Journal of Computational Mathematics, Vol.10, No.4, 1992, 291–295.

# A SHORT NOTE ON AN $L_1$-NORM MINIMIZATION ALGORITHM*

Yang Zi-qiang Y. Yuan
(*Computing Center, Academia Sinica, Beijing, China*)

**Abstract**

In this short note, examples are constructed to show that a recent algorithm given by Soliman, Christensen and Rouhi[1] may give a non-optimal solution.

In [1], a linear least absolute value (LAV) estimate algorithm is presented. The linearly constrained LAV problem has the following form.

$$\min_{\theta \in \Re^n} \|H\theta - z\|_1, \tag{1}$$

$$C\theta = d \tag{2}$$

where $H \in \Re^{m\times n}$, $z \in \Re^m$, $C \in \Re^{l\times n}$ and $d \in \Re^l$. One of the algorithms given in [1] is for solving problem (1)-(2). The algorithm can be restated as follows:

**Algorithm 1**[1]. **Step 1.** Calculate

$$\theta^* = \begin{bmatrix} H \\ C \end{bmatrix}^+ \begin{pmatrix} z \\ d \end{pmatrix}, \tag{3}$$

where $B^+$ is the Moore-Penrose generalized inverse of $B$.

**Step 2.** Compute

$$r^* = \begin{pmatrix} z \\ d \end{pmatrix} - \begin{bmatrix} H \\ C \end{bmatrix} \theta^*, \tag{4}$$

$$\bar{r} = \frac{1}{m+l} \sum_{i=1}^{m+l} r_i^*, \tag{5}$$

$$\sigma = \sqrt{\frac{1}{m-1} \sum_{i=1}^{m} (r_i^* - \bar{r})^2}. \tag{6}$$

**Step 3.** Let $J = \{j \mid |r_j^*| \le \sigma, 1 \le j \le m\}$ and

$$P_J = \sum_{j\in J} e_j e_j^T \tag{7}$$

* Received March 11, 1992.

where $e_j$ $(j = 1, \cdots, m)$ are unit vectors in $\Re^m$.

Compute the new least squares solution

$$\theta^*_{\text{new}} = \begin{bmatrix} P_J H \\ C \end{bmatrix}^+ \begin{pmatrix} P_J z \\ d \end{pmatrix}, \tag{8}$$

$$r^*_{\text{new}} = z - H\theta^*_{\text{new}}. \tag{9}$$

**Step 4.** Let $I = \{i_1, \cdots, i_{n-l}\}$ be a subset of $\{1, \cdots, m\}$ which corresponds to the $n - l$ smallest residuals. Let $P_I = \sum_{i \in I} e_i e_i^T$ and solve

$$\begin{bmatrix} P_I H \\ C \end{bmatrix} \theta = \begin{pmatrix} P_I z \\ d \end{pmatrix} \tag{10}$$

to get $\bar{\theta}$. Accept $\bar{\theta}$ as a solution.

It should be noted that definition (6) is not the usual definition for standard deviation. We use (6) because it is the definition, as we understand, used by [1]. However, our examples are also valid if the usual definition of standard deviation is used. Another point that is worth mentioning is that $r^*_{\text{new}}$ denotes first $m$ residuals of the whole system, though $\theta^*_{\text{new}}$ is the least squares solution of a reduced system.

Soliman et al. [1] also extended the above algorithm to solving nonlinear LAV problems. For more details, see [1]. Now we give a linear LAV problem for which a non-optimal solution would be given by the above algorithm.

**Example 1.** Solve problem (1)–(2) with the following data:

$$H = \begin{bmatrix} 1 & 1 \\ 1 & 2 \\ 1 & 3 \\ 1 & 4 \\ 1 & 5 \\ \varepsilon & 0 \end{bmatrix}, \quad z = \begin{pmatrix} 2 \\ 2 \\ 3 \\ 4 \\ 0 \\ 0 \end{pmatrix}, \tag{11}$$

$$C = (1 \quad 6), \quad d = (5), \tag{12}$$

where $\varepsilon \in (0, 1)$ is a very small parameter.

Our example is very similar to Example 2.1 of [1]. We have added a very small row in the example, expecting that the corresponding residual will eventually be the smallest. The original $z_5 = 3$ (as in [1]) is changed to 0 to guarantee that the fifth residual will be the only measure to be deleted. It should be noted that, unlike Example 2.1 of [1], the above example can not be viewed as a straight line data fitting problem because $\varepsilon \neq 1$. However, we can still analyze the above algorithm for problem (1)–(2) with data given by (11)–(12).

It is easy to calculate

$$\theta^* = \frac{1}{105} \begin{pmatrix} 175 \\ 30 \end{pmatrix} + O(\varepsilon^2), \tag{13}$$

which gives

$$r^* = \frac{1}{105}\begin{pmatrix} 5 \\ -25 \\ 50 \\ 125 \\ -325 \\ -175\varepsilon \\ 170 \end{pmatrix} + O(\varepsilon^2), \tag{14}$$

$$\sigma = \sqrt{\frac{1}{5}\sum_{i=1}^{6}(r_i^*)^2} = \frac{\sqrt{24880}}{105} + O(\varepsilon^2) \approx \frac{157.734}{105} + O(\varepsilon^2). \tag{15}$$

Hence, because $\varepsilon << 1$, only the fifth measure should be deleted. Therefore the algorithm will compute the least squares solution of

$$\begin{bmatrix} 1 & 1 \\ 1 & 2 \\ 1 & 3 \\ 1 & 4 \\ \varepsilon & 0 \\ 1 & 6 \end{bmatrix}\theta = \begin{pmatrix} 2 \\ 2 \\ 3 \\ 4 \\ 0 \\ 5 \end{pmatrix}. \tag{16}$$

Direct calculations give

$$\theta^*_{\text{new}} = \frac{1}{74}\begin{pmatrix} 80 \\ 49 \end{pmatrix} + O(\varepsilon^2). \tag{17}$$

Consequently, we have

$$r^*_{\text{new}} = \frac{1}{74}\begin{pmatrix} 19 \\ -30 \\ -5 \\ 20 \\ -325 \\ -80\varepsilon \end{pmatrix} + O(\varepsilon^2). \tag{18}$$

Again, because $\varepsilon << 1$, the residual $80\varepsilon/74$ is the smallest. Thus, the final linear system should be

$$\begin{bmatrix} \varepsilon & 0 \\ 1 & 6 \end{bmatrix}\theta = \begin{pmatrix} 0 \\ 5 \end{pmatrix}, \tag{19}$$

which gives the point

$$\bar{\theta} = \begin{pmatrix} 0 \\ 5/6 \end{pmatrix}. \tag{20}$$

It is not difficult to show that the optimal solution is

$$\hat{\theta} = \begin{pmatrix} 1 \\ 2/3 \end{pmatrix}. \tag{21}$$

Therefore we have shown that the above algorithm may yield a non-optimal solution.

Our technique for constructing the above example is to introduce such a small row, that it will have the least residual, and the algorithm will take this measure as an active measure. Consequently a non-optimal point would be computed. Our next example shows that even for straight line $L_1$ data fitting problems the algorithm given above may also give a non-optimal solution.

**Example 2.** Fit the data points $\{(2,1),(3,2),(4,3),(6,6)\}$ with a straight line of the form $z(x) = a_1 + a_2 x$ such that $z(0) = 1$.

For Example 2, we have

$$H = \begin{bmatrix} 1 & 2 \\ 1 & 3 \\ 1 & 4 \\ 1 & 6 \end{bmatrix}, \quad z = \begin{pmatrix} 1 \\ 2 \\ 3 \\ 6 \end{pmatrix}, \tag{22}$$

$$C = (1 \quad 0), \quad d = (1). \tag{23}$$

From (22)–(23) and (3), it follows that

$$\theta^* = \frac{1}{100} \begin{pmatrix} 5 \\ 85 \end{pmatrix}, \tag{24}$$

which gives

$$r^* = \frac{1}{100} \begin{pmatrix} -75 \\ -60 \\ -45 \\ 85 \\ 95 \end{pmatrix}, \tag{25}$$

$$\sigma = \sqrt{\frac{1}{3}\sum_{i=1}^{4}(r_i^*)^2} = \frac{1}{100}\sqrt{18475/3} \approx 0.78475. \tag{26}$$

Therefore the fourth measure should be deleted, and we compute the least squares solution of the following reduced system:

$$\begin{bmatrix} 1 & 2 \\ 1 & 3 \\ 1 & 4 \\ 1 & 0 \end{bmatrix} \theta = \begin{pmatrix} 1 \\ 2 \\ 3 \\ 1 \end{pmatrix}. \tag{27}$$

It is

$$\theta^*_{\text{new}} = \frac{1}{35} \begin{pmatrix} 23 \\ 17 \end{pmatrix}. \tag{28}$$

And we have

$$r^*_{\text{new}} = \frac{1}{35}\begin{pmatrix} -22 \\ -4 \\ 14 \\ 85 \end{pmatrix}. \tag{29}$$

Now it is quite clear that the second residual is the smallest, and from the algorithm, we should solve the linear system

$$\begin{bmatrix} 1 & 3 \\ 1 & 0 \end{bmatrix}\theta = \begin{pmatrix} 2 \\ 1 \end{pmatrix}, \tag{30}$$

which has a unique solution

$$\bar{\theta} = \begin{pmatrix} 1 \\ 1/3 \end{pmatrix}. \tag{31}$$

But the optimal solution for Example 2 is

$$\hat{\theta} = \begin{pmatrix} 1 \\ 1/2 \end{pmatrix}, \tag{32}$$

and it is easy to verify that

$$||H\bar{\theta} - z||_1 = \frac{13}{3} > \frac{7}{2} = ||H\hat{\theta} - z||_1. \tag{33}$$

Thus, again the algorithm gives a non-optimal solution.

We carried out our research in August 1991. Recently we were informed by Professor D. Naeve, Co-editor of *Computational Statistics and Data Analysis,* that a similar result was obtained by Bassett and Keonker in April 1991. The result of Basset and Keonker (1991) was submitted to *Journal of Computational Statistics and Data Analysis.*

## References

[1] S.A. Soliman, G.S. Christensen and A.H. Rouhi, A new algorithm for nonlinear $L_1$-norm minimization with nonlinear equality constraints, *Computational Stat. and Data Analysis,* 11 (1991), 97–109.

[2] G.W. Bassett and R.W. Koenker, A note on recent proposals for computing $L_1$ estimates, Report, University of Illinois at Chicago, USA, 1991.