Fuzzy Linear Regression Approach for Uncertainty Modeling in Power System State Estimation

A. K. AL-Othman

N. H. Abbasy

Electrical Engineering Department Collage of Technological Studies Al-Rawda, 73452, P.O. Box 33198, Kuwait

Abstract: -A fuzzy linear state estimation model is employed, which is based on Tanaka's fuzzy linear regression model, for modeling uncertainty in power system state estimation. Both measurements uncertainty as well as parametric uncertainty is considered by fuzzy estimator. The uncertain measurements and the parameters are expressed as fuzzy numbers with a triangular membership function that has middle and spread value reflected on the estimated states. The proposed fuzzy model is formulated as a linear optimization problem, where the objective is to minimize the sum of the spread of the states, subject to double inequality constraints on each measurement. Linear programming technique is employed to obtain the middle and the symmetric spread for every state variable. The estimated middle corresponds to the value of the estimated states. Preliminary results from application of the proposed on regression and D.C problems are promising.

Key-Words: - State Estimation, Fuzzy Linear Regression, Fuzzy Linear State Estimator (FLSE) and Measurements Uncertainty

1 Introduction

Having an accurate picture of the state of a system is an important part of the system operations. While a simple SCADA system has the ability to provide the system operators with raw information about the system operation conditions, only a state estimator has the ability of filtering the information to supply a more accurate picture of the status of the system.

The conventional purpose of state estimation is to reduce the effect of measurement errors by utilizing the redundancy available in the measurement system. In particular, the objective is to reduce the variance of the estimates and improve their overall accuracy. The other major objectives of state estimation methods include: detection of gross errors, detection of invalid topological information and detection of model parameter errors.

If the inaccuracy in the measurements is modelled by some random probability distribution function, then the set of feasible estimates can also be modelled by a probability distribution function. In another word, such estimators are probabilistic in nature. Regrettably, the statistics of the observation errors are difficult to be characterized in practice. In addition, it is questionable that such imprecision in error modelling can not be equated with randomness [1], and that the main source of imprecision can be associated with fuzziness instead of randomness [2]. Fuzzy theory can very well be deployed to in such circumstances to overcome this limitation and address such uncertainty in the modelling of such statistics, due to its ability in handling uncertainties and vagueness associated with the observation errors. Fuzzy estimators are possibilistic in nature. If the observation errors are assumed to be fuzzy due to uncertainty that is inherently present in the system, then the estimates are assumed to be a range of possible values. In such situations, it is desirable to provide not just a single 'optimal' estimate of each state variable but also an uncertainty range within which we can be assured that the 'true' state variable must lie. This is attainable by utilizing some fuzzy function to represent the estimates as fuzzy estimates with their associated uncertainty ranges as apposed to crisp estimates (single point only) produced by the conventional estimators [3].

The main theme of this report is to model the uncertainties associated with the measured quantities in a way that defines an interval (range) with respect to their nominal values. The range is governed by the tolerance, of the measuring instrument (a quantification of accuracy usually provided by the manufacturer) and other factors that are known to have direct effects on network mathematical model being used in the estimation procedure. By utilizing the proposed fuzzy linear techniques, which is known to address uncertainties very well, the confidence interval (or bounds) of the state variables can be computed.

The present author proposes an estimator based on fuzzy linear regression formulation for estimating the uncertainty interval around the system state The uncertainty is expressed in both variables. measurements and network parameters in a unified fuzzy model. The main objective is to minimize the fuzziness in the estimated states. This can be achieved by minimize the sum of spreads of all fuzzy states, subject to double inequality constraints on each measurement to guarantee that the original membership is included in the estimated membership. Linear programming has been employed to obtain the middle and the symmetric spread of every state variable. The estimated middle corresponds to the value of the estimated state, whereas the symmetric spreads in the membership functions of the state variables represents the uncertainty interval around that estimated state. Thus, the primary goal is to minimize the sums of the uncertainties around the states.

2 Uncertainty and State Estimation

The uncertainty is a parameter associated with the measurement that describes the dispersion of the values that could reasonably be attributed to the measured quantity [4]. This uncertainty reflects the lack of complete knowledge of the exact value of the quantity being measured. Theoretically, availability of complete knowledge about the measured quantity requires an infinite amount of information, which is obviously impossible. Phenomena that contribute to the uncertainty are called sources of uncertainty. According to [4], the various possible sources of uncertainty in measurements, include: incomplete definition of the measured quantity, inadequate realization of the definition of the measured quantity, non-representative sampling (sample measured may not fully represent the measured quantity), incomplete knowledge of environmental conditions, human error in reading analogue instruments approximations incorporated in the measurement procedure and finite resolution of instrumentation.

The idea of an uncertainty range is recognizable in engineering practice, where the accuracy of a particular measurement is often described as (for example) plus or minus 2 percent, rather than by quantifying the standard deviation or variance. Schweppe [5] introduced the concepts of uncertainty in the general context of engineering analysis, estimation and optimization. In [5] the concept of unknown-but-bounded errors for modeling uncertainty in estimation problems was introduced. Measurements are assumed to be inexact and have errors that are unknown but fall within a bounded range.

Fuzzy theory has also been widely used in power system state estimation. For example, Shahidehpour and et al. in [18, 19] have utilized Fuzzy theory to handle the uncertainty in decision making and power purchasing in deregulated environment. In [20, 21] authors have applied fuzzy set for multi-area generation scheduling and for optimal reactive power control respectively. As for state estimation the concept of fuzzy-logic has been employed by Shabani et. al in [22] to improve the over all performance of the WLS estimator. A hybrid WLS and fuzzy-logic estimator was developed in [22] to model residual based on possibility theory. Shahidehpour and et al., on the other, have employed fuzzy sets in conjunction with LAV (Least Absolute value) estimator and LMS (Least Medium Squares) estimator to robustly eliminate the bad data in [23]. Further more, authors in [24] have developed a fuzzy LAV estimator based on maximizing the sum of individual memberships. This fuzzy LAV estimator out performed the standard WLAV in the presence of leverage point.

3 An overview of Tanaka's Fuzzy linear regression

Fuzzy linear regression was introduced by Tanaka et. al [25] in 1982. The general form of Tanaka's formulation is given by:

 $Y_{\sim} = f(x) = A_0 + A_1 x_1 + A_2 x_2 + \dots + A_n x_n = Ax$ (1)

where Y_{i} is output (dependant fuzzy variable), $\{x_{1}, x_{2}, \dots, x_{n}\}$ is a non fuzzy set of crisp independent parameters and $\{A_{0}, A_{1}, \dots, A_{n}\}$ is a fuzzy set of symmetric members, unknowns, needs to be estimated. Each fuzzy element in that set may be represented by a symmetrical triangular membership function, shown in figure 1, defined by a middle and a spread values, p_{i} and c_{i} respectively. The middle is known as the model value and the spread denotes the fuzziness of that model value. The triangular membership function can be expressed as: Therefore, since $A_i = (p_i, c_i)$, then equation (1) may be rewritten as:

$$\mu_{\underline{A}_{i}}(a_{i}) = \begin{cases} 1 - \frac{\left| p_{i} - a_{i} \right|}{c_{i}} , p_{i} - c_{i} \leq a_{i} \leq p_{i} + c_{i} \\ 0 , otherwise \end{cases}$$
(2)

Therefore, since $A_i = (p_i, c_i)$, then equation (1) may be rewritten as:

$$Y_{2} = f(x) = (p_{0}, c_{0}) + (p_{1}, c_{1})x_{1} + \dots + (p_{n}, c_{n})x_{n}$$
(3)

The membership function of output Y_{\sim} may be given by:



Figure 1 membership function of fuzzy coefficient A

Now, by substituting equation (3) in (4), the output membership function is given as:

$$\mu_{\underline{Y}}(y) = \begin{cases} 1 - \frac{\left| \begin{array}{c} y - \sum_{i=1}^{n} p_{i} x_{i} \right|}{\sum_{i=1}^{n} c_{i} \left| x_{i} \right|} & , x_{i} \neq 0 \\ 1 & , x_{i} = 0, y_{i} = 0 \\ 0 & , x_{i} = 0, y_{i} \neq 0 \end{cases}$$
(5)

The output membership function is depicted in figure 2.

From regression point of view, equations (1-5) may be applied to *m* samples where the output can be either non-fuzzy, (certain or exact), in which no assumption of ambiguity is associated with the output or fuzzy (uncertain), where uncertainty in the out is involved due to human judgment or meters impression [20]. In this study both non fuzzy and fuzzy output will be considered.



Figure 2 membership function of output

3.1 Non- fuzzy output model [19]:

In this model, Tanaka converted regression model into a linear programming problem [19]. In this case the objective is to solve for the best parameters, i.e. A^* , such that the fuzzy output set is associated with a membership value grater than h as in ;

$$\mu_{Y_{j}}(y_{j}) \geq h, \qquad j = 1, \dots, m \tag{6}$$

where $h \in [0,1]$ is the degree of the fuzziness and is normally defined by the user.

Therefore, with equation (6) as a condition, the main objective is to find the fuzzy coefficients that minimize the spread of all fuzzy output for all data set. Note that the fuzziness in the output is due to fuzziness assumed in the system structure A^* .

Thus, given non-fuzzy data (y_i, x_i) , the fuzzy parameters $A^* = (p, c)$ may be solve for by the linear programming formulation as:

$$F_{non-fuzzy} = \min(\sum_{j=1}^{m} \sum_{i=1}^{n} c_i x_{ij})$$
(7)

Subject to:

$$y_{j} \ge \sum_{i=1}^{n} p_{i} x_{ij} - (1-h) \sum_{i=1}^{n} c_{i} x_{ij}$$
 (8)

$$y_j \le \sum_{i=1}^n p_i x_{ij} + (1-h) \sum_{i=1}^n c_i x_{ij}$$
 (9)

Note that from in (8) and (9), $\sum_{i=1}^{n} p_i x_{ij}$, defines the middle value and $\sum_{i=1}^{n} c_i x_{ij}$ defines the sympatric spread to the left, constraint (8), and to the right, constraint (9), as illustrated in figure 2. As can be seen from the figure 2, as the degree of fuzziness, h, increases the spread, c_i , increases and therefore the uncertainty associated with the p_i would increase [21].

3.2 Fuzzy output model [22]:

Due to human error and various other sources of imprecision in the measurements, the output may certainly be fuzzy. The uncertainty in the measurements is represented by a fuzzy number as $Y_{2j} = (y_j, e_j)$, where y_j is the middle value and e_j represents the uncertainty in measurement *j* as shown in figure 3.

$$\mu_{Y_{j}}(y) = 1 - \frac{\left| y_{j} - y \right|}{e_{j}}$$
(10)

An estimation of equation 10 would be:

$$\mu_{Y_{j}}(y) = 1 - \frac{\left| y_{j} - \sum_{i=1}^{n} p_{i} x_{ij} \right|}{\sum_{i=1}^{n} c_{i} \left| x_{i} \right|}$$
(11)

The above the objective of the fuzzy linear regression is to determine the fuzzy parameters A^* that minimize the sum of spread as in:

$$F_{fuzzy - output} = \min(\sum_{j=1}^{m} \sum_{i=1}^{n} c_i x_{ij}) \quad (12)$$

Subject to:

$$y_{j} \ge \sum_{i=1}^{n} p_{i} x_{ij} - (1-h) \sum_{i=1}^{n} c_{i} x_{ij} + (1-h) e_{j} (13)$$
$$y_{j} \le \sum_{i=1}^{n} p_{i} x_{ij} + (1-h) \sum_{i=1}^{n} c_{i} x_{ij} - (1-h) e_{j} (14)$$

Note that an additional term, $(1-h)e_j$, emerged in the formulation due to the introduction of fuzziness (or uncertainty) in the measurements. As mentioned, the equation (13) represents the y_j when it lies in the interval to the left of the middle value with the uncertainty with respect to it added to that interval. In the same manner, equation (14) represents the y_j when it lies in the interval to the right of the middle value with the uncertainty with respect to it added to that interval. The prove and detailed derivation for both formulation may be found in [19, 22].



Figure 3 membership function of output

4 Implementation of case studies

This section presents some typical results obtained by applying the proposed algorithms to regression examples, example 14.6, page 544 and problem 14.11, page 554 in [26], and a D.C. state estimation example 3.7, page 49 in [29].

A set of MATLABTM files has been developed to facilitate the computation of all fuzzy parameters and states. The LP problems have been solved by the function linprog() incorporated in the MATLABTM optimization toolbox [30].

4.1 Regression example 14.6 [26]:

<i>y</i> _i	x _{ij}
2.1	0.52
4.6	1.36

From the above given data objective is to obtain the fuzzy regression parameters along with their spreads. After invoking the relevant Matlab script file:

Fuzzy-LP
P ₀ : 0.5524
$P_1: 2.9762$
11.2.7702
$c_0: 0.0000$
c ₁ : 0.0000

Which match the result in that same reference [26]. Note that with zero spreads, there is a strong indication of what is called an exact fit. Therefore the parameters are called crisp in this case.

4.2 Problem 14.11, page 554 in [26]:

Now with assuming that the above data set has an ambiguity represented by the e_i in the output y_i , see problem 14.11, page 554 in [26], as follows:

(y_i, e_i)	x _{ij}
(2.1, 0.20)	0.52
(4.6, 0.35)	1.36

In order to solve for the regression parameters of the above scenario, the second formulation should be employed (equation 12 to 14), which can handle fuzziness, or uncertainty, in the output y_i .

Fuzzy-LP
$\begin{array}{c} P_0: \ 0.5524 \\ P_1: \ 2.9762 \\ c_0: \ 0.1071 \\ c_1: \ 0.1786 \end{array}$

The above solution appears to be correct when compared with the results shown in the previous example. As can be seen that the spread for both parameters is no longer zero (an indication that there are uncertainties in the estimates and the measurements). Also looking at both results, it is crucial to mentioned that the solution of the same problem shown in [31] is totally wrong and can never be justified. For both runs the degree of fuzziness used is 0.4.

4.3 D.C. state estimation example 3.7, page 49 in [29]:



Figure1. Two-bus system

As far as state estimation is concerned, both fuzzy and non-fuzzy formulation are applied on a small D.C. network, shown in figure 1. It is important to mention that there is no need for linearization, since the problem is D.C., i.e. linear. Note that the degree of fuzziness used is 0.5.

Fuzzy-LP		
$\begin{array}{rrrr} P_0: & 1.0300 \\ P_1: & -0.7324 \\ c_0: & 0.0400 \\ c_1: & -0.0247 \end{array}$		

Note that form the above solution, it appears that states (P_G and P_{L1}) both have uncertainty in them (with out us knowing the source of uncertainty). Or there could be an element of in the parameters, (i.e. Y-bus), of the given example. Hence the states are not crisp and an uncertainty interval can be constructed around the states as follows:

The lower bound of P_G is $P_G^- = 0.99$ p.u. The upper bound of P_G is $P_G^+ = 1.07$ p.u. The lower bound of P_{L1} is $P_{L1}^- = -0.7571$ p.u. The upper bound of P_{L1} is $P_{L1}^+ = -0.7077$ p.u.

Now applying the fuzzy formulation to the same example, but in this case information about the meters accuracies is available. The standard deviation of all meters is obtained from the covariance matrix given in that same example. The STD for the three meters is $STD = [0.0632 \quad 0.0316]$ 0.0316]. The outcome is:

Fuzzy-LP	-
P ₀ : 1.0300 P ₁ : -0.7324 c ₀ : 0.1032 c ₁ : -0.0563	

As can be seen, the state remains to be the same as in the previous run, however, the spreads have gotten larger to accommodate for more uncertainty in the measurements. It must be emphasized that an increase in the meters inaccuracy might cause the states not to remain constant.

5 Conclusion

An analysis of uncertainty in power system state estimation is presented in this report. The uncertainty is modelled and is assumed to be present in the system parameters and in the measurements which take into account known meter accuracies. A Fuzzy linear estimator was employed to estimate the both the states and their respective upper and lower bounds. The provision of bounds by the proposed FLSE offers useful additional information to the power system operator. Based on the preliminary testing and the promising results, form regression and D.C. examples, the proposed FLSE has proved to be effective tool for uncertainty analysis.

The next phase of research is to conduct intense investigation and apply FLSE on A.C. power systems.

Accknowlegedment:

This work was supported in full by the Publice Authority of Applied Scince & Traning, Kuwait. Grant # TS-06-01.

References:

- [1]N. Logic, E. Kyriakides, and G. T. Heydt, "Lp State Estimators for Power Systems," *Journal of Electric Power Components and Systems*, accepted for publication, 2002.
- [2] P. J. Rousseeuw and A. M. Leroy, *Robust regression and outlier detection*. New York ; Chichester: Wiley, 1987.
- [3] A. Stefanski, "A note on High- Breakdown Estimator," *Statistics and Probability*, vol. 11, pp. 353-358, 1991.
- [4] P. Ryan Thomas, *Modern regression methods*. New York ; Chichester: J. Wiley, 1997.
- [5] P. J. Rousseeuw and V. J. Yohai, "Robust regression by means of S-estimators," in *Robust* and Nonlinear Time Series Analysis, Franke, J., HKardle, W., Martin, R.D. (Eds.), Lecture Notes in Statistics, vol. 26. New York: pp. 256– 272, Springer, 1984.
- [6] F. C. Schweppe, *Uncertain dynamic systems*. Englewood Cliffs, N.J.,: Prentice-Hall, 1973.
- [7] A. Bargiela and G. Hainsworth, "Pressure and flow uncertainty in water system," *ASC J. Water Resources Planning Mang*, vol. 115 (2), pp. 212 -229, 1989.
- [8] A. Nagar and R. Powell, "LFT/SDP approach to the uncertainty analysis for state estimation of water distribution systems," *IEE Proceedings on*. *Control Theory and Applications*, vol. 149, pp. 137-142, Mar 2002.
- [9] A. K. AL-Othman and M. R. Irving, "Uncertainty Modeling in Power System State Estimation," Submitted for review to Generation, Transmission and Distribution, IEE Proceedings-, 2003.

- [10]A. Abur and M. K. Celik, "Least absolute value state estimation with equality and inequality constraints," *Power Systems, IEEE Transactions* on, vol. 8, pp. 680 - 686, May 1993.
- [11]H. Singh, F. L. Alvarado, and W.-H. E. Liu, "Constrained LAV state estimation using penalty functions," *Power Systems, IEEE Transactions on*, vol. 12, pp. 383 - 388, Feb. 1997.
- [12]K. A. Clements, P. W. Davis, and K. D. Frey, "Treatment of inequality constraints in power system state estimation," *Power Systems, IEEE Transactions on*, vol. 10, pp. 567 - 574, May 1995.
- [13]A. Monticelli, F. Wu, and M. Yen, "Multiple bad data identification for state estimation by combinatorial optimisation," *Power Delivery*, *IEEE Transactions on*, vol. PWRD-1, No.3, pp. 361-369, 1996.
- [14]L. Mili, V. Phaniraj, and P. J. Rousseeuw, "Least median of squares estimation in power systems," *Power Systems, IEEE Transactions on*, vol. 6, pp. 511 -523, 1991.