A Dual Algorithm to Solve Linear Least Absolute Value Problems
Author(s): Ronald D. Armstrong and Mabel Tam Kung
Source: *The Journal of the Operational Research Society*, Vol. 33, No. 10 (Oct., 1982), pp. 931-936

Published by: Palgrave Macmillan Journals on behalf of the Operational Research Society
Stable URL: http://www.jstor.org/stable/2580999
Accessed: 30-10-2015 12:05 UTC

J. Opl Res. Soc. Vol. 33, pp. 931 to 936, 1982
Printed in Great Britain.

0160-5682/82/100931-06$03.00/0

A Dual Algorithm to Solve Linear Least Absolute Value Problems

RONALD D. ARMSTRONG and MABEL TAM KUNG

Department of Quantitative Business Analysis, College of Business,
University of Georgia and Department of Management Science,
California State University at Fullerton

This paper presents a special purpose dual linear programming algorithm to solve a linear least absolute value multiple linear regression problem. Various computer implementations of the fundamental algorithm are discussed and compared with existing special-purpose procedures to solve the least absolute value problem. Computational results with four implementations of the algorithm are given.

INTRODUCTION

THE LEAST ABSOLUTE VALUE (LAV) criterion has been widely considered as an alternative method to least squares in fitting a linear model. LAV estimation yields the unknown parameters to minimize the sum of the absolute deviations of a given set of observations from the values predicted by the model. The two major contributing factors to the popularity of LAV regression are the sensitivity of least squares to outliers and the inappropriateness of a Gaussian error assumption. LAV is particularly well suited to problems where the error distribution is fat-tailed (e.g. Laplace or Cauchy distributed). A survey of LAV algorithmic procedures, statistical properties and empirical tests are found in Armstrong *et al.*,[1] Charnes *et al.*,[2] Robers and Ben-Israel,[3] Dielman and Pfafferberger,[4] Spyropoulos *et al.*[5] and Glover.[6]

The problem examined here can be stated as follows: Given a set of n observational measurements $(y_i, x_{i1}, x_{i2}, \ldots, x_{im})$, $i = 1, 2, \ldots, n$, determine the value for $\beta = (\beta_1, \beta_2, \ldots, \beta_m)^T$ which will

$$\underset{\beta}{\text{Mininize}} \sum_{i=1}^{n} |y_i - x_{i1}\beta_1 - x_{i2}\beta_2 - \ldots x_{im}\beta_m|. \tag{1}$$

The difficulty of solving (1) for many years contributed to less than occasional use of LAV estimation. It was not until the implementation of the linear programming algorithms on the digital computer that LAV estimates could be obtained for problems of reasonable size. Charnes *et al.*[2] appear to be first to have demonstrated that linear least absolute value problems can be rewritten as linear programming problems. Employing their result here, problem (1) is equivalent to:

$$\text{Minimize} \sum_{i=1}^{n} (P_i + N_i)$$

subject to

$$\sum_{j=1}^{m} x_{ij}\beta_j + P_i - N_i = y_i, \quad i = 1, 2, \ldots, n, \tag{2}$$

$$P_i \geqslant 0, \; N_i \geqslant 0, \qquad i = 1, 2, \ldots, n,$$

where P_i and N_i are, respectively, the positive and negative deviation associated with the ith observation.

Barrodale and Young,[7] Usow,[8] Robers and Ben-Israel,[3] Abdelmalek,[9] Spyropoulos *et al.*,[5] Barrodale and Roberts[10] and Armstrong *et al.*[1] present special purpose primal

algorithms to solve (2). Dual solution methods for a more general class of weighted LAV problems have been discussed in Robers and Ben-Israel[3] and Glover.[6] These methods utilize compact basis procedures analogous to those presented in this paper. The algorithm given here is a special purpose dual algorithm to solve (1). Variations of the dual algorithm which include multiple pivots in phase 1 or a start with a dual feasible solution are discussed.

ALGORITHM

The dual problem of (2) is:

$$\text{Maximize} \sum_{i=1}^{n} \pi_i y_i$$

subject to

$$\sum_{i=1}^{n} \pi_i x_{ij} = 0, \quad j = 1, 2, \ldots, m, \tag{3}$$

$$-1 \leqslant \pi_i \leqslant 1, \quad i = 1, 2, \ldots, n.$$

Assume a basis matrix B of dimension m by m has been identified. Rank deficiencies can easily be handled within the linear programming framework (see Ben-Israel and Charnes[11]), and a rank of m for the observation matrix X will be assumed. Define IB to be the index set of the basic variables, and the index sets NL and NU to be indicators for the non-basic variables which are respectively at their lower and upper bounds. Define π_B to be the vector of the basic variables.

A possible method for obtaining an initial solution is given as follows. All non-basic variables are set to their upper bound value, namely $+1$. The values of the basic variables are:

$$\bar{\pi}_B = -B^{-1} H_B^T$$

where

$$H_{B_j}^T = \sum_{i \in NU} \pi_i x_{ij}, \quad j = 1, 2, \ldots, m.$$

If $\bar{\pi}_B$ satisfies $-1 \leqslant \pi_i \leqslant +1$, $i \in IB$, the current basis B is feasible and the algorithm will proceed directly to phase 2 of the simplex method. Otherwise, $\bar{\pi}_B$ is infeasible and a phase 1 procedure is required to produce a feasible solution.

Phase 1

Define $\tilde{c}_B$ to be the basic cost vector in the phase 1 process. The values of $\tilde{c}_B$ are determined as follows:

$$\tilde{c}_{B_j} = \begin{cases} 0 & \text{if } -1 \leqslant \bar{\pi}_{B_j} \leqslant +1, \\ -\text{sign}\,(\bar{\pi}_{B_j}) & \text{otherwise.} \end{cases}$$

The termination criterion for this process is that all values of $\tilde{c}_B$ are equal to zero.

If the termination criterion is not satisfied, the algorithm chooses a non-basic variable, π_s, $s \in (NL \cup NU)$ to enter the basis. To accomplish this, the reduced costs of the non-basic variables, $k \in (NL \cup NU)$ are calculated. The reduced costs are:

$$z_k = \tilde{c}_B B^{-1} X_k, \; k \in (NL \cup {}^{NU}),$$

where X_k is the kth row of the observational matrix, X, of dimension n by m.

The candidates for the entering variable satisfy the following relations:

$$\begin{aligned} z_k &< 0 \text{ for } k \in NL \\ z_k &> 0 \text{ for } k \in NU. \end{aligned} \tag{4}$$

The procedure for choosing the vector to enter the basis consists of selecting the maximum of the absolute values of z_k satisfying (4). This will not, in general, give the largest improvement in the objective function value, but does give the fastest change in the objective per unit change of the incoming variable.

If the non-basic variable π_s is considered to be brought into the basis, the algorithm then calculates the amount of change required by the entering variable to force the feasibility of the leaving variable. The value of the change, θ, is obtained from finding the minimum of the following:

$$\theta = \min \begin{cases} 2; \\ \dfrac{1 - \rho\psi_j \bar{\pi}_{B_j}}{\xi_j} \quad \text{for } \xi_j \neq 0,\ \tilde{c}_{B_j} = 0,\ j = 1, 2, \ldots, m; \\ \\ \dfrac{\bar{\pi}_{B_j} + \rho\psi_j}{\xi_j} \quad \text{for } \rho\psi_j > 0,\ \tilde{c}_{B_j} \neq 0,\ j = 1, 2, \ldots, m, \end{cases} \tag{5}$$

where

$$\xi_j = B_j^{-1}(x_{s1}, x_{s2}, \ldots, x_{sm})^T, \quad j = 1, 2, \ldots, m;$$

$$\psi_j = \text{sign}(\xi_j), \quad j = 1, 2, \ldots, m;$$

and

$$\rho = \text{sign}\,(z_s).$$

If $\theta = 2$, π_s will remain non-basic but will switch to its opposite bound value. Furthermore, the values of the basic variables $\bar{\pi}_{B_j}$, $j = 1, 2, \ldots, m$ will be updated as follows:

$$\bar{\pi}_{B_j} \leftarrow \bar{\pi}_{B_j} + 2\rho\xi_j,\ j = 1, 2, \ldots, m.$$

If $\theta \neq 2$ and the minimum ratio value comes from the rth basic variable, the value of z_s will be decreased by ξ_r. If this updated value, $\hat{z}$, where $\hat{z} = z_s - \xi_r$, remains positive, no pivoting is performed. Rather, the value of $\tilde{c}_{B_r}$ will equal zero. The algorithm then recalculates θ from (5) with $\tilde{c}_{B_r} = 0$ and evaluates the basic variable to be considered to leave the basis. On the other hand, if $\hat{z}$ is negative, π_s enters the basis at the rth position and the values of the basic variables $\bar{\pi}_{B_j}$, $j = 1, 2, \ldots, m$ become:

$$\bar{\pi}_{B_j} \leftarrow \bar{\pi}_{B_j} + \rho\theta\xi_j \text{ for all } j \neq r$$

$$\bar{\pi}_{B_r} \leftarrow \pi_s\,(1 - \theta)$$

In this manner several standard pivots may be combined into one.

After the updating process is completed, the algorithm checks the feasibility level of the basic variables and continues with the above iterative procedure until $-1 \leqslant \pi_i \leqslant +1$, $i \in IB$, is satisfied. At this time the algorithm proceeds to phase 2.

Phase 2

The optimality conditions are characterized by the following:

$$\begin{aligned} z_k - y_k &\geqslant 0 \text{ for } k \in NL \\ z_k - y_k &\leqslant 0 \text{ for } k \in NU, \end{aligned} \tag{6}$$

where $z_k - y_k = Y_B B^{-1} X_k - y_k$, Y_B is the basic cost vector.

If the conditions in (6) are not satisfied, the algorithm then determines the most violating reduced cost. If $z_s - y_s$ yields the maximum violation, this means that the non-basic variable π_s is considered to enter the basic.

The algorithm then finds the basic variable to be examined by calculating the minimum of the following:

$$\theta = \min\left\{2;\ \frac{1 - \rho\psi_j\bar{\pi}_{B_j}}{\xi_j}\right\} \text{ for } \xi_j \neq 0,\ j = 1, 2, \ldots, m.$$

If $\hat{\theta} = 2$, π_s will remain as a non-basic variable but will switch to its opposite bound value. If $\hat{\theta} \neq 2$, and the minimum ratio corresponds to the rth basic variable, π_s will enter the basis to replace π_{B_r}, and the pivoting procedure of the simplex method will be carried out. In any event, the values of the basic variables will be updated:

$$\bar{\pi}_{B_j} \leftarrow \bar{\pi}_{B_j} + \rho\hat{\theta}\xi_j \text{ for all } j \neq r$$

$$\bar{\pi}_{B_r} \leftarrow \bar{\pi}_{B_r} + 2\rho\xi_r \text{ when } \hat{\theta} = 2$$

$$\bar{\pi}_{B_r} \leftarrow \pi_s(1 - \hat{\theta}) \text{ when } \hat{\theta} \neq 2.$$

The algorithm repeats the above procedure until the conditions of (6) are satisfied.

Dual feasible start

It is possible to begin the dual algorithm with a dual feasible solution and by-pass phase 1. It may be noted that setting $\pi_i = 0$, $i = 1, 2, \ldots, n$ provides a feasible solution to (3). It is, however, not an extreme point solution. The algorithm maintains dual feasibility and moves to an extreme point solution based on reduced costs. The initial basis consists of artificial variables at a zero level, and these variables are assigned an upper and lower bound of zero. Each π_i is sequentially chosen as a candidate to enter the basis. If $z_i > 0$, then π_i is a candidate to increase in value, and if $z_i \leqslant 0$, then π_i is a candidate to decrease in value. Since each π_i has an initial value of zero, the minimum ratio test must be modified to the following:

$$\theta = \min\left\{1;\ \frac{U_j - \rho\psi_j\pi_{B_j}}{\xi_j}\ \xi_j \neq 0,\ j = 1, 2, \ldots, m\right\},$$

where $U_j = 0$ if an artificial variable is in the jth basic position and $U_j = 1$ otherwise. Thus, all artificial variables will be removed from the basis if X has full column rank.

After all π_i have either moved to a bound or are in the basis, the algorithm proceeds to the previously described phase 2. Thus, although a dual feasible solution is always maintained, additional labour is required before beginning the normal simplex algorithm.

ALGORITHMIC VARIATIONS AND COMPUTATIONAL EXPERIENCE

The dual algorithms presented here were coded in FORTRAN IV and tested on the CYBER 750 computer at the University of Georgia at Athens. The codes were used to solve randomly generated problems and problems from the literature. In order to present results in a tabular form and study the effect of parameter changes, only results from a set of randomly generated problems are given. The assumed model was $X\beta = Y + \epsilon$, where the error term, ϵ, had a doubly exponential error distribution with a mean of zero and a standard deviation of 20. The X and β values were generated from uniform distributions and Y set equal to $X\beta + \epsilon$. Routines from IMSL were used for all random number generation.

The five programs utilized in the testing are now defined.

D1. This version utilized the described dual algorithm with the following modifications:

(a) A candidate list of size K was used to select the incoming L.P. variable in phase 2; i.e. the K variables with the largest reduced costs were selected and

TABLE 1. COMPUTATIONAL RESULTS FOR 5 LAV CODES

m	n	$D1$	$D2$	$D3$	$D4$	P
5	100	0.49	0.59	1.21	0.40	0.20
		(144)	(158)	(96)	(90)	(13)
10	200	4.5	3.8	4.8	3.4	1.3
		(430)	(396)	(260)	(310)	(33)
15	300	14.6	13.2	16.7	14.4	4.4
		(742)	(670)	(444)	(590)	(55)
15	400	23.6	21.1	28.3	21.1	5.6
		(944)	(910)	(605)	(760)	(56)
20	500	57.8	52.7	68.0	56.2	15.1
		(1596)	(1455)	(1048)	(1126)	(95)

The upper value in each cell is CPU time in seconds and the lower value is number of basis updates. All values are means from runs of three randomly generated problems

only these variables were considered to enter the basis until either the list became empty or no eligible candidates existed in the list. Through testing, an effective value of K was chosen to be 5.

(b) The multiple pivot technique described for phase 1 was not used.

D2. This version maintained the candidate list and utilized the multiple pivot in phase 1.

D3. This version did not use a candidate list but did include the multiple pivot technique described for phase 1.

D4. This version started with a dual feasible solution, as described in the previous section.

P. This is the primal code of Armstrong *et al.*[1]

The computational results are summarized in Table 1. The multiple pivot in phase 1 had a significant effect on phase 1 times and an overall reduction of approximately 10%. The candidate list increased the number of basis updates but reduced solution time by approximately 10%. Little difference is seen between versions D2 and D4, but the primal code was always significantly faster. These results were consistent with other testing not reported here.

All algorithms and comparisons discussed thus far have been for solving problems where an approximate solution is not readily available. There are situations where several variations in a model must be inspected and an advanced starting solution can be obtained. One such case arises when finding the best parameter subset of size q taken from a full model with m parameters. In the algorithm to solve this (see Barrodale and Young[7]), a basic dual feasible solution is available for each problem with fewer than m parameters. The dual algorithm (without phase 1) was able to out-perform the primal algorithm by a timing factor of 20–30% in this situation.

CONCLUSION

This paper has presented implementations of specialized dual methods to solve a linear least absolute value problem. Various basis entry and initialization procedures were considered. The most efficient implementations utilized an initial dual feasible start or a phase 1 with multiple pivots. Nevertheless, the specialized dual method remains somewhat slower than a specialized primal algorithm.[1] The advantage of the dual approach comes when a 'good' dual feasible solution is readily available. The dual approach has been shown[1,2] to be superior to the primal approach in this situation.

REFERENCES

[1] R. D. ARMSTRONG, E. L. FROME and D. S. KUNG (1979) A revised simplex algorithm for the absolute deviation curve-fitting problem. *Commun. Statist.* **B8**(2), 175–190.

[2] A. Charnes, W. W. Cooper and R. O. Ferguson (1955) Optimal estimation of executive compensation by linear programming. *Mgmt Sci.* **1,** 138–151.
[3] P. D. Robers and A. Ben-Israel (1969) An interval programming algorithm for discrete linear L_1 approximation problems. *J. Approx. Theory* **2,** 323–336.
[4] T. Dielman and R. Pfafferberger (1982) LAV (Least Absolute Value) estimation in linear regression: a review. *TIMS Stud. Mgmt Sci.*, Optimization in Statistics Volume (in press).
[5] K. Spyropoulos, E. Kiountouzis and A. Young (1973) Discrete approximations in the L_1 norm. *Comput. J.* **16,** 180–186.
[6] F. Glover (1978) Linear programming for weighted deviation problems using compact basis techniques. *Eur. J. Opns Res.* **2,** 360–367.
[7] I. Barrodale and A. Young (1968) Algorithms for best L_1 and L_∞ linear approximation on a discrete set. *Opns Res.* **16,** 295–306.
[8] K. H. Usow (1967) On L_1 approximation II: computation for discrete functions and discretization effects. *SIAM J. Numer. Analysis* **4,** 233–244.
[9] N. N. Abdelmalek (1971) Linear L_1 approximation for a discrete point set and L_1 solutions for overdetermined linear equations. *J. ACM* **18,** 41–47.
[10] I. Barrodale and F. D. K. Roberts (1973) An improved algorithm for discrete L_1 linear approximation. *SIAM J. Numer. Analysis* **10,** 839–848.
[11] A. Ben-Israel and A. Charnes (1968) An explicit solution of a special class of linear programming problems. *Opns Res.* **16,** 1166–1175.
[12] R. Armstrong and M. Kung (1982) An algorithm to select the best subset for a Least Absolute Value problem. *TIMS Stud. Mgmt Sci.*, Optimization in Statistics Volume (in press).