ROBUST STATE ESTIMATION IN POWER SYSTEMS
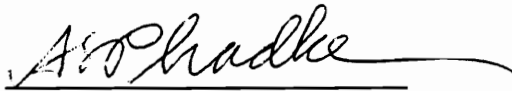
by

Viruru Phaniraj

Dissertation submitted to the Faculty of the

Virginia Polytechnic Institute and State University

in partial fulfillment of the requirements for the degree of
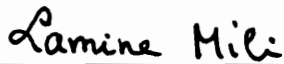
DOCTOR OF PHILOSOPHY

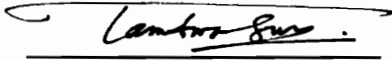in

Electrical Engineering

Approved :

A. G. Phadke, Chairman

L. M. Mili (Advisor)

S. Rahman

K. S. Tam

J. B. Birch

November, 1991

Blacksburg, Virginia

ROBUST STATE ESTIMATION IN POWER SYSTEMS

By

Viruru Phaniraj

Committee Chairman: A.G. Phadke

Electrical Engineering

(ABSTRACT)

The application of robust estimation methods to the power system state estimation problem was investigated. Techniques using both nonlinear and combinatorial optimization were considered, based on the requirements that the method developed should be statistically robust, and fast enough to be used in a real-time environment.

Some basic concepts from robust statistics are introduced. The various estimation methods considered are reviewed, and the implementation of the selected estimator is described. Simulation results for several IEEE test systems are included. Other applications of the proposed technique, such as leverage point identification in large sparse systems, and robust meter placement are described.

Acknowledgements

Table of Contents

List of Tables

List of Illustrations

Chapter 1 : Introduction

This dissertation is based on the premise that the predominant method used currently in control centers for state estimation, the Weighted Least Squares estimator (and derivatives thereof), is not adequate. Due to its ease of implementation, and high speed, this method has continued to be in wide use, despite its many drawbacks. Though alternate methods have been proposed, they have either lacked demonstrably superior statistical properties, or have been too slow to be of practical significance. In this dissertation, concepts from robust estimation theory will be introduced, in order to demonstrate systematically the weaknesses of the Least Squares and related methods. Based on this theoretical foundation, alternate estimators will be described, which possess much more desirable statistical properties than those currently used. Since the use of these estimators is in the real-time control of power systems, the implementation of these techniques to achieve maximum speed is crucial, if they are to be competitive with the Least Squares. Hence computational issues such as sparsity, parallel programming and hierarchical estimation will be addressed. The use of some of these robust methods to related subjects such as meter placement and leverage point identification will be described.

In Chapter 2, the state estimation problem is defined in terms of its reduction from a nonlinear power system model to a more tractable linearized form.

Some of the practical difficulties encountered in solving this problem in a real-time environment are described, especially in the context of how these influence the choice of an estimation algorithm. The conventional methods that have been used in control centers since the introduction of practical state estimators in the early 1970's are discussed, and the more recent attempts to address the deficiencies in these methods are described. This chapter serves as a review of the state of the art, and defines the problems and issues that this dissertation seeks to resolve.

The main reason that some of the recent developments have not been fruitful is that they have occurred without a sufficient knowledge of the progress made by statisticians in formulating a theory of robust estimation. These advances have created a strong theoretical foundation for describing the robustness and other properties of estimators. Great improvements have occurred, starting with the work of Huber, Hampel, and Tukey. In Chapter 3, the basic tenets of robust estimation theory will be described, and these tools will be used to analyze the weaknesses in the methods that were discussed in Chapter 2. Several terms which have a specific meaning in the context of robust estimation, such as the term robust itself will be defined, and concepts such as the breakdown point of an estimator, the influence function, and leverage point will be introduced here.

Chapter 4 describes a new class of estimators that have been developed recently, namely the High Breakdown Point Estimators. The theoretical properties that make these estimators attractive for use in power system state estimation will be discussed, as well as the potential problems that could be encountered in implementing these methods in large sparse

systems. Some discussion of the issues to be faced in implementation of these methods will be given, along with algorithms for both direct nonlinear minimization and combinatorial optimization.

Chapter 5 discusses the actual implementation of the Least Median of Squares estimator, with particular focus on making this method competitive (in terms of computational speed) with the Least Squares based estimators that are prevalent currently. The use of speedup techniques such as sparse matrix methods, parallel processor utilization, and the division of the task into on-line and off-line portions will be described. Test results from various IEEE test systems will be presented, especially with cases that have multiple interacting bad data. These will be contrasted with results for the same cases obtained using other estimation methods ; this will be used to illustrate the vastly superior robustness properties of the High Breakdown Point Estimators.

The issue of leverage points in power systems is the subject of Chapter 6. The definitions and concepts introduced in Chapter 3 will be used to show that leverage points are common in power system state estimation, and that without the use of appropriate precautions, they can cause the Least Absolute Value (LAV) estimator to fail. This estimator is becoming more popular, as it seems to combine the speed of the Least Squares estimator with some of the robustness properties of the High Breakdown Point Estimators. Methods for detection and identification of leverage points based on projection pursuit techniques have been devised for use in large, highly sparse systems. These provide a robust estimate of the Mahalanobis distance for such systems. Results from the application of these methods to the IEEE

test systems will be given, and these methods will be validated by testing them using the statistical programming language S-PLUS, and through Monte Carlo simulations of the LAV estimator.

Chapter 7 contains a summary of the contributions of this dissertation towards advancing the state of the art in power system state estimation. The conclusions reached as an outcome of the work described will be reviewed, both in terms of their immediate application, and their potential use as the computing power in the control center increases. As a result of this work, there are several other research topics related to state estimation that can benefit from the use of these robust methods. These include robust meter placement, hierarchical state estimation by the decomposition of large power systems into smaller sections, and alternate estimation methods based on incorporating the robust Mahalanobis distance estimates of Chapter 6 into the state estimation procedure. Chapter 7 will also discuss these issues, some of which are the subject of ongoing research at Virginia Tech. Some possible directions of future research in the area of robust power system state estimation will also be suggested.

Chapter 2 : Overview of Power System State Estimation

2.1 : Introduction

A. C power systems began as a single generator in each city serving the local customers, with the generation, transmission and distribution of power all occurring at the same voltage. With the invention of the transformer in the 1880's, it became possible to generate the power at the most economical voltage, typically 13.2 kV or less, and then to step up the voltage (usually to the 345 kV - 765 kV range) by a transformer connected to the transmission line. The voltage magnitude is reduced by transformers at the remote end of the transmission line to the level required by the loads, which ranges from 120 V for households, to 13.2 kV for industrial establishments. This development was primarily responsible for the predominance of a.c systems over d.c systems, since no d.c equivalent of the transformer existed. It also allowed power systems to become much larger, since the generating plants could now be located near the resources they required (rivers for cooling, or near coal mines for thermal plants, etc). Since then, a.c power systems have evolved into a huge interconnection of generators, transmission lines and loads. The entire continental U.S, for example, is served by three large synchronous interconnections. With this complex network in place, it became

increasingly difficult for a utility to monitor its system, which could have components located over a thousand miles from each other. Early power systems used decentralized control, where each plant operator made decisions based on local conditions ; this later evolved into a system-wide control center, where operators kept in touch with the generating plants over the phone, and co-ordinated the control of the system. The growth of power systems, especially the dramatic boom after World War II made this method of monitoring the system impractical, and utilities were quick to use computers to aid them in this process as soon as they became available. The control center became a hub where real time measurements from the remote ends of the system were received, and processed by the computers, which then provided the data for the operator to analyze. Some aspects of power system operation became highly automated, such as the load-frequency control and the economic dispatch of generators for minimizing the cost of operation.

Though these control centers were a great improvement over the prior ones, they still were not adequate. There was no method of synthesizing a global picture of the state of the system by combining all the measurements ; instead the knowledge of one area of the system was determined by measurements from that area alone. This left the process vulnerable to failure, since if some of those measurements were corrupt, there was no way to counteract this using the other measurements that were available. This situation was acceptable until events in the 1960's made it essential to improve the secure operation of the system. Analytical tools were developed which required as their input the "state" of the entire system (the state of

the system is the set of all bus voltages and phase angles) ; these tools were used to forecast loads, identify equipment being operated beyond ratings etc. Therefore it became important to devise a technique, which could utilize all the measurements available at the control center, and combine them to produce a snapshot of the system state at that instant. This process is called power system state estimation.

Power system state estimation became a topic of active research interest after the blackouts of the 1960's, especially the Northeast blackout of 1965. With the improvement in SCADA systems it became possible to gather a large number of real-time measurements at the control center, and it was envisioned that this database could be used to greatly enhance system security, as well as improve other aspects of power system operation [1]. It soon became evident to some researchers that though this database was a vast improvement over that previously available, it could not be used directly as is. The database typically consisted of the breaker status information and the real power measurements used in Automatic Generation Control ; while most security and optimization programs required as their primary input the state of the power system. What was required then was a procedure to convert the available measurements into the form desired ; in essence, a real time loadflow was needed.

The conventional steady state loadflow assumes that the injection at each load bus is known accurately, and that the generator bus voltages are known. From this it determines the state of the system; the procedure is deterministic in that the number of unknown variables is equal to the number of equations. There is no redundancy in the input data; if one of the

assumed values is inaccurate, the result will be wrong. This approach cannot be used with the real time loadflow, for several reasons. Due to the measurements being taken at remote sites and telemetered to the control center, the possibility of communication errors introducing bad data was very real. Since the output of the state estimator would be used in real-time for vital system control operations, the usual procedure of checking the data and output for plausibility would not be sufficient. It would not be fast enough, and considering the increasing size and complexity of power systems, would be humanly impossible. The other reason redundancy is required is the nature of the power system. A conventional single-phase loadflow is based on the assumption that the three-phase power system can be represented by a single-phase equivalent, in effect assuming that the system is balanced, that the π-circuit model used is accurate, and that the line parameters are known. When the output of the loadflow is used for planning studies and simulations, the consequence of small errors in the output is not immediate, nor is it drastic. This is not the case in state estimation, where the operator takes action to ensure the secure and reliable operation of the system based on the results, and where control actions are taken immediately after the results become available. It was clear then that a new method had to be developed to perform the function of the real time loadflow ; it should be based on having some redundancy in the input data, it should use the existing measurements and supplement them with as few additional meters as necessary (due to the cost of the equipment, and more importantly the cost of the communication links to the control center).

2.2 : Formulation of the Problem

The measurement set that is typically available consists of bus voltage magnitude measurements, line power flow measurements (real and reactive), and bus injection measurements (real and reactive). These are related to the desired outputs, the bus voltages and angles, by a set of simultaneous nonlinear equations, as derived below. Let the transmission line between bus i and bus j be represented by its π-circuit equivalent, which is the complex series impedance $R_{ij} + j\, X_{ij}$, and half the shunt admittance $(B_{ij}/2)$ at nodes i and j. The real power flow P_{ij} and reactive power flow Q_{ij} between bus i and bus j are then given by equations (1) and (2)

$$P_{ij} = \frac{|V_i|^2}{Z} \sin\delta_{ij} + \frac{|V_i|\,|V_j|}{Z} \sin(\theta_{ij} - \delta_{ij}) \tag{1}$$

$$Q_{ij} = \frac{|V_i|^2}{Z} (\cos\delta_{ij} - \frac{B_{ij}}{2}) - \frac{|V_i|\,|V_j|}{Z} \cos(\theta_{ij} - \delta_{ij}) \tag{2}$$

where $Z\ \angle\delta_{ij}$ is the polar representation of the impedance $R_{ij} + j\, X_{ij}$, V_i and V_j are the voltage magnitudes at nodes i and j respectively, and θ_{ij} is the phase angle difference across the line. The injection at bus i can be written in terms of the sum of all the flows out of bus i, due to Kirchoff's law, while the voltage magnitude measurements are direct measurements of a component of the state vector. (Until recently, it was not possible to measure the bus phase angles directly. An ongoing project at Virginia Tech, the phasor measurement project, has made this feasible, and some implications of this development will be addressed in a later chapter). Thus if we define the vector of measurements as **z,** of length m (the total number of

measurements), and the state vector $\mathbf{x}$ (of length 2n - 1, where n is the number of busses, since one bus is taken as the reference for the phase angles), then the system can be modeled by the equation $\mathbf{z} = h(\mathbf{x})$, where $h(\mathbf{x})$ represents the set of equations just described. The number of measurements is usually at least 2 times n, so that there is adequate redundancy. This is a nonlinear estimation problem, to determine the state vector $\hat{\mathbf{x}}$ that fits the measurements $\mathbf{z}$ optimally.

As mentioned before, there is a great deal of commonality between the loadflow and the state estimation problem. This extends to the use of the fast decoupled loadflow concepts in state estimation. Essentially, as can be seen from (1) and (2) , there is a relatively weak connection between the real power measurements and the voltage magnitudes, and between the reactive power measurements and the voltage angles. The larger estimation problem can then be divided into two smaller segments, which provides a significant computational advantage. Throughout the remainder of this dissertation, we will be using the decoupled form of the state estimator.

2.3 : Sources of Bad Data

Practical experience with operating state estimators [2] has shown that at almost every run (typically, every 15 minutes) , there is some data present that does not match the model. Van Slyck et al report that over a period of 13 years, only 5 days were completely free of bad data, with an average of between 1 and 2 % of all measurements being so classified. These discordant measurements are generally called bad data, though this term can be misleading. The implication that there is something "wrong" with the data is not necessarily true ; there are many situations in which a measurement can

be perfectly accurate, and yet be classified as bad by the estimator. In this section, some of the common causes for bad data will be reviewed, since ultimately, these provide the greatest motivation for the development of the estimators described later in this dissertation.

The most common cause of bad data is that the data is actually bad. The measurement that reaches the control center has gone through several stages of processing, each of which provides an opportunity for the introduction of an error. The first step in the measurement process is the transducer that interfaces directly to the transmission line ; either a current or a voltage transformer. These are susceptible to errors such as mis-calibration, bias, drift due to temperature variations, rounding errors, as well as device failures [3]. A more subtle source of error at this stage is the timing skew; this occurs when the data is not sampled simultaneously throughout the system. Since the overall measurement set is assembled at the control center, and is assumed to be a snapshot of the system at some instant ; if some measurements are taken at a different time than others, the picture of the system presented will not be uniform. There are enormous difficulties, both financial and technical, in synchronizing all the measurements. The phasor project mentioned earlier requires much more accurate synchronization, and solves this problem by using the Global Positioning System satellites. There have been proposals to implement a system-wide timing signal using a fiber-optic communication system at some utilities, but this is not yet economically viable.

The next stage in the collection of the measurements is the transmission of the raw data from the substations to the control center. The most common

source of error here is the noise present in the communication channels ; this effect is lessening as digital transmission technology with error-correcting codes is becoming more prevalent, but the problem persists in analog channels. Some of the larger utilities use data concentrators, which act as regional control centers, which collect the data from a large part of the system, and forward it to the control center (for example, AEP uses four data concentrators, one of which is in Roanoke). Any malfunction at these intermediate sites will impact significant portions of the data.

The other reason for discrepancies between the data and the model is inadequacies in the model. The mathematical model that is at the heart of the state estimation process embodies a series of assumptions, most of which are usually accurate. The accumulated effect of these errors however may result in the measured data not fitting the model. Some of these assumptions are :

(1) The numerical values used for various system elements in the model are subject to error. There are minor variations in line impedance due to temperature ; and in other cases the data is just not known accurately. In addition, some portions of the system model are actually a reduced equivalent of the external power system, which is based on other approximations, and uses data whose current value cannot be confirmed easily, if at all.

(2) All existing state estimation algorithms use a linearized version of the nonlinear system model, in order to reduce the computational burden. If the number of iterations used is large enough, this does not cause any error. It can result in numerical convergence problems if the

system is heavily loaded.

(3) State estimators generally use a single-phase model of the power system, which implicitly assumes a balanced 3-phase system. This is not usually true, due to factors such as the instantaneously varying load, and the presence of untransposed transmission lines.

(4) Topology errors, which are the worst kind of error to detect. An operating power system is not static ; lines are switched in and out, transformer tap settings are changed, static var systems and switched capacitors dynamically change the shunt admittance of the power system. The control center may not always be aware of these changes in time to respond to them.

In general, the errors due to approximations of one kind or another result in mild outliers, while device failures and mis-calibration cause gross errors.

2.4 : The Weighted Least Squares Estimator

The first such method was described by Schweppe *et al* in their seminal series of papers [4] in 1970, which introduced the concept of the state estimator being a filter which determined the state of the system as a weighted function of the input measurements. These papers lay the foundation for most of the work on static state estimation in the subsequent two decades. Essentially, the measurements and the state of the power system were assumed to be related by the nonlinear equation

$$\mathbf{z} = h(\mathbf{x}) + \mathbf{e} \tag{3}$$

where $\mathbf{z}$ is the measurement vector, $\mathbf{x}$ is the state vector and $\mathbf{e}$ is the measurement error vector. In the case of a conventional load flow, $\mathbf{e}$ is

assumed to be null, and the dimensions of $\mathbf{x}$ and $\mathbf{z}$ are equal. In state estimation however, the number of measurements m, is greater than the number of unknowns, n ; with m/n typically being about 2. At this point, a critical assumption was made which led to the computation of the state estimate being made much faster, and the analysis of the statistical properties of the estimator much easier. This was the assumption that the components of the error vector $\mathbf{e}$, $(e_1, e_2, \dots e_m)$ are independently distributed random variables following a Gaussian distribution with zero mean. Then the measurement error covariance matrix $\mathbf{R}$, which is the expected value of $[\mathbf{e}.\mathbf{e}^t]$ is a diagonal matrix, with its entries as

$$\mathbf{R} = \text{Diag} (\sigma_1^2, \ \sigma_2^2, \ \sigma_3^2, \dots \sigma_m^2) \tag{4}$$

From this the probability density function of $\mathbf{z}$ can be written as

$$f(\mathbf{z}) = \frac{\exp\{ -0.5 \ [\mathbf{z}\text{-}h(\mathbf{x})]^t \ \mathbf{R}^{-1} \ [\mathbf{z}\text{-}h(\mathbf{x})]\}}{(2\pi)^{m/2}(\det \mathbf{R})^{0.5}} \tag{5}$$

where det $\mathbf{R}$ is the determinant of R. The question is then to find an "optimal" $\mathbf{x}$. One of the accepted methods of doing this is the maximum likelihood estimator approach, where we seek to maximize the likelihood that the observed measurements were the values actually observed. That is, if the observed measurements are ($z_1, z_2, \dots, z_m$) , then the likelihood function

$$L = f(z_1).f(z_2).f(z_3). \ \dots \ .f(z_m) \tag{6}$$

is to be maximized. Since $\mathbf{R}$ is independent of the actual value of $\mathbf{z}$, this is equivalent to maximizing the numerator of (5), or minimizing the magnitude of the exponent term. The problem then reduces to minimizing the quadratic

function which is a classic least squares problem. If we now define $\mathbf{H(x)}$) as $d(h(\mathbf{x}))/d\mathbf{x}$, this is equivalent to finding $\hat{\mathbf{x}}$ so that $\mathbf{H}^t(\hat{\mathbf{x}}). \mathbf{R}^{-1}. [\mathbf{z}-h(\hat{\mathbf{x}})] = 0$ (this is also equivalent to minimizing $\sum_{i=1}^{m} r_i^2$, which is why this method is known as the Least Squares Estimator). $\mathbf{H}$ is then by definition the Jacobian matrix of the measurements. The problem now has been reduced to one very similar to that in conventional load flow problems, where the problem is to find $\mathbf{x}$ such that a nonlinear function of the Jacobian matrix is as close to zero as possible. This makes the use of all the computational enhancements that were devised for the solution of the load flow problem applicable to the state estimation problem. In particular the Newton Raphson iterative algorithm for the solution of nonlinear equations, and the sparse matrix techniques described by Tinney et al [5,6] can be applied directly. Since the objective function being minimized is globally convex, and has only one minima, the choice of the starting point for the state estimation iterations is important only for the speed of the solution. Regardless of the starting point, the final solution will always be the same ; this is not true of some of the estimators that will be discussed later in this chapter.

Attempts to directly implement such schemes with existing measurement sets were not as successful [7,8] as expected. There were numerical problems with the convergence of the estimator, and the issue of its bad data handling properties had not been addressed satisfactorily. However, the method was fast enough that it could be implemented on the computer systems then available, and be run every fifteen minutes or so, which was the typical cycle time used in control centers then. Operating experience revealed almost

immediately that without taking some precautions against the presence of bad data, the output of the estimator could not be relied upon. Until then, the optimization process was assumed to be a success if the final value of the objective function was sufficiently small, with little attention paid to the measurement residuals, $\mathbf{r} = \mathbf{z} - h(\hat{\mathbf{x}})$.

At first, the new procedures were based substantially on the weighted least squares method. The first step was the use of the chi-square test applied to the weighted residuals

$$\mathbf{r_w} = \frac{\mathbf{z} - h(\hat{\mathbf{x}})}{\sigma} \tag{7}$$

where σ is the vector of the diagonal elements of $\mathbf{R}$. It was shown that the weighted residuals followed a $\chi^2_{f,\alpha}$ distribution, where $\alpha = m - n$, the number of degrees of freedom. Therefore subsequent to the estimation, a detection test based on the weighted residuals would be performed, and if none of the $\mathbf{r_w}$ were greater than the threshold from the $\chi^2_{f,\alpha}$ distribution, then it was assumed that no bad data were present. The weighted residual test however does not identify the bad data in all cases ; it merely indicates the presence of bad data. A more rigorous test was devised, based on the normalized residuals $\mathbf{r_n}$

$$\mathbf{r_n} = \frac{\mathbf{r_w}}{\sqrt{\mathbf{w}}} \tag{8}$$

where $\mathbf{w}$ is the vector of the diagonal elements of the residual sensitivity matrix $\mathbf{W}$, where $\mathbf{W}$ is defined by

$$\mathbf{W} = \mathbf{I} - \mathbf{H} (\mathbf{H^t R^{-1} H})^{-1} \mathbf{H^t} \tag{9}$$

and is derived from the equation $\mathbf{r} = \mathbf{We}$. It was shown that if only one bad data point were present, it could be identified as the measurement with the largest normalized residual. In this case then, the bad data point would be deleted from the measurement set, and the estimation process repeated with the reduced set. This is computationally expensive, but was the only known way at the time to handle bad data using a least squares estimator. In principle, the normalized residuals could also handle the case of multiple non-interacting bad data ; that is if the bad data points were "far away" from each other. There was no clear definition of what constituted being "far away" ; in practice it was assumed that if the bad data occurred at nodes electrically distant from each other, they would not interact. On the other hand, if the bad data points were "close" to each other and interacted, this would result in the largest normalized residual not always corresponding to one of the bad data. In fact, a masking effect was observed, where the effect of adding the second bad data point was to make the normalized residuals for both the bad data very small, and cause false identification of a good data point as a bad input.

In summary, these methods were the fastest computationally, the simplest to implement, and if the measurement errors were normally distributed, they were optimal in the sense of having the minimum variance of the estimate. If a single bad data point was present, it could be identified by the weighted least squares test, and identified by a normalized residual test. If a bad data point were identified, the estimation would have to be redone with the bad point deleted.

2.5 : Non-Quadratic Estimators

Due to the limitations mentioned in the previous section, it was clear that though the Normalized Least Squares estimator was adequate, it would be necessary to investigate estimators that did not suffer from the same drawbacks. The basic problem with least squares based methods was in the assumption of the normality of the measurement errors ; hence it was logical to test estimators which were not optimal under the Gaussian distribution, but performed better under longer-tailed distributions, which were more common in practice. Though such estimators had already been introduced by statisticians some years previously under the general classification of M-estimators, Schweppe et al described them first in the power systems field as Non-Quadratic Estimators [9,10]. The reason for this nomenclature will be discussed in Chapter 3.

Basically these estimators minimize a function of the residuals; if the function happens to be the quadratic function, then the resulting estimator is the Least Squares estimator. For example, if the criterion is to minimize $\sum_{i=1}^{m} |r_i|$, then this becomes the Least Absolute Value (LAV) estimator. In general, these estimators put less weight on those measurements which have a large value of r_i. Therefore these measurements do not have as large an effect on the estimate, resulting in a more robust estimator. To some degree, these estimators combine the estimation and detection phases into one step, since the suspected bad data are automatically downweighted during the estimation process. The objective function used need not be smooth; in fact some of the best estimators of this type do not have derivatives at all points.

This causes some problems from a numerical standpoint, since the most common method for the solution of nonlinear equations in power systems is the Newton-Raphson technique. This cannot be used here, if the derivative of the objective function cannot always be evaluated. Also, depending on the exact nature of the function, the useful convexity property of the quadratic function may not hold true ; this implies that the objective function may have many local minima, and that it is therefore necessary to choose the initial starting point for any numerical search procedure carefully. The superior bad data handling properties of these estimators were not recognized sufficiently at that time in the power system area, due to a lack of familiarity with the robust estimation theory that had recently been developed by statisticians. These methods were not investigated as thoroughly, since their drawbacks were immediately apparent, while their advantages over the least squares were not as readily comprehended. Such estimators will be described in detail in Chapter 3.

Since the introduction of these estimators, research in the state estimation area has focused on two approaches. One approach is to use the LAV estimator, and to overcome the computational problems associated with this nonlinear minimization by reformulating it as a sequence of linear problems, each of which can be solved using Linear Programming (LP) methods [11,12]. This approach takes advantage of the progress made in the field of operations research towards solving LP problems rapidly, such as the Simplex method of Dantzig and more recently the geometric approach of Karmarkar. The sparse nature of power system models is also well suited to the application of the LP methods. However most of the research by power

system engineers has concentrated on making the estimator faster, rather than determining if this type of estimator is really suitable for application in state estimation. In Chapter 3, a theoretical analysis of this estimator will be presented, which shows that there are other drawbacks with the LAV estimator.

The other approaches that were investigated in the early 1980's concentrated on improving the Normalized Least Squares estimator. The hope was to find some test, which when applied to the residuals, could always detect and identify multiple interacting bad data. Various methods were proposed, such as the combinatorial optimization methods of Monticelli et al [13], the geometric optimization methods of Clements et al [14], and the Hypothesis Testing and Identification methods of Mili et al [15-17]. Again these method were mostly *ad hoc* procedures, developed without much attention to the intrinsic properties of the least squares estimator. These all relied on detection tests based on the residuals from the least squares, without considering the possibility that the very fact that the residuals were from a Least Squares estimate condemned their efforts to failure.

2.6 : Summary

Since the inception of static state estimation, there has been a constant introduction of new techniques, usually based on the work of statisticians. These include the bad data analysis tests for the weighted least squares estimator, the M-estimator and other non-quadratic estimators, and the Least Absolute Value Estimator. In this chapter, the fundamental concepts underlying these methods have been reviewed, though it is beyond the scope of this dissertation to describe each of these techniques exhaustively,

especially the details of their implementation. Further references, and a broad overview of power system state estimation can be found in Schweppe et al [18] and in Wu [19].

In Chapter 3, some of the tools of robust estimation theory will be introduced, and the estimators described in this chapter will be analyzed using these tools. This chapter also reflects the contributions of statisticians to the advancement of power system state estimation, albeit in many instances unknowingly.

Chapter 3 : Robust Estimation Theory

3.1 : Introduction

This chapter describes the general concepts that constitute the theory of robust estimation, which seeks to devise estimators that are relatively immune to outliers in their input. The techniques will first be illustrated on the simple case of a one-dimensional location problem ; they will then be extended to the multiple regression problems that are relevant to power system state estimation. In particular, these concepts will be introduced :

- Qualitative Robustness
- Breakdown Point
- The Influence Function
- Leverage Points and the Mahalanobis distance

The application of these ideas to power system state estimation will be discussed, along with any modifications necessitated by the nature of the power system. Various estimators will be described, starting with simple estimators of location, such as the arithmetic mean and median, and then focusing on M- and Generalized M-estimators.

3.2 : Historical Review

Robust estimation is not a recent development ; it is an old idea that has come into favor again as digital computers made practical the kind of computations it requires. A complete historical review is beyond the scope of this dissertation ; a useful reference on the subject is Beckman and Cook [20], in which it is stated that one of the first references to rejection of discordant data was made by Daniel Bernoulli in 1777. Other uses of the techniques that would today be called robust estimators, were in the scientific developments of the 18th and 19th centuries. Astronomers used a trimmed mean estimator, which discards both the highest and lowest observation, in order to get an estimate less influenced by outliers in either direction. Scientists used this form of averaging while estimating physical constants experimentally, such as the Michelson experiments to measure the speed of light, and in determining the atomic weights of the elements. When the modern Olympics were revived in 1896, events in which the scoring is subjective, such as gymnastics and diving, used the same principle, with the competitor's score being the average of the judge's scores, with the highest and lowest score disregarded.

These methods tended to be heuristic procedures, without a formal theoretical foundation. The first attempt to rectify this situation was the *parametric estimation theory* devised by Fisher in the early 1920's. This assumes an *a priori* knowledge of the probability distribution of the measurement errors, the so-called parametric model. In conventional state estimation, and in most other applications, the errors are assumed to be distributed normally, though there is generally no firm basis for this. Based

on this assumption, Fisher developed the maximum likelihood estimators which are optimal (unbiased with the minimum asymptotic variance), when the assumed parametric model is exact. He also proposed several concepts such as consistency, efficiency, and sufficiency, which measure the performance of an estimator under these assumptions.

This development was of great theoretical significance, but its implications became more relevant with the increased use of the computer to analyze data. One of the unfortunate side effects of the widespread use of the computer was that tools that were hitherto used only by statisticians who understood the underlying assumptions (such as the Gaussian distribution of the errors) were now being used widely by less knowledgeable practitioners. Another problem was that the size of the data sets being analyzed increased, as the computations were done by computer rather than manually, which made checking the data for obvious errors more difficult. The Least Squares (LS) method which was the most commonly used estimator, was also the most vulnerable in the presence of outliers. It has the advantages of ease of implementation, and having the maximum efficiency under the Gaussian assumption. However, as shown by Tukey [21], it loses its efficiency very rapidly as soon as the tails of the actual probability distribution function (p.d.f) lengthen. In particular, Tukey showed that the arithmetic mean (the one-dimensional LS estimator) becomes less efficient than the sample median (the LAV estimator) even in the case of the c.d.f being the Gaussian mixture model

$$G \;=\; (1 - \varepsilon).\, N(0,1) + \varepsilon.N(0,9) \tag{10}$$

when ε exceeds 0.08, where N(a,b) denotes the normal distribution with mean a and standard deviation b, the p.d.f of which is

$$g(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{\frac{-(x-a)^2}{2\sigma^2}} \tag{11}$$

This appears to contradict the Gauss-Markov theorem which states that the LS estimator is the Best Linear Unbiased Estimator (BLUE) for a general class of symmetric distributions with a finite variance. However all linear estimators exhibit poor performance under departures from the Gaussian model, and the LS estimator is biased for asymmetric distributions, which are much more common in practice. In fact, any finite sample drawn from a symmetric distribution appears to be asymmetric. Thus the apparent optimality of the LS estimator is non-existent under real-world conditions, and moreover, its bias can be made arbitrarily large by adding just one outlier, independent of the number of measurements. Questions about the suitability of the LS estimator grew [22,23,24,25] and this resulted in the development of the *non-parametric* estimation theory. As implied by its name, this requires no *a priori* knowledge of the probability distribution of the measurement errors. The assumptions made deal with the continuity of the distributions and the independence of the measurements.

Huber [26] formulated the *robust estimation theory*, in response to these developments. He postulated that a reasonable approximate model can usually be found, which fits only the majority of the data points. For example, by using the central limit theorem, one can say that the majority of the data follows the Gaussian distribution, with the other data points being

outliers, whose influence must be diluted. In this context, Huber [27] defined the concept of robustness as being "synonymous to insensitivity to small departure from the assumptions". Robust estimation theory therefore seeks to develop estimators which are not necessarily optimal under the assumptions made, but ones whose bias and variance remain bounded when the assumptions are not valid. Using this framework, Huber created the minimax approach to robustness, and developed the class of M-estimators (or generalized maximum likelihood estimators), which were introduced in power systems by Merrill and Schweppe [9] as non-quadratic estimators.

Continuing in this vein, Hampel [28] introduced several concepts which complement the Fisherian ones in that they assess the robustness of an estimator under departures from the assumptions. The estimator is treated as a system for which we wish to analyze the stability of the outputs, the bias and variance of the estimates, when noise affects the inputs, the assumptions. Hampel [29,30] proposed to analyze the robustness of an estimator by these measures :

(1) Qualitative robustness, which assesses the effect of small deviations from the assumptions.

(2) Global robustness, as quantified by the breakdown point, which determines the largest fraction of contamination that an estimator can handle.

(3) Local robustness, which is measured by the influence and change of variance functions, and analyzes the effects of infinitesimal deviations on the bias and variance, respectively.

3.3 : The Location problem

The simplest form of an estimation problem is the *location problem,* where the data consists of a set of numbers, and the objective is to determine the value most representative of the set. Let the measurement set be denoted by

$$\mathbf{z} = (z_1, z_2, \dots z_m) \tag{12}$$

where the measurements are assumed to independent and identically distributed according to the cumulative distribution function G. Since G is generally unknown, it is assumed to be approximated by the c.d.f F. Let the parameter being estimated be denoted by θ, and let $\hat{\theta}(\mathbf{z})$ be the estimate obtained from these m measurements. Let $\Lambda_F(\hat{\theta})$ and $\Lambda_G(\hat{\theta})$ be the cumulative distribution functions of the estimate obtained with the c.d.f 's F and G respectively.

3.3.1 : Functional Form of an Estimator

An empirical c.d.f $G_m(u)$ can be associated with the set of measurements

$$G_m(u) = \frac{1}{m} \sum_{i=1}^{m} \Delta(u - z_i) \tag{13}$$

where $\Delta(u)$ is the Heaviside step function. This empirical c.d.f tends to the true c.d.f G as m increases. Since

$$\int_{-\infty}^{\infty} u . d\Delta(u - z_i) = z_i \tag{14}$$

we can replace the sample by G_m, and the estimator by a function of G_m, which yields

$$\hat{\theta}_m(z) = \hat{\theta}_m(G_m) \tag{15}$$

If this function remains the same for all m and G_m, it is said to be a functional, and the estimator $\hat{\theta}_m(G_m)$ is also a functional. This equation will be used later in developing the influence function and other concepts. It can also be used to define several common one-dimensional estimators in their functional form.

$$\text{Arithmetic mean} = \int_{-\infty}^{\infty} u \, dG(u) \tag{16}$$

$$\text{Sample Median} = G^{-1}\left(\frac{1}{2}\right) \tag{17}$$

$$\alpha\text{-trimmed Mean} = \frac{1}{1-2\alpha} \int_{\alpha}^{1-\alpha} G^{-1}(u) \, du \tag{18}$$

The conventional definition of these estimators is

$$\text{Arithmetic Mean} = \frac{1}{m} \sum_{i=1}^{m} z_i \tag{19}$$

and the median z_{50} is defined as

$$z_{50} = \begin{cases} z_{(m+1)/2} & \text{if m is odd} \\ 0.5\,(z_{m/2}+z_{(m/2)+1}) & \text{if m is even} \end{cases} \qquad (20)$$

The α-trimmed mean of a sample with m measurements is the mean of the subsample formed by discarding the $[\alpha m]$ largest and $[\alpha m]$ smallest observations. The notation $[x]$ denotes the integer part of x, the floor function.

The following assumptions will be made regarding the estimators described in this chapter :

(1) The estimators can be replaced, at least asymptotically (as $m \rightarrow \infty$) by a functional.

(2) The estimators are Fisher consistent, i.e the estimate based on the assumed c.d.f F, $\hat{\theta}_m\,(F_m)$ converges to the true value θ as m increases.

3.3.2 : Qualitative Robustness

We are now in a position to define qualitative robustness :

A sequence of estimators $\{\,\hat{\theta}_m\,;\,m \geq 1\}$ is said to be qualitatively robust at the c.d.f F if a small deviation between F and G creates a small deviation between $\Lambda_F(\hat{\theta}_m)$ and $\Lambda_G(\hat{\theta}_m)$, for all sample size m.

This is a difficult concept to use in practice, since measuring the distance between the cumulative distribution functions F and G is not easy. Hampel [29] suggests the use of the Prohorov distance as a suitable metric.

A more practical definition of qualitative robustness given by Hampel is that a sequence of estimators $\{\,\hat{\theta}_m\,;\,m \geq 1\}$ which is a continuous sequence of continuous functions $\hat{\theta}_m(z)$ at F, is qualitatively robust at F. Thus by

verifying the continuity of the functional $\hat{\theta}_m(F_m)$, the qualitative robustness of the estimator can be assessed. Again considering the contamination model discussed before, $G = (1 - \varepsilon) F + \varepsilon H$, where H is any c.d.f and ε is the degree of contamination ($0 \leq \varepsilon \leq 1$), then the asymptotic bias is

$$b = |\hat{\theta} (G) - \hat{\theta} (F)| \tag{21}$$

Then for the estimator F to be qualitatively robust, the asymptotic bias must be continuous at $\varepsilon = 0$. The finite sample bias b_m is defined as $b_m = |\hat{\theta} (G_m) - \hat{\theta} (F_m)|$.

3.3.3 : Global Robustness and the Breakdown Point

If an estimator is qualitatively robust, its maximum possible bias must remain bounded, at least for the presence of one outlier among the m measurements. Another measure of the robustness of an estimator is the maximum fraction of measurements, ε^*, that can be made arbitrarily large with the bias remaining bounded. Formally if $z = (z_1, z_2, ... z_m)$ is a good sample, and $\hat{\theta}_m$ is the estimate calculated from z, then let z' be a sample derived from z by replacing any f out of the m measurements with arbitrary values (i.e $\varepsilon = f/m$), and $\hat{\theta}'_m$ is the estimate calculated from z'. Now consider all possible corrupted samples z' ; and for each one compute the bias, and find the maximum bias

$$b_{max} = |\hat{\theta}_m - \hat{\theta}'_m| \tag{22}$$

caused by the contamination. Then the breakdown point of the estimator, ε^* is defined as

$$\varepsilon^* = \max \{ \varepsilon = f/m \mid b_{max} \text{ is finite.}\} \tag{23}$$

The breakdown point is a worst case concept ; if there is just one sample with f_1 contaminated observations that has an infinite bias, the breakdown point is less than f_1/m, even though all other samples with f_1 contaminated observations may have finite bias. It is also independent of the choice of the f observations, the magnitude of the contamination, and the p.d.f of the good sample z. Since a robust estimator relies on the majority of the data, the maximum possible value of f is

$$f_{max} = [\frac{m-n}{2}] \tag{24}$$

and hence

$$\varepsilon^*_{max} = [\frac{m-n}{2}]/m \tag{25}$$

where n is the number of parameters being estimated ($n =1$, in the location problem). Asymptotically, as $m \to \infty$, ε^*_{max} reaches its maximum value of 50% ; estimators that attain this breakdown point are called high breakdown point estimators. For reference, the breakdown point for the estimators discussed in Section 3.3.1 is given here, while a more complete list is given in the Princeton robustness study [31], where more than 68 estimators have been analyzed.

The arithmetic mean has a breakdown point of 0 ; it is not qualitatively robust. This can be shown by taking one observation, and setting it to the value ∞ ; the mean of any set containing this sample is also ∞, independent of the size of the sample. The effect of one outlier alone is enough to make the

bias infinite, and hence the mean is not qualitatively robust.

The sample median has the maximum breakdown point in the one-dimensional case. In the worst case, where all the corrupted samples are on the same side of the original median, at least 50 % of the observations would have to be corrupted before the median is affected.

The α-trimmed median has a breakdown point that depends on the fraction of trimming, α. When α=0, the α-trimmed mean reduces to the arithmetic mean, and with α=0.5, it becomes the sample median. In fact, its breakdown point is α.

In general, the breakdown point of an estimator decreases as the number of parameters being estimated, n, increases. For example, the LAV which is the multi-dimensional equivalent of the sample median has a breakdown point of 25 % in the simple regression (n=2) case, as opposed to 50% for n=1. In Chapter 6, it will be verified through Monte Carlo simulations that the breakdown point decreases rapidly as n increases.

3.3.4 : The Influence Function

The local robustness of an estimator at a given probability distribution of the measurement errors is measured by two functions: the Influence Function (IF) and the Change-of-Variance Function (CVF), which assess the effect of a single outlier on the bias and variance, respectively.

a. Finite sample IF in the unidimensional case

The finite sample influence function IF(z ; $\hat{\theta}_m$, F) of an estimator of location $\hat{\theta}_m$ at the c.d.f F, associated with the sample $\mathbf{z} = \{z_1, z_2 \dots , z_{m-1}, z\}$ is the difference between the estimates $\hat{\theta}_m$ and $\hat{\theta}_{m-1}$, computed with and without

the observation z, and divided by the fraction of contamination $\varepsilon = 1/m$. The observations $\{z_1, \dots ,z_{m-1}\}$ follow the c.d.f F and z takes on all real values. Formally

$$IF(z ; \hat{\theta}_m, F) = m \{ \hat{\theta}_m(z_1, \dots ,z_{m-1}, z) - \hat{\theta}_{m-1}(z_1, \dots ,z_{m-1})\} \qquad (26)$$

Several local robustness measures can be derived from the influence function. The most important of them are the gross error sensitivity and the local shift sensitivity, which are discussed below.

b. *The gross error sensitivity*

The gross error sensitivity γ^* is the supremum of the absolute value of IF computed over all values of z, i.e

$$\gamma^* = \sup_z | IF(z ; \hat{\theta}_m , F)| \qquad (27)$$

Its importance stems from its connection to the maximum bias of the estimator. If the bias b is defined as $|\hat{\theta}_m - \hat{\theta}_{m-1}| \approx \varepsilon |IF|$ (the approximation is the difference between the finite sample IF and the asymptotic IF) , then the maximum value of the bias is the maximum value of $\varepsilon |IF|$. Hence

$$b_{max} = \varepsilon\gamma^* \qquad (28)$$

Hence a well-behaved robust estimator should have a bounded gross error sensitivity, and consequently a bounded bias. Such an estimator is said to be B-robust. Bounding γ^* is an important criterion in the design of new estimators.

c. *The local shift sensitivity*

The local shift sensitivity measures the effect of small variations in the

observations, which can be caused by rounding errors, for example. As noted by Hampel [30], this is equivalent to shifting an observation from a point z to a nearby point y. The effect of this shift on the estimate is equal to the difference between the values of the IF at these two points. The local shift sensitivity λ^* is a standardized version of the supremum of this difference, taken over all $y \neq z$.

$$\lambda^* = \sup_{y \neq z} \frac{|\text{IF}(y; \hat{\theta}_m, F) - \text{IF}(z; \hat{\theta}_m, F)|}{|y-z|} \qquad (29)$$

λ^* is the greatest slope of all straight lines intersecting the IF at two distinct points [32]. When the IF is differentiable for all z, the straight lines will be the tangents to IF, and we have

$$\lambda^* = \sup_z \left| \frac{\partial \text{IF}(z; \hat{\theta}_m, F)}{\partial z} \right| \qquad (30)$$

d. Asymptotic influence function

If we let m, the number of observations increase to infinity, then the fraction of contamination, $\varepsilon = 1/m$, will decrease to zero. Then it can be shown that under certain regularity conditions [33], the IF becomes the directional derivative of the functional form of the estimator, $\hat{\theta}(G)$, in the direction of G, with G being the contaminated model, $G = (1 - \varepsilon) . F + \varepsilon . \Delta(u-z)$. Formally

$$\text{IF}(z; \hat{\theta}, F) = \frac{\partial \hat{\theta}(G)}{\partial \varepsilon} \Bigg|_{\varepsilon=0} = \lim_{\varepsilon \to 0} \frac{\hat{\theta}(G) - \hat{\theta}(F)}{\varepsilon} \qquad (31)$$

The IF is a linearization of the estimator $\hat{\theta}(G)$ at $\varepsilon = 0$. It describes the local behavior of $\hat{\theta}$ in a small neighborhood of the assumed model F. It is related

to the asymptotic bias $b(\varepsilon)$ (as shown in Figure 1) by

$$b(\varepsilon) = |\hat{\theta} (G) - \hat{\theta} (F)| \approx \varepsilon |IF(z ; \hat{\theta} , F)| \tag{32}$$

and to the asymptotic variance $V(\hat{\theta} , F)$ through

$$V(\hat{\theta} , F) \approx \int IF^2(z ; \hat{\theta} , F) \, dF(z) \tag{33}$$

The asymptotic variance determines the statistical efficiency of an estimator ; the lower the variance, the more efficient the estimator. Figure 1 illustrates the link between the various robustness concepts defined previously. As suggested by Huber and Hampel, qualitative robustness implies the continuity of the maximum possible asymptotic bias $b(\varepsilon)|_{\varepsilon=0}$; the gross error sensitivity is the slope of its tangent at $\varepsilon = 0$ for well-behaved estimators, and the breakdown point is the abscissa of its first asymptote.

3.3.5 : The IF of some location estimators

Figure 2 shows the asymptotic IF at the Gaussian distribution of the three estimators of location defined in Section 3.3.1. The IF of the arithmetic mean is a straight line with slope 1, which indicates that the greater the value of a measurement, the larger its influence. Since the maximum value of the IF is unbounded, the mean has an infinite gross error sensitivity ; and hence this estimator is not robust. On the other hand, γ^* is bounded for the α-trimmed mean and for the sample median. In fact, the median has the lowest possible value of γ^* that an estimator may have ($\gamma^* = \sqrt{\pi/2}$) at the standard Gaussian distribution).

On the other hand, the sample median has an infinite value for the local shift

sensitivity, and λ^* is equal to one for the arithmetic mean. This is the smallest possible value of λ^* that an estimator may have (at the Gaussian distribution). In this regard, it is the estimator most resistant to rounding errors. On the other hand, since the sample median is computed from the central observations, it is very sensitive to such errors. Note that from a practical viewpoint, robustness against gross errors is much more critical than robustness against rounding errors.

3.3.6 : Hampel's Optimality Criterion

Clearly, it would be ideal to find an estimator that has low values for both the asymptotic bias and the asymptotic variance, at a given c.d.f F. This however poses a dilemma, since these requirements conflict ; the lower the bias, the larger the variance, and vice versa. Therefore in practice, we have to find a trade-off between robustness and efficiency. For instance, the arithmetic mean is the most efficient estimator at the Gaussian distribution but it is not robust ; whereas the sample median has a poor efficiency at this distribution, but it is the most robust estimator. This requirement is expressed by Hampel's optimality criterion, which consists of minimizing the asymptotic variance under the constraint that the gross error sensitivity is smaller than a given threshold. Estimators which satisfy such a criterion are called optimally B-robust.

3.3.7 : The Change of Variance Function

Till now, we have assessed the local robustness of the asymptotic bias of an estimator through the influence function. To analyze the local robustness of

the asymptotic variance, the corresponding function is the Change of Variance Function (CVF). It is defined as the first derivative of the asymptotic variance with respect to ε at $\varepsilon = 0$

$$\text{CVF} (z ; \hat{\theta} , F) = \frac{\partial V(\hat{\theta}, G)}{\partial \varepsilon} \Big|_{\varepsilon=0} \tag{34}$$

with

$$G = (1 - \varepsilon) F + \frac{\varepsilon}{2} [\Delta(u\text{-}z) + \Delta(u\text{+}z)] \tag{35}$$

Similar to the analysis for the bias, the analysis for the variance uses the quantity κ^* (the equivalent of γ^*) which is called the change-of-variance sensitivity. It is related to the maximum asymptotic variance by the equation

$$\underset{G \in N(F)}{\text{Sup}} \ V(\hat{\theta} , G) = V(\hat{\theta} , F) \exp\{\varepsilon\kappa^*(\hat{\theta} , F) \} \tag{36}$$

It is apparent that if κ^* is bounded, then the asymptotic variance will be bounded ; such an estimator is said to be V-robust. On the other hand, when the variance $V(\hat{\theta}, F)$ associated with the assumed distribution F is minimized with the constraint that κ^* is less than a given threshold, then the estimator is said to be optimally V-robust (the optimality criterion of Hampel).

3.4 : Leverage Points and M-estimators

As mentioned previously, some of the concepts presented in the previous sections can be used as tools to design estimators with more desirable

properties. For example, it has been shown that a bounded influence function is necessary. The first class of estimators which embodies these principles is the M-estimator. Using the notation of Chapter 2, where h(x) is the assumed model, and σ_i is the variance associated with the measurement z_i, M-estimators have been defined [33] as those which minimize the criterion

$$J(\mathbf{x}) = \sum_{i=1}^{m} \rho(\frac{z_i - h_i(\mathbf{x})}{\sigma_i}) = \sum_{i=1}^{m} \rho(r_{Wi}) \tag{37}$$

where r_{Wi} are the weighted residuals. As mentioned before, if $\rho(x) = x^2$, then this describes a Weighted Least Squares estimator. To find the solution to the above equation, we take the derivative of $J(\mathbf{x})$ with respect to x and equate it to 0. In Chapter 2, we have defined the Jacobian matrix H as $\partial h(\mathbf{x})/\partial \mathbf{x}$, and **R** as the covariance matrix associated with the measurements. Let l_i be the ith row of $\mathbf{R}^{-1/2}\mathbf{H}$; then each l_i is associated with a particular measurement, and defines a point in the factor space. We then have to solve

$$\sum_{i=1}^{m} \psi(r_{Wi})\, l_i = 0 \tag{38}$$

where $\psi(r_{Wi})$ is the derivative of $\rho(r_{Wi})$ with respect to r_{Wi}. This equation can be solved using standard numerical techniques, such as the Newton-Raphson method, or by the iterative WLS method. The ψ function is proportional to the influence function of an estimator.

Based on the properties of the ρ and ψ functions, M-estimators can be classified in three ways :

(1) Those with a convex ρ function. Examples of this class are the LS estimator, the LAV with $\rho(r) = |r|$, and the Huber estimator.

(2) Those with a non-convex $\rho(r)$, but having non-vanishing $\psi(r)$ functions. Examples of this class are the Schweppe estimator and the Muller estimator.

(3) Those with a non-convex $\rho(r)$, but having $\psi(r)$ functions that become 0 for some values of r. Examples of this class are the Hampel estimator and the Huber type skipped-mean estimator. These are also called re-descending M-estimators, since their $\psi(r)$ function increases from $r = 0$ and then decreases. Weighted LS estimators with rejection rules such as the normalized residual test also belong to this class.

The convexity of the ρ function implies that there is a unique solution to equation (38), so that the choice of starting point for an iterative solution technique is not crucial. This is not true for the redescending estimators, which have an infinite number of solutions. Table 1 gives analytical expressions for the $\rho(r)$ and $\psi(r)$ of these estimators, while Figure 3 shows the ψ function graphically. In Table 1 these are given as functions of r, the residuals ; if weighted residuals are to be used, each r should be replaced by $r_W = r/\sigma$. It can be seen that in the center of each ψ function, around r=0, all the estimators are very similar to the LS. This is done intentionally, since the LS is the most efficient estimator at the Gaussian distribution, and most distributions resemble the Gaussian in the middle. Thus, having $\psi(r) = r$ for small r gives an M-estimator increased efficiency for centrally located measurements.

Table 1 : ρ and ψ functions for some M-estimators

ESTIMATOR	DOMAIN	$\rho(\mathbf{r})$	$\psi(\mathbf{r})$
Least Squares	$r \in \Re$	r^2	r
LAV	$r \in \Re$	$|r|$	$\text{sign}(r)$
Huber	$|r| \leq b$	$0.5\, r^2$	r
	$|r| > b$	$b\,|r| - 0.5\, b^2$	$b\,\text{sign}(r)$
Muller	$|r| \leq b$	$0.5\, r^2$	r
	$|r| > b$	$0.5\, b\,|r|$	$0.5\, b\,\text{sign}(r)$
Schweppe	$|r| \leq b$	$0.5\, r^2$	r
	$|r| > b$	$2b^{3/2}\sqrt{|r|} - 1.5\, b^2$	$b^{3/2}\dfrac{\text{sign}(r)}{\sqrt{|r|}}$
Huber Type	$|r| \leq b$	$0.5\, r^2$	r
Skipped Mean	$|r| > b$	$0.5\, b^2$	0
Hampel	$|r| \leq a$	$0.5\, r^2$	r
	$a < |r| \leq b$	$a\,|r| - 0.5\, a^2$	$a\,\text{sign}(r)$
	$b < |r| \leq c$	$a\,\dfrac{c\,|r| - 0.5r^2}{c\text{-}b} - 0.5(a^2 - \dfrac{ab^2}{c\text{-}b})$	$a\,\dfrac{c\text{-}|r|}{c\text{-}b}\,\text{sign}(r)$
	$|r| > c$	$0.5\, a\,(b + c - a)$	0

3.4.1 : Robustness of M-estimators

Hampel [33] showed that the total influence function for a regression M-estimator at the Gaussian c.d.f Φ is made up of two components ; it is the product of the scalar influence function of the residuals IR, and the vector valued influence of position in the factor space IP. He showed that

$$IR(\ r_W\ ;\ \hat{x},\Phi) = \frac{\psi(r_w)}{E[\psi'(r_w)]} \tag{39}$$

and

$$IR(\ l\ ;\ \hat{x},\Lambda) = (E[ll^T])^{-1}l \tag{40}$$

Here Λ is the c.d.f of the points in the factor space, which are defined by the vector l, which are assumed to have zero mean and a covariance matrix $E[ll^T]$. For the total influence function to be bounded, both ψ and l must be bounded ; which is clearly not true for l. As a point moves further away from the bulk of the data in the factor space, the value of l increases, and hence all regression M-estimators (including the LAV) are not robust in the presence of leverage points. A single outlier can cause breakdown, if it is associated with an outlier in the factor space. Potentially influential measurements are not only those far from the others in the z direction but also those distant in the factor space, the so-called leverage points. The associated measurements have a great influence on the M-estimators, including the WLS and the LAV estimator, as we will see later on. The residuals of such measurements remain small (zero values for the LAV) even if they are corrupted by gross errors, and therefore they are not detected by hypothesis tests based on the weighted residuals.

3.4.2 : Leverage Points in Power Systems

Clearly it is important to determine if leverage points are present in the factor space of typical power systems. The LAV estimator, which is otherwise

an extremely viable option, is not suitable if leverage points are present. The factor space is the measurement Jacobian matrix of the system. A normal measurement set consists of line flows, node injections and voltage magnitudes. For a line flow measurement, there are two entries per row of the Jacobian matrix (with the decoupled assumption) ; these are proportional to the admittance of the corresponding line. An injection measurement at node i has (n_i+1) entries in the Jacobian, where n_i is the number of lines connected to node i. One of these entries is the sum of the admittances of the n_i lines ; each of the other entries has the value of the corresponding line admittance. The voltage measurements have one entry, a value of 1 in each row. Leverage points are likely to occur when the admittance of a particular line becomes relatively large (short transmission lines), or when a large number of lines are connected to a particular bus. Due to the sparsity of the system, especially at the high voltage level, most rows in the Jacobian matrix have three or four entries ; one large admittance value can therefore make that measurement become a leverage point. In Chapter 6, the issue of how to determine which measurements are leverage points is addressed, especially in the context of large sparse systems. The Mahalanobis distance is an estimate based on the LS method, used for identifying outliers in the factor space, and is subject to the masking effect similar to that seen in the residuals from the LS estimate. That is, it can detect an isolated leverage point, but if several leverage points are present, they may serve to camouflage each others presence.

It should be noted that even if all the leverage points can be identified, removing them from the measurement set is not advisable. By definition,

leverage points have an undue influence on the estimate ; if the associated measurement is bad, this is deleterious. On the other hand, a good leverage point is extremely beneficial in reducing the variance of the estimate.

3.4.3 : Generalized M-estimators

Several proposals have been made with the objective of making the M-estimators more robust against bad leverage points by bounding both the influence of residual and of position. Such estimators are referred to as the generalized M-estimators (also called the bounded-influence regression estimators). According to Hill [34], all the proposals may be written in the following form:

$$\sum_{i=1}^{m} u(l_i) \, \psi(r_{Wi} \, v(l_i)) \, l_i = 0 \tag{41}$$

where $u(l_i)$ and $v(l_i)$ are weighting functions whose role is to downweight the influence of the leverage points in the regression. The proposals are generally of two forms :

(1) The Mallows type [35], with $v(l_i)=1$; in this case, the leverage points are downweighted regardless of their residual value. This implies that the influence of position is bounded separately from the influence of the residuals, and that good leverage points will be treated just like bad ones.

(2) The Schweppe type [36], with $v(l_i) = 1 \, / \, u(l_i)$; in this case, the leverage points are downweighted only if the residuals are large. Unlike the Mallows type, the influence of position and the influence of

residuals depend on each other so that the total influence is bounded.

Since leverage points are relatively common in state estimation, the Schweppe type estimator will be analyzed in more detail. The principle is to not use a fixed tuning constant b for all the weighted residuals, but to adjust b based on the diagonal element of the W matrix. This is equivalent to comparing the normalized residuals r_{Ni} to b, since the normalized residual r_{Ni} is equal to $r_i / \sigma_i\sqrt{W_{ii}}$. Since the residual sensitivity matrix W is not robust if multiple bad leverage points are present, a more suitable replacement for the use of W_{ii} has been the subject of some research [37].

The generalized M-estimators are an improvement over the M-estimators in terms of their ability to limit the influence of leverage points. However Maronna et al [38,39] have shown that the asymptotic breakdown point ε^* of such estimators cannot exceed the reciprocal of the number of variables being estimated, i.e $\varepsilon \leq 1/n$. As a result, it tends to vanish for large n, which are common even in medium size power systems. With 100 buses, n = 199, and the breakdown point is less than 1/200 = 0.005, which is of little practical use. In larger systems, where n often exceeds 1000, it is zero for all practical purposes.

3.5 : Summary

The numerical values for some of the robustness parameters discussed in this chapter are summarized in Tables 2 and 3 [32]. Table 2 is for location estimates at the Gaussian distribution, and shows that the mean has the minimum variance, but has no robustness whatsoever, while the median is

robust, but has a significant loss of efficiency. The Huber estimator combines the robustness of the median, with an efficiency comparable to that of the mean, even when the distribution of the measurements is such that the mean is the optimal estimator.

Table 2 : Robustness Parameter Values for some Estimators

Estimator	Asymptotic Variance	Gross Error Sensitivity	Local Shift Sensitivity	Breakdown Point
Mean	1.00	∞	1.00	0.00
Median	1.571	1.25	∞	0.50
Huber (b=1)	1.107	1.46	1.46	0.50
Huber (b=1.5)	1.037	1.73	1.15	0.50

Table 3 : Location estimate variance for several estimators

Estimator	n Var($\hat{\theta}$) Gaussian	n Var($\hat{\theta}$) One-wild
Mean	1.000	6.485
Median	1.498	1.555
Huber (b = 1.5)	1.051	1.222
Hampel (a =1.7, b=3.4, c=8.5)	1.130	1.166

In Table 3 the effect of one outlier is shown ; that is, out of a total sample size of 20, 19 measurements are taken from the distribution $N(0,1)$ and one

measurement is drawn from the N(0,10) distribution. This is known as the one-wild distribution. The table gives the value for n times the variance obtained from Monte Carlo simulations [32]. It illustrates the rapid loss of efficiency suffered by the mean as soon as the measurement distribution deviates from the Gaussian; in contrast, the robust estimators are hardly affected by the addition of a single outlier.

Chapter 2 served as a review of the standard algorithms and techniques for power system state estimation. In Chapter 3, some tools have been introduced, with which the suitability of these methods for the task can be assessed. It has been shown that the Weighted Least Squares Estimator has the advantages of speed and efficiency, but is deficient in regard to its handling of multiple interacting bad data and leverage points. The class of M- and Generalized M-estimators, which includes the Least Absolute Value Estimator, is competitive with (but not as good as) the LS in speed and efficiency, and has better handling of multiple bad data. They are not, however, capable of dealing with bad leverage points. Though they are theoretically able to handle multiple interacting bad data, their low breakdown point in high-dimensional systems, such as those used in state estimation, makes them less viable. Therefore, the estimator that would be ideal in this application should have the following properties :

- A high breakdown point, even in large systems.
- Be immune to multiple bad leverage points.
- Reasonably efficient statistically.
- Competitive in computation time with the LS estimator.

In the next two chapters, we will discuss a class of estimators that fulfills

most of these requirements, namely the High Breakdown Point Estimators, and in particular, we will consider the implementation of the Least Median of Squares Estimator in the context of real-time power system state estimation. Some other aspects of the material presented in this chapter can be found in Mili, Phaniraj and Rousseeuw [40].
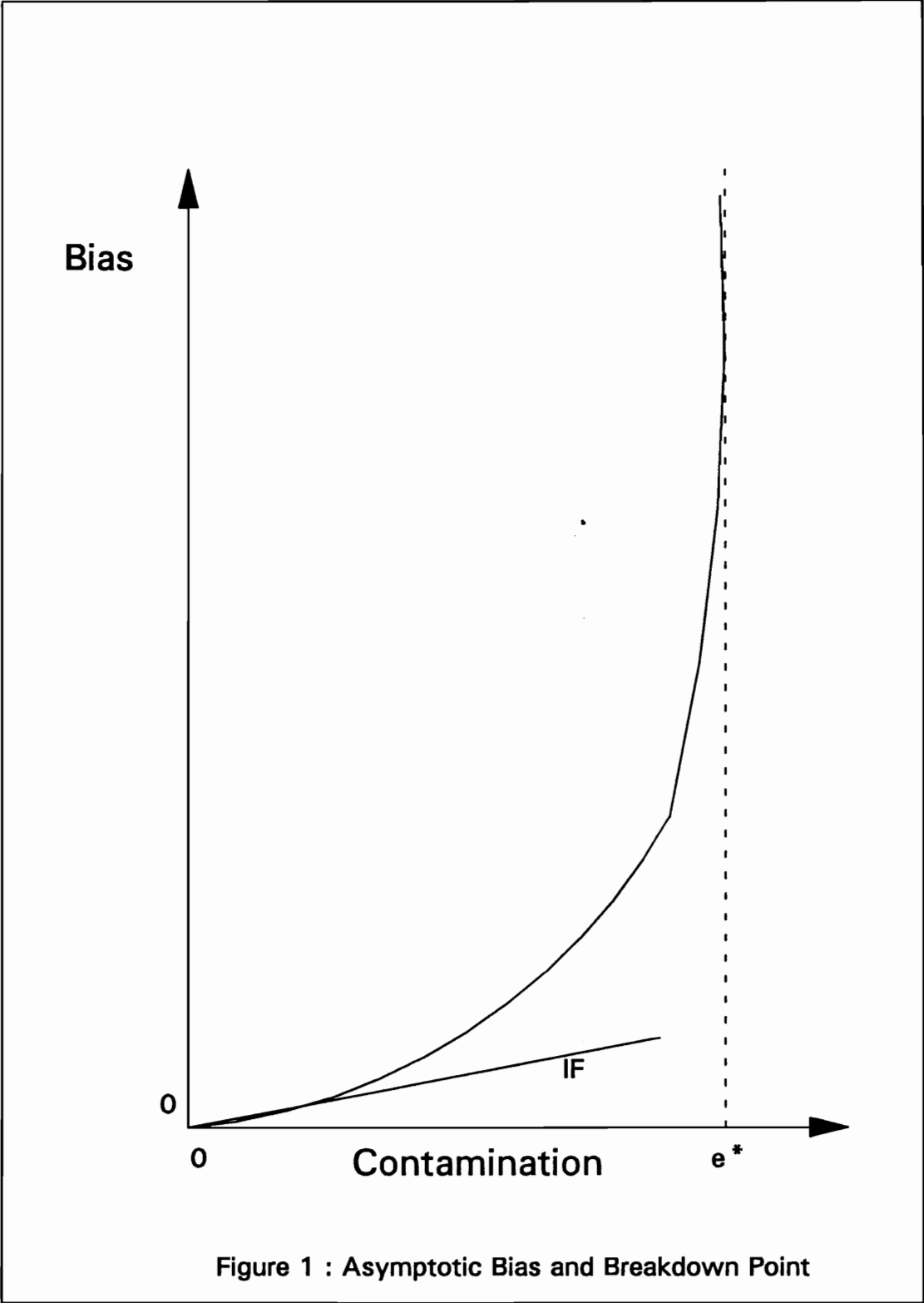
Figure 1 : Asymptotic Bias and Breakdown Point

Figure 2 : Asymptotic Influence Function For Several Estimators
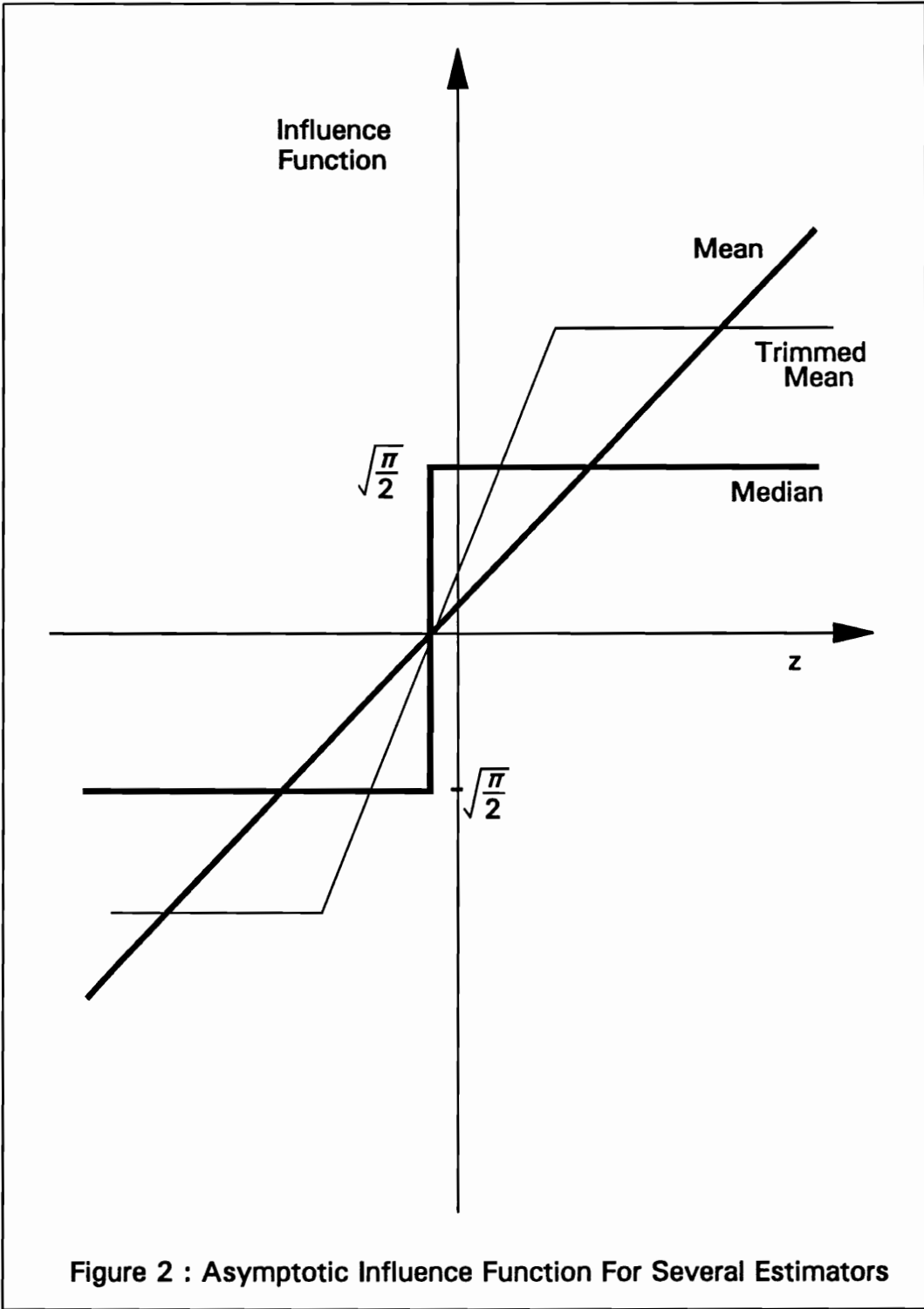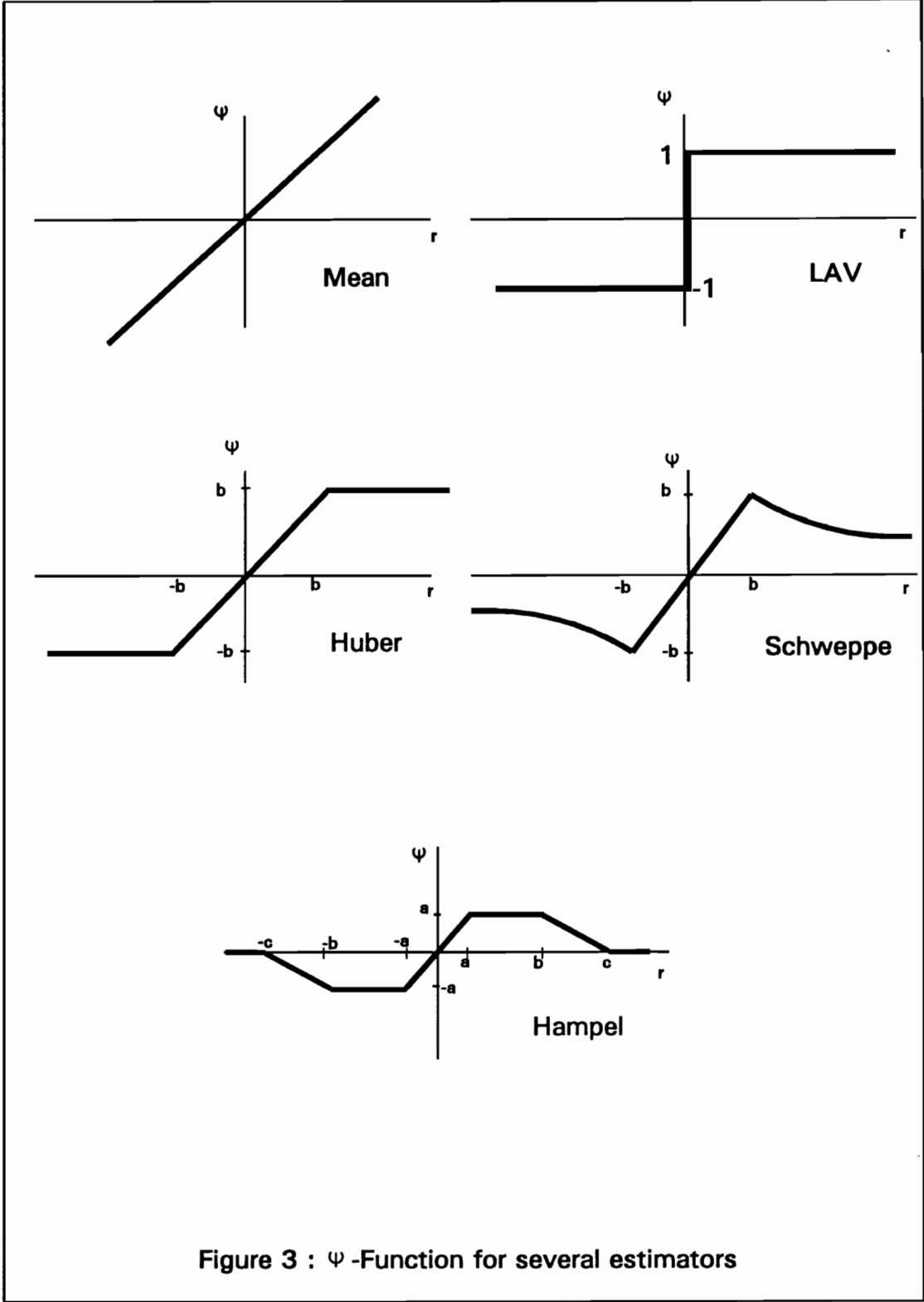
Figure 3 : Ψ -Function for several estimators

Chapter 4 : High Breakdown Point Estimators

4.1 : Introduction

This class of estimators was developed as an outcome of the work of Huber, Hampel, and others in defining such concepts as the breakdown point. The robust estimators developed as a direct result of this work, such as the M- and Generalized M-estimators, do not achieve the maximum breakdown point of 50% when used in higher dimension multiple regression problems. They were, of course, vastly superior to then existing methods. The objective in devising the estimators described in this chapter was to retain the robustness properties of previous estimators in the location and simple regression problems, while improving on them in the higher dimensional problems.

The first estimator belonging to this class is the repeated median estimator, which was developed by Siegel [41]. This was primarily of theoretical interest, in that it established that estimators with a high breakdown point in multiple regression do exist. Its main drawback was that it was not equivariant for a linear transformation of the problem. In the location case, if a constant α is added to each measurement, then the new estimate should be equal to the original estimate plus α. A further extension to this is the concept of scale equivariance, where the sample is multiplied by an arbitrary

constant as well as being shifted, $\mathbf{z'} = \alpha\mathbf{z} + \beta$. If $\theta(\mathbf{z'}) = \alpha.\theta(\mathbf{z}) + \beta$, then the estimator θ is scale-equivariant. M-estimators are location invariant, and depending on the choice of the scale parameter in the estimation, are also scale equivariant.

The repeated median estimator was followed by the least median of squares (LMS) and the least trimmed sum of squares (LTS) estimators developed by Rousseeuw [42]. These are location equivariant estimators, and will be the main focus of this chapter. Other estimators belonging to the class of High Breakdown Point Estimators (HBPE) include the s-estimators also developed by Rousseeuw [43], the MM-estimates and the τ-estimators developed by Yohai et al [44,45].

In general, these estimators can be characterized as having the maximum breakdown point, resistance to leverage points and multiple interacting bad data, a significantly lower efficiency at the Gaussian distribution, and greatly increased computational complexity. It would be likely therefore that these estimators would be used, for example, to compute a robust starting point for the subsequent application of a generalized M-estimator, or as diagnostic tools for the identification and rejection of the outliers. The efficiency of these estimators is such that a post-estimation procedure such as a least squares based estimate may be needed.

In this chapter, the properties of these estimators will be discussed, along with some aspects of the implementation of HBPE methods in large systems. Finally, a small example will be developed, to which various estimators will be applied. This will serve to demonstrate the superior performance of the HBPE estimators over the LS and the LAV, as well as other methods

proposed which purport to be able to handle multiple interacting bad data.

4.2 : High Breakdown Point Estimators

In this section, the five estimators mentioned above will be discussed in more detail.

4.2.1 : Least Median of Squares Estimator

In the location and simple regression cases (n=1 or n=2), the LMS estimator minimizes the median, rather than the sum, of the squared residuals. In multiple regression problems, it minimizes the v-th ordered squared residual, yielding the criterion

$$\min \ J(\mathbf{x}) = (\ r_W^2)_{\ v \ : \ m} \tag{42}$$

where

$$v = [\frac{m}{2}] + [\frac{n+1}{2}] \tag{43}$$

The weighted residuals are first squared and then ordered by increasing value:

$$(\ r_W^2)_{\ 1 \ : \ m} \ \leq (\ r_W^2)_{\ 2 \ : \ m} \leq \ ... \ \leq (\ r_W^2)_{\ m \ : \ m} \tag{44}$$

(In a private communication, Rousseeuw states that an alternate robust estimator can be obtained by not ordering the residuals prior to selecting the vth one. This is not of interest in the context of power systems.)

The LMS estimator has these geometrical properties [46]:

(1) In estimation of location, the LMS is the midpoint of the shortest

half of the real-valued observations. A "half" is a subset of the data that contains v successive ordered observations, where v is as defined above with $n = 1$. The shortest half is that half which has the smallest distance between its extreme observations, i.e that with the smallest difference in the set

$$(z_{v \, : \, m} - z_{1 \, : \, m}) \, , \, (z_{v+1 \, : \, m} - z_{2 \, : \, m}) \, , \, \dots \, , \, (z_{m \, : \, m} - z_{m-v+1 \, : \, m}) \, , \quad (45)$$

(2) In simple regression, the line that is the output of the LMS bisects the narrowest strip that encloses half the data points (see Figure 4).

(3) In multiple regression, the hyperplane that is the output of the LMS bisects the narrowest hyperstrip that encloses half the data points.

It is clear from properties 2 and 3 that the LMS estimate is then defined by a simple majority of the data, that which defines the narrowest (hyper) strip. This implies that the LMS can withstand a case with half the redundant measurements being outliers, which results in the maximum breakdown point ε^*_{max}.

4.2.2 : Least Trimmed Sum of Squares Estimator

The LTS estimator resembles the LMS in that it is based on the ordered squared residuals. It minimizes the sum of the h smallest squared residuals, where h is an integer that has a value between v and n. If h=n, the LTS is identical to the LS estimator ; if h=v, it achieves the maximum breakdown point. For intermediate values of h, there is a tradeoff between robustness and efficiency at the Gaussian distribution. The larger the value of h, the lower the breakdown point, but the greater the efficiency. In effect, this estimator can be tuned to some extent, so that by a judicious choice of h, a

combination of both robustness and efficiency can be achieved. The LTS has geometric properties similar to the LMS ; in the location case, where the LMS is the midpoint of the shortest half, the LTS is the mean of the observations that constitute the shortest half (if h=v).

Both the LTS and the LMS are location equivariant ; the LTS is slightly more complex computationally, since each of the h smallest squared residuals must be determined ; to evaluate the LMS, algorithms exist that can find the median without explicitly computing all the intermediate terms [43]. This implies that the computation of the objective function for the LMS has complexity $O(m)$, while for the LTS it is of complexity $O(m \log m)$. On the other hand, the LMS converges like $m^{-1/3}$, while the LTS converges like $m^{-1/2}$. The computational issues in the implementation of these HBPE will be discussed in more detail in subsequent sections.

4.2.3 : S-Estimators

S-estimators are based on estimating the scale, rather than the location of a sample. They were devised by Rousseeuw [43] in an attempt to find an estimator that combined the best properties of the LMS and LTS, namely :

(1) It should have a 50 % breakdown point, and be affine equivariant.

(2) It should converge as $m^{-1/2}$ or better.

(3) The objective function should have complexity $O(m)$ or lower.

The objective function that they minimize is

$$\min \ J(\mathbf{x}) = s\{r_1(\theta), r_2(\theta) \ ... \ , r_m(\theta)\} \tag{46}$$

where s is an estimate of the scale of the residuals. Rousseeuw defines s as

the solution to

$$\frac{1}{m} \sum_{i=1}^{m} \rho(r_i/s) = k \tag{47}$$

where $\rho(r_i/s)$ is similar to the ρ function used in M-estimators, and k is the expected value of ρ at the normal distribution, $E_\phi[\rho]$. He suggests the use of either the Tukey biweight or the Hampel as possible M-estimators to be used. S-estimators possess the exact-fit property, that is, if more than half the data fits the model exactly, the estimate will pass through those measurements exactly.

S-estimators have most of the desired statistical properties ; they present however, a much more complex computational problem than the LMS or LTS, which in turn are more difficult to implement than the LS or LAV. Methods suggested for its implementation are similar to the projection pursuit techniques of Friedman and Tukey [47].

4.2.4 : MM-Estimators

The MM estimators of Yohai [44] are also based on the estimation of a scale parameter. They resemble the S-estimator, and are in some ways an intermediate step between S-estimators and τ-estimators. The computation of an MM-estimate is done in three stages, where the first stage serves to provide a robust initial estimate for the second step, where an M-estimate of the scale of the errors is calculated. The final stage is to compute the regression coefficients, using a redescending M-estimator. This is an involved procedure, making it computationally expensive, which is its biggest

drawback. The final estimate however is affine-equivariant, statistically efficient, and has a high breakdown point. The τ-estimate created by the same researchers possesses the same properties, is easier to compute, and hence is treated in more detail next.

4.2.5 : τ-Estimators

τ-estimators were devised by Yohai and Zamar [45], and are also based on estimation of the scale. They have the robustness properties of the S-estimator, and are designed to have a higher efficiency at the Gaussian distribution. The objective function minimized is $\tau(\theta)$, where

$$\tau^2 = s^2 \frac{1}{m} \sum_{i=1}^{m} \rho_2(r_i/s) = k \tag{48}$$

where s is an estimate of the scale of the residuals, based on the function $\rho_1(r_i/s)$. If the functions ρ_1 and ρ_2 are the same, then $\tau = \sqrt{b}\, s$; if $\rho_2(u) = u^2$, then τ is the sample standard deviation. It is shown [45] that the high breakdown point properties of the τ-estimate are determined by the function ρ_1, and the efficiency at the Gaussian distribution is controlled by ρ_2. Normally, in M-estimators, the choice of the same parameters controls both the efficiency and breakdown point, so that one property cannot be maximized without adversely affecting the other. The ρ function and parameters suggested, such that the τ-estimate has both a breakdown point of 50 % and an efficiency of 95 % at the normal are given by

$$\rho(u) \;=\; \left\{ \begin{array}{ll} \dfrac{u^2}{2}\left(1 - \dfrac{u^2}{c^2} + \dfrac{u^4}{3c^4}\right) & \text{if } |u| < c \\[1.5em] \dfrac{c^2}{6} & \text{if } |u| \geq c \end{array} \right. \tag{49}$$

where c is 1.56 for ρ_1 and is 6.08 for ρ_2. The two advantages of the τ-estimators over s-estimators are :

(1) The combination of high breakdown and efficiency made possible by the independent control over the design of ρ_1 and ρ_2.

(2) The τ–estimate is easier to compute than the s-estimate, in principle. An algorithm for the τ-estimation based on an iterative WLS algorithm was proposed in [45].

4.3 : Bad Data Detection and Identification

Even though the estimators described in the previous section reject outliers during the estimation process, they are generally not efficient enough to be the only estimator used. Therefore the outliers have to be identified, so that they can be removed and a second estimation step performed, usually either the least squares or a robust M-estimator. The residuals from the HBPE are standardized by an estimate of the scale of the measurement errors; it makes no sense to use a non-robust estimate such as the standard deviation for this purpose. The most commonly used robust scale estimates are the MAD (the Median Absolute Deviation from the Median) and the length of the shortest half. These are both appropriately scaled, so that at the Gaussian distribution they have the same value as the standard deviation. The MAD of a set of numbers z is defined as

$$MAD(\mathbf{z}) = 1.483 * Med(|\mathbf{z} - Med(\mathbf{z})|) \qquad (50)$$

where $Med(\mathbf{z})$ is the median of $\mathbf{z}$. Then to test for outliers after performing the HBPE, the MAD is computed, and each residual is divided by the MAD, yielding the standardized residuals $\mathbf{r_s}$. If the value of a particular r_{si} is greater than a threshold value, typically 2.5, then the corresponding measurement z_i is identified as an outlier. Rousseeuw [46] performed Monte Carlo simulations using the LMS estimator, and determined that this procedure was not adequate. It was observed that the MAD based on an LMS estimation was relatively low, and that therefore using this resulted in the false identification of some measurements as outliers. He therefore proposed using as a scale estimate

$$\hat{\sigma}_r = C \sqrt{(r_W^2)_{v\,:\,m}} \qquad (51)$$

where $C = 1.483 \left(1 + \dfrac{5}{m-n} \right)$. This has not been justified theoretically, and is based strictly on numerical experiments with some data sets. Therefore in the state estimation simulations described in the next chapter, this technique will be used as a starting point, but it will be modified to take into account the nature of the power system data.

4.4 : Computation of the HBPE

The first question faced at this point is which estimator to use. From a computational viewpoint, the LMS and the LTS are very similar. Neither one has a derivative that can be analytically computed, so gradient-based iterative methods (such as the Newton-Raphson) are not feasible. Neither

method has a convex ρ function ; in fact it is shown in [48] that there are $O(m^2)$ local minima, so any method used must be capable of handling this. The S and τ estimators are also similar; both are much more computationally involved, though for the τ-estimator, an algorithm based on the iterated WLS has been devised. All four methods have a high breakdown point, as well as resistance against bad leverage points and multiple interacting bad data. At this point, much more experience exists with the use of the LMS than any other estimator, and an existing program was available for modification, the PROGRESS package [49]. Therefore for the remainder of this dissertation, the LMS is the primary estimator used. The τ-estimator appears to have some useful properties, and the implementation of this estimator is a topic for future work.

In the next sections, two methods for computing the LMS estimate will be discussed. Neither of these techniques are used directly in the final implementation, but they provide insight into the problems encountered in computing the LMS. The two methods are the combinatorial optimization algorithm used in PROGRESS, and a direct search method.

4.4.1 : The PROGRESS algorithm

PROGRESS is a general purpose multiple linear regression package developed by Rousseeuw, and cannot directly be used for nonlinear regression. It identifies outliers using the methods described previously, using an LMS estimate and a standardized residual based detection test. Since PROGRESS must provide a statistically efficient estimate of the regression coefficients, after rejecting the outliers, an LS estimate is

computed. The breakdown point and the other properties of this combination of two sequential estimates have not been determined, but it appears that if the LMS step rejects all the outliers, the LS stage will produce a valid, efficient estimate.

Let us first consider the PROGRESS algorithm for the LMS in the context of simple linear regression. The measurement set consists of m points of the form (x_i, y_i), and the problem is to find the line $y = ax + b$ that minimizes the least median of the squared residuals. Clearly there is an infinite number of lines of this form, each of which can be considered ; equally obvious is that it is not feasible to investigate all possible lines. The PROGRESS algorithm examines a subset of all possible lines, and uses a probabilistic argument to justify not considering all possible lines. Given that the LMS line bisects a strip that contains at least half the data points, it must pass very close to several of the data points. Also if half the data points lie exactly on a straight line, that line will be the LMS estimate. Combining these two notions, the PROGRESS algorithm considers first the set of all lines that pass through any two of the measurements exactly. This set, unlike the set of all lines, is finite; the number of such lines is $\frac{m!}{2!(m-2)!}$, excluding the possibility of duplication of lines. For small values of m, the LMS criterion can be computed for all the lines in the exact-fit set, and the line that has the lowest value of the criterion is the LMS estimate. Unfortunately, as m increases, the number of lines grows exponentially, which renders this approximation virtually useless, as shown in Table 4.

Table 4 : Growth of the number of combinations in PROGRESS

m	$\dfrac{m\,!}{2\,!\,(m\text{-}2)\,!}$
5	10
10	45
20	190
50	1225
100	4950
150	11175

Therefore for larger values of m, an alternate approach has to be used, which is known as the resampling method. Let the fraction of bad measurements be ε. If the LMS line was determined by one of these measurements, it would be corrupt, but if the subset of all possible lines contained one line based on the good measurements, that line will be the LMS estimate. Therefore the objective is to ensure, with a high degree of certainty, that at least one line is good. The analysis is as follows (with n being the number of parameters being estimated, n=2 for simple regression) : to determine a hyperplane, n points have to be selected out of the m measurements. The probability of selecting 1 bad point is ε, and hence the probability of selecting one good point is $(1-\varepsilon)$, and the chance of choosing a sample of n good points is $(1-\varepsilon)^n$. The probability that the sample will contain one bad point is therefore $1 - (1 - \varepsilon)^n$. If we are selecting k samples, the probability that all k samples are bad is $(1 - (1 - \varepsilon)^n)^k$; and the probability that at least one sample is good is $P = 1 - (1 - (1 - \varepsilon)^n)^k$. Since it is sufficient that one sample be good, this is the probability that must be considered. We can therefore rewrite this equation in terms of the minimum number of samples, k, that must be taken, in order

that P be sufficiently large. Then $(1 - P) = (1 - (1 - \varepsilon)^n)^k \Rightarrow \text{Ln}(1 - P) = \text{Ln}(1 - (1 - \varepsilon)^n)^k$, which leads to

$$k = \frac{\text{Ln}(1 - P)}{\text{Ln}(1 - (1 - \varepsilon)^n)} \tag{52}$$

This is the minimum value of k, for a given value of P. In practice, P is usually 0.95 or greater, and ε is taken as 0.05 to 0.20 depending on the expected quality of the data. Table 5 shows the values of k, for various values of n, P and ε.

Table 5 : Number of Samples required by PROGRESS

n	$\varepsilon = 0.025$ P = 0.95	$\varepsilon = 0.025$ P = 0.975	$\varepsilon = 0.05$ P = 0.95	$\varepsilon = 0.05$ P = 0.975
25	4	5	9	11
50	9	11	37	46
100	36	45	504	621
250	1679	2067	1110711	1367705

The growth in samples is still rapid, but for medium sized data-sets, the problem is still of manageable proportions. Note however the huge growth in the number of samples required by changing ε from 0.025 to 0.05, for n=250 ; a doubling in ε causes k to increase by a factor of over 5000.

In multiple regression, the PROGRESS algorithm is equivalent to taking the design matrix H, with m rows and n columns, and forming a sub-matrix, H',

by selecting randomly n rows from H, and a sub-vector, z', by selecting the corresponding measurements from z. Then the PROGRESS estimate of x for that subsample is $H'^{-1}.z'$. The LMS criterion is then found by computing the residuals, and this process is repeated for the desired number of samples.

There are several reasons why this algorithm cannot be directly used in power system state estimation:

(1) The size of the problem, typically 300-1000 unknowns, makes the number of samples k far too large to be practical, even with the relatively low value of ε used in Table 5.

(2) This algorithm implicitly assumes that the matrix H' has an inverse, which may be true in the relatively full design matrices that are common in other problems. In state estimation, with the very sparse H matrix (5 % non zero elements or less, for high voltage systems), a randomly selected H' matrix is usually singular. In simulations with the IEEE 118 bus system, with m=246, and n=117, the randomly chosen matrix H' was singular in 98 % of the cases.

The PROGRESS algorithm can be applied successfully in the case of smaller systems, such as the IEEE 14 bus test system, where the design matrix is less sparse, and the number of samples required is not impractical. To verify this, tests were performed using the 14 bus system, with 34 real power measurements. The system Jacobian was obtained by linearizing the operating equations about $|V| = 1.0$ p.u, $\theta = 0.0$ degrees ; of the 442 terms in the matrix, 81 are non-zero, and 361 are zero (18% non-zero). For comparison, in the 30-bus system with 56 measurements, there are 135 non-zero elements and 1489 zero elements (8% non-zero), and in the 118 bus

system with 247 measurements, there are 647 non-zero elements among a total of 28899 entries (2% non-zero). The percentage increases if there are more injections in the measurement set, since a line flow can only contribute 2 non-zero entries, a voltage magnitude only 1 entry, while an injection contributes as many entries as there are lines incident on that bus, plus 1.

Table 6 : PROGRESS results with IEEE 14 Bus System

Location of the bad data	Bad value (In MW)	True Value (In MW)	PROGRESS LMS Estimate	Standardized Residual
Line 5-4	63.6	148.8	67.9	105.4
Injection 4	-47.8	-132.8	-57.0	-98.7

Table 6 shows the results of one such test case ; two bad data were placed around Bus 4, which makes this a case of multiple interacting bad data, irrespective of their actual values. The measurement of the power being transmitted into Bus 4 through Line 5-4 was increased by 85.2 MW ; to account for this excess power, the value for the power injected at Bus 4 was reduced by 85 MW. Thus the two corrupted measurements agree with each other, and are therefore both interacting and conforming. Table 6 gives the LMS estimate for this case, which shows that the LMS rejected these measurements, as evidenced by their large standardized residuals, far exceeding the threshold used in PROGRESS of 2.5. This case illustrates the robust nature of the LMS, but is not a sufficiently large case in terms of system size or computation time required.

Other methods have been proposed [50], but the extension of these to multiple regression problems appears to be overly complicated. Instead, the shortcomings of the PROGRESS algorithm will be addressed in the next chapter, with a view to adapting it to power system state estimation.

4.4.2 : Direct Search Methods

In this section, the LMS estimation problem will be formulated as a nonlinear minimization problem. This formulation is complicated by two factors :

(1) The objective function is not convex, and has several local minima, as shown in the previous section. This implies that the choice of starting point is critical, and that the method must be able to handle local minima, which is not simple.

(2) The lack of an analytical expression for the derivative of the objective function limits the choice of solution method, since the most popular optimization techniques depend on the gradient information. This also makes the algorithm slower in converging.

The starting point is not as much of a problem as it seems, since in state estimation, a flat start, $|V| = 1.00$ and $\theta = 0.0$ is usually satisfactory. If this is not adequate, the state estimator can be operated in the tracking mode, where the starting point for one state estimation run is the output from the previous run. Since the state estimation is usually repeated frequently, this relies on the fact that the state of the power system does not usually change drastically between successive runs.

Three optimization methods were tested [51], namely the Hook and Jeeves

method, Rosenbrock's method and the Zangiwill-Powell method. It was found that the latter two methods were less suitable for this problem, since they were more prone to locking in on one search direction, making them less likely to "escape" from local minima. The Hook and Jeeves method, which searches all co-ordinate axes successively, was more suitable, as well as being the least complex method computationally. This method will be explained with reference to a two-dimensional minimization problem, to find the minimum of f(x, y), starting with an initial guess of (x_0, y_0).

In all multi-dimensional optimization methods, the critical component is the one-dimensional search, since the overall minimization is usually accomplished as a series of one-dimensional searches. This is especially true in the state estimation problem, for the following reason. The criterion for comparing two minimization methods is the number of objective function evaluations each requires to achieve the same degree of accuracy. In this problem, a function evaluation is extremely complex ; from the current value of the state of the system, the line flows have to be computed, and from these the injections can be evaluated. Then the residuals have to be found, and only then can the actual objective function be evaluated. In a medium sized problem, with 200 co-ordinate axes, if each one-dimensional search takes 40 function evaluations, 8000 function evaluations are required just to complete one step of the minimization.

The Fibonacci method minimizes the function f(x) over the interval [a,b] with the fewest number of function evaluations, if f(x) is convex over the interval [a,b]. The number of function evaluations to be allowed, N, is pre-specified ; at the end of one stage of the minimization, the uncertainty in the location of

the minimum is $\dfrac{b-a}{F_N}$, where F_N is the Nth Fibonacci number, the sequence F_i = $F_{i-1} + F_{i-2}$, and $F_0 = F_1 = 1$. For $N = 40$, $F_N \approx 100$ million.

Each stage of the Hook-Jeeves method in an n-dimensional minimization problem is made up of (n+1) one-dimensional searches; n of these are along the co-ordinate axes, and the last is along the net resultant direction.

Table 7 : Bad Data Locations for Case 1

Location	Bad data (p.u)	True Value (p.u)
Line 2-6	2.0 + j 1.0	0.62 + j 0.1
Line 6-2	-2.0 - j 1.0	-0.60 - j 0.07
Line 6-28	1.5 + j 2.0	0.19- j 0.0
Line 28-6	-1.5 - j 2.0	-0.19-j 0.01
Line 27-29	-0.1 + j 1.0	0.06+j 0.02
Line 29-27	0.1 - j 1.0	-0.06-j 0.02

Table 8 : Bad Data Locations for Case 2

Location	Bad data (p.u)	True Value (p.u)
Line 4-6	1.7 + j 0.35	0.70 - j 0.16
Line 6-4	-1.69 -j 0.3	-0.69 + j 0.17
Line 6-8	-0.70 - j 0.60	0.30 - j 0.08
Injection 6	-2.0 - j 1.0	0.0 + j 0.0

Table 9 : Bad Data Locations for Case 3

Location	Bad data (p.u)	True Value (p.u)
Line 9-11	2.0 + j 1.0	0.0 + j 0.163
Line 11-9	-2.0 - j 1.0	0.0 + j 0.163
Injection 11	-2.0 - j 1.0	0.0 + j 0.163

This is illustrated in Figure 5 for the two-dimensional problem, with the objective function represented by contours of constant value. This method is still very time-consuming, due to the number of function evaluations, and the complexity of each one. It was not viable for the IEEE 118 bus system, due to the excessive computation time. Hence the results given here are for the 30 bus system, with 125 measurements. This method of state estimation does not use the decoupled assumption, so there are 59 unknowns, with a redundancy of 125/59 = 2.1.

Three test cases were used ; Case 1 has 12 bad data, distributed throughout the power system ; Case 2 has eight conforming bad data, all concentrated in the vicinity of a bus with a large number of measurements near it. Case 3 has six conforming bad data, at a bus with very few measurements. The actual locations and values of the bad data are shown in Tables 7, 8 and 9.

The starting point for the estimators was the true state of the system, corrupted by the addition of zero mean Gaussian noise, with $\sigma = 0.01$ p.u for the real part of the voltage and 0.005 p.u for the imaginary part. The true values were obtained from a standard loadflow of the system. Both the estimators were used here, the LMS and the LTS with 20 percent of the

measurements rejected. In all cases 4 iterations of the Hook-Jeeves method were performed.

Table 10 shows the results for these two estimators applied to the three cases described. The estimation error is averaged over the entire system, both for voltage magnitudes and phase angles. From Table 10, it can be concluded that the LMS and LTS are feasible in terms of their ability to handle multiple interacting and conforming bad data. This would not be true of the M-estimators in these cases, since the number of bad data exceeds the global breakdown point, which would be (125/59) or 2.1 at the maximum. However the methods were not satisfactory in two aspects, their ability to find the global minima in the presence of large numbers of local minima, and the computation time they required, especially in larger systems.

Table 10 : Average Errors for the LMS and LTS Estimators

Case	LEAST TRIMMED SQUARES		LEAST MEDIAN OF SQUARES	
	ΔV (p.u)	$\Delta\theta$ (°)	ΔV (p.u)	$\Delta\theta$ (°)
1	0.034	0.178	0.027	1.00
2	0.036	0.262	0.026	1.01
3	0.038	0.189	0.029	0.71

4.5 : Three Bus Example

Figure 6 shows a three-bus example system, which is taken from [13]. Let us assume that all the lines in the system have an impedance of $0 + j\,1$, except

line 1-3, which has an impedance of 0 + j 0.2. Since we are considering the real power problem, there are two unknowns, θ_1 and θ_2, (bus 3 is taken as the reference bus). The 6 measurements are also shown in Figure 6, 4 line flows and 2 injections, while Figure 7 shows the corresponding factor space for this system. The linearized system model is then

$$
\begin{bmatrix} z_1 \\ z_2 \\ z_3 \\ z_4 \\ z_5 \\ z_6 \end{bmatrix} = \begin{bmatrix} 1 & -1 \\ -1 & 1 \\ 5 & 0 \\ 0 & 1 \\ -1 & 2 \\ -5 & -1 \end{bmatrix} \begin{bmatrix} \theta_1 \\ \theta_2 \end{bmatrix}
\tag{53}
$$

In the base case, with all angles zero, the true value of all the measurements is 0. We now introduce 2 bad data, $z_3 = -0.55$ and $z_6 = 1.0$ (note that these two measurements are leverage points, since the impedance of line 1-3 is much smaller than that of the other lines).

Table 11 : S-PLUS Example Estimates for LS, LAV and LMS

Variable	Least Squares	Least Abs. Value	Least Med. Sqrs
θ_1	-8.25°	-6.30°	0.0°
θ_2	-6.13°	3.15°	0.0°

Table 11 shows the results of the estimation for three methods, the LS, the LAV, and the LMS, while Table 12 shows the residuals for these methods. These estimates were all obtained by using the package S-PLUS [52] ; the

LAV and LS algorithms are standard, while the S-PLUS implementation of the LMS uses the PROGRESS algorithm.

From Table 12, it can be seen that the LAV estimate passes through measurement 3 exactly. This is typical of the LAV in the presence of bad leverage points.

Table 12 : S-PLUS Example Residuals for LS, LAV and LMS

Measurement	Least Squares	Least Abs. Value	Least Med. Sqrs
Line 1-2	0.037	0.055	0.0
Line 2-1	-0.037	-0.055	0.0
Line 1-3	0.171	0.000	-0.55
Line 2-3	0.107	0.055	0.0
Injection 5	0.069	0.000	0.0
Injection 6	0.172	0.395	1.0

This example will also be used to illustrate the weaknesses of other methods that were proposed to solve the problem of multiple interacting bad data. These are the Identification by Elimination method of Wu *et al* [13], the Hypothesis Testing and Identification technique of Mili *et al* [15], and the Measurement Compensation method of Slutsker [53]. The latter two methods fail if after the LS estimate, the largest normalized residual does not correspond to a bad measurement. Wu's method relies on removing the measurement with the largest normalized residual, and re-computing the estimate. The measurement with the largest residual is again removed, and

this process is repeated until all the normalized residuals are lower than the threshold value of 2.5. The value of σ_i used to standardize the residual is 0.01.

Table 13 : Example of Identification by Elimination method

Measurement #	W_{ii}	r_{NI} Estimate 1	Estimate 2	Estimate 3
1	0.90	3.42	2.90	1.54
2	0.90	-3.42	-290	-1.54
3	0.36	3.27	2.11	-2.39
4	0.96	**8.32**	---	---
5	0.74	5.70	**6.75**	---
6	0.14	6.06	3.40	-2.39

In this case, measurement 4 is falsely identified as bad at the first stage, and finally, measurements 4 and 5 are rejected ; neither one of which is bad. This is illustrated in Table 13.

4.6 : Summary

In this chapter, the class of estimators known as the High Breakdown Point Estimators has been defined. Their theoretical properties, which make them suitable for high dimension multiple regression problems, especially those with multiple bad data and bad leverage points, have been reviewed. Since

the computation of these estimators is not straightforward, some aspects of the implementation of these estimators have been discussed. The findings of this chapter are that the HBPE are ideal for the power system state estimation problem from a statistical viewpoint, but that much attention has to be paid to the way they are implemented, otherwise they are too time-consuming computationally to be of any practical use.

In the next chapter, this issue will be treated in detail, with the objective of making the HBPE feasible in quasi-real-time for systems of about 100 busses or more.
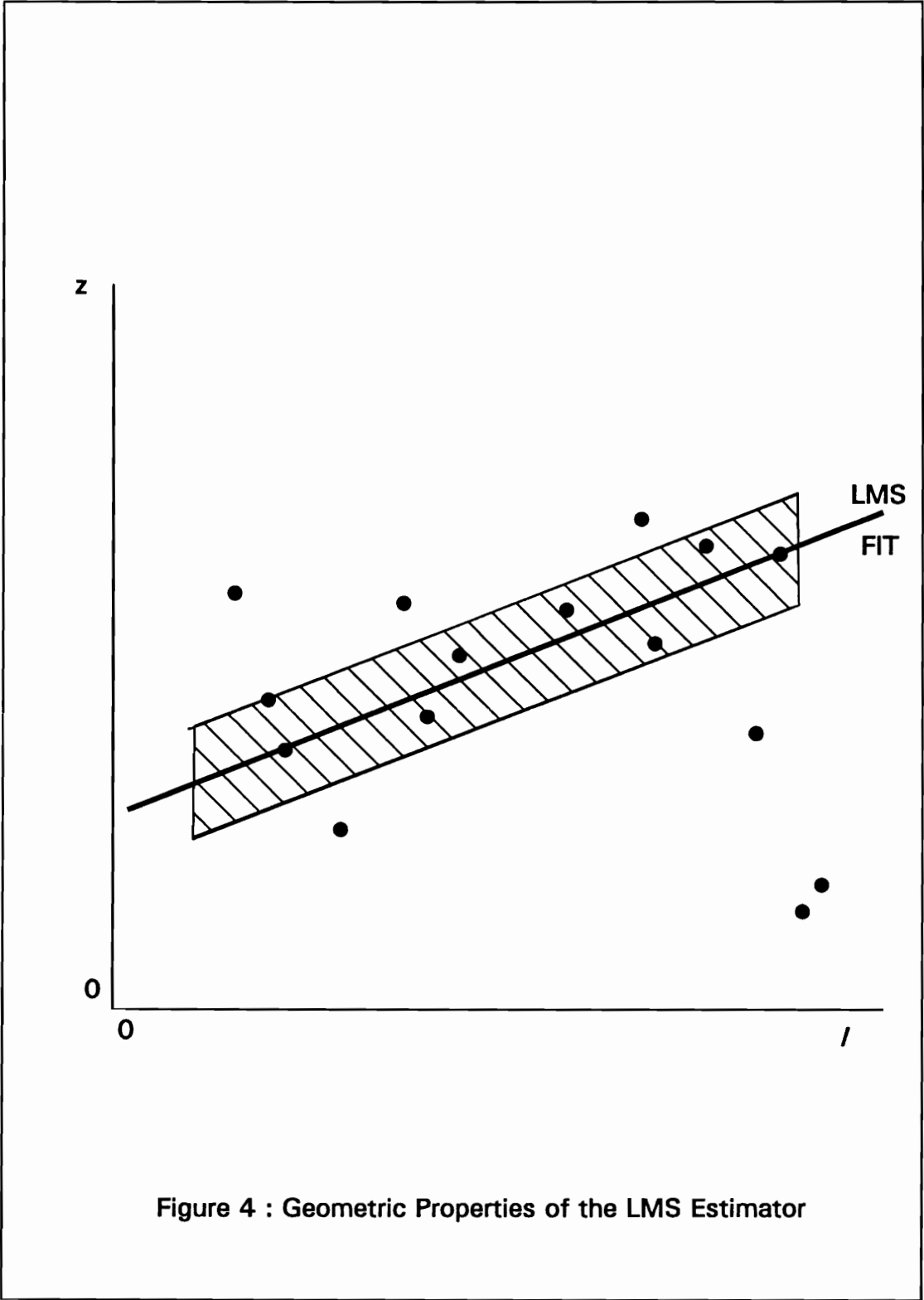
Figure 4 : Geometric Properties of the LMS Estimator
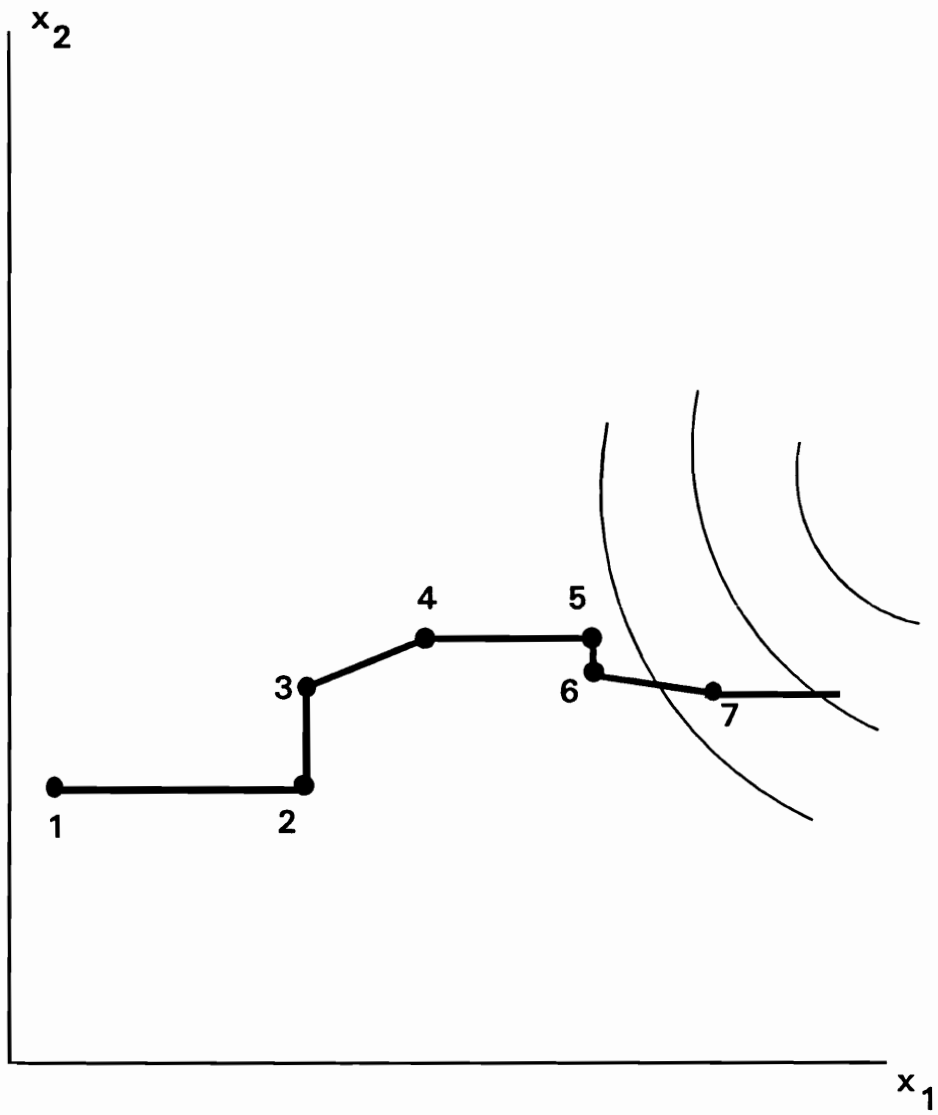
Figure 5 : Hook-Jeeves Method for Direct Optimization
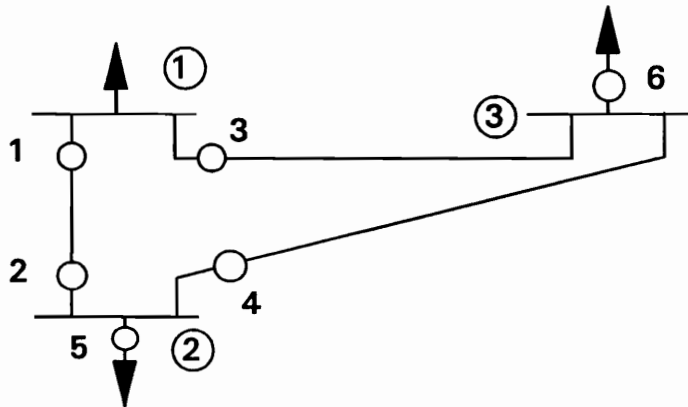
Figure 6 : Three Bus System - One Line Diagram
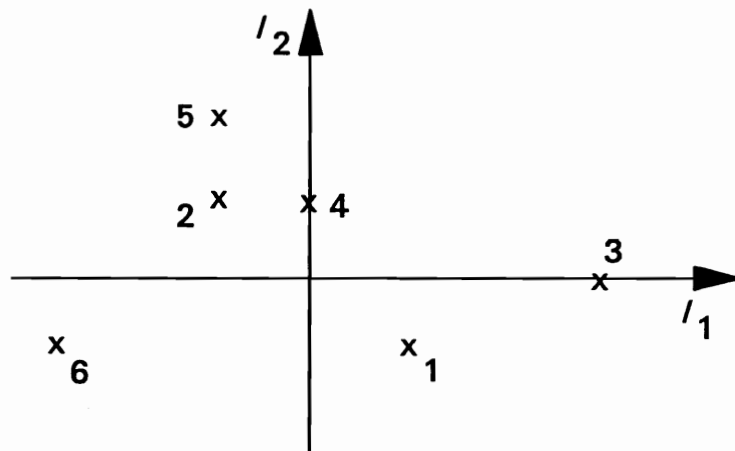
Figure 7 : Three Bus System - Factor Space

Chapter 5 : Implementation and Testing of High Breakdown Point Estimators

5.1 : Introduction

In Chapter 4, the class of estimators known as High Breakdown Point Estimators (HBPE) was introduced, and some methods proposed for their implementation were discussed. It was shown that these methods are not practical for application to real-time power system state estimation as-is. This chapter describes efforts to modify the methods, so as to make possible (and feasible) the use of the LMS in state estimation. An initial objective was to complete the estimation of the IEEE 118 bus system in about 1 minute. In order to achieve this goal, the PROGRESS algorithm had to be altered in several ways, for the following reasons :

(1) Sparse Matrix Methods : The conventional matrix inversion and linear system solution used in PROGRESS was not suited for large sparse systems.

(2) Computational Issues : These are related to the actual algorithm used, and include numerical accuracy, the use of parallel processing, input / output considerations, the division of the task into an off-line and on-line component, etc.

(3) Observability : The implicit assumption that in most cases, the sub-

matrix selected was non-singular is not true in power system state estimation.

(4) Local Redundancy and Breakdown Point : This is a consequence of the sparse design matrix, and affects the manner in which the robust estimator fails. This mechanism causes failure at levels of bad data below the global breakdown point of the estimator.

All of these issues are addressed in this chapter. Detailed simulation results are given for the 14 bus system, and computation times are given for the 14, 30 and 118 bus systems.

5.2 : Review of the Problem

As defined in Chapter 2, the power system state estimation problem consists of estimating the state of the system $\mathbf{x}$, the voltage magnitudes $\mathbf{V}$ and the phase angles θ, from a set of measurements $\mathbf{z}$, which contains line flow, bus injection and voltage magnitude measurements. An approximate model of the system, based on several assumptions previously outlined, results in a system of nonlinear equations of the form

$$\mathbf{z} = h(\mathbf{x}) \tag{54}$$

where $\mathbf{x}$ has n components, and there are m measurements that comprise $\mathbf{z}$. This system can be linearized about the normal operating point, $\mathbf{V} = 1.0$ p.u and $\theta = 0.0$ degrees, so that an even more approximate model of the system is $\Delta\mathbf{z} = \mathbf{H}.\Delta\mathbf{x}$, where $\mathbf{H}$ (known both as the design matrix and the Jacobian matrix) $= \partial\mathbf{h} / \partial\mathbf{x}$. Note that since there are more equations than unknowns, there are multiple solutions to this overdetermined system of equations.

The normal iterative procedure for the solution of this system is to assume an initial value of **x** and find the mismatch between the measurements **z** and the estimated measurements obtained from (54). This can then be used to update **x** by finding an "optimal" Δ**x** among the many solutions that exist. The new estimate of the state can then be used to restart the iteration process by computing a new mismatch vector. If necessary, the matrix **H** can be updated as well, by recomputing the partial derivatives at the new state estimate. Whether this is required is dependent on the numerical properties of the method used, as well as the physical properties of the power system parameters. For example, in lower voltage sub-transmission systems, the line X/R ratio is lower than in EHV systems. This causes numerical convergence problems if the Jacobian is not updated periodically, especially if the decoupled model is being used (**V** and θ being estimated separately).

In the PROGRESS algorithm, the "optimal" solution is found by repeatedly forming a sub-matrix from n randomly selected rows of the **H** matrix, and using this sub-matrix to compute the correction Δ**x**, and hence update the residuals. The value of Δ**x** that minimizes the objective function is chosen to update **x**, and the process continues in the same manner, until the magnitude of Δ**x** is lower than a specified value, or the maximum number of iterations has been reached.

In the next few sections, the modifications to this procedure that were needed in order to make the robust estimation techniques being proposed competitive computationally, especially for large power system models, are described.

5.3 : Sparse Matrix Methods

In almost all aspects of power system analysis, where it is necessary to deal with large systems, the sparsity of the system has to be exploited. This is particularly true when the application being considered is to be run in real-time. Sparse matrix techniques were introduced into power systems by Tinney *et al* in the late 1960's [5,6], and have been an active research topic since. In this section, a brief overview of sparse matrix methods will be given, with emphasis on the techniques that were used to speedup the LMS estimator. By definition, a sparse matrix is one whose elements are predominantly zero. As illustrated in the previous chapter, the design matrix for large EHV power systems is very sparse, with 10 % non-zero elements being rather high. In general, the larger the system, the more sparse the design matrix.

The basic operation for which sparse matrix methods are usually considered is matrix inversion. This step is required by almost every power system analysis algorithm in one form or another, and this is again particularly true in our case. The fundamental equation to be solved is of the form

$$\mathbf{A}.\mathbf{x} = \mathbf{b} \qquad\qquad (55)$$

where $\mathbf{A}$ is a square matrix, whose inverse is assumed to exist (the case where this is not true will be dealt with later), and is sparse, and $\mathbf{x}$ and $\mathbf{b}$ are vectors of the same length. The solution to this system is of course $\mathbf{b} = \mathbf{A}^{-1}.\mathbf{x}$. If b is directly computed in this manner (by explicitly inverting A, and performing the matrix multiplication $\mathbf{A}^{-1}.\mathbf{x}$) there are two problems that arise. Numerically, this is perhaps the worst way to compute $\mathbf{b}$, since it

introduces the possibility of large errors in b, especially if A is ill-conditioned. Secondly, if **A** is sparse, its inverse is generally full. Therefore from a computational viewpoint, if **A** is of dimension n x n, to determine one element of **b** requires n^2 multiplications, and to compute **b** takes a total of n^3 operations. This seems excessive, when it is considered that **A** had perhaps only $t = 0.1(n^2)$ non-zero elements. To compute **Ax** in this case would take $0.1(n^3)$ operations, while $\mathbf{A^{-1}b}$ takes 10 times as many. This ratio gets more lopsided as the system size increases, as shown previously. Clearly the inversion of **A** introduces some extra computations and problems. The question is, can b be computed without computing $\mathbf{A^{-1}}$ explicitly, in a way that minimizes the possibility of numerical problems, and introduces as few extra operations as possible ? This is the main question that was addressed by Tinney in his seminal paper [54] and by Carpentier [55] independently. A secondary issue was that if the product $\mathbf{A^{-1}b}$ had to be computed repeatedly, usually with **A** remaining constant and **b** changing, then it seemed only natural to store $\mathbf{A^{-1}}$ and use it over and over, without recomputing it each time. At the time these methods were developed, computer resources were at a premium, so storing the full matrix $\mathbf{A^{-1}}$ rather than **A** was not acceptable.

Tinney's approach consists of two steps, which when implemented together, provide a great improvement over the direct inversion of **A**. They are based on the decomposition of the matrix **A** into two triangular factors, **A** = **LU**; this operation is known by many names, the most common of which is the Cholesky decomposition. The first step, which can be used for all matrices, full or sparse, consists of a scheme whereby the sequence of operations needed to compute the decomposition could be stored, in a manner

independent of the value of **b**. It therefore allows repeated use of the inverse efficiently. The second step, which was perhaps the greatest contribution, was a method of re-ordering the equations that constitute the system Ax = b, such that the sparsity (if any) of the original system was preserved. These schemes are known as optimal ordering schemes. Not all sparse matrices lend themselves to being re-ordered in this manner; however diagonally dominant matrices, such as those encountered in the admittance matrix of a power system, and in loadflow problems can directly use these methods. The design matrix in WLS state estimation can also benefit; however the matrices used in algorithms similar to that in PROGRESS cannot use it directly. Later in this section, the modifications to Tinney's original scheme that were made in order to utilize the optimal ordering will be described.

5.3.1 : Triangular Decomposition

This is based on the standard Gaussian elimination method used to compute either the inverse of **A** or the solution to (55). The terms below the main diagonal are eliminated by row manipulation ; if necessary some re-ordering in order to maintain the numerical accuracy is also performed [56]. This is usually done by the process of pivoting, with the objective of making the matrix diagonally dominant. In this case, there is no pivoting performed, instead the re-ordering is done at a later stage, and is done to minimize the loss of sparsity. The result of this step is to transform the augmented matrix

$$\begin{bmatrix} a_{11} & a_{12} & a_{13} & \cdots & a_{1n} & b_1 \\ a_{21} & a_{22} & a_{23} & \cdots & a_{2n} & b_2 \\ a_{31} & a_{32} & a_{33} & \cdots & a_{3n} & b_3 \\ \cdot & \cdot & \cdot & \cdots & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdots & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdots & \cdot & \cdot \\ a_{n1} & a_{n2} & a_{n3} & \cdots & a_{nn} & b_n \end{bmatrix} \tag{56}$$

to

$$\begin{bmatrix} 1 & a'_{12} & a'_{13} & \cdots & a'_{1n} & b'_1 \\ & 1 & a'_{23} & \cdots & a'_{2n} & b'_2 \\ & & 1 & \cdots & a'_{3n} & b'_3 \\ & & & \cdots & \cdot & \cdot \\ & & & \cdots & \cdot & \cdot \\ & & & & \cdot & \cdot \\ & & & & 1 & b'_n \end{bmatrix} \tag{57}$$

The elimination is usually performed one column after another; Tinney recommends that it be done row-wise. Both methods yield the same final result, a reduced matrix system, whose solution can easily be found by the process of backsolving. That is,

$$x_n = b'_n, \ \ x_{n-1} = b'_{n-1} - a'_{n-1,n} x_n, \ \ x_i = b'_i - \sum_{j=i+1}^{n} a'_{ij} x_{ij} \tag{58}$$

For large n, it has been shown [5] that this method results in $n^3/3$ multiplications rather than the n^3 multiplications required by a full inversion.

The complete sequence of operations required to form this system can be stored in a matrix of the same dimensions as **A**, using the following notation :

D_i : This is the equal to the identity matrix, with the element at location (i,i) replaced by d_{ii}.

L_i : This is equal to the identity matrix, with the ith column replaced by the vector $[\ 0, 0, \ldots, 1, -l_{i+1,i}, -l_{i+2,i}, -l_{i+3,i}, \ldots -l_{n,i}]^t$.

U_i : This is equal to the identity matrix, with the ith row replaced by the row vector $[\ 0, 0, \ldots, 1, -u_{i,i+1}, -u_{i,i+2}, -u_{i,i+3}, \ldots -u_{i,n}]$.

In these matrices, the elements are some of the various multipliers used during the triangularization of **A**, and are the only numbers that are required to be stored for future use. This set is called the factor table. We have

$$\mathbf{A^{-1}b} = U_1 U_2 U_3 \ldots U_{N-1} D_N L_{N-1} D_{N-1} L_{N-2} \ldots L_2 D_2 L_1 D_1 \mathbf{b} \qquad (59)$$

Therefore if **b** is changed, a new value of **x** can be computed by carrying out the matrix multiplications indicated above. It is shown in [5] that this requires n^2 multiplications.

5.3.2 : Optimal Ordering

Unlike the situation in the previous section, this material applies only to sparse matrices. It is normal in Gaussian elimination to re-order the equations for a specific purpose; usually to improve the numerical performance of the solution by the pivoting technique. In this section, some alternate methods for re-ordering the equations will be described. These are known as Tinney's Scheme I, II and III, after their originator ; with Scheme I

being the simplest and the least effective, and Scheme III being extremely complex and hence rarely used. A modified version of Scheme II was used in the implementation of the LMS.

The objective behind all of Tinney's schemes is to minimize the number of extra non-zero terms created during Gaussian elimination. The following example illustrates the basic principle :

$$A = \begin{bmatrix} 5 & 2 & 1 & 1 \\ 10 & 14 & 0 & 0 \\ 3 & 0 & 6 & 4 \\ 1 & 0 & 0 & 7 \end{bmatrix} \qquad (60)$$

If the standard Gaussian elimination were to be performed, the first step would be to multiply Row 1 by (-2) and add the result to Row 2, which makes the new $a_{21} = 0$, however it also causes the new a_{23} and a_{24} change from zero to non-zero (to -2). Instead suppose row 4 and row 1 were interchanged. Then only a_{24} would change, since the new term a_{13} would be zero. Thus we have reduced the fill-in by 1 term, and kept the size of the factor table smaller. Note that this change would not be required for the sake of numerical performance, since the diagonal terms of the matrix are the largest in their rows. The three schemes of Tinney are heuristic methods to minimize the number of fill-in terms, and thus make the computation of $A^{-1}b$ faster, since there would be more multiplications by zero, which do not have to be performed. A secondary effect is to minimize the amount of memory required to store the factor table ; this is of far less consequence today than at the time these methods were developed.

<u>Scheme I</u> : The method used in the example above is Scheme 1. In this scheme, the rows with the maximum number of off-diagonal zero terms is made row 1, the row with the second largest number of off-diagonal zero terms is made row 2, etc. A simple justification of this is based on the number of operations conducted on each row. During the elimination, Row 1 of **A** is not altered, Row 2 is altered by adding a multiple of Row 1, Row 3 is changed by adding a multiple of Row 1 and a multiple of the updated row 2. The chances of a zero term remaining zero are higher if fewer operations are performed on it, hence the rows should be rearranged so that those with many zero terms are near the top of the matrix. This is a very simplistic algorithm, since the effects of the elimination process on the number of zero terms is not considered. For example, after the first stage of the elimination, the updated Row 2 may not have any more zero elements left, if they happen to lie beneath non-zero terms of Row 1. Of course, it is very simple to implement Scheme 1, since the only information required is the number of zero terms in each row of A. Scheme 1 represents a tradeoff between effectiveness and simplicity, with extreme simplicity attained at the cost of a significant loss of efficacy.

<u>Scheme II</u> : This scheme, and Scheme III both require some simulation of the elimination process, in order to predict where the fill-in might occur. In Scheme II, the principle is to order the rows so that the next row to be operated upon has the maximum number of zero elements. It starts by choosing as Row 1, the row with the maximum number of zeros. The first

step of the elimination process is then simulated (not performed) by finding that row which would have the maximum number of zeros left, after the appropriate multiple of Row 1 was added to it. Note that the values of the non-zero elements are not evaluated at this stage, only the number of zeros. That row is then moved to Row 2. Now the next step, namely adding a multiple of Row 2 to Rows 3 and beyond is simulated, and the row with the maximum number of zeros after this step is chosen as Row 3. The process continues in this manner till all rows have been ordered. This scheme may be described as a one-step look-ahead, where the effects of the next step are predicted, and the most beneficial outcome is made to occur. Note that this scheme does not take into account the possibility that a zero term may be created by chance when two non-zero elements are added.

Scheme III : This scheme requires a complete simulation of the elimination process. Instead of looking ahead and choosing as the next row, the row that will have the maximum number of zero terms, the next row is chosen such that this will introduce the fewest number of fill-in terms. This is recursive in nature, since the choice of row 2 depends on the choice of row 3, and that in turn depends on the choice of row 4, etc. Essentially then this scheme requires the simulation of the fill-in for all possible orderings of the rows. Again, note that the actual values are not computed during the simulation, merely the number of zero terms destroyed is tracked.

A comparative study of these three ordering schemes was performed by Stott [57], in which it was concluded that Scheme II is the most suitable choice for most power system applications. While it is slightly more complicated than

Scheme I, the extra computation in choosing an ordering is more than justified by the consequent reduction in effort needed to compute the solution. Scheme III is far more complicated than Scheme II, and does not produce consistently superior results.

Most of the tests of these schemes were performed with two types of network matrices in mind; the admittance matrix of a system, and the loadflow Jacobian. These matrices rarely have a diagonal entry of zero; this occurs only in special situations in the loadflow Jacobian [58], for example, when the voltage magnitude at one bus is determined by the reactive power injection at a bus not connected to the original bus. This is not true in the case of the submatrices drawn from the measurement Jacobian. There is no intrinsic reason for the diagonal elements of these matrices to be non-zero; and in practice it was found that they are very often zero.

The schemes just described do not handle this case, since they assume that the elimination can proceed at all stages, which is not the situation when tone of the diagonal elements becomes zero. Tinney investigated this problem [6], and his findings are used as the basis of the method employed in the implementation of the LMS. This method is somewhat unusual in that it selectively tries to increase the fill-in of the matrix, and that both rows and columns are interchanged. It is based on Scheme II, with these modifications : at the first step, the variables are divided into two classes, those whose corresponding matrix diagonal terms are zero (Type 1) and those that have non-zero diagonal entries (Type 2). The next variable to be eliminated is then the Type 2 variable that has the maximum number of non-zero entries in its row and column. The elimination of this variable is then simulated, to

check if any diagonal terms which were zero have now become non-zero, which is equivalent to reevaluating the variable type. The next variable to be eliminated is then chosen, based on the criterion just mentioned. It is postulated, but not proven analytically, that if at the end of this process, there remain some zero diagonal entries, the matrix is singular. This method works well as long as the bulk of the diagonal terms are non-zero. This condition was generally not true of the matrices used in the LMS procedure. Due to this, the method has to be modified to suit our application. The submatrix drawn is very likely to be singular ; this corresponds to the state of the system not being observable based on the subset of measurements selected. Therefore the method was modified so that at the simulation stage, the row/column interchanges were performed so as to cause maximum fill-in at the zero diagonal terms, rather than to minimize the overall fill-in. If at the end of the simulation it was found that some zero diagonal terms still existed, the matrix was assumed to be singular, and no attempt to re-order it would be made. This served as one check for singularity (other checks will be described in the section on observability). If on the other hand, the zero diagonal entries were removed at some stage, from that point onwards the method would revert to the modified Scheme II described by Tinney. This would minimize the fill-in to some extent, while guaranteeing the existence of a solution to $\mathbf{A}^{-1}\mathbf{b}$.

5.4 : Computational Issues I - Offline Methods

At this point, the PROGRESS algorithm was modified by replacing the regular Gaussian elimination routine in it with a sparse, optimally ordered

routine based on the methods just outlined. It was found that though this did indeed decrease the computation time substantially, the gain was not enough to make the method viable. In addition, it was found that a very large fraction of the sub-matrices drawn (over 90 %) were singular, and so the probabilistic analysis of the chance of selecting at least one good sample was rendered irrelevant. The entire approach to the problem had to be radically altered, since the method was not fast enough, nor was it guaranteed to be robust.

The principal advantage gained by using a sparse matrix method is that repeated solutions of the same system of equations could be obtained rapidly, when the vector **b** changed. In the context of the PROGRESS algorithm, this was not of much consequence, since each set of equations (a sub-matrix) would only be used once. Therefore a way to re-use the factored sub-matrices was needed, in order to fully exploit the sparse matrix methods. Two methods of achieving this were considered.

5.4.1 : Partial Refactorization and Sparse Vector Methods

A recent development in the area of sparse matrix methods is the Partial Matrix Refactorization technique of Brandwajn et al [59]. This was devised to handle the situation where a sparse matrix A has been factored, and the factors stored for future use. Then, if some terms in the matrix **A** change slightly, the question is whether the pre-computed factors have to be rebuilt from scratch, or can the factors be updated to allow for the effects of the change in **A**.

Techniques were developed, which allow for the updating of the factored

matrix directly. These rely on refactoring the largest submatrix that contains all the modified elements. If however, these elements are near the top of **A**, then updating the terms affected by the change will require almost as much effort as recomputing the factors. In [59], two methods are described that address this problem. By using suitable data structures, it was shown that it was possible to identify the individual terms in the factors that are to be updated, rather than an entire block. Both the algorithms proposed are based on the sparse vector methods [60]. These methods were originally developed for the purpose of solving only a portion of $\mathbf{x} = \mathbf{A}^{-1}\mathbf{b}$, so that if a few terms in **b** changed, the entire solution would not have to be recomputed.

To illustrate the use of these methods in a PROGRESS type algorithm, consider the following scenario. After randomly selecting a sub-matrix from the design matrix, and computing its inverse and objective function value (the LMS) ; instead of choosing the next sub-matrix randomly, let it be a perturbed version of the first sub-matrix. Then by using the partial refactorization methods, the new estimate and LMS value can be determined quickly as updates to the previous values. This process can be repeated, so that a large number of sub-matrices can be analyzed rapidly, each a modified version of the previous one.

The main reason why this method was not used lies in the reasoning behind the probabilistic argument for the number of sub-samples to be drawn. Quoting from Chapter 4, "*the probability that the sample will contain at least one bad point is therefore $1 - (1 - \varepsilon)^n$. If we are selecting k samples, the probability that all k samples are bad is $(1 - (1 - \varepsilon)^n)^k$*". This argument holds

if (and only if) each of the k samples is independent of the others. If one sample is created by the modification of a prior sample, then the probability that the second sample is bad is definitely a function of whether the first sample was bad. Therefore the reasoning used to determine the number of sub-samples required does not hold true if such methods based on partial refactorization are used. This method does not address the question of the singularity of the sampled sub-matrices in any way. In a later section, the possibilities for the limited use of these methods will be considered, but it is clear that they cannot be used as the primary algorithm without invalidating one of the basic assumptions behind this technique.

5.4.2 : The Offline /Online Task Division

The objective of the work described in this section is to find a method which fully utilizes the sparsity of the Jacobian matrix, and maintains the validity of the assumptions made previously regarding the independence of the samples. It is also desirable that the method reduce the amount of time spent trying to invert singular matrices.

Let us assume that the estimation process can be divided into two sections, on the following basis : *All tasks that are directly dependent on the value of the measurements are designated as on-line tasks, while all other tasks are called off-line tasks.* The significance of this seemingly inconsequential statement is that if a large portion of the overall work can be classified as off-line, and if the output of those tasks can be made available to the on-line tasks in a suitable way, then those tasks need not be performed in real-time. Therefore a large portion of the work can be done in advance, or even on a

separate computer, as long as the results are available when required.

Examining the LMS estimation problem, the following tasks can be designated as off-line tasks:

(1) The formation of the Jacobian matrix. This can be updated at regular intervals, to take into account any topology changes in the system. If so desired, the Jacobian can be formed not at the flat start $\{|\mathbf{V}| = 1.0 \text{ p.u}, \theta = 0.0\}$, but at the most recent output of the state estimator. This can hasten the convergence of the estimator, if the state does not change drastically.

(2) The selection of the sub-matrices. We can also use this opportunity to deal with the singularity problem, to a large extent. It is required that we draw at least k independent samples; instead of just selecting them randomly and then finding out in the on-line stage that many of these are singular, the selected sub-matrices can be tested for singularity off-line. If a sample is singular, it can be discarded, and another sample selected. Thus instead of having k independent samples to use at the on-line stage, there will be k independent non-singular samples. This simple mechanism greatly increases the chance of finding one good sample that is also observable.

[3] The inversion of the selected sub-matrices. If we can invert the sub-matrix, and store the inverse in an efficient manner which can be retrieved during the on-line stage, then this shifts the major computational burden from the on-line to the off-line phase of the work. The optimally ordered sparse triangularization methods of Section 5.3 are well suited for this ; in fact, they combine both Step 2 and this step.

The selected sub-matrix is triangularized as soon as it is drawn ; if during the optimal ordering or the elimination process, it is found to be singular, it is rejected. On the other hand, if it is observable, then the inverse is stored in the factored format of equation (55). This utilizes the minimization of the storage that was so important when the sparse matrix methods were devised. The on-line computation then consists of reading this inverted matrix, multiplying it by the corresponding measurements, and computing the LMS criterion. The input / output has been reduced substantially by the use of the factored format for storage, and the inverse has already been computed, so this re-structuring of the problem shifts most of the work from the on-line, real-time phase to the off-line phase.

This division of labor made it possible to complete the state estimation for systems of upto about 100 busses in times that are reasonably close to the goal that was established at the beginning of this chapter.

5.5 : Observability Analysis

A rigorous observability analysis would have been extremely important if the task division of the previous section had not been implemented. In that case, the observability analysis would be the main singularity check, and it would be performed in real-time, so speed would have been of the essence. Even though it is not as time-critical as before, it was found that some form of observability checking had to be performed, even in the off-line stage. As mentioned in Chapter 4, the randomly selected sub-matrices are singular in an overwhelming fraction of the cases, and this fraction gets larger as the

system size increases. By implementing some simple tests for observability, a large amount of effort could be saved in the off-line portion of the work. Some of the concepts developed as a result of this effort are of great significance later. A rigorous, but still very fast, observability algorithm was implemented separately by M. Cheniae as part of his M.S thesis [61]. He has incorporated this method into the off-line sample selection algorithm, and in future work, this will be one of the methods used.

5.5.1 : The Fundamental Set and Sample Selection

In this section, the concept of a *fundamental set* for each unknown variable is introduced. This concept will play a significant role in the selection of the sub-matrices for the LMS estimation, and it will be used in subsequent sections to analyze anomalous breakdown behavior, and in other applications such as meter placement and leverage point identification. It is applicable to other areas where large sparse design matrices are used.

Definition : In the multiple regression model, $z = H.x$, the fundamental set of the variable x_i, F_i, is defined as the set of those measurements which have non-zero terms in the ith column of H. Mathematically

$$F_i = \{ z_k \mid H_{k,i} \neq 0 \} \qquad (61)$$

The immediate application of the fundamental set is to change the way sub-matrices are selected. This is because of a theorem in Mili, Phaniraj and Rousseeuw [62] which shows that there is a link between the fundamental sets of the system, and its observability.

Theorem *: A necessary condition for the system to be observable for a given sample is that at least one measurement from each fundamental set be included in the sample.*

Proof : This theorem can be proven by means of a contradiction : let us assume that there exists a sample for which the system is observable, and that there is at least one fundamental set which has no members in the sample drawn. This implies that there exists at least one state variable which is not related to any of the measurements. This state variable then cannot be estimated from the sample; this however contradicts the assumption that the system is observable for that sample. ▢

This theorem can also be proven based on graph theory [63]. If there exists a sample in which at least one fundamental set F_i is not represented, then no measurement can be assigned to branches incident on the node associated with the state variable x_i. Hence no spanning tree of full rank can be found for this sample, and the system is therefore unobservable for this sample.

The application of the this theorem to the sampling process, is as follows : let us first compute the fundamental set for each state variable in the system. This can be done directly from the Jacobian with very little effort. Select randomly one measurement from F_1 ; then check whether all measurements in F_2 have been picked. If not, randomly select a measurement from F_2 (if the measurement selected has already been picked, repeat until an unused measurement is picked). At this point, repeat the process used to select a

measurement for F_2 on F_3, and then for F_4 etc. If at any stage, a fundamental set is encountered, all of whose members have already been selected, then an observable sample cannot be drawn with the existing selections. The process is restarted with F_1, and since the measurements are picked randomly, it is hoped that a new sequence of selected measurements will result, and ultimately, one measurement can be selected from each F_i, $\forall i \in 1, \ldots, n$.

This procedure results in the selection of samples that satisfy one of the necessary conditions for observability, and has resulted in a much greater proportion of infeasible samples being filtered out, before the time-consuming triangularization step. One change which further improves the process is to re-order the fundamental sets in order of increasing size before starting to select measurements. Without this change, a small fundamental set, which might happen to be near the end of the estimation process, would cause a lot of failures, since the chances that all its members had already been selected is relatively high. Now all the small sets are grouped at the beginning of the process, which improves the chances that an unused measurement can be found in each of them.

Another simple set of tests can be applied to the sample that results from this procedure, to further eliminate any unobservable samples. These tests are based on power system observability [64], and are discussed next.

5.5.2 : Post Selection Analysis

These tests are performed on the sub-matrix once it has been generated, and further reduce the chance that an unobservable sub-matrix is factored. They

are also based on necessary but not sufficient conditions for observability. That is, if any of these tests fails, the system is not observable for that sub-matrix selection; but the sub-matrix is not guaranteed to be observable by the fact that all these tests succeeded.

(1) The double-ended line test. If the line flow at both ends of a line are measured, and both these measurements are drawn in the same sample, the resulting sub-matrix is singular. This is because the each of the two corresponding rows in the sub-matrix will be the negative of the other, and therefore, by adding the first row to the second, a null row will be created in the sub-matrix. This operation does not affect the determinant of the matrix, and a matrix with a row made up of all zeros is singular. Therefore a sample containing measurements of the flow from both ends of a line is not observable, and hence the sample can be rejected without further analysis.

(2) The complete set at a node. If all lines leaving a particular node are metered, and the injection at that node is also measured, and all these measurements are selected in one sub-matrix, then that sub-matrix is singular. This is illustrated by the sub-matrix in the example below :

$$
\begin{bmatrix}
2 & 0 & -2 & 0 & 0 \\
0 & 1 & 0 & -1 & 0 \\
0 & 0 & 3 & 0 & -3 \\
6 & 0 & -2 & 0 & -4 \\
4 & 0 & 0 & 0 & -4
\end{bmatrix}
\tag{62}
$$

In the matrix above, row 1 corresponds to a measurement from bus 1 to

bus 3, row 3 to a measurement from bus 3 to bus 5, and row 5 to a measurement from bus 1 to bus 5, and row 4 is an injection at Bus 1. From examining row 4, it is clear that the lines incident on Bus 1 go to Bus 3 and Bus 5. Both those flows are present in this sub-matrix, as Rows 1 and 5 respectively. By adding Row 1 and Row 5, and subtracting the result from Row 4, a null row can be created, which implies that the matrix is singular.

(3)   The loop test. If a loop is formed by a subset of the line flow measurements, the sub-matrix is not observable. In the example of (58), Rows 1, 3 and 5 form the loop 1-3, 3-1, 5-1; and by using row manipulation, a null row can be created in the submatrix, implying singularity. In this case, if (-2) ∗ Row 1 is added to Row 5, the result is that Row 5 becomes [0 0 4 0 -4], and if -0.75 ∗ Row 5 is now added to Row 3, Row 3 becomes [0 0 0 0 0]. Hence a sub-matrix containing a loop of line flows is singular. In practice, this is a relatively difficult condition to detect quickly, but if implemented, it can help reduce the number of singular matrices processed.

In the implementation of the LMS used, tests 1 and 2 were used to reject sub-matrices before the triangular decomposition phase.

5.6 : Computational Issues II - Online Methods

The description of the off-line portion of the algorithm is now complete. It can be summarized as comprising of the following steps :

(1) Compute the base case Jacobian matrix for the system.

(2) Find the fundamental sets for all the state variables in the system.

(3) Order the fundamental sets by increasing size.

(4) Form a sample using the method of Section 5.4.1.

(5) Check whether the sample passes the tests of Section 5.4.2 ; if not. go back to Step 4.

(6) Triangularize the matrix, if this fails go back to Step 4.

(7) Store the factored matrix, and the corresponding sample, and increment the count of completed samples.

(8) If the necessary number of samples have been created, then exit. Otherwise go to Step 4. The number of samples used is 2k where k is given by equation (49) of Chapter 4.

The remainder of this section deals with the implementation of the on-line portion of the estimator.

The sole connection between the off-line and on-line portions of the estimator is the set of factored matrices stored by the off-line program, that is used as input to the on-line part. These may be close to a thousand such files, depending on the size of the system, so efficient handling of the input is the first important issue in minimizing the on-line computation time. Two changes in the way the files are handled made a significant improvement in the performance of the system. First, the files are opened by the on-line program in read-only mode, which reduces the time to open the files. Secondly, the files are written and read using unformatted FORTRAN I/O ; this lowers the actual time to read the files. For smaller systems (14 and 30 bus), where the input/output overhead was relatively significant compared to the time required for the matrix multiplication, these measures reduced the computation time by upto 30 percent. For larger systems (118 bus) the

reduction was about 10-15 percent.

Once the data has been read, the next step is to compute the estimate itself. The procedure is based on a Newton-Raphson process ; first the residuals are computed from the equation $\mathbf{r} = \mathbf{z} - h(\mathbf{x})$, where the initial value of $\mathbf{x}$ is the flat start. Then the correction to the state is found using the sparse inverse generated previously, $\Delta x = \mathbf{H_i^{-1} r}$, and the updated value is found as $\mathbf{x_j} = \mathbf{x_{j-1}} + \alpha \Delta \mathbf{x}$; i is a counter of the sample number, and j is the iteration count. α is a numerical relaxation parameter, which is set to 0.9 for smaller systems, which have a low X / R ratio. For larger systems (57 and 118 bus) α is 1.0. This is necessitated by the fact that the decoupled form of the state estimator is being used; these methods are known to have numerical problems in converging when the line resistance is significant. The iterative procedure continues until the value of Δx is sufficiently small, with an updated value for $\mathbf{r}$ being calculated and so on. Once the converged estimate is available, the objective function is computed from the residuals. Note that if gross errors are present, the iteration scheme may not converge. This is because when there are an equal number of measurements and unknowns, the estimation is very similar to a load flow, and the bad data corresponds to a solution very "far" from the flat start. Load flow methods fail to converge under such conditions, which correspond to highly loaded systems. In the load flow case, the updating of the Jacobian often helps to alleviate this problem; this option is not available for the LMS state estimation. In all the simulations performed, a case was never encountered where the state estimation diverged, and the sample was good. Therefore if the estimator fails to converge after 20 iterations, the sample is considered to be corrupt.

This does not occur in all cases of bad data, only when the errors are exceptionally large.

Once the value of the LMS criterion has been obtained for all the samples for which the procedure converged, the sample with the lowest value of the LMS is selected. This sample is used to identify the bad data, by examining the standardized residuals and comparing them to the threshold. In a practical state estimator, where statistical efficiency is a concern, the LMS estimate would be followed by another estimate, with the outliers removed from the measurement set. Possible estimators for this purpose are the WLS and an M-estimator.

5.7 : Results from the LMS estimator

In this section, detailed test results from the IEEE 14 bus system will be given, along with a summary of the results from the IEEE 30 and 118 bus systems. These results are all for the real power estimation problem, so the number of unknowns is 13, 29 and 117 , while the number of measurements is 34, 56 and 247 for the 14, 30 and 118 bus systems respectively. Table 14 summarizes the test systems, with $P = 0.95$, and f being the maximum number of bad data which the estimator is supposed to be able to handle.

5.7.1 : IEEE 14-bus system

The one-line diagram and measurements for this system are shown in Figure 8. The following test cases will be studied :

(1) Multiple Interacting Bad Data.

(2) Multiple Bad Interacting Leverage Points.

(3) Multiple Bad Data at a Bus with Low Redundancy.

Table 14 : Number of Samples for the LMS estimator.

System	n	m	Max. # of bad data,f	$\varepsilon = f / m$	k	2*k
14 bus	13	29	2	0.059	5	10
30 bus	29	56	3	0.054	14	28
118 bus	117	247	5	0.020	31	62

Table 15 : Computing Times for the LMS estimator.

Test System	Time for Data Input	CPU Time (per Sample)	CPU Time (2k Samples)
14 bus	0.60 sec	0.08 sec	1.6 sec
30 bus	1.96 sec	0.18 sec	7.0 sec
118 bus	5.12 sec	1.41 sec	92.4 sec

5.8 : Analysis of the Results

Tables 16 and 17 show the 34 measurements for cases1 and 2 for the real portion of the state estimation problem. The measurements in boldface are the bad data added. The simulation results in these tables show that the LMS estimator is able to handle types of bad data that other estimators which are popular, such as the WLS and LAV estimators cannot handle.

Table 16: LMS Results for 14 Bus System Case 1

Measurement	Measured Value	True Value	|Residuals|	|Residuals| standardized
Flow 1-2	43.9	46.4	0.039	0.015
Flow 3-2	-35.6	-36.6	0.533	0.210
Flow 5-1	**449.8**	**-23.2**	**471.4**	**185.6**
Flow 5-2	**589.1**	**-13.34**	**600.9**	**236.6**
Flow 5-4	**2513.9**	**48.75**	**2461.8**	**969.2**
Flow 5-6	**396.8**	**-19.86**	**417.3**	**164.3**
Flow 7-8	-10.77	-1.00	0.00	0.00
Flow 8-7	8.82	-10.00	0.00	0.00
Flow 9-4	-4.25	-4.08	0.016	0.061
Flow 9-7	-13.36	-13.75	1.04	0.410
Flow 9-10	-11.26	-10.75	0.021	0.008
Flow 10-9	13.29	10.76	0.012	0.005
Flow 6-11	22.40	23.39	0.037	0.015
Flow 13-6	-25.22	-25.56	0.033	0.013
Flow 14-9	1.32	0.92	0.00	0.00
Flow 10-11	-19.81	-19.76	0.027	0.011
Flow 13-12	-3.27	-3.92	0.0471	0.019
Injection 1	69.97	69.55	0.041	0.016
Injection 2	29.36	28.30	0.083	0.033
Injection 4	-7.41	-7.80	41.83	16.47
Injection 7	-0.33	0.00	1.04	0.408
Flow 2-1	-45.51	-46.37	0.00	0.00
Flow 2-4	25.57	-24.65	0.00	0.00
Flow 1-5	24.24	23.17	0.00	0.00
Flow 4-3	18.32	17.65	0.00	0.00
Flow 4-7	2.11	3.75	0.00	0.00
Flow 13-14	16.09	15.99	0.00	0.00
Injection 6	77.9	78.80	0.00	0.00
Injection 8	9.38	10.0	0.00	0.00
Injection 10	-9.79	-9.0	0.00	0.00
Injection 11	-2.93	-3.5	0.00	0.00
Injection 12	-6.63	-6.1	0.00	0.00
Injection 13	-11.92	-13.5	0.00	0.00
Injection 14	-15.39	-14.9	0.00	0.00

Table 17 : LMS Results for 14 Bus System Case 2

Measurement	Measured Value	True Value	Residuals	Residuals standardized
Flow 2-1	-46.8	46.4	-0.34	-0.85
Flow 3-2	-36.4	-36.6	-0.23	-0.58
Flow 2-4	24.3	24.6	-0.02	-0.05
Flow 1-5	23.3	23.1	-0.23	-0.58
Flow 5-2	-13.1	-13.3	-0.05	-0.1
Flow 5-4	**150.0**	**48.7**	**102.3**	**255.7**
Flow 5-6	-19.8	-19.9	0.21	0.53
Flow 4-7	3.8	3.7	0.32	0.80
Flow 8-7	10.2	10.0	-0.30	-0.75
Flow 9-7	-13.4	-13.7	0.98	2.45
Flow 9-10	-10.9	-10.7	-0.59	1.48
Flow 6-11	23.5	23.3	-0.03	0.08
Flow 13-6	-25.8	-25.6	0.24	0.60
Flow 10-11	-19.9	-19.8	0.03	0.08
Flow 13-14	15.6	16.0	-0.05	0.13
Injection 1	69.4	69.5	-0.64	1.58
Injection 4	**-150.0**	**-47.8**	**-103.1**	**-257.7**
Injection 8	10.3	10.0	-0.20	0.50
Injection 10	-9.1	-9.0	0.43	1.08
Injection 12	-6.5	-6.1	0.92	2.30
Injection 14	-14.7	-14.9	0.61	1.53
Flow 1-2	46.5	46.4	0.00	0.00
Flow 5-1	-23.4	-23.1	0.00	0.00
Flow 4-3	17.3	17.6	0.00	0.00
Flow 7-8	-10.5	-10.0	0.00	0.00
Flow 9-4	-4.2	-4.1	0.00	0.00
Flow 10-9	10.4	10.8	0.00	0.00
Flow 14-9	0.6	0.9	0.00	0.00
Flow 13-12	-3.2	-3.9	0.00	0.00
Injection 2	28.0	28.3	0.00	0.00
Injection 6	80.2	78.8	0.00	0.00
Injection 7	0.4	0.0	0.00	0.00
Injection 11	-3.6	-3.5	0.00	0.00
Injection 13	-13.1	-13.5	0.00	0.00

Table 18 : LAV and LS Residuals for Cases 1 and 2

Meas. #	LAV Residuals Case 1	LS Residuals Case 1	LAV Residuals Case 2	LS Residuals Case 2
1	0.0	-80.840	-4.6414	-4.8056
2	0.01100	80.8518	4.6524	4.81660
3	0.85457	373.026	0.0	7.8539
4	0.21781	52.1432	-11.7749	-9.7961
5	0.80026	342.499	4.64140	7.97035
6	**472.212**	**130.51**	-4.62240	-7.9513
7	**602.039**	**136.96**	-7.0022	-11.3213
8	0.0	-484.313	13.3590	2.22770
9	**2464.08**	**768.850**	**24.3829**	**14.8794**
10	**399.350**	**-110.048**	-16.4656	-13.9811
11	0.0	-377.154	0.08997	6.03223
12	-0.00000	-58.3457	0.000	-0.0200
13	0.0	58.3457	0.0	0.0200
14	0.051619	170.273	0.0	-3.4249
15	0.0	143.772	-0.08997	-6.1122
16	0.00200	81.8209	0.002	5.8181
17	0.0	-81.8189	0.0	-5.8160
18	0.0	-163.536	-16.6257	-7.7132
19	0.0	123.654	0.0	0.75616
20	6.9937	-15.2254	1.64501	3.45301
21	0.00084	152.342	16.6573	10.0939
22	-0.09749	-19.6988	-0.09749	-1.33316
23	7.15982	-36.788	1.81120	3.1386
24	0.8002	261.658	0.0	3.1647
25	0.0	225.674	0.0	-1.39183
26	0.0	611.625	**0.0**	**5.7606**
27	17.3310	161.544	0.0	4.3206
28	0.0000	175.037	0.0	0.0600
29	0.00000	58.3457	-0.0000	0.0200
30	0.0	70.5142	16.6563	4.2770
31	0.03069	11.2201	0.0	-2.3488
32	0.0	97.5907	0.0	2.58707
33	7.0623	67.1672	1.71371	2.56163
34	0.0	21.7370	0.0	0.48037

Table 19 : LMS Results for 14 Bus System Case 3

Measurement	Measured Value	True Value	Residuals	Residuals standardized
Injection 7	40.0	0.0	0.00	0.00
Injection 8	50.0	15.1	0.65	0.88

This is shown in Table 18, which shows the residuals from S-PLUS simulations for cases 1 and 2. The LAV solution for Case 1 is one of 2 solutions that was found; the residuals from the good solution are shown. The measurements are not in the same sequence as in Tables 16 and 17 ; the bad data are in boldface, so that the corresponding measurements can be identified. In the results for Case 1, the LS is seen to fail, while the LAV has several solutions. In Case 2, both the estimators fail, the LS because it has no resistance to multiple interacting bad data, and the LAV due to the multiple bad leverage points. In fact, the LAV solution passes through the bad leverage point, #26. The computation times given in Table 15 were measured on a VAX 8800 computer (for reference, this computer is about 20 % slower than an Intel 80486 based machine operating at 33 Mhz). The computing times shown do not make the use of the LMS immediately viable, however they are not unreasonable, given the benefits that they provide.

The most interesting case is that with the bad data clustered in a region of low local redundancy. From Figure 8, it can be seen that Bus 8 forms an antenna of the system, a bus that is connected to only one other bus, Bus 7 in

this system. Therefore at most 4 measurements can belong to the fundamental set of such a bus; the two flow measurements on the line, and injections at the two busses. If two of these four measurements are bad and conforming, then there will be an ambiguity in the estimation of θ_8 ; if three of the four are bad, then the estimate of θ_8 will be bad. This is independent of the global breakdown point, and the number of measurements at other locations etc. Note that this is also caused by the sparsity of the system, since all other measurements except the four mentioned do not contribute in any way to the estimation of θ_8. Therefore it is possible, due to the small fundamental set size, to cause the so-called local breakdown of the estimate. In Case 3, 2 of the 4 measurements connected to Bus 8 are bad, while 2 of the 8 measurements that influence the estimation of θ_7 are bad. No other measurements are bad, so the estimates of all other angles are good. Though some measurements connected to Bus 7 are bad, the majority are good, so θ_7 is also good. The effect of the 2 bad data is then to corrupt θ_8 alone, which is why this type of failure is called local breakdown, as shown in Table 19. Hampel [64] had suggested that this kind of situation could arise, but to our knowledge, this is the first demonstration of the failure of a High Breakdown Point Estimator below its global breakdown point. This mode of failure is not unique to power systems, but can be encountered in any system where the design matrix is sparse. In Chapter 7, the implications of the local breakdown for other aspects of power system state estimation will be examined.

5.9 : Computational Issues III - Parallel Processing

Rousseeuw[46] pointed out that the nature of the LMS algorithm used in PROGRESS was such that the bulk of the computations required were parallel, rather than serial. In effect, if the program were implemented well, by running it on an n-processor machine, the time required to complete the estimation would be close to n times less than the time it would take on a single processor. If an algorithm were completely parallelizable, the speedup would be in direct proportion to the number of processors ; on the other hand, if an algorithm was totally serial, there would be no improvement gained by adding any extra processors. In practice, no algorithm is either totally parallel, or totally serial; usually they contain sections of each type. This is true of the PROGRESS algorithm as well ; the computation of the estimate from any particular sample can be done in parallel with the estimate from another sample. There is no interaction between the two steps ; they do not contend for the same data, nor do they need the results of the other to proceed. On the other hand the determination of which sample has the lowest objective function value is serial. It requires the output estimates of all the samples, so if no amount of parallelism will be useful if one of the samples has not yet been processed.

The LMS algorithm described in this chapter has been modified to use multi-processors if they are available, and the resulting implementation was tested on a two processor machine. Table 20 shows the results of these tests, which were conducted using the 30 bus system. The total CPU time required increased slightly ; this is due to the extra overhead and communication between the processors. For example, all processors share the same console,

so if two processors attempt to write to the screen simultaneously, an access error will occur. In order to prevent such events, a system of locks and released for shared resources had to be used. Such mechanisms, which re not needed in the single processor implementation, contribute to the increase in total CPU time used. The elapsed time, which is the parameter of interest, decreased by 38 % ; if the method were totally parallelizable, a gain of 50 % would have been attained. This shows that it is feasible to use multi-processor machines for LMS state estimation, and that unlike the LS , the algorithm exploits the availability of parallel processors if present.

Table 20 : Computation Times with Dual Processors

	CPU Time (Total)	Elapsed Time
Single Processor	18 seconds	50.7 seconds
Dual Processor	24 seconds	31.3 seconds

5.10 : Summary

At the end of Chapter 4, the PROGRESS and nonlinear search methods had been reviewed, and it had been shown that the LMS possessed many desirable features in terms of its robustness, and almost every undesirable feature possible regarding its implementation. In this chapter, it has been shown that the LMS estimator is not just a theoretical curiosity, relevant only to small systems, but is a method that can be applied with good results to power system state estimation, on realistically sized systems. In order to

achieve this, various computational devices had to be employed, and those are the focus of this chapter. Some of these are techniques that are common in other areas of power systems, and had to be modified in order to be used in this application; sparse matrix methods and rapid observability analysis, for example. Some of these are new tools devised for this purpose alone, such as the fundamental set, and the division of the effort into an off-line and on-line part. The successful implementation of the LMS required a fusion of all these ideas, and that is what this chapter describes.

The test cases described show the LMS at its best, when handling the types of bad data that cause other estimation methods to fail ; multiple interacting (and in fact, conforming) bad data for the WLS with normalized residuals, and multiple bad leverage points for the LAV and the M-estimators. This chapter also reveals hitherto unsuspected weaknesses of the LMS (local breakdown) ; however it is shown that these are not a failure of the estimator used, but a consequence of the sparsity of the system.

The next chapter further develops on the foundations established in this chapter. The concept of a fundamental set will be used in the identification of leverage points, in the development of robust meter placement tools, and in other methods that seek to improve on the results described in this chapter, such as the decomposition methods for application of the LMS to much larger systems. The local breakdown and the issue of local redundancy also have a bearing on these topics, especially the robust meter placement.

Figure 8 : IEEE 14 Bus System and Measurements
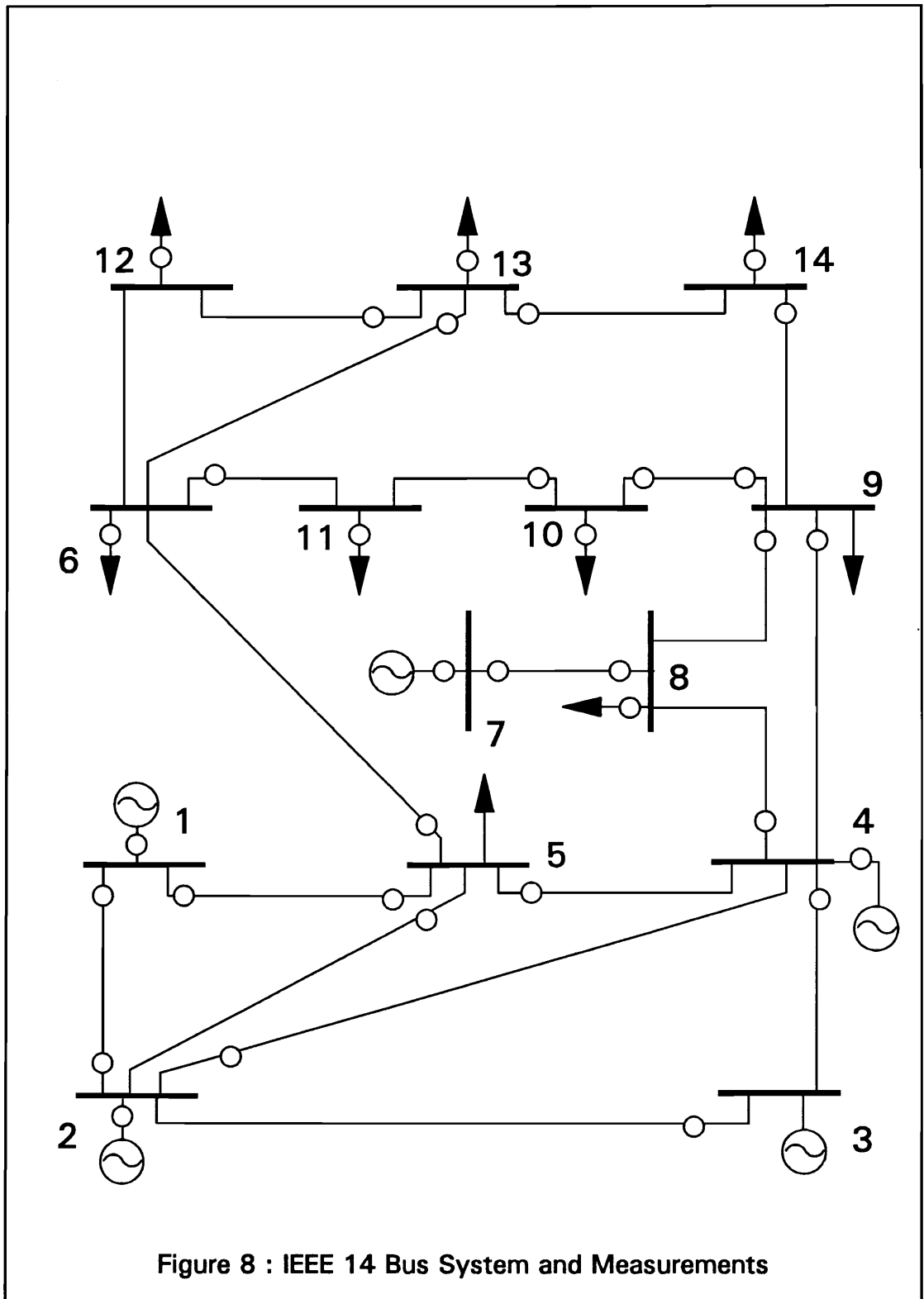
Chapter 6 : Leverage Points in Power Systems

6.1 : Introduction

There are several reasons why leverage points are important in power system state estimation. It has been shown that the Jacobian matrix used in state estimation contains leverage points, and it has also been shown that an important class of robust estimators, the M-estimators, cannot handle bad leverage points. Since the use of M-estimators in general, and the LAV estimator in particular is becoming increasingly popular [65,66] in state estimation, a thorough investigation of the effects of leverage points in power systems is warranted. Though bad leverage points are harmful to many estimators, good leverage points are very useful in reducing the variance of the estimate. Therefore it might be fruitful to place the measurements such that there are a large number of leverage points, especially if an estimator that is unaffected by the presence of bad leverage points is being used, such as the LMS. Since the LMS has a relatively poor statistical efficiency, the addition of leverage points to the measurement set could be used to provide a much-needed improvement in that regard. Finally, the Generalized M-estimators seek to handle leverage points by incorporating the influence of position in the factor space into the objective function. The conventional method used to measure the influence of position is the Mahalanobis

distance, a measure that is non-robust, and is subject to the masking effect in the presence of multiple leverage points. In this chapter, methods that produce a robust estimate of the position in the factor space for large, sparse systems will be developed, and these have potential applications in devising GM estimators that would use this estimate in place of the Mahalanobis distance.

6.2 : The Mahalanobis distance

In the regression model $\mathbf{z} = \mathbf{Hx} + \varepsilon$, let l_i be the transpose of the ith row of $\mathbf{H}$. Then the squared Mahalanobis distance, D_i^2 is given by

$$D_i^2 = (l_i - \underline{l})^t \Gamma^{-1} (l_i - \underline{l}) \qquad (63)$$

where $\underline{l}$ is the average of all the vectors l_i, and

$$\Gamma = \frac{1}{m-1} \sum_{i=1}^{m} (l_i - \underline{l})(l_i - \underline{l})^t \qquad (64)$$

It can be seen that there is a strong connection between the Mahalanobis distance and the Least Squares estimator. For instance, the estimate of the location of the l_i vectors is taken as their mean, and Γ is computed as a function of the sum of the squares of the distances of l_i from its mean. In fact in the case of multiple regression with an intercept term, there is a linear relationship between the diagonal elements of the hat matrix, $\mathbf{S} = \mathbf{I} - \mathbf{W}$, and the Mahalanobis distance [67]

$$s_{ii} = \frac{D_i^2}{n-1} + \frac{1}{n} \tag{65}$$

If only one leverage point is present in the factor space of H it can be identified as that measurement which has the largest Mahalanobis distance. If however multiple leverage points are present, they can serve to mask each others presence, just as in the largest normalized residual test. Therefore the Mahalanobis distance is not suitable for identifying leverage points in power system state estimation, where there is a very strong likelihood of multiple leverage points being present. The remainder of this chapter focuses on robust estimates of the Mahalanobis distance, and the applications thereof.

6.3 : The Minimum Volume Ellipsoid (MVE)

In [67], Rousseeuw and Van Zomeren develop a strong case against using the Mahalanobis distance for identifying leverage points in multiple regression. They then present an alternate method, the minimum volume ellipsoid, which is a high breakdown point estimator of multivariate location and covariance, and is an extension of the work of Donoho [68].

The minimum volume ellipsoid estimate for the location and scatter of a set of data is based on minimizing the determinant of the matrix C, which is a robust estimate of the covariance matrix. In higher dimensions, the volume of an ellipsoid is linearly related to this determinant, and hence the name of the algorithm. Geometrically, the minimum volume ellipsoid is that ellipsoid with minimum volume that contains within it half the data points. As is to be expected, the computation of the minimum volume ellipsoid is not simple.

Rousseeuw describes an algorithm which is an extension of the PROGRESS algorithm, i.e it is based on re-sampling.

If the data set consists of n points in p-dimensional space, an ellipsoid that contains (p+1) points is determined as the standard covariance matrix. The location estimate T is given by

$$T = \frac{1}{p+1} \sum_{i=1}^{p+1} x_i \tag{66}$$

and the covariance matrix is

$$C = \frac{1}{p} \sum_{i=1}^{n} (x_i - T)^t (x_i - T) \tag{67}$$

The ellipsoid is then inflated so that it contains h = [(n+p+1)/2] points, and the objective function, a multiple of the determinant of C is computed. This process is repeated for many sets of (p+1) points, which the number of samples to be drawn is based on the same argument that was used in PROGRESS. That is, enough samples must be drawn to guarantee, with a probability P, that at least one good subsample of (p+1) points is drawn from among the n points.

This algorithm was found to have poor numerical performance for p > 9, and therefore was not used in our simulations. Another algorithm described in [67] was modified for in power systems.

6.3.1 : Projection Pursuit Algorithm for the MVE
This algorithm is based on the work of Donoho [68]. For each point x_i, the

quantity u_i is computed, where

$$u_i = \max_v \frac{|x_i v^t - L(x_1 v^t, x_2 v^t, \ldots, x_n v^t)|}{S(x_1 v^t, x_2 v^t, \ldots, x_n v^t)} \tag{68}$$

and v is a set of vectors. At the minimum the vectors that comprise v are the vectors between each x_i and the center of the point cloud. Rousseeuw suggests using the co-ordinate-wise median of the points, M, as the center, and then $v = x_i - M$. The functions $L(x_1 v^t, x_2 v^t, \ldots, x_n v^t)$ and $S(x_1 v^t, x_2 v^t, \ldots, x_n v^t)$ are the one dimensional location and scale estimates for the set of projections of each x_i on the vector v. A robust estimate for both these quantities should be taken, so for example, the median and the MAD, or the LMS and the length of the shortest half could be used.

The algorithm then consists of successively taking each x_i, finding the value of the parameter for each member of the set of projections v for that x_i, and finding the maximum value of the parameter over that set. The u_i is a robust estimate of the Mahalanobis distance for x_i. The estimate can be improved at the cost of increased computation time by considering an enlarged set v.

6.3.2 : Implementation of the Algorithm

The projection pursuit algorithm, like many of the algorithms discussed previously, cannot be directly applied to sparse systems. There are two reasons for this:

(1) The co-ordinate wise median is not a reliable estimate of the center of the point cloud, when applied to sparse systems. Since the bulk of the points have a zero projection along any particular axis, the co-ordinate

wise median essentially reduces to taking the origin as the center in all cases.

(2) Since the center effectively is the origin, the projection $v = x_i - M$ becomes the vector x_i. Again, due to sparsity, the majority of the dot products x.v are zero. This makes the location and scale estimates in (68) both zero, and hence the bulk of the u_i become infinite.

The following modifications were used in order to compute the projection pursuit estimates for the Mahalanobis distance :

(1) Instead of computing the co-ordinate wise median over all points, the center along any particular axis is the median of those points which have a non-zero projection in that direction. In terms of the H matrix, this implies that instead of computing the median of all points in a column, only the non-zero terms are used.

(2) A similar scheme could be used for the location and the scale, where only non-zero terms are considered. Instead, a method based on the fundamental sets of the system was used. In this implementation, the set of projections used for a measurement x_i consists of the projections derived from those measurements that are members of the same fundamental set as x_i. In this way, not all zero projections are automatically discarded.

In this manner, the u_i are computed as robust estimates of the Mahalanobis distance. In Rousseeuw's original formulation, where the data is assumed to have a definite structure, outliers in the factor space are detected based on a chi-square test. Since there is not valid basis in our case for assuming that the data has such a structure, the leverage points correspond to the outlying

values of the u_i.

6.4 : Simulation Results

The algorithm outlined was applied to the IEEE 14, 30 and 118 bus systems. In Table 21, three estimates for the Mahalanobis distance of the IEEE 14 bus system are given. The first column gives the values obtained from equation (64), the conventional least-squares based definition of the Mahalanobis distance. It can be seen that there are no gross outliers in this estimate. The second and third columns contain estimates based on the projection pursuit method, using as location estimates the LMS and the median respectively. Using the identification methods described above, the values in boldface were identified as leverage points. The same results are shown in Figure 9, and this shows that the simulation results of Chapter 5 with the LAV were to be expected, since measurement 26 is a leverage point.

To test this, repeated simulations were performed with S-PLUS, with the objective of identifying leverage points. The hypothesis was that a single bad leverage point would cause breakdown of the LAV estimator. Each of the measurements was corrupted systematically, and a LAV estimate was performed. This was repeated for 500 different values of the bad data for each measurement ; if the LAV broke down, the corresponding measurement was identified as a leverage point. In the case of the 14 bus system, all 5 measurements that were singled out by both projection pursuit methods as leverage points were also identified as such by the simulations using the LAV. No other leverage points were identified.

Table 21 : Robust and Standard Mahalanobis Distances

Meas. #	Mahalanobis Distance	Robust Estimate with the LMS and the Shorth	Robust Estimate with the Median and the MAD
1	2.694675	1.509981	1.125303
2	2.859162	2.427989	1.493228
3	3.433517	1.024723	1.159266
4	1.717582	1.298616	1.167924
5	1.776187	1.035306	0.092447
6	1.824778	1.035306	0.072424
7	2.054524	0.059063	0.083235
8	3.787278	1.537044	1.290677
9	4.424230	6.149024	6.073198
10	3.558899	1.642998	2.060165
11	3.231435	1.295841	0.084731
12	3.137837	3.823320	2.289832
13	3.171221	3.177051	1.804307
14	1.667462	0.05000	0.061857
15	4.157850	2.18061	2.106199
16	3.518257	2.39986	3.019548
17	3.205394	2.453454	2.739188
18	3.646947	**12.80709**	**16.84717**
19	2.808445	5.513679	6.241782
20	2.907760	0.05842	0.06740
21	3.204530	1.714092	0.081828
22	2.473397	1.123895	0.06739
23	2.270237	2.132259	2.93067
24	3.805232	3.040953	4.099148
25	4.863077	**7.166780**	**8.878441**
26	4.991078	**10.59408**	**9.473007**
27	4.883853	**72.02291**	**89.90276**
28	4.763819	3.649050	4.245772
29	3.171221	2.097929	1.829083
30	4.033298	6.534436	5.192388
31	4.482615	**19.98782**	**21.37044**
32	4.644884	2.840712	2.944829
33	4.453237	2.487358	2.898319
34	4.458154	1.914088	1.532806

The simulations using the 30 and 118 bus systems produced similar results, with the measurements identified by the robust projection methods also being confirmed by the LAV simulations as such. These results shows that typically about 10-15 percent of the measurements are leverage points. For the larger systems, it was found that the LMS based projection methods were more conservative than the median method.

6.5 : Robust Meter Placement

As already shown, leverage point identification can be used in conjunction with the LAV, since the LAV has no resistance against bad leverage points. In this section, the use of this technique as part of a robust meter placement scheme is described, along with other methods for robust meter placement.

The basic principle that is applied in this section is that the size of all the fundamental sets is to be increased ; as a secondary objective, as many leverage points as possible should be included in the measurement set, so as to maximize the efficiency of the estimator. This is to some extent a recursive problem, since a measurement does not become a leverage point in isolation ; to categorize it as such requires a knowledge of the location of the bulk of the measurements. Therefore to solve this problem exactly, the effect of adding a new measurement on existing leverage points has to be considered. In this preliminary work, this is not considered, and the values of the robust Mahalanobis distance for the maximal measurement set are used.

Since an injection measurement contributes more entries to the fundamental sets of a system than a flow measurement (at worst, their contributions are equal), the first step is to assume all injections are metered. This contributes

to both the robustness of the system and the leverage points, since many injections are also leverage points. The problem is then to add the desired number of flow measurements so as to maximize the local breakdown of the system, while also trying to incorporate leverage points. The following algorithm was used.

Since the worst case local breakdown is determined by the smallest fundamental set, the immediate goal is to increase the size of the smallest fundamental set if possible. It may not be possible, if all the possible measurements have been metered. In that case, the next smallest set is considered. At this point, one measurement will be selected from among those possible, and added to the measurement set. Two alternate criteria were used for this, one which maximizes robustness, and the other that considers leverage.

In the robust algorithm, the size of the fundamental sets of all variables adjacent to the variable with the smallest fundamental set is evaluated. A measurement is placed so as to increase the size of the smallest "neighbor" fundamental set. Once this measurement has been placed, the size of all the fundamental sets is updated, and the next measurement is placed according to the principles just described. This process continues until the specified number of flows have been chosen.

In the leverage algorithm, the robust Mahalanobis value of each potential measurement can be considered, and the potential measurement that increases the smallest fundamental set and has the largest leverage value is chosen. This scheme is not competitive with the robust algorithm computationally, since it requires the evaluation of the robust Mahalanobis

Table 22 : Robust Meter Placement (IEEE 30 bus system)

Node #	Max. F_i	Actual F_i	$[(F_i-1)/2]$
1	7	6	2
2	13	6	2
3	7	5	2
4	13	5	2
5	7	5	2
6	22	8	3
7	7	5	2
8	7	5	2
9	10	6	2
10	19	7	3
11	4	4	1
12	16	8	3
13	4	4	1
14	7	5	2
15	13	7	3
16	7	5	2
17	7	5	2
18	7	5	2
19	7	7	3
20	7	5	2
21	7	5	2
22	10	6	2
23	7	5	2
24	10	6	2
25	10	6	2
26	4	4	1
27	13	5	2
28	10	6	2
29	7	5	2
30	7	5	2

distance. Table 22 shows the results from application of the robust meter placement algorithm to the IEEE 30 bus system. The maximum size of each fundamental set is given, along with the actual realized size. The goal of the algorithm was to produce fundamental set sizes that are approximately equal, without any excessively large ones. It can be seen that the small fundamental sets, those corresponding to variables 11, 13 and 26 have all been filled to the maximum ; while the largest two, those corresponding to variables 6 and 10 have been kept at about 35 percent of their maximum size.
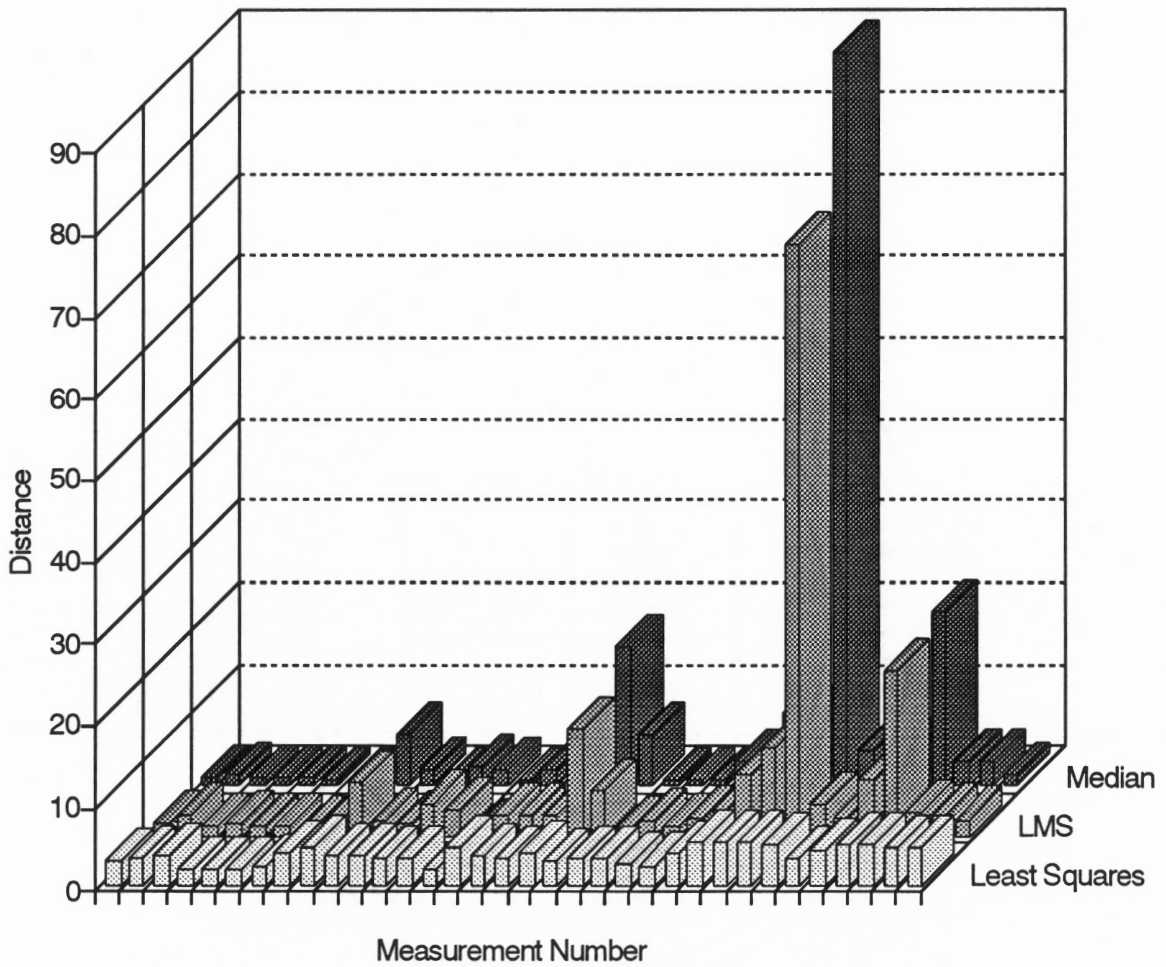
Figure 9 : Leverage Estimates for the IEEE 14 Bus System

Chapter 7 : Conclusions

7.1 : Contributions of this Dissertation

The objective of this dissertation was to demonstrate that the robust high breakdown point estimators that have recently been developed by statisticians can be used in power system state estimation. In order to accomplish this, the underlying principles behind robust estimation theory had to be demonstrated and their relevance established. It might be taken for granted by statisticians that everyone who uses the Least Squares Estimate is aware of the fact that it is not robust, and loses efficiency rapidly as the distribution of the errors becomes non-Gaussian. This was certainly not the case among power system engineers; the conventional wisdom was that if no outliers were detected by tests based on the LS estimate, then that "proved" that no outlier existed. The possibility that the presence of the outliers might affect the LS estimate in such a way that the presence of the outliers is disguised was not considered.

To remedy this situation, this dissertation, and the papers published from it always started by showing that the Least Squares is not a robust estimate. Until this became an accepted view, the need for more robust estimators would always be questioned. The first contribution of this dissertation is to take the work of many distinguished statisticians, and extend it to a procedure usable by the power system state estimation community. Existing

methods for multiple linear regression were modified to incorporate sparse matrix techniques, and to handle the nonlinear multiple regression case. The introduction of such concepts as the breakdown point and the influence function, changes (hopefully forever) the criteria by which future state estimators are evaluated. From the response to these publications, it has been a struggle, but the objective has been accomplished, in our judgement. State estimation researchers are now aware of the concepts of robust statistics, and indeed, other researchers have written papers using these concepts to extend their previous work. For example, the recent work of Abur in attempting to eliminate leverage points in the LAV by rotating the design matrix [66] whereas in his first work on the LAV [65], he states "the LAV estimator may fail to reject bad data occasionally". Re-analyzing his old simulations, he has concluded that the hitherto unexplainable failures of the LAV were actually due to bad leverage points. We do not believe that the LAV can be made sufficiently robust to be used in large systems, despite its advantages computationally, but we are gratified that based on our introduction of the concepts of leverage points to this community [40,62,69], he and other researchers [19] are re-assessing their own work.

The next contribution of this dissertation is to systematically explore the practical difficulties in implementing the LMS and other high breakdown point estimators in large sparse systems. The original PROGRESS program is limited to handling 10 unknown variables, and does not work properly in the case of sparse design matrices. In this dissertation, the concepts employed in PROGRESS have been extended to much larger systems than was originally possible, and the speed of the computation has been greatly

enhanced. In order to achieve this, simple observability tests that are extremely effective in reducing the number of unobservable cases tested have been devised and implemented. A modified version of Tinney's optimal ordering scheme for ill-conditioned, non-diagonally dominant matrices was used to further reduce the fill-in.

The concepts of fundamental sets and local breakdown that arose from this work are finding applications in areas indirectly connected to state estimation. Some of these applications will be described in the section on future research. These concepts are also making it possible to understand the mechanism by which state estimators fail, rather than leaving it as mystery. This is important in the design of the metering points, so that areas of low local redundancy are not tolerated just because of a satisfactory global redundancy.

The implementation of the algorithm to efficiently utilize multi-processor computers could become significant, as power system control centers move away from the older computers that are the mainstay currently. In fact, in 1990, the Hawaiian Electric Company put into service a control center using a network of multi-processor VAX workstations, rather than the conventional mini-computer. This is believed to be the first such installation.

The work on identification of leverage points in large sparse systems has no direct bearing on the LMS, since it is immune to the effect of bad leverage points. This work does impact the research on the LAV being performed by other researchers, in that they may use this technique to identify potential trouble spots for the LAV, and in the extreme case, remove all leverage points altogether before the estimation. The robust estimate for the Mahalanobis

distance can be used to make a more robust class of GM estimators possible, though it was also made demonstrated in this dissertation that the breakdown point of these estimators is inversely proportional to the size of the system, and hence for realistically sized power systems, the breakdown point is essentially zero. The robust Mahalanobis distance will also be applied in the normalization of the residuals from the LMS; as described in the dissertation, the heuristic scale estimate used in PROGRESS was not suitable for state estimation.

7.2 : Future Research

Some of the objectives described in this section are already being investigated here ; others are expected to be dealt with in the near future.

(1) Application of a decomposition method to make the computation time of the LMS lower, for systems of several thousand busses. By direct application of the methods described in this dissertation, systems of upto about 150 busses can be handled effectively. To solve larger systems, the design matrix is decomposed or partitioned into several overlapping matrices, each of size less than 150 busses. The LMS estimate for each partition is computed separately, and then the global estimate is found by combining the estimates from all the regions. The problems faced include the selection of the partitions, without further reducing the already low redundancy, especially at the boundary busses ; the re-combination of the estimates, and the usual computational problems that always arise when dealing with systems of this size. Preliminary results from the application of this method are very encouraging, and it is being tested on a 2418 bus system.

(2) It has been shown that the global breakdown point is an upper bound on the number of bad data the state estimator can withstand. For sparse systems, the practical limit is the local breakdown point, which is determined at a local level by the positioning of the measurements. Till now, most meter placements have just evolved, rather than being systematically designed to minimize any problems. The analytical work in this area has focused on maximizing the efficiency at the Gaussian distribution. Using a robust meter placement, an existing set of measurements can be strengthened against local breakdown by maximizing the size of the fundamental sets of the system. This involves either choosing the location of new measurements, or redistributing existing measurements. Some work on this has already been performed using heuristic optimization methods as described in Chapter 6; this will be continued, and the application of simulated annealing to solve this combinatorial optimization problem is being investigated.

(3) As mentioned in the previous section, the potential applications of the high dimensional leverage point identification method are varied, ranging from removing them from the measurement set if the LAV is being used (not recommended) ; to devising new estimators that employ the robust Mahalanobis distance as a scale estimate ; and applications in areas other than state estimation, such as the stability problem mentioned, where the method is being used as a clustering technique.

(4) The S and τ estimators mentioned in Chapter 3 have not been investigated thoroughly in terms of their application to power systems. They appear to have some very interesting properties, especially the possibility than an analytical method rather than a combinatorial method could be

used, while maintaining high breakdown in multiple regression.

(5) The extremely complex problem of topological rather than measurement errors may benefit from robust estimators. When topological errors occur, they invariably create multiple bad data, and they are almost always gross errors, rather than a minor deviation from the assumed model. Therefore the possibility of using high breakdown point estimators here seems to be feasible, especially since the existing methods are already combinatorial in nature, so the computing time competition with the Least Squares that dominated this dissertation does not arise.

(6) An improved implementation of the LMS should be investigated. Methods based on simulated annealing and neural networks / expert systems are envisioned as providing alternate optimization techniques for this purpose.

(7) Though in the case of the LMS, the direct search methods tested were not found suitable for this application ; in a private communication, David Ruppert has indicated that these methods can be used fruitfully in the implementation of S estimators. This estimator has robustness properties similar to the LMS, while it has a greater statistical efficiency.

Bibliography

[1] T. E. Dy Liacco, "The Role and Implementation of State Estimation in an Energy Management System", *Electrical Power and Energy Systems*, Volume 12, No. 2, April 1990, pp. 75-79.

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

[3] M. M. Adibi and D. K. Thorne, "Remote Measurement Calibration, *IEEE Transactions on Power Systems*, Volume PWRS-1, No. 2, May 1986, pp. 194-203.

[4] F. C. Schweppe, J. Wildes and D. B. Rom, "Power System Static State Estimation, Part I, II and III", *IEEE Transactions on Power Apparatus and Systems*, Volume PAS-89, No. 1, January 1970, pp. 120-135.

[5] W. F. Tinney and J. W. Walker, "Direct Solutions of Sparse Network Equations by Optimally Ordered Triangular Factorization", *Proceedings of the IEEE*, Volume 55, No. 11, November 1967, pp. 1801-1809.

[6] W. F. Tinney, W. L. Powell and J. W. Walker, "Programming of Sparsity-Directed Ordering Schemes", *Proceedings of the 5th PSCC Conference, 1975*.

[7] J. F. Dopazo, O. A. Klitin, G. W. Stagg and L. S. Van Slyck. "State calculation of power systems from line flow measurements, Part I", *IEEE*

Transactions on Power Apparatus and Systems, Volume PAS-89, No. 9, September 1970, pp. 1698-1708.

[8]   J. F. Dopazo, O. A. Klitin and L. S. Van Slyck. "State calculation of power systems from line flow measurements, Part II", *IEEE Transactions on Power Apparatus and Systems*, Volume PAS-91, No. 1, January 1972, pp. 145-151.

[9]   H. M. Merrill and F. C. Schweppe, "Bad data suppression in power system state estimation", *IEEE Transactions on Power Apparatus and System*, Volume PAS-90, November / December 1971, pp. 2718-2725.

[10] E. Handschin, F. C. Schweppe, J. Kohlas and A. Fiechter, "Bad data analysis for power system state estimation", *IEEE Transactions on Power Apparatus and Systems*, Volume PAS-94, No. 2, March/April 1975, pp. 329-337.

[11] W. W. Kotiuga and M. Vidyasagar, "Bad data rejection properties of weighted least absolute value techniques applied to static state estimation", *IEEE Transactions on Power Apparatus and Systems*, Volume PAS-101, No. 4, April 1982, pp. 844-853.

[12] D. M. Falcao and S. M. de Assis, "Linear Programming State Estimation: Error Analysis and Gross Error Identification", *IEEE Transactions on Power Systems*, Volume PWRS-3, No. 3, August 1988, pp. 809-815.

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

[14] K. A. Clements and P. A. Davis, "Multiple Bad Data Detectability and Identifiability: A Geometric Approach", *IEEE Transactions on Power Delivery*, Volume PWRD-1, No. 3, July 1986, pp. 355-360.

[15] L. Mili, T. 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*, Volume PAS-103, November 1984, pp. 3239-3252.

[16] L. Mili, T. Van Cutsem and M. Ribbens-Pavella, "Bad Data Identification Methods in Power System State Estimation - a Comparative Study", *IEEE Transactions on Power Apparatus and Systems*, Volume PAS-104, November 1985, pp. 3037-3049.

[17] L. Mili and T. Van Cutsem, "Implementation of the Hypothesis Testing Identification in Power System State Estimation", *IEEE Transactions on Power Systems*, Volume 3, No. 3, August 1988, pp. 887-893.

[18] F. C. Schweppe and E. Handschin, "Static State Estimation in Electric Power Systems", *Proceedings of the IEEE*, Volume 62, No. 7, July 1974, pp. 972-982.

[19] F. Wu, "Power system state estimation: a survey", *Electrical Power and Energy Systems*, Volume 12, No. 2, April 1990, pp. 80-87.

[20] R. J. Beckman and R. D. Cook, "Outlier·...s", *Technometrics*, Volume 25, No. 2, May 1983, pp. 119-163.

[21] J. W. Tukey, "A Survey of Sampling from Contaminated Distributions", in *Contributions to Probability and Statistics*, I. Olkin (ed.), Stanford University Press, 1960, pp. 448-485.

[22] E. S. Pearson and C. Chandra Sekar, "The Efficiency of Statistical Tools and a Criterion for the Rejection of Outlying Observations", *Biometrika*, Volume 28, 1936, pp. 308-320.

[23] G. E. P. Box, "Non-normality and Tests of Variance", *Biometrika*, Volume 40, 1953, pp. 318-335.

[24] F. J. Anscombe, "Rejection of Outliers", *Technometrics*, Volume 2, 1960, pp. 123-147.

[25] E. Lehmann, *Testing Statistical Hypotheses*, John Wiley, 1959.

[26] P. J. Huber, "Robust Estimation of a Location Parameter", *Annals of Mathematical Statistics*, Volume 35, 1964, pp. 73-101.

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

[28] F. R. Hampel, *Contributions to the Theory of Robust Estimation*, Ph. D Dissertation, University of California, Berkeley, 1968.

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

[30] F. R. Hampel, "The Influence Curve and its Role in Robust Estimation", *Journal of the American Statistical Association*, Volume 69, 1974, pp. 383-393.

[31] D. F. Andrews, P. J. Bickel, F. R. Hampel, P. J. Huber, W. H. Rogers and J. W. Tukey, *Robust Estimates of Location: Survey and Advances*, Princeton University Press, 1972.

[32] D. C. Hoaglin, F. Mosteller and J. W. Tukey, *Understanding Robust and Exploratory Data Analysis*, John Wiley, 1983.

[33] F. R. Hampel, E. W. Ronchetti, P. J. Rousseeuw and W. A. Stahel, *Robust Statistics : the Approach based on Influence Functions*, John Wiley, 1986.

[34] R. W. Hill, *Robust Regression when there are Outliers in the Carriers*, Ph.D dissertation, Harvard University, 1977.

[35] C. L. Mallows, *On Some Topics in Robustness*, Unpublished Memorandum, Bell Telephone Laboratories, Murray Hill, N.J., 1975.

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

[37] W. S. Krasker and R. E. Welsch, "Efficient Bounded-influence Regression Estimation", *Journal of the American Statistical Association*, Volume 77, No. 379, 1982, pp. 595-604.

[38] R. A. Maronna, "Robust M-estimators of Multivariate Location and Scatter", *Annals of Statistics*, Volume 4, No. 1, 1976, pp. 51-67.

[39] R. A. Maronna, O. H. Butos and V. J. Yohai, "Bias-and-efficiency Robustness of General M-estimators for Regression with Random Carriers", in *Smoothing Techniques for Curve Estimation*, T. Gasser and M. Rosenblass (eds.) , Lecture Notes in Mathematics 757, Springer-Verlag, 1979, pp. 91-116.

[40] L. Mili, V. Phaniraj and P. J. Rousseeuw, "Robust estimation theory for bad data diagnostics in electric power systems", in *Advances in Control and Dynamical Systems*, (edited by C. T. Leondes), Volume XXXVI, Academic Press, New York, 1991.

[41] A. F. Siegel, "Robust Regression using Repeated Medians", *Biometrika*, Volume 69, No. 1 , 1982, pp. 242-244.

[42] P. J. Rousseeuw, "Least Median of Squares Regression", *Journal of the American Statistical Association*, Volume 79, No. 388, 1984, pp. 871-880.

[43] 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. D. Martin (eds.), Lecture Notes in Statistics, No. 26, Springer Verlag, 1984, pp. 256-272.

[44] V. J. Yohai, "High Breakdown Point and High Efficiency Robust Estimates for Regression", *The Annals of Statistics*, Volume 15, No. 20, 1987, pp. 642 - 656.

[45] V. J. Yohai and R. Zamar, "High Breakdown Point Estimates of regression by means of the minimization of an efficient scale", *Journal of the American Statistical Association*, Volume 83, No. 402, June 1988, pp. 406 - 413.

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

[47] J. H. Friedman and J. W. Tukey, "A Projection Pursuit Algorithm for Exploratory Data Analysis", *IEEE Transactions on Computers,* C-23, pp. 881-889.

[48] J. M. Steele and W. L. Steiger, "Algorithms and Complexity for Least Median of Squares Regression", *Discrete Applied Mathematics,* North-Holland, Volume 14, 1986, pp. 93-100.

[49] A. M. Leroy and P. J. Rousseeuw, *PROGRESS : A program for robust regression*, Report 201, Centrum voor Statistiek en Operationeel Onderzoek, University of Brussels, 1984.

[50] H. Edelsbrunner and D. L. Souvaine, "Computing Least Median of Squares Regression Lines and Guided Topological Sweep", *Journal of the American Statistical Association*, Volume 85, No. 409, March 1990, pp. 115-119.

[51] M. S. Bazaraa and C. M. Shetty, *Nonlinear Programming : Theory and Algorithms*, John Wiley, 1979.

[52] R. A. Becker, J. M. Chambers and A. R. Wilks, *The New S Language : A Programming Environment for Data Analysis and Graphics*, Wadsworth and Brooks/Cole, 1988.

[53] I. W. Slutsker, "Bad Data Identification in Power System State Estimation based on Measurement Compensation and Linear Residual Calculation", *IEEE Transactions on Power Systems*, Volume PWRS-4, No.1, February 1989, pp. 53-60.

[54] N. Sato and W. F. Tinney, "Techniques for exploiting the sparsity of the Network Admittance Matrix", *IEEE Transactions on Power Apparatus and Systems*, Volume PAS-82, No.12, December 1963, pp. 944-950.

[55] J. Carpentier, "Ordered Eliminations", *Proceedings of the Power Systems Computation Conference*, London, 1963.

[56] A. Ralston, *A First Course in Numerical Analysis*, McGraw-Hill, New York, 1965.

[57] B. Stott and E. Hobson, "Solution of large power-system networks by ordered elimination : a comparison of ordering schemes", *Proceedings of the IEEE*, Volume 118, No.1, January 1971.

[58] J. P. Britton, "Improved Area Interchange control for Newton's method load flows", *IEEE Transactions on Power Apparatus and Systems*, Volume PAS-83, No.10, October 1969, pp. 1577-1581.

[59] S. M. Chan and V. Brandwajn, "Partial Matrix Refactorization", *IEEE Transactions on Power Systems*, Volume PWRS-1, No.1, February 1986, pp. 193-200.

[60] W. F. Tinney, S. M. Chan and V. Brandwajn, "Sparse Vector Methods", *IEEE Transactions on Power Apparatus and Systems*, Volume PAS-104, No.2, February 1985, pp. 295-301.

[61] M. Cheniae, *Observability Methods for Least Median of Squares Estimators as applied to Power Systems*, M.S thesis, Department of Electrical Engineering, Virginia Tech, July 1991.

[62] L. Mili, V. Phaniraj and P. J. Rousseeuw, "High Breakdown Point Estimation in Electric Power Systems", presented at the International Symposium on Circuits and Systems, New Orleans, May 1990.

[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*, Volume PAS-99, No. 4, July/August 1980, pp. 1534-1542.

[64] F. Hampel, "Some Problems in Statistics", *Proceedings of the First World Congress of the Bernoulli Society for Mathematical Statistics and Probability*, Tashkent, USSR, Sept. 8-14, 1986.

[65] A. Abur, "A Bad Data Identification Method for Linear Programming State Estimation", *IEEE Transactions on Power Systems*, Volume PWRS-5, No. 3, August 1990, pp. 894-901.

[66] A. Abur and M. K. Celik, "A Robust WLAV State Estimator using Transformations", presented at the IEEE PES Summer Meeting, San Diego, July 1991.

[67] P. J. Rousseeuw and B. C. Van Zomeren, "Unmasking Multivariate Outliers and Leverage Points", to appear in *Journal of the American Statistical Association*, Dec. 1989.

[68] D. L. Donoho, "Breakdown Properties of Multivariate Location Estimators", qualifying paper, Harvard University, 1982.

[69] L. Mili, V. Phaniraj and P. J. Rousseeuw, "Least Median of Squares Estimation in Power Systems", *IEEE Transactions on Power Systems*, Volume PWRS-6, No. 2, May 1991, pp. 511-523.

Vita

V. Phaniraj was born in Kurnool, India on July the Eleventh, Nineteen Hundred and Sixty Two. He received the degree of Bachelor of Technology in Electrical Engineering from the Indian Institute of Technology, Madras in July 1984, and the Master of Science degree in Electrical Engineering from Virginia Tech in July 1986.

Phani Raj. V