

N. Logic, E. Kyriakides, G. T. Heydt, " \mathcal{L}_p State Estimators for Power Systems," N. Logic, E. Kyriakides, G. T. Heydt, " L_p state estimators for power systems," Journal of Electric Power Components and Systems, accepted for publication, 2002.

L_p State Estimators for Power Systems

N. Logic
E. Kyriakides
G. T. Heydt

Arizona State University
PO Box 875706
Tempe, Arizona 85287 USA

The widely used method of least squares for state estimation is revisited. The commonly used least squares philosophy is based on the L_2 Hölder norm. The L_1 and L_∞ norms are considered for applications in power engineering. The effects of outliers in measurements and multicollinearity on state estimation are studied. An application in parameter estimation for synchronous generators is given as an example.

1. Introduction

Since Schweppe [1] introduced it three decades ago, state estimation for the real-time modeling of the electric power system has remained an extremely active and contentious area. To date, there have been more than a thousand research and development publications on new and improved methods. More recently, these have promoted dynamic, distributed, and non-WLS (Weighted Least Squares) approaches. State estimation is the process of usage of input data (measurements) and possible physical laws to calculate a set of numbers (system states) that satisfy input measurements and laws in some optimal way. For electric power systems the most important basis of all state estimators is the method of least squares. The method requires a process of the form,

$$z = Hx + \mathbf{h},$$

where z is an m vector of known measurements, H is the $m \times n$ process matrix, x is the unknown n vector of the true states, Hx is the m vector of the linear function linking measurements to states, \mathbf{h} is the m vector of random errors, m is the number of measurements, and n is the number of states. The objective is generally to somehow minimize \mathbf{h} . The Hölder norm [2] of any vector is the so-called L_p norm,

$$\|n\|_p = \left(\sum_i |n_i|^p \right)^{1/p}.$$

The Hölder norms [2] are a subset of norms which are defined as follows: let S be a linear space, and let $\|\cdot\|$ be a real-valued function defined on the elements of S such that,

- i) $\|x\| > 0$ unless $x = 0$
- ii) $\|Ix\| = |I| \|x\|$ where I is a scalar
- iii) $\|x + y\| \leq \|x\| + \|y\|$ (triangle inequality).

Then $\|\cdot\|$ defines a *norm* on S . When $p = 2$, one obtains the familiar least squares solution,

$$\begin{aligned} J &= \mathbf{h}^t \mathbf{h} \\ \frac{\partial J}{\partial x} &= 0 \\ \hat{x} &= H^+ z \end{aligned} \tag{1}$$

where H^+ is the pseudoinverse of H . In this paper, the case of $p \neq 2$ is considered and alternative cases will be applied to power engineering estimation problems. The specific cases of parameter estimation and measurement noise are discussed.

Consider an overdetermined static linear system of equations, with more equations than unknowns. This is the most often case, especially in power engineering in which measurements are usually plentiful. Overdetermined systems of linear equations only have a solution when the $\text{rank}(H)$ is the same as the $\text{rank}(H,z)$ which in practice may not be numerically determinable. State estimation is the determination of the system state vector x , so that the residual vector r or measurement error vector $\mathbf{h} = z - Hx$ will be, in a certain sense, small. It is natural to minimize the Euclidean length. This method of ordinary least squares (OLS) discovered independently by Gauss in 1795 and Legendre in 1805, is the only method promoted as a fitting principle in most literature. This method is preferred largely because the determination of the desired parameters is mathematically very simple. When certain assumptions are fulfilled (e.g. the linear model is correct, and the errors \mathbf{h} in the measurements z are independent and normally distributed), statistical assertions (e.g. about confidence intervals for the parameters) can be made, and the method is always the best linear “unbiased” estimator (BLUE) of x (according to the Gauss-Markov theorem, the least squares estimate has the minimum variance). Unfortunately, the mathematical elegance that makes OLS so popular depends on a number of fairly restrictive and often unrealistic assumptions. However, if \mathbf{h} is not normally distributed, then the OLS state estimates and inferences can be flawed. Violation of the normal distribution of the error term can occur when there are one or more outliers in the measurements. An outlier is an observation that is inconsistent with the remainder of the measurements, in other words, if the corresponding error or residual is large or very large compared with those of the majority of the other observations. Only after having found the fit can one identify which observations are outliers and which are not. One would like to state *a priori* whether there are outliers or not, in order to use an adequate fitting method.

2. Outliers in Measurements

Outlier detection methods could be direct and indirect. Many of the direct procedures in the literature are based on either sequential deletion (backward search) of outlying observations or sequential addition (forward search) of clean observations. In a backward search, the entire set of observations is initially considered and the outliers are sequentially removed by a criterion such as the largest absolute value of some transformed residual. The forward search works similarly. A small subset of the data is selected as the initial clean basis and clean observations are sequentially added to this basis. Methods using a forward search generally outperform backward search methods. Indirect methods for outlier detection are based on examination of the final weights (between 0 and 1) that the robust estimator assigns each observation. Residuals provide the most reliable signal to detect outliers. The cutoff values to declare an observation to be an outlier from the residual value may be computed by Monte Carlo simulation because the distribution of robust estimator residuals is not known.

If the outliers are successfully identified, they could be left in the analysis or they could be removed entirely from the analysis. Dismissal of these outliers from the analysis could be a missed opportunity to characterize the process at certain operating conditions. A compromise

between including and deleting the outliers is to introduce a weighting factor in (1) with low (or zero) weights for the measurements corresponding to the outliers.

In power engineering, outliers occur as a consequence of high amplitude noise in measurements (e.g., due to induction during large transients), momentary loss of measurement data, and noise occurring due to unintended signal paths and measurements. Unfortunately, many AC applications result in recurring measurement errors and “periodic noise.” Periodic errors are not uncommon in power electronic switched applications. The true outlier does occur in wide area measurements in which satellite communications channels are used and a variety of unintended signal processing occurs (as an example, in [3], locational data derived from the Global Positioning Satellite are described).

3. Robust Estimation

Robust estimators have been proposed as alternatives to OLS to downweigh observations as a function of “outlyingness” in state estimation. This can be achieved by choosing a scaling factor s and a suitable threshold bound T as in the following objective function,

$$\sum_{i=1}^m w_i (\mathbf{h}_i / s)^2 \rightarrow \min$$

with

$$w_i = \begin{cases} 1, & \text{for } \frac{|\mathbf{h}_i|}{s} \leq T \\ \frac{T}{|\mathbf{h}_i|/s}, & \text{for } \frac{|\mathbf{h}_i|}{s} > T \end{cases}.$$

Several robust estimation methods have been proposed:

- The M -estimator is the most popular [4], M stands for maximum likelihood
- R -estimator is based on the ranks of the residuals [5]
- L -estimator is based on linear combination of order statistics [6]
- Least Median of Squares (LMS) and Least Trimmed Squares (LTS) estimator approaches
- S -estimator is based on the minimization of a robust M -estimate of the residual scale
- The Generalized M -estimator (GM) attempts to downweigh the high influence points as well as large residual points
- Multi-stage GM-estimator is using different techniques in different stages so that the desirable properties of each technique can be combined
- MM -estimator is a multistage estimator which combines high breakdown with high asymptotic efficiency.

4. Multicollinearity

A second condition that potentially impacts the reliability of OLS estimates is multicollinearity, which is a near-linear dependency among the states. Multicollinearity can cause large variability in the state estimates, sometimes resulting in estimates that differ from the true values by orders of magnitude or have the incorrect sign. It is common to have states to behave simi-

larly relative to each other. In this case, the states have a degree of dependency or multicollinearity, which can result in an ill-conditioned matrix $H^T H$, meaning that the matrix inversion routines in obtaining $(H^T H)^{-1}$ can be very inaccurate. In general, multicollinearity tends to inflate the variance and absolute value of the least squares coefficients. Alternative estimation techniques that have been proposed induce a little bias by augmenting the matrix H , causing increased stability in the $H^T H$ matrix, and resulting in large reduction in the variance of the estimates. With biased estimation methods, such as ridge estimator, the objective function

$$\left\| \begin{pmatrix} H \\ II \end{pmatrix} x - \begin{pmatrix} z \\ 0 \end{pmatrix} \right\|_p \quad (1 \leq p \leq 8).$$

For the case $p = 2$, this is equivalent to the minimization of

$$\|\mathbf{h}\|_2 + I^2 \|x\|_2.$$

It is often not mentioned that there are many other fitting criteria for discrete linear approximations that, depending on the particular data, probably give better estimate. Corresponding numerical methods are available and easily realized, to minimize

$$\|\mathbf{h}\|_p = \left(\sum_{i=1}^m |\mathbf{h}_i|^p \right)^{1/p} \quad 1 \leq p < 8$$

or

$$\|\mathbf{h}\|_\infty = \max_{i=1:m} |\mathbf{h}_i|$$

with reference to x . In [2], numerous other fitting principles are explained and corresponding numerical methods implemented in FORTRAN 77 subroutines are discussed.

A concrete measurement data set (H, z) can, in principle, be estimated by any one of these methods. In each case, a different state vector x will be obtained. There are auxiliary means available that enable us to evaluate the results of the different methods. For example, one could determine relative class frequencies of the (residuals r_i)/(errors \mathbf{h}_i) or of the relative (residuals)/(errors) with reference to the z_i , and compare them for the different methods.

In power engineering measurements, multicollinearity is a consequence of unintended dependence between measurements. As an example, the measurement of frequency at two points – in a power system for a power system stabilizer (PSS) application may be multicollinear if the same reference clock is used for both measurements. Common mode error in measurements often results in multicollinearity. An example is the use of a common sequence ground for several voltage measurements. In this case, the measurements may be multicollinear.

5. L_p Estimation

Consider the features and facts regarding discrete linear L_p approximation where, for a given linear system of equations

$$z = Hx + \mathbf{h},$$

the length $r = Hx - z$ of the residual vector r in the L_p norm $\|r\|_p$ is minimized. This problem is written in the form

$$\min \|Hx - z\|_p, x \in R^n$$

In order to state that \hat{x} is a solution of above equation,

$$\|H\hat{x} - z\|_p = \min \|Hx - z\|_p, x \in R^n$$

For $H \hat{\mathbf{I}} R^{m \times n}$, it is generally assumed that $m > n$ (the number of measurements is greater than the number of states), and that the rank of H is n and not less than n , and almost always that $\|r\|_p \neq 0$, i. e., $r \neq 0$. The vector z is to be approximated in the L_p norm by an element Hx of the subspace

$$U = \{ y \hat{\mathbf{I}} R^m : y = Hx, x \hat{\mathbf{I}} R^n \}$$

is spanned by the columns of H . As U is finite dimensional and R^m is normed by $\|\cdot\|_p$, there always exists, according to a general theorem in approximation theory, an element $y \hat{\mathbf{I}} U$, minimizing $\|y - z\|_p$, so that y is a best approximation [7]. The set of best approximations is convex, since for two such elements $y_1, y_2 \in R^m$ with

$$\|y_1 - z\|_p = \|y_2 - z\|_p = \min_y \|y - z\|_p,$$

and for $0 \leq \mathbf{I} \leq 1$,

$$\|\mathbf{I}y_1 + (1 - \mathbf{I})y_2 - z\|_p \leq \mathbf{I}\|y_1 - z\|_p + (1 - \mathbf{I})\|y_2 - z\|_p = \|y_1 - z\|_p$$

Since the L_p norm is strictly convex for $1 < p < \infty$, y is unique [8]. One has to consider that this does not necessarily imply the uniqueness of a solution x for the problem; there can be more than one element x with $y = Hx$. It is possible to isolate x with minimal Euclidean length. If $\text{rank}(H) = n$, i.e. $\text{rank}(H)$ is maximal, then x is unique along with y . Even for $\text{rank}(H) = n$ the solution y , and thus x too, does not have to be unique when $p = 1$ and when $p = \infty$. In these cases the set of all solutions x is convex.

For $p = 1$, the method of the least sum of absolute deviations is used. In this case, it is known that outliers (large errors) in the measurements z have much less effect on the state vector x than for $p = 2$ [9]. Ironically, the first robust estimator pre-dates OLS by nearly a half century. The L_1 or least absolute deviations (LAD) estimator (introduced by Boscovich in 1757) is particularly well suited for those heavy-tailed distributions (e.g. double exponential) that can generate outliers. It is known that in the method of least absolute deviations ($p = 1$) the outliers or bad data in the measurements z have much less effect on the state vector x than for $p = 2$ (OLS). This method recommended LAD as one of the robust estimators applied to the electric power system. According to [10], this application is encountered on issues with existence of leverage points, which makes unwanted bad data to be selected as good data. Leverage points are measurements that have significant influence on state estimation, regardless of the relatively low weights they have in a LAD objective function. Although various methods to identify leverage points and to reduce their effect on state estimation have been suggested in the literature, this still is an active research area.

Consider the problem with $p > 2$. The solution of the problem

$$S_p(x) = \|Hx - z\|_p = \|r\|_p \rightarrow \min$$

is unique for $\text{rank}(H) = n$ and $1 < p < \infty$. Assume that x is a solution if and only if the gradient of S vanishes, i.e. $\nabla S = 0$ [8]. With

$$S(x) = \sum_{i=1}^m |r_i|^p$$

$$\partial S / \partial x_j = p \sum_{i=1}^m h_{ij} |r_i|^{p-2} r_i,$$

and thus using the notation

$$V = V(x) = \text{diag}(v_1, \dots, v_m), \quad v_i = v_i(x) = |r_i|^{p-2},$$

gives

$$\nabla S = p H^t V r = p (H^t V H x - H^t V z).$$

For the case of $p > 2$, the system of equations is no longer linear in the components of x , and Newton's method is applied,

$$x^{(t+1)} = x^{(t)} + \Delta x^{(t)}.$$

to the nonlinear system of equations $F(x) = \nabla S = 0$, where

$$\Delta x = -[F'(x)]^{-1} F(x), \text{ and } F'(x) = \frac{\partial F}{\partial x}$$

$$\nabla^2 S = p(p-1) H^t V H,$$

and thus,

$$\Delta x = -\frac{1}{p-1} (H^t V H)^{-1} H^t V r.$$

If $H^t V H$ is positive definite (which is guaranteed when $\text{rank}(H) = n$ and $r_i \neq 0, i = 1, \dots, m$) then the Newton correction $\mathbf{D}x$ gives a direction of descent for $S(x)$. This is because the Taylor series gives [8]

$$\begin{aligned} S(x + \mathbf{b}\Delta x) &= \|r\|_p + \mathbf{b}(\nabla S)^t \Delta x + O(\mathbf{b}^2) \\ &= S(x) - \frac{p\mathbf{b}}{p-1} r^t V H (H^t V H)^{-1} H^t V r + O(\mathbf{b}^2) \\ &< S(x) \end{aligned}$$

for $\tilde{N}S \neq 0$ and adequately small values of \mathbf{b} . Thus, one can use the damped Newton's method, i.e. $x^{(t+1)} = x^{(t)} + \mathbf{b}\Delta x^{(t)}$, where \mathbf{b} , starting with $\mathbf{b} = 1$, will occasionally be reduced in order to decrease S . Empirically, it turns out that for $p > 2$ and for the L_2 solution as starting value, each iteration was managed with $\mathbf{b} = 1$, whereas for $1 < p < 2$, $\mathbf{b} = p - 1 < 1$ must be used if the method is at all applicable in this case (because of the requirement that $r_i \neq 0$).

Consider the damped Newton's method for $\mathbf{b} = p - 1$, $x^{(t+1)} = (H^t V H)^{-1} H^t V z$ with $V = V(x^{(t)})$ i.e., $y = x^{(t+1)}$ can be obtained by solving the weighted L_2 problem

$$\min_x \sum_{i=1}^m w_i y_i - w_i z_i \quad \text{min}$$

where $W = \text{diag}(w_1, \dots, w_m)$, and $w_i = \sqrt{v_i} = \sqrt{v_i(x)} = w_i(x)$ (again assume $r_i \neq 0$). But, for $p > 2$ one can also interpret the damped Newton's method as the solution of the above equation. Indeed, for $\mathbf{b} = 1$,

$$\Delta x^{(t)} = \frac{-1}{p-1} (x^{(t)} - y).$$

Thus Newton's method is:

- i) Set $t = 0, x^{(0)} = 0, w^{(0)} = I$
- ii) In the t -th iteration find y as the solution of $\min_{x'} \sum_{i=1}^m W^{(t)} H x - W^{(t)} z_i$
- iii) Alternatively put

$$x^{(t+1)} = \begin{cases} y & \text{for } 1 < p \leq 2 \\ \frac{1}{p-1} [(p-2)x^{(t)} + y] & \text{for } 2 < p < \infty \end{cases}$$

iv) Unless the iteration has converged, set $t = t + 1$ and go to step ii.

State estimation literature is beginning to admit that $p = 2$ is not the only choice but merely one of many alternatives. The cases $p = 1, 8$ are treated comprehensively, as are values of p lying in between and not equal to 2. But, how does one choose among possible values of p in order to get the best fit for the linear model chosen? In typical studies concerning this matter, the data H, z , and the state vector x are constructed such that the errors \mathbf{h} or residuals $r = Hx - z$ obey a certain distribution. Empirical results then indicate for which values of p the particular distributions and corresponding states x were retrieved [11].

6. Applications in Power Engineering

It is desired to compare the L_p solution for the three most common values of p , namely 1, 2 and 8. It is of foremost importance to evaluate the three solutions based on the accuracy of the results and on the computation time required by each method. Both the accuracy and the computation time are certainly dependent on a number of factors such as the example itself and the level of measurement noise. Nevertheless, some general conclusions can be drawn.

For the purposes of this paper, a state estimation example has been selected. Authors in [12] describe a method to identify synchronous machine parameters from on-line measurements. This is achieved by obtaining measurements of the voltage and current values at the machine terminals and using the forward difference formula to calculate the current derivative values. These values are used in the developed model and the unknown parameters are estimated. For simplicity, the theory and the procedure behind this approach will not be repeated in this paper. The developed model is of the form,

$$V = -R \cdot I - L \cdot \dot{I},$$

where V, I and \dot{I} are vectors of length 5, while R and L are 5x5 matrices. References [12-15] discuss the model and formulation in detail. Figure 1 shows the synchronous machine model which is

$$\begin{bmatrix} v_a \\ v_b \\ v_c \\ -v_F \\ v_D \end{bmatrix} = - \begin{bmatrix} r_a & 0 & 0 & 0 & 0 \\ 0 & r_b & 0 & 0 & 0 \\ 0 & 0 & r_c & 0 & 0 \\ 0 & 0 & 0 & r_F & 0 \\ 0 & 0 & 0 & 0 & r_D \end{bmatrix} \begin{bmatrix} i_a \\ i_b \\ i_c \\ i_F \\ i_D \end{bmatrix} - \begin{bmatrix} \dot{i}_a \\ \dot{i}_b \\ \dot{i}_c \\ \dot{i}_F \\ \dot{i}_D \end{bmatrix} + \begin{bmatrix} \mathbf{v}_n \\ \mathbf{0} \end{bmatrix}. \quad (2)$$

In (2) the voltage is expressed in terms of both currents and flux linkages. This is not desirable and therefore one of the two variables has to be replaced. The mathematical model can be derived as,

$$\begin{bmatrix} v_0 \\ v_d \\ v_q \\ -v_F \\ v_D \end{bmatrix} = - \begin{bmatrix} r+3r_n & 0 & 0 & 0 & 0 \\ 0 & r & \omega L_q & 0 & 0 \\ 0 & -\omega L_d & r & -\omega kM_F & -\omega kM_D \\ 0 & 0 & 0 & r_F & 0 \\ 0 & 0 & 0 & 0 & r_D \end{bmatrix} \begin{bmatrix} i_0 \\ i_d \\ i_q \\ i_F \\ i_D \end{bmatrix} - \begin{bmatrix} L_0+3L_n & 0 & 0 & 0 & 0 \\ 0 & L_d & 0 & kM_F & kM_D \\ 0 & 0 & L_q & 0 & 0 \\ 0 & kM_F & 0 & L_F & M_R \\ 0 & kM_D & 0 & M_R & L_D \end{bmatrix} \begin{bmatrix} \dot{i}_0 \\ \dot{i}_d \\ \dot{i}_q \\ \dot{i}_F \\ \dot{i}_D \end{bmatrix}. \quad (3)$$

Equation (3) effectively is a formulation of the state variable equations in a form suitable for state estimation. Equation (3) is easily rearranged in the form $\dot{x} = Ax + Bu$ provided that the above equations are linear.

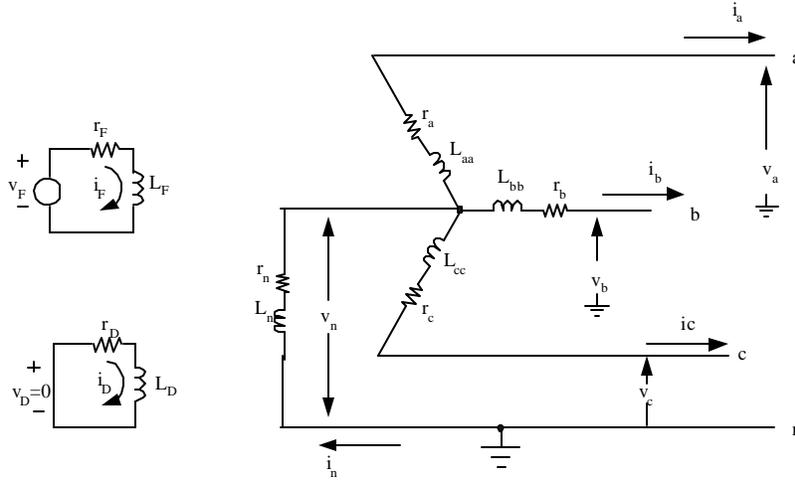


Figure 1. Schematic diagram of a synchronous machine

Two sets of measurements are used for the purposes of this example, resulting in a system of ten equations. The measurements are obtained from the simulation of the developed model, and therefore the values of all parameters are known exactly. This is of great importance, since the estimated parameters can be compared to their exact values. For illustration purposes it is desired to estimate three parameters, namely the stator phase resistance r , the equivalent field winding resistance r_F and the equivalent quadrature axis reactance L_q . Moreover, two cases will be studied. In the first case, a small amount of noise was added to the measurements. This noise was added to only one of the voltage measurements and the signal-to-noise ratio (S/N) is in the order of 7000. Essentially, the measurement sets are noise free. In the second case a more significant amount of noise is added. The S/N is in the order of 10 to 100, and the noise is added primarily to the voltage measurements. For both noise-free and noise-contaminated cases, the example eventually reduces into a matrix equation of the form,

$$H \cdot x = z,$$

where H is 10×3 , x is 3×1 and z is 10×1 .

For the $p = 1$ case, it is desired to minimize the sum of the absolute deviations of all ten equations. Let the deviation for each equation be denoted by r_i . Since r_i is either positive, negative or zero, it can be declared that

$$r_i = r_i^+ - r_i^-,$$

where $r_i^+ = 0$, $-r_i^- = 0$ and $r_i^+ \times r_i^- = 0$. The problem is solved using linear programming. After adding the artificial variables, A_i for each equation, the problem becomes,

$$\min \sum_{i=1}^{10} (r_i^+ + r_i^-) + M \cdot \sum_{i=1}^{10} A_i$$

subject to

$$[I, -I, H, -H] \cdot \begin{bmatrix} r^+ \\ r^- \\ x^+ \\ x^- \end{bmatrix} = z.$$

This formulation is readily solved by setting up a tableau according to the simplex method. There are ten initial basis variables (the artificial variables A_i), and 26 state variables in the objective function.

For the $p = 2$ case, it is desired to minimize the sum of the squares of the deviations of each row in the matrix equation. Such minimization is achieved by obtaining the pseudoinverse of matrix H ,

$$\hat{x} = H^+ \cdot z,$$

where

$$H^+ = (H^t \cdot H)^{-1} \cdot H^t$$

For the $p = 8$ case, it is desired to minimize the maximum of the absolute deviations r_i . Let the objective function be denoted by s . Therefore,

$$\min s = \max_i |r_i| = \|Ax - b\|_\infty,$$

subject to

$$\begin{bmatrix} H & e \\ -H & e \end{bmatrix} \cdot \begin{bmatrix} x \\ s \end{bmatrix} \geq \begin{bmatrix} z \\ -z \end{bmatrix},$$

where e is a vector of ones, with number of rows equal to the number of rows of H .

Since the constraints are inequality constraints, it is necessary to add both a surplus and an artificial variable to each equation, in order to solve this problem as a linear programming problem. Therefore, the linear programming problem becomes,

$$\min \left[s + M \cdot \sum_{i=1}^{20} A_i \right],$$

subject to,

$$\begin{bmatrix} H & e \\ -H & e \end{bmatrix} \cdot \begin{bmatrix} x \\ s \end{bmatrix} - [X] + [A] = \begin{bmatrix} z \\ -z \end{bmatrix},$$

where X and A are vectors that incorporate the surplus and artificial variables respectively. The initial basis variables in the tableau are the artificial variables A_i , while the total number of variables in the objective function is 44.

The results for all three cases for both low noise data and high noise data are depicted in Tables 2 and 4 respectively. The residuals for $p = 1, 2$ and 8 are computed for each set of estimated parameters and are also depicted in Tables 2 and 4. It is noted that as expected, the residuals on the diagonal are less than or equal to the other residuals in their columns. For example, for the $p = 2$ case, it is expected that the least squares residual is smaller than or equal to the least squares residuals of the $p = 1$ and $p = 8$ cases.

Tables 3 and 5 show the percent errors of the estimated parameters relative to the exact values of those parameters. An “average” error for each case is also computed to give a better idea of the accuracy of the estimation for each case. The exact values of the parameters to be estimated are shown in Table 1.

Table 1

Exact values of synchronous generator parameters to be estimated

| r (?) | r_F (?) | L_q (H) |
|----------------------|------------------------|-----------|
| 4.6×10^{-3} | 9.722×10^{-4} | 1.72 |

Table 2

Estimated machine parameters and residuals for the low noise case

| | r (?) | r_F (?) | L_q (H) | R_1 | R_2 | R_8 |
|---------|---------------------------|--------------------------|-----------|--------------------|-----------------------|--------------------|
| $p = 1$ | 4.566857×10^{-3} | 9.73697×10^{-4} | 1.719874 | 2×10^{-4} | 2×10^{-4} | 2×10^{-4} |
| $p = 2$ | 4.566857×10^{-3} | 9.73697×10^{-4} | 1.719988 | 2×10^{-4} | 1.41×10^{-4} | 1×10^{-4} |
| $p = 8$ | 4.566857×10^{-3} | 9.73697×10^{-4} | 1.719988 | 2×10^{-4} | 1.41×10^{-4} | 1×10^{-4} |

Table 3

Percent errors of estimated machine parameters for the low noise case

| | r (?) | r_F (?) | L_q (H) | “average” |
|---------|---------|-----------|-----------|-----------|
| $p = 1$ | -0.72 | 0.154 | -0.00735 | 0.2938 |
| $p = 2$ | -0.72 | 0.154 | -0.00071 | 0.2916 |
| $p = 8$ | -0.72 | 0.154 | -0.00071 | 0.2916 |

Table 4

Estimated machine parameters and residuals for the high noise case

| | r (?) | r_F (?) | L_q (H) | R_1 | R_2 | R_8 |
|---------|-------------------------|--------------------------|-----------|--------|--------|-------|
| $p = 1$ | 5.4811×10^{-3} | -228.26×10^{-4} | 1.6712 | 0.2948 | 0.1369 | 0.091 |
| $p = 2$ | 7.7669×10^{-3} | 9.1284×10^{-4} | 1.7277 | 0.2948 | 0.1103 | 0.059 |
| $p = 8$ | -57.38×10^{-3} | 0 | 1.5753 | 0.4048 | 0.1380 | 0.059 |

Table 5

Percent errors of estimated machine parameters for the high noise case

| | r (?) | r_F (?) | L_q (H) | “average” |
|---------|---------|-----------|-----------|-----------|
| $p = 1$ | 19.16 | -2448 | -2.84 | 823.3 |
| $p = 2$ | 68.85 | -6.11 | 0.45 | 25.14 |
| $p = 8$ | -1347 | 100 | -8.41 | 485.1 |

In the low noise case, it is observed that the solutions to the problem are almost identical when using any of the norms under study. Nevertheless, $p = 2$ and $p = 8$ give greater accuracy of estimation in one parameter, and a smaller residual, thus causing them to be the favored solu-

tions. The comparison is mainly based on the percent errors of each parameter and not on the resulting residual, since it is always true that [2],

$$\begin{aligned}\|x\|_2 &\leq \|x\|_1, \\ \|x\|_\infty &\leq \|x\|_1, \\ \|x\|_\infty &\leq \|x\|_2.\end{aligned}$$

This is readily verified by inspection of Tables 2 and 4, where in a given row the residuals R_i are lower as i increases.

In the high noise case the L_2 norm is clearly the best minimization norm for this particular example. The percent errors in the estimation of some of the parameters in the $p = 1$ and $p = 8$ cases are unacceptable, and therefore it seems that these cases are particularly sensitive for these noisy data. On the other hand, the method of least squares seems to be able to reduce the effect of the excessive noise in the measurements. Another important consideration in the selection of the minimization method is the computation time. Since this problem consists of only ten equations, the computation time for all three cases will be measured in microseconds and therefore, no computational burden will be imposed no matter which norm is selected. However, for larger problems, in general $p = 2$ is the fastest minimization method. To ascertain the computation time, one should consider the accuracy of the solution required and the conditioning of the H matrix.

For the L_2 norm minimization, the most popular techniques are the Cholesky factorization, the QR factorization and the Singular Value Decomposition (SVD). The SVD is the most accurate method but it is also the slowest. For the L_1 and L_∞ norm minimization, the two most widely used approaches are the linear programming based approaches and the iteratively re-weighted least squares method (IRLS). For this case, the linear programming approaches are the fastest, since they can be done in a single pass through the data, while the IRLS method requires a certain number of iterations. As a conclusion, the selection of the minimization method depends on the problem, the accuracy required and the computation time that can be allotted.

7. Conclusions

The 2-norm minimization (or method of least squares) has been the most favored minimization method in power systems. Nevertheless, it is possible to find state estimation solutions using as a basis other minimization norms such as the $p = 1$ and $p = \infty$ norms. It is possible that these techniques offer more accurate solutions depending on the application. Different minimization norms such as $p = 3, 4, \dots$ are also possible but are rejected due to their excessive computational burden.

The residuals for each row of the matrix equation under study are often an indication of the presence of outliers in the measurements. Therefore, weighting techniques can be employed to remove or downweigh the effect of outliers in the estimation of the required parameters. The selection of the L_p method is a critical factor in the effect of the outliers in the estimation process. Multicollinearity is also a condition that impacts the reliability of OLS estimates. In this case it is possible that the states behave similarly relative to each other. The resulting $H^T H$ matrix will therefore be ill-conditioned and the estimated states will be inaccurate. It is possible to employ different minimization techniques, such that some bias will be induced by augmenting the matrix

H , thus causing increased stability in the $H^t H$ matrix, and resulting in large reduction in the variance of the estimates.

The three main methods of minimization explored in this paper are $p = 1$, $p = 2$ and $p = 8$. For the 2-norm minimization, the most popular technique is the SVD since it is the most accurate. Its computational time is greater than other available methods such as the Cholesky factorization and the QR factorization but the accuracy of the method is a decisive factor in its wide use. For the 1 and infinite norm minimization, the most widely used approach is the linear programming approach. In general, the $p = 2$ minimization requires the least computation time.

A power system state estimation example was presented. Three parameters were estimated using the three aforementioned norms and the results were compared. The state estimation problem was solved using both low noise and high noise measurements. In the low noise case, it is observed that the solution to the problem is almost identical when using any of the norms under study. Nevertheless, $p = 2$ and $p = 8$ seem to give greater accuracy of estimation in one parameter, and a smaller residual, thus causing them to be the favored solutions. In the high noise case the L_2 norm is clearly the best minimization norm since it reduces the effect of the noise in the measurements.

Alternatives to the L_2 minimization have been proposed in this paper. The selection of the minimization method depends on the application, and the level of noise in the measurements. The power engineer is urged to attempt different estimation methods and compare the respective solutions for their accuracy, before reaching a definite conclusion for the particular estimation process. Different L_p methods may hold the key to a better estimation of the required parameters.

References

- [1] F. C. Schweppe, J. Wildes, "Power System Static State Estimation, Parts I, II and III," IEEE Transactions on Power Apparatus and Systems, Vol. 89, January 1970, pp. 120-135.
- [2] G. H. Golub, C. F. Van Loan, Matrix Computations, The Johns Hopkins University Press, Baltimore, 1989.
- [3] C. Mensah-Bonsu, U. Fernández, G. T. Heydt, Y. Hoverson, J. Schilleci, B. Agrawal, "Application of the Global Positioning System to the Measurement of Overhead Power Transmission Conductor Sag," IEEE Transactions on Power Delivery, accepted for publication, 2001
- [4] P. J. Huber, Robust Statistics, John Wiley and Sons, New York, 1981.
- [5] L. A. Jaeckel, "Estimating Regression Coefficients by Minimizing the Dispersion of Residuals," Annals of Mathematical Statistics, Vol. 5, 1972, pp. 1449-1458.
- [6] P. J. Bickel, "On Some Analogues to Linear Combination of Order Statistics in the Linear Model," Annals of Statistics, Vol. 1, 1973, pp. 597-616.
- [7] M. J. D. Powell, Approximation Theory and Methods, Cambridge University Press, Cambridge UK, 1981.
- [8] G. A. Watson, Approximation Theory and Numerical Methods, John Wiley and Sons, New York, 1980.
- [9] P. Bloomfield, W. L. Staiger, Least Absolute Deviations: Theory, Applications, and Algorithms, Birkhäuser, Basel, Switzerland 1983.
- [10] A. Monticelli, State Estimation in Electric Power Systems – A Generalized Approach, Kluwer Academic Publishers, Boston, 1999.

- [11] H. Späth, Mathematical Algorithms for Linear Regression, Academic Press, Boston, 1992.
- [12] E. Kyriakides, G. T. Heydt, "Synchronous Machine Parameter Estimation Using a Visual Platform," IEEE Power Engineering Society Summer Meeting, July 2001.
- [13] H. B. Karayaka, A. Keyhani, B. Agrawal, D. Selin, G. T. Heydt, "Methodology Development for Estimation of Armature Circuit and Field Winding Parameters of Large Utility Generators," IEEE Transactions on Energy Conversion, Vol. 14, No. 4, December 1999, pp. 901-908.
- [14] H. B. Karayaka, A. Keyhani, B. Agrawal, D. Selin, G. T. Heydt, "Identification of Armature Circuit and Field Winding Parameters of Large Utility Generators," IEEE Power Engineering Society Winter Meeting, Vol. 1, 1999, pp. 29-34.
- [15] P. M. Anderson, A. A. Fouad, Power System Control and Stability, Iowa State University Press, Ames, Iowa, 1977.