This article considers nonuniform support recovery via Orthogonal Matching Pursuit (OMP) from noisy random measurements. Given $m$ admissible random measurements (of which Subgaussian measurements is a special case) of a fixed $s$-sparse signal $x$ in $\mathbb{R}^n$ corrupted with additive noise, we show that under a condition on the minimum magnitude of the nonzero components of $x$, OMP can recover the support of $x$ exactly after $s$ iterations with overwhelming probability provided that $m = O(s \log n)$. This extends the results of Tropp and Gilbert (2007) [53] to the case with noise. It is a real improvement over previous results in the noisy case, which are based on mutual incoherence property or restricted isometry property analysis and require $O(s^2 \log n)$ random measurements. In addition, this article also considers sparse recovery from noisy random frequency measurements via OMP. Similar results can be obtained for the partial random Fourier matrix via OMP provided that $m = O(s(s + \log(n - s)))$. Thus, for some special cases, this answers the open question raised by Kunis and Rauhut (2008) [34], and Tropp and Gilbert (2007) [53].

© 2012 Elsevier Inc. All rights reserved.

Keywords: Orthogonal Matching Pursuit; Compressed sensing; Nonuniform sparse recovery; Support recovery

1. Introduction

Compressed sensing is a new type of sampling theory, that predicts sparse signal can be reconstructed from what was previously believed to be incomplete information [10,11,18]. In
compressed sensing, one considers the following model:
\[ y = Ax + z, \]  
(1.1)
where \( A \) is a known \( m \times n \) measurement matrix (with \( m \ll n \)) and \( z \in \mathbb{R}^m \) is a vector of measurement errors. The goal is to reconstruct the unknown signal \( x \in \mathbb{R}^n \) based on \( y \) and \( A \). Clearly, in general this task is impossible since even if \( A \) has full rank then there are infinitely many solutions to this equation. The situation dramatically changes if \( x \) is sparse, i.e., \( \|x\|_0 = \{|j : x_j \neq 0\} \) is small.

The approach for solving this problem probably comes first to mind is to search for the sparsest vector in the feasible set of possible solutions, which leads to the \( l_0 \)-minimization. However, solving \( l_0 \)-minimization directly is NP-hard in general and thus is computationally infeasible \([35,39]\). Then it is natural to consider the method of basis pursuit (BP), which can be viewed as a convex relaxation of \( l_0 \)-minimization that consists in solving the following \( l_1 \)-minimization problem:
\[
\min_{\tilde{x} \in \mathbb{R}^n} \|\tilde{x}\|_1 \quad \text{subject to } A\tilde{x} - y \in B, \tag{L1,B}
\]
where \( B \) is a bounded set determined by noise structure. For instance, \( B = \{0\} \) in the noiseless case and \( B = \{r : \|r\|_2 \leq b\} \) or \( B = \{r : \|A^*r\|_\infty \leq b_\infty\} \) in the noisy case. For any \( p \in [1, \infty) \), \( u \in \mathbb{R}^d \), denote \( \|u\|_p = \left(\sum_{j=1}^d |u_j|^p\right)^{1/p} \) and \( \|u\|_\infty = \max_j |u_j| \). \( l_1 \)-minimization problems with different types of constraints have been well studied in the literature \([7–14,18–21, 25,24,33]\). Donoho et al. \([20]\) considered constrained \( l_1 \)-minimization under \( l_2 \) constraint. Candès and Tao \([14]\) introduced the Dantzig Selector, which is a constrained \( l_1 \)-minimization under \( l_\infty \) constraint. Now it has been shown that \( l_1 \)-minimization recovers all \( s \)-sparse vectors with small or zero errors provided that the measurement matrix \( A \) satisfies a restricted isometry property (RIP) condition \( \delta_{cs} \leq C \) for some constants \( c, C > 0 \) \([14,9,12,25,7,24,37]\). Let us mention a few results, the condition \( \delta_{2s} < 0.414 \) was used in Candès \([9]\), \( \delta_{2s} < 0.453 \) in Foucart and Lai \([25]\), \( \delta_{2s} < 0.472 \) with the provision that \( s \) is either large or a multiple of 4 in Cai et al. \([7]\) and \( \delta_{2s} < 0.493 \) in Mo and Li \([37]\). For an \( m \times n \) matrix \( A \) and \( s \leq n \), the RIP constant \( \delta_s \) \([11,13,19]\) is defined as the smallest number such that for all \( s \)-sparse vectors \( \tilde{x} \in \mathbb{R}^n \),
\[
(1 - \delta_s)\|\tilde{x}\|_2^2 \leq \|A\tilde{x}\|_2^2 \leq (1 + \delta_s)\|\tilde{x}\|_2^2.
\]
Note that it is hard to check that a deterministic matrix \( A \) has a small RIP constant. So the strategy is to prove that a random matrix satisfies the RIP condition. It is now well known \([11,2,36,49]\) that many types of random measurement matrices such as Gaussian matrices or Subgaussian matrices have RIP constant \( \delta_s \leq \delta \) with overwhelming probability provided that
\[
m \geq C\delta^{-2}s \log(n/s). \tag{1.2}
\]
Up to the constant, the lower bounds for Gelfand widths of \( l_1 \)-balls \([27,26]\) show that this dependence on \( n \) and \( s \) is optimal. The RIP condition also holds for a rich class of structured random matrices \([11,49,33,48,44,47,43]\). The fast multiply partial random Fourier matrix has RIP constant \( \delta_s \leq \delta \) with very high probability provided that \( m = O(\delta^{-2}s(\log n)^4) \) \([11,49,33]\). Based on its RIP guarantees, with high probability, BP can recover every \( s \)-sparse vector with small or zero errors from \( O(s \log(n/s)) \) (or with additional log factors in \( n \)) random measurements. But in practice, BP is too expensive in large-scale applications. In fact, numerous researchers have
claimed that BP is much too slow for large-scale applications [53,22]. Thus it is necessary to use alternative iterative methods for sparse recovery that are not based on optimization.

There are many pursuit methods for sparse recovery in the literature, including Orthogonal Matching Pursuit (OMP) [42,17], Stagewise OMP [23], Regularized OMP [41], Compressive Sampling Matching Pursuit [40], Iterative Hard Thresholding [4], Subspace Pursuit [15] and many other variants. We refer the readers to [54] for an overview of these pursuit methods.

The accurate recovery of signal support is crucial to compressed sensing both in theory and in practice. If one has obtained the support of the signal, then one can recover the signal by solving a least squares problem. In practical applications, the support is usually physically more significant than the component values; see for example [3,30,31]. Refer to [51] for more discussions on sparse support recovery.

In this paper, we consider sparse support recovery from noisy random measurements via OMP. OMP is a greedy algorithm that finds the support of the signal iteratively, and reconstructs the signal using the pseudoinverse. The entire algorithm is specified in Section 2, Table 1. Compared with other alternative methods, a major advantage of OMP is its simplicity and fast implementation. Theoretical analysis of OMP for sparse support recovery to date has concentrated primarily on two types.

- Uniform recovery: results state that with high probability on the draw of the random matrix, the support of every sparse signal can be reconstructed under appropriate conditions.
- Nonuniform recovery: results state that for a given sparse signal, its support can be reconstructed with high probability on the draw of the random matrix under appropriate conditions.

Theoretical analysis of uniform recovery via OMP has concentrated primarily on two fronts. The first one has involved the notion of a mutual incoherence property (MIP) introduced by Donoho and Huo [21]. The mutual incoherence of a measurement matrix \( A \) is defined by

\[
\mu = \max_{j \neq k} \frac{|\langle A_j, A_k \rangle|}{\|A_j\|_2 \|A_k\|_2},
\]

where \( A_j \) denotes the \( j \)th column in \( A \). Tropp [52] showed that if \( \mu < \frac{1}{\sqrt{s-1}} \), then OMP can recover every \( s \)-sparse signal in \( s \) iterations in the noiseless case. For the noisy case, Donoho et al. [20] showed that under conditions on the MIP and the minimum magnitude of the nonzero components of the signal, the signal support can be recovered exactly by OMP in \( s \) iterations. Such object was also considered in [6,29]. The other kind of theoretical analysis of uniform recovery via OMP has involved the RIP of the measurement matrix. For example, in the noiseless case, it was shown that \( \delta_{s+1} < \frac{1}{\sqrt{s}} \) [16] is a sufficient condition for OMP to recover every \( s \)-sparse signal in \( s \) iterations successfully. Later, this sufficient condition was improved to \( \delta_{s+1} < \frac{1}{\sqrt{s+1}} \) [38]. For any \( s \geq 2 \), Mo and Shen [38] constructed matrix with \( \delta_{s+1} = \frac{1}{\sqrt{s}} \) such that OMP cannot recover some \( s \)-sparse signal in \( s \) iterations. The noisy case was considered in [32,50].

Unfortunately, from (1.2), we see that finding a random matrix satisfying \( \delta_{s+1} < \frac{1}{\sqrt{s+1}} \) will require \( O(s^2 \log(n/s)) \) (or with additional log factors in \( n \)) measurements. Also, the MIP condition \( \mu < \frac{1}{\sqrt{s-1}} \) would lead to \( O(s^2 \log(n/s)) \) random measurements; see [53] for Subgaussian measurements and [34] for random frequency measurements. Therefore, it is impossible to develop stronger results by way of MIP. In particular, it has been shown that when \( m \leq O(s^{3/2}) \), for most random matrices there will exist some \( s \)-sparse signal that cannot be recovered exactly via \( s \) iterations of OMP [45].
Aforehand analysis shows that uniform recovery via $s$ iterations of OMP roughly requires $O(s^2 \log(n/s))$ random measurements, and it seems that the number cannot be further improved based on RIP or MIP guarantees. Comparing with BP which requires $O(s \log(n/s))$ measurements, the measurements needed for OMP is a bit large. A good news is that Tropp and Gilbert [53] showed that for a fixed $s$-sparse signal, given $O(s \log n)$ random measurements of that signal, OMP can recover the signal in $s$ iterations with high probability in the noiseless case. However, there are fewer results for the general noisy case.

In this paper, we focus on nonuniform support recovery via OMP with random measurements corrupted by additive noise. In this case, the observed vector $y$ is given by (1.1), while $A$ is a random matrix and the noise vector $z$ is not equal to zero in general. Our first contribution is to extend the main results in [53] to the noisy case. Given $m$ Bernoulli measurements of a fixed $s$-sparse signal $x \in \mathbb{R}^n$ corrupted with additive noise, we first show that under a condition on the minimum magnitude of the nonzero coordinates, the support of $x$ can be recovered exactly via OMP in $s$ iterations with overwhelming probability provided that $m = O(s \log n)$. We consider two types of bounded noise: the $l_2$ bounded noise and the $l_\infty$ bounded noise. Once the bounded noise cases are understood, the Gaussian noise case follows easily. Furthermore, it is also shown that our results can be extended to general types of random measurement ensembles satisfying the four properties. Since our condition on the nonzero coefficients of the signal is roughly the same as that in the previous results, see for example [20, Theorem 5.1], this is a real improvement over previous results in the noisy case, which are based on MIP or RIP analysis and require $O(s^2 \log n)$ random measurements.

Note that the previous method is heavily based on that the columns of the measurement matrix $A$ are statistically independent, so the approach cannot be applied directly to the partial random Fourier matrix. Compared with Bernoulli/Gaussian matrix, the partial random Fourier matrix is preferable since it can be stored efficiently and has fast algorithms for matrix vector multiplication using a fast Fourier transform. Random frequency measurements also arise from practical possible applications [10]. Kunis and Rauhut [34,33] have studied the performance of OMP for signal recovery from random frequency measurements. They showed that the first iteration of OMP is likely to choose a correct column from the measurement matrix, given $O(s \log n)$ measurements of an $s$-sparse signal. While their numerical experiments indicate that $O(s \log n)$ measurements are sufficient for OMP to recover an $s$-sparse signal, there are fewer theoretical results. Until now, it still remains an open question to ascertain whether $O(s \log n)$ random frequency measurements are sufficient for OMP to recover an $s$-sparse signal [34,53].

Our second contribution is that we develop results on sparse support recovery from noisy random frequency measurements via OMP. Given $m$ noisy random frequency measurements of a fixed $s$-sparse signal $x$ corrupted with additive noise, we show that under a condition on the minimum magnitude of the nonzero coordinates, the support of $x$ can be recovered exactly via OMP in $s$ iterations with overwhelming probability provided that $m = O(s(s + \log(n-s)))$. In particular, when $s \leq O(\log n)$, the condition on $m$ becomes $m = O(s \log n)$. Thus, for some special cases, this answers the open question raised in [34,53].

This paper is organized as follows. In Section 2, we specify the OMP algorithm first. Then we give two lemmas which are useful to our proofs of the main results. In Section 3, we introduce our random measurement ensembles and in Section 4, we state our main results. Section 5 presents the proofs for main results related to sparse support recovery from Bernoulli measurements via OMP, and Section 6 deals with proofs for main results related to frequency random measurements.
Table 1
Orthogonal Matching Pursuit (OMP).

**INPUT:** measurement matrix \( A \), vector \( y \) and the stopping criterion.

**PROCEDURE:**

1. Set the residual \( r_0 = y \), the index set \( \Omega_0 = \emptyset \), and the iteration counter \( t = 1 \).
2. Find \( k_t = \arg \max_{j \in [n]} \| r_{t-1} \cdot A_j \| \). Augment the index set \( \Omega_t = \Omega_{t-1} \cup \{ k_t \} \).
   (if multiple maxima exist, choose the one with minimal index).
3. Solve \( x_t = \arg \min_{x \in \mathbb{R}^{|\Omega_t|}} \| A_{\Omega_t} x - y \|_2 \). Update the residual \( r_t = y - A_{\Omega_t} x_t \).
4. End if the stopping condition is achieved. Otherwise, set \( t = t + 1 \) and turn to Step 2.

**OUTPUT:** \( \Omega_t, \hat{x}_{\Omega_t} = x_t, \hat{x}_{\Omega_t^c} = 0 \).

The following notation is used throughout this paper. The set of indices of the nonzero entries of a vector \( \hat{x} \) is called the support of \( \hat{x} \) and denoted as \( \text{supp}(\hat{x}) \). For \( n \in \mathbb{R} \), denote \([n] \) to mean \( \{1, 2, \ldots, n\} \). Given an index set \( T \subset [n] \) and a matrix \( A \), \( T^c \) is the complement of \( T \) in \([n]\), \( A_T \) is the submatrix of \( A \) formed from the columns of \( A \) indexed by \( T \). For \( j \in [n] \), \( A_j \) is the \( j \)th column of \( A \). \( \hat{x}_T \) is the vector equal to \( \hat{x} \) restricted to \( T \) and zero elsewhere or a vector of \( \hat{x} \) restricted to \( T \). Write \( A^* \) to mean the conjugate transpose of a matrix \( A \), \( \lambda_{\min}(A^* A) \) and \( \lambda_{\max}(A^* A) \) to mean the smallest and largest eigenvalues of \( A^* A \), \( \sigma_{\min}(A) \) and \( \sigma_{\max}(A) \) to mean the smallest and largest singular values of \( A \). \( C > 0 \) (or \( c, c_1 \)) denotes a universal constant that might be different in each occurrence.

2. Orthogonal Matching Pursuit

OMP is an iterative greedy algorithm. At each iteration, it selects one column of \( A \) which is most correlated with the current residual. This column is then added into the set of selected columns. The residuals are updated by projecting the observation \( y \) onto the linear subspace spanned by the columns that have already been selected, and the algorithm then iterates. The algorithm is explained in Table 1.

Notice that in Step 3, one can compute that \( x_t = (A_{\Omega_t}^* A_{\Omega_t})^{-1} A_{\Omega_t}^* y \) and \( r_t = (I - P_t) y \), where \( P_t = A_{\Omega_t} (A_{\Omega_t}^* A_{\Omega_t})^{-1} A_{\Omega_t}^* \) denotes the projection onto the linear space spanned by the columns of \( A_{\Omega_t} \). It thus follows that no column is selected twice and the index set of selected columns grows at each iteration.

One need to set the stopping rule before running the algorithm. In general, there are several natural stopping criteria.

- Stop after a fixed number of iterations: \( s \).
- Stop when the residual has small \( l_2 \) norms: \( \| r_t \|_2 \leq b_2 \).
- Stop when no column explains a significant amount of energy in the residual: \( \| A^* r_t \|_\infty \leq b_\infty \).

In the noiseless case the natural stopping rule is fixing the number of iterations to be the sparsity level of the unknown signal \( x \) or \( r_t = 0 \). In this paper, we use the stopping rule \( \| r_t \|_2 \leq b_2 \) for \( l_2 \) bounded noise while \( \| A^* r_t \|_\infty \leq b_\infty \) for \( l_\infty \) bounded noise. These stopping rules are reasonable since in the special case of \( z = 0 \), they will guarantee that OMP does not select any incorrect index.

The following result is due to Cai and Wang [6, Lemma 5].

**Lemma 2.1.** Let \( T \subset [n], \Omega_t \subset T \) and \( u_t = T / \Omega_t \). Denote \( P_t = A_{\Omega_t} (A_{\Omega_t}^* A_{\Omega_t})^{-1} A_{\Omega_t}^* \). Then the minimum eigenvalue of \( A_T^* A_T \) is less than or equal to the minimum eigenvalue of \( A_{\Omega_t}^* A_{\Omega_t} \).
The maximum eigenvalue $A^*_u(I - P_t)A_u$ is greater than or equal to the maximum eigenvalue of $A^*_u(I - P_t)A_u$.

The following result gives a sufficient condition for OMP to choose a correct index at the current iteration.

**Lemma 2.2.** Let $y = Ax + z$ and $T = \text{supp}(x)$ with $|T| = s$. Assume that at the first $t$ iterations ($t < s$), OMP selects $t$ correct indices, that is, $\Omega_t \subset T$. If for some $\rho \in (0, 1)$,

$$
\|A^*_T(I - P_t)A_Tx_T\|_\infty < \rho \|A^*_T(I - P_t)A_Tx_T\|_\infty \tag{2.1}
$$

and

$$
\|x_T \setminus \Omega_t\|_2 \geq \frac{2(s - t)^{1/2}}{\lambda_{\min}(A^*_T A_T)(1 - \rho)} \|A^*(I - P_t)z\|_\infty, \tag{2.2}
$$

then

$$
\frac{\|A^*_T r_t\|_\infty}{\|A^*_T r_t\|_\infty} < 1. \tag{2.3}
$$

Furthermore, if $r_t$ does not satisfy the stopping rule, then OMP will select a true index from $T$ at the iteration $t + 1$.

**Proof.** Note that by (2.1), we have

$$
\|A^*_T r_t\|_\infty \leq \|A^*_T(I - P_t)A_Tx_T\|_\infty + \|A^*_T(I - P_t)z\|_\infty < \rho \|A^*_T(I - P_t)A_Tx_T\|_\infty + \|A^*(I - P_t)z\|_\infty
$$

and

$$
\|A^*_T r_t\|_\infty \geq \|A^*_T(I - P_t)A_Tx_T\|_\infty - \|A^*_T(I - P_t)z\|_\infty
$$

$$
\geq \|A^*_T(I - P_t)A_Tx_T\|_\infty - \|A^*(I - P_t)z\|_\infty.
$$

It thus follows that a sufficient condition for (2.3) is

$$
\|A^*_T(I - P_t)A_Tx_T\|_\infty \geq \frac{2}{1 - \rho} \|A^*(I - P_t)z\|_\infty. \tag{2.4}
$$

Since

$$
\|A^*_T(I - P_t)A_Tx_T\|_\infty = \|A^*_T(I - P_t)A_T x_T \setminus \Omega_t\|_\infty
$$

$$
\geq (s - t)^{-1/2} \|A^*_T(I - P_t)A_T x_T \setminus \Omega_t\|_2
$$

$$
\geq (s - t)^{-1/2} \lambda_{\min}(A^*_T A_T) \|x_T \setminus \Omega_t\|_2.
$$

where we have used Lemma 2.1 at the last inequality, with the assumption and by an easy computation one can get (2.4). \qed

### 3. Random measurement ensembles

Bernoulli matrix and Gaussian matrix are two important measurement ensembles used in compressed sensing.
• Bernoulli matrix. The entries of a Bernoulli matrix are independent realizations of $\pm 1/\sqrt{m}$ Bernoulli random variable, that is, each entry takes the value $1/\sqrt{m}$ or $-1/\sqrt{m}$ with equal probability.

• Gaussian matrix. Here the entries of $A$ are chosen as i.i.d Gaussian random variables with expectation 0 and variance $1/m$.

From lemmas in Section 5.1, we can see that Bernoulli/Gaussian matrix is one of the random matrices satisfying the following four properties.

(0) Independence: the columns of $A$ are statistically independent.

(1) Near normalization: for any $\epsilon > 0$, $\max_{j \in [n]} \|A_j\|_2^2 \leq (1 + \epsilon)$ holds with overwhelming probability.

(2) Joint correlation: let $\{U_t\}$ be a sequence of $s$ vectors whose $l_2$ norms do not exceed one. Let $A_j$ be a column of $A$ that is independent from this sequence. Then
\[ \mathbb{P} \left( \max_t |\langle A_j, U_t \rangle| \leq \varepsilon \right) \geq 1 - 2se^{-c\varepsilon^2m}. \]

(3) Smallest singular value: for a given $m \times s$ submatrix $A_{\Gamma}$ from $A$, the $s$th largest eigenvalue $\lambda_{\min}(A_{\Gamma}^*A_{\Gamma})$ of $A_{\Gamma}^*A_{\Gamma}$ satisfies
\[ \mathbb{P} \left( \lambda_{\min}(A_{\Gamma}^*A_{\Gamma}) \geq 1 - \delta \right) \geq 1 - \left( \frac{c_1}{\delta} \right)^s e^{-cm\delta^2}. \]

We call such random matrices admissible measurement matrices or admissible random measurement ensembles. Our proofs of the main results related to Bernoulli matrix in Section 5 are essentially based on the fact that Bernoulli matrix is an admissible measurement matrix. Therefore, our results can be extended to any admissible measurement matrices, such as sub-Gaussian random matrices with independent columns follows column-independent model [55, p. 49] or measurement ensembles appeared in [1, Section 3.2].

A random partial Fourier matrix $A \in \mathbb{C}^{m \times n}$ is of the form
\[ A = \frac{1}{\sqrt{m}} \left( e^{-2\pi i \omega_j k} \right)_{1 \leq j \leq m, 1 \leq k \leq n}, \]
where $\omega_1, \omega_2, \ldots, \omega_m$ are independent random variables having the uniform distribution on $\{0, 1/n, \ldots, (n-1)/n\}$ (or the uniform distribution on $[0, 1]$) [10,46].

4. Main results

4.1. Bernoulli matrices

**Theorem 4.1.** Suppose that $A$ is an $m \times n$ Bernoulli matrix and the noise vector $z \in \mathbb{R}^m$ is independent of $A$ with $\|z\|_2 \leq b_2$. Fix $\beta, \delta$ and $\rho$ in $(0, 1)$, and choose $m \geq Cs \max\{\rho^{-2} (1 - \delta)^{-1} \log(n\beta^{-1}), \delta^{-2} \log(\delta^{-1}\beta^{-1})\}$. Let $x \in \mathbb{R}^n$ be an arbitrary fixed $s$-sparse vector such that all its nonzero coefficients $x_i$ satisfy
\[ |x_i| \geq \frac{2b_2}{(1 - \delta)(1 - \rho)}. \]

Given the data $y = Ax + z$, the OMP algorithm with the stopping rule $\|r_t\|_2 \leq b_2$ can recover the support of $x$ in $s$ iterations with probability exceeding $1 - \beta$. 

Remark 4.2. The condition on the nonzero coefficients of the signal \((|x_i| \geq 2b_2)\) required in this theorem is roughly the same as that in [20, Theorem 5.1] (also [6, Theorem 1]), which is based on MIP analysis and requires \(O(s^2 \log n)\) random measurements.

**Theorem 4.3.** Suppose that \(A\) is an \(m \times n\) Bernoulli matrix and \(z \in \mathbb{R}^m\) is statistically independent from \(A\) such that \(\|A^*z\|_\infty \leq b_\infty\). Fix \(\beta, \delta\) and \(\rho\) in \((0, 1)\), and choose \(m \geq C_s \max\{\rho^{-2}(1-\delta)^{-1} \log(n\beta^{-1}), \delta^{-2} \log(\delta^{-1} \beta^{-1})\}\). Let \(x \in \mathbb{R}^n\) be an arbitrary fixed \(s\)-sparse vector such that all its nonzero coefficients \(x_i\) satisfy
\[
|x_i| \geq \frac{2}{(1-\delta)(1-\rho)} \left(1 + \sqrt{s \frac{1}{1-\delta}}\right) b_\infty.
\]
Given the data \(y = Ax + z\), the OMP algorithm with the stopping rule \(\|A^*r_i\|_\infty \leq b_\infty\) can return the support of \(x\) in \(s\) iterations with probability exceeding \(1 - \beta\).

Remark 4.4. The condition on the nonzero coefficients of the signal required in this theorem is roughly the same as that in [6, Theorem 4], which is based on MIP analysis and requires \(O(s^2 \log n)\) random measurements.

Let \(z \sim N(0, \sigma^2 I_m)\). Define a bounded set
\[
\mathcal{B}_1 = \left\{ z : \|z\|_2 \leq \sigma \sqrt{m + 2 \sqrt{m \log m}} \right\}.
\]
We have the following result; see [8, Lemma 5.1].

**Lemma 4.5.** The Gaussian noise vector \(z \sim N(0, \sigma^2 I_m)\) satisfies
\[
\mathbb{P}(z \in \mathcal{B}_1) \geq 1 - \frac{1}{m}.
\] (4.2)

The following result is a direct consequence of Theorem 4.1 and Lemma 4.5.

**Theorem 4.6.** Suppose that \(A\) is an \(m \times n\) Bernoulli matrix and the noise vector \(z \sim N(0, \sigma^2 I_m)\) is statistically independent from \(A\). Fix \(\beta, \delta\) and \(\rho\) in \((0, 1)\), and choose \(m \geq C_s \max\{\rho^{-2}(1-\delta)^{-1} \log(n\beta^{-1}), \delta^{-2} \log(\delta^{-1} \beta^{-1})\}\). Let \(x \in \mathbb{R}^n\) be an arbitrary fixed \(s\)-sparse vector such that all its nonzero coefficients \(x_i\) satisfy
\[
|x_i| \geq \frac{2\sigma \sqrt{m + 2 \sqrt{m \log m}}}{(1-\delta)(1-\rho)}.
\]
Given the data \(y = Ax + z\), the OMP algorithm with the stopping rule \(\|r_i\|_2 \leq \sigma \sqrt{m + 2 \sqrt{m \log m}}\) can recover the support of \(x\) in \(s\) iterations with probability exceeding \(1 - \beta - 1/m\).

4.2. Extended to admissible measurement matrices

Our results can be extended to Gaussian matrix or any admissible measurement matrix. For Gaussian matrix or admissible measurement matrix, we assume that every column of \(A\) has unit \(l_2\) norms. In fact, using Lemma 5.2 or the near normalization property, one can show that the \(l_2\) norms of every column of \(A\) is approximately equal to one with large probability. The proofs of the main results in this subsection are analogous as that for Bernoulli matrix. We omit them.
Theorem 4.7. Suppose that $A$ is an $m \times n$ Gaussian matrix or admissible measurement matrix, and $z \in \mathbb{R}^m$ is independent from $A$ with $\|z\|_2 \leq b_2$. Fix $\beta$, $\delta$ and $\rho$ in $(0, 1)$, and choose $m \geq Cs \max\{\rho^{-2}(1 - \delta)^{-1} \log(n\beta^{-1}), \delta^{-2} \log(\delta^{-1}\beta^{-1})\}$. Let $x \in \mathbb{R}^n$ be an arbitrary fixed $s$-sparse vector such that all its nonzero coefficients $x_i$ satisfy
\[
|x_i| \geq \frac{2b_2}{(1 - \delta)(1 - \rho)}.
\]
Given the data $y = Ax + z$, the OMP algorithm with the stopping rule $\|r_i\|_2 \leq b_2$ can recover the support of $x$ in $s$ iterations with probability exceeding $1 - \beta$.

Theorem 4.8. Suppose that $A$ is an $m \times n$ Gaussian matrix or admissible measurement matrix, and $z \in \mathbb{R}^m$ is independent from $A$ with $\|A^*z\|_\infty \leq b_\infty$. Fix $\beta$, $\delta$ and $\rho$ in $(0, 1)$, and choose $m \geq Cs \max\{\rho^{-2}(1 - \delta)^{-1} \log(n\beta^{-1}), \delta^{-2} \log(\delta^{-1}\beta^{-1})\}$. Let $x \in \mathbb{R}^n$ be an arbitrary fixed $s$-sparse vector such that all its nonzero coefficients $x_i$ satisfy
\[
|x_i| \geq \frac{2}{(1 - \delta)(1 - \rho)} \left(1 + \sqrt{s} \right) b_\infty.
\]
Given the data $y = Ax + z$, the OMP algorithm with the stopping rule $\|A^*r_i\|_\infty \leq b_\infty$ can return the support of $x$ in $s$ iterations with probability exceeding $1 - \beta$.

Combining Theorem 4.7 with Lemma 4.5, one gets the following result.

Theorem 4.9. Suppose that $A$ is an $m \times n$ Gaussian matrix or admissible measurement matrix, and the noise vector $z \sim N(0, \sigma^2 I_m)$ is independent from $A$. Fix $\beta$, $\delta$ and $\rho$ in $(0, 1)$, and choose $m \geq Cs \max\{\rho^{-2}(1 - \delta)^{-1} \log(n\beta^{-1}), \delta^{-2} \log(\delta^{-1}\beta^{-1})\}$. Let $x \in \mathbb{R}^n$ be an arbitrary fixed $s$-sparse vector such that all its nonzero coefficients $x_i$ satisfy
\[
|x_i| \geq \frac{2\sqrt{m + 2\sqrt{m \log m}}}{(1 - \delta)(1 - \rho)}.
\]
Given the data $y = Ax + z$, the OMP algorithm with the stopping rule $\|r_i\|_2 \leq \sigma \sqrt{m + 2\sqrt{m \log m}}$ can recover the support of $x$ in $s$ iterations with probability exceeding $1 - \beta - 1/m$.

4.3. Random partial Fourier matrices

In this section, we discuss sparse support recovery from noisy random frequency measurements via OMP.

Theorem 4.10. Suppose that $A$ is an $m \times n$ random partial Fourier matrix and the noise vector $z \in \mathbb{R}^m$ is independent of $A$ with $\|z\|_2 \leq b_2$. Fix $\beta$ in $(0, 1)$ and $\delta$ in $(0, 1/2)$. Choose
\[
m \geq Cs^{-2}s(\log(n - s) + \log \beta^{-1}).
\]
Let $x \in \mathbb{R}^n$ be an arbitrary fixed $s$-sparse vector such that all its nonzero coefficients $x_i$ satisfy
\[
|x_i| \geq \frac{2b_2}{1 - 2\delta}.
\]
Given the data $y = Ax + z$, the OMP algorithm with the stopping rule $\|r_i\|_2 \leq b_2$ can recover the support of $x$ in $s$ iterations with probability exceeding $1 - \beta$. 
Remark 4.11. (a) In the noiseless, i.e., $b_2 = 0$, the above theorem does not impose any condition on the minimum magnitude of the nonzero components of $x$.

(b) The measurement number $m$ (4.3) required in this theorem is smaller than that in [33, Theorem 4.2] (also [34, Theorem 2.7]), which is based on MIP analysis and requires $O(s^2 \log n)$ random measurements. In particular, for special cases $s \leq O(\log n)$, condition (4.3) becomes $m = O(s \log n)$. Thus, for some special cases, this answers the open question raised by Kunis and Rauhut [34] (also Tropp and Gilbert [53]).

(c) From the proof of Theorem 4.10 in Section 6, one can see that the above theorem holds simultaneously for all $s$-sparse signals which have the same support and satisfy (4.4).

Theorem 4.12. Suppose that $A$ is an $m \times n$ random partial Fourier matrix and the noise vector $z \in \mathbb{R}^m$ is independent of $A$ with $\|A^*z\|_\infty \leq b_\infty$. Fix $\beta$ in $(0, 1)$ and $\delta$ in $(0, 1/2)$. Choose $m \geq C\delta^{-2}s(s + \log(n - s) + \log \beta^{-1})$. Let $x \in \mathbb{R}^n$ be an arbitrary fixed $s$-sparse vector such that all its nonzero coefficients $x_i$ satisfy

$$|x_i| \geq \left(1 + \sqrt{\frac{s}{1 - \delta}}\right) \frac{2b_\infty}{1 - 2\delta}.$$ 

Given the data $y = Ax + z$, the OMP algorithm with the stopping rule $\|A^*r_t\|_\infty \leq b_\infty$ can return the support of $x$ with probability exceeding $1 - \beta$.

Combining Theorem 4.10 with Lemma 4.5, one gets the following result.

Theorem 4.13. Suppose that $A$ is an $m \times n$ random partial Fourier matrix, and the noise vector $z \sim N(0, \sigma^2 I_m)$ is independent from $A$. Fix $\beta$ in $(0, 1)$ and $\delta$ in $(0, 1/2)$. Choose $m \geq C\delta^{-2}s(s + \log(n - s) + \log \beta^{-1})$. Let $x \in \mathbb{R}^n$ be an arbitrary fixed $s$-sparse vector such that all its nonzero coefficients $x_i$ satisfy

$$|x_i| \geq \frac{2\sigma \sqrt{m + 2\sqrt{m \log m}}}{1 - 2\delta}.$$

Given the data $y = Ax + z$, the OMP algorithm with the stopping rule $\|r_t\|_2 \leq \sigma \sqrt{m + 2\sqrt{m \log m}}$ can recover the support of $x$ in $s$ iterations with probability exceeding $1 - \beta - 1/m$.

5. Proofs for main results related to Bernoulli matrices

5.1. Lemmas

In this subsection, we will give some basic lemmas which are useful for our proofs of our main results. We begin by introducing some basic lemmas related to Bernoulli matrix and Gaussian matrix, which together show that Bernoulli matrix and Gaussian matrix are admissible measurement matrices.

For Bernoulli matrix $A$, $\|A_j\|_2 = 1$ for every $j \in [n]$. Consequently, Bernoulli matrix satisfies the near normalization property (1). For Gaussian matrix, one can also show that it satisfies the near normalization property. For developing such a result, we introduce the following lemma, which is due to Cai [5, Lemma 4].
Lemma 5.1. If \( X \) follows a \( \chi^2 \) distribution with \( m \) degrees of freedom, then for all \( \gamma > 0 \),

\[
\mathbb{P}(X > (1 + \gamma)m) \leq \frac{1}{\gamma \sqrt{\pi m}} e^{-m(\gamma - \log(1 + \gamma))/2}.
\]

Using the above lemma, one can prove that Gaussian matrix satisfies the near normalization property.

Lemma 5.2. Let \( A \) be an \( m \times n \) Gaussian matrix. Then for all \( \gamma \in (0, 1) \),

\[
\mathbb{P}\left( \max_j \|A_j\|_2 \leq \sqrt{1 + \gamma} \right) \geq 1 - \frac{n}{\gamma \sqrt{\pi m}} e^{-m\gamma^2(3 - 2\gamma)/12}.
\]

In particular, if for some \( \epsilon \in (0, 1) \), \( m \geq c\gamma^{-2}(3 - 2\gamma)^{-1} \log\left(\frac{n}{\gamma \epsilon \sqrt{m}}\right) \), then with probability at least \( 1 - \epsilon \), we have \( \max_j \|A_j\|_2 \leq \sqrt{1 + \gamma} \).

Proof. For arbitrary \( j \in [n] \), \( m\|A_j\|_2^2 \) follows a \( \chi^2 \) distribution with \( m \) degrees of freedom. By Lemma 5.1, we have

\[
\mathbb{P}\left( \|A_j\|_2 > \sqrt{1 + \gamma} \right) \leq \frac{1}{\gamma \sqrt{\pi m}} e^{-m(\gamma - \log(1 + \gamma))/2}.
\]

Using the union bound, one has

\[
\mathbb{P}\left( \max_j \|A_j\|_2 > \sqrt{1 + \gamma} \right) \leq \frac{n}{\gamma \sqrt{\pi m}} e^{-m(\gamma - \log(1 + \gamma))/2}.
\]

By using the fact that \( \log(1 + \gamma) \leq \gamma - \gamma^2/2 + \gamma^3/3 \), we get

\[
\mathbb{P}\left( \max_j \|A_j\|_2 > \sqrt{1 + \gamma} \right) \leq \frac{n}{\gamma \sqrt{\pi m}} e^{-m\gamma^2(3 - 2\gamma)/12},
\]

which leads to the result. \( \square \)

The following result is due to Tropp and Gilbert [53, Proposition 4].

Lemma 5.3. Let \( A \) be an \( m \times n \) Bernoulli matrix or Gaussian matrix. Let \( \{U_t\} \subset \mathbb{R}^n \) be a sequence of \( s \) vectors whose \( l_2 \) norms do not exceed one. Let \( A_j \) be a column of \( A \) that is independent from this sequence. Then

\[
\mathbb{P}\left( \max_t |\langle A_j, U_t \rangle| \leq \epsilon \right) \geq 1 - 2s e^{-\epsilon^2m/2}.
\]

A similar approach as that for [2, Lemma 5.1] would yield the following result.

Lemma 5.4. Let \( A \) be an \( m \times n \) Bernoulli matrix or Gaussian matrix. Suppose that \( A_{\Gamma} \) is an \( m \times s \) submatrix from \( A \). Then with probability at least \( 1 - (c_1/\delta)^s \cdot e^{-cm\delta^5} \),

\[
(1 - \delta)\|\tilde{x}\|_2^2 \leq \|A_{\Gamma}\tilde{x}\|_2^2 \leq (1 + \delta)\|\tilde{x}\|_2^2 \quad \text{for all} \quad \tilde{x} \in \mathbb{R}^s.
\]

In particular, if \( m \geq c\delta^{-2} (s \log(c_1/\delta) + \log \epsilon^{-1}) \) for some \( \epsilon \in (0, 1) \), then with probability at least \( 1 - \epsilon \), we have (5.1).
Note that (5.1) is equivalent to
\[ 1 - \delta \leq \lambda_{\text{min}}(A^*_r A_r) \leq \lambda_{\text{max}}(A^*_r A_r) \leq 1 + \delta \]
or
\[ \sqrt{1 - \delta} \leq \sigma_{\text{min}}(A_r) \leq \sigma_{\text{max}}(A_r) \leq \sqrt{1 + \delta}. \]

5.2. Proof of Theorem 4.1

We begin with some notation and a few simplifying assumptions. Let \( T = \text{supp}(x) \). One would note that \( y = A_T x_T + z \) is statistically independent from the random matrix \( A_{T^c} \). For a vector \( r \in \mathbb{R}^m \), we denote
\[ \rho(r) = \frac{\|A_{T^c} r\|_{\infty}}{\|A_r\|_{\infty}}. \]  
(5.2)

If \( r \) is the residual vector in Step 2 of OMP, the algorithm chooses a column from \( A_T \) whenever \( \rho(r) < 1 \). Consider the event \( E_{\text{succ}} \) where the algorithm correctly identifies the true support of \( x \) after \( s \) iterations. Define the event \( \Lambda = \{ \lambda_{\text{min}}(A^*_T A_T) \geq 1 - \delta \} \). Using Lemma 5.4 one would get
\[ \mathbb{P}(\Lambda) \geq 1 - (c_1/\delta)^s \cdot e^{-c_2\delta^2}. \]  
(5.3)

From the definition of conditional probability, we have
\[ \mathbb{P}(E_{\text{succ}}) \geq \mathbb{P}(E_{\text{succ}} \cap \Lambda) = \mathbb{P}(E_{\text{succ}} | \Lambda) \cdot \mathbb{P}(\Lambda). \]  
(5.4)

We will focus on developing a lower bound on \( \mathbb{P}(E_{\text{succ}} | \Lambda) \) in the remainder of the proof.

For deriving a lower bound on \( \mathbb{P}(E_{\text{succ}} | \Lambda) \), we introduce an imaginary event. Imagine that, with the signal \( x \), the noise vector \( z \) and the restricted measurement matrix \( A_T \), we could execute \( s \) iterations of OMP with stopping rule \( \|r_t\|_2 \leq b_2 \) to obtain a sequence of residuals \( q_0, q_1, \ldots, q_{s-1} \) and a sequence of column indices \( \omega_1, \omega_2, \ldots, \omega_s \). The algorithm is deterministic, so these sequences are both functions of \( x, z \) and \( A_T \). In particular, the residuals are statistically independent of \( A_{T^c} \). Note that at iteration \( s \), if OMP exactly recovers the support \( T \), then by an easy computation, the residual \( q_s = (I - P_s)z \). Consequently, the stopping rule is satisfied and OMP stops.

With the signal \( x \), the noise vector \( z \) and the full matrix \( A \), carry out OMP with stopping rule \( \|r_t\|_2 \leq b_2 \) to obtain the actual sequence of residuals \( r_0, r_1, \ldots, r_{s-1} \) and the actual sequence of column indices \( k_1, k_2, \ldots, k_s \). Conditional on \( \Lambda \), if the algorithm selects a correct index from \( T \) at each iteration, then OMP succeeds in recovering the support \( T \) after \( s \) iterations. And after \( s \) iterations, the OMP algorithm stops. A simple induction argument as that in the proof of [53, Theorem 6] shows that this situation occurs when \( \rho(q_t) < 1 \) and \( \|q_t\|_2 > b_2 \) for each \( t = 0, 1, \ldots, s - 1 \). Consequently, we have
\[ \mathbb{P}(E_{\text{succ}} | \Lambda) \geq \mathbb{P}\left( \max_t \rho(q_t) < 1 \text{ and } \min_t \|q_t\|_2 > b_2 | \Lambda \right). \]  
(5.5)

Note that \( \{q_t\} \) is a sequence of \( s \) random vectors depending on \( A_T \) and \( z \), and it is statistically independent from \( A_{T^c} \). Our next part of the proof is to get a lower bound on the right hand side of (5.5).
Write
\[ q_t = (I - P_t)y = (I - P_t)A_Tx_T + (I - P_t)z = s_t + n_t \quad \text{for } t = 0, 1, \ldots, s - 1. \]
It is easy to see that \( s_t \) and \( n_t, t = 0 \ldots s - 1 \) are statistically independent from \( A_T^c \). We will prove that conditional on \( A, \rho(q_t) < 1 \) and \( \|q_t\|_2 > b_2 \) for each \( t = 0, 1, \ldots, s - 1 \) occurs when \( \rho(s_t) < \rho \) for each \( t = 0, 1, \ldots, s - 1 \), on which we establish that
\[ \mathbb{P} \left( \max_t \rho(q_t) < 1 \text{ and } \min_t \|q_t\|_2 > b_2 | A \right) \geq \mathbb{P} \left( \max_t \rho(s_t) < \rho | A \right). \tag{5.6} \]
Note that \( \{s_t\} \) is a sequence of \( s \) random vectors which lie in the column span of \( A_T \) and it is statistically independent from \( A_T^c \).

During the initial iteration, \( q_0 = y = A_Tx_T + z \cdot P_0 = 0 \). Since
\[ \|A^*_n\|_\infty = \max_{j \in [n]} |\langle A_j, z \rangle| \leq \max_{j \in [n]} \|A_j\|_2 \|z\|_2 \leq b_2 \]
and
\[ \|x_T\|_2 \geq \sqrt{\frac{2b_2}{(1 - \delta)(1 - \rho)}} \geq \frac{2\sqrt{s}}{\lambda_{\min}(A_T^*)A_T(1 - \rho)} \|A^*_n\|_\infty, \]
we have (2.2) for \( t = 0 \). By Lemma 2.2 and \( \rho(q_0) < \rho \) we get \( \rho(q_0) < 1 \). Note that
\[ \|q_0\|_2 = \|A_Tx_T + z\|_2 \geq \|A_Tx_T\|_2 - \|z\|_2 \geq \sqrt{\lambda_{\min}(A_T^*)\|x\|_2 - b_2} \geq \frac{2b_2}{1 - \rho} - b_2 > b_2. \]
Consequently, conditional on \( A, \rho(s_0) < \rho \) implies \( \rho(q_0) < 1 \) and \( \|q_0\| > b_2 \), which ensures that OMP selects a column from \( A_T \) at the initial iteration. Now suppose that at the first \( t \) iterations (\( t < s \)), conditional on \( A, \rho(s_j) < \rho \) implies \( \rho(q_j) < 1 \) and \( \|q_j\| > b_2 \) for every \( j = 0, 1, \ldots, t - 1 \) and that the algorithm selects \( t \) correct indices from \( T \). Denote the set of all selected indices at the current iteration is \( c_t \). Then \( c_t \subset T \). Denote \( u_t = T \setminus c_t \). Since
\[ \|A^*_n\|_\infty = \max_{j \in [n]} |\langle A_j, (I - P_t)z \rangle| \leq \max_{j \in [n]} \|A_j\|_2 \|(I - P_t)z\|_2 \leq b_2 \]
and
\[ \|x_t\|_2 \geq (s - t)^{1/2} \frac{2b_2}{(1 - \delta)(1 - \rho)} \geq \frac{2(s - t)^{1/2}}{\lambda_{\min}(A_T^*)A_T(1 - \rho)} \|A^*_n\|_\infty, \]
we have (2.2), which leads to \( \rho(q_t) < 1 \) according to Lemma 2.2 and \( \rho(s_t) < \rho \). We still need to show that \( \|q_t\|_2 > b_2 \):
\[ \|q_t\|_2 = \|(I - P_t)A_Tx_T + (I - P_t)z\|_2 \geq \|(I - P_t)A_Tx_T\|_2 - \|(I - P_t)z\|_2 \geq \|A_T^*(I - P_t)A_T\|_{\max} \|x_u\|_2 - b_2 \geq \sqrt{\lambda_{\min}(A_T^*(I - P_t)A_T)} \|x_u\|_2 - b_2 \geq \sqrt{\lambda_{\min}(A_T^*)A_T} \|x_u\|_2 - b_2 \geq \frac{2b_2}{1 - \rho} - b_2 > b_2, \]
where we have used Lemma 2.1 at the fourth inequality. Therefore, \( \rho(s_t) < \rho \) implies \( \rho(q_t) < 1 \) and \( \|q_t\|_2 > b_2 \), and OMP will select another true index from \( T \) at this iteration. For the above analysis, by induction, we prove (5.6).
Next, we will focus on bounding $\mathbb{P}(\max_t \rho(s_t) < \rho | A)$, where $\{s_t\}$ is a sequence of $s$ random vectors which lie in the column span of $A_T$ and it is statistically independent from $A_T^c$. Assume that $A$ occurs. For each $t = 0, 1, \ldots, s-1$, we denote $v_t = s_t / \|A_T^* s_t\|_2$. Note that $s_t$ lies in the range of $A_T$. By the basic properties of singular values, we have

$$\|v_t\|_2 = \frac{\|s_t\|_2 \sqrt{1 - \delta}}{\|A_T^* s_t\|_2} \leq \frac{\sqrt{1 - \delta}}{\sigma_s(A_T)} = \frac{\sqrt{1 - \delta}}{\sqrt{\lambda_{\min}(A_T^* A_T)}} \leq 1.$$ 

Since

$$\rho(s_t) = \frac{\|A_T^* s_t\|_\infty}{\|A_T^* s_t\|_2} \leq \frac{\sqrt{s} \max_j |\langle A_j, s_t \rangle|}{\|A_T^* s_t\|_2} = \frac{\sqrt{s} \max_j |\langle A_j, v_t \rangle|}{\sqrt{1 - \delta} \max_j |\langle A_j, v_t \rangle|},$$

we get

$$\mathbb{P}\left(\max_t \rho(s_t) < \rho | A\right) \geq \mathbb{P}\left(\max_{t} \max_{j \in T^c} |\langle A_j, v_t \rangle| < \frac{\rho \sqrt{1 - \delta}}{\sqrt{s}} | A \right)$$

$$= \mathbb{P}\left(\max_{j \in T^c} \max_{t} |\langle A_j, v_t \rangle| < \frac{\rho \sqrt{1 - \delta}}{\sqrt{s}} | A \right)$$

$$\geq \prod_{j \in T^c} \mathbb{P}\left(\max_{t} |\langle A_j, v_t \rangle| < \frac{\rho \sqrt{1 - \delta}}{\sqrt{s}} | A \right)$$

where we have used the independence of the columns of $A_T^c$ at the last inequality. Notice that every column of $A_T^c$ is independent from $\{v_t\}$ and $A$; then by Lemma 5.3, we get

$$\mathbb{P}\left(\max_{t} \rho(s_t) < \rho | A\right) \geq \left(1 - 2se^{-c_2 \rho^2 (1 - \delta) m / s} \right)^{n - s}.$$ (5.7)

It thus follows from (5.3)–(5.7) that

$$\mathbb{P}(E_{s\text{ucc}}) \geq \left(1 - 2se^{-c_2 \rho^2 (1 - \delta) m / s} \right)^{n - s} (1 - (c_1 / \delta)^s \cdot e^{-cm\delta^2})$$

$$\geq 1 - 2s(n - s)e^{-c_2 \rho^2 (1 - \delta) m / s} - (c_1 / \delta)^s \cdot e^{-cm\delta^2},$$

where we have used the inequality $(1 - a)^k \geq 1 - ka$, $k \geq 1$, $a \leq 1$. Note that $s(n - s) \leq n^2 / 4$. We have

$$\mathbb{P}(E_{s\text{ucc}}) \geq 1 - n^2 e^{-c_2 \rho^2 (1 - \delta) m / s} - (c_1 / \delta)^s \cdot e^{-cm\delta^2}. \quad (5.8)$$

By the assumption $m \geq Cs \max\{\rho^{-2}(1 - \delta)^{-1} \log(n\beta^{-1}), \delta^{-2} \log(\delta^{-1} \beta^{-1})\}$, one can show that the failure probability is at most $\beta$.

**Remark 5.5.** In the noiseless $(b_2 = 0)$, Theorem 4.1 was proved in [53].

### 5.3. Proof of Theorem 4.3

The proof follows the lines of Theorem 4.1. Using the same notation (with $b_2$ instead of $b_\infty$), we will develop similar estimations as (5.4)–(5.8), with (5.5) and (5.6) replaced by

$$\mathbb{P}(E_{s\text{ucc}} | A) \geq \mathbb{P}\left(\max_{t} \rho(q_t) < 1 \text{ and } \min_{t} \|A^* q_t\|_\infty > b_\infty | A \right).$$ (5.9)
and
\[
\mathbb{P} \left( \max_t \rho(q_t) < 1 \text{ and } \min_t \| A^* q_t \|_\infty > b_\infty | A \right) \geq \mathbb{P} \left( \max_t \rho(s_t) < \rho | A \right) \geq \mathbb{P} \left( \max_t \rho(s_t) < \rho | A \right) 
\]
respectively. All the estimations except (5.10) can be derived in the same way as that for Theorem 4.1.

Now we focus on proving (5.10). Similarly, we use the induction to prove that conditional on \( A, \rho(q_t) < 1 \) and \( \| A^* q_t \|_\infty > b_\infty \) for each \( t = 0, 1, \ldots, s - 1 \) occurs when \( \rho(s_t) < \rho \) for each \( t = 0, 1, \ldots, s - 1 \). Now suppose that at the first \( t \) iterations \((t < s)\), conditional on \( A, \rho(s_j) < \rho \) implies \( \rho(q_j) < 1 \) and \( \| A^* q_j \| > b_\infty \) for every \( j = 0, 1, \ldots, t - 1 \) and that the algorithm selects \( t \) correct indices from \( T \). Denote the set of all selected indices at the current iteration is \( c_t \). Then \( c_t \subset T \). Denote \( u_t = T \setminus c_t \). Note that
\[
\| P_t z \|^2 = z^* A_{c_t}(A^*_{c_t} A_{c_t})^{-1} A_{c_t} z \leq \frac{1}{\lambda_{\min}(A^*_T A_T)} \| A^*_{c_t} z \|^2 \leq \frac{s b_\infty^2}{1 - \delta}.
\]
Therefore,
\[
\| A^* n_t \|_\infty = \max_j |\langle A_j, (I - P_t) z \rangle| \leq \max_j |\langle A_j, z \rangle| + \max_j |\langle A_j, P_t z \rangle| \\
\leq \left( 1 + \sqrt{\frac{s}{1 - \delta}} \right) b_\infty.
\]
Then we get
\[
\| x_{u_t} \|_2 \geq \frac{2(s - t)^{1/2}}{(1 - \delta)(1 - \rho)} \left( 1 + \sqrt{\frac{s}{1 - \delta}} \right) b_\infty \geq \frac{2(s - t)^{1/2}}{\lambda_{\min}(A^*_T A_T)(1 - \rho)} \| A^* n_t \|_\infty.
\]
Using Lemma 2.2, we get \( \rho(q_t) < 1 \) by \( \rho(s_t) < 1 \). We still need to show that \( \| A^* q_t \|_\infty > b_\infty \):
\[
\| A^* q_t \|_\infty = \| A^* (I - P_t) A_T x_T + A^* (I - P_t) z \|_\infty \\
\geq \| A^* (I - P_t) A_T x_T \|_\infty - \| A^* (I - P_t) z \|_\infty \\
\geq \frac{1}{\sqrt{s - t}} \| A^*_{u_t} (I - P_t) A_{u_t} x_{u_t} \|_2 - \left( 1 + \sqrt{\frac{s}{1 - \delta}} \right) b_\infty \\
\geq \frac{1}{\sqrt{s - t}} \lambda_{\min}(A^*_T A_T) \| x_{u_t} \|_2 - \left( 1 + \sqrt{\frac{s}{1 - \delta}} \right) b_\infty \\
\geq \left( 1 + \sqrt{\frac{s}{1 - \delta}} \right) \left( \frac{2}{1 - \rho} - 1 \right) b_\infty > b_\infty.
\]
Therefore, $\rho(s_t) < \rho$ implies $\rho(q_t) < 1$ and $\|A^*q_t\|_\infty > b_\infty$, and OMP will select a true index from $T$ at this iteration. For the above analysis, by induction, we prove (5.10).

6. Proofs for main results related to random partial Fourier matrices

6.1. Lemmas

The following result shows that a submatrix $A_T$ of $A$ is well conditioned under mild conditions on the number of $m$ and cardinality of $T$; see [28, Theorem 4.1] (also [34, Theorem 3.3]).

**Lemma 6.1.** Let $T \subset \mathbb{R}^n$ be of size $|T| = s$ and let $A \subset \mathbb{C}^{m \times n}$ be a random partial Fourier matrix. Fix $\beta$ and $\delta$ in $(0, 1)$, and choose $m \geq C\delta^{-2}s\log(s/\beta)$. Then with probability at least $1 - \beta$, the minimal and maximal eigenvalues of $A_T^*A_T$ satisfy

$$1 - \delta \leq \lambda_{\min}(A_T^*A_T) \leq \lambda_{\max}(A_T^*A_T) \leq 1 + \delta. \quad (6.1)$$

The following concentration inequality is due to Kunis and Rauhut [34, Lemma 3.2].

**Lemma 6.2.** Let $v$ be supported on $T$ with $|T| = s$ and let $A \subset \mathbb{C}^{m \times n}$ be a random partial Fourier matrix. Then for each $j \notin T$ and $\delta > 0$,

$$P(|\langle A_j, A v \rangle| \geq \delta) \leq 4 \exp \left( -\frac{m\delta^2}{4\|v\|_2^2 + \frac{4}{3\sqrt{2}}\|v\|_1\delta} \right).$$

For any set $T \subset [n]$ with $|T| = s$, denote by $X_T$ the set of all vectors in $\mathbb{R}^n$ that are zero outside of $T$. This is a $s$-dimensional linear space to which we endow the $l_2$ norms.

**Lemma 6.3.** Let $A \subset \mathbb{C}^{m \times n}$ be a random partial Fourier matrix. For any set $T$ with cardinality less than $s$ and any $\delta > 0$, we have

$$\|A_T^*A v\|_\infty \leq \delta \|v\|_2 \quad \text{for all } v \in X_T \quad (6.2)$$

with probability exceeding

$$1 - 4(n-s) \cdot 5^s \exp \left( -\frac{m\delta^2}{16 + \frac{8}{3\sqrt{2}}\sqrt{s}\delta} \right). \quad (6.3)$$

In particular, if $m \geq C\delta^{-2}s(s + \log(n-s) + \log \beta^{-1})$ for some fixed $\beta \in (0, 1)$, then with probability exceeding $1 - \beta$,

$$\|A_T^*A v\|_\infty \leq \frac{\delta}{\sqrt{s}} \|v\|_2 \quad \text{for all } v \in X_T. \quad (6.4)$$

**Proof.** Note that $A_T^*A_T$ is linear, thus it suffices to prove (6.2) in the case $\|v\|_2 = 1$. We first choose a finite $(1/2)$-covering of the unit sphere in $X_T$, i.e., a set of points $Q \subset X_T$, with $\|q\|_2 = 1$ for all $q \in Q$, such that for all $v \in X_T$, $\|v\|_2 = 1$, we have

$$\min_{q \in Q} \|v - q\|_2 \leq 1/2.$$
According to [36, Lemma 2.2], there exists such a $Q$ with $|Q| \leq 5^s$. Using Lemma 6.2, one gets that
\[
P\left(\max_{v \in Q} \|A^*_T c A v\|_\infty \geq \frac{\delta}{2}\right) \leq \sum_{q \in Q} \sum_{j \notin T} \mathbb{P}\left(|\langle A_j, A q \rangle| \geq \frac{\delta}{2}\right)
\leq 5^s \cdot (n - s) \cdot \exp\left(-\frac{mt^2}{16 + \frac{8}{3\sqrt{2}} \|q\|_1 \delta}\right)
\leq 4(n - s) \cdot 5^s \exp\left(-\frac{m\delta^2}{16 + \frac{8}{3\sqrt{2}} \sqrt{s} \delta}\right),
\]
where at the last inequality we have used $\|q\|_1 \leq \sqrt{s} \|q\|_2$. It thus follows that with probability exceeding (6.3), we have
\[
\|A^*_T c A q\|_\infty \leq \frac{\delta}{2} \|q\|_2 \quad \text{for all } q \in Q.
\]

Now define $B$ as the smallest number such that
\[
\|A^*_T c A v\|_\infty \leq B \|v\|_2 \quad \text{for all } v \in X_T.
\]
Our goal is to show that $B \leq \delta$. Recall that for all $v \in X_T$ with $\|v\|_2 = 1$ one can choose a $q \in Q$ such that $\|q - v\| \leq 1/2$ and get that
\[
\|A^*_T c A v\|_\infty \leq \|A^*_T c A (v - q)\|_\infty + \|A^*_T c A q\|_\infty \leq B/2 + \delta/2.
\]
It thus follows from the definition of $B$ that $B \leq B/2 + \delta/2$. Therefore, $B \leq \delta$, which leads to the result. \(\square\)

6.2. Proof of Theorem 4.10

We begin with a few assumptions. Let $T = \text{supp}(x)$. We assume (6.1) and (6.4) hold for such $T$ and constant $\delta$ in the rest of the proof. Using Lemmas 6.2 and 6.3, one can conclude that this occurs with probability at least $1 - \beta$. Under these assumptions, we will show that OMP with the stopping rule $\|r_t\|_2 \leq b_2$ can recover the support $T$ in $s$ iterations. For this, we define a ratio
\[
\rho(A, T) = \max_{v \in X_T, v \neq 0} \frac{\|A^*_T v\|_\infty}{\|A^*_T A v\|_\infty}.
\]
Note that for all $v \in X_T$, by (6.1),
\[
\|A^*_T A v\|_\infty \geq \|A^*_T A T v\|_2 / \sqrt{s} \geq \lambda_{\min}(A^*_T A T) \|v_T\|_2 / \sqrt{s} \geq (1 - \delta) \|v\|_2 / \sqrt{s}.
\]
It thus follows from the above inequality and (6.4) that
\[
\rho(A, T) \leq \frac{\delta}{1 - \delta} < 1.
\] (6.5)

Assume that at the first $t(t < s)$ iterations, OMP selects $t$ correct indices, that is, $\Omega_t \subset T$. Denote $u_t = T \setminus \Omega_t$. We will prove that $\rho(r_t) < 1$ and $\|r_t\|_2 > b_2$, where $\rho(r_t)$ is defined as
in (5.2). Consequently, OMP selects another true index from $T$ at this current iteration. Since $\|z\|_2 \leq b_2$, we have
\[
\|A^*(I - P_t)z\|_\infty \leq \max_{j \in [r]} \|A_j\|_2 \|I - P_t\|_2 \|z\|_2 \leq \|z\|_2 \leq b_2.
\]

It thus follows that
\[
x_{u_t} \geq \frac{2(s - t)^{1/2}b_2}{1 - 2\delta} \geq \frac{2(s - t)^{1/2}b_2}{(1 - \delta)(1 - \rho(A, T))} \geq \frac{2(s - t)^{1/2}\|A^*(I - P_t)z\|_\infty}{\lambda_{\min}(A_T^*A_T)(1 - \rho(A, T))}.
\]

By the definition of $\rho(A, T)$ and Lemma 2.2, one has $\rho(r_t) < 1$. It remains to be shown that $\|r_t\|_\infty > b_\infty$:
\[
\|r_t\|_2 = \|(I - P_t)A_Tx_T + (I - P_t)z\|_2 \geq \|(I - P_t)A_Tx_T\|_2 - \|(I - P_t)z\|_2
\]
\[
\geq \|(I - P_t)A_{u_t}x_{u_t}\|_2 - b_2 \geq \sqrt{\lambda_{\min}(A_{u_t}^*(I - P_t)A_{u_t})}\|x_{u_t}\|_2 - b_2
\]
\[
\geq \sqrt{\lambda_{\min}(A_T^*A_T)}\|x_{u_t}\|_2 - b_2 \geq \frac{2b_2}{1 - \rho(A, T)} - b_2 > b_2,
\]
where we have used Lemma 2.1 at the fourth inequality. When all the $s$ correct indices are selected, the algorithm will stop since $\|r_s\|_2 = \|(I - P_s)z\|_2 \leq \|z\|_2 \leq b_2$.

6.3. Proof of Theorem 4.12

The proof is similar as that of Theorem 4.10. Similarly, one can prove that (6.5) holds with probability at least $1 - \beta$. Assume that (6.5) holds. We will show that OMP with the stopping rule $\|A^*r_t\|_2 \leq b_\infty$ can return the support $T$.

Assume that at the first $t(t < s)$ iterations, OMP selects $t$ correct indices. We will prove that $\rho(r_t) < 1$ and $\|A^*r_t\|_\infty > b_\infty$. Consequently, OMP selects another true index from $T$ at this current iteration. Note that by Lemma 2.1, we have
\[
\|P_tz\|_2^2 = z^*A_{\Omega_t}(A_{\Omega_t}^*A_{\Omega_t})^{-1}A_{\Omega_t}^*z \leq \frac{1}{\lambda_{\min}(A_T^*A_T)}\|A_{\Omega_t}^*z\|_2^2 \leq \frac{sb_\infty^2}{1 - \delta}.
\]

Thus we get
\[
\|A^*(I - P_t)z\|_\infty \leq \|A^*z\|_\infty + \|A^*P_tz\|_\infty \leq b_\infty + \sqrt{\frac{s}{1 - \delta}}b_\infty.
\]

Combining this inequality with
\[
x_{u_t} \geq \left(1 + \sqrt{\frac{s}{1 - \delta}}\right)\frac{2(s - t)^{1/2}b_\infty}{1 - 2\delta}
\]
\[
\geq \left(1 + \sqrt{\frac{s}{1 - \delta}}\right)\frac{2(s - t)^{1/2}b_\infty}{\lambda_{\min}(A_T^*A_T)(1 - \rho(A, T))},
\]

one gets (2.2). From that definition of $\rho(A, T)$ and using Lemma 2.2, one can prove that $\rho(r_t) < 1$. It remains to be shown that $\|A^*r_t\|_\infty > b_\infty$:
\[
\|A^*r_t\|_\infty \geq \|A^*(I - P_t)A_Tx\|_\infty - \|A^*(I - P_t)z\|_\infty
\]
\[
\geq (s - t)^{-1/2}\|A_{u_t}^*(I - P_t)A_{u_t}x_{u_t}\|_2 - \|A^*(I - P_t)z\|_\infty
\]
\[ \geq (s-t)^{-1/2}(1-\delta)\|x_u\|_2 - b_\infty \left(1 + \sqrt{\frac{s}{1-\delta}}\right) \]
\[ \geq b_\infty \left(1 + \sqrt{\frac{s}{1-\delta}}\right) \frac{1}{1-2\delta} > b_\infty, \]

where we have used Lemma 2.1 at the third inequality.

**Acknowledgments**

We would like to thank the two reviewers and the Editor for their valuable comments which improved the presentation of this paper.

**References**

