EFFECTIVENESS OF NORMALIZED LARGEST RESIDUAL TEST BASED ON BAD DATA IDENTIFICATION PROCESS

[1]MAMAN AHMAD KHAN, [2]SHAHBAZ KHAN

[1,2]Aligarh Muslim University
E-mail: [1]maman.ahmadkhan@gmail.com, [2]Khanshahbaz13@gmail.com

Abstract— The purpose of this paper is to implement a computational program to estimate the states of a power system and shows the relationship between the state estimation accuracy and the effectiveness of bad data identification process. Weighted least square (WLS) is the method chosen. In state estimation once the states are estimated the error is analyzed to detect and identify the measurements which contain errors. Detection is done with the help of chi-square test and identification process is done with largest normal residual test. This paper shows that on the application of largest normal residual test the bad data is not identified for small number of iterations.

Index Terms— State Estimation, Bad Data, Chi-square Test, Largest Normalized Residual Test, Normal Residual, Tolerance.

I. INTRODUCTION

The measured values in power system measurement system can include a very high range of error due to the reasons including disorder in communication system and auxiliary devices such as current transformer, potential transformer etc. and also because of incorrect installation and servicing [1]. However, the main source of bad data is the errors arising due to the communication network as the errors arising from the measurement devices have low magnitude and are comparable to the residuals and hence may go undetected. These measurements can affect the state estimation of power system which is an important component of power system monitoring modern energy management system (EMS) for as it convert the redundant measurements into reliable estimates of the state of an interconnected electric power system [2].

These errors of large magnitude (i.e. those arising from bad communication etc.) are known as bad data in power system which contains large error beyond typical confidence interval of random Gaussian noise. These bad data has to be removed as it affects the reliability of the estimated data. Bad data processing generally contains two procedures; one is the bad data detection which detects the presence of false data and the other is bad data identification which helps in finding which measurement contains the bad data [3].

Many mathematical techniques are available for detection and identification of bad data. In this paper we have use the chi-square test for the detection and large normalized residual test for identification purpose. It is shown that the state estimation accuracy itself improves the bad data identification.
The bad data identification is applied on IEEE 14 bus system and an inference is drawn regarding the dependency of the effectiveness of identification process on the accuracy of the state estimation process.

II. BAD DATA DETECTION

The process of state estimation works well when the noise i.e. error in measurements are simple Gaussian Errors. While some errors are obvious and can be easily detect and identified from the estimation process with the help of simple verification of the measured input data, some errors are not easily detectable and require different approach and methodologies [4]. Some of the reasons reported in literature are[5] [6] :

i) Low redundancy measurement Gaussian errors.
ii) Interactive multiple GEs
iii) The leverage point measurements.

Out of the number of methodologies reported for the identification of errors other than simple GEs, the WLS estimator associated with chi-square test and the largest normal residual test are the most widely used tests in the detection and identification process because of its simplicity and the ease of implementation on computer.

III. FORMULATION OF BAD DATA PROCESS

In the formulation first of all the WLS state estimation technique is carried out as given in [7]. Then, the bad data detection as well as identification process is been performed.

Detecting the presence of bad data is typically accomplished in this paper through the use of a Chi-square test. The measurements are assumed to be a function of the true network state and corrupted by additive white Gaussian noise (AWGN). The distribution of the sum of squared errors between the measurements and the measurement estimates, as calculated from the state estimate, should therefore conform to a Chi-square distribution [8].
A significance level is chosen and applied to the Chi-square distribution to determine a threshold, for

the Chi-square hypothesis test. If the weighted sum of squared errors is less than the hypothesis, then the sum is consistent with the expected distribution given by the assumed AWGN. If the sum is greater than the null hypothesis, bad data is assumed to exist in the measurement vector.
The identification of the location of the bad data is more challenging. As we will see, the largest normal residual test will be used in identification process.
In power system state estimation, bad data refers to measurements that have large errors beyond typical confidence interval of random Gaussian noise. The processes are briefly explained as follows:

Chi-squares Test: Consider the estimation objective function [5]:

$$J(\hat{\theta}) = r^T R^{-1} r \quad (1)$$

Where
$r = z - H(\hat{\theta})$ is defined as the estimated residual vector.
Z is the measurement vector.
$\hat{\theta}$is the system state vector (having voltage and voltage angles).
H is Jacobean matrix formed as

$$H = \frac{\partial h}{\partial \hat{\theta}}$$

H is the theoretical relationship between the measurements and the system states$\hat{\theta}$.
R is covariance matrix of the measurement errors.
Since the measurement errors are normally distributed, the estimated objective function $J(\hat{\theta})$ obeys a chi-square distribution with m-n degrees of freedom, i.e. $J(\hat{\theta}) \sim \chi^2_{m-n}$ m and n represent the number of measurements and state variables, respectively.
Therefore, bad data will be detected if

$$J(\hat{\theta}) \geq \chi^2_{(m-n),p} \quad (2)$$

Whereχ^2chii square value and p is is the detection confidence probability. So when J is greater than chi-square then bad data is present in the measurement and if J is smaller the chi-square then bad data is absent.

2) The Largest Normalized Residual Test: Normalized residuals are used for identifying bad data after state estimation is performed. The measurement residual vector 'R' could be represented as:

$$R = z - H\theta = Se \quad (3)$$

where the residual sensitivity matrix S represents the relationship between the measurement residuals and the measurement errors:

$$S = I - HGH^T R^{-1} \quad (4)$$

where we define the gain matrix $G = HTR^{-1}H$. Therefore, normalized residual vector can be represented as:

$$r^N = \frac{|r|}{\sqrt{diag\,(SR)}} \quad (5)$$

The measurement corresponding to the largest normalized residual is considered as bad data and is to be eliminated for another round of state estimation.

IV. IMPROVING THE EFFECTIVENESS OF LNR TEST

Various observations made by recent researches pointed out some failures of LNR test in identifying the bad data [5]. When the bad data lies in the range of residual values the bad data identification is not possible. Residual values range indicates the inaccuracy of a state estimation algorithm for a given set of measurements. An accurate SE algorithm has the residuals in a narrower values range as compared to an inaccurate one. Evidently, the effectiveness of bad data identification process is also a function of SE accuracy. This is so because larger the residual range, greater is the probability that a bad data goes undetected as there may be some residuals larger than the bad data itself in this case.

Tolerance values and the number of iterations in state estimation process directly affect the converged values of the measurement residuals. This paper discusses the detection and identification of bad data for various tolerance values and the number of iterations in state estimation process. Next section contains a case study of the IEEE 14 bus system [9] where the points discussed above are implemented

V. CASE STUDY

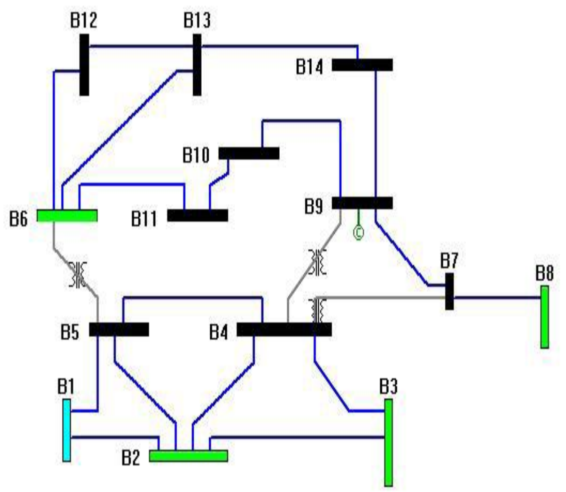

Figure 1: The IEEE 14 Bus Test System [9]

To test the performance of bad data detection and identification process, the standard IEEE 14 bus test system is used. With the help of the algorithm for bad data detection chi-square test is performed as given in section III. In this χ^2 (chi-square) is a value from table [10]taken at a confidence level (say 95%) for a given degree of freedom and is compared with J. Degree of freedom is m-n where m is number of measurement and n is number of states excluding the reference phase angle.

In this case 42 measurements have been taken and since voltage angle at bus 1 is taken as reference, number of states (voltage and voltage angles to be estimated is $(2 \times 14 - 1 = 27)$.

Thus the degree of freedom in this case is $m - n = 42 - 27 = 15$.

J is a value of summation of squares of random variables having normalized distribution function of Gaussian nature.

In this case J = summation of residuals $r = z - h$weighted by their respective covariance.

The test as described in (1-5) is carried out with bad data at bus 2 (Real power injection measurement P_2). The results for different iterations are checked and at each iteration the relation between the tolerance limit (i.e. $X_K - X_{K-1}$) and converged residual 'r' are compared. The process is repeated for various tolerance limits for such a comparison.

VI. RESULTS

After performing the chi-square test for the bad data detection we can see for the confidence level $\alpha = 0.010$ (confidence level 90%), J is 4.306 which is less than Chi-square value (i.e. 29.141) which implies that bad data is not present since $J(\hat{\theta}) \geq \chi^2_{(m-n),p}$ is not satisfied.

The measurement values and other parameters related to IEEE 14 bus test system can be found in [8]. Now when we incorporated the bad data then the chi-square test will detect the bad data and the largest normal residual test will identify the location of the bad data. For $\alpha = 0.010$ (confidence level 90.0%),as we increase the real power injection measurement at bus 2 in measurement data matrix (real power injection measurement P_2 is taken as 1.0820 instead of 0.0820). We can see that as J becomes greater than the chi-square value hence bad data is detected and also identified at bus 2 (by applying equations 1-5). Here J is 191.53 and chi square value is 29.141.

The estimated values of voltage and voltage angles at different buses are given in table 1.

Table 1: Estimated values of voltage and voltage angles

Bus No.	Voltage (pu)	Voltage Angles (Degrees)
1	1.0598	0
2	1.0436	-4.9874
3	1.007	-12.771
4	1.0118	-10.2644
5	1.0153	-8.7796
6	1.0688	-14.4804
7	1.0444	-13.2671
8	1.0788	-13.2671
9	1.0292	-14.8546
10	1.0286	-15.0709
11	1.0419	-14.8924
12	1.052	-15.3329
13	1.0454	-15.3669
14	1.0179	-16.0979

From the results we can observe that, the estimated values of the voltage and voltage angle at different buses does not change much below the tolerance level

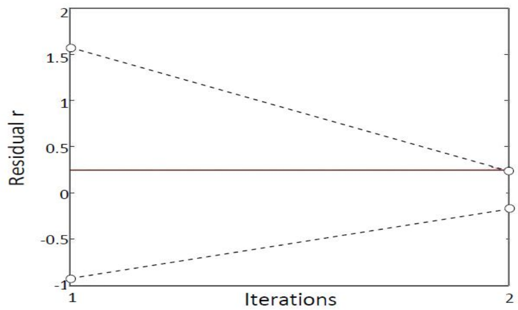

Figure a: For tolerance 1×10^{-1}

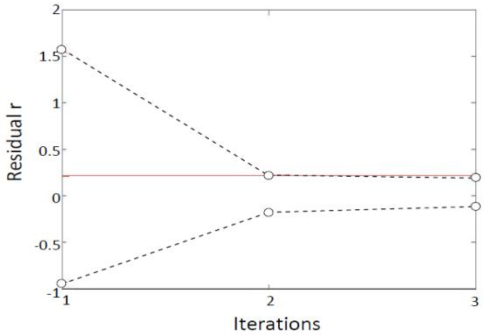

Figure b: For Tolerance 1×10^{-3}

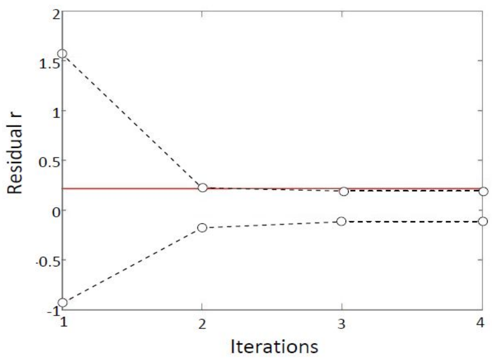

Figure c: For Tolerance 1×10^{-5}

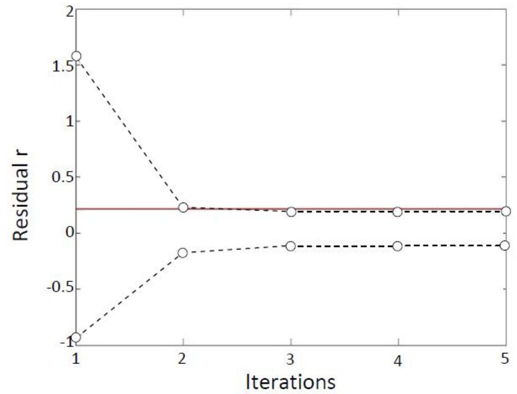

Figure d: For tolerance 1×10^{-7}

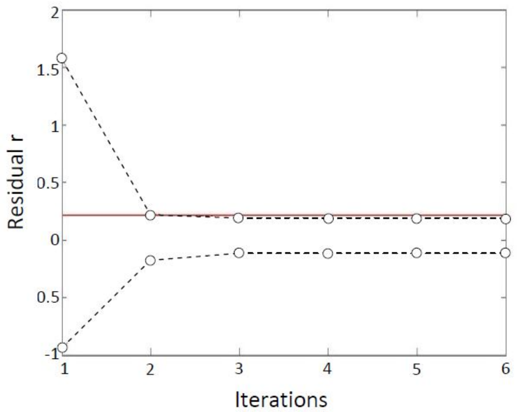

Figure e: For Tolerance 1×10^{-9}

These are the traces of the maximum and minimum values of the residuals (other than corresponding to bad data measurements) as the state estimation proceeds for different iterations. The dark line shows the bad data error. Here we see during the first iteration this bad data cannot be identified as it is less than the threshold (or under residual value range). However after the second iteration the residual range decreases and hence the bad data can be identified easily.

As the number of iterations increases the difference between two successive estimated state value $X_K - X_{K-1}$ (which is also called as tolerance) decreases.

Table 2: Bad Data Locations identified for various tolerances

Tolerance	Is Bad Data Location correctly Identified?	Location Identified (Bus No.)
10^{-1}	No	9
10^{-3}	No	17
10^{-5}	Yes	2*
10^{-7}	Yes	2*
10^{-9}	Yes	2*

*Bad data is at bus 2

Therefore as the required tolerance decreases, the correctness of bad data identification process in identifying the bad data location increases as evident from the table 2.

CONCLUSION

The paper presents relationship between the state estimation accuracy and the bad data identification process effectiveness. On the application of largest normalized residual, the bad data is not detected for small number of iterations 10^{-1}to10^{-4} . The simulation results shows that when the test is applied after the complete convergence of state estimation to a tolerance of 10^{-5} then only, the bad data are identified and when it is decreased to smaller than 10^{-5} range,the state estimation process for the identification works perfectly. This is so because as the tolerance decreases it narrowed the converged residual range also. Hence any bad data would have the residual far larger then this narrow range and is distinctly identified which would otherwise be comparable to the converged residual r values and might hence missed to be identified. Other possibilities for identifying the bad data for a given tolerance of the state estimator might be to utilize the random nature of Gaussian noise as against the bad data measurement value.

REFERENCES

[1] M. Tarafdarhagh, S.M Mahaeietol. “Improving bad data detection in state estimation of power system” IJECE, Vol 1 Number 2,2011,pp 85-92.

[2] F.C.Schwappe, J.Wildes and A.Bose “Power system static state estimation pare I,II,III”, IEEE Transaction on power apparatus and systems Vol 8,9 Number 1 pp 120-135, January 1970.

[3] Dye-Hyun Choi and Le-Xie “Fully distributed bad data processing for wide area state estimation” IEEE smartgrid com 2011.

[4] Abur and A. G. Exposito, Power systems stateestimation: theory and implementation, Marcel & Dekker Publishers, Nova York, USA, 2004.

[5] B.Carvalho and N.Bretas“Analysis of the largest normalized residual test robustness for measurements gross errors processing in the WLS state estimator” IEEE transaction on systemics, cybernetics and informatics, Volume 11,Number 7,2013.

[6] J.Monticelli, Electric power system state estimation,Proceedings of the IEEE, 88(2), 2000, 262-282.

[7] Fang Chen, XueshanHan,ZhiyuanPan,Li Han “State Estimation Model and AlgorithmIncluding PMU” IEEE conference on Mechatronic Science, Electric Engineering and Computer (MEC), 2011,page 114-117.

[8] Abur and A.G. Exposito. Power System State Estimation: Theory And Implementation. CRC Press, 2004.

[9] Power Systems Test Case Archive, Available at :http://www.ee.washington.edu/research/pstca/

[10] G. Phadke, "Synchronized phasor measurements in power systems",IEEE Computer Applications in Power, April 1993, pp. 10-15.

★★★