Characterization of Optimal Cyber Attacks on Control Systems

A Dissertation Presented for the

Doctor of Philosophy

Degree

The University of Tennessee, Knoxville

Anis Drira

December 2015

Copyright © 2015 by Anis Drira.

All rights reserved.

To my Mother, Father, and Wife

Nabiha, Noureddine, and Leila

Acknowledgments

Praise be to The Almighty, the lord of the heavens and lord of the earth Allah (God), without his will and help this work may not have come to an end.

I would like to express my profound appreciation, great gratitude, and special thanks to my advisor, Dr. Seddik M. Djouadi for his willingness to support my work, his inspiring guidance throughout my studies, valuable suggestions, constructive criticism, and persistent help which allowed me to develop my skills as a researcher and learned from him how to build a solid mathematical background. I thank him for that opportunity and all what I learned from him including a good moral ethics and to be a good engineer, a modest person, and especially a man of principles. I am certain that all what I have learned form him will continue to benefit me in my entire life.

I additionally would like to thank my committee member, Dr. Mongi Abidi. His valuable suggestions, advices, and brilliant ideas has allowed me to develop more skills as a researcher. I would further like to thank the other members of my committee: Dr. Husheng Li, and Dr. Suzanne Lenhart. I greatly appreciate their time and input to this dissertation.

Within the Lab, I owe many thanks to colleagues and friends, I express my sincere gratitude for the many conversations that have had a tremendous impact on my research and myself as a person. Samir Sahyoun, Mohamed Olama, Jin Dong, Xiao Ma, Ehsan Raoufet,Yichen Zhang, and Wassim Bara. Thank you so much.

I would like to thank my lovely wife Leila, beloved mother Nabiha, father Noureddine; dear sisters Sonia, Lilia, and Meriam, parents in-law Naima and Mohamed; brothers in-law Mohamed, Sofien, Menef, and Walid; nephews Mahmoud, and Louai; my nieces Nour, Zeineb, and Nouran. Without their encouragement, understanding, love, and support this work would not have been

possible. The motivation, patience, understanding, and endurance of my family had contributed substantially to my success.

I would like to express my appreciation to the many friends outside of my studies who have helped me through difficult times: Bilel Hadri, Hatem El Dakhakni, Watheq Al Basheer, Michele Materni, Dima Orlov, Mohamed Amine Ben Ayara, Nader Khaldi, Mohanad Allush, and Marouane Salhi. Thank you again Bilel, Sana, Marouane, Habiba, Mohaned, and Marwa for all the wonderful times.

Last but not least, I would like to deeply thank Dr. Alexander Melin, and Mr. Roger Kistner at Oak Ridge National Lab (ORNL) for their support, the wonderful research internship, and valuable discussions and suggestions.

Abstract

Securing control systems which manage the nation's critical physical infrastructures and key resources is becoming one of the highest priorities and can no longer be ignored. Several researches and incidents have confirmed the real vulnerability of these systems. Due to the vital role of control systems, the study and the development of advanced strategies to overcome these cyber-security breaches are very important. Control systems are mainly composed of sensors and actuators, which cyber-security is vital for the security of the whole control system.

Characterizing and guarding against worst case attack scenarios is important to secure control systems. This dissertation proposes to study worst case optimal attacks for various estimation and control algorithms. In particular, optimal attack signals on the ever popular least squares and weighted least squares algorithms are explicitly computed. Optimal strategies for bounded and finite energy sensor and actuator attack signals are characterized using functional analytic tools. Moreover, optimal random signal attacks on the Kalman filter are characterized. In particular, a worst case probability distribution is provided.

A new optimal attack proof estimator is derived. This estimator mitigates optimal attack strategies on the Kalman filter. Illustrative numerical studies are provided to support our methods for analyzing the worst case attack signals, including optimal attack' scenarios on a power grid network, and a radar tracking system, showing the validity of the results obtained and also the effectiveness with real applications.

Table of Contents

1 Introduction **1**

1.1 Overview 1

1.2 Literature Review 2

1.3 Main Contributions 5

1.4 Dissertation Outline 5

2 Optimal Sensor Attack for Least Squares Algorithms **7**

2.1 Optimal Sensor Attack for Least Squares 7

2.2 Optimal Sensor Attack for Weighted Least Squares 14

2.3 Optimal Sensor Attack for Recursive Least Squares 19

3 Optimal Sensor and Actuator Attacks on Finite Horizon Linear Quadratic (LQ) Control **26**

3.1 Optimal Sensor Attack Finite Energy 28

3.2 Optimal Bounded Sensor Attack Signal 35

3.3 Optimal Actuator Attack Finite Energy 39

3.4 Optimal Bounded Actuator Signal Attack 44

4 Optimal Attack Strategies on The Kalman Filter **48**

4.1 Optimal Attack on Kalman Filter Scalar Case 50

4.1.1 Optimal Attack with Δ_{y_k} independent of the Estimation Error 51

4.1.2 Optimal Attack with Δ_{y_k} dependent on the Estimation Error 54

4.1.3 Mitigation with Δ_{y_k} dependent on the Estimation Error 64

4.2 Optimal Attack on Kalman Filter Multivariable Case 70

5 Conclusions and Future Work **74**

Bibliography **76**

Vita **85**

List of Figures

2.1 Optimal Attack for Least Squares 13

2.2 Optimal Attack Signal 14

3.1 Finite Energy Sensor Attack Case: No Attack. 32

3.2 Finite Energy Sensor Attack Case: Attack Signal. 32

3.3 Finite Energy Sensor Attack Case: Optimal Attack. 33

3.4 Finite Energy Sensor Attack Case: Difference Optimal Attack and No Attack. 33

3.5 Finite Energy Sensor Attack Case: Bias Attack. 34

3.6 Finite Energy Sensor Attack Case: Difference Bias Attack and No Attack. 34

3.7 Bounded Sensor Attack Signal Case: No Attack. 36

3.8 Bounded Sensor Attack Signal Case: Attack Signal. 36

3.9 Bounded Sensor Attack Signal Case: Optimal Attack. 37

3.10 Bounded Sensor Attack Signal Case: Difference Optimal Attack and No Attack. . . 37

3.11 Bounded Sensor Attack Signal Case: Bias Attack. 38

3.12 Bounded Sensor Attack Signal Case: Difference Bias Attack and No Attack. 38

3.13 Finite Energy Actuator Attack Case: No Attack. 40

3.14 Finite Energy Actuator Attack Case: Attack Signal. 41

3.15 Finite Energy Actuator Attack Case: Optimal Attack. 41

3.16 Finite Energy Actuator Attack Case: Difference Optimal Attack and No Attack. . . 42

3.17 Finite Energy Actuator Attack Case: Bias Attack. 42

3.18 Finite Energy Actuator Attack Case: Difference Bias Attack and No Attack. 43

3.19 Bounded Actuator Signal Attack Case: No Attack. 45

3.20 Bounded Actuator Signal Attack Case: Attack Signal. 45

3.21 Bounded Actuator Signal Attack Case: Optimal Attack. 46
3.22 Bounded Actuator Signal Attack Case: Difference Optimal Attack and No Attack. . 46
3.23 Bounded Actuator Signal Attack Case: Bias Attack. 47
3.24 Bounded Actuator Signal Attack Case: Difference Bias Attack and No Attack. . . . 47

4.1 Optimal Attack Signal on the Kalman Filter 62
4.2 Optimal Attack on the Kalman Filter 63
4.3 Optimal Attack and Random Attack on Kalman Filter 64
4.4 Radar Based Target Tracking using Kalman Filter 68
4.5 Optimal Attack Signal for Mitigation Strategy 68
4.6 Radar Based Target Tracking: Optimal Attack on Kalman Filter 69
4.7 Mitigation Strategy on the Kalman Filter 69

Chapter 1

Introduction

1.1 Overview

Control systems are computer-based systems. Depending on the application, these control systems are also called Supervisory Control and Data Acquisition (SCADA), Process Control System (PCS), Distributed Control System (DCS), Cyber-Physical Systems (CPS), etc. The purpose of control systems is to control, monitor, and manage physical processes which represent a wide variety of networked information technology systems connected to the nations critical physical world infrastructures and key resources. Control systems are mainly composed of a set of networked agents, consisting of actuators, sensors, communication devices, and control processing units [1].

Control systems perform a vital function for the safe operation of the nation critical infrastructures systems ranging from relatively simple systems to incredibly complex systems such as electrical power generators, subway systems, dams, telecommunication systems, weapons systems, oil and natural gas pipelines, water and waste-water treatment, transportation systems, healthcare devices, and many others. For example, the oil and gas industry uses control systems not only to manage refining operations, monitor the gas flow and pressure at the pipelines, and control the gas transmission flow, but also to gather information about a possible leak in a pipeline, then transmit the information to the main control station to be analyzed and to determine all the possible actions the control system can perform. In this example, shutting down the pipe to stop the leak could be the best action to be performed by the control system after the information about the leak is analyzed.

Several control systems listed above are categorized as SCADA systems, which leads a SCADA security is an emerging issue, which can no longer be ignored. The U.S. Department of Homeland Security recognizes the importance of control systems security education and awareness and offers the Cyber Security Procurement Language document as a means to help asset owners integrate security into their control systems security lifecycle [2]. Not all SCADA systems are vulnerable or are at risk of attacks. However, these systems manage critical infrastructure assets that are vital to nations' economy. Whether the threats are real or perceived, it is in a nations interest to provide guidance on the protection of these assets.

There are several efforts conducted by industrial and government organizations to improve the security of control systems. The electrical sector has the North American Electric Reliability Corporation (NERC) whose mission is to ensure, develop, and enforce reliability and cybersecurity standards for control systems. The National Institute of Standards and Technology (NIST) has also published a guideline for security best practices for general IT in Special Publication 800-53. Federal agencies must meet NIST SP800-53. To address the security of control systems, NIST has also published a Guide to Industrial Control System (ICS) Security [3].The Department of Energy (DOE) has led security efforts by establishing the national SCADA test bed program and by developing a 10-year outline for securing control systems in the energy sector [2]. The International Society of Automation (ISA) whose mission is to bring together industrial cyber security experts from across the globe to develop ISA standards on control systems security and industrial automation [4].

Several control systems can be classified as life-critical systems or safety-critical systems, whose failure or their malfunction can result of a loss, harm, or severe damage to the physical system, environment, and to people. Therefore, it is critical to recognize and understand the importance of these systems and their potential vulnerabilities, and the challenges in securing these systems from any type of compromises.

1.2 Literature Review

In this section we review the current literature in the area of cyber-physical security. The study of defenceless cyber-physical systems to different types of attack has gained more and more interest

over the last decade. In [5] *denial of service* and *deception attacks* against a networked control system are presented with a semi-definite programming based solution. The *denial of service* of sensor and control data occurs when there is a jamming in the communication channels or any compromising in the data communications. The *deception attack* is defined when the integrity of the sensor and control data packets has been breached. Denial of service and deception attacks have been considered in [6–21]. The compromised integrity results when a false data is received and it is believed to be the true data. In [22] *false data injection* attack against static state estimators is introduced. Then, an extension to a generalized version of the *false data injection* attack is introduced [23]. *False data injection* attack is defined when malicious measurements are injected to mislead the state estimation process without being detected by bad measurement detection techniques. In a similar manner, *Stealthy attacks* are studied in [24–26]. In [27] a resilient control problem is studied, where the control packets sent over a communication network are corrupted by human adversaries. A receding-horizon Stackelberg control law is proposed to stabilize the control system. Attack-resilient control has been considered in [28–30]. In [31] the effect of *covert attacks* against networked control systems is investigated. The method consists of a parameterized decoupling structure which allows the covert agent to make modifications to the physical plant while remaining undetectable from the original networked controller. In [32, 33] state-estimation of a linear system with corrupted sensor measurements are studied, in which the tolerated number of attacked sensors are characterized then a decoding algorithm is proposed to recover the state of the system.

Bad measurements detection and identification have been considered in [34–48]. The generalized form is based on the small measurement error that is typically tolerated that can allow the attacker to increase the injection of the false data without being detected. In [49, 50] the effect of *replay attacks* on a control system is analyzed. The *replay attacks* is when the sensors are hijacked by the attacker, then observe, record the reading for an interval of time and repeat them afterward while attacking by injecting an exogenous control input. Considerable attention has been devoted to security issues related to cyber-physical systems in many areas such as power networks [51–58],water networks [28, 59–61], and linear network with misbehaving components [30, 62–64]. The technology of control system has being evolving over the past few decades which makes the control systems more and more vulnerable now than before. The security of these systems was not a

priority since they were traditionally stand-alone devices with logic electromechanical relays not connected to the outside world or to a network. Most efforts for protections at these stages were only focused and limited to random failure [1]. However, after these traditional controllers have been switched to microprocessors and embedded operating systems with multiple hardware and software connections and their usage over the internet and their communication and networking capabilities. Consequently, control systems inherited the vulnerability of these components. Most of these components are an open source to their protocols to public which makes it easier to the attacker to gain vital information about the system to attack and an easier access to it. However, with the new era emerged for control systems when the control systems start becoming a target to several type of attacks including but not limited to cyberattacks [2, 65–67].

In the begining of this section, we discussed the major reasons behind which the current control systems are now more vulnerable than before. In what follows; we discuss some specific incidents showing that the threat of the control systems is real. The most well-known security breach in SCADA systems is the attack on Maroochy Shire Councils sewage control system in Queensland, Australia. In the Spring of 2000, a former employee of an Australian organization that develops manufacturing software applied for a job with a local government, but he was rejected. Over a 2-month period, he continuously used a radio transmitter on as many as 46 occasions to remotely alter electronic data for particular sewage pumping stations and caused malfunctions in their control systems, ultimately releasing about 264,000 gallons of raw sewage into rivers and parks [68]. The response to the attack was very slow as the sewage operators thought it was a leak in the pipes and did not consider the possibility of an attack to their system, only after months that they discovered that spoofed controllers were activating the valves. In August 2006, two Los Angeles city employees hacked into computers controlling the city's traffic lights and disturbed signal lights causing backups and delays. In 2000, the Interior Ministry of Russia reported that hackers had seized temporary control of the system regulating gas flows in natural gas pipelines [69]. In 2008, a teenager in Poland used a modified TV remote control to control the switch tracks of trams. There were four derailments and twelve resultant injuries [70]. In June 2010, a cyber worm Stuxnet striked the Iranian nuclear facility and therefore it has demonstrated the threats that control systems are subject to sophisticated computer-based attacks to critical infrastructures, and that these attacks are not just speculations [71, 72].

There are many other reported targeted attacks [73–76]. Beside the several reported control systems attacks, there have been several studies to show the vulnerability of control systems. On March 2007, researchers at Idaho National Laboratories investigated the results of a possible cyber attack directed against a power network. The Aurora Generator Test demonstrated the ability of a cyber attack to damage a power generator turbine [77]. Other reports show how easy it is to obtain access to computers controlling physical infrastructures, [78, 79]. Also, new security audits are starting to reveal the vulnerability of major critical infrastructures. In a recent security audit, the Tennessee Valley Authority (TVA), the nations largest public power company, was found to be vulnerable to cyber attacks that could sabotage their control systems [1].

1.3 Main Contributions

The main contributions of our research can be described as follows:

1. Development of the optimal attack strategies for Least Square algorithm.
2. Development of the optimal attack strategies for Weighted Least Square algorithm.
3. Development of the optimal sensor attack strategies for Recursive Least Square algorithm in which the attacker aims at making the estimation error as large as possible.
4. Study the effect of finite energy and bounded optimal sensors and actuators attacks on Linear Quadratic (LQ) control systems.
5. Development of optimal sensor attack strategies on the Kalman filter.
6. Development of new optimal attack resistant estimator which mitigate optimal sensor attack strategies on the Kalman filter.

1.4 Dissertation Outline

This dissertation is structured as follows:

In Chapter 2, we present the derivation of the optimal sensor attack for the Least Squares, Weighted Least Squares, and Recursive Least Squares algorithms. Many control systems use least

square algorithms to estimate the states and identify plant parameters. Therefore, characterizing optimal attacks on these algorithms is of paramount importance. First, Least Squares and Weighted Least Squares are characterized and generalized to Recursive Least Squares with the view to apply the results later to the Kalman filter.

Chapter 3 presents optimal cyber-attacks on sensors and actuators. The results are anaylized and simulated for the finite horizon for linear quadratic control. We assume that the attacker is capable of spoofing sensor and actuator signals with time-varying signals. Both sensor and actuator are considered for finite energy and bounded signal attacks.

Chapter 4 presents the optimal sensor attack strategies on Kalman filter in which the attacker aims at making the estimation error as large as possible. The attacker needs to determine the probability distribution which deteriorates the estimation cost function the most. In addition, two main propositions are given considering whether the attack signal depends on the estimator error or not. We will start from the relatively easy situation, where signal attack is **independent** of the estimation error. Then we will continue to work on the general case, where the attack signal is **dependent** of the estimation error at time k. The former scenario corresponds to the case where the hacker does not have access to the estimation error of the Kalman filter, the latter case is where he has access to the estimation error. After both of the scalar cases are presented with illustrative examples, the multivariable case is presented as well. Finally, the mitigation for the scalar case for the main proposition is given considering the estimation error depends on the attack signal and a numerical example to show the effectiveness of the mitigation algorithm to counter attack the optimal sensor attack on Kalman filter.

Finally, Chapter 5 provides the conclusions and future work to be preformed which include the characterization and the implementation of optimal attacks and their mitigation on more control algorithms.

Chapter 2

Optimal Sensor Attack for Least Squares Algorithms

2.1 Optimal Sensor Attack for Least Squares

In this chapter we derive and analyze the optimal sensor attack for least squares algorithms (Least Squares, Weighted Least Squares, and Recursive Least Squares algorithms). Many control systems use least squares algorithms to estimate the states, and identify plant parameters. Therefore, characterizing optimal attacks on these algorithms is of paramount importance. First, the least squares optimal is characterized, and generalized to recursive least squares with the view to apply the results to the Kalman filter.

x is an unknown n-element vector, and y is a k-element noisy measurement vector. The goal is to find the best estimation $\hat{x}$ of x. The following matrix form represents the dependence of measurements on the unknown vector x [80]:

$$y = Hx + \upsilon \tag{2.1}$$

where υ is the measurement noise.

The measurement residual is the difference between the noisy measurements and the vector $H\hat{x}$ and is defined by [80]:

$$\varepsilon_y = y - H\hat{x} \tag{2.2}$$

The estimate $\hat{x}$ is determined based on minimizing the following cost function J:

$$\begin{aligned} J &= \varepsilon_{y_1^2} + \cdots + \varepsilon_{y_k^2} \\ &= {\varepsilon_y}^T \varepsilon_y \end{aligned} \tag{2.3}$$

By substituting ε_y in equation (2.3), the cost function J can be written as:

$$\begin{aligned} J &= (y - H\hat{x})^T (y - H\hat{x}) \\ &= y^T y - y^T H\hat{x} - \hat{x}^T H^T y + \hat{x}^T H^T H\hat{x} \end{aligned} \tag{2.4}$$

To minimize the cost function equation (2.4) with respect to $\hat{x}$, a partial derivative of J is taken then set equal to zero [80]:

$$\frac{\partial J}{\partial \hat{x}} = -y^T H - y^T H + 2\hat{x}^T H^T H = 0 \tag{2.5}$$

Thus we have:

$$\begin{aligned} \frac{\partial J}{\partial \hat{x}} &= -y^T H - y^T H + 2\hat{x}^T H^T H \\ \hat{x}^T H^T H &= y^T H \\ H^T H\hat{x} &= H^T y \\ \hat{x} &= (H^T H)^{-1} H^T y \\ \hat{x} &= H^L y \end{aligned} \tag{2.6}$$

Where H^L is the left pseudo inverse of H is valid if the number of measurement k is larger than the variables n, and measurements are linearly independent.

The new measurement with the sensor attack signal Δ_y take the following new form:

$$y_\Delta = y + \Delta_y \tag{2.7}$$

We assume that the attacker can inject into measurements signals with amplitude less than a threshold M, that is, $\|\Delta_y\| \leq M$.

Substituting the y_Δ attack in equation (2.4), yields the following new cost function:

$$\begin{aligned} J_{\Delta_y} &= \left(y + \Delta_y - HH^L\left(y + \Delta_y\right)\right)^T \left(y + \Delta_y - HH^L\left(y + \Delta_y\right)\right) \\ &= \left[\left(y + \Delta_y\right)^T \underbrace{\left(I - HH^L\right)^T \left(I - HH^L\right)}_{\left(I - HH^L\right)} \left(y + \Delta_y\right)\right] \\ &= \left[\left(y + \Delta_y\right)^T \left(I - HH^L\right) \left(y + \Delta_y\right)\right] \end{aligned} \tag{2.8}$$

The attacker optimal strategy consists of making the cost function as worst as possible, that is:

$$\max_{\Delta_y} J_{\Delta_y} \tag{2.9}$$

To determine the worst signal attack Δ_y

$$J_{\Delta_y} = \left[\left(y + \Delta_y\right)^T \left(I - HH^L\right) \left(y + \Delta_y\right)\right] \tag{2.10}$$

Let $\tilde{H} = I - HH^L$. Substituting $\tilde{H}$ in equation (2.10), yields the following:

$$\begin{aligned} J_{\Delta_y} &= \left[\left(y + \Delta_y\right)^T \tilde{H} \left(y + \Delta_y\right)\right] \\ &= y^T \tilde{H} y + \Delta_y^T \tilde{H} \Delta_y + 2 y^T \tilde{H} \Delta_y \end{aligned} \tag{2.11}$$

To maximize $J_{\Delta y}$, the term $y^T \tilde{H} y$ is omitted since it does not depend on Δ_y, yields the following:

$$\begin{aligned} \Delta_y^T \tilde{H} \Delta_y + 2 y^T \tilde{H} \Delta_y &= \left(\Delta_y^T \tilde{H} + 2 y^T \tilde{H}\right) \Delta_y \\ &= \left(\tilde{H}^T \Delta_y + 2 \tilde{H}^T y\right)^T \Delta_y \end{aligned} \tag{2.12}$$

Applying the Cauchy-Shwartz inequality, equation (2.12) can be written as:

$$\begin{aligned} \Delta_y^T \tilde{H} \Delta_y + 2 y^T \tilde{H} \Delta_y = \left(\tilde{H}^T \Delta_y + 2 \tilde{H}^T y\right)^T \Delta_y &\leq \left\| \tilde{H}^T \Delta_y + 2 \tilde{H}^T y \right\| \|\Delta_y\| \\ &\leq \left\| \tilde{H}^T \Delta_y + 2 \tilde{H}^T y \right\| M \end{aligned} \tag{2.13}$$

The upper bound in (2.13) is achieved if and only if:

$$\tilde{H}^T \Delta_y + 2\tilde{H}^T y = \beta \Delta_y \tag{2.14}$$

for some scalar β.

From (2.13) and (2.14):

$$\begin{aligned} \left\| \tilde{H}^T \Delta_y + 2\tilde{H}^T y \right\| &\leq \left\| \tilde{H}^T \Delta_y + 2\tilde{H}^T y \right\| M \\ &\leq \|\beta \Delta_y\| M \\ &\leq |\beta| M^2 \end{aligned} \tag{2.15}$$

From (2.15) the optimal attack Δ_y satisfies the following expression:

$$\begin{aligned} \left(\beta I - \tilde{H}^T\right) \Delta_y &= 2\tilde{H}^T y \\ \Delta_y &= 2\left(\beta I - \tilde{H}^T\right)^{-1} \tilde{H}^T y \end{aligned} \tag{2.16}$$

For (2.15) to hold with equality, equation (2.16) should satisfied with the scalar β with the maximal magnitude $|\beta|$. That is,

$$\begin{aligned} \max |\beta| &=: |\beta_{\max}| \\ s.t. \quad M &= \left\| 2\left(\beta I - \tilde{H}^T\right)^{-1} \tilde{H}^T y \right\| \end{aligned} \tag{2.17}$$

Solving for β yields the following:

$$\left(\beta I - \tilde{H}^T\right)^{-1} \tilde{H}^T = \left(I - \frac{\tilde{H}^T}{\beta}\right)^{-1} \beta^{-1} \tilde{H}^T \tag{2.18}$$

Using series expansion for $\left(I - \frac{\tilde{H}^T}{\beta}\right)^{-1}$, equation (2.18) can be written as:

$$\begin{aligned}\left(\beta I - \tilde{H}^T\right)^{-1}\tilde{H}^T &= \left(I + \frac{\tilde{H}^T}{\beta} + \frac{\left(\tilde{H}^T\right)^2}{\beta^2} + \frac{\left(\tilde{H}^T\right)^3}{\beta^3} + \cdots\right)\beta^{-1}\tilde{H}^T \\ &= \left(\frac{\tilde{H}^T}{\beta} + \frac{\left(\tilde{H}^T\right)^2}{\beta^2} + \frac{\left(\tilde{H}^T\right)^3}{\beta^3} + \frac{\left(\tilde{H}^T\right)^4}{\beta^4} + \cdots\right) = \sum_{j=1}^{\infty}\frac{\left(\tilde{H}^T\right)^j}{\beta^j} \quad (2.19)\end{aligned}$$

To simplify equation (2.19) we notice the following:

$$\begin{aligned}\tilde{H}\tilde{H} &= \left(I - HH^L\right)\left(I - HH^L\right) \\ &= I - 2HH^L + HH^LHH^L \\ &= I - 2H\left(H^TH\right)^{-1}H^T + H\underbrace{\left(H^TH\right)^{-1}H^TH}_{I}\left(H^TH\right)^{-1}H^T \\ &= I - H\left(H^TH\right)^{-1}H^T = I - HH^L = \tilde{H} \quad (2.20)\end{aligned}$$

Therefore, we have $\tilde{H}\tilde{H} = \tilde{H}$, and it follows that:

$$\left(\tilde{H}\tilde{H}\right)^T = \tilde{H}^T = \tilde{H}^T\tilde{H}^T \quad (2.21)$$

From equation (2.21) yields the following:

$$\tilde{H}^T = \left(\tilde{H}^T\right)^2 = \left(\tilde{H}^T\right)^3 = \left(\tilde{H}^T\right)^4 = \cdots \quad (2.22)$$

Using equation (2.22), we can simplify equation (2.19) to the following:

$$\sum_{j=1}^{\infty}\frac{\left(\tilde{H}^T\right)^j}{\beta^j} = \tilde{H}^T\sum_{j=1}^{\infty}\frac{1}{\beta^j} \quad (2.23)$$

From equation (2.52) we notice the geometric series and can be simplified to the following:

$$\sum_{j=1}^{\infty}\frac{1}{\beta^j} = \frac{1}{1-\frac{1}{\beta}} = \frac{\beta}{\beta-1} \quad |\beta| > 1 \quad (2.24)$$

Therefore, equation (2.17) can be simplified to the following:

$$M = \left\| 2\frac{\beta}{\beta - 1}\tilde{H}y \right\| \tag{2.25}$$

Solving for β we have the following:

$$\begin{aligned} M &= \left\| 2\frac{\beta}{\beta - 1}\tilde{H}y \right\| \\ M &= 2\left|\frac{\beta}{\beta - 1}\right| \left\| \tilde{H}y \right\| \\ \left|\frac{\beta}{\beta - 1}\right| &= \frac{M}{2\left\| \tilde{H}y \right\|} \\ 1 - \frac{1}{\beta} &= \pm\frac{2\left\| \tilde{H}y \right\|}{M} \end{aligned} \tag{2.26}$$

Therefore, the solution for β is:

$$|\beta| = \max\left(\frac{M}{M + 2\left\| \tilde{H}y \right\|}, \frac{M}{M - 2\left\| \tilde{H}y \right\|}\right) \tag{2.27}$$

And the optimal signal attack is given by:

$$\Delta_y = 2\left(\beta_{max} I - \tilde{H}^T\right)^{-1}\tilde{H}^T y \tag{2.28}$$

Substituting $\tilde{H}$ in equation (2.56) and (2.57), yields the following:

$$|\beta| = \max\left(\frac{M}{M + 2\left\| \left(I - HH^L\right) y \right\|}, \frac{M}{M - 2\left\| \left(I - HH^L\right) y \right\|}\right) \tag{2.29}$$

Therefore the optimal attack signal corresponds to:

$$\Delta_y = 2\left(\beta_{max} I - \left(I - HH^L\right)^T\right)^{-1}\left(I - HH^L\right)^T y \tag{2.30}$$

Numerical Example

We give an example to illustrate the performance of the optimal attack for the Least Squares algorithm. The Least Squares algorithm is used to superimpose a line on a scatter plot in the current axes. For this simulation, we chose $M = 2$.

Applying the result from eq. (2.59), we could generate the optimal attack for the Least Squares algorithm as shown in Fig. 2.1 and the attack signal as shown in Fig. 2.2.

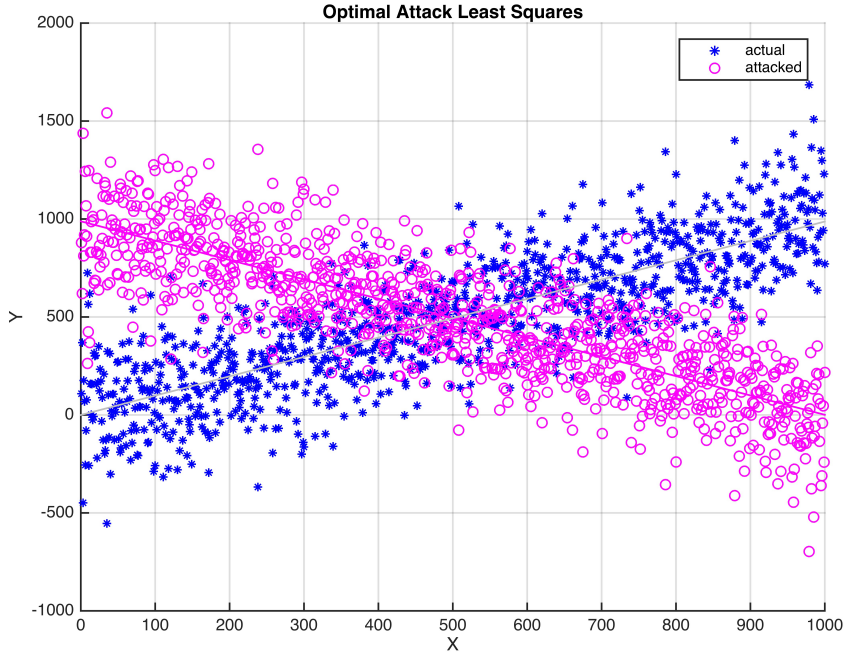


Figure 2.1: Optimal Attack for Least Squares

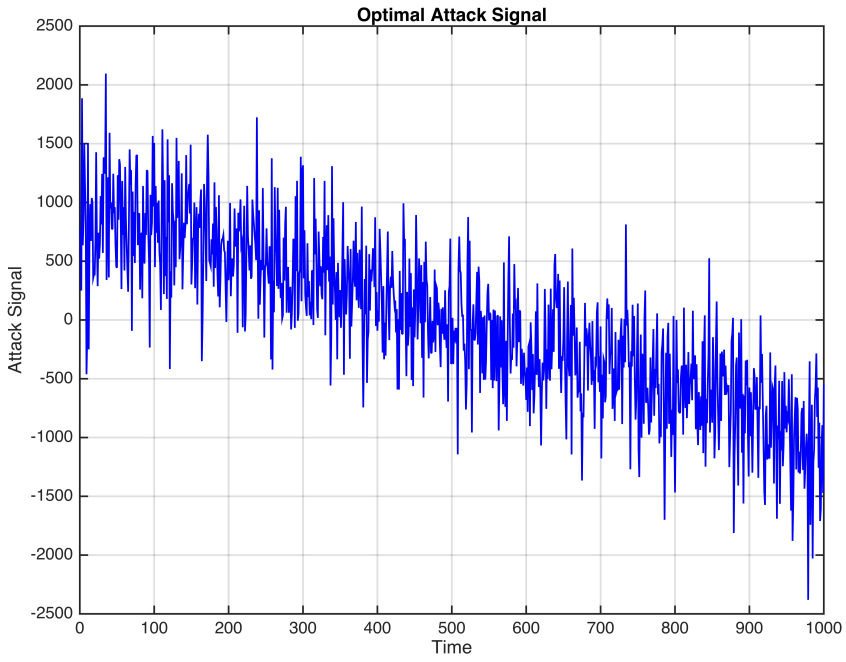


Figure 2.2: Optimal Attack Signal

Therefore, a system theoretic analysis of the effect of the optimal sensor attacks for Least Squares algorithm scalar case was presented. The characteristic equation for the Least Square algorithm is given and a numerical simulation which illustrates the effectiveness of the optimal attack technique is presented.

2.2 Optimal Sensor Attack for Weighted Least Squares

In the Least Squares section we had an equal amount of confidence in all of the measurements. In the Weighted Least Squares, the confidence in some measurements exceeds the others, therefore generalizing the Least Squares method. The variance of the measurements noise is the following [80]:

$$E(\nu_i^2) = \sigma_i^2 \qquad i = 1, 2, \cdots, k \tag{2.31}$$

where ν is the measurement noise. The assumption is that the noise for each measurement is zero-mean and independent [80].

The measurement covariance matrix is [80]:

$$R = E\left(\nu\nu^T\right) \tag{2.32}$$

The cost function for the weighted least squares can be written as [80]:

$$\begin{aligned} J &= \varepsilon_y^T R^{-1} \varepsilon_y = (y - H\hat{x})^T R^{-1} (y - H\hat{x}) \\ &= y^T R^{-1} y - \hat{x}^T H^T R^{-1} y - y^T R^{-1} H\hat{x} + \hat{x}^T H^T R^{-1} H\hat{x} \end{aligned} \tag{2.33}$$

To minimize the cost function equation (2.33) with respect to $\hat{x}$, a partial derivative of J is taken then set equal to zero [80]:

$$\begin{aligned} \frac{\partial J}{\partial \hat{x}} &= -y^T R^{-1} H + \hat{x} H^T R^{-1} H = 0 \\ H^T R^{-1} H\hat{x} &= H^T R^{-1} y \\ \hat{x} &= \left(H^T R^{-1} H\right)^{-1} H^T R^{-1} y \end{aligned} \tag{2.34}$$

By an appropriate change of variables the WLS cost function is transformed into a standard LS optimization problem. Let:

$$\bar{\Delta}_y = R^{-1/2} \Delta_y \tag{2.35}$$

$$\bar{y} = R^{-1/2} y \tag{2.36}$$

$$\bar{H} = R^{-1/2} H \tag{2.37}$$

Therefore, equations (2.33) and (2.34) can be written as the following after inserting the measurement signal corrupted with the attack signal Δ_y yields:

$$\begin{aligned} \bar{J}_{\Delta_y} &= \left(\bar{y} + \bar{\Delta}_y + \bar{H}\hat{x}\right)^T \left(\bar{y} + \bar{\Delta}_y + \bar{H}\hat{x}\right) \\ \hat{x} &= \left(\bar{H}^T \bar{H}\right)^{-1} \bar{H} \left(\bar{y} + \bar{\Delta}_y\right) \end{aligned} \tag{2.38}$$

$$
\begin{aligned}
\bar{J}_{\Delta_y} &= \left(\bar{y}+\bar{\Delta}_y+\bar{H}\left(\left(\bar{H}^T\bar{H}\right)^{-1}\bar{H}\left(\bar{y}+\bar{\Delta}_y\right)\right)\right)^T\left(\bar{y}+\bar{\Delta}_y+\bar{H}\left(\left(\bar{H}^T\bar{H}\right)^{-1}\bar{H}\left(\bar{y}+\bar{\Delta}_y\right)\right)\right) \\
&= \left(\left(I-\bar{H}\left(\bar{H}^T\bar{H}\right)^{-1}\bar{H}\right)\left(\bar{y}+\bar{\Delta}_y\right)\right)^T\left(\left(I-\bar{H}\left(\bar{H}^T\bar{H}\right)^{-1}\bar{H}\right)\left(\bar{y}+\bar{\Delta}_y\right)\right) \\
&= \left(\bar{y}+\bar{\Delta}_y\right)^T\left(I-\bar{H}\left(\bar{H}^T\bar{H}\right)^{-1}\bar{H}\right)^T\left(I-\bar{H}\left(\bar{H}^T\bar{H}\right)^{-1}\bar{H}\right)\left(\bar{y}+\bar{\Delta}_y\right) \\
&= \left(\bar{y}+\bar{\Delta}_y\right)^T\left(I-\bar{H}^T\left(\bar{H}^T\bar{H}\right)^{-1}\bar{H}^T+\bar{H}^T\left(\bar{H}^T\bar{H}\right)^{-1}\bar{H}-\bar{H}^T\left(\bar{H}^T\bar{H}\right)^{-1}\bar{H}\right)\left(\bar{y}+\bar{\Delta}_y\right) \\
&= \left(\bar{y}+\bar{\Delta}_y\right)^T\left(I-\bar{H}^T\left(\bar{H}^T\bar{H}\right)^{-1}\bar{H}\right)\left(\bar{y}+\bar{\Delta}_y\right) \qquad (2.39)
\end{aligned}
$$

Let $\tilde{H}=\tilde{H}=\left(I-\bar{H}^T\left(\bar{H}^T\bar{H}\right)^{-1}\bar{H}\right)$. Substituting $\tilde{H}$ in equation (2.39), yields the following:

$$
\begin{aligned}
\bar{J}_{\Delta_y} &= \left(\bar{y}+\bar{\Delta}_y\right)^T\tilde{H}\left(\bar{y}+\bar{\Delta}_y\right) \\
&= \bar{y}^T\tilde{H}\bar{y}+\bar{\Delta}_y^T\tilde{H}\bar{\Delta}_y+2\bar{y}^T\tilde{H}\bar{\Delta}_y \qquad (2.40)
\end{aligned}
$$

To maximize $\bar{J}_{\Delta_y}$, the term $\bar{y}^T\tilde{H}\bar{y}$ is omitted since it does not depend on $\bar{\Delta}_y$, yields the following:

$$
\begin{aligned}
\bar{\Delta}_y^T\tilde{H}\bar{\Delta}_y+2\bar{y}^T\tilde{H}\bar{\Delta}_y &= \left(\bar{\Delta}_y^T\tilde{H}+2\bar{y}^T\tilde{H}\right)\bar{\Delta}_y \\
&= \left(\tilde{H}^T\bar{\Delta}_y+2\tilde{H}^T\bar{y}\right)^T\bar{\Delta}_y \qquad (2.41)
\end{aligned}
$$

Applying the Cauchy-Shwartz inequality, equation (2.41) can be written as:

$$
\begin{aligned}
\bar{\Delta}_y^T\tilde{H}\bar{\Delta}_y+2\bar{y}^T\tilde{H}\bar{\Delta}_y = \left(\tilde{H}^T\bar{\Delta}_y+2\tilde{H}^T\bar{y}\right)^T\bar{\Delta}_y &\leq \left\|\tilde{H}^T\bar{\Delta}_y+2\tilde{H}^T\bar{y}\right\|\left\|\bar{\Delta}_y\right\| \\
&\leq \left\|\tilde{H}^T\bar{\Delta}_y+2\tilde{H}^T\bar{y}\right\|M \qquad (2.42)
\end{aligned}
$$

The upper bound in (2.42) is achieved for some scalar β if and only if:

$$
\tilde{H}^T\bar{\Delta}_y+2\tilde{H}^T\bar{y}=\beta\bar{\Delta}_y \qquad (2.43)
$$

From (2.42) and (2.43):

$$
\begin{aligned}
\left\|\tilde{H}^T\bar{\Delta}_y+2\tilde{H}^T\bar{y}\right\| &\leq \left\|\tilde{H}^T\bar{\Delta}_y+2\tilde{H}^T\bar{y}\right\|M \\
&\leq \left\|\beta\bar{\Delta}_y\right\|M \\
&\leq |\beta|\,M^2 \qquad (2.44)
\end{aligned}
$$

From (2.44) the optimal attack $\bar{\Delta}_y$ satisfies the following expression:

$$\begin{aligned}\left(\beta I - \tilde{H}^T\right)\bar{\Delta}_y &= 2\tilde{H}^T\bar{y} \\ \bar{\Delta}_y &= 2\left(\beta I - \tilde{H}^T\right)^{-1}\tilde{H}^T\bar{y}\end{aligned} \tag{2.45}$$

For (2.44) to hold with equality, equation (2.45) should satisfied with the scalar β with the maximal magnitude $|\beta|$. That is,

$$\begin{aligned}\max|\beta| &=: |\beta_{\max}| \\ s.t.\quad M &= \left\| 2\left(\beta I - \tilde{H}^T\right)^{-1}\tilde{H}^T\bar{y} \right\|\end{aligned} \tag{2.46}$$

Solving for β yields the following:

$$\left(\beta I - \tilde{H}^T\right)^{-1}\tilde{H}^T = \left(I - \frac{\tilde{H}^T}{\beta}\right)^{-1}\beta^{-1}\tilde{H}^T \tag{2.47}$$

Using series expansion for $\left(I - \frac{\tilde{H}^T}{\beta}\right)^{-1}$, equation (2.47) can be written as:

$$\begin{aligned}\left(\beta I - \tilde{H}^T\right)^{-1}\tilde{H}^T &= \left(I + \frac{\tilde{H}^T}{\beta} + \frac{\left(\tilde{H}^T\right)^2}{\beta^2} + \frac{\left(\tilde{H}^T\right)^3}{\beta^3} + \cdots\right)\beta^{-1}\tilde{H}^T \\ &= \left(\frac{\tilde{H}^T}{\beta} + \frac{\left(\tilde{H}^T\right)^2}{\beta^2} + \frac{\left(\tilde{H}^T\right)^3}{\beta^3} + \frac{\left(\tilde{H}^T\right)^4}{\beta^4} + \cdots\right) \\ &= \sum_{j=1}^{\infty}\frac{\left(\tilde{H}^T\right)^j}{\beta^j}\end{aligned} \tag{2.48}$$

To simplify equation (2.48) we notice the following:

$$\begin{aligned}\tilde{H}\tilde{H} &= \left(I - \bar{H}^T(\bar{H}^T\bar{H})^{-1}\bar{H}\right)\left(I - \bar{H}^T(\bar{H}^T\bar{H})^{-1}\bar{H}\right) \\ &= I - \bar{H}^T(\bar{H}^T\bar{H})^{-1}\bar{H} + \bar{H}^T(\bar{H}^T\bar{H})^{-1}\underbrace{\bar{H}\bar{H}^T(\bar{H}^T\bar{H})^{-1}}_{I}\bar{H} \\ &= I - \bar{H}^T(\bar{H}^T\bar{H})^{-1}\bar{H} = \tilde{H}\end{aligned} \tag{2.49}$$

Therefore, we have $\tilde{H}\tilde{H} = \tilde{H}$, and it follows that:

$$\left(\tilde{H}\tilde{H}\right)^T = \tilde{H}^T = \tilde{H}^T\tilde{H}^T \tag{2.50}$$

From equation (2.50) yields the following:

$$\tilde{H}^T = \left(\tilde{H}^T\right)^2 = \left(\tilde{H}^T\right)^3 = \left(\tilde{H}^T\right)^4 = \cdots \tag{2.51}$$

Using equation (2.51), we can simplify equation (2.48) to the following:

$$\sum_{j=1}^{\infty} \frac{\left(\tilde{H}^T\right)^j}{\beta^j} = \tilde{H}^T \sum_{j=1}^{\infty} \frac{1}{\beta^j} \tag{2.52}$$

From equation (2.52) we notice the geometric series and can be simplified to the following:

$$\sum_{j=1}^{\infty} \frac{1}{\beta^j} = \frac{1}{1-\frac{1}{\beta}} = \frac{\beta}{\beta-1} \quad |\beta| > 1 \tag{2.53}$$

Therefore, equation (2.17) can be simplified to the following:

$$M = \left\| 2\frac{\beta}{\beta-1}\tilde{H}\ \bar{y} \right\| \tag{2.54}$$

Solving for β we have the following:

$$\begin{aligned} M &= \left\| 2\frac{\beta}{\beta-1}\tilde{H}\ \bar{y} \right\| \\ M &= 2\left|\frac{\beta}{\beta-1}\right| \left\| \tilde{H}\bar{y} \right\| \\ \left|\frac{\beta}{\beta-1}\right| &= \frac{M}{2\left\| \tilde{H}\bar{y} \right\|} \\ 1-\frac{1}{\beta} &= \pm\frac{2\left\| \tilde{H}\bar{y} \right\|}{M} \end{aligned} \tag{2.55}$$

Therefore, the solution for β is:

$$|\beta| \ = \max\left(\frac{M}{M+2\left\|\tilde{H}\bar{y}\right\|}, \frac{M}{M-2\left\|\tilde{H}\bar{y}\right\|}\right) \tag{2.56}$$

And the optimal signal attack is given by:

$$\bar{\Delta}_y = 2\Big(\beta_{max}I - \tilde{H}^T\Big)^{-1}\tilde{H}^T\bar{y} \tag{2.57}$$

Substituting $\tilde{H}$ in equation (2.56) and (2.57), yields the following:

$$|\beta| \ = \max\left(\frac{M}{M+2\left\|\Big(I-\bar{H}^T\big(\bar{H}^T\bar{H}\big)^{-1}\bar{H}\Big)\,\bar{y}\right\|}, \frac{M}{M-2\left\|\Big(I-\bar{H}^T\big(\bar{H}^T\bar{H}\big)^{-1}\bar{H}\Big)\,\bar{y}\right\|}\right) \tag{2.58}$$

$$\bar{\Delta}_y = 2\Bigg(\beta_{max}I - \Big(I-\bar{H}^T\big(\bar{H}^T\bar{H}\big)^{-1}\bar{H}\Big)^T\Bigg)^{-1}\Big(I-\bar{H}^T\big(\bar{H}^T\bar{H}\big)^{-1}\bar{H}\Big)^T\bar{y} \tag{2.59}$$

Therefore, the optimal signal attack is given by:

$$\begin{aligned}\Delta_y \ &= \ 2R^{1/2}\Bigg(\beta_{max}I - \Big(I-H^TR^{-1/2}\big(H^TR^{-1}\bar{H}\big)^{-1}R^{-1/2}H\Big)^T\Bigg)^{-1} \\ &\quad \Big(I-H^TR^{-1/2}\big(H^TR^{-1}H\big)^{-1}R^{-1/2}H\Big)^TR^{-1/2}y\end{aligned} \tag{2.60}$$

2.3 Optimal Sensor Attack for Recursive Least Squares

In this section, the optimal sensor attack strategies on the Recursive Least Squares filter in which the attacker aims at making the estimation error as large as possible are derived. The attacker needs to determine the probability distribution which deteriorates the estimation cost function the most. For the Recursive Least Squares, suppose we have the following linear recursive estimator [80]:

$$\begin{aligned} y_k \ &= \ H_kx_k + v_k \\ \hat{x}_k \ &= \ \hat{x}_{k-1} + K_k\left(y_k - H_k\hat{x}_{k-1}\right)\end{aligned} \tag{2.61}$$

The noise process v_k is white, zero-mean, uncorrelated process, with known covariance matrice R_k. x_k is the state vector, Δ_{yk} is the optimal random signal attack at time k, y_k is the output vector, and H_k is the output matrix, $\hat{x}_k$ is the state estimate at time k, K_k is the estimator gain matrix to be determined.

In terms of mathematics, the optimal sensor attack corresponds to designing an attack signal such that:

$$\max_{\Delta_{yk}} \quad \min_{\hat{x}_k} \quad J_k \tag{2.62}$$

where J_k is the estimation cost function. Since a random attacker has finite resources, the attack signal is constrained to be finite energy or finite power but is otherwise random. For the sensor attack, the measurement y_k takes the following new form:

$$y_k = H_k x_k + v_k + \Delta_{y_k}, \tag{2.63}$$

To compute the estimator gain matrix K_k as a function of the optimal signal attack Δ_{yk}, first we need to compute the mean of the estimation error as a function of Δ_{yk} [80]:

$$E\left(\varepsilon_{x,k}\right) = E\left(x_k - \hat{x}_k\right) \tag{2.64}$$

where $\varepsilon_{x,k} = x_k - \hat{x}_k$ is the estimation error. Then the estimator error mean with sensor attack is:

$$\begin{aligned}
E\left(\varepsilon_{x,k}\right) &= E\left(x_k - \hat{x}_k\right) = E\left[x_k - \hat{x}_{k-1} - K_k\left(y_k - H_k\hat{x}_{k-1}\right)\right] \\
&= E\left[x_k - \hat{x}_{k-1} - K_k\left(H_k x + v_k + \Delta_{y_k} - H_k\hat{x}_{k-1}\right)\right] \\
&= E\left[x_k - \hat{x}_{k-1} - K_k H_k\left(\underbrace{x_k - \hat{x}_{k-1}}_{\varepsilon_{x,k-1}}\right) - K_k\left(v_k + \Delta_{y_k}\right)\right] \\
&= E\left[\left(I - K_k H_k\right)\left(\varepsilon_{x,k-1}\right) - K_k\left(v_k + \Delta_{y_k}\right)\right] \\
&= \left(I - K_k H_k\right) E\left(\varepsilon_{x,k-1}\right) - K_k E\left(v_k\right) - K_k E\left(\Delta_{y_k}\right)
\end{aligned} \tag{2.65}$$

The cost function J_k can be written as:

$$\begin{aligned} J_k &= E\left[(x_1-\hat{x}_1)^2\right]+\cdots+E\left[(x_n-\hat{x}_n)^2\right] \\ &= E\left({\varepsilon_{x1,k}}^2+\cdots+{\varepsilon_{xn,k}}^2\right) \\ &= E\left({\varepsilon_k}^T\varepsilon_k\right) \\ &= E\left[Tr\left(\varepsilon_k{\varepsilon_k}^T\right)\right] \\ &= TrP_k \end{aligned} \tag{2.66}$$

where P_k is estimation error covariance after the measurement y_k is received.

To minimize the cost function, we need to obtain a recursive formula for P_k as a function of the signal attack Δ_{yk}:

$$\begin{aligned} P_k &= E\left\{\left[\left(I-K_kH_k\right)\varepsilon_{k-1}-K_k\left(v_k+\Delta_{y_k}\right)\right]\left[\cdots\right]^T\right\} \\ &= \left(I-K_kH_k\right)E\left(\varepsilon_{k-1}{\varepsilon_{k-1}}^T\right)\left(I-K_kH_k\right)^T-K_kE\left(\left(v_k+\Delta_{y_k}\right){\varepsilon_{k-1}}^T\right)\left(I-K_kH_k\right)^T \\ &- \left(I-K_kH_k\right)E\left(\varepsilon_{k-1}\left(v_k+\Delta_{y_k}\right)^T\right)K_k^T-K_kE\left(\left(v_k+\Delta_{y_k}\right)\left(v_k+\Delta_{y_k}\right)^T\right)K_k^T \end{aligned} \tag{2.67}$$

Main Results

The previous equation (2.67) can be simplified since measurement noise is assumed to be independent of the estimation error and the attack signal:

$$\begin{aligned} P_k &= \left(I-K_kH_k\right)P_k(I-K_kH_k)^T+KD_kK_k^T-K_kE\left[\Delta_{y_k}{\varepsilon_k}^T\right]\left(I-K_kH_k\right)^T+KR_kK_k^T \\ &- \left(I-K_kH_k\right)E\left[\varepsilon_k\Delta_{y_k}^T\right]K_k^T \end{aligned} \tag{2.68}$$

where $R_k=E(v_k{v_k}^T)$ and $D_k=E(\Delta_{yk}{\Delta_{yk}}^T)$.

Therefore, the optimization (2.62) reduces to finding the attack signal Δ_{yk} which maximizes the following:

$$\begin{aligned} \max_{\Delta_{yk}}\left\{\min_{K_k}Tr\left(P_k\right)\right\} &= \max_{\Delta_{yk}}\left\{\min_{K_k}Tr\left(\left(I-K_kH_k\right)P_{k-1}(I-K_kH_k)^T+KD_kK_k^T+KR_kK_k^T\right.\right. \\ &\left.\left.-K_kE\left[\Delta_{y_k}{\varepsilon_k}^T\right]\left(I-K_kH_k\right)^T-\left(I-K_kH_k\right)E\left[\varepsilon_k\Delta_{y_k}^T\right]K_k^T\right)\right\} \end{aligned} \tag{2.69}$$

In the following, one main proposition is given, where the estimation error is **independent** of the attack signal Δ_y at time k.

Estimation Error Independent of Δ_y

To simplify equation (2.68) the estimation error at time $k-1$ ε_{k-1} is assumed to be independent of the measurement noise v_k , and attack signal Δ_{yk} at time k.

Proposition 2.3.1. *The optimal estimation error covariance as a function of the signal attack Δ_{yk} is given by:*

$$P_k = P_{k-1} - P_{k-1}H_k^T\left(H_kP_{k-1}H_k^T + R_k + D_k\right)^{-1}H_kP_{k-1}.$$

Proof. We rewrite (2.68) as

$$P_k \;=\; (I - K_kH_k)\,E\left(\varepsilon_{k-1}{\varepsilon_{k-1}}^T\right)(I - K_kH_k)^T + K_kE\left((v_k + \Delta_{y_k})\,(v_k + \Delta_{y_k})^T\right)K_k^T. \quad (2.70)$$

Moreover, we assume the measurement noise v_k is independent of the attack signal Δ_k at time k. Therefore, $E\left(v_k\Delta_{y_k}^T\right) = E\left(\Delta_{y_k}v_k^T\right) = 0$ and (2.70) could be further simplified to:

$$P_k \;=\; (I - K_kH_k)\,P_{k-1}(I - K_kH_k)^T + K_kD_kK_k^T + K_kR_kK_k^T, \quad (2.71)$$

where $R_k = E(v_k{v_k}^T)$ and $D_k = E(\Delta_{yk}{\Delta_{yk}}^T)$. With this expression of J_k in hand, we are ready to compute K_k by taking the derivative of the cost function with respect to K_k as:

$$\frac{\partial J_k}{\partial K_k} = 2\,(I - K_kH_k)\,P_{k-1}\left(-H_k^T\right) + 2K_kR_k + 2K_kD_k \quad (2.72)$$

In order to find K_k, we set the previous equation equal to zero and compute K_k:

$$\begin{aligned} K_k\,(R_k + D_k) &= (I - K_kH_k)\,P_{k-1}\left(H_k^T\right) \\ K_k\,(R_k + D_k) &= P_{k-1}H_k^T - K_kH_kP_{k-1}H_k^T \\ K_k\left(R_k + D_k + H_kP_{k-1}H_k^T\right) &= P_{k-1}H_k^T \end{aligned} \quad (2.73)$$

Therefore, K_k should satisfy:

$$K_k = P_{k-1}H_k^T\left(R_k + D_k + H_kP_{k-1}H_k^T\right)^{-1} \tag{2.74}$$

Let $S_k = R_k + D_k + H_kP_{k-1}H_k^T$, then K_k is:

$$K_k = P_{k-1}H_k^TS_k^{-1} \tag{2.75}$$

Substituting K_k in (2.68) we get:

$$\begin{aligned} P_k &= P_{k-1} - P_{k-1}H_k^TS_k^{-1}H_kP_{k-1} - P_{k-1}H_k^TS_k^{-1}H_kP_{k-1} + P_{k-1}H_k^TS_k^{-1}H_kP_{k-1}H_k^TS_k^{-1}H_kP_{k-1} \\ &+ P_{k-1}H_k^TS_k^{-1}R_kS_k^{-1}H_kP_{k-1} + P_{k-1}H_k^TS_k^{-1}D_kS_k^{-1}H_kP_{k-1} \end{aligned} \tag{2.76}$$

Combining the last three terms in the previous equation,

$$\begin{aligned} P_k &= P_{k-1} - 2P_{k-1}H_k^TS_k^{-1}H_kP_{k-1} + P_{k-1}H_k^TS_k^{-1}S_kS_k^{-1}H_kP_{k-1} \\ &= P_{k-1} - 2P_{k-1}H_k^TS_k^{-1}H_kP_{k-1} + P_{k-1}H_k^TS_k^{-1}H_kP_{k-1} \\ &= P_{k-1} - P_{k-1}H_k^TS_k^{-1}H_kP_{k-1} \end{aligned} \tag{2.77}$$

Finally, we can rewrite P_k as:

$$P_k = P_{k-1} - P_{k-1}H_k^T\left(H_kP_{k-1}H_k^T + R_k + D_k\right)^{-1}H_kP_{k-1} \tag{2.78}$$

□

Proposition 2.3.1 shows how the covariance matrix is affected by the sensor attack signals. It follows then that the worst case attack scenario corresponds to maximizing the following cost function:

$$J_k = Tr(P_k) \tag{2.79}$$

Next, the optimal attack problem is converted into an optimization problem described as follows:

Problem 2.3.2.

$$\begin{aligned} &\max J_k, \\ &\text{subject to } Tr\left[E\left[\Delta_{y_k}\Delta_{y_k}^T\right]\right] \leq M \end{aligned} \tag{2.80}$$

In order to solve the optimization problem (2.3.2) the Lagrangian is defined as:

$$\begin{aligned} &L = Tr\left[P_{k-1} - P_{k-1}H_k^T\left(H_kP_{k-1}H_k^T + R_k + D_k\right)^{-1}H_kP_{k-1}\right] \\ &+\lambda\left(M - Tr\left[E\left[\Delta_{y_k}\Delta_{y_k}^T\right]\right]\right) \end{aligned} \tag{2.81}$$

To avoid cumbersome computation, the scalar case is only considered:

$$\begin{aligned} L &= P_{k-1} - {P_{k-1}}^2H_k^2\left(H_k^2P_{k-1} + R_k + D_k\right)^{-1} + \lambda\left(M_k - E\left[\Delta_{y_k}^2\right]\right) \\ &= P_{k-1} - {P_{k-1}}^2H_k^2S1_k^{-1} + \lambda\left(M - E\left[\Delta_{y_k}^2\right]\right) \\ S1_k^{-1} &= H_k^2P_{k-1} + R_k + D_k \end{aligned} \tag{2.82}$$

$$\begin{aligned} \frac{\partial L}{\partial \Delta_{y_k}} &= 2{P_{k-1}}^2H_k^2E\left[\Delta_{y_k}\right]S1_k^{-2} - 2\lambda E\left[\Delta_{y_k}\right] = 0 \\ \frac{\partial L}{\partial \lambda} &= M - E\left[\Delta_{y_k}^2\right] = 0 \end{aligned} \tag{2.83}$$

From the previous equation (2.83) we have λ is equal to the following:

$$\begin{cases} \lambda = \frac{{P_{k-1}}^2H_k^2}{\left(H_k^2P_{k-1} + R_k + E\left[\Delta_{yk}^2\right]\right)^2} \\ E\left[\Delta_{yk}^2\right] = M \end{cases} \tag{2.84}$$

Therefore, we have:

$$\lambda = \frac{{P_{k-1}}^2H_k^2}{\left(H_k^2P_{k-1} + R_k + M\right)^2} \tag{2.85}$$

To show that this indeed corresponds to a maximum, the 2^{nd} derivative of L is computed:

$$\frac{\partial^2 L}{\partial \Delta_{y_k}^2} = -8E[\Delta_{y_k}]^2 {P_{k-1}}^2 H_k^2 S1_k^{-3} - 2\lambda + 2{P_{k-1}}^2 H_k^2 S1_k^{-2} \tag{2.86}$$

Replacing λ in (2.86):

$$\frac{\partial^2 L}{\partial \Delta_{y_k}^2} = -8E[\Delta_{y_k}]^2 {P_{k-1}}^2 H_k^2 S1_k^{-3} < 0 \tag{2.87}$$

Therefore, any signal Δ_{y_k} which satisfies (2.84) is optimal.

Chapter 3

Optimal Sensor and Actuator Attacks on Finite Horizon Linear Quadratic (LQ) Control

In this chapter, cyber-attack on the sensor and actuator are analyzed and results are simulated for the finite horizon linear quadratic control. We assume that the attacker is capable of spoofing sensor and actuator signals with time-varying signals. Both sensor and actuator attacks are considered for finite energy and bounded energy signals attack. We also assume that (A, B) is controllable, and (A, C) are observable. Consider the plant with the following state-space model:

$$\begin{aligned} \dot{X}(t) &= AX(t) + B_2(t)U(t), \;\; \mathrm{X}(0) = \mathrm{X}_0 \\ Z(t) &= \begin{pmatrix} C_1 X \\ U \end{pmatrix} \end{aligned} \tag{3.1}$$

The LQ control corresponds to minimizing the following quadratic performance index (cost function) [81]:

$$J = \underbrace{\tfrac{1}{2}\left(X_f^T S_f X_f\right)}_{\varphi(X_f)} + \int_{t_0}^{t_f} \underbrace{\tfrac{1}{2}\left(X^T Q X + U^T R U\right)}_{L(X,U)} dt \tag{3.2}$$

Where S_f, Q are positive semi-definite matrices ($psdf$), and R is positive definite matrix (pdf), $S_f, Q \geq 0$ and $R > 0$. The Riccati equation is [81]:

$$\dot{P} + PA - PB_2R^{-1}B_2^TP + A^TP + Q = 0 \tag{3.3}$$

Let us now consider the minimizing of the following cost function for the state-space equation 3.1

$$J = \int_t^h Z^T(\tau)Z(\tau)d\tau + X^T(h)Qx(h) \tag{3.4}$$

With $R = I$ we have the optimal controller simplifies to the following [81]:

$$U = -B_2^TPx \tag{3.5}$$

Also, the Riccati equation:

$$\dot{P} + PA - PB_2B_2^TP + A^TP + Q = 0 \tag{3.6}$$

with $P(h) = Q$ and ${C_1}^T C_1 = Q$ The cost function can be reduced using the Riccati equation, we have the following [81]:

$$J = \int_t^h X^TC_1^TC_1X + U^TU + \frac{d}{dt}\left(X^TPX\right)d\tau + X^T(t)P(t)X(t) \tag{3.7}$$

The previous equation (3.7) can be simplified to the following [81]:

$$J = \int_t^h \left(U + B_2^TPX\right)^T\left(U + B_2^TPX\right)d\tau + X^T(t)P(t)X(t) \tag{3.8}$$

For $t = 0$ (3.8) becomes [81]:

$$J = \int_0^h \left(U + B_2^TPX\right)^T\left(U + B_2^TPX\right)d\tau + X^T(0)P(0)X(0) \tag{3.9}$$

3.1 Optimal Sensor Attack Finite Energy

In this section we assume full-information, i.e., the state vector is available for measurement the sensor attack is defined as follows [82]:

$$y_\alpha(t) = x(t) + \Delta_x(t), \mathrm{t} \geq 0 \tag{3.10}$$

Therefore, the cost function J after using (3.8), the sensor attack signal $\Delta_x(t)$, and the optimal controller $U = -B_2 P(X + \Delta_x(t))$ takes the following form [82]:

$$\mathrm{J} = (X+\Delta_x(t))^T P\,(X+\Delta_x(t)) + \int_t^h \left(B_2^T P \Delta_x\right)^T \left(B_2^T P \Delta_x\right) d\tau \tag{3.11}$$

For $t = 0$ (3.11) becomes:

$$\begin{aligned} J &= (X(0)+\Delta_x(0))^T P(0)\,(X(0)+\Delta_x(0)) + \int_0^h \left(B_2^T P \Delta_x\right)^T \left(B_2^T P \Delta_x\right) d\tau \\ &= (X(0)+\Delta_x(0))^T P(0)\,(X(0)+\Delta_x(0)) + \int_0^h \left\|B_2^T P \Delta_x\right\|^2 d\tau \end{aligned} \tag{3.12}$$

Using the Rayleigh-Ritz inequality, the sensor attack satisfies the following [82]:

$$J \leq \lambda_{\max}\left(P(0)\right)\left(x_0 + \Delta_x\left(0\right)\right)^T \left(x_0 + \Delta_x\left(0\right)\right) + \int_0^h \left\|B_2^T P\left(\tau\right)\right\|^2 \|\Delta_x\left(\tau\right)\|^2 d\tau \tag{3.13}$$

Where $\lambda_{\max}\left(P(0)\right)$ is the maximal eigenvalue of $P(0)$, and $\left\|B_2^T P\left(\tau\right)\right\|^2 \equiv \lambda_{\max}\left(P B_2 B_2^T P\right)$ is the maximum singular value of $B_2^T P\left(\tau\right)$. The worst case sensor attack corresponds to the eigenvector, denoted $v_1(0)$ associated to $\lambda_{max}(P(0))$, and the singular vector, denoted $v_1(\tau)$ associated to $\left\|B_2^T P\left(\tau\right)\right\|$. Therefore, (3.13) takes the form [83]:

$$\begin{aligned} P(0)v_1(0) &= \lambda_{\max}(P(0))v_1(0), \quad \|v_1(0)\| = 1 \\ P B_2 B_2^T P(\tau)v_1(\tau) &= \left\|B_2^T P(\tau)\right\|^2 v_1(\tau), \quad \|v_1(\tau)\| = 1, \quad \tau > 0 \end{aligned} \tag{3.14}$$

From (3.14) and (3.13) the worst sensor attack signals is [82, 83]:

$$\begin{aligned}\Delta_x(0) &= \alpha_0 v_1(0) - x(0), \text{ for some scalar } \alpha_0 \\ \Delta_x(\tau) &= \alpha(\tau) v_1(\tau), \text{ for some valued function } \alpha(\cdot)\end{aligned} \tag{3.15}$$

For the finite energy case which is a special class of signals (L^2-signals). Let $\Delta_x(\cdot) \in L^2([0,h), \mathbb{R}^n)$, where $L^2([0,h), \mathbb{R}^n)$ is the space of finite energy functions defined in the interval $[0,h)$ and its values in $\mathbb{R}^n$. The assumption is that the attacker cannot expand more than an energy threshold $|\alpha_0| \leq M$ and $\|\Delta_x\|_2 \leq M$ [82, 83]. Then the worst case cost is [82]:

$$\sup_{\Delta_x} J = M\lambda_{\max}\big(P(0)\big) + M \sup_{\tau\in[0,h]} \|B_2^T P(\tau)\|^2 \tag{3.16}$$

Since

$$\sup_{\|\Delta_x\|_2 \leq M} \int_0^h \|B_2^T P \Delta_x\|^2 d\tau = \sup_{\tau\in[0,h)} M\|B_2^T P(\tau)\|^2 \tag{3.17}$$

To compute Δ_x, we need first to have ${B_2}^T P(\tau)$ have the following singular values $\sigma_1(\tau) \geq \sigma_2(\tau) \geq \cdots \geq \sigma_m(\tau)$, $\tau \in [0,h]$ and let $B_2^T P(\tau) = UDV^*(\tau)$ the singular value decomposition, Where U and V are unitary matrix, V^* is the conjugate transpose of V, and $D(\tau) = diag(\sigma_1(\tau) \geq \sigma_2(\tau) \geq \cdots \geq \sigma_m(\tau))$ [82, 83].

Let $\epsilon > 0$, by the definition of supremum there exists a subset $\mathrm{E}_\varepsilon \subset [0,h]$ of positive Lebesgue measure μ such that $\left\|B_2^T P(\tau)\right\| > \sup_{\tau\in[0,h)} \left\|B_2^T P(\tau)\right\| - \varepsilon$ for all $\tau \in \mathrm{E}_\varepsilon$. Let $\chi_{\mathrm{E}_\varepsilon}$ be the characteristic function of E_ε [83]:

$$\chi_{\mathrm{E}_\varepsilon}(\tau) = \begin{cases} 1 & \tau \in \mathrm{E}_\varepsilon \\ 0 & \tau \notin \mathrm{E}_\varepsilon \end{cases} \tag{3.18}$$

And Let $\xi_\varepsilon(\tau) = \frac{1}{\sqrt{\mu(\mathrm{E}_\varepsilon)}}(\chi_{\mathrm{E}_\varepsilon}, 0, \cdots, 0)^T$, then $\|\xi_\varepsilon(\tau)\|_2^2 = \frac{1}{\mu(\mathrm{E}_\varepsilon)}\int_{\mathrm{E}_\varepsilon} d\mu = \frac{\mu(\mathrm{E}_\varepsilon)}{\mu(\mathrm{E}_\varepsilon)} = 1$. The finite energy sensor signal attack is [82, 83]:

$$\begin{aligned}\Delta_x(\tau) &= MV(\tau)\xi_\varepsilon(\tau)\\ &= MV(\tau)\frac{1}{\sqrt{\mu(\mathrm{E}_\varepsilon)}}(\chi_{\mathrm{E}_\varepsilon}, 0, \cdots, 0)^T\\ &= M\frac{\sigma_1(\tau)\,\chi_{\mathrm{E}_\varepsilon}}{\sqrt{\mu(\mathrm{E}_\varepsilon)}}v_1(\tau)\end{aligned} \tag{3.19}$$

where $v(\cdot)$ is the first column of $V(\cdot)$ and it is the right singular vector of $B_2^T P(\cdot)$ associated to $\sigma_1(\cdot)$ Therefore, the closed-loop system for the finite energy under sensor attack can be written as [82, 83]:

$$\begin{aligned}\dot{x}(t) &= \left(A - B_2 B_2^T P(t)\right)x(t) - B_2 \underbrace{B_2^T P(t)}_{UDV^*(t)}\Delta_x(t)\\ &= \left(A - B_2 B_2^T P(t)\right)x(t) - M\frac{\sigma_1(t)\,\chi_{\mathrm{E}_\varepsilon}(t)}{\sqrt{\mu(\mathrm{E}_\varepsilon)}}B_2 u_1(t)\end{aligned} \tag{3.20}$$

Where $u_1(t)$ is the first column of $U(t)$ and represents the unit left singular vector corresponding to the maximal singular value $\sigma_1(t)$

Numerical Example

In this chapter all numerical examples we propose is by employing the distributed control scheme introduced by [84]. A numerical experiment is given to demonstrate the effectiveness of the proposed attack strategies and validate the theoretical analysis. A Power Model Network with a distributed control will be used for the simulation. We assume the attacker has access to the system parameters and the optimal sensor actuator attack for the finite [82, 83]. We assume each generator supplies a time-varying current, time-varying power, and a constant voltage [82, 83]. The state of a single generator is represented by rotor rotation angle δ and described by the following swing equation

[82–84]:

$$M\ddot{\delta} + D\dot{\delta} = P_m - P_e$$
$$\dot{\theta}_i = -\frac{D}{M}\theta_i + \frac{1}{M_i}\sum_{k-i}\frac{V^2X}{|Z|^2}(\varepsilon_i - \varepsilon_k) + \frac{1}{M_i}u_i \tag{3.21}$$

Where P_m is the mechanical power and P_e is the electrical power. M_i is the rotor inertia constant, D is the mechanical damping constant, and $Z = R + jX$ is the impedance of the transmission line. Assume all N generators operate near a stable point δ_0 and there is only small perturbation. Then define $\varepsilon_i = \delta_i - \delta_0$, based on the Kirchoff's current law. The state space equation for each generator is given as [82–84]:

$$\dot{x}_i = Ax_i + Bu_i + \sum_{k=1}^{N} C_k x_k$$
$$A = \begin{bmatrix} 0 & 1 \\ \frac{V^2X}{M_i|Z|^2} & -\frac{D}{M_i} \end{bmatrix}, B = \begin{bmatrix} 0 \\ \frac{1}{M_i} \end{bmatrix} \tag{3.22}$$

Let n_i be the number of generators that are adjacent to generator i, if k is adjacent to i then [84]

$$C_i = \begin{bmatrix} 0 & 0 \\ \frac{(n_i-1)V^2X}{M_i|Z|^2} & 0 \end{bmatrix}, C_k = 0_{2\times 2}\ , otherwise \tag{3.23}$$

This represent the effect on each local generator by states from other generators connected to it. The distributed controller employed will utilize this state information only. A suboptimal LQR is computed following [84]. The numerical results for the optimal attack from the attacker's viewpoint are shown in Figs. 3.1 to 3.6. These results are for finite energy attack signal such as $\|\Delta_x\|_2 \leq M$, where M is fixed by the user. For this simulation we chose $M = 3.4$. Looking at these figures, it is clear that the optimal attack has a greater effect and the larger damage than the random Bias attack [82].

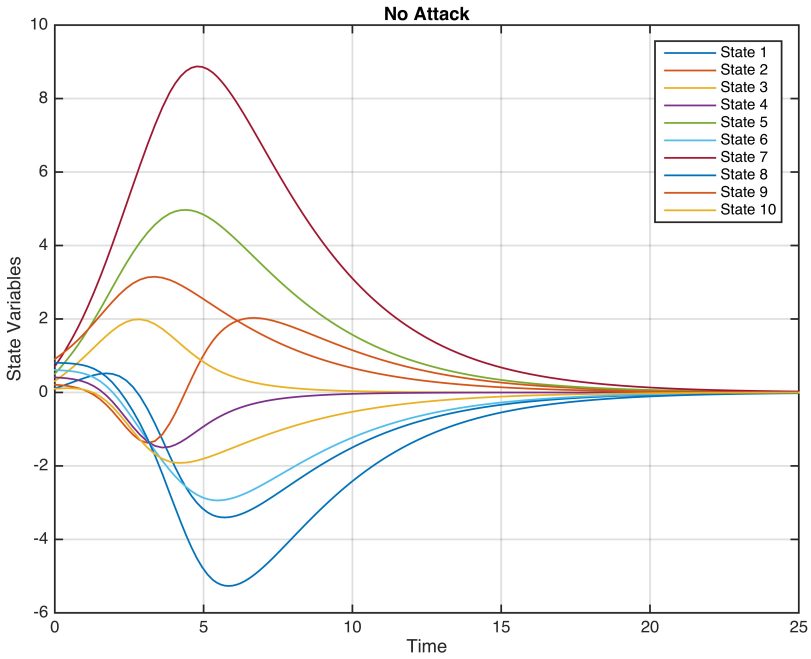


Figure 3.1: Finite Energy Sensor Attack Case: No Attack.

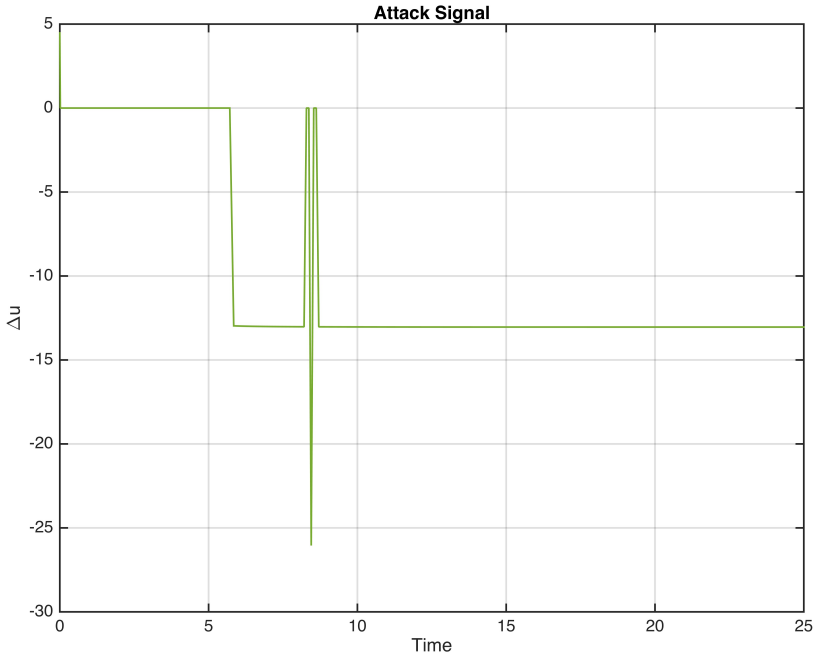


Figure 3.2: Finite Energy Sensor Attack Case: Attack Signal.

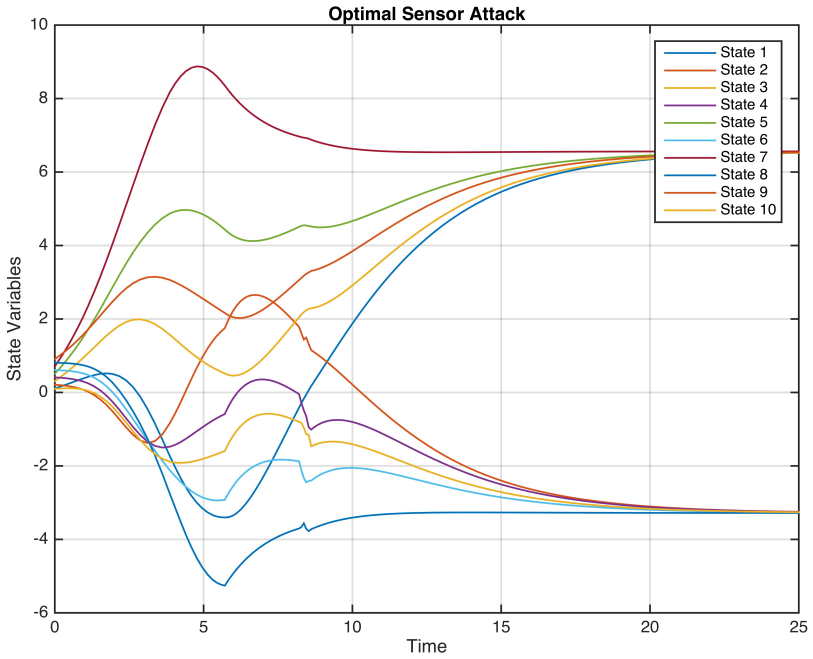


Figure 3.3: Finite Energy Sensor Attack Case: Optimal Attack.

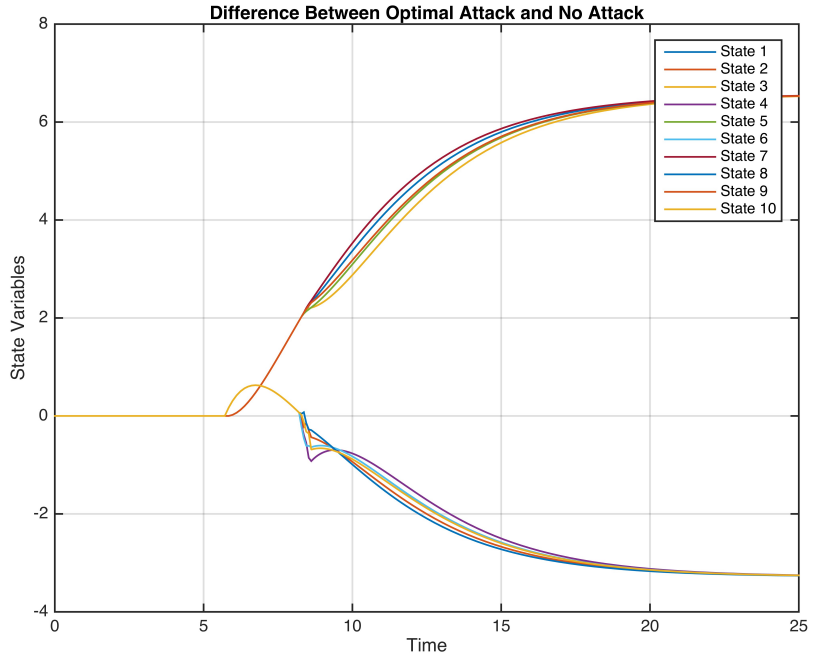


Figure 3.4: Finite Energy Sensor Attack Case: Difference Optimal Attack and No Attack.

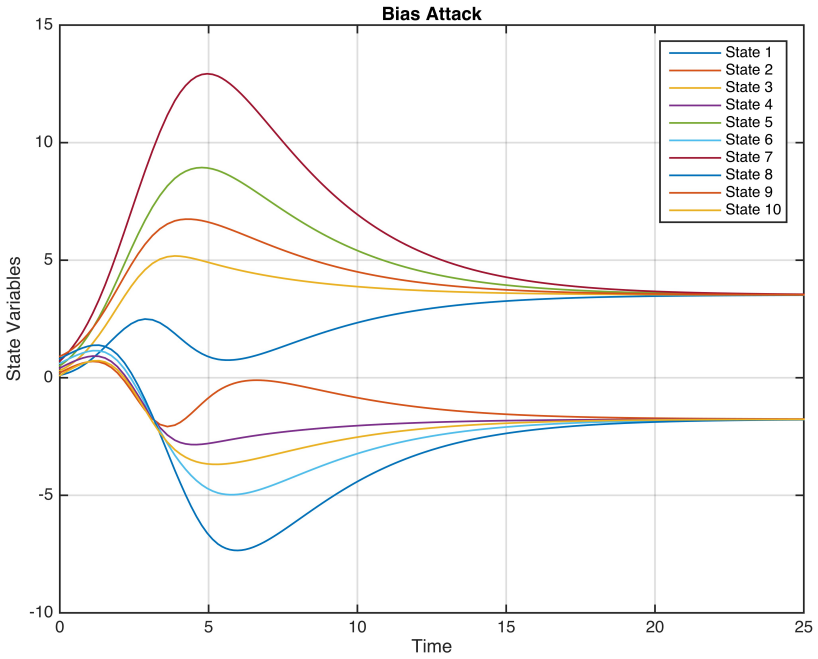


Figure 3.5: Finite Energy Sensor Attack Case: Bias Attack.

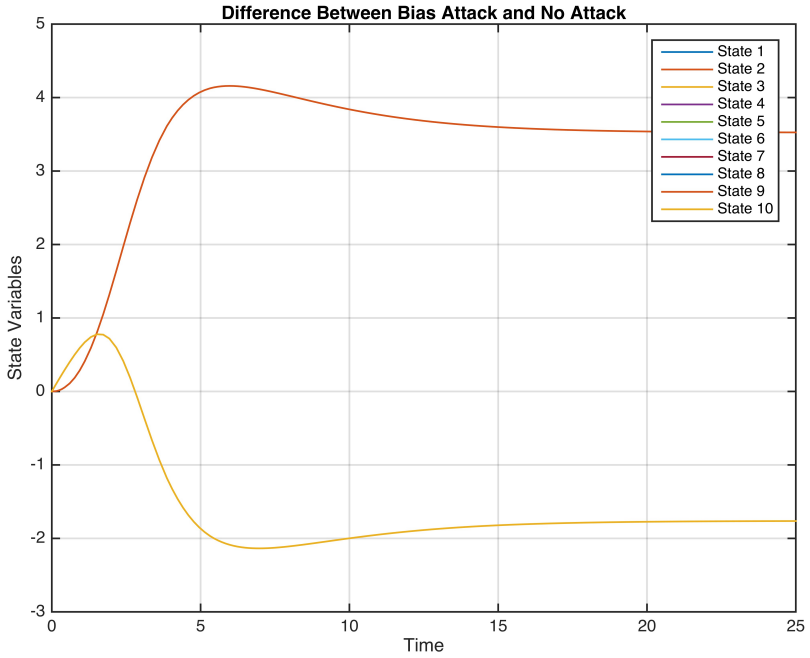


Figure 3.6: Finite Energy Sensor Attack Case: Difference Bias Attack and No Attack.

3.2 Optimal Bounded Sensor Attack Signal

For the bounded signal sensor attacks $\Delta_x \in L^\infty([0,h),\mathbb{R}^n)$, and $\|\Delta_x\|_\infty = esssup_{\tau\in[0,h)}\|\Delta_x(\tau)\|$. Therefore, the following inequalities for the sensor attack signal with $\|\Delta_x\|_\infty \leq M$ [83]:

$$\begin{aligned}\int_0^h \|B_2^T P\Delta_x\|^2 d\tau &\leq \int_0^h \|B_2^T P(\tau)\|^2 \|\Delta_x(\tau)\|^2 d\tau \\ &\leq \int_0^h \|B_2^T P(\tau)\|^2 \underbrace{\{esssup_{\tau\in[0,h)}\|\Delta_x(\tau)\|^2\}}_{\|\Delta_x\|_\infty^2} d\tau \\ &\leq M^2 \int_0^h \|B_2^T P(\tau)\|^2 d\tau\end{aligned} \tag{3.24}$$

The upper bound can be satisfied by choosing Δ_x, the attack signal as [83]:

$$\Delta_x(\tau) = MV(\tau)(1,0,\cdots,0)^T = Mv_1(\tau) \tag{3.25}$$

The bounded sensor attack signal is [83]:

$$\sup_{\|\Delta_x\|_\infty \leq M} J = M^2\lambda_{\max}\big(P(0)\big) + M^2 \int_0^h \|B_2^T P(\tau)\|^2 d\tau \tag{3.26}$$

Therefore, the closed-loop system for the bounded signal sensor attack can be written as [82, 83]:

$$\dot{x}(t) = (A - B_2B_2^T P(t))x(t) - M\sigma_1(t)B_2u_1(t), \;\; 0 \leq t < h \tag{3.27}$$

The numerical results for the optimal attack from the attacker's viewpoint are shown in Figs. 3.7 to 3.12. These results are for bounded attack signal such as $\|\Delta_x\|_2 \leq M$. Looking at these figures, it is clear that the optimal attack has a greater effect and a larger damage than the random bias attack. However, the random bias is harder to detect.

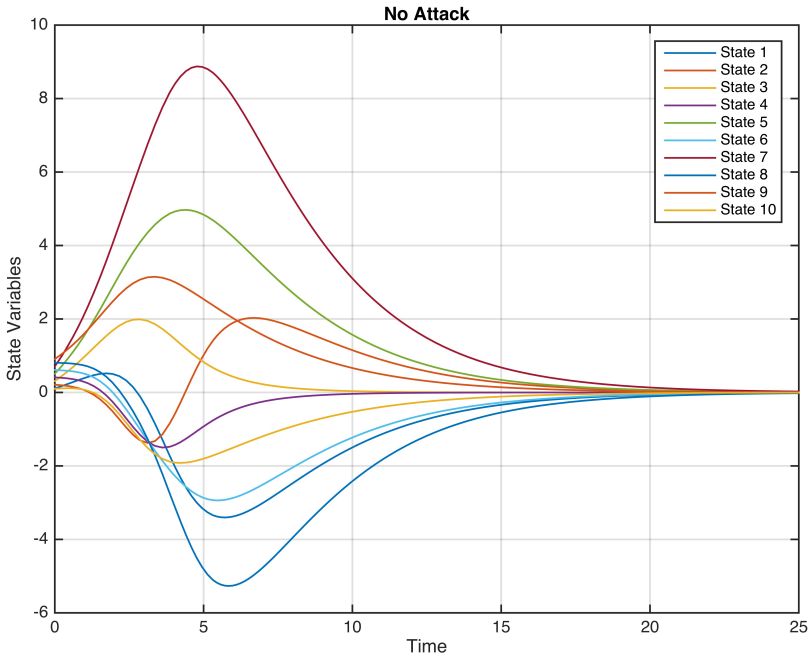


Figure 3.7: Bounded Sensor Attack Signal Case: No Attack.

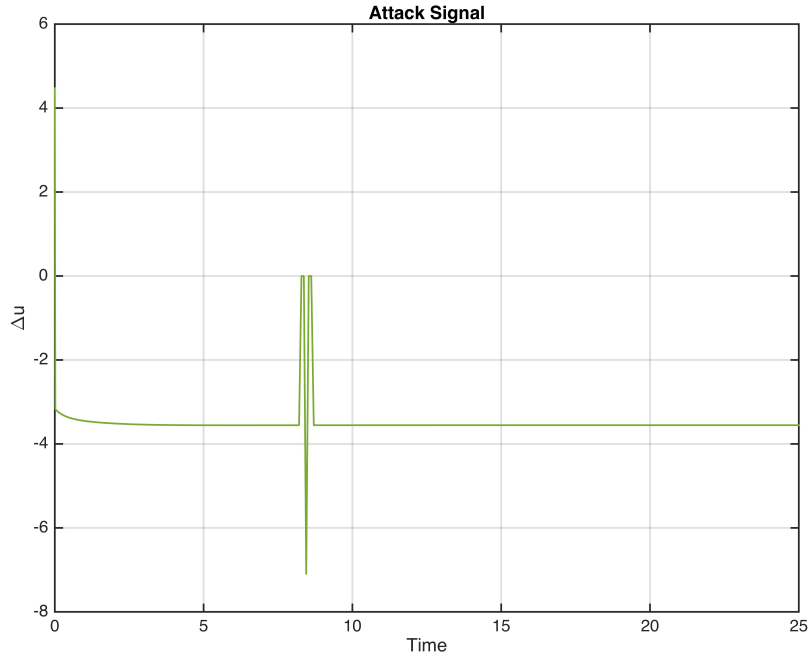


Figure 3.8: Bounded Sensor Attack Signal Case: Attack Signal.

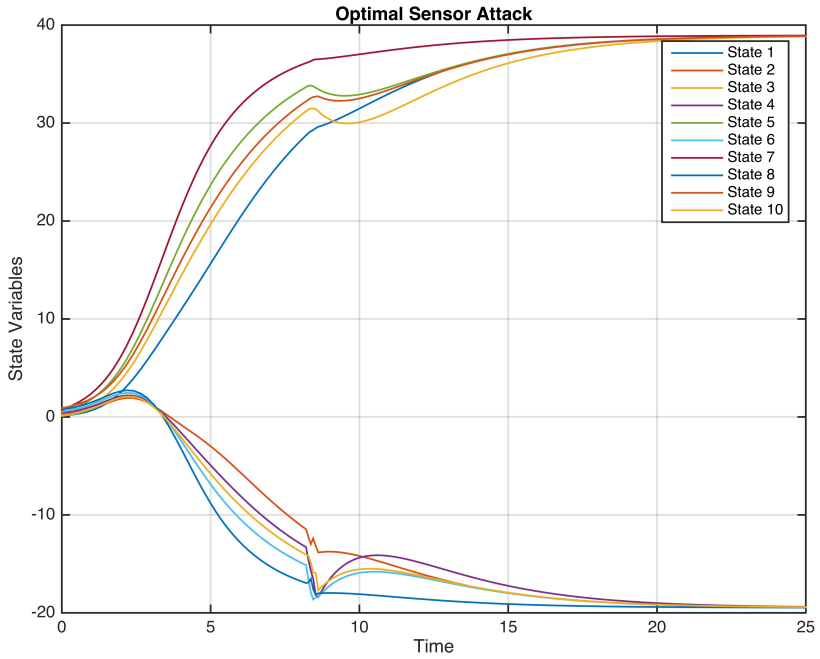


Figure 3.9: Bounded Sensor Attack Signal Case: Optimal Attack.

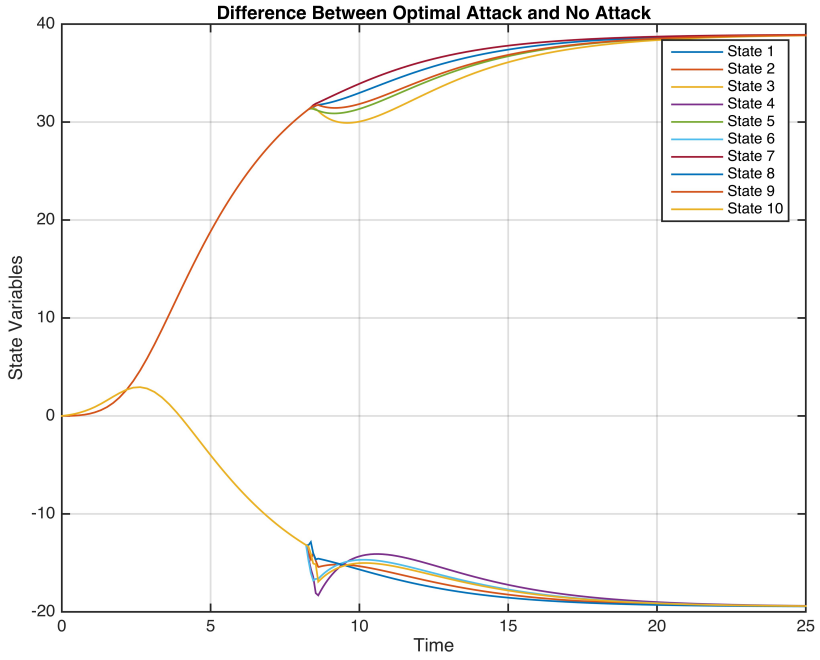


Figure 3.10: Bounded Sensor Attack Signal Case: Difference Optimal Attack and No Attack.

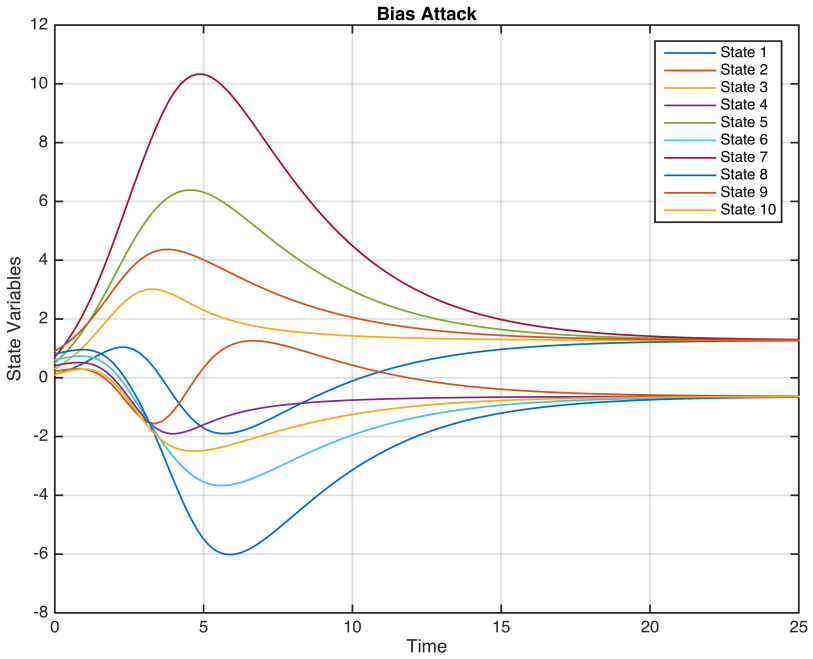


Figure 3.11: Bounded Sensor Attack Signal Case: Bias Attack.

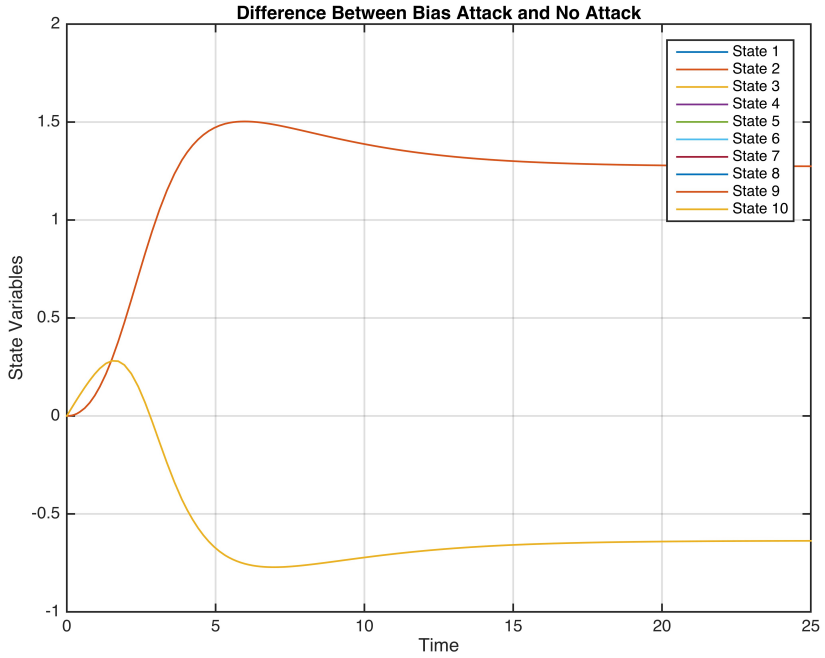


Figure 3.12: Bounded Sensor Attack Signal Case: Difference Bias Attack and No Attack.

3.3 Optimal Actuator Attack Finite Energy

In this section the actuator attack take the following form [83, 85]

$$u_\alpha(t) = u(t) + \Delta_u(t),\ t \geq 0 \tag{3.28}$$

Therefore, the cost function J with the actuator attack signal $\Delta_u(t)$ takes the following form:

$$J = x^T(t)P(t)x(t) + \int_t^h \left[(u + B_2^T Px)^T(u + B_2^T Px) + 2x^T PB_2\Delta_u\right]d\tau \tag{3.29}$$

Using $u = -B_2^T Px$, The optimal actuator cost function takes the following form [83, 85]:

$$\sup_{\Delta_u} J = x^T(t)P(t)x(t) + 2\sup_{\Delta_u}\int_t^h x^T PB_2\Delta_u(\tau)d\tau \tag{3.30}$$

For the RHS of the previous equation and with $\|\Delta_u\|_2 \leq M$, for some constant $M > 0$ and by the Cauchy-Schwarz inequality we have [83, 85]:

$$\sup_{\|\Delta_u\|_2 \leq M}\int_t^h x^T PB_2\Delta_u(\tau)d\tau \leq M\sqrt{\int_t^h \|B_2^T Px(\tau)\|^2 d\tau} \tag{3.31}$$

For a scalar $\beta > 0$, equality of previous equation is achieved when [83, 85]:

$$\Delta_u(\tau) = \beta B_2^T Px(\tau),\ \ 0 \leq t \leq \tau < h \tag{3.32}$$

Therefore, the optimal actuator signal attack takes the following form [83, 85]:

$$\Delta_u(\tau) = \frac{M}{\|B_2^T Px\|_2}\, B_2^T Px(\tau),\ \ 0 \leq t \leq \tau < h \tag{3.33}$$

Therefore, the corresponding worst case cost function [83, 85]:

$$\sup_{\Delta_u} J = 2M\|B_2^T Px\|_2 + x^T(t)P(t)x(t) \tag{3.34}$$

And the closed-loop system for the finite energy under sensor attack can be written as [83, 85]:

$$\dot{x} = \big(A - B_2 B_2^T P + 2\frac{M}{\|B_2^T P x\|_2} B_2 B_2^T P\big)x \tag{3.35}$$

The numerical results for the optimal attack from the attacker's viewpoint are shown in Figs. 3.13 to 3.18. These results are for finite energy actuator signal attack such as $\|\Delta_u\|_2 \leq M$. Looking at these figures, it is clear that the optimal attack has a greater effect and the larger damage than the random bias attack.

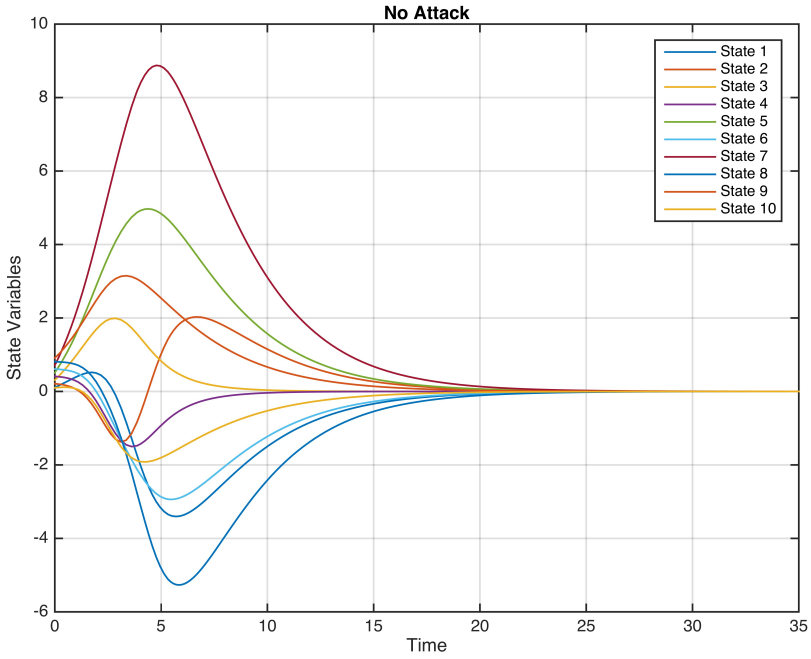


Figure 3.13: Finite Energy Actuator Attack Case: No Attack.

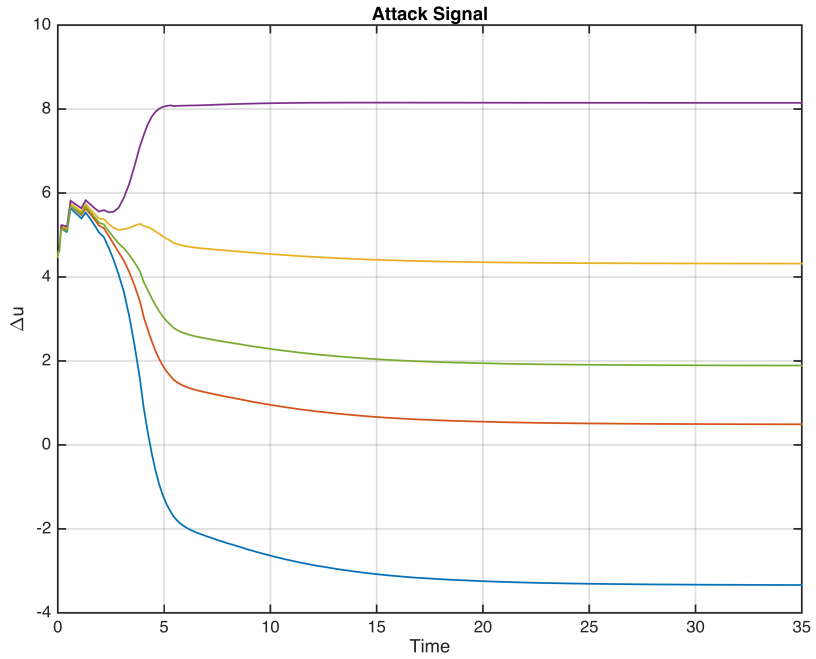


Figure 3.14: Finite Energy Actuator Attack Case: Attack Signal.

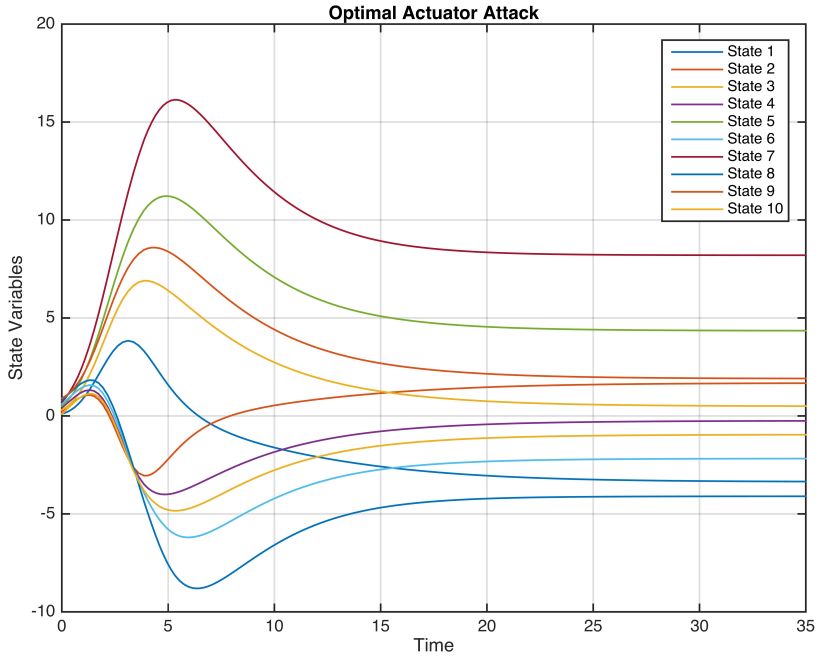


Figure 3.15: Finite Energy Actuator Attack Case: Optimal Attack.

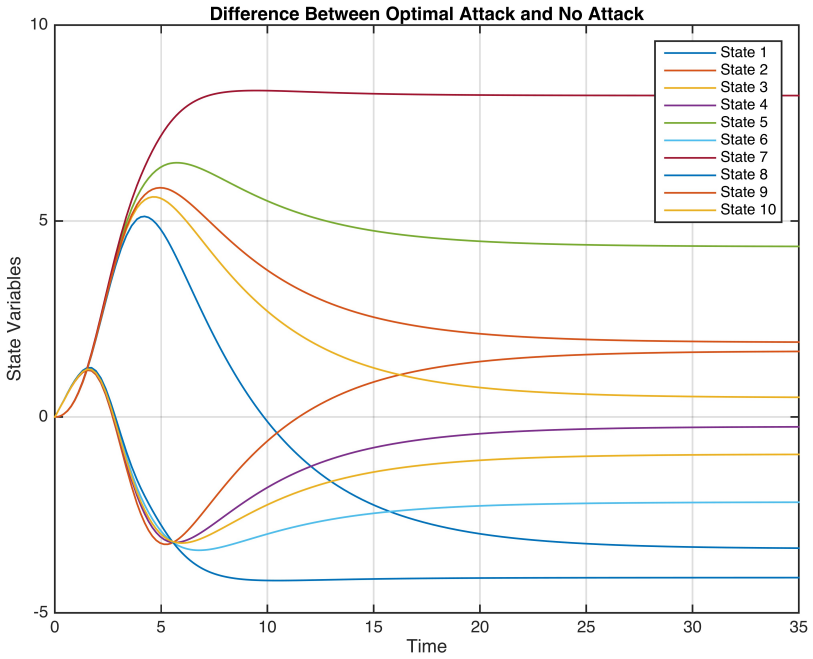


Figure 3.16: Finite Energy Actuator Attack Case: Difference Optimal Attack and No Attack.

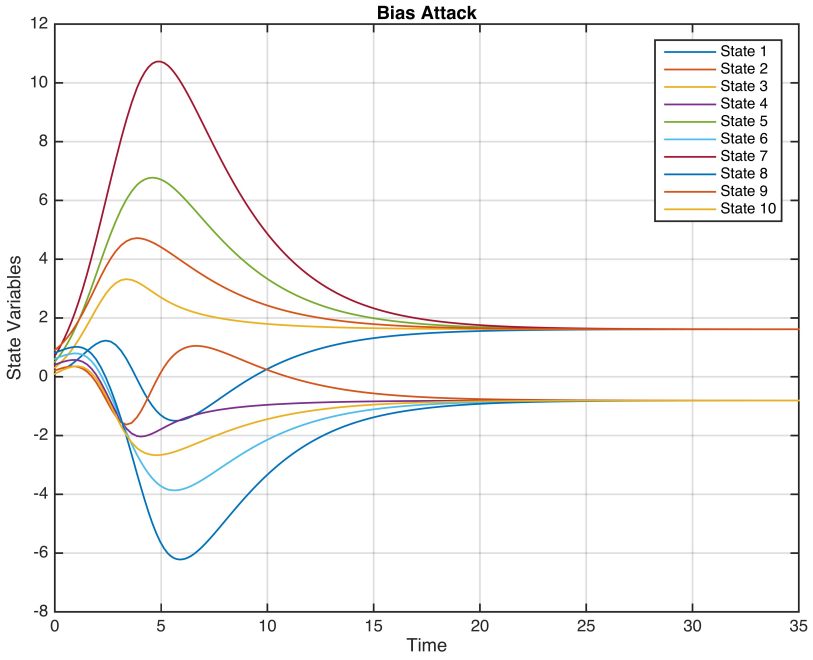


Figure 3.17: Finite Energy Actuator Attack Case: Bias Attack.

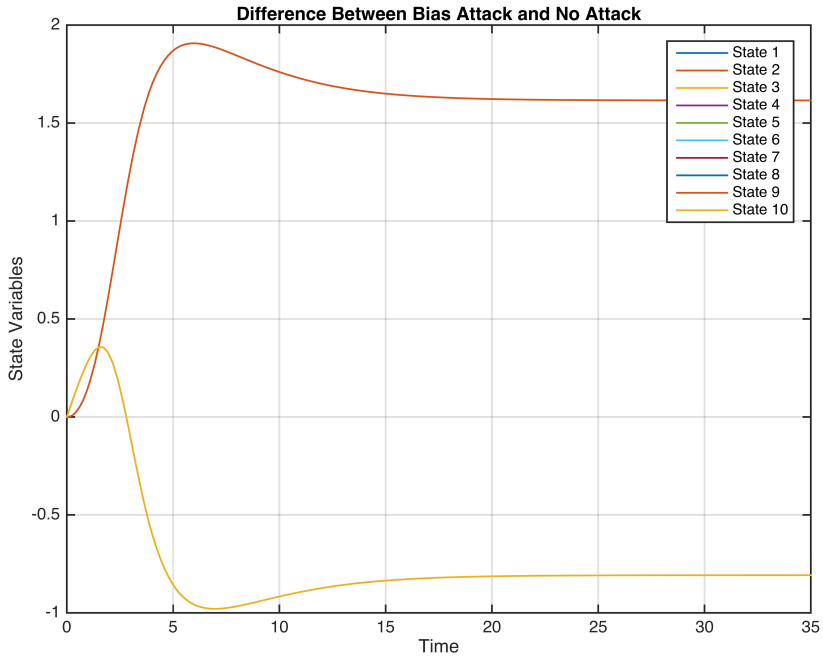


Figure 3.18: Finite Energy Actuator Attack Case: Difference Bias Attack and No Attack.

3.4 Optimal Bounded Actuator Signal Attack

For the bounded actuator signal attacks we have $\Delta_u \in L^\infty([0,h),\mathbb{R}^m)$, and $\|\Delta_x\|_\infty \leq M$. Therefore, we have [83, 85]:

$$\sup_{\|\Delta_u\|_\infty \leq M} J - x^T(t)P(t)x(t) = x^T(t)P(t)x(t) + 2 \sup_{\|\Delta_u\|_\infty \leq M} \int_t^h x^T P B_2 \Delta_u(\tau) d\tau \tag{3.36}$$

The second part of RHS the previous equation satisfies the following [83, 85]:

$$\sup_{\|\Delta_u\|_\infty \leq M} \int_t^h x^T P B_2 \Delta_u(\tau) d\tau \leq M \int_0^h \|B_2^T P x(\tau)\| d\tau \tag{3.37}$$

The upper bound can be satisfied by choosing Δ_u, the actuator attack signal as [83, 85]:

$$\Delta_u(\tau) = M \frac{B_2^T P x(\tau)}{\|B_2^T P x(\tau)\|} \tag{3.38}$$

Therefore, the closed-loop system for the bounded actuator signal attack can be written as [83, 85]:

$$\dot{x}(\tau) = \begin{cases} \big(A - B_2 B_2^T P + M \frac{B_2 B_2^T P x(\tau)}{\|B_2^T P x(\tau)\|}\big) x(\tau), \text{ on the set } E \\ A x(\tau), \text{ if } \; B_2^T P x(\tau) = 0 \;\; (\text{i.e., } \tau \notin E) \end{cases} \tag{3.39}$$

The numerical results for the optimal attack from the attacker's viewpoint are shown in Figs. 3.19 to 3.24. These results are for bounded actuator signal attack such as $\|\Delta_u\|_2 \leq M$. Looking at these figures, it is clear that the optimal attack has a greater effect and the larger damage than the random attack.

In this chapter, a system theoretic analysis of the effect of finite energy and bounded attacks on linear systems with an LQ controller was analyzed and simulated. The optimal finite energy and bounded attacks for the finite horizon LQ problem are presented for the corresponding LQ cost functions. The closed-loop systems under attack are given and a numerical simulation illustrates the effectiveness of the optimal attack is also compared to a random attack. Finally, it is clear that the optimal attack has a greater effect and the larger damage than the random attack.

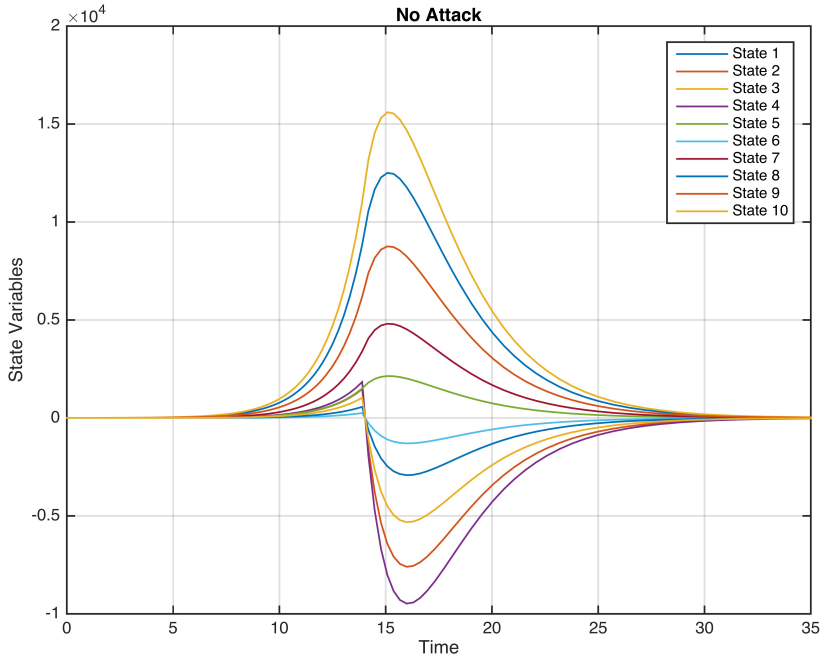


Figure 3.19: Bounded Actuator Signal Attack Case: No Attack.

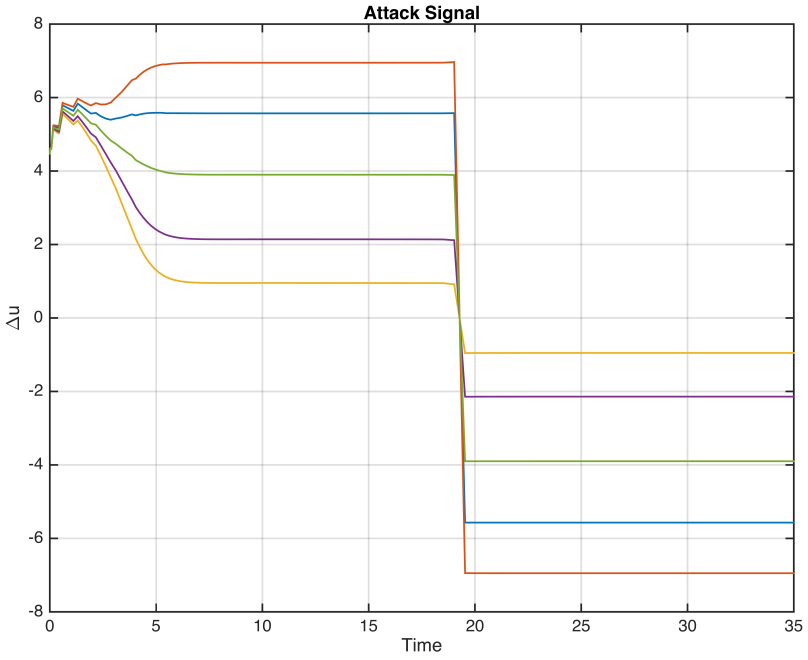


Figure 3.20: Bounded Actuator Signal Attack Case: Attack Signal.

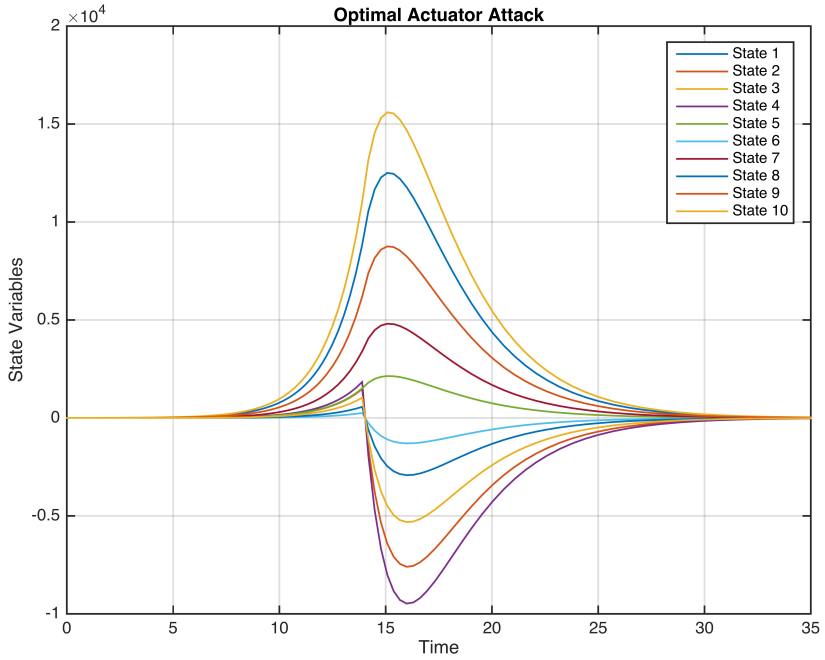


Figure 3.21: Bounded Actuator Signal Attack Case: Optimal Attack.

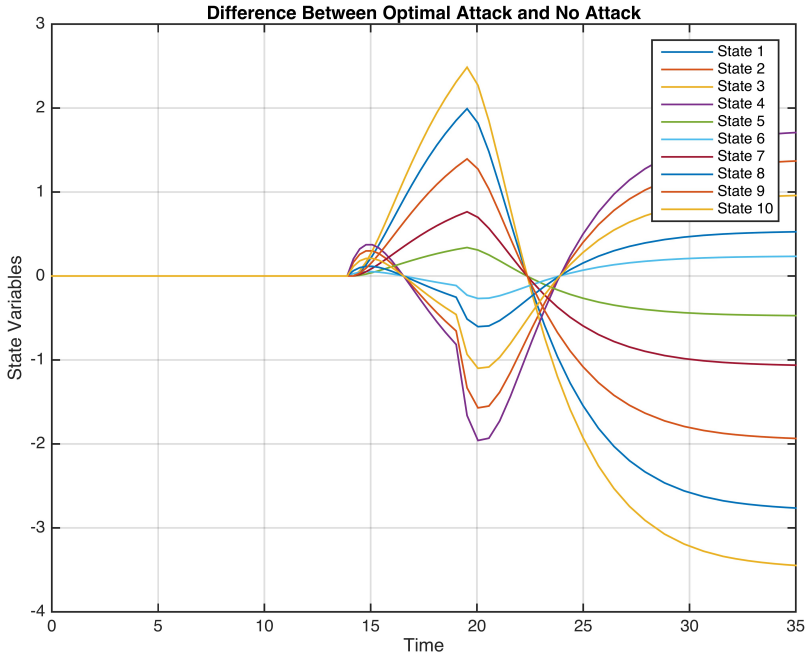


Figure 3.22: Bounded Actuator Signal Attack Case: Difference Optimal Attack and No Attack.

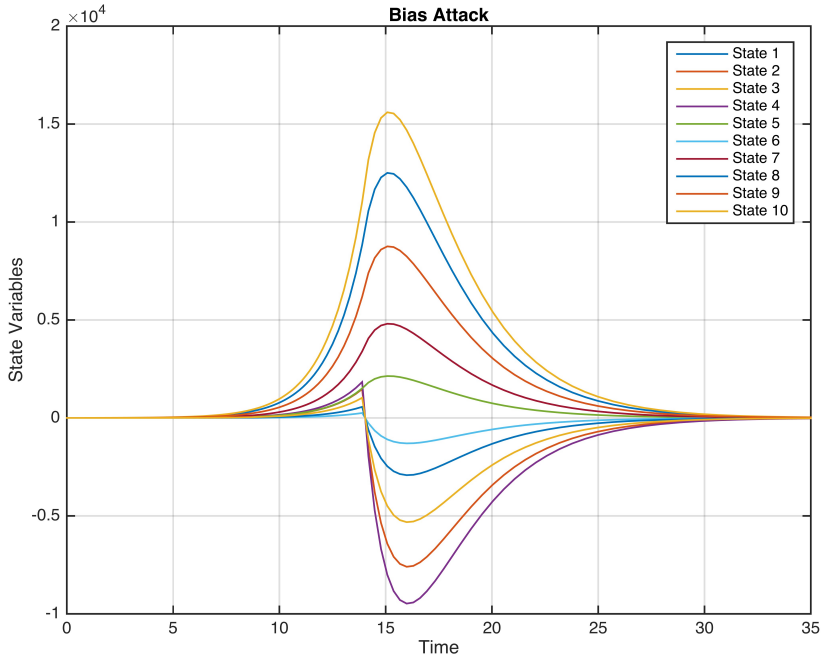


Figure 3.23: Bounded Actuator Signal Attack Case: Bias Attack.

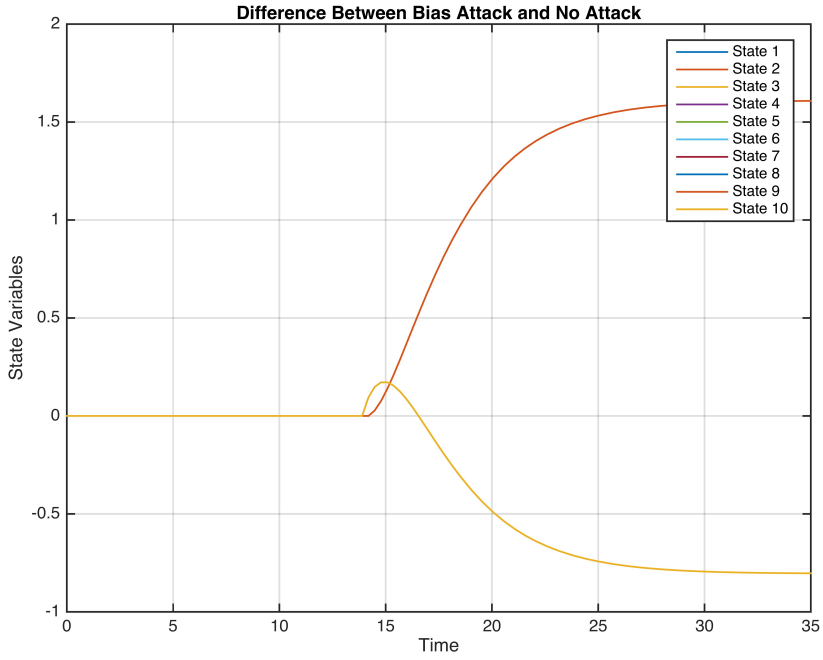


Figure 3.24: Bounded Actuator Signal Attack Case: Difference Bias Attack and No Attack.

Chapter 4

Optimal Attack Strategies on The Kalman Filter

In this chapter, the optimal sensor attack strategies on Kalman filter in which the attacker aims at making the estimation error as large as possible are derived. The attacker needs to determine the attack signal probability distribution which deteriorates the estimation cost function the most. For the Kalman filter, suppose we have the following linear discrete-time system [80]:

$$\begin{aligned} x_k &= F_{k-1}x_{k-1} + G_{k-1}u_{k-1} + w_{k-1} \\ y_k &= H_k x_k + v_k \\ \hat{x}_k^+ &= \hat{x}_k^- + K_k\left(y_k - H_k\hat{x}_k^-\right) \end{aligned} \tag{4.1}$$

The noise processes w_k and v_k are white, zero-mean, uncorrelated processes with known covariance matrices Q_k and R_k. In terms of mathematics, the optimal sensor attack corresponds to design an attack signal such as:

$$\max_{\Delta_{y_k}} \quad \min_{\hat{x}_k} \quad J_k \tag{4.2}$$

where J_k is the estimation cost function, Δ_{y_k} is the optimal random signal attack at time k, $\hat{x}_k$ is the state estimate at time k, x_k is the state vector, u_k is the control vector, y_k is the output vector, F_k is the system matrix, G_k is the input matrix, and H_k is the output matrix. Since random attacker has finite resources, the attack signal is constrained to be finite energy or finite power but

is otherwise random. Let $\hat{x}_k^-$ and $\hat{x}_k^+$ denote *priori* and *posteriori* estimates of x_k, respectively. Specially, $\hat{x}_k^-$ is the estimate using all the measurements before (but not including) time k; and $\hat{x}_k^+$ utilizes all the measurement up to and including time k.

For the sensor attack, the measurement under optimal attack signal y_{Δ_k} takes the following new form:

$$y_{\Delta_k} = H_k x_k + v_k + \Delta_{y_k}, \tag{4.3}$$

To compute the Kalman gain K_k as a function of the optimal signal attack Δ_{y_k}, first we need to compute the mean of the estimation error of the Kalman filter as a function of Δ_{y_k}:

$$E\left(\varepsilon_k^+\right) = E\left(\hat{x}_k - \hat{x}_k^+\right) \tag{4.4}$$

where $\varepsilon_k^+ = \hat{x}_k - \hat{x}_k^+$ is the estimation error. Then the estimator error mean with sensor attack is:

$$\begin{aligned}
E\left(\varepsilon_k^+\right) &= E\left[x_k - \hat{x}_k^- - K_k\left(y_k - H_k\hat{x}_k^-\right)\right] \\
&= E\left[x_k - \hat{x}_k^- - K_k\left(H_k x_k + v_k + \Delta_{y_k} - H_k\hat{x}_k^-\right)\right] \\
&= E\left[x_k - \hat{x}_k^- - K_k H_k\left(\underbrace{x_k - \hat{x}_k^-}_{\varepsilon_k^-}\right) - K_k\left(v_k + \Delta_{y_k}\right)\right] \\
&= E\left[\left(I - K_k H_k\right)\left(\varepsilon_k^-\right) - K_k\left(v_k + \Delta_{y_k}\right)\right] \\
&= \left(I - K_k H_k\right) E\left(\varepsilon_k^-\right) - K_k E\left(v_k\right) - K_k E\left(\Delta_{y_k}\right)
\end{aligned} \tag{4.5}$$

The cost function J_k can be written as:

$$\begin{aligned}
J_k &= E\left[\left(x_1 - \hat{x}_1^+\right)^2\right] + \cdots + E\left[\left(x_n - \hat{x}_n^+\right)^2\right] \\
&= E\left({\varepsilon_{x1,k}^+}^2 + \cdots + {\varepsilon_{xn,k}^+}^2\right) \\
&= E\left({\varepsilon_k^+}^T \varepsilon_k^+\right) \\
&= E\left[Tr\left(\varepsilon_k^+ {\varepsilon_k^+}^T\right)\right] \\
&= TrP_k^+
\end{aligned} \tag{4.6}$$

where P_k^+ is estimation error covariance after the measurement y_k is received.

To minimize the cost function, we need to obtain a recursive formula for P_k as a function of the signal attack Δ_{y_k}:

$$\begin{aligned} P_k^+ &= E\left\{\left[\left(I-K_kH_k\right)\varepsilon_{k-1}^- - K_k\left(v_k+\Delta_{y_k}\right)\right]\left[\cdots\right]^T\right\} \\ &= \left(I-K_kH_k\right)E\left(\varepsilon_{k-1}^-{\varepsilon_{k-1}^-}^T\right)\left(I-K_kH_k\right)^T - K_kE\left(\left(v_k+\Delta_{y_k}\right)\varepsilon_{k-1}^{-T}\right)\left(I-K_kH_k\right)^T \\ &- \left(I-K_kH_k\right)E\left(\varepsilon_{k-1}^-\left(v_k+\Delta_{y_k}\right)^T\right)K_k^T - K_kE\left(\left(v_k+\Delta_{y_k}\right)\left(v_k+\Delta_{y_k}\right)^T\right)K_k^T \end{aligned} \tag{4.7}$$

The time-update equation of P_k can also take the following more general equation [**?**]:

$$P_k^- = F_{k-1}P_{k-1}^+F_{k-1}^T + Q_{k-1} \tag{4.8}$$

After obtaining the the estimation error covariance as a function of the signal attack Δ_{y_k} (4.7), in the next section we will start with the scalar case, where two main propositions are considered. We start with the signal attack which is independent of the estimation error for both the optimal attack signal and its mitigation, then we do the same for the case where the attack signal is dependent of the estimation error.

4.1 Optimal Attack on Kalman Filter Scalar Case

The previous equation (4.7) can be simplified since measurement noise is independent of the estimation error and the attack signal:

$$\begin{aligned} P_k^+ &= \left(I-K_kH_k\right)P_k^-\left(I-K_kH_k\right)^T + KD_kK_k^T - K_kE\left[\Delta_{y_k}\varepsilon_k^{-T}\right]\left(I-K_kH_k\right)^T + KR_kK_k^T \\ &- \left(I-K_kH_k\right)E\left[\varepsilon_k^-\Delta_{y_k}^T\right]K_k^T \end{aligned} \tag{4.9}$$

where $R_k = E(v_k{v_k}^T)$ and $D_k = E(\Delta_{y_k}{\Delta_{y_k}}^T)$.

Therefore, the optimization (4.2) reduces to find the attack signal Δ_{y_k} which maximizes the following:

$$\begin{aligned} \max_{\Delta_{y_k}}\left\{\min_{K_k} Tr\left(P_k^+\right)\right\} = \max_{\Delta_{y_k}}\Big\{\min_{K_k} Tr\Big(&\left(I-K_kH_k\right)P_k^-\left(I-K_kH_k\right)^T \\ &+KD_kK_k^T + KR_kK_k^T - \left(I-K_kH_k\right)E\left[\varepsilon_k^-\Delta_{y_k}^T\right]K_k^T\Big\}\end{aligned} \tag{4.10}$$

In the following, two main propositions are given considering whether the estimation error depends on the attack signal or not. We will start from situation, where the estimation error is **independent** of the attack signal. Then we will work on the general case, where the estimation error is **dependent** of the attack signal Δ_y at time k.

4.1.1 Optimal Attack with Δ_{y_k} independent of the Estimation Error

To simplify equation (4.9) we can consider that the estimation error at time $k-1$ is independent of the measurement noise ε_k^+, and attack signal Δ_y at time k.

Proposition 4.1.1. *The optimal estimation covariance error as a function of the attack signal is given by:*

$$P_k^+ = P_k^- - P_k^- H_k^T \left(H_k P_k^- H_k^T + R_k + D_k\right)^{-1} H_k P_k^-.$$

Proof. We rewrite (4.9) as

$$P_k^+ \quad = \quad (I - K_k H_k)\, E\left(\varepsilon_k^- \varepsilon_k^{-T}\right) (I - K_k H_k)^T + K_k E\left((v_k + \Delta_{y_k})\,(v_k + \Delta_{y_k})^T\right) K_k^T \quad (4.11)$$

Moreover, we assume the measurement noise v_k is independent of the attack signal Δ_k at time k. Therefore, $E\left(v_k \Delta_{y_k}^T\right) = E\left(\Delta_{y_k} v_k^T\right) = 0$ and (4.11) could be further simplified to:

$$P_k^+ \quad = \quad (I - K_k H_k)\, P_k^- (I - K_k H_k)^T + K_k D_k K_k^T + K_k R_k K_k^T, \quad (4.12)$$

where $R_k = E(v_k {v_k}^T)$ and $D_k = E(\Delta_{y_k} {\Delta_{y_k}}^T)$.

With this expression of J_k in hand, we are ready to compute the minimum of J_k by taking its derivative as:

$$\frac{\partial J_k}{\partial K_k} = 2\,(I - K_k H_k)\, P_k^- \left(-H_k^T\right) + 2 K_k R_k + 2 K_k D_k \quad (4.13)$$

In order to find K_k, we set the previous equation equal to zero and compute K_k

$$\begin{aligned} K_k\left(R_k + D_k\right) &= \left(I - K_k H_k\right) P_k^- \left(H_k^T\right) \\ K_k\left(R_k + D_k\right) &= P_k^- H_k^T - K_k H_k P_k^- H_k^T \\ K_k\left(R_k + D_k + H_k P_k^- H_k^T\right) &= P_k^- H_k^T \end{aligned} \tag{4.14}$$

Therefore, K_k should satisfy:

$$K_k = P_k^- H_k^T \left(R_k + D_k + H_k P_k^- H_k^T\right)^{-1} \tag{4.15}$$

Let $S_k = R_k + D_k + H_k P_k^- H_k^T$, then K_k is:

$$K_k = P_k^- H_k^T S_k^{-1} \tag{4.16}$$

Substituting K_k in (4.9) we get:

$$\begin{aligned} P_k^+ &= P_k^- - P_k^- H_k^T S_k^{-1} H_k P_k^- - P_k^- H_k^T S_k^{-1} H_k P_k^- + P_k^- H_k^T S_k^{-1} H_k P_k^- H_k^T S_k^{-1} H_k P_k^- \\ &+ P_k^- H_k^T S_k^{-1} R_k S_k^{-1} H_k P_k^- + P_k^- H_k^T S_k^{-1} D_k S_k^{-1} H_k P_k^- \end{aligned} \tag{4.17}$$

Combining the last three terms in the previous equation,

$$\begin{aligned} P_k^+ &= P_k^- - 2P_k^- H_k^T S_k^{-1} H_k P_k^- + P_k^- H_k^T S_k^{-1} S_k S_k^{-1} H_k P_k^- \\ &= P_k^- - 2P_k^- H_k^T S_k^{-1} H_k P_k^- + P_k^- H_k^T S_k^{-1} H_k P_k^- \\ &= P_k^- - P_k^- H_k^T S_k^{-1} H_k P_k^- \end{aligned} \tag{4.18}$$

Finally, we can rewrite P_k^+ as:

$$P_k^+ = P_k^- - P_k^- H_k^T \left(H_k P_k^- H_k^T + R_k + D_k\right)^{-1} H_k P_k^- \tag{4.19}$$

□

Proposition 4.1.1 shows how the covariance matrix is affected by the sensor attack signals. It follows then that the worst case attack scenario corresponds to maximizing the following cost function:

$$J_k = Tr(P_k^+) \tag{4.20}$$

The optimal attack strategy is to maximize J_k as summarized in the following problem:

Problem 4.1.2.

$$\begin{aligned} &\max J_k, \\ &\text{subject to } Tr\left[E\left[\Delta_{y_k}\Delta_{y_k}^T\right]\right] \leq M \end{aligned} \tag{4.21}$$

In order to solve the optimization problem (4.1.2), the Lagrangian is first defined as:

$$L = Tr\left[P_k^- - P_k^- H_k^T\left(H_k P_k^- H_k^T + R_k + D_k\right)^{-1} H_k P_k^-\right] + \lambda\left(M - Tr\left[E\left[\Delta_{y_k}\Delta_{y_k}^T\right]\right]\right) \tag{4.22}$$

To avoid cumbersome computation, the scalar case is only considered:

$$\begin{aligned} L &= P_k^- - {P_k^-}^2 H_k^2\left(H_k^2 P_k^- + R_k + D_k\right)^{-1} + \lambda\left(M_k - E\left[\Delta_{y_k}^2\right]\right) \\ &= P_k^- - {P_k^-}^2 H_k^2 S1_k^{-1} + \lambda\left(M - E\left[\Delta_{y_k}^2\right]\right) \end{aligned} \tag{4.23}$$

where $S1_k^{-1} = H_k^2 P_k^- + R_k + D_k$

$$\begin{aligned} \frac{\partial L}{\partial \Delta_{y_k}} &= 2{P_k^-}^2 H_k^2 E\left[\Delta_{y_k}\right] S1_k^{-2} - 2\lambda E\left[\Delta_{y_k}\right] = 0 \\ \frac{\partial L}{\partial \lambda} &= M - E\left[\Delta_{y_k}^2\right] = 0 \end{aligned} \tag{4.24}$$

From the previous equation (4.24) the following identities hold:

$$\begin{cases} \lambda = \frac{P_k^{-2} H_k^2}{\left(H_k^2 P_k^- + R_k + E\left[\Delta_{y_k}^2\right]\right)^2} \\ E\left[\Delta_{y_k}^2\right] = M \end{cases} \tag{4.25}$$

Replacing P_k^- from equation (4.8), yields:

$$\begin{cases} \lambda = \frac{\left(F_{k-1} P_{k-1}^+ F_{k-1}^T + Q_{k-1}\right)^2 H_k^2}{\left(H_k^2 \left(F_{k-1} P_{k-1}^+ F_{k-1}^T + Q_{k-1}\right) + R_k + E\left[\Delta_{y_k}^2\right]\right)^2} \\ E\left[\Delta_{y_k}^2\right] = M \end{cases} \tag{4.26}$$

To show that this indeed corresponds to a maximum, the 2^{nd} derivative of L is computed:

$$\begin{aligned} \frac{\partial^2 L}{\partial \Delta_{y_k}^2} &= -8E[\Delta_{y_k}]^2 P_k^{-2} H_k^2 S1_k^{-3} \\ &- 2\lambda + 2P_k^{-2} H_k^2 S1_k^{-2} \end{aligned} \tag{4.27}$$

Replacing λ in (4.27):

$$\frac{\partial^2 L}{\partial \Delta_{y_k}^2} = -8E[\Delta_{y_k}]^2 P_k^{-2} H_k^2 S1_k^{-3} < 0 \tag{4.28}$$

Therefore, any signal Δ_{y_k} which satisfies (4.26) is optimal.

4.1.2 Optimal Attack with Δ_{y_k} dependent on the Estimation Error

Next, we will consider the general case, the estimation error at time $k-1$ is independent of the measurement noise, but is **dependent** of the attack signal Δ_y at time k.

Proposition 4.1.3. *The optimal sensor attack signal for the case, where the estimation error at time $k-1$ is **dependent** of the attack signal Δ_y at time k, comes with estimation error:*

$$\begin{aligned}
P_k^+ &= P_k^- - P_k^- H_k^T S_k^{-1} H_k P_k^- - P_k^- H_k^T S_k^{-1} E\left[\varepsilon_{x,k-1}\Delta_{y_k}^T\right] \\
&- P_k^- H_k^T S_k^{-1} H_k P_k^- - E\left[\varepsilon_{x,k-1}\Delta_{y_k}^T\right] S_k^{-1} H_k P_k^- - P_k^- H_k^T S_k^{-1} E\left[\Delta_{y_k}\varepsilon_{x,k-1}^T\right] \\
&- E\left[\varepsilon_{x,k-1}\Delta_{y_k}^T\right] S_k^{-1} E\left[\Delta_{y_k}\varepsilon_{x,k-1}^T\right] - E\left[\varepsilon_{x,k-1}\Delta_{y_k}^T\right] S_k^{-1} H_k P_k^- \\
&- E\left[\varepsilon_{x,k-1}\Delta_{y_k}^T\right] S_k^{-1} E\left[\varepsilon_{x,k-1}\Delta_{y_k}^T\right] + \left[P_k^- H_k^T S_k^{-1} + E\left[\varepsilon_{x,k-1}\Delta_{y_k}^T\right] S_k^{-1}\right] S_k \\
&\quad \left[S_k^{-1} H_k P_k^- + S_k^{-1} E\left[\varepsilon_{x,k-1}\Delta_{y_k}^T\right]\right]
\end{aligned}$$

Proof. Instead of (4.11), we get:

$$\begin{aligned}
P_k^+ &= (I - K_k H_k) P_k^- (I - K_k H_k)^T + K R_k K_k^T - K_k E\left[\Delta_{y_k}\varepsilon_k^{-T}\right](I - K_k H_k)^T + K D_k K_k^T \\
&- (I - K_k H_k) E\left[\varepsilon_k^- \Delta_{y_k}^T\right] K_k^T
\end{aligned} \tag{4.29}$$

where $R_k = E(v_k {v_k}^T)$ and $D_k = E(\Delta_{y_k} {\Delta_{y_k}}^T)$. Recall these three basic trace identities:

$$\begin{aligned}
\frac{\partial (TrAB)}{\partial A} &= B^T \\
\frac{\partial \left(TrABA^T\right)}{\partial A} &= AB^T + AB \\
\frac{\partial \left(TrBA^T\right)}{\partial A} &= B
\end{aligned} \tag{4.30}$$

We can rewrite (4.29) as:

$$\begin{aligned}
P_k^+ &= P_k^- - P_k^- H_k^T K_k^T - K_k H_k P_k^- + K_k H_k P_k^- H_k^T K_k^T - K_k E\left[\Delta_{y_k}\varepsilon_k^{-T}\right] + K_k E\left[\Delta_{y_k}\varepsilon_k^{-T}\right] H_k^T K_k^T \\
&- E\left[\varepsilon_k^- \Delta_{y_k}^T\right] K_k^T + K_k H_k E\left[\varepsilon_k^- \Delta_{y_k}^T\right] K_k^T + K R_k K_k^T + K D_k K_k^T
\end{aligned} \tag{4.31}$$

Similar as in the first case, we need to characterize the derivative of J_k with respect to K_k. We can use the relationships in (4.30) to solve for $\frac{\partial\left(TrP_k^+\right)}{\partial K_k}$:

$$\begin{aligned}\frac{\partial\left(TrP_k^+\right)}{\partial K_k} &= \frac{\partial J_k}{\partial K_k} = -P_k^- H_k^T - P_k^- H_k^T + K_k H_k P_k^- H_k^T + K_k H_k P_k^- H_k^T - E\left[\Delta_{y_k}\varepsilon_k^{-T}\right]^T \\ &+ K_k E\left[\Delta_{y_k}\varepsilon_k^{-T}\right] H_k^T + K_k E\left[\varepsilon_k^- \Delta_{y_k}^T\right]^T H_k^T + K_k H_k E\left[\Delta_{y_k}\varepsilon_k^{-T}\right]^T + 2K_k R_k \ + 2K_k D_k \\ &- E\left[\varepsilon_k^- \Delta_{y_k}^T\right] + K_k H_k E\left[\varepsilon_k^- \Delta_{y_k}^T\right] \end{aligned} \tag{4.32}$$

Therefore,

$$\begin{aligned}\frac{\partial J_k}{\partial K_k} &= -2P_k^- H_k^T + 2K_k H_k P_k^- H_k^T - 2E\left[\varepsilon_k^- \Delta_{y_k}^T\right] \ + 2K_k H_k E\left[\varepsilon_k^- \Delta_{y_k}^T\right] + 2K_k D_k + 2K_k R_k \\ &+ 2K_k E\left[\Delta_{y_k}\varepsilon_k^{-T}\right] H_k^T \end{aligned} \tag{4.33}$$

In order to find the value of K_k, we set the previous equation equal to zero and compute K_k

$$K_k\left[H_k P_k^- H_k^T + E\left[\varepsilon_k^- \Delta_{y_k}^T\right] H_k^T + H_k E\left[\varepsilon_k^- \Delta_{y_k}^T\right] + R_k + D_k\right] = \left(P_k^- H_k^T + E\left[\varepsilon_k^- \Delta_{y_k}^T\right]\right) \tag{4.34}$$

Therefore, the optimal gain K_k is:

$$K_k = \left(P_k^- H_k^T + E\left[\varepsilon_k^- \Delta_{y_k}^T\right]\right)\left[H_k P_k^- H_k^T + E\left[\varepsilon_k^- \Delta_{y_k}^T\right] H_k^T + H_k E\left[\varepsilon_k^- \Delta_{y_k}^T\right] + R_k + D_k\right]^{-1} \tag{4.35}$$

Let $S_k = H_k P_k^- H_k^T + E\left[\varepsilon_k^- \Delta_{y_k}^T\right] H_k^T + H_k E\left[\varepsilon_k^- \Delta_{y_k}^T\right] + R_k + D_k$, then K_k becomes:

$$K_k = \left(P_k^- H_k^T + E\left[\varepsilon_k^- \Delta_{y_k}^T\right]\right) S_k^{-1} \tag{4.36}$$

We can rearrange P_k^+ from equation (4.31):

$$\begin{aligned} P_k^+ &= P_k^- - P_k^- H_k^T K_k^T - K_k H_k P_k^- - K_k E\left[\Delta_{y_k}\varepsilon_k^{-T}\right] - E\left[\varepsilon_k^- \Delta_{y_k}^T\right] K_k^T \ + K_k S_k K_k^T \\ &= P_k^- - P_k^- H_k^T K_k^T - K_k H_k P_k^- - K_k E\left[\Delta_{y_k}\varepsilon_k^{-T}\right] - E\left[\varepsilon_k^- \Delta_{y_k}^T\right] K_k^T + K_k S_k K_k^T \end{aligned} \tag{4.37}$$

Therefore, the previous equation (4.37) becomes:

$$\begin{aligned} P_k^+ &= P_k^- - P_k^- H_k^T S_k^{-1} H_k P_k^- - P_k^- H_k^T S_k^{-1} E\left[\Delta_{y_k} \varepsilon_k^{-T}\right] \\ &- E\left[\varepsilon_k^- \Delta_{y_k}^T\right] S_k^{-1} H_k P_k^- - E\left[\varepsilon_k^- \Delta_{y_k}^T\right] S_k^{-1} E\left[\Delta_{y_k} \varepsilon_k^{-T}\right] \\ &= P_k^- - P_k^- H_k^T S_k^{-1}\left[H_k P_k^- + E\left[\Delta_{y_k} \varepsilon_k^{-T}\right]\right] \\ &- E\left[\varepsilon_k^- \Delta_{y_k}^T\right] S_k^{-1}\left[H_k P_k^- + E\left[\Delta_{y_k} \varepsilon_k^{-T}\right]\right] \\ &= P_k^- - \left[P_k^- H_k^T + E\left[\varepsilon_k^- \Delta_{y_k}^T\right]\right] S_k^{-1} \\ &\cdot \left[H_k P_{k-1} + E\left[\Delta_{y_k} \varepsilon_k^{-T}\right]\right] \end{aligned} \tag{4.38}$$

Substituting K_k in :

$$\begin{aligned} P_k^+ &= P_k^- - P_k^- H_k^T S_k^{-1} H_k P_k^- - P_k^- H_k^T S_k^{-1} E\left[\varepsilon_{x,k-1} \Delta_{y_k}^T\right] \\ &- P_k^- H_k^T S_k^{-1} H_k P_k^- - E\left[\varepsilon_{x,k-1} \Delta_{y_k}^T\right] S_k^{-1} H_k P_k^- - P_k^- H_k^T S_k^{-1} E\left[\Delta_{y_k} \varepsilon_{x,k-1}^T\right] \\ &- E\left[\varepsilon_{x,k-1} \Delta_{y_k}^T\right] S_k^{-1} E\left[\Delta_{y_k} \varepsilon_{x,k-1}^T\right] - E\left[\varepsilon_{x,k-1} \Delta_{y_k}^T\right] S_k^{-1} H_k P_k^- \\ &- E\left[\varepsilon_{x,k-1} \Delta_{y_k}^T\right] S_k^{-1} E\left[\varepsilon_{x,k-1} \Delta_{y_k}^T\right] + \left[P_k^- H_k^T S_k^{-1} + E\left[\varepsilon_{x,k-1} \Delta_{y_k}^T\right] S_k^{-1}\right] S_k \\ &\left[S_k^{-1} H_k P_k^- + S_k^{-1} E\left[\varepsilon_{x,k-1} \Delta_{y_k}^T\right]\right] \end{aligned} \tag{4.39}$$

□

So we can write the cost function as:

$$J_k = Tr\left(P_k^+\right) = Tr\left[P_k^- - \left[P_k^- H_k^T + E\left[\varepsilon_{x,k-1} \Delta_{y_k}^T\right]\right] S_k^{-1}\left[H_k P_k^- + E\left[\Delta_{y_k} \varepsilon_{x,k-1}^T\right]\right]\right] \tag{4.40}$$

Now the optimal attack problem is converted into an optimization problem described as follows:

Problem 4.1.4.

$$\begin{aligned} &\max J_k, \\ &\text{subject to } Tr\left[E\left[\Delta_{y_k} \Delta_{y_k}^T\right]\right] \leq M_k \end{aligned} \tag{4.41}$$

where M_k is to be determined.

In order to solve the optimization problem (4.1.4), we define the Lagrangian as:

$$L = Tr\left[P_k^- - \left[P_k^- H_k^T + E\left[\varepsilon_{x,k-1}\Delta_{y_k}^T\right]\right] S_k^{-1}\left[H_k P_k^- + E\left[\Delta_{y_k}\varepsilon_{x,k-1}^T\right]\right]\right] + \lambda\left(M_k - Tr\left[E\left[\Delta_{y_k}\Delta_{y_k}^T\right]\right]\right) \tag{4.42}$$

where λ is the Lagrange multiplier. The previous equation (4.42) can be expanded to the following form:

$$\begin{aligned} L &= Tr\left(P_k^-\right) - Tr\left(P_k^- H_k^T S_k^{-1} H_k P_k^-\right) - Tr\left(P_k^- H_k^T S_k^{-1} E\left[\Delta_{y_k}\varepsilon_{x,k-1}^T\right]\right) \\ &- Tr\left(E\left[\varepsilon_{x,k-1}\Delta_{y_k}^T\right] S_k^{-1} H_k P_k^-\right) - Tr\left(E\left[\varepsilon_{x,k-1}\Delta_{y_k}^T\right] S_k^{-1} E\left[\Delta_{y_k}\varepsilon_{x,k-1}^T\right]\right) \\ &+ \lambda\left(M_k - Tr\left[E\left[\Delta_{y_k}\Delta_{y_k}^T\right]\right]\right) \end{aligned} \tag{4.43}$$

Rearranging the previous equation:

$$\begin{aligned} L &= Tr\left(P_k^-\right) - Tr\left(S_k^{-1} H_k P_k^- P_k^- H_k^T\right) - Tr\left(S_k^{-1} E\left[\Delta_{y_k}\varepsilon_{x,k-1}^T\right] P_k^- H_k^T\right) \\ &- Tr\left(S_k^{-1} H_k P_k^- E\left[\varepsilon_{x,k-1}\Delta_{y_k}^T\right]\right) - Tr\left(S_k^{-1} E\left[\Delta_{y_k}\varepsilon_{x,k-1}^T\right] E\left[\varepsilon_{x,k-1}\Delta_{y_k}^T\right]\right) \\ &+ \lambda\left(M_k - Tr\left[E\left[\Delta_{y_k}\Delta_{y_k}^T\right]\right]\right) \end{aligned} \tag{4.44}$$

To avoid cumbersome computation, the scalar case is only considered:

$$L = P_k^- - P_k^{-2} H_k^2 S_k^{-1} - E[\varepsilon_{x,k-1}\Delta_{y_k}]^2 S_k^{-1} - 2H_k P_k^- S_k^{-1} E\left[\varepsilon_{x,k-1}\Delta_{y_k}\right] + \lambda\left(M - E\left[\Delta_{y_k}^2\right]\right) \tag{4.45}$$

where $S_k = H_k^2 P_k^- + 2E\left[\varepsilon_{x,k-1}\Delta_{y_k}\right] H_k + R_k + E\left[\Delta_{y_k}^2\right]$ Differentiating yields:

$$\begin{aligned} \frac{\partial L}{\partial \Delta_{y_k}} &= 2E[\varepsilon_{x,k-1}\Delta_{y_k}]^2 S_k^{-2} E\left[\Delta_{y_k}\right] + 2P_k^{-2} H_k^2 S_k^{-2} E\left[\Delta_{y_k}\right] - 2\lambda E\left[\Delta_{y_k}\right] \\ &+ 4H_k P_k^- S_k^{-2} E\left[\varepsilon_{x,k-1}\Delta_{y_k}\right] E\left[\Delta_{y_k}\right] = 0 \\ \frac{\partial L}{\partial \lambda} &= M_k - E\left[\Delta_{y_k}^2\right] = 0 \end{aligned} \tag{4.46}$$

Differentiating S_k yields:

$$\frac{\partial S_k}{\partial \Delta_{y_k}} = 2E\left[\Delta_{y_k}\right] \tag{4.47}$$

From the previous equation (4.46) we have λ equal to the following:

$$\lambda = \left[S_k^{-1}\left(H_k P_k^- + E\left[\varepsilon_{x,k-1}\Delta_{y_k}\right]\right)\right]^2 \tag{4.48}$$

Factorizing the equation for $\frac{\partial L}{\partial \Delta_{y_k}}$ (4.46) yields to the following:

$$\begin{aligned}\left(2S_k^{-2}E\left[\Delta_{y_k}\right]\right)\left(P_k^{-2}H_k^2 + E[\varepsilon_{x,k-1}\Delta_{y_k}]^2 + 2H_k P_k^- E\left[\varepsilon_{x,k-1}\Delta_{y_k}\right] - \lambda S_k^2\right) &= 0\\ \left(2S_k^{-2}E\left[\Delta_{y_k}\right]\right)\left[\left(H_k P_k^- + E\left[\varepsilon_{x,k-1}\Delta_{y_k}\right]\right) - \left(\sqrt{\lambda}S_k\right)\right]\left[\left(H_k P_k^- + E\left[\varepsilon_{x,k-1}\Delta_{y_k}\right]\right) + \left(\sqrt{\lambda}S_k\right)\right] &= 0\end{aligned} \tag{4.49}$$

Therefore, we have the optimal attack Δ_{y_k} satisfies the following identities:

$$\begin{cases} E\left[\Delta_{y_k}^2\right] = M_k, and \\ E\left[\Delta_{y_k}\right] = 0, or \\ E\left[\varepsilon_{x,k-1}\Delta_{y_k}\right] = -H_k P_k^- - \sqrt{\lambda}S_k, or \\ E\left[\varepsilon_{x,k-1}\Delta_{y_k}\right] = -H_k P_k^- + \sqrt{\lambda}S_k \end{cases} \tag{4.50}$$

To show that this indeed corresponds to a maximum, the 2^{nd} derivative of L is computed:

$$\begin{aligned}\frac{\partial^2 L}{\partial \Delta_{y_k}^2} &= 2E[\varepsilon_{x,k-1}\Delta_{y_k}]^2 S_k^{-2} - 8P_k^{-2}H_k^2 S_k^{-3}E[\Delta_{y_k}]^2 - 16H_k P_k^- S_k^{-3}E\left[\varepsilon_{x,k-1}\Delta_{y_k}\right]E[\Delta_{y_k}]^2 \\ &+ 4H_k P_k^- S_k^{-2}E\left[\varepsilon_{x,k-1}\Delta_{y_k}\right] + 2P_k^{-2}H_k^2 S_k^{-2} - 8E[\varepsilon_{x,k-1}\Delta_{y_k}]^2 S_k^{-3}E[\Delta_{y_k}]^2 - 2\lambda\end{aligned} \tag{4.51}$$

$$\left.\frac{\partial^2 L}{\partial \Delta_{y_k}^2}\right|_{E[\Delta_{y_k}]=0} = 2P_k^{-2}H_k^2 S_k^{-2} + 2E[\varepsilon_{x,k-1}\Delta_{y_k}]^2 S_k^{-2} + 4H_k P_k^- S_k^{-2}E\left[\varepsilon_{x,k-1}\Delta_{y_k}\right] - 2\lambda \tag{4.52}$$

Replacing λ in the previous equation (4.52) and $E\left[\Delta_{y_k}\right]=0$ gives the following:

$$\begin{aligned}\left.\frac{\partial^2 L}{\partial \Delta_{y_k}^2}\right|_{E\left[\Delta_{y_k}\right]=0} &= 2P_k^{-2}H_k^2S_k^{-2}-2P_k^{-2}H_k^2S_k^{-2}-4H_kP_k^-S_k^{-2}E\left[\varepsilon_{x,k-1}\Delta_{y_k}\right]\\ - \quad & 2E[\varepsilon_{x,k-1}\Delta_{y_k}]^2S_k^{-2}+4H_kP_k^-S_k^{-2}E\left[\varepsilon_{x,k-1}\Delta_{y_k}\right]\ +2E[\varepsilon_{x,k-1}\Delta_{y_k}]^2S_k^{-2}=0\end{aligned} \tag{4.53}$$

Therefore, the case for $E\left[\Delta_{y_k}\right]=0$ is indecisive. Replacing $E\left[\varepsilon_{x,k-1}\Delta_{y_k}\right]=-H_kP_k^--\sqrt{\lambda}S_k$ in the previous equation (4.52) gives the following:

$$\begin{aligned}\left.\frac{\partial^2 L}{\partial \Delta_{y_k}^2}\right|_{E\left[\varepsilon_{x,k-1}\Delta_{y_k}\right]=-H_kP_k^--\sqrt{\lambda}S_k} &= 2P_k^{-2}H_k^2S_k^{-2}-8P_k^{-2}H_k^2S_k^{-3}E[\Delta_{y_k}]^2\ -4\sqrt{\lambda}H_kP_k^-S_k^{-1}\\ + \quad & 2P_k^{-2}H_k^2S_k^{-2}+2\lambda+4\sqrt{\lambda}H_kP_k^-S_k^{-1}-16\sqrt{\lambda}H_kP_k^-S_k^{-2}E[\Delta_{y_k}]^2-2\lambda-4P_k^{-2}H_k^2S_k^{-2}\\ - \quad & 8\lambda S_k^{-1}E[\Delta_{y_k}]^2\ -8P_k^{-2}H_k^2S_k^{-3}E[\Delta_{y_k}]^2\ +16P_k^{-2}H_k^2S_k^{-3}E[\Delta_{y_k}]^2+16\sqrt{\lambda}H_kP_k^-S_k^{-2}E[\Delta_{y_k}]^2\end{aligned} \tag{4.54}$$

After further simplification of the previous equation (4.54) gives the following:

$$\left.\frac{\partial^2 L}{\partial \Delta_{y_k}^2}\right|_{E\left[\varepsilon_{x,k-1}\Delta_{y_k}\right]=-H_kP_k^--\sqrt{\lambda}S_k}=-8\lambda S_k^{-1}E[\Delta_{y_k}]^2<0 \tag{4.55}$$

Replacing λ and $E\left[\Delta_{y_k}^2\right]=M_k$ in the previous equation (4.55):

$$\left.\frac{\partial^2 L}{\partial \Delta_{y_k}^2}\right|_{E\left[\varepsilon_{x,k-1}\Delta_{y_k}\right]=-H_kP_k^--\sqrt{\lambda}S_k}=-8M_k\left[S_k^{-1}\left(H_kP_k^-+E\left[\varepsilon_{x,k-1}\Delta_{y_k}\right]\right)\right]^2S_k^{-1}<0 \tag{4.56}$$

If showing that (4.50) corresponds to a maximum the following holds:

$$\begin{aligned}E\left[\varepsilon_{x,k-1}\Delta_{y_k}\right] &= -H_kP_k^--\sqrt{\lambda}S_k\\ &= -H_kP_k^--S_k^{-1}S_k\left(H_kP_k^-+E\left[\varepsilon_{x,k-1}\Delta_{y_k}\right]\right)\\ &= -H_kP_k^-\end{aligned} \tag{4.57}$$

We have then for the optimal signal attack Δy_k:

$$\begin{cases} E\left[\varepsilon_{x,k-1}\Delta_{y_k}\right] = -H_k P_k^- \\ E\left[\Delta_{y_k}^2\right] = M_k \\ E\left[\varepsilon_{x,k-1}^2\right] = P_k^- \\ E\left[\varepsilon_{x,k-1}\right] = 0 \end{cases} \tag{4.58}$$

The system of equations (4.58) has many solutions in Δ_{y_k}. That is any signal Δ_{y_k} which satisfies (4.58) is optimal. To determine an optimal signal attack it suffices to exhibit one solution; A possible choice is:

$$\Delta_{y_k} = -\varepsilon_{x,k-1} H_k \tag{4.59}$$

Therefore, we have:

$$\Delta_{y_k}^2 = \varepsilon_{x,k-1}^2 H_k^2 \tag{4.60}$$

Then,

$$E\left[\Delta_{y_k}^2\right] = E\left[\varepsilon_{x,k-1}^2 H_k^2\right] = H_k^2 P_k^- = M_k \tag{4.61}$$

Numerical Example

We give an example to illustrate the performance of Proposition 4.1.3. As mentioned before, for simplicity, we only deal with the scalar case here. The derivation can be easily generalized to more complicated vector case.

We took an example from Matlab where the Kalman filter is used for object tracking to predict an object's future location, to account for noise in an object's detected location. For this simple scalar case, the state-transition matrix is also a scalar which is chosen to be 0.5. Without loss of generality, assume the white noise w_k and v_k are zero-mean with both standard deviations equal 0.1. Notice that these two noise processes are uncorrelated. We set the initial location to be 1.

Applying the result from Proposition 4.1.3, we could generate the optimal attack as shown in Fig. 4.1. The state trajectory under optimal attack is presented in Fig. 4.2. As expected, we see a large deviation from the original estimation trajectory after adding the optimal attack signal. To better illustrate the optimality of the proposed optimal attack, we compare it with another random

attack. It can be seen from Fig. 4.3 that the trajectory deviation is much larger under our optimal attack.

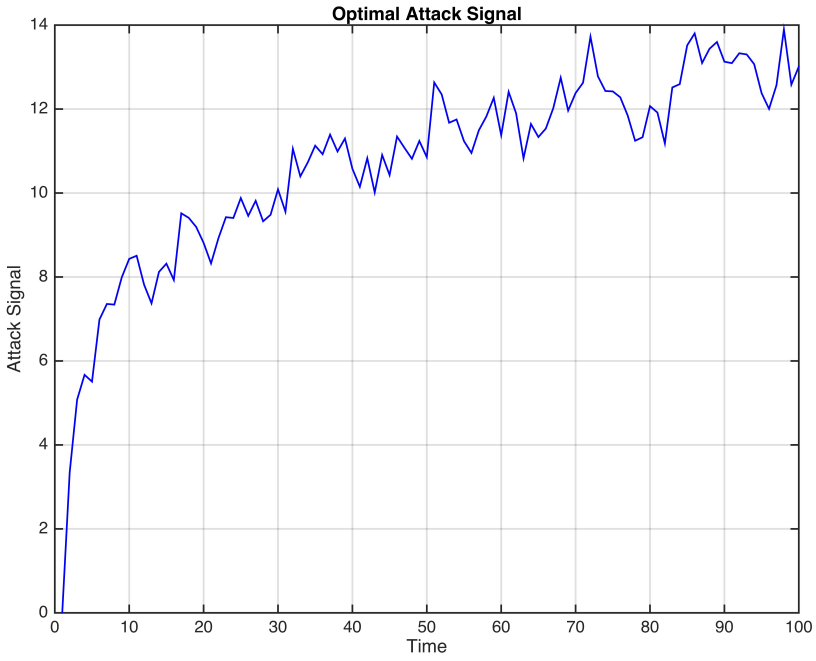


Figure 4.1: Optimal Attack Signal on the Kalman Filter

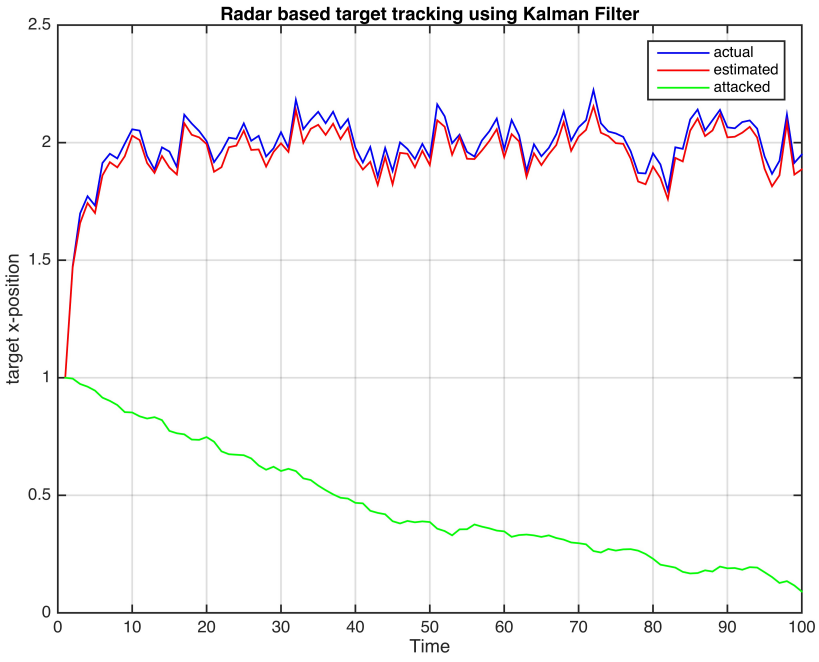


Figure 4.2: Optimal Attack on the Kalman Filter

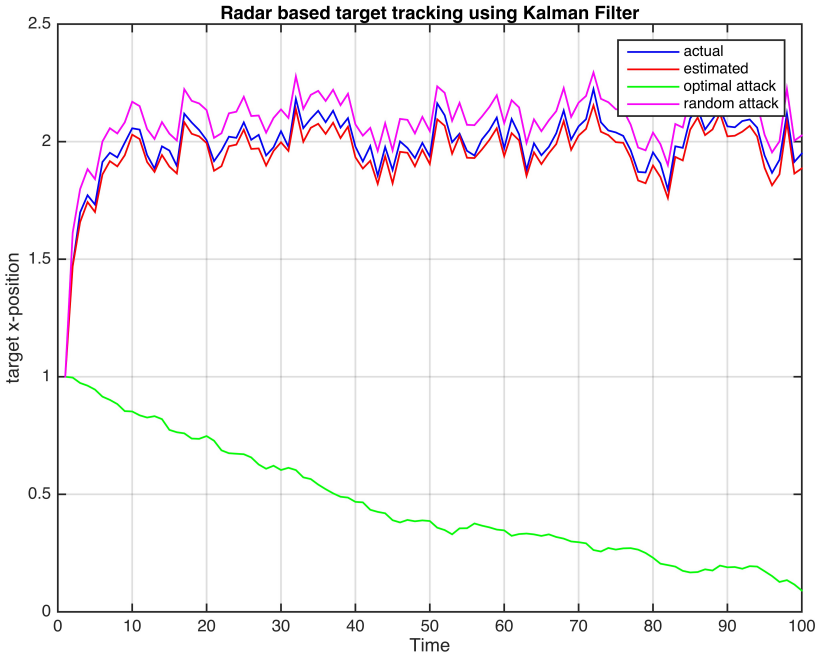


Figure 4.3: Optimal Attack and Random Attack on Kalman Filter

Therefore, a system theoretic analysis of the effect of sensor attacks on Kalman filter scalar case was presented. Specifically, the effects on the steady state response are derived. The characteristic equation for the optimal attack on Kalman filter is given and a numerical simulation which illustrates the effectiveness of the optimal attack is presented.

4.1.3 Mitigation with Δ_{y_k} dependent on the Estimation Error

We know $E\left(v_k \Delta_{y_k}^T\right) = E\left(\Delta_{y_k} v_k^T\right) = 0$. Therefore, the cost function:

$$\begin{aligned} J_k &= TrP_k^+ = Tr\left\{\left(I - K_k H_k\right) P_k^- \left(I - K_k H_k\right)^T + K R_k K_k^T - K_k E\left[\Delta_{y_k} \varepsilon_k^{-T}\right]\left(I - K_k H_k\right)^T \right. \\ &+ \left. K D_k K_k^T - \left(I - K_k H_k\right) E\left[\varepsilon_k^- \Delta_{y_k}^T\right] K_k^T \right\} \end{aligned} \tag{4.62}$$

where $R_k = E(v_k {v_k}^T)$ and $D_k = E(\Delta_{y_k} {\Delta_{y_k}}^T)$. Only the scalar case is considered, Therefore, equation (4.62) can be written as:

$$J_k = \left(I - K_k H_k\right)^2 P_k^- + K_k^2 R_k - 2\left(I - K_k H_k\right) K_k E\left[\varepsilon_k^- \Delta_{y_k}\right] \tag{4.63}$$

Now the optimal attack problem is converted into an optimization problem described as follows:

Problem 4.1.5.

$$\max_{E\left\|\Delta_{y_k}\right\|^2 \leq M} J_k \tag{4.64}$$

$$\begin{aligned} J_k &= (I - K_k H_k)^2 P_k^- + K_k^2 R_k + K_k^2 E\left[\Delta_{y_k}^2\right] - 2\left(I - K_k H_k\right) K_k E\left[\varepsilon_k^- \Delta_{y_k}\right] \\ &= N_k + K_k^2 E\left[\Delta_{y_k}^2\right] - 2\left(I - K_k H_k\right) K_k E\left[\varepsilon_k^- \Delta_{y_k}\right] \end{aligned} \tag{4.65}$$

where $N_k = (I - K_k H_k)^2 P_k^- + K_k^2 R_k$ This term from the previous equation can be written as:

$$-2\left(I - K_k H_k\right) K_k E\left[\varepsilon_k^- \Delta_{y_k}\right] = E\left[-2\left(I - K_k H_k\right) K_k \varepsilon_k^- \Delta_{y_k}\right] \tag{4.66}$$

Applying the Cauchy-Schwarz inequality yields:

$$E\left[-2\left(I - K_k H_k\right) K_k \varepsilon_k^- \Delta_{y_k}\right] \leq \sqrt{E\left[-2\left(I - K_k H_k\right) K_k \varepsilon_k^-\right]^2} \sqrt{E[\Delta_{y_k}]^2} \tag{4.67}$$

$$\begin{aligned} J_k &\leq N_k + K_k^2 E\left[\Delta_{y_k}^2\right] + \sqrt{E\left[-2\left(I - K_k H_k\right) K_k \varepsilon_k^-\right]^2} \sqrt{E[\Delta_{y_k}]^2} \\ &\leq N_k + K_k^2 M + \sqrt{E\left[-2\left(I - K_k H_k\right) K_k \varepsilon_k^-\right]^2} \sqrt{E[\Delta_{y_k}]^2} \\ &\leq \sqrt{E\left[2\left(I - K_k H_k\right) K_k \varepsilon_k^-\right]^2} \sqrt{E[\Delta_{y_k}]^2} \end{aligned} \tag{4.68}$$

The previous equation hold if and only if:

$$\Delta_{y_k} = \alpha_k \left(2\left(I - K_k H_k\right) K_k \varepsilon_k^-\right) \tag{4.69}$$

Therefore, the signal attack Δ_{y_k} should satisfy the following:

$$\begin{cases} \Delta_{y_k} = \alpha_k \left(2\left(I - K_k H_k\right) K_k \varepsilon_k^-\right) \\ E\left[\Delta_{y_k}^2\right] = M \end{cases} \tag{4.70}$$

Solving for α yields:

$$\begin{aligned} E\left[\alpha_k^2\left(4(I - K_k H_k)^2 K_k^2 \varepsilon_k^{-2}\right)\right] &= M \\ 4\alpha_k^2 (I - K_k H_k)^2 K_k^2 E\left[\varepsilon_k^{-2}\right] &= M \\ 4\alpha_k^2 (I - K_k H_k)^2 K_k^2 P_k^- &= M \end{aligned} \tag{4.71}$$

The solution for α is:

$$\alpha_k = \pm \frac{1}{\left(I - K_k H_k\right) K_k} \sqrt{\frac{M}{P_k^-}} \tag{4.72}$$

Therefore, the attack signal is:

$$\Delta_{y_k} = \pm \sqrt{\frac{M}{P_k^-}} \varepsilon_k^- \tag{4.73}$$

Replacing Δ_{y_k} in the cost function yields to:

$$J_{k1} = (I - K_k H_k)^2 P_k^- + K_k^2 R_k + K_k^2 M - 2\left(I - K_k H_k\right) K_k P_k^- \sqrt{\frac{M}{P_k^-}} \tag{4.74}$$

$$J_{k2} = (I - K_k H_k)^2 P_k^- + K_k^2 R_k + K_k^2 M + 2\left(I - K_k H_k\right) K_k P_k^- \sqrt{\frac{M}{P_k^-}} \tag{4.75}$$

Therefore, to solve for the optimal K_k is to take the partial derivative of the cost function J_k and set it equal to zero. We have the following the first partial derivative with respect to K_k:

$$\frac{\partial J_{k1}}{\partial K_k} = 2MK_k + 2P_k^- H_k \sqrt{\frac{M}{P_k^-}} K_k + 2K_k R_k - 2P_k^- H_k \left(1 - H_k K_k\right) \tag{4.76}$$

The optimal Kalman gain is:

$$K_k = \frac{P_k^- H_k + \sqrt{\frac{M}{P_k^-}} P_k^-}{M + P_k^- H_k^2 + 2H_k \sqrt{\frac{M}{P_k^-}} P_k^- + R} \tag{4.77}$$

Numerical Example

We give an example to illustrate the performance of mitigation strategy on Kalman filter. As mentioned before, for simplicity, we only deal with the scalar case here. The derivation can be easily generalized to more complicated vector case.

We took an example from Matlab where the Kalman filter is used for object tracking to predict an object's future location, to account for noise in an object's detected location. For this simple scalar case, the state-transition matrix is also a scalar which is chosen to be 0.5. Without loss of generality, assume the white noise w_k and v_k are zero-mean with both standard deviations equal 0.1. Notice that these two noise processes are uncorrelated. We set the initial location to be 1.

Applying the result from equation, the original radar tracking without attack is presented in Fig. 4.4. Then, we could generate the optimal attack signal as shown in Fig. 4.5. The state trajectory under optimal attack and mitigation are presented in Fig. 4.6 and 4.7. As expected, the mitigation strategy played an important role to mitigate the optimal attack.

Therefore, a system theoretic analysis of the effect of mitigation strategy on Kalman filter scalar case was presented. Specifically, the mitigation effects are shown after the optimal sensor attack are applied to Kalman filter. The characteristic equation for the optimal attack on Kalman filter and the mitigation new Kalman gain are given and a numerical simulation which illustrates the effectiveness of the mitigation strategy on optimal sensor attack is presented.

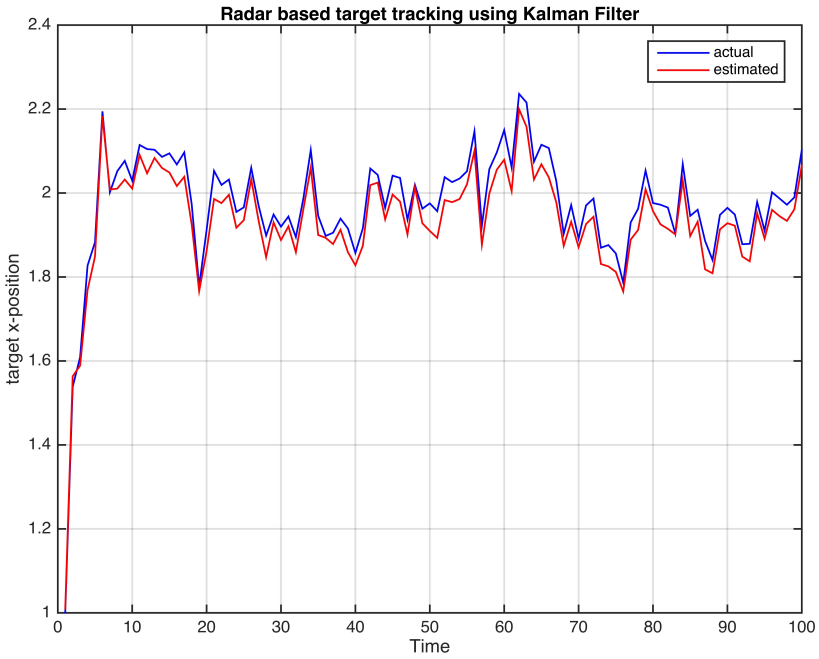


Figure 4.4: Radar Based Target Tracking using Kalman Filter

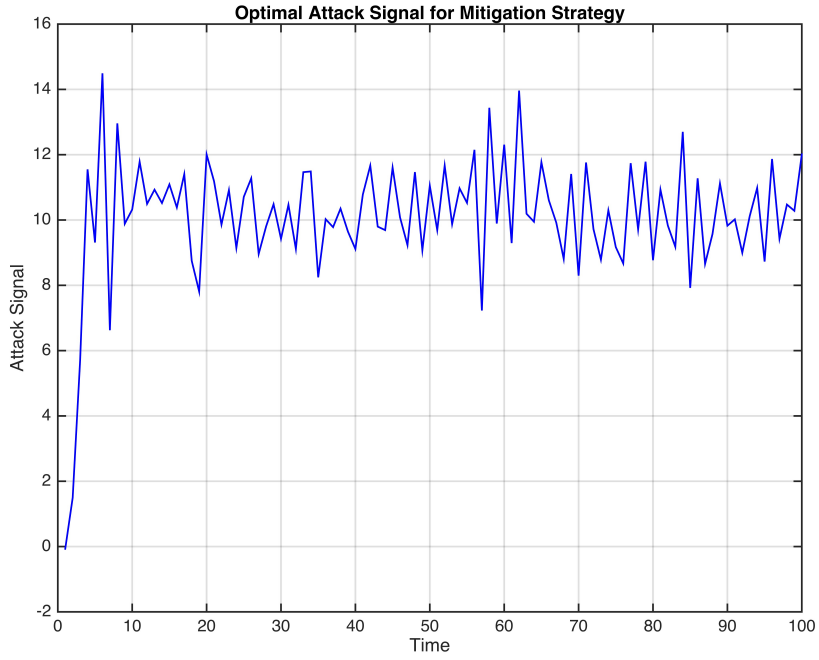


Figure 4.5: Optimal Attack Signal for Mitigation Strategy

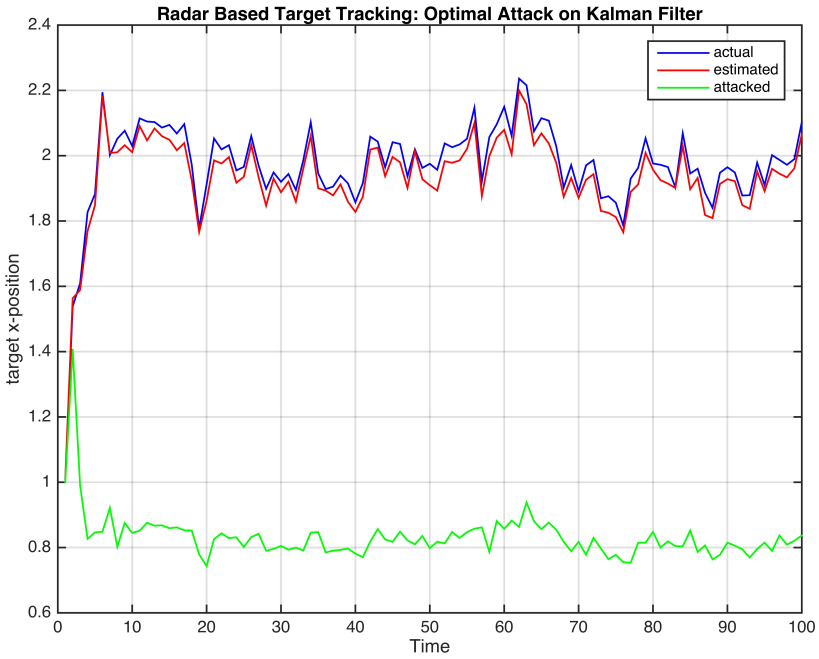


Figure 4.6: Radar Based Target Tracking: Optimal Attack on Kalman Filter

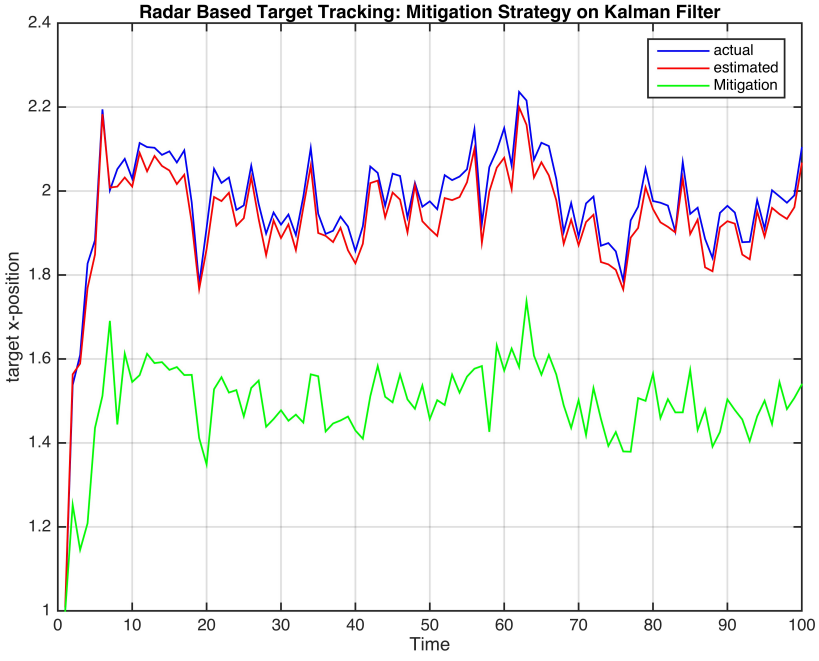


Figure 4.7: Mitigation Strategy on the Kalman Filter

4.2 Optimal Attack on Kalman Filter Multivariable Case

Now for the general case, the multivariable case is considered in this section. Let us assume that there exist matrices same size as P_k^+, $n \times n$ Ψ_k function of $\varepsilon_{x,k-1}$ components and Ω_k function of Δ_{y_k} components such that $\Psi_k \Omega_k^T = n\varepsilon_{x,k-1}\Delta_{y_k}^T$. Therefore, $\varepsilon_{x,k-1}$ can be replaced by Ψ_k, and Δ_{y_k} by Ω_k. Therefore, we have the following equation:

$$\begin{aligned} L = & Tr\left(P_{k-1}\right) - Tr\left(S_k^{-1} H_k P_k^- P_k^- H_k^T\right) - \frac{1}{n} Tr\left(S_k^{-1} E\left[\Omega_k \Psi_k^T\right] P_k^- H_k^T\right) \\ & - \frac{1}{n} Tr\left(S_k^{-1} H_k P_k^- E\left[\Psi_k \Omega_k^T\right]\right) - \frac{1}{n^2} Tr\left(S_k^{-1} E\left[\Omega_k \Psi_k^T\right] E\left[\Psi_k \Omega_k^T\right]\right) \\ & + \lambda \left(M_k - \frac{1}{n} Tr\left[E\left[\Omega_k \Omega_k^T\right]\right]\right) \end{aligned} \tag{4.78}$$

Let us take an example:

$$\begin{aligned} \varepsilon_{x,k-1}\Delta_{y_k}^T = \begin{bmatrix} e_1 \\ e_2 \\ e_3 \end{bmatrix} \begin{bmatrix} d_1 & d_2 & d_3 \end{bmatrix} = \begin{bmatrix} e_1 d_1 & e_1 d_2 & e_1 d_3 \\ e_2 d_1 & e_2 d_2 & e_2 d_3 \\ e_3 d_1 & e_3 d_2 & e_3 d_3 \end{bmatrix} \\ \frac{1}{n} \begin{bmatrix} e_1 & e_1 & e_1 \\ e_2 & e_2 & e_2 \\ e_3 & e_3 & e_3 \end{bmatrix} \begin{bmatrix} d_1 & d_2 & d_3 \\ d_1 & d_2 & d_3 \\ d_1 & d_2 & d_3 \end{bmatrix} = \begin{bmatrix} e_1 d_1 & e_1 d_2 & e_1 d_3 \\ e_2 d_1 & e_2 d_2 & e_2 d_3 \\ e_3 d_1 & e_3 d_2 & e_3 d_3 \end{bmatrix} \end{aligned} \tag{4.79}$$

Using these identity to obtain the optimality condtions:

$$\begin{aligned} \frac{\partial Tr Y^{-1} A}{\partial X} &= -\partial Tr\left(Y^{-1} A Y^{-1}\right)_c Y / \partial X \\ \frac{\partial Tr UV}{\partial X} &= \partial Tr U_c V / \partial X + \partial Tr U V_c / \partial X \end{aligned} \tag{4.80}$$

Where a matrix carrying the subscript "c" is to be regarded as a constant for further differentiation.

Which gives the following optimality conditions:

$$\begin{aligned}
\frac{\partial L}{\partial \Omega_k} &= \frac{1}{n} S_k^{-1} E\left[\Omega_k \Psi_k^T\right] P_k^- H_k^T S_k^{-1} \frac{\partial S_k}{\partial \Omega_k} - \frac{1}{n} S_k^{-1} P_k^- H_k^T E\left[\Psi_k^T\right] \\
&\quad + \frac{1}{n} S_k^{-1} H_k P_k^- E\left[\Psi_k \Omega_k^T\right] S_k^{-1} \frac{\partial S_k}{\partial \Omega_k} - \frac{1}{n} S_k^{-1} H_k P_k^- E\left[\Psi_k^T\right] \\
&\quad + \frac{1}{n^2} S_k^{-1} E\left[\Omega_k \Psi_k^T\right] E\left[\Psi_k \Omega_k^T\right] S_k^{-1} \frac{\partial S_k}{\partial \Omega_k} - \frac{1}{n^2} S_k^{-1} E\left[\Psi_k^T\right] E\left[\Psi_k \Omega_k^T\right] - \frac{2\lambda}{n} E\left[\Omega_k\right] \\
\frac{\partial L}{\partial \lambda} &= M_k - \frac{1}{n} Tr\left[E\left[\Omega_k \Omega_k^T\right]\right]
\end{aligned} \tag{4.81}$$

We know that $\frac{\partial S_k}{\partial \Omega_k} = \frac{2}{n} E\left[\Omega_k\right]$ and $E\left[\Psi_k^T\right] = 0$ therefore, equation 4.81 becomes:

$$\begin{aligned}
\frac{\partial L}{\partial \Omega_k} &= \frac{2}{n} S_k^{-1} H_k P_k^- P_k^- H_k^T S_k^{-1} E\left[\Omega_k\right] + \frac{2}{n^3} S_k^{-1} E\left[\Omega_k \Psi_k^T\right] E\left[\Psi_k \Omega_k^T\right] S_k^{-1} E\left[\Omega_k\right] \\
&\quad + \frac{2}{n^2} S_k^{-1} E\left[\Omega_k \Psi_k^T\right] P_k^- H_k^T S_k^{-1} E\left[\Omega_k\right] + \frac{2}{n^2} S_k^{-1} H_k P_k^- E\left[\Psi_k \Omega_k^T\right] S_k^{-1} E\left[\Omega_k\right] - \frac{2\lambda}{n} E\left[\Omega_k\right] = 0
\end{aligned} \tag{4.82}$$

$$\begin{aligned}
\frac{\partial L}{\partial \lambda} = M_k - \frac{1}{n} Tr\left[E\left[\Omega_k \Omega_k^T\right]\right] &= 0 \\
Tr\left[E\left[\Omega_k \Omega_k^T\right]\right] &= n M_k
\end{aligned} \tag{4.83}$$

The previous equation 4.82 can be factorized to the following:

$$\begin{aligned}
\frac{\partial L}{\partial \Omega_k} &= \frac{2}{n^3} S_k^{-1} \left(n^2 H_k P_k^- P_k^- H_k^T + n E\left[\Omega_k \Psi_k^T\right] P_k^- H_k^T \right. \\
&\left. + n H_k P_k^- E\left[\Psi_k \Omega_k^T\right] + E\left[\Omega_k \Psi_k^T\right] E\left[\Psi_k \Omega_k^T\right] - \lambda n^2 S_k^2\right) S_k^{-1} E\left[\Omega_k\right] = 0
\end{aligned} \tag{4.84}$$

Solving for λ we have:

$$\lambda = \frac{S_k^{-2}}{n^2} \left(n H_k P_k^- + E\left[\Omega_k \Psi_k^T\right]\right)^2 > 0 \tag{4.85}$$

By identification a solution to the previous equation:

$$\begin{cases}
E\left[\Omega_k \Psi_k^T\right] = E\left[\Psi_k \Omega_k^T\right]^T = -n\left(H_k P_k^- + \sqrt{\lambda} S_k\right) = -n\left(P_k^- H_k^T\right)^T - n\sqrt{\lambda}\left(S_k^T\right)^T \\
E\left[\Omega_k\right] = 0
\end{cases} \tag{4.86}$$

To check the sign of $\frac{\partial^2 L}{\partial \Omega_k^2}$ we can let $n = 1$. Therefore, the size of the matrices will be 1×1

$$\begin{aligned}\frac{\partial^2 L}{\partial \Omega_k^2} &= 2S_k^{-2}H_k^2P_k^2 + 2S_k^{-2}E[\Omega_k\Psi_k]^2 + 4S_k^{-2}H_kP_k^- E\left[\Omega_k\Psi_k\right] \\ &\quad - 8S_k^{-3}H_k^2P_k^2E[\Omega_k]^2 - 8S_k^{-3}E[\Omega_k\Psi_k]^2E[\Omega_k]^2 \\ &\quad - 16S_k^{-3}H_kP_k^- E\left[\Omega_k\Psi_k\right]E[\Omega_k]^2 - 2\lambda \end{aligned} \tag{4.87}$$

$$\left.\frac{\partial^2 L}{\partial \Omega_{y_k}^2}\right|_{E[\Omega_k]=0} = 2P_k^{-2}H_k^2S_k^{-2} + 2E[\Omega_k\Psi_k]^2S_k^{-2} + 4H_kP_k^-S_k^{-2}E\left[\Omega_k\Psi_k\right] - 2\lambda \tag{4.88}$$

$$\begin{aligned}\left.\frac{\partial^2 L}{\partial \Omega_k^2}\right|_{E[\Omega_k\Psi_k]=-H_kP_k^- - \sqrt{\lambda}S_k} &= 2P_k^{-2}H_k^2S_k^{-2} - 8P_k^{-2}H_k^2S_k^{-3}E[\Omega_k]^2 \ - 4\sqrt{\lambda}H_kP_k^-S_k^{-1} \\ &\quad + 2P_k^{-2}H_k^2S_k^{-2} + 2\lambda + 4\sqrt{\lambda}H_kP_k^-S_k^{-1} - 8P_k^{-2}H_k^2S_k^{-3}E[\Omega_k]^2 \\ &\quad - 16\sqrt{\lambda}H_kP_k^-S_k^{-2}E[\Omega_k]^2 - 2\lambda - 4P_k^{-2}H_k^2S_k^{-2} - 8\lambda S_k^{-1}E[\Omega_k]^2 \\ &\quad + 16P_k^{-2}H_k^2S_k^{-3}E[\Omega_k]^2 + 16\sqrt{\lambda}H_kP_k^-S_k^{-2}E[\Omega_k]^2 \end{aligned} \tag{4.89}$$

$$\left.\frac{\partial^2 L}{\partial \Omega_k^2}\right|_{E[\Omega_k\Psi_k]=-H_kP_k^- - \sqrt{\lambda}S_k} = -8\lambda S_k^{-1}E[\Omega_k]^2 < 0 \quad since\ \lambda > 0 \tag{4.90}$$

If showing that (4.90) corresponds to a maximum the following holds:

$$\begin{aligned} E\left[\Omega_k\Psi_k^T\right] &= -nH_kP_k^- - nS_k\sqrt{\lambda} = -nH_kP_k^- - nS_k\frac{S_k^{-1}}{n}\left(nH_kP_k^- + E\left[\Omega_k\Psi_k^T\right]\right) \\ E\left[\Omega_k\Psi_k^T\right] &= -nH_kP_k^- \end{aligned} \tag{4.91}$$

We have then for the optimal signal attack Ω_k:

$$\begin{cases} E\left[\Omega_k\Psi_k^T\right] = -nH_kP_k^- \quad H_kP_k^- \geq 0 \\ Tr\left[E\left[\Omega_k\Omega_k^T\right]\right] = nM_k \\ E\left[\Psi_k\Psi_k^T\right] = nP_k^- \\ E\left[\Psi_k\right] = 0 \end{cases} \tag{4.92}$$

The system of equations (4.58) have many solutions in Δ_{y_k}. That is any signal Δ_{y_k} which satisfies (4.58) is optimal. To determine an optimal signal attack it suffices to exhibit one solution; A possible choice is:

$$\Omega_k = -\Psi_k H_k \tag{4.93}$$

Therefore, we have:

$$\Omega_k \Omega_k^T = -\Psi_k \Psi_k^T H_k^2 \tag{4.94}$$

Then,

$$E\left[\Omega_k \Omega_k^T\right] = -E\left[\Psi_k \Psi_k^T\right] H_k^2 = -nP_k^- H_k^2 \tag{4.95}$$

Therefore, $M_k = -P_k^- H_k^2$

We have then for the optimal signal attack Ω_k:

$$\begin{cases} E\left[\Omega_k \Psi_k^T\right] = -nH_k P_k^- \quad H_k P_k^- \geq 0 \\ Tr\left[E\left[\Omega_k \Omega_k^T\right]\right] = -nP_k^- H_k^2 \\ E\left[\Psi_k \Psi_k^T\right] = nP_k^- \\ E\left[\Psi_k\right] = 0 \end{cases} \tag{4.96}$$

The optimal sensor attack strategy on the multivariable Kalman filter are characterized by (4.96). That is, any signal attack Ω_k which satisfies the identities (4.96) is optimal in the sense of maximizing the cost of the Kalman filter.

Chapter 5

Conclusions and Future Work

Securing control systems which manage the nation's critical infrastructures is extremely important. As recently highlighted the incidents and their impacts following the attacks on these control systems, it is imperative to study worst case attack scenarios which can occur to devise. Due to popular usage of least squares algorithms to estimate the states and identify plant parameters, the derivation of worst case attack strategies on these algorithms is presented in this thesis. First, the least squares optimal attack signal is characterized, then the weighted least squares optimal signal attack is characterized and generalized to the recursive least squares with the view to apply the results later to the Kalman filter. Also, the corresponding worst case cost functions for these algorithms are computed.

Optimal cyber-attack on the sensor and actuator are analyzed and numerical results are presented for the finite horizon linear quadratic control. The assumption is that the attacker is capable of spoofing sensor and actuator signals with time-varying signals. Both sensor and actuator are considered for finite energy and bounded signal attacks. Additionally, illustrative numerical results for LQ control of a power grid network are presented.

Optimal cyber-attack sensor attack strategies on the Kalman filter in which the attacker aims at making the estimation error as large as possible are derived. The attacker needs to determine the probability distribution which deteriorates the estimation cost function the most. In addition, two main propositions are given considering whether the estimation error depends on the attack signal or not. After both of the scalar cases are presented with illustrative examples, the multivariable case is presented as well. Finally, the mitigation for the scalar case is given considering the estimation

error depends on the attack signal with a numerical example, to show the effectiveness of the mitigation to counter attack the optimal sensor attack on Kalman filter. In future work, we plan to derive optimal attack mitigation strategies for the various least squares algorithms by appropriately modifying the estimation algorithms. This can be achieved by first determining the attack strategies that must deteriorate the estimation error, and then derive LS estimators which best minimize the corresponding worst case estimation error.

Another avenue for future work, is to derive mitigation strategies in the optimal finite energy and bounded signal attacks on sensors and actuators. These problems are formulated as min max problems involving the corresponding cost functions where, the max are taken with respect to the attack signal. The attackers are trying to maximize the cost, while we seek controllers and estimators that minimize the cost function.

In this thesis, we have only considered LQ control and LS estimators. Generality to other important control laws, such as, H_∞ control and maximum likelihood estimators are important research topics left for future work.

Bibliography

[1] A. A. Cárdenas, S. Amin, and S. Sastry, "Research challenges for the security of control systems.," in *HotSec*, 2008.

[2] J. Eisenhauer, P. Donnelly, M. Ellis, and M. OBrien, "Roadmap to secure control systems in the energy sector," *Energetics Incorporated. Sponsored by the US Department of Energy and the US Department of Homeland Security*, 2006.

[3] K. Stouffer, J. Falco, and K. Kent, "Guide to supervisory control and data acquisition (scada) and industrial control systems security–initial public draft," *National Institute of Standards and Technology, Gaithersburg, Maryland*, 2006.

[4] C. Robinson, "Isa99, industrial automation and control systems security-http://https://www.isa.org/isa99/."

[5] S. Amin, A. A. Cárdenas, and S. S. Sastry, "Safe and secure networked control systems under denial-of-service attacks," in *Hybrid Systems: Computation and Control*, pp. 31–45, Springer, 2009.

[6] J. Liedtke, N. Islam, and T. Jaeger, "Preventing denial-of-service attacks on a mu;-kernel for weboses," in *Operating Systems, 1997., The Sixth Workshop on Hot Topics in*, pp. 73–79, May 1997.

[7] D. Gregg, W. Blackert, D. Heinbuch, and D. Furnanage, "Assessing and quantifying denial of service attacks," in *Military Communications Conference, 2001. MILCOM 2001. Communications for Network-Centric Operations: Creating the Information Force. IEEE*, vol. 1, pp. 76–80 vol.1, 2001.

[8] J. Haggerty, Q. Shi, and M. Merabti, "Early detection and prevention of denial-of-service attacks: a novel mechanism with propagated traced-back attack blocking," *Selected Areas in Communications, IEEE Journal on*, vol. 23, pp. 1994–2002, Oct 2005.

[9] H. Nagesh and K. Sekaran, "Design and development of proactive solutions for mitigating denial-of-service attacks," in *Advanced Computing and Communications, 2006. ADCOM 2006. International Conference on*, pp. 157–162, Dec 2006.

[10] M. Hejmo, B. Mark, C. Zouridaki, and R. Thomas, "Design and analysis of a denial-of-service-resistant quality-of-service signaling protocol for manets," *Vehicular Technology, IEEE Transactions on*, vol. 55, pp. 743–751, May 2006.

[11] A. Gupta, C. Langbort, and T. Basar, "Optimal control in the presence of an intelligent jammer with limited actions," in *Decision and Control (CDC), 2010 49th IEEE Conference on*, pp. 1096–1101, Dec 2010.

[12] G. Befekadu, V. Gupta, and P. Antsaklis, "Risk-sensitive control under a class of denial-of-service attack models," in *American Control Conference (ACC), 2011*, pp. 643–648, June 2011.

[13] H. Huang, N. Ahmed, and P. Karthik, "On a new type of denial of service attack in wireless networks: The distributed jammer network," *Wireless Communications, IEEE Transactions on*, vol. 10, pp. 2316–2324, July 2011.

[14] X. Chen, S. Li, J. Ma, and J. Li, "Quantitative threat assessment of denial of service attacks on service availability," in *Computer Science and Automation Engineering (CSAE), 2011 IEEE International Conference on*, vol. 1, pp. 220–224, June 2011.

[15] R. Rejimol Robinson and C. Thomas, "Evaluation of mitigation methods for distributed denial of service attacks," in *Industrial Electronics and Applications (ICIEA), 2012 7th IEEE Conference on*, pp. 713–718, July 2012.

[16] S. Amin, X. Litrico, S. Sastry, and A. Bayen, "Cyber security of water scada systems part i: Analysis and experimentation of stealthy deception attacks," *Control Systems Technology, IEEE Transactions on*, vol. 21, pp. 1963–1970, Sept 2013.

[17] S. Amin, G. A. Schwartz, and S. S. Sastry, "Security of interdependent and identical networked control systems," *Automatica*, vol. 49, no. 1, pp. 186 – 192, 2013.

[18] C. Kwon, W. Liu, and I. Hwang, "Security analysis for cyber-physical systems against stealthy deception attacks," in *American Control Conference (ACC), 2013*, pp. 3344–3349, June 2013.

[19] V. Katkar and S. Kulkarni, "Experiments on detection of denial of service attacks using ensemble of classifiers," in *Green Computing, Communication and Conservation of Energy (ICGCE), 2013 International Conference on*, pp. 837–842, Dec 2013.

[20] V. Katkar and D. Bhatia, "Lightweight approach for detection of denial of service attacks using numeric to binary preprocessing," in *Circuits, Systems, Communication and Information Technology Applications (CSCITA), 2014 International Conference on*, pp. 207–212, April 2014.

[21] R. Muradore and D. Quaglia, "Energy-efficient intrusion detection and mitigation for networked control systems security," *Industrial Informatics, IEEE Transactions on*, vol. 11, pp. 830–840, June 2015.

[22] Y. Liu, P. Ning, and M. K. Reiter, "False data injection attacks against state estimation in electric power grids," in *Proceedings of the 16th ACM conference on Computer and communications security*, pp. 21–32, ACM, 2009.

[23] Y. Liu, P. Ning, and M. K. Reiter, "False data injection attacks against state estimation in electric power grids," *ACM Transactions on Information and System Security (TISSEC)*, vol. 14, no. 1, p. 13, 2011.

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

[25] S. D. Bopardikar and A. Speranzon, "On analysis and design of stealth-resilient control systems," in *Resilient Control Systems (ISRCS), 2013 6th International Symposium on*, pp. 48–53, IEEE, 2013.

[26] J.-Y. Keller and D. Sauter, "Monitoring of stealthy attack in networked control systems," in *2nd International Conference on Control and Fault-Tolerant Systems, SysTol'13*, pp. 462–467, 2013.

[27] M. Zhu and S. Martinez, "Stackelberg-game analysis of correlated attacks in cyber-physical systems," in *American Control Conference (ACC), 2011*, pp. 4063–4068, IEEE, 2011.

[28] S. Amin, X. Litrico, S. S. Sastry, and A. M. Bayen, "Stealthy deception attacks on water scada systems," in *Proceedings of the 13th ACM international conference on Hybrid systems: computation and control*, pp. 161–170, ACM, 2010.

[29] M. Zhu and S. Martinez, "Attack-resilient distributed formation control via online adaptation," in *Decision and Control and European Control Conference (CDC-ECC), 2011 50th IEEE Conference on*, pp. 6624–6629, Dec 2011.

[30] M. Zhu and S. Martinez, "On the performance analysis of resilient networked control systems under replay attacks," *Automatic Control, IEEE Transactions on*, vol. 59, pp. 804–808, March 2014.

[31] R. S. Smith, "A decoupled feedback structure for covertly appropriating networked control systems," *Network*, vol. 6, p. 6, 2011.

[32] H. Fawzi, P. Tabuada, and S. Diggavi, "Secure state-estimation for dynamical systems under active adversaries," in *Communication, Control, and Computing (Allerton), 2011 49th Annual Allerton Conference on*, pp. 337–344, IEEE, 2011.

[33] H. Fawzi, P. Tabuada, and S. Diggavi, "Secure estimation and control for cyber-physical systems under adversarial attacks," *Automatic Control, IEEE Transactions on*, vol. 59, no. 6, pp. 1454–1467, 2014.

[34] E. N. Asada, A. V. Garcia, and R. Romero, "Identifying multiple interacting bad data in power system state estimation," in *Power Engineering Society General Meeting, 2005. IEEE*, pp. 571–577, IEEE, 2005.

[35] J. Chen and A. Abur, "Improved bad data processing via strategic placement of pmus," in *Power Engineering Society General Meeting, 2005. IEEE*, pp. 509–513, IEEE, 2005.

[36] J. Chen and A. Abur, "Placement of pmus to enable bad data detection in state estimation," *Power Systems, IEEE Transactions on*, vol. 21, no. 4, pp. 1608–1615, 2006.

[37] E. Handschin, F. Schweppe, J. Kohlas, and A. Fiechter, "Bad data analysis for power system state estimation," *Power Apparatus and Systems, IEEE Transactions on*, vol. 94, no. 2, pp. 329–337, 1975.

[38] A. Garcia, A. Monticelli, and P. Abreu, "Fast decoupled state estimation and bad data processing," *Power Apparatus and Systems, IEEE Transactions on*, no. 5, pp. 1645–1652, 1979.

[39] S. Gastoni, G. Granelli, and M. Montagna, "Multiple bad data processing by genetic algorithms," in *Power Tech Conference Proceedings, 2003 IEEE Bologna*, vol. 1, pp. 6–pp, IEEE, 2003.

[40] T. Van Cutsem, M. Ribbens-Pavella, and L. Mili, "Bad data identification methods in power system state estimation-a comparative study," *Power Apparatus and Systems, IEEE Transactions on*, no. 11, pp. 3037–3049, 1985.

[41] A. Monticelli, *State estimation in electric power systems: a generalized approach*, vol. 507. Springer, 1999.

[42] A. Monticelli and A. Garcia, "Reliable bad data processing for real-time state estimation," *Power Apparatus and Systems, IEEE Transactions on*, no. 5, pp. 1126–1139, 1983.

[43] A. Monticelli, F. F. Wu, and M. Yen, "Mutiple bad data identfication for state estimation by combinatorial oftimization," *Power Delivery, IEEE Transactions on*, vol. 1, no. 3, pp. 361–369, 1986.

[44] V. Quintana, A. Simoes-Costa, and M. Mier, "Bad data detection and identification techniques using estimation orthogonal methods," *Power Apparatus and Systems, IEEE Transactions on*, no. 9, pp. 3356–3364, 1982.

[45] F. SCHWEPPE, "Power system state estimation, parts i, ii and iii," *IEEE Trans. Power Apparatus Syst.*, vol. 89, pp. 120–135, 1970.

[46] X. Nian-De, W. Shi-Ying, and Y. Er-Keng, "A new approach for detection and identification of multiple bad data in power system state estimation," *Power Apparatus and Systems, IEEE Transactions on*, no. 2, pp. 454–462, 1982.

[47] L. Zhao and A. Abur, "Multi area state estimation using synchronized phasor measurements," *Power Systems, IEEE Transactions on*, vol. 20, no. 2, pp. 611–617, 2005.

[48] J. Zhu and A. Abur, "Bad data identification when using phasor measurements," in *Power Tech, 2007 IEEE Lausanne*, pp. 1676–1681, IEEE, 2007.

[49] Y. Mo and B. Sinopoli, "Secure control against replay attacks," in *Communication, Control, and Computing, 2009. Allerton 2009. 47th Annual Allerton Conference on*, pp. 911–918, IEEE, 2009.

[50] Y. Mo, T. H.-J. Kim, K. Brancik, D. Dickinson, H. Lee, A. Perrig, and B. Sinopoli, "Cyber–physical security of a smart grid infrastructure," *Proceedings of the IEEE*, vol. 100, no. 1, pp. 195–209, 2012.

[51] G. Dán and H. Sandberg, "Stealth attacks and protection schemes for state estimators in power systems," in *Smart Grid Communications (SmartGridComm), 2010 First IEEE International Conference on*, pp. 214–219, IEEE, 2010.

[52] C. L. DeMarco, J. Sariashkar, and F. Alvarado, "The potential for malicious control in a competitive power systems environment," in *Control Applications, 1996., Proceedings of the 1996 IEEE International Conference on*, pp. 462–467, IEEE, 1996.

[53] A. R. Metke and R. L. Ekl, "Security technology for smart grid networks," *Smart Grid, IEEE Transactions on*, vol. 1, no. 1, pp. 99–107, 2010.

[54] A.-H. Mohsenian-Rad and A. Leon-Garcia, "Distributed internet-based load altering attacks against smart power grids," *Smart Grid, IEEE Transactions on*, vol. 2, pp. 667–674, Dec 2011.

[55] F. Pasqualetti, A. Bicchi, and F. Bullo, "A graph-theoretical characterization of power network vulnerabilities," in *American Control Conference (ACC), 2011*, pp. 3918–3923, June 2011.

[56] F. Pasqualetti, F. Dorfler, and F. Bullo, "Cyber-physical attacks in power networks: Models, fundamental limitations and monitor design," in *Decision and Control and European Control Conference (CDC-ECC), 2011 50th IEEE Conference on*, pp. 2195–2201, Dec 2011.

[57] A. Teixeira, S. Amin, H. Sandberg, K. H. Johansson, and S. S. Sastry, "Cyber security analysis of state estimators in electric power systems," in *Decision and Control (CDC), 2010 49th IEEE Conference on*, pp. 5991–5998, IEEE, 2010.

[58] S. Sridhar, A. Hahn, and M. Govindarasu, "Cyber-physical system security for the electric power grid," *Proceedings of the IEEE*, vol. 100, pp. 210–224, Jan 2012.

[59] J. Slay and M. Miller, "Lessons learned from the maroochy water breach," in *Critical Infrastructure Protection* (E. Goetz and S. Shenoi, eds.), vol. 253 of *IFIP International Federation for Information Processing*, pp. 73–82, Springer US, 2008.

[60] D. Eliades and M. Polycarpou, "A fault diagnosis and security framework for water systems," *Control Systems Technology, IEEE Transactions on*, vol. 18, pp. 1254–1265, Nov 2010.

[61] A. Teixeira, I. Shames, H. Sandberg, and K. H. Johansson, "A secure control framework for resource-limited adversaries," *Automatica*, vol. 51, pp. 135–148, 2015.

[62] S. Sundaram and C. N. Hadjicostis, "Distributed function calculation via linear iterative strategies in the presence of malicious agents," *Automatic Control, IEEE Transactions on*, vol. 56, no. 7, pp. 1495–1508, 2011.

[63] F. Pasqualetti, A. Bicchi, and F. Bullo, "Consensus computation in unreliable networks: A system theoretic approach," *Automatic Control, IEEE Transactions on*, vol. 57, no. 1, pp. 90–104, 2012.

[64] M. Zhu and S. Martínez, "On distributed convex optimization under inequality and equality constraints," *Automatic Control, IEEE Transactions on*, vol. 57, no. 1, pp. 151–164, 2012.

[65] E. Byres and J. Lowe, "The myths and facts behind cyber security risks for industrial control systems," in *Proceedings of the VDE Kongress*, vol. 116, 2004.

[66] D. Powner, "Critical infrastructure protection: Multiple efferots to secure control systems are under way, but challenges remain," 2007.

[67] R. J. Turk, *Cyber incidents involving control systems.* Idaho National Engineering and Environmental Laboratory, 2005.

[68] J. Slay and M. Miller, *Lessons learned from the maroochy water breach.* Springer, 2007.

[69] P. Quinn-Judge, "Cracks in the system," *TIME Magazine (January 9, 2002)*, 2002.

[70] J. Leyden, "Polish teen derails tram after hacking train network," *The Register*, vol. 11, 2008.

[71] J. Markoff, "A silent attack, but not a subtle one," *New York Times*, vol. 26, p. A6, 2010.

[72] J. P. Farwell and R. Rohozinski, "Stuxnet and the future of cyber war," *Survival*, vol. 53, no. 1, pp. 23–40, 2011.

[73] U. Attorney, "Willows man arrested for hacking into tehama colusa canal authority computer system," 2007.

[74] A. Greenberg, "Hackers cut cities power," *Forbes, Jaunuary*, 2008.

[75] D. Kravets, "Feds: Hacker disabled offshore oil platform leak-detection system," 2009.

[76] T. Reed, *At the abyss: an insider's history of the Cold War*. Random House LLC, 2007.

[77] J. Meserve, "Sources: Staged cyber attack reveals vulnerability in power grid," *CNN. com*, vol. 26, 2007.

[78] A. Greenberg, "Americas hackable backbone. forbes," 2007.

[79] T. Greene, "Experts hack power grid in no time," *Network World*, 2008.

[80] D. Simon, *Optimal state estimation: Kalman, H infinity, and nonlinear approaches*. John Wiley & Sons, 2006.

[81] M. Green and D. J. Limebeer, "Linear robust control," 1995.

[82] S. M. Djouadi, A. M. Melin, L. J. D. J. Ferragut, Erik M, and A. Drira, "Finite energy and bounded attacks on control system sensor signals," in *European Control Conference (ECC), 2015*, IEEE, 2015.

[83] S. Djouadi, "Finite energy and bounded cyber-attacks on closed-loop control systems," 2013.

[84] F. Borrelli and T. Keviczky, "Distributed lqr design for identical dynamically decoupled systems," *Automatic Control, IEEE Transactions on*, vol. 53, no. 8, pp. 1901–1912, 2008.

[85] S. M. Djouadi, A. M. Melin, E. M. Ferragut, J. Laska, J. Dong, *et al.*, "Finite energy and bounded actuator attacks on cyber-physical systems," in *American Control Conference (ACC), 2014*, pp. 1716–1722, IEEE, 2014.

Vita

Anis Drira was born in Hambourg, Germany in 1979, the son of Noureddine Drira and Nabiha Lejmi Drira. After graduating in 2004 from Murray State University in Murray, Kentucky where he received a Bachelor of Science degree in Engineering Physics, *magna cum laude*, he attended the University of Tennessee, Knoxville and received a Master of Science degree in 2006. He completed his Doctor of Philosophy degree in December 2015 from the department of Electrical Engineering and Computer Science. He is currently working as a research intern in Oak Ridge National Lab (ORNL) with the Electrical and Electronic Systems Research Division.