Theory and Methodology

# Newton's method for linear inequality systems

Mustafa Ç. Pınar *

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

## Abstract

We describe a modified Newton type algorithm for the solution of linear inequality systems in the sense of minimizing the $\ell_2$ norm of infeasibilities. Finite termination is proved, and numerical results are given. © 1998 Elsevier Science B.V.

*Keywords:* Linear inequalities; Piecewise-quadratic functions; Newton's method; Finiteness

## 1. Introduction and background

We consider the problem of finding a feasible point with respect to the overdetermined set of linear inequalities:

$$A^\mathrm{T} y \le c, \tag{1}$$

where $A$ is an $m \times n$ real matrix with $n > m$ and $c$ is an $n$-vector. Let

$$s := c - A^\mathrm{T} y. \tag{2}$$

We would like to compute a solution $y$ such that $s \ge 0$. This problem has applications in image reconstruction in computerized tomography; see Herman (1980). A recent account of iterative methods for this problem along with applications can be found in the forthcoming book, Censor and Zenios (1997). Linear inequality systems are also solved as auxiliary problems in linear and nonlinear optimization.

In the present paper we consider a modified Newton algorithm to find a solution that minimizes the

$\ell_2$ norm of infeasibilities with respect to (1). The problem consists in the minimization of piecewise-quadratic objective function with discontinuous second derivatives. More precisely, we consider the following minimization problem for computing a solution that minimizes the sum of squares of infeasibilities with respect to (1):

$$\min_y F(y) \equiv \frac{1}{2}\|s^-\|_2^2, \tag{3}$$

where $s^-$ is a vector whose $i$th component $s_i^- = \min\{0, s_i\}$. Clearly, if the linear inequality system (1) is consistent, a minimizer of $F$ is also a feasible solution with respect to (1) and vice versa. The algorithm proposed here is adapted from a Newton algorithm developed by Madsen and Nielsen (1990) for the minimization of the Huber function in robust linear regression analysis. The crucial observation that motivated the present paper is that the Huber function and the $\ell_2$ function are structurally identical in the context of linear inequality systems. In the present paper, we apply the ideas in Madsen and Nielsen (1990) to the $\ell_2$ solution of linear inequality systems. However, their finite termination analy-

* Fax: +90-312-266-4126; e-mail: mustafap@bilkent.edu.tr.

sis is no longer applicable here because the function $F$ of the present paper does not have bounded level sets when the underlying linear inequality system is consistent. This property holds for the Huber function under a full rank assumption and is used in the convergence analysis in Madsen and Nielsen (1990). Using some general results on the minimization of piecewise-quadratic functions from Li and Swetits (1997), we give a new proof of finite termination in the present paper.

The Madsen–Nielsen algorithm was developed to give a numerical solution to a problem in robust statistics. The robust estimation problem was studied extensively by Huber (1981). Other references on the solution of robust estimation problems include Clark and Osborne (1986) and Ekblom (1988). The Huber function can be considered as an alternative to least squares estimation. Essentially, the Huber function is a linear-quadratic function which treats arguments above (below for negative arguments) a certain positive threshold value by a linear term. Arguments which are smaller in absolute value than the threshold are treated by a quadratic term exactly as in the least squares case. This reduces the negative effect of the outliers that may be present among data points. For more details on the statistical properties of this function, the reader is referred to Huber (1981).

The approach of the present paper brings together several ideas of the optimization literature from the minimization of piecewise-quadratic functions to the solution of linear inequalities and linear programs. Piecewise-quadratic functions arise in optimization mainly in connection with the use of quadratic penalty and augmented Lagrangean functions. These functions are related by duality to proximal point methods; see e.g. Rockafellar (1976). An early reference on the minimization of piecewise-quadratic functions is Katznelson (1965) where a Newton-type algorithm to solve nonlinear resistor networks was developed. The minimization of piecewise-quadratic functions can be examined under the general title of convex quadratic programming. An excellent survey of iterative methods for solving such problems was provided in Lin and Pang (1987). However, this survey mainly focused on algorithms for strictly convex quadratic problems while some ideas on how to solve convex quadratic programs using algorithms for the strictly convex case were discussed. A related line of research in this area was carried out in a series of papers by Mangasarian (1981, 1984) and De Leone and Mangasarian (1987), where the piece-wise-quadratic function obtained from the application of quadratic penalty and augmented Lagrangean functions to linear inequalities and linear programs was minimized using a successive overrelaxation algorithm. However, unlike our approach, De Leone and Mangasarian did not directly work on the piece-wise-quadratic function. Instead, they minimized a strictly convex penalty function using slack variables in the system $A^T y \le c$. In Li and Swetits (1993), an algorithm similar to Katznelson's was provided in the context of $r$-convex approximation problems. Both Katznelson and Li and Swetits worked under the assumption (or, the condition) that the piecewise-quadratic function is strictly convex within any polyhedral set created by the hyperplanes $c - A^T y$. This ensures that the (generalized) Hessian is always non-singular. Later, in Li and Swetits (1997) the two authors developed a more sophisticated algorithm for piecewise-quadratic minimization problems where this condition is not satisfied. This is a mixture of active set and dual descent algorithm. Although it involves the use of a Newton step occasionally, in general it cannot be interpreted as a generalized Newton algorithm. Consequently, it requires a longer and more complicated convergence and finiteness analysis. Li (1997) investigated the application of conjugate gradient methods to the minimization of piecewise-quadratic functions.

Another line of work on piecewise-quadratic minimization was recently developed in Rockafellar (1987, 1990) with a view to solve large scale optimal control and stochastic programming problems using penalty methods. Rockafellar terms the problem 'linear-quadratic programming' and proposes what he calls finite-envelope methods which exploit the special structure of the piecewise-quadratic functions. These methods are based on expressing the objective function by an 'envelope' formula derived from a Lagrangean function. Structural properties of piece-wise-quadratic functions have also been studied in Sun (1992).

In Chen and Mangasarian (1995) the authors used some smooth approximations to the plus function $x^+ = \max\{x, 0\}$ to develop new iterative methods for the solution of linear and convex inequalities. While

these functions give rise to strictly convex minimization problems they are only guaranteed to have a minimizer under a strict interior point condition for the linear inequality system. Our present approach does not require such an assumption.

Against this background, our method is akin to the algorithms of Li and Swetits although it is much simpler. We make no assumption that the piecewise-quadratic function has a strictly convex quadratic representation in the entire domain. Our method uses a generalized Hessian and involves the computation of descent directions when the generalized Hessian is singular along with a particular step length control. While these ideas are essentially adapted from early work of Madsen and Nielsen on robust statistics, our contributions are

1. a more general finiteness proof than Madsen and Nielsen (1990); and
2. a demonstration of the computational viability of this approach.

On the computational front we develop a stable and efficient implementation of the proposed algorithm and compare its performance to a standard library routine for linear inequalities and linear and quadratic programs, LSSOL from Stanford University Optimization Systems Laboratory; see Gill et al. (1986). Results indicate a factor of three to four speed-up over this routine.

In the rest of the paper, we proceed as follows. In Section 2 we present the problem we consider and introduce the piecewise-quadratic minimization problem. Some elementary properties are discussed. In Section 3, we present the algorithm. Section 4 is devoted to the finite convergence analysis. Finally, in Section 5 we report numerical results on some randomly generated problems.

## 2. Basic properties of $F$

In this section we give some definitions and expose some basic properties of $F$ which will be useful in subsequent sections. We assume throughout the paper that $A$ has rank $m$. For $y \in \mathbb{R}^m$ we define the following *active set* of indices:

$$I(y) \equiv \{i \mid 1 \leq i \leq n \wedge s_i(y) \leq 0\}. \tag{4}$$

Let $A_I$ denote the submatrix of $A$ composed of those columns corresponding to the indices in the active set $I$. Similarly, we use $s_I$ the denote the vector with components corresponding to the active indices.

Consider first the solvability of (3) in general. Using Lagrangean duality, it can be shown that the dual problem to (3) is given as:

$$
\begin{aligned}
\min \quad & c^T x + \frac{1}{2}\|x\|_2^2 \\
\text{s.t.} \quad & Ax = 0, \\
& x \geq 0.
\end{aligned}
$$

Since this is a strictly convex homogeneous quadratic optimization problem, it always has a unique optimal solution. This implies that the problem (3) is well-defined.

It is evident that $F$ is composed of a finite number of quadratic functions. In each domain $D \subseteq \mathbb{R}^m$ where $I(y)$ is constant $F$ is equal to a specific quadratic function. These domains are separated by a union of hyperplanes.

$$B = \{y \in \mathbb{R}^m \mid \exists i : s_i(y) = 0\}. \tag{5}$$

Given the active set $I$, $Q_I$ is the quadratic function which equals $F$ on the subset

$$\mathscr{C}_I = \mathrm{cl}\{z \in \mathbb{R}^m \mid I(z) = I\}. \tag{6}$$

We also call $\mathscr{C}_I$ a *Q-subset* of $\mathbb{R}^m$. Notice that any $y \in \mathbb{R}^m \setminus B$ has exactly one corresponding $Q$-subset ($I = I(y)$), whereas a point $y \in B$ belongs to two or more $Q$-subsets.

$Q_I$ can be defined as follows:

$$
\begin{aligned}
Q_I(z) = \frac{1}{2}(z-y)^T \left(A_I A_I^T\right)(z-y) \\
+ F'^T(y)(z-y) + F(y).
\end{aligned}
\tag{7}
$$

Let also

$$V(y) \equiv \mathrm{span}\{a_i^T \mid i \in I(y)\}, \tag{8}$$

where $a_i$ is a column vector that denotes the $i$th row of $A$, and let $N(y)$ denote the orthogonal complement of $V(y)$. The gradient of the function $F$ is given by

$$F'(y) = -A_I s_I(y). \tag{9}$$

For $y \in \mathbb{R}^m \setminus B$, the Hessian of $F$ exists, and is given by

$$F''(y) = A_I A_I^T. \tag{10}$$

The following simple properties of the quadratics $Q_I$ will be useful in the finite termination proof in Section 4.

**Lemma 2.1.** *For any* $y \in \mathbb{R}^m$

$$V(y) = \{Q_I'' z \mid z \in \mathbb{R}^m\},$$

*and*

$$N(y) = \{\eta \mid Q_I'' \eta = 0\}.$$

**Proof.** This follows from the definitions (7), (8), (9), and (10).    □

**Lemma 2.2.** *Let* $y \in \mathbb{R}^m$ *and assume that* $Q_I$ *has a minimizer* $z$, *where* $I = I(y)$. *Then the set of minimizers* $M(Q_I)$ *of* $Q_I$ *is given by*

$$M(Q_I) = \{z + \eta \mid \eta \in N(y)\}.$$

**Proof.** The set of minimizers of $Q_I$ is the set of solutions to the linear system of equations

$$Q_I'' h = -Q_I'(y).$$

Now, the lemma follows from Lemma 2.1.    □

## 3. The Newton algorithm

The crucial design parameters of Newton's method for minimizing piecewise-quadratic functions are the following:

● How to select a generalized Hessian and descent directions?
● How to control the step length?

The answer to the first question follows from the definition (4) of the active set which includes both violated and binding constraints. In other words, locally binding constraints are treated as locally violated constraints and the algorithm proceeds by solving the Newton system corresponding to a local quadratic representation of $F$. When the generalized Hessian is non-singular we use the unique solution to the Newton system as descent direction, and perform exact line search. The question concerning the step length is answered by choosing the least norm solution as a descent direction in case the generalized Hessian is singular and using a restricted step length in this direction.

More precisely, we consider the Newton system

$$Q_I'' h = -Q_I'(y), \tag{11}$$

i.e., the system

$$\left(A_I A_I^T\right) h = A_I s_I(y). \tag{12}$$

The above Newton system may not always have a non-singular matrix $A_I A_I^T$. However, it is always consistent. In the case where the matrix $A_I A_I^T$ is singular, the question arises as to which solution to use as a descent direction. Our algorithm uses the following solutions of the Newton system in the two possible cases:

**Case 1.** $Q_I''$ is non-singular: in this case, let $h = -Q_I''^{-1} Q_I'$.

**Case 2.** $Q_I''$ is singular: in this case, compute $\hat{h}$, the solution with least 2-norm to the system (11) to be used a search direction.

The following lemma which establishes that the above choices define a descent direction for $F$ will be useful in the convergence analysis.

**Lemma 3.1.** *The search direction vectors* $h$ *and* $\hat{h}$ *computed as described above are descent directions.*

**Proof.** The result is obvious in Case 1 as a result of the positive definiteness of $Q_I''$. For Case 2, consider an eigenvalue decomposition of $Q_I''$. Since $Q_I''$ is symmetric and positive (semi)definite, it has an orthogonal eigensystem, $(\lambda_i, e_i)$, $i = 1, \ldots, m$, say. Suppose that $\lambda_i > 0$ for all $i = 1, \ldots, s$, and $\lambda_i = 0$ for $i = s + 1, \ldots, m$.

Let $\hat{h}$ be the least 2-norm solution of (11). As the system (11) is consistent, the negative gradient $-Q_I'$ has the expansion $-Q_I' = \sum_{i=1}^s \beta_i e_i$. Since the least norm solution of (11) is given by

$$\hat{h} = \sum_i^s \frac{\beta_i}{\lambda_i} e_i,$$

then, obviously, $-Q_I'^T \hat{h} = (\sum_{i=1}^s \beta_i e_i)^T (\sum_{i=1}^s [\beta_i / \lambda_i] e_i) = \sum_{i=1}^s (\beta_i^2 / \lambda_i) > 0$.    □

Regarding the line search procedure, in Case 1 the next iterate is found through a line search aiming for a zero of the directional derivative. More precisely, the next iterate is the point $y + \alpha h$, $\alpha > 0$, for which the function

$$\rho(\alpha) = F(y + \alpha h).$$

is minimized. Since $\rho$ is a convex univariate function, the problem is to find a zero of the increasing piecewise-linear smooth function $\rho'$. The solution to this problem is positive since $\rho'(0) < 0$ by the definition of $h$. We perform exact line search if the generalized Hessian is locally positive definite. Otherwise we move to the nearest breakpoint in Case 2. In other words, the line search in this case consists of simply choosing $\alpha = \alpha_1$, i.e., the smallest element of the set

$$\mathscr{X} = \left\{ \alpha > 0 \mid \exists i \in E : -a_i^{\mathrm{T}}\left( y + \alpha \hat{h} \right) + c_i = 0 \right\},$$

where $E = \{ i \mid 1 \le i \le m \wedge ( A^{\mathrm{T}}\hat{h})_i \ne 0 \}$. Clearly, the line search can be carried out in time proportional to $n$.

The reason for this step length policy is the following. At an iterate $y$ with $I = I(y)$, if the search direction is computed in Case 1, then the point $y + h$ is a minimizer of $F$ if $I(y + h) = I$ (i.e., $y + h \in \mathscr{G}_I$). This follows because $y + h$ is the unique minimizer of the quadratic $Q_I$ and $F$ is a convex function. That is, the quadratic form which contains the global minimizer has been identified. If $I(y + h) \ne I$, this is an indication that the global minimizer (the minimizer of $F$) is not in $\mathscr{G}_I$. Then the line search is performed. However, if at a given iterate $y$, the matrix $Q_I''$ is positive semi-definite the global minimizer may be in $\mathscr{G}_I$ even if $y + h$, as computed from Case 2, is not an element of $\mathscr{G}_I$. In this case, to locate the global minimizer, we only move to the smallest breakpoint in the line search phase. If the global minimizer is in $\mathscr{G}_I$, then this move will expand the active set of indices $I$ (cf. Lemma 4.5). Continuing in this fashion, we can locate the minimizer in a finite number of steps. These ideas are made precise in Section 4.

We summarize below the modified Newton algorithm:

```
stop = false
repeat
    I = I(y)
    if Q_I'' is positive definite
        find h as the unique solution to (11)
        if y + h ∈ 𝒢_I then
            y ← y + h
            stop = true
        else
            y = y + αh
        endif
    else
        if Q_I'' is positive semi-definite
            find ĥ, the least norm solution of (11)
            if y + ĥ ∈ 𝒢_I, then
                y ← y + ĥ
                stop = true
            else
                y ← y + α_1 ĥ
            endif
        endif
    until stop.
```

## 4. Convergence analysis

In this section we will show that the algorithm of Section 3 locates a minimizer of $F$ in a finite calculation. Since the level sets of the function $F$ are not necessarily bounded (as is the case with the Huber function) when the linear inequality system is consistent the finite convergence analysis in Madsen and Nielsen (1990) is not applicable. However, it is possible to give a new finite convergence proof. To this end, we cite without proof four technical results on the minimization of convex quadratic spline functions from Li and Swetits (1997). As $F$ is a convex quadratic spline function, these results are directly applicable to our case.

Consider the following unconstrained minimization of a convex quadratic spline $f(w)$:

$$-\infty < f_{\min} := \min_{w \in \mathbb{R}^m} f(w). \tag{13}$$

Let $W^*$ be the solution set of (13). That is, $W^* := \{ w \in \mathbb{R}^m : f(w) = f_{\min} \}$.

**Lemma 4.1.** *Suppose that $h(w)$ is a convex function, its gradient $h'(w)$ is Lipschitz continuous, and $0 < \beta$*

$< 1$. *Then there exists a positive constant* $\gamma$ (*depending only on h and* $\beta$) *such that*

$$\left(\frac{z^T h'(w)}{\|z\|}\right)^2 \leq \gamma(h(w) - h(w + tz)), \qquad (14)$$

*whenever* $t > 0$ *and* $0 \geq z^T h'(w + tz) \geq \beta \cdot z^T h'(w)$.

**Lemma 4.2.** *Suppose that* $f(w^{k+1}) \leq f(w^k)$ *for* $k = 0,1,\ldots$. *If there exists a subsequence* $\{k_j\}$ *such that* $\lim_{j \to \infty} \|f'(w^{k_j})\| = 0$, *then* $\lim_{k \to \infty} f(w^k) = f_{\min}$.

**Lemma 4.3.** *Suppose that* $f(w^{k+1}) \leq f(w^k)$ *for* $k = 0,1,\ldots$, *and* $\mathscr{D}$ *is a collection of finitely many positive definite matrices. If there are infinitely many $k$'s such that*

$$\left(w^{k+1} - w^k\right)^T f'\left(w^{k+1}\right) = 0$$

*and* $w^{k+1} = w^k - t_k D^k f'(w^k)$, *for* $D^k \in \mathscr{D}$, $t_k > 0$, $(15)$

*then* $\lim_{k \to \infty} f(w^k) = f_{\min}$.

**Lemma 4.4.** *If* $\lim_{k \to \infty} f(w^k) = f_{\min}$, *then there exists* $k^* \geq 1$ *such that, for* $k \geq k^*$, $w^k \in \mathscr{C}_I$ *implies* $\mathscr{C}_I \cap W^* \neq \emptyset$.

The next two technical results are also needed in the proof of finite termination.

**Lemma 4.5.** *Let* $y \in \mathbb{R}^m$ *with* $I = I(y)$. *Assume that* $z \in \mathscr{C}_I$ *is a minimizer of* $Q_I$. *Then for any minimizer* $z^*$ *of* $Q_I$, *we have* $I(z^*) \supseteq I(y)$. *Furthermore, z is a minimizer of F.*

**Proof.** Let $z^*$ be a minimizer of $Q_I$. Using Lemma 2.2, we have $(z^* - z)^T a_i = 0$ for any $i \in I(y)$. Therefore, $-b_i + a_i^T z^* = -b_i + a_i^T z \geq 0$ for any $i \in I(y)$. Hence, $I(z^*) \supseteq I(y)$. The last statement follows from (7) and convexity of $F$. $\square$

**Lemma 4.6.** *Let* $y \in \mathbb{R}^m$ *with* $I = I(y)$. *Suppose that F has a minimizer* $z \in \mathscr{C}_I$. *Let* $y'$ *be generated by one iteration of the algorithm of Section 3 starting from y. Then* $y' \in \mathscr{C}_I$, *and either* $y'$ *is a minimizer of F and the algorithm stops, or* $y' = y + \alpha_1 \hat{h}$ *with* $\alpha_1 < 1$ *and* $I(y') \supset I(y)$.

**Proof.** Since $z$ is a minimizer of $F$, this implies that $z$ is also a minimizer of $Q_I$. Hence, $Q_I'(z) = 0$. If $Q_I''$ is positive definite, then $z$ must be the unique minimizer of $Q_I$. Therefore, $y' = y + h = z \in \mathscr{C}_I$, and the algorithm stops.

If $Q_I''$ is not positive definite then $y + \hat{h}$ is a minimizer of $Q_I$ where $\hat{h}$ is the least norm solution of (11). If $y + \hat{h} \in \mathscr{C}_I$, then $y + \hat{h}$ is a minimizer of $F$ by definition and (7), and the algorithm stops. If $y + \hat{h} \notin \mathscr{C}_I$, then $y' = y + \alpha_1 \hat{h}$ with $\alpha_1 < 1$. Since $y + \hat{h}$ minimizes $Q_I$, $I(y + \hat{h}) \supseteq I(y)$ by Lemma 4.5. Hence, we have

$$I(y + \alpha \hat{h}) \supseteq I(y) \text{ for } 0 \leq \alpha \leq 1 \qquad (16)$$

by the definition of $I$ and the linearity of the problem. The definition of implies that $I(y + \alpha \hat{h}) = I(y)$ for $0 \leq \alpha < \alpha_1$, and there exists $j \in E$ such that $s_j(y + \alpha_1 h) = 0$. Now, $j \notin I(y)$ since otherwise (16) would be contradicted for $0 \leq \alpha < \alpha_1$. Therefore, the lemma is proved. $\square$

**Theorem 4.1.** *The algorithm of Section 3 computes a minimizer of F after a finite number of iterations.*

**Proof.** Let us assume that the algorithm generates an infinite sequence $\{y_k\}$. The first step in the proof is to show $\lim_{k \to \infty} F(y^k) = F_{\min}$ where $F_{\min}$ denotes the minimum value of $F$. We know that the search directions computed as discussed in Section 3 are descent directions. This follows from Lemma 3.1 in Section 3. Hence, the algorithm generates a descent sequence. Therefore, the first step of the proof follows from Lemma 4.3 if infinitely many iterates are generated by $y^{k+1} = y^k + \alpha^k h^k$ where $h^k$ is the unique solution to (11) and $\alpha^k$ is obtained from exact line search. However, this argument is not sufficient since it may happen that $y^{k+1} = y^k + \alpha_1^k \hat{h}^k$ where $\hat{h}^k$ is the minimum norm solution of (11) for $k$ large enough.

Therefore, we have the following claim.

**Claim.** $\lim_{k \to \infty} F(y^k) = F_{\min}$.

Assume that there exists $k_0$ such that $y^{k+1} = y^k + \alpha_1^k \hat{h}^k$ for $k \geq k_0$. Let $\hat{y}^k = y^k + \hat{h}^k$. If $(\hat{y}^k - y^k)^T F'(y^{k+1}) \geq \frac{1}{2}(\hat{y}^k - y^k)^T F'(y^k)$ by Lemma 4.1, there exists a positive constant $\gamma$ such that

$$\left(\frac{(\hat{y}^k - y^k)^T F'(y^k)}{\|\hat{h}^k\|}\right)^2 \leq \gamma\left(F(y^k) - F(y^{k+1})\right).$$

Since $\{\hat{h}^k\}$ is a bounded sequence, there exists a $\kappa$ such that $\|\hat{h}^k\| \leq \kappa$. Therefore, we have

$$\left(\hat{y}^k - y^k\right)^T F'(y^k) \leq \sqrt{\gamma}\,\kappa\sqrt{F(y^k) - F(y^{k+1})} . \tag{17}$$

If $(\hat{y}^k - y^k)^T F'(y^{k+1}) < \frac{1}{2}(\hat{y}^k - y^k)^T F'(y^k) < 0$ then

$$g(\theta) := \left(\hat{y}^k - y^k\right)^T F'\left(y^k + \theta(\hat{y}^k - y^k)\right)$$

$$< \frac{1}{2}\left(\hat{y}^k - y^k\right)^T F'(y^k), \quad \text{for } 0 \leq \theta \leq \alpha_1^k, \tag{18}$$

since $g(\theta)$ is a monotone function of $\theta$. By the Mean Value Theorem, there exists $0 \leq \theta_k \leq \alpha_1^k$ such that

$$\begin{aligned}
F(y^k) - F(y^{k+1}) &= -g(\theta_k)\alpha_1^k > -\frac{1}{2}\alpha_1^k \\
&\quad \left(\hat{y}^k - y^k\right)^T F'(y^k) \\
&= \frac{1}{2}\alpha_1^k\left(y^k - \hat{y}^k\right)^T F'(y^k).
\end{aligned} \tag{19}$$

Therefore, we have from (19)

$$\hat{h}^{kT} A_I A_I^T \hat{h}^k < \frac{2}{\alpha_1^k}\left(F(y^k) - F(y^{k+1})\right). \tag{20}$$

From (17) we get

$$\hat{h}^{kT} A_I A_I^T \hat{h}^k < \sqrt{\gamma}\,\kappa\sqrt{F(y^k) - F(y^{k+1})} . \tag{21}$$

Combining these two inequalities, we have, for $k \geq k_0$

$$\hat{h}^{kT} A_I A_I^T \hat{h}^k \leq \left(\frac{2}{\alpha_1^k} + \sqrt{\gamma}\,\kappa\right)\left(\left(\sqrt{F(y^k)F(y^{k+1})}\right. \right.$$

$$\left.\left. + \left(F(y^k) - F(y^{k+1})\right)\right)\right). \tag{22}$$

Since $A_I A_I^T$ is symmetric positive semidefinite, there exists a positive constant $\delta_k$ (depending only on $A_I A_I^T$) such that (cf. Luo and Luo, 1994)

$$w^T A_I A_I^T w \geq \delta_k \|A_I A_I^T w\|^2, \quad \text{for } w \in \mathbb{R}^m. \tag{23}$$

Since there are only finitely many distinct $I$, we can choose $\delta_k \equiv \delta > 0$ such that (23) holds for all $k$.

Therefore, for $k \geq k_0$,

$$\begin{aligned}
\|F'(y^k)\|^2 &= \|A_I A_I^T \hat{h}^k\|^2 \\
&\leq \frac{1}{\delta}\hat{h}^{kT} A_I A_I^T \hat{h}^k \\
&\leq \frac{1}{\delta}\left(\frac{2}{\alpha_1^k} + \frac{\sqrt{\gamma}\,\kappa}{\alpha_1^k}\right) \\
&\quad \times \left(\left(\sqrt{F(y^k) - F(y^{k+1})}\right) + \left(F(y^k)\right.\right. \\
&\qquad \left.\left. - F(y^{k+1})\right)\right).
\end{aligned}$$

Since $\lim_{k\to\infty}(F(y^k) - F(y^{k+1})) = 0$, we obtain $\lim_{k\to\infty}\alpha_1^k\|F'(y^k)\| = 0$. Now $d^k = A\hat{h}^k$ and, assume $\lim_{k\to\infty}\|F'(y^k)\| = c_1 > 0$, which requires that $\lim_{k\to\infty}\alpha_1^k = 0$. However this is impossible since $\{d^k\}$ is a sequence uniformly bounded above. Hence, we have $\lim_{k\to\infty}\|F'(y^k)\| = 0$. By Lemma 4.2, $\lim_{k\to\infty} F(y^k) = F_{\min}$.

If there is no $k_0$ such that $y^{k+1} = y^k + \alpha_1^k\hat{h}^k$ for $k \geq k_0$, then there exists a subsequence $\{k_j\}$ such that $y^{k+1} = y^k + \alpha^k h^k$ for $k = k_j$, $j = 1,2,\dots$. Since $A_I A_I^T$ is nonsingular in this case, and there are only finitely many different $I$ values, by Lemma 4.3, $\lim_{k\to\infty} F(y^k) = F_{\min}$. This completes the proof of the claim.

By Lemma 4.4 there exists $k^*$ such that for $k \geq k^*$ $y^k \in \mathscr{C}_I$ with $I = I(y^k)$ implies $\mathscr{C}_I \cap W^* \neq 0$. Let $k \geq k^*$ and $y^k \in \mathscr{C}_I$. Now, from Lemma 4.6 it follows that either $y^k + h^k$ (or, $y^k + \hat{h}^k$) is a minimizer and the algorithm stops, or $I(y^{k+1})$ is an extension of $I(y^k)$ with $y^{k+1} = y^{k+1} = y_k + \alpha_1^k\hat{h}^k$. Hence, the above argument can be repeated with $y^k$ replaced by $y^{k+1}$. Since the active set has finite cardinality, the algorithm must terminate in a finite number of iterations with a minimizer of $F$. This is clearly a contradiction and the theorem is proved. $\square$

## 5. Experimental results

In this section we report our computational experience with the Newton algorithm. We have coded the algorithm in Fortran for dense matrices using an already existing robust linear regression code in Madsen and Nielsen (1990).

The major effort in the algorithm is spent in solving the system (11). We use the AAFAC package of Nielsen (1990) to perform this. The solution is

obtained via an $LDL^T$ factorization of the matrix $C^k = A_I A_I^T$, so $D$ and $L$ are computed directly from the active columns of $A$, i.e., without squaring the condition number as would be the case if $C^k$ was first computed. The efficiency of the Newton algorithm depends critically on the fact that the difference between the active set $I(y^k)$ and $I(y^{k-1})$ is caused by a few elements. This implies that the factorization of $C^k$ can be obtained by relatively few up- and down-dates of the factorization of $C^{k-1}$. Including the cost of the matrix vector multiplications $A_I s_I$ and $Ah$ ($A\hat{h}$), and the line search the computational cost of a typical iteration is proportional to $(m^2 + 2mn + n)$. Occasionally, a refactorization is performed. The refactorization is an $O(m^3)$ process. This consists in the successive updating $LDL^T \leftarrow LDL^T + a_j a_j^T$ for all $j$ in the active set (starting with $L = I$, $D = 0$). It is considered only when some columns of $A$ leave the active set, i.e., when downdating is involved. If many columns leave we may refactorize because it is cheaper. This part of the code is based on ideas from Fletcher and Powell (1974) and Gentleman (1973).

The stopping criteria in the Newton algorithm are implemented as follows. The iterate $y + h$ is considered to be in $\mathscr{C}_I$ if

$$\left[\forall i \in I(y) : s_i(y) - (A^T h)_i \leq \tau\right]$$

and

$$\left[\forall i \notin I(y) : s_i(y) - (A^T h)_i > \tau\right].$$

Here, $\tau \approx O(\epsilon_M \|c\|_\infty)$ is used to take into account effects of rounding errors; $\epsilon_M$ denotes unit roundoff of the computer.
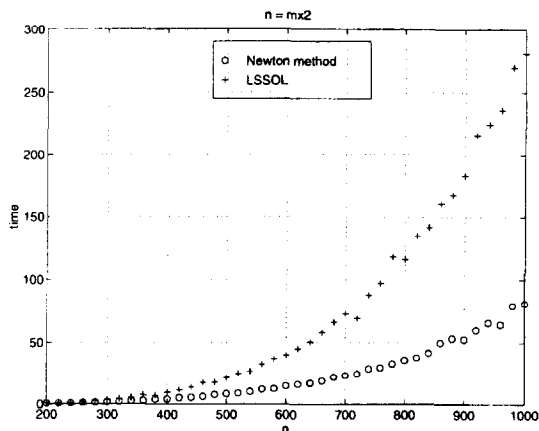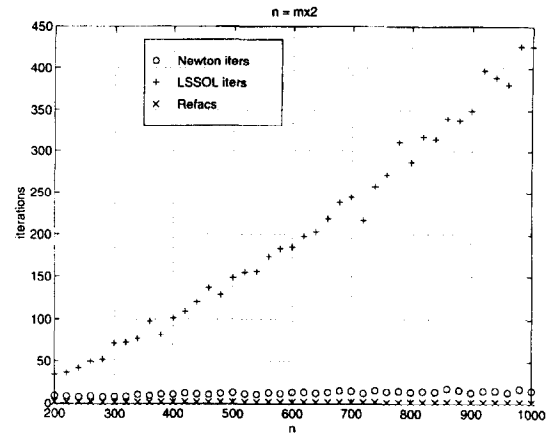


Fig. 2. Iteration results of the new method and LSSOL ($n/m = 2$).

We compare the new method with the library routine LSSOL from Stanford University Systems Optimization Laboratory available in the NAG (1987) subroutine library. LSSOL contains implementations of a numerically stable active set strategy and the simplex method for the solution of linear and quadratic programs and least squares problems. It can also find a feasible solution to a system of linear inequalities by minimizing the sum of infeasibilities. Since it does not exploit sparsity of $A$, it makes a fair comparison to our method.

To illustrate the performances of the new method and LSSOL, we generated two sets of consistent systems of dense linear inequalities randomly. In the first set, we keep the ratio $n/m$ constant at two and increase $m$ from 100 to 500 in steps of 10. We



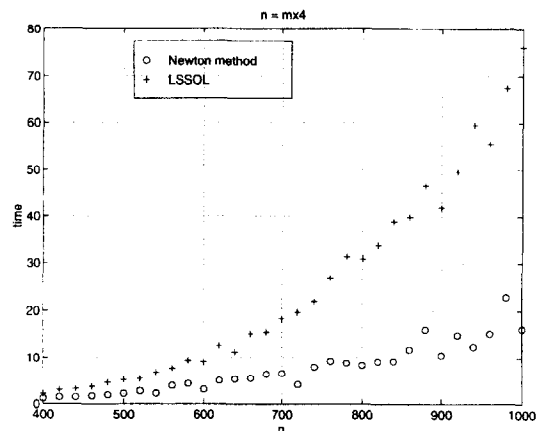Fig. 1. Run time results of the new method and LSSOL ($n/m = 2$).



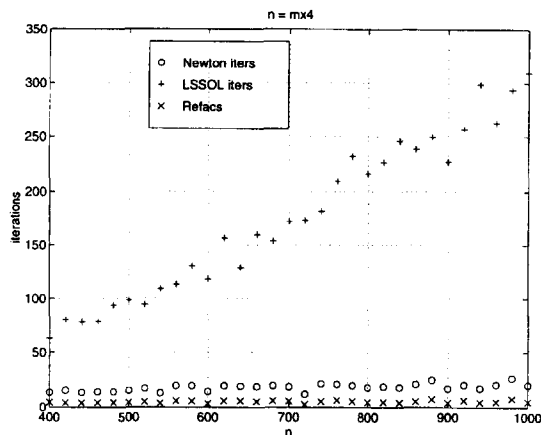Fig. 3. Run time results of the new method and LSSOL ($n/m = 4$).

Fig. 4. Iteration results of the new method and LSSOL ($n / m = 4$).

generate 10 problem instances for each pair $(m,n)$ and solve these instances using the Newton method and LSSOL, respectively. This results in the solution of 410 problems by both methods. In the second set of problems, we keep the ratio $n/m$ constant at four and vary $m$ from 100 to 250 in steps of five. Hence, we solve 310 problems by both methods. In Figs. 1–4 we report the average statistics of these runs. In Figs. 1 and 3 we plot the average run time in CPU seconds while in Figs. 2 and 4 we plot the average number of iterations of the Newton method and LSSOL along with the number of refactorizations of $A_I A_I^T$ in connection with the computations of the factors $L$ and $D$.

The absolute level of infeasibility upon termination of the Newton algorithm is measured as the $\ell_\infty$-norm of the vector $(A^T y - c)_+$ where $t_+ = \max(0, t)$ and the plus operator is applied componentwise. This measure was found to be always less than or in the vicinity of $10^{-13}$ in all test runs. These results demonstrate that the Newton algorithm is able to find an accurate feasible solution to linear inequalities efficiently. It attains a speed-up of three to four over LSSOL as the problem size is increased. This is a result of two factors:

1. The Newton method displays a slow growth rate in the number of iterations as the problems get larger.
2. The Newton method uses an almost constant number of refactorizations as the problem size is increased.

It seems that a sparse implementation of the Newton method will be a worthwhile research effort in the future.

## Acknowledgements

## References

Bramley, R., Winnicka, B., 1996. Solving linear inequalities in a least squares sense. SIAM Journal on Scientific Computing 17, 275–286.

Censor, Y., Zenios, S.A., 1997. Parallel Optimization: Theory, Algorithms and Applications. Oxford University Press, Oxford.

Chen, Ch., Mangasarian, O.L., 1995. Smoothing methods for convex inequalities and linear complementarity problems. Mathematical Programming 71, 51–69.

Clark, D.I., Osborne, M.R., 1986. Finite algorithms for Huber's $M$-estimator. SIAM Journal on Scientific and Statistical Computing 7, 72–85.

Ekblom, H., 1988. A new algorithm for the Huber estimator in linear models. BIT 28, 123–132.

Fletcher, R., Powell, M.J.D., 1974. On the modification of LDLT factorizations. Mathematics of Computation 28, 1067–1087.

Gentleman, W.M., 1973. Least squares computation by Givens transformations without square roots. Journal of the Institute of Mathematics and its Applications 12, 329–336.

Gill, P.E., Hammarling, S.J., 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. Technical report SOL 86-1. Systems Optimization Laboratory, Department of Operations Research, Stanford University, Stanford, California.

Han, S.-P., 1980. Least-squares solution of linear inequalities. Technical report TR-2141. Mathematics Research Center, University of Wisconsin-Madison.

Herman, G.T., 1980. Image Reconstruction from Projections: The Fundamentals of Computerized Tomography. Academic Press, New York.

Huber, P.J., 1981. Robust Statistics. John Wiley, New York.

Katznelson, S., 1965. An algorithm for solving nonlinear resistor networks. Bell System Technical Journal 44, 1605–1620.

De Leone, R., Mangasarian, O.L., 1987. Serial and parallel solution of large scale linear programs by augmented Lagrangian sucessive overrelaxation. In: Lecture Notes in Control and Information Sciences, Springer-Verlag.

Li, W., Swetits, J., 1993. A Newton method for convex regression, data smoothing, and quadratic programming with bounded constraints. SIAM Journal on Optimization 3, 466–468.

Li, W., Swetits, 1997. A new algorithm for strictly convex quadratic programming. SIAM Journal on Optimization (to appear).

Li, W., 1997. Conjugate gradient algorithms for convex quadratic splines. Mathematical Programming (to appear).

Lin, Y., Pang, J.-S., 1987. Iterative methods for large convex quadratic programs: A survey. SIAM Journal on Control and Optimization 25, 383–411.

Luo, X.-D., Luo, Z.-Q., 1994. Extension of Hoffman's error bound to polynomial systems. SIAM Journal on Optimization 4, 383–392.

Madsen, K., Nielsen, H.B., 1990. Finite algorithms for robust linear regression. BIT 30, 682–699.

Mangasarian, O.L., 1981. Iterative solution of linear programs. SIAM Journal on Numerical Analysis 18, 606–614.

Mangasarian, O.L., 1984. Normal solution of linear programs. Mathematical Programming Study 22, 206–216.

NAG, 1987. Fortran library routine document. Numerical Analysis Group.

Nielsen, H.B., 1990. AAFAC: A package of Fortran 77 subprograms for solving $A^TAx = c$. Report NI-90-11. Institute for Numerical Analysis, Technical University of Denmark.

Rockafellar, R.T., 1987. Linear-quadratic programming and optimal control. SIAM Journal on Control and Optimization 25, 781–814.

Rockafellar, R.T., 1990. Computational schemes for large-scale problems in extended linear-quadratic programming. Mathematical Programming, 447–474.

Rockafellar, R.T., 1976. Augmented Lagrangians and applications of the proximal point algorithm in convex programming. Mathematics of Operations Research 1, 97–116.

Sun, J., 1992. On the structure of convex piecewise-quadratic functions. Journal of Optimization Theory and Applications 72, 499–510.