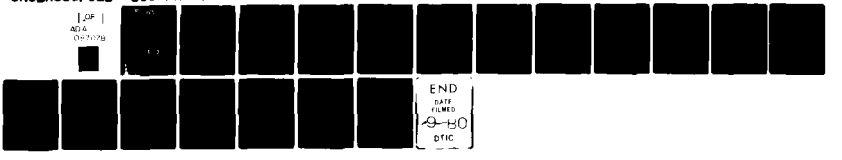
AD-A087 078 TEXAS UNIV AT AUSTIN CENTER FOR CYBERNETIC STUDIES F/G 12/1
A DUAL ALGORITHM TO SOLVE LINEAR LEAST ABSOLUTE VALUE APPROXIMA--ETC(U)
JUN 80 R ARMSTRONG, M KUNG N00014-75-C-0569
UNCLASSIFIED CCS-RR-370 NL



END
DATE FILMED
9-80
DTIC

ADA087078

LEVEL II (12)

CENTER FOR CYBERNETIC STUDIES

The University of Texas
Austin, Texas 78712

DTIC
SELECTE
JUL 24 1980
S D
C

This document has been approved for public release and sale; its distribution is unlimited.

80 7 23 126

(12

Research Report CCS 370

A DUAL ALGORITHM TO SOLVE LINEAR LEAST ABSOLUTE VALUE APPROXIMATIONS

by

R. Armstrong
M. Kung*

DTIC
ELECTE
JUL 24 1980
S D
C

June 1980

* McMaster University, Hamilton, Ontario, Canada

This research was partly supported by Project NR047-021, ONR Contract N00014-75-C-0569 and Project NR047-071, ONR Contract N00014-80-C-0242 with the Center for Cybernetic Studies, The University of Texas at Austin. Reproduction in whole or in part is permitted for any purpose of the United States Government.

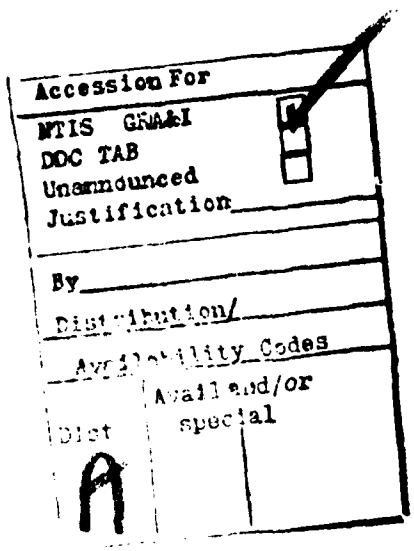
CENTER FOR CYBERNETIC STUDIES

A. Charnes, Director
Business-Economics Building, 203E
The University of Texas at Austin
Austin, TX 78712
(512) 471-1821

This document has been approved for public release and sale; its distribution is unlimited.

Abstract

This paper presents a special purpose dual linear programming algorithm to solve linear least absolute value problems. In addition, strategies involving start procedures are examined. Implementations of computer-based techniques are discussed. Computational results with three computer code versions of the algorithm are given.



1. Introduction

The least absolute value criterion has been considered as an alternative method to least squares in fitting a linear model. Least absolute value estimation yields the unknown parameters in a stochastic model as to minimize the sum of the absolute deviations of a given set of observations from the values predicted by the model.

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

$$\underset{\beta}{\text{minimize}} \sum_{i=1}^{n} |Y_i - x_{i1}\beta_1 - x_{i2}\beta_2 - \ldots x_{im}\beta_m|. \qquad (1)$$

It has been noted by Glahe and Hunt [8] that the least absolute value criterion has, in particular, interested econometricians who frequently estimate parameters of linear models with relatively small number of observations. Another advantage of the least absolute value estimator is the resistance to outliers in the data or to heavy-tailed error distributions (see, for example, Rice and White [10].) A major difference between the least squares and the least absolute value estimate is the fact that the least squares criterion always produces an unique optimum in the objective function while the least absolute value criterion at times can have multiple optimal solutions. Since most of the data used in real world applications of curve fitting problems are far from accurate, like in econometric and business research works, a method that brings out the nonuniqueness of the estimates might

at times be more preferable. However, an important handicap in the least absolute valuable criterion is that the estimates are governed by a restricted set of observations, those that are inside the convex hull of all observations.

Recently, there has been a steadily increasing interest in least absolute value estimates and their properties due to statisticians' awareness of the limitations of least squares analysis, as well as the development of computationally efficient algorithms for providing the least absolute value estimates. Charnes, Cooper and Ferguson [6] 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:

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

subject to

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

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

where P_i and N_i are, respectively, the positive and negative deviation associated with the i-th observation.

Barrodale and Young [4], Usow [14], Robers and Ben-Israel [11], Abdelmalek [1], Schlossmacher [12], Spyropoulos, Kiountouzis and Young [13], Barrodale and Roberts [3], and Armstrong, Frome and Kung [2] present special purpose primal algorithms to solve (2). The algorithm

given here is a special purpose dual algorithm to solve this problem.

2. Algorithm

The dual problem of (2) is:

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

subject to

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

$$\pi_1 \leq 1, \quad 1 = 1, 2, \ldots, n,$$

$$\pi_1 \geq -1, \quad i = 1, 2, \ldots, n.$$

Assume the basis matrix, B of dimension m by m, to be of full rank. Rank deficiencies can easily be handled within the linear programming framework (see Ben-Israel and Charnes [5]). Define IB to be the index set of the basic variables, and the index sets NL and NU to be indicators for the nonbasic variables which are respectively at their lower and upper bounds. Define π_B to be the vector of the basic variables.

The initial basic feasible solution are given as follows. All nonbasic 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 $-1 \leq \pi_i \leq +1$, $i \in IB$, 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.

2.1 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:

$$\tilde{c}_{B_j} = \begin{cases} 0 & \text{if } -1 \leq \bar{\pi}_{B_j} \leq +1,\ j \in IB \\ -\text{sign}\,(\bar{\pi}_{B_j}) & \text{otherwise,} \quad j \in IB \end{cases}$$

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 then determines a nonbasic variable, π_s, $s \in (NL \cup NU)$ to enter the basis. To accomplish this, the algorithm calculates the reduced costs of the nonbasic variables, z_k, $k \in (NL \cup NU)$ as it is done in the standard linear programming technique. The reduced costs are:

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

where X_k is the k-th row of the observational matrix, $\mathbf{X}$, of dimension n by m.

The candidates for the entering variable satisfy the following

relations:

$$z_k < 0 \quad \text{for } k \in NL$$
$$z_k > 0 \quad \text{for } k \in NU. \tag{4}$$

The procedure for choosing the vector to enter the basis consists of selecting the maximum of the absolute values of z_k. 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 nonbasic 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:

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

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

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

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

If $\theta = 2$, π_s will remain as nonbasic 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:

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

If $\theta \neq 2$ and the minimum ratio value comes from the r-th 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 calculates 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, the values of the basic variables $\bar{\pi}_{B_j}$, $j = 1, 2, \ldots, m$ become:

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

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

The algorithm then performs the pivoting procedure of the simplex method in which the nonbasic enters the basis and the basic variable π_{B_r} leaves the basis.

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

2.2 Phase 2

The optimal conditions are characterized by the following:

$$\begin{aligned} z_k - Y_k &\geq 0 \quad \text{for } k \in NL \\ z_k - Y_k &\leq 0 \quad \text{for } k \in NU \end{aligned} \tag{6}$$

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 by finding the maximum of the absolute values of the reduced costs. If $|z_s - Y_s|$ is the maximum, this means that the nonbasic variable π_s is considered to enter the basis.

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

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

If $\hat{\theta} = 2$, π_s will remain as a nonbasic variable but will switch to its opposite bound value. If $\hat{\theta} \neq 2$, and the minimum ratio corresponds to the r-th 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:

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

$$\bar{\pi}_{B_r} \leftarrow \begin{cases} \bar{\pi}_{B_r} + 2\rho\xi_r & \text{for } \hat{\theta} = 2 \\ \pi_s (1 - \hat{\theta}) & \text{for } \hat{\theta} \neq 2 \end{cases}$$

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

3. Computational Experience

The dual algorithm was coded in FORTRAN IV and tested on the CDC 6600 computer at The University of Texas at Austin. The observations were drawn from various uniform and normal distributions using a random number generator. Three versions, called PROGRAM I, PROGRAM II, and PROGRAM III were implemented.

PROGRAM I is the version of the algorithm where initially all the nonbasic variables are set to their upper bound value, namely, +1. All the computations follow the scheme of the algorithm as described earlier.

PROGRAM II initially sets the values of the nonbasic variables based on the signs of the reduced costs. The process is given:

$$\pi_i = \text{sign}\ (Y_i - Y_B\ B^{-1}\ X_i), \quad i \in NB.$$

Additionally, a chaining technique is implemented to choose the entering variable. Instead of selecting the maximum of the absolute value of the reduced costs, this technique considers the first eligible candidate, say π_s, which satisfies the conditions in (4) to become basic. Note that in this example the first chain starts from the first position and ends in the s-th position of the list of the reduced costs. The pivoting process of the algorithm will then take place. Two situations may arise:

(i) If $-1 \leq \pi_i \leq +1$, $i \in IB$, is satisfied, the algorithm will either proceed to phase 2 if the process is currently in

the phase 1 procedure or terminate if it is in the phase 2 procedure.

(ii) However, if $-1 \leq \pi_i \leq +1$, $i \in IB$, is not satisfied, the procedure continues to search for the next eligible nonbasic variable by utilizing a pointer to indicate where the previous chain ends and evaluates the reduced costs based on the new basis starting from the $(s + 1)$th position until a nonbasic variable which satisfies the conditions in (4) is found. This nonbasic variable will then enter the basis and the process continues until all the basic variables are feasible.

An advantage of this chaining technique is that the time-consuming sorting process of the maximum of the absolute reduced costs is eliminated. However, in some situations when a large number of chains has to be developed, much access time is consumed.

PROGRAM III is a version of PROGRAM II where a candidate list replaces the chaining process in the phase 2 procedure. The phase 1 procedure still maintains the chaining technique. The length of the candidate list can be easily assigned by the user. Suppose the conditions in (6) are not satisfied and the length of the candidate list is assigned to be 5, the candidates in this list will be stored in terms of the indices of the first five nonbasic variables which satisfy the conditions in (4). Moreover, the candidates are sorted in descending order based on the absolute values of their reduced costs. The process will then choose the first candidate from the list to become basic. The pivoting process of the algorithm will

be carried out. If the conditions in (6) are satisfied, the phase 2 procedure will be terminated and the current solution will be optimal. On the other hand, if the conditions in (6) are violated, the process examines the next candidate on the list by calculating its reduced cost from the new basis. If the reduced cost satisfies (4), this candidate will be brought into the basis. Otherwise, the process proceeds to examine the remaining candidates on the list. In the event when all the candidates on the list have been examined, a new list with length of five will be formulated. This process continues until the conditions in (6) are satisfied.

The candidate list structure provides an alternative to the chaining technique described earlier.

The computational results given in Table 1 are mean times and iteration counts which include the number of basis changes for a set of 10 problems with the same characteristics.

Number of Parameters	Number of Observations	PROGRAM I	PROGRAM II	PROGRAM III
10	50	223 78	158 59	154 86 L=5
10	100	408 132	331 94	595 248 L=5
10	200	1514 297	1225 215	924 381 L=4
20	200	3084 334	2969 269	2702 569 L=4
30	200	7648 435	5275 315	6655 854 L=6

The upper number in each row is the mean time in CPU milliseconds; the lower number is the mean number of iterations required; and L is the length of the candidate list.

Table 1. Computational Testings of PROGRAM I, PROGRAM II and PROGRAM III.

4. Conclusion

This paper presents a dual algorithm to solve linear least absolute value approximations. The algorithm utilizes the concept of linear programming methodology. Strategies involving start procedures are examined. The

techniques of chaining and candidate list structure are implemented. From the computational testings, these techniques enhance the efficiency of the algorithm in terms of computer time.

REFERENCES

1. Abdelmalek, N.N., "Linear L_1 Approximation for a Discrete Point Set and L_1 Solutions for Overdetermined Linear Equations", Journal of ACM, Vol. 18, 1971, pp. 41-47.

2. Armstrong, R.D., Frome, E.L. and Kung, D.S., "A Revised Simplex Algorithm for the Absolute Deviation Curve-fitting Problem", Communications in Statistics, Vol. B8(2), 1979, pp. 175-190.

3. Barrodale, I. and Roberts, F.D.K., "An Improved Algorithm for Discrete L_1 Linear Approximation", SIAM Journal on Numerical Analysis, Vol. 10, 1973, pp. 839-848.

4. Barrodale, I. and Young, A., "Algorithms for Best L_1 and L_∞ Linear Approximation on a Discrete Set", Numerical Mathematics, Vol. 8, 1966, pp. 295-306.

5. Ben-Israel, A. and Charnes, A., "An Explicit Solution of a Special Class of Linear Programming Problems", Operations Research, Vol. 16, 1968, pp. 1166-1175.

6. Charnes, A., Cooper, W.W. and Ferguson, R.O., "Optimal Estimation of Executive Compensation by Linear Programming", Management Science, Vol. 1, 1955, pp. 138-151.

7. Charnes, A. and Cooper, W.W., Management Models and Industrial Applications of Linear Programming, Volumes I and II, John Wiley & Sons, New York, 1961.

8. Glahe, F.R. and Hunt, J.G., "The Small Sample Properties of Simultaneous Equation Least Absolute Estimators Vis-a-vis Least Squares Estimators", Econometrica, Vol. 38, 1970, pp. 742-753.

9. Martin, J., Computer Data-Base Organization, Englewood Cliffs, Prentice-Hall, New Jersey, 1977.

10. Rice, J.R. and White, J.S., "Norms for Smoothing and Estimation", SIAM Review, Vol. 6, 1964, pp. 243-256.

11. Robers, P.D. and Ben-Israel, A., "An Interval Programming Algorithm for Discrete Linear L_1 Approximation Problems", Journal of Approximation Theory, Vol. 2, 1969, pp. 323-336.

12. Schlossmacher, E.J., "An Iterative Technique for Absolute Deviations Curve Fitting", Journal of the American Statistical Association, Vol. 68, 1973, pp. 857-859.

13. Spyropoulos, K., Kiountouzis, E. and Young, A., "Discrete Approximations in the L_1 Norm", The Computer Journal, Vol. 16, 1973, pp. 180-186.

14. Usow, K.H., "On L_1 Approximation II: Computation for Discrete Functions and Discretization Effects", SIAM Journal of Numerical Analysis, Vol. 4, 1967, pp. 233-244.

Unclassified
Security Classification

AD-A087078

DOCUMENT CONTROL DATA - R & D

(Security classification of title, body of abstract and indexing annotation must be entered when the overall report is classified)

1. ORIGINATING ACTIVITY (Corporate author)

Center for Cybernetic Studies
The University of Texas at Austin

2a. REPORT SECURITY CLASSIFICATION

Unclassified

2b. GROUP

3. REPORT TITLE

A Dual Algorithm to Solve Linear Least Absolute Value Approximations.

4. DESCRIPTIVE NOTES (Type of report and inclusive dates)

(9) Research rept.,

5. AUTHOR(S) (First name, middle initial, last name)

Ronald Armstrong [redacted] Mabel Kung

6. REPORT DATE

Jun[redacted]80

7a. TOTAL NO. OF PAGES

16

7b. NO. OF REFS

14

8a. CONTRACT OR GRANT NO.

N00014-75-C-0569, N00014-80-C-0242

9a. ORIGINATOR'S REPORT NUMBER(S)

CCS-370

b. PROJECT NO.

NR047-021 NR074-071

9b. OTHER REPORT NO(S) (Any other numbers that may be assigned this report)

(14) CCS-RR-370

d.

10. DISTRIBUTION STATEMENT

This document has been approved for public release and sale; its distribution is unlimited.

11. SUPPLEMENTARY NOTES

(12) 17

12. SPONSORING MILITARY ACTIVITY

Office of Naval Research (Code 434)
Washington, DC

13. ABSTRACT

This paper presents a special purpose dual linear programming algorithm to solve linear least absolute value problems. In addition, strategies involving start procedures are examined. Implementations of computer-based techniques are discussed. Computational results with three computer code versions of the algorithm are given.

DD FORM 1 NOV 68 1473 (PAGE 1)

S/N 0101-807-6811

406197

Unclassified
Security Classification

A-31408

14 KEY WORDS	LINK A		LINK B		LINK C	
	ROLE	WT	ROLE	WT	ROLE	WT
Regression Analysis						
Least Absolute Value Problems						
Linear Programming						
Goal Programming						
List Structures						