The Korean Communications
in Statistics Vol. 5, No. 3, 1998
pp. 581-589

Constrained L_1-Estimation in Linear Regression[1]

Bu-Yong Kim[2]

Abstract

An algorithm is proposed for the L_1-estimation with linear equality and inequality constraints in linear regression model. The algorithm employs a linear scaling transformation to obtain the optimal solution of linear programming type problem. And a special scheme is used to maintain the feasibility of the updated solution at each iteration. The convergence of the proposed algorithm is proved. In addition, the updating and orthogonal decomposition techniques are employed to improve the computational efficiency and numerical stability.

1. Introduction

It is well known that the L_1-estimator is robust with respect to vertical outliers, and it has received considerable attention in the literatures of the robust regression. Statistical properties of the L_1-estimator have been studied by Blattberg and Sargent (1971), Rosenberg and Carlson (1977), Pfaffenberger and Dinkel (1978), Bassett and Koenker (1978), and Dielman and Pfaffenberger (1982). Bloomfield and Steiger (1983) describe the strong consistency and some robustness properties of the L_1-estimator. A necessary condition for the consistency of the L_1-estimator is proposed by Chen and Wu (1993). Birkes and Dodge (1993) present testing of hypotheses, confidence intervals, selection of variables, and so on with the L_1-estimation. Kim (1995) investigates the robustness, in terms of breakdown point, of the L_1-estimator. We consider L_1-estimation problem in the linear regression model with linear equality and inequality constraints,

$$\boldsymbol{y} = X\boldsymbol{\beta} + \boldsymbol{\varepsilon} : \ G\boldsymbol{\beta} = \boldsymbol{g}, \ H\boldsymbol{\beta} \le \boldsymbol{h} \tag{1.1}$$

where $\boldsymbol{y}$ denotes an n-vector of response variable, X a full-rank $n \times p$ matrix of regressors, $\boldsymbol{\beta}$ a p-vector of regression parameter, $\boldsymbol{\varepsilon}$ an n-vector of random error. Matrix $G(t \times p)$ and t-vector $\boldsymbol{g}$ define linear equality constraints, and matrix $H(r \times p)$ and r-vector $\boldsymbol{h}$ define

1) This paper was supported in part by Sookmyung Women's University, 1998.
2) Professor, Department of Statistics, Sookmyung Women's University, Seoul 140-742, Korea.

linear inequality constraints.

Algorithms for constrained L_1-estimation have been developed by Armstrong and Hultz (1977), Barrodale and Roberts (1977, 1978), and Bartels and Conn (1980). Kim (1987) suggests an unconstrained L_1-estimation procedure which exhibits a dominant trend with respect to computing effort as the size of data set increases. In this article we propose a constrained L_1-estimation procedure which is an extension of Kim (1987), and prove the convergence of the proposed algorithm.

2. Proposed Algorithm

This section starts with the linear programming problem to deal with the linear constraints in L_1-estimation,

$$\begin{aligned} &\text{minimize} && \boldsymbol{\ell}'\boldsymbol{e}^{+} + \boldsymbol{\ell}'\boldsymbol{e}^{-} \\ &\text{subject to} && X\hat{\boldsymbol{\beta}} + I\boldsymbol{e}^{+} - I\boldsymbol{e}^{-} = \boldsymbol{y} \\ & && G\hat{\boldsymbol{\beta}} = \boldsymbol{g},\ \ H\hat{\boldsymbol{\beta}} + I\boldsymbol{s} = \boldsymbol{h},\ \ \boldsymbol{e}^{+} \geq 0,\ \ \boldsymbol{e}^{-} \geq 0,\ \ \boldsymbol{s} \geq 0 \end{aligned} \tag{2.1}$$

where $\boldsymbol{\ell}$ is an n-vector of all ones, $\boldsymbol{s}$ is an r-vector of slack variable, and $\boldsymbol{e}^{+}$ and $\boldsymbol{e}^{-}$ represent positive and negative deviations, respectively. This problem can be solved by any simplex type algorithms. However, when the problem has a large number of observations, the problem size becomes so large that the direct applications of those algorithms to this problem is not computationally efficient. Thus, we apply a linear scaling transformation technique to this problem. For this purpose, we employ the big-M method, and obtain the dual problem as follows,

$$\begin{aligned} &\text{maximize} && \boldsymbol{y}'\boldsymbol{w}_{(1)} + \boldsymbol{g}'\boldsymbol{w}_{(2)} + \boldsymbol{h}'\boldsymbol{w}_{(3)} \\ &\text{subject to} && X'\boldsymbol{w}_{(1)} + G'\boldsymbol{w}_{(2)} + H'\boldsymbol{w}_{(3)} = 0 \\ & && -\boldsymbol{\ell} \leq \boldsymbol{w}_{(1)} \leq \boldsymbol{\ell},\ \ -\boldsymbol{m} \leq \boldsymbol{w}_{(2)} \leq \boldsymbol{m},\ \ -\boldsymbol{m} \leq \boldsymbol{w}_{(3)} \leq 0 \end{aligned} \tag{2.2}$$

where $\boldsymbol{w}_{(1)}' = [w_1, \cdots, w_n]$, $\boldsymbol{w}_{(2)}' = [w_{n+1}, \cdots, w_{n+t}]$, $\boldsymbol{w}_{(3)}' = [w_{n+t+1}, \cdots, w_{n+t+r}]$ are vectors of dual variables. $\boldsymbol{m}$ is vector of all m which is a large positive number, and represents the cost of each artificial variable. Since this problem has similar form to the bounded constraints on $\boldsymbol{w}$ in Kim (1987), we can extend the algorithm BKIML1. Now we need an initial feasible and strictly interior point. Of course, the two-phase method may be employed to obtain an initial feasible point. But the difficulty is that we still must obtain an initial feasible point in the first phase. So we have to convert this problem to a standard format of linear programming. However, it is very cumbersome and takes a significant number of steps to compute an initial feasible solution in the two-phase method since the problem size becomes large when the bounded constraints in (2.2) are explicitly included within the constraint matrix. Furthermore, it is not guaranteed that the feasible point obtained in the first phase is a strictly interior point which is the mandatory condition for the scaling algorithm. In this context, we consider an alternative procedure to solve the problem.

For simplicity, we define several notations used in this article ; $b' = [\, y' \ g' \ h' \,]$, $S' = [\, X' \ G' \ H' \,]$, $w' = [\, w'_{(1)} \ w'_{(2)} \ w'_{(3)} \,]$, $d' = [\, \ell' \ m' \ m' \,]$, $f = [\, \ell' \ m' \ 0' \,]$. By defining $D = \text{diag}\,[\, v_i \,]$, where

$$v_i = \begin{cases} \min\{\, 1 - w_i, 1 + w_i \,\} & \text{for } i = 1, \cdots, n \\ \min\{\, 1 - w_i/m, 1 + w_i/m \,\} & \text{for } i = n+1, \cdots, n+t \\ 1 + w_i/m & \text{for } i = n+t+1, \cdots, n+t+r \end{cases}$$

and putting $w = Dz$, the problem (2.2) in w-space is simply transformed to the problem (2.3) in z-space;

$$\text{maximize} \ \{\, (Db)'z : (DS)'z = 0, -d \le Dz \le f \,\}. \tag{2.3}$$

This transformation does not affect the solution of the problem (2.2). By applying the scaling linear transformation $Db = \tilde{b}$, $DS = \tilde{S}$ at the k-th iteration, we obtain the projection of $\tilde{b}$ onto the null space of $\tilde{S}'z = 0$ such as $u^k = \{\, I - \tilde{S}(\tilde{S}'\tilde{S})^{-1}\tilde{S}' \,\}\,\tilde{b}$. Then an improved feasible point z^{k+1} is obtained by moving along the projected gradient. Denoting $\eta (0 < \eta < 1)$ to be a step length, we can update the feasible point z^k as, $z^{k+1} = z^k + \eta u^k$. Let $p^k = Du^k$, then new point is mapped back to the w-space by the transformation of $z = D^{-1}w$,

$$w^{k+1} = w^k + \eta p^k.$$

We can initialize the algorithm with $w^0 = 0$ which is a feasible point for this problem. But since $w^0_{(3)} = 0$ is boundary value, we employ a special technique to maintain the feasibility of w^k at each iteration. This can be done if we concentrate on the constraints, $-m \le w^k_{(3)} \le 0$. When it starts at the point $w^k_{(3)} = 0$ and takes a step toward a new point $w^{k+1}_{(3)}$, not all directions are allowed if the new point is to be in the feasible region. For instance, if $w^k_{(3)i} = 0$ and the i-th inequality constraint is inactive, i.e., $u^k_i > 0$, then the new point $w^{k+1}_{(3)i} = w^k_{(3)i} + \eta p^k_i$ violates the constraint $-m \le w^{k+1}_{(3)} \le 0$ since $0 < \eta < 1$. Therefore, moving in the direction i is not permitted in this case. Thus, such a variable $w^k_{(3)i}$ must be detected and the inequality constraint corresponding to that variable must be temporarily deleted from the problem when the projected gradient is computed. In this respect, we first examine the feasibility of new potential solution w^{k+1}, and delete the most inactive constraint from the inequality constraints that violate the feasibility, i.e., $w^k_{(3)i} = 0$ and $u^k_i > 0$, and repeat until only the variables which satisfy the feasibility remain. Let $\Lambda = \{\, 1, \cdots, r \,\}$,

$C=\{\, i\in\Lambda \mid w^k_{(3)i}=0,\ u_{(3)i}>0 \,\}$, then $\bar{S}$ denotes the matrix S with the i-th row elements replaced by 0 if $i\in C$. Similarly, $\bar{\boldsymbol{b}}$ and $\bar{\boldsymbol{p}}$ are the vectors $\boldsymbol{b}$ and $\boldsymbol{p}$ with the i-th element replaced by 0 if $i\in C$, respectively. Now we compute the projected gradient and hence $\bar{\boldsymbol{p}}^k$ from $\bar{S}$ and $\bar{\boldsymbol{b}}$, and update $\boldsymbol{w}^k$. At the next iteration, if we have any inequality constraints which are deleted in the previous iteration but now satisfy the feasibility, then we add the most active one back into our problem and repeat until all the variables which satisfy the feasibility are included. Lemma 1 shows that this technique leads to a usable feasible direction. In addition, to maintain strictly interior feasibility of $\boldsymbol{w}^{k+1}$, we modify the step length as,

$$\eta=\alpha/\omega^k,\quad \omega^k=\max\,[\omega^k_1,\omega^k_2,\omega^k_3],$$

where $\boldsymbol{p}'=[\,\boldsymbol{p}_{(1)}'\ \boldsymbol{p}_{(2)}'\ \boldsymbol{p}_{(3)}'\,]$, $\omega^k_1=\max_i[\max\{\, p^k_{(1)i}/(1-w^k_{(1)i}),\ -p^k_{(1)i}/(1+w^k_{(1)i})\,\}\,]$, $\omega^k_2=\max_i[\max\{\, p^k_{(2)i}/(m-w^k_{(2)i}),\ -p^k_{(2)i}/(m+w^k_{(2)i})\,\}\,]$, $\omega^k_3=\max_i[\max\{\ -\bar{p}_{(3)i}{}^k/w^k_{(3)i},\ -\bar{p}_{(3)i}{}^k/(m+w^k_{(3)i})\,\}$ if $w^k_{(3)i}\neq 0$, $\max\{\,0,\ -\bar{p}_{(3)i}{}^k/m\,\}$ if $w^k_{(3)i}=0]$. The detail steps of the algorithm for the constrained L_1-estimation are as follows.

Algorithm : L1CON

Initialization. Set the initial feasible point $\boldsymbol{w}^k=0$ with the iteration counter $k=0$.

Step 1. Let $v^k_{(1)i}=\min\{\,1-w^k_{(1)i}, 1+w^k_{(1)i}\,\}$ $(i=1,\cdots,n)$, $v^k_{(2)i}=\min\{\,1-w^k_{(2)i}/m,\ 1+w^k_{(2)i}/m\,\}$ $(i=1,\cdots,t)$, $v^k_{(3)i}=1+w^k_{(3)i}/m$ $(i=1,\cdots,r)$, and $D^k=\text{diag}\,[v^k_{(1)i}\ v^k_{(2)i}\ v^k_{(3)i}]$. Then compute $\boldsymbol{u}^k=[I-D^kS\{\,S'(D^k)^2S\,\}^{-1}S'D\,\}D\boldsymbol{b}$.

Step 2. If any $w^k_{(3)i}=0$ and $u^k_{(3)i}>0$, then delete the most inactive inequality constraint, repeatedly. And if any $w^k_{(3)i}=0$ and $u^k_{(3)i}\le 0$ among the inequality constraints which are deleted at the previous iteration, then add the most active one back and repeat until only the variables that satisfy the feasibility remain.

Step 3. If no inequality constraint is deleted in Step 2, then replace $\boldsymbol{u}^k$ for $\bar{\boldsymbol{u}}^k$. Otherwise, compute $\bar{\boldsymbol{u}}^k=[I-D^k\bar{S}\{\,\bar{S}'(D^k)^2\bar{S}\,\}^{-1}\bar{S}'D^k\,\}D^k\bar{\boldsymbol{b}}$.

Step 4. Compute $\bar{\boldsymbol{p}}^k=D^k\bar{\boldsymbol{u}}^k$. If $\|\,\bar{\boldsymbol{p}}^k\,\|_\infty<\delta$ for termination tolerance $\delta>0$, then let the constrained L_1-estimate be $\hat{\beta}=(\bar{S}'\bar{S})^{-1}\bar{S}'\bar{\boldsymbol{b}}$ and stop.

Step 5. Let $\omega^k_1=\max_i[\max\{\, p^k_{(1)i}/(1-w^k_{(1)i}),\ -p^k_{(1)i}/(1+w^k_{(1)i})\,\}\,]$, $\omega^k_2=\max_i[\max\{\, p^k_{(2)i}/(m-w^k_{(2)i}),\ -p^k_{(2)i}/(m+w^k_{(2)i})\,\}\,]$, $\omega^k_3=\max_i[\max\{\ -\bar{p}_{(3)i}{}^k/w^k_{(3)i},\ -\bar{p}_{(3)i}{}^k/$

$(m+w^k_{(3)i})\}$ if $w^k_{(3)i}\neq 0$, $\max\{0,\ -\bar{p}_{(3)i}{}^k/m\}$ if $w^k_{(3)i}=0]$, and $\omega^k=\max[\omega^k_1, \omega^k_2, \omega^k_3]$.

Update the iterate, $\boldsymbol{w}^{k+1}=\boldsymbol{w}^k+(\alpha/\omega^k)\,\bar{\boldsymbol{p}}^k$, and return to Step 1 with $k=k+1$.

3. Convergence of the Algorithm

The proposed algorithm employs a linear scaling transformation approach. In this section, it is proved that not only the approach works well, but also the algorithm converges.

Lemma 1. Each iterate $\boldsymbol{w}^k$ of the proposed algorithm is feasible.

Proof. At the initialization we set $\boldsymbol{w}^0=\boldsymbol{0}$, so the feasibility is satisfied. Assume that $\boldsymbol{w}^k$ is feasible, then since $\bar{p}_i{}^k$ corresponding to the i-th inactive inequality constraint is equal to zero, it follows that

$$\begin{aligned}\bar{S}'\boldsymbol{w}^{k+1} &= \bar{S}'\boldsymbol{w}^k+(\alpha/\omega^k)\bar{S}'\,\bar{\boldsymbol{p}}^k\\ &= \bar{S}'\boldsymbol{w}^k+(\alpha/\omega^k)\bar{S}'(D^k)^2[\bar{\boldsymbol{b}}-\bar{S}\{\bar{S}'(D^k)^2\bar{S}\}^{-1}\bar{S}'(D^k)^2\bar{\boldsymbol{b}}]\\ &= \bar{S}'\boldsymbol{w}^k.\end{aligned}$$

And for $\boldsymbol{w}^{k+1}$ to be a strictly interior feasible point, ω^k must satisfy the conditions,

$$-1\le w^k_{(1)i}+p^k_{(1)i}/\omega^k_1\le 1,\quad -m\le w^k_{(2)i}+p^k_{(2)i}/\omega^k_2\le m,\quad -m\le w^k_{(3)i}+\bar{p}_{(3)i}{}^k/\omega^k_3\le 0.$$

These conditions are explicitly written as follows,

(1) If $p^k_{(1)i}\ge 0$, then $\omega^k_1\le -p^k_{(1)i}/(1+w^k_{(1)i})$ or $\omega^k_1\ge p^k_{(1)i}/(1-w^k_{(1)i})$,
otherwise, $\omega^k_1\le p^k_{(1)i}/(1-w^k_{(1)i})$ or $\omega^k_1\ge -p^k_{(1)i}/(1+w^k_{(1)i})$.

(2) If $p^k_{(2)i}\ge 0$, then $\omega^k_2\le -p^k_{(2)i}/(m+w^k_{(2)i})$ or $\omega^k_2\ge p^k_{(2)i}/(m-w^k_{(2)i})$,
otherwise, $\omega^k_2\le p^k_{(2)i}/(m-w^k_{(2)i})$ or $\omega^k_2\ge -p^k_{(2)i}/(m+w^k_{(2)i})$.

(3) If $\bar{p}_{(3)i}{}^k>0$, then $\omega^k_3\le -\bar{p}_{(3)i}{}^k/(m+w^k_{(3)i})$ or $\omega^k_3\ge -\bar{p}_{(3)i}{}^k/w^k_{(3)i}$,
otherwise, $\omega^k_3\le -\bar{p}_{(3)i}{}^k/w^k_{(3)i})$ or $\omega^k_3\ge -\bar{p}_{(3)i}{}^k/(m+w^k_{(3)i})$ when $w^k_{(3)i}\neq 0$,
$\omega^k_3\ge -\bar{p}_{(3)i}{}^k/m$ when $w^k_{(3)i}=0$.

Consequently, we can choose ω^k as in Step 5 to maintain the feasibility. Thus, the feasibility holds for each iteration.

Theorem 1. If $\boldsymbol{p}^k=\boldsymbol{0}$ at any iteration, then $\boldsymbol{w}^k$ is optimal. Otherwise assume that $\boldsymbol{p}^k\neq\boldsymbol{0}$ for all k. Then $\{\boldsymbol{y}'\boldsymbol{w}^k_{(1)}+\boldsymbol{g}'\boldsymbol{w}^k_{(2)}+\boldsymbol{h}'\boldsymbol{w}^k_{(3)}\}$ converges in problem (2.2).

Proof. If $p^k = 0$, then $u^k = 0$. It implies that objective function is a constant for all feasible solution. That is, w^k is optimal. Now suppose $p^k \neq 0$ for all k, and prove that sequence $\{ y' w^k_{(1)} + g' w^k_{(2)} + h' w^k_{(3)} \}$ is strictly increasing. Since we can assume without loss of generality that matrix $\bar{S}$ has full column rank, the projected gradient on the original space is

$$\bar{p}^k = (D^k)^2 \bar{b} - (D^k)^2 \bar{S} \{ \bar{S}'(D^k)^2 \bar{S}) \}^{-1} \bar{S}'(D^k)^2 \bar{b}. \tag{3.1}$$

It follows from (3.1) that

$$\{ \bar{b} - (D^k)^{-2} \bar{p}^k \}' = \bar{b}'(D^k)^2 \bar{S} \{ \bar{S}'(D^k)^2 \bar{S} \}^{-1} \bar{S}'. \tag{3.2}$$

And the following relationship can be obtained from Lemma 1 and equation (3.2),

$$\begin{aligned} &\bar{S}'(w^{k+1} - w^k) = 0 \\ &\bar{b}'(D^k)^2 \bar{S} \{ \bar{S}'(D^k)^2 \bar{S} \}^{-1} \bar{S}'(w^{k+1} - w^k) = 0 \\ &\bar{b}'(w^{k+1} - w^k) = (\bar{p}^k)'(D^k)^{-2}(w^{k+1} - w^k). \end{aligned} \tag{3.3}$$

Since $(D^k)^{-2}$ is positive definite and $\bar{p}^k \neq 0$ from the assumption it follows that

$$(\bar{p}^k)'(D^k)^{-2}(w^{k+1} - w^k) = (\alpha/\omega^k)(\bar{p}^k)'(D^k)^{-2} \bar{p}^k > 0.$$

It implies that $\bar{b}'(w^{k+1} - w^k) > 0$. Similarly, it can be shown that $b'(w^{k+1} - w^k) > 0$. Therefore, we have $b' w^{k+1} > b' w^k$ at each iteration k. On the other hand, since $\{ b' w^k \}$ is bounded from above by the weak duality, $\{ y' w^k_{(1)} + g' w^k_{(2)} + h' w^k_{(3)} \}$ converges.

Theorem 2. The proposed algorithm terminates in a finite number of iterations for any termination tolerance $\delta > 0$.

Proof. It follows from (3.3) and the definition of L_2-norm that

$$\bar{b}' w^{k+1} - \bar{b}' w^k = (\alpha/\omega^k)(\bar{p}^k)'(D^k)^{-2} \bar{p}^k = (\alpha/\omega^k) \| (D^k)^{-1} \bar{p}^k \|_2^2 .$$

The convergence of $\{ \bar{b}' w^k \}$ implies that its difference sequence tends to zero, that is,

$$\{ (\alpha/\omega^k) \| (D^k)^{-1} \bar{p}^k \|_2^2 \} \to 0. \tag{3.4}$$

It follows from Step 1 and 5 of the algorithm and the definition of L_∞-norm that

$$\begin{aligned} 0 \le \omega^k &= \max \{ \omega^k_1, \omega^k_2, \omega^k_3 \} \\ &\le \max_i \{ | \bar{p}^k_i | / v^k_i \} \quad \text{if} \quad \omega^k \neq -\bar{p}^k_{(3)i} / w^k_{(3)i} \\ &= \| (D^k)^{-1} \bar{p}^k \|_\infty. \end{aligned} \tag{3.5}$$

However, if $\omega^k = -\bar{p}^k_{(3)i} / w^k_{(3)i}$, it implies that $w^{k+1}_{(3)i} = 0$. At the next iteration, this constraint is deleted if $u_{(3)i}^{k+1}>0$. But if $u_{(3)i}^{k+1}\leq 0$, then $-\bar{p}_{(3)i}^{\,k+1}/w_{(3)i}^{k+1}\leq 0$ and hence $\omega^{k+1}\neq -\bar{p}_{(3)i}^{\,k+1}/w_{(3)i}^{k+1}$. Therefore, this particular case does not affect the above result. In addition, it follows from (3.5) that

$$0\leq \alpha \| (D^k)^{-1}\bar{p}^k \|_2 \leq \alpha \| (D^k)^{-1}\bar{p}^k \|_2^2 / \| (D^k)^{-1}\bar{p}^k \|_\infty$$
$$\leq (\alpha/\omega^k) \| (D^k)^{-1}\bar{p}^k \|_2^2 .$$

Thus, the necessary condition for (3.4) is

$$\{ \| (D^k)^{-1}\bar{p}^k \|_2 \} \rightarrow 0. \tag{3.6}$$

Now, since $\max_i | \bar{p}_i^k | = \| \bar{p}^k \|_\infty \leq \| (D^k)^{-1}\bar{p}^k \|_\infty \leq \| (D^k)^{-1}\bar{p}^k \|_2$, it follows from (3.6) that $\{ p_i^k \}\rightarrow 0$, for all $i=1,\cdots,n+t+r$. Hence, the algorithm terminates in a finite number of iterations for some chosen tolerance $\delta>0$.

4. Computational Aspects

The algorithm is so computationally intractable that much effort should be made to improve the computational efficiency and numerical stability. To improve the computational efficiency one may employ updating $\bar{p}^k$ at each iteration since computing effort in the algorithm is dominated by the computation of $\bar{p}^k$, in particular, the inverse matrix of $\bar{S}'(D^k)^2\bar{S}$. The only quantity that changes from iteration to iteration is the diagonal elements of D^k. The updating procedure can be employed to compute $\{ \bar{S}'(D^{k+1})^2\bar{S} \}^{-1}$ on the basis of $\{ \bar{S}'(D^k)^2\bar{S} \}^{-1}$. Define $U=(D^{k+1})^2-(D^k)^2$ and denote by Z a set of indices corresponding to nonnull rows in U. Let $V=\{ \bar{S}'(D^k)^2\bar{S} \}^{-1}$, then

$$\begin{aligned}\{ \bar{S}'(D^{k+1})^2\bar{S} \}^{-1} &= \{ \bar{S}'(D^k)^2\bar{S}+\bar{S}'U\bar{S} \}^{-1}\\ &= (I+V\bar{S}_Z'U_{ZZ}\bar{S}_Z)^{-1}V\\ &= \{ I-V\bar{S}_Z'U_{ZZ}(I+\bar{S}_ZV\bar{S}_Z'U_{ZZ})^{-1}\bar{S}_Z \} V\\ &= V-V\bar{S}_Z'U_{ZZ}(I+\bar{S}_ZV\bar{S}_Z'U_{ZZ})^{-1}\bar{S}_ZV.\end{aligned}$$

When D^k and D^{k+1} differ only in a few elements (this situation happens when the solution is close to the optimum), that is, the dimension of U_{ZZ} is very small, the updating scheme results in significant improvement in the computational efficiency.

On the other hand, the computational instability problem may well be dealt with by the orthogonal transformation approach. It turns out that the computation of $\bar{u}^k$ (or u^k) is equivalent to computing the residuals of the weighted least squares problem:

$$\theta = \arg\min_{\theta} \| D^k \bar{b} - D^k \bar{S}\theta \|_2, \qquad \bar{u}^k = D^k \bar{b} - \bar{S}' D^k \theta.$$

One of the methods for computing $\bar{u}^k$ is to implement the orthogonal decomposition,

$$QD^k\bar{S} = [\, T' \ \ 0' \,]', \quad QD^k\bar{b} = [\, c_1' \ \ c_2' \,]'$$

where Q is orthogonal and T is upper triangular. Then $\bar{u}^k$ can be computed as follows

$$\bar{u}^k = Q[\, 0' \ \ c_2' \,]'.$$

5. Concluding Remarks

The comparison of constrained L_1-algorithms in terms of computational efficiency via simulation is not quite reasonable since the computational efficiency of constrained algorithms depends highly on the structure of constraints. However, since the constrained algorithms are the extensions of the unconstrained algorithms, it may be useful to consider the simulation studies given by Kim (1987) on the computational efficiency of the unconstrained L_1-algorithms.

References

[1] Armstrong, R. D. and Hultz, J. W. (1977). An Algorithm for a Restricted Discrete Approximation Problem in the L_1-Norm, *SIAM Journal on Numerical Analysis*, Vol. 14, 555-565.

[2] Barrodale, I. and Roberts, F. D. K. (1977). Algorithms for Restricted Least Absolute Value Estimation, *Communications in Statistics-Simula. Computa.*, Vol. 6, 353-363.

[3] Barrodale, I. and Roberts, F. D. K. (1978). An Efficient Algorithm for Discrete L_1 Linear Approximation with Linear Constraints, *SIAM Journal on Numerical Analysis*, Vol. 15, 603-611.

[4] Bartels, R. H. and Conn, A. R. (1980). Linearly Constrained Discrete L_1 Problems, *ACM Transactions on Mathematical Software*, Vol. 6, 594-608.

[5] Bassett, G. and Koenker, R. (1978). Asymptotic Theory of Least Absolute Error Regression, *Journal of the American Statistical Association*, Vol. 73, 618-622.

[6] Birkes, D. and Dodge, Y. (1993). *Alternative Methods of Regression,* John Wiley and Sons, New York.

[7] Blattberg, R. and Sargent, T. (1971). Regression with Non-Gaussian Stable Disturbances ; Some Sampling Results, *Econometrica*, Vol. 39, 501-510.

[8] Bloomfield, P. and Steiger, W. L. (1983). *Least Absolute Deviations : Theory, Applications, and Algorithms,* Birkhauser Boston, Inc.

[9] Chen, X. R. and Wu, Y. (1993). On a Necessary Condition for the Consistency of the L_1-estimates in Linear Regression Models, *Communications in Statistics - Theory Meth.*, Vol. 22, 631-639.

[10] Dielman, T. and Pfaffenberger, R. (1982). Least Absolute Value Estimation in Linear Regression ; A Review, *TIMS/Studies in the Management Sciences*, Vol. 19, 31-52.

[11] Kim, B. Y. (1987). A Robust Estimation Procedure for the Linear Regression Model, *Journal of the Korean Statistical Society*, Vol. 16, 80-91.

[12] Kim, B. Y. (1995). On the Robustness of L_1-Estimator in Linear Regression Model, *The Korean Communications in Statistics*, Vol. 2, 277-287.

[13] Pfaffenberger, R. C. and Dinkel, J. J. (1978). Absolute Deviations Curve Fitting ; An Alternative to Least Squares, *Contributions to Survey Sampling and Applied Statistics*, edited by H. A. David, Academic Press, NY, 279-294.

[14] Rosenberg, B. and Carlson, D. (1977). A Simple Approximation of the Sampling Distribution of Least Absolute Residuals Regression Estimates, *Communications in Statistics - Simulation and Computation*, Vol. B6, 421-437.