Impact of different central path neighborhoods on gross error identification in State Estimation with generalized correntropy interior point method

Hamed Moayyed
INESC TEC
Faculty of Engineering of the University of Porto
Porto, Portugal
hmoayyed@fe.up.pt

Shabnam Pesteh
INESC TEC
Faculty of Engineering of the University of Porto
Porto, Portugal
spesteh@inesctec.pt

Vladimiro Miranda
INESC TEC
Faculty of Engineering of the University of Porto
Porto, Portugal
vmiranda@inesctec.pt

Jorge Pereira
INESC TEC
Faculty of Engineering of the University of Porto
Porto, Portugal
jpereira@inesctec.pt

***Abstract*— Classical Weighted Least Squares (WLS) State estimation (SE) in power systems is known for not performing well in the presence of Gross Errors (GE). The alternative using Correntropy proved to be appealing in dealing with outliers. Now, a novel SE method, generalized correntropy interior point method (GCIP) is being proposed, taking advantage of the properties of the Generalized Correntropy and of the Interior Point Method (IPM) as solver. This paper discusses how the choice of different central path neighborhoods, an essential concept in IPM, is critical in the identification of gross errors. The simulation results indicate that a one-sided infinity norm neighborhood successfully identifies outliers in the SE problem, making GCIP a competitive method.**

Keywords*— state estimation, generalized correntropy, interior point method, gross error identification

I. INTRODUCTION

A. Motivation and Background

The large scale integration of generation from renewable sources and the expansion of State Estimation (SE) to distribution networks put back into light the need to deal with bad data arriving at the SCADA.

It is widely recognized that the Least Squares regression is sensitive to outliers and cannot deal adequately with gross errors, even in its Weighted Least Squares (WLS) version. For the same reason, methods departing from a WLS solution, to try to identify measurements with gross errors, showed poor performance when facing multiple bad data.

The alternative to methods departing from the WLS solution would be methods that would naturally deal with outliers, instead of trying to identify them in a post-processing phase. An interesting approach that has been receiving increasing attention in the adoption of a criterion based on a robust nonlinear similarity measure in a kernel space, Correntropy [1] [2]. Correntropy adopts a Gaussian kernel, which is, of course, only one among other kernel choices. Kernel shape, however, may have impact on the ability of procedures to compute correct solutions. We propose in this paper a robust state estimator based on Generalized Correntropy Criterion (GCC) that adopts the Generalized Gaussian density (GGD) function as kernel function.

B. Relevant Literature

Generalized Correntropy (GC) [3] can achieve extra flexibility through the GC parameters, which control the behavior of the induced metric, and shows a markedly better robustness than plain Correntropy. A novel method, called Generalized Correntropy Interior Point method (GCIP), answers to the SE problem in a very efficient way. GCIP is based on Generalized Error Correntropy instead of traditional Least Squares (LS) and on an interior point method instead of the conventional Gauss-Newton algorithm. This method receives an abridged description in the following sections of the paper.

Interior Point Methods (IPM) have been proposed in the past, to process SE. One of the main features of interior point methods (IPM) [4] is the generation of iterates in a neighborhood of the central path. The concept of the central path has played a crucial role in the development and analysis of IPMs. All primal-dual interior point methods require the iterates to remain in an appropriate neighborhood of the central path. This adherence to the central path promotes global convergence of the duality gap sequence and it ensures the number of iterations required to produce an appropriate solution of the problem.

Many IPMs are based on a small neighborhood or the negative infinity norm neighborhood. The first method uses full Newton steps and the iterates stay in a small neighborhood of the central path. The second method is called the large-step method, in which the iterates are allowed to move within a wide neighborhood of the central path.

C. Contributions and Organization

GCIP introduces an innovative technique to identify measurements contaminated with spike noises based on a very

978-1-7281-1156-8/19/$31.00 ©2019 IEEE

important feature occurring in IPM central path. In fact, the central path neighborhood excludes points that are too close to the boundary of the non-negative orthant [5]. Measurements contaminated with spike noises are however located far from the central path, mostly placed near the boundaries. By interactions between contaminated and clean measurements, some sound measurements may become located near the boundaries as well. GCIP verifies such suspicion by additionally evaluating the well-known weighted residuals. The results show the efficiency of GCIP compared to traditional largest normalized residual test (LNRT) [6] which employs LS as objective function and Gauss-Newton as search algorithm.

This paper is devoted to demonstrating the effect of different neighborhoods in the identification of GE. With illustrative examples, it allows recognizing which is the best neighborhood definition, leading to most accurate results in the application of GCIP. The results presented in the paper help in understanding why GCIP, relying on Generalized Correntropy as criterion and on IPM as solver, is a powerful and successful alternative to SE in environments contaminated with gross errors.

II. GENERALIZED CORRENTROPY

Information entropy is a concept from information theory. A measure of information content of a probability density function is called Entropy. Compared to the definition of Shannon [7], Renyi's entropy [8] has proven to allow working algorithms' development in an easier manner. Correntropy [1] is a generalized similarity measure between two pdfs, with a strong relation with entropy. Considering the two scalar random variables X and Y, it is defined by

$$v_\sigma(X,Y) = E[\kappa_\sigma(X-Y)] = \iint_{xy} \kappa_\sigma(x-y)p(x,y)dxdy \quad (1)$$

where κ_σ is a kernel operator with parameter (width) σ. Most of applications have adopted the Gaussian kernel G_σ, among other reasons because of its positive semidefinite property and computational tractability. Since the joint pdf $p(x,y)$ is usually unknown and only a finite number of data $\{(x_i,y_i)\}_{i=1}^N$ are available, the correntropy estimated by the Parzen windows method [9] is given by:

$$\hat{v}_\sigma(X,Y) = \frac{1}{N}\sum_{i=1}^N G_\sigma(x_i-y_i) \quad (2)$$

This was the formulation used in [2].

A very useful generalization was produced with the adoption of the GGD (generalized Gaussian density), as proposed in [3], which is given by

$$G_{c,\sigma}(e) = \frac{c}{2\sigma\Gamma(1/c)} exp\left(-\left|\frac{e}{\sigma}\right|^c\right) \quad (3)$$

where e is the error between random variables X and Y, $c > 0$ is the shape parameter, $\Gamma(.)$ is the gamma function and σ is the scale (bandwidth) parameter defined by $\sigma = \zeta\sqrt{\Gamma(1/c)\Gamma(3/c)}$. The Generalized Correntropy (GC) estimator for a finite discrete sample becomes

$$\hat{v}_{c,\sigma}(X,Y) = \frac{1}{N}\sum_{i=1}^N G_{c,\sigma}(x_i-y_i) \quad (4)$$

Associated with the GC, there is the generalized correntropy induced metric (GCIM), which depending on the parameters adopted and the region of analysis may behave like different norms. Consequently, while GCIM may vary from L_∞ to L_0, the metric induced by a Gaussian-based Correntropy only varies from L_2 to L_0. This property is demonstrated to be effective in SE with gross errors. Specifically, in the case of outliers, the ability to apply an L_0 metric (indifference in distance) is what is needed. As their residuals can be pushed to whatever value that is required without disturbing the objective function evaluation, it leads to minimizing the remaining residuals without contamination. In addition, the indifference of L_∞ metric to all components in a residual vector, except the most significant one, helps in separating outliers by providing freedom to the residual vector components' rearrangement.

This paper presents a new criterion in SE: the MGCC – maximum generalized correntropy criterion, with the same theoretical perspective as MCC: to produce a solution tending to minimize the Entropy of the error distribution while complying with the requirement that this distribution should have a zero mean, i.e., no bias present.

III. INTERIOR POINT SOLVER

Maximizing error Correntropy is tantamount to minimizing the Entropy of the error distribution, subject to the condition that the mean should be zero. Correntropy, as a cost function, has the properties of a smooth M-estimator of the kind of the Welsch estimator [10]. Thus, its successful optimization depends on both problem formulation and the algorithm used as solver.

The unconstrained Non-Linear Programming formulation of the SE problem maximizing Correntropy results in a multi-modal landscape and the traditional flat start for Newton iterations does not work in all cases, as it may be out of the region of necessary conditions for the process to converge. The algorithmic difficulties cannot provide full convergence guarantee and some rate of failure is verified [11]. This justifies experimenting with a different solving strategy, and an Interior Point Method approach has proved to have the needed behavior, as the following sections will show.

The first suggestion for using IPM in power systems has been offered in the 1980s [12], mainly applying IPM to a SE formulation associated with WLAV (weighted least absolute value) criteria [13]. IPM could assist in the rejection of some outliers and presented computing efficiency but the combination IPM/WLAV displayed other problems that led to difficulties in acceptance.

The following sections present the theoretical basis and the mathematical properties of an alternative combination of the IPM solver with the MGCC criterion. This MGCC/IPM combination forms the GCIP model for power systems state estimation and, as it will be shown, exhibits superior properties in correct estimation and multiple outlier rejection.

IV. The State Estimation Model Based on GC

The proposed state estimator model based on GC is given by the following optimization problem:

$$\max \quad J(x) = \frac{1}{m}\sum_{i=1}^{m} \frac{c}{2\sigma\Gamma(1/c)} exp\left(-\left|\frac{r_i}{\sigma}\right|^c\right) \tag{5}$$
$$\text{subject to } \begin{cases} g(x) = 0 \\ r = z - h(x) \end{cases}$$

where z is the vector of measurements, x is the vector of state variables, $h(x)$ is the nonlinear state estimation function that relates the measurements to the system states (the power flow equations producing estimated values for the measured variables), n is the number of state variables to be estimated, m is the number of available measurements, r is the vector of residuals (difference between measured and estimated values) and $g(x) : R^n \rightarrow R^{\Omega}$ are zero-injection equality constraints.

The absolute value in the proposed model (5) encounters difficulties in being solved with gradient-based methods. We introduce inequality constraints for the residual term and add non-negative slack variables to be relieved of the absolute value. The details are as follows.

Let $s \in R^m$ be defined such that

$$-s_i \le r_i \le s_i \tag{6}$$

Introduce two non-negative slack variables $u, v \in R^m$, and define two new non-negative variable vectors $p, q \in R^m$ as $p_i = \frac{1}{2}u_i$ and $q_i = \frac{1}{2}v_i$, we have

$$r_i = p_i - q_i \tag{7}$$
$$s_i = p_i + q_i \tag{8}$$

From (6) and (8), we can reach a new equivalent and differentiable system:

$$\max \ J(x) = \frac{1}{m}\sum_{i=1}^{m} \frac{c}{2\sigma\Gamma(1/c)} exp\left(-\left(\frac{p_i+q_i}{\sigma}\right)^c\right) \tag{9}$$
$$\text{subject to } \begin{cases} f(x) = 0 \\ g(x) = 0 \\ p, q \ge 0. \end{cases}$$

where $f(x) = z - h(x) - p + q$. In the following, we summarized the solution methodology and mathematical justification of GCIP model.

V. Primal-Dual Interior Point Method

The solution methodology is summarized in three main steps; introducing Lagrange multipliers, deriving the Karush-Kuhn-Tucker (KKT) and finally applying the Newton method. The details are as follows.

The Lagrangian function associated with (9) can be defined as

$$L \equiv J(x) - \alpha^T(f(x)) - \beta^T g(x) - \gamma^T q - \lambda^T p \tag{10}$$

where $\beta \in R^{\Omega}$; λ, γ and $\alpha \in R^m$ are Lagrange multipliers. Then, with deriving the Karush-Kuhn-Tucker (KKT) condition and applying the Newton method, we can express the correction equations as:

$$\gamma_i dq_i + q_i d\gamma_i = -L^{\mu}_{\gamma_i} q_i \tag{11}$$
$$\lambda_i dp_i + p_i d\lambda_i = -L^{\mu}_{\lambda_i} \tag{12}$$
$$e_i dp_1 + e_i dq_1 - d\alpha_i - d\gamma_i = -L_{q_i} \tag{13}$$
$$e_i dp_1 + e_i dq_1 + d\alpha_i - d\lambda_i = -L_{p_i} \tag{14}$$
$$\nabla g(x)\,\mathrm{d}x = -L_{\beta} \tag{15}$$
$$-\nabla h(x)\,\mathrm{d}x - \mathrm{d}p + \mathrm{d}q = -L_{\alpha} \tag{16}$$
$$(\nabla^2 h(x)\beta - \nabla^2 g(x)\alpha)dx + \nabla g(x)^T d\beta - \nabla h(x)^T d\alpha = -L_x \tag{17}$$

where $\nabla^2 h(x)$ and $\nabla^2 g(x)$ are Hessian matrices of $h(x)$ and $g(x)$; e_i is as follows

$$e_i = \frac{J(x)}{m}\left(\frac{-c(c-1)(p_i+q_i)^{c-2}}{\sigma^c} exp\left(-\left(\frac{p_i+q_i}{\sigma}\right)^c\right) + \left(\frac{c^2(p_i+q_i)^{2c-2}}{\sigma^{2c}}\right)\exp\left(-\left(\frac{p_i+q_i}{\sigma}\right)^c\right)\right) \tag{18}$$

The perturbed parameter μ is calculated as $\mu = \rho.\mathrm{Gap}/2m$ where $Gap = \gamma^T q + \lambda^T p$ and $\rho \in (0,1)$ is a centering parameter [14].

From (11-14), let $\begin{bmatrix} a_i & b_i \\ c_i & d_i \end{bmatrix} = \begin{bmatrix} q_i e_i + \gamma_i & q_i e_i \\ p_i e_i & p_i e_i + \lambda_i \end{bmatrix}^{-1}$, then the following expressions are obtained

$$dq_i = n_{1i} d\alpha_i + t_{1i} \tag{19}$$
$$dp_i = n_{2i} d\alpha_i + t_{2i} \tag{20}$$

where $n_{1i} = a_i q_i - b_i p_i; \quad n_{2i} = c_i q_i - d_i p_i$

$t_{1i} = -a_i(q_i L_{q_i} + L^{\mu}_{\gamma_i}) - b_i(p_i L_{p_i} + L^{\mu}_{\lambda_i})$, $t_{2i} = -c_i(q_i L_{q_i} + L^{\mu}_{\gamma_i}) - d_i(p_i L_{p_i} + L^{\mu}_{\lambda_i})$.

Substituting (19) and (20) into (16) then

$$\nabla h(x)\,\mathrm{d}x + L\,\mathrm{d}\alpha = v \tag{21}$$

where $L \in R^{m\times m}$ is a diagonal matrix, with the elements $L_i = -n_{1i} + n_{2i}$; $v = z - h(x) - p + q + t_1 + t_2$ and $t_1, t_2 \in R^m$ as defined in (19) and (20). Then the correction equation can be obtained according to (15), (17) and (21):

$$\begin{bmatrix} G & 0 & 0 \\ 0 & G^T & -H^T \\ H & 0 & L \end{bmatrix}\begin{bmatrix} \mathrm{d}x \\ \mathrm{d}\beta \\ d\alpha \end{bmatrix} = \begin{bmatrix} -L_{\beta} \\ -L_x \\ v \end{bmatrix} \tag{22}$$

where $G = \nabla g(x)$; $H = \nabla h(x)$ and second derivative terms ignored in (16) just as in the WLS method. The values of $d\boldsymbol{x}$, $d\boldsymbol{\beta}$, and $d\boldsymbol{\alpha}$ can be obtained from (22); then, (19) and (20) give $d\boldsymbol{p}$ and $d\boldsymbol{q}$. Eventually, $d\boldsymbol{\gamma}$ and $d\boldsymbol{\lambda}$ can be obtained from (13) and (14).

We can calculate the primal and dual step sizes by

$$\Delta_P = 0.9995\min\{\min(-q_i/dq_i : q_i < 0; -p_i/dp_i : p_i < 0), 1\} \tag{23}$$
$$\Delta_D = 0.9995\min\{\min(-\gamma_i/d\gamma_i : \gamma_i < 0; -\lambda_i/d\lambda_i : \lambda_i < 0), 1\} \tag{24}$$

which can ensure that the slack variables p and q satisfy $p, q > 0$, and the Lagrange multipliers λ and γ satisfy $\lambda, \gamma > 0$.

VI. Impact of different Central Paths in the Identification of GE

IPMs can be classified in different ways: primal, dual and primal-dual IPMs, feasible and infeasible methods, affine-scaling methods and path following IPMs. We considered the primal-dual path-following IPMs, which stipulates choosing a target on the central path and then applying the Newton method, while keeping the iterates within a neighborhood of the central path while continuously approaching the optimal set. In this way, we can consider small-step and large-step IPMs. The small-step methods work in a small neighborhood of the central path; in the large-step methods, the iterates are allowed to move within a wide neighborhood of the central path.

The new algorithm developed for SE benefits from the abovementioned feature of IPM central path following, to detect and identify measurements contaminated with spike noises. As far as one can recollect, this is the first time that this feature is used to identify GE.

The central path $\mathcal{C}$ is an arc of strictly feasible points that plays a vital role in the theory of IPM algorithms. In our algorithm, we restrict the iterations to a neighborhood of the central path and follow $\mathcal{C}$ to find a better solution. Therefore, directions calculated from any point in the neighborhood make a progress toward the solution set. The different neighborhoods of central path that were under analysis in identification of GE:

1) 2-norm

$$\mathcal{N}_{2P}(\tau) = \{(\gamma_i, q_i, \mu_i) \in \mathcal{F} | \|\gamma_i q_i - \mu\|_2 \leq \tau\mu \text{ for all } i = 1,2,\dots,m\} \quad (25)$$

$$\mathcal{N}_{2D}(\tau) = \{(\gamma_i, p_i, \mu_i) \in \mathcal{F} | \|\gamma_i p_i - \mu\|_2 \leq \tau\mu \text{ for all } i = 1,2,\dots,m\} \quad (26)$$

2) ∞-norm

$$\mathcal{N}_{\infty \mathrm{P}}(\tau) = \{(\gamma_i, q_i, \mu_i) \in \mathcal{F} | \|\gamma_i q_i - \mu\|_\infty \leq \tau\mu \text{ for all } i = 1,2,\dots,m\} = \{(\gamma_i, q_i, \mu_i) \in \mathcal{F} | \gamma_i q_i \geq \tau\mu \text{ for all } i = 1,2,\dots,m\} \quad (27)$$

$$\mathcal{N}_{\infty \mathrm{D}}(\tau) = \{(\gamma_i, p_i, \mu_i) \in \mathcal{F} | \|\gamma_i p_i - \mu\|_\infty \leq \tau\mu \text{ for all } i = 1,2,\dots,m\} = \{(\gamma_i, p_i, \mu_i) \in \mathcal{F} | \gamma_i p_i \geq \tau\mu \text{ for all } i = 1,2,\dots,m\} \quad (28)$$

3) pseudonorm

$$\mathcal{N}_{\infty P}^{-}(\tau) = \{(\gamma_i, q_i, \mu_i) \in \mathcal{F} | \|\gamma_i q_i - \mu\|_\infty^{-} \leq \tau\mu \text{ for all } i = 1,2,\dots,m\} = \{(\gamma_i, q_i, \mu_i) \in \mathcal{F} | \gamma_i q_i \geq (1-\tau)\mu \text{ for all } i = 1,2,\dots,m\} \quad (29)$$

$$\mathcal{N}_{\infty D}^{-}(\tau) = \{(\gamma_i, p, \mu_i) \in \mathcal{F} | \|\gamma_i p_i - \mu\|_\infty^{-} \leq \tau\mu \text{ for all } i = 1,2,\dots,m\} = \{(\gamma_i, p_i, \mu_i) \in \mathcal{F} | \gamma_i p_i \geq (1-\tau)\mu \text{ for all } i = 1,2,\dots,m\} \quad (30)$$

where $\tau \in (0,1)$ defines the neighborhood parameter and $\mathcal{F}$ represents the primal-dual feasible set; $\|z\|_\infty^{-} = \|z^{-}\|_\infty$ and $(z^{-})_i = min\{z_i, 0\}$.

The neighborhood excludes points that are too close to the boundary of the non-negative orthant. Measurements contaminated with spike noises are however located far from the central path. In fact, they are mostly placed near the boundaries. However, as the measurements contaminated with GE are interacting with the clean measurements, some sound measurements may become placed close to the boundaries as well. However, by having knowledge of residual values related to these measurements, one could verify this suspicion by additionally evaluating the well-known weighted residuals:

$$r_w = R^{-1/2} r \quad (31)$$

where R is the covariance matrix of the measurement error vector. This paper sets significance thresholds at a strict level of 5σ (T=5) as it corresponds to the probability of 0.00006% for two-sided p-value (+/- residuals).

It is important to mention that contrary to a classical unconstrained non-linear program formulation, the deletion of the GEs in the proposed method does not generate any computational difficulty. For a general non-linear system, where the mapping is continuously differentiable, the sequence generated by Newton method converges to a solution that is non-degenerate, i.e. the Jacobian is non-singular. In an IPM setting, convergence can be achieved even when the Jacobian approaches a singular limit and even when the solution is not unique [4].

VII. GCIP Estimator Algorithm

Table 1, summarizes the GCIP estimator algorithm in detail. Following the general logic steps of previous approaches [15] [16] it achieved significant improvements that allow flexibility in problem-solving from changing the objective function's shape parameter. A second prominent feature is the proper path following method, by selecting a well-defined neighborhood and choosing an appropriate neighborhood parameter, which improves the search of an accurate solution. But perhaps the most important feature is the novel approach introduced to detect and identify GEs.

Table I. GCIP Estimator Algorithm

Step 1: (Initialization) Set the iteration count to zero ($k = 0$) and define k_{max}; where k and k_{max} are iteration count and maximum number of iterations. Set the centering parameter $\rho \in (0,1)$ and tolerance $\varepsilon = 10^{-3}$. Choose $\sigma > 0$, $\alpha^{(0)} = \beta^{(0)} = 0$ and $p^{(0)}$, $q^{(0)}, \gamma^{(0)}, \lambda^{(0)} > 0$;

Step 2: Select a shape parameter c for the objective function;

Step 3: Compute the Complementary Gap (Gap $\equiv \gamma^T q + \lambda^T p$). If Gap $< \varepsilon$, then the optimal solution has been reached, go to Step 9; else, go to Step 4;

Step 4: Update the relaxation parameter $\mu = \rho . \frac{\mathrm{Gap}}{2m}$;

Step 5: Solve the reduced correction equation (22) for $[dx, d\beta, d\alpha]^T$ and then compute $[dp, dq, d\gamma, d\lambda]^T$;

Step 6: From (23) and (24) choose the primal and dual step-lengths (Δ_P, Δ_D);

Step 7: Set the primal and dual variables as:

$$\begin{bmatrix} x^{(k+1)} \\ p^{(k+1)} \\ q^{(k+1)} \end{bmatrix} = \begin{bmatrix} x^{(k)} \\ p^{(k)} \\ q^{(k)} \end{bmatrix} + \Delta_P \begin{bmatrix} dx \\ dp \\ dq \end{bmatrix} \quad \begin{bmatrix} \beta^{(k+1)} \\ \alpha^{(k+1)} \\ \gamma^{(k+1)} \\ \lambda^{(k+1)} \end{bmatrix} = \begin{bmatrix} \beta^{(k)} \\ \alpha^{(k)} \\ \gamma^{(k)} \\ \lambda^{(k)} \end{bmatrix} + \Delta_D \begin{bmatrix} d\beta \\ d\alpha \\ d\gamma \\ d\lambda \end{bmatrix}$$

Step 8: Set $k = k + 1$ and go to step 3;

Step 9: Choose the neighborhood parameter, τ_b, close to the boundaries and find primal-dual suspicious measurements contaminated with GE such that:

$$\varpi = \{i \mid |\gamma_i q_i - \tau_b \mu| \geq 0 \ \& \ |\lambda_i p_i - \tau_b \mu| \geq 0\}$$

Step 10: Choose a threshold T. For the identified measurements with spike noise such that: $\{\varpi \mid |r_\varpi^w| > T$; $|r_\varpi^w|$ is weighted residual of suspicious measurement$\}$ then withdraw z_ϖ from measurement set and go to Step 1, else go to Step 11;

Step 11: END

VIII. Simulations for a 3-bus system case

This section presents some illustrative insight-providing results in a 3-bus DC-model system, shown in Fig. 1, with line reactances and the indication of the measurement plan for active powers. A basic power flow case was created; Gaussian noise was added to all measurements, with zero mean and standard deviation $\delta = 0.01\ pu$.

The results for SE were obtained, for a linearized DC model, with the proposed GCIP (c=13.4, σ=6). Voltage magnitudes for all the buses are 1 pu, the angle of bus 1 is 0 rad, and the angles of buses 2 and 3 are to be estimated.

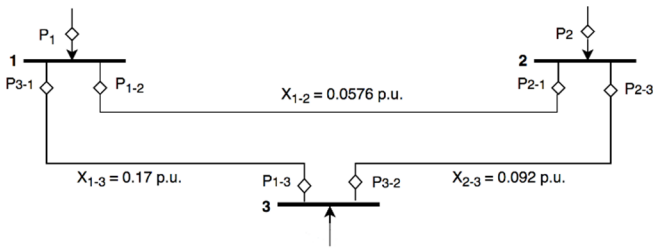

Fig. 1. The 3-bus DC test system.

A. Single gross error case

Over the basic data with noise, a gross error was superimposed on injected power in bus 2, with a value in the range of [25δ, 30δ] relative to the noise δ. Table II demonstrates the measurement scheme used for single GE case where z_f represents the power flow results and z are the measurements contaminated with noise. The corresponding results for different neighborhoods ($\mathcal{N}_2, \mathcal{N}_\infty, \mathcal{N}_\infty^-$) are then compared in Table III where errors are the difference between power flow and estimate and residuals are the difference between measured values and estimates. As suggested by experimental results, all neighborhoods could successfully identify the GE. It is clear though that infinity and pseudonorm neighborhoods could in fact detect the exact amount of added GE (error is almost zero). The results show the error is still insignificant in case of using 2-norm neighborhood in this case. Another measure to compare the results is a nodal voltage metric previously proposed in [17], given by:

$$M_V = \left(\frac{1}{n}\sum_{i=1}^{n} \left|\vec{V}_i^{\,est} - \vec{V}_i^{\,true}\right|^2\right)^{\frac{1}{2}} \tag{32}$$

where $\vec{V}_i^{\,est}$ and $\vec{V}_i^{\,true}$ are the estimated and reference (obtained from a previous power flow run) voltage values at the i-th bus. The smaller the index, the higher the performance of an estimator is. Based on [17] definition, the M_V index for L_2 neighborhood was calculated as 0.001 while this value is equal to 0.00002 for both $\mathcal{N}_\infty$ or $\mathcal{N}_\infty^-$.

TABLE II. MEASUREMENT SCHEME FOR ONE GE (BOLD)

z no.	1	2	3	4	5	6	7	8
z type	p_{1-2}	p_{2-3}	p_{1-3}	p_{2-1}	p_{3-2}	p_{3-1}	P_1	P_2
z_f	-0.43	0.10	-0.10	0.43	-0.10	0.10	-0.53	**0.53**
z	-0.44	0.11	-0.08	0.45	-0.09	0.08	-0.52	**0.78**

TABLE III. COMPARISON OF RESULTS FOR ONE GE (BOLD)

z	Estimates			Residuals			Errors		
no.	$\mathcal{N}_2$	$\mathcal{N}_\infty$	$\mathcal{N}_\infty^-$	$\mathcal{N}_2$	$\mathcal{N}_\infty$	$\mathcal{N}_\infty^-$	$\mathcal{N}_2$	$\mathcal{N}_\infty$	N_∞^-
1	-0.49	-0.43	-0.43	0.06	0.00	0.00	0.06	0.00	0.00
2	0.11	0.10	0.10	0.00	0.01	0.01	-0.01	0.00	0.00
3	-0.11	-0.10	-0.10	0.03	0.01	0.01	0.01	0.00	0.00
4	0.49	0.43	0.43	0.04	0.01	0.01	-0.06	0.00	0.00
5	-0.11	-0.10	-0.10	0.02	0.01	0.01	0.01	0.00	0.00
6	0.11	0.10	0.10	0.02	0.01	0.01	-0.01	0.00	0.00
7	-0.60	-0.53	-0.53	0.08	0.01	0.01	0.07	0.00	0.00
8	**0.60**	**0.53**	**0.53**	**0.18**	**0.25**	**0.25**	**-0.07**	**0.00**	**0.00**

B. Three gross errors case

Similarly, random noises from Gaussian distribution were added to power flows of line 1-2 and 2-3, as well as the injected power in bus 2 according to the presented values in Table IV. The results in Table V show that although GCIP is able to identify the GEs correctly in all the cases, however, infinity neighborhood is able to accurately grasp the amount of gross error in all three measurements (error is almost zero here as well). Comparing the M_v indexes as well, N_∞ has the smallest value of 0.00004 while this index is equal to 0.001, 0.004 for second and pseudonorm respectively. This proves a clear advantage of using L_∞ for GE identification.

TABLE IV. MEASUREMENT SCHEME FOR THREE GE (IN BOLD).

z no.	1	2	3	4	5	6	7	8
z type	p_{1-2}	p_{2-3}	p_{1-3}	p_{2-1}	p_{3-2}	p_{3-1}	P_1	P_2
z_f	**-0.43**	**0.10**	-0.10	0.43	-0.10	0.10	-0.53	**0.53**
z	**-0.72**	**0.39**	-0.09	0.44	-0.09	0.09	-0.53	**0.79**

TABLE V. COMPARISON OF RESULTS FOR THREE GE (IN BOLD).

z	Estimates			Residuals			Errors		
no.	$\mathcal{N}_2$	$\mathcal{N}_\infty$	$\mathcal{N}_\infty^-$	$\mathcal{N}_2$	$\mathcal{N}_\infty$	$\mathcal{N}_\infty^-$	$\mathcal{N}_2$	$\mathcal{N}_\infty$	$\mathcal{N}_\infty^-$
1	**-0.5**	**-0.4**	**-0.3**	**0.3**	**0.3**	**0.4**	**0.03**	**0.01**	**-0.09**
2	**0.1**	**0.1**	**0.1**	**0.3**	**0.3**	**0.3**	**0.00**	**0.00**	**0.02**
3	-0.1	-0.1	-0.1	0.0	0.0	0.0	0.00	0.00	-0.02
4	0.5	0.4	0.3	0.0	0.0	0.1	-0.03	-0.01	0.09
5	-0.1	-0.1	-0.1	0.0	0.0	0.0	0.00	0.00	-0.02
6	0.1	0.1	0.1	0.0	0.0	0.0	0.00	0.00	0.02
7	-0.6	-0.5	-0.4	0.0	0.0	0.1	0.03	0.00	-0.11
8	**0.6**	**0.5**	**0.4**	**0.2**	**0.3**	**0.4**	**-0.03**	**0.00**	**0.11**

IX. SIMULATIONS FOR A 30-BUS SYSTEM CASE

To confirm the preliminary results obtained with the infinity norm, a testbed was prepared based on the IEEE 30-bus system. A Monte Carlo sampling was used to create 500 complex scenarios. Each generator state was sampled as Up, or Down (probability of 0.18 to 0.23); similarly, loads could be in one of several demand levels or be disconnected with probability 0.01. This ensured a large diversity in the scenario set. Then, GEs were superimposed, in six experiments, by adding 0 to 5 simultaneous errors. The performance of the infinity norm $\mathcal{N}_\infty$ in identifying GEs is assessed by inspecting the occurrence of the following events in each scenario:

(1) False Positive (FP) or False Alarm: the measurement is labeled as a possible GE, but in fact it is not;
(2) False Negative (FN) or Missed Alarm: the measurement is labeled as clean of GE, but in fact it is contaminated.

A FN means that some GE remain undetected; a FP denotes that some measurement is labeled as possibly containing a GE while it does not. Tables VI summarizes the FN and FP rates (FNR, FPR) for all the scenarios in each set of experiment with 0 to 5 GEs. The extremely low values of these rates illustrate the high success of the IPM solver with neighborhood $\mathcal{N}_\infty$ in identifying multiple GE in a SE process. The stability of the FNR is also a strong indicator that the identification strategy based on $\mathcal{N}_\infty$ is very robust and almost never misses to pinpoint contaminated measurements, even in cases of multiple GE.

Table VI. FN and FP rates in 6 scenarios of multiple GE

GE nr.	0	1	2	3	4	5
FNR	0	0.052	0.039	0.039	0.030	0.039
FPR	0	0	0.001	0.007	0.003	0.010

X. CONCLUSION

The combination of an Interior Point Method with Generalized Correntropy can be used advantageously as a solver for the State Estimation process, not only to reach for the optimum regression solution but also as a tool to identify measurements contaminated with gross errors.

This second usage is an important novelty deserving highlighting. IPM-based models have been proposed in the past as solvers for the SE problem, but never as GE identifier. Yet, it is not a straightforward process – one of the requirements is the definition of a neighborhood to the Central Path.

This paper illustrates the fact that a neighborhood definition N_∞, based on a Chebyshev or L_∞ metric, provides the best results in gross error identification via IPM. This is not surprising, in fact, because this metric allows defining a broader neighborhood that other metrics, including the Euclidean – and gross errors are to be found apart from the Central Path of IPM.

The results presented help in shedding some light on why GCIP is a very promising and performing method for SE, even in the presence of multiple gross errors.

ACKNOWLEDGMENT

This work is integrated in the research effort supported by the ERDF – European Regional Development Fund (COMPETE 2020 Programme) and FCT- Fundação para a Ciên-cia e Tecnologia (project POCI-01-0145-FEDER-016731 INFUSE), Portugal. S. Pesteh recognizes the support of op-eration NORTE-08-5369-FSE-000043 co-funded by the Eu-ropean Social Fund (FSE) through NORTE 2020 – Programa Operacional Regional do NORTE.

REFERENCES

1. W. Liu, P. P. Pokharel, and J. C. Príncipe, "Correntropy: Properties and applications in non-Gaussian signal processing," IEEE Transactions on Signal Processing, 2007, 55(11): pp. 5286-5298.
2. V. Miranda, A. Santos, and J. Pereira, "State estimation based on correntropy: a proof of concept," IEEE Transactions on Power Systems, 2009, 24(4), pp. 1888-1889.
3. B. Chen, et al., "Generalized correntropy for robust adaptive filtering," IEEE Transactions on Signal Processing, 2016, 64(13), pp. 3376-3387.
4. M. Wright, "The interior-point revolution in optimization: history, recent developments, and lasting consequences," Bulletin of the American mathematical society, 2005, 42(1), pp. 39-56.
5. S. Roman, "Advanced Linear Algebra Third Edition," GRADUATE TEXTS IN MATHEMATICS, 2008, 1(135).
6. F. C. Schweppe, and D. B. Rom, "Power system static-state estimation, Part II: Approximate model," IEEE Transactions. on Power Apparatus and Systems, 1970(1), pp. 125-130.
7. C. E. Shannon, "A mathematical theory of communication," Bell system technical journal, 1948, 27(3), pp. 379-423.
8. Rényi, "Some fundamental questions of information theory," Selected Papers of Alfred Renyi, 1976, 2(174), pp. 526-552.
9. E. Parzen, "On estimation of a probability density function and mode," The annals of mathematical statistics, 1962, 33(3), pp. 1065-1076.
10. P. J. Huber, "Robust Statistics", ser. Wiley Series in Probability and Statistics. New York, USA: John Wiley & Sons, Inc., Feb. 1981.
11. J. Krstulovic, "Information Theoretic State Estimation in Power Systems, in Faculty of Engineering of the University of Porto. 2014.
12. K. Ponnambalam, V. Quintana, and A. Vannelli, "A fast algorithm for power system optimization problems using an interior point method," IEEE Transactions on Power Systems, 1992, 7(2), pp. 892-899.
13. J. A. Momoh, et al., "The quadratic interior point method solving power system optimization problems," IEEE Transactions on Power Systems, 1994, 9(3), pp. 1327-1336.
14. H. Wei, H. Sasaki, and T. Nagata, "An application of interior point quadratic programming algorithm to power system optimization problems," IEEJ Trans. on Power and Energy,1996, 116(2), pp. 174-180.
15. K. Clements, P. Davis, and K. Frey. "An interior point algorithm for weighted least absolute value power system state estimation," in IEEEPES Winter Meeting. 1991.
16. H. Wei, et al., "An interior point method for power system weighted nonlinear L/sub 1/norm static state estimation," IEEE Transactions on Power Systems, 1998, 13(2), pp. 617-623.
17. KEMA, "Metrics for determining the impact of phasor measurements on power system state estimation, DRAFT," Eastern Interconnection Phasor Project, January 2006.