Mismatch and resolution in compressive imaging

Albert Fannjiang*   Wenjing Liao

Department of Mathematics, University of California, Davis, CA 95616-8633.

ABSTRACT

Highly coherent sensing matrices arise in discretization of continuum problems such as radar and medical imaging when the grid spacing is below the Rayleigh threshold as well as in using highly coherent, redundant dictionaries as sparsifying operators.

Algorithms (BOMP, BLOOMP) based on techniques of band exclusion and local optimization are proposed to enhance Orthogonal Matching Pursuit (OMP) and deal with such coherent sensing matrices.

BOMP and BLOOMP have provably performance guarantee of reconstructing sparse, widely separated objects *independent* of the redundancy and have a sparsity constraint and computational cost similar to OMP's.

Numerical study demonstrates the effectiveness of BLOOMP for compressed sensing with highly coherent, redundant sensing matrices.

Keywords. Model mismatch, compressed sensing, coherence band, gridding error, redundant dictionary.

1. INTRODUCTION

Model mismatch is a fundamental issue in imaging and image processing[3]. To reduce mismatch error, it is often necessary to consider measurement matrices that are highly coherent and redundant. Such measurement matrices lead to serious difficulty in applying compressed sensing (CS) techniques.

Let us consider two examples: discretization in analog imaging and sparse representation of signals.

Consider remote sensing of point sources as depicted in figure 1. Let the noiseless signal at the point r on the sensor plane emitted by the unit source at ξ on the target plane be given by the paraxial Green function

$$
\begin{aligned}
G(r,\xi) &= \frac{e^{i\omega L}}{4\pi L} \times \exp\left(\frac{i\omega|r-\xi|^2}{2L}\right) \\
&= \frac{e^{i\omega L}}{4\pi L} \exp\left(\frac{i\omega r^2}{2L}\right)\exp\left(\frac{-i\omega r\xi}{L}\right)\exp\left(\frac{i\omega\xi^2}{2L}\right)
\end{aligned}
\tag{1}
$$

where ω is the wavenumber. Suppose that s point sources of unknown locations ξ_j and strengths $c_j, j=1,...,s$ emit simultaneously. Then the signals received by the sensors $l, l=1,...,N$ are

$$
y_l = \sum_{j=1}^{s} c_j G(r_l,\xi_j) + n_l, \quad l=1,\dots,N
\tag{2}
$$

where n_l are external noise.
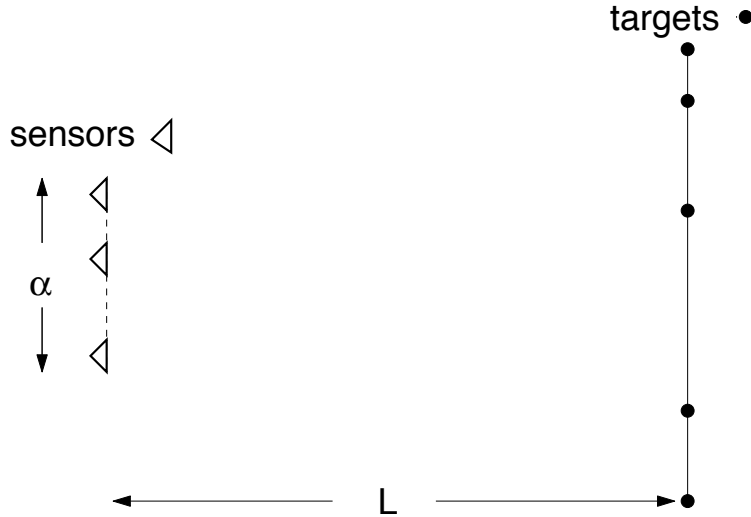
---

*Corresponding author: fannjiang@math.ucdavis.edu.

Figure 1: Sensors are distributed in an aperture of linear size A on the left ($z = 0$) and the point sources of unknown locations are distributed on the target plane on the right ($z = L$).

To cast eq. (2) in the form of finite, discrete linear inversion problem let $\mathcal{G} = \{p_1, \ldots, p_M\}$ be a regular grid of spacing ℓ smaller than the minimum distance among the targets. Consequently, each grid point has *at most* one target within the distance $\ell/2$. Write $\mathbf{x} = (x_j) \in \mathbb{C}^M$ with

$$x_j = \exp\left(\frac{i\omega p_j^2}{2L}\right) c_{j'}$$

whenever p_j is within $\ell/2$ from some target j' and zero otherwise. When a target is located at the midpoint between two neighboring grid points, we can associate either grid point with the target.

Let the data vector $\mathbf{b} = (b_l) \in \mathbb{C}^N$ be defined as

$$b_l = N^{-1/2} 4\pi L e^{-i\omega L} e^{\frac{-i\omega r^2}{2L}} y_l \tag{3}$$

and the measurement matrix be

$$\mathbf{A} = \begin{bmatrix} \mathbf{a}_1 & \ldots & \mathbf{a}_M \end{bmatrix} \in \mathbb{C}^{N \times M} \tag{4}$$

with

$$\mathbf{a}_j = \frac{1}{\sqrt{N}} \left(\exp\left(\frac{-i\omega r_k p_j}{L}\right) \right) \in \mathbb{C}^N, \quad j = 1, ..., M. \tag{5}$$

After proper normalization of noise we rewrite the problem in the form

$$\mathbf{A}\mathbf{x} + \mathbf{e} = \mathbf{b} \tag{6}$$

where the error vector $\mathbf{e} = (e_k) \in \mathbb{C}^N$ is the sum of the external noise $\mathbf{n} = (n(t_k))$ and the discretization or gridding error $\mathbf{d} = (\delta_k) \in \mathbb{C}^N$ due to approximating the locations by the grid points in $\mathcal{G}$. Obviously the discretization error decreases as the grid spacing ℓ decreases. The discretization error, however, depends *nonlinearly* on the objects and hence is not in the form of either additive or multiplicative noise.
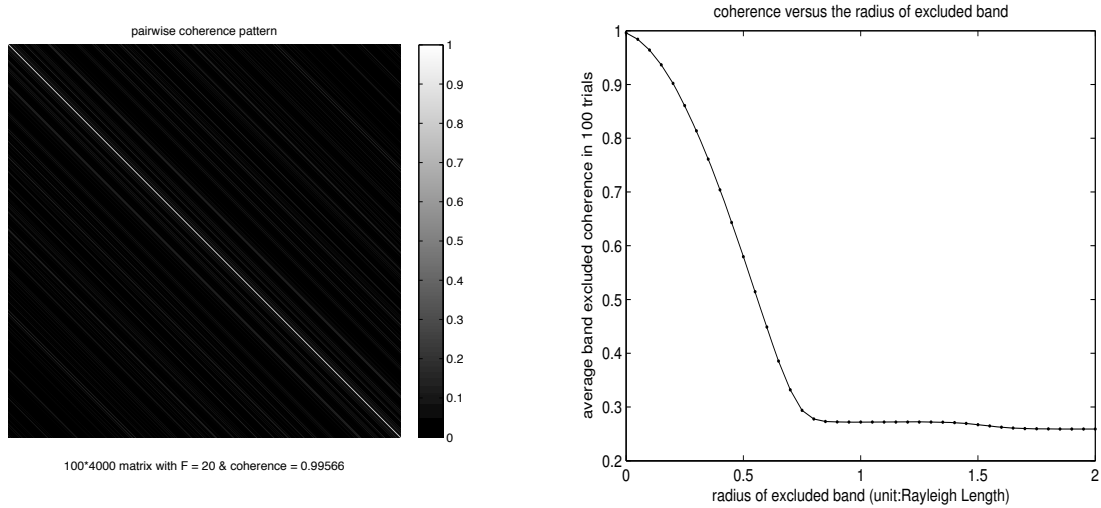
Figure 2: Left: The coherence pattern $[\mu(j,k)]$ of a 100×4000 matrix (5) with $F = 20$. Right: A semi-cross section of the coherence band averaged over 100 independent realizations of (4).

We shall consider in this paper only random sampling over an aperture α satisfying the Rayleigh criterion[1]

$$\alpha \geq \frac{L\lambda}{\ell} \tag{7}$$

where $\lambda = 2\pi/\omega$ is the wavelength. This sets the limit of the resolution

$$\ell \geq \frac{L\lambda}{\alpha} \equiv \ell_R \tag{8}$$

whose right hand side shall be referred to as the Rayleigh length (RL).

To reduce the gridding error, consider the fractional grid with spacing

$$\ell = \ell_R/F \tag{9}$$

for some large integer $F \in \mathbb{N}$ called *the refinement factor*. The relative gridding error $\|\mathbf{d}\|_2/\|\mathbf{b}\|_2$ is roughly inversely proportional to the refinement factor.

On the other hand, a large refinement factor leads to difficulty in applying compressed sensing techniques. A practical indicator of the CS performance is the mutual coherence

$$\mu(\mathbf{A}) = \max_{k \neq l} \frac{|\langle \mathbf{a}_k, \mathbf{a}_l \rangle|}{|\mathbf{a}_k||\mathbf{a}_l|}, \tag{10}$$

which increases with F as the near-by columns of the sensing matrix become highly correlated. Indeed, for $F = 1$, $\mu(\mathbf{A})$ decays like $\mathcal{O}(N^{-1/2})$ while for $F > 1$ $\mu(\mathbf{A}) = \mathcal{O}(1)$.

Figure 2 shows the coherence pattern from the one-dimensional setting. In two or three dimensions, the coherent pattern is more complicated because the coherence band corresponds to the higher dimensional neighborhood.

More generally, coherent bands can arise in sparse and redundant representation by overcomplete dictionaries. Following Duarte and Baraniuk[5] we consider the following CS problem

$$\mathbf{b} = \mathbf{\Phi}\mathbf{y} + \mathbf{e} \tag{11}$$

with a $N \times R$ i.i.d Gaussian matrix $\boldsymbol{\Phi}$ where the signal $\mathbf{y}$ is represented by a redundant dictionary $\boldsymbol{\Psi}$. For example, suppose the sparsifying dictionary is the over-sampled, redundant DFT frame

$$\Psi_{k,j} = \frac{1}{\sqrt{R}} e^{-2\pi i \frac{(k-1)(j-1)}{RF}}, \quad k = 1, ..., R, \quad j = 1, ..., RF. \tag{12}$$

where F is the redundancy factor. Writing $\mathbf{y} = \boldsymbol{\Psi}\mathbf{x}$ we have the same form (6) with $\mathbf{A} = \boldsymbol{\Phi}\boldsymbol{\Psi}$. The coherence bands of $\boldsymbol{\Psi}$ and $\mathbf{A}$ have the similar structure as shown in Figure 2.

Without extra prior information besides the object sparsity, the CS techniques can not guarantee to recover the objects. The additional prior information we impose here is that the objects are sufficiently separated with respect to the coherence band (see below for details). And we propose modified versions of Orthogonal Matching Pursuit (OMP) for handling highly coherent measurement matrices.

The rest of the paper is organized as follows. In Section 2, we introduce the algorithm, BOMP, to deal with highly coherent measurement matrices and states a performance guarantee for BOMP. In Section 3, we introduce another technique, Local Optimization, to enhance BOMP's performance and state the performance guarantee for the resulting algorithm, BLOOMP. In Section 4, we present numerical results for the two examples discussed above and compare the existing algorithms with ours. We conclude in Section 5.

2. BAND-EXCLUDED OMP (BOMP)

The first technique that we introduce to take advantage of the prior information of widely separated objects is called Band Exclusion and can be easily embedded in the greedy algorithm, Orthogonal Matching Pursuit (OMP).

Let $\eta > 0$. Define the η-coherence band of the index k as

$$B_\eta(k) = \{i \mid \mu(i, k) > \eta\}, \tag{13}$$

and the secondary coherence band as

$$B_\eta^{(2)}(k) \quad \equiv \quad B_\eta(B_\eta(k)) = \cup_{j \in B_\eta(k)} B_\eta(j) \tag{14}$$

Embedding the technique of coherence band exclusion in OMP yields the following algorithm.

Algorithm 1. Band-excluded Orthogonal Matching Pursuit (BOMP)

Input: $\mathbf{A}, \mathbf{b}, \eta > 0$

Initialization: $\mathbf{x}^0 = 0, \mathbf{r}^0 = \mathbf{b}$ and $S^0 = \emptyset$

Iteration: For $n = 1, ..., s$

    1) $i_{\max} = \arg\max_i |\langle \mathbf{r}^{n-1}, \mathbf{a}_i \rangle|, i \notin B_\eta^{(2)}(S^{n-1})$

    2) $S^n = S^{n-1} \cup \{i_{\max}\}$

    3) $\mathbf{x}^n = \arg\min_{\mathbf{z}} \|\mathbf{A}\mathbf{z} - \mathbf{b}\|_2$ s.t. $\mathrm{supp}(\mathbf{z}) \in S^n$

    4) $\mathbf{r}^n = \mathbf{b} - \mathbf{A}\mathbf{x}^n$

Output: $\mathbf{x}^s$.

We have the following performance guarantee for BOMP[8].

Theorem 1. *Let $\mathbf{x}$ be s-sparse. Let $\eta > 0$ be fixed. Suppose that*

$$B_\eta(i) \cap B_\eta^{(2)}(j) = \emptyset, \quad \forall i, j \in supp(\mathbf{x}) \tag{15}$$

and that

$$\eta(5s - 4)\frac{x_{max}}{x_{min}} + \frac{5\|\mathbf{e}\|_2}{2x_{min}} < 1 \tag{16}$$

where

$$x_{max} = \max_k |x_k|, \quad x_{min} = \min_k |x_k|.$$

Let $\hat{\mathbf{x}}$ be the BOMP reconstruction. Then $supp(\hat{\mathbf{x}}) \subseteq B_\eta(supp(\mathbf{x}))$ and moreover every nonzero component of $\hat{\mathbf{x}}$ is in the η-coherence band of a unique nonzero component of $\mathbf{x}$.

Remark 1. *In the case of the matrix (5), if every two indices in $supp(\mathbf{x})$ is more than one RL apart, then η is small for sufficiently large N, cf. Figure 2.*

When the dynamic range $x_{max}/x_{min} = \mathcal{O}(1)$, Theorem 1 guarantees approximate recovery of $\mathcal{O}(\eta^{-1})$ sparsity pattern by BOMP. Since $\eta = \mathcal{O}(N^{-1/2})$ for $N \gg 1$, the sparsity constrain by (16) has the same order of magnitude as the condition for OMP's performance[4].

The main difference between (16) and the OMP result lies in the role played by the dynamic range x_{max}/x_{min} which is absent in the condition for OMP's performance. Numerical evidence points to the sensitive dependence of BOMP's performance on dynamic range (Figure 4).

Remark 2. *Condition (15) means that BOMP has a resolution length no worse than 3 ℓ_R independent of the refinement factor. Numerical experiments show that BOMP can resolve objects separated by close to 1 ℓ_R when the dynamic range is close to 1.*

3. BAND-EXCLUDED, LOCALLY OPTIMIZED OMP (BLOOMP)

We now introduce the second technique, the *Local Optimization* (LO), to improve the performance of BOMP.

LO is a residual-reduction technique applied to the current estimate S^k of the object support. To this end, we minimize the residual $\|\mathbf{A}\hat{\mathbf{x}} - \mathbf{b}\|_2$ by varying one location at a time while all other locations held fixed. In each step we consider $\hat{\mathbf{x}}$ whose support differs from S^n by at most one index in the coherence band of S^n but whose amplitude is chosen to minimize the residual. The search is local in the sense that during the search in the coherence band of one nonzero component the locations of other nonzero components are fixed. The total number of search is $\mathcal{O}(s^2 F)$. The amplitudes of the improved estimate is carried out by solving the least squares problem. Because of the local nature of the LO step, the computation is not expensive.

Algorithm 2.   Local Optimization (LO)

Input: $\mathbf{A}, \mathbf{b}, \eta > 0, S^0 = \{i_1, \dots, i_k\}$.
Iteration: For $n = 1, 2, \dots, k$.
   1) $\mathbf{x}^n = \arg \min_{\mathbf{z}} \|\mathbf{A}\mathbf{z} - \mathbf{b}\|_2, \quad \text{supp}(\mathbf{z}) = (S^{n-1} \backslash \{i_n\}) \cup \{j_n\}, \ j_n \in B_\eta(\{i_n\})$.
   2) $S^n = \text{supp}(\mathbf{x}^n)$.
Output: S^k.

We now give a condition under which LO does not spoil the BOMP reconstruction[8].

Theorem 2. *Let $\eta > 0$ and let $\mathbf{x}$ be a s-sparse vector such that (15) holds. Let S^0 and S^k be the input and output, respectively, of the LO algorithm.*
If

$$x_{\min} > (\varepsilon + 2(s-1)\eta) \left(\frac{1}{1-\eta} + \sqrt{\frac{1}{(1-\eta)^2} + \frac{1}{1-\eta^2}} \right) \tag{17}$$

and each element of S^0 is in the η-coherence band of a unique nonzero component of $\mathbf{x}$, then each element of S^k remains in the η-coherence band of a unique nonzero component of $\mathbf{x}$.
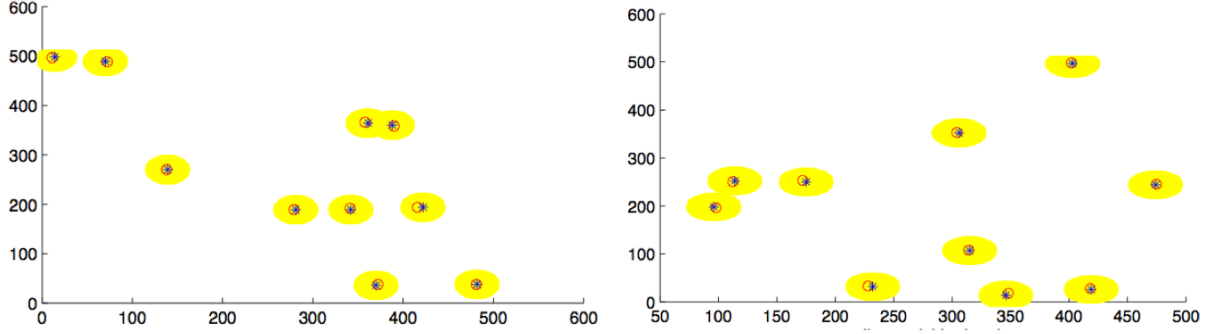
Figure 3: Two instances of BOMP reconstruction: red circles are the exact locations, blue asterisks are recovered locations and the yellow patches are the coherence bands around the objects.

Embedding LO in BOMP gives rise to the Band-excluded, Locally Optimized Orthogonal Matching Pursuit (BLOOMP).

Algorithm 3. Band-excluded, Locally Optimized Orthogonal Matching Pursuit (BLOOMP)

Input: $\mathbf{A}, \mathbf{b}, \eta > 0$
Initialization: $\mathbf{x}^0 = 0, \mathbf{r}^0 = \mathbf{b}$ and $S^0 = \emptyset$
Iteration: For $n = 1, ..., s$
    1) $i_{\max} = \arg\max_i |\langle \mathbf{r}^{n-1}, \mathbf{a}_i \rangle|, i \notin B_\eta^{(2)}(S^{n-1})$
    2) $S^n = \mathrm{LO}(S^{n-1} \cup \{i_{\max}\})$ where LO is the output of Algorithm 2.
    3) $\mathbf{x}^n = \arg\min_{\mathbf{z}} \|\mathbf{A}\mathbf{z} - \mathbf{b}\|_2$ s.t. $\mathrm{supp}(\mathbf{z}) \in S^n$
    4) $\mathbf{r}^n = \mathbf{b} - \mathbf{A}\mathbf{x}^n$
Output: $\mathbf{x}^s$.

Corollary 1. *Let $\hat{\mathbf{x}}$ be the output of BLOOMP. Under the assumptions of Theorems 1 and 2, $\mathrm{supp}(\hat{\mathbf{x}}) \subseteq B_\eta(\mathrm{supp}(\mathbf{x}))$ and moreover every nonzero component of $\hat{\mathbf{x}}$ is in the η-coherence band of a unique nonzero component of $\mathbf{x}$.*

Even though we can not improve the performance guarantee for BLOOMP, in practice the LO technique greatly enhances the success probability of recovery with respect to noise stability and dynamic range. Moreover, if Corollary 1 holds, then for all practical purposes we have the residual bound for the BLOOMP reconstruction $\hat{\mathbf{x}}$

$$\|\mathbf{b} - \mathbf{A}\hat{\mathbf{x}}\|_2 \leq c\|\mathbf{e}\|_2, \quad c \sim 1. \tag{18}$$

4. NUMERICAL RESULTS

We test the algorithms, BOMP and BLOOMP, on the two examples discussed in the Introduction.

For the first example (4)-(6), we use the refinement factor $F = 20$. For the objects $\mathbf{x}$, we use 10 randomly phased and located objects, separated by at least $3 \ell_R$. The noise is the i.i.d. Gaussian noise $\mathbf{e} \sim N(0, \sigma^2 I)$.

Figure 3 shows two instances of reconstruction by BOMP in two dimensions. The recovered objects (blue asterisks) are close to the true objects (red circles) well within the coherence bands (yellow patches).
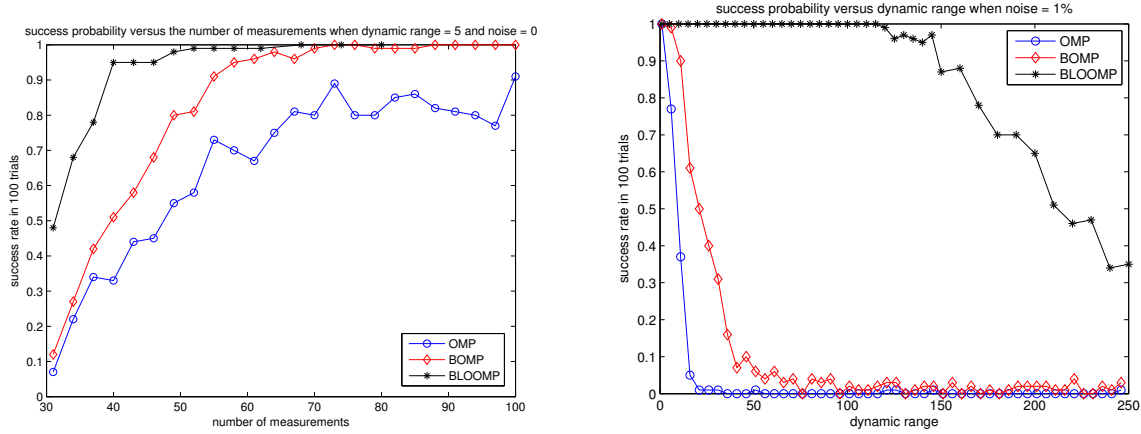
Figure 4: Success rate versus number of measurements (left, dynamic range 5, zero noise) and dynamic range (right, 1% noise) for OMP, BOMP and BLOOMP.

For the rest of simulations, we show the percentage of successes in 100 independent trials. A reconstruction is counted as a success if every reconstructed object is within $1\,\ell_R$ of the object support. This is equivalent to the criterion that the Bottleneck distance between the true support and the reconstructed support is less than $1\,\ell_R$. The result is shown in Figure 4. With 10 objects of dynamic range 5, BLOOMP requires the least number of measurements, followed by BOMP and then OMP, which does not achieve high success rate even with 100 measurements (left panel). With 100 measurements ($N = 100$) and 1% noise, BLOOMP can handle dynamic range up to 120 while BOMP and OMP can handle dynamic range about 5 and 1, respectively.

For the second example (11)-(12), we test, in addition to our algorithms, the method proposed by Duarte and Baraniuk[5] and the analysis approach of frame-adapted Basis Pursuit[2,6].

The algorithm, Spectral Iterative Hard Thresholding (SIHT)[5], assumes the model-based RIP which, in spirit, is equivalent to the assumption of well separated support in the synthesis coefficients and therefore resembles closely to our approach.

While SIHT is a synthesis method like BOMP and BLOOMP, the frame-adapted BP

$$\min \|\mathbf{\Psi}^\star \mathbf{z}\|_1 \qquad \text{s.t } \|\mathbf{\Phi}\mathbf{z} - \mathbf{b}\|_2 \le \|\mathbf{e}\|_2, \tag{19}$$

is the analysis approach[6]. Candès *et al.*[2] have established a performance guarantee for (19) provided that the measurement matrix $\mathbf{\Phi}$ satisfies the frame-adapted RIP:

$$(1 - \delta)\|\mathbf{\Psi}\mathbf{z}\|_2 \le \|\mathbf{\Phi}\mathbf{\Psi}\mathbf{z}\|_2 \le (1 + \delta)\|\mathbf{\Psi}z\|_2, \quad \|\mathbf{z}\|_0 \le 2s \tag{20}$$

for a tight frame $\mathbf{\Psi}$ and a sufficiently small δ and that the analysis coefficients $\mathbf{\Psi}^*\mathbf{y}$ are sparse or compressible.

Instead of the synthesis coefficients $\mathbf{x}$, however, the quantities of interest are $\mathbf{y}$. Accordingly we measure the performance by the relative error $\|\hat{\mathbf{y}} - \mathbf{y}\|_2 / \|\mathbf{y}\|_2$ averaged over 100 independent trials. In each trial, 10 randomly phased and located objects (i.e. $\mathbf{x}$) of dynamic range 10 and i.i.d. Gaussian $\mathbf{\Phi}$ are generated. We set $N = 100, R = 200, F = 20$ for test of noise stability and vary N for test of measurement compression.

As shown in Figure 5, BLOOMP is the best performer in noise stability (left panel) and measurement compression (right panel). BLOOMP requires about 40 measurements to achieve nearly perfect reconstruction while the other methods require more than 200 measurements. Despite the powerful error bound established in [2], the analysis approach (19) needs more than 200 measurements for accurate recovery because the analysis coefficients $\mathbf{\Psi}^*\mathbf{y}$ are typically not sparse. Here redundancy $F = 20$ produces about $2F = 40$ highly coherent columns around each synthesis coefficient and hence $\mathbf{\Psi}^*\mathbf{y}$ has about 400 significant components. In
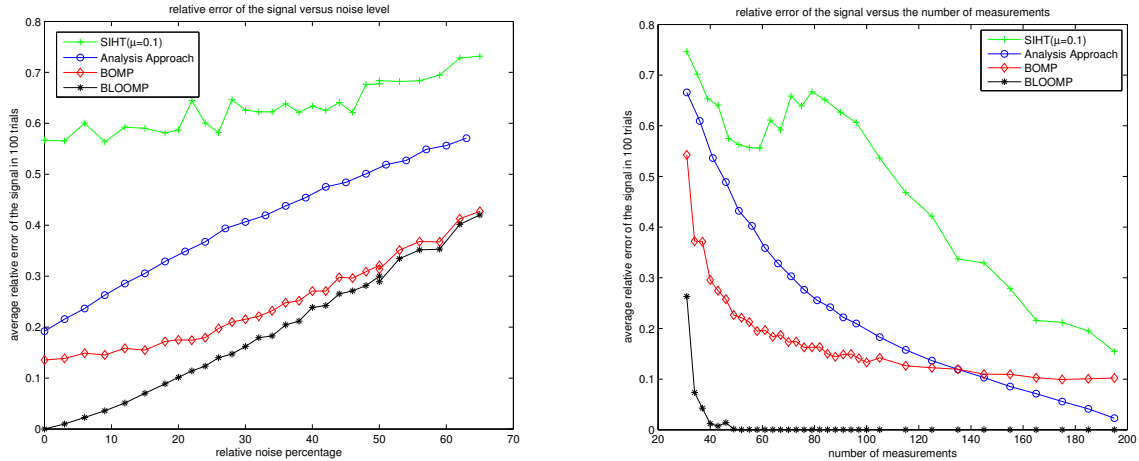
Figure 5: Relative errors versus relative noise (left) and number of measurements (right, zero noise) for dynamic range 10.

general, the sparsity of the analysis coefficients is at least $2sF$ where s is the sparsity of the *widely separated synthesis* coefficients and F is the redundancy. Thus according to the error bound of [2] the performance of the analysis approach (19) would degrade with the redundancy of the dictionary.

To understand the superior performance of BLOOMP in this set-up let us give an error bound using (18) and (20)

$$\|\mathbf{\Psi}(\mathbf{x} - \hat{\mathbf{x}})\|_2 \leq \frac{1}{1-\delta}\|\mathbf{A}(\mathbf{x} - \hat{\mathbf{x}})\|_2 \leq \frac{1}{1-\delta}\|\mathbf{b} - \mathbf{e} - \mathbf{A}\hat{\mathbf{x}}\|_2 \leq \frac{1+c}{1-\delta}\|\mathbf{e}\|_2 \tag{21}$$

where $\hat{\mathbf{x}}$ is the output of BLOOMP. This implies that the reconstruction error of BLOOMP is essentially determined by the external noise, consistent with the left and right panels of Figure 5, and is independent of the dictionary redundancy if Corollary 1 holds. In comparison, the BOMP result appears to approach an asymptote of nonzero ($\sim 10\%$) error. This demonstrates the effect of local optimization technique in reducing error. The advantage of BLOOMP over BOMP, however, disappears in the presence of large external noise (left panel).

5. CONCLUSION

We have proposed algorithms, BOMP and BLOOMP, for sparse recovery with highly coherent, redundant sensing matrices and have established performance guarantee that is *redundancy independent*. These algorithms have a sparsity constraint and computational cost similar to OMP's. Our work is inspired by the redundancy-independent performance guarantee recently established for the MUSIC algorithm for array processing.[7]

Our algorithms are based on variants of OMP enhanced by two novel techniques: band exclusion and local optimization. We have extended these techniques to various CS algorithms, including Lasso, and performed systematic tests elsewhere[8].

Numerical results demonstrate the superiority of BLO-based algorithms for reconstruction of sparse objects separated by above the Rayleigh threshold.

References

[1] Born, M. and Wolf, E., [Principles of Optics], 7-th edition, Cambridge University Press, 1999.

[2] Candès, E.J. Eldar, Y.C., Needell, D and Randall, P., "Compressed sensing with coherent and redundant dictionaries," Appl. Comput. Harmon. Anal. 31 (1), 59–73 (2011)

[3] Chi, Y., Pezeshki, A., Scharf, L. and Calderbank, R., "Sensitivity to basis mismatch in compressed sensing," International Conference on Acoustics, Speech, and Signal Processing (ICASSP). Dallas, Texas, Mar. 2010.

[4] Donoho, D.L., Elad, M. and Temlyakov, V.N., "Stable recovery of sparse overcomplete representations in the presence of noise," IEEE Trans. Inform. Theory 52, 6-18 (2006).

[5] Duarte, M.F. and Baraniuk, R.G., "Spectral compressive sensing," preprint, 2010.

[6] Elad, M., Milanfar, P. and Rubinstein, R., " Analysis versus synthesis in signal prior," Inverse Problems 23, 947-968 (2007).

[7] Fannjiang, A., "The MUSIC algorithm for sparse objects: a compressed sensing analysis," Inverse Problems 27, 035013 (2011).

[8] Fannjiang, A. and Liao, W., "Coherence pattern-guided compressive sensing with unresolved grids," arXiv:1106.5177