Stable Signal Recovery from Phaseless Measurements

Bing Gao · Yang Wang · Zhiqiang Xu

Abstract The aim of this paper is to study the stability of the $\ell_1$ minimization for the compressive phase retrieval and to extend the instance-optimality in compressed sensing to the real phase retrieval setting. We first show that $m = O(k \log(N/k))$ measurements are enough to guarantee the $\ell_1$ minimization to recover $k$-sparse signals stably provided the measurement matrix $A$ satisfies the strong RIP property. We second investigate the phaseless instance-optimality presenting a null space property of the measurement matrix $A$ under which there exists a decoder $\Delta$ so that the phaseless instance-optimality holds. We use the result to study the phaseless instance-optimality for the $\ell_1$ norm. This builds a parallel for compressive phase retrieval with the classical compressive sensing.

Keywords Phase retrieval · Sparse signals · Compressed sensing

Mathematics Subject Classification 94A12
1 Introduction

In this paper we consider the phase retrieval for sparse signals with noisy measurements, which arises in many different applications. Assume that

\[ b_j := |\langle a_j, x_0 \rangle| + e_j, \quad j = 1, \ldots, m \]

where \( x_0 \in \mathbb{R}^N, a_j \in \mathbb{R}^N \) and \( e_j \in \mathbb{R} \) is the noise. Our goal is to recover \( x_0 \) up to a unimodular scaling constant from \( b := (b_1, \ldots, b_m)^T \) with the assumption of \( x_0 \) being approximately \( k \)-sparse. This problem is referred to as the compressive phase retrieval problem [9].

The paper attempts to address two problems. Firstly we consider the stability of \( \ell_1 \) minimization for the compressive phase retrieval problem where the signal \( x_0 \) is approximately \( k \)-sparse, which is the \( \ell_1 \) minimization problem defined as follows:

\[
\min ||x||_1 \ \text{subject to} \quad \|Ax| - |Ax_0]\|_2 \leq \epsilon, \quad (1.1)
\]

where \( A := [a_1, \ldots, a_m]^T \) and \( |Ax_0| := [|\langle a_1, x_0 \rangle|, \ldots, |\langle a_m, x_0 \rangle|]^T \). Secondly we investigate instance-optimality in the phase retrieval setting.

Note that in the classical compressive sensing setting the stable recovery of a \( k \)-sparse signal \( x_0 \in \mathbb{C}^N \) can be done using \( m = \mathcal{O}(k \log(N/k)) \) measurements for several classes of measurement matrices \( A \). A natural question is whether stable compressive phase retrieval can also be attained with \( m = \mathcal{O}(k \log(N/k)) \) measurements. This has indeed proved to be the case in [6] if \( x_0 \in \mathbb{R}^N \) and \( A \) is a random real Gaussian matrix. In [8] a two-stage algorithm for compressive phase retrieval is proposed, which allows for very fast recovery of a sparse signal if the matrix \( A \) can be written as a product of a random matrix and another matrix (such as a random matrix) that allows for efficient phase retrieval. The authors proved that stable compressive phase retrieval can be achieved with \( m = \mathcal{O}(k \log(N/k)) \) measurements for complex signals \( x_0 \) as well. In [10], the strong RIP (S-RIP) property is introduced and the authors show that one can use the \( \ell_1 \) minimization to recover sparse signals up to a global sign from the noiseless measurements \( |Ax_0| \) provided \( A \) satisfies S-RIP. Naturally, one is interested in the performance of \( \ell_1 \) minimization for the compressive phase retrieval with noisy measurements. In this paper, we shall show that the \( \ell_1 \) minimization scheme given in (1.1) will recover a \( k \)-sparse signal stably from \( m = \mathcal{O}(k \log(N/k)) \) measurements, provided that the measurement matrix \( A \) satisfies the strong RIP (S-RIP) property. This establishes an important parallel for compressive phase retrieval with the classical compressive sensing. Note that in [11] such a parallel in terms of the null space property was already established.

The notion of instance optimality was first introduced in [5]. We use \( ||x||_0 \) to denote the number of non-zero elements in \( x \). Given a norm \( ||\cdot||_X \) such as the \( \ell_1 \)-norm and \( x \in \mathbb{R}^N \), the best \( k \)-term approximation error is defined as

\[
\sigma_k(x)_X := \min_{z \in \Sigma_k} \|x - z\|_X,
\]
where

$$\Sigma_k := \{ x \in \mathbb{R}^N : \|x\|_0 \leq k \}.$$  

We use $\Delta : \mathbb{R}^m \mapsto \mathbb{R}^N$ to denote a decoder for reconstructing $x$. We say the pair $(A, \Delta)$ is \textit{instance optimal of order $k$ with constant $C_0$} if

$$\| x - \Delta(Ax) \|_X \leq C_0 \sigma_k(x)_X$$  \hspace{1cm} (1.2)

holds for all $x \in \mathbb{R}^N$. In extending it to phase retrieval, our decoder will have the input $b = |Ax|$. A pair $(A, \Delta)$ is said to be \textit{phaseless instance optimal of order $k$ with constant $C_0$} if

$$\min \left\{ \| x - \Delta(|Ax|) \|_X, \| x + \Delta(|Ax|) \|_X \right\} \leq C_0 \sigma_k(x)_X$$  \hspace{1cm} (1.3)

holds for all $x \in \mathbb{R}^N$. We are interested in the following problem: \textit{Given $\| \cdot \|_X$ and $k < N$, what is the minimal value of $m$ for which there exists $(A, \Delta)$ so that (1.3) holds?}

The null space $\mathcal{N}(A) := \{ x \in \mathbb{R}^N : Ax = 0 \}$ of $A$ plays an important role in the analysis of the original instance optimality (1.2) (see [5]). Here we present a null space property for $\mathcal{N}(A)$, which is necessary and sufficient, for which there exists a decoder $\Delta$ so that (1.3) holds. We apply the result to investigate the instance optimality where $X$ is the $\ell_1$ norm. Set

$$\Delta_1(|Ax|) := \operatorname{argmin}_{z \in \mathbb{R}^N} \left\{ \| z \|_1 : |Ax| = |Az| \right\}.$$  

We show that the pair $(A, \Delta_1)$ satisfies (1.3) with $X$ being the $\ell_1$-norm provided $A$ satisfies the strong RIP property (see Definition 2.1). As shown in [10], the Gaussian random matrix $A \in \mathbb{R}^{m \times N}$ satisfies the strong RIP of order $k$ for $m = \mathcal{O}(k \log(N/k))$. Hence $m = \mathcal{O}(k \log(N/k))$ measurements suffice to ensure the phaseless instance optimality (1.3) for the $\ell_1$-norm exactly as with the traditional instance optimality (1.2).

2 Auxiliary Results

In this section we provide some auxiliary results that will be used in later sections. For $x \in \mathbb{R}^N$ we use $\|x\|_p := \|x\|_{\ell_p}$ to denote the $p$-norm of $x$ for $0 < p \leq \infty$. The measurement matrix is given by $A := [a_1, \ldots, a_m]^T \in \mathbb{R}^{m \times N}$ as before. Given an index set $I \subset \{1, \ldots, m\}$ we shall use $A_I$ to denote the sub-matrix of $A$ where only rows with indices in $I$ are kept, i.e.,

$$A_I := [a_j : j \in I]^T.$$
The matrix $A$ satisfies the Restricted Isometry Property (RIP) of order $k$ if there exists a constant $\delta_k \in [0, 1)$ such that for all $k$-sparse vectors $z \in \Sigma_k$ we have

$$(1 - \delta_k)\|z\|_2^2 \leq \|Az\|_2^2 \leq (1 + \delta_k)\|z\|_2^2.$$  

It was shown in [2] that one can use $\ell_1$-minimization to recover $k$-sparse signals provided that $A$ satisfies the RIP of order $tk$ and $\delta_{tk} < \sqrt{1 - \frac{1}{t}}$ where $t > 1$.

To investigate compressive phase retrieval, a stronger notion of RIP is given in [10]:

**Definition 2.1** (S-RIP) We say the matrix $A = [a_1, \ldots, a_m]^\top \in \mathbb{R}^{m \times N}$ has the Strong Restricted Isometry Property of order $k$ with bounds $\theta_-, \theta_+ \in (0, 2)$ if

$$\theta_- \|x\|_2^2 \leq \min_{I \subseteq [m], |I| \geq m/2} \|A_I x\|_2^2 \leq \max_{I \subseteq [m], |I| \geq m/2} \|A_I x\|_2^2 \leq \theta_+ \|x\|_2^2$$

holds for all $k$-sparse signals $x \in \mathbb{R}^N$, where $[m] := \{1, \ldots, m\}$. We say $A$ has the Strong Lower Restricted Isometry Property of order $k$ with bound $\theta_-$ if the lower bound in (2.1) holds. Similarly we say $A$ has the Strong Upper Restricted Isometry Property of order $k$ with bound $\theta_+$ if the upper bound in (2.1) holds.

The authors of [10] proved that Gaussian matrices with $m = \mathcal{O}(tk \log(N/k))$ satisfy S-RIP of order $tk$ with high probability.

**Theorem 2.1** ([10]) Suppose that $t > 1$ and $A = (a_{ij}) \in \mathbb{R}^{m \times N}$ is a random Gaussian matrix with $m = \mathcal{O}(tk \log(N/k))$ and $a_{ij} \sim \mathcal{N}(0, \frac{1}{\sqrt{m}})$. Then there exist $\theta_-, \theta_+ \in (0, 2)$ such that with probability $1 - \exp(-cm/2)$ the matrix $A$ satisfies the S-RIP of order $tk$ with constants $\theta_-$ and $\theta_+$, where $c > 0$ is an absolute constant and $\theta_-, \theta_+$ are independent of $t$.

The following is a very useful lemma for this study.

**Lemma 2.1** Let $x_0 \in \mathbb{R}^N$ and $\rho \geq 0$. Suppose that $A \in \mathbb{R}^{m \times N}$ is a measurement matrix satisfying the restricted isometry property with $\delta_{tk} \leq \sqrt{\frac{t-1}{t}}$ for some $t > 1$. Then for any

$$\hat{x} \in \left\{ x \in \mathbb{R}^N : \|x\|_1 \leq \|x_0\|_1 + \rho, \|Ax - Ax_0\|_2 \leq \epsilon \right\}$$

we have

$$\|\hat{x} - x_0\|_2 \leq c_1 \epsilon + c_2 \frac{2\sigma_k(x_0)_1}{\sqrt{k}} + c_2 \cdot \frac{\rho}{\sqrt{k}},$$

where $c_1 = \frac{\sqrt{2(1+\delta)}}{1-\sqrt{t/(t-1)}\delta}$, $c_2 = \frac{\sqrt{2\delta + \sqrt{(\sqrt{t(t-1)} - \delta)t\delta}}}{\sqrt{t(t-1)} - \delta t} + 1$. 

\[\text{Birkhäuser}\]
Remark 2.1 We build the proof of Lemma 2.1 following the ideas of Cai and Zhang [2]. The full proof is given in Appendix for completeness. It is well-known that an effective method to recover approximately-sparse signals $x_0$ in the traditional compressive sensing is to solve

$$x^\# := \arg\min_x \{ \|x\|_1 : \|Ax - Ax_0\|_2 \leq \epsilon \}. \quad (2.2)$$

The definition of $x^\#$ shows that

$$\|x^\#\|_1 \leq \|x_0\|_1, \quad \|Ax^\# - Ax_0\|_2 \leq \epsilon,$$

which implies that

$$\|x^\# - x_0\|_2 \leq C_1 \epsilon + C_2 \frac{\sigma_k(x_0)}{\sqrt{k}},$$

provided that $A$ satisfies the RIP condition with $\delta_{tk} \leq \sqrt{1 - 1/t}$ for $t > 1$ (see [2]). However, in practice one prefers to design fast algorithms to find an approximation solution of (2.2), say $\hat{x}$. Thus it is possible to have $\|\hat{x}\|_1 > \|x_0\|_1$. Lemma 2.1 gives an estimate of $\|\hat{x} - x_0\|_2$ for the case where $\|\hat{x}\|_1 \leq \|x_0\|_1 + \rho$.

Remark 2.2 In [7], Han and Xu extend the definition of S-RIP by replacing the $m/2$ in (2.1) by $\beta m$ where $0 < \beta < 1$. They also prove that, for any fixed $\beta \in (0, 1)$, the $m \times N$ random Gaussian matrix satisfies S-RIP of order $k$ with high probability provided $m = O(k \log(N/k))$.

3 Stable Recovery of Real Phase Retrieval Problem

3.1 Stability Results

The following lemma shows that the map $\phi_A(x) := |Ax|$ is stable on $\Sigma_k$ modulo a unimodular constant provided $A$ satisfies strong lower RIP of order $2k$. Define the equivalent relation $\sim$ on $\mathbb{R}^N$ and $\mathbb{C}^N$ by the following: for any $x, y, x \sim y$ iff $x = cy$ for some unimodular scalar $c$, where $x, y$ are in $\mathbb{R}^N$ or $\mathbb{C}^N$. For any subset $Y$ of $\mathbb{R}^N$ or $\mathbb{C}^N$ the notation $Y/\sim$ denotes the equivalent classes of elements in $Y$ under the equivalence. Note that there is a natural metric $D_\sim$ on $\mathbb{C}^N/\sim$ given by

$$D_\sim(x, y) = \min_{|c|=1} \|x - cy\|.$$

Our primary focus in this paper will be on $\mathbb{R}^N$, and in this case $D_\sim(x, y) = \min\{\|x - y\|_2, \|x + y\|_2\}$.

Lemma 3.1 Let $A \in \mathbb{R}^{m \times N}$ satisfy the strong lower RIP of order $2k$ with constant $\theta_-$. Then for any $x, y \in \Sigma_k$ we have

$$\|\|Ax\| - |Ay|\|_2^2 \geq \theta_- \min(\|x - y\|_2^2, \|x + y\|_2^2).$$
Proof For any \( x, y \in \Sigma_k \) we divide \( \{1, \ldots, m\} \) into two subsets:

\[
T = \{ j : \text{sign}(\langle a_j, x \rangle) = \text{sign}(\langle a_j, y \rangle) \}
\]

and

\[
T^c = \{ j : \text{sign}(\langle a_j, x \rangle) = -\text{sign}(\langle a_j, y \rangle) \}.
\]

Clearly one of \( T \) and \( T^c \) will have cardinality at least \( m/2 \). Without loss of generality we assume that \( T \) has cardinality no less than \( m/2 \). Then

\[
\| |Ax| - |Ay| \|_2^2 \geq \| A_T x - A_T y \|_2^2 \\
\geq \| x - y \|_2^2 \\
\geq \theta_- \min(\|x - y\|_2^2, \|x + y\|_2^2).
\]

\[\Box\]

Remark 3.1 Note that the combination of Lemma 3.1 and Theorem 2.1 shows that for an \( m \times N \) Gaussian matrix \( A \) with \( m = \text{O}(k \log(N/k)) \) one can guarantee the stability of the map \( \phi_A(x) := |Ax| \) on \( \Sigma_k/\sim \).

3.2 The Main Theorem

In this part, we will consider how many measurements are needed for the stable sparse phase retrieval by \( \ell_1 \)-minimization via solving the following model:

\[
\min \|x\|_1 \quad \text{subject to} \quad \| |Ax| - |Ax_0| \|_2^2 \leq \epsilon^2,
\]

where \( A \) is our measurement matrix and \( x_0 \in \mathbb{R}^N \) is a signal we wish to recover. The next theorem tells under what conditions the solution to (3.1) is stable.

Theorem 3.1 Assume that \( A \in \mathbb{R}^{m \times N} \) satisfies the S-RIP of order \( tk \) with bounds \( \theta_-, \theta_+ \in (0, 2) \) such that

\[
t \geq \max \left\{ \frac{1}{2\theta_- - \theta_-^2}, \frac{1}{2\theta_+ - \theta_+^2} \right\}.
\]

Then any solution \( \hat{x} \) for (3.1) satisfies

\[
\min \{ \| \hat{x} - x_0 \|_2, \| \hat{x} + x_0 \|_2 \} \leq c_1 \epsilon + c_2 \frac{2\sigma_k(x_0)}{\sqrt{k}},
\]

where \( c_1 \) and \( c_2 \) are constants defined in Lemma 2.1.
Proof Clearly any \( \hat{x} \in \mathbb{R}^N \) satisfying (3.1) must have

\[
\|\hat{x}\|_1 \leq \|x_0\|_1 \tag{3.2}
\]

and

\[
\|A\hat{x} - Ax_0\|_2^2 \leq \epsilon^2. \tag{3.3}
\]

Now the index set \( \{1, 2, \ldots, m\} \) is divisible into two subsets

\[
T = \{j : \text{sign}(\langle a_j, \hat{x} \rangle) = \text{sign}(\langle a_j, x_0 \rangle)\},
\]

\[
T^c = \{j : \text{sign}(\langle a_j, \hat{x} \rangle) = -\text{sign}(\langle a_j, x_0 \rangle)\}.
\]

Then (3.3) implies that

\[
\|A_T\hat{x} - A_Tx_0\|_2^2 + \|A_{T^c}\hat{x} + A_{T^c}x_0\|_2^2 \leq \epsilon^2. \tag{3.4}
\]

Here either \( |T| \geq m/2 \) or \( |T^c| \geq m/2 \). Without loss of generality we assume that \( |T| \geq m/2 \). We use the fact

\[
\|A_T\hat{x} - A_Tx_0\|_2^2 \leq \epsilon^2. \tag{3.5}
\]

From (3.2) and (3.5) we obtain

\[
\hat{x} \in \left\{ x \in \mathbb{R}^N : \|x\|_1 \leq \|x_0\|_1, \|A_Tx - A_Tx_0\|_2 \leq \epsilon \right\}. \tag{3.6}
\]

Recall that \( A \) satisfies S-RIP of order \( t^k \) and constants \( \theta_-, \theta_+ \). Here

\[
t \geq \max\left\{ \frac{1}{2\theta_- - \theta_-^2}, \frac{1}{2\theta_+ - \theta_+^2} \right\} > 1. \tag{3.7}
\]

The definition of S-RIP implies that \( A_T \) satisfies the RIP of order \( t^k \) in which

\[
\delta_{t^k} \leq \max\{1 - \theta_-, \theta_+ - 1\} \leq \sqrt{\frac{t - 1}{t}}. \tag{3.8}
\]

where the second inequality follows from (3.7). The combination of (3.6), (3.8) and Lemma 2.1 now implies

\[
\|\hat{x} - x_0\|_2 \leq c_1 \epsilon + c_2 \frac{2\sigma_k(x_0)}{\sqrt{k}},
\]

where \( c_1 \) and \( c_2 \) are defined in Lemma 2.1. If \( |T^c| \geq m/2 \) we get the corresponding result

\[
\|\hat{x} + x_0\|_2 \leq c_1 \epsilon + c_2 \frac{2\sigma_k(x_0)}{\sqrt{k}}.
\]

The theorem is now proved.
This theorem demonstrates that, if the measurement matrix has the S-RIP, the real compressive phase retrieval problem can be solved stably by $\ell_1$-minimization.

4 Phase Retrieval and Best k-term Approximation

4.1 Instance Optimality from the Linear Measurements

We introduce some definitions and results in [5]. Recall that for a given encoder matrix $A \in \mathbb{R}^{m \times N}$ and a decoder $\Delta : \mathbb{R}^m \mapsto \mathbb{R}^N$, the pair $(A, \Delta)$ is said to have instance optimality of order $k$ with constant $C_0$ with respect to the norm $X$ if

$$\|x - \Delta(Ax)\|_X \leq C_0 \sigma_k(x)_X$$

holds for all $x \in \mathbb{R}^N$. Set $\mathcal{N}(A) := \{ \eta \in \mathbb{R}^N : A\eta = 0 \}$ to be the null space of $A$. The following theorem gives conditions under which the (4.1) holds.

**Theorem 4.1** ([5]) Let $A \in \mathbb{R}^{m \times N}$, $1 \leq k \leq N$ and $\| \cdot \|_X$ be a norm on $\mathbb{R}^N$. Then a sufficient condition for the existence of a decoder $\Delta$ satisfying (4.1) is

$$\|\eta\|_X \leq C_0 \sigma_{2k}(\eta)_X, \forall \eta \in \mathcal{N}(A).$$

A necessary condition for the existence of a decoder $\Delta$ satisfying (4.1) is

$$\|\eta\|_X \leq C_0 \sigma_{2k}(\eta)_X, \forall \eta \in \mathcal{N}(A).$$

For the norm $X = \ell_1$ it was established in [5] that instance optimality of order $k$ can indeed be achieved, e.g. for a Gaussian matrix $A$, with $m = O(k \log(N/k))$. The authors also considered more generally taking different norms on both sides of (4.1). Following [5], we say the pair $(A, \Delta)$ has $(p, q)$-instance optimality of order $k$ with constant $C_0$ if

$$\|x - \Delta(Ax)\|_p \leq C_0 k^{\frac{1}{q} - \frac{1}{p}} \sigma_k(x)_q, \forall x \in \mathbb{R}^N,$$

with $1 \leq q \leq p \leq 2$. It was shown in [5] that the $(p, q)$-instance optimality of order $k$ can be achieved at the cost of having $m = O(k(N/k)^{2-2/q}) \log(N/k)$ measurements.

4.2 Phaseless Instance Optimality

A natural question here is whether an analogous result to Theorem 4.1 exists for phaseless instance optimality defined in (1.3). We answer the question by presenting such a result in the case of real phase retrieval.

Recall that a pair $(A, \Delta)$ is said to have the phaseless instance optimality of order $k$ with constant $C_0$ for the norm $\| \cdot \|_X$ if

$$\min \left\{ \|x - \Delta(|Ax|)\|_X, \|x + \Delta(|Ax|)\|_X \right\} \leq C_0 \sigma_k(x)_X$$

holds for all $x \in \mathbb{R}^N$.
Theorem 4.2 Let $A \in \mathbb{R}^{m \times N}$, $1 \leq k \leq N$ and $\| \cdot \|_X$ be a norm. Then a sufficient condition for the existence of a decoder $\Delta$ satisfying the phaseless instance optimality (4.5) is: For any $I \subseteq \{1, \ldots, m\}$ and $\eta_1 \in \mathcal{N}(A_I)$, $\eta_2 \in \mathcal{N}(A_{I^c})$ we have

$$\min\{\|\eta_1\|_X, \|\eta_2\|_X\} \leq \frac{C_0}{4} \sigma_k(\eta_1 - \eta_2)_X + \frac{C_0}{4} \sigma_k(\eta_1 + \eta_2)_X. \quad (4.6)$$

A necessary condition for the existence of a decoder $\Delta$ satisfying (4.5) is: For any $I \subseteq \{1, \ldots, m\}$ and $\eta_1 \in \mathcal{N}(A_I)$, $\eta_2 \in \mathcal{N}(A_{I^c})$ we have

$$\min\{\|\eta_1\|_X, \|\eta_2\|_X\} \leq \frac{C_0}{2} \sigma_k(\eta_1 - \eta_2)_X + \frac{C_0}{2} \sigma_k(\eta_1 + \eta_2)_X. \quad (4.7)$$

Proof We first assume (4.6) holds, and show that there exists a decoder $\Delta$ satisfying the phaseless instance optimality (4.5). To this end, we define a decoder $\Delta$ as follows:

$$\Delta(|Ax_0|) = \arg\min_{\|Ax\| = |Ax_0|} \sigma_k(x)_X.$$

Suppose $\hat{x} := \Delta(|Ax_0|)$. We have $|A\hat{x}| = |Ax_0|$ and $\sigma_k(\hat{x})_X \leq \sigma_k(x_0)_X$. Note that $\langle a_j, \hat{x} \rangle = \pm \langle a_j, x_0 \rangle$. Let $I \subseteq \{1, \ldots, m\}$ be defined by

$$I = \left\{ j : \langle a_j, \hat{x} \rangle = \langle a_j, x_0 \rangle \right\}.$$

Then

$$A_I(x_0 - \hat{x}) = 0, \quad A_{I^c}(x_0 + \hat{x}) = 0.$$

Set

$$\eta_1 := x_0 - \hat{x} \in \mathcal{N}(A_I), \quad \eta_2 := x_0 + \hat{x} \in \mathcal{N}(A_{I^c}).$$

A simple observation yields

$$\sigma_k(\eta_1 - \eta_2)_X = 2\sigma_k(\hat{x})_X \leq 2\sigma_k(x_0)_X, \quad \sigma_k(\eta_1 + \eta_2)_X = 2\sigma_k(x_0)_X. \quad (4.8)$$

Then (4.6) implies that

$$\min\{\|\hat{x} - x_0\|_X, \|\hat{x} + x_0\|_X\} = \min\{\|\eta_1\|_X, \|\eta_2\|_X\}$$

$$\leq \frac{C_0}{4} \sigma_k(\eta_1 - \eta_2)_X + \frac{C_0}{4} \sigma_k(\eta_1 + \eta_2)_X$$

$$\leq C_0 \sigma_k(x_0)_X.$$

Here the last equality is obtained by (4.8). This proves the sufficient condition.
We next turn to the necessary condition. Let \( \Delta \) be a decoder for which the phaseless instance optimality (4.5) holds. Let \( I \subseteq \{1, \ldots, m\} \). For any \( \eta_1 \in \mathcal{N}(A_I) \) and \( \eta_2 \in \mathcal{N}(A_{I'}) \) we have
\[
|A(\eta_1 + \eta_2)| = |A(\eta_1 - \eta_2)| = |A(\eta_2 - \eta_1)|.
\] (4.9)

The instance optimality implies
\[
\min \left\{ \| \Delta(\eta_1 + \eta_2) \|_X + \eta_1 + \eta_2, \| \Delta(\eta_1 - \eta_2) \|_X - (\eta_1 + \eta_2) \|_X \right\} \leq C_0 \sigma_k(\eta_1 + \eta_2) X.
\] (4.10)

Without loss of generality we may assume that
\[
\| \Delta(\eta_1 + \eta_2) \|_X + \eta_1 + \eta_2 \leq \| \Delta(\eta_1 + \eta_2) \|_X - (\eta_1 + \eta_2) \|_X.
\]

Then (4.10) implies that
\[
\| \Delta(\eta_1 + \eta_2) \|_X + \eta_1 + \eta_2 \leq C_0 \sigma_k(\eta_1 + \eta_2) X.
\] (4.11)

By (4.9), we have
\[
\| \Delta(\eta_1 + \eta_2) \|_X + \eta_1 + \eta_2 \leq \| \Delta(\eta_1 + \eta_2) \|_X - (\eta_2 - \eta_1) + 2\eta_2 \|_X \geq 2\| \eta_2 \|_X - \| \Delta(\eta_2 - \eta_1) \|_X - (\eta_2 - \eta_1) \|_X.
\] (4.12)

Combining (4.11) and (4.12) yields
\[
2\| \eta_2 \|_X \leq C_0 \sigma_k(\eta_1 + \eta_2) X + \| \Delta(\eta_2 - \eta_1) \|_X - (\eta_2 - \eta_1) \|_X.
\] (4.13)

At the same time, (4.9) also implies
\[
\| \Delta(\eta_1 + \eta_2) \|_X + \eta_1 + \eta_2 \leq \| \Delta(\eta_1 + \eta_2) \|_X + (\eta_2 - \eta_1) + 2\eta_1 \|_X \geq 2\| \eta_1 \|_X - \| \Delta(\eta_2 - \eta_1) \|_X + (\eta_2 - \eta_1) \|_X.
\] (4.14)

Putting (4.11) and (4.14) together, we obtain
\[
2\| \eta_1 \|_X \leq C_0 \sigma_k(\eta_1 + \eta_2) X + \| \Delta(\eta_2 - \eta_1) \|_X + (\eta_2 - \eta_1) \|_X.
\] (4.15)

It follows from (4.13) and (4.15) that
\[
\min \{ \| \eta_1 \|_X, \| \eta_2 \|_X \} \leq \frac{C_0}{2} \sigma_k(\eta_1 + \eta_2) X + \frac{1}{2} \min \{ \| \Delta(\eta_2 - \eta_1) \|_X - (\eta_2 - \eta_1) \|_X, \| \Delta(\eta_2 - \eta_1) \|_X + (\eta_2 - \eta_1) \|_X \} \leq \frac{C_0}{2} \sigma_k(\eta_1 + \eta_2) X + \frac{C_0}{2} \sigma_k(\eta_1 - \eta_2) X.
\]
Here the last inequality is obtained by the instance optimality of $(A, \Delta)$. For the case where
\[ \| \Delta(|A(\eta_1 + \eta_2)|) - (\eta_1 + \eta_2) \|_X \leq \| \Delta(|A(\eta_1 + \eta_2)|) + \eta_1 + \eta_2 \|_X, \]
we obtain
\[ \min(\| \eta_1 \|_X, \| \eta_2 \|_X) \leq \frac{C_0}{2} \sigma_k(\eta_1 + \eta_2)_X + \frac{C_0}{2} \sigma_k(\eta_1 - \eta_2)_X \]
via the same argument. The theorem is now proved. \( \Box \)

We next present a null space property for phaseless instance optimality, which allows us to establish parallel results for sparse phase retrieval.

**Definition 4.1** We say a matrix $A \in \mathbb{R}^{m \times N}$ satisfies the strong null space property (S-NSP) of order $k$ with constant $C$ if for any index set $I \subseteq \{1, \ldots, m\}$ with $|I| \geq m/2$ and $\eta \in \mathcal{N}(A_I)$ we have
\[ \| \eta \|_X \leq C \cdot \sigma_k(\eta)_X. \]

**Theorem 4.3** Assume that a matrix $A \in \mathbb{R}^{m \times N}$ has the strong null space property of order $2k$ with constant $C_0/2$. Then there must exist a decoder $\Delta$ having the phaseless instance optimality (1.3) with constant $C_0$. In particular, one such decoder is
\[ \Delta(|Ax_0|) = \arg\min_{|Ax| = |Ax_0|} \sigma_k(x)_X. \]

**Proof** Assume that $I \subseteq \{1, \ldots, m\}$. For any $\eta_1 \in \mathcal{N}(A_I)$ and $\eta_2 \in \mathcal{N}(A_{I^c})$ we must have either $\| \eta_1 \|_X \leq \frac{C_0}{2} \sigma_{2k}(\eta_1)_X$ or $\| \eta_2 \|_X \leq \frac{C_0}{2} \sigma_{2k}(\eta_2)_X$ by the strong null space property. If $\| \eta_1 \|_X \leq \frac{C_0}{2} \sigma_{2k}(\eta_1)_X$ then
\[ \| \eta_1 \|_X \leq \frac{C_0}{2} \sigma_{2k}(\eta_1)_X \leq \frac{C_0}{4} \sigma_k(\eta_1 - \eta_2)_X + \frac{C_0}{4} \sigma_k(\eta_1 + \eta_2)_X. \]
Similarly if $\| \eta_2 \|_X \leq \frac{C_0}{2} \sigma_{2k}(\eta_2)_X$ we will have
\[ \| \eta_2 \|_X \leq \frac{C_0}{2} \sigma_{2k}(\eta_2)_X \leq \frac{C_0}{4} \sigma_k(\eta_1 - \eta_2)_X + \frac{C_0}{4} \sigma_k(\eta_1 + \eta_2)_X. \]
It follows that
\[ \min(\| \eta_1 \|_X, \| \eta_2 \|_X) \leq \frac{C_0}{4} \sigma_k(\eta_1 - \eta_2)_X + \frac{C_0}{4} \sigma_k(\eta_1 + \eta_2)_X. \] (4.16)

Theorem 4.2 now implies that the required decoder $\Delta$ exists. Furthermore, by the proof of the sufficiency part of Theorem 4.2,
\[ \Delta(|Ax_0|) = \arg \min_{|Ax| = |Ax_0|} \sigma_k(x) \]

is one such decoder. \[ \square \]

4.3 The Case \( X = \ell_1 \)

We will now apply Theorem 4.3 to the \( \ell_1 \)-norm case. The following lemma establishes a relation between S-RIP and S-NSP for the \( \ell_1 \)-norm.

**Lemma 4.1** Let \( a, b, k \) be integers. Assume that \( A \in \mathbb{R}^{m \times N} \) satisfies the S-RIP of order \((a + b)k\) with constants \( \theta_-, \theta_+ \in (0, 2) \). Then \( A \) satisfies the S-NSP of order \( ak \) under the \( \ell_1 \)-norm with constant

\[ C_0 = 1 + \sqrt{\frac{a(1 + \delta)}{b(1 - \delta)}} \]

where \( \delta \) is the restricted isometry constant and \( \delta := \max\{1 - \theta_-, \theta_+ - 1\} < 1 \).

We remark that the above lemma is the analogous to the following lemma providing a relationship between RIP and NSP, which was shown in [5]:

**Lemma 4.2** ([5, Lemma 4.1]) Let \( a = l/k, b = l'/k \) where \( l, l' \geq k \) are integers. Assume that \( A \in \mathbb{R}^{m \times N} \) satisfies the RIP of order \((a + b)k\) with \( \delta = \delta_{(a+b)k} < 1 \). Then \( A \) satisfies the null space property under the \( \ell_1 \)-norm of order \( ak \) with constant

\[ C_0 = 1 + \sqrt{\frac{a(1 + \delta)}{b(1 - \delta)}} \]

**Proof** By the definition of S-RIP, for any index set \( I \subseteq \{1, \ldots, m\} \) with \(|I| \geq m/2\), the matrix \( A_I \in \mathbb{R}^{|I| \times N} \) satisfies the RIP of order \((a + b)k\) with constant \( \delta_{(a+b)k} = \delta := \max\{1 - \theta_-, \theta_+ - 1\} < 1 \). It follows from Lemma 4.2 that

\[ \|\eta\|_1 \leq \left(1 + \sqrt{\frac{a(1 + \delta)}{b(1 - \delta)}}\right) \sigma_{ak}(\eta)_1 \]

for all \( \eta \in \mathcal{N}(A_I) \). This proves the lemma. \[ \square \]

Set \( a = 2 \) and \( b = 1 \) in Lemma 4.1 we infer that if \( A \) satisfies the S-RIP of order \( 3k \) with constants \( \theta_-, \theta_+ \in (0, 2) \), then \( A \) satisfies the S-NSP of order \( 2k \) under the \( \ell_1 \)-norm with constant \( C_0 = 1 + \sqrt{\frac{2(1 + \delta)}{1 - \delta}} \). Hence by Theorem 4.3, there must exist a decoder that has the instance optimality under the \( \ell_1 \)-norm with constant \( 2C_0 \).

According to Theorem 2.1, by taking \( m = O(k \log(N/k)) \) a Gaussian random matrix \( A \) satisfies S-RIP of order \( 3k \) with high probability. Hence, there exists a decoder \( \Delta \) so that the pair \( (A, \Delta) \) has the the \( \ell_1 \)-norm phaseless instance optimality at the cost of \( m = O(k \log(N/k)) \) measurements, as with the traditional instance optimality.

We are now ready to prove the following theorem on phaseless instance optimality under the \( \ell_1 \)-norm.
Theorem 4.4 Let \( A \in \mathbb{R}^{m \times N} \) satisfy the S-RIP of order \( tk \) with constants \( 0 < \theta_- < 1 < \theta_+ < 2 \), where

\[
t \geq \max \left\{ \frac{2}{\theta_-}, \frac{2}{2 - \theta_+} \right\} > 2.
\]

Let

\[
\Delta(|Ax_0|) = \arg\min_{x \in \mathbb{R}^N} \{\|x\|_1 : |Ax| = |Ax_0|\}. \tag{4.17}
\]

Then \((A, \Delta)\) has the \( \ell_1 \)-norm phaseless instance optimality with constant \( C = \frac{2C_0}{2-C_0} \), where \( C_0 = 1 + \sqrt{\frac{1+\delta}{(t-1)(1-\delta)}} \) and as before

\[
\delta := \max\{1 - \theta_- - \theta_+ - 1\} \leq 1 - \frac{2}{t}.
\]

Proof of Lemma 4.1 Let \( x_0 \in \mathbb{R}^N \) and set \( \hat{x} = \Delta(|Ax_0|) \). Then by definition

\[
\|\hat{x}\|_1 \leq \|x_0\|_1 \quad \text{and} \quad |A\hat{x}| = |Ax_0|.
\]

Denote by \( I \subseteq \{1, \ldots, m\} \) the set of indices

\[
I = \left\{ j : \langle a_j, \hat{x} \rangle = \langle a_j, x_0 \rangle \right\},
\]

and thus \( \langle a_j, \hat{x} \rangle = -\langle a_j, x_0 \rangle \) for \( j \in I^c \). It follows that

\[
A_I(\hat{x} - x_0) = 0 \quad \text{and} \quad A_{I^c}(\hat{x} + x_0) = 0.
\]

Set

\[
\eta := \hat{x} - x_0 \in \mathcal{N}(A_I).
\]

We know that \( A \) satisfies the S-RIP of order \( tk \) with constants \( \theta_- , \theta_+ \) where

\[
t \geq \max \left\{ \frac{2}{\theta_-}, \frac{2}{2 - \theta_+} \right\} > 2.
\]

For the case \(|I| \geq m/2\), \( A_I \) satisfies the RIP of order \( tk \) with RIP constant

\[
\delta = \delta_{tk} := \max\{1 - \theta_- - \theta_+ - 1\} \leq 1 - \frac{2}{t} < 1.
\]

Take \( a := 1, \ b := t - 1 \) in Lemma 4.1. Then \( A \) satisfies the \( \ell_1 \)-norm S-NSP of order \( k \) with constant
$$C_0 = 1 + \sqrt{\frac{1 + \delta}{(t-1)(1 - \delta)}} < 2.$$ 

This yields

$$\|\eta\|_1 \leq C_0 \|\eta^{T^c}\|_1, \quad (4.18)$$

where $T$ is the index set for the $k$ largest coefficients of $x_0$ in magnitude. Hence $\|\eta_T\|_1 \leq (C_0 - 1)\|\eta^{T^c}\|_1$. Since $\|\hat{x}\|_1 \leq \|x_0\|_1$ we have

$$\|x_0\|_1 \geq \|\hat{x}\|_1 = \|x_0 + \eta\|_1 = \|x_0_T + x_0^{T^c} + \eta_T + \eta^{T^c}\|_1 \geq \|x_0_T\|_1 - \|x_0^{T^c}\|_1 + \|\eta^{T^c}\|_1 - \|\eta_T\|_1.$$ 

It follows that

$$\|\eta^{T^c}\|_1 \leq \|\eta_T\|_1 + 2\sigma_k(x_0)_1 \leq (C_0 - 1)\|\eta^{T^c}\|_1 + 2\sigma_k(x_0)_1$$

and thus

$$\|\eta^{T^c}\|_1 \leq \frac{2}{2 - C_0}\sigma_k(x_0)_1.$$ 

Now (4.18) yields

$$\|\eta\|_1 \leq C_0 \|\eta^{T^c}\|_1 \leq \frac{2C_0}{2 - C_0}\sigma_k(x_0)_1,$$

which implies

$$\|\hat{x} - x_0\|_1 \leq C_0 \|\eta^{T^c}\|_1 \leq \frac{2C_0}{2 - C_0}\sigma_k(x_0)_1.$$ 

For the case $|I^c| \geq m/2$ identical argument yields

$$\|\hat{x} + x_0\|_1 \leq C_0 \|\eta^{T^c}\|_1 \leq \frac{2C_0}{2 - C_0}\sigma_k(x_0)_1.$$ 

The theorem is now proved. \hfill \Box 

By Theorem 2.1, an $m \times N$ random Gaussian matrix with $m = \mathcal{O}(tk \log(N/k))$ satisfies the S-RIP of order $tk$ with high probability. We therefore conclude that the $\ell_1$-norm phaseless instance optimality of order $k$ can be achieved at the cost of $m = \mathcal{O}(tk \log(N/k))$ measurements.
4.4 Mixed-Norm phaseless Instance Optimality

We now consider mixed-norm phaseless instance optimality. Let $1 \leq q \leq p \leq 2$ and $s = 1/q - 1/p$. We seek estimates of the form

$$\min\{\|x - \Delta(|Ax|)\|_p, \|x + \Delta(|Ax|)\|_p\} \leq C_0 k^{-s} \sigma_k(x)_q \quad (4.19)$$

for all $x \in \mathbb{R}^N$. We shall prove both necessary and sufficient conditions for mixed-norm phaseless instance optimality.

**Theorem 4.5** Let $A \in \mathbb{R}^{m \times N}$ and $1 \leq q \leq p \leq 2$. Set $s = 1/q - 1/p$. Then a decoder $\Delta$ satisfying the mixed norm phaseless instance optimality (4.19) with constant $C_0$ exists if: for any index set $I \subseteq \{1, \ldots, m\}$ and any $\eta_1 \in \mathcal{N}(A_I)$, $\eta_2 \in \mathcal{N}(A_{I^c})$ we have

$$\min\{\|\eta_1\|_p, \|\eta_2\|_p\} \leq C_0 \frac{k^{-s}}{4} \left(\sigma_k(\eta_1 - \eta_2)_q + \sigma_k(\eta_1 + \eta_2)_q\right). \quad (4.20)$$

Conversely, assume a decoder $\Delta$ satisfying the mixed norm phaseless instance optimality (4.19) exists. Then for any index set $I \subseteq \{1, \ldots, m\}$ and any $\eta_1 \in \mathcal{N}(A_I)$, $\eta_2 \in \mathcal{N}(A_{I^c})$ we have

$$\min\{\|\eta_1\|_p, \|\eta_2\|_p\} \leq C_0 \frac{k^{-s}}{2} \left(\sigma_k(\eta_1 - \eta_2)_q + \sigma_k(\eta_1 + \eta_2)_q\right).$$

**Proof of Lemma 4.1** The proof is virtually identical to the proof of Theorem 4.2. We shall omit the details here in the interest of brevity. □

**Definition 4.2** (Mixed-Norm Strong Null Space Property) We say that $A$ has the mixed strong null space property in norms $(\ell_p, \ell_q)$ of order $k$ with constant $C$ if for any index set $I \subseteq \{1, \ldots, m\}$ with $|I| \geq m/2$ the matrix $A_I \in \mathbb{R}^{|I| \times N}$ satisfies

$$\|\eta\|_p \leq C k^{-s} \sigma_k(\eta)_q$$

for all $\eta \in \mathcal{N}(A_I)$, where $s = 1/q - 1/p$.

The above is an extension of the standard definition of the mixed null space property of order $k$ in norms $(\ell_p, \ell_q)$ (see [5]) for a matrix $A$, which requires

$$\|\eta\|_p \leq C k^{-s} \sigma_k(\eta)_q$$

for all $\eta \in \mathcal{N}(A)$. We have the following straightforward generalization of Theorem 4.3.

**Theorem 4.6** Assume that $A \in \mathbb{R}^{m \times N}$ has the mixed strong null space property of order $2k$ in norms $(\ell_p, \ell_q)$ with constant $C_0/2$, where $1 \leq q \leq p \leq 2$. Then there exists a decoder $\Delta$ such that the mixed-norm phaseless instance optimality (4.19) holds with constant $C_0$. 

\[ \text{Birkhäuser} \]
We establish relationships between mixed-norm strong null space property and the S-RIP. First we present the following lemma that was proved in [5].

**Lemma 4.3** ([5, Lemma 8.2]) Let \( k \geq 1 \) and \( \tilde{k} = \lceil k(\frac{N}{k})^{2-2/q} \rceil \). Assume that \( A \in \mathbb{R}^{m \times N} \) satisfies the RIP of order \( 2k + \tilde{k} \) with \( \delta := \delta_{2k+\tilde{k}} < 1 \). Then \( A \) satisfies the mixed null space property in norms \((\ell_p, \ell_q)\) of order \( 2k \) with constant \( C_0 = 2^{1/p+1/2} \sqrt{\frac{1+\delta}{1-\delta}} + 2^{1/p-1/q} \).

**Proposition 4.1** Let \( k \geq 1 \) and \( \tilde{k} = \lceil k(\frac{N}{k})^{2-2/q} \rceil \). Assume that \( A \in \mathbb{R}^{m \times N} \) satisfies the S-RIP of order \( 2k + \tilde{k} \) with constants \( 0 < \theta_- < 1 < \theta_+ < 2 \). Then \( A \) satisfies the mixed strong null space property in norms \((\ell_p, \ell_q)\) of order \( 2k \) with constant \( C_0 = 2^{1/p+1/2} \sqrt{\frac{1+\delta}{1-\delta}} + 2^{1/p-1/q} \), where \( \delta \) is the RIP constant and \( \delta := \delta_{2k+\tilde{k}} = \max\{1 - \theta_-, \theta_+ - 1\} \).

**Proof of Lemma 4.1** By definition for any index set \( I \subseteq \{1, \ldots, m\} \) with \( |I| \geq m/2 \), the matrix \( A_I \in \mathbb{R}^{|I| \times N} \) satisfies RIP of order \( 2k + \tilde{k} \) with constant \( C_0 = 2^{1/p+1/2} \sqrt{\frac{1+\delta}{1-\delta}} + 2^{1/p-1/q} \), where \( \delta \) is the RIP constant and \( \delta := \delta_{2k+\tilde{k}} = \max\{1 - \theta_-, \theta_+ - 1\} \). By Lemma 4.3, we know that \( A_I \) satisfies the mixed null space property in norms \((\ell_p, \ell_q)\) of order \( 2k \) with constant \( C_0 = 2^{1/p+1/2} \sqrt{\frac{1+\delta}{1-\delta}} + 2^{1/p-1/q} \), in other words for any \( \eta \in \mathcal{N}(A_I) \),

\[
\|\eta\|_p \leq Ck^{-s} \sigma_{2k}(\eta)_q.
\]

So \( A \) satisfies the mixed strong null space property. \( \square \)

**Corollary 4.1** Let \( k \geq 1 \) and \( \tilde{k} = k(\frac{N}{k})^{2-2/q} \). Assume that \( A \in \mathbb{R}^{m \times N} \) satisfies the S-RIP of order \( 2k + \tilde{k} \) with constants \( 0 < \theta_- < 1 < \theta_+ < 2 \). Let \( \delta := \delta_{2k+\tilde{k}} = \max\{1 - \theta_-, \theta_+ - 1\} < 1 \). Define the decoder \( \Delta \) for \( A \) by

\[
\Delta(|Ax|) = \arg\min_{|Ax| = |Ax_0|} \sigma_k(x)_q.
\] (4.21)

Then (4.19) holds with constant \( 2C_0 \), where \( C_0 = 2^{1/p+1/2} \sqrt{\frac{1+\delta}{1-\delta}} + 2^{1/p-1/q} \).

**Proof of Lemma 4.1** By the Proposition 4.1, the matrix \( A \) satisfies the mixed strong null space property in \((\ell_p, \ell_q)\) of order \( 2k \) with constant \( C_0 = 2^{1/p+1/2} \sqrt{\frac{1+\delta}{1-\delta}} + 2^{1/p-1/q} \). The corollary now follows immediately from Theorem 4.6. \( \square \)

**Remark 4.1** Combining Theorem 2.1 and Corollary 4.1, the mixed phaseless instance optimality of order \( k \) in norms \((\ell_p, \ell_q)\) can be achieved for the price of \( O(k(N/k)^{2-2/q} \log(N/k)) \) measurements, just as with the traditional mixed \((\ell_p, \ell_q)\)-norm instance optimality. Theorem 3.1 implies that the \( \ell_1 \) decoder satisfies the \((p, q) = (2, 1) \) mixed-norm phaseless instance optimality at the price of \( O(k \log(N/k)) \) measurements.
Appendix: Proof of Lemma 2.1

We will first need the following two Lemmas to prove Lemma 2.1.

**Lemma 5.1** (Sparse Representation of a Polytope [2,12]) Let $s \geq 1$ and $\alpha > 0$. Set

$$T(\alpha, s) := \left\{ u \in \mathbb{R}^n : \|u\|_{\infty} \leq \alpha, \|u\|_1 \leq s\alpha \right\}.$$ 

For any $v \in \mathbb{R}^n$ let

$$U(\alpha, s, v) := \left\{ u \in \mathbb{R}^n : \text{supp}(u) \subseteq \text{supp}(v), \|u\|_0 \leq s, \|u\|_1 = \|v\|_1, \|u\|_{\infty} \leq \alpha \right\}.$$ 

Then $v \in T(\alpha, s)$ if and only if $v$ is in the convex hull of $U(\alpha, s, v)$, i.e. $v$ can be expressed as a convex combination of some $u_1, \ldots, u_N$ in $U(\alpha, s, v)$.

**Lemma 5.2** ([1, Lemma 5.3]) Assume that $a_1 \geq a_2 \geq \cdots \geq a_m \geq 0$. Let $r \leq m$ and $\lambda \geq 0$ such that $\sum_{i=1}^r a_i + \lambda \geq \sum_{i=r+1}^m a_i$. Then for all $\alpha \geq 1$ we have

$$\sum_{j=r+1}^m a_j^\alpha \leq r \left( \frac{\sum_{i=1}^r a_i^\alpha}{r} + \frac{\lambda}{r} \right)^\alpha.$$ 

(5.1)

In particular for $\lambda = 0$ we have

$$\sum_{j=r+1}^m a_j^\alpha \leq \sum_{i=1}^r a_i^\alpha.$$ 

We are now ready to prove Lemma 2.1.

**Proof** Set $h := \hat{x} - x_0$. Let $T_0$ denote the set of the largest $k$ coefficients of $x_0$ in magnitude. Then

$$\|x_0\|_1 + \rho \geq \|\hat{x}\|_1 = \|x_0 + h\|_1$$

$$= \|x_0, T_0 + h_{T_0} + x_0, T_c\|_1$$

$$\geq \|x_0, T_0\|_1 - \|h_{T_0}\|_1 - \|x_0, T_0\|_1 + \|h_{T_0}\|_1.$$ 

It follows that

$$\|h_{T_0}\|_1 \leq \|h_{T_0}\|_1 + 2\|x_0, T_0\|_1 + \rho$$

$$= \|h_{T_0}\|_1 + 2\sigma_k(x_0)_1 + \rho.$$
Suppose that $S_0$ is the index set of the $k$ largest entries in absolute value of $h$. Then we can get
\[
\|h_{S_0}^c\|_1 \leq \|h_{T_0}^c\|_1 \leq \|h_{T_0}\|_1 + 2\sigma_k(x_0) + \rho \\
\leq \|h_{S_0}\|_1 + 2\sigma_k(x_0) + \rho.
\]
Set
\[
\alpha := \frac{\|h_{S_0}\|_1 + 2\sigma_k(x_0) + \rho}{k}.
\]
We divide $h_{S_0}^c$ into two parts $h_{S_0}^c = h^{(1)} + h^{(2)}$, where
\[
h^{(1)} := h_{S_0}^c \cdot I_{\{i : |h_{S_0}^c(i)| > \alpha/(t-1)\}}, \quad h^{(2)} := h_{S_0}^c \cdot I_{\{i : |h_{S_0}^c(i)| \leq \alpha/(t-1)\}}.
\]
A simple observation is that $\|h^{(1)}\|_1 \leq \|h_{S_0}^c\|_1 \leq \alpha k$. Set
\[
\ell := |\text{supp}(h^{(1)})| = \|h^{(1)}\|_0.
\]
Since all non-zero entries of $h^{(1)}$ have magnitude larger than $\alpha/(t-1)$, we have
\[
\alpha k \geq \|h^{(1)}\|_1 = \sum_{i \in \text{supp}(h^{(1)})} |h^{(1)}(i)| \geq \sum_{i \in \text{supp}(h^{(1)})} \frac{\alpha}{t-1} = \frac{\alpha \ell}{t-1},
\]
which implies $\ell \leq (t-1)k$. Thus we have:
\[
\langle A(h_{S_0} + h^{(1)}), Ah \rangle \leq \|A(h_{S_0} + h^{(1)})\|_2 \cdot \|Ah\|_2 \leq \sqrt{1 + \delta} \cdot \|h_{S_0} + h^{(1)}\|_2 \cdot \epsilon.
\] (5.2)
Here we apply the facts that $\|h_{S_0} + h^{(1)}\|_0 = \ell + k = tk$ and $A$ satisfies the RIP of order $tk$ with $\delta := \delta^A_{tk}$. We shall assume at first that $tk$ as an integer. Note that $\|h^{(2)}\|_\infty \leq \frac{\alpha}{t-1}$ and
\[
\|h^{(2)}\|_1 = \|h_{S_0}^c\|_1 - \|h^{(1)}\|_1 \leq k\alpha - \frac{\alpha \ell}{t-1} = (k(t-1) - \ell) \frac{\alpha}{t-1}.
\] (5.3)
We take $s := k(t-1) - \ell$ in Lemma 5.1 and obtain that $h^{(2)}$ is a weighted mean
\[
h^{(2)} = \sum_{i=1}^{N} \lambda_i u_i, \quad 0 \leq \lambda_i \leq 1, \quad \sum_{i=1}^{N} \lambda_i = 1
\]
where $\|u_i\|_0 \leq k(t-1) - \ell$, $\|u_i\|_1 = \|h^{(2)}\|_1$, $\|u_i\|_\infty \leq \alpha/(t-1)$ and $\text{supp}(u_i) \subseteq \text{supp}(h^{(2)})$. Hence
\[ \|u_i\|_2 \leq \sqrt{\|u_i\|_0 \cdot \|u_i\|_\infty} = \sqrt{k(t-1)} \cdot \|u_i\|_\infty \leq \sqrt{k(t-1)} \cdot \|u_i\|_\infty \leq \alpha \sqrt{k/(t-1)}. \]

Now for \( 0 \leq \mu \leq 1 \) and \( d \geq 0 \), which will be chosen later, set

\[ \beta_j := h_{S_0} + h^{(1)} + \mu \cdot u_j, \quad j = 1, \ldots, N. \]

Then for fixed \( i \in [1, N] \)

\[ \sum_{j=1}^{N} \lambda_j \beta_j - d \beta_i = h_{S_0} + h^{(1)} + \mu \cdot h^{(2)} - d \beta_i = (1 - \mu - d)(h_{S_0} + h^{(1)}) - d \mu u_i + \mu h. \]

Recall that \( \alpha = \frac{\|h_{S_0}\|_1 + 2\sigma_k(x_0) + \rho}{k} \). Thus

\[ \|u_i\|_2 \leq \frac{\sqrt{k/(t-1)} \alpha}{\sqrt{t-1}} \leq \frac{\|h_{S_0}\|_2 + 2\sigma_k(x_0) + \rho}{\sqrt{k/(t-1)}} \leq \frac{\|h_{S_0} + h^{(1)}\|_2}{\sqrt{t-1}} + \frac{2\sigma_k(x_0) + \rho}{\sqrt{k/(t-1)}} = \frac{z + R}{\sqrt{t-1}}, \]

where \( z := \|h_{S_0} + h^{(1)}\|_2 \) and \( R := \frac{2\sigma_k(x_0) + \rho}{\sqrt{k}} \). It is easy to check the following identity:

\[ (2d - 1) \sum_{1 \leq i < j \leq N} \lambda_i \lambda_j \|A(\beta_i - \beta_j)\|^2_2 = \sum_{i=1}^{N} \lambda_i \|A(\sum_{j=1}^{N} \lambda_j \beta_j - d \beta_i)\|^2_2 - \sum_{i=1}^{N} \lambda_i (1 - d)^2 \|A \beta_i\|^2_2, \]

provided that \( \sum_{i=1}^{N} \lambda_i = 1 \). Choose \( d = 1/2 \) in (5.5) we then have

\[ \sum_{i=1}^{N} \lambda_i \|A \left( \frac{1}{2} - \mu \right)(h_{S_0} + h^{(1)}) - \frac{\mu}{2} u_i + \mu h \|^2_2 - \sum_{i=1}^{N} \frac{\lambda_i}{4} \|A \beta_i\|^2_2 = 0. \]
Note that for $d = 1/2$,

\[
\left\| A \left( \frac{1}{2} - \mu \right) (h_{S_0} + h^{(1)}) - \frac{\mu}{2} u_i + \mu h \right\|^2_2 \\
= \left\| A \left( \frac{1}{2} - \mu \right) (h_{S_0} + h^{(1)}) - \frac{\mu}{2} u_i \right\|^2_2 \\
+ 2 \left\langle A \left( \frac{1}{2} - \mu \right) (h_{S_0} + h^{(1)}) - \frac{\mu}{2} u_i, \mu A h \right\rangle + \mu^2 \| Ah \|^2_2.
\]

It follows from $\sum_{i=1}^N \lambda_i = 1$ and $h^{(2)} = \sum_{i=1}^N \lambda_i u_i$ that

\[
\sum_{i=1}^N \lambda_i \left\| A \left( \frac{1}{2} - \mu \right) (h_{S_0} + h^{(1)}) - \frac{\mu}{2} u_i + \mu h \right\|^2_2 \\
= \sum_{i=1}^N \lambda_i \left\| A \left( \frac{1}{2} - \mu \right) (h_{S_0} + h^{(1)}) - \frac{\mu}{2} u_i \right\|^2_2 \\
+ 2 \left\langle A \left( \frac{1}{2} - \mu \right) (h_{S_0} + h^{(1)}) - \frac{\mu}{2} h^{(2)}, \mu A h \right\rangle + \mu^2 \| Ah \|^2_2 \\
= \sum_{i=1}^N \lambda_i \left\| A \left( \frac{1}{2} - \mu \right) (h_{S_0} + h^{(1)}) - \frac{\mu}{2} u_i \right\|^2_2 \\
+ \mu (1 - \mu) \left\langle A (h_{S_0} + h^{(1)}), Ah \right\rangle - \sum_{i=1}^N \frac{\lambda_i}{4} \| A \beta_i \|^2_2. \tag{5.6}
\]

Set $\mu = \sqrt{t(t-1)} - (t-1)$. We next estimate the three terms in (5.6). Noting that $\|h_{S_0}\|_0 \leq k$, $\|h^{(1)}\|_0 \leq \ell$ and $\|u_i\|_0 \leq s = k(t-1) - \ell$, we obtain

\[
\| \beta_i \|_0 \leq \|h_{S_0}\|_0 + \|h^{(1)}\|_0 + \|u_i\|_0 \leq t \cdot k
\]

and $\| (\frac{1}{2} - \mu) (h_{S_0} + h^{(1)}) - \frac{\mu}{2} u_i \|_0 \leq t \cdot k$. Since $A$ satisfies the RIP of order $t \cdot k$ with $\delta$, we have

\[
\left\| A \left( \frac{1}{2} - \mu \right) (h_{S_0} + h^{(1)}) - \frac{\mu}{2} u_i \right\|^2_2 \\
\leq (1 + \delta) \left\| (\frac{1}{2} - \mu) (h_{S_0} + h^{(1)}) - \frac{\mu}{2} u_i \right\|^2_2 \\
= (1 + \delta) \left( \left( \frac{1}{2} - \mu \right)^2 \|h_{S_0} + h^{(1)}\|^2_2 + \frac{\mu^2}{4} \|u_i\|^2_2 \right) \\
= (1 + \delta) \left( \left( \frac{1}{2} - \mu \right)^2 z^2 + \frac{\mu^2}{4} \|u_i\|^2_2 \right)
\]

and

\[
\| A \beta_i \|^2_2 \geq (1 - \delta) \| \beta_i \|^2_2 = (1 - \delta) (\|h_{S_0} + h^{(1)}\|^2 + \mu^2 \cdot \|u_i\|^2_2) \\
= (1 - \delta) (z^2 + \mu^2 \cdot \|u_i\|^2_2).
\]
Combining the result above with (5.2) and (5.4) we get

\[
0 \leq (1 + \delta) \sum_{i=1}^{N} \lambda_i \left( \left(\frac{1}{2} - \mu \right)^2 z^2 + \frac{\mu^2}{4} \|u_i\|^2_2 \right) + \mu (1 - \mu) \sqrt{1 + \delta} \cdot z \cdot \epsilon
\]

\[
- (1 - \delta) \sum_{i=1}^{N} \frac{\lambda_i}{4} (z^2 + \mu^2 \|u_i\|^2_2)
\]

\[
= \sum_{i=1}^{N} \lambda_i \left( \left(1 + \delta \right) \left(\frac{1}{2} - \mu \right)^2 - \frac{1 - \delta}{4} \right) z^2 + \frac{\delta}{2} \mu^2 \|u_i\|^2_2 \right) + \mu (1 - \mu) \sqrt{1 + \delta} \cdot z \cdot \epsilon
\]

\[
\leq \sum_{i=1}^{N} \lambda_i \left( \left(1 + \delta \right) \left(\frac{1}{2} - \mu \right)^2 - \frac{1 - \delta}{4} \right) z^2 + \frac{\delta}{2} \mu^2 \left( z + R \right)^2 \right)
\]

\[
+ \mu (1 - \mu) \sqrt{1 + \delta} \cdot z \cdot \epsilon
\]

\[
= \left( \mu^2 - \mu + \delta \left(\frac{1}{2} - \mu + \frac{1}{2(t - 1)} (\mu^2) \right) \right) z^2
\]

\[
+ \left( \mu (1 - \mu) \sqrt{1 + \delta} \cdot \epsilon + \frac{\delta \mu^2 R^2}{2(t - 1)} \right) z + \frac{\delta \mu^2 R^2}{2(t - 1)}
\]

\[
= -t \left( 2t - 1 \right) \left( 2\sqrt{t(t - 1)}\right) \left( \frac{1}{t} - \delta \right) z^2
\]

\[
+ \mu^2 \sqrt{t} \left( 1 + \delta \cdot \epsilon + \frac{\delta \mu^2 R}{2(t - 1)} \right) z + \frac{\delta \mu^2 R^2}{2(t - 1)}
\]

\[
= \frac{\mu^2}{t - 1} \left( -t \sqrt{\frac{1}{t} - \delta} z^2 + \sqrt{t(t - 1)} (1 + \delta) \epsilon + \delta R) z + \frac{\delta R^2}{2} \right),
\]

which is a quadratic inequality for \(z\). We know \(\delta < \sqrt{(t - 1)/t}\). So by solving the above inequality we get

\[
z \leq \sqrt{t(t - 1)(1 + \delta) \epsilon + \delta R) + ((\sqrt{t(t - 1)} (1 + \delta) \epsilon + \delta R)^2 + 2t(\sqrt{(t - 1)/t} - \delta) R^2)}/2t(\sqrt{(t - 1)/t})
\]

\[
\leq \frac{\sqrt{t(t - 1)(1 + \delta) \epsilon + \delta R) + 2\delta + 2t(\sqrt{(t - 1)/t} - \delta) R^2}}{2t(\sqrt{(t - 1)/t} - \delta)}.
\]

Finally, noting that \(\|h_{S_0}\|_1 \leq \|h_{S_0}\|_1 + R\sqrt{k}\), in the Lemma 5.2, if we set \(m = N\), \(r = k\), \(\delta = R\sqrt{k} \geq 0\) and \(\alpha = 2\) then \(\|h_{S_0}\|_2 \leq \|h_{S_0}\|_2 + R\). Hence

\[
\|h\|_2 = \sqrt{\|h_{S_0}\|_2^2 + \|h_{S_0}\|_2^2}
\]

\[
\leq \sqrt{\|h_{S_0}\|_2^2 + (\|h_{S_0}\|_2 + R)^2}
\]
\[
\leq \sqrt{2}\|hS_0\|_2^2 + R \leq \sqrt{2}z + R \\
\leq \frac{\sqrt{2}(1 + \delta)}{1 - \sqrt{t/(t-1)}}\epsilon + \left(\frac{\sqrt{2}\delta + \sqrt{t((t-1)/t-\delta)}\delta}{t((t-1)/t-\delta)} + 1\right)R.
\]

Substitute \(R\) into this inequality and the conclusion follows.

For the case where \(t \cdot k\) is not an integer, we set \(t^* := \lceil tk \rceil / k\), then \(t^* > t\) and \(\delta t^*_k = \delta_t k < \sqrt{\frac{t-1}{t}} < \sqrt{\frac{t^*-1}{t^*}}\). We can then prove the result by working on \(\delta t^*_k\). \(\square\)

**Acknowledgments** Yang Wang was supported in part by the AFOSR grant FA9550-12-1-0455 and NSF grant IIS-1302285. Zhiqiang Xu was supported by NSFC grant (11171336, 11422113, 11021101, 11331012) and by National Basic Research Program of China (973 Program 2015CB856000).

**References**

1. Cai, T.T., Zhang, A.: Sharp RIP bound for sparse signal and low-rank matrix recovery. Appl. Comput. Harmon. Anal. 35(1), 74–93 (2013)
2. Cai, T.T., Zhang, A.: Sparse representation of a polytope and recovery of sparse signals and low-rank matrices. IEEE Trans. Inf. Theory 60(1), 122–132 (2014)
3. Candès, E.J., Tao, T.: Decoding by linear programming. IEEE Trans. Inf. Theory 51(12), 4203–4215 (2005)
4. Candès, E.J., Romberg, J.K., Tao, T.: Stable signal recovery from incomplete and inaccurate measurements. Commun. Pure Appl. Math. 59(8), 1207–1223 (2006)
5. Cohen, A., Dahmen, W., Devore, R.: Compressed sensing and best \(k\)-term approximation. J. Am. Math. Soc. 22(1), 211–231 (2009)
6. Eldar, Y.C., Mendelson, S.: Phase retrieval: stability and recovery guarantees. Appl. Comput. Harmon. Anal. 36(3), 473–494 (2014)
7. Han, B., Xu, Z. Q.: Robustness properties of dimensionality reduction with Gaussian random matrices. arXiv:1501.01695 (2015)
8. Iwen, M., Viswanathan, A., Wang, Y.: Robust sparse phase retrieval made easy. arXiv:1410.5295 (2015)
9. Moravec, M.L., Romberg, J.K., Baraniuk, R.G.: Compressive phase retrieval. In: SPIE Proceedings, vol. 6701 (2007)
10. Voroninski, V., Xu, Z.Q.: A strong restricted isometry property, with an application to phaseless compressed sensing. Appl. Comput. Harmon. Anal. (2015). doi:10.1016/j.acha.2015.06.004
11. Wang, Y., Xu, Z.Q.: Phase retrieval for sparse signals. Appl. Comput. Harmon. Anal. 37(3), 531–544 (2014)
12. Xu, G.W., Xu, Z.Q.: On the \(\ell_1\)-sparse decomposition of signals. J. Oper. Res. Soc. China 1(4), 537–541 (2013)