Real-time Implementation of High Breakdown Point Estimators in Electric Power Systems via System Decomposition

by

Michael G. Cheniae

Dissertation submitted to the Faculty of the
Virginia Polytechnic Institute and State University
in partial fulfillment of the requirements for the degree of

DOCTOR OF PHILOSOPHY

in

Electrical Engineering

Approved:

Lamine Mili

Dr. L.M. Mili, Chairman

Dr. A.G. Phadke

Dr. C.W. Coakley

Dr. Y. Liu

Dr. R.Q. Huang

September, 1994

Blacksburg, Virginia

REAL-TIME IMPLEMENTATION OF HIGH BREAKDOWN POINT ESTIMATORS IN ELECTRIC POWER SYSTEMS VIA SYSTEM DECOMPOSITION

by

Michael G. Cheniae

Committee Chairman: L.M. Mili

Electrical Engineering

(ABSTRACT)

This dissertation presents a new, highly robust algorithm for electric power system state estimation. A graph theory-based system decomposition scheme is coupled with a high breakdown point estimator to allow reliable identification of multiple interacting bad data even in cases of conforming errors. The algorithm is inherently resistant to bad measurements in positions of leverage, makes no a priori measurement error probability distribution assumptions, and is applicable in a real-time environment.

In addition to presenting a new state estimation algorithm, the weaknesses of two prominent state determination methods are explored. The comparative advantages of high breakdown point estimators are then summarized. New theorems quantifying the previously unexamined effect system sparsity has on the exact fit point of some members of this estimator family are presented. These results serve as the catalyst for the overall state estimation algorithm presented. Numerous practical implementation issues are addressed with efficient implementation techniques described at each step.

Acknowledgements

I would like to thank Dr. L.M. Mili for his years of tutelage and financial support during my graduate studies. He initially introduced me to the subject of robust estimation and has continued to spark my interest in the field both in and out of the classroom. His dedicated approach to solving difficult research problems, in the face of sometimes discouraging results, serves as an excellent example for graduate students. Dr. C.W. Coakley has been an enthusiastic and helpful member of my Ph.D. committee. He is extremely amenable for bouncing off new ideas and has proven himself to be a very capable teacher. I wish him the best of luck in his career as a faculty member. I would also like to express my thanks to Dr. A.G. Phadke, Dr. Y. Liu, and Dr. R.Q. Huang for serving as members of my Ph.D. advisory committee.

This dissertation would not have been possible without the encouragement and support of my loving wife Mary. She has willingly made the necessary financial sacrifices and has, at times, placed some of her own studies on hold. To both of our surprise, she has also provided a new addition to our family. Our daughter, Clairisse, has been an additional challenge for two full-time students, but she has also provided us with many laughs and is truly a joy to behold.

Finally, I would like to thank my family and in-laws for their support. Although they all probably thought we were crazy to quit perfectly good jobs and return to school, they never let these thoughts be known. This ability to give their children the freedom to explore without instilling doubt is an admirable trait which Mary and I hope to mimic with our children.

Table of Contents

List of Tables

List of Illustrations

Chapter 1: Introduction

Modern electric power systems are continuously monitored by regional control centers. The goal of these centers is to ensure the efficient and reliable operation of their respective systems. An accurate knowledge of the state of the system is paramount if the control center operators are to make sound operational decisions. Providing this dependable system assessment is the role of a state estimator. A state estimator combines a system model with system measurement values sent to the control center and estimates the state of the system by minimizing an estimator's objective function.

The two most prominent estimators studied by the power community are the weighted least squares (WLS) and weighted least absolute value (WLAV) estimators. Whereas a great deal of research has dealt with improving the computing efficiency of methods based on these estimators, little attention has been paid to their statistical robustness in the presence of the actual types of errors common to power system measurements. As is shown in chapter 2, these errors cannot be realistically modeled by a short-tailed distribution function. Additionally, the presence of leverage points in power systems constitutes a further challenge to any estimator. These factors combine to make robust state estimation of power systems a very challenging problem, particularly when the relatively low measurement redundancy is considered.

With the actual nature of system errors taken into consideration, concepts from the field of robust estimation theory are used to illustrate the inherent dangers of the straightforward use of WLS and WLAV estimators in power systems. In particular, the WLS estimator's inability to identify multiple interacting outliers is noted. An improved method of identifying leverage points via a projection algorithm is then used to show that leverage points are quite common in power systems. It is an extension of the work of Vichare [52]. Having noted their prevalence, the potentially detrimental effects of

leverage points on the WLAV estimator are discussed with the interesting conclusion being reached that the LAV cannot be configured as a Schweppe-type estimator. This means that bad leverage points cannot be selectively downweighted based upon their residual values - good and bad leverage points are both necessarily downweighted. Although downweighting bad leverage points is beneficial, reduction of the estimate's accuracy results from the downweighting of good leverage points. More importantly, the breakdown point of the LAV can be reduced when extreme leverage points are essentially removed from the system due to very small measurement weights. Examples are used to illustrate these points, making clear the weaknesses of the LAV estimator in the presence of leverage points. This observation prompts the search for an estimator which possesses the inherent capability of handling leverage points.

An attractive alternative to the WLS and WLAV estimators is given by the family of high breakdown point estimators (HBPEs). This family of estimators possesses several desirable qualities: (i) they are insensitive to bad measurements in positions of leverage, and (ii) they feature the highest breakdown point an estimator can attain. This latter attribute is not based on any assumed measurement error distribution function. The major disadvantage is that they are computationally demanding. In fact, they exhibit an exponential growth in computing time relative to system size. This undesirable feature jeopardizes their serious consideration by the power community unless some means of overcoming this obstacle is found.

Previous work involving HBPEs in the statistical literature assumes the data (measurements) is in *general position* (any n rows of the m x n design matrix are linearly independent or, equivalently, any subset of n points observe the system). Such an assumption is invalid in power systems due to system sparsity. Prompted by this fact, new theorems are presented which describe the exact fit behavior of HBPEs when the data is in

reduced position (not in *general position*). These theorems generalize previous results, allowing one to make optimal use of HBPEs in linear regression problems with data in general or reduced position. More importantly, as far as this dissertation is concerned, they form the basis for a graph theory-based decomposition algorithm which, compared to a nondecomposed approach, yields a substantial increase in outlier identification capability and a simultaneous drastic decrease in computing time.

The proposed algorithm overcomes two major obstacles: (i) a vanishing breakdown point which results when applying a HBPE to a sparse system, and (ii) an exponential growth relationship between computing time and system size. Both of these results are obtained via a system decomposition scheme based on system topology. It is shown that by identifying two subsystem types, it is possible to substantially increase the outlier identification capability of a HBPE applied to a sparse system while simultaneously realizing a drastic decrease in required computing time. Many implementation considerations are analyzed with the goal of algorithm efficiency in mind. Capabilities and limitations of the algorithm are presented using IEEE test systems.

In addition to presenting a robust alternative state estimation algorithm, the dissertation considers a new approach to the overall power system state estimation problem. Instead of somewhat "blindly" applying estimators, it is recommended that one utilize system analysis tools to determine which estimator is best for various portions of the network. This allows the advantages of various estimators to be capitalized upon, while avoiding pitfalls that are easy to foresee if system characteristics and estimator behavior in these conditions are known. Overall, a more theoretically sound approach to the problem is encouraged.

This dissertation is organized as follows. Chapter 2 defines the power system state estimation problem in a statistical sense and delineates common sources of bad data or

outliers. The major weaknesses of the predominantly used estimators are presented and are illustrated by some examples. Earlier research involving the application of high breakdown point estimators is also covered. Chapter 3 presents some of the major concepts of robust estimation theory. For readers lacking a background in this area, it may be beneficial to review this section first since some terms common to the field are used throughout the dissertation. Chapter 4 presents the proposed algorithm in detail. Topics from the field of graph theory are used throughout the chapter and, for those readers lacking a familiarity with the subject, a short tutorial is available in Appendix B. Chapter 5 provides leverage point identification simulations as well as results using the proposed algorithm. The final chapter summarizes the work and suggests areas offering further research opportunities.

The study of robust estimation of power systems is valuable from both a theoretical and practical perspective. Assumptions traditionally made by statisticians when developing HBPEs do not hold for the power system model. This fact has prompted the furthering of statistical theory to account for previously unconsidered data structures. The power system model has also served to illustrate the concept of local breakdown, while the proposed decomposition algorithm contributes to the field of robust estimation by providing a means of increasing the vanishing breakdown point for sparse systems. These contributions are significant in that they are applicable to other problems involving structured regression. For these reasons, this dissertation should be of interest to persons interested in the general field of robust estimation, as well as to those specifically interested in power system state estimation.

Chapter 2: Overview of Power System State Estimation

2.1 Introduction

The state of an electric power system is defined by the voltage magnitude and phase angle at each bus. Knowledge of the state allows one to calculate the real and reactive power flow in each line - information that is critical for economic and security analysis and control. Determination of the system state is accomplished at a centralized control center where a computer system model is coupled with redundant measurement values which are telemetered to the center. The system model involves several assumptions, including balanced phases and perfectly known line parameters. In addition to errors in the system model, the measurement values involve uncertainties. The nature and source of these uncertainties will be explained shortly. Given these imprecise inputs, one uses an estimator to determine the "best" solution, i.e. one which minimizes an estimator's objective function. The "best" solution depends upon the estimator being used. Different estimators can give drastically different results for the same inputs. Concepts which allow one to explain these variations will be presented in Chapter 3. For now, it suffices to say that the nature of the measurement errors plays a major role in the variation among different estimators' solutions.

If a certain error probability distribution is assumed, one can choose an estimator which gives the most accurate solution under that assumption, i.e. the one with minimum variance. Such estimators are called Maximum Likelihood Estimators. Violation of the assumed distribution can seriously alter the solution of some of these estimators resulting in totally erroneous results. Since the actual measurement errors cannot be known, it is best to make no assumption concerning their error distribution and use an estimator whose viability is not inherently tied to it. Such an approach is taken in this dissertation.

2.2 Problem Formulation

2.2.1 Power System Overview

A power system consists of buses connected by transmission lines. A bus may have power injected into it by a generator or synchronous condensor and/or it may have power drawn from it by a load (a negative injection). Buses without any injected power are said to contain a zero injection. Figure 2.1 shows a one-line diagram of a 5-bus system.

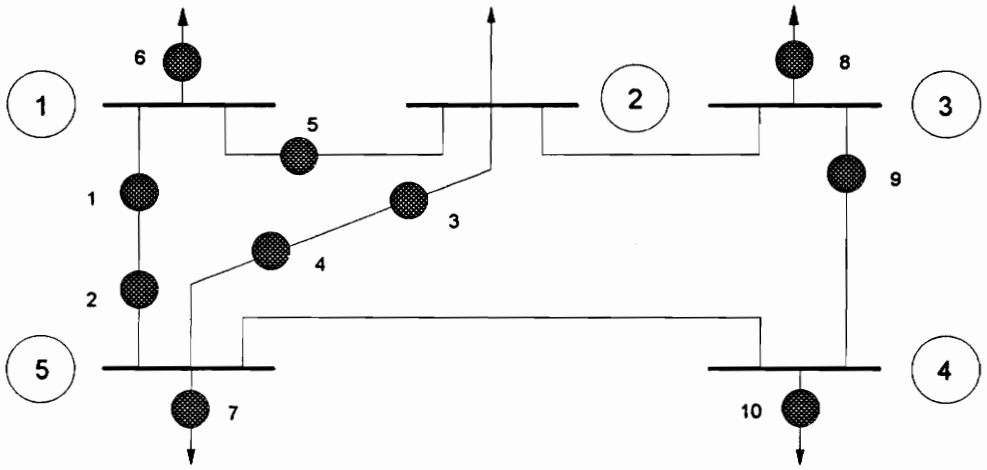


Figure 2.1 One-line Diagram of a 5-bus System

Transmission lines are typically represented by a π-equivalent model as shown in Figure 2.2, where the line connecting buses i and j is referred to as line (i, j). The *impedance* of a line, denoted by $\overline{Z}_{ij}$, is equal to the complex sum of the line resistance, R_{ij}, and the line reactance, X_{ij}, yielding

$$\overline{Z}_{ij} = R_{ij} + jX_{ij}. \tag{2.1}$$

The distributed capacitance of the line, B_{cap}, is divided in half and treated as two discrete shunt capacitors - one at each end of the line. The inverse of the line impedance is given by

$$\overline{Y}_{ij} = G_{ij} + jB_{ij}, \tag{2.2}$$

where $\overline{Y}_{ij}$ is the line *admittance*. The line *conductance*, G_{ij}, and the line *susceptance*, B_{ij}, are written as

$$G_{ij} = \frac{R_{ij}}{R_{ij}^2 + X_{ij}^2} \qquad B_{ij} = \frac{-X_{ij}}{R_{ij}^2 + X_{ij}^2}. \tag{2.3}$$

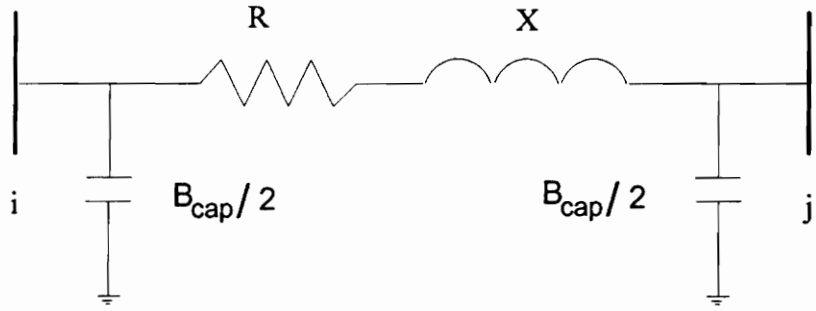


Figure 2.2 π-Equivalent Model of a Transmission Line

The voltage at bus *i* is a complex quantity which can be expressed in polar form as $V_i \angle \delta_i$. The bus voltage magnitude, which is a state variable, is often measured directly. An emerging technology developed at Virginia Tech based on Phasor Measurement Units (PMUs) [1] also allows the measurement of the voltage phase angles. Once the use of PMUs becomes more widespread, this valuable information can be incorporated into state estimation algorithms.

The real power flow from bus *i* to bus *j* on line (*i*, *j*) can be expressed as

$$P_{ij} = V_i^2 G_{ij} - V_i V_j (G_{ij} \cos\delta_{ij} + B_{ij} \sin\delta_{ij}), \tag{2.4}$$

where δ_{ij} is equal to $\delta_i - \delta_j$. The reactive power flow on line (i, j) can be expressed as

$$Q_{ij} = V_i V_j (B_{ij} \cos\delta_{ij} - G_{ij} \sin\delta_{ij}) - V_i^2 (B_{ij} + B_{cap}). \tag{2.5}$$

The real power injected at bus i is equal to

$$P_i = \sum_j P_{ij}. \tag{2.6}$$

A similar expression for the reactive power injected at a bus i is

$$Q_i = \sum_j Q_{ij}. \tag{2.7}$$

If the system measurements are expressed by the (m x 1) vector $\mathbf{z}$, and the state variables by the (m x $2n$ - 1) vector $\mathbf{x}$ (n is the number of system buses with one bus taken as the phase angle reference), the system can be modeled by

$$\mathbf{z} = h(\mathbf{x}) + \mathbf{e}. \tag{2.8}$$

Here $h(\mathbf{x})$ represents the set of equations (2.4 - 2.7) and $\mathbf{e}$ is an (m x 1) vector containing the measurement errors. Voltage magnitude measurements are also typically included in (2.8). Since the power flow equations are nonlinear, determining the state of a power system is a nonlinear estimation problem which is typically solved as a series of linearized

problems by expanding (2.8) using a first-order Taylor series. Each step of the series is then expressed as

$$\Delta \mathbf{z} = \mathbf{H}(\mathbf{x})\Delta \mathbf{x}, \tag{2.9}$$

where $\mathbf{H}(\mathbf{x})$ is the measurement Jacobian matrix. Viewed as such, the problem is one of multiple nonlinear regression without intercept.

A standard simplification to (2.9) is called the *decoupled model*. Use of this model will be made in later sections; so it is now briefly described. The Jacobian matrix associated with the real and reactive power measurements can be partitioned into four submatrices as follows:

$$\begin{bmatrix} \Delta \mathbf{P} \\ \Delta \mathbf{Q} \end{bmatrix} = \begin{bmatrix} \mathbf{H}_{P\delta} & \mathbf{H}_{PV} \\ \mathbf{H}_{Q\delta} & \mathbf{H}_{QV} \end{bmatrix} \begin{bmatrix} \Delta \delta \\ \Delta \mathbf{V} \end{bmatrix}. \tag{2.10}$$

For extremely high voltage (EHV) systems, the coupling matrices $\mathbf{H}_{PV}$ and $\mathbf{H}_{Q\delta}$ can be neglected. The conventional advantages which result are a decrease in the required computing time and memory allotment with a negligible loss of accuracy. For the proposed decomposition scheme, use of this model is also advantageous from a robustness point of view as will be explained in Section 5.

2.2.2 *Functions of a State Estimator*

Given the problem as described above, there are three major functions of a state estimator:

- provide an estimate of all metered and unmetered quantities;
- filter small errors;
- detect and identify discordant measurements or *outliers*.

As long as a system is observed by its measurements (i.e., $\mathbf{H}(\mathbf{x})$ is of full rank), any estimator can satisfy the first role. In fact, one can simply use a load flow to determine the system state and then solve for the line power flows and injections.

The presence of redundant measurements is required to justify the use of an estimator. Broadly speaking, increasing the level of redundancy increases the ability of an estimator to filter small errors and identify outliers. Unfortunately, estimators that are good at filtering small errors are relatively poor at identifying outliers, and vice versa. Since it makes no sense to worry about filtering small errors if large errors go undetected, it is necessary to first employ a robust estimator to identify the outliers and, then, use a more statistically efficient estimator to achieve an accurate estimate. These concepts will be described in detail in Chapter 3, but they are important to keep in mind as one considers the types of errors present in system measurements.

2.3 Sources of Outliers

In the introduction of this section it was stated that the nature of the measurement errors are of great concern since they are a determining factor when deciding upon an appropriate estimator. Before mentioning the specific types of measurement errors which are present in power system data, it will be helpful to describe the process by which measurement values are received at a control center. Figure 2.3 is a cursory schematic which shows the typical sequence of events necessary to receive an analog field measurement at the control center.

The first step is to use a current or voltage transformer (CT and VT respectively) to step-down the high-level current or voltage to levels which require less insulation and allow smaller, lightweight equipment (typically 0-5 amperes and 120 volts, respectively). These transformers introduce errors due to their inherent non-linear characteristics as

evidenced by their associated hysteresis curves. The transformer outputs are sent to transducers which convert the input values to 0-1 mA dc signals. Real (P) and reactive (Q) power are measured by feeding the transformer outputs into kW and kVAR transducers which each produce a ±1 mA dc signal proportional to their inputs. The transducers introduce errors due to miscalibration, and gain and zero offset drift with time and temperature [2, 3].

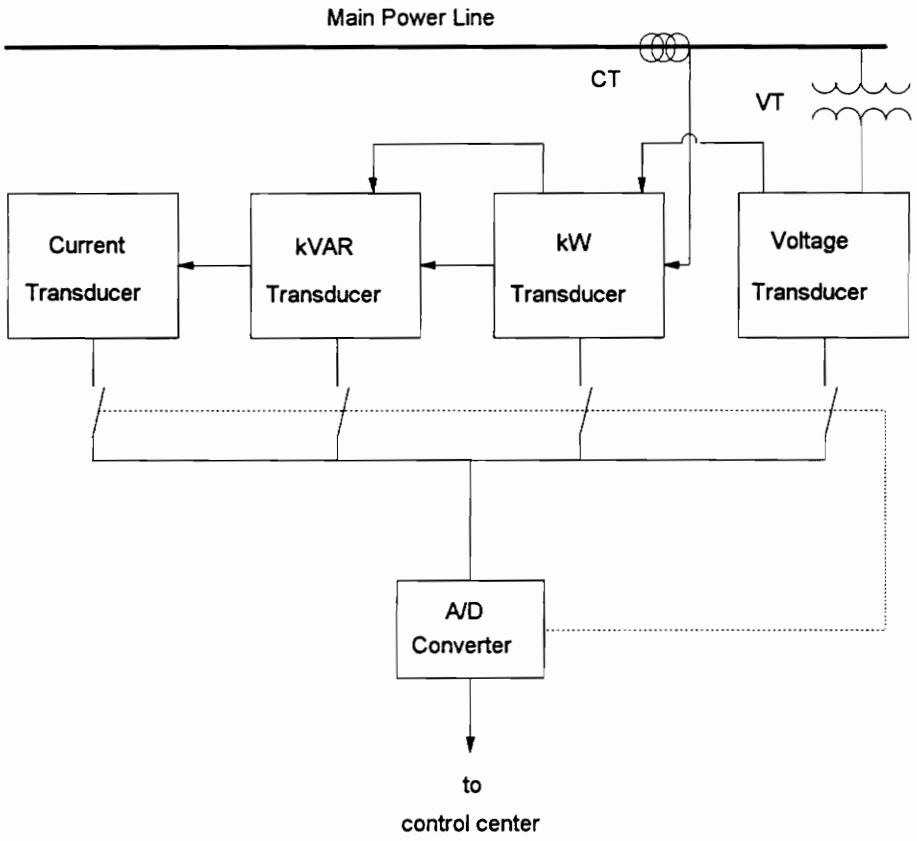


Figure 2.3 Power Line Measurements and Telemetry Schematic

The outputs of the transducers are then sampled, converted to digital form, and sent to the control center. Since the sampling cannot be done simultaneously, there is an error introduced which is referred to as timing skew. This is considered an error because the measurement set utilized by the control center is assumed to be a snapshot of the system at a particular instant and timing skew means this assumption is not completely true. Lastly, transmission of the measured values subjects them to corruption by communication channel noise, the degree of which depends upon the communication methodology and medium used.

The measurement errors described above can be grouped into three categories. First are the random errors which are related to the accuracy of the instrument. These are typically assumed to follow a gaussian distribution and are relatively small. The second type are called intermittent errors and are caused by device failures, spurious communication interference, and timing skew. The last class encompasses all other errors and are termed systematic errors. These errors completely dominate the random errors [3]. Figure 2.4 shows an actual example of systematic error effecting the voltages at a substation. Theoretically, the voltages at points T1, T2, and S1 - S4 should be equal since they are measured at the same potential. However, the cumulative effects of the error sources has caused an obvious bias in some, if not all, of the measurements. The conclusion to be drawn from this discussion is that all system measurements received at the control center are very likely biased to some extent. Those measurements with significant bias, be it due to intermittent or systematic errors, are called *outliers*.

Practical experience has shown that considering the potential presence of outliers is not an academic exercise. Van Slyck and Allemong [4] reported that over a 13 year period only 5 days existed for which no outliers were detected using conventional detection methods based on a wighted least squares estimator. They reported that an

average of 1-2% of the system measurements were deemed to be outliers. Note that these results are based on using a nonrobust estimator, so that the percentages are questionable. However, the conclusion that outliers are not uncommon remains valid.

More recently, Sierra Pacific Power reported that *nearly half* of the telemetered line MVAR readings in their system were backwards in sign [5]. In [6], a summary of errors encountered by Pacific Gas & Electric (PG&E) during the implementation of an Energy Management System (EMS) state estimator is presented. The major ones are summarized after Figure 2.4b.

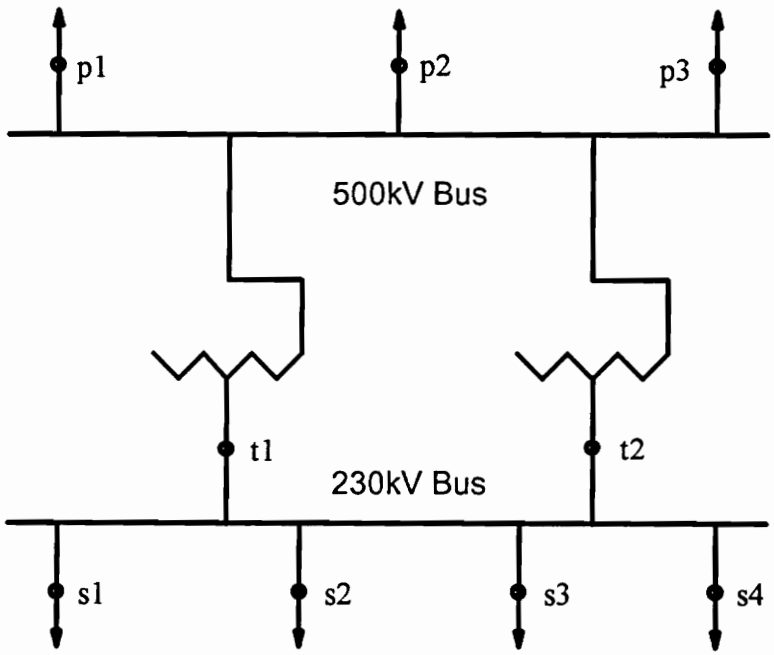


Figure 2.4a Substation Diagram

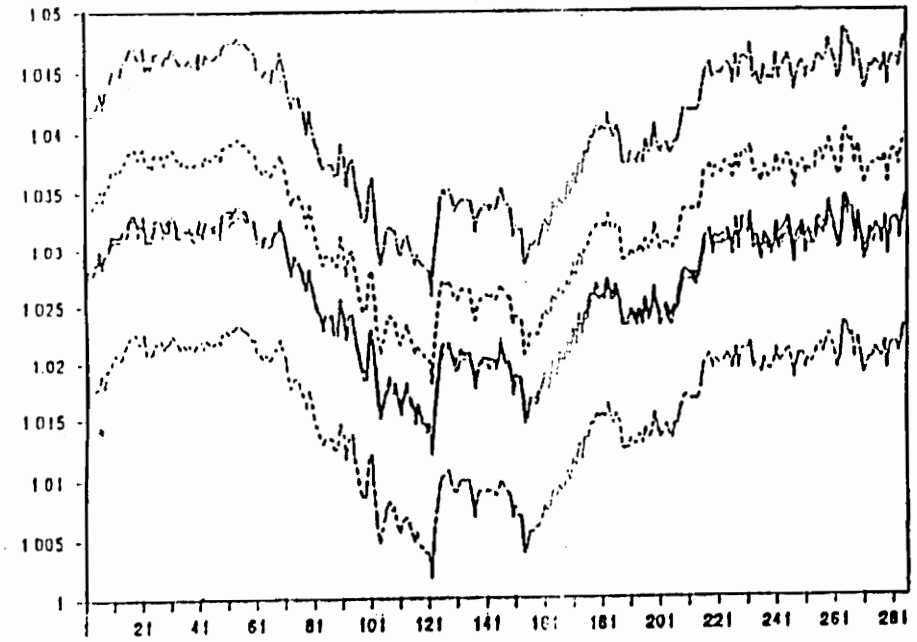


Figure 2.4b Per Unit Hourly Voltage Readings at Designated Substation Points on 230 kV Side

- **Analog Telemetry Errors:**

  - *reversed polarity conventions among neighboring utilities and within PG&E's application software* - inconsistent definition of positive power being into or out of a transformer;

  - *incorrect transducer full scale readings* - current and potential transformer tap settings are periodically updated for relaying purposes without updating the full scale readings of the EMS transducers;

  - *incorrect transducer installation* - (i) transducers are sometimes connected to normally deenergized substation auxiliary buses rather than the main bus, (ii) on

series compensated lines, the current transformer at the series capacitor side was used with the potential transformer at the transmission line side to measure the transmission line MW and MVAR flows, (iii) on Load Tap Changing (LTC) transformers, transformer low side potential was used to measure the high voltage side by multiplying the low side potential by the transformer's typical tap ratio (since LTC transformer taps change during system operation, assuming a typical ratio introduces error);

- *absent Qualifying Facility Remote Terminal Units (RTUs)* - RTUs not actually present in the field were modeled in the EMS, resulting in totally erroneous measurement values.

- **System Parameter Errors:**

  - *inconsistent transformer tap numbering system between the EMS model and actual field devices* - Most transformer tap settings were not telemetered to the control center. Seasonal field tap changes were often not reported to the EMS group. Tap numbering in the EMS model ranged from -16 to +16, while the field markings ranged from 0 to 32. This led to an inconsistency between the meaning of field and model markings even when tap changes were reported.

  - *incorrect transformer impedance modeling.*

Errors introduced by human incompetence do not follow a simple or predictable distribution; nor do those caused by significant intermittent or systematic bias. Given that there are a nontrivial number of sources of outliers, and that the resulting measurement error distribution is unlikely to follow a simple pattern, it is advisable to use an estimator whose robustness is not linked to an assumed distribution.

2.4 The Weighted Least Squares Estimator

The weighted least squares estimator (WLS) forms the basis for most power system state estimator routines currently in use. First utilized for power system state estimation by Schweppe, Wildes and Rom [7-9] in 1970, the WLS estimates the system state by minimizing the sum of the squared residuals, which in matrix form is written as

$$\min_{\mathbf{x}} J(\mathbf{x}) = \left[\mathbf{z} - h(\mathbf{x})\right]^{\mathrm{T}} \left[\mathbf{R}^{-1}\right] \left[\mathbf{z} - h(\mathbf{x})\right]. \tag{2.11}$$

Here $\mathbf{R}$ denotes the (m x m) covariance matrix of measurement errors, assumed to be independent, and is equal to

$$\mathbf{R} = \begin{bmatrix} \sigma_1^2 & & \mathbf{0} \\ & \ddots & \\ \mathbf{0} & & \sigma_m^2 \end{bmatrix}. \tag{2.12}$$

The assumed variance associated with the ith measurement is denoted by σ_i^2. The WLS is initially attractive for several reasons. First, when the regression model is linear, a closed form solution exists. For the nonlinear model given by (2.8), we solve (2.11) iteratively, for instance throught the Gauss-Newton algorithm, yielding

$$\hat{\mathbf{x}} = \left[\mathbf{H}(\mathbf{x})^{\mathrm{T}}\mathbf{R}^{-1}\mathbf{H}(\mathbf{x})\right]^{-1}\mathbf{H}(\mathbf{x})^{\mathrm{T}}\mathbf{R}^{-1}\left(\mathbf{z}-\mathbf{h}(\mathbf{x})\right). \tag{2.13}$$

Second, if the measurement errors are actually independently distributed random variables following a gaussian distribution with zero mean, the WLS is the optimal estimator in the sense that it will give the most accurate solution - the variance of its estimate is smaller than any other estimator's. In statistical parlance, it is the Maximum Likelihood Estimator for normally distributed errors. If the errors do not satisfy this assumption, especially if the tails of the actual distribution are longer, the WLS's estimate can be altered dramatically. This matter will be discussed further in Chapter 3.

Outlier detection and identification with the WLS can be accomplished using the weighted residuals

$$r_{w_i} = \frac{z_i - h_i(\hat{\mathbf{x}})}{\sigma_i} \tag{2.14}$$

or, preferably, the normalized residuals [10]

$$r_{N_i} = \frac{z_i - h_i(\hat{\mathbf{x}})}{\sigma_i w_{ii}}, \tag{2.15}$$

where $w_{ii} = \sqrt{\mathrm{diag}_{ii}(\mathbf{W})}$ and $\mathbf{W}$ is the residual sensitivity matrix given by

$$\mathbf{W} = \mathbf{I} - \mathbf{R}^{-1}\mathbf{H}(\mathbf{x})\left[\mathbf{H}(\mathbf{x})^{\mathrm{T}}\mathbf{R}^{-1}\mathbf{H}(\mathbf{x})\right]^{-1}\mathbf{H}(\mathbf{x})^{\mathrm{T}}\mathbf{R}^{-1}. \tag{2.16}$$

Handschin, Schweppe, Kohlas, and Fiechter [10] showed that the normalized residuals possess a standard normal distribution when the measurement errors are gaussian, and that

when there is only one measurement in error and the remaining measurements are exact (free from any error), the outlying measurement will possess the largest normalized residual. Normalized residuals are also fairly reliable diagnostic tools when there are multiple noninteracting outliers. Two measurements i and j are noninteracting if an error in measurement i has little effect on r_{N_j} and vice versa. An assessment of such interaction between measurement errors can be made by examining the entries of the residual sensitivity matrix or the matrix of residual correlation coefficients (**K**) [11] given by

$$\mathbf{K} = \mathrm{diag}(\mathbf{W})^{-1/2}\,\mathbf{W}\,\mathrm{diag}(\mathbf{W})^{-1/2}. \tag{2.17}$$

Whether **W** or **K** is used, the i, jth entry of the matrix is compared to a threshold to decide if the measurements are interacting.

Since the WLS residuals are linear combinations of the measurement errors, there is not necessarily a one-to-one correspondence between outliers and large normalized residuals. As such, outlier identification methods based on these residuals are incapable of reliably identifying multiple interacting outliers. Errors associated with multiple interacting outliers which essentially cancel one another are said to be conforming. The presence of such errors cannot even be *detected* by WLS residuals and, therefore, the resulting outliers are not identifiable. This holds true for even the most sophisticated outlier identification methods based on the WLS normalized residuals - the two most widely known being the combinatorial optimization method of Monticelli, Wu, and Yen [12] and the Hypothesis Testing and Identification method of Mili, Van Cutsem, and Ribbens-Pavella [13]

A final comment concerning the assumed measurement errors is in order. As its name implies, the WLS solution places different emphasis on measurements based upon

their weights, $1/\sigma_i^2$. These weights are the reciprocal of the assumed variances which are derived from the equipment manufacturer's specifications. By placing greater weight on certain measurements one is implicitly assuming an *a priori* knowledge of the "best" measurements. But identifying outliers and, hence identifying the "best" measurements, is one of the primary functions of a state estimator. To influence the estimate with unwarranted assumptions makes no sense. An excellent example of the sensitivity of methods based on the WLS to assumed weights is found in the closure of [14] where a two bus, single line system is studied. A flow measurement is located at each end of the line and an injection is measured at one of the buses. The results of numerous simulations in which one measurement was made an outlier while a variety of weights were given to the measurements are reported. As one might expect, if the measurement assumed to be the most accurate (the greatest weight) is the outlier, methods based upon the WLS often fail. Attempting to find a suitable set of measurement weights is a hopeless cause. Initial incorrect weights can lead to a solution in which no outliers are detected or are misidentified. Basing subsequent weights on these erroneous results [15] only helps to perpetuate these results. For these reasons, the algorithm presented in Chapter 4 does not weight measurements - it lets a highly robust estimator do the job of identifying outliers. It is worth noting at this point that the estimator used in this dissertation guarantees correct identification of a single outlier in the two bus case studied in [14].

2.5 The Weighted Least Absolute Value Estimator

The other major estimator receiving increasing attention in the power system community is the weighted least absolute value estimator (WLAV). This estimator minimizes the sum of the absolute value of the residuals instead of the sum of the squared

residuals as is the case with the least squares estimator. The argument leading to the introduction of the WLAV is that the least squares estimator is overly sensitive to outliers because the residuals are squared.

Whereas the WLS is the Maximum Likelihood Estimator for normally distributed errors, the WLAV is the Maximum Likelihood Estimator for data following a Laplacian or double-exponential probability distribution. This distribution has longer tails than the Gaussian; so it follows that this estimator is more robust than the WLS. However, the WLAV is a member of the class of M-estimators and is, therefore, extremely sensitive to measurements in positions of leverage. Such measurements are common in power systems and are associated with flows on (and injections incident to) relatively short lines and injections located on buses with a relatively large number of incident lines [16].

Noting this weakness, Celik and Abur [17] propose transforming the measurement Jacobian matrix so that leverage points are downweighted - thereby making the WLAV a *generalized M-estimator* (GM-estimator or bounded-influence estimator). Celik and Abur refer to this estimator as the WLAV-T and propose obtaining an appropriate transformation via a heuristic approach. An alternative method of rapidly identifying leverage points and computing appropriate row scaling factors via *projection statistics* is presented in [18]. A variation of this method is introduced in [19] and in Section 3.4 of this dissertation. In any case, downweighting all leverage points has its disadvantages. Good leverage points (i.e. measurements in positions of leverage whose observed values are accurate) are extremely beneficial in that they tend to improve the precision of the state estimate. Ideally, one would like to downweight only the bad leverage points. It is now shown that such selective downweighting is not possible for the WLAV/WLAV-T. Some material covered in Chapter 3 is repeated here for the reader's convenience.

For ease of discussion and without loss of generality, all of the measurement weights are set equal to unity. Since the weights are not applicable, the WLAV will be refered to as the LAV. The LAV objective function minimizes the sum of the absolute value of the residuals, i.e.

$$J(\mathbf{x}) = \sum_i |r_i| . \tag{2.18}$$

The function of the residuals that is being minimized, $|r_i|$, is an example of many possible such functions. Estimators which minimize some function of the residuals are called M-estimators and the function in question is denoted $\rho(r_i)$. For the LAV, a plot of the function is shown in Figure 2.5.

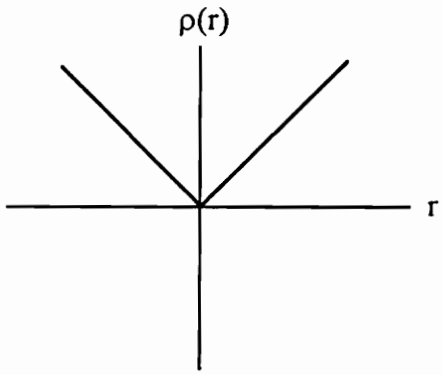


Figure 2.5 ρ(r) Function for LAV

Although in practice the LAV solution is usually computed via linear programming techniques, the solution can theorectically be found by defining the *score function* as $\psi(r) = d\rho(r)/dr$, and solving

$$\sum_i \psi(r_i/\hat{\sigma}) = 0. \tag{2.19}$$

The factor, $\hat{\sigma}$, is an estimate of the residuals' spread. From Figure 2.5 it follows that, for the LAV,

$$\psi(r_i/\hat{\sigma}) = sign(r_i/\hat{\sigma}). \tag{2.20}$$

A plot of the *sign* function is depicted in Figure 2.6.

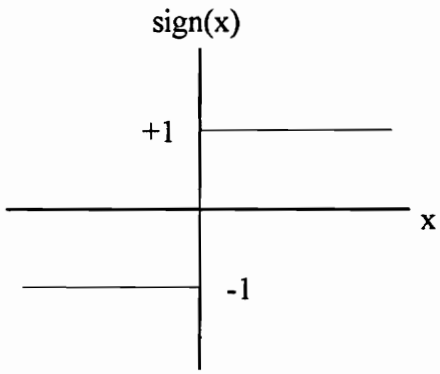


Figure 2.6 Sign(x) Function

Assuming one has obtained appropriate measurement weights, $w(\mathbf{x}_i)$, based on each measurement's outlyingness in the factor space, there are two methods to downweight the effects of leverage points in (2.19). The Mallow's method, downweights all leverage points regardless of whether they are good or bad, yielding

$$\sum_i w(\mathbf{x}_i)\psi(r_i/\hat{\sigma})\mathbf{x}_i = 0. \tag{2.21}$$

The Schweppe method strives to downweight only bad leverage points. This is accomplished by downweighting only those leverage points with large residuals, i.e.

$$\sum_i w(\mathbf{x}_i)\psi(r_i/\hat{\sigma}w(\mathbf{x}_i))\mathbf{x}_i = 0. \tag{2.22}$$

Since the score function associated with the LAV is the *sign* function, expressions (2.21) and (2.22) are identical. Applying the measurement weights to the residuals has no effect on the LAV solution. It is not possible to selectively downweight leverage points based on their residual values - good and bad leverage points are all downweighted.

Two simple examples in two dimensions serve to illustrate the unpredictable results of simply downweighting all leverage points. First, the benefits of downweighting a bad leverage point is shown. Figure 2.7 is a plot of three points fitting a line with slope equal to one with a zero intercept. A fourth point is an outlier in a position of leverage. Without application of measurement weights, the LAV solution gives a slope of 2.67 and an intercept term of -3.33. This solution is indicated by the dashed line. The leverage point clearly has a large effect on the estimate. With measurement weights computed via the method of projection statistics discussed in Section 3.4, the correct solution indicated by the solid line is found. The measurement weights used are listed in Table 2.1 which precedes the figure.

Table 2.1 Robust Measurement Weights for First LAV Example

Measurement Number	Measurement Weight
1	1.0
2	1.0
3	1.0
4	0.092

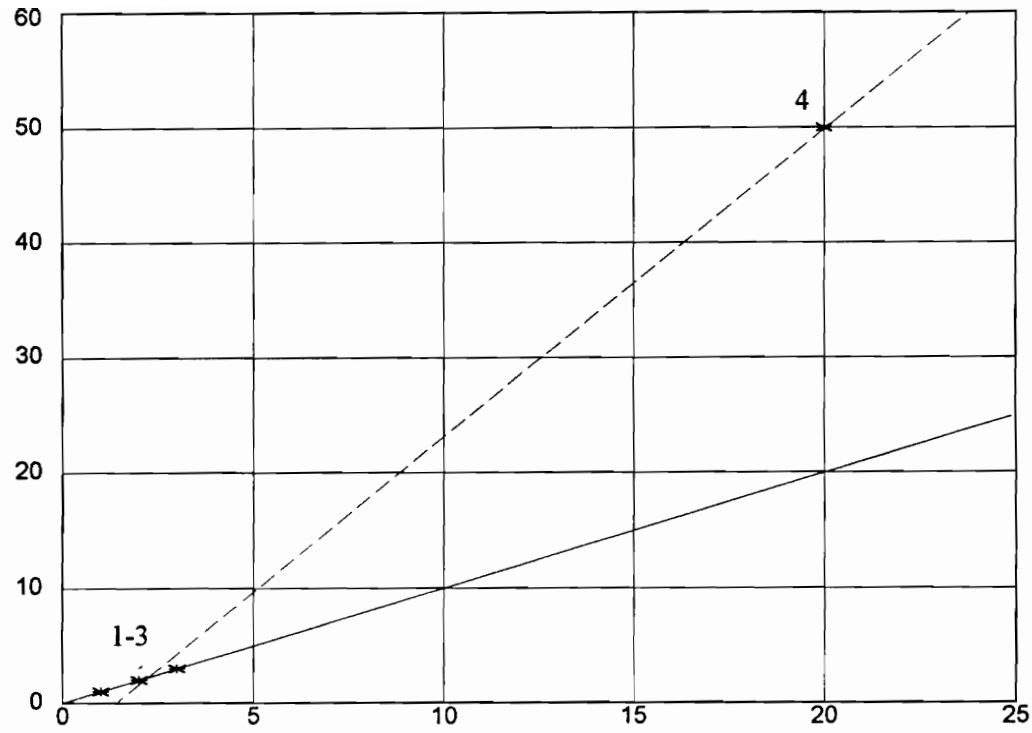


Figure 2.7 Illustration of Benefits of Downweighting a Bad Leverage Point

The second example's purpose is to show how downweighting a good leverage point can be disadvantageous. Figure 2.8 shows four points fitting the line with unity slope and zero intercept. Point six is a good leverage point. Points four and five are both bad leverage points. Without application of measurement weights, the correct solution is found by the LAV as depicted by the solid line in the figure. Application of the measurement weights indicated in Table 2.2 result in the dashed line whose parameters are a slope of -0.055 and an intercept of 2.11. In this case, downweighting the good leverage point removes its beneficial effect on the estimate and allows more mild bad leverage points to drive the solution.

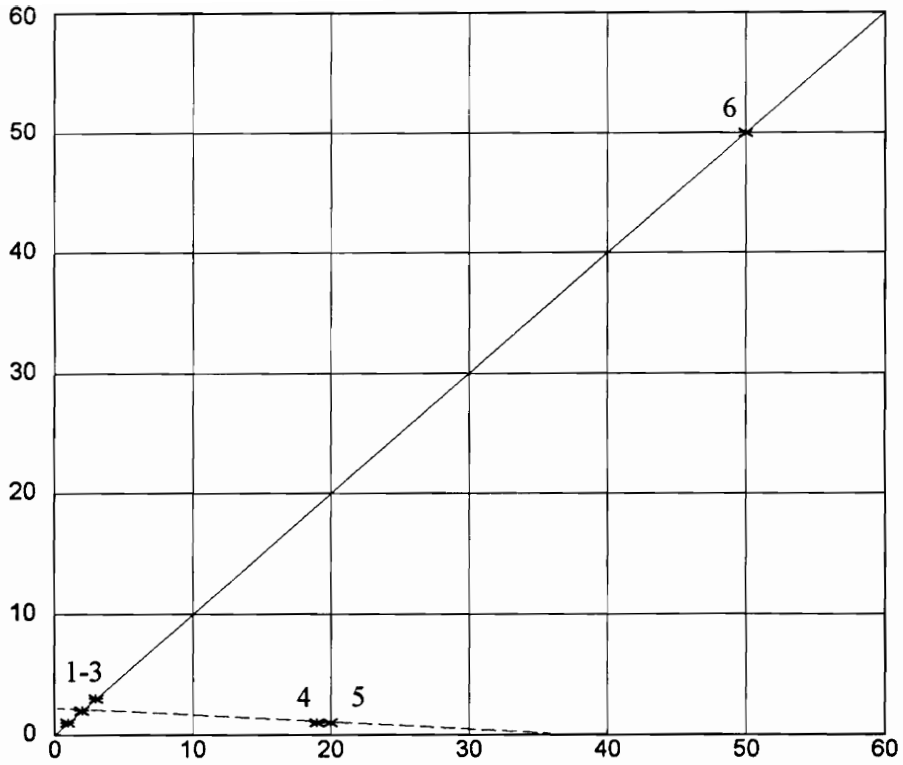


Figure 2.8 Illustration of Drawbacks of Downweighting a Good Leverage Point

Table 2.2 Robust Measurement Weights for Second LAV Example

Measurement Number	Measurement Weight
1	1.0
2	1.0
3	1.0
4	1.0
5	1.0
6	0.276

The phenomenon can also be seen using a linearized version of the IEEE-14 bus system with the measurement configuration shown in Figure 2.9. The measurement weights as determined using projection statistics are given in Table 2.3. All line parameters are those for the IEEE 14-Bus test system with the exception that lines 6-13 and 1-2 are shortened by a factor of ten. Those measurements with a weight less than one are leverage points. As expected, the measurements with the smallest weights are thosed associated with the shortened lines. The true measurement values are all zero, whereas the observed values are given in Table 2.3. Measurements with weights less than one and observed values of zero are good leverage points. Without application of the measurement weights, all of the outliers are correctly identified as seen by the first column of residuals listed in the table. With the weights applied, the LAV breaks down as evidenced by the second column of residuals.

This power system example reinforces the idea that indiscrimate downweighting of leverage points is a bad practice. Without downweighting, the exaggerated influence of the good leverage point tends to force the LAV solution to pass through it, despite the nearby presence of other outliers. This beneficial effect is lost with downweighting. For this reason, the WLAV-T estimator is not a panacea.

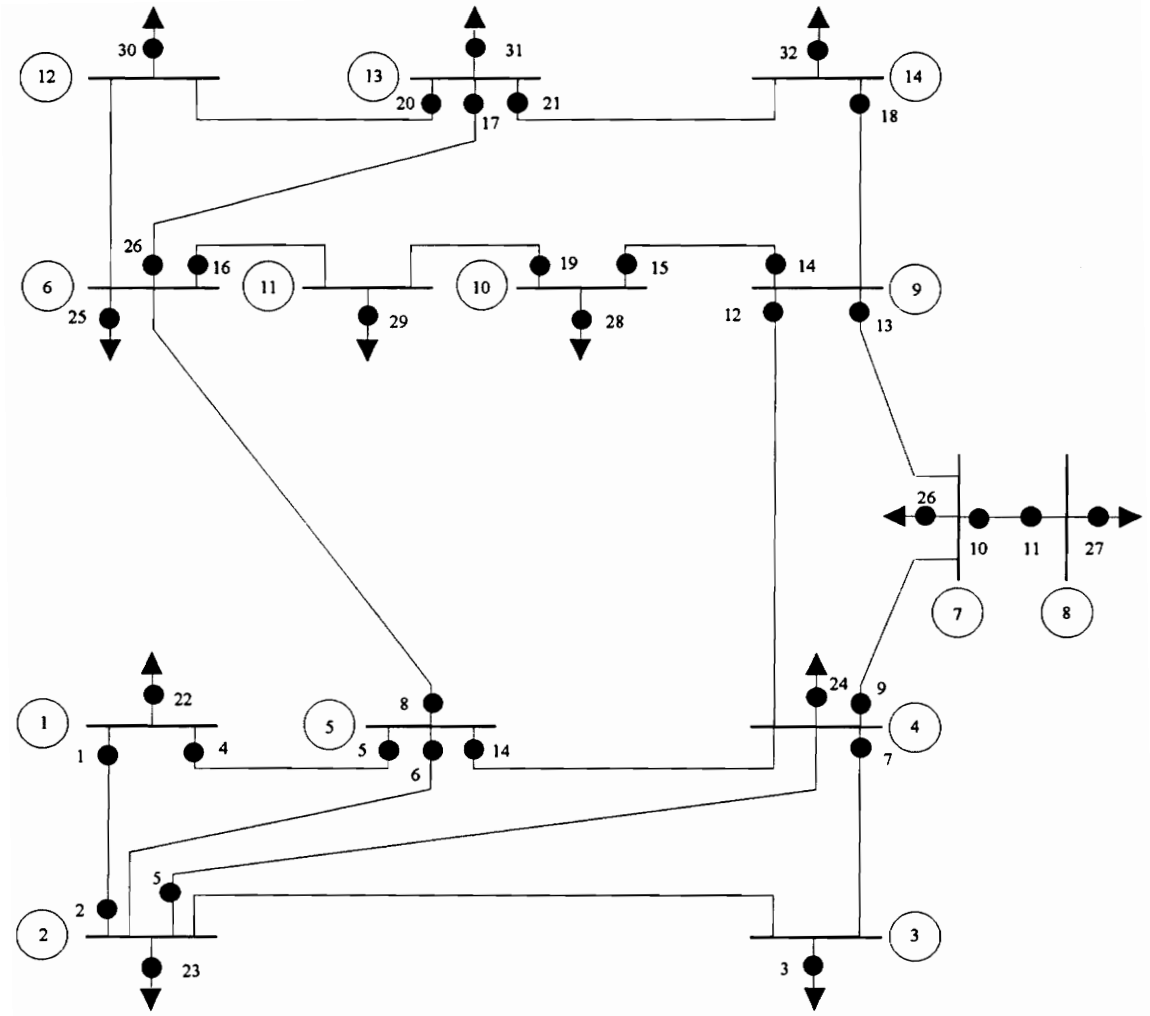


Figure 2.9 IEEE 14-Bus Test System

Table 2.3 LAV Results With/Without Application of Robust Measurement Weights

Measurement Number	Measurement Weight	Actual Value	Measured Value	Residual	
				Without Weights	With Weights
1	0.051	0.00	0.00	0.00	6.33E-07
2	0.051	0.00	0.00	0.00	-6.33E-07
3	1.000	0.00	100.00	100.00	0.00
4	1.000	0.00	0.00	0.00	0.00
5	1.000	0.00	0.00	0.00	0.00
6	1.000	0.00	0.00	0.00	4.26E-07
7	1.000	0.00	0.00	0.00	47.12
8	1.000	0.00	0.00	0.00	7.01
9	1.000	0.00	0.00	0.00	-4.77E-07
10	1.000	0.00	0.00	0.00	0.00
11	1.000	0.00	0.00	0.00	0.00
12	1.000	0.00	0.00	0.00	1.19E-15
13	1.000	0.00	0.00	0.00	0.00
14	1.000	0.00	0.00	0.00	0.00
15	1.000	0.00	0.00	0.00	0.00
16	1.000	0.00	0.00	0.00	0.00
17	0.030	0.00	0.00	0.00	-8.84E-15
18	1.000	0.00	0.00	0.00	6.05E-23
19	1.000	0.00	0.00	0.00	6.96E-16
20	1.000	0.00	0.00	0.00	-1.73E-07
21	1.000	0.00	0.00	0.00	0.00
22	0.079	0.00	0.00	0.00	0.00
23	0.138	0.00	0.00	0.00	8.54
24	1.000	0.00	0.00	0.00	0.00
25	0.075	0.00	0.00	0.00	-0.528
26	1.000	0.00	0.00	0.00	0.00
27	1.000	0.00	0.00	0.00	0.00
28	1.000	0.00	0.00	0.00	-9.65E-16
29	1.000	0.00	0.00	0.00	0.00
30	1.000	0.00	0.00	0.00	0.00
31	0.065	0.00	0.00	0.00	-1.26E-08
32	1.000	0.00	0.00	0.00	0.00

Another concern with the WLAV-T is its breakdown point when applied to sparse systems. For regression problems involving data in general position, the breakdown point of any GM-estimator (e.g. the WLAV-T) is inversely proportional to the number of unknowns [20]. There is some conjecture [17] that the sparsity of power systems will

result in the breakdown point of the WLAV-T being insensitive to system size due to the local nature of the estimation model. There is not yet any theoretical support for this claim and a simple example [21] is presented which questions its validity.

Consider the 5-bus system shown previously in Figure 2.1. The transposed measurement Jacobian matrix for the DC model of the system with all line resistances equal to zero and reactances equal to 1.0 p.u. is

$$\mathbf{H}^T = \begin{bmatrix} 1 & -1 & 0 & 0 & 1 & 2 & -1 & 0 & 0 & 0 \\ 0 & 0 & 1 & -1 & -1 & -1 & -1 & -1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 2 & 1 & -1 \\ 0 & 0 & 0 & 0 & 0 & 0 & -1 & -1 & -1 & 2 \end{bmatrix}. \tag{2.23}$$

There are no leverage points present so that no transformation is necessary. Let z_{10} be an outlier with a value of 30. All the other measurement values are zero - the true values. With bus #5 taken as the reference bus ($\theta_5 = 0$), the goal is to estimate the voltage phase angles for buses 1 through four. Note that the true solution is $\theta_i = 0$ for i = 1,...,5. The WLAV solution to the problem is not unique, i.e. multiple solutions exist. One of the possible solutions, in fact the one resulting from the solution method of Barrodale and Roberts [22], is

$$\theta^T = [0 \quad 0 \quad 10 \quad 20 \quad 0], \tag{2.24}$$

resulting in the following absolute residuals:

$$\mathbf{r}^T = [0 \quad 0 \quad 0 \quad 0 \quad 0 \quad 0 \quad 20 \quad 0 \quad 10 \quad 0]. \tag{2.25}$$

Clearly, this is an incorrect solution and it is not due to the effects of leverage. The WLAV exhibits a breakdown point of zero for a *very* small system - below the maximum a GM-estimator could attain in this case of 25%. Thus, the conjecture regarding the benefits of system sparsity on the WLAV or WLAV-T estimators is questionable. The same comments concerning the choice of measurement weights made during the WLS discussion apply to the WLAV/WLAV-T.

2.6 Previous Use of High Breakdown Point Estimators

Earlier application of HBPEs to power systems [16] was done without an understanding of the local breakdown phenomenon, although the existence of the phenomenon was noted [23]. The research was primarily concerned with decreasing the computing time necessary to implement the LMS estimator. Sparsity programming techniques were used with the majority of the work being done off-line. Actual computation of the LMS estimate was accomplished via the resampling algorithm used by Rousseeuw in his program PROGRESS [24]. The number of random samples selected was based on the same criteria as in PROGRESS without considering the effects of system sparsity. Even with these techniques, the need for system decomposition was apparent. However, without an understanding of local breakdown, the proposed scheme was not effective in increasing the estimator's breakdown point. As is now known, the breakdown point of LMS when applied directly vanishes for large power systems.

A major contribution of the pioneering work concerning the application of HBPEs was advertising the fact that the WLS and WLAV estimators suffer some serious drawbacks and that HBPEs were an alternative worth investigating. Identifying the presence of leverage points in power systems was also significant since it offered an explanation for significant outlier detection failure rates of WLS- and WLAV-based

methods. Overall, this work did not offer a viable algorithm, but it did start a fundamental reassessment of the power system state estimation problem and introduced the power community to the field of robust statistics. Additional comments regarding this earlier work can be found in Chapter 4.

Chapter 3: Estimation Theory

3.1 Introduction

To discuss the intricacies of the state estimation problem, a familiarity with the fundamentals of estimation theory is necessary. This chapter provides a sufficiently detailed overview of the field to allow quantitative comparisons of various estimators. It is not intended to be an all-encompassing review of robust estimation theory, but it does make clear the necessity of robust estimators, highlights some of the pertinent historical developments in the field, and presents a class of revised HBPEs on which the proposed system decomposition algorithm is based.

3.2 Parametric Estimation Theory

Parametric statistics assumes that the observation errors are independent and identically distributed (exactly) according to a parametric model, i.e. a probability density function. Given this assumption, Fisher [25] derived the class of Maximum Likelihood Estimators (MLEs) which are optimal in the sense that their estimate has minimum variance. As noted by Tukey in [26]:

> "A tacit hope in ignoring deviations from ideal models was that they would not matter; that statistical procedures which were optimal under the strict model would still be approximately optimal under the approximate model. Unfortunately, it turned out that this hope was often drastically wrong; even mild deviations often have much larger effects than were anticipated by most statisticians".

Specifically, Tukey (1960) considered the following example. Assume we have a large, randomly mixed batch of good and bad observations each denoted by x_i. Their total number is n and each good observation occurs with a probability (1 - ε), while each bad observation occurs with probability ε, where ε is a small number called the fraction of contamination. Let the good observations follow a $N(\mu, \sigma^2)$ distribution and the bad ones be distributed according to $N(\mu, 9\sigma^2)$. This means that all the observations have the same mean but the bad observations have a three-fold error compared to the good ones. One can equivalently write

$$F(x) = (1-\varepsilon)\Phi\left(\frac{x-\mu}{\sigma}\right) + \varepsilon\Phi\left(\frac{x-\mu}{3\sigma}\right), \tag{3.1}$$

where

$$\Phi(x) = \frac{1}{\sqrt{2\pi}}\int_{-\infty}^{x} e^{-y^2/2}dy \tag{3.2}$$

is the standard normal cumulative distribution function. Equation 3.1 represents a small departure from the idealized model in which all the observations are good.

Tukey compared two estimates of scale - the mean absolute deviation

$$d_n = \frac{1}{n}\sum|x_i - \bar{x}| \tag{3.3}$$

and the standard deviation

$$s_n = \left[\frac{1}{n}\sum(x_i - \bar{x})^2\right]^{1/2}. \tag{3.4}$$

Classical asymptotic optimality theory states that s_n is the optimal estimator for normally distributed data because its estimate converges to σ. We are interested in its performance relative to a non-optimal estimator, d_n, when the data follows a slightly modified distribution. To allow a relative comparison, the asymptotic relative efficiency (ARE) of d_n relative to s_n is defined as

$$\mathrm{ARE}(\varepsilon) = \lim_{n\to\infty} \frac{\mathrm{var}(s_n)/\left(E[s_n]\right)^2}{\mathrm{var}(d_n)/\left(E[d_n]\right)^2}, \tag{3.5}$$

which for this example gives

$$\mathrm{ARE}(\varepsilon) = \frac{\left[\dfrac{3(1+80\varepsilon)}{(1+8\varepsilon)^2} - 1\right]\Big/4}{\dfrac{\pi(1+8\varepsilon)}{2(1+2\varepsilon)^2} - 1}. \tag{3.6}$$

Table 3.1 (taken from [27]) gives the results of the asymptotic relative efficiency for a range of fraction of contamination. It suffices to take ε = 0.0018 to make the mean deviation more efficient than the standard deviation. The table also shows that d_n is better than s_n all the way up to ε = 0.5. Thus, even a slight lengthening of the tails of the underlying distribution has a dramatic effect on the variance of s_n. Using the same example, Tukey also showed that the asymptotic efficiency of the mean (the asymptotically optimal location estimator for the normal distribution) quickly loses its optimality when compared to other estimators of location.

Table 3.1 Asymptotic Efficiency of the Mean Absolute Deviation Relative to the Standard Deviation

ε	ARE(ε)
0	0.876
0.001	0.948
0.002	1.016
0.005	1.198
0.01	1.439
0.02	1.752
0.05	2.035
0.10	1.903
0.15	1.689
0.25	1.371
0.5	1.017
1.0	0.876

For a small deviation from the assumed model, the MLEs associated with the model (s_n in the scale case and the mean in the location case) quickly lose their optimality. The extreme sensitivity of the MLEs to even minor violations of the underlying assumption prompted the development of robust statistics.

3.3 Robust Estimation

3.3.1 The M- and GM-Estimators

The basic idea behind robust statistics is to grant the assumption that a parametric model which describes the majority of the data does exist, but that not all of the data fits the model exactly. Given this framework, Huber [28] notes several desirable characteristics any statistical procedure should possess. They are as follows:

- The procedure should have a reasonably good efficiency at the assumed model;
- Small deviations from the assumed model should have a minor effect on the outcome of the procedure;
- Larger deviations from the assumptions should not cause a catastrophe.

With these ideas in mind, Huber [29] generalized the class of MLEs into what are called M-estimators (M as in Maximum likelihood-type). These estimators minimize an objective function given by

$$J(\theta) = \sum_{i=1}^{n} \rho(r_i), \tag{3.7}$$

where ρ is a symmetric function with a unique minimum at zero and r_i denotes the ith residual. These estimators are solved by letting $\psi(r) = d\rho(r)/dr$ and setting

$$\sum_{i=1}^{n} \psi(r_i) = 0. \tag{3.8}$$

If there exists a probability density function $f(r)$ ($f(r)$ is termed the target distribution) such that $\rho(r) = -\ln(f(r))$, the M-estimator is the MLE associated with $f(r)$. However, no error

probability density function is assumed or necessary, i.e. an M-estimator is not necessarily a MLE.

Before discussing a major drawback of M-estimators it is necessary to define two new terms - breakdown point and leverage points. Breakdown point is considered first. Let T be a regression estimator and let θ' denote the vector of regression coefficients obtained by applying T to a set of data points represented by $\mathbf{z}$. If the true regression coefficients are denoted by θ, the bias of the estimator T is given by

$$\text{bias}(T,\mathbf{z}) = \|\theta - \theta'\| \tag{3.9}$$

The *global robustness* of an estimator is determined by the maximum fraction of data that can be made arbitrarily large without causing the estimator's maximum bias to become unbounded. This fraction is called the *breakdown point* and is denoted by ε^*. Hodges [30] and Hampel [31] first introduced this concept using the functional definition of an estimator. A more easily understood and practical definition due to Donoho and Huber [32] is presented here.

Let $\mathbf{z} = \{z_1,\ldots,z_m\}$ be a good data set and $\mathbf{z}'$ be a data set derived from $\mathbf{z}$ by replacing any f of the m measurements with arbitrary values. Applying the estimator T to $\mathbf{z}$ will result in θ since these points are all good, while applying the estimator T to $\mathbf{z}'$ results in θ'. With the maximum bias expressed as

$$\text{bias}_{\max}(T,\mathbf{z}) = \sup_{\mathbf{z}'}\|\theta - \theta'\|, \tag{3.10}$$

the breakdown point of T can then be written as

$$\varepsilon^*(T,\mathbf{z}) = \max\{f/m: \text{bias}_{\max}(T,\mathbf{z}) \text{ is finite}\}. \tag{3.11}$$

Note that this definition implies a severe test of an estimator since the outliers are considered to be in the worst possible configuration. No probability density function is assumed. As m approaches infinity, the term *asymptotic breakdown point* is used.

An estimator's breakdown point is related to another measure of robustness first mentioned by Donoho [33]; namely, the *exact fit point*. When a majority of the data exactly fit some linear equation a robust estimator should recover this equation. This desirable characteristic is known as the *exact fit property*. For an estimator possessing this property, it is of interest to know the minimum fraction of contamination required to pull the estimate away from the exact fit solution. This fraction is defined as the exact fit point and is expressed as

$$\delta^*(T,\mathbf{z}) = \min\{m/n: \exists\, \mathbf{z}' \text{ such that } T(\mathbf{z}') \neq \theta\} \tag{3.12}$$

where $\mathbf{z}$ is a sample $\{(\mathbf{x}_1,y_1),\ldots,(\mathbf{x}_n,y_n)\}$ such that $y_i = \mathbf{x}_i\theta$ for all i, and $\mathbf{z}'$ ranges over all corrupted samples where any m points of $\mathbf{z}$ are replaced by arbitrary values. For regression and scale equivariant estimators (to be defined shortly), the breakdown point is always less than or equal to the exact fit point. The exact fit point is relatively easy to determine and gives a quick indication of the robustness of an estimator.

For many estimators, certain data points have an exaggerated influence on the solution. Such points are called *leverage points* because they tend to force the solution to fit them, i.e. they exert leverage on the solution. In multiple regression with data points $(x_{i1},\ldots,x_{in},y_i)$, the points associated with the explanatory variables $(x_{i1},\ldots,x_{in})$ lie in an n-dimensional space called the *factor space*. Those points that are distant from the bulk of

the points in the factor space are leverage points. Such points are common in power systems and are associated with line flows and bus injections associated with relatively short lines and with injections on buses with a relatively high number of incident lines [16]. The magnitude of the row entries in the measurement Jacobian matrix for measurements satisfying these conditions are relatively large. Since each row vector in the matrix represents a point in the factor space of the problem, large row entries correspond to points distant from the bulk of points.

Given these definitions, it can now be stated that M-estimators possess a breakdown point of zero because they are extremely sensitive to leverage points. A single outlier in a position of leverage can make the estimator's bias go to infinity. Due to this vulnerability, the family of *generalized M-estimators* (GM-estimators or bounded-influence estimators) were developed. The intent of these estimators is to bound the influence of leverage points via a weight function w. Mallows type GM-estimators [34] replace (3.8) with

$$\sum_{i=1}^{n} w(\mathbf{x}_i)\,\psi(r_i/\hat{\sigma})\mathbf{x}_i = 0. \tag{3.13}$$

Here, leverage points are downweighted regardless of their residual value. Schweppe type GM-estimators [35] only downweight those leverage points with large residuals and are characterized by

$$\sum_{i=1}^{n} w(\mathbf{x}_i)\,\psi\left(r_i/\left(w(\mathbf{x}_i)\hat{\sigma}\right)\right)\mathbf{x}_i = 0. \tag{3.14}$$

It has been shown [20] that the breakdown point of all GM-estimators is, at best, inversely proportional to the number of regression coefficients. Whether this upper bound can be attained is not clear. Further complicating the issue, not all GM-estimators have the same breakdown point as is shown by simulations for $n = 2$ in [24], and their breakdown point depends on the design (the distribution of $\mathbf{x}$).

3.3.2 High Breakdown Point Estimators

Two important properties that will be mentioned throughout the ensuing discussion are now defined. Let $T(\mathbf{z},\mathbf{H})$ be a regression estimator that yields the state estimate $\mathbf{x}$. It is a function of the $(n + 1)$-dimensional data points, $\{(z_1,\mathbf{l}_1),\cdots,(z_m,\mathbf{l}_m)\}$ concisely represented by $(\mathbf{z},\mathbf{H})$.

Definition 1. An estimator is said to be regression equivariant if

$$T(\mathbf{z}+\mathbf{H}\mathbf{v},\mathbf{H}) = T(\mathbf{z},\mathbf{H}) + \mathbf{v}\ , \tag{3.15}$$

where $\mathbf{v}$ is any column vector. Regression equivariance is the counterpart of translation equivariance in the location case. It implies that, whenever a linear function $\mathbf{H}\mathbf{v}$ is added to the observations, the estimated state vector is shifted by the vector $\mathbf{v}$.

Definition 2. An estimator is said to be scale equivariant if

$$T(c\mathbf{z};\mathbf{H}) = cT(\mathbf{z};\mathbf{H}) \ \text{ for all } c \in \Re - \{0\}. \tag{3.16}$$

Regression and scale equivariance are two properties that a good estimator must have. We now begin the discussion of HBPEs.

The highest possible finite-sample breakdown point any regression equivariant estimator can have was derived by Rousseeuw [36] in which he considered the linear model

$$\mathbf{y} = \mathbf{X}\theta + \mathbf{e}, \tag{3.17}$$

where $\mathbf{y}$ is the vector of m observations, θ is the vector of n unknown regression parameters, $\mathbf{e}$ is a vector of m observation errors, and $\mathbf{X}$ is the m x n observation matrix. Assuming that there are no linear dependencies among the row vectors of the design matrix $\mathbf{X}$ (a condition termed *general position*), he showed that the maximum attainable breakdown point equals

$$\varepsilon^*_{\max} = \left[(m-n)/2\right]/m. \tag{3.18}$$

Here, the notation $[y]$ denotes the integer part of the real number y. Under the general position assumption, this limit is attained by the least median of squares (LMS) and least trimmed squares (LTS) estimators of Rousseeuw [36, 37], the S-estimators of Rousseeuw and Yohai [38], the MM estimators of Yohai [39], the τ-estimators of Yohai and Zamar [40], and the one-step GM estimators of Simpson, Ruppert, and Carrol [41] and Coakley and Hettmansperger [42]. Note that as m tends to infinity, $\varepsilon^*_{\max}$ approaches 50%. No estimator can exceed this limit.

The above results no longer hold when the assumption of general position is violated as it is in power systems because of system sparsity. The remainder of this section examines the effect of this condition on the LMS and LTS estimators.

The LMS and LTS estimators are defined as follows. Let $r^2_{(i)}$ denote the ith ordered squared residual, where the residuals are first squared and then ordered by increasing value: $r^2_{(1)} \leq \cdots \leq r^2_{(m)}$. The LMS estimator minimizes the νth ordered squared residual, yielding an objective function defined as

$$J(\mathbf{x}) = r^2_{(\nu)}, \tag{3.19}$$

whereas the LTS estimator minimizes the sum of the smallest ordered squared residuals up to the rank ν, yielding

$$J(\mathbf{x}) = \sum_{i=1}^{\nu} r^2_{(i)}. \tag{3.20}$$

When the m observations are in general position, Rousseeuw [36, 37] (see also Rousseeuw and Leroy [24]) has shown that for

$$\nu = \left[\frac{m}{2}\right] + \left[\frac{n+1}{2}\right], \tag{3.21}$$

the LMS and LTS reach the highest breakdown point possible for a regression equivariant estimator (see (3.18)).

Let us now consider models characterized by points in reduced position. Denote by M the maximum number of observations whose projections $(0, \boldsymbol{l}_i)$ on the factor space

lie on an (n - 1)-dimensional subspace. The following theorems [21] generalize the exact fit property of the LMS and LTS estimators for data in reduced position.

Theorem 3.1: The exact fit point δ^* of a regression equivariant estimator is at most equal to

$$\delta^*_{\max} = \left[(m - M - 1)/2\right]/m. \tag{3.22}$$

Theorem 3.2: The exact fit of the LMS and LTS estimators attains δ^*_{max} for the following quantile index ν:

$$\nu = \left[(m + M + 1)/2\right]. \tag{3.23}$$

These theorems are extended in [43] to define the breakdown point of the LTS and LMS estimators. The proofs of the above theorems are given in Appendix A and [21].

If we define the surplus s^* of the model as the minimum number of observations whose removal is required to make at least one remaining observation critical (i.e., its removal reduces the rank of the design matrix), then

$$M = m - s^* - 1. \tag{3.24}$$

The discussion of the decomposition algorithm in Chapter 4 uses s^* instead of M since its determination is easier to conceptualize in power networks. To facilitate this upcoming discussion, (3.23) and (3.22) are expressed in terms of s^*. The optimal quantile index for the LMS and LTS estimators is given by

$$\nu = m - \left\lceil \frac{s^* + 1}{2} \right\rceil, \tag{3.25}$$

for which the achieved exact fit point is

$$\delta_{max}^{*} = \left[\frac{s^{*}}{2}\right] \Big/ m. \tag{3.26}$$

3.3.2.1 Computing the LMS or LTS Estimates

Since the LMS and LTS objective functions are not differentiable, they do not possess a closed form solution. Their minimization is a combinatorial problem which is solved as follows. A minimally sized set of observations which results in a design matrix of full rank is called an elemental set. For problems meeting the general position assumption, any subset of n points is an elemental set. In reduced position problems, only a portion of these $\binom{m}{n}$ combinations satisfy the definition of an elemental set. In either case, the model is solved for each elemental set in succession. For each solution, the residuals are computed, squared, and ordered. The appropriate objective function is then evaluated. The solution corresponding to the minimum objective function is the "best" estimate in the LMS or LTS sense. Since power systems are nonlinear, solving the system state for each elemental set must be done iteratively. The Newton-Raphson algorithm is used to accomplish this task.

Given that the number of $\binom{m}{n}$ combinations quickly becomes unrealistic for problems with data in general position, Rousseeuw [24] recommends using a subset of the total possible number. He based the number of random subsets k to consider on the following formula

$$P = 1 - \left(1 - (1 - \varepsilon)^{n}\right)^{k}, \tag{3.27}$$

where P is the probability of selecting at least one noncontaminated subset when a fraction ε of outliers are assumed to be among the m data points. The formula is derived as follows. Given ε, the probability of selecting a good observation is 1 - ε. If the number of observations is large, the probability of drawing n good observations is $(1-\varepsilon)^n$. This means that the probability of having selected a contaminated set is $1-(1-\varepsilon)^n$. The probability of selecting k contaminated sets is found by raising this latest expression to the power k. Finally, the probability of selecting an uncontaminated set is given by (3.27).

There is an innate assumption of independence among the n observations as they are selected. For systems characterized by data in reduced position, the equation does not apply since each selection of a member of the subset further restricts the candidates for the next position. The degree of restriction is system dependent. The algorithm in this dissertation avoids this problem by obtaining small subsystems which allow the evaluation of all elemental sets. This ensures identification of outliers (P = 1 in (3.27)) if the breakdown point of the estimator is not exceeded.

3.3.2.2 Standardizing the LMS and LTS Residuals

Once the residuals for the state associated with the measurement set giving the minimum objective function value have been determined, it is necessary to standardize them before outliers can be determined. The standardized LMS or LTS residuals are defined as

$$r_{Si} = \frac{r_i}{\hat{\sigma}\, w_i}. \tag{3.28}$$

The first factor in the denominator is a scale estimate which estimates the dispersion of the residuals. There are a variety of candidates for this task.

The major requirement of the scale estimator is that it be robust. Several possibilities include the median absolute deviation (MAD) [24]

$$\text{MAD} = 1.4826 \underset{i}{\text{med}} \left\{ \left| x_i - \underset{j}{\text{med}}\, x_j \right| \right\}; \tag{3.29}$$

the S_m estimator [45]

$$S_m = 1.1926 \underset{i}{\text{lomed}}\, \underset{j \neq i}{\text{lomed}} \left| x_i - x_j \right|; \tag{3.30}$$

and the Q_m estimator [45]

$$Q_m = 2.2219 \left\{ \left| x_i - x_j \right|; i < j \right\}_{(k)}. \tag{3.31}$$

The low median (lomed) is the [(m+1)/2] - th order statistic while k, in the definition of Q_m, is the k-th order statistic of the $\binom{m}{2}$ interpoint distances. All of these estimators possess a 50% breakdown point. The MAD is aimed at symmetric distributions and has an asymptotic efficiency at the Gaussian model of only 37%. The estimator S_m has an asymptotic efficiency of 58% and the Q_m estimator attains an efficiency of 82%. The latter two estimators behave well even for very asymmetric distributions [45].

The number of residuals which will be obtained for the subsystems resulting from the proposed system decomposition scheme are relatively small; typically, between six and thirty. Since small samples tend to appear asymmetric, even if drawn from symmetrical distributions, the MAD is not suitable for our purposes. Although either the Q_m or S_m estimators is suitable, S_m is chosen based solely upon speed considerations.

The standardization of the residuals also makes use of weights w_i based upon projection statistics calculated for each measurement. McKean, Sheather, and Hettmansperger [46] suggest using

$$w_i = min\left\{1, (b/D_i)^2\right\}, \tag{3.32}$$

where b is a given threshold and D_i are robust distances of the m points in the n-dimensional space associated with the (m x n) Jacobian matrix **H**. Hence, if $D_i \leq b$ the corresponding residual will be standardized by a unit weight, and if $D_i > b$ it will be standardized by a weight smaller than one. In the latter case, the measurement (termed a leverage point) will have a standardized residual magnified by the factor $(D_i/b)^2$.

For models with points in general position, the constant b is equal to $\chi^2_{f,1-\alpha}$, where f equals the degrees of freedom associated with the problem; namely, m - n. For problems involving reduced position observations, there is a unique b_i term associated with each i-th observation. For power systems when the projection algorithm is used, it has been determined through the use of QQ plots [47] that an appropriate value for b_i is

$$b_i = \begin{cases} \chi^2_{2,1-\alpha} \text{ for flow measurements} \\ \chi^2_{h+1,1-\alpha} \text{ for injections,} \end{cases} \tag{3.33}$$

where h is the number of lines incident to the bus on which the injection is located. The manner in which these conclusions were reached are now discussed along with an overview of leverage point identification. The next section is an extension of [52].

3.4 Leverage Point Identification

3.4.1 Classical Method

Leverage points are defined as those points which are distant from the bulk of the points in the factor space of the problem. If each row vector of the design matrix is plotted as a point (the i-th row corresponds to the explanatory variables associated with the i-th observation), leverage points are those that are judged to be far away from the majority of such plotted points. The concept of leverage points is conceptually quite simple, but their actual identification is not.

Given a set of observations $\{\mathbf{l}_1,\ldots,\mathbf{l}_m\}$ (the row vectors of the transposed design matrix), the classical method of identifying leverage points is accomplished using Mahalanobis distances which are defined for each observation as

$$\mathrm{MD}_i^2 = (\mathbf{l}_i - \bar{\mathbf{l}})^{\mathrm{T}} \mathbf{C}^{-1} (\mathbf{l}_i - \bar{\mathbf{l}}). \tag{3.34}$$

Here,

$$\bar{\mathbf{l}} = \frac{1}{m}\sum_{i=1}^{m} \mathbf{l}_i \tag{3.35}$$

is the sample mean and the matrix $\mathbf{C}$ is given by

$$\mathbf{C} = \frac{1}{m-1}\sum_{i=1}^{m} (\mathbf{l}_i - \bar{\mathbf{l}})(\mathbf{l}_i - \bar{\mathbf{l}})^{\mathrm{T}}. \tag{3.36}$$

The latter is an unbiased version of the covariance matrix Σ. The set of points with a squared Mahalanobis distance less than or equal to a given threshold b, defines an ellipsoid

centered at $\bar{\mathbf{l}}$. The ellipsoid's principal axes are in the directions of the eigenvectors of **C** and have lengths proportional to the corresponding eigenvalues. The volume of the ellipsoid is proportional to the square-root of the determinant of **C**.

If the observations follow a multivariate normal distribution $N(\mu, \Sigma)$, the squared Mahalanobis distances are approximately distributed according to χ^2_n with n degrees of freedom. Therefore, there is a 1 - α probability that a point $\mathbf{l}_i$ will fall inside the ellipsoid defined by $MD_i^2(\mathbf{l}) \leq \chi^2_{n,1-\alpha}$. Given a value for α, one then identifies an outlier as any point whose squared Mahalanobis distance exceeds this cutoff value.

The problem with this approach is that it is based on two estimators which are not robust, although they are both statistically efficient for points distributed according to a multivariate normal distribution. In fact both estimators possess a breakdown point of zero. The norm of the sample mean can be forced to infinity by moving a single data point and the determinant of the matrix **C** can also be made to explode with a single extreme point. A small percentage of outliers can inflate the ellipsoid in order to include them in its interior, especially if they appear in clusters [48]. For this reason, the classical Mahalanobis distances are not reliable indicators of leverage points.

3.4.2 Robust Method

Donoho and Stahel [see 48] independently showd that an equivalent expression of the Mahalanobis distances is given by

$$MD_i = \max_{\|\mathbf{v}\|=1} \frac{\left|\mathbf{l}_i^T\mathbf{v} - \mathbf{L}\left(\mathbf{l}_1^T\mathbf{v},\ldots,\mathbf{l}_m^T\mathbf{v}\right)\right|}{\mathbf{S}\left(\mathbf{l}_1^T\mathbf{v},\ldots,\mathbf{l}_m^T\mathbf{v}\right)}, \tag{3.37}$$

where **L** and **S** are the sample mean and the sample standard deviation of the projections of the data points $\mathbf{l}_i$ on the vector **v**. If all possible vectors are considered, then equality of the above expression holds.

Robustification of (3.37) can be accomplished by replacing **L** and **S** with robust alternatives. Gasko and Donoho [49] and Donoho and Gasko [50] suggest using the sample median and MAD. Rousseeuw and Croux [51] and Croux and Rousseeuw [45] proposed using the S_m scale estimator instead of the MAD due to its superior efficiency at the gaussian distribution.

In practice, it is not possible to examine all possible projection vectors **v**. Gasko and Donoho [49] recommend investigating only those vectors that originate from the coordinatewise median **M** and which pass through each of the data points $\mathbf{l}_i$. The vector **M** is defined as

$$\mathbf{M} = \left[\underset{j}{\text{med}}\, \mathbf{l}_{j1}, \ldots, \underset{j}{\text{med}}\, \mathbf{l}_{jm} \right]^T. \tag{3.38}$$

Each projection vector is then given by $\mathbf{v} = \mathbf{l}_i - \mathbf{M}$, for $i = 1,\ldots,m$. This technique is known as the *projection algorithm* and the resulting distances are called *projection statistics* (PS_i). A projection statistic indicates how far the point is from the bulk of the data set in the worst one-dimensional projection [50]. Having calculated the projection statistics it is necessary to decide on a threshold value for classification of outliers. Monte Carlo simulations conducted by Rousseeuw and Van Zomeren [48] showed that the projection statistics roughly follow a chi-square distribution when the data points are in general position and drawn from a multivariate normal distribution. Modifications required for the application of these techniques to power systems are discussed next.

3.4.3 Reduced Position Models

In power systems, the projection algorithm cannot be employed directly becasue (i) the regression model is without intercept, and (ii) the Jacobian matrix is sparse. The first condition implies that the ellipsoids are constrained to have the origin as their center. We can, therefore, express the projection statistics as [52]

$$PS_i = \max_{\mathbf{v}} \frac{\left|\mathbf{l}_i^T \mathbf{v}\right|}{S_m}, \tag{3.39}$$

where $\mathbf{v} = \mathbf{l}_j$ for j = 1,...,*m*. The denominator of (3.41) should measure the spread of the $\mathbf{l}_i^T \mathbf{v}$ about the origin. This can be achieved by replacing the pairwise differences in S_m with the pairwise sums, yielding the estimate

$$S_m^{'} = 1.1926 \operatorname{lomed}_i \operatorname{lomed}_{j \neq i} \left|\mathbf{l}_i^T \mathbf{v} + \mathbf{l}_j^T \mathbf{v}\right|. \tag{3.40}$$

This estimator has the same statistical properties as S_m in the case of a fixed origin.

The other problem is raised by the sparsity of the Jacobian matrix. For a given projection vector, there will exist many measurements which do not project onto the vector. All of these zero projections result in a robust scale estimate of zero, thereby, precluding the calculation of a projection statistic. To circumvent this problem, one only considers the nonzero projections when determining the scale estimate. The only measurements which have nonzero projections on a given vector are those measurements that have a nonzero term in a column associated with one of the projection vector's nonzero coordinates. The set of measurements satisfying this constraint is called the *relevant set* for that direction [52].

Reduced position data also requires evaluation of the correct cutoff value for *each* data point, since the size of the relevant subspace may vary for different points. Determination of cutoff values given in terms of the size of the subspace were determined based on the method outlined in [48]. In this procedure, one hundred different Pθ Jacobian submatrices were generated by replacing the nonzero entries of each column by random numbers drawn from a univariate standard Gaussian distribution with zero mean and unit variance. These numbers were generated using the Box-Muller method [53]. For each submatrix, statistics for a given data point are calculated and ordered, yielding one hundred quantities. This procedure is repeated fifty times so that fifty values of each quantile of the projection statistic with index 1, 2, ..., 100 are obtained. The medians and interquantile ranges are plotted versus the corresponding quantiles of the chi-square distribution with the various degrees of freedom. By evaluating plots for various measurement configurations of the IEEE 14- and 118-bus systems, it was determined that the correct value for the number of degrees of freedom was equal to two for flow measurements and h+1 for injections, where h is the number of lines incident to the bus on which the injection is located. As far as a first order approximation in concerned, no need for a correction factor for the scale estimate S'_m is needed. Some example plots from the 14-bus system are shown in Figures 3.1 - 3.4. Each plot shows the median and interquantile values for the projection statistics plotted against the corresponding $\chi^2_{h,0.975}$ values on the abcissa. In all cases, the slope associated with the median values (□) is approximately equal to one, hence the stated cutoff values.

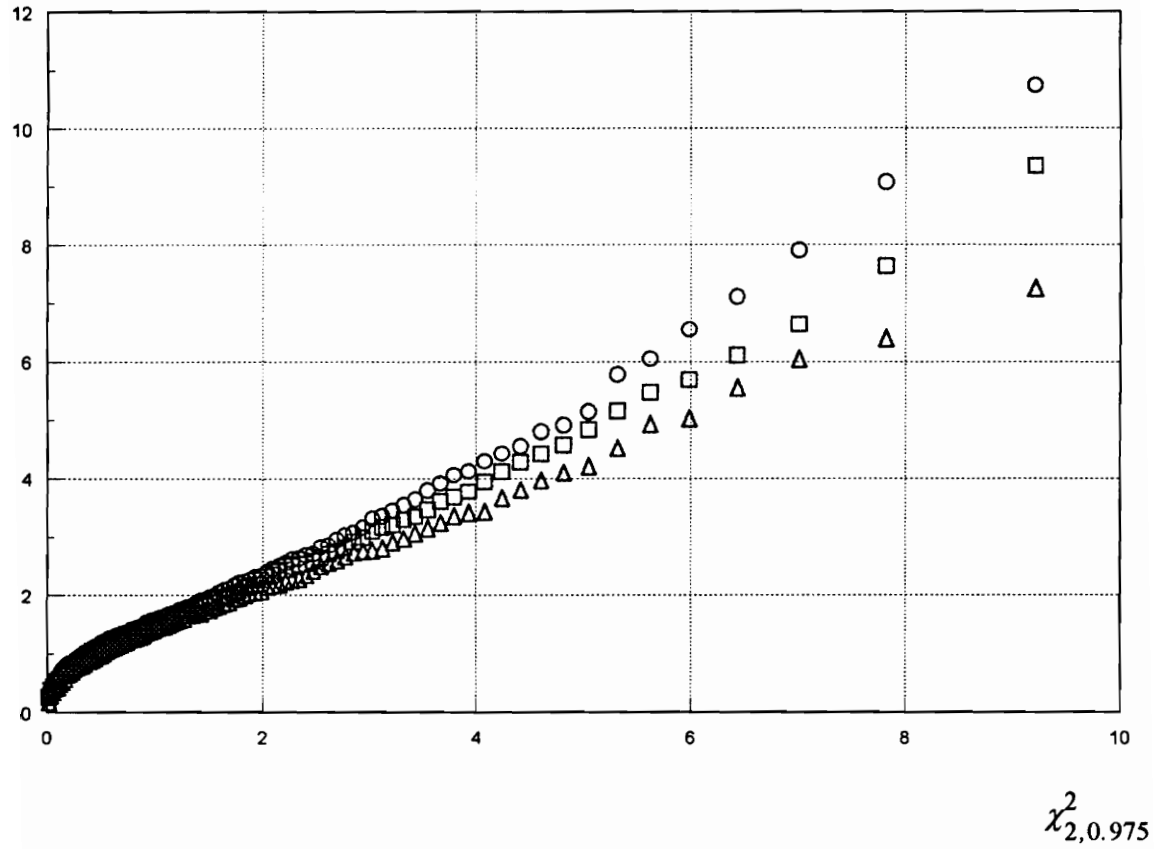


Figure 3.1 QQ Plot for Flow Measurement on Line 1-2 of 14-Bus System (Degrees of Freedom Equal to Two)

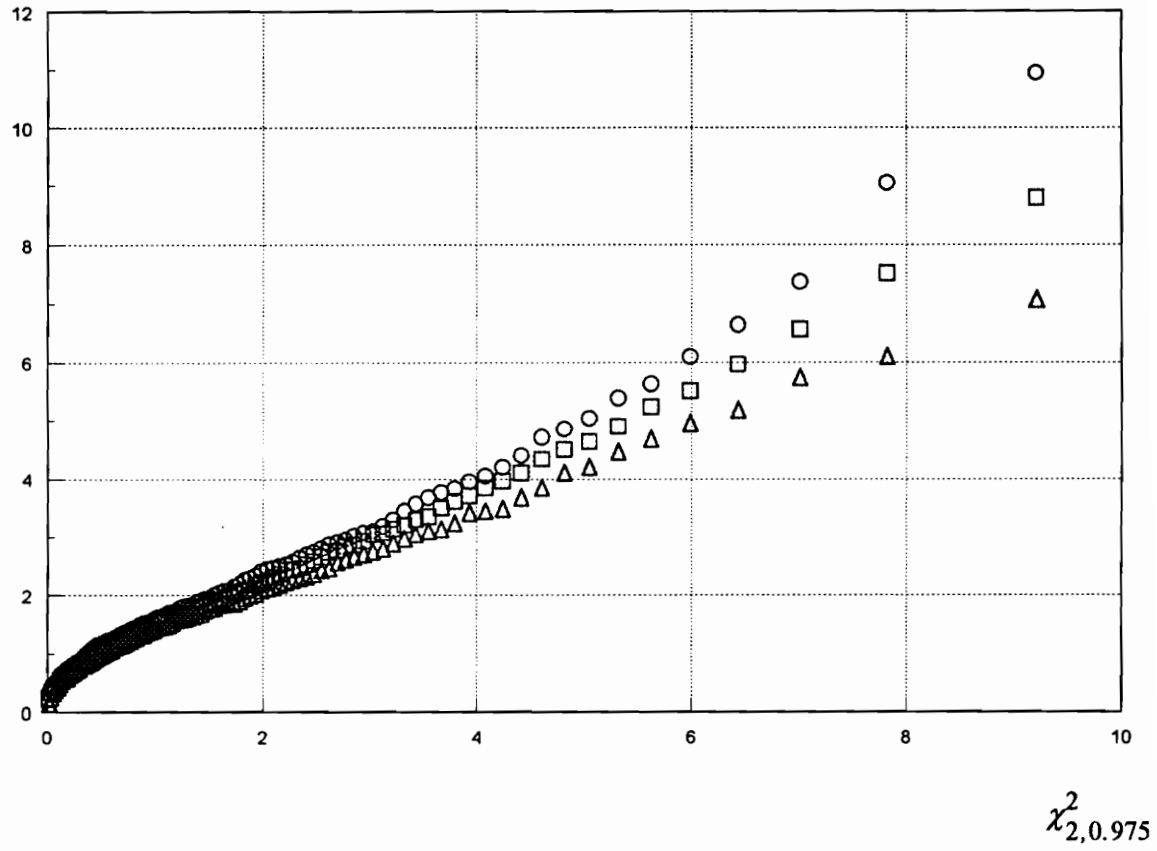


Figure 3.2 QQ Plot for Flow Measurement on Line 5 - 6 of 14-Bus System (Degrees of Freedom Equal to Two)

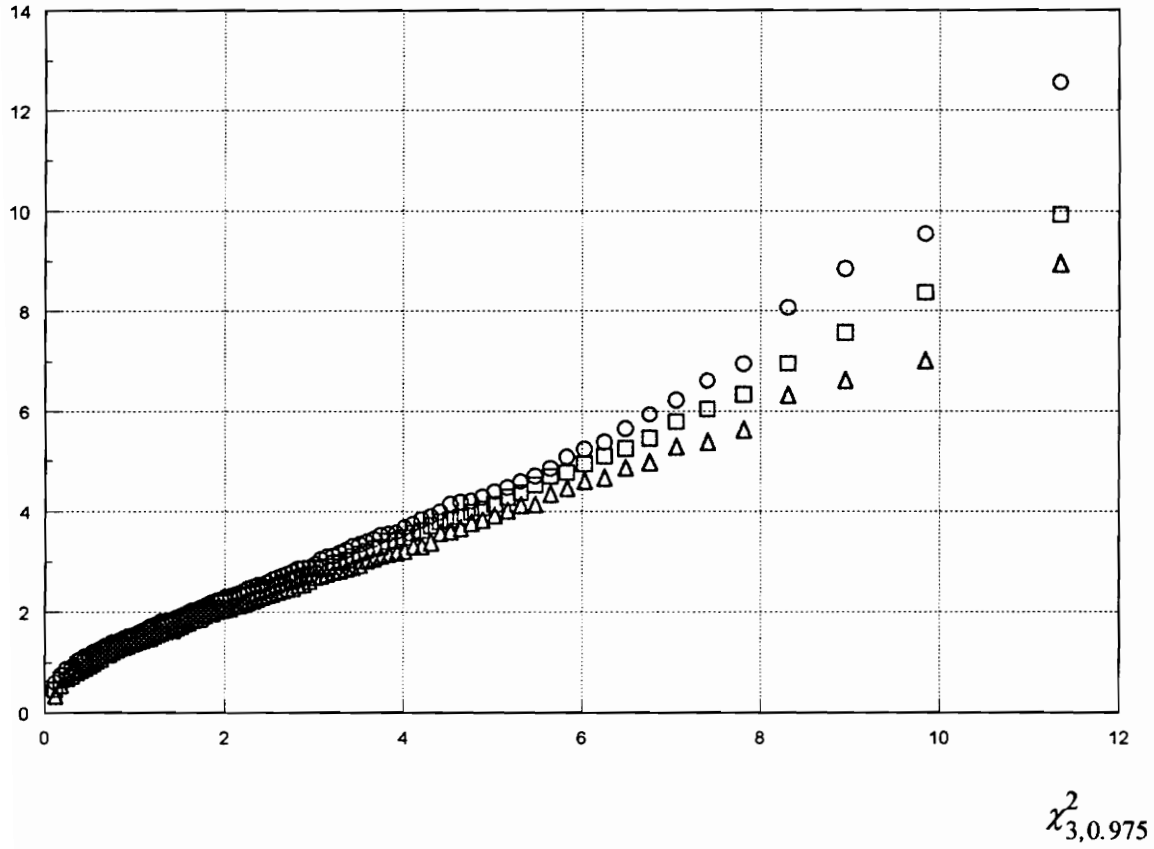


Figure 3.3 QQ Plot for Injection Measurement on Bus 1 of 14-Bus System (Degrees of Freedom Equal to Three)

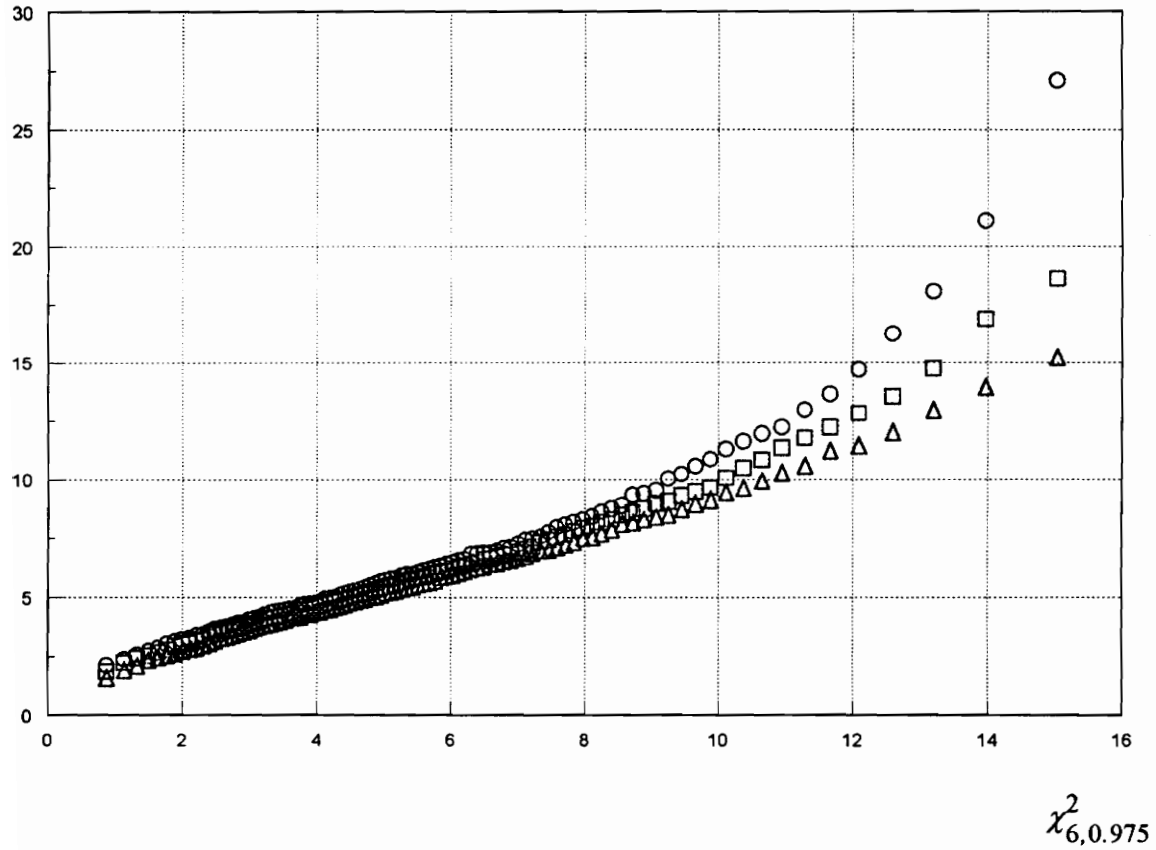


Figure 3.4 QQ Plot for Injection Measurement on Bus 4 of 14-Bus System (Degrees of Freedom Equal to Six)

Chapter 4: System Decomposition

4.1 Introduction

The major objective of this dissertation is to make the use of HBPEs a viable option for identifying outliers among power system measurements. As was noted in Chapter 3, previous work in this area was conducted without a complete understanding of the effects of system sparsity on HBPEs, so that system decomposition was seen solely as a means of realizing acceptable computing times. The method proposed by Phaniraj [21] consists of partitioning the design matrix into overlapping submatrices, each representing a subsystem of less than 150 buses. Criteria for selecting the partitions does not consider the decomposition's effect on outlier identification capability. In addition, a subsystem of 150 buses is still very large if one is to evaluate all elemental measurement sets. Although the resampling algorithm can be used to save computing time by using only a fraction of the total number of elemental measurement sets, doing so entails a nonzero probability that an uncontaminated measurement set may not be found even if one (or more) of them exists. It is preferable to make use of all of the elemental measurement sets to *guarantee* outlier identification as long as the breakdown point of the estimator is not exceeded. In addition, from an implementation standpoint, it is not clear how to generate a fraction of the total number of elemental measurement sets in an efficient, yet random, manner for problems involving data in reduced position. Due to the exponential growth relationship between the number of elemental measurement sets and system size, it is necessary to decompose the system into relatively small subsystems if all elemental sets are to be evaluated. Two subsystem types which satisfy this criteria are recommended.

In addition to choosing subsystems based on their effect on computing time, it is equally, if not more, important to consider the effect of decomposition on the estimator's

overall outlier identification capability or breakdown point. Fortuitously, the same subsystems which provide a drastic decrease in computing time also give rise to a substantial increase in outlier identification capability as compared to a nondecomposed approach. The subsystems also allow efficient generation of elemental measurement sets, rapid surplus calculations, and are readily identifiable using efficient graph theory-based algorithms.

The HBPE used in this dissertation is the LTS. Letting $r^2_{(i)}$ denote the i-th ordered squared residual, the LTS objective function is given by

$$J(\mathbf{x}) = \sum_{i=1}^{\nu} r^2_{(i)}, \tag{4.1}$$

where ν is called the quantile index. For m data points in *general position* (any set of n points observe the n state variables), Rousseeuw [24] shows that for a quantile index equal to $\nu = \lfloor m/2 \rfloor + \lfloor (n+1)/2 \rfloor$, the breakdown point of LTS is equal to $\lfloor (m-n)/2 \rfloor / m$. Note that as m tends to infinity, this latter quantity approaches 50%. Throughout this dissertation, $[x]$ means the integer part of the real variable x.

Due to the sparse nature of the measurement Jacobian matrices associated with power systems, the general position assumption is violated and the measurements are said to be in *reduced position.* As shown in [21] and in Chapter 3, using the following revised expression for the quantile index

$$\nu = m - \lfloor (s^* + 1)/2 \rfloor, \tag{4.2}$$

results in a maximum exact fit point of

$$\delta^* = \lfloor s^*/2 \rfloor / m. \tag{4.3}$$

Here s^* denotes the *measurement surplus* which is defined as the minimum number of measurements which must be removed to make at least one remaining measurement critical. Separate measurement surpluses for real power (P), reactive power (Q), and voltage magnitude measurements (V) are denoted by s_P^*, s_Q^* and s_V^*. By definition, s^* equals the minimum of these quantities. The numerator of (4.3) gives the maximum number of outliers which can be identified irrespective of their errors. This outlier identification capability will be denoted by β, that is

$$\beta = \lfloor s^*/2 \rfloor. \tag{4.4}$$

One means of obtaining an approximate LTS estimate is via elemental sets [62]. An elemental set consists of n measurements which observe the n unknowns of the system. For each elemental set, the system state is calculated and the residuals computed and squared. The system state which minimizes (4.1) is then used to identify outliers.

4.2 System Surplus

4.2.1 Necessity for Decomposition Based on System Surplus

To ensure identification of outliers in the portion(s) of the network with the lowest redundancy, it is necessary to choose the quantile index ν based upon this redundancy level. As discussed in Chapter 3, the measure of minimum redundancy is the measurement surplus s^*. Once s^* is determined, ν is determined via (4.2) allowing the identification of β outliers as given by (4.4). Before discussing a decomposition algorithm designed to increase the total number of outliers that can be identified, it is worthwhile looking at the

outlier identification capability of an HBPE applied to a typical nondecomposed system. The necessity for an alternative approach will be obvious.

The determination of s^* is a combinatorial problem involving the sequential deletion of measurement sets of increasing cardinality until a critical measurement is formed. For large systems, this straightforward approach quickly becomes infeasible as the number of combinations of measurements satisfying the setsize under consideration becomes huge. The goal here is to simply show that s^* for a typical power system is small so that applying an HBPE to a nondecomposed system makes no sense. As such, finding bounds on s^* is sufficient. Means of calculating exact s^* values for the proposed subsystems will be presented later.

Determinations of system observability and measurement criticality are made based on the topological observability requirements given by Krumpholz, Clements, and Davis [63]. They proved that a network is Pθ observable if and only if one can form a spanning tree in the system's graph by the assignment of distinct P measurements to the branches of the tree according to the following rules:

- A flow measurement can only be assigned to the edge on which it is located.
- An injection measurement can be assigned to any *one* of the edges incident to the vertex on which the injection is located.

An identical theorem holds for the QV case with an additional requirement of the presence of at least one voltage magnitude measurement. For a system to be totally observable, it must be Pθ and QV observable.

Let us now determine bounds on the surplus of a network. To facilitate the discussion, we denote three system surplus values; namely, s_V^*, s_P^*, and s_Q^*. These terms

correspond respectively to the system surplus based on the voltage magnitude measurements, the P measurements, and the Q measurements. The overall system surplus will be denoted by s^* and is, by definition, equal to the minimum of s_V^*, s_P^*, and s_Q^*.

An upper bound on s^* can be computed as follows. Without loss of generality, it is assumed that the entire network is both Pθ and QV observable. If it is not, we simply treat each observable island separately. Since a single voltage magnitude measurement is sufficient for QV observability,

$$s_V^* = m_V - 1; \tag{4.5}$$

i.e., removal of all but one voltage measurement is necessary and sufficient to make the remaining one critical. Upper bounds on the P and Q surplus values can be found based on the concept of a *fundamental set* [16].

> **Definition:** The fundamental set of a state variable is the set of measurements whose row vectors have a nonzero term in the corresponding column of the Jacobian matrix.

Since the measurements in a state variable's fundamental set are the only ones which observe the state variable, removal of all but one of these measurements will necessarily make the remaining one critical. Thus, upper bounds on s_P^* and s_Q^* are related to the fundamental set sizes of the two measurement types (FS_P and FS_Q) as follows

$$s_P^* \le \min\left|FS_P\right| - 1 \quad \text{and} \quad s_Q^* \le \min\left|FS_Q\right| - 1. \tag{4.6}$$

Combining these upper limits

$$s^* \leq \min_{\text{upper bounds}} \left\{ s_V^*, s_P^*, s_Q^* \right\}. \tag{4.7}$$

This is only an upper bound because it is sometimes possible to remove fewer P or Q measurements from different fundamental sets and cause at least one remaining measurement to become critical.

A lower bound on s^* can be determined using the graph theoretical concept of edge-connectivity. Let G denote a graph whose vertices correspond to the buses of the power system. For each transmission line flow measurement an edge of G is defined which connects the vertices of G corresponding to the buses incident to the transmission line in the actual power system. By construction there exists a one-to-one correspondence between the number of flow measurements in the system and the number of edges in G.

The edge-connectivity of a graph G, denoted by $\kappa'(G)$, is defined as the minimum number of edges which must be removed to disconnect G, i.e. break it into two components. Due to the one-to-one correspondence between the flow measurements and the edges of G, $\kappa'(G)$ equals the minimum number of flow measurements which must be removed to make the power system nonobservable from a topological standpoint. Reinstating any one of the removed measurements will restore observability and, by definition, this measurement is critical since its absence caused a loss of observability. Thus, the P (Q) surplus of the system represented by G_P (G_Q) is equal to $\kappa'(G_P) - 1$ ($\kappa'(G_Q) - 1$). Note that injection measurements have not been considered and their presence may increase the surplus of the system. As a result, the following expressions are only lower limits for systems containing measured injections:

$$s_P^* \geq \kappa'(G_P) - 1 \quad \text{and} \quad s_Q^* \geq \kappa'(G_Q) - 1. \tag{4.8}$$

The advantage of formulating the P (Q) surplus problem as one of graphical edge-connectivity lies in the fact that the determination of $\kappa'(G)$ can be accomplished using efficient non-combinatorial algorithms. One can alternatively formulate this as a maxflow/mincut network flow problem in which the edge capacities are equal to the number of flow measurements on the corresponding transmission line. Combining the upper and lower bounds one has

$$\min_{\text{lower bounds}} \left\{ s_V^*, s_P^*, s_Q^* \right\} \leq s^* \leq \min_{\text{upper bounds}} \left\{ s_V^*, s_P^*, s_Q^* \right\}. \tag{4.9}$$

Since the voltage magnitude surplus is expressed as an equality, its upper and lower bounds are identical.

Using the above ideas it is possible to quickly obtain tight upper and lower bounds on s^* for even very large power systems. Due to sparse topologies, relatively few voltage magnitude measurements, and unmeasured flows, the bounds on s^* for a power system are typically between zero and six. Since s^* is independent system size, the overall exact fit point

$$\delta^* = \lceil s^*/2 \rceil / m, \tag{4.10}$$

quickly vanishes for all but the smallest systems.

4.2.2 Subsystem Choices

The Pθ and QV cases are treated separately and, for reasons to be explained later, the evaluation of voltage magnitude measurements is conducted in a separate final step. The choice of subsystems is based upon s_P^* or s_Q^* and, since identical arguments apply to both situations, a generic s^* is used. To facilitate the discussion concerning the choice of subsystems, a measurement configuration consisting of two flow meters per line and a measured or zero injection at each bus is assumed. It is recognized that such a desirable configuration is not typically found throughout actual systems. Benefits from the proposed decomposition scheme are not dependent upon such a configuration.

The area(s) of low redundancy are detrimental in that they preclude the exploitation of greater redundancy in other portions of the network. This can be remedied by decomposing the network into sections until a point is reached for which further decomposition results in a decreased β. With this idea serving as a decomposition guideline, two subsystem types are identified to which LTS is applied separately. Since the assumed measurement configuration is symmetric (i.e. each line has the same number of flows and each bus injection is measured), one can ascertain the network surplus by examining the network topology. Using such an approach, two subsystem types are defined which (i) cause an increase in total outlier identification capability, (ii) allow application of LTS in real-time, and (iii) are readily identifiable using graph theory algorithms.

The first subsystem type is called a *radial subsystem*. These subsystems are defined by those single lines which, if removed, disconnect the network. Such lines are called *bridges* in graph theory. For example, the line connecting buses 7 and 8 of the IEEE 14-bus system depicted in Figure 4.1 is a bridge because its removal disconnects bus 8 from the remainder of the network. This is the only bridge in this system and it is

marked as ⟨8⟩. In addition to the defining line, a radial subsystem consists of the line's two incident buses, the line flow measurements, and the bus injections. With the assumed measurement configuration, each radial subsystem has an s^* equal to three - equating to a β of one. Per our decomposition guidelines, radial subsystems are identified and treated separately.

The removal of all bridges results in a set of *biconnected* components (and perhaps isolated buses which are not considered further). A biconnected component is a graph which can only be disconnected by the removal of at least two lines. In such graphs, each line appears in at least one *cycle* - a cycle being defined as a sequence whose terms are alternately distinct vertices and edges and for which the origin and terminus vertices are the same. A graph is said to be *planar* if it can be drawn in a plane without any lines crossing one another. Such a drawing is called a *planar embedding*. Figure 4.1 is a planar embedding of the 14-bus system. Note that each line, with the exception of the bridge, appears in at least one cycle. The "minimal cycles" in the embedding are those cycles which contain no others. They are marked ⟨1⟩ ··· ⟨7⟩ in Figure 4.1. Each of these

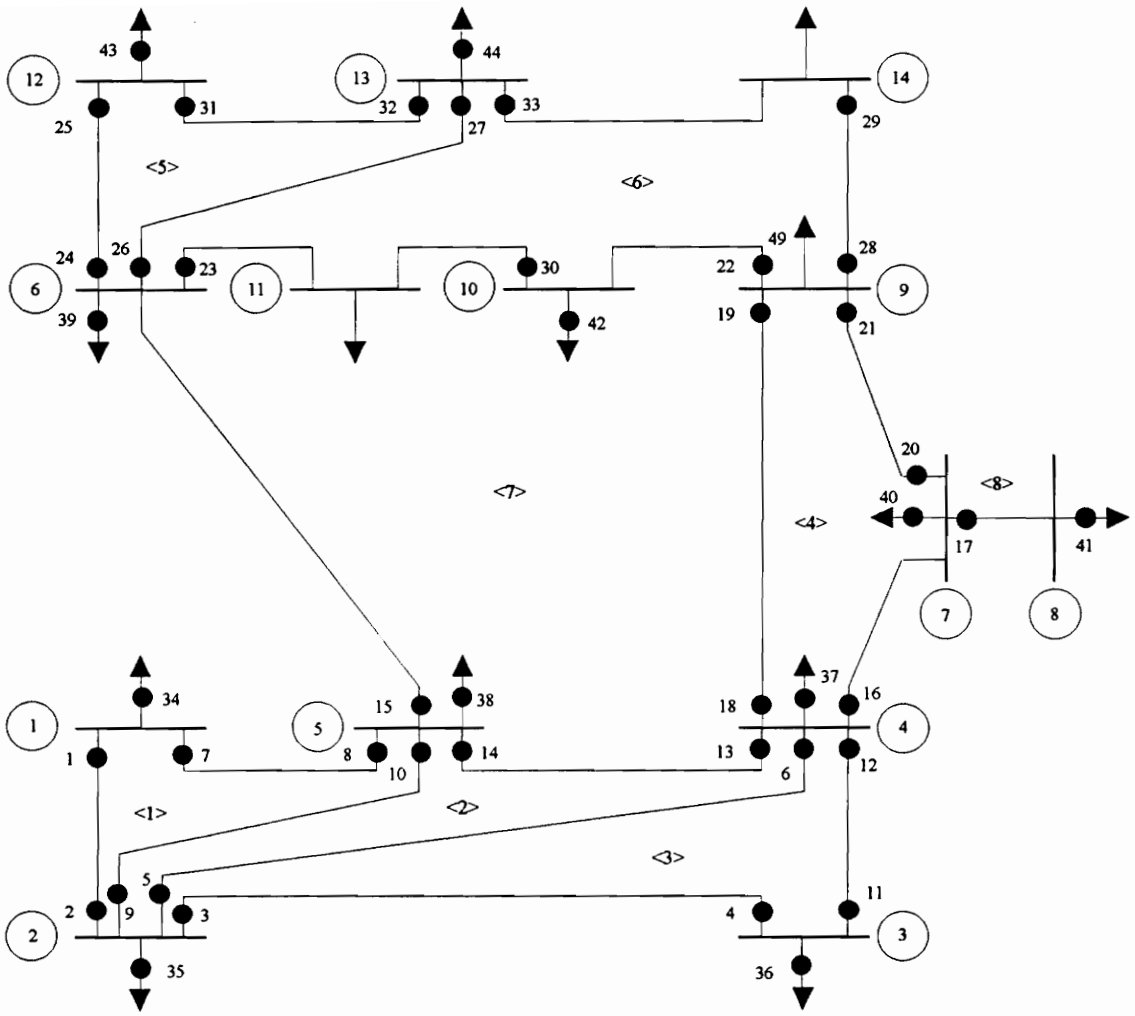


Figure 4.1 IEEE 14-bus Test System

minimal cycles defines the second type of subsystem known as a *cyclic subsystem*. These subsystems consist of the lines in the cycle along with their incident buses and the applicable line flow and injection measurements. Our motivation for using this second type of subsystem is as follows.

Based on the sparse nature of electric power systems, it is not unreasonable to assume that each minimal cycle of a component has at least one bus with only two incident

lines. If this is the case, each cyclic subsystem has a β equal to three. Since the β of the entire component is determined by one of these cycles, its β also equals three. Without further decomposition, LTS can identify three outliers in the component. On the other hand, using LTS separately on each cyclic subsystem gives one the capability of identifying three outliers in each subsystem. Since radial subsystems have a β of at most one, decomposition below the minimal cycle level is disadvantageous.

The decomposition scheme outlined above was presented using an assumed measurement configuration and topology requirement. However, the fact that a gain in outlier identification capability is realized using this method is not dependent upon an actual network satisfying these assumptions. Having introduced the general ideas behind the decomposition part of the algorithm, we will now discuss the overall algorithm in greater detail.

4.3 Description of Algorithm

After determining the buses and lines describing the radial and cyclic subsystems, one must ascertain the measurement values associated with each subsystem. Flow measurements located on subsystem lines, and injections situated on subsystem buses which did not have any incident lines cut when forming the subsystem, retain their measured values and are called *internal measurements*. Injections situated on subsystem buses which have any incident lines cut must have their values altered to account for the loss of power flow over the cut line(s). Such injections are called *salvaged injections*, while the flows on the cut lines are called *boundary flows*. To salvage an injection on bus k when the line (k, j) is cut, one uses the boundary flow (k, j) if it is available. If it is not, the flow (j, k) is utilized as the boundary flow. If neither end of the cut line is metered, the injection on bus k cannot be salvaged and is not considered as a measurement in the

subsystem. For each subsystem, the salvaged injections, boundary flows, and internal measurements comprise the subsystem's *total measurement set*. Removing the boundary flows from this set results in the subsystem's *partial measurement set*. The reason for defining these sets will become apparent shortly.

Table 4.1 lists the subsystems along with their respective total measurement sets and β values for the system of Figure 4.1. Salvaged injections are in boldface and are enclosed in parentheses with their respective boundary flows. The last column indicates the number of elemental sets associated with each subsystem.

After the subsystems and their associated measurement sets have been identified, the LTS estimate is determined for each subsystem using the method of elemental sets. Measurements from the partial measurement set are used when generating the elemental sets. The subsystems are processed in order of decreasing exact fit point, where the m in (4.3) equals the size of the subsystem's total measurement set. A decoupled model is used, allowing the Pθ and QV cases to be run in tandem. For each subsystem, one bus is arbitrarily chosen as the reference and its voltage magnitude is set equal either to 1.0 p.u. or its latest estimate. Outliers are determined using standardized residuals as discussed in Chapter 4.4. Outlying measurements are replaced by an estimated value as they are detected, unless they are a salvaged injection or have already been noted as an outlier in a previously processed subsystem (and have hence had their measured value replaced by an estimated one). If a salvaged injection is determined to be an outlier, it is not immediately clear if the injection itself and/or one or more of the boundary flows is an outlier. A statistical test which solves this dilemma is presented in Chapter 4.5. If a measurement other than a salvaged injection is noted as an outlier after it has had its measured value replaced by an estimated value, the measurement is marked as questionable since

Table 4.1 IEEE 14-Bus Decomposition Summary

Cyclic Subsystem Number	Total Measurement Set	β	Number of Elemental Sets
<1>	1, 2, 7, 8, 9, 10, 34, {**35**, 3, 5}, {**38**, 14, 15}	3	33
<2>	5, 6, 9, 10, 13, 14, {**35**, 2, 3}, {**37**, 12, 16, 18}, {**38**, 8, 15}	3	33
<3>	3-6, 11, 12, 36, {**35**, 2, 9}, {**37**, 13, 16, 18}	3	33
<4>	16, 18-21, {**37**, 6, 12, 13}, {**40**, 17}	2	19
<5>	24-27, 31, 32, 43, {**39**, 15, 23}, {**44**, 33}	3	33
<6>	22, 23, 26-30, 33, {**39**, 15, 24}, 42, {**44**, 32}	1	192
<7>	13-15, 18, 19, 22, 23, 30, {**37**, 6, 12, 16}, {**38**, 8, 10}, {**39**, 24, 26}, 42	1	336
<8>	17, {**40**, 16, 20}, 41	1	3

conflicting results have been obtained. Since this conflict is indicative of estimator breakdown, both subsystems are flagged for further investigation.

The assessment of voltage magnitude measurements is conducted after all Q measurements have been evaluated. It was necessary to delay the voltage magnitude appraisal since the decomposition scheme results in small subsystems, which may not contain a voltage magnitude measurement. The evaluation procedure proceeds as follows. Viewing the system as a whole, the largest possible QV observable islands are formed using only those Q measurements which were determined to be good, or non-questionable

ones replaced by estimated values. Since only one voltage magnitude is required for QV observability, the measurement surplus for each island is one less than the number of voltage magnitude measurements in the island. The quantile index is computed via (4.2) where m is the number of measured voltage magnitudes in the island. The elemental sets consist of each voltage magnitude used in succession as the reference voltage for the set of Q measurements used to form the island. Outlying voltages are identified through their standardized residuals.

The final output of the algorithm is (i) a list of P, Q, and voltage magnitude outliers, (ii) a list of questionable measurements, and (iii) a list of measurements which were not evaluated by the algorithm due to an insufficient surplus. The good and "corrected" measurements can then be used by a statistically efficient estimator such as the WLS or a one-step GM-estimator [41] to obtain an accurate final state estimate.

4.4 Effects of Decomposition

4.4.1 Increased Outlier Identification Capability

Due to subsystem overlap, the maximum number of outliers that can be identified is not simply equal to the sum of the outlier identification capabilities of the individual subsystems. Let β_i denote the maximum number of outliers that can be identified in the i-th subsystem and let M_i represent the total measurement set associated with the subsystem. If B is a set of outlying measurements of size $|B|$, a lower bound on the maximum number of outliers, β_G, that can be identified for the global system is

$$\beta_G = max\{|B|: |B \cap \mathrm{M}_i| \le \beta_i \quad \forall i\}, \tag{4.11}$$

i.e. the largest set of outliers such that the outlier identification capability associated with any subsystem is not exceeded. Determining this number is a combinatorial problem, but β_G is bounded by

$$\min_i\{\beta_i\} \le \beta_G \le \sum_i \beta_i . \tag{4.12}$$

These bounds represent the extremes of complete subsystem overlap and no subsystem overlap. Note that the maximum number of outliers that can be identified in the original system is equal to $\min\{\beta_i\}$, so that decomposition increases the number of outliers that can be identified for all but the most trivial systems.

Considering the application of LTS to the system of Figure 4.1 with and without decomposition, Table 4.2 gives the percentage of possible outlier combinations that do not exceed the exact fit point of any subsystem for outlier sets of increasing size. (For the nondecomposed case, the entire system can be considered a subsystem). These values are only lower bounds since the beneficial effect of correcting outlying values is not considered. The percentages apply to both the P *and* Q measurements if decoupling is used and clearly illustrate the advantages of decomposition in terms of increased outlier identification capability. Remember that the measurement errors can take any conceivable values.

4.4.2 Decreased Computing Time

The total computing time is primarily a function of (i) the number of cyclic subsystems, and (ii) the size of the cyclic subsystems. It is assumed that all elemental sets are to be evaluated for each subsystem, so that outlier identification is guaranteed as long as the exact fit point of the estimator is not exceeded in the subsystem.

Table 4.2 Comparative Outlier Identification Summary

Number of Outliers	Percentage of Possible Outlier Combinations Which Can Be Identified	
	Without Decomposition	With Decomposition
1	100.00	100.00
2	0.00	87.95
3	0.00	70.95
4	0.00	53.22
5	0.00	37.08
6	0.00	23.77
7	0.00	13.79
8	0.00	7.08
9	0.00	3.13
10	0.00	1.15

The number of radial subsystems is usually not that large since extra high voltage (EHV) networks are primarily meshed systems. The number of elemental sets for radial subsystems is simply equal to the number of measurements in the subsystem's partial measurement set. Since this number is at most four and the state determined by each elemental set can be computed directly, radial subsystems are negligible as far as total computing time is considered.

As to the number of cyclic subsystems, Euler's polyhedral formula [57] allows one to express their number in terms of the number of lines, L, and buses, N, in each

biconnected region. For a biconnected planar region, the number of cyclic subsystems, φ, is given by

$$\varphi = L - N + 1. \tag{4.13}$$

Most EHV power systems are planar or can be made so by the removal of a few lines, so that (4.13) gives a nice means of determining the number of cyclic subsystems for a given system. Given that the degree of sparsity is typically independent of system size, the growth in the number of such subsystems is approximately linear.

The determining factor in computing time is the size of the cyclic subsystems. Table 4.3 shows the number of elemental sets for various sized cyclic subsystems, assuming each bus has a measured injection and each line has the number of flow measurements indicated in the table. The growth relationship is exponential as evidenced by Figure 4.2.

Table 4.3 Growth in Elemental Measurement Sets With Cyclic Subsystem Size

	Number of Elemental Measurement Sets	
Cyclic Subsystem Size	1 Flow/Line	2 Flows/Line
3	15	33
4	52	164
5	170	765
6	534	3426
7	1631	14,917
8	4880	63,624

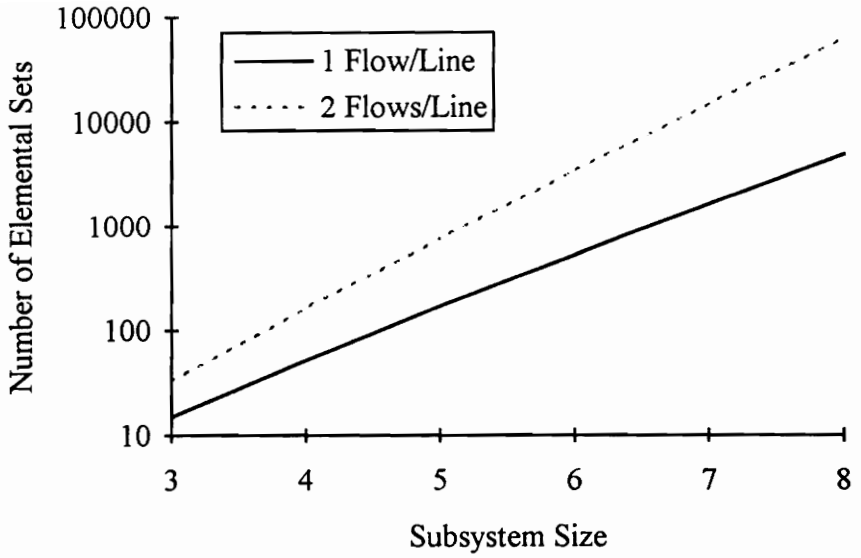


Figure 4.2 Log-Linear Plot of the Number of Elemental Measurement Sets vs. Cyclic Subsystem Size

Since the growth in computing time is a function of both the number and sizes of the cyclic subsystems, which are system dependent, it is not possible to make general statements concerning the relationship between expected computing time and system size. A large system which is highly connected can require less computing time than a much smaller one containing several large cyclic subsystems. Although a general statement relating system size and computing time is not possible, one can examine the slowest growth in computing time which can be expected by considering *maximal planar graphs*. These are planar graphs which are made nonplanar by the addition of a single line. For such graphs, the relationship between the number of lines, L, and the number of nodes, N, is given by

$$L = 3N - 6. \tag{4.14}$$

Substituting (4.14) for L in (4.13), one obtains an expression for the number of cyclic subsystems in terms of the number of system nodes; namely,

$$\varphi = 2N - 5. \tag{4.15}$$

Given that all of the cyclic subsystems for a maximal planar graph contain three nodes, (4.15) is multiplied by the applicable values from Table 4.3 to obtain Figure 4.3. This figure shows the slowest growth in the total number of elemental sets for systems of increasing size that one can expect for planar networks with the measurement configurations giving rise to the entries of Table 4.3. Since computing time is driven by the number of elemental sets that must be evalutated, Figure 4.3 gives a qualitative idea of computing time versus system size. In reality, most power systems will not be maximal planar networks so that the growth rate will be greater than indicated by the figure.

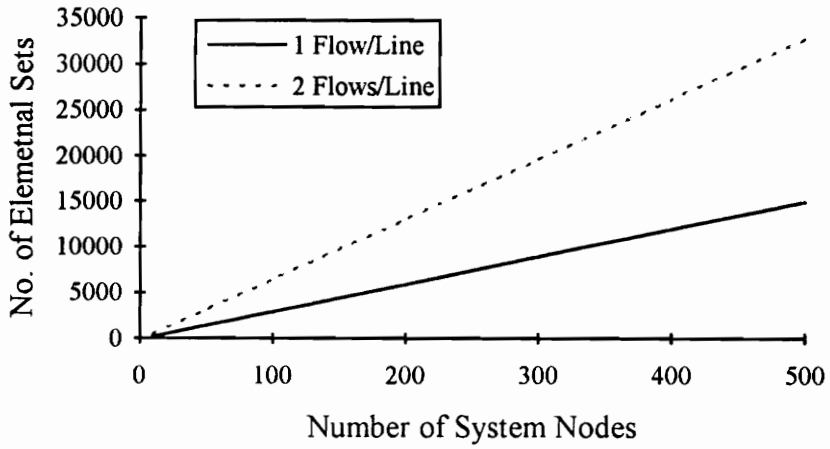


Figure 4.3 Number of Elemental Sets vs. System Size

Although the proposed decomposition scheme is based upon increasing the overall outlier identification capability of LTS, it simultaneously causes a drastic reduction in required computing time. For the nondecomposed 14-bus system of Figure 4.1, a total of 33,554,432 elemental sets must be evaluated to ensure protection against a single outlier. By decomposing, evaluation of 682 elemental sets allows the identification of 12 outliers, as we will see in Chapter 6.

4.5 Implementation Issues

4.5.1 Introduction

The actual implementation of the algorithm is described in this section. A number of ideas have been used to improve its speed. Practical matters such as resolving subsystem overlap, efficient elemental measurement set generation, breakdown detection, rapid surplus determination, and Pθ/QV decoupling are taken into consideration in the final algorithm. For the purposes of decomposition, the power system's one line diagram is viewed as a graph whose nodes correspond to the buses of the system and whose edges correspond to the system's transmission lines.

4.5.2 Subsystem Identification

An edge whose removal disconnects G is referred to as a *bridge* and corresponds to the line defining a radial subsystem. Therefore, a bridge identification algorithm [55] is used to identify the radial subsystems. This algorithm obtains its efficiency by using a depth-first search of G and has a time complexity of $O(E)$ for a graph whose edge set is denoted by E. The radial subsystems are unique.

Removal of all bridges from G results in a graph G′ in which all the edges are contained in a cycle. A *planar graph* is one that can be drawn in the plane such that no

edges intersect except at vertices. Such a drawing of the graph is called a *planar embedding*. There may exist more than one embedding for a given graph. The minimal cycles in the embedding are called *windows* and correspond to the cyclic subsystems previously defined. They are identified using a planarity testing/embedding algorithm [56]. Although not every power system is planar, it can be made so by removing relatively few edges from its graphical representation (e.g. the IEEE 14, 30, and 57 bus systems are planar while the 118 bus system only requires the removal of one line to become planar). The edge(s) that is removed to show planarity will contribute an additional cyclic subsystem. The reason for using a planarity testing algorithm to identify these subsystems is that two very efficient algorithms, both of which run in linear time, are available [60 , 61]. If [60] is used, the cyclic subsystems are determined from a dependency graph which is formed during the planarity testing phase. If the latter is used, a separate embedding algorithm due to Chiba and Nishizeki [64] is used to identify the subsystems. The computing time of this algorithm is also linear in its input. The cyclic subsystems are not necessarily unique, i.e. a variety of planar embeddings may exist.

By using these efficient and exact algorithms, one is ensured that each measurement will appear in a subsystem. Examination of the effects of various embeddings on outlier identification capability and computing speed is a subject of ongoing research.

The subsystem identification methods just described do not necessarily have to be repeated for changes in system topology. Addition of a line to the system results in one of the following conditions:

- The added line connects two buses which were previously connected by a single radial subsystem line, i.e. a parallel line is added to the system. In this case, the parallel lines constitute a new cyclic subsystem.
- The added line connects two buses between which a path consisting of only radial subsystems exists. This creates a new cyclic subsystem consisting of the added line and the radial subsystems included in the path. Note, that by definition, there is only one such path.
- The added line connects two buses which are contained in the same cyclic subsystem. This line splits the original cyclic subsystem into two parts.
- The added line connects two buses which are contained in different cyclic subsystems. In such cases, the planar embedding algorithm is run on the appropriate region to determine the new cyclic subsystems.

Removal of a line results in one of the following situations:

- The removed line is incident to two buses which are contained in cyclic subsystems. If the degree of both of the incident nodes after removal of the line is equal to or greater than two, the two subsystems bounded by the removed line are merged into one. If either of the incident nodes has a degree less than two after the line's removal, this means that a radial subsystem(s) has been formed. The radial subsystems are identified by separate searches along the path starting at each node incident to the removed line until a node of degree greater than one is reached. After the removal of these radial subsystems, the two cycles which were separated by the radial sections are merged into one.

- The removed line was a radial subsystem. Here, the subsystem is simply removed from consideration.

Thus, in only one situation is it necessary to rerun the planar embedding algorithm.

4.5.3 Elemental Measurement Set Generation

Since linear dependencies exist among the rows of the measurement Jacobian matrix corresponding to a cyclic subsystem, a means is necessary to generate those minimal measurement sets which observe the subsystem - the elemental sets. One possibility is to generate all measurement sets of the correct size and test each one to see if it observes the subsystem. As one might expect, this is not very efficient. Figure 4.4 shows the percentage of correctly sized measurement sets which observe the subsystem as a function of subsystem size. An efficient means of generating the elemental sets is

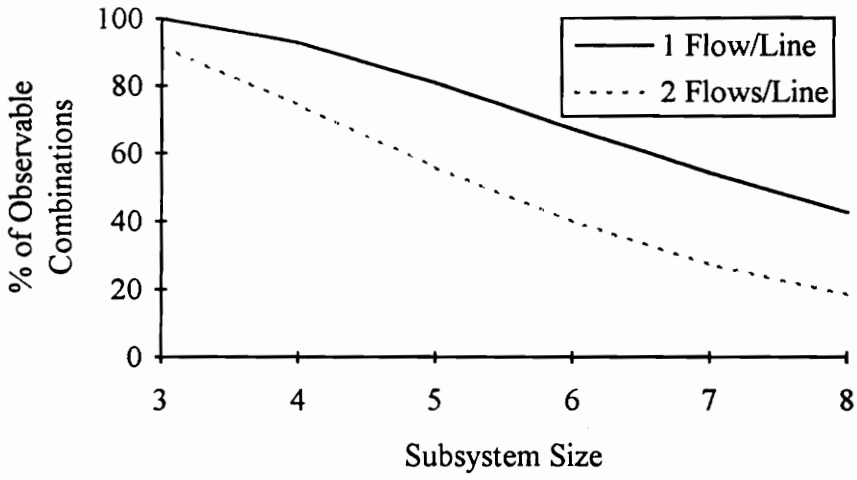


Figure 4.4 Percentage of $\binom{m}{n}$ Measurement Sets Which Observe Cyclic Subsystem vs. Subsystem Size

realized by coupling a modified topological observability algorithm with a backtracking procedure as described below.

Generating all elemental sets requires finding all vectors (z_1, z_2, ... , z_l) of given length l whose entries (measurements) satisfy an observability criterion. In a backtracking procedure, the vector is "grown" from left to right. If the partial vector has no chance of being extended, it is rejected. In this manner, one saves the effort of constructing the descendants of a clearly unsuitable partial vector [65].

At the kth stage (k = 1, l), we have a partial vector (z_1, z_2, ... , z_{k-1}) which is not inconsistent with the observability criteria. We use this vector to construct a list of all candidates for the kth position. For an element x to be a candidate means that the new partial vector (z_1, z_2, ... , z_{k-1}, x) does not yet violate the observability criteria.

If there are no candidates for the kth position, we *backtrack* by reducing k by 1, deleting z_{k-1} from the list of candidates for position k-1, and choose the new occupant of the (k-1)th position from the reduced list of candidates. If we reach $k = l$, an elemental set has been found. This set is used to calculate the state of the subsystem and the HBPE's objective function is evaluated. Upon reentry into the elemental set generation subroutine, z_l is deleted from the list of candidates for position l and we proceed as before.

To prevent duplication of elemental sets, the measurement vectors are constrained to lexicographic order. In terms of finding candidates for a particular position, the observability algorithm of Nucera and Gilles [54] is used with one simplification. Normally, one needs to check for the formation of loops during the measurement assignment process. This check is not necessary here. The number of edges with a measurement assignment must equal N-1 for an N bus subsystem to be topologically observable. Thus, the length of the vector being generated by the backtracking procedure is always less than or equal to N-1. Since it takes N edges to form a loop in an N bus

system, it is impossible for a loop to be formed during the generation of elemental sets so the customary check for loop formation can be discarded. The only check that is required is that no two measurements are assigned to the same edge. The elimination of the unnecessary check results in much faster sample generation.

4.5.4 Surplus Determination

To compute the optimal quantile index (i.e., the one which maximizes the estimator's exact fit point) it is necessary to know the exact measurement surplus of each subsystem. Simply finding upper and lower bounds for s^* as discussed in Chapter 4.2 will not suffice. The two subsystems chosen for the decomposition scheme lend themselves well to fast s^* calculation methods.

For radial subsystems, only a single measurement is required to observe the subsystem. This means that all but one measurement have to be removed from m_i to cause the remaining one to become critical. Therefore, radial subsystems have an s_i^* equal to

$$s_i^* = |m_i| - 1. \tag{4.16}$$

The surplus of a cyclic subsystem cannot be computed directly. Instead, an observability algorithm is required. In this dissertation, observability refers to topological observability as introduced in Chapter 4.2. A straightforward approach to determining s^* entails removing all combinations of measurement sets of increasing size until observability is lost. If the number of measurements removed to cause a loss of observability is the minimum number which causes this result (denote this number by d^*), it follows that restoration of any one of the measurements will restore observability and that the restored measurement is critical. The measurement surplus is then given by

$$s^* = d^* - 1. \tag{4.17}$$

The number of sets which have to be evaluated (N_S) to determine d^* by such a combinatorial approach is bounded by

$$\sum_{i=1}^{d^*-1}\binom{m}{i} + 1 \le N_s \le \sum_{i=1}^{d^*}\binom{m}{i} - 1. \tag{4.18}$$

To illustrate the inefficiency of this method, consider a cyclic subsystem consisting of five buses and fifteen measurements with an s^* equal to six ($d^* = 7$). The bounds of N_S would be

$$9949 \le N_s \le 16,383. \tag{4.19}$$

One can use a more complicated observability algorithm that detects critical measurements [66] in an attempt to decrease the number of measurement combinations that must be analyzed. If one were to do so, revised versions of (4.18) and (4.19) would be written as

$$\sum_{i=1}^{s^*-1}\binom{m}{i} + 1 \le N_s \le \sum_{i=1}^{s^*}\binom{m}{i} - 1 \tag{4.20}$$

and

$$4944 \le N_s \le 9947. \tag{4.21}$$

These bounds on relatively complicated algorithms show that computing s^* can be a burdensome task even for small subsystems consisting of simple circuits. Fortunately, the unique structure of the cyclic subsystems allows a much more efficient means of computing s^*.

For any system to be topologically observable, there must exist a unique measurement assignment for N - 1 edges, i.e. one must be able to form a spanning tree. If a system contains E edges, it follows that if E - (N - 2) edges do not have a measurement assignment available the system cannot be observable. In general, not any subset of edges of this size necessarily suffices to show a loss of observability. This is analagous to the existence of edge cutsets with cardinality greater than the edge connectivity of a graph G. Such a possiblity does not exist for cyclic subsystems. The removal of all assignable measurements for *any* two edges renders the subsystem unobservable. We can use this fact to devise a rapid means of surplus calculation for cyclic subsystems.

Associate with each edge of a cyclic subsystem a set of measurements which can be topologically assigned to it. This set consists of those flow measurements located on the edge's corresponding transmission line and those injection measurements located on the buses incident to the transmission line. The cardinality of a set indicates the number of measurements which can be assigned to the set's defining edge. To show a loss of observability it is necessary and sufficient to have two edges which do not have any assignable measurements. This condition can be achieved by removing the measurements contained within the union of any two edge measurement sets. Remember that we are interested in determining the *minimum* number of measurements which must be removed to cause a loss of observability.

Let an edge measurement set be denoted by E_i. If the minimum number of measurements which must be deleted to cause two edges to be without an assignable measurement is denoted by d', it follows that

$$d' = \min\left\{\left|E_i \cup E_j\right| \forall i, j\right\}. \tag{4.22}$$

Although removing the measurements contained in the union if these two sets will definitely cause a loss of observability, it may not be the minimum number d^* because each injection appears in two edge measurement sets. Deletion of an injection from one edge measurement set also removes it from the set associated with the adjacent edge. If the injection was the only measurement in the adjacent edge's measurement set, removing it from the subsystem leaves three, not just two, edges without any assignable measurements. Due to this fact, $d^* \leq d'$. To determine if equality holds, one determines if a tree consisting of N - 2 branches can indeed be formed. If not, measurements are reinstated until such a tree is formed. Having determined d^*, s^* can then be calculated. The efficiency of this method as compared to the other options is readily apparent. Only $\binom{N}{2}$ edge measurement set unions must be evaluated followed, at most, by a few measurement reinstations and observability checks.

4.5.5 Deciphering Salvaged Injections

A salvaged injection actually represents a set of measurements - the actual injection and the boundary flow(s). Therefore, if a salvaged injection is found to be an outlier in one or more subsystems it is necessary to determine if the injection itself is bad and/or one or more of the boundary flows is bad.

After all subsystems have been processed (i.e., outliers identified), all system flow measurements will have been classified according to the following rules:

- **Bad** - The flow was consistently identified as being an outlier in all subsystems in which it was an internal flow.
- **Good** - The flow was never determined to be an outlier in any subsystem in which it was an internal flow.
- **Questionable** - The flow was identified as an outlier in some subsystem(s) and not an outlier in other subsystem(s) in which it was an internal flow.
- **Untested** - The flow was not contained in a subsystem with sufficient surplus to warrant use of a HBPE.

Each flow measurement classified as bad is replaced by an estimated value based on the good measurements in a subsystem.

For each salvaged injection identified as an outlier, one applies the following test. If any of the salvaged injection's incident flows is untested or questionable, the actual injection is marked as questionable. This is because it is not possible to make a determination of the validity of the injection if the status of any of the incident flows is unknown. If all of the incident flows are marked as good or bad (and hence have an estimated value) Kirchoff's law is used to devise a statistical test which is sufficient to identify an outlying injection. The test will be described for real power measurements, but the same methodology applies for reactive power measurements.

Let P_i denote the actual real power injected at the ith node and let P_i^M denote its measured value. Likewise, let P_{ij} denote the actual real power flowing from node i to node j and let P_{ij}^M denote the measured flow. These measured quantities can be written as

$$\begin{aligned} P_i^M &= P_i + \varepsilon_i \\ P_{ij}^M &= P_{ij} + \varepsilon_{ij}. \end{aligned} \tag{4.23}$$

Let us assume that the measurements are independent and that their errors are normally distributed, i.e. $\varepsilon_i \sim N(0, \sigma_i^2)$ and $\varepsilon_{ij} \sim N(0, \sigma_{ij}^2)$. (Assuming we have no outlying incident flows remaining, this last assumption is not unwarranted). This allows us to write

$$\sum_j P_{ij}^M = \sum_j P_{ij} + \varepsilon, \tag{4.24}$$

where $\varepsilon \sim N(0, \Sigma\sigma_{ij}^2)$. Now,

$$\begin{aligned} P_i^M &\sim N(\mu, \sigma_i^2) \\ \sum_j P_{ij}^M &\sim N(\mu, \sum \sigma_{ij}^2). \end{aligned} \tag{4.25}$$

By Kirchoff's law,

$$P_i = \sum_j P_{ij}. \tag{4.26}$$

Since the errors in both of these terms have an assumed zero mean, we can conclude that the means of the two terms in (4.25) are equal.

Define two variables as follows:

$$\begin{aligned} X_1 &= \frac{P_i^M}{\sigma_i} \\ X_2 &= \frac{\sum_j P_{ij}^M}{\sqrt{\sum \sigma_{ij}^2}}. \end{aligned} \tag{4.27}$$

If we have removed all outliers among the incident flow measurements, these two variables should have the same N(μ, 1) distribution. Let $Y = X_1 - X_2$. If our assumption of normality is correct, i.e. the injection is not an outlier, we should have $Y \sim N(0,2)$. Alternatively, if we let

$$Y = (X_1 - X_2)/\sqrt{2}, \tag{4.28}$$

we should have $Y \sim N(0,1)$. Substituting X_1 and X_2 into the above equation we obtain

$$\frac{\left(\frac{P_i^M}{\sigma_i} - \frac{\sum_j P_{ij}^M}{\sqrt{\sum \sigma_{ij}^2}}\right)}{\sqrt{2}} \sim N(0,1). \tag{4.29}$$

Thus, if

$$\left| \left(\frac{P_i^M}{\sigma_i} - \frac{\sum_j P_{ij}^M}{\sqrt{\sum \sigma_{ij}^2}} \right) \Big/ \sqrt{2} \right| \geq \beta \tag{4.30}$$

we conclude that the injection is an outlier. If our normality assumptions are correct, β should be set equal to 3.0. Simulations indicate that this value is too low and a value of 4.5 is recommended.

This test is intentionally non-robust so that it can detect a violation of the assumption that the measurement values follow a normal distribution. It is sufficient to show that the injection is an outlier, but failure to exceed the threshold does not necessarily mean the injection is not an outlier since all outlying incident flows may not have been identified. Methods of ensuring that they have been identified and corrected, i.e. breakdown has not occurred in any subsystem containing one or more of the incident

flows, are presented in a Chapter 4.5.10. The result is that the test described in this section is only used if there is a high confidence in the validity of all the salvaged injection's incident flows.

4.5.6 Advantages of Decoupling

The number of outliers that can be identified is determined by the measurement surplus. Without decoupling, the measurement surplus (s*) is equal to the minimum of the P measurement surplus, the Q measurement surplus and the voltage magnitude surplus. A quantile index is chosen based upon s*. The **total** number of outlying measurements that can be identified is then equal to [s*/2]. For example if [s*/2] = 2, we can identify 2 outliers - 2 P **or** 2 Q measurements; 1 P and 1 Q measurement; 2 voltage magnitudes; **or** 1 voltage magnitude and 1 P or Q measurement.

Now consider the decoupled case neglecting the voltage magnitude measurements. A separate s* and quantile index are determined for the Pθ and QV cases. Let the respective surpluses be denoted by s_P* and s_Q*. Now, the total number of outliers that can be identified is equal to [s_P*/2] + [s_Q*/2]. This number is greater than for the non-decoupled case whenever the [max{s_P*, s_Q*}/2] is greater than zero. The number of outliers that can be identified in the decoupled case can never be less than for the non-decoupled case and is almost invariably greater. In the case of paired measurements, decoupling increases the P and Q outlier identification capability by a factor of two. Therefore, it is always beneficial to use the decoupled model from an outlier identification perspective.

Returning to the non-decoupled case where we are attempting to identify outlying P and Q measurements, it will be shown that decoupling is even more important when a (*j*, *i*) boundary flow is not available for an injection on the *j*th node. In such cases we would

still like to salvage the injection and can do so if an (i, j) flow pair is available by using these measurements to compute the (j, i) flows. This is accomplished as follows. The voltage angle at the ith node is assumed to be zero and the voltage magnitude is taken as 1.0 or the latest best estimate. The (i, j) flow pair is used to calculate the voltage drop across the line according to the following equation

$$\overline{V}_{ij} = \overline{Z}\left(\frac{P_{ij} - jQ_{ij}}{V_i} - jV_i\frac{B}{2}\right), \tag{4.31}$$

where $\overline{Z}$ is the line impedance and B is the total line shunt capacitance. This complex voltage drop is then applied to the assumed voltage at bus i to calculate the voltage at the jth bus

$$\overline{V}_j = V_i - \overline{V}_{ij}. \tag{4.32}$$

Now, knowing the line parameters and the voltage magnitude and phase angle at both buses one can calculate the real and reactive (j, i) flows using (2.4) and (2.5)

One concern with this method is the error in the calculated P_{ji} and Q_{ji} flows due to an erroneous P_{ij} and/or Q_{ij} flow. Expressions are now developed to show that decoupling is very important if the decomposition algorithm is to be applied to systems lacking (i, j) flows for all i-node injections. Rectangular coordinates will be used throughout the following derivations.

The voltages at buses i and j are expressed as

$$\begin{aligned} \overline{V}_i &= V_{R_i} + jV_{I_i} \\ \overline{V}_j &= V_{R_j} + jV_{I_j} \end{aligned}, \tag{4.33}$$

where the subscript R indicates the real component and I signifies the imaginary component. Since the angle at bus i is set equal to zero, $V_{I_i} = 0$. Substituting these equations into (4.32) the voltage at bus j can be written as

$$\overline{V}_j = V_{R_i} - (R + jX)\left(\frac{P_{ij} - jQ_{ij}}{V_{R_i}} - jV_{R_i}\frac{B}{2}\right). \tag{4.34}$$

Solving for the real and imaginary parts,

$$\begin{aligned} V_{R_j} &= V_{R_i} - \frac{RP_{ij}}{V_{R_i}} - \frac{XQ_{ij}}{V_{R_i}} - \frac{XB}{2}V_{R_i} \\ V_{I_j} &= \frac{RQ_{ij}}{V_{R_i}} - \frac{XP_{ij}}{V_{R_i}} + \frac{RB}{2}V_{R_i}. \end{aligned} \tag{4.35}$$

Note that the voltage at bus j is now expressed as a function of line parameters and the assumed voltage at the ith bus. The real and reactive power (j, i) flows are given by

$$P_{ji} = V_{R_j}\left[\frac{R(V_{R_j} - V_{R_i}) + X(V_{I_j} - V_{I_i})}{Z}\right] + V_{I_j}\left[\frac{R(V_{I_j} - V_{I_i}) - X(V_{R_j} - V_{R_i})}{Z}\right], \tag{4.36a}$$

and

$$Q_{ji} = V_{R_j}\left[\frac{X(V_{R_j} - V_{R_i}) - R(V_{I_j} - V_{I_i})}{Z}\right] + V_{I_j}\left[\frac{R(V_{R_j} - V_{R_i}) + X(V_{I_j} - V_{I_i})}{Z}\right] - \frac{B}{2}(V_{R_j}^2 + V_{I_j}^2). \tag{4.36b}$$

Substituting (4.35) into the above we obtain

$$P_{ji} = -P_{ij} + \frac{R}{V_{R_i}^2}\left(P_{ij}^2 + Q_{ij}^2\right) + RBQ_{ij} + \frac{RB^2V_{R_i}^2}{4} \tag{4.37}$$

and

$$\begin{aligned} Q_{ji} = {} & \frac{X}{V_{R_i}^2}\left(P_{ij}^2 + Q_{ij}^2\right) + BV_{R_i}^2\left(\frac{XB}{4} - \frac{1}{2}\right) + (XB-1)\left(Q_{ij} + \frac{B}{2}V_{R_i}^2\right) \\ & + B\left(RP_{ij} + XQ_{ij}\right) - \frac{B}{2}\left(R^2 + X^2\right)\left(\frac{P_{ij}^2 + Q_{ij}^2}{V_{R_i}^2} + BQ_{ij} + \frac{RB^2V_{R_i}^2}{4}\right). \end{aligned} \tag{4.38}$$

We now have expressions for the power flowing from bus j to bus i written in terms of an assumed voltage at bus i, measured quantities P_{ij} and Q_{ij}, and line parameters.

Our objective is to obtain equations which allow an evaluation of the sources and magnitude of the error in P_{ji} and Q_{ji} using this method of salvaging an injection. The error is equal to the difference between the true (j, i) flows $\left(P_{ji}^T \text{ and } Q_{ji}^T\right)$, which are based upon the actual voltages at buses i and j, and the P_{ji} and Q_{ji} calculated with the bus voltages determined as explained above. Let the actual (true) voltages are written as

$$\begin{aligned} \overline{V}_i^T &= V_{R_i}^T + jV_{I_i}^T \\ \overline{V}_j^T &= V_{R_j}^T + jV_{I_j}^T, \end{aligned} \tag{4.39}$$

and the estimated voltages by (4.33). The differences between the true and computed flows,

$$\begin{aligned} e_P &= P_{ji}^T - P_{ji} \\ e_Q &= Q_{ji}^T - Q_{ji} \end{aligned} \tag{4.40}$$

give expressions for the error in the computed flows using the salvaging procedure. Substitution and simplification gives

$$e_P = \left(P_{ij} - P_{ij}^T\right) + R\left(\frac{\left(P_{ij}^{T^2} + Q_{ij}^{T^2}\right)}{V_{R_i}^{T^2}} - \frac{\left(P_{ij}^2 + Q_{ij}^2\right)}{V_{R_i}^2}\right) + RB\left(Q_{ij}^T - Q_{ij}\right) + \frac{RB^2}{4}\left(V_{R_i}^{T^2} - V_{R_i}^2\right) \quad (4.41)$$

and

$$e_Q = \left(\frac{2X - BZ^2}{2}\right)\left(\frac{\left(P_{ij}^{T^2} + Q_{ij}^{T^2}\right)}{V_{R_i}^{T^2}} - \frac{\left(P_{ij}^2 + Q_{ij}^2\right)}{V_{R_i}^2}\right) + \frac{B\left(6XB - 8 - B^2Z^2\right)}{8}\left(V_{R_i}^{T^2} - V_{R_i}^2\right) + \frac{\left(4XB - B^2Z^2 - 2\right)}{2}\left(Q_{ij}^T - Q_{ij}\right) + BR\left(P_{ij}^T - P_{ij}\right) \quad (4.42)$$

Note that if the P_{ij} and Q_{ij} values are correct and $V_{R_i} = V_{R_i}^T$, the errors are equal to zero as expected.

Let us assume that $V_{R_i} = V_{R_i}^T = V$. For this case, we obtain

$$e_P' = -\left(P_{ij}^T - P_{ij}\right) + \frac{R}{V^2}\left(\left(P_{ij}^{T^2} - P_{ij}^2\right) + \left(Q_{ij}^{T^2} - Q_{ij}^2\right)\right) + RB\left(Q_{ij}^T - Q_{ij}\right) \quad (4.43)$$

and

$$e_Q' = \left(\frac{2X - BZ^2}{2V^2}\right)\left(\left(P_{ij}^{T^2} - P_{ij}^2\right) + \left(Q_{ij}^{T^2} - Q_{ij}^2\right)\right) + \frac{\left(4XB - B^2Z^2 - 2\right)}{2}\left(Q_{ij}^T - Q_{ij}\right) + BR\left(P_{ij}^T - P_{ij}\right) \quad (4.44)$$

Taking partial derivatives,

$$\frac{\partial e_P'}{\partial\left(P_{ij}^T - P_{ij}\right)} = -1.0 + \frac{R\left(P_{ij}^T + P_{ij}\right)}{V^2} \cong -1.0 \tag{4.45}$$

$$\frac{\partial e_P'}{\partial\left(Q_{ij}^T - Q_{ij}\right)} = \frac{R\left(Q_{ij}^T + Q_{ij}\right)}{V^2} + RB \cong R\left(Q_{ij}^T + Q_{ij}\right) \tag{4.46}$$

$$\frac{\partial e_Q'}{\partial\left(P_{ij}^T - P_{ij}\right)} = \left(\frac{2X - BZ^2}{2V^2}\right)\left(P_{ij}^T + P_{ij}\right) + BR \cong X\left(P_{ij}^T + P_{ij}\right) \tag{4.47}$$

and

$$\frac{\partial e_Q'}{\partial\left(Q_{ij}^T - Q_{ij}\right)} = \left(\frac{2X - BZ^2}{2V^2}\right)\left(Q_{ij}^T + Q_{ij}\right) + \frac{\left(4XB - B^2Z^2 - 2\right)}{2} \cong X\left(Q_{ij}^T + Q_{ij}\right). \tag{4.48}$$

The approximations are made under the accurate assumptions of $V \cong 1.0$, $B << X$, and P_{ij}^T and Q_{ij}^T are not large. The last assumption is valid based on a p.u. system. Of course P_{ij} and Q_{ij} can conceivably be large, but if they are that large they should be detected in the prescreeing process, in which case one would not be using them at all. One observes that the error introduced in the calculated P_{ji} by inaccurate Q_{ij} is much smaller than the error introduced in Q_{ji} by inaccurate P_{ij}. This follows from the fact that $R << X$ for EHV systems.

The conclusion is that for a non-decoupled model an outlying boundary measurement P_{ij} can cause a salvaged injection Q_j to appear as an outlier even though the injection is actually a good measurement. Decoupling is, therefore, beneficial from an outlier identification perspective for two reasons. It potentially doubles the number of

outliers that can be identified and avoids the cross-contamination between P and Q measurements as discussed above.

4.5.7 Identifying Outlying Voltage Magnitudes

Up to this point, the voltage magnitude measurements have been ignored. By decomposing the system it is almost assured that none of the subsystems will have a sufficient number of voltage magnitude measurements to allow outlier identification. In fact, some of the subsystems will probably not have any voltage magnitude measurements. By necessity, then, their evaluation must be accomplished using the nondecomposed system.

First, one may question the effect of the likely absence of a subsystem voltage magnitude measurement on Q measurement outlier identification since a voltage reference is needed for the subsystem to be QV observable. It has been determined [67, 68] that line flows are negligibly sensitive to system voltage level over a range of nominal measurement accuracy. Using this information and the fact that we are looking for outlying Q (i.e., robustness, not efficiency), the method recommended consists of using a voltage reference of 1.0 p.u. for each subsystem. If no significant changes occur in the system, the latest estimated voltage at one of the subsystem buses can be used as the reference on subsequent runs.

The evaluation of voltage magnitudes is conducted after the Q measurements have been evaluated. The Q measurements contained in subsystems in which an HBPE was used and which did not exhibit any sign of breakdown are used to form the largest possible QV observable islands. A set of good Q measurements (or bad measurements replaced by their estimated values) sufficient to observe each island is, thereby, obtained. Let V_i denote the number of voltage magnitude measurements in the i-th island. Since only one

voltage magnitude measurement is required in addition to the selected Q measurements to observe an island, the voltage magnitude measurement surplus of an island is equal to V_i - 1. If this surplus is sufficient to allow identification of outlying voltage magnitude measurements, an HBPE is used in the following manner.

Identifying outliers among the voltage magnitude measurements is a one-dimensional problem. Each of the elemental sets required for the HBPE calculation consists of the island's selected Q measurements along with one voltage magnitude. Therefore, there are only V_i elemental sets to evaluate and

$$\nu = V_i - \left[(s^* + 1)/2\right] = V_i - \left[(V_i - 1 + 1)/2\right] = V_i - \left[(V_i)/2\right]. \tag{4.49}$$

For each elemental set, the voltage magnitude in the set is used as the reference voltage and the state of the island is calculated using the Newton-Raphson algorithm. The residuals of the *measured* voltage magnitudes in the island (not all the voltage magnitudes may be measured) are then computed and the HBPE's objective function evaluated to determine which reference voltage gives the minimum value. Based upon the system state calculated using this reference voltage, outlying voltage magnitudes in the island are determined using their standardized residuals.

4.5.8 Exact Fit Point

If the state calculated using an elemental measurement set results in an objective function value below a small threshold, one concludes that the measurements give an exact fit solution. In such a case, there is no need to continue evaluating additional measurement sets. Increasing the exact fit point threshold results in fewer elemental set evaluations and, hence, decreases computing time. Of course, increasing it too much

results in accepting poor fits as exact fits thereby reducing outlier identification capability. Simulations show that an exact fit point threshold of 1.0 E-06 (as used in PROGRESS [24]) is acceptable. If anything, it is on the conservative side.

4.5.9 Residual Standardization

Once the residuals for the state associated with the measurement set giving the minimum objective function value have been determined, it is necessary to standardize them before outliers can be determined. The standardized LMS or LTS residuals are defined as

$$r_{Si} = \frac{r_i}{\hat{\sigma}\, w_i}. \tag{4.50}$$

The first term in the denominator is a scale estimate which estimates the dispersion of the residuals. Refer to Chapter 3 for calculation of this term. In cases of exact fit, the scale estimate is set equal to unity.

The standardization of the residuals also makes use of weights w_i based upon projection statistics calculated for each measurement. Their determination is also discussed in Chapter 3.

4.5.10 Zero Injections

Zero injections in a subsystem are beneficial in that they reduce the number of elemental measurement sets that must be generated to apply a HBPE to the subsystem. They are perfect measurements in that they induce a constraint due to Kirchoff's Law at the bus on which they are located. Since they are definitely good measurements (they cannot be contaminated) they are forced to appear in each elemental measurement set.

Each zero injection present in a subsystem reduces the number of places in the elemental measurement vector that must be filled causing a decrease in the total number of elemental measurement sets that are generated. Table 4.4 shows the number of measurement sets for several sized cyclic subsystems when one zero injection is present in the subsystem.

Table 4.4 Effect of A Zero Injection on the Number of Elemental Sets

Cyclic Subsystem Size	Number of Elemental Sets With One Zero Injection	
	1 Flow/Line	2 Flows/Line
3	5	8
4	20	47
5	71	240
6	235	1137
7	744	5144
8	2285	22,559

Comparison of these entries with those of Table 4.2 indicates a decrease of over 50% in the number of elemental measurement sets. Just one zero injection results in significant savings in computing time.

4.5.11 Detection of Breakdown

Although using an HBPE gives the best protection possible against breakdown, it would be nice to have some means by which a degree of confidence in our estimates can be ascertained. The simplest means is to simply compare results from adjacent subsystems. If any measurement is inconsistently identified as being an outlier the results

of the subsystems in which the measurement appears must be considered questionable. If desired, one can further evaluate these subsystems. For example, if sufficient measurement surplus exists, sequential measurement elimination with subsequent HBPE is a possibility. Even without additional evaluation, the fact that there is an indication of breakdown gives one an indication of the magnitude of contamination in that portion of the network.

Another simple check for breakdown is examination of the computed line flow(s) for a subsystem. If the flows are unrealistic, it is logical to conclude breakdown has occurred. The per unit method used in power system computations makes such an evaluation possible.

4.5.12 Parallel Processing

The implementation of an HBPE via the proposed decomposition algorithm is ideally suited for parallel processing. Each subsystem can be processed in parallel with the P and Q measurements treated separately. Although this subdivision reduces computing time somewhat, truly dramatic savings can be realized by allotting portions of the computations of the larger subsystems to different processors. By giving each processor sequential start and stop candidates for the first position in the elemental set vector (see chapter 4.5.3), each processor generates a portion of the elemental measurement sets for the subsystem and obtains a minimized objective function for those sets. These "local minima" are compared and the minimum of these is taken as the global minimum for subsystem. Since the exponential growth in the number of elemental sets is the deciding factor in the computing time of the overall algorithm, one can dramatically reduce this time using parallel processing. Systems with hundreds of processing elements capable of 13.9 gigaflops (*flops* stands for floating-point operations per second) are currently

available [69]. Increased savings in computing time will continue as advances are made in this area.

4.5.13 Estimation Efficiency

The LMS and LTS are good estimators for determining outliers, but they are not that statistically efficient for normally distributed errors. To obtain a final state estimate, it is recommended that the proposed algorithm be followed by either (1) a WLS estimate in which all measurements receive a weight of one, except for those identified as outliers which receive zero weight or (2) a Schweppe-type one-step GM-estimator [41, 70] starting from the LMS or LTS solution. The latter method is more robust than the WLS which is an advantage even if all the outliers have been removed since the errors are still probably not truly gaussian. The latter estimator possesses a high asymptotic efficiency, but this efficiency is sensitive to leverage points [70]. Evaluation of this sensitivity for large power systems with typical leverage point values is an area of potential research.

Chapter 5: Simulation Results

5.1 Introduction

Simulation results are presented for each of the three major areas studied in this dissertation. They include (1) leverage point identification using an improved projection statistics algorithm, (2) outlier identification capability of the proposed decomposition algorithm, and (3) computing time of the proposed algorithm. The reader is referred to chapter three for monte carlo simulations used to determine leverage point threshold values used in calculating the robust measurement weights associated with section 5.2.

5.2 Leverage Point Identification

The identification of leverage points using projection statistics is superior to the heuristic approach proposed by Celik and Abur [17] which is based on the observation [16] that leverage points are typically associated with flow measurements on relatively short lines and with injection measurements located on buses with a relatively high number of incident lines. This simplistic approach is unable to account for situations in which the lines incident to a bus of relatively high degree all have a fairly large impedance. In such cases, a measured injection on the bus is not necessarily a leverage point. An example illustrates this point.

Consider the IEEE 118-bus system depicted in Figure 5.1. A measurement configuration consisting of one flow measurement per line and a measured injection on each bus is assumed. Those measurements determined to be leverage points by the projection statistic algorithm described in Chapter 3 are shown as open circles. Measurements not identified as leverage points are not indicated in the figure. Observe that the injection on bus 100 is not a leverage point despite the fact the bus has seven lines

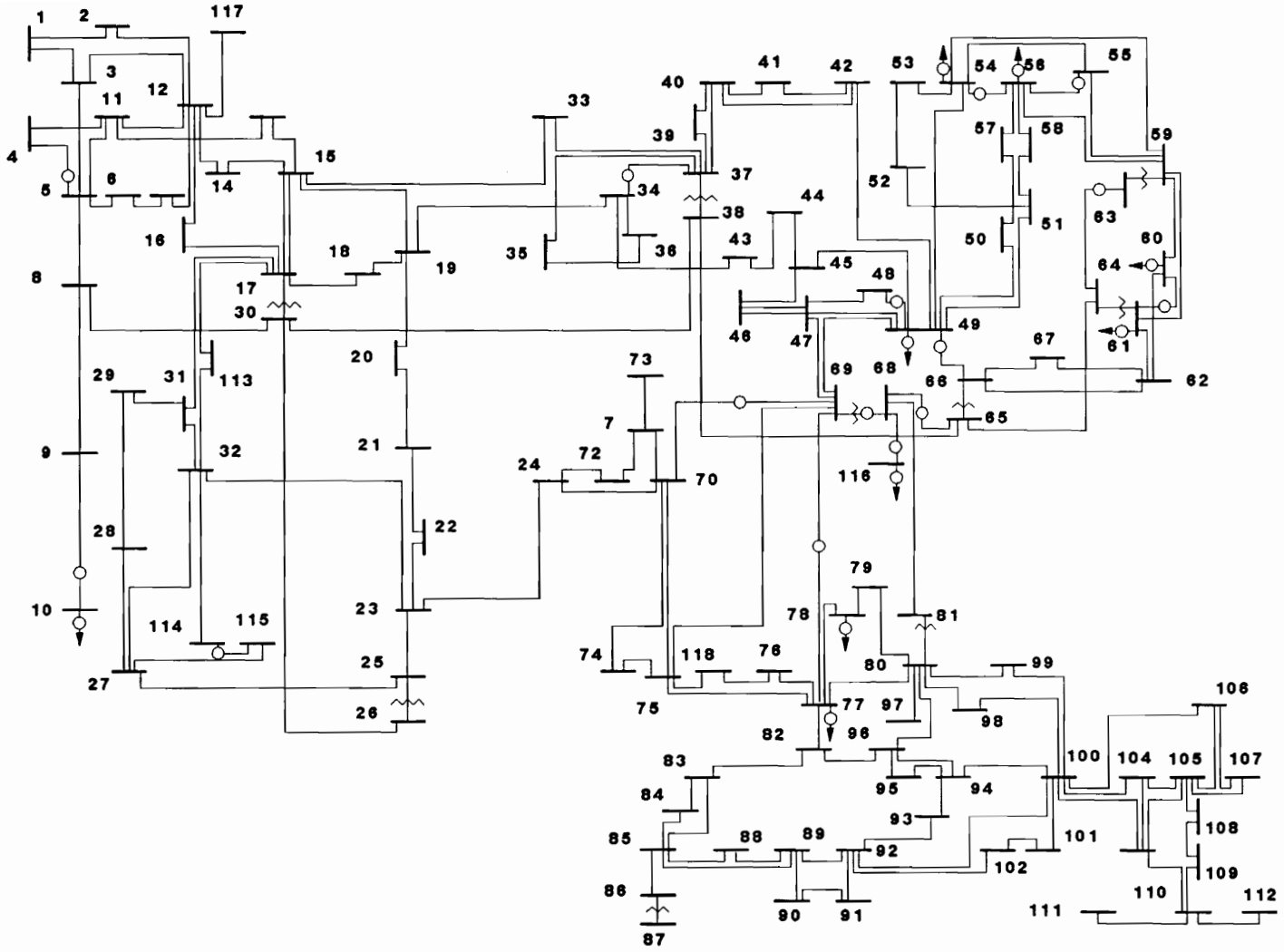


Figure 5.1 IEEE 118-Bus System Leverage Points

incident to it. Examination of the system line parameters explains why this is so - the line impedances for the lines incident to bus 100 are all relatively large. A simple heuristic approach will falsely label this injection as a leverage point because it cannot account for the line impedances.

Another point of interest related to this example is the concentration of leverage points in the vicinity of buses 48 - 70. Here, the option of removing leverage points to ensure they do not adversely affect the LAV estimator is clearly not viable. The redundancy in this section of the network would be severely diminished. In fact, observability of some buses would be lost, e.g. 116. Downweighting of all of the leverage points is possible, but examination of the corresponding measurement weights shown in Table 5.1 indicates that a nonnegligible number of measurements are essentially removed from the system due to their small weights. See measurements 102, 116, 130, 235, 240, 242, 254, and 302.

Table 5.1 IEEE 118-Bus Test System Robust Measurement Weights

Measurement Number	Fm Bus	To Bus	Measurement Weight
1	1	2	1
2	1	3	1
3	2	12	1
4	3	5	1
5	3	12	1
6	4	5	0.2114
7	4	11	1
8	5	6	1
9	5	11	1
10	6	7	1
11	7	12	1
12	8	5	1
13	8	9	1

Table 5.1 (continued)

Measurement Number	Fm Bus	To Bus	Measurement Weight
14	8	30	1
15	9	10	1
16	11	12	1
17	11	13	1
18	12	14	1
19	12	16	1
20	12	117	1
21	13	15	1
22	14	15	1
23	15	17	1
24	15	19	1
25	15	33	1
26	16	17	1
27	17	18	1
28	17	31	1
29	17	113	0.941
30	18	19	1
31	19	20	1
32	19	34	1
33	20	21	1
34	21	22	1
35	22	23	1
36	23	24	1
37	23	25	1
38	23	32	1
39	24	70	1
40	24	72	1
41	25	27	1
42	26	25	1
43	26	30	1
44	27	28	1
45	27	32	1
46	27	115	1
47	28	29	1
48	29	31	1
49	30	17	1
50	30	38	1
51	31	32	1

Table 5.1 (continued)

Measurement Number	Fm Bus	To Bus	Measurement Weight
52	32	113	1
53	32	114	1
54	33	37	1
55	34	36	1
56	34	37	0.1437
57	34	43	1
58	35	36	1
59	35	37	1
60	37	39	1
61	37	40	1
62	38	37	1
63	38	65	1
64	39	40	1
65	40	41	1
66	40	42	1
67	41	42	1
68	42	49	1
69	42	49	1
70	43	44	1
71	44	45	1
72	45	46	1
73	45	49	1
74	46	47	1
75	46	48	1
76	47	49	0.5588
77	47	69	1
78	48	49	0.3711
79	49	50	0.8999
80	49	51	1
81	49	54	1
82	49	54	1
83	49	66	0.247
84	49	66	0.247
85	49	69	1
86	50	57	1
87	51	52	1
88	51	58	1
89	52	53	1
90	53	54	1

Table 5.1 (continued)

Measurement Number	Fm Bus	To Bus	Measurement Weight
91	54	55	0.9001
92	54	56	0.0071
93	54	59	1
94	55	56	0.1437
95	55	59	1
96	56	57	1
97	56	58	1
98	56	59	1
99	56	59	1
100	59	60	1
101	59	61	1
102	60	61	0.0546
103	60	62	1
104	61	62	1
105	62	66	1
106	62	67	1
107	63	59	0.7585
108	63	64	0.2445
109	64	61	0.7786
110	64	65	1
111	65	66	1
112	65	68	0.3406
113	66	67	1
114	68	69	0.3341
115	68	81	0.5877
116	68	116	0.0109
117	69	70	0.1122
118	69	75	1
119	69	77	1
120	70	71	0.7773
121	70	74	1
122	70	75	1
123	71	72	1
124	71	73	0.509
125	74	75	1
126	75	77	1
127	75	118	1
128	76	77	1
129	76	118	0.9707

Table 5.1 (continued)

Measurement Number	Fm Bus	To Bus	Measurement Weight
130	77	78	0.0608
131	77	80	0.6683
132	77	80	0.6683
133	77	82	1
134	78	79	0.5663
135	79	80	1
136	80	96	1
137	80	97	1
138	80	98	1
139	80	99	1
140	81	80	1
141	82	83	1
142	82	96	1
143	83	84	1
144	83	85	1
145	84	85	1
146	85	86	1
147	85	88	1
148	85	89	1
149	86	87	1
150	88	89	1
151	89	90	1
152	89	90	1
153	89	92	0.5943
154	89	92	0.5943
155	90	91	1
156	91	92	1
157	92	93	1
158	92	94	1
159	92	100	1
160	92	102	1
161	93	94	1
162	94	95	1
163	94	96	1
164	94	100	1
165	95	96	1
166	96	97	1
167	98	100	1
168	99	100	1

Table 5.1 (continued)

Measurement Number	Fm Bus	To Bus	Measurement Weight
169	100	101	1
170	100	103	1
171	100	104	1
172	100	106	1
173	101	102	1
174	103	104	1
175	103	105	1
176	103	110	1
177	104	105	0.6244
178	105	106	1
179	105	107	1
180	105	108	1
181	106	107	1
182	108	109	1
183	109	110	1
184	110	111	1
185	110	112	1
186	114	115	0.3761
187	1	N/A	1
188	2	N/A	1
189	3	N/A	1
190	4	N/A	0.3199
191	5	N/A	0.3921
192	6	N/A	1
193	7	N/A	1
194	8	N/A	1
195	9	N/A	1
196	10	N/A	0.1925
197	11	N/A	1
198	12	N/A	1
199	13	N/A	1
200	14	N/A	1
201	15	N/A	1
202	16	N/A	1
203	17	N/A	0.3822
204	18	N/A	1
205	19	N/A	1
206	20	N/A	1
207	21	N/A	1

Table 5.1 (continued)

Measurement Number	Fm Bus	To Bus	Measurement Weight
208	22	N/A	1
209	23	N/A	1
210	24	N/A	1
211	25	N/A	1
212	26	N/A	1
213	27	N/A	1
214	28	N/A	1
215	29	N/A	1
216	30	N/A	1
217	31	N/A	1
218	32	N/A	0.6653
219	33	N/A	1
220	34	N/A	0.3314
221	35	N/A	1
222	36	N/A	1
223	37	N/A	0.2948
224	38	N/A	1
225	39	N/A	1
226	40	N/A	1
227	41	N/A	1
228	42	N/A	1
229	43	N/A	1
230	44	N/A	1
231	45	N/A	1
232	46	N/A	1
233	47	N/A	1
234	48	N/A	0.5651
235	49	N/A	0.0669
236	50	N/A	1
237	51	N/A	1
238	52	N/A	1
239	53	N/A	1
240	54	N/A	0.018
241	55	N/A	0.2426
242	56	N/A	0.0151
243	57	N/A	1
244	58	N/A	1
245	59	N/A	0.7763
246	60	N/A	0.0857

Table 5.1 (continued)

Measurement Number	Fm Bus	To Bus	Measurement Weight
247	61	N/A	0.1027
248	62	N/A	1
249	63	N/A	0.1962
250	64	N/A	0.2275
251	65	N/A	0.7917
252	66	N/A	0.3494
253	67	N/A	1
254	68	N/A	0.0176
255	69	N/A	0.4228
256	70	N/A	1
257	71	N/A	1
258	72	N/A	1
259	73	N/A	0.5104
260	74	N/A	1
261	75	N/A	1
262	76	N/A	1
263	77	N/A	0.0781
264	78	N/A	0.073
265	79	N/A	1
266	80	N/A	0.9889
267	81	N/A	0.8421
268	82	N/A	1
269	83	N/A	1
270	84	N/A	1
271	85	N/A	1
272	86	N/A	1
273	87	N/A	1
274	88	N/A	1
275	89	N/A	0.3106
276	90	N/A	1
277	91	N/A	1
278	92	N/A	0.2472
279	93	N/A	1
280	94	N/A	0.6545
281	95	N/A	1
282	96	N/A	0.7007
283	97	N/A	1
284	98	N/A	1
285	99	N/A	1

Table 5.1 (continued)

Measurement Number	Fm Bus	To Bus	Measurement Weight
286	100	N/A	0.82
287	101	N/A	1
288	102	N/A	1
289	103	N/A	1
290	104	N/A	0.8731
291	105	N/A	0.9301
292	106	N/A	1
293	107	N/A	1
294	108	N/A	1
295	109	N/A	1
296	110	N/A	0.9544
297	111	N/A	1
298	112	N/A	1
299	113	N/A	1
300	114	N/A	0.4304
301	115	N/A	0.54
302	116	N/A	0.011
303	117	N/A	1
304	118	N/A	1

This example indicates the superiority of determining robust measurement weights via the projection statistics algorithm versus a heuristic approach based strictly on network topology. It also shows that a state estimation algorithm capable of handling leverage points is needed. The decomposition algorithm coupled with the LTS estimator as introduced in the previous chapter is just such an algorithm. The power and limitations of this alternative algorithm are now shown via examples, tables, and graphical plots.

5.3 Proposed LTS and Decomposition Algorithm Results

5.3.1 Outlier Identification

The ability of the algorithm to identify conforming errors among interacting measurements is demonstrated using the IEEE 14-bus system of Figure 5.2. Gaussian errors are added to actual measurement values obtained via a load flow. A standard deviation of 0.02 p.u. is used. Measurements {2, 3, 5 , 9, 35} are set equal to zero to simulate a remote terminal unit (RTU) failure at bus 2; measurements {25, 31, 43} are increased by 20σ to simulate an RTU shift at bus 12; measurement 19 is set equal to -10.0; measurement 21 is set equal to 10.0; measurement 17 is set equal to -20.0; and measurement 42 is set equal to 1.0. The standardized residuals for the $P\theta$ case are given in Table 5.2 in which the number in the upper left corner of each subsystem summary indicates the order in which the subsystem was processed. All of the twelve outliers are correctly identified and each has its standardized residual underlined and marked in boldface type. Analogous results are obtained for the QV case and, if all bus voltages are measured, any set of six outlying voltage magnitudes can be identified. This yields a total of 30 outliers. Including the surplus calculations, the computing time for the $P\theta$ case on a VAXstation 4000/60 was 1.45 seconds. The total number of elemental sets evaluated was 682.

The 14-bus system is sufficiently small to list all of the subsystem results in a concise manner. Since the algorithm's principles are independent of system size, the 14-bus system results serve to illustrate the proposed method without overwhelming the reader with tables of data. Results from larger systems are not included to avoid this problem.

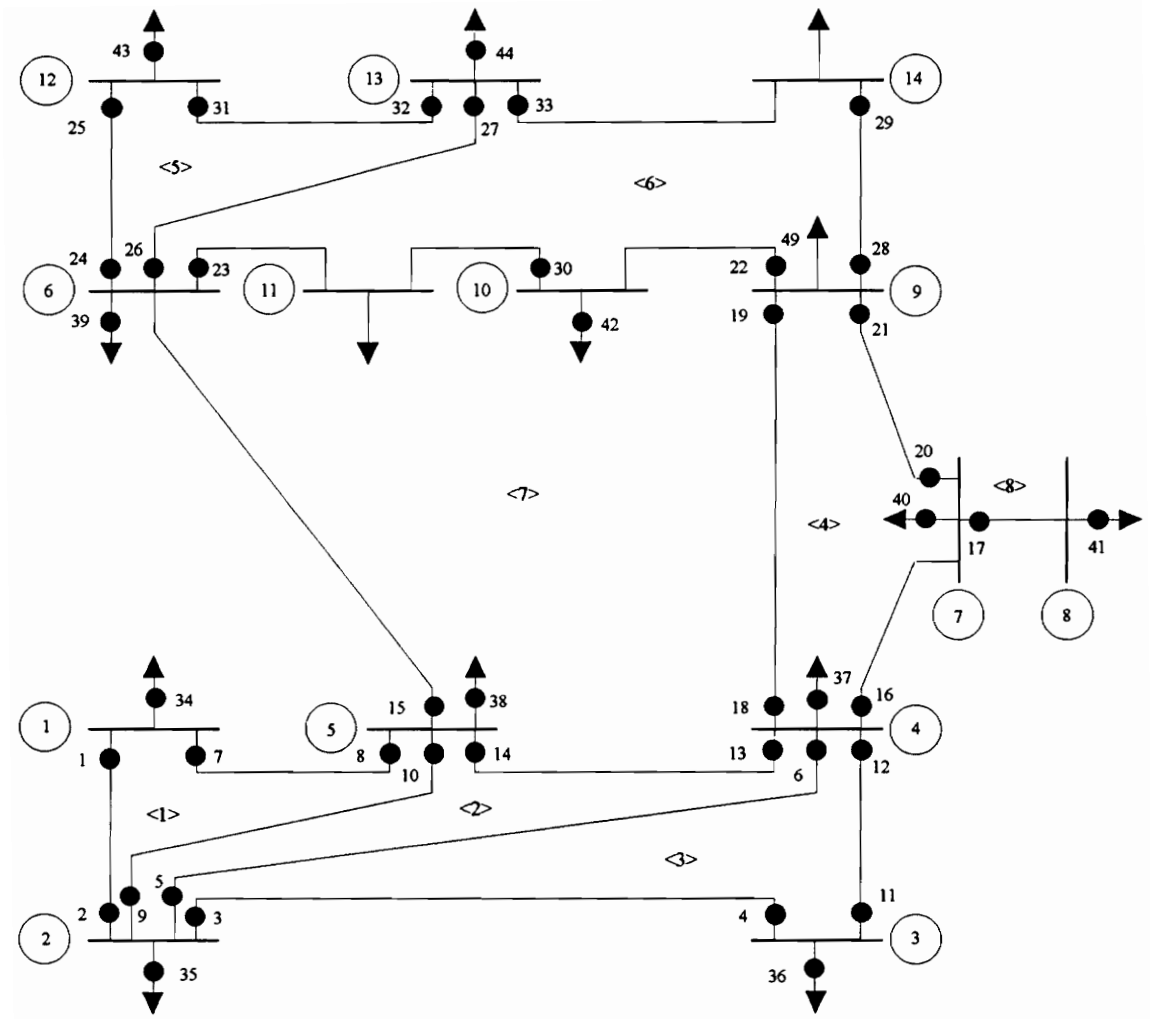


Figure 5.2 IEEE 14-bus Test System

Table 5.2 Standardized Residuals for IEEE 14 Bus Subsystems

(1) Subsystem <5>	
Meas. No.	Standardized Residual
24	-0.969
25	**19.865**
26	0.729
27	-0.018
31	**19.793**
32	0.000
39	0.306
43	**19.515**
44	0.429

(2) Subsystem <1>	
Meas. No.	Standardized Residual
1	0.000
2	**17.140**
7	-0.641
8	0.134
9	**-5.306**
10	0.310
34	0.000
35	**11.834**
38	-0.432

(3) Subsystem <3>	
Meas. No.	Standardized Residual
3	**-24.855**
4	0.000
5	**-16.154**
6	-0.450
11	0.663
12	-0.543
35	**-20.076**
36	0.000
37	0.198

(4) Subsystem <8>	
Meas. No.	Standardized Residual
17	**-14.713**
40	-0.223
41	0.230

(5) Subsystem <2>	
Meas. No.	Standardized Residual
5	-0.795
6	0.000
9	0.415
10	0.602
13	-2.307
14	2.218
35	**-37.622**
37	-0.139
38	-0.054

(6) Subsystem <4>	
Meas. No.	Standardized Residual
16	-0.700
18	0.000
19	**-74.603**
20	1.598
21	**74.268**
37	-0.564
40	0.000

(7) Subsystem <6>	
Meas. No.	Standardized Residual
22	-0.003
23	-0.009
26	-0.031
27	2.780
28	0.739
29	0.000
30	-0.013
33	-2.846
39	2.055
42	**199.439**
44	1.630

(8) Subsystem <7>	
Meas. No.	Standardized Residual
13	-3.191
14	2.962
15	3.329
18	-2.407
19	2.408
22	0.019
23	-0.016
30	-0.021
37	-0.132
38	-0.819
39	0.008
42	-0.024

5.3.2 Computing Time

The capability of the proposed algorithm to operate in a real-time environment for the IEEE 14-Bus system has been demonstrated. For the algorithm to be truly useful, its computing time when applied to larger must not be excessive. The computing time of the algorithm is determined by the number and sizes of the subsystems resulting from system decomposition. Larger subsystems require a greater number of elemental set evaluations, as was tabulated in Table 4.3. In addition to an increase in the number of elemental sets with increasing subsystem size, determining the subsystem state associated with each elemental set requires more time for larger subsystems. This is because LU factorization and backsolving is an $O(n^3/3)$ procedure, where n is the number of subsystem unknowns.

Table 5.3 lists the number of subsystems of varying size for each of the IEEE 14-, 30-, 57-, and 118-bus systems. Table 5.4 gives the time required to compute the LTS estimate of various subsystem sizes for two measurement configurations - (1) one flow per line and an injection on each node, and (2) two flows per line and an injection on each node. Figure 5.3 is a plot of these values. For measurement configuration one, it is apparent that cyclic subsystems of more than 8 - 9 buses are too computationally demanding for examination of all elemental sets. An upper limit of about 7 buses applies to the second measurement configuration. It should be noted at this point that it may be possible to decrease these times via symbolic inverses, sparsity programming, and/or QR factorization with partial updating to decrease each state solution time. Resampling can be used to decrease the number of elemental sets needing evaluation.

Current upper limits on subsystem size have been noted. Fortuitously, for a given system, the percentage of subsystems that exceed these limits is small. This fact is evident from Figure 5.4 which is a plot of the cumulative percentage of subsystems versus subsystem size for each of the test systems examined. For each of the systems,

approximately 70% of the subsystems consist of six or less buses. The fact that this percentage is independent of system size is very appealing since it means that the proposed algorithm can be used in about 70% of the subsystems for any of the systems without resorting to resampling or parallel processing in order to achieve acceptable computing times.

Table 5.3 Decomposition Summary of Test Systems

	Number of Subsystems of Size N For The Following Test Systems			
Subsystem Size N	14-Bus	30-Bus	57-Bus	118-Bus
2	1	3	3	9
3	4	4	8	19
4	0	2	3	15
5	1	0	1	8
6	0	0	2	5
7	2	2	0	2
8	0	2	2	6
9	0	1	1	4
10	0	0	0	0
11	0	0	2	0
12	0	0	1	0
13	0	0	0	1
14	0	0	0	1
15	0	0	1	0
16	0	0	0	1

Table 5.4 Subsystem Computing Times (Seconds)

Subsystem Size	Measurement Configuration One	Measurement Configuration Two
3	.01	.02
4	.03	.16
5	.18	.62
6	.50	2.88
7	1.70	14.55
8	5.84	73.49
9	20.03	*
10	67.64	*

* - time too excessive to consider listing

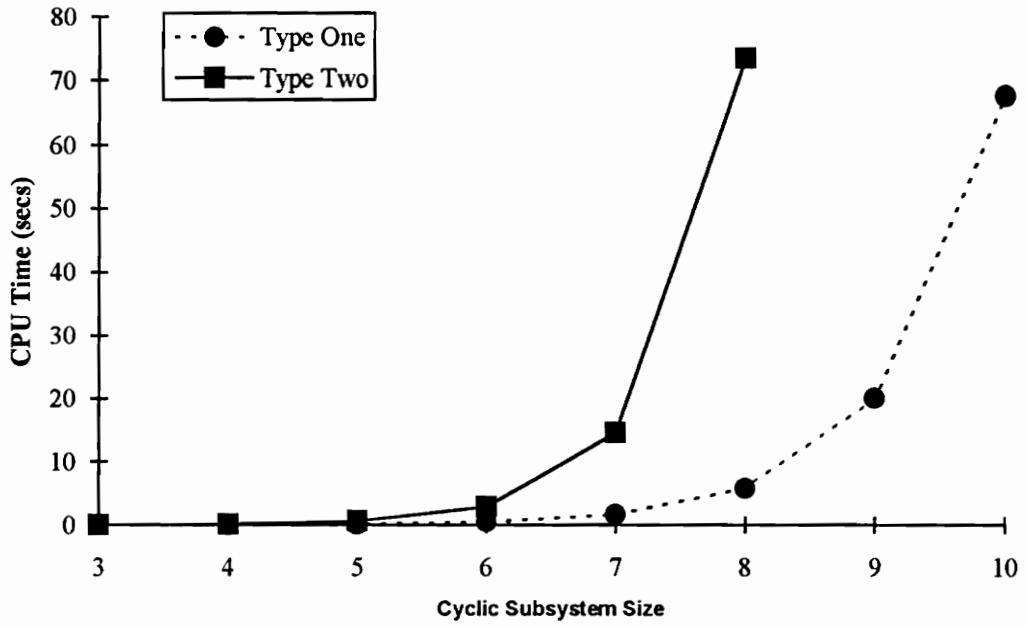


Figure 5.3 LTS Estimate Computing Time versus Subsystem Size

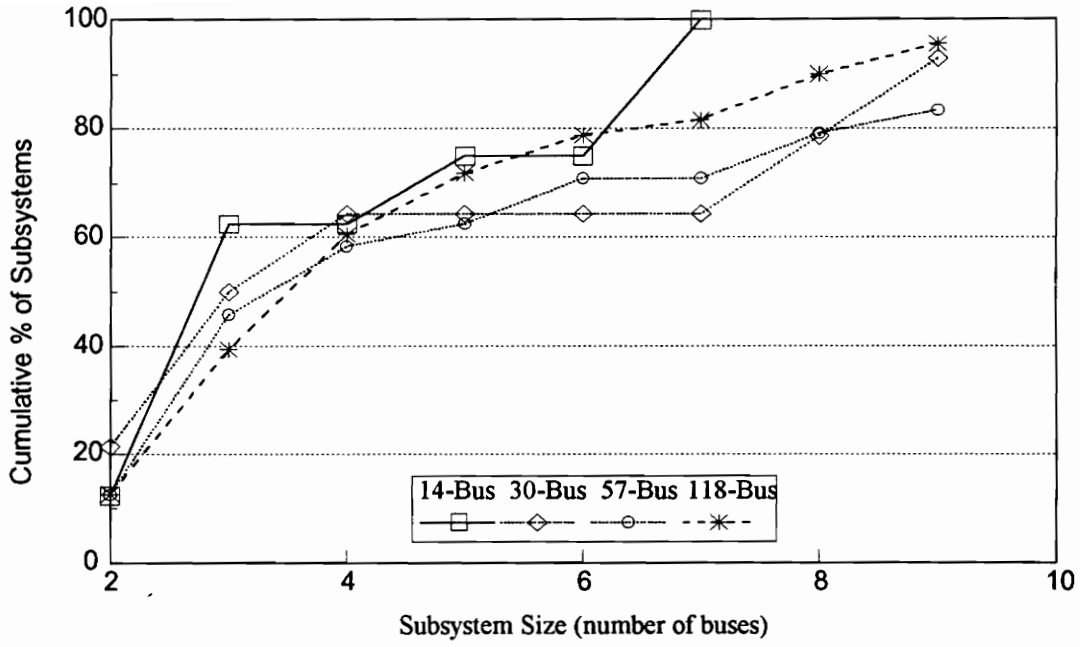


Figure 5.4 Subsystem Size Distribution Comparison for Selected IEEE Test Systems

5.4 Summary

The identification of leverage points based on projection statistics is a very powerful tool - theoretically and practically superior to heuristic methods, yet still very fast. Application of the method to IEEE test systems indicates that leverage points are quite prevalent in power systems, so that direct application of the LAV estimator is not wise. Although use of robust measurement weights mitigate the detrimental effect of bad leverage points on the LAV, their use also removes the gain in state estimation accuracy that good leverage points provide. For this reason, a state estimation algorithm inherently capable of handling leverage points is desired. The LTS estimator, coupled with system decomposition, is an attractive alternative from this perspective.

The viability of the proposed algorithm has been illustrated by showing its outlier identification capability and examining its required computing time relative to system size. In its present form, the algorithm can be used to evaluate a large percentage of a given network. With additional improvements, it may be possible to achieve substantial savings in computing time making the algorithm suitable for all subsystem sizes.

Chapter 6: Conclusions

6.1 Contributions of This Dissertation

This dissertation provides a method to conduct highly robust, real-time state estimation of electric power systems. The effect of system sparsity on members of a particular family of estimators is examined resulting in the development of new theorems decribing their exact fit point for regression problems where the general position assumption is violated. Having obtained a full understanding of the local exact fit phenomenon, a system decomposition scheme is developed using efficient graph-theoretical algorithms resulting in increased outlier identification capability and decreased computing times.

The number of outliers that the algorithm is able to identify in each subsystem is based on proofs, not simulations. There are no underlying measurement error probability distribution assumptions to be violated, no unwarranted measurement weights assumed, and the method is inherently resitant to bad measurements in positions of leverage. For the first time, it is possible to reliably identify multiple interacting outliers even if their errors are conforming.

Extensive research has gone into the practical aspects of implementing the proposed algorithm. From subsystem identification to elemental set generation, only theoretically supportable methods have been selected. Efficiency of each step has also been ensured. The algorithm is ideally suited for parallel processing. This fact can prove to be especially helpful for systems containing large cyclic subsystems. Since the algorithm is based upon system decomposition, it can be used on selected portions of a network if desired. The fact that the algorithm is based on a decomposed network has the

added benefit of limiting the effect outliers in one region have on the state estimate of other regions.

In the process of arriving at the proposed algorithm, a better understanding of the statistical challenges associated with the power system state estimation problem was achieved. Reliable outlier identification in power systems is very challenging due to (i) relatively low measurement redundancy levels, (ii) system sparsity, (iii) huge systems involving a variety of measuring devices which are not reliably installed, tested, or maintained, and (iv) the presence of leverage points. The proposed algorithm should be viewed as an additional tool by the power system engineer interested in state estimation. It, like its predecessors, has some disadvantages and is not the sole answer to the problem.

A major contribution of this research is the development/furtherance of tests which allow the quantitative analysis of a system to be made from an outlier identification standpoint. Based on the results of these analyses and a knowledge of the characteristics of various estimators, one can decide upon an overall strategy for obtaining a robust state estimate of a particular system.

For example, the ability to identify leverage points and to compute robust measurement weights based upon projection statistics is a powerful tool. Measurements in positions of leverage are of particular concern when using the WLAV, so it may be advisable to use the proposed decomposition/HBPE algorithm in these areas. Evaluating the measurement surplus in the neighborhood of leverage points determines the viability of this approach. For an operator at a control center, knowledge of leverage point locations and local measurement supluses is valuable information. It lets him know which sections of the system are more likely to have poor state estimates. Therefore, if strange results are obtained from the state estimator, he can make an educated guess as to where in the system estimator breakdown has occured.

These same diagnostic tools can be used in developing a meter placement methodology. Increasing redundancy in regions around leverage points should be a high priority if the area surplus is too low.

From a theoretical standpoint, the only way to ensure reliable identification of outliers is to significantly increase the level of measurement redundancy thereby providing a sufficient measurement surplus for a HBPE to be used. This is not a realistic option due to the huge financial investment this would entail. Several options which may offer a way out of this predicament are noted in the closing section of future research opportunities.

6.2 Future Research

A number of topics related to this dissertation deserve further investigation. Some specific areas of interest include:

- Quantify the effect of sparsity on the LAV estimator.

- Conduct simulations using parallel processing to ascertain the actual CPU time savings obtainable.

- Investigate the effect of different system embeddings on outlier identification capability.

- Develop a methodology for optimal meter placement in concert with leverage point identification and measurement surplus calculations.

- Investigate faster state solution methods such as QR-decomposition updates of the Jacobian matrix.

- Experiment with various Schweppe-type redescending GM-estimators.

- Evaluate the improvements in statistical efficiency gained by post-LMS/LTS estimation with the WLS or a Schweppe-type one-step GM estimator.

- Consider lower quantile index values for the HBPE used. Perhaps the worst case outlier error scenario is too restrictive and can be relaxed.

Appendix A

(Theorem Proofs for Reduced Position Models)

Proof of Theorem 3.1:

Consider a collection of m measurements $(z_i, \mathbf{l}_i)$. Replace f of them by outliers. Suppose that the remaining m - f measurements, which are good ones, lie exactly on an n-dimensional hyperplane $\mathcal{P}_0$. Since we are restricting ourselves to regression equivariant estimators, we may assume without loss of generality that $\mathcal{P}_0$ satisfies z = 0, implying that the good measurements are defined as $(0, \mathbf{l}_i)$. Suppose that there exists a regression equivariant estimator with an exact fit point δ* which is larger than [s*/2]/m. Put $\delta^* = \left(\left[s^*/2\right]+1\right)/m$. This would mean that for any f = [s*/2] + 1 outliers, the hyperplane $\mathcal{P}_0$ is unique and contains the majority of the measurements, which are all good. Their number is

$$m - f = m - \left(\left[s^*/2\right]+1\right) = m - \left[(s^*+2)/2\right].$$

Consider the (n - 1) - dimensional hyperplane $\mathcal{P}_2$ that contains the maximum number M of measurements. Assume they are all good. This implies that $\mathcal{P}_2$ is contained in $\mathcal{P}_0$, i.e., $\mathcal{P}_2 \subset \mathcal{P}_0$. It also follows that there are f = [s*/2] + 1 outliers among the remaining s* + 1 measurements. Let us place these outliers on an n-dimensional hyperplane $\mathcal{P}_1$, which is different from $\mathcal{P}_0$. $\mathcal{P}_1$ can always be chosen so that its intersection with $\mathcal{P}_0$ is $\mathcal{P}_2$, namely $\mathcal{P}_2 = \mathcal{P}_0 \cap \mathcal{P}_1$. Hence, in addition to the f outliers, $\mathcal{P}_1$ contains M good measurements, yielding a total of $M + f$ measurements. This number equals

$$M + f = m - s^* - 1 + \left[s^*/2\right] + 1 = m - \left(s^* - \left[s^*/2\right]\right) = m - \left[(s^*+1)/2\right].$$

Now since

$$\left[(s^*+1)/2\right] \le \left[(s^*+2)/2\right],$$

we have

$$m-\left[(s^*+1)/2\right] \ge m-\left[(s^*+2)/2\right].$$

It follows that $\mathcal{P}_1$ contains the majority of the measurements, a contradiction. Therefore, $\delta^* \le \left[s^*/2\right]/m$. □

Proof of Theorem 3.2:

By virtue of theorem 3.1, the exact fit points of LMS and LTS cannot be greater than [s*/2]/*m*, for these estimators are regression equivariant. Now let us show that the exact fit points δ* of LMS and LTS are no smaller than [s*/2]/*m*. Consider *m* good measurements that lie on a hyperplane $\mathcal{P}_0$. Let us replace [s*/2] of these measurements by outliers. Therefore, *m* - [s*/2] good measurements remain on $\mathcal{P}_0$. It follows that, when the fit is $\mathcal{P}_0$, the sum of the smallest ordered squared residuals up to the rank ν has a zero value, because $\nu = m-\left[(s^*+1)/2\right] \le m-\left[s^*/2\right]$. By definition of the surplus, there exists at least one subset $\mathcal{S}^*$ with s* measurements who's deletion from the network turns at least one remaining measurement into a critical one. If there are several subsets $\mathcal{S}^*$, pick any one of them. Suppose now that the outliers are chosen from $\mathcal{S}^*$. Let us place them on a hyperplane $\mathcal{P}_1$ passing through the largest number *M* of measurements that lie on an (n - 1)-dimensional subspace, where *M* = *m* - s* - 1. The total number of measurements contained in $\mathcal{P}_1$ is

$$(m - s^* - 1) + [s^*/2] = m - (s^* - [s^*/2]) - 1$$

which equals

$$m - [(s^* + 1)/2] - 1.$$

Since this number is equal to v - 1, it results that the vth ordered squared residual will be different from zero when the fit is $\mathcal{P}_1$. This also holds for any hyperplane passing through any group of [s*/2] outliers out of the m measurements, for $\mathcal{P}_1$ is the one that contains the largest contaminated set of measurements; it leads to the largest number of zero residuals among these bad hyperplanes. Consequently, it is only $\mathcal{P}_0$ that has the smallest value for the LMS and LTS objective functions, which is zero. Since it contains only good measurements, we conclude that the LMS and LTS can handle at least [s*/2] outliers. □

Appendix B
(Topics From Graph Theory)

Introduction

Efficient graph-theoretical algorithms have been developed to solve a wide variety of problems which, at first glance, appear purely combinatorial. For example: (1) given a group of workers with different skills and a set of jobs each requiring a subset of the skills, assign as many qualified workers to a job as possible; and (2) given a large meshed piping system where each pipe has a capacity limitation, determine the maximum flow the system can sustain. Familiarity with applied graph theory makes it possible to formulate many seemingly insurmountable problems in such a way that *good* graphical algorithms can be used. A good algorithm is one for which the number of computational steps required for its implementation on any graph G is bounded above by a polynomial in the number of edges and vertices of G. As such, graph theory has already found applications in the power systems engineering field. An excellent example is the topological observability algorithm of Nucera and Gilles [54] in which augmenting sequences are used to determine injection measurement assignments after formulating the problem as a maximum intersection of two matroids.

The decomposition methodology presented in this dissertation is based on material presented in Section 3.3.2. Implementation of the scheme makes use of a variety of a efficient graph-theoretic algorithms including a modified version of the observability algorithm mentioned above, a bridge identification algorithm [55], and a planarity testing algorithm [56]. This chapter presents a sufficient amount of background material to allow a discussion of the major algorithms used. The particulars of some of these algorithms are also presented to illustrate their efficiency and exactness. Additional concepts and terms related to graphs which are utilized in subsequent chapters are also defined. Theorems

which are deemed relevant to the purpose of the chapter are noted. See [55, 58, 59] for further information and proofs of the theorems.

Fundamentals

A graph G is an ordered triple (V(G), E(G), ψ_G) consisting of a nonempty set V(G) of *vertices* (or nodes), a set E(G), disjoint from V(G), of *edges*, and an incidence function ψ_G that associates with each edge of G an unordered pair of vertices of G. If e is an edge and u and v are vertices such that $\psi_G(e) = uv$, then e is said to *join* u and v. The two joined vertices are said to be *adjacent*. The incidence relations can be indicated via a diagram in which each pair of adjacent vertices are connected by a line. For our purposes, the buses and transmission lines of a power system's one-line diagram are viewed as a graph in which V(G) corresponds to the system buses, E(G) corresponds to the transmission lines, and ψ_G is indicated by the diagram.

Graphs can be represented in a computer by several means. The method used throughout this dissertation is via *adjacency lists*. The adjacency list of G associates with each vertex v a list of its adjacent vertices Adj(v). Figure B.1a shows the IEEE 5-bus system which will be referred to as G'. The adjacency lists of G' are given in Table B.1 following the figure.

A graph with just one vertex is called the *trivial* graph. The *degree* of a vertex v is the number of vertices adjacent to v. The minimum of all vertex degrees is denoted by δ. If $e = uv$ is an edge of G, we say that e and u (and e and v) are *incident* to each other. If e and f are distinct edges that are incident to a common vertex, then e and f are *adjacent edges*. An edge which has both ends incident to the same vertex is called a *loop*. A *simple* graph is one which has no loops and no parallel or *multiple* edges.

Two graphs G and H are *identical* or *equal* (G = H) if V(G) = V(H), E(G) = E(H), and $\psi_G = \psi_H$. *Isomorphic* graphs (G ≅ H) are graphs for which there is a one-to-one function ϕ from V(G) to V(H) such $uv \in E(G)$ if and only if $\phi(u)\phi(v) \in E(H)$. This simply

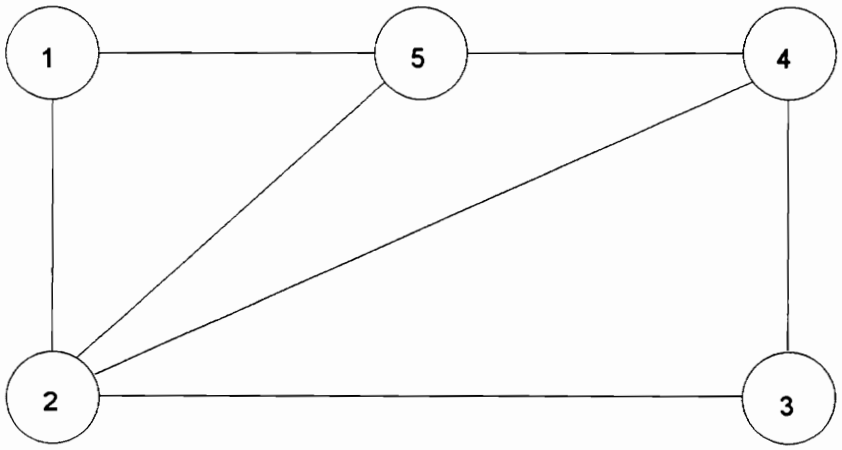


Figure B.1a IEEE 5-Bus System Topology

(Graph G')

Table B.1 Adjacency Lists for Graph G'

Vertex *v*	Adj(*v*)
1	2, 5
2	1, 3, 4, 5
3	2, 4
4	2, 3, 5
5	1, 2, 4

means that G can be obtained from the drawing of H and vice versa. A graph isomorphic to G' is shown in Figure B.1b.

A graph H is a *subgraph* of G if $V(H) \subseteq V(G)$ and $E(H) \subseteq E(G)$. This is denoted by $H \subseteq G$. If S is a nonempty subset of V(G), then the subgraph of G whose vertex set is S and whose edge set consists of those edges of G that have both ends in S is called the subgraph of G *induced* by S. Such a subgraph is denoted by G(S). The *induced subgraph* G - S is the subgraph obtained by deleting the vertices in S together with their incident edges. Figure B.2 illustrates a subgraph of G' induced by the set $S = \{v_3\}$.

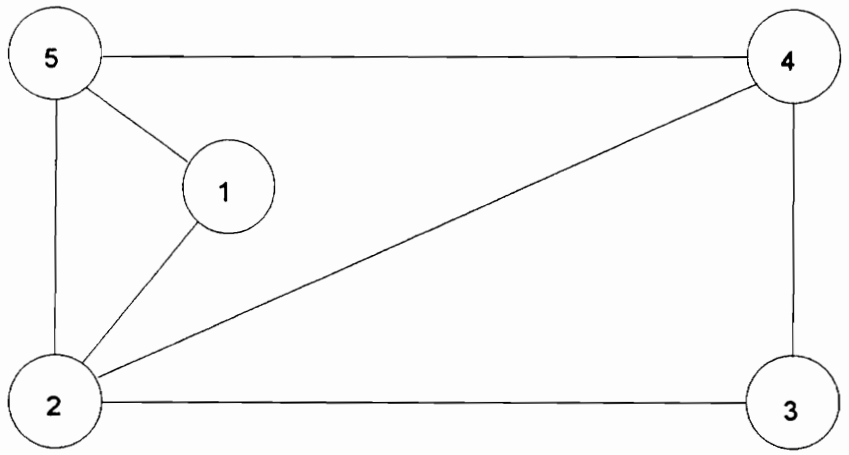


Figure B.1b Isomorphic Graph of G'

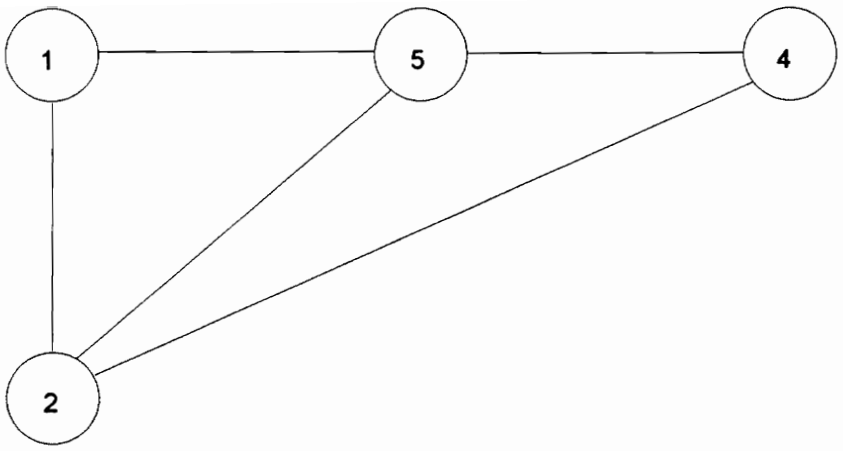


Figure B.2 Induced Subgraph of G'

A *walk* is a finite non-null sequence $W = v_0 e_1 \ldots e_k v_k$ whose terms are alternately vertices and edges, such that, for $1 \leq i \leq k$, the ends of e_i, are v_{i-1}, and v_i. Such a walk is termed a (v_0, v_k)-walk where v_0 is called the *origin* and v_k the *terminus*. The integer k is the *length* of the walk. If the edges of a walk are distinct, W is called a *trail*. If the vertices are also distinct, W is called a *path*. A path of length greater than three for which the origin and terminus are the same is called a *cycle*.

Two vertices u and v are said to be *connected* if there exists a (u, v)-path in the graph G containing u and v. One can partition V(G) into nonempty subsets $V_1, V_2, \ldots, V_\omega$ such that two vertices are connected if and only if the two vertices belong to the same set V_i. The subgraphs $G(V_i)$ are called the *components* of G. The number of components of G is denoted by $\omega(G)$. If G has only one component it is said to be connected, otherwise it is *disconnected*.

A vertex v is a *cut-vertex* in a connected graph G if G - v is disconnected. A maximal connected graph with no cut-vertices is called a *block*. The edge analog of a cut-vertex is a *cut-edge* or *bridge*, i.e., an edge e in a connected graph G is a bridge if G - e is

disconnected. Figure B.3 shows a graph G for which v_1 and v_6 are cut-vertices and v_1v_6 is a bridge. Removal of v_1v_6 disconnects G into two components. In this case, both components are blocks and are also cycles.

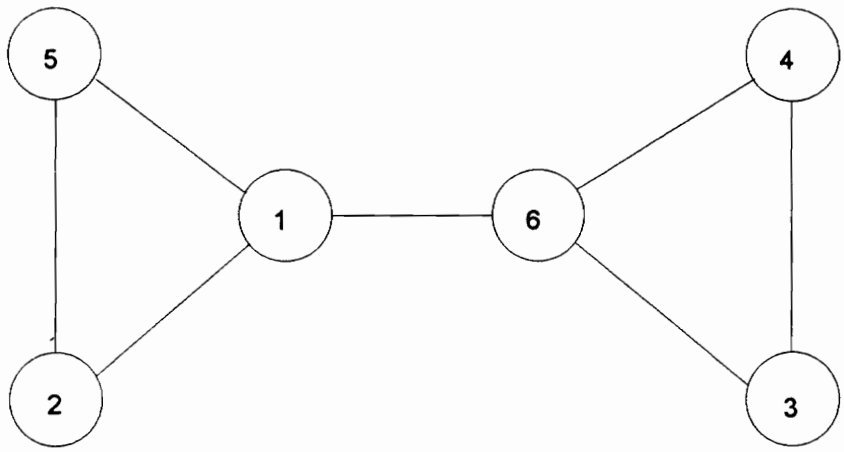


Figure B.3 Graph G

Theorem 1: An edge *e* of a connected graph G is a bridge of G if and only if *e* does not lie on a cycle of G.

Thus, graphs which do not contain bridges are graphs for which every edge lies on a cycle.

The minimum number of vertices which must be deleted from a connected graph G to disconnect it is called the *connectivity* $\kappa(G)$ of G . (Note that the removal of a vertex necessitates the removal of its incident edges as well). Similarly, the edge-connectivity $\kappa'(G)$ of G is the minimum number of edge deletions required to disconnect G. It is useful to note that $\kappa(G) \leq \kappa'(G) \leq \delta$. A graph with $\kappa(G) \geq 2$ is said to be *2-connected* or *biconnected*. In general, a graph with $\kappa(G) \geq k$ is said to be *k-connected*.

A graph which contains no cycles is called an acyclic graph. A *tree* is defined as a connected acyclic graph. Some texts call a tree incident to all vertices a *spanning tree*. An edge appearing in a tree is called a *branch*. The number of branches contained in a tree is equal to |V(G)| - 1. Trees are useful when defining the topological observability of a power system. A system provided with a set of measurements is topologically observable if and only if a spanning tree can be formed in the graphical representation of the system for which each branch is associated with a distinct system measurement. The tree symbolizes a number of linearly independent equations (due to the lack of cycles) which relate the system unknowns represented by vertices. The algebraic equivalent to such a tree is a Jacobian matrix of full rank. A spanning tree of G' is given in Figure B.4.

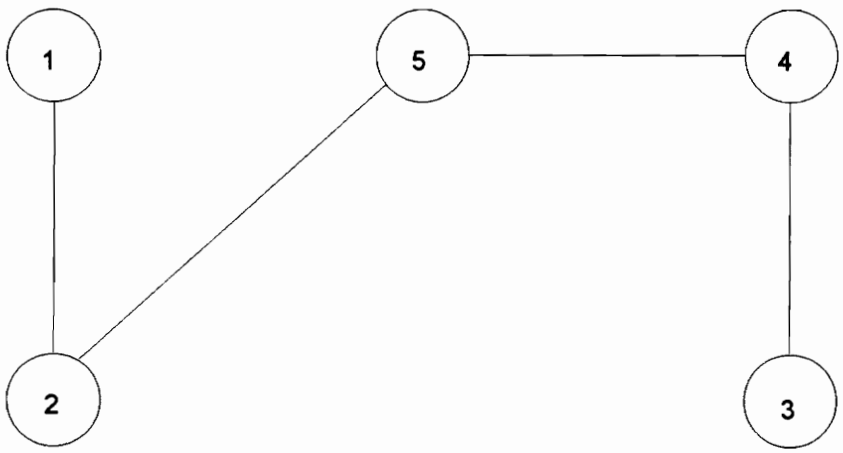


Figure B.4 Spanning Tree of G'

A *rooted tree* is a directed tree (a tree whose branches are directed edges, i.e. the vertices incident to the edge is an ordered pair) in which there exists a vertex r of T, called the root, such that for every vertex v of T, there is an r - v path in T. A vertex x in a rooted tree with root r is at *level i* if and only if the r - x path in T has length i. If a vertex

v of T is adjacent to u and u lies in the level below v, then u is called a *child* of v, and v is the *parent* of u. A vertex w is a *descendant* of v, and v is an *ancestor* of w, if the v - w path in T lies below v. A rooted tree of G' is illustrated in Figure B.5 where the root is v_1. The level of each vertex is noted next to the corresponding vertex. Vertex v_2 is a child of v_1 (v_1 is a parent of v_2) and v_3 is a descendant of v_1 (v_3 is an ancestor of v_1).

A graph that can be drawn in the plane without any edges intersecting is called a *planar graph*. A graph that is so drawn is said to be *embedded* in the plane and is called a *plane graph*. The connected pieces of the plane that remain when the vertices and edges of G are removed are called the *regions* or *faces* of G. All plane graphs have one unbounded region called the *exterior region*. For every region R in a plane embedding of G, there exists another embedding in which R is the exterior region. In other words, a planar embedding is not necessarily unique. The vertices and edges of G that are incident with a region R form a subgraph of G called the *boundary* or *window* of R. In a

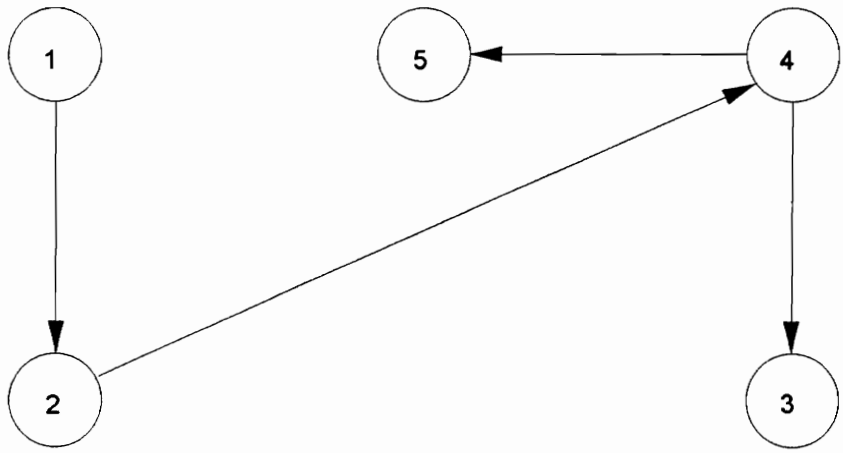


Figure B.5 Rooted Tree of G'

biconnected planar graph, each such window is a simple cycle. A *maximal planar graph* is a plane graph for which the addition of another edge makes it nonplanar. The regions of the plane graph G' are indicated in Figure B.6. R_4 is the external region in this particular embedding. The number of regions (ϕ) a connected planar graph has can be obtained from Euler's polyhedral formula,

$$|V(G)| - |E(G)| + \phi = 2.$$

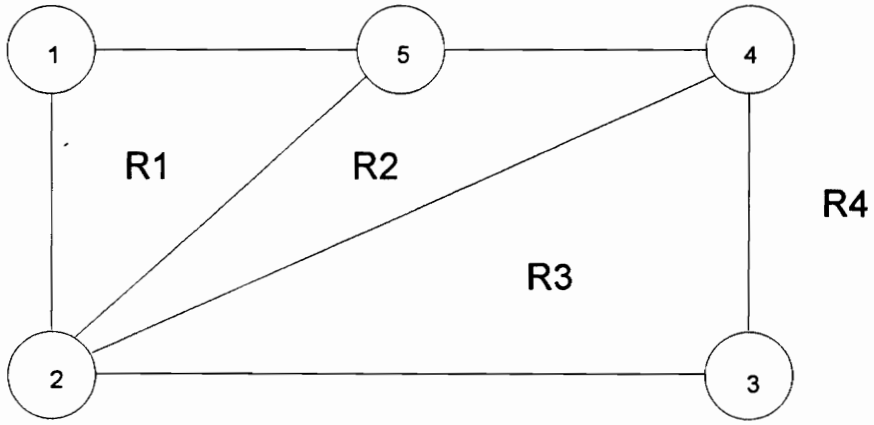


Figure B.6 A Planar Embedding of G'

Algorithms of Interest

Bridge Identification

For reasons that are explained in Section 4.2, the first part of the decomposition algorithm consists of identifying the bridges of the graph G representing the power system of interest. The algorithm is based on a *depth-first search* scheme and has a complexity of O(max{|V|, |E|}).

A depth-first search (DFS) [59] is a systematic method of visiting the vertices of a graph and serves as a basis for many graph algorithms. Given a graph G with V(G) = $\{v_1, ..., v_p\}$ represented by its adjacency lists, a DFS proceeds as follows. One vertex is chosen as the *root* or starting vertex. A vertex that is currently being visited is designated the *active vertex*, thus the root is the first active vertex. It is assigned the label 1. Next, we select a vertex adjacent to the root and set it as the active vertex, giving it the label 2. The edge connecting these two vertices is placed in a set S.

In general, let n denote the label of the current vertex in the search. Assuming that not all vertices in the component of G containing n have been visited, the DFS proceeds as follows. If there are unvisited vertices adjacent to n, select the first vertex on the adjacency list of n that has not been visited and label it with the next sequentially increasing label. This vertex becomes the active vertex and the edge just traversed is placed in the set S. If all vertices adjacent to n have been visited, we backtrack to the vertex visited prior to n and make this vertex our active vertex. This process is repeated until all vertices in G have been labeled. If they have not all been visited, G is not connected and a new root is chosen from the next component to be explored.

The label assigned to a vertex v by the DFS is called the DFS search index of v, and is denoted by dfi(v). When the DFS of G is completed, the number dfi(v) is the order in which v was first visited during the search. If G is connected, the set of edges, S, forms a spanning tree T. Each edge not contained in S is a *back edge*. These edges are indicated pictorially by broken lines.

To identify bridges in G, we need one additional term. The *lowpoint* l(u) of a vertex v is the smallest value dfi(u) of a vertex u that can be reached from v by a directed v-u path consisting of branches of T followed by at most one back edge. Figure B.7 gives

the results of a DFS of the graph of Figure B.3 along with the lowpoint values associated with each vertex.

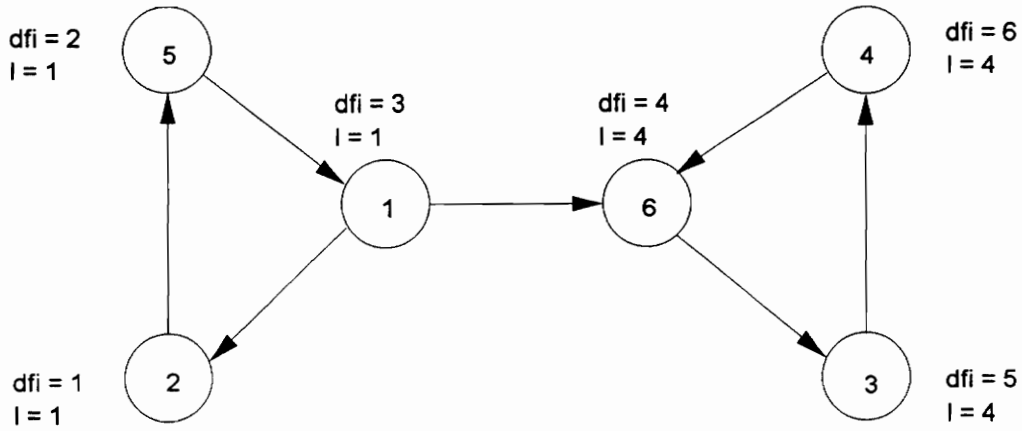


Figure B.7 Results of DFS of Graph of Figure B.3

Theorem 2: Let T be a DFS tree of a connected graph G. Then an edge uv of G, with $\text{dfi}(u) < \text{dfi}(v)$, is a bridge of G if and only if uv is an edge of T and $l(v) > \text{dfi}(u)$.

This theorem gives us a characterization of bridges that allows their identification. Based upon this criteria, examination of Figure B.7 shows that v_1v_6 is the only bridge in the graph. Note that the removal of all bridges from a graph G results in a set of biconnected components.

Planarity Testing

Two linear time planarity testing algorithms exist [57, 60]. Both are quite lengthy and complex to describe. A significant part of their explanation centers around efficient data structures which is not really the point of this section. Although not as efficient as

the linear algorithms, [57] has been implemented for the purposes of this dissertation. It is still a good algorithm with $O(|V^4|)$ complexity and is simpler to describe. The following presentation will afford the reader some familiarization with the conceptual aspects of planarity testing. Since a connected graph is planar if and only if each of its blocks is planar, subsequent discussions of planarity testing assumes that the test is being applied to a single block.

Let H be a subgraph of G. Define a relation ~ on E(G) - E(H) by $e \sim f$ if there exists a walk W in G - E(H) whose first and last edges are e and f, and where no internal vertex of W belongs to H. The subgraphs of G - E(H) induced by the resulting equivalence classes are called the *fragments* of H in G. The fragments of G' associated with a particular H are shown in Figure B.8. The subgraph H is depicted by solid lines, while the fragments are shown by dotted lines. Fragments are identified using a DFS.

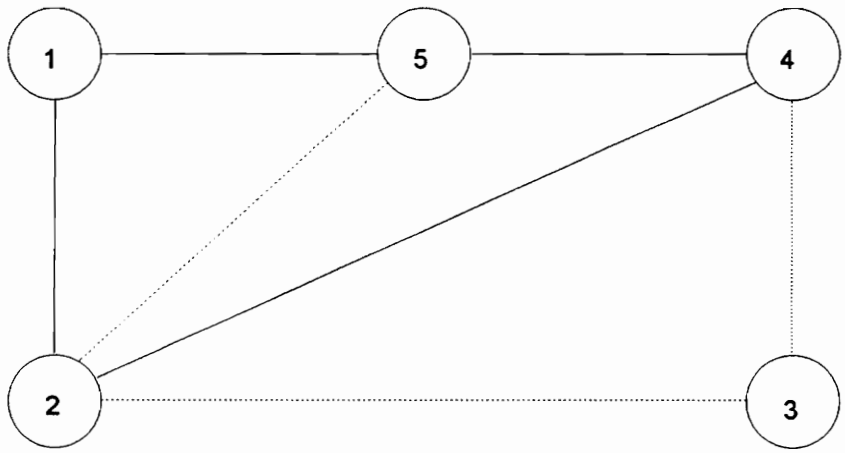


Figure B.8 Fragments Associated With a Subgraph of G'

A few more terms need to be defined before the planarity testing algorithm is presented. A subgraph H is *G-extendable* if G is planar and the plane embedding of H can

be extended to a plane embedding of G. Let H be a plane subgraph of G, and let R be a region of H. An *R-fragment* is a fragment F whose vertices belonging to H all lie on the boundary of R. The set of regions for which F is an R-fragment is denoted by R(F, H). If a plane subgraph H of a planar graph G is G-extendable, $R(F, H) \neq 0$ for every fragment of H in G. This simply means that every fragment of H can be added to H to obtain G.

The planarity testing algorithm of Demoucron, Malgrange, and Pertuiset [56] is presented below in generic form. Comments concerning each step are made where appropriate. For this listing, let $q = |E|$ and let $p = |V|$ for G(V, E). Two examples will be used to clarify the algorithm. First, the planarity of G' will be confirmed. Secondly, the nonplanarity of $K_{3,3}$ is determined. The examples are presented after the algorithm listing which is as follows.

1. [*Determine if the graph can possibly be planar based on the relationship between the number of vertices and number of edges.*]

If $q > 3p - 6$ stop; the graph cannot be planar since the relationship given for a maximal planar graph violated.

2. [*Determine an initial plane subgraph* G_1 *of* G]

Find a cycle G_1 of G and embed the cycle in the plane. Let R_1 and R_2 be the respective interior and exterior regions of G_1. The boundary of both R_1 and R_2 is, therefore, the cycle G_1. Mark all the edges of G_1 as used and let $i = 1$, and $q' = m$ where m is the number of edges in the cycle G_1.

3. [*At this point, a plane subgraph* G_i, *with regions* R_1, R_2, ..., R_{i+1} *has been determined. The boundary of every region is a cycle. This step determines if* $G_i = G$.]

IF $q - q' = 0$, stop, since G is planar. Otherwise, continue.

4. Determine the fragments of G_i in G.

5. [*For every fragment* F_k *of* G_i *in* G, *this step determines all those regions* R *of* G_i *for which the boundary of* R *contains* $V(F_k) \cap V(G_i)$.]

For each fragment F_k of G_i in G, determine the set $R(F_k, G_i)$.

6. If there is a fragment F_k of G_i in G for which $R(F_k, G_i) = 0$, stop; since G_i is not G-extendable. Otherwise, continue.

7. [*If there is a fragment* F *of* G_i *in* G for *which there is only one choice of a region in which* F *can be embedded, then this step chooses such a fragment first.*]

If there exists a fragment F_k for which $|R(F_k, G_i)| = 1$, let $R = R_j$, where $R(F_k, G_i) = \{R_j\}$. Otherwise, let $R = R_j$, for some $R_j \in R(F_k, R(F_k, G_i)$.

8. [*This step determines a plane graph* G_{i+1} *that contains* G_i *as a proper subgraph.*]

8.1 Determine a path P in F_k that connects two vertices of G_i.

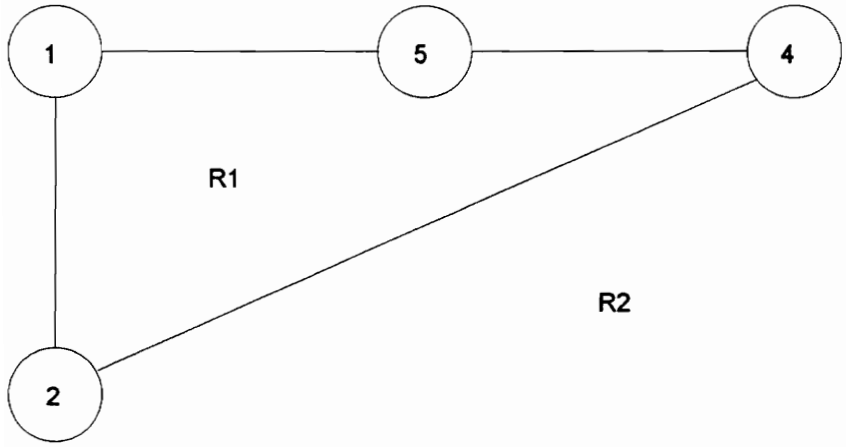
8.2 The path P splits the region containing the two vertices into two regions. Update the boundaries for these regions as follows. Let be the boundary cycle of $R = R_j$ be $v_{j1},..., v_{jl}$. Suppose P is a (v_{jr}, v_{js})-path given by $v_{jr} = x_1, x_2,..., x_l = v_{js}$. Let R_j be the region with boundary $v_{jr}, v_{jr+1},..., x_{l-1}, x_{l-2}, x_1$, and let R_{i+2} be the region whose boundary is $v_{js}, v_{js+1},..., v_{jr}, x_2, x_3,..., x_l$. All other boundaries remain unchanged. Mark all the edges of P used, let $i = i + 1$, $q' = q' + l - 1$ and return to step 3 letting G_{i+1} denote the plane subgraph obtained by drawing P in R.

Example 1: Application of algorithm to G'

After passing the test of step 1, a cycle G_1 is found (using a DFS) which divides the plane into two regions R_1 and R_2 as indicated in Figure B.9a where G_1 is shown along with its fragments, boundaries, and set of regions for which each fragment is an R-fragment. Fragment F_1 is arbitrarily chosen for embedding, resulting in G_2 as shown in Figure B.9b. There is only one fragment associated with this new graph, so it is then embedded in the plane resulting in the graph G_3 given in Figure B.9c. Since all edges have now been embedded, the graph is determined to be planar. The embedding is given by the region boundaries as indicated.

Example 2: Application of algorithm to $K_{3,3}$.

The graph known as $K_{3,3}$ is shown in Figure B.10a. Following the steps of the algorithm, we obtain the graphs and associated information of Figure B.10b-d. Once the graph G_3 is obtained, we note that the fragment F_6 cannot be embedded because the two vertices v_3 and v_6 do not have a region in common implying that adding v_3v_6 will necessarily intersect with at least one of the edges of G_3. Since the last fragment cannot be embedded, we correctly conclude that the graph $K_{3,3}$ is nonplanar.



F_1: $\{v_2, v_5\}$

F_2: $\{v_2, v_3, v_4\}$

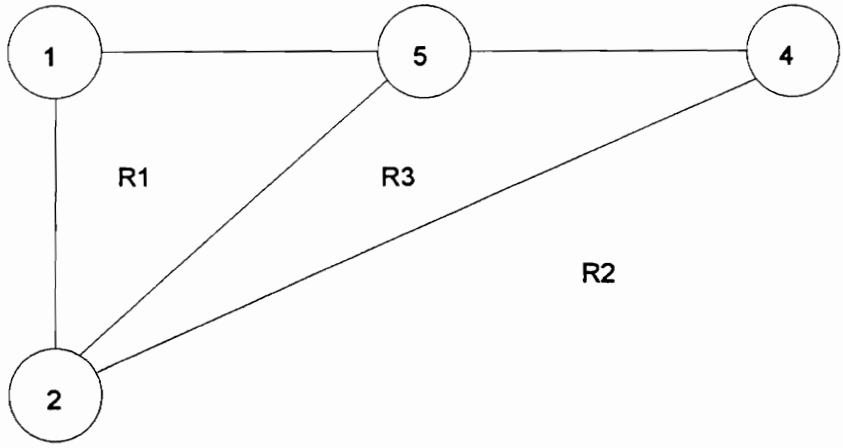
R_1 has boundary G_1

R_2 has boundary G_1

$R(F_1, G_1) = \{R_1, R_2\}$

$R(F_2, G_1) = \{R_1, R_2\}$

Figure B.9a: Graph G_1 of Example 1



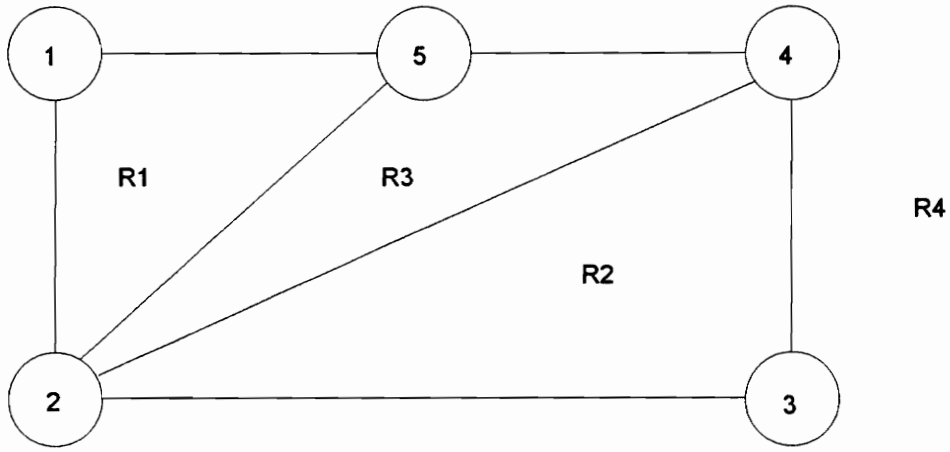
F_3: $\{v_2, v_3, v_4\}$

R_1 has boundary $\{v_1, v_2, v_5\}$

R_2 has boundary G_1

R_3 has boundary $\{v_2, v_4, v_5\}$

Figure B.9b: Graph G_2 of Example 1



R_1 has boundary $\{v_1, v_2, v_5\}$

R_2 has boundary $\{v_2, v_3, v_4\}$

R_3 has boundary $\{v_2, v_4, v_5\}$

R_4 has boundary G_1

Figure B.9c: Graph G_3 of Example 1

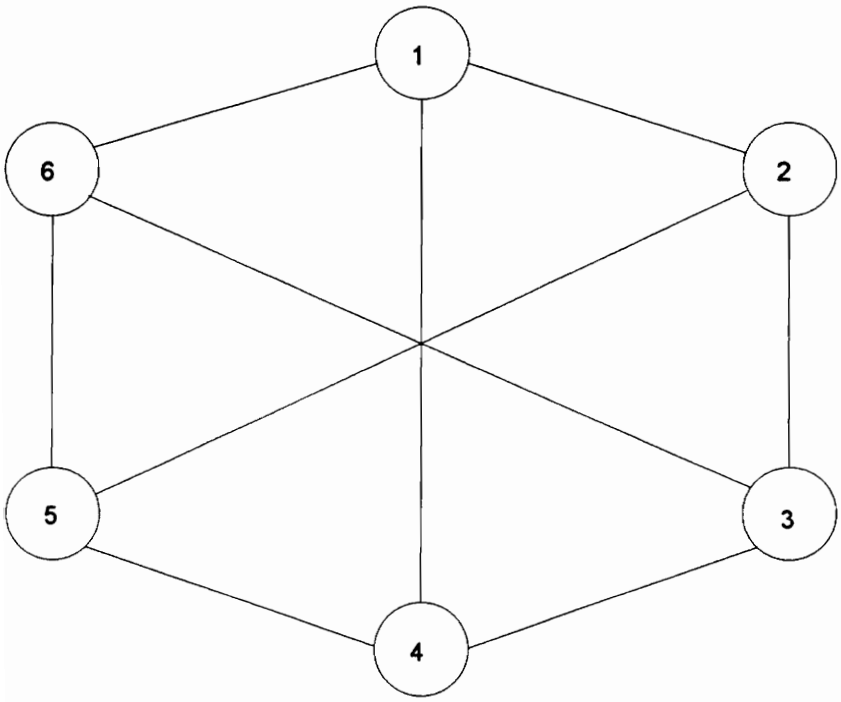
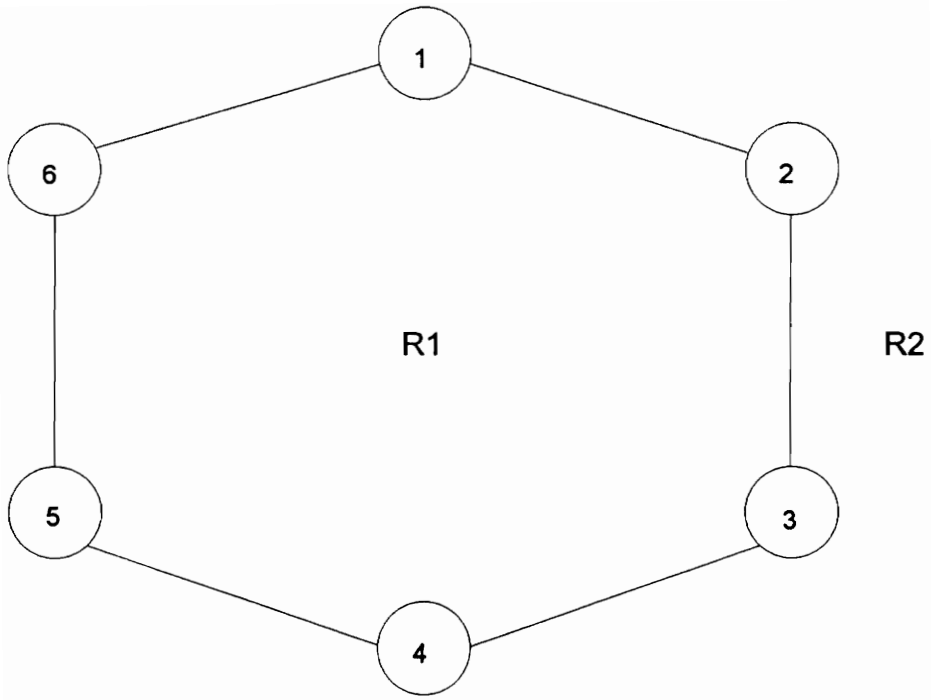


Figure B.10a: Graph $K_{3,3}$ for Example 2



F_1: $\{v_1, v_4\}$

F_2: $\{v_2, v_5\}$

F_3: $\{v_3, v_6\}$

R_1 has boundary G_1

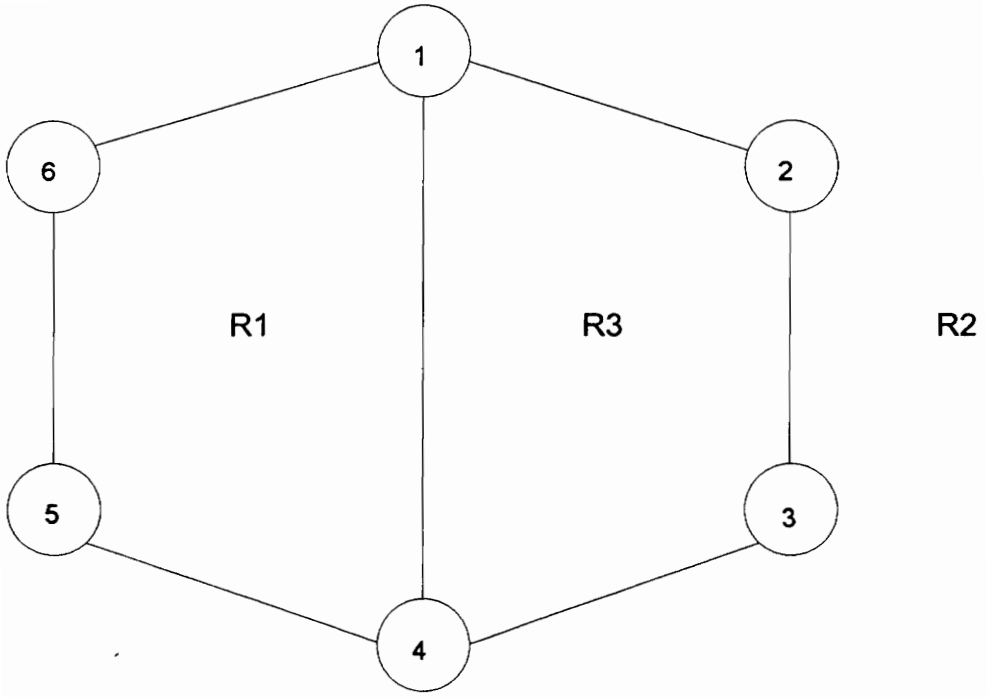
R_2 has boundary G_1

$R(F_1, G_1) = \{R_1, R_2\}$

$R(F_2, G_1) = \{R_1, R_2\}$

$R(F_3, G_1) = \{R_1, R_2\}$

Figure B.10b: Graph G_1 of Example 2



F_4: $\{v_2, v_5\}$

F_5: $\{v_3, v_6\}$

R_1 has boundary $\{v_1, v_4, v_5, v_6, v_1\}$

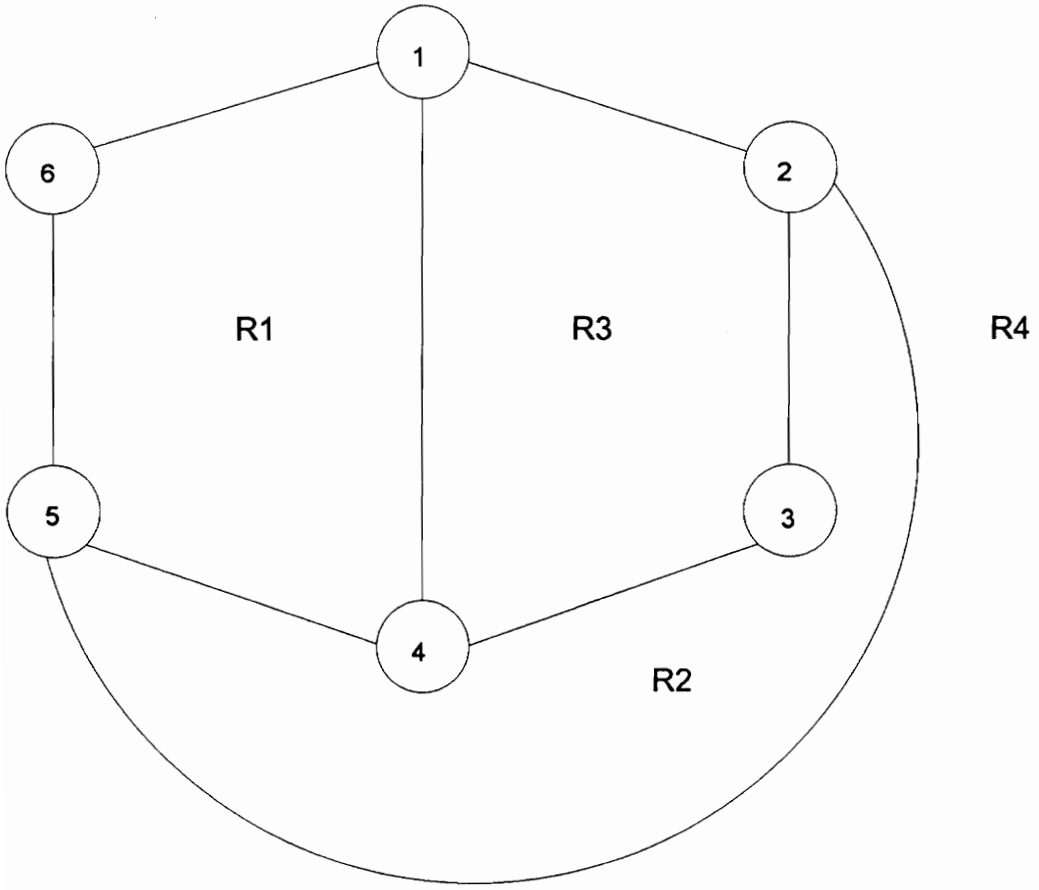
R_2 has boundary G_1

R_3 has boundary $\{v_1, v_2, v_3, v_4, v_1\}$

$R(F_4, G_2) = \{R_2\}$

$R(F_5, G_2) = \{R_2\}$

Figure B.10c: Graph G_2 of Example 2



F_6: $\{v_3, v_6\}$

R_1 has boundary $\{v_1, v_4, v_5, v_6, v_1\}$

R_2 has boundary $\{v_2, v_3, v_4, v_5, v_2\}$

R_3 has boundary $\{v_1, v_2, v_3, v_4, v_1\}$

R_4 has boundary $\{v_1, v_2, v_5, v_6, v_1\}$

$R(F_6, G_3) = \{0\}$

Figure B.10d: Graph G_3 of Example 2

Bibliography

[1] A.G. Phadke, "Synchronized Phasor Measurements in Power Systems", *IEEE Computer Applications in Power*, Vol. 6, No. 2, April 1993, pp. 10-15.

[2] M.M. Adibi and D.K. Thorne, "Remote Measurement Calibration," *IEEE Transactions on Power Systems*, vol. PWRS-1, May 1986, pp. 194-203.

[3] M.M. Adibi and J.P. Stoval, "On Estimation of Uncertainties in Analog Measurements," *IEEE Transactions on Power Systems*, vol. 5, Nov. 1990, pp. 1222-1230.

[4] L.S. Van Slyck and J. Allemong, "Operating Experience with the AEP State Estimator", *IEEE Transactions on Power Systems*, Vol. 3, No. 2, May 1988, pp. 521-528.

[5] B. Heckman, A. Jamshidi, L. Radu, G. Romero, R. Smith, and W. Snyder, "State Estimator - Practical Experience", 1993 Winter Power Meeting, Columbus, OH.

[6] M. Assadian, R.J. Goddard, H.W. Hong, and D. French, "Field Operational Experiences with on Line State Estimator", *IEEE Transactions on Power Systems*, Vol. 9, No. 1, February 1994, pp. 50-58.

[7] F.C. Schweppe, J. Wildes, "Power System Static State Estimation, Part I: Exact Model", *IEEE Transactions of Power Apparatus and Systems*, Vol. PAS-89, pp. 120-125, January 1970.

[8] F.C. Schweppe, D.B. Rom, "Power System Static State Estimation, Part II: Approximate Model", *IEEE Transactions of Power Apparatus and Systems*, Vol. PAS-89, pp. 125-130, January 1970.

[9] F.C. Schweppe, "Power System Static State Estimation, Part III: Implementation", *IEEE Transactions of Power Apparatus and Systems*, Vol. PAS-89, pp. 130-135, January 1970.

[10] E. Handschin, F.C. Schweppe, J. Kohlas, and A. Fiechter, "Bad Data Analysis for Power System State Estimation", *IEEE Transactions of Power Apparatus and Systems*, Vol. PAS-94, No. 2, March/April 1975, pp.329-337.

[11] H-J Koglin, D. Oeding, and K.D. Schmitt, "Local Redundancy and Identification of Bad Measuring Data", *Proc. 9th PSCC, Cascais*, 1987, pp. 528-534.

[12] A. Monticelli, F.F. Wu, and M. Yen, "Multiple Bad Data Identification for State Estimation By Combinatorial Optimization", *IEEE Transactions of Power Delivery*, Vol. PWRD-1, No. 3, July 1986, pp. 361-369.

[13] L. Mili, Th. Van Cutsem, and M. Ribbens-Pavella, "Hypothesis Testing Identification: A New Method for Bad Data Analysis in Power System State Estimation", *IEEE Transactions on Power Apparatus and Systems*, Vol. PAS-103, No. 11, November 1984, pp. 3229-3252.

[14] A. Monticelli and A. Garcia, "Reliable Bad Data Processing for Real-time State Estimation", *IEEE Transactions of Power Apparatus and Systems*, Vol. PAS-102, No. 5, May 1993, pp. 1126-1139.

[15] H-J Koglin, Th. Neisius, G. Beißler, and K.D. Schmitt, "Bad Data Detection and Identification", *Electrical Power & Energy Systems*, Vol. 12, No. 2, April 1990, pp. 94-103.

[16] L. Mili, V. Phaniraj and P.J. Rousseeuw, "Least Median of Squares Estimation in Power Systems," *IEEE Transactions on Power Systems*, vol. 6, pp. 511-523, May 1991.

[17] M.K. Celik and A. Abur, "A Robust WLAV State Estimator Using Transformations", *IEEE Transactions of Power Systems*, Vol. 7, No. 1, February 1992.

[18] L. Mili, N.S. Vichare, M.G. Cheniae, P.J. Rousseeuw, "Robust Mahalanobis Distances in Power System State Estimation," *Proceedings of the 6th Midwest Symposium on Circuits and Systems*, Detroit, MI, August 1993.

[19] L. Mili, M.G. Cheniae, N.S. Vichare and P.J. Rousseeuw, "Leverage Point Identification in Power System State Estimation", accepted for presentation at the IEEE PES Winter Meeting, New York, Jan., 1995.

[20] R.A. Maronna, O. Bustos, and V. Yohai, "Bias- and Efficiency-Robustness of General M-estimators for Regression With Random Carriers", in *Smoothing Techniques for Curve Estimation*, ed. T. Gasser and M. Rosenblatt, Springer Verlag, pp. 91-116.

[21] L. Mili, M.G. Cheniae, and P.J. Rousseeuw, "Robust State Estimation of Electric Power Systems", *IEEE Transactions on Circuits and Systems I: Fundamental Theory and Applications*, Vol. 41, No. 5, May 1994, pp. 349-358.

[22] I. Barrodale and F.D.K. Roberts, "Solution of an Overdetermined System of Equations in the L_1 Norm", *Communications of the ACM*, Vol. 17, No. 6, June 1974, pp. 319-320.

[23] V. Phaniraj, "Robust State Estimation in Power Systems", Ph.D. Dissertation, Virginia Polytechnic Institute and State University, November 1991.

[24] P.J. Rousseeuw and A.M. Leroy, *Robust Regression and Outlier Detection*, John Wiley, 1987.

[25] R.A. Fisher, "On The Mathematical Foundations of Theoretical Statistics", *Philos. Trans. Roy. Soc. London Ser. A*, **222**, 1922, pp. 309-368.

[26] J.W. Tukey, "A Survey of Sampling From Contaminated Distributions", *Contributions to Probability and Statistics*, I. Olkin (ed.) Stanford University Press, Stanford, CA, pp. 448-485.

[27] P.J. Huber, "Robust Statistical Procedures: Regional Conference", *Series in Applied Mathematics*, No. 27, Soc. Industr. Appl. Math., Philadelphia, Pa. 1977

[28] P.J. Huber, *Robust Statistics*, John Wiley, 1981.

[29] P.J. Huber, "Robust Estimation of a Location of Parameter", *Ann. Math. Stat.*, 35, 1964, pp. 73-101.

[30] J.L Hodges, Jr., "Efficiency in Normal Samples and Tolerance of Extreme Values for some Estimates of Location," *Proceedings of the 5th Berkeley Symposium on Mathematical Statistics Statistics and Probability*, vol. 1, University of California Press, Berkeley and Los Angeles, pp. 163-168, 1967.

[31] F.R. Hampel, "A General Qualitative Definition of Robustness," *The Annals of Mathematical Statistics*, vol. 42, pp. 1887-1896, 1971.

[32] D.L. Donoho and P. J. Huber, "The Notion of Breakdown Point," in *A Festschrift for Eric L. Lehmann*, eds. P. Bickel, K. Doksum and J.L. Hodges, Jr., Belmont, Calif.:Wadsworth.

[33] D.L. Donoho, "High Breakdown Made Simple, subtalk presented at the Oberwolfach Conference on Robust Statistics, West Germany, 9-15 September, 1984.

[34] C.L. Mallows, "On Some Topics in Robustness", unpublished memorandum, Bell Telephone Laboratories, Murray Hill, NJ

[35] R.W. Hill, "Robust Regression When There Are Outliers in the Carriers", unpublished Ph.D. dissertation, Harvard University, Boston, MA, 1977.

[36] P.J. Rousseeuw, "Least Median of Squares Regression," *Journal of the American Statistical Association*, vol. 79, pp. 871-880, Dec. 1984.

[37] P.J. Rousseeuw, "Multivariate Estimation With High Breakdown Point," paper presented at the *Fourth Pannonian Symposium on Mathematical Statistics and Probability*, Bad Tatzmannsdorf, Austria, September 4-9, 1983. Appeared in, *Mathematical Statistics and Applications, Vol. B*, edited by W. Grossmann, G. Pflug, I. Vincze, and W. Wertz, Reidel, Dordrecht, The Netherlands, pp. 283-297.

[38] P.J. Rousseeuw and V.J. Yohai, "Robust Regression by Means of S-estimators", in *Robust and Nonlinear Time Series Analysis*, J. Franke, W. Härdle, and R.E. Martin (eds), Lecture Notes in Statistics, No. 26, Springer Verlag, 1984, pp. 256-272.

[39] V.J. Yohai, "High Breakdown Point and High Efficiency Robust Estimates for Regression", *Ann. Statist.*, **15**, 1987, pp. 642-656.

[40] V.J. Yohai and R. Zamar, "High Breakdown-point Estimates of Regression by Means of the Minimization of an Efficient Scale", *J. Amer. Statist. Assoc.* **83**, 1988, pp. 406-413.

[41] D.G. Simpson, D. Ruppert, and R.J. Carroll, "On one-step GM Estimates and Stability of Inferences in Linear Regression", *J. Amer. Statist. Assoc.*, **87**, 1992, pp. 439-450.

[42] C.W. Coakley and T.P. Hettmansperger, "A Bounded Influence, High Breakdown, Efficient Regression Estimator", *J. Amer. Statist. Assoc.*, **88**, 1993, pp. 872-880.

[43] L. Mili and C.W. Coakley, *Robust Estimation in Structured Linear Regression, Technical Report No. 93-13*, Department of Statistics, Virginia Polytechnic and State University, Blacksburg, VA 24061, 1993.

[44] C.W. Coakley and L. Mili, "Exact Fit Points Under Simple Regression With Replication," *Statistics and Probability Letters*, vol. 17, pp. 265-271, 1993.

[45] C. Croux and P.J. Rousseeuw, "Time-efficient Algorithms for Two Highly Robust Estimators of Scale," *Computational Statistics*, Vol. 1, eds. Y. Dodge and J. Whittaker, Heidelberg: Physica-Verlag, 1992, pp. 411-428.

[46] J.W. McKean, S.J. Sheather and T.P. Hettmansperger, "The Use and Interpretation of Residuals Based on Robst Estimation", *J. Amer. Statist. Assoc.*, **88**, December 1993, pp. 1254-1263.

[47] L. Mili, M.G. Cheniae, N.S. Vichare and P.J. Rousseeuw, "Leverage Point Identification in Power System State Estimation", submitted to IEEE PES Summer Meeting, July 1994.

[48] P.J. Rousseeuw and B.C. van Zomeren, "Unmasking Multivariate Outliers and Leverage Points," *Journal of the American Statistical Association*, vol. 85, pp. 633-651, Sept. 1990.

[49] M. Gasko and D. Donoho, "Influential Observation in Data Analysis", *American Statistical Association, Proceedings of the Business and Economic Statistics Section*, 1982, pp. 104-110.

[50] D.L. Donoho and M. Gasko, "Breakdown Properties of Location Estimates Based on Halfspace Depth and Projected Outlyingness", *The Annals of Statistics*, Vol. 20, No. 4, 1992, pp. 1803-1827.

[51] P.J. Rousseeuw and C. Croux, "Alternatives to the Median Absolute Deviation", *Journal of the American Statistical Association*, Vol. 88, No. 424, Dec. 1993, pp. 1273-1283.

[52] N.S. Vichare, "Robust Mahalanobis Distance in Power Systems State Estimtion", Ph.D. dissertation, Virginia Polytechnic Institute and State University, January, 1993.

[53] W.H. Press, W.T. Vetterling, S.A. Teukolsky, and B.P. Flannery, *Numerical Recipes in Fortran, The Art of Scientific Computing*, Cambridge University Press, 1992.

[54] R.R. Nucera and M.L. Gilles, "Observability Analysis: A New Topological Algorithm", *IEEE Transactions of Power Systems*, Vol. 6, No. 2, May 1991, pp. 466-475.

[55] G. Chartrand and O.R. Oellermann, *Applied and Algorithmic Graph Theory*, International Series in Pure and Applied Mathematics, McGraw-Hill, 1993.

[56] G. Demoucron, Y. Malgrange, and R. Pertuiset, "Graphes Planaires: Reconnaissance et Construction de Représentation Planaires Topologiques", *Rev. Francaise Recherche Opérationnelle*, **8**, 1964, pp. 33-47.

[57] J.A. Bondy and U.S.R. Murty, *Graph Theory with Applications*, North-Holland, 1976.

[58] S. Even, *Graph Algorithms*, Computer Science Press, 1979.

[59] Tarjan, "Depth-First Search and Linear Graph Algorithms", *SIAM J. Comput.*, Vol. 1, 1972, pp. 146-160.

[60] Hopcroft and R.E. Tarjan, "Efficient Planarity Testing", *J. Assoc. Comput. Mach*, Vol. 21, No. 4, pp.549-568, 1974.

[61] Lempel, S. Even, and I. Cederbaum, "An Algorithm for Planarity Testing of Graphs", *Theory of Graphs*, International Symposium, Rome, July 1966 (P. Rosenstiel Ed.), pp. 215-232, Gordon & Breach, 1967.

[62] Hawkins, "The Accuracy of Elemental Set Approximations for Regression", *Journal of the Americal Statistical Association, Theory and Methods*, Vol. 88, No. 422, pp. 580-589, June 1993.

[63] G.R. Krumpholz, K.A. Clements and P.W. Davis, "Power System Observability: A Practical Algorithm Using Network Topology", *IEEE Transactions on Power Apparatus and Systems*, Vol. PAS-99, No. 4, July/August 1980, pp. 1534-1542.

[64] Chiba and T. Nishizeki, "A Linear Algorithm for Embedding Planar Graphs Using PQ-Trees", *Journal of Computer and System Sciences*, Vol. 30, pp. 54-56, 1985.
[65] Nijenhuis and H.S. Wilf, *Combinatoial Algorithms for Computers and Calculators*, Academic Press, 1978.
[66] Costa, T.S. Piazza, and A. Mandel, "Qualitative Methods to Solve Qualitative Problems in Power System State Estimation", *IEEE Transactions of Power Systems*, Vol. 5, No. 3, pp. 941-949, August1990.
[67] Hubbi and F. Shirazi, "A Line Flow Only Power System State Estimator", *Electric Power Systems Research*, Vol. 19, pp. 229-237, 1990.
[68] J.F. Dopazo, O.A. Kiltin and L.S. Van Slyck, "State Calculation of Power Systems From Line Flow Measurements", *IEEE Transactions on Power Systems, Part II*, PAS-91, 1972, pp. 145-151.
[69] G. Zorpette, "The Power of Parallelism", *IEEE Spectrum*, pp. 28-33, September 1992.
[70] Coakley, L. Mili, and M.G. Cheniae, "Effect of Leverage on the Finite Sample Efficiencies of High Breakdown Estimators", *Statistics and Probability Letters*, Vol. 19, 1994.
[71] Stoll, *Least-Cost Electric Utility Planning*, John Wiley & Sons, Inc., 1989.
[72] Koglin, D. Oeding, and K.D. Schmitt, "Local Redundancy and Identification of Bad Measuring Data", *Proceedings of the Ninth Power Systems Computation Conference* Cascais, Portugal, 30 August to 4 September 1987, pp. 528-534.

Vita

Michael G. Cheniae was born in Baltimore, Maryland on October 17, 1960. He graduated from the U.S. Naval Academy in May 1982 with a B.S. in electrical engineering. After completing the Navy's nuclear propulsion training program, he served as an officer in the nuclear submarine force for seven and a half years. He then resigned his commission and returned to school, receiving an M.S. in electrical engineering in June, 1991.