

Equivalence and Strong Equivalence between Sparsest and Least ℓ_1 -Norm Nonnegative Solutions of Linear Systems and Their Application

YUN-BIN ZHAO *

(First version: June 2012; Revised in November 2012, September and December 2013)

Abstract. Many practical problems can be formulated as ℓ_0 -minimization problems with nonnegativity constraints, which seek the sparsest nonnegative solutions to underdetermined linear systems. Recent study indicates that ℓ_1 -minimization is efficient for solving some classes of ℓ_0 -minimization problems. From a mathematical point of view, however, the understanding of the relationship between ℓ_0 - and ℓ_1 -minimization remains incomplete. In this paper, we further discuss several theoretical questions associated with these two problems. For instance, how to completely characterize the uniqueness of least ℓ_1 -norm nonnegative solutions to a linear system, and is there any alternative matrix property that is different from existing ones, and can fully characterize the uniform recovery of K -sparse nonnegative vectors? We prove that the fundamental strict complementarity theorem of linear programming can yield a necessary and sufficient condition for a linear system to have a unique least ℓ_1 -norm nonnegative solution. This condition leads naturally to the so-called range space property (RSP) and the ‘full-column-rank’ property, which altogether provide a broad understanding of the relationship between ℓ_0 - and ℓ_1 -minimization. Motivated by these results, we introduce the concept of the ‘RSP of order K ’ that turns out to be a full characterization of the uniform recovery of K -sparse nonnegative vectors. This concept also enables us to develop certain conditions for the non-uniform recovery of sparse nonnegative vectors via the so-called weak range space property.

Key words Linear programming, Underdetermined linear system, Sparsest nonnegative solution, Range space property, Uniform (non-uniform) recovery

*School of Mathematics, University of Birmingham, Edgbaston, Birmingham B15 2TT, United Kingdom (y.zhao.2@bham.ac.uk). This work was supported by the Engineering and Physical Sciences Research Council (EPSRC) under the grant #EP/K00946X/1.

1 Introduction

In this paper, we use $\|\cdot\|_0$ to denote the number of nonzero components of a vector. We investigate the following optimization problem with nonnegativity constraints:

$$\min\{\|x\|_0 : Ax = b, x \geq 0\}, \quad (1)$$

which is called an ℓ_0 -minimization problem or ℓ_0 -problem. It is well known that nonnegativity constraints are quite common in mathematical optimization and numerical analysis (see [1] and the references therein). Clearly, the aim of the problem (1) is to find a sparsest nonnegative solution to a system of linear equations. This problem has found so many applications in such areas as signal and image processing [2–10], machine learning [11–15], pattern recognition and computer vision [9, 16], proteomics [17], to name but a few. This problem is a special case of compressed nonnegative sparse coding [18–19], and rank minimization with positive semidefinite constraints (e.g., [10, 20–21]). It is closely related to the nonnegative matrix factorization as well [22–24].

The ℓ_0 -minimization problem is NP-hard [25]. Current theories and algorithms for ℓ_0 -minimization are mainly developed through certain heuristic methods and continuous approximations. A large amount of recent attention is attracted to the ℓ_1 -problem

$$\min\{\|x\|_1 : Ax = b, x \geq 0\} \quad (2)$$

which is efficient for solving (1) in many situations, so is the reweighted ℓ_1 -minimization method (e.g., [26–27]). In this paper, the optimal solution of the problem (2) is called the *least ℓ_1 -norm nonnegative solution* to the linear system $Ax = b$. Any linear programming solver can be used to solve the problem (2). Various specialized algorithms for this problem have also been proposed in the literature (e.g., [3, 9, 13, 28–29]).

Over the past few years, ℓ_0 -problems without nonnegativity constraints have been extensively studied in the fields of sparse signal and image processing and compressed sensing. Both theories and numerical methods have been developed for this problem (e.g., [26, 27, 30–34]). However, the sparsest solution and sparsest nonnegative solution to a linear system are very different from a mathematical point of view. The analysis and many results developed for the sparsest solution of a linear system cannot apply to the sparsest nonnegative solution straightaway. So far, the understanding of the relationship between (1) and (2), and the ℓ_1 -method-based recovery theory for sparse nonnegative vectors remains incomplete, compared with the linear systems without nonnegativity constraints. For example, the following important questions have not well addressed at present:

- (a) How to completely characterize the uniqueness of least ℓ_1 -norm nonnegative solutions to a linear system?
- (b) Empirical results indicate that when existing sufficient criteria for the efficiency of the ℓ_1 -method fail, the ℓ_1 -method still succeeds in solving ℓ_0 -problems in many such situations. The ℓ_1 -method actually performs remarkably better than what the current theories have indicated. How to interpret such a gap between the performance of the ℓ_1 -method indicated by the current theories and that demonstrated by the numerical simulations?

- (c) Are there any other matrix properties that are different from the existing ones (such as the restricted isometric property (RIP) [31, 35–36] and the null space property (NSP) [6, 37]), and can fully characterize the uniform recovery (i.e., the exact recovery of all K -sparse nonnegative vectors by a single sensing matrix)?
- (d) Is it possible to develop some theory for the recovery of sparse nonnegative vectors (representing signals or images) that may go beyond the scope of the current uniform recovery?

In general, for a given pair (A, b) , the sparsest nonnegative solution to the system $Ax = b$ is not unique. So it is important to distinguish the equivalence and the strong equivalence between (1) and (2). In this paper, ℓ_0 - and ℓ_1 -problems are said to be equivalent if the ℓ_0 -problem has an optimal solution that coincides with the unique optimal solution to the ℓ_1 -problem. We say that the ℓ_0 - and ℓ_1 -problems are strongly equivalent if the ℓ_0 -problem has a unique optimal solution that coincides with the unique optimal solution to the ℓ_1 -problem. Clearly, the ‘strong equivalence’ implies the ‘equivalence’, but the converse is not true in general. The ‘equivalence’ does not require an ℓ_0 -problem to have a unique optimal solution. As shown by our later analysis, the ‘equivalence’ concept enables us to broadly understand the relationship between ℓ_0 - and ℓ_1 -minimization, making it possible to address the aforementioned questions. Of course, the above-mentioned questions (a)-(d) can be partially addressed by applying the existing theories based on such concepts as the mutual coherence [38–40, 45], ERC [41–42], RIP [31, 35–36], NSP [6, 37], outwardly k -neighborliness property [4], and the verifiable condition [43–44]. However, these existing conditions are relatively restrictive in the sense that they imply the strong equivalence (instead of the equivalence) between ℓ_0 - and ℓ_1 -problems. For instance, Donoho and Tanner [4] have given a geometric condition, i.e., the outwardly K -neighborliness property of a sensing matrix, which guarantees that if a K -sparse nonnegative vector is a solution to the linear system $Ax = b$, then it must be the unique optimal solution to both problems (1) and (2). From a null-space perspective, Khajehnejad et al [6] have shown that K -sparse nonnegative vectors can be recovered by ℓ_1 -minimization if and only if the null space of A satisfies certain property. Thus both the outwardly K -neighborliness property [4] and the null space property [6] imply the strong equivalence between (1) and (2). We note that the strong equivalence conditions fail to explain the success of the ℓ_1 -method for solving ℓ_0 -problems with multiple optimal solutions. We also note that the uniqueness of least ℓ_1 -norm nonnegative solutions to a linear system plays a fundamental role in both theoretical and practical efficiencies of the ℓ_1 -method for solving ℓ_0 -problems. While the strong equivalence conditions are sufficient for the uniqueness of least ℓ_1 -norm nonnegative solutions to a linear system, these conditions are not the necessary condition.

The first purpose of this paper is to completely address the question (a) by developing a necessary and sufficient condition for the uniqueness of least ℓ_1 -norm nonnegative solutions to a linear system. We establish this condition through the strict complementarity theory of linear programming, which leads naturally to the new concept of the range space property (RSP) of A^T . Based on this result, we further point out that the equivalence between ℓ_0 - and ℓ_1 -problems can be theoretically interpreted by the RSP of A^T . That is, an ℓ_1 -problem is equivalent to an ℓ_0 -problem if and only if the RSP of A^T holds at an optimal solution of the ℓ_0 -problem.

While the RSP of A^T is defined locally at an individual vector (e.g., the solution to an ℓ_1 - or an ℓ_0 -problem), it provides a complete and practical checking condition for the uniqueness of least ℓ_1 -norm nonnegative solutions to a linear system, and the RSP of A^T yields a broad understanding of the efficiency of the ℓ_1 -method for solving ℓ_0 -problems (see the discussion in Sect. 3 for details). It turns out that when the strong equivalence conditions fail (e.g., the ℓ_0 -problem has multiple optimal solutions), the RSP of A^T is still able to explain the success of the ℓ_1 -method for solving ℓ_0 -problems in many such situations. Thus, the current gap between the performance of the ℓ_1 -method indicated by the existing theories and that observed from numerical simulations can be clarified in terms of the RSP of A^T , leading to a certain answer to the question (b).

Although a global RSP-type condition for the equivalence between ℓ_0 - and ℓ_1 -problems remains not clear at present, such a condition for the strong equivalence between these two problems can be developed. In Sect. 4, we further introduce a matrix property, called the RSP of order K , through which we provide a characterization of the uniform recovery of all K -sparse nonnegative vectors. As a result, the RSP of order K is a strong equivalence condition for ℓ_0 - and ℓ_1 -problems. Interestingly, the variants of this new concept make it possible to extend uniform recovery to non-uniform recovery of some sparse nonnegative vectors, to which the uniform recovery does not apply. Such an extension is important not only from a mathematical point of view, but from the viewpoint of many practical applications as well. For instance, when many columns of A are important, the sparsest solution to the linear system $Ax = b$ may not be sparse enough to satisfy the uniform recovery condition. The RSP of order K and its variants make it possible to address the aforementioned questions (c) and (d).

This paper is organized as follows. In Sect. 2, we develop a necessary and sufficient condition for a linear system to have a unique least ℓ_1 -norm nonnegative solution. In Sect. 3, we provide an efficiency analysis for the ℓ_1 -method through the RSP of A^T . In Sect. 4, we develop some (uniform and non-uniform) recovery conditions for K -sparse nonnegative vectors via the so-called RSP of order K and its variants. Conclusions are given in the last section.

2 Uniqueness of least ℓ_1 -norm nonnegative solutions

We use the following notation: Let R_+^n be the first orthant of R^n , the n -dimensional Euclidean space. Let $e = (1, 1, \dots, 1)^T \in R^n$ be the vector of ones throughout this paper. For two vectors $u, v \in R^n$, $u \leq v$ means $u_i \leq v_i$ for every $i = 1, \dots, n$, and in particular, $v \geq 0$ means $v \in R_+^n$. For a set $S \subseteq \{1, 2, \dots, n\}$, $|S|$ denotes the cardinality of S , and $S_c = \{1, 2, \dots, n\} \setminus S$ is the complement of S . For a matrix A with columns a_j , $1 \leq j \leq n$, we use A_S to denote the submatrix of A with columns $a_j, j \in S$. Similarly, x_S denotes the subvector of x with components $x_j, j \in S$. For $x \in R^n$, let $\|x\|_1 = \sum_{i=1}^n |x_i|$ denote the ℓ_1 -norm of x . For $A \in R^{m \times n}$, we use $\mathcal{R}(A^T)$ to denote the range space of A^T , i.e., $\mathcal{R}(A^T) = \{A^T u : u \in R^m\}$.

In this section, we develop a necessary and sufficient condition for x to be the unique least ℓ_1 -norm nonnegative solution to a linear system. Note that when x is the unique optimal solution to the problem (2), there is no other nonnegative solution $w \neq x$ such that $\|w\|_1 \leq \|x\|_1$. Thus

the uniqueness of x is equivalent to

$$\{w : Aw = b, w \geq 0, \|w\|_1 \leq \|x\|_1\} = \{x\}.$$

Since $x \geq 0$ and $w \geq 0$, we have $\|w\|_1 = e^T w$ and $\|x\|_1 = e^T x$. Thus the above relation can be further written as $\{w : Aw = Ax, e^T w \leq e^T x, w \geq 0\} = \{x\}$. Consider the following linear programming (LP) problem with the variable $w \in R^n$:

$$\min\{0^T w : Aw = Ax, e^T w \leq e^T x, w \geq 0\}, \quad (3)$$

which is feasible (since $w = x$ is always a feasible solution), and the optimal value of the problem is finite (equal to zero). From the above discussion, we immediately have the following observation.

Lemma 2.1 *x is the unique least ℓ_1 -norm nonnegative solution to the system $Ax = b$ if and only if $w = x$ is the unique optimal solution to the problem (3), i.e., $(w, t) = (x, 0)$ is the unique optimal solution to the following problem:*

$$\min\{0^T w : Aw = Ax, e^T w + t = e^T x, (w, t) \geq 0\} \quad (4)$$

where t is a slack variable introduced into (3).

Note that the dual problem of (4) is given by

$$\max\{(Ax)^T y + (e^T x)\beta : A^T y + \beta e \leq 0, \beta \leq 0\} \quad (5)$$

where y and β are variables. Throughout this section, we use $(s, r) \in R_+^{n+1}$ to denote the slack variables of the problem (5), i.e.,

$$s = -(A^T y + \beta e) \geq 0, \quad r = -\beta \geq 0.$$

Let us recall a fundamental theorem for LP problems. Let $B \in R^{m \times n}$ be a given matrix, and $p \in R^m$ and $c \in R^n$ be two given vectors. Consider the LP problem

$$\min\{c^T x : Bx = p, x \geq 0\}, \quad (6)$$

and its dual problem

$$\max\{p^T y : B^T y + s = c, s \geq 0\}. \quad (7)$$

Any optimal solution pair $(x, (y, s))$ to (6) and (7) satisfies the so-called complementary slackness condition: $x^T s = 0, x \geq 0$ and $s \geq 0$. Moreover, if a solution pair $(x, (y, s))$ satisfies that $x + s > 0$, it is called a strictly complementary solution pair. For any feasible LP problems (6) and (7), there always exists a pair of strictly complementary solutions.

Lemma 2.2([46]) (i) (Optimality condition) $(x, (y, s))$ is a solution pair of the LP problems (6) and (7) if and only if it satisfies the following conditions: $Bx = p, B^T y + s = c, x \geq 0, s \geq 0$, and $x^T s = 0$. (ii) (Strict complementarity) If (6) and (7) are feasible, then there exists a pair $(x^*, (y^*, s^*))$ of strictly complementary solutions to (6) and (7).

We now prove the following necessary condition for the problem (2) to have a unique optimal solution.

Lemma 2.3 *If x is the unique least ℓ_1 -norm nonnegative solution to the system $Ax = b$, then there exists a vector $\eta \in \mathcal{R}(A^T)$ satisfying*

$$\eta \in \mathcal{R}(A^T), \eta_i = 1 \text{ for } i \in J_+, \text{ and } \eta_i < 1 \text{ for } i \notin J_+, \quad (8)$$

where $J_+ = \{i : x_i > 0\}$.

Proof. Consider the problem (4) and its dual problem (5), both of which are feasible. By Lemma 2.2, there exists an optimal solution (w^*, t^*) to the problem (4) and an optimal solution (y^*, β^*) to (5) such that these two solutions constitute a pair of strictly complementary solutions. Let $(s^*, r^*) = (-A^T y^* - \beta^* e, -\beta^*)$ be the value of the associated slack variables of the dual problem (5). Then by the strict complementarity, we have

$$(w^*)^T s^* = 0, \quad t^* r^* = 0, \quad w^* + s^* > 0, \quad t^* + r^* > 0. \quad (9)$$

Since x is the unique least ℓ_1 -norm nonnegative solution to $Ax = b$, by Lemma 2.1, $(x, 0)$ is the unique optimal solution to the problem (4). Thus

$$(w^*, t^*) = (x, 0), \quad (10)$$

which implies that $w_i^* > 0$ for all $i \in J_+ = \{i : x_i > 0\}$ and $w_i^* = 0$ for all $i \notin J_+$. Thus it follows from (9) and (10) that $r^* > 0$, $s_i^* = 0$ for all $i \in J_+$, and that $s_i^* > 0$ for all $i \notin J_+$. That is,

$$\beta^* < 0, \quad (A^T y^* + \beta^* e)_i = 0 \text{ for } i \in J_+, \quad (A^T y^* + \beta^* e)_i < 0 \text{ for } i \notin J_+,$$

which can be written as

$$\beta^* < 0, \quad \left[A^T \begin{pmatrix} y^* \\ -\beta^* \end{pmatrix} - e \right]_i = 0 \text{ for } i \in J_+, \quad \left[A^T \begin{pmatrix} y^* \\ -\beta^* \end{pmatrix} - e \right]_i < 0 \text{ for } i \notin J_+.$$

By setting $\eta = A^T y^* / (-\beta^*)$, the condition above is equivalent to

$$\eta \in \mathcal{R}(A^T), \eta_i = 1 \text{ for } i \in J_+, \text{ and } \eta_i < 1 \text{ for } i \notin J_+,$$

as desired. \square

Throughout this paper, the condition (8) is called the *range space property* (RSP) of A^T at $x \geq 0$. It is worth noting that the RSP (8) can be easily checked by simply solving the following LP problem:

$$t^* = \min \left\{ t : (A_{J_+})^T y = e_{J_+}, (A_{J_{+c}})^T y = \eta_{J_{+c}}, \eta_{J_{+c}} \leq t e_{J_{+c}} \right\},$$

where $J_{+c} = \{1, 2, \dots, n\} \setminus J_+$. Clearly, $t^* < 1$ if and only if the RSP (8) holds. Lemma 2.3 shows that the RSP of A^T at x is a necessary condition for the ℓ_1 -problem to have a unique optimal solution. We now prove another necessary condition.

Lemma 2.4 *If x is the unique least ℓ_1 -norm nonnegative solution to the system $Ax = b$, then the matrix*

$$M = \begin{pmatrix} A_{J_+} \\ e_{J_+}^T \end{pmatrix} \quad (11)$$

has full column rank, where $J_+ = \{i : x_i > 0\}$.

Proof. Assume the contrary that the columns of M defined by (11) is linearly dependent. Then there exists a vector $u \in R^{|J_+|}$ such that

$$u \neq 0, \quad Mu = \begin{pmatrix} A_{J_+} \\ e_{J_+}^T \end{pmatrix} u = 0. \quad (12)$$

Let (w, t) be given by $w = (w_{J_+}, w_{J_0}) = (x_{J_+}, 0)$ and $t = 0$, where $J_0 = \{i : i \notin J_+\}$. Then it is easy to see that such defined (w, t) is an optimal solution to the problem (4). On the other hand, let us define (\tilde{w}, \tilde{t}) as follows:

$$\tilde{w} = (\tilde{w}_{J_+}, \tilde{w}_{J_0}) = (w_{J_+} + \lambda u, 0), \quad \tilde{t} = 0.$$

Since $w_{J_+} = x_{J_+} > 0$, there exists a small $\lambda \neq 0$ such that

$$\tilde{w}_{J_+} = w_{J_+} + \lambda u \geq 0. \quad (13)$$

Substituting (\tilde{w}, \tilde{t}) into the constraints of the problem (4), we see from (12) that (\tilde{w}, \tilde{t}) satisfies all those constraints. Thus (\tilde{w}, \tilde{t}) is also an optimal solution to the problem (4). It follows from (13) that $\tilde{w}_{J_+} \neq w_{J_+}$ since $\lambda u \neq 0$. Therefore, the optimal solution to (4) is not unique. However, by Lemma 2.1, when x is the unique least ℓ_1 -norm nonnegative solution to the system $Ax = b$, the problem (4) must have a unique optimal solution. This contradiction shows that M must have full column rank. \square

The next result shows that the combination of the necessary conditions developed in Lemmas 2.3 and 2.4 is sufficient for the ℓ_1 -problem to have a unique optimal solution.

Lemma 2.5 *Let $x \geq 0$ be a solution to the system $Ax = b$. If the condition (8) (i.e., the RSP of A^T) is satisfied at x , and the matrix M given by (11) has full column rank, then x is the unique least ℓ_1 -norm nonnegative solution to the system $Ax = b$.*

Proof. By Lemma 2.1, to prove that x is the unique least ℓ_1 -norm nonnegative solution to the system $Ax = b$, it is sufficient to prove that the problem (4) has a unique optimal solution $(x, 0)$. First, the condition (8) implies that there exist η and y such that

$$A^T y = \eta, \quad \eta_i = 1 \text{ for } i \in J_+, \text{ and } \eta_i < 1 \text{ for } i \notin J_+.$$

By setting $\beta = -1$, the relation above can be written as

$$(A^T y)_i + \beta = 0 \text{ for } i \in J_+, \text{ and } (A^T y)_i + \beta < 0 \text{ for } i \notin J_+, \quad (14)$$

from which we see that (y, β) satisfies all constraints of the problem (5). We now further verify that it is an optimal solution to (5). In fact, by (14), the objective value of (5) at (y, β) is

$$(Ax)^T y + (e^T x)\beta = x^T (A^T y) + (e^T x)\beta$$

$$\begin{aligned}
&= \sum_{i \in J_+} x_i (A^T y)_i + \beta \sum_{i \in J_+} x_i \\
&= -\beta \sum_{i \in J_+} x_i + \beta \sum_{i \in J_+} x_i = 0.
\end{aligned} \tag{15}$$

Since the optimal value of (4) is zero, by LP duality theory, the maximum value of the dual problem is also zero. Thus it follows from (15) that the point (y, β) satisfying (14) is an optimal solution to the problem (5).

We now prove that the optimal solution of (4) is uniquely determined under the assumption of the theorem. Assume that (w^*, t^*) is an arbitrary optimal solution to the problem (4), which of course satisfies all constraints of (4), i.e.,

$$Aw^* = Ax, \quad e^T w^* + t^* = e^T x, \quad (w^*, t^*) \geq 0. \tag{16}$$

Since (y, β) satisfying (14) is an optimal solution of (5), $((w^*, t^*), ((y, \beta), s))$ is a solution pair to (4) and (5). From (14), we see that the dual slack variables $s_i = -((A^T y)_i + \beta) > 0$ for $i \notin J_+$ and $r = -\beta = 1 > 0$. By complementary slackness property (Lemma 2.2(i)), we must have that

$$t^* = 0, \quad w_i^* = 0 \text{ for all } i \notin J_+.$$

By substituting these known components into (16) and noting that $x_i = 0$ for $i \notin J_+$, we see that the remaining components of (w^*, t^*) satisfy

$$A_{J_+} w_{J_+}^* = Ax = A_{J_+} x_{J_+}, \quad e_{J_+}^T w_{J_+}^* = e^T x = e_{J_+}^T x_{J_+}, \quad w_{J_+}^* \geq 0.$$

Since the matrix $M = \begin{pmatrix} A_{J_+} \\ e_{J_+}^T \end{pmatrix}$ has full column rank, $w_{J_+}^* = x_{J_+}$ is the unique solution to the reduced system above. Therefore, (w^*, t^*) is uniquely given by $(x, 0)$. In other words, the only optimal solution to the problem (4) is $(x, 0)$. By Lemma 2.1, x must be the unique least ℓ_1 -norm nonnegative solution to the system $Ax = b$. \square

By Lemmas 2.3, 2.4 and 2.5, we have the main result of this section.

Theorem 2.6 *Let x be a nonnegative solution to the system $Ax = b$. Then x is the unique least ℓ_1 -norm nonnegative solution to the system $Ax = b$ if and only if the RSP (8) holds at x and the matrix $M = \begin{pmatrix} A_{J_+} \\ e_{J_+}^T \end{pmatrix}$ has full column rank, where $J_+ = \{i : x_i > 0\}$.*

Clearly, when A_{J_+} has full column rank, so does the matrix M given by (11). The converse is not true. In general, when M has full column rank, it does not imply that the matrix A_{J_+} has full column rank. For instance, $M = \begin{pmatrix} A_{J_+} \\ e_{J_+}^T \end{pmatrix} = \begin{pmatrix} -1 & 1 \\ 0 & 0 \\ 1 & 1 \end{pmatrix}$ has full column rank, but $A_{J_+} = \begin{pmatrix} -1 & 1 \\ 0 & 0 \end{pmatrix}$ does not. However, when the RSP (8) holds at x , we see that $e_{J_+} = A_{J_+}^T u$ for some $u \in R^n$, in which case A_{J_+} has full column rank if and only if $\begin{pmatrix} A_{J_+} \\ e_{J_+}^T \end{pmatrix}$ has full column rank. Thus Theorem 2.6 can be restated as follows.

Theorem 2.7 *Let x be a nonnegative solution to the system $Ax = b$. Then x is the unique least ℓ_1 -norm nonnegative solution to the system $Ax = b$ if and only if the RSP (8) holds at x and the matrix A_{J_+} has full column rank, where $J_+ = \{i : x_i > 0\}$.*

The above results completely characterize the uniqueness of least ℓ_1 -norm nonnegative solutions to a system of linear equations, and thus the question (a) in Sect. 1 has been fully addressed. Note that $A \in \mathbb{R}^{m \times |J_+|}$, so when it has full column rank, we must have $\text{rank}(A_{J_+}) = |J_+| \leq m$. Thus Theorem 2.7 shows that *if the ℓ_1 -problem has a unique optimal solution x , then x must be m -sparse*. We can use the results established in this section to discuss other questions associated with ℓ_0 - and ℓ_1 -problems (see the remainder of the paper for details).

We now close this section by giving two examples to show that our necessary and sufficient condition can be easily used to check the uniqueness of least ℓ_1 -norm nonnegative solutions of linear systems.

Example 2.8 Consider the linear system $Ax = b$ with

$$A = \begin{pmatrix} 1 & 0 & -1 & -1 \\ 0 & -1 & -1 & 6 \\ 0 & 0 & -1 & 1 \end{pmatrix}, \quad b = \begin{pmatrix} 1/2 \\ -1/2 \\ 0 \end{pmatrix},$$

to which $x^* = (1/2, 1/2, 0, 0)^T$ is a nonnegative solution. It is easy to see that the submatrix A_{J_+} associated with this solution has full column rank. Moreover, by taking $y = (1, -1, 0)^T$, we have $\eta = A^T y = (1, 1, 0, -7)^T \in \mathcal{R}(A^T)$, which clearly satisfies (8). Thus the RSP of A^T holds at x^* . Therefore, by Theorem 2.7 (or Lemma 2.5), x^* is the unique least ℓ_1 -norm nonnegative solution to the system $Ax = b$.

Example 2.9 Consider the linear system $Ax = b$ with

$$A = \begin{pmatrix} 1 & 0 & -1 & 1 \\ 1 & -0.1 & 0 & -0.2 \\ 0 & 0 & -1 & 1 \end{pmatrix}, \quad b = \begin{pmatrix} 1/2 \\ -1/2 \\ 0 \end{pmatrix},$$

to which $x^* = (1/2, 10/3, 10/3, 10/3)^T$ is a least ℓ_1 -norm nonnegative solution. By taking $y = (11, -10, -12)^T$, we have $\eta = A^T y = (1, 1, 1, 1)^T \in \mathcal{R}(A^T)$. Thus the RSP of A^T holds at x^* . However, the matrix

$$A_{J_+} = \begin{pmatrix} 1 & 0 & -1 & 1 \\ 1 & -0.1 & 0 & -0.2 \\ 0 & 0 & -1 & 1 \end{pmatrix}$$

does not have full column rank. By Theorem 2.7, x^* is not the unique least ℓ_1 -norm nonnegative solution to the system $Ax = b$. In fact, we have another least ℓ_1 -norm nonnegative solution given by $\tilde{x} = (1/2, 10, 0, 0)^T$ (for which the associated A_{J_+} has full column rank, but the RSP of A^T does not hold at \tilde{x}).

3 RSP-based efficiency analysis for ℓ_1 -minimization

For linear systems without nonnegativity constraints, some sufficient conditions for the strong equivalence between ℓ_0 - and ℓ_1 -problems have been developed in the literature. If these sufficient conditions are applied directly to sparsest nonnegative solutions of linear systems, the resulting criteria would be very restrictive. For instance, by applying the mutual coherence condition, we immediately conclude that if a nonnegative solution x obeys $\|x\|_0 < (1 + 1/\mu(A))/2$ where

$\mu(A)$ denotes the mutual coherence of A (i.e., $\mu(A) = \max_{i \neq j} |a_i^T a_j| / (\|a_i\|_2 \|a_j\|_2)$ where $a_i, 1 \leq i \leq n$, are the columns of A), then x is the unique sparsest solution and the unique least ℓ_1 -norm solution to the linear system $Ax = b$. In this case, the unique sparsest nonnegative solution coincides with the unique sparsest solution and the unique least ℓ_1 -norm solution of the linear system. Clearly, such a sufficient (strong-equivalence-type) condition is too restrictive. In fact, a sparsest nonnegative solution is usually not a sparsest one of the linear system, and the sparsest nonnegative ones can be multiple (as shown by Example 3.4 in this section). Although some conditions have been developed specifically for the sparsest nonnegative solution in the literature (e.g., [3–4, 6]), these conditions still imply the strong equivalence between ℓ_0 - and ℓ_1 -problems. They can only partially explain the success of the ℓ_1 -method for solving ℓ_0 -problems. In this section, we point out that Theorem 2.6 or 2.7 can be used to broadly clarify the relationship between ℓ_0 - and ℓ_1 -problems, and thus the efficiency of the ℓ_1 -method can be more widely interpreted through the RSP of A^T . First, we have the following property for sparsest nonnegative solutions.

Lemma 3.1 *If x is a sparsest nonnegative solution to the system $Ax = b$, then $M = \begin{pmatrix} A_{J_+} \\ e_{J_+}^T \end{pmatrix}$ has full column rank, where $J_+ = \{i : x_i > 0\}$.*

Proof. Let x be a sparsest nonnegative solution to the system $Ax = b$ and let $J_+ = \{i : x_i > 0\}$. Assume by contrary that the columns of M are linearly dependent. Then there exists a vector $v \neq 0$ in $R^{|J_+|}$ such that

$$\begin{pmatrix} A_{J_+} \\ e_{J_+}^T \end{pmatrix} v = 0.$$

It follows from $e_{J_+}^T v = 0$ and $v \neq 0$ that v must have at least two nonzero components with different signs, i.e., $v_i v_j < 0$ for some $i \neq j$. Define the vector $\tilde{v} \in R^n$ as follows: $\tilde{v}_{J_+} = v$ and $\tilde{v}_i = 0$ for all $i \notin J_+$. We consider the vector $y(\lambda) = x + \lambda \tilde{v}$ where $\lambda \geq 0$. Note that $y(\lambda)_i = 0$ for all $i \notin J_+$, and that

$$Ay(\lambda) = Ax + A(\lambda \tilde{v}) = b + \lambda A_{J_+} v = b.$$

Thus $y(\lambda)$ is also a solution to the linear system $Ax = b$. By the definition of \tilde{v} , \tilde{v} has at least one negative component. Thus let

$$\lambda^* = \frac{x_{i_0}}{-\tilde{v}_{i_0}} = \min \left\{ \frac{x_i}{-\tilde{v}_i} : \tilde{v}_i < 0 \right\},$$

where λ^* must be a positive number and $i_0 \in J_+$. By such a choice of λ^* and the definition of $y(\lambda^*)$, we conclude that $y(\lambda^*) \geq 0$, $y(\lambda^*)_i = 0$ for $i \notin J_+$, and $y(\lambda^*)_{i_0} = 0$ with $i_0 \in J_+$. Thus $y(\lambda^*)$ is a nonnegative solution to the linear system $Ax = b$, which is sparser than x . This is a contradiction. Therefore, M must have full column rank. \square

By Theorem 2.6 and Lemma 3.1, we immediately have the following result.

Theorem 3.2 *ℓ_0 - and ℓ_1 -problems are equivalent if and only if the RSP (8) holds at an optimal solution of the ℓ_0 -problem. (In other words, a sparsest nonnegative solution x to the system $Ax = b$ is the unique least ℓ_1 -norm nonnegative solution to the system if and only if the RSP (8) holds at x .)*

Proof. Assume that problems (1) and (2) are equivalent. So the ℓ_0 -problem has an optimal solution x that is the unique least ℓ_1 -norm nonnegative solution to the system $Ax = b$. Thus, by Theorem 2.6 (or Lemma 2.3), the RSP (8) must hold at x . Conversely, assume that the RSP (8) holds at an optimal solution x to the ℓ_0 -problem. Since x is a sparsest nonnegative solution to the system $Ax = b$, by Lemma 3.1, the matrix $\begin{pmatrix} A_{J_+} \\ e_{J_+}^T \end{pmatrix}$ has full column rank. Thus by Lemma 2.5 (or Theorem 2.6) again, x must be the unique least ℓ_1 -norm nonnegative solution to the system $Ax = b$. Hence ℓ_0 - and ℓ_1 -problems are equivalent. \square

It should be pointed out that the equivalence between ℓ_0 - and ℓ_1 -problems are characterized implicitly in Theorem 3.2 in the sense that the RSP condition therein is defined locally at a solution of the ℓ_0 -problem. Whether or not a checkable RSP-type equivalence (instead of strong equivalence) condition can be developed for ℓ_0 - and ℓ_1 -problems remains an open question. However, Theorem 3.2 is still important from a theoretical point of view, and it can be used to explain the numerical performance of the ℓ_1 -method more efficiently than strong equivalence conditions. Since the RSP (8) at an optimal solution of ℓ_0 -problem is a necessary and sufficient condition for the equivalence between ℓ_0 - and ℓ_1 -problems, all existing sufficient conditions for strong equivalence (or equivalence) between these two problems must imply the RSP (8). However, the converse is clearly not true in general, as shown by the following example.

Example 3.3 (When existing criteria fail, the RSP may still succeed).

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

For this example, the system $Ax = b$ does not have a solution x with $\|x\|_0 = 1$. So $x^* = (1, 0, \sqrt{3}, 0, 0, 0)^T$ is a sparsest nonnegative solution of this linear system. Note that the mutual coherence $\mu(A) = \max_{i \neq j} |a_i^T a_j| / \|a_i\|_2 \|a_j\|_2 = \sqrt{2}/\sqrt{3}$. Thus the mutual coherence condition $\|x\|_0 < \frac{1}{2}(1 + 1/\mu(A)) = (\sqrt{2} + \sqrt{3})/(2\sqrt{2}) \approx 1.077$ fails for this example. The RIP [35] fails since the last two columns of A are linearly dependent. This example also fails to comply with the definition of the NSP. Let us now check the RSP of A^T at x^* . By taking $y = (\frac{1}{2} + \sqrt{3}, \frac{1}{2}, -1)^T$, we have

$$\eta = A^T y = \left(1, -\left(\frac{1}{2} + \sqrt{3}\right), 1, -\frac{1}{2}, \frac{2\sqrt{3}-1}{2\sqrt{2}}, -\frac{2\sqrt{3}-1}{2\sqrt{2}} \right)^T \in \mathcal{R}(A^T),$$

where the first and third components of η are equal to 1 (corresponding to $J_+ = \{1, 3\}$ determined by x^*) and all other components of η are less than 1. Thus the RSP (8) holds at x^* . By Theorem 3.2, ℓ_1 -minimization will find this solution.

This example indicates that even if the existing sufficient conditions fail, the RSP of A^T at a vector may still be able to confirm the success of the ℓ_1 -method when solving an ℓ_0 -problem. To further understand the efficiency of the ℓ_1 -method, let us decompose the class of linear systems with nonnegative solutions, denoted by \mathcal{G} , into three subclasses. That is, $\mathcal{G} = \mathcal{G}_1 \cup \mathcal{G}_2 \cup \mathcal{G}_3$ where \mathcal{G}_i 's are defined as follows:

- \mathcal{G}_1 The system $Ax = b$ has a *unique* least ℓ_1 -norm nonnegative solution and a *unique* sparsest nonnegative solution.

\mathcal{G}_2 The system $Ax = b$ has a *unique* least ℓ_1 -norm nonnegative solution and *multiple* sparsest nonnegative solutions.

\mathcal{G}_3 The system $Ax = b$ has *multiple* least ℓ_1 -norm nonnegative solutions.

Clearly, every linear system with a nonnegative solution falls into one of these categories. Since many existing sufficient conditions (such as the mutual coherence, RIP and NSP) imply the strong equivalence between ℓ_0 - and ℓ_1 -problems, these conditions can apply only to (and explain the efficiency of the ℓ_1 -method only for) a subclass of linear systems in \mathcal{G}_1 . However, the RSP (8) defined in this paper goes beyond this scope of linear systems. An important feature of the RSP (8) is that it does not require a linear system to have a unique sparsest nonnegative solution in order to achieve the equivalence between ℓ_0 - and ℓ_1 -problems, as shown by the next example.

Example 3.4 (The ℓ_1 -method may successfully solve ℓ_0 -problems with multiple optimal solutions.) Consider the system $Ax = b$ with

$$A = \begin{pmatrix} 0.2 & 0 & -0.3 & -0.1 & 0.5 & -0.25 \\ 0 & 0.2 & 0.5 & 0.2 & -0.9 & 0.05 \\ 0.2 & 0 & -0.3 & -0.1 & 0.5 & -0.25 \end{pmatrix}, \quad b = \begin{pmatrix} 0.1 \\ -0.1 \\ 0.1 \end{pmatrix}.$$

For this example, it is easy to verify that $Ax = b$ has multiple sparsest nonnegative solutions:

$$x^{(1)} = (0, \frac{2}{5}, 0, 0, \frac{1}{5}, 0)^T, \quad x^{(2)} = (0, 0, 0, 4, 1, 0)^T, \quad x^{(3)} = (\frac{2}{9}, 0, 0, 0, \frac{1}{9}, 0)^T.$$

Since $\|x^{(1)}\|_1 > \|x^{(3)}\|_1$ and $\|x^{(2)}\|_1 > \|x^{(3)}\|_1$, by Theorem 3.2, the RSP of A^T is impossible to hold at $x^{(1)}$ and $x^{(2)}$. So we only need to check the RSP at $x^{(3)}$. Taking $y = (5, 5/3, 0)^T$ yields $\eta = A^T y = (1, 1/3, -2/3, -1/6, 1, -7/6)^T \in \mathcal{R}(A^T)$ where the first and fifth components are 1, and all others are strictly less than 1. Thus the RSP (8) holds at $x^{(3)}$, which (by Theorem 3.2) is the unique least ℓ_1 -norm nonnegative solution to the linear system. So the ℓ_1 -method solves the ℓ_0 -problem, although the ℓ_0 -problem has multiple optimal solutions.

The following corollary is an immediate consequence of Theorem 3.2, which claims that when an ℓ_0 -problem has multiple sparsest nonnegative optimal solutions, only one of them can satisfy the RSP of A^T .

Corollary 3.5 *For any underdetermined system of linear equations, there exists at most one sparsest nonnegative solution satisfying the RSP (8).*

Example 3.4 and Theorem 3.2 show that ℓ_0 - and ℓ_1 -problems can be equivalent provided that the RSP (8) is satisfied at an optimal solution to the ℓ_0 -problem, irrespective of the multiplicity of optimal solutions to the ℓ_0 -problem. This analysis indicates that the success of the ℓ_1 -method can happen not only for a subclass of linear systems in \mathcal{G}_1 , but also for a wide range of linear systems in \mathcal{G}_2 . Since many existing conditions imply the strong equivalence between ℓ_0 - and ℓ_1 -problems, they can only explain the success of ℓ_1 -methods when solving some ℓ_0 -problems in \mathcal{G}_1 . In other words, these strong equivalence conditions cannot apply to the ℓ_0 -problems in \mathcal{G}_2 which have multiple sparsest optimal solutions, and hence they cannot interpret the numerical success of the ℓ_1 -method in these situations. However, the RSP-based analysis has shown that

the success of the ℓ_1 -method may take place not only for those problems in \mathcal{G}_1 , but for a wide range of linear systems in \mathcal{G}_2 as well. So the ℓ_1 -method can solve a wider range of ℓ_0 -problems than what the strong equivalence theory can cope with. This does explain and clarify the gap between the performance of the ℓ_1 -method observed from numerical simulations and that indicated by existing strong equivalence conditions. Thus this analysis provides a certain answer to the question (b) in Sect. 1.

Remark 3.6 It is worth noting that our analysis method and results can be easily generalized to interpret the relationship between ℓ_0 - and weighted ℓ_1 -problems. More specifically, let us consider the weighted ℓ_1 -problem

$$\min\{\|Wx\|_1 : Ax = b, x \geq 0\}, \quad (17)$$

where $W = \text{diag}(w)$ with $w \in R_+^n$ and $w > 0$. By the nonsingular linear transformation, $u = Wx$, the above weighted ℓ_1 -problem is equivalent to

$$\min\{\|u\|_1 : (AW^{-1})u = b, u \geq 0\}. \quad (18)$$

Clearly, x is the unique optimal solution to the problem (17) if and only if $u = Wx$ is the unique optimal solution to the problem (18), and u and x have the same supports. Thus any weighted ℓ_1 -problem with weight $W = \text{diag}(w)$, where w is a positive vector in R_+^n , is nothing but a standard ℓ_1 -problem with a scaled matrix AW^{-1} . As a result, applying Theorems 2.7 to the problem (18), we conclude that u is the unique optimal solution to (18) if and only if $(AW^{-1})_{J_+(u)}$ has full column rank, and there exists a vector $\zeta \in \mathcal{R}((AW^{-1})^T)$ such that $\zeta_i = 1$ for $u_i > 0$ and $\zeta_i < 1$ for $u_i = 0$. By the one-to-one correspondence between the solutions of (17) and (18), and by transforming back to the weighted ℓ_1 -problem using $u = Wx$ and $\eta = W\zeta$, we immediately conclude that x is the unique optimal solution to the weighted ℓ_1 -problem (17) if and only if (i) A_{J_+} has full column rank where $J_+ = \{i : x_i > 0\}$, and (ii) there exists an $\eta \in \mathcal{R}(A^T)$ such that $\eta_i = w_i$ for $x_i > 0$, and $\eta_i < w_i$ for $x_i = 0$. We may call the above property (ii) the *weighted RSP of A^T at x* . Thus the results developed in this paper can be easily generalized to the weighted ℓ_1 -method for ℓ_0 -problems.

Remark 3.7 The RSP-based analysis and results can be also applied to the sparsest optimal solution of the linear program (LP)

$$\min\{c^T x : Ax = b, x \geq 0\}. \quad (19)$$

The sparsest optimal solution of (19) is meaningful. For instance, in production planning scenarios, the decision variables $x_i \geq 0$, $i = 1, \dots, n$, represent what production activities/events that should take place and how much resources should be allocated to them in order to achieve an optimal objective value. The sparsest optimal solution of a linear program provides the smallest number of activities to achieve the optimal objective value. In many situations, reducing the number of activities is vital for efficient planning, management and resource allocations. We denote by d^* the optimal value of (19), which can be obtained by solving the LP by simplex methods, or interior point methods. We assume that (19) is feasible and has a finite optimal

value d^* . Thus the optimal solution set of the LP is given by $\{x : Ax = b, x \geq 0, c^T x = d^*\}$. So a sparsest optimal solution to the LP is an optimal solution to the ℓ_0 -problem

$$\min \left\{ \|x\|_0 : \begin{pmatrix} A \\ c^T \end{pmatrix} x = \begin{pmatrix} b \\ d^* \end{pmatrix}, x \geq 0 \right\}, \quad (20)$$

associated with which is the ℓ_1 -problem

$$\min \left\{ \|x\|_1 : \begin{pmatrix} A \\ c^T \end{pmatrix} x = \begin{pmatrix} b \\ d^* \end{pmatrix}, x \geq 0 \right\}. \quad (21)$$

Therefore the developed results for sparsest nonnegative solutions of linear systems in this paper can be directly applied to (20) and (21). For instance, from Theorems 2.7 and 3.2, we immediately have the following statements: *x is the unique least ℓ_1 -norm optimal solution to the problem (19) if and only if $H = \begin{pmatrix} A_{J_+} \\ c_{J_+}^T \end{pmatrix}$ has full column rank, and there is a vector $\eta \in R^n$ satisfying*

$$\eta \in \mathcal{R}([A^T, c]), \eta_i = 1 \text{ for all } i \in J_+, \text{ and } \eta_i < 1 \text{ for all } i \notin J_+ \quad (22)$$

where $J_+ = \{i : x_i > 0\}$. Moreover, a sparsest optimal solution to the problem (19) is the unique least ℓ_1 -norm optimal solution to the LP if and only if the range space property (22) holds at this optimal solution. Note that a degenerated optimal solution has been long studied since 1950s (see [47–48]) and the references therein). It is well-known that finding a degenerated optimal solution requires extra effort than nondegenerated ones. Finding the most degenerated optimal solution or the sparsest optimal solution becomes even harder. The RSP-based analysis provides a new understanding for the most degenerated or the sparsest optimal solutions of LPs.

4 Application to compressed sensing

One of the tasks in compressed sensing is to exactly recover a sparse vector (representing a signal or an image) via an underdetermined system of linear equations [31, 33–35]. In this section, we consider the exact recovery of an unknown sparse nonnegative vector x^* by ℓ_1 -minimization. For this purpose, we assume that an $m \times n$ ($m < n$) sensing matrix A and the measurements $y = Ax^*$ are available. A nonnegative solution x of the system $Ax = b$ is said to have a guaranteed recovery (or to be exactly recovered) by ℓ_1 -minimization if x is the unique least ℓ_1 -norm nonnegative solution to the linear system. To guarantee the success of recovery, the current compressed sensing theory assumes that the matrix $A \in R^{m \times n}$ ($m < n$) satisfies some conditions (e.g., the RIP or the NSP of order $2K$) which imply the following properties: (i) x^* is the unique least ℓ_1 -norm nonnegative solution to the system $Ax = y = Ax^*$ (where the components of y are measurements); (ii) x^* is the unique sparsest nonnegative solution to the system $Ax = y$. So the underlying ℓ_0 - and ℓ_1 -problems must be strongly equivalent. Most of the recovering conditions developed so far are for the so-called uniform recovery.

4.1 Uniform recovery of sparse nonnegative vectors

The exact recovery of all K -sparse nonnegative vectors (i.e., $\{x : x \geq 0, \|x\|_0 \leq K\}$) by a single sensing matrix A is called the *uniform recovery* of K -sparse nonnegative vectors. To develop a RSP-based recovery theory, let us first introduce the following concept.

Definition 4.1 (RSP of order K). Let A be an $m \times n$ matrix with $m < n$. A^T is said to satisfy the range space property of order K if for any subset $S \subseteq \{1, \dots, n\}$ with $|S| \leq K$, $\mathcal{R}(A^T)$ contains a vector η such that $\eta_i = 1$ for all $i \in S$, and $\eta_i < 1$ for all $i \in S_c = \{1, 2, \dots, n\} \setminus S$.

We first show that if A^T has the RSP of order K , then K must be bounded by the spark of A , denoted by $\text{Spark}(A)$, which is the smallest number of columns of A that are linearly dependent (see, e.g., [30, 38]).

Lemma 4.2 If A^T has the RSP of order K , then any K columns of A are linearly independent, so $K < \text{Spark}(A)$.

Proof. Let $S = \{s_1, \dots, s_K\}$, with $|S| = K$, be an arbitrary subset of $\{1, \dots, n\}$. Suppose that A^T has the RSP of order K . We now prove that A_S has full column rank. It is sufficient to show that $z_S = 0$ is the only solution to $A_S z_S = 0$. Indeed, let $A_S z_S = 0$. Then $z = (z_S, z_{S_c} = 0) \in \mathbb{R}^n$ is in the null space of A . By the RSP of order K , there exists a vector $\eta \in \mathcal{R}(A^T)$ such that every component of η_S is 1, i.e., $\eta_{s_i} = 1$ for $i = 1, \dots, K$. By the orthogonality of the null and range spaces, we have

$$z_{s_1} + z_{s_2} + \dots + z_{s_K} = z_S^T \eta_S = z^T \eta = 0. \quad (23)$$

Now let k be an arbitrary number with $1 \leq k \leq K$, and $S_k = \{s_1, s_2, \dots, s_k\} \subseteq S$. Since $|S_k| \leq |S| = K$, it follows from the definition of the RSP of order K , there exists a vector $\tilde{\eta} \in \mathcal{R}(A^T)$ with $\tilde{\eta}_{s_i} = 1$ for every $i = 1, \dots, k$ and $\tilde{\eta}_j < 1$ for every $j \notin S_k$. By the orthogonality again, it follows from $z^T \tilde{\eta} = 0$ that

$$(z_{s_1} + \dots + z_{s_k}) + (\tilde{\eta}_{s_{k+1}} z_{s_{k+1}} + \dots + \tilde{\eta}_{s_K} z_{s_K}) = 0.$$

This is equivalent to

$$(z_{s_1} + \dots + z_{s_k}) + (z_{s_{k+1}} + \dots + z_{s_K}) + [z_{s_{k+1}}(\tilde{\eta}_{s_{k+1}} - 1) + \dots + z_{s_K}(\tilde{\eta}_{s_K} - 1)] = 0$$

which, together with (23), implies that

$$(\tilde{\eta}_{s_{k+1}} - 1)z_{s_{k+1}} + \dots + (\tilde{\eta}_{s_K} - 1)z_{s_K} = 0$$

where $\tilde{\eta}_{s_i} < 1$ for $i = k+1, \dots, K$. Since such relations hold for every specified k with $1 \leq k \leq K$. In particular, for $k = K - 1$, the relation above is reduced to $(\tilde{\eta}_{s_K} - 1)z_{s_K} = 0$ which implies that $z_{s_K} = 0$ since $\tilde{\eta}_{s_K} < 1$. For $k = K - 2$, the relation above is of the form

$$(\tilde{\eta}_{s_{K-1}} - 1)z_{s_{K-1}} + (\tilde{\eta}_{s_K} - 1)z_{s_K} = 0$$

which, together with $z_{s_K} = 0$ and $\tilde{\eta}_{s_{K-1}} < 1$, implies that $z_{s_{K-1}} = 0$. Continuing this process by considering $k = K - 3, \dots, 1$, we deduce that all components of z_S are zero. Thus A_S has full column rank. By the definition of $\text{Spark}(A)$, we must have $K < \text{Spark}(A)$. \square

The RSP of order K can completely characterize the uniform recovery of all K -sparse non-negative vectors by ℓ_1 -minimization, as shown by the next result.

Theorem 4.3 Let the measurements of the form $y = Ax$ be taken. Then any $x \geq 0$ with $\|x\|_0 \leq K$ can be exactly recovered by the ℓ_1 -method (i.e., $\min\{\|z\|_1 : Az = y, z \geq 0\}$) if and only if A^T has the RSP of order K .

Proof. Assume that the RSP of order K is satisfied. Let $x^* \geq 0$ be an arbitrary vector with $\|x^*\|_0 \leq K$. Let $S = J_+ = \{i : x_i^* > 0\}$. Since $|S| = \|x^*\|_0 \leq K$, by the RSP of order K , there exists a vector $\eta \in \mathcal{R}(A^T)$ such that $\eta_i = 1$ for all $i \in S$, and $\eta_i < 1$ for all $i \in S_c$. This implies that the RSP (8) holds at $x^* \geq 0$. Moreover, it follows from Lemma 4.2 that A_S has full column rank. Hence, by Theorem 2.7, x^* is the unique least ℓ_1 -norm nonnegative solution to the system $Ax = y = Ax^*$. So x^* can be exactly recovered by the ℓ_1 -method.

Conversely, assume that any $x \geq 0$ with $\|x\|_0 \leq K$ can be exactly recovered by the ℓ_1 -method. We now prove that the RSP of order K must be satisfied. Let $S = J_+ = \{i : x_i > 0\}$. Under the assumption, x is the unique optimal solution to the ℓ_1 -problem

$$\min\{\|z\|_1 : Az = y = Ax, z \geq 0\}.$$

By Theorem 2.7, the RSP (8) holds at x , i.e., there exists a vector $\eta \in \mathcal{R}(A^T)$ such that $\eta_i = 1$ for all $i \in S = J_+$, and $\eta_i < 1$ otherwise. Since x can be any K -sparse nonnegative vectors, this implies that $S = J_+$ can be any subset of $\{1, \dots, n\}$ with $|S| \leq K$, and for every such a subset there exists accordingly a vector η satisfying the above property. By Definition 4.1, A^T has the RSP of order K . \square

Let a_j , $1 \leq j \leq n$, be the columns of A and let $a_0 = 0$. Let P denote the convex hull of a_j , $0 \leq j \leq n$. Donoho and Tanner [4] introduced the following concept: The polytope P is *outwardly K -neighborly* if every subset of K vertices not including $a_0 = 0$ spans a face of this polytope. They have shown that the polytope P is outwardly K -neighborly if and only if any nonnegative solution x to the system $Ax = b$ with $\|x\|_0 \leq K$ is the unique optimal solution to the ℓ_1 -problem. In other words, the outwardly K -neighborly property is a full geometric characterization of the uniform recovery of all K -sparse nonnegative vectors. Recently, Khajehnejad et al [6] characterized the uniform recovery by using the property of $\mathcal{N}(A)$, the null space A . They have showed that all nonnegative K -sparse vector can be exactly recovered if and only if for every vector $w \neq 0$ in $\mathcal{N}(A)$, and every index set $S \subseteq \{1, \dots, n\}$ with $|S| = K$ such that $w_{S_c} \geq 0$, it holds that $e^T w > 0$. As shown by Theorem 4.3, the RSP of order K introduced in this section provides an alternative full characterization of the uniform recovery of all K -sparse vectors. Clearly, from different perspectives, all the above-mentioned properties (outwardly K -neighborly, null space, and range space of order K) are equivalent since each of these properties is a necessary and sufficient condition for (i.e., equivalent to) the uniform recovery of all K -sparse vectors. As a result, if a matrix satisfies one of these conditions, it also satisfies the other two conditions. However, directly checking these conditions is difficult. Checking the outwardly k -property needs to check all possible K out of n vertices and verify whether all such subsets can span a face of the polytope defined by A . To check the null space property, we need to to check all possible vectors in the null space of A , and for every such a vector, we need to verify the required property for all possible K components of the vector. Similarly, for the range space property, we need to verify the individual RSP holds at every subset $S \subset \{1, 2, \dots, n\}$ with cardinality $|S| \leq K$. Clearly, none of these properties is easier to check than the others for general instances of matrices.

We now close this section by stressing the difference between the RSP of order K and the RSP (8). Such a difference can be easily seen from the following result.

Corollary 4.4 *If A^T has the RSP of order K , then any $\hat{x} \geq 0$ with $\|\hat{x}\|_0 \leq K$ is both the unique least ℓ_1 -norm nonnegative solution and the unique sparsest nonnegative solution to the linear system $Ax = y = A\hat{x}$.*

Proof. By Theorem 4.3, under the RSP of order K , any $\hat{x} \geq 0$ with $\|\hat{x}\|_0 \leq K$ can be exactly recovered by ℓ_1 -minimization, i.e., \hat{x} is the unique least ℓ_1 -norm nonnegative solution to the system $Ax = y = A\hat{x}$. We now prove that \hat{x} is also the sparsest nonnegative solution to this system. Assume that there exists another solution $z \geq 0$ such that $\|z\|_0 \leq \|\hat{x}\|_0$. Let $S = \{i : z_i > 0\}$. Since $|S| = \|z\|_0 \leq \|\hat{x}\|_0 \leq K$, by the RSP of order K , there exists an $\eta \in \mathcal{R}(A^T)$ such that $\eta_i = 1$ for all $i \in S$, and $\eta_i < 1$ for all $i \in S_c$. Thus the individual RSP (8) holds at z . By Lemma 4.2, any K columns of A are linearly independent. Since the number of the columns of A_S , where $S = \{i : z_i > 0\}$, is less than K , this implies that A_S has full column rank. By Theorem 2.7, z is also the unique least ℓ_1 -norm nonnegative solution to the system $Ax = y = A\hat{x}$. Thus $z = \hat{x}$, which implies that \hat{x} is the unique sparsest nonnegative solution to this system. \square

This result shows that the RSP of order K is more restrictive than the individual RSP (8) which is defined at a single point. The former requires that the RSP (8) hold at every K -sparse nonnegative solution. By contrast, the individual RSP (8) is only a local property, and it does not imply that the underlying linear system has a unique sparsest nonnegative solution, as we have shown in Sect. 3.

4.2 Non-uniform recovery of sparse nonnegative vectors

The purpose of the uniform recovery is to recover all k -sparse vectors. So some strong assumptions (such as the RIP, NSP and RSP of certain orders) must be imposed on the matrix. These strong assumptions imply that the unknown sparse vector x must be the unique optimal solution to both ℓ_0 - and ℓ_1 -problems (hence, the strong equivalence between these two problems are required by the uniform recovery). In this subsection, we extend the uniform recovery theory to non-uniform ones by using the RSP. The non-uniform recovery of sparse signals has been investigated by some researchers (see e.g., [4, 49]). From a geometric perspective, Donoho and Tanner [4] introduced the so-called weak neighborliness conditions for the non-uniform recovery by ℓ_1 -minimization, and they have shown under such a condition that most nonnegative K -sparse vectors can be exactly recovered by the ℓ_1 -method. Ayaz and Rauhut [49] focused on the non-uniform recovery of signals with given sparsity and given signal length by ℓ_1 -minimization, and they have provided the number of samples required to recover such signals with gaussian and subgaussian random matrices. In what follows, we introduce the so-called weak RSP of order K , a range space property of A^T , which provides a non-uniform recovery condition for some vectors which may have high sparsity level, going beyond the scope of normal uniform recoveries.

Given a sensing matrix A , Theorem 2.7 claims that a vector x^* can be exactly recovered by ℓ_1 -minimization provided that the RSP(8) hold at x^* and that the matrix A_{J_+} , where $J_+ = \{i : x_i^* > 0\}$, has full-column rank. Such an x^* is not necessarily the unique sparsest nonnegative solution to the linear system as shown by Example 3.4, and it may not even be a sparsest

nonnegative solution as well. For instance, let

$$A = \begin{pmatrix} 6 & 4 & 1.5 & 4 & -1 \\ 6 & 4 & -0.5 & 4 & 0 \\ 0 & -2 & 31.5 & -1 & -1.5 \end{pmatrix}, \quad y = \begin{pmatrix} 4 \\ 4 \\ -1 \end{pmatrix} = Ax^*$$

where $x^* = (1/3, 1/2, 0, 0, 0)^T$. It is easy to see that $\tilde{x} = (0, 0, 0, 1, 0)^T$ is the unique sparsest nonnegative solution to the system $Ax = y$, while x^* is the unique least ℓ_1 -norm nonnegative solution to the system $Ax = y$. Although x^* is not the sparsest nonnegative solution, it can be exactly recovered by the ℓ_1 -method. Because of this, it is interesting to develop a recovery theory without requiring that the targeted unknown sparse vector be a sparsest or be the unique sparsest solution to a linear system. This is also motivated by some practical applications. In fact, a real sparse signal may not be sparse enough to be recovered by the uniform recovery, and partial information for the unknown sparse vector may be available in some situations, for example, the support of an unknown vector may be known. The concept of the RSP of order K can be adapted to handle these cases. So we introduce the following concept.

Definition 4.5 (WRSP of order K) *Let A be an $m \times n$ matrix with $m < n$. A^T is said to satisfy the weak range space property (WRSP) of order K if the following two properties are satisfied:*

- (i) *There exists a subset $S \subseteq \{1, \dots, n\}$ such that $|S| = K$ and A_S has full column rank.*
- (ii) *For any subset $S \subseteq \{1, \dots, n\}$ such that $|S| \leq K$ and A_S has full column rank, the space $\mathcal{R}(A^T)$ contains a vector η such that $\eta_i = 1$ for $i \in S$, and $\eta_i < 1$ otherwise.*

The WRSP of order K only requires that the individual RSP hold for those subsets $S \subseteq \{1, \dots, n\}$ with $|S| \leq K$ and that A_S be full-column-rank, while the RSP of order K requires that the individual RSP hold for any subset $S \subseteq \{1, \dots, n\}$ with $|S| \leq K$. So the WRSP of order K is less restrictive than the RSP of order K . By Theorem 2.6, we have the following result.

Theorem 4.6 *Let the measurements of the form $y = Ax$ be taken. Suppose that there exists a subset $S \subseteq \{1, \dots, n\}$ such that $|S| = K$ and A_S has full column rank. Then A^T has the WRSP of order K if and only if any $x \geq 0$, satisfying that $\|x\|_0 \leq K$ and A_{J_+} has full-column-rank where $J_+ = \{i : x_i > 0\}$, can be exactly recovered by the ℓ_1 -minimization $\min\{\|z\|_1 : Az = y = Ax, z \geq 0\}$.*

Proof. Assume that A^T has the WRSP of order K . Let x be an arbitrary nonnegative vector such that $\|x\|_0 \leq K$ and A_{J_+} has full-column-rank, and let $S = J_+ = \{i : x_i > 0\}$. Since A^T has the WRSP of order K , there exists an $\eta \in \mathcal{R}(A^T)$ such that $\eta_i = 1$ for $i \in S = J_+$, and $\eta_i < 1$ otherwise. This implies that the RSP(8) holds at x . Since A_{J_+} has full column rank, by Theorem 2.7, x must be the unique least ℓ_1 -norm nonnegative solution to the system $Az = y$ ($= Ax$). In other words, x can be recovered by ℓ_1 -minimization. Conversely, we assume that any $x \geq 0$, satisfying that $\|x\|_0 \leq K$ and A_{J_+} has full-column-rank, can be exactly recovered by ℓ_1 -minimization. We now prove that A^T must have the WRSP of order K . In fact, let $x \geq 0$ be a vector such that $\|x\|_0 \leq K$ and A_{J_+} has full-column-rank. Denote by $S = J_+ = \{i : x_i > 0\}$. Since x can be recovered by the ℓ_1 -method, it is the unique least ℓ_1 -norm nonnegative solution to the system $Az = y = Ax$. By Theorem 2.7, the RSP (8) holds at x , i.e., there exists an

$\eta \in \mathcal{R}(A^T)$ such that $\eta_i = 1$ for $i \in J_+ = S$, and $\eta_i < 1$ otherwise. Since x can be any vector such that $\|x\|_0 \leq K$ and A_{J_+} has full column rank, this implies that the condition (ii) of Definition 4.5 holds, thus A^T has the WRSP of order K . \square

We may further relax the concept of the RSP and WRSP, especially when partial information available to the unknown vector. For instance, when $\|x\|_0 = K$ is known, we may introduce the next two concepts.

Definition 4.7 (PRSP of order K). *We say that A^T has the partial range space property (PRSP) of order K if for any subset S of $\{1, \dots, n\}$ with $|S| = K$, the range space $\mathcal{R}(A^T)$ contains a vector η such that $\eta_i = 1$ for all $i \in S$, and $\eta_i < 1$ otherwise.*

Definition 4.8 (PWRSP of order K). *A^T is said to have partial weak range space property (PWRSP) of order K if for any subset $S \subseteq \{1, \dots, n\}$ such that $|S| = K$ and A_S has full column rank, $\mathcal{R}(A^T)$ contains a vector η such that $\eta_i = 1$ for all $i \in S$, and $\eta_i < 1$ otherwise.*

Different from the RSP of order K , the PRSP of order K only requires that the individual RSP hold for the subset S with $|S| = K$. Similarly, the PWRSP of order K is also less restrictive than WRSP. Based on such definitions, we have the next result which follows from Theorem 2.7 straightaway.

Theorem 4.9 (i) *The matrix A^T has the partial range space property (PRSP) of order K if and only if any $x \geq 0$, with $\|x\|_0 = K$, can be exactly recovered by the ℓ_1 -minimization $\min\{\|z\|_1 : Az = y = Ax, z \geq 0\}$.*

(ii) *A^T has the PWRSP of order K if and only if any $x \geq 0$, satisfying that $\|x\|_0 = K$ and A_{J_+} has full-column-rank where $J_+ = \{i : x_i > 0\}$, can be exactly recovered by the ℓ_1 -minimization $\min\{\|z\|_1 : Az = y = Ax, z \geq 0\}$.*

When A_S has full column rank, we have $|S| \leq m$. Thus the WRSP and PWRSP of order K imply that $K \leq m$. Moreover, the PRSP of order K implies that $K < \text{Spark}(A)$. In fact, the proof of this fact is identical to that of Lemma 4.1. Theorems 4.6 and 4.9(ii) indicate that a portion of vectors with $\|x\|_0 \leq m$ can be recovered if a sensing matrix satisfies certain properties milder than the RSP of order K (and thus milder than the RIP and the NSP of order $2K$). Since the PRSP, WRSP and PWRSP of order K do not require that the individual RSP hold for all subsets S with $|S| \leq K$, by Theorem 4.3, these properties are non-uniform recovering conditions developed through certain range space properties of A^T .

5 Conclusions

Through the range space property, we have characterized the conditions for an ℓ_1 -problem to have a unique optimal solution, and for K -sparse vectors to be uniformly or non-uniformly recovered by the ℓ_1 -method. We have shown the following main results: (i) A nonnegative vector is the unique optimal solution to the ℓ_1 -problem if and only if the RSP holds at this vector, and the associated submatrix A_{J_+} has full column rank; (ii) All K -sparse vectors can be exactly recovered by a single sensing matrix if and only if the transpose of this matrix has the RSP of order K ; (iii) ℓ_0 - and ℓ_1 -problems are equivalent if and only if the RSP holds at an optimal solu-

tion of the ℓ_0 -problem. The RSP originates naturally from the strict complementarity property of linear programming problems. The RSP-based analysis has indicated that the uniqueness of optimal solutions of ℓ_0 -problems is not the reason for the problem being computationally tractable, and the multiplicity of optimal solutions of ℓ_0 -problems is also not the reason for the problem being hard. The RSP may hold in both situations, and hence an ℓ_0 -problem can be successfully solved by the ℓ_1 -method in both situations, provided that the RSP is satisfied at a solution of the ℓ_0 -problem. Thus the relationship between ℓ_0 - and ℓ_1 -problems and the numerical performance of the ℓ_1 -method can be broadly interpreted via the RSP-based analysis.

References

1. Chen, D., Plemmons, R.: Nonnegativity constraints in numerical analysis. Symposium on the Birth of Numerical Analysis (2007)
2. Bardsley, J., Nagy, J.: Covariance-preconditioned iterative methods for nonnegativity constrained astronomical imaging. *SIAM J. Matrix Anal. Appl.* **27**, 1184–1198 (2006)
3. Bruckstein, A.M., Elad, M., Zibulevsky, M.: On the uniqueness of nonnegative sparse solutions to underdetermined systems of equations. *IEEE Trans. Inform. Theory* **54**, 4813–4820 (2008)
4. Donoho, D.L., Tanner, J.: Sparse nonnegative solutions of underdetermined linear equations by linear programming. *Proc. Natl. Acad. Sci.* **102**, 9446–9451 (2005)
5. Donoho, D.L., Tanner, J.: Counting the faces of randomly-projected hypercubes and orthants with applications. *Discrete Comput. Geom.* **43**, 522–541 (2010)
6. Khajehnejad, M.A., Dimakis, A.G., Xu, W., Hassibi, B.: Sparse recovery of nonnegative signals with minima expansion. *IEEE Trans. Signal Proc.* **59**, 196–208 (2011)
7. O’Grady, P.D., Rickard, S.T.: Recovery of nonnegative signals from compressively samples observations via nonnegative quadratic programming. SPASR’09 (2009)
8. Slawski, M., Hein, M.: Sparse recovery by thresholded nonnegative least squares. Technical Report, Saarland University, Germany (2011)
9. Vo, N., Moran, B., Challa, S.: Nonnegative-least-square classifier for face recognition. Proc. of the 6th Inter. Symposium on Neural Networks: Advances in Neural Networks, 449-456 (2009)
10. Wang, M., Xu, W., Tang, A.: A unique nonnegative solution to an underdetermined system: from vectors to matrices. *IEEE Trans. Signal Proc.* **59**, 1107–1016 (2011)
11. Bradley, P.S., Mangasarian, O.L., Rosen, J.B.: Parsimonious least norm approximation. *Comput. Optim. Appl.* **11**, 5–21 (1998)
12. Bradley, P.S., Fayyad, U.M., Mangasarian, O.L.: Mathematical programming for data mining: formulations and challenges. *INFORMS J. Computing* **11**, 217–238 (1999)

13. He, R., Zheng, W., Hu, B., Kong, X.: Nonnegative sparse coding for discriminative semi-supervised learning. Proc. of IEEE conference on CVPR, 2849–2856 (2011)
14. Mangasarian, O.L.: Minimum-support solutions of polyhedral concave programs. Optimization **45**, 149–162 (1999)
15. Mangasarian, O.L., Machine learning via polyhedral concave minimization. Applied Mathematics and Parallel Computing-Festschrift for Klaus Ritter (H. Fischer, B. Riedmueller and S. Schaeffler eds.), Springer, Heidelberg, 175–188 (1996)
16. Szlam, A., Guo, Z., Osher, S.: A split Bregman method for nonnegative sparsity penalized least squares with applications to hyperspectral demixing. IEEE Inter. Conf. on Imaging Processing, 1917–1920 (2010)
17. Slawski, M., Hein, M.: Sparse recovery for protein mass spectrometry data. Technical Report, Saarland University, Germany (2010)
18. Hoyer, P.O.: Nonnegative sparse coding. Proc. of the 12th IEEE Workshop, 557–565 (2002)
19. Wang, F., Li, P.: Compressed nonnegative sparse coding. IEEE 10th Int. Conf. on Data Mining (ICDM), 1103–1108 (2010)
20. Recht, B., Fazel, M., Parrilo, P.A.: Guaranteed minimum-rank solutions of linear matrix equations via nuclear norm minimization. SIAM Rev. **52**, 471–501 (2010)
21. Zhao, Y.B.: An approximation theory of matrix rank minimization and its application to quadratic equations. Linear Algebra Appl. **437**, 77–93 (2012)
22. Lee, D., Seung, H.: Learning the parts of objects by nonnegative matrix factorization. Nature **401**, 788–791 (1999)
23. Park, H., Kim, H.: Nonnegative matrix factorization based on alternating nonnegativity constrained least squares and the active set method. SIAM J. Matrix Anal. Appl. **30**, 713–730 (2008)
24. Pauca, P., Piper, J., Plemmons, R.: Nonnegative matrix factorization for spectral data analysis. Linear Algebra Appl. **416**, 29–47 (2006)
25. Natarajan, B.K.: Sparse approximate solutions to linear systems. SIAM J. Comput. **24**, 227–234 (1995)
26. Candès, E., Wakin, M.B., Boyd, S.P.: Enhancing sparsity by reweighted ℓ_1 minimization. J. Fourier Anal. Appl. **14**, 877–905 (2008)
27. Zhao, Y.B., Li, D.: Reweighted ℓ_1 -minimization for sparse solutions to underdetermined linear systems. SIAM J. Optim. **22**, 1065–1088 (2012)

28. Wen, Z., Yin, W., Goldfarb, D., Zhang, Y.: A fast algorithm for sparse reconstruction based on shrinkage, subspace optimization, and continuation. *SIAM J. Sci. Comput.* **32**, 1832–1857 (2010)
29. Yin, W., Osher, S., Goldfarb, D., Darbon, J.: Bregman iterative algorithms for ℓ_1 minimization with applications to compressed sensing. *SIAM J. Imaging Sci.* **1**, 143–168 (2008)
30. Bruckstein, A.M., Donoho, D.L., Elad, M.: From sparse solutions of systems of equations to sparse modeling of signals and images. *SIAM Rev.* **51**, 34–81 (2009)
31. Candès, E., *Compressive sampling*. Proceedings of the International Congress of Mathematicians, Madrid, Spain (2006)
32. Candès, E., Romberg, J., Tao, T.: Robust uncertainty principles: Exact reconstruction from highly incomplete frequency information. *IEEE Trans. Inform. Theory* **52**, 489–509 (2006)
33. Donoho, D. L.: Compressed sensing. *IEEE Trans. Inform. Theory* **52**, 1289–1306 (2006)
34. Elad, M.: *Sparse and Redundant Representations: From Theory to Applications in Signal and Image Processing*. Springer, New York (2010)
35. Candès, E.: The restricted isometry property and its implications for compressed sensing. *Comptes Rendus de l’Académie des Sciences, Paris, Série I*, **346**, 589–592 (2008)
36. Candès, E., Tao, T.: Decoding by linear programming. *IEEE Trans. Inform. Theory* **51**, 4203–4215 (2005)
37. Daubechies, I., DeVore, R., Fornasier, M., Güntürk, S.: Iteratively reweighted squares minimization for sparse recovery. *Commun. Pure. Appl. Math.* **63**, 1-38 (2010)
38. Donoho, D.L., Elad, M., Optimality sparse representation in general (non-orthogonal) dictionaries via ℓ_1 minimization. *Proc. Natl. Acad. Sci.* **100**, 2197–2202 (2003)
39. Elad, M., Bruckstein, A.: A generalized uncertainty principle and sparse representation in pairs of bases. *IEEE Trans. Inform. Theory* **48**, 2558–2567 (2002)
40. Feuer, A., Nemirovsky, A.: On sparse representation in pairs of bases. *IEEE Trans. Inform. Theory* **49**, 1579–1581 (2003)
41. Fuchs, J.J.: On sparse representations in arbitrary redundant bases. *IEEE Trans. Inform. Theory* **50**, 1341–1344 (2004)
42. Tropp, J.A.: Greed is good: Algorithmic results for sparse approximation. *IEEE Trans. Inform. Theory* **50**, 2231–2242 (2004)
43. Juditski, A., Nemirovski, A.: On verifiable sufficient conditions for sparse signal recovery via ℓ_1 minimization. *Math. Program., Ser. B* **127**, 57–88 (2011)

44. Juditsky, A., Kilinc Karzan, F., Nemirovski, A.: Verifiable conditions of ℓ_1 -recovery for sparse signals with sign restrictions. *Math. Program. Ser. B* **127**, 89–122 (2011)
45. Gribonval, R., Nielsen, M.: Sparse decompositions in unions of bases. *IEEE Trans. Inform. Theory* **49**, 3320–3325 (2003)
46. Schrijver, A.: *Theory of Linear and Integer Programming*. Wiley, New York (1986)
47. Greenberg, H.J.: An analysis of degeneracy. *Naval Res. Logist. Quart.* **33**, 635–655 (1986)
48. Greenberg, H.J.: The use of the optimal partition in a linear programming solution for postoptimal analysis. *Oper. Res. Lett.* **15**, 179–185 (1994)
49. Ayaz, U., Rauhut, H.: Nonuniform sparse recovery with subgaussian matrices. Technical Report, University of Bonn, Germany (2011)