Necessary and Sufficient Conditions of Solution Uniqueness in 1-Norm Minimization

Hui Zhang · Wotao Yin · Lizhi Cheng

Received: 18 September 2012 / Accepted: 3 May 2014 / Published online: 26 August 2014
© Springer Science+Business Media New York 2014

Abstract This paper shows that the solutions to various 1-norm minimization problems are unique if, and only if, a common set of conditions are satisfied. This result applies broadly to the basis pursuit model, basis pursuit denoising model, Lasso model, as well as certain other 1-norm related models. This condition is previously known to be sufficient for the basis pursuit model to have a unique solution. Indeed, it is also necessary, and applies to a variety of 1-norm related models. The paper also discusses ways to recognize unique solutions and verify the uniqueness conditions numerically. The proof technique is based on linear programming strong duality and strict complementarity results.

Keywords l1 minimization · Basis pursuit · Lasso · Solution uniqueness · Strict complementarity

Mathematics Subject Classification 65K05 · 90C25

1 Introduction

There is a rich literature on analyzing, solving, and applying 1-norm minimization in the context of information theory, signal processing, statistics, machine learning, and
optimization. We are interested in when a 1-norm minimization problem has a unique solution, which is undoubtedly one of the very first questions toward any inverse problem. In compressive sensing signal recovery, having non-unique solutions means that the underlying signal cannot be reliably recovered from its measurements. In feature selection, non-unique solutions cause ambiguous selections, and thus other criteria are needed. In addition, a number of optimization methods and algorithms, in particular, those producing the solution path by varying the model parameters such as least angle regression (LASSO) [1] and parametric quadratic programming [2], can fail (or require special treatments) upon encountering solution non-uniqueness on the path. Therefore, establishing a condition for solution uniqueness is important for both the analysis and computation of 1-norm minimization.

Various sufficient conditions have been given to guarantee solution uniqueness. They include Spark [3,4], the mutual incoherence condition [5,6], the null-space property (NSP) [7,8], the restricted isometry principle (RIP) [9], the spherical section property [10], the “RIPless” property [11], and so on. Some conditions guarantee not only a unique solution but also the solution equals the original signal, provided that the signal has sufficiently few nonzero entries; this is called uniform recovery, that is, uniform over all sufficiently sparse signals. Other conditions guarantee that the unique solution equals just one given original signal, or any original signal whose signs are fixed according to the conditions; these are called non-uniform recovery. Being it uniform or not, none of them is known to be both necessary and sufficient for recovering a given solution. This paper shows that given a solution [to any one of the problems (1a)–(1d) below], our Condition 2.1 below is both necessary and sufficient for guaranteeing recovering that solution uniquely. Our condition is weaker than the sufficient conditions mentioned above, but it does not provide any uniform guarantee. We also discuss ways how to recognize a unique solution and how to verify our condition numerically.

Our proof is based on linear programming strong duality: if a linear program has a solution, so does its dual, and any primal and dual solutions must give the same objective value; in addition, there must exist a pair of strictly complementary primal and dual solutions (see textbook [12], for example). In Sect. 4, we reduce the so-called basis pursuit problem to a linear program and apply these results to establish the necessity part of our condition.

The rest of the paper is organized as follows. Section 2 states the main results of this paper. Section 3 reviews several related results. Proofs for the main results are given in Sect. 4. Section 5 discusses condition verification.

### 2 Main Results

Let $x \in \mathbb{R}^n$. Let its $\ell_1$-norm and $\ell_2$-norm be defined as $\|x\|_1 := \sum_{i=1}^n |x_i|$ and $\|x\|_2 := (\sum_{i=1}^n |x_i|^2)^{1/2}$. We study the solution uniqueness conditions for the problems of minimizing $\|x\|_1$ including the basis pursuit problem [13]

$$\min \|x\|_1 \quad \text{s.t. } Ax = b,$$

(1a)
as well as the following convex programs:

\[
\begin{align*}
\text{(1b)} & \quad \min [f_1(Ax - b) + \lambda \|x\|_1], \\
\text{(1c)} & \quad \min \|x\|_1, \quad \text{s.t.} \quad f_2(Ax - b) \leq \sigma, \\
\text{(1d)} & \quad \min f_3(Ax - b), \quad \text{s.t.} \quad \|x\|_1 \leq \tau,
\end{align*}
\]

where \(\lambda, \sigma, \tau > 0\) are scalar parameters, \(A\) is a matrix, and \(f_i(x), i = 1, 2, 3\) are strictly convex functions. LASSO [14] is a special case of problem (1b), while basis pursuit denoising [13] is a special case of problem (1c), all with \(f_i(\cdot) = \frac{1}{2}\|\cdot\|_2^2, i = 1, 2, 3\).

In general, any problem (1a)–(1d) can have more than one solution.

Let \(X, X_\lambda, Y_\sigma, Z_\tau\) denote the sets of solutions to problems (1a)–(1d), respectively. Let \(a_i\) be the \(i\)th column of \(A\) and \(x_i\) be the \(i\)th entry of \(x\). Given an index set \(I\), we frequently use \(A_I\) as the submatrix of \(A\) formed by its columns \(a_i, i \in I\), and use \(x_I\) as the subvector of \(x\) formed by entries \(x_i, i \in I\).

Our analysis makes the following assumptions:

**Assumption 2.1** Matrix \(A\) has full row rank.

**Assumption 2.2** The solution sets \(X, X_\lambda, Y_\sigma, Z_\tau\) of problems (1a)–(1d), respectively, are nonempty.

**Assumption 2.3** In problems (1b)–(1d), functions \(f_1, f_2, f_3\) are strictly convex. In addition, the constraint of problem (1d) is bounding, namely, \(\tau\) is less than or equal to \(\inf\{\|x\|_1 : f_3(Ax - b) = f_3^*\}\), where \(f_3^* := \min_{y \in \mathbb{R}^n} f(Ay - b)\).

Assumptions 2.1 and 2.2 are standard. If Assumption 2.1 does not hold and \(Ax = b\) is consistent, then the problems can be simplified; specifically, one can decompose \(A = \begin{bmatrix} A_1 \\ A_2 \end{bmatrix}\) and \(b = \begin{bmatrix} b_1 \\ b_2 \end{bmatrix}\), so that and \(A_1\) has full row rank equal to rank \((A)\), and one can replace the constraints \(Ax = b\) by \(A_1x = b_1\) and introduce functions \(\tilde{f}_i\) so that \(\tilde{f}_i(A_1x - b_1) \equiv f_i(Ax - b), i = 1, 2, 3\). Assumption 2.2 guarantees that the solutions of problems (1a)–(1d) can be attained, so the discussion of solution uniqueness makes sense. The strict convexity of \(f_1, f_2, f_3\) and the restriction on \(\tau\) in Assumption 2.3 are quite basic for solution uniqueness. Strict convexity rules out piece-wise linearity. (Note that \(f_1, f_2, f_3\) are not necessarily differentiable.) If the restriction on \(\tau\) is removed, the solution uniqueness of problem (1d) becomes solely up to \(f_3(Ax - b)\), rather than \(\|x\|_1\).

For a given vector \(x^*\), solution uniqueness is determined by the following conditions imposed on matrix \(A\), and the sufficiency of these conditions has been established in [15]. Define \(\text{supp}(x^*) := \{i \in \{1, \ldots, n\} : x_i^* \neq 0\}\).

**Condition 2.1** Under the definitions \(I := \text{supp}(x^*)\) and \(s := \text{sign}(x_I^*)\), matrix \(A \in \mathbb{R}^{m \times n}\) has the following properties:
1. submatrix $A_I$ has full column rank, and
2. there is $y \in \mathbb{R}^m$ obeying $A_I^T y = s$ and $\|A_I^T y\|_{\infty} < 1$.

The main theorem of this paper asserts that Condition 2.1 is both necessary and sufficient to the uniqueness of solution $x^*$.

**Theorem 2.1** (Solution uniqueness) *Under Assumptions 2.1–2.3, given that $x^*$ is a solution to problem (1a), (1b), (1c), or (1d), $x^*$ is the unique solution if and only if Condition 2.1 holds.*

In addition, combining Theorem 2.1 with the optimality conditions for problems (1a)–(1d), the following theorems give the necessary and sufficient conditions of unique optimality for those problems.

**Theorem 2.2** (Basis pursuit unique optimality) *Under Assumptions 2.1–2.2, $x^* \in \mathbb{R}^n$ is the unique solution to problem (1a) if and only if $Ax^* = b$ and Condition 2.1 is satisfied.*

**Theorem 2.3** (Problems (1b)–(1d) unique optimality) *Under Assumptions 2.1–2.3 and the additional assumption $f_1, f_2, f_3 \in C^1$, $x^* \in \mathbb{R}^n$ is the unique solution to problem (1b), (1c), or (1d) if and only if, respectively,

\[
\exists p^* \in \partial \|x^*\|_1, \exists p^* + \lambda A^T \nabla f_1(Ax^* - b) = 0, \quad (2a)
\]

\[
f(Ax^* - b) \leq \sigma \quad \text{and} \quad \exists p^* \in \partial \|x^*\|_1, \eta \geq 0, \quad \exists p^* + \eta A^T \nabla f_2(Ax^* - b) = 0, \quad \text{or} \quad (2b)
\]

\[
\|x^*\|_1 \leq \tau \quad \text{and} \quad \exists p^* \in \partial \|x^*\|_1, \nu \geq 0, \quad \exists \nu p^* + A^T \nabla f_3(Ax^* - b) = 0, \quad (2c)
\]

and in addition, Condition 2.1 holds.

The proofs of these theorems are given in Sect. 4.

### 3 Related Works

Since the sufficiency is not the focus of this paper, we do not go into more details of the sufficient conditions. We would just point out that several papers, such as [15,16], construct the least-squares (i.e., minimal $\ell_2$-norm) solution $\tilde{y}$ of $A_I^T y = s$ and establish $\|A_I^T \tilde{y}\|_{\infty} < 1$. Next, we review the existing results toward necessary conditions for the uniqueness of $\ell_1$ minimizer.

Work [17] considers problem (1a) with complex-valued quantities and $A$ equal to a down-sampled discrete Fourier operator, for which it establishes both the necessity and sufficiency of Condition 1 to the solution uniqueness of (1a). Their proof uses the Hahn–Banach Theorem and the Parseval Formula. Work [18] lets the entries of matrix $A$ and vector $x$ in problem (1a) have complex values and gives a sufficient condition for its solution uniqueness. In regularization theory, Condition 2.1 is used...
to derive linear error bounds under the name of range or source conditions in [19], which shows the necessity and sufficiency of Condition 1 for solution uniqueness of (1a) in a Hilbert space setting. More recently, [20] constructs the set

$$\mathbb{F} = \{ x : \| A_J^T ( A_J^T A_J)^{-1} \text{sign}(x_J) \|_\infty < 1 \text{ and rank}(A_J) = |J| \},$$

where $J := \text{supp}(x)$, and states that the set of vectors that can be recovered by (1a) is exactly characterized by the closure of $\mathbb{F}$ if the measurement matrix $A$ satisfies the general position condition: for any sign vector $s \in \{-1, 1\}^n$, the set of columns $\{ A^i \}$ of $A \in \mathbb{R}^{m \times n}$ satisfying that any $k$-dimensional affine subspace of $\mathbb{R}^m$, $k < m$, contains at most $k + 1$ points from the set $\{ s_i A^i \}$. This paper claims that the result holds without the condition.

To our knowledge, there are very few conditions addressing the solution uniqueness of problems (1b)–(1d). The following conditions in [15,21] are sufficient for $x^*$ to the unique minimizer of (1b) for $f_1(\cdot) = \frac{1}{2} \| \cdot \|_2^2$:

$$A_I^T (b - A_I x_I^*) = \lambda \cdot \text{sign}(x_I^*), \quad (3a)$$

$$\| A_I^T (b - A_I x_I^*) \|_\infty < \lambda, \quad (3b)$$

$$A_I^T \text{has full column rank.} \quad (3c)$$

However, they are not necessary in light of the following example. Let

$$A := \begin{bmatrix} 1 & 0 & 2 \\ 0 & 2 & -2 \end{bmatrix}, \quad b = \begin{bmatrix} 1 \\ 1 \end{bmatrix}, \quad \lambda = 1 \quad (4)$$

and consider solving the Lasso problem, as a special case of problem (1b):

$$\min \left[ \frac{1}{2} \| Ax - b \|_2^2 + \lambda \| x \|_1 \right]. \quad (5)$$

One gets the unique solution $x^* = [0 \ 1/4 \ 0]^T$ and $I = \text{supp}(x^*) = \{2\}$. However, the inequality in condition (3b) holds with equality. In general, conditions (3) become necessary in case $A_I$ happens to be a full rank square matrix. This assumption, however, does not apply to a sparse solution $x^*$. Nevertheless, we summarize the result in the following corollary.

**Corollary 3.1** If $x^*$ is the unique minimizer of problem (1b) with $f_1(\cdot)$ set to $\frac{1}{2} \| \cdot \|_2^2$ and if $A_I$, where $I = \text{supp}(x^*)$, is a square matrix with full rank, then the conditions given in Eqs. (3) holds.

**Proof** From Theorem 2.3, if $x^*$ is the unique minimizer of problem (1b) with $f_1(\cdot) = \frac{1}{2} \| \cdot \|_2^2$, then Condition 1 holds, so there must exist a vector $y$ such that $A_I^T y = s$ and $\| A_I^T y \|_\infty < 1$. Combining with (3a), we have

$$\lambda A_I^T y = \lambda s = A_I^T (b - A_I x^*).$$
Since $A_I$ is a full rank square matrix, we get $y = \frac{1}{\lambda}(b - A_Ix^*)$. Substituting this formula to $\|A_{I^c}^T y\|_\infty < 1$, we obtain condition (3). □

Very recently, work [22] investigates the solution uniqueness of (5).

**Theorem 3.1** ([22]) Let $x^*$ be a solution of (5) and

$$J := \{i : |\langle a_i, b - Ax^* \rangle| = \lambda \}.$$  \hspace{0.5cm} (6)

If submatrix $A_J$ is full column rank, then $x^*$ is unique. Conversely, for almost every $b \in \mathbb{R}^m$, if $x^*$ is unique, then $A_J$ is full column rank.

In Theorem 3.1, the necessity part “for almost every $b$” is new (here, “almost every $b$” means that the statement holds except possibly on a set of Lebesgue measure zero). Indeed, it is not for every $b$. An example is given in (4) with a unique solution $x^*$ and $J = \{1, 2, 3\}$, but $A_J$ does not have full column rank. On the other hand, we can figure out a special case in which the full column rankness of $A_J$ becomes necessary for all $b$ in the following corollary.

**Corollary 3.2** Let $x^*$ be a solution of problem (5), $J$ be defined in (6), and $I := \text{supp}(x^*)$. If $|J| = |I| + 1$, then $x^*$ is the unique solution if and only if $A_J$ has full column rank.

**Proof** The sufficiency part follows from Theorem 3.1. We shall show the necessity part. Following the assumption $|J| = |I| + 1$, we let $\{i_0\} = J \setminus I$. Since $x^*$ is the unique solution, from Theorem 2.3, we know that $A_I$ has full column rank. Hence, if $A_J$ does not have full column rank, then we can have $a_{i_0} = A_I \beta$ for some $\beta \in \mathbb{R}^{|I|}$. From Theorem 2.3, if $x^*$ is the unique minimizer, then Condition 2.1 holds, and in particular, there exists a vector $y$ such that $A_I^T y = s$ and $\|A_{I^c}^T y\|_\infty < 1$. Now, on one hand, as $i_0 \in I^c$, we get $1 > |\langle a_{i_0}, y \rangle| = |\langle \beta, A_I^T y \rangle| = |\langle \beta, s \rangle|$; on the other hand, as $i_0 \in J$, we also have $|\langle a_{i_0}, b - Ax^* \rangle| = \lambda$, which implies

$$1 = \frac{1}{\lambda} |\langle a_{i_0}, b - Ax^* \rangle| = \frac{1}{\lambda} |\langle \beta, A_I^T (b - Ax^*) \rangle| = |\langle \beta, s \rangle|,$$

where the last equality follows from (2a). Contradiction. □

4 Proofs of Theorems 2.1–2.3

We establish Theorem 2.1 in three steps. The first step proves the theorem for problem (1a). Since the only difference between it and Theorem 2.2 is the conditions $Ax^* = b$, we prove Theorem 2.2 first. In the second step, for problems (1b)–(1d), we show that both $\|x\|_1$ and $Ax - b$ are constant for $x$ over each of $X_\lambda$, $Y_\sigma$, and $Z_T$. Finally, Theorem 2.1 is shown for problems (1b)–(1d).

**Proof (of Theorem 2.2)** We will use $I = \text{supp}(x^*)$ and $s = \text{sign}(x^*_I)$ below.
“⇐” This part has been shown in [15, 18]. For completeness, we give a proof. Let \( y \) satisfy Condition 2.1, part 2, and let \( x \in \mathbb{R}^n \) be an arbitrary vector satisfying \( Ax = b \) and \( x \neq x^* \). We shall show \( \|x^*\|_1 < \|x\|_1 \).

Since \( A_I \) has full column rank and \( x \neq x^* \), we have \( \text{supp}(x) \neq I \); otherwise from \( A_I x^*_I = b = A_I x_I \), we would get \( x^*_I = x_I \) and the contradiction \( x^* = x \).

From \( \text{supp}(x) \neq I \), we get \( \langle b, y \rangle < \|x\|_1 \). To see this, let \( J := \text{supp}(x) \setminus I \), which is a non-empty subset of \( I^c \). From Condition 2.1, we have \( \|A_I^Ty\|_\infty = 1 \) and \( \|A_J^Ty\|_\infty < 1 \), and thus

\[
\langle x_I, A_I^Ty \rangle \leq \|x_I\|_1 \cdot \|A_I^Ty\|_\infty \leq \|x_I\|_1,
\]

\[
\langle x_J, A_J^Ty \rangle \leq \|x_J\|_1 \cdot \|A_J^Ty\|_\infty < \|x_J\|_1,
\]

(the last inequality is “\(<\)” not “\(\leq\)”) which lead to

\[
\langle b, y \rangle = \langle x, A^Ty \rangle = \langle x_I, A_I^Ty \rangle + \langle x_J, A_J^Ty \rangle < \|x_I\|_1 + \|x_J\|_1 = \|x\|_1.
\]

On the other hand, we have

\[
\|x^*\|_1 = \langle x^*_I, \text{sign}(x^*_I) \rangle = \langle x^*_I, A_I^Ty \rangle = \langle A_Ix^*_I, y \rangle = \langle b, y \rangle
\]

and thus \( \|x^*\|_1 = \langle b, y \rangle < \|x\|_1 \).

“⇒”. Assume that \( x^* \) is the unique solution to (1a). Obviously, \( Ax^* = b \).

It is easy to obtain Condition 2.1, part 1. Suppose it does not hold. Then, \( A_I \) has a nontrivial null space, and perturbing \( x^*_I \) along the null space will change the objective \( \|x^*_I\|_1 = \langle s, x^*_I \rangle \) while maintaining \( A_Ix^*_I = b \); hence, this perturbing breaks the unique optimality of \( x^* \). In more details, there exists a nonzero vector \( d \in \mathbb{R}^n \) such that \( A_IdI = 0 \) and \( dI^c = 0 \). For any scalar \( \alpha \) near zero, we have \( \text{sign}(x^*_I + \alpha d_I) = \text{sign}(x^*_I) = s \) and thus

\[
\|x^* + \alpha d\|_1 = \langle s, (x^*_I + \alpha d_I) \rangle = \langle s, x^*_I \rangle + \alpha \langle s, d_I \rangle = \|x^*\|_1 + \alpha \langle s, d_I \rangle.
\]

Since \( x^* \) is the unique solution, we must have \( \|x^*_I + \alpha d_I\|_1 > \|x^*\|_1 \) and thus \( \alpha \langle s, d_I \rangle > 0 \) whenever \( \alpha \neq 0 \). This is false as \( \alpha \) can be negative.

It remains to construct a vector \( y \) for Condition 2.1, part 2. Our construction is based on the strong convexity relation between a linear program (called the primal problem) and its dual problem, namely, if one problem has a solution, so does the other, and the two solutions must give the same objective value. (For the interested reader, this result follows from the Hahn–Banach Separation Theorem, also from the theorem of alternatives [23].)

The strong duality relation holds between (1a) and its dual problem

\[
\max_{p \in \mathbb{R}^m} \langle b, p \rangle \quad \text{s.t.} \quad \|A^Tp\|_\infty \leq 1
\]
because (1a) and (7), as a primal-dual pair, are equivalent to the primal-dual linear programs

\[
\begin{align*}
\min_{u, v \in \mathbb{R}^n} & \quad \langle 1, u \rangle + \langle 1, v \rangle \\
\text{s.t.} & \quad Au - Av = b, \quad u \geq 0, \quad v \geq 0,
\end{align*}
\]

\[
\max_{q \in \mathbb{R}^m} \langle b, q \rangle \\
\text{s.t.} & \quad -1 \leq A^T q \leq 1,
\]

respectively, where the strong duality relation holds between (8a) and (8b). By “equivalent,” we mean that one can obtain solutions from each other:

- given \(u^*, v^*\), obtain \(x^* = u^* - v^*\);
- given \(x^*\), obtain \(u^* = \max(x^*, 0), \quad v^* = \max(-x^*, 0)\);
- given \(q^*\), obtain \(p^* = q^*\);
- given \(p^*\), obtain \(q^* = p^*\).

Therefore, since (1a) has solution \(x^*\), there exists a solution \(y^*\) to (7), which satisfies \(\|x^*\|_1 = \langle b, y^* \rangle\) and \(\|A^T y^*\|_\infty \leq 1\). (One can obtain such \(y^*\) from the Hahn–Banach Separation Theorem or the theorem of alternatives rather directly.) However, \(y^*\) may not obey \(\|A^T_{I^c} y^*\|_\infty < 1\). We shall perturb \(y^*\) so that \(\|A^T_{I^c} y^*\|_\infty < 1\). We adopt a strategy similar to the construction of a strictly complementary dual solution in linear programming.

To prepare for the perturbation, we let

\[
L := \{i \in I^c : \langle a_i, y^* \rangle = -1\} \quad \text{and} \quad U := \{i \in I^c : \langle a_i, y^* \rangle = 1\}.
\]

Our goal is to perturb \(y^*\) so that \(-1 < \langle a_i, y^* \rangle < 1\) for \(i \in L \cup U\) and \(y^*\) remains optimal to (7). To this end, consider for a fixed \(\alpha > 0\) and \(t := \|x^*\|_1\), the linear program

\[
\begin{align*}
\min_{x \in \mathbb{R}^n} & \quad \sum_{i \in L} \alpha x^*_i - \sum_{i \in U} \alpha x^*_i \\
\text{s.t.} & \quad Ax = b, \quad \|x\|_1 \leq t.
\end{align*}
\]

Since \(x^*\) is the unique solution to (1a), it is the unique feasible solution to problem (9), so (9) has the optimal objective value \(\sum_{i \in U} \alpha x^*_i - \sum_{i \in L} \alpha x^*_i = 0\). By setting up equivalent linear programs like what has been done for (1a) and (7), the strong duality relation holds between (9) and its dual problem

\[
\begin{align*}
\max_{p \in \mathbb{R}^m, q \in \mathbb{R}} & \quad \langle b, p \rangle - tq \\
\text{s.t.} & \quad \|A^T p - \alpha r\|_\infty \leq q, \quad q \geq 0,
\end{align*}
\]

where \(r \in \mathbb{R}^n\) is given by

\[
r_i = \begin{cases} 
1, & i \in L, \\
-1, & i \in U, \\
0, & \text{otherwise.}
\end{cases}
\]

Therefore, (10) has a solution \((p^*, q^*)\) satisfying \(\langle b, p^* \rangle - tq^* = 0\).
According to the last constraint of (10), we have $q^* \geq 0$, which we split into two cases: $q^* = 0$ and $q^* > 0$.

i) If $q^* = 0$, we have $A^T p^* = a \alpha r$ and $\langle b, p^* \rangle = 0$.

ii) If $q^* > 0$, we let $r^* := p^*/q^*$, which satisfies $\langle b, r^* \rangle = t = \|x^*\|_1$ and $\|A^T r^* - \frac{\alpha}{q^*} r\|_{\infty} \leq 1$, or equivalently, $-1 + \frac{\alpha}{q^*} r \leq A^T r^* \leq 1 + \frac{\alpha}{q^*} r$.

Now we perturb $y^*$. Solve (10) with a sufficiently small $\alpha > 0$ and obtain a solution $(p^*, q^*)$. If case (i) occurs, we let $y^* \leftarrow y^* + p^*$; otherwise, we let $y^* \leftarrow \frac{1}{2}(y^* + r^*)$.

In both cases,

- $\langle b, y^* \rangle$ is unchanged, still equal to $\|x^*\|_1$;
- $-1 < \langle a_i, y^* \rangle < 1$ holds for $i \in L \cup U$ after the perturbation;
- for each $j \not\in L \cup U$, if $\langle a_j, y^* \rangle \in [-1, 1]$ or $\langle a_j, y^* \rangle \not\in ]-1, 1[$ holds before the perturbation, the same holds after the perturbation;

Therefore, after the perturbation, $y^*$ satisfies:

1) $\langle b, y^* \rangle = \|x^*\|_1$,

2) $\|A^T y^*\|_{\infty} \leq 1$, and

3) $\|A^T, y^*\|_{\infty} < 1$.

From 1) and 2) it follows

4) $A^T y = \text{sign}(x^*_{i})$

since $\|x^*_{i}\|_1 = \|x^*\|_1 = \langle b, y^* \rangle = \langle A_1 x^*_{i}, y^* \rangle = \langle x^*_{i}, A^T y^* \rangle$ and by Holder’s inequality $\langle x^*_{i}, A^T y^* \rangle \leq \|x^*_{i}\|_1 \|A^T y^*\|_{\infty} \leq \|x^*_{i}\|_1$, and thus

$$\langle x^*_{i}, A^T y^* \rangle = \|x^*_{i}\|_1.$$ 

which dictates 4). From 3) and 4), Condition 2.1, part 2, holds with $y = y^*$.

\[\square\]

**Proof (of Theorem 2.1 for problem (1a))** The above proof for Theorem 2.2 also serves the proof of Theorem 2.1 for problem (1a) since $Ax^* = b$ is involved only in the optimality part, not the uniqueness part. \[\square\]

Next, we show that $Ax - b$ is constant for $x$ over $X_\lambda, Y_\sigma, Z_\tau$.

**Lemma 4.1** Let $f$ be a strictly convex function. If $f(Ax - b) + \|x\|_1$ is constant on a convex set $S$, then both $Ax - b$ and $\|x\|_1$ are constant on $S$.

**Proof** It suffices to prove the case where $S$ has more than one point. Let $x_1$ and $x_2$ be any two different points in $S$. Consider the line segment $L$ connecting $x_1$ and $x_2$. Since $X$ is convex, we have $L \subset X$ and that $f(Ax - b) + \|x\|_1$ is constant on $L$. On one hand, $\|x\|_1$ is piece-wise linear over $L$; on the other hand, the strict convexity of $f$ makes it impossible for $f(Ax - b)$ to be piece-wise linear over $L$ unless $Ax_1 - b = Ax_2 - b$. Hence, we have $Ax_1 - b = Ax_2 - b$, and thus $f(Ax_1 - b) = f(Ax_2 - b)$, from which it follows $\|x_1\|_1 = \|x_2\|_1$. Since $x_1$ and $x_2$ are arbitrary two points in $S$, the lemma is proved. \[\square\]

With Lemma 4.1 we can show
Lemma 4.2 Under Assumptions 2.2 and 2.3, the following statements for problems (1b)–(1d) hold

1) $X_{\lambda}$, $Y_{\sigma}$, and $Z_{\tau}$ are convex;
2) In problem (1b), $Ax - b$ and $\|x\|_1$ are constant for all $x \in X_{\lambda}$;
3) Part 2) holds for problem (1c) and $Y_{\sigma}$;
4) Part 2) holds for problem (1d) and $Z_{\tau}$.

Proof Assumption 2.2 makes sure that $X_{\lambda}$, $Y_{\sigma}$, and $Z_{\tau}$ are all non-empty.
1) The set of solutions of a convex program is convex.
2) Since $f_1(Ax - b) + \lambda \|x\|_1$ is constant over $x \in X_{\lambda}$ and $f_1$ is strictly convex by Assumption 2.3, the result follows directly from Lemma 4.1.
3) If $0 \notin Y_{\sigma}$, then the optimal objective is $\|0\|_1 = 0$; hence, $Y_{\lambda} = \{0\}$ and the results hold trivially. Suppose $0 \notin Y_{\sigma}$. Since the optimal objective $\|x\|_1$ is constant for all $x \in Y_{\sigma}$ and $f_2$ is strictly convex by Assumption 2.3, to prove this part in light of Lemma 4.1, we shall show $f_2(Ax - b) = \sigma$ for all $x \in Y_{\sigma}$.
Assume that there is $\hat{x} \in Y_{\sigma}$ such that $f_2(A\hat{x} - b) < \sigma$. Since $f_2(Ax - b)$ is convex and thus continuous in $x$, there exists a non-empty ball $B$ centered at $\hat{x}$ with a sufficiently small radius $\rho > 0$ so that $f_2(A\bar{x} - b) < \sigma$ for all $\bar{x} \in B$. Let $\alpha := \min \{\frac{\rho}{\|A\|_2}, \frac{1}{\sqrt{2}}\} \in ]0, 1[$. We have $(1 - \alpha)\hat{x} \in B$ and also have $\| (1 - \alpha)\hat{x} \|_1 = (1 - \alpha)\|\hat{x}\|_1 < \|\hat{x}\|_1$, so $(1 - \alpha)\hat{x}$ is both feasible and achieving an objective value lower than the optimal one. Contradiction.
4) By Assumption 2.3, we have $\|x\|_1 = \tau$ for all $x \in Z_{\tau}$; otherwise, there exists $\bar{x} \in Z_{\tau}$ such that $\tau > f_3(A\bar{x} - b) \geq \inf \{\|x\|_1 : f_3(Ax - b) = f_3^*\}$, contradicting Assumption 2.3. Since the optimal objective $f_3(Ax - b)$ is constant for all $x \in Z_{\tau}$ and $f_3$ is strictly convex, the result follows from Lemma 4.1.

Proof (of Theorem 2.1 for problems (1b)–(1d)) This proof exploits Lemma 4.2. Since the results of Lemma 4.2 are identical for problems (1b)–(1d), we present the proof for problem (1b). The proofs for the other two are similar.

From Assumption 2.3, $X_{\lambda}$ is nonempty so we pick $x^* \in X_{\lambda}$. Let $b^* = Ax^*$, which is independent of specific $x^*$ (Lemma 4.2). Consider the linear program

$$
\min \|x\|_1, \quad \text{s.t. } Ax = b^*,
$$

and let $X^*$ denote its solution set.

Now, we show that $X_{\lambda} = X^*$. Since $Ax = Ax^*$ and $\|x\|_1 = \|x^*\|_1$ for all $x \in X_{\lambda}$, and conversely, any $x$ obeying $Ax = Ax^*$ and $\|x\|_1 = \|x^*\|_1$ belongs to $X_{\lambda}$, it is sufficient to show that $\|x\|_1 = \|x^*\|_1$ for any $x \in X^*$. Assuming this does not hold, then as problem (11) has $x^*$ as a feasible solution and has a finite objective, we have a nonempty $X^*$ and there exists $\tilde{x} \in X^*$ satisfying $\|\tilde{x}\|_1 < \|x^*\|_1$. However, $f(Ax^* - b) = f(b^* - b) = f(Ax^* - b)$ and $\|\tilde{x}\|_1 < \|x^*\|_1$ mean that $\tilde{x}$ is a strictly better solution than $x^*$, contradicting $x^* \in X_{\lambda}$.

Since $X_{\lambda} = X^*$, $x^*$ is the unique solution to problem (1b) if and only if it is the same to problem (11). Since (11) is in the same form of (1a), applying the part of Theorem 2.1 for (1a), which is already proved, we conclude that $x^*$ is the unique solution to problem (1b) if and only if Condition 2.1 holds.
Proof of Theorem 2.3. The proof above also serves the proof for Theorem 2.3 since (2a)–(2c) is the optimality conditions of \(x^*\) to problems (1b)–(1d), respectively, and furthermore, given the optimality of \(x^*\), Condition 2.1 is the necessary and sufficient condition for the uniqueness of \(x^*\). \(\square\)

Remark 4.1 For problems (1b)–(1d), the uniqueness of a given solution \(x^* \neq 0\) is also equivalent to a condition that is slightly simpler than Condition 2.1. To present the condition, consider the first-order optimality conditions (the KKT conditions) (2a)–(2c) of \(x^*\) to problems (1b)–(1d), respectively. Given \(x^* \neq 0\), \(\eta\) and \(\nu\) can be computed. From \(p^* \neq 0\) it follows that \(\eta > 0\). Moreover, \(\nu = 0\) if and only if \(A^T \nabla f_3(Ax^* - b) = 0\). The condition below for the case \(\nu = 0\) in problem (1d) reduces to Condition 2.1. Define

\[
P_1 := \{i : |\lambda \langle a_i, \nabla f_1(Ax^* - b) \rangle| = 1\},
\]

\[
P_2 := \{i : |\eta \langle a_i, \nabla f_2(Ax^* - b) \rangle| = 1\},
\]

\[
P_3 := \{i : |\langle a_i, \nabla f_2(Ax^* - b) \rangle| = \nu\}.
\]

By the definitions of \(\partial \|x^*\|_1\) and \(P_i\), we have \(\text{supp}(x^*) \subseteq P_i, i = 1, 2, 3\).

Condition 4.1 Under the definitions \(I := \text{supp}(x^*) \subseteq P_i\) and \(s := \text{sign}(x^*_I)\), matrix \(A_{P_i} \in \mathbb{R}^{m \times |P_i|}\) obeys

1. submatrix \(A_I\) has full column rank, and
2. there exists \(y \in \mathbb{R}^m\) such that \(A_I^T y = s\) and \(\|A_{P_I \setminus I} y\|_\infty < 1\).

Condition 4.1 only checks the submatrix \(A_{P_i}\) but not the full matrix \(A\).

It is not difficult to show that the linear programs

\[
\min \|x\|_1, \quad \text{s.t.} \ (A_{P_i})x = b^*,
\]

for \(i = 1, 2, 3\), have the solution sets \(X_\lambda, Y_\sigma, \text{and} Z_\tau\), respectively. Hence, we have

Theorem 4.1 Under Assumptions 2.1–2.3 and assuming \(f_1, f_2, f_3 \in C^1\), given that \(x^* \neq 0\) is a solution to problem (1b), (1c), or (1d), \(x^*\) is the unique solution if and only if Condition 4.1 holds for \(i = 1, 2, \text{or} 3\), respectively.

5 Recognizing and Verifying Unique Solutions

Applying Theorem 2.1, we can recognize the uniqueness of a given solution \(x^*\) to problem (1a) given a dual solution \(y^*\) (to problem (7)). Specifically, let \(J := \{i : |\langle a_i, y^* \rangle| = 1\}\), and if \(A_J\) has full column rank and \(\text{supp}(x^*) = J\), then following Theorem 2.1, \(x^*\) is the unique solution to (1a). The converse is not true since there can exist many dual solutions with different \(J\). The key is to find the one with the smallest \(J\). Several interior point methods ([24] for example) return the dual solution \(y^*\) with the smallest \(J\), so if either \(A_J\) is column-rank deficient or \(\text{supp}(x^*) \neq J\), then \(x^*\) is surely non-unique.
Corollary 5.1 Under Assumption 2.1, given a pair of primal–dual solutions \((x^*, y^*)\) to problem (1a), let \(J := \{ i : |\langle a_i, y^* \rangle| = 1 \}\). Then, \(x^*\) is the unique solution to (1a) if \(A_J\) has full column rank and \(\text{supp}(x^*) = J\). If \(y^*\) is obtained by a linear programming interior-point algorithm, the converse also holds.

Similar results also hold for (1b)–(1d) if a dual solution \(y^*\) to (11) is available.

One can directly verify Condition 2.1. Given a matrix \(A \in \mathbb{R}^{m \times n}\), a set of its columns \(I\), and a sign pattern \(s = \{ -1, 1 \}^{|I|}\), we give two approaches to verify Condition 2.1. Checking whether \(A_I\) has full column rank is straightforward.

To check part 2 of Condition 2.1, the first approach is to follow the proof of Theorem 2.1. Note that Condition 2.1 depends only on \(A, I,\) and \(s\), independent of \(x^*\). Therefore, construct an arbitrary \(x^*\) such that \(\text{supp}(x^*) = I\) and \(\text{sign}(x^*_I) = s\) and let \(b = Ax^*\). Solve problem (7) and let \(y^*\) be its solution. If \(y^*\) satisfies part 2 of Condition 2.1, we are done; otherwise, define \(L, U,\) and \(t\) by \(x^*\) as in the proof, pick a small \(\bar{\alpha} > 0\), and solve program (10) parametrically in \(\alpha \in [0, \bar{\alpha}]\). The solution is piece-wise linear in \(\alpha\) (it is possible that the solution does not exist over certain intervals of \(\alpha\)). Then, check if there is a perturbation to \(y^*\) so that \(y^*\) satisfies part 2 of Condition 2.1.

In the other approach to check that part, one can solve the convex program

\[
\min_{y \in \mathbb{R}^m} \left[ -\sum_{i \in I^c} \log(1 - \langle a_i, y \rangle) + \log(1 + \langle a_i, y \rangle) \right] \quad \text{s.t.} \quad A_I^T y = s. \quad (12)
\]

Since \(\langle a_i, y \rangle \to 1\) or \(\langle a_i, y \rangle \to -1\) will infinitely increase the objective, (12) will return a solution satisfying Condition 2.1, part 2, as long as a solution exists. In fact, any feasible solution to (12) with a finite objective satisfies Condition 2.1, part 2. To find a feasible solution, one can apply the augmented Lagrangian method, which does not require \(A_I^T y = s\) to hold at the initial point (which must still satisfy \(|\langle a_i, y \rangle| < 1\) for all \(i \in I^c\)), or one can consider the alternating direction method of multipliers (ADMM) and the equivalent problem

\[
\min_{y, z} \left[ -\sum_{i \in I^c} \log(1 - z_i) + \log(1 + z_i) \right] \quad \text{s.t.} \quad A_I^T y = s, \quad z - A_{I^c}^T y = 0. \quad (13)
\]

One can start ADMM from the origin, and the two subproblems of ADMM have closed-form solutions; in particular, the \(z\)-subproblem is separable in \(z_i\)’s and reduces to finding the zeros of 3-order polynomials in \(z_i, i \in I^c\). If (13) has a solution, ADMM will find one; otherwise, it will diverge.

It is worth mentioning that one can use alternating projection in [25] to generate test instances that fulfill Condition 2.1.

6 Conclusions

Solution uniqueness is a fundamental question in computational mathematics. For the widely used basis pursuit model and its variants, an existing sufficient condition is shown also necessary in this paper. The proof essentially exploits the fact that a
pair of feasible primal–dual linear programs has strict complementary solutions. The results shed light on numerically recognizing unique solutions and verifying solution uniqueness. Most existing conditions and sampling matrix constructions ensure unique recovery of all sufficiently sparse signals. Such uniform recovery is not required in many applications as the signals of interest are specific. Therefore, an important line of future work is to develop computationally tractable approaches that construct sampling matrices that ensure unique solutions for a set of signals. To this end, the necessary and sufficient conditions in this paper are good start points.

Acknowledgments The authors thank Prof. Dirk Lorenz for bringing References [19] and [25] to their attention. The work of H. Zhang is supported by the Chinese Scholarship Council during his visit to Rice University, and in part by NUDT Funding of Innovation B110202. The work of W. Yin is supported in part by NSF Grants DMS-0748839 and ECCS-1028790. The work of L. Cheng is supported by NSFC Grants 61271014 and 61072118.

References

1. Efron, B., Hastie, T., Johnstone, I., Tibshirani, R.: Least angle regression. Ann. Stat. 32(2), 407–499 (2004)
2. Best, M.J., Grauer, R.R.: Sensitivity analysis for mean–variance portfolio problems. Manag. Sci. 37(8), 980–989 (1991)
3. Donoho, D.L., Elad, M.: Optimally sparse representation in general (nonorthogonal) dictionaries via \( \ell_1 \) minimization. Proc. Natl. Acad. Sci. 100(5), 2197–2202 (2003)
4. 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(1), 34–81 (2009)
5. Donoho, D.L., Huo, X.: Uncertainty principles and ideal atomic decomposition. IEEE Trans. Inf. Theory 47(7), 2845–2862 (2001)
6. Elad, M., Bruckstein, A.M.: A generalized uncertainty principle and sparse representation in pairs of bases. IEEE Trans. Inf. Theory 48(9), 2558–2567 (2002)
7. Donoho, D.L.: Compressed sensing. IEEE Trans. Inf. Theory 52(4), 1289–1306 (2006)
8. Cohen, A., Dahmen, W., DeVore, R.: Compressed sensing and best k-term approximation. J. Am. Math. Soc. 22(1), 211–231 (2009)
9. Candès, E.J., Tao, T.: Decoding by linear programming. IEEE Trans. Inf. Theory 51(12), 4203–4215 (2005)
10. Zhang, Y.: Theory of compressive sensing via \( \ell_1 \)-minimization: a non-RIP analysis and extensions. J. Oper. Res. Soc. China 1(1), 79–105 (2008)
11. Candès, E.J., Plan, Y.: A probabilistic and RIPless theory of compressed sensing. IEEE Trans. Inf. Theory 57(11), 7235–7254 (2011)
12. Chvatal, V.: Linear Programming. W.H. Freeman and Company, New York (1983)
13. Chen, S.S., Donoho, D.L., Saunders, M.A.: Atomic decomposition by basis pursuit. SIAM J. Sci. Comput. 20(1), 33–61 (1998)
14. Tibshirani, R.: Regression shrinkage and selection via the LASSO. J. R. Stat. Soc. Ser. B (Methodological) 58, 267–288 (1996)
15. Fuchs, J.J.: On sparse representations in arbitrary redundant bases. IEEE Trans. Inf. Theory 50(6), 1341–1344 (2004)
16. Candès, E., Recht, B.: Simple bounds for recovering low-complexity models. Math. Program. 141, 577–589 (2013)
17. Candès, E.J., Romberg, J., Tao, T.: Robust uncertainty principles: exact signal reconstruction from highly incomplete frequency information. IEEE Trans. Inf. Theory 52(2), 489–509 (2006)
18. Tropp, J.A.: Recovery of short, complex linear combinations via \( \ell_1 \) minimization. IEEE Trans. Inf. Theory 51(4), 1568–1570 (2005)
19. Grasmair, M., Scherzer, O., Haltmeier, M.: Necessary and sufficient conditions for linear convergence of \( \ell_1 \)-regularization. Commun. Pure Appl. Math. 64(2), 161–182 (2011)
20. Dossal, C.: A necessary and sufficient condition for exact sparse recovery by $\ell_1$ minimization. Comptes Rendus Mathematique 350(1), 117–120 (2012)
21. Fuchs, J.J.: Recovery of exact sparse representations in the presence of bounded noise. IEEE Trans. Inf. Theory 51(10), 3601–3608 (2005)
22. Tibshirani, R.J.: The LASSO problem and uniqueness. Electron. J.Stat. 7, 1456–1490 (2013)
23. Dantzig, G.B.: Linear Programming and Extensions. Princeton university press, Princeton, NJ (1998)
24. Güler, O., Ye, Y.: Convergence behavior of interior-point algorithms. Math. Program. 60(1–3), 215–228 (1993)
25. Lorenz, D.A.: Constructing test instances for basis pursuit denoising. IEEE Trans. Signal Process. 61(5–8), 1210–1214 (2013)