

Robust imaging of localized scatterers using the singular value decomposition and ℓ_1 minimization

A Chai¹, M Moscoso² and G Papanicolaou³

¹Institute for Computational and Mathematical Engineering, Stanford University, Stanford, CA 94305 USA

²Gregorio Millán Institute, Universidad Carlos III de Madrid, Madrid 28911, Spain

³Department of Mathematics, Stanford University, Stanford, CA 94305 USA

E-mail: anwei@math.stanford.edu, moscoso@math.uc3m.es, papanico@math.stanford.edu

Abstract. We consider narrow band, active array imaging of localized scatterers in a homogeneous medium with and without additive noise. We consider both single and multiple illuminations and study ℓ_1 minimization-based imaging methods. We show that for large arrays, with array diameter comparable to range, and when scatterers are sparse and well separated, ℓ_1 minimization using a single illumination and without additive noise can recover the location and reflectivity of the scatterers exactly. For multiple illuminations we introduce a hybrid method which combines the singular value decomposition and ℓ_1 minimization. This method can be used when the essential singular vectors of the array response matrix are available. We show that with this hybrid method we can recover the location and reflectivity of the scatterers exactly when there is no noise in the data. Numerical simulations indicate that the hybrid method is, in addition, robust to noise in the data. We also compare the ℓ_1 minimization-based methods with others including Kirchhoff migration, ℓ_2 minimization, and multiple signal classification (MUSIC).

Submitted to: *Inverse Problems*

1. Introduction

In this paper, we study active array imaging of small scatterers in homogeneous media where narrow band signals are sent from an array of transducers and the backscattered echoes are recorded. We consider the situation in which we have a small number of scatterers (targets) that are localized, with diameter small compared to the wavelength, and are well separated so that multiple scattering between them is negligible. Under these conditions and for a fixed illumination, the imaging problem amounts to solving a linear inverse problem $A\rho = b$. Here, A is a known $N \times K$ matrix that depends on the illumination, ρ is the solution vector that contains the position and reflectivity of the scatterers, and b is the data vector without additive, measurement noise. The linear system $A\rho = b$ is underdetermined in our case, so $N \ll K$, that is, the number of array transducers N is much smaller than the number of possible scatterer locations within the image region of interest, and therefore there are many configurations of scatterers that match the data vector b .

We cannot expect, in general, to recover ρ exactly from the array measurements forming the vector b . However, if the configuration of the scatterers is simple enough (sparsity is important) and the matrix A satisfies certain conditions (e.g. A has approximately orthogonal columns), then there is a unique sparse solution ρ_0 such that $A\rho_0 = b$. By sparse, we understand a solution vector ρ_0 whose support is not too large compared to its length, that is, the number of nonzero entries M of ρ_0 is much smaller than the total number of entries K .

Similar ideas have been exploited in sparse signal processing [8, 12], especially in compressed sensing (CS) [9], in order to recover a signal from incomplete measurements of it. Motivated by advances in this field, many researchers have applied the recent developments in CS to imaging problems. Both theoretical and empirical results have been obtained [1, 23, 17, 18, 19, 24] (and references therein). In [1], the authors use CS ideas to improve the performance of a wide class of radar systems. These techniques are appropriate for monostatic, bistatic, and multistatic radar configurations, and allow simplifications that make them computationally less demanding. Similar ideas have been applied to ground penetrating radar in [23], where it is shown that only a small number of random space-frequency measurements are enough to construct an image by solving an ℓ_1 minimization problem. In [17, 18, 19], the authors show that the image quality can be significantly improved using *randomized illumination*. In [24], the authors study radar by using compressed sensing technique and a special Gabor frame as the basis of sensing matrix in equations of compressed sensing to achieve higher resolution.

The main idea in these works can be summarized as recovering the sparse solution through a special form of optimization. The exact, most sparse solution is given by an ℓ_0 minimization, where the ℓ_0 norm of a vector is the number of its nonzero entries. However, ℓ_0 minimization requires a combinatorial search that grows exponentially with K . In fact, it has been shown that it is, in general, NP-hard [34] and, therefore, computationally intractable. On the other hand, it has been shown [9] that under

certain conditions on A (the restricted isometry property (RIP)) and sparsity of ρ_0 , ℓ_1 minimization is equivalent to ℓ_0 minimization. Since ℓ_1 minimization is a convex optimization problem that can be solved by standard optimization methods the search of sparse solutions becomes computationally tractable.

In this paper we consider first active array imaging with a single, central illumination and then active array imaging with multiple illuminations. In the case of single illumination from the central transducer in the array, we solve the underdetermined linear system $A\rho = \mathbf{b}$ by ℓ_1 norm minimization, which means that we minimize $\|\rho\|_1$ with $A\rho = \mathbf{b}$ as constraint, when there is no noise. We show in Theorem 3.1 that by solving this optimization problem we can recover the locations and reflectivities of the scatterers exactly provided that certain conditions involving the array configuration and the discretization of the image area (image window) are satisfied. These conditions control indirectly the resolution of the image. The ℓ_1 norm optimization can be solved efficiently by iterative soft-thresholding algorithms. When multiple illumination vectors are used to probe the medium, instead of simply stacking the data to form a problem of larger dimension, we introduce a new hybrid approach combining the use of the singular value decomposition (SVD) of the data with ℓ_1 norm minimization. Since in active array imaging we can control the illumination vectors that can be used to probe the medium, it is best to use the top, or essential, right singular vectors of the array response matrix as illuminations to collect data. We can then project the array data onto the subspace spanned by the top left singular vectors so as to filter out the unnecessary data and noise. The ℓ_1 norm minimization is applied to this reduced linear system to obtain the sparse solution. The top or essential singular vectors can be obtained by applying the SVD to the data array response matrix, if available, or by an iterative time reversal process as recounted in Appendix B.

We also briefly review several existing methods for active array imaging and compare their performance with ℓ_1 minimization with a single illumination and the hybrid ℓ_1 method with multiple illuminations. These methods include Kirchhoff Migration (KM), ℓ_2 norm minimization, and Multiple Signal Classification (MUSIC). KM imaging applies the adjoint of A to the data to get an image, and is often used as an approximate solution to Least Squares imaging (LSQ) for large scale problems to reduce the computational cost. MUSIC is very efficient for locating well-separated scatterers and is robust to additive noise [37, 16, 3, 20]. We also discuss the relation between the hybrid ℓ_1 method for active array imaging used here and methods in passive array imaging with Multiple Measurement Vectors (MMV) that seek solution vectors that have common support. Although passive array imaging is essentially different from active array imaging where the illumination can be controlled, we compare the method used here with recent research in MMV [29, 40, 38, 27, 28].

The organization of the paper is as follows. In section 2, we first introduce the active array imaging problem. In section 3, we discuss the active array imaging using ℓ_1 norm minimization, with single illumination. In section 4, we consider imaging with multiple illuminations and introduce the hybrid- ℓ_1 approach which takes the advantages

of both the SVD and ℓ_1 norm minimization. In section 5, we briefly review KM, ℓ_2 minimization, MUSIC, and discuss the differences between active array imaging and MMV based passive array methods. In section 6, we compare the performance of ℓ_1 minimization method for single illumination with and without subsampling to KM, ℓ_2 minimization, and MUSIC assuming noise free as well as data with additive noise. In section 7, we show the results of numerical experiments comparing the performance of the hybrid ℓ_1 with other imaging methods using multiple illuminations. In the end, we conclude with a brief discussion in section 8.

2. Active array imaging

In this section, we introduce the direct and inverse scattering problems for imaging point-like scatterers with an active array in a homogeneous medium. A typical active array imaging setup is shown in Figure 1. The array is active because it emits signals from transducers at $\mathbf{x}_s \in \mathcal{A}$ and records the echoes with receivers at $\mathbf{x}_r \in \mathcal{A}$. In active array imaging we seek to locate the positions and, if possible, the reflectivities of the scatterers using the data recorded on the array.

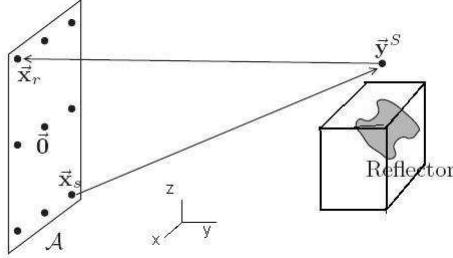


Figure 1. General setup of array imaging

Let the active array with N transducers at positions \mathbf{x}_p , $p = 1, \dots, N$, be located on the plane $z = 0$. We will use the notation \mathbf{x}_s and \mathbf{x}_r to denote the locations of the transducers that emit or receive the signals, respectively. The distance between transducers is of the order of the wavelength $\lambda = 2\pi c/\omega$ (c is the wave speed in the medium), and we denote its physical diameter by a , which is proportional to $N\lambda$. Assume that there are M scatterers located at positions $\mathbf{y}_{n_1}, \dots, \mathbf{y}_{n_M}$. Under the Born approximation, the response at \mathbf{x}_r due to a narrow band pulse of angular frequency ω sent from \mathbf{x}_s and reflected by the M scatterers is given by

$$\widehat{\Pi}(\mathbf{x}_r, \mathbf{x}_s, \omega) = \sum_{j=1}^M \alpha_j \widehat{G}_0(\mathbf{x}_r, \mathbf{y}_{n_j}, \omega) \widehat{G}_0(\mathbf{y}_{n_j}, \mathbf{x}_s, \omega), \quad (1)$$

where α_j , $j = 1, 2, \dots, M$, are the unknown reflectivities of scatterers, which we assume to be real. In (1),

$$\widehat{G}_0(\mathbf{y}, \mathbf{x}, \omega) = \frac{\exp(i\kappa|\mathbf{x} - \mathbf{y}|)}{4\pi|\mathbf{x} - \mathbf{y}|}, \quad (2)$$

with $\kappa = \omega/c$, is the free space Green's function that characterizes wave propagation from \mathbf{x} to \mathbf{y} in a homogeneous medium. The expression (1) for the array response matrix is widely used and derived from first principles in [15, 30] under the hypotheses we have introduced.

The inverse scattering problem for active array imaging is as follows. The M scatterers will be assumed to be located in a known bounded region, which is called the Image Window (IW). The IW is taken to be the parallelepiped $\mathcal{K} := [-b_x, b_x] \times [-b_y, b_y] \times [L, L + b_z]$, where L is the distance between the array and the center of the IW. 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 $\{\mathbf{y}_{n_1}, \dots, \mathbf{y}_{n_M}\} \subset \{\mathbf{y}_1, \dots, \mathbf{y}_K\}$. We will assume here that the number of scatterers M is much smaller than the number of grid points K , so $M \ll K$. Furthermore, since the number N of transducers on the array is typically much smaller than K , the positions and reflectivities of the scatterers will be given as the solution of an underdetermined linear inverse problem. We now write explicitly the form of the linear problem that relates the reflectivity at each grid point \mathbf{y}_j of the IW with the data measured at the array, for each illumination.

Using the Kronecker delta notation, we introduce the *reflectivity vector* $\boldsymbol{\rho}_0 = (\rho_{01}, \rho_{02}, \dots, \rho_{0K}) \in \mathbb{R}^K$, such that

$$\rho_{0k} = \sum_{j=1}^M \alpha_j \delta_{\mathbf{y}_k, \mathbf{y}_{n_j}}, \quad k = 1, 2, \dots, K,$$

and the *Green's function vector* $\widehat{\mathbf{g}}_0(\mathbf{y}_k, \omega)$ for each point \mathbf{y}_k in IW as

$$\widehat{\mathbf{g}}_0(\mathbf{y}_k, \omega) = (\widehat{G}_0(\mathbf{x}_1, \mathbf{y}_k, \omega), \widehat{G}_0(\mathbf{x}_2, \mathbf{y}_k, \omega), \dots, \widehat{G}_0(\mathbf{x}_N, \mathbf{y}_k, \omega))^T, \quad k = 1, 2, \dots, K,$$

where the superscript T means transpose. The vector $\widehat{\mathbf{g}}_0(\mathbf{y}_k, \omega)$ is the signal at the array due to a point source at \mathbf{y}_k . By spatial reciprocity, $\widehat{G}_0(\mathbf{x}_i, \mathbf{y}_k, \omega) = \widehat{G}_0(\mathbf{y}_k, \mathbf{x}_i, \omega)$, this vector can also be interpreted as the illumination vector of the array targeting the position \mathbf{y}_k . With $\boldsymbol{\rho}_0$ and $\widehat{\mathbf{g}}_0(\mathbf{y}_k, \omega)$ we can rewrite the model response matrix as a sum of outer products

$$\begin{aligned} \widehat{\Pi}(\omega) &\equiv [\widehat{\Pi}(\mathbf{x}_r, \mathbf{x}_s, \omega)]_{r,s=1}^N = \sum_{j=1}^M \alpha_j \widehat{\mathbf{g}}_0(\mathbf{y}_{n_j}, \omega) \widehat{\mathbf{g}}_0^T(\mathbf{y}_{n_j}, \omega) \\ &= \sum_{j=1}^K \rho_{0j} \widehat{\mathbf{g}}_0(\mathbf{y}_j, \omega) \widehat{\mathbf{g}}_0^T(\mathbf{y}_j, \omega). \end{aligned} \quad (3)$$

Note that since $M \ll K$, the vector $\boldsymbol{\rho}_0$ is sparse.

Now let $\widehat{\mathbf{f}}(\omega) = (\widehat{f}_1(\omega), \dots, \widehat{f}_N(\omega))^T$ be an illumination vector satisfying $\|\widehat{\mathbf{f}}(\omega)\|_{\ell_2} = 1$. Then, the data $\mathbf{b}(\omega)$ at the array is given by

$$\mathbf{b}(\omega) = \widehat{\Pi}(\omega) \widehat{\mathbf{f}}(\omega) = \begin{bmatrix} \sum_{s=1}^N \widehat{\Pi}(\mathbf{x}_1, \mathbf{x}_s, \omega) \widehat{f}_s(\omega) \\ \sum_{s=1}^N \widehat{\Pi}(\mathbf{x}_2, \mathbf{x}_s, \omega) \widehat{f}_s(\omega) \\ \vdots \\ \sum_{s=1}^N \widehat{\Pi}(\mathbf{x}_N, \mathbf{x}_s, \omega) \widehat{f}_s(\omega) \end{bmatrix}. \quad (4)$$

Through $\widehat{\Pi}(\omega)\widehat{\mathbf{f}}(\omega)$, we define the operator $\mathcal{A}_{\widehat{f}(\omega)}$ by

$$\mathcal{A}_{\widehat{f}(\omega)}\boldsymbol{\rho}_0 = \widehat{\Pi}(\omega)\widehat{\mathbf{f}}(\omega), \quad (5)$$

which connects the reflectivity vector $\boldsymbol{\rho}_0$ and the data (4). It is easy to see that $\mathcal{A}_{\widehat{f}(\omega)}$ is an $N \times K$ matrix whose j^{th} column is given by $\widehat{g}_{\widehat{f}}(\mathbf{y}_j, \omega)\widehat{\mathbf{g}}_0(\mathbf{y}_j, \omega)$ where

$$\widehat{g}_{\widehat{f}}(\mathbf{y}_j, \omega) = \widehat{\mathbf{g}}_0^T(\mathbf{y}_j, \omega)\widehat{\mathbf{f}}(\omega), \quad (6)$$

that is,

$$\mathcal{A}_{\widehat{f}(\omega)} = \begin{bmatrix} \widehat{g}_{\widehat{f}}(\mathbf{y}_1, \omega)\widehat{\mathbf{g}}_0(\mathbf{y}_1, \omega) & \widehat{g}_{\widehat{f}}(\mathbf{y}_2, \omega)\widehat{\mathbf{g}}_0(\mathbf{y}_2, \omega) & \cdots & \widehat{g}_{\widehat{f}}(\mathbf{y}_K, \omega)\widehat{\mathbf{g}}_0(\mathbf{y}_K, \omega) \end{bmatrix}. \quad (7)$$

Note that $\widehat{g}_{\widehat{f}}(\mathbf{y}_j, \omega)$ in (6) is the illumination seen at a point located at \mathbf{y}_j due to an illumination vector $\widehat{\mathbf{f}}(\omega)$ at the array. With this notation, active array imaging with a single illumination amounts to solving

$$\mathcal{A}_{\widehat{f}(\omega)}\boldsymbol{\rho} = \mathbf{b}(\omega) \quad (8)$$

for the unknown reflectivity vector $\boldsymbol{\rho}$. Since $N \ll K$, this is an underdetermined linear system.

When data from multiple illuminations $\widehat{\mathbf{f}}^{(j)}(\omega)$, $j = 1, \dots, \nu$, are recorded at the array, we may stack the data vectors $\mathbf{b}^{(j)}(\omega)$ from each illumination j in an augmented νN data vector

$$\mathbf{b}_\nu(\omega) = ((\mathbf{b}^{(1)}(\omega))^T, \dots, (\mathbf{b}^{(\nu)}(\omega))^T)^T.$$

Similarly, we define the augmented $\nu N \times K$ matrix

$$\mathcal{A}_\nu = (\mathcal{A}_{\widehat{f}^{(1)}(\omega)}^T, \dots, \mathcal{A}_{\widehat{f}^{(\nu)}(\omega)}^T)^T.$$

Then, active array imaging with multiple illuminations can be formulated as solving the larger set of equations

$$\mathcal{A}_\nu \boldsymbol{\rho} = \mathbf{b}_\nu(\omega), \quad (9)$$

for the reflectivity vector $\boldsymbol{\rho}$. This system is still underdetermined if $\nu N < K$.

Finally, we give a result that will be used throughout the paper which is a consequence of the destructive interference and spatial decay of the Green's functions for points that are far apart. For small arrays, $a \ll L$, this is well known [3, 4]. For more general arrays the proof is given in Appendix A.

Proposition 2.1. *For a planar array of finite size a , as in Figure 1, let \mathbf{y} and \mathbf{y}^S be two points within the IW. When $\lambda \ll |\mathbf{y} - \mathbf{y}^S| \ll L$, we have*

$$\widehat{\mathbf{g}}_0^*(\mathbf{y}, \omega)\widehat{\mathbf{g}}_0(\mathbf{y}^S, \omega) \rightarrow 0, \quad \text{as } \frac{|\mathbf{y} - \mathbf{y}^S|}{\lambda} \rightarrow \infty, \quad (10)$$

where $*$ is the conjugate transpose operator.

The above result implies that the illumination vectors $\widehat{\mathbf{g}}_0(\mathbf{y}_i, \omega)$ at each grid point \mathbf{y}_i of the IW are approximately orthogonal, that is,

$$\widehat{\mathbf{g}}_0^*(\mathbf{y}_i, \omega) \widehat{\mathbf{g}}_0(\mathbf{y}_j, \omega) = \begin{cases} \|\widehat{\mathbf{g}}_0(\mathbf{y}_i, \omega)\|_{\ell_2}^2, & \text{if } i = j, \\ \epsilon \ll 1, & \text{otherwise,} \end{cases} \quad (11)$$

if the grid points are sufficiently apart relative to the wavelength, for a large array. For small arrays, [3, 4], approximate orthogonality will hold under the condition

$$|\mathbf{y}_i - \mathbf{y}_j| \gg \max\left\{\frac{\lambda L}{a}, \lambda \left(\frac{L}{a}\right)^2\right\}. \quad (12)$$

3. Single illumination imaging with ℓ_1 norm minimization

In many inverse problems that are underdetermined we often seek the one that is sparsest. For a single illumination given by the vector $\widehat{\mathbf{f}}(\omega)$, the sparsest solution of the inverse scattering problem is given by

$$\min \|\boldsymbol{\rho}\|_{\ell_0} \quad \text{subject to} \quad \mathcal{A}_{\widehat{\mathbf{f}}(\omega)} \boldsymbol{\rho} = \mathbf{b}(\omega). \quad (13)$$

In this form of the imaging problem one may be able to exploit the sparsity of the solution $\boldsymbol{\rho}_0$. However, (13) is an NP-hard problem and therefore it has exponential complexity.

Basis Pursuit (BP) and CS [8, 9, 10, 12] methods provide an alternative convex optimization approach to this NP-hard problem under certain conditions. More precisely, the ℓ_0 norm can be replaced by the ℓ_1 norm when $\mathcal{A}_{\widehat{\mathbf{f}}(\omega)}$ has approximately orthogonal columns, in the form described in the previous section. Motivated by this, we use ℓ_1 norm minimization to capture the sparsity of $\boldsymbol{\rho}_0$, and solve the convex relaxation of (13)

$$\min \|\boldsymbol{\rho}\|_{\ell_1} \quad \text{subject to} \quad \mathcal{A}_{\widehat{\mathbf{f}}(\omega)} \boldsymbol{\rho} = \mathbf{b}(\omega). \quad (14)$$

A related problem to (14) has also been studied in [11] where array imaging of localized scatterers is done based on intensity-only measurements. It is shown that exact recovery can be achieved by minimizing the rank of a positive semidefinite matrix associated with the unknown reflectivities of the scatterers. Since this optimization problem is NP-hard, the rank of the matrix is replaced by its nuclear norm, which is a convex programming problem that can be solved in polynomial time.

3.1. Exact recovery: single illumination

To analyze (14), we write the linear constraint with a normalized matrix by redefining the unknown vector $\boldsymbol{\rho}$. The matrix $\mathcal{A}_{\widehat{\mathbf{f}}(\omega)}$ is given by (7), which has dimensions $N \times K$. We now define the vector

$$\tilde{\boldsymbol{\rho}} = (\widehat{g}_f(\mathbf{y}_1, \omega) \|\widehat{\mathbf{g}}_0(\mathbf{y}_1, \omega)\|_{\ell_2} \rho_1, \dots, \widehat{g}_f(\mathbf{y}_K, \omega) \|\widehat{\mathbf{g}}_0(\mathbf{y}_K, \omega)\|_{\ell_2} \rho_K)^T, \quad (15)$$

where $\hat{g}_f(\mathbf{y}_k, \omega)$ is defined by (6), and consider the problem

$$\min \|\tilde{\boldsymbol{\rho}}\|_{\ell_1} \quad \text{subject to} \quad \tilde{A}\tilde{\boldsymbol{\rho}} = \mathbf{b}(\omega), \quad (16)$$

where \tilde{A} is an $N \times K$ matrix whose j^{th} column is given by the normalized Green's function vector $\hat{\mathbf{g}}_0(\mathbf{y}_j, \omega)/\|\hat{\mathbf{g}}_0(\mathbf{y}_j, \omega)\|_{\ell_2}$. Now the matrix \tilde{A} does not depend on the illumination vector $\tilde{\mathbf{f}}(\omega)$ but the solution $\tilde{\boldsymbol{\rho}}$ does. After solving (16) for $\tilde{\boldsymbol{\rho}}$, we can recover $\boldsymbol{\rho} = (\rho_1, \rho_2, \dots, \rho_K)$ from (15) since the illumination vector is known. We will assume in the rest of this section that the Green's function vectors $\hat{\mathbf{g}}_0(\mathbf{y}_j, \omega)$ have unit norm.

It follows from Proposition 2.1 that $\tilde{A}^*\tilde{A}$ is a diagonally dominant matrix, approximately equal to the identity I , provided that the grid points in the IW are well separated. In this case, $\tilde{A}^*\tilde{A}$ is an identity matrix perturbed by a matrix \mathcal{E} with $\|\mathcal{E}\|_1 \ll 1$, where $\|\cdot\|_1$ is the maximum ℓ_1 norm of all columns of a given matrix. Hence, $\tilde{A}^*\tilde{A} = I + \mathcal{E}$ with $\|\mathcal{E}\|_1 \ll 1$. The bound of $\|\mathcal{E}\|_1$ depends on the discretization of the IW, the array size a and the range L , which control the resolution of the image. For example, the larger the grid spacing in the IW is, the smaller $\|\mathcal{E}\|_1$ is.

The following result tells us that the ℓ_1 norm minimization method can recover the reflectivity $\boldsymbol{\rho}_0$ exactly from noiseless data in the case of single illumination.

Theorem 3.1. *Assume that the resolution of the IW is such that*

$$\max_{i \neq j} |\hat{\mathbf{g}}_0^*(\mathbf{y}_i, \omega) \hat{\mathbf{g}}_0(\mathbf{y}_j, \omega)| < \epsilon.$$

If the number of scatterers M is such that $\epsilon M < 1/2$, then $\boldsymbol{\rho}_0$ is the unique solution to (16) with support M .

Proof: Let $T = \{n_j, 1 \leq j \leq M\}$ be the set of indices corresponding to the scatterers. Based on the resolution condition, for any $1 \leq i, j \leq M$, we have $\hat{\mathbf{g}}_0^*(\mathbf{y}_i, \omega) \hat{\mathbf{g}}_0(\mathbf{y}_j, \omega) = \delta_{ij} + (1 - \delta_{ij})\epsilon_{ij}$ with $|\epsilon_{ij}| < \epsilon$. Therefore, the submatrix \tilde{A}_T , which is formed by columns n_1, \dots, n_M of the matrix \tilde{A} , is such that $\tilde{A}_T^*\tilde{A}_T$ is of full rank and diagonally dominant. Following a similar argument as in Theorem 1.3 in [10], we only need to find a Lagrange multiplier vector $\boldsymbol{\mu}$ such that for any column \mathbf{a}_j of \tilde{A}_T

$$\boldsymbol{\mu}^* \mathbf{a}_j = \text{sign}(\rho_{0j}) = \text{sign}(\alpha_j), \quad (17)$$

and for any column \mathbf{a}_j with $j \notin T$

$$|\boldsymbol{\mu}^* \mathbf{a}_j| < 1. \quad (18)$$

Let $\boldsymbol{\mu} = \tilde{A}_T(\tilde{A}_T^*\tilde{A}_T)^{-1}\mathbf{1}$ where $\mathbf{1}$ is the vector whose entries are all 1. Then (17) is automatically satisfied because $\tilde{A}_T^*\boldsymbol{\mu} = \mathbf{1}$. For (18), choosing any column of \tilde{A} not in the submatrix \tilde{A}_T , we have

$$|\mathbf{a}_j^* \tilde{A}_T(\tilde{A}_T^*\tilde{A}_T)^{-1}\mathbf{1}| \leq \|(\tilde{A}_T^*\tilde{A}_T)^{-1}\tilde{A}_T^*\mathbf{a}_j\|_1 \leq \|(\tilde{A}_T^*\tilde{A}_T)^{-1}\|_1 \|\tilde{A}_T^*\mathbf{a}_j\|_1 \leq \frac{M\epsilon}{1 - M\epsilon} < 1,$$

where the last inequality is due to the condition $\epsilon M < 1/2$.

Remark 3.2. The resolution condition in Theorem 3.1 is related to the mutual coherent condition when solving ℓ_1 minimization problems. In our notation, the mutual coherence of \tilde{A} in (16) is

$$\mu(\tilde{A}) = \max_{i \neq j} |\hat{\mathbf{g}}_0^*(\mathbf{y}_i, \omega) \hat{\mathbf{g}}_0(\mathbf{y}_j, \omega)|.$$

Theorem 7 in [8] shows that (16) has a unique solution when \tilde{A} has full row rank and the following holds:

$$\mu(\tilde{A})\|\boldsymbol{\rho}\|_0 < \frac{1}{2}(1 + \mu(\tilde{A})).$$

Under the conditions of Theorem 3.1, this implies $\epsilon M < (1 + \epsilon)/2$. However, in imaging we do not necessarily have an \tilde{A} with full row rank.

3.2. An algorithm for ℓ_1 norm minimization.

When the data is noisy, instead of solving the constrained problem (14) directly, we solve the relaxed problem

$$\boldsymbol{\rho}_{\ell_1(\tau)} = \underset{\boldsymbol{\rho}}{\operatorname{argmin}} \{F_1(\boldsymbol{\rho}) \equiv \tau \|\boldsymbol{\rho}\|_{\ell_1} + f(\boldsymbol{\rho})\}, \quad (19)$$

where

$$f(\boldsymbol{\rho}) = \frac{1}{2} \|\mathcal{A}_{\widehat{f}(\omega)} \boldsymbol{\rho} - \mathbf{b}(\omega)\|_{\ell_2}^2. \quad (20)$$

Problems (14) and (19) can be rewritten as linear programming (LP) problems for which many efficient algorithms and packages exist [14, 39, 35]. However, standard LP algorithms are too expensive for large scale problems with dense matrices. A large amount of research has been devoted to finding new algorithms for solving such problems (see, for example,[2, 41, 33]). In this paper, we use a *shrikage-thresholding algorithm*, which is a simple extension of the classical gradient method and involves only matrix-vector multiplications followed by a shrinkage-thresholding step. The classical gradient method with shrinkage-thresholding generates a sequence of iterates $\{\boldsymbol{\rho}^{(k)}\}$ of the form

$$\boldsymbol{\rho}^{(k+1)} = \eta_{\tau\alpha_k}(\boldsymbol{\rho}^{(k)} - \alpha_k \nabla f(\boldsymbol{\rho}^{(k)})). \quad (21)$$

In (21), $\alpha_k > 0$ is the stepsize, $[\eta_a(\boldsymbol{\rho})]_i = \max(|\rho_i| - a, 0) \operatorname{sign}(\rho_i)$ is the shrinkage-thresholding operator, and $\nabla f(\boldsymbol{\rho}) = \mathcal{A}_{\widehat{f}(\omega)}^* (\mathcal{A}_{\widehat{f}(\omega)} \boldsymbol{\rho} - \mathbf{b}(\omega))$ is the gradient of (20). The convergence and performance of (21) has been studied extensively in the literature (see, e.g., [21, 7, 2] and references therein). A typical condition to ensure that the sequence $\{\boldsymbol{\rho}^{(k)}\}$ converges to the solution $\boldsymbol{\rho}_0$ is to require that $\alpha_k < 2/L$, where $L = \|\mathcal{A}_{\widehat{f}(\omega)}^* \mathcal{A}_{\widehat{f}(\omega)}\|_{\ell_2}$ is the Lipschitz constant of ∇f . The convergence rate is $O(1/k)$, so $F_1(\boldsymbol{\rho}^{(k)}) - F_1(\boldsymbol{\rho}_0) \sim O(1/k)$.

In order to accelerate the convergence, we use a *fast shrikage-thresholding algorithm* that improves the rate of convergence to $O(1/k^2)$ (see [2] for details). This algorithm speeds up the performance by computing the iterate at step $k + 1$ using not only the

previous one as in (21), but using the two previously computed ones, that is, the iterate at step $k+1$ depends on the iterates at steps k and $k-1$. The algorithm is the following:

$$\boldsymbol{\rho}^{(k)} = \eta_{\tau\alpha_k}(\boldsymbol{\xi}^{(k)} - \alpha_k \nabla f(\boldsymbol{\xi}^{(k)})), \quad (22)$$

$$\alpha_{k+1} = \frac{1 + \sqrt{1 + 4\alpha_k^2}}{2}, \quad (23)$$

$$\boldsymbol{\xi}^{(k+1)} = \boldsymbol{\rho}^{(k)} + \frac{\alpha_k - 1}{\alpha_{k+1}}(\boldsymbol{\rho}^{(k)} - \boldsymbol{\rho}^{(k-1)}), \quad (24)$$

for given $\boldsymbol{\rho}_1$, $\boldsymbol{\xi}_2 = \boldsymbol{\rho}_1$ and $\alpha_1 < 2/L$. The algorithm ends on convergence or when a maximum number of iterations is reached. This algorithm keeps the ease of implementation of (21), and its main computational effort, while improving the convergence rate.

4. The hybrid ℓ_1 method for imaging with multiple illuminations

In this section, we introduce a method for active array imaging problem with multiple illuminations. We will assume at first that the array response matrix $\widehat{P}(\omega) = [\widehat{P}(\mathbf{x}_r, \mathbf{x}_s, \omega)]_{r,s=1}^N$ is recorded and known. This means that we record at the array the signals received when each transducer emits a spherical wave at angular frequency ω . It is a complex, symmetric $N \times N$ matrix. We distinguish between the measured array response matrix $\widehat{P}(\omega)$, the data, and the model response matrix $\widehat{\Pi}(\omega)$ defined in (3) in terms of Green's functions and in the Born approximation. This distinction is necessary because multiple scattering between the scatterers may not be negligible, there may be inhomogeneities in the ambient medium, and we may not have ideal point scatterers. The model we use for inversion is still (8).

We write the singular value decomposition (SVD) of the data matrix $\widehat{P}(\omega)$ in the form

$$\widehat{P}(\omega) = \widehat{U}(\omega)\Sigma(\omega)\widehat{V}^*(\omega) = \sum_{j=1}^K \sigma_j(\omega)\widehat{U}_j(\omega)\widehat{V}_j^*(\omega). \quad (25)$$

It is a consequence of Proposition 2.1 (see also [3]) that if the scatterers are sufficiently far apart from each other, we can associate with each of the M scatterers \mathbf{y}_{n_j} a nonzero singular value $\sigma_j(\omega)$, and the corresponding singular vectors have the explicit form

$$\widehat{U}_j(\omega) = \overline{\widehat{V}_j(\omega)} \approx \frac{\widehat{\mathbf{g}}_0(\mathbf{y}_{n_j}, \omega)}{\|\widehat{\mathbf{g}}_0(\mathbf{y}_{n_j}, \omega)\|_{\ell_2}}, \quad \sigma_j(\omega) \approx \alpha_j \|\widehat{\mathbf{g}}_0(\mathbf{y}_{n_j}, \omega)\|_{\ell_2}^2, \quad j = 1, \dots, M. \quad (26)$$

The remaining singular vectors $\widehat{U}_j(\omega)$, $j = M+1, \dots, N$, span the *noise subspace* which is the orthogonal complement to the range of $\widehat{P}(\omega)\widehat{P}^*(\omega)$. The singular values corresponding to the noise subspace are essentially zero. The approximate equalities in (26) are not true when multiple scattering is not negligible or when the scatterers are not very far apart. However, the subspace spanned by the vectors $\widehat{\mathbf{g}}_0(\mathbf{y}_{n_j}, \omega)$, $n_j = 1, \dots, M$, is the same as the signal subspace $\text{span}\{\widehat{U}_1(\omega), \widehat{U}_2(\omega), \dots, \widehat{U}_M(\omega)\}$ of

the array response matrix [3]. Imaging using the SVD when the scatterers are not well separated is considered in [5].

To solve the inverse problem (8) with multiple illuminations using ℓ_1 minimization, we have to solve (14) with several constraints of the form $\mathcal{A}_{\widehat{f}^{(j)}(\omega)}\boldsymbol{\rho} = \mathbf{b}^{(j)}(\omega)$ for $j = 1, 2, \dots, \nu$ where ν is the number of illuminations. This is a special form of (9). We introduce an approach, the hybrid ℓ_1 method, that uses efficiently the SVD of the data, instead of using all possible (or all available) illuminations and the array response matrix (or the part of it available). The essential point here is that only the singular vectors corresponding to significant (non-zero) singular values of $\widehat{P}(\omega)$ should be used. The constraints in (14) take now the form

$$\widehat{U}_j^*(\omega)\mathcal{A}_{\widehat{V}_j(\omega)}\boldsymbol{\rho} = \widehat{U}_j^*(\omega)\widehat{P}(\omega)\widehat{V}_j(\omega) = \sigma_j(\omega), \quad j = 1, \dots, M, \quad (27)$$

where we use the right singular vectors $\widehat{V}_j(\omega)$ as illumination vectors, that is, $\widehat{\mathbf{f}}^{(j)}(\omega) = \widehat{V}_j(\omega)$, and project with the left singular vectors $\widehat{U}_j(\omega)$. There are two implications coming from (27). First, $\widehat{V}_j(\omega)$, $j = 1, \dots, M$, are the illuminations that focus at each scatterer, whose location is not yet known, and give rise to the strongest echoes in the array. Second, by projecting the received data onto the space spanned by $\widehat{U}_j(\omega)$, $j = 1, \dots, M$, the redundant data is filtered out, which plays the role of dimension reduction of the linear system. As we will see later, it is also important for denoising when data is contaminated by additive noise.

To analyze (27), we write it in matrix form. Denote by $\mathcal{B}(\omega)$ the following $M \times K$ matrix

$$\mathcal{B}(\omega) = \begin{bmatrix} (\widehat{\mathbf{g}}_0^T(\mathbf{y}_1, \omega)\widehat{V}_1(\omega))\overline{\widehat{\mathbf{g}}_0^*(\mathbf{y}_1, \omega)\widehat{U}_1(\omega)} & \cdots & (\widehat{\mathbf{g}}_0^T(\mathbf{y}_K, \omega)\widehat{V}_1(\omega))\overline{\widehat{\mathbf{g}}_0^*(\mathbf{y}_K, \omega)\widehat{U}_1(\omega)} \\ \vdots & \ddots & \vdots \\ (\widehat{\mathbf{g}}_0^T(\mathbf{y}_1, \omega)\widehat{V}_M(\omega))\overline{\widehat{\mathbf{g}}_0^*(\mathbf{y}_1, \omega)\widehat{U}_M(\omega)} & \cdots & (\widehat{\mathbf{g}}_0^T(\mathbf{y}_K, \omega)\widehat{V}_M(\omega))\overline{\widehat{\mathbf{g}}_0^*(\mathbf{y}_K, \omega)\widehat{U}_M(\omega)} \end{bmatrix}, \quad (28)$$

and by $\mathbf{b}_\sigma(\omega)$ the data vector $\mathbf{b}_\sigma(\omega) = (\sigma_1(\omega), \dots, \sigma_M(\omega))^T$. Then (27) can be expressed as

$$\mathcal{B}(\omega)\boldsymbol{\rho} = \mathbf{b}_\sigma(\omega). \quad (29)$$

Compared to both (8) and (9), this new linear system uses a smaller subset of the data, but with all the necessary information for imaging the scatterers included. The matrix $\mathcal{B}(\omega)$ is the reduced operator that connects the true reflectivity vector $\boldsymbol{\rho}_0$ and the essential data $\mathbf{b}_\sigma(\omega)$. With this notation, the hybrid ℓ_1 method is to recover $\boldsymbol{\rho}_0$ from the problem

$$\min \|\boldsymbol{\rho}\|_{\ell_1} \quad \text{subject to} \quad \mathcal{B}(\omega)\boldsymbol{\rho} = \mathbf{b}_\sigma(\omega). \quad (30)$$

To summarize, the hybrid ℓ_1 method takes the following three steps:

- (i) Obtain the singular vectors of the recorded array response matrix $\widehat{P}(\omega)$ corresponding to the significant singular values $\sigma_j(\omega)$, $j = 1, \dots, M$, either by SVD of the response matrix or by iterative time reversal, which is discussed in Appendix B.

- (ii) Construct the matrix $\mathcal{B}(\omega)$ given in (28) with the Green's function vectors $\widehat{\mathbf{g}}_0(\mathbf{y}_j, \omega)$, $j = 1, \dots, K$, and the M singular vectors obtained in step (i), and form the data vector $\mathbf{b}_\sigma(\omega) = (\sigma_1(\omega), \dots, \sigma_M(\omega))^T$.
- (iii) Solve the minimization problem (30).

4.1. Exact recovery: multiple illumination

To study the properties of (30), we first normalize the matrix $\mathcal{B}(\omega)$. Let $\mathcal{D}(\omega) = \text{diag}(\|\widehat{\mathbf{g}}_0(\mathbf{y}_1, \omega)\|_{\ell_2}^{-2}, \dots, \|\widehat{\mathbf{g}}_0(\mathbf{y}_K, \omega)\|_{\ell_2}^{-2})$. Then, the (r, s) entry of $\mathcal{B}(\omega)\mathcal{D}(\omega)$ becomes

$$\left(\frac{\widehat{\mathbf{g}}_0^T(\mathbf{y}_s, \omega)}{\|\widehat{\mathbf{g}}_0(\mathbf{y}_s, \omega)\|_{\ell_2}} \widehat{V}_r(\omega) \right) \overline{\left(\frac{\widehat{\mathbf{g}}_0^*(\mathbf{y}_s, \omega)}{\|\widehat{\mathbf{g}}_0(\mathbf{y}_s, \omega)\|_{\ell_2}} \widehat{U}_r(\omega) \right)}$$

and (29) can be rewritten as $\mathcal{B}(\omega)\mathcal{D}(\omega)\mathcal{D}^{-1}(\omega)\boldsymbol{\rho} = \mathbf{b}_\sigma(\omega)$. The reflectivity can be obtained by solving (30) first with constraint $\mathcal{B}(\omega)\mathcal{D}(\omega)\tilde{\boldsymbol{\rho}} = \mathbf{b}_\sigma(\omega)$, and then computing $\boldsymbol{\rho} = \mathcal{D}(\omega)\tilde{\boldsymbol{\rho}}$. Assuming this normalization has been made, we will take hereafter the vectors $\widehat{\mathbf{g}}_0(\mathbf{y}_j, \omega)$, $j = 1, \dots, K$, to have unit norms. Now, using (26), we have

$$\widehat{\mathbf{g}}_0^T(\mathbf{y}_s, \omega) \widehat{V}_r(\omega) = \widehat{U}_r^*(\omega) \widehat{\mathbf{g}}_0(\mathbf{y}_s, \omega),$$

and the $M \times K$ matrix $\mathcal{B}(\omega)$ can be rewritten as

$$\mathcal{B}(\omega) = \begin{bmatrix} (\widehat{U}_1^*(\omega) \widehat{\mathbf{g}}_0(\mathbf{y}_1, \omega))^2 & \cdots & (\widehat{U}_1^*(\omega) \widehat{\mathbf{g}}_0(\mathbf{y}_K, \omega))^2 \\ \vdots & \ddots & \vdots \\ (\widehat{U}_M^*(\omega) \widehat{\mathbf{g}}_0(\mathbf{y}_1, \omega))^2 & \cdots & (\widehat{U}_M^*(\omega) \widehat{\mathbf{g}}_0(\mathbf{y}_K, \omega))^2 \end{bmatrix}.$$

It is easy to see that the following $M \times M$ submatrix of $\mathcal{B}(\omega)$

$$\mathcal{B}_{M \times M}(\omega) = \begin{bmatrix} (\widehat{U}_1^*(\omega) \widehat{\mathbf{g}}_0(\mathbf{y}_{n_1}, \omega))^2 & \cdots & (\widehat{U}_1^*(\omega) \widehat{\mathbf{g}}_0(\mathbf{y}_{n_M}, \omega))^2 \\ \vdots & \ddots & \vdots \\ (\widehat{U}_M^*(\omega) \widehat{\mathbf{g}}_0(\mathbf{y}_{n_1}, \omega))^2 & \cdots & (\widehat{U}_M^*(\omega) \widehat{\mathbf{g}}_0(\mathbf{y}_{n_M}, \omega))^2 \end{bmatrix}$$

is diagonally dominant, and asymptotically equal to the identity matrix I , using (26), provided that the scatterers are well separated, so that $\lambda \ll |\mathbf{y}_{n_i} - \mathbf{y}_{n_j}| \ll L$.

Without loss of generality, we can assume that the scatterers are located at $\mathbf{y}_1, \dots, \mathbf{y}_M$, that is, $n_j = j$ for $j = 1, \dots, M$. Based on Proposition 2.1 and the assumption of normalized (unit) vectors $\widehat{\mathbf{g}}_0(\mathbf{y}, \omega)$, the submatrix $\mathcal{B}_{M \times M}(\omega)$ is the identity matrix perturbed by a matrix E , with $\|E\|_1 \ll 1$. The value of $\|E\|_1$ depends on the minimal distance between any two scatterers and is close to zero if they are well separated. In summary, $\mathcal{B}(\omega)$ can be written as

$$\mathcal{B} = [\mathcal{B}_{M \times M} \ S] = [I + E \ S],$$

where the matrix S is formed by all the remaining columns with $\|S\|_1 \ll 1$. Each column of S is related to the inner products between Green's function vectors at the

scatterers positions (\mathbf{y}_{n_j} , $j = 1, \dots, M$) and at the grid points with no scatterers (\mathbf{y}_k , $k \in \{1, \dots, K\} \setminus \{n_1, \dots, n_M\}$). Hence, $\|S\|_1$ is determined by the magnitude of inner products between Green's function vectors at the location of the scatterers and at the location of grid points 'closest' to the scatterers, which is the resolution of the IW.

The following result tells us that hybrid ℓ_1 method recovers the reflectivity $\boldsymbol{\rho}_0$ exactly with a mild condition on the matrix S .

Theorem 4.1. *Assume the scatterers are far apart such that (26) holds. If $\|S\|_1 < 1 - \|E\|_1$, then $\boldsymbol{\rho}_0$ is the unique solution to (30).*

Proof: The proof relies on the approximate orthogonality between any two columns of $\mathcal{B}_{M \times M}(\omega)$ and the optimality condition of ℓ_1 minimization problem given in [10]. As in the proof of Theorem 3.1, we need to show that under the assumptions of Theorem 4.1, there exists a vector $\boldsymbol{\mu}$ such that

- for each column \mathbf{c}_j of $I + E$, $\mathbf{c}_j^* \boldsymbol{\mu} = 1$, $j = 1, \dots, M$;
- for each column \mathbf{c}_j of S , $|\mathbf{c}_j^* \boldsymbol{\mu}| < 1$, $j = M + 1, \dots, K$.

Because $I + E$ is nonsingular, let $\boldsymbol{\mu} = (I + E)^{-*} \mathbf{1}$ where $\mathbf{1}$ is the vector of all 1's. Then, part one of the condition is satisfied since $(I + E)^* \boldsymbol{\mu} = \mathbf{1}$. For the second part, we have for any column vector \mathbf{c}_j of S ,

$$|\mathbf{c}_j^* \boldsymbol{\mu}| = |\mathbf{c}_j^* (I + E)^{-*} \mathbf{1}| \leq \|(I + E)^{-1} \mathbf{c}_j\|_{\ell_1} \leq \|(I + E)^{-1}\|_1 \|\mathbf{c}_j\|_{\ell_1} \leq \frac{\|S\|_1}{1 - \|E\|_1} < 1.$$

This completes our proof.

Remark 4.2. *Condition $\|E\|_1 + \|S\|_1 < 1$ relates the distances between scatterers and the resolution implied by the IW discretization. The further apart the scatterers are, the smaller $\|E\|_1$ is and, hence, less constrained $\|S\|_1$ is. To understand better the improvement in resolution given by the hybrid ℓ_1 method with multiple illuminations compared to the ℓ_1 minimization method for single illumination, we write the resolution condition $\|E\|_1 + \|S\|_1 < 1$ in terms of the inner products, so that*

$$\max_{j=1, \dots, M} \left| \sum_{i=1}^M (\widehat{U}_i^*(\omega) \widehat{\mathbf{g}}_0(\mathbf{y}_j, \omega))^2 \right| + \max_{j=M+1, \dots, K} \left| \sum_{i=1}^M (\widehat{U}_i^*(\omega) \widehat{\mathbf{g}}_0(\mathbf{y}_j, \omega))^2 \right| < 1.$$

The inner products in the first term are taken over Green's function vectors at the locations of scatterers and those in the second term are taken over Green's function vectors at the locations of scatterers and the remaining grid points. Let ε_1 and ε_2 be such that

$$\max_{i,j=1, \dots, M, i \neq j} |\widehat{U}_i^*(\omega) \widehat{\mathbf{g}}_0(\mathbf{y}_j, \omega)| < \varepsilon_1, \quad \text{and} \quad \max_{i=1, \dots, M, j=M+1, \dots, K} |\widehat{U}_i^*(\omega) \widehat{\mathbf{g}}_0(\mathbf{y}_j, \omega)| < \varepsilon_2.$$

Note that $\varepsilon_2 = \varepsilon$ in Theorem 3.1 since ε_2 is controlled by the resolution of the IW. Furthermore, $\varepsilon_1 \leq \varepsilon_2$ since we assume that the scatterers are at different grid points. Hence, we have

$$\|E\|_1 + \|S\|_1 < M\varepsilon_1^2 + M\varepsilon_2^2 < 2M\varepsilon^2.$$

If we require the upper bound of $\|E\|_1 + \|S\|_1$ to be less than 1, then $M\varepsilon^2 < 1/2$. Comparing this with the condition in Theorem 3.1 we see that with the same array and IW configuration, by using the optimal (multiple) illuminations more scatterers can be imaged with exact recovery.

5. Other methods for array imaging

In this section, we briefly review some often used methods in array imaging, which will be compared to the ℓ_1 method for single illumination and to the hybrid ℓ_1 method for multiple illuminations in sections 6 and 7, respectively. As mentioned in section 2, to form an image we need to solve for the reflectivity vector $\boldsymbol{\rho}$ from (8) or (9) depending on whether we use single or multiple illumination, respectively. For Kirchhoff migration and for ℓ_2 norm minimization we will discuss the formulation of these two methods based on the case of single illumination. They can be applied to the case of multiple illuminations to the larger set of equations (9) without any modification.

5.1. Kirchhoff migration and least squares imaging

These methods are often used in broadband imaging where they perform quite well when the medium is homogeneous or weakly scattering, leading to range resolutions proportional to the reciprocal of the bandwidth and cross range resolutions proportional to the reciprocal of the array size (for small arrays). They are also very stable and relatively insensitive to additive noise.

If the data is noise free, then the righthand side of (8) is in the range of $\mathcal{A}_{\widehat{f}(\omega)}$. However, with real data this will be at best approximately the case. Therefore, we generalize our criterion for a solution. The straightforward approach is the least squares imaging (LSQ) where one seeks the minimizer of the quadratic cost functional

$$f(\boldsymbol{\rho}) = \frac{1}{2} \|\mathcal{A}_{\widehat{f}(\omega)} \boldsymbol{\rho} - \mathbf{b}(\omega)\|_{\ell_2}^2. \quad (31)$$

This problem has an explicit solution $\boldsymbol{\rho}_{LSQ}$ in terms of the pseudo-inverse of $\mathcal{A}_{\widehat{f}(\omega)}$, denoted by $\mathcal{A}_{\widehat{f}(\omega)}^\dagger$, which has different forms depending on the dimensions of $\mathcal{A}_{\widehat{f}(\omega)}$ ($N > K$ or $N < K$). The sought solution has the well known form

$$\boldsymbol{\rho}_{LSQ} = \mathcal{A}_{\widehat{f}(\omega)}^\dagger \mathbf{b}(\omega) = \begin{cases} (\mathcal{A}_{\widehat{f}(\omega)}^* \mathcal{A}_{\widehat{f}(\omega)})^{-1} \mathcal{A}_{\widehat{f}(\omega)}^* \mathbf{b}(\omega), & \text{if } N > K \\ \mathcal{A}_{\widehat{f}(\omega)}^* (\mathcal{A}_{\widehat{f}(\omega)} \mathcal{A}_{\widehat{f}(\omega)}^*)^{-1} \mathbf{b}(\omega), & \text{if } N < K \end{cases} \quad (32)$$

where $\mathcal{A}_{\widehat{f}(\omega)}^*$ denotes the complex conjugate transpose of $\mathcal{A}_{\widehat{f}(\omega)}$. For large matrices $\mathcal{A}_{\widehat{f}(\omega)}$, the operator $(\mathcal{A}_{\widehat{f}(\omega)}^* \mathcal{A}_{\widehat{f}(\omega)})^{-1}$ in (32) is expensive to compute. Furthermore, it is often very ill-conditioned and therefore (32) is meaningless in the presence of noise. To overcome this difficulty, $(\mathcal{A}_{\widehat{f}(\omega)}^* \mathcal{A}_{\widehat{f}(\omega)})^{-1}$ is often approximated by the identity (or a diagonal matrix) and the Kirchhoff (or weighted Kirchhoff) migration image

$$\boldsymbol{\rho}_{KM} = \mathcal{A}_{\widehat{f}(\omega)}^* \mathbf{b}(\omega) \quad (33)$$

is taken as an approximation to (32). Hence, Kirchhoff migration amounts to applying the adjoint of the matrix $\mathcal{A}_{\hat{f}(\omega)}$ directly to the data.

5.2. ℓ_2 norm minimization or Tikhonov regularization

Another often used imaging method is ℓ_2 minimization, or Tikhonov regularization, in which a weighted quadratic penalty is added to the LSQ term that quantifies the fidelity to the data. The weight of the quadratic penalty provides a tradeoff between fidelity to the data and noise sensitivity. It is well known that ℓ_2 minimization gives stable reconstructions with noisy data, but it over-smoothes the images resulting in poor resolution.

For the underdetermined case ($N < K$), the solution (32) in the previous subsection can be interpreted as the one of the ℓ_2 norm minimization problem

$$\min \|\boldsymbol{\rho}\|_{\ell_2} \quad \text{subject to} \quad \mathcal{A}_{\hat{f}(\omega)} \boldsymbol{\rho} = \mathbf{b}(\omega). \quad (34)$$

This problem can be solved using Lagrange multipliers. Indeed, let us minimize the objective function

$$h(\boldsymbol{\rho}; \boldsymbol{\mu}) = \|\boldsymbol{\rho}\|_{\ell_2} + \boldsymbol{\mu}^* (\mathbf{b}(\omega) - \mathcal{A}_{\hat{f}(\omega)} \boldsymbol{\rho}), \quad (35)$$

where the vector $\boldsymbol{\mu}$ is a Lagrange multiplier, by setting the derivative of $h(\boldsymbol{\rho}; \boldsymbol{\mu})$ with respect to the unknowns $\boldsymbol{\rho}$ and $\boldsymbol{\mu}$ to zero. Then, we obtain the minimum norm solution

$$\boldsymbol{\rho}_{LSQ} = \mathcal{A}_{\hat{f}(\omega)}^* (\mathcal{A}_{\hat{f}(\omega)} \mathcal{A}_{\hat{f}(\omega)}^*)^{-1} \mathbf{b}(\omega). \quad (36)$$

When the data is contaminated by noise, we cannot fit them exactly in (34). Therefore, the constraint in (34) is replaced by $\|\mathcal{A}_{\hat{f}(\omega)} \boldsymbol{\rho} - \mathbf{b}(\omega)\|_{\ell_2} < \varepsilon$, where the constant ε is a positive real parameter related to the amount of noise in the data. In fact, if $\mathcal{A}_{\hat{f}(\omega)} \boldsymbol{\rho} + \mathbf{e}(\omega) = \mathbf{b}(\omega)$ then $\varepsilon \geq \|\mathbf{e}(\omega)\|_{\ell_2}$. When the data is contaminated by noise, the ℓ_2 norm minimization problem looks for the minimum ℓ_2 norm solution while minimizing the discrepancy between $\mathcal{A}_{\hat{f}(\omega)} \boldsymbol{\rho}$ and $\mathbf{b}(\omega)$. Hence,

$$\boldsymbol{\rho}_{\ell_2(\tau)} = \underset{\boldsymbol{\rho}}{\operatorname{argmin}} \{F_2(\boldsymbol{\rho}; \tau) \equiv \frac{\tau}{2} \|\boldsymbol{\rho}\|_{\ell_2}^2 + f(\boldsymbol{\rho})\}, \quad (37)$$

where $f(\boldsymbol{\rho})$ is given by (31). The minimum of $F_2(\boldsymbol{\rho}; \tau)$ is obtained by setting its derivative with respect to the unknown $\boldsymbol{\rho}$ to zero. The minimizer can then be written as

$$\boldsymbol{\rho}_{\ell_2(\tau)} = (\mathcal{A}_{\hat{f}(\omega)}^* \mathcal{A}_{\hat{f}(\omega)} + \tau I)^{-1} \mathcal{A}_{\hat{f}(\omega)}^* \mathbf{b}(\omega), \quad (38)$$

where I is the identity matrix. This solution is also called the Tikhonov regularized solution or the *damped least squares* solution. Compared to (32), (38) replaces the ill-conditioned operator $(\mathcal{A}_{\hat{f}(\omega)}^* \mathcal{A}_{\hat{f}(\omega)})^{-1}$ by the operator $(\mathcal{A}_{\hat{f}(\omega)}^* \mathcal{A}_{\hat{f}(\omega)} + \tau I)^{-1}$ for some positive scalar τ . For very large problems, the later, although better conditioned, is still expensive to compute and therefore the KM image (33) is still often used.

5.3. MUSIC

MUltiple SIgnal Classification (MUSIC) is a signal subspace projection method for localizing and imaging unknown scatterers based on the SVD of the $N \times N$ response matrix $\widehat{P}(\omega)$ [37, 16, 3, 20]. It is an efficient and robust algorithm for localizing well-separated scatterers that are small compared to the wavelength. Even for a single frequency and for non-homogeneous, random media, it is robust to noise and quite accurate for large arrays [3]. Generalizations of MUSIC for extended scatterers have also been developed (see, for example, [25]).

The idea of MUSIC is to project the reference illumination vectors $\widehat{\mathbf{g}}_0(\mathbf{y}_j, \omega)$, $j = 1, \dots, K$, onto the noise space using the projection operator

$$\mathcal{P} = I - \sum_{r=1}^M \widehat{U}_r \widehat{U}_r^*.$$

The search point \mathbf{y}_s corresponds to the location of a scatterer when $\mathcal{P}\widehat{\mathbf{g}}_0(\mathbf{y}_s, \omega) \approx \mathbf{0}$. Thus, we can form an image using the functional

$$\rho_{MUSIC}(\mathbf{y}_s) = \frac{\min_{1 \leq j \leq K} \|\mathcal{P}\widehat{\mathbf{g}}_0(\mathbf{y}_j, \omega)\|_{\ell_2}}{\|\mathcal{P}\widehat{\mathbf{g}}_0(\mathbf{y}_s, \omega)\|_{\ell_2}}, \quad s = 1, \dots, K, \quad (39)$$

so as to locate the points in IW which minimizes the norm of the projection \mathcal{P} onto the noise space by searching for the peaks of (39). The numerator in (39) is a normalization factor.

5.4. MMV with arbitrary illuminations

Multiple Measurement Vectors (MMV) is a matrix form of the linear inverse problem associated with passive array imaging of localized sources when multiple data sets are available at the array. Active array imaging with multiple illuminations that we consider here can also be approached as an MMV problem as we discuss in this section.

In the usual form of MMV it is assumed that a common sensing matrix \mathbf{A} is known and the multiple data received is collected in a matrix \mathbf{Y} , each column of which represents one single measurement. The unknown variable is also a matrix \mathbf{X} where all columns of \mathbf{X} have the same sparse support but possibly different nonzero values. This common support, sparse recovery problem is to solve for \mathbf{X} from the matrix-matrix equation

$$\mathbf{AX} = \mathbf{Y},$$

under the constraint that the columns of \mathbf{X} have the same support.

There are several approaches for getting the solution of the joint sparse recovery problem. For example, in [13] it is shown that orthogonal matching pursuit algorithms can find the sparsest representation of an MMV under certain conditions. In [27, 28], the authors solve the MMV problem through a generalized subspace projection approach. In [40, 38], statistical frameworks are based on certain assumptions on the distribution

of the underlying unknowns. In [29] the authors use the SVD of the data matrix \mathbf{Y} and ℓ_1 regularization. This work is the closest to ours for the case of MMV problems. However MMV deals with passive array imaging and therefore the matrix \mathbf{A} is the same for all the snapshots recorded at the array. The essential difference with active array imaging of scatterers considered here is that there is no possibility of controlling the illumination, and in particular of optimizing it. The active array scattering imaging is essentially different from the passive array MMV imaging problems as we now discuss.

We first reformulate active array imaging of scatterers as a special source localization problem. This amounts to turning the scatterers into sources that have a special amplitude and phase. This leads to a matrix to matrix equation instead of a matrix to vector one as in (9) and the solution is a matrix whose columns share the same sparse support. Assume that we have multiple illumination vectors $\widehat{\mathbf{f}}^{(j)}(\omega)$, $j = 1, 2, \dots, \nu$. For each illumination vector, we have $\mathcal{A}_{\widehat{\mathbf{f}}^{(j)}(\omega)}\boldsymbol{\rho} = \mathbf{b}^{(j)}(\omega)$, $j = 1, 2, \dots, \nu$. Now define

$$\mathcal{G} = [\widehat{\mathbf{g}}_0(\mathbf{y}_1, \omega) \cdots \widehat{\mathbf{g}}_0(\mathbf{y}_K, \omega)]$$

and

$$\boldsymbol{\rho}^{(j)} = \left(\widehat{\mathbf{g}}_{\widehat{\mathbf{f}}^{(j)}}(\mathbf{y}_1, \omega)\rho_1, \dots, \widehat{\mathbf{g}}_{\widehat{\mathbf{f}}^{(j)}}(\mathbf{y}_K, \omega)\rho_K \right)^T = D_{\widehat{\mathbf{f}}^{(j)}(\omega)}\boldsymbol{\rho}, \quad j = 1, 2, \dots, \nu, \quad (40)$$

where $D_{\widehat{\mathbf{f}}^{(j)}(\omega)} = \text{diag}(\widehat{\mathbf{g}}_{\widehat{\mathbf{f}}^{(j)}}(\mathbf{y}_1, \omega), \dots, \widehat{\mathbf{g}}_{\widehat{\mathbf{f}}^{(j)}}(\mathbf{y}_K, \omega))$. Then the scattering problem with multiple illuminations can be rewritten as an MMV problem of the form

$$\mathcal{G} \mathbf{X}_{\widehat{\mathbf{f}}} = \mathbf{Y}, \quad (41)$$

with $\mathbf{Y} = [\mathbf{b}^{(1)}, \mathbf{b}^{(2)}, \dots, \mathbf{b}^{(\nu)}]$ and $\mathbf{X}_{\widehat{\mathbf{f}}} = [\boldsymbol{\rho}^{(1)}, \boldsymbol{\rho}^{(2)}, \dots, \boldsymbol{\rho}^{(\nu)}]$ where we use the subscript $\widehat{\mathbf{f}}$ to denote dependence on the set of illuminations. The matrix \mathcal{G} does not depend on the illuminations, and the columns of $\mathbf{X}_{\widehat{\mathbf{f}}}$ have in general common support. Once this problem is solved by MMV methods the reflectivities $(\rho_1, \rho_2, \dots, \rho_K)$ are obtained from (40).

The problem is that in the MMV formulation (41) it is not clear that active array imaging has special features because it is possible to control the illuminations. To see this, we will now show how the general MMV formulation can be simplified using optimal illuminations so that (41) reduces to (30).

We will assume that the optimal illuminations have been obtained, by iterative time reversal, for example, as described in Appendix B. Because of (26), applying optimal illuminations means taking $\widehat{\mathbf{f}}^{(j)}(\omega) = \widehat{V}_j(\omega)$ and the data matrix becomes

$$\mathbf{Y}_{opt} = [\mathbf{b}^{(1)}, \dots, \mathbf{b}^{(M)}] = [\widehat{P}(\omega)\widehat{V}_1(\omega), \dots, \widehat{P}(\omega)\widehat{V}_M(\omega)] = [\sigma_1(\omega)\widehat{U}_1(\omega), \dots, \sigma_M(\omega)\widehat{U}_M(\omega)].$$

The unknown multishot matrix has the form

$$\mathbf{X}_V = \begin{bmatrix} (\widehat{\mathbf{g}}_0^T(\mathbf{y}_1, \omega)\widehat{V}_1(\omega))\rho_1 & (\widehat{\mathbf{g}}_0^T(\mathbf{y}_1, \omega)\widehat{V}_2(\omega))\rho_1 & \cdots & (\widehat{\mathbf{g}}_0^T(\mathbf{y}_1, \omega)\widehat{V}_M(\omega))\rho_1 \\ \vdots & \vdots & \vdots & \vdots \\ (\widehat{\mathbf{g}}_0^T(\mathbf{y}_K, \omega)\widehat{V}_1(\omega))\rho_K & (\widehat{\mathbf{g}}_0^T(\mathbf{y}_K, \omega)\widehat{V}_2(\omega))\rho_K & \cdots & (\widehat{\mathbf{g}}_0^T(\mathbf{y}_K, \omega)\widehat{V}_M(\omega))\rho_K \end{bmatrix}.$$

We also multiply both sides of (41), with optimal illumination so that $\mathcal{G} \mathbf{X}_V = \mathbf{Y}_{opt}$, by the adjoint of the matrix $\widehat{U}(\omega) = [\widehat{U}_1(\omega), \dots, \widehat{U}_M(\omega)]$ to obtain

$$\widehat{U}^*(\omega) \mathcal{G} \mathbf{X}_V = \begin{bmatrix} \widehat{U}_1^*(\omega) \mathcal{G} D_{\widehat{V}_1(\omega)} \boldsymbol{\rho} & \widehat{U}_1^*(\omega) \mathcal{G} D_{\widehat{V}_2(\omega)} \boldsymbol{\rho} & \cdots & \widehat{U}_1^*(\omega) \mathcal{G} D_{\widehat{V}_M(\omega)} \boldsymbol{\rho} \\ \widehat{U}_2^*(\omega) \mathcal{G} D_{\widehat{V}_1(\omega)} \boldsymbol{\rho} & \widehat{U}_2^*(\omega) \mathcal{G} D_{\widehat{V}_2(\omega)} \boldsymbol{\rho} & \cdots & \widehat{U}_2^*(\omega) \mathcal{G} D_{\widehat{V}_M(\omega)} \boldsymbol{\rho} \\ \vdots & \vdots & \vdots & \vdots \\ \widehat{U}_M^*(\omega) \mathcal{G} D_{\widehat{V}_1(\omega)} \boldsymbol{\rho} & \widehat{U}_M^*(\omega) \mathcal{G} D_{\widehat{V}_2(\omega)} \boldsymbol{\rho} & \cdots & \widehat{U}_M^*(\omega) \mathcal{G} D_{\widehat{V}_M(\omega)} \boldsymbol{\rho} \end{bmatrix}$$

and

$$\widehat{U}^*(\omega) \mathbf{Y}_{opt} = \widehat{U}^*(\omega) \widehat{U}(\omega) \text{diag}(\sigma_1(\omega), \dots, \sigma_M(\omega)) = \text{diag}(\sigma_1(\omega), \dots, \sigma_M(\omega)).$$

so that

$$\widehat{U}^*(\omega) \mathcal{G} \mathbf{X}_V = \widehat{U}^*(\omega) \mathbf{Y}_{opt} \quad (42)$$

The matrix $\widehat{U}^*(\omega) \mathcal{G}$ is $M \times K$ and the matrix \mathbf{X}_V is $K \times M$ so that the right side is $M \times M$. This is a form to which MMV problems can be reduced using the SVD of the data (as in [29]) if only M very particular snapshots are used as we do here with the optimal illuminations. The last step cannot be done for the general MMV problem since there is no possibility of controlling the illuminations.

In active array imaging by choosing the illumination vectors optimally, through the SVD of the array response matrix $\widehat{P}(\omega)$ or by iterative time reversal, the matrix $\widehat{U}^*(\omega) \mathcal{G} \mathbf{X}_V$ is diagonally dominant for any given reflectivity vector $\boldsymbol{\rho}$. In fact, we have that $\widehat{U}_i^*(\omega) \mathcal{G} D_{\widehat{V}_j(\omega)}$ is a $1 \times K$ vector that is approximately zero when $i \neq j$. To see this, notice that the k^{th} entry of $\widehat{U}_i^*(\omega) \mathcal{G} D_{\widehat{V}_j(\omega)}$ is given by $(\widehat{U}_i^*(\omega) \widehat{\mathbf{g}}_0(\mathbf{y}_k, \omega))(\widehat{\mathbf{g}}_0^T(\mathbf{y}_k, \omega) \widehat{V}_j(\omega))$, which can be rewritten as $\widehat{\mathbf{g}}_0^T(\mathbf{y}_k, \omega) \overline{\widehat{U}_i(\omega) \widehat{V}_j^*(\omega)} \widehat{\mathbf{g}}_0(\mathbf{y}_k, \omega)$. As a consequence of Proposition 2.1 and the approximate form (26) of the singular vectors of the response matrix $\widehat{P}(\omega)$, we have that $\widehat{V}_j^*(\omega) \widehat{U}_i(\omega)$ is approximately equal to zero when $i \neq j$. Thus, only when $i = j$ the matrix blocks play a role in the reconstruction. By equating the diagonal entries of the right and left side in (42), we have the following matrix-vector equation

$$\begin{bmatrix} \widehat{U}_1^*(\omega) \mathcal{G} D_{\widehat{V}_1(\omega)} \\ \vdots \\ \widehat{U}_M^*(\omega) \mathcal{G} D_{\widehat{V}_M(\omega)} \end{bmatrix} \boldsymbol{\rho} = \begin{bmatrix} \sigma_1(\omega) \\ \vdots \\ \sigma_M(\omega) \end{bmatrix},$$

which is equivalent to (29).

We see, therefore, that a direct MMV formulation will not take into account the special features of active array imaging so that the system to solve is (42) but without using the right singular vectors \widehat{V}_j , $j = 1, \dots, M$, corresponding to optimal illumination. It will thus fail to note that the off diagonal entries may be nearly zero. This will lead to a more complex problem than when we use optimal illumination and then exploit the properties of the resulting linear system.

6. Comparison of imaging methods: single illumination

In this section we present numerical simulations comparing the performance of the imaging methods described in section 5 for single illumination, that is, KM and ℓ_2 norm minimization with ℓ_1 norm minimization in section 3. We consider simulations in both two and three dimensional space. In these examples and in all that follows, we normalize spatial units by the wavelength λ . Wave propagation between any two points is modeled by the Green's function (2) in both two and three dimensional simulations.

We first consider simulations in two dimensions. We assume that five scatterers are placed within an IW of size 40×40 (see Fig. 2). The reflectivities of the scatterers are 2.96, 2.76, 2.05, 1.54 and 1.35, respectively. The IW is at a distance $L = 100$ from the linear active array. The array, located on the left, has an aperture equal to $a = 100$ and consists of 201 transducers that are half a wavelength apart. We use a uniform lattice on the IW with points separated by 1. This results in a 41×41 uniform mesh. Hence, we have 1681 unknowns and 201 measurements.

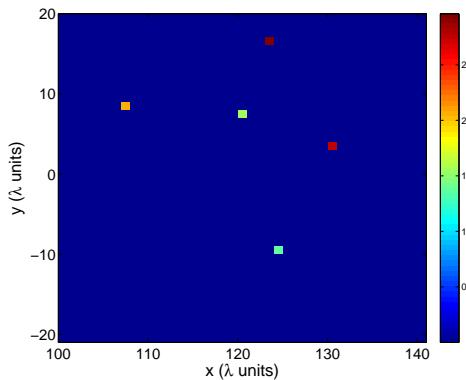


Figure 2. Original configuration of 40×40 image window with grid points separated by 1

Figures 3 (a)-(d) show the results of KM, ℓ_2 norm minimization, and ℓ_1 minimization with and without subsampling, respectively, with no noise in the data. In these examples, we pick the subsampling ratio to be 50%, and we consider single illumination $\hat{f}(\omega)$ coming from the center of the array, that is, $\hat{f}_{101}(\omega) = 1$ and $\hat{f}_p(\omega) = 0$ for any $p \neq 101$. The subsampling in the ℓ_1 minimization method means solving for the unknown reflectivity vector ρ using randomly selected rows in the constraint of (14).

For ℓ_2 norm minimization, we set the regularization parameter to $\tau = 0.3$. Note that in ℓ_2 norm minimization τ controls the trade-off between smoothness and data fidelity. Eq. (38) converges to the Moore-Penrose solution $\rho^\dagger = \mathcal{A}_{\hat{f}(\omega)}^\dagger \mathbf{b}(\omega)$ for $\tau \rightarrow 0$, and to the zero solution $\rho = \mathbf{0}$ for $\tau \rightarrow \infty$. On the other hand, in the ℓ_1 norm minimization (19) the regularization parameter controls the sparsity of the solution, with larger values of τ usually yielding sparser solutions. In contrast to ℓ_2 norm minimization, the convergence to the zero solution occurs for a finite value of τ . If $\tau \geq \|\mathcal{A}_{\hat{f}(\omega)}^* \mathbf{b}(\omega)\|_\infty$, then we obtain

the maximally sparse solution $\rho = \mathbf{0}$ [26]. For noise levels less than or equal to 10% we choose $\tau = 0.01 \|\mathcal{A}_{\hat{f}(\omega)}^* \mathbf{b}(\omega)\|_\infty$. For a large amount of noise in the data (100% noise in the figures below) we choose $\tau = 0.1 \|\mathcal{A}_{\hat{f}(\omega)}^* \mathbf{b}(\omega)\|_\infty$ so more noise is suppressed.

As discussed in section 5, both KM and ℓ_2 norm minimization over-smooth the images, leading to low resolution from which it is hard or impossible to identify the location of the five scatterers (see Figs. 3 (a) and (b)). Furthermore, as a consequence of this over-smoothing, the correct reflectivity values of the scatterers are not recovered. In contrast, when there is no noise in the data, ℓ_1 norm minimization with and without subsampling recovers the positions and reflectivities of the scatterers with precision (see Figs. 3 (c) and (d)). Note that subsampling reduces the dimension of the minimization problem and thus results in a faster computation, though it is more sensitive to noise as we will see next. We also note that the subsampling ratio is picked arbitrarily. A lower subsampling ratio may still recover the correct configuration of the IW but the minimal lower bound of subsampling is hard to determine a priori in array imaging.

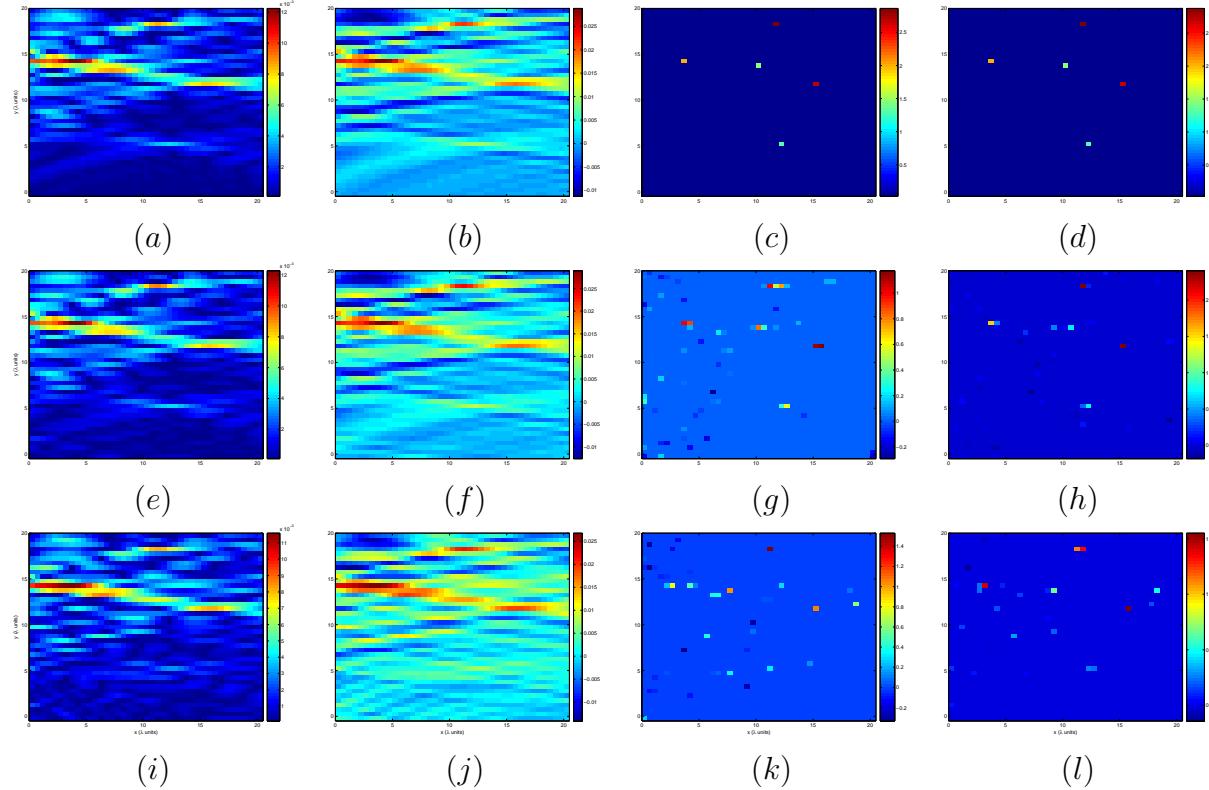


Figure 3. (a) KM with no noise in data. (b) ℓ_2 minimization with no noise in data. (c) ℓ_1 minimization applied to 50% subsamples of noiseless data. (d) ℓ_1 minimization with noiseless data. (e) KM with 10% noise. (f) ℓ_2 minimization with 10% noise. (g) ℓ_1 minimization with 10% noise and 50% subsamples of data. (h) ℓ_1 minimization with 10% noise. (i) KM with 100% noise. (j) ℓ_2 minimization with 100% noise. (k) ℓ_1 minimization with 100% noise and 50% subsamples of data. (l) ℓ_1 minimization with 100% noise.

Next, we examine the performance of KM, ℓ_2 norm minimization, and ℓ_1

minimization (with and without subsampling) when noise is added to the data. For these numerical experiments, we simulate instrument noise by adding a noise matrix $\mathcal{N}(\omega) \in \mathbb{C}^{N \times N}$ with zero mean, uncorrelated Gaussian distributed entries having variance

$$\sigma^2 = \delta \|\widehat{\Pi}(\omega)\|_F^2 / N^2,$$

proportional to the power of response matrix $\widehat{\Pi}(\omega)$ [6]. Here, $\|\cdot\|_F$ is the Frobenius matrix norm, and $\delta > 0$. The signal-to-noise ratio (SNR) of the noise contaminated data is given in db, $-10 \log_{10} \delta$.

Figures 3 (e)-(h) illustrate the results with 10% noise or SNR=10dB. The images obtained by KM and ℓ_2 minimization (Figs. 3 (e) and (f)), although very blurred, do not change too much compared to the noiseless case. On the other hand, the images obtained by ℓ_1 minimization, with and without subsampling, show now a few ghost scatterers and their resolution deteriorates a bit (Figs. 3 (g) and (h)). As a consequence, the recovered reflectivities are not as sharp as in the noiseless case. The result obtained with ℓ_1 minimization and full data is, however, still better than those obtained with KM and ℓ_2 minimization.

When we increase the amount of noise to 100% or SNR=0dB, the KM and ℓ_2 minimization images (Figs. 3 (i) and (j)) show more background noise but they still do not differ much from the noiseless images. These two methods are very stable relative to additive noise. The images obtained by ℓ_1 minimization with and without subsampling (Figs. 3 (k) and (l)) now show more ghost scatterers instead of further decreasing the resolution. In other words, the error terms in the solution of the ℓ_1 minimization method are also sparse, so the method tries to keep the resolution.

Next, we show simulations in 3D. We probe the medium with a 100×100 square array in the xz -plane. The transducers are placed on a 26×26 uniform lattice so they are 4 wavelengths apart. As in the 2D examples, we consider illumination from the center of the array. The size of the IW, K , is a $41 \times 11 \times 41$ three dimensional box, and the grid points are separated by 1. Hence, we have 18491 unknowns and 676 measurements. The number of scatterers in the image window is 5 with the same reflectivity as those in the two dimensional case. The locations of scatterers are $(3, 104, 17)$, $(10, 108, 4)$, $(-13, 100, 9)$, $(0, 110, 8)$, $(4, 107, -9)$. Because it is difficult to illustrate the results in three dimensions, we summarize them by listing the 10 largest values of the solution vector ρ , as well as their corresponding locations in the IW. We use four dimensional vectors $(\mathbf{y}_{n_j}; \rho_{n_j})$ to show the results.

Table 1 shows our results without noise in the data. As expected, we see that ℓ_1 minimization, with and without subsampling, gives the exact locations of scatterers. In these examples, we pick the subsampling ratio to be 22%. The reflectivities are also very close to the true values. KM and ℓ_2 minimization give blurred images. Their solutions are far from the true values of the reflectivities. We refer to these values so as to compare the relative strength between different scattering regions.

Table 2 shows the results with 100% additive noise in data. Compared to the results without noise, there is not much difference when we use KM or ℓ_2 minimization. On

KM	ℓ_2	$\ell_1(S)$	ℓ_1
(3, 104, 17; 0.0398)	(10, 108, 4; 0.1935)	(3, 104, 17; 2.9277)	(3, 104, 17; 2.9291)
(3, 105, 17; 0.0379)	(3, 104, 17; 0.1597)	(10, 108, 4; 2.7313)	(10, 108, 4; 2.7308)
(3, 103, 17; 0.0377)	(0, 110, 8; 0.1124)	(−13, 100, 9; 2.0229)	(−13, 100, 9; 2.0214)
(10, 108, 4; 0.0376)	(−13, 100, 9; 0.1074)	(0, 110, 8; 1.5059)	(0, 110, 8; 1.5113)
(10, 107, 4; 0.0369)	(4, 107, −9; 0.0999)	(4, 107, −9; 1.3221)	(4, 107, −9; 1.3190)
(10, 109, 4; 0.0358)	(3, 105, −11; 0.0827)	(−16, 100, 20; 0)	(−16, 100, 20; 0)
(−13, 100, 9; 0.0342)	(10, 109, 4; 0.0816)	(−17, 100, 20; 0)	(−17, 100, 20; 0)
(10, 106, 4; 0.0336)	(10, 107, 4; 0.0783)	(−18, 100, 20; 0)	(−18, 100, 20; 0)
(3, 106, 17; 0.0327)	(3, 101, 16; 0.0759)	(−19, 100, 20; 0)	(−19, 100, 20; 0)
(3, 102, 17; 0.0321)	(3, 107, 18; 0.0742)	(−20, 100, 20; 0)	(−20, 100, 20; 0)

Table 1. Comparison of the ten largest values in descending order for each method applied to the data without any additive noise. The locations of the 5 scatterers are (3, 104, 17), (10, 108, 4), (−13, 100, 9), (0, 110, 8), (4, 107, −9), and their reflectivities are, in the same order, 2.96, 2.76, 2.05, 1.54 and 1.35. The first three entries in each case give the location and the last entry gives the value of the reflectivity. Here $\ell_1(S)$ means the results obtained by ℓ_1 minimization with subsampling.

the other hand, ℓ_1 minimization applied to full data (without subsampling) still gives pretty good results regarding resolution and reflectivity. Observe that the values of the reflectivities change a lot between the 5th and 6th rows in the last column of Table 2. This means that there are ghost scatterers brought in by the high noise in the data. Note also that the result of ℓ_1 minimization with subsampled data is not as good as the one with full data. Only four scatterers appear among the top ten locations with largest values. Moreover, we could not deduce the number of scatterers based on these values since there is not a significant difference between two consecutive values. Hence, to obtain a better image with this amount of noise in the data it is necessary to have more data, which in this context means more illuminations.

7. Comparison of simulations with multiple illuminations

In this section, we present simulation results using the hybrid ℓ_1 method (30). We compare hybrid ℓ_1 images with the results obtained by the other imaging methods described in sections 3 and 5. We use a setup similar to the one in section 6. Five scatterers are placed in the IW, with locations and reflectivities equal to those given in section 6. However, in two dimensions, the array of aperture $a = 100$ has 51 equally spaced transducers so that the total amount of data collected with multiple illuminations is about the same as that with single illumination in section 6.

We note that the hybrid ℓ_1 method relies on the use of all essential singular vectors in (30), that is singular vectors with singular values clearly above the noise level. The essential singular vectors correspond one-to-one with the scatterers when they are well

KM	ℓ_2	$\ell_1(S)$	ℓ_1
(3, 104, 17; 0.0432)	(10, 108, 4; 0.02081)	(3, 104, 17; 2.5308)	(3, 104, 17; 2.9560)
(3, 103, 17; 0.0407)	(3, 104, 17; 0.1698)	(10, 108, 4; 1.8573)	(10, 108, 4; 2.6301)
(10, 108, 4; 0.0402)	(0, 110, 8; 0.1215)	(0, 110, 8; 0.8526)	(−13, 100, 9, 1.7191)
(10, 107, 4; 0.0397)	(−13, 100, 9; 0.0993)	(−13, 100, 9; 0.6993)	(0, 110, 8; 1.3480)
(3, 105, 17; 0.0393)	(10, 109, 4; 0.0976)	(5, 103, 15; 0.5978)	(4, 107, −9; 0.8096)
(10, 109, 4; 0.0381)	(3, 101, 16; 0.0844)	(10, 103, 5; 0.4984)	(−3, 100, 14; 0.3411)
(10, 106, 4; 0.0366)	(4, 107, −9; 0.0832)	(9, 104, −15; 0.4834)	(−2, 106, 4; 0.3095)
(3, 102, 17; 0.0347)	(3, 107, 18; 0.0815)	(2, 100, −20; 0.4404)	(−2, 107, 2; 0.2742)
(10, 110, 4; 0.0339)	(10, 107, 4; 0.0794)	(2, 102, −11; 0.3866)	(−18, 108, 19; 0.2731)
(3, 106, 17; 0.0329)	(11, 102, 4; 0.0783)	(8, 108, −4; 0.3654)	(4, 104, −4; 0.2717)

Table 2. Same as Table 1 but with 100% noise added to the data.

separated and, in any case, there are as many essential singular vectors as scatterers [3]. Nevertheless, we will consider what happens when the number of singular vectors used in (30) is either less or more than the number of scatterers, which means that we either omit an essential singular vector or include one from the noise subspace.

7.1. Using all essential singular vectors

We set $\tau = 0.3$ for ℓ_2 norm minimization, and $\tau = 0.01 \|\mathcal{A}_{\widehat{f}(\omega)}^* \mathbf{b}(\omega)\|_\infty$ and $\tau = 0.1 \|\mathcal{A}_{\widehat{f}(\omega)}^* \mathbf{b}(\omega)\|_\infty$ for ℓ_1 norm minimization with noiseless data and noisy data, respectively. For KM, ℓ_2 minimization and ℓ_1 minimization, we collect the data $\mathbf{b}^{(j)}(\omega) = \widehat{\Pi}(\omega) \widehat{\mathbf{f}}^{(j)}(\omega) + \mathbf{e}^{(j)}(\omega)$ generated by $\nu = 5$ arbitrary illumination vectors $\widehat{\mathbf{f}}^{(j)}(\omega)$, $j = 1, \dots, \nu$. The noise vectors $\mathbf{e}^{(j)}(\omega)$ associated to each illumination are independent of each other. Each $\widehat{\mathbf{f}}^{(j)}(\omega)$ satisfies that $\widehat{f}_p^{(j)}(\omega) = 1$ for $p = j$ and $\widehat{f}_p^{(j)}(\omega) = 0$ for $p \neq j$. With the measurements of multiple illuminations, we use the KM, ℓ_2 minimization, and ℓ_1 minimization as described in sections 3 and 5 to solve the larger set of equations given in (9).

We note that MUSIC and hybrid ℓ_1 differ from KM, ℓ_2 minimization, and ℓ_1 minimization because they do not take inputs generated by arbitrary multiple illuminations directly. They assume knowledge of the significant singular vectors of the response matrix $\widehat{P}(\omega)$ to form an image. In the simulations for MUSIC and hybrid ℓ_1 we assume that the top 5 singular values and singular vectors have been obtained. Thus, the comparison between all the images presented below is carried out with about the same amount of data as in the simulations done in section 6.

Figure 4 shows the simulation results in two dimensions. The images obtained from clean data are given in Figs. 4 (a1)-(f1). After collecting data from more illuminations, the resolution of images using KM and ℓ_2 minimization, even though still blurred, improves dramatically compared to the results shown in Figs. 3 (a) and (b). As expected,

MUSIC provides the correct location of the scatterers. Regarding the images obtained with ℓ_1 minimization (with and without subsampled data), we observe that both the location and reflectivity of the scatterers are recovered accurately, as in the single illumination case. Also the hybrid ℓ_1 method gives us perfect recovery of the scatterers, both location and reflectivity.

Figures 4 (a2)-(f2) give the results after adding 100% noise to the data. The resolution of the images obtained by KM and ℓ_2 minimization does not differ much from that obtained with clean data in Figs. 4 (a1) and (b1). These methods are very stable with respect to additive noise. On the other hand, the image obtained with MUSIC is now degraded (Fig. 4 (c2)), and ℓ_1 minimization gives rise to several ghost scatterers (Figs. 4 (d2) and(e2)). However, the hybrid ℓ_1 method still delivers sharp images (Fig. 4 (f2)). It recovers the location of the scatterers and gives accurate estimates of their reflectivities. The robustness of the hybrid ℓ_1 method relies on the selection of the top or essential singular values so that the noise in the subspace formed by singular vectors associated with small singular values is filtered out.

Next, we show simulation results in three dimensions. The imaging array is still a 100×100 square but now has 13×13 transducers placed uniformly. For each method, we list the corresponding 10 pixels with largest values of the recovered images and their locations. Same as in the tables shown in section 6, the first three values of the vector give the location of the scatterers and the fourth value gives their reflectivities. We also use 5 illumination vectors $\hat{\mathbf{f}}^{(j)}(\omega)$ to obtain the data. Table 3 summarizes the results with noiseless data. We observe that in 3D KM and ℓ_2 minimization still blur the images with 5 arbitrary illuminations. The original scatterers appear as if they are spread out. This is so, because in 3D the ratio between the number of unknowns and the number of measurements is larger than in the 2D experiments and, therefore, 5 arbitrary illuminations are not enough for KM and ℓ_2 minimization to improve their images.

On the other hand, MUSIC, ℓ_1 minimization, and hybrid ℓ_1 method are able to recover the sparse solution. The locations of the five scatterers are found exactly by these methods. At the same time, ℓ_1 minimization and the hybrid ℓ_1 method also give very good estimates of the reflectivity of each scatterer.

We show in Table 4 the results when 100% noise is added to the data generated by each illumination. As in the two dimensional case, we observe that KM and ℓ_2 minimization are very stable with respect to additive noise, although the positions of the scatterers are still blurred. Also, MUSIC and ℓ_1 minimization produce blurred positions of the scatterers when 100% noise is added to the data and 5 illuminations $\hat{\mathbf{f}}^{(j)}(\omega)$ are used. However, the estimates of reflectivity obtained by ℓ_1 minimization give us a better idea on how many (strong) scatterers are in the image window by checking the difference between consecutive estimated values. Finally, the hybrid ℓ_1 method, by combining the advantage of both SVD and ℓ_1 minimization, gives a very good solution (as good as with no noise, see Table 3). It picks out the location of scatterers exactly and gives the sparsest solution even with data contaminated by 100% noise.

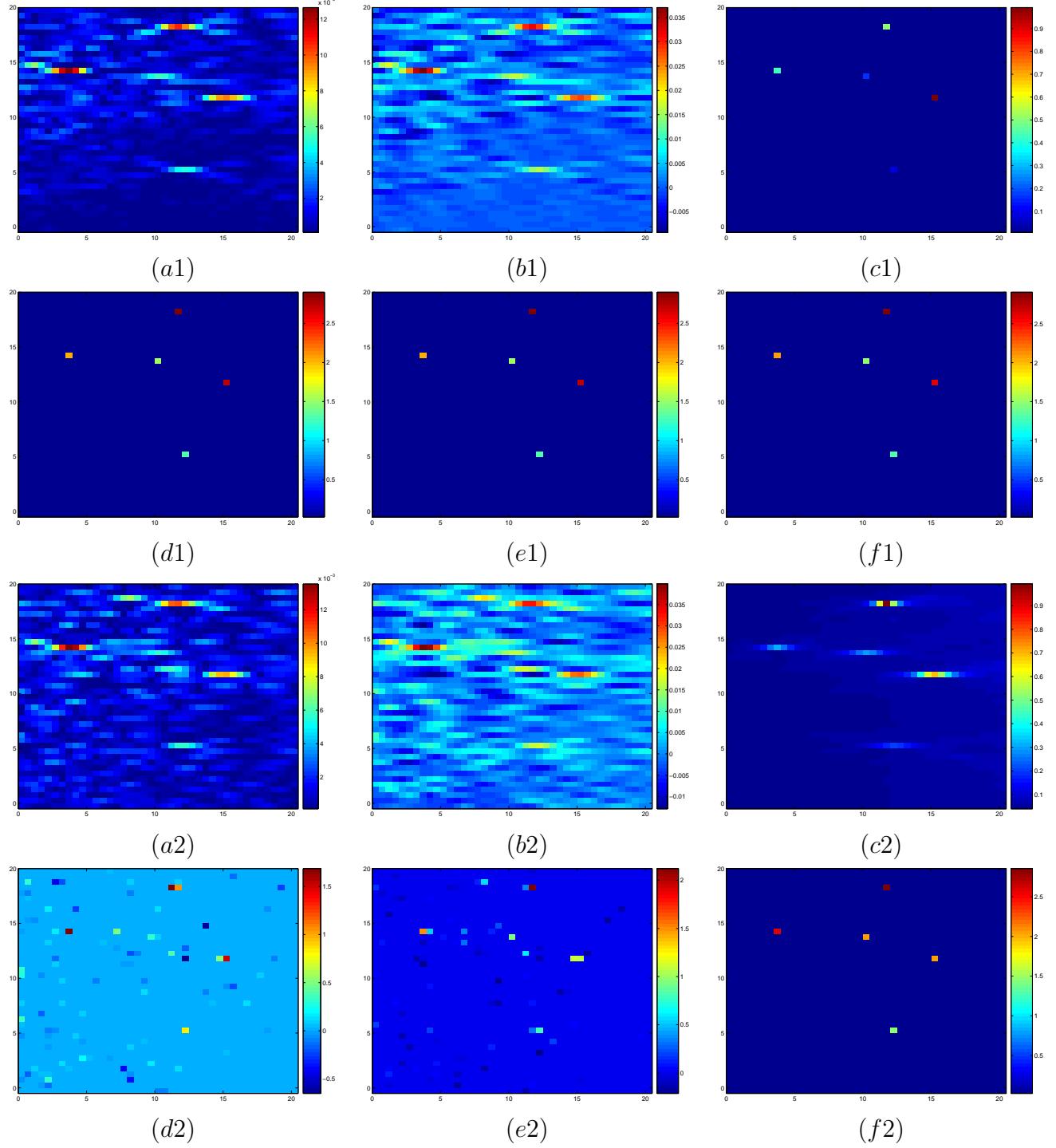


Figure 4. (a1) KM with no noise in data. (b1) ℓ_2 minimization with no noise in data. (c1) MUSIC with no noise in data. (d1) ℓ_1 minimization applied to 50% subsamples of noiseless data. (e1) ℓ_1 minimization with full and noiseless data. (f1) Hybrid ℓ_1 with no noise. (a2) KM with 100% noise. (b2) ℓ_2 minimization with 100% noise. (c2) MUSIC with 100% noise in data. (d2) ℓ_1 minimization with 100% noise and 50% subsamples of data. (e2) ℓ_1 minimization with 100% noise. (f2) Hybrid ℓ_1 with 100% noise in data. We have used 5 illuminations in KM, ℓ_2 minimization, and ℓ_1 minimization with and without subsampling.

KM	ℓ_2	MUSIC
(3, 104, 17; 0.057)	(−13, 100, 9; 0.103)	(3, 104, 17; 1)
(−13, 100, 9; 0.0538)	(3, 104, 17; 0.077)	(0, 110, 8; 0.3332)
(10, 108, 4; 0.0432)	(10, 108, 4; 0.069)	(−13, 100, 9; 0.2947)
(3, 101, 17; 0.0343)	(−13, 101, 9; 0.0529)	(4, 107, −9; 0.2742)
(3, 102, 17; 0.0342)	(0, 110, 8; 0.0473)	(10, 108, 4; 0.1664)
(3, 107, 17; 0.0335)	(−11, 105, 17; 0.0439)	(0, 109, 8; 0)
(10, 106, 4; 0.0334)	(4, 107, −9; 0.0432)	(4, 108, −9; 0)
(−11, 105, 17; 0.0327)	(0, 100, 9; 0.0375)	(10, 109, 4; 0)
(0, 110, 8; 0.0313)	(−11, 106, 17; 0.0352)	(4, 106, −9; 0)
(−11, 105, 3; 0.0308)	(17, 104, 17; 0.0338)	(10, 107, 4; 0)
$\ell_1(S)$	ℓ_1	hybrid ℓ_1
(3, 104, 17; 2.9315)	(3, 104, 17; 2.9280)	(3, 104, 17; 2.9312)
(10, 108, 4; 2.7204)	(10, 108, 4; 2.7252)	(10, 108, 4; 2.7268)
(−13, 100, 9; 2.0278)	(−13, 100, 9; 2.0309)	(−13, 100, 9; 2.04)
(0, 110, 8; 1.5062)	(0, 110, 8; 1.5093)	(0, 110, 8; 1.4845)
(4, 107, −9; 1.3125)	(4, 107, −9; 1.3137)	(4, 107, −9; 1.3176)
(−16, 100, 20; 0)	(−16, 100, 20; 0)	(−16, 100, 20; 0)
(−17, 100, 20; 0)	(−17, 100, 20; 0)	(−17, 100, 20; 0)
(−18, 100, 20; 0)	(−18, 100, 20; 0)	(−18, 100, 20; 0)
(−19, 100, 20; 0)	(−19, 100, 20; 0)	(−19, 100, 20; 0)
(−20, 100, 20; 0)	(−20, 100, 20; 0)	(−20, 100, 20; 0)

Table 3. Comparison of the ten largest values in descending order for each method applied to data with no noise. The locations of the 5 scatterers are (3, 104, 17), (10, 108, 4), (−13, 100, 9), (0, 110, 8), (4, 107, −9), and their reflectivities are, in the same order, 2.96, 2.76, 2.05, 1.54 and 1.35. The first three entries of each tuple give the location and the last entry gives the value of the reflectivity. $\ell_1(S)$ means the results obtained by ℓ_1 minimization with subsampling.

Remark 7.1. Based on the choice of the regularization parameter τ , the images resulting from ℓ_1 minimization with FISTA may be different. For an ℓ_1 minimization algorithm that does not depend on τ we refer to [33]. The choice of τ in FISTA is more important when there is noise present in data. In our simulations, we select the one we found to be the best one among values we tried. However, since the SVD that is used in hybrid ℓ_1 is also a way to filter out noise, we have more flexibility in choosing τ . Based on our simulations, any reasonable value used in the hybrid ℓ_1 method gives a clear sparse image.

7.2. Using fewer than the essential singular vectors or using some inessential ones

We show now results when some essential singular vectors used in (30) are omitted or when some singular vectors from the noise subspace are included. We will see that in

KM	ℓ_2	MUSIC
(3, 104, 17; 0.0579)	(−13, 100, 9; 0.1007)	(3, 104, 17; 1)
(−13, 100, 9; 0.0532)	(3, 104, 17; 0.0783)	(10, 108, 4; 0.09051)
(10, 108, 4; 0.0417)	(10, 108, 4; 0.0681)	(−13, 100, 9; 0.5073)
(3, 102, 17; 0.0355)	(−13, 101, 9; 0.0528)	(0, 110, 8; 0.2876)
(3, 101, 17; 0.0353)	(0, 110, 8; 0.0495)	(4, 107, −9; 0.2344)
(3, 107, 17; 0.0334)	(−11, 105, 17; 0.044)	(10, 109, 4; 0.2299)
(−11, 105, 17; 0.0326)	(4, 107, −9; 0.0401)	(10, 107, 4; 0.2112)
(0, 110, 8; 0.0318)	(17, 104, 17; 0.0368)	(0, 109, 8; 0.1788)
(10, 110, 4; 0.0312)	(1, 103, 9; 0.0353)	(3, 103, 17; 0.1679)
(−11, 104, 3; 0.0301)	(−11, 106, 17; 0.0323)	(3, 105, 17; 0.1670)
$\ell_1(S)$	ℓ_1	hybrid ℓ_1
(3, 107, 4; 2.4403)	(3, 107, 4; 2.6937)	(3, 107, 4; 2.7303)
(10, 108, 4; 2.1794)	(10, 108, 4; 2.1798)	(10, 108, 4; 2.5259)
(−13, 100, 9; 1.7689)	(−13, 100, 9; 1.8011)	(−13, 100, 9; 1.9926)
(0, 110, 8; 1.3540)	(0, 110, 8; 1.2586)	(0, 110, 8; 1.3066)
(4, 107, −9; 0.6035)	(4, 107, −9; 0.8584)	(4, 107, −9; 1.2296)
(−15, 107, 13; 0.2139)	(−14, 107, 2; 0.1875)	(−16, 100, 20; 0)
(−12, 101, 12; 0.1982)	(5, 110, 8; 0.1661)	(−17, 100, 20; 0)
(−19, 100, −14; 0.1971)	(−15, 107, 13; 0.1481)	(−18, 100, 20; 0)
(−10, 100, −8; 0.1922)	(−18, 107, −9; 0.1433)	(−19, 100, 20; 0)
(−20, 102, 5; 0.1830)	(−8, 102, −1; 0.1396)	(−20, 100, 20; 0)

Table 4. Same as Table 3 but with 100% noise added to the data.

the first case the hybrid ℓ_1 method locates as many scatterers as the essential singular vectors used in (30). On the other hand, when more than the essential singular vectors are used, the hybrid ℓ_1 method locates all the scatterers without producing ghosts. This is, in fact, a direct consequence of Theorem 4.1, which states that if the scatterers are far apart and the noise level is below a threshold the solution is unique. In the following, we show simulation results in 2D to illustrate what happens when the singular vectors used to form (30) are not all the essential ones.

In Fig. 5, we show images similar to those in Fig. 4 but using 4 essential singular vectors (left image) and 6 singular vectors (right image), all five essential ones and one from the noise subspace. We consider 100% noise in the data. In the first case only 4 scatterers are located in the image because when optimal illuminations are used each one is aimed at one scatterer at a time (see Appendix B). The one with lowest reflectivity is missing because illumination by the essential singular vector that has the smallest singular value above the noise level is missing. In the second case, all the scatterers are recovered in the image and there are no ghosts.

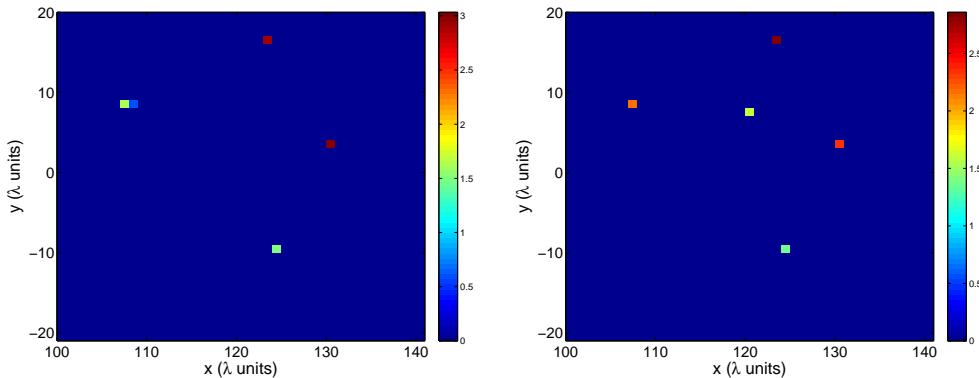


Figure 5. Left: reconstruction with the top 4 singular values and associated vectors. Right: reconstruction with the top 6 singular values and associated singular vectors. There is 100% noise in the data.

8. Conclusion

In this paper, we study array imaging of localized scatterers in a homogeneous medium with and without additive noise. We formulate the problem as an ℓ_1 minimization and compare the images formed with various benchmark imaging methods such as Kirchhoff migration, ℓ_2 minimization, and MUSIC. When only a single illumination is used to probe the medium, we show that the ℓ_1 minimization can recover the locations and reflectivities of scatterers in the image window within suitable resolution limits. This is verified with numerical simulations which show that ℓ_1 minimization is superior to Kirchhoff migration and ℓ_2 minimization for recovering the location of scatterers in a sparse configuration. We also introduce a hybrid ℓ_1 method that uses the essential information contained in the top singular vectors of the array response matrix and ℓ_1 minimization. This method reduces the dimensionality of the problem, filters out noise from the data, and keeps all the essential properties of ℓ_1 minimization for handling sparse configurations of scatterers. Furthermore, the hybrid ℓ_1 approach leads to a better resolution and a larger number of scatterers can be located within the same imaging configuration. Numerical simulations confirm that when the noise level is high, hybrid ℓ_1 minimization produces sharper images than using ℓ_1 minimization directly. Moreover, it is robust to the number of singular vectors used to reconstruct the image.

Acknowledgment. AC was supported by a Hewlett Packard Stanford Graduate Fellowship. MM was supported by the Spanish Ministry of Education and Science grant FIS2010-18473. GP was supported by AFOSR grant FA9550-11-1-0266.

Appendix A. Proof of Proposition

Without loss of generality, we fix $\mathbf{y} = (0, 0, L)$ and alter \mathbf{y}^S . The transducers are spaced by a distance $h \sim O(\lambda)$. Because $\lambda \ll |\mathbf{y}^S - \mathbf{y}| \ll L$, we have $\frac{|\mathbf{y}^S - \mathbf{y}|}{|\mathbf{x} - \mathbf{y}|} \ll 1$ and

$$\begin{aligned} |\mathbf{x} - \mathbf{y}^S| &= |\mathbf{x} - \mathbf{y}| + (\mathbf{y}^S - \mathbf{y}) \cdot \frac{\mathbf{y} - \mathbf{x}}{|\mathbf{x} - \mathbf{y}|} \\ &\quad + (\mathbf{y}^S - \mathbf{y}) \cdot \left(\frac{I}{|\mathbf{x} - \mathbf{y}|} - \frac{(\mathbf{x} - \mathbf{y}) \otimes (\mathbf{x} - \mathbf{y})}{|\mathbf{x} - \mathbf{y}|^3} \right) \cdot (\mathbf{y}^S - \mathbf{y}) + \dots \\ &\approx |\mathbf{x} - \mathbf{y}| + (\mathbf{y}^S - \mathbf{y}) \cdot \frac{\mathbf{y} - \mathbf{x}}{|\mathbf{x} - \mathbf{y}|} + o(1). \end{aligned}$$

With this first order approximation, we can estimate the inner product of Green's function vectors as follows.

$$\begin{aligned} \widehat{\mathbf{g}}_0^*(\mathbf{y}, \omega) \widehat{\mathbf{g}}_0(\mathbf{y}^S, \omega) &= \sum_{\mathbf{x}} \overline{\widehat{G}_0(\mathbf{x}, \mathbf{y}, \omega)} \widehat{G}_0(\mathbf{x}, \mathbf{y}^S, \omega) \\ &= \sum_{\mathbf{x}} \frac{\exp(-i\kappa|\mathbf{x} - \mathbf{y}|)}{4\pi|\mathbf{x} - \mathbf{y}|} \frac{\exp(i\kappa|\mathbf{x} - \mathbf{y}^S|)}{4\pi|\mathbf{x} - \mathbf{y}^S|} \\ &\simeq \frac{1}{(4\pi)^2 h^2} \int_{\Omega(\mathbf{x})} \frac{\exp(i\kappa(|\mathbf{x} - \mathbf{y}^S| - |\mathbf{x} - \mathbf{y}|))}{|\mathbf{x} - \mathbf{y}^S| |\mathbf{x} - \mathbf{y}|} d\mathbf{x} \\ &\simeq \frac{1}{(4\pi)^2 h^2} \int_{\Omega(\mathbf{x})} \frac{\exp(-i\kappa(\mathbf{y}^S - \mathbf{y}) \cdot \frac{\mathbf{x} - \mathbf{y}}{|\mathbf{x} - \mathbf{y}|})}{|\mathbf{x} - \mathbf{y}|^2} d\mathbf{x} \end{aligned}$$

where $\Omega(\mathbf{x})$ is the integral region. To estimate the inner product, now we only need to work out the estimate of the integral over $\Omega(\mathbf{x})$.

We denote

$$\mathcal{I} = \int_{\Omega(\mathbf{x})} \frac{\exp(-i\kappa(\mathbf{y}^S - \mathbf{y}) \cdot \frac{\mathbf{x} - \mathbf{y}}{|\mathbf{x} - \mathbf{y}|})}{|\mathbf{x} - \mathbf{y}|^2} d\mathbf{x}.$$

Because \mathbf{x} only varies on the 2D plane, we can use the following spherical coordinate transformation

$$\frac{\mathbf{x} - \mathbf{y}}{|\mathbf{x} - \mathbf{y}|} = (\cos \theta \sin \phi, \sin \theta \sin \phi, \cos \phi),$$

where θ is the azimuthal angle on xy -plane and ϕ is the polar angle. Based on the geometry, we have $|\mathbf{x} - \mathbf{y}| = L \sec \phi$ and Jacobian determinant of the transform from \mathbf{x} to (θ, ϕ) as

$$\frac{\partial \mathbf{x}}{\partial (\theta, \phi)} = \begin{vmatrix} -L \sin \theta \tan \phi & L \cos \theta \sec^2 \phi \\ L \cos \theta \tan \phi & L \sin \theta \sec^2 \phi \end{vmatrix} = -L^2 \sec^2 \phi \tan \phi.$$

Then integral in spherical coordinate the integral becomes

$$\mathcal{I} = \int_{\Omega(\theta, \phi)} \exp(-i\kappa(\mathbf{y}^S - \mathbf{y}) \cdot (\cos \theta \sin \phi, \sin \theta \sin \phi, \cos \phi)) \tan \phi d\theta d\phi.$$

In the above integral, $\Omega(\theta, \phi) = [0, 2\pi] \times [0, \epsilon]$ where $\epsilon = \arctan \frac{a}{2L} < \frac{\pi}{2}$ so that the integrant is nonsingular within the integral region $\Omega(\theta, \phi)$.

Next we discuss the behavior of the integral \mathcal{I} in three cases:

case 1 $\mathbf{y}^S = (0, 0, L + \eta)$ and $\mathbf{y}^S - \mathbf{y} = (0, 0, \eta)$. We have

$$\begin{aligned}\mathcal{I} &= \int_0^{2\pi} \int_0^\epsilon \exp(-i\kappa\eta \cos \phi) \times \frac{\sin \phi}{\cos \phi} d\phi d\theta \\ &= 2\pi \int_{\cos \epsilon}^1 \frac{\exp(-i\kappa\eta z)}{z} dz.\end{aligned}$$

The function $f(z) := \frac{1}{z}$ is integrable over interval $[\cos \epsilon, 1]$. Notice that $\kappa = \frac{\omega}{c} = \frac{2\pi}{\lambda}$ and $\eta = |\mathbf{y}^S - \mathbf{y}|$. Then we have $\kappa\eta = 2\pi \frac{|\mathbf{y}^S - \mathbf{y}|}{\lambda} \rightarrow \infty$ and by Riemann-Lebesgue lemma, $\mathcal{I} \rightarrow 0$.

case 2 $\mathbf{y}^S = (\eta, 0, L)$ and $\mathbf{y}^S - \mathbf{y} = (\eta, 0, 0)$. We have

$$\begin{aligned}\mathcal{I} &= \int_0^{2\pi} \int_0^\epsilon \exp(i\kappa\eta \sin \phi \cos \theta) \times \frac{\sin \phi}{\cos \phi} d\phi d\theta \\ &= \int_0^\epsilon d\phi \frac{\sin \phi}{\cos \phi} \int_0^{2\pi} d\theta \exp(-i\kappa\eta \sin \phi \cos \theta) \\ &= \int_0^\epsilon d\phi \frac{\sin \phi}{\cos \phi} \left(\int_0^\pi d\theta \exp(-i\kappa\eta \sin \phi \cos \theta) + \int_\pi^{2\pi} d\theta \exp(-i\kappa\eta \sin \phi \cos \theta) \right) \\ &= \int_0^\epsilon d\phi \frac{\sin \phi}{\cos \phi} \left(\int_0^\pi d\theta \exp(-i\kappa\eta \sin \phi \cos \theta) + \int_0^\pi d\theta \exp(i\kappa\eta \sin \phi \cos \theta) \right) \\ &= 2 \int_0^\epsilon d\phi \frac{\sin \phi}{\cos \phi} \int_0^\pi d\theta \cos(\kappa\eta \sin \phi \cos \theta) \\ &= 2\pi \int_0^\epsilon d\phi \frac{\sin \phi}{\cos \phi} \mathcal{J}_0(\kappa\eta \sin \phi) \\ &= 2\pi \int_0^{\sin \epsilon} dt \mathcal{J}_0(\kappa\eta t) \frac{t}{1-t^2}.\end{aligned}$$

In the above derivation, $\mathcal{J}_0(\cdot)$ is the type-I Bessel function of order 0. It satisfies

$$\mathcal{J}_0(x) \approx \begin{cases} 1, & 0 < x \ll 1, \\ \sqrt{\frac{2}{\pi x}} \cos\left(x - \frac{\pi}{4}\right), & x \gg \frac{1}{8}. \end{cases}$$

For small $\epsilon \ll 1$, the integral can be further simplified as

$$\begin{aligned}\mathcal{I} &= 2\pi \left(\int_0^\epsilon + \int_\epsilon^{\sin \epsilon} \right) \left(\mathcal{J}_0(\kappa\eta t) \frac{t}{1-t^2} \right) dt \\ &\approx 2\pi \int_0^\epsilon \frac{t}{1-t^2} dt + 2\pi \int_\epsilon^{\sin \epsilon} \sqrt{\frac{2}{\pi \kappa \eta t}} \cos\left(\kappa\eta t - \frac{\pi}{4}\right) \frac{t}{1-t^2} dt + o(\epsilon) \\ &= -\pi \log(1 - \epsilon^2) + \frac{2\sqrt{2\pi}}{\sqrt{\kappa\eta}} \int_\epsilon^{\sin \epsilon} \frac{\sqrt{t}}{1-t^2} \cos\left(\kappa\eta t - \frac{\pi}{4}\right) dt + o(\epsilon)\end{aligned}$$

Because $\log(1 - \epsilon^2) = -\epsilon^2 + o(\epsilon^2)$ and the second term approaches to zeros as $k\eta \rightarrow \infty$, we thus have integral \mathcal{I} is asymptotically zero.

case 3 Search point is $\mathbf{y}^S = (0, \eta, L)$ and $\mathbf{y}^S - \mathbf{y} = (0, \eta, 0)$. Similar to **case 2**, we have

$$\begin{aligned}\mathcal{I} &= \int_0^\epsilon d\phi \frac{\sin \phi}{\cos \phi} \int_0^{2\pi} d\theta \exp(-i\kappa\eta \sin \phi \sin \theta) \\ &= 2\pi \int_{\cos \epsilon}^1 \frac{\mathcal{J}_0(\kappa\eta\sqrt{1-z^2})}{z} dz.\end{aligned}$$

Based on similar derivation to **case 2** we have integral asymptotically equal to zero.

The general placement of points \mathbf{y} and \mathbf{y}^S can be bounded by the above three cases and hence we complete the proof.

Remark Appendix A.1. *The result given in Proposition 2.1 only requires that the size of the array satisfies the condition $\arctan \frac{a}{2L} < \frac{\pi}{2}$. The result applies to either large or small arrays. The behavior of $\widehat{\mathbf{g}}_0^*(\mathbf{y}, \omega)\widehat{\mathbf{g}}_0(\mathbf{y}^S, \omega)$ is well known to be a sinc function when the aperture a is small. To see how Proposition 2.1 gives the small array result, note that in the derivation of **case 2** or **case 3**, for small ϵ , integral $\int_0^\epsilon (\dots) dt$ corresponds to the case in which the aperture satisfies $0 \leq \sin \phi \leq \epsilon$. This implies that*

$$\frac{a}{2L} = \tan \phi \approx \sin \phi \ll 1 \iff a \ll L.$$

In other words, the estimate of the integral corresponds to the setting where the size of the imaging array is small. However, our result only gives the asymptotic behavior of $\widehat{\mathbf{g}}_0^(\mathbf{y}, \omega)\widehat{\mathbf{g}}_0(\mathbf{y}^S, \omega)$ and does not provide a more quantitative characterization.*

Appendix B. Iterative time reversal

In principle, the hybrid ℓ_1 method assumes that the full array response matrix $\widehat{P}(\omega)$ is available and uses its singular vectors to form an image. The singular vectors, or only the top singular vectors, can then be obtained from this matrix numerically. However, it is not necessary to have the full array response matrix available in order to obtain the top singular values and singular vectors. As noted in the previous section, an iterative time reversal acquisition process can be used for this purpose [36, 31, 32]. This iterative technique obtains the necessary part of the singular value decomposition as part of the data collection process.

The iterative time reversal procedure is a physical realization of the *power method* used in numerical linear algebra for finding eigenvectors corresponding to distinct eigenvalues of a matrix [22]. In this process, an illumination vector is first emitted into the medium. The reflected signals received at the array are then time reversed and re-emitted into the medium. After repeating these steps over the medium for a few times, the data at the array is approximately the eigenvector associated with the largest eigenvalue of the matrix $\widehat{P}(\omega)\widehat{P}^*(\omega)$, that is, the illumination vector that tends to focus on the scatterer with the largest reflectivity.

The above process of approximating the illumination vector associated with the target of largest reflectivity is summarized as follows:

- (i) given an initial illumination unit vector $\hat{\mathbf{f}}_0(\omega)$, transmit $\overline{\hat{\mathbf{f}}_0(\omega)}$,
- (ii) receive $\hat{\mathbf{f}}_{1/2}(\omega) = \widehat{P}(\omega) \overline{\hat{\mathbf{f}}_0(\omega)}$ and transmit $\overline{\hat{\mathbf{f}}_{1/2}(\omega)}$,
- (iii) receive $\hat{\mathbf{f}}_1(\omega) = \widehat{P}(\omega) \overline{\hat{\mathbf{f}}_{1/2}(\omega)} = \widehat{P}(\omega) \widehat{P}^*(\omega) \hat{\mathbf{f}}_0(\omega)$,
- (iv) repeat this process until changes are below a predetermined threshold.

By the power method applied to the time reversal matrix $\widehat{P}(\omega) \widehat{P}^*(\omega)$, the sequence of vectors $\hat{\mathbf{f}}_n(\omega)/\|\hat{\mathbf{f}}_n(\omega)\|$ converges to its top eigenvector $\widehat{U}_1(\omega)$ (if the time reversal matrix has a top eigenvalue of algebraic multiplicity one).

The next eigenvector of $\widehat{P}(\omega) \widehat{P}^*(\omega)$ can be found by repeating this process with a cancellation filter that eliminates the signal coming from the already detected scatterer. In general, to estimate the k^{th} eigenvector once the $k - 1$ previous ones have been found, we define the cancellation operator

$$\mathcal{C}_k(\hat{\mathbf{f}}) = \hat{\mathbf{f}} - \sum_{j=1}^{k-1} (\hat{\mathbf{f}}^*(\omega) \widehat{U}_j(\omega)) \widehat{U}_j(\omega),$$

and we apply it to all the illumination vectors sent to the medium during the iterative time reversal process so the components along directions $\widehat{U}_j(\omega)$, $j = 1, \dots, k - 1$, are eliminated. Moreover, because the response matrix is symmetric, the right singular vectors $\widehat{V}_j(\omega)$ are obtained by taking the complex conjugate of the left singular vectors $\widehat{U}_j(\omega)$. The process ends at a k for which the received power $\|\mathcal{C}_k(\hat{\mathbf{f}})\|^2$ is below a noise threshold.

Iterative time reversal is a very efficient acquisition method for obtaining the essential part of the array response matrix $\widehat{P}(\omega)$. It does not require any prior knowledge of the medium or the number of scatterers. Since iterative time reversal does not depend on the number of array transducers N , it is more efficient than using the SVD of the full array response matrix for large arrays. This is because the essential information for imaging with MUSIC or the hybrid ℓ_1 method is contained in the top singular vectors of response matrix.

References

- [1] Baraniuk R and Steeghs P 2007 Compressive radar imaging *IEEE Radar Conference*
- [2] Beck A and Teboulle M 2009 A Fast Iterative Shrinkage-Thresholding Algorithm for Linear Inverse Problems *SIAM J. Imag. Sci.* **2**(1) 183–202
- [3] Borcea L, Papanicolaou G, Tsogka C and Berryman J 2002 Imaging and time reversal in random media *Inverse Problems* **18** 1247–1279
- [4] Borcea L, Papanicolaou G and Tsogka C 2003 Theory and application of time reversal and interferometric imaging *Inverse Problems* **19** 139–164
- [5] Borcea L, Papanicolaou G and Tsogka C 2007 Optimal Waveform Design for Array Imaging *Inverse Problems* **23** 1973–2021
- [6] Borcea L, Papanicolaou G and Vasquez F 2008 Edge illumination and imaging of extended reflectors *SIAM J. Imag. Sci.* **1**(1) 75–114
- [7] Bredies K and Lorenz D A Linear convergence of iterative soft-thresholding *J. Fourier Anal. Appl.* **14** 813–837

- [8] Bruckstein A, Donoho D and Elad M 2009 From sparse solutions of systems of equations to sparse modeling of signals and images *SIAM Review* **51**(1) 34–81
- [9] Candès E 2006 Compressive sampling *Proc. Int. Congress of Mathematicians* Madrid, Spain
- [10] Candès E and Tao T 2005 Decoding by linear programming *IEEE Trans. Inform. Theory* **51**(12) 4203–4215
- [11] Chai A, Moscoso M and Papanicolaou G 2011 Array imaging using intensity-only measurements *Inverse Problems* **27** 015005 doi: 10.1088/0266-5611/27/1/015005
- [12] Chen S, Donoho D and Saunders M 1998 Atomic decomposition by basis pursuit *SIAM J. Sci. Comput.* **20**(6) 33–61
- [13] Chen J and Hou X 2006 Theoretical results on sparse representations of multiple measurement vectors *IEEE Trans. Signal Process.* **54**(12) 4634–4643
- [14] Dantzig G B 1963 Linear Programming and Extensions Princeton, NJ: Princeton University Press
- [15] Dassios G and Kleinman R 2000 Low Frequency Scattering Oxford University Press
- [16] Devaney A J, Marengo E A, and Gruber F K 2005 Time-reversal-based imaging and inverse scattering of multiply scattering point targets *J. Acoust. Soc. Am* **118**(5) 3129–3138
- [17] Fannjiang A, Strohmer T and Yan P 2010 Compressed remote sensing of sparse objects *SIAM J. Imaging Sci.* **3** 596–618
- [18] Fannjiang A 2010 Compressive inverse scattering I. High-frequency SIMO/MISO and MIMO measurements *Inverse Problems* **26** 035008
- [19] Fannjiang A 2011 Exact localization and superresolution with noisy data and random illumination *Inverse Problems* **27** 065012
- [20] Fannjiang A 2011 The MUSIC algorithm for sparse objects: a compressed sensing analysis *Inverse Problems* **27** 035013
- [21] Figueiredo M A T and Nowak R D 2003 An EM algorithm for wavelet-based image restoration *IEEE Trans. Image Process.* **12** (2003) 906–916
- [22] Golub G and Van Loan C F 1996 Matrix Computations, third ed. Johns Hopkins Univ. Press
- [23] Gurbuz A, McClellan J and Scott Jr W 2009 A compressive sensing data acquisition and imaging method for stepped frequency GPRs *IEEE Trans. Signal Processing* **57**(7) 2640–2650
- [24] Herman M and Strohmer T 2009 High resolution radar via compressed sensing *IEEE Trans. Signal Processing* **57**(6) 2275–2284
- [25] Hou S, Solna K, and Zhao H 2006 A direct imaging algorithm for extended targets *Inverse Problems* **22** 1151–1178
- [26] Kim S, Koh K, Lustig M, Boyd S, and Gorinvesky D 2007 An interior-point method for large-scale ℓ_1 -regularized least squares *IEEE J. Select. Topics Signal Proc.* **1**(4) 606–617
- [27] Kim J, Lee O and Ye J 2012 Compressive MUSIC: revisiting the link between compressive sensing and array signal processing *IEEE Trans. on Information Theory* **58**(1) 278–301
- [28] Lee K, Bresler Y and Junge M 2012 Subspace methods for joint sparse recovery *IEEE Trans. on Information Theory* **58**(6) 3613–3641
- [29] Malioutov D, Cetin M and Willsky A 2005 A sparse signal reconstruction perspective for source localization with sensor arrays *IEEE Trans. on Signal Processing* **53**(8) 3010–3022
- [30] Martin P A 2006 Multiple Scattering Interaction of Time-Harmonic Waves with N Obstacles Cambridge University Press
- [31] Montaldo G, Tanter M, and Fink M 2004 Real time inverse filter focusing through iterative time reversal *J. Acoust. Soc. Am.* **115**(2) 768–775
- [32] Montaldo G, Tanter M, and Fink M 2004 Revisiting iterative time reversal processing: Application to detection of multiple targets *J. Acoust. Soc. Am.* **115**(2) 776–784
- [33] Moscoso M, Novikov A, Papanicolaou G and Ryzhik L 2012 A differential equations approach to ℓ_1 -minimization with applications to array imaging *Inverse Problems* **28** 105001
- [34] Natarajan B K 1995 Sparse approximate solutions to linear systems *SIAM J. Comput.* **24**(2) 227–234

- [35] Nesterov Y E and Nemirovsky A 1994 Interior-Point Polynomial Methods in Convex Programming Philadelphia, PA: SIAM
- [36] Prada C, Thomas J, and Fink M 1995 The iterative time reversal process: Analysis of the convergence *J. Acoust. Soc. Am.* **97**(1) 62–71
- [37] Schmidt R 1986 Multiple emitter location and signal parameter estimation *IEEE Trans. Antennas Propagation* **34**(3) 276–280
- [38] Stoica P and Babu P 2012 SPICE and LIKES: Two hyperparameter-free methods for sparse-parameter estimation *Signal Processing* **92** 1580–1590
- [39] Toh K, Todd M and Tutuncu R 1999 SDPT3—a Matlab software package for semidefinite programming *Optimization Methods and Software* **11** 545–581
- [40] Wipf D and Rao B 2007 An empirical Bayesian strategy for solving the simultaneous sparse approximation problem *IEEE Trans. on Signal Processing* **55**(7) 3704–3716
- [41] Yin W, Osher S, Goldfarb D and Darbon J 2008 Bregman iterative algorithms for ℓ_1 -minimization with applications to compressed sensing *SIAM J. Imaging Sci.* **1**(1) 143–168