SIAM J. IMAGING SCIENCES
Vol. 5, No. 1, pp. 179–202

Coherence Pattern–Guided Compressive Sensing with Unresolved Grids*

Albert Fannjiang† and Wenjing Liao‡

Abstract. Highly coherent sensing matrices arise in discretization of continuum imaging problems such as radar and medical imaging when the grid spacing is below the Rayleigh threshold. Algorithms based on techniques of band exclusion (BE) and local optimization (LO) are proposed to deal with such coherent sensing matrices. These techniques are embedded in the existing compressed sensing algorithms, such as Orthogonal Matching Pursuit (OMP), Subspace Pursuit (SP), Iterative Hard Thresholding (IHT), Basis Pursuit (BP), and Lasso, and result in the modified algorithms BLOOMP, BLOSP, BLOIHT, BP-BLOT, and Lasso-BLOT, respectively. Under appropriate conditions, it is proved that BLOOMP can reconstruct sparse, widely separated objects up to one Rayleigh length in the Bottleneck distance *independent* of the grid spacing. One of the most distinguishing attributes of BLOOMP is its capability of dealing with large dynamic ranges. The BLO-based algorithms are systematically tested with respect to four performance metrics: dynamic range, noise stability, sparsity, and resolution. With respect to dynamic range and noise stability, BLOOMP is the best performer. With respect to sparsity, BLOOMP is the best performer for high dynamic range, while for dynamic range near unity BP-BLOT and Lasso-BLOT with the optimized regularization parameter have the best performance. In the noiseless case, BP-BLOT has the highest resolving power up to certain dynamic range. The algorithms BLOSP and BLOIHT are good alternatives to BLOOMP and BP/Lasso-BLOT: they are faster than both BLOOMP and BP/Lasso-BLOT and share, to a lesser degree, BLOOMP's amazing attribute with respect to dynamic range. Detailed comparisons with the algorithms Spectral Iterative Hard Thresholding (SIHT) and the frame-adapted BP demonstrate the superiority of the BLO-based algorithms for the problem of sparse approximation in terms of highly coherent, redundant dictionaries.

Key words. compressed sensing, gridding error, model mismatch, coherence band exclusion, local optimization

AMS subject classifications. 65, 68

DOI. 10.1137/110838509

1. Introduction.

Reconstruction of a high-dimensional sparse signal from sparse linear measurements is a fundamental problem relevant to imaging, inverse problems, and signal processing.

Consider, for example, the problem of spectral estimation in signal processing. Let the uncontaminated signal $y(t)$ be a linear combination of s time-harmonic components

$$\{e^{-i2\pi\omega_j t} : j = 1, \ldots, s\},$$

*Received by the editors June 24, 2011; accepted for publication (in revised form) November 21, 2011; published electronically February 23, 2012.
http://www.siam.org/journals/siims/5-1/83850.html
†Corresponding author. Department of Mathematics, University of California, Davis, CA 95616-8633 (fannjiang@math.ucdavis.edu). The research of this author was supported in part by NSF grant DMS 0908535.
‡Department of Mathematics, University of California, Davis, CA 95616-8633 (wjliao@math.ucdavis.edu). The research of this author was supported in part by NSF grant DMS 0908535.

namely

$$y(t) = \sum_{j=1}^{s} c_j e^{-i2\pi\omega_j t}, \tag{1}$$

where c_j are the amplitudes. Suppose that $y(t)$ is contaminated by noise $n(t)$ and the received signal is

$$b(t) = y(t) + n(t). \tag{2}$$

The task is to find the frequencies $\Omega = \{\omega_j\}$ and the amplitudes $\{c_j\}$ by sampling $b(t)$ at discrete times.

A standard approach to spectral estimation is to turn the problem into the linear inversion problem as follows. To fix the idea, let $t_k, k = 1, \ldots, N$, be the sample times in the unit interval $[0, 1]$. Set $\mathbf{b} = (b(t_k)) \in \mathbb{C}^N$ as the data vector. We approximate the frequencies by the unknown closest subset of cardinality s of a regular grid $\mathcal{G} = \{p_1, \ldots, p_M\}$ and write the corresponding amplitudes as $\mathbf{x} = (x_j) \in \mathbb{C}^M$, where the components of $\mathbf{x}$ equal the amplitudes $\{c_j\}$ whenever the grid points are the *nearest* grid points to the frequencies $\{\omega_j\}$ and zero otherwise. Let the measurement matrix be

$$\mathbf{A} = \begin{bmatrix} \mathbf{a}_1 & \ldots & \mathbf{a}_M \end{bmatrix} \in \mathbb{C}^{N\times M} \tag{3}$$

with

$$\mathbf{a}_j = \frac{1}{\sqrt{N}} \left(e^{-i2\pi t_k p_j}\right)_{k=1}^{N}, \quad j = 1, \ldots, M. \tag{4}$$

We cast the spectral estimation problem in the form

$$\mathbf{A}\mathbf{x} + \mathbf{e} = \mathbf{b}, \tag{5}$$

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 frequencies by the grid points in $\mathcal{G}$. By definition, the gridding error is given by

$$\mathbf{d} = \mathbf{b} - \mathbf{n} - \mathbf{A}\mathbf{x}. \tag{6}$$

Small gridding error requires that the objects be a priori close to the grid points.

Sparse reconstruction with $N, s \ll M$, where s is the sparsity of $\mathbf{x}$, has recently attracted a lot of attention in various areas thanks to the breakthroughs in compressive sensing (CS) [8, 15, 29]. The main thrust of CS is the L^1-minimization principles, Basis Pursuit (BP) and Lasso, for solution characterization. Many L^1-based algorithms as well as the alternative, greedy algorithms, which are not directly based on global optimization, require either incoherence or the restricted isometry property (RIP) to have good performances.

One commonly used characterization of incoherence in CS is in terms of the mutual coherence μ. Let the pairwise coherence between the kth and jth columns be

$$\mu(k, l) = \frac{|\langle \mathbf{a}_k, \mathbf{a}_l \rangle|}{\|\mathbf{a}_k\| \|\mathbf{a}_l\|}. \tag{7}$$

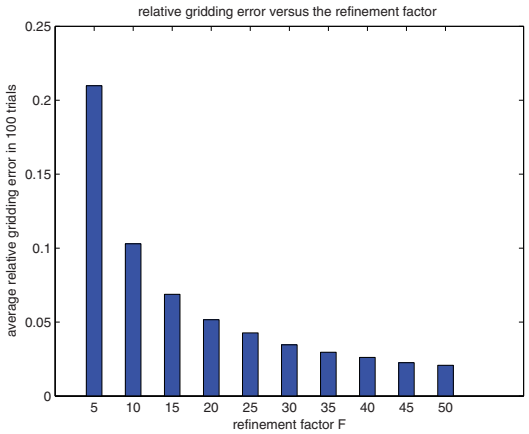


Figure 1. *The relative gridding error is roughly inversely proportional to the refinement factor.*

The mutual coherence of $\mathbf{A}$ is the maximum pairwise coherence among all pairs of columns

$$\mu(\mathbf{A}) = \max_{k \neq l} \mu(k, l). \tag{8}$$

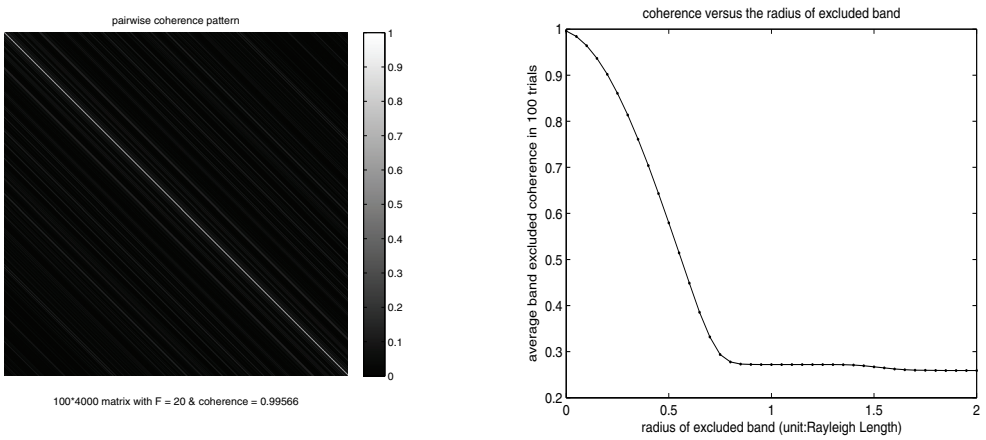
According to the theory of optimal recovery [14], for time sampling in $[0, 1]$, the minimum resolvable length in the frequency domain is unity. This is the Rayleigh threshold, and we shall refer to this length as the Rayleigh length (RL). Hence for the traditional inversion methods to work, it is essential that the grid spacing in $\mathcal{G}$ be no less than 1 RL. In the CS setting the Rayleigh threshold is closely related to the decay property of the mutual coherence [21]. Moreover, for $\mathcal{G} \subset \mathbb{Z}$ and uniformly randomly selected $t_k \in [0, 1]$ the corresponding matrix $\mathbf{A}$ is a random partial Fourier matrix which has a decaying mutual coherence $\mu = \mathcal{O}(N^{-1/2})$ and satisfies RIP with high probability [8, 28].

Without any prior information about the object support, however, the gridding error for the resolved grid can be as large as the data themselves, creating an unfavorable condition for sparse reconstruction. To reduce the gridding error, it is natural to consider the fractional grid

$$\mathbb{Z}/F = \{j/F : j \in \mathbb{Z}\} \tag{9}$$

with some large integer $F \in \mathbb{N}$ called *the refinement factor*. Figure 1 shows that the relative gridding error $\|\mathbf{d}\|_2/\|\mathbf{b}\|_2$ is roughly inversely proportional to the refinement factor. The mutual coherence, however, increases with F as the nearby columns of the sensing matrix become highly correlated.

Figure 2(a) shows the coherence pattern $[\mu(j, k)]$ of a 100×4000 matrix (4) with $F = 20$. The bright diagonal band represents a heightened correlation (pairwise coherence) between a column vector and its neighbors on both sides (about 30 elements). Figure 2(b) shows a half cross section of the coherence band across 2 RLs. Sparse recovery with large F exceeds the capability of currently known algorithms as the condition number of the 100×30 submatrix



(a) (b)

Figure 2. (a) *Coherence pattern* $[\mu(j,k)]$ *for the* 100×4000 *matrix with* $F=20$. *The off-diagonal elements tend to diminish as the row number increases. The coherence band near the diagonals, however, persists and has the average profile shown in* (b)*, where the vertical axis is the pairwise coherence averaged over* 100 *independent trials and the horizontal axis is the separation between two columns in the unit of RL.*

corresponding to the coherence band in Figure 2 easily exceeds 10^{15}. The high condition number makes stable recovery impossible.

The difficulty with unresolved grids is not limited to the problem of spectral estimation in signal processing. Indeed, the issue is intrinsic and fundamental to discretization of PDE-based inverse problems such as remote sensing and medical imaging [10, 11, 25]. While Figure 2 is typical of the coherence pattern from discretization of a one-dimensional problem, in two or three dimensions, the coherent pattern is more complicated than Figure 2. Nevertheless, the coherence band typically reflects proximity in the physical space. The proximity between the object support and its reconstruction can be described by the Bottleneck or the Hausdorff distance [18]. More generally, coherent bands can arise in sparse and redundant representation by overcomplete dictionaries (see section 6 for an example). Under this circumstance, the Bottleneck or Hausdorff distance may not have a direct physical meaning.

In any case, the hope is that if the objects are sufficiently separated with respect to the coherence band, then the problem of a huge condition number associated with unresolved grids can somehow be circumvented and the object support can be approximately reconstructed.

Under this additional assumption of widely separated objects, we propose in the present work several algorithmic approaches to recovery with unresolved grids and provide some performance guarantee for these algorithms.

The paper is organized as follows. In section 2 we introduce the technique of band exclusion (BE) to modify the Orthogonal Matching Pursuit (OMP) algorithm and obtain a performance guarantee for the improved algorithm, called Band-excluded OMP (BOMP). In section 3 we introduce the technique of local optimization (LO) and propose the algorithms Locally Optimized OMP (LOOMP) and band-excluded LOOMP (BLOOMP). In section 4 we

introduce the technique of Band-Excluded Thresholding (BET), which comes in two forms, Band-excluded Matched Thresholding (BMT) and Band-excluded Locally Optimized Thresholding (BLOT), and propose the algorithms Band-excluded Locally Optimized Subspace Pursuit (BLOSP), Band-excluded Locally Optimized Compressive Sampling Matching Pursuit (BLOCoSaMP), Band-excluded Locally Optimized Iterative Hard Thresholding (BLOIHT), and BP/Lasso with BLOT (BP/Lasso-BLOT). In section 5 we present a numerical study of the comparative advantages of various algorithms. In section 6 we compare the performance of our algorithms with that of the existing algorithms Spectral Iterative Hard Thresholding (SIHT) and coherent dictionary–based BP recently proposed in [17] and [5], respectively. We conclude in section 7.

2. Band exclusion (BE). The first technique that we introduce to take advantage of the prior information of widely separated objects is called BE and can be easily embedded in the greedy algorithm OMP [13, 27].

First let us recall a standard performance guarantee for OMP [16].

Proposition 1. *Suppose that the sparsity s of the signal vector $\mathbf{x}$ satisfies*

$$\mu(\mathbf{A})(2s-1) + 2\frac{\|\mathbf{e}\|_2}{x_{\min}} < 1, \tag{10}$$

where $x_{\min} = \min_k |x_k| = |x_s|$. Denote by $\hat{\mathbf{x}}$ the output of the OMP reconstruction. Then

$$\operatorname{supp}(\hat{\mathbf{x}}) = \operatorname{supp}(\mathbf{x}),$$

where $\operatorname{supp}(\mathbf{x})$ *is the support of* $\mathbf{x}$.

In the ideal case where $\mathbf{e} = 0$, (10) reduces to

$$\mu(\mathbf{A}) < \frac{1}{2s-1}, \tag{11}$$

which is near the threshold of OMP's capability for exact reconstruction of arbitrary objects of sparsity s.

Intuitively speaking, if the objects are *not* in each other's coherence band, then it should be possible to localize the objects *approximately* within their respective coherence bands, no matter how large the mutual coherence is.

First let us precisely define the notion of coherence band. Let $\eta > 0$. Define the η-coherence band of the index k to be the set

$$B_\eta(k) = \{i \mid \mu(i,k) > \eta\}, \tag{12}$$

and the η-coherence band of the index set S to be the set

$$B_\eta(S) = \cup_{k\in S} B_\eta(k).$$

Due to the symmetry $\mu(i,k) = \mu(k,i)\ \forall i,k$, $i \in B_\eta(k)$ if and only if $k \in B_\eta(i)$.

Denote

$$B_\eta^{(2)}(k) \equiv B_\eta(B_\eta(k)) = \cup_{j\in B_\eta(k)} B_\eta(j), \tag{13}$$

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

To imbed BE into OMP, we make the following change to the matching step:

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

meaning that the double η-band of the estimated support in the previous iteration is avoided in the current search. This is natural if the sparsity pattern of the object is such that $B_\eta(j), j \in \mathrm{supp}(\mathbf{x})$, are pairwise disjoint. We call the modified algorithm the Band-excluded Orthogonal Matching Pursuit (BOMP), which is formally stated in Algorithm 1.

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 $n = 1, \dots, s$
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. $\mathbf{x}^n = \arg\min_{\mathbf{z}} \|\mathbf{A}\mathbf{z} - \mathbf{b}\|_2$ subject to (s.t.) $\mathrm{supp}(\mathbf{z}) \in S^n$.
4. $\mathbf{r}^n = \mathbf{b} - \mathbf{A}\mathbf{x}^n$.

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

A main theoretical result of the present paper is the following performance guarantee for BOMP.

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

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

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* $\mathrm{supp}(\hat{\mathbf{x}}) \subseteq B_\eta(\mathrm{supp}(\mathbf{x}))$*, and, moreover, every nonzero component of* $\hat{\mathbf{x}}$ *is in the* η*-coherence band of a unique nonzero component of* $\mathbf{x}$*.*

Proof. We prove the theorem by induction.

Suppose $\mathrm{supp}(\mathbf{x}) = \{J_1, \dots, J_s\}$. Let $J_{\max} \in \mathrm{supp}(\mathbf{x})$ be the index of the largest component in the absolute value of $\mathbf{x}$.

In the first step,

$$\begin{aligned} |\mathbf{b}^\star \mathbf{a}_{J_{\max}}| &= |x_{J_{\max}} + x_{J_2}\mathbf{a}_{J_2}^\star \mathbf{a}_{J_{\max}} + \cdots + x_{J_s}\mathbf{a}_{J_s}^\star \mathbf{a}_{J_{\max}} + \mathbf{e}^\star \mathbf{a}_{J_{\max}}| \\ &\geq x_{\max} - x_{\max}(s-1)\eta - \|\mathbf{e}\|_2 \end{aligned} \tag{17}$$

by assumption (15). On the other hand, $\forall l \notin B_\eta(\mathrm{supp}(\mathbf{x}))$,

$$\begin{aligned} |\mathbf{b}^\star \mathbf{a}_l| &= |x_{J_1}\mathbf{a}_{J_1}^\star \mathbf{a}_l + x_{J_2}\mathbf{a}_{J_2}^\star \mathbf{a}_l + \cdots + x_{J_s}\mathbf{a}_{J_s}^\star \mathbf{a}_l + \mathbf{e}^\star \mathbf{a}_l| \\ &\leq x_{\max} s\eta + \|\mathbf{e}\|_2 \end{aligned} \tag{18}$$

by using (15) again.

Hence, if

$$(2s-1)\eta + 2\frac{\|\mathbf{e}\|_2}{x_{\max}} < 1,$$

then the right-hand side of (17) is greater than the right-hand side of (18), which implies that the first index selected by BOMP must belong to $B_\eta(\mathrm{supp}(\mathbf{x}))$.

Now suppose without loss of generality that the first $(k-1)$ indices $I_1, \dots, I_{k-1}$ selected by BOMP are in $B_\eta(J_i), J_i \in \mathrm{supp}(\mathbf{x}), i = 1, \dots, k-1$, respectively. Write the residual as

$$\mathbf{r}^{k-1} = \mathbf{b} - c_{I_1}\mathbf{a}_{I_1} - c_{I_2}\mathbf{a}_{I_2} - \cdots - c_{I_{k-1}}\mathbf{a}_{I_{k-1}}.$$

First, we estimate the coefficients $c_{I_1}, \dots, c_{I_{k-1}}$. Since $\langle \mathbf{r}^{k-1}, \mathbf{a}_{I_1} \rangle = 0$,

$$c_{I_1} = x_{J_1}\mathbf{a}_{J_1}^\star \mathbf{a}_{I_1} + x_{J_2}\mathbf{a}_{J_2}^\star \mathbf{a}_{I_1} + \cdots + x_{J_s}\mathbf{a}_{J_s}^\star \mathbf{a}_{I_1} + \mathbf{e}^\star \mathbf{a}_{I_1} - c_{I_2}\mathbf{a}_{I_2}^\star \mathbf{a}_{I_1} - \cdots - c_{I_{k-1}}\mathbf{a}_{I_{k-1}}^\star \mathbf{a}_{I_1},$$

which implies

$$|c_{I_1}| \le x_{\max} + x_{\max}(s-1)\eta + \|\mathbf{e}\|_2 + \eta(|c_{I_2}| + |c_{I_3}| + \cdots + |c_{I_{k-1}}|).$$

Likewise, we have

$$|c_{I_j}| \le x_{\max} + x_{\max}(s-1)\eta + \|\mathbf{e}\|_2 + \eta \sum_{i \ne j} |c_{I_i}|, \quad j = 1, \dots, k-1. \tag{19}$$

Let $c_{\max} = \max_{j=1,\dots,k-1} |c_{I_j}|$. Inequality (19) implies that

$$c_{\max} \le x_{\max} + x_{\max}(s-1)\eta + \|\mathbf{e}\|_2 + \eta(k-2)c_{\max}$$

and hence

$$c_{\max} \le \frac{1}{1-\eta(k-2)}[x_{\max} + x_{\max}(s-1)\eta + \|\mathbf{e}\|_2].$$

Moreover, condition (16) implies that $\eta(s-1) < \frac{1}{5}$ and $\frac{1}{1-\eta(k-2)} \le \frac{5}{4}$. Hence

$$c_{\max} \le \frac{5}{4}\left(x_{\max} + \frac{1}{5}x_{\max} + \|\mathbf{e}\|_2\right) \le \frac{3}{2}x_{\max} + \frac{5}{4}\|\mathbf{e}\|_2. \tag{20}$$

We claim that $B_\eta^{(2)}(S^{k-1})$ and $\{J_k, \dots, J_s\}$ are disjoint.

If the claim is not true, then there exists J_i for some $i \in \{k, \dots, s\}$, $J_i \in B_\eta^{(2)}(I_l)$ for some $l \in \{1, \dots, k-1\}$. Consequently, $J_i \in B_\eta^{(3)}(J_l)$ or, equivalently, $B_\eta(J_i) \cap B_\eta^{(2)}(J_l) \ne \emptyset$, which is contradictory to the assumption (15).

Now we show that the index selected in the kth step is in $B_\eta(\{J_k, \dots, J_s\})$.

On the one hand, we have

$$\begin{aligned} |\mathbf{r}^{k-1\star}\mathbf{a}_{J_i}| &= |x_{J_1}\mathbf{a}_{J_1}^\star \mathbf{a}_{J_i} + \cdots + x_{J_s}\mathbf{a}_{J_s}^\star \mathbf{a}_{J_i} + \mathbf{e}^\star \mathbf{a}_{J_i} \\ &\qquad - c_{I_1}\mathbf{a}_{I_1}^\star \mathbf{a}_{J_i} - \cdots - c_{I_{k-1}}\mathbf{a}_{I_{k-1}}^\star \mathbf{a}_{J_i}| \\ &\ge x_{\min} - \eta(s-1)x_{\max} - \|\mathbf{e}\|_2 - \eta(k-1)c_{\max}. \end{aligned} \tag{21}$$

On the other hand, we have that $\forall l \notin B_\eta^{(2)}(S^{k-1}) \cup B_\eta(\{J_k, \dots, J_s\})$,

$$
\begin{aligned}
|\mathbf{r}^{k-1\star}\mathbf{a}_l| &= |x_{J_1}\mathbf{a}_{J_1}^\star \mathbf{a}_l + \cdots + x_{J_k}\mathbf{a}_{J_k}^\star \mathbf{a}_l + \cdots + x_{J_s}\mathbf{a}_{J_s}^\star \mathbf{a}_l + \mathbf{e}^\star \mathbf{a}_l \\
&\quad - c_{I_1}\mathbf{a}_{I_1}^\star \mathbf{a}_l - \cdots - c_{I_{k-1}}\mathbf{a}_{I_{k-1}}^\star \mathbf{a}_l| \\
&\le \eta s x_{\max} + \|\mathbf{e}\|_2 + \eta(k-1)c_{\max}
\end{aligned} \tag{22}
$$

in view of $B_\eta(\{J_1, \dots, J_{k-1}\}) \subseteq B_\eta^{(2)}(S^{k-1})$ as a result of the induction assumption.

If the right-hand side of (21) is greater than the right-hand side of (22) or, equivalently,

$$
\eta(2s + 3k - 4)\frac{x_{\max}}{x_{\min}} + \left(2 + \frac{5}{2}\eta(k-1)\right)\frac{\|\mathbf{e}\|_2}{x_{\min}} < 1, \tag{23}
$$

then the kth index selected by BOMP must be in $B_\eta(\{J_k, \dots, J_s\})$ because

$$
B_\eta^{(2)}(S^{k-1}) \cap \{J_k, \dots, J_s\} = \emptyset
$$

and because the kth selected index does not belong in $B_\eta^{(2)}(S^{k-1})$ according to the band exclusion rule. Condition (16) implies (23) by setting the maximal $k = s$ in (23) and noting that $\eta(5s - 4) < 1$ under (16). ■

Remark 1. In the case of the matrix (4), if every two indices in supp($\mathbf{x}$) are more than 1 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 the $\mathcal{O}(\eta^{-1})$ sparsity pattern by BOMP.

Remark 2. The main difference between Theorem 1 and Proposition 1 lies in the role played by the dynamic range $x_{\max}/x_{\min}$ and the separation condition (15).

First, numerical evidence shows degradation in BOMP's performance with increased dynamic range (Figures 3(c)–3(f)), consistent with the prediction of (16). Dynamic range of objects clearly is an essential factor in determining the performance of recovery. This sensitivity to dynamic range can be drastically reduced by the LO technique introduced in the next section.

Second, condition (15) means that BOMP can resolve 3 RLs. Numerical experiments show that BOMP can resolve objects separated by close to 1 RL when the dynamic range is close to 1 (Figure 7).

Remark 3. The BE technique reduces the total number of inner products in the selection step 1 of Algorithm 1 (BOMP) by $\mathcal{O}(s^2\beta)$, where β is the average number of points in a coherence band. This amounts to a savings of $\mathcal{O}(s^2\beta N)$ flops.

3. Local optimization (LO). As our numerical experiments show, the main shortcoming with BOMP is in its failure to perform even when the dynamic range is only moderate.

To overcome this problem, we now introduce the second technique: *local optimization* (LO).

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 are 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 amplitude of the improved estimate is recovered by solving the least squares problem. Because of the local nature of the LO step, the computation is not expensive (see Remark 4 below).

Algorithm 2. Local Optimization (LO).

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

Output: S^k.

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 $n = 1, \ldots, s$
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. $\mathbf{x}^n = \arg\min_{\mathbf{z}} \|\mathbf{Az} - \mathbf{b}\|_2$ s.t. $\operatorname{supp}(\mathbf{z}) \in S^n$.
4. $\mathbf{r}^n = \mathbf{b} - \mathbf{Ax}^n$.

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

We now give a condition under which LO does not spoil the BOMP reconstruction of Theorem 1.

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

If

$$x_{\min} > (\varepsilon + 2(s-1)\eta)\left(\frac{1}{1-\eta} + \sqrt{\frac{1}{(1-\eta)^2} + \frac{1}{1-\eta^2}}\right), \quad \varepsilon = \|\mathbf{e}\|, \tag{24}$$

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 $\mathbf{x}$.

Proof. Because of the iterative nature of Algorithm 2, it is sufficient to show that each element of S^1 is in the η-coherence band of a unique nonzero component of $\mathbf{x}$.

Suppose $J_1 \in \operatorname{supp}(\mathbf{x})$ and $i_1 \in B_\eta(J_1)$. Let

$$r = \min_{\mathbf{z}} \|\mathbf{Az} - \mathbf{b}\|_2, \quad \operatorname{supp}(\mathbf{z}) = (S^0 \backslash \{i_1\}) \cup \{J_1\},$$
$$r' = \min_{\mathbf{z}} \|\mathbf{Az} - \mathbf{b}\|_2, \quad \operatorname{supp}(\mathbf{z}) = (S^0 \backslash \{i_1\}) \cup \{j\}, \quad j \in B_\eta(i_1) \backslash B_\eta(J_1).$$

We want to show that $r < r'$ $\forall j \in B_\eta(i_1)\backslash B_\eta(J_1)$ so that the LO step is certain to pick a new index within the η-coherence band of J_1, reducing the residual in the meantime. For the subsequent analysis, we fix $j \in B_\eta(i_1)\backslash B_\eta(J_1)$.

Reset the J_1 component of $\mathbf{x}$ to zero and denote the resulting vector by $\mathbf{x}'$. Hence the sparsity of $\mathbf{x}'$ is $s-1$. It follows from the definition of r that

$$r \le \min_{\mathbf{z}} \|\mathbf{A}\mathbf{z} - \mathbf{A}\mathbf{x}' - \mathbf{e}\|_2, \quad \text{supp}(\mathbf{z}) = \{i_2, \dots, i_k\}.$$

We also have

$$r' = \min_{\mathbf{z}} \|\mathbf{A}(\mathbf{z} - \mathbf{x}') - \mathbf{e} + x_{J_1}\mathbf{a}_{J_1} - c\mathbf{a}_j\|_2, \quad \text{supp}(\mathbf{z}) = \{i_2, \dots, i_k\}, \quad c \in \mathbb{C},$$

and hence by the law of cosine

$$r' \ge \min_{\mathbf{z},c} \sqrt{\|\mathbf{A}(\mathbf{z} - \mathbf{x}') - \mathbf{e}\|_2^2 + \|x_{J_1}\mathbf{a}_{J_1} - c\mathbf{a}_j\|_2^2 - 2|\left\langle \mathbf{A}(\mathbf{z} - \mathbf{x}') - \mathbf{e}, x_{J_1}\mathbf{a}_{J_1} - c\mathbf{a}_j\right\rangle|}, \tag{25}$$

where $\text{supp}(\mathbf{z}) = \{i_2, \dots, i_k\}$.

Because of (15), $j, J_1 \notin B_\eta(\text{supp}(\mathbf{x}') \cup \{i_2, \dots, i_k\})$. By the definition of the η-coherence band, we have

$$\left|\left\langle \mathbf{A}(\mathbf{z} - \mathbf{x}') - \mathbf{e}, x_{J_1}\mathbf{a}_{J_1} - c\mathbf{a}_j\right\rangle\right| \le (\varepsilon + \eta(s + k - 2))(|x_{J_1}| + |c|)$$

and hence by (25)

$$\begin{aligned} r' &\ge \min_{\mathbf{z},c} \sqrt{\|\mathbf{A}(\mathbf{z} - \mathbf{x}') - \mathbf{e}\|_2^2 + |x_{J_1}|^2 + |c|^2 - 2\eta|cx_{J_1}| - 2(\varepsilon + \eta(s + k - 2))(|x_{J_1}| + |c|)} \\ &\ge \sqrt{\min_{\mathbf{z}} \|\mathbf{A}(\mathbf{z} - \mathbf{x}') - \mathbf{e}\|_2^2 + \min_{c\in\mathbb{C}} [|x_{J_1}|^2 + |c|^2 - 2\eta|cx_{J_1}| - 2(\varepsilon + \eta(s + k - 2))(|x_{J_1}| + |c|)]}. \end{aligned}$$

To prove $r < r'$, it suffices to show

$$\begin{aligned} &\min_{c\in\mathbb{C}} \left[|x_{J_1}|^2 + |c|^2 - 2\eta|cx_{J_1}| - 2(\varepsilon + \eta(s + k - 2))(|x_{J_1}| + |c|)\right] \\ &= (1 - \eta^2)|x_{J_1}|^2 - 2(1 + \eta)(\varepsilon + \eta(s + k - 2))|x_{J_1}| - (\varepsilon + \eta(s + k - 2))^2 > 0, \end{aligned}$$

which leads to the inequality

$$|x_{J_1}| > (\varepsilon + (s + k - 2)\eta)\left(\frac{1}{1 - \eta} + \sqrt{\frac{1}{(1 - \eta)^2} + \frac{1}{1 - \eta^2}}\right).$$

Considering the worst case scenario, we replace $|x_{J_1}|$ by $x_{\min}$ and k by s to obtain the condition (24). ∎

Remark 4. For each $n = 1, \dots, s$, the LO algorithm involves least squares separately for each of the n coherence bands which can be accomplished by orthogonalizing the column vectors of each coherence band with respect to the column vectors of the other coherence bands and hence requires $\mathcal{O}(n^2\beta N)$ flops. In total, LO requires $\mathcal{O}(s^3\beta N)$ more flops, which is about s times the number of flops saved in BE (Remark 3).

Corollary 1. *Let* $\hat{\mathbf{x}}$ *be the output of BLOOMP. Under the assumptions of Theorems* 1 *and* 2, $\operatorname{supp}(\hat{\mathbf{x}}) \subseteq B_\eta(\operatorname{supp}(\mathbf{x}))$, *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 cannot improve the performance guarantee for BLOOMP, in practice the LO technique greatly enhances the success probability of recovery that BLOOMP has the best performance among all the algorithms tested with respect to noise stability and dynamic range (see section 5). In particular, the LO step greatly enhances the performance of BOMP with respect to dynamic range. Moreover, whenever Corollary 1 holds, for all practical purposes we have the residual bound for the BLOOMP reconstruction $\hat{\mathbf{x}}$

$$\|\mathbf{b} - \mathbf{A}\hat{\mathbf{x}}\|_2 \le c\|\mathbf{e}\|_2, \quad c \sim 1. \tag{26}$$

On the other hand, it is difficult to obtain bounds for the reconstruction error since $\|\mathbf{x} - \hat{\mathbf{x}}\|_2$ is not a meaningful error metric without *exact* recovery of an overwhelming majority of the object support.

4. Band-Excluded Thresholding (BET). The BE technique can be extended and applied to selecting s objects all at once in what is called the Band-Excluded Thresholding (BET).

We consider two forms of BET. The first is the Band-excluded Matched Thresholding (BMT), which is the band exclusion version of the One-Step Thresholding (OST) recently shown to possess CS capability under incoherence conditions [1].

For the purpose of comparison with BOMP, we give a performance guarantee for BMT under similar but weaker conditions than (15)–(16).

Algorithm 4. Band-excluded Matched Thresholding (BMT).

Input: $\mathbf{A}, \mathbf{b}, \eta > 0$
Initialization: $S^0 = \emptyset$
Iteration: For $k = 1, \ldots, s$
1. $i_k = \arg\max_j |\langle \mathbf{b}, \mathbf{a}_j \rangle|$, $j \notin B_\eta^{(2)}(S^{k-1})$.
2. $S^k = S^{k-1} \cup \{i_k\}$.

Output $\hat{\mathbf{x}} = \arg\min_{\mathbf{z}} \|\mathbf{A}\mathbf{z} - \mathbf{b}\|_2$ s.t. $\operatorname{supp}(\mathbf{z}) \subseteq S^s$.

Theorem 3. *Let* $\mathbf{x}$ *be* s*-sparse. Let* $\eta > 0$ *be fixed. Suppose that*

$$B_\eta(i) \cap B_\eta(j) = \emptyset \quad \forall i, j \in \operatorname{supp}(\mathbf{x}) \tag{27}$$

and that

$$\eta(2s-1)\frac{x_{\max}}{x_{\min}} + \frac{2\|\mathbf{e}\|_2}{x_{\min}} < 1, \tag{28}$$

where

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

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

Proof. Let $\mathrm{supp}(\mathbf{x}) = \{J_1, \ldots, J_s\}$. Let $J_{\max} \in \mathrm{supp}(\mathbf{x})$ be the index of the largest component of $\mathbf{x}$ in absolute value.

On the one hand, for $k = 1, \ldots, s$,

$$
\begin{aligned}
|\mathbf{b}^\star \mathbf{a}_k| &= |x_1 \mathbf{a}_1^\star \mathbf{a}_k + \cdots + x_{k-1}\mathbf{a}_{k-1}^\star \mathbf{a}_k + x_k + x_{k+1}\mathbf{a}_{k+1}^\star \mathbf{a}_k + \cdots + x_s \mathbf{a}_s^\star \mathbf{a}_k + \mathbf{e}^\star \mathbf{a}_k| \\
&\geq x_{\min} - (s-1)\eta x_{\max} - \|\mathbf{e}\|_2.
\end{aligned} \tag{29}
$$

On the other hand, $\forall l \notin B_\eta(\mathrm{supp}(\mathbf{x}))$,

$$
\begin{aligned}
|\mathbf{b}^\star \mathbf{a}_l| &= |x_1 \mathbf{a}_1^\star \mathbf{a}_l + x_2 \mathbf{a}_2^\star \mathbf{a}_l + \cdots + x_s \mathbf{a}_s^\star \mathbf{a}_l + \mathbf{e}^\star \mathbf{a}_l| \\
&\leq x_{\max} s\eta + \|\mathbf{e}\|_2.
\end{aligned} \tag{30}
$$

Therefore, the condition (28) implies that the right-hand side of (29) is greater than the right-hand side of (30). This means $|\mathbf{b}^* \mathbf{a}_k| > |\mathbf{b}^* \mathbf{a}_l|$ $\forall k = 1, \ldots, s, \forall l \notin B_\eta(\mathrm{supp}(\mathbf{x}))$, and hence the s highest points of $|\mathbf{b}^* \mathbf{a}_k|$ are in $B_\eta(\mathrm{supp}(\mathbf{x}))$.

From step 2 of BMT and (27) follows the second half of the statement, namely, that every nonzero component of $\hat{\mathbf{x}}$ is in the η-coherence band of a unique nonzero component of $\mathbf{x}$. ∎

Condition (27) roughly means that the objects are separated by at least 2 RLs, which is weaker than (15). In numerical simulations, however, BOMP performs far better than BMT. In other words, BMT is not a stand-alone algorithm but should instead be imbedded in other algorithms such as Subspace Pursuit (SP) [12], Compressive Sampling Matching Pursuit (CoSaMP) [26], and Normalized Iterative Hard Thresholding (NIHT) [3]. This gives rise to Band-excluded Subspace Pursuit (BSP), Band-excluded Compressive Sampling Matching Pursuit (BCoSaMP), and Band-excluded Normalized Iterative Hard Thresholding (BNIHT), whose performance we demonstrate numerically.

In addition to BMT, the second form of BET, namely Band-excluded Locally Optimized Thresholding (BLOT), can further enhance the performance in reconstruction with unresolved grids.

Algorithm 5. Band-excluded Locally Optimized Thresholding (BLOT).

Input: $\mathbf{x} = (x_1, \ldots, x_M)$, $\mathbf{A}, \mathbf{b}, \eta > 0$
Initialization: $S^0 = \emptyset$
Iteration: For $n = 1, 2, \ldots, s$
1. $i_n = \arg\max_j |x_j|$, $j \notin B_\eta^{(2)}(S^{n-1})$.
2. $S^n = S^{n-1} \cup \{i_n\}$.

Output: $\hat{\mathbf{x}} = \arg\min \|\mathbf{A}\mathbf{z} - \mathbf{b}\|_2$, $\mathrm{supp}(\mathbf{z}) \subseteq \mathrm{LO}(S^s)$, where LO is the output of Algorithm 2.

Now we state the algorithm Band-excluded Locally Optimized Subspace Pursuit (BLOSP). The Band-excluded Locally Optimized CoSaMP (BLOCoSaMP) is similar and is omitted here.

Embedding BLOT in NIHT turns out to have a nearly identical performance to embedding BLOT in Iterative Hard Thresholding (IHT) [2]. Since the latter is simpler to implement and more efficient to compute, we state the resulting algorithm, the Band-excluded Locally Optimized IHT (BLOIHT), as Algorithm 7.

Algorithm 6. Band-excluded Locally Optimized Subspace Pursuit (BLOSP).

Input: $\mathbf{A}, \mathbf{b}, \eta > 0$
Initialization: $\mathbf{x}^0 = 0, \mathbf{r}^0 = \mathbf{b}$
Iteration: For $n = 1, 2, \ldots$
1. $\tilde{S}^n = \text{supp}(\mathbf{x}^{n-1}) \cup \text{supp}(\text{BMT}(\mathbf{r}^{n-1}))$,
   where $\text{BMT}(\mathbf{r}^{n-1})$ is the output of Algorithm 4 with data $\mathbf{r}^{n-1}$.
2. $\tilde{\mathbf{x}}^n = \arg\min \|\mathbf{Az} - \mathbf{b}\|_2$ s.t. $\text{supp}(\mathbf{z}) \subseteq \tilde{S}^n$.
3. $S^n = \text{supp}(\text{BLOT}(\tilde{\mathbf{x}}^n))$, where $\text{BLOT}(\tilde{\mathbf{x}}^n)$ is the output of Algorithm 5.
4. $\mathbf{r}^n = \min_{\mathbf{z}} \|\mathbf{Az} - \mathbf{b}\|_2, \text{supp}(\mathbf{z}) \subseteq S^n$.
5. If $\|\mathbf{r}^{n-1}\|_2 \leq \varepsilon$ or $\|\mathbf{r}^n\|_2 \geq \|\mathbf{r}^{n-1}\|_2$, then quit and set $S = S^{n-1}$;
   otherwise continue iteration.

Output: $\hat{\mathbf{x}} = \arg\min_{\mathbf{z}} \|\mathbf{Az} - \mathbf{b}\|_2$ s.t. $\text{supp}(\mathbf{z}) \subseteq S$.

Algorithm 7. Band-excluded Locally Optimized Iterative Hard Thresholding (BLOIHT).

Input: $\mathbf{A}, \mathbf{b}, \eta > 0$
Initialization: $\hat{\mathbf{x}}^0 = 0, \mathbf{r}^0 = \mathbf{b}$
Iteration: For $n = 1, 2, \ldots$
1. $\mathbf{x}^n = \text{BLOT}(\mathbf{x}^{n-1} + \mathbf{A}^*\mathbf{r}^{n-1})$, where BLOT denotes the output of Algorithm 5.
2. If $\|\mathbf{r}^{n-1}\|_2 \leq \varepsilon$ or $\|\mathbf{r}^n\|_2 \geq \|\mathbf{r}^{n-1}\|_2$, then quit and set $S = S^{n-1}$;
   otherwise continue iteration.

Output: $\hat{\mathbf{x}}$.

In addition, the technique BLOT can be used to enhance the recovery capability with unresolved grids of the L^1-minimization principles, Basis Pursuit (BP), i.e.,

$$\min_{\mathbf{z}} \|\mathbf{z}\|_1 \quad \text{s.t.} \quad \mathbf{b} = \mathbf{Az}, \tag{31}$$

and Lasso, i.e.,

$$\min_{\mathbf{z}} \frac{1}{2}\|\mathbf{b} - \mathbf{Az}\|_2^2 + \lambda\sigma\|\mathbf{z}\|_1, \tag{32}$$

where σ is the standard deviation of each noise component and λ is the regularization parameter. In this case, BLOT is applied to the BP and Lasso estimates $\hat{\mathbf{x}}$ to produce an s-sparse reconstruction. The resulting algorithms are called BP-BLOT and Lasso-BLOT, respectively.

The thresholded Lasso, the Lasso followed by a hard thresholding, has been considered previously (see [24] and references therein). The novelty of our version lies in the BE and LO steps, which greatly enhance the performance in dealing with unresolved grids.

5. Numerical results.

We test our various band exclusion algorithms on the matrix

$$\frac{e^{-2\pi i(l-1)\xi_k/F}}{\sqrt{N}}, \quad k = 1, \ldots, N, \ l = 1, \ldots, RF, \tag{33}$$

where ξ_k is uniformly and independently distributed in $(0,1)$. When $F=1$, $\mathbf{A}$ is the random partial Fourier matrix analyzed in [28], and, with a sufficient number of samples, successful recovery with $\mathbf{A}$ is guaranteed with high probability. For large F, however, $\mathbf{A}$ has a high mutual coherence. Unless otherwise stated, we use $N=100$, $M=4000$, $F=20$, and the independent and identically distributed (i.i.d.) Gaussian noise $\mathbf{e}\sim N(0,\sigma^2 I)$ in our simulations. Recall the decay profile of pairwise coherence as the index separation increases (Figure 2(b)). The η-coherence band is about 0.7 RL in half-width with $\eta=0.3$.

We compare the performance of various algorithms in terms of success probability versus dynamic range, signal-to-noise ratio (SNR), number of measurements, and resolution. Here and below SNR is defined as the ratio of the 2-norm of the data vector to the 2-norm of noise. Unless otherwise stated, we use in our simulations 10 randomly phased and located objects, separated by at least 3 RLs. A reconstruction is counted as a success if every reconstructed object is within 1 RL 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 RL.

For two subsets A and B in $\mathbb{R}^d$ of the same cardinality, the Bottleneck distance $d_B(A,B)$ is defined as

$$d_B(A,B)=\min_{f\in\mathcal{M}}\max_{a\in A}|a-f(a)|,$$

where $\mathcal{M}$ is the collection of all one-to-one mappings from A to B.

For subsets in one dimension, the Bottleneck distance can be calculated easily. Let $A=\{a_1,\dots,a_n\}$ and $B=\{b_1,\dots,b_n\}$ be listed in ascending order. Then

$$d_B(A,B)=\max_j|a_j-b_j|.$$

In higher dimensions, however, it is more costly to compute the Bottleneck distance [18, 22]. The Bottleneck distance is a stricter metric than the Hausdorff distance, which does not require one-to-one correspondence between the two target sets. We do not use 2-norm to measure the error in support reconstruction as 2-norm cannot distinguish large from small displacements.

In the first set of experiments, we test various greedy algorithms equipped with the BE step (only). This includes BOMP, BSP, BCoSaMP, and BNIHT. For comparison, we also show the performance of OMP without BE.

As shown in Figure 3, BOMP has the best performance with respect to dynamic range, SNR, and, in the case of higher dynamic range (≥ 3; see Figure 3(f)), number of measurements. In the case of dynamic range equal to 1, BSP is the best performer in terms of number of measurements, followed closely by BCoSaMP and BNIHT (Figure 3(e)). In the case of dynamic range equal to 1, BOMP and OMP have almost identical performance with respect to SNR (Figure 3(c)) and number of measurements (Figure 3(e)). The performance of BSP and BCoSaMP, however, depends crucially on the BE step without which both SP and CoSaMP fail catastrophically (not shown).

In the next set of experiments, we test BLO-based algorithms. For the purpose of comparison, we also test the algorithm LOOMP, which is the same as Algorithm 3 but without BE.

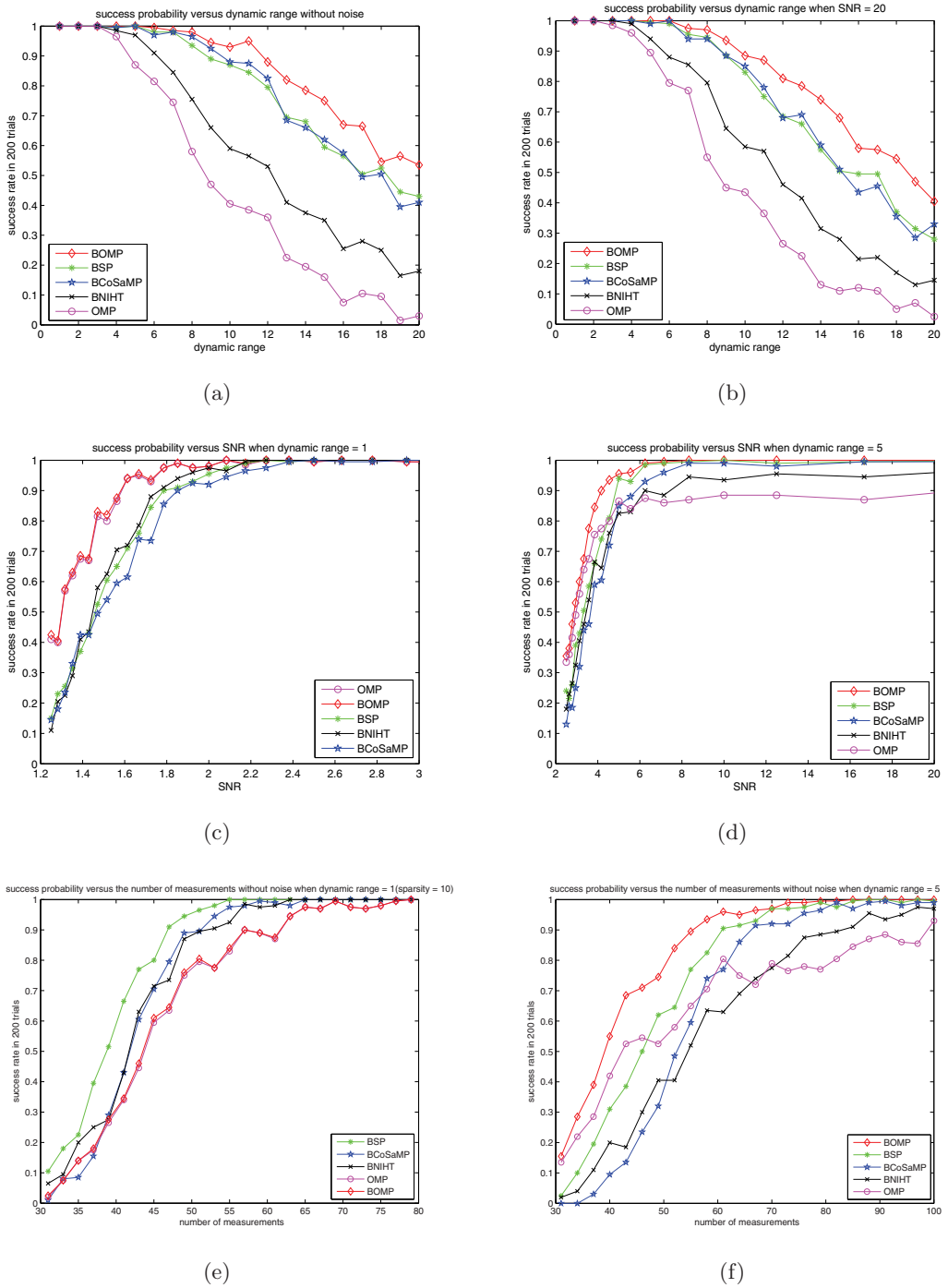
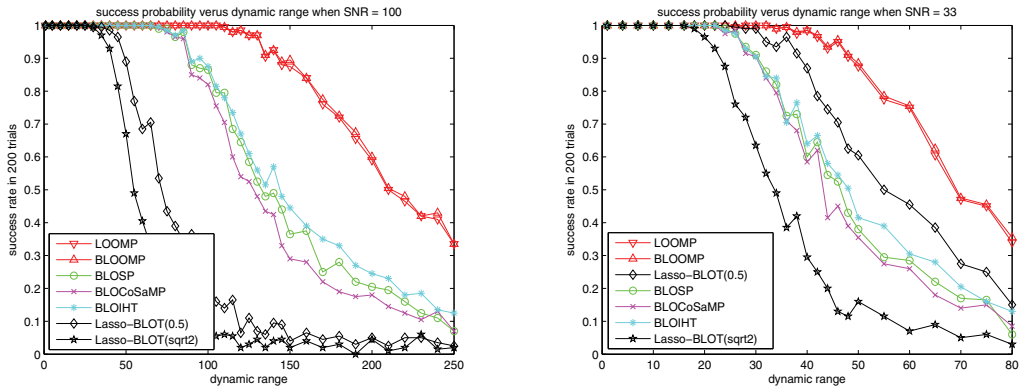


Figure 3. *Success probability of various band exclusion algorithms versus dynamic range for SNR* $= \infty$ (a) *and* 20 (b) *noise, versus SNR for dynamic range* 1 (c) *and* 5 (d)*, and versus number of noiseless measurements for dynamic range* 1 (e) *and* 5 (f).



(a) (b)

Figure 4. *Success probability of BLO-based algorithms versus dynamic range with SNR* 100 (a) *and* 33 (b).

Lasso-BLOT is implemented with the regularization parameter

$$\lambda = 0.5\sqrt{\log M} \tag{34}$$

or

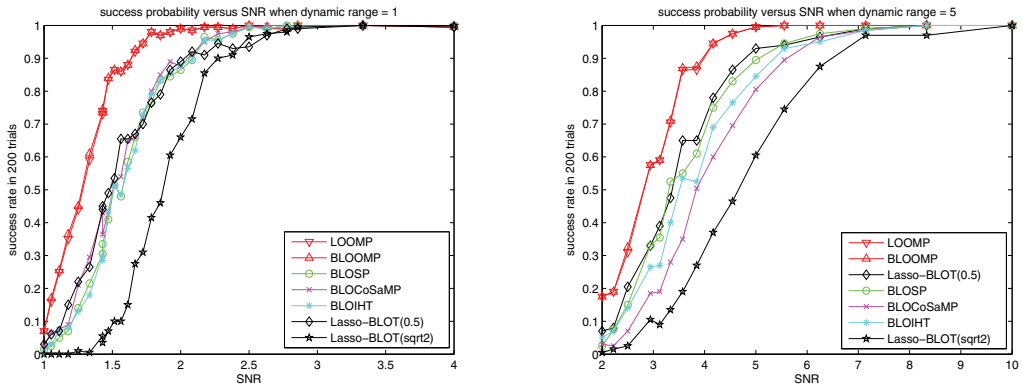
$$\lambda = \sqrt{2\log M} \tag{35}$$

which is proposed in [9]. Other larger values have been proposed in [6, 7]. Our numerical experiments indicate that for matrix (33) with large F the choice (34) is nearly optimal among all $\lambda/\sqrt{\log M} \leq 10$ and SNR ≥ 20. The superiority of the choice (34) to (35) (and other choices) manifests clearly across all performance metrics.

Figure 4 shows success probability versus dynamic range in the presence of noise. The top performers are LOOMP and BLOOMP, both of which can handle large dynamic range. In the noiseless case, the success rate for LOOMP, BLOOMP, BLOSP, BLOOMP, and BLOCoSaMP stays near unity for dynamic range up to as high as 10^{14}. For SNR = 100 (Figure 4(a)), BLOSP, BLOCoSaMP, and BLOIHT perform better than Lasso-BLOT with either (34) or (35), while for SNR = 33 (Figure 4(b)), BLOSP, BLOCoSaMP, and BLOIHT performance curves have dropped below that of Lasso-BLOT with (34). But the noise stability of Lasso-BLOT never catches up with that of LOOMP/BLOOMP even as the noise level increases, as can be seen in Figure 5.

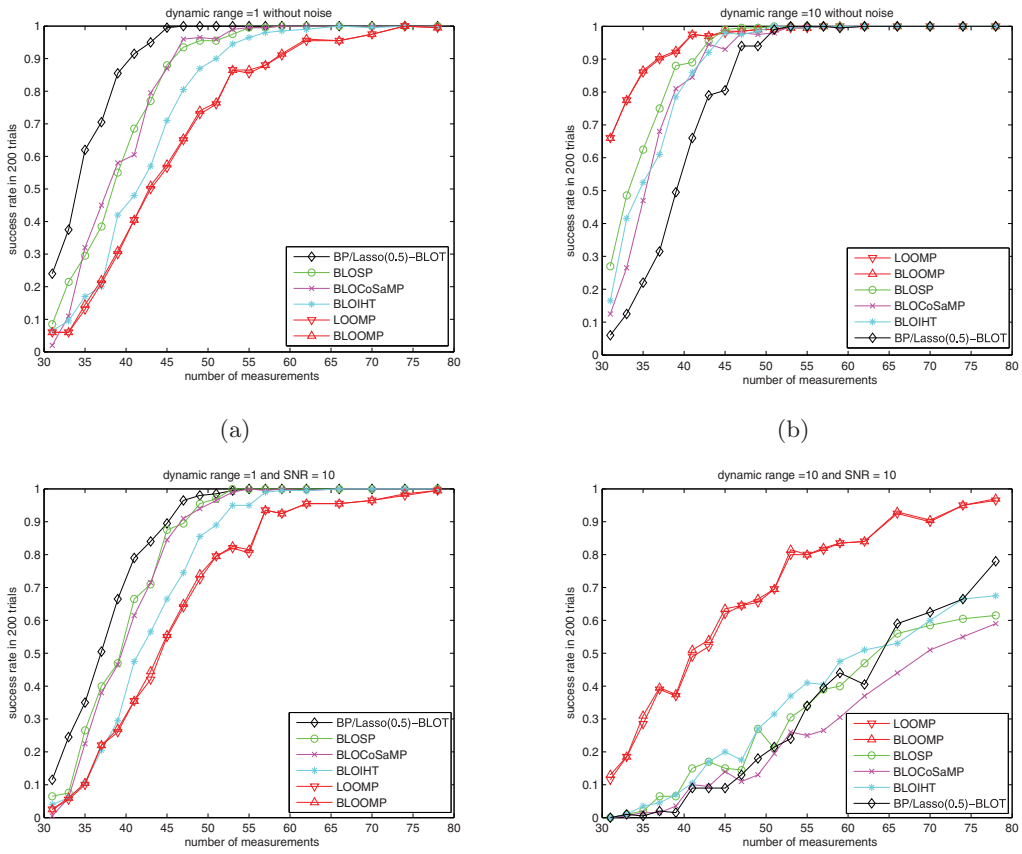
Figure 5 shows that LOOMP and BLOOMP remain the top performers with respect to noise, while Lasso-BLOT with (35) has the worst performance. Lasso-BLOT with (34), however, is a close second in noise stability. As seen in Figures 4 and 5, the performance of Lasso-BLOT depends significantly on the choice of the regularization parameter.

With respect to number of measurements (Figure 6), BP/Lasso-BLOT with (34) is the best performer, followed closely by BLOSP and BLOCoSaMP for dynamic range 1 (Figures 6(a) and 6(c)), while for dynamic range 10, BLOOMP and LOOMP perform significantly better than the rest (Figures 6(b) and 6(d)). As is clear from the comparison of Figures 6(a)



(a) (b)

Figure 5. *Success probability of BLO-based algorithms versus SNR for dynamic range* 1 (a) *and* 5 (b).



(a) (b)

(c) (d)

Figure 6. *Success probability versus the number of measurements with* (a) *dynamic range* 1, *without noise,* (b) *dynamic range* 10, *without noise,* (c) *dynamic range* 1, $SNR = 10$, *and* (d) *dynamic range* 10, $SNR = 10$.

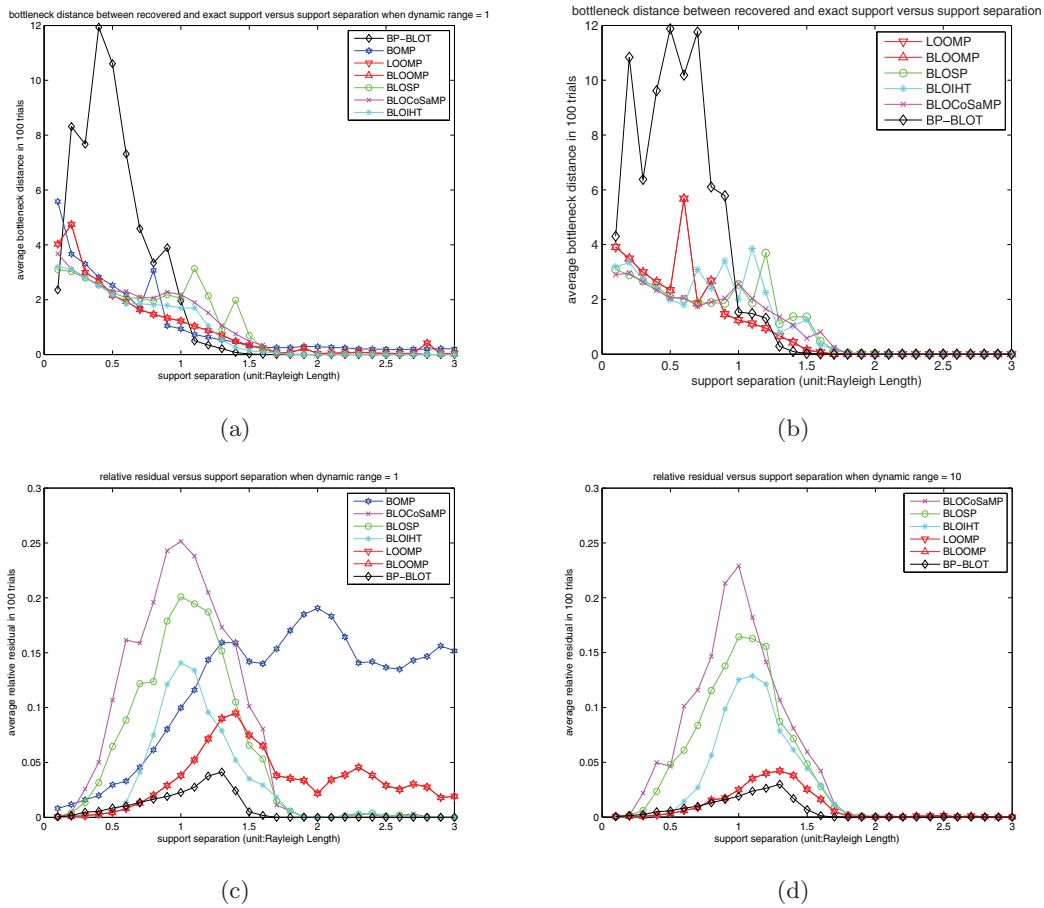
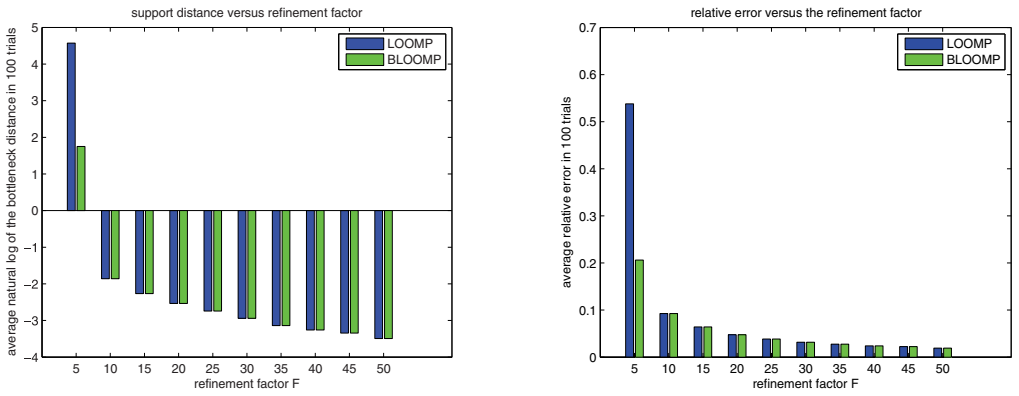


Figure 7. *The average Bottleneck distance for dynamic range* 1 (a) *and* 10 (b)*, and the relative residual for dynamic range* 1 (c) *and* 10 (d) *versus the minimum separation of objects. Note the reduction of error in LOOMP/BLOOMP for larger dynamic range when the minimum separation is greater than* 1.5 *RL.*

and 6(b), at a low level of noise the performance of BLOOMP and LOOMP improves significantly as the dynamic range increases from 1 to 10. In the meantime, the performance of BLOSP, BCoSaMP, and BLOIHT improves slightly, while the performance of BP/Lasso-BLOT deteriorates. With SNR $= 10$, however, the performance of BLOOMP and LOOMP is roughly unchanged as dynamic range increases, while the performance of all other algorithms deteriorates significantly (Figure 6(d)).

Next we compare the resolution performance of the various algorithms for 10 randomly phased objects of unit dynamic range in the absence of noise. The 10 objects are consecutively located and separated by equal length varying from 0.1 to 3 RLs. The whole object support is, however, randomly shifted for each of the 100 trials. For closely spaced objects, it is necessary to modify the BE and LO rules: If h is the object spacing, we use $h/2$ to replace 2 RLs of the original BE rule and 1 RL of the original LO rule.

Figure 7 shows the averaged Bottleneck distance between the reconstruction and the true object support (Figures 7(a) and 7(b)) and the residual (Figures 7(c) and 7(d)) as a function



(a) (b)

Figure 8. *The reconstruction error of LOOMP and BLOOMP as measured by the Bottleneck distance (*(a)*, semi-log) and the relative* 2*-norm* (b) *for* 10 *objects of dynamic range* 10 *as a function of* F.

of the object spacing. For this class of objects, BP-BLOT has the best resolution for dynamic range up to 10, followed closely by BLOIHT for dynamic range 1 and by BLOOMP/LOOMP for dynamic range 10. The *high precision* (i.e., nearly zero Bottleneck distance) resolution ranges from about 1.5 RLs for BP-BLOT to about 1.7 RLs for the rest. Consistent with what we have seen in Figure 6, the resolving power of BLOOMP/LOOMP improves significantly as the dynamic range increases, while that of BP-BLOT deteriorates. Note that in the case of unit dynamic range, BOMP recovers the support as well as BLOOMP/LOOMP does. BOMP, however, produces a high level of residual even when objects are widely separated. There is essentially no difference between the resolving power of BLOOMP and LOOMP.

It is noteworthy that the relative residuals of all tested algorithms peak at a separation of 1–1.5 RLs and decline to zero as the object separation decreases. In contrast, the average Bottleneck distances increase as the separation decreases, except for BP-BLOT when the separation drops below 0.5 RL. When the separation drops below 1 RL, the Bottleneck distance between the objects and the reconstruction indicates that the objects are not well recovered by any of the algorithms (Figures 7(a) and 7(b)). The vanishing residuals in this regime indicate nonuniqueness of sparse solutions.

Figures 4–7 show negligible difference between the performances of LOOMP and BLOOMP. To investigate their distinction more closely, we test the stability with respect to the gridding error for various F's (cf. Figure 1). We consider randomly phased objects of dynamic range 10 that are randomly located in $[0, 1000)$ and separated by at least 3 RLs. We compute the reconstruction error in the Bottleneck distance and 2-norm averaged over 100 trials and show the result in Figure 8. Evidently the advantage of BLOOMP over LOOMP lies in the cases when the refinement factor F is less than 10 and the gridding error is sufficiently large. When $F \geq 10$, the difference between their performances with respect to gridding error is negligible. On the other hand, for $F = 5$ both BLOOMP's, and LOOMP's reconstructions would have been considered a failure given the magnitudes of error in the Bottleneck distance due to a large gridding error.

6. Comparison with other algorithms in the literature. The present work is inspired by the performance guarantee established in [20] that the MUSIC algorithm aided by BMT produces a support estimate that is within 1 RL of the locations of sufficiently separated objects.

In comparison to other CS literature on coherent and redundant dictionaries, our work resembles that of [5] and [17], albeit with a different perspective.

Following [17], we consider the problem

$$\mathbf{b} = \mathbf{\Phi y} + \mathbf{e}, \tag{36}$$

where $\mathbf{\Phi}$ is an $N \times R$ i.i.d. Gaussian matrix of mean 0 and variance σ^2. The signal to be recovered is given by $\mathbf{y} = \mathbf{\Psi x}$, where $\mathbf{\Psi}$ is the oversampled, redundant DFT frame

$$\Psi_{k,j} = \frac{1}{\sqrt{R}} e^{-2\pi i \frac{(k-1)(j-1)}{RF}}, \quad k = 1, \dots, R, \quad j = 1, \dots, RF. \tag{37}$$

As before, F is the refinement factor. Combining (36) and (37), we have the same form (5) with $\mathbf{A} = \mathbf{\Phi\Psi}$. The coherence bands of $\mathbf{\Psi}$ and $\mathbf{A}$ are shown in Figure 9.

In the simulation, unless otherwise stated, we take $N = 100, R = 200, F = 20$, and $\sigma = \frac{1}{\sqrt{N}}$ so that $\mathbf{A} \in \mathbb{C}^{100\times 4000}$ as before. We use randomly located and phased $\mathbf{x} = (x_j)$ which are well separated in the sense that $|x_j - x_k| \geq 3\ \forall j \neq k$.

Spectral Iterative Hard Thresholding (SIHT), the algorithm proposed in [17], assumes the model-based RIP which, in spirit, is equivalent to the assumption of well-separated support in the synthesis coefficients. The frame-adapted BP proposed in [5],

$$\min \|\mathbf{\Psi}^\star \mathbf{z}\|_1 \qquad \text{s.t. } \|\mathbf{\Phi z} - \mathbf{b}\|_2 \leq \|\mathbf{e}\|_2, \tag{38}$$

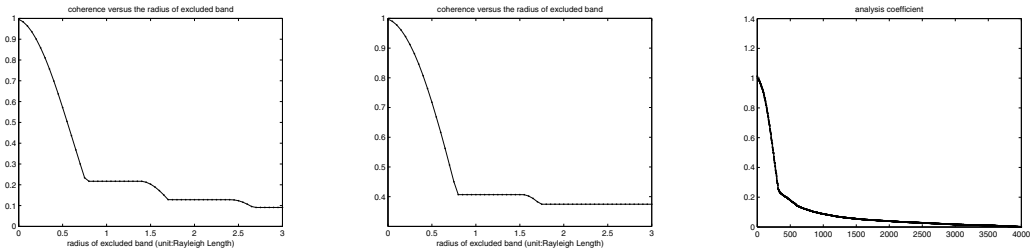
assumes a measurement matrix $\mathbf{\Phi}$ satisfying the frame-adapted RIP

$$(1-\delta)\|\mathbf{\Psi z}\|_2^2 \leq \|\mathbf{\Phi\Psi z}\|_2^2 \leq (1+\delta)\|\mathbf{\Psi} z\|_2^2, \quad \|\mathbf{z}\|_0 \leq 2s, \tag{39}$$

as well as the sparsity or compressibility of the analysis coefficients $\mathbf{\Psi}^*\mathbf{y}$. The latter condition is the main difficulty in applying (38) to the problem at hand, despite the powerful error bound established in [5].

For example, Figure 9(c) shows the absolute values of the component of the vector $\mathbf{\Psi}^*\mathbf{y}$ in order of descending magnitude. Clearly $\mathbf{\Psi}^*\mathbf{y}$ is neither sparse nor compressible. The shape of the curve can be roughly understood as follows. 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 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. The long tail of the curve is due to the fact that the pairwise coherence of $\mathbf{\Psi}$ decays slowly as the separation increases. Therefore the analysis approach (38) requires a far higher number of measurements than 100 for accurate reconstruction, which is roughly proportional to the redundancy ratio.

In the language of digital signal processing (38) is an L^1-analysis method, while the standard BP or Lasso and their BLOT-aided versions are L^1-synthesis methods. Both methods are



(a) (b) (c)

Figure 9. *The coherence bands of the DFT frame* $\mathbf{\Psi}$ (a) *and* $\mathbf{A} = \mathbf{\Phi\Psi}$ (b)*, the latter being averaged over* 100 *trials.* (c) *shows the magnitudes of the analysis coefficient vector* $\mathbf{\Psi}^*\mathbf{y}$ *in descending order of magnitude. In other words, the horizontal axis counts the number of coefficients greater than or equal to the value of the curve.*

based on the same principle of representational sparsity. In principle, the synthesis approach (such as BP, Lasso, and all BLO-based algorithms) is more general than the analysis approach since every analysis method can be recast as a synthesis method, while many synthesis formulations have no equivalent analysis form [19].

For the problem of sparse representation by redundant dictionaries, the main quantities of interest are $\mathbf{y}$, not the synthesis coefficients $\mathbf{x}$. So in our comparison experiments, 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 1 or 10 and i.i.d. Gaussian $\mathbf{\Phi}$ are generated. We compute the averaged relative errors, as dynamic range, SNR, and the number of measurements vary.

The results are shown in Figure 10. The best performers with respect to dynamic range are BLOOMP, BLOSP, and BP-BLOT achieving 10^{-16} relative error (Figure 10(a)). Contrary to the case of support recovery, the errors generally stay constant independent of dynamic range. Indeed, with the exception of SIHT, the errors are decreasing in the dynamic range from 1 to 10 (Figure 10(a)). With respect to noise, the smallest error is achieved by BLOOMP, followed closely by BLOSP and Lasso-BLOT with (34) (Figure 10(b)). BP-BLOT is the best performer with respect to sparsity followed closely by BP when the dynamic range is 1 (Figure 10(c)). When the dynamic range increases to 10, BLOOMP is the best performer, followed closely by BP-BLOT (Figure 10(d)). Note the significant reduction in error for BLOOMP when the dynamic range increases from 1 to 10. Note also that the difference between the errors committed by BP and BP-BLOT increases when the dynamic range increases from 1 to 10. The SIHT algorithm requires a much higher number of measurements to get its error down (Figure 10(c)) and produces the highest level of error with respect to dynamic range (Figure 10(a)) and noise (Figure 10(b)).

In Figure 10, we include the performance curves of OMP with various numbers of iterations (s, $2s$, $5s$) as well as the standard BP/Lasso without BLOT. Not surprisingly, the relative error of OMP reconstruction decreases as the number of iterations increases.

It is noteworthy that the BLOT technique reduces the BP/Lasso reconstruction errors: BP without BLOT produces 0.5% relative error with respect to dynamic range (Figure 10(a), not clearly visible), while BP-BLOT produces 10^{-16} relative error across a wide dynamic

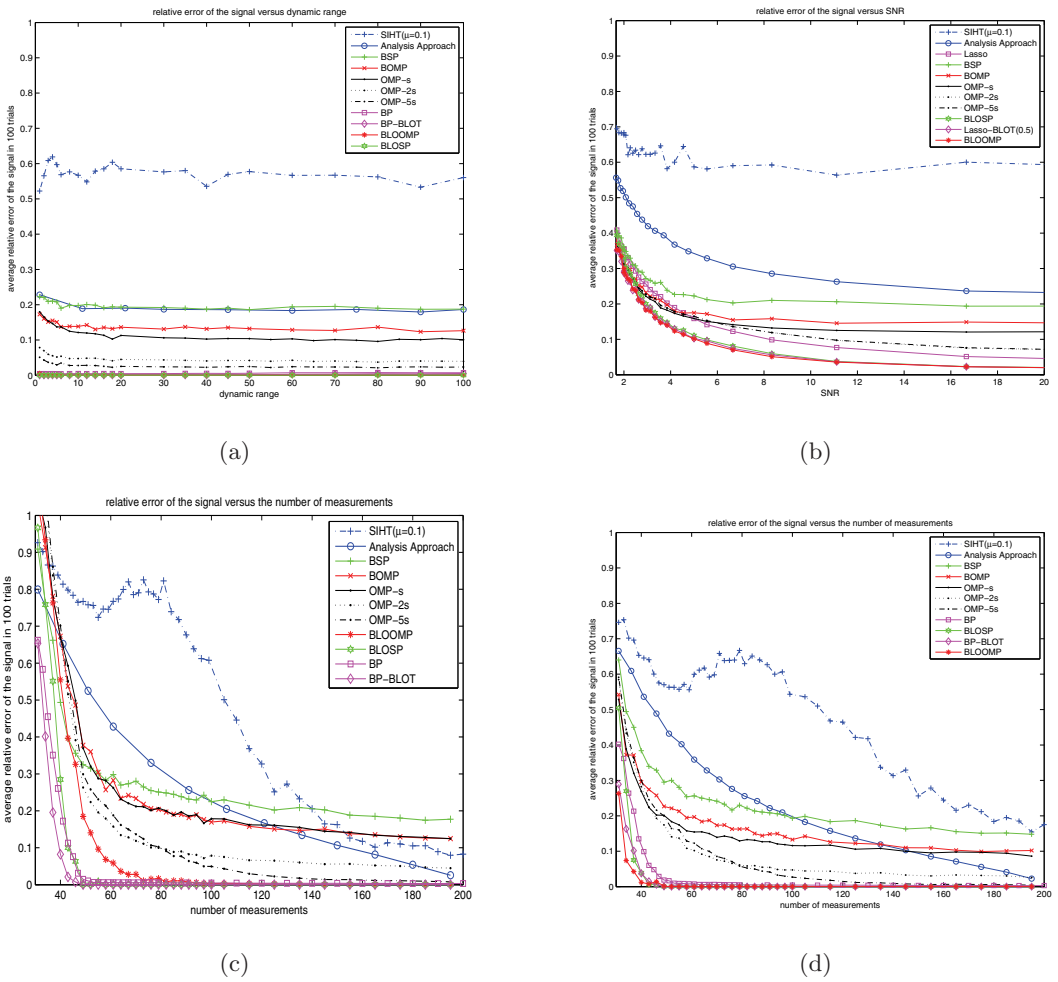


Figure 10. (a) *Relative errors versus dynamic range.* (b) *SNR in the case of dynamic range* 10. *Relative errors versus the number of measurements for dynamic range* 1 (c) *and* 10 (d). *Note the significant reduction in the number of measurements for BLOOMP (red asterisk) as the dynamic range increases from* 1 *to* 10. *The results for the frame-adapted BP* (38) *are labelled as "Analysis Approach," which requires more than* 200 *measurement data for accurate recovery.*

range. Also, BP-BLOT produces smaller errors than BP which, in turn, produces smaller errors than frame-adapted BP (38). Moreover, Lasso with the optimized parameter (34) but without BLOT produces significantly greater errors than Lasso-BLOT, BLOOMP, and BLOSP (Figure 10(b)).

To understand the superior performance of BLOOMP, let us give an error bound using (26) and (39):

$$\|\mathbf{\Psi}(\mathbf{x}-\hat{\mathbf{x}})\|_2 \le \frac{1}{1-\delta}\|\mathbf{A}(\mathbf{x}-\hat{\mathbf{x}})\|_2 \le \frac{1}{1-\delta}\|\mathbf{b}-\mathbf{e}-\mathbf{A}\hat{\mathbf{x}}\|_2 \le \frac{1+c}{1-\delta}\|\mathbf{e}\|_2, \tag{40}$$

where $\hat{\mathbf{x}}$ is the output of BLOOMP. This implies that the reconstruction error of BLOOMP is

roughly proportional to the external noise, consistent with Figure 10(b), and is independent of the dictionary redundancy. The enormous advantage of BLOOMP over BOMP, however, gradually disappears as SNR decreases (Figure 10(a)).

7. Conclusions. We have developed and tested various algorithms for sparse recovery with highly coherent sensing matrices arising in discretization of continuum imaging problems such as radar and medical imaging when the grid spacing is below the Rayleigh threshold [14].

We have introduced two essential techniques to deal with unresolved grids: band exclusion (BE) and local optimization (LO). We have embedded these techniques in various CS algorithms and performed systematic tests on them. When embedded in OMP, both BE and LO steps manifest their advantage in dealing with larger dynamic range. Moreover, when both BE and LO are implemented optimally, the increase in computational complexity by LO is more than compensated for by the decrease in complexity by BE. We have established a performance guarantee for BOMP and BLOOMP, whose performance is *grid independent*. When embedded in SP, CoSaMP, IHT, BP, and Lasso, the effects are more dramatic.

We have studied these modified algorithms from four performance metrics: dynamic range, noise stability, sparsity, and resolution. With respect to the first two metrics (dynamic range and noise stability), BLOOMP is the best performer. With respect to sparsity, BLOOMP is the best performer for high dynamic range, while for dynamic range near unity, BP-BLOT and Lasso-BLOT with the optimal regularization parameter have the best performance. BP-BLOT also has the highest resolving power up to a certain dynamic range. Lasso-BLOT's performance, however, is sensitive to the choice of regularization parameter.

One of the most surprising attributes of BLOOMP is its improved performance in terms of sparsity with *increased* dynamic range.

The algorithms BLOSP, BLOCoSaMP, and BLOIHT are good alternatives to BLOOMP and BP/Lasso-BLOT: They are faster than both BLOOMP and BP/Lasso-BLOT and share, to a lesser degree, BLOOMP's desirable attribute with respect to dynamic range.

Comparisons with existing algorithms (SIHT and frame-adapted BP) demonstrate the superiority of BLO-based algorithms for reconstruction of sparse objects separated by more than 1 RL.

Finally, to add to the debate of analysis versus synthesis [19], the performance of BLO-based algorithms for sparse, widely separated objects is independent of the refinement factor F representing redundancy, and, since the discretization error decreases with F, the reconstruction errors of the BLO-based synthesis methods also decrease with F in stark contrast to the examples presented in [19] which show that the synthesis approach degrades with redundancy.

REFERENCES

[1] 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., 12 (2010), pp. 289–307.

[2] T. Blumensath and M.E. Davies, *Iterative hard thresholding for compressed sensing*, Appl. Comput. Harmon. Anal., 27 (2009), pp. 265–274.

[3] T. Blumensath and M.E. Davies, *Normalized iterative hard thresholding: Guaranteed stability and performance*, IEEE J. Sel. Topics Signal Process., 4 (2010), pp. 298–309.

[4] A.M. Bruckstein, D.L. Donoho, and M. Elad, *From sparse solutions of systems of equations to sparse modeling of signals and images*, SIAM Rev., 51 (2009), pp. 34–81.

[5] E.J. Candès, Y.C. Eldar, D. Needell, and P. Randall, *Compressed sensing with coherent and redundant dictionaries*, Appl. Comput. Harmon. Anal., 31 (2011), pp. 59–73.
[6] E.J. Candès and Y. Plan, *Near-ideal model selection by ℓ_1 minimization*, Ann. Statist., 37 (2009), pp. 2145–2177.
[7] E.J. Candès and Y. Plan, *A probabilistic and RIPless theory of compressed sensing*, IEEE Trans. Inform. Theory, 57 (2011), pp. 7235–7254.
[8] E.J. Candès and T. Tao, *Near-optimal signal recovery from random projections: Universal encoding strategies*, IEEE Trans. Inform. Theory, 52 (2006), pp. 5406–5425.
[9] S.S. Chen, D.L. Donoho, and M.A. Saunders, *Atomic decomposition by basis pursuit*, SIAM Rev., 43 (2001), pp. 129–159.
[10] M. Cheney and B. Borden, *Fundamentals of Radar Imaging*, SIAM, Philadelphia, 2009.
[11] D. Colton and R. Kress, *Inverse Acoustic and Electromagnetic Scattering Theory*, 2nd ed., Springer, Berlin, 1998.
[12] W. Dai and O. Milenkovic, *Subspace pursuit for compressive sensing and reconstruction*, IEEE Trans. Inform. Theory, 55 (2009), pp. 2230–2249.
[13] G.M. Davis, S. Mallat, and M. Avellaneda, *Adaptive greedy approximations*, Constr. Approx., 13 (1997), pp. 57–98.
[14] D.L. Donoho, *Superresolution via sparsity constraints*, SIAM J. Math. Anal., 23 (1992), pp. 1309–1331.
[15] D.L. Donoho, *Compressed sensing*, IEEE Trans. Inform. Theory, 52 (2006), pp. 1289–1306.
[16] D.L. Donoho, M. Elad, and V.N. Temlyakov, *Stable recovery of sparse overcomplete representations in the presence of noise*, IEEE Trans. Inform. Theory, 52 (2006), pp. 6–18.
[17] M.F. Duarte and R.G. Baraniuk, *Spectral compressive sensing*, Appl. Comput. Harmon. Anal., submitted.
[18] A. Efrat and A. Itai, *Improvements on bottleneck matching and related problems using geometry*, in Proceedings of the Twelfth Annual Symposium on Computational Geometry, ACM, New York, 1996, pp. 301–310.
[19] M. Elad, P. Milanfar, and R. Rubinstein, *Analysis versus synthesis in signal prior*, Inverse Problems, 23 (2007), pp. 947–968.
[20] A. Fannjiang, *The MUSIC algorithm for sparse objects: A compressed sensing analysis*, Inverse Problems, 27 (2011), 035013.
[21] A.C. Fannjiang, T. Strohmer, and P. Yan, *Compressed remote sensing of sparse objects*, SIAM J. Imaging Sci., 3 (2010), pp. 595–618.
[22] P. Indyk and S. Venkatasubramanian, *Approximate congruence in nearly linear time*, Comput. Geom., 24 (2003), pp. 115–128.
[23] S. Mallat and Z. Zhang, *Matching pursuit with time-frequency dictionaries*, IEEE Trans. Signal Process., 41 (1993), pp. 3397–3415.
[24] N. Meinshausen and B. Yu, *Lasso-type recovery of sparse representation for high-dimensional data*, Ann. Statist., 37 (2009), pp. 2246–2270.
[25] B.K. Natarajan, *Sparse approximate solutions to linear systems*, SIAM J. Comput., 24 (1995), pp. 227–234.
[26] D. Needell and J.A. Tropp, *CoSaMP: Iterative signal recovery from incomplete and inaccurate samples*, Appl. Comput. Harmon. Anal., 26 (2009), pp. 301–329.
[27] Y.C. Pati, R. Rezaiifar, and P.S. Krishnaprasad, *Orthogonal matching pursuit: Recursive function approximation with applications to wavelet decomposition*, in Proceedings of the 27th Asilomar Conference on Signals, Systems and Computers, 1993.
[28] H. Rauhut, *Stability results for random sampling of sparse trigonometric polynomials*, IEEE Trans. Inform. Theory, 54 (2008), pp. 5661–5670.
[29] R. Tibshirani, *Regression shrinkage and selection via the lasso*, J. Roy. Statist. Soc. Ser. B, 58 (1996), pp. 267–288.