

## Imaging strong localized scatterers with sparsity promoting optimization

Anwei Chai<sup>†¶</sup>, Miguel Moscoso<sup>‡||\*\*</sup>, and George Papanicolaou<sup>§¶††</sup>

**Abstract.** We study active array imaging of small but strong scatterers in homogeneous media when multiple scattering between them is important. We use the Foldy-Lax equations to model wave propagation with multiple scattering when the scatterers are small relative to the wavelength. In active array imaging we seek to locate the positions and reflectivities of the scatterers, that is, to determine the support of the reflectivity vector and the values of its nonzero elements from echoes recorded on the array. This is a nonlinear inverse problem because of the multiple scattering. We show in this paper how to avoid the nonlinearity and form images non-iteratively through a two-step process which involves  $\ell_1$  norm minimization. However, under certain illuminations imaging may be affected by screening, where some scatterers are obscured by multiple scattering. This problem can be mitigated by using multiple and diverse illuminations. In this case, we determine solution vectors that have a common support. The uniqueness and stability of the support of the reflectivity vector obtained with single or multiple illuminations are analyzed, showing that the errors are proportional to the amount of noise in the data with a proportionality factor dependent on the sparsity of the solution and the mutual coherence of the sensing matrix, which is determined by the geometry of the imaging array. Finally, to filter out noise and improve the resolution of the images, we propose an approach that combines optimal illuminations using the singular value decomposition of the response matrix together with sparsity promoting optimization jointly for all illuminations. This work is an extension of our previous paper [5] on imaging using optimization techniques where we now account for multiple scattering effects.

**Key words.** array imaging, joint sparsity, multiple scattering, Foldy-Lax equations

**AMS subject classifications.** 49N30, 78A46

**1. Introduction.** Active array imaging when multiple scattering between the scatterers is important is challenging because it is a nonlinear inverse problem. In most applications, for example, in seismic imaging, ultrasonic non-destructive testing, synthetic aperture radar, etc., the imaging methods that are used ignore multiple scattering and deal with a linear inverse problem. This may result in some loss of resolution, especially if the imaging setup provides only partial information with, for example, a limited frequency range, limited illuminations or small arrays relative to the distance from the scatterers. In this paper, we study active array imaging at only one frequency and with single and multiple illuminations. The arrays considered are not small and could surround the scatterers. We consider the full nonlinear inverse problem when multiple scattering is included but limit the analysis to the case of imaging when the scatterers are small compared to the wavelength so that the Foldy-Lax approximation [13, 16, 17, 1, 20] can be used. Given the array data, we formulate imaging as

---

<sup>†</sup>anwei@math.stanford.edu

<sup>‡</sup>moscoso@math.uc3m.es

<sup>§</sup>papanico@math.stanford.edu

<sup>¶</sup>Department of Mathematics, Stanford University, California 94305, USA

<sup>||</sup>Gregorio Millán Institute, Universidad Carlos III de Madrid, Madrid 28911, Spain

<sup>\*\*</sup>M. Moscoso was supported by AFOSR grant FA9550-11-1-0266 and by AFOSR NSSEFF fellowship

<sup>††</sup>G. Papanicolaou was supported by AFOSR grant FA9550-11-1-0266

an underdetermined optimization problem with nonlinear constraints. It is underdetermined because the set of possible locations of the scatterers in the image regions we consider here is much larger than the array data set, as is often the case.

In this paper, we formulate the nonlinear optimization problem for imaging in two steps. In the first step, we treat the scatterers as equivalent sources of unknown locations whose strengths are also unknown but are related in a known way to the illumination, to the multiple scattering and to the underlying unknown reflectivities of the scatterers. Under specific, if somewhat conservative, hypotheses about the array imaging setup and the measurement noise level, we show that the location of the sources, that is, the scatterers, can be recovered exactly in the first step. This is because we employ an  $\ell_1$  minimization method that recovers the support of sparse solutions exactly. In the second step, once the location of the scatterers is fixed, their true reflectivities can be recovered using the known relationship to the source strengths obtained in step one. This is an explicit relation that comes from the Foldy-Lax equations, given the scatterer locations. The key to this two-step approach is the possibility of recovering exactly the locations of the scatterers in the first step. This effectively linearizes what is a nonlinear inverse problem. Our theoretical analysis is mainly based on the work by Tropp [25]. We give sufficient conditions on the imaging setup under which we can expect perfect reconstructions without noise, and conditions under which these reconstructions are stable when the data is contaminated with additive noise.

We note that the two-step imaging method we have described is applied at first to array data from a single illumination, in §3. However, because of the *screening effects*, not all scatterers can be recovered from data generated by a single illumination, in general. Moreover, using a single illumination with array imaging configurations often used in practice is not robust when data is contaminated with noise. These issues can be handled by applying the same two-step imaging method to data from multiple illuminations. For this case, a matrix version of the  $\ell_1$  minimization method is described and analyzed in §4.

When we have access to the full array response matrix, that is, when we have data for a full set of linearly independent illumination vectors, it is possible to image using its singular value decomposition (SVD). This is done in [15, 8, 19] where the authors show that, despite of the nonlinearity of the mapping from the reflectivities of the scatterers to the data received on the array, one can form an image and find the locations of the individual point scatterers by beamforming with the significant singular vectors. This is the MUSIC (Multiple Signal Classification) method, which is essentially the same as beamforming or migration using the significant singular vectors as illuminations. These illumination vectors are optimal in the sense that they result in array data with maximal power, which is proportional to the associated singular values.

We also use optimal illuminations in the  $\ell_1$  based minimization approach we introduce here. This increases the efficiency of the data collection process and lowers the impact of the noise in the data, as we show with numerical simulations in §6. This is so even when only a few optimal illuminations are used. We used optimal illuminations in [5] for a proposed hybrid  $\ell_1$  method in the context of array imaging when multiple scattering is negligible. The hybrid  $\ell_1$  method not only uses the optimal illumination but also projects the data into appropriate subspaces. This last step cannot be carried out when multiple scattering is important because after projection the sensing matrix contains unknown factors related to the multiple scattering.

We note that the problem of imaging localized scatterers from array data has already been studied from the perspective of sparse optimization, as for example in [12, 5] where, however, multiple scattering effects are neglected.

Finally, the performance of the imaging method we propose depends on the mutual coherence of the sensing matrix, defined in §3. We, therefore, carry out an analysis of the mutual coherence of the sensing matrix with different imaging configurations, with planar and spherical arrays, in §5. We show that spherical arrays give images with better resolutions and smaller upper bound of the corresponding sensing matrix than planar arrays.

We note that the formulation for imaging with non-negligible multiple scattering can also be combined with the optimization approach for imaging problems with intensity only measurements, which is studied by the authors in [4].

The organization of the paper is as follows. In the rest of this section we introduce some basic notations used throughout the paper. In §2, we formulate the array imaging problem with multiple scattering between the scatterers using the Foldy-Lax equations. In §3, we analyze the imaging problem with a single illumination, and with and without additive noise. In §4, we discuss active array imaging with multiple illuminations, where we also discuss the use of optimal illuminations and give an efficient algorithm for solving multiple measurement vector (MMV) problems. In §5, we discuss the impact of the array configuration on the performance of the methods proposed in the paper. In §6, we illustrate our results with various numerical examples under conditions much less conservative than the ones required by the theory. The proofs of all the theoretical results are given in the appendices at the end of the paper.

**Notation.** Throughout the paper, vectors are denoted by lower case letters in boldface, and matrices by capital letters in boldface. Given a vector  $\mathbf{v}$  of length  $N$ , the  $i^{\text{th}}$  entry is denoted by  $v_i$ . For a matrix  $\mathbf{Y}$  of size  $M \times N$ , the  $i^{\text{th}}$  row is denoted by  $Y_i$ , the  $j^{\text{th}}$  column by  $Y_{\cdot j}$ , and the  $(i, j)$  entry by  $Y_{ij}$ .

We will use several different norms. For  $p \geq 1$ ,  $q \geq 1$ , the  $\ell_p$  norm of vector  $\mathbf{v}$  is defined by

$$\|\mathbf{v}\|_{\ell_p} = \left( \sum_{i=1}^n |v_i|^p \right)^{1/p},$$

while the Frobenius norm of a matrix  $\mathbf{Y}$  is given by

$$\|\mathbf{Y}\|_F = \left( \sum_{i=1}^m \sum_{j=1}^n |Y_{ij}|^2 \right)^{1/2} = \sqrt{\text{trace}(\mathbf{Y}\mathbf{Y}^*)}.$$

Here,  $*$  is the conjugate transpose operator of vectors and matrices. We will use the superscript  $T$  for the transpose only operator. We will also need to use the operator norm of a matrix, defined as

$$\|\mathbf{Y}\|_{p \rightarrow q} = \max_{\mathbf{v} \neq 0} \frac{\|\mathbf{Y}\mathbf{v}\|_{\ell_q}}{\|\mathbf{v}\|_{\ell_p}},$$

and the  $(p, q)$ -norm function  $J_{p,q}(\cdot)$  defined as

$$J_{p,q}(\mathbf{Y}) = \left( \sum_{i=1}^m \|Y_i\|_{\ell_p}^q \right)^{1/q}, \quad (1.1)$$

which is simply the  $\ell_q$  norm of the vector formed by  $\ell_p$  norm of all rows of a matrix.

**2. Formulation of active array imaging.** In active array imaging we seek to locate the positions and reflectivities of a set of scatterers using the data recorded on an array  $\mathcal{A}$ . By an active array, we mean a collection of  $N$  transducers that emit spherical wave signals from positions  $\mathbf{x}_s \in \mathcal{A}$  and record the echoes with receivers at positions  $\mathbf{x}_r \in \mathcal{A}$ . The transducers are placed at distance  $h$  between them, which is of the order of the wavelength  $\lambda = 2\pi c_0/\omega$ , where  $c_0$  is the wave speed in the medium and  $\omega$  is the frequency of the probing signal.

We now introduce the direct and inverse scattering problems for imaging point-like scatterers with an active array in a homogeneous medium. We consider the case in which multiple scattering among  $M$  scatterers is not negligible. The scatterers, with unknown reflectivities  $\alpha_j \in \mathbb{C}$  and positions  $\mathbf{y}_{n_j}$ ,  $j = 1, \dots, M$ , are assumed to be located within a region of interest called the image window (IW), which is centered at a distance  $L$  from the array. We discretize the IW using a uniform grid of  $K$  points  $\mathbf{y}_j$ ,  $j = 1, \dots, K$ , and assume that each scatterer is located at one of these  $K$  grid points so that

$$\{\mathbf{y}_{n_1}, \dots, \mathbf{y}_{n_M}\} \subset \{\mathbf{y}_1, \dots, \mathbf{y}_K\}.$$

Furthermore, we assume that near-field multiple scattering effects are negligible because the scatterers are sufficiently far apart. Thus, we can use the far-field approximation

$$\widehat{G}_0(\mathbf{y}, \mathbf{x}, \omega) = \frac{\exp(i\kappa|\mathbf{x} - \mathbf{y}|)}{4\pi|\mathbf{x} - \mathbf{y}|}, \quad (2.1)$$

with  $\kappa = \omega/c_0$ , for the free-space Green's function  $\widehat{G}_0(\mathbf{y}, \mathbf{x}, \omega)$  to characterize wave propagation from point  $\mathbf{x}$  to point  $\mathbf{y}$  in the homogeneous medium.

We formulate the scattered wave field in a homogeneous medium in terms of the Foldy-Lax equations [13, 16, 17]. In this model framework, the scattered wave received at transducer  $\mathbf{x}_r$  due to a narrow band signal of angular frequency  $\omega$  sent from  $\mathbf{x}_s$  can be written as the sum of all scattered waves from the  $M$  scatterers

$$\widehat{P}(\mathbf{x}_r, \mathbf{x}_s) = \sum_{j=1}^M \widehat{\psi}_j^s(\mathbf{x}_r; \mathbf{y}_{n_1}, \dots, \mathbf{y}_{n_M}). \quad (2.2)$$

Here, and in all that follows, we will drop the dependence of waves and measurements on the frequency  $\omega$ . In (2.2),  $\widehat{\psi}_j^s(\mathbf{x}_r; \mathbf{y}_{n_1}, \dots, \mathbf{y}_{n_M})$  represents the scattered wave observed at  $\mathbf{x}_r$  due to the scatterer at position  $\mathbf{y}_{n_j}$ . It actually depends on the positions of all the scatterers  $\mathbf{y}_{n_j}$ ,  $j = 1, \dots, M$ , and it is given by

$$\widehat{\psi}_j^s(\mathbf{x}_r; \mathbf{y}_{n_1}, \dots, \mathbf{y}_{n_M}) = \alpha_j \widehat{G}_0(\mathbf{x}_r, \mathbf{y}_{n_j}) \widehat{\psi}_j^e(\mathbf{y}_{n_1}, \dots, \mathbf{y}_{n_M}), \quad (2.3)$$

where  $\hat{\psi}_j^e(\mathbf{y}_{n_1}, \dots, \mathbf{y}_{n_M})$  represents the exciting field at the scatterer located at  $\mathbf{y}_{n_j}$ . Because we ignore self-interacting fields, the exciting field at  $\mathbf{y}_{n_j}$  is equal to the sum of the incident field  $\hat{\psi}_j^{inc} := \hat{\psi}^{inc}(\mathbf{y}_{n_j}, \mathbf{x}_s)$  at  $\mathbf{y}_{n_j}$  and the scattered fields at  $\mathbf{y}_{n_j}$  due to all scatterers except for the one at  $\mathbf{y}_{n_j}$ . Hence, it is given by

$$\hat{\psi}_j^e(\mathbf{y}_{n_1}, \dots, \mathbf{y}_{n_M}) = \hat{\psi}^{inc}(\mathbf{y}_{n_j}, \mathbf{x}_s) + \sum_{k \neq j} \alpha_k \hat{G}_0(\mathbf{y}_{n_j}, \mathbf{y}_{n_k}) \hat{\psi}_k^e(\mathbf{y}_{n_1}, \dots, \mathbf{y}_{n_M}), \quad j = 1, 2, \dots, M. \quad (2.4)$$

This is a self-consistent system of  $M$  equations for the  $M$  unknown exciting fields

$$\hat{\psi}_1^e := \hat{\psi}_1^e(\mathbf{y}_{n_1}, \dots, \mathbf{y}_{n_M}), \dots, \hat{\psi}_M^e := \hat{\psi}_M^e(\mathbf{y}_{n_1}, \dots, \mathbf{y}_{n_M}),$$

which can be written in matrix form as

$$\mathbf{Z}_M(\boldsymbol{\alpha}) \boldsymbol{\Phi}^e = \boldsymbol{\Phi}^{inc}. \quad (2.5)$$

In (2.5),  $\boldsymbol{\Phi}^e = [\hat{\psi}_1^e, \dots, \hat{\psi}_M^e]^T$  and  $\boldsymbol{\Phi}^{inc} = [\hat{\psi}_1^{inc}, \dots, \hat{\psi}_M^{inc}]^T$  are vectors whose components are the exciting and incident fields on the  $M$  scatterers, respectively, and

$$(Z_M(\boldsymbol{\alpha}))_{ij} = \begin{cases} 1, & i = j \\ -\alpha_j \hat{G}_0(\mathbf{y}_{n_i}, \mathbf{y}_{n_j}), & i \neq j, \end{cases} \quad (2.6)$$

is the  $M \times M$  Foldy-Lax matrix which depends on the reflectivities  $\boldsymbol{\alpha} = (\alpha_1, \dots, \alpha_M)$ . With the solution of (2.5), we use (2.3) and (2.2) to compute the scattered data received at the array.

Note that the exciting fields  $\boldsymbol{\Phi}^e$  depend on the incident fields  $\boldsymbol{\Phi}^{inc}$  and, hence, they depend on the illumination sent from the array. To characterize it, we define the illumination vector  $\hat{\mathbf{f}} = [\hat{f}_1, \dots, \hat{f}_N]^T$  whose entries denote the strength of the signals sent from each of the  $N$  transducers in the array. We will assume that the illumination vectors are normalized, so  $\|\hat{\mathbf{f}}\|_{\ell_2} = 1$ .

To write the data received on the array in a more compact form, we define the Green's function or steering vector  $\hat{\mathbf{g}}_0(\mathbf{y})$  at location  $\mathbf{y}$  in the IW as

$$\hat{\mathbf{g}}_0(\mathbf{y}) = [\hat{G}_0(\mathbf{x}_1, \mathbf{y}), \hat{G}_0(\mathbf{x}_2, \mathbf{y}), \dots, \hat{G}_0(\mathbf{x}_N, \mathbf{y})]^T. \quad (2.7)$$

Then, given any illumination vector  $\hat{\mathbf{f}}$ , the incident field on the scatterer at position  $\mathbf{y}_{n_j}$  is equal to  $\hat{\mathbf{g}}_0^T(\mathbf{y}_{n_j}) \hat{\mathbf{f}}$ . If the illumination vector  $\hat{\mathbf{f}}$  is such that  $\hat{f}_s = 1$  and  $\hat{f}_j = 0$  for  $j = 1, \dots, N$  with  $j \neq s$ , then the incident field at  $\mathbf{y}_{n_j}$  is simply  $\hat{G}_0(\mathbf{y}_{n_j}, \mathbf{x}_s)$ .

Using (2.7), we also define the  $N \times K$  sensing matrix  $\mathcal{G}$  as

$$\mathcal{G} = [\hat{\mathbf{g}}_0(\mathbf{y}_1) \cdots \hat{\mathbf{g}}_0(\mathbf{y}_K)], \quad (2.8)$$

and the  $N \times M$  submatrix corresponding to the locations of scatterers as

$$\mathcal{G}_M = [\hat{\mathbf{g}}_0(\mathbf{y}_{n_1}) \cdots \hat{\mathbf{g}}_0(\mathbf{y}_{n_M})].$$

With this notation, the array response matrix can be written as

$$\widehat{\mathbf{P}} \equiv [\widehat{P}(\mathbf{x}_r, \mathbf{x}_s)]_{r,s=1}^N = \mathbf{G}_M \text{diag}(\boldsymbol{\alpha}) \mathbf{Z}_M^{-1}(\boldsymbol{\alpha}) \mathbf{G}_M^T, \quad (2.9)$$

and the data received on the array due to the illumination  $\widehat{\mathbf{f}}$  is

$$\mathbf{b} = \widehat{\mathbf{P}} \widehat{\mathbf{f}}. \quad (2.10)$$

Note that the response matrix in (2.9) that takes into account multiple scattering, includes the inverse of the Foldy-Lax matrix  $\mathbf{Z}_M^{-1}(\boldsymbol{\alpha})$ . When multiple scattering is negligible,  $\mathbf{Z}_M(\boldsymbol{\alpha}) = \mathbf{I}$  and we get the response matrix under the Born approximation, as shown for example in [5]. We further note that the response matrix  $\widehat{\mathbf{P}}$  given by (2.9) is symmetric.

Next, we introduce the true *reflectivity vector*  $\boldsymbol{\rho}_0 = [\rho_{01}, \dots, \rho_{0K}]^T \in \mathbb{C}^K$  such that

$$\rho_{0k} = \sum_{j=1}^M \alpha_j \delta_{\mathbf{y}_{n_j} \mathbf{y}_k}, \quad k = 1, \dots, K,$$

where  $\delta_{\cdot}$  is the classical Kronecker delta. Note that the Foldy-Lax matrix  $\mathbf{Z}_M(\boldsymbol{\alpha})$  is defined only for pairwise combinations of scatterers at  $\mathbf{y}_{n_j}$ ,  $j = 1, \dots, M$ . To formulate the inverse scattering problem, we need to extend the  $M \times M$  matrix  $\mathbf{Z}_M(\boldsymbol{\alpha})$  to a larger  $K \times K$  matrix

$$(Z(\boldsymbol{\rho}_0))_{ij} = \begin{cases} 1, & i = j \\ -\rho_{0j} \widehat{G}_0(\mathbf{y}_i, \mathbf{y}_j), & i \neq j, \end{cases} \quad (2.11)$$

which includes all pairwise combinations of the  $K$  grid points  $\mathbf{y}_j$  in the IW. With this notation, the array response matrix (2.9) can be written as

$$\widehat{\mathbf{P}} = \mathbf{G} \text{diag}(\boldsymbol{\rho}_0) \mathbf{Z}^{-1}(\boldsymbol{\rho}_0) \mathbf{G}^T. \quad (2.12)$$

Furthermore, if we define the Foldy-Lax Green's function vector  $\widehat{\mathbf{g}}_{FL}(\mathbf{y}_j)$ ,  $j = 1, \dots, K$ , as the  $j^{\text{th}}$  column of the matrix  $\mathbf{G}_{FL}(\boldsymbol{\rho}) = \mathbf{G} \mathbf{Z}^{-T}(\boldsymbol{\rho})$ , i.e.,

$$[\widehat{\mathbf{g}}_{FL}(\mathbf{y}_1) \quad \cdots \quad \widehat{\mathbf{g}}_{FL}(\mathbf{y}_K)] = \mathbf{G} \mathbf{Z}^{-T}(\boldsymbol{\rho}), \quad (2.13)$$

then (2.12) can be simplified to

$$\widehat{\mathbf{P}} = \mathbf{G} \text{diag}(\boldsymbol{\rho}_0) \mathbf{G}_{FL}^T(\boldsymbol{\rho}_0). \quad (2.14)$$

Given an illumination vector  $\widehat{\mathbf{f}}$  and the configuration of scatterers in the IW characterized by  $\boldsymbol{\rho}_0$ , the data received on the array is given by (2.10). The array imaging problem when a single illumination is used to probe the medium is to find the true reflectivity vector  $\boldsymbol{\rho}_0$  from the received data  $\mathbf{b}$ . The detailed formulation of this problem will be discussed in depth in §3. The array imaging problem that uses a collection of array data generated by different illumination vectors will be discussed in §4. In either situation, our method for active array imaging with multiple scattering is noniterative. It uses two steps to get the images: first locating the scatterers and second computing their reflectivities.

**3. Active array imaging with single illumination.** In this section, we show the formulation of active array imaging including multiple scattering when only one illumination is sent from the array to probe the medium. In this case, a single measurement vector is used to infer the location and reflectivities of the scatterers. In signal processing literature, this problem belongs to the so called *Single Measurement Vector* (SMV) problem.

For a given illumination vector  $\hat{\mathbf{f}}$ , we define the operator  $\mathcal{A}_{\hat{\mathbf{f}}}[\boldsymbol{\rho}_0]$  through the identity

$$\mathcal{A}_{\hat{\mathbf{f}}}[\boldsymbol{\rho}_0]\boldsymbol{\rho}_0 = \hat{\mathbf{P}}\hat{\mathbf{f}},$$

which connects the reflectivity vector  $\boldsymbol{\rho}_0$  and the data (2.10). It is easy to see from (2.12) that  $\mathcal{A}_{\hat{\mathbf{f}}}$  has the form

$$\mathcal{A}_{\hat{\mathbf{f}}}[\boldsymbol{\rho}] = [\hat{g}_{\hat{\mathbf{f}}}(\mathbf{y}_1)\hat{\mathbf{g}}_0(\mathbf{y}_1) \cdots \hat{g}_{\hat{\mathbf{f}}}(\mathbf{y}_K)\hat{\mathbf{g}}_0(\mathbf{y}_K)],$$

where  $\hat{g}_{\hat{\mathbf{f}}}(\mathbf{y}_j) = \hat{\mathbf{g}}_{FL}^T(\mathbf{y}_j)\hat{\mathbf{f}}$ ,  $j = 1, \dots, K$ , are scalars. With this notation, active array imaging with a single illumination amounts to solving for  $\boldsymbol{\rho}_0$  from the system of equations

$$\mathcal{A}_{\hat{\mathbf{f}}}[\boldsymbol{\rho}]\boldsymbol{\rho} = \mathbf{b}. \quad (3.1)$$

The number of transducers  $N$  is usually much smaller than the number of the grid points  $K$  in the IW and, hence, (3.1) is an underdetermined system of equations.

Although equations (3.1) are exactly of the same form as the problem studied in [5], there is an important difference. Due to the multiple scattering among the scatterers, the terms  $\hat{g}_{\hat{\mathbf{f}}}(\mathbf{y}_j)$ ,  $j = 1, \dots, K$ , contained in  $\mathcal{A}_{\hat{\mathbf{f}}}[\boldsymbol{\rho}]$  depend now on the unknown reflectivity vector  $\boldsymbol{\rho}$ . This makes equations (3.1) nonlinear with respect to  $\boldsymbol{\rho}$  and, hence, one would think that non-iterative inversion is impossible when multiple scattering is non-negligible. In fact, several nonlinear iterative methods have been proposed in the literature to solve this problem: see, for example, [8, 22]. However, as demonstrated below, by rearranging the terms in the equations, we can reformulate the problem to solve for the locations of the scatterers directly (without any iteration), and then to recover their reflectivities in a second single step.

To solve for the locations of the scatterers in one step, we introduce the *effective source vector*

$$\boldsymbol{\gamma}_{\hat{\mathbf{f}}} = \text{diag}(\boldsymbol{\rho})\mathbf{Z}^{-1}(\boldsymbol{\rho})\mathcal{G}^T\hat{\mathbf{f}}. \quad (3.2)$$

Note that the support of  $\boldsymbol{\gamma}_{\hat{\mathbf{f}}}$  is contained in the support of  $\boldsymbol{\rho}$ . Using (2.12) and (3.2), (3.1) can be rewritten as  $\mathcal{A}_{\hat{\mathbf{f}}}[\boldsymbol{\rho}]\boldsymbol{\rho} = \mathcal{G}\boldsymbol{\gamma}_{\hat{\mathbf{f}}} = \mathbf{b}$ , and the system of equations

$$\mathcal{G}\boldsymbol{\gamma}_{\hat{\mathbf{f}}} = \mathbf{b} \quad (3.3)$$

becomes linear for the new unknowns  $\boldsymbol{\gamma}_{\hat{\mathbf{f}}}$ . We point out that, unlike the problem considered in [5], when multiple scattering is not negligible, solving (3.3) may not be able to recover all the support of  $\boldsymbol{\rho}_0$ . This is not a flaw of the formulation but an implicit problem of array imaging when multiple scattering is important. Indeed, due to multiple scattering effects it is possible that one or several scalars  $\hat{g}_{\hat{\mathbf{f}}}(\mathbf{y}_j)$ ,  $j = 1, \dots, K$ , are very small or even zero and, hence, the corresponding scatterers become hidden. This is the well-known *screening effect*

which makes scatterers undetectable, and that it is manifested in our formulation making some of the components of the effective source vector  $\gamma_{\hat{\mathbf{f}}}$  arbitrarily small.

Note that, for a fixed imaging configuration, the *screening effect* depends only on the illumination vector  $\hat{\mathbf{f}}$  and the amount of noise in the data. Indeed, when the effective source at  $\mathbf{y}_j$  is below the noise level because  $\hat{\mathbf{g}}_{\hat{\mathbf{f}}}(\mathbf{y}_j)$  is small, then the corresponding scatterer cannot be detected. This motivates us, in the next section, to consider active array imaging with multiple illuminations. In this case, active array imaging is formulated as a joint sparsity recovery problem where we seek an unknown matrix whose columns share the same support. By increasing the diversity of illuminations, we minimize the *screening* and we have more chances of locating all the scatterers.

Since (3.3) is underdetermined and the effective source vector  $\gamma_{\hat{\mathbf{f}}}$  is sparse ( $M \ll K$ ), we use  $\ell_1$  minimization

$$\min \|\gamma_{\hat{\mathbf{f}}}\|_{\ell_1} \quad \text{s.t.} \quad \mathcal{G}\gamma_{\hat{\mathbf{f}}} = \mathbf{b} \quad (3.4)$$

to obtain  $\gamma_{0\hat{\mathbf{f}}}$  from noiseless data. When the data  $\mathbf{b}$  is contaminated by a noise vector  $\mathbf{e}$  with finite energy, we then seek the solution to the relaxed problem

$$\min \|\gamma_{\hat{\mathbf{f}}}\|_{\ell_1} \quad \text{s.t.} \quad \|\mathcal{G}\gamma_{\hat{\mathbf{f}}} - \mathbf{b}\|_{\ell_2} < \delta, \quad (3.5)$$

for some given positive constant  $\delta$ . Using Theorem 3.1 in [5] and Theorem 14 in [26], we have the following uniqueness and stability results.

**Theorem 3.1.** *For a given array configuration, assume that the resolution of the IW is such that*

$$\max_{i \neq j} \left| \frac{\hat{\mathbf{g}}_0^*(\mathbf{y}_i)\hat{\mathbf{g}}_0(\mathbf{y}_j)}{\|\hat{\mathbf{g}}_0(\mathbf{y}_i)\|_{\ell_2}\|\hat{\mathbf{g}}_0(\mathbf{y}_j)\|_{\ell_2}} \right| < \epsilon, \quad (3.6)$$

and there is no noise in the data. If the number of scatterers  $M$  satisfies that  $M\epsilon < 1/2$ , then  $\gamma_{0\hat{\mathbf{f}}}$  is the unique solution to (3.4).

**Theorem 3.2.** *Under the same condition (3.6) as in Theorem 3.1, if the data contain additive noise of finite energy  $\|\mathbf{e}\|_{\ell_2}$ , then the solution  $\gamma_{*\hat{\mathbf{f}}}$  to (3.5) satisfies*

$$\|\gamma_{*\hat{\mathbf{f}}} - \gamma_{0\hat{\mathbf{f}}}\|_{\ell_2} \leq \frac{\delta}{\sqrt{1 - (M-1)\epsilon}}, \quad (3.7)$$

provided  $\delta \geq \|\mathbf{e}\|_{\ell_2} \sqrt{1 + \frac{M(1-(M-1)\epsilon)}{(1-2M\epsilon+\epsilon)^2}}$ . Moreover, the support of  $\gamma_{*\hat{\mathbf{f}}}$  is fully contained in that of  $\gamma_{0\hat{\mathbf{f}}}$ , and all the components such that

$$|(\gamma_{0\hat{\mathbf{f}}})_j| > \delta/\sqrt{1 - (M-1)\epsilon} \quad (3.8)$$

are within the support of  $\gamma_{*\hat{\mathbf{f}}}$ .

**Remark 3.3.** *Theorem 3.1 gives the required condition to recover the effective source vector exactly from noiseless data. The resolution condition is based on the so called mutual coherence*

$$\mu(\mathcal{G}) = \max_{i \neq j} \left| \frac{\hat{\mathbf{g}}_0^*(\mathbf{y}_i)\hat{\mathbf{g}}_0(\mathbf{y}_j)}{\|\hat{\mathbf{g}}_0(\mathbf{y}_i)\|_{\ell_2}\|\hat{\mathbf{g}}_0(\mathbf{y}_j)\|_{\ell_2}} \right| \quad (3.9)$$

of the sensing matrix  $\mathbf{G}$ , which is determined by the array imaging configuration (array size and resolution of the IW). The mutual coherence is a measure of how linearly independent the columns of the sensing matrix are. We give analytical results regarding the impact of the array geometry on (3.9) in §5. Specifically, we show that a sensing matrix  $\mathbf{G}$  with small mutual coherence requires large arrays.

Problems (3.4) and (3.5) give the effective source vector  $\gamma_{\hat{\mathbf{f}}}$ . In a second step, we compute the true reflectivities from the solutions of these problems. According to (3.2), we need to solve a nonlinear equation and, therefore, iteration seems to be inevitable. However, it is not necessary. Let  $\Lambda_\star$  be the support of the recovered solution such that  $|\Lambda_\star| = M' \leq M$ , and  $\gamma_{\hat{\mathbf{f}}, M'}$  the solution vector on that support. From (2.13) and (3.2), we obtain

$$\gamma_{\hat{\mathbf{f}}, M'} = \text{diag}(\mathbf{Z}^{-1}(\boldsymbol{\rho}_{M'})\mathbf{G}^T\hat{\mathbf{f}})\boldsymbol{\rho}_{M'} = \text{diag}(\hat{g}_{\hat{\mathbf{f}}}(\mathbf{y}_{n_1}), \dots, \hat{g}_{\hat{\mathbf{f}}}(\mathbf{y}_{n_{M'}}))\boldsymbol{\rho}_{M'},$$

where the scalars  $\hat{g}_{\hat{\mathbf{f}}}(\mathbf{y}_{n_j}) = \hat{\mathbf{g}}_{FL}^T(\mathbf{y}_{n_j})\hat{\mathbf{f}}$ . Note that the scalars  $\hat{g}_{\hat{\mathbf{f}}}(\mathbf{y}_{n_j})$  are the exciting fields at the scatterer's positions, that is,  $\hat{g}_{\hat{\mathbf{f}}}(\mathbf{y}_{n_j}) := \hat{\psi}_j^e(\mathbf{y}_{n_1}, \dots, \mathbf{y}_{n_{M'}})$ , and that the effective sources  $\gamma_{n_j}$  are the true reflectivities  $\rho_{n_j}$  of the scatterers multiplied by the exciting fields. Hence, using (2.4), we can compute  $\hat{g}_{\hat{\mathbf{f}}}(\mathbf{y}_{n_j})$  explicitly as follows

$$\hat{g}_{\hat{\mathbf{f}}}(\mathbf{y}_{n_j}) = \hat{\mathbf{g}}_0^T(\mathbf{y}_{n_j})\hat{\mathbf{f}} + \sum_{k=1, k \neq j}^{M'} \gamma_k \hat{G}_0(\mathbf{y}_{n_j}, \mathbf{y}_{n_k}), \quad j = 1, \dots, M'. \quad (3.10)$$

Then, the true reflectivities of the scatterers are given by

$$\rho_{n_j} = \gamma_{n_j} / \hat{g}_{\hat{\mathbf{f}}}(\mathbf{y}_{n_j}), \quad j = 1, \dots, M'. \quad (3.11)$$

For the noiseless case,  $\Lambda_\star = \Lambda_0$  based on Theorem 3.1. When the data contains additive noise, we choose the support  $\Lambda_\star$  of the solution recovered by (3.5) such that all the components of  $\gamma_{\hat{\mathbf{f}}, M'}$  satisfy (3.8).

To summarize, when a single illumination is used to probe the medium, we take two steps to locate the scatterers and to obtain their reflectivities, as follows.

- Solve the  $\ell_1$  minimization problem (3.4) or (3.5) for the *effective source vector*.
- Compute the true reflectivities using (3.10) and (3.11) on the support  $\Lambda_\star$ .

We note that it is essential to carry out the second step only at those locations contained in the support  $\Lambda_\star$  given by (3.8). Otherwise, (3.11) could lead to instabilities because  $\hat{g}_{\hat{\mathbf{f}}}(\mathbf{y}_{n_j})$  can be very small at any location in the IW. On the support  $\Lambda_\star$ , this is not critical because if  $\hat{g}_{\hat{\mathbf{f}}}(\mathbf{y}_{n_j})$  is small, so is  $\gamma_{n_j}$ .

There are many fast and efficient numerical algorithms for solving (3.4) or (3.5). In the simulation study below, we use the iterative shrinkage-thresholding algorithm GelMa, described in [21], due to its flexibility with respect to the choice of the regularization parameter used in the algorithm.

**4. Imaging using multiple illumination vectors.** In the previous section we discuss a non-iterative approach for array imaging with multiple scattering when a single illumination is used. Although the proposed approach can recover the locations and reflectivities of the

scatterers exactly when the data is noiseless, it can be very sensitive to additive noise, especially when the noise level is high, leading to unreliable images. Moreover, the *screening effect* associated with multiple scattering can cause the failure of recovering some scatterers in the IW. This means that for a given illumination  $\hat{\mathbf{f}}$  the number of effective sources  $M'$  is strictly less than the number of scatterers  $M$ . These two problems can be mitigated by using multiple illuminations which can often be controlled to increase the power of the signals received at the array. We will show that by carefully choosing the illumination vectors, the use of multiple inputs makes array imaging more stable in the presence of relatively high noise and, at the same time, the screening effect is minimized.

**4.1. Imaging with multiple arbitrary illuminations.** To work with data generated by multiple (random) illumination vectors, a natural extension is to stack the data vectors  $\mathbf{b}^j$  from illuminations  $\hat{\mathbf{f}}^j$ ,  $j = 1, \dots, \nu$ , into a single  $\nu N$  vector, and to apply the approach in §3 to the augmented linear system. However, by simply stacking the data forming a larger linear system not only increases the dimensionality of the problem but also fails to exploit the intrinsic relation among the multiple data vectors. To make use of the data structure, we formulate the problem of array imaging with multiple illuminations as a joint sparsity recovery problem, also known as the *Multiple Measurement Vector* (MMV) approach. Instead of solving a matrix-vector equation for the unknown reflectivity vector, we now solve a matrix-matrix equation for an unknown matrix variable whose columns share the same sparse support but possibly different nonzero values. The MMV approach has been widely studied in passive source localization problems and other applications with success, see for example [18]. With the introduction of the *effective source vector*, MMV can also be used effectively for active array imaging when multiple scattering between scatterers is important.

Let  $\mathbf{B} = [\mathbf{b}^1 \dots \mathbf{b}^\nu]$  be the matrix whose columns are the data vectors generated by all the illuminations, and  $\mathbf{X}_0 = [\gamma_0^1 \dots \gamma_0^\nu]$  be the unknown matrix whose  $j^{\text{th}}$  column corresponds to the *effective source vector*  $\gamma_0^j$  under illumination  $\hat{\mathbf{f}}^j$ ,  $j = 1, \dots, \nu$ . Then, the MMV formulation for active array imaging is to solve for  $\mathbf{X}_0$  from the matrix-matrix equation

$$\mathcal{G}\mathbf{X} = \mathbf{B}. \quad (4.1)$$

In this framework, the sparsity of the matrix variable  $\mathbf{X}$  is characterized by the number of nonzero rows of the matrix. More precisely, we define the row-support of a given matrix  $\mathbf{X}$  by

$$\text{rowsupp}(\mathbf{X}) = \{i : \exists j \text{ s.t. } X_{ij} \neq 0\},$$

which is equivalent to

$$\text{rowsupp}(\mathbf{X}) = \{i : \|X_{i\cdot}\|_{\ell_p} \neq 0\},$$

where  $p \geq 1$ . From this definition, we see that when the matrix  $\mathbf{X}$  degenerates to a column vector, the row-support reduces to the support of the vector. The joint sparsity of  $\mathbf{X}$  is then measured by the row-wise  $\ell_0$  norm of  $\mathbf{X}$  defined by

$$\Xi_0(\mathbf{X}) = |\text{rowsupp}(\mathbf{X})|.$$

With these definitions, the sparsest solution of array imaging using multiple illuminations is given by the solution to the problem

$$\min \Xi_0(\mathbf{X}) \quad \text{s.t.} \quad \mathcal{G}\mathbf{X} = \mathbf{B}. \quad (4.2)$$

As with the  $\ell_0$  norm minimization problem in SMV, (4.2) is an NP hard problem. An alternative is to solve the convex relaxed problem

$$\min \Xi_1(\mathbf{X}) \quad \text{s.t.} \quad \mathcal{G}\mathbf{X} = \mathbf{B}, \quad (4.3)$$

where the substitution of  $\Xi_0$  by a certain function  $\Xi_1$  turns (4.2) into a tractable problem. There are many choices of  $\Xi_1$  as discussed, for example, in [7, 6, 28]. We note here that  $\Xi_1 = J_{p,1}$  for any  $p \geq 1$ , as defined in (1.1), can be used to replace the nonconvex objective function  $\Xi_0$ . We will use  $p = 2$  in the following discussion which has been studied in, for example, [7, 18, 6, 11]. Therefore, we consider the following convex relaxed problem to image the scatterers with multiple illumination vectors

$$\min J_{2,1}(\mathbf{X}) \quad \text{s.t.} \quad \mathcal{G}\mathbf{X} = \mathbf{B}. \quad (4.4)$$

As with Theorem 3.1, we have the following condition for recovery using (4.4).

**Theorem 4.1.** *For a given array configuration, assume that the resolution of the IW satisfies (3.6). If the number of scatterers  $M$  is such that  $M\epsilon < 1/2$ , then  $\mathbf{X}_0 = [\gamma_0^1 \dots \gamma_0^\nu]$  is the unique solution to (4.4).*

**Remark 4.2.** *The condition given in Theorem 4.1 is also the sufficient condition for the complete family of MMV problems that use the  $J_{p,1}$  type of objective function to convert the original non-convex problem (4.2) into a convex, solvable one. In fact, we prove Theorem 4.1 by showing  $\mathbf{X}_0$  is the unique solution to*

$$\min J_{p,1}(\mathbf{X}) \quad \text{s.t.} \quad \mathcal{G}\mathbf{X} = \mathbf{B}$$

for any  $1 < p < \infty$  in Appendix A. The case of  $p = \infty$  is studied in [28]. We also note that for the case  $p = 1$ , the resulting formulation becomes fully decoupled. Indeed, solving

$$\min J_{1,1}(\mathbf{X}) \quad \text{s.t.} \quad \mathcal{G}\mathbf{X} = \mathbf{B}$$

can be shown to be equivalent to solving  $\nu$  simple  $\ell_1$ -norm minimization problems with single illumination, and hence, this approach does not fully utilize the joint sparsity of the problem. Therefore, the support is not simultaneously recovered with  $J_{1,1}$ . This observation has also been discussed in [28] and [6].

When the collected data is contaminated by additive noise vectors  $\mathbf{e}^j$ ,  $j = 1, \dots, \nu$ , equations (4.1) become

$$\mathcal{G}\mathbf{X} = \mathbf{B} + \mathcal{E}. \quad (4.5)$$

Here,  $\mathcal{E} = [\mathbf{e}^1 \dots \mathbf{e}^\nu]$  is the matrix whose columns are independent noise vectors  $\mathbf{e}^j$  corresponding to each illumination vector  $\hat{\mathbf{f}}^j$ ,  $j = 1, \dots, \nu$ . Then, as with the the single illumination case, we seek a solution to

$$\min J_{2,1}(\mathbf{X}) \quad \text{s.t.} \quad \|\mathcal{G}\mathbf{X} - \mathbf{B}\|_F < \delta, \quad (4.6)$$

for some pre-specified constant  $\delta$ . As stated in the following result, the solution to (4.6) recovers the sparsest solution  $\mathbf{X}_0$  upon certain error bound. The result is proved using a similar approach as the one used in [28] for  $J_{\infty,1}$ . Details are given in Appendix B.

**Theorem 4.3.** *For a given array configuration, assume that the resolution of the IW satisfies (3.6). If the number of scatterers  $M$  is such that  $M\epsilon < 1/2$ , and*

$$\delta \geq \|\mathcal{E}\|_F \sqrt{1 + \frac{M(1 - (M-1)\epsilon)}{(1 - 2M\epsilon + \epsilon)^2}}, \quad (4.7)$$

then (4.6) has a unique solution  $\mathbf{X}_*$  which has row support included in that of  $\mathbf{X}_0$  and satisfies

$$\|\mathbf{X}_* - \mathbf{X}_0\|_F \leq \frac{\delta}{\sqrt{1 - (M-1)\epsilon}}. \quad (4.8)$$

Moreover, the row support of  $\mathbf{X}_*$  contains all the rows  $i$  satisfying

$$\|(\mathbf{X}_0)_i\|_{\ell_2} > \frac{\delta}{\sqrt{1 - (M-1)\epsilon}}. \quad (4.9)$$

According to Theorems 4.1 and 4.3, the performance of (4.4) and (4.6) does not depend on the number of measurements  $\nu$ . Therefore, judging from these theoretical results, there is no quantitative improvement in the conditions imposed on the imaging setup when using multiple illuminations compared to those for a single illumination. Intuitively, this is so because it is possible that measurements from different (random) illuminations may all be rather ineffective and, therefore, there would not be an advantage in using multiple measurements in such a case. However, in practice, we observe that there is in general improvement in the image, which is much better when (random) multiple illuminations are used, especially in the presence of additive noise. To explain the improved performance seen in practice, the authors in [11] carried out an average-case analysis of the underlying joint sparsity recovery problem by introducing a probability model for  $\mathbf{X}$ . They showed in that context that the probability of failing to recover the true solution vector decays exponentially with the number of measurements.

We note that the recovery condition of (4.4) and (4.6) still depends on the mutual coherence of the sensing matrix  $\mathcal{G}$ , i.e., on (3.6). As we have already remarked, this condition depends only on the configuration of the imaging problem, the array geometry and the chosen discretization of the image window IW. In §5, we discuss array configurations that lead to different conditions (3.6).

Once we obtain from (4.4) or (4.6) the matrix  $\mathbf{X}_*$ , whose columns are the effective sources corresponding to the different illuminations, we then compute in a second step the true reflectivities as follows. For each component  $i$  in the support such that (4.9) is satisfied, we compute the reflectivities  $\rho_i^j$  corresponding to each illumination  $j$  by applying (3.10) and (3.11). We then take the average  $\frac{1}{\nu} \sum_{j=1}^{\nu} \rho_i^j$  as the estimated reflectivity.

**4.2. Imaging with optimal illuminations.** In order to increase the robustness of the methods (3.4) and (3.5), and to mitigate screening effects, MMV uses data obtained from multiple

illuminations. One approach in MMV is to use multiple illuminations selected randomly. However, such illuminations may not avoid screening above a certain noise level, as we see in numerical simulations in §6. Furthermore, using random illuminations may not be very efficient because a large number of them are needed to get a significant improvement in the image.

We now introduce an approach that uses optimal illuminations within the MMV framework. The use of optimal illuminations for array imaging in homogeneous and random media has been studied in [2, 3, 5]. The optimal illuminations can be computed systematically from the singular value decomposition (SVD) of the array response matrix  $\hat{\mathbf{P}}$ , or with an iterative time reversal process as discussed in [5, 24, 23] when the full array response matrix is not available. Let the SVD of  $\hat{\mathbf{P}}$  given in (2.9) be

$$\hat{\mathbf{P}} = \hat{\mathbf{U}}\mathbf{\Sigma}\hat{\mathbf{V}}^* = \sum_{j=1}^{\tilde{M}} \sigma_j \hat{\mathbf{U}}_{\cdot j} \hat{\mathbf{V}}_{\cdot j}^*,$$

where  $\hat{\mathbf{U}}_{\cdot j}$  and  $\hat{\mathbf{V}}_{\cdot j}$  are the left and right singular vectors, respectively, and the nonzero singular values  $\sigma_j$  are given in descending order as  $\sigma_1 \geq \sigma_2 \geq \dots \geq \sigma_{\tilde{M}} > 0$ , with  $\tilde{M} \geq M$ . When there is no additive noise in the data, we have  $\tilde{M} = M$ . Let the illumination vectors be the right singular vectors  $\hat{\mathbf{V}}_{\cdot j}$ , that is,  $\hat{\mathbf{f}}^j = \hat{\mathbf{V}}_{\cdot j}$ ,  $j = 1, \dots, \nu \leq \tilde{M}$ . Then,

$$\mathbf{B}_{opt} = \mathbf{g}\mathbf{X} = \hat{\mathbf{P}}\hat{\mathbf{V}}_{\cdot 1:\nu} = [\sigma_1 \hat{\mathbf{U}}_{\cdot 1} \dots \sigma_\nu \hat{\mathbf{U}}_{\cdot \nu}] + \tilde{\mathbf{E}}. \quad (4.10)$$

All the information for imaging is contained in the matrix  $\mathbf{B}_{opt}$  given in (4.10). It is also clear that the use of optimal illuminations filters out noise in the data because it reduces the dimensionality of the resulting optimization problem without loss of essential information about the scatterers.

Recall that the singular vectors  $\hat{\mathbf{V}}_{\cdot j}$ , with  $j = 1, \dots, M$ , are the illuminations that focus at each scatterer when multiple scattering is negligible and the scatterers are well resolved by the array. The key point here is that when multiple scattering is important, these optimal illuminations still deliver most of the energy around the scatterers, but each  $\hat{\mathbf{V}}_{\cdot j}$  is no longer associated with a single scatterer only. All the scatterers are illuminated in general where multiple scattering is important. As a consequence, taking a few top singular vectors, less than  $\tilde{M}$ , is enough to locate all the scatterers and image them. Moreover, taking fewer illuminations can be beneficial since less noise is introduced into (4.10). We illustrate this observation with numerical examples in §6.

We note that, by using optimal illuminations from the SVD of the array response matrix  $\hat{\mathbf{P}}$ , we are able to make the performance of the MMV formulation deviate significantly from the average case when using random illuminations.

**4.3. A sparsity promoting algorithm.** The MMV problem (4.2) can be solved by greedy algorithms that are straightforward generalizations of orthogonal matching pursuit for the single measurement case [7, 10, 27, 14]. At each iteration, these algorithms increase the joint support set of the estimated solution by one index, until a given number of columns vectors of the sensing matrix are selected or the approximation error is below a preset threshold.

Sparse Bayesian learning approaches developed for the single measurement case have also been extended to solve (4.2) [29, 30]. Both types of methods, however, become slow when the size of the problem is large. Alternatively, (4.2) can be relaxed to the convex formulation (4.4) (or (4.6)) and then consider algorithms that are extensions of those used to solve (3.4) (or (3.5)).

For our numerical simulations we will employ an extension of an iterative algorithm proposed in [21], called GeLMA. This is a shrinkage-thresholding algorithm for solving  $\ell_1$ -minimization problems which has proven to be very efficient and whose solution does not depend on the regularization parameter that promotes sparse solutions. In our case, the algorithm deals with the penalized problem

$$L(\mathbf{X}) = \frac{1}{2} \|\mathcal{G}\mathbf{X} - \mathbf{B}\|_F^2 + \tau J_{2,1}(\mathbf{X}), \quad (4.11)$$

and is derived based on the augmented Lagrangian form

$$F(\mathbf{X}, \mathcal{Z}) = L(\mathbf{X}) + \langle \mathcal{Z}, \mathbf{B} - \mathcal{G}\mathbf{X} \rangle. \quad (4.12)$$

For any fixed matrix multiplier  $\mathcal{Z}$ , the function  $F(\mathbf{X}, \mathcal{Z})$  is convex in  $\mathbf{X}$  and thus, we can compute its minimum iteratively. At iteration  $(k+1)$ , we first fix  $\mathcal{Z} = \mathcal{Z}^{(k)}$  and we seek the minimum of the differentiable part of  $F(\mathbf{X}, \mathcal{Z}^{(k)})$  as

$$\mathbf{Y}^{(k+1)} = \underset{\mathbf{X}}{\operatorname{argmin}} \left\{ \frac{1}{2} \|\mathcal{G}\mathbf{X} - \mathbf{B}\|_F^2 + \langle \mathcal{Z}^{(k)}, \mathbf{B} - \mathcal{G}\mathbf{X} \rangle \right\}.$$

Together with  $\mathbf{X}^{(k)}$  from the previous iteration, we compute

$$\mathbf{Y}^{(k+1)} = \mathbf{X}^{(k)} + \beta \mathcal{G}^* (\mathcal{Z}^{(k)} + \mathbf{B} - \mathcal{G}\mathbf{X}^{(k)})$$

using a first order iterative gradient descent method, where  $\beta$  is the step size. Next, we consider the (non-differentiable) regularization part through minimizing

$$\min_{\mathbf{X}} \left\{ \frac{1}{2} \|\mathbf{X} - \mathbf{Y}^{(k+1)}\|_F^2 + \beta \tau J_{2,1}(\mathbf{X}) \right\}.$$

Due to the row separability of both, the Frobenius matrix norm and the function  $J_{2,1}$ , this problem can be decomposed into the following  $N$  sub-problems

$$\min_{X_{i.}} \left\{ \frac{1}{2} \|X_{i.} - Y_{i.}^{(k+1)}\|_{\ell_2}^2 + \beta \tau \|X_{i.}\|_{\ell_2} \right\}, \quad i = 1, \dots, N.$$

Each sub-problem is quadratic in  $X_{i.}$ , and there exists a closed-form solution given by

$$X_{i.}^{(k+1)} = \operatorname{sign}(\|Y_{i.}^{(k+1)}\|_{\ell_2} - \beta \tau) \frac{\|Y_{i.}^{(k+1)}\|_{\ell_2} - \beta \tau}{\|Y_{i.}^{(k+1)}\|_{\ell_2}} Y_{i.}^{(k+1)}, \quad i = 1, \dots, N,$$

which involves only a simple shrinkage-thresholding step. Finally,  $\mathcal{Z}^{(k+1)}$  is found by applying a gradient ascent method as

$$\mathcal{Z}^{(k+1)} = \mathcal{Z}^{(k)} + \beta (\mathbf{B} - \mathcal{G}\mathbf{X}^{(k)}).$$

For more details regarding the properties of this algorithm for the single measurement case, we refer to [21] and references therein. We summarize it for MMV problems in Algorithm 1.

---

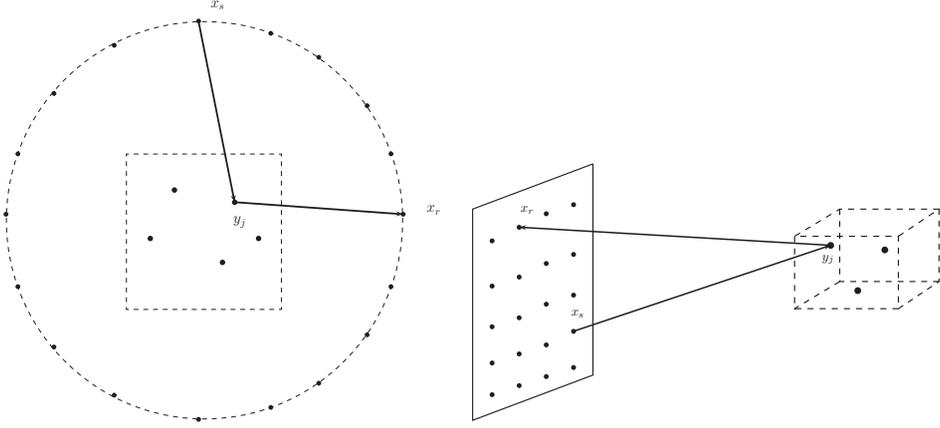
**Algorithm 1** GelMa-MMV for solving (4.12)
 

---

**Require:** Set  $\mathbf{X} = \mathbf{0}$ ,  $\mathbf{Z} = \mathbf{0}$  and pick the step size  $\beta$ , and the regularization parameter  $\tau$ 
**repeat**

 Compute the residual  $\mathbf{R} = \mathbf{B} - \mathbf{G}\mathbf{X}$ 
 $\mathbf{X} \leftarrow \mathbf{X} + \beta \mathbf{G}^*(\mathbf{Z} + \mathbf{R})$ 
 $X_i \leftarrow \text{sign}(\|X_i\|_{\ell_2} - \beta\tau) \frac{\|X_i\|_{\ell_2} - \beta\tau}{\|X_i\|_{\ell_2}} X_i, i = 1, \dots, K$ 
 $\mathbf{Z} \leftarrow \mathbf{Z} + \beta \mathbf{R}$ 
**until** Convergence
 

---



**Figure 5.1.** Schemata of spherical (left) and planar (right) arrays

**5. Array configuration and mutual coherence.** We have already discussed that the performance of sparsity promoting algorithms strongly depends on the mutual coherence of the sensing matrix, which is related to the array imaging configuration. In this section, we give some analytical results for the mutual coherence of two types of arrays that are often used in array imaging: planar arrays and spherical arrays. The schemata of these two types are illustrated in Figure 5.1. We show that under similar configurations of the IW (distance to the array and the resolution), spherical arrays give smaller upper bounds of the inner products of the normalized Green's function vectors than planar arrays in condition (3.6). We give the proofs in Appendix C.

The first result is on the estimate of the inner product when a spherical array surrounding the IW is used. It is a well-known classical result. We state it here to be self-contained.

**Proposition 5.1.** *Assume that the IW is fully surrounded by a spherical array of radius  $L$ . Given any two points  $\mathbf{y}_k$  and  $\mathbf{y}_{k'}$  in the IW such that  $\lambda \ll |\mathbf{y}_k - \mathbf{y}_{k'}| \ll L$ , we have*

$$\frac{\widehat{\mathbf{g}}_0^*(\mathbf{y}_k) \widehat{\mathbf{g}}_0(\mathbf{y}_{k'})}{\|\widehat{\mathbf{g}}_0(\mathbf{y}_k)\|_{\ell_2} \|\widehat{\mathbf{g}}_0(\mathbf{y}_{k'})\|_{\ell_2}} \approx \text{sinc}(\kappa |\mathbf{y}_k - \mathbf{y}_{k'}|) \sim \frac{1}{\kappa |\mathbf{y}_k - \mathbf{y}_{k'}|}, \quad (5.1)$$

where  $\kappa = \omega/c_0$ . Hence, the mutual coherence condition of the sensing matrix for spherical arrays is improved at the rate of the pixel size relative to the wavelength.

The second result is on the estimate when a planar array is used.

**Proposition 5.2.** *Assume a planar array of finite size and let  $\mathbf{y}_k$  and  $\mathbf{y}_{k'}$  be two points within the IW such that  $\lambda \ll |\mathbf{y}_k - \mathbf{y}_{k'}| \ll L$ . Then, we have*

$$\frac{\widehat{\mathbf{g}}_0^*(\mathbf{y}_k)\widehat{\mathbf{g}}_0(\mathbf{y}_{k'})}{\|\widehat{\mathbf{g}}_0(\mathbf{y}_k)\|_{\ell_2}\|\widehat{\mathbf{g}}_0(\mathbf{y}_{k'})\|_{\ell_2}} \sim \frac{1}{\sqrt{\kappa|\mathbf{y}_k - \mathbf{y}_{k'}|}}, \quad (5.2)$$

where  $\kappa = \omega/c_0$ . Hence, the mutual coherence condition of the sensing matrix for planar arrays is improved at the rate of square root of the pixel size relative to the wavelength.

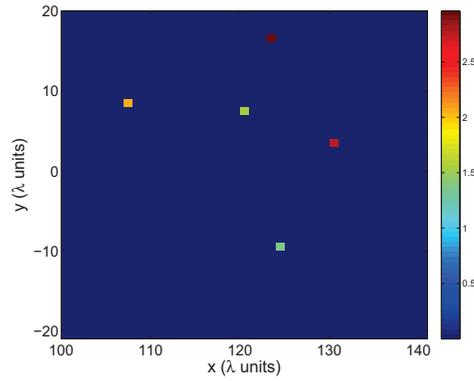
Based on these results, the upper bound of (3.9) is smaller for spherical arrays than for planar arrays. The pixel size of the IW with which good images are obtained is smaller for spherical arrays than for planar arrays. According to the analyses in §3 and §4, array imaging with spherical arrays can then locate more scatterers with higher resolution and is more robust with respect to the additive noise than array imaging with planar arrays, provided all other conditions are identical. This observation is supported by the numerical experiments.

**6. Numerical simulation.** In this section we present numerical simulations in two dimensions. The linear array consists of 100 transducers that are one wavelength  $\lambda$  apart. Five scatterers are placed within an IW of size  $41\lambda \times 41\lambda$  which is at a distance  $L = 100\lambda$  from the linear array. The amplitudes of the reflectivities of the scatterers,  $|\alpha_j|$ , are 2.96, 2.76, 2.05, 1.54 and 1.35 (see Fig. 6.1). Their phases are set randomly in each realization. We note that, given an illumination vector  $\widehat{\mathbf{f}}$  and a scatterer configuration  $\boldsymbol{\rho}_0$  with fixed amplitudes, the exact amount of multiple scattering depends on the realization of the phases in  $\boldsymbol{\rho}_0$ . For the amplitudes of the reflectivities chosen here, the amount of multiple scattering, defined by

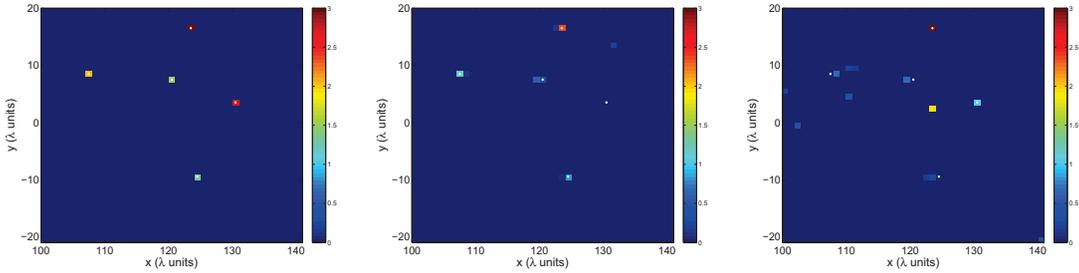
$$\frac{\|\widehat{\mathbf{P}} - \widehat{\mathbf{P}}_{ss}\|_F}{\|\widehat{\mathbf{P}}_{ss}\|_F} \times 100, \quad (6.1)$$

typically ranges between 50% and 100% in the simulations shown below. In (6.1),  $\widehat{\mathbf{P}}_{ss}$  is the response matrix without multiple scattering, computed by replacing  $\mathcal{G}_{FL}^T(\boldsymbol{\rho}_0)$  by  $\mathcal{G}^T$  in (2.14), i.e.,  $\widehat{\mathbf{P}}_{ss} = \mathcal{G} \text{diag}(\boldsymbol{\rho}) \mathcal{G}^T$ .

The five scatterers are within an IW that is discretized using a uniform grid with points separated by one wavelength  $\lambda$ . This results in a  $41 \times 41$  uniform mesh. Hence, we have 1681 unknowns and 100 measurements. Upon solution of (2.5), we generate the data received on the array using (2.3) and (2.2). We assume that all the scatterers are at one of the grid points. When off-grid scatterers exist, the mismatch is treated as noise in the data. Note that, in these simulations, the distance between grid points is small, of the order of the resolution limit dictated by the imaging setup. With this grid size, the mutual coherence (3.9) of the sensing matrix  $\mathcal{G}$  has a numerical value equal to 0.98. This, together with  $M = 5$  scatterers, clearly violates the sufficient condition for stable reconstruction required by formulations using either single illumination or multiple illuminations. However, this condition is quite conservative and we will show that the images are still good when the noise level is low in the data. Finally, we note that the obtained images depend on the realization of the random phases of the scatterers. In all the images shown below, we do not display the ones with the best quality we have seen in our numerical study. Moreover, in all the images shown, we normalize the spatial units by the wavelength  $\lambda$ .



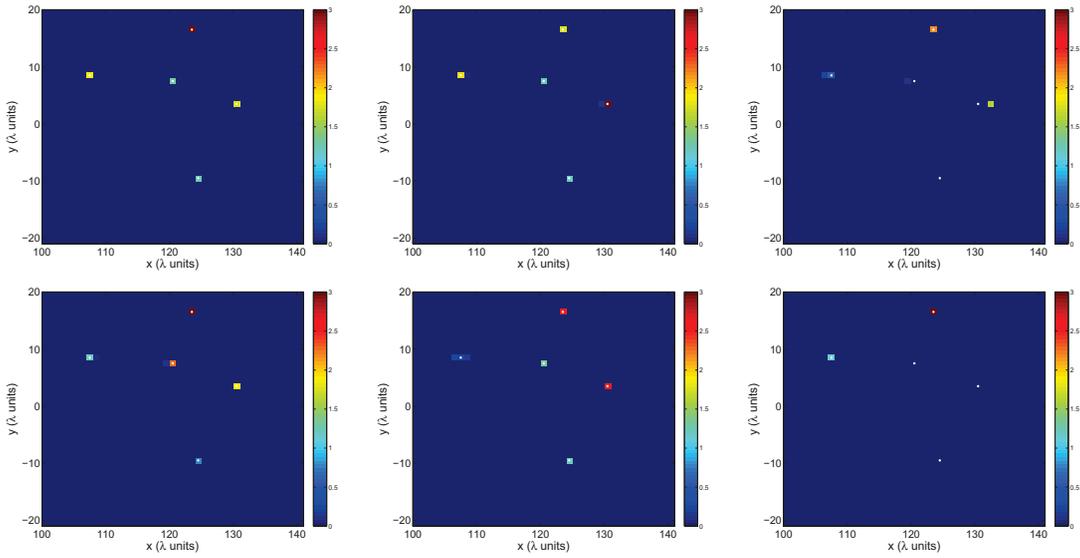
**Figure 6.1.** Original configuration of the scatterers in a  $41 \times 41$  image window with grid points separated by 1. The amplitudes of the reflectivities of the scatterers,  $|\alpha_j|$ , are 2.96, 2.76, 2.05, 1.54 and 1.35.



**Figure 6.2.** Images reconstructed by solving (3.4) and (3.5) when single illumination is used. From left to right, there is 0%, 10%, and 20% noise in the data.

Figure 6.2 shows the results of  $\ell_1$  norm minimization with 0% (left), 10% (middle) and 20% noise (right) when a single illumination coming from the center of the array is used. The exact locations of the scatterers in these images are indicated with small white dots. When there is no noise in the data,  $\ell_1$  norm minimization recovers the positions and reflectivities of the scatterers accurately. However, when 10% and 20% of noise is added to the data, the method fails to recover some of the scatterers and the images show some ghosts. Note that some scatterers are missing in the middle and right images of Figure 6.2.

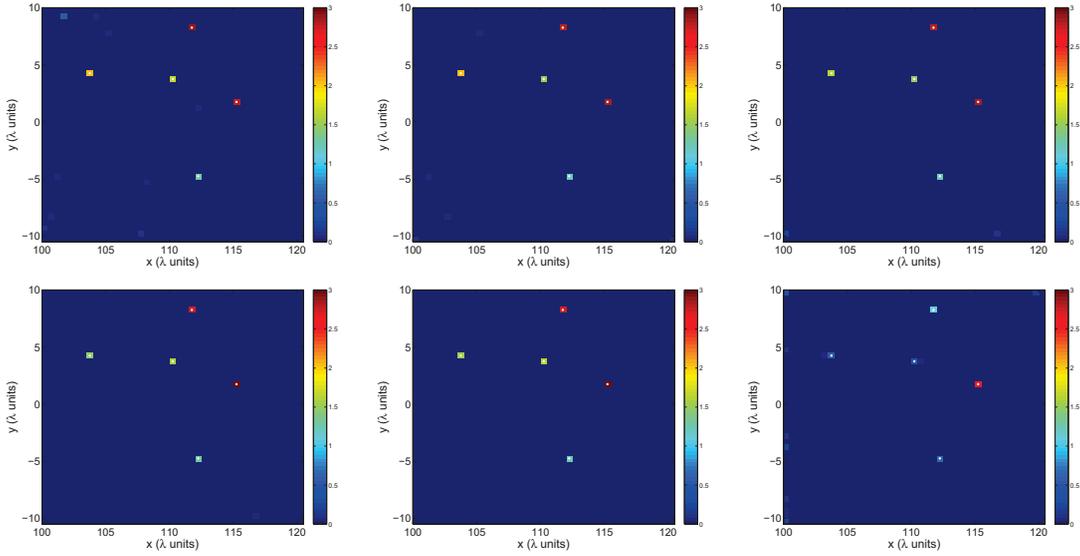
In order to stabilize the images, we study the improvement of the results when data collected with multiple illuminations are used. We consider first the case where the illuminations are randomly selected. By random illuminations we mean several illuminations coming, each one, from only one of the transducers on the array at a time, i.e.,  $\hat{f}_p = 1$  and  $\hat{f}_q = 0$  for  $q \neq p$ , with  $p$  chosen randomly at a time. Figure 6.3 shows the results of the MMV algorithm when 5 (top row) and 15 (bottom row) random illuminations are used. Additive noise at level 10% (left column), 20% (middle column) and 50% (right column) is added to the data in these numerical experiments. As expected, the images obtained with multiple illuminations are more stable with respect to additive noise. In fact, only a small number of illuminations are needed to improve the imaging performance significantly. However, it is not always true in general that more random illuminations always lead to better images. For instance, the image



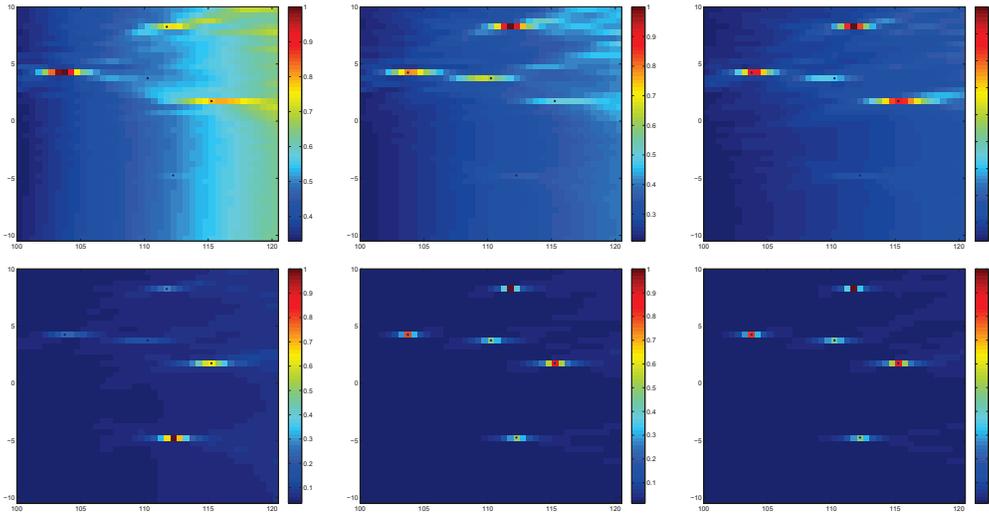
**Figure 6.3.** Images reconstructed by solving (4.6) when 5 (top row) and 15 (bottom row) random illuminations are used. From left to right, there is 10%, 20%, and 50% noise in the data.

obtained with 20% noise and 15 random illuminations (middle image of the bottom row) is worse than that obtained with 20% noise and 5 random illuminations (middle image of the top row). This is so because the illuminations are chosen randomly and “good” illuminations that lead to enough data diversity cannot be guaranteed.

Figure 6.3 indicates the importance of selecting “good” illuminations in the MMV formulation and suggests the use of optimal illuminations, especially when the signal-to-noise ratio (SNR) is low. Using optimal illuminations means taking  $\tilde{\mathbf{f}}^j = \widehat{\mathbf{V}}_{\cdot j}$ ,  $j = 1, \dots, M$ , as illuminations. These vectors can be obtained through the SVD of the array response matrix  $\widehat{\mathbf{P}}$  or by iterative time reversal. Note that, by choosing the illuminations optimally, we (i) maximize the data diversity, which also means low unnecessary redundancy of the multiple illuminations; and (ii) we reduce the noisy terms contained in the data matrix  $\mathbf{B}$ . On the other hand, we point out that, in principle, this approach would require the prior knowledge of the number of scatterers  $M$  if the noise level is high and is difficult to determine the singular values that correspond to the signal space. Hence, it is important to investigate the robustness of this approach with respect to the number of optimal illuminations used in the scheme. In Figure 6.4 we display the results when an increasing number of optimal illuminations are used. From left to right, and from top to bottom, we use 1, 2, 3, 4, 5, and 12 illuminations associated to the corresponding singular vectors  $\widehat{\mathbf{V}}_{\cdot j}$ , with  $j = 1, 2, 3, 4, 5, 12$ . We observe that this approach is very robust with respect to the number of optimal illuminations used. It is remarkable that only a few of them (2 or 3) are enough to achieve a significant improvement. Furthermore, using many more singular vectors as illuminations does not deteriorate too much the images (see the right image in the bottom row, where 12 illuminations are use). Finally, we point out that when multiple scattering is negligible all the significant singular vectors are necessary as shown in [5]. In that case, each optimal illumination aims at one scatterer at a



**Figure 6.4.** Images reconstructed by solving (4.6) when optimal illuminations are used. There is 50% noise in the data. From left to right and top to bottom, images are reconstructed by using 1, 2, 3, 4, 5, and 12 top singular vectors.



**Figure 6.5.** Images reconstructed by MUSIC. There is 50% noise in the data. From left to right and top to bottom, images are reconstructed by using 1, 2, 3, 4, 5, and 12 top singular vectors.

time, provided that the array is large enough.

We compare the images obtained with the MMV formulation and optimal illuminations, with those obtained with MUSIC. These two methods assume knowledge of the significant singular vectors of the response matrix  $\hat{\mathbf{P}}$  to form the images. Thus, the comparison between these two methods is carried out with about the same amount of data. In Figure 6.5 we show the images obtained with MUSIC when an increasing number of significant singular vectors of

the response matrix  $\widehat{\mathbf{P}}$  are used. As expected, we observe that MUSIC needs to use a number of singular vectors equal or greater to the number of scatterers. This is so, because MUSIC is a subspace based imaging technique that needs all the significant singular vectors to span the signal spaced. In other words, the complementary space representing the noise has to be correctly constructed such that no true signals should fall into. We also observe that the images formed with MUSIC do not change much when the number of singular vectors used exceeds the number of scatterers. These differences between MMV and MUSIC when both use the essential data of the response matrix  $\widehat{\mathbf{P}}$  is also observed in [18] in the context of DOA for the localization of sources with sensor arrays.

**7. Conclusion.** We give a novel approach to imaging localized scatterers with non-negligible multiple scattering between them. Our approach is non-iterative and solves the problem in two steps using sparsity promoting optimization. The uniqueness and stability of the formulations using both single and multiple illuminations are analyzed. We also propose to apply optimal illuminations to improve the robustness of the imaging methods and the resolution of the images. We show that the conditions under which the proposed methods work well are related to the configuration of the imaging problems: spherical arrays are in general more favorable than planar arrays. We illustrated the theoretical results with various numerical examples.

**Appendix A. Proof of Theorem 4.1.** We prove Theorem 4.1 by proving a more general result given below.

**Theorem A.1.** *For a given array configuration, assume that the resolution of the IW satisfies (3.6). If the number of scatterers  $M$  is such that  $M\epsilon < 1/2$ , then  $\mathbf{X}_0 = [\gamma_0^1 \dots \gamma_0^\nu]$  is the unique solution to*

$$\min J_{p,1}(\mathbf{X}) \quad \text{s.t.} \quad \mathcal{G}\mathbf{X} = \mathbf{B}$$

for any  $1 < p < \infty$ .

Clearly, Theorem 4.1 is the special case of Theorem A.1 for  $p = 2$ . The proof of Theorem A.1 is an application of the following result which is a generalization of Theorem 3.1 in [11].

**Proposition A.2.** *Let  $\Lambda$  be the set of the row support of  $\mathbf{X}_0$ , i.e.  $\Lambda = \text{rowsupp}(\mathbf{X}_0)$ . For any matrix  $\mathbf{Q}$  and any  $1 < p < \infty$ , define the matrix  $\text{sign}(\mathbf{Q})$  as the matrix whose entries are*

$$\text{sign}(Q_{ij}) = \begin{cases} \frac{|Q_{ij}|^{p-1} \text{sign}(Q_{ij})}{\|Q_{i,\cdot}\|_{\ell_q}^{p-1}}, & \|Q_{i,\cdot}\|_{\ell_q} \neq 0 \\ 0, & \|Q_{i,\cdot}\|_{\ell_q} = 0. \end{cases}$$

*In this definition,  $\text{sign}(x) = 0, \pm 1$  when  $x \in \mathbb{R}$  and  $\text{sign}(x) = \exp(i \text{angle}(x))$  when  $x \in \mathbb{C}$ . Assume that  $\mathcal{G}_\Lambda$ , the submatrix of  $\mathcal{G}$  consisting of columns with indices in  $\Lambda$ , is full rank. Then, a sufficient condition under which  $\mathbf{X}_0$  is the unique solution to*

$$\min_{\mathbf{X}} J_{p,1}(\mathbf{X}) \quad \text{s.t.} \quad \mathcal{G}\mathbf{X} = \mathbf{B}, \tag{A.1}$$

*is that there exists a matrix  $\mathbf{H} \in \mathbb{C}^{N \times \nu}$  satisfying*

$$\mathcal{G}_\Lambda^* \mathbf{H} = \text{sign}(\mathbf{X}_{0\Lambda}) \tag{A.2}$$

and

$$\|\mathbf{H}^* \widehat{\mathbf{g}}_0(\mathbf{y}_j)\|_{\ell_q} < 1, \quad j \notin \Lambda, \quad (\text{A.3})$$

where  $1/p + 1/q = 1$ , and  $\mathbf{X}_{0\Lambda}$  is the submatrix consisting of the rows of  $\mathbf{X}_0$  in  $\Lambda$ .

To prove Proposition A.2, we need the following lemma.

**Lemma A.3.** For any two matrices  $A \in \mathbb{C}^{m \times l}$  and  $B \in \mathbb{C}^{l \times n}$ , we have

$$|\text{trace}(AB)| \leq \max_{k=1, \dots, l} \|B_{\cdot k}\|_{\ell_q} J_{p,1}(A).$$

The strict inequality holds when there exists an index  $k$  such that  $\|B_{\cdot k}\|_{\ell_q} < \max_{k=1, \dots, l} \|B_{\cdot k}\|_{\ell_q}$  and  $\|A_{k \cdot}\|_{\ell_p} \neq 0$ .

*Proof.* By definition of trace, we have

$$|\text{trace}(AB)| \leq \sum_{k=1}^l |A_{k \cdot} B_{\cdot k}| \leq \sum_{k=1}^l \|A_{k \cdot}\|_{\ell_p} \|B_{\cdot k}\|_{\ell_q} \leq \max_{k=1, \dots, l} \|B_{\cdot k}\|_{\ell_q} J_{p,1}(A)$$

where we use the Hölder's inequality in the second inequality. The strict inequality clearly holds when the condition in the lemma is satisfied. ■

*Proof.* [Proof of Proposition A.2] We will show the uniqueness by contradiction. Given  $\mathbf{X}_0$ , assume there exists a matrix  $\mathbf{H}$  satisfying conditions (A.2) and (A.3), and assume that the solution to the MMV problem (A.1) is not unique, i.e. there exists another solution  $\widehat{\mathbf{X}} \neq \mathbf{X}_0$  with support  $\widehat{\Lambda}$ . Since  $\mathcal{G}_\Lambda$  is full rank, it implies that  $\widehat{\Lambda}$  cannot be a subset of  $\Lambda$ , i.e.  $\widehat{\Lambda} \setminus \Lambda \neq \emptyset$ . Based on the definition of the sign function of a matrix, we have

$$J_{p,1}(\mathbf{X}_0) = J_{p,1}(\mathbf{X}_{0\Lambda}) = \text{trace}(\text{sign}(\mathbf{X}_{0\Lambda}) \mathbf{X}_{0\Lambda}^*) = \text{trace}(\mathcal{G}_\Lambda^* \mathbf{H} \mathbf{X}_{0\Lambda}^*),$$

for all  $1 < p < \infty$ . Since the trace function is invariant with respect to matrix rotations and the transpose operation,

$$J_{p,1}(\mathbf{X}_0) = \text{trace}(\mathbf{X}_{0\Lambda} \mathbf{H}^* \mathcal{G}_\Lambda) = \text{trace}(\mathbf{H}^* \mathcal{G}_\Lambda \mathbf{X}_{0\Lambda}) = \text{trace}(\mathbf{H}^* \mathcal{G} \widehat{\mathbf{X}}),$$

where the last equality is true because  $\mathcal{G}_\Lambda \mathbf{X}_{0\Lambda} = \mathcal{G} \mathbf{X}_0 = \mathbf{B} = \mathcal{G} \widehat{\mathbf{X}}$ . Applying Lemma A.3 to  $\text{trace}(\mathbf{H}^* \mathcal{G} \widehat{\mathbf{X}})$ , we have

$$J_{p,1}(\mathbf{X}_0) = |\text{trace}(\mathbf{H}^* \mathcal{G} \widehat{\mathbf{X}})| \leq \max_{j \in \widehat{\Lambda}} \|\mathbf{H}^* \widehat{\mathbf{g}}_0(\mathbf{y}_j)\|_{\ell_q} J_{p,1}(\widehat{\mathbf{X}}).$$

If  $\widehat{\Lambda} \cap \Lambda \neq \emptyset$ , there exists at least one  $j_0$  satisfying  $j_0 \in \widehat{\Lambda}$  and  $j_0 \notin \Lambda$ . Due to (A.3),  $\|\mathbf{H}^* \widehat{\mathbf{g}}_0(\mathbf{y}_{j_0})\|_{\ell_q} < 1 = \max_{j \in \widehat{\Lambda}} \|\mathbf{H}^* \widehat{\mathbf{g}}_0(\mathbf{y}_j)\|_{\ell_q}$ . According to Lemma A.3,

$$J_{p,1}(\mathbf{X}_0) = |\text{trace}(\mathbf{H}^* \mathcal{G} \widehat{\mathbf{X}})| < \max_{j \in \widehat{\Lambda}} \|\mathbf{H}^* \widehat{\mathbf{g}}_0(\mathbf{y}_j)\|_{\ell_q} J_{p,1}(\widehat{\mathbf{X}}) = J_{p,1}(\widehat{\mathbf{X}}).$$

If  $\widehat{\Lambda} \cap \Lambda = \emptyset$ , then for any  $j \in \widehat{\Lambda}$ ,  $j \notin \Lambda$ . According to (A.3), we have  $\max_{j \in \widehat{\Lambda}} \|\mathbf{H}^* \widehat{\mathbf{g}}_0(\mathbf{y}_j)\|_{\ell_q} < 1$ . Therefore,  $J_{p,1}(\mathbf{X}_0) \leq \max_{j \in \widehat{\Lambda}} \|\mathbf{H}^* \widehat{\mathbf{g}}_0(\mathbf{y}_j)\|_{\ell_q} J_{p,1}(\widehat{\mathbf{X}}) < J_{p,1}(\widehat{\mathbf{X}})$ . In either case,  $J_{p,1}(\mathbf{X}_0) < J_{p,1}(\widehat{\mathbf{X}})$  and, hence, it contradicts that  $\widehat{\mathbf{X}}$  is also a solution of (A.1). ■

Now we will show that the multiplier  $\mathbf{H}$  satisfying (A.2) and (A.3) exists under the condition of Theorem A.1.

*Proof.* [Proof of Theorem A.1] Let  $\Lambda = \{n_j, 1 \leq j \leq M\}$  be the set of indices corresponding to the scatterers. Based on the resolution condition, we have that the inner product of the column vectors of the matrix  $\mathcal{G}$  satisfies  $\widehat{\mathbf{g}}_0^*(\mathbf{y}_i)\widehat{\mathbf{g}}_0(\mathbf{y}_j) = \delta_{ij} + (1 - \delta_{ij})\epsilon_{ij}$  with  $|\epsilon_{ij}| < \epsilon$ , for any  $1 \leq i, j \leq M$ . Therefore, the submatrix  $\mathcal{G}_\Lambda$ , composed of the columns  $n_1, \dots, n_M$  of matrix  $\mathcal{G}$ , is full column rank and satisfies that  $\mathcal{G}_\Lambda^* \mathcal{G}_\Lambda$  is full rank and diagonally dominant.

According to Proposition A.2, we need to find a matrix  $\mathbf{H}$  satisfying (A.2) and (A.3). Let  $\mathbf{H} = \mathcal{G}_\Lambda (\mathcal{G}_\Lambda^* \mathcal{G}_\Lambda)^{-1} \text{sign}(\mathbf{X}_{0\Lambda})$ . Then, the first condition is automatically satisfied because  $\mathcal{G}_\Lambda^* \mathbf{H} = \text{sign}(\mathbf{X}_{0\Lambda})$ . For the second condition, choosing any column  $j$  of  $\mathcal{G}$  not in the submatrix  $\mathcal{G}_\Lambda$ , we have

$$\begin{aligned} \|\widehat{\mathbf{g}}_0^*(\mathbf{y}_j)\mathcal{G}_\Lambda (\mathcal{G}_\Lambda^* \mathcal{G}_\Lambda)^{-1} \text{sign}(\mathbf{X}_{0\Lambda})\|_{\ell_q} &\leq \|\text{sign}(\mathbf{X}_{0\Lambda})\|_{p \rightarrow \infty} \|(\mathcal{G}_\Lambda^* \mathcal{G}_\Lambda)^{-1}\|_{\ell_1} \|\mathcal{G}_\Lambda^* \widehat{\mathbf{g}}_0(\mathbf{y}_j)\|_{\ell_1} \\ &\leq \frac{M\epsilon}{1 - M\epsilon} < 1, \end{aligned}$$

where by definition, we have  $\|\text{sign}(\mathbf{X}_{0\Lambda})\|_{p \rightarrow \infty} \leq 1$  and the last inequality is due to the resolution condition  $M\epsilon < 1/2$ . ■

**Appendix B. Proof of Theorem 4.3.** In [28], the author gives conditions for the MMV problem under which the convex relaxation formulation with functional  $J_{\infty,1}$  is robust with respect to the additive noise. In the following, we derive similar conditions for convex relaxation using  $J_{2,1}$  instead of  $J_{\infty,1}$ . This is done using the techniques developed in [25]. We first introduce some supporting results.

**Definition B.1.** Let  $f$  be a function from the complex matrix space  $\mathcal{M}$  to  $\mathbb{C}$ . The subdifferential of a function  $f$  at  $\mathbf{X} \in \mathcal{M}$  is defined as

$$\partial f(\mathbf{X}) = \{\mathbf{G} \in \mathcal{M} : f(\mathbf{Y}) \geq f(\mathbf{X}) + \text{Re} \langle \mathbf{Y} - \mathbf{X}, \mathbf{G} \rangle, \forall \mathbf{Y} \in \mathcal{M}\}.$$

**Lemma B.2.** A matrix  $\mathbf{G}$  lies in the subdifferential of  $J_{2,1}(\mathbf{X})$  at  $\mathbf{X} \in \mathbb{C}^{K \times \nu}$  if and only if its rows  $G_i \in \mathbb{C}^\nu$  satisfy

- $G_i = \frac{X_i}{\|X_i\|_{\ell_2}}$  when  $X_i \neq 0$ , and
- $\|G_i\|_{\ell_2} \leq 1$  when  $X_i = 0$ .

Moreover,  $\mathbf{G}$  is called a subgradient of  $J_{2,1}(\mathbf{X})$  at  $\mathbf{X}$ .

Then according to the definition of matrix norms, it is easy to see the subgradient of  $J_{2,1}$  satisfies the following.

**Corollary B.3.** Any subgradient  $\mathbf{G}$  of  $J_{2,1}$  satisfies  $\|\mathbf{G}\|_{2 \rightarrow \infty} = \|\mathbf{G}^*\|_{1,2} \leq 1$ .

We now give a result related to operator norms of matrices which will be used later.

**Lemma B.4.** Let  $\mathbf{A}$  be an  $m \times n$  matrix and define operator norm  $\|\cdot\|_{(2,\infty) \rightarrow F}$  as

$$\|\mathbf{A}\|_{(2,\infty) \rightarrow F} = \max_{\mathbf{C} \in \mathbb{C}^{n \times d}} \frac{\|\mathbf{A}\mathbf{C}\|_F}{\|\mathbf{C}\|_{2 \rightarrow \infty}}.$$

Then we have  $\|\mathbf{A}\|_{(2,\infty) \rightarrow F} \leq \|\mathbf{A}\|_{\infty \rightarrow 2}$ .

*Proof.* Since we have

$$\begin{aligned}
\left(\frac{\|\mathbf{A}\mathbf{C}\|_F}{\|\mathbf{C}\|_{2 \rightarrow \infty}}\right)^2 &= \frac{\sum_{i=1}^m \sum_{j=1}^d \left| \sum_{k=1}^n A_{ik} C_{kj} \right|^2}{\max_{1 \leq i \leq n} \|C_{i\cdot}\|_{\ell_2}^2} \\
&\leq \frac{\sum_{i=1}^m \left| \sum_{k=1}^n A_{ik} \|C_{k\cdot}\|_{\ell_2} \right|^2}{\max_{1 \leq i \leq n} \|C_{i\cdot}\|_{\ell_2}^2} \\
&\leq \max_{\mathbf{c} \in \mathbb{R}^n} \left( \frac{\|\mathbf{A}\mathbf{c}\|_{\ell_2}}{\|\mathbf{c}\|_{\infty}} \right)^2 \\
&\leq \|\mathbf{A}\|_{\infty \rightarrow 2}^2,
\end{aligned}$$

it is clearly that  $\|\mathbf{A}\|_{(2 \rightarrow \infty) \rightarrow F} \leq \|\mathbf{A}\|_{\infty \rightarrow 2}$ . In the derivation above, the first inequality is true because for each row index  $i$ , we have

$$\begin{aligned}
\sum_{j=1}^d \left| \sum_{k=1}^n A_{ik} C_{kj} \right|^2 &= \sum_{j=1}^d \sum_{k,k'=1}^n A_{ik} C_{kj} \bar{A}_{ik'} \bar{C}_{k'j} \\
&= \sum_{k,k'=1}^n A_{ik} \bar{A}_{ik'} \sum_{j=1}^d C_{kj} \bar{C}_{k'j} \\
&= \sum_{k,k'=1}^n A_{ik} \bar{A}_{ik'} \langle C_{k\cdot}, C_{k'\cdot} \rangle \\
&\leq \sum_{k,k'=1}^n A_{ik} \bar{A}_{ik'} \|C_{k\cdot}\|_{\ell_2} \|C_{k'\cdot}\|_{\ell_2} \\
&= \left| \sum_{k=1}^n A_{ik} \|C_{k\cdot}\|_{\ell_2} \right|^2.
\end{aligned}$$

■

To make the following discussion easier, we assume that the sensing matrix  $\mathcal{G}$  has normalized columns and we introduce some additional notations used in this appendix only. Let  $\Lambda$  be indexes of a subset of linearly independent columns of  $\mathcal{G}$ , i.e.  $\Lambda \subset \{1, 2, \dots, K\}$  such that the Green's function vectors  $\hat{\mathbf{g}}_0(\mathbf{y}_j)$ , with  $j \in \Lambda$ , are linearly independent. We denote by  $\mathcal{G}_\Lambda \in \mathbb{C}^{N \times |\Lambda|}$  the submatrix of  $\mathcal{G}$  composed of columns with indices in  $\Lambda$ , by  $\mathbf{B}_\Lambda \in \mathbb{C}^{N \times \nu}$  the best Frobenius norm approximation of the data matrix  $\mathbf{B}$  over  $\Lambda$ , and by  $\mathbf{X}_{0\Lambda} \in \mathbb{C}^{|\Lambda| \times \nu}$  the corresponding coefficient matrix synthesizing  $\mathbf{B}_\Lambda$ , i.e. such that  $\mathbf{B}_\Lambda = \mathcal{G}_\Lambda \mathbf{X}_{0\Lambda}$ . Note that  $\mathbf{X}_{0\Lambda} = \mathcal{G}_\Lambda^\dagger \mathbf{B}_\Lambda$ , with  $\mathcal{G}_\Lambda^\dagger = (\mathcal{G}_\Lambda^* \mathcal{G}_\Lambda)^{-1} \mathcal{G}_\Lambda^*$ .

Next, we give several results related to the minimizers of the Lagrange function (4.11). The proofs are analogue to those in [28] and will be skipped. Interested readers can refer to [28] or [25] for single measurement case. The first lemma is on the condition of the minimizer of (4.11).

**Lemma B.5.** *Suppose that the matrix  $\mathbf{X}_\star$  is the minimizer of (4.11) over all matrices with row-support  $\Lambda$ . A necessary and sufficient condition for  $\mathbf{X}_\star$  to be such minimizer is that*

$$\mathbf{X}_{0\Lambda} - \mathbf{X}_\star = \lambda(\mathcal{G}_\Lambda^* \mathcal{G}_\Lambda)^{-1} \mathbf{G}, \quad (\text{B.1})$$

where  $\mathbf{G} \in \partial J_{2,1}(\mathbf{X}_\star)$ . Moreover, the minimizer is unique.

Using Lemmas B.4 and B.5, we have the following estimates on the bound of the minimizer of (4.11) over a specific support.

**Lemma B.6.** *Suppose that the matrix  $\mathbf{X}_\star$  is the unique minimizer of (4.11) over all matrices with support inside  $\Lambda$ . Then, the following estimates hold:*

$$\|\mathbf{X}_{0\Lambda} - \mathbf{X}_\star\|_{2 \rightarrow \infty} \leq \lambda \|(\mathcal{G}_\Lambda^* \mathcal{G}_\Lambda)^{-1}\|_{2 \rightarrow \infty}, \quad (\text{B.2})$$

$$\|\mathcal{G}_\Lambda(\mathbf{X}_{0\Lambda} - \mathbf{X}_\star)\|_F \leq \lambda \|\mathcal{G}_\Lambda^\dagger\|_{2 \rightarrow 1}. \quad (\text{B.3})$$

The above results are on the bounds of the error between  $\mathbf{X}_\star$  and the “true” solution  $\mathbf{X}_{0\Lambda}$  when the search is restricted to a given support  $\Lambda$ . We now give a condition under which the solution  $\mathbf{X}_\star$  to (4.11) will be supported on  $\Lambda$ . For this condition, we need to use the *Exact Recovery Coefficient*

$$ERC(\Lambda) = 1 - \max_{j \notin \Lambda} \|\mathcal{G}_\Lambda^\dagger \widehat{\mathbf{g}}(\mathbf{y}_j)\|_{\ell_1}, \quad (\text{B.4})$$

introduced in [25], which measures the orthogonality between the column vectors used in  $\mathcal{G}_\Lambda$  and the remaining column vectors.

**Lemma B.7.** *Under the same condition as in Lemma B.6, if the following condition holds*

$$\|\mathcal{G}^*(\mathbf{B} - \mathbf{B}_\Lambda)\|_{2 \rightarrow \infty} \leq \lambda ERC(\Lambda), \quad (\text{B.5})$$

then the unique minimizer  $\mathbf{X}_\star$  of (4.11) is supported on  $\Lambda$ .

*Proof.* By definition,  $\mathbf{B}_\Lambda = \mathcal{G}_\Lambda \mathbf{X}_{0\Lambda}$ . Given any vector  $\mathbf{u} \in \mathbb{C}^\nu$ , we have for any  $j \notin \Lambda$ ,

$$|\langle \widehat{\mathbf{g}}^*(\mathbf{y}_j)(\mathbf{B} - \mathcal{G}_\Lambda \mathbf{X}_{0\Lambda}), \mathbf{u} \rangle| = |\langle \widehat{\mathbf{g}}^*(\mathbf{y}_j)(\mathbf{B} - \mathbf{B}_\Lambda), \mathbf{u} \rangle| \leq \|(\mathbf{B} - \mathbf{B}_\Lambda)^* \widehat{\mathbf{g}}(\mathbf{y}_j)\|_{\ell_2} \|\mathbf{u}\|_{\ell_2} \quad (\text{B.6})$$

and

$$|\langle \widehat{\mathbf{g}}^*(\mathbf{y}_j) \mathcal{G}_\Lambda(\mathbf{X}_{0\Lambda} - \mathbf{X}_\star), \mathbf{u} \rangle| = |\langle \widehat{\mathbf{g}}^*(\mathbf{y}_j) \mathcal{G}_\Lambda(\mathcal{G}_\Lambda^* \mathcal{G}_\Lambda)^{-1} \mathbf{G}, \mathbf{u} \rangle| \leq \|\mathbf{G}^* \mathcal{G}_\Lambda^\dagger \widehat{\mathbf{g}}(\mathbf{y}_j)\|_{\ell_2} \|\mathbf{u}\|_{\ell_2}. \quad (\text{B.7})$$

Since  $\mathbf{X}_\star$  is the unique minimizer among all set of matrices with support included in  $\Lambda$ , we only need to show that it is also the optimal solution among matrices with support larger than  $\Lambda$ . Let  $\boldsymbol{\zeta} \in \mathbb{C}^K$  be a standard unit vector with support on  $\{1, \dots, K\} \setminus \Lambda$ . Then,  $\mathbf{X}_\star + \boldsymbol{\zeta} \mathbf{u}^*$  is a perturbation by adding a matrix with row support disjoint from that of  $\mathbf{X}_\star$ . If we compute

the variation of (4.11) with respect to this perturbation, we obtain

$$\begin{aligned}
L(\mathbf{X}_\star + \zeta \mathbf{u}^\star, \lambda) - L(\mathbf{X}_\star, \lambda) &= \frac{1}{2} (\|\mathbf{B} - \mathcal{G}\mathbf{X}_\star - \widehat{\mathbf{g}}(\mathbf{y}_j) \mathbf{u}^\star\|_F^2 - \|\mathbf{B} - \mathcal{G}\mathbf{X}_\star\|_F^2) + \\
&\quad \lambda (J_{2,1}(\mathbf{X}_\star + \zeta \mathbf{u}^\star) - J_{2,1}(\mathbf{X}_\star)) \\
&= \frac{1}{2} \|\widehat{\mathbf{g}}(\mathbf{y}_j) \mathbf{u}^\star\|_F^2 - \operatorname{Re} \langle \mathbf{B} - \mathcal{G}\mathbf{X}_\star, \widehat{\mathbf{g}}(\mathbf{y}_j) \mathbf{u}^\star \rangle + \lambda \|\mathbf{u}\|_{\ell_2} \\
&= \frac{1}{2} \|\widehat{\mathbf{g}}(\mathbf{y}_j) \mathbf{u}^\star\|_F^2 - \operatorname{Re} \langle \widehat{\mathbf{g}}^\star(\mathbf{y}_j) (\mathbf{B} - \mathcal{G}_\Lambda \mathbf{X}_{0,\Lambda}), \mathbf{u}^\star \rangle - \\
&\quad \operatorname{Re} \langle \widehat{\mathbf{g}}^\star(\mathbf{y}_j) \mathcal{G}_\Lambda (\mathbf{X}_{0,\Lambda} - \mathbf{X}_\star), \mathbf{u}^\star \rangle + \lambda \|\mathbf{u}\|_{\ell_2} \\
&> \lambda \|\mathbf{u}\|_{\ell_2} - |\langle \widehat{\mathbf{g}}^\star(\mathbf{y}_j) (\mathbf{B} - \mathcal{G}_\Lambda \mathbf{X}_{0,\Lambda}), \mathbf{u}^\star \rangle| - |\langle \widehat{\mathbf{g}}^\star(\mathbf{y}_j) \mathcal{G}_\Lambda (\mathbf{X}_{0,\Lambda} - \mathbf{X}_\star), \mathbf{u}^\star \rangle| \\
&\geq \|\mathbf{u}\|_{\ell_2} \left( \lambda - \|(\mathbf{B} - \mathbf{B}_\Lambda)^\star \widehat{\mathbf{g}}(\mathbf{y}_j)\|_{\ell_2} - \lambda \|\mathbf{G}^\star \mathcal{G}_\Lambda^\dagger \widehat{\mathbf{g}}(\mathbf{y}_j)\|_{\ell_2} \right).
\end{aligned}$$

To show that  $L(\mathbf{X}_\star + \zeta \mathbf{u}^\star, \lambda) - L(\mathbf{X}_\star, \lambda) > 0$ , first observe that condition (B.5) implies that

$$\|(\mathbf{B} - \mathbf{B}_\Lambda)^\star \widehat{\mathbf{g}}(\mathbf{y}_j)\|_{\ell_2} \leq \|(\mathbf{B} - \mathbf{B}_\Lambda)^\star \mathcal{G}\|_{1 \rightarrow 2} = \|\mathcal{G}^\star (\mathbf{B} - \mathbf{B}_\Lambda)\|_{2 \rightarrow \infty} \leq \lambda \operatorname{ERC}(\Lambda),$$

and, at the same time, by the definition of  $\operatorname{ERC}(\Lambda)$  and using Corollary B.3, we obtain

$$\lambda \operatorname{ERC}(\Lambda) \leq \lambda (1 - \|\mathcal{G}_\Lambda^\dagger \widehat{\mathbf{g}}(\mathbf{y}_j)\|_{\ell_1}) \leq \lambda (1 - \|\mathcal{G}_\Lambda^\dagger \widehat{\mathbf{g}}(\mathbf{y}_j)\|_{\ell_1} \|\mathbf{G}\|_{2 \rightarrow \infty}) \leq \lambda (1 - \|\mathbf{G}^\star \mathcal{G}_\Lambda^\dagger \widehat{\mathbf{g}}(\mathbf{y}_j)\|_{\ell_2}).$$

Therefore,  $L(\mathbf{X}_\star + \zeta \mathbf{u}^\star, \lambda) > L(\mathbf{X}_\star, \lambda)$  which completes the proof. ■

With all the supportive results, we are now ready to prove our main result of MMV problem (4.6).

*Proof.* [Proof of Theorem 4.3] Let the support of the solution to (4.6),  $\mathbf{X}_0$ , be  $\Lambda_0$  with  $|\Lambda_0| = M$ . We denote the solution by  $\mathbf{X}_{\Lambda_0}$ , and the corresponding synthesized data matrix by  $\mathbf{B}_{\Lambda_0} = \mathcal{G}\mathbf{X}_{\Lambda_0}$ . Since (4.6) is convex, the necessary and sufficient condition for it to have a unique solution is that there exists a pair  $(\mathbf{X}_\star, \lambda_\star)$  such that the following KKT conditions are satisfied:

$$\mathbf{X}_\star = \operatorname{argmin}_{\mathbf{X}} L(\mathbf{X}, \lambda) = \frac{1}{2} \|\mathbf{B} - \mathcal{G}\mathbf{X}\|_F^2 + \lambda_\star J_{2,1}(\mathbf{X}), \quad (\text{B.8})$$

$$\|\mathbf{B} - \mathcal{G}\mathbf{X}_\star\|_F = \delta, \quad (\text{B.9})$$

$$\lambda_\star > 0. \quad (\text{B.10})$$

We first consider the following problem with additional requirement that the support is included in  $\Lambda_0$

$$\min_{\operatorname{rowsupp}(\mathbf{X}) \subset \Lambda_0} J_{2,1}(\mathbf{X}) \quad \text{s.t.} \quad \|\mathbf{B} - \mathcal{G}\mathbf{X}\|_F \leq \delta. \quad (\text{B.11})$$

Because  $\mathbf{B}_{\Lambda_0}$  is the best Frobenius norm approximation of  $\mathbf{B}$ , using Lemma B.6 we obtain

$$\begin{aligned}
\delta^2 &= \|\mathbf{B} - \mathcal{G}\mathbf{X}_\star\|_F^2 \\
&= \|\mathbf{B} - \mathbf{B}_{\Lambda_0}\|_F^2 + \|\mathbf{B}_{\Lambda_0} - \mathcal{G}\mathbf{X}_\star\|_F^2 \\
&= \|\mathbf{B} - \mathbf{B}_{\Lambda_0}\|_F^2 + \|\mathcal{G}(\mathbf{X}_{\Lambda_0} - \mathbf{X}_\star)\|_F^2 \\
&\leq \|\mathbf{B} - \mathbf{B}_{\Lambda_0}\|_F^2 + \lambda_\star^2 \|\mathcal{G}_{\Lambda_0}^\dagger\|_{2 \rightarrow 1}^2.
\end{aligned}$$

Thus, the second KKT condition (B.9) implies that

$$\lambda_\star^2 \geq \frac{\delta^2 - \|\mathbf{B} - \mathbf{B}_{\Lambda_0}\|_F^2}{\|\mathcal{G}_{\Lambda_0}^\dagger\|_{2 \rightarrow 1}^2}.$$

On the other hand, according to Lemma B.7,  $\mathbf{X}_\star$  has support on  $\Lambda_0$  if

$$\lambda_\star \geq \frac{\|\mathcal{G}^*(\mathbf{B} - \mathbf{B}_{\Lambda_0})\|_{2 \rightarrow \infty}}{ERC(\Lambda_0)}.$$

Therefore, as long as

$$\frac{\delta^2 - \|\mathbf{B} - \mathbf{B}_{\Lambda_0}\|_F^2}{\|\mathcal{G}_{\Lambda_0}^\dagger\|_{2 \rightarrow 1}^2} \geq \frac{\|\mathcal{G}^*(\mathbf{B} - \mathbf{B}_{\Lambda_0})\|_{2 \rightarrow \infty}^2}{ERC^2(\Lambda_0)},$$

$\mathbf{X}_\star$  is the optimal solution with support included in  $\Lambda_0$ . Rearranging the above inequality, we have

$$\delta^2 \geq \|\mathbf{B} - \mathbf{B}_{\Lambda_0}\|_F^2 + \frac{\|\mathcal{G}_{\Lambda_0}^\dagger\|_{2 \rightarrow 1}^2 \|\mathcal{G}^*(\mathbf{B} - \mathbf{B}_{\Lambda_0})\|_{2 \rightarrow \infty}^2}{ERC^2(\Lambda_0)}. \quad (\text{B.12})$$

By definition,

$$\|\mathcal{G}^*(\mathbf{B} - \mathbf{B}_{\Lambda_0})\|_{2 \rightarrow \infty}^2 = \left( \max_{1 \leq j \leq K} \|\widehat{\mathbf{g}}^*(\mathbf{y}_j)(\mathbf{B} - \mathbf{B}_{\Lambda_0})\|_{\ell_2} \right)^2 \leq \|\mathbf{B} - \mathbf{B}_{\Lambda_0}\|_F^2.$$

According to Propositions 3.7 and 3.9 in [25],

$$\frac{\|\mathcal{G}_{\Lambda_0}^\dagger\|_{2 \rightarrow 1}^2}{ERC^2(\Lambda_0)} \leq \frac{M(1 - (M - 1)\epsilon)}{(1 - 2M\epsilon + \epsilon)^2}.$$

Hence, we have

$$\|\mathcal{E}\|_F^2 \left( 1 + \frac{M(1 - (M - 1)\epsilon)}{(1 - 2M\epsilon + \epsilon)^2} \right) \geq \|\mathbf{B} - \mathbf{B}_{\Lambda_0}\|_F^2 + \frac{\|\mathcal{G}_{\Lambda_0}^\dagger\|_{2 \rightarrow 1}^2 \|\mathcal{G}^*(\mathbf{B} - \mathbf{B}_{\Lambda_0})\|_{2 \rightarrow \infty}^2}{ERC^2(\Lambda_0)}.$$

Therefore, condition (4.7) is sufficient for (B.12) to hold and  $\mathbf{X}_\star$  is the unique minimizer to (4.6) with support inside  $\Lambda_0$ .

Next we show that this minimizer over the support  $\Lambda_0$  is also the global minimizer to (4.6). Assume there exists another coefficient matrix  $\widehat{\mathbf{X}}$  which minimizes (4.6) and thus also satisfies the KKT conditions, especially (B.9). Then  $\mathcal{G}\mathbf{X}_\star = \widehat{\mathcal{G}}\widehat{\mathbf{X}}$  must hold. Assume this is not the case. Then since formulation (4.6) is convex, any linear combination of solutions will also be a solution. In particular,  $\frac{1}{2}(\mathbf{X}_\star + \widehat{\mathbf{X}})$  is a solution and should satisfy KKT condition (B.9). This is a contradiction because

$$\|\mathbf{B} - \frac{1}{2}\mathcal{G}\mathbf{X}_\star - \frac{1}{2}\widehat{\mathcal{G}}\widehat{\mathbf{X}}\|_F < \delta.$$

Now that both  $\mathbf{X}_*$  and  $\widehat{\mathbf{X}}$  minimize (4.6) with the same value  $\mathcal{G}\mathbf{X}_*$ . It implies that both solutions satisfy

$$\min_{\mathbf{X}} J_{2,1}(\mathbf{X}) \quad \text{s.t.} \quad \mathcal{G}\mathbf{X} = \mathcal{G}\mathbf{X}_*.$$

However, due to Theorem 4.1, when  $M\epsilon < 1/2$ , the above optimization has a unique solution. We then prove that  $\mathbf{X}_* = \widehat{\mathbf{X}}$ , i.e. the solution to (4.6) is unique.

Finally, the error bound of the minimizer compared to underlying solution is estimated as follows

$$\|\mathbf{X}_* - \mathbf{X}_0\|_F = \|(\mathcal{G}^*\mathcal{G})^{-1}\mathcal{G}^*\mathcal{G}(\mathbf{X}_* - \mathbf{X}_0)\|_F \leq \|\mathcal{G}_{\Lambda_0}^\dagger\|_{2 \rightarrow 2} \|\mathcal{G}(\mathbf{X}_* - \mathbf{X}_0)\|_F \leq \delta/\sqrt{1 - (M-1)\epsilon},$$

where we use the singular value estimate of  $\mathcal{G}_{\Lambda_0}$  given in [9] and [25]. Note that if  $\|(\mathbf{X}_0)_i\|_{\ell_2} > \delta/\sqrt{1 - (M-1)\epsilon}$  for a row  $i$ , then  $\|(\mathbf{X}_*)_i\|_{\ell_2}$  cannot be 0 and, therefore, component  $i$  is included in the recovered support. ■

**Appendix C. Proof of results in §5.** In this section, we will use  $\theta$  for azimuthal angle,  $\phi$  for polar angle and  $\Omega$  for the area of imaging array. We also assume the size of the array  $a$  is much larger than the distance  $h$  between any two neighboring transducers.

*Proof.* [Proof of Proposition 5.1] For spherical arrays of radius  $L$ , given any point  $\mathbf{x}$  on the array and  $\mathbf{y}$  in IW, we have  $|\mathbf{x} - \mathbf{y}| \approx L$ . With the continuum approximation

$$\|\widehat{\mathbf{g}}_0(\mathbf{y})\|_{\ell_2}^2 = \sum_{\mathbf{x}} \left| \frac{\exp(-i\kappa|\mathbf{x} - \mathbf{y}|)}{4\pi|\mathbf{x} - \mathbf{y}|} \right|^2 \approx \frac{1}{16\pi^2 h^2} \int_{\Omega} \frac{d\mathbf{x}}{|\mathbf{x} - \mathbf{y}|^2} = \frac{1}{16\pi^2 h^2 L^2} \times (4\pi L^2) = \frac{1}{4\pi h^2},$$

i.e. the norm of Green's function vector is constant under the spherical array. On the other hand, using continuum approximation, we have for the inner product of any two Green's function vector at  $\mathbf{y}_k$  and  $\mathbf{y}_{k'}$ ,

$$\widehat{\mathbf{g}}_0^*(\mathbf{y}_k)\widehat{\mathbf{g}}_0(\mathbf{y}_{k'}) \approx \frac{1}{16\pi^2 h^2} \int_{\Omega} \frac{\exp(i\kappa(|\mathbf{x} - \mathbf{y}_{k'}| - |\mathbf{x} - \mathbf{y}_k|))}{|\mathbf{x} - \mathbf{y}_{k'}||\mathbf{x} - \mathbf{y}_k|} d\mathbf{x},$$

where the integral is taken on the sphere of radius  $L$ , i.e.  $\Omega = \{\mathbf{x} : |\mathbf{x}| = L\}$ . Let  $\widehat{\mathbf{x}} = \frac{\mathbf{x}}{L}$  so  $|\widehat{\mathbf{x}}| = 1$  on the integral area. Because  $|\mathbf{y}| \ll L$ , we have the approximation

$$|\mathbf{x} - \mathbf{y}| = L|\widehat{\mathbf{x}} - \frac{\mathbf{y}}{L}| = L\sqrt{|\widehat{\mathbf{x}}|^2 + \frac{|\mathbf{y}|^2}{L^2} - 2\widehat{\mathbf{x}} \cdot \frac{\mathbf{y}}{L}} \approx L - \widehat{\mathbf{x}}^* \mathbf{y},$$

and therefore

$$|\mathbf{x} - \mathbf{y}_{k'}| - |\mathbf{x} - \mathbf{y}_k| \approx \widehat{\mathbf{x}}^*(\mathbf{y}_k - \mathbf{y}_{k'}).$$

Using these approximations, and since  $|\widehat{\mathbf{x}} - \mathbf{y}_{k'}/L| \approx |\widehat{\mathbf{x}} - \mathbf{y}_k/L| \approx 1$ , we have

$$\begin{aligned} \widehat{\mathbf{g}}_0^*(\mathbf{y}_k)\widehat{\mathbf{g}}_0(\mathbf{y}_{k'}) &\approx \frac{1}{16\pi^2 h^2} \int_{|\widehat{\mathbf{x}}|=1} \frac{\exp(i\kappa\widehat{\mathbf{x}} \cdot (\mathbf{y}_k - \mathbf{y}_{k'}))}{|\widehat{\mathbf{x}} - (\mathbf{y}_{k'}/L)||\widehat{\mathbf{x}} - (\mathbf{y}_k/L)|} d\widehat{\mathbf{x}} \\ &\approx \frac{1}{16\pi^2 h^2} \int_0^{2\pi} d\theta \int_0^\pi \exp(i\kappa|\mathbf{y}_k - \mathbf{y}_{k'}| \cos \phi) \sin \phi d\phi \\ &= \frac{1}{8\pi h^2} \int_0^\pi \exp(i\kappa|\mathbf{y}_k - \mathbf{y}_{k'}| \cos \phi) \sin \phi d\phi \\ &= \frac{1}{4\pi h^2} \frac{\sin \kappa|\mathbf{y}_k - \mathbf{y}_{k'}|}{\kappa|\mathbf{y}_k - \mathbf{y}_{k'}|} = \frac{1}{4\pi h^2} \text{sinc}(\kappa|\mathbf{y}_k - \mathbf{y}_{k'}|), \end{aligned}$$

where we changed the surface integral to an integral characterized by the angles  $\theta$  and  $\phi$ , with  $\phi$  the angle between  $\mathbf{y}_k - \mathbf{y}_{k'}$  and  $\hat{\mathbf{x}}$ . Using the approximate form of the norm of  $\widehat{\mathbf{g}}_0(\mathbf{y})$ , we have

$$\frac{\widehat{\mathbf{g}}_0^*(\mathbf{y}_k)\widehat{\mathbf{g}}_0(\mathbf{y}_{k'})}{\|\widehat{\mathbf{g}}_0(\mathbf{y}_k)\|_{\ell_2}\|\widehat{\mathbf{g}}_0(\mathbf{y}_{k'})\|_{\ell_2}} \approx \text{sinc}(\kappa|\mathbf{y}_k - \mathbf{y}_{k'}|).$$

■

*Proof.* [Proof of Proposition 5.2] We first calculate the norm of Green's function vector under the planar array as follows

$$\|\widehat{\mathbf{g}}_0(\mathbf{y})\|_{\ell_2}^2 \approx \frac{1}{16\pi^2 h^2} \int_{\Omega} \frac{d\mathbf{x}}{|\mathbf{x} - \mathbf{y}|^2} = \frac{1}{16\pi^2 h^2} \int_0^{2\pi} d\theta \int_0^{\phi_0} \tan \phi d\phi = -\frac{1}{8\pi h^2} \log(\cos \phi_0), \quad (\text{C.1})$$

where  $\phi_0 = \arctan(\frac{a}{2L})$  is the maximal polar angle determined by the size  $a$  of the imaging array and the distance  $L$  from the array to the IW. Using the identity  $\cos(\arctan(x)) = 1/\sqrt{1+x^2}$ , we obtain

$$\|\widehat{\mathbf{g}}_0(\mathbf{y})\|_{\ell_2}^2 \approx \frac{1}{16\pi h^2} \log\left(1 + \frac{a^2}{4L^2}\right). \quad (\text{C.2})$$

Hence, for planar arrays, the norm depends on  $a$  and  $L$  and is independent of the pixel size of the IW.

Based on the proof of Proposition 3.1 in [5], when  $\mathbf{y}_k - \mathbf{y}_{k'} \perp \mathbf{y}_k$ , it can be seen that the inner product

$$|\widehat{\mathbf{g}}_0^*(\mathbf{y}_{k'})\widehat{\mathbf{g}}_0(\mathbf{y}_k)| \sim 1/\sqrt{\kappa|\mathbf{y}_k - \mathbf{y}_{k'}|}.$$

Therefore, we only need to show below when  $\mathbf{y}_k - \mathbf{y}_{k'} \parallel \mathbf{y}_k$ , the inner product decays no worse than  $1/\sqrt{\kappa|\mathbf{y}_k - \mathbf{y}_{k'}|}$ .

According to [5], when  $|\mathbf{y}_k - \mathbf{y}_{k'}| \ll L$  and  $(\mathbf{y}_k - \mathbf{y}_{k'}) \parallel \mathbf{y}_k$ , we have that

$$\widehat{\mathbf{g}}_0^*(\mathbf{y}_k)\widehat{\mathbf{g}}_0(\mathbf{y}_{k'}) \approx \frac{1}{8\pi h^2} \int_{\cos \phi_0}^1 \frac{\exp(-i\kappa\eta z)}{z} dz,$$

where  $\eta = |\mathbf{y} - \mathbf{y}^S|$ . When  $\kappa\eta \rightarrow \infty$ , the integrand oscillates very fast provided that  $1/\kappa\eta \ll \cos \phi_0 \ll 1$ . In this case, integration by parts gives the leading asymptotic behavior as  $\kappa\eta \rightarrow \infty$ . Explicitly,

$$\int_{\cos \phi_0}^1 \frac{\exp(-i\kappa\eta z)}{z} dz = \frac{i}{\kappa\eta} \left( \exp(-i\kappa\eta) - \frac{\exp(-i\kappa\eta \cos \phi_0)}{\cos \phi_0} - \int_{\cos \phi_0}^1 \frac{\exp(-i\kappa\eta z)}{z^2} dz \right).$$

The integral on the right hand side vanishes more rapidly than the boundary terms as  $\kappa\eta \rightarrow \infty$  (to see this, integrate  $\int_{\cos \phi_0}^1 \frac{\exp(-i\kappa\eta z)}{z^2} dz$  by parts and notice that it vanishes like  $1/\kappa\eta$ ). Therefore, neglecting the integral on the right hand side, we obtain

$$\left| \int_{\cos \phi_0}^1 \frac{\exp(-i\kappa\eta z)}{z} dz \right| \sim \frac{1}{\kappa\eta \cos \phi_0} \left| \cos \phi_0 - \exp(-i\kappa\eta(\cos \phi_0 - 1)) \right| \text{ as } \kappa\eta \rightarrow \infty.$$

Thus,

$$\left| \int_{\cos \phi_0}^1 \frac{\exp(-i\kappa\eta z)}{z} dz \right| \sim \frac{1}{\kappa\eta \cos \phi_0} \sqrt{\cos^2 \phi_0 + 1 - 2 \cos \phi_0 \cos(\kappa\eta(\cos \phi_0 - 1))} \text{ as } \kappa\eta \rightarrow \infty.$$

For large arrays  $a \gg L$ , we can approximate  $\cos \phi_0 = 2L/\sqrt{a^2 + 4L^2}$  by 0 and obtain  $|\widehat{\mathbf{g}}_0^*(\mathbf{y}_k)\widehat{\mathbf{g}}_0(\mathbf{y}_{k'})| \approx 1/(\kappa\eta \cos \phi_0)$  which implies that, for large arrays, the normalized inner product decreases like  $1/(\kappa\eta \cos \phi_0 \log(\sec \phi_0))$ , as  $\kappa\eta \rightarrow \infty$ . This function depends very smoothly respect to  $\cos \phi_0$  when  $1/\kappa\eta \ll \cos \phi_0 \ll 1$ , i.e., it is almost independent of  $a/L$ .

Moreover, we find that

$$\frac{1}{\kappa\eta} \left( \frac{2}{\cos \phi_0} - 2 \right) \leq \left| \int_{\cos \phi_0}^1 \frac{\exp(-i\kappa\eta z)}{z} dz \right| \leq \frac{2}{\kappa\eta \cos \phi_0},$$

so we get the following bounds

$$\frac{1}{\kappa\eta \log(\sec \phi_0)} \left( \frac{2}{\cos \phi_0} - 2 \right) \leq \left| \frac{\widehat{\mathbf{g}}_0^*(\mathbf{y}_k)\widehat{\mathbf{g}}_0(\mathbf{y}_{k'})}{\|\widehat{\mathbf{g}}_0(\mathbf{y}_k)\|_{\ell_2} \|\widehat{\mathbf{g}}_0(\mathbf{y}_{k'})\|_{\ell_2}} \right| \leq \frac{2}{\kappa\eta \cos \phi_0 \log(\sec \phi_0)}.$$

Together with the estimate of the cases when  $\mathbf{y}_k - \mathbf{y}_{k'} \perp \mathbf{y}_k$ , we can see the inner product, when planar array is used, has decay rate  $\frac{1}{\sqrt{\kappa\eta}}$ . ■

## REFERENCES

- [1] L. BORCEA, C. TSOGKA, G. PAPANICOLAOU AND J. BERRYMAN, *Imaging and time reversal in random media*, Inverse Problems. 18 (2002), pp. 1247–1279.
- [2] L. BORCEA, G. PAPANICOLAOU AND C. TSOGKA, *Adaptive interferometric imaging in clutter and optimal illumination*, Inverse Problems. 22 (2006), pp. 1405–1436.
- [3] L. BORCEA, G. PAPANICOLAOU AND C. TSOGKA, *Optimal waveform design for array imaging*, Inverse Problems. 23(2007), pp. 1973–2021.
- [4] A. CHAI, M. MOSCOSO AND G. PAPANICOLAOU, *Array imaging using intensity-only measurements*, Inverse Problems. 27 (2011), 015005.
- [5] A. CHAI, M. MOSCOSO AND G. PAPANICOLAOU, *Robust imaging of localized scatterers using the singular value decomposition and  $\ell_1$  optimization*, Inverse Problems. 29 (2013).
- [6] J. CHEN AND X. HUO, *Theoretical results on sparse representations of multiple measurement Vectors*, IEEE Trans. Signal Processing. 54 (2006), pp. 4634–4643.
- [7] S. F. COTTER, B. D. RAO, K. ENGAN AND K. KREUTZ-DELGADO, *Sparse solutions to linear inverse problems with multiple measurement vectors*, IEEE Trans. Signal Process.. 53 (2005), pp. 2477–2488.
- [8] A. DEVANEY, E. MARENGO AND F. GRUBER, *Time-reversal-based imaging and inverse scattering of multiply scattering point targets*, J. Acoust. Soc. Am.. 118 (2005), pp. 3129–3138.
- [9] D. DONOHO, M. ELAD AND V. TEMLYAKOV, *Stable recovery of sparse overcomplete representations in the presence of noise*, IEEE Trans. Information Theory. 52 (2006), pp. 6–18.
- [10] M. DUARTE, S. SARVOTHAM, D. BARON, M. WAKIN AND R. BARANIUK, *Distributed Compressed Sensing of Jointly Sparse Signals*, Signals, Systems and Computers, 2005. Conference Record of the Thirty-Ninth Asilomar Conference on, pp. 1537–1541, doi: 10.1109/ACSSC.2005.1600024
- [11] Y. ELДАР AND H. RAUHUT, *Average Case Analysis of Multichannel Sparse Recovery Using Convex Relaxation*, IEEE Trans. Information Theory. 56 (2010), pp. 505–5019.
- [12] A. FANNJIANG, T. STROHMER AND P. YAN, *Compressed remote sensing of sparse objects*, SIAM J. Imaging Sci. 3 (2010), pp. 595–618.
- [13] L. FOLDY, *The multiple scattering of waves*, Pyhs. Rev.. 67 (1945), pp. 107–119.
- [14] R. GRIBONVAL, H. RAUHUT, K. SCHNASS AND P. VANDERGHEYNST, *Atoms of all channels, unite! Average case analysis of multi-channel sparse recovery using greedy algorithms*, Journal of Fourier analysis and Applications. 14 (2008), pp. 655–687.
- [15] F. GRUBER, E. MARENGO AND A. DEVANEY, *Time-reversal imaging with multiple signal classification considering multiple scattering between the targets*, J. Acoust. Soc. Am.. 115 (2004), pp. 3042–3047.
- [16] M. LAX, *Multiple scattering of waves*, Rev. Modern. Phys.. 23 (1951), pp. 287–310.

- [17] M. LAX, *Multiple scattering of waves II, The effective field in dense systems*, Phys. Rev.. 85 (1952), pp. 261–269.
- [18] D. MALIOUTOV, M. CETIN AND A. WILLSKY, *A sparse signal reconstruction perspective for source localization with sensor arrays*, IEEE Trans. on Signal Processing. 53 (2005), pp. 3010–3022.
- [19] E. MARENGO AND F. GRUBER, *Noniterative analytical formula for inverse scattering of multiply scattering point targets*, J. Acoust. Soc. Am.. 120 (2006), pp. 3782–3788.
- [20] P. A. MARTIN, *Multiple Scattering Interaction of Time-Harmonic Waves with  $N$  Obstacles*, Cambridge University Press, 2006.
- [21] M. MOSCOSO, A. NOVIKOV, G. PAPANICOLAOU AND L. RYZHIK, *A differential equations approach to  $\ell_1$ -minimization with applications to array imaging*, Inverse Problems. 28 (2012).
- [22] A. MARTINEZ AND Z. QIAO, *Iteratively compensating for multiple scattering in SAR imaging*, Proc. SPIE 8746, Algorithms for Synthetic Aperture Radar Imagery XX, 874603, 2013; doi:10.1117/12.2016281.
- [23] G. MONTALDO, M. TANTER, AND M. FINK, *Revisiting iterative time reversal processing: application to detection of multiple targets*, J. Acoust. Soc. Am.. 115 (2004), pp. 776–784.
- [24] C. PRADA, J. THOMAS, AND M. FINK, *The iterative time reversal process: analysis of the convergence*, J. Acoust. Soc. Am.. 97 (1995), pp. 62–71.
- [25] J. TROPP, *Just relax: Convex programming methods for subset selection and sparse approximation*, ICES Report 0404, The University of Texas at Austin, February 2004.
- [26] J. TROPP, *Just relax: Convex programming methods for identifying sparse signals in noise*, IEEE Trans. Information Theory. 52 (2006), pp. 1030–1051.
- [27] J. TROPP, A GILBERT, AND M STRAUSS, *Algorithms for simultaneous sparse approximation. Part I: Greedy pursuit*, Signal Processing. 86 (2006), pp. 572–588.
- [28] J. TROPP, *Algorithms for simultaneous sparse approximation. Part II: Convex relaxation*, Signal Processing. 86 (2006), pp. 589–602.
- [29] D. WIPF AND B. RAO, *An empirical bayesian strategy for solving the simultaneous sparse approximation problem*, IEEE Trans. Signal Process.. 55 (2007), pp. 3704–3716.
- [30] Z. ZHANG AND B. RAO, *Sparse signal recovery with temporally correlated source vectors using sparse bayesian learning*, IEEE J. Sel. Topics Signal Process.. 5 (2011), pp. 912–926.