

# Linear Programming via a Quadratic Penalty Function<sup>1</sup>

MUSTAFA Ç. PINAR

IE Department, Bilkent University, 06533 Bilkent, Ankara, Turkey

*Abstract:* We use quadratic penalty functions along with some recent ideas from linear  $l_1$  estimation to arrive at a new characterization of primal optimal solutions in linear programs. The algorithmic implications of this analysis are studied, and a new, finite penalty algorithm for linear programming is designed. Preliminary computational results are presented.

*Key Words:* Quadratic Penalty Functions. Linear Programming. Linear  $l_1$  Estimation. Characterization. Finite Algorithms.

## 1 Introduction

We consider the primal linear programming problem

$$\begin{array}{ll} \text{[P]} & \underset{x}{\text{minimize}} \quad c^T x \\ & \text{subject to} \quad Ax = b \\ & \quad \quad \quad x \geq 0 \end{array}$$

where  $x \in \Re^n$ ,  $A$  is a  $m \times n$  matrix,  $b \in \Re^m$  and  $c \in \Re^n$ , and its dual:

$$\begin{array}{ll} \text{[D]} & \underset{y}{\text{maximize}} \quad -b^T y \\ & \text{subject to} \quad A^T y + c \geq 0 \end{array}$$

where  $y \in \Re^m$ .

---

<sup>1</sup> Research supported by grant No. 11-0505 from the Danish Natural Science Research Council SNF.

The first purpose of this paper is to give a new characterization of optimal solutions to a linear program using quadratic penalty functions and some recent ideas from linear  $l_1$  estimation. The second purpose is to investigate the algorithmic implications of this result for linear programming. Consider the following piecewise quadratic functional:

$$F(x, t) \equiv tc^Tx + \frac{1}{2}r^T(x)r(x) + \frac{1}{2}x^T\Theta(x)x, \quad (1)$$

where  $r(x) = Ax - b$ ,  $t$  is a positive scalar and  $\Theta(x)$  is a diagonal matrix with diagonal entries  $\Theta_{ii}$ :

$$\Theta_{ii}(x) = \begin{cases} 1 & \text{if } x_i < 0 \\ 0 & \text{otherwise} \end{cases}, \quad (2)$$

and the unconstrained minimization problem:

$$[\mathbf{CP}] \quad \min_{x \in \mathbb{R}^n} F(x, t) \quad (3)$$

for decreasing positive values of  $t$ . Let  $x_t$  denote a minimizer of  $F(x, t)$ . It is well-known [4] that

$$\lim_{t \rightarrow 0} c^Tx_t = f^*, \quad (4)$$

where  $f^*$  is optimal value in  $[\mathbf{P}]$ .

In the present paper we characterize the solution set of  $[\mathbf{CP}]$  and show that an optimal solution of the linear program can be obtained by following any one of the infinitely many piecewise linear paths that lead to the solution set of  $[\mathbf{P}]$ . This leads to a characterization of the solution set of  $[\mathbf{P}]$ . I.e., we give a description of the solution set of  $[\mathbf{P}]$  using information from the minimization of the unconstrained function for sufficiently small  $t > 0$ . To the best of our knowledge, this is the first such result in the literature for linear programs. Following this analysis, we define a new penalty algorithm for linear programs, and analyze its finiteness. The algorithm produces infeasible primal and feasible dual iterates. The primal feasibility is obtained upon termination. A preliminary implementation and numerical results are discussed at the end of the paper. For previous work on penalty methods for linear programming see also [1, 2, 5, 11].

This analysis is made possible by adapting some recent ideas from linear  $l_1$  estimation [10]. In [10] the following problem was tackled:

$$[\mathbf{L1}] \quad \min G(x) \equiv \|E^Tx - d\|_1, \quad (5)$$

where  $E$  is  $n \times m$ ,  $d \in \mathfrak{R}^n$ . In [10] a smooth approximation of [L1] was considered:

$$[\text{SL1}] \quad \min G_\gamma(x) \equiv \sum_{i=1}^n \rho(z_i(x)) , \quad (6)$$

where  $z_i(x) = e_i^T x - d_i$ ,

$$\rho(z_i) = \begin{cases} \frac{1}{2\gamma} z_i^2 & \text{if } |z_i| \leq \gamma \\ |z_i| - \frac{\gamma}{2} & \text{otherwise} . \end{cases} \quad (7)$$

The function  $\rho$  is known as the ‘‘Huber’’ function in robust regression [8]. Clearly,  $G_\gamma$  is also a piecewise quadratic functional. By analyzing the behavior of the set of minimizers of  $G_\gamma$  for decreasing values of  $\gamma$ , characterizations of the solutions sets of both [L1] and [SL1] were given in [10]. It is precisely against this background that we develop our results in the present paper. Our proofs follow the same lines as in [10] with the necessary modifications. The contribution of the paper is to broaden the domain of application of these recent ideas, and in the process to obtain new results on the linear programming problem. For an alternative dual approach, the reader is referred to [15].

## 2 Primal Pathways to Optimal Solutions

We will assume throughout the paper that  $A$  has rank  $m$ , and that  $A$  contains no row or column that is identically zero. The following result shows that the unconstrained minimization of  $F$  is well-defined.

*Theorem 1: If [P] has a finite optimal value there exists a finite point that minimizes  $F(x, t)$  for all positive  $t$ .*

*Proof:* Assume that the conclusion is false, i.e., that there exists a sequence of points  $\{x_i\}$  with  $\lim_{i \rightarrow \infty} \|x_i\| = +\infty$  such that  $\lim_{i \rightarrow \infty} tc^T x_i = -\infty$ . The unboundedness of  $F$  also implies that there does not exist  $j \in \{1, \dots, m\}$  where  $\lim_{i \rightarrow \infty} |r_j(x_i)| = \infty$  and there does not exist  $i \in \{1, \dots, n\}$  where  $\lim_{i \rightarrow \infty} x_i = -\infty$ . However, this implies that [P] is unbounded since  $t > 0$ , which contradicts our assumption that [P] has a finite optimal value. ■

Define

$$\Phi(x) = \frac{1}{2}r^T(x)r(x) + \frac{1}{2}x^T\Theta(x)x .$$

The following is well-known; see e.g., [4, 14]:

$$\lim_{t \rightarrow 0} \Phi(x_t) = 0 . \quad (8)$$

Let us now define a *binary vector*  $\theta \in \mathfrak{R}^n$  where the entries are either 0 or 1 according to the rule:

$$\theta_i(x) = \begin{cases} 1 & \text{if } x_i < 0 \\ 0 & \text{otherwise} . \end{cases} \quad (9)$$

Hence the diagonal matrix  $\Theta$  defined earlier in (2) can be expressed

$$\Theta = \text{diag}(\theta_1, \dots, \theta_n) .$$

In what follows  $\Theta x$ , and  $\Theta \cdot x$  are both used to denote the multiplication of a vector  $x$  with the diagonal matrix  $\Theta$  to avoid confusion with  $\Theta(x)$  where  $x$  is the argument of  $\Theta$ . We denote by  $X$  the set of optimal solutions to  $[\mathbf{P}]$ .

### 2.1 The Minimizers of $F$

We observe that  $F(x, t)$  is composed of a finite number of quadratic functions. In each domain  $D \subseteq \mathfrak{R}^n$  where  $\theta(x)$  is constant  $F$  is equal to a specific quadratic function as seen from its definition. These domains are separated by the union of hyperplanes,

$$B = \{x \in \mathfrak{R}^n | \exists i: x_i = 0\} . \quad (10)$$

Given a point  $x \in \mathfrak{R}^n$  and the associated binary vector  $\theta(x)$   $Q_\theta$  is the quadratic function which equals  $F$  on the subset

$$\mathcal{C}_\theta = \text{cl}\{z \in \mathfrak{R}^n | \theta(z) = \theta\} . \quad (11)$$

$\mathcal{C}_\theta$  corresponds to an orthant of  $\mathfrak{R}^n$ . Notice that any  $x \in \mathfrak{R}^n \setminus B$  has exactly one corresponding orthant whereas a point  $x \in B$  belongs to two or more orthant. Therefore, we must specify a binary vector  $\theta$  in addition to  $x$  in order to specify which quadratic function we are currently considering as representative of  $F$ .

$Q_\theta$  can be defined as follows:

$$Q_\theta(z, t) = \frac{1}{2}(z - x)^T(A^T A + \Theta)(z - x) + F'^T(x, t)(z - x) + F(x, t) . \quad (12)$$

The gradient of the function  $F(x, t)$  is given by

$$F'(x, t) = (A^T A + \Theta)x - A^T b + tc . \quad (13)$$

For  $x \in \mathfrak{R}^n \setminus B$ , the Hessian of  $F(x, t)$  exists, and is given by

$$F''(x, t) = A^T A + \Theta . \quad (14)$$

The set of minimizers of  $F(x, t)$  is denoted by  $M_t$ . Now, we have the following lemma.

*Lemma 1: Let  $P \equiv A^T A + \Theta$ . Then for any  $x \in \mathfrak{R}^n$  the following holds:*

$$Px = 0 \Rightarrow \Theta x = 0 .$$

*Proof:*

$$x^T Px = (Ax)^T(Ax) + x^T \Theta x = \|Ax\|_2^2 + x^T \Theta x .$$

Suppose

$$\Theta_{ii} = \begin{cases} 1 & \text{for } i \in S_1 \\ 0 & \text{for } i \in S_0 . \end{cases}$$

Then, we have

$$(\Theta x)_i = \begin{cases} x_i & \text{for } i \in S_1 \\ 0 & \text{for } i \in S_0 , \end{cases}$$

and

$$x^T \Theta x = \sum_{i \in S_1} x_i^2 .$$

Now, since  $\|Ax\|_2 \geq 0$  and  $x^T \Theta x \geq 0$  it follows that

$$Px = 0 \Rightarrow x^T Px = 0 \Rightarrow Ax = 0 \quad \text{and} \quad x^T \Theta x = 0 .$$

But

$$x^T \Theta x = 0 \Rightarrow \forall i \in S_1 x_i = 0 \Rightarrow \Theta x = 0 .$$

■

*Lemma 2:  $\theta(x_t)$  is constant for  $x_t \in M_t$ . Furthermore  $(x_t)_i$  is constant for  $x_t \in M_t$  if  $\theta_i = 1$ .*

*Proof:* Let  $x_t \in M_t$  and let  $\theta = \theta(x_t)$ , i.e.,  $F(x, t) = Q_\theta(x, t)$  for  $x \in \mathcal{C}_\theta$ . If  $x \in \mathcal{C}_\theta \cap M_t$  then  $Q'_\theta(x - x_t) = 0$ . Therefore, if  $(x_t)_i < 0$  then  $x_i - (x_t)_i = 0$  by the previous Lemma. Thus  $x_i$  is constant in  $\mathcal{C}_\theta \cap M_t$ . Using the fact that  $M_t$  is connected and  $x_i$  is continuous, it is easily seen by repeating the argument above that  $x_i$  is constant in  $M_t$ . Next suppose  $(x_t)_i \geq 0$ . Then  $x_i \geq 0$  for all  $x \in M_t$  because existence of  $x \in M_t$  with  $x_i < 0$  is excluded by the convexity of  $M_t$ , continuity of  $x_i$ , and the first part of the lemma. This completes the proof. ■

Following the lemma we use the notation  $\theta(M_t) = \theta(x_t)$ ,  $x_t \in M_t$  as the binary vector corresponding to the solution set. Now, Lemma 2 has the following consequences which characterize the solution set  $M_t$ .

*Corollary 1:  $M_t$  is a convex set which is contained in one orthant:  $\mathcal{C}_\theta$  where  $\theta = \theta(M_t)$ .*

*Proof:* Follows immediately from the linearity of the problem and Lemma 2. ■

*Corollary 2: Let  $x_t \in M_t$ , and  $\theta = \theta(M_t)$ . Let  $\mathcal{N}_\theta$  be the null space of  $P \equiv A^T A + \Theta$ , where  $\Theta = \text{diag}(\theta_1, \dots, \theta_n)$ . Then*

$$M_t = (x_t + \mathcal{N}_\theta) \cap \mathcal{C}_\theta .$$

*Proof:* It follows from (13) that  $F'(x_t + u, t) = 0$  if  $u \in \mathcal{N}_\theta$  and  $x_t + u \in \mathcal{C}_\theta$ . Thus

$$M_t \supseteq (x_t + \mathcal{N}_\theta) \cap \mathcal{C}_\theta .$$

If  $x \in M_t$  then by the previous corollary  $x \in \mathcal{C}_\theta$ . Also,  $F'(x, t) = 0$ . This implies that  $P(x - x_t) = 0$ . Therefore, we have

$$M_t \subseteq (x_t + \mathcal{N}_\theta) \cap \mathcal{C}_\theta$$

which proves the result. ■

An important consequence of the previous characterization of  $M_t$  is that it provides a sufficient condition for the uniqueness of  $x_t$ .

*Corollary 3:* Let  $\theta = \theta(M_t)$ .  $x_t \in M_t$  is unique if  $\text{rank}(A^T A + \Theta) = n$ .

Notice that this condition is not necessary for uniqueness of the minimizer in [CP] as the following example demonstrates:

*Example 1:* Consider the linear program of the form [P] where

$$A = \begin{bmatrix} 1 & 2 & 3 & 1 & 0 & 0 \\ 4 & 5 & 6 & 0 & -1 & 0 \\ 7 & 8 & 9 & 0 & 0 & 1 \end{bmatrix}$$

and  $b = (7, 5, 10)^T$ , and  $c = (-75, -87, -102, 0, 5, -8)^T$ . For  $t = 1$ , the unique minimizer of  $F$  occurs at  $x_t = (-1, -2, 4, 0, 0, 5)^T$ , where  $\text{rank}(A^T A + \Theta) = 5$ .

## 2.2 Characterization of Optimal Solutions

In this section we show how the solution set  $M_t$  approximates the solution set  $X$  of [P] as  $t$  approaches 0.

Assume  $x_t \in M_t$ , and let  $\theta = \theta(M_t)$ . Let  $\mathcal{N}_\theta$  be defined as in Corollary 2.

*Lemma 3:* Let  $x_t \in M_t$ , and  $\theta = \theta(M_t)$ . Then the following linear system is consistent,

$$(A^T A + \Theta)d = c . \tag{15}$$

*Proof:* Since  $x_t$  satisfies the necessary condition for a minimizer, we have the following:

$$\mathbf{0} = (A^T A + \Theta)x_t - A^T b + tc . \quad (16)$$

Observing that  $\Theta \cdot \Theta = \Theta$ , (16) can be rewritten as:

$$(A^T \Theta) \begin{pmatrix} A \\ \Theta \end{pmatrix} x_t = (A^T \Theta) \begin{pmatrix} b \\ \mathbf{0} \end{pmatrix} - tc , \quad (17)$$

where  $\mathbf{0}$  denotes a vector identically zero in  $\mathfrak{R}^n$ . We observe that the system

$$(A^T \Theta) \begin{pmatrix} A \\ \Theta \end{pmatrix} h = (A^T \Theta) \begin{pmatrix} b \\ \mathbf{0} \end{pmatrix} \quad (18)$$

is consistent since it corresponds to normal equations for the overdetermined system:

$$\begin{pmatrix} A \\ \Theta \end{pmatrix} h = \begin{pmatrix} b \\ \mathbf{0} \end{pmatrix} . \quad (19)$$

Hence, the result. ■

Let  $d$  be a solution to (15). Then, it is easily verified by inserting (15) into (16) that  $x_t + td$  is the least squares solution to the overdetermined system of linear equations:

$$\begin{pmatrix} A \\ \Theta \end{pmatrix} h = \begin{pmatrix} b \\ \mathbf{0} \end{pmatrix} .$$

*Lemma 4:* Let  $x_t \in M_t$ , and  $\theta = \theta(M_t)$ . If the overdetermined system (19) is consistent then

$$\frac{1}{t}(Ax_t - b) = -Ad , \quad (20)$$

and



$$\frac{1}{t}\theta x_t = -\theta d \quad (21)$$

for any solution  $d$  to (15).

*Proof:* The proof follows by inserting the solution  $(x_t + td)$  into (19). ■

Now let  $d$  solve (15) and assume  $\theta(x_t + \varepsilon d) = \theta$ , i.e.,  $x_t + \varepsilon d \in \mathcal{C}_\theta$  for some  $\varepsilon > 0$ . The linearity of the problem implies  $x_t + \delta d \in \mathcal{C}_\theta$  for  $0 \leq \delta \leq \varepsilon$ . Therefore (16) and (15) show that  $(x_t + \delta d)$  is a minimizer of  $F(x, t - \delta)$ . Using Corollary 2 we have proved the following:

*Lemma 5:* Let  $x_t \in M_t$  and let  $\theta = \theta(M_t)$ . Let  $d$  solve (15). If  $\theta(x_t + \varepsilon d) = \theta$  for  $\varepsilon > 0$  then  $\theta(x_t + \delta d) = \theta$ , and

$$M_{t-\delta} = (x_t + \delta d + \mathcal{N}_\theta) \cap \mathcal{C}_\theta \quad (22)$$

for  $0 \leq \delta \leq \varepsilon$ .

*Theorem 2:* There exists  $t_0 > 0$  such that  $\theta(M_t)$  is constant for  $0 < t \leq t_0$ . Furthermore,

$$M_{t-\delta} = (x_t + \delta d + \mathcal{N}_\theta) \cap \mathcal{C}_\theta \quad \text{for } 0 \leq \delta < t \leq t_0$$

where  $\theta = \theta(M_t)$  and  $d$  solves (15).

*Proof:* Since there is only a finite number of different binary vectors the theorem is a consequence of the previous lemma. ■

The analysis shows that the minimizers of  $F$  form a family of piecewise-linear paths as a function of  $t$ .

*Corollary 4:*  $\theta(M_t)$  is a piecewise constant function of  $t$ .

*Corollary 5:* Let  $0 < t \leq t_0$ , where  $t_0$  is given in Theorem 2 and let  $\theta = \theta(M_t)$ . Then

$$\Theta \cdot (x_t + t\tilde{d}) = \mathbf{0} , \quad (23)$$

and

$$r(x_t + t\tilde{d}) = \mathbf{0} , \quad (24)$$

where  $\tilde{d}$  is any solution of (15). Furthermore,

$$\frac{1}{t}(Ax_t - b) = -A\tilde{d} , \quad (25)$$

and

$$\frac{1}{t}\Theta x_t = -\Theta\tilde{d} . \quad (26)$$

I.e.,  $r(x_t)/t$  and  $\Theta x/t$  are constant.

*Proof:* Let  $x_{t-\delta} \in M_{t-\delta}$  for  $0 \leq \delta < t$ . By Theorem 2 there exists  $d$  that solves (15) such that  $x_{t-\delta} = x_t + \delta d$ . Hence, there exists  $d^*$  that solves (15) such that  $x_t + \delta d^* \in M_{t-\delta}$  for all  $0 \leq \delta < t$ . Now, using (8)

$$\Theta \cdot (x_t + td^*) = \mathbf{0} . \quad (27)$$

Any solution  $\tilde{d}$  of (15) can be expressed as  $\tilde{d} = d^* + \eta$  where  $\eta \in \mathcal{N}(A^T A + \Theta)$ . Now, (23) follows from (27) and Lemma 1. Using (8) we have:

$$r(x_t + td^*) = \mathbf{0} . \quad (28)$$

Now, (24) follows since

$$(A^T A + \Theta)\eta = 0 \Rightarrow A^T A\eta = 0 \Rightarrow A\eta = 0 .$$

The second part follows from Lemma 4 since (23) and (24) imply that (19) is consistent. ■

We notice that if  $x_t \in M_t$  then  $y_t = r(x_t)/t$ , where  $\theta = \theta(M_t)$ , is feasible in  $[\mathbf{D}]$  as it is seen from (16). Now we recall a classical result from linear programming known as the complementary slackness theorem; see for instance [12].

*Theorem 3: Let  $x \in \mathbb{R}^n$  and  $y \in \mathbb{R}^m$  be feasible solutions in  $[\mathbf{P}]$  and  $[\mathbf{D}]$ , respectively. Then  $x$  and  $y$  are optimal solutions in their respective problems  $[\mathbf{P}]$  and  $[\mathbf{D}]$  if and only if the following conditions hold:*

$$0 < x_i \Rightarrow a_i^T y + c_i = 0 . \quad (29)$$

$$a_i^T y + c_i > 0 \Rightarrow x_i = 0 . \quad (30)$$

For the purposes of our next theorem we rewrite the constraints of  $[\mathbf{D}]$  in the form of equality constraints by introducing a non-negative vector  $u \in \mathbb{R}^m$ :

$$A^T y - u = -c .$$

Now, for  $x_t \in M_t$  we define

$$y_t \equiv r(x_t)/t , \quad (31)$$

and

$$u_t \equiv -\Theta x_t/t \quad (32)$$

where  $\theta = \theta(M_t)$ , and  $\Theta$  is defined accordingly. Then it is easy to see that  $(y_t, u_t)$  is feasible in  $[\mathbf{D}]$  from (16).

Let  $J_\theta = \{i | \theta_i = 0\}$ , and  $\mathcal{D}_\theta = \{x \in \mathbb{R}^n | x_i \geq 0 \wedge i \in J_\theta\}$ . Now we are ready to state the new characterization of optimal solution to  $[\mathbf{P}]$ .

*Theorem 4: Let  $0 < t \leq t_0$ , where  $t_0$  is given in Theorem 2 and let  $\theta = \theta(M_t)$ . Let  $x_t \in M_t$ , and  $d$  solve (15). Then*

$$M_0 \equiv X$$

where

$$M_0 = (x_t + td + \mathcal{N}_\theta) \cap \mathcal{D}_\theta \quad (33)$$

and

$$y^* = \frac{1}{t}r(x_t) , \quad u^* = -\frac{1}{t}\Theta x_t \quad (34)$$

solve [D].

*Proof:* First,  $M_0$  is non-empty as a consequence of the constant binary vector property of Theorem 2. Assume  $x_0 \in M_0$ . Then there exists a solution  $d_0$  to (15) such that  $x_0 = x_t + td_0$ . Therefore using Corollary 5

$$\Theta x_0 = \mathbf{0} . \quad (35)$$

Furthermore,  $(y^*, u^*)$  is feasible for [D]. Therefore, using (24), (32), and (35) we have

$$\begin{aligned} c^T x_0 &= x_0^T (-A^T y^* + u^*) \\ &= -x_0^T A^T y^* + x_0^T u^* \\ &= -b^T y^* . \end{aligned}$$

Now,  $x_0$  is a non-negative vector following (35) and the fact that  $x_0 \in \mathcal{D}_\theta$ . Hence,  $x_0$  and  $(y^*, u^*)$  are solutions to [P] and [D], respectively. Since this holds for any  $x_0 \in M_0$ ,  $M_0 \subseteq X$  and  $(y^*, u^*)$  solves [D].

If  $X$  is a singleton, the proof is complete. Therefore, assume the contrary. It remains to show that  $x \in M_0$  for any  $x \in X$ . Since  $x_0$  and  $(y^*, u^*)$  are primal-dual solutions it follows from Theorem 3 that  $\Theta x = \mathbf{0}$  for any  $x \in X$ . Now, let  $x \in X$  and  $x_t \in M_t$ . Hence,

$$\Theta x = \mathbf{0} . \quad (36)$$

Then using (16), (36), and the feasibility of  $x$  we have:

$$\begin{aligned} (A^T A + \Theta)(x - x_t) &= (A^T A + \Theta)x - (A^T A + \Theta)x_t \\ &= (A^T A + \Theta)x - (A^T b - tc) \\ &= tc , \end{aligned}$$

which shows that  $\frac{(x - x_t)}{t}$  solves (15). Therefore we have shown that  $x \in x_t + td + \mathcal{N}_\theta$ . Now, observing that  $x \in \mathcal{D}_\theta$  by virtue of feasibility the proof is complete. ■

Hence, all the optimal solutions to  $[\mathbf{P}]$  can be computed from any  $x_t \in M_t$  for  $t \in (0, t_0]$ . This can be performed – at least in theory – by choosing any solution  $d$  to (15) and varying  $\eta \in \mathcal{N}_\theta$  such that  $x_t + t_d + \eta \in \mathcal{D}_\theta$ .

Note that since  $r(x_t)/t$  and  $\Theta x_t/t$  are constant for all  $t \in (0, t_0]$ , no matter what  $x_t$  is picked in Theorem 4 the same pair  $(y^*, u^*)$  is obtained.

An immediate consequence of the characterization theorem is the following sufficiency condition for the uniqueness of solution in  $[\mathbf{P}]$ :

*Corollary 6:  $X$  is a singleton if  $\mathcal{N}_\theta = \{0\}$  where  $\theta = \theta(M_t)$  for  $t \in (0, t_0]$ .*

*Proof:* Since  $\mathcal{N}_\theta = \{0\}$   $x_t \in M_t$  is unique by Corollary 3. Hence  $(A^T A + \Theta)d = c$  has a unique solution,  $d_0$  say. Therefore,  $x_t + t d_0 + \mathcal{N}_\theta$  is a singleton. Hence, by Theorem 4,  $X$  is a singleton. ■

*Conjecture 1: The sufficiency condition for uniqueness of solution in the previous corollary is also necessary.*

### 3 Extended Binary Vectors

To inquire into the algorithmic implications of Theorem 4 in this section we define a new binary vector referred to as an “extended binary vector”. An “extended binary vector”  $\bar{\theta} \in \mathfrak{R}^n$  is defined as:

$$\bar{\theta}_i(x) = \begin{cases} 1 & \text{if } x_i \leq 0 \\ 0 & \text{otherwise} \end{cases} . \quad (37)$$

It is well-known that there should exist an optimal solution to  $[\mathbf{P}]$  where (at least)  $m$  components of  $x$  are zero (basic solutions), and the submatrix of  $A$  formed by picking the columns corresponding to the zero components of  $x$  has full rank. A similar property holds for the minimizers of  $F$ . Note that the two binary vector definitions only differ for those points that are on the boundary, i.e., for  $x \in B$ . We define the following *active set* of indices:

$$\mathcal{A}(x) \equiv \{i | 1 \leq i \leq n \wedge \bar{\theta}_i(x) = 1\} . \quad (38)$$

*Theorem 5: There exists a minimizer  $x_t$  of  $F(x, t)$  for which  $\text{rank}(A^T A + \bar{\Theta}(x_t)) = n$ .*

*Proof:* Let  $x$  be a minimizer of  $F$  where  $\text{rank}(A^T A + \bar{\Theta}(x)) < n$ . Therefore there exists a vector  $h \in \mathcal{N}(A^T A + \bar{\Theta}(x))$  with  $h \neq \mathbf{0}$ . Consider a point  $x + \alpha h$ ,  $\alpha \in \mathfrak{R}$ . By Lemma 1 if  $j \in \mathcal{A}(x)$ , then  $h_j = 0$ . This implies that  $x_j + h_j = x_j$ , and hence

$$\mathcal{A}(x + \alpha h) \supseteq \mathcal{A}(x) \quad (39)$$

for  $\alpha \in \mathfrak{R}$ . By definition of  $h$  and (13) it follows that if  $x + \alpha h \in \mathcal{C}_{\bar{\theta}}$  then  $F'(x + \alpha h) = F'(x) = \mathbf{0}$ , and hence we have

$$F(x + \alpha h, t) = F(x, t) \quad (40)$$

for  $x + \alpha h \in \mathcal{C}_{\bar{\theta}}$  where  $\bar{\theta} = \bar{\theta}(x)$ . By the definition of  $h$  there exists  $p \in \{1, \dots, n\} \setminus \mathcal{A}(x)$  such that  $h_p \neq 0$ . On the other hand, there exists  $\alpha \in \mathfrak{R}$  such that  $x_p + \alpha h_p = 0$ . Therefore, the active set must change along the line  $x + \alpha h$ ,  $\alpha \in \mathfrak{R}$ . The first time this happens when  $\alpha$  increases (or decreases) from zero, the point  $x + \alpha h$  is a minimizer of  $F$  as a result of (40). Further, (39) implies that the first change in the active set must be an expansion of the set. So far, it has been shown that if there exists a minimizer for which that matrix  $A^T A + \bar{\Theta}(x)$  has rank less than  $n$ , there exists another minimizer for which the corresponding active set has one more element. If the new matrix is also rank deficient, we can repeat the above process from the new point until we finally have an active set where the matrix  $A^T A + \bar{\Theta}$  has rank  $n$  since  $A^T A + I$  has rank  $n$ . ■

### 3.1 Behavior of the Set of Minimizers Near the Feasible Boundary

In this section we analyze the behavior of extended binary vectors associated with the minimizers of  $F(x, t)$  in the range  $(0, t_0]$  where  $t_0$  is as defined in Theorem 4. This is important in establishing the finite termination property of the penalty algorithm defined in section 4. First, we introduce some new concepts and definitions.

Let  $\sigma_{\theta} = \{i | \theta_i = 1\}$  for any binary vector  $\theta$ .

A “derived-extended-binary-subset” (*debs*)  $\mathcal{S}$  of a binary vector  $\theta$  (as defined in (9)) is a set of distinct extended binary vectors  $\bar{\theta}$  such that  $\sigma_{\bar{\theta}} \subseteq \sigma_{\theta}$  and there exists  $x \in \mathfrak{R}^n$  with  $\bar{\theta}(x) = \bar{\theta}$ .

An “extended-binary-set” (*eb*s)  $\mathcal{S}(M_t)$  of a set minimizers  $M_t$  is defined as the set of all distinct extended binary vectors corresponding to the elements of  $M_t$ . I.e., for any  $x_t \in M_t$   $\bar{\theta}(x_t) \in \mathcal{S}(M_t)$ . Since  $\theta(x_t)$  is constant for all  $x_t \in M_t$  clearly the *eb*s  $\mathcal{S}(M_t)$  of  $M_t$  is a *debs* of  $\theta(M_t)$  for any  $t > 0$ .

*Example 2:* In the problem of Example 1,  $\theta(M_t)$  remains constant at  $\theta \equiv (1, 1, 0, 0, 0, 0)$  for  $t \in (0, 1]$ . The sets

$$\mathcal{S}_1 \equiv \{(1, 1, 0, 0, 0, 0), (1, 1, 0, 0, 1, 0)\} ,$$

$$\mathcal{S}_2 \equiv \{(1, 1, 0, 0, 0, 0), (1, 1, 0, 1, 0, 0), (1, 1, 0, 0, 1, 0)\}$$

are sample derived-extended-binary-subset's of  $\theta = (1, 1, 0, 0, 0, 0)$ .

*Lemma 6:* If  $\mathcal{S}(M_{t_1}) = \mathcal{S}(M_{t_2})$  where  $0 < t_2 < t_1$  then  $\mathcal{S}(M_t) = \mathcal{S}(M_{t_1}) = \mathcal{S}(M_{t_2})$  for  $t_2 \leq t \leq t_1$ .

*Proof:* Let  $x_{t_1} \in M_{t_1}$ ,  $x_{t_2} \in M_{t_2}$  with  $\bar{\theta}(x_{t_1}) = \bar{\theta}(x_{t_2})$ . Define

$$x_t = (1 - \varepsilon)x_{t_2} + \varepsilon x_{t_1} ,$$

where  $\varepsilon = (t - t_2)/(t_1 - t_2)$ . Since  $0 \leq \varepsilon \leq 1$  it follows that  $\bar{\theta}(x_t) = \bar{\theta}(x_{t_1}) = \bar{\theta}(x_{t_2})$  and  $x_t$  satisfies the necessary condition (16) for a minimizer of  $F(x, t)$ . Now the result follows from the linearity of the problem and Lemma 2. ■

*Theorem 6:* There exists  $\bar{t}$  such that  $\mathcal{S}(M_t)$  is constant for  $t \in (0, \bar{t})$  where  $0 < \bar{t} \leq t_0$ .

*Proof:* Since  $\theta(M_t)$  remains constant in  $(0, t_0]$  following Theorem 2 and the number of different derived-extended-binary-subsets of  $\theta(M_t)$  is finite, the result is a consequence of the previous lemma. ■

*Example 3:* In the problem of Example 1,  $\theta(M_t)$  remains constant at  $\theta \equiv (1, 1, 0, 0, 0, 0)$  for  $t \in (0, 1]$  whereas the *ebs* remains constant at

$$\mathcal{S} \equiv \{(1, 1, 0, 0, 0, 0), (1, 1, 0, 0, 1, 0), (1, 1, 0, 0, 0, 1)\}$$

for  $t \in (0, 0.6875)$ .

*Theorem 7:* Let  $t \in (0, \bar{t})$  and  $x_t \in M_t$  with  $\bar{\theta} = \bar{\theta}(x_t)$ . Also, let  $y^* \equiv \frac{1}{t}r(x_t)$ , and  $u^* \equiv -\frac{1}{t}\bar{\theta}x_t$ . Then

$$\bar{\Theta} \cdot (x_t + td) = \mathbf{0} , \quad (41)$$

$$r(x_t + td) = \mathbf{0} , \quad (42)$$

and

$$b^T y^* + c^T (x_t + td) = 0 , \quad (43)$$

for any solution  $d$  to (15). Furthermore, if  $d$  is unique or  $x_t + td \geq \mathbf{0}$  then  $x_t + td$  solves  $[\mathbf{P}]$ .

*Proof:* Let  $t \in (0, \bar{t})$  and  $x_t \in M_t$  with  $\bar{\theta} = \bar{\theta}(x_t)$ , and  $\bar{\Theta} = \text{diag}(\bar{\theta}_1, \dots, \bar{\theta}_n)$ . Consider the system

$$(A^T A + \bar{\Theta})d = c . \quad (44)$$

This is a consistent system of linear equations as we have shown in Lemma 3. By Theorem 6 there exists  $x_t \in M_t$  such that  $\bar{\theta}(x_t) = \bar{\theta}$  for all  $t \in (0, \bar{t})$ . This implies that there exists  $d$  that solves (44) such that  $x_t + \delta d \in M_{t-\delta}$  for all  $\delta \in (0, t]$ . A consequence of this using (4) and (8) is that  $x_t + td$  solves  $[\mathbf{P}]$ , and

$$r(x_t + td) = \mathbf{0} ,$$

and

$$\bar{\Theta} \cdot (x_t + td) = \mathbf{0} .$$

Since  $d$  can be replaced by  $d + \eta$  in the above identity where  $\eta \in \mathcal{N}(A^T A + \bar{\Theta})$ , it follows that

$$r(x_t + td) = 0 , \quad (45)$$

and

$$\bar{\Theta} \cdot (x_t + td) = \mathbf{0} . \quad (46)$$

for any solution  $d$  to (44). Clearly, if the solution to (44) is unique,  $d^*$  say, then  $x_t + td^*$  solves  $[\mathbf{P}]$ .



Let

$$y^* = \frac{1}{t} r(x_t) ,$$

and

$$u^* \equiv -\frac{1}{t} \bar{\theta} x_t .$$

Let  $x_0 = x_t + td$ . Using (46) and (45) and since  $(y^*, u^*)$  is feasible for [D] we have, as in the proof of Theorem 4,

$$\begin{aligned} c^T x_0 &= x_0^T (-A^T y^* + u^*) \\ &= -x_0^T A^T y^* + x_0^T u^* \\ &= -b^T y^* . \end{aligned}$$

This completes the proof. ■

#### 4 The Penalty Algorithm

Based on the analysis of the previous sections, we now construct a penalty algorithm for linear programming.

We consider the following algorithm:

```

Choose  $t$  and compute a minimizer  $x_t$  of  $F$ 
while not STOP
    reduce  $t$ 
    compute a minimizer  $x_t$  of  $F$ 
end while.

```

In the above iteration STOP is a function that returns TRUE if the duality gap is zero (within rundoﬀ) and primal feasibility is achieved. Otherwise,  $t$  is decreased according to some criteria; see section 4.2. To complete the description

we need an algorithm to compute a minimizer of  $F$ . Such an algorithm is adapted from the Newton algorithm of [9] for robust linear regression using Huber functions. This algorithm is a standard Newton iteration with a simple line search to solve the nonlinear system of equations  $F'(x, t) = 0$ . However, special care must be given to the case where the matrix  $A^T A + \bar{\Theta}$  is rank-deficient. We give a brief description of the modified Newton algorithm below.

#### 4.1 Computing an Unconstrained Minimizer

The algorithm for computing a minimizer  $x_t$  of  $F$  is based on a modified Newton algorithm given in [9]. The idea is to inspect the orthants of  $\mathbb{R}^n$  to locate the orthant where the local quadratic  $Q_\theta$  contains its own minimizer. This is accomplished by means of the Newton iteration. At a given iterate, the Newton step is computed using the expansion (12) of  $F$ . If a unit step in this direction yields a point in the same orthant, then the global minimizer has been found. I.e., the quadratic representation of  $F$  which contains the global minimizer has been located. Otherwise, the algorithm proceeds with a line search.

A search direction  $h$  is computed by minimizing the quadratic  $Q_{\bar{\theta}}$  where  $\bar{\theta} = \bar{\theta}(x)$  and  $x$  is the current iterate. More precisely, we consider the equation

$$Q_{\bar{\theta}}'' h = -Q_{\bar{\theta}}'(x) \quad (47)$$

where  $Q_{\bar{\theta}}''$  and  $Q_{\bar{\theta}}'$  denote the Hessian and gradient of  $Q_{\bar{\theta}}$ , respectively. This system is expressed as

$$(A^T A + \bar{\Theta})h = -(A^T A + \bar{\Theta})x + A^T b - tc. \quad (48)$$

For ease of notation let  $C \equiv A^T A + \bar{\Theta}$  and  $g \equiv -Cx + A^T b - tc$ . Furthermore, let  $\mathcal{N}(C)$  denote the null space of  $C$ . If  $C$  has full rank, then  $h$  is the solution to (48). Otherwise, if the system of equations (48) is consistent, a minimum norm solution is computed. If the system is inconsistent the projection of  $g$  on  $\mathcal{N}(C)$  is computed. These choices are motivated and justified in [9]. The next iterate is found through a line search aiming for a zero of the directional derivative. This procedure is computationally cheap as a result of the piecewise-linear nature of  $F'$ . It can be shown using the analysis in [9] that the iteration is finite, i.e., after a finite number of iterations we have  $x + h \in C_{\bar{\theta}}$ . Therefore,  $x + h$  is a minimizer of  $F$  as a result of (11), (12) and the convexity of  $F$ . We summarize below the modified Newton algorithm:

```

repeat
   $\bar{\theta} = \bar{\theta}(x)$ 
  if (48) consistent then
    find  $h$  from (48)
    if  $x + h \in \mathcal{C}_{\bar{\theta}}$  then
       $x \leftarrow x + h$ 
      stop = true
    else
       $x \leftarrow x + \alpha h$  (line search)
    endif
  else
    compute  $h$  = null space projection of  $g$ 
     $x \leftarrow x + \alpha h$  (line search)
  endif
until stop.

```

#### 4.2 Reducing $t$

Let  $x_t$  be a minimizer of  $F(x, t)$  for some  $t > 0$  and  $\bar{\theta} = \bar{\theta}(x_t)$ . Consider the system

$$(A^T A + \bar{\theta})d = c. \quad (49)$$

Let  $d$  be a solution to (49). We distinguish between two cases:

*Case 1:* The duality gap  $c^T(x_t + td) + b^T y^*$  is zero but  $x_t + td$  is infeasible in  $[\mathbf{P}]$ , i.e., there exists  $j$  such that  $(x_t + td)_j < 0$ . In this case we reduce  $t$  as follows. Let  $\phi \equiv \{\alpha_k, k = 1, 2, \dots, q\}$  be the set of positive kink points where the components of  $x_t + td$  change sign, i.e., the set  $\phi = \{0 < \alpha < 1 \mid \exists i \in J \mid (x_t)_i + t\alpha d_i = 0\}$  where  $J = \{i \mid 1 \leq i \leq n \wedge d_i \neq 0\}$ . If  $\phi$  is non-empty we choose

$$\alpha^* = \min_k \alpha_k$$

and we let

$$t_{\text{next}} \equiv (1 - \alpha^*)t,$$

and

$$x_{t_{next}} \equiv x_t + \alpha^* t d$$

Otherwise, we let

$$t_{next} \equiv 0.9t ,$$

and

$$x_{t_{next}} \equiv x_t + 0.9td$$

In both cases,  $x_{t_{next}}$  is used as the starting point of the modified Newton algorithm of section 4.1 with the reduced value of  $t$ .

*Case 2:* The duality gap is not zero. This is an indication that  $t$  is not in the interval  $(0, t_0]$ . In this case we reduce  $t$  as follows. Let  $\phi \equiv \{\alpha_k, k = 1, 2, \dots, q\}$  be the set of positive kink points where the components of  $x_t + td$  change sign, i.e., the set  $\phi = \{0 < \alpha < 1 | \exists i \in J | (x_t)_i + t\alpha_i d_i = 0\}$  where  $J = \{i | 1 \leq i \leq n \wedge d_i \neq 0\}$ . The set  $\phi$  is non-empty as a consequence of Theorem 4. Let  $\alpha_1 = \min_{\alpha_i > 0, \alpha_i \in \phi} \alpha_i$  and  $\alpha_2 = \max_{\alpha_i > 0, \alpha_i \in \phi} \alpha_i$  and  $\alpha^* = \max\{0.1, 0.5(\alpha_1 + \alpha_2)\}$ . We use

$$t_{next} = (1 - \alpha^*)t ,$$

$$x_{t_{next}} \equiv x_t + \alpha^* t d .$$

For robustness we search only in the interval  $[0.1t, t]$  so that  $t_{next} \leq 0.9t$ .

## 5 Finite Convergence

In this section we show that the penalty algorithm of section 4 converges finitely. In the following analysis, an iteration of the algorithm means either a modified Newton iteration or an execution of the  $t$ -reduction procedure.

*Lemma 7:* Assume  $t \in (0, \bar{t})$ . Let  $x \in M_t$  with  $\bar{\theta} = \bar{\theta}(x)$ . Let  $d$  solve (49), and  $x_{next}$  be generated by one iteration of the penalty algorithm. Then either

$$x_{next} \equiv x + td \in X$$

and the algorithm stops, or

$$x_{next} \equiv x + \alpha^* td \in M_{t_{next}} ,$$

$$t_{next} = (1 - \alpha^*)t$$

where  $\alpha^*$  is as defined in Case 1 of the reduction procedure, and  $\mathcal{A}(x_{next})$  is an extension of  $\mathcal{A}(x)$ .

*Proof:* Let  $y = \frac{1}{t}r(x)$ . Clearly  $c^T(x_t + td) + b^T y = 0$  from Theorem 7. Hence we are in Case 1 of the reduction procedure of section 4.2. If  $x + td \geq 0$  then  $x_{next} \equiv x + td$  is a solution to [P] by Theorem 7 and the algorithm stops. Otherwise, Theorem 7 implies that  $\mathcal{A}(x) \subseteq \mathcal{A}(x + td)$ . Hence, using the definition of  $\alpha^*$ ,

$$\mathcal{A}(x + \alpha td) = \mathcal{A}(x)$$

for  $\alpha \in [0, \alpha^*)$ . Since there exists  $j \in \{1, \dots, n\} \setminus \mathcal{A}(x)$  such that  $(x + \alpha^* td)_j = 0$ ,  $\mathcal{A}(x + \alpha^* td)$  is an extension of  $\mathcal{A}(x)$ . Furthermore  $x + \alpha^* td \in \mathcal{C}_{\bar{\theta}}$ . Therefore, using the continuity of the gradient  $F'$ , (13) and the definition of  $d$ , we have

$$F'(x, t) = F'(x + \alpha^* td, (1 - \alpha^*)t) = 0 .$$

Thus,  $x_{next}$  minimizes  $F(x, (1 - \alpha^*)t)$ . ■

*Theorem 8: The penalty algorithm defined in section 4 terminates in a finite number of iterations with a primal-dual optimal pair.*

*Proof:* Let  $x \in M_t$  for some  $t > 0$ . Unless the stopping criteria are met and the algorithm stops with a primal-dual optimal pair,  $t$  is reduced by at least a factor of 0.9 as discussed in section 4.2. Since the modified Newton iteration of section 4.1 is a finite process,  $t$  will reach the range  $(0, \bar{t})$  where  $\bar{t}$  is as defined in Theorem 6 in a finite number of iterations. Now assume  $t \in (0, \bar{t})$ . From Lemma 7 either the algorithm terminates or the active set  $\mathcal{A}$  is expanded. Repeating this argument, in a finite number of iterations the matrix  $A^T A + \bar{\Theta}$  will finally have rank  $n$  since  $A$  has rank  $m$  and  $A^T A + I$  has rank  $n$ . When  $A^T A + \bar{\Theta}$  has full rank

the solution  $d$  to the system (49) is unique, and  $x_{next} = x + td$  solves [P] by Theorem 7. ■

## 6 Numerical Results

In this section we report our numerical experience with a preliminary implementation of the penalty algorithm, which does not exploit sparsity. The implementation was made using the matrix manipulation environment OCTAVE [6] on a SUN 4 Workstation. The purpose of the experiments is to test the viability of the algorithm in solving non-trivial problems. To accomplish this we choose a set of small test problems from the Netlib collection. To get an idea on the relative standing of the penalty algorithm we also compare our results to a linear programming simplex subroutine, E04MBF, from the NAG subroutine library. E04MBF is based on the package LSSOL from Stanford Systems Optimization Library. It is a Fortran 77 package for constrained linear least squares problems, linear programming and convex quadratic programming, [7]. It does not exploit sparsity. Hence, it provides a fair comparison to our numerical results. We perform this comparison only on basis of the number of iterations since (1) we do not yet have an implementation of our algorithm in Fortran 77, and (2) the cost per iteration of the simplex algorithm and the new penalty algorithm are comparable.

Note that the major effort in the Newton algorithm of section 4.1 is spent in solving the systems (48). It is observed that normally only a few entries of the diagonal matrix  $\bar{\Theta}$  change between two consecutive iterations. This implies that the factorization of  $C_k = A^T A + \bar{\Theta}_k$  at iteration  $k$  can be obtained by relatively few up- and down-dates of the factorization of  $C_{k-1}$ . Using the methods of [13] it can be verified that the computational cost of a typical iteration step is  $O(n^2)$ . Occasionally, a refactorization may be performed when there is indication of numerical instability or when the estimated computational cost of up- and down-dating the previous factorization outweighs the cost of a refactorization. This is an  $O(n^3)$  process. Since, a typical iteration of the simplex method involves  $O(m^2)$  operations, we can conclude that a typical iteration of the penalty method is somewhat more expensive than the simplex method for problems where  $n > m$ . In OCTAVE, we have not implemented the up- and down-dating of factorization. This will be done using the ideas of [13] in the future when we have a Fortran 77 implementation of the algorithm.

To initiate the algorithm, we choose a starting point  $x^0$  and  $t^0$  as follows. Let  $x$  be a solution to

$$(A^T A + I)x = A^T b - c .$$

Then  $t^0$  is chosen using the following formula:

$$t^0 = \beta \min_{x_i \neq 0} -x_i$$

where  $\beta \in (0, 1]$ . Then, we let  $x^0$  as the solution of

$$(A^T A + I)x^0 = A^T b - t^0 c .$$

Our test problem characteristics are described in Table 1 below (the source for Netlib is [3]). We consider seven problems from the Netlib collection. We also used a test problem from a civil engineering application at the Technical University of Denmark, referred to as plate. All the test problems are put into the form [P] using slack columns.

In Table 2, we show the solution statistics of the penalty method. The columns “iter” and “reduc” refer to the number of iterations and the number of reductions of the parameter  $t$ , respectively. The columns  $t^0$  and  $t^*$  report the initial and final values of  $t$ , respectively. Our stopping criteria are based on the relative duality gap:

**Table 1.** Characteristics of the test problems

Problem Name	Constraints	Variables	Slack Variables	Nonzeros in A
afiro	27	32	19	83
sc50b	50	48	30	118
sc50a	50	48	30	130
sc105	105	103	60	280
adlittle	56	97	41	383
stocfor1	117	111	54	447
blend	74	83	31	491
plate	61	73	60	209

**Table 2.** Solution statistics of the new penalty algorithm on the test set

Problem Name	iter	reduc	$t^0$	$t^*$
afiro	14	1	$2.46 \times 10^{-1}$	$2.46 \times 10^{-1}$
sc50b	30	1	$2.45 \times 10^0$	$2.45 \times 10^0$
sc50a	31	1	$4.47 \times 10^{-1}$	$4.47 \times 10^{-1}$
sc105	71	1	$2.41 \times 10^{-3}$	$2.41 \times 10^{-3}$
adlittle	143	11	$1.56 \times 10^{-2}$	$3.88 \times 10^{-6}$
stocfor1	62	7	$2.40 \times 10^{-3}$	$8.44 \times 10^{-4}$
blend	57	3	$2.08 \times 10^{-2}$	$4.13 \times 10^{-3}$
plate	7	2	$4.24 \times 10^{-2}$	$2.05 \times 10^{-2}$

$$\frac{|c^T x| - |b^T y|}{1 + |c^T x| + |b^T y|} ,$$

the values of the components of  $x$ , and the smallest component of the residual in the primal constraints:

$$r = Ax - b .$$

I.e., we stop when the above quantities are less than some tolerances. In all cases reported below, the relative duality gap tolerance is  $10^{-8}$ , and the feasibility tolerance is  $10^{-11}$ . In all test cases, the penalty algorithm achieves at least ten correct digits in the optimal objective function value with respect to the known optimal value [3].

We observe that the final value of the penalty parameter varies in the range between  $O(10^{-6})$  and  $O(1)$ . It is also noticed that the half of the problems were solved without the need to reduce  $t$ . The problem *adlittle* required the largest number of  $t$ -reduction steps.

In Table 3 below we provide a comparison of the number of iterations of the penalty algorithm with the code LSSOL of [7]. We note that the conclusions we make based on Table 3 are limited since the cost per iteration of the two algorithms are not identical. As we have remarked already, a typical iteration of the penalty algorithm involves operations proportional to  $O(n^2)$  whereas a typical iteration of the simplex algorithm can be performed in time proportional to  $O(m^2)$ .

We notice that the penalty method converged in smaller number of iterations than LSSOL. In particular, the convergence of the penalty algorithm on the civil engineering design problem *plate* is very fast (7 iterations) compared to the 150 iterations of LSSOL. The total number of iterations for the test set was 415 for the new algorithm, and 641 for LSSOL. This corresponds approximately to a value of 1.5 for the ratio LSSOL/new algorithm.

**Table 3.** Comparison of iteration numbers for the penalty method and LSSOL on the test set

Problem Name	Penalty Algorithm	LSSOL
afiro	14	19
sc50b	30	47
sc50a	31	40
sc105	71	78
adlittle	143	149
stocfor1	62	73
blend	57	85
plate	7	150
Total	415	641



## 7 Conclusions

We showed in this paper that the recent ideas from linear  $l_1$  estimation [10] can be successfully used to analyze primal continuous non-interior paths leading to the optimal set of a linear program. Interestingly, these paths yield new characterizations of optimal solutions. We adapted a finite algorithm to perform the unconstrained minimization of  $F$  from [9] where it was developed to solve problems of the form [SL1]. Using this Newton algorithm we defined a new penalty algorithm for linear programming and proved its finiteness. The computational results indicate that the algorithm is numerically stable and accurate. This suggests that further research is necessary to establish the true potential of the penalty algorithm. Several aspects of the algorithm need further study. Among those, the most important are:

- A careful implementation of numerical linear algebra
- Experimenting with different initialization procedures
- Experimenting with alternative  $t$ -reduction procedures.

*Acknowledgements:* I would like to thank K. Madsen and H. B. Nielsen for teaching me about their algorithm, and sharing their insights with me. In particular, K. Madsen arranged my two-year visit to the Institute of Mathematical Modelling where this work was done. Special thanks are also due to V. A. Barker for his careful reading of the manuscript.

## References

- [1] Bartels RH (1980) A penalty linear programming method using reduced-gradient basis-exchange techniques. *Linear Algebra Appl* 29:17–32
- [2] Bertsekas DP (1975) Necessary and sufficient conditions for a penalty method to be exact. *Mathematical Programming* 9:87–99
- [3] Bixby RE (1990) Implementing the simplex method: Part I: Introduction and Part II: The initial basis. Technical Report TR90-32, Rice University, (also in ORSA J on Computing)
- [4] Chebotarev SP (1973) Variation of the penalty coefficient in linear programming problems. *Automation and Control* 7:102–107
- [5] Conn AR (1976) Linear programming via a nondifferentiable penalty function. *SIAM J Numer Anal* 13:145–154
- [6] Eaton JW (1993) OCTAVE: A high-level interactive language for numerical computations. Manuscript University of Texas at Austin
- [7] Gill PE, Hammarling S, Murray W, Saunders MA, Wright MH (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, Stanford, CA
- [8] Huber P (1981) *Roubust statistics*. John Wiley, New York

- [9] Madsen K, Nielsen HB (1990) Finite algorithms for robust linear regression. *BIT* 30:682–699
- [10] Madsen K, Nielsen HB, Pinar MC (1993) New characterizations of  $l_1$  solutions of overdetermined linear systems. *Operations Research Letters* 16:159–166
- [11] Mangasarian OL (1986) Some applications of penalty functions in mathematical programming. *Lecture Notes in Mathematics*, No 1190. In: Conti R, De Giorgi E, Giannessi F (eds), *Optimization and Related Fields*. Springer-Verlag, Berlin 307–329
- [12] Murty KG (1976) *Linear and combinatorial programming*. John Wiley & Sons, New York
- [13] Nielsen HB (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
- [14] Pietrzykowski T (1961) Application of the steepest ascent method to linear programming. *PRACE ZAM A* 11
- [15] Pinar MC Piecewise-linear pathways to the optimal set in linear programming. Technical Report, Bilkent University, 1996

Received: February 1995

Revised version received: May 1995