

# RESOLUTION ANALYSIS OF IMAGING WITH $\ell_1$ OPTIMIZATION

LILIANA BORCEA AND ILKER KOCYIGIT \*

**Abstract.** We study array imaging of a sparse scene of point-like sources or scatterers in a homogeneous medium. For source imaging the sensors in the array are receivers that collect measurements of the wave field. For imaging scatterers the array probes the medium with waves and records the echoes. In either case the image formation is stated as a sparsity promoting  $\ell_1$  optimization problem, and the goal of the paper is to quantify the resolution. We consider both narrow-band and broadband imaging, and a geometric setup with a small array. We take first the case of the unknowns lying on the imaging grid, and derive resolution limits that depend on the sparsity of the scene. Then we consider the general case with the unknowns at arbitrary locations. The analysis is based on estimates of the cumulative mutual coherence and a related concept, which we call interaction coefficient. It complements recent results in compressed sensing by deriving deterministic resolution limits that account for worse case scenarios in terms of locations of the unknowns in the imaging region, and also by interpreting the results in some cases where uniqueness of the solution does not hold. We demonstrate the theoretical predictions with numerical simulations.

**Key words.** array imaging, sparse,  $\ell_1$  optimization, cumulative mutual coherence.

**1. Introduction.** Array imaging is an inverse problem for the wave equation, where the goal is to determine remote sources or scatterers from measurements of the wave field at a collection of nearby sensors, called the array. The problem has applications in medical imaging, nondestructive evaluation of materials, oil prospecting, seismic imaging, radar imaging, ocean acoustics and so on. There is extensive literature on various imaging approaches such as reverse time migration and its high frequency version called Kirchhoff migration [2, 3, 14], matched field imaging [1], Multiple Signal Classification (MUSIC) [30, 23], the linear sampling method [9], and the factorization method [25]. In this paper we consider array imaging using  $\ell_1$  optimization, which is appropriate for sparse scenes of unknown sources or scatterers that have small support in the imaging region.

Imaging with sparsity promoting optimization has received much attention recently, specially in the context of compressed sensing [21, 19, 29], where a random set of sensors collect data from a sparse scene. Such studies use the restricted isometry property of the sensing matrix [10] or its mutual coherence [8] to derive probability bounds on the event that the imaging scene is recovered exactly for noiseless data, or with small error that scales linearly with the noise level. The array does not play an essential role in these studies, aside from its aperture bounding the random sample of locations of the sensors, and for justifying the scaling that leads to models of wave propagation like the paraxial one [21].

A different approach proposed in [11, 12] images a sparse scattering scene using illuminations derived from the singular value decomposition (SVD) of the response matrix measured by probing sequentially the medium with pulses emitted by one sensor at a time and recording the echoes. Illuminations derived from the SVD are known to be useful in imaging [27, 6, 4, 5, 23] and they may mitigate noise. The setup in [11], which is typical in array imaging, lets the sensors be closely spaced so that sums over them can be approximated by integrals over the array aperture. We consider the same continuous aperture setup here and study the resolution of the images produced by  $\ell_1$  optimization, also known as basis pursuit and  $\ell_1$ -penalty. We address two questions: (1) How should we chose the discretization of the imaging region so that we can guarantee unique recovery of the sparse scene, at least when the unknowns lie on the grid? (2) If the imaging region is discretized on a finer grid, for which uniqueness does not hold, are there cases where the solution of the  $\ell_1$  optimization is still useful?

By studying question (1) we complement the existing results with deterministic resolution limits that account for worse case scenarios, and guarantee unique recovery of the scene for a given sparsity  $s$ . This is defined as the number of non-zero entries of the vector of unknowns or, equivalently, the number of grid points in the support of the sources/scatterers that we image. We consider a geometric setup with a small array, where wave propagation can be modeled by the paraxial approximation. We have a more general paraxial model than in [21], which takes into consideration sources/scatterers at different ranges from the array. This turns out to be important in narrow-band regimes. We also consider broad-band regimes and

---

\*Department of Mathematics, University of Michigan, Ann Arbor, MI 48109-1043.  
Email: borcea@umich.edu & ilkerk@umich.edu

show that the additional multi-frequency data improves the resolution.

It is typical in imaging with sparsity promoting optimization to assume that the unknown sources or scatterers lie on the discretization grid, meaning that they can be modeled by a sparse complex vector  $\boldsymbol{\rho} \in \mathbb{C}^N$ , where  $N$  is the number of grid points. If the unknowns lie off-grid the results deteriorate. We refer to [24, 18] for a perturbation analysis of compressed sensing with small off-grid displacements. General tight error bounds can be found in [13]. They may be quite large and increase with  $N$ . Thus, there is a trade-off in imaging with  $\ell_1$  optimization: on one hand we need a coarse enough discretization of the imaging region to ensure unique recovery of the solution, and on the other hand finer discretization to minimize modeling errors due to off-grid placement of the unknowns. This trade-off is particularly relevant in the narrow-band paraxial regime, where the resolution limits may grow significantly with the sparsity  $s$  of the scene.

At question (2) we consider fine discretizations of the imaging region, to mitigate the modeling error. The problem is then how to interpret the result  $\boldsymbol{\rho}_*$  of the  $\ell_1$  minimization, which is no longer guaranteed to be unique. We show that there are cases where the minimization may be useful. Specifically, we prove that when the unknown sources/scatterers are located at points or clusters of points that are sufficiently well separated, an  $\ell_1$  minimizer  $\boldsymbol{\rho}_*$  is supported in the vicinity of these points. While the entries of  $\boldsymbol{\rho}_*$  may not be close in the point-wise sense to those of  $\boldsymbol{\rho}$ , their average over such vicinities are close to the true values in  $\boldsymbol{\rho}$  in the case of well separated points, or the averages of the true values in the case of clusters of points. That is to say,  $\ell_1$  optimization gives an effective vector of source/scatterer amplitudes averaged locally around the points in its support. Note that question (2) was also investigated in [20], where novel algorithms for imaging well separated sources have been introduced and analyzed. Our study complements the results in [20] by analyzing directly the performance of the  $\ell_1$  minimization and  $\ell_1$ -penalty, and also considering clusters of sources/scatterers.

The paper is organized as follows. In section 2 we formulate the problem, introduce notation, and describe the relation between imaging sources vs. scatterers. Question (1) is studied in section 3. We describe the paraxial scaling regime and derive resolution bounds that depend on the sparsity  $s$  of the imaging scene. In section 4 we study question (2). In both sections we begin with the statement of results and numerical illustrations, and end with the proofs. A summary is in section 5.

**2. Formulation of the imaging problem.** We formulate first the basic problem of imaging  $s$  point-like sources with a remote array of sensors that record the incoming sound waves. The generalization to the inverse scattering problem is described in section 2.2.

Suppose that there are  $s$  unknown sources located at points  $\vec{\mathbf{y}}_j$  in the imaging region  $W \subset \mathbb{R}^3$ , emitting signals  $\hat{f}_j(\omega)$  at frequency  $\omega$ , for  $j = 1, \dots, s$ . The hat stands for Fourier transform with respect to time, and reminds us that we work in the frequency domain. The receivers are at locations  $\vec{\mathbf{x}}_r \in \mathcal{A}$ , for  $r = 1, \dots, M_r$ , where  $\mathcal{A}$  is a set on the measuring surface, called the array aperture. The sound pressure wave measured at  $\vec{\mathbf{x}}_r$  and frequency  $\omega$  is modeled by

$$\hat{p}(\omega, \vec{\mathbf{x}}_r) = \sum_{j=1}^s \hat{f}_j(\omega) \hat{G}(\omega, \vec{\mathbf{x}}_r, \vec{\mathbf{y}}_j), \quad (2.1)$$

where  $\hat{G}$  is the outgoing Green's function of Helmholtz's equation. The propagation is through a homogeneous medium with sound speed  $c$ , and the Green's function is

$$\hat{G}(\omega, \vec{\mathbf{x}}_r, \vec{\mathbf{y}}_j) = \frac{e^{ik|\vec{\mathbf{x}}_r - \vec{\mathbf{y}}_j|}}{4\pi|\vec{\mathbf{x}}_r - \vec{\mathbf{y}}_j|}, \quad (2.2)$$

where  $k = \omega/c$  is the wavenumber. The inverse source problem is to determine  $\{\hat{f}_j\}_{j=1, \dots, s}$  from the measurements (2.1) at one or more frequencies  $\omega$ .

**2.1. Imaging sources with  $\ell_1$  optimization.** To state the inverse problem as an  $\ell_1$  optimization we discretize  $W$  with a regular grid of rectangular prisms, and let  $W_N \subset \mathbb{R}^3$  be the set of  $N$  grid points denoted by  $\vec{\mathbf{z}}_j$ . The lengths of the edges of the rectangular prisms are the components of the vector  $\vec{\mathbf{h}} \in \mathbb{R}^3$ , called the mesh size. The sources may be on or off the grid. If they are on the grid, as assumed in section 3, we denote by  $\mathcal{S}$  the set of indexes of the grid points that support them. Explicitly, we define the bijective map

$J : \{1, \dots, s\} \rightarrow \mathcal{S} \subset \{1, \dots, N\}$ , such that  $\bar{\mathbf{z}}_{J(j)} = \bar{\mathbf{y}}_j$ , for  $j = 1, \dots, s$ , and  $\mathcal{S} = \{J(1), \dots, J(s)\}$ . When the sources are not on the grid, we let  $\mathcal{S}$  index the nearest points  $\bar{\mathbf{z}}_j$  to each source, as explained in more detail in section 4. We assume henceforth that  $N \gg s$ , meaning that the imaging scene is sparse.

Let  $\mathbf{d} \in \mathbb{C}^M$  be the data vector, with components  $\hat{p}(\omega_l, \bar{\mathbf{x}}_r)$ , for  $l = 1, \dots, M_\omega$  and  $r = 1, \dots, M_r$ . The number of measurements is  $M = M_\omega M_r$ . Let also  $\mathcal{G} \in \mathbb{C}^{M \times N}$  be the sensing matrix, with entries defined by  $\hat{G}(\omega_l, \bar{\mathbf{x}}_r, \bar{\mathbf{z}}_j) / \alpha_j$ , where  $\alpha_j$  normalizes the columns of  $\mathcal{G}$ , denoted by  $\mathbf{g}_j$ . Absorbing the normalization constants in the vector  $\boldsymbol{\rho}$  of unknowns, we obtain the linear system

$$\mathcal{G}\boldsymbol{\rho} = \mathbf{d}. \quad (2.3)$$

For single frequency measurements at  $\omega = \omega_1$  the normalization constant is

$$\alpha_j = \left( \sum_{r=1}^{M_r} |\hat{G}(\omega, \bar{\mathbf{x}}_r, \bar{\mathbf{z}}_j)|^2 \right)^{1/2} = \left( \sum_{r=1}^{M_r} \frac{1}{16\pi^2 |\bar{\mathbf{x}}_r - \bar{\mathbf{z}}_j|^2} \right)^{1/2},$$

so that when the sources lie on the grid there are  $s$  non-zero entries in  $\boldsymbol{\rho}$ , equal to

$$\rho_j = \alpha_{J^{-1}(j)} \hat{f}_{J^{-1}(j)}(\omega), \quad j \in \mathcal{S}. \quad (2.4)$$

Here  $J^{-1} : \mathcal{S} \rightarrow \{1, \dots, s\}$  is the inverse of the mapping  $J$ . For multiple frequency measurements we simplify the problem by letting

$$\hat{f}_j(\omega) = \hat{f}(\omega) R_j, \quad \forall j = 1, \dots, s, \quad (2.5)$$

so that all the sources emit the same known signal  $\hat{f}(\omega)$  multiplied by an unknown complex amplitude  $R_j$ . This simplification is motivated by the inverse scattering problem described in section 2.2. It keeps the same number  $N$  of unknowns as in the single frequency case, although we have more measurements. The normalization constants are

$$\alpha_j = \|\hat{f}\|_2 \left( \sum_{r=1}^{M_r} \frac{1}{16\pi^2 |\bar{\mathbf{x}}_r - \bar{\mathbf{y}}_j|^2} \right)^{1/2}, \quad \|\hat{f}\|_2 = \left( \sum_{l=1}^{M_\omega} |\hat{f}(\omega_l)|^2 \right)^{1/2},$$

and the non-zero entries of  $\boldsymbol{\rho}$  equal

$$\rho_j = \alpha_{J^{-1}(j)} R_{J^{-1}(j)}, \quad j \in \mathcal{S}. \quad (2.6)$$

Note that we could have written the multiple frequency problem for  $M_\omega N$  unknowns, the Fourier coefficients  $\hat{f}_j(\omega_l)$  of the signals emitted by the sources. However, at each frequency these have the same spatial support, so another optimization approach, known as Multiple Measurement Vector (MMV) [15] would be more appropriate. For the purpose of this paper it suffices to consider the simpler model (2.6).

The  $\ell_1$  optimization (basis pursuit) formulation of the inverse source problem is

$$\min_{\boldsymbol{\rho} \in \mathbb{C}^N} \|\boldsymbol{\rho}\|_1 \quad \text{such that} \quad \mathcal{G}\boldsymbol{\rho} = \mathbf{d}, \quad (2.7)$$

where  $\|\boldsymbol{\rho}\|_1 = \sum_{j=1}^N |\rho_j|$ . Our goal in section 3 is to determine bounds on the mesh size  $\vec{\mathbf{h}}$  so that (2.7) has a unique  $s$  sparse solution, equal to the true  $\boldsymbol{\rho}$  defined in (2.4) and (2.6). The analysis is based on the next lemma, following from [31, 32, 17].

LEMMA 2.1. *Suppose that (2.3) has an  $s$  sparse solution  $\boldsymbol{\rho}$  and that the cumulative mutual coherence  $\mu(\mathcal{G}, s)$  of matrix  $\mathcal{G}$  with columns  $\mathbf{g}_j$  of Euclidian length equal to one satisfies*

$$\mu(\mathcal{G}, s) = \max_{j=1, \dots, N} \max_{|S|=s} \sum_{q \in S, q \neq j} |\langle \mathbf{g}_q, \mathbf{g}_j \rangle| < \frac{1}{2}, \quad (2.8)$$

where  $\langle \cdot, \cdot \rangle$  denotes the usual inner product in  $\mathbb{C}^N$ , and  $S$  is a set of cardinality  $|S|$ . Then  $\boldsymbol{\rho}$  is the unique  $s$  sparse solution of (2.3) and the unique minimizer of (2.7).

To deal with noise and modeling (discretization) error, we also consider in section 4 the  $\ell_1$ -penalty problem [32]

$$\min_{\boldsymbol{\rho} \in \mathbb{C}^N} \mathcal{L}(\boldsymbol{\rho}), \quad \mathcal{L}(\boldsymbol{\rho}) = \frac{1}{2} \|\mathcal{G}\boldsymbol{\rho} - \mathbf{d}\|_2^2 + \gamma \|\boldsymbol{\rho}\|_1, \quad (2.9)$$

with parameter  $\gamma$  accounting for the trade-off between the approximation error and the sparsity of the unknown vector  $\boldsymbol{\rho}$ .

**2.2. Imaging point scatterers.** The problem of imaging a sparse scene of point scatterers at  $\vec{\mathbf{y}}_j$  for  $j = 1, \dots, s$ , can be written as one of imaging  $s$  sources, as we now explain.

Let the array probe the medium with a signal  $\hat{f}(\omega)$  emitted from the sensor at  $\vec{\mathbf{x}}_e$ . Using the Foldy-Lax model [22, 26] we write the scattered wave at  $\vec{\mathbf{x}}_r$  in the form (2.1), with effective sources at  $\{\vec{\mathbf{y}}_j\}_{j=1, \dots, s}$  emitting signals

$$\hat{f}_j(\omega) = \hat{f}(\omega) R_j \hat{u}_j(\omega). \quad (2.10)$$

Here  $R_j$  is the reflectivity of the  $j$ -th scatterer, and  $\hat{u}_j$  is the wave that illuminates it. It is given by the sum of the incident wave  $\hat{G}(\omega, \vec{\mathbf{x}}_e, \vec{\mathbf{y}}_j)$  and the wave scattered at the other points

$$\hat{u}_j(\omega) = \hat{G}(\omega, \vec{\mathbf{x}}_e, \vec{\mathbf{y}}_j) + \sum_{l=1}^s (1 - \delta_{lj}) R_l \hat{G}(\omega, \vec{\mathbf{y}}_l, \vec{\mathbf{y}}_j) \hat{u}_l(\omega), \quad \forall j = 1, \dots, s, \quad (2.11)$$

where  $\delta_{lj}$  is the Kronecker delta.

In the Born approximation we neglect the sum in (2.11), and simplify (2.10) as

$$\hat{f}_j(\omega) = \hat{f}(\omega) R_j \hat{G}(\omega, \vec{\mathbf{x}}_e, \vec{\mathbf{y}}_j). \quad (2.12)$$

This can be written in the form (2.3) with entries of  $\boldsymbol{\rho}$  like in (2.6), and slightly redefined matrix  $\mathcal{G}$  and normalization constant

$$\alpha_j = \|\hat{f}\|_2 \sqrt{\sum_{r=1}^{M_r} \frac{1}{(4\pi)^4 |\vec{\mathbf{x}}_r - \vec{\mathbf{y}}_j|^2 |\vec{\mathbf{x}}_e - \vec{\mathbf{y}}_j|^2}}.$$

Multiple scattering effects can be included by solving the Foldy-Lax equations (2.11) or, equivalently, the linear system

$$\mathbf{Q}\hat{\mathbf{u}} = \begin{pmatrix} \hat{G}(\omega, \vec{\mathbf{x}}_e, \vec{\mathbf{y}}_1) \\ \vdots \\ \hat{G}(\omega, \vec{\mathbf{x}}_e, \vec{\mathbf{y}}_s) \end{pmatrix},$$

with  $\hat{\mathbf{u}}$  the vector with components  $\hat{u}_j$ , and  $\mathbf{Q} = (Q_{jl})_{j,l=1, \dots, s}$  the matrix with entries

$$Q_{jl} = \delta_{lj} - (1 - \delta_{jl}) \hat{G}(\omega, \vec{\mathbf{y}}_l, \vec{\mathbf{y}}_j) R_l.$$

Note that  $\mathbf{Q}$  is a perturbation of the  $s \times s$  identity matrix, and depending on the magnitude of the reflectivities and the distance between the scatterers, it is invertible. Again we can write the problem in the form (2.3) with entries of  $\boldsymbol{\rho}$  like in (2.6), except that now  $\alpha_j$  are more complicated and depend on the unknown reflectivity. An elegant solution of this nonlinear problem is in [12]. It amounts to solving a source imaging problem like (2.7), to determine the locations  $\vec{\mathbf{y}}_j$ , for  $j = 1, \dots, s$ . Because there are multiple emitters the authors use an MMV approach. Then the reflectivities are estimated using the Foldy-Lax model.

Given the relation between inverse scattering and source problems describe above, we focus attention henceforth on imaging sparse scenes of sources using (2.7) or (2.9).

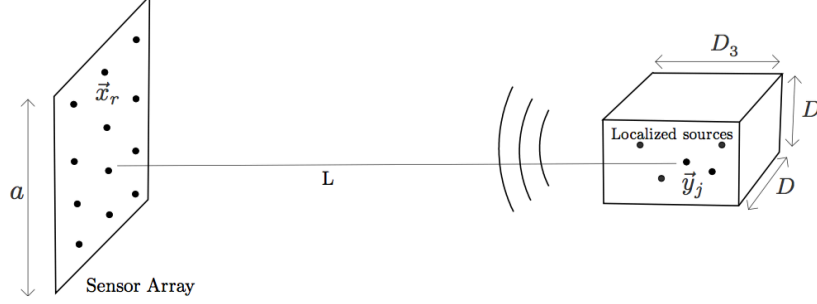


FIG. 3.1. Schematic for the paraxial setup.

**3. Imaging with small arrays.** The setup is illustrated in Figure 3.1. We consider a planar square array of aperture size  $a$ , and a coordinate system with origin at the center of the array and range axis orthogonal to it. The locations of the receivers are  $\vec{\mathbf{x}}_r = (\mathbf{x}_r, 0)$ , with  $\mathbf{x}_r = (x_{1,r}, x_{2,r})$  and  $|x_{1,r}|, |x_{2,r}| \leq a/2$ , for  $r = 1, \dots, M_r$ . The imaging region  $W$  is a rectangular prism with center on the range axis, at distance  $L$  from the array. It has a square side  $[-D/2, D/2] \times [-D/2, D/2]$  in the cross-range plane, parallel to the array, and length  $D_3$  in the range direction. The discretization of  $W$  has grid points  $\vec{\mathbf{z}}_j = (\mathbf{z}_j, z_{3,j})$ , with cross-range vector  $\mathbf{z}_j = (z_{1,j}, z_{2,j})$  and range  $z_{3,j}$ , and the mesh size is  $\vec{\mathbf{h}} = (h, h, h_3)$ . Our goal in this section is to estimate  $\vec{\mathbf{h}}$  so that the  $\ell_1$  optimization problem (2.7) determines exactly the unknown sources supported at  $\vec{\mathbf{z}}_j$  for  $j \in \mathcal{S}$ , a set of cardinality  $s$ .

We begin in section 3.1 with the scaling regime and the paraxial model of wave propagation. The resolution limits are stated and illustrated with numerical simulations in sections 3.2 and 3.3. The setup of the simulations is described in appendix A. The proofs are in section 3.4.

**3.1. Scaling regime and the paraxial model.** The scaling regime is defined by the relation between the important length scales: the wavelength  $\lambda$ , the range scale  $L$ , the aperture  $a$ , and the size  $D$  and  $D_3$  of the imaging region. We assume for now a single frequency  $\omega$ , and refer to section 3.3 for the multi frequency case where another important scale arises, the bandwidth  $B$ .

The scales are ordered as

$$\lambda \ll D \ll a \ll L, \quad D_3 \ll L, \quad (3.1)$$

and satisfy the following assumptions

$$\frac{a^2}{\lambda L} \gtrsim 1, \quad \frac{a^2 D_3}{\lambda L L} \gtrsim 1, \quad (3.2)$$

$$\frac{D^2}{\lambda L} \ll 1, \quad \frac{a^2 a D}{\lambda L L^2} \ll 1, \quad \frac{a^2}{\lambda L} \left(\frac{D_3}{L}\right)^2 \ll 1, \quad \frac{a^2}{\lambda L} \left(\frac{a}{L}\right)^2 \frac{D_3}{L} \ll 1. \quad (3.3)$$

Roughly, conditions (3.3) say that the imaging region is small enough and far enough from the array, so that we can linearize phases of the Green's functions in  $\vec{\mathbf{z}}_j$ , for all  $j = 1, \dots, N$ . Physically, this means that when viewed from the imaging region, the wave fronts appear planar. The array aperture  $a$  is small with respect to the distance  $L$  of propagation of the waves, but equations (3.2) say that the Fresnel number is large, so we have diffraction effects.

We show in appendix B that

$$\sum_{r=1}^{M_r} \hat{G}(\omega, \vec{\mathbf{x}}_r, \vec{\mathbf{z}}_j) \overline{\hat{G}(\omega, \vec{\mathbf{x}}_r, \vec{\mathbf{z}}_q)} \approx \frac{e^{ik(z_{3,j} - z_{3,q})}}{(4\pi L)^2} \sum_{r=1}^{M_r} e^{-ik \left[ \frac{|\mathbf{x}_r|^2 (z_{3,j} - z_{3,q})}{2L^2} + \frac{\mathbf{x}_r \cdot (\mathbf{z}_j - \mathbf{z}_q)}{L} \right]}, \quad (3.4)$$

where the bar denotes complex conjugate.

REMARK 3.1. The terms  $e^{ikz_3}$  in (3.4) are highly oscillatory when  $k$  is large, but can be absorbed in the vector of unknowns. This is convenient because it implies that when two points  $\vec{\mathbf{z}}_j$  and  $\vec{\mathbf{z}}_q$  are close to

each other, the inner product of the  $j$ -th and  $q$ -th columns of the scaled sensing matrix is close to one. Let  $\mathcal{Z}$  be the diagonal matrix with  $j^{\text{th}}$  entry  $e^{ikz_{3,j}}$ , for  $j = 1, \dots, N$ , and rewrite the linear system (2.3) as  $\mathcal{G}\mathcal{Z}(\mathcal{Z}^{-1}\boldsymbol{\rho}) = \mathbf{d}$ . To simplify the presentation we denote henceforth by  $\mathcal{G}$  the new sensing matrix  $\mathcal{G}\mathcal{Z}$  and by  $\boldsymbol{\rho}$  the new unknown vector  $\mathcal{Z}^{-1}\boldsymbol{\rho}$ , so that the scaled system looks the same as (2.3). The Green's function with the large phase removed is still denoted by  $\hat{G}$ , and satisfies

$$\sum_{r=1}^{M_r} \hat{G}(\omega, \vec{\mathbf{x}}_r, \vec{\mathbf{z}}_j) \overline{\hat{G}(\omega, \vec{\mathbf{x}}_r, \vec{\mathbf{z}}_q)} \approx \frac{1}{(4\pi L)^2} \sum_{r=1}^{M_r} e^{-ik \left[ \frac{|\mathbf{x}_r|^2 (z_{3,j} - z_{3,q})}{2L^2} + \frac{\mathbf{x}_r \cdot (\mathbf{z}_j - \mathbf{z}_q)}{L} \right]}. \quad (3.5)$$

Assuming that the receivers are on a square grid of spacing  $h_{\mathcal{A}}$ , satisfying the scaling relations

$$h_{\mathcal{A}} \ll \frac{\lambda L}{D}, \quad h_{\mathcal{A}} \ll \frac{\lambda L^2}{aD_3}, \quad (3.6)$$

we see that the exponential in (3.5) is approximately constant in each grid cell in  $\mathcal{A}$ . This allows us to use the continuous aperture approximation in the analysis, where the sum over  $r$  can be replaced by the integral over  $\mathcal{A} = [-a/2, a/2] \times [-a/2, a/2]$ ,

$$\sum_{r=1}^{M_r} \hat{G}(\omega, \vec{\mathbf{x}}_r, \vec{\mathbf{z}}_j) \overline{\hat{G}(\omega, \vec{\mathbf{x}}_r, \vec{\mathbf{z}}_q)} \approx \frac{1}{(4\pi L h_{\mathcal{A}})^2} \int_{\mathcal{A}} d\mathbf{x} e^{-ik \left[ \frac{|\mathbf{x}|^2 (z_{3,j} - z_{3,q})}{2L^2} + \frac{\mathbf{x} \cdot (\mathbf{z}_j - \mathbf{z}_q)}{L} \right]}. \quad (3.7)$$

With our discretization the number of unknowns is  $N = (D/h)^2 D_3/h_3$  and the problem is underdetermined when the number of measurements  $M = M_r = (a/h_{\mathcal{A}})^2$  satisfies  $M < N$  or, equivalently,

$$h_{\mathcal{A}} > a \frac{h}{D} \sqrt{\frac{h_3}{D_3}}, \quad (3.8)$$

which is consistent with (3.6) for  $h \ll D$  and  $h_3 \ll D_3$ .

**3.2. Single frequency resolution limits.** Using the paraxial model for our sensing matrix  $\mathcal{G}$ , we obtain from (3.7) and the definition of the cumulative coherence  $\mu(\mathcal{G}, s)$  that

$$\mu(\mathcal{G}, s) = \max_{j=1, \dots, N} \max_{|\mathcal{S}|=s} \sum_{q \in \mathcal{S}, q \neq j} \mathcal{U} \left( \frac{z_{1,j} - z_{1,q}}{H}, \frac{z_{3,j} - z_{3,q}}{H_3} \right) \mathcal{U} \left( \frac{z_{2,j} - z_{2,q}}{H}, \frac{z_{3,j} - z_{3,q}}{H_3} \right), \quad (3.9)$$

where

$$H = \frac{L}{ka} = \frac{\lambda L}{2\pi a}, \quad H_3 = \frac{2L^2}{ka^2} = \frac{\lambda L^2}{\pi a^2}, \quad (3.10)$$

and  $\mathcal{U}(\beta, \eta)$  is the absolute value of the Fresnel integral

$$\mathcal{U}(\beta, \eta) = \left| \int_{-1/2}^{1/2} dt e^{-i\beta t - i\eta t^2} \right|. \quad (3.11)$$

The search set of cardinality  $s$  is denoted by  $\mathcal{S}$ , to distinguish it from the set of indexes of the true support points of  $\boldsymbol{\rho}$ , called calygraphic  $\mathcal{S}$ .

As stated in Lemma 2.1, unique recovery of a sparse  $\boldsymbol{\rho}$  with the  $\ell_1$  minimization (2.7) is guaranteed when  $\mu(\mathcal{G}, s) < 1/2$ . This criterion allows us to estimate the resolution limit stated in the next two theorems.

**THEOREM 3.2.** *If the mesh size satisfies*

$$h > h^* = \frac{2\lambda L}{\pi a}, \quad h_3 > h_3^* = \frac{16\lambda L^2}{\pi a^2}, \quad (3.12)$$

$\ell_1$  optimization recovers exactly two sources located at any distinct grid points.

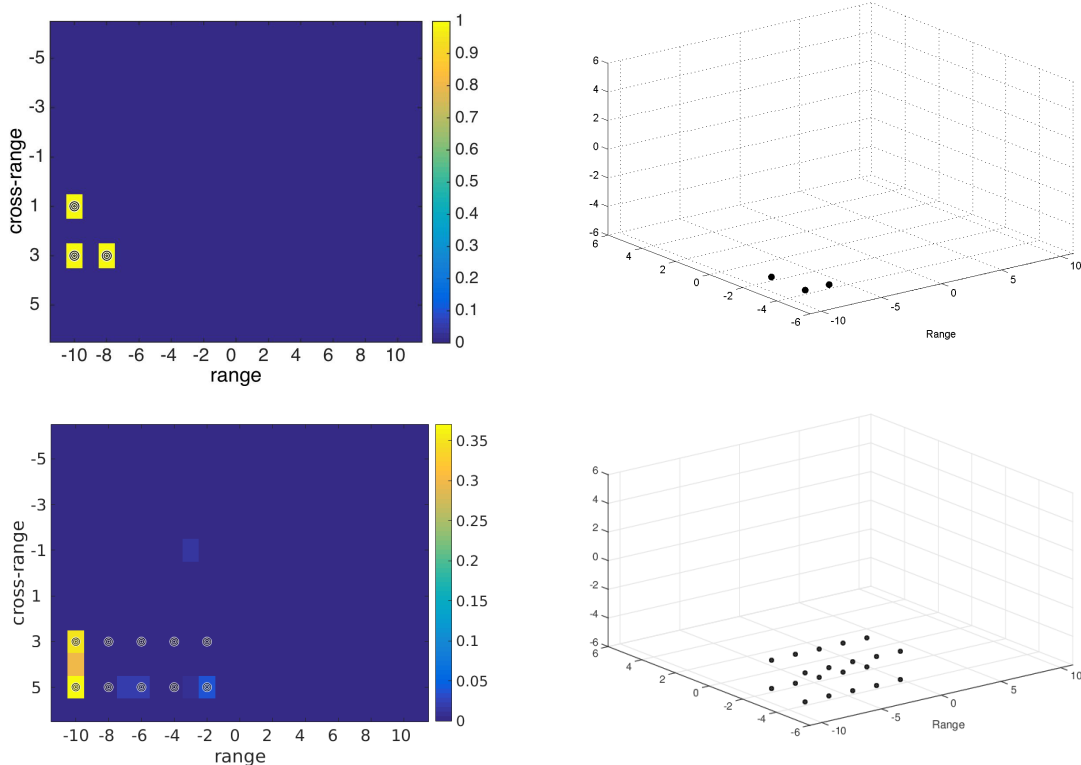


FIG. 3.2. Illustration of recovery of two imaging scenes discretized at base resolution  $(h^*, h^*, h_3^*)$ . The results on the top line are for 3 sources and on the bottom line for 20 sources. The distribution of the sources is displayed in the right column. The left column shows a cross-section of the images. The range axis is in units of  $h_3^*$  and the cross-range axis in units of  $h^*$ . The exact location of the sources is superposed on the images. They all have strength  $\rho_j = 1$ , for  $j \in S$ , and the magnitude of the reconstruction is shown with the color bar.

We call the estimates  $h^*$  and  $h_3^*$  the “base resolution”. They are the same, up to order one constants, as the well known resolution limits in array imaging [7]. We verified numerically the estimates (3.12) as follows. To check the value of  $h^*$ , we solved the optimization problem (2.7), as explained in appendix A, for data corresponding to a vector  $\boldsymbol{\rho}$  supported at any two grid points offset in cross-range. We determined the smallest  $h$  so that the relative error between  $\boldsymbol{\rho}$  and its numerical reconstruction was less than 1%. A similar estimation was done for  $h_3^*$ , with  $\boldsymbol{\rho}$  supported at points offset in range. The results were close to those in Theorem 3.2:  $0.46\lambda L/a$  for  $h^*$  and  $3\lambda L^2/a^2$  for  $h_3^*$ .

The next result states that when there are more sources to estimate, the resolution limits deteriorate. This is also illustrated in Figure 3.2, where we display images discretized at base resolution. The reconstruction is perfect for 3 sources (top line), but not for 20 sources (bottom line). How the resolution limits deteriorate with  $s$  depends on the distribution of the sources in the imaging region. Our estimate in the next theorem accounts for worse case scenarios.

**THEOREM 3.3.** *There exists a constant  $C$  of order one such that if the mesh size satisfies the conditions*

$$\frac{h/h^*}{h_3/h_3^*} = 0(1) \quad \text{and} \quad \left[ \left( \frac{h}{h^*} \right)^2 \frac{h_3}{h_3^*} \right]^{1/3} > Cs^{2/3}, \quad (3.13)$$

*$\ell_1$  optimization recovers exactly  $s$  sources located at any distinct grid points.*

The isotropic dilation of the mesh in (3.13) is for convenience, but the result generalizes to anisotropic dilations, where the mesh is stretched much more in one direction than the others. We refer to the proof in

section 3.4.1 for details on the generalization. The resolution decrease with  $s$  predicted by Theorem 3.3 may be traced to the slow decay with range offsets of the terms  $|\langle \mathbf{g}_j, \mathbf{g}_k \rangle|$  summed in  $\mu(\mathcal{G}, s)$ . This is also why

$$h_3^*/h^* = 8L/a \gg 1.$$

Sources at different ranges may have strong interaction, hence they must be further apart in order to get  $\mu(\mathcal{G}, s) < 1/2$ . In the next section we show that if the base range resolution improves, as it does with broad band data, then there is almost no resolution loss with the sparsity  $s$ .

**3.3. Broad band resolution limits.** When we have  $M_\omega$  frequency measurements, the vector of unknowns is defined as in (2.6), and the rows of the sensing matrix  $\mathcal{G}$  are indexed by the receiver-frequency pair  $(r, j) \in \{1, \dots, M_r\} \times \{1, \dots, M_\omega\}$ . Let  $\omega_o$  be the central frequency, so that

$$|\omega_j - \omega_o| \lesssim B, \quad j = 1, \dots, M_\omega, \quad (3.14)$$

where  $B$  is the bandwidth assumed to satisfy the scaling relation

$$1 \lesssim \max \left\{ \frac{D}{\lambda_o L/a}, \frac{D_3}{\lambda_o L^2/a^2} \right\} \ll \frac{\omega_o}{B} \ll \left( \frac{L}{a} \right)^2. \quad (3.15)$$

The lower bound says that  $\omega_o \gg B$ , so that  $\omega_o$  is the scale of all the measured frequencies  $\omega_j$ , for  $j = 1, \dots, M_\omega$ . The upper bound implies  $c/B \ll \lambda_o L^2/a^2$ , where we recall from the previous section that  $\lambda_o L^2/a^2$  is, up to a factor of order one, the base range resolution for single frequency measurements at  $\omega = \omega_o$ . The next theorem states that the base range resolution for multi frequency measurements is of order  $c/B$ , so (3.15) implies a gain in range resolution.

Let us assume, for simplicity of the calculations, a Gaussian signal

$$\hat{f}(\omega_j) = e^{-\frac{(\omega_j - \omega_o)^2}{4B^2}}. \quad (3.16)$$

The results should extend to any signal with bandwidth  $B$ , with modifications of the constants in the bounds. We show in appendix C that

$$\begin{aligned} \sum_{j=1}^{M_\omega} \sum_{r=1}^{M_r} |\hat{f}(\omega_j)|^2 \hat{G}(\omega_j, \vec{\mathbf{x}}_r, \vec{\mathbf{z}}_q) \overline{\hat{G}(\omega_j, \vec{\mathbf{x}}_r, \vec{\mathbf{z}}_l)} &\approx \frac{1}{(4\pi L)^2} \sum_{j=1}^{M_\omega} e^{-\frac{(\omega_j - \omega_o)^2}{2B^2} + i \frac{(\omega_j - \omega_o)}{c} (z_{3,q} - z_{3,l})} \times \\ &\sum_{r=1}^{M_r} e^{-ik_o \left[ \frac{|\mathbf{x}_r|^2 (z_{3,q} - z_{3,l})}{2L^2} + \frac{\mathbf{x}_r \cdot (\mathbf{z}_q - \mathbf{z}_l)}{L} \right]}, \end{aligned} \quad (3.17)$$

where  $k_o = \omega_o/c$  is the central wavenumber, and we proceeded as in Remark 3.1 to absorb the large phases  $e^{ik_o z_{3,q}}$  in the vector of unknowns. Assuming that the array is discretized on a mesh with spacing  $h_A$  satisfying (3.6), we approximate the sum over  $r$  by an integral over the aperture  $\mathcal{A}$ , as in the previous section. We also suppose that the frequencies  $\omega_j$  are spaced at intervals  $h_\omega$  satisfying  $h_\omega \ll c/D_3$ , so that we can write the sum over the frequencies as an integral over the bandwidth. Equation (3.17) becomes

$$\sum_{j=1}^{M_\omega} \sum_{r=1}^{M_r} |\hat{f}(\omega_j)|^2 \hat{G}(\omega_j, \vec{\mathbf{x}}_r, \vec{\mathbf{z}}_q) \overline{\hat{G}(\omega_j, \vec{\mathbf{x}}_r, \vec{\mathbf{z}}_l)} \approx \frac{\sqrt{2\pi} B}{(4\pi L h_A)^2 h_\omega} e^{-\frac{B^2 (z_{3,q} - z_{3,l})^2}{2c^2}} \times \quad (3.18)$$

$$\int_{\mathcal{A}} d\mathbf{x} e^{-ik_o \left[ \frac{|\mathbf{x}|^2 (z_{3,q} - z_{3,l})}{2L^2} + \frac{\mathbf{x} \cdot (\mathbf{z}_q - \mathbf{z}_l)}{L} \right]}, \quad (3.19)$$

and we can simplify it further by neglecting the quadratic phase in the integral over the aperture. This is because

$$\frac{k_o |\mathbf{x}|^2 |z_{3,q} - z_{3,l}|}{L^2} \lesssim O\left(\frac{k_o a^2 c}{L^2 B}\right) = O\left(\frac{\omega_o a^2}{BL^2}\right) \ll 1,$$



for range offsets in the support of order  $c/B$  of the Gaussian factor in (3.19). Integrating over the aperture and normalizing, we arrive at the following model of the products of the columns of the sensing matrix  $\mathcal{G}$ ,

$$|\langle \mathbf{g}_q, \mathbf{g}_l \rangle| = e^{-\frac{(z_{3,q} - z_{3,l})^2}{2(c/B)^2}} \left| \operatorname{sinc}\left(\frac{z_{1,q} - z_{1,l}}{2L/(k_o a)}\right) \right| \left| \operatorname{sinc}\left(\frac{z_{2,q} - z_{2,l}}{2L/(k_o a)}\right) \right|. \quad (3.20)$$

These are the terms in the cumulative coherence  $\mu(\mathcal{G}, s)$ , and the resolution limits are as stated next.

**THEOREM 3.4.** *Assume a sensing matrix  $\mathcal{G}$  with inner products of the columns defined by (3.20). If the mesh size  $\vec{\mathbf{h}} = (h, h, h_3)$  satisfies*

$$h > h^* = \frac{2}{\pi} \frac{\lambda_o L}{a}, \quad h_3 > h_3^* = \sqrt{2 \ln 2} \frac{c}{B}, \quad (3.21)$$

$\ell_1$  optimization recovers exactly any two sources on the grid. Moreover, there exists an order one constant  $C > 1$ , independent of  $s$ , such that if

$$\frac{h}{h^*}, \frac{h_3}{h_3^*} > C \ln s, \quad (3.22)$$

$\ell_1$  optimization recovers exactly any  $s$  sources on the grid.

We call  $h^*$  and  $h_3^*$  the base resolution, as in the previous section. While the cross-range resolution  $h^*$  is the same as in the single frequency case, the base range resolution  $h_3^*$  is significantly better. Moreover, there is little loss of resolution at large  $s$ . The mesh size grows at most logarithmically with  $s$ , as opposed to  $s^{2/3}$  in the single frequency case. This agrees with the known fact in array imaging that bandwidth improves the quality of images.

**3.4. Proofs.** The proofs of Theorems 3.2 and 3.3 which estimate the resolution in the single frequency case are in section 3.4.1. The proof of the broad band result in Theorem 3.4 is in section 3.4.1.

**3.4.1. Single frequency.** We begin with some basic bounds on the Fresnel integral (3.12). The simplest estimate is for  $\eta = 0$ , in which case

$$\mathcal{U}(\beta, 0) = \left| \operatorname{sinc}(\beta/2) \right| \leq \min\{1, 2/\beta\}. \quad (3.23)$$

For  $\eta \neq 0$  we can change variables and rewrite (3.12) in the form

$$\begin{aligned} \mathcal{U}(\beta, \eta) &= \frac{1}{\sqrt{\eta}} \left| \int_{\frac{\beta-\eta}{2\sqrt{\eta}}}^{\frac{\beta+\eta}{2\sqrt{\eta}}} dt e^{-it^2} \right| \leq \frac{1}{\sqrt{\eta}} \left[ \left| \int_0^{\frac{\beta+\eta}{2\sqrt{\eta}}} dt (\cos t^2 - i \sin t^2) \right| + \left| \int_0^{\frac{\beta-\eta}{2\sqrt{\eta}}} dt (\cos t^2 - i \sin t^2) \right| \right] \\ &\leq \frac{2\sqrt{2}}{\sqrt{\eta}}, \end{aligned} \quad (3.24)$$

for any  $\eta \neq 0$  and  $\beta \geq 0$ , where we used that

$$\left| \int_0^\alpha dt \cos t^2 \right| \leq 1, \quad \left| \int_0^\alpha dt \sin t^2 \right| \leq 1, \quad \forall \alpha \in \mathbb{R}.$$

The final estimate

$$\mathcal{U}(\beta, \eta) = \frac{1}{\sqrt{\eta}} \left| \int_{\frac{\beta-\eta}{2\sqrt{\eta}}}^{\frac{\beta+\eta}{2\sqrt{\eta}}} dt e^{it^2} \right| \leq \frac{\pi + 1}{\alpha}, \quad \text{for } \beta > \alpha + \eta, \quad \forall \alpha, \eta > 0, \quad (3.25)$$

follows from contour integration, as shown in appendix D.

**Proof of Theorem 3.2:** We wish to estimate  $h$  and  $h_3$  so that  $\mu(\mathcal{G}, 2) < 1/2$ . The cumulative coherence for  $s = 2$  is the same as the mutual coherence [17, 32], and in our case it takes the simple form

$$\mu(\mathcal{G}, 2) = \max_{\vec{\zeta} \in \mathbb{Z}^3, \vec{\zeta} \neq 0} \mathcal{U}\left(\frac{h\zeta_1}{H}, \frac{h_3\zeta_3}{H_3}\right) \mathcal{U}\left(\frac{h\zeta_2}{H}, \frac{h_3\zeta_3}{H_3}\right). \quad (3.26)$$

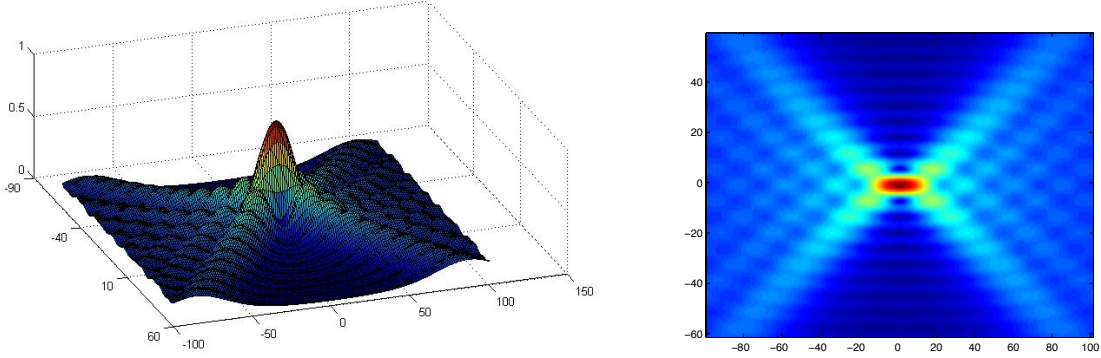


FIG. 3.3. Surface and top view display of the Fresnel integral  $\mathcal{U}(\beta, \eta)$  for  $|\beta|, |\eta| \leq 60$ . In the right plot abscissa is  $\eta$  and the ordinate is  $\beta$ .

Here we used that the sources are on the grid and denote by  $\vec{\zeta} = (\zeta_1, \zeta_2, \zeta_3)$  vectors with integer components. We display in Figure 3.3 the Fresnel integral  $\mathcal{U}(\beta, \eta)$ , and note that it is bounded above by 1, and larger than  $1/2$  for  $\vec{\zeta}$  near the origin. When  $\zeta_3 = 0$  we get from (3.23) that

$$\mathcal{U}\left(\frac{h\zeta_1}{H}, 0\right)\mathcal{U}\left(\frac{h\zeta_2}{H}, 0\right) \leq \frac{2H}{h},$$

because at least one of  $\zeta_1$  and  $\zeta_2$  is not equal to zero. If  $\zeta_3 \neq 0$  we have by (3.24)

$$\mathcal{U}\left(\frac{h\zeta_1}{H}, \frac{h_3\zeta_3}{H_3}\right)\mathcal{U}\left(\frac{h\zeta_2}{H}, \frac{h_3\zeta_3}{H_3}\right) \leq \frac{8H_3}{h_3}, \quad \forall \zeta_1, \zeta_2 \in \mathbb{Z}.$$

Thus,  $\mu(\mathcal{G}, 2)$  is guaranteed to be less than  $1/2$  if  $h > 4H = h^*$  and  $h_3 > 16H_3 = h_3^*$ . This concludes the proof of Theorem 3.2.  $\square$

**Proof of Theorem 3.3:** Note from the expression (3.9) of the cumulative coherence that it is translation invariant in  $\mathbb{R}^3$ . Thus, we can fix the origin at one source location and rewrite (3.9) as

$$\mu(\mathcal{G}, s) = \max_{|\Lambda|=s-1} \sum_{\vec{\zeta} \in \Lambda} \mathcal{U}\left(\frac{h\zeta_1}{H}, \frac{h_3\zeta_3}{H_3}\right)\mathcal{U}\left(\frac{h\zeta_2}{H}, \frac{h_3\zeta_3}{H_3}\right), \quad (3.27)$$

where  $\Lambda \subset \mathbb{Z}^3 \setminus \{0\}$  is a set of cardinality  $s - 1$ . The proof of the theorem follows from the bound on  $\mu(\mathcal{G}, s)$  stated in the next lemma and the relation  $h^* = 4H$  and  $h_3^* = 16H_3$  established above. The constant  $C$  in the lemma is the same as in (3.13).

LEMMA 3.5. *There exists constants  $C, C_1$  and  $C_2$  of order one such that*

$$\begin{aligned} \max_{|\Lambda|=s-1} \sum_{\vec{\zeta} \in \Lambda} \mathcal{U}\left(\frac{h\zeta_1}{H}, \frac{h_3\zeta_3}{H_3}\right)\mathcal{U}\left(\frac{h\zeta_2}{H}, \frac{h_3\zeta_3}{H_3}\right) &\leq 2^{5/3}C \left[ \frac{s^2}{(h/H)^2 h_3/H_3} \right]^{1/3} + \\ &C_1 \frac{H_3}{h_3} \ln s + C_2 \left(\frac{H}{h}\right)^2 \ln^2 s. \end{aligned} \quad (3.28)$$

The assumption in Theorem 3.3 that the mesh is dilated in an isotropic fashion, and the relations  $H = h^*/4$  and  $H_3 = h_3^*/16$  established above imply

$$\frac{\left[ \frac{s^2}{(h/H)^2 h_3/H_3} \right]^{1/3}}{\frac{H_3}{h_3} \ln s} = \frac{s^{2/3}}{\ln s} \left[ \left(\frac{h_3}{H_3}\right) / \left(\frac{h}{H}\right) \right]^{2/3} = O\left(\frac{s^{2/3}}{\ln s}\right),$$

and

$$\frac{\left[ \frac{s^2}{(h/H)^2 h_3/H_3} \right]^{1/3}}{\left( \frac{H}{h} \right)^2 \ln^2 s} = \frac{s^{2/3} h}{\ln^2 s H} \left[ \left( \frac{h}{H} \right) / \left( \frac{h_3}{H_3} \right) \right]^{1/3} = O\left( \frac{s^{2/3} h}{\ln^2 s H} \right).$$

Thus, the first term in the right hand side of (3.28) dominates the others for large  $s$  and  $h \gtrsim h_* = 4H$ , and the result stated in Theorem 3.3 follows. For anisotropic dilations of the mesh the logarithmic terms in (3.28) may become important. For example, when  $h_3/H_3 > (h/H)^4 s^2 / \ln^6 s$ , the last term in (3.28) dominates the bound, and we can get a unique solution for a modest mesh stretch in the cross-range direction  $h/H = O(\ln s)$  at the expense of a very large stretch in range  $h_3/H_3 = O(s^2 / \ln^2 s)$ .

**Proof of Lemma 3.5:** Let  $\Lambda_*$  be the set on which the maximum in (3.27) is achieved. It is difficult to determine  $\Lambda_*$  explicitly, so we construct another set  $\Lambda_\tau$ , which allows us to bound the cumulative coherence. We use the behavior of the Fresnel integral  $\mathcal{U}$ , shown in Figure 3.3, to guide us in the construction. We write  $\Lambda_\tau$  as the union of two sets  $\Lambda_\tau^+$  and  $\Lambda_\tau^0$ . The first set contains the left and right cones (in Figure 3.3 they are defined by the diagonal lines  $|\beta| = |\eta|$ ), where  $\mathcal{U}$  displays a slower decay, as well as a vicinity of the origin

$$\Lambda_\tau^+ = \left\{ \vec{\zeta} \in \mathbb{Z}^3 \text{ s.t. } 0 < |\zeta_3| \leq \frac{\tau}{h_3/H_3}, \quad |\zeta_1|, |\zeta_2| \leq \frac{H}{h} \left[ \frac{\tau}{\sqrt{h_3|\zeta_3|/H_3}} + \frac{h_3}{H_3} |\zeta_3| \right] \right\}. \quad (3.29)$$

The second set is for the points with  $\zeta_3 = 0$ ,

$$\Lambda_\tau^0 = \left\{ \vec{\zeta} \in \mathbb{Z}^3 \setminus \{0\} \text{ s.t. } |\zeta_1|, |\zeta_2| \leq \frac{\tau}{h/H} \right\}. \quad (3.30)$$

The parameter  $\tau > 1$  is used to control the volume of  $\Lambda_\tau = \Lambda_\tau^+ \cup \Lambda_\tau^0$ , so that it contains at least  $s$  grid points,

$$\tau = \min \left\{ \left[ \frac{3s}{8} \left( \frac{h}{H} \right)^2 \frac{h_3}{H_3} \right]^{1/3}, \frac{sh}{H}, \frac{sh_3}{H_3} \right\}. \quad (3.31)$$

The proof consists of two steps. First we derive the bound

$$\sum_{\zeta \in \Lambda_*} \mathcal{U}\left( \frac{h\zeta_{1,q}}{H}, \frac{h_3\zeta_{3,q}}{H_3} \right) \mathcal{U}\left( \frac{h\zeta_{2,q}}{H}, \frac{h_3\zeta_{3,q}}{H_3} \right) \leq \sum_{\zeta \in \Lambda_\tau} \mathcal{Z}(\vec{\zeta}), \quad (3.32)$$

where

$$\mathcal{Z}(\vec{\zeta}) = \begin{cases} \frac{2\sqrt{2}(\pi+1)H_3}{h_3|\zeta_3|}, & \text{if } \vec{\zeta} \in \Lambda_\tau^+, \\ \left[ \delta_{\zeta_1,0} + \frac{2H(1-\delta_{\zeta_1,0})}{h|\zeta_1|} \right] \left[ \delta_{\zeta_2,0} + \frac{2H(1-\delta_{\zeta_2,0})}{h|\zeta_2|} \right], & \text{if } \vec{\zeta} \in \Lambda_\tau^0. \end{cases} \quad (3.33)$$

Then we estimate the sum in the right hand side of (3.32). To prove (3.32) we show:

- (i)  $\Lambda_\tau^+$  contains at least  $s$  grid points.
- (ii) For any  $\vec{\zeta} \in \Lambda_\tau$ , we have

$$\mathcal{U}\left( \frac{h\zeta_1}{H}, \frac{h_3\zeta_3}{H_3} \right) \mathcal{U}\left( \frac{h\zeta_2}{H}, \frac{h_3\zeta_3}{H_3} \right) < \mathcal{Z}(\vec{\zeta}).$$

- (iii) For any  $\vec{\zeta} \notin \Lambda_\tau$ ,

$$\mathcal{U}\left( \frac{h\zeta_1}{H}, \frac{h_3\zeta_3}{H_3} \right) \mathcal{U}\left( \frac{h\zeta_2}{H}, \frac{h_3\zeta_3}{H_3} \right) \leq \frac{2\sqrt{2}(\pi+1)}{\tau} = \min_{\vec{\zeta} \in \Lambda_\tau^+} \mathcal{Z}(\vec{\zeta}).$$

By showing (i), we establish that there are at least as many terms to sum on the right hand side of (3.32) as on the left, and we can define a one to one map  $\mathcal{M} : \Lambda_\star \rightarrow \Lambda_\tau$ , such that  $\mathcal{M}(\vec{\zeta}) = \vec{\zeta}$  if  $\vec{\zeta} \in \Lambda_\star \cap \Lambda_\tau$  and  $\mathcal{M}(\vec{\zeta}) \in \Lambda_\tau^+$  if  $\vec{\zeta} \in \Lambda_\star \setminus (\Lambda_\star \cap \Lambda_\tau)$ . Points (ii) and (iii) ensure that

$$\mathcal{U}\left(\frac{h\zeta_1}{H}, \frac{h_3\zeta_3}{H_3}\right)\mathcal{U}\left(\frac{h\zeta_2}{H}, \frac{h_3\zeta_3}{H_3}\right) \leq \mathcal{Z}(\mathcal{M}(\vec{\zeta})), \quad \forall \vec{\zeta} \in \Lambda_\star,$$

from which (3.32) follows.

**Proof of (i):** This statement is trivial when  $\tau = sh/H$  or  $\tau = sh_3/H_3$ , because  $\Lambda_\tau^+$  has cardinality larger than  $s$ , by definition. Thus, let  $\tau = \left[\frac{3s}{8}\left(\frac{h}{H}\right)^2\frac{h_3}{H_3}\right]^{1/3}$ , and calculate the number  $s_\tau$  of grid points in  $\Lambda_\tau^+$  as

$$\begin{aligned} s_\tau &= \sum_{\vec{\zeta} \in \Lambda_\tau^+} 1 \geq \sum_{\zeta_3 \in \mathbb{Z}, 0 < |\zeta_3| \leq \tau H_3/h_3} \left[ 2 \left\lfloor \frac{H}{h} \left( \frac{\tau}{\sqrt{h_3|\zeta_3|/H_3}} + \frac{h_3}{H_3} |\zeta_3| \right) \right\rfloor \right]^2 \\ &\geq 4 \left( \frac{Hh_3}{hH_3} \right)^2 \sum_{\zeta_3 \in \mathbb{Z}, 0 < |\zeta_3| \leq \tau H_3/h_3} |\zeta_3|^2 \\ &\geq 4 \left( \frac{Hh_3}{hH_3} \right)^2 \frac{2}{3} \left( \frac{\tau H_3}{h_3} \right)^3 \\ &= \frac{8}{3} \left( \frac{H}{h} \right)^2 \frac{H_3}{h_3} \tau^3 \geq s. \end{aligned}$$

The first inequality is by definition of  $\Lambda_\tau^+$ . The factor 2 is due to the absolute values and  $\lfloor \cdot \rfloor$  denotes the integer part. The second inequality is because we omit one positive term in the sum. The third inequality is by direct summation

$$\sum_{j=1}^n j^2 = \frac{n(n+1)(2n+1)}{6} > \frac{n^3}{3}.$$

Again the factor 2 is due to the absolute values. The last inequality is by definition of  $\tau$ . This concludes the proof of (i).

**Proof of (ii):** This follows immediately from bounds (3.23) and (3.24).

**Proof of (iii):** We note from definition (3.33) of  $\mathcal{Z}(\vec{\zeta})$  and (3.29) that

$$\min_{\vec{\zeta} \in \Lambda_\tau^+} \mathcal{Z}(\vec{\zeta}) = \frac{2\sqrt{2}(\pi+1)}{\tau}.$$

Now consider an arbitrary  $\vec{\zeta} \notin \Lambda_\tau$ . Recalling definitions (3.29)-(3.30), we distinguish three cases:

1. If  $|\zeta_3| \geq \frac{\tau H_3}{h_3}$  we have by (3.24) that

$$\mathcal{U}\left(\frac{h\zeta_1}{H}, \frac{h_3\zeta_3}{H_3}\right)\mathcal{U}\left(\frac{h\zeta_2}{H}, \frac{h_3\zeta_3}{H_3}\right) \leq \frac{8}{h_3|\zeta_3|/H_3} \leq \frac{8}{\tau} < \frac{2\sqrt{2}(\pi+1)}{\tau}.$$

2. If  $\zeta_3 = 0$  we can assume without loss of generality that  $|\zeta_1| > \frac{\tau}{h/H}$  since at least one of  $\zeta_1$  and  $\zeta_2$  must satisfy this condition. We obtain from (3.23) that

$$\mathcal{U}\left(\frac{h\zeta_1}{H}, 0\right)\mathcal{U}\left(\frac{h\zeta_2}{H}, 0\right) \leq \frac{2}{h|\zeta_1|/H} < \frac{2}{\tau} < \frac{2\sqrt{2}(\pi+1)}{\tau}.$$

3. If  $0 < |\zeta_3| \leq \frac{\tau H_3}{h_3}$  we can assume without loss of generality that

$$|\zeta_1| > \frac{H}{h} \left[ \frac{\tau}{\sqrt{h_3|\zeta_3|/H_3}} + \frac{h_3}{H_3} |\zeta_3| \right],$$

since at least one of  $\zeta_1$  and  $\zeta_2$  must satisfy this condition. Then

$$\mathcal{U}\left(\frac{h}{H}\zeta_1, \frac{h_3}{H_3}\zeta_3\right) \leq \frac{\pi+1}{\tau} \sqrt{\frac{h_3}{H_3}|\zeta_3|},$$

by estimate (3.25), and using (3.24) for the other Fresnel integral we get

$$\mathcal{U}\left(\frac{h\zeta_1}{H}, \frac{h_3\zeta_3}{H_3}\right)\mathcal{U}\left(\frac{h\zeta_2}{H}, \frac{h_3\zeta_3}{H_3}\right) \leq \frac{2\sqrt{2}(\pi+1)}{\tau}.$$

This concludes the proof of (3.32).

It remains to estimate the right hand side in (3.32),

$$\sum_{\vec{\zeta} \in \Lambda_\tau} \mathcal{Z}(\vec{\zeta}) = \frac{2\sqrt{2}(\pi+1)H_3}{h_3} \sum_{\vec{\zeta} \in \Lambda_\tau^+} \frac{1}{|\zeta_3|} + \sum_{\vec{\zeta} \in \Lambda_\tau^0} \left[ \delta_{\zeta_1,0} + \frac{2H(1-\delta_{\zeta_1,0})}{h|\zeta_1|} \right] \left[ \delta_{\zeta_2,0} + \frac{2H(1-\delta_{\zeta_2,0})}{h|\zeta_2|} \right]. \quad (3.34)$$

For the first sum we have by the definition (3.29) of  $\Lambda_\tau^+$  that

$$\begin{aligned} \frac{H_3}{h_3} \sum_{\vec{\zeta} \in \Lambda_\tau^+} \frac{1}{|\zeta_3|} &\leq \frac{H_3}{h_3} \sum_{\zeta_3 \in \mathbb{Z}, 0 < |\zeta_3| \leq \frac{H_3\tau}{h_3}} \frac{1}{|\zeta_3|} \left[ 1 + \frac{2H}{h} \left( \frac{\tau}{\sqrt{h_3}|\zeta_3|/H_3} + \frac{h_3}{H_3}|\zeta_3| \right) \right]^2 \\ &\leq \sum_{\zeta_3 \in \mathbb{Z}, 0 < |\zeta_3| \leq \frac{H_3\tau}{h_3}} \left\{ 16 \left( \frac{H}{h} \right)^2 \left[ \frac{\tau^2}{(h_3|\zeta_3|/H_3)^2} + \frac{h_3}{H_3}|\zeta_3| \right] + \frac{2H_3}{h_3|\zeta_3|} \right\}. \end{aligned}$$

Now using that

$$\sum_{j=1}^n j \leq \frac{n(n+1)}{2}, \quad \sum_{j=1}^n \frac{1}{j^2} \leq \frac{\pi^2}{6}, \quad \sum_{j=1}^n \frac{1}{j} \leq 1 + \ln(j),$$

and substituting in the bound above, we get

$$\frac{H_3}{h_3} \sum_{\vec{\zeta} \in \Lambda_\tau^+} \frac{1}{|\zeta_3|} \leq 8 \left( \frac{H}{h} \right)^2 \left[ \tau + \frac{\tau^2 H_3}{h_3} \left( 1 + \frac{\pi^2 H_3}{3h_3} \right) \right] + \frac{2H_3}{h_3} \left[ 1 + \ln \left( \frac{\tau H_3}{h_3} \right) \right].$$

For the second sum in (3.34) we have

$$\begin{aligned} \sum_{\vec{\zeta} \in \Lambda_\tau^0} \left[ \delta_{\zeta_1,0} + \frac{2H(1-\delta_{\zeta_1,0})}{h|\zeta_1|} \right] \left[ \delta_{\zeta_2,0} + \frac{2H(1-\delta_{\zeta_2,0})}{h|\zeta_2|} \right] &= \sum_{\vec{\zeta} \in \Lambda_\tau^0} \left[ \delta_{\zeta_1,0}(1-\delta_{\zeta_2,0}) \frac{2H}{h|\zeta_2|} + \right. \\ &\quad \left. \delta_{\zeta_2,0}(1-\delta_{\zeta_1,0}) \frac{2H}{h|\zeta_1|} + (1-\delta_{\zeta_1,0})(1-\delta_{\zeta_2,0}) \frac{2H}{h|\zeta_1|} \frac{2H}{h|\zeta_2|} \right], \end{aligned}$$

because we cannot have both  $\zeta_1$  and  $\zeta_2 = 0$ . The first term is bounded as

$$\sum_{\vec{\zeta} \in \Lambda_\tau^0} \delta_{\zeta_1,0}(1-\delta_{\zeta_2,0}) \frac{2H}{h|\zeta_2|} = \frac{4H}{h} \sum_{\zeta_2=1}^{\lfloor \tau H/h \rfloor} \frac{1}{\zeta_2} \leq \frac{4H}{h} [1 + \ln(\tau H/h)],$$

and similar for the second term. For the last term we have

$$\sum_{\vec{\zeta} \in \Lambda_\tau^0} (1-\delta_{\zeta_1,0})(1-\delta_{\zeta_2,0}) \frac{2H}{h|\zeta_1|} \frac{2H}{h|\zeta_2|} = 4 \left( \frac{2H}{h} \right)^2 \left( \sum_{\zeta=1}^{\lfloor \tau H/h \rfloor} \frac{1}{\zeta} \right)^2 \leq \frac{16H^2}{h^2} [1 + \ln(\tau H/h)]^2.$$

Gathering the results we have for large  $s$ , and therefore large  $\tau$ ,

$$\sum_{\vec{\zeta} \in \Lambda_\tau} \mathcal{Z}(\vec{\zeta}) \leq \tilde{C} \left(\frac{H}{h}\right)^2 \frac{H_3}{h_3} \tau^2 + \tilde{C}_1 \frac{H_3}{h_3} \ln \tau + \tilde{C}_2 \left(\frac{H}{h}\right)^2 \ln^2 \tau, \quad (3.35)$$

with constant  $\tilde{C}$  close to  $16\sqrt{2}(\pi+1)$ ,  $\tilde{C}_1$  close to 2 and  $\tilde{C}_2$  close to 16. Here we used the expectation that  $h > h^* = 4H$  and  $h_3 > h_3^* = 16H$ . To finish the proof of the Lemma we obtain from definition (3.31) of  $\tau$  that

$$\frac{H_3}{h_3} \ln \tau \leq \frac{\ln(sh_3/H_3)}{h_3/H_3} \leq \frac{\ln s}{h_3/H_3} + e^{-1},$$

and similarly for  $H/h \ln \tau$ , where we used that  $\ln x/x$  attains its maximum over the interval  $[1, \infty)$  at  $x = e$ . Moreover, we note that the first term in (3.35) is negligible in comparison with the others unless  $\tau = \left[\frac{3s}{8} \left(\frac{h}{H}\right)^2 \frac{h_3}{H_3}\right]^{1/3}$ . Substituting this expression of  $\tau$  in the first term and using that  $s$  is large, we get Lemma 3.5 with constant  $C$  close to  $(3/2)^{2/3}(\pi+1)$ ,  $C_1$  close to  $\tilde{C}_1$  and  $C_2$  close to  $\tilde{C}_2$ . This concludes the proof of Theorem 3.3.  $\square$

**3.4.2. Broad band.** The proof of Theorem 3.4 is similar to that of Theorems 3.2 and 3.3, with modifications that account for the faster decay with the range offset of the inner products (3.20) of the columns of the sensing matrix.

Using the translation invariance of (3.20) and writing explicitly that the points  $\vec{z}_q$  are on the grid, we can write the cumulative coherence as

$$\mu(\mathcal{G}, s) = \max_{|\Lambda|=s-1} \sum_{\vec{\zeta} \in \Lambda} e^{-\left(\frac{h_3|\zeta_3|}{\mathcal{H}_3}\right)^2} \left| \operatorname{sinc}\left(\frac{h|\zeta_1|}{\mathcal{H}}\right) \right| \left| \operatorname{sinc}\left(\frac{h|\zeta_2|}{\mathcal{H}}\right) \right|. \quad (3.36)$$

Here  $\Lambda \in \mathbb{Z}^3 \setminus \{0\}$  is a set of cardinality  $s-1$ , as before, and we introduced the notation

$$\mathcal{H}_3 = \frac{\sqrt{2}c}{B}, \quad \mathcal{H} = \frac{2L}{k_o a}.$$

To derive the base resolution limits, let  $s=2$  in (3.36) and observe that

$$e^{-\left(\frac{h_3|\zeta_3|}{\mathcal{H}_3}\right)^2} \left| \operatorname{sinc}\left(\frac{h|\zeta_1|}{\mathcal{H}}\right) \right| \left| \operatorname{sinc}\left(\frac{h|\zeta_2|}{\mathcal{H}}\right) \right| \leq e^{-\left(\frac{h_3|\zeta_3|}{\mathcal{H}_3}\right)^2}, \quad \text{for } \zeta_3 \neq 0, \quad (3.37)$$

uniformly in  $\zeta_1, \zeta_2 \in \mathbb{Z}$ , and when  $\zeta_3 = 0$ ,

$$\left| \operatorname{sinc}\left(\frac{h|\zeta_1|}{\mathcal{H}}\right) \right| \left| \operatorname{sinc}\left(\frac{h|\zeta_2|}{\mathcal{H}}\right) \right| \leq \left[ \delta_{\zeta_1,0} + \frac{1 - \delta_{\zeta_1,0}}{h|\zeta_1|/\mathcal{H}} \right] \left[ \delta_{\zeta_2,0} + \frac{1 - \delta_{\zeta_2,0}}{h|\zeta_2|/\mathcal{H}} \right]. \quad (3.38)$$

Thus,  $\mu(\mathcal{G}, 2) < 1/2$  when  $\mathcal{H}/h < 1/2$  and  $e^{-h_3^2/\mathcal{H}_3^2} < 1/2$  or, equivalently, when

$$h > 2\mathcal{H} = h^* \quad \text{and} \quad h_3 > \sqrt{\ln(2)}\mathcal{H}_3 = h_3^*, \quad (3.39)$$

as stated in equation (3.21) of Theorem 3.4.

To prove the second statement of Theorem 3.4, for large  $s$ , we use assumption (3.22) and the relation (3.39) to write

$$\frac{h}{\mathcal{H}}, \frac{h_3}{\mathcal{H}_3} > \beta := \tilde{C} \ln s, \quad (3.40)$$

for a constant  $\tilde{C}$  that is slightly larger than  $C$ . We also define  $\mathcal{Z} : \mathbb{Z}^3 \rightarrow \mathbb{R}$  by

$$\mathcal{Z}(\vec{\zeta}) = e^{-\beta^2 \zeta_3^2} \left[ \delta_{\zeta_1,0} + \frac{1 - \delta_{\zeta_1,0}}{\beta|\zeta_1|} \right] \left[ \delta_{\zeta_2,0} + \frac{1 - \delta_{\zeta_2,0}}{\beta|\zeta_2|} \right], \quad (3.41)$$

and using (3.40) in (3.36) we get

$$\mu(\mathcal{G}, s) \leq \max_{|\Lambda|=s-1} \sum_{\vec{\zeta} \in \Lambda} \mathcal{Z}(\vec{\zeta}). \quad (3.42)$$

Now let  $\Lambda_\star$  denote the maximizing set in (3.42). We do not know it explicitly, but we can define a one-to-one mapping from  $\Lambda_\star$  to another set  $\Lambda_\beta$  which allows us to bound  $\mu(\mathcal{G}, s)$ . The construction of the set

$$\Lambda_\beta = \left\{ \vec{\zeta} \in \mathbb{Z}^3 \setminus \{0\} \quad \text{s.t.} \quad \mathcal{Z}(\vec{\zeta}) \geq \frac{1}{\beta s} \right\}, \quad (3.43)$$

is motivated by the rapid decay in range of the terms in the sum in (3.36). Explicitly, we note that when  $\vec{\zeta} \in \Lambda_\beta$  we have  $\zeta_3 = 0$  because if this were not true, definition (3.41) would give

$$\mathcal{Z}(\vec{\zeta}) \leq e^{-\beta^2 \zeta_3^2} \leq e^{-\beta^2} \leq e^{-2\beta} < e^{-\ln(\beta s)} = \frac{1}{\beta s}.$$

Here we assumed  $\beta > 2$ , which is consistent with (3.40) for large  $s$ , and since  $\beta > \ln \beta$ , we also have  $2\beta > \ln(\beta) + \ln s = \ln(\beta s)$ .

Let  $\Lambda_\beta^j$  be the intersection of  $\Lambda_\beta$  with the  $\zeta_j$  axis, for  $j = 1, 2$ . Then, the cardinality  $|\Lambda_\beta|$  of the set  $\Lambda_\beta$  satisfies

$$|\Lambda_\beta| \geq |\Lambda_\beta^1| + |\Lambda_\beta^2| = 4s,$$

because by definition (3.43),  $\zeta_j \in \Lambda_\beta^j$  means that  $|\zeta_j| \leq s$ , for  $j = 1, 2$ . Thus, there are at least  $4s$  points in  $\Lambda_\beta$ , and we can define a one to one mapping from the maximizing set  $\Lambda_\star$  to  $\Lambda_\beta$ . Moreover, since for any  $\vec{\zeta} \notin \Lambda_\beta$  we have  $\mathcal{Z}(\vec{\zeta}) < 1/(\beta s)$  we conclude from (3.42) that

$$\mu(\mathcal{G}, s) \leq \sum_{\vec{\zeta} \in \Lambda_\beta} \mathcal{Z}(\vec{\zeta}).$$

To bound the right hand side in this equation, note from definitions (3.41) and (3.43) that  $\Lambda_\beta$  is contained in the punctured disk  $\mathcal{D}_s$  of radius  $s$ ,

$$\mathcal{D}_s = \left\{ \vec{\zeta} \in \mathbb{Z}^3 \setminus \{0\} \quad \text{s.t.} \quad \zeta_3 = 0, \quad |\zeta_1|, |\zeta_2| \leq s \right\},$$

and obtain

$$\begin{aligned} \mu(\mathcal{G}, s) &\leq \sum_{\vec{\zeta} \in \mathcal{D}_s} \mathcal{Z}(\vec{\zeta}) = \prod_{j=1}^2 \sum_{|\zeta_j| \leq s} \left[ \delta_{\zeta_j, 0} + \frac{1 - \delta_{\zeta_j, 0}}{\beta |\zeta_j|} \right] - \mathcal{Z}(0) \leq \left[ 1 + \frac{2}{\beta} (1 + \ln s) \right]^2 - 1 \\ &= \frac{4 \ln s}{\beta} + \frac{4 \ln^2 s}{\beta^2} = \frac{4 \ln s}{\beta} + \frac{4}{\beta} + \left[ \frac{2}{\beta} (1 + \ln s) \right]^2. \end{aligned}$$

The proof of Theorem 3.4 is completed with the observation that we can make the bound in this estimate less than  $1/2$  by choosing the constant  $\tilde{C}$  in (3.40) large enough, independent of  $s$ .  $\square$

**4. Imaging on fine grids.** The resolution estimates in Theorems 3.2-3.4 do not account for noise and modeling errors due to off-grid placement of the sources, which may be large for coarser discretizations required by the theorems. In this section we mitigate the modeling error by discretizing the imaging region on a fine mesh, and then study the results of the  $\ell_1$  optimization. As the results in the previous section shows it is impossible to have a meaningful answer for arbitrary distributions of the sources. However, if they are located at points or clusters of points that are sufficiently far apart, the results are useful, as illustrated in Figure 4.1. In the left image we display the result of the optimization for a discretizations of the imaging region at the base resolution  $\tilde{\mathbf{h}}^\star$  defined in Theorem 3.2. If the sources were on the grid, the

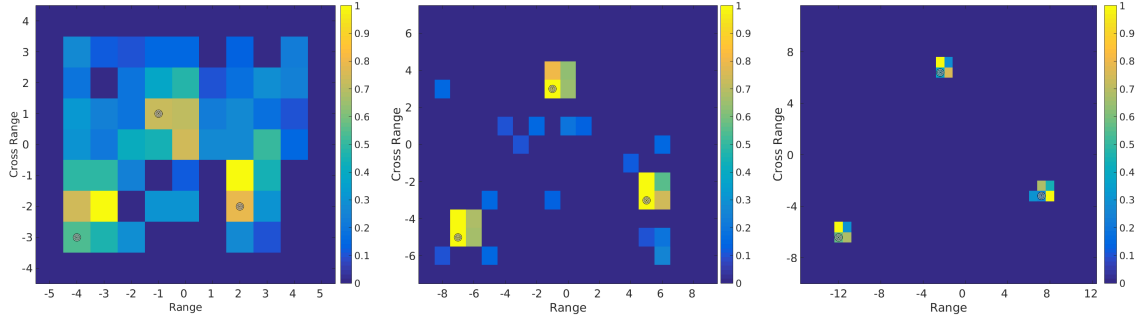


FIG. 4.1. *Effect of modeling error due to sources off-grid. From left to right: Image discretized at base resolution, at 1/2 base resolution and 1/4 base resolution.*

reconstruction would have been perfect. Here the sources are off-grid, and the reconstruction is poor due to the modeling error. The result is better in the other two plots because the modeling error is reduced by taking a smaller mesh size. While in general a discretization at  $1/4\vec{\mathbf{h}}^*$  does not guarantee a good recovery, here the result is good because the sources are far apart, and thus have little interaction. The analysis in this section formalizes this observation. We point the reader to [20] for a different study of similar ideas, and algorithms designed to take advantage of the weak interaction between the sources. Here we study directly the optimization problems (2.7) and (2.9) and consider in addition clusters of sources.

We begin in section 4.1 with the statement of results for well separated sources and then consider clusters of sources in section 4.2. The proofs are in section 4.3.

**4.1. Statement of results for well separated sources.** Let us modify the notation slightly, and call  $\mathbf{g}_{\vec{\mathbf{y}}}$  the normalized vector of Green's functions taking us from the array to the point  $\vec{\mathbf{y}}$  in the imaging region. When  $\vec{\mathbf{y}}$  is a point  $\vec{\mathbf{z}}_j$  on the grid, then  $\mathbf{g}_{\vec{\mathbf{y}}}$  is the same as  $\mathbf{g}_j$  defined before, and we have the simpler notation  $\mathbf{g}_j \equiv \mathbf{g}_{\vec{\mathbf{z}}_j}$ . We quantify the interaction between two sources located at  $\vec{\mathbf{y}}$  and  $\vec{\mathbf{y}}'$  in the imaging region  $W$  by the value of  $\left| \langle \mathbf{g}_{\vec{\mathbf{y}}}, \mathbf{g}_{\vec{\mathbf{y}}'} \rangle \right|$ . Explicitly, in terms of the semi-metric  $\mathcal{D} : W \times W \rightarrow [0, 1]$ ,

$$\mathcal{D}(\vec{\mathbf{y}}, \vec{\mathbf{y}}') = 1 - \left| \langle \mathbf{g}_{\vec{\mathbf{y}}}, \mathbf{g}_{\vec{\mathbf{y}}'} \rangle \right|, \quad \forall \vec{\mathbf{y}}, \vec{\mathbf{y}}' \in W, \quad (4.1)$$

that defines the open ball

$$\mathcal{B}_r(\vec{\mathbf{y}}) = \left\{ \vec{\mathbf{y}}' \in \mathbb{R}^3 \quad \text{s.t.} \quad \mathcal{D}(\vec{\mathbf{y}}, \vec{\mathbf{y}}') < r \right\}, \quad (4.2)$$

we say that points  $\vec{\mathbf{y}}'$  outside  $\mathcal{B}_r(\vec{\mathbf{y}})$  have a weaker interaction with  $\vec{\mathbf{y}}$  than the points in the ball,

$$\left| \langle \mathbf{g}_{\vec{\mathbf{y}}}, \mathbf{g}_{\vec{\mathbf{y}}'} \rangle \right| \leq 1 - r, \quad \forall \vec{\mathbf{y}}' \notin \mathcal{B}_r(\vec{\mathbf{y}}). \quad (4.3)$$

In these definitions it does not matter if we have single or multiple frequency measurements\*. In both cases we know from the previous section that  $\left| \langle \mathbf{g}_{\vec{\mathbf{y}}}, \mathbf{g}_{\vec{\mathbf{y}}'} \rangle \right|$  is a function of  $\vec{\mathbf{y}} - \vec{\mathbf{y}}'$  which peaks at the origin, and is monotonically decreasing in its vicinity. This means that there exists a small enough  $\bar{r}$  such that if  $r < \bar{r}$  and  $\vec{\mathbf{y}}' \in \mathcal{B}_r(\vec{\mathbf{y}})$ ,  $\vec{\mathbf{y}}$  and  $\vec{\mathbf{y}}'$  are close in Euclidian distance.

Suppose that we have  $s$  sources in the imaging region  $W$ , supported at points in the set  $\mathcal{Y} = \{\vec{\mathbf{y}}_j, \quad j = 1, \dots, s\}$ , and discretize  $W$  on a grid with  $N$  points  $\vec{\mathbf{z}}_j$  and mesh size  $\vec{\mathbf{h}}$  that is as small as needed to mitigate the modeling error. We define the interaction coefficient of the set  $\mathcal{Y}$  by

$$\mathcal{I}(\mathcal{Y}) = \max_{q=1, \dots, N} \sum_{\vec{\mathbf{y}}_j \in \mathcal{Y} \setminus \{\vec{\mathbf{z}}_q\}} \left| \langle \mathbf{g}_{\vec{\mathbf{y}}_j}, \mathbf{g}_q \rangle \right|, \quad (4.4)$$

\*The theory applies to both single frequency and multiple frequency measurements, but the numerical simulations are for a single frequency.



where  $\mathcal{N}(\vec{\mathbf{z}}_q)$  is the closest point to  $\vec{\mathbf{z}}_q$  in  $\mathcal{Y}$ , with respect to semi-metric  $\mathcal{D}$ . Here it is possible to make  $\mathcal{I}(\mathcal{Y})$  independent of the mesh by replacing the maximum with supremum over the imaging window which is an open subset of  $\mathbb{R}^3$ . Note that  $\mathcal{I}(\mathcal{Y})$  is similar to the cumulative mutual coherence  $\mu(\mathcal{G}, s)$ , except that the set  $\mathcal{Y}$  is fixed and the points in it may not be on the grid. Note also that  $\mathcal{I}(\mathcal{Y})$  is well defined even when there are multiple points in  $\mathcal{Y}$  that are closest to  $\vec{\mathbf{z}}_q$ . In such cases we let  $\mathcal{N}(\vec{\mathbf{z}}_q)$  be any one of these points without affecting the value of  $\mathcal{I}(\mathcal{Y})$ .

The next two theorems describe the support of the  $\ell_1$  minimizer. Theorem 4.1 and its corollary are for formulation (2.7) of the optimization problem, which assumes an exact model. Theorem 4.3 is for formulation (2.9) which accounts for noise and modeling error.

**THEOREM 4.1.** *Suppose that the unknown sources are supported on the fine grid at points  $\vec{\mathbf{z}}_j$  enumerated by the set  $S$  of cardinality  $s$ , and are represented by the  $s$ -sparse vector  $\boldsymbol{\rho} \in \mathbb{C}^N$ , satisfying  $\mathcal{G}\boldsymbol{\rho} = \mathbf{d}$ . The sources are assumed sufficiently far apart so that for some  $r \in (0, 1)$  the balls  $\mathcal{B}_r(\vec{\mathbf{z}}_j)$  are disjoint. Let  $\boldsymbol{\rho}_\star$  be the solution of the optimization problem (2.7) and decompose it as*

$$\boldsymbol{\rho}_\star = \boldsymbol{\rho}_\star^{(i)} + \boldsymbol{\rho}_\star^{(o)}, \quad (4.5)$$

where  $\text{supp } \boldsymbol{\rho}_\star^{(i)} \subset \bigcup_{j \in S} \mathcal{B}_r(\vec{\mathbf{z}}_j)$ , and  $\boldsymbol{\rho}_\star^{(o)}$  is supported in the complement of this union. Then,

$$\|\boldsymbol{\rho}_\star^{(o)}\|_1 \leq \frac{2\mathcal{I}(\mathcal{Y})}{r} \|\boldsymbol{\rho}_\star\|_1. \quad (4.6)$$

This theorem says that when the interaction coefficient  $\mathcal{I}(\mathcal{Y})$  is smaller than  $r/2$ , the support of the optimal  $\boldsymbol{\rho}_\star$  is concentrated in the vicinity of the sources. The next corollary quantifies the error of the reconstruction.

**COROLLARY 4.2.** *Under the same assumptions as in Theorem 4.1, the error of the reconstruction is quantified by*

$$\|\boldsymbol{\rho} - \bar{\boldsymbol{\rho}}_\star\|_1 \leq \frac{2\mathcal{I}(\mathcal{Y})}{r} \|\boldsymbol{\rho}\|_1, \quad (4.7)$$

where  $\bar{\boldsymbol{\rho}}_\star$  is the effective source vector in  $\mathbb{C}^N$  with  $j$ -th component given by

$$\bar{\rho}_{\star j} = \begin{cases} \sum_{q \in \mathcal{S}_j} \rho_{\star q}^{(i)} \langle \mathbf{g}_j, \mathbf{g}_q \rangle, & \text{for } j \in \mathcal{S}, \\ 0, & \text{for } j \notin \mathcal{S}, \end{cases} \quad (4.8)$$

where  $\rho_{\star q}^{(i)}$  denotes the  $q$ -th component of  $\boldsymbol{\rho}_\star^{(i)}$  and  $\mathcal{S}_j$  is the set<sup>†</sup> of indexes of the grid points supported in  $\mathcal{B}_r(\vec{\mathbf{z}}_j)$ .

Note that  $\bar{\boldsymbol{\rho}}_\star$  is an  $s$ -sparse vector of the same support  $\mathcal{S}$  as  $\boldsymbol{\rho}$ , but with entries given by the “weighted” sum of the components of  $\boldsymbol{\rho}_\star^{(i)}$  supported in the vicinity of each source. When the radius  $r$  is small, the complex weights  $\langle \mathbf{g}_j, \mathbf{g}_q \rangle$  are close to one, and  $\bar{\rho}_{\star j}$  is approximately the sum of the components of  $\boldsymbol{\rho}_\star^{(i)}$  supported in  $\mathcal{S}_j$ . Furthermore, (4.7) implies that when  $\frac{2\mathcal{I}(\mathcal{Y})}{r}$  is sufficiently small such that the right hand side of (4.7) is less than  $\min_{j \in \mathcal{S}} \{|\rho_j|\}$ ,  $\boldsymbol{\rho}_\star$  has a non-zero component in the  $r$ -neighborhood of every source location.

The statements of Theorem 4.1 and Corollary 4.2 are already illustrated in the right plot of Figure 4.1. Additional examples are in Figure 4.2, where we show numerical reconstructions for two sources (left plot) and five sources (right plot). We display the balls  $\mathcal{B}_r(\vec{\mathbf{y}}_j)$  in green, and the entries in  $\boldsymbol{\rho}_\star$  with stars of size proportional to their magnitude. In the left plot (see also the zoom in Figure 4.3) the sources have weak interaction  $\mathcal{I}(\mathcal{Y}) = 0.086$ , and for  $r = 0.009$  we have  $\|\boldsymbol{\rho}_\star^{(o)}\|_1 / \|\boldsymbol{\rho}_\star\|_1 = 2.4\%$  and  $\|\boldsymbol{\rho} - \bar{\boldsymbol{\rho}}_\star\|_1 / \|\boldsymbol{\rho}\|_1 = 18\%$ .

<sup>†</sup>Note that the set  $\mathcal{S}_j$  depends on  $r$ , but for simplicity we suppress the dependence in the notation.

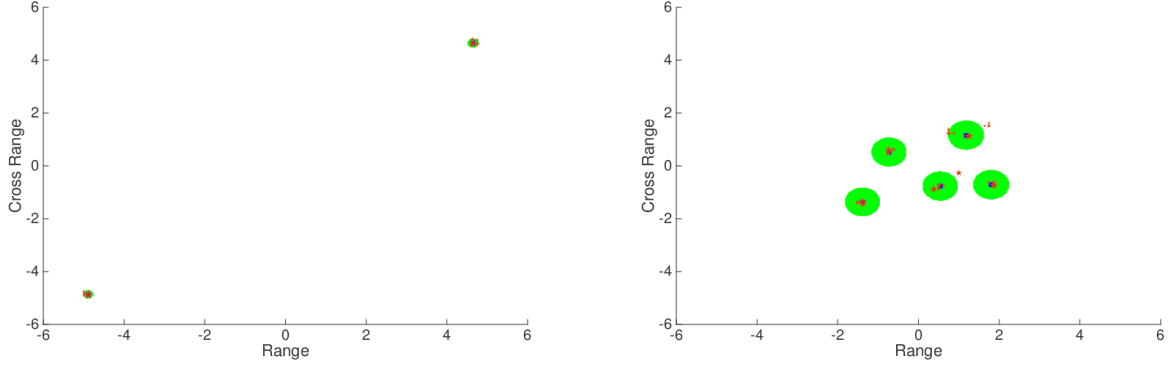


FIG. 4.2. Reconstructions of two sources (left) and five sources (right). The support of  $\rho_*$  is indicated with a star of size proportional to its magnitude. The balls  $\mathcal{B}_r(\vec{y}_j)$  are drawn in green. The radius is 0.009 in the left plot and 0.11 in the right plot. The axes are range and cross-range in units of base resolution  $h_3^*$  and  $h^*$ .

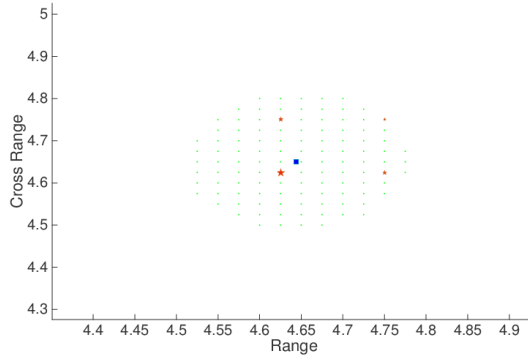


FIG. 4.3. Zoom of the image displayed in the left plot of Figure 4.2 around one of the sources shown with a blue square.

For  $r = 0.005$  the error drops to 1.14%, however  $\|\rho_*^{(o)}\|_1/\|\rho_*\|_1$  grows to 19%. That is to say, roughly 80% of the amplitude of the reconstruction  $\rho_*$  is accumulated very close near the sources. In the right plot the interaction coefficient is larger  $\mathcal{I}(\mathcal{Y}) = 1.43$ , but the reconstruction is still good,  $\|\rho_*^{(o)}\|_1/\|\rho_*\|_1 = 0.5\%$  and  $\|\rho - \rho_*\|_1/\|\rho\|_1 = 10\%$  for  $r = 0.11$ .

Note that in both simulations the support of the reconstruction is much better than predicted by Theorem 4.1, which gives a pessimistic bound for  $r > 2\mathcal{I}(\mathcal{Y})$ . A sharper estimate may be obtained under the additional assumption that all the entries in  $\rho$  are positive, by taking advantage of cancellations of the oscillatory terms in the sums analyzed in the proof of the theorem in section 4.3.1. However, this is difficult to do without making strong assumptions on the geometric distribution of the sources in the imaging region.

The next theorem considers the more general case of an inexact model, due to noisy data and sources off the grid, and uses the  $\ell_1$  penalty formulation (2.9). The result is stronger than in Theorem 3.2, as it states that the minimizer  $\rho_*$  is exactly supported in the vicinity of the sources, for large enough penalty parameter  $\gamma$ . This is somewhat expected, as increasing  $\gamma$  in (2.9) means putting more emphasis on having a smaller  $\ell_1$  norm i.e., a sparser solution. What is interesting is that the support of this sparse solution is guaranteed to be near that of the unknown sources. However, increasing  $\gamma$  comes at the cost of a larger residual  $\|\mathcal{G}\rho_* - \mathbf{d}\|_2$ , and there is no guarantee that the error of recovery of  $\rho$  is small, as in Corollary 4.2.

**THEOREM 4.3.** *Consider  $s$  sources supported in the set  $\mathcal{Y} = \{\vec{y}_j, j = 1, \dots, s\}$ , with interaction coefficient  $\mathcal{I}(\mathcal{Y}) < 1/2$ , so that we can find an  $r \in (0, 1)$  satisfying  $r > 2\mathcal{I}(\mathcal{Y})$ . Then, for sufficiently large penalty parameter  $\gamma$ , that depends on the noise and modeling error, the minimizer  $\rho_*$  of (2.9) is supported in  $\bigcup_{j=1}^s \mathcal{B}_r(\vec{y}_j)$ .*

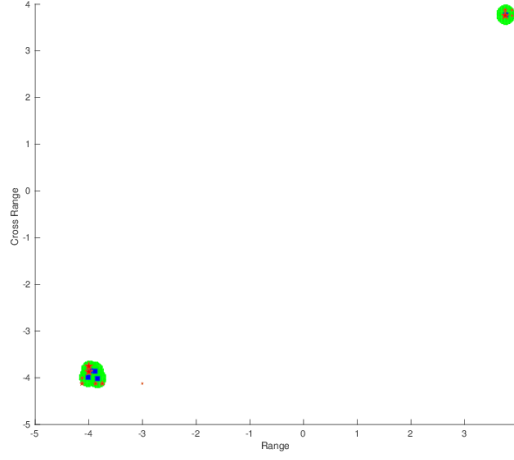


FIG. 4.4. Numerical simulation for two clusters of sources. The support of  $\rho_\star$  is indicated with a star of size proportional to its magnitude. The balls  $\mathcal{B}_r$  of radius  $r = 0.017$  are drawn in green. The axes are range and cross-range in units of base resolution  $h_3^\star$  and  $h^\star$ .

**4.2. Statement of results for clusters of sources.** Here we give the generalization of Theorem 4.1 to clusters of sources. We assume for simplicity, as in Theorem 4.1, that the sources are on the grid. The result extends to sources off the grid by modifying the proof of Theorem 4.3.

Let us define the effective support  $\mathcal{S}_\varepsilon \subset \mathcal{S}$  of  $\rho \in \mathbb{C}^N$ , for some  $\varepsilon \in (0, 1)$ , so that

$$\mathcal{Y} = \{\bar{\mathbf{y}}_1, \dots, \bar{\mathbf{y}}_s\} = \{\bar{\mathbf{z}}_j, j \in \mathcal{S}\} \subset \bigcup_{q \in \mathcal{S}_\varepsilon} \mathcal{B}_\varepsilon(\bar{\mathbf{z}}_q), \quad (4.9)$$

where

$$\mathcal{B}_\varepsilon(\bar{\mathbf{z}}_q) \cap \mathcal{B}_\varepsilon(\bar{\mathbf{z}}_l) = \emptyset \quad \forall l, q \in \mathcal{S}_\varepsilon, l \neq q.$$

More explicitly, we cover the set  $\mathcal{Y}$  of locations of the sources with disjoint balls of radius  $\varepsilon$ , centered at points in  $\mathcal{S}_\varepsilon$ . The set  $\mathcal{S}_\varepsilon$  is the support of the effective source vector  $\bar{\rho}$ , with entries defined similarly to (4.8)

$$\bar{\rho}_j = \begin{cases} \sum_{q \in \mathcal{S} \cap \mathcal{B}_\varepsilon(\bar{\mathbf{z}}_j)} \rho_q \langle \mathbf{g}_q, \mathbf{g}_j \rangle, & \text{for } j \in \mathcal{S}_\varepsilon, \\ 0, & \text{otherwise,} \end{cases} \quad (4.10)$$

Obviously  $\bar{\rho}$  depends on  $\varepsilon$ , but we suppress the dependence in the notation. When  $\varepsilon \ll 1$ , meaning that the sources are tightly clustered around the points in  $\mathcal{S}_\varepsilon$ , the effective source is approximately the sum of the entries of  $\rho$  supported in the cluster. When  $\varepsilon$  is larger the complex weights in (4.10) can be far from one and oscillatory, so there may be a lot of cancellations in the sum in (4.10). Cancellations (destructive interference of sources) can arise for tight clusters as well, when the entries in  $\rho$  in a cluster have opposite signs.

The result stated in the next theorem says that if the clusters are far apart and there is little destructive interference of the sources in the cluster, the support of the optimizer  $\rho_\star$  of (2.7) is concentrated near the sources.

**THEOREM 4.4.** *Suppose that the unknown sources are supported on the grid at points enumerated by the set  $\mathcal{S}$ , and that there is an  $\varepsilon \in (0, 1)$  for which we can define the effective support  $\mathcal{S}_\varepsilon$ . Let  $\rho_\star$  be the  $\ell_1$  minimizer of (2.7) and decompose it as  $\rho = \rho_\star^{(i)} + \rho_\star^{(o)}$ , where  $\rho_\star^{(i)}$  is supported in the disjoint union  $\bigcup_{j \in \mathcal{S}_\varepsilon} \mathcal{B}_r(\bar{\mathbf{z}}_j)$ , for  $r$  satisfying  $\varepsilon < r < 1$ , and  $\rho_\star^{(o)}$  is supported in the complement of this union. We have*

$$\|\rho_\star^{(o)}\|_1 \leq \frac{2\mathcal{I}(\mathcal{Y}_\varepsilon)}{r} \|\rho_\star\|_1 + \frac{\|\rho\|_1 - \|\bar{\rho}\|_1}{r}, \quad (4.11)$$

where  $\mathcal{Y}_\varepsilon = \{\bar{\mathbf{z}}_j, j \in \mathcal{S}_\varepsilon\}$  is assumed to satisfy  $\mathcal{I}(\mathcal{Y}_\varepsilon) < 1$ .

The assumption  $\mathcal{I}(\mathcal{Y}_\varepsilon) < 1$  is used in the proof, but for the estimate (4.11) to be useful we need  $\mathcal{I}(\mathcal{Y}_\varepsilon) < r/2 < 1/2$ . This is because by definition (4.10) of  $\bar{\rho}$  we have  $\|\bar{\rho}\|_1 \leq \|\rho\|_1$  and the bound is larger

than  $\|\boldsymbol{\rho}_\star\|_1$  for  $\mathcal{I}(\mathcal{Y}_\varepsilon) > r/2$ . As before, the estimate in the theorem is pessimistic. The numerical results are better, as illustrated by the simulation in Figure 4.4. There are five sources with amplitude equal to one, and locations indicated in the plot with blue squares. One source is isolated and the other four form a cluster, so  $\mathcal{S}_\varepsilon$  has cardinality two. The interaction coefficient of the set  $\mathcal{Y}$  is large,  $\mathcal{I}(\mathcal{Y}) = 3.02$ , because of the cluster, but  $\mathcal{I}(\mathcal{Y}_\varepsilon) = 0.096$ . The balls shown with green in the figure are for  $r = 0.017$ , and the entries in  $\boldsymbol{\rho}_\star$  are indicated with stars of size proportional to the magnitude. The error is  $\|\boldsymbol{\rho}_\star^{(o)}\|_1/\|\boldsymbol{\rho}_\star\|_1 = 0.54\%$ . The restriction of  $\boldsymbol{\rho}_\star$  to the ball containing the cluster has  $l_1$  norm 2.3. The restriction to the other ball has norm 0.96, which is close to the amplitude of the isolated source.

**4.3. Proofs.** We begin in section 4.3.1 with the proofs of Theorem 4.1 and its Corollary 4.2. Theorem 4.4 for clusters of sources is proved in section 4.3.2. The proofs of Theorems 4.1 and 4.4 are similar but the proof of Theorem 4.3 for  $l_1$  penalty reconstructions is more involved. We present it in section 4.3.3.

**4.3.1.  $l_1$  optimal reconstructions of well separated sources.** Theorem 4.1 and its corollary assume an exact model, with well separated sources on the grid, at points indexed by  $\mathcal{S}$ . Recall that  $\mathcal{S}_q$  is the set that enumerates the grid points supported in  $\mathcal{B}_r(\bar{\mathbf{z}}_q)$ , for  $q \in \mathcal{S}$ . The balls  $\mathcal{B}_r(\bar{\mathbf{z}}_q)$  are disjoint by assumption, so each nonzero entry in  $\boldsymbol{\rho}_\star^{(i)}$  is contained in exactly one ball and

$$\text{supp } \boldsymbol{\rho}_\star^{(i)} = \bigcup_{q \in \mathcal{S}} \mathcal{S}_q.$$

By definition of  $\boldsymbol{\rho}_\star$  we have  $\mathcal{G}\boldsymbol{\rho} = \mathcal{G}\boldsymbol{\rho}_\star$  or more explicitly,

$$\sum_{q \in \mathcal{S}} \rho_q \mathbf{g}_q = \sum_{q \in \mathcal{S}} \sum_{j \in \mathcal{S}_q} \rho_{\star j}^{(i)} \mathbf{g}_j + \sum_{j \in \mathcal{S}^c} \rho_{\star j}^{(o)} \mathbf{g}_j, \quad (4.12)$$

where we denote the support of  $\boldsymbol{\rho}_\star^{(o)}$  by  $\mathcal{S}^c = \{1, \dots, N\} \setminus \bigcup_{j \in \mathcal{S}} \mathcal{S}_j$ . The proof of the theorem and its corollary amounts to estimating the inner products of the left and right sides of equation (4.12) with a carefully chosen vector  $\mathbf{u}$ , as shown next.

**Proof of Theorem 4.1:** Let us define the vector

$$\mathbf{u} = \sum_{q \in \mathcal{S}} \text{sign}(\rho_q) \mathbf{g}_q, \quad (4.13)$$

where “sign” denotes the complex sign function, and take the inner product of  $\mathbf{u}$  with the left and right hand side in (4.12). We obtain

$$\mathcal{T}_L := \left| \sum_{q \in \mathcal{S}} \rho_q \langle \mathbf{g}_q, \mathbf{u} \rangle \right| = \left| \sum_{q \in \mathcal{S}} \sum_{j \in \mathcal{S}_q} \rho_{\star j}^{(i)} \langle \mathbf{g}_j, \mathbf{u} \rangle + \sum_{j \in \mathcal{S}^c} \rho_{\star j}^{(o)} \langle \mathbf{g}_j, \mathbf{u} \rangle \right| =: \mathcal{T}_R, \quad (4.14)$$

where obviously the left hand side  $\mathcal{T}_L$  equals the right hand side  $\mathcal{T}_R$ . We distinguish them here so we can bound them separately below and above.

For  $\mathcal{T}_L$  we have

$$\begin{aligned} \mathcal{T}_L &= \left| \sum_{q \in \mathcal{S}} \left[ \rho_q \text{sign}(\rho_q) + \rho_q \sum_{j \in \mathcal{S}, j \neq q} \text{sign}(\rho_j) \langle \mathbf{g}_q, \mathbf{g}_j \rangle \right] \right| \\ &= \left| \sum_{q \in \mathcal{S}} \left[ |\rho_q| + \rho_q \sum_{j \in \mathcal{S}, j \neq q} \text{sign}(\rho_j) \langle \mathbf{g}_q, \mathbf{g}_j \rangle \right] \right| \\ &\geq \sum_{q \in \mathcal{S}} |\rho_q| - \sum_{q \in \mathcal{S}} |\rho_q| \sum_{j \in \mathcal{S}, j \neq q} |\langle \mathbf{g}_q, \mathbf{g}_j \rangle|, \end{aligned}$$

where we used the definition of the sign function and the triangle inequality. Since

$$\sum_{j \in \mathcal{S}, j \neq q} |\langle \mathbf{g}_q, \mathbf{g}_j \rangle| \leq \max_{q \in \mathcal{S}} \sum_{j \in \mathcal{S}, j \neq q} |\langle \mathbf{g}_q, \mathbf{g}_j \rangle| \leq \max_{q=1, \dots, N} \sum_{j \in \mathcal{S}, j \neq q} |\langle \mathbf{g}_q, \mathbf{g}_j \rangle| = \mathcal{I}(\mathcal{Y}),$$

and  $\boldsymbol{\rho}$  is supported on  $\mathcal{S}$ , we get

$$\mathcal{T}_L \geq \|\boldsymbol{\rho}\|_1 \left[1 - \mathcal{I}(\mathcal{Y})\right]. \quad (4.15)$$

For  $\mathcal{T}_R$  we have by definition (4.13) of  $\mathbf{u}$  that

$$\mathcal{T}_R = \left| \sum_{q \in \mathcal{S}} \sum_{j \in \mathcal{S}_q} \rho_{\star j}^{(i)} \left[ \text{sign}(\rho_q) \langle \mathbf{g}_j, \mathbf{g}_q \rangle + \sum_{l \in \mathcal{S}, l \neq q} \text{sign}(\rho_l) \langle \mathbf{g}_j, \mathbf{g}_l \rangle \right] + \sum_{j \in \mathcal{S}^c} \sum_{l \in \mathcal{S}} \rho_{\star j}^{(o)} \text{sign}(\rho_l) \langle \mathbf{g}_j, \mathbf{g}_l \rangle \right|$$

and using the triangle inequality

$$\mathcal{T}_R \leq \sum_{q \in \mathcal{S}} \sum_{j \in \mathcal{S}_q} |\rho_{\star j}^{(i)}| |\langle \mathbf{g}_j, \mathbf{g}_q \rangle| + \sum_{q \in \mathcal{S}} \sum_{j \in \mathcal{S}_q} |\rho_{\star j}^{(i)}| \sum_{l \in \mathcal{S}, l \neq q} |\langle \mathbf{g}_j, \mathbf{g}_l \rangle| + \sum_{j \in \mathcal{S}^c} |\rho_{\star j}^{(o)}| \sum_{l \in \mathcal{S}} |\langle \mathbf{g}_j, \mathbf{g}_l \rangle|. \quad (4.16)$$

In the first term in (4.16) we have

$$1 - r < |\langle \mathbf{g}_j, \mathbf{g}_q \rangle| \leq 1,$$

because  $j \in \mathcal{S}_q$ . In the second term

$$\sum_{l \in \mathcal{S}, l \neq q} |\langle \mathbf{g}_j, \mathbf{g}_l \rangle| \leq \mathcal{I}(\mathcal{Y}),$$

by definition (4.4) of the interaction coefficient and the fact that  $\vec{\mathbf{z}}_q$  is the closest source point to  $\vec{\mathbf{z}}_j$ . To bound the third term in (4.16), recall that  $\mathcal{N}(\vec{\mathbf{z}}_j)$  is the closest source point to  $\vec{\mathbf{z}}_j$ , for  $j \in \mathcal{S}^c$ . Its distance from  $\vec{\mathbf{z}}_j$  satisfies  $\mathcal{D}(\vec{\mathbf{z}}_j, \mathcal{N}(\vec{\mathbf{z}}_j)) \geq d$ , by definition of  $\mathcal{S}^c$ , and therefore

$$|\langle \mathbf{g}_j, \mathbf{g}_{\mathcal{N}(\vec{\mathbf{z}}_j)} \rangle| \leq 1 - r.$$

Moreover,

$$\sum_{l \in \mathcal{S}, \vec{\mathbf{z}}_l \neq \mathcal{N}(\vec{\mathbf{z}}_j)} |\langle \mathbf{g}_j, \mathbf{g}_l \rangle| \leq \mathcal{I}(\mathcal{Y}),$$

so in the third sum in (4.16) we have

$$\sum_{l \in \mathcal{S}} |\langle \mathbf{g}_j, \mathbf{g}_l \rangle| \leq \|\boldsymbol{\rho}_\star^{(o)}\|_1 \left[1 - r + \mathcal{I}(\mathcal{Y})\right], \quad \forall j \in \mathcal{S}^c.$$

Thus, the bound on  $\mathcal{T}_R$  becomes

$$\mathcal{T}_R \leq \|\boldsymbol{\rho}_\star^{(i)}\|_1 \left[1 + \mathcal{I}(\mathcal{Y})\right] + \|\boldsymbol{\rho}_\star^{(o)}\|_1 \left[1 - r + \mathcal{I}(\mathcal{Y})\right]. \quad (4.17)$$

To complete the proof use that  $\|\boldsymbol{\rho}_\star\|_1 = \|\boldsymbol{\rho}_\star^{(i)}\|_1 + \|\boldsymbol{\rho}_\star^{(o)}\|_1$  in (4.17) and obtain from equations (4.14) and (4.15) that

$$\|\boldsymbol{\rho}_\star\|_1 \left[1 - \mathcal{I}(\mathcal{Y})\right] \leq \|\boldsymbol{\rho}\|_1 \left[1 - \mathcal{I}(\mathcal{Y})\right] \leq \|\boldsymbol{\rho}_\star\|_1 \left[1 + \mathcal{I}(\mathcal{Y})\right] - r \|\boldsymbol{\rho}_\star^{(o)}\|_1.$$

The first inequality is because  $\boldsymbol{\rho}_\star$  is the  $\ell_1$  minimizer in (2.7). Statement (4.6) follows from this equation.  $\square$

**Proof of Corollary 4.2:** We start with equation (4.12) and take inner product with vector

$$\mathbf{u} = \sum_{q \in \mathcal{S}} \sigma_q \mathbf{g}_q, \quad \sigma_q = \text{sign}(\rho_q - \bar{\rho}_{\star q}). \quad (4.18)$$

We obtain

$$\mathcal{T}_L := \left| \sum_{q \in \mathcal{S}} \rho_q \langle \mathbf{g}_q, \mathbf{u} \rangle - \sum_{q \in \mathcal{S}} \sum_{j \in \mathcal{S}_q} \rho_{\star j}^{(i)} \langle \mathbf{g}_j, \mathbf{u} \rangle \right| = \left| \sum_{j \in \mathcal{S}^c} \rho_{\star j}^{(o)} \langle \mathbf{g}_j, \mathbf{u} \rangle \right| =: \mathcal{T}_R, \quad (4.19)$$

and proceed as in the previous proof by bounding both sides of this equation.

For  $\mathcal{T}_L$  we have

$$\mathcal{T}_L = \left| \sum_{q \in \mathcal{S}} \sigma_q (\rho_q - \bar{\rho}_{*q}) + \sum_{q \in \mathcal{S}} \rho_q \sum_{j \in \mathcal{S}, j \neq q} \sigma_j \langle \mathbf{g}_q, \mathbf{g}_j \rangle - \sum_{q \in \mathcal{S}} \sum_{j \in \mathcal{S}_q} \rho_{*j}^{(i)} \sum_{l \in \mathcal{S}, l \neq q} \sigma_l \langle \mathbf{g}_j, \mathbf{g}_l \rangle \right|,$$

where we used definition (4.8) of the components of  $\bar{\rho}_*$ . We bound it as

$$\begin{aligned} \mathcal{T}_L &\geq \sum_{q \in \mathcal{S}} |\rho_q - \bar{\rho}_{*q}| - \sum_{q \in \mathcal{S}} |\rho_q| \sum_{j \in \mathcal{S}, j \neq q} |\langle \mathbf{g}_q, \mathbf{g}_j \rangle| - \sum_{q \in \mathcal{S}} \sum_{j \in \mathcal{S}_q} |\rho_{*j}^{(i)}| \sum_{l \in \mathcal{S}, l \neq q} |\langle \mathbf{g}_j, \mathbf{g}_l \rangle| \\ &\geq \|\rho - \bar{\rho}_*\|_1 - \left( \|\rho\|_1 + \|\rho_*^{(i)}\|_1 \right) \mathcal{I}(\mathcal{Y}), \end{aligned} \quad (4.20)$$

using the triangle inequality and the definition of  $\sigma_q$  and  $\mathcal{I}(\mathcal{Y})$ .

For  $\mathcal{T}_R$  we have by definition (4.18) of  $\mathbf{u}$  and the triangle inequality that

$$\mathcal{T}_R \leq \sum_{j \in \mathcal{S}^c} |\rho_{*j}^{(o)}| |\langle \mathbf{g}_j, \mathbf{g}_{\mathcal{N}(\bar{\mathbf{z}}_j)} \rangle| + \sum_{j \in \mathcal{S}^c} |\rho_{*j}^{(o)}| \sum_{q \in \mathcal{S}, \bar{\mathbf{z}}_q \neq \mathcal{N}(\bar{\mathbf{z}}_j)} |\langle \mathbf{g}_j, \mathbf{g}_q \rangle|.$$

The right hand side in this equation can be bounded as in the proof of Theorem 4.1, and the result is

$$\mathcal{T}_R \leq \|\rho_*^{(o)}\|_1 \left[ 1 - r + \mathcal{I}(\mathcal{Y}_1) \right]. \quad (4.21)$$

Now equations (4.19)-(4.21) give

$$\begin{aligned} \|\rho - \bar{\rho}_*\|_1 &\leq \left( \|\rho\|_1 + \|\rho_*^{(i)}\|_1 \right) \mathcal{I}(\mathcal{Y}) + \|\rho_*^{(o)}\|_1 \left[ 1 - r + \mathcal{I}(\mathcal{Y}_1) \right] \\ &= \left( \|\rho\|_1 + \|\rho_*\|_1 \right) \mathcal{I}(\mathcal{Y}) + \|\rho_*^{(o)}\|_1 (1 - r) \\ &\leq \left( \|\rho\|_1 + \|\rho_*\|_1 \right) \mathcal{I}(\mathcal{Y}) + \|\rho_*\|_1 \frac{2(1-r)\mathcal{I}(\mathcal{Y})}{r}, \end{aligned}$$

with the last inequality due to Theorem 4.1. Corollary 4.2 follows from this inequality and  $\|\rho_*\|_1 \leq \|\rho\|_1$ .  $\square$

**4.3.2.  $\ell_1$  optimal reconstructions of clusters of sources.** The proof of Theorem 4.4 is a slight modification of that in section 4.3.1. We begin by defining the index map  $J_\varepsilon : \mathcal{S} \rightarrow \mathcal{S}_\varepsilon$  that takes any  $j \in \mathcal{S}$  to  $J_\varepsilon(j)$ , the index of the point in  $\mathcal{S}_\varepsilon$  at the center of the ball containing  $\bar{\mathbf{z}}_j$  i.e.,  $\bar{\mathbf{z}}_j \in \mathcal{B}_\varepsilon(\bar{\mathbf{z}}_{J_\varepsilon(j)})$ . Obviously, the restriction of  $J_\varepsilon$  on  $\mathcal{S} \cap \mathcal{S}_\varepsilon$  is the identity map.

Using the definition of  $\rho_*$  and its decomposition in  $\rho_*^{(i)}$  and  $\rho_*^{(o)}$  we obtain the equivalent of equation (4.12)

$$\sum_{q \in \mathcal{S}} \rho_q \mathbf{g}_q = \sum_{q \in \mathcal{S}_\varepsilon} \sum_{j \in \mathcal{S}_q} \rho_{*j}^{(i)} \mathbf{g}_j + \sum_{j \in \mathcal{S}_\varepsilon^c} \rho_{*j}^{(o)} \mathbf{g}_j, \quad (4.22)$$

where  $\mathcal{S}_\varepsilon^c = \{1, \dots, N\} \setminus \bigcup_{q \in \mathcal{S}_\varepsilon} \mathcal{S}_q$ . We take the inner product of both sides of this equation with vector

$$\mathbf{u} = \sum_{q \in \mathcal{S}_\varepsilon} \sigma_q \mathbf{g}_q, \quad \sigma_q = \text{sign}(\bar{\rho}_q), \quad (4.23)$$

and get

$$\mathcal{T}_L := \left| \sum_{q \in \mathcal{S}} \rho_q \langle \mathbf{g}_q, \mathbf{u} \rangle \right| = \left| \sum_{q \in \mathcal{S}_\varepsilon} \sum_{j \in \mathcal{S}_q} \rho_{*j}^{(i)} \langle \mathbf{g}_j, \mathbf{u} \rangle + \sum_{j \in \mathcal{S}_\varepsilon^c} \rho_{*j}^{(o)} \langle \mathbf{g}_j, \mathbf{u} \rangle \right| =: \mathcal{T}_R, \quad (4.24)$$

where  $\mathcal{T}_L$  and  $\mathcal{T}_R$  denote the left and right hand side of the equation, as before.

For  $\mathcal{T}_L$  we have

$$\begin{aligned}
\mathcal{T}_L &= \left| \sum_{q \in \mathcal{S}_\varepsilon} \sigma_q \sum_{j \in \mathcal{S} \cap \mathcal{B}_\varepsilon(\bar{\mathbf{z}}_q)} \rho_j \langle \mathbf{g}_j, \mathbf{g}_q \rangle + \sum_{q \in \mathcal{S}_\varepsilon} \sigma_q \sum_{j \in \mathcal{S}, \bar{\mathbf{z}}_j \notin \mathcal{B}_\varepsilon(\bar{\mathbf{z}}_q)} \rho_j \langle \mathbf{g}_j, \mathbf{g}_q \rangle \right| \\
&= \left| \sum_{q \in \mathcal{S}_\varepsilon} |\bar{\rho}_q| + \sum_{q \in \mathcal{S}_\varepsilon} \sigma_q \sum_{j \in \mathcal{S}, \bar{\mathbf{z}}_j \notin \mathcal{B}_\varepsilon(\bar{\mathbf{z}}_q)} \rho_j \langle \mathbf{g}_j, \mathbf{g}_q \rangle \right| \\
&\geq \sum_{q \in \mathcal{S}_\varepsilon} |\bar{\rho}_q| - \sum_{j \in \mathcal{S}} |\rho_j| \sum_{q \in \mathcal{S}_\varepsilon, q \neq J_\varepsilon(j)} |\langle \mathbf{g}_j, \mathbf{g}_q \rangle| \\
&\geq \sum_{q \in \mathcal{S}_\varepsilon} |\bar{\rho}_q| - \sum_{j \in \mathcal{S}} |\rho_j| \mathcal{I}(\mathcal{Y}_\varepsilon).
\end{aligned}$$

The equality in the second row is by definition (4.10) of the effective source vector  $\bar{\rho}$  and definition (4.22) of  $\sigma_q$ . The bound in the third row is by the triangle inequality and in the last row by definition of  $\mathcal{I}(\mathcal{Y}_\varepsilon)$ . Thus, the left hand side of (4.24) satisfies

$$\mathcal{T}_L \geq \|\bar{\rho}\|_1 - \|\rho\|_1 \mathcal{I}(\mathcal{Y}_\varepsilon). \quad (4.25)$$

For the right hand side  $\mathcal{T}_R$  we have from the definition of  $\mathbf{u}$  and the triangle inequality

$$\begin{aligned}
\mathcal{T}_R &= \left| \sum_{q \in \mathcal{S}_\varepsilon} \sum_{j \in \mathcal{S}_q} \rho_{\star j}^{(i)} \left[ \sigma_q \langle \mathbf{g}_j, \mathbf{g}_q \rangle + \sum_{l \in \mathcal{S}_\varepsilon, l \neq q} \sigma_l \langle \mathbf{g}_j, \mathbf{g}_l \rangle \right] + \sum_{j \in \mathcal{S}_\varepsilon^c} \rho_{\star j}^{(o)} \sum_{l \in \mathcal{S}_\varepsilon} \sigma_l \langle \mathbf{g}_j, \mathbf{g}_l \rangle \right| \\
&\leq \sum_{q \in \mathcal{S}_\varepsilon} \sum_{j \in \mathcal{S}_q} |\rho_{\star j}^{(i)}| \left[ |\langle \mathbf{g}_j, \mathbf{g}_q \rangle| + \sum_{l \in \mathcal{S}_\varepsilon, l \neq q} |\langle \mathbf{g}_j, \mathbf{g}_l \rangle| \right] + \sum_{j \in \mathcal{S}_\varepsilon^c} |\rho_{\star j}^{(o)}| \sum_{l \in \mathcal{S}_\varepsilon} |\langle \mathbf{g}_j, \mathbf{g}_l \rangle|.
\end{aligned}$$

In the first term we can only say that  $|\langle \mathbf{g}_j, \mathbf{g}_q \rangle| \leq 1$ , because the points indexed by  $\mathcal{S}_q$  are all clustered around  $\bar{\mathbf{z}}_q$ . The sum in the second term is bounded by the interaction coefficient of the set  $\mathcal{Y}_\varepsilon$ , and for the last term we have

$$\begin{aligned}
\sum_{j \in \mathcal{S}_\varepsilon^c} |\rho_{\star j}^{(o)}| \sum_{l \in \mathcal{S}_\varepsilon} |\langle \mathbf{g}_j, \mathbf{g}_l \rangle| &= \sum_{j \in \mathcal{S}_\varepsilon^c} |\rho_{\star j}^{(o)}| \left[ |\langle \mathbf{g}_j, \mathbf{g}_{J_\varepsilon(j)} \rangle| + \sum_{l \in \mathcal{S}_\varepsilon, l \neq J_\varepsilon(j)} |\langle \mathbf{g}_j, \mathbf{g}_l \rangle| \right] \\
&\leq \sum_{j \in \mathcal{S}_\varepsilon^c} |\rho_{\star j}^{(o)}| \left[ 1 - r + \mathcal{I}(\mathcal{Y}_\varepsilon) \right].
\end{aligned}$$

The upper bound on  $\mathcal{T}_R$  becomes

$$\mathcal{T}_R \leq \|\rho_{\star}^{(i)}\|_1 \left[ 1 + \mathcal{I}(\mathcal{Y}_\varepsilon) \right] + \|\rho_{\star}^{(o)}\|_1 \left[ 1 - r + \mathcal{I}(\mathcal{Y}_\varepsilon) \right] = \|\rho_{\star}\|_1 \left[ 1 + \mathcal{I}(\mathcal{Y}_\varepsilon) \right] - r \|\rho_{\star}^{(o)}\|_1,$$

and substituting it in (4.24) and using (4.25), we get after some rearrangement

$$\|\bar{\rho}\|_1 - \|\rho\|_1 + \|\rho\|_1 \left[ 1 - \mathcal{I}(\mathcal{Y}_\varepsilon) \right] \leq \|\rho_{\star}\|_1 \left[ 1 + \mathcal{I}(\mathcal{Y}_\varepsilon) \right] - r \|\rho_{\star}^{(o)}\|_1. \quad (4.26)$$

The statement of Theorem 4.4 follows from this, the assumption that  $\mathcal{I}(\mathcal{Y}_\varepsilon) \leq 1$  and  $\|\rho_{\star}\|_1 \leq \|\rho\|_1$ .  $\square$

**4.3.3.  $\ell_1$  penalty reconstructions of well separated sources.** Before giving the proof of Theorem 4.3, let us introduce some notation. The sources are at points in  $\mathcal{Y} = \{\bar{\mathbf{y}}_q, q = 1, \dots, s\}$  which may be off-grid, and we let  $\mathcal{S}$  be the set of indexes of the grid points in the  $r$ -vicinity of the sources, so that

$$\bar{\mathbf{z}}_j \in \bigcup_{q=1}^s \mathcal{B}_r(\bar{\mathbf{y}}_q), \quad \forall j \in \mathcal{S}. \quad (4.27)$$

The complement of the set  $\mathcal{S}$  is  $\mathcal{S}^c = \{1, \dots, N\} \setminus \mathcal{S}$ . For any vector  $\mathbf{u} \in \mathbb{C}^N$ , we denote by  $\mathbf{u}_{\mathcal{S}}$  its restriction to the set  $\mathcal{S}$ . This is a vector of length  $|\mathcal{S}| < N$ . We also let  $\mathcal{G}_{\mathcal{S}}$  be the  $M \times |\mathcal{S}|$  matrix with  $|\mathcal{S}|$  columns  $\mathbf{g}_j$ , for  $j \in \mathcal{S}$ , and denote by  $\mathcal{P}_{\mathcal{S}}$  the orthogonal projection on the range of  $\mathcal{G}_{\mathcal{S}}$ . It is given by

$$\mathcal{P}_{\mathcal{S}} = \mathcal{G}_{\mathcal{S}} \mathcal{G}_{\mathcal{S}}^\dagger, \quad (4.28)$$

where  $\dagger$  denotes the pseudo-inverse.

The proof of Theorem 4.3 is based on the next two lemmas. The first uses results in convex analysis, specifically the sub-gradient of a convex function, defined in [28]. We need here the sub-gradient of the  $\ell_1$  norm function evaluated at  $\mathbf{u} \in \mathbb{C}^N$ , which is shown in [32] to be any vector in the set

$$\partial\|\mathbf{u}\|_1 = \{\boldsymbol{\xi} \in \mathbb{C}^N \text{ s.t. } \xi_i = \text{sign}(u_i) \text{ if } u_i \neq 0 \text{ and } |\xi_i| \leq 1 \text{ if } u_i = 0\}. \quad (4.29)$$

The second lemma estimates the Lagrange multiplier  $\gamma$  needed to prove the theorem.

LEMMA 4.5. *Let  $\boldsymbol{\rho}_\star$  minimize the augmented Lagrangian  $\mathcal{L}(\boldsymbol{\rho})$  defined in (2.9), over vectors supported in  $\mathcal{S}$ , with  $\boldsymbol{\rho}_{\star\mathcal{S}}$  its restriction to  $\mathcal{S}$ . Then, there exists a sub-gradient vector  $\boldsymbol{\xi} \in \partial\|\boldsymbol{\rho}_{\star\mathcal{S}}\|_1$  such that*

$$\mathcal{G}_{\mathcal{S}}^H \mathcal{G}_{\mathcal{S}} (\boldsymbol{\rho}_{\star\mathcal{S}} - \mathcal{G}_{\mathcal{S}}^\dagger \mathbf{d}) + \gamma \boldsymbol{\xi} = \mathbf{0}, \quad (4.30)$$

where the index  $H$  denotes the Hermitian adjoint of  $\mathcal{G}_{\mathcal{S}}$ , a matrix in  $\mathbb{C}^{|\mathcal{S}| \times N}$ . Moreover, if we let  $\mathcal{S} \subset \mathcal{S}$  be the set of  $s$  grid points that are nearest the locations  $\bar{\mathbf{y}}_j$  of the  $s$  sources, we get

$$\mathcal{G}_s^H \mathcal{G}_{\mathcal{S}} (\boldsymbol{\rho}_{\star\mathcal{S}} - \mathcal{G}_{\mathcal{S}}^\dagger \mathbf{d}) + \gamma \boldsymbol{\xi}_s = \mathbf{0}. \quad (4.31)$$

**Proof:** For any  $\boldsymbol{\rho}$  supported in  $\mathcal{S}$  we have  $\|\boldsymbol{\rho}\|_1 = \|\boldsymbol{\rho}_{\mathcal{S}}\|_1$ , and using Pythagora's theorem

$$\|\mathcal{G}\boldsymbol{\rho} - \mathbf{d}\|_2^2 = \|\mathcal{G}_{\mathcal{S}}\boldsymbol{\rho}_{\mathcal{S}} - \mathcal{P}_{\mathcal{S}}\mathbf{d}\|_2^2 + \|\mathbf{d} - \mathcal{P}_{\mathcal{S}}\mathbf{d}\|_2^2.$$

Therefore  $\mathcal{L}(\boldsymbol{\rho}) = \mathcal{L}_{\mathcal{S}}(\boldsymbol{\rho}_{\mathcal{S}}) + \frac{1}{2}\|\mathbf{d} - \mathcal{P}_{\mathcal{S}}\mathbf{d}\|_2^2$ , with  $\mathcal{L}_{\mathcal{S}}$  defined on vectors of length  $|\mathcal{S}|$ ,

$$\mathcal{L}_{\mathcal{S}}(\mathbf{x}) = \frac{1}{2}\|\mathcal{G}_{\mathcal{S}}\mathbf{x} - \mathcal{P}_{\mathcal{S}}\mathbf{d}\|_2^2 + \gamma\|\mathbf{x}\|_1, \quad \forall \mathbf{x} \in \mathbb{C}^{|\mathcal{S}|}. \quad (4.32)$$

We conclude that  $\boldsymbol{\rho}_{\star\mathcal{S}}$  is the minimizer of  $\mathcal{L}_{\mathcal{S}}$ . Then, results in convex analysis [28, 32] imply that  $\mathbf{0}$  must be an element of the sub-gradient of  $\mathcal{L}_{\mathcal{S}}$ . Equivalently, there exists a vector  $\boldsymbol{\xi} \in \partial\|\boldsymbol{\rho}_{\star\mathcal{S}}\|_1$  satisfying (4.30), where we use the expression of the projection  $\mathcal{P}_{\mathcal{S}}$ . Equation (4.31) is just the restriction of equation (4.30) to the rows indexed by  $\mathcal{S}$ .  $\square$

LEMMA 4.6. *Let  $\boldsymbol{\rho}_\star$  be the minimizer of  $\mathcal{L}(\boldsymbol{\rho})$  over vectors supported in  $\mathcal{S}$ . If  $\gamma$  satisfies*

$$\sqrt{2r}\|\boldsymbol{\rho}_{\star\mathcal{S}} - \mathcal{G}_{\mathcal{S}}^\dagger \mathbf{d}\|_1 + \|\mathcal{G}_{\mathcal{S}^c}^H (\mathcal{P}_{\mathcal{S}}\mathbf{d} - \mathbf{d})\|_\infty < \gamma \left[ 1 - \max_{j \in \mathcal{S}^c} \left| \langle \boldsymbol{\xi}_s, \mathcal{G}_s^\dagger \mathbf{g}_j \rangle \right| \right], \quad (4.33)$$

with  $\boldsymbol{\xi}$  as in Lemma 4.5,  $\boldsymbol{\rho}_\star$  is the global minimizer of  $\mathcal{L}(\boldsymbol{\rho})$  over  $\mathbb{C}^N$ .

**Proof:** To prove the lemma we show that any perturbation of  $\boldsymbol{\rho}_\star$  by a vector that is not supported in  $\mathcal{S}$  leads to an increase of the objective function  $\mathcal{L}$ . This implies that  $\boldsymbol{\rho}_\star$  is a local minimizer of  $\mathcal{L}$  in  $\mathbb{C}^N$ . That  $\boldsymbol{\rho}_\star$  is the global minimizer follows from the convexity of  $\mathcal{L}$ .

Consider an arbitrary vector  $\mathbf{v} \in \mathbb{C}^N$  and decompose it as

$$\mathbf{v} = \mathbf{u} + \mathbf{w}, \quad \text{supp } \mathbf{u} \subset \mathcal{S}, \quad \text{supp } \mathbf{w} \subset \mathcal{S}^c. \quad (4.34)$$

For small and positive  $\varepsilon$  we have from definition (2.9) and the disjoint support of  $\mathbf{u}$  and  $\mathbf{w}$  that

$$\mathcal{L}(\boldsymbol{\rho}_\star + \varepsilon \mathbf{v}) - \mathcal{L}(\boldsymbol{\rho}_\star + \varepsilon \mathbf{u}) = \varepsilon \left[ \text{real} \left( \langle \mathcal{G}\boldsymbol{\rho}_\star - \mathbf{d}, \mathcal{G}\mathbf{w} \rangle \right) + \gamma\|\mathbf{w}\|_1 \right] + \varepsilon^2 \left[ \frac{1}{2}\|\mathcal{G}\mathbf{w}\|_2^2 + \text{real} \left( \langle \mathcal{G}\mathbf{u}, \mathcal{G}\mathbf{w} \rangle \right) \right], \quad (4.35)$$

with the first term dominating the second for  $\varepsilon \ll 1$ . We write it in terms of the components  $w_j$  of  $\mathbf{w}$  as

$$\langle \mathcal{G}\boldsymbol{\rho}_\star - \mathbf{d}, \mathcal{G}\mathbf{w} \rangle = \sum_{j \in \mathcal{S}^c} \langle \mathcal{G}_{\mathcal{S}}\boldsymbol{\rho}_{\star\mathcal{S}} - \mathbf{d}, \mathbf{g}_j \rangle w_j = \sum_{j \in \mathcal{S}^c} \langle \mathcal{G}_{\mathcal{S}}\boldsymbol{\rho}_{\star\mathcal{S}} - \mathcal{P}_{\mathcal{S}}\mathbf{d}, \mathbf{g}_j \rangle w_j - \sum_{j \in \mathcal{S}^c} \langle \mathbf{d} - \mathcal{P}_{\mathcal{S}}\mathbf{d}, \mathbf{g}_j \rangle w_j, \quad (4.36)$$

where we used that  $\boldsymbol{\rho}_\star$  is supported in  $\mathcal{S}$ . We estimate next the two sums in the right hand side.



For the terms in the first sum we have

$$\langle \mathcal{G}_{\mathcal{S}} \boldsymbol{\rho}_{\star_{\mathcal{S}}} - \mathcal{P}_{\mathcal{S}} \mathbf{d}, \mathbf{g}_j \rangle = \langle \mathcal{G}_{\mathcal{S}} \boldsymbol{\rho}_{\star_{\mathcal{S}}} - \mathcal{P}_{\mathcal{S}} \mathbf{d}, \mathcal{P}_S \mathbf{g}_j \rangle + \langle \mathcal{G}_{\mathcal{S}} \boldsymbol{\rho}_{\star_{\mathcal{S}}} - \mathcal{P}_{\mathcal{S}} \mathbf{d}, (I - \mathcal{P}_S) \mathbf{g}_j \rangle, \quad (4.37)$$

where  $\mathcal{P}_S = \mathcal{G}_S \mathcal{G}_S^\dagger$  is the orthogonal projection on the range of the full rank <sup>‡</sup> matrix  $\mathcal{G}_S$ , with pseudo-inverse

$$\mathcal{G}_S^\dagger = \left( \mathcal{G}_S^H \mathcal{G}_S \right)^{-1} \mathcal{G}_S^H. \quad (4.38)$$

Substituting the expression of  $\mathcal{P}_S$  and  $\mathcal{P}_{\mathcal{S}}$  in (4.37) we get

$$\begin{aligned} \langle \mathcal{G}_{\mathcal{S}} \boldsymbol{\rho}_{\star_{\mathcal{S}}} - \mathcal{P}_{\mathcal{S}} \mathbf{d}, \mathbf{g}_j \rangle &= \langle \mathcal{G}_S^H \mathcal{G}_{\mathcal{S}} (\boldsymbol{\rho}_{\star_{\mathcal{S}}} - \mathcal{G}_{\mathcal{S}}^\dagger \mathbf{d}), \mathcal{G}_S^\dagger \mathbf{g}_j \rangle + \left\langle \mathcal{G}_{\mathcal{S}} (\boldsymbol{\rho}_{\star_{\mathcal{S}}} - \mathcal{G}_{\mathcal{S}}^\dagger \mathbf{d}), (I - \mathcal{P}_S) \mathbf{g}_j \right\rangle \\ &= -\gamma \langle \boldsymbol{\xi}_S, \mathcal{G}_S^\dagger \mathbf{g}_j \rangle + \langle \mathcal{G}_{\mathcal{S}} \boldsymbol{\alpha}, (I - \mathcal{P}_S) \mathbf{g}_j \rangle, \end{aligned} \quad (4.39)$$

with the second inequality following from Lemma 4.5, and notation  $\boldsymbol{\alpha} = \boldsymbol{\rho}_{\star_{\mathcal{S}}} - \mathcal{G}_{\mathcal{S}}^\dagger \mathbf{d}$ . Let us write explicitly the second term in (4.39)

$$\langle \mathcal{G}_{\mathcal{S}} \boldsymbol{\alpha}, (I - \mathcal{P}_S) \mathbf{g}_j \rangle = \sum_{l \in \mathcal{S}} \langle \mathbf{g}_l, (I - \mathcal{P}_S) \mathbf{g}_j \rangle \alpha_l, \quad \text{for } j \in \mathcal{S}^c. \quad (4.40)$$

Since  $l \in \mathcal{S}$ , we have by definition (4.27) that there exists  $q \in \mathcal{S}$  such that  $\bar{\mathbf{z}}_l \in \mathcal{B}_r(\bar{\mathbf{y}}_q)$ , and we can decompose  $\mathbf{g}_l$  in two parts:  $\mathbf{g}_l^\parallel$  which is along  $\mathbf{g}_q$  and  $\mathbf{g}_l^\perp$  which is orthogonal to it,

$$\mathbf{g}_l = \mathbf{g}_l^\parallel + \mathbf{g}_l^\perp, \quad \mathbf{g}_l^\parallel = \langle \mathbf{g}_q, \mathbf{g}_l \rangle \mathbf{g}_q, \quad \mathbf{g}_l^\perp = \mathbf{g}_l - \mathbf{g}_l^\parallel.$$

Clearly  $\mathbf{g}_l^\parallel$  is in the range of  $\mathcal{G}_S$ , so it is orthogonal to  $(I - \mathcal{P}_S) \mathbf{g}_j$ , and the terms in equation (4.40) satisfy

$$|\langle \mathbf{g}_l, (I - \mathcal{P}_S) \mathbf{g}_j \rangle| = |\langle \mathbf{g}_l^\perp, (I - \mathcal{P}_S) \mathbf{g}_j \rangle| \leq \|\mathbf{g}_l^\perp\|_2 \|(I - \mathcal{P}_S) \mathbf{g}_j\|_2 \leq \|\mathbf{g}_l^\perp\|_2. \quad (4.41)$$

Moreover, by Pythagora's theorem

$$\|\mathbf{g}_l^\perp\|_2^2 = 1 - \|\mathbf{g}_l^\parallel\|_2^2 = 1 - |\langle \mathbf{g}_q, \mathbf{g}_l \rangle|^2 < 1 - (1 - r)^2 = 2r - r^2 < 2r, \quad (4.42)$$

where the first inequality follows  $\bar{\mathbf{z}}_l \in \mathcal{B}_r(\bar{\mathbf{y}}_q)$  i.e.,  $\mathcal{D}(\bar{\mathbf{z}}_l, \bar{\mathbf{z}}_q) = 1 - |\langle \mathbf{g}_q, \mathbf{g}_l \rangle| < r$ . Gathering the results (4.39)-(4.42) and using the triangle inequality we get the following bound on the first sum in (4.36)

$$\left| \sum_{j \in \mathcal{S}^c} \langle \mathcal{G}_{\mathcal{S}} \boldsymbol{\rho}_{\star_{\mathcal{S}}} - \mathcal{P}_{\mathcal{S}} \mathbf{d}, \mathbf{g}_j \rangle w_j \right| \leq \left[ \gamma \max_{j \in \mathcal{S}^c} |\langle \boldsymbol{\xi}_S, \mathcal{G}_S^\dagger \mathbf{g}_j \rangle| + \sqrt{2r} \|\boldsymbol{\rho}_{\star_{\mathcal{S}}} - \mathcal{G}_{\mathcal{S}}^\dagger \mathbf{d}\|_1 \right] \|\mathbf{w}\|_1. \quad (4.43)$$

For the second sum in (4.36) we have

$$\left| \sum_{j \in \mathcal{S}^c} \langle \mathbf{d} - \mathcal{P}_{\mathcal{S}} \mathbf{d}, \mathbf{g}_j \rangle w_j \right| \leq \max_{l \in \mathcal{S}^c} |\langle \mathbf{d} - \mathcal{P}_{\mathcal{S}} \mathbf{d}, \mathbf{g}_l \rangle| \sum_{j \in \mathcal{S}^c} |w_j| = \|\mathcal{G}_{\mathcal{S}^c}^H (\mathbf{d} - \mathcal{P}_{\mathcal{S}} \mathbf{d})\|_\infty \|\mathbf{w}\|_1, \quad (4.44)$$

and putting together the results (4.43)-(4.44) we obtain from (4.36) that

$$|\langle \mathcal{G} \boldsymbol{\rho}_\star - \mathbf{d}, \mathcal{G} \mathbf{w} \rangle| \leq \left[ \gamma \max_{j \in \mathcal{S}^c} |\langle \boldsymbol{\xi}_S, \mathcal{G}_S^\dagger \mathbf{g}_j \rangle| + \sqrt{2r} \|\boldsymbol{\rho}_{\star_{\mathcal{S}}} - \mathcal{G}_{\mathcal{S}}^\dagger \mathbf{d}\|_1 + \|\mathcal{G}_{\mathcal{S}^c}^H (\mathbf{d} - \mathcal{P}_{\mathcal{S}} \mathbf{d})\|_\infty \right] \|\mathbf{w}\|_1. \quad (4.45)$$

This estimate and the triangle inequality give that the  $\varepsilon$  term in (4.35) is positive for  $\gamma$  satisfying

$$\gamma > \gamma \max_{j \in \mathcal{S}^c} |\langle \boldsymbol{\xi}_S, \mathcal{G}_S^\dagger \mathbf{g}_j \rangle| + \sqrt{2r} \|\boldsymbol{\rho}_{\star_{\mathcal{S}}} - \mathcal{G}_{\mathcal{S}}^\dagger \mathbf{d}\|_1 + \|\mathcal{G}_{\mathcal{S}^c}^H (\mathbf{d} - \mathcal{P}_{\mathcal{S}} \mathbf{d})\|_\infty,$$

<sup>‡</sup>That  $\mathcal{G}_S$  is full rank follows from the assumption  $\mathcal{I}(\mathcal{Y}) < 1/2$ . Matrix  $\mathcal{G}_{\mathcal{S}}$  is not full rank because its columns are associated to nearby points in the vicinity of the sources, as stated in (4.27).

as assumed in the lemma. Then, the perturbation of  $\boldsymbol{\rho}_\star$  by the arbitrary vector (4.34) increases the objective function  $\mathcal{L}$  and the lemma follows.  $\square$

To complete the proof of the theorem it remains to show that we can find a positive  $\gamma$  as in Lemma 4.6. We begin with the estimate

$$\begin{aligned} \max_{j \in \mathcal{S}^c} |\langle \boldsymbol{\xi}_s, \mathcal{G}_s^\dagger \mathbf{g}_j \rangle| &= \max_{j \in \mathcal{S}^c} \left| \left\langle \boldsymbol{\xi}_s, (\mathcal{G}_s^H \mathcal{G}_s)^{-1} \mathcal{G}_s^H \mathbf{g}_j \right\rangle \right| \\ &\leq \max_{j \in \mathcal{S}^c} \|\boldsymbol{\xi}_s\|_\infty \|(\mathcal{G}_s^H \mathcal{G}_s)^{-1} \mathcal{G}_s^H \mathbf{g}_j\|_1 \\ &\leq \max_{j \in \mathcal{S}^c} \|(\mathcal{G}_s^H \mathcal{G}_s)^{-1}\|_{1,1} \|\mathcal{G}_s^H \mathbf{g}_j\|_1, \end{aligned} \quad (4.46)$$

where we used (4.38) and definition (4.29) of the sub-gradient of the  $\ell_1$  norm. Now since  $j \in \mathcal{S}^c$ ,  $\vec{\mathbf{z}}_j$  is outside every ball of radius  $r$  centered at a source or, equivalently,  $|\langle \mathbf{g}_q, \mathbf{g}_j \rangle| \leq 1 - r$ , for all  $q \in \mathcal{S}$ . This and the definition of  $\mathcal{I}(\mathcal{Y})$  give

$$\|\mathcal{G}_s^H \mathbf{g}_j\|_1 = \sum_{q \in \mathcal{S}} |\langle \mathbf{g}_q, \mathbf{g}_j \rangle| \leq 1 - r + \mathcal{I}(\mathcal{Y}), \quad \forall j \in \mathcal{S}^c. \quad (4.47)$$

Moreover, for any vector  $\mathbf{u}$  supported on  $\mathcal{S}$  we have

$$\begin{aligned} \|\mathcal{G}_s^H \mathcal{G}_s \mathbf{u}\|_1 &= \sum_{q \in \mathcal{S}} \left| \sum_{j \in \mathcal{S}} \langle \mathbf{g}_q, \mathbf{g}_j \rangle u_j \right| \\ &= \sum_{q \in \mathcal{S}} \left| u_q + \sum_{j \in \mathcal{S}, j \neq q} \langle \mathbf{g}_q, \mathbf{g}_j \rangle u_j \right| \\ &\geq \sum_{q \in \mathcal{S}} |u_q| \left[ 1 - \sum_{j \in \mathcal{S}, j \neq q} |\langle \mathbf{g}_q, \mathbf{g}_j \rangle| \right] \\ &\geq [1 - \mathcal{I}(\mathcal{Y})] \|\mathbf{u}\|_1, \end{aligned}$$

so the operator norm of the inverse of  $\mathcal{G}_s^H \mathcal{G}_s$ , which exists because  $\mathcal{G}_s$  is full rank, satisfies

$$\|(\mathcal{G}_s^H \mathcal{G}_s)^{-1}\|_{1,1} \leq [1 - \mathcal{I}(\mathcal{Y})]^{-1}. \quad (4.48)$$

Putting together (4.46)-(4.48) we obtain that

$$\max_{j \in \mathcal{S}^c} |\langle \boldsymbol{\xi}_s, \mathcal{G}_s^\dagger \mathbf{g}_j \rangle| \leq \frac{1 - r + \mathcal{I}(\mathcal{Y})}{1 - \mathcal{I}(\mathcal{Y})} < 1, \quad (4.49)$$

where the second inequality is by the assumption of the theorem that  $r > 2\mathcal{I}(\mathcal{Y})$ . This shows that the right hand side in equation (4.33) in Lemma 4.6 is positive.

Finally, we show that the left hand side in equation (4.33) is bounded independent of  $\gamma$ . Clearly, the term  $\|\mathcal{G}_{\mathcal{S}^c}^H (\mathcal{P}_{\mathcal{S}} \mathbf{d} - \mathbf{d})\|_\infty$  does not depend on  $\gamma$ . It is due to the modeling error that is small when the grid is fine enough and the additive noise, that may cause  $\mathbf{d}$  to lie outside the range of  $\mathcal{G}_{\mathcal{S}}$  i.e.,  $\mathbf{d} \neq \mathcal{P}_{\mathcal{S}} \mathbf{d}$ . To bound the first term in (4.33) note that  $\mathbf{u} = \mathcal{G}_{\mathcal{S}}^\dagger \mathbf{d}$  is the minimum  $\ell_2$  norm solution of  $\mathcal{G}_{\mathcal{S}} \mathbf{u} = \mathbf{d}$  if it exists, or otherwise the minimizer of the least squares misfit  $\|\mathcal{G}_{\mathcal{S}} \mathbf{u} - \mathbf{d}\|_2$ . Then, since  $\boldsymbol{\rho}_{\star, \mathcal{S}}$  minimizes  $\mathcal{L}_{\mathcal{S}}$ , we have

$$\mathcal{L}_{\mathcal{S}}(\boldsymbol{\rho}_{\star, \mathcal{S}}) = \frac{1}{2} \|\mathcal{G}_{\mathcal{S}} \boldsymbol{\rho}_{\star, \mathcal{S}} - \mathbf{d}\|_2^2 + \gamma \|\boldsymbol{\rho}_{\star, \mathcal{S}}\|_1 \leq \mathcal{L}_{\mathcal{S}}(\mathbf{u}) = \frac{1}{2} \|\mathcal{G}_{\mathcal{S}} \mathbf{u} - \mathbf{d}\|_2^2 + \gamma \|\mathbf{u}\|_1,$$

and using that  $\|\mathcal{G}_{\mathcal{S}} \mathbf{u} - \mathbf{d}\|_2 \leq \|\mathcal{G}_{\mathcal{S}} \boldsymbol{\rho}_{\star, \mathcal{S}} - \mathbf{d}\|_2$ , we get  $\|\boldsymbol{\rho}_{\star, \mathcal{S}}\|_1 \leq \|\mathbf{u}\|_1$ . Consequently,

$$\|\boldsymbol{\rho}_{\star, \mathcal{S}} - \mathcal{G}_{\mathcal{S}}^\dagger \mathbf{d}\|_1 = \|\boldsymbol{\rho}_{\star, \mathcal{S}} - \mathbf{u}\|_1 \leq \|\boldsymbol{\rho}_{\star, \mathcal{S}}\|_1 + \|\mathbf{u}\|_1 = 2\|\mathbf{u}\|_1, \quad (4.50)$$

independent of  $\gamma$ . We conclude that we can find  $\gamma > 0$  as in Lemma 4.6, and therefore complete the proof of the theorem.  $\square$

**5. Summary.** We presented a resolution study of sensor array imaging of a sparse scene of point sources or scatterers. The setup is in the paraxial regime, where the array aperture is small with respect to the distance to the imaging region. The imaging is done with two sparsity promoting optimization methods:  $\ell_1$  optimization (basis pursuit) and  $\ell_1$ -penalty. The latter deals with noise and modeling errors. Our resolution analysis takes into account the sparse support of the unknowns. In case that they lie on the imaging grid, we obtained conditions on the grid size that guarantee their exact and unique recovery for noiseless data. This is for both single frequency and broad-band regimes, and the results show the benefit of having multiple frequency measurements. In case that the unknowns lie off-grid, we studied imaging on fine grids that mitigate the modeling error. We showed that when the unknowns are located at sufficiently far apart points in the scene, or they lie in well separated clusters, the results of imaging with sparsity promoting optimization are useful. The support of the reconstruction is near that of the unknowns and its locally averaged amplitudes approximates the true ones.

**Acknowledgments.** This work was partially supported by AFOSR Grant FA9550-15-1-0118. LB also acknowledges support from the ONR Grant N00014-14-1-0077.

**Appendix A. Numerical setup.** The simulations are for an aperture  $a = 25\lambda$ , range  $L = 1000\lambda$ . We used different sizes of imaging grids such as  $W_N = 10 \times 10 \times 20$  or  $W_N = 5 \times 64 \times 64$ , for a given mesh size  $\vec{\mathbf{h}} = (h, h, h_3)$ .

The  $\ell_1$  optimization is solved with the package [16]. For noisy data we solve (2.9) with  $\gamma$  chosen to be close to the noise level. We find the results to be very similar to those obtained from the constrained optimization

$$\min_{\boldsymbol{\rho} \in \mathbb{C}^N} \|\boldsymbol{\rho}\|_1 \quad \text{s.t.} \quad \|\mathcal{G}\boldsymbol{\rho} - \mathbf{d}\|_2 \leq \text{noise level.}$$

Because the simulations give only an approximation  $\tilde{\boldsymbol{\rho}}_*$  of the minimizer  $\boldsymbol{\rho}_*$ , we threshold the results at 1% of the maximum entry in absolute value. We say that  $\boldsymbol{\rho}$  is recovered numerically if  $\|\tilde{\boldsymbol{\rho}}_* - \boldsymbol{\rho}\|_\infty / \|\boldsymbol{\rho}\|_\infty < 1\%$ .

**Appendix B. Derivation of the paraxial model.** We have for  $\vec{\mathbf{x}} = (\mathbf{x}, 0)$  in the array and  $\vec{\mathbf{z}} = (\mathbf{z}, z_3)$  in the imaging region that

$$|\vec{\mathbf{x}} - \vec{\mathbf{y}}| = L \left[ 1 + O\left(\frac{D_3}{L}\right) + O\left(\frac{a^2}{L^2}\right) \right] \quad (\text{B.1})$$

so we can approximate the geometrical spreading factor in the Green's function by

$$\frac{1}{4\pi|\vec{\mathbf{x}} - \vec{\mathbf{y}}|} \approx \frac{1}{4\pi L}. \quad (\text{B.2})$$

The phase is given by

$$k|\vec{\mathbf{x}} - \vec{\mathbf{y}}| = kz_3 + \frac{k|\mathbf{x} - \mathbf{z}|^2}{2z_3} + \mathcal{E}(\vec{\mathbf{x}}, \vec{\mathbf{y}}), \quad (\text{B.3})$$

with remainder

$$\mathcal{E}(\vec{\mathbf{x}}, \vec{\mathbf{y}}) = -\frac{k|\mathbf{x} - \mathbf{z}|^4}{8z_3^3} + O\left(\frac{k|\mathbf{x} - \mathbf{z}|^6}{z_3^5}\right).$$

The scaling assumptions (3.3) make most of the terms in  $\mathcal{E}$  negligible, except for those that are independent of  $\vec{\mathbf{z}}$ , which cancel in the product of the Green's functions in the left hand side of equation (3.5). Thus we write

$$\mathcal{E}(\vec{\mathbf{x}}, \vec{\mathbf{y}}) \approx -\frac{k|\mathbf{x}|^4}{8L^3} + O\left(\frac{k|\mathbf{x}|^6}{L^5}\right). \quad (\text{B.4})$$

We also have

$$\frac{k|\mathbf{x} - \mathbf{z}|^2}{2z_3} = \frac{k|\mathbf{x}|^2}{2z_3} - \frac{k\mathbf{x} \cdot \mathbf{z}}{z_3} + \frac{k|\mathbf{z}|^2}{2z_3},$$

with the last term negligible by assumption (3.3). Simplifying further, we get

$$\frac{k|\mathbf{x} - \mathbf{z}|^2}{2z_3} \approx \frac{k|\mathbf{x}|^2}{2L} - \frac{k|\mathbf{x}|^2 z_3}{2L^2} - \frac{k\mathbf{x} \cdot \mathbf{z}}{L} + O\left(\frac{a^2 D_3^2}{\lambda L L^2}\right) + O\left(\frac{aDD_3}{\lambda L^2}\right), \quad (\text{B.5})$$

with the remainder negligible by (3.3).

Now let  $\bar{\mathbf{x}} = \bar{\mathbf{x}}_r$  and obtain from (B.2)-(B.5) that

$$\hat{G}(\omega, \bar{\mathbf{x}}_r, \bar{\mathbf{z}}_j) \overline{\hat{G}(\omega, \bar{\mathbf{x}}_r, \bar{\mathbf{z}}_q)} \approx \frac{e^{ik(z_{3,j} - z_{3,q})}}{(4\pi L)^2} e^{-\frac{ik|\mathbf{x}_r|^2(z_{3,j} - z_{3,q})}{2L^2} - \frac{ik\mathbf{x}_r \cdot (\mathbf{z}_j - \mathbf{z}_q)}{L}}.$$

Equation (3.5) follows by summing over  $r$ .  $\square$

**Appendix C. Derivation of the broad-band paraxial model.** Using the parabolic scaling, we find as in appendix B that

$$\hat{G}(\omega_j, \bar{\mathbf{x}}_r, \bar{\mathbf{z}}_q) \overline{\hat{G}(\omega_j, \bar{\mathbf{x}}_r, \bar{\mathbf{z}}_l)} \approx \frac{e^{ik_j(z_{3,q} - z_{3,l})}}{(4\pi L)^2} e^{-\frac{ik_j|\mathbf{x}_r|^2(z_{3,q} - z_{3,l})}{2L^2} - \frac{ik_j\mathbf{x}_r \cdot (\mathbf{z}_q - \mathbf{z}_l)}{L}}, \quad (\text{C.1})$$

where  $k_j = \omega_j/c$ . Moreover, the phase  $\Phi$  in the right hand side of (C.1) is

$$\Phi = k_j(z_{3,q} - z_{3,l}) - \frac{k_o|\mathbf{x}_r|^2(z_{3,q} - z_{3,l})}{2L^2} - \frac{k_o\mathbf{x}_r \cdot (\mathbf{z}_q - \mathbf{z}_l)}{L} + O\left(\frac{B aD}{c L}\right) + O\left(\frac{B a^2 D_3}{c L^2}\right),$$

with negligible remainder by assumption (3.15). Result (3.17) follows after substituting the approximation of  $\Phi$  in (C.1), multiplying with the Gaussian pulse, and summing over the frequencies and the receivers. We also remove the phase  $k_o(z_{3,q} - z_{3,l})$  in equation (3.17).  $\square$

**Appendix D. Proof of estimate 3.25.** Simplifying notation as

$$r_o = \frac{\beta - \eta}{2\sqrt{\eta}} > \frac{\alpha}{2\sqrt{\eta}} > 0, \quad r_1 = \frac{\beta + \eta}{2\sqrt{\eta}} = r_o + \sqrt{\eta},$$

and choosing a contour defined by the line segments at angle 0 and  $\pi/4$  with the real axis, and the circular arcs at radius  $r_o$  and  $r_1$ , we obtain

$$\int_{r_o}^{r_1} dt e^{it^2} = -\int_0^{\pi/4} d\theta \left[ r_1 e^{-r_1^2 \sin(2\theta) + i[\theta + r_1^2 \cos(2\theta)]} - r_o e^{-r_o^2 \sin(2\theta) + i[\theta + r_o^2 \cos(2\theta)]} \right] + \sqrt{i} \int_{r_o}^{r_1} dr e^{-r^2}. \quad (\text{D.1})$$

For the last term we have the estimate

$$\int_{r_o}^{r_1} dr e^{-r^2} = \int_0^{\sqrt{\eta}} ds e^{-(r_o + s)^2} \leq e^{-r_o^2} \int_0^\infty ds e^{-s^2} = \frac{\sqrt{\pi}}{2} e^{-r_o^2}, \quad (\text{D.2})$$

where we changed variables as  $r = r_o + s$ . Moreover,

$$\left| \int_0^{\pi/4} d\theta r_o e^{-r_o^2 \sin(2\theta) + i[\theta + r_o^2 \cos(2\theta)]} \right| \leq \int_0^{\pi/4} d\theta r_o e^{-4r_o^2 \theta/\pi} = \frac{\pi}{4r_o} (1 - e^{-r_o^2}), \quad (\text{D.3})$$

and similar for the other integral over  $\theta$ , because  $\sin(2\theta) \geq 4\theta/\pi$ , for all  $\theta \in (0, \pi/4)$ . The estimate (3.25) follows from (D.1)-(D.3) and the triangle inequality

$$\begin{aligned} \left| \int_{r_o}^{r_1} dt e^{it^2} \right| &\leq \frac{\sqrt{\pi}}{2} e^{-r_o^2} + \frac{\pi}{4r_1} (1 - e^{-r_1^2}) + \frac{\pi}{4r_o} (1 - e^{-r_o^2}) \\ &\leq \frac{\pi}{2r_o} + \frac{\sqrt{\pi}}{2} e^{-r_o^2} < \frac{(\pi + 1)\sqrt{\eta}}{\alpha}. \end{aligned}$$

The second inequality is because  $r_1 > r_o$ , the third inequality is by the definition of  $r_o$  and the last inequality is because  $e^{-x^2/4} < 2/(\sqrt{\pi}x)$  for any  $x > 0$ .  $\square$

## REFERENCES

- [1] A. B. BAGGEROER, W. A. KUPERMAN, AND P. N. MIKHALEVSKY, *An overview of matched field methods in ocean acoustics*, Oceanic Engineering, IEEE Journal of, 18 (1993), pp. 401–424.
- [2] B. BIONDI, *3D seismic imaging*, Society of Exploration Geophysicists, 2006.
- [3] N. BLEISTEIN, J. K. COHEN, AND J.J.W. STOCKWELL, *Mathematics of multidimensional seismic imaging, migration, and inversion*, vol. 13, Springer, 2001.
- [4] L. BORCEA, T. CALLAGHAN, J. GARNIER, AND G. PAPANICOLAOU, *A universal filter for enhanced imaging with small arrays*, Inverse Problems, 26 (2010), p. 015006.
- [5] L. BORCEA, G. PAPANICOLAOU, AND F. GUEVARA-VASQUEZ, *Edge illumination and imaging of extended reflectors*, SIAM Journal on Imaging Sciences, 1 (2008), pp. 75–114.
- [6] L. BORCEA, G. PAPANICOLAOU, AND C. TSOGKA, *Optimal illumination and wave form design for imaging in random media*, The Journal of the Acoustical Society of America, 122 (2007), pp. 3507–3518.
- [7] M. BORN AND E. WOLF, *Principles of optics: electromagnetic theory of propagation, interference and diffraction of light*, Cambridge university press, 1999.
- [8] A. M. BRUCKSTEIN, D. L. DONOHO, AND M. ELAD, *From sparse solutions of systems of equations to sparse modeling of signals and images*, SIAM review, 51 (2009), pp. 34–81.
- [9] F. CAKONI AND D. COLTON, *On the mathematical basis of the linear sampling method*, Georgian Mathematical Journal, 10 (2003), pp. 411–425.
- [10] E. J. CANDÉS AND T. TAO, *Decoding by linear programming*, Information Theory, IEEE Transactions on, 51 (2005), pp. 4203–4215.
- [11] A. CHAI, M. MOSCOSO, AND G. PAPANICOLAOU, *Robust imaging of localized scatterers using the singular value decomposition and  $l_1$  minimization*, Inverse Problems, 29 (2013), p. 025016.
- [12] ———, *Imaging strong localized scatterers with sparsity promoting optimization*, SIAM Journal on Imaging Sciences, 7 (2014), pp. 1358–1387.
- [13] Y. CHI, L. L. SCHARF, A. PEZESHKI, AND A. R. CALDERBANK, *Sensitivity to basis mismatch in compressed sensing*, Signal Processing, IEEE Transactions on, 59 (2011), pp. 2182–2195.
- [14] J. F. CLAERBOUT, *Fundamentals of geophysical data processing : with applications to petroleum prospecting*, CA : Blackwell Scientific Publications, Palo Alto, 1985.
- [15] S. F. COTTER, B. D. RAO, K. ENGAN, AND K. KREUTZ-DELGADO, *Sparse solutions to linear inverse problems with multiple measurement vectors*, Signal Processing, IEEE Transactions on, 53 (2005), pp. 2477–2488.
- [16] CVX RESEARCH, *Cvx: matlab software for disciplined convex programming, version 2.0*. <http://cvxr.com/cvx><http://cvxr.com/cvx>, August 2012.
- [17] D. L. DONOHO AND M. ELAD, *Maximal sparsity representation via  $l_1$  minimization*, Proc. Natl. Acad. Sci., 100 (2003), pp. 2197–2202.
- [18] A. FANNJIANG AND H-C TSENG, *Compressive radar with off-grid targets: a perturbation approach*, Inverse Problems, 29 (2013), p. 054008.
- [19] A. C. FANNJIANG, *Compressive inverse scattering: I. high-frequency simo/miso and mimo measurements*, Inverse Problems, 26 (2010), p. 035008.
- [20] A. C. FANNJIANG AND W. LIAO, *Coherence pattern-guided compressive sensing with unresolved grids*, SIAM Journal on Imaging Sciences, 5 (2012), pp. 179–202.
- [21] A. C. FANNJIANG, T. STROHMER, AND P. YAN, *Compressed remote sensing of sparse objects*, SIAM Journal on Imaging Sciences, 3 (2010), pp. 595–618.
- [22] L.L. FOLDY, *The multiple scattering of waves. i. general theory of isotropic scattering by randomly distributed scatterers*, Physical Review, 67 (1945), p. 107.
- [23] F. K. GRUBER, E. A. MARENGO, AND A. J. DEVANEY, *Time-reversal imaging with multiple signal classification considering multiple scattering between the targets*, The Journal of the Acoustical Society of America, 115 (2004), pp. 3042–3047.
- [24] M. A. HERMAN AND T. STROHMER, *General deviants: An analysis of perturbations in compressed sensing*, Selected Topics in Signal Processing, IEEE Journal of, 4 (2010), pp. 342–349.
- [25] A. KIRSCH, *Characterization of the shape of a scattering obstacle using the spectral data of the far field operator*, Inverse problems, 14 (1998), p. 1489.
- [26] M. LAX, *Multiple scattering of waves*, Reviews of Modern Physics, 23 (1951), p. 287.
- [27] C. PRADA, S. MANNEVILLE, D. SPOLIANSKY, AND M. FINK, *Decomposition of the time reversal operator: Detection and selective focusing on two scatterers*, The Journal of the Acoustical Society of America, 99 (1996), pp. 2067–2076.
- [28] RALPH TYRELL ROCKAFELLAR, *Convex analysis*, Princeton university press, 2015.
- [29] M. RUDELSON AND R. VERSHYNIN, *On sparse reconstruction from fourier and gaussian measurements*, Communications on Pure and Applied Mathematics, 61 (2008), pp. 1025–1045.
- [30] R. O. SCHMIDT, *Multiple emitter location and signal parameter estimation*, Antennas and Propagation, IEEE Transactions on, 34 (1986), pp. 276–280.
- [31] J. A. TROPP, *Greed is good: Algorithmic results for sparse approximation*, Information Theory, IEEE Transactions on, 50 (2004), pp. 2231–2242.
- [32] ———, *Just relax: Convex programming methods for identifying sparse signals in noise*, Information Theory, IEEE Transactions on, 52 (2006), pp. 1030–1051.