Data recovery from grossly corrupted observation via l1 minimization

Dongcai Su at 2015-9-22

Abstract

This paper studies the problem of recovering a signal vector and the corrupted noise vector from a collection of corrupted linear measurements through the solution of a l1 minimization (2.1.2), where the sensing matrix is a partial Fourier matrix whose rows are selected randomly and uniformly from rows of a full Fourier matrix. After choosing the parameter in (2.1.2) appropriately, we show that the recovery can be successful even when a constant fraction of the measurement are arbitrarily corrupted, provided that the signal vector is sparse enough. The upper-bound on the sparsity of the signal vector required in this paper is asymptotically optimal when the sensing matrix is partial bounded orthonormal system (BOS) [1] and is better than those achieved by recent literatures [2, 3] by a $\ln(n)$ factor. Furthermore, the assumptions we impose on the signal vector and the corrupted noise vector are loosest comparing to the existing literatures [2-4] which lenders our recovery guarantees are more applicable. Extensive numerical experiments based on synthesis as well as real world data are presented to verify the conclusion of the proposed theorem and to demonstrate the potential of the l1 minimization framework.

Keywords—Compressed sensing, Fourier basis, partial BOS, l1 minimization, corrupted observations

1. Introduction

Compressed sensing have been well studied [5-7] recently as a novel data sampling framework, which have attracted great attentions in scientific areas [8] as well as industrial applications [9-13]. In a CS sampling framework, a sparse n-dimensional signal $x^{(0)}$ is recovered from a collection of linear projections $b = Ax^{(0)}$, where $A$ is a $m \times n$ sensing matrix, typically, the number of measurement $m$ is significantly smaller than the dimension of $x^{(0)}$ $n$, to recover $x^{(0)}$, a standard $\ell_1$ minimization is often adopted in the CS literature:

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

(1.1)

Typical theoretical results in the CS literatures show that if $A$ obeys some properties, e.g., the restricted isometry property (RIP) [14, 15], then the solution of (1.1) can recover $x^{(0)}$ with high probability provided that $x^{(0)}$ is sparse enough [16].

It is now well-known that RIP is satisfied with high probability for a broad family of sensing

---

1 Part of this paper had been submitted to The 2016 IEEE International Symposium on Information Theory (isit 2016)
matrices, e.g., when A is the sub-gaussian matrix [1, 8, 17], then A satisfies the RIP with high probability if \( m \geq O\left(k\ln\left(\frac{en}{k}\right)\right) \). Similarly, when A is the partial bounded orthogonal system (BOS)[1], then A satisfies the RIP as long as \( m \geq O(k\ln^4(n)) \) [17], where \( k \geq \max\left\{ 1, \|x^{(0)}\|_0 \right\} \).

In many applications, the measurement vector \( b \) may also contain certain amount of dense, bounded noise, which is denoted by a \( m \)-dimensional vector \( v \), and \( b \) is expressed as: \( b = Ax + v \), with \( \|v\|_2 \leq \eta \). To recover \( x^{(0)} \) in this case, a convex optimization is usually adopted [16]:

\[
\min_x \|x\|_1, \text{ s.t. } \|Ax - b\|_2 \leq \eta \tag{1.2}
\]

It can be proved that if A satisfies the RIP, then stable recovery\(^2\) of \( x^{(0)} \) by (1.2) is still possible based on the same number of measurements as in (1.1) [14, 18-21]. Alternatively, similar exact or stable recovery results corresponding to (1.1~1.2) can also be obtained through other algorithms such as orthogonal matching pursuit (OMP) [22, 23], and the analysis is based on the mutual coherence property of the sensing matrix, which is closely related to the RIP chap 5~6 of [1].

Generally speaking, it require asymptotically more number of measurement to guarantee the RIP of \( A \) or successful recovery of (1.2) when the sensing matrix \( A \) is partial BOS than when \( A \) is Gaussian random matrix. Chap. 12 of [1].

However, as pointed out in chap.12 [1], manipulating a completely random sensing matrix might add extra storage or computation burden, for example, storing a partial Fourier basis as the sensing matrix requires far more less memories than storing a random i.i.d Gaussian matrix with the same size, besides, algorithms (e.g. alternating direction multipliers method (ADMM) [24] and Bregman method [25]) solving the \( \ell_1 \) minimization (1.1) with \( A \) as the partial Fourier basis matrix can be much faster than in general case.

Furthermore, as commented by [26, 27], completely random sensing matrix might not be relevant in real applications, for instance, the data acquisition systems in many practical applications are inherently frequent based, in other words, the sensing matrices corresponding to these data acquisition systems are essential partial Fourier basis matrices, we should just mentioned a few of important applications lie in this case as stated below:

- Magnetic Resonance Imaging (MRI). Chap. 1 of [1] [9-12]. Let \( \Omega \subset \mathbb{R}^3 \) be the interested region for imaging, the magnetization at position \( z \in \Omega \) is denoted by \( X(z) = |X(z)|\exp(-2\pi i(k(t), z)) \), where \( |X(z)| > 0 \) denotes the magnitude, \( k(t) = [k_1(t), k_2(t), k_3(t)]^T \in \mathbb{R}^3 \) is a time dependent vector and \( \langle \cdot, \cdot \rangle \) denotes the inner product. Then at time \( t \), the receiver coil integrates \( X(z) \) over \( \Omega \) and measures the value

\[
f(t) = \int_{\Omega} |X(z)|\exp(-2\pi i(k(t), z))dz.
\]

To write it in a discretization way, firstly, we represents \( \Omega \) by \( N_1 \times N_2 \times N_3 = N \) voxels, let \( x \in \mathbb{R}^N \) denotes the vector to be

\(^2\) Here stable recovery means \( \|\hat{x} - x^{(0)}\|_2 \leq C\eta \), where \( \hat{x} \) denotes the solution of (1.2), \( C \) is some positive constant.
reconstructed where \( x(j), 1 \leq j \leq N \) denotes the value of \(|X(z)|\) of the corresponding voxel. Secondly, we also we also sample \( k(t) = [k_1(t), k_2(t), k_3(t)]^T \) randomly and uniformly within a region: \( 0 \leq k_j(t) < N_j, 1 \leq j \leq 3 \) and \( t \in [1, T] \) be a integer. In this manner, one has, \( b = Ax \) . Where \( b \in C^T \) denotes the measurement vector with \( b(t) = f(t), t \in [1, T], \) and \( A \in C^{T \times N} \) is a submatrix whose rows are selected randomly and uniformly from a \( N \times N \) Fourier basis.

- Radar. In this application, at time \( t \), a source antenna sends out a properly designed electromagnetic wave, then a receive antenna measures the electromagnetic signal reflected by objects in the sky. Mathematically, let \( z \) denotes object in the sky which is parameterized by \( d \)—the distance between the object and the receive antenna, then at time \( t \), the measurement of the receive antenna reads \( b(t) = \sum_d |z(d)| \exp(-2\pi i \phi(d, t)), \) where \( |z(d)| \) denotes the magnitude of \( z(d), |z(d_0)| = 0 \) if and only if there is no object parameterized by \( d = d_0 \). \( \phi(d, t) = c d f(t) \) denotes the phase translation which depends on \( d \) and \( t \), where \( c \) is a positive constant depends on factors like wave speed, and \( f(t) \) denotes the frequency of the electromagnetic wave sent by the source antenna at time \( t \), after proper rescaling operations, let \( d \in [0, N - 1] \) be integer, let \( x \in R^N \) be a vector whose element denotes the value \( |z(d)| \) of the corresponding objects. Then we have:

\[
b(t) = \sum_d |z(d)| \exp \left( - \frac{2\pi i d f(t)}{N} \right), \quad t \in [1, T] \text{ are integers, if we choose } f(t) \text{ randomly and uniformly within a given region, then one can express the measurement vector } b \in C^T \text{ compactly as: } \ b = Ax, \text{ where } A \text{ is a } T \times N \text{ matrix whose rows are selected uniformly at random from a } N \times N \text{ Fourier basis.}
\]

- Planar X-ray. [28] Let \( \Omega \subset R^3 \) denotes the interested region, let \( f(z) \in R, z \in \Omega \) represents the X-ray reaction coefficient of the position \( z \), our goal is to reconstruct \( f(z) \). A plane is defined by \( P(d, \gamma(t)) = \{z| z^T \gamma(t) = d\} \), where \( \gamma(t) \in R^3 \) is a normal vector dependent on \( t \), a radon transform defined on the plane reads: \( R_d(g, \gamma(t)) = \int_{\Omega \cap P(d, \gamma(t))} f(z) dz \). Suppose at time \( t \), we’ve measured a collection of radon transform defined by \( \gamma(t): \ R(t) = \{R_d(g, \gamma(t))|L(\gamma(t)) \leq d \leq U(\gamma(t))\} \), where \( L(\gamma(t)), U(\gamma(t)) \) are the minimum, maximum value of \( d \) such that \( P(d, \gamma(t)) \cap \Omega \neq \emptyset \). The 1-D Fourier transform of \( R_d(g, \gamma(t)) \) along the \( d \) direction is denoted as \( \hat{R}_t(s, \gamma(t)) \), \( s \in R \). It can be proved that when \( f(z) \) satisfies certain conditions, \( \hat{R}_t(s, \gamma(t)) = \hat{f}(s\gamma(t)) \) holds, where \( \hat{f}(s\gamma) \), \( \gamma \in R^3 \) denotes the 3D Fourier transform of \( f(z), z \in R^3 \). In discretization case, if one choose \( \gamma(t) \) uniformly at random, then one has: \( b = Ax, \) where \( b \) is the measurement vector whose elements are obtained by Fourier transform of the radon transform as described above, \( A \) is a wide matrix whose rows are selected randomly and uniformly from a Fourier basis, and \( x \) is the vector whose element represent \( f(z) \) of the corresponding position \( z \).

These applications motivate us to study the signal recovery when the sensing matrix \( A \) is partial Fourier basis.

When \( A \) is partial Fourier basis and the measurement vector \( b \) contains some noise whose norm is bounded above by \( \eta \), there exist elegant recovery guarantee for (1.2): the analysis based on RIP [17] show that, if \( m \geq O(kd^4(n)) \), where \( k \) is a positive integral represents the upper-bound
Recently, the matrix can be treated as a special case, more recently, Nguyen [3] shows that when A is a partial BOS (where partial Fourier matrix can be treated as a special case), after choosing \( \theta \) appropriately as depending on \( m \) and \( n \), then recovery of \( x^{(0)} \) and \( f^{(0)} \) is possible even when a constant fraction of \( b \) is corrupted by \( f^{(0)} \), provided that \( m \geq O(\|x^{(0)}\|_0 \ln^2(n)) \).

Although these recovery guarantee results are promising, when the energy of the measurement noise grows larger, the recovery error of (1.2) may also becomes unexpectedly large, which means that a few grossly corrupted measurements can severely degrade the performance of (1.2). Unfortunately, corrupted noise and irrelevant measurements are prevalent in modern applications such as image processing, sensor networks where certain amount of measurements are grossly corrupted due to the factors like hardware flaws and environmental hazards [30-32].

And this motivated another line of works, which is called compressed sensing with corruption [2, 33]. Where the measurement vector is represented as \( b = A x^{(0)} + f^{(0)} \), here, \( f^{(0)} \) is a \( m \)-dimensional vector representing the corrupted noise, which is often assumed to be sparse and whose non-zero elements can take arbitrary values. Similar to (1.1), the signal \( x^{(0)} \) and corrupted noise \( f^{(0)} \) are recovered through the below \( \ell_1 \) minimization:

\[
\min_x \|x\|_1 + \theta \|f\|_1, \text{ s.t. } Ax + f = b
\] (1.3)

Existing works on compressed sensing with corruption. [34, 35] apply (1.3) when \( \theta = 1 \) and A is Full Fourier basis on the problem of separating the sinus and spikes, the deterministic guarantee for (1.3) based on the coherence of matrices require that the sparsity of both \( x^{(0)} \) and \( f^{(0)} \) are bounded from above by \( O(\sqrt{n}) \) as indicated in [36, 37]. Later, Candes [4] shows that when A is full Fourier basis, \( \theta = 1 \), then exact recovery of \( x^{(0)} \) and \( f^{(0)} \) is possible when the supports of \( x^{(0)} \) and \( f^{(0)} \) are uniformly random and both \( \|x^{(0)}\|_0 \) and \( \|f^{(0)}\|_0 \) is smaller than \( O(n/\sqrt{\ln(n)}) \).

Motivated by the problem of face recognition, Wright et.al [30] shows that recovery is possible even when \( \|f^{(0)}\|_0 \) grows arbitrarily close to \( m \), provided that \( \|x^{(0)}\|_0 \) is sub-linear smaller than \( m \), however, in their “cross and brouquet model”, A is a Gaussian i.i.d designed matrix, which is different from our setting.

More recently, Nguyen [3] and Li [2] shows that when A is a partial BOS (where partial Fourier matrix can be treated as a special case), after choosing \( \theta \) appropriately as depending on \( m \) and \( n \), then recovery of \( x^{(0)} \) and \( f^{(0)} \) is possible even when a constant fraction of \( b \) is corrupted by \( f^{(0)} \), provided that \( m \geq O(\|x^{(0)}\|_0 \ln^2(n)) \).

Recently, [38] studies the probabilistic recovery guarantee of a more general \( \ell_1 \) minimization:
\[
\min_{\mathbf{x}, \mathbf{f}} \|\mathbf{x}\|_1 + \|\mathbf{f}\|_1, \text{ s.t. } \mathbf{A}\mathbf{x} + \mathbf{B}\mathbf{f} = \mathbf{b}
\]  
(1.3)

Where \( \mathbf{b} = \mathbf{A}\mathbf{x}^{(0)} + \mathbf{B}\mathbf{f}^{(0)} \), \( \mathbf{A} \) and \( \mathbf{B} \) are general matrices, based on the coherence of matrices \( \mathbf{A} \) and \( \mathbf{B} \), the authors in [38] show that recovery of \( \mathbf{x}^{(0)} \) and \( \mathbf{f}^{(0)} \) is possible even when the sparsity of \( \mathbf{x}^{(0)} \) and \( \mathbf{f}^{(0)} \) scale linearly to the number of measurement \( \mathbf{m} \), provided that the signs and supports of \( \mathbf{x}^{(0)} \) and \( \mathbf{f}^{(0)} \) satisfy some random assumption.

**Our contributions.** Despite the promising theoretical results, in the proofs of the above mentioned literatures [2-4], all impose some restrictions on \( \mathbf{x}^{(0)} \) and \( \mathbf{f}^{(0)} \), e.g., random assumptions on the supports or signs of \( \mathbf{x}^{(0)} \) and \( \mathbf{f}^{(0)} \), which are unlikely satisfied in real applications. This motivates the research of this paper, for convenience, we reformulate (1.2) as (1.4) below:

\[
\min_{\mathbf{x}, \mathbf{f}} \|\mathbf{x}\|_1 + \|\mathbf{f}\|_1, \text{ s.t. } \lambda\mathbf{A}\mathbf{x} + \mathbf{f} = \mathbf{b}
\]  
(1.4)

Where \( \mathbf{A} \) is a column normalized partial Fourier matrix whose rows are chosen randomly and uniformly from an \( n \times n \) Fourier basis, it is easy to observe that solutions of (1.2) and (1.4) can be the same if we appropriately tune \( \lambda \) in (1.4) depending on \( \theta \) in (1.2).

We show that when \( \lambda = 1 \), (1.4) can recover \( \mathbf{x}^{(0)} \) and \( \mathbf{f}^{(0)} \) when a constant fraction of the measurement vector \( \mathbf{b} \) are arbitrarily corrupted by \( \mathbf{f}^{(0)} \), provided that \( m \geq O(k\ln(k)) \), here \( k = \max\{\|\mathbf{x}^{(0)}\|_0, 1\} \), this lower bound for \( m \) is better than the asymptotically optimal lower bound on \( m \) when the sensing matrix is partial BOS chap. 12 [1]. When \( \lambda = 1/\sqrt{\ln(2n/\varepsilon)} \), we show that successful recovery of (1.4) is also possible even when the number of corrupted observation \( \|\mathbf{f}^{(0)}\|_0 \) becomes arbitrarily close the total number of measurement \( m \), provided that \( m \geq O(k\ln(n)) \), this lower bound of \( m \) is asymptotically better than those achieved by recent literatures [2, 3] by a \( \ln(n) \) factor.

It’s worthy to mention that we impose no assumption of the signs of \( \mathbf{x}^{(0)} \) or \( \mathbf{f}^{(0)} \), which lends our theoretical result to be more applicable in practice. Extensive experiments based on synthesis data as well as real data faithfully validate our analysis result.

**Organization of the paper.** The organization of the remaining of the paper is as follows, section 2 provides our main results—theorem 2.1, section 3 provides the proof roadmap of theorem 2.1, with the supporting lemmas detailed in appendix A.1~A.3 which eventually leads to the final proof of theorem 2.1 as given in A.4. In section 4, we provided the experiments verify the conclusion of theorem 2.1 base on both synthesis data and the real image data obtained from the BSDS 500 database. Finally, section 5 summaries the findings of this paper with conclusions and future works.
2. Main result

2.1 notations and problem statement

Before the problem statement, we briefly introduce some notations that will be used in this section and throughout the rest of the paper.

**Notations.** Let \([n] = \{1, \ldots, n\}\) denotes an indices set if \(n\) represents a positive integer, and \(i + [n] = \{i + 1, \ldots, i + n\}\) denotes an indices set \([n]\) be right shifted by \(i\), if \(i\) denotes an integer. Let \(s_x, s_f\) denotes some indices subset of \([n], [m]\) which contain the support of \(x^{(0)}, f^{(0)}\), respectively. \(|s|\) denotes the cardinality of \(s\), here \(|\cdot|\) denotes the cardinality of a set if \(\cdot\) represents a set, \(|s|\) is interpreted similarly. Let \(\sigma_x, \sigma_f\) denote the sign vector of \(x^{(0)}(s_x), f^{(0)}(s_f)\), respectively. \(Z^+\) denotes the set of all positive integers.

Suppose we have a \(m\)-dimensional measurement vector \(b\) which can be written as:

\[
b = A x^{(0)} + f^{(0)} \tag{2.1.1}
\]

Where \(x^{(0)}, f^{(0)}\) denote the \(n\)-dimensional ground-truth signal and \(m\)-dimensional corrupted noise, respectively. \(\tilde{A}\) is a \(m \times n\) column normalized partial Fourier basis: \(\tilde{A} = \frac{1}{\sqrt{m}} F(\Lambda, [n])\), we use the below definition 2.1 to define \(\Lambda\) -- the row indices set of the sensing matrix \(\tilde{A}\).

**Definition 2.1 (random subset)** An indices set \(S = \{s_1, \ldots, s_m\}\) with cardinality \(m\) is called a random subset of \([n]\), if \(S \subseteq [n]\) and \(S\) is selected randomly and uniformly from all subset of \([n]\) with cardinality \(m\).

In this paper, our goal is to recover \(x^{(0)}, f^{(0)}\) through the below \(\ell_1\) minimization:

\[
\text{minimize}_{x, f} \|x\|_1 + \|f\|_1 \quad \text{s.t.} \quad \lambda \tilde{A} x + f = b \tag{2.1.2}
\]

E.g., let \(\hat{x}, \hat{f}\) denote the solution of (2.1.2), then \(\lambda \hat{x}, \hat{f}\) are estimations of \(x^{(0)}, f^{(0)}\), respectively, here \(\lambda > 0\) serves as a parameter of (2.1.2).

2.2 Recovery guarantee

We prove that (2.1.2) can recover \(x^{(0)}\) and \(f^{(0)}\) exactly with high probability provided that \(\tilde{A}\), \(x^{(0)}\) and \(f^{(0)}\) satisfy certain conditions, which is stated formally in below theorem 2.1.

**Theorem 2.1** if \(n\) is prime, \(\Lambda\) and \(\Lambda (s_f^\epsilon)\) are random subsets of \([n]\), \(|s_f^\epsilon| \geq \frac{32}{3} |s_x| \ln \left(\frac{2|s_x|}{\epsilon}\right)\)

and \(|s_f| \geq \frac{32}{3} |s_x| \ln \left(\frac{2|s_x|}{\epsilon}\right)\), where \(0 < \epsilon < \frac{1}{3}\) is a constant. Furthermore, if below (2.2.1) holds,

\[3\] Here we define the sign of a complex variable \(re^{i\theta}\) as \(e^{i\theta}\), where \(r \geq 0\) denotes the absolute value of this variable, in particular, we define the sign of 0 as 0.
\[
\rho_1 \sqrt{|s_f|} + \rho_2 \sqrt{|s_x|} \leq \frac{1}{2} c \sqrt{|s_f^2| - |s_x|}
\]  \hspace{1cm} (2.2.1)

Where \( \rho_1 = \frac{n}{\sqrt{m}} \lambda \), \( \rho_2 = \frac{m}{\sqrt{|s_f|}} \sqrt{6(\lambda^{-1} + \sqrt{2\ln(2|s_x|/\epsilon)}) + \sqrt{6(1 + \lambda \sqrt{2\ln(2|s_x|/\epsilon)})}} \), \( 0 < c < 1 \) is a constant, then with probability at least \( 1 - 3\epsilon \), the solution of minimization (2.1.2) can recover \( x^{(0)} \) and \( f^{(0)} \) exactly.

Assuming that \( m = C_1 n \), \( |s_f| = C_2 m \), where \( C_1, C_2 \in (0, 1) \). In other words, the total number of measurements \( m \) takes a certain percentage of the dimension \( n \), and the number of corrupted observation occupies a constant fraction of \( m \), then one have the following conclusion according to theorem 2.1:

- When \( \lambda = 1 \), theorem 2.1 shows that exact recovery by the solution of (2.1.2) is possible if \( |s_f| \leq C_3 m / \ln (|s_x|) \), provided that \( C_3 \) and \( C_2 \) are sufficiently small. Roughly speaking, when \( \lambda = 1 \), theorem 2.1 states that if \( |s_f| \leq O(m / \ln (|s_x|)) \), then (2.1.2) can recover \( x^{(0)} \) and \( f^{(0)} \) even when a constant fraction of the measurement vector \( b \) is arbitrarily corrupted, this bound is better than the optimal bound achieved by the BOS sensing matrix chap.12 [1], which is \( O(m / \ln (n)) \). However, condition (2.2.1) in theorem 2.1 prevents \( |s_f| \) from becoming arbitrary close to \( m \).

- When \( \lambda = 1 / \sqrt{\ln (2n / \epsilon)} \), theorem 2.1 shows that (2.1.2) can recover \( x^{(0)} \) and \( f^{(0)} \) if \( |s_f| \leq O(m / \ln (n)) \) be sufficiently small, as it is indicated above, this upper bound for \( |s_f| \) is also the asymptotically optimal bound when sensing matrix is partial BOS. Moreover, it’s worthy to mention that in this case, when \( n \) is sufficiently large, it allows that the number of corrupted observations \( |s_f| \) grow arbitrarily close to the total number of measurement \( m \), because in this case when \( n \to \infty \), on has \( \rho_1 \to 0 \), this guarantees condition (2.2.1) in theorem 2.1 hold.

**Existing works and our contribution**

Candes 2006 [4] shows that when \( A = F \) be the full Fourier basis matrix, then (2.1.2) can recover \( x^{(0)} \) and \( f^{(0)} \) exactly provided that \( \|x^{(0)}\|_0 + \|f^{(0)}\|_0 \leq O\left(\frac{n}{\sqrt{\ln (n)}}\right) \), the upper bound of \( \|x^{(0)}\|_0 \) in [4] can be larger than those proposed in theorem 2.1 in this paper by a factor of \( \sqrt{\ln (n)} \), however, it disallows constant fraction of corruption in the measurement vector \( b \), furthermore, there are restrictions on \( x^{(0)} \) and \( f^{(0)} \) as showed in sect. 2 of [4], e.g., the support of \( x^{(0)} \) and \( f^{(0)} \) are 2 independent random subsets of \([n]\), and the non-zero elements in \( f^{(0)} \) should obey circularly symmetric distribution in the complex plane.

Wright et. al [30] show that (2.1.2) can successfully recover \( x^{(0)} \) and \( f^{(0)} \) even if \( f^{(0)} \) corrupted almost all entries in measurement vector \( b \), provided \( \|x^{(0)}\|_0 \) is sub-linear smaller than \( m \), which is far from the bound \( |s_x| \leq O(m / \ln (n)) \) as we obtained in theorem 2.1,
Furthermore, the matrix $\tilde{A}$ is an Gaussian i.i.d designed matrix, which is different from the partial Fourier basis matrix as we consider in this paper.

Li [2] and Nguyen [3] propose a reweighted $\ell_1$ minimization:

$$\min_{x, f} \|x\|_1 + \theta \|f\|_1 \text{ s.t. } \tilde{A}x + f = b \quad (2.2.2)$$

After appropriately choosing $\theta$ depending on $m$ and $n$, they also show that recovery of $x^{(0)}$ and $f^{(0)}$ by (2.6) is possible even the corruption ratio $\frac{\|f^{(0)}\|_0}{m}$ grows arbitrarily close to 1, provided that $m \geq O(\|x^{(0)}\|_0 \ln^2(n))$, which is asymptotically worse than the bound required in theorem 2.1 by a $\ln(n)$ factor.

Besides, all the literatures mentioned above [2-4] require certain restrictions on the $x^{(0)}$ and $f^{(0)}$ in their proofs, e.g. Wright [30] assumes that the support of $f^{(0)}$ distributes randomly and uniformly, and the signs of non-zero elements in $f^{(0)}$ takes $+1, -1$ uniformly and independently. In Nguyen [3] the author assumes the signs of $x^{(0)}$ on the support of $x^{(0)}$ are independently and uniformly to be $+1$ or $-1$, and the support of $f^{(0)}$ is a random subset of $[m]$. Li [2] assumes that the sign sequences of $x^{(0)}$ and $f^{(0)}$ on their supports are independent of each other, and each is a sequence of symmetric i.i.d $\pm 1$ variables.

In this paper, the support and sign of $x^{(0)}$ as well as the sign of $f^{(0)}$ can be arbitrary, the only restriction on $f^{(0)}$ is that $\Lambda\left(\mathcal{f}_f\right)$ should be a random subset of $[n]$, to this end, one can only assume the support of $f^{(0)}$ be independent of $\Lambda$ (which is the row indices set of $\tilde{A}$ corresponding to the Fourier basis matrix $F$).

An unappealing restriction in theorem 2.1 is that it requires $n$ to be prime, which is unlikely satisfied in real applications. To the best of our knowledge, whether the conclusion of theorem 2.1 holds for more general case where $n \in \mathbb{Z}^+$ remains an open problem. However, according to the promising recovery results presented in section 4 when $n$ is not prime, we believe that restricting $n$ to be a prime is artificial and only for the convenience of the current proof in this paper.

**Stable recovery guarantee**

Finally, although in theorem 2.1, it doesn’t consider explicitly certain bounded, dense measurement noise, say $b = \tilde{A}x^{(0)} + f^{(0)} + \nu$, where $\nu$ is a $m$-dimensional vector denotes some kinds of measurement noise, e.g., $\nu$ can be a vector of Gaussian noise. In fact, using the dual vector $h \in \mathbb{C}^m$ we constructed in the next section, stable guarantee of the below quadratic programming is achievable according to theorem 3.33 of [1]:

$$\min_{x, f} \|x\|_1 + \|f\|_1 \text{ s.t. } \|\lambda \tilde{A}x + f - b\|_2 \leq \eta \quad (2.2.3)$$

Where $\|z\|_2 \leq \eta$, which states that the recovery error of (2.2.3) is proportional to $\eta$ and $\sigma_k(x^{(0)})_1$, where $\sigma_k(\cdot)_1$ is defined below as in [1, 8]:
3. Proof road map of theorem 2.1

In this section, we provide a brief roadmap for the proof of theorem 2.1, where the necessary supporting lemma given in section A.1~A.2 which lead to a final proof of theorem 2.1 as stated in section A.3.

Firstly we introduce some notations that will be used in this section and in the rest of this paper.

**Notations.** In this paper, $a:b$ denotes the set $\{a, a+1, \ldots, b\}$ where $a, b$ are 2 integers such that $a \geq b$. $M^*$ denotes the conjugate of $M$ if $M$ denotes a matrix with complex entries. $I_N$ denotes a $N \times N$ identity matrix, where $N$ is a positive integer.

The starting point of the proof is the so called “dual certification” as presented in the following lemma 3.1:

**Lemma 3.1 (dual certification [39])** $x^{(0)}$ and $f^{(0)}$ is the unique solution of (2.1.2) if and only if the following condition holds:

There exists a vector $h \in C^m$ such that,

$$
\begin{aligned}
\begin{cases}
\lambda \bar{A}^*(s_x, [m]) h = \sigma_x, h(s_f) = \sigma_f \\
\|\lambda \bar{A}^*(s_x, [m]) h\|_\infty < 1, \|h(s_f)\|_\infty < 1
\end{cases}
\end{aligned}
$$

(3.1)

And matrix $B = [\lambda \bar{A}([m], s_x), I_m([m], s_f)]$ is full rank.

Where $\lambda \bar{A}^*$ denotes the conjugate matrix of $\lambda \bar{A}$.

**Proof:** this follows from a direct application of theorem 4 in [39].

According to [30], we have the below lemma 3.2.

**Lemma 3.2 ([30])** suppose we have a $|s_x^1| + |s_f^1|$ dimensional vector $q$ satisfies,

\[
\begin{bmatrix}
\lambda \bar{A}^*(s_x^1, s_f^1) & I_{|s_x^1|} \\
\lambda \bar{A}^*(s_x^1, s_f^1) & 0
\end{bmatrix} q =
\begin{bmatrix}
-\lambda \bar{A}^*(s_x^1, s_f^1) \sigma_f \\
\sigma_x - \lambda \bar{A}^*(s_x^1, s_f^1) \sigma_f
\end{bmatrix}
\]

(3.2)

And $\|q\|_\infty < 1$, where $I_{|s_x^1|}$ denotes a $|s_x^1| \times |s_x^1|$ identity matrix. then there exists a $m$-dimensional vector $h$ as defined in (3.3) satisfies (3.1).

$h(s_i) = \sigma_i$, $h(s_f^1) = q(1:|s_f^1|)$

(3.3)

**Proof:**

According to (3.3) and (3.2), one has,

\[
\begin{aligned}
\begin{cases}
\lambda \bar{A}^*(s_x^1, s_f^1) h(s_f^1) + q(|s_f^1| + 1:|s_f^1| + |s_x^1|) = -\lambda \bar{A}^*(s_x^1, s_f^1) h(s_f^1) \\
\lambda \bar{A}^*(s_x^1, s_f^1) h(s_f^1) = \sigma_x - \lambda \bar{A}^*(s_x^1, s_f^1) h(s_f^1)
\end{cases}
\end{aligned}
\]

Which implies that:

\[
\begin{aligned}
\begin{cases}
q(|s_f^1| + 1:|s_f^1| + |s_x^1|) = -\lambda \bar{A}^*(s_x^1, [m]) h \\
\lambda \bar{A}^*(s_x^1, [m]) h = \sigma_x
\end{cases}
\end{aligned}
\]

(3.4)

Combining (3.4), (3.2) and the fact that $\|q\|_\infty < 1$ proves the conclusion of the lemma.

By lemma 3.1~lemma 3.2, we conclude that if we can prove that matrix $B$ in lemma 3.1 is full rank and then find a $q$ satisfies (3.2) and $\|q\|_\infty < 1$, then we can prove that $x^{(0)}$ and $f^{(0)}$ can be recovered by the solution of (2.1) exactly, lemma 3.3 below shows a sufficient condition for the

$$
\sigma_k(y)_1 = \min_k \|x - y\|_1, \text{s.t.} \|x\|_0 \leq k
$$

(2.2.4)
4. Experiments

similar which are not presented algorithm experiment results when n is not prime in each sections. under data to verify the conclusion of theorem 2.1, although the requirement of n is a prime in theorem I leads lemma 3.1 is full rank is provided in appendix A.3 searching for a viable F B Φ

Let Φ S : system W sparse vectors: projection operator on to L thresholding operator, similarly, if y = S(x) denotes a n-dimensional vector with y(i) = S(x(i)), 1 ≤ i ≤ n.

Lemma 3.3 (lemma 2 in [30]) Consider an underdetermined system Φq = w. let PΦ denotes the projection operator on to range(Φ∗), and let ξk(Φ) denotes the norm of PΦ, restricted to sparse vectors:

\[ ξ_k(Φ) := \sup_{\|x\|_0 ≤ k, \|x\|_1 ≤ 1} \|P_Φ x\|_0 \]  \hspace{1cm} (3.5)

Suppose ξk(Φ) < 1 and there exists a solution q₀ satisfies,

\[ Φq_0 = w, \|q_0\|_2 + \frac{ξ_k(Φ)}{1 - ξ_k(Φ)} \|S[q_0]\|_2 ≤ \sqrt{k}/2 \]  \hspace{1cm} (3.6)

Where S is the soft-thresholding operator with threshold 1/2, then there exists a solution q to system Φq = w satisfies \|q\|_∞ < 1/2.

Since \|S[q_0]\|_2 ≤ \|q_0\|_2, a sufficient condition of q₀ satisfies (3.6) is q₀ satisfies below (3.7):

\[ Φq_0 = w, \left(1 + \frac{ξ_k(Φ)}{1 - ξ_k(Φ)}\right) \|q_0\|_2 ≤ \sqrt{k}/2 \]  \hspace{1cm} (3.7)

Let

\[ Φ = \begin{bmatrix} \lambda \bar{A}^*(s_x^c, s_f^c) \mathbb{I}_{\|\bar{X}\|_1} & \text{0} \end{bmatrix}, w = \begin{bmatrix} -\lambda \bar{A}^*(s_x^c, s_f^c) \sigma_f \\ \sigma_x - \lambda \bar{A}^*(s_x^c, s_f^c) \sigma_f \end{bmatrix} \]  \hspace{1cm} (3.8)

By the discussion above, the recovery guarantee of (2.1.2) is achieved in 3 steps:

Firstly, the proof of ξk(Φ) < 1 when k = \sqrt{|s| - |s_x|} is provided in appendix A.1, and the searching for a viable q₀ satisfying (3.7) are detailed in appendix A.2. Secondly, the proof of B in lemma 3.1 is full rank is provided in appendix A.3. Finally combining results in appendix A.1~A.3 leads to a natural proof of theorem 2.1, which is provided in appendix A.4.

4. Experiments

In this section, we provide numerical experiments based on synthesis data as well as real image data to verify the conclusion of theorem 2.1, although the requirement of n is a prime in theorem 2.1 is necessary in our proof, but it is not necessary satisfied in real applications, to understanding the performance of (2.1.2) when this assumption is violated, we also show the experiment results when n is not prime in each sections. For simplicity, we set \( \lambda = 1 \) in algorithm (2.1.2) within all our experiments, when \( \lambda = 1/\sqrt{\ln(n/ε)} \), the recovery results are similar which are not presented in this paper for saving space.
4.1. Experiment based on synthesis data

In this section, we provide extensive simulations to illustrate the probability of the $\ell_1$-minimization (2.1.2) in recovering data $x^{(0)}$ and $f^{(0)}$.

The setting for $x^{(0)}$ and $f^{(0)}$ are according to theorem 2.1 as described below: $\|x^{(0)}\|_0 = 0.2n/\ln(0.2n)$, which is asymptotically the same as suggested in theorem 2.1, $\Lambda = \frac{1}{\sqrt{m}}F(\Lambda, [n])$, where $\Lambda$ is selected randomly and uniformly in $[n]$, to ensure $\Lambda(s_f)$ being a random subset of $[n]$, $s_f$ is chosen independently of $\Lambda$.

Furthermore, to violate the random distribution assumptions on the supports and signs of $x^{(0)}$ and $f^{(0)}$ which are necessary in existing literatures [2-4] etc, but not necessary in theorem 2.1 in this paper, we set $s_x = \{j, j + 1, \ldots, j + |s_x|\}$, where $j$ is selected randomly and uniformly in interval $[1, n - |s_x|]$, and without lose of generality, all entries in $x^{(0)}(s_x)$ and $f^{(0)}(s_f)$ are set to be positive. The norm of corrupted noise is set significantly larger than those of $x^{(0)}$, i.e, $\|f^{(0)}\|_2 \geq 100\|x^{(0)}\|_2$.

Random vectors $x^{(0)}$ and $f^{(0)}$ as described above are generated of varying parameters $\vartheta_m$, $\vartheta_f$ and $n$ to test the recoverability of algorithm (2.1.2). Where $\vartheta_m \coloneqq \frac{|s_x|}{n}$ denotes the measurement rate, $\vartheta_f \coloneqq \frac{|s_f|}{m}$ denotes the corruption rate, and $n$ denotes the dimension of the signal $x^{(0)}$.

The algorithm (2.1.2) is declared to succeed in each run if the relative recovery error (RRE) defined below is less than $10^{-8}$.

$$RRE = \frac{\|x^{(0)} - f^{(0)}\|_2}{\|x^{(0)}\|_2}$$

More specifically, different parameters $\vartheta_m$, $\vartheta_f$ and $n$ are set within a range as described below: $\vartheta_m \in \{0.1, 0.2, \ldots, 1\}$ takes 10 different values, $\vartheta_f \in \{0.05, 0.15, 0.25, 0.35\}$ takes 4 different values, $n$ takes different possible values within 2 different sets: in the first scenario, $n$ takes 20 prime values within the interval $[128, 512]$ while in the second scenario, $n$ takes 20 non-prime integers which are chosen randomly and uniformly from interval $[128, 256]$.

The succeed rate of recovery of algorithm (2.1.2) in these 2 different scenarios are summarized in the 2 heat-maps of figure (4.1) respectively. In figure (4.1), the vertical axis in the heat-map indicates different values of $\vartheta_m$(the measurement rate), while the horizontal axis indicates different values of $\vartheta_f$ (the corruption rate).
For each fixed parameters pair \((\theta_m, \theta_f, n)\), 25 independent runs of minimization (2.1.2) are performed in order to calculate the succeed rate, thus, each element in the heat maps of figure 4.1 is achieved through \(20 \times 25 = 500\) independent runs of (2.1.2). e.g., the element located on the first row, first column of the left heat map in figure 4.1 indicates the average succeed rate \((\text{the probability of RRE} < 10^{-8} \text{ occurs})\) of 20 different parameter pairs \(\{(\theta_m, \theta_f, n)|\theta_m = 0.1, \theta_f = 0.05, n \text{ is prime belong to } [128, 512]\}\).

For each fixed parameter pair \((\theta_m, \theta_f, n)\) we perform 25 independent runs and hence we totally performed \(20 \times 25 = 500\) independent runs before we obtaining the average succeed rate value in this particular location of the heat map. Elements in the right heat-map can be interpreted similarly.

Figure (4.1) the recovery performance of (2.1.2) based on synthesis data. The left heat-map indicates the average recovery rate with different measurement rate \(\theta_m\) (indicated on the vertical axis) and different corruption rate \(\theta_f\) (indicated on the horizontal axis) when the signal dimension \(n\) is prime value in interval \([128, 512]\). The right heat-map is interpreted similarly when \(n\) is not prime integer in interval \([128, 512]\).

As showed in figure (4.1), when the measurement rate \(\theta_m \geq 0.9\) and the corruption rate \(\theta_f \leq 0.15\), then the successful recovery rate of (2.1.2) is larger than 81% in most case, which validates the conclusion of theorem 2.1, also note that the recovery performance of (2.1.2) is not severely degraded when the assumption of \(n\) is prime is violated, as we can see in figure (4.1), there are only slightly difference between the left and the right heat map which is caused by the difference between the distribution of primes and non-prime integers in the interval \([128, 512]\).

4.2 Experiments based on the real data

In this section, the signal to be recovered \(x^{(0)}\) is chosen from the real world data, more specifically, data in \(x^{(0)}\) is from an image patch selected randomly from the natural images in the BSDS-500 database, then elements of the image patch are stacked together to form \(x^{(0)}\). 10 typical image patches selected in this manner are showed in figure (4.2.1). Without loss of
generality, the size of the image path is set different values as $8 \times 8$, $16 \times 16$ and $32 \times 32$, the corresponding dimension of $x^{(0)}$ are set differently as 64, 256 and 1024.

To measure the recovery performance of minimization (2.1.2) in this section, instead of recover $x^{(0)}$, we recover the Fourier transform of $x^{(0)}$(which is denoted as $\hat{x}^{(0)}$) based on partial measurement of $x^{(0)}$, let $b = x^{(0)}(\Lambda) + f^{(0)}$, where $\Lambda$ is a random subset of $[n]$, to ensure $\Lambda (s^\top_i)$ also be a random subset of $[n]$, we set $s_i$ be independent of $\Lambda$, then the minimization (2.1.2) to recover $\hat{x}^{(0)}$ is wrote as:

$$\text{minimize} \|x\|_1 + \|f\|_1 \quad \text{s.t.} \quad A x + f = \tilde{A} \hat{x}^{(0)} + f^{(0)} \quad (4.2.1)$$

Where $\tilde{A} = \frac{1}{\sqrt{m}} F^\dagger (\Lambda_m , [n])$ denotes the partial matrix of the conjugate Fourier basis with normalized columns, it is easily to verify that theorem 2.1 also applies to (4.2.1) since lemma A.1.1 applies when the measurement matrix is $F^\dagger$.

The norm of corrupted noise is set significantly larger than $\|x^{(0)}\|_2$ as in the previous section, i.e,

$$\|f^{(0)}\|_2 \geq 100 \|x^{(0)}\|_2,$$

finally, the recovery performance is measured by the signal relative recovery error (SRRE):

$$\text{SRRE} = \frac{\|x - \hat{x}^{(0)}\|_2}{\|x^{(0)}\|_2} \quad (4.2.2)$$

Figure (4.2.1) displays the process of the recovery of a typical image patch mentioned above.

$\Rightarrow$ Partial data missing

$\Leftrightarrow$ Recovered via l1 minimization

$\Leftrightarrow$ Add corrupted noise

Figure (4.2.1) demonstration of the data recovery based on the real image data, the upper left image shows a typical $32 \times 32$ image patch randomly selected from BSDS 500 data base, the
upper-right image shows the original image patch with 20% missing data (showed as the dark pixels), the lower-right image shows the upper-right image with 25% measurement are corrupted (showed as the bright pixels), finally the lower-left image shows the recovered image via the $\ell_1$ minimization (2.1.2) using the lower-right noisy image as input, in this example, the recovery relative error is 20.04%.

Unlike in the previous section where the signal to be recovered is designed to be sparse, $x^{(0)}$ here can only be treated as approximately sparse, for convenience, we let $\frac{\sigma_k(x^{(0)})}{\|x^{(0)}\|_2}$ be a k-sparse indicator of a non-zero vector $x^{(0)}$ which reflects how good $x^{(0)}$ can be approximated by a k-sparse vector, where the meaning of $\sigma_k(x^{(0)})_1$ is defined by (2.2.4).

It is easily to see that k-sparse indicator of a vector equals to 0 if and only if the cardinality of the vector is less than or equal to k. Moreover, the k-sparse indicator $\frac{\sigma_k(x^{(0)})}{\|x^{(0)}\|_2}$ also reflects a theoretical upper-bound of the relative recovery error on the convex programming (2.2.3). Figure (4.2.2) shows the k-sparse indicator curves of 3 different types of signal: 1. The signal is a Gaussian random vector; 2. The signal is a sparse random vector as described in the previous section; 3. The signal is $x^{(0)}$ as described above in this section.

Figure (4.2.2) the average value of $\frac{\sigma_k(y)_1}{\|y\|_2}$ when y is the signal selected in 3 different cases. The horizontal axis indicates $\frac{k}{n}$ where n denotes the dimension of y, and $1 \leq k \leq n$ be positive integers, the vertical axis denotes the average value of $\frac{\sigma_k(y)_1}{\|y\|_2}$, i.e., the black dashed curve, the blue dashed curve and the red solid curve indicate the average value of $\frac{\sigma_k(y)_1}{\|y\|_2}$ of 200 random signals, 200 synthesis signal $x^{(0)}$ as defined in section 4.1 and 200 Fourier transform of 32 × 32 image patch signal defined in this section, respectively.
In figure (4.2.2), we observe that the k-sparse indicator curve of $\hat{x}^{(0)}$ defined in this section (the dashed blue curve) decreases more rapidly than the curve of the random signal (the dashed dark curve) which suggests that $\hat{x}^{(0)}$ we used in this section can be treated approximately a k-sparse vector if $k$ is sufficiently large. Also notice that we have $\sigma_k(x^{(0)})/\|x^{(0)}\|_2 > 0$ whenever $k < n$, this suggests that $\hat{x}^{(0)}$ can be represented as a k-sparse vector plus a bounded non-zero vector.

The simulation results of this section is summarized in the heat-maps of figure (4.2.3), where each element in a heat-map indicates the average SRRE value (obtained by averaging the SRRE of 200 image patch selected randomly from BSDS-500) for a particular $\theta_m$, $\theta_e$ pair value.

Figure (4.2.3) performance of (2.1.2) based on the real world data. Each element in the heat-maps indicates the average SRRE value obtained by 200 independent runs based on the real world signal $x^{(0)}$, with a particular parameters pair $(n, \theta_m, \theta_e)$. The vertical axis of the heat-maps indicates different value of $\theta_m$ (the measurement rate), the horizontal axis of the heat-maps indicates different value of $\theta_e$ (the corruption rate), and the title of the heat-maps indicates the dimensions of the input signal $x^{(0)}$. E.g., the element on the first row, first column of the left-most heat-map indicates the average SRRE based on 200 $8 \times 8$ different image patches randomly selected from BSDS 500, with $\theta_m = 0.1$ and $\theta_e = 0.05$

It can be seen from figure (4.2.3) that minimization (2.1.2) works well on recovering real image data when the measurement ratio $\theta_m$ is large enough and the corruption ratio $\theta_e$ is small enough (e.g., the average relatively recovery error of $x^{(0)}$ is less than 12% when $\theta_m \geq 0.9$ and $\theta_e \leq 15\%$), and the SRRE value displayed in figure (4.2.2) gradually decrease as $\theta_m$ decreases and $\theta_e$ increases, this verify the conclusion of theorem 2.1 even when $n$ (the length of signal) is not prime.

Notice that the recovery result in figure (4.2.2) is not as good as those in figure (4.1.1), where in figure (4.1.1) one obtain accurate recovery results with certain probability, but in figure (4.2.2), the SRRE value is always larger than a positive constant, this can be explained by the fact that
\( \tilde{x}^{(0)} \) is only approximately sparse as suggested in the k-sparse indicator curve in figure (4.2.2).

5. Conclusion and future works

In summary, we prove that exact recovery of signal is possible from linear measurements \( b = Ax^{(0)} + f^{(0)} \) even when a constant fraction of measurement \( b \) is arbitrarily corrupted by \( f^{(0)} \), unlike in the existing literatures [2, 3, 30], the assumptions on the signs and supports of the \( f^{(0)} \) and \( x^{(0)} \) are kept to be minimal: we only require the support of \( f^{(0)} \) be chosen such that \( \Lambda \left( s_f \right) \) is a random subset of \([m]\). Extensive experiments based on the synthesis data as well as the real world data faithfully verify the conclusion of theorem 2.1 even when \( n \) is not prime, furthermore, we’ve also observed similar experimental results when the sensing matrix \( A \) is chosen as partial Bounded Orthogonal System [1] other than partial Fourier basis. These observations drive us to believe that similar conclusion as in theorem 2.1 might hold for general integer \( n \) and when \( A \) is belong to a broader family—the partial BOS, and the generalization of theorem 2.1 in such kinds are left in future works.

Acknowledgement

The author wish to thank his wife Haixia Yu for her mental as well as material supports over the past 8 years.

Appendix A—supporting lemmas for theorem 2.1

For convenience, firstly we introduce 2 definitions which would be used later in this section.

**Definition A.1 (Bounded Orthonormal System chap.12 [1])** We call \( \Psi' = \{ \psi_1, ..., \psi_N \} \) a bounded orthonormal system (BOS) with constant \( K \) if it satisfies (A.1):

\[
\begin{align*}
\psi_j^* \psi_k &= \begin{cases} 
0, & \text{if } j \neq k \\
1, & \text{if } j = k \quad \forall \ j, k \in [N] 
\end{cases} \\
\|\psi_j\|_\infty &\leq K 
\end{align*} 
\] 

(A.1)

**Definition A.2 (Steinhaus sequence, [1])** A complex random variable which is uniformly distributed on the torus \( T = \{ z \in \mathbb{C}, |z| = 1 \} \) is called a Steinhaus variable. A sequence \( \epsilon = (\epsilon_1, ..., \epsilon_N) \) of independent Steinhaus variables is called a Steinhaus sequence.

A.1 Proving \( \xi_k(\Phi) < 1 \)

**Lemma A.1.1 ([40])** let \( f \) be a \( n \)-dimensional non-zero complex vector, \( \tilde{f} \) be the Fourier transform of \( f \), suppose \( n \) is prime, then,

\[
\|f\|_0 + \|\tilde{f}\|_0 \geq n + 1 
\] 

(A.1.1)

**Theorem A.1.1** if \( k \leq |s_f^*| - |s_x| \) and \( n \) is prime then, \( \xi_k(\Phi) < 1 \) holds.

Proof:
Event $\xi_k(\Phi) < 1$ is equivalent to the event that a $k$-sparse $(|s^c_{1}| + |s^c_{2}|)$-dimensional vector $y$ cannot be represented by the column space of $\Phi^{*} = \begin{bmatrix} \lambda \tilde{A}(s^c_{1}, s^c_{2}) & \lambda \tilde{A}(s^c_{1}, s^c_{2}) \\ \lambda \tilde{A}(s^c_{1}, s^c_{2}) & \lambda \tilde{A}(s^c_{1}, s^c_{2}) \end{bmatrix}$, suppose the inverse is true, i.e., let $z$ be a $n$-dimensional non-zero vector satisfies,

$$\begin{bmatrix} \lambda \tilde{A}(s^c_{1}, s^c_{2}) & \lambda \tilde{A}(s^c_{1}, s^c_{2}) \\ \lambda \tilde{A}(s^c_{1}, s^c_{2}) & \lambda \tilde{A}(s^c_{1}, s^c_{2}) \end{bmatrix} z = y,$$

such that $\|y\|_0 \leq k$ \hspace{1cm} (A.1.2)

Let $\tilde{z}$ denotes the Fourier transform of $z$, since $\lambda$ is a positive constant, (A.1.2) implies that,

$$\|y\|_0 = ||\tilde{z}(\Lambda (s^c_{1}))||_0 + ||\tilde{z}(\Lambda (s^c_{2}))||_0 + ||\tilde{z}(\Lambda (s^c_{2}))||_0 + ||\tilde{z}(\Lambda (s^c_{2}))||_0 + ||\tilde{z}(\Lambda (s^c_{2}))||_0 - ||\tilde{z}(\Lambda (s^c_{1}))||_0 - ||\tilde{z}(\Lambda (s^c_{2}))||_0 - ||\tilde{z}(\Lambda (s^c_{2}))||_0 - $$

$$\|\tilde{z}(\Lambda (s^c_{1}))\|_0 - |s^c_{1}| > |s^c_{1}| - |s^c_{1}|$$ \hspace{1cm} (A.1.3)

Which leads to a contradiction since $\|y\|_0 \leq k \leq |s^c_{1}| - |s^c_{1}|$, where the last inequality of (A.1.3) we use the fact $||\tilde{z}(\Lambda (s^c_{1}))||_0 - ||\tilde{z}(\Lambda (s^c_{2}))||_0 - ||\tilde{z}(\Lambda (s^c_{2}))||_0 - $.

\section*{A.2 Finding a viable $q_0$ satisfies (3.7)}

To find a viable $q_0$ satisfies (3.7), firstly we find a $h_0$ as below:

$$\begin{cases} h_0(s^c_{1}) = \sigma_f \\ h_0(s^c_{2}) = \lambda^{-1} \tilde{A}^t(s^c_{1}, s^c_{2}) \sigma_f \end{cases} \hspace{1cm} (A.2.1)$$

Where $\tilde{A}^t(s^c_{1}, s^c_{2}) = \tilde{A}(s^c_{1}, s^c_{2})(\tilde{A}^t(s^c_{1}, s^c_{2}) \tilde{A}(s^c_{1}, s^c_{2}))^{-1}$ represents the Peron-Moore inverse of $\tilde{A}(s^c_{1}, s^c_{2})$, $\sigma_f = \sigma_x - \lambda \tilde{A}(s^c_{1}, s^c_{2}) \sigma_f$. Then we obtain $q_0$ as in (A.2.2), which is stated below,

$$\begin{cases} q_0(1; |s^c_{1}|) = h_0(s^c_{1}) \\ -q_0(1; |s^c_{1}| + 1; |s^c_{1}| + 1; |s^c_{1}|) = \lambda \tilde{A}(s^c_{1}, [m]) h_0 \end{cases} \hspace{1cm} (A.2.2)$$

It is easy to see that $q_0$ as defined in (A.2.2) satisfies $\Phi q_0 = w$, where $\Phi, w$ are defined as in (3.8). The remaining part is to bound $\|q_0\|_2$

We’ll firstly bound $\|h_0\|_2$, and then bound $\|q_0\|_2$ according to (A.2.2).

\textbf{Lemma A.2.1 (Theorem 12.12 of [1])} Let $A \in C^{m \times n}$ be random sampling matrix associated to a Bounded Orthogonal System (BOS) with constant $K \geq 1$. Let $s \subset [n]$ be an indices set. Then, for $\delta \in (0, 1)$, the normalized matrix $\tilde{A} = \frac{1}{\sqrt{m}} A$ satisfies $\|\tilde{A}([m], s) \tilde{A}([m], s) - I_{|s|}\|_2 \leq \delta$, with probability at least $1 - 2|s| \exp \left(-\frac{3m \delta^2}{8K^2 |s|}\right)$.

\textbf{Lemma A.2.2} if $\Lambda$ and $\Lambda (s^c_{1})$ are random subsets of $[n]$, then,

$$\text{prob} \left( \|\tilde{A}^t(s^c_{1}, s^c_{2})\|_2 \leq \sqrt{\frac{|s^c_{1}|}{m}} \sqrt{\frac{3}{2}} \sqrt{\frac{m}{|s^c_{1}|}} \right) \geq 1 - 2 \delta \hspace{1cm} (A.2.3)$$

holds for any $\epsilon \in (0, \frac{1}{2})$, provided that, $|s^c_{1}| \geq \frac{32}{3} |s_x| \ln \left(\frac{24s_x}{\epsilon}\right)$ and $|s^c_{1}| \geq \frac{32}{3} |s_x| \ln \left(\frac{24s_x}{\epsilon}\right)$.

\textbf{Proof}:

Substituting the constant $K=1$ and letting $\delta = 1/2$ into lemma A.2.1, one has,

$$\text{prob} \left( \|\tilde{A}^t(s^c_{1}, s^c_{2}) \tilde{A}(s^c_{1}, s^c_{2}) - I_{|s^c_{1}|}\|_2 \leq \frac{1}{2} \right) \geq 1 - \epsilon \hspace{1cm} (A.2.4)$$

provided that,

$$|s^c_{1}| \geq \frac{32}{3} |s_x| \ln \left(\frac{24s_x}{\epsilon}\right) \hspace{1cm} (A.2.5)$$
According to (A.2.1), the conclusion of the lemma follows immediately from (A.2.9) and letting $\gamma = \frac{1}{2}$, we have, \[\left|\sqrt{2|s_x|\ln(2|s_x|/\varepsilon)}\right| \geq 1 - 2\exp\left(-\ln(2|s_x|/\varepsilon)\right) = 1 - \frac{1}{|s_x|}, \quad 1 \leq j \leq |s_x| \] (A.2.10) then applying a union bound on (A.2.10) and using the fact $\frac{|s_x|}{\sqrt{m}} \leq 1$ prove (A.2.9). ■

**Lemma A.2.5** if conditions in lemma A.2.4 holds, then, \[\text{prob}(\|\sigma_x^c\|_2 \leq 1 + \lambda \sqrt{2\ln(2|s_x|/\varepsilon)}) \geq 1 - \varepsilon \] (A.2.11) for any $0 < \varepsilon < 1$.

Proof:

Since $\sigma_x = \sigma_x - \lambda\tilde{A}^*(s_x, s_t)\sigma_t$, by triangle inequality: $\|\sigma_x^c\|_2 \leq \|\sigma_x\|_2 + \lambda\|\tilde{A}^*(s_x, s_t)\sigma_t\|_2$, then the conclusion of the lemma follows immediately from (A.2.9). ■

**Lemma A.2.6** (bounding $\|h_0(s_f^c)\|_2$) If conditions in lemma A.2.2 hold, $h_0$ is defined in (A.2.1), then, \[\text{prob}\left(\|h_0(s_f^c)\|_2 \leq \frac{m}{\sqrt{|s_f|}} \sqrt{6\left(\lambda^{-1} + \sqrt{2\ln(2|s_x|/\varepsilon)}\right)} \right) \geq 1 - 3\varepsilon \] (A.2.12)

Proof:

According to (A.2.1), $h_0(s_f^c) = \lambda^{-1}\tilde{A}^*(s_f^c, s_x)\sigma_x^c$, (A.2.12) follows from combining (A.2.3) and (A.2.11) and then applying a union bound. ■
**Lemma A.2.7** (bounding $\|h_0\|_2$) if conditions in lemma A.2.2 holds, $h_0$ is defined in (A.2.1), one has,

$$\text{prob}\left(\|h_0\|_2 \leq \sqrt{|s_f|} + \frac{m}{\sqrt{|s_f|}} \sqrt{6} (\lambda^{-1} + \sqrt{2\ln(2|s_x|/\epsilon)}) \sqrt{|s_x|}\right) \geq 1 - 3\epsilon \quad \text{(A.2.13)}$$

Proof:

By the triangle inequality, one has,

$$\|h_0\|_2 \leq \|h_0(s_f)\|_2 + \|h_0(s_f^\parallel)\|_2 \quad \text{(A.2.14)}$$

The conclusion of the lemma follows from (A.2.1) and Lemma A.2.6.

Now we are in the position of bounding $\|q_0\|_2$, as stated in the below lemma A.2.8.

**Lemma A.2.8** (bounding $\|q_0\|_2$) If conditions in lemma A.2.2 hold, $h_0$ and $q_0$ are defined in (A.2.1) and (A.2.2), respectively, then,

$$\text{prob}\left(\|q_0\|_2 \leq \rho_1 \sqrt{|s_f|} + \rho_2 \sqrt{|s_x|}\right) \geq 1 - 3\epsilon \quad \text{(A.2.15)}$$

where $\rho_1 = \frac{n}{m} \lambda$, $\rho_2 = \frac{m}{\sqrt{|s_f|}} \sqrt{6} (\lambda^{-1} + \sqrt{2\ln(2|s_x|/\epsilon)}) + \sqrt{6} (1 + \lambda \sqrt{2\ln(2|s_x|/\epsilon)}) \sqrt{\frac{n}{m}}$

Proof:

By triangle inequality,

$$\|q_0\|_2 \leq \|q_0(1:|s_f|)\|_2 + \|q_0(|s_f^\parallel| + 1:|s_f^\parallel| + |s_f^\parallel|)\|_2 \quad \text{(A.2.16)}$$

According to (A.2.2), $q_0(1:|s_f|) = h_0(s_f^\parallel)$ and by lemma A.2.6, one has,

$$\text{prob}\left(\|q_0(1:|s_f|)\|_2 \leq \frac{m}{\sqrt{|s_f|}} \sqrt{6} (\lambda^{-1} + \sqrt{2\ln(2|s_x|/\epsilon)}) \sqrt{|s_x|}\right) \geq 1 - 3\epsilon \quad \text{(A.2.17)}$$

since $q_0(|s_f^\parallel| + 1:|s_f^\parallel| + |s_f^\parallel|) = -\lambda \hat{A}(s_f^\parallel, [m]) h_0$ according to (A.2.2), using the fact that

$$\|\lambda \hat{A}(s_f^\parallel, [m]) h_0\|_2 \leq \lambda \sqrt{\frac{n}{m}} \|h_0\|_2 \quad \text{(A.2.18)}$$

By lemma A.2.7, one has,

$$\text{prob}\left(\|q_0(|s_f^\parallel| + 1:|s_f^\parallel| + |s_f^\parallel|)\|_2 \leq \frac{m}{\sqrt{|s_f|}} \lambda \sqrt{|s_f|} + \frac{m}{\sqrt{|s_f|}} \sqrt{6} (1 + \lambda \sqrt{2\ln(2|s_x|/\epsilon)}) \sqrt{|s_x|}\right) \geq 1 - 3\epsilon \quad \text{(A.2.19)}$$

combining (A.2.17) and (A.2.19) prove the conclusion of the lemma.

**Theorem A.2.1** If conditions in lemma A.2.2 hold, let $q_0$ be defined in (A.2.2) which satisfies $\Phi q_0 = w$, if

$$\rho_1 \sqrt{|s_f|} + \rho_2 \sqrt{|s_x|} \leq \frac{1}{2} (1 - \xi_k(\Phi)) \sqrt{k} \quad \text{(A.2.20)}$$

where $k = |s_f^\parallel| - |s_x|$, $\rho_1$, $\rho_2$ are defined in lemma A.2.8. Then one has,

$$\text{prob}\left(\|q_0\|_2 + \frac{\xi_k(\Phi)}{1 - \xi_k(\Phi)} \|S[q_0]\|_2 \leq \sqrt{k}/2\right) \geq 1 - 3\epsilon \quad \text{(A.2.21)}$$

where $0 < \epsilon < \frac{1}{3}$ is a constant.

Proof:

According to theorem A.1.1, one has $0 < \xi_k(\Phi) < 1$, since $\|S[q_0]\|_2 \leq \|q_0\|_2$, to show

$$\|q_0\|_2 + \frac{\xi_k(\Phi)}{1 - \xi_k(\Phi)} \|S[q_0]\|_2 \leq \sqrt{k}/2,$$ it is sufficient to show $\|q_0\|_2 \leq \frac{1 - \xi_k(\Phi)}{2} \sqrt{k}$. (A.2.21) follows immediately from (A.2.20) and (A.2.15) in lemma A.2.8.
A.3 proving that B in lemma 3.1 is full rank

**Theorem A.3.1** if ∧ and ∧ (s_f) are random subsets of [n], |s_f| ≥ \(\frac{32}{3} |s_x| \ln\left(\frac{2|s_x|}{\varepsilon}\right)\), then with probability at least \(1 - \varepsilon\), matrix \(B = [\lambda \tilde{A}([m], s_x), I_m([m], s_f)]\) is full rank.

Proof: to show B is full rank, it's sufficient to show that matrix \(C = B^*B\) is non-singular, in other words, the minimum eigenvalue of C is positive. Since,

\[
C = \begin{bmatrix}
\lambda^2 \tilde{A}^*(s_x, [m]) \tilde{A}([m], s_x) & \lambda \tilde{A}^*(s_x, s_f) \\
\lambda \tilde{A}(s_f, s_x) & I_{|s_f|}
\end{bmatrix}
\]  \hspace{1cm} (A.3.1)

By Schur complement decomposition, one has,

\[
HCHT = \begin{bmatrix}
\lambda^2 \tilde{A}^*(s_x, s_f) \tilde{A}(s_f, s_x) & 0 \\
0 & I_{|s_f|}
\end{bmatrix}
\]  \hspace{1cm} (A.3.2)

Where \(H = \begin{bmatrix}
I_{|s_x|} & -\lambda \tilde{A}^*(s_x, s_f) \\
0 & I_{|s_f|}
\end{bmatrix}\) is a non-singular matrix, by lemma A.2.1, one has

\[
\text{prob}(\|\lambda^2 \tilde{A}^*(s_x, s_f) \tilde{A}(s_f, s_x)\|_{2-2} \geq \frac{\lambda^2}{2}) \geq 1 - \varepsilon \hspace{1cm} (A.3.3)
\]

Holds if \(|s_f| \geq \frac{32}{3} |s_x| \ln\left(\frac{2|s_x|}{\varepsilon}\right)\).

The conclusion of this theorem is proved by combining (A.3.2) and (A.3.3). \(\blacksquare\)

A.4 Proof of the theorem 2.1

Proof of theorem 2.1:

Firstly, matrix \(B = [\lambda \tilde{A}([m], s_x), I_m([m], s_f)]\) is proved to be full rank with probability at least \(1 - \varepsilon\) in Theorem A.3.1.

Furthermore, defining \(q_0\) as in (A.2.2), let \(k = |s_f| - |s_x|\) and \(c = 1 - \xi_k(\Phi)\) then according to Theorem A.2.1, we conclude that \(q_0\) satisfies (3.8) with probability at least \(1 - 3\varepsilon\), which proves the conclusion of the theorem according to Lemma 3.1~lemma 3.3. \(\blacksquare\)

Reference

1. Foucart, S. and H. Rauhut, *A mathematical introduction to compressive sensing*. 2013: Springer.
2. Li, X., *Compressed sensing and matrix completion with constant proportion of corruptions*. Constructive Approximation, 2013. 37(1): p. 73-99.
3. Nguyen, N.H. and T.D. Tran, *Exact recoverability from dense corrupted observations via-minimization*. Information Theory, IEEE Transactions on, 2013. 59(4): p. 2017-2035.
4. Candes, E.J. and J. Romberg, *Quantitative robust uncertainty principles and optimally sparse decompositions*. Foundations of Computational Mathematics, 2006. 6(2): p. 227-254.
5. Donoho, D.L., *Compressed sensing*. Information Theory, IEEE Transactions on, 2006. 52(4): p. 1289-1306.
6. Candès, E.J., J. Romberg, and T. Tao, *Robust uncertainty principles: Exact signal reconstruction
from highly incomplete frequency information. Information Theory, IEEE Transactions on, 2006. 52(2): p. 489-509.

7. Rauhut, H., Random sampling of sparse trigonometric polynomials. Applied and Computational Harmonic Analysis, 2007. 22(1): p. 16-42.

8. Eldar, Y.C. and G. Kutyniok, Compressed sensing: theory and applications. 2012: Cambridge University Press.

9. Sparse, M., The application of compressed sensing for rapid MR imaging Lustig Michael; Donoho David; Pauly John M. Magnetic resonance in medicine: official journal of the Society of Magnetic Resonance in Medicine/Society of Magnetic Resonance in Medicine, 2007. 58(6): p. 1182-95.

10. Haldi, J.P., D. Hernando, and Z.-P. Liang, Compressed-sensing MRI with random encoding. Medical Imaging, IEEE Transactions on, 2011. 30(4): p. 893-903.

11. Murphy, M., et al., Fast-SPIRiT compressed sensing parallel imaging MRI: scalable parallel implementation and clinically feasible runtime. Medical Imaging, IEEE Transactions on, 2012. 31(6): p. 1250-1262.

12. Vasanawala, S.S., et al., Improved Pediatric MR Imaging with Compressed Sensing 1. Radiology, 2010. 256(2): p. 607-616.

13. Herman, M. and T. Strohmer, High-resolution radar via compressed sensing. Signal Processing, IEEE Transactions on, 2009. 57(6): p. 2275-2284.

14. Candès, E.J., J.K. Romberg, and T. Tao, Stable signal recovery from incomplete and inaccurate measurements. Communications on pure and applied mathematics, 2006. 59(8): p. 1207-1223.

15. Candès, E.J. and T. Tao, Decoding by linear programming. Information Theory, IEEE Transactions on, 2005. 51(12): p. 4203-4215.

16. Candès, E.J., The restricted isometry property and its implications for compressed sensing. Comptes Rendus Mathematique, 2008. 346(9): p. 589-592.

17. Rudelson, M. and R. Vershynin, On sparse reconstruction from Fourier and Gaussian measurements. Communications on Pure and Applied Mathematics, 2008. 61(8): p. 1025-1045.

18. Ben-Haim, Z., Y.C. Eldar, and M. Elad, Coherence-based performance guarantees for estimating a sparse vector under random noise. Signal Processing, IEEE Transactions on, 2010. 58(10): p. 5030-5043.

19. Donoho, D.L. and M. Elad, On the stability of the basis pursuit in the presence of noise. Signal Processing, 2006. 86(3): p. 511-532.

20. Donoho, D.L., M. Elad, and V.N. Temlyakov, Stable recovery of sparse overcomplete representations in the presence of noise. Information Theory, IEEE Transactions on, 2006. 52(1): p. 6-18.

21. Haupt, J. and R. Nowak, Signal reconstruction from noisy random projections. Information Theory, IEEE Transactions on, 2006. 52(9): p. 4036-4048.

22. Tropp, J. and A.C. Gilbert, Signal recovery from partial information via orthogonal matching pursuit. 2005, Citeseer.

23. Needell, D. and R. Vershynin, Signal recovery from incomplete and inaccurate measurements via regularized orthogonal matching pursuit. Selected Topics in Signal Processing, IEEE Journal of, 2010. 4(2): p. 310-316.
24. Boyd, S., et al., Distributed optimization and statistical learning via the alternating direction method of multipliers. Foundations and Trends® in Machine Learning, 2011. 3(1): p. 1-122.

25. Yin, W., et al., Bregman iterative algorithms for \ell_1-minimization with applications to compressed sensing. SIAM Journal on Imaging Sciences, 2008. 1(1): p. 143-168.

26. Candes, E. and J. Romberg, Sparsity and incoherence in compressive sampling. Inverse problems, 2007. 23(3): p. 969.

27. Do, T.T., T.D. Tran, and L. Gan. Fast compressive sampling with structurally random matrices. in Acoustics, Speech and Signal Processing, 2008. ICASSP 2008. IEEE International Conference on. 2008. IEEE.

28. Stein, E.M. and R. Shakarchi, Fourier analysis: an introduction. Vol. 1. 2011: Princeton University Press.

29. Candes, E.J. and Y. Plan, A probabilistic and RIPless theory of compressed sensing. Information Theory, IEEE Transactions on, 2011. 57(11): p. 7235-7254.

30. Wright, J. and Y. Ma, Dense error correction via-minimization. Information Theory, IEEE Transactions on, 2010. 56(7): p. 3540-3560.

31. Wright, J., et al., Robust face recognition via sparse representation. Pattern Analysis and Machine Intelligence, IEEE Transactions on, 2009. 31(2): p. 210-227.

32. Candes, E.J., et al., Robust principal component analysis? Journal of the ACM (JACM), 2011. 58(3): p. 11.

33. Foygel, R. and L. Mackey, Corrupted sensing: Novel guarantees for separating structured signals. Information Theory, IEEE Transactions on, 2014. 60(2): p. 1223-1247.

34. Donoho, D.L. and P.B. Stark, Uncertainty principles and signal recovery. SIAM Journal on Applied Mathematics, 1989. 49(3): p. 906-931.

35. Donoho, D.L. and X. Huo, Uncertainty principles and ideal atomic decomposition. Information Theory, IEEE Transactions on, 2001. 47(7): p. 2845-2862.

36. Donoho, D.L. and M. Elad, Optimally sparse representation in general (nonorthogonal) dictionaries via \ell_1 minimization. Proceedings of the National Academy of Sciences, 2003. 100(5): p. 2197-2202.

37. Gribonval, R. and M. Nielsen, Sparse representations in unions of bases. Information Theory, IEEE Transactions on, 2003. 49(12): p. 3320-3325.

38. Pope, G., A. Bracher, and C. Studer, Probabilistic recovery guarantees for sparsely corrupted signals. Information Theory, IEEE Transactions on, 2013. 59(5): p. 3104-3116.

39. Fuchs, J.-J., On sparse representations in arbitrary redundant bases. Information Theory, IEEE Transactions on, 2004. 50(6): p. 1341-1344.

40. Tao, T., An uncertainty principle for cyclic groups of prime order. arXiv preprint math/0308286, 2003.