Electric Power Systems Research 176 (2019) 105937

Contents lists available at ScienceDirect

Electric Power Systems Research

journal homepage: www.elsevier.com/locate/epsr

A new interior point solver with generalized correntropy for multiple gross error suppression in state estimation

Shabnam Pesteh[a], Hamed Moayyed[b], Vladimiro Miranda[a,c,*], Jorge Pereira[d], Victor Freitas[b,e], A. Simões Costa[e], J.B.A. London Jr[f]

[a] INESC TEC and Faculty of Engineering of the University of Porto, Portugal
[b] INESC TEC, Porto, Portugal
[c] INESC P&D Brasil, Brazil
[d] INESC TEC and Faculty of Economy of the University of Porto, Portugal
[e] UFSC - Federal University of Santa Catarina, Florianópolis, SC, Brazil
[f] USP - University of São Paulo, São Carlos School of Engineering, São Carlos, SP, Brazil

ARTICLE INFO

Keywords:
State estimation
Gross error processing
Information theoretic learning
Generalized
Correntropy
Interior point method
Leverage measurements
Critical measurements

ABSTRACT

This paper provides an answer to the problem of State Estimation (SE) with multiple simultaneous gross errors, based on Generalized Error Correntropy instead of Least Squares and on an interior point method algorithm instead of the conventional Gauss–Newton algorithm. The paper describes the mathematical model behind the new SE cost function and the construction of a suitable solver and presents illustrative numerical cases. The performance of SE with the data set contaminated with up to five simultaneous gross errors is assessed with confusion matrices, identifying false and missed detections. The superiority of the new method over the classical Largest Normalized Residual Test is confirmed at a 99% confidence level in a battery of tests. Its ability to address cases where gross errors fall on critical measurements, critical sets or leverage points is also confirmed at the same level of confidence.

1. Introduction

The problem of State Estimation (SE) departing from a data set contaminated with gross errors (GE) remains a challenging issue in the case of multiple errors and when GE fall on leverage points or critical measurements and sets. There are mathematical difficulties, but the sensitivity to bad data of the solving method generally used (Gauss–Newton iterations) stands as a basic hurdle to be coped with. This paper provides an innovative answer that deviates from the orthodox approach in two combined perspectives: the association of modelling the cost function based on Generalized Error Correntropy instead of Least Squares, with an interior point method solver instead of a Gauss–Newton algorithm. This blend achieves a remarkable level of accuracy in detecting, identifying and quantifying gross errors.

The definition of Generalized Correntropy is first introduced, departing from the concepts of Correntropy [1] developed within the framework of Information Theoretic Learning. Briefly, the objective of SE becomes no longer to minimize the (sum of the) squares of the errors, but to minimize the information content of the error distribution. The earliest proof of concept that such a model would work in Power Systems SE, based on Information Theory concepts, was described in Ref. [2], where Correntropy was first adopted.

Generalized Correntropy includes Correntropy as a particular case, but it has properties that allow a better matching between criterion and solving algorithms, as the paper will show. These properties were explored by designing an interior point algorithm [3] optimizing the Generalized Correntropy, whose derivation is presented in a following section of the paper.

The novel combination Generalized Correntropy + Interior Point (GCIP) has unique properties that allow the following:

(a) Detecting, identifying and quantifying/correcting multiple gross errors.
(b) Dealing naturally with errors (namely, gross errors) in leverage and critical measurements and sets.
(c) Remaining competitive in terms of computing effort for base cases with no gross errors, yet much more efficient than classical approaches for cases of multiple GE.

* Corresponding author at: INESC TEC and Faculty of Engineering of the University of Porto, Portugal.
E-mail addresses: spesteh@inesctec.pt (S. Pesteh), hamed.moayed@gmail.com (H. Moayyed), vmiranda@inesctec.pt (V. Miranda), jpereira@inesctec.pt (J. Pereira), victorfreitas89@yahoo.com.br (V. Freitas), simoes.costa@ufsc.br (A. Simões Costa), jbalj@sc.usp.br (J.B.A. London).

https://doi.org/10.1016/j.epsr.2019.105937
Received 6 April 2019; Received in revised form 18 June 2019; Accepted 7 July 2019
Available online 17 July 2019
0378-7796/

Such properties have sound mathematical foundations, with abridged reference in the following sections. The paper is mainly devoted to providing numerical evidence of such properties, which are highlighted in comparison with the performance of a conventional industrial approach (LNRT — Largest Normalized Residual test) for identifying and correcting measurements with gross errors. A following paper will dissect the mathematical foundations supporting the favorable properties of the novel method.

The paper includes didactic examples, to provide insight and allow plotting of functions, and results from extensive testing of the new model on an adaptation of the IEEE 30-bus system, in a diversity of load and generation scenarios. These tests satisfy two conditions: (a) a comparison, in terms of quality of results, with LNRT, and (b) a robust statistical testbed, instead of solely presenting individual cases of occasional success.

These results confirm the statistically significant superiority over the industrial conventional method for a confidence level of 99%, taken from a very large sample of experiments. A deeper analysis of the method as a diagnosis process confirms such assessment, counting incorrect and correct detections of gross errors (as “false alarms” and “missed alarms”) in a confusion matrix [4].

The GCIP model explores favorable properties that allow the method to swiftly deal with errors at leverage points and in critical measurements, something that has been looked at with caution, and even considered not-possible. The paper provides experimental confirmation that the method can deal with such problems. An investigation on this success led to the derivation of mathematical justifications for the behavior of GCIP in these cases. These proofs are quite extensive and will be provided in length in a separate publication — however, the experimental results are undeniable: with its formulation, GCIP naturally deals with problems that a Newton–Raphson model elicits.

The favorable properties of the GCIP approach — robustness and competitiveness in computing effort — make it a robust candidate to be integrated in an industrial environment.

2. State estimation with gross errors

The search for a suitable cost function that would allow SE from data sets contaminated with GE (also denoted as bad data) has more than four decades [5]. GE can also be seen as outliers, strongly deviating from a pattern that would be a consequence of the Kirchhoff Laws applied in the power networks. There is an early recognition that the Minimum Square Error (MSE) criterion, also denoted as Least Squares, has inconvenient properties in dealing with outliers — in fact, instead of ignoring these odd data, MSE puts unnecessary weight on them and therefore the estimation becomes deviated from a true pattern, in the attempt of compensating for such outliers.

Two main approaches can be identified in the literature, in the attempt to circumvent such difficulties and identify gross errors: we will denote them as: (a) the *a posteriori* approach, and (b) the *a priori* approach.

2.1. A posteriori approaches

In this approach, a first estimation is performed using an SE-like model. Most typically, applications explored by the industry are, in one way or another, referring to a classical, almost canonical model, best systematized in Ref. [6]. Then, a sort of diagnosis-and-correction procedure is applied, departing from the residuals (differences between measured and estimated values). The process tries to identify data corresponding to measurements contaminated with gross errors, in order to eliminate them from the data set and re-run a classical SE procedure again, using only “healthy data” plus some form of pseudo-measurements. The technique denoted as LNRT — Largest Normalized Residual Test [7] is one of the best known and used in industrial applications, following these principles. Even in recent publications, in a diversity of variants such as in Refs. [8,9], the *a posteriori* approach is commonly used.

2.2. A priori approaches

In this approach, the fundamental idea is to perform a SE directly on the original (contaminated) data, with a model that somehow, in a natural way, pushes gross errors out of the data sub-set contributing to the estimation. This process would not require, in principle, any two- or three-step procedure, as the *a posteriori* process does. To achieve such result, again two broad approaches were tried: (1) using artificial neural networks and (2) adopting cost functions that, in some way, could be related to a family of M-estimators [10].

2.2.1. Artificial neural networks (ANN)

Feed-forward artificial neural networks in SE were proposed as early as 1990 [11]. More recently, a different architecture for ANN was proposed, in the form of auto-associative ANN or autoencoders [12,13], and adopting Correntropy as the cost function of the training process.

An autoencoder learns a data pattern, when subject to training. Therefore, the cost function may also be interpreted as measuring a distance between data and a supporting manifold for the pattern learned from training. Maximizing Correntropy, the similarity between the output of the ANN and the pattern learned is measured under a specific non-Euclidean metric — the so-called CIM (Correntropy Induced Metric), in a smooth variation from metric L0 (indifference) to L2 (Euclidean) — while Least Squares represents a distance only measured in an Euclidean space (metric L2). Autoencoder approaches were shown to be successful, but display some drawbacks, among which the need for off-line training and the limitation in their practical size for a meaningful and accurate training, suggesting that they may find their niche of application in distributed architectures based on a mosaic of small intertwined local networks.

2.2.2. Non-Euclidean metrics

Attempts were made to replace the classical Least Squares criterion by a different criterion that would become insensitive to outliers, as early as in Refs. [14,15]. In perspective, these were attempts to address the problem from the point of view of the adoption of M-estimators. Their mathematical expressions move away from the quadratic form of Least Squares and, therefore, it may be said that they resort to non-Euclidean metrics when evaluating the distance between a vector of measurements and the supporting manifold for the solutions of the Kirchhoff equations; this is, in a way, a geometrical interpretation of the process. The non-quadratic criteria proposed were justified, in most cases, based on engineering judgment about the desired properties for a criterion that would be able to adequately deal with gross errors.

2.2.3. Information theoretic learning criteria

The adoption of a Maximum Correntropy Criterion (MCC) as the SE cost function and the interpretation it provides in terms of information content remaining in the error probability density function (pdf), shed new light over the State Estimation process. The basic principle is that a SE process should result in an error distribution with mode at zero and minimal Entropy. As Correntopy is related with the value of a pdf at zero, maximizing Correntropy achieves the minimization of Entropy as desired. This interpretation was first taken advantage of in Ref. [2], with the adoption of an MCC — Maximum Correntropy Criterion, being later extensively explored in Ref. [16]; in Ref. [17], this type of cost criterion was adopted but did not explore the Information Theoretic interpretation. In both latter works, a Gauss–Newton solver was developed. In Ref. [18], an evolution based on MCC with Laplace kernel functions and interior point method solver was presented. Finally, in Ref. [19] one finds a proposal to use GGD (generalized Gaussian density), as a family of kernel functions. However, the purpose of this latter work is modelling noise in measurements from a large sample of past

information. Also, a number of solvers should be ready to be used at the same time, which is not practical.

2.3. Difficulties with leverage points

Errors in some particular measurements were shown to be very influential and have the potential to mislead the computing into an erroneous solution — the location of such measurements received the label of leverage points. Early attempts with M-estimators [20,21] already had the intention of specifically down-weighting bad (with error) leverage points. M-estimators are known to be resistant to unusual observations (outliers), but sensitive to high leverage points; however, it is shown in Ref. [22]. that certain M-estimators with redescending influence functions [23] are resistant to high leverage outliers.

While the Weighted Least Squares (WLS) method provided some outlier rejection capability based on re-weighted residuals, an interesting competing method was also proposed, the Weighted Least Absolute Value (WLAV) estimator [24], with recent renewed interest, for instance in Refs. [25,26]. It was early shown that WLAV might fail even when there is a single bad leverage point and, in the event of multiple bad leverage points, it is still more vulnerable due to masking effects. Moreover, it is essential to be aware that since leverage points without gross errors (good leverage points) are capable of considerably improving the estimation reliability, the best solution is not always down-weighting all leverage points.

In order to effectively suppress bad data on leverage points, the Least Median of Squares (LMS) was extended to nonlinear power system problems [27], providing a substantial improvement to state estimators that were prone to a masking effect of multiple bad leverage points. In Ref. [28], an M-estimator was improved on the basis of projection statistics in the formulation of the so-called Schweppe-Huber GM-estimator. Later [29], had proposed a robust estimator based on maximum agreements among measurements. Furthermore [30], introduced a mixed integer nonlinear programming approach for a robust estimator. And recently [31] presented a Schweppe-type Huber generalized maximum-likehood estimator claimed to be robust in processing GE in leverage points. However, while the performances of these estimator approaches are good even for identifying gross errors on leverage points, the associated computational burden seems to be a limiting factor for their practical applications to large-scale power systems.

2.4. Difficulties with critical measurements

Critical measurements have been defined in relation to observability: a measurement is critical when its loss makes the system unobservable. This concept is not put in question. However, there is a commonly repeated assertion that “the residuals of critical measurements will always be zero, and therefore errors in critical measurements cannot be detected or identified”. Also, it has been argued that when a GE falls on a measurement in a critical set, it is possible to detect that there is a GE somewhere but it is not possible to identify which measurement is contaminated.

This is true, given that the model behind the calculation is a classical Least Squares model. But if a different mathematical representation is adopted, so that the calculation of residuals exhibits different properties, it may be possible that the variance for an estimated residual in a critical measurement is not zero, even if the variances of the measurement and its estimated value are equal — it depends on covariances not being null. If so, then this property may be used to identify gross errors in critical measurements. It is tantamount to saying that the property of “error undetectability” belongs to the method, not to the measurement.

2.5. Difficulties with Gauss–Newton solvers

In geometrical terms, Correntropy is similar to the Welsch M-estimator [32] and shares its advantages and drawbacks. One of the most unpleasant characteristics is that such cost functions exhibit a possibly large number of local optima. This rendered the use of Gauss–Newton solvers problematic, because they demand specific necessary conditions to converge to the global optimum. It is however not possible to guarantee that the traditional flat start first iterative solution (all voltages equal to 1 p.u. and all angles equal to 0) enforces such conditions. This is why the proof of concept in Ref. [2] resorted to a meta-heuristic to find the correct solution for the SE problem in an AC model.

The strategy to circumvent this problem, using as initial iteration point the MSE solution and forcing a smooth transition between a Euclidean metric and the CIM, was not 100% successful, and it has been possible to demonstrate that there are cases of failure of such processes [16]. The technique became competitive with the LNRT approach, but not significantly better if not aided by a complementary process.

This is why in Ref. [16] a mix strategy was adopted: ANN in the form of denoising autoencoders [33] provided a filtering and correction of the measurement vector, generating a suitable starting point for the iterations of a subsequent Gauss–Newton solver acting under a maximum correntropy criterion (MCC). This led to a 100% recognition of gross errors, when two simultaneous large errors are present in the measurement set, clearly beating the LNRT method — but had the drawback of demanding the adoption of a multitude of autoencoders, with the burden of fine-tuning their training and of retraining each time any topologic change would occur in the network.

In any case, for all approaches, a quick degradation of performance can still be witnessed when more than two gross errors are present.

3. Generalized correntropy

Entropy is a measure of information content of a probability density function. Renyi’s entropy [34] has proven to allow the development of working algorithms in an easier fashion than the definition of Shannon [35]. Correntropy [1] is a generalized measure of similarity between two pdf, with a strong relation with entropy. Given the two scalar random variables X and Y, it is defined by

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

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 [36] is given by:

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

As referred to in Section 2.2.3, the correntropy of the distribution of residuals was introduced in Ref. [2] as the cost function for the state estimation problem. This allows the interpretation that the SE process aims at discovering the set of estimated measurements whose distribution is as similar as possible to the distribution of measurements.

A very useful generalization was produced with the adoption of the GGD (generalized Gaussian density), as proposed in Ref. [37], plotted in Fig. 1, which is given by

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

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 $\sigma = \zeta\sqrt{\Gamma(1/c)\Gamma(3/c)}$.

Special cases of this GGD include: $c = 1$ — Laplace density function; $c = 2$ — Gaussian density function: $c \to +\infty$ – uniform distribution over $(-\sigma, \sigma)$. The Generalized Correntropy (GC) estimator for a finite

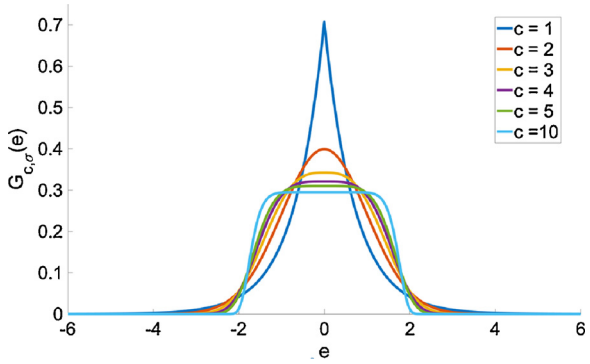

Fig. 1. GGD graph for several values of c when $\zeta = 1$.

discrete sample becomes

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

Associated with the GC, there is the generalized correntropy induced metric (GCIM), which may behave like different norms depending on the parameters adopted and the region of analysis. Accordingly, GCIM may vary from L_∞ to L_0, while the metric induced by a Gaussian-based Correntropy only varies from L_2 to L_0. This property proves to be helpful in SE with gross errors. In particular, the capacity to apply an L_0 metric (indifference in distance) in the case of outliers is what is needed, because their residuals can be pushed to whatever value is necessary without disturbing the objective function evaluation, thus allowing the remaining residuals to be minimized without contamination. Furthermore, the L_∞ metric is indifferent to all components in a residual vector, except the most significant one — this helps in separating outliers by providing freedom to the rearrangement of residual vector components.

In summary, this paper proposes the adoption of 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.

4. Interior point solver

The difficulties previously met with Correntropy (and smooth M-estimators of the kind), for problems with GE, derive from problem formulation — as an unconstrained non-linear program — as well as parametrization and the algorithm used — classical Gauss–Newton iterations. This formulation is plagued with local optima and the traditional iteration flat start solution does not work in all cases, as it may be out of the region of necessary conditions for the Newton process to converge. The information theory behind the process is robust but the algorithmic difficulties cannot provide full convergence guarantee and some rate of failure is verified [16]. To circumvent such difficulties, this paper reports the successful adoption of an Interior Point method (IPM) solving strategy.

Since the 80's one finds suggestions for using IPM in power systems problems [38–40]; in particular, several researchers applied IPM to SE, usually associated with WLAV (weighted least absolute value) criteria [41–44]. Some adepts of the WLAV criteria claimed that it had some good behavior in the rejection of outliers, but this has not been sufficiently quantified. In terms of computing efficiency, IPM was shown to be competitive.

In the following sections, a formulation of an IPM solver for the MGCC criterion will be detailed. This combination of MGCC with IPM forms the GCIP model for power systems state estimation.

4.1. The state estimation model based on GC

The formulation of the state estimator based on GC follows paths similar to previous work, as referred to above. The following optimization problem is set up:

$$\max J(x) = \frac{1}{m}\sum_{i=1}^{m} \frac{c}{2\sigma\Gamma(1/c)} exp\left(-\left|\frac{r_i}{\sigma}\right|^c\right) \tag{5}$$

$$\text{subject to } \begin{cases} g(x) = 0 \\ r = z - h(x) \end{cases}$$

where:

z is the vector of measurements (mx1)

x is the vector of state variables (nx1)

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)

$g(x) : R^n \rightarrow R^\Omega$ are zero-injection equality constraints, Ω is the number of equality constraints.

In order to be relieved of the absolute value, introducing inequality constraints for the residual term is required. These may be transformed in equalities by adding non-negative slack variables.

4.2. Producing a standard cost function

Define a new variable vector $s \in R^m$ such that

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

By introducing two slack variable vectors such that $u,\ v \in R^m \ge 0$, and defining 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

$$r_i = p_i - q_i \tag{7}$$

$$s_i = p_i + q_i \tag{8}$$

Considering the upper bound of Eq. (6) which is defined by Eq. (8), a new equivalent system based on standard form of non-linear programming is obtained:

$$\max J(x) = \frac{1}{m}\sum_{i=1}^{m} \frac{c}{2\sigma\Gamma(1/c)} exp\left(-\left(\frac{p_i + q_i}{\sigma}\right)^c\right) \tag{9}$$

$$\text{subject to } \begin{cases} f(x) = 0 \\ g(x) = 0 \\ p,\ q \ge 0. \end{cases}$$

where $f(x) = z - h(x) - p + q$.

4.3. Primal-dual interior point method

Introducing Lagrange multipliers $\beta \in R^\Omega$; λ, γ and $\alpha \in R^m$, we can construct the following Lagrangian function associated with Eq. (9)

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

Then, we can derive the following Karush–Khun–Tucker (KKT) conditions for system (9) as follows

$$L_{\gamma_i} \equiv \gamma_i q_i = 0 \tag{11}$$

$$L_{\lambda_i} \equiv \lambda_i p_i = 0 \tag{12}$$

$$L_{q_i} \equiv J(x)\left(\frac{-c(p_i + q_i)^{c-1}}{m\sigma^c}\right) exp\left(-\left(\frac{p_i + q_i}{\sigma}\right)^c\right) - \alpha_i - \gamma_i = 0 \tag{13}$$

$$L_{p_i} \equiv J(x)\left(\frac{-c(p_i+q_i)^{c-1}}{m\sigma^c}\right)exp\left(-\left(\frac{p_i+q_i}{\sigma}\right)^c\right)+\alpha_i-\lambda_i=0 \tag{14}$$

$$L_\beta \equiv g(x)=0 \tag{15}$$

$$L_\alpha \equiv f(x)=0 \tag{16}$$

$$L_x \equiv \nabla g(x)^T\beta-\nabla h(x)^T\alpha=0 \tag{17}$$

Given the special nonlinearity in complementary conditions (11) and (12), it is not possible to solve directly the above KKT equations by Newton's method. For example, by Newton's method, Eq. (10) solves as:

$$\gamma_i dq_i+q_i d\gamma_i=-\gamma_i q_i \tag{18}$$

If $q_i^{(k)}$ in Eq. (11) becomes zero at the k -th iteration, dq_i^k is equal to zero, from Eq. (15), and hence $q_i^{k+1}=q_i^k+dq_i^k=0$. This means, once q_i falls on the boundary of a feasible region, it is stuck at that point, preventing the convergence of the algorithm. It is necessary to modify the Newton formulation in a way that zero variables become nonzero in subsequent iteration. This can be done by replacing the complementarity equation with a perturbed complementary [45]. Thus, we introduce a perturbation factor $\mu>0$ to relax (11) and (12) as:

$$L_{\gamma_i}^{\mu} \equiv \gamma_i q_i-\mu=0 \tag{19}$$

$$L_{\lambda_i}^{\mu} \equiv \lambda_i p_i-\mu=0 \tag{20}$$

The 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 [45].

By applying the Newton method into the KKT optimality conditions, we can express the correction equations as:

$$\gamma_i dq_i+q_i d\gamma_i=-L_{\gamma_i}^{\mu} \tag{21}$$

$$\lambda_i dp_i+p_i d\lambda_i=-L_{\lambda_i}^{\mu} \tag{22}$$

$$e_i dp_1+e_i dq_1-d\alpha_i-d\gamma_i=-L_{q_i} \tag{23}$$

$$e_i dp_1+e_i dq_1+d\alpha_i-d\lambda_i=-L_{p_i} \tag{24}$$

$$\nabla g(x)dx=-L_\beta \tag{25}$$

$$-\nabla h(x)dx-dp+dq=-L_\alpha \tag{26}$$

$$(\nabla^2 h(x)\beta-\nabla^2 g(x)\alpha)dx+\nabla g(x)^T d\beta-\nabla h(x)^T d\alpha=-L_x \tag{27}$$

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

$$e_i=\frac{J(x)}{m}\left(\frac{-c(c-1)(p_i+q_i)^{c-2}}{\sigma^c}\exp\left(-\left(\frac{p_i+q_i}{\sigma}\right)^c\right)+\left(\frac{c^2(p_i+q_i)^{2c-2}}{\sigma^{2c}}\right)\exp\left(-\left(\frac{p_i+q_i}{\sigma}\right)^c\right)\right) \tag{28}$$

From Eqs. (20–23), 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

$$dq_i=n_{1i}d\alpha_i+t_{1i} \tag{29}$$

$$dp_i=n_{2i}d\alpha_i+t_{2i} \tag{30}$$

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

$t_{1i}=-a_i(q_i L_{q_i}+L_{\gamma_i}^{\mu})-b_i(p_i L_{p_i}+L_{\lambda_i}^{\mu})$

and

$t_{2i}=-c_i(q_i L_{q_i}+L_{\gamma_i}^{\mu})-d_i(p_i L_{p_i}+L_{\lambda_i}^{\mu})$.

Substituting Eqs. (29) and (30) into Eq. (26) then

$$\nabla h(x)dx+Ad\alpha=\nu \tag{31}$$

where $A\in R^{m\times m}$ is a diagonal matrix, with the elements

$$A_i=-n_{1i}+n_{2i}v=z-h(x)-p+q+t_1+t_2$$

and $t_1,\ t_2\in R^m$ as defined in Eqs. (29) and (30). Then, according to Eqs. (25), (27) and (31) and ignoring the second derivative terms in Eq. (27) just as in the WLS method, the reduced correction Eq. (24) is obtained,

$$\begin{bmatrix}G & 0 & 0\\ 0 & G^T & -H^T\\ H & 0 & A\end{bmatrix}\begin{bmatrix}dx\\ d\beta\\ d\alpha\end{bmatrix}=\begin{bmatrix}-L_\beta\\ -L_x\\ \nu\end{bmatrix} \tag{32}$$

where $G=\nabla g(x)$; and $H=\nabla h(x)$. The values of $d\boldsymbol{x}$, $d\boldsymbol{\beta}$, and $d\boldsymbol{\alpha}$ can be obtained from Eq. (32); then, Eqs. (29) and (30) give $d\boldsymbol{p}$ and $d\boldsymbol{q}$. Eventually, $d\boldsymbol{\gamma}$ and $d\boldsymbol{\lambda}$ can be obtained from Eqs. (23) and (24).

We can calculate the primal and dual step sizes by Ref. [44]

$$\Delta_P=0.9995\min\{\min(-q_i/dq_i:\ q_i<0;\ -p_i/dp_i:\ p_i<0),\ 1\} \tag{33}$$

$$\Delta_D=0.9995\min\{\min(-\gamma_i/d\gamma_i:\ \gamma_i<0;\ -\lambda_i/d\lambda_i:\ \lambda_i<0),\ 1\} \tag{34}$$

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

4.4. The role of central path in the identification of GE

The new algorithm benefits from a very important feature occurring in 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 $\mathscr{C}$ 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 $\mathscr{C}$ to find a better solution. Therefore, directions calculated from any point in the neighborhood make a progress toward the solution set. The neighborhood of $\mathscr{C}$ is defined by

$$\mathscr{N}_{1\tau}=\{(\gamma,\ q,\ \mu)\in\mathscr{F}|\gamma_i q_i\geq\tau\mu\ \text{for all}\ i=1,2,...,m\} \tag{35}$$

$$\mathscr{N}_{2\tau}=\{(\lambda,\ p,\ \mu)\in\mathscr{F}|\lambda_i p_i\geq\tau\mu\ \text{for all}\ i=1,2,...,m\} \tag{36}$$

where $\tau\in(0,1)$ defines the neighborhood parameter and $\mathscr{F}$ represents the primal-dual feasible set. 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}:

$$r^w=R^{-1/2}r \tag{37}$$

where R is the covariance matrix of the measurement error vector. Considering weighted residuals to detect GEs does not bring any difficulty in case of critical and leverage outliers, as the residuals are not necessarily small for such points in GCIP formulation. In addition, although for critical measurements there is no redundant information, zero injection and inequality constraints used in the formulation work the same way as pseudo measurements and provide an extra redundancy for the system. These results are fully demonstrated in Ref. [49] and will be presented in a coming publication.

Notice that not considering the normalized residuals as a verification method in here is due to the fact that measurement error distributions are not constrained by any assumption and can be either Gaussian or non-Gaussian. By comparing the absolute value $|r^w|$ of the suspicious measurement identified by the primal-dual criterion with a defined threshold T, the measurement is either marked as GE, if $|r^w|>T$ or non-GE, otherwise. Measurements labeled as GE are then withdrawn from the measurement set and the states are re-estimated.

Notice that the statistical significance is used in here to determine

the T value in the sense that the algorithm is prepared to ignore all results that fail to reach this standard. Statistical significance is often expressed in multiples of the standard deviation of a normal distribution. Such criterion is still independent from any Gaussian or non-Gaussian assumptions for the measurement error distributions; therefore, based on central limit theorem, even if the distributions are not normal, the estimates of the means are going to be close to normally distributed. This paper sets significance thresholds at a strict level of 5σ as it corresponds to the probability of 0.00006% for two-sided p-value (+/− residuals). In fact, p-value corresponds to the probability of detecting a measurement as GE when it is in fact a sound measurement, given that the null hypothesis is true [46].

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 [47].

4.5. Special favorable properties of GCIP

The combination of Generalized Correntropy with an Interior Point method (GCIP) has important properties that make it a powerful and robust method, superior to many other approaches. Among these properties, the following are of the utmost importance:

A There is a theoretical optimum value for the shape parameter c, and therefore there is no need for trial and error experiments. It can be proved that for c = 13.4, the strict convexity range of GC objective function is in its maximum value and so the solutions found by the algorithm in the search space are global solutions.

B The GCIP solving process, as designed, naturally introduces a distortion or transformation of the search space in a way that errors in measurements in leverage points appear no longer causing spurious contaminations, disturbing the estimation process. It can be shown that, in GCIP, all data points are forced to belong to a large cluster of evenly spread measurements, by a coordinate transformation that is embedded in the method. This causes an effect similar to the method argued in Ref. [48], but without imposing external transformations or adding extra procedures.

The coordinate transformation induced by the application of GCIP leads to defining a new hat matrix ***K'***, distinct from the hat matrix usually defined in WLS, i.e. its residual sensitivity matrix. ***K'*** is idempotent, non-symmetrical, and GCIP adaptively makes the value of any diagonal element of ***K'*** as close as possible to its expected value. The scaling factors used are not prescribed as suggested in Ref. [48] and are assigned according to the spread of the row vectors in matrix ***H***. Therefore, the method naturally evolves in a way that mitigates the effect of leverage measurements.

A The identification of errors in critical measurements becomes possible, because problems in solving with the WLS/Newton–Raphson approach disappear with GCIP. The co-variance matrix of the estimated measurement vector can be shown to derive from the new ***K'*** hat matrix and the scaling and translation introduced by the method. It can then be demonstrated that, contrary to WLS, the estimated residual of a critical measurement is not zero in GCIP, which allows the identification of gross errors in such measurements.

The authors have developed detailed mathematical derivations that justify the above three properties. Because they are lengthy, they are omitted from this paper and will be published in a separate paper, solely devoted to the favorable features of GCIP and their mathematical foundations.

The experimental results described in this paper provide evidence that such properties become useful in dealing with gross errors occurring is such classical difficult cases. It must be recalled that the concepts of leverage and critical measurements evolved from the classical Least Squares criterion + Newton–Raphson optimizer and are a consequence of such formulation and of interpretations over it.

This paper presents, in the form of numerical results from a sample of hundreds of experiences, referred to in Section 6 below, solid evidence of results in line with the expectations generated by the properties above. In particular, Sections 6.5 and 6.6 present the experiments and result analysis for special cases of gross errors in leverage points and critical measurements. These results translate the favorable properties of GCIP, which have a provable mathematical foundation, into numeric experimental evidence.

Table 1
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 ($\text{Gap} \equiv \gamma^T q + \lambda^T p$). 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 $\mu = \rho . \frac{\text{Gap}}{2m}$;

Step 5: Solve the reduced correction Eq. (32) for $[dx,\ d\beta,\ d\alpha]^T$ and then compute $[dp,\ dq, d\gamma,\ d\lambda]^T$;

Step 6: From Eqs. (33) and (34) choose the primal and dual step-lengths (Δ_P, Δ_D);

Step 7: Set the primal and dual variables as:

$$\begin{bmatrix} x^{(k+1)} \\ p^{(k+1)} \\ q^{(k+1)} \end{bmatrix} = \begin{bmatrix} x^{(k)} \\ p^{(k)} \\ q^{(k)} \end{bmatrix} + \Delta_P \begin{bmatrix} dx \\ dp \\ dq \end{bmatrix}$$

$$\text{and} \quad \begin{bmatrix} \beta^{(k+1)} \\ \alpha^{(k+1)} \\ \gamma^{(k+1)} \\ \lambda^{(k+1)} \end{bmatrix} = \begin{bmatrix} \beta^{(k)} \\ \alpha^{(k)} \\ \gamma^{(k)} \\ \lambda^{(k)} \end{bmatrix} + \Delta_D \begin{bmatrix} d\beta \\ d\alpha \\ d\gamma \\ d\lambda \end{bmatrix}$$

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: $\{\varpi \mid |r^w_\varpi| > T\ ;\ |r^w_\varpi|$ 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

4.6. GCIP estimator algorithm

The detailed GCIP estimator algorithm is summarized in Table 1. Although following the general logic steps of previous approaches [41,44], it includes important improvements that allow for flexibility in solving different problems from changing the shape parameter of the objective function. 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. This method is based on the interesting property of the proposed algorithm in which the neighborhood excludes the points that are too close to the boundary of the non-negative orthant (hyperoctant). Hence, the measurements that are located near the boundaries are suspected to be contaminated with GEs. Weighted residuals are then used in order to verify such suspicion so that the data can be cleaned before SE is run one more time. Experimental results confirm the efficiency introduced by this approach.

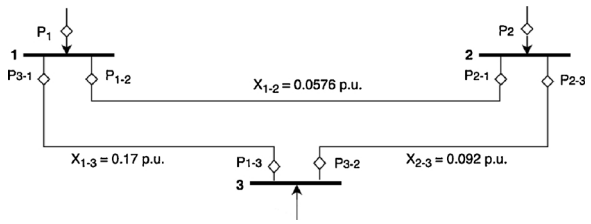

Fig. 2. The 3-bus DC test system.

5. Simulations for a 3-bus system case

5.1. Testbed

This section presents some illustrative insight-providing results in a 3-bus DC-model system, shown in Fig. 2, 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 $\delta = 0.01\,pu$. but not allowing values so large that could be seen as gross errors.

The results for SE were obtained, for a linearized DC model, with the classical WLS and with GCIP ($c = 13.4$, $\zeta = 6$). Voltage magnitudes for all the buses are 1 pu, the angle of Bus 1 of 0 rad, and the angles of Buses 2 and 3 are to be estimated. For DC experiments, the solutions were reached in a single iteration (internal loop) for both methods. Notice that the SE results of GCIP in the following section do not show the impact of identification and cleaning GEs yet.

5.2. Three simultaneous gross errors

The didactic experiment consists of establishing a stress test (3 GE in a 3-bus system), with the GE having an amplitude of 25δ relative to the noise component.

Table 2 shows how GCIP detected and corrected all three errors, and how its solution is very close to the power flow solution (taken as reference), while the WLS process failed (as expected).

The relationship between the objective function $J(x)$ of the GCIP estimator, the constraints and x (voltage angles of bus 2 and bus 3) is depicted in Fig.3 ($\zeta = 6$). In Fig. 4, the projection on the (θ_2, θ_3) space shows the GCIP solution (x) very close to the reference power flow solution and on one of the lines that jointly represent the feasible region of the IP model, while the WLS solution is apart.

6. Results with a 30-bus system

6.1. Testbed

A comprehensive testbed for algorithm behavior was prepared based on the IEEE 30-bus system from Matpower database [50] with a load added to bus 11, by changing generation and load values.

A Monte Carlo sampling was used to create 500 complex scenarios.

Table 2
Comparison of results for three GE (in gray).

Measure-ments	Power flow	Measured values	Estimates		Residuals		Errors	
			GCIP	WLS	GCIP	WLS	GCIP	WLS
p_{1-2}	**-0.43**	**-0.72**	**-0.46**	**-0.89**	**0.26**	**0.17**	**0,03**	**0,46**
p_{2-3}	**0.10**	**0.39**	**0.10**	**0.24**	**0.29**	**0.15**	**0**	**-0,14**
p_{1-2}	-0.10	-0.09	-0.10	-0.13	0.01	0.04	0	0,03
p_{2-1}	0.43	0.44	0.43	0.34	0.02	0.10	0	0,09
p_{3-2}	-0.10	-0.09	-0.10	-0.24	0.01	0.15	0	0,14
p_{2-1}	0.10	0.09	0.08	0.06	0.01	0.04	0,02	0,04
P_1	-0.53	-0.53	-0.56	-0.60	0.04	0.07	0,03	0,07
P_2	**0.53**	**0.79**	**0.56**	**0.79**	**0.23**	**0.00**	**-0,03**	**-0,26**

In grey/bold: measurements/results directly affected by gross error.

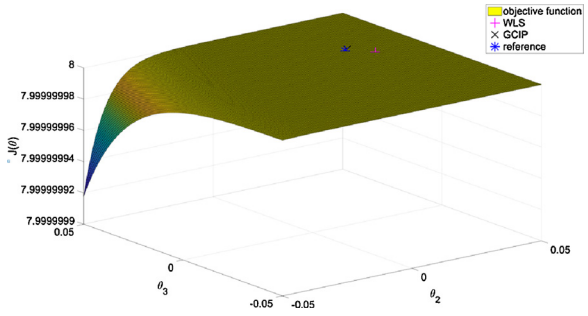

Fig. 3. 3D view of the objective function of GCIP for the 3-bus DC system in the flat start ($\theta_2 = 0$, $\theta_3 = 0$) area in the problem with three GEs.

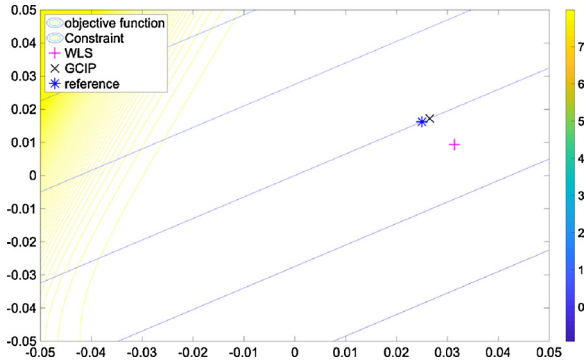

Fig. 4. 2D projection of the objective function and constraints in Fig. 3. The lines correspond to the feasible area in the IP model. The GCIP solution × is over a constraint and close to the reference solution ∗, while the WLS solution + is far away.

Each generator could be sampled as in one of the two states: operating or failed (Up, or Down with a high probability of 0.18–0.23); 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.

The definition of measurement plans for each scenario as well as the location of specific gross errors was dealt with care, to ensure fairness and avoid bias in the experiment outcomes. Four measurement plans were defined, with different redundancy levels Λ_r (ratio between actual number of measurements and number of state variables). The measurement plans were built to have critical measurements or critical sets appear in controlled percentages, which are higher for reduced redundancy rates. In each case, a power flow was run, Gaussian noise was randomly added to measurements with zero mean, standard deviation of δ, $0.02\,pu$ for active and reactive power and $0.003\,pu$ for voltage magnitudes (noise sampled under other distribution could be used, not relevant to the essence of the results presented).

Then, GEs were superimposed, in six experiments, by adding 0–5 simultaneous errors. This addition was made randomly but under supervision, making sure that the percentage of cases of GE falling in critical sets and leverage points was in line with the theoretical probability of this happening (counting the number of possible measurement combinations and the number of combinations falling in critical sets). The process of identifying of leverage points is described below, in Section 6.5.

Per experiment (for each Λ_r), 500 scenarios were built.

For each scenario, GCIP ($c = 13.4$, $\zeta = 6$) iterated twice (steps 2–10 in Table 1) until convergence, while WLS + LNRT was allowed to iterate with progressive removal of the largest suspect measurement in each run (removal threshold set to 4 as commonly done in SE problems), until a stopping criterion is met, or convergence failed. In our experience, parameter values of c fixed at values close enough to 13.4 (e.g., 12) do not provide a discernable difference in the quality of results.

Table 3
Average iteration number for 500 cases in each set of experiments with 0–5 GE in the four redundancy-level scenarios.

Λ_r	3.2		3		2.8		2.6	
No. GE	GCIP	LNRT	GCIP	LNRT	GCIP	LNRT	GCIP	LNRT
0	2	6	2	6	2	6	2	6
1	5	13	5	13	5	13	5	13
2	5	20	5	20	5	20	5	21
3	5	27	5	28	5	28	5	28
4	5	36	5	36	5	36	5	37
5	5	46	5	45	5	45	5	46

The results were then analyzed by examining the comparative performance of both methods, in the whole set of scenarios as well as on selected subsets, to analyze specific cases of GE falling on critical measurements or critical sets, or in leverage points.

All tests were run on a MacBook Pro machine equipped with a two Intel Core i5, CPU clocked at 2.5 GHz and 8 GBs of RAM. It used macOS 10.13.5 and 64-bit version of Matlab 2017a as well as Matpower package for steady-state power system simulations [50].

6.2. Performance in the complete set of scenarios

As Tables 3 and 4 show, GCIP is very stable in terms of the number of iterations and computing time, and both methods exhibit comparable running times, for average results over the complete set of scenarios. GCIP exhibits no disadvantage in terms of computing burden.

These numbers of iterations are computed until the algorithm stops, irrespective of achieving convergence or not, and irrespective of reaching the correct solution or not.

An observed valuable property of the GCIP algorithm lies in that the number of iterations is almost insensitive to the degree of difficulty of the problem (as measured by the number of GE involved). This is clearly shown in Fig. 5, illustrating the average and maximum number of iterations in 500 cases for each method, for the case of a (good) redundancy of $\Lambda_r = 3.2$ (similar plots for other cases).

GCIP iterations never exceed 7 internal loops, while in LNRT there are scenarios reaching 117 internal iterations, when there are 5 gross errors present in the sampled measurement set. Table 5 presents the maximum number of iterations for all the tested redundancy levels.

6.3. Estimation accuracy in the complete set

To evaluate the performance of an estimator we made use of a nodal voltage metric previously proposed in Ref. [51], given by:

$$M_V = \left(\frac{1}{n} \sum_{i=1}^{n} |\vec{V}_i^{est} - \vec{V}_i^{true}|^2 \right)^{\frac{1}{2}} \tag{38}$$

where $\vec{V}_i^{est}$ and $\vec{V}_i^{true}$ are the estimated and reference (obtained from a previous power flow run) complex voltage values at the i-th bus. The

Table 4
Average CPU time (s), for 500 cases in each set of experiments with four redundancy-level scenarios.

Λ_r	3.2		3		2.8		2.6	
No. GE	GCIP	LNRT	GCIP	LNRT	GCIP	LNRT	GCIP	LNRT
0	0.06	0.02	0.08	0.02	0.06	0.02	0.08	0.07
1	0.10	0.06	0.12	0.04	0.12	0.09	0.14	0.08
2	0.11	0.06	0.13	0.09	0.14	0.10	0.14	0.10
3	0.11	0.07	0.13	0.11	0.14	0.14	0.14	0.11
4	0.12	0.08	0.15	0.14	0.14	0.16	0.14	0.13
5	0.12	0.16	0.15	0.16	0.14	0.19	0.14	0.17

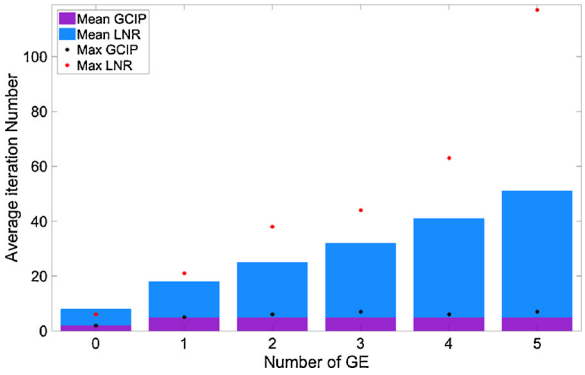

Fig. 5. Average and maximum iteration number for 500 cases in each set of experiment with 0–5 GE with measurement redundancy of 3.2.

Table 5
Maximum iteration number for 500 cases in each set of experiment with 0–5 GE with four redundancy-level scenarios.

Λ_r	3.2		3		2.8		2.6	
No. GE	GCIP	LNR	GCIP	LNR	GCIP	LNR	GCIP	LNR
0	2	6	2	6	2	6	2	6
1	5	21	5	23	5	21	6	21
2	6	38	6	30	6	44	6	42
3	7	44	6	41	6	51	7	52
4	6	63	7	58	6	66	7	73
5	7	117	7	82	7	83	7	102

M_V index was calculated for 500 cases in each set of experiments with 0 to 5 GE with four redundancies scenarios.

As a first remark, GCIP converged in 100% cases, in all experiments and for all levels of redundancy rate Λ_r. On the other hand, the LNRT procedure failed to converge in 212 out of 500 cases for the high redundancy level of 3.2 (the progressive withdrawal of suspicious measurements led at some point to an unsolvable problem).

The failure of the LNRT procedure is known to be due to several factors: (i) GEs in measurements in low redundancy scenarios (critical measurements or pertaining to critical sets of measurements); (ii) interactive multiple GEs; (iii) GE in highly influential measurements, i.e., affecting the convergence of the SE process, the so-called leverage point measurements [52].

A 99% confidence interval for the convergence rate was calculated based on the dimension of the sample, leading to the representation in Fig. 6 of confidence intervals in all experiments. The LNRT procedure, at a confidence level of 99%, will fail to converge in at least 53% of the cases, in the experiment conditions.

To produce a fair comparison, in terms of accuracy and for all these

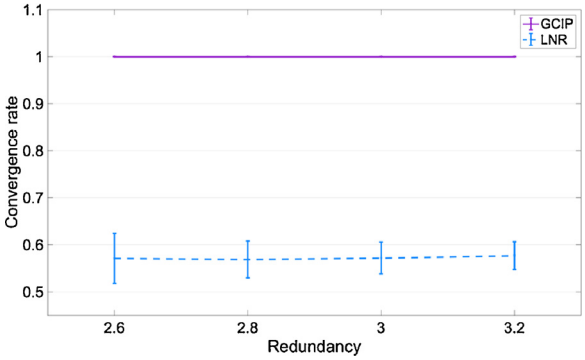

Fig. 6. Confidence interval for convergence rate for 500 cases in each set of experiment with 0–5 GE with four redundancies scenarios.

Table 6
Average M_V index for 500 cases in each set of experiment with 0–5 GE with four redundancy-level scenarios.

Λ_r	3.2		3		2.8		2.6	
No. GE	GCIP	LNRT	GCIP	LNRT	GCIP	LNRT	GCIP	LNRT
0	0.005	0.005	0.005	0.005	0.005	0.005	0.005	0.006
1	0.005	0.008	0.005	0.009	0.006	0.009	0.007	0.009
2	0.006	0.010	0.005	0.011	0.006	0.011	0.009	0.014
3	0.007	0.013	0.007	0.013	0.007	0.013	0.011	0.016
4	0.007	0.014	0.008	0.015	0.009	0.017	0.014	0.020
5	0.008	0.016	0.008	0.018	0.010	0.018	0.019	0.021

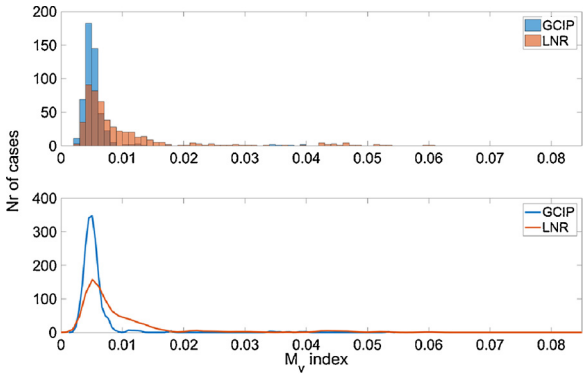

Fig. 7. Estimates for the density of M_V index for $\Lambda_r = 3.2$, 2 GEs.

cases, the WLS estimation results obtained, or the results before the verification of the lack convergence conditions, were retained as the LNRT outcome.

This produced the results in Table 6, showing the average M_V index value obtained, with a net superiority of the GCIP algorithm.

Fig. 7 is also enlightening. The top plot shows the histogram of M_V indexes for both methods when we have a redundancy level of $\Lambda_r = 3.2$ and only 2 random gross errors. The bottom plot represents their probability density estimates, calculated by a kernel smoothing function method [53]. The estimate is based on a Gaussian kernel function, and is evaluated at equally-spaced points, that cover the range of the M_V indexes for each scenario and per SE approach. These estimates allow one to assess, for each method, the probability (risk) of reaching an M_V index worse than a specified threshold.

Tables 7 and 8 show the numerical value of such risk, associated with thresholds 0.01 and 0.02, by computing the area under the estimated density function. The conclusion is clear: GCIP is a much more robust method, with a much smaller risk of producing a bad result ("bad" meaning with a M_V index worse that some reasonable threshold).

6.4. Efficiency in the detection and identification of GE

In a final comparison test, the detection of gross errors was treated

Table 7
Risk (%) associated with having M_V larger than 0.01.

Λ_r	3.2		3		2.8		2.6	
No. GE	GCIP	LNRT	GCIP	LNRT	GCIP	LNRT	GCIP	LNRT
0	0.00	0.00	0.01	0.19	0.00	0.25	0.21	0.41
1	1.70	13.40	2.95	16.96	3.57	17.37	3.61	19.83
2	1.72	25.46	3.36	31.17	6.93	33.66	9.51	41.66
3	14.47	36.97	16.25	40.44	17.13	41.30	27.06	47.90
4	19.30	43.88	19.63	45.65	25.67	50.86	33.35	61.17
5	22.16	47.72	26.78	51.04	31.14	51.24	45.44	53.99

Table 8
Risk (%) associated with having M_V larger than 0.02.

Λ_r	3.2		3		2.8		2.6	
No. GE	GCIP	LNRT	GCIP	LNRT	GCIP	LNRT	GCIP	LNRT
0	0	0	0	0	0	0	0	0
1	0.00	5.00	0.80	5.94	0.88	6.95	2.00	7.07
2	1.00	9.14	1.40	11.53	1.80	13.55	6.59	17.39
3	1.80	16.18	1.81	16.30	2.20	18.27	9.94	26.77
4	2.76	21.67	2.90	26.12	4.88	28.08	13.15	37.28
5	3.97	27.35	5.41	32.43	5.58	34.18	16.44	36.89

as a binary classifier, by defining a threshold T for residuals: if $|r_i^w| \leq T$, the corresponding measurement i is accepted; if $|r_i^w| > T$, an alarm is triggered, and the measurement is said to be possibly contaminated by a gross error. For the purpose of this paper, T was set to the value T = 5 (see Section 4.4.). Then, for each scenario, the occurrence of the following events was inspected:

a True Positive (TP): the method labels the measurement as a GE, and it actually is.
b False Positive (FP) or False Alarm: the measurement is labeled as a GE, but in fact it is not.
c True Negative (TN): the method labels the measurement as clean of GE, and it is true.
d False Negative (FN) or Missed Alarm: the measurement is labeled as clean of GE, but in fact it is contaminated.

False Positives and False Negatives do not have symmetrical consequences. A FN means that some GE remained undetected and, therefore, the estimation is (strongly) contaminated and may be considered wrong, for practical purposes. A FP means that some measurement was labeled as possibly containing a GE while it did not – but this usually does not strongly affect the quality of the estimation, because such measurement will have a reduced weight and small contribution to the final result.

The FP and FN rates, denoted FPR and FNR, are defined as:

$$FPR = \frac{FP}{TN + FP} \text{ and } FNR = \frac{FN}{FN + TP} \tag{39}$$

FPR and FNR were computed for all the scenarios in each set of experiment with 0–5 GEs. Figs. 8 and 9 show the confidence intervals at 99% for each set of experiments with different redundancies.

As it can be seen, GCIP performs extremely well, in comparison to LNRT. The confidence intervals for GCIP are much smaller than the ones for LNRT and the FPR and FNR rates are always close to zero for GCIP; this means LNRT is much more prone to generate false alarms or missed alarms for bad data. The small confidence interval and values near zero show the robustness of the new method in detecting and identifying the gross errors.

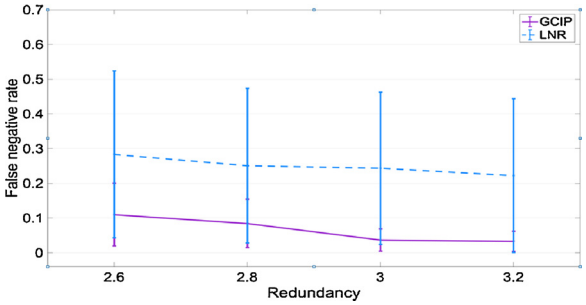

Fig. 8. Confidence interval at 99% of FNR, for 500 cases in each set of experiments with 0–5 GE, in the four Λ_r redundancy-level scenarios.

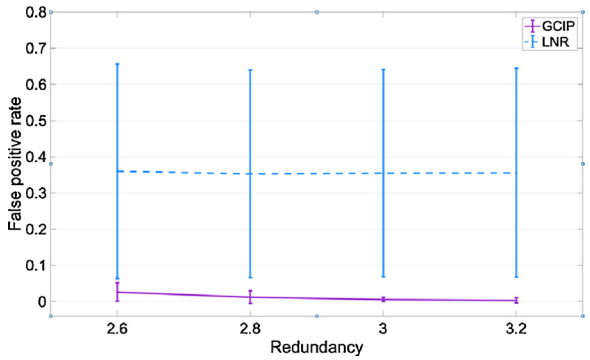

Fig. 9. Confidence interval at 99% of FPR, for 500 cases in each set of experiment with 0–5 GE, in the four Λ_r redundancy-level scenarios.

6.5. GE in critical measurements and leverage points

The Achilles heel of many proposed SE processes is the performance in cases traditionally identified as difficult: errors falling in critical measurements, sets or measurements in leverage points. This has been recently discussed, regarding the vulnerability of the LNRT [55]. Thus, particular care was taken in observing the performance of GCIP in such cases. To achieve that, the set of scenarios was screened and divided in the following sub-sets:

- Group A: cases including neither leverage nor critical measurements or sets contaminated with GE.
- Group B: GEs in measurements at leverage points;
- Group C: GEs in critical measurements or critical sets;
- Group D: GEs in leverage points and in critical measurements or critical sets simultaneously.

To identify leverage points (which depend on the measurement plan defined, measurement weighting and system parameters), the method suggested in Ref. [56] was adopted: an indicator, called Undetectability Index (UI), is used to qualify all the measurements of a metering system from a geometrical point of view. In Ref. [56], the authors showed that the set of measurements with high UI contains the critical measurements and leverage points. In general, the UI values have to be compared with a cutoff value. It was shown in Ref [56]. that critical measurements belonging to the range space of the Jacobian matrix have an infinite UI. However, the suggested algorithm does not identify critical sets. Therefore, in order to map the dataset onto the four above-mentioned groups, the method in Ref. [57]. was adopted in order to identify critical measurement sets. The combination of these techniques allowed us to identify leverage points and their occurrence (or not) associated with critical measurements or critical sets.

6.6. Convergence in the special cases

The GCIP algorithm is very stable and converges in 100% of the scenarios regardless of the incidence of GEs, for all groups A–D (even if some GE is missed), while LNRT is prone to fail in at least 46.3% cases. This rate can reach as low as 13% when there are GEs located in leverage and critical measurements simultaneously. The intervals of confidence at 99% for GCIP and LNRT, for each subset A to D, are depicted in Fig. 10. The estimation accuracy associated with of each method is again evaluated by comparing the average M_V indexes of all the four redundancy-level scenarios for groups A, B, C, and D. The comparisons shown in Table 9 provide evidence of a clear superiority of GCIP over LNRT.

It is important to note that, as the locations of GEs were assigned randomly, the percentage of cases in groups B, C, and D are relatively smaller than group A. For all redundancies and different numbers of

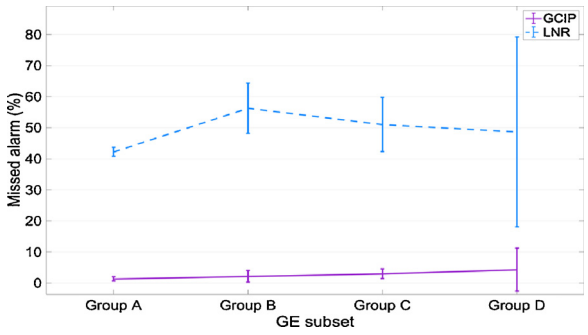

Fig. 10. Confidence interval of percentage of the missed alarms to total existing GEs in all the four redundancy-level scenarios with 1–5 simultaneous GEs categorized by the location of GEs.

Table 9
Average M_V index of all the four redundancy-level scenarios with 1–5 simultaneous GEs, by location of GEs.

Λ_r	Group A		Group B		Group C		Group D	
GE	GCIP	LNR	GCIP	LNR	GCIP	LNR	GCIP	LNR
1	0.006	0.008	0.006	0.023	0.008	0.033	N/A	N/A
2	0.006	0.011	0.006	0.021	0.014	0.034	0.007	0.048
3	0.007	0.013	0.008	0.020	0.018	0.031	0.004	0.040
4	0.009	0.015	0.008	0.028	0.018	0.034	0.006	0.040
5	0.010	0.016	0.010	0.029	0.035	0.039	0.038	0.052

added GE, the percentage of cases in groups A, B, C, and D are respectively 92, 3.8, 4.1, and 0.1. Clearly, the co-incidence of GE in leverage measurements and critical sets simultaneously in a single scenario (group D) is much lower than other experimental groups and it does not happen in all redundancy-level experiments. Those experiments that did not include this condition are pointed as N/A in Table 9 (obviously, when we have only one GE).

6.7. Detection and identification efficiency in special cases

Figs. 10 and 11 represent the 99% confidence intervals for the percentage of the missed alarms and false alarms to total existing GEs, for all the data sub-sets, considering the total in all four redundancy-level scenarios, with 1–5 simultaneous GE.

The plots show the results for each data group separately. As suggested in the figures, one can be 99% certain that the true population mean is contained in the ranges calculated for each category. As it can be seen from the plots, the confidence interval of Group A (which is a large sample) is quite narrow.

This happens because much more certainty (narrower confidence

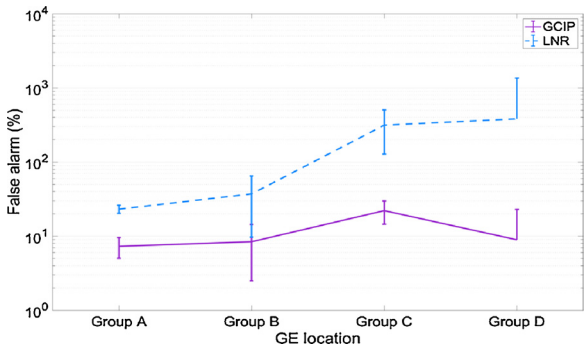

Fig. 11. Confidence interval of percentage of false alarms to total existing GEs in all the four redundancy-level scenarios with 1–5 simultaneous GEs categorized by the location of GEs.

interval) may be estimated with the calculation of the mean of larger samples, which is a non-biased estimator of an assumed true distribution. Constructing confidence intervals can suffer with smaller samples such as groups B, C and D.

A common solution to compute confidence intervals on non-normally distributed and small datasets is by using the Bootstrap method [54]. This method resamples the data with replacement many times and performs a statistic analysis (average values in this work) in each iteration. Due to the central limit theorem, the distribution of means will always approach a normal distribution and that can be used to compute accurate confidence intervals.

Fig. 10 suggests, at 99% confidence, that the missed alarm percentage using GCIP method is less than 2%, 4.1%, 4.6%, and 11.5% for groups A, B, C, and D respectively, noting that group A constitutes 92% of total scenarios. These percentages are in the sense of conditional probability, i.e. they are estimates *given that* we are in a specific special set (a subset of all events).

These values are much higher for LNRT. Furthermore, as the confidence intervals of two methods do not overlap, one can conclude, with 99% confidence, that the true means of these two methods differ significantly. In particular, given the severe impact, in the SE, of having false negatives, the performance of GCIP in Groups B and D, where GE fall on leverage points, is remarkable. The same can be said for Group C, where GE fall on critical measurements or critical sets – something that cannot be dealt with the LNRT model.

Fig. 11 shows another perspective of the statistically significant superiority of the GCIP method, at a 99% confidence level: no more than 31% false alarms in the worst-case scenario — significantly lower than what is obtained with LNRT. Fig. 11 is plotted in logarithmic scale through the y-axis so that the most foreseeable range for the mean of the rate of false alarms for LNRT can be easily seen and compared with the ones of GCIP.

The base value for calculating percentages was the number of actually existing gross errors. As the number of false alarms (or false positives) can be significantly higher, the calculated percentages may well be orders of magnitude higher that 100% - this is evident for the false positives generated by the LNRT process.

7. Conclusions

A single gross error in a measurement set, in power system state estimation, is handled quite well in most cases (but not in critical sets or measurements) by classical approaches, such as the Largest Normalized Residual Test, based on the weighted least squares approach, which is the most used in industrial applications. The same cannot be said for simultaneous multiple errors. Moreover, the application of that and other proposed methods is quite unsuccessful in particular cases, such as of errors occurring in critical sets.

Previous work departing from a least squares approach and adopting Correntropy as the cost function have produced a sound theoretical model explaining where to find the optimal solution of a SE problem with gross errors — however, not how to find it. The problem usually becomes plagued with local optima, and convergence with classical Gauss–Newton iterations is problematic and may fail, for the lack of a consistently good starting point for iterations. Also, attempts using Correntropy and Interior Point method faced problems because of the need to reduce the size of Parzen windows.

This paper presented work demonstrating that an alternative method could be built, based on two evolutions: 1. Adopting Generalized Correntropy as cost function, with a very large shape parameter c, and 2. Adopting an Interior Point solving strategy, moving away from the Gauss–Newton classical approach, keeping a separate representation of objective function and constraints. This combination, with its favorable properties, is the essence and novelty of the GCIP model.

Generalized Correntropy, based on generalized Gaussian kernels, keeps the fundamental Information Theoretical basis and properties of formulations using Correntropy, but allows re-shaping the optimization landscape in a more favorable manner. Furthermore, the Generalized Correntropy definition depends on setting some parameters, but theoretical results indicate an optimal shape parameter to be used in GCIP, so there is no need for parameter tuning.

A toy problem presented in the paper visually illustrates some of the interesting properties of GCIP. However, in order to claim that this is a viable alternative, it was necessary to demonstrate a competitive performance in terms of computing effort and a superior performance in the quality of results. The term of comparison, in this paper, in a carefully designed testbed, was the classical Largest Normalized Residual Test, because this is the technique most used in industrial applications.

One of the remarkable results obtained by GCIP is its ability to correctly solve problems with multiple gross errors falling in critical measurements and sets, something that LNRT and other methods are unable to address. Another remarkable feature is the robustness in face of gross errors falling on leverage points, something that methods such as Least Absolute Error have been shown to be vulnerable to. GCIP processes these features in a natural way and in a competitive computing time. Consequently, in an operational environment, GCIP will be a reliable tool that offers the same quality as LNRT for simple cases and provides quality answers in the more difficult cases. The quantification of this advantage was achieved in two ways: by evaluating the risk, posed by both approaches, of producing an unsatisfactory estimation — and by the comparison in the rate of production of missed alarms (failing to detect a gross error) and false alarms (wrongly indicating an error-free measurement as being contaminated by a gross error). In all cases, the superiority of GCIP is demonstrated by the non-overlapping intervals of confidence estimated for the results obtained.

The results obtained provide experimental validation to special properties of GCIP, which allow the method to deal with cases that were previously considered very difficult or impossible to solve and for which no practical solution existed so far. The derivation of such properties will be approached in a coming publication — but the evidence that results are appearing in cases where previously one would think not possible, should immediately make one realize that a mathematical explanation must be behind them.

Some additional future lines of work and reporting are the following:

(a) dissecting the mathematical justifications for the properties of the GCIP model;
(b) showing the effects of changing shape and size of Parzen windows, confirming theoretical expectations;
(c) showing the robustness of GCIP to distinct noise distributions (actually, in GCIP there are no assumptions on the shape of the distribution of the errors, contrarily to WLS, where a Gaussian distribution is assumed for an optimal estimation);
(d) numerically showing that the method scales up nicely with network size and with the number of gross errors.

As for network size, a general characteristic of electric networks is that the effect of perturbations is mostly local, so SE algorithms will perform statistically better in larger networks, for the same number of randomly assigned gross errors (the probability of simultaneous gross errors falling in interfering measurements becomes smaller). Therefore, tests on a 30-bus system are significant and justified, and nothing out of the ordinary is to be expected in larger networks.

The work reported, therefore, presents experimental evidence supporting the development of a new reliable, robust and competitive tool, combining Generalized Correntropy with an Interior Point solver, for a state estimation process resilient to gross errors, including in leverage and critical measurements, apt to be integrated in control centers and used in an operation environment.

Conflict of interest

None.

Acknowledgments

This work is integrated in the research effort supported by the ERDF – European Regional Development Fund (COMPETE 2020 Programme) and FCT - Fundação para a Ciência e Tecnologia (project POCI-01-0145-FEDER-016731 INFUSE), Portugal. A.S. Costa also recognizes the support of CNPq (projects 312037/2013-9 and 400799/2014-6), Brazil. S. Pesteh recognizes the support of operation NORTE-08-5369-FSE-000043 co-funded by the European Social Fund (FSE) through NORTE 2020 - Programa Operacional Regional do NORTE. J. B. A. London Jr. recognizes the support of FAPESP (projects 2017/22297-6), Brazil.

References

[1] W. Liu, P. Pokharel, J. Principe, Correntropy: properties and applications in non-Gaussian signal processing, IEEE Trans. Signal Process. 55 (November (11)) (2007) 5286–5298, https://doi.org/10.1109/TSP.2007.896065.
[2] V. Miranda, A. Santos, J. Pereira, State estimation based on correntropy: a proof of concept, IEEE Trans. Power Syst. 24 (November (4)) (2009) 1888–1889, https://doi.org/10.1109/TPWRS.2009.2030117.
[3] M.H. Wright, The interior-point revolution in optimization: history, recent developments, and lasting consequences, Bull. Am. Math. Soc. 42 (September (1)) (2004) 39–56, https://doi.org/10.1090/S0273-0979-04-01040-7.
[4] D.C. Montgomery, G.C. Runger, Applied Statistics and Probability for Engineers, John Wiley & Sons, Inc., 2003.
[5] H.M. Merrill, F.C. Schweppe, Bad data suppression in power system static state estimation, IEEE Trans. Power Appar. Syst. PAS-90 (6) (1971) 2718–2725.
[6] O. Alsac, N. Vempati, B. Stott, A. Monticelli, Generalized state estimation, IEEE Trans. Power Syst. 13 (August (3)) (1998) 1069–1075, https://doi.org/10.1109/59.709101.
[7] F.C. Schweppe, D.B. Rom, Power system static-state estimation, part II: approximate model, IEEE Trans. Power Appar. Syst. PAS-89 (January (1)) (1970) 125–130, https://doi.org/10.1109/TPAS.1970.292679.
[8] W. Zheng, W. Wu, A. Gomez-Exposito, B. Zhang, Y. Guo, Distributed robust bilinear state estimation for power systems with nonlinear measurements, IEEE Trans. Power Syst. 32 (1) (2017) 499–509, https://doi.org/10.1109/TPWRS.2016.2555793.
[9] Y. Lin, A. Abur, Robust state estimation against measurement and network parameter errors, IEEE Trans. Power Syst. 33 (5) (2018) 4751–4759, https://doi.org/10.1109/TPWRS.2018.2794331.
[10] P.J. Huber, Robust statistics, Wiley Series in Probability and Statistics, John Wiley & Sons, Inc., New York, USA, 1981 February.
[11] A.A. da Silva, V. Quintana, G.K.H. Pang, A pattern analysis approach for topology determination, bad data correction and missing measurement estimation in power systems, Proceedings of the Twenty-Second Annual North American Power Symposium (1990) 363–372, https://doi.org/10.1109/NAPS.1990.151390.
[12] V. Miranda, J. Krstulovic, H. Keko, C. Moreira, J. Pereira, Reconstructing missing data in state estimation with autoencoders, IEEE Trans. Power Syst. 27 (May (2)) (2012) 604–611, https://doi.org/10.1109/TPWRS.2011.2174810.
[13] J. Krstulovic, V. Miranda, A. Simões Costa, J. Pereira, Towards an auto-associative topology state estimator, IEEE Trans. Power Syst. 28 (August (3)) (2013) 3311–3318, https://doi.org/10.1109/TPWRS.2012.2236656.
[14] F. Zhuang, R. Balasubramanian, Bad data suppression in power system state estimation with a variable quadratic-constant criterion, IEEE Trans. Power Appar. Syst. PAS-104 (July (4)) (1985) 857–863, https://doi.org/10.1109/TPAS.1985.319086.
[15] A. Simões Costa, J.G. Rolim, P.W. Aitchinson, Iterative bad-data suppression applied to state estimators based on the augmented matrix method, Electr. Power Syst. Res. 20 (March (3)) (1991) 205–213, https://doi.org/10.1016/0378-7796(91)90065-U.
[16] J. Krstulovic, Information Theoretic State Estimation in Power Systems. PhD Dissertation, April Faculty of Engineering of the University of Porto, 2014, http://hdl.handle.net/10216/72756.
[17] W. Wu, Y. Guo, B. Zhang, A. Bose, S. Hongbin, Robust state estimation method based on maximum exponential square, IET Gener. Transm. Distrib. 5 (November (11)) (2011) 1165–1172, https://doi.org/10.1049/iet-gtd.2011.0100.
[18] Y. Chen, F. Liu, G. He, S. Mei, Y. Fu, Maximum exponential absolute value approach for robust state estimation, IEEE International Conference on in Power System Technology (POWERCON) (2012), https://doi.org/10.1109/PowerCon.2012.6401317 November.
[19] Y. Chen, F. Liu, S. Mei, J. Ma, Toward adaptive robust state estimation based on MCC by using the generalized Gaussian density as kernel functions, Electr. Power Energy Syst. 71 (October) (2015) 297–304, https://doi.org/10.1016/j.ijepes.2015.03.011.
[20] P.J. Huber, Robust estimation of a location parameter, Ann. Math. Stat. 35 (1) (1964) 73–101.
[21] E. Handschin, et al., Bad data analysis for power system state estimation, IEEE Trans. Power Appar. Syst. 94 (2) (1975) 329–337, https://doi.org/10.1109/T-PAS.1975.31858.
[22] C. Yu, W. Yao, Robust linear regression: a review and comparison, Commun. Stat. Simul. Comput. 46 (8) (2017), https://doi.org/10.1080/03610918.2016.1202271.
[23] B. Mendes, D.E. Tyler, Constrained M-estimation for regression, in: H. Rieder (Ed.), Robust Statistics, Data Analysis, and Computer Intensive Methods, Lecture Notes in Statistics, vol. 109, Springer, New York, 1996.
[24] M. Irving, R. Owen, M. Sterling, Power-system state estimation using linear programming, Proc. Inst. Electr. Eng. 125 (September (9)) (1978), https://doi.org/10.1049/piee.1978.0206.
[25] A. Majumdar, B.C. Pal, Bad data detection in the context of leverage point attacks in modern power networks, IEEE Trans. Smart Grid 9 (3) (2018) 2042–2054, https://doi.org/10.1109/TSG.2016.2605923.
[26] Y. Chen, F. Liu, S. Mei, J. Ma, A robust WLAV state estimation using optimal transformations, IEEE Trans. Power Syst. 30 (4) (2015) 2190–2191, https://doi.org/10.1109/TPWRS.2014.2358452.
[27] L. Mili, V. Phaniraj, P.J. Rousseeuw, Least median of squares estimation in power systems, IEEE Trans. Power Syst. 6 (May (2)) (1991) 511–523, https://doi.org/10.1109/59.76693.
[28] L. Mili, et al., Robust state estimation based on projection statistics [of power systems], IEEE Trans. Power Syst. 11 (May (2)) (1996) 1118–1127, https://doi.org/10.1109/59.496203.
[29] S. Gastoni, G. Granelli, M. Montagna, Robust state-estimation procedure based on the maximum agreement between measurements, IEEE Trans. Power Syst. 19 (November (4)) (2004) 2038–2043, https://doi.org/10.1109/TPWRS.2004.836190.
[30] M. Irving, Robust state estimation using mixed integer programming, IEEE Trans. Power Syst. 23 (August (3)) (2008) 1519–1520, https://doi.org/10.1109/TPWRS.2008.926721.
[31] J. Zhao, L. Mili, A framework for robust hybrid state estimation with unknown measurement noise statistics, IEEE Trans. Ind. Inform. 14 (May (5)) (2018) 1866–1875.
[32] W.S. Krasker, R.E. Welsch, Efficient bounded-influence regression estimation, J. Am. Stat. Assoc. 71 (379) (1982), https://doi.org/10.2307/2287719.
[33] J. Krstulović, V. Miranda, Denoising auto-associative measurement screening and repairing, ISAP 2015 - 18th International Conference on Intelligent System Application to Power Systems (2015), https://doi.org/10.1109/ISAP.2015.7325548.
[34] A. Renyi, Some fundamental questions of information theory, Selected Papers of Alfred Renyi vol. 2, Akademia Kiado, Budapest, 1976, pp. 526–552.
[35] C.E. Shannon, A mathematical theory of communication, Bell Syst. Tech. J. 27 (3) (1948) 379–423.
[36] E. Parzen, On the estimation of a probability density function and mode, Ann. Math. Stat. 33 (September (3)) (1962) 1065–1076.
[37] B. Chen, L. Xing, H. Zhao, N. Zheng, J.C. Principe, Generalized correntropy for robust adaptive filtering, IEEE Trans. Signal Process. 64 (13) (2016) 3376–3387, https://doi.org/10.1109/TSP.2016.2539127.
[38] K. Ponnambalam, V. Quintana, A. Vannelli, A fast algorithm for power system optimization problems using an interior point method, IEEE Power Industry Computer Application Conference (1991), https://doi.org/10.1109/PICA.1991.160607.
[39] J.A. Momoh, S.X. Guo, E.C. Ogbuobiri, R. Adapa, The quadratic interior point method solving power system optimization problems, IEEE Trans. Power Syst. 9 (3) (1994) 1327–1336, https://doi.org/10.1109/59.336133.
[40] H. Wei, H. Sasaki, T. Nagata, An application of interior point quadratic programming algorithm to power system optimization problems, IEEE Trans. Power Energy 116 (2) (1996) 174–180, https://doi.org/10.1109/59.486104.
[41] K.A. Clements, P.W. Davis, K.D. Frey, An interior point algorithm for weighted least absolute value power, system state estimation, IEEE Winter Power Meeting (1991) P. 91-WM 235-2 PWRS.
[42] K.A. Clements, P.W. Davis, K.D. Frey, Treatment of inequality constraints in power system state estimation, IEEE Trans. Power Syst. 10 (2) (1995) 567–574, https://doi.org/10.1109/59.387890.
[43] H. Singh, F.L. Alvarado, Weighted least absolute value state estimation using interior point methods, IEEE Trans. Power Syst. 9 (3) (1994) 1478–1484, https://doi.org/10.1109/59.336114.
[44] H. Wei, et al., An interior point method for power system weighted nonlinear L_1 norm static state estimation, IEEE Trans. Power Syst. 13 (2) (1998) 617–623, https://doi.org/10.1109/59.667390.
[45] A.S. El-Bakry, R.A. Tapia, T. Tsuchiya, Y. Zhang, On the formulation and theory of the Newton interior-point method for nonlinear programming, J. Optim. Theory Appl. 89 (3) (1996) 507–541, https://doi.org/10.1007/BF02275347.
[46] S. Vaughan, Scientific Inference: Learning From Data, 1st ed., Cambridge University Press, Cambridge, UK, 2013.
[47] S.J. Wright, Primal-Dual Interior-Point Methods, SIAM, 1997, https://doi.org/10.1137/1.9781611971453.ch1.
[48] A.M.K. Celik, A. Abur, A robust wlav state estimator using transformations, IEEE Trans. Power Syst. 7 (1) (1992) 106–113, https://doi.org/10.1109/59.141693.
[49] S. Pesteh, H. Moayyed, V. Miranda, GCIP: a robust state estimation method for dad data in leverage and critical measurements, INESC TEC Internal Report 12/03/2019, Project INFUSE - PTDC/EEI-EEL/5949/2014, Mar 2019. Submitted to EPSR As “Favorable Properties of Interior Point Method and Generalized Correntropy in Power System State Estimation”, (2019).
[50] R.D. Zimmerman, C.E. Murillo-Sánchez, R.J. Thomas, MATPOWER: steady-state operations, planning and analysis tools for power systems research and education, IEEE Trans. Power Syst. 26 (February (1)) (2011) 12–19, https://doi.org/10.1109/

TPWRS.2010.2051168.
[51] KEMA, Metrics for determining the impact of phasor measurements on power system state estimation, DRAFT, Eastern Interconnection Phasor Project, (2006) January.
[52] A.J. Monticelli, Electric power system state estimation, Proc. IEEE 88 (2) (2000) 262–282, https://doi.org/10.1109/5.824004.
[53] T. Hastie, R. Tibshirani, J. Friedman, Kernel smoothing methods, The Elements of Statistical Learning. Springer Series in Statistics, Springer, New York, NY, 2009.
[54] N.I. Fishera, P. Hall, Bootstrap algorithms for small samples, J. Stat. Plan. Inference 27 (2) (1991) 157–169, https://doi.org/10.1016/0378-3758(91)90013-5.
[55] J. Zhao, L. Mili, Vulnerability of the largest normalized residual statistical test to leverage points, IEEE Trans. Power Syst. 33 (July (4)) (2018), https://doi.org/10.1109/TPWRS.2018.2831453.
[56] R.A.S. Benedito, L.F.C. Alberto, N.G. Bretas, J.B.A. London Jr, Power system state estimation: undetectable bad data, Int. Trans. Electr. Energy Syst. 24 (2014) 91–107, https://doi.org/10.1002/etep.1744.
[57] A.S. Costa, T.S. Piazza, Qualitative methods to solve qualitative problems in power system state estimation, IEEE Trans. Power Syst. 5 (3) (1990), https://doi.org/10.1109/59.65924.