ℓ_1 solution of linear inequalities

MUSTAFA Ç. PINAR

*Department of Industrial Engineering, Bilkent University,
06533 Ankara, Turkey*

AND

BINTONG CHEN

*Department of Management and Systems, Washington State University,
Pullman, WA 99164-4736, USA*

[Received 24 September 1996 and in revised form 23 December 1997]

The numerical solution of a possibly inconsistent system of linear inequalities in the ℓ_1 sense is considered. The non-differentiable ℓ_1 norm minimization problem is approximated by a piecewise quadratic Huber smooth function. A continuation algorithm is designed to find an ℓ_1 solution of the inequality system. In the case where the linear inequality system is consistent, a solution is obtained by solving any smoothed problem. Otherwise, the algorithm is shown to terminate in a finite number of iterations. We also consider an alternative smoothing scheme which shares similar properties with the first one, but results in an improved computational performance of the continuation algorithm on inconsistent systems. Numerical experiments are conducted to test the efficiency of the algorithm.

1. Introduction

Consider a system of linear inequalities:

$$Ax \leq b, \quad (1)$$

where $A \in \Re^{m \times n}$ with $m \geq n$, $b \in \Re^m$, and $x \in \Re^n$. Similar to a system of equations, the above inequality problem has many applications in data analysis, set separation problems, and image reconstructions. There is a vast literature on solving (1) by iterative projection methods. A recent survey and detailed treatment of these methods can be found in Censor & Zenios (1997).

In general, the inequality system (1) may be inconsistent. Therefore, it is natural to find a least norm solution—a solution with minimum violations to the inequality system measured in some norm. Define the residual vector $r(x)$ as

$$r(x) = Ax - b.$$

The least norm solution of (1) is to find a vector x that minimizes $\|r_+(x)\|$, where the plus function z_+ is a vector with i th component given by $\max\{0, z_i\}$. Past literature has mainly focused on the ℓ_2 norm solution of the problem (Han (1980), Bramley & Winnicka (1996), Pinar (1996)), partly because the objective function $\|r_+(x)\|_2$ is once differentiable. However, in many cases, the ℓ_1 norm solution may provide a more robust solution than the ℓ_2 norm solution (Bennett & Mangasarian (1992), Mangasarian (1984)). The objective of the paper is to design a continuation method to find the least ℓ_1 norm solution of the inequality system (1). The ℓ_1 problem can be formulated as the following non-differentiable optimization problem:

[P]

$$\min_x G(x) \equiv \|r_+(x)\|_1 = e^T r_+(x),$$

where e denotes a vector with all components unity. This problem is similar to the problem of computing ℓ_1 solutions of overdetermined linear systems of equations. Background on the ℓ_1 solution of overdetermined linear systems can be found in the book by Watson (1980). It is well known that **P** can be reformulated as a linear program, which in turn can be solved by any standard linear programming method such as the simplex method and various interior point methods. However, the reformulation is done by introducing additional m variables and thus increases the size of the original problem. It is also possible to solve the dual problem to **P**, which is again a linear program. In this paper we subscribe to the view that a method specially tailored for **P** may be a viable alternative. In Section 5, we provide some experimental evidence to support this view.

Our paper is partly motivated by the smoothing method recently proposed by Chen & Mangasarian (1995, 1996), where they propose a class of smoothing functions to approximate the plus function z_+ . The degree of approximation is controlled by a smoothing parameter and the smoothing function approaches the plus function as the parameter approaches zero. With the smoothed plus function, both ℓ_1 and ℓ_2 norm problems are turned into smooth optimization problems and thus can be solved by many traditional unconstrained optimization techniques. Chen and Mangasarian then showed that the solutions of the smooth problems are good approximate solutions to the inequality system when the smoothing parameter is sufficiently small.

In this paper, we choose to use a different smoothing function, known as the Huber function (to be described in detail in the next section) to smooth the plus function. We then design a continuation method to solve the ℓ_1 problem **P**. By exploring the special structure of the Huber function, we characterize the solution set of the smoothed problem and analyze its behaviour as the smoothing parameter approaches zero. The Huber smooth function, which is only once differentiable, is shown to have the following advantages over other continuously differentiable smooth functions, including the smooth function used in Chen & Mangasarian (1995):

1. The smoothed problem always has a solution, which is not true for general smooth functions. Indeed, the smoothed problem proposed in Chen & Mangasarian (1995) does not have a solution for such a simple inequality as $x \leq 0$.
2. If the inequality system (1) is consistent, then any solution of the smoothed problem is a solution of the problem (1) for any smoothing parameter. Hence, the solution of the first smoothed problem serves as a test for consistency of the linear inequality system.
3. The continuation method based on the Huber function is shown to converge finitely for any inequality system. For many other continuously differentiable smooth functions, however, this result is true only if the solution set has a nonempty interior point (see Chen & Mangasarian (1995)), which requires the inequality system to be at least consistent.

An alternative Huber function for smoothing problem **P** was suggested by an anonymous referee. We included this function into the paper and implemented the continuation method using this function as well. The function suffers from some serious drawbacks as discussed

in Section 2.2, and experimentally demonstrated on consistent systems of linear inequalities in Section 5. However, it yields a good performance for inconsistent linear inequality systems. The results of the paper carry over to this alternative Huber function, *mutatis mutandis*. Therefore, we point out some issues resulting from the use of the alternative Huber function, and do not reiterate all the results for this case.

The current paper is also partially based on earlier work on the ℓ_1 solution of an overdetermined system of linear equations using the Huber smooth function (Li & Swetits (1995), Madsen & Nielsen (1993), Madsen *et al* (1996), Madsen & Nielsen (1990), Madsen *et al* (1994)). However, the above papers work on a symmetric approximation whereas the approximation used in the present paper is unsymmetric. In addition, the finite convergence result is proved under less restrictive assumptions than the above papers. The algorithm of the present paper is also related to the least norm algorithm of Mangasarian (1984) where the ℓ_1 solution of a possibly inconsistent linear inequality system is considered using an SOR algorithm.

The paper is organized as follows. Section 2 introduces the smoothed $\mathbf{P}$ problem based on the Huber function together with the dual problems of both $\mathbf{P}$ and the smoothed problem. The relationship between the primal and dual problems and other related results are presented. Section 3 analyzes the behaviour of the solution set of the smoothed problem as the smoothing parameter approaches zero. Section 4 constructs a continuation method to solve $\mathbf{P}$ and the method is shown to converge finitely. Section 5 reports results of numerical experiments for the continuation method.

2. The structure of the Huber problem and duality

In this section we examine some structural properties of the Huber problem and important duality relations.

The Huber function with smoothing parameter $\gamma > 0$ is given by

$$p(z, \gamma) = \begin{cases} 0 & \text{if } z \leq 0 \\ \frac{1}{2\gamma} z^2 & \text{if } 0 < z < \gamma \\ z - \frac{\gamma}{2} & \text{if } z \geq \gamma. \end{cases}$$

Clearly, $p(z, \gamma)$ is once differentiable with respect to z and approaches z_+ as γ approaches 0. As a result, problem $\mathbf{P}$ can be approximated by the following smoothed problem, called the Huber problem:

[HP $^\gamma$]

$$\min_x G_\gamma(x) \equiv \sum_{i=1}^n p(r_i(x), \gamma). \quad (2)$$

Both $\mathbf{P}$ and $\mathbf{HP}^\gamma$ have well structured dual problems and their primal–dual relationship is instrumental to many of the results presented here. Following early work by Mangasarian (1984), we can derive the dual of $\mathbf{P}$ as the following linear program:

[D]

$$\begin{aligned} & \text{minimize} && b^T y \\ & \text{s.t.} && A^T y = 0 \\ & && 0 \leq y \leq e \end{aligned}$$

where $y \in \mathbb{R}^m$ is the corresponding dual variable. In addition, the primal and dual optimal solutions are related by the following complementary slackness conditions (see for instance Murty (1976)):

THEOREM 1 Let $x \in \mathbb{R}^n$ and $y \in \mathbb{R}^m$ be a feasible solution to **D**. Then x and y are optimal solutions of their respective problems if and only if the following conditions are satisfied:

$$\begin{aligned} y_i &= 0 && \text{if } r_i(x) < 0 \\ 0 \leq y_i \leq 1 && \text{if } r_i(x) = 0 \\ y_i &= 1 && \text{if } r_i(x) > 0 \end{aligned}$$

for all $i = 1, \dots, m$.

The Huber problem **HP** $^\gamma$ can also be formulated as the following quadratic program by introducing additional variables $u, v \in \mathbb{R}^m$.

[HPQ $^\gamma$]

$$\begin{aligned} & \text{minimize} && \frac{1}{2\gamma} u^T u + e^T v \\ & \text{s.t.} && Ax - b \leq u + v \\ & && u \leq \gamma e \\ & && v \geq 0. \end{aligned}$$

One can easily verify that x solves **HP** $^\gamma$ if and only if x solves **HPQ** $^\gamma$.

The dual of **HP** $^\gamma$ can be derived either directly following related papers (Li *et al* (1992), Li & Swetits (1995), Mangasarian (1984), Mangasarian & Meyer (1979), Michelot & Bougeard (1994)) or based on the above quadratic formulation **HPQ** $^\gamma$. It turns out to be a quadratically perturbed version of **D**:

[HD $^\gamma$]

$$\begin{aligned} & \text{minimize} && b^T y + \frac{1}{2}\gamma y^T y \\ & \text{s.t.} && A^T y = 0 \\ & && 0 \leq y \leq e. \end{aligned}$$

Clearly, **HD** $^\gamma$ is feasible and the objective function is quadratic, strictly convex, and bounded below. By the well known Frank–Wolfe theory (Frank & Wolfe (1956)), **HD** $^\gamma$ has a unique optimal solution. In addition, the following complementary slackness results hold (see Li & Swetits (1995) for a related result):

THEOREM 2 Let $x \in \mathbb{R}^n$ and $y \in \mathbb{R}^m$ be a feasible solution to **HD** $^\gamma$. Then x and y are

optimal solutions of their respective problems if and only if the following conditions are satisfied:

$$\begin{aligned} y_i &= 0 & \text{if } r_i(x) < 0 \\ y_i &= r_i(x)/\gamma & \text{if } 0 \leq r_i(x) \leq \gamma \\ y_i &= 1 & \text{if } r_i(x) > \gamma \end{aligned}$$

for all $i = 1, \dots, m$.

More interestingly, it has been shown (Mangasarian (1984), Mangasarian & Meyer (1979)) that the solution of the Huber dual $\mathbf{HD}^\gamma$ is constant for all γ sufficiently small.

THEOREM 3 There exists a $\gamma_0 > 0$ such that the solution y of $\mathbf{HD}^\gamma$ is constant for all $0 \leq \gamma \leq \gamma_0$. In particular, y is the least 2-norm solution of problem **D**.

2.1 Structural properties

To facilitate our presentation, we first define a ‘binary vector’ $s \in \mathfrak{R}^m$ such that

$$s_i(x) = \begin{cases} 0 & \text{if } r_i(x) < \gamma \\ 1 & \text{otherwise.} \end{cases} \quad (3)$$

Define also a diagonal matrix W as follows:

$$W_{ii}(x) = \begin{cases} 1 & \text{if } 0 < r_i(x) < \gamma \\ 0 & \text{otherwise,} \end{cases} \quad (4)$$

for all $i = 1, \dots, m$. That is, W is related to s as follows:

$$W_{ii}(x) = \begin{cases} s_i & \text{if } r_i(x) \leq 0 \\ 1 - s_i & \text{otherwise,} \end{cases} \quad (5)$$

for all $i = 1, \dots, m$. We refer to W as a ‘binary matrix’. Now, we can recast the problem **P** in the form:

$$\min_x G_\gamma(x) = \frac{1}{2\gamma} r^T(x) W(x) r(x) + s^T(x) [r(x) - \frac{1}{2}\gamma s(x)]. \quad (6)$$

We drop the argument x when the meaning is clear from the context. We denote by X the set of optimal solutions to **P**.

Using the notation s and W , we can restate Theorems 2 and 3 as follows.

THEOREM 4 Let $x \in \mathfrak{R}^n$. Let $y \in \mathfrak{R}^m$ be a feasible solution to $\mathbf{HD}^\gamma$. Then x and y are optimal solutions of their respective problems if and only if

$$y = \frac{1}{\gamma} W(x) r(x) + s(x). \quad (7)$$

In addition, there exists a $\gamma_0 > 0$ such that the unique solution y is constant for all $0 < \gamma \leq \gamma_0$, and y is the least 2-norm solution of problem **D**.

It is evident that G_γ is composed of a finite number of quadratic functions. In each domain D where $W(x)$ and $s(x)$ are constant G_γ is equal to a specific quadratic function. These domains are separated by the following union of hyperplanes,

$$B = \{x \in \mathbb{R}^n \mid \exists i : r_i(x) = 0 \wedge r_i(x) = \gamma\}. \quad (8)$$

A binary matrix W is *feasible* at x if

$$\forall \varepsilon > 0 \exists z \in \mathbb{R}^n \setminus B : \|x - z\| < \varepsilon \wedge W = W(z). \quad (9)$$

Similarly, a binary vector s is *feasible* at x if

$$\forall \varepsilon > 0 \exists z \in \mathbb{R}^n \setminus B : \|x - z\| < \varepsilon \wedge s = s(z). \quad (10)$$

If W is a feasible binary matrix and s is a feasible binary vector at some point x then $Q_{w,s}(x, \gamma)$ is the quadratic function which equals G_γ on the subset

$$C_{w,s}^\gamma = \text{cl}\{z \in \mathbb{R}^n \mid s(z) = s \wedge W(z) = W\}. \quad (11)$$

We also call $C_{w,s}^\gamma$ a Q -subset of $\mathbb{R}^n$. Notice that any $x \in \mathbb{R}^n \setminus B$ has exactly one corresponding Q -subset ($W = W(x)$), whereas a point $x \in B$ belongs to two or more Q -subsets.

$Q_{w,s}$ can be defined as follows:

$$Q_{w,s}(z, \gamma) = \frac{1}{2\gamma} (z - x)^T (A^T W A) (z - x) + G_\gamma'^T(x) (z - x) + G_\gamma(x). \quad (12)$$

The gradient of the function G_γ is given by

$$G_\gamma'(x) = A^T \left[\frac{1}{\gamma} W(x) r(x) + s(x) \right]. \quad (13)$$

For $x \in \mathbb{R}^n \setminus B$, the Hessian of G_γ exists, and is given by

$$G_\gamma''(x) = \frac{1}{\gamma} A^T W(x) A. \quad (14)$$

Note that $\mathcal{N}(A^T W A) = \mathcal{N}(W A)$, i.e., the matrices $A^T W A$ and $W A$ share the same null space since W is idempotent (i.e., $W W = W$). We also denote by X_γ the set of minimizers of $G_\gamma(x)$.

To state the Huber problem in the form (6) we could have equally used a different binary vector and matrix definition. In particular, we can define an ‘extended binary matrix’ $\bar{W}$ such that:

$$\bar{W}_{ii}(x) = \begin{cases} 1 & \text{if } 0 \leq r_i(y) \leq \gamma \\ 0 & \text{otherwise,} \end{cases} \quad (15)$$

and an extended binary vector $\bar{s}$:

$$\bar{s}_i(x) = \begin{cases} 0 & \text{if } r_i(y) \leq \gamma \\ 1 & \text{otherwise,} \end{cases} \quad (16)$$

for all $i = 1, \dots, m$. Note that the two binary matrix and binary vector definitions only

differ for those points that are on the boundary, i.e., for $x \in B$. The reason for two separate definitions is the following. In posing the Huber problem and computing its first derivative, we obtain the same results using both definitions of the binary vector and the associated binary matrix. However, this leads to different results when computing second derivatives of the function G_γ . To make a finitely convergent algorithm we need the extended definitions (15) and (16). On the other hand, the definitions (3) and (4) allow us to state an important result (Theorem 7) which is instrumental in the analysis of the algorithm.

We start by establishing the existence of a solution to the Huber problem.

THEOREM 5 There exists a solution for problem $\mathbf{HP}^\gamma$ for all $\gamma > 0$. In addition, any solution of $\mathbf{HP}^\gamma$ is a solution of $\mathbf{P}$ if the inequality system (1) is consistent.

Proof. Since $\mathbf{HP}^\gamma$ and $\mathbf{HPQ}^\gamma$ are equivalent, it suffices to show that $\mathbf{HPQ}^\gamma$ has a solution. Notice that the feasible region $\Omega = \{(x, u, v) \mid Ax - b \leq u + v, u \leq \gamma, v \geq 0\}$ is a nonempty polyhedral set, the objective function is quadratic and convex, and bounded below on Ω . Therefore, it follows from the Frank–Wolfe theory (Frank & Wolfe (1956)) that a solution exists. In addition, if the inequality system (1) is consistent, $u = v = 0$ in the optimal solution of $\mathbf{HPQ}^\gamma$. As a result, x solves the ℓ_1 problem $\mathbf{P}$. $\square$

The above result indicates that the Huber problem always has a solution like the ℓ_1 problem itself. In addition, if the inequality system (1) is consistent, it suffices to solve only one Huber problem for any $\gamma > 0$. However, this nice property is not shared by all smooth functions. Consider the smooth function proposed by Chen & Mangasarian (1995)

$$p(z, \gamma) = z + \gamma \log(1 + e^{-z/\gamma}),$$

which is the double integration of the sigmoid function $1/(1 + e^{-z/\gamma})$. We note that Chen & Mangasarian (1995) use $\alpha = \frac{1}{\gamma}$ as their smoothing parameter and formulate their smoothed problem accordingly. It is not difficult to see that the resulting smoothed problem does not have a solution even for such a simple inequality as $x \leq 0$.

Now, we have the following properties of the solution set of $\mathbf{HP}^\gamma$.

LEMMA 1 $s(x_\gamma)$ ($W(x_\gamma)$) is constant for $x_\gamma \in X_\gamma$. Furthermore $r_i(x_\gamma)$ is constant for $x_\gamma \in X_\gamma$ for i such that $W_{ii}(x_\gamma) = 1$.

Proof. The result is a consequence of Theorem 4, and the uniqueness of the dual solution y . $\square$

Following Lemma 6 we let $W(X_\gamma) = W(x_\gamma)$ and $s(X_\gamma) = s(x_\gamma)$, $x_\gamma \in X_\gamma$, be the binary matrix and the binary vector corresponding to the solution set. Now, we can use Lemma 1 to characterize the solution set X_γ .

COROLLARY 1 X_γ is a convex set which is contained in one Q -subset: $\mathcal{C}_{w,s}^\gamma$ where $W = W(X_\gamma)$ and $s = s(X_\gamma)$.

Proof. This follows from the linearity of the problem and Lemma 6. $\square$

2.2 Another Huber smoothing function

An anonymous referee indicated that we could have equally used the following function

$$q(z, \gamma) = \begin{cases} 0 & \text{if } z \leq -\gamma \\ \frac{1}{4\gamma}z^2 + \frac{1}{2}z + \frac{1}{4}\gamma & \text{if } -\gamma < z < \gamma \\ z & \text{if } z \geq \gamma \end{cases}$$

as a smooth approximation to **P**. It is easy to see that the function q is obtained by shifting the original Huber function p leftwards by γ . To pose the minimization problem in the form (6) we modify the binary matrix definition as follows. Let

$$w_i(x) = \begin{cases} 1 & \text{if } -\gamma < r_i(x) < \gamma \\ 0 & \text{otherwise,} \end{cases} \quad (17)$$

for all $i = 1, \dots, m$. Then, we define W as the diagonal matrix $\text{diag}(w_1, \dots, w_m)$. The binary vector s retains its original definition. Now, we have the following smooth problem:

$$\min_x H_\gamma(x) = \frac{1}{4\gamma} r^T(x) W(x) r(x) + w^T [\frac{1}{2}r + \frac{1}{4}\gamma e] + s^T(x) r(x). \quad (18)$$

The dual of this problem turns out to be the following quadratic program

$$\begin{aligned} & \text{minimize} && y^T (b - \gamma e) + \gamma y^T y \\ & \text{s.t.} && A^T y = 0 \\ & && 0 \leq y \leq e. \end{aligned}$$

We can now state the equivalent of Theorem 4 as follows.

THEOREM 6 Let $x \in \mathbb{R}^n$ and $y \in \mathbb{R}^m$ be a feasible solution to the dual problem. Then x and y are optimal solutions of their respective problems if and only if

$$y = \frac{1}{2\gamma} W(x) r(x) + \frac{1}{2} w(x) + s(x). \quad (19)$$

In addition, there exists a $\gamma_0 > 0$ such that the unique solution y is constant for all $0 < \gamma \leq \gamma_0$, and y is the least 2-norm solution of problem **D**.

In contrast to p , the use of q in the smoothing approximation yields a smooth minimization problem whose solutions do not necessarily coincide with the solutions of a consistent linear inequality system. To see this, consider the simple inequality system $x \leq 0$. Obviously, although all of the non-positive half-line $\mathbb{R}^-$ is the set of solutions to both the inequality and the associated problem **HP'**, only the set $\{x : x \leq -\gamma\}$ gives the set of solutions to the smooth problem obtained from q . In addition, although guaranteed to have a minimizer unlike the Chen–Mangasarian case, for large values of γ problem (18) may have a solution which fails to be feasible for a consistent system of linear inequalities. As an example, consider the linear inequality system $x \geq 0$ and $2x \leq 0$. Obviously the unique solution is $x = 0$. However, for $\gamma = 1$ problem (18) has the solution $x_\gamma = -1/5$.

3. Behaviour of the minimizers of G_γ for small γ

Assume $x_\gamma \in X_\gamma$, and $x_{\gamma-\delta} \in X_{\gamma-\delta}$ for some $0 < \delta < \gamma$. If $s(X_{\gamma-\delta}) = s(X_\gamma)$ and $W(X_{\gamma-\delta}) = W(X_\gamma)$ then, by linearity of the problem we have that $s(X_{\gamma-\varepsilon}) = s(X_\gamma)$ and $W(X_{\gamma-\varepsilon}) = W(X_\gamma)$ for $0 \leq \varepsilon \leq \delta$. Since we have a finite number of possibilities for s and W , we have the following theorem.

THEOREM 7 There exists $\gamma_0 > 0$ such that $W(X_\gamma)$ and $s(X_\gamma)$ are constant for $0 < \gamma \leq \gamma_0$.

We denote by $\mathcal{W}(X_\gamma)$ the set of all distinct extended binary matrices corresponding to the elements of X_γ . That is, for any $x_\gamma \in X_\gamma$ $\overline{W}(x_\gamma) \in \mathcal{W}(X_\gamma)$. Similarly, we define $\mathcal{S}(X_\gamma)$ as the set of all distinct extended binary vectors corresponding to the elements of X_γ . That is, for any $x_\gamma \in X_\gamma$ $\overline{s}(x_\gamma) \in \mathcal{S}(X_\gamma)$. Under the previous definitions, the following is a consequence of the linearity of the problem and Lemma 1.

LEMMA 2 If $\mathcal{W}(X_{\gamma_1}) = \mathcal{W}(X_{\gamma_2})$ and $\mathcal{S}(X_{\gamma_1}) = \mathcal{S}(X_{\gamma_2})$ where $0 < \gamma_2 < \gamma_1$ then $\mathcal{W}(X_\gamma) = \mathcal{W}(X_{\gamma_1}) = \mathcal{W}(X_{\gamma_2})$ and $\mathcal{S}(X_\gamma) = \mathcal{S}(X_{\gamma_1}) = \mathcal{S}(X_{\gamma_2})$ for $\gamma_2 \leq \gamma \leq \gamma_1$.

THEOREM 8 There exists $\bar{\gamma}$ such that $\mathcal{W}(X_\gamma)$ and $\mathcal{S}(X_\gamma)$ are constant for $\gamma \in (0, \bar{\gamma})$ where $0 < \bar{\gamma} \leq \gamma_0$.

Proof. Since $W(X_\gamma)$ and $s(X_\gamma)$ remain constant in $(0, \gamma_0]$ following Theorem 7 and the number of different extended binary vectors and matrices is finite the result is a consequence of Lemma 10. $\square$

EXAMPLE Consider the problem with the following data:

$$A = \begin{pmatrix} -1 & 0 \\ 0 & 1 \\ 1 & 0 \\ 0 & -1 \\ -1 & 1 \\ 1 & 1 \end{pmatrix}$$

and $b = (-1, -1, -1, -1, 1, 0)^T$. The solution set associated with the ℓ_1 problem in this example is

$$X = \{x \mid x = \lambda x^1 + (1 - \lambda)x^2 \quad \forall \lambda \in [0, 1]\}$$

where $x^1 = (-1/2, 1/2)^T$ and $x^2 = (1, -1)^T$ with an optimal value of 4. For $0 < \gamma \leq 1$, we have $W(X_\gamma) = \text{diag}(0, 0, 0, 0, 0, 1)$ and $s(X_\gamma) = (1, 1, 1, 1, 0, 0)$. However for $1/2 \leq \gamma \leq 1$, one has

$$X_\gamma = \{x \mid x = \lambda z^1 + (1 - \lambda)z^2 \quad \forall \lambda \in [0, 1]\}$$

where $z^1 = (\gamma - 1, 1 - \gamma)^T$ and $z^2 = (1 - \gamma, \gamma - 1)^T$. Notice that for $\gamma = 1$, the unique minimizer is $(0, 0)^T$. However, we have for $1/2 < \gamma < 1$

$$\mathcal{W}(X_\gamma) = \{\text{diag}(0, 0, 0, 0, 0, 1); \text{diag}(1, 1, 0, 0, 0, 1); \text{diag}(0, 0, 1, 1, 0, 1)\}$$

and

$$\mathcal{S}(X_\gamma) = \{(1, 1, 1, 1, 0, 0); (0, 0, 1, 1, 0, 0); (1, 1, 0, 0, 0, 0)\}.$$

For $\gamma = 1/2$, one has

$$\mathcal{W}(X_\gamma) = \{\text{diag}(0, 0, 0, 0, 0, 1); \text{diag}(1, 1, 0, 0, 0, 1); \text{diag}(0, 0, 1, 1, 1, 1)\}$$

with

$$\mathcal{S}(X_\gamma) = \{(1, 1, 1, 1, 0, 0); (0, 0, 1, 1, 0, 0); (1, 1, 0, 0, 0, 0)\}.$$

Finally, in the interval $0 < \gamma < 1/2$, we have

$$X_\gamma = \{x \mid x = \lambda t^1 + (1 - \lambda)t^2 \quad \forall \lambda \in [0, 1]\}$$

where $t^1 = (-1/2, 1/2)^T$ and $t^2 = (1 - \gamma, \gamma - 1)^T$ with

$$\mathcal{W}(X_\gamma) = \{\text{diag}(0, 0, 0, 0, 0, 1); \text{diag}(0, 0, 0, 0, 1, 1); \text{diag}(1, 1, 0, 0, 0, 1)\}$$

with

$$\mathcal{S} = \{(1, 1, 1, 1, 0, 0); (1, 1, 1, 1, 0, 0); (0, 0, 1, 1, 0, 0)\}.$$

In this example $\gamma_0 = 1$ whereas $\bar{\gamma} = 1/2$.

For convenience in our next result, define the following set

$$\mathcal{D}(x) = \{z \mid r_i(z) \leq 0 \text{ for all } i \in \sigma_-(x) \text{ and } r_i(z) \geq 0 \text{ for all } i \in \sigma_+(x)\},$$

where $\sigma_-(x) = \{i \mid r_i(x) < 0\}$ and $\sigma_+(x) = \{i \mid r_i(x) > \gamma\}$. Let $\gamma \in (0, \bar{\gamma})$ and $x_\gamma \in X_\gamma$ with $\bar{W} = \bar{W}(x_\gamma)$ and $\bar{s} = \bar{s}(x_\gamma)$. Then x_γ satisfies

$$A^T \bar{W} r(x_\gamma) = -\gamma A^T \bar{s}. \quad (20)$$

Now, consider the system

$$(A^T \bar{W} A)d = -\frac{1}{\gamma} A^T \bar{W} r(x_\gamma). \quad (21)$$

The linear system (21) is always consistent as it corresponds to the normal equations associated with $\bar{W} A h = \frac{1}{\gamma} \bar{W} r(x_\gamma)$. By Theorem 8 there exists $x_\gamma \in X_\gamma$ such that $\bar{W}(x_\gamma) = \bar{W}$ and $\bar{s}(x_\gamma) = \bar{s}$ for all $\gamma \in (0, \bar{\gamma})$, which implies, using (20) and (21), that there exists d that solves (21) such that $x_\gamma + \delta d \in X_{\gamma-\delta}$ for all $\delta \in (0, \gamma)$. Therefore, using the continuity of r we have that $x_\gamma + \gamma d$ solves $\mathbf{P}$, and $\bar{W} r(x_\gamma + \gamma d) = 0$. Since d can be replaced by $d + \eta$ in the above identity where $\eta \in \mathcal{N}(A^T \bar{W} A)$, it follows that

$$\bar{W} r(x_\gamma + \gamma d) = 0 \quad (22)$$

for any solution d to (21). Clearly, if the solution to (21) is unique, d^* say, then $x_\gamma + \gamma d^*$ solves $\mathbf{P}$. Hence, we have proved the following:

THEOREM 9 Let $\gamma \in (0, \bar{\gamma})$ and $x_\gamma \in X_\gamma$ with $\bar{W} = \bar{W}(x_\gamma)$. Then

$$\bar{W} r(x_\gamma + \gamma d) = 0 \quad (23)$$

for any solution d to (21). Furthermore, if d is unique or $x_\gamma + \gamma d \in \mathcal{D}(x_\gamma)$, then $x_\gamma + \gamma d \in X$.

4. A continuation method

Based on the analysis of the previous sections, we construct a continuation method to solve the ℓ_1 problem. The algorithm is similar to the algorithm of Madsen *et al* (1996) for robust linear regression. However, the two methods differ in the γ reduction strategy, which leads to a quite different argument for the finite convergence proof. The continuation method is also related to the least norm algorithm of Mangasarian (1984). The algorithm of Mangasarian (1984) consists of solving the dual to $\mathbf{HD}'$ using a SOR iterative scheme to compute a least norm solution to $\mathbf{D}$. Our algorithm computes both a least norm solution to the dual problem $\mathbf{D}$ and an optimal solution to the primal problem $\mathbf{P}$ by using a finite Newton-type method.

The continuation method is summarized below:

```

Choose an initial  $\gamma > 0$  and compute a minimizer  $x_\gamma$  of  $G_\gamma$ 
while not STOP
    reduce  $\gamma$ 
    compute a minimizer  $x_\gamma$  of  $G_\gamma$ 
end while.

```

We now describe each element of the algorithm in detail: the procedure of computing a minimizer of G_γ , the procedure of reducing γ , and the stopping criteria.

4.1 Computing a minimizer of G_γ

The procedure for computing a minimizer x_γ of G_γ is adapted from the modified Newton algorithm given in Madsen & Nielsen (1990) for robust linear regression using Huber functions. It solves $G'_\gamma(x) = 0$, a system of piecewise linear equations, by a modified Newton's method with a specialized line search procedure.

A search direction h is computed using the linear system

$$(A^T \overline{W} A)h = -A^T [\overline{W}r(x) + \gamma \bar{s}]. \quad (24)$$

If $A^T \overline{W} A$ has full rank, then h is the solution to (24). The algorithm proceeds with a piecewise linear one-dimensional search along h . Otherwise, if the system of equations (24) is consistent, a minimum norm solution is computed. The next iterate is found by moving to the first kink point α_1 along h , i.e., the smallest value of α where $\overline{W}(x + \alpha) \neq \overline{W}(x)$. If the system is inconsistent, a descent direction h is computed by replacing $A^T \overline{W} A$ with a suitable positive definite matrix. Following this, a piecewise linear one-dimensional search along h is performed. The one-dimensional search procedure is computationally cheap as a result of the piecewise-linear nature of G'_γ . To keep the paper at a manageable length we refer to Madsen & Nielsen (1990) for details of the algorithm. However, the following finite convergence result is needed for our paper and the result can be shown by a suitable modification of the convergence analysis in Madsen & Nielsen (1990).

LEMMA 3 Let $\gamma > 0$. A minimizer x_γ of G_γ can be found in finite iterations by the modified Newton method (similar to that described in Madsen & Nielsen (1990)).

4.2 Reduction of γ and stopping criteria

We describe the γ reduction procedure and the stopping criteria in detail. Let x_γ be a minimizer of G_γ for some $\gamma > 0$ and $\bar{W} = \bar{W}(x_\gamma)$ with $\bar{s} = \bar{s}(x_\gamma)$. The logical switch STOP evaluates TRUE if $Ax_\gamma \leq b$. In that case, the algorithm stops with a feasible solution to the linear inequality system. Otherwise, the algorithm proceeds to execute the γ reduction procedure. Based on the analysis of Section 4, $\bar{W}r(x_\gamma + \gamma d) = 0$ for any d that solves equation (21) and γ sufficiently small ($0 < \gamma \leq \bar{\gamma}$). In addition, if $x_\gamma + \gamma d \in \mathcal{D}(x_\gamma)$ (complementary to y) then $x_\gamma + \gamma d \in X$. Therefore, we construct the γ reduction strategy based on whether $\bar{W}r(x_\gamma + \gamma d) = 0$.

Case 1

$\bar{W}r(x_\gamma + \gamma d) \neq 0$. In this case, γ can be reduced by any nonzero factor and $\bar{\gamma}$ is reached in finite steps. A practical reduction procedure is described in Section 5.

Case 2

$\bar{W}r(x_\gamma + \gamma d) = 0$. In this case, we check whether $x_\gamma + \gamma d \in \mathcal{D}(x_\gamma)$. That is, for all i such that $r_i(x_\gamma) < 0$ we check whether $r_i(x_\gamma + \gamma d) \leq 0$, and for all i such that $r_i(x_\gamma) > \gamma$ whether $r_i(x_\gamma + \gamma d) \geq 0$. If the answer is affirmative, STOP evaluates TRUE and an ℓ_1 solution of the inequality system is obtained. In addition, the dual optimal solution y can be obtained by formula (7). Otherwise, let $\phi \equiv \{\alpha_k, k = 1, 2, \dots, q\}$ be the set of positive kink points where the components of r change sign, i.e., the set

$$\phi = \phi_1 \cup \phi_2$$

where

$$\phi_1 = \{0 < \alpha_i < 1 \mid \exists i \in J^+ \wedge r_i(x_\gamma) + \alpha_i a_i d = \gamma - \alpha_i\},$$

and

$$\phi_2 = \{0 < \alpha_i < 1 \mid \exists i \in J^- \wedge r_i(x_\gamma) + \alpha_i a_i d = 0\}$$

with

$$J^+ = \{i \mid 1 \leq i \leq m \wedge r_i(x_\gamma) > \gamma \wedge a_i d \neq 0\},$$

and

$$J^- = \{i \mid 1 \leq i \leq m \wedge r_i(x_\gamma) < 0 \wedge a_i d \neq 0\}.$$

We choose

$$\delta^* = \min_k \alpha_k$$

followed by setting

$$\gamma_{\text{next}} = \gamma - \delta^*.$$

Note that if the stopping criterion is not satisfied by $x_\gamma + \gamma d$ then δ^* is strictly positive.

4.3 Finite termination

Clearly, unless the stopping criteria are met and the algorithm stops with a primal–dual optimal pair, the above reduction procedure ensures that γ is reduced by a nonzero factor. Since the modified Newton method to find a minimizer of G_γ is a finite process, γ enters the range $(0, \bar{\gamma})$ in a finite number of iterations unless the algorithm stops. Therefore, to prove the finite convergence of the continuation method, it suffices to show that the method is finite once $\gamma \in (0, \bar{\gamma})$. We define the following *active set* of indices:

$$\bar{I}(x) \equiv \{i \mid 1 \leq i \leq n \wedge \bar{W}_{ii}(x) = 1\}. \quad (25)$$

LEMMA 4 Assume $0 < \gamma < \bar{\gamma}$. Let $x = x_\gamma$ and d solve (21) with $W = \bar{W}(x)$ and $s = \bar{s}(x)$. Then either $x + \gamma d \in X$ and the continuation method stops, or $x_{\text{next}} \equiv x + \delta^* d \in X_{\gamma_{\text{next}}}$, and the active set expands, i.e., $\bar{I}(x_{\text{next}}) \supset \bar{I}(x)$.

Proof. Since $0 < \gamma < \bar{\gamma}$, $\bar{W}r(x + \gamma d) = 0$ by Theorem 9. Thus, the reduction of γ follows Case 2. By Theorem 9, if $x + \gamma d \in \mathcal{D}(x_\gamma)$ then $x + \gamma d \in X$ and the algorithm stops with an ℓ_1 solution. Otherwise, the theorem implies that $\bar{I}(x) \subseteq \bar{I}(x + \gamma d)$ and therefore by the linearity of r , $\bar{I}(x) \subseteq \bar{I}(x + \delta d)$ for all $0 \leq \delta \leq \gamma$. Hence, using the definition of δ^* , we have $\bar{I}(x + \delta d) = \bar{I}(x)$ for all $0 \leq \delta < \delta^*$ and $\bar{I}(x + \delta^* d) \supset \bar{I}(x)$. In addition, we have that $x + \delta d \in X_{\gamma-\delta}$ for all $0 \leq \delta < \delta^*$. By the continuity of the gradient G'_γ , it follows that $x_{\text{next}} \in X_{\gamma-\delta^*}$. $\square$

Based on Lemma 4 either the algorithm terminates or the active set $\bar{I}$ is expanded. Since the active set has a finite dimension, the continuation method has to terminate in a finite number of iterations. Therefore we have proved the following theorem:

THEOREM 10 The continuation method terminates in a finite number of iterations with a primal–dual optimal pair.

Notice that the above finite convergence result is shown under no assumption on the matrix A of the inequality system (1). This is an improvement over many previous finite convergence results for the ℓ_1 solution of overdetermined systems of equations (Li & Swetits (1995), Madsen & Nielsen (1993), Madsen *et al* (1996), Madsen & Nielsen (1990), Madsen *et al* (1994)), where A was assumed to have full rank.

4.4 Numerical stability and implementation issues

The continuation method involves the solution of systems of the general form

$$(A^T \bar{W} A)h = A^T (\eta \bar{W} r + \beta s) \quad (26)$$

where the pair $(\eta, \beta) = (-1, -\gamma)$ corresponds to (24) and $(\eta, \beta) = (-1/\gamma, 0)$ corresponds to (21). Now, following Nielsen (1991) we consider a singular value decomposition of $\bar{W} A$ of rank k , $k \leq n$:

$$\bar{W} A = U \Sigma V^T, \quad (27)$$

where $U \in \mathbb{R}^{m \times q}$ and $V \in \mathbb{R}^{n \times q}$ satisfy $U^T U = V^T V = I$ and $\Sigma = \text{diag}(\sigma_1, \dots, \sigma_q)$ with positive singular values σ_j . Using (27) in (26) we get

$$V \Sigma^2 V^T h = \eta V \Sigma U^T r + \beta A^T s. \quad (28)$$

If $k < n$ and the system is consistent, we compute the minimum norm solution as

$$h_m = \eta V \Sigma^{-1} U^T r + \beta V \Sigma^{-2} V^T A^T s. \quad (29)$$

For the case $k = n$, h_m reduces to the unique solution. Expression (29) shows that if some singular values are much smaller than others, then the term involving $A^T s$ may have a marked effect on the solution and may lead to severe loss of accuracy. Following Björck (1990) we can say that the part of the solution corresponding to $\eta A^T \bar{W} r$ can be found with a relative order of magnitude $n \epsilon_M \kappa(\bar{W} A)$ where ϵ_M denotes the computer unit roundoff and $\kappa(\bar{W} A)$ is the condition number of $\bar{W} A$. When $\beta = -\gamma$, we essentially face a problem with condition number $(\kappa(\bar{W} A))^2$.

There are two measures to alleviate the effects of possible ill-conditioning on the computation of h . (i) We use an LDL^T factorization of $A^T \bar{W} A$ based on a QR -type factorization of $\bar{W} A$. This combines some ideas of Gentleman (1973) and Fletcher & Powell (1974). (ii) We use one step of iterative refinement on the computed solution $\tilde{h}$. To this effect, let Δ denote the correction defined by

$$h = \tilde{h} + \Delta.$$

We solve for Δ from the system

$$A^T \bar{W} A \Delta = \xi$$

where $\xi = A^T (\eta W r + \beta s) - A^T \bar{W} A \tilde{h}$.

The algorithm described in Sections 4.1 and 4.2 was implemented in Fortran 77 without exploiting sparsity. We refer to this implementation as LIPACK. The implementation is largely based on the ideas of Nielsen (1990, 1991). The dominant work in the algorithm is the factorization of the matrix $A^T \bar{W} A$, which is required to solve the linear system (21) and to find the minimizer of G_γ (system (24)). An important feature of the algorithm is that only a few entries of the matrix $\bar{W}$ change between two consecutive iterations. This suggests that one should use the information obtained from the matrix $A^T \bar{W} A$ in the previous iteration to perform the inversion of $A^T \bar{W} A$ in the current iteration. This fact has also been exploited in all active set and simplex codes. In LIPACK this is achieved by maintaining an LDL^T factorization of $A^T \bar{W} A$ by a sequence of rank-one updates (or, downdates). When the cost of updating the factorization is higher than a refactorization or when there is an indication of numerical instability, a refactorization is performed. This is done only occasionally. When the system (24) is inconsistent, the projection of g on the null space of C is computed and used as a search direction. To perform these tasks the AAFAC package of Nielsen (1990) was used in LIPACK. AAFAC is a collection of Fortran 77 subroutines to solve linear systems of the form $A^T A x = c$ based on the LDL^T factorization of the matrix. It also performs rank-one up- and downdates of L and D as well as rank and condition estimation.

The solution of (21) is performed after a minimizer of G_γ is obtained. This implies that the LDL^T factors from the previous Newton iteration are available.

Initialization

To initiate the solution of the first Huber problem, we solve the linear system

$$A^T A x = A^T b.$$

Let x^0 be any solution. Let $r^0 \equiv Ax^0 - b$. We choose the initial value of γ , γ^0 , using the formula

$$\gamma^0 = \max_{i=1,\dots,m} |r_i^0|.$$

Implementation of the γ reduction procedure when $\overline{W}r(x_\gamma + \gamma d) \neq 0$ (Case 1)

In this case we reduce γ as follows. Let $\mathbf{c}(\gamma - \delta)$ denote the number of changes in the active set from $\overline{T}(x_\gamma)$ to $\overline{T}(x_\gamma + \delta d)$. We use bisection to find a value $\bar{\delta}$ of δ such that $\mathbf{c}(\gamma - \bar{\delta}) \approx \frac{1}{2}\mathbf{c}(\gamma)$, and use

$$\gamma_{\text{next}} = \gamma - \bar{\delta}.$$

For robustness we search only in the interval $[0.1\gamma, \gamma)$ so that $\gamma_{\text{next}} \leq 0.9\gamma$.

LIPACK is available for distribution as a standard Fortran 77 subroutine.

5. Numerical results

The purpose of this section is to demonstrate the viability of the proposed algorithm using both smoothing functions p and q .

To test the performance of the algorithm, we have randomly generated two sets of test problems: (i) consistent and overdetermined linear inequality systems, and (ii) inconsistent and overdetermined linear inequality systems with known optimal solutions so as to satisfy the optimality conditions stated in Theorem 1. For the inconsistent case we have generated two separate sets of well-conditioned and ill-conditioned test examples. We have compared our results to one of the best available dense linear programming packages in the software domain, LSSOL of Gill *et al* (1986). The purpose of this comparison is not to claim any clear superiority over an alternative approach via linear programming. Rather, we intend to give the reader a feel of how the proposed algorithms perform relative to an off-the-shelf dense linear programming package. A healthier comparison should certainly be made between a sparsity exploiting implementation of the proposed algorithms and state-of-the-art software for linear programming.

For consistent systems, we have used the FP option of LSSOL which signals a linear feasibility problem to LSSOL. For inconsistent systems, we have used both primal and dual linear programming formulations of $\mathbf{P}$ as input to LSSOL. We have observed that LSSOL performs four times faster on average on the dual formulation $\mathbf{D}$. Therefore, we only report results of LSSOL obtained on the dual formulation.

We begin with consistent systems. In Figs 1 and 2 below we report the average run time, and the average number of iterations of the smoothing algorithm using p , referred to as ‘MN’, the smoothing algorithm using q , referred to as ‘Shifted Huber’, and finally LSSOL for ten problems of a given size. Figures 1 and 2 indicate that the algorithm based on p is quite competitive with both LSSOL and the algorithm based on q .

In Figs 3 and 4 we use some well-conditioned but inconsistent inequality systems. Here the continuation method based on the function q seems to outperform the other algorithms. To complete the statistics, we note that the number of refactorizations used by MN increases from 9 to 13 as the problem size gets bigger, from 29 to 60 for γ -reduction steps on these problems. The corresponding figures for Shifted Huber are 3 and 6 refactorizations, with 26 to 43 reductions on average.

FIG. 1. Run time results of the smoothing methods and LSSOL for consistent systems.

In our final results we generate ill-conditioned inconsistent linear inequalities by weighting the rows of A using different constants. This leads to increasingly ill-conditioned examples as the size of the problems increases. We were not able to solve any of these problems using LSSOL. Therefore, we report our results using MN and Shifted Huber in Figs 5 and 6. We notice a significant degradation in the speed of both codes. Here, the number of refactorizations used by MN increases from 5 to 132 with the problem size, and the number of reductions from 48 to 136 while in Shifted Huber the respective numbers are 4 and 55 refactorizations and 42 to 111 γ -reductions. On the other hand, the algorithms always return solutions with the accuracy the condition of the problem permits. For instance, for the initial sizes starting from 100, the condition of $\overline{W}A$ on termination is of the order of 10^5 . For such problems, both codes return fully accurate (15 digits in double precision) solutions. For larger problems, we observe that the condition number of $\overline{W}A$ on termination

FIG. 2. Iteration results of the smoothing methods and LSSOL for consistent systems.

FIG. 3. Run time results of the smoothing methods and LSSOL for inconsistent and well-conditioned systems.

is around 10^8 where the codes return solutions with approximately 8 correct digits in the objective function value. Finally for the largest problems towards the end of the spectrum, the condition number of $\overline{W}A$ reaches approximately 10^{11} in which case we get around 5 accurate digits. These experimental results are in line with the analysis of Section 4.4. Our analysis indicated that during the reduction steps we essentially face a problem with condition number $\kappa(\overline{W}A)$. Our final accuracy corroborates this view.

We observed that the final value of γ before termination is usually between 10^{-2} and 10^{-4} while we occasionally encounter test cases where γ is reduced to 10^{-8} . However, this does not seem to have any correlation with the ill-conditioning since we observe such γ values for well-conditioned cases as well. The smooth subproblems become harder to solve when we force ill-conditioning into the problem as signalled by the analysis of Section 4.4 since subproblems involve the solution of Newton systems with the term $A^T s$.

FIG. 4. Iteration results of the smoothing methods and LSSOL for inconsistent and well-conditioned systems.

FIG. 5. Run time results of the smoothing methods for inconsistent and ill-conditioned systems.

These preliminary results indicate that a specialized method for $\mathbf{P}$ may indeed be more appropriate. A future version of LIPACK that exploits sparsity seems to be a worthwhile undertaking.

Acknowledgements

We gratefully acknowledge the careful work of two anonymous referees. In particular, one referee suggested the alternative Huber function which led to an improved performance of the proposed algorithm on inconsistent systems. This research was supported by NATO Collaborative Research Grant CRG-94-0609.

FIG. 6. Iteration results of the smoothing methods for inconsistent and ill-conditioned systems.

REFERENCES

- BENNETT, K. & MANGASARIAN, O. L. 1992 Robust linear programming discrimination of two linearly inseparable sets. *Optimiz. Methods Software* **1**, 23–34.
- BJÖRCK, Å. 1990 *Least Squares Methods*, Vol. 1 of *Handbook of Numerical Analysis* (P. G. Ciarlet and J. L. Lions, eds). Amsterdam: Elsevier, pp 465–652.
- BRAMLEY, R. & WINNICKA, B. 1996 Solving linear inequalities in a least squares sense. *SIAM J. Sci. Comput.* **17**, 275–286.
- CENSOR, Y. & ZENIOS, S. A. 1997 *Parallel Optimization: Theory, Algorithms and Applications*. New York: Oxford University Press.
- CHEN, C. & MANGASARIAN, O. L. 1995 Smoothing methods for convex inequalities and linear complementarity problems. *Math. Programming* **71**, 51–70.
- CHEN, C. & MANGASARIAN, O. L. 1996 A class of smoothing functions for nonlinear and mixed complementarity problems. *Comput. Optimiz. Appl.* **5**, 97–138.
- FLETCHER, R. & POWELL, M. J. D. 1974 On the modification of LDL^T factorizations. *Math. Comput.* **28**, 1067–1087.
- FRANK, M. & WOLFE, P. 1956 An algorithm for quadratic programming. *Naval Res. Logistics Quart.* **3**, 95–110.
- GENTLEMAN, W. M. 1973 Least squares computation by Givens transformations without square roots. *J.I.M.A.* **12**, 329–336.
- GILL, P. E., HAMMARLING, S., MURRAY, W., SAUNDERS, M. A., & WRIGHT, M. H. 1986 User's guide for LSSOL (Version 1.0): a Fortran package for constrained linear least squares and convex quadratic programming. *Report SOL 86-1*, Department of Operations Research, Stanford University.
- HAN, S.-P. 1980 Least-squares solution of linear inequalities. *Technical Report TR-2141*, Mathematics Research Center, University of Wisconsin-Madison.
- LI, W., PARDALOS, P., & HAN, C. G. 1992 Gauss–Seidel method for least distance problems. *J. Optimiz. Theory Applic.* **75**, 487–500.
- LI, W. & SWETTIS, J. 1995 Linear ℓ_1 estimator and Huber M-estimator. *Technical Report*, Old Dominion University.
- MADSEN, K. & NIELSEN, H. B. 1990 Finite algorithms for robust linear regression. *BIT* **30**, 682–699.
- MADSEN, K. & NIELSEN, H. B. 1993 A finite smoothing algorithm for linear ℓ_1 estimation. *SIAM J. Optimiz.* **3**, 223–235.
- MADSEN, K., NIELSEN, H. B., & PINAR, M. Ç. 1994 New characterizations of ℓ_1 solutions of overdetermined linear systems. *Operations Res. Lett.* **16**, 159–166.
- MADSEN, K., NIELSEN, H. B., & PINAR, M. Ç. 1996 A new finite continuation algorithm for linear programming. *SIAM J. Optimiz.* **6**, 600–616.
- MANGASARIAN, O. L. 1984 Normal solutions of linear programs. *Math. Programming Study* **22**, 206–216.
- MANGASARIAN, O. L. & MEYER, R. R. 1979 Nonlinear perturbations of linear programs. *SIAM J. Control Optimiz.* **17**, 745–752.
- MICHELOT, C. & BOUGEARD, M. L. 1994 Duality results and proximal solutions of the Huber M-estimator problem. *Appl. Math. Optimiz.* **30**, 203–221.
- MURTY, K. G. 1976 *Linear and Combinatorial Programming*. New York: John Wiley & Sons.
- NIELSEN, H. B. 1990 AAFAC: a package of Fortran 77 subprograms for solving $A^T Ax = c$. *Report NI-90-11*, Institute for Numerical Analysis, Technical University of Denmark.
- NIELSEN, H. B. 1991 Implementation of a finite algorithm for linear ℓ_1 estimation. *Report NI-91-01*, Institute for Numerical Analysis, Technical University of Denmark.
- PINAR, M. Ç. 1996 Newton's method for linear inequality systems. *Technical Report*, Bilkent University; *Europ. J. Oper. Res.* **107**, No 3, 710–719.
- WATSON, G. A. 1980 *Approximation Theory and Numerical Methods*. New York: John Wiley & Sons.