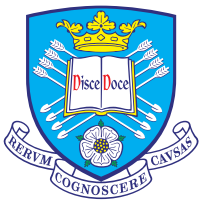
Information-Theoretic Data Injection Attacks on the Smart Grid



The University Of Sheffield.

Ke Sun

Department of Automatic Control and Systems Engineering
University of Sheffield

This dissertation is submitted for the degree of
Doctor of Philosophy

January 2020

Abstract

Power system state estimation is essential and crucial for power system operation. The operator uses the estimated states for decision making and anomaly detection. To facilitate the upgrade from classical power system to smart grid, a large amount of advanced sensing and communication infrastructure is built to improve the efficiency of the communication and control within power systems. But these sensing and communication infrastructures also expose the smart grid to cyber threats. One of the cyber threats that smart grid faces is the data injection attacks, in which the attacker compromises the measurements that are used for state estimation to mislead the operator.

In this thesis, we use information-theoretic measures to quantify the disruption caused by the attacks and the probability of detection induced by the attacks. Specifically to minimize the amount of information acquired by the operator from the measurements about the state variables describing the states of the grid, the attacker minimizes the mutual information between the state variables and the compromised measurements. Also to bypass the likelihood ratio test set by the operator, the attacker minimizes the Kullback-Leibler (KL) divergence between the distribution of measurements with attack and without attack to minimize the probability of detection. The stealth attacks achieve these two contradictive objectives by minimizing the sum of them, and closed-form expression for the optimal Gaussian attack is proposed.

To decrease the probability of detection induced by the stealth attacks, the equal sum in the objective of stealth attacks is generalized to a weighted sum by introducing a weighting parameter to the KL divergence term that represents probability of detection. Closed-form expression is proposed for the optimal generalized stealth attacks when the weighting parameter is larger or equal to one, i.e. when the attacker prioritizes the probability of detection over the disruption. Additionally a closed-form expression of the resulting probability of detection is obtained, but the expression does not give explicit insight into the relation between the probability of detection and the weighting parameter. As a result, a concentration inequality upper bound is proposed for the probability of detection to inform the design guidelines for the corresponding weighting parameter.

To construct the (generalized) stealth attacks, the attacker requires the second order statistics, i.e. the covariance matrix, of the state variables. When the attacker only gets access to a limited number of samples of the state variables, the attacker estimates the covariance matrix of the state variables by the sample covariance matrix of the state variables. Random matrix theory tools are employed to characterize the ergodic performance of the attacks using the sample covariance matrix for both the asymptotic scenario and the non-asymptotic scenario. Given the fact that there is no closed-form expression for the distribution of eigenvalues of random matrices under the non-asymptotic scenario, a closed-form expression is not available for the ergodic performance. Instead, an upper bound is proposed for the ergodic performance, for which a simple convex optimization needs to be solved to compute it. For the asymptotic case, a closed-form expression is provided for the ergodic performance of the attacks using the sample covariance matrix.

Keywords: Stealth, data injection attacks, information-theoretic measures, imperfect knowledge, learning

Acknowledgements

I would express my gratitude to my advisor, Dr. Iñaki Esnaola. I think the choice of working with Iñaki is one of the best choices I made in my life. His attitude and persistence towards "real science" inspire me quite a lot during the four years. It is him that transforms me from a person who is afraid of math to a person who enjoys math. Thanks for Iñaki setting an example for my future academic life.

I also would like to acknowledge the help from my collaborators: Dr. Samir M. Perlaza from the Institut National de Recherche en Informatique, Automatique et Mathématiques Appliquées (INRIA), Dr. Antonia M. Tulino from Nokia Bell Labs, Holmdel and University degli Studi di Napoli Federico II, and Dr. H. Vincent Poor from Princeton University. Without their advice and instructions, this work would not be as good as what it is now.

I would like to express my appreciation to my parents, Dr. Yuedong Sun and Mrs. Dongxia Yin. Thanks for giving me life and bringing me up. Thanks for the unconditional support and love from them. They provide me the courage to meet the difficulties and challenges in my life and study.

I would like to thank the information theory team at the Department of Automatic Control and Systems Engineering (ACSE), University of Sheffield: Cristian Gene, Miguel Arrieta, and Billy Casbolt. The talks and discussions with you are always inspiring and helpful. I really enjoy the time with you.

I would also like to thank my friends: Zhenglin Li, Sikai Zhang, Yuanlin Gu, Ning Wang, Yaxin Li, Yiwen Wang, Li Wang, Mengyao Xu, the staff team of Momentum gym, and .... Thanks for making my life in Sheffield full of fun and providing kindly support to me. The time with all of you is my treasure in my life.

In the end, I would like to acknowledge the scholarship support from ACSE and China Scholarship Council. Thanks for making my dream of studying in the UK come true.

There are many thanks I want to say to many friends. What I obtained in four years is not only the doctor degree, but the time with all of you.

Ke Sun

Table of Contents

List of Figures

List of Symbols

Notation	Description
$\mathbb{1}_{\{\cdot\}}$	indicator function for event given in $\{\cdot\}$
A^m	vector of random variables representing the attacks of dimension $m \times 1$
$\tilde{A}^m$	vector of random variables representing the attacks constructed using sample covariance matrix
$\mathbf{a}$	realization of A^m
$\mathbf{0}$	vector of all zero elements with proper dimension
$\mathbf{c}$	vector of errors that are injected into the estimated state variables
$D(\cdot\|\cdot)$	Kullback-Leibler divergence between two distributions
f_{X^n}	probability density function of variable X^n
$f_{X^n}(\mathbf{x})$	value of the probability density function of variable X^n at $\mathbf{x}$
$\mathsf{F}^{n_0}_{\tilde{\mathbf{\Lambda}}}(x)$	empirical cumulative distribution function of the diagonal elements of $\tilde{\mathbf{\Lambda}}$ when $n = n_0$
$\mathsf{F}_{\tilde{\mathbf{\Lambda}}}(x)$	asymptotic eigenvalue distribution of $\tilde{\mathbf{\Lambda}}$
$\mathbf{I}_m$	identical matrix of dimension $m \times m$
$I(\cdot;\cdot)$	mutual information between two variables
$\mathbf{H}/\mathbf{H}_{\mathbf{x}}$	Jacobian observation matrix of size $m \times n$ at operation point $\mathbf{x}$ (sometimes $\mathbf{x}$ is neglected)
$H_i\left(X^n\right)$	relation between measurement Y_i (i-th term in Y^m) and vector of state variables X^n
$H\left(X^n\right)^m$	$H\left(X^n\right)^m = [H_1(X^n), H_2(X^n), \cdots, H_m(X^n)]^{\mathrm{T}}$
k	number of samples of state variables that are available to the attacker
m	dimension of the vector of measurements
n	dimension of the vector of state variables

Notation	Description
n_0	number of state variables to be estimated within actual power system (for asymptotic scenario of random matrix theory)
N	number of buses in the power system
P_{X^n}	probability distribution of state variables X^n, similar for P_{Y^m}, $P_{Y_A^m}$
$\mathsf{P_D}$	probability of detection
p	rank of matrix $\mathbf{H}\boldsymbol{\Sigma}_{XX}\mathbf{H}^{\mathrm{T}}$
P_i	net active power injected into bus i
P_{ij}	line active power flow from bus i to bus j
Q_i	net reactive power injected into bus i
Q_{ij}	line reactive power flow from bus i to bus j
V_i	voltage magnitude of bus i in power system
θ_i	phase angle of bus i in the system
$\boldsymbol{\Lambda}_s$	matrix of eigenvalues of $\mathbf{H}\boldsymbol{\Sigma}_{XX}\mathbf{H}^{\mathrm{T}}$ in decreasing order
$\boldsymbol{\Lambda}_p$	matrix of non-zero eigenvalues of $\mathbf{H}\boldsymbol{\Sigma}_{XX}\mathbf{H}^{\mathrm{T}}$ in decreasing order
$\mathbf{V}$	matrix of eigenvector of $\mathbf{H}\boldsymbol{\Sigma}_{XX}\mathbf{H}^{\mathrm{T}}$ in the corresponding order of $\boldsymbol{\Lambda}_s$
$\tilde{\boldsymbol{\Lambda}}$	$\tilde{\boldsymbol{\Lambda}} \triangleq \frac{1}{\sigma^2}\boldsymbol{\Lambda}_p$
X^n	vector of random variables representing the state variables of dimension $n \times 1$
$\hat{X}^n$	an estimate of X^n
$\hat{X}^{n^\star}$	an estimate of X^n that is optimal for some cost function
$\mathbf{x}$	realization of X^n
$\hat{\mathbf{x}}$	an estimate of $\mathbf{x}$
$\hat{\mathbf{x}}^\star$	an estimate of $\mathbf{x}$ that is optimal for some cost function
Y^m	vector of random variables representing the measurements without attack of dimension $m \times 1$
$\mathbf{y}$	realization of Y^m
Y_A^m	vector of random variables representing the measurements with attack of dimension $m \times 1$
$\mathbf{y}_a$	realization of Y_A^m

Notation	**Description**
Z^m	vector of system noises of dimension $m \times 1$
$\mathbf{Z}_l$	random matrix of size $(k-1)\times l$ with i.i.d. standard Gaussian entries
α	probability of Type I error in hypothesis testing
β	probability of Type II error in hypothesis testing
$\mathcal{H}_0/\mathcal{H}_1$	null hypothesis / alternative hypothesis in hypothesis testing
$\mathcal{N}(\boldsymbol{\mu}, \boldsymbol{\Sigma})$	Gaussian distribution with mean $\boldsymbol{\mu}$ and covariance $\boldsymbol{\Sigma}$
σ^2	variance of system noise
$\boldsymbol{\Sigma}_{XX}$	covariance matrix of the state variables
λ	weight parameter for probability of detection
$\lambda_i(\mathbf{A})$	i-th eigenvalue of matrix $\mathbf{A}$ in defined order
$\lambda_{\min}(\mathbf{A})$	minimum eigenvalue of matrix $\mathbf{A}$
$\lambda_{\max}(\mathbf{A})$	maximum eigenvalue of matrix $\mathbf{A}$
$\lambda_{\mathbf{A}}$	unordered eigenvalue of $\mathbf{A}$
$s_{\min}(\mathbf{A})$	minimum singular value of matrix $\mathbf{A}$
$s_{\max}(\mathbf{A})$	maximum singular value of matrix $\mathbf{A}$
τ	detection threshold for likelihood ratio test
$\mathcal{S}_+^m$	set of all $m \times m$ positive semi-definite matrices
$\boldsymbol{\Sigma}_{AA}^{\star}$	covariance of optimal attack construction strategy for stealth attack or generalized stealth attack
diag$\{\cdot\}$	diagonal matrix with diagonal terms given in brackets
$\mathbb{E}[\cdot]$	expected value
$\mathbb{R}$	real number

Abbreviations

AC Alternating Current

AED Asymptotic Eigenvalue Distribution

AWGN Additive White Guassian Noise

DC Direct Current

DIA Data Injection Attack

EED Empeirical Eigenvalue Distribution

i.i.d. Independent and Identically Distributed

KL Kullback-Leibler

LRT Likelihood Ratio Test

LS Least Squares

MAP Maximum A Posteriori

ML Maximum Likelihood

MMSE Minimum Mean Squared Error

MSE Mean Squared Error

OPF Optimal Power Flow

PCA Principal Component Analysis

p.d.f. Probability Density Function

RMT Random Matrix Theory

RTU Remote Terminal Unit

SCADA Supervisory Control and Data Acquisition

SVD Singular Vector Decomposition

SVM Support Vector Machine

WLS Weighted Least Squares

Chapter 1

Introduction

1.1 Background and Motivation

Electricity is key in all aspects of industrial and daily activity, and therefore, electrification becomes a crucial indicator of the level of modernization and development of a society [1]. However the current power system architecture was proposed more than one hundred years ago, and its functioning and technology have not been significantly updated since its inception. In its classical setting power plants generate power, then the generated power is delivered by the transmission system in the form of high voltage, and the distribution system distributes the power to the users. Due to the limited power storage capability of the grid, the electricity generated by the generators has to be consumed by the users immediately. Otherwise the unbalance between generation and consumption can lead to significant performance degradation and eventually the collapse of the power system. So the operator of the power system schedules the generation and dispatches the power carefully according to the consumption of the loads and the structure of the power system. For that reason, larger grids give the operator more flexibility to regulate the power systems, which in turn makes the operation of the power systems more efficient and manageable. For example power systems with large capacity have high reliability and good peak regulation ability [1]. However power systems with large capacity also exhibit some risks. One of the problems is that accidents spread easily, and as a result, the large power systems are prone to collapse [1]. For example in 2003 a transmission line failure and the following cascading outages led to the largest power failure in North American history [2]. The continuous increase of electricity consumption (3.2 % increment in 2016 [3]) moves electricity grids towards larger-scale implementation to include more generators and users, which makes the management and control of such large systems increasingly challenging for the operator.

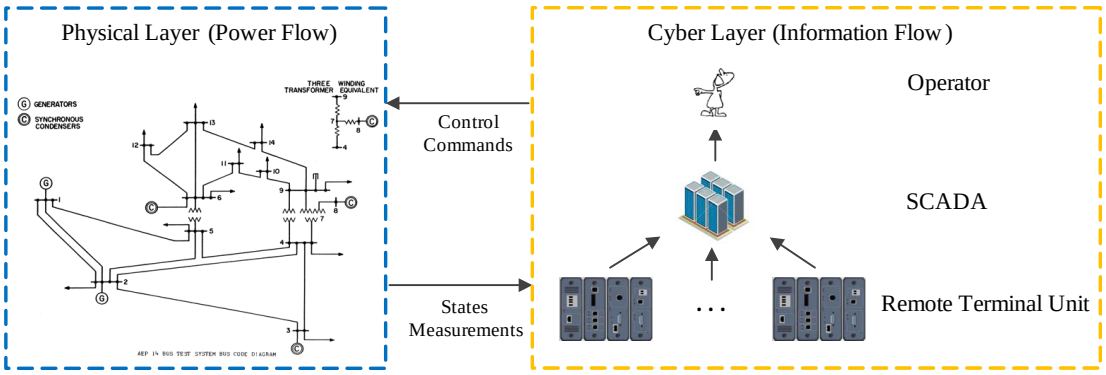


Fig. 1.1. Interconnection between the power system in the physical layer and the control units in the cyber layer.

To guarantee the reliable and efficient operation of power systems, the operator needs to know the state of the system and to make decisions based on the observed states. The state monitoring procedure relies on the sensors or remote terminal units (RTUs) distributed across the power network. These sensors provide measurements that contain information about the state of the power system, which are utilized by the operator to verify the state of the power system and to schedule the generation to meet the consumption. The measurements produced by the sensors are transmitted to the Supervisory Control and Data Acquisition (SCADA) system through a public or dedicated communication system. Then the operator in the control center uses the measurements in the SCADA system to determine the state of the power system and to make decisions based on the available state information, such as determining the optimal power flow (OPF) and detecting abnormalities. As the measurements are usually noisy linear or nonlinear observations of the state, the operator performs state estimation to retrieve information about the state, i.e. to obtain an estimation of the state based on the measurements. With the estimated state, the operator manages the power system and detects abnormalities within the system.

The interconnection between the power system in the physical layer and the sensing and decision-making in the cyber layer is depicted in Fig. 1.1. The cyber layer monitors the states of the physical layer through the RTUs, which forms the information flow; the physical layer implements the decisions made by the operator in the cyber layer to operate the power system, which forms the power flow. The cooperation and coherence of the physical layer and the cyber layer guarantee the efficient and reliable operation of the power system.

Power system has been moving toward the *smart grid* in recent years, a solution in which various technologies are implemented to make the power system more efficient, more intelligent, and more resilient. The definition of the smart grid is given below.

Definition 1.1 ([4]). *The Smart Grid is an electric system that uses information, two-way, cyber-secure communication technologies, and computational intelligence in an integrated fashion across electricity generation, transmission, substations, distribution, and consumption to achieve a system that is clean, safe, secure, reliable, resilient, efficient, and sustainable. The Smart Grid has the following characteristics.*

- *Self-healing: repair or removal of potentially faulty equipment from service before it fails and reconfigure the system automatically;*
- *Flexible: interconnect the distributed generations and energy storage units in the system rapidly and safely;*
- *Predictive: use different kinds of tools, such as machine learning, to predict the most likely events in the power system;*
- *Interactive: provide information to both the operator and the customers to allow them to play an active role;*
- *Optimized: know the state of major components in the system and optimize the system based on the state;*
- *Secure: guarantee the security of all critical assets in the system, both physical security and cyber security.*

This definition highlights that advanced communication technology and renewable generation sources are two key enabling components to make the grid "smart". Instead of the one-way information flow in which only the operator acquires information, the advanced communication infrastructure allows two-way communication between the operator and the users, which makes the power generation, transmission, and consumption more efficient [5]. As a complement to central power plants, the distributed generators that use renewable resources, such as wind turbines, enable green and cost-effective power generation. Also distributed generation makes power systems more resilient to incidents, as the users are still supplied by distributed generation option in the event of the main power plant failing. In the U.S.A., the capacity of distributed generators is one-sixth of the capacity of the nation's existing centralized power plants [6]. These technologies make power generation greener and more efficient, but they also make the management of power systems more challenging. For example, the power generated by wind turbines is highly affected by the wind speed, which is hard to predict accurately and to control efficiently. To monitor the state of distributed generations, the operator deploys more advanced communication infrastructure and places more sensors in the power system to monitor the state of the power system for reliable and efficient operation.

As the number of distributed generators and sensors increases, the interconnection between the physical layer and the cyber layer needs to be more efficient and speedier. Advanced communication infrastructure makes the efficient and rapid interconnection possible, but they also expose the smart grid to the cyber threats, such as computer viruses and data injection attacks, c.f. [7] and [8]. The frailty of the cyber layer in the smart grid affects the security of the physical power system directly. In 2015 attackers hacked the power supply system of Ukraine using the BlackEnergy virus, and shut down 30 generation substations, which made about 225,000 residents lose power [9]. Although some protection approaches were set up against the cyber threats after the failure, a similar attack happened again in 2017 [10].

Data injection attacks (DIAs) are one of the main cyber security threats that the smart grid faces. Unlike the BlackEnergy virus whose consequences are easy to be observed by the operator, DIAs aim to disrupt the power system in a covert fashion. Specifically, DIAs disrupt the state estimation procedure implemented by the operator in the control center by compromising the sensors in the system. Therefore the attacker can inject an additional term into the true measurements, and mislead the operator with fake measurements. It is shown in [7] and [8] that when the attacks are designed in a "*smart*" way, the anomaly detection procedure set by the operator is unable to distinguish the attacks. Detailed information about DIAs is covered in Section 2.2.

The cyber security threats to which the smart grid is exposed are not well-understood yet, and therefore, practical security solutions need to come forth as a multidisciplinary effort combining technologies such as cryptography, machine learning, and information-theoretic security [11]. Information-theoretic tools are well-suited to analyze power systems by leveraging the stochastic description of the state variables. Information-theoretic measures also provide fundamental limits for the information acquisition between the cyber layer and the physical layer. For example, a sensor placement strategy that accounts for the amount of information acquired by the sensing infrastructure is studied in [12], in which the operator maximizes the mutual information between the state variables and the gathered measurements. Also information-theoretic privacy guarantees for smart meter users of power systems are proposed in [13–15] for memoryless stochastic processes and in [16] for general random processes.

In this thesis, we investigate DIAs using information-theoretic measures to quantify the disruption caused by the attacks and the resulting probability of detection. The mutual information between the state variables and the measurements obtained by the sensors determines the amount of information that the operator obtains about the state variables from the measurements. As a result, the attacker

minimizes the mutual information between the state variables and the compromised measurements to minimize the amount of information acquired by the operator from the measurements. Meanwhile, the attacker minimizes the Kullback-Leibler (KL) divergence between the distribution of measurements with attacks and without attacks to minimize the probability of detection. Stealth attacks achieve these two contradictive objectives by minimizing the sum of them, which is generalized by adopting a weighted sum of the objective later. Then the performance of the stealth attacks when the attacker has imperfect knowledge about the system is analyzed using tools from random matrix theory (RMT).

1.2 Overview of Thesis

The remaining part of the thesis is divided into five chapters.

- **Chapter 2 Literature Review**

  Chapter 2 introduces the commonly adopted state estimation problem framework for general settings and the specific observation model arising in power systems. Based on the observation model, we introduce the classical power system state estimation and anomaly detection approaches. Afterward, state estimation and anomaly detection formulation are extended to the Bayesian framework, in which the state variables are modeled by some given distribution. In the end, we provide the fundamental formulation of the DIAs and the attack construction for different estimation and detection frameworks, including the least squares (LS) framework, Bayesian framework, and data-driven framework.

- **Chapter 3 Information-Theoretic Stealth Attacks**

  In Chapter 3, we propose an information-theoretic framework for DIAs under Bayesian estimation framework with linearized dynamics. Specifically the mutual information between the state variables and the compromised measurements is treated as the disruption objective of the attack, and the resulting probability of detection is characterized by the KL divergence between the distribution of measurements with attack and without attack. The stealth attacks combine the mutual information objective and the KL divergence objective by summing these two objectives. A closed-form expression of the stealth attacks construction is provided.

- **Chapter 4 Generalized Information-Theoretic Stealth Attacks**

  In Chapter 4, the stealth attacks construction in chapter 3 is generalized by adopting a weighted sum of the mutual information objective and the KL

divergence objective, in which a weighting parameter is assigned to the KL divergence objective. The optimal attack construction is characterized for the case in which more weight is given to the detection constraint. Additionally a closed-form expression of the resulting probability of detection is obtained, but the expression does not give explicit insight into the relation between the probability of detection and the weighting parameter. As a result, a concentration inequality upper bound is proposed for the probability of detection to inform the design guidelines for the corresponding weighting parameter.

- **Chapter 5 Learning Requirements for Stealth Attacks**

  In Chapter 5, we investigate the attack construction when an attacker does not have perfect knowledge of the distribution of the state variables. Specifically the attacker only gets access to a limited number of realizations of the state variables and uses the sample covariance matrix of the samples to construct the attacks. RMT tools are employed to characterize the ergodic performance of the attacks using the sample covariance matrix for both the asymptotic scenario and the non-asymptotic scenario. Given the fact that it is challenging to characterize the sample covariance matrices under the non-asymptotic scenario, a closed-form expression is not available for the ergodic performance. Instead an upper bound is proposed for the ergodic performance, for computing which a simple convex optimization problem needs to be solved. For the asymptotic case, a closed-form expression is provided for the ergodic performance of the attacks using a sample covariance matrix.

- **Chapter 6 Conclusions and Future Work**

  The thesis ends with Chapter 6, which contains the conclusion and potential future work.

1.3 Disseminated Results

The results from this research are disseminated in following papers.

Journal paper:

- **K. Sun**, I. Esnaola, S.M. Perlaza, and H.V. Poor, "Stealth attacks on the smart grid," *IEEE Trans. Smart Grid* (Early Access), 2019.

Book chapter:

- I. Esnaola, S.M. Perlaza, and **K. Sun**, "Bayesian attacks," in *Advanced Data Analytics for Power Systems*, A. Tajer, S.M. Perlaza and H.V. Poor, Eds., Cambridge University Press, Cambridge, UK, 2020 (to appear).

Conference paper:

- **K. Sun**, I. Esnaola, A.M. Tulino and H.V. Poor, "Learning requirements for stealth attacks", in *Proc. IEEE Int. Conf. on Acoust., Speech, and Signal Process.*, Brighton, UK, May 2019, pp. 8102-8106. (**invited paper**)
- **K. Sun**, I. Esnaola, S.M. Perlaza, and H.V. Poor, "Information-theoretic attacks in the smart grid," in *Proc. IEEE Int. Conf. on Smart Grid Commun.*, Dresden, Germany, Oct. 2017, pp. 455-460.

Poster Presentation:

- **K. Sun**, I. Esnaola, A.M. Tulino and H.V. Poor, "Learning requirements for stealth attacks", in *Proc. 5th London Symp. on Inform. Theory* , London, UK, May 2019.
- **K. Sun**, I. Esnaola, "Information theoretical attack in electricity grids", in *ACSE PGR Symp. (Departmental Ph.D. Symp.)*, Sheffield, UK, Oct. 2016.

Oral Presentation:

- **K. Sun**, I. Esnaola, A.M. Tulino and H.V. Poor, "Learning requirements for stealth attacks", in *Proc. IEEE Int. Conf. on Acoust., Speech, and Signal Process.*, Brighton, UK, May 2019, pp. 8102-8106.
- **K. Sun**, I. Esnaola, , A.M. Tulino and H.V. Poor, "Learning requirements for stealth attacks", in *ACSE PGR Symp. (Departmental Ph.D. Symp.)*, Sheffield, UK, Feb. 2019.
- **K. Sun**, I. Esnaola, S.M. Perlaza, and H.V. Poor, "Information-theoretic attacks in the smart grid," in *Proc. IEEE Int. Conf. on Smart Grid Commun.*, Dresden, Germany, Oct. 2017, pp. 455-460.

Chapter 2

Literature Review

2.1 State Estimation in Power System

Economic and efficient operation of the power system requires the operator to get access to the correct state of the power system. For this reason, the operator of the power system places large numbers of sensors in the system to monitor it. The measurements obtained by the sensors are transferred to the SCADA system through the communication network to support decision-making at the energy control center. Using the measurement data, the operator averts major system failures and regional blockout. Before making security assessments or taking control actions, a reliable estimate of the existing state of the system must be determined [17]. The procedure of estimating the states from the measurements is called state estimation.

The power system state estimation procedure has three main tasks [18, 19]:

- *Observability Analysis*: To determine if a unique estimate for any state of the system can be obtained.
- *State Estimation*: To determine an optimal estimate for any state of the system in real-time.
- *Bad Data Detection*: To detect measurement errors and identify bad data, and to eliminate them if possible.

In the following, we introduce the observation model for the power system state estimation problem first, and the bad data detection based on the estimated states or the obtained measurements. Then we review the state estimation and detection approaches within a Bayesian framework. We end by providing an overview of some advanced estimation and detection approaches for the power system.

2.1.1 Classical Power System State Estimation and Anomaly Detection

Power System Observation Model

The observation model is the relation between the measurements obtained from a system and the state variables describing the state of this system, which is given by

$$Y^m = \begin{bmatrix} Y_1 \\ Y_2 \\ \vdots \\ Y_m \end{bmatrix} = \begin{bmatrix} H_1(X_1, X_2, \cdots, X_n) \\ H_2(X_1, X_2, \cdots, X_n) \\ \vdots \\ H_m(X_1, X_2, \cdots, X_n) \end{bmatrix} + \begin{bmatrix} Z_1 \\ Z_2 \\ \vdots \\ Z_m \end{bmatrix} = H(X^n)^m + Z^m \tag{2.1}$$

where $X^n \in \mathbb{R}^n$ is a vector of random variables describing the state of the system; $H(X^n)^m = [H_1(X^n), H_2(X^n), \cdots, H_m(X^n)]^{\mathrm{T}} \in \mathbb{R}^m$ with $H(X^n)^m : \mathbb{R}^n \to \mathbb{R}^m$ denoting the nonlinear or linear relation between the measurements Y^m and the state variables X^n; and $Z^m \in \mathbb{R}^m$ is the additive noise introduced by the sensors, which is usually modeled by a Gaussian distribution [17, 20].

In power systems, the phase angle and voltage magnitude of the buses are the state variables that are difficult to measure directly using the sensors, but they are required by the operator to verify the state of the grid and make corresponding decisions. So the phase angle and voltage magnitude of buses are usually chosen to be the state variables that need to be estimated from the measurements. Within this setting, a power system containing N buses is described by $2N-1$ state variables, i.e., N bus voltage magnitudes and $N-1$ bus phase angle, for which the phase angle of a chosen reference bus is set to a known value, usually 0. Sometimes the voltage magnitude of the reference node is also set to a known value, usually 1.0, to simplify the state estimation calculation, and the voltage magnitude of the other nodes is expressed as a percent or *per unit* of the reference value. For example, if a base voltage of 120 kV is chosen, i.e. the voltage magnitude of the reference node is 120kV, the voltage of 108 kV is expressed as 0.90 per unit. Here except the direct current (DC) state estimation that will be covered later in this section, we do not specify the voltage magnitude of the reference node in this thesis.

Assuming bus 1 to be the reference bus, the vector of state variables has the following form

$$X^n = [\theta_2, \theta_3, \ldots, \theta_N, V_1, V_2, \ldots, V_N]^{\mathrm{T}}, \tag{2.2}$$

where $\theta_i \in [-\pi, +\pi]$ and $V_i \in \mathbb{R}^+$ are the phase angle and voltage magnitude of bus i, respectively [1]. In power systems, the measurements are of different types. The commonly measured variables are line power flows and bus power injections, i.e.

$$Y^m = [P_1, \ldots, P_n, Q_1, \ldots, Q_n, \ldots P_{ij} \ldots, \ldots Q_{ij} \ldots]^{\mathrm{T}}, \tag{2.3}$$

where $P_i \in \mathbb{R}$ and $Q_i \in \mathbb{R}$ are the net active power and reactive power injected into bus i, respectively; and $P_{ij} \in \mathbb{R}$ and $Q_{ij} \in \mathbb{R}$ are the active power flow and reactive power flow from bus i to bus j, respectively. Here when P_i and Q_i are of negative values, it implies that bus i consumes active and reactive power from the main grid; when P_i and Q_i are of positive values, it implies that bus i injects active and reactive power into the main grid. Similarly when P_{ij} and Q_{ij} are of negative values, it implies that bus i receives active and reactive power from bus j; when P_{ij} and Q_{ij} are of positive values, it implies that bus i transmits active and reactive power to bus j.

Here we use a 4-Bus system from [17, Problem 15.10] as an example to illustrate the nonlinear observation function $H\left(X^n\right)^m$. The topology and parameters of the 4 bus system are provided in Fig. 2.1, in which bus 1 is chosen to be the reference node. The real line flow P_{12} between bus 1 and bus 2 is given by

$$P_{12} = \mathrm{real}\left(\frac{\left(V_2\angle\theta_2 - V_1\angle 0\right)^2}{Z_{12}}\right) = -\mathrm{real}\left(\left(V_2\angle\theta_2 - V_1\angle 0\right)^2 Y_{12}\right). \tag{2.4}$$

where Z_{12} is the impedance for the branch connecting bus 1 and bus 2, and $Y_{12} = -1/Z_{12}$ is the negative of the admittance of the branch. Changing the polar coordinate in (2.4) to Cartesian coordinate yields

$$P_{12} = -|V_1|^2 G_{12} + |V_1 V_2 Y_{12}| \cos\left(\delta_{12} + \theta_2 - \theta_1\right), \tag{2.5}$$

in which G_{12} is the real part of Y_{12}, and δ_{12} is the argument of Y_{12}, i.e.

$$Y_{12} = |Y_{12}|\angle\delta_{12} = |Y_{12}|\cos\delta_{12} + j|Y_{12}|\sin\delta_{12} = G_{12} + jB_{12}. \tag{2.6}$$

Instead of taking the real part in (2.4), the reactive power flow follows from taking the imaginary part of (2.4). The net active power injection of bus 1 follows immediately

[1] Without confusion, we use subscript n to denote the dimension of the vector of state variables. The exact dimension of the vector is determined by the operator. For example, the operator can estimate the phase angle, or the voltage magnitude, or both of them, so we have $n = N - 1$, $n = N$, or $n = 2N - 1$. Here we do not specify the variables that are estimated by the operator. Similarly we use subscript m to denote the dimension of the vector of measurements. The value of m changes when the number of measurements used in the state estimation changes.

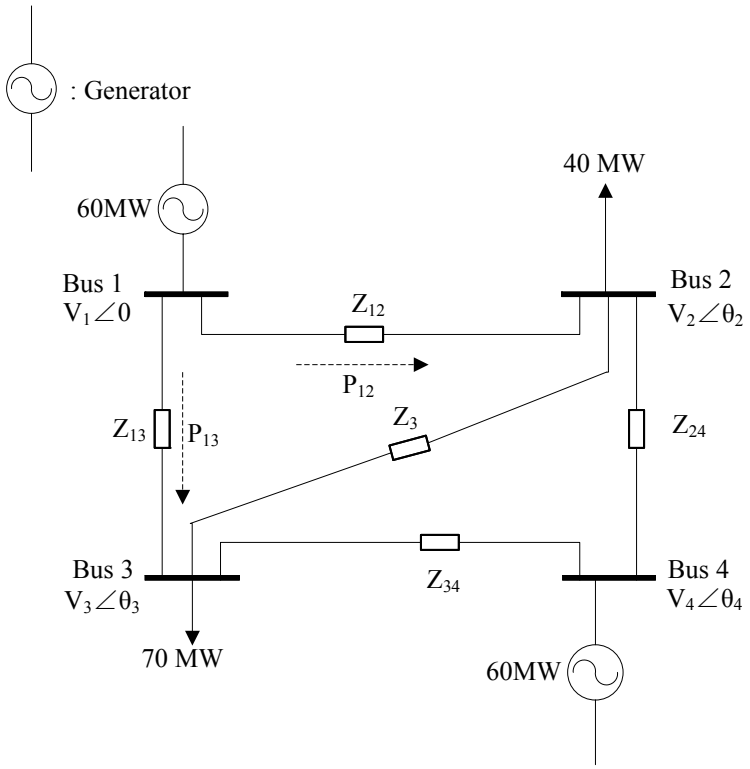


Fig. 2.1. 4-Bus system.

by

$$P_1 = P_{12} + P_{13} + (P_{inj})_1 - (P_{consump})_1, \tag{2.7}$$

where $(P_{inj})_1$ is the power generated by the generator at bus 1, which is 60 MW for the 4 Bus system; and $(P_{consump})_1$ is the power consumption at bus 1.

Under a general setting, the real power flow and the reactive power flow of the branch connecting bus i and bus j are given by

$$P_{ij} = -|V_i|^2 G_{ij} + |V_i V_j Y_{ij}| \cos\left(\delta_{ij} + \theta_j - \theta_i\right) \tag{2.8}$$

$$Q_{ij} = -\left(|V_i|^2 \left(\frac{B'_{ij}}{2} - B_{ij}\right) + |V_i V_j Y_{ij}| \sin\left(\delta_{ij} + \theta_j - \theta_i\right)\right), \tag{2.9}$$

respectively, where Y_{ij} is the negative of the admittance of the branch connecting bus i and bus j, which is given by

$$Y_{ij} = |Y_{ij}| \angle \delta_{ij} = |Y_{ij}| \cos\delta_{ij} + j|Y_{ij}| \sin\delta_{ij} = G_{ij} + jB_{ij}; \tag{2.10}$$

and $B'_{ij}/2$ is the line-charging susceptance of the branch connecting bus i and bus j. The real power and reactive power injected into bus i are given by

$$P_i = |V_i|^2 G_{ii} + \sum_{j=1, j\neq i}^{n} |V_i V_j Y_{ij}| \cos\left(\delta_{ij} + \theta_j - \theta_i\right) \tag{2.11}$$

$$Q_i = -\left(|V_i|^2 B_{ii} + \sum_{j=1, j\neq i}^{n} |V_i V_j Y_{ij}| \sin\left(\delta_{ij} + \theta_j - \theta_i\right)\right), \tag{2.12}$$

respectively, where

$$Y_{ii} = -\sum_{j=1, j\neq i}^{n} Y_{ij} = |Y_{ii}| \cos\delta_{ii} + j|Y_{ii}| \sin\delta_{ii} = G_{ii} + jB_{ii}. \tag{2.13}$$

State Estimation with Linearized Dynamics

When the nonlinear relation between the state variables and the measurements is considered for power system case, the state estimation is called "*alternating current (AC) state estimation*". Given the nonlinearity of the observation functions $H(X^n)$ in power systems, the state estimation is difficult to implement, even under some specific assumptions about the distribution of the state variables. So the nonlinear observation functions are often linearized at some operation point to simplify the state estimation problem. The observation functions $H\left(X^n\right)^m$ with linearized dynamics are given by

$$Y^m = \mathbf{H}\mathbf{x} + Z^m, \tag{2.14}$$

where $\mathbf{H} \in \mathbb{R}^{m\times n}$ is the Jacobian matrix of $H\left(X^n\right)^m$ for operation point $\mathbf{x}$, which is given by

$$\mathbf{H} = \frac{\partial}{\partial X^n} H\left(X^n\right)^m|_{X^n=\mathbf{x}} = \begin{bmatrix} \frac{\partial H_1}{\partial X_1} & \frac{\partial H_1}{\partial X_2} & \cdots & \frac{\partial H_1}{\partial X_n} \\ \frac{\partial H_2}{\partial X_1} & \frac{\partial H_2}{\partial X_2} & \cdots & \frac{\partial H_2}{\partial X_n} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial H_m}{\partial X_1} & \frac{\partial H_m}{\partial X_2} & \cdots & \frac{\partial H_m}{\partial X_n} \end{bmatrix} |_{X^n=\mathbf{x}}. \tag{2.15}$$

For the vector of measurements given in (2.3) and the vector of state variables in (2.21), the Jacobian matrix for state estimation is given by

$$\begin{bmatrix} \vdots \\ P_i \\ \vdots \\ \vdots \\ Q_i \\ \vdots \\ \vdots \\ P_{ij} \\ \vdots \\ \vdots \\ Q_{ij} \\ \vdots \end{bmatrix} = \begin{bmatrix} & \vdots & & & \vdots & \\ \cdots & \frac{\partial}{\partial\theta_j}P_i & \cdots & \cdots & \frac{\partial}{\partial V_j}P_i & \cdots \\ & \vdots & & & \vdots & \\ \cdots & \frac{\partial}{\partial\theta_j}Q_i & \cdots & \cdots & \frac{\partial}{\partial V_j}Q_i & \cdots \\ & \vdots & & & \vdots & \\ \cdots & \frac{\partial}{\partial\theta_j}P_{ij} & \cdots & \cdots & \frac{\partial}{\partial V_j}P_{ij} & \cdots \\ & \vdots & & & \vdots & \\ \cdots & \frac{\partial}{\partial\theta_j}Q_{ij} & \cdots & \cdots & \frac{\partial}{\partial V_j}Q_{ij} & \cdots \\ & \vdots & & & \vdots & \end{bmatrix} \begin{bmatrix} \vdots \\ \theta_i \\ \vdots \\ \vdots \\ V_i \\ \vdots \end{bmatrix} + Z^m, \tag{2.16}$$

where

$$\frac{\partial}{\partial\theta_j}P_i = \begin{cases} -|V_iV_jY_{ij}|\sin\left(\delta_{ij}+\theta_j-\theta_i\right), & i\neq j \\ \sum_{l\neq i}^{n}|V_iV_lY_{il}|\sin\left(\delta_{il}+\theta_l-\theta_i\right), & i=j \end{cases} \tag{2.17}$$

$$\frac{\partial}{\partial\theta_j}Q_i = \begin{cases} -|V_iV_jY_{ij}|\cos\left(\delta_{ij}+\theta_j-\theta_i\right), & i\neq j \\ \sum_{l\neq i}^{n}|V_iV_lY_{il}|\cos\left(\delta_{il}+\theta_l-\theta_i\right), & i=j \end{cases} \tag{2.18}$$

$$\frac{\partial}{\partial\theta_l}P_{ij} = \begin{cases} -|V_iV_jY_{ij}|\sin\left(\delta_{ij}+\theta_j-\theta_i\right), & l=j \\ |V_iV_jY_{ij}|\sin\left(\delta_{ij}+\theta_j-\theta_i\right), & l=i \end{cases} \tag{2.19}$$

$$\frac{\partial}{\partial\theta_l}Q_{ij} = \begin{cases} -|V_iV_jY_{ij}|\cos\left(\delta_{ij}+\theta_j-\theta_i\right), & l=j \\ |V_iV_jY_{ij}|\cos\left(\delta_{ij}+\theta_j-\theta_i\right), & l=i \end{cases}. \tag{2.20}$$

The linearized observation function in (2.14) is further simplified by setting the bus magnitude of every bus to 1.0 per unit, and by ignoring the shunt elements and the branch resistances, which leads to the *DC state estimation*. As a result, the vector of state variables of the DC state estimation case is given by

$$X^n = [\theta_2, \theta_3, \ldots, \theta_N]^{\mathrm{T}}, \tag{2.21}$$

i.e. the state variables to be estimated are the phase angles of the buses. Within the DC state estimation, only the real power injections and the real power flows are

considered, i.e. the vector of measurements is given by

$$Y^m = [P_1, \ldots, P_n, \ldots P_{ij} \ldots]^{\mathrm{T}}. \tag{2.22}$$

The real power flow from bus i to bus j in (2.8) is given by

$$P_{ij} = \frac{\theta_i - \theta_j}{B_{ij}}, \tag{2.23}$$

and the real power injection at bus i is given by

$$P_i = \sum_{j=1, j\neq i}^{N} P_{ij} \tag{2.24}$$

for all buses connected to bus i with $i = 1, \ldots, N$.

Classical Power System State Estimation

The aim of the state estimator $g : \mathbb{R}^m \to \mathbb{R}^n$ is to obtain an estimate of the state variables that minimizes a cost function $c : \mathbb{R}^m \to \mathbb{R}$ that describes the cost of the estimate $\hat{\mathbf{x}}$ with respect to real state $\mathbf{x}$.

For the observation model with linearized dynamics given by

$$Y^m = \mathbf{H}\mathbf{x} + Z^m \tag{2.25}$$

and the cost function given by

$$c\left(\hat{\mathbf{x}}, \mathbf{x}\right) = c\left(\hat{\mathbf{x}}, \mathbf{y}(\mathbf{x})\right) = \|\mathbf{y} - \mathbf{H}\hat{\mathbf{x}}\|_{\ell_2}^2 \tag{2.26}$$

with $\hat{\mathbf{x}}$ denoting an estimate of state variables and $\|\cdot\|_{\ell_2}$ denoting the ℓ_2 norm of the vector given in $\cdot$, the resulting estimator is given by

$$\hat{\mathbf{x}}^\star = \arg\min_{\hat{\mathbf{x}}} \|\mathbf{y} - \mathbf{H}\hat{\mathbf{x}}\|_{\ell_2}^2 = \left(\mathbf{H}^{\mathrm{T}}\mathbf{H}\right)^{-1}\mathbf{H}^{\mathrm{T}}\mathbf{y}, \tag{2.27}$$

which is the *LS estimate*. Specifically when $Z^m \sim \mathcal{N}(\mathbf{0}, \mathbf{\Sigma}_{ZZ})$,

$$\hat{\mathbf{x}}^\star = \arg\min_{\hat{\mathbf{x}}} \|\mathbf{y} - \mathbf{H}\hat{\mathbf{x}}\|_{\ell_2}^2 = \left(\mathbf{H}^{\mathrm{T}}\mathbf{\Sigma}_{ZZ}^{-1}\mathbf{H}\right)^{-1}\mathbf{H}^{\mathrm{T}}\mathbf{\Sigma}_{ZZ}^{-1}\mathbf{y}, \tag{2.28}$$

which is the *weighted least squares (WLS) estimate*.

Anomaly Detection Based on LS/WLS Estimation

As mentioned before, one of the objectives of the power system state estimation is bad data detection, i.e. the operator has to decide to accept or reject the measurements. Accepting the measurements means that the operator trusts the measurements obtained from the grid and uses them to dispatch the power or optimize the power flow; rejecting means that the measurements are not reliable and the system is operating in an abnormal condition.

Given the observation model in (2.14), the anomaly detection approaches are mainly residual-based. The residual of the LS or WLS estimation is given by

$$r = \|\mathbf{y} - \mathbf{H}\hat{\mathbf{x}}^{\star}\|_{\ell_2}^2, \tag{2.29}$$

where r is the residual. Using the residual, bad data detection is cast as a hypothesis testing problem with hypotheses

$$\begin{aligned} &\mathcal{H}_0 : r < \tau \qquad \text{no bad data} \\ &\mathcal{H}_1 : r \geq \tau \qquad \text{bad data presents,} \end{aligned} \tag{2.30}$$

where τ is a detection threshold set by the operator.

When the noise term is assumed to follow a zero mean multivariate Gaussian distribution with independent entries, the normalized residual, which is given by

$$r_n = \left(\mathbf{y} - \mathbf{H}\hat{\mathbf{x}}^{\star}\right)^{\mathrm{T}} \boldsymbol{\Sigma}_{ZZ}^{-1} \left(\mathbf{y} - \mathbf{H}\hat{\mathbf{x}}^{\star}\right), \tag{2.31}$$

is another option for anomaly detection. It is easy to show that

$$r_n \sim \chi_2^{m-n}, \tag{2.32}$$

where χ_2^{m-n} is the chi-square distribution with $m-n$ degrees of freedom. That being the case the hypothesis testing problem for the anomaly detection is given by

$$\begin{aligned} &\mathcal{H}_0 : r_n \in \chi_2^{m-n}(\alpha) \qquad \text{no bad data} \\ &\mathcal{H}_1 : r_n \notin \chi_2^{m-n}(\alpha) \qquad \text{bad data exists,} \end{aligned} \tag{2.33}$$

where $\alpha \in [0, 1]$ is the significance level, i.e. the probability of false alarm; and $\chi_2^{m-n}(\alpha)$ is the critical region of the χ_2^{m-n} distribution when the significance level is α.

2.1.2 State Estimation and Dad Data Detection within Bayesian Framework

Unlike the deterministic setting for the state variables in (2.14), the state variables are described by a vector of random variables X^n within the Bayesian framework. Also the conditional distribution of Y^m given X^n, i.e. $P_{Y^m|X^n}$, needs to be set to fit the observation process.

The Bayesian framework has two main advantages. Firstly the Bayesian framework takes the prior information about the state variables into consideration. The prior information about the state variables is represented by the distribution of the state variables, i.e. P_{X^n}. Through the conditional distribution $P_{Y^m|X^n}$, the prior knowledge about the state variables changes the statistical structure of the measurements. Secondly the Bayesian framework provides a probabilistic modeling of the state variables. The deterministic setting regards the state variable as an unknown but fixed parameter. But the state variables are considered as unknown and random variables within the Bayesian framework, which allows the uncertainty of the state variable and matches the real applications more.

State Estimation within Bayesian Framework

The observation model with linearized dynamics for the Bayesian framework is given by

$$Y^m = \mathbf{H}X^n + Z^m, \tag{2.34}$$

where X^n is the vector of random variables describing the states of the grid. Here the condition distribution $P_{Y^m|X^n}$ follows directly from the observation model in (2.34). That being the case, the vector of state variables X^n and the measurement vector Y^m are dependent with known joint distribution $P_{X^nY^m}$.

The commonly adopted cost function for the state estimation within the Bayesian framework is the mean squared error (MSE), which is given by

$$c = \mathbb{E}\left[c\left(X^n, \hat{X}^n\right)\right] = \mathbb{E}\left[\|X^n - \hat{X}^n\|_{\ell_2}^2\right], \tag{2.35}$$

where $\hat{X}^n$ is an estimate of X^n. The optimal estimator that achieves the minimum mean squared error (MMSE) is given by

$$\hat{X}^{n\star} = \arg\min_{\hat{X}^n} \mathbb{E}\left[\|X^n - \hat{X}^n\|_{\ell_2}^2\right] = \mathbb{E}\left[X^n|Y^m\right]. \tag{2.36}$$

When a realization $\mathbf{y}$ of measurement vector Y^m is available, the optimal estimator is given by

$$\hat{\mathbf{x}}^\star = \mathbb{E}\left[X^n | Y^m = \mathbf{y}\right]. \tag{2.37}$$

Specifically when $X^n \sim \mathcal{N}(\mathbf{0}, \mathbf{\Sigma}_{XX})$, the MMSE estimator is given by

$$\hat{\mathbf{x}}^\star = \mathbf{M}\mathbf{y}, \tag{2.38}$$

where $\mathbf{M}$ is given by

$$\mathbf{M} = \mathbf{\Sigma}_{XX}\mathbf{H}^{\mathrm{T}}\left(\mathbf{H}\mathbf{\Sigma}_{XX}\mathbf{H}^{\mathrm{T}} + \mathbf{\Sigma}_{ZZ}\right)^{-1} \tag{2.39}$$

with $\mathbf{\Sigma}_{ZZ}$ denoting the covariance matrix of the system noise.

Detection within Bayesian Framework

The detection within the Bayesian framework is usually cast as an M-ary hypothesis testing problem, in which the operator has to decide among M possible statistical situations describing the observations. Here we focus on the binary hypothesis testing problem, i.e. $M = 2$, described by

$$\begin{aligned} \mathcal{H}_0 :&\quad Y^m \sim P_0, \quad \text{versus} \\ \mathcal{H}_1 :&\quad Y^m \sim P_1, \end{aligned} \tag{2.40}$$

where P_0 and P_1 are two probability distributions. The null hypothesis $Y^m \sim P_0$ in $\mathcal{H}_0$ means that the measurements follow distribution P_0, and similarly the alternative hypothesis $\mathcal{H}_1$ means that the measurements follow distribution P_1. In the following without loss of generality, we assume that the null hypothesis $\mathcal{H}_0$ states that the system is safe, and the alternative hypothesis $\mathcal{H}_1$ states that the system is under an abnormal condition.

Two types of error exist in binary hypothesis testing: Type I error and Type II error. Type I error, or "*false alarm*", is the event that rejects a true null hypothesis, and the probability of Type I error is usually denoted by α. Type II error, or "*miss*", is the event that accepts a false null hypothesis, and the probability of Type II error is usually denoted by β. Table 2.1 summarizes the relation between Type I error and Type II error in the binary hypothesis testing.

Here we adhere to the Neyman-Person hypothesis testing framework, in which the decision rule aims to minimize the probability of Type II error with a given constraint on the probability of Type I error, i.e. maximize the probability of anomaly detection

Table 2.1. Relations between Type I error and Type II error in binary hypothesis testing

	Accept $\mathcal{H}_0$	Reject $\mathcal{H}_0$
$\mathcal{H}_0$ is true	$\surd$	Type I error (false alarm)
$\mathcal{H}_1$ is true	Type II error (miss)	$\surd$

with a given constraint on the probability of false alarm. The following lemma shows the optimality of likelihood ratio test (LRT) for hypothesis testing problem given in (2.40) within the Neyman-Person framework.

Lemma 2.1. *[21, Proposition II.D.1: Neyman-Pearson Lemma] Compared with the other tests that achieve probability of Type I error $\alpha \leq \alpha'$ in the hypothesis testing problem (2.40), the LRT given by*

$$L(\mathbf{y}) = \frac{f_{P_0}(\mathbf{y})}{f_{P_1}(\mathbf{y})} \underset{\mathcal{H}_1}{\overset{\mathcal{H}_0}{\gtrless}} \tau \tag{2.41}$$

achieves the minimum probability of Type II error, where τ is the value that achieves probability of Type I error equals to α' in (2.41); and $f_{P_0}(\cdot)$ and $f_{P_1}(\cdot)$ are the probability density functions (p.d.f.s) of distributions P_0 and P_1, respectively.

The Neyman-Person lemma states that the LRT is the optimal test in the sense that it maximizes the probability of detection for a given constraint on the probability of false alarm. However for the non-asymptotic setting, i.e. when the number of samples is finite, it is challenging to obtain the probability of detection under LRT. The following lemma characterizes the asymptotic probability of Type II error, i.e. probability of miss, for the LRT.

Lemma 2.2. *[22, Theorem 11.8.3: Chernoff-Stein Lemma] For the LRT given in (2.41), for any $\epsilon \in (0, 1/2)$,*

$$\lim_{k\to\infty} \frac{1}{k} \log \beta_k^\epsilon = -D(P_0 \| P_1), \tag{2.42}$$

where β_k^ϵ is the minimum probability of Type II error β when the probability of Type I error $\alpha < \epsilon$ and k samples are available; and $D(\cdot\|\cdot)$ is the KL divergence.

In a nutshell, the Chernoff-Stein Lemma states that the logarithm of the averaged minimum probability of Type II error β for any probability of Type I error α smaller than one half asymptotically converges to the negative of the KL divergence between the distributions of the two hypotheses for LRT in (2.41).

2.1.3 Advanced State Estimation

In addition to the conventional state estimation setting introduced before, there are some other advanced state estimation approaches for the power system case, such as nonlinear LS and dynamic state estimation.

The state estimation approaches proposed above rely on a linearized observation model. For the observation model with nonlinear dynamics, the cost function given in (2.26) is given by

$$c\left(\hat{\mathbf{x}}, \mathbf{x}\right) = c\left(\mathbf{y}, \mathbf{y}(\hat{\mathbf{x}})\right) = \|\mathbf{y} - H\left(\hat{\mathbf{x}}\right)\|_{\ell_2}^2. \tag{2.43}$$

Iterative approaches, such as the Gauss-Newton method, are employed to solve the nonlinear problem above. In these approaches, the estimate of the state variables is updated at each iteration based on the residual in a given iteration. However such iterative approaches are sensitive to the choice of the initial point, and the convergence is not guaranteed [23]. For this problem, [24] and [25] reformulate the nonlinear state estimation problem as a semidefinite programming problem, in which the complex expression of the bus voltages is regarded as state variables, i.e. $\mathbf{x} = \mathbf{v}$ in (2.43) with $\mathbf{v}$ denoting the vector of bus voltages in complex expression. The resulting estimator is given by

$$\hat{\mathbf{v}} = \arg\min_{\hat{\mathbf{v}}} \sum_{i=1}^{m} \left(\mathbf{y}_i - H_i\left(\hat{\mathbf{v}}\right)\right)^2 \tag{2.44}$$

$$s.t. \quad \mathbf{i} = \mathbf{Y}\mathbf{v}, \tag{2.45}$$

where $\mathbf{i}$ is the vector of bus currents, and $\mathbf{Y}$ is the admittance matrix which is determined by power system parameters [17, pp.32]. Note that the power flows and bus injections are quadratic function of bus voltage $\mathbf{v}$, the estimator in (2.44) is reformulated as

$$\hat{\mathbf{V}} = \arg\min_{\hat{\mathbf{V}}} \sum_{i=1}^{m} \left(\mathbf{y}_i - \mathrm{tr}(\mathbf{H}_i\hat{\mathbf{V}})\right)^2 \tag{2.46}$$

$$s.t. \quad \hat{\mathbf{V}} \succcurlyeq \mathbf{0}$$

$$\mathrm{rank}\left(\hat{\mathbf{V}}\right) = 1,$$

where

$$\mathbf{V} = \mathbf{v}_e\mathbf{v}_e^{\mathrm{T}}, \quad \text{with} \quad \mathbf{v}_e = \left[\mathrm{real}(\mathbf{v}^{\mathrm{T}}) \; ; \quad \mathrm{imaginary}(\mathbf{v}^{\mathrm{T}})\right]^{\mathrm{T}}; \tag{2.47}$$

and $\mathbf{H}_i$ is a matrix obtained from admittance matrix $\mathbf{Y}$, which is determined by power system parameters, c.f. [24, (11a) - (11e)]. [24] and [25] show that finding

the estimator in (2.46) is a relaxed semidefinite programming problem, and propose the necessary and sufficient condition for the existence of the optimal solution. Specifically [25] solves the semidefinite programming problem in a decentralized pattern.

The states of the power system evolve over time, so dynamic modeling of the state evolution allows dynamic state estimation [23]. The evolution of the state variables is modeled by

$$\mathbf{x}(t+1) = \mathbf{F}(t)\mathbf{x}(t) + \mathbf{z}, \tag{2.48}$$

where $\mathbf{x}(t)$ is the vector of state variables at time $t \in \mathbb{R}^+$, and $\mathbf{F}(t)$ is the state transition matrix. Specifically when $\mathbf{F} = \mathbf{I}$, the state evolution in (2.48) changes to a "*random walk*" [26]. Another approach to include the state dynamics into account is the state space model given by

$$\mathbf{x}(t+1) = \mathbf{F}(t)\mathbf{x}(t) + \mathbf{B}(t)\mathbf{u}(t) + \mathbf{z} \tag{2.49}$$

$$\mathbf{y}(t+1) = \mathbf{C}(t)\mathbf{x}(t) + \tilde{\mathbf{z}}, \tag{2.50}$$

where $\mathbf{B}(t)$ is the actuation matrix at time t, $\mathbf{u}(t)$ is the input to the system at time t, $\mathbf{C}(t)$ is the observation matrix at time t, and $\tilde{\mathbf{z}}$ is the additive white Guassian noise (AWGN) in the measurements. The state space model allows considering the state estimation problem in a control-theoretic framework. Given the sate space model or state evolution model, Kalman filter techniques obtain the estimate of the state variables [26].

There are also some other approaches for state estimation. For example, the operator can infer the topology of the power system when information about the status of the breakers in the power system is available, and estimate the state variables based on the inferred topology.

2.2 Data Injection Attacks

In the following, we introduce DIAs and their different approaches to the attack formulation. We first focus on DIAs targeting on the LS estimation with different residual-based detection approaches. Then we show DIAs construction and detection within the Bayesian framework. In the end, we review the data-driven approach for DIAs construction and detection, such as machine learning approaches and statistical learning approaches.

2.2.1 DIAs Against LS Estimation and Residual-Based Detection

DIAs are a kind of cyber threats that target state estimation of power systems, and they were first proposed by [7] and [8]. Therein, DIAs aim to disrupt the state estimation by compromising the measurements available to the operator, i.e. injecting some extra terms into the true measurements, which are modeled as, from (2.14),

$$Y_A^m = \mathbf{H}\mathbf{x} + Z^m + \mathbf{a}, \tag{2.51}$$

where $\mathbf{a} \in \mathbb{R}^m$ is the attack injected by the attacker. Furthermore, [7] and [8] prove the following lemma.

Lemma 2.3. *For any attacks constructed as*

$$\mathbf{a} = \mathbf{H}\mathbf{c}, \ \forall \mathbf{c} \in \mathbb{R}^n \tag{2.52}$$

the attacks are undetectable under the residual detection given in (2.30).

Proof. For any attacks constructed as $\mathbf{a} = \mathbf{H}\mathbf{c}$, the residual given in (2.29) of LS estimation for the observation model (2.51) is the same as the residual for the observation model (2.14), i.e.

$$\|\mathbf{y}_a - \mathbf{H}\hat{\mathbf{x}}_a\| = \|\mathbf{y} + \mathbf{a} - \mathbf{H}(\hat{\mathbf{x}} + \mathbf{c})\| = \|\mathbf{y} - \mathbf{H}\hat{\mathbf{x}} + \mathbf{a} - \mathbf{H}\mathbf{c}\| = \|\mathbf{y} - \mathbf{H}\hat{\mathbf{x}}\|, \tag{2.53}$$

where $\mathbf{c}$ is the extra term injected into the estimated state variables via attack $\mathbf{a}$, $\hat{\mathbf{x}}_a$ is the estimated state variables using compromised measurements $\mathbf{y}_a$, and $\mathbf{y}_a$ is a realization of compromised measurement Y_A^m. Note that the result in (2.53) holds for any vector norm. This implies that the residual-based detection approaches are easily bypassed by the attacks given in (2.52), as the residual of LS estimation is unchanged. □

The structure $\mathbf{a} = \mathbf{H}\mathbf{c}$ implies that the attacker needs to get access to the Jacobian matrix $\mathbf{H}$, which is defined in (2.15). Although the assumptions for the DIAs, i.e. perfect knowledge of system and capability of compromising measurements, are strong, the attacker that launched the cyber attack towards Ukraine in 2015 truly meets these assumptions [27].

The DIAs compromise the measurements at three different periods: the sensing period, the communicating period, and the SCADA processing period. The attacks that target these three periods are shown by A1, A2, and A3 in Fig. 2.2, respectively

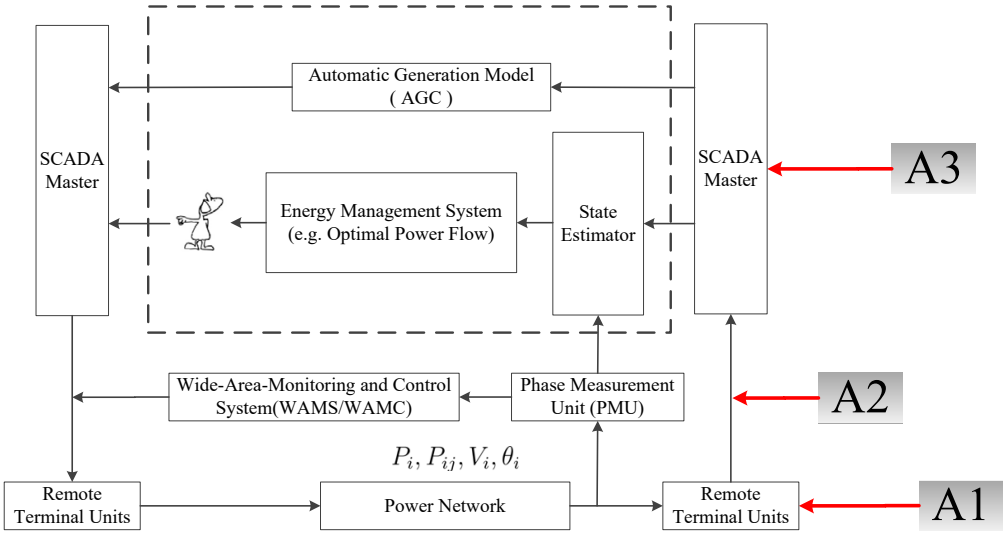


Fig. 2.2. Data injection attack launch points of the power system network.

[28]. Although Fig. 2.2 represents the attacks on the power system, it is easy to extend this framework to any cyber-physical systems or networked systems [29]. Without triggering the residual-based alarm, DIAs are also capable of destabilizing control systems, for which the dynamic of state variables are modeled by a state space model [30]. Also the attack targeting at the transmission system is extended to the distribution system of the power system, which usually has low transform ratio at the transformers in the system [31].

The DIAs proposed above target the observation model with linearized dynamics, which allows closed-form expressions and analytical results for the attack construction. For the observation model with nonlinear dynamics, i.e. AC state estimation, [32] shows that the optimal attack in (2.52) is usually not the optimal in the AC case, but it still has acceptable performance.

In the following part, we review the DIAs for the LS estimation and residual-based detection at first. Then two assumptions for the attacker, i.e. ability to compromise meters and perfect information of systems, are relaxed. Since changing the state variables affects the decisions made by the operator, the DIAs that target some specific decisions, such as OPF, are covered in the Extended DIAs subsection. We finish by reviewing the state of the art detection and attack protection mechanisms again DIAs.

Sparse Attacks

One of the assumptions made by [7] and [8] is that the attacker has the capability to compromise meters in the power system. However compromising sensors in power

systems is usually costly for the attacker. As a result, the attacker has to minimize or limit the number of sensors that need to be hacked. Also the operator protects some of the sensors in the system, which implies that the attacker cannot compromise the measurements from the protected sensors. This problem is called *sparse attack construction problem* by [7], which is given by

$$\begin{aligned} \min_{\mathbf{a}} \quad & \|\mathbf{a}\|_{\ell_0} && (2.54)\\ \text{s.t.} \quad & \mathbf{a} = \mathbf{H}\mathbf{c} && (2.55)\\ & \mathbf{a}_i = 0 \ \text{ for all } \ i \notin \mathcal{S}_a, && (2.56) \end{aligned}$$

where $\|\mathbf{a}\|_{\ell_0}$ is the ℓ_0 norm of vector $\mathbf{a}$ and $\mathcal{S}_a$ is the set of compromisable sensors. However, it is difficult to solve this problem as the problem is usually NP-hard. To solve this problem, [7] proposes an equivalent expression for the detection constrain in (2.55), which is given by

$$\mathbf{a} = \mathbf{H}\mathbf{c} \Longleftrightarrow \mathbf{B}\mathbf{a} = \mathbf{0}, \tag{2.57}$$

where $\mathbf{B} = \mathbf{H}\left(\mathbf{H}^{\mathrm{T}}\mathbf{H}\right)^{-1}\mathbf{H}^{\mathrm{T}} - \mathbf{I}_m$. Then some heuristic or greedy algorithms, such as the matching pursuit algorithm [33] and the orthogonal matching pursuit [34], are used to find a sparse attack vector subject to the unobservable constraint in (2.57). The heuristic algorithms are also proposed by [35] and [36] to solve this problem. Specifically [35] provides an upper bound for the number of sensors that need to be compromised to construct sparse attacks. The minimum number of sensors that need to be compromised is also studied in [37] and [38]. [37] shows that the necessary and sufficient condition for the attack to be undetectable under residual detection is that the Jacobian matrix describing the relation between the measurements from the uncompromisable sensors and the state variables is rank deficient. And [38] obtains the same result in [37] via a graph-theoretical approach.

Unlike the heuristic algorithms, [39] relaxes the ℓ_0 norm of the attack vector in (2.54) to the ℓ_1 norm, which turns this problem into a convex problem given by

$$\begin{aligned} \min_{\mathbf{a}} \quad & \|\mathbf{a}\|_{\ell_1} \\ \text{s.t.} \quad & \mathbf{a} = \mathbf{H}\mathbf{c} \\ & \mathbf{a}_i = 0 \text{ for all } i \notin \mathcal{S}_a. && (2.58) \end{aligned}$$

In [39], the sparsity of the attack vector and the sparsity of the compromised state variables are achieved simultaneously by the reweight approach proposed by [40]. The ℓ_1 relaxation approach is also utilized in [41], [42], and [43] to construct sparse attacks.

Usually the solution to the ℓ_1 norm optimization problem is not the same as the original ℓ_0 norm problem. However [42] proves that when there is no measurements about the power injections or the power injections are not used in the state estimation, the solution to the ℓ_1 norm problem is the same as the solution for the ℓ_0 norm case.

Finding the minimal number of sensors need to be compromised not only leads to sparse attack construction for the attacker, but also helps the operator to evaluate the vulnerability of the power system. [41] uses the minimal number of meters needed to be compromised for injecting some certain error into a specific state variable as an index for the cyber-security of the power system. Also the minimal number of meters needed to be compromised to launch an unobservable attack is used as an index of security for the power system in [44] and for the joint system of power system and communication system in [45].

Attack Construction with Incomplete System Information

The other assumption made by [7] and [8] is that the attacker needs to know the Jacobian matrix $\mathbf{H}$, which is determined by the power system topology, the system parameters, and the operation point. Incomplete system information known by the attacker results in a mismatch between the actual Jacobian matrix $\mathbf{H}$ and the Jacobian matrix $\tilde{\mathbf{H}}$ the attacker has, which is modeled as

$$\tilde{\mathbf{H}} = \mathbf{H} + \Delta\mathbf{H}, \tag{2.59}$$

where $\Delta\mathbf{H}$ is the mismatch. The mismatch $\Delta\mathbf{H}$ increases the probability of the attack being detected. For example, [46] shows that when the attacker has limited information about the admittance of branches, the mismatch $\Delta\mathbf{H}$ truly increases the probability of detection under residual-based detection. The enough condition for the attack to be undetectable under residual detection is also proposed in [46].

For the case that the attacker only has perfect knowledge about a part of the grid, [47], [48], and [49] show that the attack is still able to be stealthy. [47] proves that when the line incidence matrix and the line admittances of the attack region in the system are perfectly known by the attacker, the attack is still undetectable when the state variables representing the boundary buses of the region change the same amount. The result in [47] is extended to the AC state estimation case in [50]. The choosing of such an attack region is proposed [48] with the consideration of using less information of the power system. Specifically when it is not feasible for the attack to be stealthy using partial information, the *framing attack* proposed by [49] is able to make the attack stealthy. When the operator chooses to remove the abnormal part from the measurements during the state estimation, the framing attack makes the

operator removing some key measurements to guarantee the stealthy of the attacks. Also the imperfect knowledge of the power system further impacts the performance of the attacks on the electricity market. [51] studies the impact of imperfect attacks on the electricity market, in which the state variables and the system model are modeled with uncertainty. Especially, the optimal stochastic guarantees for the attack and the resulting economic impacts are proposed.

Extended DIAs

The objectives of DIAs are not limited to just disrupting state estimation. The compromised state variables further affect the decisions made by the operator, such as OPF. The operator chooses the configuration that minimizes the generation and distribution cost for power systems, and the resulting power flow is called OPF. Also some constraints, such as the balance between generation and consumption, are added in the problem of determining the OPF. The impact of compromising the generation of the generator on the DC OPF is analyzed in [52], in which the generations of generators are the decision variables for the OPF. Also [53] illustrates the impact of DIAs on the OPF within an integrated simulation platform.

Changing the estimated state variables to some value also brings economic benefits to the attacker within the electricity market. [54] shows that changing the state variables using DIAs leads to the change of nodal price in the electricity market, and then the attacker makes profits during the virtual bidding period. The impact of DIAs on the locational marginal prices is studied in [55]. Furthermore [56] considers the benefits that the attacker obtained from the electricity market when the attacker only has some samples of the measurements, for which the attacker has to infer the system information from the samples. [57] shows that the operator cheats the users in the power system to pay extra for the electricity bill by changing the status of the breakers in the power system without causing any security threats. Except for the induced finical profits, DIAs are also able to mask some physical faults in the power system. For example, the measurements corrupted by the attacks make the operator ignore the outage of the transmission line [58, 59].

The interaction between attacks from the attacker and defense from the operator forms a game, in which the attacker updates the attack strategy according to the information the attacker has about the operator and the same for the operator. [60] studies the behavior of the attacker and the operator using game theory, in which the attacker tries to make profit from the electricity market and the operator tries to protect the sensors in the system to detect the attack. [61] considers the same scenario when two attackers and one operator in the system. It is worth to note that

when the attackers are uncoordinated, the effects of these attackers are eliminated by each other [62].

The target measurements of the attacker are not limited to the power flow or the power injection. Other than compromising the usual measurements in power systems, the on or off status of the switches and the load information in the power system are also potential targets for the attacker. Changing the statue of the switches results in the changes of the power system topology, which leads to the "*topology attack*" proposed by [63]. Also [63] proposes the sufficient condition for the topology attacks to be undetectable under residual detection. Compromising of the load information deviates the frequency of the power system away from the normal value [64], which is named "*load alter attack*" by [65]. The timestamp of phase measurement units (PMUs) is also a target for the attacker, this leads to a delay of transmission of measurements [66]. Only changing the timestamp of PMUs guarantees that the attacks are still undetectable under the residual detection. Also the energy trading or energy transmission between different parts in the power system is a potential target of DIAs. [67] chooses the distributed energy routing that guarantees the economic operation of the grid with multiple demanders or suppliers as the target to attack. Specifically, the corruption of the energy supply quantity, the energy request quantity and the link state of energy transmission is studied in a simulation-based way. The impact of DIAs on the load sharing between microgrids is analyzed in [68], in which the region and the sufficient condition for the stability of the microgrids under DIAs is proposed.

Detection and Protection

As the countermeasures for DIAs, different kinds of detection approaches are implemented by the operator to detect the attacks. The compromised measurements lead to the derivation of preset variables from the nominal values. For example, [69] shows that when the attacker compromises the measurements of currents, current angles or voltage angles, the calculated impedances of the branches in the power system are different from the setting known to the operator. The update of the classical detection approaches in (2.30) and (2.33) also helps the operator to detect the attacks. [70] proposes a trimmed least squares based detection approach, and several least trimmed square detectors are implemented together to increase the accuracy of attack detection. When further information or actions are available for the operator, the operator uses this information to detect the attack. For example, load forecasts are utilized in [71] and [72] to detect attacks. Also [73] shows that when the operator knows the location about the protected meters and the compromisable meters, only the topology information is sufficient for the operator to know whether

the attack is undetectable or not. [74] considers the scenario that the operator has the ability to shut down one of the transmission lines to detect the attack, and it shows that the attacks are detectable when the connectivity of the grid is larger than two.

The detection approaches reviewed above are implemented in a centralized way, also the decentralized state estimation is another option for the operator to detect the attacks. [75] proposes an adaptive partitioning approach for the grid to implement decentralized state estimation to detect an attack. Partitioning the grid increases the sensitivity of the chi-squared detection in the separated system. The decentralized detection approach is also proposed in [76] for the attack targeting at the electro-mechanical swing dynamics of the generator. Also [77] proposes a distributed detection and estimation approach for simultaneously DIAs and Jamming attacks.

For the dynamic setting of state estimation in (2.49) and (2.50), [78] designs a detection method for state estimation with Kalman filter, in which a new chi-squared metric for residual and an Euclidean distance metric for residual are proposed to detect the attack. To detect the replay attack that changes the current measurements to some historical measurements, [79] shows that when the operator of the power system adds watermark, i.e. white noise, into the input signal of the system, the attack is detectable. Here the design of the optimal watermark is cast as an optimization problem, in which the probability of attack detection is maximized and the disruption caused by the watermark is constrained.

Except for the passive detection approach, the active protecting approaches are also implemented by the operator to protect the system. Through these approaches the operator increases the difficulty of the attacker launching the attack or makes the undetectable attack impossible. The measurements gathered by the PMUs are with timestamps, which are difficult to corrupt. Therefore, using the PMUs to collect measurements is an effective tool to protect the power system. [80] shows that placing PMU at a bus not only guarantees the security of this bus, but also the buses connected to the bus with PMU. The number of PMUs needed to protect the power system is studied in [81] and [82]. [81] shows that the number of PMUs should be larger than the number of state variables to detect the attacks that are undetectable under residual detection. Also [82] shows that DIAs are detectable when the magnitude of the buses is protected securely.

Other than using the PMUs to guarantee the integrity of the measurements, encrypting the measurements in the power system also guarantees that the measurements that the operator obtained are accurate. [83] uses the McEliece cryptographic schemes to resist the DIAs. Also coding the measurements is helpful for the operator to protect the measurements [84]. When the attacks passed through the detection

mechanism, the operator is able to implement corresponding approaches to mitigate or eliminate the effect of the undetected attacks. For example, [85] studies mitigating the physical overload induced by attacks using a corrective dispatch.

2.2.2 DIAs within Bayesian Framework

The DIAs proposed by [7] and [8] are undetectable for the LS estimation in (2.27) and residual detection in (2.30). Unlike [7] and [8], the DIAs within the Bayesian estimation framework, which is covered in Section 2.1.2, are considered by [44], [86], [87], and [88].

Within the Bayesian framework, the distribution of the measurements changes from the distribution under normal condition to the distribution under attack. So the detection of attacks is cast as a hypothesis testing problem given in (2.40). Under this setting, the probability of detection for the attack is given by

$$\mathsf{P_D} \triangleq \int_{\mathcal{S}} \mathrm{d}P_{Y_A^m}, \tag{2.60}$$

where $P_{Y_A^m}$ is the distribution of the measurements under attack and $\mathcal{S}$ is the set of all the realizations of Y_A^m that being detected by the detection approach. Given the fact that the LRT is of fastest decay rate for probability of miss [44] and is optimal in the sense that it achieves the maximum probability of detection (Lemma 2.1), the LRT is the most commonly adopted detection approach for the attack [44, 86–88]. On the other hand, the disruption caused by attacks is measured by the extra MMSE induced by the attack in [44] and [86]. Information-theoretic measures are adopted by [87] and [88] to quantify the disruption, in which the attacker minimizes the amount of information that the measurements contained about the state variables.

The caused disruption and the probability of detection are two contradictive objectives with the Bayesian framework. In [44] and [86], the tradeoff between disruption and probability of detection is cast as an optimization problem given by

$$\begin{aligned} &\max_{\mathbf{a}} \ \|\mathbf{Ma}\|^2 \\ &\text{s.t. } \mathsf{P_D} \leq \tau' \end{aligned} \qquad \text{or} \qquad \begin{aligned} &\min_{\mathbf{a}} \ \mathsf{P_D} \\ &\text{s.t. } \|\mathbf{Ma}\|^2 \geq \tau' \end{aligned}$$

in which $\mathbf{M}$ is given in (2.39) and τ' is a threshold set by the operator, i.e. maximize the distortion with a constraint on the probability of detection or minimize the probability of detection with a constraint on disruption. For the information-theoretic attacks adopted by [87] and [88], the information loss caused by the attacks and the asymptotic probability of detection is summed up, in which a weighting parameter is assigned to the objective representing the probability of detection. The weighting

parameter reflects the preference of the attacker and allows the attacker tuning the probability of detection. Details for [87] and [88] are provided in Chapter 3 and Chapter 4.

For the Bayesian framework that is considered in [44] and [86], the tradeoff between disruption and probability of detection is also considered by [89] under the dynamic setting. Furthermore the Bayesian framework with dynamics for packet substitution attack and the extra package injection attack is considered by [90], in which a two-step approach, i.e. predication-correction, is proposed for the attack detection and state estimation.

2.2.3 Data-Driven DIAs

Power systems are of large scale and are with lots of sensors placed across the systems. As a result, there are large numbers of measurements generated every day. Using these measurements, the attacks are still able to be stealthy even when the attacker has incomplete information about the power system. The available measurements allow the attacker constructing the attacks via learning, in which the attacker learns the system information from the measurements. The learning of power system topology information from bus injection measurements is studied in [91], in which the learning is cast as a maximum a posteriori problem of the system parameters under sparsity constraints. Except for the topology information, some other information is also being extracted from the measurements. For example, the power flow measurements or the power injection measurements are used by the attacker to infer the operation point in [31]. Learning the system information is also studied in [92], [93], and [94], in which the statistical behavior of the measurements is extracted by statistical tools, such as principal component analysis (PCA) in [92] and independent component analysis in [93], to construct stealth attacks.

As countermeasures for DIAs, the statistical behavior of the state variables or the measurements are used by the operator to detect the attacks, as the distribution of the state variables or the measurements changes when the attacks launch. [95] proposes a detection approach for AC state estimation using the difference between the distribution of measurements obtained from historical data and the distribution at current time, in which two different measures, i.e. absolute distance and KL divergence, are utilized to quantify the differences in distributions. The change of distribution is also utilized by [96] to detect the attack when the state variables are modeled by a Gaussian Markov random field. Specifically the decentralized detection method proposed by [96] compares the marginal distributions of state variables under normal condition and under abnormal condition to locate the attack.

Expect the statistical quantify of the difference in distribution, such as KL divergence, the LRT is also a powerful tool to distinguish the distributions. The generalized LRT is adopted in [97] for the observation model with white noise and in [98] for the observation model with colored noise. Specifically the generalized LRT in [97] achieves the minimal averaged maximum delay of the detection. The likelihood of the state variables is also utilized to detect the attack by [99], in which the behavior of the attacker is modeled as a Markov decision process.

When the power system operates at steady status, the measurements at different time instants are of small difference, so the matrix of measurements at different time instants is of low rank. The low rank characteristic of the matrix of measurements and the sparsity of the attacks are used by [100] to detect the attack. Also a compromised measurement recovery approach is proposed in [101] using the low rank characteristic of the matrix of measurements. However the rank minimization in [101] is a NP hard problem, so the nuclear norm is utilized to approximate the rank operation. The data recovery in the dynamic setting is studied by [102].

As a powerful and useful tool to classify and cluster data, the machine learning approach is also capable to detect the attack. Different machine learning algorithms, such as support vector machine (SVM) and Adaboost, are compared in [103] for attack detection. The distributed SVM and PCA are utilized by [104] to distinguish the abnormality of the grid. Some other machine learning algorithms, such as common path mining [105], density ratio estimation [106], and margin setting algorithm [107], are also capable to detect the attacks. As a powerful tool in machine learning, the neural network is also utilized to detect the DIAs, which lead to a black box modeling problem. Given the training measurements, [108] uses the deep neural network to detect the attacks in the system. Other than using the training measurements directly, [109] uses the wavelet transform of the time series of the estimated state variable as features and trains the deep neural network to detect the attacks.

2.3 Summary

In this chapter, the state estimation problem is formulated for the power system cases, including the classical state estimation and anomaly detection, as well as the Bayesian case estimation and detection. Within the presented estimation and abnormal data detection framework, the DIAs construction is posed as a sparse attack construction problem for the sensor constrained case. Similarly, the case with imperfect knowledge is introduced and analyzed within this setting. We finish by providing an overview of learning approaches for DIAs construction and detection.

Therein, the attacker learns the information about the system from a limited number of training samples.

Chapter 3

Information-Theoretic Stealth Attacks

In this chapter, DIAs that utilize information-theoretic measures as merit metrics are proposed. Specifically, the attacker minimizes the mutual information between the state variables and the compromised measurements to minimize the amount of information that the operator obtained from the measurements about the states variables. Meanwhile the attacker minimizes the KL divergence between the distribution of measurements with attack and the distribution without attack to minimize the asymptotic probability of detection. The proposed stealth attacks minimize the sum of these two information-theoretic objectives. Closed-form expression of the stealth attacks is obtained for the Gaussian state variables and Gaussian attacks case.

3.1 Bayesian Framework for State Estimation

The observation model with linearized dynamics for power system state estimation problem is given in (2.34), i.e.

$$Y^m = \mathbf{H}X^n + Z^m, \tag{3.1}$$

where $X^n \in \mathbb{R}^n$ is the vector of random variables describing the true state of the system; $\mathbf{H} \in \mathbb{R}^{m\times n}$ is the Jacobian matrix, which is defined in (2.15); $Y^m \in \mathbb{R}^m$ is the vector of random variables containing the measurements available to the attacker; and $Z^m \in \mathbb{R}^m$ is the AWGN introduced by the sensors in the power system [17, 20],

The work in Chapter 3 is published in "K. Sun, I. Esnaola, S.M. Perlaza, and H.V. Poor, "Information-theoretic attacks in the smart grid," in *Proc. IEEE Int. Conf. on Smart Grid Commun.*, Dresden, Germany, Oct. 2017, pp. 455-460.".

i.e. the vector of random variables Z^m follows a multivariate Gaussian distribution $\mathcal{N}(\mathbf{0}, \sigma^2\mathbf{I}_m)$, in which σ^2 is the noise variance.

3.1.1 State Variable Model

In the Bayesian framework that is reviewed in Section 2.1.2, the state variables are described by a vector of random variables and follow a given distribution. Modeling the distribution is arbitrary and has been tackled using different approaches in the literature. The state variables are modeled by a joint Gaussian distribution in [44] and [86], and by a general distribution in [110] and [111] for the power system state estimation scenario. The statistical modeling of the state variables within power systems is studied by [112] and [113] when the bus voltage magnitudes are chosen to be state variables. It is shown in [112] that the bus voltages of a low voltage distribution system in the north-west of England are well described by a multivariate Gaussian distribution, and [113] shows that the Gaussianity of the voltage magnitude for a 6-Bus circuit is a sensible modeling assumption. In [114] it is shown that the measurements, i.e. active power consumption and reactive power consumption, from real power grids also follow a joint Gaussian distribution. The Gaussian distribution also has some other advantages. For example, the Gaussian distribution has the maximum entropy, i.e. maximum uncertainty, among all the real-valued distributions that have the same variance, which makes the modeling of the state variables more robust [22].

As a result, here the state variables are assumed to follow a multivariate Gaussian distribution denoted by

$$X^n \sim \mathcal{N}(\mathbf{0}, \mathbf{\Sigma}_{XX}), \tag{3.2}$$

where $\mathbf{\Sigma}_{XX} \in \mathcal{S}_+^n$ is the covariance matrix of the state variables with $\mathcal{S}_+^n$ denoting the set of positive semi-definite matrices of dimension $n \times n$. Consequently, from (3.1), the measurement vector also follows a multivariate Gaussian distribution given by

$$Y^m \sim \mathcal{N}(\mathbf{0}, \mathbf{\Sigma}_{YY}), \tag{3.3}$$

where $\mathbf{\Sigma}_{YY} = \mathbf{H}\mathbf{\Sigma}_{XX}\mathbf{H}^{\mathrm{T}} + \sigma^2\mathbf{I}_m$ is the covariance matrix of the measurements.

3.1.2 Random Attack Model

The observation model for the case in which the measurements are compromised is given by

$$Y_A^m = \mathbf{H}X^n + Z^m + A^m, \tag{3.4}$$

where $A^m \in \mathbb{R}^m$ is the attack vector [7]. Given the stochastic nature of the state variables, it is reasonable for the attacker to pursue a stochastic attack construction strategy. As a result, an attack vector independent of the state variables is constructed as

$$A^m \sim P_{A^m}, \tag{3.5}$$

where P_{A^m} is the distribution of A^m.

The i-th element of the vector of the compromised measurements is given by

$$(Y_A)_i = (\mathbf{H})_i X^n + Z_i + A_i, \tag{3.6}$$

in which $(\mathbf{H})_i$ is the i-th row of $\mathbf{H}$, $(Y_A)_i$ is the i-th elements of Y_A^m. It is shown in [115] that Gaussian distribution is the distribution that achieves

$$\min_{\mathbb{E}[A_i]=\sigma_a^2<\infty} I\left(X^n; (Y_A)_i\right) \tag{3.7}$$

for $i = 1, \ldots, m$, where $I(\cdot;\cdot)$ is the mutual information between two state variables given in $\cdot$, which is defined in Definition A.12; and σ_a^2 is the variance of random variable A_i. It is worth to point out that this result holds for any distribution of the state variables, i.e. for any P_{X^n}. This implies that the additive attack distribution that minimizes the mutual information between the vector of state variables and the compromised measurements under a fixed covariance for the attack term is Gaussian. For this reason, in this thesis we adopt a Gaussian random attack framework. While the Gaussian distribution guarantees the minimization of the mutual information under second order constraints, we do not have a formal justification when detection constraints are introduced. However, we adopt the Gaussian attack construction for the rest of this thesis.

In the following, an attack vector independent of the state variables is constructed following a multivariate Gaussian distribution denoted by

$$A^m \sim \mathcal{N}(\mathbf{0}, \mathbf{\Sigma}_{AA}), \tag{3.8}$$

where $\mathbf{\Sigma}_{AA} \in \mathcal{S}_+^m$ is the covariance matrix of the attack vector. As a result of the linearity in (3.4) and the Gaussianity of the attack vector, the compromised measurements, i.e. Y_A^m, follow a multivariate Gaussian distribution described as

$$Y_A^m \sim \mathcal{N}(\mathbf{0}, \mathbf{\Sigma}_{Y_A Y_A}), \tag{3.9}$$

where $\mathbf{\Sigma}_{Y_A Y_A} = \mathbf{H}\mathbf{\Sigma}_{XX}\mathbf{H}^{\mathrm{T}} + \sigma^2\mathbf{I}_m + \mathbf{\Sigma}_{AA}$.

It is worth noting that the independence of the attack vector with respect to the state variables implies that the attacker does not need to know the joint distribution of the state variables and the measurements to construct the attack vector. Knowledge of the second order moments of the state variables and the variance of the AWGN introduced by the observation process suffice to construct the attack. This assumption significantly reduces the difficulty of the attack construction. Later we show in Theorem 3.1 that the variance of the AWGN introduced by the observation process is not required to construct the Gaussian attacks.

3.1.3 Attack Detection Formulation

The detection problem within the Bayesian framework is usually cast as a hypothesis testing given in (2.40). Given the distribution of the measurements without attack and under attack in (3.3) and (3.9), respectively, the attack detection problem is cast into a hypothesis testing problem with hypotheses

$$\begin{aligned} &\mathcal{H}_0 : Y^m \sim \mathcal{N}(\mathbf{0}, \mathbf{\Sigma}_{YY}), \quad \text{versus} \\ &\mathcal{H}_1 : Y^m \sim \mathcal{N}(\mathbf{0}, \mathbf{\Sigma}_{Y_A Y_A}). \end{aligned} \tag{3.10}$$

The null hypothesis $\mathcal{H}_0$ describes the case in which the power system is not compromised, while the alternative hypothesis $\mathcal{H}_1$ describes the case in which the power system is under attack.

The Neyman-Pearson Lemma ([116], or Lemma 2.1) states that for a fixed probability of Type I error, the LRT achieves the minimum probability of Type II error β, when compared with any other tests with an equal or smaller probability of Type I error α. In view of this, a LRT is chosen as the attack detection strategy. The LRT between $\mathcal{H}_0$ and $\mathcal{H}_1$ takes the following form

$$L(\mathbf{y}) = \frac{f_{Y_A^m}(\mathbf{y})}{f_{Y^m}(\mathbf{y})} \underset{\mathcal{H}_0}{\overset{\mathcal{H}_1}{\gtrless}} \tau, \tag{3.11}$$

where $\mathbf{y} \in \mathbb{R}^m$ is a realization of the vector of random variables modeling the measurements, $f_{Y_A^m}$ and f_{Y^m} denote the p.d.f. of Y_A^m and Y^m, respectively, and τ is the decision threshold set by the operator to meet the false alarm constraint.

3.2 Information-Theoretic Objectives

3.2.1 Disruption Measure

The information-theoretic attacks aim to disrupt the state estimation by minimizing the amount of information that the measurements contained about the state variables. By doing so, the amount of information retrieved from the measurements about the state variables during the state estimation is minimized. So the operator obtains less information about the state variables, and the estimation, forecasting, and control the operator conducts with these measurements are carried out with less information about the state of the system.

The mutual information between two random variables is a measure of the amount of information obtained about one random variable through observing the other random variable. Consequently, the amount of information that the vector of measurements contains about the vector of state variables is determined by the mutual information between the vector of state variables and the vector of measurements. Information measures have previously been used to quantify the amount of information acquired by different monitoring systems in a smart grid context. For instance, in [12] mutual information is used to quantify the amount of information obtained by PMUs from the grid about the state variables. Similarly, mutual information is used in [13] and [14] to quantify the amount of information leaked by smart meters in the power system. Except for the smart grid context, the mutual information is also utilized in [117] for parameter setting in machine learning algorithm.

Capitalizing on the Bayesian framework and the observation model under attack in (3.4), the attacker constructs the attack vector, i.e. chooses the distribution of the attack vector, in such a way that it

$$\min_{A^m} I(X^n; Y_A^m). \tag{3.12}$$

This is equivalent to guaranteeing that the amount of information that the operator acquires about the state variables X^n by observing Y^m is minimized.

Minimizing the mutual information between the state variables and the compromised measurements leads to an increase in the MMSE defined in (2.35). Specifically

[118] proves that for the linear system given by

$$Y^m = \sqrt{\mathrm{SNR_L}}\mathbf{H}X^n + Z^m \tag{3.13}$$

with $\mathrm{SNR_L}$ denoting the Signal-to-Noise Ratio of the observation model in linear scale, the mutual information is connected with MMSE, as a function of $\mathrm{SNR_L}$, by the following equality.

$$\frac{\mathrm{d}}{\mathrm{d}\ \mathrm{SNR_L}} I\left(X^n; \sqrt{\mathrm{SNR_L}}\mathbf{H}X^n + Z^m\right) = \frac{1}{2}\mathrm{MMSE}\left(\mathrm{SNR_L}\right). \tag{3.14}$$

Injecting extra terms, i.e. launching attacks, that are independent with respect to the state variables is equivalent to decreasing the $\mathrm{SNR_L}$ of the observation model. When the mutual information is a concave and monotonically increasing function [1] of the $\mathrm{SNR_L}$, a decrease in the $\mathrm{SNR_L}$ leads to an increase in the derivative of the mutual information as a result of the concavity of mutual information, which leads to an increase in MMSE.

The conclusion that minimizing the mutual information between the state variables and the compromised measurements leads to an increase in MMSE also follows from Corollary A.2 in Appendix A, which states that

$$\mathbb{E}\left[\left(X^n - \hat{X}^n(Y_A^m)\right)^2\right] \geq \frac{1}{2\pi e} e^{2h(X^n|Y_A^m)} \tag{3.15}$$

$$= \frac{1}{2\pi e} e^{2\left(h(X^n) - I(X^n;Y_A^m)\right)}, \tag{3.16}$$

in which $\hat{X}^n(Y_A^m)$ is any estimator using the compromised measurements Y_A^m, $h(X^n|Y_A^m)$ is the conditional entropy of X^n given Y_A^m, and $h(X^n)$ is the entropy of X^n. Given the fact $h(X^n)$ is only determined by the power system, it is easy to see that minimizing the mutual information between the state variables and the compromised measurements leads to an increase in MMSE, which is the minimum value of the expression that is on the left-hand side of (3.15).

3.2.2 Detection Measure

Except maximizing the disruption represented by mutual information, the attacker also wants to minimize the probability of attack being detected. Note that the LRT

[1] It is clear that the mutual information is a monotonically increasing function of $\mathrm{SNR_L}$, the concavity of the function needs further proof. The idea is that when $\mathrm{SNR_L} = 0$, the mutual information is 0; when $\mathrm{SNR_L} \to \infty$, the mutual information saturates at the value of $\min\left(H\left(X^n\right), H\left(\sqrt{\mathrm{SNR_L}}\mathbf{H}X^n + Z^m\right)\right)$. The rate of the increase in mutual information decreases as SNR increases. So we think the mutual information is a concave function of SNR.

attack detection is given by

$$L(\mathbf{y}) = \frac{f_{Y_A^m}(\mathbf{y})}{f_{Y^m}(\mathbf{y})} \underset{\mathcal{H}_0}{\overset{\mathcal{H}_1}{\gtrless}} \tau, \tag{3.17}$$

where $Y_A^m \sim \mathcal{N}(\mathbf{0}, \mathbf{\Sigma}_{Y_A Y_A})$, $Y^m \sim \mathcal{N}(\mathbf{0}, \mathbf{\Sigma}_{YY})$, and $\mathbf{\Sigma}_{Y_A Y_A} = \mathbf{\Sigma}_{YY} + \mathbf{\Sigma}_{AA}$. As a result, the probability of detection under finite scenario is given by

$$\mathsf{P_D} \triangleq \mathbb{E}\left[\mathbb{1}_{\left\{L(Y_A^m) \geq \tau\right\}}\right], \tag{3.18}$$

where $\mathbb{1}_{\{.\}}$ is the indicator function. However the probability of detection is involved to characterize without closed-form expression for $\mathbf{\Sigma}_{AA}$, as the expected value is taking with respect to random variables Y_A^m. Instead of minimizing the probability of detection, the attacker can minimize the expected value of the likelihood ratio between $P_{Y_A^m}$ and $P_{Y_A^m}$ to minimize the probability of detection as in [44]. But the relation between the expected value of likelihood ratio and the probability of detection is not easy to characterize, as the distribution of the likelihood ratio is involved.

As a result, here we use the asymptotic value of the likelihood ratio to characterize the probability of detection. For the hypothesis testing problem given in (3.10), the probability of detection equals to one minus the probability of miss, or equals to one minus the probability of Type II error. For the LRT given in (3.11), the Chernoff-Stein Lemma ([22], or Lemma 2.2) states that for any probability of Type I error α smaller than one half, the logarithm of the averaged minimum value of probability of Type II error β asymptotically converges to the inverse of the KL divergence between the distributions of the two hypotheses. Therefore, for the attacker, minimizing the asymptotic detection probability is equivalent to maximizing the probability of Type II error, which is achieved by

$$\min_{A^m} D(P_{Y_A^m} || P_{Y^m}), \tag{3.19}$$

where $P_{Y_A^m}$ and P_{Y^m} denote the probability distributions of Y_A^m and Y^m, respectively. Minimizing the KL divergence ensures that the effect of the attacks on the induced distribution over the measurements is minimized, i.e. the attack is stealthy [119].

The KL divergence between two probability distributions is a measure of the statistical difference between the distributions. As such, it is a practical measure to quantify the deviation of the measurement statistics with respect to the statistics under normal operating conditions. For instance, in [96] it is used to test abnormal behaviors on the grid. For the hypothesis testing problem in (3.11), a small value

of the KL divergence between $P_{Y_A^m}$ and P_{Y^m} implies that on average the attack is unlikely to be detected by the LRT set by the attacker for a fixed value of τ.

3.3 Stealth Attack Construction

In this information-theoretic setting, the attacker aims to minimize the mutual information between the state variables and the compromised measurements, i.e. $I(X^n; Y_A^m)$, and the asymptotic probability of detection via $D(P_{Y_A^m}||P_{Y^m})$ as the asymptotic probability of detection is $\mathsf{P_D} \approx 1-\exp\left\{-D(P_{Y_A^m}||P_{Y^m})\right\}$, simultaneously. Following the approach in [119], the attacker constructs the utility function

$$I(X^n; Y_A^m) + D(P_{Y_A^m}||P_{Y^m}) \tag{3.20}$$

for the attack. The attacker minimizes this utility function to disrupt the estimation and bypass the detection set by the operator simultaneously.

Note that

$$\begin{aligned}
&I(X^n; Y_A^m) + D(P_{Y_A^m}||P_{Y^m}) && (3.21)\\
&= \int f_{X^n Y_A^m} \log \frac{f_{X^n Y_A^m}}{f_{X^n} f_{Y_A^m}} \, \mathrm{d}\mathbf{x}\mathrm{d}\mathbf{y}_a + \int f_{Y_A^m} \log \frac{f_{Y_A^m}}{f_{Y^m}} \, \mathrm{d}\mathbf{y}_a && (3.22)\\
&= \int f_{X^n Y_A^m} \log \frac{f_{X^n Y_A^m}}{f_{X^n} f_{Y_A^m}} \, \mathrm{d}\mathbf{x}\mathrm{d}\mathbf{y}_a + \int f_{X^n Y_A^m} \log \frac{f_{Y_A^m}}{f_{Y^m}} \, \mathrm{d}\mathbf{x}\mathrm{d}\mathbf{y}_a && (3.23)\\
&= \int f_{X^n Y_A^m} \log \frac{f_{X^n Y_A^m}}{f_{X^n} f_{Y^m}} \, \mathrm{d}\mathbf{x}\mathrm{d}\mathbf{y}_a && (3.24)\\
&= D\left(P_{X^n Y_A^m}||P_{X^n} P_{Y^m}\right), && (3.25)
\end{aligned}$$

where (3.22) follows from taking the definition of KL divergence in Definition A.10 and the definition of mutual information in Definition A.12 into (3.21), in which $f_{X^n Y_A^m}$ are the joint p.d.f. of (X^n, Y_A^m); (3.23) follows from extending the integration domain of the last term in (3.22) from all the realizations of Y_A^m to all the realizations of (X^n, Y_A^m) without changing the value of the integration; (3.24) follows from summing the two terms in (3.23); (3.25) follows from the fact that (3.24) coincides with the KL divergence between $P_{X^n Y_A^m}$ and $P_{X^n} P_{Y^m}$ with $P_{X^n Y_A^m}$ denoting the joint distribution of (X^n, Y_A^m), see Definition A.10.

In view of this, minimizing $I(X^n; Y_A^m) + D(P_{Y_A^m}||P_{Y^m})$ is posed as the following optimization problem:

$$\min_{A^m} \; D(P_{X^n Y_A^m}||P_{X^n} P_{Y^m}). \tag{3.26}$$

Note that in the Bayesian framework the state variables and the compromised measurements follow a joint multivariate Gaussian distribution given by

$$(X^n, Y_A^m) \sim \mathcal{N}(\mathbf{0}, \mathbf{\Sigma}), \tag{3.27}$$

where the block covariance matrix has the following structure:

$$\mathbf{\Sigma} = \begin{bmatrix} \mathbf{\Sigma}_{XX} & \mathbf{\Sigma}_{XX}\mathbf{H}^{\mathrm{T}} \\ \mathbf{H}\mathbf{\Sigma}_{XX} & \mathbf{H}\mathbf{\Sigma}_{XX}\mathbf{H}^{\mathrm{T}} + \sigma^2\mathbf{I}_m + \mathbf{\Sigma}_{AA} \end{bmatrix}. \tag{3.28}$$

Proposition 3.1. [22] *The mutual information between the vectors of random variables $X^n \sim \mathcal{N}(\mathbf{0}, \mathbf{\Sigma}_{XX})$ and $Y_A^m \sim \mathcal{N}(\mathbf{0}, \mathbf{\Sigma}_{Y_AY_A})$ is given by*

$$I(X^n; Y_A^m) = \frac{1}{2}\log\frac{|\mathbf{\Sigma}_{XX}||\mathbf{\Sigma}_{Y_AY_A}|}{|\mathbf{\Sigma}|}, \tag{3.29}$$

in which $|\cdot|$ is the determinant of the matrix given in $\cdot$.

Proof. Note that

$$I(X^n; Y_A^m) \tag{3.30}$$
$$= \mathbb{E}_{X^nY_A^m}\left[\log\frac{f_{X^nY_A^m}}{f_{X^n}f_{Y_A^m}}\right] \tag{3.31}$$
$$= \frac{1}{2}\mathbb{E}_W\Big[-\mathbf{w}^{\mathrm{T}}\mathbf{\Sigma}^{-1}\mathbf{w} - \log|\mathbf{\Sigma}| + \mathbf{x}^{\mathrm{T}}\mathbf{\Sigma}_{XX}^{-1}\mathbf{x} + \log|\mathbf{\Sigma}_{XX}| + \mathbf{y}_a^{\mathrm{T}}\mathbf{\Sigma}_{Y_AY_A}^{-1}\mathbf{y}_a + \log|\mathbf{\Sigma}_{Y_AY_A}|\Big] \tag{3.32}$$
$$= \frac{1}{2}\log\frac{|\mathbf{\Sigma}_{XX}||\mathbf{\Sigma}_{Y_AY_A}|}{|\mathbf{\Sigma}|} - \frac{1}{2}\mathbb{E}_W\left[-\mathrm{tr}(\mathbf{\Sigma}^{-1}\mathbf{w}\mathbf{w}^{\mathrm{T}}) + \mathrm{tr}(\mathbf{\Sigma}_{XX}^{-1}\mathbf{x}\mathbf{x}^{\mathrm{T}}) + \mathrm{tr}(\mathbf{\Sigma}_{Y_AY_A}^{-1}\mathbf{y}_a\mathbf{y}_a^{\mathrm{T}})\right] \tag{3.33}$$
$$= \frac{1}{2}\log\frac{|\mathbf{\Sigma}_{XX}||\mathbf{\Sigma}_{Y_AY_A}|}{|\mathbf{\Sigma}|}, \tag{3.34}$$

where (3.31) follows from taking the definition of mutual information in Definition A.12 into (3.30); (3.32) follows from combining the Gaussianity of X^n, Y_A^m, and (X^n, Y_A^m) with (3.31), in which $W = [X^n; Y^m]^{\mathrm{T}}$ is distributed as $\mathcal{N}(\mathbf{0}, \mathbf{\Sigma})$, and $\mathbf{w}$ is a realization of random variable W; (3.33) is obtained by taking the constant terms outside the expectation; (3.34) follows from the fact that

$$\mathbb{E}_W\left[\mathrm{tr}(\mathbf{\Sigma}^{-1}\mathbf{w}\mathbf{w}^{\mathrm{T}})\right] = \mathrm{tr}\left(\mathbf{\Sigma}^{-1}\mathbb{E}_W\left[\mathbf{w}\mathbf{w}^{\mathrm{T}}\right]\right) = \mathrm{tr}\left(\mathbf{\Sigma}^{-1}\mathbf{\Sigma}\right) = m + n \tag{3.35}$$

and applying the same operation in (3.35) to the terms $\mathrm{tr}(\mathbf{\Sigma}_{XX}^{-1}\mathbf{x}\mathbf{x}^{\mathrm{T}})$ and $\mathrm{tr}(\mathbf{\Sigma}_{Y_AY_A}^{-1}\mathbf{y}_a\mathbf{y}_a^{\mathrm{T}})$. □

Proposition 3.2. [22] *The KL divergence between two m-dimensional multivariate Gaussian distributions* $P_0 = \mathcal{N}(\mathbf{0}, \mathbf{\Sigma}_0)$ *and* $P_1 = \mathcal{N}(\mathbf{0}, \mathbf{\Sigma}_1)$ *is given by*

$$D(P_0 \| P_1) = \frac{1}{2}\left(\log \frac{|\mathbf{\Sigma}_1|}{|\mathbf{\Sigma}_0|} - m + \mathrm{tr}(\mathbf{\Sigma}_1^{-1}\mathbf{\Sigma}_0)\right). \tag{3.36}$$

Proof. Note that

$$\begin{aligned}
D(P_0 \| P_1) &= \mathbb{E}_V\left[\log f_{\mathcal{N}(\mathbf{0},\mathbf{\Sigma}_0)}(\mathbf{v}) - \log f_{\mathcal{N}(\mathbf{0},\mathbf{\Sigma}_1)}(\mathbf{v})\right] && (3.37)\\
&= \frac{1}{2}\mathbb{E}_V\left[-\log|\mathbf{\Sigma}_0| - \mathbf{v}^{\mathrm{T}}\mathbf{\Sigma}_0^{-1}\mathbf{v} + \log|\mathbf{\Sigma}_1| + \mathbf{v}^{\mathrm{T}}\mathbf{\Sigma}_1^{-1}\mathbf{v}|\right] && (3.38)\\
&= \frac{1}{2}\log\frac{|\mathbf{\Sigma}_1|}{|\mathbf{\Sigma}_0|} + \frac{1}{2}\mathbb{E}_V\left[\mathbf{v}^{\mathrm{T}}\mathbf{\Sigma}_1^{-1}\mathbf{v} - \mathbf{v}^{\mathrm{T}}\mathbf{\Sigma}_0^{-1}\mathbf{v}|\right] && (3.39)\\
&= \frac{1}{2}\log\frac{|\mathbf{\Sigma}_1|}{|\mathbf{\Sigma}_0|} + \frac{1}{2}\mathbb{E}_V\left[\mathrm{tr}(\mathbf{\Sigma}_1^{-1}\mathbf{v}\mathbf{v}^{\mathrm{T}}) - \mathrm{tr}(\mathbf{\Sigma}_0^{-1}\mathbf{v}\mathbf{v}^{\mathrm{T}})\right] && (3.40)\\
&= \frac{1}{2}\left(\log\frac{|\mathbf{\Sigma}_1|}{|\mathbf{\Sigma}_0|} - m + \mathrm{tr}(\mathbf{\Sigma}_1^{-1}\mathbf{\Sigma}_0)\right), && (3.41)
\end{aligned}$$

where (3.37) follows from the definition of KL divergence in Definition A.10, in which V is a multivariate Gaussian random variable follows distribution P_0 and $\mathbf{v}$ is the realization of V; (3.38) follows from taking the Gaussian p.d.f. of V into (3.37); (3.39) is obtained by taking the constant terms outside the expectation; (3.40) follows from the fact that

$$\mathbf{v}^{\mathrm{T}}\mathbf{\Sigma}_1^{-1}\mathbf{v} = \mathrm{tr}\left(\mathbf{v}^{\mathrm{T}}\mathbf{\Sigma}_1^{-1}\mathbf{v}\right) = \mathrm{tr}\left(\mathbf{\Sigma}_1^{-1}\mathbf{v}\mathbf{v}^{\mathrm{T}}\right) \tag{3.42}$$

and apply the operation in (3.42) also to the term $\mathbf{v}^{\mathrm{T}}\mathbf{\Sigma}_0^{-1}\mathbf{v}$; (3.41) is obtained via the same approach in (3.35).

□

Combining (3.36) and (3.26) yields

$$\begin{aligned}
D(P_{X^nY_A^m} || P_{X^n}P_{Y^m}) &= \frac{1}{2}\left(\log\frac{|\tilde{\mathbf{\Sigma}}|}{|\mathbf{\Sigma}|} - (m+n) + \mathrm{tr}((\tilde{\mathbf{\Sigma}})^{-1}\mathbf{\Sigma})\right)\\
&= \frac{1}{2}\left(-\log|(\tilde{\mathbf{\Sigma}})^{-1}\mathbf{\Sigma}| - (m+n) + \mathrm{tr}((\tilde{\mathbf{\Sigma}})^{-1}\mathbf{\Sigma})\right), \qquad (3.43)
\end{aligned}$$

where $\tilde{\mathbf{\Sigma}}$ is the covariance matrix of $P_{X^n}P_{Y^m}$ and is given by

$$\tilde{\mathbf{\Sigma}} = \begin{bmatrix} \mathbf{\Sigma}_{XX} & \mathbf{0} \\ \mathbf{0} & \mathbf{\Sigma}_{YY} \end{bmatrix}.$$

Note that

$$(\tilde{\boldsymbol{\Sigma}})^{-1}\boldsymbol{\Sigma}=\begin{bmatrix} \boldsymbol{\Sigma}_{XX}^{-1} & \mathbf{0} \\ \mathbf{0} & \boldsymbol{\Sigma}_{YY}^{-1} \end{bmatrix}\begin{bmatrix} \boldsymbol{\Sigma}_{XX} & \boldsymbol{\Sigma}_{XX}\mathbf{H}^{\mathrm{T}} \\ \mathbf{H}\boldsymbol{\Sigma}_{XX} & \boldsymbol{\Sigma}_{Y_AY_A} \end{bmatrix}=\begin{bmatrix} \mathbf{I}_n & \mathbf{H}^{\mathrm{T}} \\ \boldsymbol{\Sigma}_{YY}^{-1}\mathbf{H}\boldsymbol{\Sigma}_{XX} & \boldsymbol{\Sigma}_{YY}^{-1}\boldsymbol{\Sigma}_{Y_AY_A} \end{bmatrix} \tag{3.44}$$

and

$$|(\tilde{\boldsymbol{\Sigma}})^{-1}\boldsymbol{\Sigma}| = |\mathbf{I}_n||\mathbf{I}_m + \boldsymbol{\Sigma}_{YY}^{-1}\boldsymbol{\Sigma}_{AA} + \sigma^2\boldsymbol{\Sigma}_{YY}^{-1}|, \tag{3.45}$$

an equivalent expression for the optimization problem in (3.26) is obtained in the following lemma.

Lemma 3.1. *The optimization problem in (3.26) is equivalent to*

$$\underset{\boldsymbol{\Sigma}_{AA}\in\mathcal{S}_+^m}{min} \quad \left[\mathrm{tr}(\boldsymbol{\Sigma}_{YY}^{-1}\boldsymbol{\Sigma}_{AA}) - \log|\boldsymbol{\Sigma}_{AA} + \sigma^2\mathbf{I}_m|\right]. \tag{3.46}$$

Proof. Taking (3.44) and (3.45) into (3.43) and neglecting the constant term $\log|\boldsymbol{\Sigma}_{YY}^{-1}|$ yields the result. □

In the following we show that the optimization problem in Lemma 3.1 is a convex optimization problem.

Proposition 3.3. *The optimization problem given by (3.46) is equivalent to minimizing a convex function within a convex set.*

Proof. The trace operator is a linear operator, and $-\log|\boldsymbol{\Sigma}_{AA} + \sigma^2\mathbf{I}_m|$ is a convex function of the positive semi-definite matrix $\boldsymbol{\Sigma}_{AA}$ [120, pp. 74]. Therefore, the objective function in (3.46) is a convex function of $\boldsymbol{\Sigma}_{AA}$.

Since $\mathcal{S}_+^m$ forms a convex set, the result follows immediately. □

Before introducing the closed-form expression for the stealth attacks, the first order condition for convex functions is proposed to aid the proof.

Proposition 3.4. *[120, pp. 69] Suppose f is differentiable (i.e., its gradient ∇f exists at each point in **dom** f, which is open). Then f is convex if and only if **dom** f is convex and*

$$f(y) \geq f(x) + \nabla f(x)^T(y - x) \tag{3.47}$$

*holds for all $x, y \in$ **dom** f, where **dom** f represents the domain of function f.*

Using Proposition 3.4, it is easy to show that when a convex function only has one critical point x_0, i.e. $\nabla f(x_0) = 0$, then x_0 is the global minimum of the convex function.

The following theorem provides the closed-form expression for the stealth attacks.

Theorem 3.1. *The solution to the attack construction optimization problem (3.46) is the covariance matrix* $\mathbf{\Sigma}_{AA}^{\star} = \mathbf{H}\mathbf{\Sigma}_{XX}\mathbf{H}^{\mathrm{T}}$.

Proof. Taking the derivative of $\mathrm{tr}(\mathbf{\Sigma}_{YY}^{-1}\mathbf{\Sigma}_{AA})$ with respect to $\mathbf{\Sigma}_{AA}$ yields [121, Sec. 17.5.2]

$$\frac{\partial}{\partial \mathbf{\Sigma}_{AA}}\Big(\mathrm{tr}(\mathbf{\Sigma}_{YY}^{-1}\mathbf{\Sigma}_{AA})\Big) = 2\mathbf{\Sigma}_{YY}^{-1} - \mathrm{diag}(\mathbf{\Sigma}_{YY}^{-1}).$$

Similarly we have

$$\begin{aligned}\frac{\partial}{\partial \mathbf{\Sigma}_{AA}}\Big(\log|\mathbf{\Sigma}_{AA} + \sigma^2\mathbf{I}_m|\Big) &= |\mathbf{\Sigma}_{AA} + \sigma^2\mathbf{I}_m|^{-1}\frac{\partial|\mathbf{\Sigma}_{AA} + \sigma^2\mathbf{I}_m|}{\partial \mathbf{\Sigma}_{AA}} &(3.48)\\ &= 2(\mathbf{\Sigma}_{AA} + \sigma^2\mathbf{I}_m)^{-1} - \mathrm{diag}(\mathbf{\Sigma}_{AA} + \sigma^2\mathbf{I}_m)^{-1}. &(3.49)\end{aligned}$$

from [121, Sec. 17.5.3]. So taking the derivative of the objective function in (3.46) with respect to $\mathbf{\Sigma}_{AA}$ yields

$$\begin{aligned}&\frac{\partial}{\partial \mathbf{\Sigma}_{AA}}\Big(\mathrm{tr}(\mathbf{\Sigma}_{YY}^{-1}\mathbf{\Sigma}_{AA}) - \log|\mathbf{\Sigma}_{AA} + \sigma^2\mathbf{I}_m|\Big)\\ &= 2\mathbf{\Sigma}_{YY}^{-1} - \mathrm{diag}(\mathbf{\Sigma}_{YY}^{-1}) - 2(\mathbf{\Sigma}_{AA} + \sigma^2\mathbf{I}_m)^{-1} + \mathrm{diag}(\mathbf{\Sigma}_{AA} + \sigma^2\mathbf{I}_m)^{-1} &(3.50)\\ &= 2\left(\mathbf{\Sigma}_{YY}^{-1} - (\mathbf{\Sigma}_{AA} + \sigma^2\mathbf{I}_m)^{-1}\right) - \left(\mathrm{diag}(\mathbf{\Sigma}_{YY}^{-1}) - \mathrm{diag}(\mathbf{\Sigma}_{AA} + \sigma^2\mathbf{I}_m)^{-1}\right). &(3.51)\end{aligned}$$

Notice that the only critical point is $\mathbf{\Sigma}_{YY} = \mathbf{\Sigma}_{AA}^{\star} + \sigma^2\mathbf{I}_m$, i.e. $\mathbf{\Sigma}_{AA}^{\star} = \mathbf{H}\mathbf{\Sigma}_{XX}\mathbf{H}^{\mathrm{T}}$. The result follows immediately from combining this result with Proposition 3.3 and Proposition 3.4. □

Interestingly, the optimal attack construction depends only on the second order moments of the state variables, i.e. $\mathbf{\Sigma}_{XX}$, and the Jacobian matrix $\mathbf{H}$. The variances of the noise terms are not required to construct the attacks. This implies that the quality of the measurements from the sensing infrastructure has a limited impact on the attack construction. The Jacobian matrix $\mathbf{H}$ is determined by the topology of the network and the admittance of the branches, which are of minor difference under different operation conditions.

In a practical setting, the covariance matrix of the state variables is usually estimated through the historical data of the state variables. Therefore, historical data of the state variables is central to the proposed attack construction. From a practical point of view, making historical data and the topology of the grid available to the public poses a security threat to the operator. However, the extent to which historical data aids the attack construction remains to be determined. In fact, due to practical and operational constraints, it is safe to assume that the attacker gets access to only partial information about the second order statistics of the state variables.

In Chapter 5, the attacker is assumed to only get access to a limited number of samples of the state variables. Other than the exact covariance matrix $\mathbf{\Sigma}_{XX}$, the sample covariance matrix is employed to construct the attack. Also the performance of the attack using the sample covariance matrix is analyzed in Chapter 5 using RMT tools.

The following proposition characterizes the mutual information loss and the KL divergence for the stealth attacks in Theorem 3.1.

Proposition 3.5. *The mutual information loss induced by the stealth attacks in Theorem 3.1 is given by*

$$I(X^n;Y^m) - I(X^n;Y_A^m) = \frac{1}{2}\left(\log|\mathbf{\Sigma}_{YY}| - \log|\sigma^2\mathbf{I}_m| - \log|2\mathbf{I}_m + \sigma^2\mathbf{\Sigma}_{YY}^{-1}|\right). \quad (3.52)$$

Proof. Note that

$$\begin{aligned} I(X^n;Y^m) - I(X^n;Y_A^m) &= \frac{1}{2}\left(\log\frac{|\mathbf{\Sigma}_{YY}||\mathbf{\Sigma}_{XX}|}{|\mathbf{\Sigma}_N|} - \log\frac{|\mathbf{\Sigma}_{Y_AY_A}||\mathbf{\Sigma}_{XX}|}{|\mathbf{\Sigma}|}\right) \\ &= \frac{1}{2}\left(\log\frac{|\mathbf{\Sigma}_{YY}||\mathbf{\Sigma}_{XX}|}{|\mathbf{\Sigma}_N|} - \log\frac{|\mathbf{\Sigma}_{Y_AY_A}||\mathbf{\Sigma}_{XX}|}{|\mathbf{\Sigma}|}\right) \\ &= \frac{1}{2}\left(\log\frac{|\mathbf{\Sigma}_{YY}||\mathbf{\Sigma}_{Y_AY_A} - \mathbf{H}\mathbf{\Sigma}_{XX}\mathbf{H}^{\mathrm{T}}|}{|\mathbf{\Sigma}_{Y_AY_A}|} - \log|\sigma^2\mathbf{I}_m|\right) \\ &= \frac{1}{2}\left(\log|\mathbf{\Sigma}_{YY}| - \log|\sigma^2\mathbf{I}_m| - \log|2\mathbf{I}_m + \sigma^2\mathbf{\Sigma}_{YY}^{-1}|\right), \end{aligned}$$

where $\mathbf{\Sigma}_N$ is the covariance of the joint Gaussian distribution of (X^n, Y^m), i.e. the distribution of (X^n, Y^m) under normal condition, which is given by

$$\mathbf{\Sigma}_N = \begin{bmatrix} \mathbf{\Sigma}_{XX} & \mathbf{\Sigma}_{XX}\mathbf{H}^{\mathrm{T}} \\ \mathbf{H}\mathbf{\Sigma}_{XX} & \mathbf{H}\mathbf{\Sigma}_{XX}\mathbf{H}^{\mathrm{T}} + \sigma^2\mathbf{I}_m \end{bmatrix}. \quad (3.53)$$

□

Proposition 3.6. *The KL divergence induced by the stealth attack in Theorem 3.1 is given by*

$$D(P_{Y_A^m}||P_{Y^m}) = \frac{1}{2}\left(m - \sigma^2 tr(\mathbf{\Sigma}_{YY}^{-1}) - \log|2\mathbf{I}_m - \sigma^2\mathbf{\Sigma}_{YY}^{-1}|\right). \quad (3.54)$$

Proof. Note that

$$\begin{aligned} D(P_{Y_A^m}||P_{Y^m}) &= \frac{1}{2}\left(\log\frac{|\mathbf{\Sigma}_{YY}|}{|\mathbf{\Sigma}_{Y_AY_A}|} - m + \mathrm{tr}(\mathbf{\Sigma}_{YY}^{-1}\mathbf{\Sigma}_{Y_AY_A})\right) \\ &= \frac{1}{2}\Big(\mathrm{tr}(2\mathbf{I}_m - \sigma^2\mathbf{\Sigma}_{YY}^{-1}) - \log|2\mathbf{I}_m - \sigma^2\mathbf{\Sigma}_{YY}^{-1}| - m\Big) \\ &= \frac{1}{2}\Big(m - \sigma^2\mathrm{tr}(\mathbf{\Sigma}_{YY}^{-1}) - \log|2\mathbf{I}_m - \sigma^2\mathbf{\Sigma}_{YY}^{-1}|\Big). \end{aligned}$$

□

3.4 Numerical Simulation

The IEEE 30-Bus test system and IEEE 118-Bus test system are used to simulate the DC state estimation setting in which the bus voltage magnitudes are set to 1.0 per unit, c.f. (2.23) and (2.24). Here the bus voltage angles are chosen to be the state variables, and the power injections and the power flows in both directions are used as the measurements. The Jacobian matrix $\mathbf{H}$ is determined by the branch reactances of the grid and it is computed using MATPOWER [122].

The optimal attack construction in Theorem 3.1 shows that the covariance matrix of the attack is a function of the covariance matrix of the state variables. The result in Theorem 3.1 holds for any positive semi-definite covariance matrix. Since covariance matrices of weakly stationary random processes are Toeplitz, here to simplify the simulation, a specific Toeplitz matrix with exponential decay parameter ρ is adopted [86]. The Toeplitz matrix of dimension $n \times n$ with exponential decay parameter ρ is given by $\mathbf{\Sigma}_{XX} = [s_{ij} = \rho^{|i-j|}; i, j = 1, 2, \ldots, n]$, i.e.

$$\mathbf{\Sigma}_{XX} = \begin{pmatrix} 1 & \rho & \rho^2 & \ldots & \rho^{n-2} & \rho^{n-1} \\ \rho & 1 & \rho & \ldots & \rho^{n-3} & \rho^{n-2} \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ \rho^{n-2} & \rho^{n-3} & \rho^{n-4} & \ldots & 1 & \rho \\ \rho^{n-1} & \rho^{n-2} & \rho^{n-3} & \ldots & \rho & 1 \end{pmatrix}. \tag{3.55}$$

The parameter ρ reflects the correlation strength between the state variables, which is the correlation strength between the buses connected by branches in the power system. Under this setting, the utility function of the optimal attack is a function of the correlation strength ρ and the noise variance σ^2. We define the Signal-to-Noise Ratio (SNR) to be

$$\mathrm{SNR} = 10\log_{10}\left(\frac{\mathrm{tr}(\mathbf{H}\mathbf{\Sigma}_{XX}\mathbf{H}^{\mathrm{T}})}{m\sigma^2}\right). \tag{3.56}$$

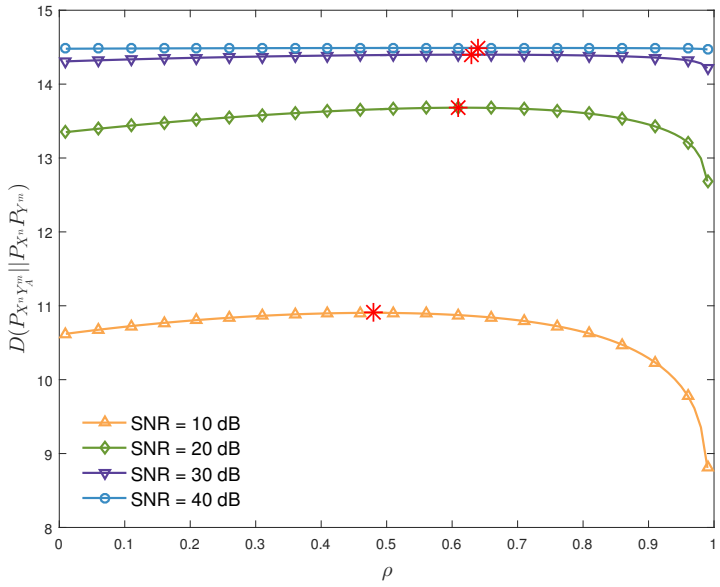


Fig. 3.1. Performance of the stealth attack in terms of the utility function in (3.26) for different values of ρ and SNR on IEEE 30-Bus test system.

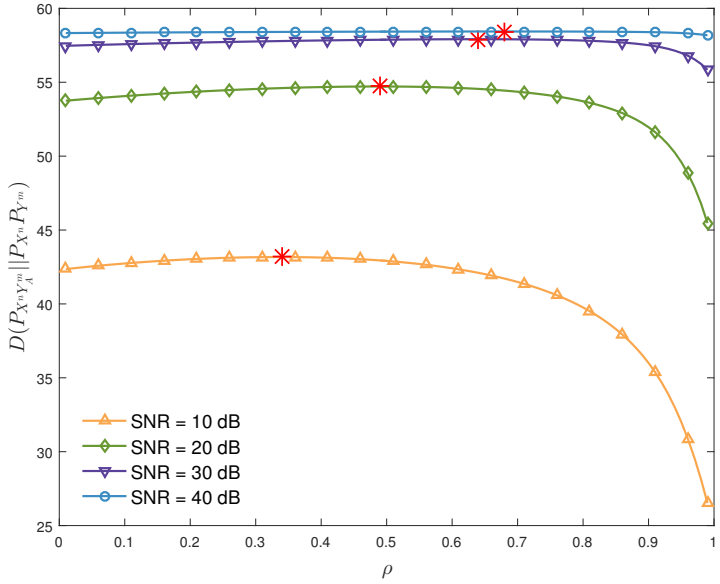


Fig. 3.2. Performance of the stealth attack in terms of the utility function in (3.26) for different values of ρ and SNR on IEEE 118-Bus test system.

As a result, the utility function is a function of the correlation strength ρ and the SNR at which the grid operates.

3.4.1 Performance of Stealth Attacks

The performance of the optimal attack as measured by the utility function given by (3.26) is shown in Fig. 3.1 and Fig. 3.2 for IEEE 30-Bus test system and IEEE 118-Bus test system, respectively, in which the maximum value of the utility function,

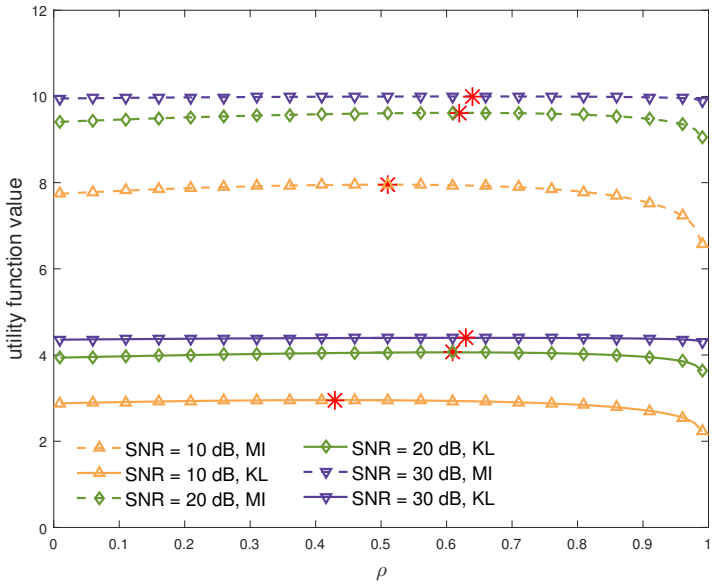


Fig. 3.3. Performance of the stealth attack in terms of mutual information (MI) and KL divergence for different values of ρ and SNR on IEEE 30-Bus test system.

i.e. the worst performance of the attack vector, is represented by a star. Surprisingly, the performance of the attack is non-monotonic with the correlation strength ρ. The simulations show that higher values of SNR yield worse performance for the attacker. Moreover, the performance of the attack is insensitive to the correlation strength, ρ, for a wide range of correlation values and only becomes significant when the correlation strength is large. For low and medium range values of the SNR, the performance of the attack is governed by the SNR and the correlation strength does not play a significant role. In the high SNR regime, the performance of the attack does not change significantly with the value of the correlation strength. This observation contrasts with linearly encoded Gaussian communication systems in which the impact of correlation is significant even for the cases in which the correlation strength is low [123]. Furthermore, the performance gain that benefits from the high correlation strength is more obvious in the power system of large scale.

The tradeoff between the disruption and the probability of attack detection is shown in Fig. 3.3 and Fig. 3.4 for IEEE 30-Bus test system and 118-Bus test system, respectively. The performance of the attack is analyzed in terms of the mutual information, $I(X^n; Y_A^m)$, and the KL divergence, $D(P_{Y_A^m}||P_{Y^m})$, that the attack induces. Interestingly, the performance of both objectives of the utility function is similar and there is no significant difference in the effect of the SNR or the correlation strength. This suggests that the tradeoff between disruption and detection achieved by the optimal attack construction does not change significantly with different system parameters. It is only when the value of the correlation strength is high that the performance gain obtained in terms of mutual information grows faster than the

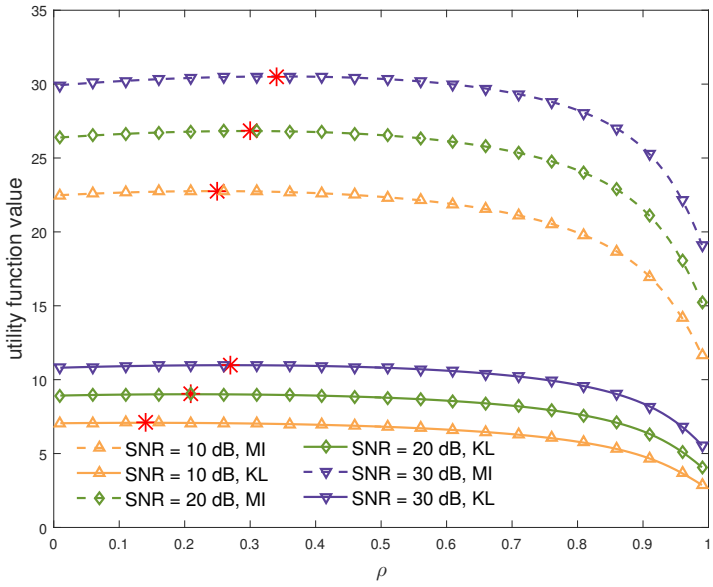


Fig. 3.4. Performance of the stealth attack in terms of mutual information (MI) and KL divergence for different values of ρ and SNR on IEEE 118-Bus test system.

performance gain obtained from the KL divergence improvement. This is more significant in larger power system. From a practical point of view, this suggests that the attacker expects to inflict a similar disruption on the grid for a given probability of detection regardless of the system parameters ρ and SNR. Furthermore, the performance of the attack is more sensitive to the correlation strength in larger power system, which is the same as the results from Fig. 3.1 and Fig. 3.2.

3.4.2 MMSE Degradation and Probability of Detection Induced by Stealth Attacks

As proposed in Section 3.2, the objective of the stealth attacks is two-fold. On one hand, the attacker minimizes the mutual information between the state variables and the compromised measurements to minimize the amount of information that the operator obtained from the measurements about the state variables. As stated in (3.14), minimizing the mutual information leads to an increase in the MMSE of the estimation. The MMSE degradation induced by the stealth attacks, i.e. $\mathbf{Ma}$ with $\mathbf{M}$ given in (2.39), is shown in Fig. 3.5 for IEEE 14-Bus test system and in Fig. 3.6 for IEEE 30-Bus test system, in which $10,000$ and $20,000$ realizations are generated for 14-Bus system and 30-Bus system, respectively. It is shown that the stealth attacks have better performance on MMSE degradation when the correlation strength ρ is of high value and have quite steady performance when ρ is of low and medium value, which coincides with the performance of the attacks on the mutual information objective. Also when SNR is of high value, the stealth attacks have

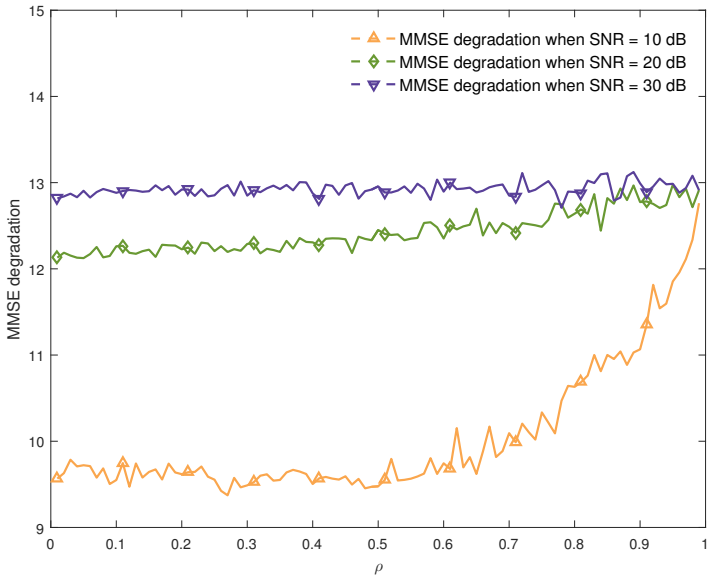


Fig. 3.5. MMSE degradation induced by stealth attack for different values of ρ and SNR on IEEE 14-Bus test system.

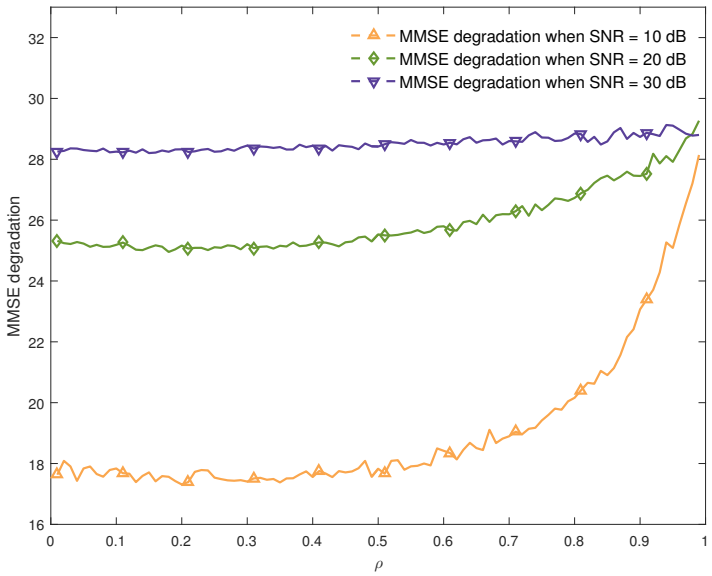


Fig. 3.6. MMSE degradation induced by stealth attack for different values of ρ and SNR on IEEE 30-Bus test system.

better performance on MMSE degradation, but it is also shown in Fig. 3.3 that the probability of detection of the attacks is also high when SNR is of high value. So the attacker needs to tradeoff carefully between the mutual information objective and the probability of detection objective.

On the other hand, the attacker minimizes the asymptotic probability of detection by minimizing the KL divergence given in (3.19). For the finite case, the probability of detection is given by (3.18), which is the probability that the attacks trigger the LRT given in (3.11). The finite probability of detection in (3.18) and the asymptotic

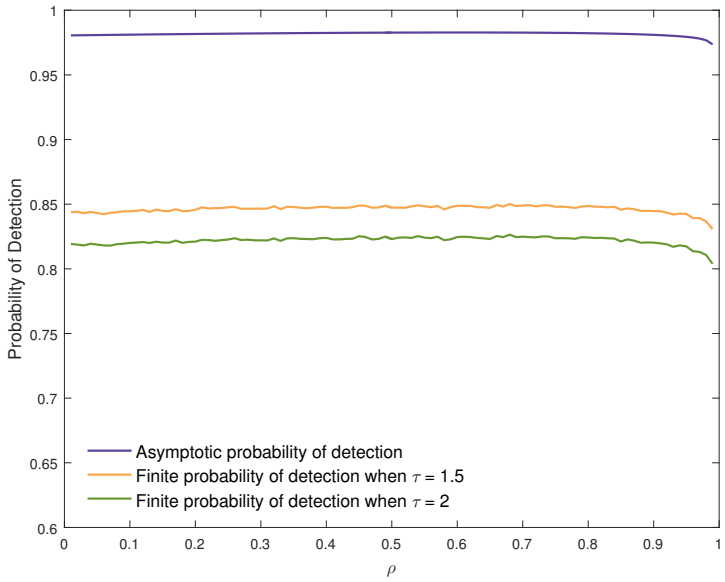


Fig. 3.7. Probability of detection of stealth attack for different values of ρ, $\tau = 1.5$ and $\tau = 2$ on IEEE 30-Bus test system when SNR = 20 dB.

probability given by $\mathsf{P_D} \approx 1 - \exp\left\{-D(P_{Y_A^m}||P_{Y^m})\right\}$ are depicted in Fig. 3.7 for IEEE 30-Bus test system, in which 20,000 realizations are generated for the finite probability of detection. The detection threshold τ in LRT given by (3.11) is chosen to be 1.5 and 2, and the SNR is set to be 20 dB. Except that the asymptotic probability is higher than the finite probability of detection, the finite probability of detection is similar to the asymptotic probability of detection, i.e. quiet steady for for wide range of ρ and decrease a little bit when ρ is of large value. But both the asymptotic probability of detection and the finite probability of detection of the stealth attacks are high. This implies that when the operator chooses to put less importance to the probability of false alarm and focuses on the probability of detection, i.e. chooses a small value for τ, the stealth attack is easy to be detected.

In next chapter, we introduce the generalized stealth attacks, which allow the attacker set preference between the mutual information objective and the probability of detection objective. As a result, the attacks lead to a lower probability of detection for both the asymptotic case and the finite case when the attacker prioritizes the probability of detection objective.

3.5 Summary

In this chapter, the stealth attacks construction within the Bayesian framework is proposed using the information-theoretic measures. Specifically the attacker minimizes the amount of information that the operator obtained about the state variables from the measurements, which is achieved by minimizing the mutual

information between the state variables and the compromised measurements. On the other hand, the attacker minimizes the asymptotic probability of detection under LRT via minimizing the KL divergence between the distribution with attack and without attack. The stealth attacks achieve these two contradictive objectives simultaneously by summing them up, which is a convex optimization for the Gaussian attacks case. Closed-form expression is obtained for the Gaussian stealth attacks.

Chapter 4

Generalized Information-Theoretic Stealth Attacks

In this chapter, the stealth attacks in last chapter are generalized by introducing a weighting parameter to the KL divergence objective that represents the probability of detection. The weighting parameter allows the attacker to adopt different tradeoff strategies between the mutual information and the probability of detection. Closed-form expressions for the generalized stealth attacks and the resulting probability of detection are proposed for the case that the attacker emphasizes the probability of detection objective, i.e. the weighting parameter is larger than one. To provide explicit insight into the relation between the probability of detection and the weighting parameter, a concentration inequality upper bound is proposed for the probability of detection, which provides a guideline to the attacker for choosing the weighting parameter.

4.1 Generalized Stealth Attacks

The aim of the stealth attacks is to minimize the mutual information between the state variables and the compromised measurements, i.e. minimize $I(X^n; Y_A^m)$, and to minimize the asymptotic probability of detection by minimizing the KL divergence between the distribution of the compromised measurements and the distribution of the uncompromised measurements, i.e. minimizing $D(P_{Y_A^m} \| P_{Y^m})$. The attacker combines these two objectives by

$$I(X^n; Y_A^m) + D(P_{Y_A^m} \| P_{Y^m}) = D(P_{X^n Y_A^m} \| P_{X^n} P_{Y^m}). \tag{4.1}$$

The work in Chapter 4 is published in "K. Sun, I. Esnaola, S.M. Perlaza, and H.V. Poor, "Stealth attacks on the smart grid," *IEEE Trans. Smart Grid* (Early Access), 2019.".

The resulting optimization problem to construct the stealth attacks is given by

$$\min_{A^m} D(P_{X^n Y_A^m} \| P_{X^n} P_{Y^m}). \tag{4.2}$$

Therein, it is shown that this is a convex optimization problem and the covariance matrix of the optimal Gaussian attack is $\mathbf{\Sigma}_{AA}^{\star} = \mathbf{H}\mathbf{\Sigma}_{XX}\mathbf{H}^{\mathrm{T}}$. The simulation on IEEE test system shows that the stealth attack has a good performance on the mutual information, but it also has a high probability of detection.

To address the issue of high probability of detection, a parameter that weights the detection term in the cost function on the left-hand side of (4.1) is introduced to allow the attacker tuning the probability of detection. The resulting optimization problem is given by

$$\min_{A^m} I(X^n; Y_A^m) + \lambda D(P_{Y_A^m} \| P_{Y^m}), \tag{4.3}$$

where $\lambda \geq 1$ governs the weight given to each objective in the cost function. It is interesting to note that for the case in which $\lambda = 1$ the proposed cost function boils down to the effective secrecy proposed in [119] and the attack construction in (4.3) coincides with that in Theorem 3.1. For $\lambda > 1$, the attacker adopts a conservative approach and prioritizes remaining undetected over minimizing the amount of information acquired by the operator. By increasing the value of λ, the attacker decreases the probability of detection at the expense of increasing the amount of information acquired by the operator via the measurements.

The following lemma proposes an equivalent expression for the optimization problem in (4.3) for the Gaussian state variables and Gaussian attacks case.

Lemma 4.1. *The optimization problem in (4.3) is equivalent to the optimization problem given by*

$$\min_{\mathbf{\Sigma}_{AA} \in \mathcal{S}_+^m} -(\lambda - 1) \log |\mathbf{\Sigma}_{YY} + \mathbf{\Sigma}_{AA}| - \log |\mathbf{\Sigma}_{AA} + \sigma^2 \mathbf{I}_m| + \lambda tr(\mathbf{\Sigma}_{YY}^{-1} \mathbf{\Sigma}_{AA}). \tag{4.4}$$

Proof. Combing the objective function in (4.3) with Proposition 3.1 and Proposition 3.2 yields

$$\begin{aligned} & I(X^n; Y_A^m) + \lambda D(P_{Y_A^m} \| P_{Y^m}) \\ &= \frac{1}{2} \log \frac{|\mathbf{\Sigma}_{XX}||\mathbf{\Sigma}_{Y_A Y_A}|}{|\mathbf{\Sigma}|} + \frac{\lambda}{2} \left(\log \frac{|\mathbf{\Sigma}_{YY}|}{|\mathbf{\Sigma}_{Y_A Y_A}|} - m + \mathrm{tr}(\mathbf{\Sigma}_{YY}^{-1} \mathbf{\Sigma}_{Y_A Y_A}) \right) \qquad (4.5) \\ &= \frac{1 - \lambda}{2} \log |\mathbf{\Sigma}_{YY} + \mathbf{\Sigma}_{AA}| - \frac{1}{2} \log |\mathbf{\Sigma}_{AA} + \sigma^2 \mathbf{I}_m| + \frac{\lambda}{2} \mathrm{tr}(\mathbf{\Sigma}_{YY}^{-1} \mathbf{\Sigma}_{AA}) + c, \qquad (4.6) \end{aligned}$$

where (4.6) follows from expanding the logarithm terms in (4.5) and combining the similar terms, in which c is a constant that is only determined by system parameters, i.e. it is not a function of $\mathbf{\Sigma}_{AA}$. □

We now proceed to solve the optimization problem above. First, note that the optimization domain $\mathcal{S}_+^m$ is a convex set. The following proposition characterizes the convexity of the cost function.

Proposition 4.1. *Let $\lambda \geq 1$. Then the cost function in the optimization problem in (4.4) is convex.*

Proof. Note that the term $-\log|\mathbf{\Sigma}_{AA}+\sigma^2\mathbf{I}_m|$ is a convex function on $\mathbf{\Sigma}_{AA} \in \mathcal{S}_+^m$ [120, pp. 74]. Additionally, $-(\lambda-1)\log|\mathbf{\Sigma}_{YY}+\mathbf{\Sigma}_{AA}|$ is a convex function on $\mathbf{\Sigma}_{AA} \in \mathcal{S}_+^m$ when $\lambda \geq 1$. Since the trace operator is a linear operator and the sum of convex functions is still a convex function, it follows that the cost function in (4.4) is convex on $\mathbf{\Sigma}_{AA} \in \mathcal{S}_+^m$. □

As a result, the optimization problem in (4.4) is an optimization problem that minimizes a convex function within a convex set. The following theorem provides a closed-form expression for the solution of the optimization problem given in (4.4).

Theorem 4.1. *Let $\lambda \geq 1$. Then the solution to the optimization problem in (4.4) is*

$$\mathbf{\Sigma}_{AA}^{\star} = \frac{1}{\lambda}\mathbf{H}\mathbf{\Sigma}_{XX}\mathbf{H}^{\mathrm{T}}. \tag{4.7}$$

Proof. Denote the cost function in (4.4) by $f(\mathbf{\Sigma}_{AA})$. Following the approach utilized in the proof of Theorem 3.1 and taking the derivative of the cost function in (4.4) with respect to $\mathbf{\Sigma}_{AA}$ yield

$$\begin{aligned}\frac{\partial f(\mathbf{\Sigma}_{AA})}{\partial \mathbf{\Sigma}_{AA}} =& -2(\lambda-1)(\mathbf{\Sigma}_{YY}+\mathbf{\Sigma}_{AA})^{-1} - 2(\mathbf{\Sigma}_{AA}+\sigma^2\mathbf{I}_m)^{-1} \\ &+2\lambda\mathbf{\Sigma}_{YY}^{-1} + (\lambda-1)\mathrm{diag}((\mathbf{\Sigma}_{YY}+\mathbf{\Sigma}_{AA})^{-1}) \\ &+\mathrm{diag}((\mathbf{\Sigma}_{AA}+\sigma^2\mathbf{I}_m)^{-1}) - \lambda\mathrm{diag}(\mathbf{\Sigma}_{YY}^{-1}).\end{aligned} \tag{4.8}$$

Note that the critical point satisfies

$$(\lambda-1)(\mathbf{\Sigma}_{YY}+\mathbf{\Sigma}_{AA})^{-1} + (\mathbf{\Sigma}_{AA}+\sigma^2\mathbf{I}_m)^{-1} - \lambda\mathbf{\Sigma}_{YY}^{-1} = \mathbf{0}, \tag{4.9}$$

which has a solution given by $\mathbf{\Sigma}_{AA}^{\star} = \frac{1}{\lambda}\mathbf{H}\mathbf{\Sigma}_{XX}\mathbf{H}^{\mathrm{T}}$. The result follows immediately from combining this result with Proposition 3.4 and Proposition 4.1. □

The generalized stealth attacks given in Theorem 4.1 are also the solution to the optimization problem given by

$$\min_{\mathbf{\Sigma}_{AA}\in\mathcal{S}_+^m} I(X^n;Y_A^m) \tag{4.10}$$

$$\text{s.t. } D(P_{Y_A^m}\|P_{Y^m})\leq\delta, \tag{4.11}$$

where δ is an upper bound for the KL divergence between the distribution of measurements with attacks and without attacks, and

$$\delta\leq D(P_{Y_{A^\star}^m}\|P_{Y^m}), \tag{4.12}$$

in which $D(P_{Y_{A^\star}^m}\|P_{Y^m})$ is the KL divergence achieved by the stealth attacks, i.e. when $\mathbf{\Sigma}_{AA}^\star=\mathbf{H}\mathbf{\Sigma}_{XX}\mathbf{H}^{\mathrm{T}}$. Here the setting in (4.12) guarantees that the asymptotic probability of detection of the attacks is smaller than that of the stealth attacks. The Lagrangian $\mathcal{L}:\mathbb{R}^{m\times m}\times\mathbb{R}\to\mathbb{R}$ associated with the optimization problem in (4.10) and (4.11) is given by

$$\mathcal{L}\left(\mathbf{\Sigma}_{AA},\lambda\right)=I(X^n;Y_A^m)+\lambda\left(D(P_{Y_A^m}\|P_{Y^m})-\delta\right), \tag{4.13}$$

in which λ behaves as the Lagrange multiplier. It is easy to see that $\mathbf{\Sigma}_{AA}^\star=\frac{1}{\lambda}\mathbf{H}\mathbf{\Sigma}_{XX}\mathbf{H}^{\mathrm{T}}$ is the saddle point of the Lagrangian function $\mathcal{L}\left(\mathbf{\Sigma}_{AA},\lambda\right)$, as

$$\frac{\partial}{\partial\mathbf{\Sigma}_{AA}}\mathcal{L}\left(\mathbf{\Sigma}_{AA},\lambda\right)=\frac{\partial}{\partial\mathbf{\Sigma}_{AA}}\left(I(X^n;Y_A^m)+\lambda D(P_{Y_A^m}\|P_{Y^m})\right), \tag{4.14}$$

which is characterized in Theorem 4.1. Then the optimality of the generalized stealth attacks follows from the Karush-Kuhn-Tucker conditions [120, pp.243].

Theorem 4.1 shows that the generalized stealth attacks share the same structure of the stealth attacks in Theorem 3.1 up to a scaling factor determined by λ. The solution in Theorem 4.1 holds for the case in which $\lambda\geq 1$, and therefore, lacks full generality. However the case in which $\lambda<1$ yields unreasonably high probability of detection, which indicates that the proposed attack construction is indeed of practical interest in a wide range of state estimation settings. Furthermore the optimization problem in (4.4) results in a non-convex problem when $\lambda<1$ and the solution obtained above no longer holds.

For the bi-objective optimization problem with objectives

$$\min_{A^m} I(X^n;Y_A^m) \qquad \text{and} \qquad \min_{A^m} D(P_{Y_A^m}\|P_{Y^m}), \tag{4.15}$$

changing the value of λ yields different solutions on the Pareto front of the optimization problem in (4.4). This implies that the solution moves along the Pareto front when the attacker changes the value of λ. We will show this in the numerical results in Section 4.3.4. For any $\lambda \geq 1$, Theorem 4.1 guarantees that the generalized stealth attack is the only Parento efficient solution, i.e. the attack construction that minimizes the mutual information subject to the probability of detection constraint being satisfied. By increasing the value of λ the attacker places more importance on the probability of detection than on the mutual information which results in a more conservative attack that disrupts less but is more difficult to detect.

Theorem 4.1 also shows that the resulting attack construction is remarkably simple to implement provided that the information about the system is available to the attacker. Indeed, the attacker only requires access to the linearized Jacobian measurement matrix $\mathbf{H}$ and the second order statistics of the state variables, but the variance of the noise introduced by the sensors is not necessary, this is the same as the stealth attacks in Theorem 3.1. To obtain the Jacobian, a malicious attacker needs to know the topology of the grid, the admittances of the branches, and the operation point of the system. On the other hand, the second order statistics of the state variables can be estimated using historical data, i.e. the covariance matrix of the state variables can be approximated by the sample covariance matrix. Later in Chapter 5, we will show that the attack construction with a sample covariance matrix of the state variables obtained with historical data is always suboptimal when the attacker has a limited number of training samples, but is asymptotically optimal when the size of the training data grows to infinity.

For the generalized stealth attack given in Theorem 4.1, the mutual information induced by the attack is given in the following corollary.

Corollary 4.1. *The mutual information between the vector of state variables and the vector of compromised measurements induced by the optimal attack construction is given by*

$$I(X^n; Y_A^m) = \frac{1}{2} \log \left| \mathbf{H}\boldsymbol{\Sigma}_{XX}\mathbf{H}^{\mathrm{T}} \left(\sigma^2 \mathbf{I}_m + \frac{1}{\lambda} \mathbf{H}\boldsymbol{\Sigma}_{XX}\mathbf{H}^{\mathrm{T}} \right)^{-1} + \mathbf{I}_m \right|. \tag{4.16}$$

Proof. Combining the generalized stealth attacks given in (4.7) with Proposition 3.1 yields the result. □

Corollary 4.1 shows that the mutual information increases monotonically with λ and that it asymptotically converges to $I(X^n; Y^m)$, i.e. the case in which there is no attack. While the evaluation of the mutual information as shown in Corollary 4.1 is straightforward, the computation of the associated probability of detection

yields involved expressions that do not provide much insight. For that reason, the probability of detection of optimal attacks is treated in the following section.

4.2 Probability of Detection

The asymptotic probability of detection of the generalized stealth attacks characterized in Theorem 4.1 is governed by the KL divergence as described in Lemma 2.2. However in the non-asymptotic case, determining the probability of detection is difficult, and therefore, choosing a value of λ that provides the desired probability of detection is a challenging task. In this section, we first provide a closed-form expression of the probability of detection by direct evaluation and show that the expression does not provide any practical insight over the choice of λ that achieves the desired detection performance. That being the case, we then provide an upper bound on the probability of detection, which in turn provides a lower bound on the value of λ that achieves the desired probability of detection.

4.2.1 Direct Evaluation of Probability of Detection

As stated in Section 3.1.3, the detection within the Bayesian framework is cast as a hypothesis testing problem with hypotheses

$$\begin{aligned} &\mathcal{H}_0 : Y^m \sim \mathcal{N}(\mathbf{0}, \mathbf{\Sigma}_{YY}), \quad \text{versus} \\ &\mathcal{H}_1 : Y^m \sim \mathcal{N}(\mathbf{0}, \mathbf{\Sigma}_{Y_A Y_A}). \end{aligned} \tag{4.17}$$

The LRT between $\mathcal{H}_0$ and $\mathcal{H}_1$, which is given by

$$L(\mathbf{y}) = \frac{f_{Y_A^m}(\mathbf{y})}{f_{Y^m}(\mathbf{y})} \underset{\mathcal{H}_0}{\overset{\mathcal{H}_1}{\gtrless}} \tau, \tag{4.18}$$

is chosen by the operator to detect the attacks due to the optimality from Neyman-Pearson lemma in Lemma 2.1.

For the LRT with threshold τ in (4.18), the probability of detection of the attacks is the probability that the likelihood ratio between $P_{Y_A^m}$ and P_{Y^m} is larger than τ for any realization of Y_A^m. As a result, the detection based on the LRT with threshold τ in (4.18) yields a probability of detection given by

$$\mathsf{P_D} \triangleq \mathbb{E}\left[\mathbb{1}_{\left\{L(Y_A^m) \geq \tau\right\}}\right], \tag{4.19}$$

where $\mathbb{1}_{\{.\}}$ is the indicator function. The following theorem particularizes the above expression to the optimal attack construction described in Theorem 4.1.

Theorem 4.2. *The probability of detection of the LRT in (4.18) for the attack construction in (4.7) is given by*

$$\mathsf{P}_{\mathsf{D}}(\lambda) = \mathbb{P}\left[(U^p)^{\mathrm{T}}\boldsymbol{\Delta}U^p \geq \lambda\left(2\log\tau + \log\left|\mathbf{I}_p + \lambda^{-1}\boldsymbol{\Delta}\right|\right)\right], \tag{4.20}$$

where $p = rank(\mathbf{H}\boldsymbol{\Sigma}_{XX}\mathbf{H}^{\mathrm{T}})$*,* $U^p \in \mathbb{R}^p$ *is a vector of random variables with distribution* $\mathcal{N}(\mathbf{0}, \mathbf{I}_p)$*, and* $\boldsymbol{\Delta} \in \mathbb{R}^{p\times p}$ *is a diagonal matrix with entries given by* $(\boldsymbol{\Delta})_{i,i} = \lambda_i(\mathbf{H}\boldsymbol{\Sigma}_{XX}\mathbf{H}^{\mathrm{T}})\lambda_i(\boldsymbol{\Sigma}_{YY}^{-1})$*, where* $\lambda_i(\mathbf{A})$ *with* $i = 1, \ldots, p$ *denotes the* i*-th eigenvalue of matrix* $\mathbf{A}$ *in descending order.*

Proof. Taking the Gaussian p.d.f. of Y_A^m, i.e. $Y_A^m \sim \mathcal{N}(\mathbf{0}, \boldsymbol{\Sigma}_{YY} + \boldsymbol{\Sigma}_{AA}^{\star})$ with $\boldsymbol{\Sigma}_{AA}^{\star}$ given in (4.7), into the probability of detection given in (4.19) yields

$$\begin{aligned}
\mathsf{P}_{\mathsf{D}}(\lambda) &= \int_{\mathcal{S}} \mathrm{d}P_{Y_A^m} && (4.21)\\
&= \frac{1}{(2\pi)^{\frac{m}{2}}\left|\boldsymbol{\Sigma}_{Y_AY_A}\right|^{\frac{1}{2}}}\int_{\mathcal{S}} \exp\left\{-\frac{1}{2}\mathbf{y}^{\mathrm{T}}\boldsymbol{\Sigma}_{Y_AY_A}^{-1}\mathbf{y}\right\}\mathrm{d}\mathbf{y}, && (4.22)
\end{aligned}$$

where the integration domain $\mathcal{S}$ only contains the realizations of Y_A^m that yield a likelihood ratio value larger than τ, i.e.

$$\begin{aligned}
\mathcal{S} &= \{\mathbf{y} \in \mathbb{R}^m : L(\mathbf{y}) \geq \tau\} && (4.23)\\
&= \left\{\mathbf{y} \in \mathbb{R}^m : \frac{\left|\boldsymbol{\Sigma}_{YY}\right|^{\frac{1}{2}}}{\left|\boldsymbol{\Sigma}_{Y_AY_A}\right|^{\frac{1}{2}}}\exp\left\{-\frac{1}{2}\mathbf{y}^{\mathrm{T}}\boldsymbol{\Sigma}_{Y_AY_A}^{-1}\mathbf{y} + \frac{1}{2}\mathbf{y}^{\mathrm{T}}\boldsymbol{\Sigma}_{YY}^{-1}\mathbf{y}\right\} \geq \tau\right\} && (4.24)\\
&= \left\{\mathbf{y} \in \mathbb{R}^m : \mathbf{y}^{\mathrm{T}}\boldsymbol{\Delta}_0\mathbf{y} \geq 2\log\tau + \log\left|\mathbf{I}_m + \boldsymbol{\Sigma}_{AA}\boldsymbol{\Sigma}_{YY}^{-1}\right|\right\}, && (4.25)
\end{aligned}$$

where (4.24) follows from taking $Y_A^m \sim \mathcal{N}(\mathbf{0}, \boldsymbol{\Sigma}_{YY} + \boldsymbol{\Sigma}_{AA}^{\star})$ into (4.18); and (4.25) follows from taking logarithm on both sides of the inequality in (4.24) with $\boldsymbol{\Delta}_0 \triangleq \boldsymbol{\Sigma}_{YY}^{-1} - \boldsymbol{\Sigma}_{Y_AY_A}^{-1}$.

Let $\boldsymbol{\Sigma}_{YY} = \mathbf{U}_{YY}\boldsymbol{\Lambda}_{YY}\mathbf{U}_{YY}^{\mathrm{T}}$ where $\boldsymbol{\Lambda}_{YY} \in \mathbb{R}^{m\times m}$ is a diagonal matrix containing the eigenvalues of $\boldsymbol{\Sigma}_{YY}$ in descending order and $\mathbf{U}_{YY} \in \mathbb{R}^{m\times m}$ is a unitary matrix whose columns are the eigenvectors of $\boldsymbol{\Sigma}_{YY}$ ordered matching the order of the eigenvalues. Noticing that

$$\boldsymbol{\Sigma}_{AA}^{\star} = \frac{1}{\lambda}\mathbf{H}\boldsymbol{\Sigma}_{XX}\mathbf{H}^{\mathrm{T}} = \frac{1}{\lambda}\left(\boldsymbol{\Sigma}_{YY} - \sigma^2\mathbf{I}_m\right) \tag{4.26}$$

and

$$\boldsymbol{\Sigma}_{Y_AY_A} = \boldsymbol{\Sigma}_{YY} + \boldsymbol{\Sigma}_{AA}^{\star} = \mathbf{H}\boldsymbol{\Sigma}_{XX}\mathbf{H}^{\mathrm{T}} + \sigma^2\mathbf{I}_m + \frac{1}{\lambda}\mathbf{H}\boldsymbol{\Sigma}_{XX}\mathbf{H}^{\mathrm{T}} \tag{4.27}$$

are also diagonalized by $\mathbf{U}_{YY}$.

Applying the change of variable

$$\mathbf{y}_1 \triangleq \mathbf{U}_{YY}\mathbf{y} \tag{4.28}$$

in (4.22) results in

$$\begin{aligned}
\mathsf{P}_\mathsf{D}(\lambda) &= \frac{1}{(2\pi)^{\frac{m}{2}}\left|\mathbf{\Sigma}_{Y_AY_A}\right|^{\frac{1}{2}}}\int_{\mathcal{S}}\exp\left\{-\frac{1}{2}\mathbf{y}^{\mathrm{T}}\mathbf{\Sigma}_{Y_AY_A}^{-1}\mathbf{y}\right\}\mathrm{d}\mathbf{y} && (4.29)\\
&= \frac{1}{(2\pi)^{\frac{m}{2}}\left|\mathbf{\Sigma}_{Y_AY_A}\right|^{\frac{1}{2}}}\int_{\mathcal{S}}\exp\left\{-\frac{1}{2}(\mathbf{U}_{YY}\mathbf{y})^{\mathrm{T}}\mathbf{\Lambda}_{Y_AY_A}^{-1}\mathbf{U}_{YY}\mathbf{y}\right\}\mathrm{d}\mathbf{y} && (4.30)\\
&= \frac{1}{(2\pi)^{\frac{m}{2}}\left|\mathbf{\Sigma}_{Y_AY_A}\right|^{\frac{1}{2}}}\int_{\mathcal{S}_1}\exp\left\{-\frac{1}{2}\mathbf{y}_1^{\mathrm{T}}\mathbf{\Lambda}_{Y_AY_A}^{-1}\mathbf{y}_1\right\}\mathrm{d}\mathbf{y}_1, && (4.31)
\end{aligned}$$

where $\mathbf{\Lambda}_{Y_AY_A} \in \mathbb{R}^{m\times m}$ denotes the diagonal matrix containing the eigenvalues of $\mathbf{\Sigma}_{Y_AY_A}$ in descending order. Also the integration domain $\mathcal{S}$ is changed into $\mathcal{S}_1$, which is given by

$$\mathcal{S}_1 = \left\{\mathbf{y}_1 \in \mathbb{R}^m : \mathbf{y}_1^{\mathrm{T}}\mathbf{\Delta}_1\mathbf{y}_1 \geq 2\log\tau + \log\left|\mathbf{I}_m + \mathbf{\Lambda}_{AA}\mathbf{\Lambda}_{YY}^{-1}\right|\right\}, \tag{4.32}$$

where $\mathbf{\Delta}_1 \triangleq \mathbf{\Lambda}_{YY}^{-1} - \mathbf{\Lambda}_{Y_AY_A}^{-1}$ and $\mathbf{\Lambda}_{AA}$ denotes the diagonal matrix containing the eigenvalues of $\mathbf{\Sigma}_{AA}$ in descending order.

Further applying the change of variable

$$\mathbf{y}_2 \triangleq \mathbf{\Lambda}_{Y_AY_A}^{-\frac{1}{2}}\mathbf{y}_1 \tag{4.33}$$

for (4.31) results in

$$\begin{aligned}
\mathsf{P}_\mathsf{D}(\lambda) &= \frac{1}{(2\pi)^{\frac{m}{2}}\left|\mathbf{\Sigma}_{Y_AY_A}\right|^{\frac{1}{2}}}\int_{\mathcal{S}_1}\exp\left\{-\frac{1}{2}\mathbf{y}_1^{\mathrm{T}}\mathbf{\Lambda}_{Y_AY_A}^{-1}\mathbf{y}_1\right\}\mathrm{d}\mathbf{y}_1 && (4.34)\\
&= \frac{1}{(2\pi)^{\frac{m}{2}}\left|\mathbf{\Sigma}_{Y_AY_A}\right|^{\frac{1}{2}}}\int_{\mathcal{S}_1}\exp\left\{-\frac{1}{2}\left(\mathbf{\Lambda}_{Y_AY_A}^{-\frac{1}{2}}\mathbf{y}_1\right)^{\mathrm{T}}\mathbf{\Lambda}_{Y_AY_A}^{-\frac{1}{2}}\mathbf{y}_1\right\}\mathrm{d}\mathbf{y}_1 && (4.35)\\
&= \frac{1}{\sqrt{(2\pi)^m}}\int_{\mathcal{S}_2}\exp\{-\frac{1}{2}\mathbf{y}_2^{\mathrm{T}}\mathbf{y}_2\}\mathrm{d}\mathbf{y}_2, && (4.36)
\end{aligned}$$

with the transformed integration domain given by

$$\mathcal{S}_2 = \left\{\mathbf{y}_2 \in \mathbb{R}^m : \mathbf{y}_2^{\mathrm{T}}\mathbf{\Delta}_2\mathbf{y}_2 \geq 2\log\tau + \log\left|\mathbf{I}_m + \mathbf{\Delta}_2\right|\right\}, \tag{4.37}$$

with

$$\mathbf{\Delta}_2 \triangleq \mathbf{\Lambda}_{AA}\mathbf{\Lambda}_{YY}^{-1}. \tag{4.38}$$

Setting $\mathbf{\Delta} \triangleq \lambda \mathbf{\Delta}_2$, the result given by

$$\operatorname{rank}(\mathbf{\Delta}) = \operatorname{rank}(\mathbf{\Delta}_2) = \operatorname{rank}(\mathbf{\Lambda}_{AA}\mathbf{\Lambda}_{YY}^{-1}) = \operatorname{rank}(\mathbf{\Lambda}_{AA}) = \operatorname{rank}(\mathbf{H}\mathbf{\Sigma}_{XX}\mathbf{H}^{\mathrm{T}}) \quad (4.39)$$

follows immediately from the fact that $\mathbf{\Lambda}_{YY}$ is of full rank and diagonal. Combining (4.39) and (4.37) yield the theorem. □

Theorem 4.2 shows that the probability of detection is equivalent to the probability that a weighted sum of independent χ^2 random variables exceeds a certain threshold. In our setting, the threshold is determined by the tradeoff parameter λ. Notice that the left-hand term $(U^p)^{\mathrm{T}}\mathbf{\Delta}U^p$ in (4.20) is a weighted sum of independent χ^2 distributed random variables with a single degree of freedom where the weights are determined by the diagonal entries of $\mathbf{\Delta}$ which depend on the second order statistics of the state variables, the Jacobian measurement matrix, and the variance of the noise; i.e. the attacker has no control over this term. The right-hand side contains in addition λ and τ, and therefore, the probability of attack detection is described as a function of the parameter λ.

The probability of false alarm of the LRT given in (3.11) for the attack construction in (4.7) is given by

$$\mathsf{P}_{\mathsf{FA}} \triangleq \mathbb{E}\left[\mathbb{1}_{\{L(Y^m)\geq\tau\}}\right]. \quad (4.40)$$

As we focus on the probability of detection in this thesis, the probability of false alarm is proved in Appendix C using the same approach in Theorem 4.2, which is also equivalent to the probability that a weighted sum of independent χ^2 random variables exceeds a certain threshold.

Unfortunately, no closed-form expression is available for the distribution of positively weighted sum of independent χ^2 random variables with one degree of freedom [124]. To solve this problem, some moment matching approximation techniques approximate this distribution by matching the moment of this distribution to some order, such as the Lindsay-Pilla-Basak (LPB) method [125]. But the resulting expressions are complex and the relation of the probability of detection with λ is difficult to describe analytically following this course of action. As a result, both the closed-form expression in Theorem 4.2 and the moment matching methods provide no detailed insight into the relation between the probability of detection and λ. So there is no clue for the attacker to choose the suitable λ that achieves a certain probability of detection.

In the following, an upper bound on the probability of attack detection is derived. The upper bound is then used to provide a simple lower bound on the value λ that achieves the desired probability of detection.

4.2.2 Upper Bound for Probability of Detection

In this section, a lower bound for λ that yields an upper bound on the probability of detection is proposed using concentration inequality result. This gives the attacker a guideline in choosing suitable λ to achieve a certain probability of detection. The concentration inequality result, or χ^2 tail inequality, that is used in the proof later is provided first.

Lemma 4.2. [126, Proposition 1.1] *Let $\mathbf{A} \in \mathbb{R}^{n\times n}$ be a matrix, and let $\widetilde{\mathbf{\Sigma}} = \mathbf{A}^{\mathrm{T}}\mathbf{A}$. Let Q^n be an isotropic multivariate Gaussian random vector with mean zero. For all $t > 0$,*

$$\mathbb{P}\left[(Q^n)^{\mathrm{T}}\widetilde{\mathbf{\Sigma}}Q^n > \mathrm{tr}(\widetilde{\mathbf{\Sigma}}) + 2\sqrt{\mathrm{tr}(\widetilde{\mathbf{\Sigma}}^2)t} + 2\|\widetilde{\mathbf{\Sigma}}\|_\infty t\right] \leq e^{-t}, \tag{4.41}$$

where $\|\widetilde{\mathbf{\Sigma}}\|_\infty$ is the spectral norm of matrix $\tilde{\mathbf{\Sigma}}$.

The isotropic multivariate Gaussian random vector implies that $Q^n \sim \mathcal{N}(\mathbf{0}, \sigma_1^2\mathbf{I}_n)$, i.e. the elements of Q^n share the same variance σ_1^2. For any diagonal matrices, the spectral norm of the matrix is the maximum entries of the matrix, which is equivalent to the infinity norm of the matrix.

Note that when $\sigma_1^2 = 1$ in Lemma 4.2 the expected value of $(Q^n)^{\mathrm{T}}\widetilde{\mathbf{\Sigma}}Q^n$ is $\mathrm{tr}(\widetilde{\mathbf{\Sigma}})$. So the concentration inequality result in Lemma 4.2 shows that the probability of $(Q^n)^{\mathrm{T}}\widetilde{\mathbf{\Sigma}}Q^n$ deviating the mean by $2\sqrt{\mathrm{tr}(\widetilde{\mathbf{\Sigma}}^2)t} + 2\|\widetilde{\mathbf{\Sigma}}\|_\infty t$ is upper bounded by e^{-t} for all $t > 0$. Interestingly the left-hand term $(U^p)^{\mathrm{T}}\mathbf{\Delta}U^p$ in (4.20) is identical to $(Q^n)^{\mathrm{T}}\widetilde{\mathbf{\Sigma}}Q^n$ when $\widetilde{\mathbf{\Sigma}} = \mathbf{\Delta}$. As a result, an upper bound for the probability of detection can be obtained using the concentration inequality result in Lemma 4.2.

In the following, the concentration inequality upper bound for probability of detection is proposed.

Theorem 4.3. *Let $\tau > 1$ be the decision threshold of the LRT. For any $t > 0$ and $\lambda \geq \max(\lambda^\star(t), 1)$ then the probability of attack detection satisfies*

$$\mathsf{P}_\mathsf{D}(\lambda) \leq e^{-t}, \tag{4.42}$$

where $\lambda^\star(t)$ is the only positive solution of λ satisfying

$$2\lambda\log\tau - \frac{1}{2\lambda}\mathrm{tr}(\mathbf{\Delta}^2) - 2\sqrt{\mathrm{tr}(\mathbf{\Delta}^2)t} - 2\|\mathbf{\Delta}\|_\infty t = 0. \tag{4.43}$$

Proof. We start with the result of Theorem 4.2 which gives

$$\mathsf{P}_{\mathsf{D}}(\lambda) = \mathbb{P}\left[(U^p)^{\mathrm{T}}\boldsymbol{\Delta}U^p \geq \lambda\left(2\log\tau + \log\left|\mathbf{I}_p + \lambda^{-1}\boldsymbol{\Delta}\right|\right)\right]. \tag{4.44}$$

Note that $\boldsymbol{\Delta}$ is a diagonal matrix, we now proceed to expand the term $\log|\mathbf{I}_p + \lambda^{-1}\boldsymbol{\Delta}|$ using the Taylor series expansion, which results in

$$\begin{aligned}
&\log\left|\mathbf{I}_p + \lambda^{-1}\boldsymbol{\Delta}\right| \\
&= \sum_{i=1}^{p}\log\left(1 + \lambda^{-1}(\boldsymbol{\Delta})_{i,i}\right) &(4.45)\\
&= \sum_{i=1}^{p}\left(\sum_{j=1}^{\infty}\left(\frac{(\lambda^{-1}(\boldsymbol{\Delta})_{i,i})^{2j-1}}{2j-1} - \frac{(\lambda^{-1}(\boldsymbol{\Delta})_{i,i})^{2j}}{2j}\right)\right). &(4.46)
\end{aligned}$$

Since

$$(\boldsymbol{\Delta})_{i,i} = \lambda_i(\mathbf{H}\boldsymbol{\Sigma}_{XX}\mathbf{H}^{\mathrm{T}})\lambda_i(\boldsymbol{\Sigma}_{YY}^{-1}) = \frac{\lambda_i(\mathbf{H}\boldsymbol{\Sigma}_{XX}\mathbf{H}^{\mathrm{T}})}{\lambda_i(\mathbf{H}\boldsymbol{\Sigma}_{XX}\mathbf{H}^{\mathrm{T}}) + \sigma^2} \leq 1 \tag{4.47}$$

for $i = 1, \ldots, p$ and $\lambda \geq 1$, then $\lambda^{-1}(\boldsymbol{\Delta})_{i,i} \leq 1$ and

$$\frac{(\lambda^{-1}(\boldsymbol{\Delta})_{i,i})^{2j-1}}{2j-1} - \frac{(\lambda^{-1}(\boldsymbol{\Delta})_{i,i})^{2j}}{2j} \geq 0, \text{ for } j \in \mathbb{Z}^+. \tag{4.48}$$

Thus, (4.46) is lower bounded by the second order Taylor expansion, i.e.,

$$\begin{aligned}
\log|\mathbf{I}_p + \boldsymbol{\Delta}| &\geq \sum_{i=1}^{p}\left(\lambda^{-1}(\boldsymbol{\Delta})_{i,i} - \frac{(\lambda^{-1}(\boldsymbol{\Delta})_{i,i})^2}{2}\right) &(4.49)\\
&\geq \frac{1}{\lambda}\mathrm{tr}(\boldsymbol{\Delta}) - \frac{1}{2\lambda^2}\mathrm{tr}(\boldsymbol{\Delta}^2). &(4.50)
\end{aligned}$$

Substituting (4.50) in (4.44) yields

$$\mathsf{P}_{\mathsf{D}}(\lambda) \leq \mathbb{P}\left[(U^p)^{\mathrm{T}}\boldsymbol{\Delta}U^p \geq \mathrm{tr}(\boldsymbol{\Delta}) + 2\lambda\log\tau - \frac{1}{2\lambda}\mathrm{tr}(\boldsymbol{\Delta}^2)\right]. \tag{4.51}$$

Note that $\mathbb{E}\left[(U^p)^{\mathrm{T}}\boldsymbol{\Delta}U^p\right] = \mathrm{tr}(\boldsymbol{\Delta})$, and therefore, evaluating the probability in (4.51) is equivalent to evaluating the probability of $(U^p)^{\mathrm{T}}\boldsymbol{\Delta}U^p$ deviating $2\lambda\log\tau - \frac{1}{2\lambda}\mathrm{tr}(\boldsymbol{\Delta}^2)$ from the mean. In view of this and using Lemma 4.2, the right-hand side in (4.51) is upper bounded by

$$\begin{aligned}
\mathsf{P}_{\mathsf{D}}(\lambda) &\leq \mathbb{P}\left[(U^p)^{\mathrm{T}}\boldsymbol{\Delta}U^p \geq \mathrm{tr}(\boldsymbol{\Delta}) + 2\sqrt{\mathrm{tr}(\boldsymbol{\Delta}^2)t} + 2\|\boldsymbol{\Delta}\|_\infty t\right] &(4.52)\\
&\leq e^{-t} &(4.53)
\end{aligned}$$

for $t > 0$ when

$$2\lambda \log \tau - \frac{1}{2\lambda}\mathrm{tr}(\mathbf{\Delta}^2) \geq 2\sqrt{\mathrm{tr}(\mathbf{\Delta}^2)t} + 2\|\mathbf{\Delta}\|_\infty t \tag{4.54}$$

is satisfied.

The expression in (4.54) is rewritten as

$$(2\log\tau)\lambda^2 - \left(2\sqrt{\mathrm{tr}(\mathbf{\Delta}^2)t} + 2\|\mathbf{\Delta}\|_\infty t\right)\lambda - \frac{1}{2}\mathrm{tr}(\mathbf{\Delta}^2) \geq 0. \tag{4.55}$$

When $\tau \geq 1$, it is clear that the quadratic equation on the left-hand side of (4.55) is convex and goes through the point $\left(0, -\frac{1}{2}\mathrm{tr}(\mathbf{\Delta}^2)\right)$. Then the quadratic equation on the left-hand side of (4.55) is satisfied with equality for two values of λ, one is strictly negative and the other one is strictly positive denoted by $\lambda^*(t)$, which is given by

$$\lambda^\star(t) = \frac{\left(2\sqrt{\mathrm{tr}(\mathbf{\Delta}^2)t} + 2\|\mathbf{\Delta}\|_\infty t\right) + \sqrt{\left(2\sqrt{\mathrm{tr}(\mathbf{\Delta}^2)t} + 2\|\mathbf{\Delta}\|_\infty t\right)^2 + 4\mathrm{tr}(\mathbf{\Delta}^2)\log\tau}}{4\log\tau}. \tag{4.56}$$

The result follows by noticing that the left-hand term of (4.55) increases monotonically for $\lambda \geq \lambda^\star(t)$ and choosing $\lambda \geq \max(\lambda^\star(t), 1)$. This concludes the proof. □

It is interesting to note that for large values of λ the probability of detection decreases exponentially fast with λ. We show in the numerical results section below that the regime in which the exponentially fast decrease kicks-in does not align with the saturation of the mutual information loss induced by the attack.

The assumption that $\tau > 1$ is a realistic setting. When $\tau = 1$, the probability of false alarm is one for the LRT given in (4.18), as the likelihood ratio always equals to one for every realization of Y^m. Considering the unreasonable high probability of false alarm, the operator of the power system has to set $\tau > 1$. Usually λ is of some small or moderate value, as high value of λ results in a low probability of detection.

The expression in (4.46) is also lower bounded by the higher order Taylor expansion, which leads to a tighter lower bound. However, the Taylor expansion of higher order also increases the order of the inequality given in (4.50). This, in turn, results in a $\lambda^\star$ that is less intuitive. The generalization to higher order terms is straightforward.

The addition of the requirement that $\lambda \geq 1$ into $\lambda \geq \lambda^*(t)$ comes from the definition of the generalized stealth attacks, in which λ has to be greater than or equal to one. Usually for the power system $\lambda^*(t)$ is greater than one, especially for the power systems of large scale. This is from the fact that the $\sqrt{\mathrm{tr}(\mathbf{\Delta}^2)t} + 2\|\mathbf{\Delta}\|_\infty t$

term in the nominator of (4.56) is usually larger than $2\log\tau$ for small and moderate values of τ, and increases as the size of the grid increases.

4.3 Numerical Evaluation

In this section, we present simulations to evaluate the performance of the proposed attack strategy in practical state estimation settings. In particular, the IEEE 14-Bus, 30-Bus, and 118-Bus test systems are considered in the simulation. In state estimation with linearized dynamics, the Jacobian measurement matrix is determined by the operation point, see (2.15). We assume a DC state estimation scenario, in which we set the resistances of the branches to 0 and the bus voltage magnitude to 1.0 per unit, c.f. (2.23) and (2.24). Note that in this setting it is sufficient to specify the network topology, the branch reactances, real power flow, and the power injection values to fully characterize the system. Specifically, we use the IEEE test system framework provided by MATPOWER [122]. We choose the bus voltage angles to be the state variables, and use the power injection and the power flows in both directions as the measurements.

As stated in Section 4.2.1, there is no closed-form expression for the distribution of a positively weighted sum of independent χ^2 random variables, which is required to calculate the probability of detection of the generalized stealth attacks as shown in Lemma 4.2. For that reason, we use the LPB method and the MOMENTCHI2 package [127] to numerically evaluate the probability of attack detection.

The simulation setting is the same as in Section 3.4. The covariance matrix of the state variables is assumed to be a Toeplitz matrix with exponential decay parameter ρ, where the exponential decay parameter ρ determines the correlation strength between different entries of the state variable vector. The performance of the generalized stealth attack is a function of weight given to the detection term in the attack construction cost function, i.e. λ, the correlation strength between state variables, i.e. ρ, and the SNR of the power system which is defined in (3.56).

4.3.1 Performance of Generalized Stealth Attacks

Fig. 4.1 and Fig. 4.2 depict the performance of the optimal attack construction given in (4.7) for different values of ρ with SNR = 10 dB and SNR = 20 dB, respectively, when $\lambda = 2$ and $\tau = 2$. Interestingly, the performance of the attack construction does not change monotonically with correlation strength, which suggests that the correlation among the state variables does not necessarily provide an advantage to the attacker. Admittedly, for a small or moderate value of ρ, the performance

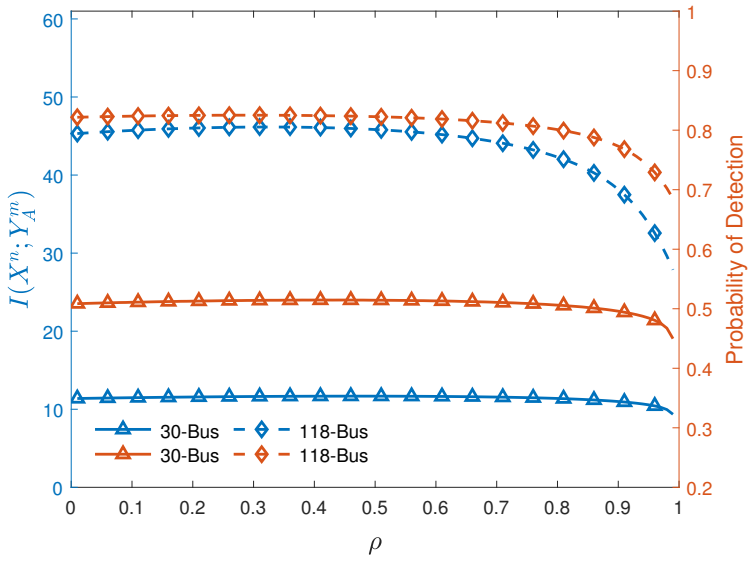


Fig. 4.1. Performance of the generalized stealth attack in terms of mutual information and probability of detection for different values of ρ when $\lambda = 2$, $\tau = 2$, and SNR $= 10$ dB.

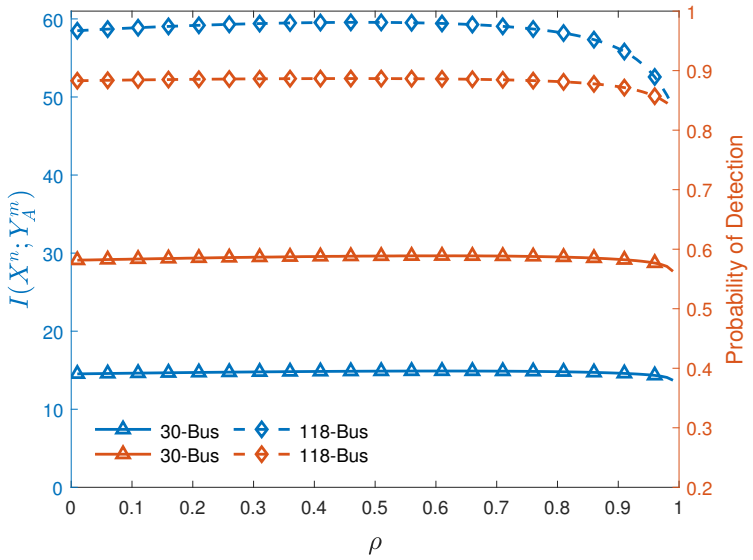


Fig. 4.2. Performance of the generalized stealth attack in terms of mutual information and probability of detection for different values of ρ when $\lambda = 2$, $\tau = 2$, and SNR $= 20$ dB.

of the attack does not change significantly with ρ for both objectives. This effect is more noticeable in the high SNR scenario. However, for large values of ρ the performance of the attack improves significantly in terms of both mutual information and probability of detection. Moreover, the advantage provided by large values of ρ is more significant for the 118-Bus system than for the 30-Bus system, which indicates that the correlation between the state variables is easier to exploit for the attacker in large systems.

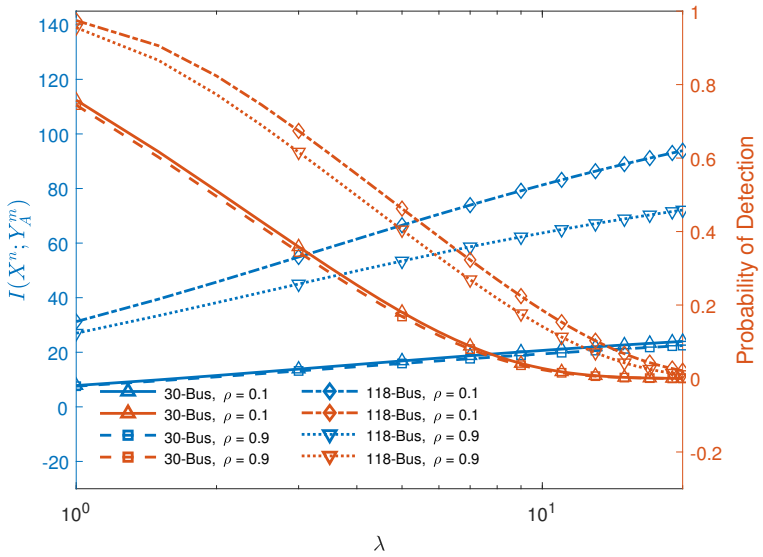


Fig. 4.3. Performance of the generalized stealth attack in terms of mutual information and probability of detection for different values of λ and system size when $\rho = 0.1$, $\rho = 0.9$, SNR $= 10$ dB and $\tau = 2$.

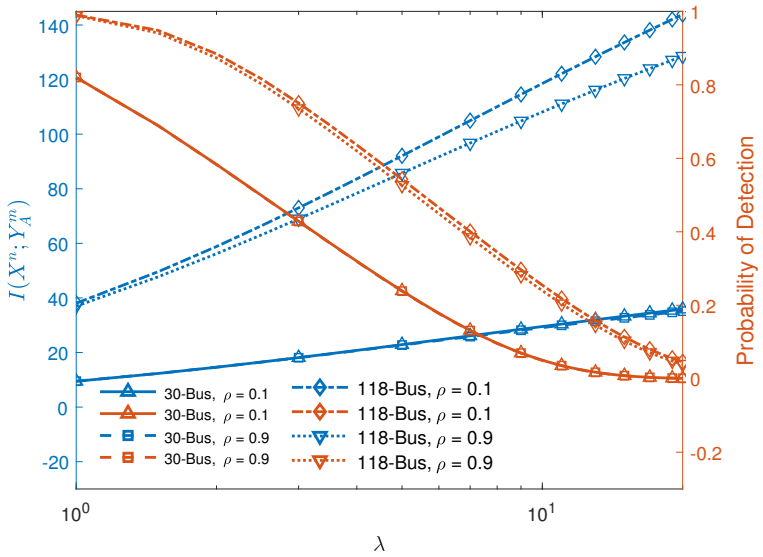


Fig. 4.4. Performance of the generalized stealth attack in terms of mutual information and probability of detection for different values of λ and system size when $\rho = 0.1$, $\rho = 0.9$, SNR $= 20$ dB and $\tau = 2$.

Fig. 4.3 and Fig. 4.4 depict the performance of the optimal attack construction for different values of λ and ρ with SNR $= 10$ dB and SNR $= 20$ dB, respectively, when $\tau = 2$. As expected, larger values of the parameter λ yield smaller values of the probability of attack detection while increasing the mutual information between the state variables vector and the compromised measurement vector. We observe that the probability of detection decreases approximately linearly with respect to $\log \lambda$ for moderate values of λ. On the other hand, Theorem 4.3 states that for large values of

λ the probability of detection decreases exponentially fast to zero. However, for the range of values of λ in which the decrease of probability of detection is approximately linear with respect to $\log \lambda$, there is no significant reduction in the rate of growth of mutual information. In view of this, the attacker needs to choose the value of λ carefully as the convergence of the mutual information to the asymptote $I(X^n; Y^m)$ is slower than that of the probability of detection to zero.

The comparison between the 30-Bus and 118-Bus systems shows that for the smaller size system the probability of detection decreases faster to zero while the rate of growth of mutual information is smaller than that on the larger system. This suggests that the choice of λ is particularly critical in large size systems as smaller size systems exhibit a more robust attack performance for different values of λ. The effect of the correlation between the state variables is significantly more noticeable for the 118-bus system. While there is a performance gain for the 30-bus system in terms of both mutual information and probability of detection due to the high correlation between the state variables, the improvement is more noteworthy for the 118-bus case. Remarkably, the difference in terms of mutual information between the case in which $\rho = 0.1$ and $\rho = 0.9$ increases as λ increases, which indicates that the cost in terms of mutual information of reducing the probability of detection is large in the small values of correlation.

The performance of the upper bound given by Theorem 4.3 on the probability of detection for different values of λ and ρ when $\tau = 2$ and SNR $= 10$ dB is shown in Fig. 4.5. Similarly, Fig. 4.6 depicts the upper bound with the same parameters but with SNR $= 20$ dB. As shown by Theorem 4.3 the bound decreases exponentially fast for large values of λ. Still, there is a significant gap in the probability of attack detection evaluated numerically. This is partially due to the fact that our bound is based on the concentration inequality in Lemma 4.2 which introduces a gap of more than an order of magnitude. Interestingly, the gap decreases when the value of ρ increases although the change is not significant. More importantly, the bound is tighter for lower values of SNR for both 30-bus and 118-bus systems.

4.3.2 MMSE Degradation Induced by Generalized Stealth Attacks

As stated in Section 3.2.1, the minimization of the mutual information between the state variables and the measurements leads to an increase in the MMSE of the estimation. Fig. 4.7 and Fig. 4.8 depict the MMSE degradation induced by the generalized stealth attacks on IEEE 14-Bus test system and IEEE 30-Bus test system when $\rho = 0.1$ and $\rho = 0.9$ for different values of SNR and λ, in which the setting is

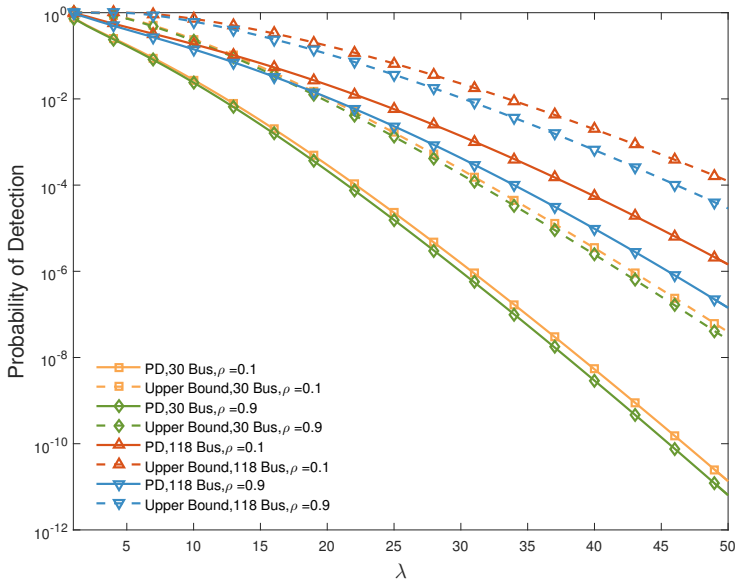


Fig. 4.5. Upper bound on probability of detection given in Theorem 4.3 for different values of λ when $\rho = 0.1$ or 0.9, SNR $= 10$ dB, and $\tau = 2$.

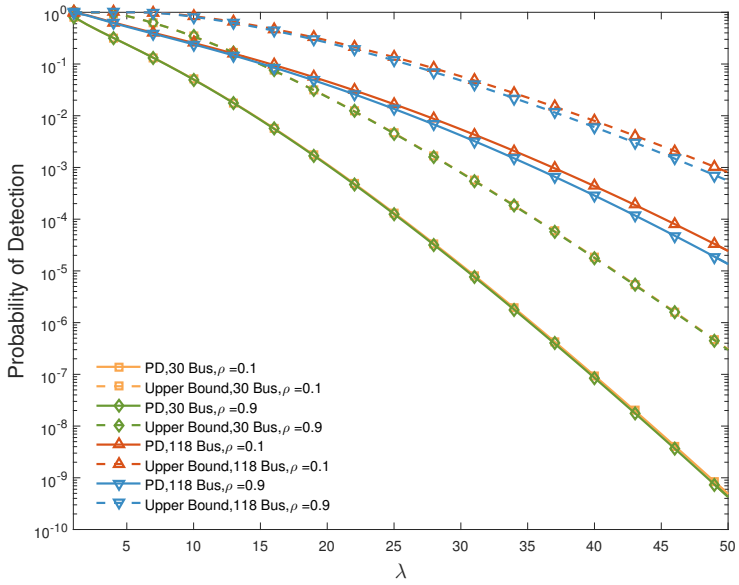


Fig. 4.6. Upper bound on probability of detection given in Theorem 4.3 for different values of λ when $\rho = 0.1$ or 0.9, SNR $= 20$ dB, and $\tau = 2$.

the same as the setting in Section 3.4.2. It is shown that the MMSE degradation is a monotonically decreasing function of λ, adding the fact that the mutual information is a monotonically increasing function of λ, this verifies the conclusion that the minimization of mutual information leads to an increase in MMSE. Comparing Fig. 4.7 and Fig. 4.8 with Fig. 4.3 and Fig. 4.4, it is found that although the mutual information between the state variables and the compromised measurements increases linearly with $\log \lambda$, the MMSE degradation induced by the attacks decreases faster when λ is of small and medium values and starts to flat when λ continues to increase.

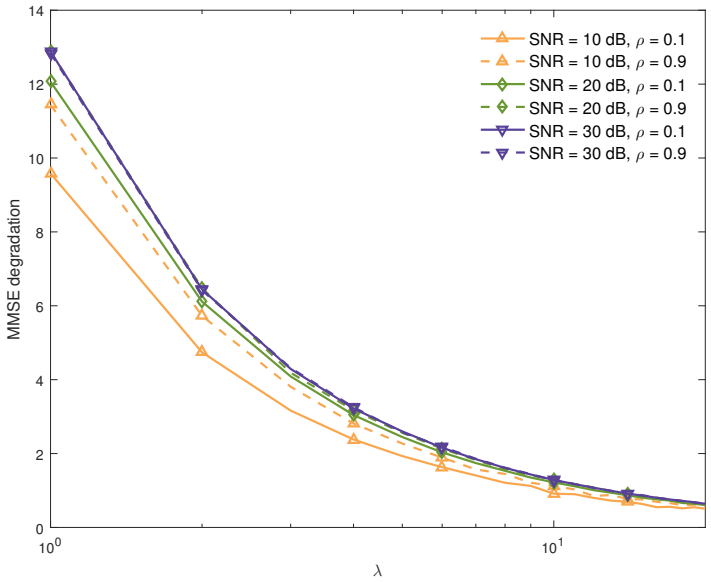


Fig. 4.7. MMSE degradation induced by the generalized stealth attacks on IEEE 14-Bus test system when $\rho = 0.1$ and $\rho = 0.9$ for different values of SNR and λ.

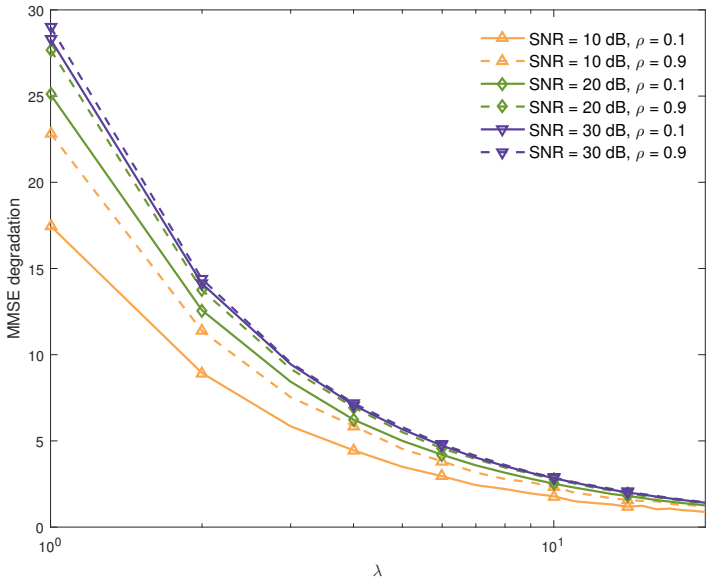


Fig. 4.8. MMSE degradation induced by the generalized stealth attacks on IEEE 30-Bus test system when $\rho = 0.1$ and $\rho = 0.9$ for different values of SNR and λ.

So when the attacker chooses a large value for λ, the performance of the attacks on MMSE is not as good as the performance of the attacks on mutual information.

4.3.3 Sensitivity of Attacks to System Information

As shown in Theorem 4.1 the construction of generalized stealth attacks requires knowledge of the linearized Jacobian matrix $\mathbf{H}$. In practical settings, it is reasonable to assume that the attacker only has access to imperfect system information and that

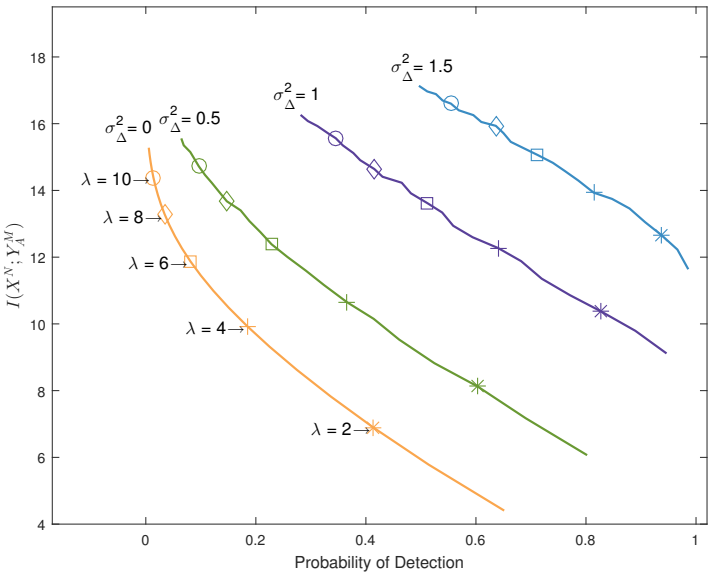


Fig. 4.9. Performance of generalized stealth attack in terms of mutual information and probability of detection for different values of σ^2_Δ and λ on IEEE 14-Bus system when $\rho = 0.1$, $\tau = 2$, and SNR = 20 dB. The marker represents the same value of λ is used in the attack construction.

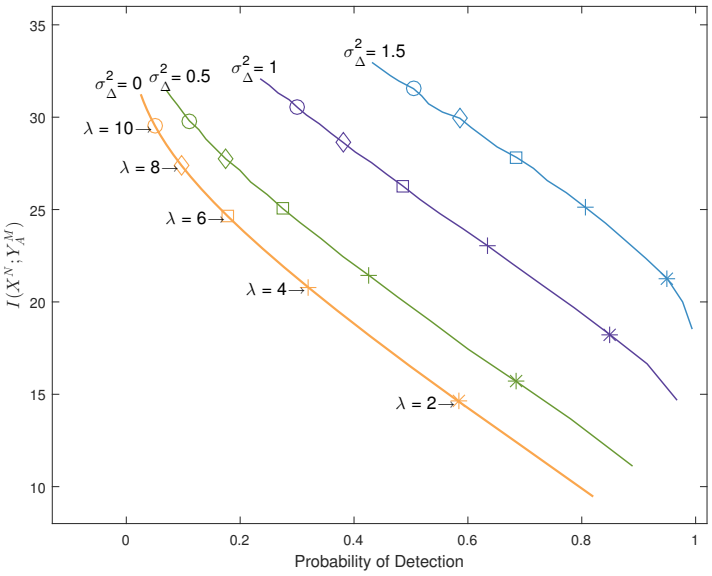


Fig. 4.10. Performance of generalized stealth attack in terms of mutual information and probability of detection for different values of σ^2_Δ and λ on IEEE 30-Bus system when $\rho = 0.1$, $\tau = 2$, and SNR = 20 dB. The marker represents the same value of λ is used in the attack construction.

the Jacobian matrix available during the attack construction is not the real one. For that reason, in the following we numerically analyze the attack performance when the Jacobian matrix is not perfectly known by the attacker and instead a postulated mismatched Jacobian matrix is employed by the attacker. Specifically, we model the Jacobian matrix available to the attacker as $\mathbf{H} + \Delta\mathbf{H}$ where $\Delta\mathbf{H}$ is a matrix

modeling the mismatch introduced by imperfect system information. Uncertainty about the operating point and the dynamics of the power system suggest that it is reasonable to assume a random mismatch framework. That being the case, we model the mismatch as a random matrix with entries given by

$$(\Delta\mathbf{H})_{i,j} = \begin{cases} \Delta, & \text{for } (\mathbf{H})_{i,j} \neq 0 \\ 0, & \text{otherwise,} \end{cases} \tag{4.57}$$

for $i = 1, \ldots, m$, and $j = 1, \ldots, n$, where each entry of the matrix are determined by the independent random variables $\Delta \sim \mathcal{N}(0, \sigma_\Delta^2)$.

Fig. 4.9 depicts the performance of the generalized stealth attack in terms of the mutual information and the probability of detection for different values of σ_Δ^2 and λ on the IEEE 14-Bus system when $\rho = 0.1$, $\tau = 2$, and SNR $= 20$ dB. We generate 100 realizations of the mismatched Jacobian matrix per point and for each realization of the Jacobian matrix we evaluate 1000 realizations of the state variables. The curve corresponding to the perfect Jacobian matrix case, i.e. $\sigma_\Delta^2 = 0$, describes the Pareto optimal front. As expected, when the mismatch of the Jacobian matrix that is available to the attacker increases the performance decreases and moves away from the Pareto front. Interestingly, the performance decrease is smooth and the shape of the curve does not change significantly, which suggests that the tradeoff behavior between the disruption introduced by the attacker and the probability of detection is maintained in the mismatched case. It is also worth noticing that, with larger values of mismatch, the probability of detection increases faster than the mutual information decreases. In view of this, it seems that the stealth of the attack is more severely impacted by the imperfect system information than the disruption. Fig. 4.10 depicts the performance on the IEEE 30-Bus system with the same parameters. Comparing 14-Bus system and 30-Bus system, it easy to see that the performance of the attacker, both mutual information and probability of detection, decreases when mismatch, i.e. σ_Δ^2, increases. But the decrease of performance on IEEE 30-Bus test system is more slowly than that in IEEE 14-Bus test for both the mutual information and the probability of detection. This implies that the uncertainty of $\mathbf{H}$ affects the larger network less, i.e. larger networks exhibit more robust attack construction scenarios and the construction of stealth attacks is simpler in larger networks.

4.3.4 Performance and Sensitivity under AC State Estimation

In the AC state estimation case the iterative estimation methods require a nominal operation point that is updated for each iteration. When the attacker has perfect

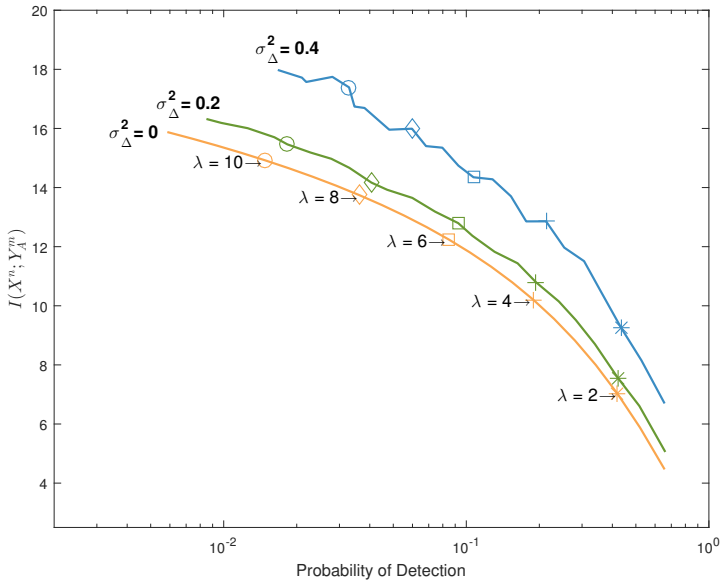


Fig. 4.11. Performance of generalized stealth attack in terms of mutual information and probability of detection for different values of $\widetilde{\sigma}^2_\Delta$ and λ on IEEE 14-Bus system when $\rho = 0.1$, $\tau = 2$, and SNR $= 20$ dB. The marker represents the same value of λ is used in the attack construction.

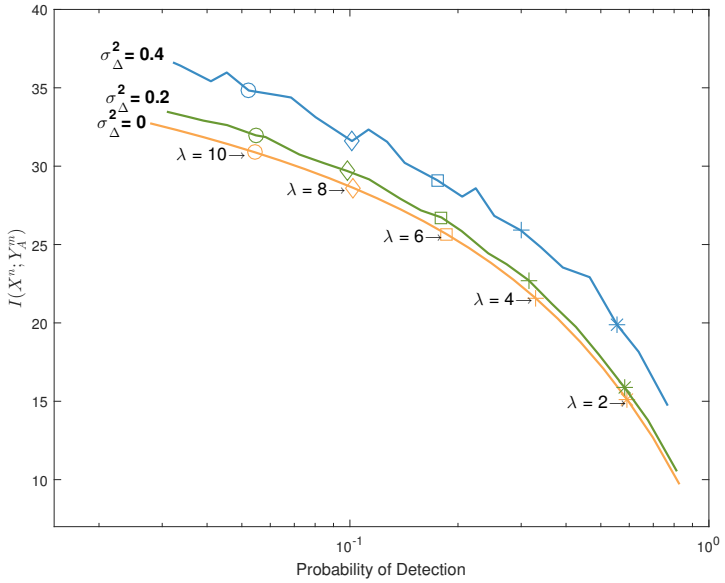


Fig. 4.12. Performance of generalized stealth attack in terms of mutual information and probability of detection for different values of $\widetilde{\sigma}^2_\Delta$ and λ on IEEE 30-Bus system when $\rho = 0.1$, $\tau = 2$, and SNR $= 20$ dB. The marker represents the same value of λ is used in the attack construction.

information about the operation point in each iteration, i.e. perfect information about Jacobian matrix $\mathbf{H}$ in each iteration, the resulting mutual information and probability of detection follow from Corollary 4.1 and Theorem 4.2 directly. *In the following, we study the impact of imperfect nominal operation point information on*

the attack performance. In particular, the generalized stealth attacks are constructed as $A_0^m \sim \mathcal{N}(\mathbf{0}, \frac{1}{\lambda}\mathbf{H}_0\mathbf{\Sigma}_{XX}\mathbf{H}_0^{\mathrm{T}})$, where $\mathbf{H}_0$ is the Jacobian matrix at the nominal operation point $\mathbf{x}_0$ and is given by

$$\mathbf{H}_0 = \frac{\partial}{\partial X^n} H(X^n)^m|_{X^n=\mathbf{x}_0}, \tag{4.58}$$

with $H(X^n) \in \mathbb{R}^m$ denoting the vector of random variables induced by the nonlinear relation between the state variables and the measurements. To model the imperfect knowledge of the nominal point, the nominal linearization point is perturbed with random variable $\widetilde{\Delta} \sim \mathcal{N}(\mathbf{0}, \widetilde{\sigma}_\Delta^2\mathbf{I})$ resulting in the Jacobian matrix $\tilde{\mathbf{H}}$ given by

$$\tilde{\mathbf{H}} = \frac{\partial}{\partial X^n} H(X^n)^m|_{X^n=\mathbf{x}_0+\widetilde{\Delta}}. \tag{4.59}$$

Note that the introduction of this random perturbation gives us a way to control the strength of the perturbation, i.e. the uncertainty over the nominal linearization point, and as a result, we study the sensitivity of the attacks under AC state estimation by changing the variance $\widetilde{\sigma}_\Delta^2$ in the simulations.

Fig. 4.11 depicts the performance of the generalized stealth attacks in terms of the mutual information and the probability of detection for different values of σ_Δ^2 and λ on the IEEE 14-Bus system when $\rho = 0.1$, $\tau = 2$, and SNR = 20 dB. Similarly Fig. 4.12 shows the performance of the attacks under the same setting on IEEE 30-Bus system. We generate 200 realizations of $\widetilde{\Delta}$ per point and for each realization of $\widetilde{\Delta}$ we evaluate 2000 realizations of the state variables. The curve corresponding to the case when $\widetilde{\sigma}_\Delta^2 = 0$ describes the performance of the attacks with perfect knowledge of the nominal operation point. As expected, when there is less accurate knowledge about the nominal operation point, i.e. $\widetilde{\sigma}_\Delta^2$ increases, the performance of the attack A_0^m decreases. Interestingly the performance decrease translates in a larger value of mutual information for all cases. However, the change in probability of detection is not as significant, to the extent that in some cases the probability of detection decreases. Note that this is different from the results in Section 4.3.3, in which both the mutual information and probability of detection decrease simultaneously as a result of imperfect system information. For all cases, overall the attack performance decreases when perfect operation point is not available. Interestingly, the stealth of the attacks is more robust for the IEEE 30-Bus system than for the IEEE 14-Bus system, which suggests that the attacker is better positioned to cope with system uncertainty for larger networks.

4.4 Summary

The stealth attacks in Chapter 3 are generalized by adding a weighting parameter to the KL divergence term that represents the probability of detection in the objective of stealth attacks. The introducing of the weighting parameter allows the attacker taking different tradeoff strategies between the two contradictive objectives, i.e. mutual information and probability of detection. The closed-form expression is proposed for the generalized stealth attacks when the attacker prioritizes the probability of detection over the caused disruption, i.e. the weighted parameter is larger than or equal to one. Changing the value of the weighting parameter allows the closed-form solution moves along the Pareto front between the mutual information and probability of detection. Closed-form expression is also proposed for the probability of detection of the generalized stealth attacks, but it provides no insight into the relation between the probability of detection and the weighting parameter. To that end, a concentration inequality upper bound for the probability of detection is proposed to provide a guideline for the attacker to choose the weighting parameter.

Chapter 5

Learning Requirements for Stealth Attacks

Both the stealth attacks in Theorem 3.1 and the generalized stealth attacks in Theorem 4.1 require the Jacobian matrix and the second order statistics of the state variables, i.e. $\mathbf{H}$ and $\mathbf{\Sigma}_{XX}$, to construct the attacks. In Section 4.3.3 and Section 4.3.4, we numerically evaluate the impact of having access to an imperfect Jacobian matrix on the performance of the attacks. Therein, we show that the attacks constructed using an imperfect Jacobian matrix impose performance degradation on both mutual information and probability of detection. In this chapter, we analyze the impact of imperfect knowledge of the second order statistics of the state variables on the performance of the attacks. Specifically, we focus on the scenario that the attacker only gets access to a limited number of samples of the state variables, and estimates the second order statistics of the state variables via the sample covariance matrix of the samples. Here we use RMT tools to characterize the performance of the attacks constructed using the sample covariance matrix in substitution of the second order moments for the non-asymptotic case and the asymptotic case.

5.1 Stealth Attacks Using Imperfect Second Order Statistics

5.1.1 Statistical Learning Setting

Theorem 3.1 shows that the stealth attack construction is given by $\mathbf{\Sigma}^{\star}_{AA} = \mathbf{H}\mathbf{\Sigma}_{XX}\mathbf{H}^{\mathrm{T}}$, which implies that the attacker needs the Jacobian matrix, i.e. $\mathbf{H}$, and the covariance

The non-asymptotic part in Chapter 5 is published in "K. Sun, I. Esnaola, A.M Tulino and H.V. Poor, "Learning requirements for stealth attacks", in *Proc. IEEE Int. Conf. on Acoust., Speech, and Signal Process.*, Brighton, UK, May 2019, pp. 8102-8106.".

matrix of the state variables, i.e. $\mathbf{\Sigma}_{XX}$, to construct attacks. In the following, we study the performance of the attack when the second order statistics are not perfectly known by the attacker but the linearized Jacobian measurement matrix is known. We model the partial knowledge of the attacker about the covariance matrix by assuming that the attacker has access to a given number of samples of the state variables. Specifically, we assume that the training data set consisting of k state variable realizations $\{X_i^n\}_{i=1}^k$ is available to the attacker. Since the sample covariance matrix is an unbiased estimate of the covariance matrix of the state variables and is asymptotically optimal, the sample covariance matrix is computed to estimate the second order statistics from the training data. That being the case, the sample covariance matrix of k realizations is given by

$$\mathbf{S}_{XX} = \frac{1}{k-1}\sum_{i=1}^{k} X_i^n (X_i^n)^{\mathrm{T}}. \tag{5.1}$$

Given the optimal expression in Theorem 3.1, the stealth attacks constructed using the sample covariance matrix follow a multivariate Gaussian distribution given by

$$\tilde{A}^m \sim \mathcal{N}(\mathbf{0}, \mathbf{\Sigma}_{\tilde{A}\tilde{A}}), \tag{5.2}$$

where $\mathbf{\Sigma}_{\tilde{A}\tilde{A}} = \mathbf{H}\mathbf{S}_{XX}\mathbf{H}^{\mathrm{T}}$ is the covariance matrix of $\tilde{A}^m$.

Recall that the objective of the stealth attacks is to minimize the cost function given by

$$\min_{A^m} D(P_{X^n Y_A^m} \| P_{X^n} P_{Y^m}). \tag{5.3}$$

With the estimated statistics in (5.1), the KL divergence in (5.3) conditioned on the covariance matrix obtained from the training data becomes

$$D(P_{X^n Y_{\tilde{A}}^m | S_{XX}} \| P_{X^n} P_{Y^m} | P_{S_{XX}}), \tag{5.4}$$

where $P_{X^n Y_{\tilde{A}}^m | S_{XX}}$ is the conditional joint distribution of $(X^n, Y_{\tilde{A}}^m)$, in which

$$Y_{\tilde{A}}^m = \mathbf{H}X^n + Z^m + \tilde{A}^m \tag{5.5}$$

and $\mathbf{\Sigma}_{\tilde{A}\tilde{A}} = \mathbf{H}\mathbf{S}_{XX}\mathbf{H}^{\mathrm{T}}$; and $P_{S_{XX}}$ is the distribution of S_{XX}.

5.1.2 Suboptimality of the Learning Attacks

The following lemma shows that the objective function in (5.3) for exact statistics is a lower bound on the KL divergence conditioned on the training data given by (5.4).

Before introducing the lower bound, Jensen's inequality is reviewed first to aid the later proof.

Lemma 5.1. *[22, Theorem 2.6.2] If f is a convex function and X is a random variable, then*

$$\mathbb{E}\left[f(X)\right] \geq f\left(\mathbb{E}\left[X\right]\right). \tag{5.6}$$

If f is strictly convex, the equality in (5.6) implies that $X = \mathbb{E}\left[X\right]$ with probability 1.

Using Jensen's inequality obtains the following lemma.

Lemma 5.2. *The conditional divergence for the attack vector construction with covariance matrix $\mathbf{\Sigma}_{\tilde{A}\tilde{A}} = \mathbf{H}\mathbf{S}_{XX}\mathbf{H}^{\mathrm{T}}$ in (5.4) is lower bounded by the divergence in (5.3) with $\mathbf{\Sigma}^{\star}_{AA} = \mathbf{H}\mathbf{\Sigma}_{XX}\mathbf{H}^{\mathrm{T}}$, that is*

$$D(P_{X^nY^m_{\tilde{A}}|S_{XX}}\|Q_{X^nY^m}|P_{S_{XX}}) \geq D(P_{X^nY^m_{A^\star}}\|Q_{X^nY^m}), \tag{5.7}$$

where $P_{X^nY^m_{A^\star}}$ is the joint distribution of $(X^n, Y^m_{A^\star})$ when the optimal attack is constructed, and $Q_{X^nY^m} = P_{X^n}P_{Y^m}$.

Proof. We have that

$$\begin{aligned}
&D\left(P_{X^nY^m_{\tilde{A}}|S_{XX}}\|Q_{X^nY^m}|P_{S_{XX}}\right)\\
&= D\left(P_{X^nY^m_{\tilde{A}}|S_{XX}}\|Q_{X^nY^m|S_{XX}}|P_{S_{XX}}\right) && (5.8)\\
&= \mathbb{E}_{S_{XX}}\left[D(P_{X^nY^m_{\tilde{A}}|S_{XX}=S}\|Q_{X^nY^m|S_{XX}=S})\right] && (5.9)\\
&= \frac{1}{2}\mathbb{E}_{S_{XX}}\left[\mathrm{tr}(\mathbf{\Sigma}^{-1}_{YY}\mathbf{\Sigma}_{\tilde{A}\tilde{A}})\right] - \frac{1}{2}\mathbb{E}_{S_{XX}}\left[\log|\mathbf{\Sigma}_{\tilde{A}\tilde{A}} + \sigma^2\mathbf{I}_m|\right] - \frac{1}{2}\log|\mathbf{\Sigma}^{-1}_{YY}| && (5.10)\\
&\geq \frac{1}{2}\mathrm{tr}\left(\mathbf{\Sigma}^{-1}_{YY}\mathbf{\Sigma}^{\star}_{AA}\right) - \frac{1}{2}\log\left|\mathbf{\Sigma}^{\star}_{AA} + \sigma^2\mathbf{I}_m\right| - \frac{1}{2}\log\left|\mathbf{\Sigma}^{-1}_{YY}\right| && (5.11)\\
&= D\left(P_{X^nY^m_{A^\star}}\|Q_{X^nY^m}\right), && (5.12)
\end{aligned}$$

where (5.8) follows from the independence of X^n and Y^m with respect to S_{XX}; (5.9) follows from the definition of conditional divergence in Appendix A.11; (5.10) follows from taking the Gaussianity of $(X^n, Y^m_{\tilde{A}})$, i.e.

$$(X^n, Y^m_{\tilde{A}}) \sim \mathcal{N}\left(\mathbf{0}, \begin{bmatrix} \mathbf{\Sigma}_{XX} & \mathbf{\Sigma}_{XX}\mathbf{H}^{\mathrm{T}} \\ \mathbf{H}\mathbf{\Sigma}_{XX} & \mathbf{H}\mathbf{\Sigma}_{XX}\mathbf{H}^{\mathrm{T}} + \sigma^2\mathbf{I}_m + \mathbf{\Sigma}_{\tilde{A}\tilde{A}} \end{bmatrix}\right), \tag{5.13}$$

and the Gaussianity of $Q_{X^nY^m}$ into Proposition 3.2; and (5.11) follows from Jensen's inequality and the fact that $-\log|\mathbf{V}|$ is a convex function of $\mathbf{V} \in \mathcal{S}^m_+$. The proof completes. □

Lemma 5.2 shows that the KL divergence achieved by the attack conditioned on the training data is higher than the performance of the attack construction with exact statistics. This implies that the attacks constructed using sample covariance matrix are always *suboptimal*. However, the performance of the attack constructed by the sample covariance matrix converges asymptotically in k to that of the attack constructed by the exact covariance matrix. Therefore, the speed of convergence needs to be characterized to analyze the performance of the attacks constructed using the sample covariance matrix.

The sample covariance matrix of samples from zero mean multivariate Gaussian distribution follows a central Wishart distribution, c.f. (B.2). As a result, the sample covariance matrix in (5.1) is a random matrix with central Wishart distribution given by

$$\mathbf{S}_{XX} \sim \frac{1}{k-1} W_n(k-1, \mathbf{\Sigma}_{XX}), \tag{5.14}$$

the ergodic counterpart of the cost function in (5.3) is defined in terms of the conditional KL divergence given by

$$\mathbb{E}_{\mathbf{S}_{XX}} \left[D\left(P_{X^n Y_A^m | \mathbf{S}_{XX}} \| P_{X^n} P_{Y^m} \right) \right]. \tag{5.15}$$

The ergodic cost function characterizes the expected performance of the attack averaged over the realizations of the training data. In the next section, we introduce an upper bound for the ergodic performance of attacks constructed using sample covariance matrix under the non-asymptotic setting.

5.2 Bounds for Non-asymptotic Ergodic Performance

As just stated before, the sample covariance matrix follows the central Wishart distribution gave in (5.14), and therefore, the cost function in (5.3) is a random variable. Here we focus on the ergodic performance, i.e. the expected value of the distribution, which describes the expected performance of the attack constructed using sample covariance matrix. In the following, RMT tools are used to characterize the distribution of the performance in (5.3) for the non-asymptotic case and the asymptotic case. As covered in Appendix B, the asymptotic analysis of RMT shows the limiting distribution of a function of the eigenvalues of random matrix when the dimensions of the random matrix go to infinity with a given ratio between the

dimensions. On the other hand, the non-asymptotic analysis of RMT focuses on the condition that the dimension and number of samples are of finite values.

In this section, we focus on the non-asymptotic scenario. Specifically RMT tools are utilized to propose an upper bound for the ergodic performance of the attacks using the sample covariance matrix and study it within the non-asymptotic scenario. The proposed upper bound converges to the ergodic performance when the number of samples increases. Adding the fact that the performance of the attacks using sample covariance matrix is lower bounded by the performance of the attacks using the exact covariance matrix, i.e. Lemma 5.2, the proposed upper bound and the performance of the attacks using the exact covariance matrix regulates the ergodic performance of the attacks.

In the next subsection, some auxiliary results for Wishart matrices are presented to aid the derivation of the upper bound later.

5.2.1 Auxiliary Non-asymptotic Results using RMT

The standard Gaussian matrix is a matrix whose entries are independent standard normal random variables, and therefore, the maximum singular value of the standard Gaussian matrix is a random variable. The following lemma shows that under certain moment constraints the maximum singular value of the Gaussian matrix is a sub-gaussian random variable with variance smaller than one.

Before we introduce the lemma, two auxiliary definitions are presented.

Definition 5.1. *A function* $f : \mathbb{R}^{(k-1)\times l} \to \mathbb{R}$ *is a* C*-Lipschitz function when*

$$|f(\mathbf{A}) - f(\mathbf{B})| \leq C\|\mathbf{A} - \mathbf{B}\|^2 \tag{5.16}$$

holds for any matrix $\mathbf{A}$ *and* $\mathbf{B}$ *in the domain of* f*.*

Definition 5.2. *A random variable* X *is said to be sub-gaussian with variance proxy* σ_p^2 *if* $\mathbb{E}[X] = 0$ *and it satisfies*

$$P[X > t] \leq \exp\left(-\frac{t^2}{2\sigma_p^2}\right) \tag{5.17}$$

for all $t \geq 0$*.*

The following lemma proves the sub-gaussianity of the maximum singular value of the standard Gaussian matrix.

Lemma 5.3. *Let $\mathbf{Z}_l$ be a $(k-1)\times l$ matrix whose entries are independent standard normal random variables, then*

$$\operatorname{var}\left(s_{\max}(\mathbf{Z}_l)\right) \leq 1, \tag{5.18}$$

where $\operatorname{var}(\cdot)$ *denotes the variance and* $s_{\max}(\mathbf{Z}_l)$ *is the maximum singular value of* $\mathbf{Z}_l$*.*

Proof. Note that $s_{\max}(\mathbf{Z}_l)$ is a 1-Lipschitz function of matrix $\mathbf{Z}_l$, the maximum singular value of $\mathbf{Z}_l$ is concentrated around the mean given by $\mathbb{E}[s_{\max}(\mathbf{Z}_l)]$ [128, Proposition 5.34] . Then for $t \geq 0$, it holds that

$$\mathbb{P}\left[s_{\max}(\mathbf{Z}_l) - \mathbb{E}[s_{\max}(\mathbf{Z}_l)] > t\right] \leq \exp\{-t^2/2\}. \tag{5.19}$$

Therefore $s_{\max}(\mathbf{Z}_l) - \mathbb{E}[s_{\max}(\mathbf{Z}_l)]$ is a sub-gaussian random variable with variance proxy $\sigma_p^2 \leq 1$. The lemma follows from the fact that the variance of a zero-mean sub-gaussian random variable is smaller than the variance proxy of this random variable, i.e. $\operatorname{var}\left(s_{\max}(\mathbf{Z}_l)\right) \leq \sigma_p^2$. □

Note that a standard central Wishart random matrix is the product of a Gaussian random matrix and its transpose. So a given eigenvalue of the standard central Wishart matrix is equivalent to the square of the corresponding singular value of the standard Gaussian matrix. The following lemma provides bounds for the expected values of the eigenvalues of standard central Wishart matrix.

Lemma 5.4. *Let $\mathbf{W}_l$ denote a central Wishart matrix distributed as $\frac{1}{k-1}W_l(k-1,\mathbf{I}_l)$, then the non-asymptotic expected value of the extreme eigenvalues of $\mathbf{W}_l$ are bounded by*

$$\left(1-\sqrt{\frac{l}{k-1}}\right)^2 \leq \mathbb{E}\left[\lambda_{\min}(\mathbf{W}_l)\right] \tag{5.20}$$

and

$$\mathbb{E}\left[\lambda_{\max}(\mathbf{W}_l)\right] \leq \left(1+\sqrt{\frac{l}{k-1}}\right)^2 + \frac{1}{k-1}, \tag{5.21}$$

where $\lambda_{\min}(\mathbf{W}_l)$ and $\lambda_{\max}(\mathbf{W}_l)$ denote the minimum eigenvalue and maximum eigenvalue of $\mathbf{W}_l$, respectively.

Proof. Note that [128, Theorem 5.32]

$$\sqrt{k-1} - \sqrt{l} \leq \mathbb{E}[s_{\min}(\mathbf{Z}_l)] \tag{5.22}$$

and

$$\sqrt{k-1}+\sqrt{l} \geq \mathbb{E}[s_{\max}(\mathbf{Z}_l)], \tag{5.23}$$

where $s_{\min}(\mathbf{Z}_l)$ is the minimum singular value of $\mathbf{Z}_l$. Given the fact that $\mathbf{W}_l = \frac{1}{k-1}\mathbf{Z}_l^{\mathrm{T}}\mathbf{Z}_l$, then it holds that

$$\begin{aligned}
\mathbb{E}[\lambda_{\min}(\mathbf{W}_l)] &= \frac{\mathbb{E}\left[s_{\min}(\mathbf{Z}_l)^2\right]}{k-1} && (5.24)\\
&= \frac{\mathbb{E}\left[s_{\min}(\mathbf{Z}_l)\right]^2 + \mathrm{var}(s_{\min}(\mathbf{Z}_l))}{k-1} && (5.25)\\
&\geq \frac{\mathbb{E}\left[s_{\min}(\mathbf{Z}_l)\right]^2}{k-1} && (5.26)
\end{aligned}$$

and

$$\begin{aligned}
\mathbb{E}[\lambda_{\max}(\mathbf{W}_l)] &= \frac{\mathbb{E}\left[s_{\max}(\mathbf{Z}_l)^2\right]}{k-1} && (5.27)\\
&= \frac{\mathbb{E}[s_{\max}(\mathbf{Z}_l)]^2 + \mathrm{var}(s_{\max}(\mathbf{Z}_l))}{k-1} && (5.28)\\
&\leq \frac{\mathbb{E}\left[s_{\max}(\mathbf{Z}_l)\right]^2 + 1}{k-1}, && (5.29)
\end{aligned}$$

where (5.29) follows from Lemma 5.3. Combining (5.22) with (5.26), and (5.23) with (5.29), respectively, yields the lemma. □

5.2.2 Upper Bound for Non-asymptotic Ergodic Performance

The ergodic attack performance given in (5.15) is expanded as

$$\begin{aligned}
\mathbb{E}\left[f(\mathbf{\Sigma}_{\tilde{A}\tilde{A}})\right] &= \frac{1}{2}\mathbb{E}\left[\mathrm{tr}(\mathbf{\Sigma}_{YY}^{-1}\mathbf{\Sigma}_{\tilde{A}\tilde{A}}) - \log\left|\mathbf{\Sigma}_{\tilde{A}\tilde{A}} + \sigma^2\mathbf{I}_m\right| - \log\left|\mathbf{\Sigma}_{YY}^{-1}\right|\right]\\
&= \frac{1}{2}\Big(\mathrm{tr}\left(\mathbf{\Sigma}_{YY}^{-1}\mathbf{\Sigma}_{AA}^{\star}\right) - \log\left|\mathbf{\Sigma}_{YY}^{-1}\right| - \mathbb{E}\left[\log\left|\mathbf{\Sigma}_{\tilde{A}\tilde{A}} + \sigma^2\mathbf{I}_m\right|\right]\Big), \quad (5.30)
\end{aligned}$$

where (5.30) follows from the fact that $\mathbb{E}\left[\mathrm{tr}(\mathbf{\Sigma}_{YY}^{-1}\mathbf{\Sigma}_{\tilde{A}\tilde{A}})\right] = \mathrm{tr}(\mathbf{\Sigma}_{YY}^{-1}\mathbf{\Sigma}_{AA}^{\star})$ due to the linearity of the trace operator, see (B.3). The assessment of the ergodic attack performance boils down to evaluating the last term in (5.30). Closed-form expressions for this term are provided in [129] for the same case considered in this chapter. However, the resulting expressions are involved and are only computable for small dimensional settings. For systems with a large number of dimensions, such as power systems, the expressions are computationally prohibitive. To circumvent this challenge we propose a lower bound on the last term that yields an upper bound on the ergodic attack performance.

Before presenting the lower bound, we provide the following auxiliary convex optimization result, in which a lower bound is proposed for the expected value of the logarithm of the determinant of an identity matrix plus the inverse of a standard Wishart matrix.

Lemma 5.5. *Let* $\mathbf{W}_p$ *denote a central Wishart matrix distributed as* $\frac{1}{k-1}W_p(k-1,\mathbf{I}_p)$ *and let* $\mathbf{B}=\mathrm{diag}(b_1,\ldots,b_p)$ *denote a positive definite diagonal matrix. Then*

$$\mathbb{E}\left[\log\left|\mathbf{B}+\mathbf{W}_p^{-1}\right|\right]\geq\sum_{i=1}^{p}\log\left(b_i+1/x_i^{\star}\right), \tag{5.31}$$

where $x_i^{\star}$ *is the solution to the convex optimization problem given by*

$$\min_{\{x_i\}_{i=1}^p}\ \sum_{i=1}^{p}\log\left(b_i+1/x_i\right) \tag{5.32}$$

$$\text{s.t.}\quad \sum_{i=1}^{p}x_i=p \tag{5.33}$$

$$\max\left(x_i\right)\leq\left(1+\sqrt{p/(k-1)}\right)^2+1/(k-1) \tag{5.34}$$

$$\min\left(x_i\right)\geq\left(1-\sqrt{p/(k-1)}\right)^2. \tag{5.35}$$

Proof. Note that

$$\mathbb{E}\left[\log\left|\mathbf{B}+\mathbf{W}_p^{-1}\right|\right]=\sum_{i=1}^{p}\mathbb{E}\left[\log\left(b_i+\frac{1}{\lambda_i(\mathbf{W}_p)}\right)\right] \tag{5.36}$$

$$\geq\sum_{i=1}^{p}\log\left(b_i+\frac{1}{\mathbb{E}[\lambda_i(\mathbf{W}_p)]}\right) \tag{5.37}$$

where in (5.36), $\lambda_i(\mathbf{W}_p)$ is the i-th eigenvalue of $\mathbf{W}_p$ in decreasing order; (5.37) follows from Jensen's inequality due to the convexity of $\log\left(b_i+\frac{1}{x}\right)$ for $x>0$ when $b_i>0$.

Finding the minimum value for the expression in (5.37) yields a lower bound for the left-hand side expression of (5.36), which is formulated as the objective in (5.32). Constraint (5.33) follows from the fact that $\mathbb{E}[\mathrm{trace}(\mathbf{W}_p)]=p$, and constraints (5.34) and (5.35) follow from Lemma 5.4. This completes the proof.

□

The following theorem provides a lower bound for the last term in (5.30), and therefore, it enables us to characterize the ergodic attack performance.

Theorem 5.1. *Let* $\mathbf{\Sigma}_{\tilde{A}\tilde{A}}=\mathbf{H}\mathbf{S}_{XX}\mathbf{H}^T$ *with* $\mathbf{S}_{XX}$ *distributed as* $\frac{1}{k-1}W_n(k-1,\mathbf{\Sigma}_{XX})$ *and denote by* $\mathbf{\Lambda}_p=\mathrm{diag}(\lambda_1,\ldots,\lambda_p)$ *the diagonal matrix containing the nonzero*

eigenvalues of $\mathbf{H}\boldsymbol{\Sigma}_{XX}\mathbf{H}^T$ *in decreasing order. Then*

$$\mathbb{E}\left[\log|\boldsymbol{\Sigma}_{\tilde{A}\tilde{A}}+\sigma^2\mathbf{I}_m|\right] \geq \left(\sum_{i=0}^{p-1}\psi(k-1-i)\right)-p\log(k-1)+\sum_{i=1}^{p}\log\left(\frac{\lambda_i}{\sigma^2}+\frac{1}{\lambda_i^\star}\right)+2m\log\sigma, \quad (5.38)$$

where $\psi(\cdot)$ *is the Euler digamma function,* $p=\operatorname{rank}(\mathbf{H}\boldsymbol{\Sigma}_{XX}\mathbf{H}^T)$*, and* $\{\lambda_i^\star\}_{i=1}^p$ *is the solution to the optimization problem given by (5.32) - (5.35) with* $b_i=\frac{\lambda_i}{\sigma^2}$*,* for $i=1,\ldots,p$*.*

Proof. We proceed by noticing that

$$\begin{aligned}
&\mathbb{E}\left[\log|\boldsymbol{\Sigma}_{\tilde{A}\tilde{A}}+\sigma^2\mathbf{I}_m|\right]\\
&=\mathbb{E}\left[\log\left|\frac{1}{\sigma^2}\boldsymbol{\Lambda}_s^{\frac{1}{2}}\frac{\mathbf{Z}_m^{\mathrm{T}}\mathbf{Z}_m}{k-1}\boldsymbol{\Lambda}_s^{\frac{1}{2}}+\mathbf{I}_m\right|\right]+2m\log\sigma && (5.39)\\
&=\mathbb{E}\left[\log\left|\frac{1}{\sigma^2}\boldsymbol{\Lambda}_s\frac{\mathbf{Z}_m^{\mathrm{T}}\mathbf{Z}_m}{k-1}+\mathbf{I}_m\right|\right]+2m\log\sigma && (5.40)\\
&=\mathbb{E}\left[\log\left|\frac{\boldsymbol{\Lambda}_p}{\sigma^2}\frac{\mathbf{Z}_p^{\mathrm{T}}\mathbf{Z}_p}{k-1}+\mathbf{I}_p\right|\right]+2m\log\sigma && (5.41)\\
&=\mathbb{E}\left[\log\left|\frac{\mathbf{Z}_p^{\mathrm{T}}\mathbf{Z}_p}{k-1}\right|+\log\left|\frac{\boldsymbol{\Lambda}_p}{\sigma^2}+\left(\frac{\mathbf{Z}_p^{\mathrm{T}}\mathbf{Z}_p}{k-1}\right)^{-1}\right|\right]+2m\log\sigma && (5.42)\\
&\geq\left(\sum_{i=0}^{p-1}\psi(k-1-i)\right)-p\log(k-1)+\sum_{i=1}^{p}\log\left(\frac{\lambda_i}{\sigma^2}+\frac{1}{\lambda_i^\star}\right)+2m\log\sigma, && (5.43)
\end{aligned}$$

where (5.39) follows from (B.3), i.e.

$$\boldsymbol{\Sigma}_{\tilde{A}\tilde{A}}=\mathbf{H}\mathbf{S}_{XX}\mathbf{H}^{\mathrm{T}}\sim\frac{1}{k-1}W_m(k-1,\mathbf{H}\boldsymbol{\Sigma}_{XX}\mathbf{H}^{\mathrm{T}}), \quad (5.44)$$

and

$$\mathbf{H}\mathbf{S}_{XX}\mathbf{H}^{\mathrm{T}}\overset{d}{=}\mathbf{V}\boldsymbol{\Lambda}_s^{\frac{1}{2}}\frac{\mathbf{Z}_m^{\mathrm{T}}\mathbf{Z}_m}{k-1}\boldsymbol{\Lambda}_s^{\frac{1}{2}}\mathbf{V}^{\mathrm{T}}, \quad (5.45)$$

in which $\boldsymbol{\Lambda}_s$ and $\mathbf{V}$ are the matrix of eigenvalues in decreasing order and the unitary matrix of the corresponding eigenvector, respectively, of $\mathbf{H}\boldsymbol{\Sigma}_{XX}\mathbf{H}^{\mathrm{T}}$, and $\overset{d}{=}$ denotes equality in distribution; (5.40) follows from Sylvester's determinant identity, which states that

$$|\mathbf{I}+\mathbf{C}\mathbf{D}|=|\mathbf{I}+\mathbf{D}\mathbf{C}| \quad (5.46)$$

for matrices $\mathbf{C}$ and $\mathbf{D}$ of proper dimensions; (5.41) follows from the fact that $\boldsymbol{\Lambda}_s$ is a rank deficient matrix with rank p; (5.42) follows from the non-singular property of

Wishart matrix when $k-1 \geq p$; (5.43) follows from [130, Theorem 2.11] and Lemma 5.5. This completes the proof. □

Substituting Theorem 5.1 into (5.30) yields the following upper bound for the expected value of the ergodic performance.

Theorem 5.2. *The ergodic attack performance given in (5.30) is upper bounded by*

$$\mathbb{E}\left[f(\boldsymbol{\Sigma}_{\tilde{A}\tilde{A}})\right] \leq \frac{1}{2}\Bigg(\operatorname{tr}\left(\boldsymbol{\Sigma}_{YY}^{-1}\boldsymbol{\Sigma}_{AA}^{\star}\right) - \log\left|\boldsymbol{\Sigma}_{YY}^{-1}\right| - 2m\log\sigma - \left(\sum_{i=0}^{p-1}\psi(k-1-i)\right) + p\log(k-1) - \sum_{i=1}^{p}\log\left(\frac{\lambda_i}{\sigma^2}+\frac{1}{\lambda_i^{\star}}\right)\Bigg). \quad (5.47)$$

Proof. The proof follows immediately from combing Theorem 5.1 with (5.30). □

Using the same approach as in Lemma 5.5 and Theorem 5.1, a lower bound for the performance in (5.30) can be derived. However the simulation on IEEE 30-Bus test system shows that the obtained lower bound is better than the performance of the attack using exact covariance matrix. This goes against the result in Lemma 5.2, which states that the performance of the attacks using sample covariance matrix is worse than the one using exact covariance. The derivation for the lower bound is provided in Appendix D for completion.

5.3 Explicit Expression for Asymptotic Ergodic Performance

In the following section, we analyze the asymptotic performance of the attacks constructed using the sample covariance matrix. By studying the asymptotic setting, we are able to provide a closed-form expression for the asymptotic ergodic attack performance. Before introducing the asymptotic ergodic performance, some auxiliary definitions under asymptotic scenario are provided.

5.3.1 Auxiliary Asymptotic Results from RMT

Definition 5.3. *[130] The η-transform of a nonnegative random variable X is*

$$\eta_X(\gamma) = \mathbb{E}\left[\frac{1}{1+\gamma X}\right], \quad (5.48)$$

where $\gamma \geq 0$ and thus $1 \geq \eta_X(\gamma) > 0$.

Definition 5.4. *[130] The Shannon transform of a nonnegative random variable X is defined as*

$$\mathcal{V}_X(\gamma) = \mathbb{E}\left[\log(1+\gamma X)\right]. \tag{5.49}$$

The Shannon transform is linked to the η-transform by noticing that

$$\frac{\partial}{\partial\gamma}\mathcal{V}_X(\gamma) = \big(1-\eta_X(\gamma)\big)\frac{\log e}{\gamma}. \tag{5.50}$$

Example 5.1. *The Shannon transform of the Marčenko-Pastur law $f_\beta(\cdot)$ in (B.6) is given by*

$$\mathcal{V}(\gamma) = \log\left(1+\gamma-\frac{1}{4}\mathcal{F}\left(\gamma,\frac{1}{\beta}\right)\right)+\beta\log\left(1+\frac{\gamma}{\beta}-\frac{1}{4}\mathcal{F}\left(\gamma,\frac{1}{\beta}\right)\right)-\frac{\beta\log e}{4\gamma}\mathcal{F}\left(\gamma,\frac{1}{\beta}\right) \tag{5.51}$$

with

$$\mathcal{F}(\gamma,\beta) = \left(\sqrt{\gamma\left(1+\sqrt{\frac{1}{\beta}}\right)^2+1}-\sqrt{\gamma\left(1-\sqrt{\frac{1}{\beta}}\right)^2+1}\right)^2. \tag{5.52}$$

Definition 5.5. *[130] The asymptotic/empirical eigenvalue distribution (AED or EED), $\mathsf{F}_{\mathbf{A}}(\cdot)$, of an $n\times n$ Hermitian random matrix $\mathbf{A}$ is defined as*

$$\mathsf{F}_{\mathbf{A}}(x) = \lim_{n\to\infty}\frac{1}{n}\sum_{i=1}^{n}\mathbb{1}_{\{\lambda_i(\mathbf{A})\le x\}}, \tag{5.53}$$

where $\lambda_1(\mathbf{A}),\ldots,\lambda_n(\mathbf{A})$ are the eigenvalues of $\mathbf{A}$.

5.3.2 Explicit Ergodic Performance

Before characterizing the ergodic performance of the attack constructed with the sample covariance matrix in the asymptotic case, the following theorem provides an equivalent distribution for the random variable describing the performance of the attack in (5.30).

Here without loss of generality, we assume that the rank of matrix $\mathbf{H}\boldsymbol{\Sigma}_{XX}\mathbf{H}^{\mathrm{T}}$ is equal to n. The rational of this assumption comes from the observability check set by the operator, which guarantees that $\mathbf{H}$ is a full rank matrix with $m\ge n$ for the state estimation procedure. As a result, it holds that

$$\operatorname{rank}(\mathbf{H}\boldsymbol{\Sigma}_{XX}\mathbf{H}^{\mathrm{T}}) = \operatorname{rank}(\boldsymbol{\Sigma}_{XX}). \tag{5.54}$$

For power systems, the covariance matrix of the state variables is usually of full rank, which implies that any random variable in the vector of state variables is not a linear combination of the other random variables. This also implies that the distribution of any random variable in the vector of state variables is not singular.

Theorem 5.3. *The performance of the attack using the sample covariance matrix is equivalent in distribution to the random variable given by*

$$f(\mathbf{\Sigma}_{\tilde{A}\tilde{A}}) \overset{d}{=} \frac{1}{2}\left(tr\left(\left(\tilde{\mathbf{\Lambda}}+\mathbf{I}_n\right)^{-1}\tilde{\mathbf{\Lambda}}\frac{\mathbf{Z}_n^T\mathbf{Z}_n}{k-1}\right)+\log\left|\tilde{\mathbf{\Lambda}}+\mathbf{I}_n\right|-\log\left|\tilde{\mathbf{\Lambda}}\frac{\mathbf{Z}_n^T\mathbf{Z}_n}{k-1}+\mathbf{I}_n\right|\right), \quad (5.55)$$

where $\tilde{\mathbf{\Lambda}} \triangleq \frac{1}{\sigma^2}\mathbf{\Lambda}_p \in \mathbb{R}^n$.

Proof. Note that

$$\begin{aligned}
&f(\mathbf{\Sigma}_{\tilde{A}\tilde{A}}) \\
&= \frac{1}{2}\left(\mathrm{tr}(\mathbf{\Sigma}_{YY}^{-1}\mathbf{\Sigma}_{\tilde{A}\tilde{A}}) - \log|\mathbf{\Sigma}_{\tilde{A}\tilde{A}}+\sigma^2\mathbf{I}_m| - \log|\mathbf{\Sigma}_{YY}^{-1}|\right) && (5.56)\\
&\overset{d}{=}\frac{1}{2}\left(\mathrm{tr}\left(\mathbf{\Sigma}_{YY}^{-1}\mathbf{V}\mathbf{\Lambda}_s^{\frac{1}{2}}\frac{\mathbf{Z}_m^{\mathrm{T}}\mathbf{Z}_m}{k-1}\mathbf{\Lambda}_s^{\frac{1}{2}}\mathbf{V}^{\mathrm{T}}\right)+\log|\mathbf{\Sigma}_{YY}|-\log\left|\mathbf{V}\mathbf{\Lambda}_s^{\frac{1}{2}}\frac{\mathbf{Z}_m^{\mathrm{T}}\mathbf{Z}_m}{k-1}\mathbf{\Lambda}_s^{\frac{1}{2}}\mathbf{V}^{\mathrm{T}}+\sigma^2\mathbf{I}_m\right|\right) && (5.57)\\
&\overset{d}{=}\frac{1}{2}\left(\mathrm{tr}\left(\left(\mathbf{\Lambda}_s+\sigma^2\mathbf{I}\right)^{-1}\mathbf{\Lambda}_s\frac{\mathbf{Z}_m^{\mathrm{T}}\mathbf{Z}_m}{k-1}\right)+\log\left|\mathbf{\Lambda}_s+\sigma^2\mathbf{I}_m\right|-\log\left|\mathbf{V}\mathbf{\Lambda}_s^{\frac{1}{2}}\frac{\mathbf{Z}_m^{\mathrm{T}}\mathbf{Z}_m}{k-1}\mathbf{\Lambda}_s^{\frac{1}{2}}\mathbf{V}^{\mathrm{T}}+\sigma^2\mathbf{I}_m\right|\right) && (5.58)\\
&\overset{d}{=}\frac{1}{2}\left(\mathrm{tr}\left(\left(\mathbf{\Lambda}_s+\sigma^2\mathbf{I}\right)^{-1}\mathbf{\Lambda}_s\frac{\mathbf{Z}_m^{\mathrm{T}}\mathbf{Z}_m}{k-1}\right)+\log\left|\frac{\mathbf{\Lambda}_s}{\sigma^2}+\mathbf{I}\right|-\log\left|\frac{\mathbf{\Lambda}_s}{\sigma^2}\frac{\mathbf{Z}_m^{\mathrm{T}}\mathbf{Z}_m}{k-1}+\mathbf{I}_m\right|\right) && (5.59)\\
&\overset{d}{=}\frac{1}{2}\left(\mathrm{tr}\left(\left(\tilde{\mathbf{\Lambda}}+\mathbf{I}\right)^{-1}\tilde{\mathbf{\Lambda}}\frac{\mathbf{Z}_n^{\mathrm{T}}\mathbf{Z}_n}{k-1}\right)+\log\left|\tilde{\mathbf{\Lambda}}+\mathbf{I}\right|-\log\left|\tilde{\mathbf{\Lambda}}\frac{\mathbf{Z}_n^{\mathrm{T}}\mathbf{Z}_n}{k-1}+\mathbf{I}_n\right|\right) && (5.60)
\end{aligned}$$

where (5.56) follows from the definition of objective $f(\mathbf{\Sigma}_{\tilde{A}\tilde{A}})$ in (5.30); (5.57) follows from the fact that $\mathbf{\Sigma}_{\tilde{A}\tilde{A}} = \mathbf{H}\mathbf{S}_{XX}\mathbf{H}^{\mathrm{T}} \sim \frac{1}{k-1}W_m(k-1, \mathbf{H}\mathbf{\Sigma}_{XX}\mathbf{H}^{\mathrm{T}})$, so it holds that

$$\mathbf{H}\mathbf{S}_{XX}\mathbf{H}^{\mathrm{T}} \overset{d}{=} \mathbf{V}\mathbf{\Lambda}_s^{\frac{1}{2}}\frac{\mathbf{Z}_m^{\mathrm{T}}\mathbf{Z}_m}{k-1}\mathbf{\Lambda}_s^{\frac{1}{2}}\mathbf{V}^{\mathrm{T}}; \quad (5.61)$$

Given the fact that $\mathbf{\Sigma}_{YY} = \mathbf{H}\mathbf{\Sigma}_{XX}\mathbf{H}^{\mathrm{T}}+\sigma^2\mathbf{I}$ shares the same eigenvectors as $\mathbf{H}\mathbf{\Sigma}_{XX}\mathbf{H}^{\mathrm{T}}$, (5.58) follows from apply cyclic permutation for the trace term in (5.57); (5.59) follows from Sylvester's determinant identity; (5.60) follows from the fact $\mathbf{\Lambda}_s$ is a rank deficient matrix with rank n. This completes the proof. □

The ergodic performance of the attack is the expected value of the performance with respect to the distribution given in (5.55). To obtain the asymptotic ergodic performance in (5.55), the asymptotic behavior of diagonal matrix $\tilde{\mathbf{\Lambda}} \in \mathbb{R}^n$ needs to be defined. Let n_0 denote the number of state variables of the power system. For example, when the voltage angles of the buses are chosen to be the state variables

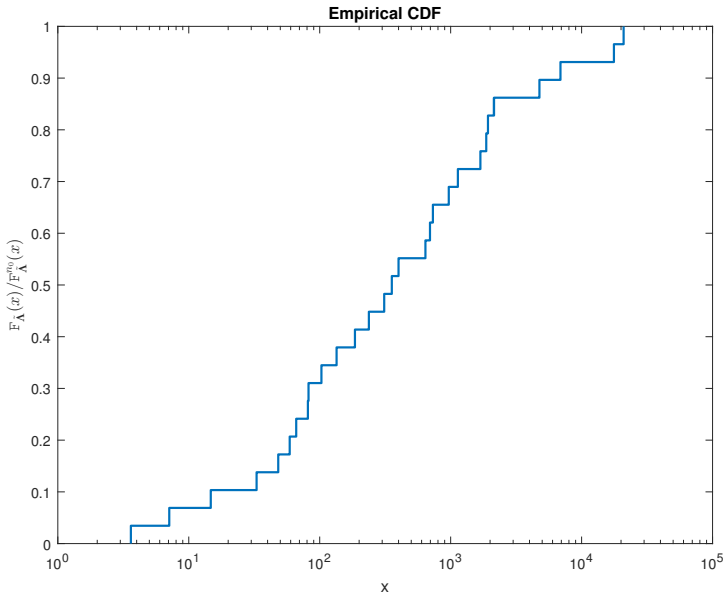


Fig. 5.1. An example for AED of $\tilde{\mathbf{\Lambda}}$, or empirical c.d.f. of $\tilde{\mathbf{\Lambda}}$ when $n = n_0$.

there are 29 state variables for the IEEE 30-Bus test system, which implies that $n_0 = 29$. As a result, there are 29 positive eigenvalues of the matrix $\mathbf{H}\mathbf{\Sigma}_{XX}\mathbf{H}^{\mathrm{T}}$. The empirical cumulative distribution function (c.d.f.) of the diagonal elements of $\tilde{\mathbf{\Lambda}}$ when $n = n_0$ is given by

$$\mathsf{F}_{\tilde{\mathbf{\Lambda}}}^{n_0}(x) = \frac{\sum_{i=1}^{n_0} \mathbb{1}_{\{x \leq \lambda_i(\tilde{\mathbf{\Lambda}})\}}}{n_0}, \tag{5.62}$$

which is obtained from the parameters of the power system. When $n \to \infty$, the AED of $\tilde{\mathbf{\Lambda}}$, i.e. $\mathsf{F}_{\tilde{\mathbf{\Lambda}}}(x)$, is the limiting distribution of $\mathsf{F}_{\tilde{\mathbf{\Lambda}}}^{n_0}(x)$, i.e.

$$\mathsf{F}_{\tilde{\mathbf{\Lambda}}}(x) = \lim_{n\to\infty} \frac{\sum_{i=1}^{n} \mathbb{1}_{\{x \leq \lambda_i(\tilde{\mathbf{\Lambda}})\}}}{n}. \tag{5.63}$$

Here we define that

$$\mathsf{F}_{\tilde{\mathbf{\Lambda}}}(x) = \lim_{n\to\infty} \frac{\sum_{i=1}^{n} \mathbb{1}_{\{x \leq \lambda_i(\tilde{\mathbf{\Lambda}})\}}}{n} = \frac{\sum_{i=1}^{n_0} \mathbb{1}_{\{x \leq \lambda_i(\tilde{\mathbf{\Lambda}})\}}}{n_0}, \tag{5.64}$$

which states that the AED of $\tilde{\mathbf{\Lambda}}$ is the same as the distribution of eigenvalues when $n = n_0$. Fig. 5.1 provides an illustrative example for AED of $\tilde{\mathbf{\Lambda}}$, or empirical c.d.f. of $\tilde{\mathbf{\Lambda}}$ when $n = n_0$.

The following proposition provides a closed-form expression for the Shannon transform of the AED of the attack covariance matrix.

Lemma 5.6. *Let $k \to \infty$ with $\frac{n}{m} \to \alpha$ and $\frac{k-1}{n} \to \beta$. Then it holds that*

$$\frac{1}{n}\mathbb{E}\left[\log\left|\tilde{\mathbf{\Lambda}}\frac{\mathbf{Z}_n^T\mathbf{Z}_n}{k-1}+\mathbf{I}_n\right|\right] \to \mathcal{V}_{\tilde{\mathbf{\Lambda}}}(\eta)-\beta\log\eta+\beta(\eta-1)\log e, \tag{5.65}$$

where $\to$ denotes almost surely convergence to, η denotes the η-transform of $\widetilde{\mathbf{Z}}_n\tilde{\mathbf{\Lambda}}\widetilde{\mathbf{Z}}_n^T$ evaluated at $\gamma=1$ and satisfies

$$\frac{1}{\beta}=\frac{1-\eta}{1-\eta_{\tilde{\mathbf{\Lambda}}}(\eta)}, \tag{5.66}$$

for which $\widetilde{\mathbf{Z}}_n \triangleq \frac{\mathbf{Z}_n}{\sqrt{k-1}}$ is a random matrix of size $(k-1)\times n$ with i.i.d. entries distributed as $\mathcal{N}(0,\frac{1}{k-1})$; and $\tilde{\mathbf{\Lambda}}$ is a diagonal matrix of size $n\times n$ whose AED is given by (5.62).

Proof. Note that

$$\begin{aligned}
\frac{1}{n}\mathbb{E}\left[\log\left|\tilde{\mathbf{\Lambda}}\frac{\mathbf{Z}_n^{\mathrm{T}}\mathbf{Z}_n}{k-1}+\mathbf{I}_n\right|\right] &= \frac{1}{n}\mathbb{E}\left[\log\left|\mathbf{I}_{k-1}+\frac{\mathbf{Z}_n}{\sqrt{k-1}}\tilde{\mathbf{\Lambda}}\frac{\mathbf{Z}_n^{\mathrm{T}}}{\sqrt{k-1}}\right|\right] && (5.67)\\
&\to \beta\mathbb{E}\left[\log\left(1+\lambda_{\widetilde{\mathbf{Z}}_n\tilde{\mathbf{\Lambda}}\widetilde{\mathbf{Z}}_n^{\mathrm{T}}}\right)\right] && (5.68)\\
&\to \beta\mathcal{V}_{\widetilde{\mathbf{Z}}_n\tilde{\mathbf{\Lambda}}\widetilde{\mathbf{Z}}_n^{\mathrm{T}}}(1) && (5.69)\\
&\to \mathcal{V}_{\tilde{\mathbf{\Lambda}}}(\eta)-\beta\log\eta+\beta(\eta-1)\log e && (5.70)
\end{aligned}$$

where (5.67) follows from Sylvester's determinant identity; (5.68) follows from denoting the unordered eigenvalues of matrix $\mathbf{A}$ by $\lambda_{\mathbf{A}}$; (5.69) follows from the definition of Shannon transform given by Definition 5.4; (5.70) follows from [130, Theorem 2.39] directly. □

Lemma 5.6 characterizes the Shannon transform of the attack covariance matrix, in which the η-transform of $\widetilde{\mathbf{Z}}_n\tilde{\mathbf{\Lambda}}\widetilde{\mathbf{Z}}_n^{\mathrm{T}}$ denoted by η is solved from (5.66). The following proposition shows that (5.66) always has a unique solution of η for any $\beta \geq 0$.

Proposition 5.1. *Let $n_0 \geq 1$, $\beta \in [0,\infty)$, and the AED of $\tilde{\mathbf{\Lambda}}$ is given by (5.62). Then η in (5.66) has a unique solution.*

Proof. Given the fact that the unordered eigenvalue distribution of $\tilde{\mathbf{\Lambda}}$ is given by (5.62), it follows that the η-transform of $\tilde{\mathbf{\Lambda}}$ at $\gamma=\eta$ is given by

$$\eta_{\tilde{\mathbf{\Lambda}}}(\eta)=\mathbb{E}_{\tilde{\Lambda}}\left[\frac{1}{1+\eta\tilde{\Lambda}}\right]=\sum_{i=1}^{n^0}\frac{1}{n^0}\frac{1}{1+\frac{\eta}{\sigma^2}\lambda_i}, \tag{5.71}$$

where $\tilde{\Lambda}$ is a random variable distributed as the AED of $\tilde{\mathbf{\Lambda}}$, i.e. $\tilde{\Lambda} \sim \mathsf{F}_{\tilde{\mathbf{\Lambda}}}(x)$. After some algebraic manipulation (5.66) can be expressed as

$$\beta\eta - \frac{1}{n^0}\left(\sum_{i=1}^{n^0} \frac{1}{1+\frac{\eta}{\sigma^2}\lambda_i}\right) = \beta - 1. \tag{5.72}$$

Note that the range of η is within the interval $[0,1]$, the left-hand term of (5.72) is a monotonically increasing function of $\eta \in (0,1]$ and its range contains the value $\beta - 1$. This completes the proof. □

It is worth pointing out that Proposition 5.1 can be easily extended for any positive semi-definite matrix $\tilde{\mathbf{\Lambda}}$. Taking Lemma 5.6 into the expected value of $f(\mathbf{\Sigma}_{\tilde{A}\tilde{A}})$ in (5.55) yields the following theorem, which characterizes the asymptotic ergodic performance is characterized.

Theorem 5.4. *Let $k \to \infty$ with $\frac{n}{m} \to \alpha$ and $\frac{k-1}{n} \to \beta$, then the ergodic performance of the stealth attacks*

$$\bar{f}_n \triangleq \frac{1}{n} f(\mathbf{\Sigma}_{\tilde{A}\tilde{A}}) \tag{5.73}$$

converges almost surely to $\bar{f}_\infty$

$$\bar{f}_\infty \triangleq \frac{1}{2}\left(\frac{\tilde{\Lambda}}{\tilde{\Lambda}+1}\lambda_{\widetilde{\mathbf{Z}}_n^T\widetilde{\mathbf{Z}}_n} + \log\left(\tilde{\Lambda}+1\right) - \log\left(1+\lambda_{\widetilde{\mathbf{Z}}_n\tilde{\mathbf{\Lambda}}\widetilde{\mathbf{Z}}_n^T}\right)\right) \tag{5.74}$$

with

$$\mathbb{E}\left[\bar{f}_\infty\right] = \frac{1}{2}\Big(\Theta + \Xi\Big) - \frac{1}{2}\left(\mathcal{V}_{\tilde{\mathbf{\Lambda}}}(\eta) - \beta\log\eta + \beta(\eta-1)\log e\right), \tag{5.75}$$

where

$$\Theta \triangleq \lim_{n\to\infty}\frac{1}{n}\mathrm{tr}\left(\left(\tilde{\mathbf{\Lambda}}+\mathbf{I}\right)^{-1}\tilde{\mathbf{\Lambda}}\right) = \frac{1}{n^0}\sum_{i=1}^{n^0}\frac{\lambda_i}{\lambda_i+\sigma^2} \tag{5.76}$$

and

$$\Xi \triangleq \lim_{n\to\infty}\frac{1}{n}\log\left|\tilde{\mathbf{\Lambda}}+\mathbf{I}\right| = \frac{1}{n^0}\sum_{i=1}^{n^0}\log\left(\frac{\lambda_i}{\sigma^2}+1\right) \tag{5.77}$$

are determined by the distribution given in (5.62).

Proof. Starting from (5.60), there is

$$\bar{f}_\infty = \lim_{n\to\infty} \frac{1}{n} f(\mathbf{\Sigma}_{\tilde{A}\tilde{A}}) \tag{5.78}$$

$$= \frac{1}{2n}\left(\mathrm{tr}\left(\left(\tilde{\mathbf{\Lambda}}+\mathbf{I}\right)^{-1}\tilde{\mathbf{\Lambda}}\frac{\mathbf{Z}_n^{\mathrm{T}}\mathbf{Z}_n}{k-1}\right)+\log\left|\tilde{\mathbf{\Lambda}}+\mathbf{I}\right|-\log\left|\tilde{\mathbf{\Lambda}}\frac{\mathbf{Z}_n^{\mathrm{T}}\mathbf{Z}_n}{k-1}+\mathbf{I}_n\right|\right) \tag{5.79}$$

$$\to \frac{1}{2}\left(\frac{\tilde{\Lambda}}{\tilde{\Lambda}+1}\lambda_{\widetilde{\mathbf{Z}}_n^{\mathrm{T}}\widetilde{\mathbf{Z}}_n}+\log\left(\tilde{\Lambda}+1\right)-\log\left(1+\lambda_{\widetilde{\mathbf{Z}}_n\tilde{\mathbf{\Lambda}}\widetilde{\mathbf{Z}}_n^{\mathrm{T}}}\right)\right), \tag{5.80}$$

where (5.80) follows from the same operation in (5.68).

Adding the fact that $\mathbb{E}[\lambda_{\widetilde{\mathbf{Z}}_n^{\mathrm{T}}\widetilde{\mathbf{Z}}_n}] = 1$, the expected value of $\bar{f}_\infty$ follows from taking Lemma 5.6 into (5.80) and is given by

$$\mathbb{E}\left[\bar{f}_\infty\right] = \frac{1}{2}\Big(\Theta+\Xi\Big) - \frac{1}{2}\left(\mathcal{V}_{\tilde{\mathbf{\Lambda}}}(\eta) - \beta\log\eta + \beta(\eta-1)\log e\right). \tag{5.81}$$

This completes the proof. □

5.4 Numerical Results

The numerical simulations are implemented on the IEEE 30-Bus and 118-Bus test system, where the Jacobian matrix $\mathbf{H}$ is obtained using MATPOWER [122]. The construction of the vector of measurements is the same as Section 4.3. For the construction of the stealth attack the covariance matrix of the state variables is chosen to be a Toeplitz matrix with exponential decay parameter ρ as in (3.55). Specifically, the Toeplitz matrix of dimension $n \times n$ with exponential decay parameter ρ is given by $\mathbf{\Sigma}_{XX} = [s_{ij} = \rho^{|i-j|}; i, j = 1, 2, \ldots, n]$. We define the SNR as

$$\mathrm{SNR} = 10\log_{10}\left(\frac{\mathrm{tr}(\mathbf{H}\mathbf{\Sigma}_{XX}\mathbf{H}^{\mathrm{T}})}{m\sigma^2}\right). \tag{5.82}$$

To verify the results in Lemma 5.2, we generate 100 realizations for the sample covariance matrix distributed as $\mathbf{S}_{XX} \sim \frac{1}{k-1}W_n(k-1, \mathbf{\Sigma}_{XX})$. Fig. 5.2 shows the minimum objective value in (5.4) among the 100 realizations under different numbers of training samples. It is easy to see that when the number of training samples increases the minimum value of the performance converges to the optimal value of the perfect knowledge scenario, i.e. when the attacker knows the covariance matrix $\mathbf{\Sigma}_{XX}$ exactly. However the performance of the attacks using sample covariance is always above the optimal value, which is consistent with Lemma 5.2. Here we only provide the results on IEEE 30-Bus test system and SNR = 20 dB, the simulations for the IEEE 118-Bus test system and the other parameter for the SNR show the same result as the simulation using the IEEE 30-Bus test system.

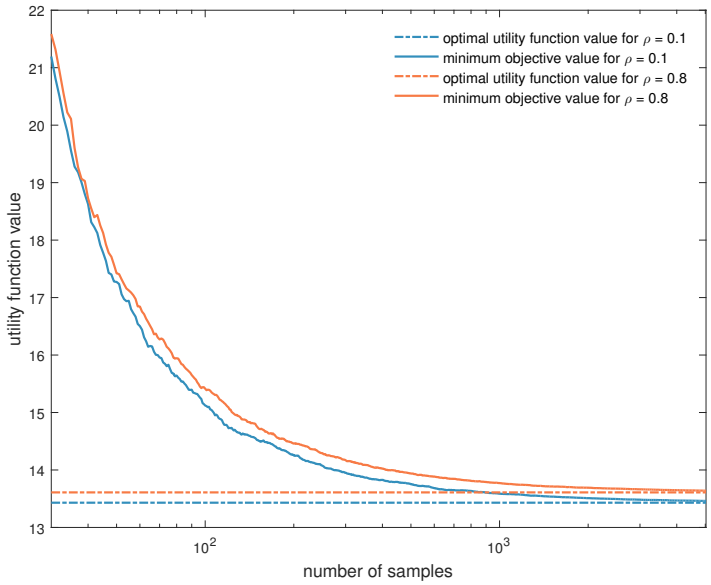


Fig. 5.2. The minimum objective value in (5.4) for $\rho = 0.1$ and $\rho = 0.8$ for SNR = 20 dB on IEEE 30-Bus test system.

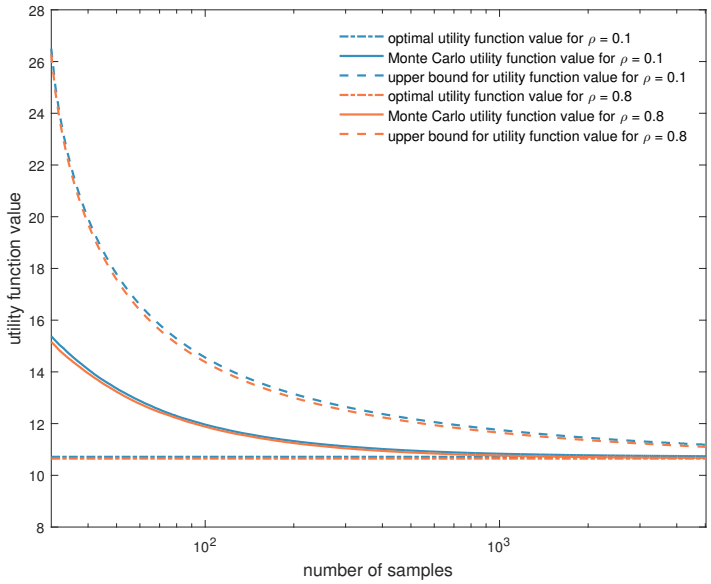


Fig. 5.3. Performance of the upper bound in Theorem 5.2 for $\rho = 0.1$ and $\rho = 0.8$ on IEEE 30-Bus test system when SNR = 10 dB.

5.4.1 Simulation for Upper Bound in Theorem 5.2

Fig. 5.3 depicts the performance of the upper bound in Theorem 5.2 for $\rho = 0.1$ and $\rho = 0.8$ on IEEE 30-Bus test system when SNR = 10 dB, in which the Monte Carlo utility function value is obtained by averaging over 100 realizations of the sample covariance matrix. Similarly Fig. 5.4 and Fig. 5.5 show the performance of the bound when SNR = 20 dB and SNR = 30 dB, respectively. It is easy to see that the proposed upper bound is tight when the number of training samples is large for

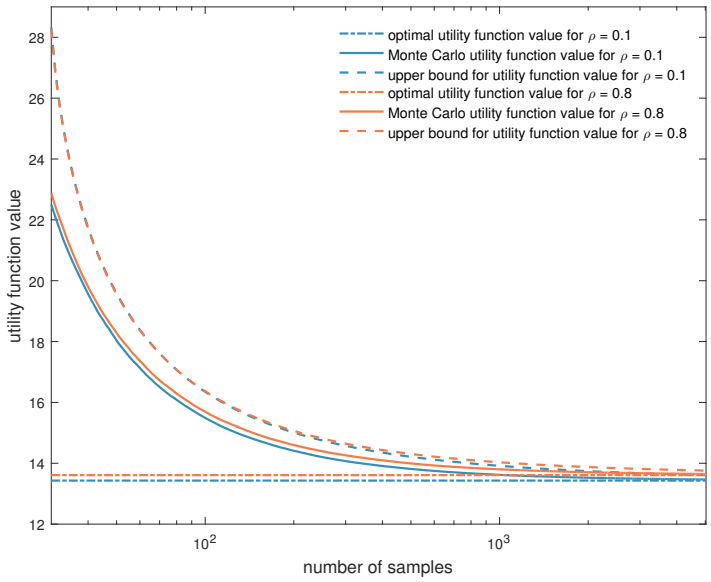


Fig. 5.4. Performance of the upper bound in Theorem 5.2 for $\rho = 0.1$ and $\rho = 0.8$ on IEEE 30-Bus test system when SNR = 20dB.

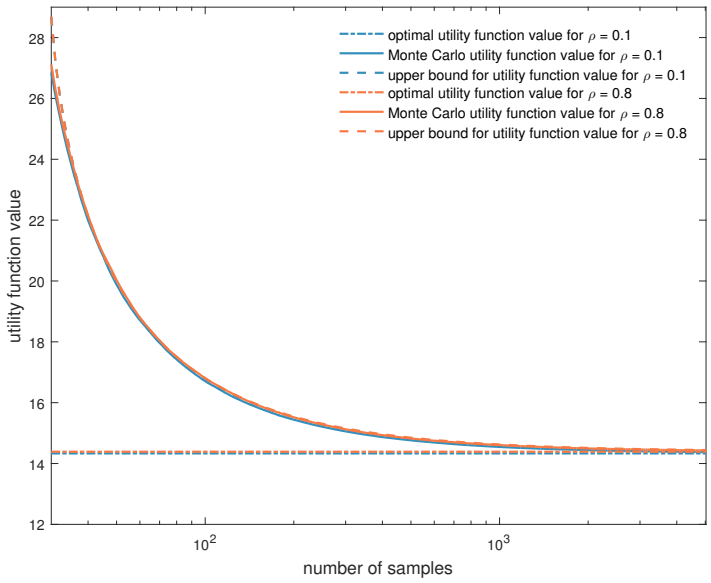


Fig. 5.5. Performance of the upper bound in Theorem 5.2 for $\rho = 0.1$ and $\rho = 0.8$ on IEEE 30-Bus test system when SNR = 30dB.

different values of SNR. Especially when the SNR is high the proposed upper bound is almost the same as the performance obtained through Monte Carlo. Interestingly unlike the performance obtained via Monte Carlo, the upper bound is quite steady under different values of SNR. This implies that when the SNR of the power system is high, the attacks have a higher probability of detection, but the upper bound in Theorem 5.2 is tighter.

Fig. 5.6 to Fig. 5.8 depicts the performance of the upper bound in Theorem 5.2 for $\rho = 0.1$ and $\rho = 0.8$ on IEEE 118-Bus test system when SNR = 10 dB, 20 dB

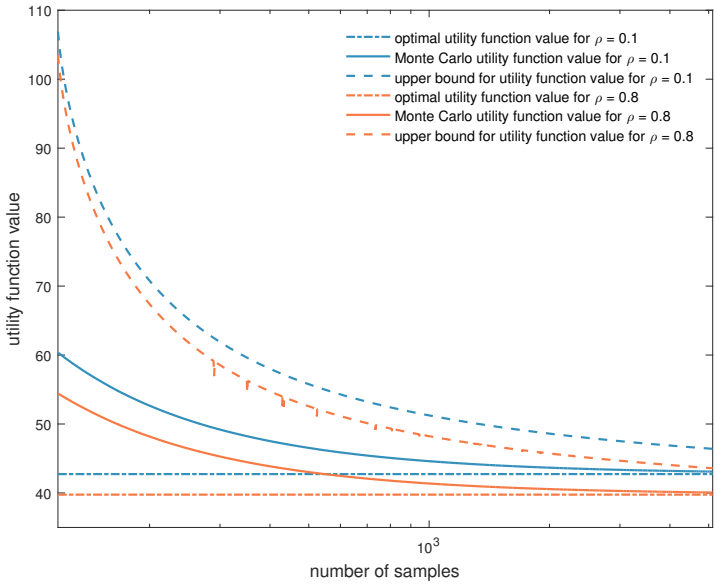


Fig. 5.6. Performance of the upper bound in Theorem 5.2 for $\rho = 0.1$ and $\rho = 0.8$ on IEEE 118-Bus test system when SNR = 10dB.

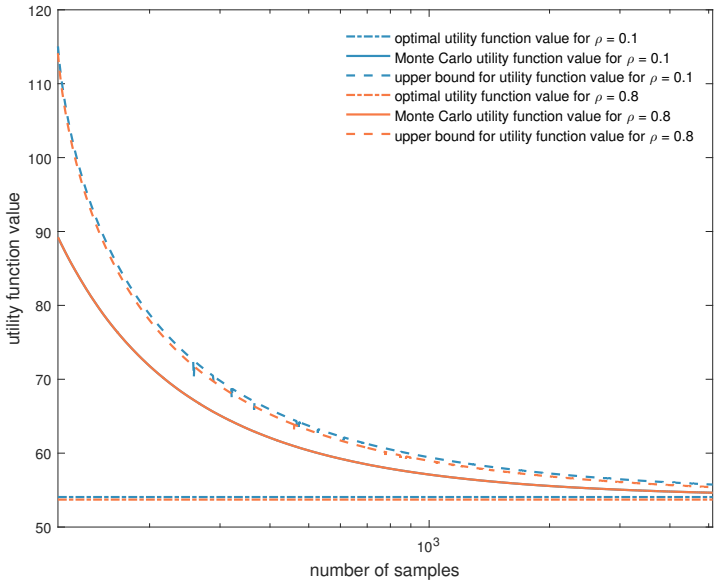


Fig. 5.7. Performance of the upper bound in Theorem 5.2 for $\rho = 0.1$ and $\rho = 0.8$ on IEEE 118-Bus test system when SNR = 20dB.

and 30 dB, respectively, in which the Monton Carlo utility function value is obtained by averaging over 200 realizations of the sample covariance matrix. Similar to the simulation on IEEE 30-Bus test system, the bound is tighter when SNR is high and is insensitive to the change of SNR. The tradeoff between the probability of detection and the tightness of the bound still exists for the 118-Bus test system.

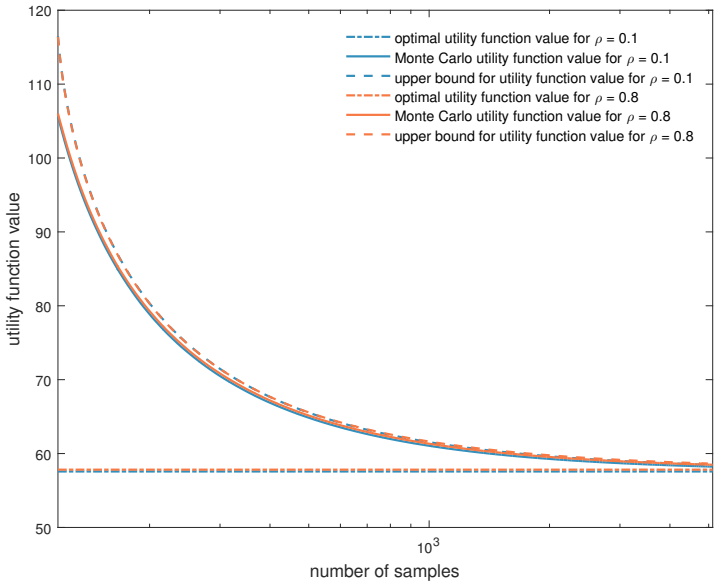


Fig. 5.8. Performance of the upper bound in Theorem 5.2 for $\rho = 0.1$ and $\rho = 0.8$ on IEEE 118-Bus test system when SNR = 30dB.

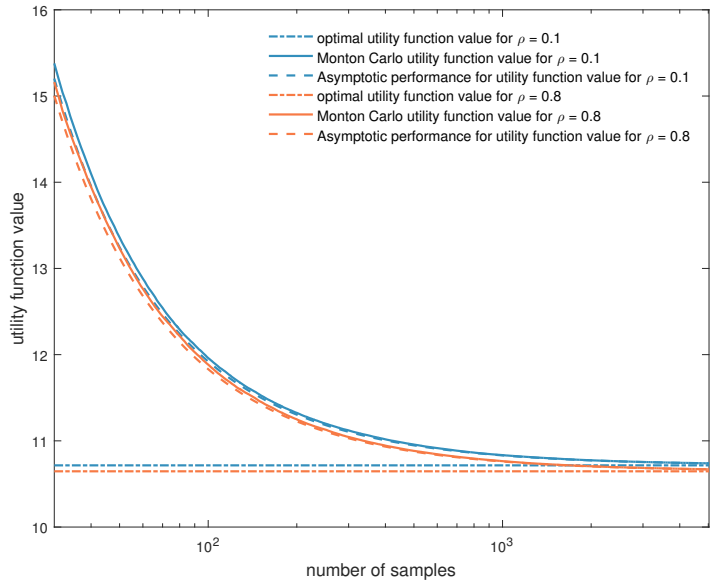


Fig. 5.9. Performance of the asymptotic performance in Theorem 5.4 for $\rho = 0.1$ and $\rho = 0.8$ on IEEE 30-Bus test system when SNR = 10dB.

5.4.2 Simulation for Asymptotic Performance in Theorem 5.4

Fig. 5.9 depicts the asymptotic performance of the attacks in Theorem 5.4 for $\rho = 0.1$ and $\rho = 0.8$ on IEEE 30-Bus test system when SNR = 10 dB, in which the Monte Carlo utility function value is obtained by averaging over 100 realizations of the sample covariance matrix. Similarly Fig. 5.10 and Fig. 5.11 show the asymptotic performance of the attacks in Theorem 5.2 when SNR = 20 dB and SNR = 30

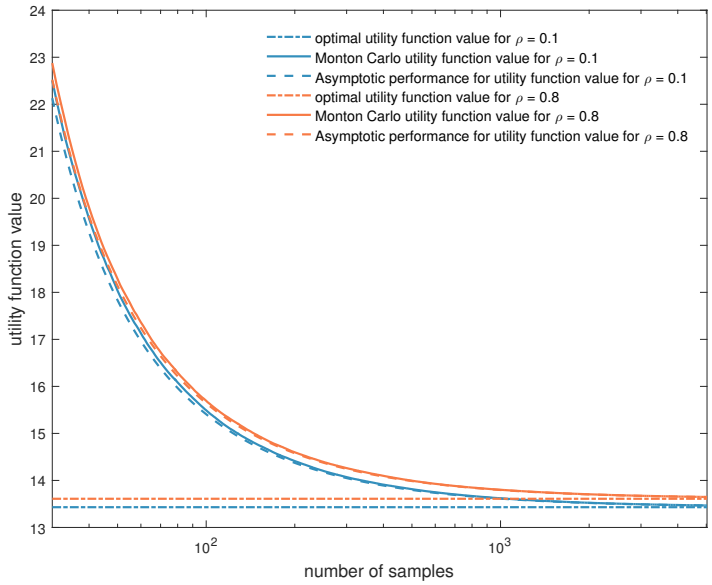


Fig. 5.10. Performance of the asymptotic performance in Theorem 5.4 for $\rho = 0.1$ and $\rho = 0.8$ on IEEE 30-Bus test system when SNR = 20dB.

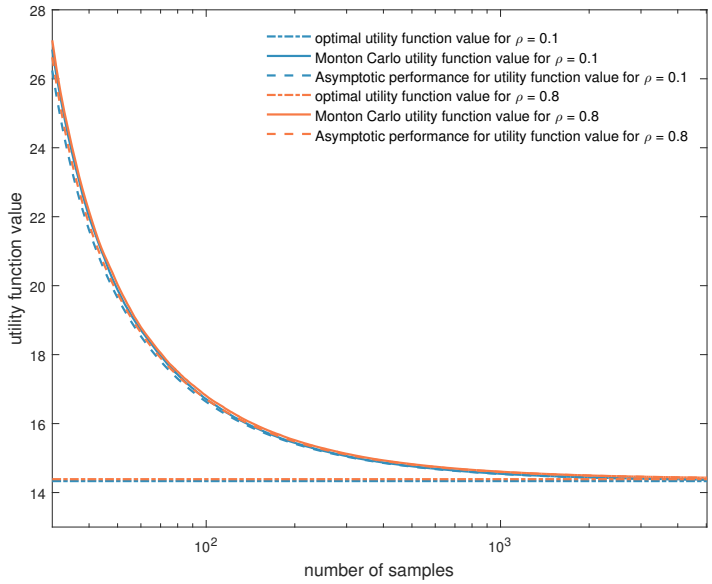


Fig. 5.11. Performance of the asymptotic performance in Theorem 5.4 for $\rho = 0.1$ and $\rho = 0.8$ on IEEE 30-Bus test system when SNR = 30dB.

dB, respectively. It is shown that the asymptotic performance is quite close to the performance obtained via Monte Carlo approach for different values of the correlation strength ρ. When the number of samples is high, i.e. β is of high values, the asymptotic performance is almost the same as the performance from Monton Carlo. Furthermore the asymptotic performance is closer to the performance from Monte Carlo with the SNR increases.

Fig. 5.12 to Fig. 5.14 depict the performance of the asymptotic performance in Theorem 5.4 for $\rho = 0.1$ and $\rho = 0.8$ on IEEE 118-Bus test system when SNR

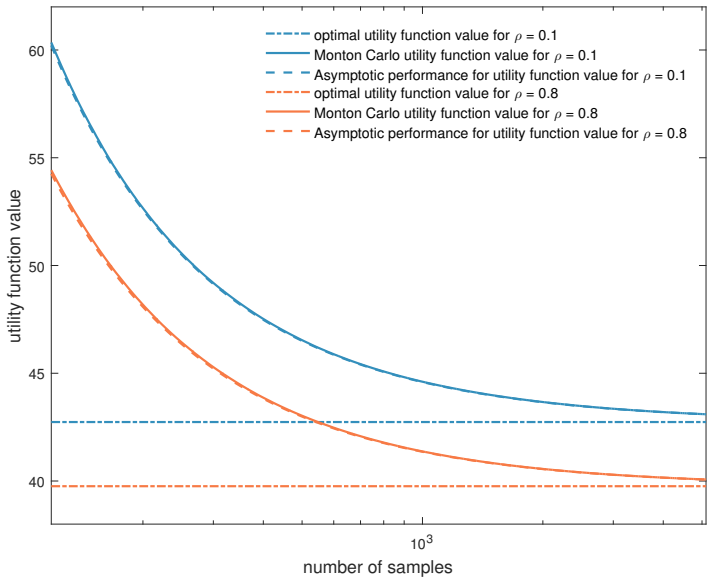


Fig. 5.12. Performance of the asymptotic performance in Theorem 5.4 for $\rho = 0.1$ and $\rho = 0.8$ on IEEE 118-Bus test system when SNR = 10dB.

= 10 dB, 20 dB, and 30 dB, respectively. Compared with the simulation results for IEEE 30-Bus test system, the asymptotic performance in IEEE 118 Bus test system is closer to the results from Monton Carlo. This implies that the asymptotic performance is closer to the finite performance of attacks on the power system of large scale, which suggests that the attacker approximates the finite performance better in a power system of larger scale. This observation is consistent with the results of RMT covered in Appendix 5, which states that with the dimension of sample covariance matrix increases, the distribution of unordered eigenvalues of sample covariance matrix under the finite case is closer to the limiting distribution of it under the asymptotic case.

5.5 Summary

In this chapter, the performance of the attacks is analyzed using RMT tools for the case that the attacker has imperfect knowledge about the second order statistics of the state variables. Specifically the attacker only gets access to a limited number of samples of the state variables, and estimates the second order statistics of the state variables via the sample covariance matrix of the samples. RMT tools are employed to characterize the ergodic performance of the attacks constructed using sample covariance matrix for both the non-asymptotic scenario and the asymptotic scenario. Given the fact that the distribution of the singular values of random matrices is challenging to characterize under the non-asymptotic scenario, an upper bound is proposed for the ergodic performance of the attacks using sample covariance matrix,

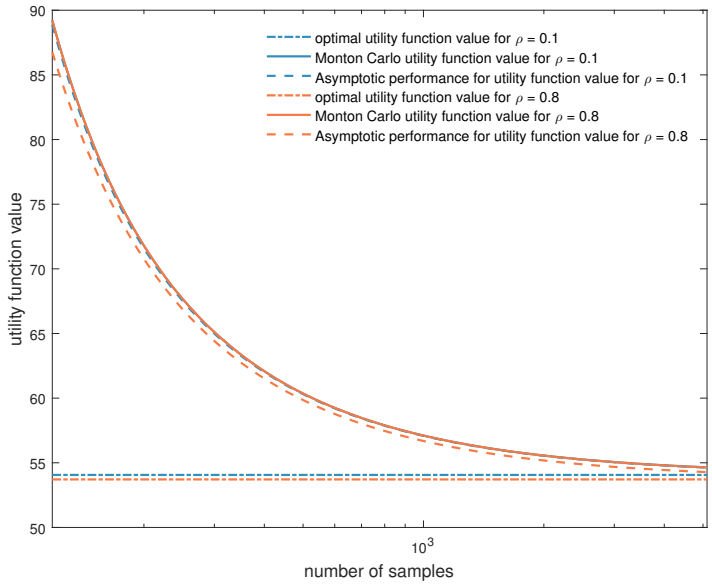


Fig. 5.13. Performance of the asymptotic performance in Theorem 5.4 for $\rho = 0.1$ and $\rho = 0.8$ on IEEE 118-Bus test system when SNR = 20dB.

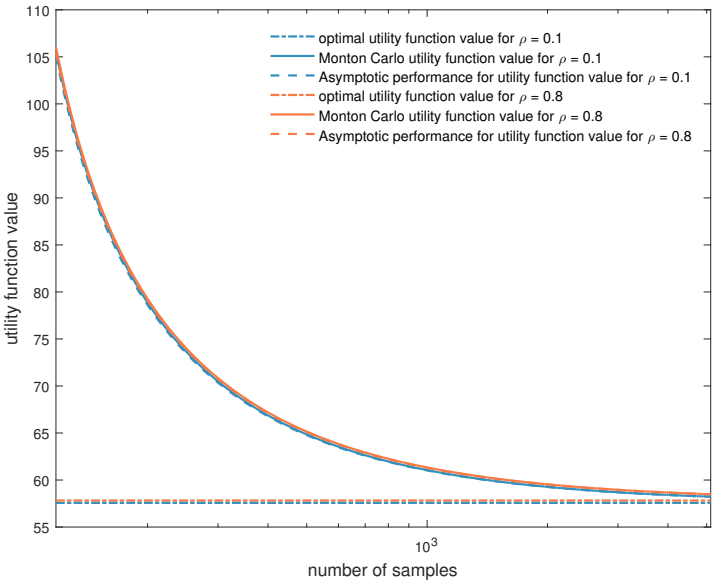


Fig. 5.14. Performance of the asymptotic performance in Theorem 5.4 for $\rho = 0.1$ and $\rho = 0.8$ on IEEE 118-Bus test system when SNR = 30dB.

in which a simple convex optimization needs to be solved. As a result, the ergodic performance of the attacks using sample covariance matrix is regulated by the the proposed upper bound and the performance of the attacks when perfect knowledge is known by the attacker. For the asymptotic scenario, an equivalent distribution is proposed for the performance of the attacks using sample covariance matrix, and the closed-form expression for the ergodic performance is provided via the equivalent distribution.

Chapter 6

Conclusions and Future Work

6.1 Conclusions

In this thesis, the information-theoretic measures are utilized to quantify the disruption caused by DIAs and the probability of detection induced by DIAs. Specifically the mutual information between the state variables and the compromised measurements is minimized to reduce the amount of information acquired by the operator about the state variables from the obtained measurements. On the other hand, the minimization of the probability of detection induced by the attacks is characterized by minimizing the KL divergence between the distribution of measurements under attacks and without attacks. The stealth attacks sum up these two contradictive objectives, and the closed-form expression for the stealth attacks is proposed for the Gaussian attack case.

To achieve lower probability of detection, the stealth attacks are generalized by assigning a weighting parameter to the KL divergence term in the objective of stealth attacks. When the weighting parameter is larger than one, the attacker is conservative and prioritizes the probability of detection over the caused disruption. A closed-form expression is proposed for the generalized stealth attacks when the weighting parameter is larger than one. Also closed-form expression is proposed for the resulting probability of detection. To provide explicit insight into the relation between the probability of detection and the weighting parameter, a concentration inequality upper bound is proposed for the probability of detection, which provides a guideline to the attacker for choosing the weighting parameter.

The (generalized) stealth attacks require the second order statistics of the state variables to construct the attacks. The requirement is relaxed for the scenario that the attacker only gets access to a limited number of samples of the state variables. Specifically the attacker estimates the second order statistics of the state variables

via the sample covariance matrix of the samples of the state variables. RMT tools are used to characterize the ergodic performance of the attacks constructed using sample covariance matrix for the asymptotic scenario and non-asymptotic scenario. Given the randomness of sample covariance matrices under the non-asymptotic scenario, closed-form expression is not available for the ergodic performance. Instead an upper bound is proposed for the ergodic performance, for which a simple convex optimization needs to be solved to compute it. For the asymptotic case a closed-form expression is provided for the ergodic performance of the attacks using a sample covariance matrix.

6.2 Future Work

This section discusses the open research directions that arise based on the work presented in this thesis.

Asymptotic Characterization of Variance of Performance of Attacks using Sample Covariance Matrix

In Chapter 5 we use RMT to characterize the ergodic performance of the attacks using sample covariance matrix both asymptotically and non-asymptotically. In section 5.3 we propose the equivalent distribution for the performance of the attacks using sample covariance matrix, and the explicit expression for the ergodic performance under the asymptotic setting. Specifically Theorem 5.4 describes the ergodic performance, i.e. the expected value, of the equivalent distribution given in Theorem 5.3. The central limit theorem results for linear spectral statistics of random matrices can be utilized to characterize the asymptotic variance for the distribution given in Theorem 5.3.

Sparse Information-Theoretic Attacks

As reviewed in section 2.2, compromising the sensors in the system is costly for the attacker, as a result, the attacker wants to minimize or constrain the number of sensors that need to be hacked. Two different scenarios can be considered here. Firstly when the attacker can only compromise the sensors within a certain subset of all the sensors, and the cardinality of the subset is K, then the attack construction problem is the same as the one in (3.26) or (4.3), in which $m - K$ constraints are added to guarantee that the rest of the sensors are not compromised. This problem is still a convex optimization problem, as the constraints still form a convex set. Secondly when the attacker has the ability to compromise at most K meters in the system, and the attacker has the ability to compromise any meters within the system,

the attack construction problem is equivalent to a subset selection problem, in which the attacker chooses a K-cardinality subset of all the sensors that maximizes the caused distortion with a constraint on the probability of detection.

Decentralized Information-Theoretic Attacks

The stealth attacks in Chapter 3 and the generalized stealth attacks in Chapter 4 construct the attacks in a centralized pattern, i.e. there is only one attacker in the power system. For the condition that there are multiple attackers in the system, the collaboration between the attackers is modeled by a game, in which the attackers are the players. Specifically, each of the attackers in the system has the ability to compromising part of all the sensors in the power system, and each attacker only injects attacks into the sensors that are compromisable for this attacker. In this game, the attackers aim to maximize the distortion caused by the attackers with the constraint that the probability of detection is smaller than a threshold.

Appendix A

Information Theory

Discrete Case

Definition A.1. *The* **entropy** $H(X)$ *of a discrete random variable* X *is defined as*

$$H(X) = -\sum_{x \in \mathcal{X}} p(x) \log p(x), \tag{A.1}$$

where $\mathcal{X}$ *is the alphabet of* X*.*

- Entropy is a measure of the uncertainty of a random variable, it is also a measure of the amount of information required on the average to describe the random variable.
- We use the convention that $0 \log 0 = 0$.
- Entropy is expressed in *bits* when the log is to base 2, and in *nats* when the base is the Euler's number.
- $H(X) \geq 0$.
- $H(X) = \mathbb{E}_X \left[\log \frac{1}{p(X)}\right]$.

Definition A.2. *The* **joint entropy** $H(X, Y)$ *of a pari of discrete random variables* (X, Y) *with a joint distribution* $p(x, y)$ *is defined as*

$$H(X, Y) = -\sum_{x \in \mathcal{X}} \sum_{y \in \mathcal{Y}} p(x, y) \log p(x, y), \tag{A.2}$$

where $\mathcal{Y}$ *is the alphabet of* Y*.*

- $H(X, Y) = \mathbb{E}_{X,Y} \left[\log \frac{1}{p(X,Y)}\right]$.

Definition A.3. *If* $(X, Y) \sim p(x, y)$*, the* **conditional entropy** $H(Y|X)$ *is defined as*

$$\begin{aligned} H(Y|X) &= \sum_{x \in \mathcal{X}} p(x) H(Y|X = x) && (A.3) \\ &= -\sum_{x \in \mathcal{X}} p(x) \sum_{y \in \mathcal{Y}} p(y|x) \log p(y|x) && (A.4) \\ &= -\sum_{x \in \mathcal{X}} \sum_{y \in \mathcal{Y}} p(x, y) \log p(y|x) && (A.5) \\ &= -\mathbb{E}_{X,Y} \left[\log p(Y|X) \right]. && (A.6) \end{aligned}$$

- $H(X, Y) = H(X) + H(Y|X) = H(Y) + H(X|Y)$.

Definition A.4. *The* **relative entropy** *or* **KL divergence** *between two probability mass functions* $p(x)$ *and* $q(x)$ *is defined as*

$$\begin{aligned} D(p\|q) &= \sum_{x \in \mathcal{X}} p(x) \log \frac{p(x)}{q(x)} && (A.7) \\ &= \mathbb{E}_{X \sim p} \left[\log \frac{p(x)}{q(x)} \right]. && (A.8) \end{aligned}$$

- The expectation is taken with respect to distribution $p(x)$.
- We use the convention that $0 \log \frac{0}{0} = 0$, $0 \log \frac{0}{q} = 0$, and $p \log \frac{p}{0} = 0$.
- $D(p\|q) \geq 0$, and with equality only when $p(x) = q(x)$ for all $x \in \mathcal{X}$.
- KL divergence or relative entropy measures the "distance" between distributions, but it is not a true distance since it is not symmetric and does not satisfy the triangle inequality.
- In statistics, KL divergence arises as an expected logarithm of the likelihood ratio.

Definition A.5. *The* **condition divergence** *between two probability mass functions* $p(Y|X)$ *and* $q(Y|X)$ *is defined as*

$$\begin{aligned} D\left(p(Y|X)\|q(Y|X)|P_X\right) &= \mathbb{E}_X \left[D\left(p(Y|X = x)\|q(Y|X = x)\right) \right] && (A.9) \\ &= \sum_{x \in \mathcal{X}} p(x) D\left(p(Y|X = x)\|q(Y|X = x)\right). && (A.10) \end{aligned}$$

Definition A.6. *Consider two random variables* X *and* Y *with a joint probability mass function* $p(x, y)$ *and marginal probability mass functions* $p(x)$ *and* $p(y)$*. The*

mutual information $I(X;Y)$ *is the relative entropy between the joint distribution and the product distribution* $p(x)p(y)$*:*

$$I(X;Y) = \sum_{x\in\mathcal{X}}\sum_{y\in\mathcal{Y}} p(x,y)\log\frac{p(x,y)}{p(x)p(y)} \tag{A.11}$$

$$= D(p(x,y)\|p(x)p(y)) \tag{A.12}$$

$$= \mathbb{E}_{X,Y}\left[\log\frac{p(X,Y)}{p(X)p(Y)}\right]. \tag{A.13}$$

- Mutual information is a measure of the amount of information that one random variable contains about another random variable. It is the reduction in the uncertainty of one one random variable due to the knowledge of the other, i.e. $H(X) - H(X|Y)$.
- $I(X;Y) = H(X) - H(X|Y) = H(Y) - H(Y|X) = H(X) + H(Y) - H(X,Y)$.
- $I(X;Y) = I(Y;X)$, i.e. mutual information is symmetric.

Continuous Case

Definition A.7. *The* **differential entropy** $h(X)$ *of a continuous random variable* X *with density* $f(x)$*is defined as*

$$h(X) = -\int_S f(x)\log f(x), \tag{A.14}$$

where S *is the support of* X*.*

- $h(X)$ can be negative, such as the differential entropy of uniform distribution.
- Translation does not change the differential entropy, i.e. $h(X+c) = h(X)$.
- $h(aX) = h(X) + \log|a|$ for single variate case, and $h(\mathbf{AX}) = h(\mathbf{X}) + \log|\mathbf{A}|$ for multivariate case.

Definition A.8. *The* **differential entropy** $h(X)$ *of a set* $X_1,\ldots,X_n$ *of continuous random variable with density* $f(x_1,\ldots,x_n)$ *is defined as*

$$h(X_1,\ldots,X_n) = -\int f(x^n)\log f(x^n). \tag{A.15}$$

Definition A.9. *If* (X,Y) *has a joint density function* $f(x,y)$*, the* **conditional differential entropy** $h(Y|X)$ *is defined as*

$$h(Y|X) = -\int f(x,y)\log f(y|x)\mathrm{d}x\mathrm{d}y. \tag{A.16}$$

- $h(X,Y) = h(X) + h(Y|X) = h(Y) + h(X|Y)$ when any of the differential entropies are not infinite.

Definition A.10. *The* **relative entropy** *or* **KL divergence** $D(f\|g)$ *between two densities* f *and* g *is defined as*

$$D(f\|g) = \int f \frac{f}{g}. \tag{A.17}$$

- We use the convention that $0 \log \frac{0}{0} = 0$.
- $D(f\|g)$ is finite only if the support of f is contained in the support of g.
- $D(f\|g) \geq 0$ with equality iff $f = g$ almost everywhere.

Definition A.11. *The* **condition divergence** *between two probability density functions* $f(Y|X)$ *and* $g(Y|X)$ *is defined as*

$$D\left(f(Y|X)\|g(Y|X)|P_X\right) = \mathbb{E}_X\left[D\left(f(Y|X=x)\|g(Y|X=x)\right)\right] \tag{A.18}$$

$$= \int f(x) D\left(f(Y|X=x)\|g(Y|X=x)\right). \tag{A.19}$$

Definition A.12. *The* **mutual information** $I(X;Y)$ *between two random variables with joint density* $f(x,y)$ *is defined as*

$$I(X;Y) = \int f(x,y) \log \frac{f(x,y)}{f(x)f(y)} \mathrm{d}x\mathrm{d}y. \tag{A.20}$$

- $I(X;Y) = H(X) - H(X|Y) = H(Y) - H(Y|X) = H(X) + H(Y) - H(X,Y)$.
- $I(X;Y) = D(f(x,y)\|f(x)f(y))$.
- $I(X;Y) \geq 0$ with equality iff X and Y are independent.

Connection between Information Theory and Estimation Theory

Here we observe a random variable Y that is related to X by the conditional distribution $p(y|x)$. From Y, we calculate a function $g(Y) = \hat{X}$, where $\hat{X}$ is an estimate of X and takes in values in $\hat{\mathcal{X}}$.

Theorem A.1 (Fano's inequality)**.** *For any estimator* $\hat{X}$ *such that* $X \to Y \to \hat{X}$ *forms a Markov chain, with* $P_e = P(X \neq \hat{X})$*, we have*

$$H(P_e) + P_e \log |\mathcal{X}| \geq H(\hat{X}|X) \geq H(X|Y). \tag{A.21}$$

This inequality can be weakened to

$$1 + P_e \log |\mathcal{X}| \geq H(X|Y) \tag{A.22}$$

or

$$P_e \geq \frac{H(X|Y) - 1}{\log |\mathcal{X}|}. \tag{A.23}$$

Corollary A.1. *Let $P_e = P(X \neq \hat{X})$, and let $\hat{X} : \mathcal{Y} \to \mathcal{X}$; then*

$$H(P_e) + P_e(\log |\mathcal{X}| - 1) \geq H(X|Y). \tag{A.24}$$

For the continuous variable case, let X be a random variable with differential entropy $h(X)$ (in nats), and let $\hat{X}$ be an estimate of X, and let $\mathbb{E}\left[(X - \hat{X})^2\right]$ be the expected estimation error.

Theorem A.2. *For any random variable and estimator $\hat{X}$,*

$$\mathbb{E}\left[(X - \hat{X})^2\right] \geq \frac{1}{2\pi e} e^{2h(X)} \tag{A.25}$$

with equality iff X is Gaussian and $\hat{X}$ is the mean of X.

Corollary A.2. *Given side information Y and estimator $\hat{X}(Y)$, it follows that*

$$\mathbb{E}\left[(X - \hat{X}(Y))^2\right] \geq \frac{1}{2\pi e} e^{2h(X|Y)}. \tag{A.26}$$

Appendix B

Random Matrix Theory

Introduction to Random Matrix Theory

RMT is a subject that analyzes the behavior of random matrices, mainly spectral of the random matrices, i.e. eigenvalues or eigenvectors of the random matrices. The landmark contributions to the theory of random matrix of Wishart (1928), Wigner (1955), and Marčenko and Pastur (1967) were motivated to a large extent by practical experimental problem [130]. Due to the limited space in the thesis, here we focus on the Wishart matrix, which is utilized in this thesis to characterize the distribution of the sample covariance matrix of samples from a multivariate Gaussian distribution. Let X^l denote the vector of dimension $l \times 1$ whose entries are normal random variables, i.e. $X^l \sim \mathcal{N}(\mathbf{0}, \mathbf{\Sigma}_{XX})$, then $\mathbf{Z}_l = [X_1^l, \cdots, X_i^l, \cdots, X_k^l]^{\mathrm{T}}$ is the matrix of dimension $k \times l$ composed of k realizations of X^l. The sample covariance matrix of X^l using k samples is given by

$$\mathbf{S}_{XX} = \frac{1}{k-1} \sum_{i=1}^{k} X_i^l (X_i^l)^{\mathrm{T}}. \tag{B.1}$$

As a result of the Gaussianity of X^l, the sample covariance matrix $\mathbf{S}_{XX}$ follows a central Wishart distribution given by

$$\mathbf{S}_{XX} \sim \frac{1}{k-1} W_l \left(k-1, \mathbf{\Sigma}_{XX}\right), \tag{B.2}$$

where $W_l \left(k-1, \mathbf{\Sigma}_{XX}\right)$ denotes Wishart distribution with degree of freedom $k-1$. When the mean vector of X^l is $\mathbf{0}$, the Wishart distribution is called central Wishart distribution. The Wishart distribution has many properties, here we only list some properties that are used in this thesis. Further details about the properties are available at [130], [131], and [132].

Proposition B.1. *Let* $\mathbf{S}_{XX} \sim \frac{1}{k-1} W_l\left(k-1, \mathbf{\Sigma}_{XX}\right)$*, then it holds that*

$$\mathbb{E}\left[\mathbf{S}_{XX}\right] = \mathbf{\Sigma}_{XX}. \tag{B.3}$$

Proposition B.2. *Let* $\mathbf{S}_{XX} \sim \frac{1}{k-1} W_l\left(k-1, \mathbf{\Sigma}_{XX}\right)$ *with* $\mathbf{\Sigma}_{XX}$ *being a full rank matrix, then it holds that*

$$P\left[\left|\mathbf{\Sigma}_{XX}\right| = 0\right] = 0 \tag{B.4}$$

when $k - 1 \geq l$*.*

Proposition B.3. *Let* $\mathbf{S}_{XX} \sim \frac{1}{k-1} W_l\left(k-1, \mathbf{\Sigma}_{XX}\right)$ *and* $\mathbf{E} \in \mathbb{R}^{q \times l}$*, then it holds that*

$$\mathbf{E}\mathbf{S}_{XX}\mathbf{E}^T \sim \frac{1}{k-1} W_q\left(k-1, \mathbf{E}\mathbf{\Sigma}_{XX}\mathbf{E}^T\right). \tag{B.5}$$

Asymptotic Analysis of Random Matrix Theory

The analytical results for the random matrices, like Wishart matrices, are mainly categorized into two kinds, non-asymptotic results and asymptotic results. The asymptotic results mainly focus on the scenario that $k - 1 \to \infty$, $l \to \infty$ and $\frac{k-1}{l} \to \beta$, which allows closed-form expressions for analytical purposes. As mentioned before, the researches of the random matrix mainly focus on the spectral analysis. The unordered eigenvalue of Wishart matrices following $\frac{1}{k-1} W_l\left(k-1, \mathbf{I}_l\right)$ is characterized by the Marčenko-Pastur law [130, pp.7], which is given by

$$f_\beta(x) = \begin{cases} \max\{1-\beta, 0\}\delta(x) + \beta\frac{\sqrt{(b-x)(x-a)}}{2\pi x}, & \text{when } a \leq x \leq b \\ 0, & \text{elsewhere} \end{cases} \tag{B.6}$$

with

$$a = \left(1 - \sqrt{1/\beta}\right)^2, \qquad b = \left(1 + \sqrt{1/\beta}\right)^2. \tag{B.7}$$

For the maximum eigenvalue of the Wishart matrix $W_l\left(k-1, \mathbf{I}_l\right)$, it is shown in [131, Theorem 5.21] that the recentred and rescaled maximum eigenvalue W_{max} follows the Tracy-Widom law of order 1, that is,

$$\frac{\lambda_{max} - \mu_{max}}{\sigma_{max}} \xrightarrow{\mathcal{D}} W_{max} \sim F_1, \tag{B.8}$$

where λ_{max} is the maximum eigenvalue; μ_{max} and σ_{max} are the recenter mean and the rescale standard deviation for the maximum eigenvalue, respectively, which are

given by

$$\mu_{max} = \left(\sqrt{k-2} + \sqrt{l}\right)^2, \quad \sigma_{max} = \left(\sqrt{k-2} + \sqrt{l}\right)\left(1/\sqrt{l} + 1/\sqrt{k-2}\right)^{\frac{1}{3}}, \tag{B.9}$$

and F_1 is the Tracy-Widom law of order 1. For the minimum eigenvalue of the Wishart matrix $W_l(k-1, \mathbf{I}_l)$, it is shown in [133] that the recentered and rescaled minimum eigenvalue W_{min} follows the Tracy-Widom law of order 1, that is,

$$\frac{\lambda_{min} - \mu_{min}}{\sigma_{min}} \xrightarrow{\mathcal{D}} W_{min} \sim F_1, \tag{B.10}$$

where λ_{min} is the minimum eigenvalue; μ_{min} and σ_{min} are the recenter mean and the rescale standard deviation for the minimum eigenvalue, respectively, which are given by

$$\mu_{min} = \left(\sqrt{k - \frac{1}{2}} - \sqrt{l + \frac{1}{2}}\right)^2, \quad \sigma_{min} = \left(k - \frac{1}{2}\right)^{1/2}\left(l + \frac{1}{2}\right)^{-1/6}. \tag{B.11}$$

For the asymptotic scenario, a key point here is the rate of convergence of the distributions from the finite case to the asymptotic limit. Although there is no analytical solution to this question, the convergence of distribution is usually quite fast. Some simulations can help us to understand the rate of convergence. Fig. B.1 compares the histogram from Monte Carlo simulation with the Marčenko-Pastur law in (B.6) when $\beta = 2$, in which 10,000 realizations are generated. With n increases, the distribution of the unordered eigenvalues of Wishart distribution converges to the Marčenko-Pastur law. Especially when $l = 20$, these two distributions are quite close to each other. This implies that although the Marčenko-Pastur law is for asymptotic scenario, it fits the finite scenario quiet well.

Non-asymptotic Analysis of Random Matrix Theory

The non-asymptotic results for the random matrices mainly show the statistical behavior of the eigenvalues of the matrices when k and n are of finite numbers. Unlike the asymptotic scenario, the results in the non-asymptotic scenario usually provide bounds for the statistical behavior of the matrices, instead of the closed-form expression. As there are loads of research results on the non-asymptotic behavior of the random matrix, here only the results that are used in this thesis are introduced. For a random matrix with independent standard normal entries, the expected value and the variance of the singular value are bounded by the following theorem and corollary.

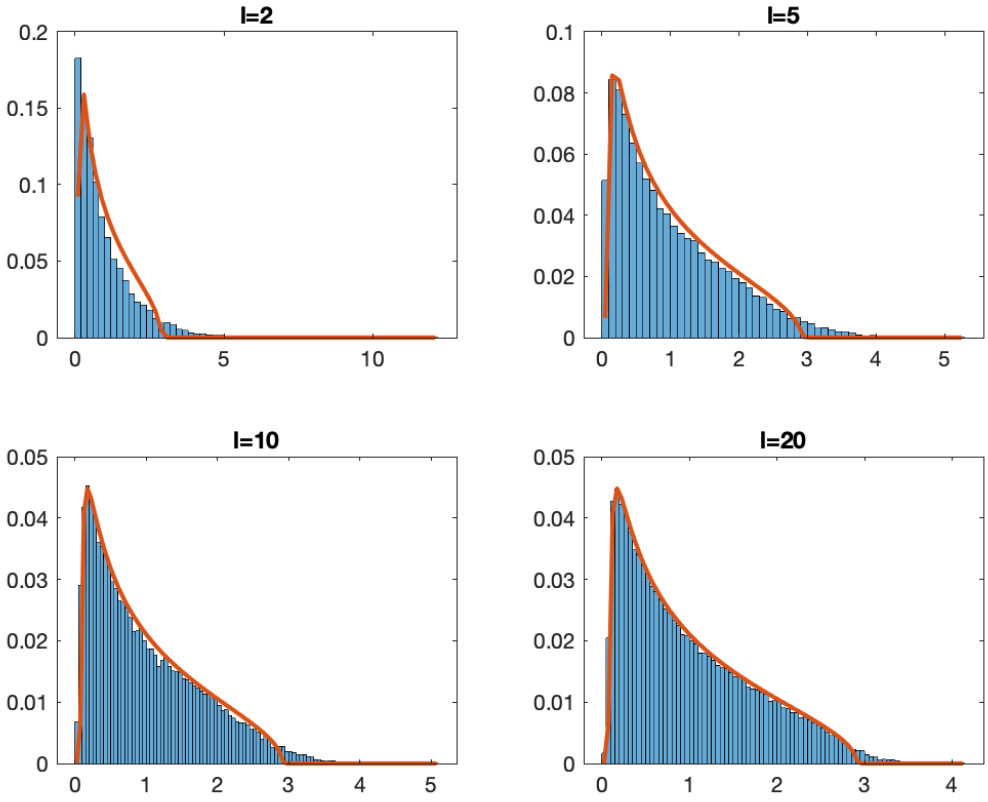


Fig. B.1. Comparison between the histogram from Monte Carlo simulation and the Marčenko-Pastur law in (B.6) when $\beta = 2$.

Theorem B.1 (Theorem 5.32 in [128]). *Let $\mathbf{Z}_l$ be a $(k-1)\times l$ matrix whose entries are independent standard normal random variables. Then*

$$\sqrt{k-1} - \sqrt{l} \leq \mathbb{E}\left[s_{\min}\left(\mathbf{Z}_l\right)\right] \leq \mathbb{E}\left[s_{\max}\left(\mathbf{Z}_l\right)\right] \leq \sqrt{k-1} + \sqrt{l}. \tag{B.12}$$

Corollary B.1 (Corollary 5.35 in [128]). *Let $\mathbf{Z}_l$ be a $(k-1)\times l$ matrix whose entries are independent standard normal random variables. Then for every $t \geq 0$, with probability at least $1 - 2\exp(-t^2/2)$ one has*

$$\sqrt{k-1} - \sqrt{l} - t \leq s_{\min}\left(\mathbf{Z}_l\right) \leq s_{\max}\left(\mathbf{Z}_l\right) \leq \sqrt{k-1} + \sqrt{l} + t. \tag{B.13}$$

Applications of Random Matrix Theory to Power Systems

Although RMT is a powerful tool, it has not been widely utilized to solve the power system problems. The power system is usually of big scale, so the state variables govern the states of the power system, or the measurements representing the condition

of the power system, are of high dimension. This makes the asymptotic results of RMT work quite well in the finite but large dimensional case, which is a result of the fast rate of convergence of asymptotic results. The works that employ random matrix theory tools to solve power system problem mainly include [134–137] and [138]. The single-ring law of the product of Gaussian random matrices is used in [134] to visualize the high-dimensional data in the power system. Also the validity of the Marčenko-Pastur Law, kernel density estimation, and the ring law on the IEEE test system is studied by [135], [136], and [137], respectively. Furthermore the correlation between the state variables in the power system is characterized in [138] via spectral analysis of random matrices.

Appendix C

Probability of False Alarm

Lemma C.1. *The probability of false alarm of the LRT in (3.11) for the attack construction in (4.7) is given by*

$$\mathsf{P}_{\mathsf{FA}}(\lambda) = \mathbb{P}\left[(U^p)^{\mathrm{T}}\bar{\mathbf{\Delta}}U^p \geq 2\log\tau + \log\left|\mathbf{I}_p + \lambda^{-1}\mathbf{\Delta}\right|\right], \tag{C.1}$$

where $\bar{\mathbf{\Delta}} \in \mathbb{R}^{p\times p}$ *is a diagonal matrix with entries given by* $(\bar{\mathbf{\Delta}})_{i,i} = \lambda_i(\mathbf{\Sigma}_{AA})\lambda_i(\mathbf{\Sigma}^{-1}_{Y_AY_A})$.

Proof. The probability of false alarm of the stealth attack is given by

$$\mathsf{P}_{\mathsf{FA}}(\lambda) = \int_{\tilde{\mathcal{S}}} \mathrm{d}P_{Y^m} \tag{C.2}$$

$$= \frac{1}{(2\pi)^{\frac{m}{2}}\left|\mathbf{\Sigma}_{YY}\right|^{\frac{1}{2}}} \int_{\tilde{\mathcal{S}}} \exp\left\{-\frac{1}{2}\mathbf{y}^{\mathrm{T}}\mathbf{\Sigma}^{-1}_{YY}\mathbf{y}\right\}\mathrm{d}\mathbf{y}, \tag{C.3}$$

where the integration domain is given by

$$\tilde{\mathcal{S}} = \{\mathbf{y}\in\mathbb{R}^m : L(\mathbf{y}) \geq \tau\} \tag{C.4}$$

$$= \{\mathbf{y}\in\mathbb{R}^m : \mathbf{y}^{\mathrm{T}}\mathbf{\Delta}_0\mathbf{y} \geq 2\log\tau + \log|\mathbf{I}_m + \mathbf{\Sigma}_{AA}\mathbf{\Sigma}^{-1}_{YY}|\} \tag{C.5}$$

with $\mathbf{\Delta}_0 \triangleq \mathbf{\Sigma}^{-1}_{YY} - \mathbf{\Sigma}^{-1}_{Y_AY_A}$. Applying the change of variable $\mathbf{y}_1 \triangleq \mathbf{U}_{YY}\mathbf{y}$ in (4.22) results in

$$\mathsf{P}_{\mathsf{FA}}(\lambda) = \frac{1}{(2\pi)^{\frac{m}{2}}\left|\mathbf{\Sigma}_{YY}\right|^{\frac{1}{2}}} \int_{\tilde{\mathcal{S}}_1} \exp\left\{-\frac{1}{2}\mathbf{y}_1^{\mathrm{T}}\mathbf{\Lambda}^{-1}_{YY}\mathbf{y}_1\right\}\mathrm{d}\mathbf{y}_1 \tag{C.6}$$

with the integration domain $\tilde{\mathcal{S}}_1$ given by

$$\tilde{\mathcal{S}}_1 = \left\{\mathbf{y}_1\in\mathbb{R}^m : \mathbf{y}_1^{\mathrm{T}}\mathbf{\Delta}_1\mathbf{y}_1 \geq 2\log\tau + \log|\mathbf{I}_m + \mathbf{\Lambda}_{AA}\mathbf{\Lambda}^{-1}_{YY}|\right\}, \tag{C.7}$$

where $\boldsymbol{\Delta}_1 \triangleq \boldsymbol{\Lambda}_{YY}^{-1} - \boldsymbol{\Lambda}_{Y_AY_A}^{-1}$. Further applying the change of variable $\mathbf{y}_2 \triangleq \boldsymbol{\Lambda}_{YY}^{-\frac{1}{2}}\mathbf{y}_1$ in (4.31) results in

$$\mathsf{P}_{\mathsf{FA}}(\lambda) = \frac{1}{\sqrt{(2\pi)^m}} \int_{\tilde{\mathcal{S}}_2} \exp\{-\frac{1}{2}\mathbf{y}_2^{\mathrm{T}}\mathbf{y}_2\}\mathrm{d}\mathbf{y}_2, \tag{C.8}$$

with the transformed integration domain given by

$$\tilde{\mathcal{S}}_2 = \left\{\mathbf{y}_2 \in \mathbb{R}^m : \mathbf{y}_2^{\mathrm{T}}\bar{\boldsymbol{\Delta}}\mathbf{y}_2 \geq 2\log\tau + \log|\mathbf{I}_m + \boldsymbol{\Delta}_2|\right\}, \tag{C.9}$$

with

$$\bar{\boldsymbol{\Delta}} \triangleq \boldsymbol{\Lambda}_{AA}\boldsymbol{\Lambda}_{Y_AY_A}^{-1}. \tag{C.10}$$

The proof completes. □

Appendix D

Non-asymptotic Lower Bound

To provide the non-asymptotic lower bound for the performance in (5.30), it boils down to providing upper bound for the last term in (5.30). The following lemma provides an upper bound for the last term in (5.30).

Lemma D.1. *Let* $\mathbf{W}_p$ *denote a central Wishart matrix distributed as* $\frac{1}{k-1}W_p(k-1, \mathbf{I}_p)$ *and let* $\mathbf{B} = \mathrm{diag}(b_1, \ldots, b_p)$ *denote a positive definite diagonal matrix. Then*

$$\mathbb{E}\left[\log\left|\mathbf{I}_p + \mathbf{B}\mathbf{W}_p\right|\right] \leq \sum_{i=1}^{p} \log\left(1 + b_i \tilde{x}_i^{\star}\right), \tag{D.1}$$

where $\tilde{x}_i^{\star}$ *is the solution to the convex optimization problem given by*

$$\max_{\{\tilde{x}_i\}_{i=1}^{p}} \sum_{i=1}^{p} \log\left(1 + b_i \tilde{x}_i\right) \tag{D.2}$$

$$s.t. \quad \sum_{i=1}^{p} \tilde{x}_i = p \tag{D.3}$$

$$\max\left(\tilde{x}_i\right) \leq \left(1 + \sqrt{p/(k-1)}\right)^2 + 1/(k-1) \tag{D.4}$$

$$\min\left(\tilde{x}_i\right) \geq \left(1 - \sqrt{p/(k-1)}\right)^2. \tag{D.5}$$

Proof. Note that

$$\mathbb{E}\left[\log\left|\mathbf{I}_p + \mathbf{B}\mathbf{W}_p\right|\right] = \sum_{i=1}^{p} \mathbb{E}\left[\log\left(1 + b_i \lambda_i(\mathbf{W}_p)\right)\right] \tag{D.6}$$

$$\leq \sum_{i=1}^{p} \log\left(1 + b_i \mathbb{E}\left[\lambda_i(\mathbf{W}_p)\right]\right) \tag{D.7}$$

where (D.7) follows from Jensen's inequality due to the concavity of $\log\left(1 + b_i x\right)$ for $x > 0$. Constraint (D.3) follows from the fact that $\mathbb{E}[\mathrm{trace}(\mathbf{W}_p)] = p$, and constraints (D.4) and (D.5) follow from Lemma 5.4. This completes the proof.

□

The following theorem characterizes the lower bound for the performance in (5.30).

Theorem D.1. *The ergodic attack performance given in (5.30) is upper bounded by*

$$\mathbb{E}\left[f(\mathbf{\Sigma}_{\tilde{A}\tilde{A}})\right] \geq \frac{1}{2}\left(tr\left(\mathbf{\Sigma}_{YY}^{-1}\mathbf{\Sigma}_{AA}^{\star}\right) - \log\left|\mathbf{\Sigma}_{YY}^{-1}\right| - 2m\log\sigma - \sum_{i=1}^{p}\log\left(1+\frac{\lambda_i}{\sigma^2}\tilde{\lambda}_i^{\star}\right)\right), \quad (D.8)$$

where $\{\lambda_i^{\star}\}_{i=1}^{p}$ *is the solution to the optimization problem given by (D.2) - (D.5) with* $b_i = \frac{\lambda_i}{\sigma^2}$

Proof. The proof follows immediately from combing Lemma D.1 with (5.30). □

References

[1] J. D. Glover, M. S. Sarma, and T. Overbye, *Power System Analysis & design, SI version*, Cengage Learning, Aug. 2012.

[2] U.S.-Canada Power System Outage Task Force, *Final Report on the August 14, 2003 Blackout in the United States and Canada: Causes and Recommendations*, 2004.

[3] International Energy Agency, "Electricity statistics," https://www.iea.org/statistics/electricity/, 2018.

[4] H. Gharavi and R. Ghafurian, "Smart grid: The electric energy system of the future," *Proc. IEEE*, vol. 99, no. 6, pp. 917–921, Jun. 2011.

[5] Energy Department of U.S.A, "Grid modernization and the smart grid," https://www.energy.gov/oe/activities/technology-development/grid-modernization-and-smart-grid.

[6] Environmental Protection Agency of United States, "Distributed generation of electricity and its environmental impacts," https://www.epa.gov/energy/distributed-generation-electricity-and-its-environmental-impacts.

[7] Y. Liu, P. Ning, and M. K. Reiter, "False data injection attacks against state estimation in electric power grids," in *Proc. ACM Conf. on Computer and Communications Security*, Chicago, IL, USA, Nov. 2009, pp. 21–32.

[8] Y. Liu, P. Ning, and M. K. Reiter, "False data injection attacks against state estimation in electric power grids," *ACM Trans. Inf. Syst. Secur.*, vol. 14, no. 1, pp. 13:1–13:33, Jun. 2011.

[9] D. Alderson and R. Di Pietro, "Operational technology: Are you vulnerable?," *Governance Directions*, vol. 68, no. 6, pp. 339–343, Jul. 2016.

[10] British Broadcasting Corporation, "Ukraine power cut 'was cyber-attack'," https://www.bbc.co.uk/news/technology-38573074.

[11] Royal Society, *Progress and Research in Cybersecurity: Supporting a Resilient and Trustworthy System for the UK*, 2016.

[12] Q. Li, T. Cui, Y. Weng, R. Negi, F. Franchetti, and M. D. Ilić, "An information-theoretic approach to PMU placement in electric power systems," *IEEE Trans. Smart Grid*, vol. 4, no. 1, pp. 446–456, Mar. 2013.

[13] D. Varodayan and A. Khisti, "Smart meter privacy using a rechargeable battery: Minimizing the rate of information leakage," in *Proc. IEEE Int. Conf. on Acoust., Speech and Signal Process.*, Prague, Czech Republic, May 2011, pp. 1932–1935.

[14] L. Sankar, S.R. Rajagopalan, S. Mohajer, and H.V. Poor, "Smart meter privacy: A theoretical framework," *IEEE Trans. Smart Grid*, vol. 4, no. 2, pp. 837–846, Jun. 2013.

[15] O. Tan, D. Gündüz, and H. V. Poor, "Increasing smart meter privacy through energy harvesting and storage devices," *IEEE J. Sel. Areas Commun.*, vol. 31, no. 7, pp. 1331–1341, Jul. 2013.

[16] M. Arrieta and I. Esnaola, "Smart meter privacy via the trapdoor channel," in *Proc. IEEE Int. Conf. on Smart Grid Commun.*, Dresden, Germany, Oct. 2017, pp. 227–282.

[17] J. J. Grainger and W. D. Stevenson, *Power System Analysis*, McGraw-Hill, 1994.

[18] S. Cui, Z. Han, S. Kar, T. T. Kim, H. V. Poor, and A. Tajer, "Coordinated data-injection attack and detection in the smart grid: A detailed look at enriching detection solutions," *IEEE Signal Process. Mag.*, vol. 29, no. 5, pp. 106–115, Sep. 2012.

[19] A. Tajer, S. Kar, H. V. Poor, and S. Cui, "Distributed joint cyber attack detection and state recovery in smart grids," in *Proc. IEEE Int. Conf. on Smart Grid Commun.*, Brussels, Belgium, Oct. 2011, pp. 202–207.

[20] A. Abur and A. G. Expósito, *Power System State Estimation: Theory and Implementation*, CRC Press, Mar. 2004.

[21] H. V. Poor, *An Introduction to Signal Detection and Estimation*, Springer, New York, 1994.

[22] T. M. Cover and J. A. Thomas, *Elements of Information Theory*, John Wiley & Sons, Nov. 2012.

[23] G. B. Giannakis, V. Kekatos, N. Gatsis, S. Kim, H. Zhu, and B. F. Wollenberg, "Monitoring and optimization for power grids: A signal processing perspective," *IEEE Signal Process. Mag.*, vol. 30, no. 5, pp. 107–128, Sep. 2013.

[24] H. Zhu and G. B. Giannakis, "Estimating the state of AC power systems using semidefinite programming," in *Proc. North American Power Symp.*, Boston, MA, USA, Aug. 2011, pp. 1–7.

[25] H. Zhu and G. B. Giannakis, "Multi-area state estimation using distributed SDP for nonlinear power systems," in *Proc. IEEE Int. Conf. on Smart Grid Commun.*, Tainan, Taiwan, Nov. 2012, pp. 623–628.

[26] A. Monticelli, "Electric power system state estimation," *Proc. IEEE*, vol. 88, no. 2, pp. 262–282, Feb. 2000.

[27] G. Liang, S. R. Weller, J. Zhao, F. Luo, and Z. Y. Dong, "The 2015 Ukraine blackout: Implications for false data injection attacks," *IEEE Trans. Power Syst.*, vol. 32, no. 4, pp. 3317–3318, Jul. 2017.

[28] A. Giani, S. Sastry, K. H. Johansson, and H. Sandberg, "The VIKING project: An initiative on resilient control of power networks," in *Proc. 2nd Int. Symp. on Resilient Control Syst.*, Idaho Falls, ID, USA, Aug. 2009, pp. 31–35.

[29] H. Sandberg, S. Amin, and K. H. Johansson, "Cyberphysical security in networked control systems: An introduction to the issue," *IEEE Control Syst. Mag.*, vol. 35, no. 1, pp. 20–23, Feb. 2015.

[30] Y. Mo and B. Sinopoli, "False data injection attacks in control systems," in *Proc. 1st Workshop on Secure Control Syst.*, Stockholm, Sweden, Apr. 2010.

[31] R. Deng, P. Zhuang, and H. Liang, "False data injection attacks against state estimation in power distribution systems," *IEEE Trans. Smart Grid*, vol. 10, no. 3, pp. 2871–2881, May 2019.

[32] G. Hug and J. A. Giampapa, "Vulnerability assessment of AC state estimation with respect to false data injection cyber-attacks," *IEEE Trans. Smart Grid*, vol. 3, no. 3, pp. 1362–1370, Sep. 2012.

[33] S. G. Mallat and Z. Zhang, "Matching pursuits with time-frequency dictionaries," *IEEE Trans. Signal Process.*, vol. 41, no. 12, pp. 3397–3415, Dec. 1993.

[34] Y. C. Pati, R. Rezaiifar, and P. S. Krishnaprasad, "Orthogonal matching pursuit: Recursive function approximation with applications to wavelet decomposition," in *Proc. 27th Asilomar Conf. on Signals, Syst. and Comput.*, Pacific Grove, CA, USA, Nov. 1993, pp. 40–44.

[35] G. Dán and H. Sandberg, "Stealth attacks and protection schemes for state estimators in power systems," in *Proc. IEEE Int. Conf. on Smart Grid Commun.*, Gaithersburg, MD, USA, Oct. 2010, pp. 214–219.

[36] Q. Yang, J. Yang, W. Yu, D. An, N. Zhang, and W. Zhao, "On false data-injection attacks against power system state estimation: Modeling and countermeasures," *IEEE Trans. Parallel Distrib. Syst.*, vol. 25, no. 3, pp. 717–729, Mar. 2014.

[37] S. Bi and Y. J. Zhang, "Defending mechanisms against false-data injection attacks in the power system state estimation," in *Proc. IEEE Global Commun. Conf. Workshops*, Houston, TX, USA, Dec. 2011, pp. 1162–1167.

[38] S. Bi and Y. J. Zhang, "Graphical methods for defense against false-data injection attacks on power system state estimation," *IEEE Trans. Smart Grid*, vol. 5, no. 3, pp. 1216–1227, May 2014.

[39] T. T. Kim and H. V. Poor, "Strategic protection against data injection attacks on power grids," *IEEE Trans. Smart Grid*, vol. 2, no. 2, pp. 326–333, Jun. 2011.

[40] E. J. Candès, M. B. Wakin, and S. P. Boyd, "Enhancing sparsity by reweighted ℓ_1 minimization," *Journal of Fourier Analysis and Applications*, vol. 14, no. 5-6, pp. 877–905, Oct. 2008.

[41] H. Sandberg, A. Teixeira, and K. H. Johansson, "On security indices for state estimators in power networks," in *Proc. 1st Workshop on Secure Control Syst.*, Stockholm, Sweden, Apr. 2010.

[42] K. C. Sou, H. Sandberg, and K. H. Johansson, "On the exact solution to a smart grid cyber-security analysis problem," *IEEE Trans. Smart Grid*, vol. 4, no. 2, pp. 856–865, Jun. 2013.

[43] M. Ozay, I. Esnaola, F. T. Y. Vural, S. R. Kulkarni, and H. V. Poor, "Sparse attack construction and state estimation in the smart grid: Centralized and distributed models," *IEEE J. Sel. Areas Commun.*, vol. 31, no. 7, pp. 1306–1318, Jul. 2013.

[44] O. Kosut, L. Jia, R. J. Thomas, and L. Tong, "Malicious data attacks on the smart grid," *IEEE Trans. Smart Grid*, vol. 2, no. 4, pp. 645–658, Dec. 2011.

[45] O. Vuković, K. C. Sou, G. Dán, and H. Sandberg, "Network-aware mitigation of data integrity attacks on power system state estimation," *IEEE J. Sel. Areas Commun.*, vol. 30, no. 6, pp. 1108–1118, Jul. 2012.

[46] M. A. Rahman and H. Mohsenian-Rad, "False data injection attacks with incomplete information against smart power grids," in *Proc. IEEE Global Commun. Conf.*, Anaheim, CA, USA, Dec. 2012, pp. 3153–3158.

[47] X. Liu and Z. Li, "Local load redistribution attacks in power systems with incomplete network information," *IEEE Trans. Smart Grid*, vol. 5, no. 4, pp. 1665–1676, Jul. 2014.

[48] X. Liu, Z. Bao, D. Lu, and Z. Li, "Modeling of local false data injection attacks with reduced network information," *IEEE Trans. Smart Grid*, vol. 6, no. 4, pp. 1686–1696, Jul. 2015.

[49] J. Kim, L. Tong, and R. J. Thomas, "Subspace methods for data attack on state estimation: A data driven approach," *IEEE Trans. Signal Process.*, vol. 63, no. 5, pp. 1102–1114, Mar. 2015.

[50] X. Liu and Z. Li, "False data attacks against ac state estimation with incomplete network information," *IEEE Trans. Smart Grid*, vol. 8, no. 5, pp. 2239–2248, Sep. 2017.

[51] A. Tajer, "False data injection attacks in electricity markets by limited adversaries: Stochastic robustness," *IEEE Trans. Smart Grid*, vol. 10, no. 1, pp. 128–138, Jan. 2019.

[52] A. Teixeira, H. Sandberg, G. Dán, and K. H. Johansson, "Optimal power flow: Closing the loop over corrupted data," in *Proc. Amer. Control Conf.*, Montreal, Canada, Jun. 2012, pp. 3534–3540.

[53] M. A. Rahman, E. Al-Shaer, and R. G. Kavasseri, "A formal model for verifying the impact of stealthy attacks on optimal power flow in power grids," in *Proc. ACM/IEEE Int. Conf. on Cyber-Physical Syst.*, Berlin, Germany, Apr. 2014, pp. 175–186.

[54] L. Xie, Y. Mo, and B. Sinopoli, "False data injection attacks in electricity markets," in *Proc. IEEE Int. Conf. on Smart Grid Commun.*, Gaithersburg, MD, USA, Oct. 2010, pp. 226–231.

[55] L. Jia, J. Kim, R. J. Thomas, and L. Tong, "Impact of data quality on real-time locational marginal price," *IEEE Trans. Power Syst.*, vol. 29, no. 2, pp. 627–636, Mar. 2014.

[56] S. Tan, W. Song, M. Stewart, J. Yang, and L. Tong, "Online data integrity attacks against real-time electrical market in smart grid," *IEEE Trans. Smart Grid*, vol. 9, no. 1, pp. 313–322, Jan. 2018.

[57] G. Liang, S. R. Weller, F. Luo, J. Zhao, and Z. Y. Dong, "Generalized FDIA-based cyber topology attack with application to the Australian electricity market trading mechanism," *IEEE Trans. Smart Grid*, vol. 9, no. 4, pp. 3820–3829, Jul. 2018.

[58] X. Liu, Z. Li, X. Liu, and Z. Li, "Masking transmission line outages via false data injection attacks," *IEEE Trans. Inf. Forensics Security*, vol. 11, no. 7, pp. 1592–1602, Jul. 2016.

[59] J. Zhang and L. Sankar, "Physical system consequences of unobservable state-and-topology cyber-physical attacks," *IEEE Trans. Smart Grid*, vol. 7, no. 4, pp. 2016–2025, Jul. 2016.

[60] M. Esmalifalak, G. Shi, Z. Han, and L. Song, "Bad data injection attack and defense in electricity market using game theory study," *IEEE Trans. Smart Grid*, vol. 4, no. 1, pp. 160–169, Mar. 2013.

[61] A. Sanjab and W. Saad, "Data injection attacks on smart grids with multiple adversaries: A game-theoretic perspective," *IEEE Trans. Smart Grid*, vol. 7, no. 4, pp. 2038–2049, Jul. 2016.

[62] A. Sanjab and W. Saad, "Smart grid data injection attacks: To defend or not?," in *Proc. IEEE Int. Conf. on Smart Grid Commun.*, Miami, FL, USA, Nov. 2015, pp. 380–385.

[63] J. Kim and L. Tong, "On topology attack of a smart grid: Undetectable attacks and countermeasures," *IEEE J. Sel. Areas Commun.*, vol. 31, no. 7, pp. 1294–1305, Jul. 2013.

[64] Y. Wu, Z. Wei, J. Weng, X. Li, and R. H. Deng, "Resonance attacks on load frequency control of smart grids," *IEEE Trans. Smart Grid*, vol. 9, no. 5, pp. 4490–4502, Sep. 2018.

[65] S. Amini, F. Pasqualetti, and H. Mohsenian-Rad, "Dynamic load altering attacks against power system stability: Attack models and protection schemes," *IEEE Trans. Smart Grid*, vol. 9, no. 4, pp. 2862–2872, Jul. 2018.

[66] S. Barreto, M. Pignati, G. Dán, J. Le Boudec, and M. Paolone, "Undetectable timing-attack on linear state-estimation by using rank-1 approximation," *IEEE Trans. Smart Grid*, vol. 9, no. 4, pp. 3530–3542, Jul. 2018.

[67] J. Lin, W. Yu, X. Yang, G. Xu, and W. Zhao, "On false data injection attacks against distributed energy routing in Smart Grid," in *Proc. IEEE/ACM 3rd Int. Conf. on Cyber-Physical Sys.*, Beijing, China, Apr. 2012, pp. 183–192.

[68] H. Zhang, W. Meng, J. Qi, X. Wang, and W. X. Zheng, "Distributed load sharing under false data injection attack in an inverter-based microgrid," *IEEE Trans. Ind. Electron.*, vol. 66, no. 2, pp. 1543–1551, Feb. 2019.

[69] S. Pal, B. Sikdar, and J. H. Chow, "Classification and detection of pmu data manipulation attacks using transmission line parameters," *IEEE Trans. Smart Grid*, vol. 9, no. 5, pp. 5057–5066, Sep. 2018.

[70] Y. Chakhchoukh and H. Ishii, "Enhancing robustness to cyber-attacks in power systems through multiple least trimmed squares state estimations," *IEEE Trans. Power Syst.*, vol. 31, no. 6, pp. 4395–4405, Nov. 2016.

[71] S. Sridhar and M. Govindarasu, "Model-based attack detection and mitigation for automatic generation control," *IEEE Trans. Smart Grid*, vol. 5, no. 2, pp. 580–591, Mar. 2014.

[72] A. Ashok, M. Govindarasu, and V. Ajjarapu, "Online detection of stealthy false data injection attacks in power system state estimation," *IEEE Trans. Smart Grid*, vol. 9, no. 3, pp. 1636–1646, May 2018.

[73] Y. Zhao, A. Goldsmith, and H. V. Poor, "Minimum sparsity of unobservable power network attacks," *IEEE Trans. Automat. Contr.*, vol. 62, no. 7, pp. 3354–3368, Jul. 2017.

[74] S. Wang and W. Ren, "Stealthy false data injection attacks against state estimation in power systems: Switching network topologies," in *Proc. Amer. Control Conf.*, Portland, OR, USA, Jun. 2014, pp. 1572–1577.

[75] T. Liu, Y. Gu, D. Wang, Y. Gui, and X. Guan, "A novel method to detect bad data injection attack in smart grid," in *Proc. IEEE Conf. on Comput. Commun. Workshops*, Turin, Italy, Apr. 2013, pp. 49–54.

[76] T. R. Nudell, S. Nabavi, and A. Chakrabortty, "A real-time attack localization algorithm for large power system networks using graph-theoretic techniques," *IEEE Trans. Smart Grid*, vol. 6, no. 5, pp. 2551–2559, Sep. 2015.

[77] Y. Guan and X. Ge, "Distributed attack detection and secure estimation of networked cyber-physical systems against false data injection attacks and jamming attacks," *IEEE Trans. Signal Inf. Process. Netw.*, vol. 4, no. 1, pp. 48–59, Mar. 2018.

[78] K. Manandhar, X. Cao, F. Hu, and Y. Liu, "Detection of faults and attacks including false data injection attack in smart grid using kalman filter," *IEEE Trans. Control of Netw. Syst.*, vol. 1, no. 4, pp. 370–379, Dec. 2014.

[79] Y. Mo, S. Weerakkody, and B. Sinopoli, "Physical authentication of control systems: Designing watermarked control inputs to detect counterfeit sensor outputs," *IEEE Control Syst. Mag.*, vol. 35, no. 1, pp. 93–109, Feb. 2015.

[80] Q. Yang, D. An, R. Min, W. Yu, X. Yang, and W. Zhao, "On optimal pmu placement-based defense against data integrity attacks in smart grid," *IEEE Trans. Inf. Forensics Security*, vol. 12, no. 7, pp. 1735–1750, Jul. 2017.

[81] B.R. Bobba, M.K. Rogers, Q. Wang, H. Khurana, and Overbye J.T. Nahrstedt, K. and, "Detecting false data injection attacks on DC state estimation," in *Proc. 1st Workshop on Secure Control Syst.*, Stockholm, Sweden, Apr. 2010.

[82] K. C. Sou, H. Sandberg, and K. H. Johansson, "Data attack isolation in power networks using secure voltage magnitude measurements," *IEEE Trans. Smart Grid*, vol. 5, no. 1, pp. 14–28, Jan. 2014.

[83] A. Abdallah and X. S. Shen, "Efficient prevention technique for false data injection attack in smart grid," in *Proc. IEEE Int. Conf. on Commun.*, Kuala Lumpur, Malaysia, May 2016, pp. 1–6.

[84] F. Miao, Q. Zhu, M. Pajic, and G. J. Pappas, "Coding schemes for securing cyber-physical systems against stealthy data injection attacks," *IEEE Trans. Control of Netw. Syst.*, vol. 4, no. 1, pp. 106–117, Mar. 2017.

[85] L. Che, X. Liu, and Z. Li, "Mitigating false data attacks induced overloads using a corrective dispatch scheme," *IEEE Trans. Smart Grid*, vol. PP, no. 99, pp. 1–1, 2018.

[86] I. Esnaola, S. M. Perlaza, H. V. Poor, and O. Kosut, "Maximum distortion attacks in electricity grids," *IEEE Trans. Smart Grid*, vol. 7, no. 4, pp. 2007–2015, Jul. 2016.

[87] K. Sun, I. Esnaola, S.M. Perlaza, and H.V. Poor, "Information-theoretic attacks in the smart grid," in *Proc. IEEE Int. Conf. on Smart Grid Commun.*, Dresden, Germany, Oct. 2017, pp. 455–460.

[88] K. Sun, I. Esnaola, S.M. Perlaza, and H.V. Poor, "Stealth attacks on the smart grid," *IEEE Trans. Smart Grid*, 2019.

[89] R. Zhang and P. Venkitasubramaniam, "Stealthy control signal attacks in linear quadratic gaussian control systems: Detectability reward tradeoff," *IEEE Trans. Inf. Forensics Security*, vol. 12, no. 7, pp. 1555–1570, Jul. 2017.

[90] N. Forti, G. Battistelli, L. Chisci, and B. Sinopoli, "A bayesian approach to joint attack detection and resilient state estimation," in *Proc. IEEE Conf. on Decision and Control*, Las Vegas, NV, USA, Dec. 2016, pp. 1192–1198.

[91] X. Li, H. V. Poor, and A. Scaglione, "Blind topology identification for power systems," in *Proc. IEEE Int. Conf. on Smart Grid Commun.*, Vancouver, Canada, Oct. 2013, pp. 91–96.

[92] Z. H. Yu and W. L. Chin, "Blind false data injection attack using PCA approximation method in smart grid," *IEEE Trans. Smart Grid*, vol. 6, no. 3, pp. 1219–1226, May 2015.

[93] M. Esmalifalak, H. Nguyen, R. Zheng, L. Xie, L. Song, and Z. Han, "A stealthy attack against electricity market using independent component analysis," *IEEE Syst. J.*, vol. 12, no. 1, pp. 297–307, Mar. 2018.

[94] S. Xie, J. Yang, K. Xie, Y. Liu, and Z. He, "Low-sparsity unobservable attacks against smart grid: Attack exposure analysis and a data-driven attack scheme," *IEEE Access*, vol. 5, pp. 8183–8193, Mar. 2017.

[95] G. Chaojun, P. Jirutitijaroen, and M. Motani, "Detecting false data injection attacks in AC state estimation," *IEEE Trans. Smart Grid*, vol. 6, no. 5, pp. 2476–2483, Sep. 2015.

[96] R. Moslemi, A. Mesbahi, and J. M. Velni, "A fast, decentralized covariance selection-based approach to detect cyber attacks in smart grids," *IEEE Trans. Smart Grid*, vol. 9, no. 5, pp. 4930–4941, Sep. 2018.

[97] S. Li, Y. Yılmaz, and X. Wang, "Quickest detection of false data injection attack in wide-area smart grids," *IEEE Trans. Smart Grid*, vol. 6, no. 6, pp. 2725–2735, Nov. 2015.

[98] B. Tang, and S. Kay, and H. He, "Detection of false data injection attacks in smart grid under colored Gaussian noise," in *Proc. IEEE Conf. on Commun. and Network Security*, Philadelphia, PA, USA, Oct. 2016, pp. 172–179.

[99] Y. Hao, M. Wang, and J. H. Chow, "Likelihood analysis of cyber data attacks to power systems with Markov decision processes," *IEEE Trans. Smart Grid*, vol. 9, no. 4, pp. 3191–3202, Jul. 2018.

[100] L. Liu, M. Esmalifalak, and Z. Han, "Detection of false data injection in power grid exploiting low rank and sparsity," in *Proc. IEEE Int. Conf. on Commun.*, Budapest, Hungary, Jun. 2013, pp. 4461–4465.

[101] L. Liu, M. Esmalifalak, Q. Ding, V. A. Emesih, and Z. Han, "Detecting false data injection attacks on power grid by sparse optimization," *IEEE Trans. Smart Grid*, vol. 5, no. 2, pp. 612–621, Mar. 2014.

[102] C. Liu, J. Wu, C. Long, and Y. Wang, "Dynamic state recovery for cyber-physical systems under switching location attacks," *IEEE Trans. Control of Netw. Syst.*, vol. 4, no. 1, pp. 14–22, Mar. 2017.

[103] R. C. Borges Hink, J. M. Beaver, M. A. Buckner, T. Morris, U. Adhikari, and S. Pan, "Machine learning for power system disturbance and cyber-attack discrimination," in *Proc. Int. Symp. on Resilient Control Syst.*, Denver, CO, USA, Aug. 2014, pp. 1–8.

[104] M. Esmalifalak, L. Liu, N. Nguyen, R. Zheng, and Z. Han, "Detecting stealthy false data injection using machine learning in smart grid," *IEEE Syst. J.*, vol. 11, no. 3, pp. 1644–1652, Sep. 2017.

[105] S. Pan, T. Morris, and U. Adhikari, "Developing a hybrid intrusion detection system using data mining for power systems," *IEEE Trans. Smart Grid*, vol. 6, no. 6, pp. 3104–3113, Nov. 2015.

[106] Y. Chakhchoukh, S. Liu, M. Sugiyama, and H. Ishii, "Statistical outlier detection for diagnosis of cyber attacks in power state estimation," in *Proc. IEEE Power and Energy Soc. General Meeting*, Boston, MA, USA, Jul. 2016, pp. 1–5.

[107] Y. Wang, M. M. Amin, J. Fu, and H. B. Moussa, "A novel data analytical approach for false data injection cyber-physical attack mitigation in smart grids," *IEEE Access*, vol. 5, pp. 26022–26033, 2017.

[108] Y. He, G. J. Mendis, and J. Wei, "Real-time detection of false data injection attacks in smart grid: A deep learning-based intelligent mechanism," *IEEE Trans. Smart Grid*, vol. 8, no. 5, pp. 2505–2516, Sep. 2017.

[109] J. J. Q. Yu, Y. Hou, and V. O. K. Li, "Online false data injection attack detection with wavelet transform and deep neural networks," *IEEE Trans. Ind. Informat.*, vol. 14, no. 7, pp. 3271–3280, Jul. 2018.

[110] A. Tajer, S. Sihag, and K. Alnajjar, "Non-linear state recovery in power system under bad data and cyber attacks," *Journal of Modern Power Systems and Clean Energy*, Jul. 2019.

[111] S. Sihag and A. Tajer, "Power system state estimation under model uncertainty," *IEEE J. Sel. Topics Signal Process.*, vol. 12, no. 4, pp. 593–606, Aug. 2018.

[112] C. Genes, I. Esnaola, S. M. Perlaza, L. F. Ochoa, and D. Coca, "Recovering missing data via matrix completion in electricity distribution systems," in *Proc. IEEE Workshop on Signal Process. Advances in Wireless Commun.*, Edinburgh, UK, Jul. 2016, pp. 1–6.

[113] A.K. Ghosh, D.L. Lubkeman, M.J. Downey, and R.H. Jones, "Distribution circuit state estimation using a probabilistic approach," *IEEE Trans. Power Syst.*, vol. 12, no. 1, pp. 45–51, Feb. 1997.

[114] N. C. Woolley and J. V. Milanovic, "Statistical Estimation of the Source and Level of Voltage Unbalance in Distribution Networks," *IEEE Trans. Power Del.*, vol. 27, no. 3, pp. 1450–1460, Jul. 2012.

[115] I. Shomorony and A. S. Avestimehr, "Worst-case additive noise in wireless networks," *IEEE Trans. Inf. Theory*, vol. 59, no. 6, pp. 3833–3847, Jun. 2013.

[116] J. Neyman and E. S. Pearson, "On the problem of the most efficient tests of statistical hypotheses," in *Breakthroughs in Statistics*, Springer Series in Statistics, pp. 73–108. Springer New York, 1992.

[117] J. Denzler and C. M. Brown, "Information theoretic sensor data selection for active object recognition and state estimation," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 24, no. 2, pp. 145–157, Feb. 2002.

[118] D. Guo, S. Shamai, and S. Verdú, "Mutual information and minimum mean-square error in Gaussian channels," *IEEE Trans. on Inf. Theory*, vol. 51, no. 4, pp. 1261–1282, Apr. 2005.

[119] J. Hou and G. Kramer, "Effective secrecy: Reliability, confusion and stealth," in *Proc. IEEE Int. Symp. on Infor. Theory*, Honolulu, HI, USA, Jun. 2014, pp. 601–605.

[120] S. Boyd and L. Vandenberghe, *Convex Optimization*, Cambridge University Press, Mar. 2004.

[121] G. A. F Seber, *A Matrix Handbook for Statisticians*, John Wiley & Sons, 2008.

[122] R. D. Zimmerman, C. E. Murillo-Sánchez, and R. J. Thomas, "MATPOWER: Steady-state operations, planning, and analysis tools for power systems research and education," *IEEE Trans. Power Syst.*, vol. 26, no. 1, pp. 12–19, Feb. 2011.

[123] I. Esnaola, A. M. Tulino, and J. Garcia-Frias, "Linear analog coding of correlated multivariate Gaussian sources," *IEEE Trans. Commun.*, vol. 61, no. 8, pp. 3438–3447, Jun. 2013.

[124] D. A. Bodenham and N. M. Adams, "A comparison of efficient approximations for a weighted sum of chi-squared random variables," *Stat Comput*, vol. 26, no. 4, pp. 917–928, Jul. 2016.

[125] B. G. Lindsay, R. S. Pilla, and P. Basak, "Moment-based approximations of distributions using mixtures: Theory and applications," *Ann. Inst. Stat. Math.*, vol. 52, no. 2, pp. 215–230, Jun. 2000.

[126] D. Hsu, S.M. Kakade, and T. Zhang, "A tail inequality for quadratic forms of subgaussian random vectors," *Electron. Commun. in Probab.*, vol. 17, no. 52, pp. 1–6, 2012.

[127] D. Bodenham, "*Momentchi2: Moment-Matching Methods for Weighted Sums of Chi-Squared Random Variables*. (2016) [Online]," Available: https://cran.r-project.org/web/packages/momentchi2/index.html.

[128] R. Vershynin, "Introduction to the non-asymptotic analysis of random matrices," in *Compressed Sensing: Theory and Applications*, Y. Eldar and G. Kutyniok, Eds., chapter 5, pp. 210–268. Cambridge University Press, Cambridge, UK, 2012.

[129] G. Alfano, A. M. Tulino, A. Lozano, and S. Verdú, "Capacity of MIMO channels with one-sided correlation," in *Proc. IEEE Int. Symp. on Spread Spectrum Techn. and Appl.*, Sydney, Australia, Aug. 2004.

[130] A. M. Tulino and S. Verdú, *Random Matrix Theory and Wireless Communications*, Now Publishers Inc, 2004.

[131] Z. Bai and J. W. Silverstein, *Spectral Analysis of Large Dimensional Random Matrices*, Springer Series in Statistics. Springer-Verlag, New York, 2 edition, 2010.

[132] M. Bilodeau and D. Brenner, *Theory of Multivariate Statistics*, Springer Science & Business Media, Jan. 2008.

[133] D. Paul, "Asymptotic distribution of the smallest eigenvalue of Wishart (N, n) when $N, n \to \infty$ such that $N/n \to 0$," in *Nonparametric Statistical Methods and Related Topics*, pp. 423–458. Word Scientific, Sep. 2011.

[134] D. Cai, X. He, Z. Yu, L. Wang, G. Xie, and Q. Ai, "3d power-map for smart grids - An integration of high-dimensional analysis and visualization," in *Proc. Int. Conf. on Renewable Power Generation*, Beijing, China, Oct. 2015, pp. 1–5.

[135] X. He, R. C. Qiu, Q. Ai, L. Chu, X. Xu, and Z. Ling, "Designing for situation awareness of future power grids: An indicator system based on linear eigenvalue statistics of large random matrices," *IEEE Access*, vol. 4, pp. 3557–3568, 2016.

[136] X. He, Q. Ai, R. C. Qiu, W. Huang, L. Piao, and H. Liu, "A big data architecture design for smart grids based on random matrix theory," *IEEE Trans. Smart Grid*, vol. 8, no. 2, pp. 674–686, Mar. 2017.

[137] X. He, L. Chu, R. C. Qiu, Q. Ai, and Z. Ling, "A novel data-driven situation awareness approach for future grids-Using large random matrices for big data modeling," *IEEE Access*, vol. 6, pp. 13855–13865, 2018.

[138] X. Xu, X. He, Q. Ai, and R. C. Qiu, "A correlation analysis method for power systems based on random matrix theory," *IEEE Trans. Smart Grid*, vol. 8, no. 4, pp. 1811–1820, Jul. 2017.