

NECESSARY AND SUFFICIENT CONDITIONS FOR NOISELESS SPARSE RECOVERY VIA CONVEX QUADRATIC SPLINES*

MUSTAFA Ç. PINAR†

*This paper is dedicated to the memory of Hans Bruun Nielsen (1943–2015) and to
Kaj Madsen on the occasion of his 75th birthday*

Abstract. The problem of exact recovery of an individual sparse vector using the Basis Pursuit (BP) model is considered. A differentiable Huber loss function (a convex quadratic spline) is used to replace the ℓ_1 -norm in the BP model. Using the theory of duality and classical results from quadratic perturbation of linear programs, a necessary condition for exact recovery leading to a negative result is given. An easily verifiable sufficient condition is also presented.

Key words. exact recovery of a sparse vector, basis pursuit, Huber loss function, strictly convex quadratic programming, linear programming, convex quadratic splines, ℓ_1 -norm, quadratic perturbation

AMS subject classifications. 65K05, 90C05, 90C20, 41A15, 94A12

DOI. 10.1137/18M1185375

1. Introduction. The purpose of this paper is to give necessary and sufficient conditions (of different type) for exact recovery of a sparse vector by means of ℓ_1 -norm minimization, an approach which is part of the literature on compressive sensing [2, 7]. To the best of the author's knowledge, the extensive literature on compressive sensing has not considered the approach of the present paper, where an instance of a convex quadratic spline function, namely, Huber loss function (see, e.g., [6] for a reference on convex quadratic splines), is studied with a view to recovering an individual sparse vector by solving strictly convex quadratic programming problems. Studies on connections between ℓ_1 estimation and Huber loss function [13] minimization date from at least two decades ago (e.g., [15, 16, 17]). The present paper is circulated in the hope that these connections can be reused for sparse recovery. A recent connection between sparse recovery and Huber loss was established in [21], which uses the Huber function (and its generalization) as the basis of a minimax-concave penalty in the context of regularized least squares for sparse recovery. Since the 1-norm regularization tends to underestimate large components of a sparse vector, other nonconvex sparsity-inducing regularization terms have been proposed. The reference [21] provides an extensive list of the literature on nonconvex regularizers. The emphasis and techniques of [21] are different from those of the present paper as they are derived from convex analysis and lead to saddle point computations involving proximal algorithms.

For the most part, references of a theoretical nature in the area of sparse recovery are concerned with the nullspace condition and the Restricted Isometry Property (RIP) (in particular with random matrices satisfying the conditions requisite for recovery); see, e.g., [2, 4, 5, 10, 14] for exact ℓ_1 recovery of all sparse signals with K nonzero entries when no measurement noise is present. Instead, the present paper will rather deal with the recovery of an individual sparse vector.

*Received by the editors May 3, 2018; accepted for publication (in revised form) by S. Le Borne December 10, 2018; published electronically February 12, 2019.

<http://www.siam.org/journals/simax/40-1/M118537.html>

†Department of Industrial Engineering, Bilkent University, 06800 Ankara, Turkey (mustafap@bilkent.edu.tr).

The problem of interest in the present paper comes from the Basis Pursuit (BP) approach to sparse recovery (there is a vast literature on the subject, which cannot be reviewed within this brief paper; see, e.g., [4, 8, 9, 11], or [10] for an in-depth monograph on the problem of compressive sensing) and is the following problem, referred to as [SL1]:

$$(1.1) \quad \min_x \{\|x\|_1 : Ax = Au\},$$

where $A \in \mathbb{R}^{m \times n}$ ($m < n$), $u \in \mathbb{R}^n$ is a sparse vector with $\text{nnz}(u) < m$ nonzero components (the notation $\text{nnz}(u)$ is used to denote the number of nonzero elements of the vector u), which is a reasonable assumption, since according to Theorem 3.1 of [10], if [SL1] has a unique solution, then the cardinality of the support of the unique solution is at most equal to m . Hence, in any case, the BP model results in at most m nonzeros. Therefore, it is plausible to look for recovery of a vector with cardinality of support less than m . For ease of reference later, define $y \equiv Au$. It is also assumed that A has full rank (equal to m). The notation sgn refers to the sign vector (or scalar) with components in $\{0, \pm 1\}$ such that

$$\text{sgn}(t) = \begin{cases} 0 & \text{if } t = 0, \\ 1 & \text{if } t > 0, \\ -1 & \text{if } t < 0. \end{cases}$$

The term “solution to a problem” in the present paper refers to an optimal solution. Let X^0 denote the set of optimal solutions to (1.1).

DEFINITION 1. *Exact individual recovery by BP is said to take place if the unique solution to (1.1) is u .*

A result close to the work of the present paper (at least in the second part since the nullspace-type conditions are not studied in the present paper) is the following (see Theorem 4.30 of [10]; it is credited to [11] and [22] and shown to be necessary in [3] for partial Fourier matrices).

THEOREM 1. *Given a matrix $A \in \mathbb{R}^{m \times n}$, a vector $x \in \mathbb{R}^n$ with support S is the unique minimizer of $\|z\|_1$ subject to $Az = Ax$ (i.e., exact individual recovery takes place) if and only if one of the following equivalent conditions holds:*

- (a) $\sum_{j \in S} |\text{sgn}(x_j)v_j| < \|v_S\|_1$ for all $v \in \text{Null}(A) \setminus \{0\}$.
- (b) A_S is injective, and there exists a vector $h \in \mathbb{R}^m$ such that

$$(A^T h)_j = \text{sgn}(x_j) \text{ for } j \in S, \quad |(A^T h)_\ell| < 1, \quad \ell \in \bar{S}.$$

Results similar to the one above are also studied in other papers; see, e.g., [12, 24].

In the rest of the paper, after some preliminaries in section 2, easy-to-verify sufficient conditions are presented in section 3, while section 4 discusses a necessary condition leading to a (negative) result for exact recovery of sparse vectors with the cardinality of support less than the number of observations (under some mild assumptions). Section 5 gives a numerical illustration. To the best of the author’s knowledge, the main results of the paper in Theorem 5 and Corollary 4 are novel contributions to the literature on sparse recovery.

2. Preliminaries. The Lagrange dual problem to [SL1] plays an important role in the necessary condition that will be presented later:

$$(2.1) \quad \max_{\lambda \in \mathbb{R}^m} \{y^T \lambda : \|A^T \lambda\|_\infty \leq 1\}.$$

Needless to say, both problems are equivalent to a linear program; in the case of [SL1], some additional variables are used. Furthermore, both are feasible and bounded, and hence strong duality holds. The following result is an easy consequence of optimality conditions in linear optimization (it can be derived by several means; see Lemma 2.1 of [15], which can be adapted verbatim to the purposes of the present paper).

LEMMA 1. *Let $\bar{x} \in \mathbb{R}^n$ and $\bar{\lambda} \in \mathbb{R}^m$. Then \bar{x} and $\bar{\lambda}$ are solutions of (1.1) and (2.1), respectively, if and only if $A\bar{x} = y$, $\|A^T\bar{\lambda}\|_\infty \leq 1$, and the following conditions hold:*

$$(2.2) \quad \begin{aligned} \bar{x}_i &\geq 0 & \text{if } (A^T\bar{\lambda})_i &= 1, \\ \bar{x}_i &\leq 0 & \text{if } (A^T\bar{\lambda})_i &= -1, \\ \bar{x}_i &= 0 & \text{if } -1 < (A^T\bar{\lambda})_i < 1. \end{aligned}$$

The conditions discussed here involve a slightly different problem based on a convex quadratic spline function, namely, the Huber loss function depending on a tuning constant $\gamma > 0$:

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

The Huber loss problem that will be treated is referred to as [HL1]:

$$(2.3) \quad \min_x \{\|x\|_H : Ax = y\},$$

where, by abuse of notation, $\|x\|_H$ is used to denote $\sum_{i=1}^n \rho(x_i)$, although the Huber loss function is not a norm (it does not satisfy the triangle inequality). However, the function ρ (and hence $\|x\|_H$) is a convex differentiable piecewise quadratic polynomial. The Lagrange dual problem [DHL1] is as follows:

$$(2.4) \quad \max_{\lambda \in \mathbb{R}^m} \left\{ y^T \lambda - \frac{\gamma}{2} \|A^T \lambda\|_2^2 : \|A^T \lambda\|_\infty \leq 1 \right\}.$$

By virtue of strict concavity of the objective function (A has full rank), the dual problem (2.4) has a unique solution. The following characterization of primal-dual solutions easily follows from Karush–Kuhn–Tucker optimality conditions.

LEMMA 2. *Let $\bar{x} \in \mathbb{R}^n$ and $\bar{\lambda} \in \mathbb{R}^m$. Then \bar{x} and $\bar{\lambda}$ are solutions of (2.3) and (2.4), respectively, if and only if $A\bar{x} = y$, $\|A^T\bar{\lambda}\|_\infty \leq 1$, and the following conditions hold:*

$$(2.5) \quad \begin{aligned} \bar{x}_i &\geq \gamma & \text{if } (A^T\bar{\lambda})_i &= 1, \\ \bar{x}_i &\leq -\gamma & \text{if } (A^T\bar{\lambda})_i &= -1, \\ \bar{x}_i &= \gamma(A^T\bar{\lambda})_i & \text{if } -1 < (A^T\bar{\lambda})_i < 1. \end{aligned}$$

The optimality conditions of the previous lemma can be reiterated in what follows using more compact notation; define $s^\gamma \in \{0, \pm 1\}^n$ as follows:

$$s_i^\gamma(t) = \begin{cases} 0 & \text{if } |t| < \gamma, \\ 1 & \text{if } t \geq \gamma, \\ -1 & \text{if } t \leq -\gamma, \end{cases}$$

with the diagonal matrix $W^\gamma(\cdot)$ derived from $s^\gamma(\cdot)$ using $W_{ii}^\gamma = 1 - (s_i^\gamma)^2$ for $i = 1, \dots, n$. Note that the gradient of $\|x\|_H$ at a point x can be expressed as $\frac{1}{\gamma} W^\gamma(x)x + s^\gamma(x)$. Then the following corollary can be stated.

COROLLARY 1. Let $\bar{x} \in \mathbb{R}^n$ and $\bar{\lambda} \in \mathbb{R}^m$. Then \bar{x} and $\bar{\lambda}$ are solutions of (2.3) and (2.4), respectively, if and only if $A\bar{x} = y$, $\|A^T\bar{\lambda}\|_\infty \leq 1$, and the following equation holds:

$$(2.6) \quad \frac{1}{\gamma} W^\gamma(\bar{x})\bar{x} + s^\gamma(\bar{x}) - A^T\bar{\lambda} = 0.$$

Let X^γ denote the optimal solution set to (2.3) (by extension, let X^0 denote the solution set of problem (1.1)). Then the following results are easy to prove (see the proofs of Lemma 2 and Corollary 9 of [17] or Lemmas 3 and 4 of [16], from which the following results are obtained after evident modifications). Nonetheless, the proof is included to make the paper self-contained.

LEMMA 3. $s^\gamma(\bar{x})$ is constant for $\bar{x} \in X^\gamma$. Furthermore, \bar{x}_i is constant for $\bar{x} \in X^\gamma$ if $s_i^\gamma(\bar{x}) = 0$.

Proof. Let $x^\gamma \in X^\gamma$ along with $s = s^\gamma(x^\gamma)$ and $W^\gamma = W^\gamma(x^\gamma)$ with dual solution λ^* to (2.4). By (2.6) one has

$$(2.7) \quad \frac{1}{\gamma} W^\gamma x^\gamma + s^\gamma - A^T \lambda^* = 0.$$

Define $\mathcal{C}_s^\gamma = \text{cl}\{z \in \mathbb{R}^n | s^\gamma = s\}$. Note that for any $x \in \mathcal{C}_s^\gamma$ the function $\|z\|_H$ can be locally represented by the quadratic function

$$Q_s^\gamma(z) = \frac{1}{2\gamma} (z - x)^T W^\gamma (z - x) + \left[\frac{1}{\gamma} W^\gamma x + s^\gamma \right]^T (z - x) + \|x\|_H.$$

If $z \in \mathcal{C}_s^\gamma \cap X^\gamma$, then one can rewrite the quadratic $Q_s^\gamma(z)$ using an expansion around x^γ as

$$Q_s^\gamma(z) = \frac{1}{2\gamma} (z - x^\gamma)^T W^\gamma (z - x^\gamma) + \left(\frac{1}{\gamma} W^\gamma x^\gamma + s^\gamma \right)^T (z - x^\gamma) + \|x^\gamma\|_H.$$

Thus, from the optimality conditions for the problem of minimizing Q_s^γ over all z satisfying $Az = y$, one has that $(Q_s^\gamma)'(z) - A^T \lambda^* = 0$, which is equivalent to

$$\frac{1}{\gamma} W^\gamma (z - x^\gamma) + \frac{1}{\gamma} W^\gamma x^\gamma + s^\gamma = A^T \lambda^*.$$

Since by (2.7) $\frac{1}{\gamma} W^\gamma x^\gamma + s^\gamma = A^T \lambda^*$ one gets $W^\gamma (z - x^\gamma) = 0$. Therefore, if $|x_i^\gamma| < \gamma$, then $z_i = x_i^\gamma$. Thus, x_i^γ is constant for $x^\gamma \in \mathcal{C}_s^\gamma \cap X^\gamma$ if $|x_i^\gamma| < \gamma$. Now, let U be a subset of \mathbb{R}^n such that $U \cap \mathcal{C}_s^\gamma \neq \emptyset$. If $U \cap X^\gamma \neq \emptyset$, then there should exist points x in $U \cap X^\gamma$ such that $|x_i| < \gamma$ since x_i is a continuous function of itself, and X^γ is a convex set. Hence x_i is constant in $U \cap X^\gamma$ because of the argument given above. Therefore, if $|x_i^\gamma| < \gamma$, then $z_i = x_i^\gamma$ for any $z \in U \cap X^\gamma$. Repeating this argument, due to the connectedness of X^γ , one obtains that x_i is constant for $x \in X^\gamma$ if $|x_i| < \gamma$. In other words, small components of any minimizer are constant.

Now, let $x \in X^\gamma$. Fix a $j \in \{1, \dots, n\}$. If $|x_j| < \gamma$, then by the above paragraph x_j is constant for any $x \in X^\gamma$. Assume $x_j \geq \gamma$. If there exists $z \in X^\gamma$ with $z_j < \gamma$, then there exists $w \in X^\gamma$ with $|w_j| < \gamma$ due to convexity of X^γ and continuity. However, this is a contradiction to the above paragraph. Hence $y_j \geq \gamma$ for all $y \in X^\gamma$. The proof for $y_j \leq -\gamma$ is similar. \square

The result is saying that the small components (less than γ) have constant value for all minimizers, while the large components remain large and do not change sign.

Following the above lemma, one can refer to $s^\gamma(X^\gamma)$ (and, by extension, $W^\gamma(X^\gamma)$) for the entire solution set X^γ .

Remark 1. One could equivalently define s^γ based on the entries of $A^T \bar{\lambda}$; i.e., one would have

$$s_i^\gamma((A^T \bar{\lambda})_i) = \begin{cases} 0 & \text{if } |(A^T \bar{\lambda})_i| < 1, \\ (A^T \bar{\lambda})_i & \text{otherwise.} \end{cases}$$

LEMMA 4. Let $\bar{x} \in X^\gamma$. X^γ is a singleton if the matrix $\begin{pmatrix} W^\gamma(\bar{x}) \\ A \end{pmatrix}$ has rank equal to n .

Proof. If the matrix $\begin{pmatrix} W^\gamma(\bar{x}) \\ A \end{pmatrix}$ has rank equal to n , then the system

$$\begin{pmatrix} W^\gamma(\bar{x}) \\ A \end{pmatrix} z = 0$$

has the trivial solution $z = 0$. Now let \bar{x}^γ be another point in X^γ . By the previous lemma $W^\gamma(\bar{x}^\gamma) = W^\gamma(\bar{x})$ with $s^\gamma(\bar{x}^\gamma) = s^\gamma(\bar{x})$, which are referred to as W and s for notational convenience. By (2.6) one has

$$\frac{1}{\gamma} W \bar{x}^\gamma + s = A^T \lambda^*$$

and

$$\frac{1}{\gamma} W \bar{x} + s = A^T \lambda^*,$$

where λ^* is the unique dual solution. The above two equations imply $W(\bar{x} - \bar{x}^\gamma) = 0$. From $A\bar{x} = y$ and $A\bar{x}^\gamma = y$, one obtains $A(\bar{x} - \bar{x}^\gamma) = 0$. Hence, one gets

$$\begin{pmatrix} W \\ A \end{pmatrix} (\bar{x} - \bar{x}^\gamma) = 0,$$

which implies that $\bar{x} = \bar{x}^\gamma$. □

Let λ^* denote the least AA^T -norm solution to (2.1) where it is meant that λ^* solves the problem

$$\min_{\lambda \in \Lambda} \frac{1}{2} \|\lambda\|_{AA^T}^2,$$

where Λ denotes the solution set of (2.1), and $\|\lambda\|_{AA^T}^2 \equiv \lambda^T AA^T \lambda$. Henceforth, the least AA^T -norm solution to (2.1) will be referred to as the *normal* solution to (2.1).

The following is a consequence of a classical result by Mangasarian and Meyer [18].

LEMMA 5. Let λ^γ be the unique solution of the strictly convex quadratic programming problem (2.4). Then there exists $\delta > 0$ such that λ^γ is the least AA^T -norm solution of (2.1) for $0 < \gamma \leq \delta$.

COROLLARY 2. Let λ be a solution to (2.4). Then \bar{x} is a solution to (2.3) if and only if \bar{x} satisfies $A\bar{x} = y$ and the following conditions:

$$(2.8) \quad \begin{aligned} \bar{x}_i &\geq \gamma && \text{if } (A^T \lambda)_i = 1, \\ \bar{x}_i &\leq -\gamma && \text{if } (A^T \lambda)_i = -1, \\ \bar{x}_i &= \gamma(A^T \lambda)_i && \text{if } -1 < (A^T \lambda)_i < 1. \end{aligned}$$

The following result is proved similarly to Lemma 2.9 of [15].

LEMMA 6. *If the solution λ^δ of (2.4) is the normal solution λ^* to (2.1) for some $\gamma = \delta > 0$, then the solution λ^γ of (2.4) is the normal solution of (2.1) for any $0 < \gamma \leq \delta$ (i.e., $\lambda^\gamma = \lambda^*$ for $0 < \gamma \leq \delta$).*

Proof. One has $x^\delta \in X^\delta$ if and only if $Ax^\delta = y$ and

$$(2.9) \quad \begin{aligned} x_i^\delta &\geq \delta && \text{if } (A^T \lambda^*)_i = 1, \\ x_i^\delta &\leq -\delta && \text{if } (A^T \lambda^*)_i = -1, \\ x_i^\delta &= \delta(A^T \lambda^*)_i && \text{if } -1 < (A^T \lambda^*)_i < 1. \end{aligned}$$

On the other hand, one has $x^0 \in X^0$ if and only if $Ax^0 = y$ and

$$(2.10) \quad \begin{aligned} x_i^0 &\geq 0 && \text{if } (A^T \lambda^*)_i = 1, \\ x_i^0 &\leq 0 && \text{if } (A^T \lambda^*)_i = -1, \\ x_i^0 &= 0 && \text{if } -1 < (A^T \lambda^*)_i < 1. \end{aligned}$$

Now, let $0 < \gamma \leq \delta$, $x^0 \in X^0$, $\bar{x} = (1 - \frac{\gamma}{\delta})x^0 + \frac{\gamma}{\delta}x^\delta$. Then it is easy to verify by direct computation that

$$(2.11) \quad \begin{aligned} \bar{x}_i &\geq \gamma && \text{if } (A^T \lambda^*)_i = 1, \\ \bar{x}_i &\leq -\gamma && \text{if } (A^T \lambda^*)_i = -1, \\ \bar{x}_i &= \gamma(A^T \lambda^*)_i && \text{if } -1 < (A^T \lambda^*)_i < 1, \end{aligned}$$

and $A\bar{x} = y$. □

An immediate consequence of the previous result is the following important invariance property.

COROLLARY 3. *If the solution λ^δ of (2.4) is the normal solution λ^* to (2.1) for some $\gamma = \delta > 0$, then for all $\gamma \in (0, \delta]$ $s^\gamma(X^\gamma)$ (and, by extension, $W^\gamma(X^\gamma)$) remains unchanged.*

As a final remark before closing the preliminaries, the problem [HL1] can be solved numerically (as was done to produce the numerical illustration of section 5) as the equivalent convex quadratic programming problem in variables x, p, q (cf. [20]):

$$\min \frac{1}{2\gamma} \sum_{i=1}^n p_i^2 + \sum_{i=1}^n \left(q_i - \frac{\gamma}{2} \right)$$

subject to

$$\begin{aligned} -p - q &\leq x \leq p + q, \\ p &\leq \gamma e, q \geq 0, \\ Ax &= y, \end{aligned}$$

where e denotes a vector of all ones.

3. A sufficient condition. The following results building up to the sufficient condition are obtained, mutatis mutandis, from the proofs of earlier results which will be pointed out explicitly for the interested reader.

THEOREM 2. *Suppose that the solution of (2.4) for some $\gamma = \delta$ is the normal solution to (2.1). Then*

$$(3.1) \quad X^\tau = \frac{\tau - \alpha}{\beta - \alpha} X^\beta + \frac{\beta - \tau}{\beta - \alpha} X^\alpha \text{ for } 0 \leq \alpha < \tau < \beta \leq \delta,$$

where $\mu X + \nu Y = \{\mu x + \nu y : x \in X, y \in Y\}$.

The proof is obtained using arguments similar to those of the proofs of Theorems 3.3 and 3.4 of [15]. It is given in the appendix for the reader's convenience. The only noticeable change in the proof of Theorem 3.4 of [15] is that one needs to work with the matrix $\begin{pmatrix} \bar{W}^\gamma(x) \\ A \end{pmatrix}$, where \bar{W}^γ is the diagonal matrix whose i th diagonal entry is equal to one if $|x_i| \leq \gamma$ and to zero otherwise.

Using the above theorem, one immediately obtains the following key result.

THEOREM 3. *Suppose that the solution of (2.4) for some $\gamma = \delta$ is the normal solution λ^* of (2.1). Let $\gamma \in (0, \delta]$. If the Huber loss problem (2.3) has a solution x^γ with $\begin{pmatrix} W^\gamma(x^\gamma) \\ A \end{pmatrix}$ having rank equal to n , then problem (1.1) has a unique solution x^0 which is given as $x^0 = x^\gamma + \gamma v$, where v is the unique solution to the linear system of equations*

$$(3.2) \quad \begin{pmatrix} W^\gamma(x^\gamma) \\ A \end{pmatrix} h = \begin{pmatrix} s^\gamma(x^\gamma) - A^T \lambda^* \\ 0 \end{pmatrix}.$$

Proof. Let $\alpha = 0$, $\beta = \delta$, and $0 < \tau < \delta$ in the previous theorem. One has $X^\tau \supset \frac{\tau}{\delta} X^\delta + (1 - \frac{\tau}{\delta}) X^0$. Since the matrix $\begin{pmatrix} W^\tau(x^\tau) \\ A \end{pmatrix}$ has rank equal to n by hypothesis, then X^τ is a singleton. If X^τ is a singleton, i.e., $X^\tau = \{x^\tau\}$, then $X^0 = \{x^0\}$ and $X^\delta = \{x^\delta\}$ are both singletons. One has that x^τ satisfies the optimality equations

$$\frac{1}{\tau} W^\tau(x^\tau) x^\tau + s^\tau(x^\tau) - A^T \lambda^* = 0,$$

$$A x^\tau = y.$$

As x^τ is an affine function of τ by the above argument, and since x^0 satisfies

$$W^\tau(x^\tau) x^0 = 0$$

by virtue of its optimality and the fact that λ^* solves (2.1), one can replace x^τ in the above equations by $x^0 - \tau v$, where v is the unique solution to the linear system

$$\begin{pmatrix} W^\tau(x^\tau) \\ A \end{pmatrix} h = \begin{pmatrix} s^\tau(x^\tau) - A^T \lambda^* \\ 0 \end{pmatrix}. \quad \square$$

Now, the sufficient condition for sparse recovery can be stated.

COROLLARY 4. *Suppose that the solution of (2.4) for some $\gamma = \delta$ is the normal solution λ^* of (2.1). Let $\gamma \in (0, \delta]$. If the Huber loss problem (2.3) has a solution x^γ with $\begin{pmatrix} W^\gamma(x^\gamma) \\ A \end{pmatrix}$ having rank equal to n , and $\text{sgn}(x^\gamma + \gamma v) = \text{sgn}(u)$, where v is the unique solution to the linear system of equations*

$$(3.3) \quad \begin{pmatrix} W^\gamma(x^\gamma) \\ A \end{pmatrix} h = \begin{pmatrix} s^\gamma(x^\gamma) - A^T \lambda^* \\ 0 \end{pmatrix},$$

then exact individual recovery by BP takes place.

Remark 2. One can easily check the sufficient condition by solving at most two strictly convex quadratic programs for two sufficiently small values of γ , and a linear system of equations if the W^γ matrix remains unchanged for these two values of γ .

Adjoining a regularity condition, one can render the sufficient condition necessary as well.

DEFINITION 2. A solution $\bar{\lambda}$ to (2.1) is said to be nondegenerate if there exists a solution \bar{x} to (1.1) such that

$$(3.4) \quad \begin{aligned} \bar{x}_i &> 0 & \text{if } (A^T \bar{\lambda})_i &= 1, \\ \bar{x}_i &< 0 & \text{if } (A^T \bar{\lambda})_i &= -1, \\ \bar{x}_i &= 0 & \text{if } -1 < (A^T \bar{\lambda})_i < 1. \end{aligned}$$

Define $\mathcal{A}^\gamma(x) = \{i : |x_i| \leq \gamma\}$, $\mathcal{N}^\gamma(x) = \{i : |x_i| > \gamma\}$, with $\mathcal{A}^0(x) = \{i : x_i = 0\}$, $\mathcal{N}^0(x) = \{i : |x_i| > 0\}$. The proof of the following result follows easily from the necessary modifications to the proof of Theorem 3.6 [15].

THEOREM 4. Suppose that the normal solution λ^* of (2.1) is nondegenerate, and that exact individual recovery by BP takes place. Then there exists a $\delta > 0$ such that there is a unique Huber loss solution x^γ for $0 < \gamma \leq \delta$ with $\mathcal{A}^\gamma(x^\gamma) = \mathcal{A}^0(u)$ and $\mathcal{N}^0(u) = \mathcal{N}^\gamma(x^\gamma)$.

Proof. By hypothesis, u is the unique solution of (1.1). Let $\delta > 0$ be such that the solution λ^γ of 2.4) is the normal solution λ^* to (2.1). Since λ^* is nondegenerate by hypothesis, one has

$$(3.5) \quad \begin{aligned} u_i &> 0 & \text{if } (A^T \lambda^*)_i &= 1, \\ u_i &< 0 & \text{if } (A^T \lambda^*)_i &= -1, \\ u_i &= 0 & \text{if } -1 < (A^T \lambda^*)_i < 1. \end{aligned}$$

Let W be the diagonal matrix with $W_{ii} = 1$ for $i : |(A^T \lambda^*)_i| < 1$. Then the matrix $\begin{pmatrix} W \\ A \end{pmatrix}$ has full rank. To see why, assume $\begin{pmatrix} W \\ A \end{pmatrix} z = 0$. By (3.5) there exists a sufficiently small ϵ such that

$$(3.6) \quad \begin{aligned} (u + \epsilon z)_i &> 0 & \text{if } (A^T \lambda^*)_i &= 1, \\ (u + \epsilon z)_i &< 0 & \text{if } (A^T \lambda^*)_i &= -1, \\ (u + \epsilon z)_i &= 0 & \text{if } -1 < (A^T \lambda^*)_i < 1. \end{aligned}$$

Thus, $u + \epsilon z$ solves (1.1). But uniqueness of u implies $z = 0$, and hence the claim holds. Now, since the system (with $s = s^\gamma(A^T \lambda^*)$)

$$\begin{pmatrix} W \\ A \end{pmatrix} v = \begin{pmatrix} s - A^T \lambda^* \\ 0 \end{pmatrix}$$

has a unique solution v , then $x^\gamma = u - \gamma v$ uniquely solves (2.3) for $\gamma \in (0, \delta]$ with $\mathcal{A}^\gamma(x^\gamma) = \mathcal{A}^0(u)$ and $\mathcal{N}^0(u) = \mathcal{N}^\gamma(x^\gamma)$. \square

Compare the above result to Theorem 1 above, and notice that the dual condition in part (b) is in fact a nondegeneracy condition on the dual vector h .

4. A necessary condition. In this section, the following reasonable assumptions (reasonable since A can be reduced by suitable elimination to a full rank matrix, and if problem [SL1] does not have a unique solution, one cannot even begin to talk about exact recovery) will be in force in addition to the initial assumptions of full rank on A and that the number $\text{nnz}(u)$ of nonzero components of u is less than m , which are stated as well.

ASSUMPTION 1. $\text{nnz}(u) < m$.

ASSUMPTION 2. A has full rank ($= m$).

ASSUMPTION 3. Problem (1.1) has a unique solution.

THEOREM 5. *Under Assumptions 1–3, suppose that the solution of (2.4) for some $\gamma = \delta$ is the normal solution λ^* of (2.1). Let $\gamma \in (0, \delta]$. If exact individual recovery by BP takes place, Λ is not a singleton, or, equivalently the normal solution to (2.1) is not the unique solution to (2.1).*

Proof. Let Λ be a singleton. Under the premises of the theorem, solving problem (2.4) for $\gamma \in (0, \delta]$, one gets the normal solution λ^* to (2.1). Basic theory of linear programming implies that the unique (optimal) solution λ^* to (2.1) is also an extreme point of the polyhedral set $D = \{\lambda \in \mathbb{R}^m : A^T \lambda \leq 1, A^T \lambda \geq -1\}$. Since any extreme point of the set D has at least m of the defining inequalities of D holding as equalities, the cardinality of the set $I(\lambda^*) = \{i : |(A^T \lambda^*)_i| = 1\}$ is at least equal to m (see, e.g., [1]). By the Goldman–Tucker strict complementarity theorem [23] and Assumption 3, the unique solution \bar{x} to (1.1) has the property that

$$\bar{x}_i \neq 0 \quad \forall i \in I(\lambda^*).$$

Hence $\text{nnz}(\bar{x}) \geq m$, which renders exact individual recovery by BP impossible. \square

The above result provides an interesting (and hitherto unnoticed, to the best of the author's knowledge) link between the problem of sparse recovery and the optimal set of the dual problem (2.1) via a least norm solution of the dual problem. In particular one has that solving the dual problem for $\gamma = 0$ will never yield a unique solution if the solution is sparse. One naturally wonders whether the statement is also sufficient. However, under the conditions stated (Assumptions 1–3) it turns out that the converse of the above statement is not necessarily true. A counterexample is provided in the next section in Example 3. On the other hand, a consequence of part (a) of Theorem 1 in the introduction is that if a vector $u \in \mathbb{R}^n$ with support S satisfies the condition in part (a), then for all vectors $u' \in \mathbb{R}^n$ with support $S' \subset S$ and $\text{sgn}(u')_{S'} = \text{sgn}(u)_{S'}$, exact individual recovery holds (see Remark 4.27 of [10] on p. 92). Then it follows that for any vector $u' \in \mathbb{R}^n$ with support $S' \subset S$ and $\text{sgn}(u')_{S'} = \text{sgn}(u)_{S'}$, the optimal set Λ of (2.1) is not a singleton. Hence, the necessary condition of Theorem 5 has a repercussion on the exact recovery of all such vectors u' regardless of their numerical values, provided the sign pattern of the nonzero entries (reduced in number) is preserved.

The above result can be equivalently rephrased as follows.

COROLLARY 5. *Under the premises of the previous theorem, if Λ is a singleton, or, equivalently, if the normal solution to (2.1) is the unique solution to (2.1), then exact individual recovery by BP of sparse vector u with $\text{nnz}(u) < m$ is impossible.*

DEFINITION 3. *An instance of [SL1] is primal nondegenerate if at any point x the vectors $\{(\begin{smallmatrix} e_j \\ A_j \end{smallmatrix}) : j \in \mathcal{A}^0(x)\}$ are linearly independent (e_j denotes the j th basis vector in \mathbb{R}^n and A_j the j th column of A).*

DEFINITION 4. *An instance of the [SL1] is dual nondegenerate if at any point λ satisfying $\|A^T \lambda\|_\infty \leq 1$: $\text{card}(\{i : |(A^T \lambda)_i| = 1\}) \leq m$.*

Remark 3. According to [19], primal and dual nondegeneracy imply unique primal and dual optimal solutions; therefore, one can reiterate the previous corollary under these strong assumptions. Attention is drawn to the similarity of the conditions imposed below to part (b) of Theorem 1.

COROLLARY 6. *Under Assumptions 1–2, if [SL1] is primal and dual nondegenerate, then the normal solution to (2.1) is the unique solution to (2.1), and exact individual recovery by BP of sparse vector u with $\text{nnz}(u) < m$ is impossible.*

5. Numerical illustration.

Example 1. Consider a 5×15 matrix A (randomly generated) where the columns are given below (rounded to three digits of accuracy); the first six columns of A are

$$\begin{pmatrix} -0.313 & 0.328 & 0.464 & -1.830 & -0.732 & -0.972 \\ -0.721 & -1.299 & -1.375 & 0.320 & 1.187 & -1.153 \\ 2.213 & -0.625 & -0.218 & -0.962 & -0.108 & -0.263 \\ -1.431 & 2.099 & 0.673 & 0.169 & 1.649 & 1.268 \\ 0.071 & -0.123 & -1.761 & -0.001 & -0.315 & -0.295 \end{pmatrix},$$

the next six columns are

$$\begin{pmatrix} -0.394 & 0.935 & -0.759 & 0.000 & -0.909 & 0.344 \\ 0.710 & 1.017 & -0.724 & -0.999 & 0.006 & -0.995 \\ -0.527 & -1.037 & 1.982 & 0.826 & 1.004 & -0.752 \\ 1.805 & -1.248 & -1.594 & -0.812 & -0.802 & -0.543 \\ -1.133 & -1.224 & -0.056 & 0.147 & -0.605 & 1.294 \end{pmatrix},$$

and, finally, the last three columns are

$$\begin{pmatrix} -0.662 & -0.486 & 1.884 \\ -0.907 & 0.663 & 1.571 \\ 0.756 & 1.200 & 0.844 \\ 1.283 & -0.381 & -0.714 \\ -0.194 & 0.778 & -0.567 \end{pmatrix}.$$

Let u be the zero vector except for a one in the 12th position. The vector $y = Au$ is equal to the 12th column of A . The critical value δ for γ is approximately equal to 0.35. For example, at $\gamma = 0.035$ one gets the normal solution λ^* to (2.1):

$$\lambda^* = (0.110 \quad -0.233 \quad -0.239 \quad -0.115 \quad 0.378)^T,$$

and the unique Huber loss minimizer

$$x^\gamma = \begin{pmatrix} -0.007 \\ 0.007 \\ -0.011 \\ -0.002 \\ -0.022 \\ -0.001 \\ -0.025 \\ -0.007 \\ -0.008 \\ 0.006 \\ -0.017 \\ 0.937 \\ -0.009 \\ -0.005 \\ -0.017 \end{pmatrix}$$

with

$$s^\gamma = (0 \ 0 \ 0 \ 0 \ 0 \ 0 \ 0 \ 0 \ 0 \ 0 \ 0 \ 1 \ 0 \ 0 \ 0)^T$$

with a full rank matrix $\begin{pmatrix} W^\gamma \\ A \end{pmatrix}$. The unique solution v to the system

$$\begin{pmatrix} W^\gamma \\ A \end{pmatrix} h = \begin{pmatrix} s^\gamma - A^T \lambda^* \\ 0 \end{pmatrix}$$

is obtained as

$$v = \begin{pmatrix} 0.205 \\ -0.200 \\ 0.319 \\ 0.065 \\ 0.639 \\ 0.033 \\ 0.718 \\ 0.205 \\ 0.227 \\ -0.184 \\ 0.477 \\ 1.804 \\ 0.263 \\ 0.157 \\ 0.492 \end{pmatrix},$$

from which $\text{sgn}(x^\gamma + \gamma v) = \text{sgn}(u)$ and $x^\gamma + \gamma v$ yields u . The sufficient condition is thus satisfied. With respect to the necessary condition, solving problem (1.1), one gets the dual vector

$$(0.012 \ -0.629 \ 0.096 \ -0.084 \ 0.307)^T,$$

entirely different from the normal solution reported above.

Example 2. The example is for a 6×20 matrix. The first six columns are

$$\begin{pmatrix} 0.337 & 0.171 & -1.382 & 0.506 & -1.419 & -0.718 \\ 1.245 & -0.706 & 0.007 & -0.793 & -0.461 & -0.307 \\ -1.294 & -0.257 & -0.233 & 1.430 & 0.393 & 1.290 \\ 0.179 & 1.456 & -1.174 & -0.096 & 0.000 & -0.596 \\ -0.854 & 0.223 & -0.631 & 1.846 & -0.696 & -1.524 \\ -0.605 & 0.491 & 0.196 & 0.814 & -1.176 & -0.962 \end{pmatrix},$$

the next six columns are

$$\begin{pmatrix} 1.176 & -0.192 & -0.139 & -0.347 & -0.636 & 0.671 \\ 1.431 & -1.062 & 0.104 & 1.251 & -0.199 & -0.102 \\ -0.450 & -0.866 & 1.233 & -0.179 & 0.425 & -1.487 \\ 0.801 & 0.606 & -1.360 & 0.417 & 0.040 & -0.857 \\ -0.549 & -1.220 & -0.143 & -0.222 & 0.537 & 0.157 \\ -0.069 & -0.424 & 1.353 & 0.024 & 0.808 & 0.423 \end{pmatrix},$$

and the last eight columns are

$$\begin{pmatrix} 0.038 & 0.427 & 0.752 & -1.262 & -0.749 & 0.141 & 1.873 & -0.791 \\ 0.198 & -0.366 & 1.105 & -1.285 & 1.164 & -0.860 & -1.547 & -1.581 \\ -0.597 & 0.811 & -0.871 & -1.318 & 0.606 & -1.900 & 0.030 & 0.742 \\ 0.040 & 0.672 & -1.295 & 0.657 & 1.406 & 0.581 & -0.185 & 1.110 \\ 0.530 & -0.876 & 0.186 & 2.042 & -0.140 & 1.124 & -0.483 & 2.195 \\ -0.036 & -0.351 & -0.011 & -1.249 & 1.294 & 0.641 & 1.076 & 0.039 \end{pmatrix}.$$

The sparse vector u to be recovered has nonzeros equal to $-1.000, 2.000, 1.000$ at the 6th, 9th, and 19th positions. The vector y is then

$$(2.313, -1.031, 1.206, -2.309, 0.754, 4.744)^T.$$

In this example, already for $\gamma \leq 0.22$ one gets the normal solution

$$\lambda^* = (9.177252 \times 10^{-4}, -0.112, -0.203, -0.161, 0.077, 0.779)^T.$$

For $\gamma = 0.0025$, the sign vector $s^\gamma(x^\gamma)$ of the unique minimizer x^γ is

$$(0, 0, 0, 0, -1, -1, 0, 0, 1, 0, 0, 0, 0, 0, 0, 0, 0, 1, 0)^T,$$

where

$$x^\gamma = \begin{pmatrix} -0.001 \\ 7.4264 \times 10^{-4} \\ 8.4498 \times 10^{-4} \\ 0.001 \\ -0.006 \\ -0.991 \\ -7.333 \times 10^{-4} \\ -5.690 \times 10^{-4} \\ 1.995 \\ -4.254 \times 10^{-4} \\ 0.001 \\ 0.002 \\ 2.6465 \times 10^{-4} \\ -0.001 \\ 6.7064 \times 10^{-4} \\ -0.001 \\ 0.001 \\ 0.002 \\ 0.998 \\ 1.1715 \times 10^{-4} \end{pmatrix}.$$

Note that the s^γ is not equal to the sign vector of u ! Solving for the unique solution v from the system with the full rank matrix on the left,

$$\begin{pmatrix} W^\gamma \\ A \end{pmatrix} h = \begin{pmatrix} s^\gamma - A^T \lambda^* \\ 0 \end{pmatrix},$$

one obtains

$$v = \begin{pmatrix} 0.443 \\ -0.297 \\ -0.338 \\ -0.591 \\ 2.511 \\ -3.495 \\ 0.293 \\ 0.228 \\ 2.113 \\ 0.170 \\ -0.600 \\ -0.794 \\ -0.106 \\ 0.573 \\ -0.268 \\ 0.511 \\ -0.517 \\ -0.975 \\ 0.700 \\ -0.047 \end{pmatrix},$$

from which $x^\gamma + \gamma v$ results exactly in u , verifying the sufficient sign condition first, i.e., $\text{sgn}(x^\gamma + \gamma v) = \text{sgn}(u)$. Again, λ^* is not the only dual solution: an extreme point solution is

$$\bar{\lambda} = (0.115, -0.183, -0.180, -0.421, 0.311, 0.539)^T.$$

Example 3. This example shows that the condition of having multiple dual solutions to problem (2.1) is not sufficient for exact individual recovery. The matrix $A \in \mathbb{R}^{4 \times 5}$ (with full rank equal to 4) is given as follows:

$$\begin{pmatrix} 1 & 2 & 0 & 1 & 0 \\ 0 & 2 & 3 & 0 & 0 \\ 0 & 0 & 4 & 0 & -8 \\ 1 & 4 & 3 & 0 & 7 \end{pmatrix},$$

while the sparse vector to be recovered is $u = (0, 0, 0, 7, 1)^T$. The normal solution to (2.1) is

$$\lambda^* = (0.840, -0.660, 0.125, 0.160)^T,$$

whereas an extreme point optimal solution is obtained as

$$\lambda^0 = (1, -0.5, 0.125, 0)^T.$$

Hence, Λ is not a singleton. However, solving the BP problem (1.1), one gets the unique solution $x^\# = (1, 3, -2, 0, 0)$ instead of u . Thus, no exact individual recovery is observed.

Appendix A. Proof of Theorem 2. Let $\text{ext}X^\gamma$ denote the set of extreme points of X^γ .

Before delving into the proof of Theorem 2, the following result, which is lifted from [15] without proof in the interest of brevity (the proof is an almost verbatim

repetition of the proof of the corresponding result of [15]), is stated since it will be used below.

THEOREM 6. *Let δ be such that the solution to (2.4) for $\gamma = \delta$ is the normal solution λ^* to (2.1). Then there exists a positive constant η such that the Hausdorff distance $H(X^\gamma, X^{\gamma'})$ between X^γ and $X^{\gamma'}$ satisfies*

$$(A.1) \quad H(X^\gamma, X^{\gamma'}) \leq \eta |\gamma - \gamma'| \quad \text{for } \gamma, \gamma' \in [0, \delta].$$

Now, to start the proof of Theorem 2, assume, as in Lemma 6, that λ^* is the solution to (2.4) for $0 < \gamma \leq \delta$. System (2.5) reduces to (2.2) for $\gamma = 0$. Therefore, using Lemma 1 and Corollary 2, $\bar{x} \in X^\gamma$ if and only if system (2.5) holds for $0 \leq \gamma \leq \delta$.

Let $0 \leq \alpha < \tau < \beta \leq \delta$, $x^\alpha \in X^\alpha$, $x^\beta \in X^\beta$, and $x^\tau := \frac{\tau-\alpha}{\beta-\alpha}x^\beta + \frac{\beta-\tau}{\beta-\alpha}x^\alpha$. Then conditions (2.5) hold for $\gamma = \alpha$, $\bar{x} = x^\alpha$. Since $\frac{\tau-\alpha}{\beta-\alpha} > 0$ one gets

$$(A.2) \quad \begin{aligned} \frac{\tau-\alpha}{\beta-\alpha}x_i^\beta &\geq \frac{\tau-\alpha}{\beta-\alpha}\beta && \text{if } (A^T\lambda^*)_i = 1, \\ \frac{\tau-\alpha}{\beta-\alpha}x_i^\beta &\leq -\frac{\tau-\alpha}{\beta-\alpha}\beta && \text{if } (A^T\lambda^*)_i = -1, \\ \frac{\tau-\alpha}{\beta-\alpha}x_i^\beta &= \frac{\tau-\alpha}{\beta-\alpha}\beta(A^T\lambda^*)_i && \text{if } -1 < (A^T\lambda^*)_i < 1. \end{aligned}$$

By the same token, one has

$$(A.3) \quad \begin{aligned} \frac{\beta-\tau}{\beta-\alpha}x_i^\alpha &\geq \frac{\beta-\tau}{\beta-\alpha}\alpha && \text{if } (A^T\lambda^*)_i = 1, \\ \frac{\beta-\tau}{\beta-\alpha}x_i^\alpha &\leq -\frac{\beta-\tau}{\beta-\alpha}\alpha && \text{if } (A^T\lambda^*)_i = -1, \\ \frac{\beta-\tau}{\beta-\alpha}x_i^\alpha &= \frac{\beta-\tau}{\beta-\alpha}\alpha(A^T\lambda^*)_i && \text{if } -1 < (A^T\lambda^*)_i < 1. \end{aligned}$$

Adding the corresponding inequalities in (A.2) and (A.3), one obtains the conditions (2.5) for $\gamma = \tau$. Furthermore, it is clear that $Ax^\tau = y$. Hence, $x^\tau \in X^\tau$. Therefore, it has been established that

$$(A.4) \quad X^\tau \supset \frac{\tau-\alpha}{\beta-\alpha}X^\beta + \frac{\beta-\tau}{\beta-\alpha}X^\alpha.$$

Let \bar{W}^γ be the diagonal matrix with the i th diagonal entry equal to one if $|x_i| \leq \gamma$, and to zero otherwise. By the same token, $\bar{s}^\gamma(\cdot)$ is defined as

$$\bar{s}_i^\gamma(t) = \begin{cases} 0 & \text{if } |t| \leq \gamma, \\ 1 & \text{if } t > \gamma, \\ -1 & \text{if } t < -\gamma. \end{cases}$$

CLAIM. *x is an extreme point of X^γ if and only if $C := (\bar{W}^\gamma(x))_A$ has rank equal to n .*

Proof. Suppose $x \in X^\gamma$ with dual vector λ and C has rank less than n . Then there exists a vector $h \neq 0$ such that $Ch = 0$. Pick a small positive number ε such that $(x_i \pm \varepsilon h_i) > \gamma$ if $x_i > \gamma$ and $(x_i \pm \varepsilon h_i) < -\gamma$ if $x_i < -\gamma$. Therefore, one has

$$\frac{1}{\gamma}\bar{W}^\gamma(x \pm \varepsilon h) + \bar{s}(x \pm \varepsilon h) - A^T\lambda = \frac{1}{\gamma}\bar{W}^\gamma x + \bar{s}(x) - A^T\lambda = 0.$$

Therefore, $x \pm \varepsilon h \in X^\gamma$. As $x = \frac{1}{2}(x + \varepsilon h) + \frac{1}{2}(x - \varepsilon h)$, x is not an extreme point of X^γ .

On the other hand, let $x \in X^\gamma$ and C have rank equal to n , and suppose $u, v \in X^\gamma$ such that $x = \alpha u + (1 - \alpha)v$ for some $\alpha \in (0, 1)$. Then it is a consequence of (2.5) that $x_i = \gamma(A^T \lambda)_i$ if and only if $u_i = v_i = \gamma(A^T \lambda)_i$, and, furthermore $Ax = Au = Av$, which implies that $Cx = Cu = Cv$. Since C has full rank, then $x = u = v$, and thus x is an extreme point of X^γ . This completes the proof of the claim.

Now, assume δ is such that for $\gamma = \delta$, the solution of (2.4) is the normal solution to (2.1). Suppose $\alpha < \beta \leq \delta$ and $\tau = \zeta\alpha + (1 - \zeta)\beta$ for $\zeta \in (0, 1)$. Let x^α and x^β be extreme points of X^α and X^β , respectively. By the previous development above it follows that $x^\tau := \zeta x^\alpha + (1 - \zeta)x^\beta$ is a member of X^τ . If $\bar{W}^\alpha(x^\alpha) = \bar{W}^\beta(x^\beta)$, then it follows that $\bar{W}^\tau(x^\tau) = \bar{W}^\alpha(x^\alpha) = \bar{W}^\beta(x^\beta)$. Therefore, x^τ is an extreme point of X^τ . Now, define the set of diagonal matrices $\mathcal{W}_\gamma = \{\bar{W}^\gamma(x) : x \in \text{ext} X^\gamma\}$. Let $\bar{W} \in \cup_{\gamma > 0} \mathcal{W}_\gamma$. Then, by the above argument, the set $\{\gamma : \bar{W} \in \mathcal{W}_\gamma\}$ is a segment of a line. Therefore, there exists a small positive constant ϵ such that either

$$(!) \quad \bar{W} \in \mathcal{W}_\gamma \text{ for } 0 < \gamma \leq \epsilon,$$

or

$$(!!) \quad \bar{W} \notin \mathcal{W}_\gamma \text{ for } 0 < \gamma \leq \epsilon.$$

Since $\cup_{\gamma > 0} \mathcal{W}_\gamma$ is a finite set, one can choose ϵ such that one of the above alternatives holds for every $\bar{W} \in \cup_{\gamma > 0} \mathcal{W}_\gamma$. Define $\mathcal{W}^* = \{\bar{W} \in \cup_{\gamma > 0} \mathcal{W}_\gamma : (!) \text{ holds}\}$. Let $0 < \alpha < \beta \leq \epsilon$ and $\alpha < \tau < \beta$. Let $x^\tau \in \text{ext} X^\tau$. Then $\bar{W} := \bar{W}^\tau(x^\tau) \in \mathcal{W}^*$. By definition of \mathcal{W}^* there exist $x^\alpha \in X^\alpha$ and $x^\beta \in X^\beta$ such that $\bar{W} = \bar{W}^\alpha(x^\alpha) = \bar{W}^\beta(x^\beta)$. It was proven above that $\bar{x}^\tau \in X^\tau$ and $\bar{W}^\tau(\bar{x}^\tau) = \bar{W}$, where

$$\bar{x}^\tau := \frac{\tau - \alpha}{\beta - \alpha} x^\beta + \frac{\beta - \tau}{\beta - \alpha} x^\alpha.$$

Hence, $C\bar{x}^\tau = Cx^\tau$. Since C has full rank, $\bar{x}^\tau = x^\tau$. This implies

$$x^\tau := \frac{\tau - \alpha}{\beta - \alpha} x^\beta + \frac{\beta - \tau}{\beta - \alpha} x^\alpha,$$

which further means that $\text{ext} X^\tau \subseteq \bar{X} := \frac{\tau - \alpha}{\beta - \alpha} X^\beta + \frac{\beta - \tau}{\beta - \alpha} X^\alpha$.

Now, it is evident that $\|x\|_H \rightarrow \infty$ as $\|x\| \rightarrow \infty$, which implies that X^γ is a bounded set. Since X^γ is the set of dual solutions to the strictly convex quadratic programming problem (2.4), it is a convex polyhedral subset of \mathbb{R}^n . Since any point in the polytope X^τ is a convex combination of extreme points of X^τ , any point in X^τ is a convex combination of elements of \bar{X} . Since \bar{X} is a convex set, one obtains

$$X^\tau \subset \frac{\tau - \alpha}{\beta - \alpha} X^\beta + \frac{\beta - \tau}{\beta - \alpha} X^\alpha.$$

Combined with (A.4), one gets

$$X^\tau = \frac{\tau - \alpha}{\beta - \alpha} X^\beta + \frac{\beta - \tau}{\beta - \alpha} X^\alpha$$

for $\alpha > 0$. By Theorem 6, as $\alpha \rightarrow 0^+$, $X^\gamma \rightarrow X^0$. Therefore, one has

$$X^\tau = \frac{\tau - \alpha}{\beta - \alpha} X^\beta + \frac{\beta - \tau}{\beta - \alpha} X^\alpha$$

for $\alpha \geq 0$. □

Acknowledgments. The author would like to thank two anonymous referees for suggestions and corrections that improved the presentation. Special thanks are due to Can Kızılkale for his help with the counterexample in Example 3.

REFERENCES

- [1] D. BERTSIMAS AND J. TSIKSIKLIS, *Introduction to Linear Optimization*, Athena Scientific, Belmont, MA, 1997.
- [2] K. BRYAN AND T. LEISE, *Making do with less: An introduction to compressed sensing*, SIAM Rev., 55 (2013), pp. 547–566, <https://doi.org/10.1137/110837681>.
- [3] E. J. CANDÈS, J. ROMBERG, AND T. TAO, *Robust uncertainty principles: Exact signal reconstruction from highly incomplete frequency information*, IEEE Trans. Inform. Theory, 52 (2006), pp. 489–509.
- [4] E. J. CANDÈS AND T. TAO, *Decoding by linear programming*, IEEE Trans. Inform. Theory, 52 (2006), pp. 4203–4215.
- [5] A. D’ASPREMONT AND L. EL GHAOU, *Testing the nullspace property using semidefinite programming*, Math. Program., 127 (2011), pp. 123–144.
- [6] B. T. CHEN, K. MADSEN, AND SH. ZHANG, *On the characterization of quadratic splines*, J. Optim. Theory Appl., 124 (2005), pp. 93–111.
- [7] S. S. CHEN, D. L. DONOHO, AND M. A. SAUNDERS, *Atomic decomposition by basis pursuit*, SIAM J. Sci. Comput., 20 (1998), pp. 33–61, <https://doi.org/10.1137/S1064827596304010>.
- [8] D. L. DONOHO, *Compressed sensing*, IEEE Trans. Inform. Theory, 51 (2005), pp. 1289–1306.
- [9] D. L. DONOHO, *For most large underdetermined systems of linear equations the minimal ℓ_1 -norm solution is also the sparsest solution*, Comm. Pure Appl. Math., 59 (2006), pp. 797–829.
- [10] S. FOUCART AND H. RAUHUT, *A Mathematical Introduction to Compressive Sensing*, Springer, New York, 2013.
- [11] J. J. FUCHS, *On sparse representations in arbitrary redundant bases*, IEEE Trans. Inform. Theory, 50 (2004), pp. 1341–1344.
- [12] J. C. GILBERT, *On the solution uniqueness characterization in the L_1 norm and polyhedral gauge recovery*, J. Optim. Theory Appl., 172 (2015), pp. 70–101.
- [13] P. J. HUBER AND E. RONCHETTI, *Robust Statistics*, John Wiley and Sons, New York, 2009.
- [14] A. JUDITSKY AND A. NEMIROVSKI, *On verifiable sufficient conditions for sparse signal recovery via ℓ_1 minimization*, Math. Program. Ser. B, 127 (2011), pp. 57–88.
- [15] W. LI AND J. J. SWETTIS, *The linear ℓ_1 estimator and the Huber M -estimator*, SIAM J. Optim., 8 (1998), pp. 457–475, <https://doi.org/10.1137/S1052623495293160>.
- [16] K. MADSEN AND H. B. NIELSEN, *A finite smoothing algorithm for linear ℓ_1 estimation*, SIAM J. Optim., 3 (1993), pp. 223–235, <https://doi.org/10.1137/0803010>.
- [17] K. MADSEN, H. B. NIELSEN, AND M. Ç. PINAR, *New characterizations of ℓ_1 solutions to over-determined systems of linear equations*, Oper. Res. Lett., 16 (1994), pp. 159–166.
- [18] O. L. MANGASARIAN AND R. R. MEYER, *Nonlinear perturbation of linear programs*, SIAM J. Control Optim., 17 (1979), pp. 745–752, <https://doi.org/10.1137/0317052>.
- [19] M. R. OSBORNE, *Finite Algorithms in Optimization and Data Analysis*, John Wiley and Sons, Chichester, UK, 1985.
- [20] M. Ç. PINAR, *Duality in robust linear regression using Huber’s M -estimator*, Appl. Math. Lett., 10 (1997), pp. 65–70.
- [21] I. SELESNICK, *Sparse regularization via convex analysis*, IEEE Trans. Signal Process., 65 (2017), pp. 4481–4494.
- [22] J. A. TROPP, *Recovery of short, complex linear combinations via ℓ_1 minimization*, IEEE Trans. Inform. Theory, 51 (2005), pp. 1568–1570.
- [23] A. C. WILLIAMS, *Complementarity theorems for linear programming*, SIAM Rev., 12 (1970), pp. 135–137, <https://doi.org/10.1137/1012015>.
- [24] H. ZHANG, W. YIN, AND L. CHENG, *Necessary and sufficient conditions of solution uniqueness in 1 -norm minimization*, J. Optim. Theory Appl., 164 (2015), pp. 109–122.