One condition for solution uniqueness and robustness of both l1-synthesis and l1-analysis minimizations

Hui Zhang · Ming Yan · Wotao Yin

Abstract The ℓ_1-synthesis model and the ℓ_1-analysis model recover structured signals from their under-sampled measurements. The solution of former is a sparse sum of dictionary atoms, and that of the latter makes sparse correlations with dictionary atoms. This paper addresses the question: when can we trust these models to recover specific signals? We answer the question with a condition that is both necessary and sufficient to guarantee the recovery to be unique and exact and, in presence of measurement noise, to be robust. The condition is one–for–all in the sense that it applies to both of the ℓ_1-synthesis and ℓ_1-analysis models, to both of their constrained and unconstrained formulations, and to both the exact recovery and robust recovery cases. Furthermore, a convex infinity–norm program is introduced for numerically verifying the condition. A comprehensive comparison with related existing conditions are included.

Keywords exact recovery · robust recovery · ℓ_1-analysis · ℓ_1-synthesis · sparse optimization · compressive sensing

1 Introduction

With both synthesis and analysis models of sparse signal recovery [1], one is interested in when the recovery is successful, namely, whether the solution is unique and whether the solution error is proportional to the amount of noise in the measurements.

Various *sufficient* conditions have been proposed to guarantee successful recovery. Let us first discuss two types of recovery conditions. The *uniform recovery* conditions guarantee not only the successful recovery of one sparse signal but all signals that are sufficiently sparse, irrelevant of the locations of their nonzero entries.

Communicated by xxx

Hui Zhang
Department of Mathematics and Systems Science, College of Science,
(also affiliated with the state key laboratory for high performance computation,)
National University of Defense Technology,
Changsha, Hunan, China, 410073.
E-mail: hhuuii.zhang@gmail.com

Ming Yan and Wotao Yin
Department of Mathematics,
University of California,
Los Angeles, CA 90095, USA.
E-mail: yanm@math.ucla.edu (M. Yan) and wotaoyin@math.ucla.edu (W. Yin)

The *non-uniform recovery conditions*, however, focus on the recovery of a restricted set of sufficiently sparse signals, for example, the sparse signals with a specific support. Since non-uniform recovery conditions cover fewer signals, they are in general weaker and thus easier to hold than uniform recovery conditions. In addition, some non-uniform recovery conditions, especially those for fixed signal support, are easier to numerically verify, whereas all the existing uniform recovery conditions are numerically intractable to verify on given sensing matrices. On the other hand, several kinds of random matrices (such as those with i.i.d. subgaussian entries) satisfy the uniform recovery conditions with high probability. The uniform conditions are useful in designing randomized linear measurements, whereas non-uniform conditions are useful on deterministic linear measurements on restricted sets of signals.

Non-uniform conditions for ℓ_1-synthesis minimization include the non-uniform dual certificate condition [2] and the "RIPless" property [3]. Well-known examples of uniform conditions include the restricted isometry principle [4], the null space condition [5], the spherical section property [6], and others.

Because ℓ_1-analysis minimization takes a more general form than ℓ_1-synthesis minimization, some of the above non-uniform recovery conditions have been extended to the analysis case; recent works [7,8,9,10,11, 12] have made significant contributions.

This paper studies the so-called *dual certificate* condition, a type of non-uniform condition. We show that if a signal is a solution to any model among (1a)–(1c) described below, there exists a necessary and sufficient condition, same for any of the three models, that guarantees that the signal will be uniquely recovered. While this result has been partially known in previous work for ℓ_1 minimization as a sufficient condition, we establish three new results:

- the condition previously known to be sufficient is in fact necessity;
- the condition guarantees robustness to noise. That is, under this condition, if the observed data is contaminated by arbitrary noise, the solution to either (1b) or (1c) is robust to the noise in the sense the solution error is proportional to the Euclidean norm of the noise;
- nearly the same condition but imposed on the support of the largest $|I|$ entries of an approximately sparse signal guarantees its robust recovery by model (1c).

These results complete the theory of non-uniform recovery of the ℓ_1-synthesis and ℓ_1-analysis models.

The proposed condition is compared to existing non-uniform conditions in the literature, most of which are stronger than ours and are thus sufficient yet not necessary. Note that some of those stronger ones also give additional properties that ours does not. Technically, a part of our analysis is inspired by existing results in [13,7,14,15,16,17], which bear certain similarity among themselves and will be mentioned in later sections.

The rest of the paper is organized as follows. Sections 2 and 3 formulate the problem and state the main results. Section 4 reviews several related results. Section 5 discusses condition verification. Proofs for the main results are given in sections 6, 7, and 8.

2 Problem Formulation and Contributions

2.1 Notation

We equip $\mathbb{R}^n$ with the canonical scalar product $\langle \cdot, \cdot \rangle$ and Euclidean norm $\|\cdot\|_2$. We let $|\cdot|$ return the cardinality if the input is a set or the absolute value if the input is a number. For any $x \in \mathbb{R}^n$, $\mathrm{supp}(x) = \{k : 1 \leq k \leq n, x_k \neq 0\}$ is the index set of the non-zero entries of x. $\mathrm{sign}(x)$ is the vector whose ith entry is the sign of

x_i, taking a value among $+1$, -1, and 0. For any $p \geq 1$, the ℓ_p-norm of $x \in \mathbb{R}^n$ is

$$\|x\|_p = \left(\sum_{i=1}^{n} |x_i|^p \right)^{1/p},$$

its ℓ_0-"norm" is $\|x\|_0 = |\mathrm{supp}(x)|$, and its ℓ_∞-norm is $\|x\|_\infty = \max\{|x_i| : i = 1, \cdots, n\}$. For $x \in \mathbb{R}^n$ and $I \subset \{1, 2, \cdots, n\}$, x_I denotes the vector formed by the entries x_i of x for $i \in I$, and I^c is the complement of I. Similarly, A_I is the submatrix formed by the columns of A indexed by I. A^T is the transpose of A. We use A_I^T for the transpose of submatrix A_I, not a submatrix of A^T. For square matrix A, $\lambda_{\max}(A)$ and $\lambda_{\min}(A)$ denote its largest and smallest eigenvalues, respectively, $\mathrm{Cond}(A)$ denotes its condition number, and $\|A\|$ denotes its spectral norm. The null and column spaces of A are denoted by $\mathrm{Ker}(A)$ and $\mathrm{Im}(A)$, respectively.

2.2 Problem formulations

Let $x^* \in \mathbb{R}^n$ be a signal of interest. This paper studies when ℓ_1 minimization can uniquely and robustly recover x^* from its linear measurements

$$b = \Phi x^* + w,$$

where $\Phi \in \mathbb{R}^{m \times n}$ is a certain matrix and $w \in \mathbb{R}^m$ is noise. We focus on the setting $m \leq n$.

The results of this paper cover the following ℓ_1 minimization formulations:

$$\underset{x}{\text{minimize}} \, \|\Psi^T x\|_1, \quad \text{subject to } \Phi x = b, \tag{1a}$$

$$\underset{x}{\text{minimize}} \, \|\Phi x - b\|_2^2 + \lambda \|\Psi^T x\|_1, \tag{1b}$$

$$\underset{x}{\text{minimize}} \, \|\Psi^T x\|_1, \quad \text{subject to } \|\Phi x - b\|_2 \leq \delta, \tag{1c}$$

where δ, λ are positive parameters. Model (1a) is used if no noise is present, i.e., $w = 0$. If $\Psi = Id$, the identify matrix, models in (1) are referred to as the ℓ_1 (or more generally, ℓ_1-synthesis) models. If $\Psi \neq Id$, they are referred to as the ℓ_1-analysis models, which are recently reviewed in [1].

In ℓ_1-synthesis models, the signal of interest is synthesized as $x^* = Dc$, where D is a certain dictionary and c is the sparse coefficients. The ℓ_1-analysis model, including the cosparse analysis model [18] and the total variation model [19] as widely known examples, has recently attracted a lot of attention. The underlying signal is expected to make sparse correlations with the columns (atoms) in an (overcomplete) dictionary Ψ, i.e., $\Psi^T x^*$ is sparse; see [20, 18, 21].

2.3 Geometry

It is worth noting that the underlying geometry of solution uniqueness is rather clear. Indeed, it was characterized in terms of polytope faces by Donoho in [22], and also of null spaces and tangent cones by Chandrasekaran et al. in [23]. However, there lacks a concrete *if-and-only-if* condition for one to check and the robustness results are difficult to obtain from geometry.

Let us describe the geometry of uniqueness recovery under the simple setting: $\Psi = I$. Then, x^* is the unique solution to (1a) if the affine set $\mathcal{F} = \{x : \Phi x = b\}$ touches the ℓ_1 polytope $\{x : \|x\|_1 = \gamma\}$, for some $\gamma > 0$, at a *unique* point on a face of the polytope. The uniqueness for (1b) and (1c) is similar except $\mathcal{F}$ is replaced by a "tube" $\mathcal{T} = \{x : \|\Phi x - b\|_2 \leq \sigma\}$. In both cases, the uniqueness is certified by (i) a kernel condition that ensures touching the face at just one point, and (ii) the existence of a hyperplane (dual certificate) that separates $\mathcal{F}$ or $\mathcal{T}$ from the polytope that creates a positive gap between them outside the

face. The same geometry applies to ℓ_1-analysis minimization but involves the more complicated polytope $\{x : \|\Psi^T x\|_1 = \gamma\}$; see [24,25].

This work not only formalizes these geometrical explanations with a concrete *if-and-only-if* condition, but also establishes robustness bounds under the same condition, which cannot be explain by the above geometry. Suppose that noise is present in the measurements; then, the point where $\mathcal{F}$ or $\mathcal{T}$ touches the polytope will move, and the point may or may not stay in the same face in the noise-free case. The above geometry cannot clearly explain how far the point will move due to noise.

3 Main Condition and Results

3.1 Main Condition

Condition 1 *Given $\bar{x} \in \mathbb{R}^n$, index sets $I = \mathrm{supp}(\Psi^T \bar{x}) \subset \{1, \cdots, l\}$ and $J = I^c$ satisfy*

(1) $\mathrm{Ker}(\Psi_J^T) \bigcap \mathrm{Ker}(\Phi) = \{0\}$;

(2) There exists $y \in \mathbb{R}^l$ such that $\Psi y \in \mathrm{Im}(\Phi^T)$, $y_I = \mathrm{sign}(\Psi_I^T \bar{x})$, and $\|y_J\|_\infty < 1$.

It is not difficult to understand the condition as we shall explain below.

Condition 1 part (1) says that there does *not* exist any nonzero Δx satisfying both $\Psi_J^T \bar{x} = \Psi_J^T(\bar{x} + \Delta x)$ and $\Phi\bar{x} = \Phi(\bar{x} + \Delta x)$. Otherwise, there exists a nonempty interval $\mathcal{I} = [\bar{x} - \alpha\Delta x, \bar{x} + \alpha\Delta x]$ for some sufficiently small $\alpha > 0$ so that $\Phi x = \Phi\bar{x}$ and $\|\Psi^T x\|_1$ is constant for $x \in \mathcal{I}$; hence $\bar{x}$ cannot be the unique minimizer. This condition implies that the role of ℓ_1 minimization is "limited" to recovering the support set I; if an oracle gives us I, we must solely rely on Ψ, Φ, and b recover $\bar{x}$, since $\|\Psi^T x\|_1$ is locally linear near $\bar{x}$ over $\{x : \mathrm{supp}(\Psi^T x) = I\}$ and thus lacks the ability to pick $\bar{x}$ out. In addition, part (1) of the condition also implies the existence of ρ and τ such that

$$\|\Psi_I^T x\|_2 \leq \rho\|\Psi_J^T x\|_1 + \tau\|\Phi x\|_2 \tag{2}$$

for all $x \in \mathbb{R}^n$, which will be used later in Theorem 3.

Condition 1 part (2) states the existence of a strictly-complementary *dual certificate* y. To see this, let us check a part of the optimality conditions of (1a): $0 \in \Psi\partial\| \cdot \|_1(\Psi^T x) - \Phi^T\beta$, where vector β is the Lagrangian multipliers; we can rewrite the condition as $0 = \Psi y - \Phi^T\beta$ where $y \in \partial\| \cdot \|_1(\Psi^T x)$, which translates to $y_I = \mathrm{sign}(\Psi_I^T x)$ and $\|y_J\|_\infty \leq 1$. This y certifies the optimality of $\bar{x}$. For solution uniqueness and/or robustness, we shall later show that the strict inequality $\|y_J\|_\infty < 1$ is necessary.

It is worth mentioning a variant of Condition 1 as follows.

Condition 2 *Given $\bar{x} \in \mathbb{R}^n$, let $I = \mathrm{supp}(\Psi^T \bar{x})$. There exists a nonempty index set $J \subseteq I^c$ such that the index sets I, J and $K = (I \bigcup J)^c$ satisfy*

(1) $\mathrm{Ker}(\Psi_J^T) \bigcap \mathrm{Ker}(\Phi) = \{0\}$;

(2) There exists $y \in \mathbb{R}^l$ such that $\Psi y \in \mathrm{Im}(\Phi^T)$, $y_I = \mathrm{sign}(\Psi_I^T \bar{x})$, $\|y_J\|_\infty < 1$, and $\|y_K\|_\infty \leq 1$.

In Condition 2, a smaller J relaxes part (2) but gives a larger $\mathrm{Ker}(\Psi_J^T)$ and thus tightens part (1). Although Condition 2 allows a more flexible J than Condition 1, we shall show that they are equivalent.

The comparisons between Condition 1 and those in the existing literature are given in Section 4 below.

3.2 Main Results

Depending on the specific models in (1), we need the following assumptions:

Assumption 1 *Matrix Φ has full row-rank.*

Assumption 2 $\lambda_{max}(\Psi\Psi^T) = 1.$

Assumption 3 *Matrix Ψ has full row-rank.*

Assumptions 1 and 3 are standard to avoid redundancy. Assumption 2 is non-essential as we can scale any general Ψ by multiplying $\frac{1}{\sqrt{\lambda_{\max}(\Psi\Psi^T)}}$. Below we state our main results and delay their proofs to Sections 6–8.

Theorem 1 (Uniqueness) *Under Assumption 1, let $\hat{x}$ be a solution to problem (1a) or (1b), or under Assumptions 1 and 3, let $\hat{x}$ be a solution to problem (1c). The following are equivalent:*

1) Solution $\hat{x}$ is unique;

2) Condition 1 holds for $\bar{x} = \hat{x}$;

3) Condition 2 holds for $\bar{x} = \hat{x}$.

This theorem states that Conditions 1 and 2 are equivalent, and they are necessary and sufficient for a solution $\hat{x}$ to problem (1a), or to problem (1b), or to problem (1c) to be unique. To state our next result on robustness, we let

$$r(J) := \sup_{u \in \text{Ker}(\Psi_J^T) \backslash \{0\}} \frac{\|u\|_2}{\|\Phi u\|_2}.$$

Part (1) of Condition 1 ensures that $0 < r(J) < +\infty$. If $\Psi = I$, then $u \in \text{Ker}(\Psi_J^T)\backslash\{0\}$ is a sparse nonzero vector with maximal support J^c, so $r(J)$ is the inverse of the minimal singular value of the submatrix Φ_{J^c}. Below we claim Condition 1 ensures the robustness of problems (1b) and (1c) to arbitrary noise in b.

Theorem 2 (Robustness to measurement noise) *Under Assumptions 1-3, given an original signal $x^* \in \mathbb{R}^n$, let $I = \text{supp}(\Psi^T x^*)$ and $J = I^c$. For arbitrary noise w, let $b = \Phi x^* + w$ and $\delta = \|w\|_2$. If Condition 1 is met for $\bar{x} = x^*$, then*

1) For any $C_0 > 0$, there exists constant $C_1 > 0$ such that every minimizer $x_{\delta,\lambda}$ of problem (1b) using parameter $\lambda = C_0\delta$ satisfies

$$\|\Psi^T(x_{\delta,\lambda} - x^*)\|_1 \leq C_1\delta;$$

2) Every minimizer x_δ of problem (1c) satisfies

$$\|\Psi^T(x_\delta - x^*)\|_1 \leq C_2\delta.$$

The constraints c_1 and C_2 are given as follows. Define

$$\beta = (\Phi\Phi^T)^{-1}\Phi\Psi y, \quad C_3 = r(J)\sqrt{|I|} \quad and \quad C_4 = \frac{1 + \text{Cond}(\Psi)\|\Phi\|C_3}{1 - \|y_J\|_\infty},$$

with which we have

$$C_1 = 2C_3 + C_0\|\beta\|_2 + \frac{(1 + C_0\|\beta\|_2/2)^2 C_4}{C_0},$$

$$C_2 = 2C_3 + 2C_4\|\beta\|_2.$$

Remark 1 From the results of Theorem 2, it is straightforward to derive ℓ_1 or ℓ_2 bounds for $(x_{\delta,\lambda} - x^*)$ and $(x_\delta - x^*)$ under Assumption 2.

Remark 2 Since C_0 is free to choose, one can choose the optimal $C_0 = \sqrt{\frac{4C_4}{4\|\beta\|_2 + C_4\|\beta\|_2^2}}$ and simplify C_1 to

$$C_1 = 2C_3 + C_4\|\beta\|_2 + \sqrt{C_4^2\|\beta\|_2^2 + 4C_4\|\beta\|_2} \leq 2C_3 + 2C_4\|\beta\|_2 + 2,$$

which becomes very similar to C_2. This reflects the equivalence between problems (1b) and (1c) in the sense that given λ, one can find δ so that they have the same solution, and vice versa.

Remark 3 Both C_1 and C_2 are the sum of $2C_3$ and other terms. $2C_3$ alone bounds the error when $\Psi^T x_{\delta,\lambda}$ (or $\Psi^T x_\delta$) and $\Psi^T x^*$ have matching signs. Since C_3 does not depend on y_J, part (2) of Condition 1 does not play any role, whereas part (1) plays the major role. When the signs of $\Psi^T x_{\delta,\lambda}$ (or $\Psi^T x_\delta$) and $\Psi^T x^*$ do *not* match, the remaining terms in C_1 and C_2 are involved, and they are affected part (2) of Condition 1; in particular, $\|y_J\|_\infty < 1$ plays a big role as C_4 is inversely proportional to $1 - \|y\|_\infty$. Also, since there is no knowledge about the support of $\Psi^T x_{\delta,\lambda}$, which may or may not equal to that of $\Psi^T x^*$, C_4 inevitably depends the global properties of Ψ and Φ. In contrast, C_3 only depends on the restricted property of Φ.

When $\Psi^T x^*$ is only approximately sparse, we need to choose I as the support of the $|I|$ largest entries of $\Psi^T x^*$ in Condition 1. Then we have the following robustness result of approximately sparse signals for (1c).

Theorem 3 (Robustness for approximately sparse signals) *Under Assumption 3, given an original signal $x^* \in \mathbb{R}^n$ with $|I|$ largest entries of $\Psi^T x^*$ supported on I, let $J = I^c$. For arbitrary noise w, let $b = \Phi x^* + w$ and $\delta = \|w\|_2$. If Condition 1 is met on I, i.e., $\mathrm{Ker}(\Psi_J^T) \bigcap \mathrm{Ker}(\Phi) = \{0\}$, and there exist $y \in \mathbb{R}^l$ and $\beta \in \mathbb{R}^m$ such that $\Psi y = \Phi^T \beta$, and*

$$y_I = \mathrm{sign}(\Psi_I^T x^*), \ \|y_J\|_\infty < 1. \tag{3}$$

Then there exist ρ and τ such that

$$\|\Psi_I^T x\|_2 \le \rho \|\Psi_J^T x\|_1 + \tau \|\Phi x\|_2, \tag{4}$$

for all $x \in \mathbb{R}^n$, and every minimizer x_δ of $\|\Psi^T x\|_1$ subjective to $\|\Phi x - b\|_2 \le \delta$ satisfies

$$\|\Psi^T (x_\delta - x^*)\|_2 \le \frac{2(1+\rho)}{1 - \|y_J\|_\infty} \|\Psi_J^T x^*\|_1 + \left(\frac{2(1+\rho)\|\beta\|_2}{1 - \|y_J\|_\infty} + 2\tau \right) \delta. \tag{5}$$

Remark 4 With Assumption 3, part (1) of Condition 1 is equivalent to the existence of ρ and τ in (4) for all $x \in \mathbb{R}^n$ with ρ and τ. When part (1) of Condition 1 is satisfied, the existence of ρ and τ is easy to show. If part (1) is not satisfied, then there exist $x \ne 0$ such that $\Psi_J^T x = 0$ and $\Phi x = 0$. Assumption 3 gives us that $\Psi_I^T x \ne 0$. Thus $\|\Psi_I^T x\|_2 > 0 = \rho \|\Psi_J^T x\|_1 + \tau \|\Phi x\|_2$ for any ρ and τ.

Condition (3) can be relaxed into

$$\|y_I - \mathrm{sign}(\Psi_I^T x^*)\|_2 \le \theta_1, \ \|y_J\|_\infty < 1.$$

In this case, denote $\mu_1 := \rho\theta_1 + \|y_J\|_\infty, \mu_2 := \tau\theta_1 + \|\beta\|_2$. If $\mu_1 < 1$, then every minimizer x_δ of $\|\Psi^T x\|_1$ subjective to $\|\Phi x - b\|_2 \le \delta$ satisfies

$$\|\Psi^T (x_\delta - x^*)\|_2 \le \frac{2(1+\rho)}{1 - \mu_1} \|\Psi_J^T x^*\|_1 + \left(\frac{2(1+\rho)\mu_2}{1 - \mu_1} + 2\tau \right) \delta.$$

Remark 5 Theorem 3 is inspired by several existing results in the standard ℓ_1-synthesis case; for example Theorem 3.1 in [16] and Theorem 4.33 and Exercise 4.17 in [17]. The result here can be viewed as an extension from the ℓ_1-synthesis case to the ℓ_1-analysis case.

4 Related Works

In the case of $\Psi = Id$, Condition 1 is well known in the literature for ℓ_1 (or ℓ_1-synthesis) minimization. It is initially proposed in [2] as a sufficient condition for the ℓ_1 solution uniqueness. For problems (1b) and (1c), [26, 27] present sufficient but non-necessary conditions for solutions uniqueness. Later, its necessity is established in [14] for model (1b) and then in [15] for all models in (1), assuming $\Psi = Id$ or equal to an orthogonal basis. The solution robustness of model (1b) is given under the same condition in [14]. Below we restrict our literature review to results for the ℓ_1-analysis model.

4.1 Previous Uniqueness Conditions

Papers [7,18,8,9,28] establish the uniqueness of the ℓ_1-analysis model and some use stronger conditions than ours. Our purpose in the comparison is restricted to uniqueness and robustness to arbitrary noise. The reader should be aware that some of them also imply stronger results such as sign consistency by Conditions 4 and 6 below; while the rest such as Conditions 3 and 5 below are not used to derive stronger results yet.

The following condition in [18] guarantees the solution uniqueness for problem (1a):

Condition 3 *Given $\bar{x}$, let Q be a basis matrix of $\mathrm{Ker}(\Phi)$, and $I = \mathrm{supp}(\Psi^T \bar{x})$. The followings are met:*
(1) $\Psi_{I^c}^T Q$ is full column rank;
(2)$\|(Q^T \Psi_{I^c})^+ Q^T \Psi_I \mathrm{sign}(\Psi_I^T \bar{x})\|_\infty < 1$.

Here A^+ is the pseudo-inverse of matrix A defined as $A^+ = A^T(AA^T)^{-1}$.

Paper [8] proposes the following condition for the solution uniqueness and robustness for problems (1a) and (1b) (the robustness requires the non-zero entries of $\Psi_I^T x$ to be sufficiently large compared to noise).

Condition 4 *For a given $\bar{x}$, index sets $I = \mathrm{supp}(\Psi^T \bar{x})$ and $J = I^c$ satisfy:*
(1)$\mathrm{Ker}(\Psi_J^T) \bigcap \mathrm{Ker}(\Phi) = \{0\}$;
(2)Let $A^{[J]} = U(U^T \Phi^T \Phi U)^{-1}U^T$ and $\Omega^{[J]} = \Psi_J^+(\Phi^T \Phi A^{[J]} - Id)\Psi_I$, where U is a basis matrix of $\mathrm{Ker}(\Psi_J^T)$.
Then

$$IC(\mathrm{sign}(\Psi_I^T \bar{x})) := \min_{u \in \mathrm{Ker}(\Psi_J)} \|\Omega^{[J]}\mathrm{sign}(\Psi_I^T \bar{x}) - u\|_\infty < 1.$$

According to [8], Conditions 3 and 4 do not contain each other.

The following example shows that Conditions 3 and 4 are both stronger than Conditions 1 and 2. Let

$$\Psi = \begin{pmatrix} 10.5 & 1 & 10 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}, \quad \Phi = \begin{pmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}, \quad \hat{x} = \begin{pmatrix} 1 \\ -1 \\ -10 \end{pmatrix}, \quad b = \begin{pmatrix} -1 \\ -10 \end{pmatrix}.$$

It is straightforward to verify that Conditions 1 and 2 hold. However, Conditions 3 and 4 fail to hold. Indeed, we have $\Psi^T \tilde{x} = (10.5, 0, 0)^T$ and $I = \{1\}$. $Q = (1, 0, 0)^T$ is a basis matrix of $\mathrm{Ker}(\Phi)$. Thus

$$\|(Q^T \Psi_{I^c})^+ Q^T \Psi_I \mathrm{sign}(\Psi_I^T \tilde{x})\|_\infty = \left\| \left(\frac{10.5}{101}, \frac{105}{101}\right)^T \right\|_\infty = \frac{105}{101}.$$

Hence, Condition 3 does not hold. Furthermore, $U = (1, -1, -10)^T$ is a basis matrix of $\mathrm{Ker}(\Psi_J^T)$, and the definition of $\Omega^{[J]}$ gives us $\Omega^{[J]} = (\frac{10.5}{101}, \frac{105}{101})^T$. Therefore, $IC(\mathrm{sign}(\Psi_I^T \tilde{x})) = \frac{105}{101} > 1$, so Condition 4 does not hold either. Paper [8] also presents sufficient conditions for solution uniqueness, which are reviewed in [15] and shown to be not necessary.

During the time we are preparing this manuscript, work [28] gives the same condition as Condition 1 and shows its sufficiency. However. The necessity of Condition 1 or 2 is never discussed in the literature.

4.2 Previous Robustness Conditions

Turning to solution robustness, [7,9,28] have studied the robustness of problems (1b) and (1c) in the Hilbert-space setting. Translating to the finite dimension, the condition in [7] is equivalent to Condition 2. Under Condition 2, work [7] shows the existence of constant C (not explicitly given) such that the solution $x_{\delta,\lambda}$ to (1b) obeys $\|\Psi^T(x_{\delta,\lambda} - x^*)\|_2 \leq C\delta$ when λ is set proportional to the noise level δ. [28] gives an explicit formula of C in $\|x_{\delta,\lambda} - x^*\|_2 \leq C\delta$ for solution $x_{\delta,\lambda}$ to (1b). In order to obtain an explicit formula for C, [9] introduces the following:

Condition 5 *Assume that $\Psi\Psi^T$ is invertible and let $\hat{\Psi} = (\Psi\Psi^T)^{-1}\Psi$. Given $\bar{x}$, the following two statements hold:*

(1) There exists some $y \in \partial\|\cdot\|_1(\Psi^T\bar{x})$ such that $\Psi y \in \text{Im}(\Phi^T)$;

(2) For some $t \in (0,1)$, letting $I(t) = \{i : |y_i| > t\}$, the mapping $\hat{\Phi} := \Phi|_{\text{Span}\{\hat{\Psi}_i : i \in I(t)\}}$ is injective.

Under this condition, the solutions to (1b) and (1c) are subject to error bounds whose constants depend on t, $\hat{\Phi}$, and other quantities.

Proposition 1 *Condition 5 is stronger than Condition 2.*

Proof Let $J = I(t)^c$; then we have $\|y_J\|_\infty \leq t < 1$ from the definition of $I(t)$. It remains to show that $\text{Ker}(\Psi_J^T) \bigcap \text{Ker}(\Phi) = \{0\}$. For any $x \in \text{Ker}(\Psi_J^T)$, we have

$$x = (\Psi\Psi^T)^{-1}\Psi\Psi^T x = (\Psi\Psi^T)^{-1}\Psi_J\Psi_J^T x + (\Psi\Psi^T)^{-1}\Psi_{J^c}\Psi_{J^c}^T x = (\Psi\Psi^T)^{-1}\Psi_{I(t)}\Psi_{I(t)}^T x.$$

Since Φ restricted to $\text{Span}\{\hat{\Psi}_i : i \in I(t)\} = \text{Im}((\Psi\Psi^T)^{-1}\Psi_{I(t)})$ is injective, we have what we need. $\square$

Definition 5 in paper [8] provides a much stronger condition below that strengthens Condition 4 by dropping the dependence on the Ψ-support (see the definition of $RC(I)$ below).

Condition 6 *Given $\bar{x}$, index sets $I = \text{supp}(\Psi^T\bar{x})$ and $J = I^c$ satisfy:*

(1) $\text{Ker}(\Psi_J^T) \bigcap \text{Ker}(\Phi) = \{0\}$;

(2) Letting $\Omega^{[J]}$ be given as in Condition 4,

$$RC(I) := \max_{p \in \mathbb{R}^{|I|}, \|p\|_\infty \leq 1} \min_{u \in Ker(\Psi_J)} \|\Omega^{[J]}p - u\|_\infty < 1.$$

Under this condition, a nice error bound and a certain kind of "weak" sign consistency (between $\Psi^T x_{\delta,\lambda}$ and $\Psi^T x^*$) are given provided that problem (1b) is solved with the parameter $\lambda = \frac{\rho\|w\|_2 c_J}{2(1-RC(I))}$ for some $\rho > 1$, where $c_J = \|\Psi_J^+\Phi^T(\Phi A^{[J]}\Phi^T - Id)\|_{2,\infty}$. When $1 - RC(I)$ gets close to 0, this λ can become too large than it should be.

5 Verifying the Conditions

In this section, we present a method to verify Condition 1. Our method includes two steps:

(**Step 1:**) Let $\Phi = U\Sigma V^T$ be the singular value decomposition of Φ. Assume $V = [v_1, \cdots, v_n]$. Since Φ has full row-rank, we have $\text{Ker}(\Phi) = \text{Span}\{v_{m+1}, \cdots, v_n\}$ and $Q = [v_{m+1}, \cdots, v_n]$ as a basis of $\text{Ker}(\Phi)$. We verify that $\Psi_J^T Q$ has full row-rank, ensuring part (1) of Condition 1.

(**Step 2:**) Let $u_1 = -Q^T\Psi_I\text{sign}(\Psi_I^T\bar{x})$ and $A = Q^T\Psi_J$. Solve the convex problem

$$\underset{u \in \mathbb{R}^{|J|}}{\text{minimize}} \|u\|_\infty, \quad \text{subject to } Au = u_1. \tag{6}$$

If the optimal objective of (6) is strictly less than 1, then part (2) of Condition 1 holds. In fact, we have:

Proposition 2 *Part (2) of Condition 1 holds if and only if (6) has an optimal objective < 1.*

Proof Let $\hat{u}$ be a minimizer of (6). Assume $\|\hat{u}\|_\infty < 1$. We consider the vector y composed by $y_I = \text{sign}(\Psi_I^T\bar{x})$ and $y_J = \hat{u}$. To show part (2) of Condition 1, it suffices to prove $\Psi y \in \text{Im}(\Phi^T)$, or equivalently, $Q^T\Psi y = 0$. Indeed,

$$Q^T\Psi y = Q^T\Psi_J y_J + Q^T\Psi_I y_I = Q^T\Psi_J\hat{u} + Q^T\Psi_I y_I = 0.$$

The converse is obvious. $\square$

Convex program (6) is similar in form to one in [8, Definition 4] though they are used to verify different conditions.

6 Proof of Theorem 1

We establish Theorem 1 in two steps. Our first step proves the theorem for problem (1a) only. The second step proves Theorem 1 for problems (1b) and (1c).

6.1 Proof of Theorem 1 for problem (1a)

The equivalence of the three statements is shown in the following orders: 3) $\Longrightarrow$ 1) $\Longrightarrow$ 2) $\Longrightarrow$ 3).

3) $\Longrightarrow$ 1). Consider any perturbation $\hat{x} + h$ where $h \in \mathrm{Ker}(\Phi)\backslash\{0\}$. Take a subgradient $g \in \partial\|\cdot\|_1(\Psi^T\hat{x})$ obeying $g_I = \mathrm{sign}(\Psi_I^T\hat{x}) = y_I$, $g_K = y_K$, and $\|g_J\|_\infty \leq 1$ such that $\langle g_J, \Psi_J^T h\rangle = \|\Psi_J^T h\|_1$. Then,

$$\|\Psi^T(\hat{x}+h)\|_1 \geq \|\Psi^T\hat{x}\|_1 + \langle\Psi g, h\rangle \tag{7a}$$

$$= \|\Psi^T\hat{x}\|_1 + \langle\Psi g - \Psi y, h\rangle \tag{7b}$$

$$= \|\Psi^T\hat{x}\|_1 + \langle g - y, \Psi^T h\rangle \tag{7c}$$

$$= \|\Psi^T\hat{x}\|_1 + \langle g_J - y_J, \Psi_J^T h\rangle \tag{7d}$$

$$\geq \|\Psi^T\hat{x}\|_1 + \|\Psi_J^T h\|_1(1 - \|y_J\|_\infty), \tag{7e}$$

where (7b) follows from $\Psi y \in \mathrm{Im}(\Phi^T) = \mathrm{Ker}(\Phi)^\perp$ and $h \in \mathrm{Ker}(\Phi)$, (7d) follows from the setting of g, and (7e) is an application of the inequality $\langle x, y\rangle \leq \|x\|_1\|y\|_\infty$ and $\langle g_J, \Psi_J^T h\rangle = \|\Psi_J^T h\|_1$. Since $h \in \mathrm{Ker}(\Phi)\backslash\{0\}$ and $\mathrm{Ker}(\Psi_J^T)\bigcap\mathrm{Ker}(\Phi) = \{0\}$, we have $\|\Psi_J^T h\|_1 > 0$. Together with the condition $\|y_J\|_\infty < 1$, we have $\|\Psi^T(\hat{x}+h)\|_1 > \|\Psi^T\hat{x}\|_1$ for every $h \in \mathrm{Ker}(\Phi)\backslash\{0\}$ which implies that $\hat{x}$ is the unique minimizer of (1a).

1) $\Longrightarrow$ 2). For every $h \in \mathrm{Ker}(\Phi)\backslash\{0\}$, we have $\Phi(\hat{x}+th) = \Phi\hat{x}$ and can find t small enough around 0 such that $\mathrm{sign}(\Psi_I^T(\hat{x}+th)) = \mathrm{sign}(\Psi_I^T\hat{x})$. Since $\hat{x}$ is the unique solution, for small and nonzero t we have

$$\|\Psi^T(\hat{x})\|_1 < \|\Psi^T(\hat{x}+th)\|_1 = \|\Psi_I^T(\hat{x}+th)\|_1 + \|\Psi_{I^c}^T(\hat{x}+th)\|_1$$

$$= \langle\Psi_I^T(\hat{x}+th), \mathrm{sign}(\Psi_I^T(\hat{x}+th))\rangle + \|t\Psi_{I^c}^T h\|_1$$

$$= \langle\Psi_I^T\hat{x} + t\Psi_I^T h, \mathrm{sign}(\Psi_I^T\hat{x})\rangle + \|t\Psi_{I^c}^T h\|_1$$

$$= \langle\Psi_I^T\hat{x}, \mathrm{sign}(\Psi_I^T\hat{x})\rangle + t\langle\Psi_I^T h, \mathrm{sign}(\Psi_I^T\hat{x})\rangle + \|t\Psi_{I^c}^T h\|_1$$

$$= \|\Psi^T(\hat{x})\|_1 + t\langle\Psi_I^T h, \mathrm{sign}(\Psi_I^T\hat{x})\rangle + \|t\Psi_{I^c}^T h\|_1.$$

Therefore, for any $h \in \mathrm{Ker}(\Phi)\backslash\{0\}$, we have

$$\langle\Psi_I^T h, \mathrm{sign}(\Psi_I^T\hat{x})\rangle < \|\Psi_{I^c}^T h\|_1. \tag{9}$$

If the condition $\mathrm{Ker}(\Psi_{I^c}^T)\bigcap\mathrm{Ker}(\Phi) = \{0\}$ does not hold, we can choose a nonzero vector $h \in \mathrm{Ker}(\Psi_{I^c}^T)\bigcap\mathrm{Ker}(\Phi)$. We also have $-h \in \mathrm{Ker}(\Psi_{I^c}^T)\bigcap\mathrm{Ker}(\Phi)$. Then we have $\langle\Psi_I^T h, \mathrm{sign}(\Psi_I^T\hat{x})\rangle < 0$ and $-\langle\Psi_I^T h, \mathrm{sign}(\Psi_I^T\hat{x})\rangle < 0$, which is a contradiction.

It remains to show the existence of y in item (2) of Condition 1. This part is in spirit of the methods in papers [14] and [15], which are based on linear programming strong duality. We take $\hat{y}$ with restrictions $\hat{y}_I = \mathrm{sign}(\Psi_I^T\hat{x})$ and $\hat{y}_{I^c} = 0$. If such $\hat{y}$ satisfies $\Psi\hat{y} \in \mathrm{Im}(\Phi^T)$, then the existence has been shown. If $\Psi\hat{y} \notin \mathrm{Im}(\Phi^T) = \mathrm{Ker}(\Phi)^\perp$, then we shall construct a new vector to satisfy part (2) of Condition 1. Let Q be a basis matrix of $\mathrm{Ker}(\Phi)$. We have that $a := Q^T\Psi\hat{y}$ must be a nonzero vector. Consider the following problem

$$\underset{z\in\mathbb{R}^l}{\mathrm{minimize}} \|z\|_\infty \quad \text{subject to } Q^T\Psi z = -a \text{ and } z_I = 0. \tag{10}$$

For any minimizer $\hat{z}$ of problem (10), we have $\Psi(\hat{y} + \hat{z}) \in \text{Ker}(\Phi)^\perp = \text{Im}(\Phi^T)$ and $(\hat{y} + \hat{z})_I = \hat{y}_I = \text{sign}(\Psi_I^T \hat{x})$. Thus, we shall show that the objective of problem (10) is strictly less than 1. To this end, we rewrite problem (10) in an equivalent form as:

$$\underset{z}{\text{minimize}} \, \|z_{I^c}\|_\infty \quad \text{subject to } Q^T \Psi_{I^c} z_{I^c} = -a, \tag{11}$$

whose Lagrange dual problem is

$$\underset{p}{\text{maximize}} \langle p, a \rangle \quad \text{subject to } \|\Psi_{I^c}^T Qp\|_1 \leq 1. \tag{12}$$

Note that $Qp \in \text{Ker}(\Phi)$ and $|\langle p, a \rangle| = |\langle p, Q^T \Psi \hat{y} \rangle| = |\langle p, Q^T \Psi_I \text{sign}(\Psi_I^T \hat{x}) \rangle| = |\langle \Psi_I^T Qp, \text{sign}(\Psi_I^T \hat{x}) \rangle|$. By using (9), for any p we have

$$|\langle p, a \rangle| = \begin{cases} |\langle \Psi_I^T Qp, \text{sign}(\Psi_I^T \hat{x}) \rangle| = 0, & \text{if } Qp = 0; \\ |\langle \Psi_I^T Qp, \text{sign}(\Psi_I^T \hat{x}) \rangle| < \|\Psi_{I^c}^T Qp\|_1 \leq 1, \text{otherwise.} \end{cases}$$

Hence, problem (12) is feasible, and its objective value is strictly less than 1. By the linear programming strong duality property, problems (10) and (11) also have solutions, and their the objective value is strictly less than 1, too. This completes the proof.

2) $\implies$ 3). Let $J = I^c$ and $K = \emptyset$; then Condition 2 follows.

The proof of 3) $\implies$ 1) is a standard technique in compressed sensing community.

6.2 Proof of Theorem 1 for problems (1b) and (1c)

Lemma 1 *Let $\gamma > 0$. If $\gamma\|\Phi x - b\|_2^2 + \|\Psi^T x\|_1$ is constant on a convex set Ω, then both $\Phi x - b$ and $\|\Psi^T x\|_1$ are constant on Ω.*

Proof It suffices to prove the case where the convex set has more than one point. Suppose x_1 and x_2 are arbitrary two *different* points in Ω. Consider the line segment L connecting x_1 and x_2. By the convexity of set Ω, we know $L \subset \Omega$. Thus, $\hat{c} = \gamma\|\Phi x - b\|_2^2 + \|\Psi^T x\|_1$ is a constant on L. If $\Phi x_1 - b \neq \Phi x_2 - b$, then for any $0 < \alpha < 1$, we have

$$\gamma\|\Phi(\alpha x_1 + (1 - \alpha)x_2) - b\|_2^2 + \|\Psi^T(\alpha x_1 + (1 - \alpha)x_2)\|_1 \tag{13a}$$

$$= \gamma\|\alpha(\Phi x_1 - b) + (1 - \alpha)(\Phi x_2 - b)\|_2^2 + \|\alpha(\Psi^T x_1) + (1 - \alpha)(\Psi^T x_2)\|_1 \tag{13b}$$

$$< \alpha(\gamma\|\Phi x_1 - b\|_2^2 + \|\Psi^T x_1\|_1) + (1 - \alpha)(\gamma\|\Phi x_2 - b\|_2^2 + \|\Psi^T x_2\|_1) \tag{13c}$$

$$= \alpha\hat{c} + (1 - \alpha)\hat{c} = \hat{c}, \tag{13d}$$

where the strict inequality follows from the *strict* convexity of $\gamma\|\cdot\|_2^2$ and the convexity of $\|\Psi^T x\|_1$. This means that the points $\alpha x_1 + (1 - \alpha)x_2$ on L attain a lower value than $\hat{c}$, which is a contradiction. Therefore, we have $\Phi x_1 - b = \Phi x_2 - b$, from which it is easy to see $\|\Psi^T x_1\|_1 = \|\Psi^T x_2\|_1$. $\qquad\square$

We let X_λ and Y_δ denote the sets of solutions to problems (1b) and (1c), respectively; moreover, we assume that these two sets are nonempty. Then, from Lemma 1, we have the following result.

Corollary 1 *In problem (1b), $\Phi x - b$ and $\|\Psi^T x\|_1$ are constant on X_λ; in problem (1c), $\Phi x - b$ and $\|\Psi^T x\|_1$ are constant on Y_δ.*

Proof Since $\|\Phi x - b\|_2^2 + \lambda\|\Psi^T x\|_1$ is constant over X_λ, the result follows directly from Lemma 1 for problem (1b). For problem (1c), if $0 \in Y_\delta$, then we have $Y_\delta = \{0\}$ because of the full row-rankness of Ψ. The result holds trivially. Suppose $0 \notin Y_\delta$. Since the optimal objective $\|\Psi^T x\|_1$ is constant for all $x \in Y_\delta$, we have to show that $\|\Phi x - b\|_2^2 = \delta$ for all $x \in Y_\delta$. If there exist a nonzero $\hat{x} \in Y_\delta$ such that $\|\Phi \hat{x} - b\|_2^2 < \delta$, we can find a non-empty ball $\mathcal{B}$ centered at $\hat{x}$ with a sufficiently small radius $\rho > 0$ such that $\|\Phi \tilde{x} - b\|_2^2 < \delta$ for all $\tilde{x} \in \mathcal{B}$. Let $\alpha = \min\{\frac{\rho}{2\|\hat{x}\|_2}, \frac{1}{2}\} \in (0,1)$. We have $(1-\alpha)\hat{x} \in \mathcal{B}$ and $\|(1-\alpha)\Psi^T \hat{x}\|_1 < \|\Psi^T \hat{x}\|_1$, which is a contradiction. $\qquad\square$

Proof (Proof of Theorem 1 for problems (1b) and (1c)) This proof exploits Corollary 1. Since the results of Corollary 1 are identical for problems (1b) and (1c), we present the proof for problem (1b) only.

By assumption, X_λ is nonempty so we pick $\hat{x} \in X_\lambda$. Let $b^* = \Phi\hat{x}$, which is independent of the choice of $\hat{x}$ according to Corollary 1. We introduce the following problem

$$\underset{x}{\text{minimize}} \|\Psi^T x\|_1, \quad \text{subject to } \Phi x = b^*, \tag{14}$$

and let X^* denote its solution set.

Now, we show that $X_\lambda = X^*$. Since $\Phi x = \Phi\hat{x}$ and $\|\Psi^T x\|_1 = \|\Psi^T \hat{x}\|_1$ for all $x \in X_\lambda$ and conversely any x obeying $\Phi x = \Phi\hat{x}$ and $\|\Psi^T x\|_1 = \|\Psi^T \hat{x}\|_1$ belongs to X_λ, it suffices to show that $\|\Psi^T x\|_1 = \|\Psi^T \hat{x}\|_1$ for any $x \in X^*$. Assuming this does *not* hold, then since problem (14) has $\hat{x}$ as a feasible solution and has a finite objective, we have a nonempty X^* and there exists $\tilde{x} \in X^*$ satisfying $\|\Psi^T \tilde{x}\|_1 < \|\Psi^T \hat{x}\|_1$. But, $\|\Phi\tilde{x} - b\|_2 = \|b^* - b\|_2 = \|\Phi\hat{x} - b\|_2$ and $\|\Psi^T \tilde{x}\|_1 < \|\Psi^T \hat{x}\|_1$ mean that $\tilde{x}$ is a strictly better solution to problem (1b) than $\hat{x}$, contradicting the assumption $\hat{x} \in X_\lambda$.

Since $X_\lambda = X^*$, $\hat{x}$ is the unique solution to problem (1b) if and only if it is the unique solution to problem (14). Since problem (14) is in the same form of problem (1a), applying the part of Theorem 1 for problem (1a), which is already proved, we conclude that $\hat{x}$ is the unique solution to problem (1b) if and only if Condition 1 or 2 holds. $\qquad\square$

7 Proof of Theorem 2

Lemma 2 *Assume that vectors $\bar{x}$ and y satisfy Condition 1. Let $I = \text{supp}(\Psi^T \bar{x})$ and $J = I^c$. We have*

$$\|\Psi^T x - \Psi^T \bar{x}\|_1 \leq C_3 \|\Phi(x - \bar{x})\|_2 + C_4 d_y(x, \bar{x}), \quad \forall x, \tag{15}$$

where $d_y(x, \bar{x}) := \|\Psi^T x\|_1 - \|\Psi^T \bar{x}\|_1 - \langle \Psi y, x - \bar{x}\rangle$ is the Bregman distance of function $\|\Psi^T \cdot\|_1$, the absolute constants C_3, C_4 are given in Theorem 2.

Proof This proof is divided into two parts. They are partially inspired by [7].

1. this part shows that for any $u \in \text{Ker}(\Psi_J^T)$,

$$\|\Psi^T x - \Psi^T \bar{x}\|_1 \leq \left(1 + \frac{C_3\|\Phi\|}{\sqrt{\lambda_{\min}(\Psi\Psi^T)}}\right)\|\Psi^T(x - u)\|_1 + C_3\|\Phi(x - \bar{x})\|_2. \tag{16}$$

2. this part shows that

$$f(x) := \min\left\{\|\Psi^T(x - u)\|_1 : u \in \text{Ker}(\Psi_J^T)\right\} \leq (1 - \|y_J\|_\infty)^{-1} d_y(x, \bar{x}). \tag{17}$$

Using the definition of C_4, combining (16) and (17) gives (15).

Part 1. Let $u \in \mathrm{Ker}(\Psi_J^T)$. By the triangle inequality of norms, we get

$$\|\Psi^T x - \Psi^T \bar{x}\|_1 \le \|\Psi^T (x - u)\|_1 + \|\Psi^T (u - \bar{x})\|_1. \tag{18}$$

Since $\bar{x} \in \mathrm{Ker}(\Psi_J^T)$, we have $u - \bar{x} \in \mathrm{Ker}(\Psi_J^T)$ and thus $\|u - \bar{x}\|_2 \le r(J)\|\Phi(u - \bar{x})\|_2$, where $r(J) < +\infty$ follows from part (1) of Condition 1. Using the fact that $\mathrm{supp}(\Psi^T(u - \bar{x})) = I$, we derive that

$$\|\Psi^T (u - \bar{x})\|_1 \le \sqrt{|I|}\|\Psi^T (u - \bar{x})\|_2 \tag{19a}$$

$$\le \sqrt{|I|}\|u - \bar{x}\|_2 \tag{19b}$$

$$\le \sqrt{|I|}\, r(J)\|\Phi(u - \bar{x})\|_2 \tag{19c}$$

$$= C_3 \|\Phi(u - \bar{x})\|_2, \tag{19d}$$

where we have used the assumption $\lambda_{\max}(\Psi\Psi^T) = 1$ and the definition $C_3 = r(J)\sqrt{|I|}$. Furthermore,

$$\|\Phi(u - \bar{x})\|_2 \le \|\Phi(x - u)\|_2 + \|\Phi(x - \bar{x})\|_2 \tag{20a}$$

$$\le \|\Phi\|\|x - u\|_2 + \|\Phi(x - \bar{x})\|_2 \tag{20b}$$

$$\le \frac{\|\Phi\|\|\Psi^T(x - u)\|_2}{\sqrt{\lambda_{\min}(\Psi\Psi^T)}} + \|\Phi(x - \bar{x})\|_2 \tag{20c}$$

$$\le \frac{\|\Phi\|\|\Psi^T(x - u)\|_1}{\sqrt{\lambda_{\min}(\Psi\Psi^T)}} + \|\Phi(x - \bar{x})\|_2. \tag{20d}$$

Therefore, we get (16) after combining (18), (19), and (20).

Part 2. Since $\langle \Psi y, \bar{x} \rangle = \|\Psi^T \bar{x}\|_1$ implies $d_y(x, \bar{x}) = \|\Psi^T x\|_1 - \langle \Psi y, x \rangle$, it is equivalent to proving

$$f(x) \le (1 - \|y_J\|_\infty)^{-1}(\|\Psi^T x\|_1 - \langle \Psi y, x \rangle).$$

Since $u \in \mathrm{Ker}(\Psi_J^T)$ is equivalent to $\Psi_J^T u = 0$, the Lagrangian of the minimization problem in (17) is

$$L(u, v) = \|\Psi^T(x - u)\|_1 + \langle v, \Psi_J^T u \rangle = \|\Psi^T(x - u)\|_1 + \langle \Psi_J v, u - x \rangle + \langle \Psi_J v, x \rangle.$$

Then, $f(x) = \min_u \max_v L(u, v)$. Following the minimax theorem, we derive that

$$\begin{aligned} f(x) &= \max_v \min_u L(u, v) \\ &= \max_v \min_u \{\|\Psi^T(x - u)\|_1 + \langle \Psi_J v, u - x \rangle + \langle \Psi_J v, x \rangle\} \\ &= \max_w \min_u \{\|\Psi^T(x - u)\|_1 + \langle w, u - x \rangle + \langle w, x \rangle : w \in \mathrm{Im}(\Psi_J)\} \\ &= \max_w \{\langle w, x \rangle : w \in \partial\|\Psi^T \cdot \|_1(0) \cap \mathrm{Im}(\Psi_J)\} \\ &= \max_w \{\langle c\Psi y + w, x \rangle : w \in \partial\|\Psi^T \cdot \|_1(0) \cap \mathrm{Im}(\Psi_J)\} - \langle c\Psi y, x \rangle, \ \forall c > 0 \\ &= c \max_w \{\langle \Psi y + w, x \rangle : w \in c^{-1}\partial\|\Psi^T \cdot \|_1(0) \cap \mathrm{Im}(\Psi_J)\} - c\langle \Psi y, x \rangle, \ \forall c > 0. \end{aligned}$$

Let

$$c = (1 - \|y_J\|_\infty)^{-1}$$

and $Z_J = \{z \in \mathbb{R}^l : z_I = 0\}$. Since $\|y_J\|_\infty < 1$ from part (2) of Condition 1, we have $c < +\infty$ and get

$$\left(y + c^{-1}\partial\|\cdot\|_1(0) \cap Z_J\right) \subset \partial\|\cdot\|_1(0),$$

from which we conclude

$$\left(\Psi y + c^{-1}\partial\|\Psi^T \cdot \|_1(0) \cap \mathrm{Im}(\Psi_J)\right) \subset \partial\|\Psi^T \cdot \|_1(0).$$

Hence, for any $w \in c^{-1}\partial\|\Psi^T \cdot \|_1(0) \cap \mathrm{Im}(\Psi_J)$, it holds $\Psi y + w \subset \partial\|\Psi^T \cdot \|_1(0)$, which by the convexity of $\|\Psi^T \cdot \|_1$ implies $\langle \Psi y + w, x \rangle \le \|\Psi^T x\|_1$. Therefore, $f(x) \le c(\|\Psi^T x\|_1 - \langle \Psi y, x \rangle)$. $\qquad \square$

Lemma 3 *([13], Theorem 3; [14], Lemma 3.5) Suppose that $x^* \in \mathbb{R}^n$ is a fixed vector obeying $\mathrm{supp}(\Psi^T x^*) = I$ and that there are vectors satisfying $y \in \partial\|\cdot\|_1(\Psi^T x^*)$ and $\Psi y = \Phi^T \beta$. Then for every $\delta > 0$ and every data vector b satisfying $\|\Phi x^* - b\|_2 \le \delta$, the following two statements hold:*

1)Every minimizer $x_{\delta,\lambda}$ of problem (1b) satisfies $d_y(x_{\delta,\lambda}, x^*) \le \frac{(\delta + \lambda\|\beta\|_2/2)^2}{\lambda}$ and $\|\Phi x_{\delta,\lambda} - b\|_2 \le \delta + \lambda\|\beta\|_2$;

2)Every minimizer x_δ of problem (1c) satisfies $d_y(x_\delta, x^*) \le 2\delta\|\beta\|_2$.

From $\Psi y = \Phi^T \beta$ and the full-rankness of Φ, we have $\beta = (\Phi\Phi^T)^{-1}\Phi\Psi y$.

Proof (Proof of Theorem 2) Firstly, we derive that

$$\|\Psi(x_{\delta,\lambda} - x^*)\|_1 \le C_3 \|\Phi(x_{\delta,\lambda} - x^*)\|_2 + C_4 d_y(x_{\delta,\lambda}, x^*) \tag{22a}$$

$$\le C_3\|\Phi x_{\delta,\lambda} - b\|_2 + C_3\|\Phi x^* - b\|_2 + C_4 d_y(x_{\delta,\lambda}, x^*) \tag{22b}$$

$$\le C_3(\delta + \lambda\|\beta\|_2) + C_3\delta + C_4\frac{(\delta + \lambda\|\beta\|_2/2)^2}{\lambda}, \tag{22c}$$

where the first and the third inequalities follow from Lemmas 2 and 3, respectively. Substituting $\lambda = C_0\delta$ and collecting like terms in (22c), we obtain the first part of Theorem 2. The second part can be proved in the same way. $\qquad\square$

8 Proof of Theorem 3

Let $w := x_\delta - x^*, v := \Psi^T w$. Then by the minimality of $\|\Psi^T x_\delta\|_1$, we derive

$$\begin{aligned}
\|\Psi^T x^*\|_1 \ge \|\Psi^T x_\delta\|_1 &= \|v + \Psi^T x^*\|_1 = \|v_I + \Psi_I^T x^*\|_1 + \|v_J + \Psi_J^T x^*\|_1 \\
&\ge \langle v_I + \Psi_I^T x^*, \mathrm{sign}(\Psi_I^T x^*)\rangle + \|v_J\|_1 - \|\Psi_J^T x^*\|_1 \\
&= \langle v_I, \mathrm{sign}(\Psi_I^T x^*)\rangle + \|\Psi_I^T x^*\|_1 + \|v_J\|_1 - \|\Psi_J^T x^*\|_1.
\end{aligned}$$

Rearranging the above inequality and using the fact that $\|\Psi^T x^*\|_1 = \|\Psi_I^T x^*\|_1 + \|\Psi_J^T x^*\|_1$ yield

$$\|v_J\|_1 \le 2\|\Psi_J^T x^*\|_1 + |\langle v_I, \mathrm{sign}(\Psi_I^T x^*)\rangle|. \tag{24}$$

In addition, we have

$$|\langle v_I, \mathrm{sign}(\Psi_I^T x^*)\rangle| = |\langle y_I, v_I\rangle| \le |\langle y, v\rangle| + |\langle y_J, v_J\rangle|.$$

In what follows, we further bound the right-hand side of the above inequality term by term. First of all, in view of the optimization constraint and the feasible of both x_δ and x^*, we have

$$\|\Phi w\|_2 = \|\Phi(x_\delta - x^*)\|_2 \le \|\Phi x_\delta - b\|_2 + \|\Phi x^* - b\|_2 \le 2\delta.$$

Using the condition $\Psi y = \Phi^T \beta$, we obtain

$$\begin{aligned}
|\langle y, v\rangle| = |\langle y, \Psi^T w\rangle| &= |\langle \Psi y, w\rangle| = |\langle \Phi^T \beta, w\rangle| \\
&= |\langle \beta, \Phi w\rangle| \le \|\beta\|_2 \cdot \|\Phi w\|_2 \le 2\delta\|\beta\|_2.
\end{aligned}$$

At last, $|\langle y_J, v_J\rangle| \le \|y_J\|_\infty \|v_J\|_1$. Collecting these upper bounds, we derive

$$|\langle v_I, \mathrm{sign}(\Psi_I^T x^*)\rangle| \le \|y_J\|_\infty \|v_J\|_1 + 2\|\beta\|_2\delta$$

Using (24) and noting that $\|y_J\|_\infty < 1$, we get

$$\|v_J\|_1 \le \frac{2}{1 - \|y_J\|_\infty}\|\Psi_J^T x^*\|_1 + \frac{2\|\beta\|_2}{1 - \|y_J\|_\infty}\delta. \tag{26}$$

Now, using the inequality (4), we get

$$\|v_I\|_2 \leq \rho\|v_J\|_1 + \tau\|\Phi w\|_2 \leq \frac{2\rho}{1 - \|y_J\|_\infty}\|\Psi_J^T x^*\|_1 + \left(\frac{2\rho\|\beta\|_2}{1 - \|y_J\|_\infty} + 2\tau\right)\delta. \qquad (27)$$

Finally, combing (26) and (27), we obtain

$$\|v\|_2 \leq \|v_I\|_2 + \|v_J\|_2 \leq \|v_I\|_2 + \|v_J\|_1$$
$$\leq \frac{2(1 + \rho)}{1 - \|y_J\|_\infty}\|\Psi_J^T x^*\|_1 + \left(\frac{2(1 + \rho)\|\beta\|_2}{1 - \|y_J\|_\infty} + 2\tau\right)\delta.$$

This completes the proof. $\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad$ □

Acknowledgements

References

1. Michael Elad, Peyman Milanfar, and Ron Rubinstein. Analysis versus synthesis in signal priors. *Inverse Problems*, 23(3):947, 2007. 1, 2.2

2. J. J. Fuchs. On sparse representations in arbitrary redundant bases. *IEEE Transactions on Information Theory*, 50(6):1341–1344, June 2004. 1, 4

3. Emmanuel J. Candes and Yaniv Plan. A probabilistic and RIPless theory of compressed sensing. *IEEE Transactions on Information Theory*, 57(11):7235–7254, November 2011. 1

4. E. Candes. The restricted isometry property and its implications for compressed sensing. *Comptes Rendus Mathematique*, 346(9-10):589–592, May 2008. 1

5. Albert Cohen, Wolfgang Dahmen, and Ronald Devore. Compressed sensing and best k-term approximation. *Journal of the American Mathematical Society*, pages 211–231, 2009. 1

6. Y. Zhang. Theory of compressive sensing via ℓ_1-minimization: a non-RIP analysis and extensions. *Journal of the Operations Research Society of China*, 1(1):79–105, 2013. 1

7. Markus Grasmair. Linear convergence rates for Tikhonov regularization with positively homogeneous functionals. *Inverse Problems*, 27(7):075014, 2011. 1, 4.1, 4.2, 7

8. S. Vaiter, G. Peyre, C. Dossal, and J. Fadili. Robust sparse analysis regularization. *IEEE Transactions on Information Theory*, 59(4):2001–2016, 2013. 1, 4.1, 4.1, 4.1, 4.2, 5

9. M. Haltmeier. Stable signal reconstruction via ℓ^1 -minimization in redundant, non-tight frames. *IEEE Transactions on Signal Processing*, 61(2):420–426, 2013. 1, 4.1, 4.2

10. Deanna Needell and Rachel Ward. Stable image reconstruction using total variation minimization. *SIAM Journal on Imaging Sciences*, 6(2):1035–1058, 2013. 1

11. Deanna Needell and Rachel Ward. Near-optimal compressed sensing guarantees for total variation minimization. *IEEE Transactions on Image Processing*, 22(10):3941–3949, 2013. 1

12. Felix Krahmer and Rachel Ward. Stable and robust sampling strategies for compressive imaging. *IEEE Transactions on Image Processing*, 23(2):612–622, 2014. 1

13. Martin Burger and Stanley Osher. Convergence rates of convex variational regularization. *Inverse Problems*, 20(5):1411, 2004. 1, 3

14. Markus Grasmair, Markus Haltmeier, and Otmar Scherzer. Necessary and sufficient conditions for linear convergence of ℓ_1-regularization. *Communications on Pure and Applied Mathematics*, 64(2):161–182, 2011. 1, 4, 6.1, 3

15. Hui Zhang, Wotao Yin, and Lizhi Cheng. Necessary and sufficient conditions of solution uniqueness in ℓ_1-norm minimization. *Journal of Optimization Theory and Applications*, in press, 2014. 1, 4, 4.1, 6.1

16. M. Hügel, H Rauhut, and T. Strohme. Remote sensing via ℓ_1-minimization. *Foundations of Computational Mathematics*, 14:115–150, 2014. 1, 5

17. S. Foucart and H. Rauhut. *A mathematical introduction to compressive sensing*. Applied and Numerical Harmonic Analysis. Birkhäuser, 2013. 1, 5

18. S. Nam, M. E. Davies, M. Elad, and R. Gribonval. The cosparse analysis model and algorithms. *Applied and Computational Harmonic Analysis*, 34(1):30–56, January 2013. 2.2, 4.1

19. L. Rudin, S. Osher, and E. Fatemi. Nonlinear total variation based noise removal algorithms. *Physica D*, 60:259–268, 1992. 2.2

20. Emmanuel J. Candès, Yonina C. Eldar, Deanna Needell, and Paige Randall. Compressed sensing with coherent and redundant dictionaries. *Applied and Computational Harmonic Analysis*, 31(1):59–73, October 2011. 2.2

21. Yulong Liu, Tiebin Mi, and S. Li. Compressed sensing with general frames via optimal-dual-based ℓ_1-analysis. *IEEE Transactions on Information Theory*, 58(7):4201–4214, 2012. 2.2

22. David L Donoho and Jared Tanner. Sparse nonnegative solution of underdetermined linear equations by linear programming. *Proceedings of the National Academy of Sciences of the United States of America*, 102(27):9446–9451, 2005. 2.3

23. Venkat Chandrasekaran, Benjamin Recht, Pablo A Parrilo, and Alan S Willsky. The convex geometry of linear inverse problems. *Foundations of Computational Mathematics*, 12(6):805–849, 2012. 2.3

24. M. Kabanava and H. Rauhut. Analysis ℓ_1-recovery with frames and gaussian measurements. *arxiv:1306.1356*, 2013. 2.3

25. S Foucart. Stability and robustness of ℓ_1-minimizations with weibull matrices and redundant dictionaries. *Linear Algebra and its Applications*, 441:4–21, 2014. 2.3

26. J. J. Fuchs. Recovery of exact sparse representations in the presence of bounded noise. *IEEE Transactions on Information Theory*, 51(10):3601–3608, October 2005. 4

27. Ryan J. Tibshirani. The lasso problem and uniqueness. *Electronic Journal of Statistics*, 7:1456–1490, 2013. 4

28. J. Fadili, G. Peyré, S. Vaiter, C-A. Deledalle, and J. Salmon. Stable recovery with analysis decomposable priors. In *Proceedings of Signal Processing with Adaptive Sparse Structured Representations (SPARS)*, 2013. 4.1, 4.1, 4.2