M-Arctan estimator based on the trust-region method

Yacine Hassaïne
French transmission system operator (RTE)
Yacine.Hassaine@rte-france.com

Benoît Delourme
RTE
Benoit.Delourme@rte-france.com

Pierre Hausheer
RTE

Patrick Panciatici
RTE
Patrick.panciatici@rte-france.com

Eric Walter
L2S (CNRS-SUPELEC-U.Paris Sud)
Walter@lss.supelec.fr

Abstract - In this paper a new approach is proposed to increase the robustness of the classical L_2-norm state estimation, to achieve this task a new formulation of the Levemberg Marquardt algorithm based on the trust-region method is applied to a new M-estimator, denoted M-Arctan. Results obtained on IEEE network up to 300 buses are presented

Keywords - State estimation, robustness, trust region method, M-estimator.

1 Introduction

Electrical power problems provides challenging applications to mathematics in general, and state estimation in particular. In the early seventies, Schweppe [14] introduced *static state estimation* in the context of power network analysis. L_2-norm estimation combined to the Gauss-Newton algorithm was the approach mainly used. This contribution is also related to *static state estimation*. The state vector is thus assumed invariant ([12], [11] and [6]) ; several cost functions were studied : L. Mili *et al.* [10] used Huber's M-estimator, R. Baldick *et al.* [1] used a non-quadratic cost function closely related to Huber's M-estimator and M.K Çelik *et al.* [3] proposed a weighted L_1 norm estimator using constrained optimization. In the most general case, the state vector may consist of a mixture of real and Boolean variables [7]. Here, the state vector is supposed to be real valued.

In this paper we propose a new algorithm to increase the robustness of the classical L_2 state estimator based on the Gauss-Newton algorithm. To achieve this task, two types of robustness are defined. (i) *Statistical robustness*, aims at decreasing the influence of outliers on the estimation quality [8]. (ii) *Numerical robustness*, aims at guaranteeing convergence of the algorithm in ill conditioned environment or in the presence of outliers. The outliers may correspond to bad measurements, topology errors or erroneous characteristics of branches. The non linear model is given by

$$\mathbf{z} \quad = \quad \mathbf{h}(\mathbf{x}) + \boldsymbol{\epsilon}, \tag{1}$$

where $\mathbf{z}$ is the m-dimensional measurement vector, $\mathbf{x}$ is the n-dimensional state vector, $\mathbf{h}$ is a nonlinear vector function and $\boldsymbol{\epsilon}$ is the m-dimensional vector of measurement errors. The L_2 estimator is known to be very sensitive to outliers in contrast with the M-estimators and the Gauss-Newton algorithm applied to L_2 estimation reaches its limits when the optimum is ill conditioned or singular [13]. In this paper an examples illustrating the problem is given in a context of *maximum power transfer capacity*. To increase statistical robustness, a new strictly convex M-estimator, denoted *M-Arctan*, is introduced. To improve numerical robustness, a trust-region algorithm is used [5] with a new approach to calculating the Lagrange multiplier μ. The algorithm obtained is applied to the M-Arctan estimator to obtain an estimation that is both statistically and numerically robust.

The paper is organized as follow. Section 2 introduces the problem of ill conditioning and details the special case of maximum power transfer capacity. Section 3 propose the new M-estimator denoted M-Arctan. Section 4 details the Levemberg Marquardt algorithm based on the trust-region method. Finally the approach proposed is tested in Section 5.

2 Problems of ill conditioned optimum

Several problems may cause the divergence of state estimation based on a Newtonian algorithm. In this section, one of them namely the problem of *maximum power transfer capacity* is studied. Consider a network with two busses connected by a branch with a known admittance y. Assume that a power flow measurement and a phase reference on one of the two busses are available. Assume further that voltage magnitude is fixed in the two busses. The objective of this example is to estimate the unknown phase θ. Under these conditions, one can write

$$T = v^2 y \sin(\theta) + \epsilon, \tag{2}$$

where ϵ is a white noise, T is the active measurement, v is the know voltage magnitude and θ is the unknown phase. To estimate θ, we minimize the quadratic cost function

$$J(\theta) = (T - v^2 y \sin(\theta))^2, \tag{3}$$

The minimizer $\widehat{\theta}$ of (3) satisfies the stationarity condition

$$\frac{dJ}{d\theta} = 0, \tag{4}$$

so

$$-v^2 y \cos(\widehat{\theta})(T - v^2 y \sin(\widehat{\theta})) = 0, \tag{5}$$

which implies either

$$\cos\widehat{\theta} = 0, \tag{6}$$

or

$$\sin\widehat{\theta} = \frac{T}{v^2 y}. \tag{7}$$

Usually, the iterative optimization procedure is started at $\widehat{\theta}_0 = 0$, and the Gauss-Newton algorithm ensures convergence, toward the second solution given by (7). However, if

$$T \quad > \quad v^2 y, \tag{8}$$

this solution is unfeasible. The algorithm can then converge only towards the first solution, i.e. $\cos\widehat{\theta} = 0$. However since

$$\frac{dT}{d\theta} = -v^2 y \cos\theta, \tag{9}$$

it becomes impossible to compute the inverse of the approximate Hessian of the Gauss-Newton algorithm. This phenomenon appears whenever a critical measurement of power flow turn out to be higher than the maximum power transfer capacity, equal to $v^2 y$ MW in the present example.

We can easily generalize to a matrix formulation. Consider the general model given by (1). If the active power flow measurement z_i is greater than the maximum power transfer capacity, and if there is a solution, then it must satisfy

$$h_i(\widehat{\mathbf{x}}) \quad = \quad \max_{\mathbf{x}}(h_i(\mathbf{x})), \tag{10}$$

so

$$\frac{dh_i}{d\mathbf{x}}|_{\mathbf{x}=\widehat{\mathbf{x}}} = \mathbf{0}. \tag{11}$$

Thus the row of the jacobian matrix corresponding to the critical measurement exceeding maximum power transfer capacity is zero. This implies that the matrix $\mathbf{H}$ is rank deficient. Under this condition, the Gauss-Newton algorithm cannot converge.

Two factors can create a problem of maximum power transfer capacity: (i) a low local redundancy level which allows the appearance of critical measurements, (ii) errors in the data.

3 M-Arctan estimator

The basic idea of M-estimators [8] is to consider a large class of functions in defining the cost function. This class of functions has the following expression

$$J(\mathbf{x}) = \sum_{i=1}^{m} \rho(r_{\mathrm{w}i}(\mathbf{x})), \tag{12}$$

where ρ is symmetric with a single minimum at zero. For example, Huber's and Tuckey's M-estimators [8] are statisticaly much more robust than the classical L_2-norm estimator. However these two least M-estimators are not strictly convex,i.e. eigenvalues of the Hessian matrix may be zero. Under this condition the classical Gauss-Newton algorithm is not applicable. L. Mili et al. [10] applied the *re-weighted least squares algorithm* to Huber's M-estimator. J. B. Birch [2] compared these two methods (in the location model) in terms of their rates of convergence to the appropriate solution. He showed that the re-weighted least squares algorithm has a *linear* rate of convergence. As known, the Gauss-Newton algorithm has a *quadratic* rate of convergence [4], thus it converges much more quickly.

In this paper, we introduce a new type of M-estimator which called M-Arctan. Its cost function is given by

$$J(\mathbf{x}) = \sum_{i=1}^{m} \rho(\mathbf{z} - \mathbf{h}(\mathbf{x})), \tag{13}$$

with

$$\rho(r) = \frac{2\lambda}{\pi} r \arctan(\frac{\pi r}{2\lambda}) - \frac{2\lambda^2}{\pi^2}\ln(1 + \frac{(\pi r)^2}{4\lambda^2}), \tag{14}$$

the first derivative is

$$\rho(r_{\mathrm{w}})' = \frac{2\lambda}{\pi}\arctan(\frac{\pi r_{\mathrm{w}}}{2\lambda}), \tag{15}$$

and the second derivative is

$$\rho(r_{\mathrm{w}})'' = \frac{4\lambda^2}{(\pi r_{\mathrm{w}})^2 + 4\lambda^2}. \tag{16}$$

In (14), $\lambda \in \mathbb{R}_+^*$. Note that $\rho(.)$ is strictly convex as $\frac{d^2\rho(r)}{dr^2} > 0$, which always allows the application of the Gauss-Newton algorithm, contrary to Huber's and Tuckey's M-estimators.

4 Trust-region method applied to M-Arctan estimation

The quadratic approximation of the cost function used by the Newtonian approaches is only valid in the neighborhood of the current point. However, the inversion of the Hessian or of its approximation, in particular when the Hessian is close to singular, may lead to a result far from the neighborhood of the current point, as in the example of maximum-power transfer capacity. Under these conditions the trust-region method becomes relevant. The *Levemberg and Marquardt algorithm with the trust region method* [5] is then used to improve the convergence of the Gauss-Newton algorithm. The resulting algorithm is based on: (i) the use of a quadratic approximation in the neighborhood of the current point, which guarantees the validity of the approximation, (ii) a Levemberg-Marquardt regularization of the Hessian matrix.

At each iteration k, the following optimization problem is solved

$$\left\{ \begin{array}{l} \min J_{\text{app}}(\mathbf{s}_k) \\ \|\mathbf{s}_k\| \leq \delta_k, \end{array} \right. \tag{17}$$

with

$$J_{\text{app}}(\mathbf{s}_k) = J(\mathbf{x}_k) + \nabla J(\mathbf{x}_k)^{\mathrm{T}}\mathbf{s}_k + \frac{1}{2}\mathbf{s}_k^{\mathrm{T}}\nabla^2 J(\mathbf{x}_k)\mathbf{s}_k, \tag{18}$$

where $\nabla J(\mathbf{x}_k)$ is the gradient vector of the cost function, $\nabla^2 J(\mathbf{x}_k)$ is the Hessian matrix (or its approximation), $\mathbf{s} = \widehat{\mathbf{x}}_{k+1} - \widehat{\mathbf{x}}_k$ and δ_k is the radius of the trust region.

The goal of this approach is to guarantee convergence to a stationary point and a local minimizer of the cost function. This method was applied, by K. Clements *et al.* [16], to L_2-norm state estimation. In this paper, we propose a new approach to compute the Lagrange multiplier and apply the resulting method to the M-Arctan estimation, Thus improving the numerical and statistical robustness.

4.1 Definition of trust region

The Karush-Kuhn-Tucker (KKT) conditions [4] to (17) imply

$$\begin{array}{rcl} \nabla J_{\text{app}}(\mathbf{s}) + \mu \mathbf{I}\mathbf{s} & = & 0, \\ \mu\left(\|\mathbf{s}\|^2 - \delta^2\right) & = & 0, \end{array} \tag{19}$$

where $\mu \in \mathbb{R}^+$ is the Lagrange multiplier. The second condition of (19), known as *complementarity constraint*, means that :

- if the constraint is not active in $\mathbf{s}$, i.e. $\|\mathbf{s}\| \leq \delta$, then $\mu = 0$.
- if $\mu > 0$, then the constraint is active in this point and $\|\mathbf{s}\| = \delta$.

Thus two cases have to be considered. If $\mu > 0$, then

$$\begin{array}{rcl} (\nabla^2 J + \mu \mathbf{I})\, \mathbf{s}(\mu) & = & -\nabla J, \\ \|\mathbf{s}(\mu)\| & = & \delta. \end{array} \tag{20}$$

Using Gauss-Newton approximation, we get

$$\begin{array}{rcl} (\mathbf{H}^{\mathrm{T}}\mathbf{Q}\mathbf{H} + \mu \mathbf{I})\, \mathbf{s}(\mu) & = & \mathbf{H}^{\mathrm{T}}\mathbf{Q}\mathbf{r}, \\ \|\mathbf{s}(\mu)\| & = & \delta. \end{array} \tag{21}$$

If $\mu = 0$ then (20) becomes

$$\nabla^2 J\, \mathbf{s}(\mu) \quad = \quad -\nabla J, \tag{22}$$

which corresponds to the classical Newton algorithm. The Gauss-Newton algorithm can also be applied

$$(\mathbf{H}^{\mathrm{T}}\mathbf{Q}\mathbf{H})\, \mathbf{s}(\mu) \quad = \quad \mathbf{H}^{\mathrm{T}}\mathbf{Q}\mathbf{r}. \tag{23}$$

It remains to be seen how $\mathbf{s}$ and μ are updated. The following sections are devoted to these problems.

4.2 Radius of trust region

The idea is to quantify the validity of the quadratic approximation (see [15]) using the *actual reduction* denoted ared and the *predicted reduction* denoted pred

$$\text{ared} \quad = \quad J(\mathbf{x}_k) - J(\mathbf{x}_{k+1}) \tag{24}$$

$$\text{pred} \quad = \quad J(\mathbf{x}_k) - J_{\text{app}}(\mathbf{x}_{k+1}). \tag{25}$$

The radius of the trust region is updated via two iterations:

1. The first iteration is at the current point. A tuning parameter of the algorithm is $t \in]0, 1]$:
   - If ared $< t$.pred
     the cost function approximation is considered as invalid thus the trust-region radius is too large. It is then necessary to decrease δ, by dividing it by two for example, and to solve again the problem (17) with this new δ.
   - If ared $> t$.pred
     In this case the effective minimization of the cost function is sufficient, thus the value of δ is accepted.
2. The second iteration corresponds to the main iteration on the state vector $\mathbf{x}$. Consider two more tuning parameters u and v such that $0 < t < u < v < 1$. δ was accepted in the precedent iteration so

   $$\text{ared} > t.\text{pred}.$$

   - If ared $< u$.pred
     the minimization is considered effective but the cost function approximation is not good enough because the trust region radius is too large. δ should then be decreased, by dividing it by two for example.
   - If ared $> v$.pred
     the cost function approximation is considered so good the radius of the trust region δ is increased for the following iteration.

4.3 Updating of the Lagrange multiplier

In this section we propose a new approach to updating the multiplier μ, which is also regarded as a regulation parameter ; it is chosen such that $\mathbf{s}(\mu)$ is solution of (20), so

$$\begin{array}{rcl} \|\ \mathbf{s}(\mu)\ \|^2 & = & \|(\nabla^2 J + \mu \mathbf{I})^{-1} \nabla J\|^2, \\ & = & (\nabla J)^T ((\nabla^2 J + \mu \mathbf{I})^{\mathrm{T}} (\nabla^2 J + \mu \mathbf{I}))^{-1} \nabla J, \end{array} \tag{26}$$

where $\nabla^2 J$ is a real symmetric matrix, thus diagonalizable on $\mathbb{R}^n$:

$$\nabla^2 J \quad = \quad \mathbf{\Omega}^{\mathrm{T}} \mathbf{D} \mathbf{\Omega}, \tag{27}$$

and $\mathbf{\Omega}$ is orthogonal, i.e $\mathbf{\Omega}^t \mathbf{\Omega} = \mathbf{I}$, and $\mathbf{D} = \mathrm{diag}(\lambda_1, \ldots, \lambda_n)$, with $\lambda_{i \in \{1,\ldots,n\}}$ the eigenvalues of $\nabla^2 J$.
Then we can write

$$\begin{array}{rcl} \|\mathbf{s}(\mu)\|^2 & = & (\nabla J)^{\mathrm{T}} [(\mathbf{\Omega}^{\mathrm{T}} \mathbf{D} \mathbf{\Omega} + \mu \mathbf{I})^{\mathrm{T}} \\ & & (\mathbf{\Omega}^{\mathrm{T}} \mathbf{D} \mathbf{\Omega} + \mu \mathbf{I})]^{-1} \nabla J, \\ & = & (\mathbf{\Omega} \nabla J)^{\mathrm{T}} [(\mathbf{D} + \mu \mathbf{I})^2]^{-1} \mathbf{\Omega} \nabla J, \end{array} \tag{28}$$

or

$$\|\mathbf{s}(\mu)\|^2 = \sum_{i=1}^{n} \frac{\gamma_i^2}{(\lambda_i + \mu)^2} \ \ (\text{with}\, \gamma_i = (\mathbf{\Omega} \nabla J)_i). \tag{29}$$

It is very difficult to solve the second equation in (20). Actually it is not necessary to impose $\| \mathbf{s} \| = \delta$. Following the recommendations of J.E. Dennis [5], one may instead test whether

$$\frac{3}{4}\delta \leq \| \mathbf{s}(\mu) \| \leq \frac{3}{2}\delta. \tag{30}$$

Define a function $\varphi(\mu)$ from $\mathbb{R}$ to $\mathbb{R}$ such that

$$\varphi(\mu) = \| \mathbf{s}(\mu) \| - \delta. \tag{31}$$

Taking into account the structure of (29), J.E. Denis proposes in [5] to approximate φ in the neighborhood of the current point μ_k by a rational function

$$q_k(\mu) = \frac{\alpha_k}{\beta_k + \mu} - \delta. \tag{32}$$

where $(\alpha_k, \beta_k) \in \mathbb{R}^2$. When the problem is ill conditioned, the eigenvalues of the Hessian matrix are very different. As a consequence, the approximation (32) is less accurate and this may lead to a negative value to μ. The proposed approach guarantees the positivity of this coefficient. The basic idea of this approach is to contract the initial interval I_0 to which belongs μ until condition (30) is satisfied. As suggested in appendix A, the initial interval is taken as

$$I_0 \quad = \quad [0, \tfrac{4\|\nabla J\|}{3\delta}]. \tag{33}$$

Take $I = [\underline{\mu}, \overline{\mu}]$, with $\underline{\mu}$ and $\overline{\mu}$ chosen such that:

$$\begin{cases} \|\mathbf{s}(\mu)\| < \frac{3}{4}\delta & \text{si} \quad \mu > \overline{\mu}, \\ \|\mathbf{s}(\mu)\| > \frac{3}{2}\delta & \text{si} \quad \mu < \underline{\mu}. \end{cases} \tag{34}$$

At each iteration k :

$$\begin{cases} q_k(\underline{\mu}_k) & = & \varphi(\underline{\mu}_k), \\ q_k(\overline{\mu}_k) & = & \varphi(\overline{\mu}_k). \end{cases} \tag{35}$$

Define

$$\begin{cases} \Delta\varphi & = & \varphi(\overline{\mu}_k) - \varphi(\underline{\mu}_k), \\ \Delta\mu_k & = & \overline{\mu}_k - \underline{\mu}_k. \end{cases} \tag{36}$$

The conditions (35) are used to compute α_k and β_k, so

$$\begin{cases} \varphi(\underline{\mu}_k) & = & \frac{\alpha_k}{\beta_k + \underline{\mu}_k} - \delta, \\ \varphi(\overline{\mu}_k) & = & \frac{\alpha_k}{\beta_k + \overline{\mu}_k} - \delta, \end{cases} \tag{37}$$

where

$$\begin{cases} \beta_k = \frac{\alpha_k}{\varphi(\underline{\mu}_k) + \delta} - \underline{\mu}_k = \frac{\alpha_k}{\varphi(\overline{\mu}_k) + \delta} - \overline{\mu}_k, \\ \alpha_k = \frac{(\varphi(\underline{\mu}_k) + \delta)(\varphi(\overline{\mu}_k) + \delta)}{\Delta\varphi / \Delta\mu_k}. \end{cases} \tag{38}$$

The parameter μ_k is updated so that $q_k(\mu_{k+1}) = 0$. The final algorithm is thus

$$\begin{array}{rcl} \mu_{k+1} & = & \underline{\mu}_k - \frac{\|\mathbf{s}(\overline{\mu}_k)\|}{\delta} \left(\frac{\varphi(\underline{\mu}_k)}{\Delta\varphi / \Delta\mu_k} \right), \\ & \text{or} & \\ & = & \overline{\mu}_k - \frac{\|\mathbf{s}(\underline{\mu}_k)\|}{\delta} \left(\frac{\varphi(\overline{\mu}_k)}{\Delta\varphi / \Delta\mu_k} \right). \end{array} \tag{39}$$

Equation (29) shows that the function $\mu \mapsto \|\mathbf{s}(\mu)\|$ is strictly decreasing. Therefore

- If $\|\mathbf{s}(\mu_{k+1})\| > \frac{3}{2}\delta$, then μ should be increased. The new search interval for μ is then

$$I_{k+1} = I_k \cap]\mu_{k+1}, \frac{4\|\nabla J\|}{3\delta}].$$

- If $\|\mathbf{s}(\mu_{k+1})\| < \frac{3}{4}\delta$ then μ should be decreased, the new search interval for μ is then

$$I_{k+1} = I_k \cap [0, \mu_{k+1}[.$$

Note that the algorithm (39) guarantees that

$$\mu_{k+1} \in [\underline{\mu}_k, \overline{\mu}_k]. \tag{40}$$

The proof is in Appendix B.

4.4 Structure of the algorithm

This algorithm is as follows

choose $\epsilon > 0$, υ, ν and t,
initialize δ,
if $\nabla J(\mathbf{s}) > \epsilon$

compute $\nabla J(\mathbf{s})$, and $\nabla^2 J(\mathbf{s})$
initialize ared $= 0$ and pred $= 1$,
while (ared $<$ pred),
if $\|\nabla^2 J^{-1} \nabla J\|_2 > \delta$

compute μ and $\mathbf{s}(\mu)$,
else if
$\mu = 0$,
compute $\mathbf{s}(0)$,
end
compute ared and pred,
if ared $<$ t pred, then $\delta = \frac{\delta}{2}$,
end
if ared $> \upsilon$ pred, then $\delta = 2\delta$,
if ared $< \nu$ pred, then $\delta = \frac{\delta}{2}$,

end

Remark 1 *Normalization of* $\mathbf{s}(\mu)$

The elements of $\mathbf{s}(\mu)$ correspond to variation of the voltage phases and magnitudes in all the electrical busses of the network. Since the ranges and units of these quantities are different, we recommend a normalization for example, by dividing $\delta\mathbf{v}$ by its nominal value v_n and $\delta\boldsymbol{\theta}$ by 2π. This considerably improves the speed of the algorithm given in Section 4.4.

Remark 2 *Computation of* λ

The parameter λ of the M-estimator has still to be chosen. Several strategies are possible. First, λ may be taken as a fixed value suggested by past experiments [10]. The approach proposed in [9] may also be used. In this approach λ is taken equal to $\nu\widehat{\sigma}$, with $1 < \nu < 1.8$ and $\widehat{\sigma}$ a robust estimator (with respect to outliers) of the standard deviation, given by : $\widehat{\sigma} = \frac{\mathrm{med}_i(|r_{\mathrm{w}i} - \mathrm{med}_i r_{\mathrm{w}i}|)}{0.7}$. To choose λ, we may assume that there are no more than a given percentage of outliers, say 5% and look for the value of λ that is such that a percentage $\mathcal{P}_\lambda$ of the residuals are treated as regular (Here $\mathcal{P}_\lambda = 95\%$).

5 Tests

The approach proposed here has been implemented and tested on the IEEE 14, 57, 118 and 300 bus networks. Results on 14 and 300 bus networks in the presence of maximum power transfer capacity problems will be reported here. The variables ϵ, ν, υ and t are fixed as follows: $\epsilon = 10^{-3}$, $\nu = 0.9$, $\upsilon = 0.5$ and $t = 0.1$. Other values may of course be considered as well. Note that in all of the following tests the Gauss-Newton algorithm diverges. Noise-free simulation data were corrupted by noise according to the equation

$$z_{\mathrm{a}}(i) = z_{\mathrm{r}}(i)\,(1 + N(i)), \tag{41}$$

where $z_{\mathrm{a}}(i)$ is the ith component of the actual vector of measurements, $z_{\mathrm{r}}(i)$ is the corresponding noise-free datum and $N(i)$ is a realization of Gaussian white noise with zero mean and standard deviation $\sigma = 4\%$. For L_2 and M-Arctan estimators the Hessian matrix is computed according to the Gauss-Newton approximation,i.e.

$$\nabla^2 J = \mathbf{H}^{\mathrm{T}}\mathbf{R}^{-1}\mathbf{H}.$$

Finally, the algorithm stops when

$$\|\frac{dJ}{d\mathbf{x}}\|_2 \leq 10^{-4}.$$

5.1 Tests on a IEEE-14 bus network

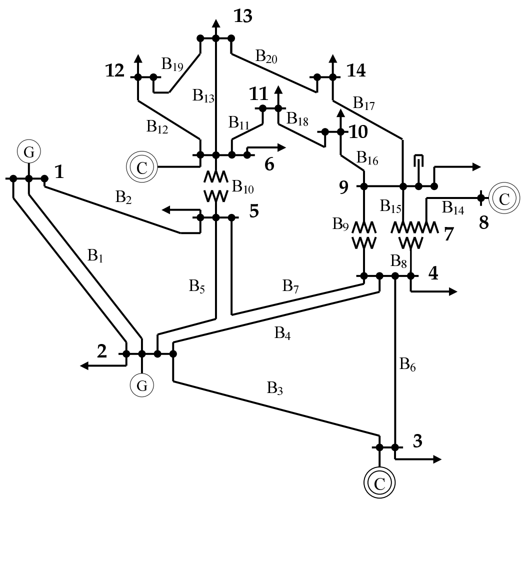

Figure 1: IEEE 14-bus network

In this test, a maximum power transfer capacity problem is introduced in the branches B_1 et B_2 (see Figure 1). To this effect, we eliminate the injection measurement on bus 2, reduce by 25% the voltage magnitude measurement at bus 2, divide the admittance of B_1 and B_2 by 10 and finally use the following active and reactive measurements : Active injections measurements are taken at busses :

$$\{3, 4, 5, 6, 8, 9, 10, 11, 12, 13\}$$

Reactive injections measurements are taken at busses :

$$\{3, 4, 6, 8, 9, 11, 12, 13, 14\}$$

Active power flows are measured at origin of branches :

$$\{\mathrm{B}_3, \mathrm{B}_5, \mathrm{B}_6, \mathrm{B}_7, \mathrm{B}_9, \mathrm{B}_{10}, \mathrm{B}_{11}, \mathrm{B}_{14}, \mathrm{B}_{16}, \mathrm{B}_{18}, \mathrm{B}_{20}\}.$$

Active power flows are measured at the end of branches :

$$\{\mathrm{B}_4, \mathrm{B}_8, \mathrm{B}_{12}, \mathrm{B}_{13}, \mathrm{B}_{14}, \mathrm{B}_{15}, \mathrm{B}_{16}, \mathrm{B}_{17}, \mathrm{B}_{19}\}.$$

Reactive power flows are measured at the origin of branches :

$$\{\mathrm{B}_3, \mathrm{B}_4, \mathrm{B}_5, \mathrm{B}_6, \mathrm{B}_7, \mathrm{B}_8, \mathrm{B}_9, \mathrm{B}_{10}, \mathrm{B}_{11}, \mathrm{B}_{13}, \mathrm{B}_{14}, \mathrm{B}_{15}, \mathrm{B}_{16}, \mathrm{B}_{17}, \mathrm{B}_{18}, \mathrm{B}_{19}, \mathrm{B}_{20}\}.$$

Reactive power flows are measured at the end of branches :

$$\{\mathrm{B}_{12}, \mathrm{B}_{14}, \mathrm{B}_{16}\}.$$

Table 1 presents the L_2 norm of the estimation error $\|\widehat{\mathbf{v}} - \mathbf{v}^*\|$, where $\widehat{\mathbf{v}}$ is the vector of the voltage magnitude estimates and $\mathbf{v}^*$ is the true value of the voltage magnitude vector. Note that the M-Arctan estimator allows a better estimation than the L_2 estimator ; thus the M-Arctan estimator is statistically more robust. The formulation proposed for updating λ generally yields better convergence properties. As usual, the residual values during the first iterations is too large. A small value for λ then makes convergence more difficult and introduce an ill conditioning in the gain matrix ($\mathbf{H}^{\mathrm{T}}\mathbf{R}^{-1}\mathbf{H}$). It is thus better to start the algorithm with a large value of λ that corresponds to a better conditioning. This is particularly important when iterating far from the final solution.

Estimator	$\|\widehat{\mathbf{v}} - \mathbf{v}^*\|_2$
L_2	68.6350
M-Arctan $\lambda = 2$	28.3497
M-Arctan ($\mathcal{P}_\lambda = 98\%$)	13.5604

Table 1: Quadratic norm of the estimation error for L_2 estimator, M-Arctan ($\lambda = 2$) and M-Arctan ($\mathcal{P}_\lambda = 98\%$).

Tables 2, 3 and 4 show the evolution of the adjustment parameters respectively for L_2, M-Arctan ($\lambda = 2$) and M-Arctan ($\mathcal{P}_\lambda = 98\%$) estimators. The number of iterations is always less than 20. All these tests are performed with a normalization of $\mathbf{s}(\mu)$ as explained in Remark 1 ; otherwise, the number of iterations increases to 61 for L_2 norm estimator and to 45 for the M-Arctan estimator (with $\lambda = 2$).

Iteration	δ	μ	J	∇J
1	0.05	1.12E+04	95.254	9.75E+02
2	0.1	1.36E+03	61.268	2.17E+02
3	0.2	2.31E+02	43.153	40.035
4	0.4	5.8003	31.236	25.695
5	0.8	0.1446	25.096	33.69
6	1.6	0.049598	24.537	3.6419
7	0.8	0.0067583	24.433	4.6639
8	0.2	0.11661	24.422	0.6217
9	0.1	0.059331	24.419	0.26568
10	0.05	0.077929	24.418	0.0096762
11	0.05	0.03392	24.418	0.0093672
12	0.00625	0.11872	24.418	0.008258
13	0.0015625	0.16209	24.418	0.0015631
14	0.00039063	0.062718	24.418	0.0005053

Table 2: L_2 norm estimator, applied to the IEEE 14 bus network in the presence of maximum power transfer capacity problems

Iteration	δ	μ	J	∇J
1	0.05	9.60E+03	45.852	4.93E+02
2	0.05	2.98E+03	27.839	3.39E+02
3	0.05	1.09E+03	20.623	135.04
4	0.1	9.61E+01	17.545	17.295
5	0.2	1.05E+00	15.733	7.4221
6	0.4	2.53E-01	15.519	2.8633
7	0.8	1.08E-01	15.473	0.18196
8	0.8	9.07E-02	15.413	0.79894
9	0.8	5.39E-02	15.382	0.48623
10	0.2	3.57E-02	15.367	0.51133
11	0.05	5.53E-02	15.365	0.0094461
12	0.025	4.02E-02	15.365	0.0036724
13	0.00625	5.69E-03	15.365	0.0021831
14	0.00078125	1.69E-01	15.365	0.00037657

Table 3: M-Arctan estimator (with $\lambda = 2$), applied to the IEEE 14 bus network in the presence of maximum power transfer capacity problems

Iteration	δ	μ	J	∇J	λ
1	0.05	1.25E+04	60.739	7.03E+02	3.6607
2	0.1	9.59E+02	26.726	2.16E+02	1.8561
3	0.1	1.15E+02	13.673	80.265	1.1572
4	0.05	1.04E+01	12.418	65.511	1.2676
5	0.05	5.01E+00	9.8054	33.871	0.97179
6	0.05	2.65E+00	9.9831	9.5612	1.018
7	0.1	8.71E-01	9.9342	0.23299	1.013
8	0.2	3.89E-01	9.896	0.17367	1.0087
9	0.4	1.84E-01	9.8421	0.21186	1.0033
10	0.8	8.00E-02	9.7488	0.32105	0.99419
11	0.8	5.69E-02	9.5953	0.34614	0.97868
12	0.4	4.98E-02	9.515	0.14088	0.97054
13	0.2	3.77E-02	9.4986	0.044499	0.96898
14	0.05	2.01E-02	9.4942	0.019415	0.96851
15	0.003125	4.05E-02	9.4932	0.0066656	0.96839
16	0.003125	7.52E-02	9.4929	0.0014973	0.96834
17	0.0015625	2.59E-02	9.4929	0.00050473	0.96834

Table 4: M-Arctan estimator ($\mathcal{P}_\lambda = 98\%$), applied to the IEEE 14 bus network in the presence of maximum power transfer capacity problems

5.2 Test on a IEEE 300 bus network

This test presents the results of a state estimation on the IEEE 300 bus network. A problem of maximum power transfer capacity is again artificially introduced on branch Q_{297}. To achieve this task the injection measurement on bus 195 is eliminated, the voltage measurement at the same bus is reduced about 25%, the admittance of branch Q_{297} is divided by 100 and a topology error is created by disconnecting branches Q_{272} and Q_{274}. Table 5 shows that the estimator M-arctan ($\mathcal{P}_\lambda = 98\%$) converges after 37 iterations. The initial value of λ is 33.319 and the final one is 0.26, indicative of the improvement in the quality of the model.

Itération	δ	μ	J	∇J	λ
1	0.00625	1.37E+08	50849	5.52E+05	33.319
2	0.05	4.58E+06	44945	1.76E+05	32.257
3	0.1	1.31E+06	33520	86186	27.216
4	0.2	2.82E+05	19710	45221	19.159
5	0.4	3.77E+04	9425.9	23061	13.354
6	0.1	1.09E+05	3484.7	16734	8.2973
7	0.00625	4.00E+06	2512.1	20530	6.8061
8	0.05	1.20E+05	2355.5	6832.1	6.6875
9	0.1	3.79E+04	2013.4	2459.2	6.0145
10	0.2	1.19E+04	1520.3	1654.4	4.9131
11	0.4	3.32E+03	1005.5	1127.7	4.1916
12	0.8	4.89E+02	461.1	1351.6	2.7388
13	0.8	1.44E+02	100.55	873.7	1.2069
14	0.025	1.46E+04	23.981	767.39	0.65348
15	0.00625	6.12E+04	16.679	519.43	0.4935
16	0.0015625	2.35E+05	14.972	507.64	0.49117
17	0.003125	5.28E+04	14.33	307.25	0.47579
18	0.00625	19555	13.781	156.92	0.46523
19	0.0125	6.95E+03	12.784	60.479	0.43822
20	0.025	2.19E+03	11.484	44.463	0.41346
21	0.05	6.22E+02	9.9627	36.79	0.38476
22	0.1	1.02E+02	8.1753	51.977	0.3328
23	0.2	1.06E+01	7.1913	25.106	0.33258
24	0.4	1.37E+00	6.3577	32.967	0.29864
25	0.4	7.23E-02	5.8423	22.657	0.27748
26	0.8	0.026868	5.7169	9.1835	0.26943
27	1.6	0.011534	5.6576	3.8712	0.26616
28	1.6	0.0057402	5.622	1.647	0.26479
29	0.2	5.00E-03	5.6078	0.70491	0.26421
30	0.05	0.0021694	5.6046	0.2919	0.26397
31	0.0125	0.0027304	5.6032	0.12524	0.26387
32	0.003125	0.009731	5.6027	0.054092	0.26382
33	0.003125	5.66E-03	5.6024	0.023523	0.2638
34	0.00078125	0.0035017	5.6023	0.010168	0.2638
35	0.00019531	0.0027475	5.6022	0.0043846	0.26379
36	9.77E-05	0.0047545	5.6022	0.0018956	0.26379
37	2.44E-05	0.0014677	5.6022	0.00082261	0.26379

Table 5: M-arctan estimator ($\mathcal{P}_\lambda = 98\%$), applied to the IEEE 300 bus network in the presence of maximum power transfer capacity problems

6 Conclusions

In this paper a new convex M-estimator, the M-Arctan, combined with the Levenberg-Marquardt algorithm has been proposed for robust state estimation. A new approach has been suggested for updating the Lagrange multiplier. The effectiveness of the algorithm proposed has been illustrated on the standard IEEE 14 and 300-bus networks and in prepense of maximum power transfer capacity problems.

REFERENCES

[1] R. Baldick, K.A. Clements, Z.P. Dzigal, and P.W. Davis. Implementating nonquadratic objective functions for state estimation and bad data rejection. *IEEE Transactions on Power Systems*, 12:376–382, February 1997.

[2] J. B. Birch. Some convergence properties of iterated reweighted least squares in the location model. *Commun. Statis. Simula. Computa*, B9(4):359–369, 1980.

[3] M. K. Çelik and A. Abur. Use of scaling in wlav estimation of power system state. *IEEE Transactions on Power Systems*, pages 684–692, May 1992.

[4] J.C. Culioli. *Introduction à l'optimisation*. Ellipses, Paris, 1989.

[5] J.E. Dennis and R.E. Schnabel. *Numerical methods for uncontrained minimization and nonlinear equations*. SIAM, Philadelphia, 1996.

[6] J. Kohlas A. Fiechter E. Handschin, F.C. Schweppe. Bad data analysis for power system estimation. *IEEE Transactions on Power Apparatus and system*, Mars/April 1975.

[7] Y. Hassaine, B. Delourme, M. Sidorkiewicz, and E. Walter. Identification of the topology of the power system networks via mixed integer inequalities. *Proc. IEEE Power Enginering Society*, June 2004.

[8] P.J. Huber. *Robust statistics*. John Wiley and Sons, New York, 1981.

[9] L. Ljung. *System identification, theory for the user*. Prentice-Hall, Englewood Cliffs, 1987.

[10] L. Mili, G. Steeno, F. Dobraca, and D. French. A robust estimation method for topology error identification. *IEEE Transactions on Power Systems*, 14:1469–1476, November 1999.

[11] A. Monticelli, O. Alsac, N. Vempati, and B. Stott. Generalized state estimation. *IEEE Transactions on Power Systems*, 13:1069–1075, August 1998.

[12] J. M. C. Pereira. *A state estimation approach for distribution networks considering uncertainties and switching*. PhD dissertation, Faculty engineering of the university of Prto, Portugal, 2001.

[13] B. T. Polyak. *Introduction to optimisation*. Optimisation Software, New York, 1987.

[14] F.C. Schweppe, J. Wildes, and D.B. Rom. Power system static state estimation. *IEEE Transactions on Power Apparatus and Systems*, pages 120–135, January 1970.

[15] D.C. Sorensen. Newton's method with a model trust region modification. *SIAM J. Numer. Anal*, pages 409–426, April 1982.

[16] S.Pajic and K.A. Clements. Globally convergent state estimation via the trust region method. *IEEE Bologna Power Tech Conference*, June 2003.

A Initialization interval for μ

The initial interval I for μ is taken equal to $[0, \frac{4\|\nabla J\|}{3\delta}]$ because if we supposed (30) is satisfied then

$$\mu \leq \frac{4\|\nabla J\|}{3\delta}. \tag{42}$$

The proof that $\frac{4\|\nabla J\|}{3\delta}$ is an upper bound for μ is by *reductio ab absurdo*. Assume that

$$\mu > \frac{4\|\nabla J\|}{3\delta}.$$

Then

$$\|\mathbf{s}(\mu)\|^2 = \sum_{i=1}^{n} \frac{\gamma_i^2}{(\lambda_i+\mu)^2} \leq \frac{\|\nabla J\|^2}{\mu^2} < (\frac{3}{4}\delta)^2.$$

However, according to (30),

$$\|\mathbf{s}(\mu)\| \geq \frac{3}{4}\delta.$$

Thus

$$\mu \leq \frac{4\|\nabla J\|}{3\delta}.$$

This implies that

$$\mu \in [0, \frac{4\|\nabla J\|}{3\delta}].$$

B Bound for μ_{k+1}

We have $\Delta\mu_k = \overline{\mu}_k - \underline{\mu}_k \geq 0$, and

$$\Delta\varphi = \varphi(\overline{\mu}_k) - \varphi(\underline{\mu}_k) = \|\mathbf{s}(\overline{\mu}_k)\| - \|\mathbf{s}(\underline{\mu}_k)\| \leq 0,$$

but $\mu \mapsto \|\mathbf{s}(\mu)\|$ is strictly decreasing.
However

$$\mu_{k+1} = \underline{\mu}_k - \frac{\|\mathbf{s}(\overline{\mu}_k)\|}{\delta}(\frac{\varphi(\underline{\mu}_k)}{\Delta\varphi/\Delta\mu_k}),$$

and

$$\varphi(\underline{\mu}_k) = \|\mathbf{s}(\underline{\mu}_k)\| - \delta \geq \frac{3}{2}\delta - \delta = \frac{1}{2}\delta \geq 0,$$

then

$$\mu_{k+1} \geq \underline{\mu}_k.$$

We also have

$$\mu_{k+1} = \overline{\mu}_k - \frac{\|\mathbf{s}(\underline{\mu}_k)\|}{\delta}(\frac{\varphi(\overline{\mu}_k)}{\Delta\varphi/\Delta\mu_k}).$$

and

$$\varphi(\overline{\mu}_k) = \|\mathbf{s}(\overline{\mu}_k)\| - \delta \leq \frac{3}{4}\delta - \delta = -\frac{1}{4}\delta \leq 0.$$

Therefore

$$\mu_{k+1} \leq \overline{\mu}_k.$$

Globally

$$\mu_{k+1} \in [\underline{\mu}_k, \overline{\mu}_k].$$