

Exact localization and superresolution with noisy data and random illumination

This article has been downloaded from IOPscience. Please scroll down to see the full text article.

2011 Inverse Problems 27 065012

(<http://iopscience.iop.org/0266-5611/27/6/065012>)

View [the table of contents for this issue](#), or go to the [journal homepage](#) for more

Download details:

IP Address: 71.142.210.102

The article was downloaded on 21/05/2011 at 17:54

Please note that [terms and conditions apply](#).

Exact localization and superresolution with noisy data and random illumination

Albert C Fannjiang

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

E-mail: fannjiang@math.ucdavis.edu

Received 18 August 2010, in final form 25 February 2011

Published 20 May 2011

Online at stacks.iop.org/IP/27/065012

Abstract

This paper studies the problem of exact localization of multiple objects with noisy data. The crux of the proposed approach consists of random illumination. Two recovery methods are analyzed: the Lasso and the one-step thresholding (OST). For independent random probes, it is shown that both recovery methods can localize exactly $s = \mathcal{O}(m)$, up to a logarithmic factor, objects where m is the number of data. Moreover, when the number of random probes is large the Lasso with random illumination has a performance guarantee for superresolution, beating the Rayleigh resolution limit. Numerical evidence confirms the predictions and indicates that the performance of the Lasso is superior to that of the OST for the proposed setup with random illumination.

(Some figures in this article are in colour only in the electronic version)

1. Introduction

Two-point resolution is a standard criterion for evaluation of imaging systems, i.e. the ability of the imaging system to distinguish two closely located point objects. The smallest resolvable distance ℓ between two objects, called the (two-point) resolution length, is then defined as a metric of the resolving power of the imaging system. Let A be the aperture of the imaging system, z_0 the distance to the objects and λ the wavelength. The classical Rayleigh resolution criterion then states

$$\frac{A\ell}{z_0\lambda} = \mathcal{O}(1) \quad (1)$$

where there is some arbitrariness in the constant depending on the precise definition of the minimum resolvable length ℓ .

For noisy data, such a criterion is more difficult to apply as determination of ℓ becomes a statistical problem. One option would be to formulate the two-point resolution problem as a statistical-hypothesis-testing problem (one versus two objects), see [23, 33] and references therein. However, it is cumbersome to generalize this approach to multiple point objects.

In this paper, we first study the resolution issue from the perspective of exact, simultaneous localization of multiple point objects. We evaluate an imaging method by saying that it can exactly localize s (sparsity) randomly distributed point objects mutually separated by a minimum distance ℓ with high probability. In addition to reconsidering the issue of resolution, we seek an approach that can recover a high number $s = \mathcal{O}(m)$ of objects where m is the number of data, with resolution ℓ far below what is dictated by the Rayleigh resolution limit (1) (see remark 3). This latter effect is called superresolution.

Consider the noisy data model:

$$Y = \Phi X + E, \quad \|E\|_2 \leq \varepsilon, \quad (2)$$

where $X \in \mathbb{C}^N$ is the object to be recovered, $Y \in \mathbb{C}^m$ is the data vector and $E \in \mathbb{C}^m$ represents noise. We assume that Φ has unit-norm columns. This can always be realized by redefining the object vector X .

Sparse object reconstruction for this model can be broken into two steps: localization (i.e. support recovery) and strength estimation. For underdetermined systems, the former, being combinatorial in nature, is by far more difficult than the latter which is a straightforward inversion if the former is exact. The former step is called *model selection* in linear regression and machine learning theory [3, 5, 6, 10, 12, 26, 28, 35, 41] from which one of the reconstruction methods studied in the present paper originates.

Exact localization with noisy data is challenging. Many reconstruction methods guarantee stability (i.e. the reconstruction error bounded by a constant multiple of the noise level) but not necessarily exact localization. Orthogonal matching pursuit (OMP) is a simple greedy algorithm with proven guarantee of exact localization for sufficiently small noise and worst-case coherence.

A basic quantity for stability analysis in compressed sensing is the notion of coherence. Let the worst-case coherence $\mu(\Phi)$ be defined as

$$\mu(\Phi) = \max_{i \neq j} \frac{|\Phi_j^* \Phi_i|}{\|\Phi_j\|_2 \|\Phi_i\|_2}. \quad (3)$$

A standard result is as follows [18].

Proposition 1. *Consider the signal model (2). Suppose the sparsity s of the real-valued object vector $X \in \mathbb{R}^N$ satisfies*

$$s < \frac{1}{2} \left(1 + \frac{1}{\mu} \right) - \frac{\varepsilon}{\mu X_{\min}}, \quad X_{\min} = \min_{i \in \mathcal{S}} |X_i|.$$

Denote by \hat{X}^ε the output of OMP which stops as soon as the residual error (in ℓ^2 -norm) is no greater than ε . Then

(i) \hat{X}^ε has the correct support, i.e.

$$\text{supp}(\hat{X}^\varepsilon) = \text{supp}(X);$$

(ii) \hat{X}^ε approximates the true object vector

$$\|\hat{X}^\varepsilon - X\|_2^2 \leq \frac{\varepsilon^2}{1 - \mu(s - 1)}.$$

The general lower bound [17, 40]

$$\sqrt{\frac{N - m}{m(N - 1)}} \leq \mu$$

for the mutual coherence of any $m \times N$ matrix Φ implies that the sparsity s allowed by proposition 1 is $\mathcal{O}(\sqrt{m})$ for $N \gg m$.

A main purpose of the paper is to explore the utility of two other methods from compressed sensing theory, the one-step thresholding (OST) [3] and the Lasso [15, 35], that have the potential for exact localization of a much higher number $\mathcal{O}(m)$ of objects.

The OST, proposed in [3], involves just one matrix multiplication plus *thresholding*. Compute $Z = \Phi^* Y$ and determine the set of points

$$\hat{\mathcal{S}} = \{i \in \{1, \dots, N\} : |Z_i| > \tau_*\}$$

for some threshold τ_* . In other words, the OST is the linear processor of matched field processing (MFP) plus a thresholding step [2]. On the other hand, the linear processor of MFP is the same as the first iterate of OMP. Consequently, OST has even lower complexity than OMP which is its main appeal.

For the OST's performance guarantee, we need the notion of average coherence defined as [3]

$$\nu(\Phi) = \frac{1}{N-1} \max_{j'} \left| \sum_{j \neq j'} \Phi_{j'}^* \Phi_j \right|$$

in addition to the worst-case coherence.

The following is the performance guarantee for OST [3].

Proposition 2. *Consider the signal model (2). Assume that $X \in \mathbb{R}^N$ is drawn from the generic s -sparse ensemble of real-valued objects. Assume E to be distributed as $\text{CN}(0, \sigma^2 \mathbf{I})$, the complex Gaussian random vectors with the covariance matrix $\sigma^2 \mathbf{I}$.*

Suppose

$$\mu(\Phi) \leq \frac{c_1}{\sqrt{m}} \leq \frac{1}{\sqrt{10 \log N}} \quad (4)$$

for some $c_1 > 0$ (which may depend on $\log N$) and

$$\nu(\Phi) \leq \frac{12\mu(\Phi)}{\sqrt{m}}. \quad (5)$$

Assume $\|X\|_2 = 1$. Define the threshold

$$\tau_* = 4\sqrt{\log N} \max\{\sigma, 12\mu\sqrt{2}\}. \quad (6)$$

Suppose the number of objects obeying

$$s \leq \frac{m}{2 \log N} \quad (7)$$

and that

$$X_{\min} = \min_{i \in \mathcal{S}} |X_i| > 2\tau_*. \quad (8)$$

Then, the OST with threshold τ_* satisfies $\mathbb{P}(\hat{\mathcal{S}} \neq \mathcal{S}) \leq 9/N$.

In other words, for sufficiently small worst-case coherence (4) and average coherence (5) and noise (8), OST can exactly localize $\mathcal{O}(m)$ objects, up to a logarithmic factor with high probability. Once the support is exactly recovered, an estimate \hat{X} can be obtained by pseudo-inversion on the object support \mathcal{S} .

The other method studied in this paper is the Lasso [35]. The Lasso estimate \hat{X} is defined as the solution to

$$\min_Z \frac{1}{2} \|Y - \Phi Z\|_2^2 + \gamma \sigma \|Z\|_1, \quad \gamma > 0, \quad (9)$$

where γ is a regularization parameter.

The following sufficient condition for exact localization by the Lasso is given by [12].

Proposition 3. Consider the signal model (2). Assume that $X \in \mathbb{R}^N$ is drawn from the generic s -sparse ensemble of real-valued objects. Assume E to be distributed as $\text{CN}(0, \sigma^2 \mathbf{I})$.

Suppose that Φ obeys the coherence property

$$\mu(\Phi) \leq \frac{a_0}{\log N} \quad (10)$$

with some positive constant a_0 . Suppose

$$s \leq \frac{c_0 N}{\|\Phi\|_2^2 \log N} \quad (11)$$

for some positive constant c_0 . Let S be the support of X and suppose

$$X_{\min} > 8\sigma \sqrt{2 \log N}. \quad (12)$$

Then the Lasso estimate \hat{X} with $\gamma = 2\sqrt{2 \log N}$ obeys

$$\text{supp}(\hat{X}) = \text{supp}(X) \quad (13)$$

$$\text{sign}(\hat{X}) = \text{sign}(X) \quad (14)$$

with probability at least $1 - 2N^{-1}((2\pi \log N)^{-1/2} + sN^{-1}) - \mathcal{O}(N^{-2 \log 2})$.

Some comparison between propositions 3 and 2 is in order. Both deal with randomly distributed objects. Both (4) and (10) are sufficiently weak assumptions for most imaging problems. Also (12) and (8) are similar when $\mu = \mathcal{O}(\sigma)$. The lower bounds for the success probabilities are comparable up to a logarithmic factor. The main technical assumption of proposition 3 is (11), while for proposition 2 it is (5). When the operator norm $\|\Phi\|_2$ obeys the bound $\|\Phi\|_2 = \mathcal{O}(N/m)$ condition (11) is comparable to (7).

A drawback to proposition 2 is that the thresholding rule (6) requires the precise knowledge of μ which can only be calculated numerically. As we will see, the Lasso-based method also has a better numerical performance than the OST does (cf figures 5 and 7).

To realize the potential of the two above results in imaging, we consider the idea of *random illumination* (RI) for point scatterers (figure 1). We show that a suitable condition of RI enables us to (i) obtain a guaranteed exact localization of $s = \mathcal{O}(m)$, up to a logarithmic factor, objects and to (ii) harness the superresolution capability (i.e. breaking the Rayleigh resolution limit (1)).

Previously we have studied the problems of imaging point scatterers [21, 24] using coherence and operator-norm bounds. We demonstrate that the imaging performance can be significantly improved by RI. In particular, suitable RI leads to superresolution.

However, both propositions 2 and 3 share the following common drawbacks: (i) they are restricted to random objects; (ii) they do not address the reconstruction error when the error level is above threshold and exact localization is unattainable; (iii) they are limited to the i.i.d. Gaussian noise model. Issue (i) is pertinent particularly to imaging extended objects whose supports are clearly not random. Issue (ii) is related to robustness with respect to a wider range of error. Issue (iii) arises in optics where the Poisson or shot noise model is more appropriate.

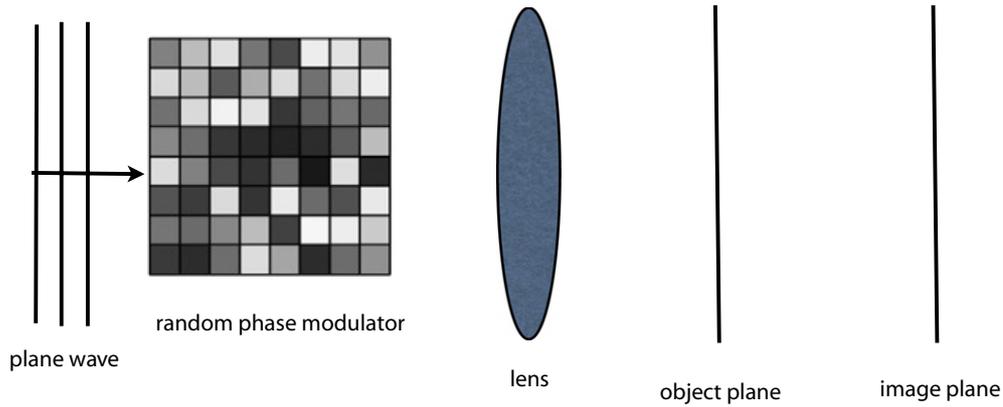


Figure 1. The imaging geometry for point objects.

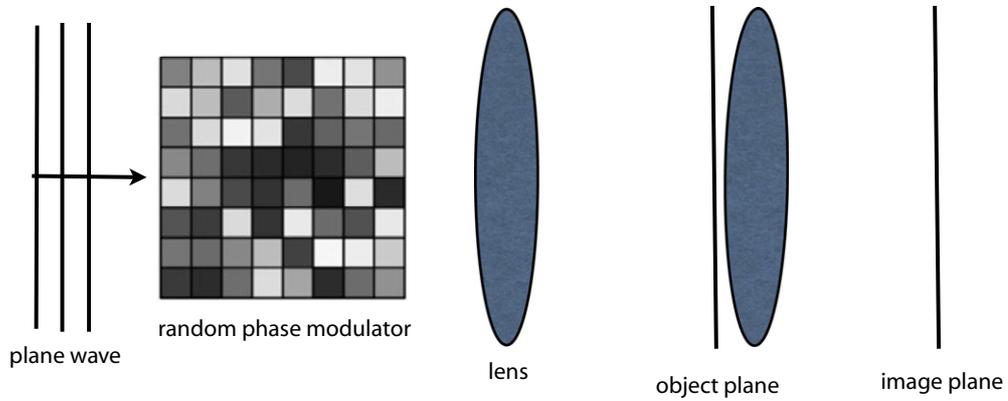


Figure 2. The imaging geometry for extended objects.

The standard compressed sensing method that is without any of the above limitations is the basis pursuit denoising (BPDN)

$$\min_Z \|Z\|_1, \quad \text{s.t. } \|Y - \Phi Z\|_2 \leq \varepsilon \tag{15}$$

[15]. BPDN, of course, is equivalent to the Lasso (55) for an appropriately chosen γ .

The performance guarantee for BPDN is typically given in terms of the restricted isometry property (RIP) due to Candès and Tao [13]. Precisely, let the sparsity s of a vector $Z \in \mathbb{C}^N$ be the number of nonzero components of Z and define the restricted isometry constant (RIC) $\delta_s \in [0, 1]$ to be the *smallest* nonnegative number such that the inequality

$$(1 - \delta_s)\|Z\|_2^2 \leq \|\Phi Z\|_2^2 \leq (1 + \delta_s)\|Z\|_2^2 \tag{16}$$

holds for all $Z \in \mathbb{C}^N$ of sparsity at most s . BPDN has the following performance guarantee [11].

Proposition 4. *Suppose the RIC satisfies the bound*

$$\delta_{2s} < \sqrt{2} - 1. \tag{17}$$

Then, the BPDN minimizer \hat{X} is unique and satisfies the error bound

$$\|\hat{X} - X\|_2 \leq C_1 s^{-1/2} \|X - X^{(s)}\|_1 + C_2 \varepsilon$$

where $X^{(s)}$ is the best s -sparse approximation of X and C_1, C_2 are absolute constants depending on δ_{2s} only.

In proposition 4, BPDN does not guarantee the exact recovery of the discrete support X , which is less important for extended objects, but also does not have any of the limitations mentioned above for propositions 2 and 3.

The plan for the rest of the paper is as follows. In section 2, we review the forward scattering problem and the paraxial approximation. We describe the setup of RI in the paraxial regime. In section 3 we state and prove the main results. In section 4, we analyze the performance of BPDN with RI for extended objects and discuss the issue of resolution in imaging extended objects. In section 5 we give the worst-case coherence bounds. In section 6, we give the average coherence bound. In section 7, we give an operator norm bound of $\mathcal{O}(N/m)$ required to guarantee a nearly optimal performance for the Lasso. In section 8, we present numerical simulations to verify the predictions and show the superiority of the Lasso over the OST for the setup of RI. We also present numerical results for extended objects. We conclude in section 9.

2. Point scatterers and paraxial approximations

Let \mathcal{L} be a finite square lattice of spacing ℓ in the object plane $\{z = 0\} \subset \mathbb{R}^3$:

$$\mathcal{L} = \{\mathbf{r}_l : l = 1, \dots, N\} = \{(i\ell, j\ell) : i, j = 1, \dots, \sqrt{N}\}, \quad l = (i-1)\sqrt{N} + j, \quad (18)$$

and suppose that s point scatterers are located at grid points of \mathcal{L} . The total number of grid points N is a perfect square.

Let $\tau_j \in \mathbb{C}, l = 1, \dots, N$, be the reflectivity of the scatterers. The scattered field u^s obeys

$$u^s(\mathbf{r}) = \sum_{j=1}^N \tau_j G(\mathbf{r}, \mathbf{r}_j) (u^i(\mathbf{r}_j) + u^s(\mathbf{r}_j)) \quad (19)$$

for any $\mathbf{r} \notin \{\mathbf{r}_k : \tau_k \neq 0\}$ where u^i is the incident field and

$$G(\mathbf{r}, \mathbf{r}') = \frac{e^{i\omega|\mathbf{r}-\mathbf{r}'|}}{4\pi|\mathbf{r}-\mathbf{r}'|}, \quad \forall \mathbf{r} \neq \mathbf{r}' \in \mathbb{R}^3, \quad (20)$$

is the Green function of the operator $-(\Delta + \omega^2)$.

In the Born scattering approximation, u^s on the right-hand side of (19) is neglected, resulting in

$$u^s(\mathbf{r}) = \sum_{j=1}^N \tau_j G(\mathbf{r}, \mathbf{r}_j) u^i(\mathbf{r}_j). \quad (21)$$

Let $\mathbf{a}_j, j = 1, \dots, n$, be the locations of the sensors in the sensor plane $\{z = z_0\} \subset \mathbb{R}^3$ and write $\mathbf{a}_j = (\xi_j, \eta_j, z_0)$ where ξ_j and η_j are chosen independently and uniformly from the discrete subset of $[0, A]$:

$$\mathcal{D} = \left\{ \frac{qA}{\sqrt{N}} : q = 1, \dots, \sqrt{N} \right\} \quad (22)$$

where A is the aperture of the sensor array.

In the Fresnel approximation under the condition

$$\frac{(A + \ell\sqrt{N})^4}{\lambda z_0^3} \ll 1 \quad (23)$$

the Green function G can be approximated by

$$G_{\text{par}}(\mathbf{r}, \mathbf{a}) = \frac{e^{i\omega z_0}}{4\pi z_0} e^{i\omega|x-\xi|^2/(2z_0)} e^{i\omega|y-\eta|^2/(2z_0)}, \quad \mathbf{r} = (x, y, 0), \quad \mathbf{a} = (\xi, \eta, z_0), \quad (24)$$

called the paraxial Green function.

In the subsequent analysis we assume both the Born and paraxial approximations in the scattering model.

A main ingredient of the proposed approach is RI which has recently been used extensively for wavefront reconstruction and imaging [1, 19, 30]. Here, we consider random *phase* modulation (RPM) which is a random perturbation of the phase of a wavefront while maintaining the amplitude of the near-field beam almost constant. The advantage of phase modulation, compared to amplitude modulation, is the lossless energy transmission of an incident wavefront through the modulator. In optics, RPM can be created by random phase plates, digital holograms or liquid crystal panels [8, 34].

3. Main results for point objects

We assume that as a result of p independent realizations of random phase modulators the incident field at the grid points can be represented as $e^{i\theta_{kj}}$, $k = 1, \dots, p$, $j = 1, \dots, N$, where θ_{kj} are i.i.d. *uniform* random variables in $[0, 2\pi]$ (i.e. circularly symmetric). The information about θ_{kj} is incorporated in the sensing matrix (figure 1).

Let the scattered field u_k^s be measured and collected by n sensors located at \mathbf{a}_l , $l = 1, \dots, n$. Let $X = (\tau_j)_1^N \in \mathbb{C}^N$ be the object vector and $Y = (Y_i) = (u_k^s(\mathbf{a}_j)) \in \mathbb{C}^{np}$, $i = (k-1)n + j$, $j = 1, \dots, n$, the data vector.

After proper normalization, the data vector Y can be written as (2) with the sensing matrix Φ being the column-normalized version of $[G_{\text{par}}(\mathbf{a}_l, \mathbf{r}_j)u_k^i(\mathbf{r}_j)]$, i.e.

$$\phi_{ij} = \frac{1}{\sqrt{np}} e^{i\omega|x_j-\xi_l|^2/(2z_0)} e^{i\omega|y_j-\eta_l|^2/(2z_0)} e^{i\theta_{kj}}, \quad i = (k-1)n + l. \quad (25)$$

Here $m = np$ is the number of data.

Our first result is a performance guarantee for the OST with RI in the diffraction-limited case satisfying the Rayleigh resolution criterion.

Theorem 1. *Let*

$$N^2 \leq \frac{\delta}{2} e^{K^2/2}, \quad \delta, K > 0. \quad (26)$$

Suppose

$$np \geq 40K^4 \log N \quad (27)$$

and

$$\frac{A\ell}{\lambda z_0} = 1. \quad (28)$$

Then with a probability of at least

$$1 - 2\delta - 4t\sqrt{\frac{2}{\pi}} - \frac{4}{\sqrt{p}} - \frac{4}{\sqrt{n}} - 8Ne^{-12t^2\sqrt{\frac{N-1}{np}}}, \quad \forall t > 0, \quad (29)$$

OST with the threshold (6) can localize exactly s objects satisfying (7)–(8).

Remark 1. The constants δ and K in (26) are controlling parameters. δ can be adjusted to control the lower bound (29) for success probability and then K can be adjusted to control the number of grid points in the computation domain and the number of data.

For example, suppose $\delta = 1\%$ is acceptable. Then, (26) with $K = 10$ implies a computation domain of up to $0.1e^{25}/\sqrt{2}$ grid points.

Proof. The proof of theorem 1 relies on proposition 2 and the following three lemmas.

Lemma 1. Under (26), the worst-case coherence satisfies

$$\mathbb{P} \left\{ \mu(\Phi) \leq \frac{aK\sqrt{2}}{\sqrt{p}} + \frac{2K^2}{\sqrt{np}} \right\} \geq 1 - 2\delta \quad (30)$$

where a is given by (35).

In particular, if (28) holds, then $a = 0$ and (30) becomes

$$\mathbb{P}\{\mu(\Phi) \leq 2K^2/\sqrt{np}\} \geq 1 - 2\delta. \quad (31)$$

The proof of lemma 1 is given in section 5. The utility of estimate (30) lies in the situation where both the aperture and the sensor number are limited but the number of probe waves is exceedingly large (see remark 3). For the proof of theorem 1 we need estimate (31).

Lemma 2. Under the assumption (28),

$$\mathbb{P} \left[\mu(\Phi) \geq \frac{2t_1 t_2}{\sqrt{np}} \right] \geq \left(1 - 2t_1 \sqrt{\frac{2}{\pi}} - \frac{4}{\sqrt{p}} \right) \left(1 - 2t_2 \sqrt{\frac{2}{\pi}} - \frac{4}{\sqrt{n}} \right). \quad (32)$$

Lemma 2 is an easy consequence of the Berry–Esseen theorem and its proof is given in section 5.2.

Lemma 3. Let (28) hold true. Then for any $c > 0$,

$$\mathbb{P} \left\{ \nu(\Phi) \leq \frac{c}{np} \right\} \geq 1 - 8N e^{-\frac{c}{2} \sqrt{\frac{N-1}{np}}}. \quad (33)$$

The proof of lemma 3 is given in section 6.

First of all, by the upper bound (30) for the worst-case coherence and setting $c_1 = 2K^2$ in (4), the first inequality of (4) holds with a probability of at least $1 - 2\delta$. The second inequality of (4) follows from (26) and (27) and holds with a probability of at least $1 - 2\delta$.

Second, the lower bound (32) for the worst case coherence, with $t_1 = t_2 = t$, and the upper bound (33), with $c = 24t^2$, for the average coherence imply that (5) holds with a probability of at least

$$1 - 4t \sqrt{\frac{2}{\pi}} - \frac{4}{\sqrt{p}} - \frac{4}{\sqrt{n}} - 8N e^{-12t^2 \sqrt{\frac{N-1}{np}}}.$$

This completes the proof of theorem 1. \square

Our second result is a performance guarantee for the Lasso with RI.

Theorem 2. Let (26) hold and suppose

$$\frac{aK\sqrt{2}}{\sqrt{p}} + \frac{2K^2}{\sqrt{np}} \leq \frac{a_0}{\log N} \quad (34)$$

where

$$a = \max_{j \neq j'} |\mathbb{E}(e^{t\xi_l \omega(x_{j'} - x_j)/z_0}) \mathbb{E}(e^{t\eta_l \omega(y_{j'} - y_j)/z_0})|. \quad (35)$$

Assume that the s objects are real valued and satisfy (12) and

$$s \leq \frac{c_0 np}{2 \log N}. \quad (36)$$

Then the Lasso estimate \hat{X} with $\gamma = 2\sqrt{2 \log N}$ has the same support as X with a probability of at least

$$1 - 2\delta - \rho n(n-1) \frac{\pi}{2} \sqrt{\frac{np-1}{N}} - 2n^2 p(p-1) e^{-\frac{N}{(np-1)^2}} - 2N^{-1}((2\pi \log N)^{-1/2} + sN^{-1}) - \mathcal{O}(N^{-2 \log^2}). \quad (37)$$

Remark 2. While it requires that $N \gg np$ for the bound (29) to approach unity, it demands a much stronger assumption $N \gg \max\{pn^5, p^2n^2\}$ for the bound (37) to behave the same way. Numerical evidence indicates the latter to be a pessimistic estimate.

For the special case of a single sensor $n = 1$, the probability lower bound (37) is substantially improved and requires $N \gg p^2$ to approach unity. On the other hand, for (29) to approach unity, it is necessary that $n \rightarrow \infty$ (hence $n = 1$ is not an option).

Remark 3. The superresolution effect can occur when the number p of random probes is large. Consider, for example, the case of $n = 1$ and hence the aperture A is essentially zero. Since $a \leq 1$, the conditions

$$\frac{K\sqrt{2} + 2K^2}{\sqrt{p}} \leq \frac{a_0}{\log N}$$

and

$$s \leq \frac{c_0 p}{2 \log N}$$

imply that the Lasso with $\gamma = 2\sqrt{2 \log N}$ recovers exactly the support of s objects with a probability of at least that given by (37).

This superresolution effect should be compared to that with deterministic near-field illumination [20].

Proof. The proof of theorem 2 uses proposition 3, lemma 1 and the following operator-norm bound.

Lemma 4. We have

$$\mathbb{P} \left\{ \|\Phi\|_2^2 < \frac{2N}{np} \right\} \geq 1 - \rho n(n-1) \frac{\pi \sqrt{np-1}}{2\sqrt{N}} - 2n^2 p(p-1) e^{-\frac{N}{(np-1)^2}}. \quad (38)$$

On one hand, lemma 1 and (34) imply that (10) holds with a probability of at least $1 - 2\delta$.

On the other hand, lemma 4 and (36) imply that (11) holds with a probability of at least that given by the right-hand side of (38).

Combining the two and using proposition 3 we obtain the desired statement of theorem 2. \square

To further demonstrate the advantage of RI, let us consider the imaging setup of *multistatic responses* (MR) which consists of an array of n fixed transceivers which are both sources and sensors (i.e. transceivers). One by one, each transceiver of the array emits an impulse and the entire array of transceivers records the echo. Each transmitter–receiver pair gives rise to a datum and there are altogether n^2 data forming a data matrix called the MR matrix. By the

reciprocity of the wave equation, the MR matrix is symmetric and hence has at most $n(n+1)/2$ degrees of freedom.

Recalling the coherence and operator norm bounds established in [24] and using proposition 3 as in the proof of theorem 2 (below), we have the following result [24] analogous to theorem 2.

Proposition 5. *Let the locations of the n transceivers be i.i.d. uniform random variables in $[0, A]^2$. Let (27) and (28) hold true.*

Suppose

$$n \geq \frac{K^2 \log N}{a_0}$$

and that the s real-valued objects satisfy (12) and

$$s \leq \frac{c_0 n(n+1)}{4 \log N}. \quad (39)$$

Then, the Lasso estimate \hat{X} with $\gamma = 2\sqrt{2 \log N}$ has the same support as X with a probability of at least

$$1 - 2\sqrt{2\delta} - \frac{\rho n^{5/2}(n+1)^{5/2}}{\pi 2^{5/2} N^{1/2}} - 2N^{-1}((2\pi \log N)^{-1/2} + sN^{-1}) - \mathcal{O}(N^{-2 \log 2}). \quad (40)$$

Remark 4. The main drawback of the lower bound (40) lies in the third term which requires $N \gg n^{10}$ to diminish.

More generally, one can consider the case of p transmitters and n receivers, all randomly and independently distributed in $[0, A]^2$. Then an extension of the bound (40), which is omitted here, requires $N \gg n^5 p^5$ (cf [21]).

A fair comparison of proposition 5 and theorem 2 would be to set $p = (n+1)/2$ and match their degrees of freedom, i.e. $n(n+1)/2$. However, proposition 5 does not guarantee superresolution when (28) is violated preventing the worst-case coherence from being sufficiently small due to the deterministic nature of the illumination. Also, the probability lower bound (40) has a less favorable scaling behavior ($N \gg n^{10}$) than (37) for $p = (n+1)/2$ ($N \gg n^6$, cf remark 2). Indeed, the numerical simulations show that the recovery with RI has a higher success rate than the MR recovery (figures 3 and 4).

4. Sparse extended objects

We extend the above results to the case of sparse extended objects here (figure 3).

We pixelate the sparse extended object with N pixels \square_j , $j = 1, \dots, N$, of size ℓ to create a piecewise constant approximation of the object. The centers of the pixels are identified as \mathcal{L} given in (18). Let $O(\mathbf{r})$ be the original object and O_ℓ its ℓ -discretization, i.e.

$$O_\ell = \sum_{j=1}^N \mathbb{I}_{\square_j} O(\mathbf{r}_j)$$

where \mathbb{I}_{\square_j} is the indicator function of the pixel \square_j . We reconstruct the discrete approximation O_ℓ by determining the object function restricted to \mathcal{L} , denoted still by $X = (O(\mathbf{r}_j))$, by compressed sensing techniques.

Under the RI u_k^i , pixel \square_j now produces a signal at the sensor \mathbf{a}_l of the form

$$O(\mathbf{r}_j) \int_{\square_j} G_{\text{par}}(\mathbf{r}, \mathbf{a}_l) e^{-i\omega x^2/(2z_0)} e^{-i\omega y^2/(2z_0)} u_k^i(x, y) dx dy$$

where the quadratic phase factors are due to the presence of a parabolic lens immediately after the object plane (figure 2). This lens is introduced here to simplify our analysis. In practice, the lens is not needed and should have a negligible effect on performance.

As for the case of point objects we assume that as a result of the RPM u_k^i takes a constant value $e^{i\theta_{kj}}$ in the pixel \square_j and that θ_{kj} are i.i.d. random variables in $[0, 2\pi]$ as a result of RPM.

The total signal produced by O_ℓ and detected at the sensor \mathbf{a}_l is

$$\begin{aligned} & \sum_j O(\mathbf{r}_j) e^{i\theta_{kj}} \int_{\square_j} G_{\text{par}}(\mathbf{r}, \mathbf{a}_l) e^{-i\omega x^2/(2z_0)} e^{-i\omega y^2/(2z_0)} dx dy \\ &= \sum_j O(\mathbf{r}_j) e^{i\theta_{kj}} e^{i\omega \xi_l^2/(2z_0)} e^{i\omega \eta_l^2/(2z_0)} e^{-i\omega \xi_l x_j/z_0} e^{-i\omega \eta_l y_j/z_0} \int_{\square} e^{-i\omega \xi_l x/z_0} e^{-i\omega \eta_l y/z_0} dx dy, \end{aligned}$$

plus an error term E_{kl} which includes the discretization error and external noise where \square denotes the square of size ℓ centered at the origin. Since

$$\int_{\square} e^{-i\omega \xi_l x/z_0} e^{-i\omega \eta_l y/z_0} dx dy = \frac{2z_0}{\omega \xi_l \ell} \sin\left(\frac{\omega \xi_l \ell}{2z_0}\right) \frac{2z_0}{\omega \eta_l \ell} \sin\left(\frac{\omega \eta_l \ell}{2z_0}\right) \equiv g(\mathbf{a}_l) \quad (41)$$

independent of the pixel index, we can normalize the data by dividing the signal at sensor l by this number as long as

$$\frac{\xi_l \ell}{\lambda z_0}, \frac{\eta_l \ell}{\lambda z_0} < 1, \quad \forall l = 1, \dots, n. \quad (42)$$

Dividing the data further by the phase factors $e^{i\omega \xi_l^2/(2z_0)} e^{i\omega \eta_l^2/(2z_0)}$ and \sqrt{np} , we write the signal model as (2) with the sensing matrix element

$$\phi_{ij} = \frac{1}{\sqrt{np}} e^{i\theta_{kj}} e^{-i\omega \xi_l x_j/z_0} e^{-i\omega \eta_l y_j/z_0}, \quad i = (k-1)n + l. \quad (43)$$

The difference between the signals produced by O and its discretization O_ℓ is the discretization error E_{disc} . How small must ℓ be in order for the ℓ_2 -norm of the discretization error E_{disc} to be less than, say, ε after rewriting the signal model as (2)? This can be estimated as follows.

First, by the inequality $\|E_{\text{disc}}\|_2 \leq \|E_{\text{disc}}\|_\infty \sqrt{np}$ it suffices to show $\|E_{\text{disc}}\|_\infty \leq \varepsilon/\sqrt{np}$. Since

$$u_k^i(\mathbf{r}) = \sum_{j=1}^N e^{i\theta_{kj}} \mathbb{I}_{\square_j}(\mathbf{r})$$

is the illumination field, the uncontaminated signal detected by the sensor \mathbf{a}_l in the absence of external noise in the signal model (2) is

$$(\mathcal{F}O)_i = \frac{1}{g(\mathbf{a}_l)} \sum_{j=1}^N e^{i\theta_{kj}} \int_{\square_j} O(x, y) e^{-i\omega \xi_l x_j/z_0} e^{-i\omega \eta_l y_j/z_0} dx dy, \quad (44)$$

for $i = (k-1)n + l$. On the other hand, we have

$$(\mathcal{F}O_\ell)_i = \frac{1}{g(\mathbf{a}_l)} \sum_{j=1}^N e^{i\theta_{kj}} O(\xi_j, y_j) \int_{\square_j} e^{-i\omega \xi_l x_j/z_0} e^{-i\omega \eta_l y_j/z_0} dx dy$$

for $i = (k-1)n + l$. By definition

$$E_{\text{disc}} = \mathcal{F}O - \mathcal{F}O_\ell \in \mathbb{C}^{pn}$$

and hence

$$\|E_{\text{disc}}\|_\infty \leq \frac{\|O - O_\ell\|_{L^1}}{\min_l |g(\mathbf{a}_l)|} \quad (45)$$

where $\|\cdot\|_{L^1}$ denotes

$$\|f\|_{L^1} = \int |f(x, y)| dx dy,$$

i.e. the norm of the function space L^1 . Therefore, we have the following statement.

Lemma 5. *If*

$$\|O - O_\ell\|_{L^1} \leq \frac{\varepsilon}{\sqrt{np}} \min_l |g(\mathbf{a}_l)|, \quad (46)$$

then

$$\|E_{\text{disc}}\|_2 \leq \varepsilon.$$

Remark 5. The presence of the factor $(np)^{-1/2}$ in (46) is due to the transition from L^1 function space norm to the discrete ℓ_2 -norm.

Since the sensing matrix (25) for the point objects can be written as

$$\mathbf{D}_1 \Phi \mathbf{D}_2$$

where Φ is as (43) and

$$\begin{aligned} \mathbf{D}_1 &= \text{diag}(e^{i\omega\xi_l^2/(2z_0)} e^{i\omega\eta_l^2/(2z_0)}) \\ \mathbf{D}_2 &= \text{diag}(e^{i\omega x_j^2/(2z_0)} e^{i\omega y_j^2/(2z_0)}) \end{aligned}$$

are diagonal, unitary matrices, all the preceding results, including theorems 1 and 2, can be proved for the sensing matrix (43) by minor modification of the previous arguments.

However, the object vector $X = (O(\mathbf{r}_j))$ of an extended object generally does not fall into the category of *random point* objects assumed in either proposition 2 or 3 since by definition the discrete approximation of an extended object must cluster in aggregates and its amplitude typically changes continuously. So we take an alternative approach below by resorting to the minimization principle (15) of BPDN.

The RIC for a structured sensing matrix such as (43) is difficult to estimate directly except for the case of single shot ($p = 1$) and the case of one sensor ($n = 1$). For the one-sensor case, (43) with $(\xi_l, \eta_l) = (0, 0)$ is the complex-value version of the random i.i.d. Bernoulli matrix:

$$\phi_{kj} = \frac{1}{\sqrt{p}} e^{i\theta_{kj}}, \quad (47)$$

whose RIC can easily be estimated by the same argument given in [4]. The single-sensor imaging setup resembles that of Rice's single-pixel camera [19] which employs a discrete random screen instead of a random phase modulator.

For the single-shot case, the sensing matrix (43) is equivalent to the random partial Fourier matrix, modulo an unitary diagonal matrix, and the standard RIP estimate [29] requires the Rayleigh criterion (28) to be met which guarantees (42) with probability 1. However, a small probability of $\mathbf{a}_l = (\xi_l, \eta_l)$ exists falling near the boundary of the aperture and hence a small value of $|g(\mathbf{a}_l)|$. Normalizing the data by $|g(\mathbf{a}_l)|$ then carries a small risk of magnifying the errors.

For the general setup with multiple shots and sensors, we use the mutual coherence to bound the RIC trivially as follows.

Proposition 6. *For any $s \in \mathbb{N}$ we have*

$$\delta_s \leq \mu(\Phi)(s - 1).$$

Combining lemma 1, propositions 4 and 6 we obtain the following result.

Theorem 3. Under (26), the RIC bound (17) holds true with a probability of at least $1 - 2\delta$ for the sensing matrix (43) and sparsity up to

$$s < \frac{1}{2} + \left(\frac{1}{\sqrt{2}} - \frac{1}{2} \right) \left(\frac{aK\sqrt{2}}{\sqrt{p}} + \frac{2K^2}{\sqrt{np}} \right)^{-1} \tag{48}$$

where

$$a = \max_{j \neq j'} |\mathbb{E}(e^{i\xi_l \omega(x_{j'} - x_j)/z_0}) \mathbb{E}(e^{i\eta_l \omega(y_{j'} - y_j)/z_0})|$$

cf (35).

Furthermore, suppose the total error in the data is $E = E_{\text{disc}} + E_{\text{ext}}$ where E_{disc} and E_{ext} are, respectively, the discretization error and the external noise. Then, the reconstruction \hat{X} by BPDN satisfies the error bound

$$\|\hat{X} - X\|_2 \leq C_1 s^{-1/2} \|X - X^{(s)}\|_1 + C_2 (\|E_{\text{disc}}\|_2 + \|E_{\text{ext}}\|_2) \tag{49}$$

for all s satisfying (48).

Remark 6. Since BPDN does not guarantee exact localization, an appropriate metric for resolution can be formulated in terms of the smallest pixel size ℓ_{\min} and largest sparsity s such that (49) holds true with both the discretization error E_{disc} and $s^{-1/2} \|X - X^{(s)}\|_1$ being reasonably small.

The right definition of ‘small errors’, however, is problem specific. The discrete norms (ℓ_1 - or ℓ_2 -norm) tend to go up simply because the effective sparsity increases. Hence, the right metric of reconstruction error should be properly normalized by the size of the object. For example, consider the special case when X is s -sparse. Then, we can rewrite (49) as

$$s^{-1/2} \|\hat{X} - X\|_2 \leq C_2 \|E_{\text{ext}}\|_2 s^{-1/2} + C_2 \|E_{\text{disc}}\|_2 s^{-1/2} \tag{50}$$

whose left-hand side is a measure of the reconstruction error *per pixel of size* ℓ .

Below the diffraction limit $A\ell/(\lambda z_0) < 1$ ($A \neq 0$), one can reduce the discretization error by reducing the pixel size according to lemma 5. On the other hand, the sparsity s increases in proportion to ℓ^{-2} for a two-dimensional extended object. To satisfy (48) the smallest admissible pixel size ℓ_{\min} is bounded from below roughly by

$$\ell_{\min} \gtrsim a^{1/2} p^{-1/4} \tag{51}$$

meaning that the minimum super-resolved scale decreases at least as fast as the negative quarter power of the number of RI.

For the diffraction-limited case $a = 0$, we have instead

$$\ell_{\min} \gtrsim n^{-1/4} p^{-1/4}$$

which is more favorable than (51) for $n \gg 1$. However, the discretization error bound (lemma 5) is less useful in this case.

5. Worst-case coherence bound

5.1. Proof of lemma 1: upper bound

Proof. Summing over $\mathbf{a}_l, l = 1, \dots, n$, we obtain

$$\sum_{k=1}^p \sum_{l=1}^n \phi_{i,j'}^* \phi_{ij} = e^{i\omega(x_j^2 + y_j^2 - x_{j'}^2 - y_{j'}^2)/(2z_0)} \frac{1}{np} \sum_{k=1}^p e^{i(\theta_{kj} - \theta_{kj'})} \sum_{l=1}^n e^{i\xi_l \omega(x_{j'} - x_j)/z_0} e^{i\eta_l \omega(y_{j'} - y_j)/z_0}. \tag{52}$$

We estimate the two summations separately.

First, consider the summation over RI $k = 1, \dots, p$. Define the random variables $A_l, B_l, l = 1, \dots, n$, as

$$A_l = \cos [\theta_{kj} - \theta_{kj'}] \quad (53)$$

$$B_l = \sin [\theta_{kj} - \theta_{kj'}] \quad (54)$$

and let

$$S_p = \sum_{l=1}^p (A_l + iB_l). \quad (55)$$

To estimate S_p , we recall the Hoeffding inequality [27].

Proposition 7. *Let $A_1 + iB_1, \dots, A_p + iB_p$ be independent random variables. Assume that $A_l, B_l \in [a_l, b_l], l = 1, \dots, p$, almost surely. Then, we have*

$$\mathbb{P}[|S_p - \mathbb{E}S_p| \geq pt] \leq 4 \exp \left[-\frac{p^2 t^2}{\sum_{l=1}^p (b_l - a_l)^2} \right] \quad (56)$$

for all positive values of t .

We apply the Hoeffding inequality to S_p with $a_l = -1, b_l = 1, l = 1, \dots, p$, and

$$t = K \sqrt{\frac{2}{p}}, \quad K > 0,$$

to obtain

$$\mathbb{P} \left[p^{-1} \left| \sum_{k=1}^p e^{i(\theta_{kj} - \theta_{kj'})} \right| \geq K \sqrt{\frac{2}{p}} \right] \leq 4 e^{-K^2/2}. \quad (57)$$

Note the dependence of S_p on $\theta_{kj} - \theta_{kj'}$ and the symmetry: $|S_p(\theta_{kj'} - \theta_{kj})| = |S_p(\theta_{kj} - \theta_{kj'})|$. As a consequence, there may be $N(N-1)/2$ different values of S_p . By union bound with (57), we obtain

$$\mathbb{P} \left[p^{-1} \max_{j \neq j'} \left| \sum_{k=1}^p e^{i(\theta_{kj} - \theta_{kj'})} \right| \geq K \sqrt{\frac{2}{p}} \right] \leq 2N(N-1) e^{-K^2/2} \leq \delta \quad (58)$$

by (26).

Next, consider the summation, denoted by T_n , over the sensor locations $l = 1, \dots, n$ in (52):

$$T_n = \sum_{l=1}^n e^{i\xi_l \omega(x_{j'} - x_j)/z_0} e^{i\eta_l \omega(y_{j'} - y_j)/z_0}.$$

By the same argument we obtain

$$\mathbb{P} \left[\max_{j' \neq j} n^{-1} |T_n - \mathbb{E}T_n| \geq K \sqrt{\frac{2}{n}} \right] \leq 2N(N-1) e^{-K^2/2}$$

and hence

$$\mathbb{P} \left[\max_{j' \neq j} \frac{1}{n} |T_n| \geq a + K \sqrt{\frac{2}{n}} \right] \leq \delta, \quad a = \max_{j \neq j'} \frac{1}{n} |\mathbb{E}T_n| \quad (59)$$

by (26).

By the mutual independence of ξ_l and η_l we have

$$\begin{aligned} a &= \max_{j \neq j'} \frac{1}{n} |\mathbb{E} T_n| = \max_{j \neq j'} \frac{1}{n} \left| \sum_{l=1}^n \mathbb{E} \left(e^{i \xi_l \omega(x_{j'} - x_j)/z_0} \right) \mathbb{E} \left(e^{i \eta_l \omega(y_{j'} - y_j)/z_0} \right) \right| \\ &= \max_{j \neq j'} |\mathbb{E} (e^{i \xi_l \omega(x_{j'} - x_j)/z_0}) \mathbb{E} (e^{i \eta_l \omega(y_{j'} - y_j)/z_0})| \end{aligned}$$

since $\xi_l, \eta_l, l = 1, \dots, n$, are independently identically distributed.

Combining (59) and (58) and noting the independence of these two events, we obtain

$$\mu(\Phi) \leq \frac{aK\sqrt{2}}{\sqrt{p}} + \frac{2K^2}{\sqrt{np}}$$

with a probability of at least $1 - 2\delta$.

Simple calculation with the uniform distribution on the set \mathcal{D} given in (22) yields

$$|\mathbb{E} (e^{i \xi_l \omega(x_{j'} - x_j)/z_0}) \mathbb{E} (e^{i \eta_l \omega(y_{j'} - y_j)/z_0})| = 0, \quad j' \neq j, \quad (60)$$

if (28) holds. In this case,

$$\mu(\Phi) \leq 2K^2/\sqrt{np}$$

with probability $1 - 2\delta$.

5.2. Proof of lemma 2: lower bound

Proof. The Berry–Esseen theorem [25] states that the distribution of the sum of m independent and identically distributed zero-mean random variables normalized by its standard deviation differs from the unit Gaussian distribution by at most $C\rho/(\sigma^2\sqrt{m})$, where σ^2 and ρ are respectively the variance and the absolute third moment of the parent distribution, and C is a distribution-independent absolute constant which is not greater than 0.7655 [32].

We apply the Berry–Esseen theorem to the two summations, denoted by S_p and T_n respectively, on the right-hand side of (52).

The complex-valued random variables involved can be treated as \mathbb{R}^2 -valued random variables. Under (28) the variance of these random variables is $1/2$ and the absolute third moment is $4/(3\pi)$.

Let F_1, F_2 be the cumulative distributions of the real and imaginary parts of $p^{-1/2}S_p$ and G_1, G_2 the cumulative distributions of the real and imaginary parts of $n^{-1/2}T_n$. Let Ψ be the cumulative distribution of the standard normal random variable. We have by the Berry–Esseen theorem

$$\sup_t |F_i(t) - \Psi(t)| \leq \frac{C8\sqrt{2}}{3\pi\sqrt{p}}, \quad i = 1, 2, \quad (61)$$

$$\sup_t |G_i(t) - \Psi(t)| \leq \frac{C8\sqrt{2}}{3\pi\sqrt{n}}, \quad i = 1, 2. \quad (62)$$

Since $C \leq 0.7655$, we can replace the right-hand side of (61) and (62) by $p^{-1/2}$ and $n^{-1/2}$ respectively for the sake of notational simplicity. Hence

$$\begin{aligned} |F_i(t) - F_i(-t)| &\leq |\Psi(t) - \Psi(-t)| + \frac{2}{\sqrt{p}} \\ |G_i(t) - G_i(-t)| &\leq |\Psi(t) - \Psi(-t)| + \frac{2}{\sqrt{n}} \end{aligned}$$

$\forall t$. For small $t > 0$ we can bound the above expressions by

$$|F_i(t) - F_i(-t)| \leq t\sqrt{\frac{2}{\pi}} + \frac{2}{\sqrt{p}}$$

$$|G_i(t) - G_i(-t)| \leq t\sqrt{\frac{2}{\pi}} + \frac{2}{\sqrt{n}}$$

which imply

$$\mathbb{P}[p^{-1/2}|S_p| \leq t\sqrt{2}] \leq 2t\sqrt{\frac{2}{\pi}} + \frac{4}{\sqrt{p}}$$

$$\mathbb{P}[n^{-1/2}|T_n| \leq t\sqrt{2}] \leq 2t\sqrt{\frac{2}{\pi}} + \frac{4}{\sqrt{n}}$$

and consequently

$$\mathbb{P}\left[\frac{|S_p T_n|}{np} \geq \frac{2t_1 t_2}{\sqrt{np}}\right] \geq \left(1 - 2t_1\sqrt{\frac{2}{\pi}} - \frac{4}{\sqrt{p}}\right) \left(1 - 2t_2\sqrt{\frac{2}{\pi}} - \frac{4}{\sqrt{n}}\right) \quad (63)$$

which is what we want to prove. \square

6. Average coherence bound: proof of lemma 3

Proof. Write

$$v(\Phi) = \frac{1}{N-1} \max_{j'} \left| \sum_{l=1}^n \sum_{k=1}^p \sum_{j \neq j'} \phi_{ij'}^* \phi_{ij} \right|, \quad i = (k-1)n + l,$$

and consider the sums over k and j simultaneously with a fixed j' and fixed n sensor locations. This is a summation of $p \cdot N$ independent random variables $\phi_{ij'} \phi_{ij}$ each bounded by $n^{-1} p^{-1}$ in absolute value. Note that

$$\mathbb{E}_{\xi, \eta} [\phi_{ij'}^* \phi_{ij}] = 0, \quad \forall j, j', i, \quad (64)$$

since θ_{kj} are uniformly distributed in $[0, 2\pi]$. Applying Hoeffding inequality with

$$t = \frac{c_1}{(N-1)^{1/2} p^{3/2} n}, \quad c > 0,$$

we have

$$\mathbb{P}_{\xi, \eta} \left[\frac{1}{N-1} \left| \sum_{k=1}^p \sum_{j \neq j'} \phi_{ij'} \phi_{ij} \right| \geq \frac{c_1}{(N-1)^{1/2} p^{1/2} n} \right] \leq 4e^{-c_1^2} \quad (65)$$

where $\mathbb{P}_{\xi, \eta}$ is the probability conditioned on fixed $\xi = (\xi_j)$, $\eta = (\eta_j) \in \mathbb{R}^n$. In analyzing the sum over $l = 1, \dots, n$ we restrict to the event

$$\mathcal{A} = \left\{ \Theta = [\theta_{kj}] : \frac{1}{N-1} \left| \sum_{k=1}^p \sum_{j \neq j'} \phi_{ij'} \phi_{ij} \right| < \frac{c_1}{np^{1/2}(N-1)^{1/2}} \text{ for almost all sensor locations} \right\}.$$

Since there are at most N possible sensor locations, by (65)

$$\mathbb{P}(\mathcal{A}^c) \leq 4N e^{-c_1^2} \quad (66)$$

where \mathcal{A}^c denotes the complement of \mathcal{A} .

Let

$$Z_{j'l} = \frac{1}{N-1} \sum_{k=1}^p \sum_{j \neq j'} \phi_{ij}^* \phi_{ij}$$

and $\mathbb{E}_{\mathcal{A}}$ is the expectation conditioned on the event \mathcal{A} .

We proceed with the following estimate:

$$\begin{aligned} & \mathbb{P} \left[\max_{j'} \left| \sum_{l=1}^n (Z_{j'l} - \mathbb{E}_{\mathcal{A}} Z_{j'l}) \right| \geq \frac{c_2}{\sqrt{np(N-1)}} \right] \\ &= \mathbb{P}_{\mathcal{A}} \left[\max_{j'} \left| \sum_{l=1}^n (Z_{j'l} - \mathbb{E}_{\mathcal{A}} Z_{j'l}) \right| \geq \frac{c_2}{\sqrt{np(N-1)}} \right] \mathbb{P}(\mathcal{A}) \\ & \quad + \mathbb{P}_{\mathcal{A}^c} \left[\max_{j'} \left| \sum_{l=1}^n (Z_{j'l} - \mathbb{E}_{\mathcal{A}} Z_{j'l}) \right| \geq \frac{c_2}{\sqrt{np(N-1)}} \right] \mathbb{P}(\mathcal{A}^c) \\ &\leq \mathbb{P}_{\mathcal{A}} \left[\max_{j'} \left| \sum_{l=1}^n (Z_{j'l} - \mathbb{E}_{\mathcal{A}} Z_{j'l}) \right| \geq \frac{c_2}{\sqrt{np(N-1)}} \right] + 4N e^{-c_1^2}, \quad c_1, c_2 > 0 \end{aligned} \quad (67)$$

by (66) where $\mathbb{P}_{\mathcal{A}}$ and $\mathbb{P}_{\mathcal{A}^c}$ are respectively the probabilities conditioned on the events \mathcal{A} and \mathcal{A}^c .

Applying Hoeffding's inequality with

$$t = \frac{c_2}{p^{1/2}(N-1)^{1/2}n^{3/2}}$$

to estimate the first term on the right-hand side of (67), we obtain

$$\mathbb{P}_{\mathcal{A}} \left[\left| \sum_{l=1}^n (Z_{j'l} - \mathbb{E}_{\mathcal{A}} Z_{j'l}) \right| \geq \frac{c_2}{\sqrt{np(N-1)}} \right] \leq 4 e^{-c_2^2/(2c_1)^2}.$$

Maximizing over $j' = 1, \dots, m$ and using the union bound we then arrive at

$$\mathbb{P}_{\mathcal{A}} \left[\max_{j'} \left| \sum_{l=1}^n (Z_{j'l} - \mathbb{E}_{\mathcal{A}} Z_{j'l}) \right| \geq \frac{c_2}{\sqrt{np(N-1)}} \right] \leq 4N e^{-c_2^2/(2c_1)^2}. \quad (68)$$

Using (67) and (68) with

$$c_2 = c \sqrt{\frac{N-1}{np}}, \quad c_2 = 2c_1^2, \quad c > 0,$$

we have

$$\mathbb{P} \left[\max_{j'} \left| \sum_{l=1}^n (Z_{j'l} - \mathbb{E}_{\mathcal{A}} Z_{j'l}) \right| \geq \frac{c}{np} \right] \leq 8N e^{-\frac{c}{2} \sqrt{\frac{N-1}{np}}}, \quad c > 0,$$

which is what we set out to prove.

Note that

$$\mathbb{E}_{\Theta} Z_{j'l} = \frac{1}{n} \mathbb{E} (e^{t\xi_l \omega(x_{j'} - x_j)/z_0}) \mathbb{E} (e^{t\eta_l \omega(y_{j'} - y_j)/z_0}), \quad \forall j' = 1, \dots, m, \quad l = 1, \dots, n, \quad j' \neq j,$$

where \mathbb{E}_{Θ} is the expectation conditioned on $\Theta = (\theta_{kj}) \in \mathbb{C}^{p \times n}$. If

$$\frac{1}{\rho} = \frac{A\ell}{\lambda z_0} \in \mathbb{N},$$

then

$$\mathbb{E}_{\Theta} Z_{j'l} = 0, \quad \forall j' = 1, \dots, m, \quad l = 1, \dots, n,$$

and hence

$$\mathbb{E}_{\mathcal{A}} Z_{j'l} = 0, \quad \forall j' = 1, \dots, m, \quad l = 1, \dots, n. \quad \square$$

7. Operator norm bound: proof of lemma 4

Proof. It suffices to show that the matrix Φ satisfies

$$\left\| \frac{np}{N} \Phi \Phi^* - \mathbf{I}_{np} \right\|_2 < 1 \tag{69}$$

where \mathbf{I}_{np} is the $np \times np$ identity matrix with the corresponding probability bound. Since the diagonal elements of $\frac{np}{N} \Phi \Phi^*$ are unity, (69) would in turn follow from

$$\mu(\Phi^*) < \frac{1}{np - 1} \tag{70}$$

by the Gershgorin circle theorem.

The pairwise coherence has the form

$$\frac{np}{N} \sum_{j=1}^N \phi_{ij} \phi_{i'j}^* = \frac{1}{N} e^{t\omega(\xi_l^2 + \eta_l^2 + \xi_{l'}^2 - \xi_{l'}^2 - \eta_{l'}^2)/(2z_0)} \sum_{j=1}^N e^{t\omega x_j(\xi_{l'} - \xi_l)/z_0} e^{t\omega y_j(\eta_{l'} - \eta_l)/z_0} e^{t(\theta_{kj} - \theta_{k'j})}.$$

There are two cases: (i) $k \neq k'$, (ii) $k = k', l \neq l'$.

For case (i), $\theta_{kj} - \theta_{k'j}$ are independent random variables for $j = 1, \dots, N$. Applying Hoeffding's inequality to

$$Z_N \equiv \sum_{j=1}^N e^{t\omega x_j(\xi_{l'} - \xi_l)/z_0} e^{t\omega y_j(\eta_{l'} - \eta_l)/z_0} e^{t(\theta_{kj} - \theta_{k'j})}$$

we obtain

$$\mathbb{P} \left[\frac{1}{N} |Z_N| \geq t \right] \leq 4 e^{-Nt^2}. \tag{71}$$

Setting $t = \alpha/\sqrt{N}$, we have

$$\mathbb{P} \left[\left| \frac{np}{N} \sum_{j=1}^N \phi_{ij} \phi_{i'j}^* \right| \geq \frac{\alpha}{\sqrt{N}} \right] \leq 4 e^{-\alpha^2}$$

and thus

$$\mathbb{P} \left[\sup_{\substack{k \neq k' \\ \forall l, l'}} \left| \frac{np}{N} \sum_{j=1}^N \phi_{ij} \phi_{i'j}^* \right| \geq \frac{\alpha}{\sqrt{N}} \right] \leq 2n^2 p(p - 1) e^{-\alpha^2} \tag{72}$$

by the union bound.

For case (ii), $\theta_{kj} - \theta_{k'j} = 0$ and Z_N becomes a geometric series

$$Z_N = \frac{e^{t\omega(\xi_{l'} - \xi_l)(x_1 + \sqrt{N}\ell)/z_0} - e^{t\omega(\xi_{l'} - \xi_l)x_1/z_0}}{1 - e^{t\omega(\xi_{l'} - \xi_l)\ell/z_0}} \times \frac{e^{t\omega(\eta_{l'} - \eta_l)(y_1 + \sqrt{N}\ell)/z_0} - e^{t\omega(\eta_{l'} - \eta_l)y_1/z_0}}{1 - e^{t\omega(\eta_{l'} - \eta_l)\ell/z_0}}.$$

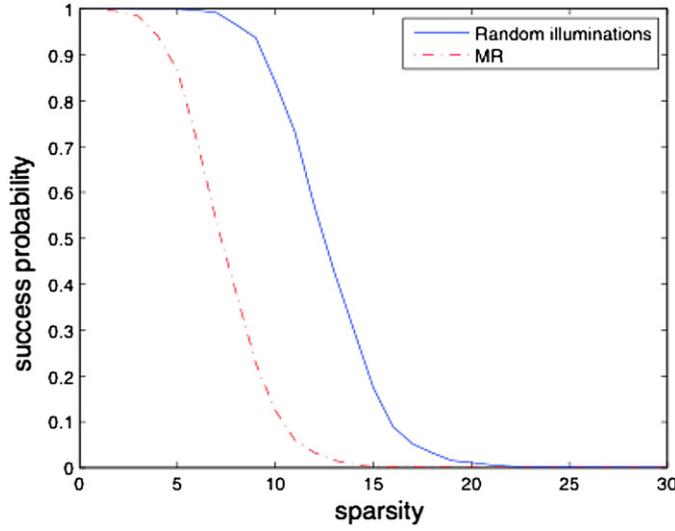


Figure 3. The Subspace Pursuit performance comparison between RI with $n = 11, p = 6$ and MR with $n = 11$. The vertical axis is for the success probability and the horizontal axis is for the number of objects. The success probability is estimated from 1000 independent trials.

Thus,

$$\frac{np}{N} \left| \sum_{j=1}^N \phi_{ij} \phi_{i'j}^* \right| \leq \frac{1}{N} \left| \frac{\sin \frac{\omega \ell \sqrt{N} (\xi_{i'} - \xi_i)}{2z_0}}{\sin \frac{\omega \ell (\xi_{i'} - \xi_i)}{2z_0}} \right| \cdot \left| \frac{\sin \frac{\omega \ell \sqrt{N} (\eta_{i'} - \eta_i)}{2z_0}}{\sin \frac{\omega \ell (\eta_{i'} - \eta_i)}{2z_0}} \right|.$$

Let

$$\kappa = \min_{l \neq l'} \min_{j \in \mathbb{Z}} \left\{ \left| \frac{\ell(\xi_{l'} - \xi_l)}{\lambda z_0} - j \right|, \left| \frac{\ell(\eta_{l'} - \eta_l)}{\lambda z_0} - j \right| \right\}.$$

Clearly κ is nonzero with probability 1. For $l \neq l'$, the probability density functions (PDF) for the random variables

$$\frac{\ell(\xi_{l'} - \xi_l)}{\lambda z_0}, \quad \frac{\ell(\eta_{l'} - \eta_l)}{\lambda z_0}$$

are either the symmetric triangular distribution or its self-convolution supported on $[-2\rho^{-1}, 2\rho^{-1}]$. In either case, their PDFs are bounded by ρ . Hence, the probability that $\{\kappa > \beta\}$ for small $\beta > 0$ is larger than

$$(1 - 2\rho\beta)^{n(n-1)/2} > 1 - \beta\rho n(n-1)$$

where the exponent counts the number of distinct unordered pairs (l, l') . Note that the above analysis is independent of $k = k'$. Since $\sin \theta \geq \theta, \forall \theta \in [0, \pi/2]$ we have that

$$\mathbb{P} \left[\sup_{\substack{k=k' \\ l \neq l'}} \frac{np}{N} \left| \sum_{j=1}^N \phi_{ij} \phi_{i'j}^* \right| \geq \frac{\pi^2}{4N\beta^2} \right] \leq \beta\rho n(n-1). \tag{73}$$

Setting

$$\max \left\{ \frac{\alpha}{\sqrt{N}} \right\} < \frac{\pi^2}{4N\beta^2} = \frac{1}{np-1} \tag{74}$$

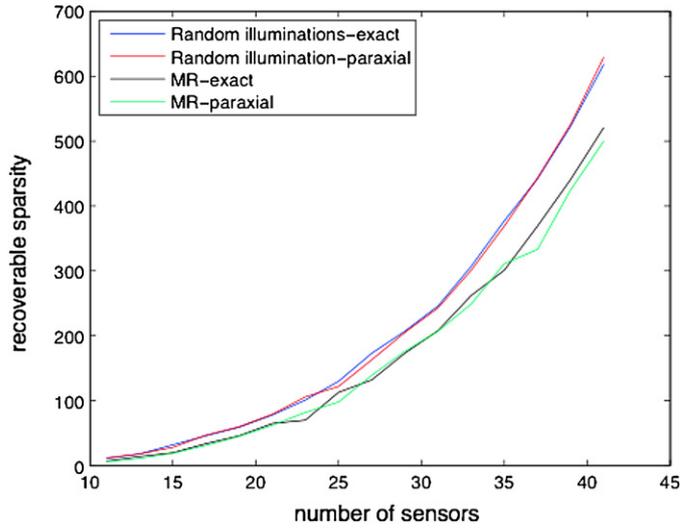


Figure 4. The numbers of recoverable (by Subspace Pursuit) objects for RI with $p = (n+1)/2$ and MR as n varies. The curves indicate a quadratic behavior predicted by the theory. The difference between recoveries with the exact and paraxial Green functions is negligible in both the RI and MR setups.

and using (72) and (73) we have

$$\mathbb{P} \left[\sup_{i \neq i'} \frac{np}{N} \left| \sum_{j=1}^N \phi_{ij} \phi_{i'j}^* \right| \geq \frac{1}{np-1} \right] \leq \beta \rho n(n-1) + 2n^2 p(p-1) e^{-\alpha^2}. \quad (75)$$

As a consequence,

$$\mathbb{P} \left[\sup_{i \neq i'} \frac{np}{N} \left| \sum_{j=1}^N \phi_{ij} \phi_{i'j}^* \right| \geq \frac{1}{np-1} \right] < \rho n(n-1) \frac{\pi}{2} \sqrt{\frac{np-1}{N}} + 2n^2 p(p-1) e^{-\frac{N}{(np-1)^2}}$$

by maximizing the right-hand side of (75) under constraint (74). \square

8. Numerical simulations

We use two numerical settings: the diffraction-limited case when (28) is satisfied (figures 3–6) and the under-resolved case when the ratio in (28) is smaller than unity (figure 7).

For the diffraction-limited case we set $z_0 = 10000$ and $\lambda = 0.1$ for the search domain $[-250, 250]^2$ with $\ell = 10$. The targets are i.i.d. uniform random points in the grid with amplitudes in the range $[1, 2]$. We randomly select sensor locations from $[-50, 50]^2$ with the aperture $A = 100$ satisfying (28). With these parameters

$$\frac{(A + \ell\sqrt{N})^4}{\lambda z_0^3} \approx 1.3$$

condition (23) is barely satisfied. Although the Lasso or Basis Pursuit solvers generally have better performance, for the sake of the speed of computing figures 3–7 we have used the Matlab code *Subspace Pursuit* (SP) (available at <http://igorcarron.googlepages.com/cscodes>).

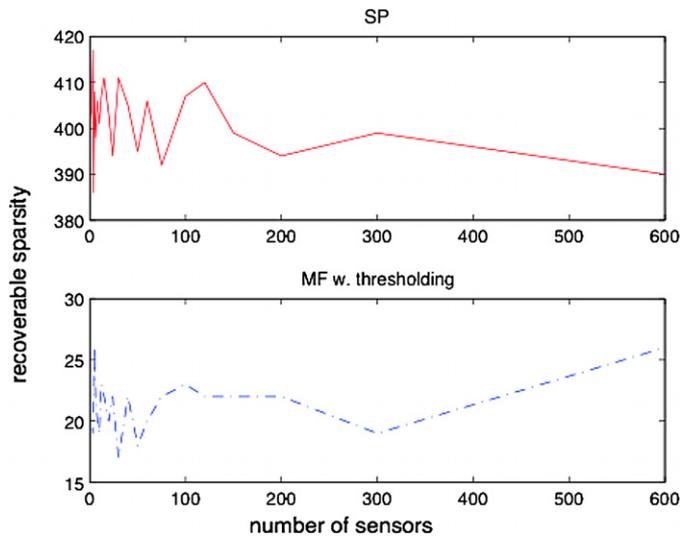


Figure 5. The number of recoverable objects as a function of the number of sensors $n = 1, 2, 3, 4, 5, 6, 8, 10, 12, 15, 20, 24, 25, 30, 40, 50, 60, 75, 100, 120, 150, 200, 300, 600$ with $np = 600$ fixed. The top panel is for the Subspace Pursuit and the bottom panel for OST. The left ends of both curves indicate superresolution.

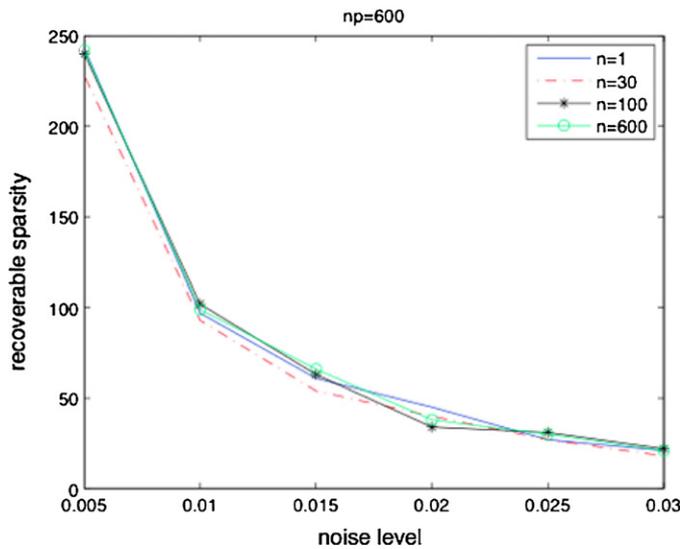


Figure 6. Noisy recovery by Subspace Pursuit for $n = 1, 30, 100, 600$ and with $np = 600$ fixed. The noise is given by the circularly random Gaussian noise of magnitude $\sigma \|Y\|_2$ where σ is the horizontal coordinate. Note that in this case $\mathbb{E}\|E\|_2^2 = np\sigma^2\|Y\|_2^2$.

We use the true Green function (20) in the computation of scattered waves and in the recovery of the exact Green function as well as its paraxial approximation to construct the sensing matrix (for comparison). In other words, we allow model mismatch between the forward and inversion steps.

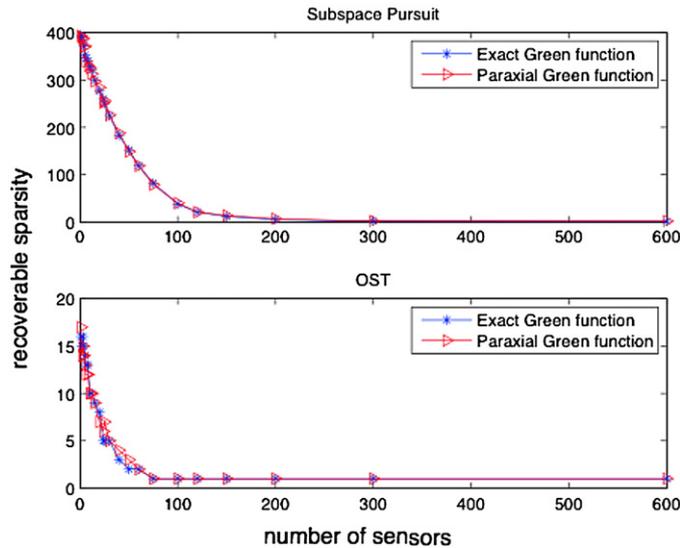


Figure 7. The number of recoverable objects in the under-resolved case as a function of the number of sensors $n = 1, 2, 3, 4, 5, 6, 8, 10, 12, 15, 20, 24, 25, 30, 40, 50, 60, 75, 100, 120, 150, 200, 300, 600$ with $np = 600$ fixed.

In the first set of simulations, we compare the performances of SP for two imaging setups: one with RI and the other with MR. As figure 3 shows, the RI setup has a higher success probability than the MR setup. Another comparison is shown in figure 4 in terms of the number of recoverable objects over a range of n . The quadratic behavior is consistent with the prediction of (39) and (36). The difference between the exact and paraxial Green function recoveries is negligible in both the RI and MR setups. For a given n , SP with the RI setup recovers a higher number of objects than SP with the MR setup does.

Figure 5 compares the performances of SP (top panel) and OST (bottom panel) in terms of the number of recoverable objects for a fixed $np = 600$ but variable n . Clearly, SP can recover far more objects exactly than the OST does. For a fixed np the performance for each method appears relatively constant over the whole range of n . For small n , the performance curves of both methods indicate superresolution. As noise level increases SP performance decays (figure 6).

To further understand the superresolution effect of RI, we consider the setup with $z_0 = 25000$, $\lambda = 0.4$ for which the ratio in (28) is 0.1. This is an under-resolved case whose performance is shown in figure 7. In contrast to the diffraction-limited case (figure 5), the number of recoverable objects in the under-resolved case decays rapidly as p decreases (n increases). To maintain high performance in the under-resolved case, it is necessary that $p \gg 1$. The number of recoverable objects is calculated based on 90% recovery of 100 independent trials.

We demonstrate in figures 8–11 the performance for extended objects in the presence of external noise of the form

$$\frac{\alpha}{\sqrt{2}}(v_1 + \iota v_2) \frac{\|Y\|_2}{\sqrt{np}}, \quad \alpha = 5\%, 20\%$$

where α is the percentage of noise in each entry of the data vector and v_1, v_2 are i.i.d. uniform random variables in $[0, 1]$.

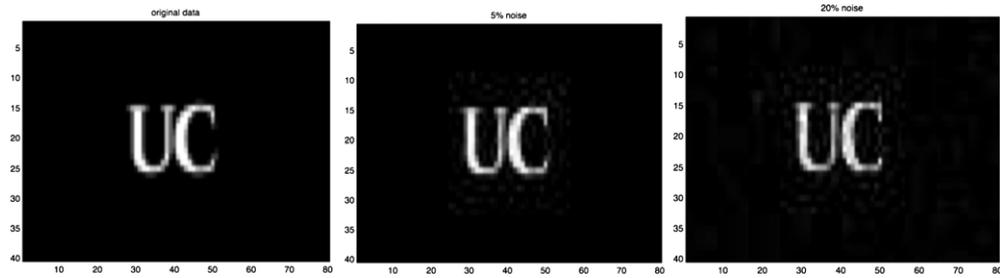


Figure 8. The original 40×80 pixel image (left) and the BPDN reconstructions (middle panel, 5% noise; right panel 20% noise) with one sensor and 500 RI.

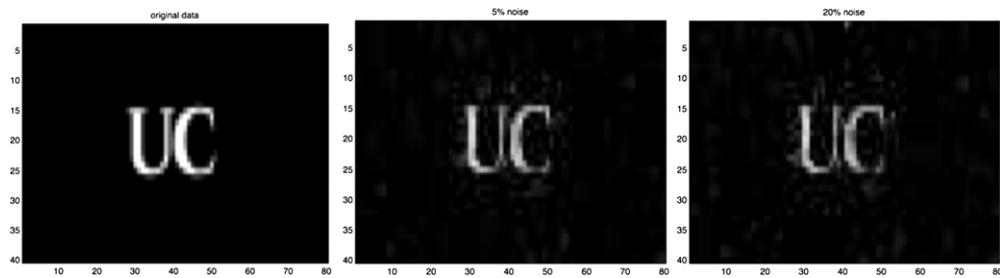


Figure 9. The original 40×80 pixel image (left) and the BPDN reconstructions (middle panel, 5% noise; right panel, 20% noise) with one illumination and 500 randomly distributed sensors.

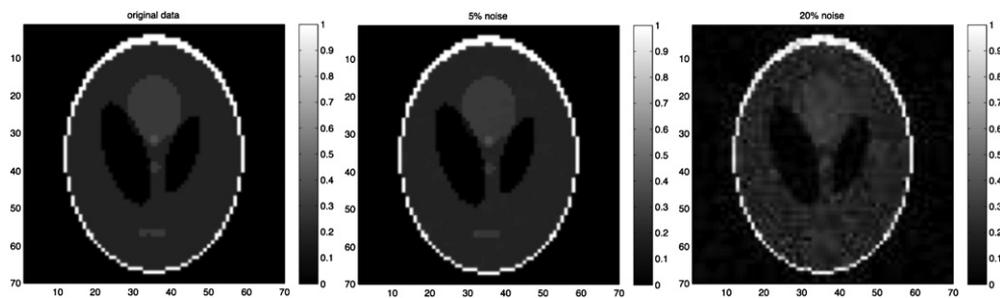


Figure 10. The original 70×70 pixel image (left), the Shepp–Logan phantom and the TV-minimization reconstructions (middle panel, 5% noise; right panel, 20% noise) with one sensor and 1000 RI.

Figure 8 shows the original 40×80 pixel image (left) and its reconstructions (middle panel, 5% noise; right panel, 20% noise) by the BPDN solver YALL1 (<http://yall1.blogs.rice.edu/>) using one sensor and 500 RI, while figure 9 shows the results with one illumination and 500 randomly distributed sensors.

Figure 10 shows the original 70×70 pixel image (left), the Shepp–Logan phantom and its reconstructions (middle panel, 5% noise; right panel, 20% noise) by the total-variation minimization [14, 31] solver TVAL3 (<http://www.caam.rice.edu/optimization/L1/TVAL3/>)

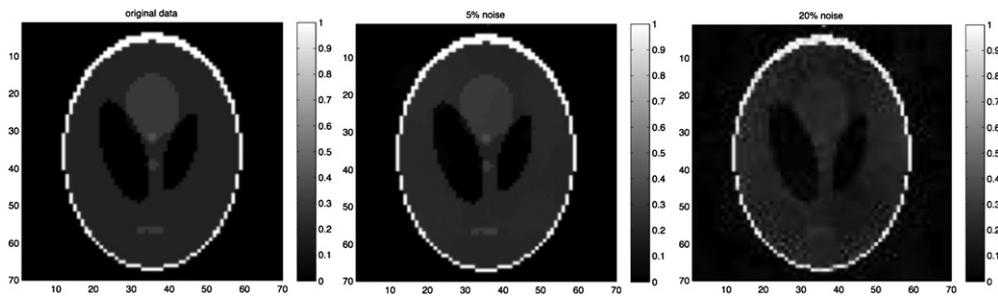


Figure 11. The original 70×70 pixel image (left), the Shepp–Logan phantom and the TV-minimization reconstructions (middle panel, 5% noise; right panel, 20% noise) with one illumination and 1000 randomly distributed sensors.

using one sensor and 1000 RI, while figure 11 shows the results with one illumination and 1000 randomly distributed sensors.

For the one-illumination reconstructions (figures 9 and 11), the classical resolution criterion (28) is met. Note, however, that the Shepp–Logan phantom is not in the class of sparse extended objects analyzed in section 4 because the object support covers more than 50% of the domain (only the gradient is sparse). As a result, the same percentage of noise represents a greater amount of noise in the case of Shepp–Logan phantom and has a more serious effect on performance (figures 10 and 11, right panels).

9. Conclusion

We have proposed a new approach to superresolving point and extended objects based on random illumination and compressed sensing reconstruction.

We have proved that in the diffraction-limited case both the Lasso and the OST with random illumination can exactly localize $s = \mathcal{O}(m)$ objects where the number of data m is the product of the numbers of random probes and sensors. For the under-resolved case where the Rayleigh resolution limit is broken, the Lasso still has a similar performance guarantee if the number of random illuminations is sufficiently large. It is possible to extend the OST result to the under-resolved case which is omitted here to simplify the presentation.

Numerical evidence supports our theoretical prediction and confirms the superiority of the Lasso to the OST in the setup with random illumination.

We have also shown that the BPDN is suitable for imaging extended objects and has provided numerical examples to demonstrate its performance.

The superresolution effect with random illumination revealed here contrasts with the subwavelength resolution with deterministic near-field illumination studied in [20].

Finally, we note that in our approach it is essential to measure the wave field. For *intensity-only* measurements, additional techniques such as interferometry or phase retrieval methods are necessary for object reconstruction.

Acknowledgments

I am grateful to Mike Yan for producing figures 3–7 and Hsiao-Chieh Tseng for producing figures 8–11 of section 8. The research is partially supported by the NSF grant DMS-0908535.

References

- [1] Almero P F, Pedrini G, Gundu P N, Osten W and Hanson S G 2011 Enhanced wavefront reconstruction by random phase modulation with a phase diffuser *Opt. Laser Eng.* **49** 252–7
- [2] Baggeroer A B, Kuperman W A and Mikhalevsky P N 1993 An overview of matched field methods in ocean acoustics *IEEE J. Ocean. Eng.* **18** 401–24
- [3] Bajwa W, Calderbank R and Jafarpour S 2010 Why Gabor Frames? Two fundamental measures of coherence and their role in model selection *J. Commun. Net.* **12** 289–307
- [4] Baraniuk R, Davenport M, DeVore R and Wakin M 2008 A simple proof of the restricted isometry property for random matrices *Constructive Approx.* **28** 253–63
- [5] Barron A, Birgé L and Massart P 1999 Risk bounds for model selection via penalization *Probab. Theory Relat. Fields* **113** 301–413
- [6] Birgé L and Massart P 2001 Gaussian model selection *J. Eur. Math. Soc.* **3** 203–68
- [7] Brady D J, Choi K, Marks D L, Horisaki R and Lim S 2009 Compressive holography *Opt. Express* **17** 13040–9
- [8] Bräuer R, Wojak U, Wyrowski F and Bryngdahl O 1991 Digital diffusers for optical holography *Opt. Lett.* **16** 1427–9
- [9] Bruckstein A M, Donoho D L and Elad M 2009 From sparse solutions of systems of equations to sparse modeling of signals *SIAM Rev.* **51** 34–81
- [10] Bunea F, Tsybakov A B and Wegkamp M H 2007 Sparsity oracle inequalities for the Lasso *Electron. J. Stat.* **1** 169–94
- [11] Candès E J 2008 The restricted isometry property and its implications for compressed sensing *C. R. Acad. Sci., Paris I* **346** 589–92
- [12] Candès E J and Plan Y 2009 Near-ideal model selection by ℓ_1 minimization *Ann. Stat.* **37** 2145–77
- [13] Candès E J and Tao T 2005 Decoding by linear programming *IEEE Trans. Inf. Theory* **51** 4203–15
- [14] Chambolle A and Lions P-L 1997 Image recovery via total variation minimization and related problems *Numer. Math.* **76** 167–88
- [15] Chen S S, Donoho D L and Saunders M A 1998 Atomic decomposition by basis pursuit *SIAM J. Sci. Comput.* **20** 33–61
- [16] Dai W and Milenkovic O 2009 Subspace pursuit for compressive sensing and reconstruction *IEEE Trans. Inf. Theory* **55** 2230–49
- [17] Delsarte P, Goethals J M and Seidel J J 1975 Bounds for systems of lines and Jacobi polynomials *Phil. Res. Rep.* **30** 91–105 issue in honour of C J Bouwkamp
- [18] Donoho D L, Elad M and Temlyakov V N 2006 Stable recovery of sparse overcomplete representations in the presence of noise *IEEE Trans. Inf. Theory* **52** 6–18
- [19] Duarte M, Davenport M, Takhar D, Laska J, Sun T, Kelly K and Baraniuk R 2008 Single-pixel imaging via compressive sampling *IEEE Signal Process. Mag.* **25** 83–91
- [20] Fannjiang A C 2009 Compressive imaging of subwavelength structures *SIAM J. Imag. Sci.* **2** 1277–91
- [21] Fannjiang A C 2010 Compressive inverse scattering: I. High-frequency SIMO/MISO and MIMO measurements *Inverse Problems* **26** 035008
- [22] Fannjiang A C 2010 The MUSIC algorithm for sparse objects: a compressed sensing analysis arXiv:1006.1678
- [23] Fannjiang A and Solna K 2007 Broadband resolution analysis for imaging with measurement noise *J. Opt. Soc. Am. A* **24** 1623–32
- [24] Fannjiang A, Strohmer T and Yan P 2010 Compressed remote sensing of sparse objects *SIAM J. Imaging Sci.* **3** 596–618
- [25] Feller W 1970 *An Introduction to Probability Theory and its Applications* vol II, 2nd edn (New York: Wiley)
- [26] Greenshtein E 2006 Best subset selection, persistence in high-dimensional statistical learning and optimization under ℓ_1 -constraint *Ann. Stat.* **34** 2367–86
- [27] Hoeffding W 1963 Probability inequalities for sums of bounded random variables *J. Am. Stat. Assoc.* **58** 13–30
- [28] Meinshausen N and Bühlmann P 2006 High-dimensional graphs and variable selection with the Lasso *Ann. Stat.* **34** 1436–62
- [29] Rauhut H 2008 Stability results for random sampling of sparse trigonometric polynomials *IEEE Trans. Inf. Theory* **54** 5661–70
- [30] Romberg J 2009 Compressive sensing by random convolution *SIAM J. Imaging Sci.* **2** 1098–128
- [31] Rudin L I, Osher S and Fatemi E 1992 Nonlinear total variation based noise removal algorithms *Physica D* **60** 259–68
- [32] Senatov V V 1998 *Normal Approximation: New Results, Methods, and Problems* (Utrecht: Walter de Gruyter)
- [33] Shahram M and Milanfar P 2004 Imaging below the diffraction limit: a statistical analysis *IEEE Trans. Image Process.* **13** 677–89

- [34] Shirai T and Wolf E 2004 Coherence and polarization of electromagnetic beams modulated by random phase screens and their changes on propagation in free space *J. Opt. Soc. A* **21** 1907–16
- [35] Tibshirani R 1996 Regression shrinkage and selection via the lasso *J. R. Stat. Soc. B* **58** 267–88
- [36] Tolstoy A 1993 *Matched Field Processing in Underwater Acoustics* (Singapore: World Scientific)
- [37] Tropp J A 2004 Greed is good: algorithmic results for sparse approximation *IEEE Trans. Inf. Theory* **50** 2231–42
- [38] Tropp J A 2006 Just relax: convex programming methods for identifying sparse signals in noise *IEEE Trans. Inf. Theory* **52** 1030–51
- Tropp J A 2009 *IEEE Trans. Inf. Theory* **55** 917–8 (corrigendum)
- [39] Tropp J A 2008 On the conditioning of random subdictionaries *Appl. Comput. Harmon. Anal.* **25** 1–24
- [40] Welch L 1974 Lower bounds on the maximum cross-correlation of signals *IEEE Trans. Inf. Theory* **20** 397–9
- [41] Zhao P and Yu B 2006 On model selection consistency of Lasso *J. Mach. Learn. Res.* **7** 2541–63