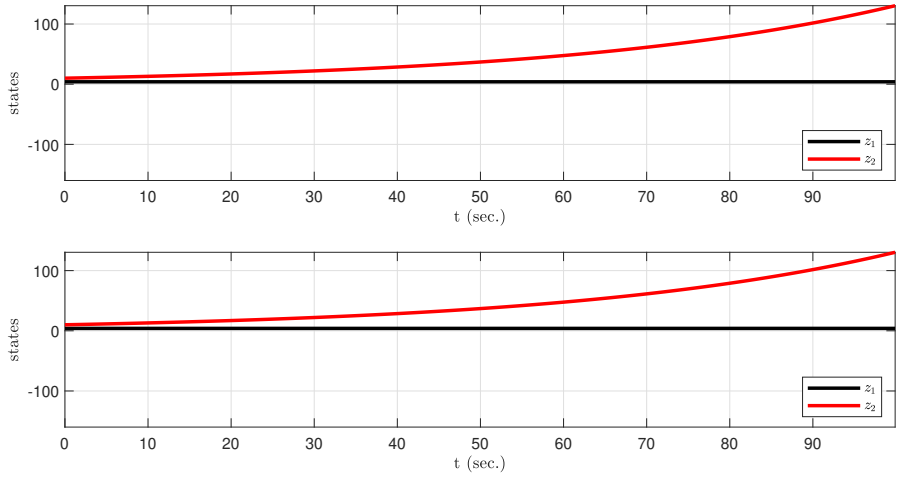

Figure 8: State trajectories of the system for the analytical attack (top) and the observer-based attack (bottom)

TABLE 4: Values of g, $z_2(100)$ for different $u_1(t) = k_1$

$\boldsymbol{k_1}$	$\boldsymbol{g}$	$\boldsymbol{z_2(100)}$
-1	0.025	130.77
-2	0.02	86.66
-3	0.015	58.74
-4	0.01	40.92
-5	0.005	29.46

5.2 ATTACK DETECTION

Reconsider the SISO model of the petro-chemical plant given in the previous section. To demonstrate the efficacy of the proposed detection system, the zero dynamics attack to the plant (54) is recreated in a simulation environment. The identified bilinear model of the plant and the synthesized attack signal $u_2^*(t)$ from the previous section are used to recreate

the attack scenario.

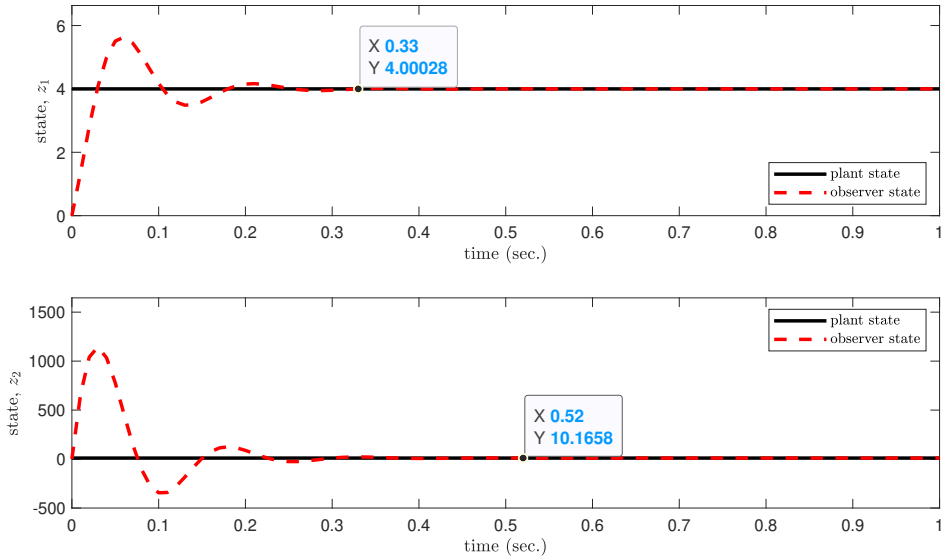

Figure 9: Observer tracking the plant states z_1 (top) and z_2 (bottom) under no attack

An observer-based detection system is designed using the method proposed in Chapter 4. It is notable that in Section 5.1, the linear part (N_0, C) of the system (53) is not detectable. But according to Theorem 4.5.1, (N_0, C) must be detectable to satisfy $F_0 < 0$. To resolve this issue, one can choose either u_1 or u_2 to be constant and then combine the corresponding coupling term with the linear term to get a modified $(\bar{N}_0, C)$ which is detectable. For example, let $u_1 = a$, $a \in \mathbb{R}$, then (53) is detectable if

$$det\left(\begin{bmatrix} \frac{a}{V_1} & -\frac{a}{V_1} \\ -\frac{a}{V_2} & \frac{a+\alpha}{V_2} \end{bmatrix}\right) \neq 0$$

$$a \neq 0.$$

Therefore, according to Table 2, the system is detectable for $u_1 \in \mathbb{R}_-$. The detectable linear part $(\bar{N}_0, C)$ of the SISO realization (54) is obtained after setting u_1 to a constant -1. In that case, the error dynamics become

$$\dot{e} = [(N_0 + N_1(-1)) - (L_0 + L_1(-1))C] + (N_2 + L_2C)u_2$$

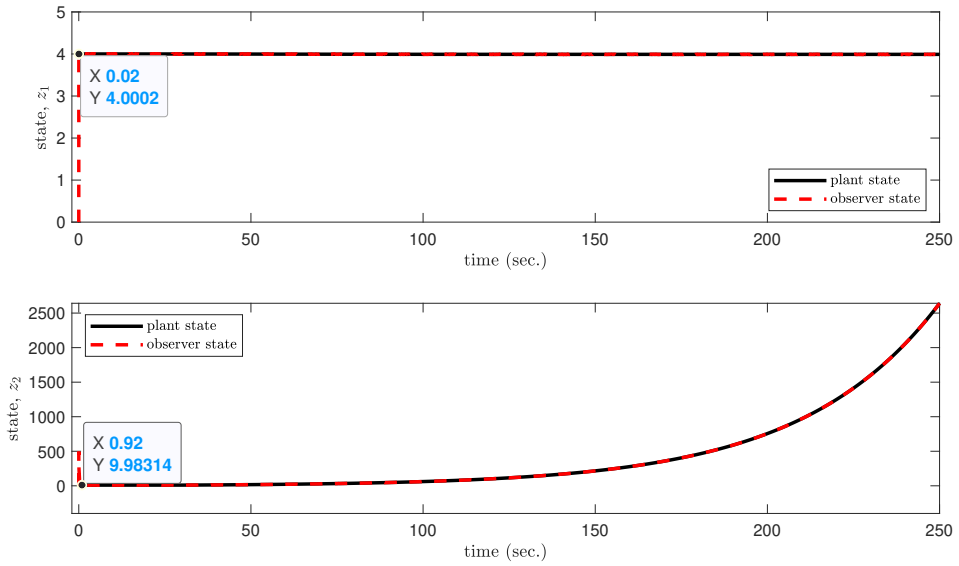

Figure 10: Observer tracking the plant states z_1 (top) and z_2 (bottom) when attack is initiated at $t = 30$ seconds

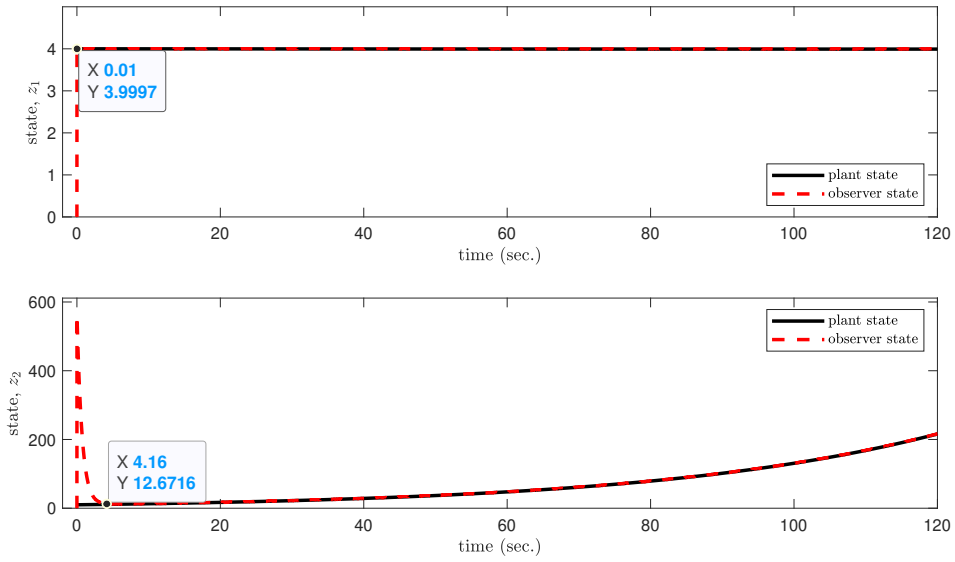

Figure 11: Observer tracking the plant states z_1 (top) and z_2 (bottom) when attack and observer are initiated at the same moment (worst case)

$$= (\bar{N_0} + \bar{L_0}C) + (N_2 + L_2C)u_2,$$

where $\bar{N_0} = N_0 + N_1(-1)$ and $\bar{L_0} = L_0 + L_1(-1)$. Assuming the Lyapunov candidate $V = e^T Pe$, it follows that

$$\dot{V} = e^T \underbrace{(\bar{N_0}^T P - C^T S_0^T + P\bar{N_0} - S_0 C)}_{F_0} e + e^T \underbrace{(N_2^T P - C^T S_2^T + PN_2 - S_2 C)}_{F_2} e.$$

The observer gains for the petro-chemical plant are calculated by solving a semidefinite programming. The semidefinite programming is solved using optimization software MOSEK, version 10.1.21 [46] and MATLAB $R2022b$. The machine configuration used during the optimization code's implementation is Intel(R) $Core(TM)i7-8565U$ CPU @ 1.80 GHz, and 16.0 GB RAM capacity. In the numerical computation, the followings are set as constraints,

$$F_0 < 0,$$

$$F_2 = 0,$$

$$P \leq 0.01 \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}.$$

and the gains are found to be

$$L_0 = \begin{bmatrix} 30.2702 & 61.2556 \end{bmatrix}^T$$
$$L_1 = \begin{bmatrix} -0.0102 & -1975.3693 \end{bmatrix}^T$$
$$L_2 = \begin{bmatrix} 0.0100 & 0.0000 \end{bmatrix}^T.$$

Figures 10 and 11 show the performance of the observer with no attack and during an attack initiated at $t = 30$ seconds, respectively. In both figures, the initial state for the observer was set to zero. It is evident that the observer states converge to the corresponding plant states in less than 0.72 seconds for this particular equilibrium. Figure 11 shows the

worst case scenario, where the observer and attack are initiated simultaneously. In this case, the observer requires much longer (around 4.16 seconds) to converge to the true z_2 state (bottom plot). By the time the observer states converge to the plant states, the unmonitored plant state z_2 has already deviated by 19% from its nominal value, which is the worst case. An alarm can be triggered for general cases when the observer state crosses a certain threshold value. An isolation procedure can then be initiated to defend the plant. For example, suppose the threshold is set to $z_e \pm 5\% z_e$. If any state of the state vector crosses the respective preset upper or lower threshold values, it will trigger an alarm. The detector also needs to consider the observer's transient response time to avoid false alarms. For example, the alarm is trigged at time t if the observer states $\hat{z} \geq (z_e \pm 5\% z_e)$ and $t \geq T_{conv}$ to avoid false alarm. T_{conv} is chosen so that it is greater than the longest convergence time of the states at equilibrium. Incorporating the transient time constraint with the maximum allowable deviation of the state in the defense mechanism will cover both the general and worst-case attack scenarios.

CHAPTER 6

CONCLUSIONS

This dissertation had four main objectives as stated in Section 1.6. Each objective is restated below with a summary of the solution presented in the dissertation.

1. Show how in general an adversary can successfully execute a malicious zero dynamics attack on an unknown bilinear system.

Chapter 3 demonstrated the first objective. The first step in the attack procedure was to perform system identification. Once the system was identified, the model was used to design the attack signal. Two approaches were developed: an observer-based approach and an analytical approach.

2. Demonstrate by simulation a zero dynamics attack using a bilinear model of a petro-chemical plant.

The second objective of simulating the zero dynamics attack was illustrated in Section 5.1 in Chapter 5 using a model of a petro-chemical plant. As part of the demonstration, the identification algorithm was applied first on a set of input-output data after the plant reaches an equilibrium and the identified system was used to generate attack signals. Both attack approaches, the observer-based approach, and the analytical approach, were demonstrated numerically and found to be effective.

3. Develop methods to detect a zero dynamics attack on a bilinear system.

A zero dynamics attack detection technique was proposed in Chapter 4. A bilinear observer-based attack detection method was developed, and the convergence of the observer error to zero were ensured using Lyapunov stability theory. The observer gain determination problem was formulated as a semidefinite programming problem and solved using convex optimization toolsets. The required conditions for the existence of such an observer were provided.

4. Demonstrate by simulation the effectiveness of these attack detection methods on the petro-chemical plant.

It was demonstrated in Section 5.2 how to design an observer to detect zero dynamics attacks on this plant. Its performance under normal operation and under zero dynamics attack were found to be effective in both cases.

BIBLIOGRAPHY

[1] Y. Adachi, Y. Funahashi, Existence of bilinear observers for bilinear systems, *Inf. Sci.*, 19 (1979) 67–80.

[2] A. Banimerian, K. Khorasani, N. Meskin, Monitoring and detection of malicious adversarial zero dynamics attacks in cyber-physical systems, *IEEE Conf. on Control Technology and Applications*, Montreal, Canada, 2020, pp. 726–731.

[3] A. Banimerian, K. Khorasani, N. Meskin, A special class of zero dynamics cyber-attacks for SISO time-delay systems, *Proc. of 60th IEEE Conf. on Decision and Control*, Austin, TX, 2021, pp. 4182–4187.

[4] S. Berhe, H. Unbehauen, Bilinear continuous-time systems identification via Hartley-based modulating functions, *Automatica*, 34 (1998) 99–503.

[5] S. Boyd, L. E. Ghaoui, E. Feron, V. Balakrishnan, *Linear Matrix Inequalities in System and Control Theory*, SIAM, Philadelphia, PA, 1994.

[6] S. Boyd, L. Vandenberghe, *Convex Optimization*, Cambridge University Press, Cambridge, UK, 2004.

[7] L Cao, X. Jiang, Y. Zhao, S. Wang, D. You, X. Xu, A survey of network attckes on cyber-physical systems, *IEEE Access*, 8 (2020) 44219–44227.

[8] Y. Chen, S. Kar, J. M. F. Moura, Dynamic attack detection in cyber-physical systems with side initial state information, *IEEE Trans. Automat. Control*, 62 (2017) 4618–4624.

[9] W. J. Culver, On the existence and uniqueness of the real logarithm of a matrix, *Proc. Amer. Math. Soc.*, 17 (1966) 1146–1151.

[10] H. Dai, N. K. Sinha, Robust identification of systems using block-pulse functions, *IEE Proceedings D (Control Theory and Applications)*, 139 (1992) 308–316.

[11] G. B. Dantzig, The simplex method, Santa Monica, CA: RAND Corporation, 1956.

[12] A. C. B. de Oliveira, M. Siami, E. D. Sontag, Bilinear dynamical networks under malicious attack: An efficient edge protection method, *Proc. 2021 American Control Conf.*, New Orleans, LA, 2021, pp. 1210–1215.

[13] D. Ding, Q. Han, Y. Xiang, X. Ge, X. Zhang, A survey on security control and attack detection for industrial cyber-physical systems, *Neurocomputing*, 275 (2018) 1674–1683.

[14] F. R. Gantmacher, *The Theory of Matrices*, vol. 1, Chelsea Publishing Company, New York, 1959.

[15] J. Giraldo, D. Urbina, Á. Cárdenas, J. Valente, M. Faisal, J. Ruths , N. Tippenhauer, H. Sandberg, R. Candell, A survey of physics-based attack detection in cyber-physical systems, *ACM Computing Surveys*, 51 (2018) 1–36.

[16] W. S. Gray, L. A. Duffaut Espinosa, and M. Thitsa, Left Inversion of Analytic Nonlinear SISO Systems via Formal Power Series Methods, *Automatica*, 50 (2014) 2381–2388.

[17] W. S. Gray, K. Ebrahimi-Fard, A. Schmeding, Universal zero dynamics: SISO case, *Proc. 55th Conf. on Information Sciences and Systems*, Baltimore, MD, 2021.

[18] W. S. Gray, L. A. Duffaut Espinosa, M. A. Haq, Universal zero dynamics attacks using only input-output data, *Proc. 2022 American Control Conf.*, Atlanta, GA, 2022, pp. 4985–4991.

[19] J. Ha, H. Shim, Study on realizable generalized hold functions as a countermeasure against zero dynamics attack, *Proc. 58th IEEE Conf. on Decision and Control*, Nice, France, 2019, pp. 5362–5367.

[20] M. A. Haq, W. S. Gray, Zero dynamics attacks on unknown bilinear systems: Case study, *Proc. 7th IEEE International Conf. on Industrial Cyber-Physical Systems*, St. Louis, MO, 2024.

[21] M. A. Haq, I. B. Kucukdemiral, H_∞ observer design for linear time-delay systems, *Proc. 9th International Conf. on Electrical and Electronics Engineering*, Bursa, Turkey, 2015, pp. 848–853.

[22] M. A. Haq, I. B. Kucukdemiral, Observer design with better delay margin for linear time-delay systems, *IU-JEEE*, 16, 2016, pp. 2065–2071.

[23] C. Hawng, M. Chen, Analysis and parameter identification of bilinear systems via shifted Legendre polynomials, *Int. J. Control*, 44 (1986) 351–362.

[24] A. Hoehn, P. Zhang, Detection of covert attacks and zero dynamics attacks in cyber-physical systems, *2016 American Control Conf.*, 2016, Boston, MA, pp. 302–307.

[25] A. Howell, J. K. Hedrick, Nonlinear observer design via convex optimization, *Proc. 2002 American Control Conf.*, Anchorage, AK, 2002, pp. 2088–2093.

[26] A. Isidori, *Nonlinear Control Systems*, 3rd Ed., Springer-Verlag, London, 1995.

[27] A. Isidori, The zero dynamics of a nonlinear system: From the origin to the latest progresses of a long successful story, *Eur. J. Control*, 19 (2013) 369–378.

[28] A. N. Jha, A. S. Saxena, V. S. Rajmani, Parameter estimation algorithms for bilinear and non-linear systems using Walsh functions—recursive approach, *Int. J. Syst. Sci.*, 23 (1992) 283–290.

[29] J.-N. Juang, Continuous-time bilinear system identification, *Nonlinear Dynam.*, 39 (2005) 79–94.

[30] J.-N. Juang, C.-H. Lee, Continuous-time bilinear system identification using single experiment with multiple pulses, *Nonlinear Dynam.*, 69 (2012) 1009–1021.

[31] T. Kailath, *Linear Systems*, Prentice-Hall, Inc., Englewood Cliffs, NJ, 1980.

[32] R. E. Kalman, Lyapunov functions for the problem of Lur'e in automatic control, *Proc. Nat. Acad. Sci.*, 49 (1963) 201–205.

[33] R. E. Kalman, On a new characterization of linear passive systems, *Proc. First Annual Allerton Conf. on Communication, Control and Computing*, 1963, Monticello, IL, pp 456–470.

[34] N. Karmarkar, A new polynomial time algorithm for linear programming, *Combinatorica*, 4 (1984), 373–395.

[35] B. Kim, K. Ryu, J. Back, A generalized hold based countermeasure against zero-dynamics attack with application to DC-DC converter, *IEEE Access*, 10 (2022) 44923–44933.

[36] D. Kim, K. Ryu, J. H. Kim, J. Back, Zero assignment via generalized sampler: A countermeasure against zero-dynamics attack, *IEEE Access*, 9 (2021) 109932–109942.

[37] Y. Li, J. Zhang, Q. Wu, *Adaptive sliding mode neural network control for nonlinear systems*, 1st Ed., Academic Press, London, 2018.

[38] C. Liu, Y. Shih, Analysis and parameters estimation of bilinear systems via Chebyshev polynomials, *J. Frank. Inst.*, 317 (1984) 373–382.

[39] L. Ljung, *System Identification: Theory for the User*, 2nd Ed., Prentice Hall, Englewood Cliffs, NJ, 1987.

[40] L. Ljung, Perspectives on system identification, *IFAC Proc. Volumes*, 41 (2008) 7172–7184.

[41] Y. Mao, E. Akyol, Z Zhang, A novel defense stategy against zero-dynamics attacks in multi-agent systems, *Proc. 58th IEEE Conf. on Decision and Control*, Nice, France, 2019, pp. 3563–3568.

[42] Y. Mao, E. Akyol, Detectability of cooperative Zero-dynamics attack, *Proc. 56th Annual Allerton Conf. on Communication, Control, and Computing*, Allerton Park, IL, 2018, pp. 227–234.

[43] Y. Mao, H. Jafarnejadsani, P. Zhao, E. Akyol, N. Hovakimyan, Detectability of intermittent zero-dynamics attack in networked control systems, *Proc. 58th IEEE Conf. on Decision and Control*, Nice, France, 2019, pp. 5605–5610.

[44] R. R. Mohler, *Bilinear Control Processes: with Applications to Engineering, Ecology, and Medicine*, 1st Ed., Academic Press, New York and London, 1973.

[45] R. R. Mohler, Bilinear techniques and petrochemical applications, *IFAC Proc. Automatic Control in Petrolium, Petrolium and Desalination Industries*, Kuwait, 1986, vol. 19, pp. 33–36.

[46] *Mosek Modelling Cookbook*, Release 3.3.0, MOSEK Aps, 2024, available at https://www.mosek.com/documentation.

[47] A. Padoan, A. Astolfi, Nonlinear system identifcation for autonomous systems via functional equations methods, *Proc. 2016 American Control Conf.*, 2016, Boston, MA, USA, pp. 1814–1819.

[48] P. M. Pardalos, V. Yatsenko, *Optimization and Control of Bilinear Systems: Theory, Algorithm, and Applications*, Springer, New York, 2008.

[49] M. N. O. Sadiku, Y. Wang, S. Cui, S. M. Musa, Cyber-physical systems: A literature review, *Eur. Sci. J.*, 36 (2017) 52–58.

[50] J. Schoukens, L. Ljung, Nonlinear system identification: A user-oriented road map, *IEEE Control Syst. Mag.*, 39 (2019) 28–99.

[51] Y. I. Son, J. H. Seo, Observer design for bilinear systems with unknown inputs, *Proc. of the KIEE Conf.*, 1996, pp. 927–929.

[52] H. J. Sussmann, Minimal realizations and canonical forms for bilinear systems, *J. Frank. Inst.*, 301 (1976) 593–604.

[53] H. J. Sussmann, Semigroup representations, bilinear approximation of input-output maps, and generalized inputs, *Mathematical Systems Theory*, Springer Berlin Heidelberg, Berlin, Heidelberg, 1976.

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

[55] Z. Wang, H. Zhang, X. Cao, E. Liu, H. Li, J. Zhang, Modelling and detection scheme for zero-dynamics attack on wind power system, *IEEE Trans. Smart Grid*, 15 (2024) 934–943.

[56] S. Weerakkody, X. Liu, B. Sinopoli, Robust structural analysis and design of distributed control systems to prevent zero dynamics attacks, *Proc. 56th IEEE Conf. on Decision and Control*, Melbourne, Australia, 2017, pp. 1356–1361.

[57] V. A. Yakubovich, The solution of certain matrix inequlities in automatic control theory, *Soviet Math. Dokl.*, In Russian 3 (1961) 620–623.

[58] V. A. Yakubovich, Solution of certain matrix inequlities encountered in nonlinear control theory, *Soviet Math. Dokl.*, 5 (1964) 652–656.

[59] D. Ye, T. Zhang, G. Guo, Stochastic coding detection scheme in cyber-physical systems against replay attack, *Inf. Sci.*, 481 (2019) 432–444.

APPENDIX A

MATLAB CODE

A.1 MATLAB CODE FOR SIMULATING A ZERO DYNAMICS ATTACK

```
%% Example 5.1
%% Zero dynamics attack simulation to a petro-chemical plant
%% Part 1: Bilinear system identification
%% Part 2: Attack synthesis
%% MATLAB Version 2022b
clc;
clear all;
clear figure;

%% Petro-chemical plant parameters
v1=100; % Tank 1 volume
v2=200; % Tank 2 volume
c2=6;   % positive coefficient

%% State-space description of the petro-chemical plant
N0=[0 0; 0 c2/v2];
N1=[1/v1 -1/v1;-1/v2 1/v2];
N2=[1/v1 0; 0 0];
C=[1 0];
```

```
x0_initial=[4; 10];
u_fixed=-1 ;  % Adversary changes u1=-1 at the onset of ZDA

%% SISO bilinear system to be identified
Ac=N0+N1*u_fixed; % Ac is the modified state transition matrix
   (N0) for the SISO bilinear system.
Nc1= N2;          % Nc1 is the coupling matrix (N1) for the
   SISO bilinear system.
desired_sampling_time=0.1;
desired_vi=-C*Ac*x0_initial/(C*Nc1*x0_initial);  % amplitude u*
    based on SISO bilinear system
tmax1=100;

%% Part 1: Bilinear system ID Algorithm
%%  Defining bilinear ID algorithm parameters
l0=5;
l=2;
alpha=size(Ac,1)+1;
beta=size(Ac,1)+1;
p=size(Ac,1)+1;
r=1; % number of inputs
m=1; % number of outputs
delta_t=desired_sampling_time;
sample_time=desired_sampling_time;
vi=desired_vi; % input magnitude
```

```
%% Generating input sequence for bilinear system identification
uu=zeros(1,20);
uu(1)=1;
for i=1:p+5 %p+1
index=l0+(i-1)*l+1;
uu(index)=1;
end
t=0:sample_time:sample_time*(length(uu)-1) ;
u.time=t';
u.signals.values=vi*uu';

%% Generating output data for the bilinear identification
   algorithm
u_sim=u;
Ac_sim=Ac;
C_sim=C;
Nc_sim=Nc1;
x0_sim=x0_initial;
y=sim('SimulinkBlock1',t(end));
y_time=y.y.time;
yy=y.y.Data;
length(yy)
Y1=yy(2:(alpha+beta-1)+1);
length(Y1)
```

```
%% STEP 1: Buliding Hankel Matrix
H1=[]; % H1 is the Hankel matrix
row_length=alpha*m;
p1=1;
for j=1:beta
h(1:row_length,1:r)=Y1(j+p1-1:j+p1-1+row_length-1,:);
H1=[H1 h];
end
rank_HankelMatrix=rank(H1)
H1

%% STEP 2: Calculating singular value decomposition (SVD)
format long
[U1 S1 V1]=svd(H1);
rank(U1)
v1=[];
format short
for i=1:min(size(S1))
ss1=S1(i,i);
v1=[v1 ss1];
end
tol=10^-6;
vv1=find(abs(v1)>tol);
n_est1=length(vv1);          % estimated order of the system
```

```
%% STEP 3: Estimating output matrix C_est1, or \hat{C} in the
   dissertation
C_est1=U1(1:m,1:n_est1);      % estimated C.
U1_up=U1(1:end-1,1:n_est1);
U1_down=U1(2:end,1:n_est1);
A_est=pinv(U1_up)*U1_down;
eig_A_est=eig(A_est)
VV2=V1';
CM1=S1(1:n_est1,1:n_est1)*VV2(1:n_est1,1:n_est1);
x1=CM1(:,1)

%% STEP 4: Formulating two groups of ordered members
%%  Determining YY, or  \mathcal{y}  in the dissertation
for i =1:p
index=(l0+(i-1)*l+1)+1;
hh=yy(index:index+n_est1-1);
YY(:,i)=pinv(U1_up)*hh;
end
%%  Determining XX, or  \mathcal{x} in the dissertation
XX(:,1)=(A_est)^(l0-1)*x1;
for i=2:p
XX(:,i)=(A_est)^(l-1)*YY(:,i-1);
end
```

```
%% STEP 5:  Estimating discrete coupling matrix A1_est, or (\
   hat{M}_1) in the dissertation
A1_est=YY*pinv(XX);
eig(A1_est);
eig(expm(Ac+Nc1*vi));

%% STEP 6: Estimating Ac, Nc, x0 , or \hat{N}_0, \hat{N}_1, \
   hat{z}_0 in the dissertation
x0_est=inv(A1_est)*x1;
Ac_est=(1/delta_t)*logm(A_est);  % estimated Ac.
eig(Ac);
eig_Ac_est=eig(Ac_est);
Nc_est=(1/vi)*(((1/delta_t)*(1/1)*logm((A1_est)^1)-Ac_est));  %
   estimating Nc
eig_Nc=eig(Nc1);
eig_Nc_est=eig(Nc_est);

%% Determining the similarity transformation matrix
Q_est=[C_est1;C_est1*Ac_est];
Q=[C;C*Ac];
Phi=pinv(Q)*Q_est;

%% Matching the estimated matrices with the original matrices
   through similarity transformation
Ac
```

```
Ac_est2=(Phi)*Ac_est*inv(Phi)

Nc1

Nc_est2=(Phi)*Nc_est*inv(Phi)

C

C_est2=C_est1*inv(Phi)

x0_initial

x0_est2=(Phi)*x0_est

format long

Ac_est

Nc_est

C_est1

x0_est


%% Part 2: Zero dynamics attack simulation using identified
   system
% actual plant

Ac_sim=Ac;

C_sim=C;

Nc_sim=Nc1;

x0_sim=x0_initial;


% Identified plant

Ac_sim_id=Ac_est;

C_sim_id=C_est1;

Nc_sim_id=Nc_est;
```

```
x0_sim_id=x0_est;

gain_num=-C_est1*Ac_est;

gain_den=C_est1*Nc_est;


%% Observer-based zero dynamics attack simulation

t=tmax1;

simulation=sim('SimulinkBlock3',t(end));

y_time=simulation.y.time;

yy4=simulation.states.Data;

yo4=simulation.y.Data;

u2=simulation.input_id.Data;

yo4_bound=C_sim*[x0_sim*01+x0_sim];

index4=max(find(yo4<=yo4_bound));


%% purging data to plot

index=index4

index_state=index%tmax_state/sample_time;

y_time_state_plot=y_time(1:index_state);

y_time_plot=y_time(1:index);

% yo3_plot=yo3(1:index);

% yy3_plot=yy3(1:index_state,:);

yo4_plot=yo4(1:index);

yy4_plot=yy4(1:index_state,:);

u2_plot=u2(1:index);
```

```
%% Analytical zero dynamics attack
syms s
partfrac((0.0025*s^2-1.5011*s-0.0301)/(s*(s-0.025)),s,'
   FactorMode','complex')
t5=(0:sample_time:tmax1)';%y_time_state_plot;
length(t5)
exponential_input1=1.204-2.7050375*exp(0.025.*t5);
exponential_input=[t5 exponential_input1];
simulation=sim('SimulinkBlock2',t5(end))
y_time=simulation.y.time;
yy5=simulation.states.Data;
yo5=simulation.y.Data;
u5=simulation.exp_input.Data;
yo5_bound=C_sim*[x0_sim*10+x0_sim];
index5=max(find(yo5<=yo5_bound));
u5_plot=u5(1:index5);

%% purging data to plot
axis_scaling=40;
index=min(index5,index4)
index_state=index%tmax_state/sample_time;
y_time_state_plot=y_time(1:index_state);
y_time_plot=y_time(1:index);
yo5_plot=yo5(1:index);
yy5_plot=yy5(1:index_state,:);
```

```
u5_plot=u5(1:index);


%% purging data to plot
axis_scaling=40;
index=min(index5,index4)
index_state=index%tmax_state/sample_time;
y_time_state_plot=y_time(1:index_state);
y_time_plot=y_time(1:index);
yo5_plot=yo5(1:index);
yy5_plot=yy5(1:index_state,:);
u5_plot=u5(1:index);


%% Output plot
figure
plot(y_time_plot,yo5_plot,'k--','LineWidth',2)
hold on
plot(y_time_plot,yo4_plot,'r-','LineWidth',2)
hold off
grid on
legend('Using $u_2^*$ from analytical approach', 'Using $u_2^*$
    from observer-based approach','Interpreter','latex','
   Location','southwest')
xlabel('t (sec.)','Interpreter','latex');
ylabel('output ($y$)', 'Interpreter','latex')
```

```
title('Outputs $y$ when the plant is under a zero dynamics
   attack','Interpreter','latex')

%% Attack intput plot
figure
plot(y_time_plot,u5_plot,'k--','LineWidth',2.3)
hold on
plot(y_time_plot,u2_plot,'r-','LineWidth',1.5)
hold off
grid on
legend('$u_2^*$ from analytical approach', '$u_2^*$ from
   observer-based approach','Location','southwest','Interpreter
   ','latex')
xlabel('t (sec.)','Interpreter','latex');
ylabel('Input ($u_2^*$)','Interpreter','latex')
title('Attack inputs $u_2^*$ applied to the plant','Interpreter
   ','latex' )



%% states trajectory plot
figure
subplot(211)
plot(y_time_state_plot,yy5_plot(:,1),'k-','LineWidth',2)
hold on
```

```
plot(y_time_state_plot,yy5_plot(:,2),'r-','LineWidth',2)
hold off
axis([0,max(y_time_plot), -axis_scaling*max(yy5_plot(:,1)),max(
   yy5_plot(:,2))])
grid on
legend('$z_1$', '$z_2$','Location','southeast','Interpreter','
   latex')
xlabel('t (sec.)','Interpreter','latex');
ylabel('states','Interpreter','latex')
subplot(212)
plot(y_time_state_plot,yy4_plot(:,1),'k-','LineWidth',2)
hold on
plot(y_time_state_plot,yy4_plot(:,2),'r-','LineWidth',2)
hold off
grid on
axis([0,max(y_time_plot), -axis_scaling*max(yy4_plot(:,1)),max(
   yy4_plot(:,2))])
legend('$z_1$', '$z_2$','Location','southeast','Interpreter','
   latex')
xlabel('t (sec.)','Interpreter','latex');
ylabel('states','Interpreter','latex')
sgtitle('State trajectories of the system for the analytical
   attack (top) and the observerbased attack (bottom)','
   Interpreter','latex')
```

A.1.1 Associated simulink block diagram for Example 5.1

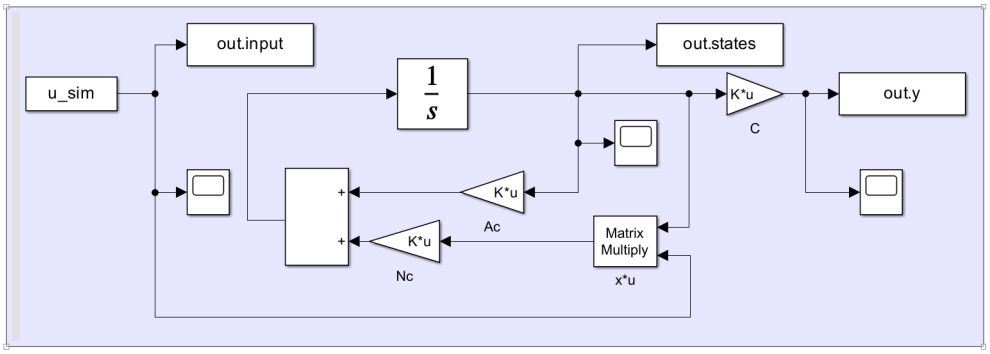

Figure 12: Simulink diagram for generating input-output data for the bilinear system identification algorithm ("*SimulinkBlock*1" in A.1 MATLAB code)

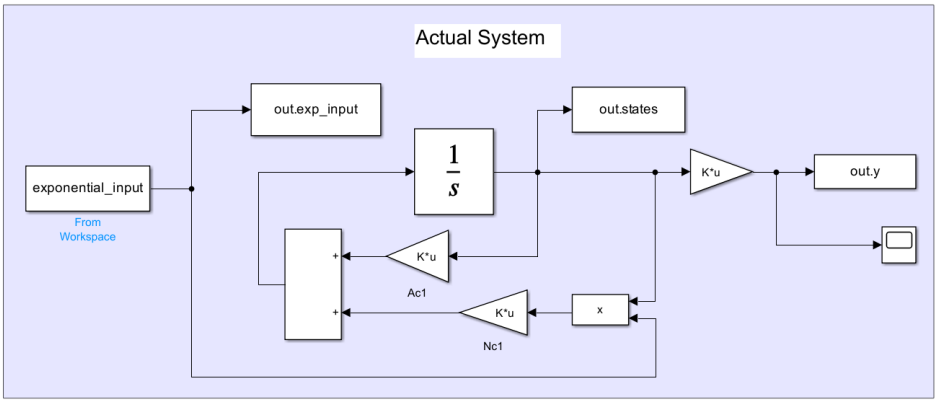

Figure 13: Simulink diagram for the analytical attack ("*SimulinkBlock*2" in A.1 MATLAB code)

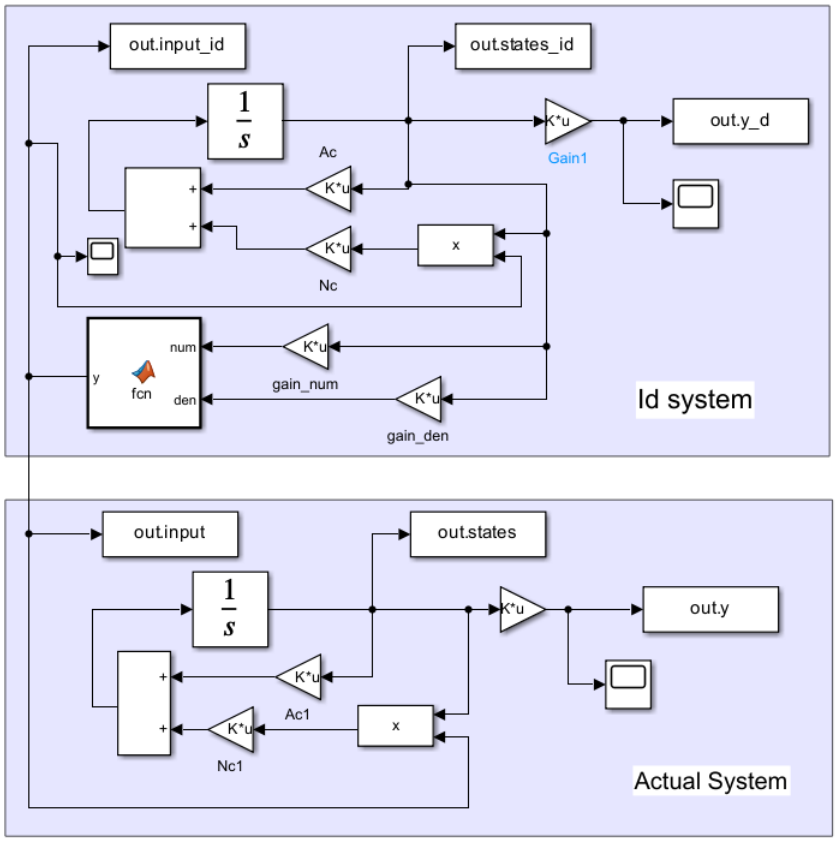

Figure 14: Simulink diagram for the observer-based attack ("*SimulinkBlock3*" in A.2 MATLAB code)

A.2 MATLAB CODE FOR DESIGNING AN ATTACK DETECTOR

```
%% Example 5.2
%% THE MATLAB CODE FOR DETERMINING THE OBSERVER GAINS USING
   SEMIDEFINITE PROGRAMING.
%% SOFTWARE PACKAGES: MATLAB R2022b
%%                    MOSEK VERSION 10.1.21
%% PC ARCHITECTURE    PCWIN64
%% EXAMPLE 5_2
clc;
```

```
clear all;
clf;

%% Petro-chemical plant parameters
v1=100;
v2=200;
c2=6;

%% State space matrices of the petro-chemical plant
N0=[0 0; 0 c2/v2];
N1=[1/v1 -1/v1;-1/v2 1/v2];
N2=[1/v1 0; 0 0];
C=[1 0];
x0=[4; 10];
u1_fixed=-1;
n=size(x0,1);

%% Semidefinite programming (SDP) variables
P=sdpvar(n); % PSD MATRIX
S0=sdpvar(n,1);
S1=sdpvar(n,1);
S2=sdpvar(n,1);

%% Constraints for the SDP
F0=N0'*P-C'*S0'+P*N0-S0*C;
```

```
F1=(N1'*P-C'*S1'+P*N1-S1*C)*u1_fixed;
F2=N2'*P-C'*S2'+P*N2-S2*C;
F=[P>=0, F0<=0, F1==0, F2==0, P<=0.01*eye(n)];

%% Solving the SDP
ops = sdpsettings('solver','mosek','verbose',1,'debug',1);
ops = sdpsettings('mosek.MSK_IPAR_BI_MAX_ITERATIONS',2342342);
optimize(F);

format long
Pfeasible=value(P);
S0feasible=value(S0);
S1feasible=value(S1);
S2feasible=value(S2);
rcond(Pfeasible);

%% Observer gains
P1=Pfeasible;
L0=pinv(P1)*S0feasible;
L1=pinv(P1)*S1feasible;
L2=pinv(P1)*S2feasible;

%% Observer performance when there is no attack to the plant
sampleTime=0.01;
Tmax=1%3;
```

```
t=0:sampleTime:Tmax;
x0_initial_observer=[0;0];
%% when at equilibrium
u_e1_sim=[t' -10*ones(length(t),1)];
u_e2_sim=[t' -15*ones(length(t),1)];
x0_initial_plant=[4;10];
simulation=sim('SimulinkBlock4', Tmax);
x_plant=simulation.plant_states.Data;
x_observer=simulation.observer_states.Data;
figure
subplot(211)
plot(t,x_plant(:,1),'-k','LineWidth',1.5)
hold on
plot(t, x_observer(:,1),'--r','LineWidth',1.5)
hold off
axis([0, Tmax, 0, max(x_observer(:,1))+1])
xlabel('time (sec.)','Interpreter','latex')
ylabel('state, $z_1$','Interpreter','latex')
legend('plant state', 'observer state','Interpreter','latex',
   Location='southeast')

subplot(212)
plot(t,x_plant(:,2),'-k','LineWidth',1.5)
hold on
plot(t, x_observer(:,2),'--r','LineWidth',1.5)
```

```
hold off
axis([0, Tmax, -500, max(x_observer(:,2))+500])
xlabel('time (sec.)','Interpreter','latex')
ylabel('state, $z_2$','Interpreter','latex')
legend('plant state', 'observer state','Interpreter','latex',
   Location='northeast')
sgtitle('Observer tracking the plant states $z_1$ (top) and $
   z_2$ (bottom) under no attack','Interpreter','latex')


%% Observer perfomance when the plant is under attack
sampleTime=0.01;
Tmax1=30;
Tmax2=250
t1=0:sampleTime:Tmax1;
t2=Tmax1+sampleTime:sampleTime:Tmax2;
t=[t1 t2];
x0_initial_observer=[0;0];

u_e1_sim=[t1' -10*ones(length(t1),1); %% when at equilibrium
          t2'  -1*ones(length(t2),1)]; %% when under attack

u_e2_sim=[t1' -15*ones(length(t1),1);          %% when at
   equilibrium
```

```
            t2'   1.204-2.7050375*exp(0.025.*(t2-Tmax1+sampleTime)
               ')]; %% when under attack


x0_initial_plant=[4;10];
simulation=sim('SimulinkBlock4', Tmax2);
x_plant=simulation.plant_states.Data;
x_observer=simulation.observer_states.Data;
figure
subplot(211)
plot(t,x_plant(:,1),'-k','LineWidth',1.5)
hold on
plot(t, x_observer(:,1),'--r','LineWidth',1.5)
hold off
axis([-2, 250, 0, max(x_plant(:,1))+1])
xlabel('time (sec.)','Interpreter','latex')
ylabel('state, $z_1$','Interpreter','latex')
legend('plant state', 'observer state','Interpreter','latex',
   Location='southeast')


subplot(212)
plot(t,x_plant(:,2),'-k','LineWidth',1.5)
hold on
plot(t, x_observer(:,2),'--r','LineWidth',1.5)
hold off
axis([-2, 250, -400, max(x_plant(:,2))+1])
```

```
xlabel('time (sec.)','Interpreter','latex')
ylabel('state, $z_2$','Interpreter','latex')
legend('plant state', 'observer state','Interpreter','latex',
   Location='northwest')
sgtitle('Observer tracking the plant states $z_1$ (top) and $
   z_2$ (bottom) when attack is initiated at $t = 30$ seconds',
   'Interpreter','latex')


%% Worst case: When the observer and attack are initiated
   simultaniously
sampleTime=0.01;
Tmax=120;
t=0:sampleTime:Tmax;
x0_initial_observer=[0;0];
u_e1_sim=[t' -1*ones(length(t),1)];
u_e2_sim=[t' 1.204-2.7050375*exp(0.025.*t')];
x0_initial_plant=[4;10];
simulation=sim('SimulinkBlock4', Tmax);
x_plant=simulation.plant_states.Data;
x_observer=simulation.observer_states.Data;
figure
subplot(211)
plot(t,x_plant(:,1),'-k','LineWidth',1.5)
hold on
```

```
plot(t, x_observer(:,1),'--r','LineWidth',1.5)
hold off
axis([-2, 120, 0, max(x_observer(:,1))+1])
xlabel('time (sec.)','Interpreter','latex')
ylabel('state, $z_1$','Interpreter','latex')
legend('plant state', 'observer state','Interpreter','latex',
   Location='southeast')

subplot(212)
plot(t,x_plant(:,2),'-k','LineWidth',1.5)
hold on
plot(t, x_observer(:,2),'--r','LineWidth',1.5)
hold off
axis([-2, 120, 0, max(x_observer(:,2))+50])
xlabel('time (sec.)','Interpreter','latex')
ylabel('state, $z_2$','Interpreter','latex')
legend('plant state', 'observer state','Interpreter','latex',
   Location='northeast')
sgtitle('Observer tracking the plant states $z_1$ (top) and $
   z_2$ (bottom) when attack and observer are initiated at the
   same moment (worst case)','Interpreter','latex')
```

A.2.1 Associated Simulink block diagram for Example 5.2

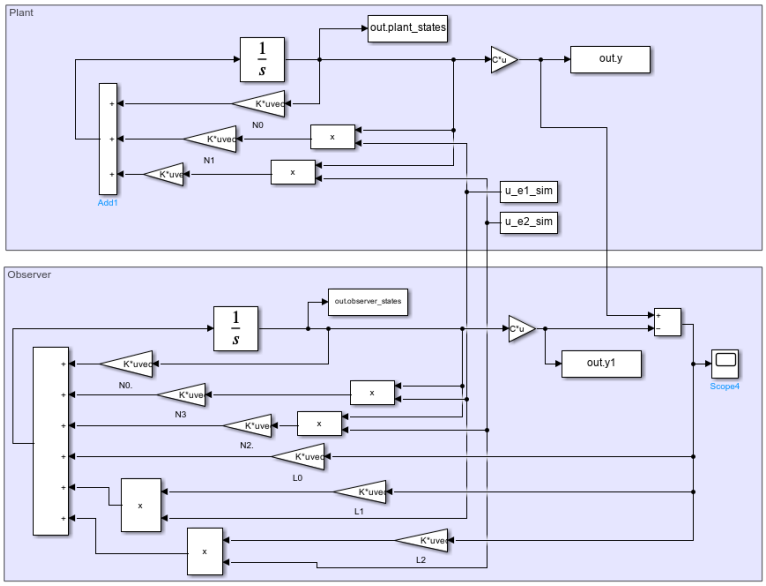

Figure 15: Simulink diagram for observer-based zero dynamics attack detector ("*SimulinkBlock*4" in A.2 MATLAB code)

VITA

Mohammad Aminul Haq
Department of Electrical and Computer Engineering
Old Dominion University
Norfolk, VA 23529

Education

- M.Sc. Control and Automation Engineering, Yildiz Technical University, Istanbul, Turkey, 2014-2016.
- B.Sc. Electrical and Electronic Engineering, Rajshahi University of Engineering and Technology, Rajshahi, Bangladesh, 2004-2009.

Ph.D. Dissertation

Zero Dynamics Attack on Unknown Bilinear Systems: Vulnerability and Detection, Old Dominion University, August 2024.
Advisor: W. S. Gray

Recent Publications

1. M. A. Haq and W. S. Gray, "Zero Dynamics Attacks on Unknown Bilinear Systems: Case Study", Proc. 7th IEEE International Conference on Industrial Cyber-Physical Systems, St. Louis, USA, 2024.
2. W. S. Gray, L. A. Duffaut Espinosa, and M. A. Haq, "Estimating Relative Degree of Nonlinear Systems Using Generating Series", Proc. 61st Conference on Decision and Control, Cancun, Mexico, 2022, pp 7333–7338.

Academic Positions

Old Dominion University, Department of Electrical and Computer Engineering, Norfolk, VA

- Teaching Assistant - Introduction to Discrete-Time Signal Processing (ECE 381), Linear System Analysis (ECE 302), Automatic Control System (ECE 695).

Awards and Honors

- Research grant, Center for Research and Training (CRT), Uttara University, 2019.
- Turkiye Burslari Scholarship, Turkey, 2012-2016.

Typeset using LaTeX.