

# Coherence-Based Performance Guarantees of Orthogonal Matching Pursuit

Yuejie Chi<sup>†</sup> and Robert Calderbank<sup>‡</sup>

<sup>†</sup>Department of Electrical and Computer Engineering, The Ohio State University, USA

<sup>‡</sup>Department of Electrical and Computer Engineering, Duke University, USA

Email: chi@ece.osu.edu, robert.calderbank@duke.edu

**Abstract**—In this paper, we present coherence-based performance guarantees of Orthogonal Matching Pursuit (OMP) for both support recovery and signal reconstruction of sparse signals when the measurements are corrupted by noise. In particular, two variants of OMP either with known sparsity level or with a stopping rule are analyzed. It is shown that if the measurement matrix  $X \in \mathbb{C}^{n \times p}$  satisfies the strong coherence property, then with  $n \gtrsim \mathcal{O}(k \log p)$ , OMP will recover a  $k$ -sparse signal with high probability. In particular, the performance guarantees obtained here separate the properties required of the measurement matrix from the properties required of the signal, which depends critically on the minimum signal to noise ratio rather than the power profiles of the signal. We also provide performance guarantees for partial support recovery. Comparisons are given with other performance guarantees for OMP using worst-case analysis and the sorted one step thresholding algorithm.

**Index Terms**—Compressive Sensing (CS), Orthogonal Matching Pursuit (OMP), worst-case coherence, average coherence, support recovery, signal reconstruction

## I. INTRODUCTION

Sparse signal processing is a fundamental task in many applications involving high-dimensional data. Many of the recent advances in Compressive Sensing (CS) [1], [2] have been centered around reconstructing a sparse signal  $\beta \in \mathbb{C}^p$  with a few nonzero entries, from a number of linear measurements that is much smaller than the signal dimension, possibly corrupted by noise  $\eta$ , given as  $y = X\beta + \eta$ , where  $X \in \mathbb{C}^{n \times p}$  is the measurement matrix. Two main classes of algorithms have been successfully applied, one is convex optimization based algorithms such as Basis Pursuit (BP) [3], the other one is greedy pursuit based algorithms such as Orthogonal Matching Pursuit (OMP) [4]. In particular, the latter class is appealing and competitive in practice due to its simplicity and low computational cost [5].

The performance of OMP can be characterized either in a worst-case sense or in an average (probabilistic) sense. Define the worst coherence of the measurement matrix  $X$  as  $\mu = \max_{i \neq j} |\langle x_i, x_j \rangle|$ , where  $x_i$  denotes the  $i$ th unit-norm column of  $X$ . It is shown in [4] that if  $\mu < \frac{1}{2k-1}$ , then OMP recovers any  $k$ -sparse vector  $\beta$  from the noiseless measurement  $y = X\beta$ , and this result is confirmed to be sharp in

[6]. It is further studied in [7] that given the amplitudes of the nonzero entries of  $\beta$  are not too small, OMP recovers the support of the signal from noisy observations. From the Welch bound [8] which gives  $\mu \gtrsim \mathcal{O}(n^{-1/2})$ , in order to recover all  $k$ -sparse vectors, the number of measurements is required to satisfy  $n \gtrsim \mathcal{O}(k^2)$ . It is demonstrated in [9] that, if the Restricted Isometry Property (RIP) of  $X$  of order  $k+1$  is satisfied with constant smaller than  $1/(3\sqrt{k})$ , then OMP recovers any  $k$ -sparse signal, again from the noiseless measurements  $y = X\beta$ . Given a random matrix satisfies the RIP of order  $k$  with constant  $\delta_k$  provided that  $n \gtrsim \mathcal{O}(k \log(p/k)/\delta_k^2)$ , then the number of measurements is required  $n \gtrsim \mathcal{O}(k^2 \log(p/k))$ . These results all suffer from what is called “square-root bottleneck”. Alternatively, instead of aiming to recover all  $k$  sparse signals using OMP, it is shown that for a fixed sparse vector  $\beta$ , a randomly drawn measurement matrix  $X$  from i.i.d. normal entries can recover  $\beta$  with high probability with  $n \gtrsim \mathcal{O}(k \log p)$  measurements from the noiseless measurements  $y = X\beta$ . However, it does not provide a practical way to design or validate the usefulness of a measurement matrix.

In this paper, we aim to use the strong coherence properties proposed in [10] to capture the performance of OMP. The strong coherence properties require that the worst-case coherence of the measurement matrix  $X$  is sufficiently small, and also that the average coherence, defined as  $\nu = \frac{1}{p-1} \max_i |\sum_{j:j \neq i} \langle x_i, x_j \rangle|$ , is small relative to the worst-case coherence. Different from the worst-case sense, we aim to provide conditions on the measurement matrix  $X$  that are easily verifiable in contrast to RIP, and that succeeds in support recovery and reconstruction of sparse signals with high probability in the presence of noise. We show that if  $X$  satisfied the strong coherence property, then with  $n \gtrsim \mathcal{O}(k \log p)$  measurements, OMP recovers a  $k$ -sparse signal with high probability. In particular, the performance of OMP depends on the smallest signal-to-noise ratio  $\text{SNR}_{\min}$  determined by the smallest nonzero entry instead of the relative strengths of the nonzero entries of the signal.

The rest of the paper is organized as follows. Section II introduces two coherence parameters of the measurement matrix and provides a detailed description of OMP. Section III

gives the main theorems on the performance guarantees of OMP for support recovery and signal reconstruction under the strong coherence property. Section IV prepares for the proof and Section V proves the main theorems. Finally we conclude with discussions in Section VI.

## II. TWO FUNDAMENTAL COHERENCE PARAMETERS

Suppose we are given a measurement vector as  $y = X\beta + \eta$ , where  $X$  is an  $n \times p$  unit-norm measurement matrix,  $\beta \in \mathbb{C}^p$  is a  $k$ -sparse vector,  $y \in \mathbb{C}^n$  is the measurement and  $\eta \in \mathbb{C}^n$  is the noise. We define two coherence properties of  $X = [x_1, \dots, x_p]$  below. The first is worst-case coherence:

$$\mu = \max_{i \neq j} |\langle x_i, x_j \rangle| = \|X^H X - I\|_\infty,$$

which captures the correlation between different columns of  $X$ . The second is average coherence:

$$\nu = \frac{1}{p-1} \max_i \left| \sum_{j:j \neq i} \langle x_i, x_j \rangle \right| = \frac{1}{p-1} \|(X^H X - I)1\|_\infty.$$

which captures the average correlation between one column of  $X$  and the remaining columns of  $X$ .

We say a measurement matrix  $X$  satisfies the strong coherence property if the equation below holds:

$$\mu \leq \frac{1}{240 \log p}, \quad \nu \leq \frac{\mu}{\sqrt{n}}. \quad (1)$$

It is known that Gaussian random matrices satisfy the strong coherence property with high probability as long as  $n \gtrsim \mathcal{O}((\log p)^4)$  [11]. Several families of deterministic matrices are also known to satisfy the strong coherence property, including Gabor frames [10], Kerdock code sets [12], and Delsarte-Goethals code sets [12]. Notice that the condition on average coherence  $\nu \leq \frac{\mu}{\sqrt{n}}$  can be achieved with essentially no cost via “wiggling”, i.e. flipping the signs of the columns of  $X$  [11]. The “wiggling” procedure doesn’t change  $\mu$  and  $\|X\|_2$ .

The goal of this paper is to present performance guarantees on OMP for both support recovery and signal reconstruction under the assumption that  $X$  satisfies the strong coherence property, when the measurements are corrupted by noise. In particular, two variants of OMP that differ in the way they terminate the iterations are analyzed, i.e. Algorithm 1 with known sparsity level, and Algorithm 2 with a stopping rule.

## III. PERFORMANCE GUARANTEES OF OMP

We define the minimum-to-average ratio MAR and the  $t$ th-largest-to-average ratio  $\text{LAR}_{(t)}$  of the signal  $\beta$  respectively as

$$\text{MAR} = \frac{|\beta|_{\min}^2}{\|\beta\|_2^2/k}, \quad \text{LAR}_{(t)} = \frac{|\beta|_{(t)}^2}{\|\beta\|_2^2/k}, \quad (2)$$

where  $|\beta|_{(t)}$  is the  $t$ th largest absolute value of  $\beta$ ,  $|\beta|_{\min}$  is the smallest nonzero absolute value of  $\beta$ , and  $k$  is the sparsity

---

### Algorithm 1 OMP with a fixed number of iterations

---

- 1: Input: an  $n \times p$  matrix  $X$ , a vector  $y \in \mathbb{C}^n$ , and a sparsity level  $k$
  - 2: Output: an estimate  $\hat{\mathcal{S}}$  of the true model  $\mathcal{S}$
  - 3: Initialization:  $\hat{\mathcal{S}}_0 :=$  empty set, residual  $r_0 = y$
  - 4: **for**  $t := 1 : k$  **do**
  - 5:    $f := X^H r_{t-1}$
  - 6:    $i := \arg \max_j |f_j|$
  - 7:    $\hat{\mathcal{S}}_t := \hat{\mathcal{S}}_{t-1} \cup \{j\}$
  - 8:    $r_t := y - X_{\hat{\mathcal{S}}_t} (X_{\hat{\mathcal{S}}_t}^H X_{\hat{\mathcal{S}}_t})^{-1} X_{\hat{\mathcal{S}}_t}^H y$
  - 9: **end for**
  - 10:  $\hat{\mathcal{S}} := \hat{\mathcal{S}}_k$
- 

---

### Algorithm 2 OMP with a stopping rule

---

- 1: Input: an  $n \times p$  matrix  $X$ , a vector  $y \in \mathbb{C}^n$ , and a threshold level  $\delta$
  - 2: Output: an estimate  $\hat{\mathcal{S}}$  of the true model  $\mathcal{S}$
  - 3: Initialization:  $\hat{\mathcal{S}}_0 :=$  empty set, residual  $r_0 = y$ , set the iteration counter  $t = 1$
  - 4: **while**  $\|X^H r_{t-1}\|_\infty > \delta$  **do**
  - 5:    $f := X^H r_{t-1}$
  - 6:    $i := \arg \max_j |f_j|$
  - 7:    $\hat{\mathcal{S}}_t := \hat{\mathcal{S}}_{t-1} \cup \{j\}$
  - 8:    $r_t := y - X_{\hat{\mathcal{S}}_t} (X_{\hat{\mathcal{S}}_t}^H X_{\hat{\mathcal{S}}_t})^{-1} X_{\hat{\mathcal{S}}_t}^H y$
  - 9:    $t := t + 1$
  - 10: **end while**
  - 11:  $\hat{\mathcal{S}} := \hat{\mathcal{S}}_{t-1}$
- 

level of  $k$ . The signal-to-noise ratio SNR and minimum signal-to-noise ratio  $\text{SNR}_{\min}$  are defined respectively as

$$\text{SNR} = \frac{\|\beta\|_2^2}{\mathbb{E}\|\eta\|_2^2}, \quad \text{SNR}_{\min} = \frac{|\beta|_{\min}^2}{\mathbb{E}\|\eta\|_2^2/k}. \quad (3)$$

#### A. Performance Guarantee for Support Recovery

We have the following theorem for OMP with knowledge of sparsity level  $k$  in Algorithm 1.

**Theorem 1.** *Suppose  $X$  satisfies the strong coherence property for any  $p \geq 128$ , and  $\eta \sim \mathcal{CN}(0, \sigma^2 I)$ . If the sparsity level of  $\beta$  satisfies*

$$k \leq \min \left\{ \frac{p}{c_2^2 \|X\|_2^2 \log p}, \frac{1}{c_1^2 \mu^2 \log p} \right\} \quad (4)$$

for  $c_1 = 50\sqrt{2}$  and  $c_2 = 104\sqrt{2}$ , and its nonzero entries satisfy

$$|\beta|_{(t+1)} > \frac{2\sigma\sqrt{(1+\alpha)\log p}}{1 - c_1\mu\sqrt{(k-t)\log p}}, \quad (5)$$

or write differently, that is

$$\text{LAR}_{(t+1)} > \frac{4(1+\alpha)}{(1 - c_1\mu\sqrt{(k-t)\log p})^2} \cdot \left( \frac{k \log p}{n \text{SNR}} \right), \quad (6)$$

for  $0 \leq t \leq k-1$  and  $\alpha \geq 1$ , then the OMP algorithm in Algorithm 1 successfully finds the support of  $\beta$  with probability at least  $1 - k(p^\alpha \pi)^{-1} - 2p^{-2 \log 2} - 4p^{-1}$ .

For the OMP algorithm with a stopping rule in Algorithm 2, we have the following theorem.

**Theorem 2.** *If  $X$  satisfies the strong coherence property for any  $p \geq 128$ , and  $\eta \sim \mathcal{CN}(0, \sigma^2 I)$ . If the sparsity level of  $\beta$  satisfies (4) and its nonzero entries satisfy (6) for  $\alpha \geq 1$ , and choose  $\delta = \sigma \sqrt{(1+\alpha) \log p}$ , then the OMP algorithm in Algorithm 2 successfully finds the support of  $\beta$  with probability at least  $1 - (k+1)(p^\alpha \pi)^{-1} - 2p^{-2 \log 2} - 4p^{-1}$  in  $k$  iterations.*

Since  $\text{MAR} \leq \text{LAR}_{(t+1)}$  for all  $0 \leq t \leq k-1$ , we have the following corollary respectively from Theorem 1 and Theorem 2.

**Corollary 3.** *If  $X$  satisfies the strong coherence property for any  $p \geq 128$ , and  $\eta \sim \mathcal{CN}(0, \sigma^2 I)$ . If the sparsity level of  $\beta$  satisfies (4), and it satisfies*

$$\text{MAR} > \frac{4(1+\alpha)}{(1 - c_1 \mu \sqrt{k \log p})^2} \cdot \left( \frac{k \log p}{n \text{SNR}} \right), \quad (7)$$

for  $\alpha \geq 1$ , then the support of  $\beta$  is successfully recovered by OMP with probability at least  $1 - k(p^\alpha \pi)^{-1} - 2p^{-2 \log 2} - 4p^{-1}$  using Algorithm 1, and with probability at least  $1 - (k+1)(p^\alpha \pi)^{-1} - 2p^{-2 \log 2} - 4p^{-1}$  by Algorithm 2 with  $\delta = \sigma \sqrt{(1+\alpha) \log p}$  in  $k$  iterations.

Let  $\theta = c_1 \mu \sqrt{k \log p} \in (0, 1)$ , then (7) implies the sparsity level  $k$  satisfies

$$k < \frac{(1-\theta)^2}{4(1+\alpha)} \cdot \frac{\text{SNR}_{\min}}{\log p}. \quad (8)$$

Combining with (4), we have

$$k < \max_{0 < \theta < 1} \min \left\{ \frac{n(1-\theta)^2 \text{SNR}_{\min}}{4(1+\alpha) \log p}, \frac{\theta^2}{c_1^2 \mu^2 \log p}, \frac{p}{c_2^2 \|X\|_2^2 \log p} \right\}, \quad (9)$$

where the first term is determined by  $\text{SNR}_{\min}$ , which is signal dependent; the second term and the third term are determined by the worst-case coherence and the spectral norm of the measurement matrix  $X$ . If  $X$  is a tight frame,  $\|X\|_2^2 = p/n$ , the third term becomes  $k < \mathcal{O}(n/\log p)$ . From the Welch bound  $\mu$  is lower bounded as  $\mu \gtrsim \mathcal{O}(n^{-1/2})$ , we write the worst-case coherence as  $\mu = c_3 n^{-1/\gamma}$  for some  $c_3 > 0$  and  $\gamma \geq 2$ . Therefore the maximum sparsity level is determined by the second term in (9), yielding  $k \lesssim \mathcal{O}((n/\log p)^{2/\gamma})$ , and when  $\gamma = 2$  this gives  $k \lesssim \mathcal{O}(n/\log p)$ . In particular, the sparsity level  $k$  doesn't depend on the profile of signal strength of  $\beta$ , i.e. MAR of the signal.

We have another corollary on partial recovery.

**Corollary 4.** *If  $X$  satisfies the strong coherence property for any  $p \geq 128$ , and  $\eta \sim \mathcal{CN}(0, \sigma^2 I)$ . If the sparsity level of*

$\beta$  satisfies (4), and its nonzero entries satisfy (6) for  $0 \leq t \leq k'-1 \leq k-1$  and  $\alpha \geq 1$ , then the OMP algorithm in both Algorithm 1 and Algorithm 2 successfully selects  $k'$  entries from the support of  $\beta$  with probability at least  $1 - k'(p^\alpha \pi)^{-1} - 2p^{-2 \log 2} - 4p^{-1}$ .

It is worth noting that it is not necessarily the support of the  $k'$ -largest entries that is recovered from the first  $k'$  iterations. The next corollary provides the condition on detecting the  $k'$ -largest entries.

**Corollary 5.** *If  $X$  satisfies the strong coherence property for any  $p \geq 128$ , and  $\eta \sim \mathcal{CN}(0, \sigma^2 I)$ . If the sparsity level of  $\beta$  satisfies (4), and its nonzero entries satisfy*

$$|\beta|_{(t+1)} > \frac{|\beta|_{(t+2)} + 2\sigma \sqrt{(1+\alpha) \log p}}{1 - c_1 \mu \sqrt{(k-t) \log p}}, \quad (10)$$

for  $0 \leq t \leq k'-1 \leq k-1$  and  $\alpha \geq 1$ , then the OMP algorithm in both Algorithm 1 and Algorithm 2 successfully selects  $k'$  largest entries from the support of  $\beta$  with probability at least  $1 - k'(p^\alpha \pi)^{-1} - 2p^{-2 \log 2} - 4p^{-1}$ .

### B. Performance Guarantees for Signal Reconstruction

Furthermore, we could reconstruct the amplitude of the signal  $\beta$  by first reconstructing the amplitude on the detected support  $\Pi$  via least-squares estimation as

$$\hat{z} = X_{\Pi}^{\dagger} y,$$

then  $\hat{\beta}$  is obtained by filling in the zero entries. We have the following theorem.

**Theorem 6.** *If  $X$  satisfies the strong coherence property for any  $p \geq 128$ , and  $\eta \sim \mathcal{N}(0, \sigma^2 I)$ . If the sparsity level of  $\beta$  satisfies (4) and its nonzero entries satisfy (6) for  $0 \leq t \leq k-1$  and  $\alpha \geq 1$ , then the  $\ell_2$  norm error of the signal reconstructed by least-squares estimation on the support recovered by the OMP algorithm satisfies*

$$\|\hat{\beta} - \beta\|_2^2 \leq 4(1+\alpha)k\sigma^2 \log p$$

with probability at least  $1 - k(p^\alpha \pi)^{-1} - 2p^{-2 \log 2} - 4p^{-1}$  using Algorithm 1, and with probability at least  $1 - (k+1)(p^\alpha \pi)^{-1} - 2p^{-2 \log 2} - 4p^{-1}$  using Algorithm 2.

### C. Comparison with Other Results

We now compare our bound with the performance guarantee of OMP for support recovery provided in [7], which we have modified slightly for complex Gaussian noise. In order to select exactly the correct support with probability at least  $1 - (k+1)(p^\alpha \pi)^{-1}$  for the OMP Algorithm 2 with the stopping rule  $\delta = \sigma \sqrt{(1+\alpha) \log p}$ , the signal  $\beta$  needs to satisfy

$$\text{MAR} > \frac{4(1+\alpha)}{(1 - (2k-1)\mu)^2} \cdot \left( \frac{k \log p}{n \text{SNR}} \right), \quad (11)$$

therefore the sparsity level of  $\beta$  satisfies

$$k < \max_{0 < \theta < 1} \min \left\{ \frac{n(1-\theta)^2 \text{SNR}_{\min}}{4(1+\alpha) \log p}, \frac{1}{2} + \frac{\theta}{2\mu} \right\}.$$

The first term is the same as that in (9), but the second term gives  $k \sim \mathcal{O}(\mu^{-1})$ , therefore  $k \lesssim \mathcal{O}(n^{1/\gamma})$ . We achieved a much tighter bound (9) by sacrificing the probability of success to  $1 - (k+1)(p^\alpha \pi)^{-1} - 2p^{-2 \log 2} - 4p^{-1}$ .

We also compare with the performance guarantee of the Sorted One Step Thresholding (SOST) algorithm analyzed in [10], which outputs the index set of the  $k$ -largest entries in absolute values of  $f = X^H y$  from line 5 in Algorithm 1. By rephrasing Theorem 4 in [10], in order to select the correct support with probability at least  $1 - 6p^{-1}$ , the sparsity level of  $\beta$  satisfies

$$k < \max_{0 < \theta < 1} \min \left\{ \frac{n(1-\theta)^2 \text{SNR}_{\min}}{16 \log p}, \frac{\theta^2}{800\mu^2 \log p} \cdot \frac{1}{\text{MAR}}, \frac{n}{2 \log p} \right\}. \quad (12)$$

Compare to (9), we see that when  $X$  is not a tight frame, it is the third term that degrades the performance guarantees. On the other hand, the SOST algorithm performs poorly when the MAR is of the signal is much smaller than 1, as seen from the second term in (12).

#### IV. PREPARATIONS FOR PROOF

##### A. Statistical Orthogonality Condition (StOC)

The Statistical Orthogonality Condition (StOC) for a measurement matrix  $X$  is first introduced in [10].

**Definition 1.** Let  $\bar{\Pi} = (\pi_1, \dots, \pi_p)$  be a random permutation of  $\{1, \dots, p\}$ , and define  $\Pi = (\pi_1, \dots, \pi_k)$  and  $\Pi^c = (\pi_{k+1}, \dots, \pi_p)$  for any  $k \leq p$ . Then the matrix  $X$  is said to satisfy the  $(k, \epsilon, \delta)$ -StOC, if there exist  $\epsilon, \delta \in [0, 1)$  such that the inequalities

$$\|(X_{\bar{\Pi}}^H X_{\bar{\Pi}} - I)z\|_{\infty} \leq \epsilon \|z\|_2, \quad (13)$$

$$\|X_{\Pi^c}^H X_{\Pi} z\|_{\infty} \leq \epsilon \|z\|_2, \quad (14)$$

hold for every fixed  $z \in \mathbb{C}^k$  with probability exceeding  $1 - \delta$ , with respect to  $\bar{\Pi}$ .

We have the following proposition rephrased from [10] stating that the StOC is satisfied with high probability if  $X$  satisfies the strong coherence property.

**Proposition 1** ([10]). *If the measurement matrix  $X$  satisfies the strong coherence property, then it satisfies  $(k, \epsilon, \delta)$ -StOC for  $k \leq n/(2 \log p)$ , with  $\epsilon = 10\mu\sqrt{2 \log p}$  and  $\delta \leq 4p^{-1}$ .*

If (13) and (14) hold for a realization of permutation  $\bar{\Pi}$ , then for  $t \leq k$ , let  $\Pi_t = (\pi_1, \dots, \pi_t)$  and  $\Pi_t^c = (\pi_{t+1}, \dots, \pi_k)$ , if (13) and (14) hold for every  $z \in \mathbb{C}^k$ , so

that  $\Pi_t \cup \Pi_t^c = \Pi$  and  $\Pi_t \cap \Pi_t^c = \emptyset$ . For every  $z \in \mathbb{C}^t$ , we have

$$\left\| \begin{bmatrix} X_{\Pi_t}^H X_{\Pi_t} - I_t & X_{\Pi_t}^H X_{\Pi_t^c} \\ X_{\Pi_t^c}^H X_{\Pi_t} & X_{\Pi_t^c}^H X_{\Pi_t^c} - I_{k-t} \end{bmatrix} \begin{bmatrix} z \\ 0 \end{bmatrix} \right\|_{\infty} \leq \epsilon \|z\|_2,$$

from (13), therefore

$$\|(X_{\Pi_t}^H X_{\Pi_t} - I_t)z\|_{\infty} \leq \epsilon \|z\|_2,$$

$$\|X_{\Pi_t^c}^H X_{\Pi_t} z\|_{\infty} \leq \epsilon \|z\|_2.$$

Moreover, from (14) we have

$$\|X_{\Pi^c}^H X_{\Pi} z\|_{\infty} = \left\| \begin{bmatrix} X_{\Pi^c}^H X_{\Pi_t} & X_{\Pi^c}^H X_{\Pi_t^c} \end{bmatrix} \begin{bmatrix} z \\ 0 \end{bmatrix} \right\|_{\infty} \leq \epsilon \|z\|_2.$$

##### B. Conditioning of random submatrices

We need the following proposition that shows a random submatrix of  $X$  is well-conditioned with high probability, which is essentially due to Tropp [13], and first presented in the form below by Candès and Plan [14].

**Proposition 2** ([13], [14]). *Let  $\bar{\Pi} = (\pi_1, \dots, \pi_p)$  be a random permutation of  $\{1, \dots, p\}$ , and define  $\Pi = (\pi_1, \dots, \pi_k)$  for any  $k \leq p$ . Then for  $q = 2 \log p$  and  $k \leq p/(4\|X\|_2^2)$ , we have*

$$\begin{aligned} & (\mathbb{E} [\|X_{\bar{\Pi}}^H X_{\bar{\Pi}} - I\|_2^q])^{1/q} \\ & \leq 2^{1/q} \left( 30\mu \log p + 13\sqrt{\frac{2k\|X\|_2^2 \log p}{p}} \right). \end{aligned} \quad (15)$$

with respect to the random permutation  $\bar{\Pi}$ .

The following proposition [14] states a probabilistic bound on the extreme singular values of a random submatrix of  $X$ , by applying Markov's inequality

$$\Pr(\|X_{\bar{\Pi}}^H X_{\bar{\Pi}} - I\|_2 \geq 1/2) \leq 2^q \mathbb{E} [\|X_{\bar{\Pi}}^H X_{\bar{\Pi}} - I\|_2^q]$$

to Proposition 2.

**Proposition 3** ([14]). *Let  $\bar{\Pi} = (\pi_1, \dots, \pi_p)$  be a random permutation of  $\{1, \dots, p\}$ , and define  $\Pi = (\pi_1, \dots, \pi_k)$  for any  $k \leq p$ . Suppose that  $\mu(X) \leq 1/(240 \log p)$  and  $k \leq p/(c_2^2 \|X\|_2^2 \log p)$  for numerical constant  $c_2 = 104\sqrt{2}$ , then we have*

$$\Pr(\|X_{\bar{\Pi}}^H X_{\bar{\Pi}} - I\|_2 \geq 1/2) \leq 2p^{-2 \log 2}.$$

Notice that  $\|X_{\bar{\Pi}}^H X_{\bar{\Pi}} - I\|_2 = \max\{\lambda_{\max}(X_{\bar{\Pi}}^H X_{\bar{\Pi}}) - 1, 1 - \lambda_{\min}(X_{\bar{\Pi}}^H X_{\bar{\Pi}})\}$ , where  $\lambda_{\max}(X_{\bar{\Pi}}^H X_{\bar{\Pi}})$  and  $\lambda_{\min}(X_{\bar{\Pi}}^H X_{\bar{\Pi}})$  are the maximum and minimum eigenvalues of  $X_{\bar{\Pi}}^H X_{\bar{\Pi}}$ , i.e. all the eigenvalues of  $X_{\bar{\Pi}}^H X_{\bar{\Pi}}$  are bounded in  $[1/2, 3/2]$ . If for a realization of permutation  $\bar{\Pi}$ ,  $\|X_{\bar{\Pi}}^H X_{\bar{\Pi}} - I\|_2 \geq 1/2$ , we have

$$\|(X_{\bar{\Pi}}^H X_{\bar{\Pi}})^{-1}\|_2 \leq 2 \quad \text{and} \quad \|X_{\bar{\Pi}}(X_{\bar{\Pi}}^H X_{\bar{\Pi}})^{-1}\|_2 \leq \sqrt{2}.$$

Moreover, for  $t \leq k$  and  $\Pi_t = (\pi_1, \dots, \pi_t)$ , we have

$$\|X_{\Pi_t}^H X_{\Pi_t} - I_t\|_2 \geq 1/2,$$

since the eigenvalues of  $X_{\Pi_t}^H X_{\Pi_t}$  are majorized by the eigenvalues of  $X_{\Pi}^H X_{\Pi}$ .

### C. Correlated Gaussian Noise

Let  $P \in \mathbb{C}^{n \times n}$  be a projection matrix such that  $P^2 = P$ . Since  $\eta \sim \mathcal{CN}(0, \sigma^2 I_n)$  is i.i.d. complex Gaussian noise,  $X^H P \eta \sim \mathcal{CN}(0, \sigma^2 X^H P X)$  is also Gaussian distributed, but is correlated with covariance matrix  $\sigma^2 X^H P X$ . We want to bound  $\Pr(\|X^H P \eta\|_{\infty} \geq \tau)$  for some  $\tau > 0$ . First, we need the Sidak's lemma [15] below.

**Lemma 4** (Sidak's lemma). *Let  $[X_1, \dots, X_n]$  be a vector of random multivariate normal variables with zero means, arbitrary variances  $\sigma_1^2, \dots, \sigma_n^2$  and an arbitrary correlation matrix. Then, for any positive numbers  $c_1, \dots, c_n$ , we have*

$$\Pr(|X_1| \leq c_1, \dots, |X_n| \leq c_n) \geq \prod_{i=1}^n \Pr(|X_i| \leq c_i).$$

Since  $X^H P \eta \sim \mathcal{CN}(0, \sigma^2 X^H P X)$ , then each  $x_i^H P \eta \sim \mathcal{CN}(0, \sigma_i^2)$ , where  $\sigma_i^2 = \sigma^2 x_i^H P x_i \leq \sigma^2$ . Then

$$\Pr(|x_i^H P \eta| \leq \tau) = 1 - \frac{1}{\pi} e^{-\tau^2/\sigma_i^2} \geq 1 - \frac{1}{\pi} e^{-\tau^2/\sigma^2}.$$

Following Sidak's lemma, for  $\tau > 0$  we have

$$\begin{aligned} \Pr(\|X^H P \eta\|_{\infty} \leq \tau) &\geq \prod_{i=1}^p \Pr(|x_i^H P \eta| \leq \tau) \\ &\geq \left(1 - \frac{1}{\pi} e^{-\tau^2/\sigma^2}\right)^p \\ &\geq 1 - \frac{p}{\pi} e^{-\tau^2/\sigma^2}, \end{aligned}$$

provided the RHS is greater than zero. We have the proposition below.

**Proposition 5.** *Let  $\eta$  be a random vector with i.i.d.  $\mathcal{CN}(0, \sigma^2)$  entries,  $P$  be a projection matrix, and  $X$  be a unit-column matrix, then for  $\tau > 0$  we have*

$$\Pr(\|X^H P \eta\|_{\infty} \leq \tau) \geq 1 - \frac{p}{\pi} e^{-\tau^2/\sigma^2},$$

provided the RHS is greater than zero.

Now let  $\tau = \sigma \sqrt{(1 + \alpha) \log p}$  for  $\alpha \geq 1$ , we have

$$\Pr\{\|X^H P \eta\|_{\infty} \leq \sqrt{\sigma^2(1 + \alpha) \log p}\} \geq 1 - (p^\alpha \pi)^{-1}.$$

## V. PROOF OF MAIN RESULTS

We first write the data vector  $\beta$  as  $\beta = P_{\Pi} z$ , where  $z \in \mathbb{C}^k$  is a deterministic vector, and  $P_{\Pi} \in \mathbb{R}^{p \times k}$  is a partial identity matrix composed of columns indexed by  $\Pi$ . Then the measurement vector can be written as

$$y = X \beta + \eta = X P_{\Pi} z + \eta = X_{\Pi} z + \eta,$$

where  $X_{\Pi}$  denotes the submatrix of  $X$  composed of columns indexed by  $\Pi$ .

We note that in OMP, the residual  $r_t$ ,  $t = 0, \dots, k-1$  is orthogonal to the selected columns in previous iterations, so in each iteration a new column will be selected. Define a subset  $\Pi_t$  which contains  $t$  variables that are selected at the  $t$ th iteration. Then  $P_t = X_{\Pi_t} (X_{\Pi_t}^H X_{\Pi_t})^{-1} X_{\Pi_t}^H$  is the projection matrix onto the linear subspace spanned by the columns of  $X_{\Pi_t}$ , and we assume  $P_0 = 0$ .

We want to prove  $\Pi_t \subset \Pi$  by induction. First at  $t = 0$ ,  $\Pi_t = \emptyset \subset \Pi$ . Assume at iteration  $t$ ,  $\Pi_t \subset \Pi$ , then the residual  $r_t$  can be written as

$$r_t = (I - P_t)y = (I - P_t)X_{\Pi} z + (I - P_t)\eta \triangleq s_t + n_t.$$

Let  $M_{\Pi}^t = \|X_{\Pi}^H s_t\|_{\infty}$ ,  $M_{\Pi^c}^t = \|X_{\Pi^c}^H s_t\|_{\infty}$  and  $N_t = \|X^H n_t\|_{\infty}$ , then a sufficient condition for  $\Pi_{t+1} \subset \Pi$ , i.e. for OMP to select a correct variable at the next iteration is that

$$M_{\Pi}^t - M_{\Pi^c}^t > 2N_t, \quad (16)$$

since

$$\|X_{\Pi}^H r_t\|_{\infty} \geq M_{\Pi}^t - N_t > M_{\Pi^c}^t + N_t \geq \|X_{\Pi^c}^H r_t\|_{\infty}.$$

### A. Bounding $M_{\Pi}^t$ and $M_{\Pi^c}^t$

Define the event

$$\mathcal{G}_1 = \{X \text{ satisfies the } (k, \epsilon, \delta)\text{-StOC}\},$$

that happens with probability at least  $1 - 4p^{-1}$  with respect to  $\bar{\Pi}$  from Proposition 1, and the event

$$\mathcal{G}_2 = \{\|X_{\Pi}^H X_{\Pi} - I\|_2 \leq 1/2\},$$

which happens at least  $1 - 2p^{-2 \log 2}$  with respect to  $\bar{\Pi}$  from Proposition 3. Let the event  $\mathcal{G} = \mathcal{G}_1 \cap \mathcal{G}_2$ . From the above discussions the event  $\mathcal{G}$  holds with probability at least  $1 - 4p^{-1} - 2p^{-2 \log 2}$  with respect to  $\bar{\Pi}$ .

Now we bound  $M_{\Pi}^t$  and  $M_{\Pi^c}^t$  under the event  $\mathcal{G}$ . Let  $\Pi_t^c = \Pi \setminus \Pi_t$  be the set of yet to be selected indices of the support of  $\beta$ , and  $\beta_{\Pi_t^c} = z_{\Pi_t^c}$  be the corresponding coefficients. Since  $(I - P_t)X_{\Pi} z \in \mathcal{R}(X_{\Pi_t^c})$  belongs to the linear subspace spanned by the columns of  $X_{\Pi_t^c}$ , we can find a vector  $w$  of dimension  $(k - t)$  such that  $X_{\Pi_t^c} w = (I - P_t)X_{\Pi} z$ , where the vector  $w$  can be written as

$$\begin{aligned} w &= (X_{\Pi_t^c}^H X_{\Pi_t^c})^{-1} X_{\Pi_t^c}^H (I - P_t) X_{\Pi} z \\ &= z_{\Pi_t^c} - (X_{\Pi_t^c}^H X_{\Pi_t^c})^{-1} X_{\Pi_t^c}^H P_t X_{\Pi} z. \end{aligned}$$

We need the following lemma on eigenvalue majorization from [7].

**Lemma 6** ([7]). *The minimum and maximum eigenvalues of  $X_{\Pi_t^c}^H (I - P_t) X_{\Pi_t^c}$  are bounded as*

$$\begin{aligned} \lambda_{\min}(X_{\Pi_t^c}^H (I - P_t) X_{\Pi_t^c}) &\geq \lambda_{\min}(X_{\Pi}^H X_{\Pi}), \\ \lambda_{\max}(X_{\Pi_t^c}^H (I - P_t) X_{\Pi_t^c}) &\leq \lambda_{\max}(X_{\Pi}^H X_{\Pi}). \end{aligned}$$

The readers are referred to [7] for the proof. Since we have

$$\|w\|_2 \leq \|(X_{\Pi_t^c}^H X_{\Pi_t^c})^{-1}\|_2 \|X_{\Pi_t^c}^H (I - P_t) X_{\Pi_t^c} z_{\Pi_t^c}\|_2 \quad (17)$$

$$\leq 2\|X_{\Pi}^H X_{\Pi}\|_2 \|z_{\Pi_t^c}\|_2 \leq 3\|z_{\Pi_t^c}\|_2, \quad (18)$$

where (17) follows from Lemma 6, and (18) follows from Proposition 3. Also,

$$\begin{aligned} \|X_{\Pi_t^c}^H P_t X_{\Pi_t^c} z_{\Pi_t^c}\|_\infty &= \|X_{\Pi_t^c}^H X_{\Pi_t} (X_{\Pi_t}^H X_{\Pi_t})^{-1} X_{\Pi_t}^H X_{\Pi_t^c} z_{\Pi_t^c}\|_\infty \\ &\leq \epsilon \|(X_{\Pi_t}^H X_{\Pi_t})^{-1} X_{\Pi_t}^H X_{\Pi_t^c} z_{\Pi_t^c}\|_2 \\ &\leq \epsilon \|(X_{\Pi_t}^H X_{\Pi_t})^{-1}\|_2 \|X_{\Pi_t}^H X_{\Pi_t^c}\|_2 \|z_{\Pi_t^c}\|_2 \\ &\leq \epsilon \|z_{\Pi_t^c}\|_2, \end{aligned}$$

therefore  $M_{\Pi}^t$  can be bounded as

$$\begin{aligned} M_{\Pi}^t &= \|X_{\Pi_t^c}^H X_{\Pi_t^c} z_{\Pi_t^c} - X_{\Pi_t^c}^H P_t X_{\Pi_t^c} z_{\Pi_t^c}\|_\infty \\ &\geq \|z_{\Pi_t^c}\|_\infty - \|(X_{\Pi_t^c}^H X_{\Pi_t^c} - I) z_{\Pi_t^c}\|_\infty - \|X_{\Pi_t^c}^H P_t X_{\Pi_t^c} z_{\Pi_t^c}\|_\infty \\ &\geq \|z_{\Pi_t^c}\|_\infty - 2\epsilon \|z_{\Pi_t^c}\|_2. \end{aligned} \quad (19)$$

where (19) follows from (13). Next,  $M_{\Pi^c}^t$  can be bounded as

$$\begin{aligned} M_{\Pi^c}^t &= \|X_{\Pi^c}^H (I - P_t) X_{\Pi} z\|_\infty \\ &= \|X_{\Pi^c}^H X_{\Pi_t^c} w\|_\infty \\ &\leq \epsilon \|w\|_2 \leq 3\epsilon \|z_{\Pi_t^c}\|_2. \end{aligned} \quad (20)$$

where (20) follows from (14).

### B. Bounding $N_t$

Conditioned on the event  $\mathcal{G}$ , for each  $P_t$ , since  $I - P_t$  is also a projection matrix, define the event

$$\mathcal{H}_t = \{N_t \leq \sigma \sqrt{(1 + \alpha) \log p}\}, \quad t = 0, \dots, k-1. \quad (21)$$

Then from Proposition 5, it happens with probability at least  $1 - (p^\alpha \pi)^{-1}$  with respect to  $\eta$ . We further define the event  $\mathcal{H} = \bigcap_{t=0}^{k-1} \mathcal{H}_t$ , then from the union bound  $\Pr(\mathcal{H}|\mathcal{G}) = \Pr(\mathcal{H}) \geq 1 - k(p^\alpha \pi)^{-1}$ . Similarly, for the event  $\mathcal{H}' = \bigcap_{t=0}^k \mathcal{H}_t$ , then from the union bound  $\Pr(\mathcal{H}'|\mathcal{G}) = \Pr(\mathcal{H}') \geq 1 - (k+1)(p^\alpha \pi)^{-1}$ .

### C. Proof of Theorem 1 and Theorem 2

Define the event  $\mathcal{I} = \mathcal{G} \cap \mathcal{H}$ , from the above discussions we have  $\Pr(\mathcal{I}) \geq 1 - k(p^\alpha \pi)^{-1} - 2p^{-2 \log 2} - 4p^{-1}$ . Now we are ready to analyze the OMP performance under the event  $\mathcal{G}$ . We want to prove  $\Pi_t \subset \Pi$  by induction.

Substituting the bounds (19), (20) and (21) into (16), it is sufficient that at the  $t$ th iteration

$$\|z_{\Pi_t^c}\|_\infty > 5\epsilon \|z_{\Pi_t^c}\|_2 + 2\sigma \sqrt{(1 + \alpha) \log p}. \quad (22)$$

Note that  $\|z_{\Pi_t^c}\|_\infty \geq |\beta|_{(t+1)}$ ,  $\|z_{\Pi_t^c}\|_2 \leq \sqrt{k-t} |\beta|_{(t+1)}$ , (22) is satisfied by

$$|\beta|_{(t+1)} > 5\epsilon \sqrt{k-t} |\beta|_{(t+1)} + 2\sigma \sqrt{(1 + \alpha) \log p}.$$

Since  $k < 1/(c_1^2 \mu^2 \log p)$  and  $\epsilon = 10\mu\sqrt{2 \log p}$ , this is equivalent to the condition in (6) for  $0 \leq t \leq k-1$ , therefore

a correct variable is selected at the  $t$ th iteration,  $\Pi_t \subset \Pi$ . Since the sparsity level of  $\beta$  is  $k$ , the OMP algorithm in Algorithm 1 successfully finds the support of  $\beta$  in  $k$  iterations under the event  $\mathcal{I}$ , and we have proved Theorem 1.

Now we define the event  $\mathcal{I}' = \mathcal{H}' \cap \mathcal{G}$ , where  $\mathcal{I}' \subset \mathcal{I}$  which happens with probability at least  $1 - (k+1)(p^\alpha \pi)^{-1} - 2p^{-2 \log 2} - 4p^{-1}$ . Conditioned on the event  $\mathcal{I}'$ , in order to prove Theorem 2, we need to further show that  $\|X^H r_t\|_\infty > \delta$  for  $0 \leq t \leq k-1$  so that the algorithm doesn't stop early, and  $\|X^H r_k\|_\infty \leq \delta$  so that the algorithm stops at the  $k$ th iteration. While the latter is obvious from the definition of  $\mathcal{H}_k$ , for the first inequality we have

$$\begin{aligned} \|X^H r_t\|_\infty &\geq M_{\Pi}^t - N_t \\ &\geq \|z_{\Pi_t^c}\|_\infty - 2\epsilon \|z_{\Pi_t^c}\|_2 - \sigma \sqrt{(1 + \alpha) \log p} \end{aligned} \quad (23)$$

$$\begin{aligned} &> 3\epsilon \|z_{\Pi_t^c}\|_2 + \sigma \sqrt{(1 + \alpha) \log p} \\ &\geq \sigma \sqrt{(1 + \alpha) \log p} = \delta, \end{aligned} \quad (24)$$

where (23) follows from (19) and (21), and (24) follows from (22).

**Remark:** The proof of Corollary 4 is straightforward by early-terminating the induction procedure at the  $k'$ th iteration.

### D. Proof of Corollary 5

Again we prove by induction. First at  $t = 0$ ,  $\Pi_t = \emptyset \subset \Pi$ . Assume at iteration  $t$ , the OMP algorithm has successfully detected the  $t$  largest entries of  $|\beta|$ . For  $i \in \Pi_t^c$  that corresponds to the  $t+1$ th largest entry of  $|\beta|$ , we have

$$\begin{aligned} |x_i^H r_t| &\geq |z_i| - 2\epsilon \|z_{\Pi_t^c}\|_2 - \sigma \sqrt{(1 + \alpha) \log p}, \\ &= |\beta|_{(t+1)} - 2\epsilon \|z_{\Pi_t^c}\|_2 - \sigma \sqrt{(1 + \alpha) \log p} \end{aligned}$$

from a simple variation of (19). On the other hand, for  $j \in \Pi_t^c$  that corresponds to the rest undetected entries of  $\beta$ , we have

$$\begin{aligned} |x_j^H r_t| &\leq |z_j| + 2\epsilon \|z_{\Pi_t^c}\|_2 + \sigma \sqrt{(1 + \alpha) \log p}, \\ &\leq |\beta|_{(t+2)} + 2\epsilon \|z_{\Pi_t^c}\|_2 + \sigma \sqrt{(1 + \alpha) \log p}. \end{aligned}$$

To detect the  $t+1$ th largest entries it is sufficient to have

$$|\beta|_{(t+1)} - |\beta|_{(t+2)} \geq 4\epsilon \sqrt{k-t} |\beta|_{(t+1)} + 2\sigma \sqrt{(1 + \alpha) \log p}.$$

This is satisfied when (10) by simply plugging it into the above equation.

### E. Proof of Theorem 6

Conditioned on the event that the support is successfully recovered, since

$$\hat{z} = X_{\Pi}^\dagger y = X_{\Pi}^\dagger (X_{\Pi} z + \eta) = z + (X_{\Pi}^H X_{\Pi})^{-1} X_{\Pi}^T \eta,$$

where  $z = \beta_{\Pi}$  is the non-zero part of  $\beta$ , and

$$\begin{aligned} &\|(X_{\Pi}^H X_{\Pi})^{-1} X_{\Pi}^H \eta\|_2^2 \\ &\leq \|(X_{\Pi}^H X_{\Pi})^{-1}\|_2^2 \|X_{\Pi}^H \eta\|_2^2 \\ &\leq 4k \|X_{\Pi}^H \eta\|_\infty^2 \leq 4(1 + \alpha) k \sigma^2 \log p, \end{aligned} \quad (25)$$

it follows that

$$\|\hat{\beta} - \beta\|_2^2 = \|\hat{z} - z\|_2^2 \leq 4(1 + \alpha)k\sigma^2 \log p.$$

## VI. CONCLUSION

In this paper, we provide coherence-based performance guarantees of two variants of OMP for both support recovery and signal reconstruction of sparse signals when the measurements are corrupted by noise. It is shown that if  $X$  satisfied the strong coherence property, then with  $n \gtrsim \mathcal{O}(k \log p)$ , OMP recovers a  $k$ -sparse signal with high probability. In particular, the guarantees obtained here separate the properties required of the measurement matrix from the properties required of the signal. The resilience of OMP to variability in relative strength of the entries of the signal might be an advantage in applications like multi-user detection in wireless communications because it makes power control less critical [16], [17].

## ACKNOWLEDGEMENTS

The work of Y. Chi and R. Calderbank was supported in part by NSF under Grants NSF CCF-0915299 and NSF CCF-1017431.

## REFERENCES

- [1] D. Donoho, "Compressed sensing," *IEEE Trans. Info. Theory*, vol. 52, no. 4, pp. 1289–1306, Apr. 2006.
- [2] E. Candés and T. Tao, "Decoding by linear programming," *IEEE Transactions on Information Theory*, vol. 51, no. 12, pp. 4203–4215, Dec. 2005.
- [3] S. S. Chen, D. L. Donoho, and M. A. Saunders, "Atomic decomposition by basis pursuit," *SIAM Rev.*, vol. 43, no. 1, pp. 129–159, Jan. 2001.
- [4] J. A. Tropp, "Greed is good: algorithmic results for sparse approximation," *IEEE Transactions on Information Theory*, vol. 50, no. 10, pp. 2231–2242, 2004.
- [5] S. Kunis and H. Rauhut, "Random sampling of sparse trigonometric polynomials, ii. orthogonal matching pursuit versus basis pursuit," *Foundations of Computational Mathematics*, vol. 8, no. 6, pp. 737–763, 2007.
- [6] T. Cai, L. Wang, and G. Xu, "Stable recovery of sparse signals and an oracle inequality," *IEEE Trans. Info. Theory*, vol. 56, no. 7, pp. 3516–3522, Jul. 2010.
- [7] T. Cai and L. Wang, "Orthogonal matching pursuit for sparse signal recovery with noise," *IEEE Transactions on Information Theory*, vol. 57, no. 7, pp. 1–26, 2011.
- [8] L. R. Welch, "Lower bounds on the maximum cross correlation of signals," *IEEE Trans. on Info. Theory*, vol. 20, no. 3, pp. 397–399, May 1974.
- [9] M. A. Davenport and M. B. Wakin, "Analysis of orthogonal matching pursuit using the restricted isometry property," *IEEE Trans. Info. Theory*, vol. 56, no. 9, pp. 4395–4401, Sep. 2010.
- [10] W. U. Bajwa, R. Calderbank, and S. Jafarpour, "Why gabor frames? two fundamental measures of coherence and their role in model selection," *J. Commun. Netw.*, vol. 12, no. 4, pp. 289–307, Aug. 2010.
- [11] W. Bajwa, R. Calderbank, and D. Mixon, "Two are better than one: Fundamental parameters of frame coherence," *Appl. Comput. Harmon. Anal.*, p. to appear, 2012.
- [12] R. Calderbank and S. Jafarpour, "Reed muller sensing matrices and the lasso," in *Proceedings of the 6th international conference on Sequences and their applications*, 2010, pp. 442–463.
- [13] J. A. Tropp, "On the conditioning of random subdictionaries," *Applied and Computational Harmonic Analysis*, vol. 25, no. 1, pp. 1–24, 2008.
- [14] E. J. Candés and Y. Plan, "Near-ideal model selection by  $\ell_1$  minimization," *Annals of Statistics*, vol. 37, no. 5A, pp. 2145–2177, 2009.
- [15] Z. Sidak, "Rectangular confidence regions for the means of multivariate normal distributions," *J. of Amer. Stat. Assn.*, vol. 12, no. 4, pp. 626–633, Jun. 1967.
- [16] Y. Xie, Y. Chi, L. Applebaum, and R. Calderbank, "Compressive demodulation of mutually interfering signals," in *2012 Statistical Signal Processing Workshop*, Ann Arbor, MI, Aug. 2012.
- [17] Y. Chi, Y. Xie, and R. Calderbank, "Compressive demodulation of mutually interfering signals," *Technical Report*, Aug. 2012.