# One condition for all: solution uniqueness and robustness of $\ell_1$-synthesis and $\ell_1$-analysis minimizations

Hui Zhang* Ming Yan† Wotao Yin†

June 9, 2013

**Abstract**

The $\ell_1$-synthesis and $\ell_1$-analysis models recover structured signals from their undersampled measurements. The solution of the former model is often a sparse sum of dictionary atoms, and that of the latter model often 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 necessary and sufficient condition that guarantees the recovery to be unique and exact and that also guarantees the recovery is robust in presence of measurement noise. The condition is one-for-all in the sense that it applies to both of the $\ell_1$-synthesis and $\ell_1$-analysis models, and to both of their constrained and unconstrained formulations. Furthermore, a convex infinity-norm program is introduced for numerically verifying the condition. The comparison with related existing conditions are included.

**Keywords:** exact recovery, robust recovery, $\ell_1$-analysis, $\ell_1$-synthesis, sparse optimization, compressive sensing

## 1 Introduction

Let $x^* \in \mathbb{R}^n$ be a signal of interest. This paper studies when $\ell_1$ minimization can uniquely and robustly recover $x^*$ from its linear measurements

$$b = \Phi x^* + w, \tag{1}$$

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

The results of this paper cover the following $\ell_1$ minimization formulations:

$$\underset{x}{\text{minimize}}\ \|\Psi^T x\|_1, \quad \text{subject to } \Phi x = b, \tag{2a}$$

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

$$\underset{x}{\text{minimize}}\ \|\Psi^T x\|_1, \quad \text{subject to } \|\Phi x - b\|_2 \le \delta, \tag{2c}$$

where $\delta, \lambda$ are positive parameters. When $\Psi = Id$, the identify matrix, models in (2) are referred to as the $\ell_1$ (or more generally, $\ell_1$-synthesis) models. When $\Psi \neq Id$, they are referred to as the $\ell_1$-analysis models (see [6] for a recent overview).

*Department of Mathematics and Systems Science, College of Science, National University of Defense Technology, Changsha, Hunan, China, 410072. Email: hhuuii.zhang@gmail.com

†Department of Computational and Applied Mathematics, Rice University, Houston, Texas 77005, USA. Emails: yanm@ucla.edu and wotao.yin@rice.edu

In synthesis models, the signal of interest is synthesized as $x^* = Dc$, where $D$ is a certain dictionary and $c$ has the sparse coefficients. The analysis model recently attracts a lot of attention. It is assumed in [3, 14, 13] that the underlying signal makes sparse correlations with the columns (atoms) in a dictionary $\bar{D}$, i.e., $\bar{D}^T x^*$ is sparse. The signal recover model is (2b) with $\Psi$ set to $\bar{D}$.

For both synthesis and analysis models, one is interested in when the recovery is successful, that is, the solution is unique and the solution error is proportional to the amount of noise. There are several non-universal (applied to specific sparse signals) and universal (applied to all sparse signals) conditions addressing questions in various forms for $\ell_1$-synthesis minimization; examples include the non-universal dual certificate condition [7] and the "RIPless" property [4], and universal conditions such as the restricted isometry principle [2], the null space condition [5], the spherical section property [20], and others. Since $\ell_1$-analysis minimization takes a more general form than $\ell_1$-synthesis minimization, some of the above conditions have been extended to the analysis case; recent works [9, 18, 11, 16, 15, 12] have made significant contributions.

Regarding a specific signal $x^*$, this paper establishes a necessary and sufficient condition that guarantees the unique solution of any model in (2) and also that the solutions of models (2b) and (2c) are robust to any noise added to $b$. A method based on $\ell_\infty$ minimization that verifies the condition is presented. In addition, the proposed condition is compared to other conditions in the literature, most of which are stronger than ours and are thus sufficient but not necessary. Certain parts of our proofs are inspired by [1, 9, 10, 19].

**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 $x \in \mathbb{R}^n$, $\mathrm{supp}(x) = \{k : 1 \le k \le n, x_k \neq 0\}$ is the index set of the non-zero entries of $x$. $\mathrm{sign}(x)$ is the vector whose $i$th entry is the sign of $x_i$, taking a value among $+1$, $-1$, and $0$. For any $p \ge 1$, the $\ell_p$-norm of $x \in \mathbb{R}^n$ is

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

its $\ell_0$-"norm" is $\|x\|_0 = |\mathrm{supp}(x)|$, and its $\ell_\infty$-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.

**Outline.** The rest of the paper is organized as follows. Section 2 states the main results of this paper. Section 3 reviews several related results. Section 4 discusses condition verification. Proofs for the main results are given in sections 5 and 6.

# 2 Main condition and results

## 2.1 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 $\|y_J\|_\infty < 1$.*

Part (1) of the condition says that there does *not* exist any nonzero $\Delta x$ 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 linear for $x \in \mathcal{I}$; hence $\bar{x}$ cannot be the unique minimizer. 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 (2a): $0 \in \Psi\partial\|\cdot\|_1(\Psi^T x) - \Phi^T\beta$, where vector $\beta$ is the Lagrangian multipliers; we can rewrite the condition as $0 = \Psi y - \Phi^T\beta$, with $y \in \partial\|\cdot\|_1(\Psi^T x)$, which translates to $y_I = \mathrm{sign}(\Psi_I^T x)$ and $\|y_J\|_\infty \le 1$. This $y$ certifies the optimality of $\bar{x}$. For solution uniqueness and/or robustness, we shall later show that $\|y_J\|_\infty < 1$ strictly is needed.

A variant of Condition 1 is given as follows, which is equivalent to Condition 5 below from [9].

**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 $\|y_K\|_\infty \le 1$.*

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.

## 2.2 Main results

Depending on the specific models in (2), we need the following assumptions:

**Assumption 1.** *Matrix $\Phi$ has full row-rank.*

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

**Assumption 3.** *Matrix $\Psi$ has full row-rank.*

Assumptions 1 and 3 are standard. Assumption 2 is non-essential. We can scale a general $\Psi$ by multiplying it with $\frac{1}{\sqrt{\lambda_{\max}(\Psi\Psi^T)}}$. Below we state our main results, whose proofs are given in sections 5 and 6.

**Theorem 1** (Uniqueness)**.** *Under Assumption 1, let $\hat{x}$ be a solution to problem (2a) or (2b), or under Assumptions 1 and 3, let $\hat{x}$ be a solution to problem (2c). The followings 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 (2a), or to problem (2b), or to problem (2c) to be unique. To state our next result on robustness, we let

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

Part (1) of Condition 1 ensures that $0 < r(J) < +\infty$. If $\Psi = I$, then $u \in \mathrm{Ker}(\Psi_J^T)\setminus\{0\}$ is a sparse nonzero vector with maximal support $J^c$, so $r(J)$ is the inverse of the minimal singular value of the submatrix $\Phi_{J^c}$. Below we claim Condition 1 ensures the robustness of problems (2b) and (2c) to arbitrary noise in $b$.

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

*1) For any $C_0 > 0$, there exists constant $C_1 > 0$ such that every minimizer $x_{\delta,\lambda}$ of problem (2b) using parameter $\lambda = C_0\delta$ satisfies*

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

*2) Every minimizer $x_\delta$ of problem (2c) satisfies*

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

*Moreover, defining*

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

*we can let*

$$C_1 = 2C_3 + C_0\|\beta\|_2 + \frac{(1 + C_0\|\beta\|_2/2)^2 C_4}{C_0},$$
$$C_2 = 2C_3 + 2C_4\|\beta\|_2.$$

**Remark 1.** *From the results of Theorem 2, it is straightforward to derive $\ell_1$ or $\ell_2$ 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*

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

*which becomes very similar to $C_2$. This reflects the equivalence between problems* (2b) *and* (2c) *in the sense that given $\lambda$, one can find $\delta$ 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 $\Psi$ and $\Phi$. In contrast, $C_3$ only depends on the restricted property of $\Phi$.*

# 3 Related works

In the case of $\Psi = Id$, Condition 1 is well known in the literature for $\ell_1$ (or $\ell_1$-synthesis) minimization. It is initially proposed in [7] as a sufficient condition for the $\ell_1$ solution uniqueness. For problems (2b) and (2c), [8, 17] present sufficient but non-necessary conditions for solutions uniqueness. Later, its necessity is established in [10] for model (2b) and then in [19] for all models in (2), assuming $\Psi = Id$ or equal to an orthogonal basis. The solution robustness of model (2b) is given under the same condition in [10]. Below we restrict our literature review to results for the $\ell_1$-analysis model.

## 3.1 Previous uniqueness conditions

Papers [9, 14, 18, 11] cover the uniqueness of the $\ell_1$-analysis model and use stronger conditions than ours. The following condition in [14] guarantees the solution uniqueness for problem (2a):

**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)$\|(Q^T\Psi_{I^c})^+ Q^T \Psi_I \mathrm{sign}(\Psi_I^T \bar{x})\|_\infty < 1$.*

Paper [18] proposes the following condition for the solution uniqueness and robustness for problems (2a) and (2b) (the robustness requires the non-zero entries of $\Psi_I^T x$ to be sufficient large compared to noise).

**Condition 4.** *For a given $\bar{x}$, index sets $I = \operatorname{supp}(\Psi^T \bar{x})$ and $J = I^c$ satisfy:*

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

*(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* $\operatorname{Ker}(\Psi_J^T)$. *Then*

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

According to [18], 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 $\operatorname{Ker}(\Phi)$. Thus

$$\|(Q^T \Psi_{I^c})^+ Q^T \Psi_I \operatorname{sign}(\Psi_I^T \tilde{x})\|_\infty = \left\| \left( \frac{10.5}{101}, \frac{105}{101} \right)^T \right\|_\infty = \frac{105}{101}. \tag{3}$$

Hence, Condition 3 does not hold. Furthermore, $U = (1, -1, -10)^T$ is a basis matrix of $\operatorname{Ker}(\Psi_J^T)$, and the definition of $\Omega^{[J]}$ gives us $\Omega^{[J]} = (\frac{10.5}{101}, \frac{105}{101})^T$. Therefore, $IC(\operatorname{sign}(\Psi_I^T \tilde{x})) = \frac{105}{101} > 1$, so Condition 4 does not hold either. Paper [18] also presents sufficient conditions for solution uniqueness, which are reviewed in [19] and shown to be not necessary.

## 3.2 Previous robustness conditions

Turning to solution robustness, [9, 11] have studied the robustness of problems (2b) and (2c) in the Hilbert-space setting. Translating to the finite dimension, the conditions in [9] read:

**Condition 5.** *Given $\bar{x}$, the following two statements hold:*

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

*(2) There exists $y \in \partial \| \cdot \|_1 (\Psi^T \bar{x})$ such that $\Psi y \in \operatorname{Im}(\Phi^T)$ and $J = \{i : |y_i| < 1\} \neq \emptyset$.*

This condition is equivalent to Condition 2. Under Condition 5, [9] shows the existence of constant $C$ (not explicitly given) such that the solution $x_{\delta,\lambda}$ to (2b) obeys $\|\Psi^T (x_{\delta,\lambda} - x^*)\|_2 \le C\delta$ when $\lambda$ is set proportional to the noise level $\delta$. In order to obtain an explicit formula for $C$, [11] introduces the following:

**Condition 6.** *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 $\Psi y \in \operatorname{Im}(\Phi^T)$;*

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

Under this condition, the solutions to (2b) and (2c) are subject to error bounds whose constants depend on $t$, $\hat{\Phi}$, the lower frame bound of $\Psi$, and other quantities.

**Proposition 1.** *Condition 6 is stronger than Condition 2.*

*Proof.* Let $J = I(t)^c$; then we have $\|y_J\|_\infty \le t < 1$ from the definition of $I(t)$. It remains to show that $\mathrm{Ker}(\Psi_J^T) \bigcap \mathrm{Ker}(\Phi) = \{0\}$. For any $x \in \mathrm{Ker}(\Psi_J^T)$, 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 $\Phi$ restricted to $\mathrm{Span}\{\hat{\Psi}_i : i \in I(t)\} = \mathrm{Im}((\Psi\Psi^T)^{-1}\Psi_{I(t)})$ is injective, we have what we need. □

Paper [18] provides a much stronger condition below that strengthens Condition 4 by dropping the dependence on the $\Psi$-support (see the definition of $RC(I)$ below).

**Condition 7.** *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) Letting $\Omega^{[J]}$ be given as in Condition 4,*

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

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 (2b) 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 $\lambda$ can become too large than it should be.

# 4 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 $\Phi$. Assume $V = [v_1, \cdots, v_n]$. Since $\Phi$ has full row-rank, we have $\mathrm{Ker}(\Phi) = \mathrm{Span}\{v_{m+1}, \cdots, v_n\}$ and $Q = [v_{m+1}, \cdots, v_n]$ as a basis of $\mathrm{Ker}(\Phi)$. We verify that $\Psi_J^T Q$ has full row-rank, ensuring part (1) of Condition 1.

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

$$\underset{u \in \mathbb{R}^{|J|}}{\mathrm{minimize}} \|u\|_\infty, \quad \text{subject to } Au = u_1. \tag{4}$$

If the optimal objective of (4) 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 (4) has an optimal objective* $< 1$.

*Proof.* Let $\hat{u}$ be a minimizer of (4). Assume $\|\hat{u}\|_\infty < 1$. We consider the vector $y$ composed by $y_I = \mathrm{sign}(\Psi_I^T\bar{x})$ and $y_J = \hat{u}$. To show part (2) of Condition 1, it suffices to prove $\Psi y \in \mathrm{Im}(\Phi^T)$, 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. □

Convex program (4) is similar in form to one in [18] though they are used to verify different conditions.

# 5 Proof of Theorem 1

We establish Theorem 1 in two steps. Our first step proves the theorem for problem (2a) only. The second step proves Theorem 1 for problems (2b) and (2c).

## 5.1 Proof of Theorem 1 for problem (2a)

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 $\|g_J\| \le 1$ such that $\langle g_J, \Psi_J^T h\rangle = \|\Psi_J^T h\|_1$. Then,

$$
\begin{aligned}
\|\Psi^T(\hat{x}+h)\|_1 &\ge \|\Psi^T\hat{x}\|_1 + \langle \Psi g, h\rangle && \text{(5a)}\\
&= \|\Psi^T\hat{x}\|_1 + \langle \Psi g - \Psi y, h\rangle && \text{(5b)}\\
&= \|\Psi^T\hat{x}\|_1 + \langle g - y, \Psi^T h\rangle && \text{(5c)}\\
&= \|\Psi^T\hat{x}\|_1 + \langle g_J - y_J, \Psi_J^T h\rangle && \text{(5d)}\\
&\ge \|\Psi^T\hat{x}\|_1 + \|\Psi_J^T h\|_1(1-\|y_J\|_\infty), && \text{(5e)}
\end{aligned}
$$

where (5b) follows from $\Psi y \in \mathrm{Im}(\Phi^T) = \mathrm{Ker}(\Phi)^\perp$ and $h \in \mathrm{Ker}(\Phi)$, (5d) follows from the setting of $g$, and (5e) is an application of the inequality $\langle x, y\rangle \le \|x\|_1\|y\|_\infty$ 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 (2a).

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

$$
\begin{aligned}
\|\Psi^T(\hat{x})\|_1 < \|\Psi^T(\hat{x}+th)\|_1 &= \|\Psi_I^T(\hat{x}+th)\|_1 + \|\Psi_{I^c}^T(\hat{x}+th)\|_1 && \text{(6a)}\\
&= \langle \Psi_I^T(\hat{x}+th), \mathrm{sign}(\Psi_I^T(\hat{x}+th))\rangle + \|t\Psi_{I^c}^T h\|_1 && \text{(6b)}\\
&= \langle \Psi_I^T\hat{x} + t\Psi_I^T h, \mathrm{sign}(\Psi_I^T\hat{x})\rangle + \|t\Psi_{I^c}^T h\|_1 && \text{(6c)}\\
&= \langle \Psi_I^T\hat{x}, \mathrm{sign}(\Psi_I^T\hat{x})\rangle + t\langle \Psi_I^T h, \mathrm{sign}(\Psi_I^T\hat{x})\rangle + \|t\Psi_{I^c}^T h\|_1 && \text{(6d)}\\
&= \|\Psi^T(\hat{x})\|_1 + t\langle \Psi_I^T h, \mathrm{sign}(\Psi_I^T\hat{x})\rangle + \|t\Psi_{I^c}^T h\|_1. && \text{(6e)}
\end{aligned}
$$

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

$$\langle \Psi_I^T h, \mathrm{sign}(\Psi_I^T\hat{x})\rangle < \|\Psi_{I^c}^T h\|_1. \tag{7}$$

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 [10] and [19], 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

$$\underset{z\in\mathbb{R}^l}{\mathrm{minimize}}\, \|z\|_\infty \quad \text{subject to } Q^T\Psi z = -a \text{ and } z_I = 0. \tag{8}$$

For any minimizer $\hat{z}$ of problem (8), we have $\Psi(\hat{y}+\hat{z}) \in \mathrm{Ker}(\Phi)^\perp = \mathrm{Im}(\Phi^T)$ and $(\hat{y}+\hat{z})_I = \hat{y}_I = \mathrm{sign}(\Psi_I^T\hat{x})$. Thus, we shall show that the objective of problem (8) is strictly less than 1. To this end, we rewrite problem (8) in an equivalent form as:

$$\underset{z}{\mathrm{minimize}}\, \|z_{I^c}\|_\infty \quad \text{subject to } Q^T\Psi_{I^c} z_{I^c} = -a, \tag{9}$$

whose Lagrange dual problem is

$$\underset{p}{\mathrm{maximize}} \langle p, a\rangle \quad \text{subject to } \|\Psi_{I^c}^T Qp\|_1 \le 1. \tag{10}$$

Note that $Qp \in \mathrm{Ker}(\Phi)$ and $|\langle p, a\rangle| = |\langle p, Q^T\Psi\hat{y}\rangle| = |\langle p, Q^T\Psi_I \mathrm{sign}(\Psi_I^T\hat{x})\rangle| = |\langle \Psi_I^T Qp, \mathrm{sign}(\Psi_I^T\hat{x})\rangle|$. By using (7), for any $p$ we have

$$|\langle p, a\rangle| = \begin{cases} |\langle \Psi_I^T Qp, \mathrm{sign}(\Psi_I^T\hat{x})\rangle| = 0, & \text{if } Qp = 0; \\ |\langle \Psi_I^T Qp, \mathrm{sign}(\Psi_I^T\hat{x})\rangle| < \|\Psi_{I^c}^T Qp\|_1 \le 1, & \text{otherwise.} \end{cases} \tag{11}$$

Hence, problem (10) is feasible, and its objective value is strictly less than 1. By the linear programming strong duality property, problems (8) and (9) also have solutions, and their the objective value is strictly less than 1, too. This completes the proof.

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

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

## 5.2 Proof of Theorem 1 for problems (2b) and (2c)

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

*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 $\Omega$. Consider the line segment $L$ connecting $x_1$ and $x_2$. By the convexity of set $\Omega$, 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

$$\gamma\|\Phi(\alpha x_1 + (1-\alpha)x_2) - b\|_2^2 + \|\Psi^T(\alpha x_1 + (1-\alpha)x_2)\|_1 \tag{12a}$$

$$= \gamma\|\alpha(\Phi x_1 - b) + (1-\alpha)(\Phi x_2 - b)\|_2^2 + \|\alpha(\Psi^T x_1) + (1-\alpha)(\Psi^T x_2)\|_1 \tag{12b}$$

$$< \alpha(\gamma\|\Phi x_1 - b\|_2^2 + \|\Psi^T x_1\|_1) + (1-\alpha)(\gamma\|\Phi x_2 - b\|_2^2 + \|\Psi^T x_2\|_1) \tag{12c}$$

$$= \alpha\hat{c} + (1-\alpha)\hat{c} = \hat{c}, \tag{12d}$$

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 $\|\Psi^T x_1\|_1 = \|\Psi^T x_2\|_1$. □

We let $X_\lambda$ and $Y_\delta$ denote the sets of solutions to problems (2b) and (2c), respectively; moreover, we assume that these two sets are nonempty. Then, from Lemma 1, we have the following result.

**Corollary 1.** *In problem (2b), $\Phi x - b$ and $\|\Psi^T x\|_1$ are constant on $X_\lambda$; in problem (2c), $\Phi x - b$ and $\|\Psi^T x\|_1$ are constant on $Y_\delta$.*

*Proof.* Since $\|\Phi x - b\|_2^2 + \lambda\|\Psi^T x\|_1$ is constant over $X_\lambda$, the result follows directly from Lemma 1 for problem (2b). For problem (2c), if $0 \in Y_\delta$, then we have $Y_\delta = \{0\}$ because of the full row-rankness of $\Psi$. 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. □

*Proof of Theorem 1 for problems (2b) and (2c).* This proof exploits Corollary 1. Since the results of Corollary 1 are identical for problems (2b) and (2c), we present the proof for problem (2b) only.

By assumption, $X_\lambda$ 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

$$\underset{x}{\text{minimize}}\ \|\Psi^T x\|_1, \quad \text{subject to } \Phi x = b^*, \tag{13}$$

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 $X_\lambda$, 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 (13) 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 (2b) than $\hat{x}$, contradicting the assumption $\hat{x} \in X_\lambda$.

Since $X_\lambda = X^*$, $\hat{x}$ is the unique solution to problem (2b) if and only if it is the unique solution to problem (13). Since problem (13) is in the same form of problem (2a), applying the part of Theorem 1 for problem (2a), which is already proved, we conclude that $\hat{x}$ is the unique solution to problem (2b) if and only if Condition 1 or 2 holds. □

# 6 Proof of Theorem 2

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

$$\|\Psi^T x - \Psi^T \bar{x}\|_1 \le C_3 \|\Phi(x - \bar{x})\|_2 + C_4 d_y(x, \bar{x}), \quad \forall x, \tag{14}$$

*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 [9].

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

$$\|\Psi^T x - \Psi^T \bar{x}\|_1 \le \left(1 + \frac{C_3\|\Phi\|}{\sqrt{\lambda_{\min}(\Psi\Psi^T)}}\right)\|\Psi^T(x-u)\|_1 + C_3\|\Phi(x-\bar{x})\|_2. \tag{15}$$

2. this part shows that

$$f(x) := \min\left\{\|\Psi^T(x-u)\|_1 :\ u \in \operatorname{Ker}(\Psi_J^T)\right\} \le (1 - \|y_J\|_\infty)^{-1} d_y(x,\bar{x}). \tag{16}$$

Using the definition of $C_4$, combining (15) and (16) gives (14).

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

$$\|\Psi^T x - \Psi^T \bar{x}\|_1 \le \|\Psi^T(x-u)\|_1 + \|\Psi^T(u-\bar{x})\|_1. \tag{17}$$

Since $\bar{x} \in \operatorname{Ker}(\Psi_J^T)$, we have $u - \bar{x} \in \operatorname{Ker}(\Psi_J^T)$ 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 $\operatorname{supp}(\Psi^T(u-\bar{x})) = I$, we derive that

$$\begin{aligned}
\|\Psi^T(u-\bar{x})\|_1 &\le \sqrt{|I|}\|\Psi^T(u-\bar{x})\|_2 && (18a)\\
&\le \sqrt{|I|}\|u-\bar{x}\|_2 && (18b)\\
&\le \sqrt{|I|}\, r(J)\|\Phi(u-\bar{x})\|_2 && (18c)\\
&= C_3\|\Phi(u-\bar{x})\|_2, && (18d)
\end{aligned}$$

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

$$\|\Phi(u - \bar{x})\|_2 \leq \|\Phi(x - u)\|_2 + \|\Phi(x - \bar{x})\|_2 \tag{19a}$$
$$\leq \|\Phi\|\|x - u\|_2 + \|\Phi(x - \bar{x})\|_2 \tag{19b}$$
$$\leq \frac{\|\Phi\|\|\Psi^T(x - u)\|_2}{\sqrt{\lambda_{\min}(\Psi\Psi^T)}} + \|\Phi(x - \bar{x})\|_2 \tag{19c}$$
$$\leq \frac{\|\Phi\|\|\Psi^T(x - u)\|_1}{\sqrt{\lambda_{\min}(\Psi\Psi^T)}} + \|\Phi(x - \bar{x})\|_2. \tag{19d}$$

Therefore, we get (15) after combining (17), (18), and (19).

*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

$$f(x) \leq (1 - \|y_J\|_\infty)^{-1}(\|\Psi^T x\|_1 - \langle\Psi y, x\rangle). \tag{20}$$

Since $u \in \operatorname{Ker}(\Psi_J^T)$ is equivalent to $\Psi_J^T u = 0$, the Lagrangian of the minimization problem in (16) is

$$L(u, v) = \|\Psi^T(x - u)\|_1 + \langle v, \Psi_J^T u\rangle = \|\Psi^T(x - u)\|_1 + \langle\Psi_J v, u - x\rangle + \langle\Psi_J v, x\rangle. \tag{21}$$

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

$$f(x) = \max_v \min_u L(u, v) = \max_v \min_u\{\|\Psi^T(x - u)\|_1 + \langle\Psi_J v, u - x\rangle + \langle\Psi_J v, x\rangle\} \tag{22a}$$
$$= \max_w \min_u\{\|\Psi^T(x - u)\|_1 + \langle w, u - x\rangle + \langle w, x\rangle : w \in \operatorname{Im}(\Psi_J)\} \tag{22b}$$
$$= \max_w\{\langle w, x\rangle : w \in \partial\|\Psi^T \cdot \|_1(0) \cap \operatorname{Im}(\Psi_J)\} \tag{22c}$$
$$= \max_w\{\langle c\Psi y + w, x\rangle : w \in \partial\|\Psi^T \cdot \|_1(0) \cap \operatorname{Im}(\Psi_J)\} - \langle c\Psi y, x\rangle,\ \forall c > 0 \tag{22d}$$
$$= c\max_w\{\langle\Psi y + w, x\rangle : w \in c^{-1}\partial\|\Psi^T \cdot \|_1(0) \cap \operatorname{Im}(\Psi_J)\} - c\langle\Psi y, x\rangle,\ \forall c > 0. \tag{22e}$$

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

$$\big(y + c^{-1}\partial\| \cdot \|_1(0) \cap Z_J\big) \subset \partial\| \cdot \|_1(0), \tag{23}$$

from which we conclude

$$\big(\Psi y + c^{-1}\partial\|\Psi^T \cdot \|_1(0) \cap \operatorname{Im}(\Psi_J)\big) \subset \partial\|\Psi^T \cdot \|_1(0). \tag{24}$$

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

**Lemma 3.** *([1], Theorem 3; [10], Lemma 3.5) Suppose that $x^* \in \mathbb{R}^n$ is a fixed vector obeying* $\operatorname{supp}(\Psi^T x^*) = I$ *and that there are vectors satisfying $y \in \partial\| \cdot \|(\Psi^T x^*)$ and $\Psi y = \Phi^T\beta$. Then for every $\delta > 0$ and every data vector $b$ satisfying $\|\Phi x^* - b\|_2 \leq \delta$, the following two statements hold:*

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

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

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

*Proof of Theorem 2.* Firstly, we derive that

$$\|\Psi(x_{\delta,\lambda} - x^*)\|_1 \le C_3\|\Phi(x_{\delta,\lambda} - x^*)\|_2 + C_4 d_y(x_{\delta,\lambda}, x^*) \tag{25a}$$

$$\le C_3\|\Phi x_{\delta,\lambda} - b\|_2 + C_3\|\Phi x^* - b\|_2 + C_4 d_y(x_{\delta,\lambda}, x^*) \tag{25b}$$

$$\le C_3(\delta + \lambda\|\beta\|_2) + C_3\delta + C_4\frac{(\delta + \lambda\|\beta\|_2/2)^2}{\lambda}, \tag{25c}$$

where the first and the third inequalities follow from Lemmas 2 and 3, respectively. Substituting $\lambda = C_0\delta$ and collecting like terms in (25c), we obtain the first part of Theorem 2. The second part can be proved in the same way. □


## Acknowledgements

The work of H. Zhang is supported by China Scholarship Council during his visit to Rice University, and in part by Graduate School of NUDT under Funding of Innovation B110202, Hunan Provincial Innovation Foundation for Postgraduate CX2011B008, and NSFC Grants No.61271014 and No.61072118. The work of M. Yan is supported in part by the Center for Domain-Specific Computing (CDSC) funded by NSF grants CCF-0926127 and ARO/ARL MURI grant FA9550-10-1-0567. The work of W. Yin is supported in part by NSF grants DMS-0748839 and ECCS-1028790 They thank Rachel Ward and Xiaoya Zhang for their discussions.


## References


[1] M. Burger and S. Osher, *Convergence rates of convex variational regularization*, Inverse Problems, 20 (2004), p. 1411.

[2] E. Candes, *The restricted isometry property and its implications for compressed sensing*, Comptes Rendus Mathematique, 346 (2008), pp. 589–592.

[3] E. J. Candès, Y. C. Eldar, D. Needell, and P. Randall, *Compressed sensing with coherent and redundant dictionaries*, Applied and Computational Harmonic Analysis, 31 (2011), pp. 59–73.

[4] E. J. Candes and Y. Plan, *A probabilistic and RIPless theory of compressed sensing*, IEEE Trans. Inf. Theor., 57 (2011), pp. 7235–7254.

[5] A. Cohen, W. Dahmen, and R. Devore, *Compressed sensing and best k-term approximation*, J. Amer. Math. Soc, (2009), pp. 211–231.

[6] M. Elad, P. Milanfar, and R. Rubinstein, *Analysis versus synthesis in signal priors*, Inverse Problems, 23 (2007), p. 947.

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

[8] J. J. Fuchs, *Recovery of exact sparse representations in the presence of bounded noise*, IEEE Trans. Inf. Theor., 51 (2005), pp. 3601–3608.

[9] M. Grasmair, *Linear convergence rates for Tikhonov regularization with positively homogeneous functionals*, Inverse Problems, 27 (2011), p. 075014.

[10] M. Grasmair, M. Haltmeier, and O. Scherzer, *Necessary and sufficient conditions for linear convergence of $\ell_1$-regularization*, Comm. Pure Appl. Math., 64 (2011), pp. 161–182.

[11] M. Haltmeier, *Stable signal reconstruction via $\ell^1$ -minimization in redundant, non-tight frames*, Signal Processing, IEEE Transactions on, 61 (2013), pp. 420–426.

[12] F. Krahmer and R. Ward, *Compressive imaging: stable and robust recovery from variable density frequency samples*, arXiv:1210.2380, (2012).

[13] Y. Liu, T. Mi, and S. Li, *Compressed sensing with general frames via optimal-dual-based $\ell_1$ -analysis*, Information Theory, IEEE Transactions on, 58 (2012), pp. 4201–4214.

[14] S. Nam, M. E. Davies, M. Elad, and R. Gribonval, *The cosparse analysis model and algorithms*, Applied and Computational Harmonic Analysis, 34 (2013), pp. 30–56.

[15] D. Needell and R. Ward, *Near-optimal compressed sensing guarantees for total variation minimization*, arXiv:1210.3098, (2012).

[16] D. Needell and R. Ward, *Stable image reconstruction using total variation minimization*, arXiv:1202.6429, (2012).

[17] R. J. Tibshirani, *The lasso problem and uniqueness*, arXiv:1206.0313, (2012).

[18] S. Vaiter, G. Peyre, C. Dossal, and J. Fadili, *Robust sparse analysis regularization*, Information Theory, IEEE Transactions on, 59 (2013), pp. 2001–2016.

[19] H. Zhang, W. Yin, and L. Cheng, *Necessary and sufficient conditions of solution uniqueness in $\ell_1$ minimization*, arXiv:1209.0652, (2012).

[20] Y. Zhang, *Theory of compressive sensing via $\ell_1$-minimization: a non-RIP analysis and extensions*, Journal of the Operations Research Society of China, 1 (2013), pp. 79–105.