Robust State-Estimation Procedure using a Least Trimmed Squares Pre-processor

Yang Weng, Student Member, IEEE and Rohit Negi, Member, IEEE and Qixing Liu, Student Member, IEEE and Marija D. Ilić, Fellow, IEEE

Abstract—Based on real-time measurements, Static State Estimation serves as the foundation for monitoring and controlling the power grid. The popular weighted least squares with largest normalized residual removed, gives satisfactory performance when dealing with single or multiple uncorrelated bad data. However, when the bad data are correlated or bounded, this estimator has poor performance in detecting bad data, which leads to erroneous deleting of normal measurements. Similar to the Least Trimmed Squares (LTS) method of robust statistics, this paper considers a state estimator built on random sampling. However, different from previous robust estimators, which stop after estimation, we regard the LTS estimator as a pre-processor to detect bad data. A subsequent post-processor is employed to eliminate bad data and re-estimate the state. The new method has been tested on the IEEE standard power networks with random bad data insertions, showing improved performance over other proposed estimators.

I. INTRODUCTION

An electric power grid is a complex network, composed of generation, transmission and distribution systems. Two fundamental problems make power grid operations challenging. Firstly, it must monitor the voltages and powers at various buses. Then, based on the monitoring results, proper control must be applied to maintain stability of the power grid.

In today’s interconnected power grid, Supervisory Control and Data Acquisition (SCADA) systems, which collect real-time data to feed a state estimator, are key components for monitoring the power grid. State estimation is defined as the procedure of obtaining the complex phasor voltages at all buses of the grid since these are sufficient to determine the operation condition of the grid. In practice, direct measurement of complex phasor voltages at every bus is currently very expensive. Thus, state estimation uses a redundant measurement set, including several bus voltages, bus real and reactive power injections, and branch reactive power flows. (Increased deployment of Phasor Measurement Units (PMU) will also allow direct measurement of the bus phasor).

Yang Weng is with the Department of Electrical and Computer Engineering, Carnegie Mellon University, Pittsburgh, PA, 15213 USA e-mail: yangweng@andrew.cmu.edu
Rohit Negi is a professor at Carnegie Mellon University, Pittsburgh, PA, with the department of Electrical and Computer Engineering, Carnegie Mellon University, Pittsburgh, PA, 15213 USA e-mail: negi@ece.cmu.edu
Qixing Liu is with the Department of Electrical and Computer Engineering, Carnegie Mellon University, Pittsburgh, PA, 15213 USA e-mail: lqx@cmu.edu
Marija D. Ilić is a professor at Carnegie Mellon University, Pittsburgh, PA, with a joint appointment in the Electrical and Computer Engineering and Engineering and Public Policy Departments. She is also the Honorary Chaired Professor for Control of Future Electricity Network Operations at Delft University of Technology in Delft, The Netherlands e-mail: milic@ece.cmu.edu

The state estimation was initially formulated as a Weighted Least Square (WLS) error problem, due to the assumption of Gaussian background noise. Although a WLS solution is easy to obtain, it is suboptimal when bad data appear. Bad data refers to gross (non-Gaussian) measurement error. Therefore, dealing with the Gaussian noise as well as bad data is important. Bad data occurs due to equipment failure, finite accuracy, infrequency of instrument calibration and measurement scaling procedure at the control center. Besides, telecommunication errors and incorrect topology information may also cause bad data.

Some bad data are easy to spot, such as a voltage magnitude several orders larger than the expected value. However, not all bad data are easy to detect. To test the existence of bad data in the measurement set, power system engineers employ the chi-square test. Upon the failure to pass the test, the measurement with the largest normalized residual is eliminated [1], [2]. This testing-eliminating procedure continues until the test is passed 1. It can be shown that, if there is only an error in the measurement set, the maximum residual corresponds to the error with the same index. This Largest Normalized Residual Remover (LNRR) works well if there are independent uncorrelated bad data. However, as the assumption of independence does not hold in general, when correlated bad data does appear, the LNRR may fail. Furthermore, LNRR does not work satisfactorily when the multiple bad data are bounded with respect to the Gaussian noise standard deviation. These drawbacks indicate a need for an improved bad data removal method.

Several methods to deal with bad data are known in the literature [2]–[10]. In particular, Handschin et al. [3] introduced a grouped residual search strategy that can remove all suspected bad data at once. In [4], a non-quadratic state estimator which minimizes the sum of absolute value of residual was studied and shown to have the ability to reject bad data. In [5], Jeu-Min et al. proposed a bad data suppression method, based on adapting the covariance matrix to the residuals. In [6], a method, based on hypothesis testing identification (HTI) was derived. However, as the initial selection of suspected bad measurements is based on normalized residual, the HTI is inefficient if some bad data result in small normalized residuals. [7] did a survey which summarizes three classes of bad data (BD) identification methods, namely, the classes of identification by elimination (IBE), the non-
quadratic criteria (NQC) and hypothesis testing identification (HTI). Combinatorial Optimization Identification (COI) [9], [10] was developed based on the framework of decision theory to identify multiple interacting bad data. For example, Asada et al. [9] proposed an intelligent tabu-based search strategy for multiple interacting bad data identification.

The effectiveness of the estimators above are different when detecting different types of bad data, and there is no universal solution. In the practical setting, with both Gaussian background noise as well as bad data, no strategy can be proved to be optimal.

Robust Estimation procedure from statistics deal with bad data in an alternative manner. They target a large number of bad data based on a purely statistical definition of “break down point”. This is defined as the proportion of arbitrarily large observations that an estimator can handle, before giving an arbitrarily large estimation error [11]. Least Median Square (LMS) [11], [12] and Least Trimmed Square (LTS) [13] are two typical robust estimators in terms of high breakdown point. LMS is achieved by minimizing the median of the residual, while the LTS minimizes the sum of the smallest $k$ squared residuals. Recently, a robust state estimation procedure based on Maximum Agreement between Measurements was also proposed [14].

Different from the previous papers, which deal with arbitrary large bad data, our result focuses on bad data that are comparable to the Gaussian noise (3 to 40 standard deviations) which is common in power systems. In such a scenario, the intuitive idea that a large residual corresponds to large bad data at the corresponding index would be seriously violated. This is because the linear combination of the Gaussian noise and the multiple bad data would disturb the ordering of the residuals. Therefore, in this paper, we use LTS as a pre-processor, which does random sampling and trimmed squares minimization. The LTS estimate provides a coarse map of the bad data locations, which can be used to discard some measurements. Then, a new estimate can be found as a WLS solution of the remaining measurement set.

We validate our method through simulation using the standard IEEE 39-bus system [15], [16]. The simulation result demonstrates the proposed estimator’s ability to handle bad data.

The rest of the paper is organized as follows. Section II introduces the system model and the classical estimator. Section III shows the main algorithm; Section IV presents the simulation result and Section V concludes the paper.

II. PRELIMINARIES

A. Model

The state estimation problem considers the power system model of $N$ buses as,

$$z = h(x) + e,$$  \hspace{1cm} (1)

where $z = (z_1, z_2, \cdots, z_m)^T$, $x = (x_1, x_2, \cdots, x_n)^T$ and $e = (e_1, e_2, \cdots, e_m)^T$ denote the measurement vector, state vector and measurement error vector respectively, with $z_i, x_j, e_k \in \mathcal{R}$ for $i = 1, 2, \cdots, m$, $j = 1, 2, \cdots, n$ and $k = 1, 2, \cdots, m$. Here, $m$ is the number of measurement, while $n$ is the number of state. The measurement set is usually redundant, to guarantee the observability of the system (i.e. $m > n$). Further, the state vector, can be written in terms of bus voltage magnitudes and phase angles, as $x = (\delta_1, \delta_2, \cdots, \delta_N, |V|_1, |V|_2, \cdots, |V|_N)^T$, with $n = 2N$. The function $h(\cdot)$ includes the network topology, and parameters. A state estimator is a function $\hat{x} = G(z)$, which provides an estimate $\hat{x}$ of the state $x$ using (1). However, the model (1) is computationally expensive for state estimation in large power systems due to the non-linear $h(\cdot)$. Therefore, a linearized DC power flow model, which considers only the real power is often used as an approximation [17], [18]. Below, we consider the DC power flow linear model.

$$z = Hx + e$$  \hspace{1cm} (2)

where $H = (h_{i,j})_{m \times n}$ is the measurement Jacobian, describing the linear relationship between the observation vector $z$ and the state vector $x$. We assume that the matrix $H$ is full ranked, so that state is observable. In analysis, three statistical criteria are commonly used for state estimation: the maximum likelihood criterion, the weighted least-square criterion, and the minimum variance criterion [19]. When the meter error is assumed to be independent, normally distributed with zero mean, these criteria all result in the estimator,

$$\hat{x} = (H^T R^{-1} H)^{-1} H^T R^{-1} z$$  \hspace{1cm} (3)

where $R$ is a diagonal matrix of $(\sigma^2_i)$ (noise variances of measurements.)

B. Bad data processing with LNRR

When bad data, superposed on the Gaussian noise, appears, the estimator (1) is no longer optimal. Therefore, techniques for bad data detection have been developed to improve state estimation [8], [19], in such scenarios. We discuss the LNRR technique.

To determine whether a measurement set is contaminated with bad data, engineers employ the statistical chi-square test. Intuitively, if the background noise is small, the sum of squared differences between the estimated measurement and the true measurement will be small. Rigorously, if the background noises follow the Gaussian distribution, it can be mathematically shown that $L(\hat{x}) = ||r||^2_2$ follows a $\chi^2(\nu)$-distribution with $\nu = m - n$ degrees of freedom, where $r = z - Hx$ is the "residual". Therefore, for a suitable $\tau = \tau(\alpha)$, the comparison test $L(\hat{x}) \geq \tau$ detects bad data, with a false alarm probability of $1 - \alpha$.

If bad data has been detected in the measurement set, removing the bad data is the next phase for LNRR state estimator. The intuition for the technique comes from analyzing the residual as follows [20].
where \( K = \mathbf{H}(\mathbf{H}^T \mathbf{R}^{-1} \mathbf{H})^{-1} \mathbf{H}^T \mathbf{R}^{-1} \) is a projection matrix, \((I - K) \cdot \mathbf{H} = \mathbf{0}\) and the projection matrix \( S = I - K \) is called the residual sensitivity matrix. Now, for mean zero Gaussian noise,

\[
E(r) = E(S \cdot e) = S \cdot E(e) = 0
\]

\[
Cov(r) = \Omega = E[rr^T] = S \cdot E(ee^T) \cdot S^T = SR
\]

LNRR analysis assume that bad data only occurs in one measurement \( k \), i.e. \( e_j = 0, j \neq k \), and that Gaussian noise is negligible. Using (4), the normalized residual for the erroneous measurement \( k \) can be shown to be the largest among all normalized residuals, as below,

\[
r_j^N = \frac{S_{jk} \cdot e_k}{\sqrt{R_{jj}} \sqrt{S_{jj}}} = \frac{\Omega_{jk} \cdot e_k}{\sqrt{\Omega_{jj} R_{kk}}} \leq \frac{\sqrt{\Omega_{jk}} \cdot e_k}{\sqrt{\Omega_{jj} R_{kk}}} = \frac{S_{kk} \cdot e_k}{\sqrt{R_{kk} S_{kk}}} = r_k^N
\]

Therefore, by removing the measurement corresponding to the largest normalized residual, we can eliminate the bad data. However, the proof shows that, it is hard to justify the LNRR method, when multiple bad data have magnitude of only 3 to 40 standard deviations of the Gaussian noise. In the next section, we present an algorithm, to deal with this scenario.

III. LTS WITH AGREEMENT REMOVER (LTSR)

A. Proposed algorithm

We propose an algorithm called LTSR for state estimation, in the presence of multiple bad data. The algorithm has two phases: Phase 1 (Step 1, 2 and 3) detect the location of bad data, using the LTS algorithm as a pre-processor and the agreement method [14] for detection. Phase 2 estimate the state using the remaining (good) measurements.

**Algorithm:** State Estimation with LTSR

- In Step 1, instead of using a subset of size \( n \) as suggested in [13], we use \( n \) measurements for WLS. As noted in the Appendix A, usually \( m/n \) is approximated around 4. Thus, choosing \( m/2 \) measurements results in an observable system, with high probability. By using more measurements than \( n \), the estimate is less affected by Gaussian noise.
- In Step 2, the LTS estimate, which minimizes the sum of the \( v \) smallest (ordered) squared residuals \( \sum_{j=1}^{v} r^2(j) \) is found, where
  \[
v = \lceil (m + n + 1)/2 \rceil
\]
- In Step 3, bad data is detected using the residual as,
  \[
  \frac{|r_k|}{\sigma_k} > t \quad \forall k \in [1, \ldots, m].
\]
  where \( r = z - \mathbf{H}\hat{x}_{\text{lt}} \) is the estimated residual from step 2 and \( \sigma_k \) is the standard deviation of the Gaussian noise in the \( k \)-th measurement. \( t \) can be chosen as 5, so that residuals affected by only Gaussian noise are not removed.
- In Step 4, we assume that bad data have been removed, so that WLS can be used for state re-estimation.

**Remark** The LTS with agreement remover method proposed here attempts to tradeoff the impact of Gaussian noise and the bad data through the use of threshold \( t \).

B. Example

In this part, we present the example of a 3 bus system from [12] in Figure 1, with a DC model. All lines are assumed to have zero resistance and 1.0 p.u. reactance, except for Line 1 - 3, which has a susceptance of 5 p.u.. \( \theta_3 = 0 \) for the reference bus. The linear system model can be represented as,

\[
\mathbf{z}_i = l_{i1}\theta_1 + l_{i2}\theta_2 + e_i \quad \text{where} \quad \mathbf{H} = \begin{bmatrix} 1 & -1 & 1 & 0 & -1 & -1 \\ -1 & 1 & 0 & 1 & 2 & -1 \end{bmatrix}
\]

Due to the linearity of the model, w.l.o.g., we can assume that \( x = 0 \). The standard deviations \( \sigma_i \) of Gaussian noise in each measurement set is 0.01 p.u. Two measurements \#3 and \#6 are assumed to be contaminated with error values of -0.55 p.u. and 1.0 p.u., respectively. Since these are much larger than...
\[ \sigma_i \], these two measurements are bad data. The state estimation must estimate the two bus voltage angles \( \theta_1 \) and \( \theta_2 \) from a set of 6 real power measurements. As two bad data mask each other in both \( W_{ii} \) and \( r_N \) values (see Table I), LNRR fails to identify them, since LNRR successively removes measurement \#4 and \#5 instead. As noted in [7], this wrong solution is found as the optimal one by the combinatorial optimization method [10], as well as by the geometric method [21]. Further, recall, that the HTI method fails if any of the bad data are missed [7].

<table>
<thead>
<tr>
<th>Meas. #</th>
<th>( W_{ii} )</th>
<th>( r_N ); 1st</th>
<th>( r_N ); 2nd</th>
<th>( r_N ); 3rd</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>0.90</td>
<td>3.42</td>
<td>2.90</td>
<td>1.54</td>
</tr>
<tr>
<td>2</td>
<td>0.90</td>
<td>-3.42</td>
<td>-2.90</td>
<td>-1.54</td>
</tr>
<tr>
<td>3</td>
<td>0.36</td>
<td>3.27</td>
<td>2.11</td>
<td>-2.39</td>
</tr>
<tr>
<td>4</td>
<td>0.96</td>
<td>8.32</td>
<td></td>
<td></td>
</tr>
<tr>
<td>5</td>
<td>0.74</td>
<td>5.70</td>
<td>6.75</td>
<td></td>
</tr>
<tr>
<td>6</td>
<td>0.14</td>
<td>6.06</td>
<td>3.40</td>
<td>-2.39</td>
</tr>
</tbody>
</table>

As opposed to the above methods, our LTSR method works correctly. In LTSR, the random sampling is used to obtain 15 sample sets of measurements. Of these, 9 sets will contain bad data. In step 2, 15 state estimates are obtained. LTS minimizes the sum of squares of the smallest \( v = \lfloor (m + n + 1)/2 \rfloor = 4 \) residuals, for each of these estimates. For the 9 estimates computed by bad data, the LTS metric is nonzero because only two measurements result in zero residual. For the 6 sets, which are free of bad data, the correct state is found and so, only 2 residual are nonzero (Those with bad data). Thus, the trimming to \( r = 4 \) in the LTS metrics result in a zero value. So, picking the least of these metrics, correctly identifies the 6 sets free of bad data. Thus, in step 3, bad data are identified and correctly removed. In step 4, the correct state estimate is found.

C. How many samples are enough?

The new method is combinatorial and requires a prohibitive computation time if all possible sample sets are examined\(^2\). However, as seen from the example, it is sufficient to find only one (or a few) set which are free of bad data, among the many that have that feature.

For a general network, with number of bad data \( s \ll m \), the average number of sample sets needed can be approximated as

\[
\frac{m^s}{(m-s)^2} \approx 2^s
\]

(The distribution of good sets is geometrically distributed with mean \( 2^s \).) Since better performance is obtained with more sample sets, we suggest using around \( 3 \times 2^5 \) sample sets in Step 1.

IV. NUMERICAL RESULT

Simulation study was carried out on the IEEE 39-bus system shown in Figure 2. The data was pre-processed by using the MATLAB Power System Simulation Package (MATPOWER) [15] [16]. This involves solving the power flow equations in MATPOWER, which produces the true value of the static system state (voltage magnitudes and phase angles). Then, with the computed state and the system matrix \( H \), we compute the error free measurements \( Hx \). These measurements include transmission line apparent power injection from (and to) each bus that it connects, the apparent power on each bus and the direct voltage magnitude and phase angle measurements of each bus. Next, we generated the Gaussian noise, based on the variance provided by MATPOWER. It is assumed that the voltage magnitude and phase angle measurements from PMUs are more accurate than the other conventional meter measurements. Finally, the bad data is set to be 4 to 40 times the standard deviation of the Gaussian noise.

A. Simulation Results

In this simulation, besides our proposed LTSR estimator, several other state estimators such as LNRR, Maximum agreement [14] and Least Trimmed Square (without removal) are employed for comparison purpose. In LTS, Maximum agreement and LTSR methods, 1000 random sample sets of \( \frac{m^2}{2} \) measurements each, are obtained for computing the state
estimate candidates. A lower bound is also computed by assuming that an oracle provides WLS with the locations of the bad data. We randomly insert bad data (numbers $s = 1$ to $12$) to measurements. For each value of $s$, 100 bad data insertions were created to obtain the mean performance. The variance of the bad data was assumed to be 20 times that of the Gaussian noises. For the LTSR method, the thresholds is set at $t = 5$.

The performance measure is the mean square error per state as below:

$$
\epsilon_{\text{state}} = \frac{1}{n} \left\{ \sum_{l=1}^{n/2} E[|V_{l,\text{est}} - V_{l,\text{true}}|^2] + \sum_{l=1}^{n/2} E[|\delta_{l,\text{est}} - \delta_{l,\text{true}}|^2] \right\}
$$

where $V_{\text{est}}$ and $V_{\text{true}}$ are the estimated and true voltage magnitudes, and $\delta_{\text{est}}$ and $\delta_{\text{true}}$ represent the estimated and true voltage phases.

Figure 3 is the performance comparison for the 39 bus case. The diamond dashed line is the lower bound, where an oracle points out the bad data location. The dashed line is the performance of LNRR. Notice that its performance is far from the lower bound. The dashed square line represents the method of Maximum agreement, and the dashed circle line represents the LTS. These two estimators are more robust to bad data, but their performance does not meet the lower bound. The solid line shows the performance of the proposed LTSR method. When the number of bad data $s$ is smaller than 6, the LTSR line is close to the lower bound. This is because we use $1000 \approx 2^{10}$ sample set for LTS, so that the probability of hitting upon a good set is large. Simulation shows (not prescribed here), that larger values of $s$ can be handled by increasing the number of sample sets.

### B. Computing Time

The proposed estimator was computed on a desktop computer with AMD Athlon 64 bit Dual Core Processor and 4.00 GB RAM. The state estimate computation time of LTSR is shown in Table II.

#### Table II

<table>
<thead>
<tr>
<th>Power grid</th>
<th>State Computation Time</th>
</tr>
</thead>
<tbody>
<tr>
<td>39 bus</td>
<td>0.1923s</td>
</tr>
</tbody>
</table>

The computation of matrix pseudo-inverse, required by WLS, is not considered, because it is assumed that the grid topology changes much slower compared to the change of state. Thus, these inverses can be computed offline and stored. Storage requires memory which is shown in Table III.

#### Table III

<table>
<thead>
<tr>
<th>Power grid</th>
<th>Required Memory</th>
</tr>
</thead>
<tbody>
<tr>
<td>39 bus</td>
<td>112,083 KB</td>
</tr>
</tbody>
</table>

### V. Conclusions

In this paper, a static state estimation method with bad data processing capability is introduced. By techniques such as random sampling and bad data filtering, the influence of bad data are largely eliminated. In addition, weighted least square estimation is employed after bad data has been eliminated. Numerical results show that this method performs better than LNRR, Maximum Agreement and pure Least Trimmed Square, often coming close to the lower bound.

### Appendix

#### A. Number of measurements used in Matpower

In the state estimation package of Matpower [15] [16], Table IV lists the number of states $n$ and number of measurements $m$.

#### Table IV

<table>
<thead>
<tr>
<th>Number of buses</th>
<th>$n$</th>
<th>$m$</th>
</tr>
</thead>
<tbody>
<tr>
<td>4</td>
<td>8</td>
<td>32</td>
</tr>
<tr>
<td>9</td>
<td>18</td>
<td>72</td>
</tr>
<tr>
<td>14</td>
<td>28</td>
<td>136</td>
</tr>
<tr>
<td>30</td>
<td>60</td>
<td>284</td>
</tr>
<tr>
<td>39</td>
<td>78</td>
<td>340</td>
</tr>
<tr>
<td>57</td>
<td>114</td>
<td>548</td>
</tr>
<tr>
<td>118</td>
<td>236</td>
<td>1216</td>
</tr>
<tr>
<td>300</td>
<td>600</td>
<td>2844</td>
</tr>
</tbody>
</table>

Thus, $m/n \approx 4$ in these cases.

### Acknowledgment

This work was supported in part by US NSF awards 0931978, 0831973 and 0347455.
REFERENCES


Yang Weng (S’08) received his B.E. in Electrical Engineering in 2006 from Huazhong University of Science and Technology, Wuhan, China. and a M.S. in Statistics from University of Illinois at Chicago in 2009. Currently, he is a graduate student in the Department of Electrical and Computer Engineering at Carnegie Mellon University, Pittsburgh, PA. His research interest includes power systems, smart grid, information theory and wireless communications.

Rohit Negi (M’00) received the B.Tech. degree in Electrical Engineering from the Indian Institute of Technology, Bombay, India in 1995. He received the M.S. and Ph.D. degrees from Stanford University, CA, USA, in 1996 and 2000 respectively, both in Electrical Engineering. He has received the President of India Gold medal in 1995. Since 2000, he has been with the Electrical and Computer Engineering department at Carnegie Mellon University, Pittsburgh, PA, USA, where he is a Professor. His research interests include network analysis, communications systems, information theory.

QiXing Liu (S’09) received his B.Sc. in Electrical Engineering in 2007 from Zhejiang University, Hangzhou, China. Currently, he is a graduate student in the Department of Electrical and Computer Engineering at Carnegie Mellon University, Pittsburgh, PA. His research interest includes smart grid, wide-area control and regulation of large-scale electric power systems.

Marija D. Ili´c (M’80-SM’86-F’99) is currently a Professor at Carnegie Mellon University, Pittsburgh, PA, with a joint appointment in the Electrical and Computer Engineering and Engineering and Public Policy Departments. She is also the Honorary Chaired Professor for Control of Future Electricity Network Operations at Delft University of Technology in Delft, The Netherlands. She was an assistant professor at Cornell University, Ithaca, NY, and tenured Associate Professor at the University of Illinois at Urbana-Champaign. She was then a Senior Research Scientist in Department of Electrical Engineering and Computer Science, Massachusetts Institute of Technology, Cambridge, from 1987 to 2002. She has 30 years of experience in teaching and research in the area of electrical power system modeling and control. Her main interest is in the systems aspects of operations, planning, and economics of the electric power industry. She has co-authored several books in her field of interest. Prof. Ili´c is an IEEE Distinguished Lecturer.