FOR AUTHOR'S CORRECTION ONLY

Computing (2003)
Digital Object Identifier (DOI) 10.1007/s00607-003-0035-1

Computing
Printed in Austria

Finite Computation of the ℓ_1 Estimator from Huber's M-Estimator in Linear Regression

M. Ç. Pınar, Ankara

Received September 17, 2003; revised September 25, 2003
Published online: ■ ■ ■

Abstract

We review and extend previous work on the approximation of the linear ℓ_1 estimator by the Huber M-estimator based on the algorithms proposed by Clark and Osborne [7], and Madsen and Nielsen [12]. Although the Madsen-Nielsen algorithm is a promising one, it is guaranteed to terminate finitely under certain assumptions. We describe a variant of the Madsen-Nielsen algorithm to compute the ℓ_1 estimator from the Huber M-estimator in a finite number of steps without any restrictive steps nor assumptions. Summary computational results are given.

Keywords: Multiple linear regression, the ℓ_1 estimator, huber's M-estimator, finite algorithms.

1 Introduction

Consider the linear model

$$r = A^T x - b \tag{1}$$

where $x \in \Re^n$, $b \in \Re^m$ is the vector of dependent observations, $A \in \Re^{n\times m}$ (with $m > n$) is the matrix of independent observations and $r \in \Re^m$ is the vector of residuals. The purpose of this work is to review and extend algorithms for computing the linear ℓ_1 estimator using Huber's M-estimator in (1). The linear ℓ_1 estimation problem consists of finding a vector $x^* \in \Re^n$ to the following minimization problem:

[L1]

$$\text{minimize} \quad G(x) \equiv \|r\|_1 \equiv \|A^T x - b\|_1. \tag{2}$$

The notation $\|z\|_1$ is used to denote the ℓ_1 norm which is the sum of absolute values of the components of z, i.e., $\|z\|_1 = \sum_{i=1}^n |z_i|$. The linear ℓ_1 estimation problem is more difficult than the the least squares problem where the ℓ_2 norm is used. The ℓ_2 estimation problem admits a closed form solution whereas the ℓ_1 estimation problem is of a combinatorial nature as it can be recast as a linear programming problem.

S	W	0	1	1	5	0	3	B	Dispatch: 6.11.2003	Journal: Computing	No. of pages: 20
Journal number				Manuscript number					Author's disk received ☑ Used ☑	Corrupted ☐	Mismatch ☐ Keyed ☐

The ℓ_2 estimator is known to be the maximum likelihood estimator under the assumption that the residuals $r = A^T x - b$ have independent identical normal distributions. However, the ℓ_2 estimator is quite sensitive to deviations from this assumption, and the presence of a few outliers among data points may have a significant effect. The interest for the ℓ_1 estimator stems from its robustness in the face of outliers as discussed in [9]. Hence, despite the increase in computational difficulty compared to the ℓ_2 case, the ℓ_1 estimator was studied also extensively; see e.g. [4, 21] for a review of developments until 1982. There has been renewed interest in the ℓ_1 estimator as evidenced by the emergence of recent ideas in [8, 19, 20, 22].

An alternative robust estimator which does not involve nonsmooth optimization was proposed by Huber [9] as the minimizer x^* of the function

$$\Phi(x) = \sum_{i=1}^{m} \phi(r_i, \gamma) \tag{3}$$

where

$$\phi(r_i, \gamma) = \begin{cases} \frac{1}{2\gamma} r_i^2 & \text{if } |r_i(x)| \le \gamma \\ |r_i| - \frac{\gamma}{2} & \text{if } |r_i(x)| \ge \gamma, \end{cases} \tag{4}$$

and γ is a positive scalar to be estimated from the data. This estimator was shown by Huber to be a maximum likelihood estimator for a perturbed normal distribution and became known as Huber's M-estimator. Interestingly, there exist intimate relationships between the ℓ_1 estimator and the Huber M-estimator. This is to be expected since

$$\lim_{\gamma \to 0} \phi(r_i, \gamma) = |r_i|. \tag{5}$$

This fact has been noticed and studied by essentially three research groups resulting in the papers by Clark [6], Clark and Osborne [7], Madsen and Nielsen [11, 12], Madsen, Nielsen and Pınar [13, 15] and Li and Swetits [10], with valuable insights and algorithms for computing both the M-estimator and the ℓ_1 estimator. The first finite algorithm for computing the ℓ_1 estimator from the Huber M-estimator was proposed by Clark and Osborne. Later, this algorithm was extended by Madsen and Nielsen and Madsen, Nielsen and Pınar. The Madsen-Nielsen algorithm was reported in [12] to be up competitive with the Barrodale-Roberts implementation of the simplex algorithm for the ℓ_1 estimation problem [3], a significant contribution considering that the Barrodale-Roberts algorithm is regarded as one of the most efficient algorithms in this area. This was testimony to the promise of the new approach. Later, Madsen, Nielsen and Pınar [15] used this algorithm to solve linear programming problems, extending both the theory and practice of the new algorithm. All these algorithms have guaranteed finite termination under some restrictive assumptions. Later, Li and Swetits proposed a recursive variant of the Madsen and Nielsen algorithm and proved its finiteness property under the full rank assumption on A. Under the light of the above discussion the purpose of the present paper is to review the computational ties between the Huber M-estimator and the linear ℓ_1 estimator and to give a new finite algorithm for computing the ℓ_1 estimator from the Huber M-estimator. The new algorithm consists of a simple modification of the Madsen-Nielsen algorithm, and terminates finitely without any assumptions. It is inspired from the original Clark-Osborne algorithm. Comparative computational results with the modified Madsen-Nielsen algorithm show that it is competitive with the most successful implementations of the simplex type algorithms.

The plan of this paper is as follows. In Section 2, we will give basic properties of the ℓ_1 estimator, some key results on Huber's M-estimator and the connections between these two, respectively. We will study the algorithmic contributions of Clark and Osborne, and Madsen and Nielsen in Section 3 and 4, respectively. We give further results on the connection between the two estimators in Section 5. We propose an extension of the Madsen-Nielsen algorithm in Section 6 and prove its finite termination property. We also summarize computational experience with the modified algorithm in Section 6.

2 Properties

In this section we review some relevant properties of the ℓ_1 estimator, the Huber M-estimator and their connections, in this order, respectively.

2.1 The ℓ_1 Estimator

The ℓ_1 estimator is characterized by the following necessary and sufficient condition for optimality [21]: x is an ℓ_1 estimator iff there exists $\lambda_j \in [-1, 1]$ such that

$$\sum_{j \in \mathscr{A}_0(x)} \lambda_j a_j + \sum_{j \notin \mathscr{A}_0(x)} a_j s_j = 0 \tag{6}$$

where $\mathscr{A}_0(x) = \{j : r_j(x) = 0\}$ and s_j is defined for all j as

$$s_j(x) = \begin{cases} -1 & \text{if } r_i(x) < 0 \\ 0 & \text{if} |r_i(x)| = 0 \\ 1 & \text{if } r_i(x) > 0. \end{cases} \tag{7}$$

An interesting duality result links the ℓ_1 estimator with linear programming. It can be shown using Lagrange duality that the dual problem to [L1] is given as

[D1]

$$\begin{aligned} \max \quad & -b^T y \\ s.t. \quad & Ay = 0 \\ & -\mathbf{1} \leq y \leq \mathbf{1} \end{aligned}$$

where $-\mathbf{1}$ and $\mathbf{1}$ denote vectors with components -1 and 1, respectively. Furthermore, x solves [L1] and y solves the dual problem if and only if y satisfies $Ay = 0$ and the following conditions hold

$$r_j(x) < 0 \Longrightarrow y_j = -1, \tag{8}$$
$$r_j(x) > 0 \Longrightarrow y_j = 1, \tag{9}$$

and

$$-1 < y_j < 1 \Longrightarrow r_j(x) = 0. \tag{10}$$

It is easy to notice that these conditions are fully equivalent to the optimality condition (6).

2.2 The Huber M-Estimator

Define for a given threshold $\gamma > 0$ the sign vector

$$s^\gamma(x) = [s_1^\gamma(x), \ldots, s_m^\gamma(x)] \tag{11}$$

with

$$s_i^\gamma(x) = \begin{cases} -1 & \text{if } r_i(x) \le -\gamma \\ 0 & \text{if } |r_i(x)| < \gamma \\ 1 & \text{if } r_i(x) \ge \gamma. \end{cases} \tag{12}$$

If $s = s^\gamma(x)$ then we also denote W_s the $m \times m$ diagonal matrix whose ith diagonal entry is given by $1 - s_i^2$. Alternatively, we will also use $W(x)$ to denote the diagonalmatrix associated with $s^\gamma(x)$ directly. Now, the Huber M-estimation problem can be recast as the following minimization problem:

[SL1]

$$\text{minimize } G_\gamma(x) \equiv \frac{1}{2\gamma} r^T W_s r + s^{\gamma T}[r - \frac{1}{2}\gamma s^\gamma] \tag{13}$$

where the argument x is dropped for notational convenience. Clearly, G_γ measures the "small" residuals ($|r_i(x)| < \gamma$) by their squares while the "large" residuals are measured by the ℓ_1 function. Thus, G_γ is a piecewise quadratic function, and it is continuously differentiable in $\Re^n$.

G_γ is composed of a finite number of quadratic functions. In each domain $D \subseteq \Re^n$ where $s^\gamma(x)$ is constant G_γ is equal to a specific quadratic function as seen from the above definition. These domains are separated by the following union of hyperplanes,

$$B_\gamma = \{x \in \Re^n | \exists\, i : |r_i(x)| = \gamma\}. \tag{14}$$

A sign vector s is γ-*feasible* at x if

$$\forall\, \varepsilon > 0\; \exists\, z \in \Re^n \setminus B_\gamma : \|x - z\| < \varepsilon \;\wedge\; s = s^\gamma(z). \tag{15}$$

If s is a γ-feasible sign vector at some point x then Q_s is the quadratic function which equals G_γ on the subset

$$\mathscr{C}_s^\gamma = \mathrm{cl}\{z \in \Re^n | s^\gamma(z) = s\}. \tag{16}$$

$\mathscr{C}_s^\gamma$ is called a *Q-subset* of $\Re^n$. Notice that any $x \in \Re^n \setminus B_\gamma$ has exactly one corresponding Q-subset ($s = s^\gamma(x)$), whereas a point $x \in B_\gamma$ belongs to two or more Q-subsets. Therefore, we must in general give a sign vector s in addition to x in order to specify which quadratic function we are currently considering as representative of G_γ.

Q_s can be defined as follows:

$$Q_s(z) = \frac{1}{2\gamma}(z - x)^T (AW_sA^T)(z - x) + G_\gamma'^T(x)(z - x) + G_\gamma(x). \tag{17}$$

The gradient of the function G_γ is given by

$$G_\gamma'(x) = A[\frac{1}{\gamma} W_s r + s] \tag{18}$$

where s is a γ-feasible sign vector at x. For $x \in \Re^n \setminus B_\gamma$, the Hessian of G_γ exists, and is given by

$$G_\gamma''(x) = \frac{1}{\gamma} AW_sA^T. \tag{19}$$

The set of indices corresponding to "small" residuals

$$\mathscr{A}_\gamma(z) = \{i | 1 \le i \le m \;\wedge\; |r_i(z)| \le \gamma\} \tag{20}$$

is called the γ-*active* set at z. The set of minimizers of G_γ is denoted by M_γ.

Interestingly, there exists a simple duality link between the Huber M-estimation problem and quadratic programming. More precisely, it can be shown using Lagrange duality (see e.g., [17]) that the dual of the Huber M-estimation is the following quadratic program:

[D2]

$$\begin{array}{ll} \max & -b^T y - \dfrac{\gamma}{2} y^T y \\ s.t. & Ay = 0 \\ & -\mathbf{1} \le y \le \mathbf{1} \end{array}$$

Furthermore, the optimal solutions x^* of [SL1] and its dual are related by the identity:

$$y^* = \frac{1}{\gamma} W_s r(x^*), \tag{21}$$

where $s = s^\gamma(x^*)$. As y^* is the unique solution to the dual problem (the dual problem is a strictly concave maximization problem) we have the following simple but important consequences of the duality result.

Lemma 1 *$s^\gamma(x_\gamma)$ is constant for $x_\gamma \in M_\gamma$. Furthermore $r_i(x_\gamma)$ is constant for $x_\gamma \in M_\gamma$ if $s_i^\gamma = 0$.*

Following the lemma we use the notation $s^\gamma(M_\gamma) = s^\gamma(x_\gamma), x_\gamma \in M_\gamma$ as the sign vector corresponding to the solution set.

Based on the work of Mangasarian and Meyer [16], it can be shown that the point y^* defined in (21) is a least norm solution of the linear program [D1] provided that $\gamma > 0$ is sufficiently small. Li and Swetits [10] use this result to give a recursive procedure to compute the ℓ_1 estimator from Huber's M-estimator.

2.3 Connections between the ℓ_1 Estimator and the Huber M-Estimator

The purpose of this section is to summarize some key relationships between the linear ℓ_1 estimator and the Huber M-estimator. In particular, the solution set of the M-estimation problem allows a description of the solution set of the ℓ_1 estimation problem.

Assume $x_\gamma \in M_\gamma$, and let $s = s^\gamma(M_\gamma)$ and $W = W_s$. Then x_γ is a solution to the following system of linear equations:

$$AW\,A^T x_\gamma = AW\,b - \gamma As. \tag{22}$$

Now, assume that $x_\gamma + \delta h$ is a minimizer of $G_{\gamma-\delta}$ with $s^\gamma(x_\gamma + \delta h) = s$. Thus, we can write

$$AW\,A^T(x_\gamma + \delta h) = AW\,b - (\gamma - \delta)As.$$

This implies that h solves the system

$$AW\,A^T h = As. \tag{23}$$

This system of linear equations is always consistent since it is equivalent to the following system:

$$AW\,A^T h = -\frac{1}{\gamma} AW\,r(x_\gamma)$$

which corresponds to normal equations associated with $W\,A^T h = -\frac{1}{\gamma} W\,r(x_\gamma)$.

Next, we state an important result without proof from [13]. Let $\mathscr{S}$ denote the set of minimizers of [L1] and $\mathscr{D}_s^0 = \{x | r_i(x) \leq 0, i \in \sigma_-(s) \wedge r_i(x) \geq 0, i \in \sigma_+(s)\}$ where $\sigma_+(s) = \{i | s_i = 1\}$ and $\sigma_-(s) = \{i | s_i = -1\}$. Let $\sigma(s)$ denote the complement of $\sigma_-(s) \cup \sigma_+(s)$ with respect to $\{1, \ldots, m\}$.

Theorem 1 *(a) There exists* $\gamma_0 > 0$ *such that* $s^\gamma(M_\gamma)$ *is constant for* $0 < \gamma \leq \gamma_0$.

(b) For $0 < \gamma \leq \gamma_0$, *where* γ_0 *is given in (a), let* $s = s^\gamma(M_\gamma)$, *and let* $\mathcal{N}_s$ *denote the orthogonal complement to span*$\{a_i | s_i = 0\}$. *If* $x_\gamma \in M_\gamma$, *and d solves* (23) *then*

$$M_0 \equiv \mathscr{S}$$

where

$$M_0 = (x_\gamma + \gamma d + \mathcal{N}_s) \cap \mathscr{D}_s^0, \tag{24}$$

and

$$y^* = \frac{1}{\gamma} W_s r(x_\gamma) + s \tag{25}$$

solves [D1].

The above theorem gives a description of the set of ℓ_1 estimators from the set of M-estimators for small enough values of γ. We will use this result in our description of the Madsen-Nielsen algorithm and the variant of it we will propose.

3 The Clark-Osborne Continuation Algorithm

The Clark-Osborne algorithm is a continuation algorithm which was not originally intended as a device for solving the linear ℓ_1 estimation problem. Its prescribed use was to compute the Huber M-estimator for suitable values of γ starting from a large enough value so that the γ-active set includes all the indices. In otherwords, the Clark-Osborne algorithm begins with a large value of γ to mimic the ℓ_2 estimator and decreases γ until its desired value by following the piecewise linear path of Huber M-estimators. In this section we give a slightly modified version of this algorithm, tailored to compute the ℓ_1 estimator.

To carry on with a preliminary description of this algorithm we give a new sign vector definition:

$$s_{\gamma i}(x) = \begin{cases} -1 & \text{if } r_i(x) < -\gamma \\ 0 & \text{if } |r_i(x)| \leq \gamma \\ 1 & \text{if } r_i(x) > \gamma. \end{cases} \tag{26}$$

We will refer to s_γ as an "extended" sign vector. Notice that s^γ and s_γ differ only for those residuals that are on the boundary B. The Clark-Osborne algorithm works with the above definition of a sign vector rather than (12). In this section we will refer to the sign vector s_γ associated with the unique Huber M-estimator as the "optimal extended sign vector". We assume in this section that A has full rank.

The key idea that motivates the Clark-Osborne algorithm is the linear system (23). Assume the Huber M-estimator x_γ is unique and that it is non-degenerate, i. e. , at any value of γ the set $\{i|r_i(x_\gamma) = \gamma\}$ is a singleton. Clark shows that if the M-estimator is unique the matrix $AW\,A^T$ has full rank, c. f. Lemma 6 of [6]. Since there exists a continuum of values of γ for a finite set of possible γ-feasible sign vectors s_γ, one can immediately deduce from our analysis of the previous section that that s_γ remains constant by intervals. The intervals corresponding to sign vectors constitute"segments" of the piecewise linear path of M-estimators. We refer to these as the "sign intervals".

The algorithm consists of following the unique path of M-estimators using the linear system of equations (23). Under the assumption of uniqueness, and the nondegeneracy of M-estimators, the Clark-Osborne algorithm traces the piecewise linear segments of this path. They use the nondegeneracy assumption to show that when moving from one segment to another, at the change of segment the adjacent sign vectors differ by a single entry. Furthermore, the sign vector obtained from this single change is the optimal extended sign vector of the next segment.

In the rest of this section we will make the Clark-Osborne algorithm mathematically precise.

The basic algorithm can be formulated as follows:

```
find the ℓ2 estimator
choose initial γ
  repeat
    Compute h from (23)
    Decrease γ along h
  until γ = 0.
```

The ℓ_2 estimator is found as the solution x_{ls} of the linear system:

$$AA^T x = Ab.$$

The parameter γ is initialized to $\max_{j=1,\dots,m} |r_j(x_{ls})|$. The next step in the algorithm is to trace the path of M-estimators. To do this, one computes the unique solution h to the system

$$AW\,A^T h = As$$

where $s = s_\gamma(x_\gamma)$ and $W = W(x_\gamma)$ with $x_\gamma = x_{ls}$ for initialization. Let $x_{\gamma-\delta} \equiv x_\gamma + \delta h$ and $r(\gamma-\delta) \equiv r(x_\gamma) + \delta Ah$. The algorithm finds the smallest of $\delta > 0$ where one of the components of $r(\gamma-\delta)$ changes status, i.e., where

$|r_j(\gamma-\delta)| = \gamma-\delta$ for some j, $1 \le j \le m$. More precisely, let $\{\delta_i\}$ $i = 1, \ldots, K$, with $\delta_1 < \delta_2 < \cdots < \delta_K$, be the set of points in $(0, \gamma)$ where $|r_j(\gamma-\delta)| = \gamma-\delta$ for some j. Then γ is replaced with $\gamma - \delta_1$, x_γ is replaced with $x_\gamma + \delta_1 h$, s is updated as $s_\gamma(x_\gamma)$, and the loop is repeated.

We summarize the steps of this algorithm below.

find x_{ls} from $AA^T x = Ab$

choose $\gamma = \max_{j=1,\ldots,m} \| r_j(x_{ls})|$

  find $s = s_\gamma(x_\gamma)$ and $W = W_s$

repeat

  compute h from $AW\ A^T h = As$

  compute $d = Ah$

  compute $\delta_i^+ = \dfrac{\gamma - r_i(x_\gamma)}{1 + d_i}$ for $i \in \sigma(s) \cup \sigma_+(s)$

  compute $\delta_i^- = \dfrac{-\gamma - r_i(x_\gamma)}{-1 + d_i}$ for $i \in \sigma(s) \cup \sigma_-(s)$

    find $\delta = \min_i\{\delta_i^+, \delta_i^-\}$

    $x_\gamma \leftarrow x_\gamma + \delta h$

    $\gamma \leftarrow \gamma - \delta$

  find $s = s_\gamma(x_\gamma)$ and $W = W_s$

  until $\gamma = 0$

Notice that the algorithm stops with an ℓ_1 estimator (the unique ℓ_1 estimator) if $\delta = \gamma$ since the uniqueness of the M-estimator for sufficiently small $\gamma > 0$ implies the uniqueness of the ℓ_1 estimator; see [10].

Example 1 *Consider the example problem with* $r(x_1, x_2) = (3x_1 + 2x_2, 4x_1 - 4, 3x_2 - 3, 2x_1 + 3x_2 - 5, 7.5x_1 + 7x_2 - 20)^T$ *from* [6]. *The* ℓ_1 *estimator corresponding to this problem is* $(1, 1)^T$. *The least squares solution is* $x_{ls} = (1.0135, 13892)^T$ *with* $r(x_{ls}) = (5.8188, 0.0539, 1.1675, 1.1345, 0.2674]$. *We choose* $\gamma = 5.8188$ *and initialize* $s = (1, 0, 0, 0, 0)^T$. *We solve the system* (23) *which in this case gives*

$$\begin{pmatrix} 76.25 & 58.5 \\ 58.5 & 67 \end{pmatrix} \begin{pmatrix} h_1 \\ h_2 \end{pmatrix} = \begin{pmatrix} 3 \\ 2 \end{pmatrix}. \tag{27}$$

The unique solution is $h = (0.0498, -0.0136)^T$. *We find* $\delta = 4.3551$. *Therefore,* $\gamma \leftarrow \gamma - \delta = 1.4637$ *with* $x = x_{ls} + \delta h = (1.2304, 1.3298)^T$ *and the corresponding residual vector* $r = r(x_{ls}) + \delta d = (6.3507, 0.9216, 0.9893, -1.4501, -1.4637)^T$. *Notice that the optimal extended sign vector is* $(1, 0, 0, 0, 0)^T$ *in the sign interval*

$[1.4637, 5.8188]$. *Now, we update* s *to become* $s = (1, 0, 0, 0, -1)^T$. *We solve the linear system* (23) *again*:

$$\begin{pmatrix} 20 & 6 \\ 6 & 18 \end{pmatrix} \begin{pmatrix} h_1 \\ h_2 \end{pmatrix} = \begin{pmatrix} -4.5 \\ -5 \end{pmatrix}. \tag{28}$$

The solution is $h = (-0.1574, -0.2253)^T$. *We compute* $\delta = 1.4637$. *Since* $\delta = \gamma$, *the algorithm stops with* $x \leftarrow x + \gamma h = (1, 1)^T$ *as the* ℓ_1 *estimator.*

The finiteness of this algorithm depends on the following property proved by Clark and Osborne:

Theorem 2 *If* s_γ *is the optimal extended sign vector for* $\gamma > \bar{\gamma}$ *but fails to be optimal for* $\gamma < \bar{\gamma}$, *the difference being caused by the size of a single residual* r_k, *then the sign vector* s'_γ *with* $\sigma(s') = \sigma(s) \setminus \{k\}$ *or* $\sigma(s') = \sigma(s) \cup \{k\}$ *is optimal for some* $\gamma < \bar{\gamma}$.

This gives the algorithm a look-ahead ability in that at the change of intervals the algorithm knows what the optimal sign vector will be in the next interval. Now, since the algorithm moves from one optimal sign vector to another (the adjacent one) while decreasing γ (c.f. Theorem 2), and since x_γ is a piecewise linear function of γ under the absence of degeneracy (c.f. Theorem 2. 6 of [7]), the algorithm never repeats an optimal extended sign vector. As the number of distinct sign vectors is finite, the algorithm terminates finitely.

However, when the difference alluded to in the theorem above is caused by more than one residual we are no longer sure of the optimal extended sign vector in the next interval. To overcome this difficulty, Clark and Osborne propose a finite partitioning algorithm to find the M-estimator for a slightly smaller γ value than the current one and continue the algorithm from this point. However, the expression "slightly smaller value" is numerically ill-defined, and Clark and Osborne do not incorporate this finite partitioning algorithm into their implementations.

An important feature of the Clark-Osborne algorithm is the update of a suitable factorization of the symmetric, positive definite matrix AWA^T at the change of sign intervals. Since there is only one entry that changes in the matrix W at a change of interval and that the matrix always retains its positive definiteness, the factorization can be updated in a stable and efficient way by means of orthogonal transformations; see [7] for details.

4 The Madsen-Nielsen Algorithm

The Madsen-Nielsen algorithm is essentially an extension of the Clark-Osborne algorithm. The main difference between the two is that the Madsen-Nielsen algorithm does not require a unique path of M-estimators and does not stay on the path(s) of M-estimators. Although no analytical result is available to support the superiority of the Madsen-Nielsen algorithm over the Clark-Osborne algorithm, the former was shown experimentally to be significantly faster than the well-known Barrodale-Roberts simplex ℓ_1 algorithm. No such experimental result is available for the Clark-Osborne algorithm to date.

It can be easily shown using the results of Section 2.3 that the M-estimators form a family of piecewise linear paths. The algorithm then consists of the following steps. First, an M-estimator for some initial value of γ is computed. This is done using a finite, modified Newton algorithm earlier proposed by Madsen and Nielsen [11]. Then, using a solution to (23) the paths of M-estimators for decreasing values of γ are explored. However, unlike the Clark-Osborne algorithm the Madsen-Nielsen algorithm never moves to a point where there is a change of sign vectors. Instead, the algorithm allows a larger reduction in γ than the nearest end point of a sign interval. With the new value of γ, the modified Newton algorithm is invoked using a projected initial guess at the M-estimator for the new value of γ. This is repeated until suitable termination criteria are satisfied.

Notice that the most critical departure from the Clark-Osborne continuation scheme is that the Madsen-Nielsen algorithm leaves the paths of M-estimators to return to them later.

The basic algorithm can be formulated as follows:

choose initial γ

repeat

compute an M-estimator x_γ

decrease γ

until $\gamma = 0$

The algorithm has three main components: (1) stopping criterion, (2) computation of an M-estimator, (3) decreasing γ. We study these components in the above order.

4.1 Stopping Criteria

The original Madsen-Nielsen algorithm in [12] used the same stopping criteria as the Clark-Osborne algorithm. Later Madsen, Nielsen and Pınar in [15] use different stopping criteria which consist of checking the duality gap and complementarity as follows. Let $x_\gamma \in M_\gamma$ for some $\gamma > 0$ with $s = s_\gamma(x_\gamma)$ and $y_\gamma = \frac{1}{\gamma} W_s r(x_\gamma)$. Let h be a solution to the system $AW_sA^T h = As$. Let $x_0 = x_\gamma + \gamma h$. The algorithm stops with output x_0 if

$$G(x_0) + b^T y_\gamma = 0, \tag{29}$$

and

$$s_i r_i(x_0) \geq 0, \ \forall\, i \in \sigma_+(s) \cup \sigma_-(s). \tag{30}$$

Clearly, x_0 and y_γ that satisfy these criteria are optimal solutions to [L1] and [D1], respectively as these criteria are equivalent to the optimality condition (6) in the ℓ_1 estimation problem.

The problem with both termination criteria is that there is nothing that guarantees that an arbitrary solution h to (23) satisfies these conditions. Theorem 1 guarantees the existence of such a solution h for sufficiently small $\gamma > 0$ under the condition that we use the sign vector definition (4) to compute s. However, no information is conveyed in this theorem as to which solution to (23) to compute. In the special case where the M-estimator is unique and AW_sA^T has full rank (A needs to have full rank for this to hold) then the above stopping criteria lead to a finite termination argument. For implementation, one usually computes a basic solution or a least-norm solution of (23). But, there is no analytical result to justify such choices.

4.2 Computing an M-estimator

The Newton method of Madsen and Nielsen [11] is a modified Newton method with a line search procedure. We will refer to this algorithm as the MN algorithm for convenience.

The MN algorithm consists of inspecting the domains $\mathscr{C}_s^\gamma$ to find the quadratic representation of G_γ where the global minimizer is located. A search direction h is computed by minimizing the quadratic $Q_s(x)$ where s is the sign vector of the current iterate. More precisely, let x be the current iterate and $s = s(x)$ and $W = W(x)$, we consider the system of equations

$$Q_s''h = -Q_s'(x). \tag{31}$$

This system is expressed as

$$(AW\ A^T)h = -A[Wr(x) + \gamma s]. \tag{32}$$

Clearly, $x + h$ minimizes the quadratic Q_s for any h that solves (32). For ease of notation let $C \equiv AW\ A^T$. Furthermore, let $\mathscr{N}(C)$ denote the null space of C. If C has full rank, then h is the unique solution to (32). The algorithm checks whether $x + h \in \mathscr{C}_s^\gamma$. If the answer is affirmative, the algorithm stops with $x + h$ as the minimizer of G_γ. Otherwise, it proceeds with a piecewise linear one-dimensional search along h. If the system of equations (32) is consistent, a minimum norm solution is computed. The algorithm checks whether $x + h \in \mathscr{C}_s^\gamma$ and stops with $x + h$ as the minimizer if the answer is affirmative. Otherwise, the next iterate is found by moving to the first kinkpoint α_1 along h, i.e., the smallest value of α where $s_\gamma(x + \alpha) \neq s_\gamma(x)$. Notice that if h is the least norm solution of (32) the the point $x + h$ is the orthogonal projection of x onto the set of minimizers of the quadratic Q_s.

If the system is inconsistent a suitable descent direction h is computed and a piecewise linear one-dimensional search along h is performed. Madsen and

Nielsen showed that under the full rank assumption on A the iteration is finite, i.e., after a finite number of iterations we have $x + h \in \mathscr{C}_s^\gamma$ and therefore, $x + h$ is a minimizer of G_γ.

Recently, Chen and Pınar [5] proposed a modification of this algorithm and proved finite termination without the full rank assumption on A. The modified algorithm allows any solution of the system (32) to be used as a descent direction as long as its norm is bounded by a constant times the norm of the minimum norm solution h_m while the original algorithm is restricted to the use of a least norm solution in the consistent case. Furthermore, in this case, the original algorithm moved to the first kink point along the search direction whereas the modified algorithm prescribes a line search along this direction. With these computational enhancements Chen and Pınar proved that the modified MN algorithm stops at an M-estimator after a finite number of iterations. The proof of this result is quite involved. Therefore, the interested reader is referred to [5] for details.

4.3 Reduction of γ

Assume $\gamma \notin (0, \gamma_0]$ as defined in Theorem 1. Let x_γ be an M-estimator corresponding to the present value of γ. Let $x_{\gamma-\delta} \equiv x_\gamma + \delta h$ and $r(\gamma - \delta) \equiv r(x_\gamma) + \delta Ah$. The algorithm finds the smallest of $\delta > 0$ where one of the components of $r(\gamma - \delta)$ changes status, i.e., where $|r_j(\gamma - \delta)| = \gamma - \delta$ for some j, $1 \leq j \leq m$. More precisely, let $\{\delta_i\}$ $i = 1, \ldots, K$, with $\delta_1 < \delta_2 < \cdots < \delta_K$, be the set of points in $(0, \gamma)$ where $|r_j(\gamma - \delta)| = \gamma - \delta$ for some j. Then γ is replaced with $\gamma - \delta$ where $\delta > \delta_1$, x is replaced with $x_\gamma + \delta h$, s is updated as $s_\gamma(x)$, and the modified Newton algorithm is invoked with x as the starting point.

Note that there is some flexibility involved in the choice of δ in the reduction strategy as long as a change of interval is assured. Madsen and Nielsen[12] describe a strategy based on inspecting the points of interval change $\{\delta_i\}$ as in the Clark-Osborne and picking δ according to some heuristic criteria. Another heuristic method is described in Madsen, Nielsen and Pınar [15]. The important point here is to find a good heuristic that decreases γ neither too fast nor too slowly. This is usually problem dependent, but the two heuristics mentioned above seem to give good average performances.

As in the Clark-Osborne algorithm, the efficiency of the Madsen-Nielsen algorithm strongly depends on the efficient solution of linear systems (23) and (32). Both these systems involve the same symmetric, positive (semi) definite matrix $AW\,A^T$. However, the modified Newton algorithm may allow more than one index to change its sign unlike the Clark-Osborne case. Nielsen [18] describes a software package for updating LDL^T factors of $AW\,A^T$ in a stable and efficient way within the modified Newton (MN) algorithm. When the M-estimator has been computed using the MN algorithm, the system (23) is solved to check optimality and reduce γ. Since the factors of $AW\,A^T$ from the last MN iteration are available, no update or refactorization is needed at that stage.

5 Further Results

In this section, we give some further results that are useful in the analysis of the extension of the Madsen-Nielsen algorithm. We use $S(M_\gamma)$ to denote the set of all distinct extended sign vectors corresponding to the elements of M_γ. That is, for any $x_\gamma \in M_\gamma$ $s_\gamma(x_\gamma) \in S(M_\gamma)$.

The following result is a consequence of the linearity of the problem.

Lemma 2 *If* $S(M_{\gamma_1}) = S(M_{\gamma_2})$ *where* $0 < \gamma_2 < \gamma_1$ *then* $S(M_\gamma) = S(M_{\gamma_1}) = S(M_{\gamma_2})$ *for* $\gamma_2 \le \gamma \le \gamma_1$.

Theorem 3 *There exists* $\bar{\gamma}$ *such that* $S(M_\gamma)$ *are constant for* $\gamma \in (0, \bar{\gamma})$ *where* $0 < \bar{\gamma} \le \gamma_0$.

Proof: Since $s^\gamma(M_\gamma)$ remains constant in $(0, \gamma_0]$ following Theorem 1 and the number of different extended sign vectors is finite, the result is a consequence of the previous lemma. □

The above result indicates that when γ is sufficiently small, the boundaries of the set of M-estimators also remain constant. In other words, the set of extended sign vectors corresponding to M-estimators remain constant. This property allows us to prove the following important result.

Theorem 4 *Let* $\gamma \in (0, \bar{\gamma})$ *and* $x_\gamma \in M_\gamma$ *with* $s = s_\gamma(x_\gamma)$ *and* $W = W_s$. *Then*

$$Wr(x_\gamma + \gamma h) = 0 \tag{33}$$

for any solution h *to* (35). *Furthermore, if*

$$s_i r_i(x_\gamma + \gamma h) \ge 0, \ \forall\, i \in \sigma_+(s) \cup \sigma_-(s) \tag{34}$$

then $x_\gamma + \gamma h$ *solves [L1].*

Proof: Let $\gamma \in (0, \bar{\gamma})$ and $x_\gamma \in M_\gamma$ with $s = s_\gamma(x_\gamma)$ and $W = W_s$. Consider the system

$$(AW\,A^T)h = As. \tag{35}$$

This is a consistent system of linear equations as we have shown in Section 2.3. By Theorem 3 there exists $x_\gamma \in M_\gamma$ such that $s_\gamma(x_\gamma) = s$ for all $\gamma \in (0, \bar{\gamma})$. This implies that there exists h that solves (35) such that $x_\gamma + \delta h \in M_{\gamma-\delta}$ for all $\delta \in (0, \gamma]$. A consequence of this using the continuity of r and (5) is that $x_\gamma + \gamma h$ solves [L1], and $Wr(x_\gamma + \gamma h) = 0$. Since h can be replaced by $h + \eta$ in the above identity where $\eta \in \mathcal{N}(AW\,A^T)$, it follows that

$$Wr(x_\gamma + \gamma h) = 0. \tag{36}$$

Now, define $y_\gamma = \frac{1}{\gamma} Wr(x_\gamma) + s$. It is easy to verify that if (34) holds, $x_\gamma + \gamma h$ and y_γ satisfy the complementarity condition. Since y_γ is feasible for [D1], this implies that $x_\gamma + \gamma h$ is an ℓ_1 estimator. □

The theorem says that the "small" residuals in the sense of definition (26) are approaching zero as γ approaches zero using any solution to (35) provided γ is sufficiently small.

Under a certain regularity assumption on the ℓ_1 problem it is possible to relate the magnitude of $\bar{\gamma}$ to the nonzero optimal residuals magnitudes of the ℓ_1 solution.

Theorem 5 *Let x be an ℓ_1 estimator with $s = s(x)$. If for some solution h to the system*

$$AW_sA^Th = As \tag{37}$$

we have

$$\|W_sA^Th\|_\infty \leq 1, \tag{38}$$

then, there exists $x_\gamma \in M_\gamma$ with $s_\gamma(x_\gamma) = s$ for all $\gamma \in (0, \xi)$ where $\xi \leq \min\{|r_i(x)| : i \in \sigma_-(s) \cup \sigma_+(s)\}$.

Proof: Let $s = s(x)$ and $\delta = \min\{|r_i(x)| : i \in \sigma_-(s) \cup \sigma_+(s)\}$. The linear system (37) is consistent following (6). By the regularity assumption we have $\|WA^Th\| \leq 1$ for any solution h to the system. Choose $0 < \xi \leq \delta$ so that for all $0 < \gamma \leq \xi$,

$$r_i(x) - \gamma(A^Th)_i > \xi, \ \ i \in \sigma_+(s), \tag{39}$$

$$r_i(x) - \gamma(A^Th)_i < -\xi, \ \ i \in \sigma_-(s). \tag{40}$$

Now using (37) and the fact that $W_s(A^Tx - b) = 0$ we have:

$$\begin{aligned} 0 =& AW_sA^T(-\gamma h) + \gamma As \\ =& AW_s(A^T(x - \gamma h) - b) + \gamma As. \end{aligned}$$

Since $\|W_sA^Th\|_\infty \leq 1$, using (39) and (40) we have $s_\gamma(x - \gamma h) = s$. Hence, $x - \gamma h \in M_\gamma$. □

Following this theorem, we can expect to decrease γ to the level of the smallest nonzero optimal residual(s) to enter the final sign interval of M-estimators.

6 An Extension of the Madsen-Nielsen Algorithm

Before going into the details of the algorithm to be proposed below, it is instructive to examine how the theory developed in Section 5 motivates the algorithm.

Notice that for γ sufficiently small ($\gamma \in (0, \bar{\gamma})$) the point $x_\gamma + \gamma h$ gives $W(r(x_\gamma + \gamma h) = 0$ regardless of the choice of h. Hence, if $x_\gamma + \gamma h$ is complementary to $y_\gamma \equiv \frac{1}{\gamma}Wr(x_\gamma) + s$, then $(y_\gamma, x_\gamma + \gamma h)$ is clearly a primal-dual optimal pair. If

$W(r(x_\gamma + \gamma h) = 0$ but $x_\gamma + \gamma h$ and y_γ are not complementary, we move to the smallest positive point along h where a change of sign occurs. If $\gamma \in (0, \bar{\gamma})$ this leads to an expansion of the active set. Continuing this way, the algorithm stops in a finite calculation. If $W(r(x_\gamma + \gamma h) \neq 0$, we reduce γ exactly as in Madsen and Nielsen [12] or, as in [15]. As far as the finite termination arguments are concerned it suffices that γ is replaced by $\beta\gamma$ where $\beta \in (0, 1)$ in this case.

More precisely, we propose the following algorithm:

Choose γ and compute a minimizer x_γ of G_γ(call MN)
while not STOP
  find $s = s_\gamma(x_\gamma)$ and $W = W_s$
  compute h from $AW\,A^T h = As$
  if $Wr(x_\gamma + \gamma h) = 0$ **then**
    compute $d = Ah$
    compute $\delta_i^+ = \dfrac{\gamma - r_i(x_\gamma)}{1 + d_i}$ for $i \in \sigma_+(s)$
    compute $\delta_i^- = \dfrac{-\gamma - r_i(x_\gamma)}{d_i - 1}$ for $i \in \sigma_-(s)$
    find $\delta = \min_i\{\delta_i^+, \delta_i^-\}$
    $x_\gamma \leftarrow x_\gamma + \delta h$
    $\gamma \leftarrow \gamma - \delta$
  else
    reduce $\gamma(\gamma \leftarrow \beta\gamma)$
    $x \leftarrow x_\gamma + (1 - \beta)\gamma h$
    compute a minimizer x_γ of $G_\gamma$starting from x (call MN)
  endif
end while.

In the above iteration STOP is a function that returns TRUE if

$$s_i r_i(x_\gamma + \gamma h) \geq 0,\ \forall\, i \in \sigma_+(s) \cup \sigma_-(s). \tag{41}$$

Notice that when the condition $Wr(x_\gamma + \gamma h) = 0$ holds the new algorithm uses a strategy similar to the Clark-Osborne algorithm. However, no restriction about uniqueness of x_γ nor non-singularity of the matrix $AW\,A^T$ is required. Furthermore, we do not impose any requirements on the solution h of $AW\,A^T h = As$ as far as the proof of finite termination is concerned.

Example 2 *Consider the following example problem from* [10]. *We have* $r(x_1, x_2) = (x_1 + 8x_2, x_1 - 8x_2, 2x_2, 17x_2 - 1)^T$. *The* ℓ_1 *estimator corresponding to this problem is unique*, $x = (0, 0)^T$. *Interestingly*, *for* $\gamma \in (0, 4/21]$, *the Huber* M*-estimator is not unique*. *Let* $\gamma = 3/21$. *It can be verified easily that*

$x_\gamma = (1/10, 3/84)^T$ is an M-estimator for $\gamma = 3/21$, with $r(x_\gamma) = (0.3857, -0.1857, 0.0714, -0.3929)^T$. This corresponds to $s_\gamma(x_\gamma) = (1, -1, 0, -1)^T$. We solve the system (23) *which is in this case*:

$$\begin{pmatrix} 0 & 0 \\ 0 & 4 \end{pmatrix} \begin{pmatrix} h_1 \\ h_2 \end{pmatrix} = \begin{pmatrix} 0 \\ -1 \end{pmatrix}. \tag{42}$$

The least norm solution of this system is $h = (0, -1/4)^T$. This gives $d = A^T h = (-2, 2, -0.5, -4.25)^T$. The point $r(x_\gamma) + \gamma d$ gives $W(r(x_\gamma) + \gamma d) = 0$ but does not satisfy complementarity criterion (41). *Evaluating the kink points $\{\delta_i^+\}$ and $\{\delta_i^-\}$ we find $\delta = 0.0429$. Hence, $\gamma \leftarrow \gamma - \delta = 0.1$, the algorithm moves to $x = x_\gamma + \delta h = (0.1, 0.025)^T$ with $r = r(x_\gamma) + \delta d = (0.3, -0.1, 0.05, -0.575)^T$. The extended sign vector associated with this point* (x *is the Huber M-estimator for $\gamma = 0.1$*) *is $(1, 0, 0, -1)^T$. We form* (23) *again*:

$$\begin{pmatrix} 1 & -8 \\ -8 & 68 \end{pmatrix} \begin{pmatrix} h_1 \\ h_2 \end{pmatrix} = \begin{pmatrix} 1 \\ -9 \end{pmatrix} \tag{43}$$

with the unique solution $h = (-1, -0.25)$ and $d = A^T h = (-3, 1, -0.5, -4.25)^T$. This time, the point $x \leftarrow x + \gamma h = (0, 0)^T$ yields $r \leftarrow r + \gamma d = (0, 0, 0, -1)$, which satisfies the termination criteria. Thus, the algorithm stops with the unique ℓ_1 estimator in two iterations starting from the Huber M-estimator at $\gamma = 3/21$.

6.1 Finite Convergence

In this section we show that the algorithm of Section 6 converges finitely.

Lemma 3 *Assume $\gamma \in (0, \bar{\gamma})$. Let $x \in M_\gamma$* with $s = s_\gamma(x)$. *Let h solve* (35), *and x_{next} be generated by one iteration of the algorithm. Then either*

$$x_{next} \equiv x + \gamma h \in \mathscr{S}$$

and the algorithm stops, *or*

$$x_{next} \equiv x + \delta h \in M_{\gamma_{next}},$$

$$\gamma_{next} = \gamma - \delta$$

where δ is as defined in the algorithm, and $\mathscr{A}_{\gamma_{next}}(x_{next})$ is an extension of $\mathscr{A}_\gamma(x)$.

Proof: Let $y = \frac{1}{\gamma} Wr(x) + s$. Clearly $Wr(x + \gamma h) = 0$ from Theorem 4. If $x + \gamma h$ and y are complementary then $x_{next} \equiv x + \gamma h$ is a solution to [L1] by Theorem 4 and the algorithm stops. Otherwise, Theorem 4 implies that $\mathscr{A}_\gamma(x) \subseteq \mathscr{A}_0(x + \gamma h)$. Hence, using the definition of δ,

$$\mathscr{A}_{\gamma-\alpha}(x + \alpha h) = \mathscr{A}_\gamma(x)$$

for $\alpha \in [0, \delta)$. Since there exists $j \in \{1, \dots, m\} \setminus \mathscr{A}_\gamma(x)$ such that $|r_j(x + \delta h)| = \gamma - \delta$, $\mathscr{A}_{\gamma-\alpha}(x + \delta h)$ is an extension of $\mathscr{A}_\gamma(x)$. Furthermore $x + \delta d \in \mathscr{C}_s^{\gamma-\delta}$. Therefore, using the continuity of the gradient G'_γ, (18) and the definition of h, we have

$$G'_\gamma(x) = G'_{\gamma-\delta}(x + \delta h) = 0.$$

Thus, x_{next} minimizes $G_{\gamma-\delta}$. □

Theorem 6 *The algorithm defined in Section* 6 *terminates in a finite number of iterations with an* ℓ_1 *estimator.*

Proof: Let $x \in M_\gamma$ for some $\gamma > 0$. Unless the stopping criteria are met and the algorithm stops with a primal-dual optimal pair, γ is reduced by a nonzero factor. Since the modified Newton iteration of Section 4.2 is a finite process, γ enters the range $(0, \bar{\gamma})$ where $\bar{\gamma}$ is as defined in Theorem 3 in a finite number of iterations unless the algorithm stops. Now assume $\gamma \in (0, \bar{\gamma})$. From Lemma 3 either the algorithm terminates or the γ-active set $\mathscr{A}_\gamma$ is expanded. Repeating this argument, the algorithm should stop with an ℓ_1 estimator since the γ-active set has finite cardinality. □

6.2 Computational Behavior

A software system that implements the original algorithm of Madsen, Nielsen and Pınar, called LPASL1, was developed in [14], and later modified by the present author to include the changes proposed above. In preliminary tests, it was found that the additional precautions proposed above for finite convergence did not cause a discernible slowdown of the algorithm. Recently, while the present paper was under review, Shi and Lukas [20] introduced a new reduced gradient type algorithm for the ℓ_1 estimation problem, and reported extensive comparative computational results with the most important ℓ_1 codes available in the public domain and our modification of LPASL1. These include the algorithm ACM551 of [1], ACM552 of [3], the algorithm AFK of [2] which are considered to be the fastest ℓ_1 codes available. While Shi and Lukas' new reduced gradient algorithm turns out to be the fastest in a wide range of computational tests with randomly generated over-determined linear systems with up to 2430 equations and 1215 unknowns, modified LPASL1 is quite competitive with the aforementioned well-established codes. We give a brief summary of the cpu time results in Table 1 where 10 instances were solved for each size, and average cpu seconds reported. Shi and Lukas [20] indicate that the problems are degenerate. The sign $-$ indicates that AFK was not able to solve all 10 problems. However, modified LPASL1 was not able to solve successfully all 10 of the largest 2430×1215 instances, due to numerical difficulties. We believe that the reasons for this behavior are related to the degenerate nature of some test problems because for non-degenerate 2430×1215 instances, LPASL1 successfully obtained an optimal solution in competitive CPU times; see [20]. This point is to be examined in more detail in further research.

Table 1 Summary Computational Results

Size	ACM551	ACM552	AFK	modified LPASL1
480×240	13.4	4.79	209	0.69
720×360	60.4	18.4	1193	2.39
1080×540	287	87.1	–	9.06
1620×810	1299	340	–	30.2

7 Conclusions

We studied the finite computation of the ℓ_1 estimator from Huber's M-estimator. We have reviewed and extended the contributions of Clark and Osborne, and Madsen and Nielsen to give a new finite algorithm to compute the ℓ_1 estimator from the Huber M-estimator. The new method has guaranteed finite termination property without any restrictive assumptions. In particular, we removed the assumption of full rank on the matrix A, an assumption which is also present in the recent paper by Li and Swetits [10]. This reference also gives a recursive algorithm of a somewhat different nature than the algorithms of the present paper to compute the ℓ_1 estimator from the M-estimator based on the least norm solution of the dual linear program [D1] although no computational experience or any evidence to the efficiency of this algorithm is reported. We also summarized promising results of comparative computational tests obtained with the modified Madsen-Nielsen algorithm.

Acknowledgement

The author is indebted to K. Madsen, H.B., Nielsen and B., Chen for collaboration and many useful discussions.

References

[1] Abdelmalek, N. N.: Algorithm 551, A Fortran subroutine for the L_1 solution of overdetermined systems of linear equations. ACM Trans. Math. Software *6*, 229–230 (1980).

[2] Armstrong, R. D., Frome, E. L., Kung, D. S.: A revised simplex algorithm for the absolute deviation curve fitting problem. Comm. Statist. B *8*, 175–190 (1979).

[3] Barrodale, I., Roberts, F.: An improved algorithm for discrete L_1 approximation. SIAM J. Numer. Anal. *10*, 839–848 (1972).

[4] Bloomfield, P., Steiger, W. L.: Least Absolute Deviations: Theory, Applications and Algorithms, Boston: Birkäuser (1983).

[5] Chen, B., Pınar, M.Ç.: On Newton's method for Huber M-estimation problems in robust linear regression. BIT *38*, 674–684 (1998).

[6] Clark., D.: The mathematical structure of Huber's M-estimator. SIAM J. on Scientific and Statistical Computing *6*, 209–219 (1985).

[7] Clark, D.I., Osborne, M.R.: Finite algorithms for Huber's M-estimator. SIAM J. on Scientific and Statistical Computing *7*, 72–85 (1986).

[8] Coleman, T., Li, Y.: A globally and quadratically convergent algorithm for the linear ℓ_1 problem. Math. Prog. *56*, 189–222 (1992).

[9] Huber, P.J.: Robust Statistics. New York, Wiley (1981).

[10] Li, W., Swetits, J. J.: Linear ℓ_1 estimator and Huber M-estimator. SIAM J. on Optimization *8*, 457–475 1997.

[11] Madsen, K., Nielsen, H. B.: Finite algorithms for robust linear regression. BIT *30*, 682–699 (1990).

[12] Madsen, K., Nielsen, H. B.: A finite smoothing algorithm for linear ℓ_1 estimation. SIAM J. on Optimization *3*, 223–235 (1993).
[13] Madsen, K., Nielsen, H. B., Pınar, M. Ç.: New characterizations of ℓ_1 solutions of overdetermined systems of linear equations. Oper. Res. Letters *16*, 159–166 (1994).
[14] Madsen, K., Nielsen, H. B., Pınar, M. Ç.: User's guide to LPASL1: A Fortran 77 package, based on a continuation algorithm for dense linear programming. Lyngby, DK, Technical University of Denmark 1994.
[15] Madsen, K., Nielsen, H. B., Pınar, M. Ç.: A new finite continuation algorithm for linear programming. SIAM J. on Optim. *6*, 600–616 (1996).
[16] Mangasarian, O.L., Meyer, R. R.: Nonlinear perturbations of linear programs. SIAM J. on Control Optim. *25*, 583–595 (1979).
[17] Michelot, C., Bougeard, M. L.: Duality results and proximal solutions of the Huber *M*-estimator problem. Appl. Math. Optim. *30*, 203–221 (1994).
[18] Nielsen, H. B.: AAFAC: A Fortran 77 pacakge for solving $AA^Tx - c$, Technical Report NI-90-11. Lyngby, DK Technical University of Denmark, 1990.
[19] Ruzinski, S. A., Olsen, E. T.: L_1 and L_∞ minimization via a variant of Karmarkar's algorithm, IEEE Trans. on Acous. Speech and Sig. Proc. *37*, 245–253 (1989).
[20] Shi, M., Lukas, M. A.: An L_1 estimation algorithm with degeneracy and linear constraints. Computational Statistics and Data Analysis *39*, 35–55 (2002).
[21] Watson., G. A.: Approximation Theory and Numerical Methods. John Wiley & Sons, New York 1980.
[22] Zhang, Y.: Primal-dual interior point approach to computing L_1 and L_∞ solutions of overdetermined linear systems. J. Optim. Theory Appl. *77*, 323–341 (1993).

Mustafa Ç. Pınar
Department of Industrial Engineering
Bilkent University
06533 Ankara, Turkey
e-mail: mustafap@bilkent.edu.tr

Offprint Order

Journal: Computing MS No.: 1503 First Author: **M. . Pınar**

50 **Offprints** of each paper are supplied **free of cost**; the order of additional offprints against payment must be sent in when returning the corrected proofs

The prices quoted are valid only for authors ordering offprints for their private use.

Please write clearly in capital letters!

NEW **When you order offprints against payment, you are entitled to receive in addition a pdf file of your article for your own personal use.** As this pdf file is sent to you by e-mail, please insert the e-mail address here:

In addition to the free offprints, I hereby order against payment (in quantities of 100 only)

☐ **100** ☐ **200** ☐ **300** ☐ **400** **offprints**

Offprints should be sent to:

(Name of person or institution)

(Address)

Payment will be made by:

(Name of person or institution)

(Address)

(Purchase Order No.) ______________ (Date/Signature of author) ______________

☐ Please bill me **(please do not pay for offprints before receipt of invoice!)**

☐ Please charge my credit card ☐ Eurocard / Mastercard ☐ American Express

☐ Visa ☐ Diners Club

No.: ________________ Valid until: ___/___

Signature: ______________

(In all separate correspondence concerning this order please quote the Journal's title, MS No., and First Author.)

Price list for offprints* Prices include carriage charges (surface mail). Prices are subject to change without notice.

***Customers in foreign EU countries:** Please state your V.A.T registration number if applicable. Otherwise we have to add 10% V.A.T. to the list prices.

V.A.T. registration number: ______________

Pages (Figs. incl./excl.)	100 copies	200 copies	300 copies	400 copies
	EUR	EUR	EUR	EUR
☐ 1–4	158,–	284,–	402,–	523,–
☐ 5–8	285,–	492,–	709,–	925,–
☐ 9–12	350,–	622,–	894,–	1163,–
☐ 13–16	433,–	792,–	1145,–	1497,–
☐ 17–24	592,–	1067,–	1544,–	2020,–
☐ 25–32	711,–	1278,–	1844,–	2407,–

Please return your order form to:
Springer-Verlag Wien, Production Department, **Sachsenplatz 4-6, P.O. Box 89, A-1201 Wien, Austria**

Copyright Transfer Statement

(Author's signature)

Instruction to printer	Mark	Examples: In the text	Examples: In the margin
Character to be corrected	/	Latter to be corrected	e /
Group of characters to be corrected	H	Letters to be correcting	ed H
Several identical characters to be corrected	/	Cauncil ar Cammission	o ///
Differentiation of several errors in the same paragraph		There are many faults in this line	
Character or word to be deleted		Commission and and Parliament	
Character or word to be added		A word missing	is
Superior character required		The Court's judgment.	(1) /
Omitted text to be added (see copy)		1. January 12. December	Out see copy
Inferior character required		H_2SO	4 /
Change to italic		Ad infinitum	ital.
Change italic characters to roman		*status quo*	rom.
Change capitals to lower case		UNESCO	l.c.
Change to capitals or small capitals		Robert burns, AD 1759-96	Caps. s.c.
Change to bold face		This word needs emphasising!	bold
To be letter-spaced		[illegible]	/
Correct horizontal alignment		This line is crooked	/
Text to be raised or lowered		This is line uneven	/ /
Text to be aligned (to the left)		This text is to be aligned	/
Text to be aligned (to the right)		This text is to be aligned	/
Text to be centred		This text is to be centred	/
Take back to previous line		This hyph- en is unnecessary	/
Text to run on (no new paragraph)		... line. No new paragraph here	/
Take forward to next line		This hyp- hen is badly placed	/
Create new paragraph		... line. A new paragraph should begin here	/
Close up		A sp ace is wrong here	/
Equalise space	/	This spacing is very uneven	/
Add space between words		A spaceis missing here	# /
Reduce space between words		These spaces are too big!	/
Add space between lines	#	These lines are too close together #	#
Reduce space between lines		These lines are too far apart.	/
Stet (let original text stand)	.....	This text was corrected in error	
Transpose characters		These letters are transposed	/
Transpose words		These are words transposed	/
Transpose lines		transposed These lines are	/

NB: A correction made in the text must always have a corresponding mark in the margin, otherwise it may be overlooked when the corrections are made. The same marks should be used, where appropriate, by copy-editors marking up copy. Where instructional words are used in marginal marks, e.g. 'ital.', 'bold', etc., they must always be encircled to show that they are not to be printed.

33,3 % cheaper for you ...

As a Springer-author you are now entitled to receive a 33,3 % price reduction on the list price of **all books** published by Springer-Verlage Wien, Berlin/Heidelberg, New York, London, Paris, Barcelona, Tokyo and Hong Kong as well as Böhlau/Wien, Physica, Birkhäuser und Steinkopff. For your order please use this order form. Orders have to be sent directly to Springer-Verlag Vienna.

Als Autor/in des Springer-Verlages erhalten Sie 33,3 % Rabatt auf den Ladenpreis der **gesamten Buchproduktion** der Springer-Verlage Wien, Berlin/Heidelberg, New York, London, Paris, Barcelona, Tokyo und Hong Kong sowie der angeschlossenen Verlage Böhlau/Wien, Physica, Birkhäuser und Steinkopff. Bitte bestellen Sie mit diesem Bestellschein. Ihre Bestellung senden Sie bitte ausschließlich an den Springer-Verlag Wien.

For detailed information about titles published by Springer-Verlag please search our homepage.
Nähere Informationen über das Programm des Springer-Verlages finden Sie auf unserer Homepage.
http://www.springer.at

Order Form / Bestellschein

Springer-Verlag, Order Department, Sachsenplatz 4–6, P.O. Box 89, A-1201 Vienna, Fax +43.1.330 24 26
Springer-Verlag, Auslieferung, Sachsenplatz 4–6, Postfach 89, A-1201 Wien, Fax +43.1.330 24 26

I order herewith / Ich bestelle hiermit:

copy/ies Expl.	ISBN ISBN	Author Autor	Title Titel

Please copy this order form for your next orders. Bitte kopieren Sie diesen Bestellschein für Ihre weiteren Bestellungen.

o Please bill me / Bitte liefern Sie gegen Rechnung

o Please charge my credit card / Bitte belasten Sie meine Kreditkarte

o VISA o MASTERCARD o AMEX o DINERS

Card No. / Karten-Nr. .... Expiry date / Gültig bis ....

NAME / NAME ....

ADDRESS / ADRESSE ....

....

DATE / DATUM ....

SIGNATURE / UNTERSCHRIFT ....