State Estimation Based on Correntropy: A Proof of Concept

Vladimiro Miranda, *Fellow, IEEE*, André Santos, and Jorge Pereira

***Abstract*—This letter proposes a new concept applied to state estimation based on replacing traditional regression models by a criterion of maximizing error correntropy introducing a novel way to identify and correct large errors.**

***Index Terms*—Correntropy, entropy, Parzen windows, state estimation.**

I. Introduction

THIS letter opens the discussion on an alternative to regression models based on least squares or minimum square error (MSE) criterion in state estimation (SE) [1]. MSE is an optimal approach to estimate parameters only if the underlying distribution of errors is Gaussian. A gross error will distort this distribution and may be seen as an outlier that should be isolated and removed. With the MSE criterion, this is not possible: any gross error will contaminate the estimation of all parameters and all other errors. The new paradigm for SE will be based on maximizing the information that one can extract from the available measurements. The information content of a probability density function (pdf) is measured by its entropy, and therefore, instead of looking just at the variance of the error distribution, the new paradigm will look at properties related with its entropy.

This letter presents a *proof of concept* in terms of a theoretical model and examples on how the adoption of entropy-related concepts allow the natural identification and correction of gross errors.

II. Entropy, Parzen Windows, and Correntropy

An ideal method would lead to an error distribution with minimum entropy—a Dirac function. If this Dirac function is centered at zero, it will mean that all errors are zero. A way to measure entropy of an error distribution represented by its pdf is Renyi's [2] quadratic entropy definition of a discrete probability distribution $\mathrm{P} = (\mathrm{p}_1, \ldots, \mathrm{p}_\mathrm{n})$. This definition can be generalized for a continuous random variable Y with pdf $\mathrm{f}_\mathrm{Y}(\mathrm{z})$ as follows:

$$\mathrm{H}_{\mathrm{R2}} = -\log \sum_{\mathrm{k}=1}^{\mathrm{N}} \mathrm{p}_\mathrm{k}^2, \;\; \mathrm{H}_{\mathrm{R2}} = -\log \int_{-\infty}^{+\infty} \mathrm{f}_\mathrm{Y}{}^2(\mathrm{z})\mathrm{dz}. \quad (1)$$

The estimation of the pdf of data from a sample constituted by discrete points $\mathbf{y}_\mathrm{i} \in \mathrm{R}^\mathrm{M}, \mathrm{i} = 1, \ldots, \mathrm{N}$ in a M-dimensional space may be done by the Parzen window method [3]. This technique uses a kernel function centered on each point; it looks at a point as being locally described by a probability density Dirac function, which is replaced or approximated by a continuous set whose density is represented by the kernel. If a Gaussian kernel $\mathrm{G}(.,.)$ is used, the expression of the estimation $\hat{\mathrm{f}}_\mathrm{Y}$ for the real pdf f_Y of a set of N points is a summation of individual contributions

$$\hat{\mathrm{f}}_\mathrm{Y}(\mathbf{z}) = \frac{1}{\mathrm{N}} \sum_{\mathrm{i}=1}^{\mathrm{N}} \mathrm{G}(\mathbf{z} - \mathbf{y}_\mathrm{i}, \sigma^2 \mathbf{I}) \quad (2)$$

where $\sigma^2\mathbf{I}$ is the covariance matrix (here assumed with independent and equal variances in all dimensions).

The information theoretic learning approach combines (1) with (2) to obtain an operational function representing the entropy of the error distribution that can be massaged into an algorithm [4]. However, a related concept based on a generalized similarity measure called correntropy [5] leads to similar results and much faster algorithms. The maximum correntropy criterion (MCC) may be translated by

$$\mathrm{MCC}(\varepsilon) \Leftrightarrow \max \frac{1}{\mathrm{N}} \sum_{\mathrm{i}=1}^{\mathrm{N}} \mathrm{G}(\varepsilon_\mathrm{i}, \sigma^2 \mathrm{I}). \quad (3)$$

This has the effect of maximizing the value of the error pdf at zero, which will tend to approach a Dirac function at that location. However, the properties of MCC [5] make it behave like MSE for small errors and become insensitive to large errors—it is this effect that will be beneficial to detect gross errors, while keeping the known properties of SE based on least squares for small errors.

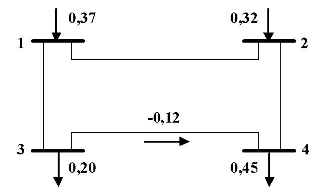

Fig. 1. Measured values of active power flows in busbars and in line 3–4.

III. Illustration for DC State Estimation

The desired effect can be found in a DC model of an SE problem (DC-SE). Take the system and data represented in Fig. 1, where all lines have an impedance of j0.1 p.u. A gross error has been introduced in line 3–4, because the flow coherent with power injections has been reversed.

In the DC model, the errors $\boldsymbol{\varepsilon}$ are given by

$$\varepsilon = \mathbf{Z} - \hat{\mathbf{Z}} = \mathbf{Z} - \mathbf{M}\hat{\boldsymbol{\theta}} \quad (4)$$

where $\mathbf{Z}$ is the vector of measurements from the SCADA, $\hat{\mathbf{Z}}$ is the vector of calculated values, $\mathbf{M}$ is the matrix of equation co-

Manuscript received January 21, 2009. First published August 25, 2009; current version published October 21, 2009. Paper no. PESL-00008-2009.

V. Miranda and A. Santos are with INESC Porto and with the Faculty of Engineering of the University of Porto, Porto, Portugal (e-mail: vmiranda@inescporto.pt; afsantos@inescporto.pt).

J. Pereira is with INESC Porto and with the Faculty of Economy of the University of Porto, Porto, Portugal (e-mail: jpereira@inescporto.pt).

Color versions of one or more of the figures in this paper are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/TPWRS.2009.2030117

TABLE I
ESTIMATED POWER INJECTIONS AND LINE FLOW

	Meas.	MSE	MCC	ε in MSE	ε in MCC
P_1	0.37	0.339048	0.359998	0.030952	0.010002
P_2	0.32	0.330952	0.310002	-0.010952	0.009998
P_3	-0.20	-0.272857	-0.210006	0.072857	0.010005
P_4	-0.45	-0.397143	-0.459995	-0.052857	0.009995
F_{34}	-0,12	0.047619	0.099995	-0.137619	-0.219995

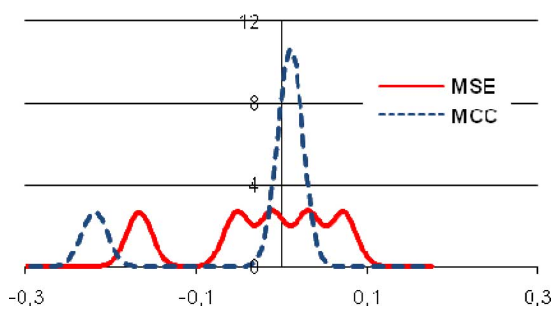

Fig. 2. Pdf of errors in both cases: the MCC criterion concentrates errors near 0 and defines one outlier. Parzen windows size in (2): $\sigma = 0.01$. The x-axis represents error values in p.u.

efficients (from the DC model), and $\hat{\boldsymbol{\theta}}$ is the vector of estimated nodal voltage angles. The MSE criterion corresponds to the application of the classical least squares solution to this problem. The application of the MCC criterion to DC-SE leads to the following objective function:

$$\mathrm{MaxC} = \frac{1}{\mathrm{N}} \sum_{\mathrm{j}=1}^{\mathrm{N}} \frac{1}{\sigma\sqrt{2\pi}} \mathrm{e}^{-(1/2\sigma^2)(\mathrm{z_j} - \mathbf{M}_\mathrm{j}\boldsymbol{\theta})^2} \tag{5}$$

where $\mathbf{M}_\mathrm{j}$ is line j of $\mathbf{M}$, N is the number of measurements (and errors), and σ is the size of the Parzen windows. Equation (5) is a nonlinear criterion requiring an iterative algorithm.

The application of both models to Fig. 1 leads to the estimated power injections and line flow in Table I, with errors ε (solution with a gradient algorithm and $\sigma = 0{,}05$).

It is clear from Fig. 2 that the MSE criterion distributed the gross error to all other errors: all estimation is contaminated, which is a well-known effect. Instead, the MCC criterion ignored the measurement with a gross error and concentrated all other errors close to 0. It can be shown that it gives the same result as for an SE using MSE when only the four first data are taken as available measurements and the fifth one is excluded.

This remarkable result gets confirmation in a full AC problem. A test has been conducted with the IEEE-RST 24-bus system [6] with minor modifications. A set of measurements was composed from a power flow solution, and a gross error is introduced in line 3–9 by reversing its active power flow from 4 to -4 MW (an error of 8). The SE problem was solved with the MSE criterion using a least squares routine and was also solved maximizing the MCC criterion using an EPSO algorithm [7] and $\sigma = 0.4$. Fig. 3 compares the pdf of the error distribution for both criteria and shows the error for each of the 50 measurements considered. The MCC criterion led to the correction of the gross error in 8 MW (measurement P39) and produced a state estimation with no other significant errors, while the least squares criterion led to errors in a large number of buses and line flows. The error pdf for MCC approaches a Dirac function at 0, and the gross error does not influence the estimation.

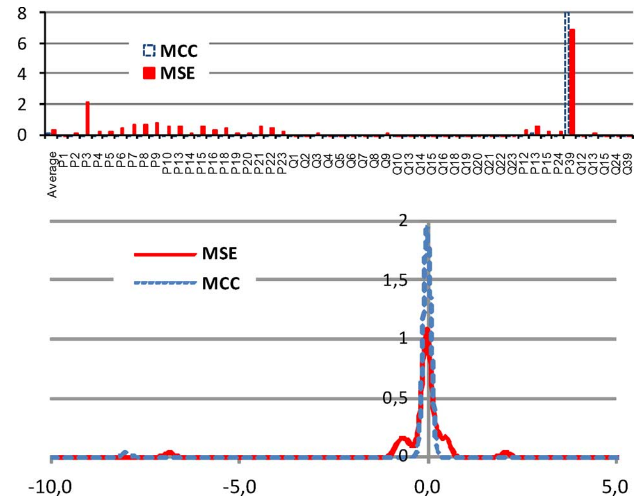

Fig. 3. Top: abs. errors (in MW) assigned by each method to each of the 50 measurements. Bottom: pdf of errors with MSE and MCC. Parzen windows size in (2): $\sigma = 0.1$. The x-axis represents error values in MW.

IV. CONCLUSIONS

This letter provides experimental results as *proof of concept* to the adoption of a new criterion in SE. One proposes that correntropy, instead of least squares, shall be used as the function of errors to be optimized. Correntropy behaves like MSE (or a L2 metric) for small errors but like a L0 metric for large errors and allows the natural isolation of gross errors as outliers, while keeping small random errors compensated. The task now will be to develop an efficient algorithm for online use taking advantage of these properties. Because correntropy is a differentiable function, suitable Newton algorithms are likely to come to light soon.

REFERENCES

[1] A. Abur and A. G. Expósito, *Power System State Estimation: Theory and Implementation*, 1st ed. New York: Marcel Dekker, 2004.
[2] A. Renyi, "Some fundamental questions of information theory," in *Selected Papers of Alfred Renyi*. Budapest, Hungary: Akademia Kiado, 1976, vol. 2, pp. 526–552.
[3] E. Parzen, "On the estimation of a probability density function and the mode," *Ann. Math Statist.*, vol. 33, no. 3, pp. 1065–1076, Sep. 1962.
[4] D. Erdogmus and J. C. Principe, "Generalized information potential criterion for adaptive system training," *IEEE Trans. Neural Netw.*, vol. 13, pp. 1035–1044, Sep. 2002.
[5] W. Liu, P. Pokharel, and J. Principe, "Correntropy: Properties and applications in non-Gaussian signal processing," *IEEE Trans. Signal Process.*, vol. 55, no. 11, pp. 5286–5298, Nov. 2007.
[6] IEEE PES Task Force, "IEEE Reliability Test System"," *IEEE Trans. Power App. Syst.*, vol. PAS-98, no. 6, pp. 2047–2054, Nov./Dec. 1979.
[7] V. Miranda, H. Keko, and A. J. Duque, "Stochastic star communication topology in evolutionary particle swarms (EPSO)," *Int. J. Comput. Intell. Res.*, vol. 4, no. 2, pp. 105–116, 2008.