Operations Research Letters 16 (1994) 159–166

operations research letters

New characterizations of ℓ_1 solutions to overdetermined systems of linear equations

Kaj Madsen[a], Hans Bruun Nielsen[a], Mustafa Ç. Pınar[b,*,1]

[a] *Institute of Mathematical Modelling, Numerical Analysis Group, Building 305, Technical University of Denmark, DK-2800 Lyngby, Denmark*

[b] *Department of Industrial Engineering, Bilkent University, Ankara, Turkey*

(Received 31 January 1994; revised 1 May 1994)

Abstract

New characterizations of the ℓ_1 solutions to overdetermined systems of linear equations are given. The first is a polyhedral characterization of the solution set in terms of a special sign vector using a simple property of the ℓ_1 solutions. The second characterization is based on a smooth approximation of the ℓ_1 function using a "Huber" function. This allows a description of the solution set of the ℓ_1 problem from any solution to the approximating problem for sufficiently small positive values of an approximation parameter. A sign approximation property of the Huber problem is also considered and a characterization of this property is given.

Key words: ℓ_1 optimization; Overdetermined linear systems; Non-smooth optimization; Smoothing; Huber functions; Characterization

1. Introduction

The main purpose of this work is to give new characterizations of solutions to the following non-smooth optimization problem:

$$\text{[L1]} \qquad \text{minimize } G(x) \equiv \| A^{\mathrm{T}}x - b \|_1, \tag{1}$$

where $x \in \mathbb{R}^n$, $b \in \mathbb{R}^m$ and $A \in \mathbb{R}^{n \times m}$ with $m > n$. The solutions to [L1] are referred to as ℓ_1 solutions to an overdetermined linear system. We alternatively refer to this problem as the "linear ℓ_1 minimization" or, simply the "linear ℓ_1" problem. Let

$$r(x) = A^{\mathrm{T}}x - b, \tag{2}$$

and define a *sign vector* s with components s_i such that

$$s_i(x) = \begin{cases} -1 & \text{if } r_i(x) < 0, \\ 0 & \text{if } r_i(x) = 0, \\ 1 & \text{if } r_i(x) > 0. \end{cases} \tag{3}$$

In general a sign vector is any vector $s \in \mathbb{R}^m$ with components $s_i \in \{-1, 1, 0\}$.

* Corresponding author. e-mail: nimp@unidhp.uni-c.dk.

[1] This work was done while this author was visiting the Institute of Mathematical Modelling of the Technical University of Denmark under a two-year post-doctoral fellowship provided through Danish Natural Sciences Research Council (SNF) grant no. 11-0505.

SSDI 0167-6377(94)00034-4

In the present paper we analyze the solution set of [L1] by examining the structure of sign vectors associated with the solutions. The main results of the paper can now be summarized as follows. In the first part of the paper in Section 2 we give a new polyhedral description of the solution set of [L1] using a special sign vector we refer to as the *minimal sign vector* of the solution set of [L1]. This result is given in Theorem 3. In the second part of the paper in Section 3 we characterize the solution set of [L1] in terms of the solution set of an approximating smooth problem, "the Huber problem" [2]. We also establish conditions under which the approximating problem yields a sign vector that coincides with the minimal sign vector of the solution of set of [L1]. These results are given in Theorem 6 and Theorem 7, respectively. To the best of our knowledge all the main results of the present paper are novel.

2. The structure of the solution set of [L1]

In this section, we describe some properties of the solution set of [L1] that are essential for our subsequent analysis. We assume without loss of generality throughout the paper that A has rank n, and that every column a_i of A is non-zero. Otherwise, the problem could easily be reformulated to have these properties.

We begin with the well-known characterization of an ℓ_1 solution to an overdetermined linear system. For any sign vector s we define

$$W_2 = \operatorname{diag}(w_1, \dots, w_m), \tag{4}$$

where

$$w_i = 1 - s_i^2. \tag{5}$$

Theorem 1. *A vector $x \in \mathbb{R}^n$ solves* [L1] *if and only if there exists $d \in \mathbb{R}^m$ such that*

$$AW_0 d + As_0 = 0, \tag{6}$$

$$\| W_0 d \|_\infty \leqslant 1, \tag{7}$$

where $s_0 = s(x)$, $W_0 = W_{s_0}$.

Proof. See [7, Theorem 6.1, pp. 118–119]. □

Clearly, the statement of the theorem is equivalent to the duality correspondence between [L1] and the following linear program

$$\text{[NormLP]} \quad \max_{y} \; b^{\mathrm{T}} y \quad \text{subject to} \quad Ay = 0, \quad -e \leqslant y \leqslant e$$

where $y \in \mathbb{R}^m$ and $e = (1, \dots, 1)$.

Let $\mathcal{S}$ denote the set of solutions to [L1], and let Ω be the set of all $x^i \in \mathcal{S}$ such that rank $\{a_j^{\mathrm{T}} \mid r_j(x^i) = 0\} = n$. Ω is non-empty by Theorem 6.2 of [7]. Now, we have the following description of the solution set of [L1].

Theorem 2. *Let $\mathcal{K}$ denote the convex hull of all $x^i \in \mathcal{S}$ such that* rank$\{a_j^{\mathrm{T}} \mid r_j(x^i) = 0\} = n$. *Then $\mathcal{S} \equiv \mathcal{K}$.*

Proof. See [7, Theorem 6.3, p. 120]. □

In the remainder of this section we characterize the solution set $\mathcal{S}$ entirely in terms of a special sign vector.

2.1. Sign structure of the solution set of [L1]

The following simple result allows a sign characterization of the solution set of [L1]. The result is also mentioned in [7, Example 5, p. 121]. We state it in a slightly different form without proof since the proof is a simple exercise using the convexity of the function $|r_j|$. Let $\mathcal{A}_0(x) \doteq \{i \mid r_i(x) = 0\}$ where $x \in \mathcal{S}$.

Lemma 1. *Let $x^1, x^2 \in \mathcal{S}$. Then, $s_j(x^1)s_j(x^2) \geqslant 0$ for any $j \in \{1, \dots, m\}$.*

Theorem 2 and Lemma 1 have the following consequences:

Corollary 1. *Let $x^1, x^2 \in \mathcal{S}$, and let $x = \alpha x^1 + (1-\alpha)x^2$, where $0 < \alpha < 1$. Then $s(x) = s(x^1) \oplus s(x^2)$ where*

$$s_j \oplus s_j' = \begin{cases} s_j + s_j' & \text{if } s_j s_j' = 0, \\ s & \text{if } s_j = s_j' = s. \end{cases} \tag{8}$$

Corollary 2. *For any* $x \in \mathscr{S}$, *there exists* $\Omega' \subseteq \Omega$ *such that* $s(x) = \bigoplus_{x^i \in \Omega'} s(x^i)$.

Proof. Since Ω is non-empty the result follows from the previous development. □

For any sign vector $s = (s_1, \ldots, s_m)$, we define $\sigma_s = \{i \mid s_i = 0\}$, $\bar{\sigma}_s^+ = \{i \mid s_i = 1\}$, and $\bar{\sigma}_s^- = \{i \mid s_i = -1\}$. Clearly, $\sigma_s \cup \bar{\sigma}_s = \{1, \ldots, m\}$, where $\bar{\sigma}_s = \bar{\sigma}_s^+ \cup \bar{\sigma}_s^-$. Let also

$$\mathscr{C}_s^0 = \mathrm{cl}\,\{x \in \mathbb{R}^n \mid s(x) = s\} \tag{9}$$

and

$$\mathscr{D}_s^0 = \mathrm{cl}\,\{x \in \mathbb{R}^n \mid s_j(x) = s_j, j \in \bar{\sigma}_s\}. \tag{10}$$

Note that $\mathscr{D}_s^0 = \mathbb{R}^n$ if $\bar{\sigma}_s$ is empty. Now, let $\bar{s} = \bigoplus_{x^i \in \Omega} s(x^i)$. We note that if $\mathscr{S}$ is a singleton $\bar{s} = s(x)$ where $\mathscr{S} = \{x\}$. We refer to $\bar{s}$ as the "minimal" sign vector of $\mathscr{S}$ since for any $x \in \mathscr{S}$ such that $s(x) = \bar{s}$, $|\mathscr{A}_0(x)| \leqslant |\mathscr{A}_0(x')|$ for any $x' \in \mathscr{S}$.

Corollary 3. $\mathrm{rank}\{a_i^{\mathrm{T}} \mid i \in \sigma_{\bar{s}}\} = n$ *iff* $\mathscr{S}$ *is a singleton.*

Proof. Necessity follows using the same argument as in the proof of the previous corollary. For the converse, let $x^1, x^2 \in \mathscr{S}$, where $x^1 \neq x^2$. Since $W_{\bar{s}} A^{\mathrm{T}}(x^1 - x^2) = 0$ this implies $\{a_i^{\mathrm{T}} \mid i \in \sigma_{\bar{s}}\}$ do not span $\mathbb{R}^n$. □

Corollary 4. *Let* $x \in \mathscr{S}$. *If* $s(y) = s(x)$ *then* $y \in \mathscr{S}$.

Proof. Follows from Theorem 1. □

Corollary 5. *There exists* $\bar{x} \in \mathscr{S}$ *with* $s(\bar{x}) = \bar{s}$.

Proof. The result is obvious if $\mathscr{S}$ is a singleton. Otherwise, for all $j \in \{1, \ldots, m\}$ there exists $x^j \in \Omega$ such that $\bar{s}_j = s_j(x)$. Define $\bar{x} = \sum_{i=1}^{p} x^i/p$ where p is the number of such distinct points. By construction $s(\bar{x}) = \bar{s}$. Now, by Theorem 2 $\bar{x} \in \mathscr{S}$. □

Now, we can give the following alternative polyhedral characterization of $\mathscr{S}$.

Theorem 3. $\mathscr{S} \equiv \mathscr{C}_{\bar{s}}^0$.

Proof. The result is evident if $\mathscr{S}$ is a singleton. Otherwise, by the previous corollary there exists $\bar{x} \in \mathscr{S}$ with $s(\bar{x}) = \bar{s}$. Now, by Corollary 4 $\{x \in \mathbb{R}^n \mid s(x) = \bar{s}\} \subseteq \mathscr{S}$. Now, by continuity, $\mathscr{C}_{\bar{s}}^0 \subseteq \mathscr{S}$ since $\mathscr{S}$ is closed.

Now, let $x \in \mathscr{S}$. Let $s_0 = s(x)$. If $s_0 = \bar{s}$, there is nothing to prove. Otherwise, using the definition of $\bar{s}$ and Lemma 1, $\sigma_{\bar{s}} \subset \sigma_{s_0}$ and $\bar{s}_i r_i(x) \geqslant 0$ for all $i \in \bar{\sigma}_s$. This implies that $x \in \mathscr{C}_{\bar{s}}^0$. □

Corollary 6. $\mathscr{S} \subseteq \mathscr{D}_{\bar{s}}^0$.

Proof. Follows from $\mathscr{C}_{\bar{s}}^0 \subseteq \mathscr{D}_{\bar{s}}^0$. □

Example 1. Consider the following problem

$$\text{minimize } G(x) \equiv |x| + |x - 3|$$

where $A = (1, 1)$ and $b = (0, 3)^{\mathrm{T}}$. The solution set is the interval $[0, 3]$ with $s(0) = (0, -1)$ and $s(3) = (1, 0)$, $\bar{s} = s(0) \oplus s(3) = (1, -1)$. In this case, $\mathscr{C}_{\bar{s}}^0 = \mathscr{D}_{\bar{s}}^0 = \mathscr{S} = [0, 3]$.

3. An approximation of [L1]

In [4] the first two authors showed that a minimizer of G can be estimated by solving a sequence of approximating smooth problems, each of which depends on a parameter $\gamma > 0$. These problems are defined as follows. Define for a given threshold $\gamma > 0$ the sign vector

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

with

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

If $s = s^\gamma(x)$ then we also denote W_s by $W_\gamma(x)$, or W_γ if no confusion is possible.

Now, the non-differentiable problem [L1] is approximated by the smooth "Huber problem", [2],

$$\text{[SL1]} \qquad \text{minimize } G_\gamma(x) \equiv \frac{1}{2\gamma} r^{\mathrm{T}} W_\gamma r + s^{\gamma\mathrm{T}}\left[r - \frac{1}{2}\gamma s^\gamma\right] \tag{13}$$

where the argument x is dropped for notational convenience. Clearly, G_γ measures the "small" residuals ($|r_i(x)| < \gamma$) by their squares while the "large" residuals are measured by the ℓ_1 function. Thus, G_γ is a piecewise quadratic function, and it is continuously differentiable in $\mathbb{R}^n$. In [4] the first two authors showed that when $\gamma \to 0_+$ then any solution to [SL1] is close to a solution to [L1]. Furthermore, in a more recent work [5], it was shown that dual solutions to [L1] and [NormLP] can be detected directly when γ is below a certain (problem dependent) threshold $\gamma_0 > 0$. In the same reference, a finite algorithm based on the above ideas is developed to solve linear programming problems of the form [NormLP] where the right-hand side is not necessarily zero.

3.1. The structure of the solution set of [SL1]

The structure of the function G_γ and its minimizers have been previously studied in [1, 3–5]. Therefore, we are not concerned with a detailed study of the properties of [SL1]. Instead, we describe some properties of this problem, which are essential to our subsequent development. In particular, we characterize the solution set of [SL1], and we give a new characterization of the solution set of [L1] in terms of the solution set of [SL1].

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

$$B_\gamma = \{x \in \mathbb{R}^n \mid \exists i\colon |r_i(x)| = \gamma\}. \tag{14}$$

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

$$\forall \varepsilon > 0 \, \exists z \in \mathbb{R}^n \backslash B_\gamma\colon$$
$$\|x - z\| < \varepsilon \wedge s = s^\gamma(z). \tag{15}$$

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

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

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

Q_s^γ can be defined as follows:

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

The gradient of the function G_γ is given by

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

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

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

The set of indices corresponding to "small" residuals

$$A_\gamma(z) = \{i \mid 1 \leqslant i \leqslant m \wedge |r_i(z)| \leqslant \gamma\} \tag{20}$$

is called the *γ-active* set at z and the subspace

$$\mathscr{V}_\gamma(z) = \mathrm{span}\{a_i \mid i \in \mathscr{A}_\gamma(z)\} \tag{21}$$

is called the *γ-active* subspace at z. The set of minimizers of G_γ is denoted by M_γ. In [1] it is shown that there exists a minimizer $x_\gamma \in M_\gamma$ for which $\mathscr{V}_\gamma(x_\gamma) = \mathbb{R}^n$.

The following three results were proved in [5] for the more general problem

$$\text{minimize } F(x) \equiv \|A^{\mathrm{T}}x - b\|_1 + c^{\mathrm{T}}x \tag{22}$$

where c is a vector of appropriate dimension. Naturally, they also apply to [L1]. In the interest of clarity we reproduce the proofs here.

Lemma 2. *$s^\gamma(x_\gamma)$ is constant for $x_\gamma \in M_\gamma$. Furthermore $r_i(x_\gamma)$ is constant for $x_\gamma \in M_\gamma$ if $s_i^\gamma = 0$.*

Proof. Let $x_\gamma \in M_\gamma$ and let $s = s^\gamma(x_\gamma)$, i.e., $G_\gamma(x) = Q_s^\gamma(x)$ for $x \in \mathscr{C}_s^\gamma$. If $x \in \mathscr{C}_s^\gamma \cap M_\gamma$ then $Q_s^{\gamma''}(x)(x - x_\gamma) = 0$. Therefore, if $|r_i(x_\gamma)| < \gamma$ then $a_i^{\mathrm{T}}(x - x_\gamma) = 0$ (see (17)), and hence $r_i(x) = r_i(x_\gamma)$. Thus r_i is constant in $\mathscr{C}_s^\gamma \cap M_\gamma$. Using the fact that M_γ is connected and r_i is continuous, it is easily seen by repeating the argument above that r_i is constant in M_γ. Next suppose $r_i(x_\gamma) \geqslant \gamma$. Then $r_i(x) \geqslant \gamma$ for all $x \in M_\gamma$ because existence of $x \in M_\gamma$ with $r_i(x) < \gamma$ is excluded by the convexity of M_γ, the continuity of r_i, and the first part of the lemma. Similarly, $r_i(x_\gamma) \leqslant -\gamma \Rightarrow r_i(x) \leqslant -\gamma$ for $x \in M_\gamma$. This completes the proof. □

Following the lemma we use the notation $s^\gamma(M_\gamma) = s^\gamma(x_\gamma)$, $x_\gamma \in M_\gamma$ as the sign vector corresponding to the solution set. Lemma 2 has the following consequences which characterize the solution set M_γ.

Corollary 7. *M_γ is a convex set which is contained in one Q-subset: $\mathscr{C}_s^\gamma$ where $s = s^\gamma(M_\gamma)$.*

Proof. Follows immediately from the linearity of the problem and Lemma 2. □

Corollary 8. *Let $x_\gamma \in M_\gamma$, and $s = s^\gamma(M_\gamma)$. Let $\mathscr{N}_s$ be the orthogonal complement of $\mathscr{V}_s = \operatorname{span}\{a_i^{\mathrm{T}} \mid s_i = 0\}$. Then*

$$M_\gamma = (x_\gamma + \mathscr{N}_s) \cap \mathscr{C}_s^\gamma.$$

Proof. It follows from (18) that $G_\gamma'(x_\gamma + u) = 0$ if $u \in \mathscr{N}_s$ and $x_\gamma + u \in \mathscr{C}_s^\gamma$. Thus

$$M_\gamma \supseteq (x_\gamma + \mathscr{N}_s) \cap \mathscr{C}_s^\gamma.$$

If $x \in M_\gamma$ then $r_i(x) = r_i(x_\gamma)$ for $s_i = 0$, and hence $x - x_\gamma \in \mathscr{N}_s$. Therefore, Corollary 7 implies

$$M_\gamma \subseteq (x_\gamma + \mathscr{N}_s) \cap \mathscr{C}_s^\gamma$$

which proves the result. □

An important consequence of the previous characterization of M_γ is that it provides a sufficient condition for the uniqueness of x_γ. This result given below in Corollary 9 is related to Lemma 6 in the paper by Clark [1]. The difference between the two approaches stems from the fact that Clark uses the following sign vector s_γ with components

$$s_{\gamma i}(x) = \begin{cases} -1 & \text{if } r_i(x) < -\gamma, \\ 0 & \text{if } |r_i(x)| \leqslant \gamma, \\ 1 & \text{if } r_i(x) > \gamma. \end{cases} \tag{23}$$

Corollary 9. *Let $s = s^\gamma(M_\gamma)$. $x_\gamma \in M_\gamma$ is unique if* $\operatorname{rank}\{a_i^{\mathrm{T}} \mid s_i = 0\} = n$.

Example 2. Note that the condition in the previous lemma is not necessary for uniqueness of x_γ. To see this consider the problem of Example 1 with $\gamma = 1.5$. The unique minimizer occurs at $x_\gamma = 1.5$ where $s^\gamma = (1, -1)$.

3.2. "Huber" characterization of the solution set of [L1]

In this section we show how the solution set M_γ approximates the solution set $\mathscr{S}$ of the linear ℓ_1 problem as γ approaches 0.

Assume $x_\gamma \in M_\gamma$, and let $s = s^\gamma(M_\gamma)$. Let $\mathscr{V}_s$ and $\mathscr{N}_s$ be defined as in Corollary 8.

Since x_γ satisfies the necessary condition for a minimizer,

$$0 = AW_s(A^{\mathrm{T}}x_\gamma - b) + \gamma As \tag{24}$$

the following linear system is consistent,

$$(AW_sA^{\mathrm{T}})d = As. \tag{25}$$

Now let d solve (25) and assume $s^{\gamma-\varepsilon}(x_\gamma + \varepsilon d) = s$, i.e., $x_\gamma + \varepsilon d \in \mathscr{C}_s^{\gamma-\varepsilon}$ for some $\varepsilon > 0$. The linearity of the problem implies $x_\gamma + \delta d \in \mathscr{C}_s^{\gamma-\delta}$ for $0 \leqslant \delta \leqslant \varepsilon$. Therefore (24) and (25) show that $(x_\gamma + \delta d)$ is a minimizer of $G_{\gamma-\delta}$. Using Corollary 8 we have the following lemma.

Lemma 3. *Let $x_\gamma \in M_\gamma$ and let $s = s^\gamma(M_\gamma)$. Let d solve (25). If $s^{\gamma-\varepsilon}(x_\gamma + \varepsilon d) = s$ for $\varepsilon > 0$ then $s^{\gamma-\delta}(x_\gamma + \delta d) = s$, and*

$$M_{\gamma-\delta} = (x_\gamma + \delta d + \mathscr{N}_s) \cap \mathscr{C}_s^{\gamma-\delta} \tag{26}$$

for $0 \leqslant \delta \leqslant \varepsilon$.

Theorem 4. *There exists $\gamma_0 > 0$ such that $s^\gamma(M_\gamma)$ is constant for $0 < \gamma \leqslant \gamma_0$. Furthermore,*

$$M_{\gamma-\delta} = (x_\gamma + \delta d + \mathscr{N}_s) \cap \mathscr{C}_s^{\gamma-\delta} \quad \text{for } 0 \leqslant \delta < \gamma \leqslant \gamma_0$$

where $s = s^\gamma(M_\gamma)$ and d solves (25).

Proof. Since there is only a finite number of different sign vectors the theorem is a consequence of the previous lemma. □

Let $\mathcal{N}(C)$ denote the null space of an arbitrary matrix C.

Corollary 10. *Let $0 < \gamma \leqslant \gamma_0$, where γ_0 is given in Theorem 4 and let $s = s^\gamma(M_\gamma)$. Then*

$$W_s r(x_\gamma + \gamma \tilde{d}) = 0 \tag{27}$$

where $\tilde{d}$ is any solution to (25).

Proof. Let $x_{\gamma-\delta} \in M_{\gamma-\delta}$ for $0 \leqslant \delta < \gamma$. By Theorem 4 there exists d that solves (25) such that $x_{\gamma-\delta} = x_\gamma + \delta d$. Therefore, using the definition of r we have

$$\| W_s[A^\mathrm{T}(x_\gamma + \delta d) - b] \|_\infty < \gamma - \delta. \tag{28}$$

Any solution $\tilde{d}$ to (25) can be expressed as $\tilde{d} = d + \eta$ where $\eta \in \mathcal{N}(AW_sA^\mathrm{T})$. Now, $\mathcal{N}(AW_sA^\mathrm{T}) \equiv \mathcal{N}(W_sA^\mathrm{T})$ since $W_sW_s = W_s$. Hence, we have

$$\| W_s[A^\mathrm{T}(x_\gamma + \delta d) - b] \|_\infty < \gamma - \delta, \tag{29}$$

or equivalently,

$$\| W_s r(x_\gamma + \delta \tilde{d}) \|_\infty < \gamma - \delta. \quad \Box \tag{30}$$

We notice that if $x_\gamma \in M_\gamma$ then $y_\gamma = -(W_s r(x_\gamma)/\gamma + s)$, where $s = s^\gamma(M_\gamma)$, is feasible in [NormLP] as it is seen from (24). Now we recall a classical result from linear programming known as the complementary slackness theorem. This result is simply a restatement of Theorem 1, which is more convenient for our purposes; see for instance [6].

Theorem 5. *Let $x \in \mathbb{R}^n$ and $y \in \mathbb{R}^m$. Then x and y are optimal solutions in their respective problems if and only if y is feasible in* [NormLP] *and the following conditions hold:*

$$-1 < y_i < 1 \ \Rightarrow \ r_i(x) = 0, \tag{31}$$

$$r_i(x) > 0 \ \Rightarrow \ y_i = -1, \tag{32}$$

$$r_i(x) < 0 \ \Rightarrow \ y_i = +1. \tag{33}$$

Next, we state and prove the first main result of this section.

Theorem 6. *Let $0 < \gamma \leqslant \gamma_0$, where γ_0 is given in Theorem 4 and let $s = s^\gamma(M_\gamma)$. Let $x_\gamma \in M_\gamma$, and d solve* (25)*. Then*

$$M_0 \equiv \mathcal{S}$$

where

$$M_0 = (x_\gamma + \gamma d + \mathcal{N}_s) \cap \mathcal{D}_s^0, \tag{34}$$

and

$$y^* = -\left(\frac{1}{\gamma} W_s r(x_\gamma) + s\right) \tag{35}$$

solves [NormLP].

Proof. First, M_0 is non-empty as a consequence of the constant sign property of Theorem 4. Assume $x_0 \in M_0$. Then there exists a solution d_0 to (25) such that $x_0 = x_\gamma + \gamma d_0$. Therefore using Corollary 10 $\sigma_s \subseteq \mathcal{A}_0(x_0)$. Now the linearity and Theorem 4 imply that $x_{\gamma-\delta} = x_\gamma + \delta d_0 \in M_{\gamma-\delta}$ for $0 \leqslant \delta \leqslant \gamma$. Since $s^\gamma(x_\gamma) = s^{\gamma-\delta}(x_{\gamma-\delta})$ for $0 \leqslant \delta < \gamma$ the continuity of r gives

$$r_i(x_0) \neq 0 \ \Rightarrow \ \operatorname{sign}(r_i(x_0)) = \operatorname{sign}(r_i(x_{\gamma-\delta})) = s_i = -y_i^*, \tag{36}$$

for δ close to γ. Furthermore, y^* is feasible for [NormLP]. Therefore

$$G(x_0) = -r(x_0)^\mathrm{T} y^* = -x_0^\mathrm{T} A y^* + b^\mathrm{T} y^* = b^\mathrm{T} y^*.$$

Hence, x_0 and y^* are solutions to [L1] and [NormLP], respectively. Since this holds for any $x_0 \in M_0$, $M_0 \subseteq \mathcal{S}$ and y^* solves [NormLP].

If $\mathcal{S}$ is a singleton, the proof is complete. Therefore, assume $\mathcal{S}$ is not a singleton. What remains to be shown is that $x \in M_0$ for any $x \in \mathcal{S}$. Since x_0 and y^* are primal-dual solutions it follows from condition (31) that $\sigma_s \subseteq \mathcal{A}_0(x)$ for any $x \in \mathcal{S}$. Now, let $x \in \mathcal{S}$ and $x_\gamma \in M_\gamma$. Since $\sigma_s \subseteq \mathcal{A}_0(x)$, we have the following:

$$W_s(A^\mathrm{T}x - b) = 0. \tag{37}$$

Then using (24) and (37) we have

$$AW_sA^{\mathrm{T}}\frac{(x-x_\gamma)}{\gamma}=\frac{1}{\gamma}AW_sA^{\mathrm{T}}x-\frac{1}{\gamma}AW_sA^{\mathrm{T}}x_\gamma$$
$$=\frac{1}{\gamma}AW_sb-\frac{1}{\gamma}(AW_sb-\gamma As)$$
$$=As,$$

which shows that $(x - x_\gamma)/\gamma$ solves (25). Therefore we have shown that $x \in x_\gamma + \gamma d + \mathcal{N}_s$. Using conditions (32) and (33), the following sign accordance holds:

$$s_i \neq 0 \;\Rightarrow\; s_i r_i(x) \geqslant 0.$$

Therefore, $x \in \mathcal{D}_s^0$ for any $x \in \mathcal{S}$. Hence, $x \in M_0$. This completes the proof. □

Following Theorem 6, all the ℓ_1 solutions to an overdetermined linear system and all the "Huber" solutions are linked by a solution d to (25) for sufficiently small positive values of the parameter γ. The following is now an immediate corollary of the Theorem 6.

Corollary 11. *$M_\gamma = (x_0 - \gamma d - \mathcal{N}_s) \cap \mathcal{C}_s^\gamma$ for $\gamma \in (0, \gamma_0]$ where $x_0 \in \mathcal{S}$ and d solves* (25).

Another immediate consequence of the characterization theorem is the following corollary.

Corollary 12. *$\mathcal{S}$ is a singleton if* rank $\{a_i^{\mathrm{T}} \mid i \in \sigma_s\} = n$ *where $s = s^\gamma(M_\gamma)$ for $\gamma \in (0, \gamma_0]$.*

Proof. Since rank $\{a_i^{\mathrm{T}} \mid i \in \sigma_s\} = n$ $x_\gamma \in M_\gamma$ is unique by Corollary 9. This also implies that $\mathcal{N}_s = \{0\}$. Hence $(AW_sA^{\mathrm{T}})d = As$ has a unique solution, d_0 say. Therefore, $x_\gamma + \gamma d_0 + \mathcal{N}_s$ is a singleton. Hence, by Theorem 6, $\mathcal{S}$ is a singleton. □

Our final results concern the following question of sign identity: "If and when s as defined in Theorem 4 coincides with the minimal sign vector $\bar{s}$ of $\mathcal{S}$?" The following sample problem from [1] illustrates the sign identity.

Example 3. Consider the problem

$$\min G(x) = |3x_1 + 2x_2| + |4x_1 - 4| + |3x_2 - 3| + |2x_1 + 3x_2 - 5| + |8x_1 + 7x_2 - 20|.$$

$\mathcal{S} = \Omega = \{x^1\} = (1, 1)^{\mathrm{T}}$ and $s(x^1) = (1, 0, 0, 0, -1)^{\mathrm{T}}$ whereas for $0 < \gamma < 1.23$, $x_\gamma = (1 + 3\gamma/16, 1 + 2\gamma/9)^{\mathrm{T}}$, with $s^\gamma(x_\gamma) = (1, 0, 0, 1, -1)^{\mathrm{T}}$. If "8" is changed to "7.5" for $0 < \gamma < 1.34$, $s^\gamma(x_\gamma) = (1, 0, 0, 0, -1)^{\mathrm{T}}$ thereby giving sign identity.

Recall that when $\mathcal{S}$ has a unique sign vector, s^* say, $\bar{s}$ reduces to s^* by definition. The following result which is a by-product of the proof of Theorem 6 gives a partial answer to the question of sign identity. The sign identity property is also mentioned in [1]. In this connection Corollary 13 below offers an alternative statement to Theorem 6 of [1] by using the concept of a minimal sign vector.

Corollary 13. *Let $0 < \gamma \leqslant \gamma_0$, where γ_0 is given in Theorem 4 and let $s = s^\gamma(M_\gamma)$. Then $\sigma_s \subseteq \sigma_{\bar{s}}$, $\bar{\sigma}_{\bar{s}}^+ \subseteq \bar{\sigma}_s^+$, and $\bar{\sigma}_{\bar{s}}^- \subseteq \bar{\sigma}_s^-$ where $\bar{s}$ is the minimal sign vector of $\mathcal{S}$.*

In [1] no conditions are specified under which the sign identity is expected to hold. In our final theorem we give alternative characterizations the sign identity property. Let Y^* denote the set of optimal solutions to [NormLP].

Theorem 7. *Let $0 < \gamma \leqslant \gamma_0$, where γ_0 is given in Theorem 4 and let $s = s^\gamma(M_\gamma)$. Let $\bar{s}$ be the minimal sign vector of $\mathcal{S}$. Then the following statements are equivalent:*

(1) $s = \bar{s}$
(2) *For all $i \in \bar{\sigma}_s$, $y_i = -s_i$ for all $y \in Y^*$*
(3) *For all $i \in \bar{\sigma}_s$, there exists $x \in \mathcal{S}$ such that $s_i(x) = s_i$*
(4) *There exists $d \in \mathbb{R}^n$ that solves*

$$(AW_{\bar{s}}A^{\mathrm{T}})d = A\bar{s} \tag{38}$$

such that $\| W_{\bar{s}}A^{\mathrm{T}}d \|_\infty < 1$.

Proof. The equivalence of (2) and (3) follows from the complementarity theorem of Goldman and Tucker (see e.g., [8]). Now, clearly (1) and (3) are equivalent using the previous corollary.

(1) ⇒ (4): This follows immediately from Corollary 11 where x_0 satisfies $s(x_0) = \bar{s} = s$.

(4) ⇒ (1): The system (38) is consistent following Theorem 1 and Corollary 5. Now, let $\bar{x}$ be

a solution to [L1] such that $s(\bar{x}) = \bar{s}$. Let $\delta = \min\{|r_i(\bar{x})|: i \in \bar{\sigma}_{\bar{s}}\}$. Choose $0 < \gamma_0 \leqslant \delta$ so that for all $0 < \gamma \leqslant \gamma_0$,

$$r_i(\bar{x}) - \gamma(A^{\mathrm{T}}d)_i \geqslant \gamma_0, \quad i \in \bar{\sigma}_{\bar{s}}^{+}, \tag{39}$$

$$r_i(\bar{x}) - \gamma(A^{\mathrm{T}}d)_i \leqslant -\gamma_0, \quad i \in \bar{\sigma}_{\bar{s}}^{-}. \tag{40}$$

Now using (38) and the fact that $W_{\bar{s}}(A^{\mathrm{T}}\bar{x} - b) = 0$ we have

$$0 = AW_{\bar{s}}A^{\mathrm{T}}(-\gamma d) + \gamma A\bar{s}$$
$$= AW_{\bar{s}}(A^{\mathrm{T}}(\bar{x} - \gamma d) - \mathrm{b}) + \gamma A\bar{s}.$$

Since $\| W_{\bar{s}}A^{\mathrm{T}}d \|_{\infty} < 1$, using (39) and (40) we have $s^{\gamma}(\bar{x} - \gamma d) = \bar{s}$. Hence, $\bar{x} - \gamma d \in M_{\gamma}$. By Theorem 4, $s = \bar{s}$. This proves the theorem. □

The following corollary gives a necessary condition for the uniqueness of solution in [NormLP].

Corollary 14. *If Y^* is a singleton $s = \bar{s}$.*

In example 3 above it can be verified that Clause (2) of Theorem 7 fails to hold since the associated linear program [NormLP] has two extreme solutions $y^1 = (-1, -3/4, -2/3, -1, 1)^{\mathrm{T}}$, and $y^2 = (-1, -11/12, -1, -2/3, 1)$.

Acknowledgements

We would like to thank an anonymous referee for several suggestions and corrections that improved the paper, and in particular for pointing out to us Clause (4) of Theorem 7.

References

[1] D. Clark, "The mathematical structure of Huber's M-estimator", *SIAM J. Sci. Statist. Comput.* **6**, 209–219 (1985).
[2] P. Huber, *Robust Statistics*, Wiley, New York, 1981.
[3] K. Madsen and H.B. Nielsen, "Finite algorithms for robust linear regression", *BIT* **30**, 682–699 (1990).
[4] K. Madsen and H.B. Nielsen, "A finite smoothing algorithm for linear ℓ_1 estimation", *SIAM J. Optimization* **3**, 223–235 (1993).
[5] K. Madsen, H.B. Nielsen and M.Ç. Pınar, "A new finite continuation algorithm for linear programming", Technical Report, Institute for Numerical Analysis, Technical University of Denmark, Lyngby 2800, Denmark, 1993.
[6] K.G. Murty, *Linear and Combinatorial Programming*, Wiley, New York, 1976.
[7] G.A. Watson, *Approximation Theory and Numerical Methods*, Wiley, New York, 1980.
[8] A.C. Williams, "Complementarity theorems for linear programming", *SIAM Review* **12**, 135–137 (1970).