ROBUST STATE ESTIMATION BASED ON PROJECTION STATISTICS

L. Mili M. G. Cheniae N. S. Vichare P. J. Rousseeuw

Department of Electrical Engineering
Virginia Polytechnic Institute
and State University
Blacksburg, VA 24061

Mathematics Department
Universitaire Instelling Antwerpen
Universiteitsplein 1
B–2610 Antwerp, Belgium

Abstract – This paper describes a fast and robust method for identifying the leverage points of a linearized power system state estimation model. These are measurements whose projections on the space spanned by the row vectors of the weighted Jacobian matrix, the so-called factor space, do not follow the pattern of the bulk of the point cloud. In other words, their projections are outliers in the factor space. The proposed method is implemented through a new version of the projection algorithm that accounts for the sparsity of the Jacobian matrix. It assigns to each data point a projection statistic defined as the maximum of the standardized projections of the point cloud on some directions passing through the origin. Based on these projection statistics, a robustly weighted Schweppe-type GM-estimator is defined, which can be computed by a reweighted least squares algorithm. The computational efficiency and the robustness of the method are demonstrated on the IEEE-14 bus and the 118-bus systems.

1 – INTRODUCTION

The leverage points of a linearized power system state estimation model are measurement points $(z_i, \underline{\ell}_i)$ whose vectors $\underline{\ell}_i$ define outliers in the n–dimensional "factor space" spanned by the rows $\underline{\ell}_j^T$ of the $m \times n$ weighted Jacobian matrix. By this we mean that their $\underline{\ell}_i$ do not follow the pattern of the majority of the point cloud. It is important to note that there are two types of leverage points. The first type consists of bad leverage points, which are leverage points contaminated by gross errors in z_i and/or $\underline{\ell}_i$. They can ruin any state estimator that minimizes the sum of a certain function of the residuals, the so-called M-estimators. This class includes the now popular weighted least absolute value (WLAV) estimator (see Çelik and Abur [1], El-Keib and Singh [2], among others) along with the weighted least squares (WLS) estimator followed by some rejection rules based on the weighted or the normalized residuals [3]. The second type consists of good leverage points, which significantly enhance the accuracy of M–estimators whenever they are not downweighted. Hence the need to identify the leverage points and to classify them as good or bad by plotting robust regression residuals (for instance least median of squares residuals [4–6]) versus robust distances of the data points in the factor space, as proposed by Rousseeuw and Van Zomeren [7] and illustrated in the closure of [5].

95 WM 216-2 PWRS A paper recommended and approved by the IEEE Power System Engineering Committee of the IEEE Power Engineering Society for presentation at the 1995 IEEE/PES Winter Meeting, January 29, to February 2, 1995, New York, NY. Manuscript submitted December 28, 1993; made available for printing January 11, 1995.

Leverage point identification is carried out by computing a kind of distance from each point in the factor space to the center of the cloud. This distance is standardized to account for the scatter of the point cloud. The theoretical framework for calculating such a distance is provided by multivariate statistics, which deals with estimation of location and covariance matrices. An introductory account of this field can be found in Anderson [8]. Under the assumption that the point cloud follows a multivariate Gaussian distribution, the classical method compares the Mahalanobis distances with a cutoff value which is a quantile of a chi-square distribution. Although these distances are affine equivariant and very easy to calculate, they suffer from one major drawback: they are prone to the masking effect of multiple outliers. This is due to the fact that they hinge on the sample mean and the sample covariance estimator, which are not robust. The same conclusion can be drawn for the diagonal elements of the hat matrix, because they are monotonically related to the classical Mahalanobis distances.

This paper investigates an alternative method that was initiated independently by Stahel [9] and Donoho [10] (for a description, see [11] pp. 238–239). The method proceeds by projecting the point cloud on all possible one-dimensional directions, and assigning to each data point the maximum of the corresponding standardized projections, called its *projection statistic*. Since not all directions can be considered, Gasko and Donoho [12] proposed to explore only those directions that originate from the coordinatewise median and that pass through the data points. However, this method cannot be applied as such to power systems. Indeed, it needs to be adapted to the absence of an intercept term in the state estimation model, and to the sparsity of the factor space. An algorithm that does just that has been developed and tested on several systems. Because it involves only univariate statistics, the method is very fast and compatible with real-time applications. Once the projection statistics are calculated, they are used to define weights for robustifying M-estimators, yielding so-called GM-estimators (G as in generalized) [3, 21, 22]. In this paper, we will use a Schweppe-type GM-estimator with the Huber psi-function [14], abbreviated as SHGM. It can be computed easily by a reweighted least squares algorithm. The advantages of this method over other alternatives are manifold: (i) it does not downweight good leverage points while bounding the influence of the bad ones; (ii) the iterations can be started from the flat voltage profile without sacrificing the robustness of the method; (iii) it is as fast as the conventional algorithm that calculates the WLS estimator; (iv) it can be implemented through a simple modification of the conventional WLS algorithm.

The paper is organized as follows. Section 2 defines the classical estimators of location and covariance matrices, gives the expression of Mahalanobis distances, and advocates the use of the projection algorithm as a more robust method. Section 3 deals with the application of the projection algorithm to power systems. Section 4 reports on Monte Carlo simulations and cutoff values for leverage point identification. Section 5 defines the SHGM estimator and describes the reweighted least squares algorithm. Section 6 illustrates the performance of the method on the IEEE 14-bus system. Some useful definitions and theorems related to affine equivariance and the breakdown point are given in the appendices.

2 – ESTIMATORS OF MULTIVARIATE LOCATION AND COVARIANCE

Consider m data points defined by the set $\{\underline{\ell}_1,\cdots,\underline{\ell}_m\}$ of n–dimensional vectors. We intend to estimate the center and the dispersion of the point cloud by means of a multivariate location estimator and a covariance matrix, respectively.

2.1 Classical estimators

Suppose that the set $\{\underline{\ell}_1,\cdots,\underline{\ell}_m\}$ follows a multivariate normal distribution $N(\underline{\mu}, \underline{\Sigma})$ with mean value $\underline{\mu}$ and covariance matrix $\underline{\Sigma}$. Its probability density function is given by

$$f(\underline{\ell}) = (2\pi)^{-m/2} (\det(\underline{\Sigma}))^{-1/2} \exp\{-\tfrac{1}{2} (\underline{\ell}-\underline{\mu})^T \underline{\Sigma}^{-1} (\underline{\ell}-\underline{\mu})\}. \quad (1)$$

Under the Gaussian model, the maximum likelihood estimate of $\underline{\mu}$ is the sample mean,

$$\underline{\bar{\ell}} = \frac{1}{m} \sum_{i=1}^{m} \underline{\ell}_i , \quad (2)$$

and that of $\underline{\Sigma}$ is the sample covariance matrix. The latter can be made unbiased by replacing the factor $1/m$ by $1/(m-1)$, yielding

$$\underline{C} = \frac{1}{m-1} \sum_{i=1}^{m} (\underline{\ell}_i - \underline{\bar{\ell}}) (\underline{\ell}_i - \underline{\bar{\ell}})^T . \quad (3)$$

Unbiasedness means that, at the model $N(\underline{\mu}, \underline{\Sigma})$, the expectation of $\underline{C}$ is equal to the matrix $\underline{\Sigma}$. Both estimators $\underline{\bar{\ell}}$ and $\underline{C}$ are affine equivariant, that is, do not depend on the coordinate system chosen (see Appendix A) and are statistically efficient at the Gaussian distribution. Unfortunately, this comes at the expense of the breakdown point, which is equal to zero. Indeed, by moving a single data point, the norm of the sample mean may be brought to infinity and the sample covariance matrix may explode (see Appendix B).

2.2 Mahalanobis distances

In order to identify the outliers of the point cloud in n–dimensions, we need a measure of the distance of each point with respect to the bulk of the point cloud. The conventional measure is provided by the Mahalanobis distance, which makes use of the sample mean and the sample covariance matrix given by (2) and (3), respectively. The Mahalanobis distance MD_i of a point $\underline{\ell}_i$ is defined as

$$MD_i = \sqrt{(\underline{\ell}_i - \underline{\bar{\ell}})^T \underline{C}^{-1} (\underline{\ell}_i - \underline{\bar{\ell}})} . \quad (4)$$

Note that the set of points $\underline{\ell}$ satisfying $MD_i^2(\underline{\ell}) \le b$ defines an ellipsoid centered at $\underline{\bar{\ell}}$. Its principal axes are in the directions of the eigenvectors $\underline{v}_j$ of $\underline{C}$ and have lengths equal to $2\,(b\,\lambda_j)^{1/2}$, where λ_j are the corresponding eigenvalues. The volume of the ellipsoid is proportional to the square-root of the determinant of $\underline{C}$, which is equal to $(\lambda_1\lambda_2\cdots\lambda_m)^{1/2}$.

Suppose now that the vectors $\underline{\ell}_i$ are Gaussian with a distribution $N(\underline{\mu}, \underline{\Sigma})$. In that case, the MD_i^2 have approximately a chi-square (χ_n^2) distribution with n degrees of freedom. Therefore, there is a probability of approximately $1-\alpha$ that a point $\underline{\ell}_i$ will fall inside the tolerance ellipsoid satisfying $MD_i^2(\underline{\ell}) \le \chi^2_{n,1-\alpha}$. This provides the rationale of the classical method, which tags as outliers all data points whose Mahalanobis distances are larger than a cutoff value, say larger than $(\chi^2_{n,0.975})^{1/2}$. Because the MD_i are based on non-robust statistics, they are prone to the masking effect of multiple outliers, especially when the latter appear in clusters [7]. In the latter case, the sample mean is attracted by the outliers and the sample covariance matrix is inflated, so that the tolerance ellipsoid covers in part or in totality the set of outliers, inducing the failure of the MD_i to reveal some or all of them (see Section 4.3).

2.3 The hat matrix

In regression, a popular method for leverage diagnostics makes use of the diagonal elements Λ_{ii} of the weighted hat matrix $\underline{\Lambda}$ given by

$$\underline{\Lambda} = \underline{R}^{-1/2}\,\underline{H}\,(\underline{H}^T\,\underline{R}^{-1}\,\underline{H})^{-1}\,\underline{H}^T\,\underline{R}^{-1/2} , \quad (5)$$

where $\underline{R}$ is a given diagonal matrix and $\underline{R}^{-1/2}\,\underline{H} = [\underline{\ell}_1,\cdots,\underline{\ell}_m]^T$. The method flags as outliers all those points whose Λ_{ii} are larger than twice their mean value, which is equal to n/m (Huber [14]). Unfortunately, this method inherits the weaknesses of the squared Mahalanobis distances stated above. As shown by Rousseeuw and Leroy [11], the Λ_{ii} are monotically related to the latter through $\Lambda_{ii} = MD_i^2/(m-1) + 1/m$ when the regression has an intercept, and through $\Lambda_{ii} = MD_i^2/m$ when the regression has no intercept, as in power systems.

2.4 The projection statistic

The first affine equivariant multivariate estimator with a high breakdown point was proposed independently by Stahel [9] and Donoho [10] (for a description, see [11] pp. 238–239). It is motivated by the following equivalent expression of the Mahalanobis distances:

$$MD_i = \max_{\|\underline{v}\|=1} \frac{|\underline{\ell}_i^T\,\underline{v} - L(\underline{\ell}_1^T\,\underline{v},\cdots,\underline{\ell}_m^T\,\underline{v})|}{S(\underline{\ell}_1^T\,\underline{v},\cdots,\underline{\ell}_m^T\,\underline{v})} , \quad (6)$$

where L and S are the sample mean and the sample standard deviation of the projections of the data points $\underline{\ell}_i$ on the direction of the vector $\underline{v}$. Here the equality (6) holds when all possible directions $\underline{v}$ are considered. A natural way to robustify (6) is to use robust estimators of location and scale for L and S, respectively. Gasko and Donoho [12] and Donoho and Gasko [13] suggest utilizing the sample median for L and the median-absolute-deviation (MAD) for S. The latter is defined as

$$MAD = 1.4826 \operatorname*{med}_i |\underline{\ell}_i^T\,\underline{v} - \operatorname*{med}_j (\underline{\ell}_j^T\,\underline{v})| . \quad (7)$$

To overcome the fact that the MAD is aimed at symmetric distributions, and to obtain a better statistical efficiency, Rousseeuw and Croux [15] and Croux and Rousseeuw [16] proposed the following highly robust estimator of scale:

$$S_m = 1.1926 \operatorname*{lomed}_i \operatorname*{lomed}_{j\ne i} |\underline{\ell}_i^T\,\underline{v} - \underline{\ell}_j^T\,\underline{v}| . \quad (8)$$

In (8) lomed denotes a low median defined as the $[(m+1)/2]$-th order statistic out of m numbers. Here, $[x]$ denotes the integer part of the real number x.

Since in practice not all the directions can be investigated, Gasko and Donoho [12] advocate the use of the *projection algorithm*. The idea is to investigate only those directions $\underline{v}$

that originate from the coordinatewise median $\underline{M}$ and pass through each of the data points $\underline{\ell}_i$. Formally, we have $\underline{v} = \underline{\ell}_i - \underline{M}$, where

$$\underline{M} = [\text{median}_j\, \ell_{j1}, \cdots, \text{median}_j\, \ell_{jm}]^T , \qquad (9)$$

and where j runs from 1 to m. The resulting distances calculated with (8) will be called *projection statistics* or, for short, PS_i. Note that, when employing a scale estimate that is affine equivariant, such as S_m or MAD, $\underline{v}$ need not be of unit length. Following Donoho and Gasko [13], we may say that the projection statistic of a point $\underline{\ell}_i$ indicates how far the point is from the bulk of the data set in the worst one-dimensional projection. The advantage of these distances over other alternatives is that they are very fast to compute, even in high dimensions. However, this comes at a price, which is the loss of affine equivariance (see Appendix A). But this is not a disadvantage in power systems because we simply intend to use these distances as a diagnostic tool to identify the leverage points from the original Jacobian matrix, which has been neither transformed nor scaled, or to bound their influence on the M–estimators.

3 – PROJECTION STATISTICS IN POWER SYSTEMS

3.1 Structure of the factor space

In power system state estimation, the *m*–dimensional measurement vector $\underline{z}$ is related to the *n*–dimensional state vector $\underline{x}$ through a nonlinear model given by

$$\underline{z} = \underline{h}(\underline{x}) + \underline{e} , \qquad (10)$$

where $\underline{e}$ is the *m*–dimensional error vector that accounts for the uncertainty in the measurements and the model. It is usually assumed to be a random vector with zero mean and a known covariance matrix $\underline{R} = \text{diag}(\sigma_1^2, \cdots, \sigma_m^2)$. If we perform a first order Taylor series expansion of (10) about an operating point $\underline{x}_0$, we get

$$\Delta \underline{z} = \underline{H}(\underline{x}_0)\, \Delta \underline{x} . \qquad (11)$$

Here $\underline{H}(\underline{x}) = \partial \underline{h}(\underline{x}) / \partial \underline{x}$ is the $m \times n$ Jacobian matrix, which is a very sparse matrix. Let $\{\underline{\ell}_i^T , i = 1, \cdots, m\}$ be the set of the m row vectors of the weighted Jacobian matrix $\underline{R}^{-1/2}\, \underline{H}(\underline{x})$. Each of the m vectors $\underline{\ell}_i$ defines a point in the *n*–dimensional factor space of the regression. There is a one–to–one correspondence between the coordinate axes of this space and the state variables. The main characteristic of the $\underline{\ell}_i$ is that they have few non–zero components. This means that each of them lies on a vector subspace with a small dimension, much smaller than n, termed the *relevant subspace* of that data point. Its dimension is equal to the number of the non–zero entries of that vector. For instance, the dimension of the relevant subspace is equal to 4 for a power flow measurement and usually does not exceed 14 for a power injection measurement, whereas n is typically in the hundreds. When the decoupled model is considered, these dimensions are reduced by a factor of two. In that case, the vector $\underline{\ell}_i$ of a power flow measurement lies exactly on a plane passing through the origin. This plane contains the axes associated with the voltage phase angles at the end buses of the branch on which the flow measurement is located.

3.2 Projection algorithm in power systems

In power systems, the projection algorithm cannot be employed as such because (i) the regression model is without intercept and (ii) the Jacobian matrix is sparse. Indeed, the first condition implies that the ellipsoids are constrained to have the origin as their center. Under this constraint, the projection statistics are defined as

$$PS_i = \max_{\underline{v}} \frac{|\underline{\ell}_i^T\, \underline{v}|}{S_m} , \qquad (12)$$

where $\underline{v} = \underline{\ell}_j$ for j = 1,···,m. However, the denominator of (12) should measure the spread of the $\underline{\ell}_i^T\, \underline{v}$ about the origin. This can be achieved by replacing the pairwise differences in (8) by pairwise sums, yielding the estimate

$$S_m' = 1.1926\, \text{lomed}_i\, \text{lomed}_{j \neq i}\, |\underline{\ell}_i^T\, \underline{v} + \underline{\ell}_j^T\, \underline{v}| . \qquad (13)$$

It has been verified that S_m' has the same statistical properties as S_m in this situation of a fixed origin. For instance, we obtain the same breakdown point and statistical efficiency.

The other problem is raised by the sparsity of the Jacobian matrix. It stems from the fact that the relevant subspace of a point $\underline{\ell}_i$ has a very small dimension. Consequently, the bulk of the point cloud projects at the origin on any direction $\underline{v} = \underline{\ell}_i$, yielding a zero value for the scale estimate S_m' as well as for the MAD, which precludes the calculation of the projection statistics.

One way to solve this problem is to realize that the data points that project at the origin on a given direction $\underline{v} = \underline{\ell}_i$ are irrelevant for $\underline{\ell}_i$ in the sense that they do not bring any information about its outlyingness. Therefore, they should be disregarded when calculating (12) along that direction. For that reason, the vectors $\underline{\ell}_j$ that have a non-zero projection along a direction $\underline{v}$ will be termed the *relevant set* for that direction. In fact, when $\underline{v} = \underline{\ell}_i$, the relevant set is the union of the fundamental sets of the state variables to which the vector $\underline{\ell}_i$ is related. Recall that a fundamental set of a state variable x_i includes all those measurements which observe that state variable, that is, which are functions of it (Mili *et al.* [3, 5]).

3.3 Example of the IEEE 14–bus system

Let us illustrate the concepts of relevant subspace and the relevant set of a data point using IEEE 14–bus system. The one-line diagram of the system is displayed in Figure 1. This system is provided with 34 pairs of real and reactive power measurements and 5 voltage measurements. A decoupled state estimation model is used. In the following, FL i–j denotes a real or a reactive power flow measurement from Bus #i to Bus #j and IN i denotes a real or a reactive power injection measurement at Bus #i.

For instance, consider the real power measurement FL 1–2. The associated vector $\underline{\ell}_{1-2}$ of the weighted P–θ Jacobian submatrix (its entries are the partial derivatives of the real power measurements P with respect to the phase angles θ) lies on a plane containing the coordinate axes associated with Buses #1 and 2. This plane is the relevant subspace of that data point and is of dimension two. The relevant set of points that have a non-zero projection on $\underline{\ell}_{1-2}$ are the $\underline{\ell}_i$ of the real power flow measurements on all the lines that are incident to Buses #1 and 2 and all the real power injection measurements on Buses #1, 2, and 4, yielding a total of 10 data points. These are the $\underline{\ell}_i$ of FL 1–2, 2–1, 1–5, 5–1, 5–2, 2–4, 3–4, and of IN 1, 2, and 4. The vectors $\underline{\ell}_j$ of the other real power measurements have projections that collapse at the origin. When the reactive power measurement FL 1–2 is considered instead, the projections of the voltage measurements at Buses #1 and 2 are to be included too, yielding a total of 12 data points in the relevant set of $\underline{\ell}_{1-2}$.

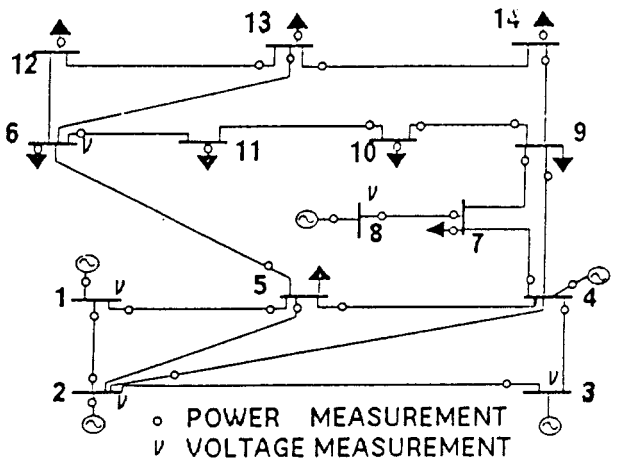

Figure 1. One–line diagram of the 14–bus system.

As a second example, let us determine the relevant subspace and the relevant set for the real power injection measurement IN 2. Its relevant subspace contains the coordinate axes associated with Buses # 1 through 5, implying that it has dimension 5. As for its relevant set, it includes the ℓ_i of FL 1–2, 2–1, 5–2, 2–4, 3–2, 4–3, 5–4, 1–5, 5–1, 5–6, 9–4, 4–7, and of IN 1, 2, 4, 6, 7, yielding a total of 17 data points.

4 – LEVERAGE POINT IDENTIFICATION

4.1 Cutoff values for the projection algorithm

Once the projection statistics have been calculated, they should be compared with cutoff values in order to identify the leverage points. To determine the appropriate cutoff values for the scale estimate S_m', extensive Monte Carlo simulations have been carried out on the IEEE 14– and 118–bus systems. These simulations were conducted along the same lines as those outlined in Rousseeuw and Van Zomeren [17], where an outlier in the factor space is defined as a point that departs from Gaussianity. This definition is supported by the following facts: when the point cloud determined by the ℓ_i follows excatly a multivariate Gaussian distribution, then (i) an M-estimator can handle gross errors affecting the z values and (ii) the distribution of the classical Mahalanobis distances is known (see Section 2.2).

Under Gaussianity, the results reported in [17] suggest that the projection statistics would roughly follow a chi–square distribution. However, it is expected that the number ν of degrees of freedom would be much smaller than n, the dimension of the factor space. In addition, it would probably vary from one data point to another because the dimension of the relevant subspace on which a data point lies varies across the system. As a first guess, ν is put equal to the dimension of the relevant subspace, that is, equal to the number of the non-zero entries of ℓ_i.

To check the validity of this conjecture, Monte Carlo simulations and Q–Q plots (Q stands for quantile) have been carried out according to the following procedure. One hundred different P–θ Jacobian submatrices are randomly generated by replacing the non–zero entries of each column by numbers drawn from a univariate standard Gaussian distribution with zero mean and unit variance. These numbers have been generated using the Box–Muller method (see Press *et al.* [18]). Note that the sparsity structure of these matrices remains the same throughout the simulations because the zero entries are untouched. For each of these submatrices, the projection statistics for a given data point are calculated and ordered, yielding 100 quantiles. The whole procedure is then repeated 50 times so that 50 values of each quantile of the projection statistic with index 1, 2,$\cdots$, 100 are obtained. Finally, the medians and the interquantile ranges of the empirical projection statistic quantiles are plotted versus the corresponding quantiles of the chi–square distribution with the appropriate degree of freedom, ν. Specifically, $\nu = 2$ for a power flow measurement and $\nu = k + 1$ for a power injection measurement at a bus with k incident branches.

As an example, the Q–Q plots for the projection statistics associated with FL 1–2 of the IEEE 14-bus system and IN 24 of the IEEE 118-bus system are displayed in Figures 2 and 3, respectively. The figures reveal that the interquantile ranges are very small for the first 87 quantiles and they increase significantly for larger quantiles. A second observation that can be infered from the plots is that the medians of the first pth empirical quantiles of the projection statistics are roughly located on a straight line with slope one, where p = 99 for FL 1–2 and p = 95 for IN 24. In the latter case, there is a sudden change in the shape of the graph at that quantile. This indicates that the projection statistics behave in first

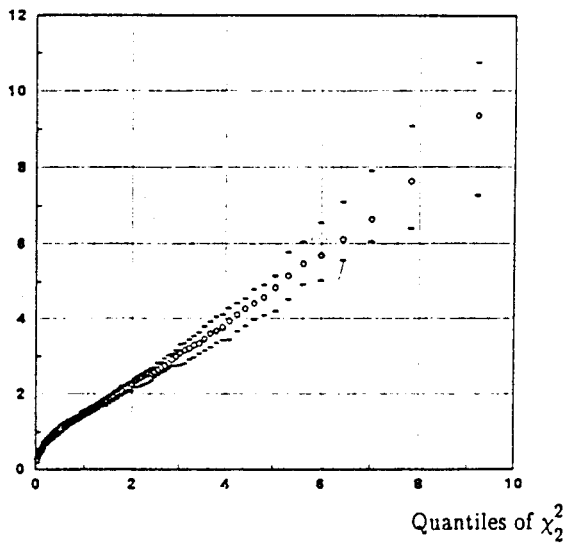

Figure 2. Q–Q plot of the projection statistics for FL 1–2 of the IEEE 14–bus system.

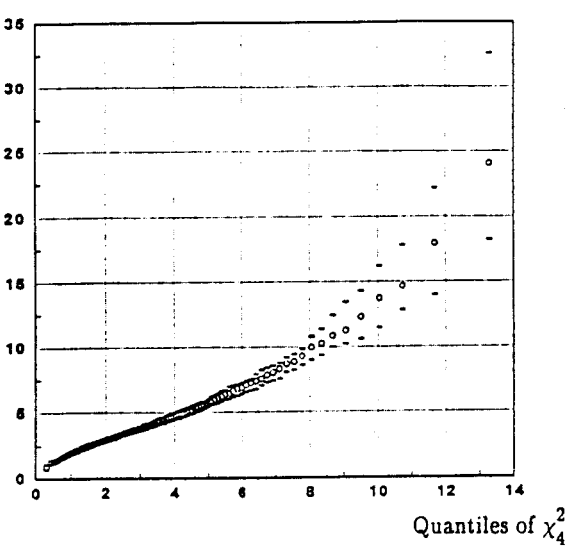

Figure 3. Q–Q plot of the projection statistics for IN 24 of the IEEE 118–bus system.

approximation like a chi–square random variable. This result has been confirmed in all the simulated cases. It allows us to flag as a leverage point any measurement that satisfies

$$PS_i > b_i = \chi^2_{\nu,0.975} , \qquad (14)$$

where ν is the number of non–zero entries in the corresponding row of the Jacobian matrix.

4.2 Robust distances

To improve the statistical efficiency at the Gaussian model, we may calculate robust distances by downweighting the identified outliers. They are given by

$$RD_i = \sqrt{\ell_i^T \, \underline{C}^{-1} \, \ell_i} , \qquad (15)$$

where

$$\underline{C} = \sum_{i=1}^{m} w_i \, \ell_i \, \ell_i^T / \left(\sum_{i=1}^{m} w_i \right) , \qquad (16)$$

$$w_i = \min \left\{ 1 , \left(\frac{b_i}{PS_i} \right)^2 \right\} , \qquad (17)$$

and where b_i is the cutoff value defined by (14). Lopuhaã and Rousseeuw [19] showed that the estimator (16) based on these weights inherits a high breakdown point. Another method consists of deleting the outliers one–by–one by decreasing values of PS_i, where at each deletion the observability of the system is verified, for instance through a graph-theoretic method [20]. Upon completion of the procedure, $\underline{C}$ is calculated through (16) by assigning zero weights ($w_i = 0$) to the removed outliers.

4.3 Example of a 3–bus system

Figure 4 displays the one–line diagram of a 3–bus system for which a dc model is considered. All the lines are assumed to have a zero resistance and a reactance of 1 pu except for Line 1–2 which has a reactance of 0.1 pu. Suppose that the 7 real power measurements that are represented by circles in Figure 4 have unit variances, yielding $\underline{R}_P = \underline{I}$. Let θ_i be the voltage phase angle at Bus #i for 1 = 1, 2, 3. When Bus #3 is taken as the reference bus for the nodal voltage phasors, yielding $\theta_3 = 0$, the weighted P–θ Jacobian submatrix assessed at the flat voltage profile (V = 1 pu and $\theta = 0°$ at all the buses) becomes

$$\underline{R}_P^{1/2} \, \underline{H}_{P\theta} = \underline{H}_{P\theta} = \begin{bmatrix} 10 & 1 & -1 & 0 & 0 & 11 & -1 \\ -10 & 0 & 0 & -1 & 1 & -10 & -1 \end{bmatrix}^T . \qquad (18)$$

The factor space associated with this matrix is a plane that contains 7 points whose coordinates are the entries of the column vectors ℓ_i of $\underline{H}_{P\theta}^T$. This plane is displayed in Figure 5. We observe that the points #1 and 6 lie far away from the others; they are outliers in the factor space. The associated measurements are leverage points. The Mahalanobis distances of the 7 data points are given in Table 1. They are all well below the cutoff value of $(\chi^2_{2,0.975})^{1/2} = 2.72$, even for the two outliers. As a result, all the points fall within the 97.5% tolerance ellipse centered at the origin and defined by

$$MD_i^2 = \ell_i^T \, \underline{C}^{-1} \, \ell_i = \chi^2_{\nu,0.975} , \qquad (19)$$

where

$$\underline{C} = \frac{1}{m} \sum_{i=1}^{m} \ell_i \, \ell_i^T = \frac{1}{m} \, \underline{H}^T \, \underline{H} , \qquad (20)$$

and where m = 7 and $\nu = 2$. The method fails to identify the outliers. The same conclusion holds when using the diagonal elements of the hat matrix instead. As seen in Table 1, they are all smaller than the threshold 2n/m = 0.57. Both methods are hence prone to the masking effect of multiple outliers. By contrast, the projection algorithm is able to identify the two outliers, which have projection statistics larger than $\chi^2_{2,0.975} = 7.38$. For comparison, the robust distances RD_i obtained after the deletion of the two outliers are also given.

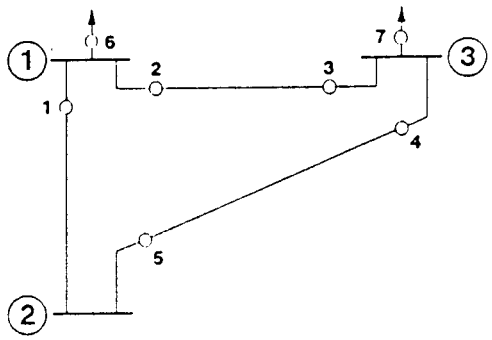

Figure 4. One–line diagram of the 3–bus system.

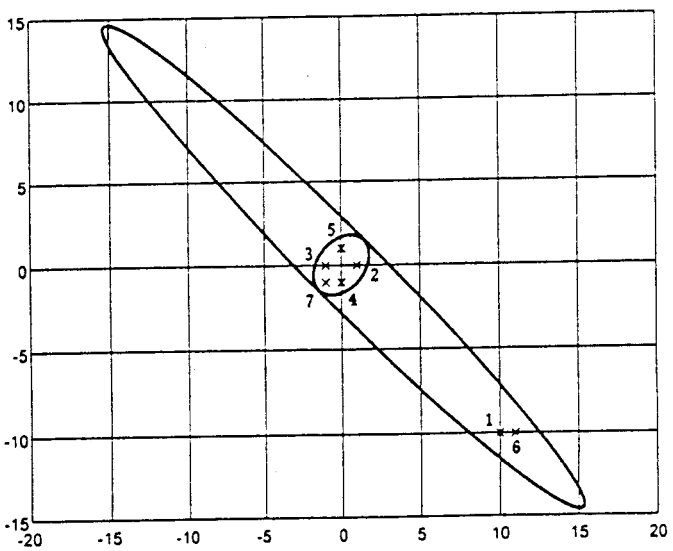

Figure 5. Plot of the 97.5% tolerance ellipses in the factor space for the 3–bus system.

Table 1. Classical and robust distances for the 3–bus system.

Meas. #	Λ_{ii}	MD_i	PS_i	RD_i
1	0.50	1.88	<u>8.39</u>	10.00
2	0.11	0.89	0.84	0.61
3	0.11	0.89	0.84	0.61
4	0.12	0.94	0.84	0.61
5	0.12	0.94	0.84	0.61
6	0.55	1.96	<u>8.82</u>	10.51
7	0.47	1.82	1.68	0.71

5 – ROBUSTIFYING THE M–ESTIMATORS

The weights $w_i = w_i(\ell_i)$ given by (17) can be used to bound the influence of the leverage points on the M–estimators.

To see how this can be achieved, let us first recall the definition of an M-estimator and then analyze its influence function. An M-estimator minimizes an objective function defined as

$$J(\underline{x}) = \sum_{i=1}^{m} \rho\,(r_i / \sigma_i) \ , \qquad (21)$$

where $\rho(\cdot)$ is a real-valued function that is even, non-decreasing for positive residuals, and $\rho(0) = 0$. An M-estimate is hence a solution of $\partial J(\underline{x})/\partial \underline{x} = \underline{0}$, yielding

$$\sum_{i=1}^{m} \underline{\ell}_i \, \psi\,(r_i / \sigma_i) = \underline{0} \ , \qquad (22)$$

where $\psi(u) = \partial\rho(u)/\partial u$ is an odd function (refer to [3, 5, 21] for typical examples of ψ). A first step toward the robustification of an M-estimator is to pick a bounded ψ function. Although such an action effectively bounds the influence of outlying z_i with inlying $\underline{\ell}_i$ (known as vertical outliers), it gives, however, no protection against bad leverage points. Indeed, the latter may completely determine the fit; their residuals may remain small regardless of their z values.

One way to overcome this difficulty is to multiply the vector $\underline{\ell}_i$ in (22) by a weight function w_i that decreases for outlying $\underline{\ell}_i$. This yields the so-called Mallows-type GM-estimators [21]. They are solutions of

$$\sum_{i=1}^{m} w_i \, \underline{\ell}_i \, \psi\,(r_i / \sigma_i) = \underline{0} \qquad (23)$$

and minimize

$$J(\underline{x}) = \sum_{i=1}^{m} w_i \, \rho\,(r_i / \sigma_i) \ . \qquad (24)$$

An obvious drawback of such a proposal is that a leverage point is downweighted irrespective of its residual, that is, irrespective of being good or bad. But a good leverage point significantly improves the accuracy of the estimates and hence, should not be downweighted. This can be achieved by instead using a Schweppe-type GM-estimator [21,22]. The latter is a solution of

$$\sum_{i=1}^{m} w_i \, \underline{\ell}_i \, \psi\left(\frac{r_i}{\sigma_i \, w_i}\right) = \underline{0} \qquad (25)$$

and minimizes

$$J(\underline{x}) = \sum_{i=1}^{m} w_i^2 \, \rho\left(\frac{r_i}{\sigma_i \, w_i}\right) . \qquad (26)$$

Note that in his original proposal made in [22], Schweppe suggests to put $w_i^2 = 1 - \Lambda_{ii}$, where Λ_{ii} are the diagonal elements of the hat matrix defined by (5). However, in our work we will use the w_i given by (17) instead, because these weights can deal with multiple outliers. To achieve good accuracy (statistical efficiency) when the measurement errors are Gaussian, the ρ function is chosen to be quadratic up to a certain value, implying that $\psi\,(r_i/(\sigma_i\, w_i)) = r_i/(\sigma_i\, w_i)$. This leads to the cancellation of the w_i in both the implicit equation (25) and the objective function (26). Consequently, a measurement point with a small standardized residual, $r_{Si} = r_i/(\sigma_i\, w_i)$, is not downweighted, be it leverage or not.

To overcome the difficulty raised by the existence of non-robust multiple solutions, we advocate the use of the Huber ψ function defined as

$$\psi(r_{Si}) = \begin{cases} r_{Si} \ , & |r_{Si}| \le c \\ c \ \mathrm{sign}(r_{Si}) \ , & \text{elsewhere} \end{cases} \qquad (27)$$

which corresponds to a ρ function given by

$$\rho(r_{Si}) = \begin{cases} \frac{1}{2} r_{Si}^2 \ , & |r_{Si}| \le c \\ c\,|r_{Si}| - c^2/2 \ , & \text{elsewhere} \end{cases} \qquad (28)$$

An interesting feature of this estimator, termed Schweppe-Huber GM-estimator or SHGM estimator for short, is that as c tends to zero, it reduces to the generalized least absolute value (GLAV) estimator and as c tends to infinity, it becomes the WLS estimator. As a result, if c is too small, the good leverage points may be severely downweighted (see the remark at the end of this section) whereas if c is too large, the bias of the estimates, though bounded, may be significant. A good choice for c would be a value that ranges between 1 and 3.

Having defined the estimator, we need now to derive an algorithm that finds a solution for (25). To this end, we suggest the use of the reweighted least squares iterative algorithm because it is less prone to numerical problems than the Newton's method. This algorithm can be derived as follows. First, divide and multiply the ψ function in (25) by r_{Si} and put the result in a matrix form, which yields

$$\underline{H}^T \, \underline{R}^{-1} \, \underline{Q} \, \underline{r} = \underline{0} \ . \qquad (29)$$

Here $q(r_{Si}) = \psi(r_{Si})/r_{Si}$, $\underline{Q} = \mathrm{diag}(q(r_{Si}))$, and $\underline{r} = \underline{z} - \underline{h}(\underline{x})$. Second, perform a first order Taylor series expansion of $\underline{h}(\underline{x})$ about $\underline{x}^{(k)}$, which gives

$$\underline{h}(\underline{x}) = \underline{h}(\underline{x}^{(k)}) + \underline{H}(\underline{x}^{(k)}) \, (\underline{x} - \underline{x}^{(k)}) \ . \qquad (30)$$

Finally, substitute (29) into (28) to get

$$\Delta \underline{x}^{(k)} = (\underline{H}^{(k)T} \, \underline{R}^{-1} \underline{Q}^{(k)} \, \underline{H}^{(k)})^{-1} \, \underline{H}^{(k)T} \, \underline{R}^{-1} \underline{Q}^{(k)} \underline{r}^{(k)} \ , \qquad (31)$$

where the superscipt k denotes the iteration step. We see that (31) is identical to the conventional algorithm that solves the WLS estimator except for the presence of the weight matrix $\underline{Q}$, which changes from one iteration to the next. Simulation results showed that this algorithm usually converges in few iterations (3 to 4 iterations) from the flat start. In addition, by using the Jacobian matrix assessed at the flat voltage profile, the w_i need to be computed only once, which can be done off-line. Note that this algorithm is different from the one suggested by Handschin *et al.* [22].

Remark. The Schweppe form of the GLAV estimator is defined by $\rho(r_i/(\sigma_i\, w_i)) = |r_i|/(\sigma_i\, w_i)$ and $\psi(r_i/(\sigma_i\, w_i)) = \psi\,(r_i/\sigma_i) = \mathrm{sign}(r_i)$, which means that (25) is identical to (22). Therefore, the Schweppe form reduces to the Mallows form in this case. As a result, the weights assigned by the GLAV to good leverage points nearly vanish as the leverage points become extreme (for instance by greatly shortening the lines), implying their virtual elimination from the measurement set. This in turn decreases the number of outliers that can be handled by this estimator.

6 – SIMULATION RESULTS

Let us now give some simulation results. They are performed on the IEEE 14-bus system displayed in Figure 1 and described in Section 3.3. Two cases have been simulated. Case 1 in Table 2 displays the projection statistics of the rows of the original P–θ Jacobian submatrix $\underline{H}_{P\theta}$ assessed at the flat voltage profile. This constitutes our base case. We observe that 3 projection statistics emerge from the rest with relatively large values. These are those of FL 5–4, IN 2 and IN 4. This comes as no surprise because Buses #2 and 4 have a large number of incident branches whereas Line 5–4 has a relatively large admittance.

Table 2. Projection statistics of $\underline{H}_{pJ}$ assessed at $1 \angle 0°$ pu.

Meas.	Case 1	Case 2	Meas.	Case 1	Case 2
FL 1–2	6.84	6.84	FL 6–11	1.79	1.53
FL 2–1	6.84	6.84	FL 13–6	2.91	39.81
FL 3–2	1.65	1.65	FL 14–9	0.67	0.55
FL 2–4	1.97	2.30	FL 10–11	2.23	2.23
FL 1–5	0.73	0.73	FL 13–12	0.87	0.87
FL 5–1	0.73	0.73	FL 13–14	0.83	0.83
FL 5–2	1.81	1.82	IN 1	7.26	7.26
FL 4–3	1.95	2.28	IN 2	12.09	12.09
FL 5–4	10.30	10.30	IN 4	16.04	16.04
FL 5–6	1.72	1.48	IN 6	7.74	44.00
FL 4–7	1.82	2.13	IN 7	3.58	3.58
FL 7–8	2.41	1.73	IN 8	2.41	1.73
FL 8–7	2.41	1.73	IN 10	3.80	3.80
FL 9–4	0.64	0.75	IN 11	3.26	3.26
FL 9–7	2.01	1.79	IN 12	2.01	1.44
FL 9–10	2.87	2.87	IN 13	3.81	41.14
FL 10–9	2.87	2.87	IN 14	0.63	0.69

Table 3. Cutoff values for the injection measurements.

Meas.	χ^2_ν	Meas.	χ^2_ν
IN 1	9.35	IN 10	9.35
IN 2	12.83	IN 11	9.35
IN 4	14.45	IN 12	9.35
IN 6	12.83	IN 13	11.14
IN 7	11.14	IN 14	9.35
IN 8	7.38		

In Case 2 displayed in Table 2, Line 6–13 is made 10 times shorter, implying that its admittance is magnified tenfold. As expected, the projection statistics of the measurements associated with this line are increased significantly. Indeed, they jump from 2.91 to 39.81 for FL 13–6, from 7.74 to 44.00 for IN 6, and from 3.81 to 41.14 for IN 13. The projection statistics of the other data points remain nearly unchanged.

In both cases, the cutoff values for the flow measurements are equal to 7.38 whereas those for the injection measurements are given in Table 3.

Under the conditions of Case 2 and under the assumption that the true state is the flat voltage profile, V = 1 pu and $\theta = 0°$, 5 outliers have been introduced in the measurement set. They consist of 2 vertical outliers on FL 5–2 and FL 10–11 and 3 bad leverage point on FL 5–4, FL 13–6, and IN 13. Their z values have been changed from 0 pu to the following pu values: 80 pu for FL 5–2 and FL 5–4, –70 pu for FL 10–11, and 100 pu for FL 13–6 and IN 13, making the latter two ones conforming outliers. The z values of the remaining measurements are drawn from a Gaussian distribution with zero mean and unit variance. Based on the linearized model $\underline{z}_P = \underline{H}_{P\theta}\,\underline{\theta} + \underline{e}_P$ with $\underline{R}_P = \underline{I}$, the SHGM implicit equation given by (25) is solved by means of the algorithm (31) with c = 2.7. As for the w_i, they are evaluated by (17) while using the projection statistics listed in Table 2 under Case 2. The SHGM residuals are displayed in Table 4 along with those of the Huber M-estimator. The latter is obtained by solving (22) through (31) with $w_i = 1$ for all the measurements. We observe that the SHGM estimator properly rejects all the outliers. By contrast, the Huber M–estimator breaks down; it rejects the good leverage point IN 6 (its residual is equal to 91.13) while keeping the outlier IN 13 (its residual is equal to 0.25).

7 – CONCLUSIONS

A method is proposed for identifying the leverage points of a linearized power system state estimation model. It consists of computing the projection statistic of each point, and comparing it to a statistical cutoff value. This new algorithm is more suitable for power system models in that it accounts for the sparsity of the Jacobian matrix. It is shown that only a few data points in the factor space need to be considered when calculating the standardized projections. Cutoff values for leverage point identification have been found through Q-Q plots and Monte Carlo simulations. Because it involves only univariate estimators of scale, the method is very fast and compatible with real–time applications, even for very large systems. Based on these projection statistics, a robustly weighted version of the Schweppe–type GM–estimator can be defined. It is computed by a fast algorithm, which is a simple modification of the conventional WLS iterative method.

Table 4. Weight. residuals of the Huber and SHGM estima ors.

Meas.	HUBER	SHGM	Meas.	HUBER	SHGM
FL 1–2	–0.11	–0.04	FL 6–11	–2.35	–0.96
FL 2–1	0.05	–0.10	FL 13–6	6.50	98.00
FL 3–2	2.25	1.74	FL 14–9	–3.39	–0.85
FL 2–4	–0.28	–0.33	FL 10–11	–67.55	–68.19
FL 1–5	1.14	–0.77	FL 13–12	–2.76	–0.55
FL 5–1	–1.27	–0.91	FL 13–14	–3.31	–0.61
FL 5–2	78.24	78.74	IN 1	0.84	0.62
FL 4–3	–1.61	–1.01	IN 2	1.58	1.40
FL 5–4	73.96	75.81	IN 4	2.24	1.33
FL 5–6	2.20	0.71	IN 6	91.13	0.88
FL 4–7	–1.47	–1.25	IN 7	0.87	0.67
FL 7–8	–0.62	–0.55	IN 8	0.04	–0.03
FL 8–7	0.22	0.15	IN 10	0.82	0.35
FL 9–4	0.84	0.77	IN 11	–0.81	–1.56
FL 9–7	0.93	0.98	IN 12	0.57	0.00
FL 9–10	–0.67	–0.84	IN 13	0.25	96.65
FL 10–9	–1.03	–0.86	IN 14	0.47	0.31

APPENDIX A: Equivariance Properties of an Estimator

We are interested in the class of estimators that enjoy some equivariance properties, especially affine equivariance. The latter guarantees that the estimator does not depend on the coordinate system chosen.

Definition 1. A multivariate location estimator T is said to be translation equivariant if

$$T(\underline{\ell}_1 + \underline{b}, \cdots, \underline{\ell}_m + \underline{b}) = T(\underline{\ell}_1, \cdots, \underline{\ell}_m) + \underline{b} \qquad (28)$$

for all $b \in \mathbb{R}^n$.

Definition 2. A multivariate location estimator T is said to be affine equivariant if

$$T(\underline{A}\,\underline{\ell}_1 + \underline{b}, \cdots, \underline{A}\,\underline{\ell}_m + \underline{b}) = \underline{A}\,T(\underline{\ell}_1, \cdots, \underline{\ell}_m) + \underline{b} \qquad (29)$$

for all $\underline{b} \in \mathbb{R}^n$ and all non–singular $n \times n$ matrices $\underline{A}$.

Definition 3. A covariance estimator $\underline{C}$ is said to be affine equivariant if

$$\underline{C}(\{\underline{A}\,\underline{\ell}_1 + \underline{b}, \cdots, \underline{A}\,\underline{\ell}_m + \underline{b}\}) = \underline{A}\,\underline{C}(\{\underline{\ell}_1, \cdots, \underline{\ell}_m\})\,\underline{A}^T \qquad (30)$$

for all $\underline{b} \in \mathbb{R}^n$ and all non–singular $n \times n$ matrices $\underline{A}$.

APPENDIX B: Breakdown Point of Equivariant Estimators

Donoho and Huber [23] have proposed the following definition for the finite–sample breakdown point. Consider a sample $\mathcal{Z} = \{\underline{\ell}_1, \cdots, \underline{\ell}_m\}$ of m data points in n dimensions. Construct a contaminated sample $\mathcal{Z}^*$ obtained from $\mathcal{Z}$ by replacing f data points $\underline{\ell}_i$ by arbitrary values. Let T be a multivariate location estimator that maps $\mathcal{Z}$ into $T(\mathcal{Z})$. Let the maximum bias be the supremum of $\|T(\mathcal{Z}^*) - T(\mathcal{Z})\|$ taken over all samples $\mathcal{Z}^*$, where $\|\cdot\|$ denotes the Euclidean norm. The

breakdown point $\epsilon^*_m(T,\mathcal{X})$ is the maximum fraction of contamination $\epsilon = f/m$ associated with a finite bias. This would imply that, when ϵ is strictly larger than $\epsilon^*_m(T,\mathcal{X})$, there exist contaminated samples $\mathcal{X}^*$ with $f = m\epsilon$ altered points for which the estimate $T(\mathcal{X}^*)$ is carried over all bounds by the outliers. As m grows to infinity, $\epsilon^*_m(T,\mathcal{X})$ tends asymptotically to a limit that is no larger than 1/2 for any translation equivariant estimator. But what is the upper bound of the finite-sample breakdown point? Lopuhaä and Rousseeuw [19] showed that $\epsilon^*_m(T,\mathcal{X}) \leq [(m-1)/2]/m$.

Unlike a multivariate location estimator which breaks down only in one way, there are two different ways for a covariance estimator to break down. Indeed, it can either implode or explode by a fraction ϵ of arbitrary points. Therefore, the breakdown point of $\underline{C}$ is defined as the maximum fraction of outliers that can neither drive the smallest eigenvalue of $\underline{C}(\mathcal{X}^*)$ arbitrarily close to 0 nor take its largest eigenvalue to infinity.

Under the assumption of linear independency of every set of n vectors $\underline{\ell}_i$, known as the general position assumption [11, 24], the finite-sample breakdown point that any affine equivariant covariance estimator may have, is no larger than $[(m-n-1)/2]/m$ (Davies [24]). This upper bound is much smaller when the general position is violated, as is the case in power systems (Mili *et al.* [6]).

ACKNOWLEDGEMENT

The support of NSF under grants ECS-9009099 and ECS-9257204 is gratefully acknowledged.

REFERENCES

[1] M. K. Çelik and A. Abur, "Use of Scaling in WLAV Estimation of Power System Sates," *IEEE Transactions on Power Systems*, Vol. 7, No. 2, May 1992, pp. 684–692.

[2] A. A. El-Keib and H. Singh, "Fast Linear Programming State Estimation Using the Dual Formulation," *IEEE Transactions on Power Systems*, Vol. 7, No. 2, May 1992, pp. 620–628.

[3] L. Mili, V. Phaniraj, and P. J. Rousseeuw, "Robust Estimation Theory for Bad Data Diagnostics in Electric Power Systems," in *Control and Dynamic Systems: Advances in Theory and Applications, Vol. 37: Advances in Industrial Systems*, edited by C. T. Leondes, Academic Press, 1990.

[4] P. J. Rousseeuw, "Least Median of Squares Regression," *Journal of the American Statistical Association*, Vol. 79, 1984, pp. 871–880.

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

[6] L. Mili, M. G. Cheniae, and P. J. Rousseeuw, "Robust State Estimation of Electric Power Systems," *IEEE Transactions on Circuits and Systems*, Vol. 41, No. 5, May 1994, pp. 349–358.

[7] P. J. Rousseeuw and B. C. Van Zomeren, "Unmasking Multivariate Outliers and Leverage Points," *Journal of the American Statistical Association*, Vol. 85, No. 411, Sept. 1990, pp. 633–651.

[8] T. W. Anderson, *An Introduction to Multivariate Statistical Analysis*, John Wiley, 2nd edition, 1984.

[9] W. A. Stahel, "Robuste Schätzungen: Infinitesimale Optimalität und Schätzungen von Kovarianzmatrizen," Ph.D. Thesis, E.T.H. Zürich, Switzerland, 1981.

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

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

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

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

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

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

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

[17] P. J. Rousseeuw and B. C. Van Zomeren, "Robust Distances: Simulations and Cutoff Values," in *Directions in Robust Statistics and Diagnostics*, Part II, edited by W. Stahel and S. Weisberg, Springer-Verlag, 1991, pp. 195-203.

[18] W. H. Press, B. P. Flannery, S. A. Teukolsky, W. T. Vetterling, *Numerical Recipes: The Art of Scientific Computing*, Cambridge University Press, 1989.

[19] H. P. Lopuhaä and P. J. Rousseeuw, "Breakdown Points of Affine Equivariant Estimators of Multivariate Location and Covariance Matrices," *The Annals of Statistics*, Vol. 19, No. 1, 1991, pp. 229-248.

[20] K. A. Clements, "Observability Methods and Optimal Meter Placement," *Electric Power & Energy Systems*, Vol. 12, No. 2, April 1990, pp. 88–93.

[21] F. R. Hampel, E. M. Ronchetti, P. J. Rousseeuw, and W. A. Stahel, *Robust Statistics: The Approach Based on Influence Functions*, John Wiley, 1986.

[22] 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*, Vol. PAS-94, No. 2, March/April 1975, pp. 329–337.

[23] D. L. Donoho and P. J. Huber, "The Notion of Breakdown Point," in *A Festschrift for Erich L. Lehmann*, eds. P. Bickel, K. Doksum and J. L. Hodges, Jr., Belmont, Calif.: Wadsworth, 1983.

[24] P. L. Davies, "Asymptotic Behaviour of S-Estimates of Multivariate Location Parameters and Dispersion Matrices," *The Annals of Statistics*, Vol. 15, No. 3, 1987, pp. 1269–1292.

BIOGRAPHY

Lamine Mili received a B.S. degree from the Swiss Federal Institute of Technology, Lausanne, in 1976, an M.S. degree from the University of Tunis, in 1983, and a Ph.D. degree from the University of Liège, Belgium, in 1987. He is presently an Associate Professor of Electrical Engineering at Virginia Tech. He is the recipient of an NSF Young Investigator award, 1992.

Michael G. Cheniae received a B.S. in Electrical Engineering from the United States Naval Academy in 1982 and an M.S. degree from Virginia Tech in 1991. Presently, he is completing his Ph.D. in Electrical Engineering at Virginia Tech. His research interests include graph theory, robust estimation theory, and power system analysis and control.

Nitin S. Vichare received the B.S. degree from the University of Washington in 1988, the M.S. and the Ph.D. degree from Virginia Tech in 1990 and 1993, respectively. His research interests include robust estimation theory and power system analysis and control.

Peter J. Rousseeuw is Professor of Statistics at the Universitaire Instelling Antwerpen, Belgium. A former member of the Mathematical Sciences Research Institute in Berkeley, Dr. Rousseeuw performed his Ph.D. research at the Swiss Federal Institute of Technology, Zürich, Switzerland.

DISCUSSION

D. M. VINOD KUMAR (Dept. of Elect. Engg., Regional Engineering College, Warangal)

The authors are to be congratulated for using Projection statistics for the robust state estimation and the paper is well written. The discusser offers the following comments on the paper.

(1) In this paper, leverage point is identified by cutoff values for the projection algorithm and robust distance. By utilizing this method can we identify the leverage points in the critical measurements.

(2) For the execution of state estimation redundancy is one of the important factor. What value of redundancy is considered by the authors.

(3) Is authors tested this method on any practical power system.

(4) Is this method is suitable for on-line applications. If yes, specify the CPU times.

Manuscript received March 1, 1995.

Ali Abur (Texas A&M University, CS, TX 77843) The authors ought to be congratulated for a very interesting and timely paper containing several novel ideas. This discusser has the following comments and questions.

1. A computational procedure to compute the projection statistics (PS_i of eq.(12)), is presented below. This discusser failed to reproduce the results of Table 2 by using this procedure, while the values given in Table 1 were obtained closely. The computational procedure used is as follows:

   Let $L = R^{-1/2} \cdot H = [\ell_1, \ell_2, \ldots, \ell_m]^T$ and

   $$\Omega = L \cdot L^T = \begin{bmatrix} \ell_1^T\ell_1 & \ell_1^T\ell_2 & \ldots & \ell_1^T\ell_m \\ \ell_2^T\ell_1 & \ell_2^T\ell_2 & \ldots & \ell_2^T\ell_m \\ & & \ddots & \\ \ell_m^T\ell_1 & \ell_m^T\ell_2 & \ldots & \ell_m^T\ell_m \end{bmatrix}$$

   Then, the following pseudo code describes the computations:

   do k=1, m
   do i=1, m
   do j=1, m
   if(j.eq.i) then
   else
   $x(j) = |\ \Omega(i,k) + \Omega(j,k)\ |$
   endif
   enddo
   $y(i) = \text{lomed}\ [\ x(1), x(2), \ldots, x(m)\ ]$
   enddo
   $sm(k) = 1.1926\ \text{lomed}\ [\ y(1), y(2), \ldots, y(m)\ ]$
   enddo

   do i=1, m
   $PS(i) = \max\ [\ \frac{\Omega(i,1)}{sm(1)}, \frac{\Omega(i,2)}{sm(2)}, \ldots, \frac{\Omega(i,m)}{sm(m)}\ ]$
   enddo

   The function "lomed" is defined as in the paper, i.e. the $[(m+1)/2]$-th order statistic, $[x]$ representing the integer part of x. Any comments on the validity of the above procedure for computing the projection statistics would be appreciated.

2. The interior point method when applied to the state estimation problem, can also be considered as a reweighted least squares method. The diagonal weighting matrix which is modified at each iteration is inversely proportional to the measurement residuals. Do the Schweppe-Huber GM estimator and the interior point method have similar convergence characteristics ? Do the authors see any advantage in using one over the other ?

Manuscript received March 2, 1995.

L. Mili, M. G. Cheniae, N. S. Vichare, P.J. Rousseeuw: We would like to thank the discussers for their interest in our paper and for their valuable comments. We will respond to each discusser individually.

Dr. Abur:

1. In our simulations, we have multiplied the factor 1.1926 in the expression (13) by a small sample correction factor f_m. As shown by Croux and Rousseeuw [16], f_m makes the scale estimator S'_m approximately unbiased. This factor f_m is defined as

m	2	3	4	5	6	7	8	9
f_m	0.743	1.851	0.954	1.351	0.993	1.198	1.005	1.131

   for $m \le 9$, and for $m > 9$ it is given by

   $$f_m = \frac{m}{m - 0.9} \qquad \text{for m odd}$$

   $$f_m = 1 \qquad \text{for m even.}$$

   The Fortran code of an efficient algorithm for computing S_m is listed in [16]. Note that we can construct a version of our procedure which avoids the use of any correction factor. This is accomplished by determining different cutoff values for the projection statistics PS_i by means of the following simple algorithm:

Step 1: Replace each non-zero entry of the Jacobian matrix by a random value drawn from a univariate standard Gaussian distribution N(0, 1) with zero mean and unit variance. If the value falls within the interval [-0.1, 0.1], just draw another. (In this way, the simulated matrix has the same sparsity as the actual Jacobian matrix.)

Step 2: For this new matrix, calculate all the projection statistics PS_i (for i = 1,...,m, where m is the number of measurements in the system);

Step 3: Repeat Steps 1 and 2 one thousand times. In the jth simulated case (j = 1,..., 1000) compute all the $PS_i(j)$. Therefore, for each i we obtain 1000 values of $PS_i(j)$. (For our purposes, we only need to store the 25 largest of them, which saves memory space);

Step 4: Compute a cutoff value b_i as the 97.5 % quantile of $\{PS_i(j); j = 1, ..., 1000\}$ for i = 1,...,m .

Obviously, any correction factor in PS_i will now cancel out in (17). There are also other advantages. We do not need to approximate the distribution of PS_i by a fixed type of theoretical distribution such as the chi-squared. Also, we do not have to compute the degrees of freedom in each point, because the new method automatically takes the local topology into account.

2. We agree with Dr. Abur that an approximate GLAV solution can be obtained by means of the reweighted least squares algorithm defined by (31) while putting c in (28) equal to a very small yet positive value. Simulation results showed that the algorithm still converges for c set to 10^{-6}. It remains to be seen how this algorithm compares with the one proposed by Singh and Alvarado [A1] that solves the dual formulation of the GLAV estimator through the interior point method. Due to time constraints, we were not able to carry out such a comparison. As mentioned in the paper, we do not recommand the use of the GLAV estimator because it nearly eliminates good leverage points. This is why we prefer to use a GM-step.

Dr. Kumar:

1. An underlying question raised by Dr. Kumar is the following: Can a critical measurement be a leverage point? The answer is negative because the deletion of a critical measurement from the data set will reduce the rank of the Jacobian matrix $\underline{H}$ by one; the remaing vectors $\underline{\ell}_i$ will not span the whole factor space of dimension n. A subsequent question is to know whether either measurement of a critical pair can be a leverage point. The answer is again negative because locally there is only one surplus measurement; the deletion of a critical pair decreases the rank of $\underline{H}$ to n - 1. In fact, a necessary condition for a measurement to be a leverage point is that the deletion of that measurement along with any other in the system will not make $\underline{H}$ singular. In other words, the local measurement surplus must be at least two.

2. For the SHGM estimator to be able to handle two conforming outliers located on the same bus, the local redundancy must be large enough. For instance, the SHGM estimator will break down if there are two conforming outliers on FLP 11-6 and on INP 6 of the IEEE 14-bus system shown in Fig. 1. However, it can handle them if we add a measurement on FLP 6-5 or on INP 5.

3. Test on pratical power systems are not yet completed.

4. The proposed estimation method is indeed suitable for on-line applications, even for very large systems. The projection statistics can be calculated off-line by processing the Jacobian matrix assessed at the flat voltage profile. The on-line part of the method, namely the iterative algorithm given by (31), is as fast as the conventional algorithm that computes the WLS estimator. Fast decoupled approximations similar to the ones proposed for the WLS estimator are under study. They will be described in a forthcoming paper.

References

[A1] H. Singh and F. L. Alvarado, "Weighted Least Absolute State Estimation Using Interior Point Methods", *IEEE Transactions on Power Systems*, Vol. 9, No. 3, August 1994, pp. 1478-1484.

Manuscript received April 17, 1995.