Article:
Salehi Dobakhshari, A, Terzija, V and Azizi, S orcid.org/0000-0002-9274-1177 (2022) Normalized Deleted Residual Test for Identifying Interacting Bad Data in Power System State Estimation. IEEE Transactions on Power Systems, 37 (5). pp. 4006-4016. ISSN 0885-8950

https://doi.org/10.1109/TPWRS.2022.3144316

Normalized Deleted Residual Test for Identifying Interacting Bad Data in Power System State Estimation

Ahmad Salehi Dobakhshari, *Member, IEEE,* V. Terzija, *Fellow, IEEE,* S. Azizi, *Senior Member, IEEE*

***Abstract*—The Largest Normalized Residual Test (LNRT) has been widely utilized in commercial Power System State Estimation (PSSE) software for bad data identification. The LNRT has proved effective in dealing with single bad data as well as multiple non-interacting and multiple interacting but non-conforming bad data. However, it is known for a long time that when two bad data are both interacting and conforming, i.e. their errors are in agreement, the LNRT may fail to identify either one. Moreover, it has been shown recently that even two interacting and non-conforming bad data can cause the failure of the LNRT. Drawing on sensitivity analysis in linear regression, we develop normalized deleted residuals for suspected measurements so that the agreement in measurement errors are broken. Therefore, the LNRT for normalized deleted residuals will be able to identify the actual bad data point. Furthermore, in the case of AC PSSE, the method does not require calculation of a new hat matrix when a measurement is deleted from the data set. This makes the method computationally cost-effective. Simulation results for identifying different conforming and non-conforming interacting bad data proves that the proposed method can enhance the effectiveness of the LNRT.**

***Index Terms*—Bad data, Largest Normalized Residual Test (LNRT), Power system operation, SCADA, State estimation.**

I. Introduction

POWER System State Estimation (PSSE) is the essential part of modern energy management systems. The Supervisory Control And Data Acquisition (SCADA) system gathers various measurements from substations across the grid in real time. Using these measurements, PSSE software often employs the weighted-least-squares (WLS) method to obtain an estimate of the system state [1]. Moreover, one of the main benefits of the PSSE in practice has been the identification of bad data in telemetered measurements [2]. The Chi-square test firstly determines if bad data is present in the measurement set and if the test is positive, the Largest Normalized Residual Test (LNRT) identifies the bad data in a sequential manner [1]. A thorough review of various algorithms for handling corrupted measurements in PSSE is provided in [3].

The LNRT is reported to be successful in identification of a single bad data as well as multiple non-interacting bad data. Two bad data are called interacting if their residuals are significantly correlated. Quantitatively, the degree of significance depends on the network and measurement topology as well as the desired level of selectivity among measurements [1]. It is shown mathematically that in the case of a single bad data, the largest normalized residual (LNR) among the measurements corresponds to the erroneous measurement [2]. Following the same rationale, it is easy to show that if multiple bad data are non-interacting, the LNR will always correspond to an erroneous measurement. In practise, if two bad data belong to different areas of the network, they can be analyzed separately. This has been used in [4] in order to carry out the LNRT for different sets of interacting suspected measurements, simultaneously. However, it has already been shown in [1] that the LNRT fails in the case of multiple interacting and conforming bad data, where measurement errors are in agreement so that circuit equations such as KCL still holds. A three-bus network in [1] exemplifies the case of two interacting and conforming bad data, which satisfy the same KCL constraint as in the case of good measurements. Therefore after executing the LNRT, a good data is identified as bad data. The LNRT is also shown to be vulnerable to multiple interacting and non-conforming bad data [5]. The authors in [5] also show that a pair of bad interacting leverage measurements can yield a swamping effect such that good measurements are identified as bad while leaving the true bad data unnoticed. In [6] the authors determine the LNRT failure zone for two interacting bad data k and l such that the LNR corresponds to another measurement $i \neq k, l$. This analysis reveals that in 10% of all possible measurement errors for two interacting bad data k and l, another measurement i is incorrectly identified as bad data by the LNRT. The problem concerned in this paper is how to improve the LNRT so that two interacting bad data do not cause failure of this test.

In [7] a combinatorial optimization algorithm is devised to deal with multiple bad data identification. A geometric approach is introduced in [8] where the residual vector is projected onto two orthogonal subspaces defined by suspected measurements. However, it is shown in [9] that both of these methods fail to identify multiple bad data in certain cases.

Hypothesis testing identification (HTI) [10], [11] was proposed to address the shortcomings of the LNRT by collective analysis of all bad data simultaneously. In contrast to the LNRT, the HTI uses a top-bottom approach by forming an initial set of suspected bad data and then excluding good data from this set. The method, however, is vulnerable if the initial suspected set does not involve all bad data as it makes use of the normalized residuals for this choice, and hence there is a possibility of missing one or more bad data whose normalized residuals are small [1]. Besides, compared to the LNRT, HTI has not been adopted as widely in commercial PSSE software.

Robust methods based on non-quadratic objective functions [9], [12]–[17] have also been suggested as alternatives to the LNRT in PSSE. Compared to the residual-based methods such as HTI and the LNRT, the advantage of the robust methods is that they reject outliers automatically, and therefore do not need a post-processing stage [1]. However, a universal non-

quadratic objective function that works for different power systems is hard to find [18] and in the presence of multiple interacting bad data these functions may fall behind the LNRT [6]. Recent advances on robust methods are making them attractive for real-world applications [19].

Compared to HTI and robust estimators, the LNRT is still the most popular method for bad data identification in practice. This paper intends to extend the power of the LNRT to identify two interacting bad data, whether conforming or non-conforming. This is accomplished by the introduction of normalized deleted residuals, a novel concept in PSSE. Deleted residuals are defined in [20] to show how influential a point is in a statistical regression. The deleted residual of measurement i, $r_{i(i)}$, is defined as $z_i - \hat{z}_{i(i)}$, where z_i is the measurement and $\hat{z}_{i(i)}$ is the prediction of measurement i when measurement i is left out of the regression. Deleted residuals appear for example in Cook's distance [21] as a measure for identifying influential measurements in a linear regression model. In [22], the concept of deleted residuals is articulated based on the residual sensitivity matrix of the measurements. We will extend this concept in the context of PSSE in order to overcome the difficulties faced by the LNRT. The concept, however, is different from the innovation analysis in forecasting-aided state estimation (FASE) [23]–[25]. The problem concerned in the present paper is not how to deal with smearing effect; rather, we are presenting a method for resolving swamping effect where the largest normalized residual belongs to a healthy measurement. In FASE, however, this is not the case as the largest normalized residual is assumed to belong to an erroneous measurement. The strength of FASE in contrast to LNRT, which deals with gross errors one by one, is resolving the smearing effect and removing bad data altogether.

The main idea of this paper is to use normalized deleted residuals in order to break the interaction between bad data, which can be conforming or non-conforming. After a suspected bad data is deleted, the other bad data will no longer be masked and the LNRT is guaranteed to identify the latter bad data similar to the case of single or multiple non-interacting bad data. In the next step, the other bad data is correctly identified by the LNRT. The proposed method does not alter the LNRT function in PSSE software; rather, it can be adopted as an add-on to PSSE software in order to counteract the impact of interacting bad data on the LNRT.

The contributions of the paper can be summarized as follows.

- Developing the concept of normalized deleted residuals in order to find the normalized residual when a certain measurement is removed, without the need to re-run the PSSE and recalculate the hat matrix.
- Introducing two theorems to limit the search space for interacting bad data and therefore to reduce the computation time of bad data identification.
- Proposing a novel algorithm as an extension of the LNRT in order to address the swamping effect faced by the LNRT.

The rest of this paper is organized as follows. In Section II, a background of interacting bad data in linear regression analysis and incapability of the LNRT to reject them are presented. Section III introduces normalized deleted residuals in the context of AC PSSE. Section IV presents the proposed modification to the LNRT in order to enable the identification of interacting bad data. Case studies demonstrate the application of the proposed method in Section V, followed by conclusions in Section VI.

II. Background

In this section, the concept of bad data and its identification by the LNRT are first reviewed since the proposed method in this paper is built on residual covariance matrix employed in the LNRT. Next, interacting bad data are examined in the context of linear regression. Finally, the difficulties of interacting and conforming bad data in the DC PSSE, i.e. the PSSE based on the dc power flow model, are examined. The problems that arise in linear regression and DC PSSE are shown later to be also of relevance to the AC PSSE, i.e. when exact AC power flow model is considered.

A. State Estimation and Bad Data Identification

Consider m measurements provided by SCADA system. They are related to the system state by a nonlinear function:

$$\mathbf{z} = \mathbf{h}(\mathbf{x}) + \mathbf{e} \tag{1}$$

where $\mathbf{z_{m\times 1}}$ is the measurement vector, including active and reactive power flows through transmission lines and transformers as well as bus voltage magnitudes. In this equation, $\mathbf{x_{n\times 1}}$ is the system state, including bus voltage magnitudes and phase angles. $\mathbf{e_{m\times 1}}$ is the measurement error vector, usually modeled as Gaussian noise, i.e. $\mathbf{e} \sim \mathcal{N}(\mathbf{0}, \mathbf{R})$. Equation (1) can be linearized as:

$$\Delta\mathbf{z} = \mathbf{H}\Delta\mathbf{x} + \mathbf{e} \tag{2}$$

which can be solved in an iterative manner by the WLS estimator as:

$$\Delta\mathbf{x^t} = (\mathbf{H^T R^{-1} H})^{-1}\mathbf{H^T R^{-1}}\Delta\mathbf{z} \tag{3}$$

where $\mathbf{H} = \frac{\partial \mathbf{h}}{\partial \mathbf{x}}|_{\mathbf{x^t}}$ and $\Delta\mathbf{z} = \mathbf{z} - \mathbf{h}(\mathbf{x^t})$ at iteration t. The vector $\mathbf{x}$ is iteratively updated as $\mathbf{x^{t+1}} = \mathbf{x^t} + \mathbf{\Delta x^t}$ until $\|\Delta\mathbf{x^t}\| < \epsilon$, at which point the estimated linearized measurements are defined as:

$$\Delta\widehat{\mathbf{z}} = \mathbf{H}\Delta\widehat{\mathbf{x}} \tag{4}$$

where $\Delta\widehat{\mathbf{x}}$ is the converged linearized state. Substituting (3) into (4) leads to:

$$\Delta\widehat{\mathbf{z}} = \mathbf{K}\Delta\mathbf{z} \tag{5}$$

where $\mathbf{K}$ is called hat matrix for putting a hat on $\Delta\mathbf{z}$ and defined as:

$$\mathbf{K} = \mathbf{H}(\mathbf{H^T R^{-1} H})^{-1}\mathbf{H^T R^{-1}} \tag{6}$$

The measurement residual vector is defined as:

$$\mathbf{r} = \mathbf{z} - \mathbf{h}(\widehat{\mathbf{x}}) = \mathbf{\Delta z} - \mathbf{\Delta}\widehat{\mathbf{z}} = \mathbf{S}\mathbf{\Delta z} \tag{7}$$

where $\mathbf{S} = \mathbf{I} - \mathbf{K}$ is called the residual sensitivity matrix whose properties lead to [1]:

$$\mathbf{r} = \mathbf{S}\mathbf{e} \tag{8}$$

From (8) and properties of $\mathbf{S}$ it can be shown that $\mathbf{r} \sim \mathcal{N}(\mathbf{0}, \mathbf{SR})$ [1]. Therefore, normalized residuals can be defined as:

$$r_i^N = \frac{|r_i|}{\sqrt{\Omega_{ii}}} \tag{9}$$

where $\mathbf{\Omega} = \mathbf{SR}$ is the residual covariance matrix. Equation (9) can also be written as

$$r_i^N = \frac{|r_i|}{\sigma_i\sqrt{S_{ii}}} \tag{10}$$

It is assumed in (10) that measurements are not correlated and $R_{ii} = \sigma_i^2$ where σ_i^2 is the variance of measurement i. From (9) it can be seen that $\frac{r_i}{\sqrt{\Omega_{ii}}} \sim \mathcal{N}(0,1)$. The LNRT identifies measurement i as erroneous if $\max_l r_l^N = r_i^N > 3$ based on the following theorem [2]:

Theorem 1. *When measurement i is the only bad data point in the measurement set, $r_k^N \leq r_i^N$ for $k = 1, ..., m$.*

B. Interacting Bad Data

The concept of interacting bad data [1] can be illustrated by a graphical display of one-dimensional data. Fig. 1 shows 7 data points (x_i, z_i), $i = 1, ..., 7$. It is intended to estimate the parameters a and b of a straight line $z = ax + b$, which best fits the data points. The true regression line as well as the estimated regression line are reflected in this figure. It can be seen that due to the presence of bad data points 6 and 7, the estimated regression line inclines toward these points. Moreover, data point 5 has the largest residual. This causes data point 5, which is good, to have the largest residual.

In this example, the coefficient matrix is as follows:

$$\mathbf{H} = \begin{bmatrix} 0.5 & 1 & 1.5 & 2 & 2.5 & 3 & 3.5 \\ 1 & 1 & 1 & 1 & 1 & 1 & 1 \end{bmatrix}^{\mathbf{T}} \tag{11}$$

The hat matrix is obtained by (6) and the residual vector is calculated as:

$$\mathbf{r} = \mathbf{z} - \mathbf{H}\hat{\mathbf{x}} = (\mathbf{I} - \mathbf{K})\mathbf{z} = \mathbf{S}\mathbf{z} \tag{12}$$

where $\hat{\mathbf{x}}$ contains the estimated slope and intercept of the regression line and $\mathbf{I}$ is the identity matrix. Normalized residuals are obtained by (10).

Fig. 2 demonstrates the concept of deleted residuals. In this figure, data point 7 is omitted and a new regression line is estimated based on the remaining 6 data points. Let use define $r_{j(i)}$ as the residual of measurement j when measurement i is deleted. Now, it can be seen that the residual of data point 6 when data point 7 is deleted, i.e. $r_{6(7)}$, has the largest value among $r_{i(7)}, i = 1, ..., 6$. This comes as no surprise since it is well known that in the presence of a single bad data, the LNR corresponds to the actual bad data. Table I summarizes the results, where normalized residuals has been calculated considering a standard deviation of $\sigma = 0.1$ for z_i measurements.

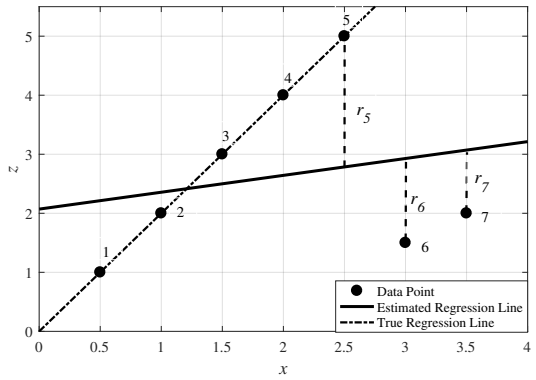

Fig. 1. Linear regression in presence of two interacting bad data.

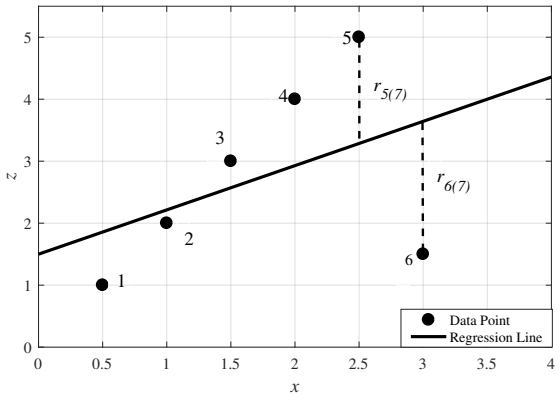

Fig. 2. Linear regression after deleting data point 7.

TABLE I
DATA POINTS AND RESIDUALS FOR LINEAR REGRESSION EXAMPLE

i	x_i	z_i	r_i	S_{ii}	$\lvert r_i^N \rvert$	$r_{i(7)}$	$S_{ii(7)}$	$\lvert r_{i(7)}^N \rvert$
1	0.5	1	-1.21	0.54	16.6	-0.85	0.48	12.4
2	1	2	-0.35	0.72	4.2	-0.21	0.71	2.6
3	1.5	3	0.50	0.82	5.5	0.43	0.82	4.7
4	2	4	1.35	0.86	14.7	1.07	0.82	11.8
5	2.5	5	2.21	0.82	24.4	1.71	0.71	20.4
6	3	1.5	-1.42	0.72	16.9	-2	0.47	31.0
7	3.5	2	-1.07	0.54	14.6	–	–	–

In Table I, $r_{i(7)}$, $S_{ii(7)}$ and $r_{i(7)}^N$ are obtained after deleting the last row of $\mathbf{H}$ in (11). As can be seen, data point 5 is identified as bad data by the LNRT. Nonetheless, if data point 7 is removed from the measurement set, the LNRT successfully identifies data point 6 as bad data.

C. Conforming Bad Data

A simple three-bus network shown in Fig. 4 is used to demonstrate the concept of conforming bad data detailed in [1]. Normalized deleted residuals are calculated for this case in order to identify bad data. It is assumed that bus voltage amplitudes are known and active power measurements are used to estimate the phase angles of bus voltages. All lines are identical and each line has an impedance of $j0.1$ pu. Table II provides the true as well as the actual SCADA measurements.

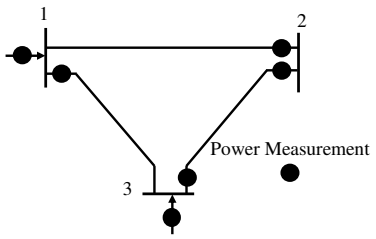

Fig. 3. Single-line diagram of three-bus network [1].

TABLE II
MEASUREMENTS AND DC STATE ESTIMATION RESULTS FOR THREE-BUS EXAMPLE

i	$Meas.$	z_i^{true}	z_i	$\lvert r_i^N \rvert$
1	P_{13}	0.634	0.634	3.247
2	P_{21}	-0.666	-0.666	2.711
3	P_{23}	-0.034	-0.034	5.807
4	P_{32}	0.034	0.134	5.558
5	P_1	1.3	1.3	1.027
6	P_3	-0.6	-0.5	2.755

TABLE III
NORMALIZED DELETED RESIDUALS FOR THREE-BUS EXAMPLE

i	$Meas.$	$\lvert r_{1(i)}^N \rvert$	$\lvert r_{2(i)}^N \rvert$	$\lvert r_{3(i)}^N \rvert$	$\lvert r_{4(i)}^N \rvert$	$\lvert r_{5(i)}^N \rvert$	$\lvert r_{6(i)}^N \rvert$
1	P_{13}	–	2.58	6.23	5.54	2.83	2.16
2	P_{21}	2.71	–	6.68	5.55	0.15	2.93
3	P_{23}	3.87	4.02	–	4.71	0.6	1.08
4	P_{32}	3.16	2.11	5.03	–	2.18	5.82
5	P_1	5.59	4.22	6.53	6.43	–	3.76
6	P_3	3.57	3.19	6.54	7.70	1.46	–

The difference is in measurements of P_3 and P_{32}, which are bad data. It should be noted that bad data still satisfy KCL at bus 3 and therefore P_{13} will not be useful in determining bad data. Measurements P_3 and P_{32} are called conforming bad data. Normalized residuals are calculated for this example as shown in Table II. As P_{23} has the largest normalized residual, the LNRT fails to identify either P_3 or P_{32} as bad data; it incorrectly identifies good measurement P_{23} as an erroneous measurement.

Table III presents normalized deleted residuals for each measurement, i.e. the normalized residuals obtained after removing each single measurement. We define $r_{j(i)}^N$ as the normalized residual of measurement j when measurement i is removed from the measurement set. If we examine rows 4 and 6, which correspond to bad data, we will find that measurement 4 has the LNR if measurement 6 is removed. In other words $r_{4(6)}^N = 7.70$. This is greater than all the figures in column 3 for $r_{3(i)}^N$. This shows that measurement 4 is the actual erroneous measurement not 3. This can be a basis for identifying actual bad data 4 and 6.

III. NORMALIZED DELETED RESIDUALS IN AC PSSE

The two examples given in Section II demonstrate how useful deleted residuals can be in identifying interacting bad data. However, they focused on linear regression and DC PSSE, while in practice the AC PSSE is applied, where accuracy of different meters should be taken into account using a separate variance (σ_k^2) for each measurement k. Normalized deleted residuals in AC state estimation can be calculated in the same way as Section II. To this end, for each measurement i, the PSSE can be run without this measurement. Accordingly, deleted residuals for the remaining measurements j, i.e. $r_{j(i)}$, as well as the residual sensitivity matrix excluding measurement i, i.e $\mathbf{S_{(i)}}$, are obtained. Finally, normalized deleted residuals $r_{j(i)}^N$ are calculated.

The following theorem states that the original measurement residuals as well as the original residual covariance matrix ($\mathbf{\Omega}$) are sufficient for finding the normalized deleted residual for each measurement.

Theorem 2. *The normalized residual of measurement j, when measurement i ($i \neq j$) is deleted from the measurement set, can be directly obtained from:*

$$r_{j(i)}^N = \frac{r_j - \dfrac{\Omega_{ij} r_i}{\Omega_{ii}}}{\sqrt{\Omega_{jj} - \dfrac{\Omega_{ij}^2}{\Omega_{ii}}}} \tag{13}$$

Proof. See Appendix A.

We call $r_{j(i)}^N$ the normalized deleted residual of measurement j when measurement i is omitted from the measurement set. To the best of the authors' knowledge, the above concept of normalized deleted residual expressed through (13) is new in the field of regression analysis and power system state estimation.

Corollary 1. *If measurement i is deleted from the measurement set, it will not be necessary to calculate a new residual sensitivity matrix, i.e. $\mathbf{S_{(i)}}$, in order to obtain normalized deleted residuals $r_{j(i)}^N$ for j=1,...,m.*

According to (13), $r_{j(i)}^N$ includes two other elements of the residual covariance matrix in addition to Ω_{jj}, which is the only element of this matrix that is used for the LNRT. It is worth noting that when i and j are not interacting, i.e. $\Omega_{ij} = 0$, the normalized deleted residual $r_{j(i)}^N$ reduces to the normalized residual r_j^N.

An important result regarding the normalized deleted residual is reflected in Corollary 1. The most computational effort in the LNRT is devoted to calculating the residual sensitivity matrix [1]. Corollary 1 ensures that the primary residual covariance matrix related to the original measurement set is sufficient for calculating normalized deleted residuals.

IV. PROPOSED ALGORITHM

In this section, a new algorithm for identifying two interacting bad data is proposed. The LNRT successfully identifies gross error in measurements unless two bad data p and q cause good measurement i to have the LNR. This was demonstrated in the two examples studied in Section II. Building upon the two theorems presented, the proposed algorithm overcomes this shortcoming of the LNRT. Our assumption is that although the measurement set may contain many erroneous measurements, the interaction between them is confined to pairs of them and it is unlikely that three of them are interacting. Form Theorem 1, if p or q are deleted from the measurement set, the other one is identified as a bad data. In this condition,

measurement i is no longer affected by the swamping effect [26].

Now let us assume the LNR corresponds to measurement i. The first step of the proposed algorithm is that for the identified measurement with the LNR, all possible p, q pairs which may swamp i are identified. The following theorem provides a necessary condition for identifying at least one of the measurements in this pair.

Theorem 3. *If two bad data p and q swamp a good data i such that $r^N_{max} = r^N_i$, then the following holds:*

$$Max\{|S_{ip}|, |S_{iq}|\} > \frac{r^N_i \sqrt{S_{ii}\sigma_i}}{2e_{max}} \tag{14}$$

where e_{max} is the maximum error that measurements p or q may experience.

Proof. See Appendix B.

In practice, gross errors with very large magnitudes ($> 100\sigma$) are easily filtered out in the pre-processing stage, and are unlikely to enter the state estimator [4]. Some literature consider 15% of the meter range as the threshold for gross errors identified during pre-filtering of state estimation [27]. These criteria can be used for quantifying e_{max}.

According to Theorem 3, it is sufficient to gather all the measurements whose residual sensitivity with respect to i is greater than the right-hand side of (14), where i is the measurement with the LNR. For every measurement satisfying this inequality (e.g. p), the other member of the pair (e.g. q) can be found by the following theorem.

Theorem 4. *If two bad data p and q swamp a good data i such that $r^N_{max} = r^N_i$, then one obtains (15), which is presented at the bottom of this page.*

Proof. See Appendix C.

The proposed algorithm is carried out in following steps.

Step 1- Set $I = \varnothing$, where I is the identified bad data set.

Step 2- After removing I from the measurement set, run SE and find measurement residuals (r_j) and residual sensitivity matrix ($\mathbf{S}$).

Step 3- Find $\max_{l \notin I} r^N_l$=r^N_i. If $r^N_i < 3$ then end the algorithm, otherwise continue.

Step 4- Define set $\mathbb{S}_{i1}$, including at least one of the measurements (p) that are interacting with measurement i and may cause healthy i to have the LNR:

$$\mathbb{S}_{i1} = \{j \notin i \mid |S_{ij}| > \epsilon^i_1\} \tag{16}$$

where ϵ^i_1 is defined by the right-hand side of (14).

Step 5- Calculate $r^N_{j(i)}$ for all $j \in \mathbb{S}_{i1}$ using (13).

Step 6- Find $\max_{j \in \mathbb{S}_i} r^N_{j(i)} = r^N_{s(i)}$. If $r^N_{s(i)} > 3$ go to Step 7. If $r^N_{s(i)} < 3$ then we will know that i is actually an erroneous measurement and it has not been swamped by other bad data (p and another q); otherwise we would have had $r^N_{p(i)} > 3$. Remove i from the measurement set, $I \leftarrow I \cup \{i\}$, and go to Step 2.

Step 7- Update $\mathbb{S}_{i1}$ as follows:

$$\mathbb{S}_{i2} = \mathbb{S}_{i1} \cup \left(\bigcup_{p \in \mathbb{S}_{i1}} \mathbb{S}_p \right) \tag{17}$$

where

$$\mathbb{S}_p = \{q \notin I \cup \{i, p\} \mid |S_{pq}| > \epsilon^p_2\} \tag{18}$$

where ϵ^p_2 is defined by the right-hand side of (15). By (17),(18) we ensure that the pair of bad data (p, q) are included in $\mathbb{S}_{i2}$.

7-1) If there is a pair of interacting bad data $\{p, q\} \in \mathbb{S}_{i2}$ such that $\max_{j \in \mathbb{S}_{i2} \cup i} r^N_{j(q)} = r^N_{p(q)} > 3$ and $\max_{j \in \mathbb{S}_{i2} \cup i} r^N_{j(p)} = r^N_{q(p)} > 3$ then include p and q into bad data set by $I \leftarrow I \cup \{p, q\}$ and go to Step 2. If more than such a pair is found, pick the pair with the largest normalized deleted residual. It can be concluded that i is swamped by p and q, i.e. i is healthy but bad data p and q caused r^N_i to have the LNR.

7-2) If there is $k \in \mathbb{S}_{i2}$ such that $\max_{j \in \mathbb{S}_{i2} \cup i} r^N_{j(k)} = r^N_{i(k)} > 3$ and $\max_{j \in \mathbb{S}_{i2} \cup i} r^N_{j(i)} = r^N_{k(i)} > 3$, then include i into bad data set by $I \leftarrow I \cup \{i\}$ and go to Step 2. k is either bad or swamped by other interacting bad data. Either way, it will be dealt with in next cycles.

The proposed algorithm can therefore be considered as an extension of the LNRT in order to handle the swamping effect. A value of 300σ has been adopted for e_{max} in a conservative manner compared to [4], [27]. On the one hand, a larger value for e_{max} ensures that the proposed algorithm is able to counteract interacting bad data and on the other hand leads to a smaller threshold for picking the elements of $\mathbb{S}_{i1}$ and $\mathbb{S}_{i2}$. Based on the system-specific pre-filtering process and engineering judgment a sound value for e_{max} can be selected.

Similar to the case of PSSE, handling multiple outliers by using deleted residuals is known to be complicated in the regression literature. In this regard, there are three inherent problems [22]. The first is how do we determine the size of the subset of jointly influential observations? The second problem is computational. Should m outliers be examined among n data points in search of outliers, for each subset of size m, there will be $\binom{n}{m}$ possible subsets. The third problem with the multiple observations case is graphical. In contrast to the case of single outlier identification, the multiple observations case is not well suited to graphical representations, especially for large n and m.

V. Case Studies

Two power networks studied in previous literature are investigated. A three-bus network similar to Section II is

$$|S_{pq}| > \frac{\sqrt{\left(|S_{ip}|\sqrt{\frac{\Omega_{qq}}{\Omega_{ii}}} - |S_{iq}|\sqrt{\frac{\Omega_{pp}}{\Omega_{ii}}}\right)^2 + 4\left(S_{pp} - |S_{ip}|\sqrt{\frac{\Omega_{pp}}{\Omega_{ii}}}\right)\left(S_{qq} - |S_{iq}|\sqrt{\frac{\Omega_{qq}}{\Omega_{ii}}}\right)}}{2} - \frac{|S_{ip}|\sqrt{\frac{\Omega_{qq}}{\Omega_{ii}}} + |S_{iq}|\sqrt{\frac{\Omega_{pp}}{\Omega_{ii}}}}{2} \tag{15}$$

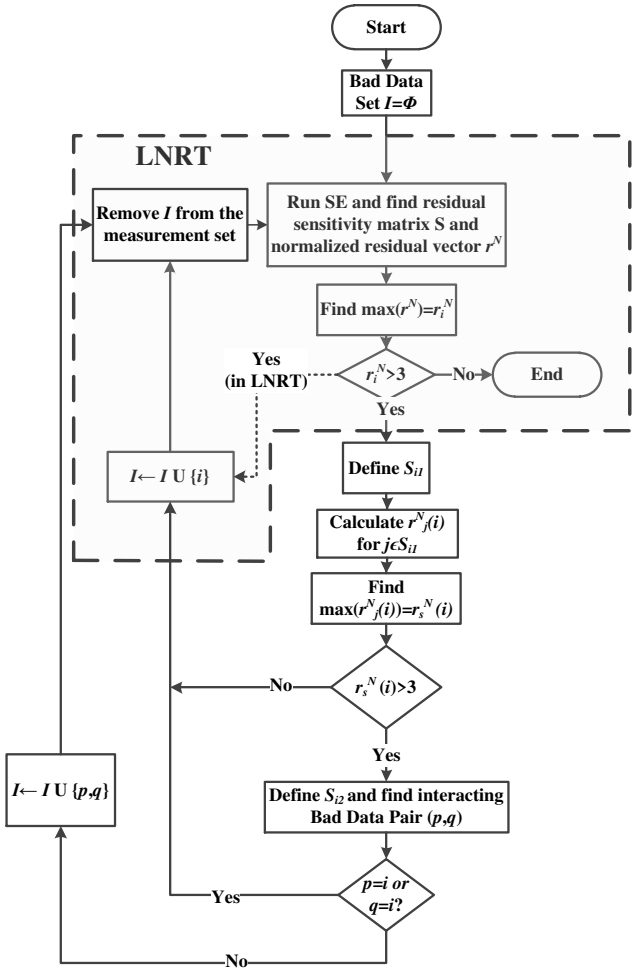

Fig. 4. Flowchart of the proposed algorithm.

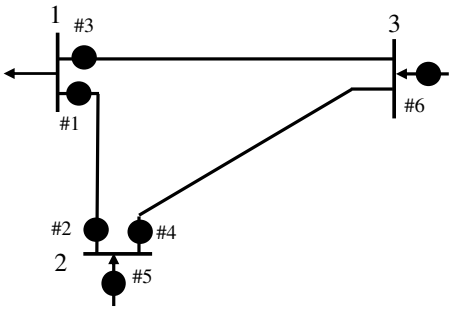

Fig. 5. Single-line diagram of 3-bus network in [7].

adopted to demonstrate the algorithm feasibility, especially in the presence of two bad leverage measurements. The IEEE 30-bus network allows for investigating the case of multiple interacting and non-interacting data as is the case in practice.

A. Three-bus Network Testing

The proposed algorithm is applied to the three-bus network studied in [7], [9]. The single-line diagram and measurements of this network are shown in Fig. 5.

TABLE IV
NORMALIZED DELETED RESIDUALS FOR CASE STUDY IN [7]

i	$\lvert r^N_{1(i)}\rvert$	$\lvert r^N_{2(i)}\rvert$	$\lvert r^N_{3(i)}\rvert$	$\lvert r^N_{4(i)}\rvert$	$\lvert r^N_{5(i)}\rvert$	$\lvert r^N_{6(i)}\rvert$
1	-	39.606	39.606	36.515	0	0
2	44.557	-	14.852	27.386	61.237	20.412
3	65.744	50.572	-	28.702	25.148	9.325
4	61.394	52.623	22.496	-	17.739	27.735
5	51.64	77.46	23.355	23.355	-	0
6	57.735	57.735	25.82	40.825	25.82	-

1) Case Study in [7]: DC PSSE is considered by means of injection and flow measurements reflected in Fig. 5. All lines are assumed to be lossless and have 0.1 pu reactance. The true flow and injection measurement values are zero and standard deviation of all measurement is 0.01 pu. Measurements #2 and #5 in Fig. 3 are bad with a value of 1.0 pu. The normalized residual vector is $\mathbf{r}^N = [58.66, 54.99, 26.21, 31.7, 27.82, 10.37]^T$ and therefore the LNRT identifies measurement #1 as bad data at first, and next rejects measurements #2, #3 and #4, successively.

In the proposed algorithm, since $r^N_{max} = r^N_1$ we enter Step 4 (i=1). With $e_{max} = 300\sigma$, i.e. any measurement error that is larger than 300 times its standard deviation can be identified in the pre-processing stage [4], ϵ^i_1 is calculated as 0.09. Based on (14), $\mathbb{S}_{i1}$ includes all measurements except #1 and #4. Following Step 6, we have $\max_{j\in\mathbb{S}_{i1}} r^N_{j(1)} = r^N_{2(i)} = r^N_{3(i)} = 39.6$. Therefore, we enter Step 7 and calculate $\mathbb{S}_{i2}$. For example, when $p = 2$ and $q = 5$ are considered in (15), we have $\epsilon^p_2 = 0.11$ and $S_{pq} = -0.29$ so that #2 and #5 are included in $\mathbb{S}_{i2}$. The other pair satisfying (15) is (#4,#5). Theses two pairs are suspected to swamp measurement #1. Now to investigate these two pairs according to in Step 7-1, normalized deleted residuals are calculated as shown in Table IV. $r_{2(5)}$ and $r_{5(2)}$ satisfy the conditions in Step 7-1 and therefore are identified as bad data. After measurements #2 and #5 are removed, all residuals will be zero and therefore the algorithm stops.

2) Case Study in [9]: The network in this case study is the same as the previous case, except that line 1-3 is shortened and has a reactance of 0.2 pu. This makes measurements #3 and #6 leverage points [9]. Now consider these two measurements have a value of -1 and 1 pu, respectively. It should be noted that this pair of bad data are conforming as the KCL at bus #6 remains the same. The normalized residual vector is $\mathbf{r}^N = [9.2, 9.2, 9.1, 10.9, 2.4, 5.43]^T$ and therefore the LNRT identifies measurement #4 as bad data at first, and next rejects measurement #5. The two bad leverage points make also [7], [8] identify measurements #4 and #5 as bad data points.

The proposed algorithm starts with the normalized residuals and proceeds with the normalized deleted residuals as shown in Table V. A closer look at this table shows how the situation is much more complicated than the previous case in Table IV. Two pairs of bad data (#1,#2) and (#3,#6) qualify as the pair of bad data in Step 7-1. However, the largest normalized deleted residuals belongs to pair (#3,#6) as $r_{3(6)} = 16.67$ and $r_{6(3)} = 14.97$, while $r_{1(2)} = r_{2(1)} = 11.01$. Therefore, according to Step 7-1, pair (#3,#6) are the bad data points. It can be seen that the proposed algorithm identifies the correct pair of bad data even if both of them are leverage points.

TABLE V
NORMALIZED DELETED RESIDUALS FOR CASE STUDY IN [9]

i	$\lvert r^N_{1(i)}\rvert$	$\lvert r^N_{2(i)}\rvert$	$\lvert r^N_{3(i)}\rvert$	$\lvert r^N_{4(i)}\rvert$	$\lvert r^N_{5(i)}\rvert$	$\lvert r^N_{6(i)}\rvert$
1	-	11.01	7.7498	9.61	1.8011	5.1131
2	11.01	-	7.7498	9.61	1.8011	5.1131
3	7.8774	7.8774	-	11.03	4.7112	14.974
4	7.7019	7.7019	9.3073	-	7.2894	7.5901
5	9.0846	9.0846	9.9743	12.863	-	6.4628
6	9.0388	9.0388	16.667	12.091	4.2471	-

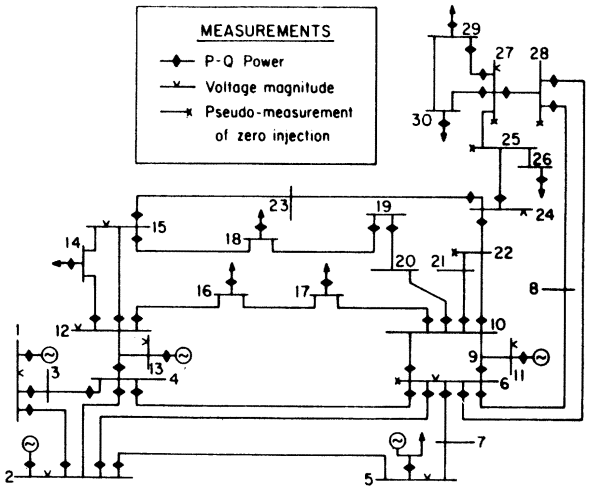

Fig. 6. Single-line diagram of 30-bus network [6], [7], [10].

B. IEEE 30-bus Network Testing

The IEEE 30-bus test system is investigated in two scenarios already studied in the literature and two scenarios introduced in this paper to demonstrate strengths and limitations of the proposed algorithm. The network single-line diagram and measurement configuration is shown in Fig. 6. It should be noted that LNRT fails in all of the scenarios studied below.

1) Case Study in [7], [10]: The IEEE 30-bus test system is investigated in presence of multiple bad data reflected in Table VI. Compared to the three-bus network studied above, both active and reactive interacting bad data are present in the vicinity of bus 1. The first 8 largest normalized residuals are presented in Table VII where it can be seen that healthy measurement P_{2-1} is swamped and identified as bad data in the LNRT. The LNRT proceeds with removing healthy measurements P_2 and P_{1-3}, leading to a completely distorted system state in the vicinity of bus 1.

The proposed algorithm, in contrast, checks for interacting bad data which may have swamped P_{2-1}. It forms $\mathbb{S}_{i1} = \{P_{1-2}, P_1, P_2\}$ and next calculates $r^N_{j(i)}$ for $j \in \mathbb{S}_{i1}$. To validate (13), Table VIII compares $r^N_{j(i)}$ calculated by (13) and $r^N_{j(i)}$ calculated by non-linear state estimation. It can be seen that the every linearized $r^N_{j(i)}$ calculated by (13) is in good agreement with its non-linear counterpart. According to Step 6 of the algorithm, we arrive in Step 7, where $\mathbb{S}_{i2}$ is calculated as $\mathbb{S}_{i2} = \{P_{1-2}, P_{4-3}, P_1, P_2\}$. At Step 7-1, the algorithm successfully identifies P_1 and P_{1-2} as the pair of bad data and subsequently removes them. At the next cycle, Q_{1-2} is identified as bad measurement and the result of checking on the swamping effect reveals that this measurement has not been swamped by other bad data, and therefore Q_{1-2} is removed, next. The next cycles are carried out similarly, resembling the LNRT. Table IX reflects the bad data identified and removed at each cycle of the proposed algorithm.

The proposed algorithm takes 33.1 seconds compared to 12.6 seconds for the LNRT. It is also worth noting that the computation time of the proposed method increases linearly with the system size in the same way as the LNRT. This is due to the fact that interaction between measurements is confined to the neighborhood of the bad data.

2) Case Studies in [6]: The first case in [6] is similar to Table VI but only with the first 4 measurements. The performance of the proposed algorithm in this case is similar to Table IX but only including first, second and fifth rows. The second case of multiple bad data in [6] is reflected in Table X. In this case, P_{2-5}, Q_{2-5} swamp healthy measurement P_5 and the LNRT identifies it as bad data. It should be noted that

TABLE VI
BAD DATA IN THE IEEE 30 BUS NETWORK [10]

Measurement	Actual Value (MW/MVAR)	Measured Value (MW/MVAR)
P_{1-2}	177.3	0
Q_{1-2}	-25.7	22
P_1	261.2	0
Q_1	-27.1	22
P_{24-25}	-0.5	20
Q_{24-25}	2.5	20
P_{29}	-2.4	-12
Q_{29}	-0.9	-12

TABLE VII
LARGEST NORMALIZED RESIDUAL FOR THE STUDIED CASE IN [10]

Measurement	Actual Value
P_{2-1}	79.2
P_1	75.5
P_{1-3}	51.5
P_{1-2}	45.6
P_2	39.8
Q_{2-1}	24.1
Q_1	21.9
...	....

TABLE VIII
COMPARISON BETWEEN THE OUTPUT OF (13) AND $r^N_{j(i)}$ CALCULATED BY NON-LINEAR STATE ESTIMATION FOR THE CASE STUDY IN [7], [10]

j	$r^N_{j(i)}$ by (13)	$r^N_{j(i)}$ by non-linear state estimation
P_{1-2}	25.033	24.827
P_1	42.928	42.496
P_2	70.512	70.905

TABLE IX
BAD DATA IDENTIFICATION BY THE PROPOSED ALGORITHM IN THE SCENARIO OF [7], [10]

Cycle No.	Removed Measurement
1	P_1,P_{1-2}
2	Q_{1-2}
3	P_{24-25}
4	Q_{24-25}
5	Q_1
6	Q_{29}
7	P_{29}

TABLE X
BAD DATA IN THE IEEE 30 BUS NETWORK [6]

Measurement	Actual Value (MW/MVAR)	Measured Value (MW/MVAR)
P_{2-5}	82.6	186.4
Q_{2-5}	2.8	101.7
P_{12-15}	17.6	69.2
Q_{12-15}	7.0	56.1
P_{24-25}	-0.5	19.0
Q_{24-25}	2.5	22.4
P_{29}	-2.4	-12.1
Q_{29}	-0.9	-10.2

TABLE XI
BAD DATA IDENTIFICATION BY THE PROPOSED ALGORITHM IN THE SCENARIO OF [6]

Cycle No.	Removed Measurement
1	P_{2-5}, Q_{2-5}
2	P_{12-15}
3	Q_{12-15}
4	Q_{24-25}
5	P_{24-25}
6	P_{29}
7	Q_{29}

active and reactive power measurements have conventionally been treated as non-interacting measurements. Quantitatively, $S_{P_{2-5},Q_{2-5}}$=-0.02. Although these measurements are hardly interacting, let alone conforming, in this case the wrong measured current through line $2-5$ causes P_5 to have the largest normalized residual. The proposed algorithm, however, enters Step 7 and identifies pair of P_{2-5}, Q_{2-5} as bad data. The next cycles of the proposed algorithm is reflected in Table XI, where the proposed algorithm reduces to the LNRT.

3) Multiple Interacting Bad Data Including Leverage Measurements: Four interacting bad data $\{P_{27-30}, Q_{27-30}, P_{30}, Q_{30}\}$ are studied to investigate a case of bad leverage measurements, i.e. P_{30} and Q_{30}. The active and reactive power measurements have been increased by 20 MW and 20 MVAR, respectively, thereby creating a case of interacting bad data due to leaving KCL at bus 30 intact. Looking at Fig. 6 one can confirm the limited redundancy for estimating voltage phasor at bus 30. Table XII presents the results obtained by the proposed method as well as the LNRT for this case. While the LNRT incorrectly identifies healthy measurement Q_{29} as bad data due to the swamping effect, the proposed algorithm identifies $\{Q_{29-30}, Q_{30}\}$, simultaneously. The proposed method, at the second and third cycles identifies P_{27-30} and P_{30}, respectively. The LNRT, however, identifies Q_{29} at its first cycle due to measurement Q_{30} being a leverage measurement, and at the next cycles removes 5 other healthy measurements, aggravating the state estimate.

4) Impact of critical measurement sets: Consider the first case in [6], where there are 4 bad data listed in the first 4 rows of Table VI. Also, assume that P_{1-3} and Q_{1-3} in Fig. 6 are absent from the measurement set. In this situation, bad data constitute half of the power flow measurements in the area, since $\{Flow_{1-2}, Flow_{2-1}, Inj_1, Inj_2\}$ form a critical measurement set; i.e., removal of them makes the system unobservable. Table XIII presents the outcomes of the

TABLE XII
PROPOSED ALGORITHM VERSUS THE LNRT FOR ACTUAL BAD DATA $\{P_{27-30}, Q_{27-30}, P_{30}, Q_{30}\}$

	Proposed Algorithm	LNRT
Cycle No.	Removed Measurement	Removed Measurement
1	Q_{27-30}, Q_{30}	Q_{29}
2	P_{27-30}	P_{29}
3	P_{30}	P_{27-30}
4	–	Q_{26}
5	–	P_{27}
6	–	Q_{24-25}

proposed algorithm as well as the LNRT. It can be seen that limited measurement redundancy has led to the failure of the proposed algorithm. It is worth noting that the LNRT also fails in this condition. There are similar cases in Fig. 6 where two bad data cannot be identified due to the limited redundancy. This includes $\{Flow_{12-16}, Flow_{20-17}, Inj_{16}, Inj_{17}\}$, $\{Flow_{12-16}, Flow_{27-30}, Inj_{27}, Inj_{29}, Inj_{30}\}$, etc.

This comes as no surprise as the situation can be exemplified by a simple regression problem. Consider the problem of fitting a regression line, where the goal is to find α and β in the model $y = \alpha x + \beta$, using 4 points (x_i,y_i), i=1,..,4. Assume that two of these measurements are erroneous. Fig. 7 depicts why it is impossible to identify the pair of bad data in this case. One can dismiss any two points as bad data and have a regression line using the remaining 2 points. For example, if we consider measurements $\{1, 2\}$ as bad data, regression line is fitted by points $\{3, 4\}$. Conversely, if we consider measurements $\{3, 4\}$ as bad data, regression line is fitted by points $\{1, 2\}$. Another measurement point is therefore required to decide which pair is actual bad data. A similar situation can arise in PSSE. When there is a critical set of at most 4 measurements, 2 bad data cannot be identified. It should be noted that this is not the limitation of the proposed method per se; rather, any other method would fail to identify a pair of bad data in this condition due to the lack of redundancy. A pair of active and reactive power flow, however, can be identified by either the LNRT or the proposed method. For example, pair of bad data $\{P_{16}, Q_{16}\}$ in the former set above or $\{P_{29}, Q_{29}\}$ in the latter set (See Tables VI, X) are identifiable as they are not a member of 4-tuple critical set.

Another condition for the success of the proposed algorithm is the one also needed for the success of the LNRT. The condition is that none of two interacting bad data should belong to a critical pair, i.e. pair of measurements whose removal will make the system unobservable [1]. Assume that measurements (a, b) form a critical pair and actual bad data are measurements (b, c). As $r_a^N = r_b^N$ [1], and this equality also holds after removing c, i.e. $r_a^N(c)$=$r_b^N(c)$, the proposed algorithm cannot distinguish between a and b in identifying bad data. Examples of critical pairs in Fig. 6 include $\{Flow_{12-14}, Inj_{14}\}$, $\{Flow_{15-23}, Flow_{24-23}\}$, etc. It should be pointed out that this is also the limitation of the LNRT as well as any other method due to the limited redundancy in the measurement set.

In cases where bad data appear in critical sets, it will be useful if critical sets are already identified as discussed in [28]–[31]. It should be noted that as long as the removed

TABLE XIII
FAILURE OF BOTH LNRT AND THE PROPOSED ALGORITHM WHEN BAD DATA ARE PART OF A CRITICAL SET.

	Proposed Algorithm	LNRT
Cycle No.	Removed Measurement	Removed Measurement
1	P_{2-1}	P_{2-1}
2	P_{1-2},P_2	P_2
3	Q_{2-1}	P_5
4	Q_2	Q_{2-1}
5	Q_1	Q_2
6	–	P_{3-4}
7	–	P_1

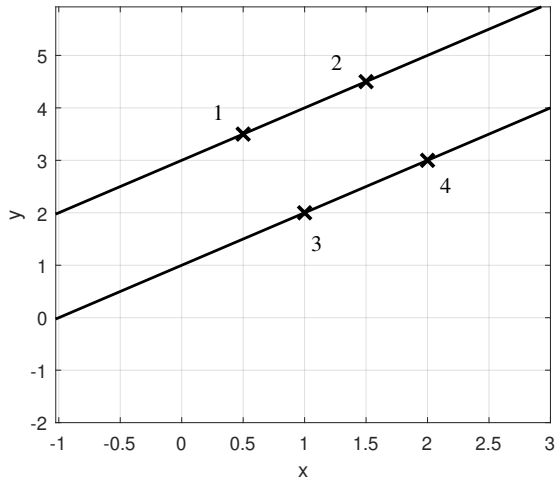

Fig. 7. Fitting a regression line using 4 measurement points of which two are bad data.

measurements are bad data, and this is possible with enough redundancy [1], [8], there is no concern over redundancy reduction. However, similar to the LNRT, it is useful to identify critical sets and modify the algorithm so that when bad data belong to a critical set of measurements, identifiability of bad data had already been taken care of.

VI. CONCLUSION

This paper presents a novel bad data identification algorithm. The concept of normalized deleted residuals is introduced to make the proposed algorithm computationally effective using the same residual sensitivity matrix employed by the largest normalized residual test (LNRT). When the measurement with the largest normalized residual is identified, two possible measurements that are interacting with that measurement are identified and examined for swamping effect. Two theorems are introduced to limit the search space for identifying these two measurements. Different cases in which the LNRT fails to identify interacting bad data are collected from the literature. The proposed algorithm successfully identifies interacting bad data in these cases, while in the case of non-interacting bad data the proposed algorithm reduces to the LNRT. Therefore, the proposed algorithm can be used as an add-on to the existing power system state estimation software for identifying interacting bad data.

APPENDIX A
PROOF OF THEOREM 2

Let us rewrite (2) as $\mathbf{y} = \mathbf{X}\boldsymbol{\beta}+\tilde{\mathbf{e}}$ where $\mathbf{y} = \mathbf{R}^{-\frac{1}{2}}\boldsymbol{\Delta}\mathbf{z}$, $\mathbf{X} = \mathbf{R}^{-\frac{1}{2}}\mathbf{H}$, $\boldsymbol{\beta} = \Delta\mathbf{x}$ and $\tilde{\mathbf{e}} = \mathbf{R}^{-\frac{1}{2}}\mathbf{e}$. It can easily be seen that (3) can be written as $\hat{\boldsymbol{\beta}} = (\mathbf{X^T X})^{-1}\mathbf{X^T y}$ at the final iteration and the hat matrix defined in (4) can be related to the transformed hat matrix as $\tilde{\mathbf{K}} = \mathbf{X}(\mathbf{X^T X})^{-1}\mathbf{X^T} = \mathbf{R}^{-\frac{1}{2}}\mathbf{K}\mathbf{R}^{\frac{1}{2}}$, given that $\mathbf{R}$ is diagonal. It can be shown that with $\tilde{\mathbf{S}} = \mathbf{I} - \tilde{\mathbf{K}}$, we have $S_{ii} = \tilde{S}_{ii}$ but $\tilde{S}_{ij} = \frac{\sqrt{R_{jj}}}{\sqrt{R_{ii}}}S_{ij}$.

Lemma A1. If $\mathbf{X_{(i)}}$ denotes $\mathbf{X}$ excluding the ith row and $\mathbf{X_i}$ denotes the ith row of $\mathbf{X}$ then

$$(\mathbf{X_{(i)}^T X_{(i)}})^{-1} = (\mathbf{X^T X})^{-1} + \frac{(\mathbf{X^T X})^{-1}\mathbf{X_i^T X_i}(\mathbf{X^T X})^{-1}}{\mathbf{S_{ii}}} \quad \text{(A.1)}$$

where $\mathbf{S} = \mathbf{I} - \mathbf{K}$ has already been defined in (7).

Proof. See [22].

If measurement i is removed, the estimated state in (3) changes to:

$$\hat{\boldsymbol{\beta}}_{(i)} = (\mathbf{X_{(i)}^T X_{(i)}})^{-1}\mathbf{X_{(i)}^T y_{(i)}} \quad \text{(A.2)}$$

where the subscript (i) denotes the removal of ith measurement, that is $\mathbf{X_{(i)}}$ and $\mathbf{y_{(i)}}$ are obtained after removing the ith row of $\mathbf{X}$ and $\mathbf{y}$, respectively. Substituting (A.1) into (A.2) leads to:

$$\hat{\boldsymbol{\beta}}_{(i)} = \left((\mathbf{X^T X})^{-1} + \frac{(\mathbf{X^T X})^{-1}\mathbf{X_i^T X_i}(\mathbf{X^T X})^{-1}}{\mathbf{S_{ii}}}\right)(\mathbf{X^T y} - \mathbf{X_i^T y_i}) \quad \text{(A.3)}$$

which can be simplified as

$$\hat{\boldsymbol{\beta}}_{(i)} = \hat{\boldsymbol{\beta}} - \frac{(\mathbf{X^T X})^{-1}\mathbf{X_i^T}\tilde{\mathbf{r}}_\mathbf{i}}{S_{ii}} \quad \text{(A.4)}$$

where $\tilde{r}_i = \frac{r_i}{\sigma_i}$ since $\tilde{\mathbf{r}} = \tilde{\mathbf{S}}\mathbf{y} = \mathbf{R}^{-\frac{1}{2}}\mathbf{S}\mathbf{R}^{\frac{1}{2}}\mathbf{R}^{-\frac{1}{2}}\boldsymbol{\Delta}\mathbf{z} = \mathbf{R}^{-\frac{1}{2}}\mathbf{r}$. Multiplying both sides of (A.4) by $\mathbf{X_j}$, that is the jth row of $\mathbf{X}$, yields:

$$\hat{y}_j - \hat{y}_{j(i)} = \frac{-\tilde{S}_{ji}\tilde{r}_i}{S_{ii}} \quad \text{(A.5)}$$

which can lead to the following by adding $y_j - y_j$ to the left-hand side of (A.5):

$$\tilde{r}_{j(i)} = \tilde{r}_j - \frac{\tilde{S}_{ji}\tilde{r}_i}{S_{ii}} \quad \text{(A.6)}$$

which can be written as

$$\begin{aligned} r_{j(i)} &= r_j - \sqrt{R_{jj}}\frac{\tilde{S}_{ji}\frac{r_i}{\sqrt{R_{ii}}}}{S_{ii}} \\ &= r_j - \frac{S_{ji}r_i}{S_{ii}} \end{aligned} \quad \text{(A.7)}$$

Now, it is sufficient to obtain $S_{jj(i)}$ for normalizing $\tilde{r}_{j(i)}$ in (A.6). $S_{jj(i)} = 1 - K_{jj(i)}$ where $K_{jj(i)}$ is expressed by definition as:

$$K_{jj(i)} = \mathbf{X_j}(\mathbf{X_{(i)}^T X_{(i)}})^{-1}\mathbf{X_j^T} \quad \text{(A.8)}$$

which can be written based on (A.1) as

$$\begin{aligned} K_{jj(i)} &= \mathbf{X_j}\left[(\mathbf{X^T X})^{-1} + \frac{(\mathbf{X^T X})^{-1}\mathbf{X_i^T X_i}(\mathbf{X^T X})^{-1}}{\mathbf{S_{ii}}}\right]\mathbf{X_j^T} \\ &= K_{jj} + \frac{\tilde{K}_{ji}\tilde{K}_{ij}}{S_{ii}} \end{aligned} \tag{A.9}$$

which leads to:

$$S_{jj(i)} = S_{jj} - \frac{\tilde{S}_{ij}^2}{S_{ii}} = \frac{\frac{R_{jj}}{R_{ii}}S_{ij}^2}{S_{ii}} = S_{jj} - \frac{\Omega_{ij}}{\Omega_{ii}}S_{ij} \tag{A.10}$$

Now r_j^N can be obtained by using (A.6) and (A.10) as:

$$\begin{aligned} r_{j(i)}^N &= \frac{r_{j(i)}}{\sqrt{R_{jj}\sqrt{S_{jj(i)}}}} = \frac{r_j - \frac{S_{ji}r_i}{S_{ii}}}{\sqrt{R_{jj}}\sqrt{S_{jj} - \frac{\Omega_{ij}}{\Omega_{ii}}S_{ij}}} \\ &= \frac{r_j - \frac{\Omega_{ji}r_i}{\Omega_{ii}}}{\sqrt{\Omega_{jj} - \frac{\Omega_{ij}^2}{\Omega_{ii}}}} \end{aligned} \tag{A.11}$$

which proves Theorem 2 given the symmetry of $\mathbf{\Omega}$.

Appendix B
Proof of Theorem 3

Assume that p and q are the only two bad data in the measurement set and $r_{max}^N = r_i^N$.

$$\begin{aligned} r_i^N &\approx \frac{|S_{ip}e_p + S_{iq}e_q|}{\sigma_i\sqrt{S_{ii}}} \\ &< \frac{|S_{ip}e_p|}{\sigma_i\sqrt{S_{ii}}} + \frac{|S_{iq}e_q|}{\sigma_i\sqrt{S_{ii}}} \\ &< \frac{2Max\{|S_{ip}|,|S_{iq}|\}e_{max}}{\sigma_i\sqrt{S_{ii}}} \end{aligned} \tag{B.1}$$

where $e_{max} = Max\{|e_p|,|e_q|\}$. Now, (B.1) can be rearranged as:

$$Max\{|S_{ip}|,|S_{iq}|\} > \frac{r_i^N\sqrt{S_{ii}}\sigma_i}{2e_{max}} \tag{B.2}$$

which proves (14).

As the first line of (B.1) is an approximation, let us investigate the impact of the neglected terms. The exact expression for r_i^N in (B.1) is as follows.

$$r_i^N = \frac{\left|\sum_j S_{ij}e_j\right|}{\sigma_i\sqrt{S_{ii}}} = \frac{\left|\sum_{k\in\{Good\ Data\}} S_{ik}e_k + S_{ip}e_p + S_{iq}e_q\right|}{\sigma_i\sqrt{S_{ii}}} \tag{B.3}$$

Given the independence of measurement errors in good data, we have

$$\sum_{k\in\{Good\ Data\}} S_{ik}e_k \sim \mathcal{N}(0, \sum_{k\in\{Good\ Data\}} S_{ik}^2\sigma_k^2) \tag{B.4}$$

Given the fact that the residual sensitivity matrix $\mathbf{S}$ is idempotent [1], we have $\sum_j S_{ij}^2 = S_{ii}$ and therefore (B.4) can be approximated as follows:

$$\sum_{k\in\{Good\ Data\}} S_{ik}e_k \sim \mathcal{N}(0, [S_{ii} - S_{ip}^2 - S_{iq}^2]\bar{\sigma}^2) \tag{B.5}$$

where $\bar{\sigma}^2$ is the root mean square of variances of measurements. By using (B.5), (B.1) can be rewritten as follows:

$$\begin{aligned} r_i^N &< \frac{2Max\{|S_{ip}|,|S_{iq}|\}e_{max}}{\sigma_i\sqrt{S_{ii}}} + \frac{\left|\sum_{k\in\{Good\ Data\}} S_{ik}e_k\right|}{\sigma_i\sqrt{S_{ii}}} \\ &< \frac{2Max\{|S_{ip}|,|S_{iq}|\}e_{max}}{\sigma_i\sqrt{S_{ii}}} + \frac{3\sqrt{[S_{ii} - S_{ip}^2 - S_{iq}^2]}\bar{\sigma}}{\sigma_i\sqrt{S_{ii}}} \\ &< \frac{2Max\{|S_{ip}|,|S_{iq}|\}e_{max}}{\sigma_i\sqrt{S_{ii}}} + \frac{3\bar{\sigma}}{\sigma_i} \end{aligned} \tag{B.6}$$

Accordingly, (B.2) is rewritten as follows:

$$Max\{|S_{ip}|,|S_{iq}|\} > \frac{r_i^N\sqrt{S_{ii}}\sigma_i}{2e_{max}} - \frac{3\bar{\sigma}}{2e_{\max}}\sqrt{S_{ii}} \tag{B.7}$$

A comparison between (B.2) and (B.7) reveals that neglecting the second term in the right-hand side of (B.7) is a good approximation given that e_{max} is set to 300 times the standard deviation and $S_{ii} < 1$ [1].

Appendix C
Proof of Theorem 4

Assume that p and q are the only two bad data in the measurement set such that $r_p^N < r_i^N$ and $r_q^N < r_i^N$. The former inequality can be extended as:

$$|\sqrt{S_{pp}}e_p + \frac{S_{pq}e_q}{\sqrt{S_{pp}}}| < \frac{|S_{ip}e_p + S_{iq}e_q|}{\sqrt{S_{ii}}}\frac{\sigma_{pp}}{\sigma_{ii}} \tag{C.1}$$

On the one hand we always have the following inequality:

$$\sqrt{S_{pp}}|e_p| - \frac{|S_{pq}||e_q|}{\sqrt{S_{pp}}} \le |\sqrt{S_{pp}}e_p + \frac{S_{pq}e_q}{\sqrt{S_{pp}}}| \tag{C.2}$$

On the other hand the following inequality always holds, similarly:

$$\frac{|S_{ip}e_p + S_{iq}e_q|}{\sqrt{S_{ii}}} \le \frac{|S_{ip}||e_p|}{\sqrt{S_{ii}}} + \frac{|S_{iq}||e_q|}{\sqrt{S_{ii}}} \tag{C.3}$$

From (C.1), (C.2) and (C.3) we have:

$$\sqrt{S_{pp}}|e_p| - \frac{|S_{pq}||e_q|}{\sqrt{S_{pp}}} < \left(\frac{|S_{ip}||e_p|}{\sqrt{S_{ii}}} + \frac{|S_{iq}||e_q|}{\sqrt{S_{ii}}}\right)\frac{\sigma_{pp}}{\sigma_{ii}} \tag{C.4}$$

which can be rewritten as:

$$|S_{pq}| > (S_{pp} - |S_{ip}|\sqrt{\frac{\Omega_{pp}}{\Omega_{ii}}})|\frac{e_p}{e_q}| - |S_{iq}|\sqrt{\frac{\Omega_{pp}}{\Omega_{ii}}} \tag{C.5}$$

where $\mathbf{\Omega} = \mathbf{SR}$ is the residual covariance matrix. A similar equation to (C.5) can be written based on $r_q^N < r_i^N$ as:

$$|S_{pq}| > (S_{qq} - |S_{iq}|\sqrt{\frac{\Omega_{qq}}{\Omega_{ii}}})|\frac{e_q}{e_p}| - |S_{ip}|\sqrt{\frac{\Omega_{qq}}{\Omega_{ii}}} \tag{C.6}$$

Given that $S_{pp} - |S_{ip}|\sqrt{\frac{\Omega_{pp}}{\Omega_{ii}}} > 0$ and $S_{qq} - |S_{iq}|\sqrt{\frac{\Omega_{qq}}{\Omega_{ii}}} > 0$ [1], the minimum value for $|S_{pq}|$ will be the intersection of

two diagrams as functions of $|\frac{e_p}{e_q}|$ in (C.5) and (C.6):

$$\begin{aligned}(S_{pp} - |S_{ip}|\sqrt{\frac{\Omega_{pp}}{\Omega_{ii}}})|\frac{e_p}{e_q}| - |S_{iq}|\sqrt{\frac{\Omega_{pp}}{\Omega_{ii}}} = \\ (S_{qq} - |S_{iq}|\sqrt{\frac{\Omega_{qq}}{\Omega_{ii}}})|\frac{e_q}{e_p}| - |S_{ip}|\sqrt{\frac{\Omega_{qq}}{\Omega_{ii}}}\end{aligned} \tag{C.7}$$

which can be written as the following quadratic equation:

$$\begin{aligned}&A|\frac{e_p}{e_q}|^2 + B|\frac{e_p}{e_q}| + C = 0 \\ &A = S_{pp} - |S_{ip}|\sqrt{\frac{\Omega_{pp}}{\Omega_{ii}}} \\ &B = |S_{ip}|\sqrt{\frac{\Omega_{qq}}{\Omega_{ii}}} - |S_{iq}|\sqrt{\frac{\Omega_{pp}}{\Omega_{ii}}} \\ &C = -(S_{qq} - |S_{iq}|\sqrt{\frac{\Omega_{qq}}{\Omega_{ii}}})\end{aligned} \tag{C.8}$$

The feasible root of (C.8) is:

$$x = \frac{\sqrt{B^2 - 4AC} - B}{2A} \tag{C.9}$$

which should be put into (C.7) to give the minimum of $|S_{pq}|$ as follows:

$$\begin{aligned}|S_{pq}| &> Ax - |S_{iq}|\sqrt{\frac{\Omega_{pp}}{\Omega_{ii}}} \\ &= \frac{\sqrt{B^2 - 4AC} - B}{2} - |S_{iq}|\sqrt{\frac{\Omega_{pp}}{\Omega_{ii}}}\end{aligned} \tag{C.10}$$

Substituting (C.8) in (C.10) yields (15).

REFERENCES

[1] A. Abur and A. G. Exposito, *Power system state estimation: theory and implementation*. CRC Press, 2004.

[2] A. Monticelli, *State estimation in electric power systems: a generalized approach*. Springer Science & Business Media, 2012.

[3] M. B. Do Coutto Filho, J. C. S. de Souza, and J. D. Glover, "Roots, achievements, and prospects of power system state estimation: A review on handling corrupted measurements," *Int. Trans. Elect. Ener. Syst.*, vol. 29, no. 4, p. e2779, 2019.

[4] Y. Lin and A. Abur, "A highly efficient bad data identification approach for very large scale power systems," *IEEE Trans. Power Syst.*, vol. 33, no. 6, pp. 5979–5989, 2018.

[5] J. Zhao and L. Mili, "Vulnerability of the largest normalized residual statistical test to leverage points," *IEEE Trans. Power Syst.*, vol. 33, no. 4, pp. 4643 – 4646, 2018.

[6] T. Van Cutsem, M. Ribbens-Pavella, and L. Mili, "Bad data identification methods in power system state estimation-a comparative study," *IEEE Trans. Power Appar. Syst.*, no. 11, pp. 3037–3049, 1985.

[7] A. Monticelli, F. F. Wu, and M. Yen, "Mutiple bad data identification for state estimation by combinatorial optimization," *IEEE Trans. Power Del.*, vol. 1, no. 3, pp. 361–369, 1986.

[8] K. Clements and P. Davis, "Multiple bad data detectability and identifiability: a geometric approach," *IEEE Trans. Power Del.*, vol. 1, no. 3, pp. 355–360, 1986.

[9] L. Mili, V. Phaniraj, and P. J. Rousseeuw, "Least median of squares estimation in power systems," *IEEE Trans. Power Syst.*, vol. 6, no. 2, pp. 511–523, 1991.

[10] T. Van Cutsem, M. Ribbens-Pavella, and L. Mili, "Hypothesis testing identification: A new method for bad data analysis in power system state estimation," *IEEE Trans. Power Appar. Syst.*, no. 11, pp. 3239–3252, 1984.

[11] L. Mili and T. Van Cutsem, "Implementation of the hypothesis testing identification in power system state estimation," *IEEE Trans. Power Syst.*, vol. 3, no. 3, pp. 887–893, 1988.

[12] M. K. Celik and A. Abur, "A robust WLAV state estimator using transformations," *IEEE Trans. Power Syst.*, vol. 7, no. 1, pp. 106–113, 1992.

[13] L. Mili, M. Cheniae, N. Vichare, and P. J. Rousseeuw, "Robust state estimation based on projection statistics [of power systems]," *IEEE Trans. Power Syst.*, vol. 11, no. 2, pp. 1118–1127, 1996.

[14] M. Irving, "Robust state estimation using mixed integer programming," *IEEE Trans. Power Syst.*, vol. 23, no. 3, pp. 1519–1520, 2008.

[15] Y. Chen, J. Ma, P. Zhang, F. Liu, and S. Mei, "Robust state estimator based on maximum exponential absolute value," *IEEE Trans. Smart Grid*, vol. 8, no. 4, pp. 1537–1544, 2015.

[16] A. S. Dobakhshari, S. Azizi, M. Abdolmaleki, and V. Terzija, "Linear lav-based state estimation integrating hybrid scada/pmu measurements," *IET Gen., Transm. & Distrib.*, vol. 14, no. 8, pp. 1583–1590, 2020.

[17] A. S. Dobakhshari, M. Abdolmaleki, V. Terzija, and S. Azizi, "Robust hybrid linear state estimator utilizing scada and pmu measurements," *IEEE Trans. Power Syst.*, 2020.

[18] H. M. Merrill and F. C. Schweppe, "Bad data suppression in power system static state estimation," *IEEE Trans. Power Appar. Syst.*, no. 6, pp. 2718–2725, 1971.

[19] J. Zhao, S. Fliscounakis, P. Panciatici, and L. Mili, "Robust parameter estimation of the French power system using field data," *IEEE Trans. Smart Grid*, vol. 10, no. 5, pp. 5334–5344, 2018.

[20] D. A. Belsley, E. Kuh, and R. E. Welsch, *Regression diagnostics: Identifying influential data and sources of collinearity*. John Wiley & Sons, 1980.

[21] R. D. Cook, "Detection of influential observation in linear regression," *Technometrics*, vol. 19, no. 1, pp. 15–18, 1977.

[22] S. Chatterjee and A. S. Hadi, *Sensitivity analysis in linear regression*. John Wiley & Sons, 2009.

[23] A. L. Da Silva, M. Do Coutto Filho, and J. Cantera, "An efficient dynamic state estimation algorithm including bad data processing," *IEEE Trans. Power Syst.*, vol. 2, no. 4, pp. 1050–1058, 1987.

[24] M. B. Do Coutto Filho and J. C. S. de Souza, "Forecasting-aided state estimation-part I: Panorama," *IEEE Trans. Power Syst.*, vol. 24, no. 4, pp. 1667–1677, 2009.

[25] M. B. Do Coutto Filho, J. C. S. de Souza, and R. S. Freund, "Forecasting-aided state estimation-part II: Implementation," *IEEE Trans. Power Syst.*, vol. 24, no. 4, pp. 1678–1685, 2009.

[26] G. A. Seber and A. J. Lee, *Linear regression analysis*. John Wiley & Sons, 2012.

[27] D. Singh, R. Misra, V. Singh, and R. Pandey, "Bad data pre-filter for state estimation," *Int. Jour. Electr. Power & Energy Syst.*, vol. 32, no. 10, pp. 1165–1174, 2010.

[28] M. D. C. Filho, J. S. De Souza, and M. T. Schilling, "Handling critical data and observability," *Elect. Power Compon. and Syst.*, vol. 35, no. 5, pp. 553–573, 2007.

[29] K. C. Sou, H. Sandberg, and K. H. Johansson, "Computing critical k-tuples in power networks," *IEEE Trans. Power Syst.*, vol. 27, no. 3, pp. 1511–1520, 2012.

[30] M. B. Do Coutto Filho, J. C. S. de Souza, and J. E. V. Tafur, "Quantifying observability in state estimation," *IEEE Trans. Power Syst.*, vol. 28, no. 3, pp. 2897–2906, 2013.

[31] A. A. Augusto, M. B. Do Coutto Filho, J. C. S. de Souza, and M. A. R. Guimaraens, "Branch-and-bound guided search for critical elements in state estimation," *IEEE Trans. Power Syst.*, vol. 34, no. 3, pp. 2292–2301, 2018.