3 Compressive Sensing Theory for Optical Systems Described by a Continuous Model

Albert Fannjiang

Contents

Introduction

A monochromatic wave u propagating in a heterogeneous medium is governed by the following Helmholtz equation:

$$\Delta u(\mathbf{r}) + \omega^2(1 + \nu(\mathbf{r}))u(\mathbf{r}) = 0, \quad \mathbf{r} \in \mathbb{R}^d,\ d = 2, 3 \tag{3.1}$$

where $\nu \in \mathbb{C}$ describes the medium heterogeneities. For simplicity, we choose the physical units such that the wave velocity is unity and the wavenumber equals the frequency ω.

The data used for imaging are the scattered field $u^s = u - u^i$ governed by

$$\Delta u^s + \omega^2 u^s = -\omega^2 \nu u, \tag{3.2}$$

or equivalently the Lippmann–Schwinger integral equation:

$$u^s(\mathbf{r}) = \omega^2 \int_{\mathbb{R}^3} \nu(\mathbf{r}')\ (u^i(\mathbf{r}') + u^s(\mathbf{r}'))G(\mathbf{r}, \mathbf{r}')d\mathbf{r}'. \tag{3.3}$$

Here

$$G(\mathbf{r}, \mathbf{r}') = \begin{cases} \dfrac{e^{i\omega|\mathbf{r}-\mathbf{r}'|}}{4\pi|\mathbf{r}-\mathbf{r}'|}, & d = 3 \\ \dfrac{i}{4} H_0^{(1)}(\omega|\mathbf{r}-\mathbf{r}'|), & d = 2, \end{cases} \tag{3.4}$$

is the Green function for the background propagator $(\Delta + \omega^2)^{-1}$, where $H_0^{(1)}$ is the zeroth-order Hankel function of the first kind.

We consider two far-field imaging geometries: paraxial and scattering. In the former, both the object plane and the image plane are orthogonal to the optical axis, while in the latter emission and detection of light can take any directions. We take u^s as the measured data in the former and the scattering amplitudes (see Equation 3.7) as the measured data in the latter (Figure 3.1).

- *Paraxial geometry*: For simplicity, let us state the 2D version. Let $\{z = z_0\}$ be the object line and $\{z = 0\}$ the image line. With $\mathbf{r} = (x, z_0)$, $\mathbf{r}' = (x', 0)$, we have

$$u^s(x, z_0) = Ce^{i\omega x^2/(2z_0)} \int_{\mathbb{R}} v(x', 0)(u^i(x', 0) + u^s(x', 0))e^{i\omega(x')^2/(2z_0)} e^{-i\omega x x'/z_0} dx', \tag{3.5}$$

where C is a complex number.

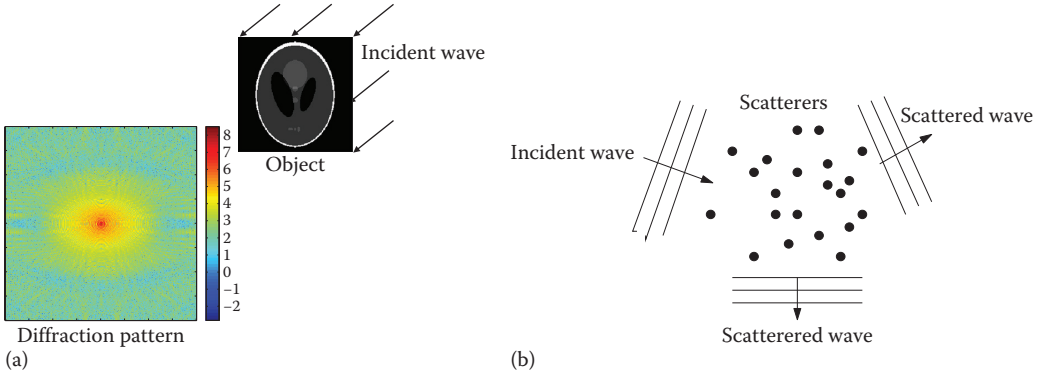


Figure 3.1 Two imaging geometries: (a) diffraction and (b) scattering.

- *Scattering geometry*: The scattered field has the far-field asymptotic (Born and Wolf 1999)

$$u^{\mathrm{s}}(\mathbf{r}) = \frac{e^{i\omega|\mathbf{r}|}}{|\mathbf{r}|^{(d-1)/2}}\left(A(\hat{\mathbf{r}},\hat{\mathbf{d}}) + \mathcal{O}\left(\frac{1}{|\mathbf{r}|}\right)\right), \quad \hat{\mathbf{r}} = \frac{\mathbf{r}}{|\mathbf{r}|}, \quad d = 2,3, \tag{3.6}$$

  where the scattering amplitude A has the dimension-independent form

$$A(\hat{\mathbf{r}},\hat{\mathbf{d}}) = \frac{\omega^2}{4\pi}\int_{\mathbb{R}^d} v(\mathbf{r}')(u^{\mathrm{i}}(\mathbf{r}') + u^{\mathrm{s}}(\mathbf{r}'))e^{-i\omega\mathbf{r}'\cdot\hat{\mathbf{r}}}d\mathbf{r}' \,. \tag{3.7}$$

Note that since u in (3.5) and (3.7) is part of the unknown due to multiple scattering, the inverse problem is a nonlinear one. To deal with multiple scattering effects in compressive sensing, it is natural to split the inverse problem into two stages: In the first stage, we recover the *masked* objects

$$V(x) = v(x,0)(u^{\mathrm{i}}(x,0) + u^{\mathrm{s}}(x,0))e^{i\omega x^2/(2z_0)} \quad \text{(paraxial geometry)},$$

$$V(\mathbf{r}) = v(\mathbf{r})(u^{\mathrm{i}}(\mathbf{r}) + u^{\mathrm{s}}(\mathbf{r})) \quad \text{(scattering geometry)},$$

with the Fourier-like integrals in (3.5) and (3.7) as the sensing operators. In the second stage, we recover the true objects from the masked objects.

For most of this chapter, however, we will focus on the first stage or make the Born approximation to linearize the imaging problem and turn to the multiple scattering effect only in the "Inverse Multiple Scattering" section.

Outline

In the "Review of Compressive Sensing" section, we review the basic elements of compressive sensing theory including basis pursuit (BP) and greedy algorithms (orthogonal matching pursuit [OMP], in particular). We place greater emphasis on the incoherence properties than on the restricted isometry property because the former is much easier to estimate than the latter, even though the latter can also be established in several settings as we will see throughout this chapter. One thing to keep in mind about incoherence is that it is far beyond the standard

notion of coherence parameter, which is the worst case metric (see Equation 3.17). The incoherence properties are fully expressed in the Gram matrix of the sensing matrix, also known as the coherence pattern. The second thing noteworthy about incoherence is that the standard performance guarantees expressed in terms of the coherence parameter often underestimate the actual performance of algorithms. Its usefulness primarily lies in providing a guideline for designing measurement schemes.

In the "Fresnel Diffraction with Pixel Basis" section, we consider the Fresnel diffraction with the pixel basis. The pixel basis, having a finite, definite size, is emphatically *not* suitable for point-like objects. Indeed, in order to build incoherence in the sensing matrix, it is imperative that the wavelength be shorter than the grid spacing. In other words, the pixel basis is suitable only for objects that are decomposable into "smooth" parts relative to the wavelength. The sparsity priors then come in two kinds: (1) there are few such parts with 1-norm as proxy and (2) there are few changes from part to part with the total variation as proxy (the "Total Variation Minimization" section). In the context of Fourier measurement, we introduce the notion of constrained joint sparsity to connect these two sparse priors and discuss basis pursuit (the "BPDN for Joint Sparsity" section) and orthogonal matching pursuit for joint sparsity (the "OMP for Joint Sparsity" section).

In contrast to the pixelated objects, point objects naturally do not live on grids. Such a problem arises in applications, for example, discrete spectral estimation among others. There is this fundamental tradeoff in using a grid to image point objects with the standard theory of compressive sensing: the finer the grid, the better the point objects are captured but the worse the coherence parameter becomes. In the "Fresnel Diffraction with Point Objects" section, we use the notion of coherence band to analyze the coherence pattern and design new compressive sensing algorithms for imaging well separated, off-grid point objects. In addition to off-grid point objects, the coherence-band techniques are also useful for imaging objects that admit a sparse representation in highly redundant dictionaries. One celebrated example is the single-pixel camera (SPC) discussed briefly in the "Redundant Dictionaries" section.

In the "Fresnel Diffraction with Littlewood–Paley Basis" section, we discuss Fresnel diffraction with sparse representation on the Littlewood–Paley basis, which is a slowly decaying wavelet basis in stark contrast to the pixel basis and the point-like objects. In this basis, the sensing matrix has hierarchical structures completely decoupled over different scales. In the "Near-Field Diffraction with Fourier Basis" section, we discuss near-field diffraction in terms of angular spectrum, which works out nicely with the Fourier basis.

In the "Inverse Scattering" section, we consider inverse scattering with the pixelated as well as point objects. Here, we focus on the design of sampling schemes (the "Sampling Schemes" section) and various coherence bounds for different schemes (the "Coherence Bounds" section).

In the "Inverse Multiple Scattering" section, we discuss multiple scattering of point objects and the appropriate techniques for solving the nonlinear inverse problem. The keys are the combination of the coherence-band and the joint sparsity techniques developed earlier.

In the "Inverse Scattering with Zernike Basis" section, we discuss inverse scattering with extended objects sparsely represented in the Zernike basis. In the "Interferometry with Incoherent Sources" section, we discuss interferometry with incoherent sources in astronomy. As a consequence of the celebrated

van Cittert–Zernike theorem, the resulting sensing matrix has a similar structure to that for scattering with multiple inputs and outputs. The difference between them lies in the fact that for interferometry the inputs and outputs are necessarily correlated while for scattering the inputs and outputs can be independent. As a result, the (in)coherence properties of interferometry are more subtle and it is an ongoing problem to search for the optimal sensor arrays in optical interferometry in astronomy.

Review of Compressive Sensing

A distinctive advantage of compressive sensing is accounting for the finite, discrete nature of measurement by appropriately discretizing the object domain.

By a slight abuse of notation, we use $\|\cdot\|_p$ to denote the p-norm ($p \geq 1$) of functions as well as vectors, that is,

$$\|f\|_p = \left(\int |f(\mathbf{r})|^p d\mathbf{r} \right)^{1/p}, \quad f \in L^p, \tag{3.8}$$

$$\|\mathbf{f}\|_p = \left(\sum_{j=1}^{N} |f_j|^p \right)^{1/p}, \quad \mathbf{f} \in \mathbb{C}^N, \tag{3.9}$$

and $\|\mathbf{f}\|_0$ (the sparsity) denotes the number of nonzero components in a vector $\mathbf{f}$.

By discretizing the right-hand side of (3.5) or (3.7) and selecting a discrete set of data on the left-hand side, we rewrite the continuous models in the form of linear inversion

$$\mathbf{g} = \mathbf{\Phi f} + \mathbf{e}, \tag{3.10}$$

where the error vector $\mathbf{e} \in \mathbb{C}^M$ is the sum of the external noise $\mathbf{n}$ and the discretization error $\mathbf{d}$ due to model mismatch. By definition, the discretization error $\mathbf{d}$ is given by

$$\mathbf{d} = \mathbf{g} - \mathbf{n} - \mathbf{\Phi f}. \tag{3.11}$$

Consider the principle of basis pursuit denoising (BPDN) convex program

$$\min \|\mathbf{h}\|_1, \quad \text{s.t. } \|\mathbf{g} - \mathbf{\Phi h}\|_2 \leq \|\mathbf{e}\|_2 = \epsilon. \tag{3.12}$$

When $\epsilon = 0$, (3.12) is called basis pursuit. With the right choice of the parameter λ, BPDN is equivalent to the unconstrained convex program that is called Lasso (Tibshirani 1996)

$$\min_{\mathbf{z}} \frac{1}{2} \| \mathbf{g} - \mathbf{\Phi z} \|_2^2 + \lambda \varepsilon \| \mathbf{z} \|_1, \tag{3.13}$$

Both BPDN (3.12) and Lasso (3.13) are convex programs and have numerically efficient solvers (Chen et al. 2001, Boyd and Vandenberghe 2004, Bruckstein et al. 2009).

A fundamental notion in compressed sensing under which BP yields a unique exact solution is the restrictive isometry property (RIP) due to Candès and

Tao (2005). Precisely, let the restricted isometry constant (RIC) δ_s be the smallest nonnegative number such that the inequality

$$\kappa(1-\delta_s)\,\|\,\mathbf{h}\,\|_2^2 \le \|\,\mathbf{\Phi h}\,\|_2^2 \le \kappa(1+\delta_s)\,\|\,\mathbf{h}\,\|_2^2$$

holds for all $\mathbf{h} \in \mathbb{C}^N$ of sparsity at most s and some constant $\kappa > 0$. RIP means a sufficiently small δ_{2s} (see Equation 3.14).

Now we recall a standard performance guarantee under RIP.

Theorem 3.1 (*Candès 2008*) *Suppose the RIC of* $\mathbf{\Phi}$ *satisfies the inequality*

$$\delta_{2s} < \sqrt{2} - 1 \tag{3.14}$$

with $\kappa = 1$. *Then the solution* $\mathbf{f}_*$ *of BPDN* (*3.12*) *satisfies*

$$\|\mathbf{f}_* - \mathbf{f}\|_2 \le C_1\, s^{-1/2}\, \|\mathbf{f} - \mathbf{f}^{(s)}\|_1 + C_2 \epsilon \tag{3.15}$$

for some constants C_1 *and* C_2. *where* $\mathbf{f}^{(s)}$ *consists of the s largest components, in magnitude, of* $\mathbf{f}$.

Remark 3.1 For general $\kappa \neq 1$, we consider the normalized version of (3.10)

$$\frac{1}{\sqrt{\kappa}}\mathbf{g} = \frac{1}{\sqrt{\kappa}}\mathbf{\Phi f} + \frac{1}{\sqrt{\kappa}}\mathbf{e}$$

and obtain from (3.15) that

$$\|\,\mathbf{f}_* - \mathbf{f}\,\|_2 \le C_1 s^{-1/2}\,\|\,\mathbf{f} - \mathbf{f}^{(s)}\,\|_1 + C_2\,\frac{\epsilon}{\sqrt{\kappa}}. \tag{3.16}$$

Note however that neither BPDN nor Lasso is an algorithm by itself and there are many different algorithms for solving these convex programs. Some solvers are available online, for example, YALL1 and the open-source code *L1-MAGIC* (http://users.ece.gatech.edu/~justin/llmagic/).

Besides convex programs, greedy algorithms are an alternative approach to sparse recovery. A widely known greedy algorithm is the OMP (Pati et al. 1993, Davis et al. 1997).

Algorithm 3.1 Orthogonal Matching Pursuit (OMP)

Input: $\mathbf{\Phi}$, $\mathbf{g}$.

Initialization: $\mathbf{f}^0 = 0$, $\mathbf{r}^0 = \mathbf{g}$ and $S^0 = \emptyset$

Iteration: For $j = 1, \ldots, s$

(1) $i_{\max} = \arg\max_i |\langle \mathbf{r}^{j-1}, \mathbf{\Phi}_i \rangle|$, $i \notin S^{j-1}$

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

(3) $\mathbf{f}^j = \arg\min_{\mathbf{h}} \|\mathbf{\Phi h} - \mathbf{g}\|_2$ s.t. $\text{supp}(\mathbf{h}) \subseteq S^j$

(4) $\mathbf{r}^j = \mathbf{g} - \mathbf{\Phi f}^j$

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

OMP has a performance guarantee in terms of the coherence parameter defined by

$$\mu(\mathbf{\Phi}) = \max_{k \neq l} \mu(k,l), \quad \mu(k,l) = \frac{|\Phi_k^\dagger \Phi_l|}{\|\Phi_k\| \|\Phi_l\|} \tag{3.17}$$

where Φ_k is the kth column of $\mathbf{\Phi}$, $\mu(k, l)$ is the pairwise coherence parameter and the totality $[\mu(k, l)]$ is the *coherence pattern* of the sensing matrix $\mathbf{\Phi}$. Here and in the following, † denotes the conjugate transpose.

Theorem 3.2 (*Donoho et al. 2006*) *Suppose that the sparsity s of the signal vector* $\mathbf{f}$ *satisfies*

$$\mu(\mathbf{\Phi})(2s-1) + 2\frac{\|\mathbf{e}\|_2}{f_{\min}} < 1 \tag{3.18}$$

where $f_{\min} = \min_k |f_k|$. *Denote by* $\mathbf{f}_*$, *the output of the OMP reconstruction. Then*

- $\mathbf{f}_*$ *has the correct support, that is,* $\text{supp}(\mathbf{f}_*) = \text{supp}(\mathbf{f})$ *where* $\text{supp}(\mathbf{f})$ *is the support of* $\mathbf{f}$.
- $\mathbf{f}_*$ *approximates the object vector in the sense that*

$$\|\mathbf{f}_* - \mathbf{f}\|_2 \leq \frac{\|\mathbf{e}\|}{\sqrt{1 + \mu - \mu s}}. \tag{3.19}$$

Incoherence or RIP often requires randomness in the sensing matrix, which can come from the randomness in sampling as well as in illumination. Between the two metrics, incoherence is far more flexible and easier to verify for a given sensing matrix. However, performance guarantees in terms of the coherence parameter such as (3.18) of Theorem 3.2 tend to be conservative.

Fresnel Diffraction with Pixel Basis

As a first example, we consider the imaging equation (3.5) for Fresnel diffraction. We shall write (3.5) in the discrete form (3.10) by discretizing the right-hand side of (3.5) and selecting a discrete set of scattered field data for the left-hand side.

We approximate the masked object

$$V(x) = v(x)u(x,0)e^{i\omega x^2/(2z_0)} \tag{3.20}$$

by the discrete sum on the scale ℓ

$$V_\ell(x) = \sum_{k=1}^{N} b\left(\frac{x}{\ell} - k\right) V(\ell k), \quad V(\ell k) = v(\ell k) u(\ell k, 0) e^{i\omega \ell^2 k^2/(2z_0)} \tag{3.21}$$

where

$$b(x) = \begin{cases} 1, & x \in \left[-\frac{1}{2}, \frac{1}{2}\right] \\ 0, & \text{else.} \end{cases} \tag{3.22}$$

is the localized pixel "basis." We assume that V_ℓ is a good approximation of the masked object for sufficiently small ℓ in the sense that $\lim_{\ell \to 0} \|V - V_\ell\|_1 = 0$.

Moreover, we assume that V_ℓ is sparse in the sense that relatively few components $V(k\ell)$ are significant compared to the number of grid points N. Note that sparse objects in the pixel basis are *not* point-like. Point objects typically induce large gridding errors and require techniques beyond standard compressive sensing reviewed in the "Review of Compressive Sensing" section (cf. the "Fresnel Diffraction with Point Objects" section).

To proceed, we make the Born approximation and set $u^i(x, 0) = 1$ (i.e., normal incidence of plane wave).

Let x_j, $j = 1, \ldots, M$ be the sampling points on the image/sensor line and define

$$\xi_j = \frac{\omega \ell x_j}{2\pi z_0}, \quad j = 1, \ldots, M. \tag{3.23}$$

Set the discretized, unknown vector $\mathbf{f} \in \mathbb{C}^N$ as

$$f_k = v(\ell k) e^{i\omega \ell^2 k^2/(2z_0)}, \quad k = 1, \ldots, N$$

and the data vector $\mathbf{g} \in \mathbb{C}^M$ as

$$g_j = \frac{u^s(x_j, z_0)}{C\ell \hat{b}(\xi_j)} e^{-i\omega x_j^2/(2z_0)}, \quad j = 1, \ldots, M$$

where

$$\hat{b}(\xi) = \int b(x) e^{-i2\pi x\xi}\, dx = \frac{\sin(\pi\xi)}{\pi\xi}. \tag{3.24}$$

As a result, (3.5) can be expressed as (3.10) with the sensing matrix

$$\mathbf{\Phi} = [\Phi_1 \cdots \Phi_N] \in \mathbb{C}^{M \times N}, \quad \mathbf{\Phi}_k = [e^{-2\pi i \xi_j k}]_{j=1}^{M}, \quad k = 1, \ldots, N. \tag{3.25}$$

A sensing matrix whose columns have the same 2-norm (as in (3.25)) tends to enjoy better performance in compressive sensing reconstruction.

When ξ_j are independent uniform random variables on [−1/2, 1/2], (3.25) is the celebrated random partial Fourier matrix that is among a few examples with a relatively sharp bound on the RIP as given in the following.

Theorem 3.3 (*Rauhut 2008*) *Suppose*

$$\frac{\mathrm{M}}{\ln M} \geq c\delta^{-2} k \ln^2 k \ln N \ln \frac{1}{\varepsilon}, \quad \varepsilon \in (0,1) \tag{3.26}$$

for given sparsity k where c is an absolute constant. Then the restricted isometry constant of the matrix (3.25) *satisfies the bound*

$$\delta_k < \delta$$

with probability at least $1 - \varepsilon$.

Remark 3.2 To apply Theorem 3.3 in the context of Theorem 3.1, we can set $k = 2s$ and $\delta = \sqrt{2} - 1$. Equation 3.26 then implies that it would take roughly $\mathcal{O}(s)$, modulo some logarithmic factors, amount of measurement data for BPDN to succeed in the sense of (3.15).

On the other hand, the coherence parameter μ typically scales as $\mathcal{O}(M^{-1/2})$ as we will see in Theorem 3.5; therefore, in view of the condition (3.18) in Theorem 3.2, the amount of needed data is $\mathcal{O}(s^2)$, significantly larger than $\mathcal{O}(s)$ for $1 \ll s \ll N$.

While this observation is usually valid in the case of OMP, it needs not apply to other greedy algorithms such as subspace pursuit (BP) whose performance guarantee requires $\mathcal{O}(s)$, up to logarithmic factor, amount of data (Dai and Milenkovic 2009).

The fact that ξ_j are independent uniform random variables on [–1/2, 1/2] implies that x_j are independent uniform random variables on [$-A/2$, $A/2$] with

$$A = \frac{2\pi z_0}{\omega \ell} \tag{3.27}$$

in view of (3.23). Viewing ℓ as the resolution length of the imaging setup, we obtain the resolution criterion

$$\ell = \frac{2\pi z_0}{A\omega} \tag{3.28}$$

which is equivalent to the classical Abbe or Rayleigh criterion.

Now let us estimate the discretization error vector $\mathbf{d}$ in (3.11). Define the transformation $\mathcal{T}$ by

$$(\mathcal{T}V)_j = \frac{1}{\ell \hat{b}(\xi_j)} \int V(x') e^{-2\pi i \xi_j x'/\ell} dx',$$

cf. (3.7). By definition

$$\mathbf{d} = \mathcal{T}V - \mathcal{T}V_\ell$$

we have

$$\|\mathbf{d}\|_\infty \leq \frac{\| V - V_\ell \|_1}{\ell \min_j |\hat{b}(\xi_j)|}, \quad \hat{b}(\xi) = \frac{\sin(\pi\xi)}{\pi\xi}. \tag{3.29}$$

For $\xi \in [-1/2, 1/2]$, $\min |\hat{b}(\xi)| = 2/\pi$ and $\max |\hat{b}(\xi)| = 1$. Hence

$$\|\mathbf{d}\|_2 \le \|\mathbf{d}\|_\infty \sqrt{M} \le \frac{\pi\sqrt{M}}{2\ell} \| V - V_\ell \|_1, \tag{3.30}$$

and

$$\frac{\|\mathbf{d}\|_2}{\|\mathbf{g}\|_2} \le \frac{\pi C \sqrt{M} \|V - V_\ell\|_1}{2\sqrt{\sum_{j=1}^{M} |u^s(x_j)|^2}}$$

which can be made arbitrarily small by setting ℓ sufficiently small while holding M fixed and maintaining the relation (3.28).

Total Variation Minimization

If the masked object V is better approximated by a piecewise (beyond the scale ℓ) constant function V_ℓ, then the sparsity prior can be enforced by the discrete total variation

$$\|\mathbf{h}\|_{\text{tv}} \equiv \sum_j |\Delta h(j)|, \quad \Delta h(j) = h_{j+1} - h_j.$$

Instead of (3.12), we consider a different convex program, which is called total variation minimization (TV-min)

$$\min \|\mathbf{h}\|_{\text{tv}}, \quad \text{s.t. } \|\mathbf{g} - \mathbf{\Phi h}\|_2 \le \varepsilon. \tag{3.31}$$

cf. (Rudin et al. 1992, Rudin and Osher 1994, Chambolle and Lions 1997, Chambolle 2004, Candès et al. 2006).

For 2D objects $h(i, j)$, $i, j = 1, \ldots, n$, let $\mathbf{h} = (h_p)$ be the vectorized version with index $p = j + (i - 1)n$. The 2D discrete (isotropic) total variation is given by

$$\|\mathbf{h}\|_{\text{tv}} \equiv \sum_{i,j} \sqrt{|\Delta_1 h(i, j)|^2 + |\Delta_2 h(i, j)|^2},$$

$$\Delta_1 h(i, j) = (h(i + 1, j) - h(i, j), \quad \Delta_2 h(i, j) = h(i, j + 1) - h(i, j)).$$

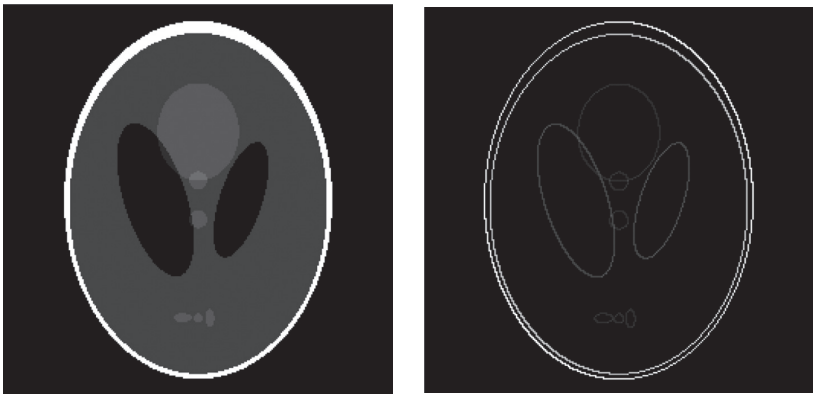
Figures 3.2 and 3.3 are a numerical demonstration of TV-min reconstruction of the 2D object (the phantom). Figure 3.2 shows the original image and its gradient, which is sparse compared to the original dimensionality. Figure 3.3 shows the reconstruction with BPDN (a) and TV-min (b). TV-min performs well as expected because the TV-sparsity is the correct prior for the object. On the other hand, BPDN performs poorly because the L1-sparsity is the wrong prior.

BPDN for Joint Sparsity

The close relationship between (3.31) and (3.12) can be seen from the following equation for the 1D setting:

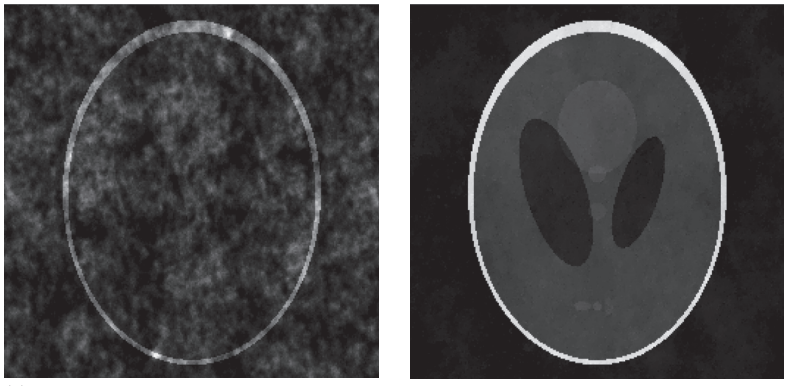
$$\left(e^{2\pi i \xi_j} - 1\right) g_j = \sum_k e^{-2\pi i \xi_j k} (f_{k+1} - f_k).$$

In other words, the new data vector $\tilde{\mathbf{g}} = ((e^{2\pi i \xi_j} - 1) g_j)$, the new noise vector $\tilde{\mathbf{e}} = ((e^{2\pi i \xi_j} - 1) e_j)$, and the new object vector $\tilde{\mathbf{f}} = (f_{k+1} - f_k)$ are related via the



(a) (b)

Figure 3.2 The original 256 × 256 Shepp–Logan phantom (a), the Shepp–Logan phantom and the magnitudes of its gradient with sparsity $s = 2184$ (b). (From Fannjiang, A., *Math. Mech. Complex Syst.*, 1, 81, 2013a. With permission.)



(a) (b)

Figure 3.3 BPDN reconstruction without external noise (a) and TV-min reconstruction with 5% noise (b). (From Fannjiang, A., *Math. Mech. Complex Syst.*, 1, 81, 2013a. With permission.)

same sensing matrix as for BPDN. Clearly, $|\tilde{e}_j| \le 2|e_j|, j = 1,\ldots,M$. Moreover, if e_j are independently and identically distributed, then $\tilde{e}_j$ are also independently and identically distributed with variance

$$\mathbb{E}\,|e_j|^2 = \mathbb{E}\,|e^{2\pi i\xi_j} - 1|^2 \times \mathbb{E}\,|e_j|^2 = 2\mathbb{E}\,|e_j|^2$$

when ξ_j is the uniform random variable over [–1/2, 1/2]. Hence, for large M, the new noise magnitude is $\|\tilde{\mathbf{e}}\|_2 \approx \sqrt{2}\,\|\mathbf{e}\|_2$. Here and in the following, $\mathbb{E}$ denotes the expected value.

A similar relationship exists in the 2D case. Let $\mathbf{f}_j = \Delta_j \mathbf{f}$ that satisfies the linear constraint

$$\Delta_1 \mathbf{f}_2 = \Delta_2 \mathbf{f}_1. \tag{3.32}$$

Define

$$\mathbf{g}_1 = [(e^{2\pi i\xi_j} - 1)g_j], \quad \mathbf{g}_2 = [(e^{2\pi i\eta_j} - 1)g_j]$$
$$\mathbf{e}_1 = [(e^{2\pi i\xi_j} - 1)e_j], \quad \mathbf{e}_2 = [(e^{2\pi i\eta_j} - 1)e_j]$$

where ξ_j, η_j, $j = 1, \ldots, M$ are independent uniform random variables over $[-1/2, 1/2]$. Then $\mathbf{F} = [\mathbf{f}_1, \mathbf{f}_2] \in \mathbb{C}^{N\times 2}$, $\mathbf{G} = [\mathbf{g}_1, \mathbf{g}_2] \in \mathbb{C}^{M\times 2}$, and $\mathbf{E} = [\mathbf{e}_1, \mathbf{e}_2]$ are related through

$$\mathbf{G} = [\mathbf{\Phi}\mathbf{f}_1, \mathbf{\Phi}\mathbf{f}_2] + \mathbf{E}$$

subject to the linear constraint (3.32). This formulation calls for the L^1-minimization (Fannjiang 2013a)

$$\min \|[\mathbf{h}_1, \mathbf{h}_2]\|_{2,1}, \quad \text{s.t. } \|\mathbf{G} - [\mathbf{\Phi}\mathbf{h}_1, \mathbf{\Phi}\mathbf{h}_2]\|_{\mathrm{F}} \le \|\mathbf{E}\|_{\mathrm{F}}, \tag{3.33}$$

subject to the constraint

$$\Delta_2\mathbf{h}_1 = \Delta_1\mathbf{h}_2 \tag{3.34}$$

where

$\|\cdot\|_{\mathrm{F}}$ is the Frobenius norm

$\|\cdot\|_{2,1}$ is the mixed (2, 1)-norm (Benedek and Panzone 1961, Kowalski 2009)

$$\|\mathbf{X}\|_{2,1} = \sum_j \|\mathrm{row}_j(\mathbf{X})\|_2. \tag{3.35}$$

The reason for minimizing the mixed (2, 1)-norm in (3.33) is that $\mathbf{f}_1$ and $\mathbf{f}_2$ share the same sparsity pattern, which should be enforced.

To get a more clear idea about $\|\mathbf{E}\|_{\mathrm{F}}$, we apply the same analysis as mentioned earlier and obtain

$$\|\mathbf{e}_i\|_2^2 \approx \mathbb{E}\|\mathbf{e}_i\|_2^2 = 2\mathbb{E}\|\mathbf{e}\|_2^2, \quad i = 1, 2,$$

for sufficiently large M.

The convex program (3.33) through (3.34) is an example of BPDN with constrained joint sparsity. More generally, suppose that the columns of the unknown multivectors $\mathbf{F} \in \mathbb{C}^{N\times J}$ share the same support and are related to the data multivectors $\mathbf{G} \in \mathbb{C}^{M\times m}$ and the noise multivectors $\mathbf{E} \in \mathbb{C}^{M\times J}$ via

$$\mathbf{G} = [\mathbf{\Phi}_1\mathbf{f}_1, \mathbf{\Phi}_2\mathbf{f}_2, \ldots, \mathbf{\Phi}_J\mathbf{f}_J] + \mathbf{E} \tag{3.36}$$

subject to the linear constraint $\mathcal{L}\mathbf{F} = 0$.

For this setting, the following formulation of BPDN with joint sparsity is natural

$$\min \|\mathbf{H}\|_{2,1}, \quad \text{s.t. } \|\mathbf{G} - [\mathbf{\Phi}_1\mathbf{h}_1, \mathbf{\Phi}_2\mathbf{h}_2, \ldots, \mathbf{\Phi}_J\mathbf{h}_J]\|_{\mathrm{F}} \le \varepsilon, \quad \text{s.t. } \mathcal{L}\mathbf{H} = 0, \tag{3.37}$$

with $\varepsilon = \|\mathbf{E}\|_{\mathrm{F}}$.

OMP for Joint Sparsity

Next, we present an algorithmic extension of OMP for joint sparsity (Cotter et al. 2005, Chen and Huo 2006, Tropp et al. 2006) to the setting with multiple sensing matrices (3.36) (Fannjiang 2013a).

Algorithm 3.2 OMP for joint sparsity

Input: $\{\mathbf{\Phi}_j\}$, $\mathbf{g}$, $\varepsilon > 0$

Initialization: $\mathbf{f}^0 = 0$, $\mathbf{R}^0 = \mathbf{G}$ and $\mathcal{S}^0 = \emptyset$

Iteration: For $k = 1, 2, 3, \ldots$

(1) $i_{\max} = \arg\max_i \sum_{j=1}^{J} | \Phi_{j,i}^{\dagger} R_j^{k-1} |$, where $\Phi_{j,i}^{\dagger}$ is the conjugate transpose of ith column of $\mathbf{\Phi}_j$,

(2) $\mathcal{S}^k = \mathcal{S}^{k-1} \cup \{i_{\max}\}$

(3) $\mathbf{F}^k = \arg\min \|[\mathbf{\Phi}_1\mathbf{h}_1, \ldots, \mathbf{\Phi}_J\mathbf{h}_J] - \mathbf{G}\|_{\mathrm{F}}$ s.t. $\mathrm{supp}(\mathbf{H}) \subseteq S^k$

(4) $\mathbf{R}^k = \mathbf{G} - [\mathbf{\Phi}_1\mathbf{f}_1^k, \ldots, \mathbf{\Phi}_J\mathbf{f}_J^k]]$

(5) Stop if $\sum_j R_{j2}^k \le \varepsilon$.

Output: $\mathbf{F}^k$.

Note that the linear constraint $\mathcal{L}$ is not enforced in Algorithm 3.2. The idea is to first find the support of the multivectors without taking into account of the linear constraint, and, in the second stage, follow the support recovery with least squares

$$\mathbf{F}_* = \arg\min_{\mathbf{H}} \|\mathbf{G} - [\mathbf{\Phi}_1\mathbf{h}_1, \ldots, \mathbf{\Phi}_J\mathbf{h}_J]\|_F, \quad \text{s.t. } \mathrm{supp}(\mathbf{H}) \subseteq \mathrm{supp}(\mathbf{F}^\infty),\ \mathcal{L}\mathbf{H} = 0 \quad (3.38)$$

where $\mathbf{F}^\infty$ is the output of Algorithm 3.2.

For more discussion and applications of constrained joint sparsity, the reader is referred to Fannjiang (2013a) where the performance guarantees similar to Theorems 3.1 and 3.2 are proved for constrained joint sparsity.

Fresnel Diffraction with Point Objects

A major problem with discretizing the object domain shows up when the objects are point-like. In this case, it is unrealistic to assume that the objects are located exactly on the grid as the forceful matching between the point objects and the grid can create detrimental errors. Without additional prior information, the gridding error due to the mismatch between the point object locations and the grid points can be as large as the data itself, resulting in a low signal-to-noise ratio (SNR).

We call the grid spacing ℓ given in (3.28), the *Resolution length* (RL), which is the natural unit for resolution analysis. In the RL unit, the object domain grid becomes a subset of the integer grid $\mathbb{Z}$.

In the case of point objects, to refine the standard grid and reduce discretization error, we consider a fractional grid

$$\mathbb{Z}/F = \{j/F : j \in \mathbb{Z}\} \quad (3.39)$$

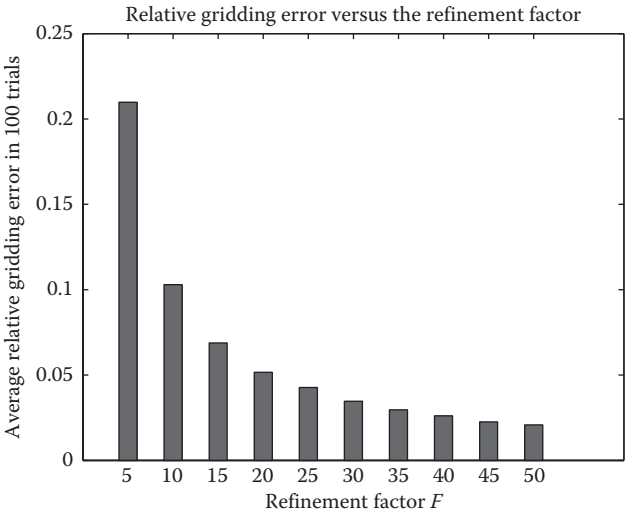


Figure 3.4 The relative gridding error is roughly inversely proportional to the refinement factor. (From Fannjiang, A. and Liao, W., *SIAM J. Imaging Sci.*, 5, 179, 2012a. Copyright © 2012 Society for Industrial and Applied Mathematics. Reprinted with permission.)

where $F \in \mathbb{N}$ is called *the refinement factor.* The random partial Fourier matrix (3.25) now takes the form

$$\mathbf{\Phi} = \left[e^{-i2\pi\xi_j k/F} \right], \tag{3.40}$$

where $\xi_j \in [-1/2,1/2]$ are independent uniform random variables. In the following numerical examples, we shall consider both deterministic (see Equation 3.45) as well as random sampling schemes.

As shown in Figure 3.4, the relative gridding error $\|\mathbf{d}\|/\|\mathbf{\Phi f}\|$ is roughly inversely proportional to the refinement factor F.

Figure 3.5 shows the coherence pattern $[\mu(j, k)]$ of a 100×4000 matrix (3.40) with $F = 20$ (Figure 3.5a). The bright diagonal band represents a heightened correlation (pairwise coherence) between a column vector and its neighbors on both sides (about 30). Figure 3.5b shows a half cross section of the coherence band across two RL, averaged over 100 independent trials. In general sparse recovery with large F exceeds the capability of currently known algorithms as the condition number of the 100×30 submatrix corresponding to the coherence band in Figure 3.5 easily exceeds 10^{15}. The high condition number makes stable recovery impossible. While Figure 3.5 is typical of the coherence pattern of 1D sensing matrices, the coherence pattern for two or three dimensions is considerably more complicated depending on how the objects are vectorized.

Band-Excluded, Locally Optimized Orthogonal Matching Pursuit

To overcome the conundrum of a highly coherent sensing matrix due to a refined grid, we have to go beyond the coherence parameter and study the coherence pattern of the sensing matrix.

The coherence pattern of a sensing matrix can be described in terms of the notion of *coherence band* defined in the following. Let $\eta > 0$. Define the η-coherence band of the index k as

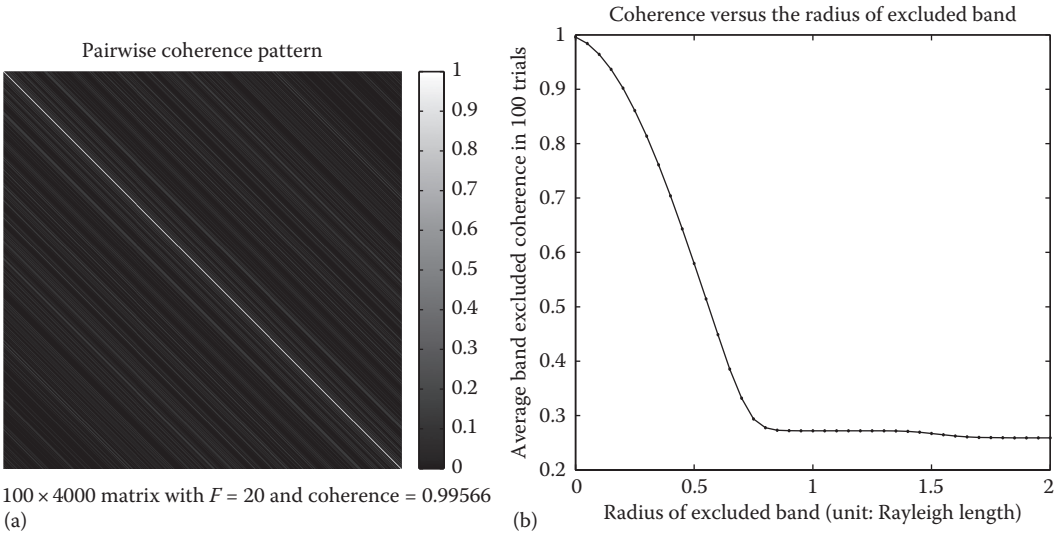


Figure 3.5 Coherence pattern [$\mu(j, k)$] for the 100×4000 matrix with $F = 20$ (a). The off-diagonal elements tend to diminish as the row number increases. The coherence band near the diagonals, however, persists and has the average profile shown in panel (b) where the vertical axis is the pairwise coherence averaged over 100 independent trials and the horizontal axis is the distance between two object points. (From Fannjiang, A. and Liao, W., *SIAM J. Imaging Sci.*, 5, 179, 2012a. Copyright © 2012 Society for Industrial and Applied Mathematics. Reprinted with permission.)

$$B_\eta(k) = \{i \mid \mu(i, k) > \eta\}, \tag{3.41}$$

and the double coherence band as

$$B_\eta^{(2)}(k) \equiv B_\eta(B_\eta(k)) = \cup_{j \in B_\eta(k)} B_\eta(j) \tag{3.42}$$

The first technique for taking advantage of the prior information of well separated objects is called band exclusion (BE) and can be easily embedded in the greedy algorithm, OMP.

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

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

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 $B_\eta(j), j \in \text{supp}(\mathbf{f})$ are pairwise disjoint. We call the modified algorithm as the band-excluded orthogonal matching pursuit (BOMP), as stated in Algorithm 3.3.

Algorithm 3.3 Band-Excluded Orthogonal Matching Pursuit (BOMP)

Input: $\mathbf{\Phi}$, $\mathbf{g}$, $\eta > 0$

Initialization: $\mathbf{f}^0 = 0$, $\mathbf{r}^0 = \mathbf{g}$, and $S^0 = \emptyset$

Iteration: For $j = 1, \ldots, s$

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

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

(3) $\mathbf{f}^j = \arg\min_{\mathbf{h}} \|\mathbf{\Phi h} - \mathbf{g}\|_2$ s.t. $\text{supp}(\mathbf{h}) \subseteq S^j$

(4) $\mathbf{r}^j = \mathbf{g} - \mathbf{\Phi f}^j$

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

The following theorem gives a (pessimistic) performance guarantee for BOMP.

Theorem 3.4 (*Fannjiang and Liao 2012a*) *Let* $\mathbf{f}$ *be s-sparse. Let* $\eta > 0$ *be fixed. Suppose that*

$$B_\eta(i) \cap B_\eta^{(2)}(j) = \emptyset, \quad \forall i, j \in \text{supp}(\mathbf{f}) \tag{3.43}$$

and that

$$\eta(5s-4)\frac{f_{\max}}{f_{\min}} + \frac{5\|\mathbf{e}\|_2}{2f_{\min}} < 1 \tag{3.44}$$

where

$$f_{\max} = \max_k |f_k|, \quad f_{\min} = \min_k |f_k|.$$

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

Remark 3.3 Condition (3.43) means that BOMP guarantees to resolve 3 RLs. In practice, BOMP can resolve objects separated by close to 1 RL when the dynamic range is nearly 1.

Remark 3.4 A main difference between Theorems 3.2 and 3.4 lies in the role played by the dynamic range $f_{\max}/f_{\min}$ and the separation condition (3.43).

Another difference is approximate recovery of support in Theorem 3.4 versus exact recovery of support in Theorem 3.2(a). In contrast to the F-independent nature of approximate support recovery, exact support recovery would probably be highly sensitive to the refinement factor F. That is, as F increases, the chance of missing some points in the support set also increases. As a result, the error of reconstruction $\|f^s - f\|_2$ tends to increase with F (as evident in Figure 3.7).

A main shortcoming with BOMP is in its failure to perform even when the dynamic range is even moderately greater than unity. To overcome this problem, we introduce the second technique: the *local optimization* (LO) that is a residual-reduction technique applied to the current estimate S^k of the object support (Fannjiang and Liao 2012a).

Algorithm 3.4 Local Optimization (LO)

Inputs: $\mathbf{\Phi}$, $\mathbf{g}$, $\eta > 0$, $S^0 = \{i_1, \ldots, i_k\}$.

Iteration: For $j = 1, 2, \ldots, k$.

(1) $\mathbf{f}^j = \arg\min_{\mathbf{h}} \|\mathbf{\Phi h} - \mathbf{g}\|_2$, $\mathrm{supp}(\mathbf{h}) = (S^{j-1} \setminus \{i_j\}) \cup \{i'_j\}, i'_j \in B_\eta(\{i_j\})$.

(2) $S^j = \mathrm{supp}(\mathbf{f}^j)$.

Output: S^k.

In other words, given a support estimate S^0, LO fine-tunes the support estimate by adjusting each element in S^0 within its coherence band in order to minimize the residual. The object amplitudes for the improved support estimate are obtained by solving the least-squares problem. Because of the local nature of LO, the computation is efficient.

Embedding LO in BOMP gives rise to the band-excluded, locally optimized orthogonal matching pursuit (BLOOMP).

Algorithm 3.5 Band-excluded, Locally Optimized Orthogonal Matching Pursuit (BLOOMP)

Input: $\mathbf{\Phi}$, $\mathbf{g}$, $\eta > 0$

Initialization: $\mathbf{f}^0 = 0$, $\mathbf{r}^0 = \mathbf{g}$ and $S^0 = \emptyset$

Iteration: For $j = 1, \ldots, s$

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

(2) $S^j = \mathrm{LO}(S^{j-1} \cup \{i_{\max}\})$, where $\mathrm{LO}(S^{j-1} \cup \{i_{\max}\})$ is the output of Algorithm 3.4 with $(S^{j-1} \cup \{i_{\max}\})$ as input.

(3) $\mathbf{f}^j = \arg\min_{\mathbf{h}} \|\mathbf{\Phi h} - \mathbf{g}\|_2$ s.t. $\mathrm{supp}(\mathbf{h}) \in \boldsymbol{S}^j$

(4) $\mathbf{r}^j = \mathbf{g} - \mathbf{\Phi f}^j$

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

The same BLO technique can be used to enhance the other well-known iterative schemes such as SP, CoSaMP (Needell and Tropp 2009), and compressed iterative hard thresholding (IHT) (Blumensath and Davies 2009, 2010), and the resulting algorithms are denoted by BLOSP, BLOCoSaMP, and BLOIHT, respectively, in the following numerical results. We refer the reader to Fannjiang and Liao (2012a) for the details and descriptions of these algorithms. MATLAB code of Algorithm 3.5 is available on-line at https://www.math.ucdavis.edu/~fannjiang/home/codes/BLOOMPcode.

Band-Excluding Thresholding

A related technique that can be used to enhance BPDN/Lasso for off-grid objects is called the band-excluding, locally optimized thresholding (BLOT).

Algorithm 3.6 Band-Excluding, Locally Optimized Thresholding (BLOT)

Input: $\mathbf{f} = (f_1, \ldots, f_N)$, $\mathbf{\Phi}$, $\mathbf{g}$, $\eta > 0$.

Initialization: $S^0 = \emptyset$.

Iteration: For $j = 1, 2, \ldots, s$.

(1) $i_j = \arg\max |f_k|, k \notin B_\eta^{(2)}(S^{j-1})$.

(2) $S^j = S^{j-1} \cup \{i_j\}$.

Output: $\mathbf{f}^s = \arg\min \|\mathbf{\Phi h} - \mathbf{g}\|_2$, $\text{supp}(\mathbf{h}) \subseteq \text{LO}(S^s)$, where LO is the output of Algorithm 3.4.

Numerical Examples

For numerical demonstration in Figures 3.6 and 3.7, we use *deterministic*, equally spaced sampling with

$$\xi_j = -\frac{1}{2} + \frac{j}{M}, \quad j = 1, \ldots, M \tag{3.45}$$

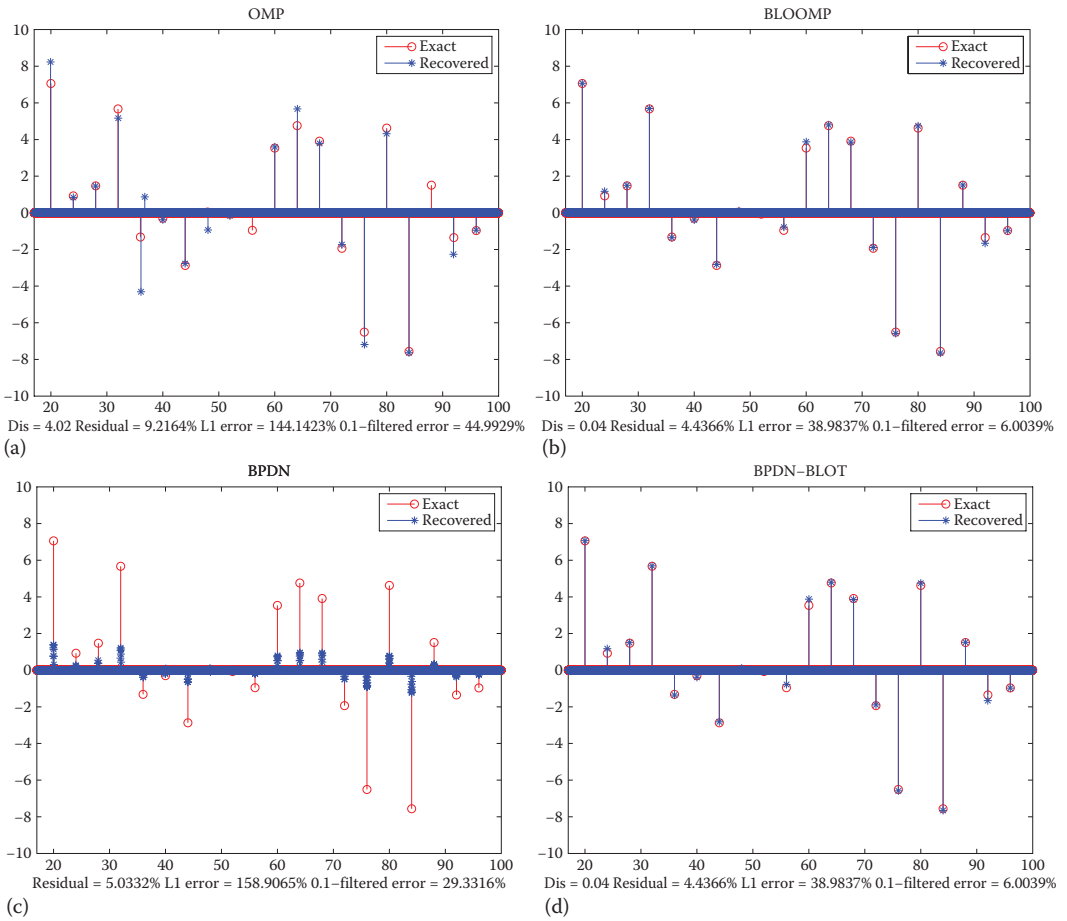


Figure 3.6 Reconstruction by (a) OMP, (b) BLOOMP, (c) BPDN, and (d) BPDN-BLOT of the real part of 20 randomly phased spikes with $F = 50$, SNR = 20. (From Fannjiang, A. and Liao, W., Super-resolution by compressive sensing algorithms, in *IEEE Proceedings of Asilomar Conference on Signals, Systems and Computers*, 2012b. With permission.)

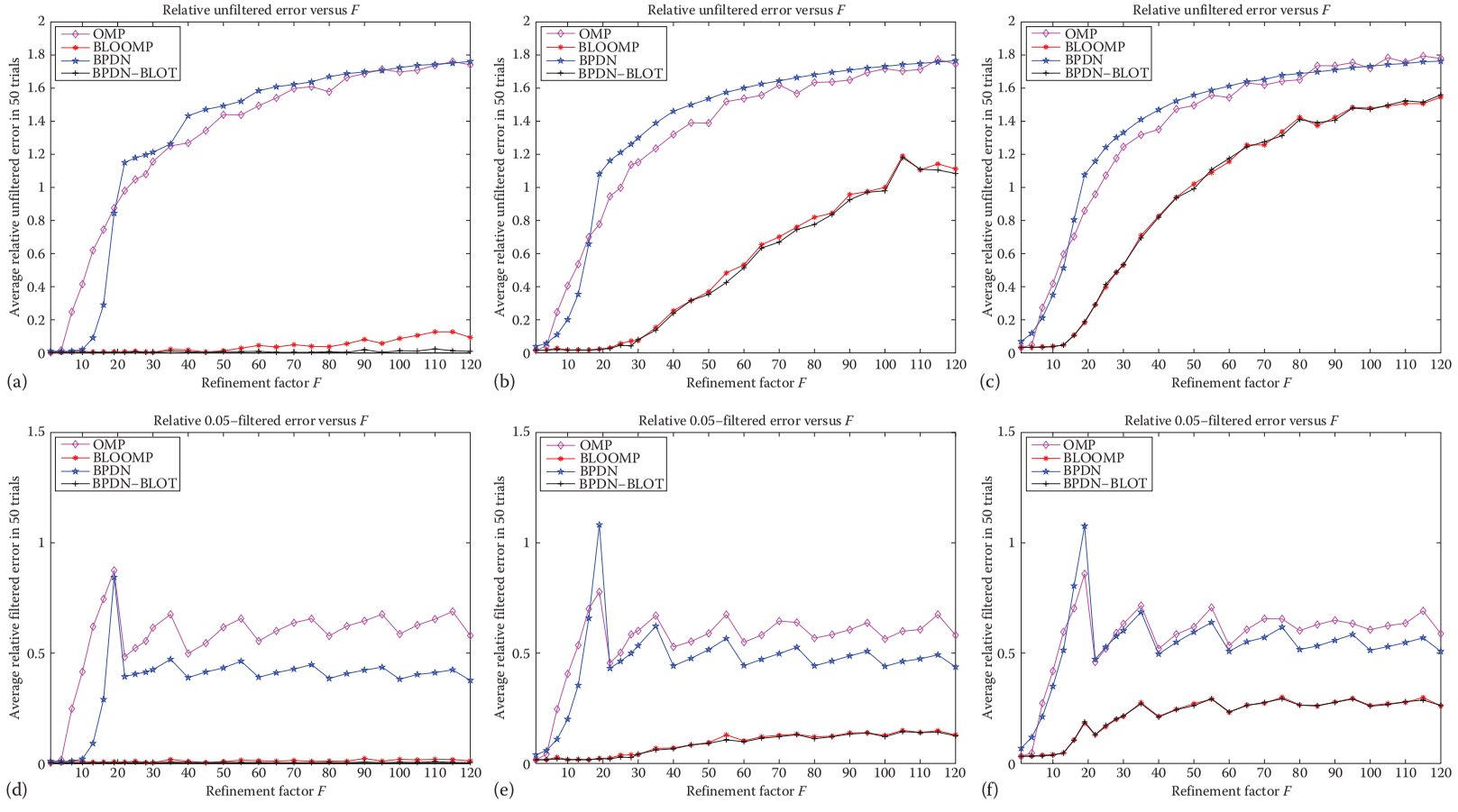


Figure 3.7 Relative errors in reconstruction by OMP, BLOOMP, BP, and BP-BLOT as F varies without (top) or with (bottom) filtering. (a) SNR = 100, $\eta = 0$; (b) SNR = 20, $\eta = 0$; (c) SNR = 10, $\eta = 0$; (d) SNR = 100, $\eta = 0.05\ell$; (e) SNR = 20, $\eta = 0.05\ell$; (f) SNR = 10, $\eta = 0.05\ell$. (From Fannjiang, A. and Liao, W., Super-resolution by compressive sensing algorithms, in *IEEE Proceedings of Asilomar Conference on Signals, Systems and Computers*, 2012b. With permission.)

and $\mathbf{\Phi} \in \mathbb{C}^{M\times FM}$ with $M = 150$, $F = 50$ to recover 20 randomly distributed and randomly phased point objects (spikes) separated by at least 4 RL.

Figure 3.6a and b shows how the BLO technique corrects the error of OMP due to the unresolved grid. In particular, several misses are recaptured and false detections removed. Figure 3.6c and d shows how the BLOT technique improves the BPDN estimate. In particular, BLOT has the effect of "trimming the bushes" and "growing the real trees." Figure 3.7a through c shows the relative error of reconstruction as a function of F by OMP, BPDN, BLOOMP, and BPDN-BLOT with the same setup and three different SNRs. For all SNRs, BLOOMP and BPDN-BLOT produce drastically fewer errors compared to OMP and BPDN.

The growth of relative error with F reflects the sensitivity of the reconstruction error alluded to in Remark 3.4. Note that the reconstruction error in the *discrete* norm cannot distinguish how far off the recovered support is from the true object support. The discrete norm treats any amount of support offset equally. An easy remedy to the injudicious treatment of support offset is to use instead the *filtered error norm* $\left\|\mathbf{f}_\eta^s - \mathbf{f}_\eta\right\|_2$, where $\mathbf{f}_\eta$ and $\mathbf{f}_\eta^s$ are, respectively, $\mathbf{f}$ and $\mathbf{f}^s$ convoluted with an approximate delta-function of width 2η.

Clearly the filtered error norm is more stable to support offset, especially if the offset is less than η. If every spike of $\mathbf{f}^s$ is within η distance from a spike of $\mathbf{f}$ *and* if the amplitude differences are small, then the η-filtered error is small. As shown in Figure 3.7d through f, averaging over $\eta = 5\%$ RL produces acceptable filtered error for any refinement factor relative to the external noise. This suggests that both BPDN-BLOT and BLOOMP recover the object support on average within 5% of 1 RL, a significant improvement over the theoretical guarantee of Theorem 3.4.

Next, we consider the unresolved partial Fourier matrix (3.40) with random sampling points to demonstrate the flexibility of the techniques. Let $\xi_j \in [-1/2, 1/2]$, $j = 1, \ldots, M$ be independent uniform random variables with $M = 100$, $N = 4000$, and $F = 20$. The test objects are 10 randomly phased and distributed objects, separated by at least 3 RL. As in Theorem 3.4, a recovery is counted as a success if every reconstructed object is within 1 RL of the object support.

Figure 3.8 compares the success rates (averaged over 200 trials) of the BLO-enhanced schemes (BLOOMP, BLOSP, BLOCoSaMP, and BLOIHT) and the BLOT-enhanced scheme (Lasso-BLOT). Lasso-BLOT is implemented with the regularization parameter (Chen et al. 2001)

$$\lambda = 0.5\sqrt{\log N} \quad \text{(black curves with diamonds)} \tag{3.46}$$

or

$$\lambda = \sqrt{2\log N} \quad \text{(black curves with stars)} \tag{3.47}$$

The empirically optimal choice (3.46) (labeled as Lasso-BLOT (0.5)) has a much improved performance over the choice (3.47). Clearly, BLOOMP is the best performer in noise stability and dynamic range among all tested algorithms.

Highly Redundant Dictionaries

Our discussion in the "Fresnel Diffraction with Point Objects" section so far is limited to point-like objects. But the methods presented earlier are also applicable to a wide variety of cases where the objects have sparse representations by redundant dictionaries, instead of orthogonal bases.

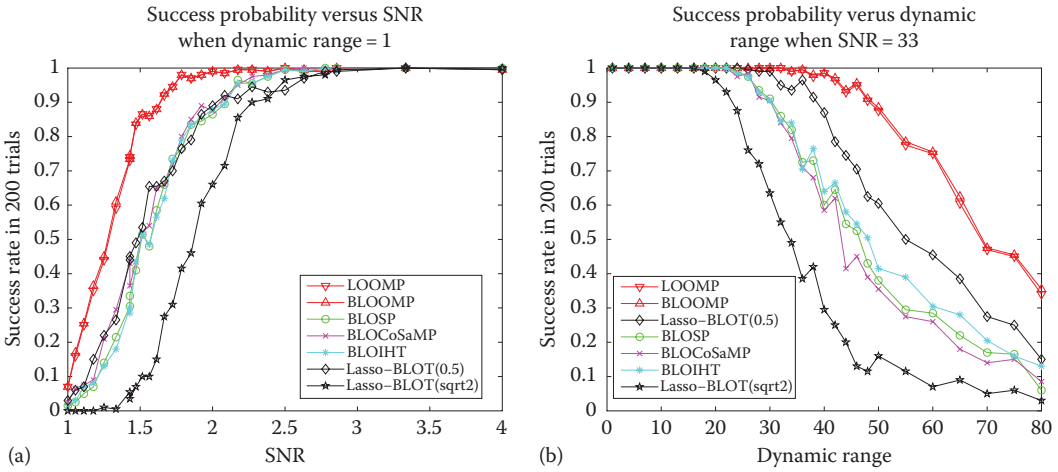


Figure 3.8 Success probability versus (a) SNR for dynamic range = 1 and (b) dynamic range for SNR = 33. Here, LOOMP is a simplified version of BLOOMP and has nearly identical performance curves. (From Fannjiang, A. and Liao, W., *SIAM J. Imaging Sci.*, 5, 179, 2012a. With permission.)

Suppose that the object is sparse in a highly redundant dictionary, which by definition tends to represent an object by fewer number of elements than a nonredundant one does. For example, one can combine different orthogonal bases into a dictionary that can sparsify a wider class of objects than any individual base can. On the other hand, a redundant dictionary tends to produce a larger coherence parameter and be ill suited for compressive sensing. This is the same kind of conundrum about off-grid point-like objects.

One of the most celebrated examples of optical compressive sensing is the SPC depicted in Figure 3.9. In the single-pixel camera (SPC), measurement diversity comes entirely from the digital micromirror device (DMD) instead of the sensor array. The DMD consists of an array of electrostatically actuated micromirrors. Each mirror can be positioned in one of the two states (±12°). Light reflected from mirrors in the +12-state only is then collected and focused by the lens and subsequently detected by a single optical sensor. For each and every measurement, the DMD is randomly and independently reconfigured. The resulting measurement matrix **A** has independently and identically distributed entries.

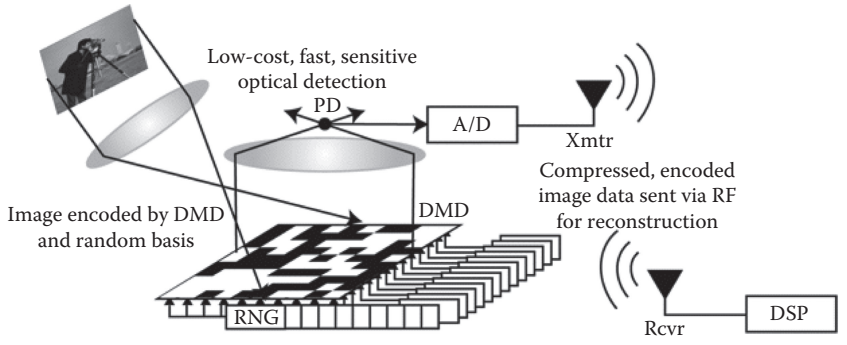


Figure 3.9 Single-pixel camera block diagram. (Courtesy of Rice Single-Pixel Camera Project, http://dsp.rice.edu/cscamera.)

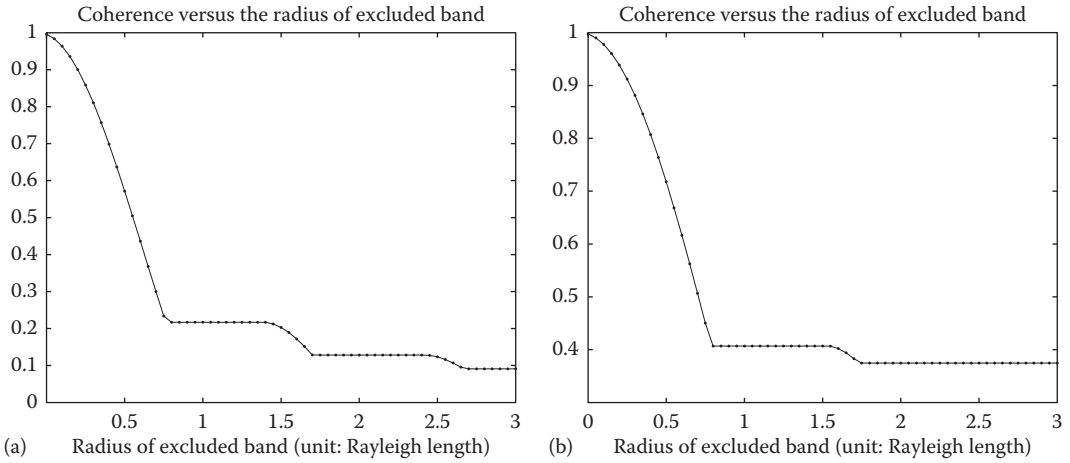


Figure 3.10 The coherence bands of the redundant Fourier frame $\mathbf{\Psi}$ (a) and $\mathbf{\Phi} = \mathbf{A\Psi}$ (b), the latter being averaged over 100 realizations of $\mathbf{A}$. (From Fannjiang, A. and Liao, W., *SIAM J. Imaging Sci.*, 5, 179, 2012a. Copyright © 2012 Society for Industrial and Applied Mathematics. Reprinted with permission.)

Suppose that the object is sparse in terms of a highly redundant dictionary. For simplicity of presentation, consider an 1D object sparse in an overcomplete Fourier frame (i.e., a dictionary that satisfies the frame bounds [Daubechies 1992]) with entries

$$\Psi_{k,j} = \frac{1}{\sqrt{R}} e^{-2\pi i((k-1)(j-1)/RF)}, \quad k = 1,\ldots,R,\ j = 1,\ldots,RF, \tag{3.48}$$

that includes harmonic as well as nonharmonic modes as its columns, where F is the redundant factor and R is a large integer. In other words, the object can be written as $\mathbf{\Psi f}$ with a sufficiently sparse vector $\mathbf{f}$. The final sensing matrix then becomes

$$\mathbf{\Phi} = \mathbf{A\Psi}. \tag{3.49}$$

The coherence bands of $\mathbf{\Psi}$ and $\mathbf{\Phi}$ are shown in Figure 3.10 from which we see that, as in Figure 3.5, the coherence radius is less than 1 RL. The same BLO- and BLOT-based techniques can be applied to (3.49); see Fannjiang and Liao (2012a) for numerical results and performance comparison with other techniques for off-grid objects (Candès et al. 2011, Candès and Fernandez-Granda 2013, 2014, Duarte and Baraniuk 2013, Tang et al. 2013).

Fresnel Diffraction with Littlewood–Paley Basis

Opposite to the localized pixel basis, the Littlewood–Paley basis is slowly decaying nonlocal modes based on the wavelet function

$$\psi(x) = (\pi x)^{-1}(\sin(2\pi x) - \sin(\pi x)), \tag{3.50}$$

which has a compactly supported Fourier transform

$$\hat{\psi}(\xi) = \int \psi(x) e^{-i2\pi\xi x}\, dx = \begin{cases} 1, & \frac{1}{2} \le |\xi| \le 1 \\ 0, & \text{otherwise.} \end{cases} \tag{3.51}$$

The following functions

$$\psi_{p,q}(x) = 2^{-p/2}\psi(2^{-p}x - q), \quad p, q \in \mathbb{Z} \tag{3.52}$$

form an orthonormal wavelet basis in $L^2(\mathbb{R})$ (Daubechies 1992). Expanding the masked object V (3.20) in the Littlewood–Paley basis, we write

$$V(x) = \sum_{p,q\in\mathbb{Z}} V_{p,q}\psi_{p,q}(x). \tag{3.53}$$

The main point of the subsequent discussion is to design a sampling scheme such that the resulting sensing matrix has desirable compressive sensing properties (Fannjiang 2009).

Let $\{2^p\colon p = -p_*, -p_* + 1, \ldots, p_*\}$ be the dyadic scales present in (3.53), $\{q\colon |q| \le N_p\}$ the modes present on the scale 2^p, and $2M_p + 1$ the number of measurements corresponding to the scale 2^p. Let

$$k = \sum_{j=-p_*}^{p'-1} (2M_j + 1) + q', \quad |q'| \le M_{p'}, \quad |p'| \le p_* \tag{3.54}$$

be the index for the sampling points. Throughout this section, k is determined by p', q' by (3.54). Let x_k be the sampling points and set the normalized coordinates

$$\frac{x_k \omega \ell}{2\pi z_0} = \xi_k, \quad k = 1, \ldots, M \tag{3.55}$$

where, as shown in the following, ℓ is a resolution parameter and $\xi_k \in [-1/2, 1/2]$ are determined in the following; cf. (3.23). This means that the aperture (i.e., the sampling range of x_k) is again given by (3.27).

Let $\mathbf{g} = (g_k)$ be the data vector with

$$g_k = C^{-1} u^s(x_k, z_0) e^{-i\omega x_k^2/(2z_0)}.$$

Direct calculation with (3.5) and (3.55) then gives

$$g_k = \sum_{p,q\in\mathbb{Z}} 2^{p/2} V_{p,q} e^{-i2\pi\xi_k \ell^{-1} 2^p q} \hat{\psi}(\xi_k \ell^{-1} 2^p), \quad k = 1, \ldots, M. \tag{3.56}$$

Let $\mathbf{f} = (f_l)$ be the object vector with

$$f_l = (-1)^q 2^{p/2} V_{p,q}$$

where the indices are related by

$$l = \sum_{j=-p_*}^{p-1} (2N_j + 1) + q.$$

Suppose that

$$\ell \le 2^{-p_*-1} \tag{3.57}$$

That is, 2ℓ is less than or equal to the smallest scale in the wavelet presentation (3.53).

Let $\zeta_{p',q'}$ be independent, uniform random variables on [–1/2, 1/2] and let

$$\xi_k = \frac{\ell}{2^{p'}} \cdot \begin{cases} 1/2 + \zeta_{p',q'}, & \zeta_{p',q'} \in [0,1/2] \\ -1/2 + \zeta_{p',q'}, & \zeta_{p',q'} \in [-1/2,0]. \end{cases} \tag{3.58}$$

where k is determined by (3.54). By the assumption (3.57), we have

$$\xi_k \in [-/2,\ 1/2], \quad \forall p' \ge -p_*.$$

More specifically, by (3.55), we have

$$x_k \in \frac{2\pi z_0}{\omega 2^{p'}} \left(\left[-1, -\frac{1}{2} \right] \cup \left[\frac{1}{2}, 1 \right] \right),$$

that is, the sampling regions for different dyadic scales indexed by p' are disjoint with the ones for the smaller scales on the outer skirt of the aperture, taking up a bigger portion of the aperture. The resulting sampling points are geometrically concentrated near (but not exactly at) the center of the aperture.

Let the sensing matrix elements be

$$\Phi_{k,l} = (-1)^q \hat{\psi}(\xi_k 2^p \ell^{-1}) e^{-i2\pi \xi_k 2^p q/\ell}. \tag{3.59}$$

We claim that $\Phi_{k,l} = 0$ for $p \ne p'$. This is evident from (3.58) and the following calculation:

$$\ell^{-1} \xi_k 2^p = 2^{p-p'} \cdot \begin{cases} 1/2 + \zeta_{p',q'}, & \zeta_{p',q'} \in [0,1/2] \\ -1/2 + \zeta_{p',q'}, & \zeta_{p',q'} \in [-1/2,0]. \end{cases} \tag{3.60}$$

For $p \ne p'$, the absolute value of (3.60) is either greater than 1 or less than 1/2 and hence (3.60) is outside the support of $\hat{\psi}$.

On the other hand, for $p = p'$, (3.60) is inside the support of $\hat{\psi}$ and so

$$\Phi_{k,l} = e^{-i2\pi q \zeta_{p,q'}}, \quad |q'| \le M_p, \quad |q| \le N_p, \tag{3.61}$$

which constitute the same random partial Fourier matrix that we have seen before. In other words, under the assumption (3.57) the sensing matrix $\mathbf{\Phi} = [\Phi_{k,l}] \in \mathbb{C}^{M \times N}$, with $N = \sum_{|p| \le p_*} (2N_p + 1)$ and $M = \sum_{|p| \le p_*} (2M_p + 1)$, is block diagonal with each block (indexed by p) in the form of the random partial Fourier matrix, representing the sensing matrix on the dyadic scale 2^p.

Near-Field Diffraction with Fourier Basis

Consider near-field diffraction by a periodic, extended object (e.g., diffraction grating) where the evanescent modes as well as the propagation modes are taken

into account. Since we cannot apply the paraxial approximation, we resort to the Lippmann–Schwinger equation (3.3).

Suppose the masked object function is sparse in the Fourier basis

$$V(x) = \sum_{j=-\infty}^{\infty} \hat{V}_j e^{i2\pi jx/L} \tag{3.62}$$

where L is the period and only s modes have nonzero amplitudes. Suppose that $\hat{V}_j = 0$ for $j \neq 1, \ldots, N$.

The 2D Green function can be expressed by the Sommerfeld integral formula (Born and Wolf 1999)

$$G(\mathbf{r}) = \frac{i}{4\pi} \int e^{i\omega(|z|\beta(\alpha)+x\alpha)} \frac{d\alpha}{\beta(\alpha)}, \quad \mathbf{r} = (z, x) \tag{3.63}$$

where

$$\beta(\alpha) = \begin{cases} \sqrt{1-\alpha^2}, & |\alpha| < 1 \\ i\sqrt{\alpha^2-1}, & |\alpha| > 1 \end{cases}. \tag{3.64}$$

The integrand in (3.63) with real-valued β (i.e., $|\alpha| < 1$) corresponds to the homogeneous wave, and that with imaginary-valued β (i.e., $|\alpha| > 1$) corresponds to the evanescent (inhomogeneous) wave, which has an exponential-decay factor $e^{-\omega|z|\sqrt{\alpha^2-1}}$. Likewise, the 3D Green function can be represented by the Weyl integral formula (Born and Wolf 1999).

The signal arriving at the sensor located at $(0, x)$ is given by the Lippmann–Schwinger equation with (3.63)

$$\int G(z_0, x - x')V(x')dx' = \frac{i}{2\omega} \sum_j \frac{\hat{V}_j}{\beta_j} e^{i\omega z_0 \beta_j} e^{i\omega \alpha_j x} \tag{3.65}$$

where

$$\alpha_j = \frac{2\pi j}{L\omega}, \quad \beta_j = \beta(\alpha_j). \tag{3.66}$$

The subwavelength structure is encoded in $\hat{V}_j$ with $\alpha_j > 1$ corresponding to the evanescent modes.

Let $(0, x_k)$, $x_k = \xi_k L$, $k = 1, \ldots, M$ be the coordinates of the sampling points where $\xi_k \in [-1/2, 1/2]$. In other words, L is also the aperture (i.e., the sampling range for x_k). To set the problem in the framework of compressed sensing, we set the vector $\mathbf{f} = (f_j) \in \mathbb{C}^N$ as

$$f_j = \frac{ie^{i\omega z_0 \beta_j}}{2\omega\beta_j} \hat{V}_j. \tag{3.67}$$

To avoid a vanishing denominator in (3.67), we assume that $\alpha_j \neq 1$ and hence $\beta_j \neq 0$, $\forall j \in \mathbb{Z}$. This is the case, for instance, when $L\omega/(2\pi)$ is irrational.

This gives rise to the sensing matrix $\mathbf{\Phi}$ with the entries

$$\Phi_{kj} = e^{i\omega\alpha_j x_k} = e^{i2\pi j\xi_k}, \quad k = 1,\ldots,M,\ j = 1,\ldots,N \tag{3.68}$$

which again is the random partial Fourier matrix.

A source of instability lurks in the expression (3.67) where β_j may be complex valued, corresponding to the evanescent modes. Stability in inverting the relationship (3.67) requires limiting the number of the evanescent modes involved in (3.67). Here, the transition is however not clear-cut. For example, if we demand that

$$|e^{i\omega z_0\beta_j}| \geq e^{-2\pi} \tag{3.69}$$

as the criterion for stable modes, then the stable modes include $|\alpha_j| \leq 1$ as well as $|\alpha_j| > 1$ such that

$$\omega|\beta_j|z_0 \leq 2\pi \tag{3.70}$$

or equivalently

$$\frac{|j|}{L} \leq \sqrt{\frac{\omega^2}{4\pi^2} + \frac{1}{z_0^2}} \tag{3.71}$$

In other words, the number of *stably* resolvable modes is proportional to the probe frequency and inversely proportional to the distance z_0 between the sensor array and the object. As z_0 drops below the wavelength, the subwavelength Fourier modes of the object can be stably recovered. This is the idea behind the near-field imaging systems such as scanning microscopy.

Inverse Scattering

In the inverse scattering theory, the scattering amplitude is the observable data, and the main objective then is to reconstruct ν from the knowledge of the scattering amplitude.

Pixel Basis

To obtain a sensing matrix with compressive sensing properties, we first make the Born approximation in (3.7) and neglect the scattered field u^s on the right-hand side of (3.7). Our purpose here is to demonstrate how to coordinate the incidence direction and the sampling direction and create a favorable sensing matrix.

Consider the incidence field

$$u^i(\mathbf{r}) = e^{i\omega\mathbf{r}\cdot\hat{\mathbf{d}}} \tag{3.72}$$

where $\hat{\mathbf{d}}$ is the incident direction. Under the Born approximation, we have from (3.7) that

$$A(\hat{\mathbf{r}},\hat{\mathbf{d}}) = A(\mathbf{s}) = \frac{\omega^2}{4\pi}\int_{\mathbb{R}^d} v(\mathbf{r}')e^{-i\omega\mathbf{r}'\cdot\mathbf{s}}\,d\mathbf{r}' \tag{3.73}$$

where $\mathbf{s} = \hat{\mathbf{r}} - \hat{\mathbf{d}}$ is the scattering vector.

We proceed to discretize the continuous system (3.73) as before. Consider the discrete approximation of the extended object ν:

$$v_\ell(\mathbf{r}) = \sum_{\mathbf{q}\in\mathbb{Z}_N^2} b\left(\frac{\mathbf{r}}{\ell} - \mathbf{q}\right) v(\ell\mathbf{q}) \tag{3.74}$$

where

$$b(\mathbf{r}) = \begin{cases} 1, & \mathbf{r} \in \left[-\frac{1}{2}, \frac{1}{2}\right]^2 \\ 0, & \text{else.} \end{cases} \tag{3.75}$$

is the pixel basis.

Define the target vector $\mathbf{f} = (f_j) \in \mathbb{C}_N$ with $f_j = \nu(\ell\mathbf{p})$, $\mathbf{p} = (p_1, p_2) \in \mathbb{Z}_N^2$, $j = (p_1 - 1)\sqrt{N} + p_2$. Let ω_l and $\hat{\mathbf{d}}_l$ be the probe frequencies and directions, respectively, and let $\hat{\mathbf{r}}_l$ be the sampling directions for $l = 1, \ldots, M$. Let $\mathbf{g}$ be the data vector with

$$g_l = \frac{4\pi A(\hat{\mathbf{r}}_l - \hat{\mathbf{d}}_l)}{\omega^2 \hat{b}((\ell\omega_l/2\pi)(\hat{\mathbf{r}}_l - \hat{\mathbf{d}}_l))}.$$

Then the sensing matrix takes the form

$$\Phi_{lj} = e^{i\omega_l \ell \mathbf{q}\cdot(\hat{\mathbf{d}}_l - \hat{\mathbf{r}}_l)}, \quad \mathbf{q} = (q_1, q_2) \in \mathbb{Z}_N^2, \quad j = (q_1 - 1)\sqrt{N} + q_2. \tag{3.76}$$

Sampling Schemes

Our strategy is to construct a sensing matrix analogous to the random partial Fourier matrix. To this end, we write the (l, j)-entry of the sensing matrix in the form

$$e^{i\pi(j_1\xi_l + j_2\zeta_l)}, \quad j = (j_1 - 1)\sqrt{N} + j_2, \quad j_1, j_2 = 1, \ldots, \sqrt{N}, l = 1, \ldots, M$$

where ξ_l, ζ_l are independently and uniformly distributed in $[-1, 1]$. We write (ξ_l, ζ_l) in the polar coordinates ρ_l, ϕ_l as

$$(\xi_l, \zeta_l) = \rho_l(\cos\phi_l, \sin\phi_l), \quad \rho_l = \sqrt{\xi_l^2 + \zeta_l^2} \le \sqrt{2} \tag{3.77}$$

and set

$$\omega_1(\cos\theta_l - \cos\tilde{\theta}_l) = \sqrt{2}\rho_l\Omega\cos\phi_l$$

$$\omega_1(\sin\theta_l - \sin\tilde{\theta}_l) = \sqrt{2}\rho_l\Omega\sin\phi_l$$

where Ω is a parameter to be determined later (3.91). Equivalently, we have

$$-\sqrt{2}\omega_l \sin\frac{\theta_l - \tilde{\theta}_l}{2} \sin\frac{\theta_l + \tilde{\theta}_l}{2} = \Omega\rho_l\cos\phi_l \tag{3.78}$$

$$-\sqrt{2}\omega_l \sin\frac{\theta_l - \tilde{\theta}_l}{2} \cos\frac{\theta_l + \tilde{\theta}_l}{2} = \Omega\rho_l\sin\phi_l. \tag{3.79}$$

This set of equations determines the single-input-(θ_l, ω_l)-single-output-$\tilde{\theta}_l$ mode of sampling.

The following implementation of (3.78) through (3.79) is natural. Let the sampling angle $\tilde{\theta}_l$ be related to the incident angle θ_l via

$$\theta_l + \tilde{\theta}_l = 2\phi_l + \pi, \tag{3.80}$$

and set the frequency ω_l to be

$$\omega_l = \frac{\rho_l \Omega}{\sqrt{2}\sin((\theta_l - \tilde{\theta}_l)/2)}. \tag{3.81}$$

Then the entries (3.76) of the sensing matrix $\mathbf{\Phi}$ have the form

$$e^{i\sqrt{2}\Omega\ell(j_1\xi l + j_2\xi l)}, \quad l = 1,\ldots,n,\ j_1, j_2 = 1,\ldots,\sqrt{N}. \tag{3.82}$$

By the square symmetry of the problem, it is clear that the relation (3.80) can be generalized to

$$\theta_l + \tilde{\theta}_l = 2\phi_l + \eta\pi, \quad \eta \in \mathbb{Z}. \tag{3.83}$$

On the other hand, the symmetry of the square lattice should not play a significant role and hence we expect the result to be insensitive to any *fixed* $\eta \in \mathbb{R}$, independent of l, as long as (3.81) holds. Indeed, this is confirmed by numerical simulations.

Let us focus on two specific measurement schemes.

Backward Sampling

This scheme employs Ω—band limited probes, that is, $\omega_l \in [-\Omega, \Omega]$. This and (3.81) lead to the following constraint:

$$\left|\sin\frac{\theta_l - \tilde{\theta}_l}{2}\right| \geq \frac{\rho_l}{\sqrt{2}}. \tag{3.84}$$

The simplest way to satisfy (3.80) and (3.84) is to set

$$\phi_l = \tilde{\theta}_l = \theta_l + \pi, \tag{3.85}$$

$$\omega_l = \frac{\rho_l \Omega}{\sqrt{2}} \tag{3.86}$$

$l = 1, \ldots, n$. In this case, the scattering amplitude is always sampled in the backscattering direction. This resembles the synthetic aperture imaging, which has been previously analyzed under the paraxial approximation in Fannjiang et al. (2010). In contrast, the forward scattering direction with $\tilde{\theta}_l = \theta_l$ almost surely violates the constraint (3.84).

Forward Sampling

This scheme employs single frequency probes no less than Ω:

$$\omega_l = \gamma\Omega, \quad \gamma \geq 1,\ l = 1, \ldots, n. \tag{3.87}$$

To satisfy (3.83) and (3.81), we set

$$\theta_l = \phi_l + \frac{\eta\pi}{2} + \arcsin\frac{\rho_l}{\gamma\sqrt{2}} \tag{3.88}$$

$$\tilde{\theta}_l = \phi_l + \frac{\eta\pi}{2} - \arcsin\frac{\rho_l}{\gamma\sqrt{2}} \tag{3.89}$$

with $\eta \in \mathbb{Z}$. The difference between the incident angle and the sampling angle is

$$\theta_l - \tilde{\theta}_l = 2\arcsin\frac{\rho_l}{\gamma\sqrt{2}}, \tag{3.90}$$

which diminishes as $\gamma \to \infty$. In other words, in the high frequency limit, the sampling angle approaches the incident angle. This resembles the setting of x-ray tomography.

In summary, let ξ_l, ζ_l be independently and uniformly distributed in $[-1, 1]$ and let (ρ_l, ϕ_l) be the polar coordinates of (ξ_l, ζ_l), that is,

$$(\xi_l, \zeta_l) = \rho_l\,(\cos\phi_l, \sin\phi_l).$$

Then with

$$\Omega\ell = \frac{\pi}{\sqrt{2}}, \tag{3.91}$$

both forward and backward samplings give rise to the random partial Fourier sensing matrix.

Coherence Bounds for Single Frequency

As in the "Fresnel Diffraction with Point Objects" section, we let the point scatterers be continuously distributed over a finite domain, not necessarily on a grid. Any computational imaging would involve some underlying, however refined, grid. Hence, let us assume that there is an underlying, possibly highly refined and unresolved, grid of spacing $\ell \ll \omega_l^{-1}$ (the reciprocal of probe frequency).

We shall focus on the monochromatic case with $\omega_l = \omega$, $l = 1, \ldots, M$. We recall that the sensing matrix is continuous of the form (3.76), which now becomes

$$\phi_{lj} = e^{i\omega\ell\mathbf{p}\cdot(\tilde{\mathbf{d}}_l - \tilde{\mathbf{r}}_l)}, \quad j = (p_1 - 1)\sqrt{N} + p_2, \quad \mathbf{p} \in \mathbb{Z}_N^2. \tag{3.92}$$

In other words, the measurement diversity comes entirely from the variations of the incidence and detection directions. We assume that the n incident directions and the m detection directions are each independently chosen according to some distributions with the total number of data $M = nm$ fixed.

Theorem 3.5 (2D case) *Suppose the incident and sampling angles are randomly, independently, and identically distributed according to the probability density functions* $f^i(\theta) \in C^1$ *and* $f^s(\theta) \in C^1$*, respectively. Suppose*

$$N \le \frac{\varepsilon}{8} e^{K^2/2}, \quad \varepsilon, K > 0. \tag{3.93}$$

Set $L = \ell|\mathbf{p} - \mathbf{p}|$ *for any* $\mathbf{p}, \mathbf{q} \in \mathbb{Z}_N^2$*. Then the sensing matrix satisfies the pairwise coherence bound*

$$\mu_{\mathbf{p},\mathbf{q}} < \left(\bar{\mu}^i + \frac{\sqrt{2K}}{\sqrt{n}}\right)\left(\bar{\mu}^s + \frac{\sqrt{2K}}{\sqrt{m}}\right). \tag{3.94}$$

with probability greater than $(1-\varepsilon)^2$*, where*

$$\bar{\mu}^i \le c(1+\omega L)^{-1/2} \sup_\theta \left\{ \left| f^i(\theta) \right|, \left| \frac{d}{d\theta} f^i(\theta) \right| \right\}, \tag{3.95}$$

$$\bar{\mu}^s \le c(1+\omega L)^{-1/2} \sup_\theta \left\{ \left| f^s(\theta) \right|, \left| \frac{d}{d\theta} f^s(\theta) \right| \right\}, \tag{3.96}$$

with a constant c.

In 3D, the coherence bound can be improved with a faster decay rate in terms of $\omega L \gg 1$ as stated in the following.

Theorem 3.6 (3D case) *Assume* (3.93)*. Suppose the incidence and sampling directions, parameterized by the polar angle* $\theta \in [0, \pi]$ *and the azimuthal angle* $\phi \in [0, 2\pi]$*, are randomly, independently, and identically distributed. Let* $f^i(\theta) \in C^1$ *and* $f^s(\theta) \in C^1$ *be the marginal density functions of the incident and sampling polar angles, respectively.*

Let $L = \ell|\mathbf{p} - \mathbf{p}|$*. Then the sensing matrix satisfies the pairwise coherence bound*

$$\mu_{\mathbf{p},\mathbf{q}} < \left(\bar{\mu}^i + \frac{\sqrt{2K}}{\sqrt{n}}\right)\left(\bar{\mu}^s + \frac{\sqrt{2K}}{\sqrt{m}}\right). \tag{3.97}$$

with probability greater than $(1-\varepsilon)^2$*, where*

$$\bar{\mu}^i \le c(1+\omega L)^{-1} \sup_\theta \left\{ \left| f^i(\theta) \right|, \left| \frac{d}{d\theta} f^i(\theta) \right| \right\} \tag{3.98}$$

$$\bar{\mu}^s \le c(1+\omega L)^{-1} \sup_\theta \left\{ \left| f^s(\theta) \right|, \left| \frac{d}{d\theta} f^s(\theta) \right| \right\}. \tag{3.99}$$

Remark 3.5 The original statements of the theorems (Fannjiang 2010b, Theorems 1 and 6) have been adapted to the present context of off-grid objects. The original proofs, however, carry over here verbatim upon minor change of notation.

Remark 3.6 When the sampling directions are randomized and the incidence directions are deterministic, then the coherence bounds (3.94) and (3.97) hold with the first factor on the right-hand side removed.

According to Remark 3.6, we have the pairwise coherence bound:

$$\text{(2D)}\quad \mu_{\mathbf{p},\mathbf{q}} \le c(1+\omega L)^{-1/2} \sup_{\theta}\left\{\left|f^s(\theta)\right|, \left|\frac{d}{d\theta} f^s(\theta)\right|\right\} + \frac{\sqrt{2}K}{\sqrt{M}} \tag{3.100}$$

$$\text{(3D)}\quad \mu_{\mathbf{p},\mathbf{q}} \le c(1+\omega L)^{-1} \sup_{\theta}\left\{\left|f^s(\theta)\right|, \left|\frac{d}{d\theta} f^s(\theta)\right|\right\} + \frac{\sqrt{2}K}{\sqrt{M}} \tag{3.101}$$

which is an estimate of the coherence pattern of the sensing matrix. Hence, if L is unresolvable (i.e., $\omega L \le 1$), the corresponding pairwise coherence parameter is high, and if L is well resolved (i.e., $\omega L \gg 1$), the corresponding pairwise coherence parameter is low. A typical coherence band has a coherence radius $\mathcal{O}(\omega^{-1})$ according to (3.100) and (3.101).

Therefore, if the point objects are well separated in the sense that any pair of objects are larger than ω^{-1}, then the same BLO- and BLOT-based techniques discussed in the "Fresnel Diffraction with Point Objects" section can be used to recover the masked object support and amplitudes. For a simple illustration, Figure 3.11 shows two instances of reconstruction by BOMP. The recovered objects (blue asterisks) are close to the true objects (red circles) well within the coherence bands (yellow patches).

Inverse Multiple Scattering

In this section, we present an approach to compressive imaging of multiply scattering point scatterers. First consider the multiple scattering effect with just a single illumination, that is, $n = 1$ and $M = m$.

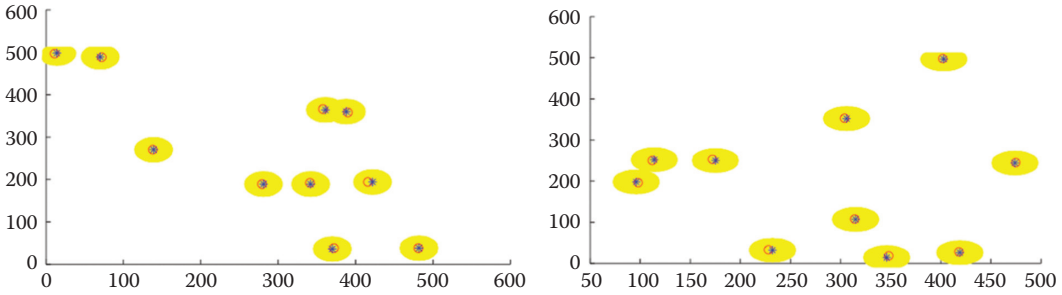


Figure 3.11 Two instances of BOMP reconstruction: red circles are the exact locations, blue asterisks are recovered locations, and the yellow patches are the coherence bands around the objects.

Note that the original object support is the same as the masked object support. With the support accurately recovered, let us consider how to unmask the objects and recover the true objects.

Define the incidence and full field vectors at the locations of the objects:

$$\mathbf{u}^i = (u^i(\mathbf{r}_1),\ldots,u^i(\mathbf{r}_s))^T \in \mathbb{C}^s$$

$$\mathbf{u} = (u(\mathbf{r}_1),\ldots,u(\mathbf{r}_s))^T \in \mathbb{C}^s.$$

Let $\mathbf{\Gamma}$ be the $s \times s$ matrix

$$\mathbf{\Gamma} = [(1 - \delta_{jl})G(\mathbf{r}_j, \mathbf{r}_l)]$$

and $\mathcal{V}$ the diagonal matrix

$$\mathcal{V} = \mathrm{diag}(v_1, \ldots, v_s).$$

The full field is determined by the Foldy–Lax equation (Mishchenko et al. 2006)

$$\mathbf{u} = \mathbf{u}^i + \omega^2\mathbf{\Gamma}\mathcal{V}\mathbf{u} \tag{3.102}$$

from which we obtain the full field

$$\mathbf{u} = (\mathbf{I} - \omega^2\mathbf{\Gamma}\mathcal{V})^{-1}\mathbf{u}^i \tag{3.103}$$

and the masked objects

$$\begin{aligned}\mathbf{f} = \mathcal{V}\mathbf{u} &= \mathcal{V}(\mathbf{I} - \omega^2\mathbf{\Gamma}\mathcal{V})^{-1}\mathbf{u}^i \\ &= (\mathbf{I} - \omega^2\mathcal{V}\mathbf{\Gamma})^{-1}\mathcal{V}\mathbf{u}^i\end{aligned} \tag{3.104}$$

provided that ω^{-2} is not an eigenvalue of $\mathbf{\Gamma}\mathcal{V}$.

Hence, by solving (3.104) we have

$$(\mathbf{I} - \omega^{-2}\mathcal{V}\mathbf{\Gamma})\mathbf{f} = \mathcal{V}\mathbf{u}^i. \tag{3.105}$$

The true objects ν can then be recovered by solving (3.105) as

$$v = \frac{\mathbf{f}}{\omega^2\mathbf{\Gamma}\mathbf{f} + \mathbf{u}^i} \tag{3.106}$$

where the division is carried out entrywise (Hadamard product).

Joint Sparsity

With the total number of data $M = nm$ fixed, the coherence bounds (3.94) and (3.97) are optimized with $n \sim m \sim \sqrt{M}$. To take advantage of this result, we should deploy multiple incidence fields for which the formula (3.106) is no longer valid.

Multiple illuminations give rise to multiple data vectors $\mathbf{g}_j$, and multiple masked object vectors $\mathbf{f}_j$, $j = 1, \ldots, n$, each of which is masked by an unknown field $\mathbf{u}_j$. However, all masked object vectors give rise to the same sensing matrix

$$\Phi_{lj} = e^{-i\omega\ell\mathbf{p}\cdot\hat{\mathbf{r}}^l}, \quad j = (p_1 - 1)\sqrt{N} + p_2, \mathbf{p} \in \mathbb{Z}_N^2.$$

Since every masked object vector shares the same support as the true object vector, this is a suitable setting for the application of joint sparsity techniques discussed in the "BPDN for joint sparsity" and "OMP for joint sparsity" sections.

Compiling the masked object vectors as $\mathbf{F} = [\mathbf{f}_1, \ldots, \mathbf{f}_n] \in \mathbb{C}^{m\times n}$ and the data vectors as $\mathbf{G} = [\mathbf{g}_1, \ldots, \mathbf{g}_n] \in \mathbb{C}^{m\times n}$, we obtain the imaging equations

$$\mathbf{G} = \mathbf{\Phi F} + \mathbf{E} \tag{3.107}$$

where $\mathbf{E}$ accounts for noise. When the true objects are widely separated, we have two ways to proceed as follows.

BPDN-BLOT for Joint Sparsity

In the first approach, we use BPDN for joint sparsity (3.37) with $\mathbf{\Phi}_j = \mathbf{\Phi}$, $\forall j$, $\mathcal{L} = 0$ to solve the imaging equation (3.107). Let $\mathbf{F}_* = (\mathbf{f}_{1*}, \ldots, \mathbf{f}_{n*})$ be the solution. We then apply the BLOT technique (Algorithm 3.5) to improve $\mathbf{F}_*$ (i.e., trim the bushes and grow the trees). In order to enforce the joint sparsity structure, we modify Algorithm 3.5 as follows.

First, we modify the LO algorithm to account for joint sparsity.

Algorithm 3.7 LO for Joint Sparsity

Inputs: $\mathbf{\Phi}_1, \ldots, \mathbf{\Phi}_n, \mathbf{G}, \eta > 0, S^0 = \{i_1, \ldots, i_s\}$.

Iteration: For $k = 1, 2, \ldots, s$.

(1) $\mathbf{F}^k = \arg\min \|[\mathbf{\Phi}_1\mathbf{h}_1, \ldots, \mathbf{\Phi}_n\mathbf{h}_n] - \mathbf{G}\|_F$ s.t. $\cup_j \text{supp}(\mathbf{h}_j) \subseteq (S^{k-1} \setminus \{i_k\}) \cup \{i_k'\}$, $i_k' \in B_\eta(\{i_k\})$.

(2) $S^k = \text{supp}(\mathbf{F}^k)$.

Output: S^s.

Next, we modify the BLOT algorithm to account for joint sparsity.

Algorithm 3.8 BLOT for Joint Sparsity

Input: $\mathbf{f}_1, \ldots, \mathbf{f}_n, \mathbf{\Phi}_1, \ldots, \mathbf{\Phi}_n, \mathbf{G}, \eta > 0$.

Initialization: $S^0 = \emptyset$.

Iteration: For $k = 1, 2, \ldots, s$.

(1) $i_k = \arg\max_j \|\mathbf{f}_j\|_2, k \notin B_\eta^{(2)}(S^{k-1})$.

(2) $S^k = S^{k-1} \cup \{i_k\}$.

Output: $\mathbf{F}_* = \arg\min \|[\mathbf{\Phi}_1\mathbf{h}_1, \ldots, \mathbf{\Phi}_n\mathbf{h}_n] - \mathbf{G}\|_F$, $\cup_j \text{supp}(\mathbf{h}_j) \subseteq \text{JLO}(S^s)$, where $\text{JLO}(S^s)$ is the output of Algorithm 3.7 with the sth iterate S^s of BLOT as input.

BLOOMP for Joint Sparsity

In the second approach, we propose the following joint sparsity version of BLOOMP.

Algorithm 3.9 BLOOMP for Joint Sparsity

Input: $\mathbf{\Phi}_1, \ldots, \mathbf{\Phi}_1, \mathbf{G}, \eta > 0$

Initialization: $\mathbf{F}^0 = 0$, $\mathbf{R}^0 = \mathbf{G}$ and $S^0 = \emptyset$

Iteration: For $k = 1, \ldots, s$

(1) $i_{\max} = \arg\max_i \sum_{j=1}^{J} | \Phi_{j,i}^{\dagger} \mathbf{r}_j^{k-1} |, i \notin B_{\eta}^{(2)}(S^{k-1})$, where $\Phi_{j,i}^{\dagger}$ = conjugate transpose of $\mathrm{col}_i(\mathbf{\Phi}_j)$.

(2) $S^k = \mathrm{JLO}(S^{k-1} \cup \{i_{\max}\})$, where JLO is the output of Algorithm 3.7.

(3) $[\mathbf{f}_1^k, \ldots, \mathbf{f}_n^k] = \arg\min_{\mathbf{H}} \|[\mathbf{\Phi}_1 \mathbf{h}_1, \ldots, \mathbf{\Phi}_n \mathbf{h}_n] - \mathbf{G}\|_{\mathrm{F}}$ s.t. $\cup_j \mathrm{supp}(\mathbf{h}_j) \subseteq S^k$

(4) $[\mathbf{r}_1^k, \ldots, \mathbf{r}_n^k] = \mathbf{G} - [\mathbf{\Phi}_1 \mathbf{f}_1^k, \ldots, \mathbf{\Phi}_n \mathbf{f}_n^k]$

Output: $\mathbf{F}_* = [\mathbf{f}_1^s, \ldots, \mathbf{f}_n^s]$.

After the first stage of either approach, we obtain an estimate of the object support as well as the amplitudes of masked objects. In the second stage, we estimate the true object amplitudes. If we use the formula (3.106) for each incident wave $\mathbf{u}_j^{\mathrm{i}}$, we end up with n amplitude estimates

$$\frac{\mathbf{f}_{j*}}{\omega^2 \Gamma \mathbf{f}_{j*} + \mathbf{u}_j^{\mathrm{i}}}, \quad j = 1, \ldots, n$$

that are typically inconsistent. Least squares is the natural way to solve this overdetermined system and obtain the object estimate.

$$v_* = \arg\min_{\mathbf{v}} \sum_{j=1}^{n} \| (\omega^2 \Gamma \mathbf{f}_{j*} + \mathbf{u}_j^i)\mathbf{v} - \mathbf{f}_{j*} \|_2^2 .$$

Inverse Scattering with Zernike Basis

In this section, we discuss a basis for representing extended objects in the scattering geometry and its application to compressive inverse scattering. We shall make the Born approximation.

A well-known orthogonal basis for representing an extended object with a compact support (e.g., the unit disk) is the product of Zernike polynomials R_n^m and trigonometric functions

$$V_n^m(x, y) = V_n^m(\rho\cos\theta, \rho\sin\theta) = R_n^m(\rho) e^{im\theta}, \quad x^2 + y^2 \le 1 \tag{3.108}$$

where $m \in \mathbb{Z}$, $n \in \mathbb{N}$, $n \ge |m|$, and $n - |m|$ are even. We refer to V_n^m as the Zernike functions of order (m, n) (Born and Wolf 1999). These Zernike functions are very useful in optics because the lowest few terms of a Zernike expansion have a

simple optical interpretation (Dai and Milenkovic 2008). In addition, a Zernike expansion usually has a superior rate of convergence (hence sparser) compared with other expansions such as a Bessel-Fourier or Chebyshev-Fourier expansion (Boyd and Yu 2011; Boyd and Petschek 2014).

We show now that the Zernike basis also results in a better coherence parameter (hence better resolution) than the pixel basis. The Zernike polynomials are given explicitly by the formula

$$R_n^m(\rho) = \frac{1}{\left(\frac{n-|m|}{2}\right)\rho^{|m|}}\left[\frac{d}{d(\rho^2)}\right]^{\frac{n-|m|}{2}}\left[(\rho^2)^{\frac{n+|m|}{2}}(\rho^2-1)^{\frac{n-|m|}{2}}\right] \tag{3.109}$$

which are nth degree polynomials in ρ and normalized such that $R_n^m(1) = 1$ for all permissible values of m, n. The Zernike polynomials satisfy the following properties:

$$\int_0^1 R_n^m(\rho)R_{n'}^m(\rho)\rho\, d\rho = \frac{\delta_{nn'}}{2(n+1)} \tag{3.110}$$

$$\int_0^1 R_n^m(\rho)J_m(u\rho)\rho\, d\rho = (-1)^{\frac{n-m}{2}}\frac{J_{n+1}(u)}{u} \tag{3.111}$$

where J_{n+1} is the $(n + 1)$-order Bessel function of the first kind. As a consequence of (3.110), the Zernike functions satisfy the orthogonality property

$$\int_{x^2+y^2\le 1} \overline{V_n^m(x,y)}V_{n'}^{m'}(x,y)dx\, dy = \frac{\pi}{n+1}\delta_{mm'}\delta_{nn'}. \tag{3.112}$$

Writing $\mathbf{s} = s(\cos\phi, \sin\phi)$, let us compute the matrix element for the scattering amplitude (3.73) as follows:

$$\begin{aligned}\int_{x^2+y^2\le 1} \overline{V_n^m(x,y)}e^{-i\omega\mathbf{s}\cdot(x,y)}dx\, dy &= \int_0^1\int_0^{2\pi} e^{i\omega s\rho\cos(\phi+\theta)}R_n^m(\rho)e^{-im\theta}\, d\theta\rho\, d\rho \\ &= \int_0^1\int_0^{2\pi} e^{i\omega s\rho\cos\theta}e^{\ im\theta}\, d\theta R_n^m(\rho)\rho\, d\rho e^{im\phi} \\ &= 2\pi i^n e^{im\phi}\int_0^1 J_{\boldsymbol{m}}(\omega s\rho)R_n^m(\rho)\rho\, d\rho \end{aligned} \tag{3.113}$$

by the definition of Bessel function

$$J_m(z) = \frac{1}{\pi i^m}\int_0^\pi e^{iz\cos\theta}\cos(m\theta)d\theta.$$

Using the property (3.111), we then obtain from (3.113) that

$$\int_{x^2+y^2\le 1} \overline{V_n^m(x,y)} e^{-i\omega \mathbf{s}\cdot(x,y)} dx\, dy = 2\pi i^m (-1)^{\frac{n-m}{2}} e^{im\phi} \frac{J_{n+1}(\omega s)}{\omega s} \tag{3.114}$$

which are the sensing matrix elements with all permissible m, n. Note that the columns of the sensing matrix are indexed by the permissible $m \in \mathbb{Z}$, $n \in \mathbb{N}$ with the constraint that $n \ge |m|$ and $n - |m|$ are even.

Let the scattering vector $\mathbf{s} = \hat{\mathbf{r}} - \hat{\mathbf{d}}$ be parameterized as

$$\mathbf{s}_{jk} = s_j(\cos\phi_k, \sin\phi_k), \quad j,k = 1,\ldots,\sqrt{M}$$

such that $\{\phi_k\}$ are independently and identically distributed uniform random variables on $[0, 2\pi]$ according to the uniform distribution and that $\{s_j\}$ are independently distributed on $[0, 2\omega]$ according to the linear density function $f(r) = r/2$. As a result, $z_j = \omega s_j$ are independently and identically distributed on $[0, 2\omega]$ according to a linear density function.

Calculation of the coherence parameter between the columns corresponding to $(m, n) \ne (m', n')$ gives the following expression:

$$\left(\frac{1}{\sqrt{M}} \sum_{j=1}^{\sqrt{M}} \frac{J_{n+1}(\omega s_j)}{\omega s_j} \frac{J_{n'+1}(\omega s_j)}{\omega s_j}\right)\left(\frac{1}{\sqrt{M}} \sum_{k=1}^{\sqrt{M}} e^{i(m-m')\phi_k}\right).$$

We recall that for $p, q \in \mathbb{N}$ (Abramowitz and Stegun 1972, formula 11.4.6)

$$\int_0^\infty J_p(z) J_q(z) \frac{dz}{z} = \begin{cases} 0, & p \ne q \\ \frac{1}{2p}, & p = q \end{cases}. \tag{3.115}$$

For $M \gg 1$, we have by the law of large numbers

$$\begin{aligned} \frac{1}{\sqrt{M}} \sum_{j=1}^{\sqrt{M}} \frac{J_{n+1}(\omega s_j)}{\omega s_j} \frac{J_{n'+1}(\omega s_j)}{\omega s_j} &\sim \mathbb{E}\left[\frac{J_{n+1}(\omega r)}{\omega r} \frac{J_{n'+1}(\omega r)}{\omega r}\right] \\ &= \frac{1}{2\omega^2} \int_0^{2\omega} J_{n+1}(z) J_{n'+1}(z) \frac{dz}{z} \end{aligned} \tag{3.116}$$

and

$$\begin{aligned} \frac{1}{\sqrt{M}} \sum_{k=1}^{\sqrt{M}} e^{i(m-m')\phi k} \sim \mathbb{E} e^{i(m-m')\phi} &= \int_0^{2\pi} e^{i(m-m')\phi} g(\phi) d\phi \\ &= \delta_{mm'}. \end{aligned} \tag{3.117}$$

When $m \ne m'$, the two columns are orthogonal and the pairwise coherence parameter is zero. When $n \ne n'$, the right-hand side of (3.116) becomes $\mathcal{O}(\omega^{-3})$ in view of (3.115) and the fact that the Bessel functions $J_n(z)$ decay like $z^{-1/2}$ for $z \gg 1$. From

(3.115) and (3.116) with $n = n'$, we see that the 2-norm of the columns is $\mathcal{O}(\omega^{-2})$. After dividing (3.116) with $n \neq n'$ by the 2-norm of the columns, the coherence parameter scales at worst like ω^{-1} (for $m = m'$, $n \neq n'$).

Notice that this decay date of the coherence parameter is faster than the $\omega^{-1/2}$ behavior in (3.94) through (3.95). Hence, imaging with the Zernike basis possesses better resolution capability than with the pixel basis, all else being equal.

Interferometry with Incoherent Sources

In this last section, we discuss the compressive sensing application to optical interferometry in astronomy, which has a similar mathematical structure to that of the inverse scattering (3.92) under the Born approximation.

In astronomy, interferometry often deals with signals emitted from incoherent sources. In this section, we present the compressive sensing approach to such a problem. With the help of the van Cittert–Zernike theorem, the sensing matrix has a structure not unlike what we discussed earlier.

Suppose the field of view is small enough to be identified with a planar patch of the celestial sphere $\mathcal{P} \subset \mathbb{R}^2$, called the object plane. Let $I(\mathbf{s})$ be the radiation intensity from the point $\mathbf{s}$ on the object plane $\mathcal{P}$. Let n antennas be located in a square of size L on the sensor plane parallel to $\mathcal{P}$ with locations $L\mathbf{r}_j$, $j = 1, \ldots, n$, where $\mathbf{r}_j \in [0, 1]^2$. Then by the van Cittert–Zernike theorem (Born and Wolf 1999), the measured visibility $v(\mathbf{r}_j - \mathbf{r}_k)$ is given by the Fourier integral

$$v(\mathbf{r}_j - \mathbf{r}_k) = \int_{\mathcal{P}} I(\mathbf{s}) e^{i\omega \mathbf{s}\cdot(\mathbf{r}_j - \mathbf{r}_k)L} d\mathbf{s}. \tag{3.118}$$

Consider the discrete approximation of the extended object I with the pixel basis on the grid $\ell\mathbb{Z}_N^2$:

$$I_\ell(\mathbf{r}) = \sum_{\mathbf{q}\in\mathbb{Z}_N^2} b\left(\frac{\mathbf{r}}{\ell} - \mathbf{q}\right) I(\ell\mathbf{q}) \tag{3.119}$$

where b is given in (3.75) and

$$\mathbb{Z}_N^2 = \{\mathbf{p} = (p_1, p_2) : p_1, p_2 = 1, \ldots, \sqrt{N}\}. \tag{3.120}$$

Substituting (3.119) into (3.118), we obtain the discrete sum

$$v(\mathbf{r}_j - \mathbf{r}_k) = \ell^2 \hat{b}\left(\frac{\omega\ell L}{2\pi}(\mathbf{r}_k - \mathbf{r}_j)\right) \sum_{l=1}^{N} I_l e^{i\omega \mathbf{p}\cdot(\mathbf{r}_j - \mathbf{r}_k)\ell L}, \tag{3.121}$$

where l, $\mathbf{p}$ are related by $l = (p_1 - 1)\sqrt{N} + p_2$ and

$$\hat{b}(\xi, \eta) = \frac{\sin(\pi\xi)}{\pi\xi} \frac{\sin(\pi\eta)}{\pi\eta}.$$

For every pair (j, k) of sensors, we measure and collect the interferometric datum $v(\mathbf{r}_j - \mathbf{r}_k)$ and we want to determine I from the collection of $n(n - 1)$ real-valued data.

Let us rewrite Equation (3.121) in the form (3.10). In contrast to (3.28), we set

$$\ell = \frac{\pi}{\omega L} \tag{3.122}$$

to account for the "two-way" structure in the imaging equation (3.121). Note that ℓ is the resolution length on the celestial sphere and hence dimensionless.

Let $\mathbf{f} = (f_i) \in \mathbb{R}^N$ be the unknown object vector, that is, $f_i = \ell^2 I_i$. Let $\mathbf{g} = (g_l) \in \mathbb{R}^M$, $M = n(n-1)/2$

$$g_l = \frac{1}{\hat{b}((\mathbf{r}_k - \mathbf{r}_j)/2)} \times \begin{cases} \Re[v(\mathbf{r}_j - \mathbf{r}_k)], & l = (2n-j)(j-1)/2 + k, \quad j < k = 1, \ldots, n \\ \Im[v(\mathbf{r}_j - \mathbf{r}_k)], & l = n(n-1)/2 + (2n-j)(j-1)/2 + k, \quad j < k = 1, \ldots, n \end{cases}$$

be the data vector where $\Re$ and $\Im$ stand for, respectively, the real and imaginary parts. The sensing matrix $\mathbf{\Phi} \in \mathbb{R}^{M \times N}$ now takes the form

$$\mathbf{\Phi}_{il} = \begin{cases} \cos[2\pi \mathbf{p}_l \cdot (\mathbf{r}_j - \mathbf{r}_k)], & i = (2n-j)(j-1)/2 + k, \quad j < k \\ \sin[2\pi \mathbf{p}_l \cdot (\mathbf{r}_j - \mathbf{r}_k)], & i = n(n-1)/2 + (2n-j)(j-1)/2 + k, \quad j < k \end{cases} \tag{3.123}$$

which is no longer the simple random partial Fourier matrix for 2D as the baselines $\mathbf{r}_j - \mathbf{r}_k$ are related to one another. Nevertheless, (3.123) has a similar structure to that of the inverse scattering (3.92) when the transmitters and receivers are colocated. Note that as $(\mathbf{r}_k - \mathbf{r}_j)/2 \in [-1/2, 1/2]^2$, the denominator $\hat{b}((\mathbf{r}_k - \mathbf{r}_j)/2)$ in the definition of g_l does not vanish.

Next, we give an upper bound for the coherence parameter. For the pairwise coherence for columns i, i' corresponding to $\mathbf{p}, \mathbf{p} \in \mathbb{Z}_N^2$, we have the following calculation:

$$\begin{aligned} \mu(i,i') &= \frac{2}{n(n-1)} \left| \sum_{j<k} \cos[2\pi \mathbf{p} \cdot (\mathbf{r}_j - \mathbf{r}_k)] \cos[2\pi \mathbf{p}' \cdot (r_j - r_k)] \right. \\ &\quad \left. + \sin[2\pi \mathbf{p} \cdot (\mathbf{r}_j - \mathbf{r}_k)] \sin[2\pi \mathbf{p}' \cdot (\mathbf{r}_j - \mathbf{r}_k)] \right| \\ &= \frac{2}{n(n-1)} \left| \sum_{j<k} \cos[2\pi (\mathbf{p} - \mathbf{p}') \cdot (\mathbf{r}_j - \mathbf{r}_k)] \right| \\ &= \frac{1}{n(n-1)} \left| \sum_{j \neq k} \cos[2\pi (\mathbf{p} - \mathbf{p}') \cdot (\mathbf{r}_j - \mathbf{r}_k)] \right| \end{aligned}$$

First, we claim

$$\mu(i,i') = \frac{1}{n(n-1)} \left| \left| \sum_{j=1}^{n} e^{i2\pi (\mathbf{p} - \mathbf{p}') \cdot \mathbf{r}_j} \right|^2 - n \right|.$$

This follows from the calculation

$$\left|\sum_{j=1}^{n} e^{i2\pi(\mathbf{p}-\mathbf{p}')\cdot\mathbf{r}_j}\right|^2 - n = \sum_{j\neq k} e^{i2\pi(\mathbf{p}-\mathbf{p}')\cdot(\mathbf{r}_j-\mathbf{r}_k)}$$

$$= \sum_{j\neq k} \cos\left[2\pi(\mathbf{p}-\mathbf{p}')\cdot(\mathbf{r}_j-\mathbf{r}_k)\right] + i\sin\left[2\pi(\mathbf{p}-\mathbf{p}')\cdot(\mathbf{r}_j-\mathbf{r}_k)\right]$$

$$= \sum_{j\neq k} \cos\left[2\pi(\mathbf{p}-\mathbf{p}')\cdot(\mathbf{r}_j-\mathbf{r}_k)\right]$$

Some modification of the arguments for Theorems 3.5 and 3.6 leads to the following coherence bound.

Theorem 3.7 *Assume that the total number of grid point N satisfies the bound*

$$N \leq \frac{\varepsilon}{2} e^{K^2/2}. \tag{3.124}$$

with some constants δ *and K. Suppose that the sensor locations* $\mathbf{r}_j$, $j = 1,\ldots,n$ *are independent uniform random variables on* $[0, 1]^2$. *Then the coherence parameter* μ *satisfies the bound*

$$\mu(\Phi) \leq \frac{\left|2K^2 - 1\right|}{n-1}. \tag{3.125}$$

with probability greater than $1 - 2\varepsilon$.

In other words, with high probability, the coherence parameter for the uniform distribution decays as n^{-1}. A central problem in interferometry is the design of an optimal array, see Fannjiang (2013b) for a discussion from the perspective of compressed sensing.

Acknowledgments

The research is supported in part by NSF grant DMS-1413373 and Simons Foundation grant 275037.

References

Abramowitz, M. and I. Stegun, *Handbook of Mathematical Functions* (New York: Dover, 1972).

Benedek, P. and R. Panzone, The space ℓ^p with mixed norm, *Duke Math. J.* 28 (1961): 301–324.

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

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

Born, M. and E. Wolf, *Principles of Optics*, 7th edn. (Cambridge, U.K.: Cambridge University Press, 1999).

Boyd, J.P. and R. Petschek, The relationships between Chebyshev, Legendre and Jacobi polynomials: The generic superiority of Chebyshev polynomials and three important exceptions, *J. Sci. Comput.* 59 (2014): 1–27.

Boyd, J.P. and F. Yu, Comparing six spectral methods for interpolation and the Poisson equation in a disk: Radial basis functions, Logan-Shepp ridge polynomials, Fourier-Bessel, Fourier-Chebyshev, Zernike polynomials, and double Chebyshev series, *J. Comput. Phys.* 230 (2011): 1408–1438.

Boyd, S. and L. Vandenberghe, *Convex Optimization*. (Cambridge, U.K.: Cambridge University Press, 2004).

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

Candès, E.J., The restricted isometry property and its implications for compressed sensing, *C. R. Acad. Sci., Paris, Ser. I.* 346 (2008): 589–592.

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

Candès, E.J. and C. Fernandez-Granda, Super-resolution from noisy data, *J. Fourier Anal. Appl.* 19(6) (2013): 1229–1254.

Candès, E.J. and C. Fernandez-Granda, Towards a mathematical theory of super-resolution, *Commun. Pure Appl. Math.* 67(6) (2014): 906–956.

Candès, E.J., J. Romberg, and T. Tao, Robust uncertainty principles: Exact signal reconstruction from highly incomplete frequency information, *IEEE Trans. Inf. Theory* 52 (2006): 489–509.

Candès, E.J. and T. Tao, Decoding by linear programming, *IEEE Trans. Inf. Theory* 51 (2005): 4203–4215.

Chambolle, A., An algorithm for total variation minimization and applications, *J. Math. Imaging Vis.* 20 (2004): 89–97.

Chambolle, A. and P.-L. Lions, Image recovery via total variation minimization and related problems, *Numer. Math.* 76 (1997): 167–188.

Chen, J. and X. Huo, Theoretical results on sparse representations of multiple-measurement vectors, *IEEE Trans. Signal Process.* 54 (2006): 4634–4643.

Chen, S.S., D.L. Donoho, and M.A. Saunders, Atomic decomposition by basis pursuit, *SIAM Rev.* 43 (2001): 129–159.

Cotter, S.F., B.D. Rao, K. Engan, and K. Kreutz-Delgado, Sparse solutions to linear inverse problems with multiple measurement vectors, *IEEE Trans. Signal Process.* 53 (2005): 2477–2488.

Dai, G.-M. and V.N. Mahajan, Orthonormal polynomials in wavefront analysis: error analysis, *Appl. Opt.* 47 (2008): 3433–3445.

Dai, W. and O. Milenkovic, Subspace pursuit for compressive sensing: Closing the gap between performance and complexity, *IEEE Trans. Inf. Theory* 55 (2009): 2230–2249.

Davis, G.M., S. Mallat, and M. Avellaneda, Adaptive greedy approximations, *J. Construct. Approx.* 13 (1997): 57–98.

Daubechies, I., *Ten Lectures on Wavelets.* (Philadelphia, PA: SIAM, 1992).

Donoho, D.L., M. Elad, and V.N. Temlyakov, Stable recovery of sparse overcomplete representations in the presence of noise, *IEEE Trans. Inf. Theory* 52 (2006): 6–18.

Duarte, M.F. and R.G. Baraniuk, Spectral compressive sensing, *Appl. Comput. Harmon. Anal.* 35 (2013): 111–129.

Fannjiang, A., Compressive imaging of subwavelength structures, *SIAM J. Imaging Sci.* 2 (2009): 1277–1291.

Fannjiang, A., Compressive inverse scattering I. High-frequency SIMO/MISO and MIMO measurements, *Inverse Probl.* 26 (2010a): 035008.

Fannjiang, A., Compressive inverse scattering II. SISO measurements with Born scatterers, *Inverse Probl.* 26 (2010b): 035009.

Fannjiang, A., TV-min and greedy pursuit for constrained joint sparsity and application to inverse scattering, *Math. Mech. Complex Syst.* 1 (2013a): 81–104.

Fannjiang, A. and W. Liao, Coherence-pattern guided compressive sensing with unresolved grids, *SIAM J. Imaging Sci.* 5 (2012a): 179–202.

Fannjiang, A. and W. Liao, Super-resolution by compressive sensing algorithms, in *IEEE Proceedings of Asilomar Conference on Signals, Systems and Computers*, 2012b.

Fannjiang, A., T. Strohmer, and P. Yan, Compressed remote sensing of sparse objects, *SIAM J. Imaging Sci.* 3 (2010): 596–618.

Fannjiang, C., Optimal arrays for compressed sensing in snapshot-mode interferometry, *Astron. Astrophys.* 559 (2013b): A73–A84.

Kowalski, M., Sparse regression using mixed norms, *Appl. Comput. Harmon. Anal.* 27 (2009): 303–324.

Mishchenko, M.I., L.D. Travis, and A.A. Lacis, *Multiple Scattering of Light by Particles: Radiative Transfer and Coherent Backscattering.* (Cambridge, U.K.: Cambridge University Press, 2006).

Needell, D. and J.A. Tropp, CoSaMP: Iterative signal recovery from incomplete and inaccurate samples, *Appl. Comput. Harmon. Anal.* 26 (2009): 301–329.

Pati, Y.C., R. Rezaiifar, and P.S. Krishnaprasad, Orthogonal matching pursuit: Recursive function approximation with applications to wavelet decomposition, in *Proceedings of the 27th Asilomar Conference in Signals, Systems and Computers*, 1993.

Rauhut, H., Stability results for random sampling of sparse trigonometric polynomials, *IEEE Trans. Inf. Theory* 54 (2008): 5661–5670.

Rudin, L. and S. Osher, Total variation based image restoration with free local constraints, in *Proceedings of IEEE ICIP*, Vol. 1, 1994, pp. 31–35.

Rudin, L.I., S. Osher, and E. Fatemi, Nonlinear total variation based noise removal algorithms, *Phys. D* 60 (1992): 259–268.

Tang, G., B. Bhaskar, P. Shah, and B. Recht, Compressed sensing off the grid, *IEEE Trans. Inf. Theory* 59 (2013): 7465–7490.

Tibshirani, R., Regression shrinkage and selection via the lasso, *J. R. Stat. Soc. Ser. B* 58 (1996): 267–288.

Tropp, J.A., Greed is good: Algorithmic results for sparse approximation, *IEEE Trans. Inf. Theory* 50 (2004): 2231–2242.

Tropp, J.A., A.C. Gilbert, and M.J. Strauss, Algorithms for simultaneous sparse approximation. Part I: Greedy pursuit, *Signal Process* (*Special Issue on Sparse Approximations in Signal and Image Processing*) 86 (2006): 572–588.

Taylor & Francis
Taylor & Francis Group
http://taylorandfrancis.com