RSP-Based Analysis for Sparsest and Least ℓ_1-Norm Solutions to Underdetermined Linear Systems

Yun-Bin Zhao *IEEE member*

Abstract—Recently, the worse-case analysis, probabilistic analysis and empirical justification have been employed to address the fundamental question: When does ℓ_1-minimization find the sparsest solution to an underdetermined linear system? In this paper, a deterministic analysis, rooted in the classic linear programming theory, is carried out to further address this question. We first identify a necessary and sufficient condition for the uniqueness of least ℓ_1-norm solutions to linear systems. From this condition, we deduce that a sparsest solution coincides with the unique least ℓ_1-norm solution to a linear system if and only if the so-called *range space property* (RSP) holds at this solution. This yields a broad understanding of the relationship between ℓ_0- and ℓ_1-minimization problems. Our analysis indicates that the RSP truly lies at the heart of the relationship between these two problems. Through RSP-based analysis, several important questions in this field can be largely addressed. For instance, how to efficiently interpret the gap between the current theory and the actual numerical performance of ℓ_1-minimization by a deterministic analysis, and if a linear system has multiple sparsest solutions, when does ℓ_1-minimization guarantee to find one of them? Moreover, new matrix properties (such as the *RSP of order K* and the *Weak-RSP of order K*) are introduced in this paper, and a new theory for sparse signal recovery based on the RSP of order K is established.

Index Terms—Underdetermined linear system, sparsest solution, least ℓ_1-norm solution, range space property, strict complementarity, sparse signal recovery, compressed sensing.

I. INTRODUCTION

Many problems across disciplines can be formulated as the problem of finding the sparsest solution to underdetermined linear systems. For instance, many data types arising from signal and image processing can be sparsely represented and the processing tasks (e.g., compression, reconstruction, separation, and transmission) often amount to the problem

$$\min\{\|x\|_0 : Ax = b\},$$

where A is an $m \times n$ full-rank matrix with $m < n$, b is a vector in R^m, and $\|x\|_0$ denotes the number of nonzero components of the vector x. The system $Ax = b$ is underdetermined, and thus it has infinitely many solutions. The problem above seeking the sparsest solution to a linear system is called ℓ_0-minimization in the literature, to which even the profound linear algebra algorithms do not apply. The ℓ_0-minimization is known to be NP-hard [30]. From a computational point of view, it is natural to consider its ℓ_1-approximation:

$$\min\{\|x\|_1 : Ax = b\},$$

based on which various computational methods for sparse solutions of linear systems have been proposed ([11], [10], [26], [39], [4], [37], [41]). We note that the ℓ_1 type minimization is a long-lasting research topic in the field of numerical analysis and optimization ([1], [2], [3], [29]). However, it has a great impact when it was first introduced by Chen, Donoho and Saunders [11] in 1998 to attack ℓ_0-minimization problems arising from signal and imaging processing.

A large amount of empirical results ([11], [9], [10], [4], [23], [20], [41]) have shown that, in many situations, ℓ_1-minimization and its variants can locate the sparsest solution to underdetermined linear systems. As a result, it is important to rigorously address the fundamental question: When does ℓ_1-minimization solve ℓ_0-minimization? This question motivates one to identify the conditions for the 'equivalence' of these two problems. Let us first clarify what we mean by 'ℓ_0- and ℓ_1-minimization problems are equivalent' in this paper.

Definition 1.1. (i) ℓ_0- and ℓ_1-minimization problems are said to be equivalent if there exists a solution to ℓ_0-minimization that coincides with the unique solution to ℓ_1-minimization. (ii) ℓ_0- and ℓ_1-minimization problems are said to be strongly equivalent if the unique solution to ℓ_0-minimization coincides with the unique solution to ℓ_1-minimization.

Thus the *equivalence* does not require that ℓ_0-minimization have a unique solution. In fact, an underdetermined linear system may have multiple sparsest solutions. Clearly, the *strong equivalence* implies the *equivalence*, but the converse is not true. Currently, the understanding of the relationship between ℓ_0- and ℓ_1-minimization is mainly focused on the strong equivalence ([21], [17], [9], [7], [8], [15], [38], [10], [4], [33], [27], [40]). The study of the strong equivalence is motivated by the newly developed compressed-sensing theory ([5], [16]) which in turn stimulates the development of various sufficient criteria for the strong equivalence between ℓ_0- and ℓ_1-minimization.

The first work on the strong equivalence ([21], [17], [22], [25], [24], [35], [28]) claims that if a solution to the system $Ax = b$ satisfies that $\|x\|_0 < \frac{1}{2}(1 + \frac{1}{\mu(A)})$, where $\mu(A)$ is the mutual coherence of A, then such a solution is unique to both ℓ_0- and ℓ_1-minimization. The mutual-coherence-based analysis are very conservative. To go beyond this analysis, the restricted isometry property (RIP) ([9], [5]), null space property (NSP) ([12], [40]), ERC condition ([35], [24]), and other conditions ([24], [40], [27]) were introduced over the past few years. The RIP and NSP have been extensively investigated, and fruitful results on the strong equivalence between ℓ_0- and ℓ_1-minimization in both noisy and noiseless data cases have been developed in the literature. For instance, in noiseless cases, it

has been shown by E. Candès that the RIP of order $2k$ implies that ℓ_0- and ℓ_1-minimization are strongly equivalent whenever the linear system has a solution satisfying $\|x\|_0 \leq k$. The same conclusion holds if certain NSP is satisfied ([40], [12], [13]). However, RIP- and NSP-type conditions remain restrictive, compared to what the simulation has actually shown ([10], [20]). Candès and Romberg [6] (see also [7], [8]) initiated a probabilistic analysis for the efficiency of the ℓ_1-method. Following their work, various asymptotic and probabilistic analyses of ℓ_1-minimization have been extensively carried out by many researchers (e.g., [15], [18], [9], [20], [31]). The probabilistic analysis does demonstrate that ℓ_1-minimization has more capability of finding sparse solutions of linear systems than what is indicated by state-of-the-art strong-equivalence sufficient criteria.

In this paper, we introduce the so-called range space property (RSP), a new matrix property governing the equivalence of ℓ_0- and ℓ_1-minimization and the uniform recovery of sparse signals. The RSP-based theory is motivated very naturally by the needs of a further development of the theory for sparse signal recovery and practical applications.

First, we note that the existing sufficient criteria for the strong-equivalence of ℓ_0- and ℓ_1-minimization are very restrictive. These criteria cannot sufficiently interpret the actual numerical performance of the ℓ_1-method in many situations. They can only apply to a class of linear systems with unique sparsest and unique least ℓ_1-norm solutions. In practice, the signal to recover may not be sparse enough, and the linear systems arising from applications may have multiple sparsest solutions. The existing strong-equivalence-based theory fails in these situations. (See Examples 3.4 and 3.6 in this paper.) For instance, when the matrix A is constructed by concatenating several matrices, the linear system often has multiple sparsest solutions. In this case, we are still interested in finding a sparsest representation of the signal b in order for an efficient compression, storing and transmission of the signal. Our purpose in this case is to find one sparsest solution of the underlying linear system. The *equivalence* of ℓ_0- and ℓ_1-minimization can guarantee to achieve this goal. The RSP turns out to be an appropriate angle to approach the equivalence of ℓ_0- and ℓ_1-minimization.

Second, from a mathematical point of view, there are sufficient reasons to develop the RSP-based theory. The mutual coherence theory was developed from the Gram matrix $A^T A$ where A has normalized columns, the RIP of order k is a property of the submatrices $A_S^T A_S$ with $|S| \leq k$, and the NSP is a property of the null space $\mathcal{N}(A)$. The RSP (introduced in this paper) is a property of the range space $\mathcal{R}(A^T)$. Since ℓ_1-minimization is a linear programming problem, the optimality and/or the dual theory for this problem will unavoidably lead to the RSP-based theory.

The RSP-based theory goes beyond the existing theory to guarantee not only the strong equivalence but also the equivalence between ℓ_0- and ℓ_1-minimization. It applies to a broader class of linear systems than the existing theory, and it enables us to address the following questions: How to interpret the numerical performance of ℓ_1-minimization more efficiently than the current theory? If a linear system has multiple sparsest

solutions, when does ℓ_1-minimization guarantee to find one of them? How to deterministically interpret the efficiency and limit of ℓ_1-minimization for locating the sparsest solution to linear systems? Can we further develop a theory for sparse signal recovery by using certain new matrix property rather than existing ones?

To address these questions, our initial goal is to completely characterize the uniqueness of least ℓ_1-norm solutions to a linear system. It is the classic strict complementarity theorem of linear programming that enables us to achieve this goal. Theorem 2.10 developed in this paper claims that x *is the unique least ℓ_1-norm solution to the linear system $Ax = b$ if and only if the so-called range space property (RSP) and a full-rank property hold at x.*

Many questions associated with the ℓ_1-method, including the above-mentioned ones, can be largely addressed from this theorem, and a new theory for sparse signal recovery can be developed as well (see sections 3 and 4 for details). For instance, the equivalence of ℓ_0- and ℓ_1-minimization can be immediately obtained through this result. Our Theorem 3.3 in this paper claims that *a sparsest solution to a linear system is the unique least ℓ_1-norm solution to this system if and only if it satisfies the range space property (irrespective of the multiplicity of sparsest solutions)*. We note that the 'if' part of this result (i.e., the sufficient condition) was actually obtained by Fuchs [24]. The 'only if' part (i.e., the necessary condition) is shown in the present paper. It is worth mentioning that Donoho [14] has characterized the (strong) equivalence of ℓ_0- and ℓ_1-minimization from a geometric (topological) point of view, and Dossal [19] has shown that Donoho's geometric result can be characterized by extending Fuchs' analysis.

The RSP-based analysis in this paper shows that the uniqueness of sparsest solutions to linear systems is not necessary for the equivalence of ℓ_0- and ℓ_1-minimization, and the multiplicity of sparsest solutions may not prohibit the equivalence of these two problems as well. The RSP is the only condition that determines whether or not a sparsest solution to a linear system has a guaranteed recovery by ℓ_1-minimization. The RSP-based analysis can also explain the numerical behavior of ℓ_1-minimization more efficiently than the strong-equivalence-based analysis.

Moreover, we establish several new recovery theorems, based on such matrix properties as the RSP and Weak-RSP of order K. Theorem 4.2 established in this paper states that *any K-sparse signal can be exactly recovered by ℓ_1-minimization if and only if A^T has the RSP of order K*. Thus the RSP of order K can completely characterize the uniform recovery of K-sparse signals. The Weak-RSP-based recovery can be viewed as an extension of the uniform recovery. The key feature of this extended recovery theory is that the uniqueness of sparsest solutions to $Az = y$ may not be required, where the vector y denotes the measurements.

This paper is organized as follows. A necessary and sufficient condition for the uniqueness of least ℓ_1-norm solutions to linear systems is identified in section 2. The RSP-based equivalence analysis for ℓ_0- and ℓ_1-minimization is given in section 3, and the RSP-based recovery theory is developed in section 4.

Notation. Let R^n be the n-dimensional Euclidean space, and R^n_+ the first orthant of R^n. For $x, y \in R^n$, $x \leq y$ means $x_i \leq y_i$ for every $i = 1, ..., n$. Given a set $J \subseteq \{1, 2, ..., n\}$, the symbol $|J|$ denotes the cardinality of J, and $J_c = \{1, 2, ..., n\} \backslash J$ is the complement of J. For $x \in R^n$, $\text{Supp}(x) = \{i : x_i \neq 0\}$ denotes the support of x, $\|x\|_1 = \sum_{i=1}^{n} |x_j|$ denotes the ℓ_1-norm of x, and $|x| = (|x_1|, ..., |x_n|)^T \in R^n$ stands for the absolute vector of x. Given a matrix $A = (a_1, ..., a_n)$ where a_i denotes ith column of the matrix, we use A_S to denote a submatrix of A, with columns a_i, $i \in S \subseteq \{1, 2, ..., n\}$, and x_S stands for the subvector of $x \in R^n$ with components x_i, $i \in S$. Throughout the paper, $e = (1, 1, ..., 1)^T \in R^n$ denotes the vector of ones.

II. UNIQUENESS OF LEAST ℓ_1-NORM SOLUTIONS

Let us first recall a classic theorem for linear programming (LP) problems. Consider the LP problem

$$(P) \qquad \min\{c^T x : \quad Qx = p, \quad x \geq 0\},$$

and its dual problem

$$(DP) \qquad \max\{p^T y : Q^T y + s = c, \ s \geq 0\},$$

where Q is a given $m \times n$ matrix, and $p \in R^m$ and $c \in R^n$ are two given vectors. Suppose that (P) and (DP) have finite optimal values. By strong duality (optimality), $(x^*, (y^*, s^*))$ is a solution pair to the linear programming problems (P) and (DP) if and only if it satisfies the conditions

$$Qx^* = p, \ \ x^* \geq 0, \ Q^T y^* + s^* = c, \ s^* \geq 0, \ \ c^T x^* = p^T y^*,$$

where $c^T x^* = p^T y^*$ can be equivalently written as $x_i^* s_i^* = 0$ for every $i = 1, ..., n$, which is called the complementary slackness property. Moreover, the result below claims that when (P) and (DP) are feasible, there always exists a solution pair $(x^*, (y^*, s^*))$ satisfying $x^* + s^* > 0$, which is called a strictly complementary solution pair.

Lemma 2.1 ([32]). *Let (P) and (DP) be feasible. Then there exists a pair of strictly complementary solutions to (P) and (DP).*

We now develop some necessary conditions for x to be the unique least ℓ_1-norm solution to the system $Ax = b$.

A. Necessary range property of A^T

Throughout this section, let x be a given solution to the system $Ax = b$. Note that for any other solution z to this system, we have that $A(z - x) = 0$. So any solution z can be represented as $z = x + u$ where $u \in \mathcal{N}(A)$, the null space of A. Thus we consider the following two sets:

$$C = \{u : \|u + x\|_1 \leq \|x\|_1\}, \quad \mathcal{N}(A) = \{u : Au = 0\}.$$

Clearly, C depends on (A, b, x). Since (A, b, x) is assumed to be given in this section, we use C, instead of $C(A, b, x)$, for simplicity

First, we have the following straightforward observation, to which a simple proof is outlined in Appendix.

Lemma 2.2. *The following three statements are equivalent:*
(i) x is the unique least ℓ_1-norm solution to the system $Ax = b$.

(ii) $C \bigcap \mathcal{N}(A) = \{0\}$.

(iii) $(u, t) = (0, |x|)$ *is the unique solution to the system*

$$Au = 0, \ \sum_{i=1}^{n} t_i \leq \|x\|_1, \ |u_i + x_i| \leq t_i, \ i = 1, ..., n. \quad (1)$$

It is evident that $(u, t) = (0, |x|)$ is the unique solution to (1) if and only if it is the unique solution to the LP problem

$$
\begin{aligned}
(LP_1) \qquad \min \quad & 0^T t \\
\text{s.t.} \quad & Au = 0, \\
& \sum_{i=1}^{n} t_i \leq \|x\|_1, \\
& |u_i + x_i| \leq t_i, \ i = 1, ..., n,
\end{aligned}
$$

where u and t are variables. By introducing slack variables $\alpha, \beta \in R^n_+$ and $r \in R_+$, we can further write (LP_1) as

$$
\begin{aligned}
(LP_2) \qquad \min \quad & 0^T t \\
\text{s.t.} \quad & u_i + x_i - t_i + \alpha_i = 0, \ i = 1, ..., n, \quad (2) \\
& -u_i - x_i - t_i + \beta_i = 0, \ i = 1, ..., n, \quad (3) \\
& r + \sum_{i=1}^{n} t_i = \|x\|_1, \\
& Au = 0, \\
& \alpha \in R^n_+, \ \beta \in R^n_+, \ r \geq 0,
\end{aligned}
$$

which is always feasible, since $u = 0, t = |x|, \alpha = |x| - x, \beta = |x| + x$ and $r = 0$ satisfy all constraints. Almost all variables of (LP_2) are nonnegative except u. The nonnegativity of t follows from (2) and (3). For the convenience of analysis, we now transform (LP_2) into a form with all variables nonnegative. Note that u, satisfying (2) and (3), is bounded. In fact,

$$-2\|x\|_1 \leq -x_i - t_i \leq u_i \leq t_i - x_i \leq 2\|x\|_1, \ i = 1, ..., n.$$

Denote by $M := 2\|x\|_1 + 1$. Then $u' = Me - u \geq 0$ for any u satisfying (2) and (3). Thus by the substitution $u = Me - u'$, problem (LP_2) can be finally written as

$$
\begin{aligned}
(LP_3) \qquad \min \quad & 0^T t \\
\text{s.t.} \quad & (Me - u') - t + \alpha = -x, \\
& -(Me - u') - t + \beta = x, \\
& A(Me) - Au' = 0, \\
& e^T t + r = \|x\|_1, \\
& (u', t, \alpha, \beta, r) \in R^{4n+1}_+.
\end{aligned}
$$

That is,

$$
\begin{aligned}
\min \quad & 0^T t \\
\text{s.t.} \quad &
\end{aligned}
$$

$$
\begin{bmatrix} -I & -I & I & 0 & 0 \\ I & -I & 0 & I & 0 \\ -A & 0 & 0 & 0 & 0 \\ 0 & e^T & 0 & 0 & 1 \end{bmatrix}
\begin{bmatrix} u' \\ t \\ \alpha \\ \beta \\ r \end{bmatrix}
=
\begin{bmatrix} -x - Me \\ x + Me \\ -M(Ae) \\ \|x\|_1 \end{bmatrix},
$$

$$(u', t, \alpha, \beta, r) \geq 0.$$

From the above discussion, we have the next observation, to which the proof is given in Appendix.

Lemma 2.3. *The following three statements are equivalent:*

(i) $(u^*, t^*) = (0, |x|)$ *is the unique solution to* (LP_1).

(ii) $(u^*, t^*, \alpha^*, \beta^*, r^*) = (0, |x|, |x| - x, |x| + x, 0)$ *is the unique solution to* (LP_2).

(iii) $(u'^*, t^*, \alpha^*, \beta^*, r^*) = (Me, |x|, |x| - x, |x| + x, 0)$ *is the unique solution to* (LP_3).

This lemma shows that (LP_1), (LP_2) and (LP_3) are equivalent in the sense that if one of them has a unique solution, so do the other two. Their unique solutions are explicitly given in terms of x. Note that the dual problem of (LP_3) is given by

$$(DLP_3) \quad \max \quad -(x + Me)^T(y - y') - Me^T A^T y'' + \omega \|x\|_1,$$
$$\text{s.t.} \quad -(y - y') - A^T y'' \le 0, \tag{4}$$
$$-(y + y') + \omega e \le 0, \tag{5}$$
$$y \le 0, \tag{6}$$
$$y' \le 0, \tag{7}$$
$$\omega \le 0 \tag{8}$$

where y, y', y'' and ω are the dual variables. Let $s^{(1)}$, $s^{(2)}$, $s^{(3)}$, $s^{(4)} \in R_+^n$ and $s \in R_+^1$ denote the nonnegative slack variables associated with the constraints (4) through (8), respectively, i.e.,

$$\begin{cases} s^{(1)} = (y - y') + A^T y'', \\ s^{(2)} = (y + y') - \omega e, \\ s^{(3)} = -y, \\ s^{(4)} = -y', \\ s = -\omega. \end{cases} \tag{9}$$

We now prove that if x is the unique least ℓ_1-norm solution to the system $Ax = b$, then $\mathcal{R}(A^T)$, the range space of A^T, must satisfy certain property.

Theorem 2.4. *If x is the unique least ℓ_1-norm solution to the system $Ax = b$, then there exist $y, y' \in R^n$ and $\omega \in R$ satisfying*

$$\begin{cases} y - y' \in \mathcal{R}(A^T), \\ \omega < y_i + y'_i, \ y_i < 0, \ y'_i < 0 & \text{for } x_i = 0, \\ y_i = 0, \ y'_i = \omega & \text{for } x_i < 0, \\ y_i = \omega, \ y'_i = 0 & \text{for } x_i > 0. \end{cases} \tag{10}$$

Proof. Assume that x is the unique least ℓ_1-norm solution to the system $Ax = b$. By Lemmas 2.2 and 2.3, the problem (LP_3) has a unique solution given by

$$(u'^*, t^*, \alpha^*, \beta^*, r^*) = (Me, \ |x|, \ |x| - x, \ |x| + x, \ 0). \tag{11}$$

Since both (LP_3) and its dual problem (DLP_3) are feasible, by Lemma 2.1, there exists a strictly complementary solution pair to (LP_3) and (DLP_3), denoted by $((\bar{u}', \bar{t}, \bar{\alpha}, \bar{\beta}, \bar{r}), (y, y', y'', \omega))$. Since the solution to (LP_3) is unique, we must have

$$(\bar{u}', \bar{t}, \bar{\alpha}, \bar{\beta}, \bar{r}) = (u'^*, t^*, \alpha^*, \beta^*, r^*). \tag{12}$$

As defined by (9), we use $(s^{(1)}, s^{(2)}, s^{(3)}, s^{(4)}, s) \in R_+^{4n+1}$ to denote the slack variables associated with constraints (4) through (8) of (DLP_3). By strict complementarity, we have

$$(\bar{u}')^T s^{(1)} = 0, \ \bar{t}^T s^{(2)} = 0, \ \bar{\alpha}^T s^{(3)} = 0, \ \bar{\beta}^T s^{(4)} = 0, \bar{r}s = 0, \tag{13}$$

and

$$\bar{u}' + s^{(1)} > 0, \ \bar{t} + s^{(2)} > 0, \ \bar{\alpha} + s^{(3)} > 0, \bar{\beta} + s^{(4)} > 0, \bar{r} + s > 0. \tag{14}$$

First, we see that $s^{(1)} = 0$, since $\bar{u}' = u'^* = Me > 0$. By the definition of $s^{(1)}$, it implies that

$$A^T y'' = -(y - y'). \tag{15}$$

From (11), we see that

$$\begin{array}{llll} t_i^* = x_i > 0, & \alpha_i^* = 0, & \beta_i^* = 2x_i > 0 & \text{for } x_i > 0, \\ t_i^* = |x_i| > 0, & \alpha_i^* = 2|x_i| > 0, & \beta_i^* = 0 & \text{for } x_i < 0, \\ t_i^* = 0, & \alpha_i^* = 0, & \beta_i^* = 0 & \text{for } i \notin \text{Supp}(x), \\ r^* = 0. \end{array}$$

Therefore, it follows from (12), (13) and (14) that

$$\begin{array}{llll} s_i^{(2)} = 0, & s_i^{(3)} > 0, & s_i^{(4)} = 0 & \text{for } x_i > 0, \\ s_i^{(2)} = 0, & s_i^{(3)} = 0, & s_i^{(4)} > 0 & \text{for } x_i < 0, \\ s_i^{(2)} > 0, & s_i^{(3)} > 0, & s_i^{(4)} > 0 & \text{for } i \notin \text{Supp}(x), \\ s > 0. \end{array}$$

By the definition of these slack variables, the (strictly complementary) solution vector (y, y', y'', ω) of (DLP_3) satisfies (15) and

$$\begin{array}{llll} \omega - (y_i + y'_i) = 0, & y_i < 0, & y'_i = 0 & \text{for } x_i > 0, \\ \omega - (y_i + y'_i) = 0, & y_i = 0, & y'_i < 0 & \text{for } x_i < 0, \\ \omega - (y_i + y'_i) < 0, & y_i < 0, & y'_i < 0 & \text{for } i \notin \text{Supp}(x), \\ \omega < 0. \end{array}$$

Clearly, the condition '$\omega < 0$' is redundant in the above system, since it is implied from other conditions of the system. Thus we conclude that (y, y', y'', ω) satisfies (15) and the following properties:

$$\begin{array}{lll} \omega < y_i + y'_i, \ y_i < 0, \ y'_i < 0 & \text{for } x_i = 0, \\ y_i = 0, \ y'_i = \omega & \text{for } x_i < 0, \\ y_i = \omega, \ y'_i = 0 & \text{for } x_i > 0, \end{array}$$

which is exactly the condition (10), by noting that (15) is equivalent to $y - y' \in \mathcal{R}(A^T)$. $\square$

Therefore, (10) is a necessary condition for x to be the unique least ℓ_1-norm solution to the system $Ax = b$. This condition arises naturally from the strict complementarity of LP problems. We now further point out that (10) can be restated more concisely. The proof of this fact is given in Appendix.

Lemma 2.5. *Let $x \in R^n$ be a given vector. There exists a vector $(y, y', \omega) \in R^{2n+1}$ satisfying (10) if and only if there exists a vector $\eta \in \mathcal{R}(A^T)$ satisfying that $\eta_i = 1$ for all $x_i > 0$, $\eta_i = -1$ for all $x_i < 0$, and $|\eta_i| < 1$ for all $x_i = 0$.*

Note that when $\eta \in \mathcal{R}(A^T)$, both $-\eta$ and $\gamma\eta$ are also in $\mathcal{R}(A^T)$, where γ is any real number. Thus the following three conditions are equivalent: (i) There is an $\eta \in \mathcal{R}(A^T)$ with $\eta_i = 1$ for $x_i > 0$, $\eta_i = -1$ for $x_i < 0$, and $|\eta_i| < 1$ for $x_i = 0$. (ii) There is an $\eta \in \mathcal{R}(A^T)$ with $\eta_i = -1$ for $x_i > 0$, $\eta_i = 1$ for $x_i < 0$, and $|\eta_i| < 1$ for $x_i = 0$. (iii) There is an $\eta \in \mathcal{R}(A^T)$ with $\eta_i = \gamma$ for $x_i > 0$, $\eta_i = -\gamma$ for $x_i < 0$, and $|\eta_i| < |\gamma|$ for $x_i = 0$. The key feature here is that $\eta \in \mathcal{R}(A^T)$ has equal components (with value $\gamma \ne 0$) corresponding to positive components of x, and has equal components (with

value $-\gamma$) corresponding to all negative components of x, and absolute values of other components of η are strictly less that $|\gamma|$. So if a linear system has a unique least ℓ_1-norm solution, the range space of A^T must satisfy the above-mentioned 'nice' property.

B. Necessary full-rank condition

In order to completely characterize the uniqueness of least ℓ_1-norm solutions to the system $Ax = b$, we need to establish another necessary condition. Let x be a solution to the system $Ax = b$, and denote by $J_+ = \{i : x_i > 0\}$ and $J_- = \{i : x_i < 0\}$. We have the following lemma.

Lemma 2.6. *The matrix*

$$H = \begin{pmatrix} A_{J_+} & A_{J_-} \\ -e_{J_+}^T & e_{J_-}^T \end{pmatrix} \quad (16)$$

has full column rank if and only if the matrix below has full column rank

$$G = \begin{pmatrix} -I_{|J_+|} & 0 & -I_{|J_+|} & 0 \\ 0 & I_{|J_-|} & 0 & -I_{|J_-|} \\ A_{J_+} & A_{J_-} & 0 & 0 \\ 0 & 0 & e_{J_+}^T & e_{J_-}^T \end{pmatrix}, \quad (17)$$

where 0's are zero submatrices with suitable sizes, and $I_{|J|}$ denotes the $|J| \times |J|$ identity matrix.

Proof. Adding the first $|J_+| + |J_-|$ rows of G into its last row yields the following matrix G'. Following that, by similar column operations, G' can be further reduced to matrix G'':

$$G' = \begin{bmatrix} -I_{|J_+|} & 0 & -I_{|J_+|} & 0 \\ 0 & I_{|J_-|} & 0 & -I_{|J_-|} \\ A_{J_+} & A_{J_-} & 0 & 0 \\ -e_{J_+}^T & e_{J_-}^T & 0 & 0 \end{bmatrix}$$

$$\rightarrow G'' = \begin{bmatrix} 0 & 0 & -I_{|J_+|} & 0 \\ 0 & 0 & 0 & -I_{|J_-|} \\ A_{J_+} & A_{J_-} & 0 & 0 \\ -e_{J_+}^T & e_{J_-}^T & 0 & 0 \end{bmatrix}.$$

Note that upper-right block of G'' is a nonsingular square matrix, and the lower-left block is H. Since any elementary row and column operations do not change the (column) rank of a matrix. Thus H has full column rank if and only if G has full column rank. $\square$

We now prove the next necessary condition for the uniqueness of least ℓ_1-norm solutions to the system $Ax = b$.

Theorem 2.7. *If x is the unique least ℓ_1-norm solution to the system $Ax = b$, then the matrix H, defined by (16), has full column rank.*

Proof. Assume the contrary that the columns of H are linearly dependent. Then by Lemma 2.6, the columns of matrix G, given by (17), are also linearly dependent. Hence, there exists a vector $d = (d_1, d_2, d_3, d_4) \neq 0$, where $d_1, d_3 \in R^{|J_+|}$ and $d_2, d_4 \in R^{|J_-|}$, such that

$$Gd = \begin{pmatrix} -I_{|J_+|} & 0 & -I_{|J_+|} & 0 \\ 0 & I_{|J_-|} & 0 & -I_{|J_-|} \\ A_{J_+} & A_{J_-} & 0 & 0 \\ 0 & 0 & e_{J_+}^T & e_{J_-}^T \end{pmatrix} \begin{pmatrix} d_1 \\ d_2 \\ d_3 \\ d_4 \end{pmatrix} = 0.$$

Note that $z = (z_1, z_2, z_3, z_4)$, where

$$\begin{cases} z_1 = Me_{J_+} > 0, \\ z_2 = Me_{J_-} > 0, \\ z_3 = x_{J_+} > 0, \\ z_4 = -x_{J_-} > 0, \end{cases} \quad (18)$$

is a solution to the system

$$\begin{pmatrix} -I_{|J_+|} & 0 & -I_{|J_+|} & 0 \\ 0 & I_{|J_-|} & 0 & -I_{|J_-|} \\ A_{J_+} & A_{J_-} & 0 & 0 \\ 0 & 0 & e_{J_+}^T & e_{J_-}^T \end{pmatrix} \begin{pmatrix} z_1 \\ z_2 \\ z_3 \\ z_4 \end{pmatrix}$$
$$= \begin{pmatrix} -x_{J_+} - Me_{J_+} \\ x_{J_-} + Me_{J_-} \\ M(A_{J_+}e_{J_+} + A_{J_-}e_{J_-}) \\ \|x_{J_+}\|_1 + \|x_{J_-}\|_1 \end{pmatrix}. \quad (19)$$

Since $z > 0$, there exists a small number $\lambda \neq 0$ such that $\widetilde{z} = (\widetilde{z}_1, \widetilde{z}_2, \widetilde{z}_3, \widetilde{z}_4) = z + \lambda d \geq 0$ is also a (nonnegative) solution to the system (19), where $\widetilde{z}_i = z_i + \lambda d_i, i = 1, ..., 4$. Clearly, $\widetilde{z} \neq z$ since $\lambda \neq 0$ and $d \neq 0$. We now construct two solutions to (LP_3) as follows. Let

$$u' = \begin{bmatrix} z_1 \\ z_2 \\ Me_{J_0} \end{bmatrix} \in R^n, t = \begin{bmatrix} z_3 \\ z_4 \\ 0 \end{bmatrix} \in R^n,$$

$$\alpha = \begin{bmatrix} 0 \\ 2z_4 \\ 0 \end{bmatrix} \in R^n, \beta = \begin{bmatrix} 2z_3 \\ 0 \\ 0 \end{bmatrix} \in R^n,$$

and $r = 0$, where z_1, z_2, z_3 and z_4 are given by (18), and $J_0 = \{i : x_i = 0\}$. It is not difficult to see that the vector $(u', t, \alpha, \beta, r)$ defined above is exactly the one

$$(u' = Me, \ t = |x|, \ \alpha = |x| - x, \ \beta = |x| + x, \ r = 0).$$

On the other hand, we can also define

$$\widetilde{u}' = \begin{bmatrix} \widetilde{z}_1 \\ \widetilde{z}_2 \\ Me_{J_0} \end{bmatrix} \in R^n, \widetilde{t} = \begin{bmatrix} \widetilde{z}_3 \\ \widetilde{z}_4 \\ 0 \end{bmatrix} \in R^n,$$

$$\widetilde{\alpha} = \begin{bmatrix} 0 \\ 2\widetilde{z}_4 \\ 0 \end{bmatrix} \in R^n, \widetilde{\beta} = \begin{bmatrix} 2\widetilde{z}_3 \\ 0 \\ 0 \end{bmatrix} \in R^n,$$

and $\widetilde{r} = 0$. Clearly, $(\widetilde{u}', \widetilde{t}, \widetilde{\alpha}, \widetilde{\beta}, \widetilde{r})$ is in R_+^{4n+1}, and by a straightforward verification, we can show that this vector satisfies all constraints of (LP_3). So $(\widetilde{u}', \widetilde{t}, \widetilde{\alpha}, \widetilde{\beta}, \widetilde{r})$ is also a solution to (LP_3). This implies that (LP_3) has two different solutions: $(u', t, \alpha, \beta, r) \neq (\widetilde{u}', \widetilde{t}, \widetilde{\alpha}, \widetilde{\beta}, \widetilde{r})$. Under the assumption of the theorem, however, it follows from Lemmas 2.2 and 2.3 that (LP_3) has a unique solution. This contradiction shows that H, defined by (16), must have full column rank. $\square$

Combining Lemma 2.5, Theorems 2.4 and 2.7 yields the next result.

Theorem 2.8. *If x is the unique least ℓ_1-norm solution to the system $Ax = b$, then (i) the matrix $\begin{pmatrix} A_{J_+} & A_{J_-} \\ -e_{J_+}^T & e_{J_-}^T \end{pmatrix}$ has full column rank, and (ii) there exists a vector η such that*

$$\begin{cases} \eta & \in & \mathcal{R}(A^T), \\ \eta_i & = & 1 & \text{for all} & x_i > 0, \\ \eta_i & = & -1 & \text{for all} & x_i < 0, \\ |\eta_i| & < & 1 & \text{for all} & x_i = 0. \end{cases} \quad (20)$$

In this paper, the condition (ii) above is called the *range space property* (RSP) of A^T at x. It is worth noting that

checking the RSP of A^T at x is very easy. It is equivalent to solving the LP problem

$$\begin{aligned} \min \quad & \tau \\ \text{s.t.} \quad & A_{J_+}^T y = e_{J_+}, \\ & A_{J_-}^T y = -e_{J_-}, \\ & A_{J_0}^T y = \eta_{J_0}, \ |\eta_{J_0}| \le \tau e_{J_0}, \end{aligned}$$

where $J_0 = \{i : x_i = 0\}$. Clearly, the RSP of A^T at x holds if and only if the optimal value of the LP problem above satisfies $\tau < 1$.

C. A necessary and sufficient condition

Clearly, if (20) holds, there is a vector u such that $\begin{pmatrix} e_{J_+} \\ -e_{J_-} \end{pmatrix} = \begin{pmatrix} A_{J_+}^T \\ A_{J_-}^T \end{pmatrix} u$, and thus under condition (20), we have $\text{rank}\begin{pmatrix} A_{J_+} & A_{J_-} \\ -e_{J_+}^T & e_{J_-}^T \end{pmatrix} = \text{rank}(A_{J_+} \ A_{J_-})$. Thus condition (i) in Theorem 2.8 can be simplified to that the matrix $(A_{J_+} \ A_{J_-})$ (i.e., $A_{Supp(x)}$) has full column rank.

As we have seen from the above analysis, it is the strict complementarity theory of linear programming that leads to the necessary conditions in Theorem 2.8 which is established for the first time in this paper. The strict complementarity theory can be also used to prove that the converse of Theorem 2.8 is true, i.e., (i) together with (ii) in Theorem 2.8 is also sufficient for ℓ_1-minimization to have a unique solution. However, we omit this proof since this sufficiency was obtained already by Fuchs [24], while his analysis was based on convex quadratic optimization, instead of strict complementarity.

Theorem 2.9 (Fuchs [24]). *If the system $Ax = b$ has a solution x satisfying (20) and the columns of A corresponding to the support of x are linearly independent, then x is the unique least ℓ_1-norm solution to the system $Ax = b$.*

The necessary condition (Theorem 2.8) developed in this paper is the key to interpret the efficiency and limit of ℓ_1-minimization in locating/recovering sparse signals (see the discussion in the next section). Each of Theorems 2.8 and 2.9 alone can only give a half picture of the uniqueness of the least ℓ_1-norm solutions to a linear system. Theorems 2.8 and 2.9 together yield the following complete characterization.

Theorem 2.10. *x is the unique least ℓ_1-norm solution to the system $Ax = b$ if and only if the RSP (20) holds at x, and the matrix $(A_{J_+} \ A_{J_-})$ has full column rank, where $J_+ = \{i : x_i > 0\}$ and $J_- = \{i : x_i < 0\}$.*

This necessary and sufficient condition provides a good basis to interpret the relationship of ℓ_0- and ℓ_1-minimization, the internal mechanism of the ℓ_1-method, and their roles in compressed sensing.

III. RSP-BASED ANALYSIS FOR THE EQUIVALENCE OF ℓ_0- AND ℓ_1-MINIMIZATION

In this section, we focus on the condition for the equivalence of ℓ_0- and ℓ_1-minimization. Through this and the next sections, we will see how Theorem 2.10 can be used to interpret the efficiency and limit of the ℓ_1-method in finding/recovering sparse solutions of linear systems. For the convenience of discussion, we clarify the concept below.

Definition 3.1. *A solution x to the system $Ax = b$ is said to have a guaranteed recovery (or to be exactly recovered) by the ℓ_1-method if x is the unique least ℓ_1-norm solution to this system.*

Note that when matrix $(A_{J_+} \ A_{J_-})$ has full column rank, the number of its columns (i.e., $|J_+| + |J_-| = \|x\|_0$) is less than or equal to the number of its rows. Therefore, the following fact is implied from Theorem 2.10.

Corollary 3.2. *If x is the unique least ℓ_1-norm solution to the system $Ax = b$, where $A \in R^{m \times n}$ with $m < n$, then $\|x\|_0 \le m$.*

This result shows that when ℓ_1-minimization has a unique solution, this solution must be at least m-sparse. In other words, any $x \in R^n$ that can be exactly recovered by ℓ_1-minimization must be a sparse vector with $\|x\|_0 \le m$. This property of the ℓ_1-method justifies its role as a sparsity-seeking method. Corollary 3.2 also implies that any x with sparsity $\|x\|_0 > m$ is definitely not the unique least ℓ_1-norm solution to a linear system, and hence there is no guaranteed recovery for such a solution by ℓ_1-minimization. Note that Gaussian elimination or other linear algebra methods can also easily find a solution with $\|x\|_0 \le m$, although there is no guarantee for $\|x\|_0 < m$. Thus what we really want from the ℓ_1-method is a truly sparse solution with $\|x\|_0 < m$ if such a solution exists. So it is natural to ask when the ℓ_1-method finds a sparsest solution. By Theorem 2.10, we have the following result that completely characterizes the equivalence of ℓ_0- and ℓ_1-minimization.

Theorem 3.3. *Let $x \in R^n$ be a sparsest solution to the system $Ax = b$. Then x is the unique ℓ_1-norm solution to this system if and only if the range space property defined by (20) holds at x.*

Proof. On one hand, Theorem 2.10 claims that when a solution x is the unique least ℓ_1-norm solution, the RSP (20) must be satisfied at this solution. On the other hand, when x is the sparsest solution to $Ax = b$, the column vectors of A corresponding to the support of x must be linearly independent (i.e., $A_{Supp(x)}$ has full column rank), since otherwise at least one of the columns of $A_{Supp(x)}$ can be represented by other columns, and hence a solution sparser than x can be found, leading to a contradiction. So the matrix $(A_{J_+} \ A_{J_-})$ always has full column rank at any sparsest solution x of the system $Ax = b$. Thus by Theorem 2.10 again, to guarantee a sparsest solution to be the unique least ℓ_1-norm solution, the only condition required is the RSP. The desired result follows. $\square$

Although Theorem 3.3 is a special case of Theorem 2.10, it is powerful enough to encompass all existing sufficient conditions for the strong equivalence of ℓ_0- and ℓ_1-minimization as special cases, and it goes beyond the scope of these conditions. To show this, let us first decompose the underdetermined linear systems into three categories as follows:

- *Group 1*: The system has a *unique* least ℓ_1-norm solution and a *unique* sparsest solution.
- *Group 2*: The system has a *unique* least ℓ_1-norm solution and *multiple* sparsest solutions.
- *Group 3*: The system has *multiple* least ℓ_1-norm solutions.

Clearly, every linear system falls into one and only one of these groups. By Theorem 2.10, the linear systems in Group 3 do not satisfy either the RSP or full-rank property at any of its solutions. Thus the guaranteed recovery by ℓ_1-minimization can only possibly happen within Groups 1 and 2. Since many existing sufficient conditions for the equivalence of ℓ_0- and ℓ_1-minimization actually imply the strong equivalence between these two problems, these conditions can only apply to a subclass of linear systems in Group 1. The following three conditions are widely used in the literature:

- Mutual coherence condition ([17], [25], [21], [22]) $\|x\|_0 \leq \frac{1}{2}(1 + \frac{1}{\mu(A)})$, where $\mu(A)$ is defined by $\mu(A) = \max_{1 \leq i,j \leq m, i \neq j} |a_i^T a_j|/(\|a_i\|_2 \cdot \|a_j\|_2)$.
- Restricted isometry property (RIP) [9]. The matrix A has the restricted isometry property (RIP) of order k if there exists a constant $0 < \delta_k < 1$ such that $(1 - \delta_k)\|z\|_2^2 \leq \|Az\|_2^2 \leq (1 + \delta_k)\|z\|_2^2$ for all k-sparse vector z.
- Null space property (NSP) ([12], [13]). The matrix A has the NSP of order k if there exists a constant $\vartheta > 0$ such that $\|h_\Lambda\|_2 \leq \vartheta \frac{\|h_{\Lambda_c}\|_1}{\sqrt{k}}$ holds for all $h \in \mathcal{N}(A)$ and all $\Lambda \subseteq \{1, 2, ..., n\}$ such that $|\Lambda| \leq k$. The NSP can be defined in other ways (see e.g., [40]).

Some other important conditions can be also found in the literature, such as the accumulative coherence condition ([35], [17]), exact recovery coefficient (ERC) [35], and others (see e.g., [40], [27]). Theorem 3.3 shows that *all existing conditions for the equivalence between ℓ_0- and ℓ_1-minimization in the literature imply the RSP (20),* since the RSP is not only a sufficient, but also a necessary condition for the equivalence between ℓ_0- and ℓ_1-minimization problems.

A remarkable difference between the RSP and many existing sufficient conditions is that the RSP does not require the uniqueness of sparsest solutions. Even if a linear system has multiple sparsest solutions, ℓ_1-minimization can still guarantee to solve an ℓ_0-minimization problem, provided that the RSP holds at a sparsest solution of the problem.

Example 3.4. (RSP *does not require the uniqueness of sparsest solutions*). Consider the linear system $Ax = b$ with

$$A = \begin{pmatrix} 1 & 0 & -2 & 5 \\ 0 & 1 & 4 & -9 \\ 1 & 0 & -2 & 5 \end{pmatrix}, \text{ and } b = \begin{pmatrix} 1 \\ -1 \\ 1 \end{pmatrix}.$$

It is easy to see that the system $Ax = b$ has multiple sparsest solutions: $x^{(1)} = (1, -1, 0, 0)^T$, $x^{(2)} = (0, 1, -1/2, 0)^T$, $x^{(3)} = (0, 4/5, 0, 1/5)^T$, $x^{(4)} = (0, 0, 2, 1)^T$, $x^{(5)} = (1/2, 0, -1/4, 0)^T$ and $x^{(6)} = (4/9, 0, 0, 1/9)^T$. We now verify that the RSP holds at $x^{(6)}$. It is sufficient to find a vector $\eta = (1, \eta_2, \eta_3, 1)$ in the range space of A^T with $|\eta_2| < 1$ and $|\eta_3| < 1$. Indeed, by taking $u = (1, 4/9, 0)^T$, we have that $\eta = A^T u = (1, 4/9, -2/9, 1)^T$. Thus the RSP holds at $x^{(6)}$ which, by Theorem 3.3, has a guaranteed recovery by ℓ_1-minimization. So ℓ_0- and ℓ_1-minimization problems are equivalent. It is worth noting that the mutual coherence, RIP and NSP cannot apply to this example, since the system has multiple sparsest solutions.

It is easy to check that among all 6 sparsest solutions of the above example, $x^{(6)}$ is the only sparsest one satisfying the RSP. This is not a surprise, since by Theorem 3.3 any sparsest solution satisfying the RSP must be the unique least ℓ_1-norm solution. Thus we have the following corollary.

Corollary 3.5. *For any given underdetermined linear system, there exists at most one sparsest solution satisfying RSP (20).*

The next example shows that even for a problem in Group 1, many existing sufficient conditions may fail to confirm the strong equivalence of ℓ_1- and ℓ_0-minimization, but the RSP can successfully confirm this.

Example 3.6. Consider the system $Ax = b$ with

$$A = \begin{pmatrix} \frac{1}{\sqrt{2}} & 0 & \frac{1}{\sqrt{3}} & -\frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} & 0 \\ \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{3}} & -\frac{1}{\sqrt{3}} & 0 & -\frac{1}{\sqrt{2}} \\ 0 & \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{3}} & \frac{1}{\sqrt{6}} & -\frac{1}{\sqrt{2}} & -\frac{1}{\sqrt{2}} \end{pmatrix}, b = \begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix}$$

Clearly, $x^* = (0, 0, \sqrt{3}, 0, 0, 0)$ is the unique sparsest solution to this linear system. Note that $a_2^T a_6 = -1$ which implies that $\mu(A) = 1$. The mutual coherence condition $\|x^*\|_0 < \frac{1}{2}(1 + \frac{1}{\mu(A)}) = 1$ fails. Since the second and last columns are linearly dependent, the RIP of order 2 fails. Note that $\eta = (0, 1, 0, 0, 0, 1)^T \in \mathcal{N}(A)$ does not satisfy the null space property of order 2. So the NSP of order 2 also fails. However, we can find a vector $\eta = (\eta_1, \eta_2, 1, \eta_4, \eta_5, \eta_6)$ in $\mathcal{R}(A^T)$ with $|\eta_i| < 1$ for all $i \neq 3$. In fact, by simply taking $y = (\frac{1}{\sqrt{3}}, \frac{1}{\sqrt{3}}, \frac{1}{\sqrt{3}})$, we have

$$\eta = A^T y = \left(\sqrt{\frac{2}{3}}, \sqrt{\frac{2}{3}}, 1, \frac{1 - \sqrt{3} - \sqrt{2}}{3\sqrt{2}}, 0, -\sqrt{\frac{2}{3}} \right).$$

Thus the RSP (20) holds at x^*. By Theorem 3.3, ℓ_0- and ℓ_1-minimization are equivalent, and thus x^* has a guaranteed recovery by ℓ_1-minimization.

From the above discussion, we have actually shown, by Theorem 3.3, that the equivalence between ℓ_0- and ℓ_1-minimization can be achieved not only for a subclass of problems in Group 1, but also for a subclass of problems in large Group 2. Since many existing sufficient conditions imply the strong equivalence between ℓ_0- and ℓ_1-minimization which can be achieved only for a subclass of problems in Group 1, these conditions cannot apply to linear systems in Group 2, and hence cannot explain the success of ℓ_1-minimization for solving ℓ_0-minimization with multiple sparsest solutions. The simulation shows that the ℓ_1-method performs much better than what has predicted by those strong-equivalence-type conditions. Such a gap between the current theory and the actual performance of the ℓ_1-method can be clearly interpreted and identified now by our RSP-based analysis. This analysis indicates that the uniqueness of sparsest solutions is not necessary for an ℓ_0-minimization to be solved by the ℓ_1-method, and the multiplicity of sparsest solutions of a linear system does not prohibit the ℓ_1-method from solving an ℓ_0-minimization as well. When many existing sufficient conditions fails (as shown by Examples 3.4 and 3.6), the RSP-based analysis shows that the ℓ_1-method can continue its success in solving ℓ_0-minimization problems in many situations. Thus it does show that the actual success rate of ℓ_1-minimization for solving ℓ_0-minimization problems is certainly higher than what has indicated by the strong-equivalence-based theory. Moreover,

the RSP-based theory also sheds light on the limit of ℓ_1-minimization. Failing to satisfy the RSP, by Theorem 3.3 a sparsest solution is definitely not the unique least ℓ_1-norm solution, and hence there is no guaranteed recovery for such a solution by ℓ_1-minimization.

IV. Compressed sensing: RSP-based sparsity recovery

So far, the sparsity recovery theory has been developed by various approaches, including the ℓ_1-method (i.e., the so-called basis pursuit) and heuristic methods such as the (orthogonal) matching pursuit (e.g. [36], [20]). In this section, we show how Theorems 2.10 and 3.3 can be used to develop recovery criteria for sparse signals. Suppose that we would like to recover a sparse vector x^*. To serve this purpose, the so-called sensing matrix $A \in R^{m \times n}$ with $m < n$ is constructed, and the measurements $y = Ax^*$ are taken. Then we solve the ℓ_1-minimization problem $\min\{\|x\|_1 : Ax = y\}$ to obtain a solution $\hat{x}$. The compressed sensing theory is devoted to addressing, among others, the following questions: What class of sensing matrices can guarantee the exact recovery $\hat{x} = x^*$, and how sparse should x^* be in order to achieve the recovery success? To guarantee an exact recovery, A is constructed to satisfy the following three conditions:

(C1) $Ax = y$ has a unique least ℓ_1-norm solution $\hat{x}$.
(C2) x^* is the unique sparsest solution to the system $Ax = y$.
(C3) These two solutions are equal.

Clearly, satisfying (C1)–(C3) actually requires that ℓ_0- and ℓ_1-minimization are strongly equivalent. Many existing recovery theories comply with this framework. For instance, if A satisfies the RIP of order $2k$ with $\delta_{2k} < \sqrt{2} - 1$ (see [5]), or if A satisfies the NSP of order $2k$ (see [12], [13]), then there exists a constant ϑ such that for any $x \in R^n$, it holds

$$\|x - x^*\|_2 \leq \vartheta \sigma_k(x)/\sqrt{k}, \tag{21}$$

where $\sigma_k(x) = \min\{\|x - z\|_1 : \|z\|_0 \leq k\}$, the ℓ_1-norm of the $n - k$ smallest components of x. This result implies that if $Ax = y$ has a solution satisfying $\|x\|_0 \leq k$, it must be equal to x^*, and it is the unique sparsest solution to the system $Ax = y = Ax^*$.

A. RSP-based uniform recovery

Recall that the spark of a given matrix, denoted by spark(A), is the smallest number of columns of A that are linearly dependent [17]. The exact recovery of all k-sparse vectors (i.e., $\{x : \|x\|_0 \leq k\}$) by a single sensing matrix A is called *uniform recovery*. It is well known that the RIP and NSP of order $2k$ can uniformly recover k-sparse vectors, where $k < \frac{1}{2}$spark(A). We now characterize the uniform recovery by a new concept defined as follows.

Definition 4.1. (RSP of order K) *Let $A \in R^{m \times n}$ with $m < n$. The matrix A^T is said to satisfy the range space property of order K if for any disjoint subsets S_1, S_2 of $\{1, ..., n\}$ with $|S_1| + |S_2| \leq K$, the range space $\mathcal{R}(A^T)$ contains a vector η such that $\eta_i = 1$ for all $i \in S_1$, $\eta_i = -1$ for all $i \in S_2$, and $|\eta_i| < 1$ for all other components.*

This concept can be used to characterize the uniform recovery, as shown by the next result.

Theorem 4.2. (i) *If A^T has the RSP of order K, then any K columns of A are linearly independent.* (ii) *Assume that the measurements of the form $y = Ax$ are taken. Then any x with $\|x\|_0 \leq K$ can be exactly recovered by ℓ_1-minimization if and only if A^T has the RSP of order K.*

Proof. (i) Let $S \subseteq \{1, ..., n\}$ be any subset with $|S| = K$. We denote the elements of S by $\{s_1, ..., s_K\}$. We prove that the columns of A_S are linearly independent. It is sufficient to show that $z_S = 0$ is the only solution to $A_S z_S = 0$. In fact, assume $A_S z_S = 0$. Then $z = (z_S, z_{S_c} = 0) \in R^n$ is in the null space of A. Consider the disjoint sets $S_1 = S$, and $S_2 = \emptyset$. By the RSP of order K, there exists a vector $\eta \in \mathcal{R}(A^T)$ with $\eta_i = 1$ for all $i \in S_1 = S$. By the orthogonality of $\mathcal{N}(A)$ and $\mathcal{R}(A^T)$, we have $0 = z^T \eta = z_S^T \eta_S + z_{S_c}^T \eta_{S_c} = z_S^T \eta_S$, i.e.,

$$\sum_{j=1}^{K} z_{s_j} = 0. \tag{22}$$

Now we consider any k with $1 \leq k \leq K$, and the pair of disjoint sets:

$$S_1 = \{s_1, s_2, ..., s_k\}, \quad S_2 = \{s_{k+1}, ..., s_K\}.$$

By the RSP of order K, there exists an $\eta \in \mathcal{R}(A^T)$ with $\eta_{s_i} = 1$ for every $i = 1, ..., k$ and $\eta_{s_i} = -1$ for every $i = k+1, ..., K$. By orthogonality again, it follows from $z^T \eta = 0$ that

$$(z_{s_1} + \cdots + z_{s_k}) - (z_{s_{k+1}} + \cdots + z_K) = 0,$$

which holds for every k with $1 \leq k \leq K$. It follows from these relations, together with (22), that all components of z_S must be zero. This implies that any K columns of A are linearly independent.

(ii) First we assume that the RSP of order K is satisfied. Let x be an arbitrary vector with $\|x\|_0 \leq K$. Let $S_1 = J_+ = \{i : x_i > 0\}$ and $S_2 = J_- = \{i : x_i < 0\}$. Clearly, S_1 and S_2 are disjoint, and $|S_1| + |S_2| \leq K$, by the RSP of order K, there exists a vector $\eta \in \mathcal{R}(A^T)$ such that $\eta_i = 1$ for all $i \in S_1$, $\eta_i = -1$ for all $i \in S_2$, and $|\eta_i| < 1$ for all other components. This implies that the RSP (20) holds at x. Also, it follows from (i) that any K columns of A are linearly independent, and thus any $|S_1| + |S_2|$ ($\leq K$) columns of A are linearly independent. So the matrix $(A_{S_1} \ A_{S_2})$ has full column rank. By Theorem 3.3 (or 2.10), x is the unique least ℓ_1-norm solution to the equation $Az = y$. So x can be exactly recovered by ℓ_1-minimization.

Conversely, assume that any K-sparse vector can be exactly recovered by ℓ_1-minimization. We prove that A^T satisfies the RSP of order K. Indeed, let x be a K-sparse vector, and let y be the measurements, i.e., $y = Ax$. Under the assumption, x can be exactly recovered by the ℓ_1-method, so x is the unique least ℓ_1-norm solution to the system $Az = y$. By Theorem 2.8 (or 2.10), the RSP (20) holds at x. This implies that there exists $\eta \in \mathcal{R}(A^T)$ such that $\eta_i = 1$ for $i \in S_1$, $\eta_i = -1$ for $i \in S_2$, and $|\eta_i| < 1$ for all other components, where $S_1 = J_+ = \{i : x_i > 0\}$ and $S_2 = J_- = \{i : x_i < 0\}$. Since x can be any K-sparse vectors, the above property holds for

any disjoint subsets $S_1, S_2 \subseteq \{1, ..., n\}$ with $|S_1| + |S_2| \leq K$. Thus the RSP of order K holds. $\quad\square$

The theorem above shows that *the RSP of order K is a necessary and sufficient condition for the exact recovery of all K-sparse vectors*. Thus the RSP of order K completely characterizes the uniform recovery by ℓ_1-minimization. It is worth mentioning that Donoho [14] has characterized the exact recovery condition from a geometric perspective, i.e., by the so-called 'k-neighborly' property. Zhang [40] has characterized the uniform recovery by using null space property of A, and Juditsky and Nemirovski [27] have also proposed a condition for the uniform recovery based on their function $\gamma(A)$. The RSP of order K is an alternative characterization of the uniform recovery. It is worth stressing that although the RSP (20) at an individual sparsest solution does not imply the uniqueness of the sparsest solution (as shown in section 3), the RSP of order K is more restrictive than the individual RSP (20). In fact, as indicated by the next corollary, the RSP of order K complies with the recovery conditions (C1)–(C3).

Corollary 4.3. *Let $A \in R^{m \times n}$ with $m < n$. If A^T has the RSP of order K, then any $\widetilde{x} \in R^n$ with $\|\widetilde{x}\|_0 \leq K$ is both the unique least ℓ_1-norm solution and the unique sparsest solution to the system $Ax = y = A\widetilde{x}$.*

Proof. By Theorem 4.2, any $\widetilde{x}$ with $\|\widetilde{x}\|_0 \leq K$ can be exactly recovered by ℓ_1-minimization. Thus, $\widetilde{x}$ is the unique least ℓ_1-norm solution. We now assume that there exists another solution x' to $Ax = A\widetilde{x}$ with $\|x'\|_0 \leq \|\widetilde{x}\|_0 \leq K$. Let $S_1 = J'_+ = \{i : x'_i > 0\}$ and $S_2 = J'_- = \{i : x'_i < 0\}$. By the definition of the RSP of order K, we see that the RSP (20) holds at x'. Since any K columns of A are linearly independent (by Theorem 4.2), so are the columns of $(A_{J'_+} \; A_{J'_-})$. Thus, by Theorem 2.10, x' is the unique least ℓ_1-norm solution. Therefore, $x' = \widetilde{x}$, which shows that $\widetilde{x}$ is the unique sparsest solution to the system $Ax = A\widetilde{x}$. $\quad\square$

Does a sensing matrix with the RSP of certain order exist? The answer is obvious, as demonstrated by the next result, to which the proof is given in Appendix.

Lemma 4.4. *Suppose that one of the following holds:*

(i) $K \leq \left\lfloor \frac{1}{2}\left(1 + \frac{1}{\mu(A)}\right) \right\rfloor$.

(ii) *The matrix A has the RIP of order $2K$ with constant $\delta_{2K} < \sqrt{2} - 1$.*

(iii) *The matrix A has the NSP of order $2K$.*

Then the matrix A^T has the RSP of order K.

Therefore, the existence of RSP matrices is not an issue. In fact, any sufficient conditions for the uniform recovery (not just those listed in Lemma 4.4) imply the RSP of certain order.

B. Beyond the uniform recovery

Theorem 4.2 implies that it is impossible to uniformly recover all k-sparse vectors with $k \geq \text{spark}(A)$ by a single matrix A. In order to recover a k-sparse vector with high sparsity, for instance, $m > k > \frac{1}{2}\text{spark}(A)$, we should relax the recovery condition (C2) which, according to Theorem 2.10, is not a necessary condition for a vector to be exactly recovered by ℓ_1-minimization. Let us first relax this condition by dropping the uniqueness requirement of the sparsest solution

to the equation $Ax = y$, where y denotes the measurements. Then we have the following immediate result.

Proposition 4.5. *Let A satisfy the following property: For any k-sparse vector x^* with $Ax^* \neq 0$, x^* is a sparsest solution to the system $Ax = Ax^*$. Then $k < \text{spark}(A)$.*

Proof. Note that for any sparsest solution x to the system $Ax = b = Ax^* \neq 0$, the columns of $A_{Supp(x)}$ are linearly independent (since otherwise, a sparser solution than x can be found). Thus under the condition, we conclude that any k columns of A are linearly independent. So $k < \text{spark}(A)$. $\quad\square$

The proposition above shows that even if we relax condition (C2) by only requiring that x^* be a sparsest solution (not necessarily the unique sparsest solution) to the system $Ax = y = Ax^*$, $\text{spark}(A)$ is an unattainable upper bound for the uniform recovery. This fact was initially observed by Donoho and Elad [17], who had developed the mutual coherence condition to guarantee the recovery success by ℓ_1-minimization. Note that the uniform recovery by a matrix with RIP or NSP of order $2k$ can recover all k-sparse vectors with $k < \text{spark}(A)/2$. Thus, from a mathematical point of view, it is interesting to study how a sparse vector with $\|x\|_0 \geq \text{spark}(A)/2$ can be possibly recovered. This is also motivated by some practical applications, where an unknown signal might not be sparse enough to fall into the range $\|x\|_0 < \text{spark}(A)/2$. Hence the uniform recovery conditions (such as the RIP or NSP of order $2k$) does not apply to these situations. Theorem 2.10 makes it possible to handle such a situation by abandoning the recovering principle (C2). This theorem shows that any solution, satisfying the individual RSP (20) and full-rank property, has a guaranteed recovery by ℓ_1-minimization. To satisfy these conditions, the targeted signal does not have to be the sparsest solution to a linear system, as shown by the next example.

Example 4.6. Let

$$A = \begin{pmatrix} 6 & -4 & 3 & 4 & -2 \\ 6 & -4 & -1 & 4 & 0 \\ 0 & 2 & 3 & -1 & -3 \end{pmatrix}, \; y = \begin{pmatrix} 4 \\ 4 \\ -1 \end{pmatrix}.$$

It is easy to check that $x^* = (1/3, -1/2, 0, 0, 0)^T$ satisfies the RSP (20) and full-rank property. Thus, by Theorem 2.10, x^* is the unique least ℓ_1-norm solution to the system $Ax = y$. Thus x^* can be exactly recovered by ℓ_1-minimization although it is not the sparsest one. It is evident that $\widetilde{x} = (0, 0, 0, 1, 0)^T$ is the unique sparsest solution (with $\|\widetilde{x}\|_0 = 1$) for this linear system. (It is worth noting that $\widetilde{x}$ cannot be recovered since the RSP (20) does not hold at this solution.)

Therefore Theorem 2.10 makes it possible to develop an extended uniform recovery theory. Toward this goal, we introduce the following matrix property.

Definition 4.7 (Weak-RSP of order K). *Let $A \in R^{m \times n}$ with $m < n$. A^T is said to satisfy the weak range space property of order K if (i) there exists a pair of disjoint subsets $S_1, S_2 \subseteq \{1, ..., n\}$ such that $|S_1| + |S_2| = K$ and $(A_{S_1} \; A_{S_2})$ has full column rank, and (ii) for any disjoint $S_1, S_2 \subseteq \{1, ..., n\}$ such that $|S_1| + |S_2| \leq K$ and $(A_{S_1} \; A_{S_2})$ has full column rank, the space $\mathcal{R}(A^T)$ contains a vector η such that $\eta_i = 1$ for $i \in S_1$, $\eta_i = -1$ for $i \in S_2$, and $|\eta_i| < 1$ otherwise.*

The essential difference between this concept and the RSP of order K is that the RSP of order K requires that the individual RSP hold for any disjoint subsets S_1, S_2 of $\{1, ..., n\}$ with $|S_1| + |S_2| \leq K$, but the Weak-RSP of order K requires that the individual RSP hold only for those disjoint subsets S_1, S_2 satisfying that $|S_1| + |S_2| \leq K$ and $\begin{pmatrix} A_{S_1} & A_{S_2} \end{pmatrix}$ has full-column-rank. So the RSP of order K implies the Weak-RSP of K, but the converse is not true in general. Based on this concept, we have the next result that follows from Theorem 2.10 immediately.

Theorem 4.8. (i) *If A^T has the Weak-RSP of order K, then $K \leq m$.*

(ii) *Assume that the measurements of the form $y = Ax$ are taken. Then any x, with $\|x\|_0 \leq K$ and $\begin{pmatrix} A_{J_+} & A_{J_-} \end{pmatrix}$ being full-column-rank, can be exactly recovered by ℓ_1-minimization if and only if A^T has the Weak-RSP of order K.*

The bound $K \leq m$ above follows directly from the condition (i) of Definition 4.7. It is not difficult to see a remarkable difference between Theorems 4.8 and 4.2. Theorem 4.2 claims that all vectors with sparsity $\|x\|_0 \leq K$ can be exactly recovered via a sensing matrix with the RSP of order K, where $K < \text{spark}(A)$ which is an unattainable upper bound for any uniform recovery. Different from this result, Theorem 4.8 characterizes the exact recovery of a part (not all) of vectors within the range $1 \leq \|x\|_0 \leq m$. This result makes it possible to use a matrix with the Weak-RSP of order K, where $\text{spark}(A)/2 \leq K < m$, to exactly recover some sparse vectors in the range $\text{spark}(A)/2 \leq \|x\|_0 < K$, to which the current uniform-recovery theory is difficult to apply.

It is worth stressing that the Weak-RSP-based recovery has abandoned the uniqueness requirement (C2) of the sparsest solution to a linear system, and has built the recovery theory on condition (C1) only. Thus the guaranteed recovery can be naturally extended into the range $[\text{spark}(A)/2, m)$. Of course, only some vectors (signals) in this range can be exactly recovered, i.e., those vectors with the RSP (20) and the full-rank property. Both Theorems 4.2 and 4.8 have shed light on the limit of the recovering ability of ℓ_1-minimization.

Finally, we point out that although checking the RSP at a given point x is easy, checking the RIP, NSP, and RSP of certain order for a matrix is generally difficult. From a practical point of view, it is also important to develop some verifiable conditions (see e.g., [27], [34]).

V. Conclusions

The uniqueness of least ℓ_1-norm solutions to underdetermined linear systems plays a key role in solving ℓ_0-minimization problems and in sparse signal recovery. Combined with Fuchs' theorem, we have proved that a vector is the unique least ℓ_1-norm solution to a linear system if and only if the so-called range space property and full-rank property hold at this vector. This complete characterization provides immediate answers to several questions in this field. The main results in this paper were summarized in Theorems 2.10, 3.3, 4.2 and 4.8. These results have been developed naturally from the classic linear programming theory, and have been benefited by distinguishing between the equivalence and strong equivalence of ℓ_0- and ℓ_1-minimization. The RSP-based analysis

in this paper is useful to explore a broad equivalence between ℓ_0- and ℓ_1-minimization, and to further understand the internal mechanism and capability of the ℓ_1-method for solving ℓ_0-minimization problems. Moreover, we have introduced such new matrix properties as the RSP of order K and the Weak-RSP of order K. The former turns out to be one of the mildest conditions governing the uniform recovery, and the latter may yield an extended uniform recovery.

It is worth mentioning that the discussion in this paper was focused on the sparse signal recovery without noises. Some open questions are worthwhile to address in the future, such as how the RSP can be used to analyze the sparse signal recovery with noises, and how the RSP-based analysis can be possibly used to establish a lower bound for measurements in compressed sensing.

References

[1] R. Bartels, A. Conn, and J. Sinclair, Minimization techniques for piecewise differentiable functions: The ℓ_1 solution to an overdetermined linear system, *SIAM J. Numer. Anal.*, 15 (1978), pp. 224–241.
[2] P. Bradley, O.L. Mangasarian and J. Rosen, Parsimonious least norm approximation, *Comput. Optim. Appl.*, 11 (1998), pp. 5–21.
[3] P. Bradley, U. Fayyad and O.L. Mangasarian, Mathematical programming for data mining: formulations and challenges, *INFORMS J. Comput.*, 11 (1999), pp. 217–238.
[4] A. Bruckstein, D. Donoho and M. Elad, From sparse solutions of systems of equations to sparse modeling of signals and images, *SIAM Rev.*, 51 (2009), pp. 34–81.
[5] E. Candès, Compressive sampling, *International Congress of Mathematicians*, Vol. III, 2006, pp.1433–1452.
[6] E. Candès and J. Romberg, Practical signal recovery from random projections, in *Wavelet XI, Proc. SPIE Conf.* 5914, 2005.
[7] E. Candès, J. Romberg and T. Tao, Robust uncertainty principles: Exact signal reconstruction from highly impcomplete frequency information, *IEEE Trans. Inform. Theory*, 52 (2006), pp. 489–509.
[8] E. Candès, J. Romberg and T. Tao, Stable signal recovery from incomplete and inaccurate measurements, *Commun. Pure Appl. Math.,* 59 (2006), pp. 1207–1223.
[9] E. Candès and T. Tao, Decoding by linear programming, *IEEE Trans. Inform. Theory*, 51 (2005), pp. 4203–4215.
[10] E. Candès, M. Wakin and S. Boyd, Enhancing sparsity by reweighted ℓ_1 minimization, *J. Fourier Anal. Appl.*, 14 (2008), pp. 877–905.
[11] S. Chen, D. Donoho, and M. Saunders, Atomic decomposition by basis pursuit, *SIAM J. Sci. Comput.*, 20 (1998), pp. 33–61.
[12] A. Cohen, W. Dahmen and R. Devore, Compressed sensing and best k-term aproximation, *J. Amer. Math. Soc.*, 22 (2009), pp. 211–231.
[13] M. Davenport, M. Duarte, Y. Eldar and G. Kutyniok, Introduction to compressive sensing, Chapter in *Compressed Sensing: Theory and Applications*, Cambridge University Press, 2012, pp. 1-68.
[14] D. Donoho, Neighborly polytopes and sparse solutions of underdetermined linear equations, Techinical report 2005-04, Stanford Univ., 2005.
[15] D. Donoho, For most large underdetermined systems of linear equations, the minimal ℓ_1-norm solution is also the sparsest soluton, *Commun. Pure Appl. Math.,* 59 (2006), pp. 797–829.
[16] D. Donoho, Compressed sensing, *IEEE Trans. Inform. Theory*, 52 (2006), pp. 1289–1306.
[17] D. Donoho and M. Elad, Optimality sparse representation in general (non-orthogonal) dictionaries via ℓ_1 minimization, *Proc. Natl. Acad. Sci.*, 100 (2003), pp. 2197–2202.
[18] D. Donoho and J. Tanner, Sparse nonnegative solutions of underdetermined linear equations by linear programming, *Proc. Natl. Acad. Sci.*, 102 (2005), pp. 9446–9451.
[19] C. Dossal, A necessary and sufficient condition for exact sparse recovery by ℓ_1 minimization, Preprint Hal-00164738, 2007.
[20] M. Elad, *Sparse and Redundant Representations: From Theory to Applications in Signal and Image Processing*, Springer, New York, 2010.
[21] M. Elad and A. Bruckstein, A generalized uncertainty priciple and sparse representation in pairs of bases, *IEEE Trans. Inform. Theory*, 48 (2002), pp. 2558–2567.

[22] A. Feuer and A. Nemirovski, On sparse representation in pairs of bases, *IEEE Trans. Inform. Theory*, 49 (2003), pp. 1579–1581.

[23] S. Foucart and M. Lai, Sparsest solutions of underdetermined linear systems via ℓ_p-minimization for $0 < q \le 1$, *Appl Comput. Harmon. Anal.*, 26 (2009), pp. 395–407.

[24] J. Fuchs, On sparse representations in arbitrary redundant bases, *IEEE Trans. Inform. Theory,* 50 (2004), pp.1341–1344.

[25] R. Gribonval and M. Nielsen, Sparse decompositions in unions of bases, *IEEE Trans. Inform. Theory,* 49 (2003), pp. 3320–3325.

[26] E. Hale, W. Yin and Y. Zhang, Fixed-point continuation for ℓ_1-minimization: methodology and convergence, *SIAM J. Optim.*, 19 (2008), pp.1107–1130.

[27] A. Juditski and A. Nemirovski, On verifiable sufficient conditions for sparse signal recovery via ℓ_1 minimization, *Math. Program. Ser. B*, 127 (2011), pp. 57–88.

[28] D. Malioutov, M. Cetin and A. Willsky, Optimal sparse representation in general overcomplete bases, *IEEE Inter. Conf. Acoustic, Speech, & Signal Processing-ICASSP*, Montreal, Canada, May 2004.

[29] O.L. Mangasarian, Minimum-support solutions of polyhedral concave programs, *Optimization*, 45 (1999), pp. 149–162.

[30] B. Natarajan, Sparse approximate solutions to linear systems, *SIAM J. Comput.*, 24 (1995), pp. 227–234.

[31] G. Pfander, H. Rauhut, and J.A. Tropp, The restricted isometry property for time-frequency structured random matrices, Preprint, 2011.

[32] A. Schrijver, *Theory of Linear and Integer Programming*, Wiley, New York, 1986.

[33] Y. Sharon, J. Wright and Y. Ma, Computation and relaxation of conditions for equivalence between ℓ_1 and ℓ_0 minimization, Technical Report, 2008.

[34] G. Tang and A. Nehorai, Verfiable and computational performance analysis of sparsity recovery, Technical report, 2011.

[35] J. Tropp, Greed is good: Algorithmic results for sparse approximation, *IEEE Trans. Inform. Theory*, 50 (2004), pp. 2231–2242.

[36] J. Tropp and A. Gilbert, Signal recovery from random measurements via orthogonal mathcing pursuit, *IEEE Trans. Inform. Theory*, 53 (2007), pp. 4655–4666.

[37] J. Tropp and S. Wright, Computational methods for sparse solution of linear inverse problems, in *Proc. of the IEEE,* 98 (2010), pp. 948–958.

[38] Y. Tsaig and D. Donoho, Breakdown of equivalence between the minimal ℓ_1-norm solution and the sparsest solution, *Signal Process.,* 86 (2006), pp. 533–548.

[39] W. Yin, S. Osher, D. Goldfarb and J. Darbon, Bregman iterative algorithms for ℓ_1 minimization with applications to compressed sensing, *SIAM J. Imaging Sci.,* 1(2008), pp. 43–168.

[40] Y. Zhang, Theory of compressive sensing via ℓ_1-mimimization: A non-RIP analysis and extensions, *JORC*, 1 (2013), pp.79-105.

[41] Y.-B. Zhao and D. Li, Reweighted ℓ_1-minimization for sparse solutions to underdetermined linear systems, *SIAM J. Optim.,* 22 (2012), pp. 1065–1088.

Appendix

Proof of Lemma 2.2: The equivalence between (i) and (ii) is obvious. The equivalence between (ii) and (iii) can be easily verified as well. Suppose that (ii) holds. We now show that $(u,t) = (0,|x|)$ is the only solution to (1). In fact, let (u,t) be an arbitrary solution to (1). Since $|u_i + x_i| \le t_i$ for all $i = 1,...,n$, adding all these inequalities leads to $\|u + x\|_1 \le \sum_{i=1}^{n} t_i$ which, combined with the second inequality of (1), yields $\|u + x\|_1 \le \|x\|_1$. This implies that $u \in C \bigcap \mathcal{N}(A)$. Since $C \bigcap \mathcal{N}(A) = \{0\}$, we must have $u = 0$. Substituting $u = 0$ into (1), we see that t must be $|x|$. In other words, under (ii), $(u,t) = (0,|x|)$ is the only solution to (1). The converse can be verified easily as well.

Proof of Lemma 2.3: Note that slack variables of (LP_2) are uniquely determined by (u,t) as follows:

$$\alpha = t - x - u, \quad \beta = t + x + u, \quad r = \|x\|_1 - \sum_{i=1}^{n} t_i. \quad (23)$$

There is a one-to-one correspondence between feasible points of (LP_1) and (LP_2), i.e., (u,t) is feasible to (LP_1) if and only if (u,t,α,β,r), where (α,β,r) is given by (23), is feasible to (LP_2). Since both problems have zero objectives, any feasible point is optimal. Thus (i) and (ii) are equivalent. Also, there exists a one-to-one correspondence between feasible points of (LP_2) and (LP_3). In fact, it is evident that (u,t,α,β,r) is feasible to (LP_2) if and only if (u',t,α,β,r) is feasible to (LP_3), where $u' = Me - u \ge 0$ (the nonnegativity follows from the definition of M). Note that both problems have zero objectives. Thus (ii) and (iii) are also equivalent.

Proof of Lemma 2.5: First, we assume that (y,y',ω) satisfies (10). Set $\eta = (y - y')/\omega$. We immediately see that $\eta \in \mathcal{R}(A^T)$, and $\eta_i = (y_i - y_i')/\omega = 1$ for every $x_i > 0$ (since $y_i = \omega$ and $y_i' = 0$ for this case). Similarly we have $\eta_i = -1$ for every i with $x_i < 0$. For $x_i = 0$, since $\omega < y_i + y_i'$ and both y_i and y_i' are negative, it follows that $|\eta_i| = |y_i - y_i'|/|\omega| < |y_i + y_i'|/|\omega| < 1$.

Conversely, assume that there is a vector $\eta \in \mathcal{R}(A^T)$ such that $\eta_i = 1$ for all $x_i > 0$, $\eta_i = -1$ for all $x_i < 0$, and $|\eta_i| < 1$ for all $x_i = 0$. We now construct a vector (y,y',ω) satisfying (10). Indeed, let us first set $\omega = -1$, and then set $y_i = 0$, $y_i' = -1$ for $x_i < 0$, and $y_i = -1, y_i' = 0$ for $x_i > 0$. For those i with $x_i = 0$, since $|\eta_i| < 1$, there exists a constant ε_i such that $0 < \varepsilon_i < (1 - |\eta_i|)/2$, and thus we define y_i and y_i' as follows:

$$\begin{cases} y_i = -\varepsilon_i - \eta_i \text{ and } y_i' = -\varepsilon_i & \text{if } \eta_i > 0, \\ y_i = -\varepsilon_i \text{ and } y_i' = \eta_i - \varepsilon_i & \text{otherwise.} \end{cases} \quad (24)$$

From the above construction, it is easy to see that $y - y' = -\eta$. Thus $y - y' \in \mathcal{R}(A^T)$. To verify that (y,y',ω) satisfies (10), it is sufficient to show that $-1 = \omega < y_i + y_i'$, $y_i < 0$, $y_i' < 0$ for all $x_i = 0$. Indeed, we see from (24) that both y_i and y_i' are negative, and

$$|y_i + y_i'| = \begin{cases} |(-\varepsilon_i - \eta_i) + (-\varepsilon)| \le 2\varepsilon_i + |\eta_i| & \text{if } \eta_i > 0 \\ |(-\varepsilon) + (\eta_i - \varepsilon_i)| \le 2\varepsilon_i + |\eta_i| & \text{otherwise,} \end{cases}$$

which by the definition of ε_i implies that $|y_i + y_i'| < 1$. Since $y_i < 0$ and $y_i' < 0$, this implies that $0 > y_i + y_i' > -1 = \omega$. Thus (y,y',ω) constructed as above does satisfy (10).

Proof of Lemma 4.4: The mutual-coherence condition implies that any x with $\|x\|_0 < (1+1/\mu(A))/2$ is both the unique sparsest and the unique least ℓ_1-norm solutions to the system $Az = y = Ax$. Let $S_1 = \{i : x_i > 0\}$ and $S_2 = \{i : x_i < 0\}$. By Theorem 2.8, there exists a vector $\eta \in \mathcal{R}(A^T)$ satisfying the RSP (20) at x, i.e., $\eta_i = 1$ for $i \in S_1$, $\eta_i = -1$ for $i \in S_2$, and $|\eta_i| < 1$ otherwise. Since x here can be any sparse vector with $\|x\|_0 \le K^0 =: \left\lfloor \frac{1}{2}\left(1 + \frac{1}{\mu(A)}\right) \right\rfloor$, the above-defined S_1 and S_2 can be any disjoint subsets S_1, S_2 of $\{1,...,n\}$ with $|S_1| + |S_2| \le K^0$. Thus A^T has the RSP of at least order K^0. Both the RIP and NSP of order $2K$ imply that any sparse vector x with $\|x\|_0 \le K$ is the unique sparsest solution and the unique least ℓ_1-norm solution to the system $Az = y = Ax$. By the same analysis above, it implies that the RSP of order K holds.

Yun-Bin Zhao Yun-Bin Zhao received his PhD degree from the Institute of Applied Mathematics, Chinese Academy of Sciences (CAS) in 1998. From June 1998 to February 2001 he was a postdoctoral research fellow with the Institute of Computational Mathematics, CAS, and with the Department of Systems Engineering & Engineering Management, Chinese University of Hong Kong. From 2001 to 2003, He was an assistant professor in the Academy of Mathematics and Systems Science (AMSS), CAS, and from 2003 to 2007 he was an associate professor in AMSS. He joined the University of Birmingham in 2007, as a lecturer of mathematical optimization. Since 2013, he is a senior lecturer in the same university. He serves as an associate editor of *Applied Mathematics and Computation*. His research interests include the operations research, computational optimization, convex analysis, numerical linear algebra, and their applications.