Security in Power System State Estimation

Ankur Majumdar

Thesis submitted for the degree of

Doctor of Philosophy

Imperial College London

Department of Electrical and Electronic Engineering

Control and Power Research Group

May 2016

Declaration of Originality

I hereby certify that all material in this dissertation is my own work, and, where appropriate, contributions from other people have been properly acknowledged.

Ankur Majumdar

May 2016

Copyright Declaration

Abstract

With the power system evolving from passive to a more active system there is an incorporation of information and communication infrastructures in the system. The measurement data are more prone to tampering from attackers for mala fide intentions. Therefore, security and reliability of distribution have become major concerns. State estimation (SE), being the core function of the energy/distribution management system (EMS/DMS), has become necessary in order to operate the system efficiently and in a controlled manner.

Although SE is a well-known task in transmission systems, it is usually not a common task in unbalanced distribution systems due to the difference in design and operation philosophy. This thesis addresses these issues and investigates the distribution system state estimation with unbalanced full three-phase modelling. The formulation, based on weighted least squares estimation, is extended to include the open/closed switches as equality constraints.

This research then explores the vulnerabilities of the state estimation problem against attacks associated with leverage measurements. Detecting gross error particularly for leverage measurements have been found to be difficult due to low residuals. The thesis presents and discusses the suitability of externally studentized residuals compared to traditional residual techniques.

Additionally, the masking/swamping phenomenon associated with multiple leverages makes the identification of gross error even more difficult. This thesis proposes a robust method of identifying the high leverages and then detecting gross error when the leverage measurements are compromised. All algorithms are validated in different IEEE test systems.

Contents

List of Tables

List of Figures

List of Abbreviations

VVC	Volt var control
ICT	Information and communication technology
PMU	Phasor measurement units
RTU	Remote terminal units
EMS	Energy Management Systems
DMS	Distribution Management Systems
TNO	Transmission network operator
DNO	Distribution network operator
DSSE	Distribution system state estimation
BDD	Bad data detection
OLTC	On-load tap changer
Y	Star
Δ	Delta
WLS	Weighted least squares
SE	State estimation
MLE	Maximum likelihood estimation
DRGP	Diagnostic robust generalized potentials
GSR	Generalized studentized residuals

Acknowledgements

First of all, I would like to thank my supervisor Prof. Bikash C. Pal for his continuous support and guidance. I am grateful to him for providing me this opportunity and having patience and confidence on me throughout my PhD studies. I would also like to acknowledge his invaluable comments and suggestions from time to time.

I would like to acknowledge the financial support provided to me by the Research Council UK Energy Programme, UK and the Department of Science and Technology (DST), India under grant EP/H040331/1 to carry out this research activity.

I would take this opportunity to thank all my past and present colleagues like Sara N., Sara M., Yashodhan, Diptargha, Hossein, Georgios, Yousef, Jose, James, Vincenzo, Abhinav, Claudia and many others for their help and support and providing a wonderful and enjoyable working environment.

My friends from India like Tapabrata, Debashis, Soumya, Dipanjan and many others also deserve mention.

Finally, I owe my biggest gratitude to my parents and family members. Without their constant support and encouragement I would have not been able to achieve this.

Dedicated to

my parents

Chapter 1

Introduction

1.1 Motivation and Objectives

Power networks all over the world are undergoing a significant scale of development. They are gradually changing from passive systems to smart and active systems, where the performance and flexibility of operation are improved. The major drivers are shift of technology of generation towards renewables (mainly solar and wind) and new forms of demand such as electric transportation, district heating etc. Due to the change of operation philosophy, the loads have become smart and the small industrial, commercial and domestic customers can feed in energy to the system and thus can participate in demand response functions. Therefore, this uncertain nature of generation and new type of demand need to be dealt with by more active energy management strategy [1]. There is an increasing adoption of smart instrumentation such as phasor measurement units (PMUs), intelligent metering etc. in transmission networks and smart meters in distribution networks with information and communication technology (ICT) infrastructure. As a result, the integrity of data and information is exposed to risk and the power system is more prone to malicious attacks from adversaries. Tampered data will obviously affect the outcome of network control and computing functions such as state estimation, security analysis, volt var control (VVC) etc.

To enable the effective control of the power network, the states of the system need

to be observed properly. The energy/distribution management systems (EMS/DMS) will play a crucial role in the control and operation of smart power systems. Central to every EMS/DMS are two functional blocks: the state estimator and the control scheduling block. The state estimation provides a real-time estimate of system states, based on the measurements obtained from meters and sensors in the remote terminal units (RTUs).

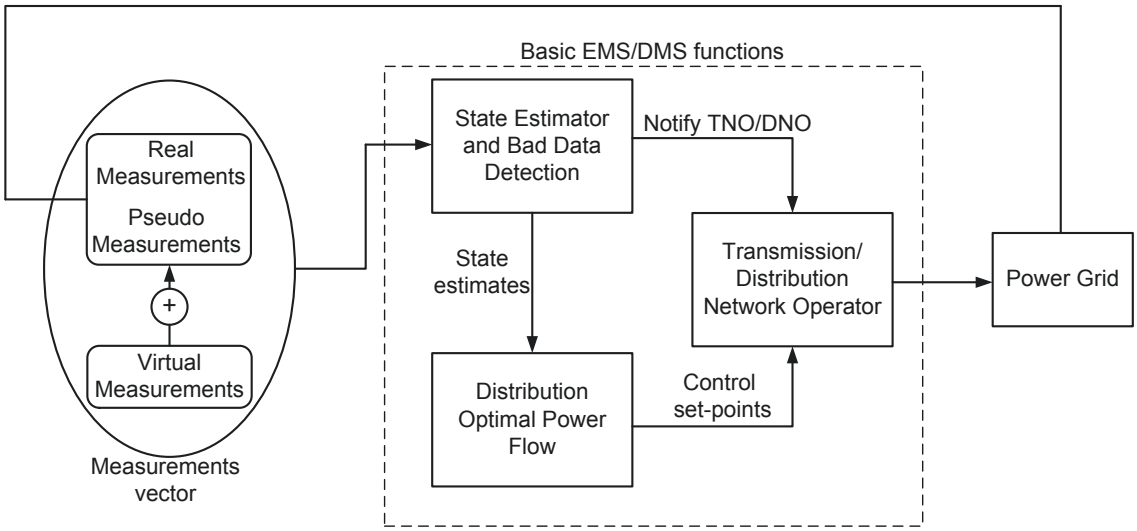

Figure 1.1: A typical energy/distribution management system (EMS/DMS) architecture

A typical energy/distribution management system architecture is shown in Figure 1.1. It shows that the control set-points for the transmission/distribution network operator are decided by the states estimated from the state estimation block. It will help the EMS/DMS to take a host of operational decisions in case of contingency and cascaded tripping etc. Thus the states of the system need to be monitored and observed effectively. Hence, state estimation has become an important and necessary function of modern network operation.

The research proposed in this thesis develops the distribution state estimation on unbalanced systems, addresses the vulnerabilities of the state estimation problem and explores and develops a new methodology for gross error detection against attacks from adversaries.

1.2 Literature review

With growing number of controllable devices distribution system state estimation is becoming popular for distribution system operation. The solution methodology mainly focusses on weighted weast squares (WLS) estimation technique. But the majority of distribution systems operate under varying degrees of unbalance. Hence, this has paved the way for the need of unbalanced three-phase state estimation rather than single-phase state estimation.

1.2.1 State estimation of distribution systems

There has been growing literature in distribution system state estimation.

Many authors [2–5] have proposed branch current based state estimators for distribution system. They take branch currents as state variables. The method is particularly useful for radial networks. The method of branch current decouples the state estimation problem into three sub problems, one for each phase. This makes it computationally efficient.

To obtain a unique estimate of system states and to take care of the measurement errors it is required to have the number of measurements larger than the number of states. This is called redundancy. Due to their large and radial nature distribution systems are not metered properly. Due to the lack of this measurement redundancy in the distribution network, the zero-injection nodes are considered as zero-injection measurements. Lin and Teng [6] proposed the method of Lagrange multipliers to incorporate the zero-injections and proposed a current-based decoupled SE based on rectangular co-ordinates methodology to solve the problem.

Baran [7] mentioned that the SE results can improve the forecasted load data by using real-time measurements. However, when there are limited real measurements the accuracy of SE depends on the accuracy of the forecasted load data. It is proposed that state estimation can be used to improve the data needed for real-time monitoring of distribution systems.

Lin et al. [3] devised an efficient method for treating the current magnitude

measurements into equivalent current phasors. The constant gain matrix and a decoupled form for current measurements developed here produced a robust and efficient solution. It also needs minimal storage requirement due to the constant gain matrix. Wang and Schulz [4] discussed the decoupling in case of branch current state estimation and about the impact of branch power flow, current magnitude and voltage measurements on the accuracy of SE results.

Li [8] has presented a distribution system state estimation based on the WLS approach and three phase modelling techniques. Several factors such as load error correlation, pseudo measurement errors, location of real-time meters, accurate load measurements improve the state estimates.

Wang and others in [5] studied the state estimation based on branch currents when the different types of DGs are added such as PQ-type, PV-type, PI-type and PQV-type and discussed the impact on accuracy of SE on location and size of DGs. The DGs were treated as real-time measurements and the state estimates were found to be closer to the true value compared to the state estimates without DGs.

Lu and others [9] have proposed a three-phase current based estimator that sought to minimise the WLS objective. The advantage of having a constant gain matrix was achieved through minimal storage requirement and less computational time. Here, power, current and voltage measurements are converted to their equivalent currents, and as a result the Jacobian terms are constant and equal to the admittance matrix elements.

Baran and Kelley [2] have introduced an algorithm that takes branch currents as state variables. This method is very efficient and works well in radial distribution systems. A better computation speed and filtering properties without losing accuracy were achieved through feeder reduction method.

Due to lack of measurements the load demands in the distribution system are taken from historical load forecast data or load curves. This creates sufficient uncertainty. A probabilistic approach to DSSE based on the probabilistic radial flow algorithm is proposed in [10] by Ghosh and others. This algorithm also takes into account the non-normality of states, radial nature of the system, low ratio of real-time measurements to states and the load diversity. The effects of load correlation

and confidence interval were also discussed.

Lubkeman et al [11] presented the distribution system SE on a field circuit and proved that the state estimation algorithm is viable on practical systems. The state estimates were found to give improved solutions over load flows.

With gradual deployment of phasor measurement units (PMUs) in networks the state estimation process is also changing. With this development the measurement system is gradually changing to a phasor-only system. Jones [12] have mentioned a three-phase linear state estimation with synchronised phasor measurements and applied on a practical system.

The distribution system operates under various degrees of unbalance. This coupled with topological uncertainties has an impact on the accuracy of state estimates. References [13–16] have demonstrated this. The full system modelling allows the detection of dangerous imbalances and better estimation of current operating point. It can also have better bad data rejection capability. Zero injection is taken as constraints in the nodes having no generation, load and measurement.

PMUs and smart meters are being incorporated into the modern distribution network. A full three-phase linear estimator based on PMU and smart meter measurements is proposed by Haughton and Heydt in [17]. The linear estimator produces better estimates and thus would help in decision making of distribution systems.

Thukaram and others [18] at first ensured the observability of the network by graph theory and then used a robust forward-backward propagation method based on nodal current injections to obtain the state estimates.

1.2.2 Bad data detection methodology

However, with the advent of deregulation in the power network and the introduction of distributed generations (DG) and smart meters [19,20], having an efficient and accurate state estimate is becoming more and more necessary.

With such integration of distributed generations and ICT infrastructure, the future smart power grid is no longer a physical system only but rather a cyber-physical system [21–23]. It has been argued in [21] that an adversary can inject

malicious data into the system without being detected by classical bad data detection techniques. In the case that the adversary performs an unobservable attack, which cannot be detected by classical bad data detection techniques, it is important to know how vulnerable the power system operation is to these attacks. The sparse attack vector can be designed by an adversary by controlling enough meters.

Smart meters with their two-way data and communication flow are easily prone to attacks from adversaries [24, 25]. Reference [24] further explains different types of attack that can be synthesized in relation to strong and weak attack regimes. The unobservable attacks exist in case of strong attack regime while in case of weak attack regime the gross errors can be detected by generalized likelihood ratio test and minimum mean squares estimator.

Reference [25] illustrates the strategies for malicious attacks to be incorporated such as unobservable attacks, minimum size unobservable attacks and minimum residue energy attacks and henceforth, uses the concept of generalized likelihood ratio test (GLRT) to detect the gross error with $L1$ norm regularization.

The issue of bad data detection has been addressed in the literature as in [26,27] by the χ^2 distribution and χ^2-test. Many people [26, 27] have used the largest normalized residuals (LNR) to detect bad data in single bad data environment or in case of multiple non-interacting bad data.

The authors in [28] and [29] have demonstrated further the impact of inaccurate parameters and untransposed lines not only on the accuracy of the estimated quantities but on the bad data rejection capability as well. This led to the requirement of synchronised phasor measurements at the distribution level.

Many authors [30] have presented the identification procedure of leverage measurements by various distance measures like Mahalanabis distance and by Projection Statistics [31]. [30] argues that with projection statistics the state estimation results are much closer to load flow results than without projection statistics. Mili et al. [31] used a Schweppe-Huber type generalized estimator based on projection statistics, which is a simple modification of WLS estimator.

In recent years, there have been growing interests in the false data injection to power system and dealing with those attacks and the vulnerabilities [24, 25, 32–34].

The basic idea of false data injection attack is to add a non-zero attack vector into the measurements [24]. It has been reported in the literature [24, 25, 32–34] how an adversary can synthesize an attack vector just to bypass the normalized residual test in the dc state estimator.

Chen and Abur [35] have proposed the placement of phasor measurement units (PMUs) for the detection of bad data. It has been proposed that minimising the number of strategically placed PMUs will improve system observability by eliminating criticality of measurements and thus improve bad data detection.

Narvaez and others [36] have presented the concept of robust distances to detect the difference between good and bad leverage measurements. They argued that a robust estimator has better performance than the WLS estimator even in the presence of bad data and bad leverage points.

Khwanram and others [37] have used Particle Swarm Optimization (PSO) to detect multiple bad data in state estimation. They used the PSO to minimize the number of bad data while maintaining the measurement redundancy. They, however, argued that the parameters of the technique need to be selected efficiently to reach a global optimum solution.

Qingyu Yang and others [32] have considered the bad data in measurements as injection of bad data and proposed the method of false data injection attacks on state estimation. It reports how to optimize the number of measurements to be tampered in order to compromise a given number of state variables.

Suzhi Bi and Yin Jun Zhang [33] have elucidated the defensive mechanism procedures against false data injection attacks in the state estimation. They have proposed a greedy algorithm which produces optimal solution for the sequential protection of state estimation against the malicious tampering. They also report that the optimal algorithm has less computational complexity compared to the sub-optimal algorithm.

Liu, Ning and Reiter [38] have summarised the concept of false data injection attacks against state estimation in electric power grid. They have discussed the different types of false data injection attacks in terms of two scenarios - access to meters and access to resources. They have also reported the impacts of these two

scenarios on state estimates when the meters or state variables are targeted.

Lin and Pan [39] have explained a static state estimation approach regarding bad data detection and identification. They employ a modified covariance matrix and reduce the effect of bad data on state estimates. In addition to that they proposed a gap statistic method to identify bad measurements.

Tarali in [40] have presented a bad data detection technique in two-stage state estimation using phasor measurements. The conventional state estimates from the first stage are used with PMU measurements in the second stage to get the new estimates. The PMU measurements will improve the system observability and thus help in bad data detection.

Hug and Giampapa [34] show how a false data injection attack at the RTU level can be hidden by tampering with a number of measurement data and they assess the vulnerability associated with this threat in terms of number of measurements to be attacked.

However, none of [24, 25, 32–34] have addressed the situation when a particular influential or leverage measurement is compromised. Hence, this provides the basis to have an identification procedure against such an unobservable attack.

1.3 Outline of the thesis

Following the introduction and literature review, the thesis is organised as follows:

Chapter 2 points out the difference between transmission systems and distribution systems. It provides a detailed modelling of various components of three-phase systems such as lines, transformers, switches and loads. It further describes the three-phase test systems used in this research.

Chapter 3 provides a generic formulation for three-phase state estimation. The technique based on weighted least squares estimation has been discussed in detail. The results and simulations on different test systems have been illustrated for various switch configurations also. The results for a part of this chapter have been published

in [41].

Chapter 4 discusses the different diagnostic techniques for the detection of gross error. It also compares the different residual techniques and the robust estimation techniques. It introduces the concepts of internally and externally studentized residuals and discusses their effectiveness in detecting bad data.

Chapter 5 proposes a technique for bad data detection when the leverage measurement points are attacked by an adversary. The mathematical formulation of the technique has been discussed in detail. The proposed technique has also been compared with the traditional detection techniques and the advantages of the method have been justified. The results of this chapter have been submitted for publication [42] and is currently under review. This was first submitted on December 2014.

Chapter 6 recapitulates the contributions of this thesis and presents a brief overview of future research directions.

1.4 Contributions of the thesis

The contributions of this research can be summarised as follows:

- The distribution system state estimation literature is mainly focused on balanced system architecture. This thesis discusses the necessity of unbalanced system component modelling, develops the detailed modelling of three-phase unbalanced systems and discusses the advantages.

- A robust three-phase unbalanced state estimation model is presented. The thesis investigates the suitability of three-phase SE on different load types, transformer connections and switch configurations. It also investigates the existing transmission system state estimation techniques and assesses their suitability to unbalance distribution system. The extended modelling of 13-bus and 123-bus has been carried out and the DSSE is implemented on these standard distribution systems.

- The thesis studies the possible vulnerabilities of SE in detail. It also introduces and discusses the suitability of externally studentized residuals, a technique

used in statistics for outlier detection, in power system bad detection for the first time.

- Due to the integration of intelligent and smart metering and communication infrastructures, the modern power system is becoming more and more vulnerable. The vulnerabilities in regards to leverage measurements have not been properly addressed before. This research proposes and develops a new technique to identify the gross error in case of leverage measurements and discusses the advantages of the technique. It has been tested on 14-bus and 123-bus systems.

1.4.1 Contributions to other thesis/research

This thesis's extensive three-phase modelling framework has been used in other PhD thesis/research as detailed below.

- The exhaustive three-phase modelling of distribution systems has led to parallel contributions in a thesis in the group which explores and investigates the estimation of discrete transformer taps in distribution systems by hybrid particle swarm optimization and then by ordinal optimization. The ordinal optimization has the capability of providing the accurate tap estimation in less computation time and at the same time has the ability of handling huge computational complexity. This has led to publications [43, 44] and one paper submitted for publication [45]. The development of ordinal optimization for three-phase state estimation has enabled us to form a strong background for the detection of discrete transformer tap error which is under process.

- The three-phase modelling framework has also benefited a research on probabilistic operation of three-phase distribution network in collaboration with a lecturer at University of Bradford. It explores the uncertain operation of unbalanced distribution systems with increased solar input under active network management schemes of coordinated voltage control and power factor control. This has led to one publication [46].

1.5 List of publications

Journal

- **Majumdar, A.**; Pal, B.C., "Bad Data Detection in the Context of Leverage Point Attacks in Modern Power Networks," *IEEE Transactions on Smart Grid*, Revision under review, Submitted on December 2014
- Nanchian, S.; **Majumdar, A.**; Pal, B.C., "Three-Phase State Estimation Using Hybrid Particle Swarm Optimization," *IEEE Transactions on Smart Grid*, vol.PP, no.99, pp.1-1 doi: 10.1109/TSG.2015.2428172, 2015
- Nanchian, S.; **Majumdar, A.**; Pal, B.C., "Ordinal Optimization Technique for Three Phase Distribution Network State Estimation with Discrete and Continuous Variables," *IEEE Transactions on Smart Grid*, Revision under review.
- Mokryani, G.; **Majumdar, A.**; Pal, B.C., "A Probabilistic Method for the Operation of Three-Phase Unbalanced Active Distribution Networks," *IET Renewable Power Generation*, 2016

Conference

- **Majumdar, A.**; Pal, B.C., "A Three-Phase State Estimation in Unbalanced Distribution Networks with Switch Modelling," *IEEE First International Conference on Control, Measurement and Instrumentation*, Kolkata, India, 2016
- Nanchian, S.; **Majumdar, A.**; Pal, B.C.; Mobsby, D.; MacLeman, D.F., "Transformer tap estimation using hybrid particle swarm optimization," *IEEE PES General Meeting*, Washington D.C., 2014

Chapter 2

Modelling of Three Phase Unbalanced System

The electric power distribution system is the final stage of electric power system. It is connected to the transmission system at the distribution substation. The substation transformers lower the transmission voltage to medium or low voltage. The distribution system carries the power to the distribution transformers. The distribution transformers lower the voltage level further to carry the electric power to customer premises. The distribution system consists of unsymmetrical network components and unbalanced load. The distribution system can be unbalanced due to many reasons: The loads connected may single phase loads such as the lighting loads or single phase induction motors. Although the distribution systems are designed as balanced statistically the distribution system may be unbalanced due to the presence of different laterals drawing different currents. Hence, the single line representation for an unbalanced distribution system is not appropriate. Therefore, the full three-phase modelling of the network components is necessary.

Over the years, there has been significant research on three-phase load flow in distribution systems [47]. However, most of the operational i.e., control and contingency, decisions based on state estimation have been applied to distribution systems assumed to be balanced [48]. Moreover, the loads considered are constant power and Y-connected loads and the different status (closed/open) of the switches

are not considered [9,18]. The following sections describe the three phase modelling of various components of the network such as line, transformers, switches and loads.

2.1 Difference between transmission system and distribution system state estimation

The design and topology of distribution systems are quite different from transmission systems. The distribution system is typically characterised by unbalanced systems, shorter lines with high R/X ratios and hence suffers more losses compared to a transmission network. The distribution system is usually spread over a large geographical area and is radial in nature. As a result, many consumer or domestic loads remain unmeasured and a large portion of the network remains unmonitored. This poses a serious challenge to the observability of the network and the state estimator module to provide reasonable estimates. Therefore, the planning and operation of distribution system(s) is different from transmission systems.

The methodologies adopted in the transmission systems cannot be duplicated in the distribution system. Since a large portion of the distribution system is unmeasured the system is underdetermined. To overcome this difficulty the unmeasured/unmetered loads are treated as pseudo measurements and their measurements are derived from typical load curves, historical data of the feeders and transformer loadings. These pseudo measurements will have more uncertainties associated with them compared to that of real measured data. Although recently automated metering infrastructure (AMI) have come into the picture it is still not practical to install AMRs in every location. There are actually more pseudo measurements than real measurements in the distribution system state estimation (DSSE). Since the R/X ratio is high the decoupling of state estimation problem into $P-\delta$ and $Q-V$ equations is not possible. The distribution system is more prone to unbalances due to 1-ph and 3-ph loads and due to the presence of 1-ph, 2-ph and 3-ph laterals. Moreover,being large,the distribution system is prone to unbalance faults and these can also create unbalance in the system.

This sets up the motivation to explore and develop the robust three phase state estimation on unbalanced systems.

2.2 Three-phase line modelling

The distribution system consists of untransposed overhead lines and underground cables which can be three-phase or single and/or two-phase laterals. Figure 2.2 shows a three-phase distribution system with single and two phase laterals. This combined with the unbalanced loads (single, two or three-phase loads) contribute to the unbalanced nature of the system. Due to the untransposed nature of the lines, the single phase/line representation of lines does not work. Thus, it is essential to compute the impedance of the lines accurately. A modified Carson's equation is applied to compute the self and mutual impedance of the lines [49].

$$Z_{ii} = r_i + 0.095 + j0.121 \times \left(ln\frac{1}{GMR_i} + 7.934 \right) \Omega/\text{mile} \tag{2.1}$$

$$Z_{ij} = 0.095 + j0.121 \times \left(ln\frac{1}{D_{ij}} + 7.934 \right) \Omega/\text{mile} \tag{2.2}$$

Where,

Z_{ii} Self-impedance of conductor i in Ω/mile.

Z_{ij} Mutual impedance between conductors i and j in Ω/mile.

r_i Resistance of conductor i in Ω/mile.

GMR_i Geometric mean radius of conductor i in feet.

D_{ij} Distance between conductors i and j in feet.

The modified Carson's equation also takes into account the ground return path (neutral conductor) for the unbalanced currents.

The modified Carson's equations (2.1) and (2.2) for a three phase overhead or underground circuit which consists of $neut$ neutral conductors forces the resulting impedance matrix $(3 + neut) \times (3 + neut)$. However, for most applications, it is necessary to have the 3×3 phase impedance matrix. Therefore, $(3+neut)\times(3+neut)$ impedance matrix is broken down to 3×3 matrices by Kron's reduction as given

in (2.3). In this approach, all the lines will be modelled by 3×3 phase impedance matrices and for two phase and single phase lines the missing phases are modelled by setting the impedance element to zero.

$$Z_{ij,abc} = [Z_{ij}] - Z_{ineut}[Z_{neutneut}]^{-1} Z_{neutj} \tag{2.3}$$

Where,

$Z_{ij,abc}$ Phase-impedance matrix.

Z_{ineut} Mutual impedance matrix between conductor i and *neut* neutral conductors in Ω/mile.

$Z_{neutneut}$ Self-impedance matrix of *neut* neutral conductors in Ω/mile.

Z_{neutj} Mutual impedance matrix between *neut* neutral conductors and conductor j in Ω/mile.

Therefore, for each line between two nodes, there will be a 3×3 matrix instead of a single element for a single phase balanced system. Hence, the resultant Y-bus matrix of the system will be of $(n \times 3) \times (n \times 3)$. The structure of the Y-bus matrix is shown in (2.4).

$$Y = \begin{bmatrix} Y_{11}^{aa} & Y_{11}^{ab} & Y_{11}^{ac} & \cdots & Y_{1n}^{aa} & Y_{1n}^{ab} & Y_{1n}^{ac} \\ Y_{11}^{ba} & Y_{11}^{bb} & Y_{11}^{bc} & \cdots & Y_{1n}^{ba} & Y_{1n}^{bb} & Y_{1n}^{bc} \\ Y_{11}^{ca} & Y_{11}^{cb} & Y_{11}^{cc} & \cdots & Y_{1n}^{ca} & Y_{1n}^{cb} & Y_{1n}^{cc} \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots & \vdots \\ Y_{n1}^{aa} & Y_{n1}^{ab} & Y_{n1}^{ac} & \cdots & Y_{nn}^{aa} & Y_{nn}^{ab} & Y_{nn}^{ac} \\ Y_{n1}^{ba} & Y_{n1}^{bb} & Y_{n1}^{bc} & \cdots & Y_{nn}^{ba} & Y_{nn}^{bb} & Y_{nn}^{bc} \\ Y_{n1}^{ca} & Y_{n1}^{cb} & Y_{n1}^{cc} & \cdots & Y_{nn}^{ca} & Y_{nn}^{cb} & Y_{nn}^{cc} \end{bmatrix} \tag{2.4}$$

2.3 Transformer modelling

The distribution system generally consists of feeder and distribution transformers which provide the final voltage transformation to the loads. The three phase transformers are modeled by an admittance matrix which depends on the connection

type. A transformer can be Y-Y, Y-Δ, Δ-Δ. In balanced systems, the transformers are modelled as single phase leakage impedances. Thus the single phase/line representation will work perfectly fine. However, in the analysis of the distribution feeder for three phase, it is required to model the various three phase transformer connections correctly. The comprehensive calculations of three phase transformers and their various connections can be found in references [49, 50]. While forming the Y-bus, a transformer can be considered as one element between two nodes of the system. Therefore, the transformer contributes to a 6×6 block in the Y-matrix. The transformer nodal admittance matrix can be calculated from the current-voltage relationship of transformer, which is given by

$$\begin{bmatrix} I_p^{abc} \\ I_s^{abc} \end{bmatrix} = \begin{bmatrix} Y_{pp}^{abc} & Y_{ps}^{abc} \\ Y_{sp}^{abc} & Y_{ss}^{abc} \end{bmatrix} \begin{bmatrix} V_p^{abc} \\ V_s^{abc} \end{bmatrix} \tag{2.5}$$

where,

I_p^{abc}, V_p^{abc} are the primary side current and line-to-neutral voltage vectors for the three phases.

I_s^{abc}, V_s^{abc} are the secondary side current and line-to-neutral voltage vectors for the three phases.

The nodal admittance matrix is formed of the sub-matrices Y_{pp}^{abc}, Y_{ps}^{abc}, Y_{sp}^{abc} and Y_{ss}^{abc}. Depending on the connection of three phase transformers on the primary and secondary sides the sub-matrices Y_{pp}^{abc}, Y_{ps}^{abc}, Y_{sp}^{abc} and Y_{ss}^{abc} will vary. Thus the nodal admittance matrix for the transformers will change as explained in [50]. The nodal admittance matrix also depends on whether the transformer connection is step-up or step-down. The nodal admittance matrix components for some of the connections

Table 2.1: Nodal admittance matrix for step-down transformers

Primary	**Secondary**	$\mathbf{Y_{pp}^{abc}}$	$\mathbf{Y_{ss}^{abc}}$	$\mathbf{Y_{ps}^{abc}}$	$\mathbf{Y_{sp}^{abc}}$
Y_g	Y_g	Y_I	Y_I	-Y_I	-Y_I
Y_g	Δ	Y_I	Y_{II}	Y_{III}	Y_{III}^T
Y	Δ	Y_{II}	Y_{II}	Y_{III}	Y_{III}^T
Δ	Δ	Y_{II}	Y_{II}	-Y_{II}	-Y_{II}
Y_g	Y	Y_{II}	Y_{II}	-Y_{II}	-Y_{II}
Y	Y_g	Y_{II}	Y_{II}	-Y_{II}	-Y_{II}
Y	Y	Y_{II}	Y_{II}	-Y_{II}	-Y_{II}
Δ	Y_g	Y_{II}	Y_I	Y_{III}	Y_{III}^T
Δ	Y	Y_{II}	Y_{II}	Y_{III}	Y_{III}^T

of step-down transformer are shown in Table 2.1. where,

$$Y_I = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} y_t \tag{2.6}$$

$$Y_{II} = \frac{1}{3} \begin{bmatrix} 2 & -1 & -1 \\ -1 & 2 & -1 \\ -1 & -1 & 2 \end{bmatrix} y_t \tag{2.7}$$

$$Y_{III} = \frac{1}{\sqrt{3}} \begin{bmatrix} -1 & 1 & 0 \\ 0 & -1 & 1 \\ 1 & 0 & -1 \end{bmatrix} y_t \tag{2.8}$$

where y_t is the transformer leakage impedance in per unit.

2.4 Switch modelling

Switches are considered as branches with zero impedance. It is assumed that the status of the switches, i.e. closed or open, are known beforehand. The operational constraints for the switches are considered as equality constraints as given by $c_{eq} = 0$ in equation (3.18) of the original problem formulation. This is described in detail in Section 3.3.

- When the switch between bus i and bus j is assumed closed for branch i-j, the voltages and angles for bus i and bus j and phase ph for all the three phases are equal.

$$\begin{aligned} V_i^{ph} - V_j^{ph} &= 0 \\ \delta_i^{ph} - \delta_j^{ph} &= 0 \end{aligned} \tag{2.9}$$

- When the switch is assumed open between bus i and bus j, the active and reactive power flow to the switch will be zero.

$$\begin{aligned} P_{ij}^{ph} &= 0 \\ Q_{ij}^{ph} &= 0 \end{aligned} \tag{2.10}$$

2.5 Load modelling

The loads in distribution systems are generally unbalanced. The loads are three-phase, two-phase or single-phase. They can be connected in grounded Y or ungrounded Δ configuration. From the point of view of electricity usage, loads can be broadly classified as constant power, constant impedance or constant current loads. They are commonly represented as power consumed per phase and considered to be line-to-neutral for Y-loads and line-to-line for Δ-loads. The typical ZIP models for Y and Δ loads are shown in (2.11) and (2.13).

$$P_L^{ph} = P_n^{ph}\left[c_1^P + c_2^P\left(\frac{V^{ph}}{V_n}\right) + c_3^P\left(\frac{V^{ph}}{V_n}\right)^2\right] \tag{2.11}$$

$$Q_L^{ph} = Q_n^{ph}\left[c_1^Q + c_2^Q\left(\frac{V^{ph}}{V_n}\right) + c_3^Q\left(\frac{V^{ph}}{V_n}\right)^2\right] \tag{2.12}$$

$$P_L^{ph12} = P_n^{ph12}\left[c_1^P + c_2^P\left(\frac{V^{ph12}}{\sqrt{3}V_n}\right) + c_3^P\left(\frac{V^{ph12}}{\sqrt{3}V_n}\right)^2\right] \tag{2.13}$$

$$Q_L^{ph12} = Q_n^{ph12}\left[c_1^Q + c_2^Q\left(\frac{V^{ph12}}{\sqrt{3}V_n}\right) + c_3^Q\left(\frac{V^{ph12}}{\sqrt{3}V_n}\right)^2\right] \tag{2.14}$$

Where, $ph12 = ab, bc, ca$.

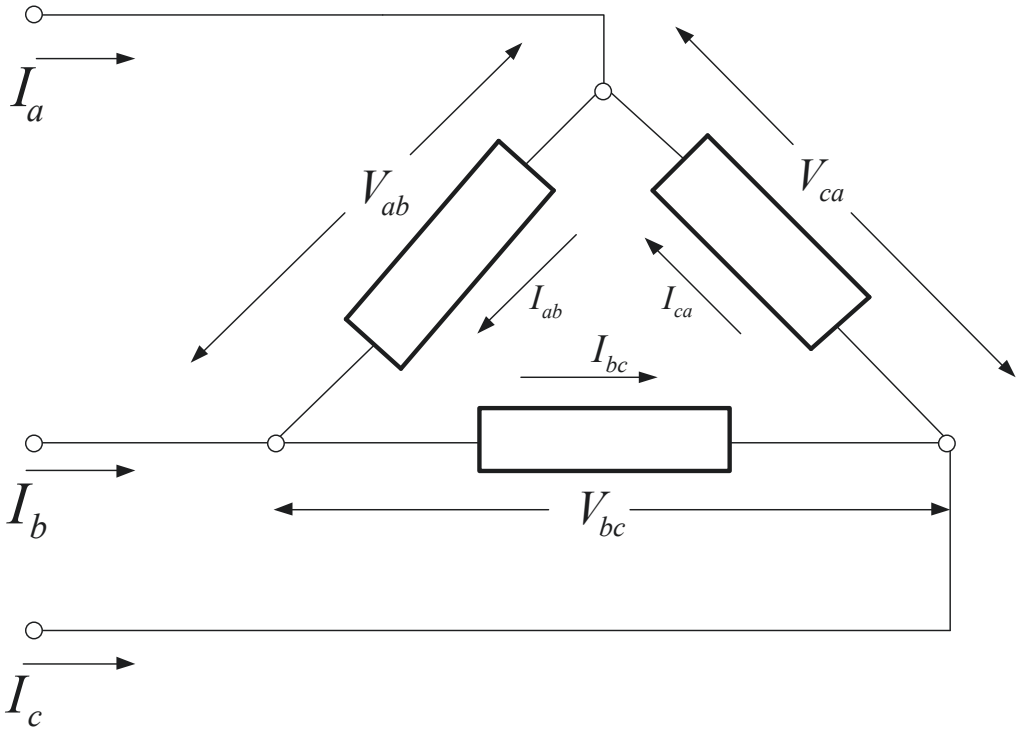

Figure 2.1: Delta-connected three-phase load

Figure 2.1 shows a typical Δ-connected three phase load. The voltage magnitudes are line-to-neutral for the state estimation formulation, which is discussed in the next chapter. Therefore, in case of delta loads, the equivalent wye powers are calculated at each iteration in order to calculate the active and reactive power at each node. This is illustrated in the following steps.

- Calculate line-to-neutral voltage for Δ loads

$$\begin{bmatrix} V_i^{ab} \\ V_i^{bc} \\ V_i^{ca} \end{bmatrix} = \begin{bmatrix} V_i^a \angle \delta_i^a - V_i^b \angle \delta_i^b \\ V_i^b \angle \delta_i^b - V_i^c \angle \delta_i^c \\ V_i^c \angle \delta_i^c - V_i^a \angle \delta_i^a \end{bmatrix} \tag{2.15}$$

- Read the active and reactive power of Δ loads

- Calculate the line currents of Δ loads

$$I_{ab} = \left(\frac{P_{ab} + jQ_{ab}}{V_{ab} \angle \delta_{ab}} \right)^* \tag{2.16}$$

- Calculate the current at each phase

$$\begin{bmatrix} I_a \\ I_b \\ I_c \end{bmatrix} = \begin{bmatrix} 1 & 0 & -1 \\ -1 & 1 & 0 \\ 0 & -1 & 1 \end{bmatrix} \begin{bmatrix} I_{ab} \\ I_{bc} \\ I_{ca} \end{bmatrix} \tag{2.17}$$

- Calculate the equivalent line-to-neutral active and reactive powers

$$\begin{aligned} V_a I_a^* &= P_a + jQ_a \\ V_b I_b^* &= P_b + jQ_b \\ V_c I_c^* &= P_c + jQ_c \end{aligned} \tag{2.18}$$

For the phases where the loads are non-existent, the active and reactive power values are set to zero for those particular phases.

2.6 Measurements

The distribution system normally covers a large geographical area. Hence, it is not possible to place meters at all nodes and lines. Hence, the redundancy of distribution systems is usually far less than that of transmission systems. However, it is required to make the system observable in order to solve the state estimation. With a given set of measurements the system is said to be observable if a unique estimate of the states can be found. For the system to be observable the number of measurements should be more than the number of state variables. The graph theory method has been carried out to ensure the observability of the network. Therefore, the load data taken from historical load data profiles are taken as pseudo measurements and zero-injection buses are considered as virtual measurements.

2.6.1 Load flow calculation

In general, current-injection or interior point based three-phase load flow is performed to generate the input to state estimation. These are taken as true values for measurements. Gaussian distributed random noise components are added to these true values to generate the measurements. The real measurements are assumed

to have 1%-3% error while the pseudo measurements are assumed to be of 20%-50% error. The variances of the measurement error are then calculated based on equation (3.12) as explained in the next chapter. However, the zero-injections or the virtual measurements and the switches are taken as equality constraints. This calculation is described in detail in the next chapter.

2.7 Test systems

The standard IEEE test systems of 13-bus and 123-bus have been studied. The SE algorithms in this thesis have been tested on these systems. The system data are taken from [51] and [52]. The characteristics of the systems are detailed below.

2.7.1 IEEE 13-bus system

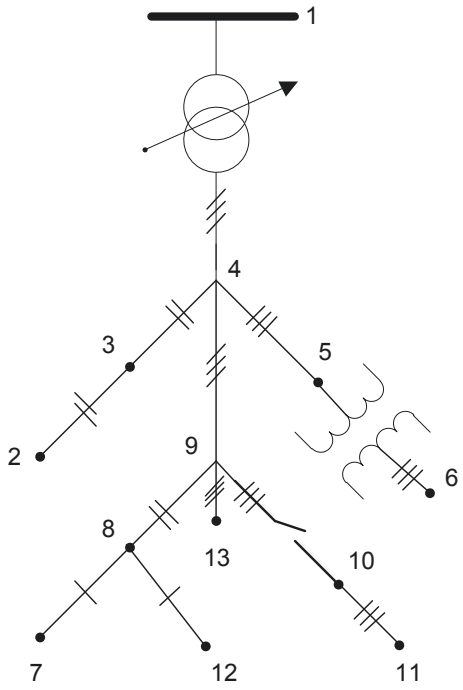

Figure 2.2: IEEE-13 bus unbalanced distribution system

The feeders are small yet they show some interesting characteristics. Figure 2.2 shows the system.

- Short and relatively high loaded for a 4.16 kV feeder.

- One substation voltage regulator consisting of three single phase units connected in Y.
- Both overhead and underground lines are present with a variety of phasing.
- It has shunt capacitors.
- It has one transformer: grounded Y-grounded Y
- Unbalanced spot and distributed loads are present.
- The loads are of constant power, constant current and constant impedance type and are Y and/or Δ connected.

The complete system data is given in Appendix A.

2.7.2 IEEE 123-bus system

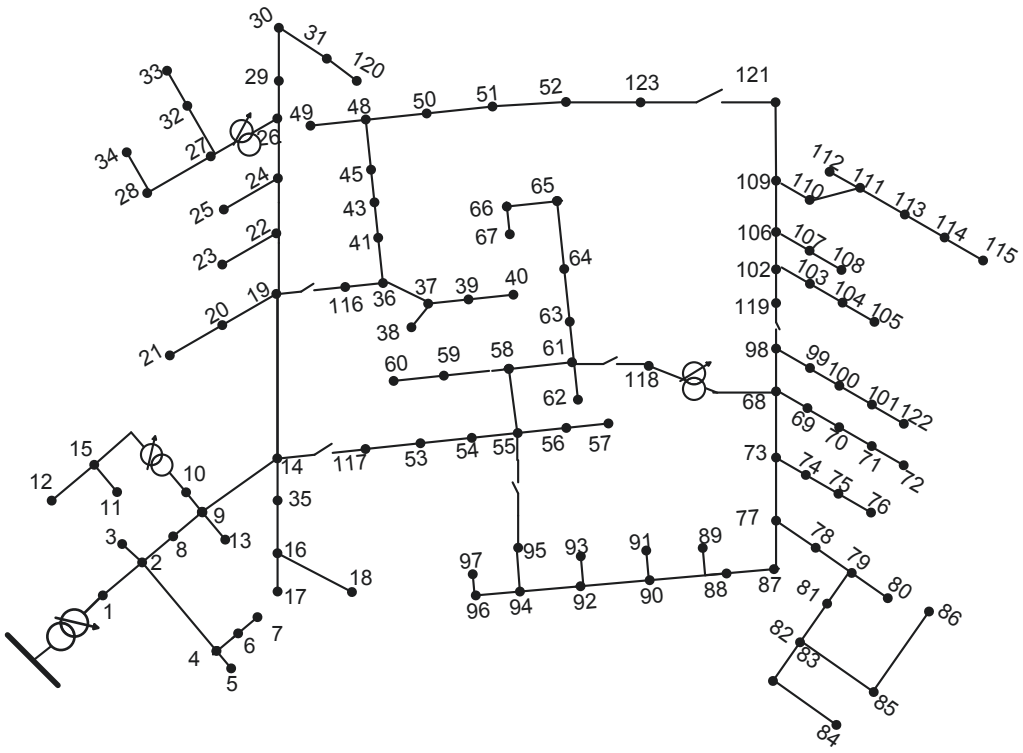

Figure 2.3: IEEE 123 bus system

The system test feeder has a nominal voltage of 4.16 kV and the system has some remarkable features. The system is shown in Figure 2.3.

- The system consists of overhead and underground lines.
- The system is unbalanced due to the presence of single, double or three phasing and unbalanced loads with constant power, constant current and current impedance types.
- There are spot loads only.
- Switching options to allow alternate ways of power flow.
- There are four tap tranformers present.

The complete system data has been provided in Appendix B.

2.8 Conclusion

This chapter presents in detail the three-phase modelling of various components such as distribution lines, transformers, switches and three-phase loads. The standard IEEE distribution systems like IEEE-13 bus and IEEE-123 systems have been discussed in detail. The systems contain all types of ZIP loads. The capacitors are considered as constant impedance loads, which are voltage dependent. The next chapter describes the state estimation formulation based on this unbalanced three-phase modelling.

Chapter 3

State Estimation of Unbalanced Distribution Systems

With the influx of phasor measurement units (PMUs), intelligent metering etc. in transmission systems and smart meters with information and communication technology (ICT) infrastructure in distribution systems, power systems now-a-days need to be monitored and controlled efficiently. To enable this, the states of the system need to be observed properly. This would help to influence the operational decisions and thus, to avoid contingency and cascaded tripping. It is done through an energy/ distribution management system (EMS/DMS) function- the state estimation (SE) [20, 53]. Figure 1.1 in Chapter 1 elucidates the importance of state estimation function in the transmission or distribution network operation. This function estimates the bus voltages and angles based on the available measurements, network data and topology information obtained from the supervisory control and data acquisition (SCADA) system.

In transmission systems, the state estimation concept is well established but in distribution systems due to the absence of sufficient measurements and unbalanced and asymmetric nature of the system, it was not mandatory to have a state estimation function as it involves significant complexity and computational time. But with growing number of controllable devices and the incorporation of smart meters in the system, state estimation is becoming important in distribution network operation.

Unlike the transmission system, the majority of distribution systems operate under varying degrees of unbalance. Moreover, the distribution system is radial in nature and has a higher R/X ratio. Therefore, the fast decoupled method causes numerical instability when applied to distribution systems [26]. Hence, this has paved the way for the need of unbalanced three-phase state estimation rather than single-phase state estimation.

To achieve accurate estimates of the state variables, this chapter presents a weighted least squares based estimator with the detailed modelling of the system components and different types of loads and also considering the different operational status of the switches in Section 3.3.

3.1 Overview of State Estimation

State estimation (SE) is a process of determining the states (voltage magnitudes and angles) of the network based on the available measurements and network topology information and parameter data. The measurements are prone to errors. The SE processes a set of redundant measurements and finds out the most optimal state of the system and thus takes care of the errors.

3.2 Maximum Likelihood Estimation

The state estimation methodology determines the most likely states of the system based on the measurements available in the system. In statistics, maximum likelihood estimation (MLE) is a method of estimating the parameters (or states) of a statistical model. For example, for a normal distribution, the parameters (mean and variance) are estimated with MLE from the knowledge of some sample data. The MLE selects the values of the parameters by maximising the given likelihood function. Here, in state estimation context, the measurement errors are generally assumed to follow a normal distribution. The joint probability density function (pdf) of all the measurements are formed and hence, an optimisation problem is solved to maximise the likelihood function.

Let x denote the state vector, comprising voltage magnitudes and angles. The non-linear measurement model relating the state vector and measurement vector $z_{meas} = [z_1\ z_2\ ...\ z_m]^T$ can be written as

$$z_{meas} = h_{func}(x) + \mathbf{e} \tag{3.1}$$

where $\mathbf{e} \backsim N(0, R)$ is a zero mean Gaussian noise with measurement error covariance matrix R. $h_{func}(x)$ is a vector of measurement functions. Equation (3.1) relates the state variables x to the measurement vector z_{meas}. Considering all the measurements to be independently and identically distributed, the joint pdf is a product of the individual pdfs of all the measurements.

$$f(z_{meas}|x) = f(z_1)f(z_2)\dots f(z_m) \tag{3.2}$$

The function given in equation (3.2) is the maximum likelihood function. The MLE maximises this likelihood function to get the maximum-likelihood estimate $\hat{x}$.

To simplify the optimization process the logarithm of the likelihood function is taken rather than only the likelihood function. Since *log* is a monotonically increasing function, maximizing the log-likelihood function is equivalent to maximizing the likelihood function. Therefore, from a power system perspective, the MLE can be defined as minimizing the negative of log-likelihood function

$$minimize : -log(f(z_{meas}|x)) \tag{3.3}$$

3.2.1 Generic Weighted Least Squares (WLS) Estimation

If we consider z_{meas} to be normally distributed,

$$f(z_{meas}|x) = \frac{1}{\sqrt{(2\pi)^m detR}} e^{-\frac{1}{2}(z_{meas}-h_{func}(x))^T R^{-1}(z_{meas}-h_{func}(x))} \tag{3.4}$$

The equation ((3.3)) can be written as

$$minimize : -\sum_{i=1}^{m} log(f(z_i)) \tag{3.5}$$

$$minimize : \frac{1}{2}\sum_{i=1}^{m}(\frac{z_i - \mu_i}{\sigma_i})^2 + \frac{m}{2}\ln(2\pi) + \sum_{i=1}^{m}\ln\sigma_i \tag{3.6}$$

which is equivalent to minimizing the first term of equation (3.6), where μ_i is the expected value E(z_i) and σ_i^2 is the variance of measurement error in z_i. The first term in equation 3.6 can be interpreted as squares of measurement errors $(z_i - \mu_i)$ weighted by σ_i^{-2}. Hence, this state estimation is also called as Weighted Least Square (WLS) state estimation. The minimization problem can be rewritten as:

$$\underset{x}{minimize} : J = (z_{meas} - h_{func}(x))^T R^{-1}(z_{meas} - h_{func}(x)) \tag{3.7}$$

The solution to this above problem can be solved by Newton's method. The details of the solution is discussed below. Here, $R = Cov(\mathbf{e}) = E(\mathbf{ee^T}) = diag(\sigma_1^2, \sigma_2^2, \ldots, \sigma_m^2)$ is defined as the error covariance matrix, where $e_i = z_i - h_i(x)$ is the error in the measurement and σ_i^2 is the variance of the i^{th} measurement.

To minimize the cost function in equation (3.7), the first-order derivatives should equate to zero. Hence, the first-order optimality condition can be written as

$$g_{derv}(x) = \frac{\partial J}{\partial x} = -H^T(x)R^{-1}(z_{meas} - h_{func}(x)) = 0 \tag{3.8}$$

Since, $g_{derv}(x)$ is a non-linear function, equation (3.8) can be solved by numerical methods only. Thus, the Taylor's series expansion of g_{derv} around the state variable vector x^k gives

$$g_{derv}(x) = g_{derv}(x^k) + \frac{\partial g_{derv}(x^k)}{\partial x}(x - x^k)^T + \frac{1}{2!}\frac{\partial^2 g_{derv}(x^k)}{\partial x^2}((x - x^k)^2)^T + \ldots = 0 \tag{3.9}$$

Neglecting the terms for 2nd and higher order derivatives, the above equation (an over-determined system) is solved by Gauss-Newton's method.

$$x^{k+1} = x^k - [G(x^k)]^{-1} g_{derv}(x^k) \tag{3.10}$$

So at each iteration, this results in solving the following equation

$$[G(x^k)]\Delta x^{k+1} = H^T(x^k)R^{-1}[z_{meas} - h_{func}(x^k)] \tag{3.11}$$

where, in the k^{th} iteration,

Δx^{k+1} $= x^{k+1} - x^k$

$H(x^k)$ $= [\frac{\partial h_{func}}{\partial x}]_{x^k}$ is the Jacobian matrix

$G(x^k) \quad = \frac{\partial g_{derv}(x^k)}{\partial x} = H^T(x^k)R^{-1}H(x^k)$ is the Gain matrix.

The measurement function can be created by having the network data and the telemetered measurements. The network data includes the information about the network topology, network parameters, transformer parameters and the loads. The required measurements for the state estimation can be classified under three broad categories:

- Actual Measurements: These are telemetered measurements which include voltage magnitudes, power injections, line currents and real and reactive power flows. The accuracy of these measurements depend on the accuracy of their meters.
- Pseudo Measurements: In the distribution system due to the large size of the system, it is very difficult to have load measurements at each and every bus. Hence, the loads are estimated based on previous load profile data with high variance of error.
- Virtual Measurements: Zero injections are considered as measurements with zero power injections. The zero injections are included in the measurement vector to improve the measurement redundancy and thus the observability of the network. One knows with more certainty the measurement values for the zero injection measurements. Hence, these measurements will have higher weights or very low variance.

The weights associated with real measurements, pseudo measurements and virtual measurements are different. In fact, the weight associated with each measurement is different. As the errors have been assumed to obey normal probability distribution the standard deviation of the errors can be computed as

$$\sigma_i = \frac{\mu_i \times \%error}{3 \times 100} \tag{3.12}$$

where, μ_i is taken as the true value and it is assumed that a $\pm 3\sigma$ deviation around the mean according to the property of normal distribution as shown in Figure 3.1.

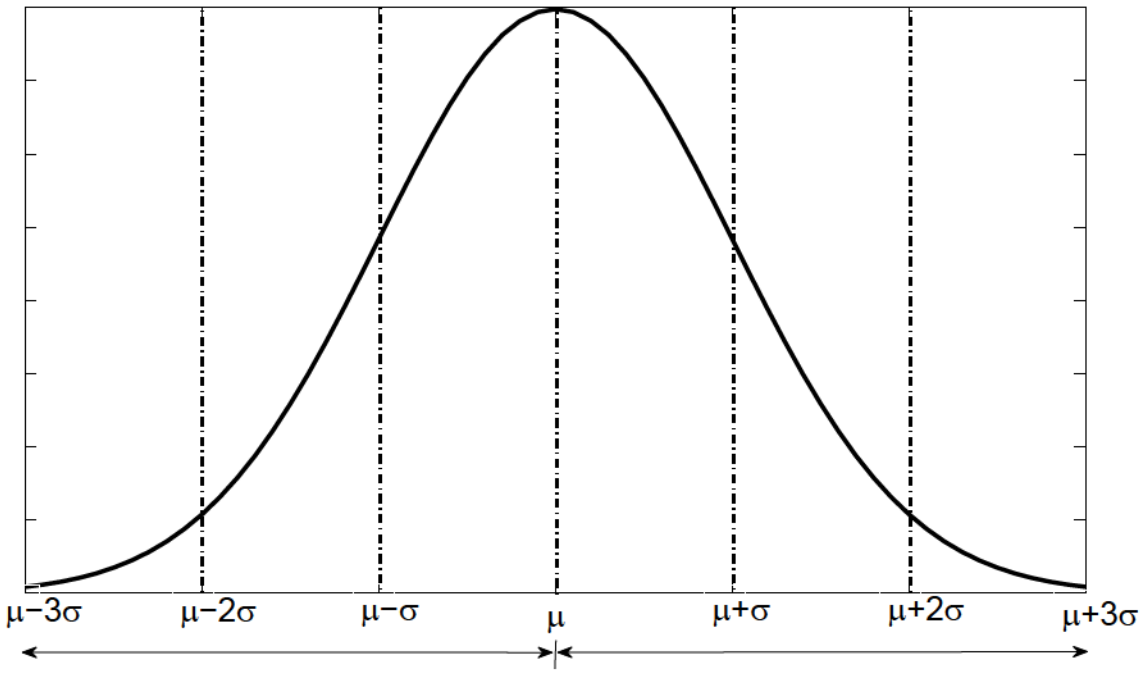

Figure 3.1: A Normal Distribution Curve

Equality Constrained Augmented Matrix Approach

The variance associated with virtual measurements such as zero injections are very low, whereas variance associated with pseudo measurements are very high. This may lead to ill-conditioning of the Gain matrix. So, to avoid this, the state estimation problem can be formulated as a constrained optimization problem.

$$\underset{\mathbf{x}}{minimize} : J = \frac{1}{2} r^T R^{-1} r \tag{3.13}$$

subject to:

$$c_{eq}(x) = 0$$

$$r - z_{meas} + h_{func}(x) = 0$$

r is the vector of residuals taken as explicit variables. The virtual measurements and the open/closed operational constraints of the switches are considered as equality constraints. These have been explained in the next section. The resulting Lagrangian will have two sets of Lagrange multipliers:

$$L = J - \lambda^T c_{eq}(x) - \mu^T (r - z_{meas} + h_{func}(x)) \tag{3.14}$$

Linearising the normal equations, the following system of equations are obtained:

$$\begin{bmatrix} R & H & 0 \\ H^T & 0 & C^T \\ 0 & C & 0 \end{bmatrix} \begin{bmatrix} \mu \\ \Delta x \\ \lambda \end{bmatrix} = \begin{bmatrix} \Delta z^k_{meas} \\ 0 \\ -c_{eq}(x^k) \end{bmatrix} \tag{3.15}$$

The coefficient matrix in equation (3.15) is called the Hachtel's matrix. Since the Hachtel's matrix is very sparse, solving the above enlarged system is not particularly expensive. However, the condition number of the Hachtel's matrix can be improved by scaling the residual matrix or in other words, by multiplying α^{-1}, where α is a scalar, to the co-variance matrix R in the Hachtel's matrix [27]. This results in the new Hachtel's matrix, which is given by

$$\begin{bmatrix} \alpha^{-1}R & H & 0 \\ H^T & 0 & C^T \\ 0 & C & 0 \end{bmatrix} \begin{bmatrix} \mu \\ \Delta x \\ \lambda \end{bmatrix} = \begin{bmatrix} \Delta z^k_{meas} \\ 0 \\ -c_{eq}(x^k) \end{bmatrix} \tag{3.16}$$

3.3 Mathematical modelling of unbalanced distribution system state estimation

3.3.1 Problem Formulation for distribution systems

The state estimation is a process which estimates real-time states of the system (voltage magnitudes and angles). The transmission system is, however, a special case of unbalanced system where, the system is balanced and hence, the number of state variables and equations are reduced. The problem for three phase unbalanced system can be looked at as a constrained non-linear optimization problem with the following objective function

$$J = [z_{meas} - h_{func}(x)]^T R^{-1} [z_{meas} - h_{func}(x)] \tag{3.17}$$

Subject to:

$$c_{eq}(x) = 0 \tag{3.18}$$

$$c_{ineq}(x) \leq 0 \tag{3.19}$$

Where,

x State variables such as voltage magnitudes and angles.

m Number of measurements per phase.

R Measurement error covariance matrix,

z_{meas} $= \left[z_1^a\ z_1^b\ z_1^c\ \dots\ z_i^a\ z_i^b\ z_i^c\ \dots\ z_m^a\ z_m^b\ z_m^c\right]^T$.

$z_i^{a,b,c}$ Measured value of i^{th} measurement.

$h_{func}(x)$ vector of measurement as a function of state x

$c_{eq}(x)$ vector of zero injection measurements and switch operational constraints.

$c_{ineq}(x)$ vector of inequality constraints.

In three phase system

$$x = \left[\delta_1^{ph} \cdots\ \delta_i^{ph} \cdots \delta_n^{ph}\ V_1^{ph} \cdots\ V_i^{ph} \cdots V_n^{ph}\right]^\top,$$

where,

$$V_i^{ph} = \begin{bmatrix} V_i^a \\ V_i^b \\ V_i^c \end{bmatrix},\ \delta_i^{ph} = \begin{bmatrix} \delta_i^a \\ \delta_i^b \\ \delta_i^c \end{bmatrix}$$

are the three-phase voltage magnitude and voltage angle at bus i respectively.

The measurements are usually considered subject to random errors due to biases, drifts or wrong connections of the measurement devices, i.e. meters. It is assumed that the measurement errors are identically and independently distributed. Hence, the covariance matrix of the errors is given by

$R{=}Cov(\mathbf{e}){=}E(\mathbf{e}\mathbf{e}^\mathbf{T}){=}diag(\sigma_1^2, \dots \sigma_i^2, \dots, \sigma_m^2)$, where $\sigma_i^2 = \left[(\sigma_i^a)^2\ (\sigma_i^b)^2\ (\sigma_i^c)^2\right]^\top$ and σ_i^2 is the variance of the i^{th} measurement error.

In three-phase system, the real power injection P_i^{ph} and reactive power injection Q_i^{ph} equations at bus i for phase ph can be written as:

$$P_i^{ph} = V_i^{ph} \sum_{l=1}^{3} \sum_{j=1}^{n} V_j^l \left[G_{ij}^{ph,l} \cos\left(\delta_i^{ph} - \delta_j^l\right) + B_{ij}^{ph,l} \sin\left(\delta_i^{ph} - \delta_j^l\right)\right] \tag{3.20}$$

$$Q_i^{ph} = V_i^{ph} \sum_{l=1}^{3} \sum_{j=1}^{n} V_j^l \left[G_{ij}^{ph,l} \sin\left(\delta_i^{ph} - \delta_j^l\right) - B_{ij}^{ph,l} \cos\left(\delta_i^{ph} - \delta_j^l\right)\right] \tag{3.21}$$

Where $G + jB$ is the system admittance matrix, n is number of buses and l is the number of phases that can be 1, 2 or 3 phase. The branch real power flow P_{ij}^{ph} and reactive power flow Q_{ij}^{ph} equations from bus i to bus j for phase ph can be written as follows:

$$P_{ij}^{ph} = V_i^{ph} \sum_{l=1}^{3} V_i^l \left[G_{ij}^{ph,l} \cos\left(\delta_i^{ph} - \delta_i^l\right) + B_{ij}^{ph,l} \sin\left(\delta_i^{ph} - \delta_i^l\right) \right] - V_i^{ph} \sum_{l=1}^{3} V_j^l \left[G_{ij}^{ph,l} \cos\left(\delta_i^{ph} - \delta_j^l\right) + B_{ij}^{ph,l} \sin\left(\delta_i^{ph} - \delta_j^l\right) \right] \quad (3.22)$$

$$Q_{ij}^{ph} = -V_i^{ph} \sum_{l=1}^{3} V_i^l \left[G_{ij}^{ph,l} \sin\left(\delta_i^{ph} - \delta_i^l\right) - B_{ij}^{ph,l} \cos\left(\delta_i^{ph} - \delta_i^l\right) \right] - V_i^{ph} \sum_{l=1}^{3} V_j^l \left[G_{ij}^{ph,l} \sin\left(\delta_i^{ph} - \delta_j^l\right) - B_{ij}^{ph,l} \cos\left(\delta_i^{ph} - \delta_j^l\right) \right] \quad (3.23)$$

Where,

V_i^l Voltage magnitude of phase l at bus i.

δ_i^l Angle of phase l in bus i.

3.3.2 Equality constraints $c_{eq}(x)$

The equality constraints are the set of equations corresponding to virtual measurements.

$$0 = P_{Gi}^{ph} - P_{Di}^{ph} = P_i^{ph} \quad (3.24)$$

$$0 = Q_{Gi}^{ph} - Q_{Di}^{ph} = Q_i^{ph} \quad (3.25)$$

Where P_{Gi}^{ph} and Q_{Gi}^{ph} are the real and reactive power injected at bus i respectively, the load demand at the same bus is represented by P_{Di}^{ph} and Q_{Di}^{ph}.

The operational constraints for open/closed switches are considered to equality constraints. These are a set of equations given as

- Closed switch

$$\begin{aligned} V_i^{ph} - V_j^{ph} &= 0 \\ \delta_i^{ph} - \delta_j^{ph} &= 0 \end{aligned} \quad (3.26)$$

- Open switch

$$\begin{aligned} P_{ij}^{ph} &= 0 \\ Q_{ij}^{ph} &= 0 \end{aligned} \tag{3.27}$$

3.3.3 Inequality constraints $c_{ineq}(x)$

These are the set of constraints on state variables that represent the system operational and security limits, such as setting upper and lower limits for control variables. The constraints are as follows:

- **Bus voltage** - Voltage magnitudes at each bus in the network:

$$V_{min,i}^{ph} \leq V_i^{ph} \leq V_{max,i}^{ph} \tag{3.28}$$

- **Bus angle** - The bus angle at each bus in the network:

$$-\delta_{min,i}^{ph} \leq \delta_i^{ph} \leq \delta_{max,i}^{ph} \tag{3.29}$$

The above equations are solved by the primal-dual interior point method. The optimization problem has both equality and inequality constraints. Hence, the method of primal-dual interior point with slack variables for the inequality constraints is used.

$\Delta x^{k+1} = x^{k+1} - x^k$ and $H(x^k) = [\frac{\partial h_{func}}{\partial x}]_{x^k}$ is the Jacobian matrix of dimension $3m \times (n-3)$. Where, n is the total number of buses in the system. The angles of a particular bus (bus #1) are taken as references, such as 0° for phase a, -120° for phase b and 120° for phase c. $h_{func}(x)$ is the measurement function of measurements as a function of state variables. A measurement can be voltage magnitude, real and reactive power flows, branch currents, real and reactive power injection. A typical distribution system does not have many measurements. Most of the measurements are branch current measurements. There are very few branch power measurements , a very few injection measurements and voltage magnitude is only measured in the substation bus. So most of the measurements are loads or pseudo measurements and some are virtual or zero-injection measurements. The solution methodology as

discussed in Section 3.2 requires the calculation of the Jacobian matrix H at each iteration. The Jacobian matrix is the matrix of partial derivatives of measurement functions with respect to the state variables of the system. The Jacobian matrix can be expressed as:

$$H = \begin{bmatrix} \frac{\partial h_1(x)}{\partial \delta_2} & \cdots & \frac{\partial h_1(x)}{\partial \delta_n} & \frac{\partial h_1(x)}{\partial V_1} & \cdots & \frac{\partial h_1(x)}{\partial V_n} \\ \frac{\partial h_2(x)}{\partial \delta_2} & \cdots & \frac{\partial h_2(x)}{\partial \delta_n} & \frac{\partial h_2(x)}{\partial V_1} & \cdots & \frac{\partial h_2(x)}{\partial V_n} \\ \vdots & \vdots & \vdots & \vdots & & \\ \vdots & \vdots & \vdots & \vdots & & \\ \frac{\partial h_m(x)}{\partial \delta_2} & \cdots & \frac{\partial h_m(x)}{\partial \delta_n} & \frac{\partial h_m(x)}{\partial V_1} & \cdots & \frac{\partial h_m(x)}{\partial V_n} \end{bmatrix} \tag{3.30}$$

The measurement functions and the derivation of Jacobian elements related to Newton's method are described in details in Section 3.4.

3.4 Measurement model

Considering the vector z_{meas} as a set of measurements and vector x as the state vector, the non-linear measurement functions $h_{func}(x)$ are given below.

Measurement Function

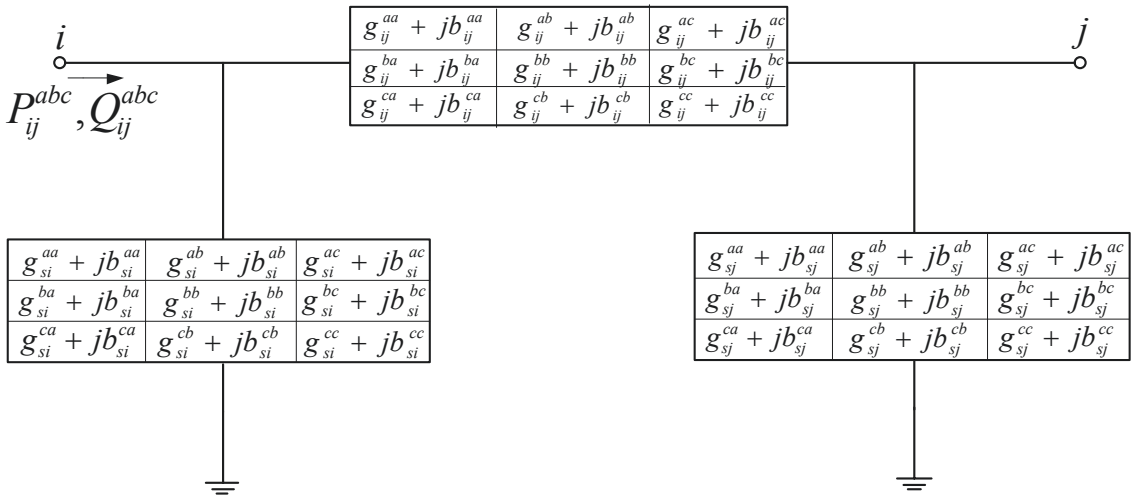

Figure 3.2: Two-port pi-model of a network

The general two-port π-model for the network branches for a transmission system is shown in Figure 3.2. The same model can be expanded to the three-phase distribution system such that each element of the π-model for network branches is a 3x3 matrix rather than a single element. This model has been used to relate the state vector to each type of measurements [49].

Bus Power Injection

The real and reactive power injections at the i^{th} bus for phase ph are given by:

$$P_i^{ph} = \sum_{l=1}^{3}\sum_{j=1}^{n} V_i^{ph} V_j^{l} [G_{ij}^{ph,l}\cos(\delta_i^{ph} - \delta_j^{l}) + B_{ij}^{ph,l}\sin(\delta_i^{ph} - \delta_j^{l})] \tag{3.31}$$

$$Q_i^{ph} = -\sum_{l=1}^{3}\sum_{j=1}^{n} V_i^{ph} V_j^{l} [B_{ij}^{ph,l}\cos(\delta_i^{ph} - \delta_j^{l}) - G_{ij}^{ph,l}\sin(\delta_i^{ph} - \delta_j^{l})] \tag{3.32}$$

Considering the two-port π-model of the network branches, the line power flows between bus i and bus j for a phase ph are given by:

$$P_{ij}^{ph} = \sum_{l=1}^{3} [V_i^{ph} V_i^{l} (g_{ij}^{ph,l} + g_{si}^{ph,l}) - V_i^{ph} V_j^{l} (g_{ij}^{ph,l}\cos(\delta_i^{ph} - \delta_j^{l})) + b_{ij}^{ph,l}\sin(\delta_i^{ph} - \delta_j^{l})] \tag{3.33}$$

$$Q_{ij}^{ph} = -\sum_{l=1}^{3} [V_i^{ph} V_i^{l} (b_{ij}^{ph,l} + b_{si}^{ph,l}) + V_i^{ph} V_j^{l} (g_{ij}^{ph,l}\sin(\delta_i^{ph} - \delta_j^{l}) + b_{ij}^{ph,l}\cos(\delta_i^{ph} - \delta_j^{l}))] \tag{3.34}$$

The line current between bus i and bus j for a phase ph is given by:

$$I_{ij}^{ph} = \frac{\sqrt{(P_{ij}^{ph})^2 + (Q_{ij}^{ph})^2}}{V_i^{ph}} \tag{3.35}$$

As the above equations are non-linear, so the vector measurement functions are linearised around an operating point. Hence, it is required to compute the Jacobian matrix whose elements are the first order derivatives of the measurement functions with respect to the state variables. The Jacobian elements corresponding to the real

and reactive power injections are:

$$\begin{aligned}\frac{\partial P_i^{ph}}{\partial \delta_i^l}_{l=ph} &= \sum_{\substack{j=1\\ j\neq i}}^{n} V_i^{ph} V_j^{ph} [-G_{ij}^{ph,ph} \sin(\delta_i^{ph} - \delta_j^{ph}) + B_{ij}^{ph,ph} \cos(\delta_i^{ph} - \delta_j^{ph})] \\ &+ \sum_{\substack{l=1\\ l\neq ph}}^{3} \sum_{j=1}^{n} V_i^{ph} V_j^{l} [-G_{ij}^{ph,l} \sin(\delta_i^{ph} - \delta_i^{l}) + B_{ij}^{ph,l} \cos(\delta_i^{ph} - \delta_j^{l})] \\ &= -Q_i^{ph} - (V_i^{ph})^2 B_{ii}^{ph,ph}\end{aligned} \tag{3.36}$$

$$\frac{\partial P_i^{ph}}{\partial \delta_i^l}_{l\neq ph} = V_i^{ph} V_i^{l} [G_{ii}^{ph,l} \sin(\delta_i^{ph} - \delta_i^{l}) - B_{ii}^{ph,l} \cos(\delta_i^{ph})] \tag{3.37}$$

$$\frac{\partial P_i^{ph}}{\partial \delta_j^l} = V_i^{ph} V_j^{l} [G_{ij}^{ph,l} \sin(\delta_i^{ph} - \delta_j^{ph}) - B_{ij}^{ph,l} \cos(\delta_i^{ph} - \delta_j^{l})] \tag{3.38}$$

$$\begin{aligned}\frac{\partial P_i^{ph}}{\partial V_i^l}_{l=ph} &= 2V_i^{ph} G_{ii}^{ph,ph} + \sum_{\substack{j=1\\ j\neq i}}^{n} V_j^{ph} [G_{ij}^{ph,ph} \cos(\delta_i^{ph} - \delta_j^{ph}) + B_{ij}^{ph,ph} \sin(\delta_i^{ph} - \delta_j^{ph})] \\ &+ \sum_{\substack{l=1\\ l\neq ph}}^{3} \sum_{j=1}^{n} V_j^{l} [G_{ij}^{ph,l} \cos(\delta_i^{ph} - \delta_j^{l}) + B_{ij}^{ph,l} \sin(\delta_i^{ph} - \delta_j^{l})] + P_n(c_2^p(\frac{1}{V_n}) + c_3^p(\frac{V}{V_n^2})) \\ &= \frac{P_i^{ph}}{V_i^{ph}} + V_i^{ph} G_{ii}^{ph,l} + P_n(c_2^p(\frac{1}{V_n}) + c_3^p(\frac{V}{V_n^2}))\end{aligned} \tag{3.39}$$

$$\frac{\partial P_i^{ph}}{\partial V_j^l} = V_i^{ph} [G_{ij}^{ph,l} \cos(\delta_i^{ph} - \delta_j^{l}) + B_{ij}^{ph,l} \sin(\delta_i^{ph} - \delta_j^{l})] \tag{3.40}$$

$$\begin{aligned}\frac{\partial Q_i^{ph}}{\partial \delta_i^l}_{l=ph} &= -\sum_{\substack{j=1\\ j\neq i}}^{n} V_i^{ph} V_j^{ph} [-B_{ij}^{ph,l} \sin(\delta_i^{ph} - \delta_j^{ph}) - G_{ij}^{ph,l} \cos(\delta_i^{ph} - \delta_j^{ph})] \\ &- \sum_{\substack{l=1\\ l\neq ph}}^{3} \sum_{j=1}^{n} V_i^{ph} V_j^{l} [-B_{ij}^{ph,l} \sin(\delta_i^{ph} - \delta_i^{l}) - G_{ij}^{ph,l} \cos(\delta_i^{ph} - \delta_j^{l})] \\ &= P_i^{ph} - (V_i^{ph})^2 G_{ii}^{ph,ph}\end{aligned} \tag{3.41}$$

$$\frac{\partial Q_i^{ph}}{\partial \delta_i^l}_{l=ph} = -V_i^{ph} V_i^{l} [B_{ii}^{ph,l} \sin(\delta_i^{ph} - \delta_i^{l}) + G_{ii}^{ph,l} \cos(\delta_i^{ph} - \delta_i^{l})] \tag{3.42}$$

$$\frac{\partial Q_i^{ph}}{\partial \delta_j^l} = -V_i^{ph} V_j^{l} [B_{ij}^{ph,l} \sin(\delta_i^{ph} - \delta_j^{l}) + G_{ij}^{ph,l} \cos(\delta_i^{ph} - \delta_j^{l})] \tag{3.43}$$

$$\begin{aligned}\frac{\partial Q_i^{ph}}{\partial V_i^l}_{l=ph} &= -2V_i^{ph}B_{ii}^{ph,ph} - \sum_{\substack{j=1\\ j\neq i}}^{n} V_j^{ph}[B_{ij}^{ph,ph}\cos(\delta_i^{ph}-\delta_j^{ph}) - G_{ij}^{ph,ph}\sin(\delta_i^{ph}-\delta_j^{ph})] \\ &\quad - \sum_{\substack{l=1\\ l\neq ph}}^{3}\sum_{j=1}^{n} V_j^{l}[B_{ij}^{ph,l}\cos(\delta_i^{ph}-\delta_j^{ph}) - G_{ij}^{ph,l}\sin(\delta_i^{ph}-\delta_j^{ph})] + Q_n(c_2^Q(\frac{1}{V_n}) + c_3^Q(\frac{V}{V_n^2})) \\ &= -\frac{Q_i^{ph}}{V_i^{ph}} - V_i^{ph}B_{ii}^{ph,l} + Q_n(c_2^Q(\frac{1}{V_n}) + c_3^Q(\frac{V}{V_n^2}))\end{aligned} \tag{3.44}$$

$$\frac{\partial Q_i^{ph}}{\partial V_j^l} = -V_i^{ph}[B_{ij}^{ph,l}\cos(\delta_i^{ph}-\delta_j^{l}) - G_{ij}^{ph,l}\sin(\delta_i^{ph}-\delta_j^{l})] \tag{3.45}$$

The Jacobian elements corresponding to the real and reactive power flows between bus i and bus j are:

$$\begin{aligned}\frac{\partial P_{ij}^{ph}}{\partial \delta_i^l}_{l=ph} &= -\sum_{l=1}^{3} V_i^k V_j^l[-g_{ij}^{ph,l}\sin(\delta_i^{ph}-\delta_j^{l}) + b_{ij}^{ph,l}\cos(\delta_i^{ph}-\delta_j^{l})] \\ &= -Q_{ij}^{ph} - (V_i^{ph})^2[b_{ij}^{ph,l} + b_{si}^{ph,l}]\end{aligned} \tag{3.46}$$

$$\frac{\partial P_{ij}^{ph}}{\partial \delta_i^l}_{l\neq ph} = V_i^{ph}V_i^{l}[g_{ij}^{ph,l}\sin(\delta_i^{ph}-\delta_i^{l}) - b_{ij}^{ph,l}\cos(\delta_i^{ph}-\delta_i^{l})] \tag{3.47}$$

$$\frac{\partial P_{ij}^{ph}}{\partial \delta_j^l} = -V_i^{ph}V_j^{l}[g_{ij}^{ph,l}\sin(\delta_i^{ph}-\delta_j^{l}) - b_{ij}^{ph,l}\cos(\delta_i^{ph}-\delta_j^{l})] \tag{3.48}$$

$$\begin{aligned}\frac{\partial P_{ij}^{ph}}{\partial V_i^l}_{l=ph} &= V_i^{ph}[g_{ij}^{ph,ph} + g_{si}^{ph,l}] \\ &\quad + \sum_{l=1}^{3}[V_i^l g_{ij}^{ph,ph} + g_{si}^{ph,l} - V_j^l[g_{ij}^{ph,l}\cos(\delta_i^{ph}-\delta_j^{l}) + b_{ij}^{ph,l}\sin(\delta_i^{ph}-\delta_j^{l})]] \\ &= \frac{P_{ij}^{ph}}{V_i^{ph}} + V_i^{ph}[g_{ij}^{ph,ph} + g_{si}^{ph,l}]\end{aligned} \tag{3.49}$$

$$\frac{\partial P_{ij}^{ph}}{\partial V_i^l}_{l\neq ph} = V_i^{ph}[g_{ij}^{ph,l}\cos(\delta_i^{ph}-\delta_i^{l}) + b_{ij}^{ph,l}\sin(\delta_i^{ph}-\delta_i^{l}) + g_{si}^{ph,l}] \tag{3.50}$$

$$\frac{\partial P_{ij}^{ph}}{\partial V_j^l} = -V_i^{ph}[g_{ij}^{ph,l}\cos(\delta_i^{ph}-\delta_j^{l}) + b_{ij}^{ph,l}\sin(\delta_i^{ph}-\delta_j^{l})] \tag{3.51}$$

$$\frac{\partial Q_{ij}^{ph}}{\partial \delta_i^l}_{l=ph} = -\sum_{l=1}^{3} V_i^{ph} V_j^l [g_{ij}^{ph,l} \cos(\delta_i^{ph} - \delta_j^l) + b_{ij}^{ph,l} \sin(\delta_i^{ph} - \delta_j^l)] \tag{3.52}$$

$$= P_{ij}^{ph} - (V_i^{ph})^2 [g_{ij}^{ph,ph} + g_{si}^{ph,l}]$$

$$\frac{\partial Q_{ij}^{ph}}{\partial \delta_i^l}_{l\neq ph} = V_i^{ph} V_i^l [-g_{ij}^{ph,l} \cos(\delta_i^{ph} - \delta_i^l) - b_{ij}^{ph} \sin(\delta_i^{ph} - \delta_i^l)] \tag{3.53}$$

$$\frac{\partial Q_{ij}^{ph}}{\partial \delta_j^l} = -V_i^{ph} V_j^l [-g_{ij}^{ph,l} \cos(\delta_i^{ph} - \delta_j^l) - b_{ij}^{ph,l} \sin(\delta_i^{ph} - \delta_j^l)] \tag{3.54}$$

$$\frac{\partial Q_{ij}^{ph}}{\partial V_i^l}_{l=ph} = -V_i^{ph} [b_{ij}^{ph,l} + b_{si}^{ph,l}]$$

$$+ \sum_{l=1}^{3} [-V_i^l (b_{ij}^{ph,l} + b_{si}^{ph,l}) - V_j^l [g_{ij}^{ph,l} \sin(\delta_i^{ph} - \delta_j^l) - b_{ij}^{ph,l} \cos(\delta_i^{ph} - \delta_j^l)]] \tag{3.55}$$

$$= \frac{Q_{ij}^{ph}}{V_i^{ph}} - V_i^{ph} [b_{ij}^{ph,l} + b_{si}^{ph,l}]$$

$$\frac{\partial Q_{ij}^{ph}}{\partial V_i^l}_{l\neq ph} = V_i^{ph} [g_{ij}^{ph,l} \sin(\delta_i^{ph} - \delta_i^l) - b_{ij}^{ph,l} \cos(\delta_i^{ph} - \delta_i^l) + b_{si}^{ph,l}] \tag{3.56}$$

$$\frac{\partial Q_{ij}^{ph}}{\partial V_j^l} = -V_i^{ph} [g_{ij}^{ph,l} \sin(\delta_i^{ph} - \delta_j^l) - b_{ij}^{ph,l} \cos(\delta_i^{ph} - \delta_j^l)] \tag{3.57}$$

The Jacobian elements corresponding to the branch current flows between bus i and bus j are:

$$\frac{\partial I_{ij}^{ph}}{\partial \delta_i^l} = \frac{real(I_{ij}^{ph})}{|I_{ij}^{ph}|} [V_i^l g_{ij}^{ph,l} \sin \delta_i^l + b_{ij}^{ph,l} \cos \delta_i^l]$$

$$+ \frac{imag(I_{ij}^{ph})}{|I_{ij}^{ph}|} [V_i^l - g_{ij}^{ph,l} \cos \delta_i^l + b_{ij}^{ph,l} \sin \delta_i^l] \tag{3.58}$$

$$\frac{\partial I_{ij}^{ph}}{\partial \delta_j^l} = \frac{real(I_{ij}^{ph})}{|I_{ij}^{ph}|} [V_j^l - g_{ij}^{ph,l} \sin \delta_j^l - b_{ij}^{ph,l} \cos \delta_j^l]$$

$$+ \frac{imag(I_{ij}^{ph})}{|I_{ij}^{ph}|} [-V_j^l - g_{ij}^{ph,l} \cos \delta_j^l + b_{ij}^{ph,l} \sin \delta_j^l] \tag{3.59}$$

$$\frac{\partial I_{ij}^{ph}}{\partial V_i^l} = \frac{real(I_{ij}^{ph})}{|I_{ij}^{ph}|} [-g_{ij}^{ph,l} \cos \delta_i^l + b_{ij}^{ph,l} \sin \delta_i^l]$$

$$+ \frac{imag(I_{ij}^{ph})}{|I_{ij}^{ph}|} [-g_{ij}^{ph,l} \sin \delta_i^l - b_{ij}^{ph,l} \cos \delta_i^l] \tag{3.60}$$

$$\frac{\partial I_{ij}^{ph}}{\partial V_j^l} = \frac{real(I_{ij}^{ph})}{|I_{ij}^{ph}|}[g_{ij}^{ph,l}\cos\delta_j^l - b_{ij}^{ph,l}\sin\delta_j^l] + \frac{imag(I_{ij}^{ph})}{|I_{ij}^{ph}|}[g_{ij}^{ph,l}\sin\delta_j^l + b_{ij}^{ph,l}\cos\delta_j^l] \tag{3.61}$$

The Jacobian elements corresponding to the voltage magnitude measurements at bus i are:

$$\frac{\partial V_i^{ph}}{\partial \delta_i^l} = 0 \tag{3.62}$$

$$\frac{\partial V_i^{ph}}{\partial \delta_j^l} = 0 \tag{3.63}$$

$$\frac{\partial V_i^{ph}}{\partial V_i^l}_{l=ph} = 1 \tag{3.64}$$

$$\frac{\partial V_i^{ph}}{\partial V_i^l}_{l\neq ph} = 0 \tag{3.65}$$

$$\frac{\partial V_i^{ph}}{\partial V_j^l}_{i\neq j} = 0 \tag{3.66}$$

where, n is the number of buses and l is the phase index $G_{ij}^{ph,l} + B_{ij}^{ph,l}$ is the ij^{th} element of the bus admittance matrix between phase ph and phase l and $g_{ij}^{ph,l} + b_{ij}^{ph,l}$ is the admittance of the series branch connecting bus i and bus j between phase ph and phase l.

The Jacobian H is used in every iteration to find the Gain matrix and calculate the correction vector.

3.5 Case studies and discussions

3.5.1 Simulation results

IEEE 13-bus system

A standard IEEE-13 bus distribution system has been used here. The feeders are small yet they show some interesting characteristics as discussed in Section 2.7 of Chapter 2. The system represents a typical distribution system with voltage

magnitude measurements only at the substation, with more branch current measurements than power flow measurements, and all loads are considered as pseudo measurements. Phase c for bus 3 and phase a and c for bus 10 are zero injection or virtual measurements.

Figure 2.2 of Chapter 2 shows the typical IEEE 13-bus system. A WLS state estimator is coded in Matlab and tested on the standard IEEE-13 bus system and run on a system with Intel Xeon processor @3.33GHz and 12 GB RAM. No Matlab toolbox was used. The complete system data are given in [51], [52] and Appendix A.

The overhead lines and underground cables are modelled by modified Carson's equations. The loads are modelled in ZIP model and the three-phase transformer is configured as a grounded Y-Y connection. Measurements have been generated using normal distribution curve with load flow values as true or mean values and standard deviation. Each measurement is taken from the distribution curve randomly and this experiment is performed a number of times in a Monte Carlo approach. One such case has been shown here, in the results. It is assumed that the measurements are taken from independent meters placed at different locations (nodes and branches). Hence, the measurement errors are assumed to be independent and identically distributed. The zero injections are considered as equality constraints. The voltage magnitudes are set to operate within 5% of the nominal values and the voltage angles within $-30°$ to $+30°$. The switch between buses 9 and 10 is assumed to be closed. Therefore, equality constraints in the state estimation formulation are used as shown in equation (2.9). The loads on nodes 2, 9 and 10 are delta-configured loads. The constant impedance loads are on nodes 2 and 10, while nodes 10 and 7 have constant current loads.

Table 3.1 present the load flow values of the state variables and Table 3.2 shows their estimates. The error in real measurement is assumed to be 3% and the error in pseudo measurement is assumed to be 20%. The load flow values are considered to be the true values. Figure 3.3 show the true and estimated voltage magnitudes of the three phases for the IEEE-13 bus model. For the ease of Matlab coding it is assumed that at each iteration the missing phases are of voltage magnitude value

Table 3.1: Load Flow Results

	Angle(in degrees)			Voltage magnitude(in pu)		
Bus No.	ph a	ph b	ph c	ph a	ph b	ph c
1	0	-120	120	1	1	1
2	-	-119.565	119.357	-	0.9888	1.0063
3	-	-119.6173	119.4225	-	0.9893	1.0062
4	-0.2412	-119.6838	119.5482	0.9943	0.9923	1.0054
5	-0.264	-119.7004	119.5467	0.9940	0.9916	1.0045
6	-0.4976	-119.8652	119.3859	0.9960	0.9930	1.0082
7	-	-	119.3912	-	-	1.0069
8	-0.6488	-	119.414	0.9886	-	1.008
9	-0.631	-119.2481	119.4222	0.9893	0.9936	1.0092
10	-0.631	-119.2481	119.4222	0.9893	0.9936	1.0092
11	-0.6725	-119.258	119.4517	0.9881	0.9939	1.0083
12	-0.6242	-	-	0.9868	-	-
13	-0.6317	-119.2475	119.4221	0.9890	0.9936	1.0091

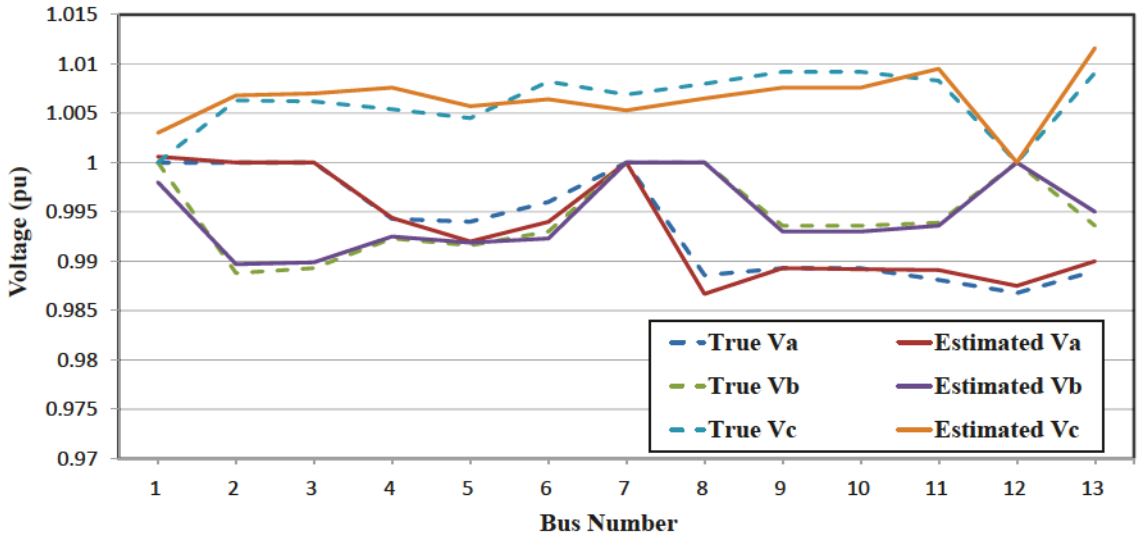

Figure 3.3: True and estimated voltages for IEEE 13 bus system

Table 3.2: State Estimates

	Angle estimates(in degrees)			Voltage estimates(in pu)		
Bus No.	ph a	ph b	ph c	ph a	ph b	ph c
1	0	-120	120	1.0006	0.9980	1.0030
2	-	-119.566	119.5905	-	0.9897	1.0068
3	-	-119.6192	119.6037	-	0.9899	1.0070
4	-0.2794	-119.6875	119.6423	0.9944	0.9925	1.0076
5	-0.3005	-119.7033	119.6340	0.9920	0.9919	1.0057
6	-0.5364	-119.8673	119.4340	0.9940	0.9923	1.0064
7	-	-	119.4432	-	-	1.0053
8	-0.6496	-	119.4655	0.9867	-	1.0065
9	-0.6319	-119.2681	119.4736	0.9893	0.9930	1.0076
10	-0.6319	-119.2681	119.4736	0.9892	0.9930	1.0076
11	-0.6519	-119.2604	119.4922	0.9891	0.9936	1.0095
12	-0.6253	-	-	0.9875	-	-
13	-0.6323	-119.2676	119.4736	0.9900	0.9950	1.0116

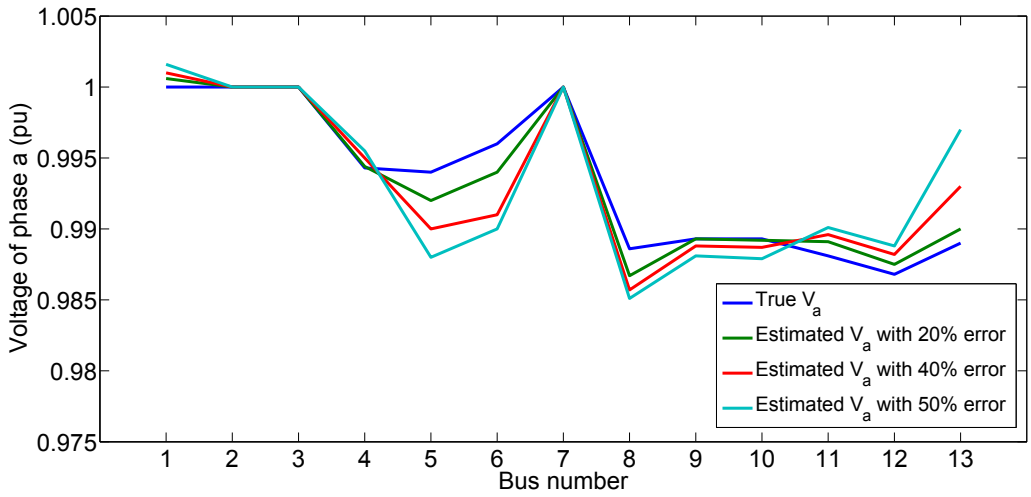

Figure 3.4: True and estimated voltages for phase a with 20% 40% and 50% error

equal to 1. However, the state estimation process is performed with 40% and 50% error in pseudo measurements as well. Figure 3.4 shows that for cases when the error in pseudo measurement is large the state estimates are less accurate. The figure shows that when the error is 20% the estimates for phase a are closest to the true value compared to other cases. In Tables 3.1 and 3.2, the missing phases have been represented by dashes. There is a closed switch between bus 9 and 10. Figure 3.3 and Tables 3.1 and 3.2 show that the voltage magnitude values and voltage angles remain the same across the closed switch. The obtained results have been found to be satisfactory within the allowable tolerance ($\pm 3\sigma$) from equation (3.12).

IEEE 123-bus system

The IEEE 123-bus system has also been considered as a case study. The system test feeder has a nominal voltage of 4.16 kV and the system has some remarkable features as mentioned in Section 2.7 of Chapter 2.

The lines are modelled according to Carson's equations as in [49] and the loads are modelled in ZIP-model. There is one on-load tap changer (OLTC) (the secondary of OLTC is on node 1), three step voltage regulators (between nodes 10 and 15, 118 and 68, 26 and 27) and shunt capacitor banks (on nodes 84, 89, 91 and 93). The switches in the system provide optimal configuration options. A three-phase transformer is modelled as three individual single-phase transformers. The tap changers are considered to have fixed taps. The details of the system are given in [51, 54]. The topology of the system is shown in Figure 2.3 of the last chapter. The cases for two cases have been shown here. The one with switch between 19-116 closed and between 52-121 open. In the second case, the switch between 52-121 closed and between 19-116 open. The switches between buses 14 and 117, 61 and 118, 98 and 119 are considered closed in both the cases.

As for the previous case study, measurements have been generated using normal distribution curve with load flow values as true or mean values and standard deviation. Each measurement is taken from the distribution curve randomly and this experiment is performed a number of times in a Monte Carlo approach. One such case has been shown here, in the results. It is assumed that the measurement

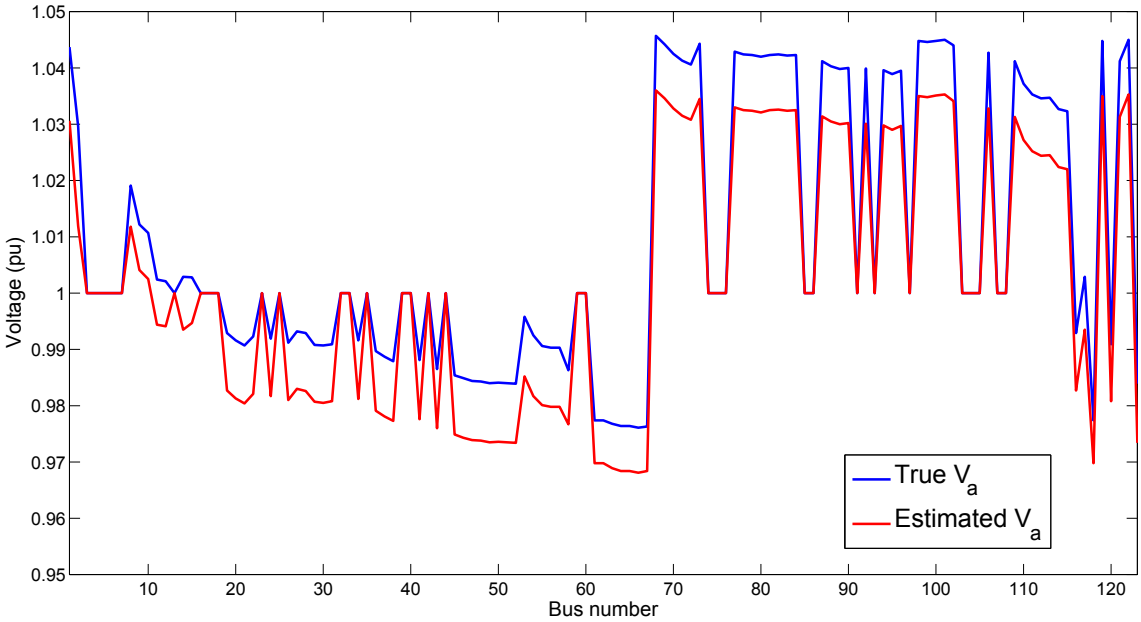

Figure 3.5: True and estimated voltages for IEEE 123 bus system

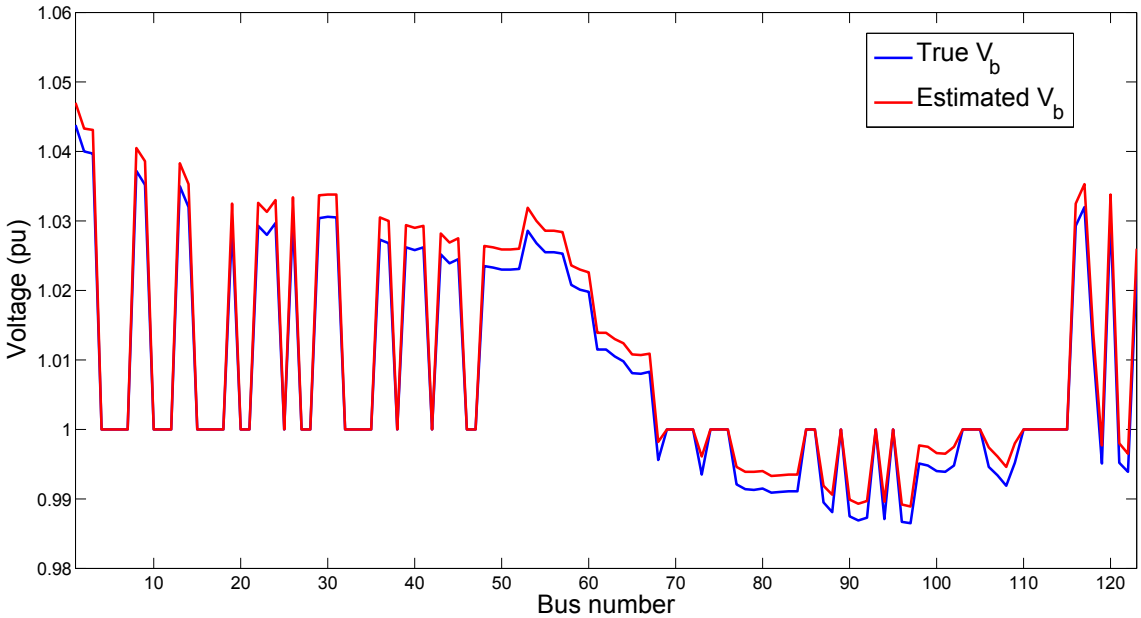

Figure 3.6: True and estimated voltages for IEEE 123 bus system

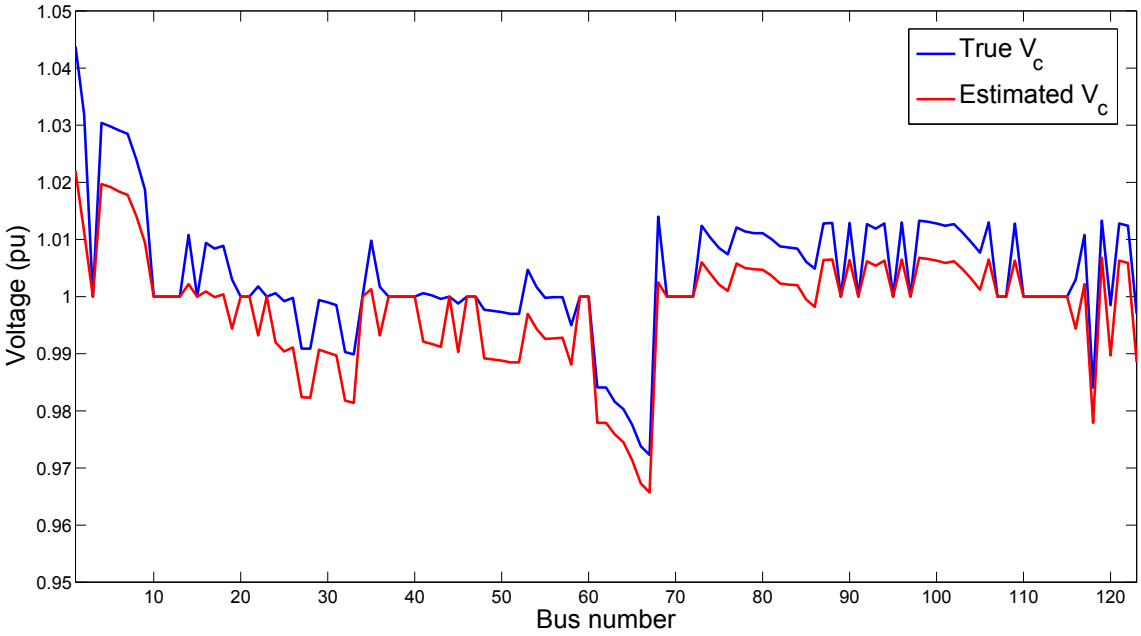

Figure 3.7: True and estimated voltages for IEEE 123 bus system

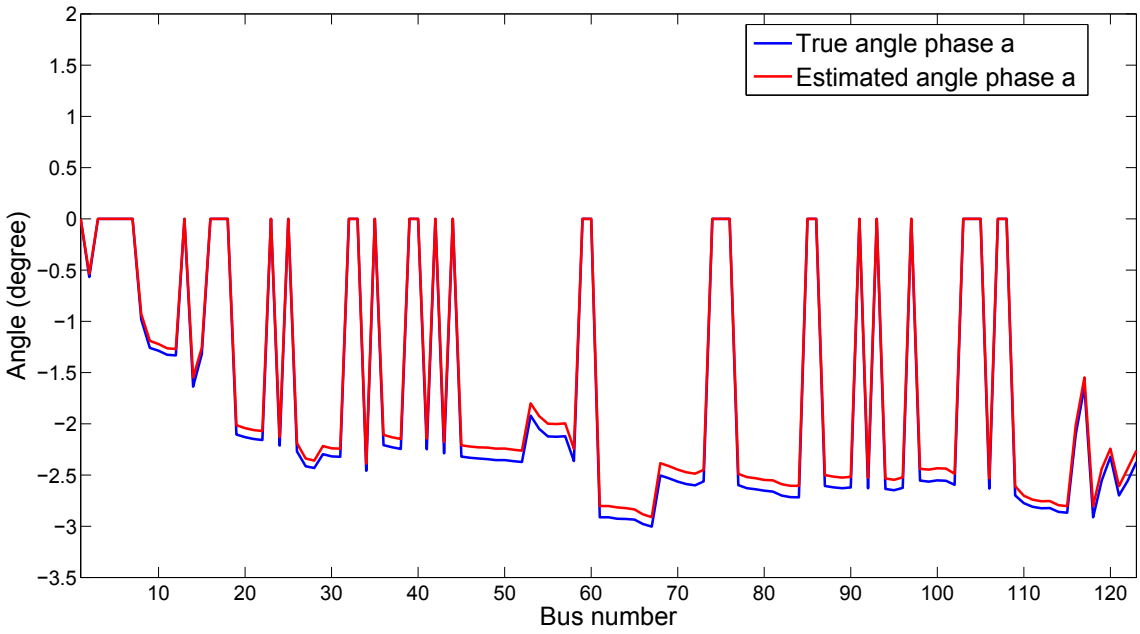

Figure 3.8: True and estimated voltage angles for IEEE 123 bus system

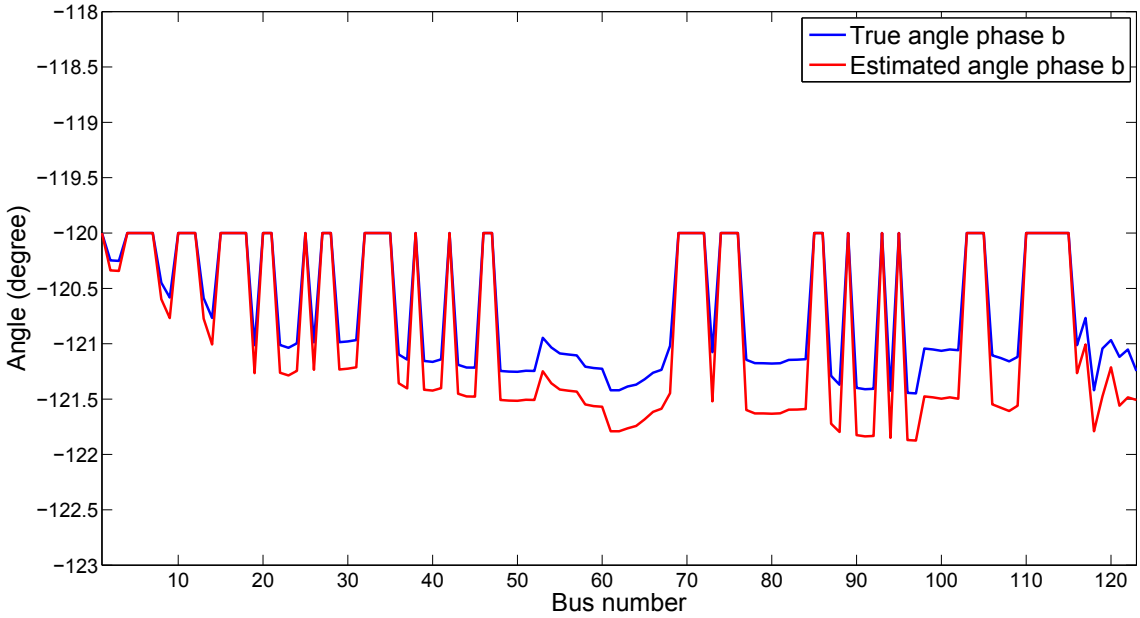

Figure 3.9: True and estimated voltage angles for IEEE 123 bus system

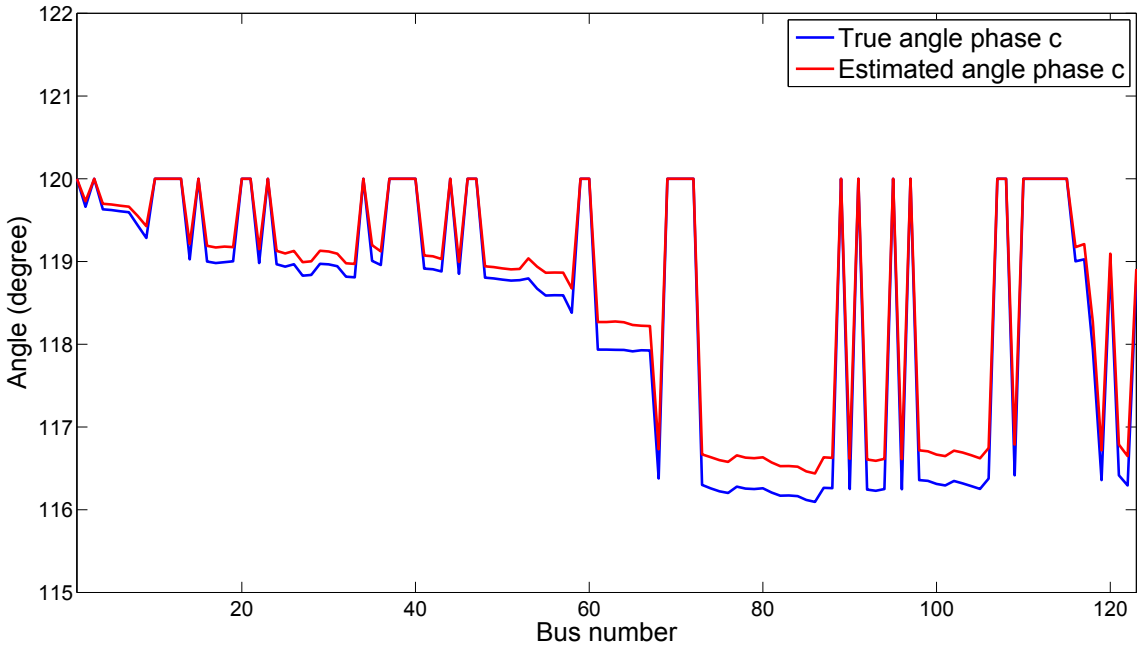

Figure 3.10: True and estimated voltage angles for IEEE 123 bus system

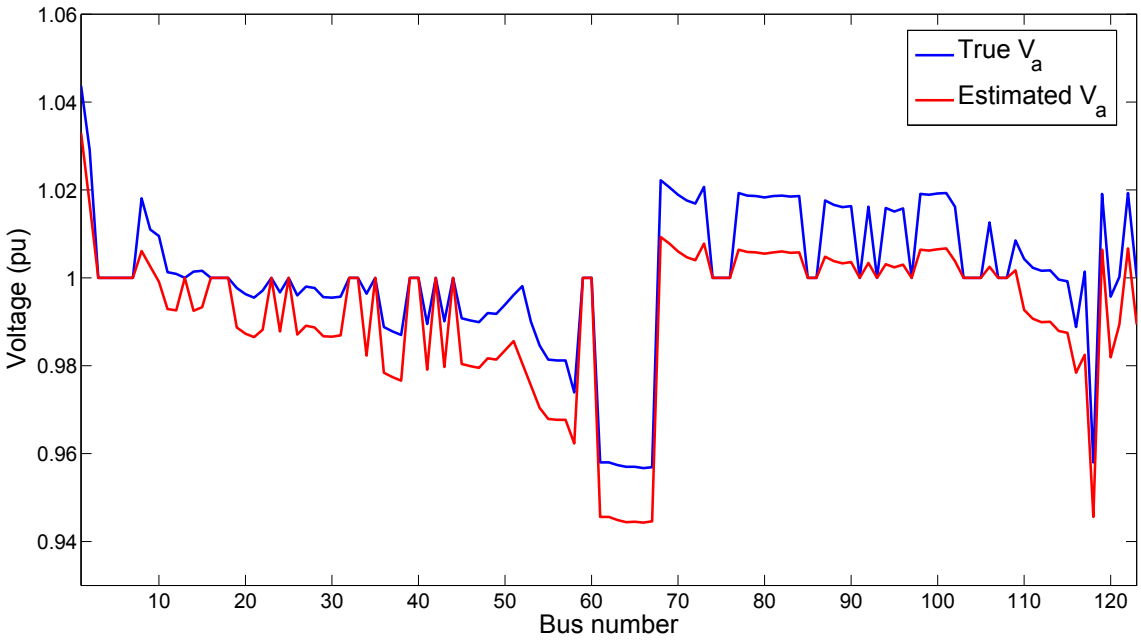

Figure 3.11: True and estimated voltages for IEEE 123 bus system with changed switch status

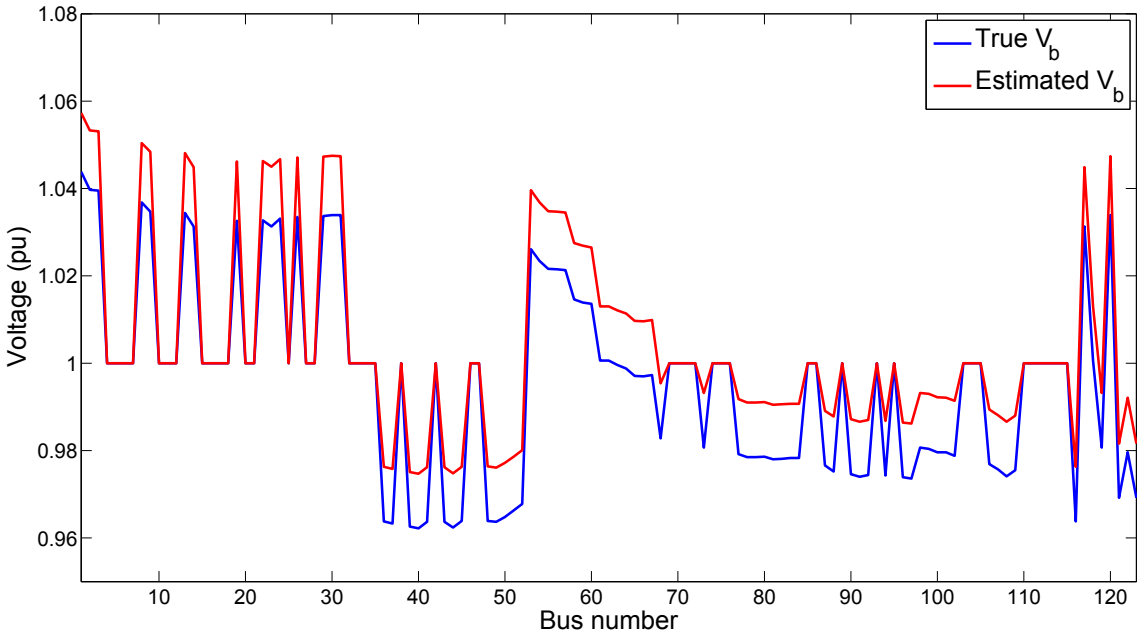

Figure 3.12: True and estimated voltages for IEEE 123 bus system with changed switch status

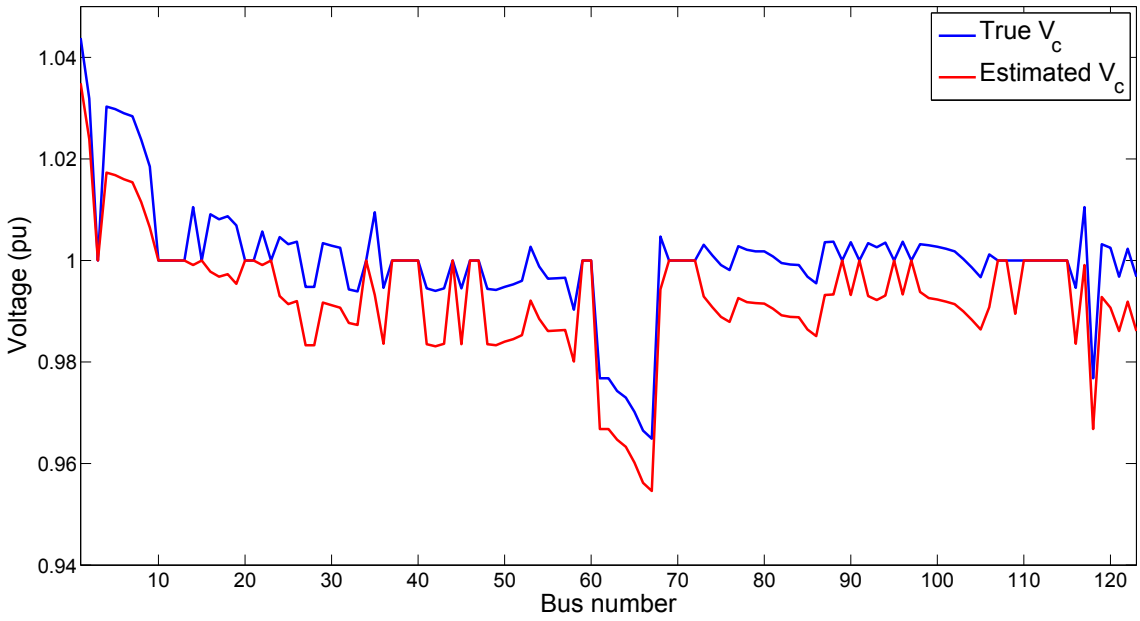

Figure 3.13: True and estimated voltages for IEEE 123 bus system with changed switch status

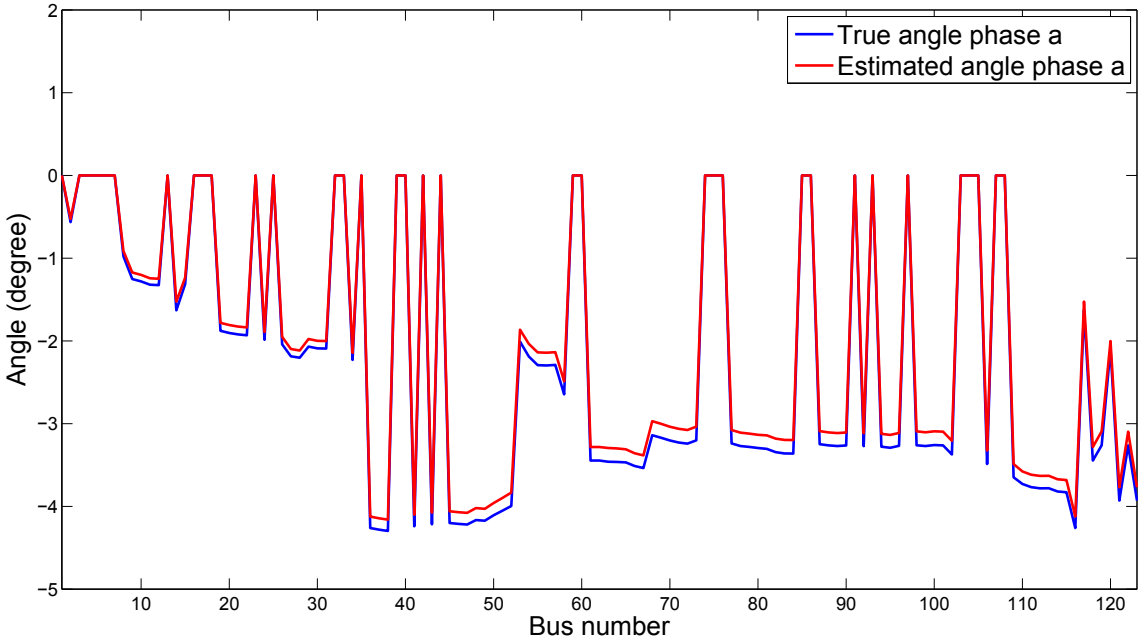

Figure 3.14: True and estimated voltage angles for IEEE 123 bus system with changed switch status

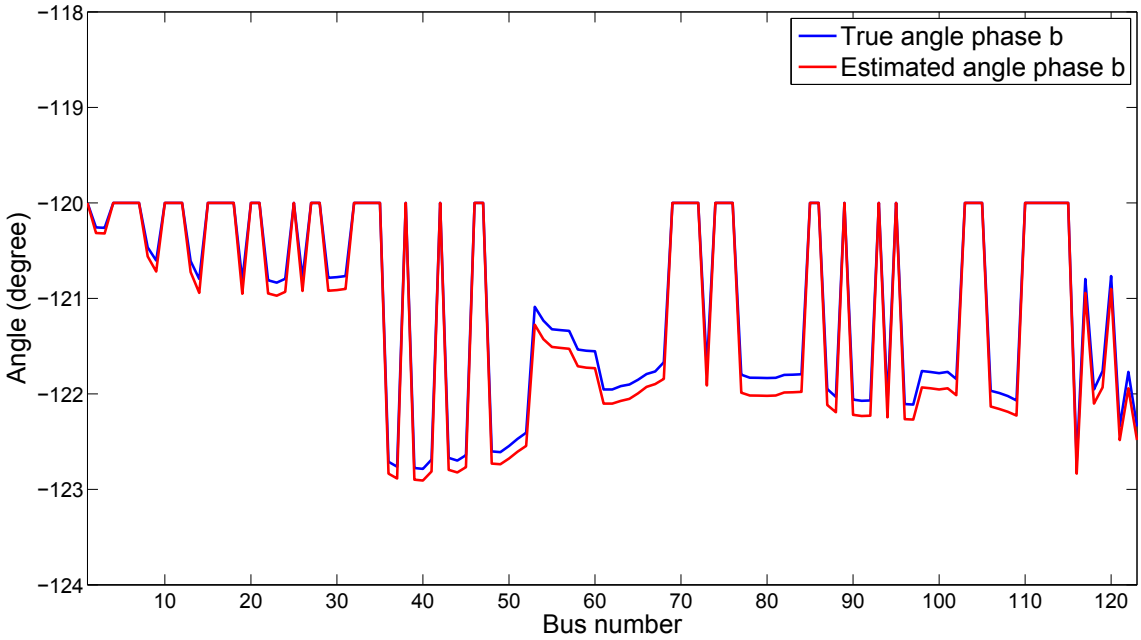

Figure 3.15: True and estimated voltage angles for IEEE 123 bus system with changed switch status

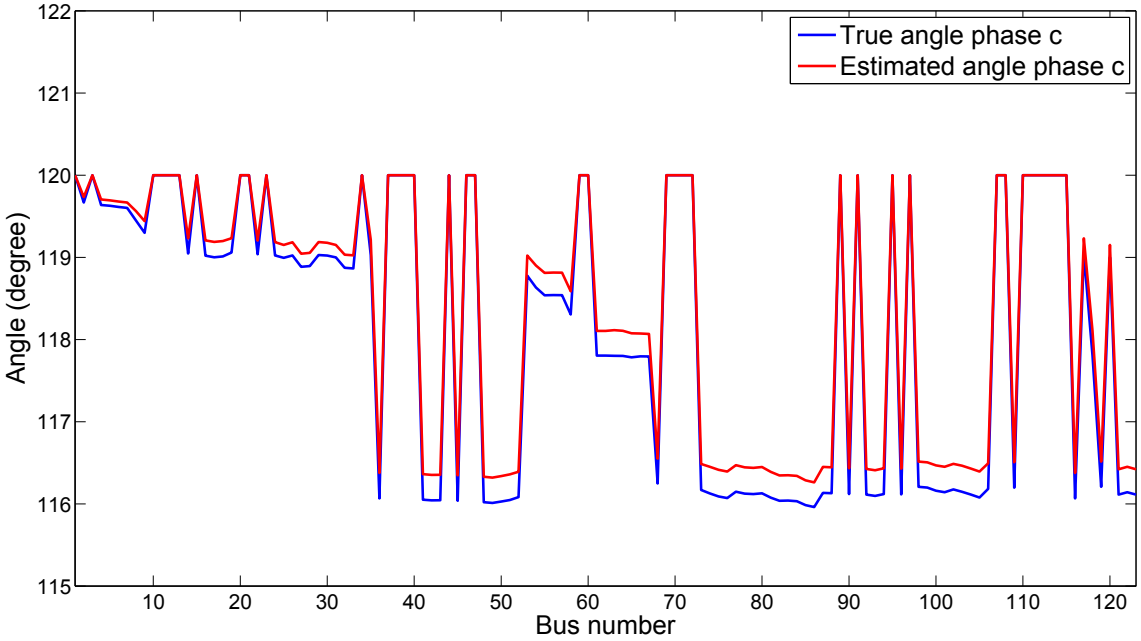

Figure 3.16: True and estimated voltage angles for IEEE 123 bus system with changed switch status

errors are independent and identically distributed. The zero injections are considered as equality constraints. The voltage magnitudes are set to operate within 5% of the nominal values and the voltage angles within $-30°$ to $+30°$. The error in real measurement is assumed to be 3% and the error in pseudo measurement is assumed to be 40%. Figure 3.5, 3.6, 3.7 show the true and estimated voltage magnitudes of the three phases for the IEEE-123 bus model for case I. Whereas, Figure 3.11, 3.12, 3.13 show the true and estimated voltage magnitudes of the three phases when the switch status are changed. Figure 3.8, 3.9, 3.10 and Figure 3.14, 3.15, 3.16 further show the true and estimated voltage angles of the three phases for case I and the case when the switch status is changed respectively. The obtained results have been found to be satisfactory within the allowable tolerance ($\pm 3\sigma$) from equation (3.12).

3.6 Conclusions

This chapter presents a WLS three-phase state estimation based on detailed modelling of the different components of three phase system considering both the star and delta-configured loads. The method achieved a reliable solution to the state estimation problem. Simulation results on IEEE 13-bus and IEEE 123-bus distribution system showed the effectiveness of the approach and the SE results have been compared with the load flow results. The cases with different switch status have also been implemented on the 123-bus system. The reliable state estimation results provides the basis for control and monitoring of modern distribution systems.

However, the state estimates are affected by the accuracy of the measurements-real and pseudo. Thus, a large measurement error can result in wrong distribution operation and control decisions. The accuracy of the measurement meters are compromised due to various reasons. This is discussed in the next chapter. The measurements can also be compromised by an attacker for his own benefits. The attacker disguises the attack in such a way that the traditional bad data detection techniques are unable to identify them. This poses a serious challenge to the EMS/DMS. The next chapters discuss the vulnerabilities and a detection technique based on diagnostic robust generalized potential and studentized residuals.

Chapter 4

Vulnerabilities associated with State Estimation

One of the essential benefits of using a state estimator is to detect, identify and correct measurement errors. This is known as bad data detection. Depending on the state estimation procedure, bad data processing can be carried out as part of the estimation process or as a post-estimation process. However, irrespective of the process, detection of bad data can only be done if there are enough redundant measurements in the system. Redundant measurements can be removed from the measurement matrix without making the system unobservable. With a given set of measurements the system is said to be observable if a unique estimate of the states can be found. The network observability is ensured by graph theory method prior to the estimation process. So, when there is an error in a redundant measurement, this can be detected by statistical tests based on measurement residuals [55]. However, removal of a critical measurement will lead the system to an unobservable system.

Measurements may contain errors due to various reasons. Random errors usually exist in measurements due to the finite accuracy of the meters and the telecommunication medium. Large measurement errors can also occur when the meters have biases, drifts or wrong connections. Telecommunication system failures or noise caused by unexpected interference also lead to large deviations in recorded measurements. Apart from these the state estimator may be affected by incorrect topology infor-

mation which can be subsequently interpreted as bad data. With the integration of PMUs, smart metering and communication infrastructure into the system, the modern power system is gradually becoming more and more cyber-physical rather than only physical system. As more and more advanced communication and cyber technologies are getting incorporated, the possibility of an adversary to tamper with the meter data to cause the state estimator to produce wrong estimates is also increasing [24, 25].

Over the years, state estimation has been developed to deal with gross error in data because of inaccuracy of the measurements. Any tampering with data and/or maliciously operating switch will also result in gross errors in data. If an adversary gains control of the switches/circuit breakers he/she can change the topology of the system completely. This would result in faulty measurement data. So in principle, the effect of malicious attack can be detected through bad data detection. Depending on the state estimation methodology bad data detection can be part of state estimation process or a post estimation computation as shown in Figure 4.1. As long as these errors are part of over measured systems (more measurements than the number of states to be estimated) and do not belong to the critical measurement and leverage points (measurements that significantly influence the state estimation solution), eliminating them to get a clear and accurate estimate is not difficult. However, if these bad data belong to the meters in the leverage set they need to be handled carefully. The leverage measurements help in improving the state variable estimates of the system by providing sufficient redundancy. The critical measurements are those whose removal affects the system observability. References [56] and [57] have discussed about protection of some or all of basic measurements or critical k-tuples in the system. A critical k-tuple is a set of measurements for which, if all the measurements of the set are lost, then the network becomes unobservable. However, the leverage measurements are not protected. The leverages can occur both in transmission and distribution networks [58]. The leverage measurements are explained in Section 4.1. This requires to develop a methodology to deal with the situation.

A schematic similar to Figure 1.1 in Chapter 1 of a typical energy/distribution

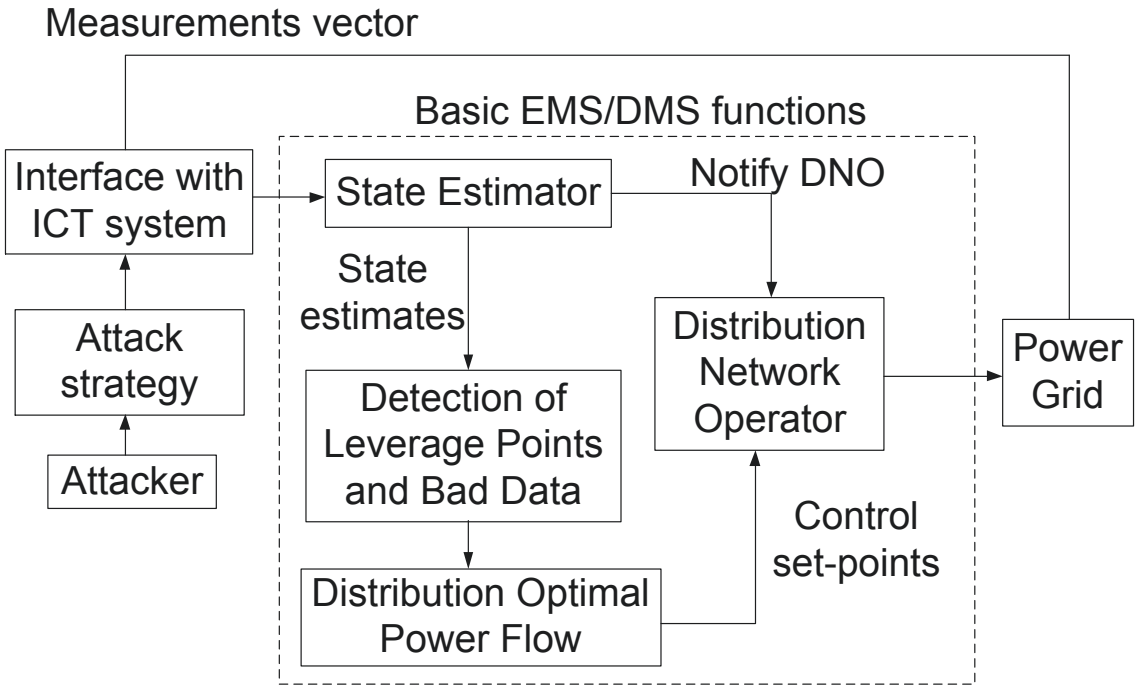

Figure 4.1: A typical energy/distribution management system architecture

management system with ICT infrastructure is shown in Figure 4.1.

Traditionally, the detection of bad data has been carried out by largest normalized residuals (LNR), which performs pretty well when there is a single bad measurement or multiple non-interacting bad measurements in the system [59]. However, it fails in case of influential or leverage measurements [20].

References [26] and [27] have proposed a χ^2 test for the identification of bad measurements. In the χ^2 test if the WLS state estimation objective function value is more than a predefined threshold then the presence of bad data is suspected. Otherwise, there is no bad data in the measurement set.

Chen and Abur have proposed the method of placement of PMUs to enable bad data detection in state estimation [35]. The placement of PMUs at strategic locations will eliminate measurement criticality and thus help in bad data detection.

References [36] and [60] have devised the concepts of robust distances and influence functions in regression analysis of measurement equations to identify leverage points in a system.

But none of these references discussed the issue of detecting bad data in leverage data points. Therefore, security associated with the state estimator has become a

matter of particular concern and hence, robust methods have to be resorted to in detection, identification and elimination of bad data from state estimation.

4.1 Leverage Points and Bad Data

Leverage points

The state estimation problem as discussed in Chapter 3 is linearized around an operating point and is expressed as the following regression model.

$$\Delta z = H\Delta x + \mathbf{e} \tag{4.1}$$

where z is considered the output of the regression model and x is the regressor vector, predictor or the factor in the regression model and $\mathbf{e}$ is the random error vector, which are random and assumed to be independently and identically distributed (*i.i.d.*), in the regression model. The matrix H is known as the coefficient or regressor matrix. The detection, assessment and understanding of influential points are the main areas of study in the regression model building. The factor variables or the explanatory variables in the regression model are solved by least squares estimation as in equations (3.17) and (3.18). From equation (3.11), the estimated measurement vector is derived as

$$\Delta\hat{z} = H(H^T R^{-1} H)^{-1} H^T R^{-1} \Delta z \tag{4.2}$$

Or,

$$R^{-1/2}\Delta\hat{z} = R^{-1/2}H((R^{-1/2}H)^T R^{-1/2}H)^{-1}(R^{-1/2}H)^T R^{-1/2}\Delta z \tag{4.3}$$

$$\Delta\hat{\tilde{z}} = \tilde{H}(\tilde{H}^T\tilde{H})^{-1}\tilde{H}^T\Delta\tilde{z} \tag{4.4}$$

Where,

$$\Delta\hat{\tilde{z}} = R^{-1/2}\Delta\hat{z},\ \tilde{H} = R^{-1/2}H,\ \Delta\tilde{z} = R^{-1/2}\Delta z$$

Therefore,

$$K = \tilde{H}(\tilde{H}^T\tilde{H})^{-1}\tilde{H}^T \tag{4.5}$$

where, K is called the hat or high-leverage matrix.

The matrix K shows some remarkable properties. It is symmetric ($K = K^T$) and idempotent ($K.K.....K = K$). Hence, the diagonal element K_{ii} can be expressed as

$$K_{ii} = K_{ii}^2 + \sum_{i \neq j} K_{ij}^2 \tag{4.6}$$

where, K_{ij} is the non-diagonal element. Therefore, the value of the diagonal element lies between 0 and 1.

A large diagonal entry of the hat matrix implies that the particular measurement has more leverage or influence on the estimated states than others and they are referred to as *leverage points*. If the influence is high enough the corresponding diagonal entry may be close to 1. In other words, according to equation (4.1), each observation (Δz_i,H_i) is a point in the factor space of regression, where H_i is a row of the H matrix. When there is an outlier in the X-space or H_i-space or the regressor variable space, it is said to have an undue influence on the state estimates and is called a leverage measurement.

Bad data

The concept of bad data and outliers go hand in hand in the context of regression jargon. *Bad data* usually refers to an erroneous measurement due to various reasons. Due to the integration of PMUs, intelligent and smart metering with ICT infrastructure, the modern power network is a cyber-physical system rather than a physical system. It uses the telecommunication medium for data transfer. Bad data or gross errors can occur during the data transfer over the SCADA telemetry system. Telecommunication system failures or noise caused by unexpected interference also lead to large deviations in recorded measurements. So, these bad data or gross errors can be looked as outliers in the measurement space. However, a measurement, which may or may not contain errors, such as leverage points, may also appear as outliers due to the structure of the corresponding regression equation. As a result, it is essential to differentiate the leverage points from bad data and identify the error, if any, in leverage points. In the modern power system, more and more advanced

communication and cyber technologies are getting incorporated [20]. Therefore, the possibility of an adversary to tamper with the measurements to drive the state estimator to wrong estimates is also high. Theoretically, the bad data detection (BDD) technique using normalized residual is a post-estimation process. Essentially, the largest normalized residual (LNR) method is used to detect, identify and eliminate bad measurement data. The largest normalized residual refers to the test where the largest normalized residual corresponds to the bad measurement data. Normalized residual based approaches for identification of bad data have been reported in [27] and [26]. In the case of one erroneous measurement data, the largest normalized residual works perfectly fine. It has been reported in the literature [26, 27] that LNR also works on both non-interacting and interacting non-conforming multiple bad measurement data. However, it fails to detect the bad data if there are multiple interacting and conforming bad data [59],where the errors are in agreement, and if they are part of the leverage set. Moreover, the residuals are given as

$$r = \Delta z - \Delta \hat{z} \tag{4.7}$$

Eq.(4.7) can be rewritten as

$$r = (\mathbf{I} - K)\Delta z \tag{4.8}$$

Therefore, the measurement residuals with large diagonal entries of the hat matrix are small even if it is contaminated with gross error.

4.2 Attack Strategies

In power systems, the state estimator as mentioned in Figure 4.1 takes three kinds of inputs-the meter measurement data (power injection and power flow), the network topology information data (on/off status of switches) and the parameter data (branch impedance and variances of measurement errors). Typically, these inputs are either sent from meters to control center or stored in the databases. It is assumed that the adversary can access and manipulate all the three kinds of inputs.

The leverage measurements occur when there are injection measurements on a bus, which has a larger number of branches connected to it compared to others,

injection measurements on a bus incident to branches with very different impedances, and the line power flow measurements on relatively short lines. In large meshed distribution systems these leverages can occur due to the presence of line power flow measurements on short lines [58] and also due to the lower redundancy of measurements. An adversary can take advantage of this situation and attack the high leverage points to influence the estimates of the state variables of the system and hence, can hide the attack from being detected. The leverage points affected by gross errors are called bad leverage points. Though bad leverage points are harmful to many estimators, good leverage points are particularly useful in improving the variance of the estimates.

4.2.1 Attacking power flow measurements

Power flow measurements are normally placed between buses to monitor the flow of the branches. Leverage power flow measurements are formed when the measurements are placed on relatively short or long lines. An attacker, if he/she intends to make the attack invisible, makes changes to the value of impedance of the branch by applying Theorem 2 and rule 1 and rule 2 as given in [61].

4.2.2 Attacking power injection measurements

Power injection measurements are placed at a bus to monitor the active and reactive power injections from a load or generations at a particular bus. A node/bus is particularly vulnerable to leverage attack if that has more connecting branches connected to it or in other words there are more non-zero elements in that row of the H as in (3.30) matrix compared to other rows. If an adversary wishes to attack an injection leverage measurements he/she should increase the particular diagonal element of the hat matrix to make the attack undetectable by applying Theorem 2 and rule 1 and rule 2 as given in [61].

To make a successful attack, the attacker makes changes to the impedance of a branch by applying Theorem 2 and rule 1 and rule 2. The theorems 1,2 and 3 are stated in Appendix D. Theorem 1 states how a successful attack can be made

on a measurement by changing K_{ii}. While Theorem 2 shows how much K_{ii} has to be increased to make an attack on a single measurement z_i. To make an attack on multiple measurements, the attacker will perturb the measurements one at a time and apply Theorem 2 repeatedly. Finally, Theorem 3 suggests how to increase the value of K_{ii}.

4.3 Masking and Swamping

The leverage points in regression studies carry with them an inherent difficulty. When there are more than one influential point, some of them may remain undetected. This phenomenon is known as masking. On the other hand, some of the non-influential points may be wrongly detected as influential points, which is known as swamping. The masking/swamping can be explained by the following equation.

The residual for the i^{th} measurement with two high leverages at z_i and z_k is expressed as

$$r_i = (1 - K_{ii})\Delta z_i - K_{ik}\Delta z_k - \sum_{\substack{j=1 \\ j\neq i, j\neq k}}^{m} K_{ij}\Delta z_j \tag{4.9}$$

So, if the first two terms in equation (4.9) are opposite in sign a bad leverage may appear like a good leverage. This is known as masking. On the other hand, if the second and the third terms in the same equation add up to a large value the good leverage may become a bad leverage. This is known as swamping.

The masking and swamping phenomena have been reported in the literature as in Hawkins, Bradu and Kass (HBK) data, Brownlee's stack loss data [62, 63], Hadi and Simonoff (HS) data, Belgian Telephone data etc. In the HBK data there are 75 observations, 14 high leverage points and 10 outliers with points 11-14 are swamped cases. The Brownlee's data shows, however, that there are 21 observations with 4 outliers (cases 1,3,4,21) and 4 high leverage points (cases 1,2,3,21). There are two points which are masked and on the other hand point 17 is swamped. The masking/swamping phenomenon can influence the final outcome of the detection procedure and result in faulty detection of bad data or leverage point. This swamping or masking phenomenon is, however, not present when there is only one influential

measurement. This case is similar to the largest normalized residual (LNR) test to identify outliers. Hence, when there are multiple outliers or bad data or influential points the largest normalized residual test is deemed unsuitable. To the best of my knowledge, the masking/swamping phenomenon has not been investigated in the context of power system state estimation. This sets up the motivation to devise a method by which the high leverage points, low leverage points and outliers or bad data are completely separated and identified.

The next sections provide a review of the different diagnostic techniques related to identification of gross error and discusses the suitability of externally studentized residuals for this purpose.

4.4 Diagnostics based on Residual Analysis

Multiple linear regression model building is one of the standard problems in chemometrics [64]. The linearised measurement equations for state estimation can be interpreted as regression modelling. Each measurement can be considered as a point in the $(n+1)$-dimensional regression plane. The term regression diagnostics has been used for a collection of methods for the identification of influential points, and for the identification of violations of the assumptions of least-squares. The residuals are defined as

$$r = \Delta z - \Delta \hat{z} \tag{4.10}$$

The residuals give a measure of the presence of bad data in the system. Thus, analysis of residuals provide a diagnostic for bad data detection. There are different kinds of residuals based on mathematical definition. These are highlighted in the following subsections.

4.4.1 Normalized residuals

Researchers have also addressed the issue by χ^2-test as in [27], by Hypothesis Testing Identification (HTI) as mentioned in [27] and some researchers have also used normalized residual for the detection of bad data [27].

In lines with the WLS state estimation, the errors are assumed to be normally distributed as $e_i \backsim N(0, R_{ii})$ and hence, the residuals are also normally distributed as $r \backsim N(0, \Omega)$ where, $\Omega = SR$, where $S = I - K$. Therefore, the normalized residuals of the measurements can be determined as

$$r_i^N = \frac{|r_i|}{\sqrt{\Omega_{ii}}} = \frac{|r_i|}{\sqrt{R_{ii}S_{ii}}} \tag{4.11}$$

The normalized residuals will then have Standard Normal Distribution $r_i^N \backsim N(0, 1)$.

4.4.2 χ^2-test for bad data detection

The measurement residual is a normally distributed variable and the residuals are independent. Therefore

$$J = \sum_{i=1}^{3m} R_{ii}^{-1}(z_i - h_i(x))^2 \tag{4.12}$$

$$= \sum_{i=1}^{3m} R_{ii}^{-1} r_i^2 \tag{4.13}$$

According to statistical properties, the objective function, being the weighted sum of the residuals will follow a χ^2 distribution with $(3m - n)$ degrees of freedom. If the estimated value of this objective function $J(\hat{x})$ is more than a threshold for a certain detection confidence probability of 97.5% then the measurements are said to have bad data. Otherwise, the measurements are error free. This is known as the χ^2-test for identification of bad data.

4.4.3 Largest Normalized Residual

The largest normalized residual (LNR) will correspond to the erroneous data if there is only one bad data in the measurement system. Hence, if the largest normalized residual is more than a threshold (say 3) then the system is said to have erroneous measurement. Otherwise the measurements are considered to be free from error.

The LNR test is able to detect the bad data if there is only one bad data in the system [26, 27]. It even works on both non-interacting and interacting non-

conforming multiple bad measurement data. However, it fails when there are multiple interacting conforming bad data, where the errors are in agreement [59].

4.4.4 DFFITS

In statistical regression, some of the data points exert more influence on the regression characteristics than the other points. Data point with large residuals and high leverages can distort the accuracy and shape of the regression. Researchers have addressed this problem by applying many diagnostic measures. DFFITS is a diagnostic measure to identify influential points. It is defined as the change in the estimated value of the measurement obtained when the influential data point is left out. It is further standardized by dividing by the estimated standard deviation of the fit at that point. It is defined by

$$DFFITS_i = \frac{\hat{z}_i - \hat{z}_i^{-i}}{\hat{\sigma}_{-i}\sqrt{K_{ii}}} \tag{4.14}$$

$$DFFITS_i = \frac{h_i^T(\hat{x} - \hat{x}^{-i})}{\hat{\sigma}_{-i}\sqrt{K_{ii}}} \tag{4.15}$$

where,

K_{ii} is the leverage value of the point

$\hat{\sigma}_{-i}$ is the estimated standard deviation without the influential point

$\hat{x}^{-i}$ are the fitted state variables without the point in question

$\hat{z}^{-i}$ are the estimated measurement values without the point in question

By definition, DFFITS is the influence of an observation on its own fitted value. It has also been referred to as Welsch and Kuh's distance in some literatures [65, 66].

4.4.5 DFBETA

DFBETA is the diagnostic measure which is defined by the change in the estimated value of the state variable or regression coefficient obtained when the influential data point in question is left out.

The mathematical expression for DFBETA has the general form

$$DFBETA_i = \hat{x}_i - \hat{x}_i^{-i} = \frac{(H^T H)^{-1} h_i^T \hat{e}_i}{1 - K_{ii}} \tag{4.16}$$

It is measure of the influence of an observation on a particular regression coefficient.

4.4.6 Cook's distance

Cook's distance is a commonly used measure for the influence of a data point named after the American statistician R. Dennis Cook [67]. Cook's distance is defined as the influence of an observation on all fitted values. It can be expressed as

$$CD_i = \frac{(\hat{x}^{-i} - \hat{x})^T (H^T H)(\hat{x}^{-i} - \hat{x})}{p\hat{\sigma}^2} \tag{4.17}$$

$$CD_i = \frac{\hat{\sigma}_{-i}}{p\hat{\sigma}^2} DFFITS_i \tag{4.18}$$

where,

p is the number of fitted parameters.

$\hat{x}$ are the fitted state variables.

$\hat{x}^{-i}$ are the fitted state variables without the point in question.

H is the Jacobian matrix as mentioned in the last chapter.

4.4.7 Studentized residuals

The diagnostics for single case influential observations are ineffective in case of multiple influential observations due to masking/swamping effects. Masking is said to occur if a bad data point, in the presence of other bad data points, appears as a good data point. Similarly swamping is said to occur if a good data point, in the presence of other bad data points, behaves as a bad data point. The phenomenon of masking and/or swamping has been explained in Section 4.3. Let the set of deleted cases be D and the set of remaining cases be R. When a group of observations is deleted

$$K_{ii}^{-(D)} = h_i^T (H_R^T H_R)^{-1} h_i$$

$K_{ii}^{-(D)}$ is the i^{th} diagonal element of the $H(H_R^T H_R)^{-1}H^T$ matrix. Most of the outlier detection methods separate the clean observations from the potential outliers.

When an additional point i is added to the set R, according to [66, 68]

$$K_{ii}^{-(D)+i} = h_i^T(H_R^T H_R + h_i h_i^T)^{-1}h_i = \frac{K_{ii}^{-(D)}}{1 + K_{ii}^{-(D)}}$$

The new state variables with the additional point i in the set R is given by

$$\begin{aligned}\Delta\hat{x}_{R+i} &= (H_R^T H_R + h_i h_i^T)^{-1}(H_R^T \Delta z_R + h_i \Delta z_i) \\ &= \Delta\hat{x}_R + \frac{(H_R^T H_R)^{-1}h_i}{1 + K_{ii}^{-(D)}} r_{st,i}^*\end{aligned}$$

Let $r_i^{-(D)}$ be the i^{th} deletion residual.

$$r_{st,i}^{*,R+i} = \frac{r_i^{-(D)}}{\hat{\sigma_R}\sqrt{1 + K_{ii}^{-(D)}}}$$

The variances of the observations in the basic subset and outside the basic subset are given [69] as:

$$1 - h_i^T(H_R^T H_R)^{-1}h_i,\ i \in R$$
$$1 + h_i^T(H_R^T H_R)^{-1}h_i,\ i \notin R$$

The studentized residuals for the two subsets are given as

$$\frac{r_i^{-(D)}}{\hat{\sigma_R}\sqrt{1 - h_i^T(H_R^T H_R)^{-1}h_i}},\ i \in R$$
$$\frac{r_i^{-(D)}}{\hat{\sigma_R}\sqrt{1 + h_i^T(H_R^T H_R)^{-1}h_i}},\ i \notin R$$

Internally and externally studentized residuals

The externally studentized residual has a clear advantage over standardized residuals. The standardized residual includes the i^{th} observation, which could be an outlier, which influences the least square function. But an externally studentized residual removes the i^{th} observation while calculating the variance estimate. Mathematically the internally studentized residual is given as

$$r_{st,i} = \frac{r_i}{\hat{\sigma}\sqrt{1 - K_{ii}}} \tag{4.19}$$

For a standardized residual, it is standardized by dividing by $\hat{\sigma}$, the ordinary standard deviation estimate. But in case of externally studentized residual, it is divided by a factor $\hat{\sigma}_{(i)}$, which is the standard deviation estimate in an estimation model with the i^{th} data deleted. In other words, while the numerator and denominator are not independent in case of standardized residuals, they are independent in case of externally studentized residuals in the expression for residuals. Thus, in general, externally studentized residuals will be more effective in detecting outliers.

4.5 Robust estimators

The WLS state estimator assumes that the measurement errors are independent and identically normally distributed. However, due to the presence of bad data the performance of WLS estimator reduces considerably. The robust estimation techniques are used to address these issues. The concept of robust estimation was first introduced by Huber [70]. Mili et al. [71] were the first to apply in power system. The class of robust estimators for power system are called M-estimators. The objective for a generalised M-estimator is given by

$$min\ J = \sum_{i=1}^{m} \phi(r_i) \tag{4.20}$$

where, m is the number of measurements. Depending on the function ϕ, different types of estimators have been explored.

4.5.1 Quadratic Constant estimator

The quadratic constant (QC) estimator behaves like WLS inside the threshold and takes a constant value outside the threshold.

$$\phi(r_i) = \begin{cases} r_i^2\ if\ \ |r_i| \leq c_{tune} \\ c_{tune}^2\ otherwise \end{cases} \tag{4.21}$$

Although the QC estimator has better bad data rejection properties, the objective function is non-convex, which gives rise to convergence and computational difficulties.

4.5.2 Square Root estimator

The objective function of a square root (SR) estimator is given by

$$\phi(r_i) = \begin{cases} r_i^2 \ if \ \ |r_i| \leq c_{tune} \\ 4c_{tune}^{3/2}\sqrt{r_i} - 3c_{tune}^2 \ otherwise \end{cases} \tag{4.22}$$

The SR estimator with no bad data reduces to WLS. But, the ϕ function is proportional to the square root of the residual if there is a bad data. Its bad data rejection property is between SHGM and QC estimators. Therefore, it is computationally similar to WLS technique.

4.5.3 Schweppe-Huber Generalized M-estimator

The Schweppe-Huber Generalized M (SHGM) estimator combines both WLAV and WLS estimators. The ϕ function is given by

$$\phi(r_i) = \begin{cases} \frac{1}{2}r_i^2 \ if \ \ |r_i| \leq c_{tune}w_{factor,i} \\ c_{tune}w_{factor,i}\,|r_i| - \frac{1}{2}c_{tune}^2 w_{factor,i}^2 \ otherwise \end{cases} \tag{4.23}$$

It behaves like the WLS estimator for small values of residuals. However, it behaves like the WLAV estimator outside the threshold. The performance of this estimator depends on the tuning parameter c_{tune} and weighting factor $w_{factor,i}$. The value of the tuning factor c_{tune} usually is between 1 and 4.

4.5.4 Least Absolute Value estimator

The weighted least absolute value (WLAV) estimator is based on minimizing the sum of the absolute values of the weighted residuals. It is less sensitive to the presence of bad data. The WLAV estimator can simultaneously detect and reject bad data. It is expressed as

$$\phi(r_i) = |r_i| \tag{4.24}$$

However, the complexity and computational time makes it impractical to use for real time state estimation applications [72].

Robust estimators are extremely handy in terms of bad data rejection. However, it carries with it some inherent downsides. These robust estimators may sometimes get stuck in the local minima. Over and above that, these estimators may suffer from slow convergence or sometimes divergence. In case of poor redundancy, which is quite possible in distribution systems, there is a possibility of numerically unobservable solution. As a result, there is a high risk of wrong identification of bad data.

This gives a motivation to apply WLS estimator in this research. WLS works well on the assumption that the errors are normally distributed. Moreover, the WLS is computationally fast and results in the global minimum solution.

4.6 Conclusions

This chapter presents the vulnerabilities of the modern power system. It emphasises particularly on the specific vulnerability in relation to the leverage measurement data points. The attack strategies associated with power injection and power flow measurements and the masking/swamping phenomenon have been investigated. The different types of residual diagnostic techniques and robust estimation methods have been explored. The externally studentized residuals have been found particularly useful for multiple influential points from the discussions. They are more effective in detecting the outliers. If the measurement errors are normally distributed the weighted least squares estimation works well due to its fast computational capability and ability to reach global solution. On the lines of vulnerabilities discussed in this chapter, the next chapter proposes a detection technique for identification of gross error.

Chapter 5

Bad Data Identification against Leverage Point Attacks

This chapter presents the concept of robust generalized potentials and proposes a technique to diagnose bad data and leverage measurements simultaneously from the rest of the regression data. The power system state estimation measurement equations in this context are regarded as linearized regression equations of state variables at each operating point.

In other technology areas, the gross measurement errors have been treated as outliers in factor space of linear regression analysis [64, 73].

However, sometimes some influential measurements called leverage points may resemble outliers in factor space as they lie outside the regression line [27,74]. Therefore, it is hard to identify errors in leverage points. As a result, it is necessary to distinguish the leverage points from the outliers.

It has been reported that there are a number of ways one can identify the leverage points from the diagonal elements of the hat matrix: Mahalanobis distance (MD) of measurements, projection statistics (PS) [26, 27] etc.

Reference [60] has devised the concept of influence function as a combination of influence of residuals and influence of position in factor space. Thus, looking at the influence function one can identify bad data even for influential measurements.

The concept of finding outliers in multivariate data has been suggested in [73]

and the concept of identifying outliers based on different residual diagnostics in linear regression models of a system has been reported in other applications [64]. However, they have not addressed the masking and swamping effect of leverage points when there are multiple leverage points.

In the field of applied statistics, there has been research on identifying the multiple high leverage points in multivariate analysis. It has been pointed out that due to the presence of more than one high leverage point, the leverage structure may change in such a way that the leverage diagnostics for single leverage point like twice-the-mean rule, thrice-the-mean rule, Cook's distance, Welsch and Kuh's distance etc. will not be able to identify the real high leverage points.

Nurunnabi, Hadi and Imon in [63] have used a modified Cook's distance and Habshah, Norazan and Imon [62] have proposed a robust diagnostic potential to address this issue. Reference [62] have applied the technique on Hawkins, Bradu and Kass data and Brownlee's stack loss data to illustrate the simultaneous identification of outliers or erroneous data and high leverage points.

To the best of my knowledge, this methodology has never been applied in power system bad data detection context. This chapter presents a robust bad data detection technique when the leverage measurements are compromised and shows that it can take care of masking/swamping phenomenon in sparse systems like power systems that existing methods cannot. This methodology is applied, for the first time, to robustly detect bad data in regards to state estimation of power system.

The primary motivation of this chapter is driven by such possible scenarios when the hacking of the data in meters is related to the measurements of the leverage points. The next sections of this chapter propose the novel technique of identifying gross error in those cases and has been tested on standard IEEE test networks.

5.1 Detection of Leverage and Bad data points

Leverage values are normally denoted as measures of influential observations in the X-space. The X-space is the space of regressor variables. The hat matrix in (4.2) gives a measure of the influence of a particular measurement. The ones which

have higher influence are called high leverages and ones which have lower influence are called low leverages. The twice-the-mean rule and thrice-the-mean rule on the diagonal elements of the hat matrix have been reported in the literature to identify the leverage points. Reference [63] has mentioned the Cook's distance and Welsch and Kuh's distance to detect and identify the single leverage point. The Mahalanobis distance based on the projection pursuit algorithm for minimum volume ellipsoid cannot be applied to sparse systems. Since the electric power system is a sparse system, the projection pursuit algorithm has to be modified in order to be applied to the sparse power system. However, due to masking or swamping effect it becomes difficult to identify the group of high leverage points.

5.2 Diagnostic robust generalized potentials

This technique, an adaptive approach to identify the group of leverage points, is a unified approach of diagnostic and robust approaches. The robust approach identifies the suspected high leverage points and the diagnostic approach confirms the above suspicion. The robust approach identifies the leverage points by the corresponding potentials of the data. The potential of a data is defined by Hadi [75] as the diagonal element of the hat matrix with the i^{th} data deleted. It is denoted by

$$pot_{ii} = h_i^T (H_{(i)}^T H_{(i)})^{-1} h_i \tag{5.1}$$

The points having a potential value more than the robust cut-off $Median(pot_{ii}) + c.MAD(pot_{ii})$ is said to be a high leverage point, where, MAD is the median absolute deviation from the median and c is a constant equal to 2 or 3. However, this method is not robust against swamping. Habshah et al [62] have proposed a robust method to identify high leverage points. The robust Mahalanobis distance (RMD_i) is defined as

$$RMD_i = \sqrt{[h_i - H_c]^T [C(H)]^{-1} [h_i - H_c]} \tag{5.2}$$

where, H_c is the mean of the l points for which determinant of the covariance matrix (MCD) is minimum or H_c is the centre of the minimum volume ellipsoid (MVE) covering these points, and $C(H)$ is the corresponding covariance matrix.

The cut-off value for a normal distributed multivariate data is $\sqrt{\chi^2_{n,\alpha}}$, but, for general non-normal data the cut-off value as suggested in [62, 66] is given by

$$Median(RMD_i) + 3MAD(RMD_i) \tag{5.3}$$

The observations are grouped in two sets. Those which have robust Mahalanobis distance greater than the cut-off as in Eq.(5.3) are considered to be in set D and the rest in set R. The robust potentials for the observations in two sets are given as

$$pot^*_{ii} = \begin{cases} \frac{K^{-(D)}_{ii}}{1-K^{-(D)}_{ii}} \ \forall \ i \in R \\ K^{-(D)}_{ii} \ \forall \ i \in D \end{cases} \tag{5.4}$$

$K^{-(D)}_{ii}$ denotes the i^{th} diagonal element of the hat matrix with data as in set D deleted. There exists no theoretical distribution for pot^*_{ii} and hence, there is no finite upper bound. However, [62, 66] suggested a suitable confidence bound type cut-off like

$$Median(pot^*_{ii}) + c.MAD(pot^*_{ii}) \tag{5.5}$$

The Mahalanobis distances of the multivariate data are first calculated. The Mahalanobis distance, however, is prone to the masking effect of multiple leverage data points [31]. Fig. 5.1 shows the step-by-step procedure for the identification of leverage points.

1. The robust Mahalanobis distances of the observations of the multi-variate data are carried out based on minimum volume ellipsoid (MVE) or minimum covariance determinant (MCD). Conceptually, MVE is the ellipsoid with minimum volume that contains l data points. MCD is, however, the minimum of the determinant of the covariance matrix which contains l points. l is typically equal to $[3m/2] + 1$ (where $3m$ is the number of data points). MVE has been considered here.

2. The multi-variate data are grouped into two separate subsets R and D. The observations which have a distance higher than the cut-off as in (5.3) are deleted from the main set and kept in a separate set called the deleted set D. The rest of the data are kept as it is in a set called R.

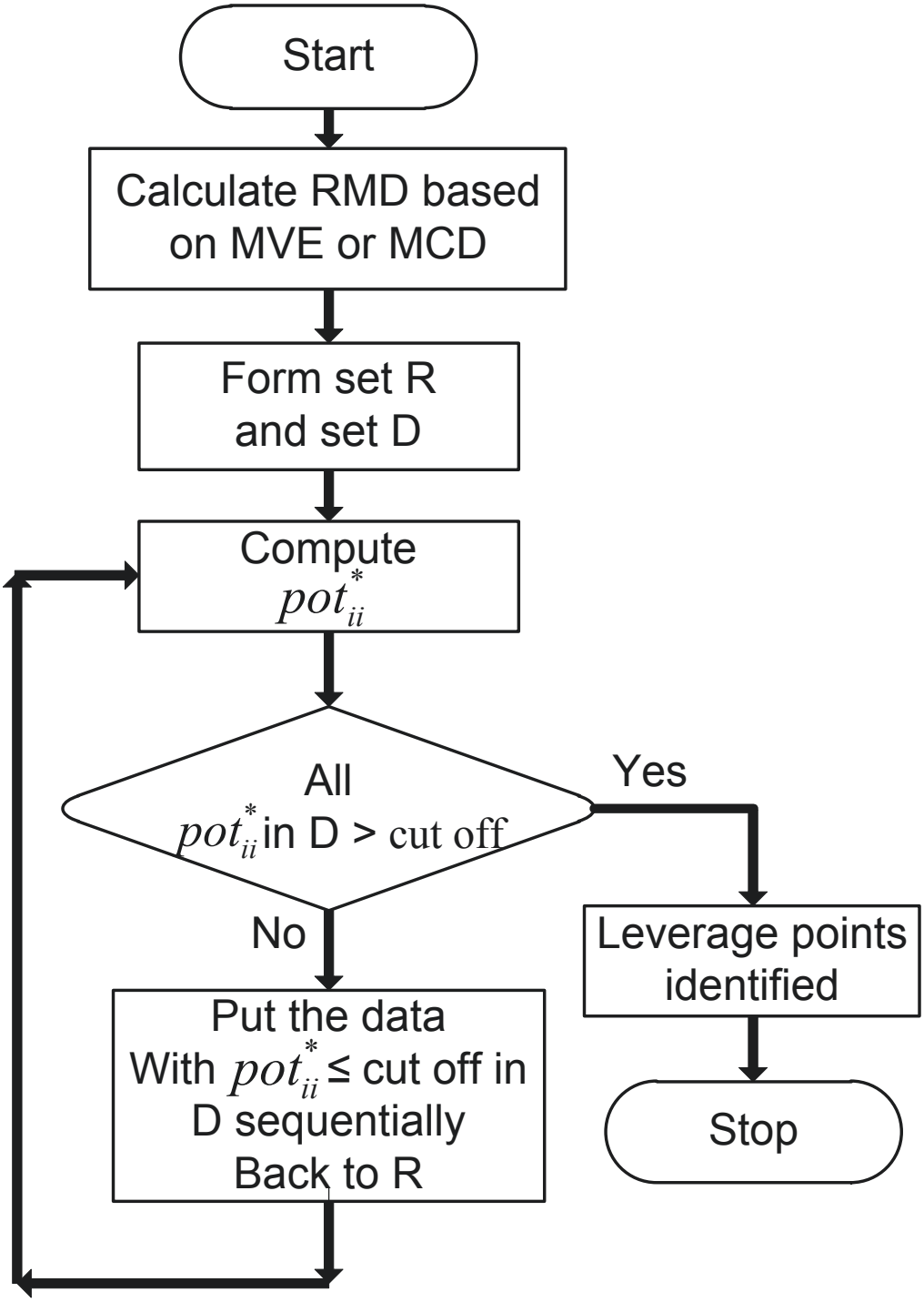

Figure 5.1: A flowchart showing the identification of leverage points

3. The generalized robust potentials for both the sets are computed.

4. If all the observations in the deleted set D have their generalized potentials higher than the cut-off, then the leverage points are identified. If not, data are put back to set R sequentially starting with the one which has the least

generalized robust potential value.

5. The generalized potential values are recalculated with the new subsets.

6. This process continues till all the data in the set D have generalized potential values more than the cut-off.

By this process, the masking and swamping effects, if present, are completely taken care of and the high leverage and non-leverages are separated from each other.

5.3 Identification of gross error and high leverage points

The measurements in a generic power system can be easily tampered with for nefarious purposes. The physical meters in the system can be compromised by introducing a large error by intelligent hackers. As discussed in Section 4.1, the residuals as in (4.8) for leverage measurements are close to zero. The important class of M-estimators, including the LAV estimator, cannot handle bad leverage data points. Hence, it is very difficult to identify the gross error in case of leverage measurements. Other estimators like LMS, LTS, RLS, Iterative RLS, BOFOLS etc. are computationally intensive.

The residuals in the measurement data are functionally related to the leverage values of the data. This method is a combination of direct and indirect approach of multiple outlier detection. The low leverages and high leverages are separated first based on DRGP and then generalized studentized residuals (GSR) is calculated for the entire data set to identify the outliers. So, an outlier in set R will not be confused with an outlier in set D. They are defined as

$$r^*_{st,i} = \begin{cases} \dfrac{r_i^{-(D)}}{\hat{\sigma}_{R-i}\sqrt{1-K_{ii}^{-(D)}}} \ \forall\ i \in R \\ \dfrac{r_i^{-(D)}}{\hat{\sigma}_{R}\sqrt{1+K_{ii}^{-(D)}}} \ \forall\ i \in D \end{cases} \tag{5.6}$$

where, $\hat{\sigma}^2$ is the least squares estimate of variance. $r_i^{-(D)}$ represents the residual of i^{th} measurement with D data set deleted. R is the data set without the high leverages.

The GSR is a form of a Student's t-statistic with $(3m - n - 3 - 1)$ degrees of freedom and 97.5% detection confidence probability. One could, therefore, use a t-table to get the exact cut-off values. But since the degrees of freedom are usually quite large, the rule of thumb that absolute value of externally stuentized residuals is greater than 3 is used [76]. The GSR is a type of an externally studentized residual. This is a way of determining the i^{th} residual except the i^{th} observation. If the i^{th} observation is a serious outlier it may influence the least square function and may influence to move it close to the i^{th} observation. So, if it is removed, the i^{th} residual on the new model will indicate that this observation is an extreme value. The mathematical background for the studentized residuals are given in Section 4.4.7 in the previous chapter. All the observations for both the data sets are then plotted in a DRGP-GSR plot. High leverage points are the points which have higher DRGP values and bad data are those data which have higher GSR values. This leverage-residual plot shows that most of the data will be clustered around the origin and the masking/swamping effects do not come into picture. The DRGP-GSR plot clearly separates and identifies the bad measurement data and high leverages. Even if high leverage measurements are adulterated with gross errors the graphical plot clearly identifies the measurement errors. Based on this concept, the next section shows some case studies both for power transmission system and power distribution system and thus justifies the effectiveness of the procedure.

5.4 Case Studies

The problem formulation shown in Section 2 is a three-phase formulation suitable for generic distribution systems. However, the formulation for balanced transmission systems can be taken as a special case of the above formulation, where, the number of state variables and the number of equations as given in Chapter 3 are reduced due to the balanced nature of the system. The voltage magnitudes and angles for a

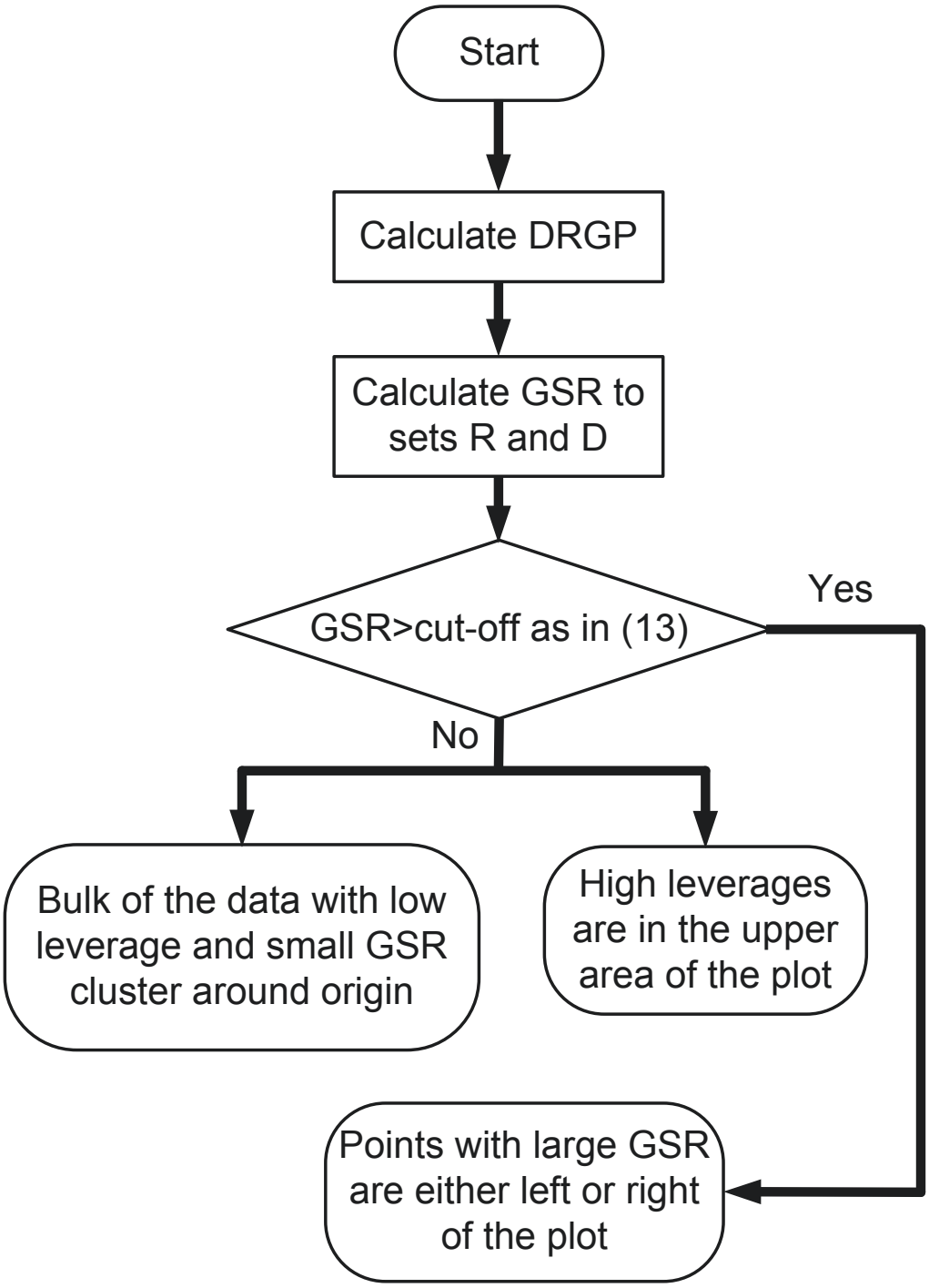

Figure 5.2: A schematic diagram of the DRGP-GSR plot

particular bus will be the same for the three different phases. The proposed approach has been performed on test systems: a small illustrative example, the IEEE 14-bus system and the IEEE 123-bus distribution system. The algorithm was implemented in MATLAB and run on a system with Intel Xeon processor @3.33 GHz and 12 GB RAM.

5.4.1 Illustrative example

Figure 5.3 shows a basic four bus system with possible power injection and branch power flow measurements. All branches are assumed to have a reactance of j0.1 pu. The state variables of the system are considered as voltage magnitude and voltage angles of buses. Since, there are four buses in the system altogether there are eight state variables. However, the voltage angle for bus #1 is taken as the reference. Table 5.1 presents the measurements for the given system. The system, currently, has no leverage points. However, if the line between 1-2 is shortened by decreasing the reactance to j0.01 pu, the measurements flow 1-2 and inj 1 become isolated leverage points. An attacker can introduce a leverage point attack on the system by tampering with the reactance of the line 1-2, should he/she wishes to attack inj 1 and/or flow 1-2. If the line 2-3 is shortened and the injection measurement is on bus 1 instead of 3, the measurements flow 3-2 and inj 3 will become leverage points.

These two measurements become bad leverage points in the factor space. The largest normalized residuals (LNR) method fails to identify these two bad leverage points. It turns out from Table 5.1 that the generalized studentized residuals clearly detects and identifies the bad measurements in case of leverage points. The value of the studentized residual corresponding to the bad measurements with respect to other measurements is much higher compared to that of the normalized residual with respect to other measurements. Table 5.2 further shows the masking/swamping effect of leverage points, if any. It also compares the leverage diagnostics of diagonal elements of the hat matrix with the DRGP technique proposed in Section 5.2. It depicts that while the leverage measure (diagonal element of the hat matrix) fails to identify the leverage points due to masking/swamping effect the DRGP technique can easily identify them. Table 5.3 and Table 5.4 present the results for the active

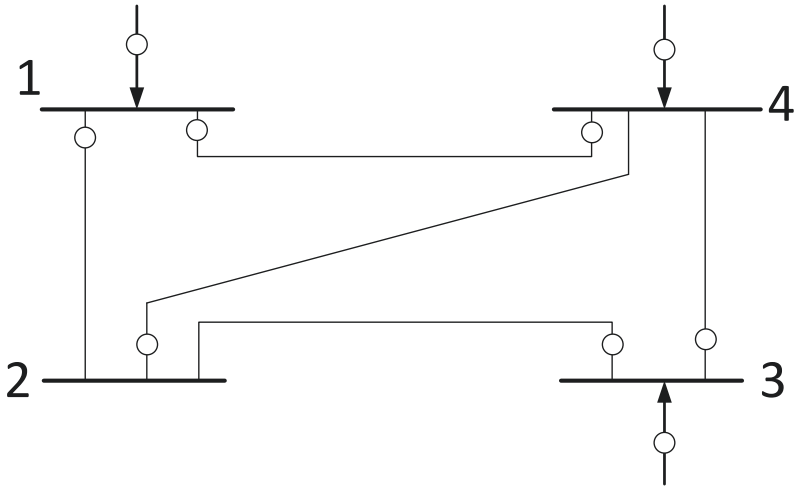

Figure 5.3: A 4-bus system for illustrative example

power injection and reactive power flow measurements when the line 2-3 is shortened.

Table 5.1: Real power measurements and residuals for the 4-bus system when line 1-2 is shortened

Measurement type	Measurement with no bad data	Measurement with bad data	Normalized/ Internally studentized residuals	$\|GSR\|$ (2.228)
flow 1-2	1.50882	**1.00892**	0.550	**2.578**
flow 1-4	0.49119	0.49119	0.4793	0.4328
flow 2-4	0.33966	0.33966	0.2987	0.578
flow 3-2	-0.56915	-0.56915	1.2921	1.374
flow 3-4	-0.23084	-0.23084	0.2373	0.4328
flow 4-1	-0.49119	-0.49119	0.7821	0.8921
inj 1	2.00011	**1.50011**	0.3034	**2.781**
inj 3	-0.800	-0.800	0.5082	0.7811
inj 4	-0.600	-0.600	0.6821	0.852

5.4.2 IEEE 14-bus system

Figure 5.5 shows a typical IEEE 14-bus system. It is a typical meshed transmission network. The network parameters and load data are given in [77] and Appendix C. There are five generation buses in the system. The loads are modelled as a com-

Table 5.2: Leverage points and masking/swamping effect for real power measurements when line 1-2 is shortened

Measurement type	Masking or Swamping effect	Leverage (0.726)	DRGP (0.823)	Bad Data
flow 1-2	No	0.3172	**0.8763**	Yes
flow 1-4	No	0.2988	0.3126	No
flow 2-4	Yes	0.6309	0.4312	No
flow 3-2	No	0.3180	0.2182	No
flow 3-4	No	0.3257	0.5278	No
flow 4-1	No	0.5587	0.6721	No
inj 1	No	0.3272	**0.8450**	Yes
inj 3	No	0.2238	0.2994	No
inj 4	Yes	0.4592	0.2994	No

Table 5.3: Real power measurements and residuals for the 4-bus system when line 2-3 is shortened

Measurement type	Measurement with no bad data	Measurement with bad data	Normalized/ Internally studentized residuals	$\|GSR\|$ (2.228)
flow 1-2	1.50882	1.50882	0.5813	0.1243
flow 1-4	0.49119	0.49119	0.5343	0.7923
flow 2-4	0.33966	0.33966	0.2453	1.265
flow 3-2	-0.56915	**-0.17119**	0.497	**2.567**
flow 3-4	-0.23084	-0.23084	1.2643	1.8809
flow 4-1	-0.49119	-0.49119	0.8702	0.811
inj 1	2.00011	2.00011	0.7982	0.1284
inj 3	-0.800	**-0.400**	0.530	**2.879**
inj 4	-0.600	-0.600	0.7033	0.4252

Table 5.4: Leverage points and masking/swamping effect for real power measurements when line 2-3 is shortened

Measurement type	Masking or Swamping effect	Leverage (0.726)	DRGP (0.823)	Bad Data
flow 1-2	No	0.3810	0.4491	No
flow 1-4	No	0.3279	0.4318	No
flow 2-4	No	0.3692	0.3268	No
flow 3-2	Yes	0.6523	**0.856**	Yes
flow 3-4	No	0.5781	0.7284	No
flow 4-1	No	0.5432	0.3067	No
inj 1	Yes	0.4789	0.3104	No
inj 3	No	0.4872	**0.894**	Yes
inj 4	No	0.5890	0.4321	No

bination of constant impedance (Z), constant current (I) and constant power (P) loads, which is known as the ZIP model. The measured variables are power injection and branch power flows. The measurements are shown in Table 5.7. The measurements are generated by adding random Gaussian noise to the single-phase load flow results. he gross errors are generated by changing the value of the corresponding diagonal element of the hat matrix K_{ii}. The change in the K_{ii} value reflects a change in the corresponding measurement z_i. The details are given in Appendix D.

The sample high and low leverage points are shown by arrow marks in Figure 5.4. In the figure, the line flow measurement flow 5-4 is a high leverage measurement. To make a successful attack, the attacker makes changes to the impedance of the branch 5-4 by applying Theorem 2 and rule 1 and rule 2.

The cut-off values for all the potential values and the studentized residuals are shown in Table 5.7. It shows that DRGP correctly identifies the leverage data points while the potential and the leverage values (i.e. diagonal entries of the hat matrix) fails to identify the leverage measurements correctly and instead swamps some non-leverage measurements as leverage and masks some leverage measurements as non-leverage for 14-bus system. Table 5.6 justifies the fact, with some key measurements

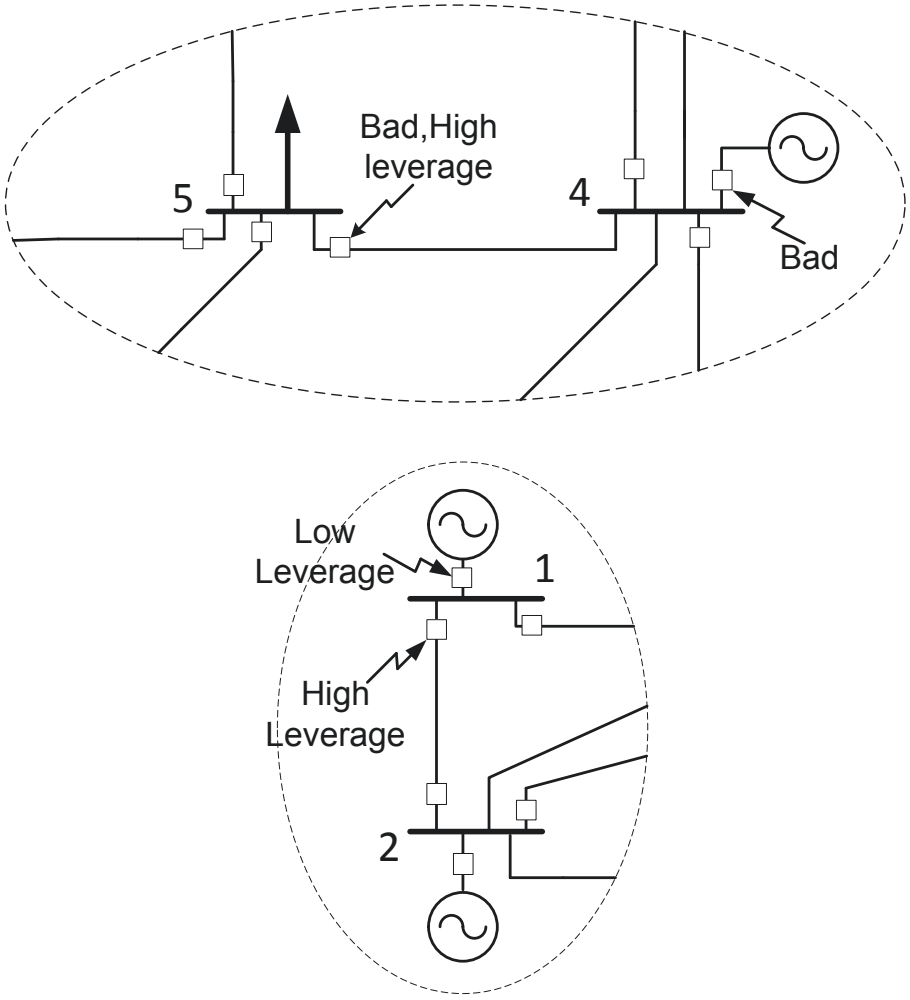

Figure 5.4: The sample high and low leverage points in IEEE-14 bus system

shown with text arrows in Figure 5.7, 5.8, that DRGP technique with GSR properly identifies the bad data for leverage measurements, however, the normalized residuals fail to do so. Table 5.7 further shows the GSR of the measurements and thus, validates the effectiveness of the strategy. Table 5.8 justifies the fact that DRGP is robust against swamping or masking effect. While the robust Mahalanobis distance masks some high leverage points as low leverages, the DRGP identifies all the high leverages correctly. The above strategy is robust against the size of the system and can be applied to larger standard systems such as IEEE-30 and IEEE-118 bus system. The next subsection provides the results for a standard large but meshed distribution 123-bus system.

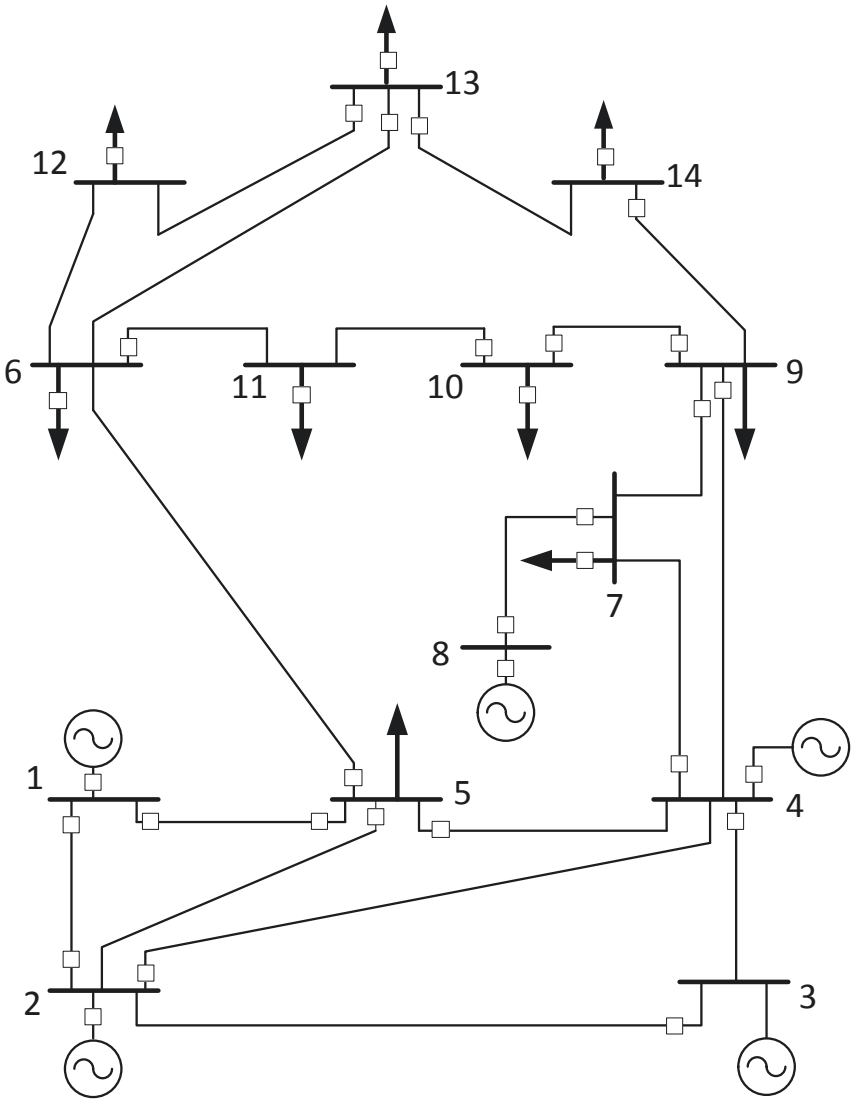

Figure 5.5: IEEE-14 bus system

Table 5.5: Comparison of studentized residuals with other residuals for 14-bus system

Measurement	Semi-studentized residuals (2.3)	Internally studentized residuals (3.0)	Externally studentized residuals (3.0)	DFFITS (1.782)	Cook's distance (1.00)
flow 2-1	0.8864	0.8208	0.9445	0.4176	0.4279
flow 3-2	2.2952	1.4276	2.1169	0.6811	0.7892
flow 2-4	0.8099	0.9821	0.349	0.9031	0.1404
flow 1-5	0.4656	0.5793	1.5759	1.2042	0.4107
flow 5-2	2.9818	**3.7311**	2.0244	1.4321	0.1201
flow 5-4	**2.7264**	**3.1437**	**5.4399**	0.6478	0.4197
flow 5-6	1.7476	1.3681	0.3057	1.1573	0.1691
flow 4-7	0.8080	0.6435	1.9444	0.7921	0.3198
flow 8-7	0.6419	0.8092	1.2097	0.7695	0.4180
flow 9-7	1.0385	1.4952	0.4564	1.4502	0.3179
flow 9-10	0.1676	0.1280	0.4745	1.2998	0.0981
flow 6-11	0.7222	0.8211	1.1889	0.8931	.4193
flow 13-6	0.4754	0.1704	0.5142	1.672	0.1801
flow 10-11	0.7417	0.7411	1.0561	0.8701	0.3153
flow 13-14	1.5130	1.3711	1.3913	0.7982	0.3172
inj 1	1.6799	**3.4143**	**6.0186**	0.7921	0.8793
inj 4	0.3914	0.3719	**4.3473**	1.2983	0.9168
inj 8	0.6934	0.5489	0.9061	1.4042	0.9082
inj 10	0.5051	0.4301	0.3792	1.2763	0.4193
inj 12	0.0713	0.1032	0.0393	0.6822	.3812
inj 14	1.8547	1.432	2.0724	1.4731	0.8932
flow 1-2	2.0240	2.4191	2.0664	0.7291	0.8911
flow 5-1	1.6224	1.4522	1.5051	0.4321	0.7821
flow 4-3	1.8094	1.2480	2.3539	0.4126	0.1794
flow 7-8	1.6933	1.3421	2.1443	1.3279	0.7891
flow 9-4	0.3187	0.2819	1.1517	1.2792	0.6871
flow 10-9	0.1167	0.3179	0.057	0.9110	0.7981
flow 14-9	0.9419	0.3183	0.3926	0.2479	0.4729

Continued on next page

Table 5.5: *Continued from previous page...*

Measurement	Semi-studentized residuals (2.3)	Internally studentized residuals (3.0)	Externally studentized residuals (3.0)	DFFITS (1.782)	Cook's distance (1.00)
flow 13-12	0.2060	0.1261	1.3136	1.593	0.7911
inj 2	2.6871	0.4271	1.9129	0.495	0.4792
inj 6	0.2100	0.2721	2.3154	1.110	0.6871
inj 7	1.4043	1.3211	0.234	0.4729	0.8862
inj 11	0.4903	0.4302	0.1057	0.4380	0.6621
inj 13	0.7720	0.8711	1.342	0.1793	0.6911

Table 5.6: The GSR-DRGP approach and LNR approach

Measurement	Normalized Residuals	Leverages identified by DRGP	GSR-DRGP	Bad Data
flow 5-4	2.7264	Yes	5.4399	Yes
inj 4	0.3914	Yes	4.3473	Yes
flow 1-2	2.0240	Yes	2.0664	No
inj 2	2.6871	No	1.9129	No
inj 1	1.6799	No	6.0186	Yes

5.4.3 Distribution system

The IEEE 123-bus test distribution system has also been considered for this study. The network parameters and load data are obtained from [51, 54]. The topologies of the test systems are shown in Figure 5.6. The voltage level of the system is 4.16 kV. There are both three-phase and single-phase loads. Thus, the system is inherently unbalanced. The three-phase loads are either star or delta connected. The loads are either constant current or constant impedance or constant power. The loads in the system have been modelled as ZIP-model. The test system consists of both overhead lines and underground cables. The overhead lines and

underground cables have been modelled with modified Carson's equations [49]. The distribution feeder is either three-phase or three-phase with a grounded neutral or single or two-phase laterals. Therefore, the impedance of each overhead line or underground cable is represented as either a 3x3 or a 4x4 matrix compared to a single element in single phase representation. However, the 4x4 matrix for three-phase lines with grounded neutral is converted to 3x3 matrix by Kron's reduction [49]. A three-phase transformer is modelled as three individual single-phase transformers. The tap changers are considered to have fixed taps. The switches between buses 14 and 117, 61 and 118, 19 and 116, 98 and 119 are considered closed.

The IEEE-123 test system has been modified to incorporate some leverage data points in the measurement data set. Injection measurements are added on buses 14, 19 and 55 and the lines 9-14 and 19-22 are made short. The switches between buses 55 and 95 and 123 and 121 are closed. This makes the network meshed in nature.

The measurements are generated by adding random Gaussian noise to the three-phase load flow results. The percentage error in real measurements is 3-5% and that in pseudo measurements is 20%. The gross errors are generated by changing the value of K_{ii} as explained in Section 4.2. Appendix D states the details.

The main advantage of this method is that it can separate and simultaneously identify the bad data points (outliers) and the leverages and, therefore, can be easily applied to the measurement set even if the high leverages are affected by gross error. These are reported here.

Table 5.7: Generalized potentials and studentized residuals for 14 bus system

Measurement No.	Measurement	Leverage (0.758)	DRGP (0.927)	GSR(3.0)
1	flow 2-1	0.5907	0.0297	-0.9445
2	flow 3-2	0.1943	0.05	2.1169
3	flow 2-4	0.6339	0.3652	-0.349
4	flow 1-5	**0.8152**	0.0677	-1.5759
5	flow 5-2	0.6124	0.5727	2.0244
6	**flow 5-4 (bad, high leverage)**	0.2519	**1.6442**	**5.4399**
7	flow 5-6	0.23	0.5057	-0.3057

Continued on next page

Table 5.7: *Continued from previous page...*

Measurement No.	Measurement	Leverage (0.758)	DRGP (0.927)	GSR(3.0)
8	flow 4-7	0.0729	0.1049	-1.9444
9	flow 8-7	0.461	0.4617	-1.2097
10	flow 9-7	0.5467	0.363	-0.4564
11	flow 9-10	0.4373	0.6811	-0.4745
12	flow 6-11	0.3406	0.4543	-1.1889
13	flow 13-6	0.5156	0.5602	0.5142
14	flow 10-11	0.5181	0.3177	1.0561
15	flow 13-14	0.5432	0.3027	-1.3913
16	**inj 1 (bad, low leverage)**	**0.7782**	0.0777	**6.0186**
17	**inj 4 (bad)**	**0.9065**	**0.9884**	**-4.3473**
18	inj 8	0.1671	0.0932	-0.9061
19	inj 10	0.5674	0.1214	-0.3792
20	inj 12	0.189	0.2737	0.0393
21	inj 14	0.0955	0.582	-2.0724
22	**flow 1-2 (good, high leverage)**	**0.8927**	**2.5896**	2.0664
23	flow 5-1	0.3601	0.0423	1.5051
24	flow 4-3	0.367	0.2795	-2.3539
25	flow 7-8	0.5296	0.3688	2.1443
26	flow 9-4	0.6162	0.2918	1.1517
27	flow 10-9	0.2802	0.4598	-0.057
28	flow 14-9	0.5296	0.4396	0.3926
29	flow 13-12	0.3329	0.2044	-1.3136
30	inj 2	**0.9115**	0.9173	-1.9129
31	inj 6	0.6663	0.0108	2.3154
32	inj 7	0.2052	0.6888	0.234
33	inj 11	0.4908	0.117	0.1057
34	inj 13	0.4658	0.0744	-1.342

Table 5.8: Masking or Swamping Effect for 14-bus system

Measurement	Identified by RMD	Identified by DRGP	Actual leverages	Bad Data
flow 5-4	No	Yes	Yes	Yes
inj 4	Yes	Yes	Yes	Yes
flow 1-2	Yes	Yes	Yes	No
inj 2	No	No	No	No
inj 1	Yes	No	No	Yes

Table 5.9: DRGP and GSR for 123 bus system

Measurement	Leverage (0.736)	DRGP (0.853)	GSR(3.0)	Bad Data
inj 55	0.854	1.7924	4.586	Yes
flow 54-55	0.756	1.8595	3.673	Yes
inj 14	0.675	0.9595	-4.457	Yes
flow 9-14	0.812	1.2595	2.0670	No
inj 67	0.478	0.5595	5.5465	Yes
inj 36	0.798	0.657	1.967	No

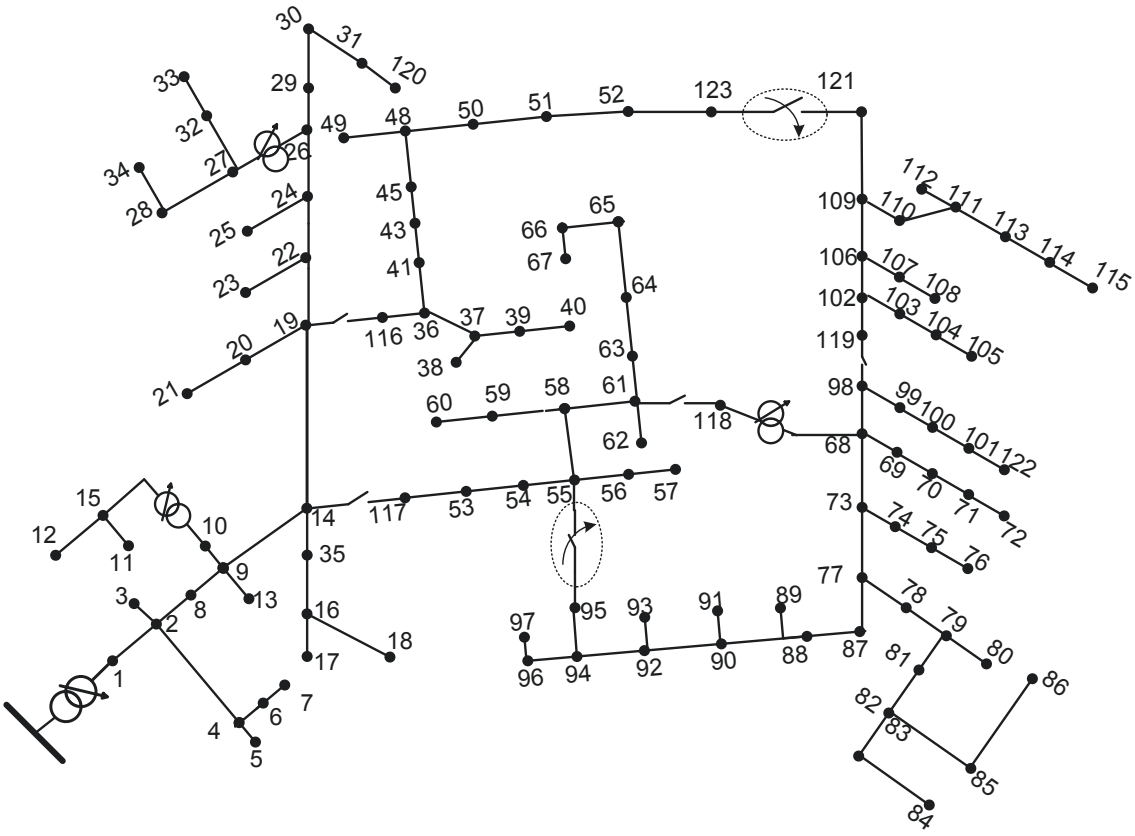

Figure 5.6: IEEE 123-bus distribution system

5.5 Discussions

The DRGP vs GSR graphs for the 14 bus and 123 bus systems are shown in Figure 5.8 and Figure 5.10 respectively. The positions of high leverage points, low leverage points, outliers on high leverage points and outliers on low leverage points are shown clearly. The high leverages and the bad data points are shown in red in the figures. As the bulk of the data are low leverages with low residuals, most of the data points lie around the origin. The points with high leverages are located in the upper area of the plot and the data points with large residuals lie either in the left or right of the plot. This is explained in the schematic in Figure 5.2. The measurements marked in red are highlighted in bold in Table 5.7. Table 5.9 shows the measurements marked in red for 123-bus system. The high leverage measurement (flow 5-4 in 14-bus system and inj 55 and flow 54-55 in 123-bus system) which contains gross error are located at the top right corner of the graph. The low leverage (inj 1 in 14-bus system and inj 67 in 123-bus system) with gross error is located at the extreme right

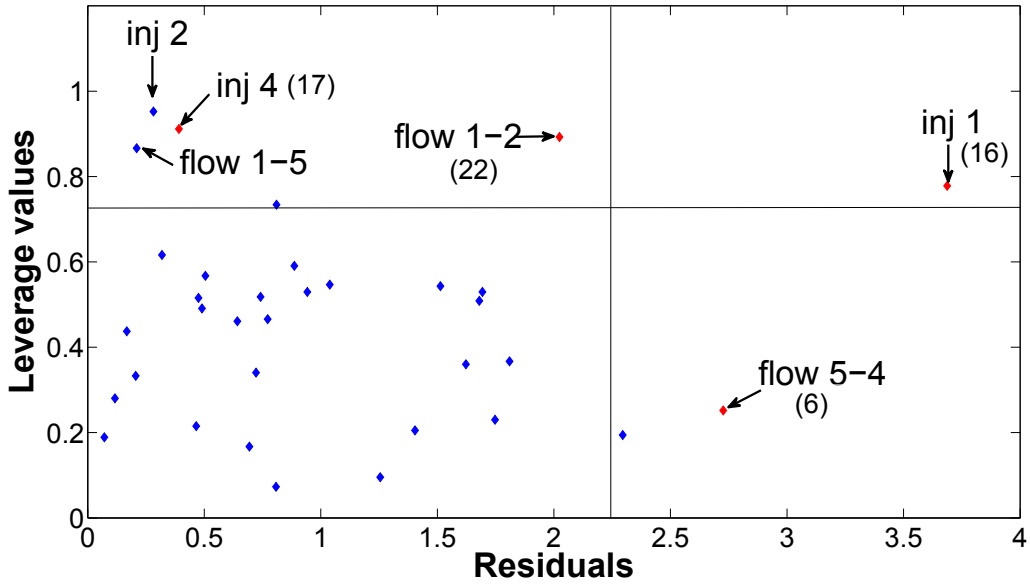

Figure 5.7: Leverage vs Residual plot for 14 bus system

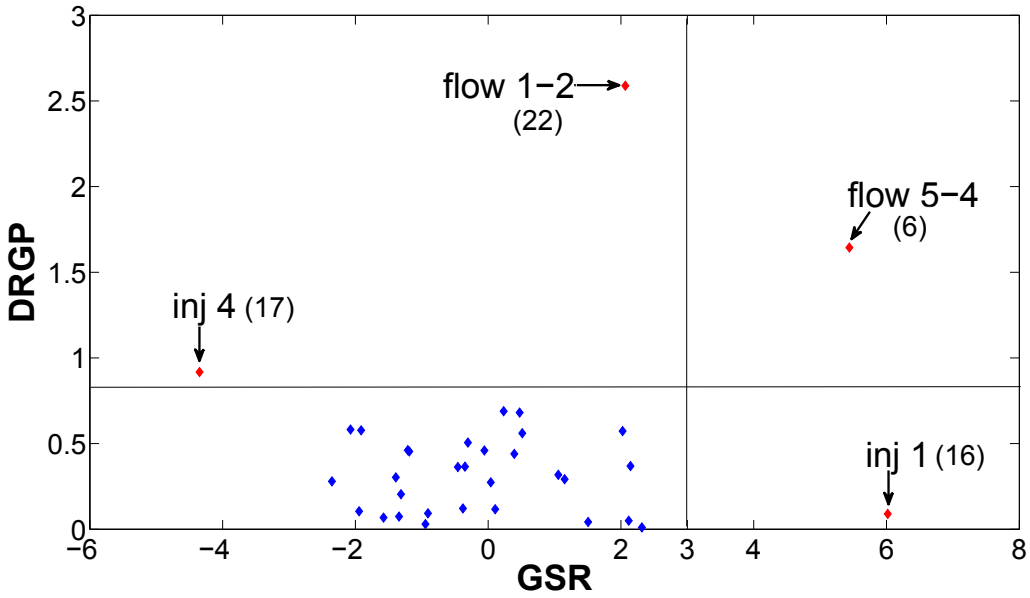

Figure 5.8: DRGP vs GSR plot for 14 bus system

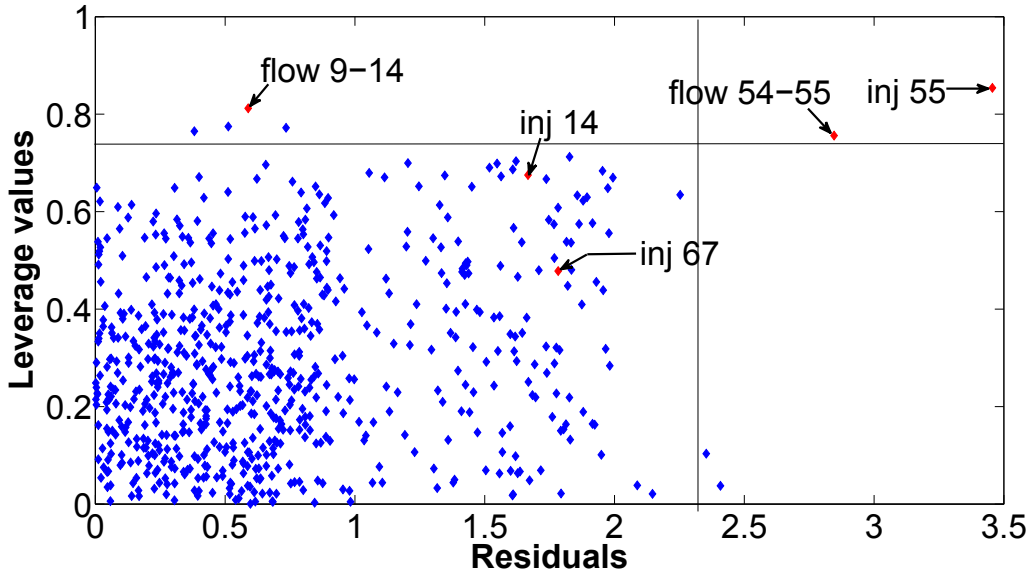

Figure 5.9: Leverage vs Residual plot for 123 bus system

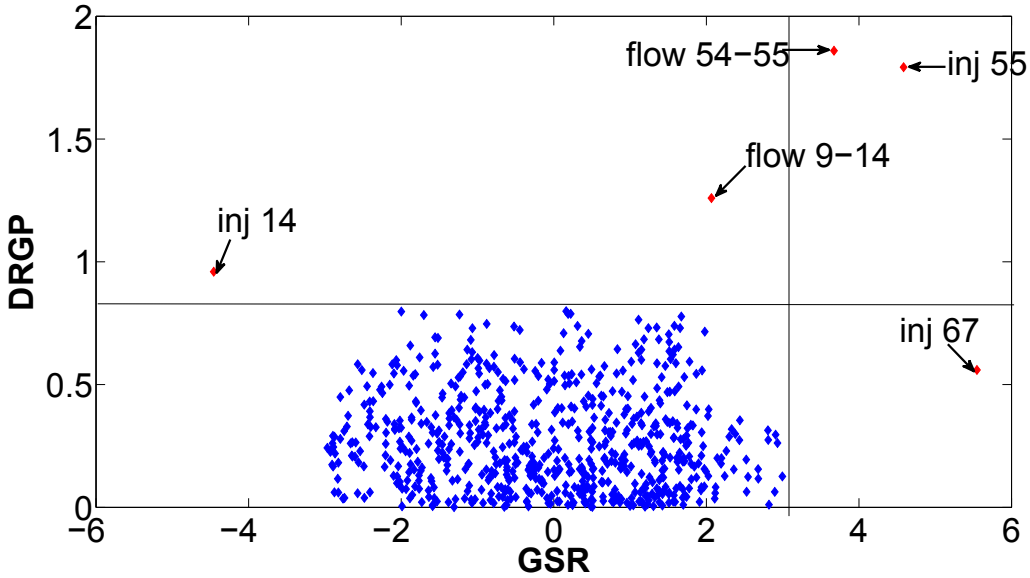

Figure 5.10: DRGP vs GSR plot for 123 bus system

end of the $x - axis$ of the graph. The high leverages (flow 1-2 in 14-bus system and flow 9-14 in 123-bus system) which are not contaminated with gross errors are located at the top of the graph. Figure 5.7 and Figure 5.9 show the plot of the leverage values (i.e. diagonal entries of the hat matrix) against the square of the normalized residuals. The same cases shown in red in Figure 5.8 and Figure 5.10 are shown in red here. However, here, the cases (inj 1, inj 2 and flow 1-5 in 14-bus system and inj 14 in 123-bus system) are swamped and the case (flow 4-3 in 14-bus system and inj 67 in 123-bus system) shows a large normalized residual. It is evident from the figure that it is difficult to differentiate the outliers from the high leverage points. Due to masking/swamping effect some measurements are misrepresented as high leverages and vice versa. The key measurement points are shown with red data points and text arrows in the Figures 5.7- 5.10.

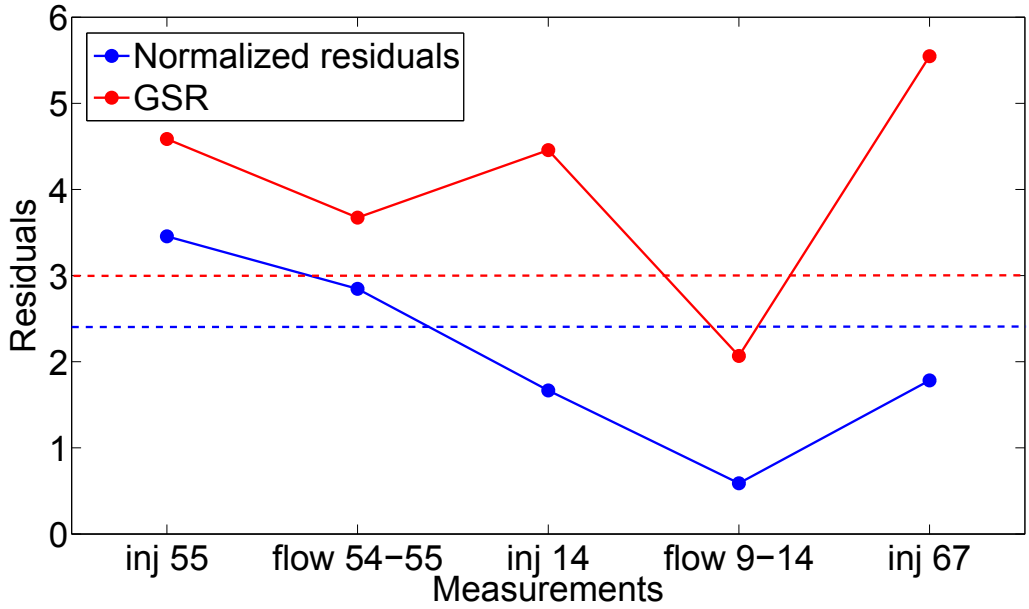

Figure 5.11: Comparison of GSR and normalized residuals of key measurements for 123-bus system

The above method has been applied to a small 4-bus example, balanced 14-bus system and unbalanced IEEE 123-bus systems. The generalized studentized residual has been used instead of the normalized residuals to identify the bad data. Even if the normalized/internally studentized residuals are low the GSR for the false data is significant. However, the above method has been compared with the normalized residual test to identify bad data. Figure 5.8, 5.10 and Figure 5.7, 5.9 justify the

effectiveness of the above algorithm. Table 5.5 compares the normalized/internally studentized residuals and other measures with externally studentized residuals for the 14-bus case, while Figure 5.11 shows the comparison of normalized residuals and GSR for key measurements in case of 123-bus distribution system. From both cases it can be inferred that while the largest normalized residual test fails to separate the outliers from the high leverages and thus fails to identify the bad data when there are multiple influential data points, the simultaneous technique of DRGP and GSR clearly separates the high leverages, low leverages and measurement outliers from one other and also prevents the masking or swamping effect in the presence of multiple influential data points. Thus the method has the capability to deal with deliberate man-made attack.

5.6 Conclusions

It is always necessary to detect erroneous measurements in active power networks. Due to growing deployment of ICT and automation technologies to operate modern power systems the measurements can be tampered for mala fide intentions. The attacker will always try to influence the states of the system by hiding the attack from the detection algorithm, which is possible if the high leverage measurements are especially targeted. The high leverages can occur in both transmission and distribution networks.

The research reported here has used the concept of regression analysis to identify the outliers and influential measurements in the system. It has been found that identifying the bad data for leverage measurements is particularly difficult due to the low value of residuals even if they are infected with gross errors. In addition, if there are multiple leverage measurements some of the high leverage measurement points may be masked or swamped. Hence, in order to take care of this masking and swamping effect, the concept of diagnostic-robust generalized potential has been proposed to separate the leverage measurements from rest of the measurements and then the studentized residuals are applied on the measurements to identify the bad data for multiple high leverage measurements. Moreover, even if there are large errors in high

leverage measurements it will be possible to identify them. Comprehensive results and comparative studies on both transmission and distribution systems/balanced and unbalanced systems further show the advantages of this methodology against other existing residual techniques to identify bad data against leverage attack. The proposed method can assist the EMS/DMS in taking control and operation decisions in these scenarios.

Chapter 6

Conclusion and Future Work

6.1 Summary of thesis contributions

This chapter summarises the research contributions of this thesis and discusses the insights gained from this research. The issues and challenges associated with DSSE have been identified and addressed in detail. The vulnerabilities in regards to state estimation and bad data detection in relation to leverage or influential measurements have been explored and algorithms have been developed to address these issues.

Due to the increasing automation and integration of ICT infrastructure in the modern power system, there is an increasing need for visibility of system states. Thus, state estimation has become an important function not only in transmission level but also in distribution level. Due to the topological and structural characteristics of the distribution system, it is inherently unbalanced. Over and above that, a large portion of distribution system is unmetered/unmeasured. The observability of the network is achieved by including pseudo measurements or loads in the measurement set.

First of all, the different components of a three-phase unbalanced system such as lines, transformers, loads and switches are modelled in detail. The difference between distribution systems and transmission systems and the challenges related to unbalanced distribution system modelling are addressed. Due to their non-transposed

nature distribution lines are modelled by Carson's equation. Three-phase transformers with different configurations will have different nodal admittance matrices. The different types of loads with star and delta configurations are modelled by ZIP models. Chapter 3 develops the WLS based state estimation algorithm for the full three-phase network. The zero injection measurements and switches are considered as equality constraints. The efficacy of the approach has been demonstrated by applying on IEEE 13 bus and IEEE 123 bus systems. The state estimates are found to vary with the percentage error in pseudo measurements. It has been shown that the estimates are less accurate when the percentage errors are high. The chapter further explores the suitability of the approach by changing the switch statuses of the IEEE 123 bus system.

The exhaustive modelling of three-phase distribution systems can be applied in a variety of power system applications and decision-making framework. It can be very useful for not only state estimation but also in other applications such as transformer tap estimation [43, 45, 78], bad data detection, security analysis [79], volt var control [80] and active network management schemes in distribution systems [46, 81].

The modern power network is undergoing a significant change. The smart instrumentation like phasor measurement units, intelligent metering, smart metering etc. are making the network active and smart. The information and communication infrastructure has become an integral part of the network. As a result, the security and integrity of data is at stake. An attacker, who has knowledge of the network, can exploit the vulnerabilities of the system by compromising the measurement meters. In particular, the attacker can take advantage of the low residuals of leverage measurement points to make the attack successful. The possible attack strategies for this attack have been explored. Chapter 4 studies the vulnerabilities of the network, both transmission and distribution, and further explains the suitability of studentized residuals against the different residual and robust estimation techniques for identification of bad data in relation to the leverage points.

The multiple leverages with inherently low residuals may also suffer from swamping or masking - a phenomenon where the low leverage point may appear to be high and the high to be low. This makes the identification of gross errors even more

difficult. Chapter 5 uses the concept of regression analysis to identify outliers and influential points. It proposes a robust method of diagnostic robust generalized potentials to identify the leverages. The methodology separates the high leverages from non-leverages and then applies the generalized studentized residuals. This completely nullifies the masking/swamping effects. The proposed method has been applied on a small 4-bus example, the IEEE 14 bus system and IEEE 123 bus system and the results justify the effectiveness of the approach. The above method can, therefore, help the EMS/DMS to make control and operation decisions.

6.2 Future Work

Although the main issues associated with state estimation and bad data detection have been addressed in the thesis, future research directions will focus on the following aspects of the problem:

- Tap positions play an important role in distribution system operation. An attacker can inject gross error into the tap position measurements. The reactive power flow through transformer will vary significantly when a tap measurement is in error. The reactive power flow being on a short line has the possibility to be a leverage measurement. Therefore, an extension of this work is to have a robust detection technique for bad tap measurements which is under process.

- Another aspect of future research is to incorporate the correlations among loads and correlation between loads and the real measurements, where the co-variance matrix R will be non-diagonal.

- The future extension of this work will be developing a decentralised state estimation in micro-grid scenario. The topology of the distribution changes due to interruptions, feeder maintenance etc. Most of the micro-grids have stand alone generators to support the demand in them. The idea is to develop the decentralised framework when those micro-grids are connected to the grid.

Bibliography

[1] C. Gomez-Quiles, A. Gomez-Exposito, and A. de la Villa Jaen, "State estimation for smart distribution substations," *IEEE Transactions on Smart Grid*, vol. 3, pp. 986–995, June 2012.

[2] M. E. Baran and A. W. Kelley, "A branch-current based state estimation for distribution systems," *IEEE Transactions on Power Systems*, 1995.

[3] W.-M. Lin, J.-H. Teng, and S.-J. Chen, "A highly efficient algorithm in treating current measurements for the branch-current-based distribution state estimation," *IEEE Transactions on Power Delivery*, vol. 16, pp. 433–439, Jul 2001.

[4] H. Wang and N. N. Schulz, "A revised branch current based distribution system state estimation algorithm and meter placement impact," *IEEE Transactions on Power Systems*, 2004.

[5] S. Wang, X. Cui, Z. Li, and M. Shahidehpour, "An improved branch current-based three-phase state estimation algorithm for distribution systems with dgs," in *IEEE Innovative Smart Grid Technologies - Asia (ISGT Asia), 2012*, pp. 1–6, May 2012.

[6] W.-M. Lin and J.-H. Teng, "State estimation for distribution systems with zero-injection constraints," *IEEE Transactions on Power Systems*, 1996.

[7] M. Baran and A. Kelley, "State estimation for real-time monitoring of distribution systems," *IEEE Transactions on Power Systems*, vol. 9, pp. 1601–1609, Aug 1994.

[8] K. Li, "State estimation for power distribution system and measurement impacts," *IEEE Transactions on Power Systems*, 1996.

[9] C. Lu, J. Teng, and W.-H. Liu, "Distribution system state estimation," *IEEE Transactions on Power Systems*, vol. 10, pp. 229–240, Feb 1995.

[10] A. Ghosh, D. Lubkeman, M. Downey, and R. Jones, "Distribution circuit state estimation using a probabilistic approach," *IEEE Transactions on Power Systems*, vol. 12, pp. 45–51, Feb 1997.

[11] J. Z. David L. Lubkeman and A. K. Ghosh, "Field results for a distribution circuit state estimator implementation," *IEEE Transactions on Power Delivery*, 2000.

[12] K. Jones, J. Thorp, and R. Gardner, "Three-phase linear state estimation using phasor measurements," in *Power and Energy Society General Meeting (PES), 2013 IEEE*, pp. 1–5, July 2013.

[13] C. Hansen and A. Debs, "Power system state estimation using three-phase models," *IEEE Transactions on Power Systems*, vol. 10, pp. 818–824, May 1995.

[14] A. Ranković, B. M. Maksimović, and A. T. Sarić, "A three-phase state estimation in active distribution networks," *International Journal of Electrical Power & Energy Systems*, vol. 54, pp. 154–162, 2014.

[15] C. Muscas, S. Sulis, A. Angioni, F. Ponci, and A. Monti, "Impact of different uncertainty sources on a three-phase state estimator for distribution networks," *IEEE Transactions on Instrumentation and Measurement*, vol. 63, pp. 2200–2209, Sept 2014.

[16] P. P. Barbeiro, H. Teixeira, J. Krstulovic, J. Pereira, and F. Soares, "Exploiting autoencoders for three-phase state estimation in unbalanced distributions grids," *Electric Power Systems Research*, vol. 123, pp. 108–118, 2015.

[17] D. Haughton and G. Heydt, "A linear state estimation formulation for smart distribution systems," *IEEE Transactions on Power Systems*, vol. 28, pp. 1187–1195, May 2013.

[18] D. Thukaram, J. Jerome, and C. Surapong, "A robust three-phase state estimation algorithm for distribution networks," *Electric Power Systems Research*, vol. 55, no. 3, pp. 191–200, 2000.

[19] S. Depuru, L. Wang, V. Devabhaktuni, and N. Gudi, "Smart meters for power grid-challenges, issues, advantages and status," in *IEEE/PES Power Systems Conference and Exposition (PSCE), 2011*, pp. 1–7, March 2011.

[20] P. Koponen, L. D. Saco, N. Orchard, T. Vorisek, J. Parsons, C. Rochas, A. Z. Morch, V. Lopes, and M. Togeby, "Definition of smart metering and applications and identification of benefits," *Deliverable D3 of the European Smart Metering Alliance ESMA (available at www. esma-home. eu, members area)*, 2008.

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

[22] F. Dorfler, F. Pasqualetti, and F. Bullo, "Distributed detection of cyber-physical attacks in power networks: A waveform relaxation approach," in *49th Annual Allerton Conference on Communication, Control, and Computing (Allerton), 2011*, pp. 1486–1491, Sept 2011.

[23] G. Heydt, "The next generation of power distribution systems," *IEEE Transactions on Smart Grid*, vol. 1, pp. 225–235, Dec 2010.

[24] O. Kosut, L. Jia, R. Thomas, and L. Tong, "Malicious data attacks on the smart grid," *IEEE Transactions on Smart Grid*, vol. 2, pp. 645–658, Dec 2011.

[25] L. Jia, O. Kosut, R. Thomas, and L. Tong, "Malicious data attacks on smart grid state estimation: Attack strategies and countermeasures," in *First IEEE*

International Conference on Smart Grid Communications (SmartGridComm), 2010, pp. 220–225, Oct 2010.

[26] A. Monticelli, *State Estimation in Electric Power Systems, A Generalized Approach*. Luwer's power Electronics and power Systems Series, 2004.

[27] A. Abur and A. Exposito, *Power System State Estimation, Theory and Implementation*. CRC Press, 2004.

[28] A. Meliopoulos, B. Fardanesh, and S. Zelingher, "Power system state estimation: modeling error effects and impact on system operation," in *System Sciences, 2001. Proceedings of the 34th Annual Hawaii International Conference on*, pp. 682–690, Jan 2001.

[29] S. Zhong and A. Abur, "Effects of nontransposed lines and unbalanced loads on state estimation," in *Power Engineering Society Winter Meeting, 2002. IEEE*, vol. 2, pp. 975–979 vol.2, 2002.

[30] S. L. Els, A. D. Els, J. Jordaan, and R. Zivanovic, "Projection statistics for power system state estimation," in *Africon, 1999 IEEE*, vol. 2, pp. 783–786 vol.2, 1999.

[31] L. Mili, M. Cheniae, N. Vichare, and P. Rousseeuw, "Robust state estimation based on projection statistics [of power systems]," *IEEE Transactions on Power Systems*, vol. 11, pp. 1118–1127, May 1996.

[32] Q. Yang, J. Yang, W. Yu, N. Zhang, and W. Zhao, "On a hierarchical false data injection attack on power system state estimation," in *Global Telecommunications Conference (GLOBECOM 2011), 2011 IEEE*, pp. 1–5, Dec 2011.

[33] S. Bi and Y. Zhang, "Defending mechanisms against false-data injection attacks in the power system state estimation," in *GLOBECOM Workshops (GC Wkshps), 2011 IEEE*, pp. 1162–1167, Dec 2011.

[34] G. Hug and J. Giampapa, "Vulnerability assessment of ac state estimation with respect to false data injection cyber-attacks," *IEEE Transactions on Smart Grid*, vol. 3, pp. 1362–1370, Sept 2012.

[35] J. Chen and A. Abur, "Placement of pmus to enable bad data detection in state estimation," *IEEE Transactions on Power Systems*, vol. 21, pp. 1608–1615, Nov 2006.

[36] A. Narvaez and S. Grijalva, "Robust state estimator applied to the ecuadorian electric power system," in *Transmission and Distribution Conference and Exposition: Latin America, 2008 IEEE/PES*, pp. 1–6, Aug 2008.

[37] J. Khwanram and P. Damrongkulkamjorn, "Multiple bad data identification in power system state estimation using particle swarm optimization," in *Electrical Engineering/Electronics, Computer, Telecommunications and Information Technology, 2009. ECTI-CON 2009. 6th International Conference on*, vol. 01, pp. 2–5, May 2009.

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

[39] J.-M. Lin and H.-Y. Pan, "A static state estimation approach including bad data detection and identification in power systems," in *Power Engineering Society General Meeting, 2007. IEEE*, pp. 1–7, June 2007.

[40] A. Tarali and A. Abur, "Bad data detection in two-stage state estimation using phasor measurements," in *Innovative Smart Grid Technologies (ISGT Europe), 2012 3rd IEEE PES International Conference and Exhibition on*, pp. 1–8, Oct 2012.

[41] A. Majumdar and B. Pal, "A three-phase state estimation in unbalanced distribution networks with switch modelling," in *First IEEE International Conference on Control, Measurement and Instrumentation, Kolkata, India*, 2016.

[42] A. Majumdar and B. Pal, "Bad data detection in the context of leverage point attacks in modern power networks," *IEEE Transactions on Smart Grid*, Revision under review.

[43] S. Nanchian, A. Majumdar, and B. Pal, "Three-phase state estimation using hybrid particle swarm optimization," *IEEE Transactions on Smart Grid*, vol. PP, no. 99, pp. 1–1, 2015.

[44] S. Nanchian, A. Majumdar, and B. Pal, "Transformer tap estimation using hybrid particle swarm optimization," in *IEEE PES General Meeting, Washington D.C.*, 2014.

[45] S. Nanchian, A. Majumdar, and B. Pal, "Ordinal optimization technique for three phase distribution network state estimation with discrete and continuous variables," *IEEE Transactions on Smart Grid*, Revision under review.

[46] G. Mokryani, A. Majumdar, and B. Pal, "A probabilistic method for the operation of three-phase unbalanced active distribution networks," *IET Renewable Power Generation*, Revision under review.

[47] J. Martinez and J. Mahseredjian, "Load flow calculations in distribution systems with distributed resources. a review," in *Power and Energy Society General Meeting, 2011 IEEE*, pp. 1–8, July 2011.

[48] R. Singh, B. Pal, and R. Jabr, "Statistical representation of distribution system loads using gaussian mixture model," *IEEE Transactions on Power Systems*, vol. 25, pp. 29–37, Feb 2010.

[49] W. H. Kersting, *Distribution system modeling and analysis*. CRC press, 2012.

[50] P. Xiao, D. Yu, and W. Yan, "A unified three-phase transformer model for distribution load flow calculations," *IEEE Transactions on Power Systems*, vol. 21, pp. 153–159, Feb 2006.

[51] "Distribution test feeders." http://ewh.ieee.org/soc/pes/dsacom/testfeeders/index.html.

[52] W. H. Kersting, "Radial distribution test feeders," in *Power Engineering Society Winter Meeting, 2001. IEEE*, vol. 2, pp. 908–912, IEEE, 2001.

[53] S. S. S. R. Depuru, L. Wang, and V. Devabhaktuni, "Smart meters for power grid: Challenges, issues, advantages and status," *Renewable and sustainable energy reviews*, vol. 15, no. 6, pp. 2736–2742, 2011.

[54] W. Kersting, "Radial distribution test feeders," in *Power Engineering Society Winter Meeting, 2001. IEEE*, vol. 2, pp. 908–912 vol.2, 2001.

[55] E. Caro, A. Conejo, R. Minguez, M. Zima, and G. Andersson, "Multiple bad data identification considering measurement dependencies," *Power Systems, IEEE Transactions on*, vol. 26, pp. 1953–1961, Nov 2011.

[56] S. Bi and Y. J. Zhang, "Graphical methods for defense against false-data injection attacks on power system state estimation," *IEEE Transactions on Smart Grid*, vol. 5, pp. 1216–1227, May 2014.

[57] K. C. Sou, H. Sandberg, and K. Johansson, "Computing critical k -tuples in power networks," *IEEE Transactions on Power Systems*, vol. 27, pp. 1511–1520, Aug 2012.

[58] M. Gol and A. Abur, "Lav based robust state estimation for systems measured by pmus," *IEEE Transactions on Smart Grid*, vol. 5, pp. 1808–1814, July 2014.

[59] E. Asada, A. Garcia, and R. Romero, "Identifying multiple interacting bad data in power system state estimation," in *Power Engineering Society General Meeting, 2005. IEEE*, pp. 571–577 Vol. 1, June 2005.

[60] L. Mili, V. Phaniraj, and P. Rousseeuw, "Least median of squares estimation in power systems," *IEEE Transactions on Power Systems*, vol. 6, pp. 511–523, May 1991.

[61] S. Tan, W.-Z. Song, M. Stewart, and L. Long, "Lpattack: Leverage point attacks against state estimation in smart grid," in *Global Communications Conference (GLOBECOM), 2014 IEEE*, pp. 643–648, Dec 2014.

[62] M. Habshah, M. Norazan, and A. Rahmatullah Imon, "The performance of diagnostic-robust generalized potentials for the identification of multiple high

leverage points in linear regression," *Journal of Applied Statistics*, vol. 36, no. 5, pp. 507–520, 2009.

[63] A. Nurunnabi, A. S. Hadi, and A. Imon, "Procedures for the identification of multiple influential observations in linear regression," *Journal of Applied Statistics*, vol. 41, no. 6, pp. 1315–1331, 2014.

[64] M. Meloun and J. Militky, "Detection of single influential points in ols regression model building," *Analytica Chimica Acta*, vol. 439, no. 2, pp. 169–191, 2001.

[65] D. A. Belsley, E. Kuh, and R. E. Welsch, *Regression diagnostics: Identifying influential data and sources of collinearity*, vol. 571. John Wiley & Sons, 2005.

[66] A. Rahmatullah Imon, "Identifying multiple influential observations in linear regression," *Journal of Applied statistics*, vol. 32, no. 9, pp. 929–946, 2005.

[67] R. D. Cook and S. Weisberg, *Residuals and influence in regression*. New York: Chapman and Hall, 1982.

[68] C. R. Rao, *Linear statistical inference and its applications*, vol. 22. John Wiley & Sons, 2009.

[69] A. S. Hadi and J. S. Simonoff, "Procedures for the identification of multiple outliers in linear models," *Journal of the American Statistical Association*, vol. 88, no. 424, pp. 1264–1272, 1993.

[70] P. J. Huber *et al.*, "Robust estimation of a location parameter," *The Annals of Mathematical Statistics*, vol. 35, no. 1, pp. 73–101, 1964.

[71] L. Mili, M. Cheniae, and P. Rousseeuw, "Robust state estimation of electric power systems," *Circuits and Systems I: Fundamental Theory and Applications, IEEE Transactions on*, vol. 41, pp. 349–358, May 1994.

[72] R. A. Jabr and B. C. Pal, "Iteratively reweighted least-squares implementation of the wlav state-estimation method," *IEE Proceedings - Generation, Transmission and Distribution*, vol. 151, pp. 103–108, Jan 2004.

[73] W. J. Egan and S. L. Morgan, "Outlier detection in multivariate analytical chemical data," *Analytical chemistry*, vol. 70, no. 11, pp. 2372–2379, 1998.

[74] L. Mili, M. Cheniae, N. S. Vichare, and P. Rousseeuw, "Robust state estimation based on projection statistics of power systems," *IEEE Transactions on Power Systems*, vol. 11, pp. 1118–1127, May 1996.

[75] A. S. Hadi, "A new measure of overall potential influence in linear regression," *Computational Statistics & Data Analysis*, vol. 14, no. 1, pp. 1–27, 1992.

[76] "Advanced regression diagnostic methods." https://onlinecourses.science.psu.edu/stat501/sites/onlinecourses.science.psu.edu.stat501/files/pt3_adv_regression_diagnostics.pdf.

[77] "Power systems test case archive." http://www.ee.washington.edu/research/pstca/.

[78] G. N. Korres, P. J. Katsikas, and G. C. Contaxis, "Transformer tap setting observability in state estimation," *Power Systems, IEEE Transactions on*, vol. 19, no. 2, pp. 699–706, 2004.

[79] W. Fan, "Security of state estimation in the smart grid."

[80] M. Biserica, Y. Besanger, R. Caire, O. Chilard, and P. Deschamps, "Neural networks to improve distribution state estimation-volt var control performances," *IEEE Transactions on Smart Grid*, vol. 3, pp. 1137–1144, Sept 2012.

[81] "Distribution system state estimation will impact operation of the smart grid." http://smartgrid.ieee.org/newsletters/june-2015/distribution-system-state-estimation-will-impact-operation-of-the-smart-grid.

Appendix A

IEEE 13-bus system data

Table A.1: Line configuration data

Configuration	**Type**	**Phasing**	**Phase**	**Neutral**	**Spacing**
			ACSR	**ACSR**	**ID**
601	overhead	BACN	556,500 26/7	4/0 6/1	500
602	overhead	CABN	4/0 6/1	4/0 6/1	500
603	overhead	CBN	1/0	1/0	505
604	overhead	ACN	1/0	1/0	505
605	overhead	CN	1/0	1/0	510
606	underground	ABCN	250,000 AA, CN	None	515
607	underground	AN	1/0 AA, TS	1/0 Cu	520

Table A.2: Line segment data

Bus A	**Bus B**	**Length(ft.)**	**Configuration**
4	3	500	603
4	5	500	602
5	6	0	XFM-1
3	2	300	603
1	4	2000	601
8	12	800	607
4	9	2000	601
9	8	300	604
9	13	1000	601
9	10	0	Switch
8	7	300	605
10	11	500	606

Table A.3: Transformer data

	kVA	**kV (high)**	**kV (low)**	$R(\%)$	$X(\%)$
Substation	5000	115-Δ	4.16-Gr. Y	1	8
XFM-1	500	4.16-Gr. Y	0.48-Gr. Y	1.1	2

Table A.4: Capacitor data

Bus	**ph a**	**ph b**	**ph c**
	kVAr	**kVAr**	**kVAr**
11	200	200	200
7	-	-	100
Total	200	200	300

Table A.5: Regulator data

Regulator ID	**1**		
Line segment	1-4		
Location	50		
Phases	A-B-C		
Connection	3-ph,LG		
Monitoring phase	A-B-C		
Bandwidth	2 volts		
PT ratio	20		
Primary CT rating	700		
Compensator settings	ph a	ph b	ph c
R-setting	3	3	3
X-setting	9	9	9
Voltage level	122	122	122

Table A.6: Load data

Bus	**Load**	**ph a**	**ph a**	**ph b**	**ph b**	**ph c**	**ph c**
	Model	**kW**	**kVAr**	**kW**	**kVAr**	**kW**	**kVAr**
6	Y-PQ	160	110	120	90	120	90
3	Y-PQ	0	0	170	125	0	0
2	Δ-Z	0	0	230	132	0	0
12	Y-Z	128	86	0	0	0	0
9	Δ-PQ	385	220	385	220	385	220
11	Y-PQ	485	190	68	60	290	212
10	Δ-I	0	0	0	0	170	151
7	Y-I	0	0	0	0	170	80

Appendix B

IEEE 123-bus system data

Table B.1: Line configuration data

Configuration	Type	Phasing	Phase	Neutral	Spacing
			ACSR	ACSR	ID
1	overhead	ABCN	336,400 26/7	4/0 6/1	500
2	overhead	CABN	336,400 26/7	4/0 6/1	500
3	overhead	BCAN	336,400 26/7	4/0 6/1	500
4	overhead	CBAN	336,400 26/7	4/0 6/1	500
5	overhead	BACN	336,400 26/7	4/0 6/1	500
6	overhead	ACBN	336,400 26/7	4/0 6/1	500
7	overhead	ACN	336,400 26/7	4/0 6/1	505
8	overhead	ABN	336,400 26/7	4/0 6/1	505
9	overhead	AN	1/0	1/0	510
10	overhead	BN	1/0	1/0	510
11	overhead	CN	1/0	1/0	510
12	underground	ABC	1/0 AA, CN	1/0	515

Table B.2: Line segment data

Bus A	Bus B	Length(ft.)	Configuration
2	3	175	10
2	4	250	11
2	8	300	1

Continued on next page

Table B.2: Line segment data

Bus A	**Bus B**	**Length(ft.)**	**Configuration**
4	5	200	11
4	6	325	11
6	7	250	11
8	9	200	1
9	13	225	10
9	10	225	9
9	14	300	1
10	15	425	9
14	35	150	11
14	19	825	2
15	12	250	9
15	11	250	9
16	17	375	11
16	18	350	11
19	20	250	9
19	22	300	2
20	21	325	9
22	23	525	10
22	24	250	2
24	25	550	11
24	26	275	2
26	27	350	7
26	29	200	2
27	28	275	7
27	32	225	11
28	34	500	9
29	30	300	2
30	31	350	2
31	120	200	2
32	33	300	11
35	16	100	11
36	37	650	8

Continued on next page

Table B.2: Line segment data

Bus A	**Bus B**	**Length(ft.)**	**Configuration**
36	41	250	1
37	38	300	9
37	39	250	10
39	40	325	10
41	42	325	11
41	43	250	1
43	44	500	10
43	45	200	1
45	46	200	9
45	48	250	1
46	47	300	9
48	49	150	4
48	50	250	4
50	51	250	4
51	52	250	4
53	54	200	1
54	55	125	1
55	56	275	1
55	58	350	3
56	57	275	1
58	59	250	10
58	61	750	3
59	60	250	10
61	62	550	5
61	63	250	12
63	64	175	12
64	65	350	12
65	66	425	12
66	67	325	12
68	69	200	9
68	73	275	3
68	98	250	3

Continued on next page

Table B.2: Line segment data

Bus A	**Bus B**	**Length(ft.)**	**Configuration**
69	70	275	9
70	71	325	9
71	72	275	9
73	74	275	11
73	77	200	3
74	75	350	11
75	76	400	11
77	78	400	6
77	87	700	3
78	79	100	6
79	80	225	6
79	81	475	6
81	82	475	6
82	83	250	6
82	85	675	11
83	84	250	6
85	86	475	11
87	88	450	6
88	89	175	9
88	90	275	6
90	91	225	10
90	92	225	6
92	93	300	11
92	94	225	6
94	95	275	9
94	96	300	6
96	97	200	10
98	99	275	3
99	100	550	3
100	101	300	3
101	122	800	3
102	103	225	11

Continued on next page

Table B.2: Line segment data

Bus A	**Bus B**	**Length(ft.)**	**Configuration**
102	106	275	3
103	104	325	11
104	105	700	11
106	107	225	10
106	109	325	3
107	108	575	10
109	110	450	9
109	121	1000	3
110	111	300	9
111	112	575	9
111	113	125	9
113	114	525	9
114	115	325	9
116	36	375	4
1	2	400	1
117	53	400	1
118	68	350	6
119	102	250	3
14	117	0	Switch
19	116	0	Switch
61	118	0	Switch
98	119	0	Switch

Table B.3: Transformer data

	kVA	**kV (high)**	**kV (low)**	$R(\%)$	$X(\%)$
Substation	5000	115-Δ	4.16-Gr. Y	1	8
XFM-1	150	4.16-Gr. Y	0.48-Gr. Y	1.27	2.72

Table B.4: Three phase switches

Bus A	**Bus B**	**Normal**
14	117	closed
19	116	closed
61	118	closed
98	119	closed
55	95	open
123	121	open

Table B.5: Capacitor data

Bus	**ph a**	**ph b**	**ph c**
	kVAr	**kVAr**	**kVAr**
84	200	200	200
89	50	-	-
91	-	50	-
93	-	-	50
Total	250	250	250

Table B.6: Regulator data

Regulator ID	**1**			**Regulator ID**	**3**		
Line segment	1-2			**Line segment**	26-27		
Location	1			**Location**	27		
Phases	A-B-C			**Phases**	A-C		
Connection	3-ph,Y			**Connection**	2-ph,LG		
Monitoring phase	A			**Monitoring phase**	A,C		
Bandwidth	2 volts			**Bandwidth**	1 volts		
PT ratio	20			**PT ratio**	20		
Primary CT rating	700			**Primary CT rating**	50		
Compensator settings	ph a			**Compensator settings**	ph a	ph c	
R-setting	3			**R-setting**	0.4	0.4	
X-setting	7.5			**X-setting**	0.4	0.4	
Voltage level	120			**Voltage level**	120	120	
Regulator ID	**2**			**Regulator ID**	**4**		
Line segment	10-15			**Line segment**	118-68		
Location	10			**Location**	68		
Phases	A			**Phases**	A-B-C		
Connection	1-ph,LG			**Connection**	3-ph,LG		
Monitoring phase	A			**Monitoring phase**	A,B,C		
Bandwidth	2 volts			**Bandwidth**	2 volts		
PT ratio	20			**PT ratio**	20		
Primary CT rating	50			**Primary CT rating**	300		
Compensator settings	ph a			**Compensator settings**	ph a	ph b	ph c
R-setting	0.4			**R-setting**	0.6	1.4	0.2
X-setting	0.4			**X-setting**	1.3	2.6	1.4
Voltage level	120			**Voltage level**	124	124	124

Table B.7: Load data

Bus	**Load**	**ph a**	**ph a**	**ph b**	**ph b**	**ph c**	**ph c**
	Model	**kW**	**kVAr**	**kW**	**kVAr**	**kW**	**kVAr**
1	Y-PQ	0	0	0	0	0	0
2	Y-PQ	40	20	0	0	0	0
3	Y-PQ	0	0	20	10	0	0
4	Y-PQ	0	0	0	0	0	0
5	Y-PQ	0	0	0	0	40	20
6	Y-I	0	0	0	0	20	10
7	Y-Z	0	0	0	0	40	20
8	Y-PQ	20	10	0	0	0	0
9	Y-PQ	0	0	0	0	0	0
10	Y-PQ	40	20	0	0	0	0
11	Y-I	20	10	0	0	0	0
12	Y-Z	40	20	0	0	0	0
13	Y-PQ	0	0	20	10	0	0
14	Y-PQ	0	0	0	0	0	0
15	Y-PQ	0	0	0	0	0	0
16	Y-PQ	0	0	0	0	0	0
17	Y-PQ	0	0	0	0	40	20
18	Y-PQ	0	0	0	0	20	10
19	Y-PQ	0	0	0	0	0	0
20	Y-PQ	40	20	0	0	0	0
21	Y-I	40	20	0	0	0	0
22	Y-PQ	0	0	0	0	0	0
23	Y-Z	0	0	40	20	0	0
24	Y-PQ	0	0	0	0	0	0
25	Y-PQ	0	0	0	0	40	20
26	Y-PQ	0	0	0	0	0	0
27	Y-PQ	0	0	0	0	0	0
28	Y-PQ	0	0	0	0	0	0
29	Y-I	40	20	0	0	0	0
30	Y-Z	40	20	0	0	0	0
31	Y-PQ	0	0	0	0	40	20

Continued on next page

Table B.7: Load data

Bus	**Load**	**ph a**	**ph a**	**ph b**	**ph b**	**ph c**	**ph c**
	Model	**kW**	**kVAr**	**kW**	**kVAr**	**kW**	**kVAr**
32	Y-PQ	0	0	0	0	20	10
33	Y-PQ	0	0	0	0	20	10
34	Y-I	40	20	0	0	0	0
35	Y-Z	0	0	0	0	40	20
36	D-PQ	40	20	0	0	0	0
37	Y-PQ	0	0	0	0	0	0
38	Y-Z	40	20	0	0	0	0
39	Y-I	0	0	20	10	0	0
40	Y-PQ	0	0	20	10	0	0
41	Y-PQ	0	0	0	0	0	0
42	Y-PQ	0	0	0	0	20	10
43	Y-PQ	20	10	0	0	0	0
44	Y-Z	0	0	40	20	0	0
45	Y-PQ	0	0	0	0	0	0
46	Y-I	20	10	0	0	0	0
47	Y-PQ	20	10	0	0	0	0
48	Y-I	35	25	35	25	35	25
49	Y-Z	70	50	70	50	70	50
50	Y-PQ	35	25	70	50	35	25
51	Y-PQ	0	0	0	0	40	20
52	Y-PQ	20	10	0	0	0	0
53	Y-PQ	40	20	0	0	0	0
54	Y-PQ	40	20	0	0	0	0
55	Y-PQ	0	0	0	0	0	0
56	Y-Z	20	10	0	0	0	0
57	Y-PQ	0	0	20	10	0	0
58	Y-PQ	0	0	0	0	0	0
59	Y-I	0	0	20	10	0	0
60	Y-PQ	0	0	20	10	0	0
61	Y-PQ	20	10	0	0	0	0
62	Y-PQ	0	0	0	0	0	0

Continued on next page

Table B.7: Load data

Bus	Load	ph a	ph a	ph b	ph b	ph c	ph c
	Model	kW	kVAr	kW	kVAr	kW	kVAr
63	Y-Z	0	0	0	0	40	20
64	Y-PQ	40	20	0	0	0	0
65	Y-I	0	0	75	35	0	0
66	D-Z	35	25	35	25	70	50
67	Y-PQ	0	0	0	0	75	35
68	Y-PQ	0	0	0	0	0	0
69	Y-PQ	20	10	0	0	0	0
70	Y-PQ	40	20	0	0	0	0
71	Y-PQ	20	10	0	0	0	0
72	Y-PQ	40	20	0	0	0	0
73	Y-PQ	0	0	0	0	0	0
74	Y-PQ	0	0	0	0	40	20
75	Y-Z	0	0	0	0	40	20
76	Y-PQ	0	0	0	0	40	20
77	D-I	105	80	70	50	70	50
78	Y-PQ	0	0	40	20	0	0
79	Y-PQ	0	0	0	0	0	0
80	Y-Z	40	20	0	0	0	0
81	Y-PQ	0	0	40	20	0	0
82	Y-PQ	0	0	0	0	0	0
83	Y-PQ	40	20	0	0	0	0
84	Y-PQ	0	0	0	0	20	10
85	Y-PQ	0	0	0	0	20	10
86	Y-PQ	0	0	0	0	40	20
87	Y-PQ	0	0	20	10	0	0
88	Y-PQ	0	0	40	20	0	0
89	Y-PQ	40	20	0	0	0	0
90	Y-PQ	0	0	0	0	0	0
91	Y-I	0	0	40	20	0	0
92	Y-PQ	0	0	0	0	0	0
93	Y-PQ	0	0	0	0	40	20

Continued on next page

Table B.7: Load data

Bus	**Load**	**ph a**	**ph a**	**ph b**	**ph b**	**ph c**	**ph c**
	Model	**kW**	**kVAr**	**kW**	**kVAr**	**kW**	**kVAr**
94	Y-PQ	0	0	0	0	0	0
95	Y-PQ	40	20	0	0	0	0
96	Y-PQ	0	0	20	10	0	0
97	Y-PQ	0	0	20	10	0	0
98	Y-PQ	0	0	0	0	0	0
99	Y-PQ	40	20	0	0	0	0
100	Y-PQ	0	0	40	20	0	0
101	Y-Z	0	0	0	0	40	20
102	Y-PQ	0	0	0	0	0	0
103	Y-PQ	0	0	0	0	20	10
104	Y-PQ	0	0	0	0	40	20
105	Y-PQ	0	0	0	0	40	20
106	Y-PQ	0	0	0	0	0	0
107	Y-PQ	0	0	40	20	0	0
108	Y-PQ	0	0	40	20	0	0
109	Y-PQ	0	0	0	0	0	0
110	Y-PQ	40	20	0	0	0	0
111	Y-PQ	0	0	0	0	0	0
112	Y-PQ	20	10	0	0	0	0
113	Y-I	20	10	0	0	0	0
114	Y-Z	40	20	0	0	0	0
115	Y-PQ	20	10	0	0	0	0
116	Y-PQ	0	0	0	0	0	0
117	Y-PQ	0	0	0	0	0	0
118	Y-PQ	0	0	0	0	0	0
119	Y-PQ	0	0	0	0	0	0
120	Y-PQ	0	0	0	0	0	0
121	Y-PQ	0	0	0	0	0	0
122	Y-PQ	0	0	0	0	0	0
123	Y-PQ	0	0	0	0	0	0

Appendix C

IEEE 14-bus system data

Table C.1: Bus data

	Active load	**Reactive load**	**Active generation**	**Reactive generation**
Bus	**kW**	**kVAR**	**kW**	**kVAR**
1	0	0	232.4	-16.9
2	21.7	12.7	40	42.4
3	94.2	19	0	23.4
4	47.8	-3.9	0	0
5	7.6	1.6	0	0
6	11.2	7.5	0	12.2
7	0	0	0	0
8	0	0	0	17.4
9	29.5	16.6	0	0
10	9	5.8	0	0
11	3.5	1.8	0	0
12	6.1	1.6	0	0
13	13.5	5.8	0	0
14	14.9	5	0	0

Table C.2: Line data

From	To	R(pu)	X(pu)	B/2	Tap
1	2	0.01938	0.05917	0.0264	1
1	5	0.05403	0.22304	0.0246	1
2	3	0.04699	0.19797	0.0219	1
2	4	0.05811	0.17632	0.017	1
2	5	0.05695	0.17388	0.0173	1
3	4	0.06701	0.17103	0.0064	1
4	5	0.01335	0.04211	0	1
4	7	0	0.20912	0	0.978
4	9	0	0.55618	0	0.969
5	6	0	0.25202	0	0.932
6	11	0.09498	0.1989	0	1
6	12	0.12291	0.25581	0	1
6	13	0.06615	0.13027	0	1
7	8	0	0.17615	0	1
7	9	0	0.11001	0	1
9	10	0.03181	0.0845	0	1
9	14	0.12711	0.27038	0	1
10	11	0.08205	0.19207	0	1
12	13	0.22092	0.19988	0	1
13	14	0.17093	0.34802	0	1

Appendix D

Theorems on Attack Strategy

Let, $\bar{H}$ is defined as $\bar{H} = R^{-1/2}H$, and $\bar{H}_i$ is the i^{th} row of $\bar{H}$

Theorem 1: Let ϵ be the threshold and $\sigma^2_{i=1,\ldots,3m}$ be the variance of errors in the $J(\hat{x})$ test. Given any set of measurements z, it is guaranteed to pass the $J(\hat{x})$ test when $\sum_{i=1}^{3m}(1-K_{ii})\sum_{j=1}^{3m}(z_j^2/\sigma_j^2) \leq \epsilon$.

Theorem 2: Suppose the original set of measurements z can bypass the $J(\hat{x})$ test. When the measurement z_i in z is perturbed into $z_i^{attacked}$ by the attacker, there always exists a new value $K_{ii}^{attacked} \in (K_{ii}, 1]$, such that the new measurement set $z^{attacked}$ is guaranteed to bypass the $J(\hat{x})$ test.

Theorem 3: Let K_{ii} be the i^{th} diagonal element of hat matrix K, then,

$$(1-K_{ii})^2 \leq \frac{\left\| \begin{bmatrix} \bar{H}_p \\ \bar{H}_f \end{bmatrix} \right\|_2^2}{\left\| \bar{H}_i^T \right\|_2^2}$$

where $\bar{H}$ is partitioned as: $\bar{H} = \begin{bmatrix} \bar{H}_p{}^T \bar{H}_i{}^T \bar{H}_f{}^T \end{bmatrix}^T$.

An attacker can increase the value of K_{ii} by just increasing the l_2-norm of $\bar{H}_i^T$. Since, $\bar{H}_i = 1/\sigma_i.H_i$, it gives rise to three rules

Rule 1: Increase the absolute values of elements in H_i.

Rule 2: Decrease the value of σ_i.

Rule 3: Increase the number of non-zero elements in H_i.

The proofs of the theorems are given in [61].

Therefore, there is a relationship between the measurement z_i and the corresponding K_{ii}. Let the attacked measurement be denoted by $z_i^{attacked}$ and the attacked corresponding diagonal element of the hat matrix be $K_{ii}^{attacked}$. Then, $\Delta K_{ii} = K_{ii}^{attacked} - K_{ii}$. Hence, the change in the value of K_{ii} reflects a change in the value of the corresponding z_i.

A smaller σ_i indicates a higher accuracy measurement. A higher accuracy measurement is more likely to become a leverage measurement and thus has a higher chance of getting attacked. From Theorem 2, it is clear that a small change in the value of K_{ii} can make the attack successful against measurements with larger value of K_{ii}. Hence, the leverage measurements are more susceptible to successful attacks.