Connotations, Characteristics and New Identification Method of Leverage Measurements

Youzhi Liu¹, Hui Zi¹, Jian Xiao¹, Chunqi Yang², Yanqiu Hou² and Jikeng Lin²,*
¹Guangzhou Power Supply Bureau of Guangdong Grid Co, Ltd, Guangzhou, China
²Department of Electrical Engineering, Tongji University, Shanghai, China

*Corresponding author email: 1552347@tongji.edu.cn

Abstract. Accurately identifying the bad data belonging to the leverage measurements can greatly improve the precision of state estimation. In this paper, we presented the algebraic and geometric characteristics of leverage measurements and proposed a new method to accurately identify the leverage measurements. First, we proved that a measurement infinitely close to the Jacobi matrix column space will be a leverage measurement. Second, we proposed a new method for leverage measurements identification using the modified generalized potential index (GPI). Finally, two cases verified the validity of the proposed method. Compared with other methods, the new method can identify multiple leverage measurements accurately.

Keywords: Leverage measurements; Algebraic and geometric characteristics; Generalized measurement potential index; State estimation.

1. Introduction
The performance of state estimation directly affects the analysis and decision of power grid dispatching. In order to improve the accuracy of the estimation, it is necessary to accurately identify and get rid of various bad data. However, the accurate identification of bad data, especially leverage bad data, has been a challenge for years [1-3].

There are two kinds of definitions of leverage measurements: 1) view the state estimation as the statistical application of regression analysis algorithms in the power system, and thus define the leverage measurements to be the outliers of the regression model in X-space; 2) define leverage measurements according to those diagonal entries close to zero in residual sensitivity matrix. Therefore, there are generally two kinds of methods for identifying leverage measurements based on the definitions above: 1) The first category is based on the first definition. Each row vector of the Jacobi matrix corresponds to a coordinate (a factor) in the multi-dimensional space. The distances between each pair of different factors are used to determine whether the measurement is a leverage measurement or not [5], [4] proposed a method for identifying leverage measurements based on the Mahalanobis Distance (MD). [6] proposed a method based on the MD with minimal covariance determinant (MCD). Since the Jacobi matrix of state estimation is always sparse, the methods may result in the existence of multiple sets consisting of partial measurements whose determinant of the covariance matrix is zero, causing the failure of the calculation or the multiplicity of identification solutions. [7] proposed a projection-statistics-based leverage measurement identification method. It follows that most of the abovementioned methods only consider how to identify a single leverage measurement, which needs to further verify their feasibility and accurateness under the circumstance of multiple leverage measurements. 2) The second category is based on the second definition. This category of methods directly identifies leverage measurements based on the magnitude of the diagonal elements of hat matrix. [8] presented a
method for identifying leverage measurements based on the ratio of the norm of the error vector to that of the residual vector. [9] constructed generalized potential index (GPI) based on the value of diagonal element of the hat matrix to comprehensively measure the effect of multiple leverage measurements on the regression model. [10] introduced this method into power systems, but it is based on the MCD and actually has the similar drawback as mentioned in [6].

Apart from the above leverage measurement identification methods, a number of papers have further investigated the bad data identification of leverage measurements. [11] pointed out that LNRT (last normalized residual test) based on least squares state estimation may mistakenly identify bad data when there are leverage measurements. [7] further presented the SHGM (Schweppe-Huber GM-estimator) estimator. [12] constructed exponential objective functions of standardized residuals to achieve automatic identification and suppression of bad data on leverage measurements.

In this paper, we present the algebraic and geometric characteristics of leverage measurements, and proved that a measurement infinitely closed to the Jacobi matrix column space must be a leverage measurement. We also propose a new method for leverage measurements identification using the modified generalized potential index (GPI). Finally, the cases verify the validity of the proposed method. Compared with other methods, the new method can identify multiple leverage measurements accurately. The following parts of this paper are organized as follows. Section 2 introduces the algebraic and geometric characterization of the leverage measurements. Section 3 proposes a new method for leverage measurements identification based on GPI. In section 4, two cases verify the validity of the method proposed in this paper. Section 5 summarizes this article.

2. Algebraic and Geometric Characterization of the Leverage Measurements

Since leverage measurements have a crucial impact on the accuracy of state estimation, this section summarizes and presents the algebraical and geometrical characteristics of the leverage measurements, which grasps the seemingly-abstract insight of leverage measurements.

2.1. Algebraic Characterization of Leverage Measurement

For a linear regression model \( Y = X \beta + \varepsilon \), solved by the least square method, the estimated values of the observation and the residual vector respectively are [10]:

\[
\hat{Y} = K Y = K (Y_{true} + \varepsilon) = K Y_{true} + K \varepsilon \tag{1}
\]

\[
\hat{\varepsilon} = Y - \hat{Y} = (I - K) Y \tag{2}
\]

where \( Y_{true} \) is the true value of the observation vector, \( \varepsilon \) is the random measurement error vector and \( K = H (H^T R^{-1} H)^{-1} H^T R^{-1} \) is called the hat matrix, where \( R^{-1} \) is the weight matrix of the measurements. When the \( i \)-th measurement is regarded as a leverage measurement and the \( j \)-th measurement is a non-leverage one, (1) can be further expressed as:

\[
\hat{y}_i = k_{yi} y_{i,true} + k_{yi} \varepsilon_i + \sum_j (k_{yi} y_{j,true} + k_{yi} \varepsilon_j)(j \neq i) \tag{3}
\]

\[
\hat{y}_j = k_{yj} y_{j,true} + k_{yj} \varepsilon_j + \sum_l (k_{yj} y_{l,true} + k_{yj} \varepsilon_l)(l \neq j) \tag{4}
\]

From (3) and (4), we can see that the corresponding diagonal element of the leverage measurement is much larger than non-diagonal counterparts (i.e., the diagonal predominates), and accordingly, the error value of this leverage measurement has a much greater influence on the estimated value than other measurements. In contrast, the diagonal element of the non-leverage measurement \( j \) is not numerically different from the non-diagonal element (the diagonal doesn’t predominate), and thus the magnitude of the estimated value \( \hat{y}_j \) is jointly affected by the errors of multiple measurements.

2.2. Geometric Characteristics of Leverage Measurements

For an observable system, the rank of the Jacobi matrix \( \text{rank}(H) = n \), where \( n \) denotes the number of
the state variables. The vector space in which measurements vectors are located can be uniquely decomposed into two mutually orthogonal sub-spaces: the column space of the Jacobi matrix, denoted by \( \mathbb{R}(H) \) and its orthogonal space, denoted by \( \mathbb{R}^\perp(H) \) [8].

The least-square solution of \( \Delta \vec{z} = H \Delta \vec{x} + \vec{e} \) is the projection of \( \Delta \vec{z} \) on \( \mathbb{R}(H) \), i.e. \( \Delta \hat{\vec{z}} = P \Delta \vec{z} \), the projection matrix is: \( P = H(H^T R^{-1} H)^{-1} H^T R^{-1} \). The projection matrix on \( \mathbb{R}^\perp(H) \) is \( P^\perp = I - P = I - H(H^T R^{-1} H)^{-1} H^T R^{-1} \).

Project the measurement error vector on \( \mathbb{R}(H) \) and on \( \mathbb{R}^\perp(H) \) respectively:

\[
e = e_u + e_o
\]

\[
e_u = Pe, \quad e_o = (I - P)e
\]

The measurement residual could be further deduced as [8]:

\[
r = \Delta \vec{z} - \Delta \hat{\vec{z}} = (I - P)e_o
\]

We can see from (7) that the residual can only reflect the projection of \( e \) on the \( \mathbb{R}^\perp(H) \)(i.e., \( e_o \)). When the projection is small, the magnitude of the measurement error cannot be reflected by the value of the residual, which indicates that the error is undetectable, and the degree of undetectability can be measured by undetectable indicators [8]:

\[
UI_i = \frac{\|e_o\|}{\|e_o\|_e}
\]

The larger \( UI_i \) is, the more likely the residual cannot reflect the actual measurement error. In the most extreme situation when \( e_o \to 0 \) i.e., \( e \in \mathbb{R}(H) \), \( UI_i \to \infty \), which is called a “critical measurement” and not necessarily a leverage measurement in [8]. We assume that “critical measurements” are also leverage measurements.

**Proposition:** If \( e_o \to 0 \), \( e \in \mathbb{R}(H) \), \( UI_i \to \infty \), the measurement \( i \) must be a leverage measurement.

**Proof:** (6), (7) show that the projection matrix \( P \) is in accord with the hat matrix \( K \) in the regression model, so the hat matrix \( K \) can be used to replace the projection matrix \( P \).

Let the measurement weight matrix \( R^{-1} = I \) ( \( I \) is an identity matrix), and assume the error \( e = b \cdot \delta_i \), and \( \delta_i = [0, 0, \ldots, j, \ldots, 0]^T \). Then (8) can be expressed as:

\[
UI_i = \left\| K \cdot \delta_i \right\| / \left\| (I - K) \cdot \delta_i \right\| = \left( k_{i1}, k_{i2}, \ldots, k_i, \ldots, k_m \right) \left\| [-k_{i1}, -k_{i2}, \ldots, 1 - k_i, \ldots, -k_m]^T \right\|)
\]

Since \( K \) is idempotent [1], (9) can be further simplified as:

\[
UI_i = \left( k_i / k_i^2 + k_i^2 + \ldots + k_i^2 + \ldots + k_i^2 + 1 - 2k_i \right)^{1/2} = \left[ 1 / (1 - k_i) - 1 \right]^{1/2}
\]

From (10), we can see that there is a direct quantitative relationship between \( k_i \) and \( UI_i \). The larger \( k_i \) is, the more likely the corresponding measurement \( i \) is to be a leverage measurement. Especially, when \( k_i \to 1 \), \( UI_i \to \infty \), which implies that this measurement is a critical measurement. By the judgement rule of a leverage measurement, if \( k_i \) is two or three times larger than the average of the diagonal entries of hat matrix, the measurement \( i \) is surely judged to be a leverage measurement. Thus, the proposition is proven.

The above proposition means that \( UI_i \to \infty \) while \( k_i \to 1 \), the measurement \( i \) has very large leverage and correspondingly it has more impact on the accuracy of state estimation. So its precision is very critical. Therefore, depending on the magnitude of \( UI_i \), we can better determine measurements whose precision need to be improved in order to improve estimation accuracy.
3. A Modified Method for Identifying Leverage Measurements Based on the GPI

As noted in Section 1, the method presented in [9] [10], which claimed to be able to avoid so-called "masking effect" and "swapping effect", is based on the MCD of the covariance matrix to construct the initial suspicious leverage measurements set. While for practical engineering problems such as power systems where Jacobi matrices are inherently sparse, the MCD method will fail in calculating or have the multiple solution results for the initial sets, which is also against with the uniqueness of the identification for leverage measurements in a specific system.

This paper uses the projection statistic-based (PS-based) MD proposed in the literature [7] to construct a modified method to replace the MCD method for the determination of the initial suspicious leverage measurements set. The method avoids the calculation of covariance matrices and directly adopts the magnitude of the projection of each row vector on the relevant subspace, which can greatly improve the accuracy and success rate of the GPI-based method for the identification of leverage measurements. The process of the method is as follows:

1) Suspicious leverage measurements detection based on the projection statistic (PS) method
The identification of suspicious leverage measurements is achieved based on the magnitude of the projection of each row vector of the Jacobi matrix on its relevant subspace [7].

\[ PS_i = \max_k |h_i^T \cdot h_k| / s_k \]  

where \( s_k = 1.1926 \cdot \text{lowmed}(\text{lowmed} |h_i^T \cdot h_k + h_j^T \cdot h_k|) \)  

where lowmed( ) denotes the low median and \( h_i^T, h_j^T, h_k^T \) represent the row vectors of row \( i, j \) and \( k \) of the Jacobi matrix \( H \), respectively. (11) computes the maximum value of the projection of Jacobi matrix row vectors on its relevant subspaces, and \( s_k \) can weigh the degree of dispersion of \( h_i^T \cdot h_k \). Since the calculation of the projection statistic includes no operation on the covariance matrix of row vectors, even in the case of a very sparse Jacobi matrix, it can be successfully carried out and effectively identify the suspicious leverage measurements.

2) Dynamic detection of leverage measurements
The set \( R \) is retained to store the non-leverage measurements and the set \( D \) stores the suspicious leverage measurements identified in step 1). The total number of measurements is \( m \). If the number of suspicious leverage measurements in \( D \) is \( d \) and the number of measurements in \( R \) is \( m-d \), the hat matrix can be decomposed into:

\[ K = \begin{bmatrix} U_R & V \\ V & U_D \end{bmatrix} \]  

where \( U_R = H_R(H^T H)^{-1} H_R^T \) and \( U_D = H_D(H^T H)^{-1} H_D^T \) are symmetric square matrices with the order of \( m-d \) and \( d \), respectively and \( V = H_R(H^T H)^{-1} H_D^T \) is a \((m-d) \times d\) matrix. \( H_R \), \( H_D \) is the Jacobi sub-matrix formed by the row vectors corresponding to the measurements in \( R \) and \( D \), respectively.

Define the index:

\[ w_i^{(D)} = h_i(H_R^T H_R)^{-1} h_i \]  

where \( w_i^{(D)} \) is the \( i \)-th diagonal element of the matrix \( H(H_R^T H_R)^{-1} H^T \) and \( h_i \) is the \( i \)-th row vector in the Jacobi matrix \( H \). \(|D|\) denotes the number of elements in the set \( D \), and \( w_i^{(D)} \) can measure the distance between the measurement \( i \) in set \( D \) and the center of samples in set \( R \) in the X-space.

If we further put a measurement \( i \) (originally belongs to set \( R \)) into set \( D \), we can obtain:

\[ w_i^{(D+1)} = h_i(H_R^T H_R)^{-1} h_i^T + [h_i(H_R^T H_R)^{-1} h_i^T]^T \left[ 1 - h_i(H_R^T H_R)^{-1} h_i^T \right] = w_i^{(D)} / [1 - w_i^{(D)}] \]  

Therefore, the generalized potential index (GPI) is defined as:
\begin{equation}
    p^*_n = \begin{cases} 
    w_n^{(d)} \left[ 1 - (1 - w_n^{(d)}) \right] & \text{if } i \in R \\
    w_n^{(d)} & \text{else } i \in D 
    \end{cases}
\end{equation}

And the threshold for the leverage measurement based on confidence intervals are [9]:

\begin{equation}
    \text{P}_{\text{thres}} = \text{median}(p^*_n) + c \cdot \text{mad}(p^*_n)
\end{equation}

where, median\((p^*_n)\) is the median of \(p^*_n(i=1,2,...,m)\), and mad\((p^*_n)\) denotes the expression:

\[ \text{mad}(p^*_n) = \text{median} | \text{median}(p^*_n) - \text{median}(p^*_n) | \]

and \(c\) is a prescribed constant, usually taken as 2 or 3.

If \(p^*_n > \text{P}_{\text{thres}}\), the measurement \(i\) is a leverage measurement. If the GPI \(p^*_n\) of all measurements in \(D\) are greater than \(\text{P}_{\text{thres}}\), all measurements in \(D\) are considered to be the final leverage measurements, and the identification process terminates; if the GPI \(p^*_n\) for some measurement \(i\) in \(D\) is less than \(\text{P}_{\text{thres}}\), then put the measurement with the smallest \(p^*_n\) into set \(R\), and repeat the process until all \(p^*_n\) measurements in \(D\) is larger than \(\text{P}_{\text{thres}}\).

4. Simulation Results

4.1. Geometric Characterization of Leverage Measurements

The geometric characterization of the leverage measurements is verified on a 2-bus system.

\begin{figure}[h]
\centering
\includegraphics[width=0.5\textwidth]{2-bus-system.png}
\caption{One-line diagram of the 2-bus system.}
\end{figure}

The resistances of line 1 and line 2 in Figure 1 are both zero, and the resistances of the two lines are j0.5 and j0.10, respectively. \(z_1\) and \(z_2\) are the active power measurements. Take bus 2 as the reference bus and \(\theta_2 = 0\). Let the measurements weight matrix \(R^{-1} = I\), then the hat matrix \(K\) (or projection matrix \(P\)) is:

\begin{equation}
    K = \begin{bmatrix}
    0.038 & 0.192 \\
    0.192 & 0.961
    \end{bmatrix}
\end{equation}

The diagonal entry of \(K\) corresponding to measurements \(z_1\) and \(z_2\) are \(k_1 = 0.038\) and \(k_2 = 0.961\), which is close to 1, and thus \(z_2\) is a leverage measurement. The UI for \(z_1\) and \(z_2\) are 0.199 and 4.96, respectively, also verifies that measurement with very large UI must be a leverage measurement.

If the actual phase angle of the voltage at bus 1 is 0.23, i.e., the true measurement vector:

\begin{equation}
    z_{\text{true}} = H \cdot x_{\text{true}} = [0.46, 2.30]^T
\end{equation}

If we manually add a gross error \(e\) which twice as many as the true value on \(z_2\), that is:

\begin{equation}
    e = [0.2z_2]^T
\end{equation}

Then the measurement vector is \(z = [0.46, 6.90]^T\), and the residuals (i.e., the projection of the measurement error on \(\mathbb{R}^\perp(H)\)) obtained with the least-square estimation \(r = [6.01, 0.63]^T\). And the projection of \(e\) on \(\mathbb{R}(H)\) is \(r = [1.34, 6.72]^T\). The projections of \(e\) on \(\mathbb{R}^\perp(H)\) and \(\mathbb{R}(H)\) are shown in Figure 2. From Figure 2, we can see that even though the gross error of \(z_2\) is relatively large, the projection of the residual vector \(r\) on the direction of \(z_2\) will be small for \(z_2\) is close to the column space \(\mathbb{R}(H)\).
4.2. The Performance of the Proposed Method for the Leverage Measurements Identification

The performance of proposed method based on GPI for the leverage measurements identification is tested using the IEEE 118-node system. In the IEEE 118-bus system, the injection measurements are fully configured and each of all lines is equipped with a single-end measurement. The results of the identification of different methods are shown in Figure 3.

*The MCD method fully fails due to the singularity of the covariance matrix in the IEEE 118-bus system. As can be seen from Figure 3, the proposed method in this paper eliminates misidentification of leverage measurements (14, 15, 16, 17 18 and 23) compared to that of the method directly based on the value of diagonal entries of the hat matrix. This means that our method can effectively avoid the “swapping effect” of leverage measurements. On the other hand, we also see that leverage measurement identification by the MCD method completely fails in this case because of the singularity of the covariance matrix (the MCD of the covariance determinant is not equal to 0). The reason of which is mainly attributed to: The MCD-based method is to find the subset of multiple row vectors in the Jacobi matrix whose covariance matrix has the minimal determinant, and consider the set consisting of these selected row vectors in factor space to be the “points cloud” in which normal (non-leverage) points are located, and treat the other vectors farther away from the point cloud to be leverage measurements. However, in IEEE 118-bus system, the number of non-zero elements in each row of the Jacobi matrix is much fewer than that of zero ones, and most of the non-zero elements of the Jacobi matrix don’t lie at the same column, which will unavoidably result in finding multiple matrixes whose determinants are all equal to zero while seeking the set of row vectors with the smallest determinant of the covariance matrix. Moreover, the
selection of the “points cloud” is not unique, and thus the calculation of MD is not unique either. As a result, we will find multiple sets of identification, thus making the identification untrustworthy and the corresponding method fully fails in this case.

5. Conclusion
This paper study the issues related to the leverage measurement, including specifically the unification of the different definitions, the algebraic and geometric properties and a method of identifying leverage measurements accurately. In this paper, we first analyze the consistency of the connotations of the two definitions of leverage measurements. We then proceed to analyze the algebraic and geometric perspectives of the properties and it is proved in propositional form that a measurement infinitely close to the Jacobi matrix column space can only be a leverage measurement. On the basis of the above contents, a new method for accurate leverage measurement identification based on the modified generalized measurement potential index is proposed, which is able to correctly identify multiple leverage measurements. The simulation results verify the feasibility and correctness of the proposed method.

Acknowledgments
Research on improving the normal measurement rate of state estimation and power flow calculation accuracy (GZJKJXM20170034)

Reference
[1] A. Abur, A. Gomez-Exposito 2004 Power System State Estimation: Theory and Implementation (New York: Marcel Dekker).
[2] Do Coutto, Filho M B, de Souza J C S 2014 Enhanced bad data processing by phasor-aided state estimation (IEEE Transactions on Power Systems vol 29) pp 2200-2209
[3] Lin Y, Abur A 2016 A New Framework for Detection and Identification of Network Parameter Errors (IEEE Transactions on Smart Grid vol 2) pp 1698-1706.
[4] Monticelli A, Wu F F, Yen M 2010 Multiple Bad Data Identification for State Estimation by Combinatorial Optimization (IEEE Power Engineering Review vol 6) pp 73-74
[5] Mili L, Cheniae M G, Rousseeuw P J 1994 Robust state estimation of electric power systems (IEEE Transactions on Circuits and Systems I: Fundamental Theory and Applications vol 41) pp 349-358
[6] Rousseeuw P J, Driessen K V 1999 A fast algorithm for the minimum covariance determinant estimator (Technometrics vol 41) pp 212-223
[7] Mili L, Cheniae M G, Vichare N S, et al 1996 Robust state estimation based on projection statistics of power systems (IEEE Transactions on Power Systems vol 11) pp 1118-1127
[8] Bretas N G, London J B A, Alberto L F C, et al 2009 Geometrical approach on masked gross errors for power systems state estimation (IEEE Power & Energy Society General Meeting) pp 1-7
[9] Habshah M, Norazan M R, Rahmatullah Imon A H M 2009 The performance of diagnostic-robust generalized potentials for the identification of multiple high leverage points in linear regression (Journal of Applied Statistics vol 36) pp 507-520
[10] Majumdar A, Pal B C 2016 Bad data detection in the context of leverage point attacks in modern power networks (IEEE Transactions on Smart Grid vol 9) pp 2042-2054
[11] Zhao J, Mili L 2018 Vulnerability of the largest normalized residual statistical test to leverage points (IEEE Transactions on Power Systems vol 2018) pp 4643-4646
[12] Wu W, Guo Y, Zhang B, et al 2011 Robust state estimation method based on maximum exponential square (IET generation, transmission & distribution vol 5) pp 1165-1172.