USE OF SCALING IN WLAV ESTIMATION OF POWER SYSTEM STATES

Mehmet K. Çelik and Ali Abur

Department of Electrical Engineering
Texas A&M University
College Station, Texas 77843

Abstract: Scaling has been widely used in linear programming problems for numerical stability and computational efficiency. This paper investigates the effectiveness of scaling in avoiding the leverage points in weighted least absolute value (WLAV) estimation of power system states. Existence of leverage points was claimed to be the reason for the WLAV estimator failing to reject bad data in the measurements. Simulation results indicate that scaling not only helps to reduce the leveraging effects but also improves the computational performance by decreasing the required number of Simplex iterations.

Keywords: Linear Programming, Least Absolute Value Estimation, Outliers, Power System State Estimation, Leverage Points, Scaling

I. INTRODUCTION

Since the time when Merrill and Schweppe [1] introduced the nonquadratic cost function, the problem of making the power system state estimator more robust has been the subject of many papers. The robustness of an estimator is defined as its ability to remain insensitive to bad data in the measurements.

The weighted least absolute value (WLAV) estimator was first considered by Irving, Owen and Sterling [2] as such a robust estimator. Early papers reported results confirming the bad data rejection properties of the WLAV estimator [3]. However, it was later shown by Falcão and Assis [4] that the WLAV estimator may indeed fail to reject bad data under certain measurement configurations. This failure was explained by the existence of what is referred to as "leverage points of regression" in the measurement models (discussion of Mili to [4]). It has been shown that the leverage measurements are exactly satisfied by the WLAV estimator and errors in these measurements will not be detected [5, 7]. Least median of squares estimator [5] has been studied recently as an alternative and more robust estimator for power systems [6]. Use of linear transformations to eliminate the leveraging effects has also been proposed [7].

The computational complexity is another issue of concern when considering an on-line application function such as state estimation [8, 9, 10]. The proposed transformation of [7] involved axis rotations which introduced extra fill-ins in the measurement equations. In this paper, a much simpler transformation which preserves the original sparsity structure of the measurement equations is investigated. It involves the row/column scaling of the measurement Jacobian. Various scaling methods were suggested for the linear programming problems in the past [11, 12]. However, no method has so far been singled out as the best for all applications. The performance of scaling methods depends heavily on the problem structure and therefore should be carefully evaluated for the case of the power system state estimation.

This paper presents the results of employing a particular scaling strategy in WLAV state estimation of various test systems of sizes as large as 1500 buses. Apart from the effects of scaling on the leverage points and the robustness of the estimator, effects on computational speed and the dependence of these effects on the measurement types are also investigated. It is demonstrated that the use of scaling can significantly decrease the required number of Simplex iterations at each state estimation iteration. Both the CPU timings and iteration counts are given for easy comparison.

II. SCALING LINEAR PROGRAMMING PROBLEMS

Efficient solution of linear programming (LP) problems has been shown to depend upon the proper scaling of its constraint equations. Two reasons are usually given for the common use of scaling in LP problem solutions [11]:

- the number of Simplex iterations required by the LP problem solution is reduced;

- the condition number of the basis, thus, the numerical stability of the LP problem is improved.

Consider the LP problem given below:

$$minimize \ c^T \cdot x \tag{1}$$

$$subject \ to \ [B:I] \cdot x \ = \ y \tag{2}$$

$$x \ \geq \ 0 \tag{3}$$

The constraint equation, Eq. (2), can be scaled as:

$$S_r[B:I]S_c \cdot \tilde{x} = S_r \cdot y \tag{4}$$

where S_r and S_c are $(m \times m)$ and $[(m+n) \times (m+n)]$ diagonal matrices for row and column scaling of the constraint matrix, respectively. B is an $(m \times n)$ matrix. I denotes the $(m \times m)$ unity matrix. x and y are the state and righthand-side vectors with $(m+n)$ and m components, respectively. $\tilde{x}$ is equal to $S_c^{-1} \cdot x$.

To preserve the unitary columns of I in Eq. (2), the corresponding entries of S_c can be chosen such that:

$$S_{cj} = S_{ri}^{-1} \quad for \quad i = 1,...,m; \quad j = i+n \tag{5}$$

Then Eq. (4) can be rewritten as:

$$[\tilde{B} : I] \cdot \tilde{x} = \tilde{y} \tag{6}$$

where :

$$[\tilde{B} : I] = S_r \cdot [B : I] S_c \tag{7}$$
$$\tilde{y} = S_r \cdot y \tag{8}$$

Among several possible ways of choosing the scales (the elements of S_r and S_c), normalization is computationally the most inexpensive [11]. In this method, each column is divided by the largest entry in that column, after which the same procedure is applied to the rows. Since the columns $(n+1)$ through $(n+m)$ contain unit (1.0) values in the original constraint matrix, the corresponding entries of S_c are initially chosen as 1.0. Subsequent to row scaling, these entries of S_c have to be equated to the corresponding entries of S_r^{-1} in order to restore the unity matrix of Eq. (6).

III. WLAV ESTIMATION USING SCALING

The linearized measurement equations can be written as:

$$z = H \cdot x + e \tag{9}$$

where z is the $(m \times 1)$ measurement mismatch vector and x is the $(n \times 1)$ state update vector. H denotes the $(m \times n)$ measurement Jacobian matrix. e is the error vector with m components.

WLAV estimation of x can be equivalently written as the following LP problem [15]:

$$minimize \quad \omega^T \cdot (u + v) \tag{10}$$
$$subject \ to \ H \cdot x + I \cdot u - I \cdot v = z \tag{11}$$
$$u, v \geq 0$$

The constraint equation, Eq. (11), can be scaled as:

$$S_r \cdot H \cdot S_c \cdot \tilde{x} + S_r \cdot I \cdot S_r^{-1} \cdot \tilde{u} - S_r \cdot I \cdot S_r^{-1} \cdot \tilde{v} = S_r \cdot z \tag{12}$$

where S_r is an $m \times m$ diagonal matrix of row scales and S_c an $n \times n$ diagonal matrix of column scales for the columns of H.

Note that the new slack variables $\tilde{u}$, $\tilde{v}$ are related to the measurement residuals u, v through the row scaling matrix S_r. Therefore, once the modified LP problem of Eq. (10,12) is solved, the solution of the original problem can be recovered as follows:

$$x = S_c \cdot \tilde{x} \tag{13}$$
$$u = S_r^{-1} \cdot \tilde{u} \tag{14}$$
$$v = S_r^{-1} \cdot \tilde{v} \tag{15}$$

IV. EFFECTS OF SCALING ON THE LEVERAGE POINTS

A brief overview of the concept of leverage points will be given next. A detailed description can be found in [5, 6, 7, 13]. The least squares measurement estimate, $\hat{z}$, in a linear regression model of the form:

$$z = H \cdot x + e \tag{16}$$

will be given by:

$$\hat{z} = K \cdot z \tag{17}$$

where $K = H(H^T H)^{-1} H^T$ is called the "hat" matrix [14].

The elements of the hat matrix K_{ij} represent the influence of the measurement j on the estimate of measurement i. It can be shown that $0 \leq K_{ii} \leq 1$ [5]. Those measurements having K_{ii}'s close to unity are referred to as "leverage" points and the errors in such measurements are difficult to identify. The WLAV estimator may also fail to reject these measurements even when they are grossly erroneous. Rows that contain elements which are significantly different than the ones in the remaining rows tend to have high values of K_{ii}. In power systems, the following conditions have been identified for causes of leverage points [6]:

- flow and/or injection measurements at the terminals of relatively short lines;

- injection measurements at buses which have a large number of immediate neighbors.

When scaling is used, the original matrix H will be replaced by $S_r H S_c$ yielding a new hat matrix $\check{K}$ given by:

$$\check{K} = S_r H (H^T S_r^2 H)^{-1} H^T S_r^T \tag{18}$$

Even though Eq. (18) suggests that column scaling, S_c, has no effect on the elements of $\check{K}$, this is actually not the case. The row scales (elements of S_r) are chosen based on the largest row entry in $H \cdot S_c$. They are functions of the choice of S_{ci}. Hence, the subsequent choice of S_{ri}'s and the resulting $\check{K}$ will be affected by choosing different S_{ci}'s. Simulations using different systems will be presented in section VI to demonstrate these dependencies in an empirical manner.

V. LEVERAGE POINTS IN WLAV ESTIMATION

The effects of leverage measurements on the (weighted) least squares measurement estimates, $\hat{z}$, is easily understood through the hat matrix analysis; although, the diagonal elements, K_{ii}, of the hat matrix are not a robust measure for determining leverage points by any means [5]. Furthermore, this analysis is true only for the (weighted) least squares estimator. However, it has been empirically shown that the WLAV estimator also fails to reject the bad data in the leverage measurements [5, 7]. In what follows, an example will be used to demonstrate this failure of the WLAV estimator in the presence of leverage points.

At each iteration of the equivalent linear programming (LP) problem of WLAV estimation, two decisions are made. First one is the selection of the variable to enter the basis by finding the nonbasic variable with the most negative relative cost. In the case all relative costs are nonnegative, the current solution

is optimal and the LP is terminated. The second decision is the selection of the variable to leave the basis (pivot). The basic variable with the minimum ratio of $rat = xa/yd$, for $yd > 0$, is chosen as the pivot. xa and yd denote the current basic solution and the updated coefficients of the column to enter the basis, respectively. In the case all $yd < 0$, the problem is unbounded.

Table 1: Pivots and relative costs without scaling

LP Iter No.	Basic Variables and Pivots						Rel. Costs of Nonbasic Var.s		
1	s_1	s_2	s_3	s_4	s_5	s_6	x_1	x_2	x_3
xa		1.4	0.5			0.2	-2.2	-0.3	-3.5*
yd	-	7.7	2.4	-	-	1.0			
rat		0.1*	0.2			0.2			
2	s_1	x_3	s_3	s_4	s_5	s_6	x_1	x_2	s_2
xa	0.5		0.1		0.1	0.-	-2.2	-3.2*	NA
yd	2.5	NA	2.4	-	1.0	1.0			
rat	0.2	.	0.-		0.1	0.-*			
3	s_1	x_3	s_3	s_4	s_5	x_2	x_1	s_6	s_2
xa	0.4		0.-	0.1			-2.2*	NA	NA
yd	2.4	NA	2.5	1.0	-	NA			
rat	0.2		0.-*	0.1					
4	s_1	x_3	x_1	s_4	s_5	x_2	s_3	s_6	s_2
xa	0.4			0.1			-0.4*	1.0	0.5
yd	1.0	NA	NA	0.400	-	NA			
rat	0.4			0.2*					
5	s_1	x_3	x_1	s_3	s_5	x_2	s_4	s_6	s_2
xa	0.3				0.-		NA	-0.-*	0.-
yd	2.5	NA	NA	-	1.0	NA			
rat	0.1				0.-*				
6	s_1	x_3	x_1	s_3	s_6	x_2	s_4	s_5	s_2
							NA	NA	0.-

Table 2: Pivots and relative costs with scaling

LP Iter No.	Basic Variables and Pivots						Rel. Costs of Nonbasic Var.s		
1	s_1	s_2	s_3	s_4	s_5	s_6	x_1	x_2	x_3
xa	0.5		0.5	0.1			-1.*	-0.3	-0.8
yd	1.0	-	1.0	0.5	-	-			
rat	0.5		0.5	0.2*					
2	s_1	s_2	s_3	x_1	s_5	s_6	s_4	x_2	x_3
xa		1.4	0.3			0.8	NA	-0.3	-0.8*
yd	-	1.0	0.3	NA	-	0.5			
rat		1.4	1.0*			1.6			
3	s_1	s_2	x_3	x_1	s_5	s_6	s_4	x_2	s_3
xa	0.3				0.4		NA	-0.3*	NA
yd	0.3	-	NA	NA	0.5	-			
rat	1.0				0.8*				
4	s_1	s_2	x_3	x_1	x_2	s_6	s_4	s_5	s_3
xa		1.2				0.2	-3.2*	0.7	-2.1
yd	-	6.0	NA	NA	NA	3.2			
rat		0.2				0.1*			
5	s_1	s_2	x_3	x_1	x_2	s_4	s_6	s_5	s_3
xa	0.2					0.-	NA	0.7	-0.5*
yd	1.0	-	NA	NA	NA	0.5			
rat	0.2					0.1*			
6	s_1	s_2	x_3	x_1	x_2	s_3	s_6	s_5	s_4
							NA	0.7	NA

Consider the 3 bus system with 3 line flows and 3 volt-

age measurements as shown in Fig. 1. Line $2 - 3$ is relatively shorter than the rest. Measurement 2, which corresponds to
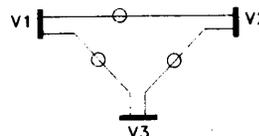

Figure 1: Line diagram of the 3 bus system

the line flow measurement for line $2 - 3$, is a suspected leverage point. Bad data is introduced to this measurement and the estimates are obtained by using the WLAV and WLAV with scaling (WLAV-S) estimators. The relative costs of the nonbasic variables and xa, yd as well as their ratio, rat, of the basic variables are given for the first $Q - V$ iteration of the decoupled SE in Tables 1 and 2 for the WLAV and WLAV-S estimators, respectively. The numbers in Tables 1 and 2 are given to the nearest single decimal point due to space limitation. Those nonzero numbers less than 0.1 are indicated by 0.— sign to differentiate them from zero. yd_i which are less than zero are shown by (−) sign. Both the variables to enter the basis and the pivots are marked by an * sign. During the first three LP iterations, only the state variables are allowed to enter the basis. Once the state variables are in the basis, they are not allowed to leave it. After the third iteration, phase 2 of the revised Simplex procedure is carried out [9]. During this phase, the slack variable, s_i, that corresponds to either u_i or v_i, which has left the basis is not allowed to reenter it (for nondegenerate solutions, no cycling occurs). For the case where no scaling is used, the slack variable, s_2, corresponding to measurement 2 (bad data as well as a leverage point) leaves the basis at the first LP iteration. At this iteration, although xa_2 is large (an indication of bad data), yd_2 is also very large, leading to the minimum ratio. Moreover, its relative cost never becomes negative enough for it to reenter the basis during phase 2. When scaling is used, yd_2 remains small during all LP iterations. As seen in Table 2, s_2 never leaves the basis. Thus, the bad data (measurement number 2) is rejected. Note that xa is dependent on the righthand-side vector and yd on the column elements of a variable in the original Jacobian matrix (both are updated versions with respect to the current basis). yd_i will tend to be large for a leverage measurement and leave the basis if the corresponding xa_i is not large enough to keep rat_i larger than the rest. Scaling improves the robustness of the WLAV estimator by spreading the elements of the Jacobian matrix evenly and making yd smaller. However, many different scaling methods have been proposed for linear programming and their performance is highly dependent on the problem structure as stated before [11]. Among the empirical methods (normalization, using geometric or arithmetic means) suggested in [11], normalization was found to be the most effective in terms of reducing the leveraging effects and the CPU times.

VI. SIMULATIONS

Simulations are carried out using the weighted least absolute value (WLAV) method and the WLAV with scaling (WLAV-S) method described in section III. The WLAV estimator used for comparison is based on the algorithm given in [9] which uses a reduced dimension basis and vertex skipping proposed by Barrodale and Roberts [15]. The state estimation

Table 3: Results for the AEP 14 bus system

| Measurement | | Residuals | | | |
Number i	Type	WLAV	WLAV-S	K_{ii}	$\tilde{K}_{ii}$
1	Fl 1-2	-0.016	-0.034	0.210	0.187
2	Fl 1-5	0.001	0.000.	0.087	0.365
3	Fl 2-3	0.003	0.000	0.105	0.193
4	Fl 2-4	-0.006	0.000 .	0.095	0.452
5	Fl 2-5	-0.011	0.054	0.096	0.339
6	Fl 3-4	0.000	0.006	0.170	0.239
7	Fl 4-5	0.000	0.000	* 0.940	0.545
8	Fl 4-7	0.000	-0.004	0.606	0.388
9	Fl 4-9	0.063	0.026	0.079	0.744
10	Fl 5-6	0.030	-0.081	0.419	0.492
11	Fl 6-11	0.042	0.000	0.209	0.336
12	Fl 6-12	-0.002	0.005	0.155	0.258
13	Fl 6-13	0.000	0.000	0.485	0.297
14	Fl 7-8	0.002	0.000	0.371	0.269
15	Fl 9-7	-0.269	0.030	0.525	0.334
16 (BD)	Fl 9-10	0.224	-0.437	0.239	0.290
17 (BD)	Fl 10-9	-0.221	0.418	0.239	0.290
18	Fl 9-14	0.000	-0.015	0.494	0.370
19	Fl 10-11	-0.061	0.000	0.161	0.231
20	Fl 11-10	0.073	-0.052	0.161	0.231
21	Fl 12-13	0.000	-0.008	0.120	0.402
22	Fl 13-14	0.065	0.000	0.459	0.428
23	In 1	0.000	-0.015	0.691	0.277
24	In 2	0.000	0.041	* 0.825	0.378
25	In 3	0.003	-0.009	0.715	0.301
26	In 6	0.000	0.001	* 0.806	0.659
27	In 8	0.000	-0.004	0.579	0.420
28 (BD)	In 9	0.000	0.345	* 0.799	0.668
29	In 10	0.000	0.000	0.542	0.339
30	In 11	0.012	0.002	0.604	0.322
31	In 12	-0.004	0.000	0.720	0.408
32	Vl 1	0.004	0.048	0.218	0.412
33	Vl 2	0.000	-0.001	0.211	0.651
34	Vl 3	-0.004	0.000	0.226	0.336
35	Vl 6	0.000	0.000	0.303	0.660
36	Vl 8	0.002	0.000	0.337	0.490

Table 4: Results for the AEP 30 bus system

| Measurement | | Residuals | | | |
Number i	Type.	WLAV	WLAV-S	K_{ii}	$\tilde{K}_{ii}$
22	Fl 15-18	0.011	-0.002	0.189	0.208
23	Fl 18-19	0.074	-0.001	0.251	0.203
24 (BD)	Fl 19-20	0.000	0.266	0.436	0.229
26	Fl 10-20	-0.043	0.000	0.354	0.253
40	Fl 8-28	0.005	-0.027	0.109	0.349
44	In 5	0.003	0.006	* 0.737	0.348
55	In 18	0.078	0.024	0.632	0.276
56 (BD)	In 19	0.000	0.321	* 0.691	0.472
83	Vl 19	-0.011	0.000	0.147	0.598
84	Vl 20	0.014	-0.031	0.172	0.393

Table 5: Results for the AEP 57 bus system

| Measurement | | Residuals | | | |
Number i	Type	WLAV	WLAV-S	K_{ii}	$\tilde{K}_{ii}$
3	Fl 3-4	0.000	-0.072	0.533	0.244
8	Fl 8-9	0.000	0.079	0.650	0.230
12 (BD)	Fl 9-12	-0.310	4.259	0.027	0.445
17	Fl 1-16	-0.098	0.000	0.110	0.330
25	Fl 12-13	-0.031	-0.132	0.272	0.438
26 (BD)	Fl 12-16	0.000	-0.389	0.228	0.245
27	Fl 12-17	-0.018	0.000	0.112	0.289
83	In 9	0.000	0.126	0.874	0.612
84 (BD)	In 12	0.020	0.269	0.836	0.435
86	In 16	0.406	0.050	0.700	0.375
91	In 21	0.034	-0.002	0.680	0.572

problem is formulated in the polar coordinate system using the $P - Q$ decoupling assumptions in [16]. Thus, the linearized active and reactive measurements are solved alternatingly until convergence. Convergence tolerances for the state variables, i.e., voltage magnitudes and phase angles, are taken as 0.001 p.u. The measurements are generated by adding low variance noise to the exact values calculated by using a power flow program.

One of the main objectives of the simulations was to test the robustness of the WLAV-S algorithm in the presence of bad data. Of particular interest are the multiple interacting bad data in the measurements corresponding to leverage points because WLAV has been shown to fail to reject them [7]. The highly influential measurement, z_i, on its estimate, $\hat{z}_i$, i.e., a leverage point, is situated far away from the bulk of the measurements in the factor space of regression [6]. The original constraint matrix in the WLAV formulation of the power system estimation may have elements of different magnitudes because of the reasons stated in section IV. Since the purpose of scaling applied in this study is to make all the elements of the constraint matrix nearly of the same magnitude, the number of the leverage points is expected to decrease. To investigate the

behavior of WLAV-S in the presence of bad data in leverage points, the test cases given in [7] are considered here.

Multiple interacting bad data in measurements corresponding to leverage points are simulated for AEP's 14, 30 and 57 bus systems and the results, reported for the reactive problem, are presented in Tables 3, 4 and 5. The first two columns in these tables give the measurement number and type. The bad data measurements are indicated by the "BD" sign next to their number. Columns 3 and 4 list the linearized residuals, u and v, for WLAV and $\tilde{u}$ and $\tilde{v}$ for WLAV-S at the last SE iteration, respectively. To further investigate the effect of scaling on leverage points, the diagonal elements of the "hat" matrix, K_{ii}, are given in column 5 for H and $\tilde{K}_{ii}$ in column 6 for the scaled Jacobian. K_{ii} and $\tilde{K}_{ii}$ which are higher that the threshold $2n/m$ are marked by an asterisk as suspected leverage points. The choice of this threshold is explained in [7, 14]. Since bad data are introduced to measurements of the reactive $(Q - V)$ problem, m stands for the total number of Q-flows, Q-injections and V-magnitude measurements and n for the number of voltage magnitudes, i.e., number of buses.

Table 3 presents the results for the 14 bus system. Four

Table 6: Test system measurement configurations

Test Systems	Line Flows	Bus Injections	Voltage Magnitudes	Reactive Redundancy
118-A	179	34	28	2.04
245-A	346	86	74	2.07
498-A	687	169	144	2.01
986-A	1321	367	312	2.03
1512-A	1974	555	472	1.98
118-B	178	26	15	1.86
245-B	346	43	63	1.84
498-B	687	145	80	1.83
986-B	1321	186	80	1.61
1512-B	1974	182	80	1.48

Table 7: Relative CPU times

Test Systems	Alg. 1	Alg. 2	Alg. 3	Alg. 4
118-A	1.	0.804	0.780	0.612
245-A	1.	0.743	0.907	0.618
498-A	1.	0.635	0.737	0.522
986-A	1.	0.496	0.391	0.241
1512-A	1.	0.480	0.389	0.293
118-B	1.	0.715	0.959	0.721
245-B	1.	0.744	1.032	0.818
498-B	1.	0.642	0.753	0.472
986-B	1.	0.605	0.840	0.576
1512-B	1.	0.621	0.946	0.626

measurements, namely 7, 24, 26 and 28, are marked as suspected leverage points. Multiple interacting bad data are introduced to three (16, 17 and 28) of the measurements at bus 9, one of which is a suspected leverage point. The estimate given by WLAV is biased as WLAV fails to reject measurement 28, i.e., its residual is zero. As given at column 5, WLAV-S rejects all three bad data, i.e, their residuals are all nonzero. The number of suspected leverage points is also decreased from 4 to 0 after after scaling has been applied.

The Tables 4 and 5 present only the measurements of interest for their high K_{ii} (or $\bar{K}_{ii}$) values and the measurements corrupted with bad data. Two interacting bad data (24 and 56), one of which is again a suspected leverage point, are introduced to the measurements of 30 bus system. As it can be seen from the linearized residuals given in columns 3 and 4 of Table 4, WLAV-S rejects both of the bad data measurements; whereas, WLAV incorrectly accepts both measurements as good. Table 5 lists similar results for 57 bus system where there are three interacting bad data at bus 12. WLAV again fails to reject all three. However, WLAV-S provides an unbiased estimate by assigning nonzero residuals to all of them. The number of K_{ii} larger than the $2n/m$ threshold decrease from 18 to 3 for the 30 bus system; whereas, increase from 3 to 4 for the 57 bus system.

K_{ii}'s of the rows corresponding to the voltage measurements are found to be sensitive to the value they are normalized

to. When normalized to 1.0 during the initial trial simulations, scaling resulted in highly influential voltage measurements for all three cases. Therefore, the normalization value of these rows had to be adjusted carefully. In all of the examples presented here, the rows corresponding to the voltage measurements are normalized to 0.5.

The second objective of the simulations was to investigate the effect of scaling on the number of LP iterations and the total CPU time. For this purpose large scale power systems with 118, 245, 498, 986 and 1512 buses are studied. The two different measurement configurations tested for each system and the redundancies for the $Q - V$ problem are given in Table 6. The first group of measurement sets (group A) has a redundancy close to 2 as well as a relatively high number of injection measurements. For the second group (group B), the redundancy and the number of injection measurements drop as the system size increases.

The effects of scaling on LP iterations is tested with the following algorithms:

1. **Algorithm 1** : The algorithm given in [9] which exploits the special structure of the constraint matrix, effectively reducing the dimension of the basis from m to n, and which uses vertex skipping proposed in [15].

2. **Algorithm 2** : Algorithm 1 with the initial LP iterations limited.

3. **Algorithm 3** : Same as algorithm 1 except the the constraint matrix is scaled as described in section III.

4. **Algorithm 4** : Algorithm 3 with the initial LP iterations limited.

The reason for limiting the LP iterations is to reduce the total number of LP iterations until SE converges. The limit is gradually relaxed becoming ineffective at higher SE iterations. The limits in this study were found by trial and error.
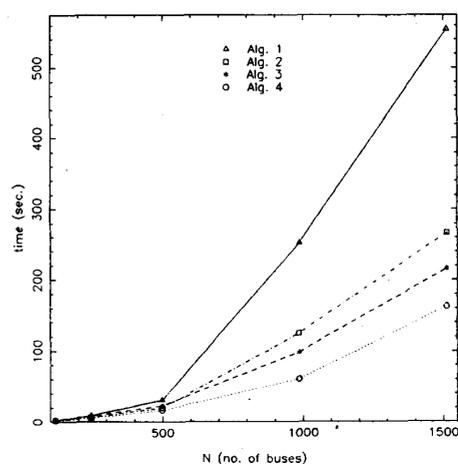

Figure 2: CPU times for data sets in group A

The simulations for these large scale power systems are run on a MIPS M/120 computer using the four algorithms listed above. The total CPU times required by algorithms 2, 3 and 4 relative to algorithm 1 are given in Table 7. In Figure 2,
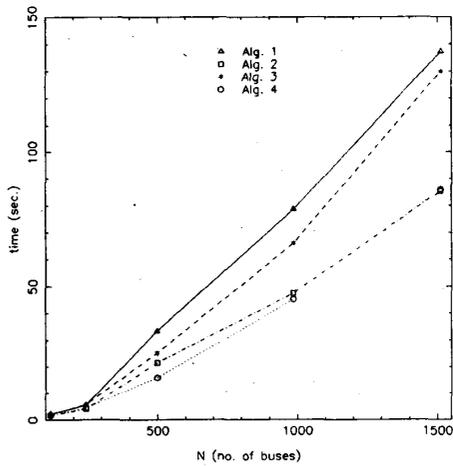
Figure 3: CPU times for data sets in group B

the actual CPU times are plotted against the number of the buses for the measurement sets in group A. Figure 3 presents the actual CPU times versus the number of buses for the data sets in group B. The CPU times seem to increase as a function of the number of buses squared for the measurement sets in group A. They are also significantly reduced by the application of scaling. For the measurement sets in group B, the CPU times are observed to increase linearly with growing system size. Also the CPU time gain obtained by scaling is not as high for group B as is for group A. The reductions in CPU times can be explained by the following. A properly scaled LP problem is expected to reach the optimal solution in fewer LP iterations because relative costs of the non-basic variables provide a better indication of their contribution in minimizing the cost function, resulting in the best local move [11].

To further show the effect of scaling on the numerical behavior of LP, the total number of LP iterations for the active and reactive problems at each SE iteration are given for groups A and B in Tables 8 and 9, respectively. As seen in Table 8, scaling results in significant reduction in the total number of

Table 8: Effects of scaling on data sets in group A

S.E. Iter. No.	118-A No. of LP Ite.		245-A No. of LP Ite.		498-A No. of LP Ite.		986-A No. of LP Ite.		1512-A No. of LP Ite.	
	Al. 1	Al. 3	Al. 1	Al. 3	Al. 1	Al. 3	Al. 1	Al. 3	Al. 1	Al. 3
1	84	82	159	160	314	286	621	559	993	761
2	61	47	117	92	208	168	424	357	678	538
3	29	19	43	45	83	55	209	102	295	223
4	10	5		10	25	12	73	19	103	62
5							18	7	10	16
6								1		

LP iterations. The effect of scaling is not that pronounced for the data sets in group B. However, results plotted in Figure 3 reveal that scaling is still effective in reducing the actual CPU time. These observations can be explained by the improvement in the numerical stability of LP after scaling is carried out. Ill-

Table 9: Effects of scaling on data sets in group B

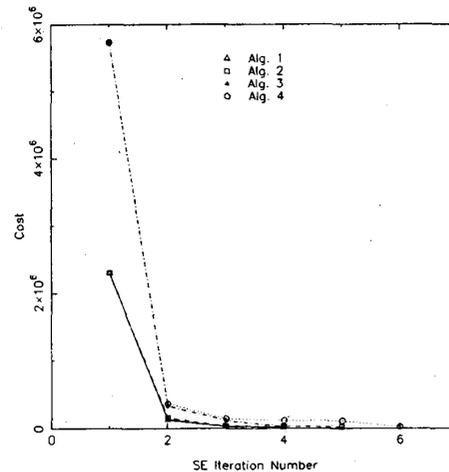| S.E. Iter. No.	118-B No. of LP Ite.		245-B No. of LP Ite.		498-B No. of LP Ite.		986-B No. of LP Ite.		1512-B No. of LP Ite.	
	Al. 1	Al. 3	Al. 1	Al. 3	Al. 1	Al. 3	Al. 1	Al. 3	Al. 1	Al. 3
1	71	88	125	131	287	293	395	441	517	589
2	64	65	80	91	194	181	262	281	326	379
3	28	27	32	32	80	47	108	82	152	133
4	2	9	1	6	27	8	47	28	64	33
5	1					1	10	3	10	8
6							1			

Figure 4: Costs for 118-B bus system

conditioning of the information matrix used for the solution of the weighted least squares estimation has been reported for measurement configurations with a high number of injection measurements [17]. It has also been shown in [17] that the condition number of the normal equations, defined as $||A||||A^{-1}||$ (where A is the information matrix and $|| \cdot ||$ is L_1 norm), is significantly higher for injections-only measurement sets compared to measurement sets containing only line flows. The results plotted in Figures 2 and 3 support these observations. Scaling seems to be more effective in reducing the CPU time and number of LP iterations for measurement configurations with a high number injection measurements. It has been also stated in [18] that the round-off errors of an ill-conditioned LP problem increase quite fast as the basis is refactorized several times before reaching the optimal solution. These round-off errors result in the reduction of the computational efficiency of LP because of the increase in the number of insignificant nonzeros. Scaling serves to decrease the round-off errors which accumulate because of elements of different magnitude. It also serves to decrease the number of significant non-zero elements deleted incorrectly by the zero check done during refactorization process. Another parameter is used during the refactorization to control the choice of pivots, referred to as u in [17]. Low values of u help maintaining sparsity; whereas, high values favor numerical stability. A value of $u = 0.1$ is used both in WLAV and WLAV-S algorithms. The parameter u seems to be

working more efficiently in maintaining the sparsity when the constraint matrix is scaled as described above. Thus, there is still a reduction in CPU times; although, there is no significant reduction in the number of LP iterations for the data sets in group B, where the constraint matrix isn't "ill-conditioned".

Finally, the cost functions for the four algorithms for the measurement set B of 118 bus system plotted in Figure 4 show that all four algorithms converge to the same optimal solution.

VII. CONCLUSIONS

In this paper, the effects of row/column scaling by normalization on the performance of the WLAV estimator are investigated. It is shown that the number of LP iterations will be significantly reduced through proper scaling. Scaling also helps to reduce the hat matrix diagonal entries corresponding to leverage measurements. The simulation results indicate that the effects of scaling on the leverage points and on the CPU times are highly dependent on the measurement configurations. Optimal scaling strategies valid for a variety of measurement configurations are currently under investigation.

ACKNOWLEDGEMENTS

References

[1] H. M. Merrill and F. C. Schweppe, "Bad Data Suppression in Power System Static Estimation", *IEEE Trans. on Power Apparatus and Systems*, Vol. PAS-90, pp. 2718-2725, Nov./Dec. 1971.

[2] M. R. Irving, R. C. Owen and M. J. H. Sterling, "Power System State Estimation Using Linear Programming", *Proceedings of IEE*, Vol. 125, pp. 879-885, 1978.

[3] W. W. Kotiuga and M. Vidyasagar, "Bad Data Rejection Properties of Weighted Least Absolute Value Techniques Applied to Static State Estimation", *IEEE Trans. on Power Apparatus and Systems*, Vol. PAS-101, No. 4, pp. 844-851, April 1982.

[4] D. M. Falcão and S. M. de Assis, "Linear Programming State Estimation: Error Analysis and Gross Error Identification", IEEE PES 1987 Summer Meeting, San Francisco, Paper No. 462-5.

[5] P. J. Rousseeuw and A. M. Leroy, "Robust Regression and Outlier Detection", John Wiley & Sons, 1987.

[6] L. Mili, V. Phaniraj and P. Rousseeuw, "Least Median of Squares Estimation in Power Systems", *IEEE Trans. on Power Systems*, Vol. PES-6, pp. 511-523, May 1991.

[7] M. K. Çelik and A. Abur, "A Robust WLAV State Estimator Using Transformations", IEEE PES 1991 Winter Meeting, New York, Paper No. 236-0.

[8] K. A. Clements, P. W. Davis and K. D. Frey, "An Efficient Algorithm for Computing the Weighted Least Absolute Value Estimate in Power System Static State Estimation",

[9] A. Abur and M. K. Çelik, "A Fast Algorithm for the Weighted Least Absolute Value State Estimation", *IEEE Trans. on Power Systems*, Vol. PES-6, pp. 1-8, February 1991.

[10] K. A. Clements, P. W. Davis and K. D. Frey, "An Interior Point Algorithm for Weighted Least Absolute Value Power System State Estimation", IEEE PES 1991 Winter Meeting, New York, Paper No. 235-2.

[11] J. A. Tomlin, "On Scaling Linear Programming Problems", Mathematical Programming Study 4, pp. 146-166, North Holland Publishing Co., 1975.

[12] M. Benichou, J. M. Gauthier, G. Hentges, G. Ribiere, "The Efficient Solution of Large-Scale Linear Programming Problems", *Mathematical Programming*, Vol. 13, pp. 280-322, 1977.

[13] F. R. Hampel, E. M. Ronchetti, P. J. Rousseeuw and W. A. Stahel, "Robust Statistics: The Approach Based on Influence Functions", John Wiley & Sons, 1986.

[14] D. C. Hoaglin and R. E. Welsch, "The Hat Matrix in Regression and ANOVA", *The American Statistician*, Vol. 32, No. 32, February 1978.

[15] I. Barrodale and F. D. K. Roberts, "An Improved Algorithm for Discrete ℓ_1 Linear Approximation ", *SIAM Journal Numer. Anal.*, Vol. 10, No. 5, pp. 839-848, October 1973.

[16] J. J. Allemong, L. Radu and A. M. Sasson, "A Fast and Reliable State Estimation Algorithm for AEP's New Control Center". *IEEE Trans. on Power Apparatus and Systems*, Vol. PAS-101, pp. 933-944, April 1982.

[17] J. W. Gu, K. A. Clements, G. R. Krumpholz and P. W. Davis, "The Solution of Ill-Conditioned Power System State Estimation Problems via the Method of Peters and Wilkinson", PICA Conference Proceedings, pp. 239-246, 1983.

[18] H. Muller-Merbach, "On Round-Off Errors in Linear Programming", Springer-Verlag, 1970.

Mehmet Kemal Çelik (S '90) received his B.S. degree from Boğaziçi University, İstanbul, Turkey in 1986 and his M.S. degree from Virginia Polytechnic Institute and State University in 1988. He has been studying for his Ph.D. degree in electrical engineering at Texas A&M University since Fall 1988.

Ali Abur (M '86, SM '90) received the B.S. degree from METU, Turkey in 1979, the M.S. and Ph.D. degrees from The Ohio State University, Columbus, Ohio, in 1981 and 1985 respectively. Since late 1985, he has been an Assistant Professor at Texas A&M University, College Station, Texas. His research interests are in computer applications to the power system problems.

Proceedings of the IFAC Int. Symp. on Power Systems and Power Plant Control, Aug.22-25, 1989, Seoul, Korea, pp. 785-790.

Discussion

H. Singh and Y. Hu (University of Wisconsin-Madison) and **A.A. El-Keib** (The University of Alabama, Tuscaloosa, AL) The authors are to be commended for their paper which presents the benefits of scaling LP problems in power system state estimation. Scaling has long been used for improving the solution of linear programs [11] and also in power system state estimation [A1,A2]. However, its benefit for dealing with leverage points in power system state estimation is an idea that has been used here by the authors.

The reduction in LP iterations by proper scaling is attributed to the fact that relative costs of non-basic variables provide a better indication of their contribution in minimizing the cost function. This is true for scale independent pivoting criteria such as the 'most negative relative cost' criteria used by the authors. However, there are alternative pivoting criteria which are independent of scale, such as the one based on 'largest decrease in the objective function.' In general, the latter criterion results in a lower number of LP iterations. However, in view of the computational overhead involved, it is seldom used. How would this criterion compare with the one used by the authors, if the overhead involved in scaling was accounted for, particularly in case of optimal scaling methods [11]? The effect of adjustment of normalization values on $K'_{ii}s$ involves some computational overhead. Is this reflected in the CPU times given in the paper?

Also, the effective rejection of bad data by the WLAV state estimator is reflected by large non-zero values of associated slack variables resulting in a large objective function value. A lower objective function value would indicate failure to reject bad data (the associated slack variable being zero). This seems to indicate that the usual pivoting criteria used in linear programming may not be the best suited for dealing with leverage points. Could the authors comment on this?

The measurement redundancies used by the authors in Table 6 for the reactive sub-problem are rather low, particularly for group B. These would be even lower for the active sub-problem. The ability of the state estimator to reject bad data would suffer in this case.

How do the authors explain the fact that even though the number of LP iterations increases with scaling for group B, the CPU time decreases? For example, in case of the 1512 bus system the total LP iterations given in Table 9 are 1069 and 1142 before and after scaling respectively.

Figure 4 illustrates the cost functions of the four algorithms converging to the same optimal solution. It should be clarified that this will not be the case in general. If leverage points are present in the measurements, the corresponding residuals before and after scaling would be different (as illustrated in the paper), resulting in different objective function values.

Finally, it is not obvious that the same scaling strategy would work best for the two stated objectives of the paper. Did the authors observe any tradeoff between the two?

[A1] A.A. El-Keib and H. Singh, "Fast Linear Programmming State Estimation using the Dual Formulation," *Paper No.91WM276-6-PWRS IEEE Winter Power Meeting, Feb.3-7, 1991, New York.*

[A2] A.A. El-Keib, J. Nieplocha, H. Singh and D. Maratukulam, "A Decomposed State Estimation Technique Suitable for Parallel Processor Implementation," *Paper No.91SM464-8-PWRS IEEE Summer Power Meeting, July 28-Aug.2, 1991, San Diego.*

Lamine Mili and Nitin S. Vichare (Virginia Polytechnic Institute and State University, Blacksburg, VA): The paper advocates the scaling of the Jacobian matrix to improve the numerical stability and the overall performance of the simplex algorithm. This claim is supported by many simulation results performed on several test systems. To alleviate ill-conditioning, the statistical literature (e.g. Belsley *et al.* [A1], p. 183) advocates to make the columns of the Jacobian matrix unit length instead of dividing the column elements by the largest entry. Could the author elaborate on this point?

Concerning the effect of scaling on leverage points, we agree with the authors that row normalization may decrease the number of leverage points by bringing them toward the bulk of the point cloud. However, extensive simulations performed on three IEEE test systems, namely the 14–, 30–, and 118–bus systems, have shown that, in general, scaling does not eliminate leverage points. Since the diagonal elements of the hat matrix $\underline{K}$ may fail to reveal multiple leverage points grouped in clusters (see Ref. [6]), we have based leverage diagnostic on a very robust algorithm that implements the Minimum Volume Ellipsoid (MVE) estimator [A2]. The estimates are robust distances d_i given by

$$d_i = \sqrt{(\underline{\ell}_i - \underline{\ell})^T \underline{C}^{-1} (\underline{\ell}_i - \underline{\ell})} ,$$

where $\underline{\ell}$ is a vector containing the coordinates of the center of the ellipsoid with minimum volume that covers $[\frac{m + n + 1}{2}]$ data points; these points constitute the bulk of the point cloud in the factor space. Here $\underline{C}$ denotes the covariance matrix associated with MVE and $\underline{\ell}^T_i$ is the ith row of the Jacobian matrix. The leverage points are those points having large values (say larger than 3) in the standardized distances

$$d_{Si} = (d_i - \underset{i}{median}(d_i))/ MAD ,$$

with

$$MAD = \underset{j}{median} |d_j - \underset{i}{median}(d_i)| .$$

Table A1. Leverage points for the IEEE 30–bus system.

Before scaling		After Scaling	
Leverage Point	Standard. Distances	Leverage Point	Standard. Distances
FLP 6–28	3.35	FLP 28–8	3.09
FLP 28–6	3.35	INP 17	4.55
INP 17	3.75	FLP 10–20	6.21
FLP 24–22	4.03	INP 9	13.83
FLP 10–22	5.31		
INP 28	5.48		
FLP 6–8	5.93		
FLP 4–6	7.74		
FLP 6–4	7.74		
FLP 10–21	10.49		
INP 6	29.37		
INP 22	84.18		

The MVE method is able to identify a large fraction of leverage points, be them isolated or in clusters. Table A1 shows leverage diagnostic performed on the IEEE 30–bus system. This system has all the real power injections (INP) set at values less than 0.3 p.u. except those at buses 5, 17 and 18 which are set about 1 p.u.; the reactive power injections are negligible. Before scaling, MVE identifies 12 leverage points out of 56 real power measurements; they show up with standardized distances larger than 3. Note that all these leverage points are associated with short lines except for INP 17 which have a relatively high loading. This indicates that leverage points may also be induced by large power flows and injections. In fact, simulations showed that the points in the factor space are not stationary but move as the loading

conditions evolve. After scaling, the picture changes drastically as new leverage points replace old ones. As shown in the last two columns of Table A1, all but INP 17 have been replaced by 3 new leverage points. Similar results have been found in the IEEE 118–bus system. For example at the flat voltage profile, 44 leverage points out of 247 real power measurements have been identified. After scaling, 4 previously identified leverage points are still present along with 23 new leverage points.

ACKNOWLEDGEMENT

REFERENCES

[A1] D. A. Belsley, E. Kuh, and R. E. Welsch, *Regression Diagnostics: Identifying Influential Data and Sources of Collinearity*, John Wiley, 1980.

[A2] L. Mili, N. S. Vichare, V. Phaniraj, and P. J. Rousseeuw, "Leverage Point Identification in Power System State Estimation", to be submitted.

[A3] P. J. Rousseeuw, and A. M. Leroy, *Robust Regression and Outlier Detection*, John Wiley, 1987.

M.K. Çelik and A. Abur We thank the discussers for their questions and comments on our paper. We will respond to their discussions in the order they are received:

To: L.Mili and N.S. Vichare:

The authors investigated the heuristic scaling methods in [11] in an effort to evaluate the normalization, arithmetic and geometric mean methods for the case of WLAV estimation of the power system states. In addition, various combinations of these methods, namely arithmetic and geometric scaling methods followed by normalization, were also studied to find out their effect on leverage points and the number of LP iterations per SE iteration. The only difference between the methods presented in [11] and the ones the authors studied is that column scaling is carried out prior to row scaling. The merit of these methods is their simplicity compared to the rotational transformations proposed in [7] and the fact that the original sparsity of the measurement Jacobian is preserved without any new fill-ins which the method of [7] suffered from. The computational overhead for these methods were found to be insignificant. Moreover, the normalization method was found to be the most effective in reducing the number of LP iterations and the leverage points. The method of making the columns of the Jacobian matrix of unit length prior to row scaling is one of the many scaling methods suggested for Linear Programming and deserves investigation along with the optimal scaling strategies in [11].

The diagonal entries of the hat matrix are not used in the method of the paper and are reported just to show the effect of the scaling method studied on the leverage points.

The authors agree with the fact that new leverage points may appear due to the scaling method used in the paper; however, in all the simulations carried out with different measurement configurations, the total number of leverage points consistently decreased. The results reported by the discussers with the leverage point diagnostic tool they have used, seem to agree with our experience.

As stated by the discussers, the leverage points will move as the operating conditions evolve, since the structure of the measurement Jacobian will change. In this study, the decoupled state estimation formulation is used and therefore the measurement Jacobian is assumed to be constant at its flat start value. Thus, the scaling procedure is carried out only once at the outset. However, the same scaling strategy can be repeatedly applied to modify the updated Jacobian at each state estimation iteration, if the coupled formulation is used. Since the scaling overhead is negligible, this is not expected to significantly impact the overall computational burden.

To: H. Singh, Y. Hu and A.A.El-Keib:

As stated above, the profile of the run times revealed that the computational burden of the scaling method used in this paper was insignificant compared to the overall CPU times. The steepest-edge algorithm where the nonbasic columns are normalized using the ℓ_2 norm, is computationally so expensive that the CPU time gains are not comparable to the significant reduction in the number of LP iterations achieved. However, this method is not scaling independent as suggested by the discussers, because the column-row scaling method of the paper modifies the Jacobian by reducing the dispersion of the magnitude of the elements of the original model. The K_{ii}'s are not used in the algorithm of the paper. Therefore, their computation are not reported in the CPU times given.

The WLAV estimator fails to identify and reject bad data if there exists an alternative solution which yields a lower objective function by rejecting some of the "good" data and accepting the bad measurements. In such a case, the accepted bad measurements are referred to as leverage points. Therefore, the Simplex pivoting logic is deceived by the leverage points. In the paper, the aim is to modify the jacobian's structure in order to completely eliminate the leverage points, thus allowing the Simplex algorithm to converge to the true solution in the absence of the leverage points.

As given in Table 6, the number of the voltage measurements for the measurement sets in group B is very small. Thus, the redundancies for the active problem are almost the same as the ones for the reactive problem. Besides, keeping the number of line flow measurements high while keeping the others low for the simulations presented in this section was done on purpose to study the effect of scaling method used on different types of measurement configurations. The reasons why the CPU times decrease for group B are given in the last part of section VI.

The cost functions in Figure 4 are given to show that the four algorithms converge to the same SE solution in the absence of bad data. The cost functions that are plotted, are calculated as $c_0 = \omega^T (z - h(\hat{x}))$ and represent the weighted sum of the absolute residuals at the converged SE solution. In case of bad data, depending on which n out of m measurements are chosen as perfect by the LP solution, there will be small discrepancies between the objective functions of the WLAV and WLAV-S solutions. This is due to the fact that leverage points are always present in the power system measurement sets. However, when there are multiple interacting bad data corresponding to leverage points, then there may also be discrepancies between the objective functions c_0 of the WLAV and WLAV-S estimators, because as shown in the paper, WLAV estimator may fail to reject bad data in the leverage points in such cases.

Please see our response to Mili and Vichare above, regarding the issue of scaling strategies.