The Korean Communications
in Statistics Vol. 7, No. 1, 2000
pp. 47-54

Algorithm for the Constrained Chebyshev Estimation in Linear Regression[1]

Bu-yong Kim[2]

Abstract

This article is concerned with the algorithm for the Chebyshev estimation with/without linear equality and/or inequality constraints. The algorithm employs a linear scaling transformation scheme to reduce the computational burden which is induced when the data set is quite large. The convergence of the proposed algorithm is proved. And the updating and orthogonal decomposition techniques are considered to improve the computational efficiency and numerical stability.

1. Introduction

The Chebyshev estimator, which is also called the L_∞-estimator or the least maximum absolute deviation estimator, is considered as an alternative to the least squares estimator. It is attractive for short-tailed underlying distribution cases since it is a maximum likelihood estimator when the distribution of error is uniform. Rice and White(1964) and Appa and Smith(1973) introduced the properties of the Chebyshev estimator.

In this article we consider the unconstrained and constrained Chebyshev estimation problem in the multiple linear regression. The constrained problem includes the linear equality and inequality constraints as follows,

$$\boldsymbol{y} = X\boldsymbol{\beta} + \boldsymbol{\varepsilon}, \quad \boldsymbol{l} \le G\boldsymbol{\beta} \le \boldsymbol{u}, \tag{1.1}$$

where $\boldsymbol{y}$ denotes an n-vector of response variable, X represents a full-rank $n\times p$ matrix of regressor variables including an intercept term, $\boldsymbol{\beta}$ is a p-vector of regression parameters, $\boldsymbol{\varepsilon}$ is an n-vector of random errors, G is a $g\times p$ constraint matrix, $\boldsymbol{l}$ is a g-vector of lower bound, and $\boldsymbol{u}$ is a g-vector of upper bound.

The following are special cases to the problem (1.1). (i) if $l_i = -\infty$ and $u_i = \infty$ for all i, it is the ordinary unconstrained estimation, (ii) if $l_i = -\infty$ and $u_i \neq \infty$ for some i, it has one-sided constraints from above, (iii) if $l_i \neq -\infty$ and $u_i = \infty$ for some i, it has one-sided constraints from below, (iv) if $l_i = 0$ and $u_i = \infty$ for some i, it has non-negative constraints, (v) if $l_i = u_i$ for some i, it has equality constraints.

The Chebyshev estimator does not have a closed-form solution, so the optimization techniques such as

1) This paper was supported in part by Sookmyung Women's University, 1999.
2) Department of Statistics, Sookmyung Women's University, Seoul 140-742, Korea.

linear programming method has to be employed. Sklar and Armstrong
(1983, 1984) proposed algorithms for the simple linear regression model, while Barrodale and Phillips(1975) and Abdelmalek(1977) suggested algorithms for the multiple model. Since those algorithms are based on the simplex method, large amount of computation is required when the data set has many regressors or observations. This article proposes an algorithm which employs the linear scaling transformation scheme to reduce the amount of computation. Also the convergence of the algorithm is proved. Furthermore, the updating and orthogonal decomposition techniques are considered to improve the computational efficiency and numerical stability of the proposed algorithm.

2. Proposed Algorithm

It is necessary to take the linear programming approach in order to deal with the linear equality and inequality constraints. The linear programming problem for the Chebyshev estimation with linear constraints is formulated as follows

$$\underset{\xi\in\Omega_\xi}{\text{minimize}}\ \boldsymbol{c}'\xi,\quad \Omega_\xi = \{\ A'\xi \ge \boldsymbol{a} :\ \xi \text{ unrestricted}\ \}, \tag{2.1}$$

where

$$\boldsymbol{c}=\begin{bmatrix} \mathbf{0}_p \\ 1 \end{bmatrix},\quad \xi=\begin{bmatrix} \widetilde{\boldsymbol{\beta}} \\ \lambda \end{bmatrix},\quad A=\begin{bmatrix} X' & -X' & G' & -G' \\ \boldsymbol{\ell}' & \boldsymbol{\ell}' & \mathbf{0}_g' & \mathbf{0}_g' \end{bmatrix},\quad \boldsymbol{a}=\begin{bmatrix} \boldsymbol{y} \\ -\boldsymbol{y} \\ \boldsymbol{l} \\ -\boldsymbol{u} \end{bmatrix},$$

and $\mathbf{0}_p=(0,\cdots,0)'\in R^p$, $\mathbf{0}_g=(0,\cdots,0)'\in R^g$, $\boldsymbol{\ell}=(1,\cdots,1)'\in R^n$, $\widetilde{\boldsymbol{\beta}}$ is a p-vector of the estimate, and $\lambda(\ge 0)$ denotes the value of the maximum absolute residual that is to be minimized.

The problem (2.1) may be readily solved by any variants of simplex method. However, since the constraint matrix is of large dimension especially when the number of regressors is large, it requires a great deal of computation to obtain solution by the simplex-type algorithms. In order to deal with this computational inefficiency problem, a linear scaling transformation scheme can be employed at each iteration. Advantages of this scheme has been shown by Sherali, Skarpness, and Kim(1988). To adapt the scheme, the dual problem corresponding to the primal problem is constructed as a canonical form,

$$\underset{\zeta\in\Omega_\zeta}{\text{maximize}}\ \boldsymbol{a}'\zeta,\quad \Omega_\zeta = \{\ \zeta \ge \mathbf{0} :\ A\zeta = \boldsymbol{c}\ \},$$

where $\zeta\in R^{2(n+g)}$ denotes the dual variable. The polytope Ω_ζ is called the feasible region. If Ω_ζ is nonempty, then the problem is feasible. This scheme starts with an initial feasible solution, which in practice can be obtained by adding one artificial variable(γ) and assigning big m to the corresponding cost vector of the new variable as follows,

$$\underset{\gamma\in\Omega_\gamma}{\text{maximize}}\ \boldsymbol{b}'\gamma,\quad \Omega_\gamma = \{\ \gamma \ge \mathbf{0} :\ B\gamma = \boldsymbol{c}\ \}, \tag{2.2}$$

where $\boldsymbol{b}=\begin{bmatrix} \boldsymbol{a} \\ m \end{bmatrix}$, $\boldsymbol{\gamma}=\begin{bmatrix} \zeta \\ \gamma \end{bmatrix}$, and $B=[A \,\vdots\, \boldsymbol{c}-A\boldsymbol{\ell}\,]$. Since the value m vanishes at the optimum if the original problem is feasible, a solution of the enlarged problem is also a solution of the original problem. Clearly $2(n+g)+1$ dimensional vector of all ones can be the initial feasible solution.

The algorithm starts with a given feasible solution $\boldsymbol{\gamma}=(\gamma_1,\cdots,\gamma_m)'>\mathbf{0}$, $m=2(n+g)+1$. At every iteration, it employs the linear scaling transformation which yields the following change of variables

$$\boldsymbol{\gamma}=D\boldsymbol{\tau}, \quad D=\text{diag}\{\ \gamma_1,\cdots,\gamma_m\}.$$

Under this transformation, the problem (2.2) is reformulated in the new $\boldsymbol{\tau}$ coordinates,

$$\underset{\tau\in\Omega_\tau}{\text{maximize}}\ \boldsymbol{b}'D\boldsymbol{\tau}, \quad \Omega_\tau=\{\ \boldsymbol{\tau}\geq\mathbf{0}:\ BD\boldsymbol{\tau}=\boldsymbol{c}\}. \tag{2.3}$$

In this transformed space, the feasible solution $\boldsymbol{\gamma}_{\langle 0\rangle}$ is mapped to $\boldsymbol{\tau}_{\langle 0\rangle}=\boldsymbol{\ell}$, and the projection $\boldsymbol{p}$ of the gradient of the objective function in terms of $\boldsymbol{\tau}$ coordinates onto the null space of BD can be described as

$$\boldsymbol{p}=\{\ I-DB'(BD^2B')^{-1}BD\}\,D\boldsymbol{b}. \tag{2.4}$$

Since X is assumed to be of full column rank, B is of full row rank. Hence BD^2B' is nonsingular since $\boldsymbol{\gamma}>\mathbf{0}$. A step length η is taken from the current iterate $\boldsymbol{\tau}=D^{-1}\boldsymbol{\gamma}=\boldsymbol{\ell}$ along the projected gradient direction $\boldsymbol{p}$ to achieve the maximum increase of the objective function. Thereafter, the $\boldsymbol{\tau}$ solution is mapped back into the $\boldsymbol{\gamma}$ space via $\boldsymbol{\gamma}=D\boldsymbol{\tau}$ and $\boldsymbol{d}=D\boldsymbol{p}$, resulting in a new $\boldsymbol{\gamma}$ solution with an improved objective function value. This amounts to taking a step along $\boldsymbol{d}$, yielding the new iterate such as

$$\boldsymbol{\gamma}_{\text{new}}=\boldsymbol{\gamma}_{\text{old}}+\eta\boldsymbol{d}. \tag{2.5}$$

The step length η should be chosen so that the feasibility of new point $\boldsymbol{\gamma}_{\text{new}}$ is maintained,

$$\eta=\delta x, \quad 1/x=\underset{i=1,\ldots,m}{\text{maximum}}\ \{-d_i/\gamma_i\}>0, \quad 0<\delta<1. \tag{2.6}$$

The feasibility of $\boldsymbol{\gamma}_{\text{new}}$ is proved by Lemma 1. The algorithm consists mainly of generating a sequence of points $\boldsymbol{\gamma}_{\langle 0\rangle}, \boldsymbol{\gamma}_{\langle 1\rangle},\cdots,\boldsymbol{\gamma}_{\langle k\rangle}$.

It is proved by Theorem 1 that if $\boldsymbol{p}=\mathbf{0}$ for some $\boldsymbol{\gamma}>\mathbf{0}$, then any feasible solution is optimal, and if $\boldsymbol{p}\geq\mathbf{0}$ and $\boldsymbol{p}\neq\mathbf{0}$ for some $\boldsymbol{\gamma}$, then the problem is unbounded. Note that the algorithm terminates since the latter condition does not occur under the specified assumptions on (1.1). So the algorithm continues until the following termination criterion is satisfied.

$$\| \boldsymbol{p} \|_\infty < \omega, \text{ for small enough } \omega > 0,$$

where $\| \cdot \|_\infty$ denotes the L_∞-norm. It can be easily verified by Theorem 2 that this criterion works well.

It follows from (2.4) that $D\boldsymbol{b}$ is in the orthogonal complement of the null space of BD if $\boldsymbol{p} = \mathbf{0}$. There exists a vector $\boldsymbol{\xi}$ such that

$$\boldsymbol{\xi}' BD = \boldsymbol{b}' D \tag{2.7}$$

since the orthogonal complement of the null space of a matrix is the row space of that matrix. Furthermore, $\boldsymbol{\xi}$ is the vector of primal variables corresponding to the constraint $BD\boldsymbol{\tau} = \boldsymbol{c}$ of the problem (2.3). And the scaling leaves the primal with respect to the problem (2.2) unchanged. Since BD^2B' is nonsingular from the assumption, (2.7) can be rewritten as

$$\boldsymbol{\xi} = (BD^2B')^{-1} BD^2 \boldsymbol{b}.$$

Therefore, current estimate $\widetilde{\boldsymbol{\beta}}$ consists of the first p entries of the vector $\boldsymbol{\xi}$. The steps of proposed algorithm are described in detail as follows.

[Algorithm : CHEBCON]

Initialization : Set the iteration counter $k=0$, and let $\boldsymbol{\gamma}_{\langle 0 \rangle} = (1, \dots, 1)'$ be the initial feasible solution. Choose $\delta \in [0.97,\ 0.99]$.

Step 1 : Given $\boldsymbol{\gamma}_{\langle k \rangle}$, define $D_{\langle k \rangle} = \text{diag}\{ \gamma_{\langle k \rangle_1}, \dots, \gamma_{\langle k \rangle_m} \}$. Compute the projected gradient $\boldsymbol{p}_{\langle k \rangle}$, and the direction of motion $\boldsymbol{d}_{\langle k \rangle}$,

$$\boldsymbol{p}_{\langle k \rangle} = \{ I - D_{\langle k \rangle} B' (BD^2_{\langle k \rangle} B')^{-1} BD_{\langle k \rangle} \} D_{\langle k \rangle} \boldsymbol{b}, \quad \boldsymbol{d}_{\langle k \rangle} = D_{\langle k \rangle} \boldsymbol{p}_{\langle k \rangle}.$$

Step 2 : If $\| \boldsymbol{p}_{\langle k \rangle} \|_\infty < \omega$ for some chosen tolerance $\omega > 0$, then go to Step 5.

Step 3 : Determine the step length $\eta_{\langle k \rangle}$; $\eta_{\langle k \rangle} = \delta x$, $1/x = \underset{i=1,\cdots,m}{\text{maximum}} \{ -d_{\langle k \rangle i} / \gamma_{\langle k \rangle i} \} > 0.$

Step 4 : Set the new iterate; $\boldsymbol{\gamma}_{\langle k+1 \rangle} = \boldsymbol{\gamma}_{\langle k \rangle} + \eta_{\langle k \rangle} \boldsymbol{d}_{\langle k \rangle}$. Increment k by one and return to Step 1.

Step 5 : Compute the primal solution; $\boldsymbol{\xi} = (BD^2_{\langle k \rangle} B')^{-1} BD^2_{\langle k \rangle} \boldsymbol{b}$, pick the first p entries of the vector $\boldsymbol{\xi}$ as the Chebyshev estimate $\widetilde{\boldsymbol{\beta}}$, and stop.

3. Convergence of the Proposed Algorithm

The convergence of the algorithm is proved under an additional assumption that the problem is bounded and feasible. Noting (2.4) and (2.5), we see that in order to prove the convergence of the algorithm, we need to verify that $\{ \boldsymbol{p}_{\langle k \rangle} \} \to \mathbf{0}$, and that any convergent subsequence indexed by K, $\{ \boldsymbol{\gamma}_{\langle k \rangle} \}_K \to \boldsymbol{\gamma}^*$ generated by the algorithm satisfies $\boldsymbol{\gamma}^* > \mathbf{0}$.

[Lemma 1] It is satisfied that $\boldsymbol{\gamma}_{\langle k \rangle} \in \Omega_\gamma$, $\Omega_\gamma = \{ \boldsymbol{\gamma} \geq \mathbf{0} : B\boldsymbol{\gamma} = \boldsymbol{c} \}$ in the steps of the algorithm.

[Proof] The initial feasible solution is $\boldsymbol{\gamma}_{\langle 0 \rangle} = (1, \ldots, 1)' \in \Omega_\gamma$. Suppose that $\boldsymbol{\gamma}_{\langle k \rangle} \in \Omega_\gamma$, then

$$\begin{aligned} B\boldsymbol{\gamma}_{\langle k+1 \rangle} &= B(\boldsymbol{\gamma}_{\langle k \rangle} + \eta_{\langle k \rangle} D_{\langle k \rangle} \boldsymbol{p}_{\langle k \rangle}) \\ &= B\boldsymbol{\gamma}_{\langle k \rangle} + \eta_{\langle k \rangle} \{ BD^2_{\langle k \rangle} \boldsymbol{b} - BD^2_{\langle k \rangle} B' (BD^2_{\langle k \rangle} B')^{-1} BD^2_{\langle k \rangle} \boldsymbol{b} \} \\ &= \boldsymbol{c}. \end{aligned}$$

Also it is clear from (2.6) that $\boldsymbol{\gamma}_{\langle k+1 \rangle} > \mathbf{0}$. Those imply that $\boldsymbol{\gamma}_{\langle k+1 \rangle} \in \Omega_\gamma$, hence each iterate $\boldsymbol{\gamma}_{\langle k \rangle}$ is the feasible solution.

[Lemma 2] The duality holds between the primal and dual problems.

[Proof] Since, from Lemma 1, the iterate $\boldsymbol{\gamma}_{\langle k \rangle}$ is feasible, the relationship

$$\bar{\boldsymbol{\xi}}' A \bar{\boldsymbol{\zeta}} = \bar{\boldsymbol{\xi}}' B \bar{\boldsymbol{\gamma}}_{\langle k \rangle} = \bar{\boldsymbol{\xi}}' \boldsymbol{c}$$

holds for any primal and dual feasible solutions $\bar{\boldsymbol{\xi}}$, $\bar{\boldsymbol{\zeta}}$, and $\bar{\boldsymbol{\gamma}}_{\langle k \rangle}$ respectively. Also, it follows from the constraints of the primal that

$$\boldsymbol{c}' \bar{\boldsymbol{\xi}} = \bar{\boldsymbol{\zeta}}' A' \bar{\boldsymbol{\xi}} \geq \bar{\boldsymbol{\zeta}}' \boldsymbol{a}.$$

Therefore, the weak duality holds. On the other hand, let $\boldsymbol{\xi}^*$ and $\boldsymbol{\gamma}^*$ be the primal and dual optimal solutions, respectively. Then

$$\begin{aligned} \boldsymbol{b}' \boldsymbol{\gamma}^* &= \boldsymbol{b}' D_{\langle k \rangle} \boldsymbol{\ell} \\ &= \boldsymbol{\xi}^{*\prime} B D_{\langle k \rangle} \boldsymbol{\ell} \\ &= \boldsymbol{c}' \boldsymbol{\xi}^* \end{aligned}$$

from (2.7). Thus, the strong duality holds.

[Theorem 1] If $\boldsymbol{p}_{\langle k \rangle} = \mathbf{0}$, then $\boldsymbol{\gamma}_{\langle k \rangle}$ is optimal. Otherwise, assume that $\boldsymbol{p}_{\langle k \rangle} \neq \mathbf{0}$ for all k. Then the sequence $\{ \boldsymbol{b}' \boldsymbol{\gamma}_{\langle k \rangle} \}$ converges.

[Proof] Let $\boldsymbol{p}_{\langle k\rangle}=\mathbf{0}$, then $\xi' B = \boldsymbol{b}'$ from (2.7). Suppose $\boldsymbol{z}$ is any feasible point. Then $\boldsymbol{b}'\boldsymbol{z} = \xi' B\boldsymbol{z} = \xi'\boldsymbol{c}$. It is clear that the objective function $\boldsymbol{b}'\boldsymbol{z}$ is constant on Ω_γ since $\xi'\boldsymbol{c}$ does not depend on $\boldsymbol{z}$, and hence $\boldsymbol{\gamma}_{\langle k\rangle}$ is optimal. However, if $\boldsymbol{p}_{\langle k\rangle}\neq\mathbf{0}$, then we need to show that the sequence $\{\boldsymbol{b}'\boldsymbol{\gamma}_{\langle k\rangle}\}$ is strictly increasing. We can rewrite the definition of $\boldsymbol{d}$ as follows

$$\boldsymbol{d} = D^2\boldsymbol{b} - D^2B'(BD^2B')^{-1}BD^2\boldsymbol{b}$$

and

$$\boldsymbol{b} - D^{-2}\boldsymbol{d} = B'(BD^2B')^{-1}BD^2\boldsymbol{b}. \tag{3.1}$$

From the feasibility in Lemma 1, we can construct the following equation,

$$\boldsymbol{b}'D^2B'(B'D^2B)^{-1}B(\boldsymbol{\gamma}_{\langle k+1\rangle} - \boldsymbol{\gamma}_{\langle k\rangle}) = 0 \tag{3.2}$$

It follows from (3.1) and (3.2) that

$$(\boldsymbol{b} - D^{-2}\boldsymbol{d})'(\boldsymbol{\gamma}_{\langle k+1\rangle} - \boldsymbol{\gamma}_{\langle k\rangle}) = 0.$$

Therefore,

$$\begin{aligned}\Delta_{\langle k\rangle} \equiv \boldsymbol{b}'(\boldsymbol{\gamma}_{\langle k+1\rangle} - \boldsymbol{\gamma}_{\langle k\rangle}) &= \boldsymbol{d}'D^{-2}(\boldsymbol{\gamma}_{\langle k+1\rangle} - \boldsymbol{\gamma}_{\langle k\rangle}) \\ &= \eta_{\langle k\rangle}\boldsymbol{p}_{\langle k\rangle}'\boldsymbol{p}_{\langle k\rangle} \\ &> 0\end{aligned} \tag{3.3}$$

since $\eta_{\langle k\rangle} > 0$ and $\boldsymbol{p}_{\langle k\rangle}\neq\mathbf{0}$. It is clear that the sequence $\{\boldsymbol{b}'\boldsymbol{\gamma}_{\langle k\rangle}\}$ is strictly increasing. Thus, the sequence converges since it is bounded from above by weak duality.

[Theorem 2] The sequence $\boldsymbol{\gamma}_{\langle k\rangle}$ generated by the algorithm converges.

[Proof] The difference of the objective functions between the k-th and $(k+1)$-th iterations is described as

$$\Delta_{\langle k\rangle} = \eta_{\langle k\rangle}\|\boldsymbol{p}_{\langle k\rangle}\|_2^2. \tag{3.4}$$

Since, from Theorem 1, the sequence $\{\boldsymbol{b}'\boldsymbol{\gamma}_{\langle k\rangle}\}$ converges, its difference sequence (3.4) tends to zero, that is,

$$\lim_{k\to\infty}\eta_{\langle k\rangle}\|\boldsymbol{p}_{\langle k\rangle}\|_2^2 = 0.$$

Therefore, $\|\boldsymbol{p}_{\langle k\rangle}\|_\infty$ is to converge to zero. Finally, since $\boldsymbol{\gamma}_{\langle k\rangle}$ belongs to the compact set Ω_γ, there exists a convergent subsequence $\{\boldsymbol{\gamma}_{\langle k\rangle}\}_K \to \boldsymbol{\gamma}^*$. For any such subsequence, we have $\boldsymbol{\gamma}^* > \mathbf{0}$ since the step length η is chosen so that this condition should be met. Consequently, the sequence $\boldsymbol{\gamma}_{\langle k\rangle}$ converges and the proof is complete.

4. Computational Aspects

In order to improve the algorithm with respect to the computational efficiency, one may update the projection $\boldsymbol{p}_{\langle k \rangle}$ at each iteration since the computational effort in the algorithm is dominated by the computation of $\boldsymbol{p}_{\langle k \rangle}$, in particular, the inverse of matrix $BD^2_{\langle k \rangle}B'$. The only quantity that changes from iteration to iteration is the diagonal elements of $D_{\langle k \rangle}$. So updating procedure can be employed to compute the matrix $(BD^2_{\langle k+1 \rangle}B')^{-1}$,

$$\begin{aligned}(BD^2_{\langle k+1 \rangle}B')^{-1} = \; & (BD^2_{\langle k \rangle}B')^{-1} - (BD^2_{\langle k \rangle}B')^{-1}B_G'Z_{GG} \\ & \times \{ I + B_G(BD^2_{\langle k \rangle}B')^{-1}B_G'Z_{GG} \}^{-1}B_G(BD^2_{\langle k \rangle}B')^{-1},\end{aligned}$$

where $Z = U - V$, $U = diag\{ \gamma^2_{\langle k+1 \rangle_i} \}$, $V = diag\{ \gamma^2_{\langle k \rangle_i} \}$, and G is the index set of the nonnull rows in Z.

On the other hand, the numerical instability problem may well be dealt with by the orthogonal decomposition approach. It turns out that the computation of $\boldsymbol{p}_{\langle k \rangle}$ is equivalent to computing the residuals of the weighted least squares problem,

$$\begin{aligned}\hat{\boldsymbol{\theta}} &= \arg\min_{\theta} \|D_{\langle k \rangle}\boldsymbol{b} - D_{\langle k \rangle}B'\boldsymbol{\theta}\|_2, \\ \boldsymbol{p}_{\langle k \rangle} &= D_{\langle k \rangle}\boldsymbol{b} - D_{\langle k \rangle}B'\hat{\boldsymbol{\theta}}.\end{aligned}$$

There are several methods for computing the vector $\boldsymbol{p}_{\langle k \rangle}$. One of them is to implement the orthogonal decomposition

$$QD_{\langle k \rangle}B' = \begin{bmatrix} T_1 \\ \cdots \\ 0 \end{bmatrix}, \qquad QD_{\langle k \rangle}\boldsymbol{b} = \begin{bmatrix} c_1 \\ \cdots \\ c_2 \end{bmatrix},$$

where Q is orthogonal matrix and T_1 is upper triangular. Then $\boldsymbol{p}_{\langle k \rangle}$ may simply be computed as

$$\boldsymbol{p}_{\langle k \rangle} = Q\begin{bmatrix} 0 \\ \cdots \\ c_2 \end{bmatrix}.$$

5. Concluding Remarks

The proposed algorithm can deal with any types of constraints in the Chebyshev estimation of the multiple regression model. By employing the linear scaling transformation scheme, the algorithm improves computational efficiency when the data set is quite large. Furthermore, it suggests the orthogonal decomposition technique to cope with numerical instability problem that may occur at the iterations.

References

[1] Abdelmalek, N. N. (1977). The Discrete Linear Restricted Chebyshev Approximation. *Nordisk Tidskrift for Information-behandling*, 17, 249-261.

[2] Appa, G. and Smith, C. (1973). On L_1 and Chebyshev Estimation. *Mathematical Programming*, 5, 73-87.

[3] Barrodale, I. and Phillips, C. (1975). Algorithm 495 : Solution of an Overdetermined System of Linear Equations in the Chebyshev Norm, *ACM Transactions on Mathematical Software*, 1, 264-270.

[4] Rice, J. R. and White, J. S. (1964). Norms for Smoothing and Estimation. *SIAM Review*, 6, 243-256.

[5] Sherali, H. D., Skarpness, B. O., and Kim, B. Y. (1988). "An Assumption-Free Convergence Analysis for a Perturbation of the Scaling Algorithm for Linear Programs, with Application to the L_1 Estimation Problem," *Naval Research Logistics*, vol. 35, 473-492.

[6] Sklar, M. G. and Armstrong, R. D. (1983). A Linear Programming Algorithm for the Simple Model for Discrete Chebychev Curve Fitting. *Computers and Operations Research*, 10, 237-248.

[7] Sklar, M. G. and Armstrong, R. D. (1984). An Algorithm for Discrete Chebychev Curve Fitting for the Simple Model using a Dual Linear Programming Approach. *Commun. Statist. - Simula. Computa.*, 13, 555-569.